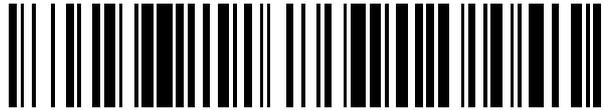


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 464 341**

51 Int. Cl.:

**H04M 3/53**

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **15.12.2009 E 09179315 (8)**

97 Fecha y número de publicación de la concesión europea: **19.02.2014 EP 2337326**

54 Título: **Procedimiento y dispositivo para resaltar objetos seleccionados en mensajes de imagen y video**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**02.06.2014**

73 Titular/es:

**DEUTSCHE TELEKOM AG (100.0%)  
Friedrich-Ebert-Allee 140  
53113 Bonn, DE**

72 Inventor/es:

**RUNGE, DIPL.-ING., FRED y  
SCHEERBARTH, THOMAS**

74 Agente/Representante:

**MORGADES MANONELLES, Juan Antonio**

**ES 2 464 341 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 y dispositivo para resaltar objetos seleccionados en mensajes de imagen y de vídeo

5 La presente invención hace referencia a un sistema de mensajes de vídeo y/o de imagen, y en particular se refiere a un procedimiento y a una cierta disposición, mediante los cuales poder transmitir a un destinatario mensajes en el marco del sistema, por ejemplo mediante un buzón de correo electrónico, comprendiendo dichos mensajes información de imagen del interlocutor. En lo que concierne a los mensajes transmitidos al destinatario, se respeta su originalidad en la medida de lo posible, y se proporcionan mejorando su contenido de información, poniendo especial relieve en el interlocutor del mensaje de imagen y/o de vídeo.

## ESTADO DE LA TÉCNICA

15 Hace largo tiempo que los sistemas de mensajería (Mobilbox, Voicemail) son conocidos en el mercado (véase por ejemplo, <http://www.tmobile.de/mobilbox>). En función de la estructura del sistema, asimismo a menudo se almacenan mensajes de voz, imagen, y/o vídeo adjuntos a mensajes de texto (adjuntos de correo electrónico junto con datos del remitente (por ejemplo, reconocimiento del remitente (por ejemplo, CLI, HLR, dirección del remitente, entre otros)) como mensajes de vídeo completos en buzones de correo electrónico de vídeo, basándose en la videotelefonía, análogamente a los buzones de correo electrónico de voz existentes desde hace largo tiempo, o a los contestadores digitales automáticos de llamadas.

20 Asimismo, como estado de la técnica conocido en general se presupone la utilización de avatares en aplicaciones animadas por ordenador, dotados de funcionalidades que posibilitan emitir textos utilizando la síntesis de voz, derivándose y visualizándose convenientemente el movimiento de los labios. Con medios para la detección facial y el establecimiento de partes del rostro que queden rellenas con elementos faciales determinados (por ejemplo, labios), es posible introducir el movimiento registrado con una cámara de los labios de un interlocutor en la zona de los labios de una foto inmóvil, que puede adoptar el rol de un avatar (véase [http://www.logitech.com/index.cfm/webcam\\_communications/webcams/devices/3056&cl=de.de](http://www.logitech.com/index.cfm/webcam_communications/webcams/devices/3056&cl=de.de); Video Effects).

25 En el documento EP1648151A1 se publica que para el procesado posterior se almacenan mensajes destinados a un sistema de traducción automática de mensajes adicionales. Sin embargo, el punto fuerte a este respecto reside en la evaluación de la información textual y fonética, contenida en los mensajes almacenados. Ello es válido asimismo para mensajes de imagen, que en caso necesario contienen textos en la imagen o en metadatos. A este respecto, no se describe la modificación automática del contenido de la imagen.

30 La visualización de las imágenes en los mensajes de imagen y en los mensajes de vídeo se lleva a cabo en dos dimensiones, puesto que el procedimiento de grabación utilizado en los terminales de telecomunicaciones se basa en el equipamiento de dicho terminal con una única cámara. Un equipamiento adicional con más de dos cámaras, en combinación con procedimientos de transmisión que requieran un mayor ancho de banda, es fácil de comprender, pero ello no es aconsejable en las unidades compactas, debido a su reducido ancho de la base y a la reducida calidad en 3D a consecuencia de ello. De este modo, la visualización tridimensional del interlocutor de un mensaje de imagen y/o de vídeo es posible únicamente con una elevada pérdida de calidad.

