



①9



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

①1 Número de publicación: **2 284 475**

⑤1 Int. Cl.:
G10L 21/02 (2006.01)

①2

TRADUCCIÓN DE PATENTE EUROPEA

T3

⑧6 Número de solicitud europea: **00902355 .7**

⑧6 Fecha de presentación : **07.01.2000**

⑧7 Número de publicación de la solicitud: **1141948**

⑧7 Fecha de publicación de la solicitud: **10.10.2001**

⑤4 Título: **Método y aparato para la supresión del ruido de manera adaptativa.**

③0 Prioridad: **07.01.1999 US 115245 P**

④5 Fecha de publicación de la mención BOPI:
16.11.2007

④5 Fecha de la publicación del folleto de la patente:
16.11.2007

⑦3 Titular/es: **Tellabs Operations, Inc.**
1415 West Diehl Road, MS 16
Naperville, Illinois 60563, US

⑦2 Inventor/es: **Chandran, Ravi;**
Dunne, Bruce, E. y
Marchok, Daniel, J.

⑦4 Agente: **Morgades Manonelles, Juan Antonio**

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 la supresión del ruido de manera adaptativa.

5 La presente invención se refiere a la supresión de ruidos en los sistemas de telecomunicaciones. En particular, la presente invención se refiere a eliminar ruido en sistemas de un solo canal o en canales únicos en sistemas de canales múltiples.

10 La mejora de la calidad de la voz es una función importante en sistemas de comunicaciones por voz. Los teléfonos móviles, por ejemplo, se utilizan con frecuencia en presencia de niveles elevados de ruido ambiental de fondo presente en los vehículos en marcha. El ruido de fondo provoca una degradación significativa de la calidad de la voz en el receptor, haciendo que la voz sea prácticamente ininteligible. En tales circunstancias, pueden emplearse técnicas de mejora de la voz para mejorar la calidad de la voz recibida, aumentando así la satisfacción del cliente y facilitando tiempos de conversación más largos.

15 Los sistemas de supresión del ruido antiguos utilizaban normalmente algunas variaciones de la sustracción espectral. En la Figura 1 se muestra un ejemplo de un sistema de supresión del ruido 100 que usa la sustracción espectral. En primer lugar, se efectúa una descomposición espectral de la señal de voz con ruido de entrada 102 usando la batería de filtros 104. La batería de filtros 104 puede ser una batería de filtros de paso de banda tal como, por ejemplo, los
20 filtros de paso de banda descritos en R. J. McAulay y M. L. Malpass, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, n° 2, (Abr. de 1980), págs. 137-145.

25 En este contexto, el ruido se refiere a cualquier señal no deseada presente en la señal de voz que incluye: 1) ruido ambiental de fondo; 2) reverberación como la debida a reflexiones acústicas o reflexiones eléctricas en híbridos; 3) ruido mecánico y/o eléctrico añadido debido a equipos específicas tales como ruido de fondo de cinta en un sistema de reproducción de voz; y 3) no linealidades debidas, por ejemplo, a recortes de la señal o a cuantificación por compresión de la voz.

30 La batería de filtros 104 descompone la señal en diferentes bandas de frecuencia. Para cada banda, se realizan mediciones de la potencia que se actualizan continuamente en el estimador de potencia de la señal y de potencia del ruido 106. Estas mediciones de la potencia se usan para determinar la relación señal-ruido (SNR) de cada banda. El detector de actividad de voz 108 se usa para distinguir periodos de actividad de habla de periodos de silencio. La potencia de ruido de cada banda de frecuencia se actualiza solo durante los silencios mientras que la potencia de la
35 señal de ruido se rastrea en todo momento. Para cada banda de frecuencia, se calcula un factor de ganancia (atenuación) en el controlador de ganancia 110 basado en la SNR de la banda para atenuar la señal del multiplicador de ganancia 112. Así, cada banda de frecuencia de la señal de voz ruidosa de entrada se atenúa según esa SNR. En este contexto, la señal de voz se refiere a una señal de audio que puede contener voz, música u otras señales de audio que transportan información (por ej., tonos DTMF, pausas silenciosas y ruido).

40 Un enfoque más sofisticado también puede usar un nivel de SNR general además de los valores SNR individuales para calcular los factores de ganancia de cada banda. La SNR general se calcula en el estimador de la SNR general 114. Los cálculos del factor de ganancia de cada banda se realizan en el controlador de ganancia 110. La atenuación de las señales de diferentes bandas se consigue multiplicando la señal de cada banda por el factor de ganancia correspondiente del multiplicador de ganancia. Las bandas de SNR baja se atenúan más que las bandas de SNR elevada. La cantidad de
45 atenuación es también mayor si la SNR general es baja. El intervalo dinámico posible de la SNR de la señal de entrada es grande. Como tal, el sistema de mejora de voz debe ser capaz de manipular tanto señales de voz muy limpias de teléfonos de la red fija así como de voz con mucho ruido de teléfonos móviles. Después del proceso de atenuación, las señales de las diferentes bandas se recombinan en una sola señal de salida limpia 116. La señal de salida 116 resultante
50 tendrá una calidad general percibida mejorada.

55 En este contexto, el sistema de mejora de la voz se refiere a un aparato o dispositivo que mejora la calidad de la señal de voz en lo que se refiere a la percepción humana o en lo que se refiere a otros criterios tales como la precisión del reconocimiento por un dispositivo de reconocimiento de voz, suprimiendo, enmascarando, cancelando o eliminado el ruido o eliminando de otro modo los efectos adversos del ruido. Los sistemas de mejora de la voz incluyen aparatos o dispositivos que modifican una señal de entrada de formas tales como por ejemplo: 1) generar una señal de voz de ancho de banda más ancho a partir de una señal de voz con un ancho de banda estrecho; 2) separar una señal de entrada en diversas señales de salida basadas en determinados criterios, por ej., separación de la voz de diferentes hablantes en donde una señal contiene una combinación de las señales de voz de los hablantes; 3) y procesar (por ejemplo
60 mediante conversión) diferentes "porciones" de una señal de entrada por separado y/o de forma diferente, en donde una "porción" puede ser una porción de la señal de entrada en el tiempo (por ejemplo, en sistemas de teléfono con altavoz) o bien pueden incluir bandas de frecuencia determinadas (por ej., sistemas de audio que refuerzan la base), o ambos.

65 La descomposición de la señal ruidosa que contiene voz de entrada también puede realizarse usando técnicas de transformada de Fourier o técnicas de transformada de ondícula. En la Figura 2 se muestra el uso de técnicas de transformada discreta de Fourier (mostradas como el bloque Ventanas y FFT 202). Aquí un bloque de muestras de entrada se transforma en el dominio de frecuencia. La magnitud de los complejos elementos del dominio de frecuencia

se atenúan en la unidad de atenuación 208 en base al principio de sustracción espectral descrito anteriormente. La fase de los elementos complejos de dominio de frecuencia se deja sin cambios. Los complejos elementos del dominio de frecuencia se vuelven a transformar entonces en el dominio de tiempo a través de una transformada discreta de Fourier inversa en el bloque IFFT 204, produciendo la señal de salida 206. En lugar de técnicas de transformada de Fourier, pueden usarse técnicas de transformada de ondulica para descomponer la señal de entrada.

Puede usarse un detector de actividad de voz con sistemas de supresión del ruido. Dicho detector de actividad de voz se presenta, por ejemplo, en la patente estadounidense núm. 4.351.983 de Crouse *et al.* En tales detectores, la potencia de la señal de entrada se compara con un nivel umbral variable. Siempre que se sobrepase el umbral, el sistema asume que hay voz presente. De lo contrario, se asume que la señal contiene solo ruido de fondo.

En la mayoría de aplicaciones de mejora de la voz, es deseable minimizar el retardo en el procesamiento. De ese modo, el uso de técnicas de transformada de Fourier o de ondulica para la descomposición espectral no es deseable debido a que dichas técnicas introducen grandes retardos al acumular un bloque de muestras para su procesamiento.

También es deseable una complejidad computacional baja ya que el sistema de supresión de ruido de la red puede procesar simultáneamente múltiples canales de voz independientes. Además, se prefiere limitar los tipos de cálculos para la adición, sustracción y multiplicación para facilitar una implantación directa de accesorios digitales así como para minimizar el procesamiento en una implantación basada en procesador de señal digital de punto fijo. La división es computacionalmente intensiva en procesadores de señal digital y también resulta molesto para la implantación directa de equipos digitales. Finalmente, los requisitos de almacenamiento en memoria de cada canal deben minimizarse debido a la necesidad de procesar simultáneamente múltiples canales de voz independientes.

Las técnicas de mejora de la voz, también debe tratar tonos de información tales como tonos DTMF (multifrecuencia de tono doble). Los tonos DTMF los generan normalmente teléfonos de teclas/marcación por tonos cuando se pulsa cualquiera de los botones. El teclado ampliado del teléfono por tonos posee 16 teclas: (1, 2, 3, 4, 5, 6, 7, 8, 9, 0, *, #, A, B, C, D). Las teclas están dispuestas en una estructura de cuatro por cuatro. Al pulsar una de las teclas se hace que un circuito eléctrico genere dos tonos. Como se muestra en la Tabla 1, existe un tono de baja frecuencia para cada hilera y un tono de alta frecuencia para cada columna. Así, las frecuencias de hilera se refieren como el Grupo Bajo y las frecuencias de columna, el Grupo Alto. De este modo, pueden generarse dieciséis combinaciones únicas de tonos usando solo ocho tonos únicos. En la Tabla 1 se muestran las teclas y las frecuencias nominales correspondientes. (A pesar de que se ha descrito en relación con tonos DTMF, los principios que se describen en relación con la presente invención son aplicables a todas las señales en banda. En este contexto, una señal en banda se refiere a cualquier tipo de señal tonal dentro del ancho de banda usado normalmente para la transmisión de voz como, por ejemplo, tonos de fax, tonos de marcado, tonos de señal ocupada, y tonos DTMF).

TABLA 1

Frecuencias de hilera de teclas de teléfono por tonos (Grupo Bajo) y de columna (Grupo Alto)

Bajo \ Alto (Hz)	1209	1336	1477	1633
697	1	2	3	A
770	4	5	6	B
852	7	8	9	C
941	*	0	#	D

Los tonos DTMF tienen normalmente una duración inferior a 100 milisegundos (ms) y pueden ser tan breves como 45 ms. Dichos tonos pueden transmitirse durante las llamadas telefónicas a sistemas de contestador automático de diversos tipos. Dichos tonos se generan por un circuito DTMF independiente cuya salida se añade a la señal de voz procesada antes de su transmisión.

Por lo general, las señales DTMF pueden transmitirse con una velocidad máxima de diez dígitos/segundo. En esta velocidad máxima, para cada intervalo de tiempo de 100 ms, el generador de tono doble debe generar señales de marcado por tonos con una duración de al menos 45 ms y no superior a 55 ms, y a continuación permanecer en silencio durante el resto del intervalo de tiempo. Cuando no se transmite a la velocidad máxima, un par de tonos puede tener una duración cualquiera, pero cada par de tonos debe estar separado del siguiente par por al menos 40 ms.

Sin embargo, en los sistemas de mejora de voz antiguos, con frecuencia los tonos DTMF se suprimían parcialmente. La supresión de tonos DTMF se producía debido a que detectores de actividad de voz y/o detectores de tonos DTMF requieren cierto retardo antes de poder determinar la presencia de una señal. Cuando se detectaba la presencia de una señal, se producía un tiempo de demora antes de que los factores de ganancia de las bandas de frecuencia adecuadas alcanzaran sus valores correctos (altos). Frecuentemente, este tiempo de reacción hacía que la parte inicial de los

tonos se suprimiera considerablemente. Así, tonos DTMF de corta duración pueden acortarse todavía más mediante el sistema de mejora de la voz.

En la Figura 7 se muestra una señal de entrada 702 que contiene un tono de 697 Hz 704 con una duración de 45 ms (360 muestras). Inicialmente, la señal de salida 703 se suprime considerablemente, hasta que el detector de actividad de voz detecta la presencia de señal. A continuación, el factor de ganancia 708 aumenta gradualmente para evitar la atenuación.

Así, el resultado es una versión abreviada del tono de entrada, que en este ejemplo, no cumple los requisitos generales mínimos de duración de tonos DTMF.

Como resultado del recorte de los tonos DTMF, el receptor puede no detectar correctamente los tonos DTMF debido a que los tonos no cumplen los requisitos mínimos de duración. Como puede observarse en la Figura 7, el factor de ganancia 708 nunca alcanza su valor máximo de unidad debido a que depende de la SNR de la banda. Esto hace que la señal de salida 706 siempre se atenúe ligeramente, lo cual puede ser suficiente para evitar que la potencia de la señal cumpla el umbral de receptor del detector DTMF. Además, los factores de ganancia de diferentes bandas de frecuencia pueden ser suficientemente diferentes como para aumentar la diferencia en las amplitudes de los tonos dobles. Esto además aumenta la probabilidad de que el receptor no detecte correctamente los tonos DTMF.

Los inconvenientes descritos anteriormente estaban presentes en sistemas antiguos de supresión de ruido. El sistema descrito, por ejemplo, en las patentes estadounidenses núm. 4.628.529, 4.630.304, y 4.630.305 de Borth *et al.*, estaba diseñado para operar en entornos de ruido de fondo elevado. Sin embargo, se prefiere el funcionamiento bajo un amplio intervalo de condiciones de SNR. Además, la división de software se usa en los métodos de Borth.

Las operaciones de división computacionalmente intensivas también se usan en la patente estadounidense núm. 4.454.609 de Kates. La utilización de estimadores de amplitud logarítmica espectral de error cuadrático medio mínimo tales como los descritos en la patente estadounidense núm. 5.012.519 de Adlersberg *et al.* también son computacionalmente intensivos. Además, el sistema descrito por Adlesberg usa la transformada de Fourier para la descomposición espectral que introduce un retardo no deseable. Además, a pesar de que se presenta un generador de tonos DTMF en el informe de aplicación de Texas Instruments, "DTMF Tone Generation and Detection: An Implementation Using the TMS320C54x", "1997, págs. 5-12, 20, A-1, A-2, B-1, B-2, no existe ningún sistema que amplíe y/o regenere tonos DTMF suprimidos.

Desde hace tiempo el sector tiene la necesidad de contar con un sistema de supresión del ruido con una complejidad computacional baja.

Además, desde hace tiempo el sector tiene la necesidad de contar con un sistema de supresión del ruido capaz de ampliar y/o regenerar tonos DTMF parcialmente suprimidos.

