



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 338 418**

51 Int. Cl.:
G10L 15/20 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04100134 .8**

96 Fecha de presentación : **16.01.2004**

97 Número de publicación de la solicitud: **1445759**

97 Fecha de publicación de la solicitud: **11.08.2004**

54 Título: **Procedimiento adaptivo al usuario para la modelización de ruidos.**

30 Prioridad: **10.02.2003 DE 103 05 369**

45 Fecha de publicación de la mención BOPI:
07.05.2010

45 Fecha de la publicación del folleto de la patente:
07.05.2010

73 Titular/es: **SIEMENS AKTIENGESELLSCHAFT**
Wittelsbacherplatz 2
80333 München, DE

72 Inventor/es: **Major, Andreas Ralph y**
Schröer, Andreas

74 Agente: **Zuazo Araluze, Alexander**

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

Procedimiento adaptivo al usuario para la modelización de ruidos.

5 Los procedimientos técnicos para reconocer la voz humana se utilizan hoy día en múltiples aparatos móviles como por ejemplo teléfonos móviles, asistentes personales digitales (Personal Digital Assistants, PDAs) y wearables o portables (reproductor MP3, terminales móviles-relojes, etc.). Al respecto se utiliza entre otros un reconocimiento de voz independiente del orador basado en la modelización Hidden-Markov (HMM) de los fonemas (o de otras unidades de sonido) de una voz y del fondo de ruido o del silencio.

10 Entonces dan lugar los dos puntos que siguen a que la calidad del reconocimiento se vea fuertemente perjudicada: En primer lugar, poseen los canales de audio (micrófono y convertidor analógico/digital) de los aparatos antes citados típicamente una curva característica muy diferente, por ejemplo con respecto a la resolución y precisión de la conversión A/D y al ruido que resulta en una toma de audio. En segundo lugar, se utilizan todos los aparatos citados en entornos acústicamente muy diferentes, es decir, por ejemplo en el automóvil, en la oficina, en la calle, etcétera. Por
15 ello es muy difícil modelizar el fondo de ruido en el reconocimiento de voz.

Un procedimiento estándar para la modelización del fondo de ruido consiste en que en la generación de los datos del modelo se tengan en cuenta muchas expresiones con distinto fondo de ruido. Este llamado proceso de entrenamiento
20 se realiza entonces sobre un banco de datos que contiene tomas procedentes de diversos entornos acústicos, como por ejemplo automóvil, oficina, etcétera. El inconveniente entonces es que las distintas curvas características de los canales de audio no pueden ser captadas por los aparatos móviles y que es imposible representar de manera adecuada todos los entornos acústicos.

25 Otra posibilidad adicional consiste en captar modelizaciones para ruidos especiales en el vocabulario activo del reconocedor (modelización Garbage o de desecho). No obstante, este tipo de modelización tiene el inconveniente de que sólo antes y después de la palabra a reconocer pueden reproducirse ruidos secundarios sobre estas palabras Garbage o de desecho. El reconocimiento de la palabra pronunciada se realiza a continuación dificultado por ruidos de fondo.

30 Otra posibilidad utilizada en muchos casos para asegurar una buena modelización del ruido de fondo son procedimientos para reducir el ruido que intentan suprimir ruidos de fondo en los datos de audio. No obstante, la reducción de los ruidos no está en competencia con los demás procedimientos aquí descritos, ya que los mismos no se excluyen mutuamente y la reducción de ruidos se basa en otra forma de proceder totalmente distinta: La misma modifica los datos de audio y no formula ninguna hipótesis sobre la modelización del ruido de fondo.

35 Además es posible captar la acústica de una sala y con ello adaptar el canal de audio un poco a una sala muy determinada.

Por el documento de patente US, 5,970,446 se conoce la consideración en el reconocimiento de voz de un perfil de
40 ruido de fondo adecuado al actual fondo de ruido.

Partiendo de ello, la invención tiene como tarea básica incrementar la calidad del reconocimiento de voz y garantizar un reconocimiento de voz seguro, en particular también con diferentes fondos de ruido.

45 Esta tarea se resuelve mediante las invenciones indicadas en las reivindicaciones independientes. Ventajosas configuraciones mejoradas resultan de las reivindicaciones subordinadas.