35 Asimismo, el aumento de visualización de avatares bidimensionales, que tal como se describe anteriormente se pueden generar a partir de fotografías, en animaciones 3D con ayuda de dos imágenes tomadas simultáneamente, únicamente puede realizarse con restricciones y con pérdida de calidad.

40 Al mismo tiempo, los procedimientos para la visualización tridimensional de imágenes y/o vídeos revelan avances cada vez más notorios (véase <http://www.elektroniknet.de/?id=3200&type=98>), de modo que asimismo en este caso surge la necesidad de la visualización tridimensional de mensajes de vídeo / imagen.

45 Asimismo, es conocida la visualización bidimensional de avatares animados generados virtualmente, que internamente en el sistema existen como modelos tridimensionales completos, y que por ejemplo reproducen textos en un sistema que comprende una pantalla y un altavoz, sin que a este respecto sea necesario generar simultáneamente una imagen derecha y una imagen izquierda (DE102004014189A1).

50 A partir del documento WO 2006/047347, se conoce un procedimiento mediante el que el usuario puede asignar individualmente un AVATAR a un mensaje. Sin embargo, ni dicho planteamiento está automatizado, ni se modifican los datos existentes de la imagen.

55 La visualización tridimensional auténtica de objetos o de imágenes se basa en proporcionar casi simultáneamente dos imágenes bidimensionales, estando prefijada una imagen para el ojo izquierdo y una segunda imagen para el ojo derecho.

60

La grabación con cámaras desplazadas lateralmente para la generación de imágenes tridimensionales es un proceso conocido en general, por lo que en este documento no se describirá. Asimismo, existen métodos destinados a la generación de imágenes tridimensionales artificiales a partir de una única imagen bidimensional y partiendo de modelos existentes, mediante cuya base se conforma un modelo tridimensional completo de objetos bidimensionales registrados / analizados manualmente, que se representa con las coordenadas tridimensionales de un mayor número de puntos en la superficie de los objetos. Algunos ejemplos de ello pueden encontrarse en los documentos EP1412917B1 y EP0991023B1.

Para la observación de la visualización tridimensional existen distintos procedimientos provistos de medios auxiliares, por ejemplo, mallas lenticulares, métodos de anaglifos, lentes polarizantes, etc. Asimismo, con la visión cruzada o la visión paralela es posible la observación mediante dos imágenes prefijadas para los ojos izquierdo y derecho sin tener que utilizar medios auxiliares.

En el documento US4925294A1 se ha intentado describir un procedimiento general para la generación de una imagen tridimensional a partir de una única imagen bidimensional, para ello separando distintos elementos frontales y del fondo de la imagen, y mediante el conocimiento de sus características tridimensionales, a partir de la imagen bidimensional generando respectivamente una imagen derecha y una izquierda. En la descripción siguiente, siempre se requieren dos imágenes bidimensionales para el proceso de generación de la imagen tridimensional; a partir de las imágenes bidimensionales se genera respectivamente una imagen bidimensional derecha y una imagen bidimensional izquierda, es decir, en el procedimiento descrito posteriormente no es preciso disponer del conocimiento descrito de las propiedades de los objetos o de los modelos de objetos individuales.

### RESUMEN DE LA INVENCION

El objetivo de la presente invención es resaltar la persona que emite el mensaje, como interlocutor virtual, buscando imágenes a partir de mensajes adicionales y de su animación.

Dicho objetivo se alcanza mediante una invención con las características de las reivindicaciones independientes.

En particular, se visualizan los interlocutores en una representación tridimensional, y en el sistema descrito a continuación basándose el interlocutor real en la emisión simultánea o casi simultánea (intervalo de unos pocos milisegundos, apenas perceptible para el observador) de dos imágenes bidimensionales del interlocutor del mensaje, prefijándose una imagen para el ojo izquierdo y una segunda imagen para el ojo derecho.