La presente invención, según se define en las reivindicaciones anexas, se centra en un aparato para mejorar la calidad de una señal de comunicaciones en un sistema de comunicaciones. Comprende un procesador adaptado para efectuar los pasos de dividir la señal de comunicación en una pluralidad de señales de banda de frecuencia; generando una primera señal de potencia para una primera señal de banda de frecuencia, la cual se basa en estimar durante un primer periodo de tiempo la potencia de la primera señal de banda de frecuencia;

generando una segunda señal de potencia para la primera señal de banda de frecuencia, la cual se basa en estimar durante un segundo periodo de tiempo inferior al primer periodo de tiempo la potencia de la primera señal de banda de frecuencia;

generando una señal de condición que representa una condición de la primera señal de banda de frecuencia en respuesta a relaciones predeterminadas entre la primera y la segunda señal de potencia; ajustar la ganancia de la señal de banda de frecuencia en respuesta a la señal de condición para generar una señal de banda de frecuencia ajustada; y combinar la señal de banda de frecuencia ajustada con al menos una señal de banda de frecuencia adicional para generar una señal de comunicación ajustada. La señal de condición es una relación ruido-senal (NSR) generada en respuesta a una relación aritmética predeterminada que comprende al menos una adición, sustracción y multiplicación pero sin división, entre los valores derivados de la primera y segunda señal de potencia.

La invención también comprende un aparato para mejorar una señal de comunicación de un sistema de comunicaciones, que comprende un procesador adaptado para realizar el método anteriormente descrito. Pueden adaptarse aparatos y métodos de la invención a tonos de información del proceso tales como tonos DTMF;

a tonos DTMF parcialmente suprimidos; y para regenerar tonos DTMF parcialmente suprimidos.

Los aparatos según la presente invención pueden utilizar una batería de filtros de paso de banda para dividir la señal ruidosa de entrada que contiene voz en bandas de frecuencia independientes. Para determinar si la señal de entrada contiene voz, tonos DTMF o silencio, puede usarse un detector conjunto de actividad de voz y actividad DTMF (JVADAD).

En la práctica de la invención, la relación ruido- señal general media (NSR) de la señal de entrada se calcula normalmente a partir de estimaciones de la potencia media de la señal ruidosa en la señal de entrada durante la actividad de voz y la potencia media de ruido durante el silencio. En lugar de medir directamente la señal de ruido y la potencia de ruido de cada banda de frecuencia como se realiza normalmente en sistemas de supresión del ruido, se realizan dos mediciones indirectas de la potencia de cada banda. Dichas medidas de potencia se denominan potencia a largo plazo y potencia a corto plazo. Estas mediciones se realizan en el estimador de potencia a largo plazo y a corto plazo. La potencia a largo plazo es una versión a escala de la potencia de ruido de la banda. La potencia de ruido a corto plazo es una versión a escala de la potencia de la señal ruidosa de la banda. Estas medidas de potencia a escala pueden usarse para reducir el intervalo dinámico necesario para una implantación de punto fijo. Esto tiene como resultado un rendimiento superior de la supresión de ruido que se aproxima al de una implantación de punto flotante. Los procesos de estimación de la potencia se adaptan en función de la actividad de la señal indicada por el JVADAD. El número de cálculos necesarios para la medición de la potencia se reduce significativamente por el submuestreo de las señales de cada banda de frecuencia antes de medir la potencia.

Puede usarse un adaptador de NSR para adaptar la NSR de cada banda de frecuencia en función de las medidas de potencia a largo y corto plazo, indicando el JVADAD, la NSR general y la actividad de la señal. La adaptación de la NSR se realiza sin división usando un error de predicción calculado como función de las mediciones a corto y largo plazo y NSR general. Un controlador de ganancia utiliza dichos valores de NSR para determinar los factores de ganancia de cada banda de frecuencia. Entonces el multiplicador de ganancia puede realizar la atenuación de cada banda de frecuencia. Finalmente, las señales procesadas en las bandas de frecuencia separadas se suman en el combinador para producir la señal de salida limpia.

El método anteriormente descrito para adaptar los valores NSR durante la conversación es diferente del usado en presencia de tonos DTMF. Para los tonos DTMF, el ajuste rápido de los valores NSR para las bandas de frecuencia adecuadas que contienen los tonos DTMF maximiza la cantidad de tonos DTMF que pasan a través de forma transparente. En el caso de la voz, los valores NSR se adaptan preferiblemente con mayor lentitud para corresponderse a la naturaleza de las señales de voz.

Otras características y ventajas de la presente invención resultarán evidentes a partir de la siguiente descripción detallada, las cuales deberán considerarse junto con los dibujos que la acompañan, en donde:

En la Figura 1 se presenta un diagrama sinóptico de un sistema de supresión del ruido típico.

En la Figura 2 se presenta un diagrama sinóptico de otro sistema de supresión del ruido típico.

En la Figura 3 se presenta un diagrama sinóptico de un aparato de supresión de ruido según una realización de la presente invención.

En la Figura 4 se presenta un diagrama sinóptico de un aparato para determinar la NSR en un aparato según la presente invención.

En la Figura 5 se presenta un diagrama de flujo en el que se ilustra un método para ampliar los tonos DTMF en un aparato según la presente invención.

En la Figura 6 se presenta un diagrama de flujo en el que se ilustra un método para regenerar los tonos DTMF en un aparato según la presente invención.

En la Figura 7 se presentan unos gráficos que ilustran la supresión de los tonos DTMF en los sistemas de mejora de la voz.

En la Figura 8 se presentan unos gráficos que ilustran la ampliación en tiempo real de los tonos DTMF.

En la Figura 9 se presenta un diagrama sinóptico de un detector conjunto de actividad de voz y actividad DTMF en un aparato según la presente invención.

Volviendo ahora a la Figura 3, en cuya Figura se presenta un diagrama sinóptico de un aparato de supresión de ruido 300. Se presenta una batería de filtros 302, un detector de actividad de voz 304, un contador de persistencia 305, y un estimador de NSR (relación ruido-senal) general 306. También están presentes un estimador de potencia 308, un adaptador de NSR 310, un controlador de ganancia 314 y un combinador 315. La realización que se ilustra en la Figura 3 también presenta una señal de entrada $x(n)$ 316 y señales de salida $x_k(n)$ 318, una detección conjunta de actividad de voz y señal de detección de actividad DTMF 320. En la Figura 3 también se presenta un generador de tono DTMF 321. El resultado del estimador NSR general 306 es la NSR general (" $NSR_{general}(n)$ ") 322.

Las estimaciones de potencia 323 son el resultado del estimador de potencia 308. Los valores NSR adaptados 324 se producen por el adaptador de NSR 310. Los factores de ganancia 326 se producen por el controlador de ganancia 312. Las señales atenuadas 328 se producen del multiplicador de ganancia 314. Los tonos DTMF 329 se producen por el generador de tonos DTMF 321.

En la Figura 3 también se ilustra que el estimador de potencia 308 puede incluir opcionalmente un circuito de submuestreo 330 y que el estimador de potencia 308 puede pasar opcionalmente los cálculos de potencia 323 al controlador de ganancia 312.

En la realización ilustrada de la Figura 3, la batería de filtros 302 recibe la señal de entrada 316. La tasa de muestreo de la señal de voz en, por ejemplo, aplicaciones de telefonía es normalmente 8 kHz con un ancho de banda Nyquist de 4 kHz. Puesto que el canal de transmisión posee normalmente un intervalo de 300-3400 Hz, la batería de filtros 302 puede estar diseñada para pasar solo señales dentro de dicho intervalo. Como ejemplo, la batería de filtros 302 puede utilizar una batería de filtros de paso de banda. Puede usarse una batería de filtros 302 de velocidades múltiples o de una velocidad. Una implantación de la batería de filtros 302 de una velocidad utiliza la estructura de filtro de muestreo de frecuencia (FSF). La realización preferida usa una batería de resonadores formada por una serie de filtros de respuesta de impulso infinito ("IIR") de bajo orden. Esta batería de resonadores puede considerarse una versión modificada de la estructura FSF y posee diversas ventajas con respecto a la estructura FSF. La batería de resonadores no requiere el filtro de memoria intensiva de la estructura FSF y, como resultado de ello, requiere menos cálculos. También se elimina la utilización de signos alternantes en la estructura FSF, lo cual tiene como resultado una complejidad computacional reducida.

La función de transferencia del k^{th} resonador puede obtenerse, por ejemplo:

$$H_k(z) = \frac{g_k[1 - r_k \cos(\theta_k)z^{-1}]}{[1 - 2r_k \cos(\theta_k)z^{-1} + r_k^2 z^{-2}]} \quad (1)$$

En la ecuación (1), la frecuencia central de cada resonador se especifica a través de θ_k . El ancho de banda del resonador se especifica a través de r_k . El valor de g_k se usa para ajustar la ganancia DC de cada resonador. Para una batería de resonadores formada por 40 resonadores que abarcan aproximadamente el intervalo de 300-3400 Hz, las siguientes son especificaciones adecuadas para las funciones de transferencia del resonador siendo $k = 3, 4, \dots, 42$:

$$r_k = 0.965 \quad (2a)$$

$$\theta_k = \frac{2\pi k}{100} \quad (2b)$$

$$g_k = 0.01 \quad (2c)$$

La entrada a la batería de resonadores se denomina $x(n)$ mientras que la salida del resonador k^{th} se denomina $x_k(n)$, en donde n es el tiempo de muestra.

El factor de ganancia 326 de la banda de frecuencia k^{th} puede calcularse una vez cada T muestras como:

$$G_k(n) = \begin{cases} 1 - NSR_k(n) & , \quad n = 0, T, 2T, \dots \\ G_k(n-1) & , \quad n = 1, 2, \dots, T-1, T+1, \dots, 2T-1, \dots \end{cases} \quad (3)$$

Cuando el factor de ganancia 326 de cada banda de frecuencia se calcula una vez cada T muestras, la ganancia está "submuestreada" puesto que no se calcula en cada muestra. (Tal y como se indica mediante las líneas discontinuas de las Figuras 1-4, el dispositivo pertinente puede producir diversos elementos diferentes de datos, por ejemplo factores de ganancia 326. Las diversas salidas se corresponden preferiblemente a las diversas subbandas en las que se divide la señal de entrada 316. El factor de ganancia oscilará entre un pequeño valor positivo, ε y 1 debido a que los valores NSR están limitados a permanecer dentro del intervalo $[0, 1-\varepsilon]$. Ajustar el límite inferior de la ganancia en ε se reducen los efectos del "ruido musical" y permite una transparencia limitada de la señal de fondo.

La atenuación de la señal $x_k(n)$ de la banda de frecuencia k^{th} se consigue multiplicando $x_k(n)$ por su factor de ganancia correspondiente, $G_k(n)$, en cada muestra. La suma de las señales atenuadas resultantes, $y(n)$, es la señal de salida limpia 328. La suma de las señales atenuadas 328 puede expresarse matemáticamente como:

$$y(n) = \sum_k G_k(n)x_k(n) \quad (4)$$

Las señales atenuadas 328 también pueden estar a escala, por ejemplo, pueden incrementarse o amplificarse para su posterior transmisión.

La potencia, $P(n)$ en la muestra n , de una señal de tiempo discreto $u(n)$, se calcula aproximadamente mediante filtrado de paso bajo de la señal rectificada de onda completa. Puede usarse un filtro IIR de primer orden para el filtrado de paso bajo, tal como, por ejemplo:

$$P(n) = \beta P(n-1) + \alpha |u(n)| \quad (5)$$

Este filtro IIR posee la siguiente función de transferencia:

$$H(z) = \frac{\alpha}{1 - \beta z^{-1}} \quad (6)$$

La ganancia DC de este filtro es $H(1) = \frac{\alpha}{1 - \beta}$.

Se hace referencia al coeficiente, β , como constante de decaimiento.

El valor de la constante de decaimiento determina el tiempo que es necesario para que el valor actual (no cero) de la potencia decaiga hasta una pequeña fracción del valor actual si la entrada es cero, es decir, $u(n) = 0$. Si la constante de decaimiento, β , está cerca de la unidad será necesario un tiempo relativamente largo para que el valor de potencia decaiga. Si β está próximo a cero, entonces será necesario un tiempo relativamente corto para que el valor de potencia decaiga. Así, la constante de decaimiento también representa la rapidez con la que se olvida el valor de potencia antiguo y la rapidez con la que se incorpora la potencia de las muestras de entrada más recientes. Así, unos valores mayores de β tienen como resultado una ventana de promediado efectivo más larga. En este contexto, los cálculos de potencia usando una ventana de promediado efectivo relativamente larga son cálculos de potencia a largo plazo, mientras que los cálculos de potencia usando una ventana de promediado efectivo relativamente corta son cálculos de la potencia a corto plazo.

En función de la señal de interés, será adecuado un promediado más largo o más corto para el cálculo de la potencia. La potencia de la voz, que posee un perfil que cambia rápidamente, podría calcularse adecuadamente usando un β más pequeño. Puede considerarse que el ruido es fijo para periodos de tiempo más largos que la voz. Por lo tanto, se calculará preferiblemente la potencia del ruido con mayor precisión usando una ventana de promediado más larga (gran β).

La realización preferida para el cálculo de la potencia reduce significativamente la complejidad computacional al submuestrear la señal de entrada con objeto del cálculo de la potencia. Esto significa que solo se usa una de cada T muestras para actualizar la potencia $P(n)$. Entre dichas actualizaciones, el cálculo de potencia se mantiene constante. Este procedimiento puede expresarse matemáticamente como

$$P(n) = \begin{cases} \beta P(n-1) + \alpha |u(n)| & , \quad n = 0, 2T, 3T, \dots \\ P(n-1) & , \quad n = 1, 2, \dots, T-1, T+1, \dots, 2T-1, \dots \end{cases} \quad (7)$$

El filtro IIR de paso bajo de primer orden se usa preferiblemente para el cálculo de la potencia del ruido de fondo media general, y una medición de la potencia a corto y largo plazo para cada banda de frecuencia. También se usa preferiblemente para mediciones de potencia en el VAD 304. El submuestreo puede conseguirse a través del uso de, por ejemplo, un circuito de submuestreo 330 conectado al estimador de potencia 308.

La SNR general (“SNR_{general}(n)”) en la muestra n se define como:

$$SNR_{general}(n) = \frac{P_{SIG}(n)}{P_{BN}(n)} \quad (8)$$

en donde $P_{SIG}(n)$ y $P_{BN}(n)$ son la potencia de la señal ruidosa media durante la conversación y la potencia del ruido de fondo medio durante los silencios, respectivamente. La SNR general se usa para influir en la cantidad de eliminación excesiva de la señal en cada banda de frecuencia. La eliminación excesiva mejora la calidad de la voz percibida, especialmente bajo condiciones de SNR general baja. La eliminación excesiva de la señal se consigue usando el valor

ES 2 284 475 T3

de SNR general para influir en el adaptador de NSR 310. Además, la supresión insuficiente en el caso de condiciones de SNR general elevada puede usarse para evitar la atenuación innecesaria de la señal. De este modo se evita la distorsión de la voz bajo condiciones de SNR elevada en donde el ruido de bajo nivel queda enmascarado efectivamente por la voz. Los detalles de la supresión excesiva y supresión insuficiente se describen a continuación.

