

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 600 227**

51 Int. Cl.:

G10L 17/00 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.12.2008 PCT/EP2008/010478**

87 Fecha y número de publicación internacional: **17.06.2010 WO10066269**

96 Fecha de presentación y número de la solicitud europea: **10.12.2008 E 08875076 (5)**

97 Fecha y número de publicación de la concesión europea: **12.10.2016 EP 2364495**

54 Título: **Procedimiento para verificar la identidad de un orador y medio legible por ordenador y ordenador relacionados**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
07.02.2017

73 Titular/es:

**AGNITIO S.L. (100.0%)
Calle Gran Vía 39 - 8a Planta
28013 Madrid, ES**

72 Inventor/es:

**SANCHEZ ASENJO, MARTA;
GUTIERREZ, ALFREDO;
MARTIN DE LOS SANTOS DE LAS HERAS,
ALBERTO y
GARCIA GOMAR, MARTA**

74 Agente/Representante:

MILTENYI, Peter

ES 2 600 227 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Procedimiento para verificar la identidad de un orador y medio legible por ordenador y ordenador relacionados.

5 La presente solicitud se refiere a un procedimiento para verificar la identidad de un orador en base a la voz del orador.

10 La verificación de la identidad del orador se utiliza, por ejemplo, para acceder a sistemas de banca en línea o cualquier otro sistema en el que tenga que verificarse la identidad del orador. La verificación de la identidad del orador se refiere a la situación en la que alguien pretende tener una cierta identidad, y se requiere comprobar que la persona tiene, de hecho, esta identidad. La identificación del orador basada en la voz del orador tiene ventajas particulares, ya que, de la voz de un orador pueden extraerse datos de voz biométricos con un grado de precisión tal que es prácticamente imposible imitar la voz de otra persona por cualquier otro orador con un grado de precisión suficiente con el fin de realizar un fraude.

15 La técnica anterior EP 0 533 396 describe un sistema para la verificación del orador que verifica varias expresiones del mismo orador y comprueba si las expresiones son idénticas.

20 El objetivo de la presente invención es presentar un procedimiento y un aparato, que aumente además la seguridad de la verificación de la identidad de un orador.

Este objetivo se consigue con el procedimiento de la reivindicación 1, el medio legible por ordenador de la reivindicación 9 y el ordenador de la reivindicación 10.

25 De acuerdo con el procedimiento para verificar la identidad de un orador, primero se recibe una expresión de voz. Esta expresión de voz se analiza utilizando datos de voz biométricos para verificar que la voz del orador corresponde a la identidad del orador que se va a verificar. Además, se realiza una o más etapas en las que se verifica que la expresión de voz recibida no está falsificada. Puede pensarse que la expresión de voz puede falsificarse debido a que la expresión de voz de la identidad del orador que se requiere verificar se graba y después se reproduce. Esto puede realizarse con el fin de simular tener una cierta identidad, por ejemplo, para obtener acceso a un sistema, el cual está protegido mediante la verificación de identidad. En tal caso, la prueba de datos de voz biométricos confirmará positivamente la identidad ya que la voz coincide con la identidad pretendida. El acceso o cualquier otro derecho, sin embargo, deben negarse ya que no es la persona correcta la que intenta tener acceso a un sistema.

35 Antes de la recepción de la expresión de voz puede solicitarse una expresión de voz dentro del procedimiento. Se puede pedir a un orador, por ejemplo, que pronuncie una palabra, número o frase determinada que se le proporcione dentro de la ejecución del procedimiento (en la misma sesión) o se le indique una palabra o una frase de contraseña acordada con él previamente (es decir, antes de la ejecución del procedimiento).

40 Con el fin de comprobar la identidad de un orador, pueden llevarse a cabo pruebas muy elaboradas y detalladas que, sin embargo, provocan que las personas se sientan molestas por extensos y prolongados procedimientos de verificación cuando, por ejemplo, se intenta acceder a un sistema o a la concesión de cualquier otro derecho. Tales procedimientos de verificación de identidad molestos no son prácticos y, por lo tanto, tiene que encontrarse una manera que, por una parte, sea conveniente para los oradores cuya identidad se verifica y, por otra parte, evite el fraude de la verificación de identidad.

50 El procedimiento se refiere a la etapa de determinar si la expresión de voz está falsificada. En ese tipo de comprobaciones, no se determina que la voz está falsificada (por ejemplo, por un imitador de voz), sino si la expresión de voz está falsificada en base a una voz auténtica. Una expresión de voz falsificada, en general, puede ser cualquier expresión de voz que no se produzca en el momento de la verificación de la identidad de la persona a la que pertenece la voz, sino que puede ser, por ejemplo, una expresión que haya sido grabada previamente (por ejemplo, en secreto) y se repita después para la verificación de la identidad. Dicha grabación puede realizarse, por ejemplo, con un micrófono situado a una cierta distancia del orador (por ejemplo, en una posición alejada, tal como más de 10 ó 5 cm), o puede estar situado muy cerca del orador, por ejemplo, en un teléfono (típicamente menos de 10 ó 5 cm).

60 Además, una expresión de voz falsificada puede ser una expresión compuesta por una pluralidad de expresiones (cortas) que se compongan para una expresión más larga obteniéndose, de este modo, un contenido semántico que nunca fue grabado. Si, por ejemplo, durante la grabación de la voz de una persona se pronuncian diferentes números o dígitos en una cierta voz de orden, pueden componerse expresiones correspondientes a cada dígito en un orden diferente, de manera que pueda producirse cualquier combinación de números que pueda solicitarse por el sistema de verificación. Mientras que en aquellos casos la voz es correcta la expresión de voz está falsificada.

Otra posibilidad de falsificación de una expresión de voz puede ser en el caso de una voz generada sintéticamente. Puede prepararse o regularse un generador de voz para imitar un tipo de voz particular, de manera que dicha expresión de voz pueda falsificarse con dicho generador de voz.

5 Otra opción que puede considerarse como forma de falsificación de una expresión de voz puede ser el caso en que se roba una expresión de voz almacenada en un sistema informático. En un sistema informático puede almacenarse una expresión de voz almacenada recibida por ejemplo para la preparación o durante una sesión anterior, por ejemplo, que se utilice para verificar la identidad de un orador tal como se describe aquí. Si se roba dicha expresión de voz, ésta puede reproducirse, generando de este modo una expresión de voz falsificada.

10 Con el fin de hacer que el sistema sea lo más conveniente posible para los oradores, se prefiere que la verificación de que la expresión de voz no esté falsificada se lleve a cabo sólo después de que se haya verificado la voz del orador.

15 Sin embargo, pueden llevarse a cabo también ciertas pruebas en paralelo tales como, por ejemplo, una prueba pasiva para verificar que la expresión de voz no está falsificada, una vez que se recibe una expresión de voz para la verificación de la identidad del orador.

20 Finalmente, en el procedimiento se lleva a cabo una etapa que acepta la identidad de un orador que se verifica o bien no acepta la identidad del orador que se verifica. Si puede verificarse que la voz del orador corresponde al orador, cuya identidad se verifica, y que la expresión de voz no ha sido falsificada, entonces la identidad del orador puede aceptarse para ser verificada. En este caso, por ejemplo, puede concederse acceso a un sistema protegido y, si no, denegarse o realizarse etapas adicionales con el fin de determinar si de hecho la expresión de voz no está falsificada.

25 En una realización preferida, se procesa la expresión de voz recibida con el fin de determinar si está falsificada o no sin procesar ninguna otra expresión de voz. La verificación se basa, por lo tanto, en una expresión de voz que puede comprobarse por indicios de que la expresión de voz está falsificada. En otras etapas de la verificación de que la expresión de voz recibida no está falsificada, sin embargo, pueden procesarse otras expresiones de voz antes o después de esta sub-etapa en la que sólo se procesa la expresión de voz recibida.

30 La sub-etapa específica se refiere al procesamiento sin ninguna otra expresión de voz sólo hasta que se ha llegado a una conclusión anterior de si la expresión de voz recibida está falsificada o no. Esto, sin embargo, no tiene por qué ser la conclusión final al respecto.

35 Este tipo de comprobación puede formar parte de una prueba pasiva de falsificación, ya que no requiere ninguna entrada adicional de un orador durante la sesión de verificación de identidad.

40 En una realización preferida, cualquier prueba sobre si una expresión de voz está falsificada o no es inicialmente sólo una prueba pasiva, es decir, una que no requiere que un orador proporcione cualquier expresión de voz adicional. En caso de que en esta prueba pasiva no se encuentre ninguna indicación de falsificación, el orador es aceptado. Esto es especialmente útil para tener un procedimiento que sea conveniente para un gran número de oradores sin intención de fraude. Esto, sin embargo, se requiere que la prueba pasiva sea capaz de detectar muchos tipos de pistas de que la expresión de voz pueda estar falsificada. Por lo tanto, la prueba pasiva, en otra realización preferida, es capaz de detectar diferentes tipos de pistas de que una expresión de voz pueda estar falsificada.

