(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 371 548**

(51) Int. CI.:
G10L 21/02 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **04006719 .1**
(96) Fecha de presentación: **19.03.2004**
(97) Número de publicación de la solicitud: **1465160**
(97) Fecha de publicación de la solicitud: **06.10.2004**

(54) Título: **PROCEDIMIENTO DE ESTIMACIÓN DE RUIDO USANDO APRENDIZAJE INCREMENTAL BAYESIANO.**

(30) Prioridad:
31.03.2003 US 403638

(45) Fecha de publicación de la mención BOPI:
05.01.2012

(45) Fecha de la publicación del folleto de la patente:
05.01.2012

(73) Titular/es:
MICROSOFT CORPORATION
ONE MICROSOFT WAY
REDMOND, WASHINGTON 98052, US

(72) Inventor/es:
Acero, Alejandro;
Deng, Li y
Droppo, James G.

(74) Agente: **Carpintero López, Mario**

DESCRIPCIÓN

Procedimiento de estimación de ruido usando aprendizaje incremental bayesiano

Antecedentes de la invención

La presente invención se refiere a la estimación del ruido. En particular, la presente invención se refiere a la estimación del ruido en señales usadas en el reconocimiento de patrones.

Un sistema de reconocimiento de patrones, tal como un sistema de reconocimiento del habla, toma una señal de entrada e intenta descodificar la señal para hallar un patrón representado por la señal. Por ejemplo, en un sistema de reconocimiento del habla, una señal de habla (a menudo denominada una señal de prueba) es recibida por el sistema de reconocimiento y es descodificada para identificar una cadena de palabras representadas por la señal de habla.

Las señales de entrada están habitualmente corrompidas por alguna forma de ruido. Para mejorar las prestaciones del sistema de reconocimiento de patrones, a menudo es deseable estimar el ruido en la señal ruidosa.

En el pasado, se han usado algunos entornos para estimar el ruido en una señal. En un entorno, se usan algoritmos en lotes que estiman el ruido en cada trama de la señal de entrada, independientemente del ruido hallado en otras tramas en la señal. Las estimaciones individuales de ruido se promedian luego entre sí para formar un valor de consenso del ruido para todas las tramas. En un segundo entorno, se usa un algoritmo recursivo que estima el ruido en la trama actual en base a estimaciones de ruido para una o más tramas anteriores o sucesivas. Tales técnicas recursivas admiten que el ruido cambie lentamente a lo largo del tiempo.

En una técnica recursiva, se supone que una señal ruidosa es una función no lineal de una señal limpia y de una señal de ruido. Para ayudar en el cálculo, esta función no lineal se aproxima a menudo por una expansión truncada en serie de Taylor, que se calcula alrededor de algún punto de expansión. En general, la expansión en serie de Taylor proporciona sus mejores estimaciones de la función en el punto de expansión. Así, la aproximación por serie de Taylor es sólo tan buena como la selección del punto de expansión. En la técnica anterior, sin embargo, el punto de expansión para la serie de Taylor no estaba optimizado para cada trama. Como resultado, la estimación del ruido producido por los algoritmos recursivos ha sido menos que ideal.

Se han usado técnicas de máxima probabilidad (ML) y de máximo a posteriori (MAP) para la estimación de puntos secuenciales del ruido no estacionario, usando un modelo no lineal iterativamente linealizado para el entorno acústico.

La técnica de ML se ilustra en el documento de la técnica anterior de L. Deng et al. “Recursive noise estimation using iterative stochastic approximation for stereo-based robust speech recognition” [“Estimación recursiva del ruido usando aproximación estocástica iterativa para el reconocimiento de voz robusto con base estéreo”], págs. 81 a 84, Taller del IEEE de Reconocimiento y Comprensión Automática del Habla, 2001. ASRU’01, 9 al 13 de diciembre de 2001.

La técnica de MAP se ilustra en el documento de la técnica anterior de L. Deng et al. “Log-domain speech feature enhancement using sequential MAP noise estimation and a phase-sensitive model of the acoustic environment” [“Mejora de características del habla en el dominio del registro usando estimación secuencial de ruido de MAP y un modelo sensible a la fase del entorno acústico”], págs. 1813 a 1816, anales de ICSLP 2002: 7ª conferencia internacional sobre el procesamiento del lenguaje hablado, 16 al 20 de septiembre de 2002.

En general, usando un sencillo modelo Gaussiano para la distribución del ruido, la estimación de MAP proporcionaba una mejor calidad de la estimación del ruido. Sin embargo, en la técnica de MAP, los parámetros de media y varianza asociados a la técnica anterior del ruido Gaussiano se fijan a partir de un segmento de cada emisión de prueba libre de habla. Para el ruido no estático, esta aproximación puede no reflejar debidamente estadísticas anteriores realistas del ruido.

Resumen de la invención

Es el objeto de la invención proporcionar un procedimiento mejorado para estimar el ruido en una señal ruidosa, y un correspondiente medio y sistema legible por ordenador, que sean más efectivos para estimar el ruido en señales de patrones.

Este objeto es resuelto por la invención, según se reivindica en las reivindicaciones independientes.

Las realizaciones preferidas se definen en las reivindicaciones dependientes.

Un nuevo enfoque de la estimación del ruido no estático usa el aprendizaje incremental de Bayes. En un aspecto, esta técnica puede definirse como que supone una distribución anterior del ruido variable en el tiempo, donde la estimación del ruido, que puede ser definida por hiperparámetros (media y varianza), se actualiza recursivamente usando una aproximación posterior calculada en una etapa precedente en el tiempo o en las tramas. En otro aspecto, esta técnica

puede definirse como estimar sucesivamente, para cada trama, el ruido en cada trama, de modo tal que una estimación del ruido para una trama actual se base en una aproximación Gaussiana de la probabilidad de datos para la trama actual y una aproximación Gaussiana del ruido en una secuencia de tramas anteriores.

Breve descripción de los dibujos

La FIG. 1 es un diagrama en bloques de un entorno de cálculo en el cual puede ponerse en práctica la presente invención.

La FIG. 2 es un diagrama en bloques de un entorno alternativo de cálculo en el cual puede ponerse en práctica la presente invención.

La FIG. 3 es un diagrama de flujo de un procedimiento de estimación del ruido en una realización de la presente invención.

La FIG. 4 es un diagrama en bloques de un sistema de reconocimiento de patrones en el cual puede usarse la presente invención.

Descripción detallada de realizaciones ilustrativas

La FIG. 1 ilustra un ejemplo de un entorno 100 de un sistema informático adecuado en el cual puede implementarse la invención. El entorno 100 de sistema informático es sólo un ejemplo de un entorno informático adecuado y no está concebido para sugerir ninguna limitación en cuanto al alcance del uso o a la funcionalidad de la invención. Tampoco debería interpretarse que el entorno informático 100 tenga alguna dependencia o requisito con respecto a cualquiera, o a una combinación, de los componentes ilustrados en el entorno operativo ejemplar 100.

La invención es operativa con otros numerosos entornos o configuraciones de sistema informático de propósito general o de propósito especial. Los ejemplos de sistemas informáticos, entornos y / o configuraciones bien conocidos, que pueden ser adecuados para su uso con la invención incluyen, pero no se limitan a, los ordenadores personales, los ordenadores servidores, los dispositivos de mano o portátiles, los sistemas multiprocesadores, los sistemas basados en microprocesadores, los equipos de sobremesa, los equipos electrónicos programables de consumo, los ordenadores personales en red, los miniordenadores, los ordenadores centrales, los sistemas de telefonía, los entornos informáticos distribuidos que incluyen a cualquiera de los sistemas o dispositivos anteriores, y similares.