La potencia de la señal ruidosa media se calcula preferiblemente durante la actividad de voz, tal y como se indica mediante el VAD 304, según la fórmula:

$$P_{SIG}(n) = \begin{cases} \beta_{SIG} P_{SIG}(n-1) + \alpha_{SIG} |x(n)|, & n = 0, 2T, 3T, \dots \\ P_{SIG}(n-1), & n = 1, 2, \dots, T-1, T+1, \dots, 2T-1, \dots \end{cases} \quad (9a)$$

en donde $x(n)$ es la señal de entrada con ruido que contiene la voz.

La potencia del ruido de fondo medio se calcula preferiblemente según la fórmula:

$$P_{BN}(n) = \begin{cases} \max[\beta_{BN} P_{BN}(n-1) + \alpha_{BN} |x(n)|, P_{BN,max}], & n = 0, 2T, 3T, \dots \\ P_{BN}(n-1), & n = 1, 2, \dots, T-1, T+1, \dots, 2T-1, \dots \end{cases} \quad (9b)$$

en donde no se permite que $P_{BN}(n)$ sobrepase $P_{BN,max}(n)$.

Durante el silencio o actividad de tono DTMF, según indica el VAD 304, la medición de la potencia de la señal ruidosa media se mantiene preferiblemente constante, es decir:

$$P_{SIG}(n) = P_{SIG}(n-1). \quad (10a)$$

Durante la actividad de voz o de tono DTMF según indique el VAD, la medición de la potencia de la señal ruidosa media se mantiene preferiblemente constante, es decir,

$$P_{BN}(n) = P_{BN}(n-1) \quad (10b)$$

Si el intervalo de las muestras de entrada se normaliza en ± 1 , los valores adecuados para los parámetros constantes usados en la realización preferida son

$$P_{BN,max} = 180/8159 \quad (11a)$$

$$\alpha_{SIG} = \alpha_{BN} = T/16000 \quad (11b)$$

$$\beta_{SIG} = \beta_{BN} = 1 - T/16000 \quad (11c)$$

en donde $T = 10$ en un posible periodo de submuestreo.

El nivel de potencia del ruido de fondo medio está limitado preferiblemente a $P_{BN,max}$ por dos motivos. En primer lugar, $P_{BN,max}$ representa el peor caso típico de un escenario de ruido en telefonía móvil. En segundo lugar, $P_{SIG}(n)$ y $P_{BN}(n)$ se usarán en el adaptador NSR 310 para influir en el ajuste de la NSR de cada banda de frecuencia. Limitar $P_{BN}(n)$ proporciona un medio para controlar la cantidad de influencia que tiene la SNR general en el valor de la NSR de cada banda.

En la realización preferida, se calcula la NSR general 322 en lugar de la SNR general. La NSR general 322 es más adecuada para la adaptación de los valores de la NSR de la banda de frecuencia individual. Como un cálculo directo de la NSR general 322 implica una división computacionalmente intensiva de $P_{BN}(n)$ entre $P_{SIG}(n)$, la realización preferida usa un enfoque que proporciona una aproximación adecuada de la NSR general 322.

Además, la definición de la NSR se amplía para ser negativa para indicar niveles de NSR general 322 muy elevados del siguiente modo:

$$NSR_{overall}(n) = \begin{cases} \nu_1 P_{BN}(n) & , & P_{SIG}(n) < \kappa_1 P_{BN}(n) \\ \nu_2 P_{BN}(n) & , & P_{SIG}(n) \geq \kappa_2 P_{BN}(n) \\ \nu_3 [P_{BN}(n) - P_{SIG}(n)] & , & \kappa_2 P_{BN}(n) > P_{SIG}(n) \geq \kappa_3 P_{BN}(n) \end{cases} \quad (12a)$$

Una realización de la invención utiliza $\nu_1 = 2,9127$, $\nu_2 = 1,45635$, $\nu_3 = 0,128$, $\kappa_1 = 10$, $\kappa_2 = 14$ y $\kappa_3 = 20$. En este caso, el intervalo de $NSR_{general}(n)$ 322 es:

$$-0.128 \leq NSR_{general}(n) \leq 0.064. \quad (12b)$$

El límite superior de $NSR_{general}(n)$ 322 en esta realización se obtiene limitando $P_{BN}(n)$ para que sea como máximo $P_{BN,max}(n)$. El límite inferior obedece al hecho de $P_{BN}(n) - P_{SIG}(n) \geq -1$.

(Puesto que se supone que el intervalo de señal de entrada se normaliza en ± 1 , tanto $P_{BN}(n)$ como $P_{SIG}(n)$ siempre están entre 0 y 1).

La medición de la potencia a largo plazo, $P_{LT}^k(n)$ en la muestra n , para la banda de frecuencia k^{th} es proporcional al nivel de potencia de ruido actual en esa banda. Se trata de una versión amplificada del nivel actual de potencia del ruido. La cantidad de amplificación se predetermina para evitar o reducir el subdesbordamiento en una implantación de punto fijo del filtro IIR usado para el cálculo de la potencia. El subdesbordamiento puede producirse debido a que el intervalo dinámico de la señal de entrada en una banda de frecuencia durante el silencio es bajo. La potencia a largo plazo de la banda de frecuencia k^{th} se calcula preferiblemente solo durante el silencio según indique el VAD 304 usando el siguiente filtro IIR de paso bajo de primer orden:

$$P_{LT}^k(n) = \begin{cases} \beta_{LT} P_{LT}^k(n-1) + \alpha_{LT} |x_k(n)| & , & n = 0, 2T, 3T, \dots \\ P_{LT}^k(n-1) & , & n = 1, 2, \dots, T-1, T+1, \dots, 2T-1, \dots \end{cases} \quad (13)$$

En este caso, la potencia a largo plazo no podría actualizarse durante la actividad del tono DTMF o durante la actividad de voz. Sin embargo, a diferencia de la voz, la actividad del tono DTMF afecta solo a unas pocas bandas de frecuencia. Así, en una realización alternativa, los cálculos de potencia a largo plazo correspondientes a las bandas de frecuencia que no contienen los tonos DTMF se actualizan durante la actividad de tono DTMF. En esta realización, los cálculos de potencia a largo plazo de bandas de frecuencia que contienen los tonos DTMF se mantienen constantes, es decir:

$$P_{LT}^k(n) = P_{LT}^k(n-1). \quad (14)$$

Debe tenerse en cuenta que la medición de la potencia a largo plazo está preferiblemente submuestreada con un periodo T . Un periodo de submuestreo adecuado es $T = 10$ muestras. Un conjunto adecuado de coeficientes de filtro para la ecuación (13) es:

$$\alpha_{LT} = T/160 \quad (15a)$$

$$\beta_{LT} = 1 - T/16000 \quad (15b)$$

En esta realización, la ganancia DC del filtro de medición de la potencia a largo plazo es $H_{LT}(1) = 100$. Esta gran ganancia DC proporciona el incremento necesario para evitar o reducir la posibilidad de subdesbordamiento de la medición de la potencia a largo plazo.

El cálculo de la potencia a corto plazo utiliza una ventana de promediado más corta que el cálculo de la potencia a largo plazo. Si el cálculo de la potencia a corto plazo se realizó usando un filtro IIR con coeficientes fijos como en la ecuación (7), la potencia podría variar rápidamente para rastrear las variaciones de potencia de la señal durante la conversación. Durante el silencio, las variaciones se reducirían, aunque continuarían siendo más que las de la medición

de la potencia a largo plazo. Así, el intervalo dinámico necesario para esta medición de potencia podría ser elevado si se utilizan coeficientes fijos. Sin embargo, al hacer que el coeficiente numerador del filtro IIR sea proporcional a la NSR de la banda de frecuencia, se hace que la medición de la potencia rastree el nivel de potencia del ruido en la banda. La posibilidad de desbordamiento se reduce o se elimina, lo cual tiene como resultado una medición de la potencia más precisa.

La realización preferida usa un filtro IIR adaptable de primer orden para calcular la potencia a corto plazo $P_{ST}^k(n)$ en la banda de frecuencia k^{th} , una vez cada T muestras:

$$P_{ST}^k(n) = \begin{cases} \beta_{ST} P_{ST}^k(n-1) + \alpha_{ST} NSR_k(n) |x_k(n)|^2, & n = 0, 2T, 3T, \dots \\ P_{ST}^k(n-1), & n = 1, 2, \dots, T-1, T+1, \dots, 2T-1, \dots \end{cases} \quad (16)$$

en donde $NSR_k(n)$ es la relación ruido-síñal (NSR) de la banda de frecuencia k^{th} en la muestra n . Este filtro IIR es adaptable puesto que el coeficiente numerador de la función de transferencia de este filtro es proporcional a $NSR_k(n)$ lo cual depende del tiempo y está adaptado en el adaptador de NSR 310. Este cálculo de la potencia se realiza preferiblemente en todo momento, independientemente de la actividad de la señal indicada por el VAD 304.

Un periodo de submuestreo adecuado para la medición de la potencia puede ser, por ejemplo, $T = 10$ muestras. Por ejemplo, coeficientes de filtro adecuados pueden ser:

$$\alpha_{ST} = 1 \quad (17a)$$

$$\beta_{ST} = 1 - T/128. \quad (17b)$$

En esta realización, la ganancia DC del filtro IIR usado para el cálculo de la potencia a corto plazo es $H_{ST}(1) = 12,8$.

A continuación se describe el método de adaptación de los valores de NSR cuando los tonos DTMF están ausentes. La NSR de una banda de frecuencia se adapta preferiblemente en base a la potencia a largo plazo, $P_{LT}(n)$, y la potencia a corto plazo, $P_{ST}(n)$, que corresponde a dicha banda y a la NSR general, $NSR_{general}(n)$ 322.

En la Figura 4 se ilustra el proceso de adaptación NSR para una banda de frecuencia individual. En la Figura 4 se presenta el adaptador del factor de compensación 402, el estimador de potencia a largo plazo 308a, el estimador de potencia a corto plazo 308b y el compensador de potencia 404. También se muestran el factor de compensación 406, el cálculo de potencia a largo plazo 323a y el cálculo de potencia a corto plazo 323b. También se muestra el error de predicción 408.

El estimador de la NSR general 306 es común a todas las bandas de frecuencia. En la realización preferida, el adaptador del factor de compensación 402 es también común a todas las bandas de frecuencia para conseguir eficiencia computacional. Sin embargo, por lo general, el adaptador del factor de compensación 402 puede estar diseñado para ser diferente para las diferentes bandas de frecuencia. Durante el silencio, el cálculo de potencia a corto plazo 323b de una banda de frecuencia es una medida del nivel de potencia de ruido. Durante la conversación, la potencia a corto plazo 323b predice el nivel de potencia del ruido.

Debido a que el ruido de fondo está casi fijo durante breves periodos de tiempo, la potencia a largo plazo 323a, que se mantiene constante durante ráfagas de voz, proporciona un buen cálculo de la potencia del ruido real preferiblemente tras su compensación con un escalar. La compensación escalar es beneficiosa debido a que la potencia a largo plazo 323a es una versión amplificada del nivel de potencia del ruido actual. Así, la diferencia entre la potencia a corto plazo 323b y la potencia a largo plazo compensada, proporciona un medio para ajustar la NSR. Esta diferencia se denomina el error de predicción 408. El signo del error de predicción 408 puede usarse para aumentar o disminuir la NSR sin realizar una división.

La adaptación NSR para la frecuencia k^{th} puede realizarse en el adaptador NSR 310 del siguiente modo durante la voz y el silencio (pero preferiblemente no durante la actividad de tono DTMF):

$$NSR_k(n) = \begin{cases} \max[0, NSR_k(n-1) - \Delta] & , P_{ST}(n) - C(n)P_{LT}(n) > 0 \\ \min[1 - \varepsilon, NSR_k(n-1) + \Delta] & , \text{de lo contrario} \end{cases} \quad (18)$$

cuando el factor de compensación (que se adapta en el adaptador del factor de compensación) para la potencia a largo plazo se obtiene mediante:

$$C(n) = \frac{H_{sr}(l)}{H_{lr}(l)} + NSR_{general}(n) \quad (19)$$

En la ecuación (18), el signo del error de predicción $P_{ST}(n) - C(n)P_{LT}(n)$, se usa para determinar la dirección del ajuste de $NSR_k(n)$. En esta realización, la cantidad de ajuste se determina en base a la actividad de la señal indicada por el VAD. La realización preferida usa un gran Δ durante la voz y un pequeño Δ durante el silencio. La potencia de la voz varía rápidamente y es adecuada una Δ mayor para rastrear rápidamente las variaciones. Durante el silencio, el ruido de fondo normalmente varía lentamente, y por ello basta un pequeño valor de Δ . Además, la utilización de un valor pequeño de Δ evita que los picos de ruido de corta duración hagan que la NSR aumente demasiado, lo que permitiría que el pico de ruido se filtrara a través del sistema de supresión del ruido.

A continuación se indica un conjunto adecuado de parámetros para su uso en la ecuación (18) cuando $T = 10$

$$\varepsilon = 0.05 \quad (20a)$$

$$\Delta = \begin{cases} 0.025 & \text{durante la conversación} \\ 0.00625 & \text{durante el silencio} \end{cases} \quad (20b)$$

En la realización preferida, el adaptador de la NSR adapta la NSR según el estado del VAD y la diferencia entre la potencia del ruido y la señal. A pesar de que esta realización preferida usa solo el signo de la diferencia entre la potencia del ruido y de la señal, la magnitud de esta diferencia también puede usarse para variar la NSR. Además, el adaptador de la NSR puede variar la NSR según uno o más de los siguientes. 1) el estado del VAD (por ej., un indicador VAD que indica voz o ruido); 2) la diferencia entre la potencia del ruido y la potencia de la señal; 3) una relación de la potencia del ruido a señal (NSR instantánea); y 4) la diferencia entre la NSR instantánea y una NSR previa. Por ejemplo, Δ puede variar en base a uno o más de dichos cuatro factores. Al adaptar Δ en base a la NSR instantánea, se proporciona un efecto de "suavización" o "promediado" al cálculo de NSR adaptada. En una realización, Δ puede variarse según la siguiente tabla (Tabla 1.1):

TABLA 1.1

Tabla de búsqueda de posibles valores de Δ usados para variar la NSR adaptada

	Magnitud de diferencia entre una NSR previa y una NSR instantánea durante la conversación	Δ
Durante la conversación	$ diferencia < 0,025$	0
	$0,025 < diferencia \leq 0,3$	0,025
	$ diferencia > 0,3$	0,05
Durante el silencio	$ diferencia < 0,00625$	0
	$0,00625 < diferencia \leq 0,3$	0,00625
	$ diferencia > 0,3$	0,01

La NSR general, $NSR_{general}(n)$, también puede ser un factor en la adaptación de la NSR a través del factor de compensación $C(n)$, dado por la ecuación (19). Un nivel de NSR general mayor tiene como resultado el sobreénfasis de la potencia a largo plazo para todas las bandas de frecuencia. Esto hace que todos los valores se adapten en niveles más elevados.

En consecuencia, esto haría que el factor de ganancia fuera inferior a los niveles de NSR general más elevados. La calidad percibida de la voz se mejora mediante esta eliminación excesiva bajo niveles de ruido de fondo más elevados.

