

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 478 627**

51 Int. Cl.:

G10L 15/22 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.11.2007 E 07120606 (4)**

97 Fecha y número de publicación de la concesión europea: **16.04.2014 EP 1927980**

54 Título: **Método para clasificar el lenguaje hablado en sistemas de diálogo hablado**

30 Prioridad:

01.12.2006 DE 102006057159

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

22.07.2014

73 Titular/es:

**DEUTSCHE TELEKOM AG (100.0%)
FRIEDRICH-EBERT-ALLEE 140
53113 BONN, DE**

72 Inventor/es:

**ECKERT, MARTIN;
ENGLERT, ROMAN;
JOHANNSEN, WIEBKE;
RUNGE, FRED y
VAN BALLEGOOY, MARKUS**

74 Agente/Representante:

MORGADES MANONELLES, Juan Antonio

ES 2 478 627 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

La presente invención se refiere a un método de clasificación del idioma hablado.

5 CAMPO DE LA INVENCION

10 Cuando se crea una aplicación de idioma multilingüe con clasificación automática del idioma del usuario ("Identificación de idioma"), se plantea la cuestión de cómo se puede invitar al interlocutor a que se exprese libremente en su lengua materna. Tal como se describirá a continuación, existen diversos métodos para clasificar el idioma hablado, basándose todos ellos exclusivamente en el uso de un único procedimiento para procesar el flujo acústico y/o del habla fonética. Puesto que dichas técnicas tienen adecuadamente su principal área de utilización en sistemas de diálogo de voz, se encuentran asimismo sustancialmente disponibles otras técnicas conocidas de procesamiento de voz, tales como, por ejemplo, el reconocimiento de voz y/o la verificación del hablante que sirven para reconocer de palabras y/o frases habladas o verificar o identificar el hablante. Se dispone además de datos relacionados con el reconocimiento de usuarios adicionales en los sistemas de diálogo de voz que pueden proporcionar, explícita o implícitamente, indicaciones adicionales de la lengua hablada (por ejemplo, alemán, inglés, francés, español, turco, etc.) que un usuario particular de una aplicación de diálogo de voz puede utilizar. La utilización por separado de técnicas para clasificar la lengua hablada es aún parcialmente defectuosa. Por lo tanto, se debe tener como objetivo, posiblemente mediante el uso conjunto de técnicas adicionales, reducir dicho índice de error.

20 Se conoce el diseño general de los sistemas de diálogo que se utilizan en lingüística y, de un modo similar, la interacción multimodal con por lo menos un usuario (véase asimismo <http://www.w3.org/TR/2000/WD-voice-intro-20001204/>). Las aplicaciones monolingües de diálogo de voz no satisfacen completamente los requisitos de un grupo de clientes multilingüe. Por este motivo, se están desarrollando aplicaciones de diálogo de voz que utilizan la técnica de clasificación del idioma (o identificación del idioma), que se basan en el tipo de expresión del interlocutor para identificar el idioma que habla (véanse las figuras 2a y 2b), a fin de, cuando sea posible directamente tras la primera expresión de la salida de voz, y las gramáticas de reconocimiento de voz, cambiar al idioma correspondiente. Para ello, se debe notificar al usuario de que existe la posibilidad de utilizar la aplicación en un idioma distinto del idioma base.

25 Un procedimiento habitual para notificar a un usuario de dicha posibilidad de utilizar una aplicación de voz en un idioma distinto de un idioma principal es la utilización de un mensaje voz indicando la información correspondiente. Por lo tanto, tras el saludo inicial, se podría preguntar al usuario: "*To use the service in English, say English; pour le service en Français, dites Français, um den deutschen Dienst zu nutzen, sagen Sie Deutsch,...*" (véase [7]). Dependiendo de la respuesta del usuario, se cambia la aplicación de diálogo, sin utilizar técnicas adicionales de reconocimiento de voz, al idioma correspondiente.

30 Un método adicional para emitir un mensaje de voz sin extender sustancialmente la salida se da a conocer en el documento 2006E00083.

35 Si se utiliza la técnica de identificación del idioma, el usuario ya no tiene que indicar explícitamente la denominación del idioma, sino que puede responder directamente a una pregunta apropiada para el objetivo de la aplicación de diálogo en su lengua materna. La formulación de esta pregunta debe hacer referencia a dicha posibilidad.

40 Los mensajes de voz a los que se hace referencia se pueden generar tanto con la ayuda del idioma indicado como utilizando la síntesis de voz (conversión de texto en voz). Los sistemas actuales de conversión de texto en voz utilizan asimismo métodos para reproducir textos secuenciales de idioma mixto acústicamente de un modo secuencial, adaptando la pronunciación a las peculiaridades fonéticas de diversos idiomas, (Idioma mixto inglés, véase asimismo <http://www.loquendo.com/es/news/whitepaper-ML.htm>).

45 La clasificación del idioma o "Identificación del idioma" (L-ID) en general se realiza inmediatamente después de la primera pantalla de aviso basándose en la primera expresión del interlocutor. Por lo tanto, el punto más importante consiste en invitar al interlocutor a que se exprese de entrada en su lengua materna.

50 Una vez el interlocutor se ha expresado, se determina generalmente mediante la clasificación del idioma qué lengua ha utilizado. Si se ha identificado de un modo fiable uno de los posibles idiomas extranjeros, se realizan automáticamente los ajustes necesarios que dependen del idioma de la aplicación del diálogo y se cambia al idioma de diálogo correspondiente.

55 Se conocen asimismo métodos en los que el reconocimiento de voz y/o la clasificación del idioma no son únicamente aplicables en tiempo real a un flujo acústica de voz, sino a una región de memoria intermedia con la información acústica digitalizada en forma de regiones de almacenamiento o archivos.

60 Los documentos WO 97/43707A y EP 693 828 dan a conocer un sistema de reconocimiento del idioma con un método de reconocimiento de voz y un método de clasificación del idioma.

Es sustancialmente posible activar la gramática para un reconocedor de voz y un diálogo de voz en diversos idiomas simultáneamente, de tal modo que las respuestas se puedan reconocer asimismo simultáneamente en una pluralidad de idiomas. Sin embargo, en el caso de vocabularios relativamente grandes, ello provoca un aumento en las posibilidades de confusión entre los elementos individuales de la gramática, por lo que se utilizan gramáticas multilingües únicamente en casos seleccionados.

SUMARIO DE LA INVENCION

Constituye un objetivo de la presente invención proporcionar un método y un sistema eficientes de identificación de voz.

El objetivo se alcanza mediante un método y un dispositivo según las reivindicaciones independientes.

Asimismo, la presente invención comprende conceptos relacionados la vinculación útil de los resultados de los métodos que se utilizan en la clasificación de los idiomas en el contexto de la identificación de la lengua hablada (por ejemplo, alemán, inglés, francés, español, turco, etc.) y el reconocimiento de voz en el contexto de la conversión del idioma hablado en cadenas de caracteres, estando dichas vinculaciones destinadas a aumentar la fiabilidad de la decisión con respecto a un idioma con el que se realizará un diálogo hablado entre el sistema y el usuario, en comparación con una utilización por separado de los métodos.

Asimismo, los resultados de las técnicas de clasificación del idioma y de reconocimiento de voz comprenden un cierto grado de incertidumbre de tal modo que, por ejemplo, resulta totalmente posible que el sistema de clasificación del idioma reconozca una expresión hablada en alemán como una expresión en alemán, pero en un sistema de reconocimiento de voz ajustado para el alemán no proporcione resultado alguno y un sistema de reconocimiento de voz ajustado para el inglés proporcione una palabra parecida o una secuencia de palabras con significado como resultado menos plausible, lo que provocaría incertidumbre al tomar la decisión de un idioma en particular. En este caso, por ejemplo, pueden ser necesarias nuevas medidas de diálogo para clarificar la lengua hablada. Se pueden expresar dichas incertidumbres, por ejemplo, como probabilidades o valores de confianza.

Es asimismo posible, por ejemplo, la clasificación de una expresión en lengua alemana como inglesa y que el reconocimiento de secuencias de palabras con significado con un proceso de reconocimiento de voz determine alemán. En este último caso, un sistema debe avisar antes de continuar analizando el flujo de voz a fin de obtener unos resultados plausibles.

En el caso más simple, se pueden tomar las decisiones de un modo natural con la lógica de toma de decisiones implementada en el sistema cuando un sistema de reconocimiento de voz ajustado a un idioma en particular proporciona resultados útiles y el sistema de clasificación del idioma confirma el idioma preestablecido.