A fin de resaltar la representación de un interlocutor, se buscan mensajes de actualidad con imágenes bidimensionales procedentes de antiguos mensajes de imagen y/o de vídeo del mismo interlocutor y disponibles en un sistema de buzón de correo electrónico, para ello evaluando los datos del remitente, se extraen dichos datos de las imágenes y se adapta su tamaño a un tamaño unitario de visualización del interlocutor. A este respecto, se procesan distintas perspectivas (imágenes de la izquierda y de la derecha) de las grabaciones bidimensionales del interlocutor, adaptando las diferencias existentes cromáticas y/o de brillo, a fin de generar una imagen tridimensional. Para

a) la visualización tridimensional de un interlocutor de un mensaje actual de vídeo bidimensional, se extraen imágenes de dicho interlocutor procedentes de imágenes individuales de un mensaje de vídeo y/o adicionalmente de antiguos mensajes de vídeo / imagen del mismo interlocutor (evaluando los datos del remitente), se adapta su tamaño a un tamaño unitario de visualización de dicho interlocutor, y en la misma grabación de vídeo, en antiguas grabaciones de vídeo y/o imágenes del interlocutor, se procesan perspectivas diversas existentes de grabaciones del interlocutor, adaptando las diferencias existentes cromáticas y/o de brillo, a fin de generar una imagen tridimensional de dicho interlocutor.

b) la visualización tridimensional de un interlocutor de mensajes actuales de texto y/o de voz, se evalúan los datos del remitente del mensaje (por ejemplo, correo electrónico, Voicemail,...) y se buscan en el buzón de voz mensajes de imagen bidimensionales y/o mensajes de vídeo bidimensionales y/o imágenes tridimensionales ya elaboradas de dicho interlocutor y se extraen dichos datos de los mensajes, se adapta su tamaño a un tamaño unitario de visualización de dicho interlocutor, y se procesan perspectivas diversas de grabaciones del interlocutor, adaptando las diferencias existentes cromáticas y/o de brillo, a fin de generar una imagen tridimensional de dicho interlocutor, en caso necesario actualizada.

c) A continuación, se efectúa una modificación de los datos de imagen y/o de vídeo del mensaje recibido, resaltando la zona de la imagen relevante del remitente con los datos generados en los apartados a) y b).

A fin de poder asignar imágenes a unidades determinadas (por ejemplo, personas), se almacenan dichas unidades junto con identificadores de dirección / remitente (por ejemplo, *Home Location Register* (HLR), *Calling Line Identification* (CLI), direcciones IP, direcciones de correo electrónico, identificadores personales,...) o datos de la

dirección similares, siendo ello el caso por ejemplo en mensajes de vídeo y consecuentemente para los datos de imagen contenidos en el mismo.

5 En una forma posible de realización, es imprescindible convertir distintos identificadores / direcciones de una persona en un identificador unívoco, cuando por ejemplo puede efectuarse la entrega de un mensaje vía distintos canales de comunicación.

Para detalles adicionales, véanse las reivindicaciones.

10 Puntualizando, se trata de un sistema digital de mensajería, en el que se almacenan múltiples vinculados a personas en una memoria. A este respecto, se trata preferentemente como mínimo de un servidor fijo, que está en condiciones de acceder a dicha memoria. Naturalmente, se deben considerar asimismo los terminales móviles. Los mensajes pueden ser SMS, mensajes de correo electrónico, MMS, mensajes de imagen / vídeo, mensajes de voz, etc., que preferentemente se reciben en un servidor centralizado, a fin de, por ejemplo, prepararlos de nuevo a través de una interfaz Web, o bien para enriquecerlos, de modo que con aplicaciones Cliente o aplicaciones Cliente/Servidor puedan convertirse en bidimensionales o tridimensionales y visualizarse. Asimismo, naturalmente puede concebirse que todas las operaciones se ejecuten en un terminal móvil y se adjunte un vídeo al mensaje, que se visualice en una PDA, en un teléfono, etc. Los mensajes están provistos de direcciones del remitente (por ejemplo, el correo electrónico, el número de teléfono, etc.) y una parte del mensaje comprende asimismo datos digitales de la imagen o del vídeo. A este respecto, se trata de datos con distintos complementos, como ficheros gráficos o de vídeo adjuntos.