ES 2 284 475 T3

5 Cuando el $NSR_{general}(n)$ 322 es negativo, lo cual sucede bajo condiciones de SNR general muy elevada, el valor NSR para cada banda de frecuencia en esta realización se adapta a cero. Así, la supresión insuficiente de niveles muy bajos de ruido se consigue debido a que dichos bajos niveles de ruido quedan efectivamente enmascarados por la voz. La relación entre la NSR general 322 y la NSR adaptada 324 en las diversas bandas de frecuencia puede describirse como una relación proporcional debido a que a medida que la NSR general 322 aumenta, la NSR adaptada 324 de cada banda aumenta.

10 En la realización preferida, $H_{LT}(1)=100$ y $H_{ST}(1)=12, 8$, de manera que $H_{ST}(1)/H_{LT}(1)=0,128$ en la ecuación (19).

Puesto que $-0,128 \leq NSR_{general}(n) \leq 0,064$ el intervalo del factor de compensación es:

$$0 \leq C(n) \leq 0.192 \quad (21)$$

15 Así, en esta realización, la potencia a largo plazo se sobreenfatiza al menos 1,5 veces respecto a su valor actual bajo condiciones de SNR baja. Bajo condiciones de SNR alta, la potencia a largo plazo se desenfatisa siempre que $C(n) \leq 0,128$.

20 Durante la actividad del tono DTMF según se indica en el VAD 304, el proceso de adaptar los valores de la NSR usando ecuaciones (18) y (19) para las bandas de frecuencia que contienen los tonos no es adecuado. Para las bandas que no contienen los tonos DTMF activos, (18) y (19) se continúan usando preferiblemente durante la actividad de tono DTMF.

25 Tan pronto como se detecta actividad DTMF, los valores NSR de las bandas de frecuencia que contienen tonos DTMF se ajustan preferiblemente a cero hasta que la actividad DTMF deja de detectarse. Al final de la actividad DTMF, puede permitirse que los valores NSR se adapten según se ha descrito anteriormente.

30 El detector de actividad de voz ("VAD") 304 determina si la señal de entrada contiene voz o silencio. Preferiblemente, el VAD 304 es un detector de actividad de voz y de actividad DTMF ("JVADAD"). La detección de actividad de voz y de actividad DTMF puede proceder independientemente y las decisiones de los dos detectores se combinan para formar una decisión final. Por ejemplo, como se muestra en la Figura 9, el JVADAD 304 puede incluir un detector de actividad de voz 304a, un detector de actividad DTMF 304b, y un circuito de determinación 304c. En una realización, el VAD 304a envía una señal de detección de voz 902 al circuito de determinación 304c y el detector de actividad DTMF envía una señal de detección DTMF 904 al circuito de determinación 304c. El circuito de determinación 304c determina entonces, en base a la señal de detección de voz 902 y señal de detección 904, si hay presente voz, actividad DTMF o silencio en la señal de entrada 316. El circuito de determinación 304c puede determinar el contenido de la señal de entrada 316, por ejemplo, en base a la lógica presentada en la Tabla 2 (a continuación).

40 En este contexto, el silencio se refiere a la ausencia de voz o de actividad DTMF, y puede incluir ruido.

45 El detector de actividad de voz puede producir un solo indicador, VAD 320, que se ajusta, por ejemplo, a uno si la voz se considera activa y cero en el caso contrario. El detector de actividad DTMF determina un indicador, por ejemplo DTMF=1, si se detecta actividad DTMF y DTMF=0 en el caso contrario. En la siguiente tabla (Tabla 2) se expone la lógica que puede usarse para determinar si hay actividad DTMF o actividad de voz:

TABLA 2

Lógica para su uso con JVADA

DTMF	VAD	Decisión
0	0	Silencio
0	1	Voz
1	0	Actividad DTMF presente
1	1	Actividad DTMF presente

50 Cuando se pulsa un botón de un teléfono de marcación por tonos, se generan un par de tonos. Uno de los tonos pertenecerá al siguiente conjunto de frecuencias: {697, 770, 852, 941} en Hz y una será del conjunto {1209, 1336, 1477, 1633} en Hz, según se ha indicado anteriormente en la Tabla 1. Estos conjuntos de frecuencias se denominan frecuencias del grupo bajo y del grupo alto, respectivamente. Así, son posibles dieciséis pares de tonos correspondien-

tes a 16 teclas de un teclado de teléfono ampliado. Es necesario que los tonos se reciban en $\pm 2\%$ de dichos valores nominales. Debe tenerse en cuenta que dichas frecuencias se han seleccionado cuidadosamente para reducir la cantidad de interacción armónica. Además, para la adecuada detección de un par de tonos, la diferencia de amplitud entre los tonos (denominada "twist") debe estar dentro de los 6 dB.

Un algoritmo de detección DTMF adecuado para la detección de tonos DTMF en el JVADAD 304 es una versión modificada del algoritmo de Goertzel. El algoritmo de Goertzel es un método recursivo de realizar la transformada discreta de Fourier (DFT) y es más eficiente que la DFT o FFT para pequeños números de tonos. A continuación se describe con mayor detalle la detección de tonos DTMF y la regeneración y ampliación de tonos DTMF.

La detección de actividad de voz se realiza preferiblemente usando las medidas de potencia en la primera región formante de la señal de salida $x(n)$. En el contexto de la señal de voz de telefonía, la primera región formante se define en el intervalo de aproximadamente 300-850 Hz. Se usa una medición de potencia a largo y una a corto plazo de la primera región formante con ecuaciones de diferencia obtenidas de:

$$P_{1st,ST}(n) = \beta_{1st,ST} P_{1st,ST}(n-1) + \alpha_{1st,ST} \left| \sum_{k \in F} x_k(n) \right| \quad (22)$$

$$P_{1st,LT}(n) = \begin{cases} \beta_{1st,LT,1} P_{1st,LT}(n-1) + \alpha_{1st,LT,1} \left| \sum_{k \in F} x_k(n) \right|, & \text{if } P_{1st,LT}(n) < P_{1st,ST}(n) \\ \beta_{1st,LT,2} P_{1st,LT}(n-1) + \alpha_{1st,LT,2} \left| \sum_{k \in F} x_k(n) \right|, & \text{if } P_{1st,LT}(n) \geq P_{1st,ST}(n) \end{cases} \quad (23)$$

en donde F representa el conjunto de bandas de frecuencia dentro de la primera región formante. Se prefiere la primera región formante pues contiene una gran proporción de la energía de voz y proporciona un medio adecuado para la detección precoz del inicio de un impulso de voz.

La medición de la potencia a largo plazo rastrea el nivel de ruido de fondo en el primer formante de la señal. La medición de la potencia a corto plazo rastrea el nivel de señal de voz en primer formante de la señal. Parámetros adecuados para las medidas de potencia del primer formante a corto y largo plazo son:

$$\alpha_{1st,LT,1} = 1/16000 \quad (24a)$$

$$\beta_{1st,LT,1} = 1 - \alpha_{1st,LT,1} \quad (24b)$$

$$\alpha_{1st,LT,2} = 1/256 \quad (24c)$$

$$\beta_{1st,LT,2} = 1 - \alpha_{1st,LT,2} \quad (24d)$$

$$\alpha_{1st,ST} = 1/128 \quad (24e)$$

$$\beta_{1st,ST} = 1 - \alpha_{1st,ST} \quad (24f)$$

El VAD 304 también puede utilizar el contador de persistencia, h_{VAD} 305. El contador de persistencia 305 se usa para mantener el estado de la salida del VAD 320 fija durante periodos breves cuando la potencia del primer formante

cae a niveles bajos. La potencia del primer formante puede caer a niveles bajos durante paradas breves y también durante los sonidos consonánticos de la conversación. La salida del VAD 320 se mantiene fija para evitar la supresión accidental de la voz. El contador de persistencia 305 puede actualizarse del siguiente modo:

$$h_{VAD} = \begin{cases} h_{VAD,max} & \text{if } P_{1st,ST}(n) > \mu P_{1st,LT}(n) + P_0 \\ \max[0, h_{VAD} - 1] & \text{de lo contrario} \end{cases} \quad (25)$$

cuando los valores adecuados para los parámetros (cuando el intervalo de $x(n)$ se normaliza en ± 1) son, por ejemplo:

$$\mu = 1.75 \quad (26)$$

$$P_0 = 16/8159 \quad (27)$$

El valor de $h_{VAD,max}$ preferiblemente corresponde a alrededor de 150-250 ms, es decir $h_{VAD,max} \in [1200, 2000]$. Se considera que la voz es activa (VAD=1) cuando se satisface la siguiente condición:

$$h_{VAD} > 0 \quad (28)$$

De lo contrario, se considera que la voz no está presente en la señal de entrada (VAD=0).

A continuación se describen los aparatos y métodos preferidos para la detección de tonos DTMF, en el JVADAD por ejemplo.

A pesar de que la realización preferida utiliza un aparato y método para detectar tonos DTMF, los principios descritos en relación con tonos DTMF son aplicables a todas las señales en banda. En este contexto, una señal en banda es cualquier tipo de señal tonal dentro del ancho de banda usado normalmente para la transmisión de voz. Las señales en banda ejemplares incluyen tonos de fax, tonos DTMF, tonos de marcado y tonos de línea ocupada.

Dado un bloque de muestras N (cuando se escoge adecuadamente N) de la señal de entrada, $u(n)$, $n = 0, 1, \dots, N-1$, el aparato puede comprobar la presencia de un tono próximo a una frecuencia determinada, ω_0 , mediante correlación de las muestras de entrada con un par de tonos en cuadratura en la frecuencia de prueba ω_0 . Los resultados de correlación pueden usarse para calcular la potencia de la señal de entrada alrededor de la frecuencia de prueba. Este procedimiento puede expresarse mediante las siguientes ecuaciones:

$$R_{\omega_0} = \sum_{n=0}^{N-1} u(n) \cos \omega_0 n \quad (29)$$

$$I_{\omega_0} = \sum_{n=0}^{N-1} u(n) \sin \omega_0 n \quad (30)$$

$$P_{\omega_0} = R_{\omega_0}^2 + I_{\omega_0}^2 \quad (31)$$

La ecuación (3) proporciona el cálculo de la potencia, P_{ω_0} , alrededor de la frecuencia de prueba ω_0 . La complejidad computacional del procedimiento indicado en (29)-(31) puede reducirse a la mitad usando un algoritmo de Goertzel modificado. Esto se describe a continuación:

$$w(n) = 2 \cos \omega_0 w(n-1) - w(n-2) + u(n), \quad n = 0, 1, 2, \dots, N-1 \quad (32)$$

$$w(N) = 2 \cos \omega_0 w(N-1) - w(N-2) \quad (33)$$

$$P_{\omega_0} = w^2(N) + w^2(N-1) - 2 \cos \omega_0 w(N) w(N-1) \quad (34)$$

Debe tenerse en cuenta que las condiciones iniciales para la recursión en (32) son $w(-1) = w(-2) = 0$. El procedimiento anterior de las ecuaciones (32)-(34) se realiza preferiblemente para cada una de las ocho frecuencias DTMF y sus segundos armónicos para un bloque determinado de muestras N . Los armónicos secundarios son las frecuencias que son el doble de los valores de las frecuencias DTMF. Dichas frecuencias se comprueban para garantizar que las señales de voz (que no poseen una estructura armónica) no se confunden con tonos DTMF. El algoritmo de Goertzel analiza preferiblemente bloques de longitud $N = 102$ muestras. A una velocidad de muestreo preferida de 8 kHz, cada bloque contiene señales con una duración de 12,75 ms. Se realizan preferiblemente las siguientes pruebas de validez para detectar la presencia de un par de tonos DTMF válidos en un bloque de muestras N :

- (1) La potencia de la frecuencia más fuerte del Grupo Bajo y la frecuencia más fuerte del Grupo Alto deben ser superiores a determinados umbrales.
- (2) La potencia de la frecuencia más fuerte del Grupo Bajo debe ser superior a los otros tres valores de potencia del Grupo Bajo en una determinada relación umbral.
- (3) La potencia de la frecuencia más fuerte del Grupo Alto debe ser superior a los otros tres valores de potencia del Grupo Alto en una determinada relación umbral.
- (4) La relación de la potencia de la frecuencia más fuerte del Grupo Bajo y la potencia de la frecuencia más fuerte del Grupo Alto deben estar comprendidas dentro de unos límites superiores e inferiores determinados.
- (5) La relación de los valores de potencia de la frecuencia más fuerte del Grupo Bajo y su armónico secundario debe superar una relación umbral determinada.
- (6) La relación de los valores de potencia de la frecuencia más fuerte del Grupo Alto y su armónico secundario debe superar una relación umbral determinada.

Si se pasan las pruebas de validez señaladas, debe realizarse una prueba de confirmación para garantizar que el par de tonos DTMF detectado es estable durante un periodo de tiempo suficiente. Para confirmar la presencia de un par de tonos DTMF, debe detectarse el mismo par de tonos DTMF para confirmar que está presente un par de tonos DTMF válido durante un periodo de tiempo suficiente tras un bloque de silencio según las especificaciones usadas, por ejemplo, para tres bloques consecutivos de aproximadamente 12,75 ms).

Para proporcionar una detección mejorada de tonos DTMF, se usa preferiblemente un algoritmo de detección de Goertzel modificado. Esto se consigue aprovechando la batería de filtros 302 del aparato de supresión de ruido 300 que ya ha dividido la señal de entrada en bandas de frecuencia diferentes. Cuando se usa el algoritmo de Goertzel para calcular la potencia próxima a una frecuencia de prueba, ω_0 , sufre un rechazo pobre de la potencia fuera de la proximidad de ω_0 . En el aparato mejorado 300, para calcular la potencia próxima a una frecuencia de prueba ω_0 , el aparato 300 usa la salida del filtro de paso de banda cuya banda de paso contiene ω_0 . Al aplicar el algoritmo de Goertzel a las señales que han pasado la banda, se consigue un rechazo excelente de la potencia en las frecuencias fuera de la proximidad de ω_0 .

Debe tenerse en cuenta que el aparato 300 usa preferiblemente las pruebas de validez descritas anteriormente en, por ejemplo, el JVADAD 304. El aparato puede o no usar la prueba de confirmación según se ha descrito anteriormente. En la realización preferida, se usa un método más sofisticado (que la prueba de confirmación) adecuado con objeto de ampliar o regenerar el tono DTMF. Las pruebas de validez se realizan preferiblemente en la porción de Detección de Actividad DTMF del detector conjunto de Actividad de Voz y Actividad DTMF 304.

A continuación se describe un método y aparato para la ampliación en tiempo real de tonos DTMF en relación con las Figuras 5 y 8. A pesar de que la realización preferida usa un aparato y método para ampliar los tonos DTMF, los principios descritos en relación con tonos DTMF son aplicables a todas las señales en banda. En este contexto, una señal en banda es cualquier tipo de señal tonal dentro del ancho de banda usado normalmente para la transmisión de voz. Las señales en banda ejemplares incluyen tonos de fax, tonos DTMF, tonos de marcado y tonos de línea ocupada.