Para minimizar el número de decisiones erróneas con respecto a un idioma determinado, se debe utilizar una lógica de toma de decisiones a fin de controlar la plausibilidad de los vínculos de los resultados.

Mediante el método conjunto de múltiples etapas descrito según la presente invención, en comparación con la utilización por separado de clasificación del idioma, se alcanza una reducción sustancial del índice de errores en la clasificación del idioma hablado en los sistemas de diálogo de voz. Resulta ventajoso que no se requieran recursos adicionales, pero que se alcancen unos índices satisfactorios al obtener el resultado final simplemente utilizando una combinación adicional de los sistemas de reconocimiento de voz ya presentes en los sistemas de diálogo de voz.

Al combinar la técnica de clasificación del idioma con las posibilidades de un sistema de diálogo de voz y otras técnicas utilizadas en el mismo (por ejemplo, reconocimiento de voz), se mejora en gran medida la fiabilidad del resultado (el idioma determinado).

BREVE DESCRIPCION DE LOS DIBUJOS

Para una mejor comprensión de las formas de realización, estos se describirán a continuación haciendo referencia a los dibujos, en los que:

la figura 1 representa un sistema de diálogo por voz con clasificación del idioma;

la figura 2 muestra una representación de una secuencia para analizar la señal de voz a fin de cambiar el idioma utilizando el sistema de clasificación del idioma;

la figura 2b representa una secuencia para analizar la señal de voz cuando se continúa el diálogo en el idioma principal;

la figura 3 muestra un diagrama de flujo del proceso de múltiples etapas para la clasificación automática del idioma;

la figura 4 muestra un diagrama de flujo del análisis de la primera expresión del interlocutor;

5 la figura 5a muestra un diagrama de flujo del método para realizar la pregunta con respecto a un cambio de idioma con respecto a la figura 2;

la figura 5b muestra un diagrama de flujo para la escalada multilingüe con respecto a la figura 2b;

10 la figura 6 muestra un diagrama de flujo de un método reducido para la clasificación automática del idioma;

la figura 7 muestra una tabla con las denominaciones de los idiomas;

la figura 8 muestra una tabla con un ejemplo de matriz de decisiones para la etapa 1;

15 la figura 9 muestra una tabla con un ejemplo de matriz de decisiones para la etapa 2a;

la figura 10 muestra una tabla con un ejemplo de matriz de decisiones para la etapa 2b;

DESCRIPCIÓN DE LAS FORMAS DE REALIZACIÓN PREFERIDAS:

20 Para alcanzar el objetivo, un sistema de diálogo de voz 6 presenta unos medios para clasificar por lo menos dos idiomas 3 y por lo menos un sistema de reconocimiento de voz 2, según la figura 1. El sistema de reconocimiento de voz consiste tanto en por lo menos un reconocedor de voz que puede procesar por lo menos dos idiomas en paralelo y/o en secuencia como en una pluralidad de reconocedores de voz 2, 4a, respectivamente, que son aptos para procesar distintos idiomas. Los elementos adicionales son un dispositivo para la lógica de toma de decisiones 7 y una red 5 a través de la que se transmiten los datos de la persona que habla.

25 El (los) reconocedor(es) de voz 2, 4a proporcionan por lo menos los valores y/o señales de resultado siguientes que tienen el significado siguiente:

- 30 • coincidencia ("*match*") (se ha reconocido una entrada de voz válida)
- no coincidencia ("*nomatch*") (la entrada de voz presente no es válida) y/o
- 35 • no entrada ("*noinput*") (no se ha dicho nada, ha caducado el período de espera de una expresión del usuario)

El sistema de clasificación del idioma (o "sistema de identificación del idioma") 3 proporciona por lo menos los valores y/o señales de resultado siguientes que tienen el significado siguiente:

- 40 • por ejemplo, alemán, inglés, francés u otra denominación e/o identificador de por lo menos un idioma clasificado
- desconocido (idioma desconocido)

Además, se pueden producir unos valores (valores de probabilidad, valores de confianza) que reflejen la fiabilidad de los resultados (es decir, el valor de confianza).

45 Tal como se ha descrito y explicado en la sección anterior, según la presente invención, los resultados que proporcionan los métodos de clasificación del idioma 3 son asimismo los resultados de los métodos de reconocimiento de voz 2 y 4. Asimismo se analiza 3 y 4 la expresión que se transfiere al sistema de clasificación del idioma 3 mediante el reconocimiento de voz. En la primera etapa, se analiza únicamente si el resultado es coincidencia (se ha reconocido una entrada válida), no coincidencia (no se ha reconocido una entrada válida) o no entrada (no se ha dicho nada). En este punto no es importante el contenido real de la entrada. Si la clasificación del idioma 3 devuelve un idioma distinto del idioma principal activo en ese momento, la señal que se encuentra en la memoria se analiza por segunda vez mediante el reconocimiento de voz 4, aunque esta vez se basa en la gramática del idioma reconocido por la clasificación del idioma 3. Asimismo en este caso, los dos resultados de estado fundamentales son: coincidencia (entrada válida) o no coincidencia (entrada no válida).

La figura 2a representa esquemáticamente la secuencia y el flujo de la señal del análisis de la señal de voz. El ejemplo seleccionado presenta una secuencia en la que un interlocutor de habla inglesa se conecta a una aplicación de diálogo de voz de habla alemana: un interlocutor de habla inglesa llama a una aplicación de diálogo de voz alemana y habla en inglés desde el principio. El reconocedor de voz 2, que se ha ajustado a alemán no reconoce la expresión y por lo tanto emite una no coincidencia (entrada no válida). La clasificación del idioma 3 clasifica la expresión como inglés. La señal que se encuentra en la memoria intermedia 4 se analiza de nuevo utilizando un reconocedor de voz ajustado a inglés y se reconoce (coincidencia, entrada válida). Este puede ser otro reconocedor de voz o el reconocedor de voz ajustado ahora a inglés en los modelos de gramática y acústicos, pero que se ajustó en un principio a alemán en el primer método descrito. El diálogo continúa en inglés. Cabe indicar que los módulos

individuales pueden funcionar asimismo en paralelo, en particular si se dispone de una potencia de cálculo suficiente.

5 La figura 2b representa la continuación del método en el caso en que se ha identificado correctamente el idioma. Se continúa el diálogo utilizando el idioma preestablecido.

10 La decisión de cómo proceder después de la entrada de voz depende, por lo tanto, de la combinación de los resultados de la clasificación del idioma 3 y el reconocimiento de voz 2. Si el interlocutor 1 no proporciona una expresión activa, el reconocimiento de voz devuelve el valor no entrada (falta entrada). Se debe tener en cuenta asimismo este caso también cuando se toma una decisión. El modo en que se continuará el proceso tras la evaluación de la primera expresión se puede adaptar individualmente y depende principalmente de los recursos disponibles. En principio, se debe decidir si además de, por ejemplo, la aplicación de diálogo de voz alemana, deberían existir aplicaciones de diálogo de voz estructuradas idénticamente en los idiomas de destino. La ventaja de este enfoque identificado como general en la presente memoria es que todos los interlocutores se mantienen en el diálogo hablado y no se produce un aumento del volumen de llamadas. Un inconveniente es la traducción compleja a otros idiomas puesto que, aparte de la creación del mensaje de voz, debe tenerse en cuenta asimismo que se tienen que generar las gramáticas de los idiomas de destino. En el caso de aplicaciones de idioma con cambios constantes de contenidos, aumenta en consecuencia la necesidad de mantenimiento.

20 Si se esperan únicamente unos pocos interlocutores en lengua extranjera, en determinadas circunstancias, no se justifica el esfuerzo necesario para generar la aplicación de diálogo en lengua extranjera. En este caso, un procedimiento más simple podría consistir en la transferencia a un asesor que hablase la lengua extranjera. A continuación se describirán soluciones y ejemplos de aplicaciones para ambos procedimientos. Se pueden prever asimismo soluciones mixtas.

25 El objetivo de la identificación conjunta de la lengua hablada en el procedimiento general es asignar el interlocutor al diálogo hablado que está configurado en su lengua materna. Ello se realiza con el siguiente método de múltiples etapas (véase asimismo la figura 3).