45 De acuerdo con una realización particular, se lleva a cabo una prueba activa de falsificación que requiere una entrada adicional del orador, solamente en caso de que la prueba pasiva de falsificación haya dado una indicación de que la expresión de voz pueda estar falsificada.

A continuación, se explican algunas de las posibles comprobaciones de una prueba pasiva de falsificación.

50 En una comprobación que forma parte de una prueba pasiva, la grabación de la voz en una posición alejada puede detectarse determinando un índice de modulación de voz de la expresión de voz. De este modo, puede identificarse ruido adicional o ruido de convolución, lo que puede ser una pista para la grabación de la expresión de voz en una posición alejada (más de 5 ó 10 cm de distancia de la boca del orador). Además, puede tenerse en cuenta una relación de intensidad de la señal en dos bandas de frecuencia, una presentando un rango de frecuencias más bajas que la otra, para la detección de una grabación en una posición alejada. Se ha encontrado que tal relación proporciona un indicador útil de una grabación en una posición alejada puesto que las componentes de las frecuencias más bajas generalmente están más reforzadas en la posición alejada que en la posición cercana. En una realización preferida, puede utilizarse una combinación del índice de modulación de la expresión y de una relación de baja frecuencia/alta frecuencia para identificar falsificaciones.

En otra comprobación que forma parte de una prueba pasiva, puede evaluarse la prosodia con el fin de comprobar, por ejemplo, si la pronunciación de una palabra corresponde a su posición en una frase. Puede comprobarse, por ejemplo, si una palabra que se encuentra al principio o al final de una frase es pronunciada de esa manera. En el habla natural, la pronunciación de una misma palabra al principio, a la mitad y al final de una frase es ligeramente diferente. Estas pronunciaciones particulares pueden comprobarse mediante la evaluación de la prosodia. De este modo es posible identificar, por ejemplo, un generador de voz sintética, que normalmente no pueden proporcionar una prosodia natural y, por otra parte, puede ser posible detectar una expresión de voz editada en la que partes de expresión de voz más pequeñas se componen en una expresión de voz más grande.

Además, en una comprobación que forma parte de una prueba pasiva, puede investigarse una determinada marca de agua acústica en una expresión de voz. Las expresiones de voz que están almacenadas en un sistema informático pueden incorporar marcas de agua acústica. De esta manera, puede garantizarse que se puedan identificar expresiones de voz robadas, cuando se intenta identificar dichas marcas de agua acústicas. Una marca de agua acústica, puede ser, por ejemplo, una señal particular a una frecuencia específica o un rango de frecuencias (pequeño) que no afecte durante la repetición, pero que pueda identificarse, por ejemplo, mediante un análisis de Fourier que proporcione la señal particular en el rango de frecuencias o frecuencia específica.

Otra posible comprobación en una prueba pasiva es una comprobación de discontinuidades en el ruido de fondo. Aquí, por ejemplo, puede calcularse un perfil de ruido de fondo para diferentes intervalos de tiempo, tal como, por ejemplo, intervalos de tiempo de 1 a 5 o de 2 a 3 segundos y puede compararse el perfil de ruido de fondo de diferentes intervalos de tiempo. Si existen grandes diferencias, esto puede ser una indicación de, por ejemplo, una expresión de voz editada o una grabación en una posición alejada en un entorno con mucho ruido de fondo o variable.

El resultado de las diferentes comprobaciones de una prueba pasiva puede combinarse de diferentes maneras. Pueden combinarse, por ejemplo, de manera lógica con operaciones Y y/u O. Dado que las diferentes comprobaciones suelen identificar diferentes tipos de falsificación, preferiblemente se combinan de manera que, si cualquier comprobación indica que puede darse una falsificación, el orador no es aceptado directamente sin pruebas anteriores o no es aceptado en absoluto.

En otra realización preferida se solicita y se recibe una segunda expresión de voz. Esto corresponde a una prueba activa de falsificación. La solicitud puede realizarse por cualquier medio adecuado tal como, por ejemplo, una conexión telefónica mediante la cual se reciba la primera expresión de voz. La petición solicita preferiblemente que un orador repita la expresión de voz recibida justo antes. Después de recibir la segunda expresión de voz, se procesa la primera expresión de voz y la segunda expresión de voz con el fin de determinar una coincidencia exacta de las dos expresiones de voz. En caso de que, por ejemplo, una expresión de voz esté falsificada mediante la reproducción de una expresión de voz grabada, estas dos expresiones de voz coincidirán exactamente en ciertos aspectos. La coincidencia exacta de dos expresiones de voz puede determinarse en base a parámetros específicos de expresión de voz, tal como un GMM o cualquier otra característica de frecuencia que se extraiga de cada una de las expresiones de voz.

Se ha encontrado que, si una misma persona repite el mismo texto, son comunes variaciones menores. Esto puede deberse a pronunciaciones ligeramente diferentes o a un ruido de fondo distinto. Sin embargo, si la expresión de voz se reproduce a partir de una expresión de voz grabada, estos elementos no varían y, por lo tanto, tratar de determinar una coincidencia exacta es un medio útil para identificar que una expresión de voz se reproduce y, de hecho, es una expresión de voz previamente grabada.

Para la prueba de coincidencia exacta mencionada anteriormente esto es, por lo tanto, ventajoso ya que el contenido semántico de la segunda expresión de voz solicitada es idéntico al de la expresión de voz recibida. Sin embargo, el contenido semántico puede ser diferente y sólo ser idéntica una parte del contenido semántico y la coincidencia exacta determinarse únicamente por esa parte.

En la determinación de una coincidencia exacta, también es posible comparar una expresión de voz recibida con una expresión de voz que se ha recibido durante una fase de registro o de preparación con ese orador, es decir, antes de la recepción de la expresión de voz para la verificación de la identidad. Si cualquier otra persona grabó en secreto dicha expresión de voz para reproducirla posteriormente, esto será detectado. Igualmente, la determinación de una coincidencia exacta puede realizarse respecto a una expresión de voz recibida previamente en otra sesión de verificación de identidad, pero después del registro o la preparación, tal como por ejemplo una sesión en la que la identidad fue verificada hace unos días. Dicha prueba para una coincidencia exacta con una expresión de voz recibida en una sesión de verificación de la identidad anterior o con una expresión de voz recibida durante el registro o la preparación puede realizarse también como parte de la prueba pasiva de falsificación que se ha mencionado anteriormente y a continuación.

En cualquier prueba de una coincidencia exacta que se ha mencionado anteriormente y a continuación puede determinarse también que las dos expresiones de voz que se comparan tienen por lo menos algún grado de similitud con el fin de evitar un resultado de una prueba de una coincidencia exacta en la que dos expresiones de voz son completamente diferentes, por ejemplo, en contenido semántico. El grado de similitud puede determinarse a partir de características extraídas de dos expresiones de voz.

En un posible escenario de fraude, esto puede juzgarse para variar sintéticamente la segunda expresión de voz, de manera que no sea exactamente igual a la primera expresión de voz. Estas variaciones pueden realizarse, por ejemplo, con la adición de ruido blanco. Otra posibilidad es estirar o comprimir ciertas partes de la expresión de voz imitando de ese modo una prosodia diferente. Cuando se realiza una prueba de una coincidencia exacta pueden llevarse a cabo diferentes comprobaciones para identificar una coincidencia exacta. Una de esas comprobaciones puede ser capaz, por ejemplo, de ignorar cualquier ruido blanco añadido mientras que otra comprobación puede no verse afectada por el estiramiento o la compresión de la expresión de voz. Los resultados de las diferentes comprobaciones de coincidencia exacta preferiblemente se combinan de manera lógica, por ejemplo, mediante una operación O tal que cualquier comprobación que indique una coincidencia exacta lleve a la conclusión final de la prueba de una coincidencia exacta.

Además, se combina preferiblemente una prueba para una coincidencia exacta con una prueba adicional para la verificación del orador basada en la segunda expresión de voz. En caso de que la segunda expresión de voz sea alterada sintéticamente, la prueba para la verificación del orador puede fallar ya que las alteraciones son demasiado profundas. Por lo tanto, la combinación de una verificación del orador y de una prueba de coincidencia exacta se complementan entre sí de manera ventajosa para identificar expresiones falsificadas.