En relación con la Figura 8, que ilustra el concepto de ampliar un tono en tiempo real, el tono de la señal de entrada 802 comienza alrededor de la muestra 100 y finaliza alrededor de la muestra 460, con una duración de alrededor de 45 ms. El indicador de actividad de tono 804, mostrado en el gráfico central, indica si se ha detectado un tono en el último bloque de, por ejemplo $N = 102$ muestras. Este indicador es cero hasta la muestra 250 punto en el que aumenta a uno.

Esto significa que el bloque desde la muestra 149 a la muestra 250 se comprobó y se encontró que contenía actividad de tono. Debe tenerse en cuenta que el bloque anterior desde la muestra 47 a la muestra 148 se analizó y se constató que no contenía actividad de tono aunque parte del bloque contenía el tono de entrada (el porcentaje de un bloque que debe contener un tono DTMF para poder fijar el indicador de actividad de tono para detectar un tono en un umbral predeterminado, por ejemplo). Se considera que este bloque contiene una pausa. En los siguientes dos bloques de muestras también se encontró que contenía actividad de tono en la misma frecuencia.

ES 2 284 475 T3

Así, tres bloques consecutivos de muestras contienen actividad de tono tras una pausa que confirma la presencia de un tono de la frecuencia que está siendo comprobada. (Debe tenerse en cuenta que, en la realización preferida, la presencia de tono del grupo bajo y de un tono del grupo alto debe confirmarse simultáneamente para confirmar la actividad DTMF).

La señal de salida 806 muestra cómo se amplía el tono de entrada incluso después de que el tono de entrada desaparezca alrededor de la muestra 460. Esta ampliación del tono se realiza en tiempo real y el tono ampliado tiene preferiblemente la misma fase, frecuencia y amplitud que el tono de entrada original.

El método preferido amplía un tono en una forma de fase continua según se discute a continuación. En la realización preferida, el tono ampliado continuará manteniendo la amplitud del tono de entrada. El método preferido aprovecha la información obtenida cuando se usa el algoritmo de Goertzel para la detección del tono DTMF. Por ejemplo, en el caso de un tono de entrada:

$$u(n) = A_0 \sin(\omega_0 n + \phi) \quad (35)$$

Pueden usarse las ecuaciones (32) y (33) del algoritmo de Goertzel para obtener los dos estados $w(N-1)$ y $w(N)$. Para valores lo suficientemente grandes de N , puede demostrarse que con las siguientes aproximaciones:

$$w(N-1) = B_0 \sin(N\omega_0 + \phi - \pi/2) \quad (36)$$

$$w(N) = B_0 \sin((N+1)\omega_0 + \phi - \pi/2) \quad (37)$$

donde

$$B_0 = \frac{A_0}{\sin \omega_0} \sum_{i=0}^{N-1} \sin^2(\omega_0 i) \quad (38)$$

Puede observarse que $w(N-1)$ y $w(N)$ contienen dos muestras consecutivas de senoide con frecuencia ω_0 . La fase y amplitud de este senoide posee preferiblemente una relación determinista con la fase y amplitud del senoide de entrada $u(n)$. Así, el generador de tono DTMF 321 puede generar un senoide usando un oscilador recursivo que coincide con la fase y amplitud del senoide de entrada $u(n)$ para tiempos de muestreo superiores a N usando el siguiente procedimiento:

(a) Calcula la siguiente muestra consecutiva del senoide con amplitud B_0 :

$$w(N+1) = (2 \cos \omega_0) w(N) - w(N-1) \quad (39)$$

(b) Genera dos muestras consecutivas de un senoide, $w'(n)$ con amplitud A_0 y fase ϕ utilizando $w(N-1)$, $w(N)$ y $w(N+1)$:

$$w'(N+1) = \frac{\cos \omega_0}{\sin \omega_0} w(N) - \frac{1}{\sin \omega_0} w(N-1) \quad (40)$$

$$w'(N+2) = \frac{\cos \omega_0}{\sin \omega_0} w(N+1) - \frac{1}{\sin \omega_0} w(N) \quad (41)$$

(c) Usa un oscilador recursivo para generar todas las muestras consecutivas del senoide para $j = 3, 4, 5, \dots$:

$$w'(N+j) = (2 \cos \omega_0) w'(N+j-1) - w'(N+j-2) \quad (42)$$

Puede usarse la secuencia $w'(N+j)$, $j = 1, 2, 3, 4, 5, \dots$ para ampliar el senoide de entrada $u(n)$ más allá de la muestra N .

Puesto que las dos frecuencias de tono DTMF las determina el detector de actividad DTMF, por ejemplo, puede usarse el procedimiento de las ecuaciones (39)-(42) para ampliar cada uno de los dos tonos. La ampliación de los tonos se realizará mediante una combinación ponderada de la señal de entrada con los tonos generados. Se usa preferiblemente una combinación ponderada para evitar cambios repentinos en la amplitud de la señal debido a una ligera incompatibilidad de amplitud y/o frecuencia entre los tonos de entrada y los tonos generados que producen ruido impulsivo. La combinación ponderada se realiza preferiblemente del siguiente modo:

$$y(n) = [1 - \rho(n)]u(n) + \rho(n)[w'_L(n) + w'_H(n)], \quad n = N+1, N+2, N+3, \dots \quad (43)$$

en donde $u(n)$ es la señal de entrada, $w'_L(n)$ es el tono generado del grupo bajo, $w'_H(n)$ es el tono generado del grupo alto, y $\rho(n)$ es un parámetro de ganancia que aumenta linealmente de 0 a 1 durante un breve periodo de tiempo, preferiblemente 5 ms o menos.

En el sistema de supresión de ruido, $x(n)$ es la muestra de entrada en el momento n en la batería de resonadores 302. La batería de resonadores 302 divide esta señal en un conjunto de señales de paso de banda $\{x_k(n)\}$. Recuperar la ecuación (4) anterior:

$$y(n) = \sum_k G_k(n)x_k(n) \quad (44)$$

Tal y como se ha descrito anteriormente, $G_k(n)$ y $x_k(n)$ son el factor de ganancia y la señal de paso de banda de la banda de frecuencia k^{th} , respectivamente, y $y(n)$ es la salida del aparato de supresión del ruido 300. El conjunto de señales de paso de banda $\{x_k(n)\}$ puede denominarse colectivamente como señal de entrada al método de ampliación del tono DTMF.

Debe tenerse en cuenta que el aparato de supresión de ruido 300 no introduce ningún retardo de bloque cuando se usa la ampliación del tono DTMF debido a que la muestra de entrada actual del aparato de supresión de ruido 300 se procesa tan pronto se recibe. Puesto que el método de detección DTMF funciona en bloques de muestras N , los autores definen el bloque actual de muestras N como las últimas muestras N recibidas, es decir, muestras $\{x(n-N), x(n-N+1), \dots, x(n-1)\}$. El bloque anterior estará formado por las muestras $\{x(n-2N), x(n-2N+1), \dots, x(n-N-1)\}$.

Volviendo ahora a la Figura 5, en cuya Figura se presenta un método ejemplar 500 para ampliar los tonos DTMF. Para determinar si están presentes tonos DTMF, se aplican preferiblemente pruebas de validez del método de detección DTMF a cada bloque. Si se detecta un par de tonos DTMF válidos, el dígito correspondiente se descodifica según la Tabla 1. En la realización preferida, se usan los dígitos descodificados que produce el detector de actividad DTMF (por ejemplo, el JVADAD) para los bloques actuales producidos y para tres bloques previos. En este contexto, la salida i -ésima del detector de actividad DTMF es D_i , con una i mayor correspondiente a una salida más reciente. De este modo, se hará referencia a los cuatro bloques de salida como D_i (es decir, D_1 , D_2 , D_3 y D_4). En la realización preferida, cada bloque de salida puede tener diecisiete posibles valores: los dieciséis valores posibles del teclado extendido y un valor que indica que no está presente ningún tono DTMF. Los bloques de salida D_i pueden transmitirse al generador de tonos DTMF 321 en la detección de actividad de voz y en la señal de detección de actividad DTMF 320. Se utilizará preferiblemente la siguiente Tabla de decisiones (Tabla 3) para implantar el método de ampliación de tono DTMF 500:

TABLA 3

Ampliación de los tonos DTMF

Condición	Acción
($D_3 = D_2 = D_1$) y (D_3, D_2, D_1 válido) y ((D_4 no válido) o ($D_4 \neq D_3$))	Eliminar los siguientes 3 bloques consecutivos
(D_4 válido) y (D_3, D_2, D_1 no válido y/o no igual)	Conjunto $G_L(n) = 1$ y $G_H(n) = 1$
($D_4 = D_3$) y (D_4, D_3 válido) y ($D_3 \neq D_2$) y (D_2, D_1 no válido y/o no igual)	Sustituir gradualmente el siguiente bloque con tonos DTMF generados usando la ecuación (46)
($D_4 = D_3 = D_2$)	Generar tonos DTMF para sustituir los tonos transmitidos
El resto de casos	Se permitió que todos los factores de ganancia varían según determine el aparato de supresión de ruido

Cuando se detecta el primer bloque que contiene un par de tonos DTMF válidos, se ajustan a uno dos factores de ganancia del sistema de supresión de ruido, $G_L(n)$ y $G_H(n)$ correspondientes las L^m y H^m bandas de frecuencia que contienen los tonos del grupo bajo y del grupo alto, respectivamente, por ejemplo, en la ecuación (4), es decir,

$$y(n) = \sum_k G_k(n) x_k(n), \quad G_L(n) = 1, G_H(n) = 1 \quad (45)$$

Esto corresponde a los pasos 504 y 506 de la Figura 5. Al ajustar a uno dichos factores de ganancia se garantiza que el aparato de supresión de ruido 300 no suprime los tonos DTMF después de ese punto. Tras ese bloque, si uno o dos bloques siguientes no tienen como resultado el mismo dígito descodificado, se permite que los factores de ganancia vuelvan a variar según determine el sistema de supresión de ruido, según indica el paso 508 de la Figura 5.

Cuando los dos primeros bloques consecutivos que contienen dígitos válidos idénticos se descodifican después de un bloque que no contiene tonos DTMF, se genera el par de tonos adecuado que corresponde al dígito, por ejemplo, usando las ecuaciones (39)-(42), y se usan para sustituir gradualmente los tonos de entrada. Esto corresponde a los pasos 510 y 512 de la Figura 5. Los tonos DTMF 329 se generan preferiblemente en el generador de tonos DTMF 321. La sustitución se realiza preferiblemente reduciendo la contribución de la señal de entrada, $x(n)$, y aumentando la contribución de los tonos generados, $w'_L(n)$ y $w'_H(n)$, a la señal de salida, $y(n)$, en las siguientes muestras M ($j = 1, 2, 3, \dots, M$) del siguiente modo:

$$y(n+j) = [1 - \rho(n+j)] \sum_k G_k(n) x_k(n) + \rho(n+j) [w'_L(n) + w'_H(n)] \quad (46)$$

$$\rho(n+j) = j/M \quad (47)$$

Debe tenerse en cuenta que no es necesaria ninguna división en la ecuación (47).

Comenzando con $\rho(n) = 0$, puede usarse la relación $\rho(n+j+1) = \rho(n+j) + 1/M$ para actualizar el valor de ganancia de cada muestra.

Un valor ejemplar de M es 40.

Así, en una realización preferida, tras recibir los primeros dos bloques con dígitos válidos idénticos, las primeras muestras M del siguiente bloque se sustituyen gradualmente por tonos DTMF 329 generados de modo que después de las muestras M , la salida $y(n) = w'_L(n) + w'_H(n)$. Tras muestras M , los tonos generados se mantienen hasta que deja de detectarse un par de tonos DTMF en un bloque. En tal caso, el retardo en la detección de la señal del tono DTMF (debido a, por ejemplo, la longitud del bloque) se compensa con el retardo en la detección del final de la señal de un tono DTMF. Como resultado de ello, el tono DTMF se extiende a través del uso de los tonos DTMF 329 generados.

En una realización alternativa, los tonos generados continúan después de que deje de detectarse un tono DTMF por ejemplo durante aproximadamente un bloque y medio después de que no se detecte en un bloque un par de tonos DTMF. En esta realización, puesto que JVADAD puede requerir aproximadamente un bloque para detectar un par de tonos DTMF, el generador de tonos DTMF extiende el tono DTMF aproximadamente un bloque más allá del par de tonos DTMF actual. Así, en el caso improbable de que un par de tonos DTMF tenga la longitud mínima detectable, la salida del tono DTMF debe tener al menos la longitud del tono de entrada mínimo. Sea cual sea la realización que se utilice, la longitud de tiempo necesaria para detectar el par de tonos DTMF puede variar según el método de detección del JVADAD y la longitud de bloque usada.

En consecuencia, el periodo de ampliación adecuada también variará.

Cuando tres o más bloques consecutivos contienen dígitos válidos, el generador de tonos DTMF 321 genera tonos DTMF 329 para sustituir los tonos DTMF de entrada. Esto corresponde a los pasos 513 y 514 de la Figura 5.

Cuando el generador de tonos DTMF ha ampliado el par de tonos DTMF, la señal de entrada se atenúa durante un tiempo adecuado, por ejemplo, durante aproximadamente tres bloques de 12,75 ms consecutivos, para garantizar que existe una pausa suficiente tras la señal DTMF de salida. Esto corresponde a los pasos 515 y 516 de la Figura 5.

Durante el periodo de atenuación, la salida viene dada por

$$y(n) = \rho(n) \sum_k G_k(n) x_k(n) \quad (48)$$

en donde $\rho(n) = 0,02$ es una opción adecuada. Después de los tres bloques, $\rho(n) = 1$, y el aparato de supresión del ruido permite determinar los factores de ganancia hasta que vuelve a detectarse actividad DTMF (según indica el paso 508 de la Figura 5).

Debe tenerse en cuenta que es posible que el bloque actual contenga actividad DTMF a pesar de que se programe que el bloque actual esté suprimido tal y como en la ecuación (48). Si esto sucede, por ejemplo, cuando los pares de tonos DTMF están espaciados por el periodo de tiempo mínimo permitido. Si la señal de entrada 316 contiene tonos DTMF legítimos, los dígitos estarán espaciados normalmente por al menos tres bloques consecutivos de silencio. Así, generalmente solo el primer bloque de muestras de un par de tonos DTMF válidos sufrirá supresión. Sin embargo, esto estará compensado por la ampliación del tono DTMF.

Volviendo ahora a la Figura 6, en cuya Figura se presenta un método para regenerar los tonos DTMF 329. La regeneración del tono DTMF es una alternativa a la ampliación del tono DTMF. A pesar de que la realización preferida usa un aparato y tonos DTMF regeneradores, los principios discutidos en relación con tonos DTMF son aplicables a todas las señales en banda. En este contexto, una señal en banda es cualquier tipo de señal tonal dentro del ancho de banda usado normalmente para la transmisión de voz.