30 Después de la primera expresión del interlocutor, en los casos en que la decisión relativa a la clasificación del idioma indica probablemente o definitivamente el idioma actual, se continúa el diálogo sin cambio de idioma alguno (continúa sin cambiar de idioma). Si valora como cierta una asignación a un idioma que no es el idioma actual, el interlocutor recibe información con respecto al cambio de idioma y el diálogo continúa en el nuevo idioma de destino (información sobre el cambio de idioma y continuación del diálogo). En ambos casos, la clasificación del idioma se desactiva en este punto.

40 Si, sin embargo, se identifica únicamente una tendencia a otro idioma de destino, se realiza una pregunta directa en cuanto a si el interlocutor desea continuar en el idioma de destino reconocido (pregunta, cambio de idioma). Se analiza la respuesta a dicha pregunta en la segunda etapa (etapa 2a), asimismo con la ayuda del sistema de clasificación del idioma. Dependiendo de la respuesta a la pregunta y/o la combinación con el resultado de la clasificación, se mantiene o se cambia el idioma de diálogo. Si la entrada del interlocutor no proporciona ninguna asignación definitiva, se puede transferir el interlocutor a un asesor que debe dominar tanto el idioma actual como el idioma alternativo que posiblemente se ha determinado. Si el resultado de la clasificación del idioma en la primera etapa es tan incierta que no se puede reconocer una tendencia clara a un idioma (por ejemplo, debido a la falta de entrada (no entrada), el interlocutor entra en la escalada multilingüe. Una vez aquí, se le pregunta una vez más en todos los idiomas posibles previstos (por ejemplo, tres) para expresar su preferencia. Aquí también, en la segunda etapa (etapa 2b), se utiliza de nuevo la clasificación del idioma. Dependiendo del resultado, en este caso también, se mantiene el diálogo o se transfiere a un asesor que, en el mejor de los casos, domina todos los posibles idiomas de destino.

50 Los siguientes párrafos describen cómo los valores del resultado controlan la decisión del sistema. La figura 4 muestra los detalles de ello.

55 Tal como se ha descrito anteriormente, el método según la presente invención se basa en la clasificación automática del idioma, de tal modo que se puede asignar el interlocutor al idioma de destino correcto. En general, existe una pluralidad de aspectos secundarios para los que se utiliza la clasificación del idioma. A continuación se describirá detalladamente el modo preciso de funcionamiento de dichas tres etapas. Se utilizan las denominaciones según la tabla de la figura 8 para los idiomas implementados. El método según la presente invención se basa en el análisis conjunto de los resultados iniciales de por lo menos un método de reconocimiento de voz, la clasificación del idioma y por lo menos un posible segundo método de reconocimiento de voz posterior. Puesto que los clasificadores de este tipo producen siempre con un cierto porcentaje de errores en el reconocimiento, todas las posibles combinaciones de las tres etapas a analizar deben examinarse con precisión para permitir que el sistema reaccione con sentido a la situación. Aquí se tienen en cuenta algunas reglas fundamentales:

60 Si se ha obtenido un resultado de reconocimiento de voz válido, en caso de duda se utiliza puesto que en este caso se debe suponer que el interlocutor ha hablado en el idioma preestablecido (por defecto).

A continuación se cambia el idioma automáticamente y sin consultas si los tres resultados del análisis apoyan esta tesis.

Si se puede identificar una tendencia clara con respecto a un idioma, se pregunta al interlocutor si se desea cambiar de idioma.

5 Si no se puede aplicar ninguna de las reglas descritas anteriormente, en la etapa siguiente se debe utilizar la escalada multilingüe.

10 La tabla de la figura 8 representa las posibles combinaciones y las reacciones del sistema correspondientes de la primera etapa del método de clasificación conjunta del idioma. En la columna 2 (L-ID), se proporciona únicamente el primer idioma extranjero; la reacción es idéntica para cada idioma extranjero adicional. El diagrama de flujo de la figura 4 aclara de nuevo la secuencia del árbol de decisiones.

15 Una vez el interlocutor ha expresado su requisito, en la primera etapa, que es la etapa 1, se realiza un análisis mediante el reconocedor de voz actual (2). Si el valor resultante es una no entrada (falta entrada), se utilizará una estrategia de escalada multilingüe (véase también la tabla de la figura 10) para invitar al interlocutor (1) a que hable (idealmente en su lengua materna). A continuación se describirá detalladamente la escalada multilingüe.

20 Si el reconocedor de voz (2) devuelve una coincidencia (entrada válida), existe una probabilidad elevada de que el idioma sea el idioma preestablecido (por defecto), es decir, el idioma actualmente activo. Sin embargo, se puede prever la detección de errores, de tal modo que a continuación se procede al análisis de la clasificación del idioma (4). Si ello devuelve asimismo como resultado un idioma preestablecido (por defecto), continúa el diálogo sin cambio de idioma y se puede desactivar la clasificación del idioma (3).

25 Lo mismo ocurre con las expresiones para las que no se puede asignar el idioma (desconocido). Esta decisión se basa en que un análisis preciso de la clasificación del idioma (3) no es posible en este caso.

30 Sin embargo, puesto que existía una expresión válida como resultado del primer proceso de reconocimiento de voz (2), se continúa el diálogo basándose en dicha expresión. Si se mediante el resultado de la clasificación del idioma (3) se descubre que otro idioma se encuentra implicado (por ejemplo, idioma 1/ idioma 2; por defecto / idioma 1), la expresión que se mantiene en la memoria intermedia (4) se analiza de nuevo mediante un reconocedor de voz, esta vez con los modelos de gramática y acústicos del idioma de destino. Si el resultado es una coincidencia (entrada válida), es probable que el interlocutor hay utilizado un idioma distinto al idioma preestablecido. Por lo tanto, se le ofrece el cambio al idioma de destino reconocido. Esta oferta se realiza tanto en el nuevo idioma de destino así como en el idioma establecido actual. Si el segundo proceso de reconocimiento de voz devuelve una no coincidencia (entrada no válida) al utilizar el contenido de la memoria intermedia (4), en este caso también, se devuelve coincidencia (entrada válida) mediante el primer proceso de reconocimiento de voz (2). Se continúa el diálogo sin cambio de idioma. En el caso de que el primer proceso de reconocimiento de voz (2) devuelva una no coincidencia (entrada no válida), se utiliza también la clasificación del idioma (3). Si ello devuelve el valor desconocido (no se pudo determinar el idioma) o por defecto (el idioma preestablecido), no resulta útil un nuevo análisis mediante el segundo proceso de reconocimiento de voz utilizando la memoria intermedia (4). Se debe seguir la ruta de escalada multilingüe, es decir, se pregunta al interlocutor en todos los idiomas disponibles (por ejemplo, tres), para que repita su preferencia. En otros casos, se realiza un análisis adicional mediante el segundo proceso de reconocimiento de voz utilizando la memoria intermedia (4) en el idioma de destino reconocido. Si ello devuelve una coincidencia (entrada válida), las tres etapas de reconocimiento devuelven el resultado coherente que es una expresión en un idioma distinto al idioma preestablecido (por defecto). Se informa al interlocutor mediante el cambio de idioma y se continúa el diálogo en el nuevo idioma de destino. Si el segundo proceso de reconocimiento de voz devuelve una no coincidencia (entrada no válida), utilizando el contenido de la memoria intermedia (4), no se puede realizar una declaración con esta claridad. Sin embargo, puesto que permanece una tendencia importante en la dirección de un nuevo idioma de destino, también en este caso, se ofrece al interlocutor un cambio de idioma.

50 En este punto, cabe indicar que, en la forma de realización este método supone la posibilidad de cambiar fácilmente a otro modo de acción si se percibe que un cambio no solicitado a un nuevo idioma resulta demasiado complicado. Se implementa la matriz de decisiones en el sistema de tal modo que se puede realizar la alteración de un sistema de reacción mediante una simple entrada de tabla. Ello es válido de un modo similar para la matriz de decisiones, etapa 2a (tabla de la figura 9) y la etapa 2b (tabla de la figura 10).