Correspondientemente, se utilizan en un procedimiento para el reconocimiento de voz automático, apoyado por ordenador, varios perfiles de fondos de ruido, que representan respectivos fondos de ruido diferentes. Cuando se
50 encuentra en el banco de datos con perfiles de ruido de fondo un perfil de ruido de fondo adecuado como modelo de ruido para el fondo de ruido actual, se elige este perfil de ruido de fondo y se realiza el reconocimiento de voz teniendo en cuenta el perfil de ruido de fondo. Cuando en el banco de datos no se encuentra ningún perfil de ruido de fondo adecuado como modelo de ruido para el actual fondo de ruido, se confecciona un perfil de ruido de fondo para este fondo de ruido y el reconocimiento de voz se realiza considerando el perfil de ruido de fondo confeccionado. La
55 elección de los perfiles de ruido de fondo a partir del banco de datos se realiza preferiblemente de forma automática en base a determinadas prescripciones, pero también puede realizarse manualmente por parte de un usuario. Estas prescripciones pueden contener y/o tener en cuenta por ejemplo los ruidos de fondo medidos en ese momento, pero también alternativa o adicionalmente otros parámetros, como por ejemplo la hora del día o un determinado entorno en el que se encuentra en ese momento el dispositivo que realiza reconocimiento de voz.

60 El reconocimiento de voz se realiza entonces considerando el perfil de ruido de fondo elegido.

Preferiblemente se analiza para seleccionar el perfil de ruido de fondo el fondo de ruido actual y se elige el perfil de ruido de fondo considerando este análisis. Preferiblemente se elige entonces naturalmente el perfil de ruido de fondo
65 que representa mejor el fondo de ruido actual.

Alternativa o adicionalmente puede no obstante también realizar el usuario la elección del perfil de ruido de fondo o influir sobre la misma.

Otra posibilidad alternativa o adicional consiste en la consideración de otros parámetros. Así puede por ejemplo estar insertado el dispositivo con el que se realiza el reconocimiento de voz en un soporte de fijación de un vehículo. En el procedimiento se detecta entonces que el dispositivo está colocado en el soporte de fijación de un vehículo, para lo cual el dispositivo dispone de los correspondientes medios. A continuación se elige un perfil de ruido de fondo que en particular esté adaptado para un reconocimiento de voz en un automóvil.

Preferiblemente pueden confeccionarse los perfiles de ruido de fondo mediante el dispositivo utilizado para el procedimiento. Para ello se capta un fondo de ruido y el dispositivo confecciona un perfil de ruido de fondo para este fondo de ruido. De esta manera puede adaptarse el procedimiento a los más diversos fondos de ruido y con ello adaptarse a las necesidades del usuario.

Además es ventajoso que se detecte con ayuda de un Voice Activity Detector (VAD, detector de actividad de voz) si existe un fondo con ruido o si existe voz. Esta detección puede realizarse tanto al elegir un perfil de ruido de fondo como también al confeccionar un nuevo perfil de ruido de fondo.

Un dispositivo que está preparado para realizar el procedimiento antes descrito presenta en particular respectivos medios que están equipados para realizar las distintas etapas del procedimiento. Configuraciones mejoradas preferentes del dispositivo resultan análogamente a las configuraciones mejoradas preferentes del procedimiento. El dispositivo es preferiblemente un aparato terminal móvil en forma de un teléfono móvil, PDA o wearable (portable).

Un producto de programa para una instalación de tratamiento de datos que contiene secciones de códigos, con las que puede ejecutarse uno de los procedimientos descritos sobre la instalación de tratamiento de datos, puede ejecutarse mediante implementación adecuada del procedimiento en un lenguaje de programación y conversión a un código que pueda ejecutar la instalación de tratamiento de datos. Las secciones de códigos se memorizan para ello. Al respecto se entiende bajo un producto de programa el programa como producto comercializable. El mismo puede existir en cualquier forma, como por ejemplo en papel, en un soporte de datos legible por ordenador o distribuido a través de una red.

Otras ventajas y características esenciales de la invención resultan de la descripción de un ejemplo de ejecución en base al dibujo. Al respecto muestra

figura 1 la señal de audio de las palabras “encuentro en la estación”;

figura 2 un diagrama secuencial para la confección de un perfil de ruido de fondo para un fondo de ruido;

figura 3 un diagrama secuencial para un procedimiento para el reconocimiento de voz.