La invención puede describirse en el contexto general de instrucciones ejecutables por ordenador, tales como módulos de programa ejecutados por un ordenador. En general, los módulos de programa incluyen rutinas, programas, objetos, componentes, estructuras de datos, etc., que realizan tareas específicas o implementan tipos específicos de datos abstractos. Las tareas realizadas por los programas y módulos se describen más adelante, y con ayuda de figuras. Los expertos en la técnica pueden implementar la descripción y / o las figuras en el presente documento como instrucciones ejecutables por ordenador, que pueden realizarse en cualquier forma de medio legible por ordenador expuesto más adelante.

La invención también puede ponerse en práctica en entornos informáticos distribuidos donde las tareas son realizadas por dispositivos de procesamiento remoto que están enlazados a través de una red de comunicaciones. En un entorno informático distribuido, los módulos de programa pueden localizarse en medios de almacenamiento de ordenador, tanto locales como remotos, incluyendo dispositivos de almacenamiento de memoria.

Con referencia a la FIG. 1, un sistema ejemplar para implementar la invención incluye un dispositivo informático de propósito general en forma de un ordenador 110. Los componentes del ordenador 110 pueden incluir, pero no se limitan a, una unidad 120 de procesamiento, una memoria 130 del sistema y un bus 121 del sistema que acopla diversos componentes del sistema, incluyendo la memoria del sistema, con la unidad 120 de procesamiento. El bus 121 del sistema puede ser cualquiera de diversos tipos de estructuras de bus, incluyendo un bus de memoria o controlador de memoria, un bus periférico y un bus local que use cualquiera entre una gran variedad de arquitecturas de bus. A modo de ejemplo, y no de limitación, tales arquitecturas incluyen el bus de Arquitectura Estándar Industrial (ISA), el bus de Arquitectura de Micro Canal (MCA), el bus ISA Mejorado (EISA), el bus local de la Asociación de Estándares de Electrónica de Vídeo (VESA) y el bus de Interconexión de Componentes Periféricos (PCI), también conocido como el bus Entresuelo.

El ordenador 110 incluye habitualmente una gran variedad de medios legibles por ordenador. Los medios legibles por ordenador pueden ser cualquier medio disponible al que pueda acceder el ordenador 110, e incluyen medios tanto volátiles como no volátiles, y medios tanto extraíbles como no extraíbles. A modo de ejemplo, y no de limitación, los medios legibles por ordenador pueden incluir medios de almacenamiento de ordenador y medios de comunicación. Los medios de almacenamiento de ordenador incluyen medios tanto volátiles como no volátiles, tanto extraíbles como no extraíbles, implementados en cualquier procedimiento o tecnología para el almacenamiento de la información, tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos. Los medios de

almacenamiento de ordenador incluyen, pero no se limitan a, memoria RAM, memoria ROM, memoria EEPROM, memoria flash u otra tecnología de memoria, CD-ROM, discos versátiles digitales (DVD) u otro almacenamiento en disco óptico, casetes magnéticos, cinta magnética, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio que pueda usarse para almacenar la información deseada y al cual pueda accederse desde el ordenador 110. Los medios de comunicación realizan habitualmente instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos en una señal modulada de datos tal como una onda portadora u otro mecanismo de transporte, e incluyen cualquier medio de entrega de información. El término "señal modulada de datos" significa una señal que tiene una o más de sus características fijadas o cambiadas de tal manera como para codificar información en la señal. A modo de ejemplo, y no de limitación, los medios de comunicación incluyen medios cableados tales como una red cableada o conexión de cableado directo, y medios inalámbricos tales como medios acústicos, de Frecuencia de Radio, infrarrojos u otros medios inalámbricos. Las combinaciones de cualquiera de los anteriores también deberían incluirse dentro del alcance de los medios legibles por ordenador.

La memoria 130 del sistema incluye medios de almacenamiento de ordenador en forma de memoria volátil y / o no volátil, tal como la memoria de sólo lectura (ROM) 131 y la memoria de acceso aleatorio (RAM) 132. Un sistema 133 básico de entrada / salida (BIOS), que contiene las rutinas básicas que ayudan a transferir información entre los elementos dentro del ordenador 110, tal como durante el arranque, se almacena habitualmente en la memoria ROM 131. La memoria RAM 132 contiene habitualmente datos y / o módulos de programa que son inmediatamente accesibles para, y / o están actualmente siendo empleados en operaciones por, la unidad 120 de procesamiento. A modo de ejemplo, y no de limitación, la FIG. 1 ilustra el sistema operativo 134, los programas 135 de aplicación, otros módulos 136 de programa y los datos 137 de programa.

El ordenador 110 también puede incluir otros medios de almacenamiento de ordenador, extraíbles o no extraíbles, volátiles o no volátiles. Sólo a modo de ejemplo, la FIG. 1 ilustra un controlador 141 de disco rígido que lee de, o escribe en, medios magnéticos no extraíbles y no volátiles, un controlador 151 de disco magnético que lee de, o escribe en, un disco magnético 152 extraíble y no volátil, y un controlador 155 de disco óptico que lee de, y escribe en, un disco óptico 156 extraíble, no volátil, tal como un CD-ROM u otros medios ópticos. Otros medios de almacenamiento de ordenador, extraíbles o no extraíbles, volátiles o no volátiles, que pueden usarse en el entorno operativo ejemplar incluyen, pero no se limitan a, los casetes de cinta magnética, las tarjetas de memoria flash, los discos versátiles digitales, la cinta de vídeo digital, la memoria RAM de estado sólido, la memoria ROM de estado sólido, y similares. El controlador 141 de disco rígido está habitualmente conectado con el bus 121 del sistema a través de una interfaz de memoria no extraíble tal como la interfaz 140, y el controlador 151 de disco magnético y el controlador 155 de disco óptico están habitualmente conectados con el bus 121 del sistema por una interfaz de memoria extraíble, tal como la interfaz 150.

Los controladores y sus medios asociados de almacenamiento de ordenador, expuestos anteriormente e ilustrados en la FIG. 1, proporcionan el almacenamiento de instrucciones legibles por ordenador, estructuras de datos, módulos de programa y otros datos para el ordenador 110. En la FIG. 1, por ejemplo, el controlador 141 de disco rígido se ilustra como almacenando el sistema operativo 144, los programas 145 de aplicación, otros módulos 146 de programa y los datos 147 de programa. Obsérvese que estos componentes pueden bien ser los mismos que, o bien ser distintos a, el sistema operativo 134, los programas 135 de aplicación, los otros módulos 136 de programa y los datos 137 de programa. El sistema operativo 144, los programas 145 de aplicación, los otros módulos 146 de programa y los datos 147 de programa reciben números distintos aquí para ilustrar que, como mínimo, son copias distintas.

Un usuario puede ingresar comandos e información en el ordenador 110 mediante dispositivos de entrada tales como un teclado 162, un micrófono 163 y un dispositivo señalador 161, tal como un ratón, bola de rastreo o panel táctil. Otros dispositivos de entrada (no mostrados) pueden incluir una palanca de juegos, un panel de juegos, una antena satelital, un escáner, o similares. Estos y otros dispositivos de entrada están frecuentemente conectados con la unidad 120 de procesamiento a través de una interfaz 160 de entrada de usuario que está acoplada con el bus del sistema, pero puede estar conectada por otra interfaz y otras estructuras de bus, tales como un puerto paralelo, un puerto de juegos o un bus universal en serie (USB). Un monitor 191 u otro tipo de dispositivo de visualización también está conectado con el bus 121 del sistema mediante una interfaz, tal como una interfaz 190 de vídeo. Además del monitor, los ordenadores también pueden incluir otros dispositivos periféricos de salida tales como los altavoces 197 y la impresora 196, que pueden conectarse a través de una interfaz periférica 190 de salida.