60 Se alcanza la etapa 2a (tabla de la figura 9) del método cuando se puede reconocer una tendencia clara a un idioma. En este caso, se pregunta al interlocutor si desea continuar en la segunda lengua reconocida. Puesto que todavía existe incertidumbre con respecto a la lengua de destino y no se puede suponer que el interlocutor entiende realmente la lengua extranjera, se plantea la cuestión, por ejemplo, en dos idiomas. Si el interlocutor se encuentra en, por ejemplo, una aplicación de diálogo de voz en alemán y si existe la probabilidad de que haya hablado en inglés, la pregunta correspondiente es: "Do you want to continue in English? Möchten Sie in Englisch weitermachen?". La respuesta a dicha pregunta puede la puede dar el interlocutor en cualquiera de los (o tres) idiomas proporcionados ya que en este punto, se implementa una gramática que comprende modelos acústicos aptos, que comprende "sí / no" en todos los idiomas previstos. Debido a que se está utilizando una gramática

universal, se proporciona el segundo proceso de reconocimiento de voz, ya que no se puede disponer de un resultado nuevo.

5 La cuestión del cambio de idioma representa simultáneamente una peculiaridad adicional de la etapa 2a puesto que el requerimiento real del interlocutor se "interrumpe" por dicha pregunta intermedia. En el caso de una coincidencia anterior (entrada válida) en la etapa 1, se almacena el contenido del reconocedor de voz para poder continuar - tras responder a la pregunta sobre el cambio de idioma - desde el mismo lugar, ya que se ha interrumpido el diálogo.

10 Por lo tanto, en el caso de la matriz de decisiones 2a (tabla de la figura 9), debe tenerse en cuenta asimismo la salida de los dos procesos de reconocimiento de voz. Puesto que la cuestión del cambio de idioma se puede alcanzar únicamente si cualquiera de los dos procesos de reconocimiento devuelven una coincidencia (entrada válida) o en ambos procesos de reconocimiento devuelven una no coincidencia (entrada no válida), cuando se establece la matriz de decisiones, resulta suficiente una visión general con la coincidencia / no coincidencia (entrada válida / no válida), es decir, en este punto no se debe tomar ninguna decisión sobre cuál de los dos procesos de reconocimiento proporciona el resultado o se utiliza en el procedimiento adicional. La condición de coincidencia (entrada válida) se sustituye en este caso con las respuestas posibles "sí" y "no", ya que representan el deseo explícito del interlocutor y se deben tener en cuenta en todos los casos. Por lo tanto, en estos casos, no se continúa el análisis con la clasificación del idioma.

20 Una peculiaridad adicional consiste en que es teóricamente posible - aunque muy poco probable - que la clasificación del idioma reconozca, desde la primera etapa, un idioma distinto al de la etapa 2a. Por lo tanto, se debe tener en cuenta dicha información en los sitios indicados. Los dos estados de entrada mostrados en la figura 5a indican las posibles combinaciones resultantes de la etapa 1, que originan la pregunta sobre el cambio de idioma. En general, cabe indicar que en el ejemplo de especificación del método, en caso de duda, se supone que el interlocutor es una persona que utiliza el idioma de preestablecido (por defecto).

25 En resumen, cabe indicar que, tras la etapa 2a, se puede desactivar la clasificación del idioma y, posteriormente, cada combinación de valores de los resultados debe originar un estado en el que se puede proporcionar el deseo del interlocutor. En este punto, cabe indicar que este método representa una forma de realización posible.

30 La escalada multilingüe, denominada etapa 2b en la presente memoria (véase la tabla de la figura 10) es mucho menos compleja en su estructura que la etapa 2a en el que se realiza la pregunta de un cambio de idioma. Ello se basa en que el flujo de diálogo no se interrumpe mediante una pregunta adicional. Por otro lado, los estados de entrada son mucho menos complejos. La escalada multilingüe se puede alcanzar únicamente con las combinaciones: no coincidencia - por defecto (entrada no válida - idioma predeterminado), no coincidencia - desconocido (entrada no válida - no se pudo determinar el idioma) y no entrada (sin entrada) (véase la figura 8). Por lo tanto, se puede suponer sustancialmente que es un interlocutor que utiliza el idioma actual o que el interlocutor ha hablado en un idioma que no está implementado. En este caso, puesto que no existe ninguna declaración que permita realizar un análisis, se requerirá en todo caso que el usuario hable de nuevo. Por lo tanto, no es necesario un análisis de los resultados de la primera etapa. Puesto que, en todo caso tras la escalada multilingüe se ha obtenido la clasificación conjunta del idioma, en el caso de que no se pueda realizar una asignación renovada, el interlocutor se transfiere a un asesor con conocimientos de los idiomas extranjeros correspondientes

45 Los tres estados de entrada mostrados en la figura 5b indican las posibles combinaciones de resultados de la etapa 1. En general, cabe indicar que en el ejemplo de especificación del método, en caso de duda, se supone que el interlocutor es una persona que utiliza el idioma de preestablecido (por defecto).

50 En este punto, cabe indicar de nuevo que los métodos descritos en la presente memoria que utilizan la matriz de decisiones representan una forma de realización posible.

55 El método general descrito en las últimas secciones representa una solución en la que se realiza el diálogo hablado en todos los idiomas de destino. En particular, en el caso de que el contenido cambie rápidamente, ello supone un gran esfuerzo, que no se justifica en todos los casos. Una alternativa al mismo consiste en la transferencia a un asesor en el "procedimiento abreviado" (véase asimismo la figura 6) que domine el idioma correspondiente. Se podrían prever asimismo soluciones mixtas en las que únicamente las ramas de diálogo más frecuentes estén representadas en todos los idiomas.

60 Se utiliza asimismo la clasificación automática del idioma para la transferencia específica a un asesor con conocimientos de los idiomas extranjeros correspondientes. El principio no difiere del método general descrito anteriormente, ya que aquí también, en la primera etapa, el interlocutor llega a una aplicación de diálogo de voz multilingüe. Es únicamente después de la primera expresión que el sistema puede asignar la lengua y decidir dónde se debe conectar el interlocutor.

65 En principio, con el método reducido descrito en la presente memoria todos los interlocutores que hablen en un idioma extranjero deben conectarse con un asesor que hable su propia lengua materna. Además, se debe impedir

que los interlocutores que hablan el idioma principal se conecten innecesariamente con el asesor. Por consiguiente, el método se estructura en dos etapas.

5 En la primera etapa, los interlocutores que pueden asignarse inequívocamente a un idioma se conectan según la configuración predeterminada. Los interlocutores que utilizan el idioma principal se transfieren a la aplicación del diálogo de voz correspondiente y los interlocutores que hablan en un idioma extranjero se transfieren a un asesor que habla su lengua materna.

10 En función del resultado del análisis de la primera clasificación conjunta del idioma, se pregunta directamente a todos los demás interlocutores si desean hablar en otro idioma o se les pasa a la escalada multilingüe. Estas dos etapas se realizan de un modo similar a las etapas descritas en el método general. En lugar de transferirse a la aplicación de diálogo de voz correspondiente, también en este caso, únicamente se transfiere el interlocutor que habla en el idioma principal a la aplicación de diálogo de voz, mientras que todos los demás interlocutores se transfieren a un asesor en el idioma nacional correspondiente. Si todavía no se ha determinado el idioma, incluso
15 después de la segunda etapa de la clasificación conjunta del idioma, en un caso ideal el interlocutor se conecta con un asesor que domina los tres idiomas. Si ello no resulta posible (por ejemplo, porque se incorporan muchos más idiomas), se debe seleccionar un asesor que domine los posibles idiomas clave (por ejemplo, inglés o español).

20 Se debe comprender que las partes, o incluso la totalidad, de la presente invención se pueden realizar mediante programas informáticos que funcionan junto con el hardware.

La forma de realización preferida no representa una limitación. Pretende únicamente facilitar la comprensión. La protección se define mediante las reivindicaciones siguientes.