Las señales en banda ejemplares incluyen tonos de fax, tonos DTMF, tonos de marcado y tonos de línea ocupada.

Por ejemplo, la regeneración del tono DTMF puede realizarse en el generador de tonos DTMF 321. El método de ampliación introduce un retardo muy corto (aproximadamente un bloque en la realización ilustrada) pero es ligeramente más complicado porque las fases de los tonos se corresponden para la adecuada detección de los tonos DTMF. El método de regeneración introduce un retardo mayor (unos pocos bloques en la realización ilustrada) aunque es más sencillo puesto que no requiere que los tonos generados se correspondan a la fase de los tonos de entrada. El retardo introducido en cualquier caso es temporal y se produce solo para tonos DTMF. El retardo hace que una pequeña cantidad de la señal tras los tonos DTMF se suprima para garantizar pausas suficientes tras un par de tonos DTMF. La regeneración DTMF también puede hacer que se suprima un solo bloque de señal de voz después un segundo de un par de tonos DTMF. Puesto que este es un caso muy improbable y solo las primeras muestras N de voz sufren la supresión, no es probable la pérdida de información útil.

Sin embargo, cuando se realiza la ampliación DTMF, el conjunto de señales $\{X_k(n)\}$ puede denominarse colectivamente como la entrada al método de regeneración DTMF. Cuando se generan tonos DTMF 329, la señal de salida del combinador 315 es:

$$y(n) = \rho_1(n) \sum_k G_k x_k(n) + \rho_2(n) [w'_L(n) + w'_H(n)] \quad (49)$$

en donde $\sum_k G_k x_k(n)$ es la salida del multiplicador de ganancia, $w'_L(n)$ y $w'_H(n)$ son los tonos bajos y altos generados (de haberlos), y $\rho_1(n)$ y $\rho_2(n)$ son factores de ganancia adicionales.

Cuando no hay ninguna señal DTMF presente en la señal de entrada, $\rho_1(n) = 1$ y $\rho_2(n) = 0$. Durante la regeneración de un par de tonos DTMF, $\rho_2(n) = 1$. Si se suprime la señal de entrada (para garantizar el silencio tras el final de un par de tonos DTMF regenerados o durante la regeneración del par de tonos DTMF), entonces $\rho_1(n)$ se ajusta en un valor pequeño, por ej., $\rho_1(n) = 0,02$. Preferiblemente se usan dos osciladores recursivos 332 para regenerar los tonos de grupo alto y bajo adecuados correspondientes al dígito descodificado.

Con referencia continuada a la Figura 6, en una realización ejemplar, la regeneración de los tonos DTMF usa los bloques de salida actuales y los cinco previstos del detector de actividad de tono DTMF (por ej., en el JVADAD), dos indicadores, y dos contadores. Puede hacerse referencia a los cinco bloques de salida previos y el actual como D1, D2, D3, D4, D5, y D6, respectivamente. Los indicadores, el indicador ELIMINAR y el indicador GENTONOS se describen a continuación en relación con la acción que producen en el generador de tonos DTMF 321, el combinador 315, y/o el multiplicador de ganancia 314 para realizar:

ELIMINAR	Acción
1	Eliminar la salida del aparato de supresión de ruido ajustando $\rho_1(n)$ en un valor pequeño, por ej., $\rho_1(n) = 0,02$ en la ecuación (49)
0	Conjunto $\rho_1(n) = 1$

GENTONOS	Acción
1	Generar tonos DTMF y darles salida ajustando $\rho_2(n) = 1$
0	Detener la generación de tonos DTMF y ajustar $\rho_2(n) = 0$

Contador	Objeto
<i>wait_count</i>	Cuenta el número de bloques a eliminar desde el punto en donde un par de tonos DTMF fue detectado en primer lugar
<i>sup_count</i>	cuenta el número de bloques a eliminar desde el final de una regeneración de par de tonos DTMF

En la inicialización, todos los indicadores y contadores están ajustados a cero preferiblemente. La siguiente Tabla (Tabla 4) ilustra una realización ejemplar del método de regeneración de tonos DTMF 600:

TABLA 4

Regeneración de los tonos DTMF

Condición	Acción
(D6 válido) y (D5, D4, D3, D2, D1 son no válidos y/o no iguales)	SUPPRESS = 1 <i>wait_count</i> = 40
(D6 = D5 = D4) y (D6, D5, D4 válido) y (D3, D2, D1 no válidos y/o no iguales)	GENTONES = 1
(D3 = D2 = D1) y (D3, D2, D1 válido) y (D6, D5, D4 no válidos y/o no iguales)	GENTONES = 0 <i>sup_count</i> = 4
(VAD = 1) y (<i>sup_count</i> = 0)	SUPPRESS = 0 <i>wait_count</i> = 0
(GENTONES = 0) y (<i>wait_count</i> = 0)	SUPPRESS = 0
(GENTONES = 0) y (<i>wait_count</i> > 0)	Decrement <i>wait_count</i>
<i>sup_count</i> > 0	Decrement <i>sup_count</i>

Debe tenerse en cuenta que las condiciones de la Tabla 4 no son necesariamente exclusivas. Así, en la realización preferida, se comprueba cada condición en el orden presentado en la Tabla 4 al final de un bloque (con la excepción de las condiciones 1-3, que son mutuamente exclusivas). La acción correspondiente se lleva entonces al siguiente bloque si la condición es verdadera. Por lo tanto, pueden tomarse múltiples acciones al principio de un bloque. Al igual que en la ampliación de tono DTMF, se usa preferiblemente $N = 102$ para la detección de tonos DTMF para su utilización con el aparato y método de regeneración de tonos DTMF.

A continuación se presenta una descripción del método de regeneración de tonos preferido. Cuando se detecta primero un par de DTMF válido en un bloque de muestras N , la salida del sistema de supresión de ruido se suprime ajustando $\rho_1(n)$ en un valor pequeño, por ej., $\rho_1(n) = 0,02$. Esto se indica satisfaciendo la primera condición de la Tabla 4 y ajustando la marca ELIMINAR en un valor de 1, y corresponde a los pasos 602 y 604 de la Figura 6. Después de encontrar que tres bloques consecutivos contienen el mismo dígito válido, los tonos DTMF, $w'_L(n)$ y $w'_H(n)$, correspondientes al dígito recibido se generan y se pasan a la salida, es decir $\rho_1(n) = 0,02$ y $\rho_2(n) = 1$. Esto corresponde a satisfacer la segunda condición de la Tabla 4 y ajustando el indicador GENTONOS en 1, y los pasos 606 y 608 de la Figura 6. La regeneración del tono DTMF continúa preferiblemente hasta que no se detecta el par DTMF en el bloque actual.

Los tonos DTMF 329 generados pueden producirse de forma continuada durante un tiempo suficiente (cuando el par DTMF deja de detectarse en el bloque actual), por ejemplo en tres o cuatro bloques más (para garantizar que se envía una duración suficiente de los tonos DTMF).

Al igual que con el método de ampliación de tono DTMF, la regeneración del tono DTMF puede producirse durante un periodo de tiempo adicional, por ejemplo, un bloque y medio o un bloque de muestras N , para garantizar que los tonos DTMF cumplen los estándares de duración mínimos. En la realización ilustrada en la Tabla 4, los tonos DTMF 329 se generan en 3 bloques cuando los tonos DTMF dejan de detectarse. Esto corresponde al cumplimiento de la condición 3 de la Tabla 4, y a los pasos 610 y 612 de la Figura 6. Debe tenerse en cuenta que a pesar de que se ajuste *sup-count* a 4 cuando 3 bloques no DTMF consecutivos siguen a 3 bloques DTMF idénticos, válidos, consecutivos, *sup-count* se elimina en los pasos 614 y 616 antes de eliminar ningún bloque (de modo que se suprimen 3 bloques, no 4). Después de esto, se transmite un periodo de silencio con la duración suficiente, es decir, $\rho_1(n) = 0,02$ y $\rho_2(n) = 0$. Esto puede ser, por ejemplo, una longitud de cuatro bloques de 12,75.

Mientras tanto, el detector de actividad DTMF (preferiblemente como parte del JVADAD) continúa funcionando durante la transmisión de los tonos generados y del silencio. Si se recibe un dígito válido mientras se está transmitiendo el último bloque de los tonos DTMF regenerados 329 y/o el silencio, se generan los tonos DTMF adecuados para dicho dígito y se transmiten tras completar el periodo de silencio. Si no se reciben dígitos válidos durante ese periodo, continúa suprimiéndose la salida durante un periodo de espera. Durante este periodo de espera, si cualquiera de los indicadores del JVADAD es uno, es decir $VAD=1$ o $DTMF=1$, el periodo de espera finaliza de inmediato. Si se finaliza el periodo de espera debido a la actividad de la voz ($VAD=1$), la salida se determina por el sistema de supresión del ruido con $\rho_1(n) = 1$ y $\rho_2(n) = 0$, por ejemplo ajustando el indicador ELIMINAR igual a 0 (según se indica si se satisface la condición 4 de la Tabla 4). Si se finaliza el periodo de espera por la actividad DTMF ($DTMF=1$), entonces continúa la supresión de la señal de entrada, por ejemplo ajustando el indicador ELIMINAR igual a 1 (según se indica si se satisface la condición 1 de la Tabla 4). Una condición de $VAD = 1$ corresponde a los pasos 618 y 620 de la Figura 6 mientras que una condición de $DTMF = 1$ corresponde a los pasos 602 y 604 de la Figura 6. Periodos de espera ejemplares van desde alrededor de medio segundo a un segundo (alrededor de 40 a 80 bloques). El periodo de espera se usa para evitar la fuga de pequeñas cantidades de tonos DTMF desde la señal de entrada. El uso del *wait_count* facilita el recuento del número de bloques a eliminar desde el punto en el que se detecta en primer lugar un par de tonos DTMF. Esto corresponde a los pasos 622 y 624 de la Figura 6.

Cuando no hay señales DTMF presentes, $\rho_1(n) = 1$ y $\rho_2(n) = 0$. En la realización actual, siempre que se detecte un par de tonos DTMF en un bloque, la salida del sistema de supresión de ruido se suprime, por ejemplo, ajustando $\rho_1(n)$ en un valor pequeño por ej., $\rho_1(n) = 0,02$. En la realización descrita en la Tabla 4, $\rho_1(n)$ se ajusta en un valor pequeño ajustando ELIMINAR igual a 1. Al final de cada bloque de muestras N , si ELIMINAR es igual a 1, entonces para las siguientes muestras N , $\rho_1(n) = 0,02$. Al final de cada bloque, si se determina que los tonos DTMF deben regenerarse durante el siguiente bloque (por ejemplo si $GENTONOS = 1$), entonces $\rho_2(n) = 1$. El generador de tonos 321 utiliza *wait_count* y los indicadores del JVADAD para determinar si continúa la supresión de la señal de entrada durante el periodo de espera.

Si no se detecta voz ni un tono DTMF durante el periodo de espera, entonces *wait_count* se reduce a 0, entonces la condición por defecto de $\rho_1(n) = 1$ y $\rho_2(n) = 0$ se ajusta preferiblemente (según los pasos 626 y 628 de la Figura 6).

Los métodos de ampliación del tono DTMF y de regeneración del tono DTMF se describen por separado. Sin embargo, es posible combinar la ampliación del tono DTMF y la regeneración del tono DTMF en un método y/o aparato.

A pesar de que los métodos de ampliación y regeneración tonos DTMF se describen en la presente memoria con un sistema de supresión del ruido. Dichos métodos también pueden usarse con otros sistemas de mejora de la voz tales como sistemas de control de la ganancia adaptables, cancelación de la reverberación y sistemas de eliminación de la reverberación. Además, la ampliación y regeneración de tonos DTMF descritos en esta memoria resultan especialmente útiles cuando no puede tolerarse un retardo. Sin embargo, si el retardo es tolerable, por ej., si se tolera un retardo de 20 ms en un sistema de mejora de la voz (que puede ser el caso si el sistema de mejora de la voz funciona en conjunción con un dispositivo de compresión de la voz), la ampliación y/o regeneración de los tonos puede no ser necesaria. Sin embargo, un sistema de mejora de la voz que no cuenta con un detector DTMF puede escalar los tonos de forma inapropiada. Cuando está presente un detector DTMF, el aparato y método de eliminación del ruido puede detectar la presencia de los tonos y determinar los factores de escalado para unificar las subbandas adecuadamente.

En relación a las Figuras 3 y 4, la batería de filtros 302, JVADAD 304, el contador de persistencia 305, el estimador de la NSR 306, el estimador de potencia 308, el adaptador de la NSR 310, el controlador de ganancia 312, el multiplicador de ganancia 314, el adaptador del factor de compensación 402, el estimador de la potencia a largo plazo 308a, el estimador de potencia a corto plazo 308b, el compensador de potencia 404, el generador de tonos DTMF 321, osciladores 332, circuito de submuestreo 330, y combinador 315 pueden implantarse usando lógica combinatoria y secuencial, un ASIC, a través del software aplicado por una CPU, un chip DSP, u otros similares. Los elementos mencionados anteriormente pueden formar parte de equipos usados para realizar otras funciones operativas. Las señales de entrada, las bandas de frecuencia, las mediciones y estimaciones de la potencia, factores de ganancia, NSR y NSR adaptadas, indicadores, errores de predicción, factores de compensación, contadores, y constantes pueden almacenarse en registros, RAM, ROM u otros similares, y pueden generarse a través del software, a través de una estructura de datos situada en un dispositivo de memoria tal como RAM o ROM, entre otros.