25 Un elemento adicional del sistema constituye una unidad para la identificación de las zonas de la imagen con rostros y/o personas en los datos digitales de imagen o de vídeo, que presenten la misma dirección del remitente. Generalmente, para ello se ejecutan procedimientos conocidos de reconocimiento facial / de la imagen, que se desarrollan en un procesador.

30 Una unidad adicional sirve para la determinación de zonas de la imagen relevantes del remitente a partir de zonas identificadas de la imagen, basándose en imágenes de referencia que están almacenadas en una memoria. A este respecto, se debe establecer la imagen del remitente, a fin de poder generar a partir de la misma un modelo bidimensional / tridimensional de un interlocutor virtual. Dicho tipo de procedimientos pueden basarse en imágenes de referencia, asignadas (por ejemplo, mediante una tabla) a una dirección del remitente. En el caso de que no exista, se puede averiguar la imagen del remitente basándose en procedimientos de reconocimiento de voz, teniendo en cuenta los datos de voz contenidos en el mensaje, en datos de voz de referencia o los datos de voz contenidos en el vídeo, que pueden compararse entre sí. A este respecto, se comparan los datos de voz comprendidos en un vídeo, por ejemplo, con muestras / modelos de voz procedentes de mensajes, entregados por el mismo remitente. El procedimiento puede ejecutarse en un procesador estándar utilizando el software correspondiente.

40 En una forma de realización alternativa, se determina la frecuencia de aparición de zonas de la imagen en mensajes con la misma dirección de remitente. En este caso, puede realizarse el proceso accediendo al buzón de correo electrónico, de modo que se establezca que la persona que aparece con más frecuencia sea el remitente.

45 A continuación, se efectúa una modificación de los datos de imagen y/o de vídeo del mensaje recibido, resaltando o realizando la animación de la zona de la imagen relevante del remitente. Dicha zona se puede poner de relieve utilizando distintos colores, brillos o tamaños. Asimismo, se puede realizar una animación en una máquina bidimensional o tridimensional, que también puede tratarse de un software provisto de por lo menos un procesador, determinándose la zona de la imagen relevante del remitente, a fin de generar en dicha zona un modelo gráfico digital de un interlocutor que reproduzca el contenido de dicho mensaje, para ello realizando la animación en la zona de su boca.

50 En una forma de realización preferida, dicho procesado para resaltar ciertos elementos se puede cancelar y volver al estado anterior, mediante datos sobre superposiciones o bien facilitando una copia de seguridad. Asimismo, en lo que concierne a zonas que hayan sido modificadas, resulta posible realizar una copia de seguridad y por lo tanto volver a copiarlas.

60 Por otra parte, existe una unidad (procesador) dotada de una interfaz para el almacenamiento de la zona de la imagen bidimensional relevante del remitente que se haya identificado, junto con la dirección de remitente y parámetros que caractericen la perspectiva de la imagen de dicho remitente. A este respecto, utilizándose procedimientos conocidos de reconocimiento de imagen, se determina la perspectiva y la postura corporal del remitente. Asimismo, se conoce una interfaz para la transmisión de la zona de la imagen bidimensional relevante del remitente, así como de zonas adicionales relevantes del remitente procedentes de antiguos mensajes almacenados en la memoria, a un módulo a fin de generar datos de la imagen tridimensionales.

5 Por otra parte, se conocen unos medios para la determinación de la perspectiva / orientación de la imagen bidimensional, a partir de proporciones corporales representadas. A fin de generar un modelo tridimensional, se requiere disponer de medios para la búsqueda de una segunda imagen bidimensional para la percepción tridimensional. A este respecto, se considera una segunda y necesaria orientación, para ello evaluando las proporciones corporales que puedan valorarse automáticamente y/o evaluando los cambios de posición respecto al fondo. Asimismo, éstas se caracterizan a su vez por unos parámetros, que describen otra perspectiva en el mismo plano horizontal.