En otra realización preferida, se procesa la expresión de voz recibida y la segunda expresión de voz recibida con el fin de determinar una coincidencia exacta de las dos expresiones de voz o una parte de las mismas, respectivamente, y la segunda expresión de voz se procesa adicionalmente mediante una prueba pasiva tal como en una sub-etapa particular sin procesar cualquier otra expresión de voz o datos determinados de la misma, con el fin de verificar que la segunda expresión de voz no está falsificada. Estas dos etapas de procesamiento se llevan a cabo de manera independiente una de la otra y/o en paralelo entre sí. Esto aumenta la velocidad de procesamiento y, por lo tanto, la conveniencia y también la exactitud del procedimiento de verificación ya que los resultados de las dos pruebas pueden combinarse de manera lógica con el fin de determinar si las expresiones de voz están falsificadas o no. Dependiendo del resultado de las dos pruebas, pueden tomarse diferentes acciones tales como aceptación, rechazo o etapas de procesamiento adicionales.

En un procedimiento ventajoso particular, se intenta comprobar la vida del orador (que es un ejemplo de una prueba activa de falsificación). Esta prueba proporciona una determinación altamente fiable de si una expresión de voz recibida está falsificada o no, pero, por otra parte, provoca mucha incomodidad para un orador lo cual es molesto para los oradores y es indeseable para oradores no fraudulentos. En el presente procedimiento es preferible, por lo tanto, tener otras pruebas menos molestas previamente o no tener pruebas anteriores previamente (lo que proporcionaría resultados sólo menos fiables).

Puede comprobarse la vida del orador, por ejemplo, proporcionando un grupo de por lo menos 100, 500, 1000, 2000 o 5000 o más frases almacenadas que puedan enviarse de manera apropiada al orador. Éstas pueden enviarse, por ejemplo, mediante la reproducción de audio a través de una conexión telefónica, o mediante el envío de un mensaje electrónico por correo electrónico o SMS o similar. La frase es preferiblemente una frase que no se haya utilizado antes durante una nueva fase de registro o de preparación del orador, que puede haberse realizado antes de que se lleve a cabo el procedimiento para la verificación de la identidad con el fin de asegurar que dicha frase todavía no la ha dicho el orador y, por lo tanto, podría no haber sido grabada previamente.

La selección de la frase puede realizarse al azar. Además, puede comprobarse que, para una misma identidad que deba verificarse nunca se utiliza la misma frase dos veces. Después de que se ha seleccionado dicha frase, se pide al orador que diga la frase seleccionada y puede recibirse otra expresión de voz. Se prefiere utilizar una frase que comprenda una pluralidad de palabras tales como por lo menos 3, 4 ó 5 palabras con el fin de garantizar que dicha frase nunca ha sido pronunciada por el orador anterior.

Tras haberse recibido otra expresión de voz, primero se lleva a cabo una etapa de reconocimiento de voz con el fin de determinar el contenido semántico de la otra expresión de voz, para determinar que el contenido semántico de la expresión de voz recibida corresponde al de la frase seleccionada. Hay que señalar aquí que en la verificación de la voz del orador generalmente se suprime cualquier contenido semántico y se utilizan solamente características individuales de una voz que son comúnmente independientes de contacto de semántico, mientras que, cuando se determina el contenido semántico puede suprimirse cualquier característica particular de la voz con el fin de determinar solamente el contenido semántico independiente de la voz.

Además, los datos de voz biométricos se utilizan para verificar que la voz del orador corresponde a la identidad que ha de ser verificada en base a la otra expresión de voz.

5 Combinando estas dos pruebas, se determina en primer lugar que un orador vivo es capaz en este momento de pronunciar una frase en particular a petición, de manera que se minimiza la posibilidad de que la otra expresión de voz recibida haya sido registrada previamente y, en segundo lugar, se verifica la identidad del orador en base a la misma expresión de voz.

10 En otras realizaciones preferidas, es posible que las distintas etapas se dispongan de manera que el procedimiento realice uno, dos, tres o más bucles, en el que, en cada bucle se solicite, se reciba y se procese otra expresión de voz. El procesamiento de dicha otra expresión de voz recibida preferiblemente tiene uno, dos, tres o todos de un grupo de sub-etapas que comprende: utilizar datos de voz biométricos para verificar que la voz del orador corresponde a la identidad del orador, cuya identidad se verifica en base a otra expresión de voz recibida; determinar una coincidencia exacta de la otra expresión de voz recibida con cualquier expresión de voz recibida previamente durante la realización del procedimiento, es decir, en una sesión (todas las expresiones de voz recibidas anteriormente, la expresión de voz previa recibida finalmente, las dos últimas expresiones de voz recibidas anteriormente, etc.), determinar la falsificación de la otra expresión de voz recibida sin procesar cualquier otra expresión de voz para esta sub-etapa particular y comprobar la vida del orador.

20 Cualquiera de los procedimientos descritos anteriormente o a continuación proporciona un resultado que es indicativo de que el orador ha de ser aceptado o rechazado. Este resultado puede utilizarse para conceder o denegar el acceso a un sistema protegido, tal como, por ejemplo, un sistema de banca telefónica o un sistema de acceso a banca en línea basado en Internet que pueda gestionar adicionalmente transmisiones de voz.

25 Otras aplicaciones del procedimiento son posibles, así como, por ejemplo, en un procedimiento para informar a una persona de un evento y un procedimiento para recibir información acerca de un evento tal como el que se describe en la solicitud internacional con número de solicitud PCT/EP2008/002778.

30 Además, el procedimiento puede utilizarse en un procedimiento para generar un medio y/o estado temporalmente limitado y/o de uso limitado, un procedimiento para obtener un medio y/o estado temporalmente limitado y/o de uso limitado tal como se describe en la solicitud internacional con número de solicitud PCT/EP2008/002777.

35 También, el procedimiento puede utilizarse en un procedimiento para la localización de una persona, un sistema para la localización de una persona tal como se describe en la solicitud internacional con número de solicitud PCT/EP2008/003768.

40 El procedimiento se lleva a cabo preferiblemente mediante un ordenador o es implementado en el mismo. Este ordenador puede formar parte de un sistema informático. El ordenador o sistema informático puede formar parte de un sistema de servicio telefónico que proporcione algunos servicios tales como un servicio de banca telefónica, para lo cual el acceso está restringido y se requiere que la restricción tenga que superarse por identificación.

45 El procedimiento puede realizarse en una llamada telefónica entrante recibida por un orador o cualquier otro tipo de comunicación capaz de transmitir datos de audio. Dicha llamada telefónica o comunicación inicia una sesión para la verificación de la identidad de un orador.

La presente invención se refiere también a un medio legible por ordenador que tiene instrucciones en el mismo las cuales, cuando se ejecutan en un ordenador, realizan cualquiera de los procedimientos descritos anteriormente o a continuación.

50 Igualmente, la invención se refiere a un sistema informático que tiene el citado medio legible por ordenador.

55 Las expresiones del orador pueden haberse proporcionado antes de la realización del procedimiento para la verificación de la identidad del orador (en una fase de preparación o registro) con el fin de evaluar tales expresiones de voz, de manera que pueden extraerse datos de voz biométricos de los mismos. Esos datos de voz biométricos pueden utilizarse para la verificación de que la voz del orador corresponde con el orador cuya identidad se verifica.

60 Los datos de voz biométricos pueden extraerse de una expresión de voz mediante un análisis de frecuencia de la voz. A partir de secuencias de expresión de voz de, por ejemplo, 20 ó 30 milisegundos, puede realizarse la transformada de Fourier y, a partir de su envolvente, pueden extraerse datos de voz biométricos. A partir de múltiples de tales secuencias de voz de la transformada de Fourier puede generarse un modelo de voz denominado Modelo de Mezcla Gaussiana, (GMM). Sin embargo, puede utilizarse cualquier otro dato de voz que permita distinguir una voz de otra voz debido a características de voz. Además, pueden considerarse características de voz que tengan en cuenta que la expresión de voz se refiere al contenido semántico específico. Por ejemplo, pueden

utilizarse modelos ocultos de Markov (HMM) que tengan en cuenta las probabilidades de transmisión entre diferentes Modelos de Mezcla Gaussiana, cada uno de los cuales se refiere a un sonido o una letra dentro de una palabra.

5 En las figuras se describen algunas realizaciones preferidas de la presente invención. Esas figuras solamente muestran algunos ejemplos, y no son limitativos de la invención. En las figuras adjuntas, se muestra:

Figuras 1 a 6: diagramas de flujo específicos de un procedimiento para realizar un procedimiento para verificar la identidad de un orador.

10

Cada figura muestra ciertas etapas de una sesión en la que se verifica la identidad de un orador.