REIVINDICACIONES

1. El aparato para mejorar la calidad de una señal de comunicaciones en un sistema de comunicaciones, que comprende:
 - los medios para dividir la señal de comunicación en una pluralidad de señales de banda de frecuencia;
 - los medios para generar una primera señal de potencia para la primera señal de banda de frecuencia, la cual se basa en estimar durante un primer periodo de tiempo la potencia de la primera señal de banda de frecuencia;
 - los medios para generar una segunda señal de potencia para la primera señal de banda de frecuencia, la cual se basa en estimar durante un segundo periodo de tiempo inferior al primer periodo de tiempo la potencia de la primera señal de banda de frecuencia;
 - los medios para generar una relación ruido-señal (NSR) que representa una condición de la primera señal de banda de frecuencia en respuesta a una relación aritmética predeterminada que comprende al menos una adición, sustracción y multiplicación pero sin división, entre los valores derivados de la primera y segunda señal de potencia;
 - los medios para ajustar la ganancia de la señal de banda de frecuencia en respuesta a la NSR para generar una señal de banda de frecuencia ajustada; y
 - los medios para combinar la señal de banda de frecuencia ajustada con al menos una señal de banda de frecuencia adicional para generar una señal de comunicación ajustada.
2. El aparato según la Reivindicación 1 en donde las primeras señales de potencia, las segundas señales de potencia y las señales de condición se generan mediante una implantación de punto fijo.
3. El aparato según la Reivindicación 1 o la Reivindicación 2 en donde la NSR comprende el cálculo de una relación que comprende la potencia de la señal ruidosa media durante dicha conversación y la potencia del ruido de fondo medio durante la ausencia de dicha conversación de una de las señales de banda de frecuencia.
4. El aparato según cualquiera de las Reivindicaciones precedentes en donde la relación aritmética incluye ajustar los valores de las señales de condición mediante incrementos predeterminados.
5. El aparato según la Reivindicación 4 en donde se incrementan los valores de las NSR en respuesta a la primera de las relaciones aritméticas predeterminadas y en donde se disminuyen los valores de las señales de condición en respuesta a la segunda de las relaciones aritméticas predeterminadas.
6. El aparato según cualquiera de las Reivindicaciones precedentes en donde la relación aritmética comprende el ajuste mediante un factor de compensación de una señal del grupo compuesto por una de las primeras señales de potencia y una de las segundas señales de potencia.
7. El aparato según la Reivindicación 6 en donde se genera una señal de ruido general que representa una relación entre los componentes de la señal y del ruido de la señal de comunicación y en donde se genera el factor de compensación en respuesta a la señal de ruido general.
8. El aparato según cualquiera de las Reivindicaciones precedentes en donde se genera una señal de detección de voz en respuesta a la detección del componente de conversación de la señal de comunicación y en donde se varia el valor de los incrementos en respuesta a la señal de detección de voz.
9. El aparato según cualquiera de las Reivindicaciones precedentes que comprende un procesador seleccionado del grupo compuesto por lógica combinatorial y secuencial, un circuito integrado para aplicaciones específicas, una unidad de procesamiento central que ejecuta el software y un procesador de señal digital que ejecuta el software.
10. El aparato según cualquiera de las Reivindicaciones precedentes en donde las primeras señales de potencia se generan durante la ausencia de dicha conversación.
11. El aparato según cualquiera de las Reivindicaciones precedentes en donde la segunda señal de potencia se genera durante dicha conversación y ruido.
12. Un método para mejorar la calidad de una señal de comunicaciones en un sistema de comunicaciones, que comprende los pasos de:
 - dividir la señal de comunicación en una pluralidad de señales de banda de frecuencia;
 - generar una primera señal de potencia para la primera señal de banda de frecuencia, la cual se basa en estimar durante un primer periodo de tiempo la potencia de las primeras señales de banda de frecuencia;

ES 2 284 475 T3

generar una segunda señal de potencia para las primeras señales de banda de frecuencia, la cual se basa en estimar durante un segundo periodo de tiempo inferior al primer periodo de tiempo la potencia de las primeras señales de banda de frecuencia;

- 5 generar una relación ruido-señal (NSR) que representa una condición de la primera señal de banda de frecuencia en respuesta a una relación aritmética predeterminada que comprende al menos una adición, sustracción y multiplicación pero sin división, entre los valores derivados de las primeras y segundas señales de potencia;

- 10 ajustar la ganancia de la señal de banda de frecuencia en respuesta a la NSR para generar una señal de banda de frecuencia ajustada; y

combinar la señal de banda de frecuencia ajustada con al menos una señal de banda de frecuencia adicional para generar una señal de comunicación ajustada.

- 15 13. Un método según la Reivindicación 12 en donde dichas primeras señales de potencia, segundas señales de potencia y señales de condición se generan mediante una implantación de punto fijo.

- 20 14. Un método según la Reivindicación 12 o Reivindicación 13 en donde la relación entre los componentes de la señal y del ruido comprende el cálculo de una relación que comprende la potencia de la señal ruidosa media durante dicha conversación y la potencia del ruido de fondo medio durante la ausencia de dicha conversación de las señales de banda de frecuencia.

- 25 15. Un método según cualquiera de las Reivindicaciones 12 a 14 en donde la relación aritmética incluye ajustar los valores de las señales de condición mediante incrementos predeterminados.

16. Un método según la Reivindicación 15 en donde se incrementan los valores de las NSR en respuesta a la primera de las relaciones aritméticas predeterminadas y en donde se disminuyen los valores de las señales de condición en respuesta a la segunda de las relaciones aritméticas predeterminadas.

- 30 17. Un método según cualquiera de las Reivindicaciones 12 a 16 en donde la relación aritmética comprende el ajuste mediante un factor de compensación de una señal del grupo compuesto por una de las primeras señales de potencia y una de las segundas señales de potencia.

- 35 18. Un método según la Reivindicación 17 que incluye el paso de generar una señal de ruido general que representa una relación entre los componentes de la señal y del ruido de la señal de comunicación y en donde se genera el factor de compensación en respuesta a la señal de ruido general.

- 40 19. Un método según cualquiera de las Reivindicaciones 12 a 18 que incluye el paso de generar una señal de detección de voz en respuesta a la detección del componente de conversación de la señal de comunicación y en donde se varia el valor de los incrementos en respuesta a la señal de detección de voz.

20. Un método según cualquiera de las Reivindicaciones 12 a 19 en donde la generación de las primeras señales de potencia comprende generar las primeras señales de potencia durante la ausencia de dicha conversación.

- 45 21. Un método según cualquiera de las Reivindicaciones 12 a 20 en donde la generación de las segundas señales de potencia comprende generar las segundas señales de potencia durante dicha conversación y ruido.