REIVINDICACIONES

- 5 1. Método para la clasificación de por lo menos dos idiomas en un sistema automático de diálogo (6) que procesa la entrada de voz digitalizada, que comprende las etapas siguientes:
la utilización de por lo menos un método de reconocimiento de voz (2, 4a) y por lo menos un método de clasificación del idioma (3) en la entrada de voz digitalizada, para identificar el idioma de la entrada de voz mediante el análisis lógico (7) de los resultados de la métodos,
10 caracterizado porque dicho método de clasificación del idioma y el proceso de reconocimiento de voz se realizan en paralelo en el tiempo.
- 15 2. Método según la reivindicación anterior, en el que se utiliza el método de clasificación del idioma tras el método de reconocimiento de voz si el método de reconocimiento de voz no ha podido reconocer la entrada de voz, para clasificar la entrada de voz y, si la clasificación se realizó correctamente, se cambia el método de reconocimiento de voz y/o por lo menos un parámetro del método de reconocimiento de voz, de tal modo que el reconocimiento de voz se realiza con el método de reconocimiento de voz cambiado y/o modificado.
- 20 3. Método según cualquiera de las reivindicaciones anteriores, en el que los resultados de por lo menos un método de reconocimiento de voz y/o por lo menos un método de clasificación del idioma relacionado con la entrada de voz comprenden por lo menos una condición de resultado.
- 25 4. Método según cualquiera de las reivindicaciones anteriores, en el que el análisis lógico de los resultados devueltos de por lo menos un método de reconocimiento de voz y, si es necesario, de por lo menos un método de clasificación del idioma relacionado con la entrada de voz comprenden por lo menos una cadena de caracteres y/o un símbolo digital y/o una cadena de símbolos.
- 30 5. Método según cualquiera de las reivindicaciones anteriores, en el que la entrada de voz se almacena temporalmente en forma digitalizada.
- 35 6. Método según la reivindicación anterior, en el que además los datos de voz digitalizados determinadas a partir de la entrada de voz se almacenan temporalmente, pudiendo comprender los datos de voz uno o más de los elementos siguientes: datos de características extraídas, datos de audio.
- 40 7. Método según cualquiera de las dos reivindicaciones anteriores, en el que los datos almacenados temporalmente se describen mediante por lo menos un método de reconocimiento de voz.
- 45 8. Método según la reivindicación anterior, en el que los datos almacenados temporalmente se describen mediante por lo menos un método de clasificación del idioma.
- 50 9. Método según cualquiera de las 3 reivindicaciones anteriores, en el que los datos almacenados temporalmente en el almacenamiento temporal se describen mediante por lo menos un método que es apto para preprocesar el idioma hablado digitalizado, en particular mediante métodos de cancelación de eco, métodos reductores del ruido y métodos de codificación.
- 55 10. Método según cualquiera de las reivindicaciones anteriores, en el que el resultado del análisis mediante la lógica de toma de decisiones influye en el curso posterior del diálogo con el usuario.
- 60 11. Método según cualquiera de las reivindicaciones anteriores, en el que se establecen una lógica de toma de decisiones y/o partes de la misma en forma de reglas deterministas cambiables administrativas y/o programáticas.
- 65 12. Método según cualquiera de las reivindicaciones anteriores, en el que la lógica de toma de decisiones y/o partes de la misma se han creado recopilando y analizando datos de voz estadísticamente significativos, en particular mediante redes neuronales y/o modelos ocultos de Markov y/o lógica difusa.
13. Procedimiento según cualquiera de las reivindicaciones anteriores, que comprende las etapas siguientes:
- si se ha obtenido un resultado válido del reconocimiento de voz válido, este se mantiene para diálogos futuros, puesto que se puede suponer en este caso que el interlocutor se ha expresado en el idioma preestablecido (por defecto);
- el idioma únicamente se altera automáticamente y sin consultas si ha fallado un primer intento de reconocimiento con un método de reconocimiento de voz en una primer idioma, el método de clasificación idioma ha reconocido un nuevo idioma y si un segundo intento de reconocimiento con un método de reconocimiento de voz en el nuevo idioma reconocido por el método de clasificación idioma se ha realizado con éxito,
- si no se puede aplicar ninguna de las reglas anteriores, se debe utilizar una escalada multilingüe en una etapa siguiente.

14. Método según la reivindicación anterior, en el que si existe una probabilidad clara de un idioma mediante el reconocimiento de un idioma con una cierta probabilidad y una clasificación correspondiente, se pregunta al interlocutor si desea cambiar de idioma.
- 5 15. Método según cualquiera de las dos reivindicaciones anteriores, en el que los resultados del reconocimiento son los siguientes:
 - una coincidencia (se reconoció una entrada válida); ello se puede determinar asimismo mediante la probabilidad y/o valores de confianza;
 - una no coincidencia (no se reconoció una entrada válida); ello se puede determinar asimismo mediante la
 10 probabilidad y/o valores de confianza; o
 - una no entrada (no se dijo nada); en la que, en función de estos resultados, se cambia el idioma, se mantiene el mismo o se inician diálogos con el usuario.
16. Dispositivo para clasificar por lo menos dos idiomas en un sistema de diálogo automático (6) que procesa el
 15 habla digitalizada, que presenta unos medios y una configuración de los medios que permiten la utilización de por lo menos un método de reconocimiento de voz (2, 4a) y un método de clasificación del idioma (3) en la entrada de voz digitalizada para reconocer el idioma de la entrada mediante un análisis lógico (7) de los resultados del método, caracterizado porque los medios se adaptan de tal modo que el método de clasificación de voz y el reconocimiento de voz se realizan en paralelo en el tiempo.
- 20 17. Dispositivo según la reivindicación anterior del dispositivo, en el que los medios y el dispositivo correspondiente se configuran de tal modo que el método de clasificación del idioma se utiliza después del método de reconocimiento de voz, si el método de reconocimiento de voz no ha podido reconocer la entrada de voz, para a continuación clasificar la entrada y, si la clasificación se ha realizado correctamente, se cambia el método de reconocimiento de voz y/o por lo menos un parámetro del método de reconocimiento de voz de tal modo que el reconocimiento de voz se realiza con el método de reconocimiento de voz modificado y/o por lo menos un parámetro modificado del método de reconocimiento de voz.
- 25 18. Dispositivo según cualquiera de las reivindicaciones anteriores del dispositivo, en el que los medios se configuran de tal modo que los resultados de por lo menos un método de reconocimiento de voz y/o por lo menos un método de clasificación del idioma relacionado con la entrada de voz comprenden por lo menos una condición de resultado.
- 30 19. Dispositivo según cualquiera de las reivindicaciones anteriores del dispositivo, en el que los medios se configuran de tal modo que el análisis lógico de los resultados devueltos de por lo menos un método de reconocimiento de voz y, si es necesario, de por lo menos un método de clasificación del idioma relacionado con la entrada de voz comprenden por lo menos una cadena de caracteres y/o un símbolo digital y/o una cadena de símbolos.
- 35 20. Dispositivo según cualquiera de las reivindicaciones anteriores del dispositivo, en el que los medios se configuran de tal modo que la entrada de voz se almacena temporalmente en forma digitalizada.
- 40 21. Dispositivo según la reivindicación anterior del dispositivo, en el que los medios se configuran de tal modo que los vectores de características adicionales y/u otros datos digitalizados que se han determinado a partir de la entrada de voz y se han almacenado temporalmente, en el que los datos de voz pueden comprender uno o más de los elementos siguientes: datos de características extraídas, datos de audio.
- 45 22. Dispositivo según la reivindicación anterior del dispositivo, en el que los medios se configuran de tal modo que los datos almacenados temporalmente se describen mediante por lo menos un método de reconocimiento de voz.
- 50 23. Dispositivo según la reivindicación anterior del dispositivo, en el que los medios se configuran de tal modo que los datos almacenados temporalmente se describen mediante por lo menos un método de clasificación del idioma.
- 55 24. Dispositivo según cualquiera de las 3 reivindicaciones anteriores del dispositivo, en el que los medios se configuran de tal modo que los datos almacenados temporalmente en el almacenamiento temporal se describen mediante por lo menos un método que es apto para preprocesar el idioma hablado digitalizado.
- 60 25. Dispositivo según cualquiera de las reivindicaciones anteriores del dispositivo, en el que los medios se configuran de tal modo que el resultado del análisis mediante la lógica de toma de decisiones influye en el curso posterior del diálogo con el usuario.
- 65 26. Dispositivo según cualquiera de las reivindicaciones anteriores del dispositivo, en el que los medios se configuran de tal modo que se establecen una lógica de toma de decisiones y/o partes de la misma en forma de reglas deterministas que se pueden cambiar a partir de un punto de vista administrativo y/o programático.

27. Dispositivo según cualquiera de las reivindicaciones anteriores del dispositivo, en el que los medios se configuran de tal modo que la lógica de toma de decisiones y/o partes de la misma se han creado recopilando y analizando datos de voz estadísticamente significativos, en particular métodos de cancelación de eco, métodos reductores del ruido y métodos de codificación.