El ordenador 110 puede funcionar en un entorno en red usando conexiones lógicas con uno o más ordenadores remotos, tales como un ordenador remoto 180. El ordenador remoto 180 puede ser un ordenador personal, un dispositivo de mano, un servidor, un encaminador, un ordenador personal en red, un dispositivo a la par u otro nodo común de red, y habitualmente incluye muchos de, o todos, los elementos descritos anteriormente con respecto al ordenador 110. Las conexiones lógicas ilustradas en la FIG. 1 incluyen una red de área local (LAN) 171 y una red de área amplia (WAN) 173, pero también pueden incluir otras redes. Tales entornos de red son comunes en oficinas, redes de ordenadores de ámbito empresarial, intranets e Internet.

Cuando se usa en un entorno de red LAN, el ordenador 110 se conecta con la LAN 171 a través de una interfaz de red o adaptador 170. Cuando se usa en un entorno de red WAN, el ordenador 110 incluye habitualmente un módem 172 u otro medio para establecer comunicaciones por la red WAN 173, tal como Internet. El módem 172, que puede ser interno o externo, puede conectarse con el bus 121 del sistema mediante la interfaz 160 de entrada de usuario, u otro mecanismo adecuado. En un entorno en red, los módulos de programa ilustrados con respecto al ordenador 110, o partes del mismo, pueden almacenarse en el dispositivo remoto de almacenamiento en memoria. A modo de ejemplo, y no de limitación, la FIG. 1 ilustra los programas 185 de aplicación remota como residentes en el ordenador remoto 180. Se apreciará que las conexiones de red mostradas son ejemplares y que pueden usarse otros medios para establecer un enlace de comunicaciones entre los ordenadores.

La FIG. 2 es un diagrama en bloques de un dispositivo móvil 200, que es un entorno informático ejemplar. El dispositivo móvil 200 incluye un microprocesador 202, una memoria 204, componentes 206 de entrada / salida (E / S) y una interfaz 208 de comunicación para comunicarse con ordenadores remotos u otros dispositivos móviles. En una realización, los componentes precitados se acoplan para la comunicación entre sí por un bus 210 adecuado.

La memoria 204 se implementa como memoria electrónica no volátil, tal como memoria de acceso aleatorio (RAM) con un módulo de resguardo por batería (no mostrado), de modo tal que la información almacenada en la memoria 204 no se pierda cuando se apaga la alimentación general al dispositivo móvil 200. Una parte de la memoria 204 está preferiblemente adjudicada como memoria direccionable para la ejecución de programas, mientras que otra parte de la memoria 204 se usa preferiblemente para el almacenamiento, tal como para simular almacenamiento en un controlador de disco.

La memoria 204 incluye un sistema operativo 212, y programas 214 de aplicación, así como un almacén 216 de objetos. Durante el funcionamiento, el sistema operativo 212 es preferiblemente ejecutado por el procesador 202 a partir de la memoria 204. El sistema operativo 212, en una realización preferida, es un sistema operativo de marca WINDOWS® CE, disponible comercialmente en la Corporación Microsoft. El sistema operativo 212 está preferiblemente diseñado para dispositivos móviles, e implementa características de bases de datos que pueden ser utilizadas por las aplicaciones 214 a través de un conjunto de interfaces y procedimientos de programación de aplicaciones expuestas. Los objetos en el almacén 216 de objetos son mantenidos por las aplicaciones 214 y el sistema operativo 212, al menos parcialmente, en respuesta a llamadas a las interfaces y procedimientos de programación de aplicaciones expuestas.

La interfaz 208 de comunicación representa a numerosos dispositivos y tecnologías que permiten al dispositivo móvil 200 enviar y recibir información. Los dispositivos incluyen módems cableados e inalámbricos, receptores satelitales y sintonizadores de difusión, para nombrar sólo unos pocos. El dispositivo móvil 200 también puede conectarse directamente con un ordenador para intercambiar datos con el mismo. En tales casos, la interfaz 208 de comunicación puede ser un transceptor infrarrojo o una conexión de comunicación en serie o en paralelo, todos los cuales son capaces de transmitir información de transferencia por flujo.

Los componentes 206 de entrada / salida incluyen una gran variedad de dispositivos de entrada, tales como una pantalla sensible al tacto, botones, rodillos y un micrófono, así como una gran variedad de dispositivos de salida, que incluyen un generador de audio, un dispositivo vibratorio y un visor. Los dispositivos enumerados anteriormente son a modo de ejemplo y no necesariamente están todos presentes en el dispositivo móvil 200. Además, otros dispositivos de entrada / salida pueden adosarse a, o hallarse en, el dispositivo móvil 200, dentro del alcance de la presente invención.

En un aspecto de la presente invención, se proporcionan un sistema y un procedimiento que estiman el ruido en señales de reconocimiento de patrones. Para hacer esto, la presente invención usa un algoritmo recursivo para estimar el ruido en cada trama de una señal ruidosa, en base, en parte, a una estimación de ruido hallada para al menos una trama vecina. En la presente invención, se estima el ruido para una única trama usando el aprendizaje incremental de Bayes, donde se supone una distribución anterior del ruido variable a lo largo del tiempo y se actualiza recursivamente una estimación del ruido usando una aproximación para el ruido posterior calculado en una trama anterior. Mediante este proceso recursivo, la estimación del ruido puede rastrear el ruido no estático.

Sea $y^t_1 = y_1, y_2,..., y_p..., y_t$ una secuencia de datos de observación de habla ruidosa, expresados en el dominio de registro (tal como log-spectra o cepstra), y de los que se supone que tienen valores escalares, sin pérdida de generalidad. Los datos y^t_1 se usan para estimar secuencialmente la secuencia de ruido corruptor $n^t_1 = n_1, n_2,... ,... ,n_t$ con la misma longitud *t* de datos. Dentro del entorno de aprendizaje Bayesiano, se supone que el conocimiento acerca del ruido *n* (tratado como un parámetro desconocido) está contenido en una distribución a-priori dada de *p(n)*. Si la secuencia de ruido es estática, es decir, las propiedades estadísticas del ruido no cambian a lo largo del tiempo, entonces la inferencia convencional de Bayes (es decir, el cálculo del posterior) sobre el parámetro *n* del ruido en cualquier momento puede lograrse mediante la regla de Bayes de “modalidad en lotes”:

$$p(n|y_1^t) = \frac{p(y_1^t|n)p(n)}{\int_\Theta p(y_1^t|n)p(n)dn},$$

donde Θ es una región admisible del espacio de parámetros de ruido. Dado $p(n|y_1^t)$, cualquier estimación del ruido n es posible, en principio. Por ejemplo, una estimación puntual convencional de MAP del ruido *n* se calcula como un máximo global o local del posterior $p(n|y_1^t)$. La estimación de error de mínimos cuadrados medios (MMSE) es la expectativa acerca del posterior $p(n|y_1^t)$.