50

55

60

65

FIG. 1

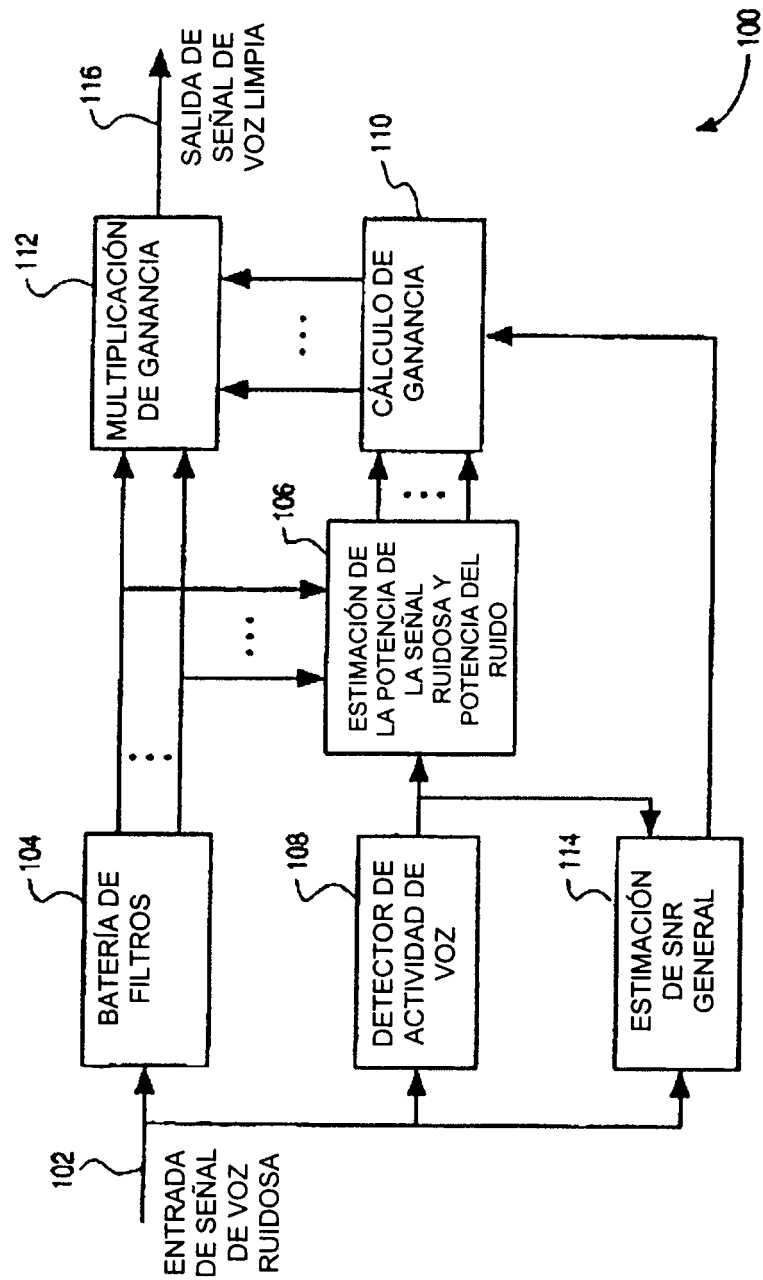


FIG. 2

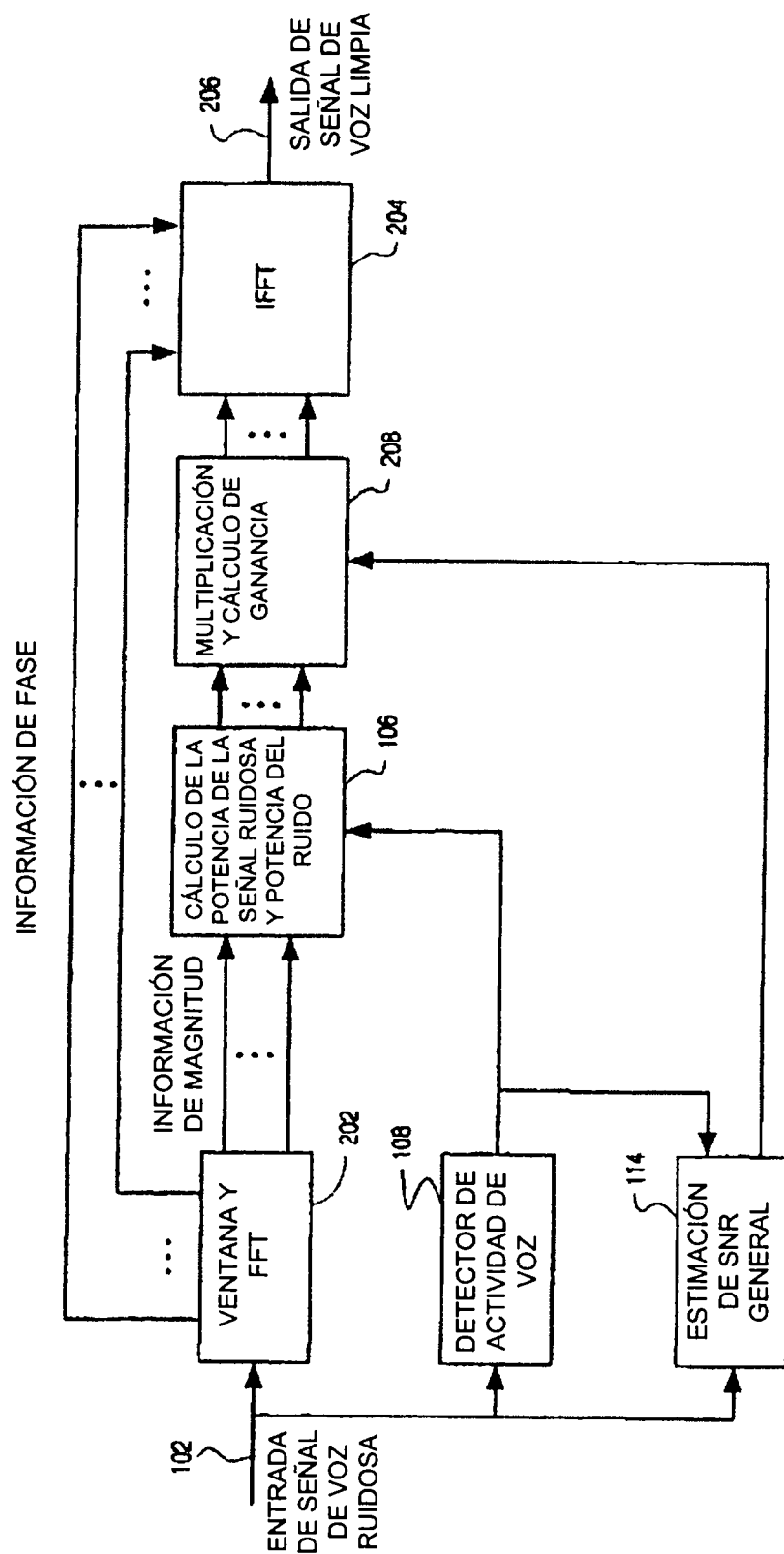


FIG. 3

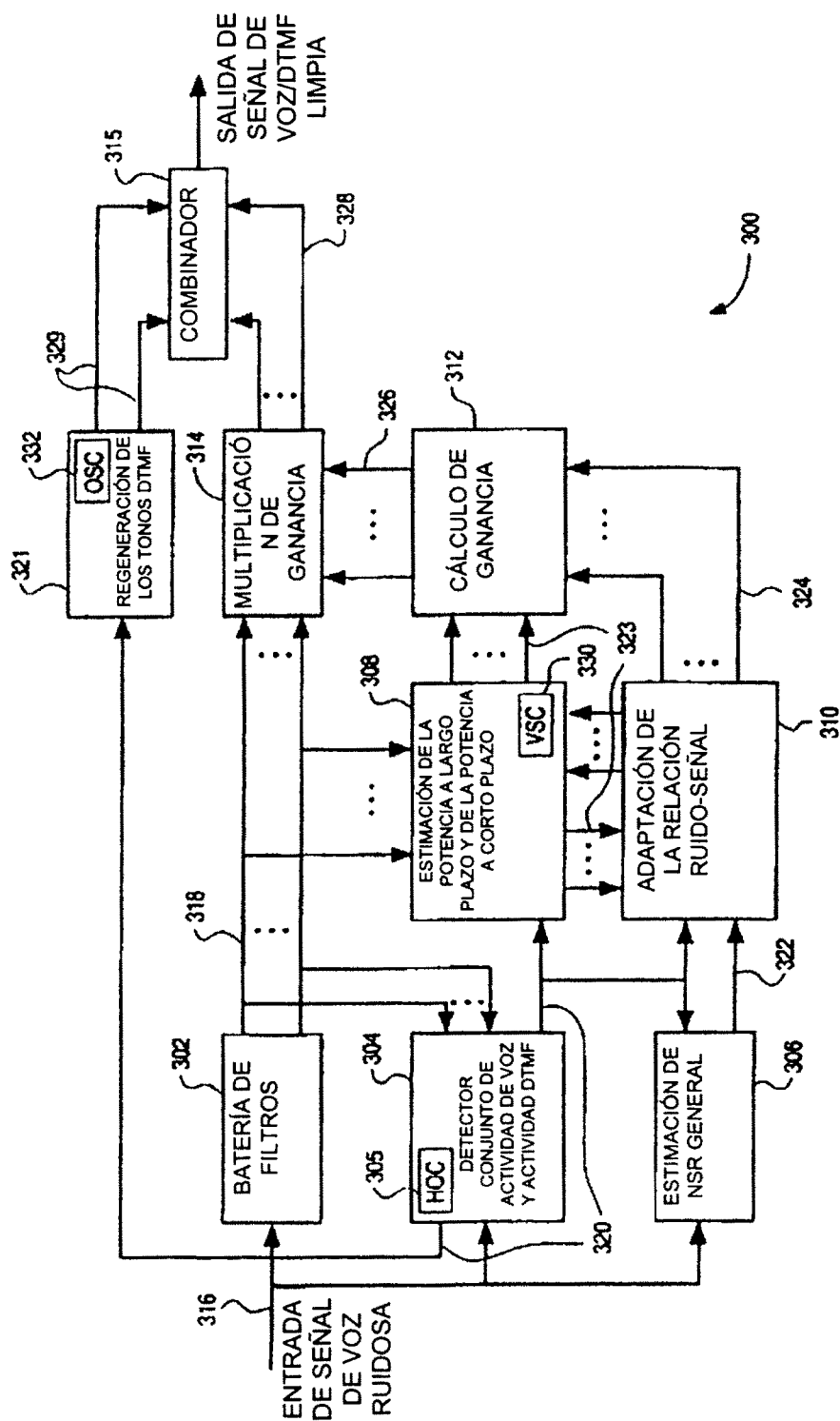


FIG. 4

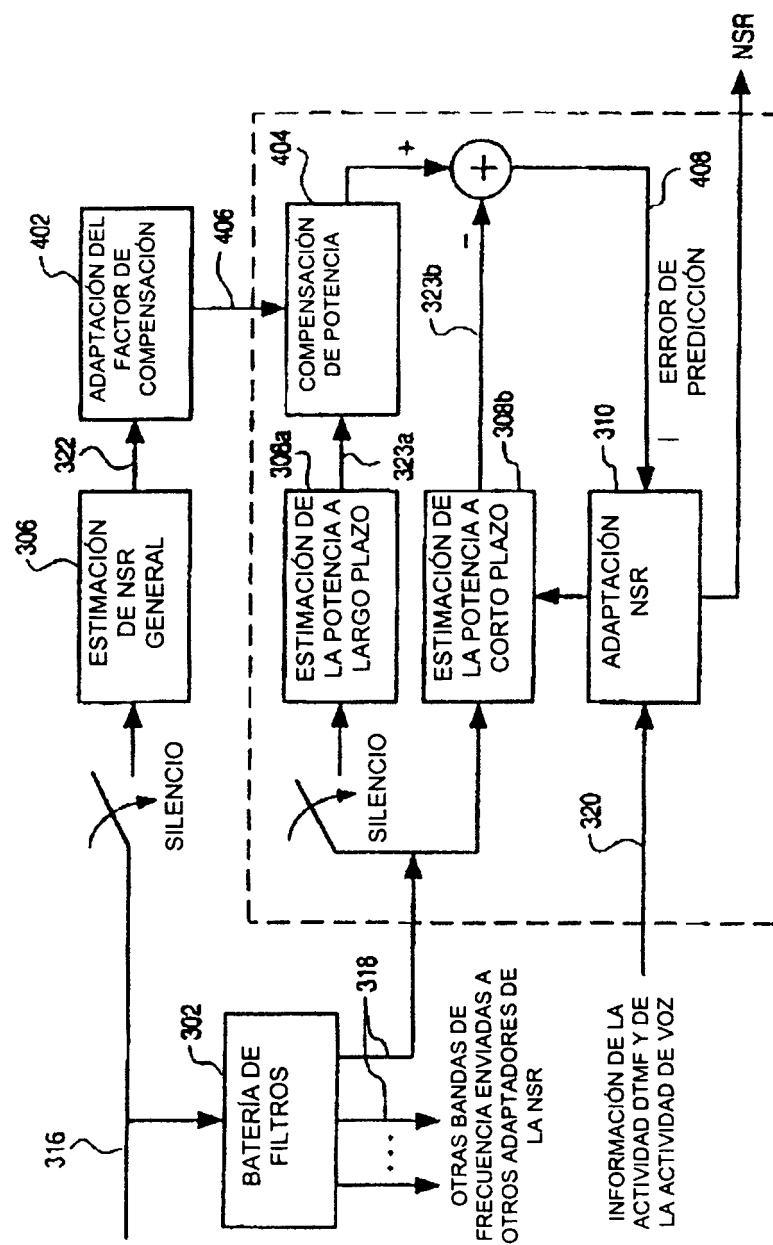


FIG. 5

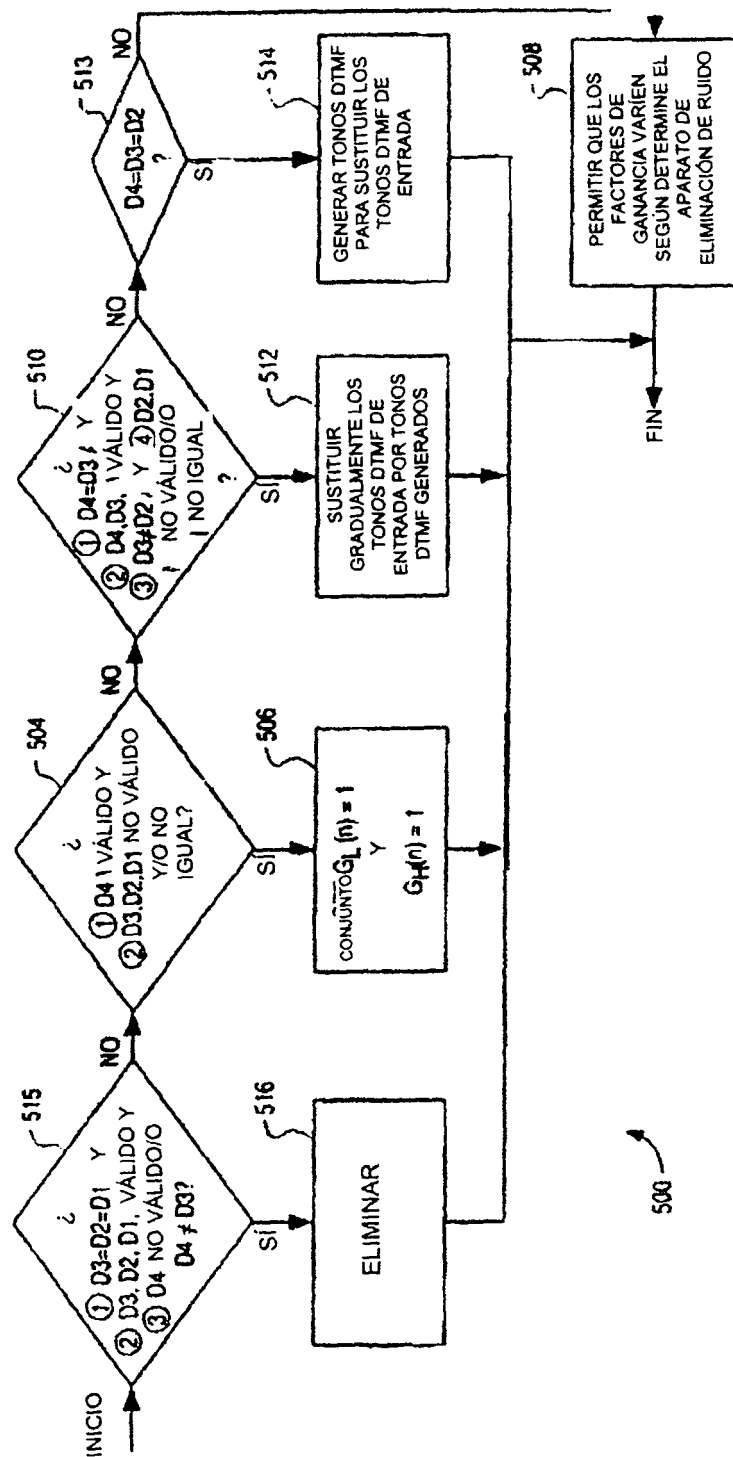


FIG. 6

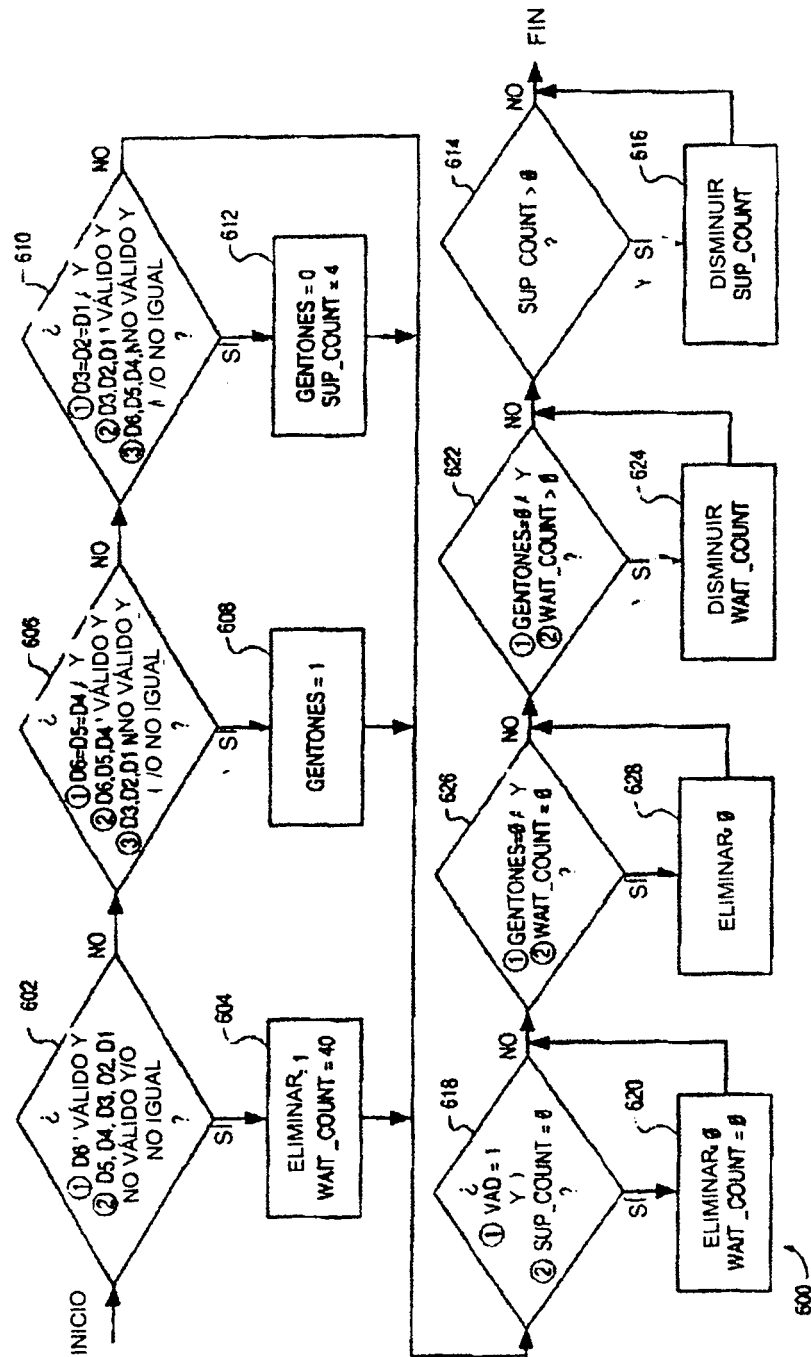


FIG. 7

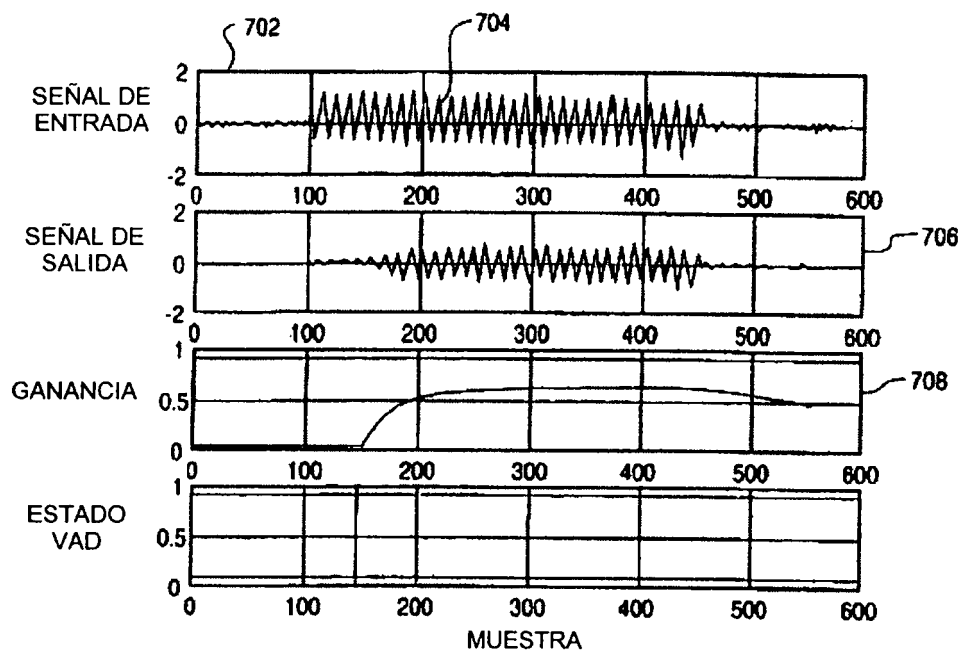


FIG. 8

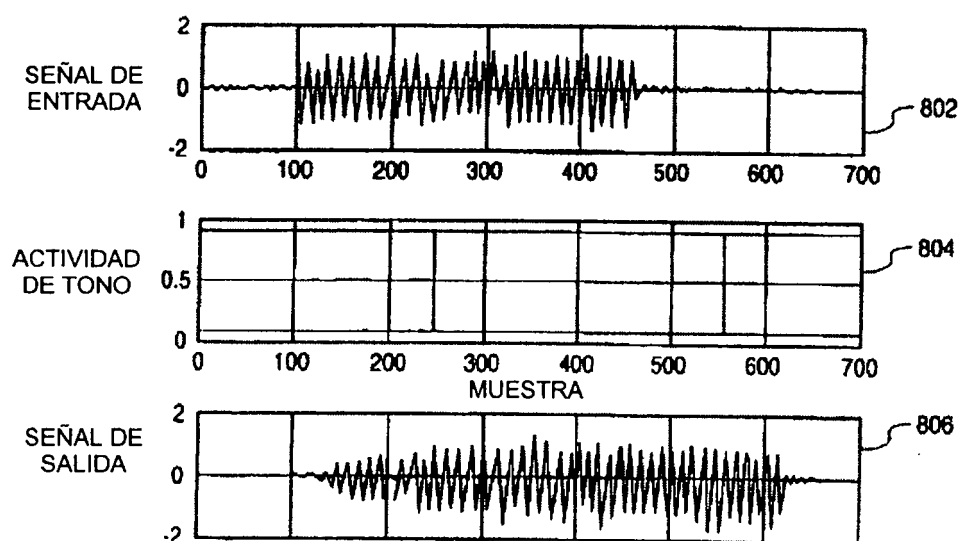


FIG. 9