15 En la figura 1a, en el punto 10, se lleva a cabo la verificación del orador. En esta etapa, se acaba de recibir una expresión de voz en la misma sesión y se utilizan datos de voz biométricos (tal como un GMM o un HMM) para verificar que la voz de este orador corresponde al orador, cuya identidad se verifica. La verificación del orador puede basarse en datos (por ejemplo, un modelo de voz) que se almacenan en una base de datos, y que se extraen de expresiones de voz de los oradores durante una fase de registro o de preparación.

20 Durante la verificación del orador se verifica un orador particular, lo que significa que se adopta una identidad y esta identidad debe ser verificada. Con la información de identidad a mano, la cual puede basarse, por ejemplo, en un nombre del orador, un número de teléfono de una llamada telefónica entrante o similar, los datos de voz biométricos particulares se recuperan de una base de datos y se utilizan en el procesamiento de una expresión de voz recibida con el fin de verificar que la voz del orador corresponde al orador cuya identidad se verifica.

25 El resultado de la verificación del orador da un resultado lógico, que es positivo o negativo (sí/no) e indica si se ha verificado o no la identidad. Esto se muestra en la etapa 11 de la figura 1a. Si la identidad no se ha verificado, el orador es rechazado en el punto 14. Si la identidad puede verificarse, debe tenerse en cuenta que la expresión de voz recibida puede estar falsificada, por ejemplo, grabada previamente. Por lo tanto, en el punto 12 se lleva a cabo una prueba pasiva de falsificación. Una prueba pasiva es una que no necesita que el orador proporcione activamente ninguna otra expresión de voz en ese momento, sino que se basa solamente en la expresión de voz recibida en esta etapa de verificación del orador 10. Esta prueba pasiva de falsificación, en particular, es ventajosa puesto que no se requiere ninguna entrada adicional del orador, lo que permite una manera de determinar si la expresión de voz recibida puede estar falsificada o no, sin oradores molestos que, sin embargo, no tienen la intención de cometer un fraude. Dado que, sin embargo, un orador es aceptado directamente en caso de que la prueba pasiva 12 no indique ninguna sospecha de falsificación, esta prueba pasiva preferiblemente es capaz de comprobar múltiples tipos de falsificación. Por lo tanto, esta prueba puede llevar a cabo una comprobación para la determinación de una grabación en una posición alejada, anomalías en la prosodia, presencia de una marca de agua, discontinuidades en el fondo, tal como se ha explicado anteriormente, u otro tipo de comprobación. Si cualquier comprobación indica una falsificación, en la etapa 13 se concluirá que la expresión de voz está falsificada.

40

Si no pueden encontrarse indicaciones de que la expresión de voz fue falsificada, el orador es aceptado (véase el punto 16). Si se ha encontrado que la expresión de voz fue falsificada, entonces el orador puede ser rechazado o pueden adoptarse más medidas (véase punto 15). El tipo particular de acción (rechazo u otras etapas) puede llevarse a cabo en función del tipo de comprobación pasiva que indica que una expresión de voz fue falsificada. Diferentes comprobaciones pueden funcionar con una fiabilidad diferente en relación con la detección de expresiones de voz falsificadas. Si una comprobación (muy) fiable indica que se trata de una falsificación, el usuario puede ser rechazado directamente. Si una comprobación menos fiable indica que se trata de una falsificación, pueden adoptarse nuevas medidas (tal como se ha explicado anteriormente o a continuación, tal como una prueba activa de falsificación) con el fin de confirmar la anulación de la detección de una expresión de voz falsificada.

50

En la figura 1b se muestra un enfoque alternativo en el cual se realiza una verificación del orador y una prueba pasiva de falsificación (etapas 18 y 19) independientemente una de la otra y/o en paralelo. Ambas etapas se basan en una expresión de voz recibida en la etapa 17, que significa una misma expresión de voz. La verificación del orador en el punto 18, y la prueba pasiva de falsificación en el punto 19, cada una de las cuales permite una decisión de si el orador será aceptado o no, se combinan de manera lógica. Si ambas pruebas resultan positivas, el orador es aceptado (véase el punto 22). Si la etapa de verificación 20 es negativa, el orador es rechazado independientemente del resultado del punto 21 (véase el punto 24). Si en el punto 20 se obtiene un resultado positivo y en el punto 21 uno negativo, el orador puede ser rechazado en el punto 23, o pueden adoptarse otras medidas con el fin de determinar si el orador debe o no debe ser aceptado o rechazado. La acción particular tomada en la etapa 23 puede depender del tipo de comprobación particular que indica la falsificación en la etapa 19, 21 tal como se ha explicado anteriormente para la etapa 15.

60

Aunque en las figuras 4, 5 y 6 se muestra el mismo esquema inicial que el de las etapas 10 a 13 de la figura 1 esas etapas pueden ser sustituidas por las etapas de la figura 1b.

5 La figura 2 muestra una realización particularmente ventajosa, en la que, después de la verificación del orador en el punto 30, se decide si se verifica o no la identidad en el punto 31. Si la identidad no se verifica, el orador es rechazado (punto 32). Si la identidad se verifica, entonces, antes de aceptar al orador, se solicita al orador que proporcione otra expresión de voz en la etapa 33, la cual se recibe en el punto 34. Esta expresión de voz se procesa de nuevo para la verificación del orador en el punto 35 y, si en esta etapa no puede verificarse la identidad del orador, entonces el orador es rechazado en el punto 37. Si el resultado de la prueba en el punto 36 es positiva, entonces se procede a la etapa 38 donde se comprueba si las dos expresiones de voz recibidas en el punto 30 y 35 tienen una coincidencia exacta. Si éste es el caso, entonces en el punto 39 se determina que una o ambas expresiones de voz están falsificadas y, por lo tanto, el orador es rechazado en el punto 40. De lo contrario, es aceptado en el punto 41.

15 Dicho procedimiento es más complicado para un orador, puesto que tiene que proporcionar por lo menos dos expresiones de voz. Sin embargo, proporciona un buen grado de certidumbre para la cuestión de si la expresión de voz está falsificada o no. Este buen grado de certidumbre proviene, en particular, de la combinación de la etapa de verificación del orador de la segunda expresión de voz con la determinación de una coincidencia exacta, ya que un intento de pasar por la prueba de coincidencia exacta variando la segunda expresión de voz puede conducir a un rechazo al no pasar la prueba de verificación del orador 35.

20 La figura 3 muestra otro ejemplo particular, en el que, después de la verificación del orador en los puntos 50 y 51, que puede dar lugar al punto de rechazo 52, en el punto 53 se realiza una detección de vida. Aquí, la detección de vida se lleva a cabo directamente después de la etapa de verificación del orador, de manera que no deben adoptarse etapas previas. La detección de vida puede considerarse especialmente molesta para los oradores, ya que se requiere proporcionar otra entrada del orador de manera que pueda detectarse algún tipo de inteligencia por parte del orador. Si en el punto 54 se determina que el orador está vivo, éste es aceptado en el punto 56 y, de lo contrario, es rechazado en el punto 55.

30 En la figura 4 se muestra un ejemplo en el que se realizan pruebas activas de falsificación después de una prueba pasiva de falsificación. Esto corresponde al caso en que, en la figura 1, en un punto 15, se adoptan otras etapas. En la figura 4 se verifica un orador en los puntos 60 y 61, y es rechazado en el punto 62 en caso de que la identidad no pueda ser verificada. Si se verifica la identidad, entonces se lleva a cabo la prueba pasiva de falsificación en el punto 63. El resultado de la misma se comprueba en el punto 64. Si se determina que la expresión de voz no estaba falsificada, entonces el procedimiento procedería al punto 73 (véase A rodeada). Si se comprueba que la expresión de voz puede estar falsificada, entonces el orador no es rechazado directamente, sino que se adoptan medidas adicionales. En el ejemplo particular, se solicita otra expresión del orador en el punto 65 y se recibe en el punto 66. Esta expresión de voz recibida adicionalmente se comprueba a través de la etapa de verificación del orador en 67. Si la identidad no puede verificarse adicionalmente a partir de esta expresión de voz, el orador es rechazado en el punto 69 y, de lo contrario, se procede a determinar una coincidencia exacta en el punto 70. Si se encuentra una coincidencia exacta (véase el punto 71), entonces el orador es rechazado en el punto 72 y, de lo contrario, se procede a la aceptación 73. En la figura 4 se muestra una alternativa para la etapa de aceptación 73, que indica que antes de aceptar un orador puede llevarse a cabo una detección de vida 74. En la etapa 75 se decide si el orador se considera que está vivo o no y después, si esta prueba resulta positiva, el orador es aceptado en la etapa 77 y de lo contrario es rechazado en la etapa 76.