Sin embargo, cuando la secuencia de ruido no es estática y los datos de entrenamiento de habla ruidosa y^t_1 se presentan secuencialmente como en las más prácticas aplicaciones de mejora de características del habla, se necesitan nuevas técnicas de estimación del ruido a fin de rastrear las estadísticas del ruido que está cambiando a lo largo del tiempo. En una aplicación iterativa, la regla de Bayes puede escribirse como:

$$p(n_t|y_1^t) = \frac{1}{C_t}p(y_t|y_1^{t-1}, n_t)p(n_t|y_1^{t-1}),$$

donde

$$C_t = p(y_1^t|y_1^{t-1}) = \int_\Theta p(y_t|y_1^{t-1}, n_t)p(n_t|y_1^{t-1})dn_t.$$

Suponiendo la independencia condicional entre el habla ruidosa y_t y su pasada y_1^{t-1} dado n_t, o $p(y_t|y_1^{t-1}, n_t) = p(y_t|n_t)$, y suponiendo fluidez en el posterior: $p(n_t|y_1^{t-1}) \approx p(n_{t-1}|y_1^{t-1})$, la ecuación anterior puede escribirse como:

$$p(n_t|y_1^t) \approx \frac{1}{C_t}p(y_t|n_t)p(n_{t-1}|y_1^{t-1}). \qquad (1)$$

El aprendizaje incremental del ruido no estático puede establecerse ahora con el uso repetido de la Ec. 1 según lo siguiente. Inicialmente, en ausencia de datos *y* de habla ruidosa, la función de distribución de probabilidad posterior proviene de la $p(n_0|y_0) = p(n_0)$ anterior conocida, donde $p(n_0)$ se obtiene del análisis de tramas conocidas de sólo ruido, y se supone Gaussiana. Entonces, el uso de la Ec. 1 para t = 1 produce:

$$p(n_1|y_1) \approx \frac{1}{C_1}p(y_1|n_1)p(n_0), \qquad (2)$$

y para *t* = 2 produce:

$$p(n_2|y_1, y_2) \approx \frac{1}{C_2}p(y_2|n_2)p(n_1|y_1),$$

usando el valor $p(n_1|y_1)$ ya calculado a partir de la Ec. 2. Para *t* = 3, la Ec. 1 se convierte en

$$p(n_3|y_1^3) \approx \frac{1}{C_3}p(y_3|n_3)p(n_2|y_1, y_2),$$

y así sucesivamente. Este proceso genera así recursivamente una secuencia de posteriores (siempre que se disponga

de $p(y_t|n_t)$:

$$p(n_1|y_1),\ p(n_2|y_1^2),\ \ldots, p(n_\tau|y_1^\tau),\ \ldots, p(n_t|y_1^t), \ldots \quad (3)$$

que proporciona una base para efectuar la inferencia incremental de Bayes sobre la secuencia n_1^t de ruido no estático. El principio general de la inferencia incremental de Bayes expuesto hasta ahora se aplicará ahora a un modelo específico de distorsión acústica, que proporciona los datos $p(y_t|n_t)$ de la función de distribución de probabilidad trama a trama, y con la hipótesis simplificadora de que el ruido anterior sea Gaussiano.

Según se aplica al ruido, el aprendizaje incremental de Bayes actualiza la distribución "anterior" actual acerca del ruido usando la posterior, dados los datos observados hasta el pasado más reciente, dado que esta posterior es la información más completa acerca del parámetro precedente al momento actual. Este procedimiento se ilustra en la FIG. 3, donde en una primera etapa una señal ruidosa 300 se divide en tramas. En la etapa 302, para cada trama se aplica el aprendizaje incremental de Bayes, donde una estimación del ruido de cada trama supone una distribución anterior del ruido variable a lo largo del tiempo, y la estimación del ruido se actualiza recursivamente usando una aproximación para el ruido posterior calculada en una trama de un momento anterior. Por lo tanto, la secuencia posterior en la Ec. 3 se convierte en una secuencia anterior variable a lo largo del tiempo (es decir, la evolución anterior) para parámetros distributivos del ruido de interés (con el desfase temporal de una trama en el tamaño). En una realización, la etapa 302 puede incluir calcular la probabilidad $p(y_t|n_t)$ de datos para la trama actual, usando a la vez una estimación del ruido en una trama precedente, preferiblemente la trama inmediatamente precedente, lo que supone la fluidez en la posterior, según lo indicado por la Ec. 1.

Para la probabilidad $p(y_t|n_t)$ de datos, que es no Gaussiana (y que se describirá en breve), la posterior es necesariamente no Gaussiana. Una aplicación sucesiva de la Ec. 1 daría como resultado una rápida combinación expansiva de las posteriores previas, y llevaría a formas inmanejables. Se necesitan aproximaciones para superar la inmanejabilidad. La aproximación que se usa es aplicar la expansión en serie de Taylor de primer orden para linealizar la relación no lineal entre y_t y n_t. Esto lleva a una forma Gaussiana de $p(y_t|n_t)$. Por lo tanto, la función de distribución de probabilidad $p(n_{\tau+1})$ del ruido variable a lo largo del tiempo, que se hereda de la posterior para la historia pasada de datos $p(n_\tau|y_1^\tau)$ puede aproximarse por la Gaussiana:

$$\begin{aligned} p(n_\tau|y_1^\tau) &= \frac{1}{(2\pi)^{1/2}\sigma_{n_\tau}} \exp\left[-\frac{1}{2}\left(\frac{n_\tau - \mu_{n_\tau}}{\sigma_{n_\tau}}\right)^2\right] \\ &\doteq \mathcal{N}[n_\tau; \mu_{n_\tau}, \sigma_{n_\tau}^2], \end{aligned} \quad (4)$$

donde μ_{nt} y σ^2_{nt} se llaman los hiperparámetros (media y varianza) que caracterizan la función anterior de distribución de probabilidad. Luego la secuencia posterior en la Ec. 3, calculada a partir de la regla recursiva de Bayes de la Ec. 1 ofrece una manera sensata de determinar la evolución temporal de los hiperparámetros, lo que se describe más adelante.

Se proporcionarán ahora los modelos de distorsión acústica y de habla neta para calcular la probabilidad $p(y_t|n_t)$ de datos. Primero supongamos un modelo mezcla de Gaussianos invariante en el tiempo para espectros de registro del habla neta χ:

$$p(x) = \sum_m p(m)\mathcal{N}[x; \mu_x(m), \sigma_x^2(m)]. \quad (5)$$

Puede usarse luego un sencillo modelo de distorsión acústica no lineal en el dominio de registro-espectral:

$$\exp(y) = \exp(x) + \exp(n), \quad y = x + g(n - x), \quad (6)$$

donde la función no lineal es:

$$g(z) = \log[1 + \exp(z)].$$

A fin de obtener una forma útil para la probabilidad $p(y_t|n_t)$ de datos, se usa una expansión en serie de Taylor para linealizar la no linealidad *g* en la Ec. 6. Esto da el modelo linealizado de

$$y \approx x + g(n_0 - \mu_x(m_0)) + g'(n_0 - \mu_x(m_0))(n - n_0), \quad (7)$$

donde n_0 es el punto de expansión de la serie de Taylor y el coeficiente de expansión de la serie de primer orden puede calcularse fácilmente como:

$$g'(n_0 - \mu_x(m_0)) = \frac{\exp(n_0)}{\exp[\mu_x(m_0)] + \exp(n_0)}.$$

Al evaluar las funciones *g* y *g'* en la Ec. 7, el valor χ de habla neta se toma como la media ($\mu_\chi(m_0)$) del componente Gaussiano m_0 de mezcla "óptima".