5 28. Dispositivo según cualquiera de las reivindicaciones anteriores del dispositivo, en el que los medios se configuran de tal modo que se pueden realizar las etapas siguientes:

10 - si se ha obtenido un resultado válido del reconocimiento de voz válido, este se mantiene para diálogos futuros, puesto que se puede suponer en este caso que el interlocutor se ha expresado en el idioma preestablecido (por defecto);

- el idioma únicamente se altera automáticamente y sin consultas si ha fallado un primer intento de reconocimiento con un método de reconocimiento de voz en un primer idioma, el método de clasificación idioma ha reconocido un nuevo idioma y si un segundo intento de reconocimiento con un método de reconocimiento de voz en el nuevo idioma reconocido por el método de clasificación idioma se ha realizado con éxito,

15 - si no se puede aplicar ninguna de las reglas anteriores, se debe utilizar una escalada multilingüe en una etapa siguiente.

20 29. Dispositivo según la reivindicación anterior del dispositivo, en el que los medios se configuran de tal modo que si existe una probabilidad clara de un idioma mediante el reconocimiento de un idioma con una cierta probabilidad y una clasificación correspondiente, se pregunta al interlocutor si desea cambiar de idioma.

30. Dispositivo según cualquiera de las dos reivindicaciones anteriores del dispositivo, en el que los medios se configuran de tal modo que los resultados del reconocimiento son los siguientes:

25 - una coincidencia (se reconoció una entrada válida); ello se puede determinar asimismo mediante la probabilidad y/o valores de confianza;

- una no coincidencia (no se reconoció una entrada válida); ello se puede determinar asimismo mediante la probabilidad y/o valores de confianza; o

- una no entrada (no se dijo nada); en la que, en función de estos resultados, se cambia el idioma, se mantiene el mismo o se inician diálogos con el usuario.

30 31. Soporte de datos con una estructura de datos que, después de la carga en el ordenador, permite realizar un método según cualquiera de las reivindicaciones anteriores del método.