45 Puede verificarse el contenido semántico de la expresión de voz recibida en el punto 66. Esto significa que se comprueba que el contenido semántico de la expresión recibida en el punto 66 se ajusta al contenido semántico que se requiere en el apartado 65. Este examen puede realizarse en el punto 66, 67 o 70. Si el contenido semántico no corresponde, el orador puede ser rechazado o el procedimiento vuelve de nuevo a la etapa 65 solicitando de nuevo una expresión de voz.

50 La figura 5 muestra otro ejemplo especialmente ventajoso en términos de conveniencia para los oradores y de seguridad respecto a la verificación de la identidad.

55 En la etapa 80 se verifica un orador en base a una expresión de voz recibida en esta etapa. Si en la etapa 81 la identidad del orador no se verifica, el orador es rechazado en el punto 82. En caso de que se verifique la identidad, primero se realiza una prueba pasiva de falsificación 83. Dado que esta prueba pasiva no necesita ninguna entrada adicional del orador, ésta no afecta a la conveniencia del sistema para un orador que no pretende cometer un fraude. Si en la etapa 84 se determina que la expresión de voz no está falsificada, el orador se considera aceptado 85. En tal caso, un orador no nota ningún cambio del sistema respecto a la etapa de introducción de verificación si la expresión de voz recibida está falsificada. En caso de que en la etapa 84 se determine que la expresión de voz está falsificada o puede estar falsificada, el procedimiento pasa a la etapa 86, donde se solicita otra expresión al orador la

cual se recibe en la etapa 87. En la etapa 88, esta expresión de voz recibida adicionalmente es procesada para la verificación del orador. Si la identidad del orador, que se va a verificar, no puede ser identificada en la etapa 89 el orador es rechazado en la etapa 90.

5 Si la identidad puede verificarse positivamente, entonces el procedimiento sigue en las etapas 91 y 92. Ambas etapas pueden llevarse a cabo en paralelo. Sin embargo, también pueden llevarse a cabo una después de la otra. Es preferible, sin embargo, llevar a cabo las dos etapas de manera independiente una de la otra, y/o en paralelo dado que, de este modo, los resultados de las dos pruebas 91 y 92 pueden ser evaluados en combinación. Esto se muestra en la figura 5, donde en las etapas 93 y 94 se obtienen dos resultados posibles, siendo uno positivo y
10 siendo uno negativo sobre la cuestión de si cualquier expresión de voz, en particular la segunda expresión de voz, está falsificada o no. Si ambas pruebas determinan que la expresión de voz no está falsificada, entonces se procede a la aceptación en el punto 95. En este caso, debe suponerse que la prueba en la etapa 84 era errónea.

15 Realizando la prueba pasiva para falsificación también sobre la segunda expresión de voz en la etapa 91, se garantiza que se identifica y se tiene en cuenta cualquier indicio de falsificación presente sólo en la segunda expresión de voz, que puede ser diferente del tipo de pista determinada en la primera emisión de voz.

20 Si ambas pruebas 93 y 94 proporcionan un resultado negativo, entonces se procede al rechazo en el punto 96. En caso de que la prueba en la etapa 93 y 94 den resultados contradictorios, entonces puede realizarse después la prueba más profunda tras la B en el círculo. Aquí se realiza, además, una detección de vida en la etapa 97, que conduce entonces al rechazo o aceptación final 99, 100 en base al resultado en el punto 98.

25 Esta realización resulta conveniente para una gran cantidad de oradores que no tienen ninguna intención de fraude y que se consideran para aceptación 85. Para aquellos oradores que, sin embargo, han sido clasificados erróneamente como que utilizan expresiones de voz falsificadas en la etapa 84, se lleva a cabo el grupo de pruebas 91 y 92 con el fin de poder revertir el fallo de la etapa 84. Sin embargo, si no puede tomarse ninguna decisión clara (aceptación 95 o rechazo 96) entonces puede llevarse a cabo una prueba más avanzada para la detección de vida con el fin de lograr la decisión final. En la realización de la figura 5, pueden llevarse a cabo en cascada tres pruebas o grupos de pruebas diferentes (punto 84, punto combinado 93, 94 y punto 98) con el fin de obtener un número
30 mínimo de falsos rechazos y un alto grado de seguridad para determinar fraude, mientras que, al mismo tiempo, se ofrecen un enfoque práctico para la mayoría de los oradores.

35 En la realización de la figura 5, el contenido semántico de la expresión de voz recibida en el punto 87 puede comprobarse para ver si encaja o no con el contenido semántico de la expresión de voz requerida en el apartado 86. Si el contenido semántico no encaja, el procedimiento puede rechazar el orador o volver a la opción 86, de manera que se reciba otra expresión de voz.

40 La figura 6 muestra otro ejemplo particular preferido, que incluye un bucle en las etapas del procedimiento. Se llevan a cabo unas etapas 110 a 119 de manera similar a las etapas 80 a 89. Después, sin embargo, en la etapa 121 se realiza una determinación de una coincidencia exacta en el punto 120 y su evaluación con la posibilidad de rechazo en el punto 122. Después, se lleva a cabo una prueba pasiva de falsificación en el punto 123 y se evalúa en el punto 124 con la posibilidad de aceptación en 125. La combinación de las etapas 120 y 121 con la combinación de 123 y 124 también puede llevarse a cabo en orden inverso con etapas 123 y 124 realizadas previamente. Sin embargo, se prefiere realizar previamente la determinación de la coincidencia exacta en el punto 120, de manera que en
45 cualquier caso puede realizarse un rechazo en el punto 122 en caso de que se determine una coincidencia exacta.

Si la prueba 123 da un resultado positivo respecto a la cuestión de falsificación, entonces el procedimiento vuelve a la etapa 116, en la cual se solicita otra expresión.

50 De esta manera, se recibe una nueva expresión de voz, que puede comprobarse tal como se ha explicado anteriormente. En caso de que, por ejemplo, se utilicen de manera fraudulenta dos grabaciones de expresiones de voz diferentes, entonces la primera determinación en el punto 120 puede no indicar falsificación en la etapa 121. Sin embargo, si se recibe entonces una tercera expresión de voz en el segundo paso del bucle, entonces la tercera expresión de voz será una coincidencia exacta con la primera o la segunda expresión de voz recibida, lo que puede
55 determinarse entonces en la etapa 120. Por lo tanto, en la etapa 120 la determinación de una coincidencia exacta puede realizarse respecto a la expresión de voz recibida al final en la etapa 116, con cualquier otra expresión de voz previamente recibida (en la misma sesión), o las dos últimas, o tres últimas, o las cuatro últimas expresiones de voz recibidas. De esta manera, en caso de que esté presente más de una expresión de voz grabada, la misma puede utilizarse para determinar una coincidencia exacta en 120 y para identificar la falsificación en la etapa 121.

60 Tal como puede apreciarse en la figura 6, la identificación de una coincidencia exacta conduce al rechazo. La prueba pasiva de falsificación en la etapa 123 no conduce directamente a un rechazo, ya que se ha encontrado que la

citada prueba es menos fiable. Por lo tanto, con el fin de evitar un falso rechazo se proporciona un bucle, lo que aumenta la comodidad para los oradores, dándoles otra oportunidad.

REIVINDICACIONES

1. Procedimiento para verificar la identidad de un orador en base a la voz de oradores, que comprende las etapas de:

- 5
- a) recibir una expresión de voz de una palabra o una frase de contraseña;
 - b) utilizar datos de voz biométricos para verificar (10) que la voz de los oradores corresponde con el orador cuya identidad se va a verificar en base a la expresión de voz recibida;
 - c) rechazar al orador si la voz de los oradores no se verifica;

10 y, sólo si se verifica la voz de los oradores:

- 15
- d) verificar (12, 13) que la expresión de voz recibida no está falsificada, lo que comprende determinar si la expresión de voz puede estar falsificada y aceptar al orador, si se determina que la expresión de voz no está falsificada y solicitar una segunda expresión de voz, sólo si se determina que la expresión de voz puede estar falsificada, verificar que la segunda expresión de voz corresponde al orador cuya identidad se verifica y rechazar al orador, si no se verifica que la segunda expresión de voz corresponde al orador cuya identidad se verifica y, sólo si se verifica que la segunda expresión de voz corresponde al orador cuya identidad se verifica, después de haber verificado la identidad del orador, determinar si existe una coincidencia exacta de las dos expresiones de voz;

y sólo si no se determina que existe una coincidencia exacta de las dos expresiones de voz:

- 25
- e) aceptar (16) la identidad de los oradores que se verifica en caso de que se determine que no existe una coincidencia exacta de las dos expresiones de voz y, de otro modo, no aceptar (15) la identidad de los oradores que se verifica.