En el procedimiento para el reconocimiento de voz se entrenan con la mayor precisión posible en relación con el canal de audio y el perfil de usuario perfiles de ruido de fondo adecuados (modelos de silencio) en un aparato terminal. Estos perfiles de ruido de fondo se utilizan entonces en el reconocimiento. Se procede en dos etapas, que a continuación se describirán con exactitud.

Primeramente inicia el usuario del aparato terminal un entrenamiento de ruido, con lo que se capta el fondo de ruido actual. En una realización de este entrenamiento en una sala silenciosa, corresponde la captación al ruido del canal de audio. Si adicionalmente se inicia una captación durante por ejemplo una visita a una feria, refleja la captación el ruido de la feria inclusive el ruido del canal de audio. Pueden confeccionarse adicionalmente otras cualesquiera captaciones, por ejemplo en el automóvil, en el tren, en una calle transitada, etcétera. Las captaciones de ruido de fondo pueden realizarlas el usuario en cualquier momento y en cualquier lugar.

La figura 1 muestra el proceso de la adaptación del ruido, que se inicia explícitamente por parte del usuario. Una vez que el usuario ha iniciado el entrenamiento, puede el mismo detener de nuevo la captación de ruido tras un cierto espacio de tiempo. Al respecto puede tratarse de varios segundos hasta varios minutos. Una vez que el usuario ha detenido de nuevo la captación, se modeliza el ruido y se memoriza el perfil de ruido de fondo así obtenido en un banco de datos.

En cada proceso de reconocimiento del reconocedor de voz del aparato terminal se consideran y comparan todas las captaciones de entorno existentes o bien los perfiles de ruido de fondo que representan a las mismas para el fondo de ruido actual en cada caso (modelización del silencio). Esto se realiza de la siguiente manera: cuando inicia el usuario el reconocimiento de voz, se analiza primeramente el fondo de ruido. Para ello puede aprovecharse que por lo general dura de uno a dos segundos hasta que el usuario comienza a hablar.

Un ejemplo de ello se indica en la figura 1. Aquí puede reconocerse la señal de las palabras “encuentro en la estación”. La misma contiene pausas claras al comienzo y al final de toda la captación, así como pausas más cortas entre las distintas palabras. Al comienzo de la forma de onda, está marcada la zona más segura A para el reconocimiento del fondo de ruido. Esta zona contiene el ruido del canal de audio, así como el actual fondo de ruido. La zona se analiza para el reconocimiento de ruido. La diferencia entre sí una forma de onda es ruido o voz puede detectarse por ejemplo mediante la energía de la posición en la captación.

ES 2 338 418 T3

5 El comienzo de la voz en la expresión puede encontrarse con un procedimiento para la detección de voz en forma de un Voice Activity Detector (detector de actividad de voz), por ejemplo mediante umbrales de energía. En base a la captación así obtenida del actual fondo de ruido al comienzo del proceso actual de reconocimiento, se elige el perfil de ruido de fondo que mejor encaja de entre los perfiles existentes. Con el perfil de ruido de fondo ahora elegido se realiza el reconocimiento. Esto da lugar a que en fondos de ruido típicos para el usuario no se presenten ninguno o pocos reconocimientos incorrectos.

10 Si el fondo de ruido actual no es similar a ninguno de los ruidos ya captados, puede recomendársele al usuario entrenar con este fondo de ruido. No obstante, esto puede realizarse a elección automáticamente, es decir sin consultar. La recomendación para el usuario puede realizarse visualmente y/o mediante señal de audio.

15 La secuencia esquemática del reconocimiento de ruido se representa de nuevo en base a la figura 3. Aquí arranca el usuario el reconocedor de voz. La captación se realiza como fondo de ruido hasta que se detecta que a partir del tramo actual de la captación, se trata de voz. Para ello se utiliza un Voice Activity Detector (detector de actividad de voz). Entonces se decide si en el banco de datos se encuentra ya un perfil de ruido de fondo adecuado como modelo de ruido o no. Si existe el perfil de ruido de fondo, se elige éste para el reconocimiento y comienza el reconocimiento propiamente dicho de las palabras expresadas por el usuario. Si por el contrario no existe ningún perfil de ruido de fondo adecuado, se recomienda al usuario modelizar el ruido y registrarlo en la base de datos. Alternativamente puede realizarse la modelización del fondo de ruido desconocido automáticamente. Si el usuario elige el entrenamiento, se modeliza el fondo de ruido y a continuación se realiza inmediatamente el reconocimiento de la expresión con el nuevo modelo. Si el usuario rechaza la modelización, se realiza el proceso de reconocimiento en base a un modelo por defecto.