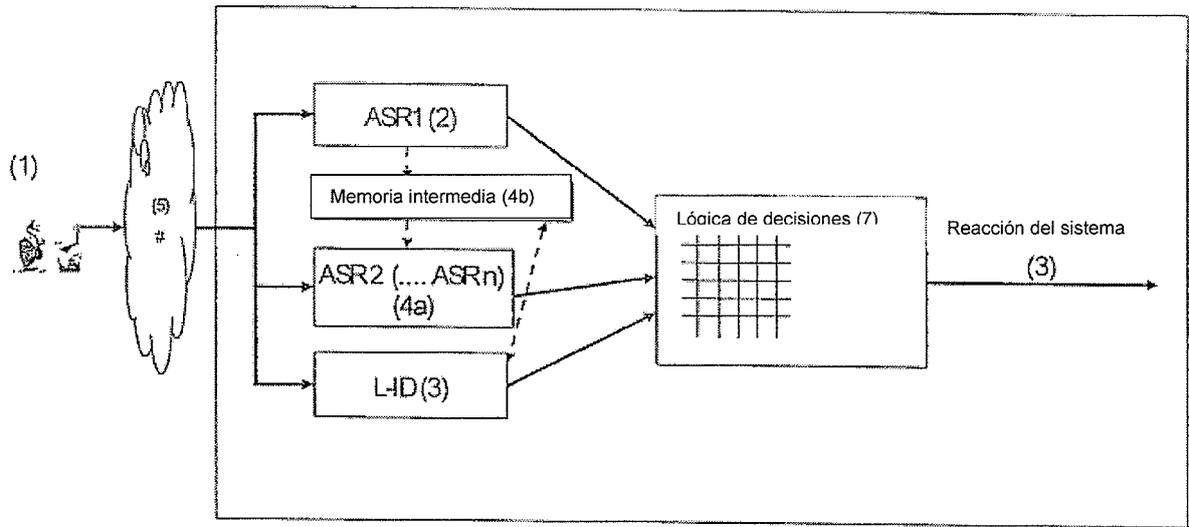


Fig.1

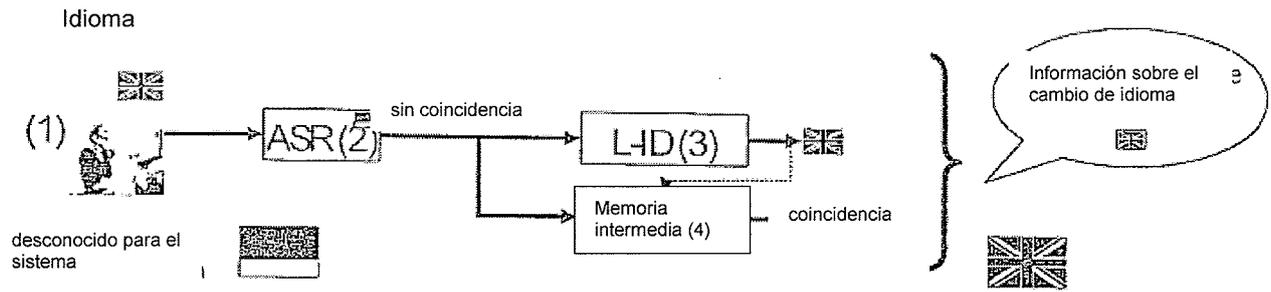


Fig.2 a

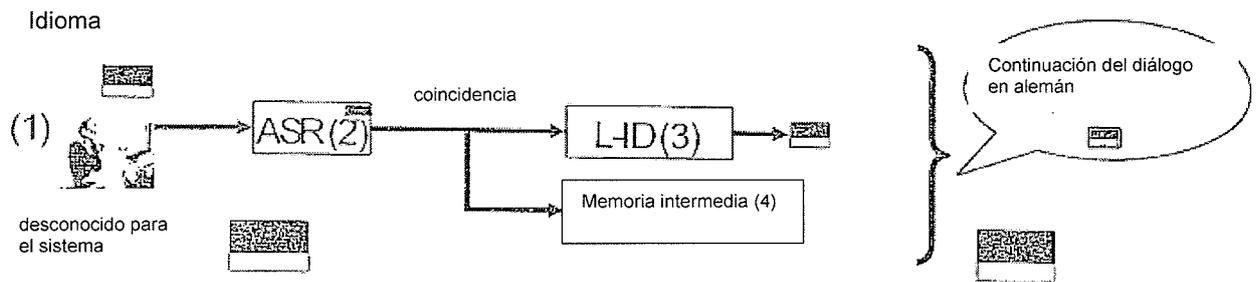


Fig.2b

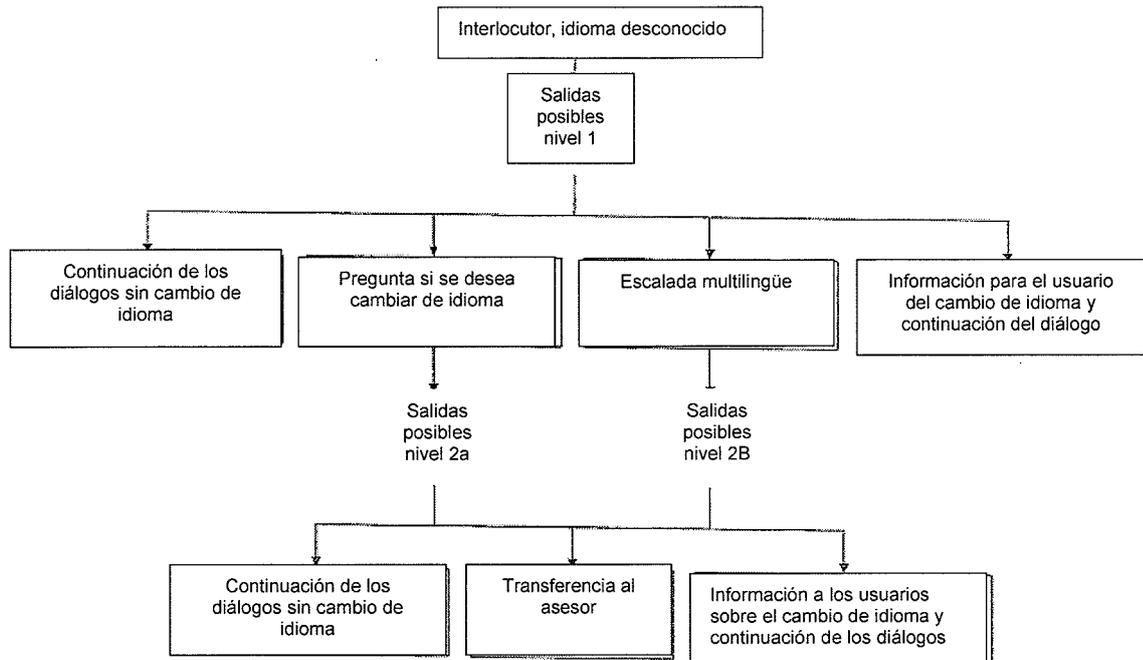


Fig.3

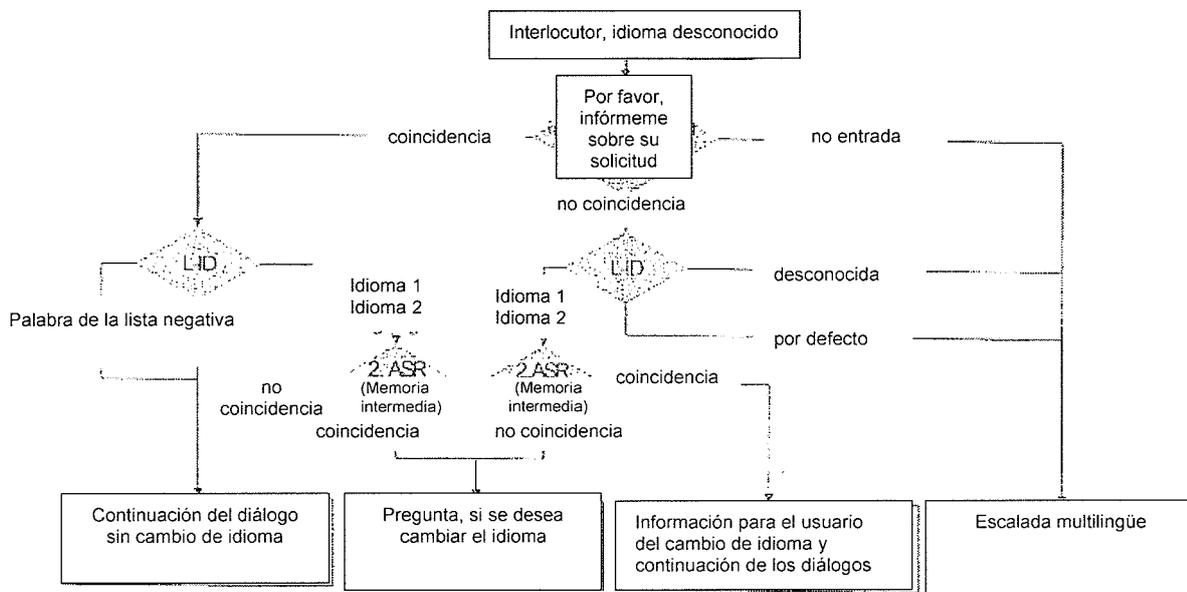


Fig.4

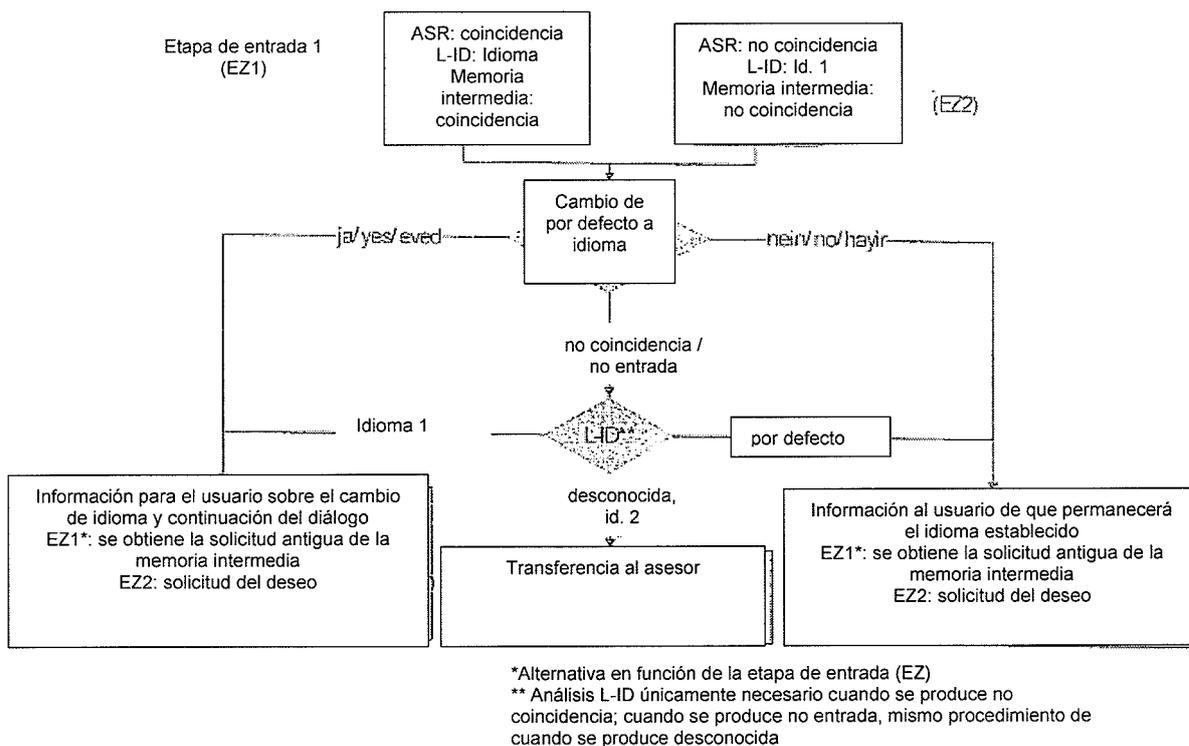


Fig.5a

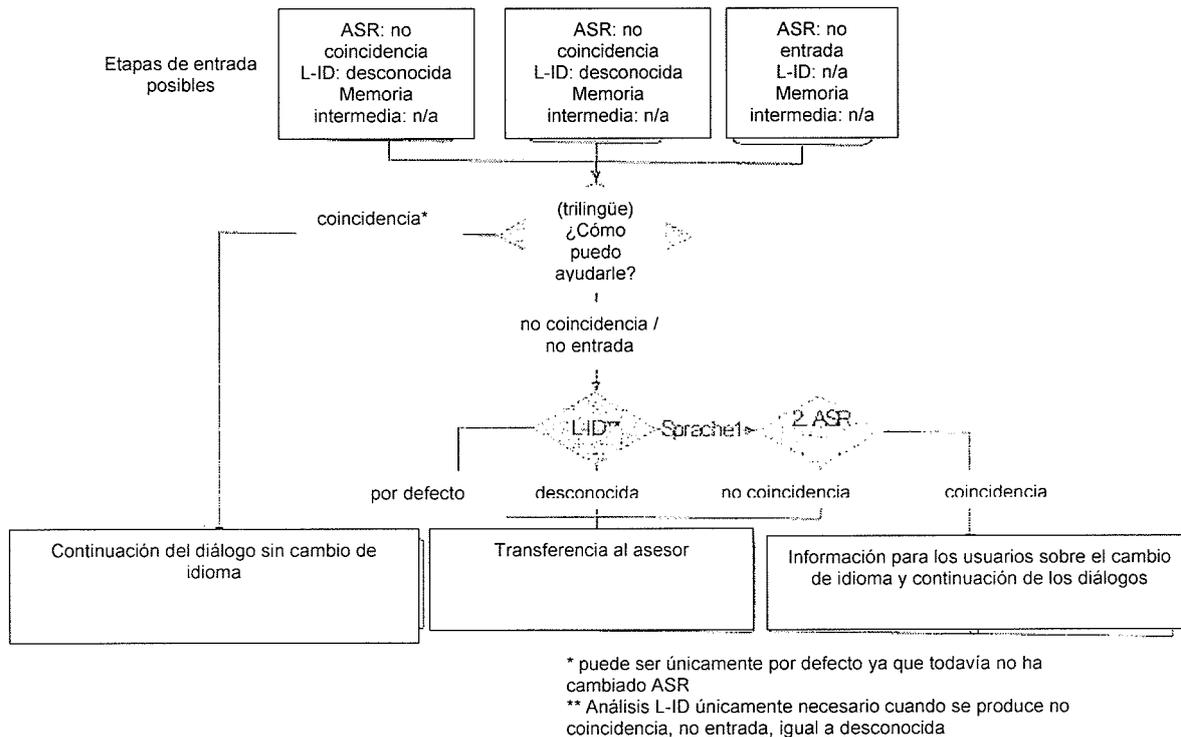


Fig.5b

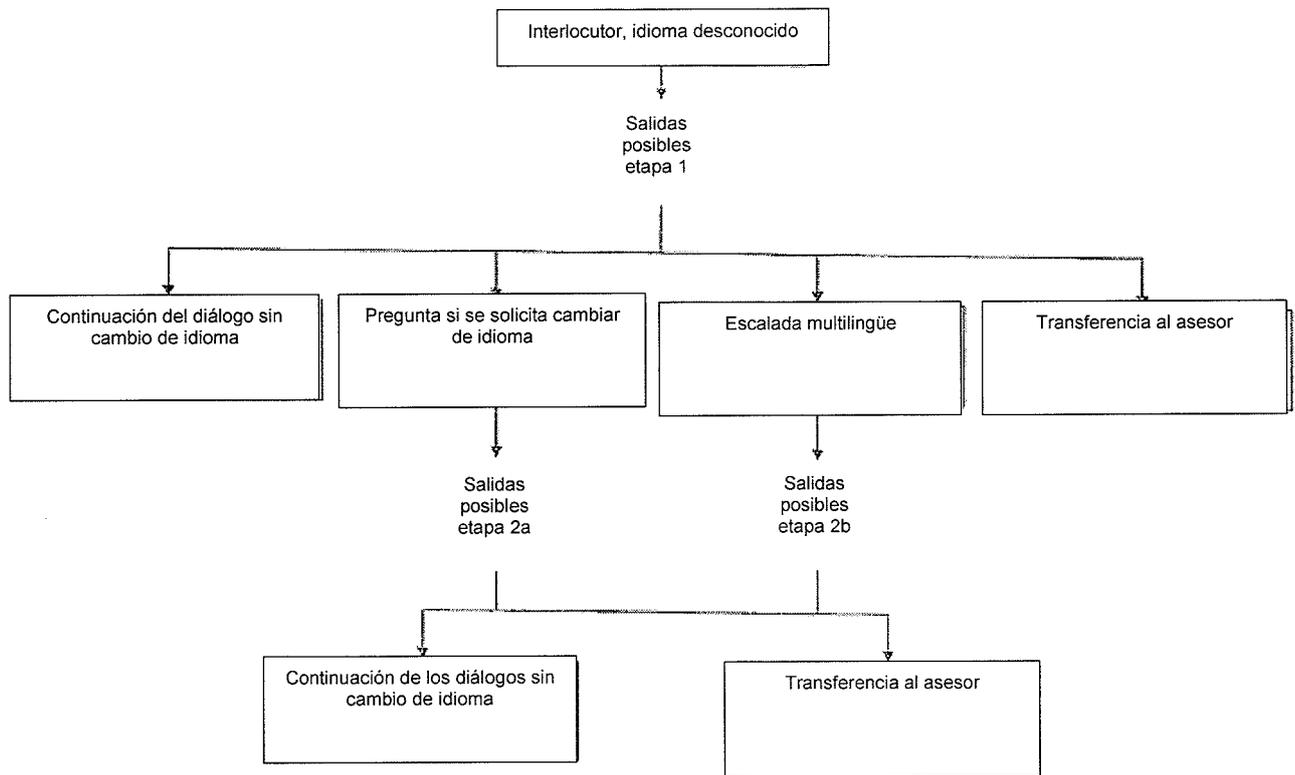


Fig.6

Descripción	Comentarios	Ejemplo
por defecto	Idioma preestablecido, que está activo cuando se inicia el diálogo	Alemán
idioma 1	Idioma en el que se proporciona adicionalmente el idioma de diálogo	Inglés
idioma N*	Uno de los idiomas en los que se proporciona adicionalmente el idioma de diálogo	Turco

* El número de idiomas adicionales y la secuencia de numeración es opcional

Fig. 7

LASR (con idioma por defecto preestablecido)	L-ID (Identificación del idioma)	2. ASR (ajustado a idioma 1, aplicación en memoria intermedia)	Validación	Reacción del sistema
coincidencia	por defecto (idioma preseleccionado)	n/a	Idioma preestablecido seguro	Continuación sin cambio de idioma *
coincidencia	Idioma 1	coincidencia	Tendencia al idioma 1	Solicitud de cambio de idioma **
coincidencia	Idioma 1	sin coincidencia	Idioma preestablecido (por defecto) factible	Continuación sin cambio de idioma *
coincidencia	desconocida (idioma desconocido)	n/a	Idioma preestablecido (por defecto) muy probable	Continuación sin cambio de idioma *
sin coincidencia	por defecto (idioma preestablecido)	n/a	Resultado dudoso	Escalada multilingüe ***
sin coincidencia	idioma	coincidencia	Tendencia definida al idioma 1	Información para los usuarios sobre el cambio de idioma y continuación del diálogo
sin coincidencia	idioma 1	sin coincidencia	Tendencia al idioma 1	Pregunta sobre el cambio de idioma
sin coincidencia	desconocida (idioma desconocido)	n/a	Resultado dudoso	Escalada multilingüe ***
No entrada	n/a	n/a	Resultado dudoso	Escalada multilingüe ***

* se desactivará la identificación del idioma de entrada tras esta etapa,
 ** se continúa con la etapa 2a de la matriz de decisiones
 *** se continúa con la etapa 2b de la matriz de decisiones
 n/a sin análisis: No se realizó ningún análisis, ya que el 1.ASR proporcionó el valor no entrada o la clasificación del idioma como clasificador de idiomas como valor devuelto del idioma del proceso 1. clasificador de idioma (2) y, por lo tanto, no es necesaria la utilización de un segundo proceso clasificador de idioma del contenido de la memoria intermedia (4).

Fig. 8

ES 2 478 627 T3