2. Procedimiento de acuerdo con la reivindicación 1, en el que dicha determinación de si la expresión de voz puede estar falsificada comprende una, tres, cuatro o todas las siguientes comprobaciones:

- 30
- procesar (12) la expresión de voz recibida sin procesar cualquier otra expresión de voz o datos determinados de la misma con el fin de determinar si la expresión de voz recibida está falsificada o no;
 - determinar un índice de modulación de voz y/o una relación entre la intensidad de la señal en dos bandas de frecuencia diferentes de la expresión de voz recibida para determinar una grabación de una voz en una posición alejada;
 - evaluar la prosodia de la expresión de voz recibida;
 - comprobar marcas de agua en la expresión de voz recibida y
 - detectar discontinuidades en el ruido de fondo.

40 3. Procedimiento de acuerdo con la reivindicación 1, en el que el contenido semántico de la segunda expresión de voz solicitada o una parte de la misma es idéntico al de la expresión de voz recibida o una parte de la misma.

45 4. Procedimiento de acuerdo con la reivindicación 1 ó 3, en el que la expresión de voz recibida y la segunda expresión de voz recibida se procesan (92) con el fin de determinar una coincidencia exacta y la segunda expresión de voz se procesa (91) a través de una prueba pasiva de falsificación y/o sin procesamiento de cualquier otra expresión de voz o datos determinados de la misma con el fin de verificar que la segunda expresión de voz recibida no está falsificada, en el que las dos etapas de procesamiento (91, 92) se llevan a cabo de manera independiente una de la otra y/o en paralelo una respecto a la otra y los resultados de las etapas de procesamiento se combinan de manera lógica con el fin de determinar si cualquier expresión de voz está falsificada o no.

50 5. Procedimiento de acuerdo con la reivindicación 4, en el que la combinación lógica de los resultados de las etapas de procesamiento se utiliza para decidir si se lleva a cabo o no una prueba de vida (98) del orador y en el que se lleva a cabo una prueba de vida del orador sólo en caso de que las dos etapas de procesamiento (91, 92) den resultados contradictorios en relación con la cuestión de si por lo menos la segunda expresión de voz está falsificada o no, en el que la vida se determina:

- 55
- seleccionando una frase con un sistema que tiene un grupo de por lo menos 100, 500, 1000, 2000 o 5000 frases almacenadas;
 - solicitar al orador que diga la frase seleccionada;
 - 60 recibir otra expresión de voz;
 - utilizar medios de reconocimiento de voz para determinar que el contenido semántico de la otra expresión de voz corresponde con el de la frase seleccionada; y

utilizar datos de voz biométricos para verificar que la voz de los oradores corresponde con el orador cuya identidad se verifica en base a la otra expresión de voz.

5 6. Procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 4, en el que la etapa c) comprende la sub-etapa de:

determinar la vida (54, 75, 98) del orador, en el que la vida se determina:

10 seleccionando una frase con un sistema que tiene un grupo de por lo menos 100, 500, 1000, 2000 o 5000 frases almacenadas;

solicitar al orador que diga la frase seleccionada;

recibir otra expresión de voz;

utilizar medios de reconocimiento de voz para determinar que el contenido semántico de la otra expresión de voz corresponde con el de la frase seleccionada; y

15 utilizar datos de voz biométricos para verificar que la voz de los oradores corresponde con el orador cuya identidad se verifica en base a la otra expresión de voz.

20 7. Procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 6, en el que el procedimiento realiza uno o más bucles (116 a 124), en el que, en cada bucle, otra expresión de voz se solicita (116), se recibe (117) y se procesa, en el que el procesamiento de la otra expresión de voz recibida comprende además preferiblemente una, dos, tres o todas las siguientes sub-etapas:

utilizar datos de voz biométricos para verificar (118, 119) que la voz de los oradores corresponde con la identidad del orador cuya identidad se verifica en base a la otra expresión de voz recibida;

25 determinar una coincidencia exacta (120) de la otra expresión de voz recibida con una expresión de voz recibida previamente;

determinar (123) una falsificación de la otra expresión de voz recibida en base a la otra expresión de voz recibida sin procesar cualquier otra expresión de voz; y

30 determinar (53) la vida de un orador.

8. Procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 7, en el que el procedimiento proporciona un resultado que es indicativo de que el orador es aceptado o rechazado.

35 9. Medio legible por ordenador, que tiene instrucciones almacenadas en el mismo las cuales, cuando se ejecutan en un ordenador, realizan cualquiera de los procedimientos de las reivindicaciones 1 a 8.