25 La invención posee las ventajas de que un sistema de reconocimiento de voz alcanza mejores resultados, ya que

- la calidad de un canal de audio malo de un aparato tiene un papel secundario en el proceso de reconocimiento. Esto es debido a que el canal de audio, inclusive el entorno de ruido, se modelizan;
- los entornos de ruido se tienen en cuenta a la vez en el proceso de reconocimiento, lo cual reduce el empeoramiento del resultado del reconocimiento causado por los mismos. El reconocimiento de voz funciona óptimamente en los entornos de ruido dinámicos definidos por el usuario, ya que los mismos se modelizan a la vez,
- los perfiles de ruido de fondo, una vez se han registrado y entrenado, los elige automáticamente el programa de reconocimiento de voz para el fondo de ruido existente en ese momento. Esto está en contraposición con el comportamiento de los reconocedores utilizados hasta ahora, que en todo caso posibilitan la elección explícita, cualitativa de un entorno de ruido no específico del usuario y del canal de audio.

40 Es característica la posibilidad de definición individual de fondos de ruido por parte del usuario, lo que posibilita una calidad de reconocimiento óptima, en el entorno de utilización típico para él mismo de un aparato móvil.

45

50

55

60

65

REIVINDICACIONES

1. Procedimiento para el reconocimiento de voz con un banco de datos con varios perfiles de ruido de fondo, que
5 representan distintos fondos de ruido,

- en el que cuando en el banco de datos se encuentra un perfil de ruido de fondo adecuado como modelo de
ruido para el fondo de ruido actual, se elige este perfil de ruido de fondo y se realiza el reconocimiento de
voz teniendo en cuenta el perfil de ruido de fondo,

10 - en el que cuando en el banco de datos no se encuentra ningún perfil de ruido de fondo adecuado como
modelo de ruido para el actual fondo de ruido, se confecciona un perfil de ruido de fondo para este fondo
de ruido y el reconocimiento de voz se realiza considerando el perfil de ruido de fondo confeccionado.

15 2. Procedimiento según la reivindicación 1,

- en el que se analiza el fondo de ruido actual,

20 - en el que el perfil de ruido de fondo se elige teniendo en cuenta este análisis.

3. Procedimiento según una de las reivindicaciones precedentes,

25 en el que con ayuda de un Voice-Activity-Detector (detector de actividad de voz) se detecta si existe un fondo de
ruido.

30 4. Dispositivo, en particular aparato terminal móvil, equipado para realizar un procedimiento según una de las
reivindicaciones precedentes, para lo que el mismo presenta en cada caso medios para realizar las distintas etapas de
procedimiento.

35 5. Producto de programa para una instalación de tratamiento de datos que contiene secciones de códigos con las
que puede realizarse un procedimiento según una de las reivindicaciones 1 a 3 sobre la instalación de tratamiento de
datos.