1.ASR y 2.ASR Etapa 1 **	1.ASR	L-ID	L-ID Etapa 1	Reacción del sistema
coincidencia	ja/yes/eved	n/a	no significativa	Información cambio de idioma y continuación *
coincidencia	nein/no/hayir	n/a	no significativa	Información se mantiene el idioma y continuación *
coincidencia	sin coincidencia	por defecto	no significativa	Información se mantiene el idioma y continuación *
coincidencia	sin coincidencia	idioma 1	idioma 1	Información cambio de idioma y continuación *
coincidencia	sin coincidencia	idioma 1	idioma 2	Transferencia al asesor
coincidencia	sin coincidencia	idioma 2	idioma 1	Transferencia al asesor
sin coincidencia	sin coincidencia	idioma 2	idioma 2	Información cambio de idioma y continuación *
sin coincidencia	sin coincidencia	idioma desconocido	no significativa	Transferencia al asesor
sin coincidencia	no entrada	n/a	no significativa	Transferencia al asesor
sin coincidencia	ja/yes/eved	n/a	no significativa	Información cambio de idioma y solicitud **
sin coincidencia	nein/no/hayir	n/a	no significativa	Información se mantiene el idioma y solicitud **
sin coincidencia	sin coincidencia	por defecto	no significativa	Información se mantiene el idioma y solicitud **
sin coincidencia	sin coincidencia	idioma 1	idioma 1	Información cambio de idioma y solicitud **
sin coincidencia	sin coincidencia	idioma 1	idioma 2	Transferencia al asesor
sin coincidencia	sin coincidencia	idioma 2	idioma 1	Transferencia al asesor
sin coincidencia	sin coincidencia	idioma 2	idioma 2	Información cambio de idioma y solicitud **
sin coincidencia	sin coincidencia	idioma desconocido	no significativa	Transferencia al asesor
sin coincidencia	no entrada	n/a	no significativa	Transferencia al asesor

* Se recoge la antigua solicitud de la memoria intermedia para continuar los diálogos, ** *sin coincidencia* normal – Etapa de escalada en diálogo monolingüe

** Según la tabla 10.2, los resultados de 1.ASR y 2 ASR siempre coinciden para alcanzar la etapa 2a (coincidencia o no coincidencia)

Fig. 9

1.ASR	L-ID	L-ID Etapa 1	Reacción del sistema
coincidencia	n/a	n/a	Continuación sin cambio de idioma *
sin coincidencia	por defecto	n/a	Transferencia al asesor
sin coincidencia	idioma 1	coincidencia	Información cambio de idioma y continuación
sin coincidencia	idioma 1	sin coincidencia	Transferencia al asesor
sin coincidencia	desconocida	n/a	Transferencia al asesor
no entrada	n/a	n/a	Transferencia al asesor
* si se observa una declaración válida, el diálogo continuará inmediatamente, existía anteriormente una tendencia clara al idioma por defecto			

Fig. 10