La Ec. 7 define una transformación lineal de las variables aleatorias χ a *y* (después de fijar *n*). En base a esta transformación, obtenemos la función de distribución de probabilidad sobre γ a continuación, a partir de la función de distribución de probabilidad sobre χ (Ec. 5) con una aproximación de Laplace:

$$\begin{aligned} p(y_t|n_t) &= \sum_m p(m)\mathcal{N}[y_t; \mu_y(m,t), \sigma_y^2(m,t)] \\ &\approx \mathcal{N}[y_t; \mu_y(m_0,t), \sigma_y^2(m_0,t)], \end{aligned} \quad (8)$$

donde el componente de mezcla óptima está determinado por

$$m_0 = \arg\max_m \mathcal{N}[y_t; \mu_y(m,t), \sigma_y^2(m,t)],$$

y donde la media y la varianza de las Gaussianas aproximadas son

$$\begin{aligned} \mu_y(m_0,t) &= \mu_x(m_0) + g_{m_0} + g'_{m_0} \times (n_t - n_0) \\ \sigma_y^2(m_0,t) &= \sigma_x^2(m_0) + g'^2_{m_0}\sigma_{n_t}^2. \end{aligned} \quad (9)$$

Como se mostrará más adelante, la estimación Gaussiana para $p(y_t|n_t)$ se usa para desarrollar ese algoritmo. Aunque lo precedente usó una expansión en serie de Taylor y una aproximación de Laplace para proporcionar una estimación Gaussiana para $p(y_t|n_t)$, debería entenderse que pueden usarse otras técnicas para proporcionar una estimación Gaussiana sin apartarse de la presente invención. Por ejemplo, además de usar una aproximación de Laplace en la Ec. 8, pueden usarse técnicas numéricas para la aproximación o un modelo de mezcla Gaussiana (con un número pequeño de componentes).

Puede proporcionarse ahora un algoritmo para estimar la media variable a lo largo en el tiempo y la varianza en el ruido anterior. Dada la forma Gaussiana aproximada para $p(y_t|n_t)$, como en la Ec. 8, y para $p(n_t|y^\tau_1)$, como en la Ec. 4, puede proporcionarse el algoritmo para determinar la evolución anterior del ruido, expresada como estimaciones secuenciales de los hiperparámetros variables a lo largo del tiempo de la media μ_{nt} y la varianza σ^2_{nt}. Reemplazando las Ec. 4 y 8 en la Ec. 1, puede obtenerse lo siguiente:

$$\begin{aligned} & \mathcal{N}(n_t; \mu_{n_t}, \sigma^2_{n_t}) \\ \propto \quad & \mathcal{N}[y_t; \mu_y(m_0, t), \sigma^2_y(m_0, t)]\mathcal{N}(n_{t-1}; \mu_{n_{t-1}}, \sigma^2_{n_{t-1}}) \\ \approx \quad & \mathcal{N}[g'_{m_0} n_{t-1}; \mu_1, \sigma^2_y(m_0, t)]\mathcal{N}(n_{t-1}; \mu_{n_{t-1}}, \sigma^2_{n_{t-1}}) \quad (10) \end{aligned}$$

donde $\mu_1 = y_t - \mu_x(m_0) - g_{m0} + g'_{m0}\, n_0$, y se usó la hipótesis de fluidez del ruido. Las medias y varianzas, respectivamente, de los lados izquierdo y derecho coinciden en la Ec. 10 para obtener las fórmulas de evolución anterior:

$$\mu_{n_t} = \frac{g'_{m_0}\ \bar{\mu}_1\ \sigma^2_{n_{t-1}} + \mu_{n_{t-1}}\sigma^2_y(m_0, t-1)}{g'^2_{m_0}\sigma^2_{n_{t-1}} + \sigma^2_y(m_0, t-1)}, \qquad (11)$$

$$\sigma^2_{n_t} = \frac{\sigma^2_y(m_0, t-1)\sigma^2_{n_{t-1}}}{g'^2_{m_0}\sigma^2_{n_{t-1}} + \sigma^2_y(m_0, t-1)},$$

donde $\bar{\mu}_1 = y_t - \mu_x(m_0) - g_{m0} + g'_{m0}\, \mu_{nt\text{-}1}$. Al establecer la Ec. 11, se usa la media anterior del momento previo como el punto de expansión de la serie de Taylor para el ruido; es decir, $n_0 = \mu_{nt\text{-}1}$. También se usó el resultado, bien establecido, en el cálculo Gaussiano (fijando $a_1 = g'_{m0}$):

$$\mathcal{N}(ax; \mu_1, \sigma_1^2)\mathcal{N}(x; \mu_2, \sigma_2^2) = \frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 + K\right],$$

donde

$$\mu = \frac{a\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{a^2\sigma_2^2 + \sigma_1^2}; \quad \sigma^2 = \frac{\sigma_1^2\sigma_2^2}{a^2\sigma_2^2 + \sigma_1^2}.$$

En base a un conjunto de hipótesis simplificadas pero efectivas, se usa el apareo aproximado recursivo del término cuadrático de la regla de Bayes para obtener con éxito las fórmulas de evolución anterior del ruido, según se resume en la Ec. 11. Se ha hallado que la estimación media del ruido es medida más precisamente por la reducción de error del Sistema de Medición de Distancias, mientras que la información de varianza puede usarse para proporcionar una dosis de fiabilidad.

Las técnicas de estimación del ruido descritas anteriormente pueden usarse en una técnica de normalización del ruido o en la eliminación del ruido, según lo expuesto en una solicitud de patente titulada PROCEDIMIENTO DE REDUCCIÓN DEL RUIDO USANDO VECTORES DE CORRECCIÓN EN BASE A ASPECTOS DINÁMICOS DEL HABLA Y LA NORMALIZACIÓN DEL RUIDO, solicitud de Nº de Serie 10 / 117.142, registrada el 5 de abril de 2002. La invención también puede usarse más directamente como parte de un sistema de reducción del ruido en el cual el ruido estimado identificado para cada trama se elimina de la señal ruidosa para producir una señal limpia tal como se describe en la solicitud de patente titulada MODELO NO LINEAL DE OBSERVACIÓN PARA ELIMINAR EL RUIDO DE SEÑALES CORROMPIDAS, solicitud de Nº de Serie 10 / 237.163, registrada el 6 de septiembre de 2002.

La FIG. 4 proporciona un diagrama en bloques de un entorno en el cual puede utilizarse la técnica de estimación del ruido de la presente invención para efectuar la reducción del ruido. En particular, la FIG. 4 muestra un sistema de reconocimiento del habla en el cual puede usarse la técnica de estimación del ruido de la presente invención para reducir el ruido en una señal de entrenamiento usada para entrenar un modelo acústico y / o para reducir el ruido en una señal de prueba que se aplica ante un modelo acústico para identificar el contenido lingüístico de la señal de prueba.

En la FIG. 4, un altavoz 400, bien un entrenador o bien un usuario, habla hacia un micrófono 404. El micrófono 404 también recibe ruido adicional desde una o más fuentes 402 de ruido. Las señales de audio detectadas por el micrófono 404 se convierten en señales eléctricas que se suministran al convertidor 406 de analógico a digital.

Aunque el ruido adicional 402 se muestra como ingresando a través del micrófono 404 en la realización de la FIG. 4, en otras realizaciones, el ruido adicional 402 puede añadirse a la señal de habla de entrada como una señal digital después del convertidor 406 de analógico a digital.

El convertidor 406 de analógico a digital convierte la señal analógica del micrófono 404 en una serie de valores digitales. En varias realizaciones, el convertidor 406 de analógico a digital muestrea la señal analógica a 16 kHz y 16 bits por muestra, creando por ello 32 kilooctetos de datos de habla por segundo. Estos valores digitales se suministran a un constructor 407 de tramas que, en una realización, agrupa los valores en tramas de 25 milisegundos que comienzan cada 10 milisegundos.