10. Ordenador que tiene un medio legible por ordenador como el de la reivindicación 9.

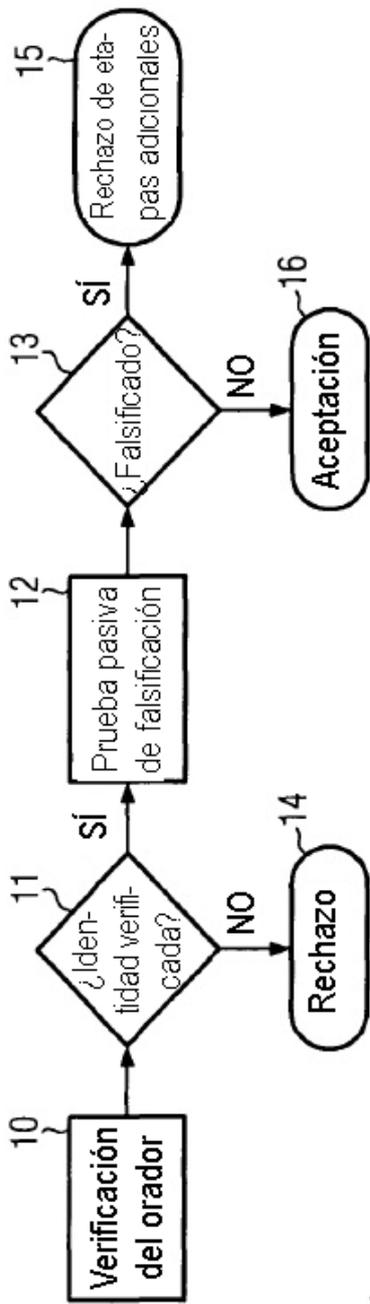


FIG. 1a

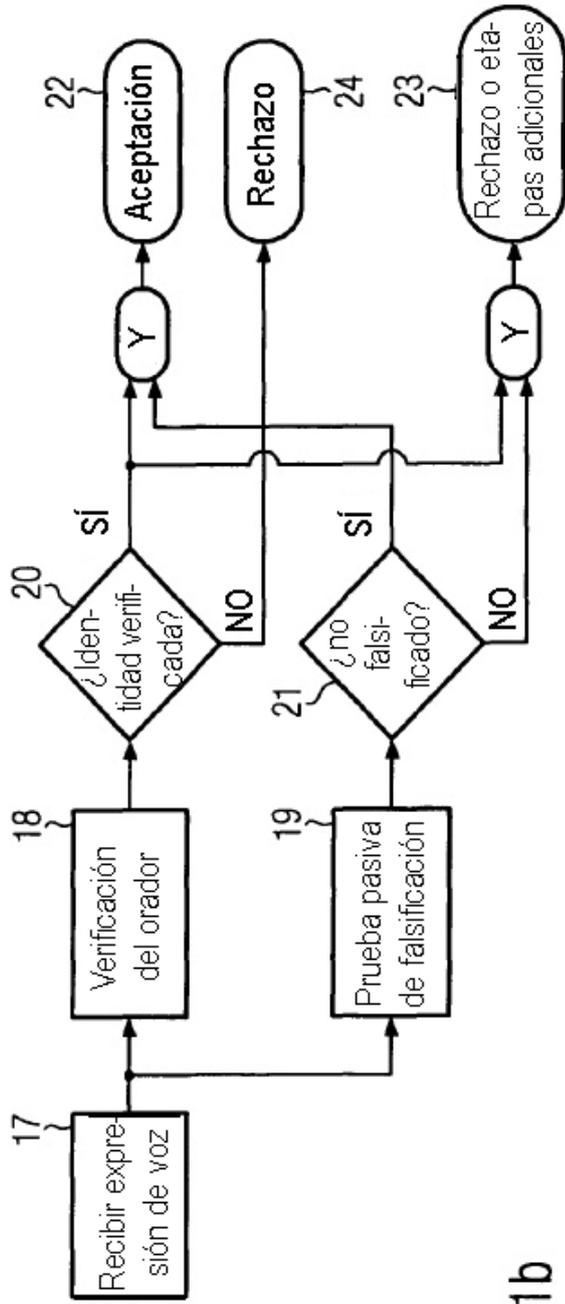


FIG. 1b

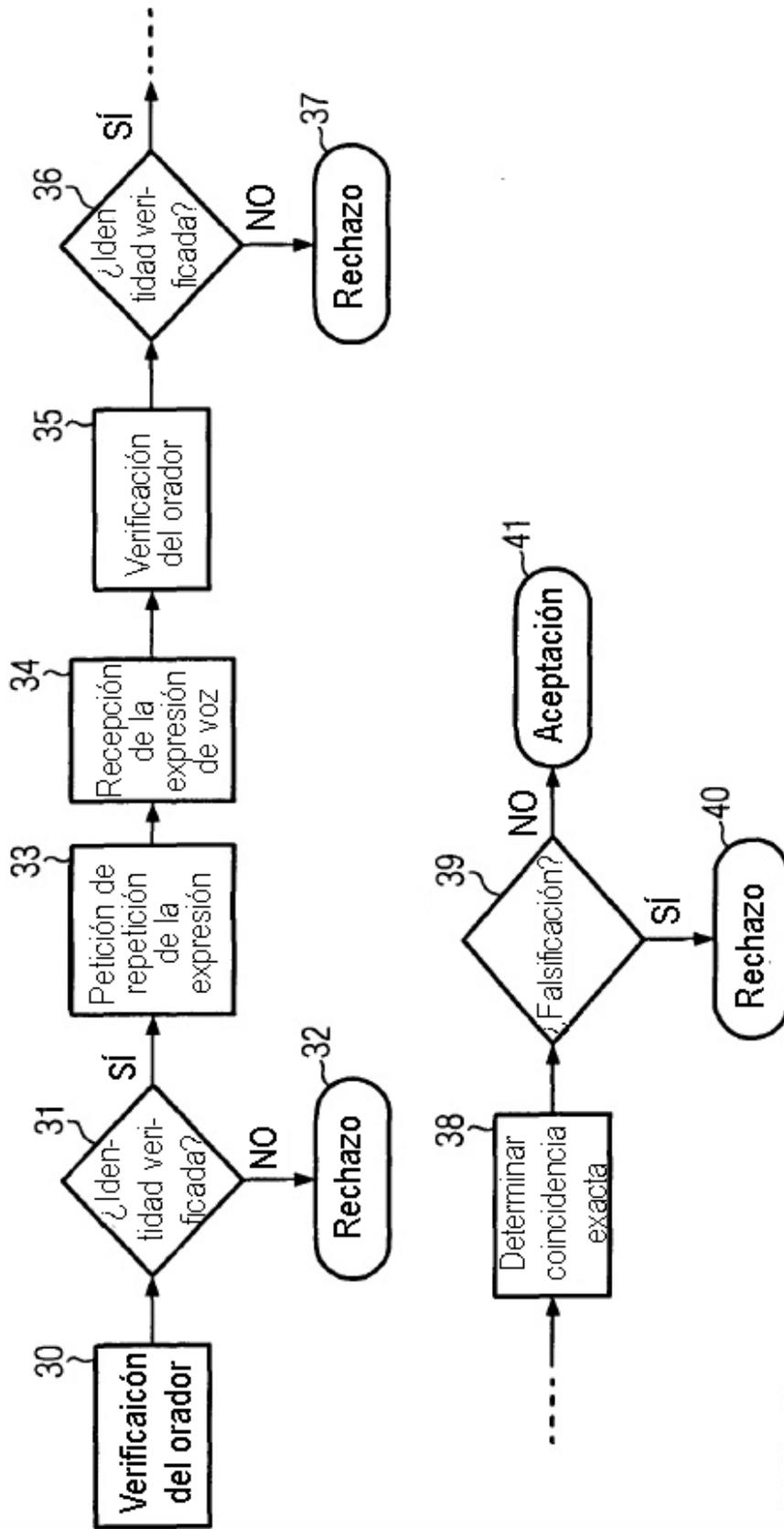