10 Por otra parte, se emplean unos medios para la generación de por lo menos una representación tridimensional del interlocutor del mensaje, complementando la imagen bidimensional original con una segunda imagen copiada de dicha imagen, en la que se haya recortado la imagen del interlocutor y se haya reemplazado por una imagen desde otra perspectiva.

15 Diseñando el proceso de búsqueda de la segunda imagen bidimensional de modo que en la búsqueda en primer lugar se busque en el mensaje actual las imágenes (por ejemplo, en una imagen de vídeo como secuencia de imágenes) desde otra perspectiva, preferentemente empleando procedimientos de reconocimiento facial en el caso de la presencia de diversas personas.

20 En el caso de que en el mensaje actual no existan imágenes adicionales aprovechables desde otra perspectiva, la búsqueda se efectúa en la memoria, preferentemente en la memoria de archivo permanente del sistema de mensajería. Este proceso se puede efectuar, por ejemplo, como proceso periódico por lotes por la noche, de modo que se presenten múltiples datos ya antes de realizar la consulta. A este respecto, se realiza preferentemente la verificación de las proporciones de características de la imagen del remitente en las imágenes encontradas adicionalmente desde otra perspectiva, a fin de comprobar el cumplimiento de los límites prefijados. Dichos límites pueden estar determinados por la posición, los colores, la resolución o la calidad de la imagen. Otros valores límite describen un plano de grabación horizontal (ángulo de grabación en la vertical) y un plano de grabación vertical (ángulo de grabación en la horizontal).

30 Se ha descrito un sistema de buzón de correo electrónico basado en Internet, que comprende memorias de mensajes vinculados con personas, y que adicionalmente está provisto de por lo menos un módulo para el almacenamiento de datos de dirección del remitente y con ello datos de imagen relacionados de remitentes de mensajes, para la recolección de dichos datos de imagen y para la generación de datos gráficos tridimensionales del remitente a partir de datos gráficos bidimensionales, a fin de que el usuario se ponga de relieve en los datos de imagen.

35 El procedimiento para resaltar el remitente de un mensaje o zonas del fondo en mensajes de imagen o de vídeo se caracteriza, porque

- 40 - se almacenan las imágenes adjuntas del mensaje o los datos de vídeo como secuencia de imágenes, conjuntamente con los datos del remitente,
- en estas imágenes, se identifican y detectan ciertas zonas, que representan rostros y/o personas,
- se comparan dichas zonas con por lo menos una imagen de referencia del remitente del mensaje correspondiente procedente del módulo de memoria, a fin de identificar una zona de la imagen que sea relevante del remitente,
- 45 - en el caso de identificar una zona de la imagen como relevante del remitente:

- o Almacenamiento de la zona de la imagen identificada como relevante del remitente juntamente con datos de dirección de dicho remitente,
- 50 o Transmisión de la zona de la imagen relevante del remitente, así como de zonas adicionales relevantes del remitente procedentes de antiguos mensajes almacenados en la memoria, a un módulo a fin de generar datos de la imagen tridimensionales.
- o Generación o adaptación de un modelo tridimensional existente del remitente en la memoria indicada anteriormente, para ello guardando la zona de la imagen relevante remitente que se ha obtenido en la memoria indicada anteriormente del sistema de buzón de correo electrónico.
- 55 o Registro de los datos que caracterizan las distintas perspectivas / orientación del modelo del remitente en la memoria indicada anteriormente del sistema de buzón de correo electrónico.
- o Búsqueda de una segunda imagen bidimensional en relación con la imagen recibida, en el marco de una segunda orientación necesaria para la percepción tridimensional, evaluando las proporciones corporales y buscando una imagen del remitente que esté caracterizada por parámetros que describan una perspectiva adicional en el mismo plano horizontal.
- 60

- A este respecto, en primer lugar se busca en el mensaje actual (por ejemplo, Videomail) las imágenes desde otra perspectiva, en caso necesario empleando procedimientos de reconocimiento facial en el caso de la presencia de diversas personas.