Las tramas de datos creados por el constructor 407 de tramas se suministran al extractor 408 de características, que extrae una característica de cada trama. Los ejemplos de módulos de extracción de características incluyen módulos para realizar la Codificación Predictiva Lineal (LPC), el cepstrum obtenido de la LPC, la Predicción Lineal Perceptiva (PLP), la extracción de características del modelo de Auditorio y la extracción de características de Coeficientes de Mel-Cepstrum de Frecuencia (MFCC). Obsérvese que la invención no se limita a estos módulos de extracción de características y que pueden usarse otros módulos dentro del contexto de la presente invención.

El módulo de extracción de características produce un flujo de vectores de características, cada uno de los cuales está asociado a una trama de la señal de habla. Este flujo de vectores de características se suministra al módulo 410 de reducción del ruido, que usa la técnica de estimación del ruido de la presente invención para estimar el ruido en cada trama.

La salida del módulo 410 de reducción del ruido es una serie de vectores “limpios” de características. Si la señal de entrada es una señal de entrenamiento, esta serie de vectores “limpios” de características se suministra a un entrenador 424, que usa los vectores “limpios” de características y un texto 426 de entrenamiento para entrenar un modelo acústico 418. Las técnicas para entrenar tales modelos son conocidas en la técnica, y no se requiere una descripción de ellas para la comprensión de la presente invención.

Si la señal de entrada es una señal de prueba, los vectores “limpios” de características se suministran a un descodificador 412, que identifica una secuencia más probable de palabras en base al flujo de vectores de características, un léxico 414, un modelo lingüístico 416 y el modelo acústico 418. El procedimiento específico usado para la descodificación no es importante para la presente invención y puede usarse cualquiera de varios procedimientos conocidos para la descodificación.

La secuencia más probable de palabras hipotéticas se proporciona a un módulo 420 de medición de confianza. El módulo 420 de medición de confianza identifica qué palabras son las más probables de haber sido indebidamente identificadas por el reconocedor del habla, en base, en parte, a un modelo acústico secundario (no mostrado). El módulo 420 de medición de confianza proporciona luego la secuencia de palabras hipotéticas a un módulo 422 de salida, junto con identificadores que indican qué palabras pueden haber sido indebidamente identificadas. Los expertos en la técnica reconocerán que el módulo 420 de medición de confianza no es necesario para la puesta en práctica de la presente invención.

Aunque la FIG. 4 ilustra un sistema de reconocimiento del habla, la presente invención puede usarse en cualquier sistema de reconocimiento de patrones y no está limitada al habla.

Aunque la presente invención ha sido descrita con referencia a realizaciones específicas, los operarios expertos en la técnica reconocerán que pueden hacerse cambios en la forma y en el detalle sin apartarse del alcance de la invención.

REIVINDICACIONES

1. Un procedimiento para estimar el ruido en una señal ruidosa, comprendiendo el procedimiento:

dividir (300) la señal ruidosa en tramas; y

determinar (302) una estimación del ruido para una trama usando el aprendizaje incremental de Bayes, basándose la estimación del ruido en una aproximación Gaussiana, e incluyendo parámetros que definen tanto una media como una varianza de la distribución anterior del ruido, donde se supone una distribución anterior del ruido variable a lo largo del tiempo, y se actualiza recursivamente una estimación del ruido usando una aproximación para el ruido posterior calculado en una trama precedente, en base a una aplicación iterativa de la regla de Bayes.

2. El procedimiento de la reivindicación 1, en el cual la determinación de una estimación del ruido comprende:

determinar una estimación del ruido para una primera trama de la señal ruidosa usando una aproximación para el ruido posterior calculado en una trama precedente;

determinar una estimación de probabilidad de datos para una segunda trama de la señal ruidosa; y

usar la estimación de probabilidad de datos para la segunda trama y la estimación de ruido para la primera trama, a fin de determinar una estimación del ruido para la segunda trama.

3. El procedimiento de la reivindicación 2, en el cual la determinación de la estimación de probabilidad de datos para la segunda trama comprende usar la estimación de probabilidad de datos para la segunda trama en una ecuación que se basa en parte en una definición de la señal ruidosa como una función no lineal de una señal limpia y de una señal ruidosa.

4. El procedimiento de la reivindicación 3, en el cual la ecuación se basa adicionalmente en una aproximación a la función no lineal.

5. El procedimiento de una de las reivindicaciones 2 a 4, en el cual la aproximación es igual a la función no lineal en un punto definido en parte por la estimación del ruido para la primera trama.

6. El procedimiento de la reivindicación 5, en el cual la aproximación es una expansión en serie de Taylor.

7. El procedimiento de la reivindicación 6, en el cual la aproximación comprende adicionalmente adoptar una aproximación de Laplace.

8. El procedimiento de una de las reivindicaciones 2 a 4, en el cual el uso de la estimación de probabilidad de datos para la segunda trama comprende usar la estimación de ruido para la primera trama como un punto de expansión para una expansión en serie de Taylor de una función no lineal.

9. El procedimiento de una de las reivindicaciones 1 a 4, en el cual el uso de una aproximación para el ruido posterior comprende usar una aproximación Gaussiana.

10. El procedimiento de la reivindicación 1, en el cual la determinación de la estimación del ruido comprende determinar sucesivamente una estimación del ruido para cada trama.

11. El procedimiento de la reivindicación 1, en el cual la etapa de determinación comprende:

estimar sucesivamente para cada trama el ruido en cada trama, de modo tal que una estimación del ruido para una trama actual se base en una aproximación Gaussiana de la probabilidad de datos para la trama actual y en una aproximación Gaussiana del ruido en una secuencia de tramas anteriores.

12. El procedimiento de la reivindicación 11, en el cual la estimación del ruido en cada trama comprende usar una ecuación que se basa en parte en una definición de la señal ruidosa como una función no lineal de una señal limpia y de una señal ruidosa, para determinar la aproximación de la probabilidad de datos en la trama actual.

13. El procedimiento de la reivindicación 12, en el cual la ecuación se basa adicionalmente en una aproximación a la función no lineal.

14. El procedimiento de la reivindicación 13, en el cual la aproximación es igual a la función no lineal en un punto definido en parte por la estimación del ruido para la trama anterior.

15. El procedimiento de la reivindicación 14, en el cual la aproximación es una expansión en serie de Taylor.

16. El procedimiento de la reivindicación 15, en el cual la aproximación incluye adicionalmente una aproximación de

Laplace.

17. Un medio legible por ordenador que incluye instrucciones legibles por un ordenador que, cuando se implementan, causan que el ordenador realice cualquiera de los procedimientos de las reivindicaciones 1 a 16.