FIG. 2

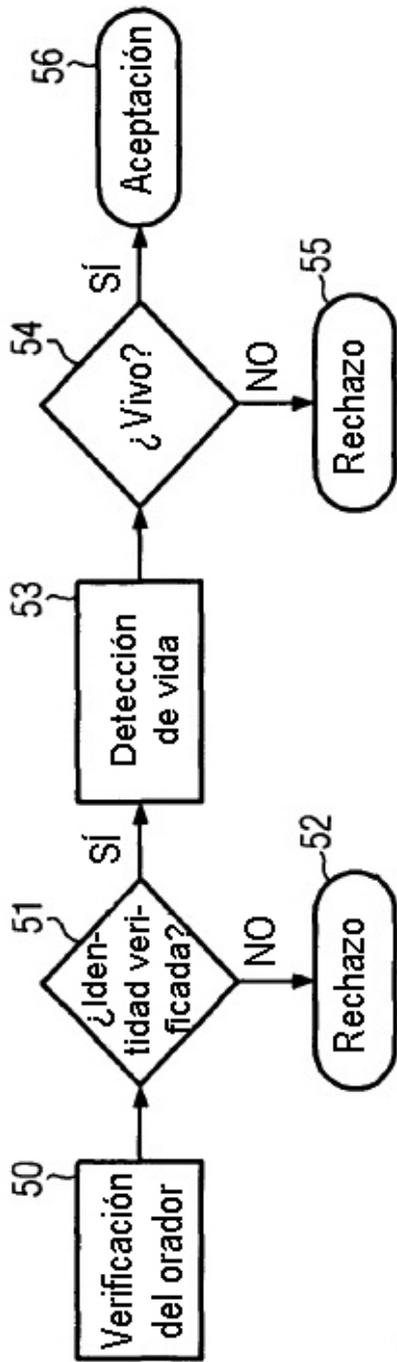


FIG. 3

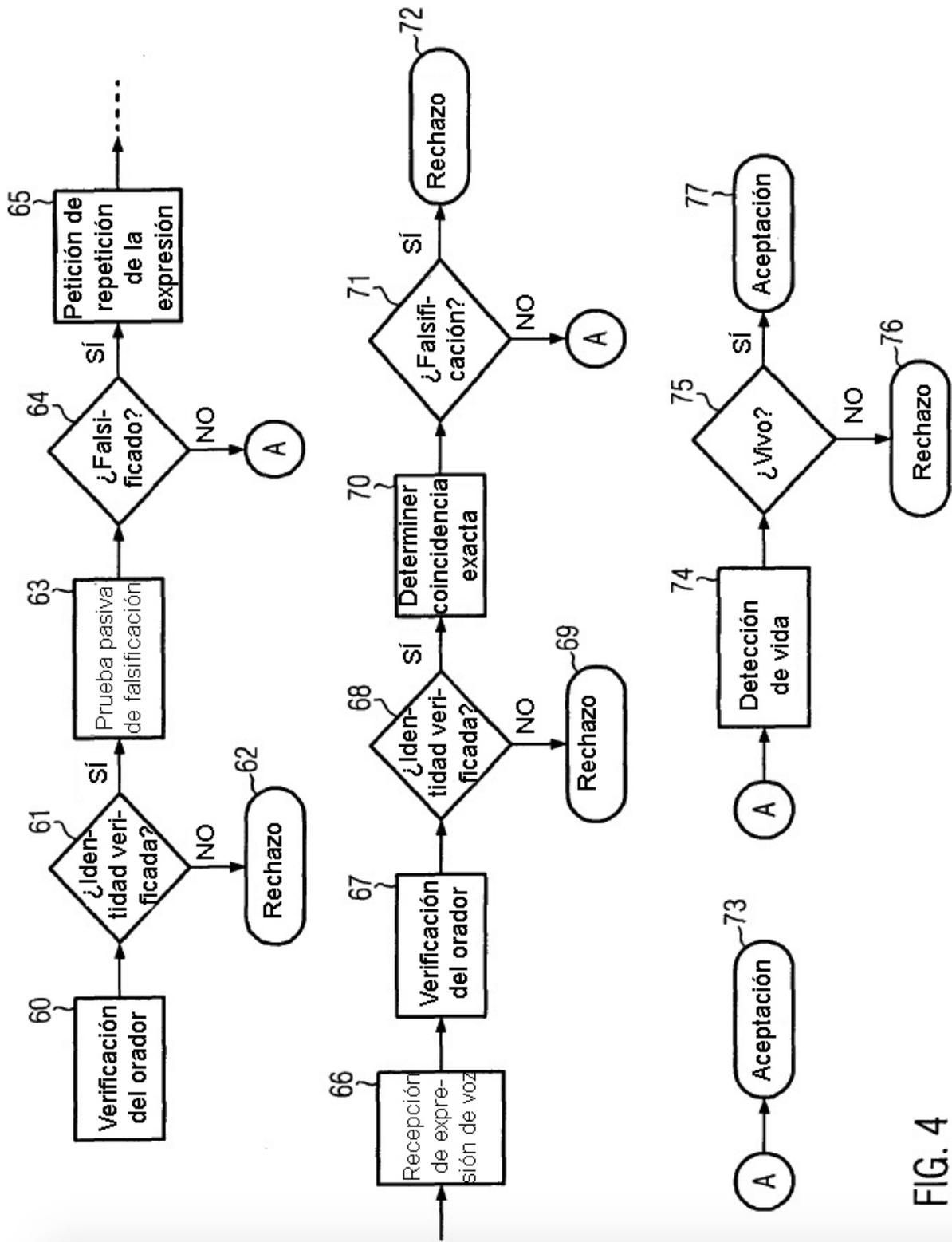


FIG. 4

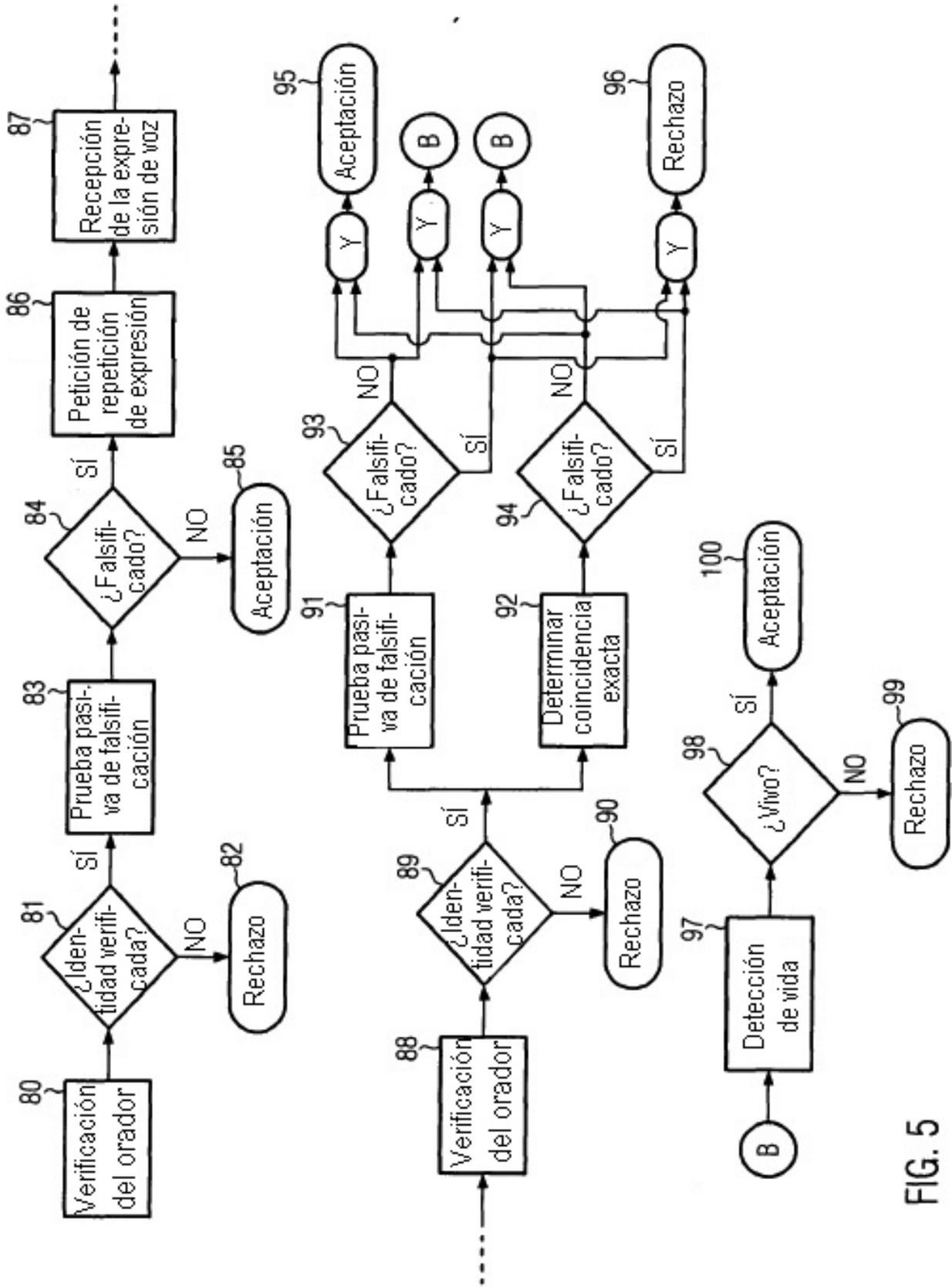


FIG. 5

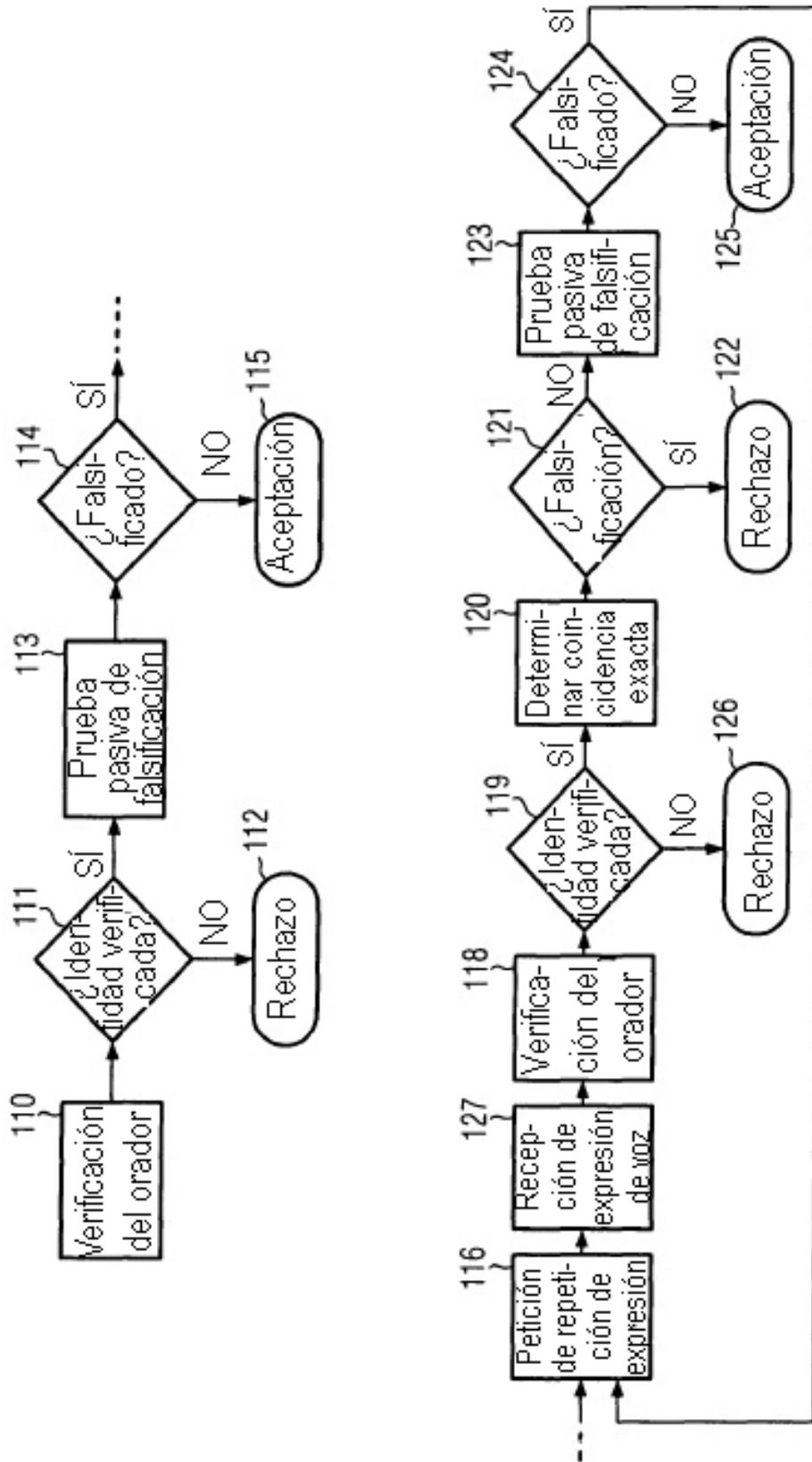


FIG. 6