- 5
  - En el caso de que en el mensaje actual no existan imágenes adicionales aprovechables desde otra perspectiva, la búsqueda se efectúa en la memoria de archivo permanente del sistema de correo electrónico.
  - A este respecto, la verificación de las proporciones de características de la imagen del remitente en las imágenes encontradas adicionalmente desde otra perspectiva, a fin de comprobar el cumplimiento de los límites prefijados, que describen el plano de grabación horizontal (ángulo de grabación en la vertical) y el plano de grabación vertical (ángulo de grabación en la horizontal).
- 10
  - Adaptación del tamaño y del color de la imagen encontrada y que se debe insertar.
  - Generación de por lo menos una representación tridimensional del interlocutor del mensaje, complementando la imagen bidimensional original con una segunda imagen copiada de dicha imagen, en la que se haya recortado la imagen del interlocutor y se haya reemplazado por una imagen desde otra perspectiva.
- 15
  - Modificación de la información de imagen del mensaje recibido y puesta de relieve del interlocutor del mensaje, para ello introduciendo por lo menos una parte del modelo tridimensional en la perspectiva obtenida en lugar de la zona de la imagen relevante del remitente de la segunda imagen.

20 Un aspecto adicional constituye un sistema de buzón de correo electrónico basado en Internet, que comprende memorias de mensajes vinculados con personas, y que adicionalmente está provisto de por lo menos un módulo para el almacenamiento de datos de dirección del remitente y con ello datos de imagen relacionados de remitentes de mensajes, para la recolección de dichos datos de imagen y para la generación de datos gráficos tridimensionales del remitente a partir de datos gráficos bidimensionales.

25 Se ha descrito un sistema de buzón de correo electrónico basado en Internet, para el almacenamiento y la obtención de mensajes de imagen o de vídeo, en el que el interlocutor del mensaje queda resaltado de modo especial, que comprende memorias de mensajes vinculados con personas, y que adicionalmente está provisto de por lo menos un módulo para el almacenamiento de datos de dirección del remitente y con ello datos de imagen relacionados de remitentes de mensajes, para la recolección de dichos datos de imagen y para la generación de datos gráficos tridimensionales del remitente a partir de datos gráficos bidimensionales.

30 Adicionalmente, evaluando los datos referidos a la ubicación del remitente, se puede crear el fondo del interlocutor en función de su localización.

35 El sistema de buzón de correo 30 representado en la figura 1, la generación tridimensional 60, el reconocimiento de voz 70 y el reconocimiento facial 80 pueden encontrarse totalmente en local o en parte en el terminal. En el caso que el sistema de buzón de correo electrónico, inclusive 90, se localice en el terminal, están disponibles en 90 en caso necesario únicamente datos del remitente, con los que el usuario del terminal tiene contacto.

40 En un sistema de buzón de correo basado en una red 30, bajo ciertas circunstancias se encuentran disponibles asimismo datos del remitente (por ejemplo, imágenes, vídeos, datos de dirección) procedentes de mensajes que se enviaron a otros destinatarios.

45 El modelo tridimensional de un interlocutor comprende múltiples imágenes bidimensionales, que se almacenan junto con ciertos datos, que describen la perspectiva/orientación de una imagen bidimensional individual.

50 La ventaja del modo de proceder descrito anteriormente reside en el hecho de que en mensajes de imagen/vídeo con datos de imagen bidimensionales (imágenes, vídeos) del remitente y del interlocutor, se resalta dicho interlocutor de modo especial, de manera que el destinatario, incluso en el caso de que el fondo sea ambiguo (por ejemplo, existencia de diversas personas en la imagen), sea capaz de identificar unívocamente el interlocutor del mensaje. Adicionalmente, es posible revalorizar ciertos mensajes, que originalmente no contengan información de imagen, complementándolos con datos generados de imagen/vídeo resaltando el interlocutor o el fondo focalizado.

55 Si se convierte un objeto identificado contenido en el mensaje bidimensional en un objeto tridimensional, se obtiene la ventaja de poder resaltar adicionalmente dicho objeto focalizado ante un fondo bidimensional o en un entorno bidimensional.

## DESCRIPCIÓN DE LAS FIGURAS

60 A continuación, se describen las figuras detalladamente.

En la figura 1 se representa una configuración esquemática del sistema de buzón de correo electrónico / mensajería con distintos buzones de correo electrónico y terminales móviles.