18. Un sistema adaptado para realizar uno cualquiera de los procedimientos de las reivindicaciones 1 a 16.

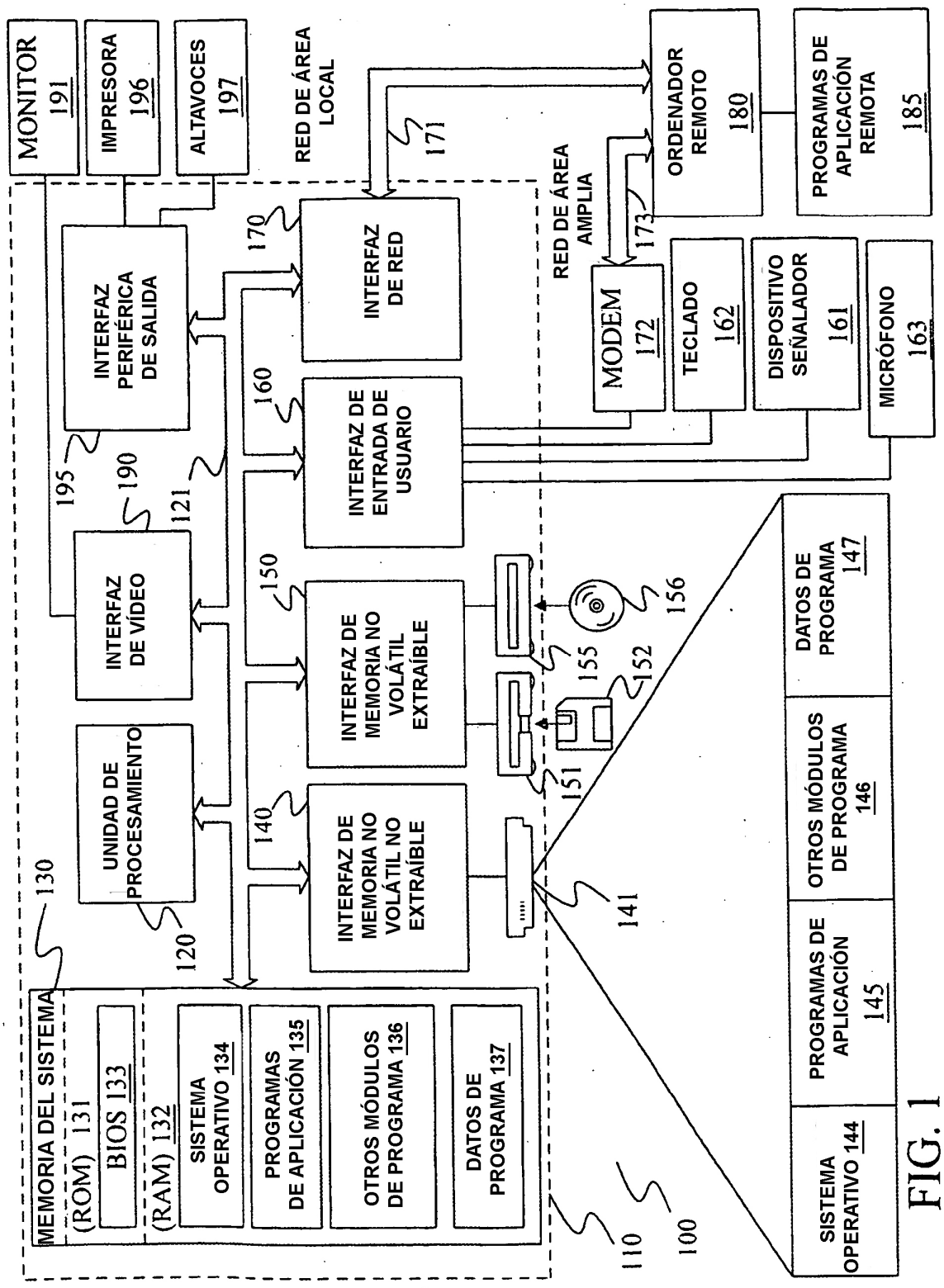
MONITOR 191
IMPRESORA 196
ALTAVOCES 197
RED DE ÁREA LOCAL
171
ORDENADOR REMOTO 180
PROGRAMAS DE APLICACIÓN REMOTA 185
RED DE ÁREA AMPLIA
173
INTERFAZ PERIFÉRICA DE SALIDA
170
INTERFAZ DE RED
MODEM 172
TECLADO 162
DISPOSITIVO SEÑALADOR 161
MICRÓFONO 163
195
190
121
160
INTERFAZ DE ENTRADA DE USUARIO
INTERFAZ DE VÍDEO
150
INTERFAZ DE MEMORIA NO VOLÁTIL EXTRAÍBLE
156
155
152
151
UNIDAD DE PROCESAMIENTO
140
INTERFAZ DE MEMORIA NO VOLÁTIL NO EXTRAÍBLE
130
120
141
MEMORIA DEL SISTEMA
(ROM) 131
BIOS 133
(RAM) 132
SISTEMA OPERATIVO 134
PROGRAMAS DE APLICACIÓN 135
OTROS MÓDULOS DE PROGRAMA 136
DATOS DE PROGRAMA 137
DATOS DE PROGRAMA 147
OTROS MÓDULOS DE PROGRAMA 146
PROGRAMAS DE APLICACIÓN 145
SISTEMA OPERATIVO 144
110
100

FIG. 1

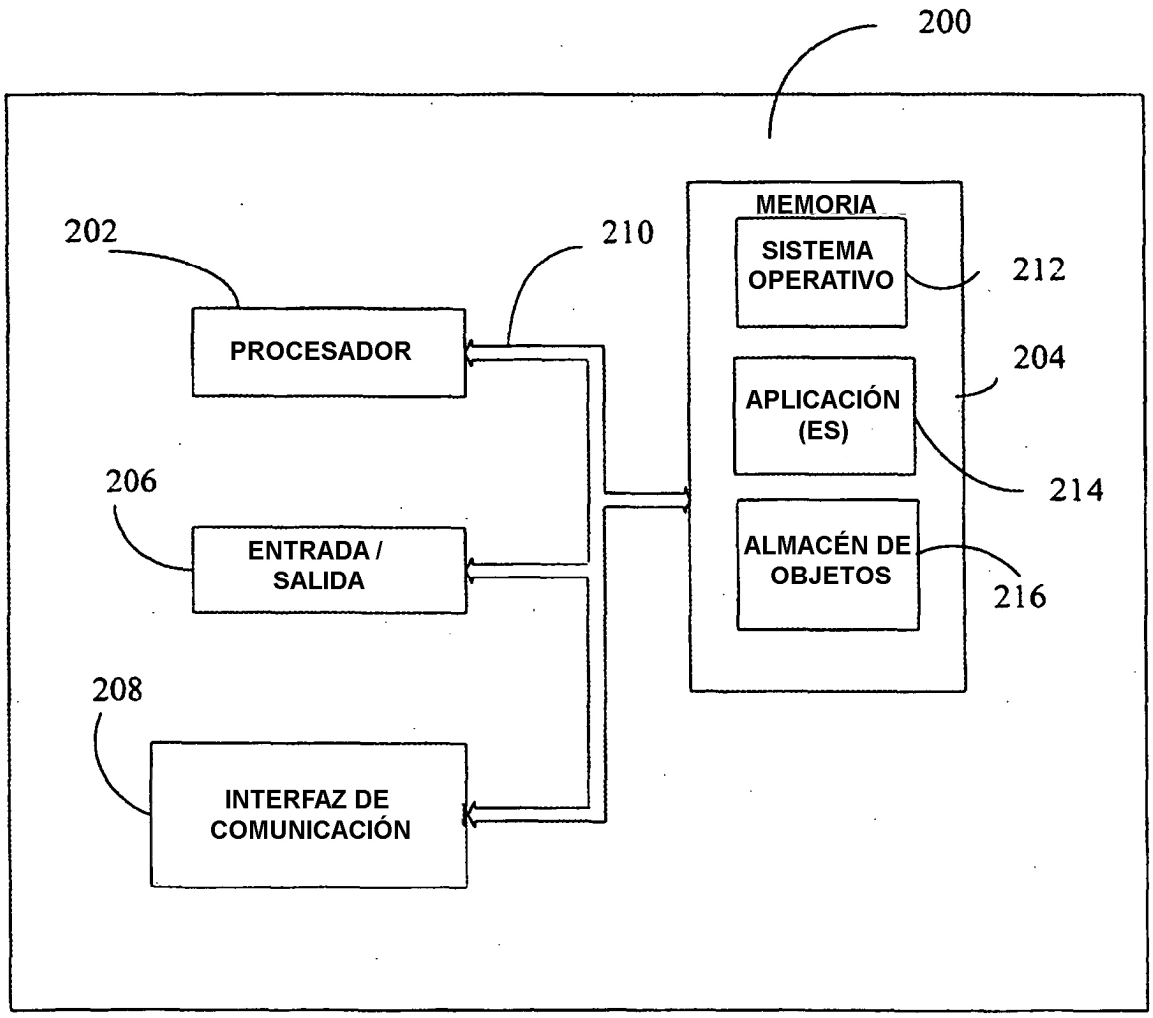
200
MEMORIA
SISTEMA OPERATIVO
212
204
APLICACIÓN (ES)
214
ALMACÉN DE OBJETOS
216
202
PROCESADOR
210
206
ENTRADA / SALIDA
208
INTERFAZ DE COMUNICACIÓN