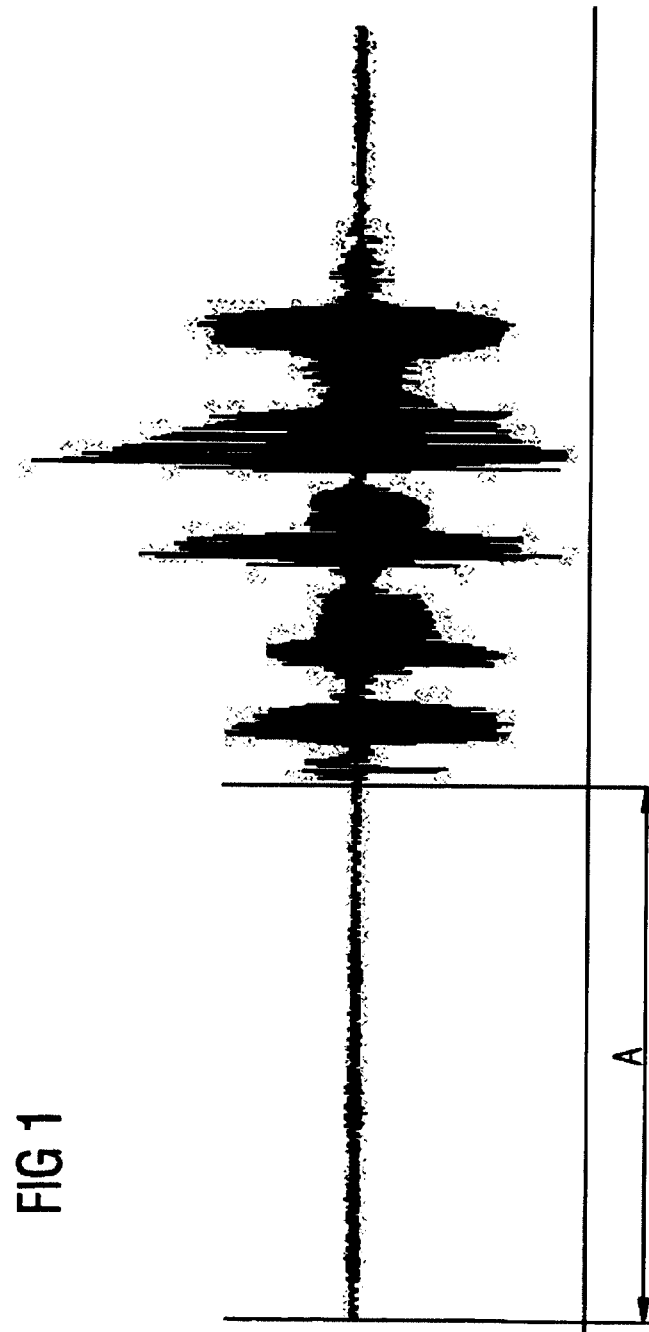


FIG 2

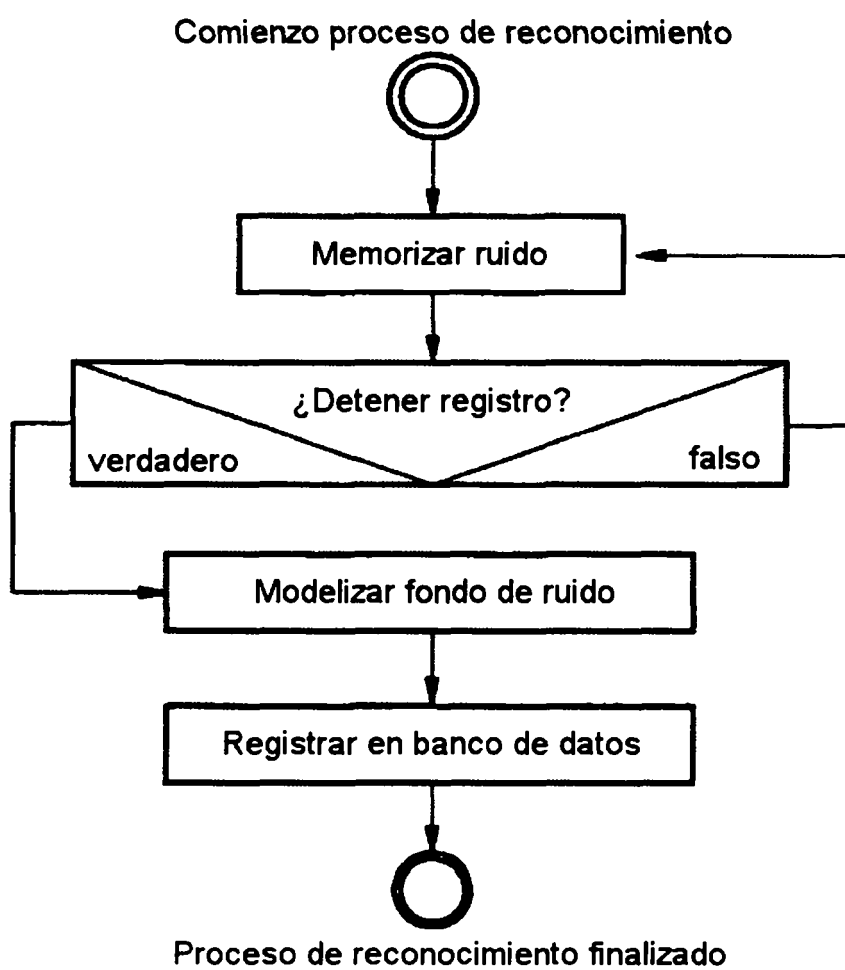


FIG 3 Comienzo proceso de reconocimiento