FIG. 2

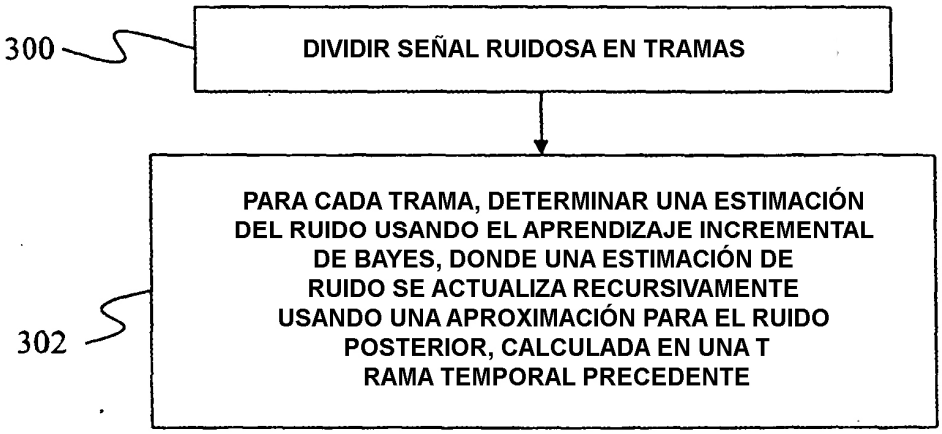
300
DIVIDIR SEÑAL RUIDOSA EN TRAMAS
302
PARA CADA TRAMA, DETERMINAR UNA ESTIMACIÓN DEL RUIDO USANDO EL APRENDIZAJE INCREMENTAL DE BAYES, DONDE UNA ESTIMACIÓN DE RUIDO SE ACTUALIZA RECURSIVAMENTE USANDO UNA APROXIMACIÓN PARA EL RUIDO POSTERIOR, CALCULADA EN UNA T
RAMA TEMPORAL PRECEDENTE

FIG. 3

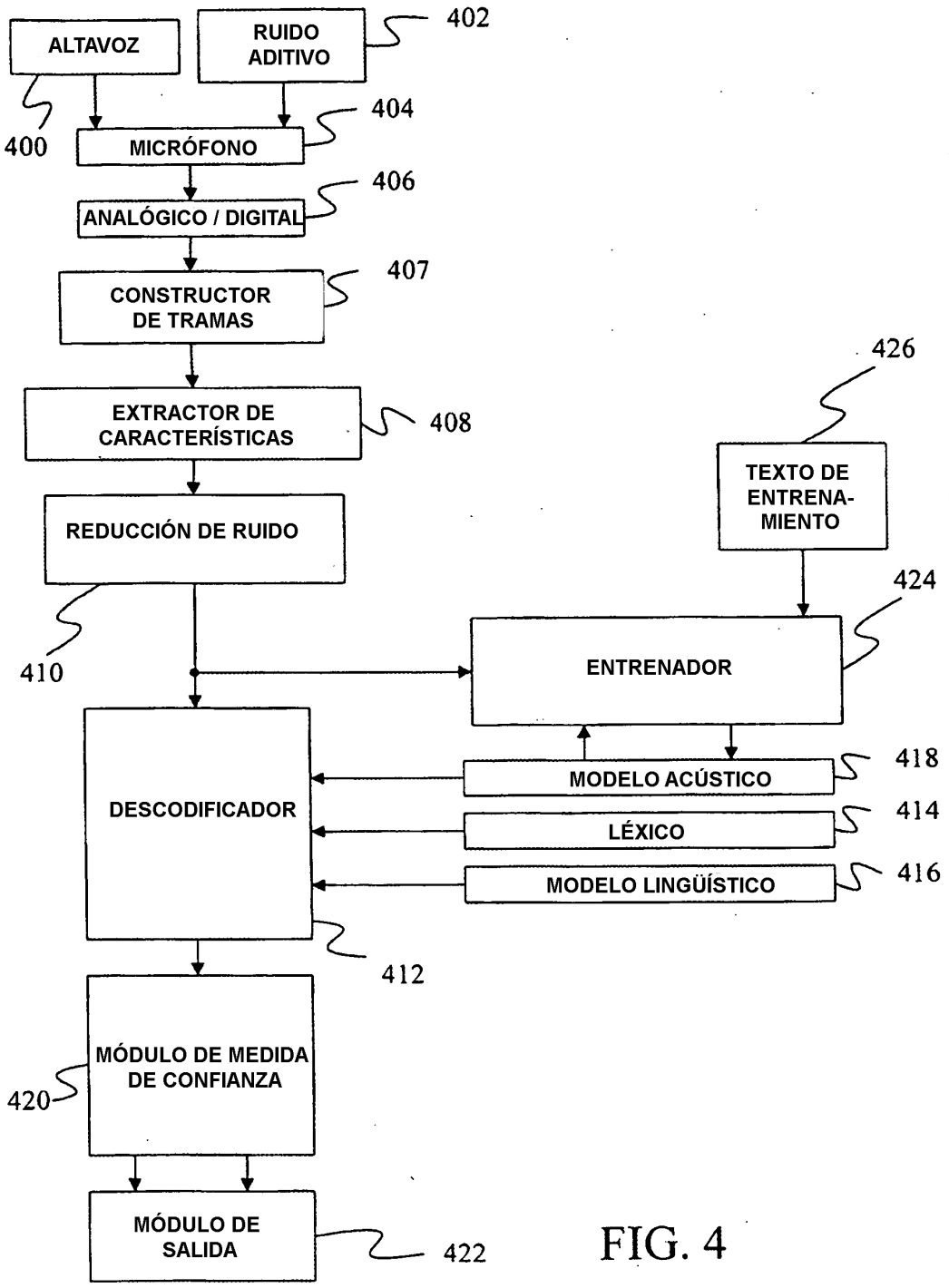
ALTAVOZ
RUIDO ADITIVO
402
400
MICRÓFONO
404
ANALÓGICO / DIGITAL
406
CONSTRUCTOR DE TRAMAS
407
EXTRACTOR DE CARACTERÍSTICAS
408
REDUCCIÓN DE RUIDO
410
426
TEXTO DE ENTRENA-MIENTO
424
ENTRENADOR
DESCODIFICADOR
MODELO ACÚSTICO
418
LÉXICO
414
MODELO LINGÜÍSTICO
416
412
MÓDULO DE MEDIDA DE CONFIANZA
420
MÓDULO DE SALIDA
422

FIG. 4