En las figuras 2a-3b se representa un procedimiento esquemático para la determinación de imágenes con el objetivo de su representación tridimensional.

### DESCRIPCIÓN DETALLADA

5 En el sistema descrito a continuación, una representación tridimensional auténtica del interlocutor real dentro de los datos de la imagen y/o vídeo se basa en proporcionar simultáneamente o casi simultáneamente dos imágenes bidimensionales del interlocutor del mensaje, estando prefijada una imagen para el ojo izquierdo y una segunda imagen para el ojo derecho.

10 Dicho objetivo se alcanza mediante los pasos siguientes:

15 d) para la representación tridimensional de un interlocutor de mensajes de imagen bidimensionales actuales, se buscan antiguos mensajes de imagen y/o de vídeo del mismo interlocutor y disponibles en un sistema de buzón de correo electrónico, para ello evaluando los datos del remitente, se extraen dichos datos de las imágenes y se adapta su tamaño a un tamaño unitario de visualización del interlocutor, y se procesan distintas perspectivas (imágenes de la izquierda y de la derecha) de las grabaciones bidimensionales del interlocutor, adaptando las diferencias existentes cromáticas y/o de brillo, a fin de generar una imagen tridimensional.

20 e) para la representación tridimensional de un interlocutor de mensajes actuales de vídeo bidimensional procedentes de imágenes individuales de un mensaje de vídeo y/o adicionalmente de antiguos mensajes de vídeo / imagen del mismo interlocutor (evaluando los datos del remitente), se extraen las imágenes del interlocutor, se adapta su tamaño a un tamaño unitario de visualización de dicho interlocutor, y en la misma grabación de vídeo, en antiguas grabaciones de vídeo y/o imágenes del interlocutor, se procesan perspectivas diversas existentes de grabaciones del interlocutor, adaptando las diferencias existentes cromáticas y/o de brillo, a fin de generar una imagen tridimensional de dicho interlocutor.

25 f) para la representación tridimensional de un interlocutor de mensajes actuales de texto y/o de voz, se evalúan los datos del remitente del mensaje (por ejemplo, correo electrónico, Voicemail,...) y en el sistema de buzón de correo electrónico se buscan mensajes de imagen bidimensionales y/o mensajes de vídeo bidimensionales y/o imágenes tridimensionales ya elaboradas de dicho interlocutor y se extraen dichos datos de los mensajes (en el caso de que no se haya realizado ya dicho proceso y por lo tanto no existan ni se hayan almacenado), se adapta su tamaño a un tamaño unitario de visualización de dicho interlocutor, y se procesan perspectivas diversas de grabaciones del interlocutor, adaptando las diferencias existentes cromáticas y/o de brillo, a fin de generar una imagen tridimensional de dicho interlocutor, en caso necesario actualizada.

30 g) para la representación tridimensional de un interlocutor de mensajes, sustituir el interlocutor bidimensional real de mensajes de imagen y/o vídeo por una imagen tridimensional cualquiera existente de otro interlocutor, en caso necesario preajustable por parte del receptor.

35 h) para la representación tridimensional de un interlocutor de mensajes actuales de voz y/o de texto, emplear por lo menos una imagen tridimensional cualquiera existente, en caso necesario preajustable por parte del receptor, por lo menos de otro interlocutor o de otra figura artificial (por ejemplo, un avatar).

40 A fin de poder asignar imágenes a unidades determinadas (por ejemplo, personas), se almacenan dichas unidades junto con identificadores de dirección / remitente (por ejemplo, *Home Location Register* (HLR), *Calling Line Identification* (CLI), direcciones IP, direcciones de correo electrónico, identificadores personales,...) o datos de la dirección similares, siendo ello el caso por ejemplo en mensajes de vídeo y consecuentemente para los datos de imagen contenidos en el mismo.

45 En caso necesario, resulta imprescindible mezclar diversas identificaciones/direcciones de una persona, convirtiéndolas en un identificado unívoco.

50 En las imágenes/vídeos (43a, 44a) de los mensajes, además del interlocutor puede que se aprecien personas adicionales, de modo que resulta práctico verificar el rostro del interlocutor del mensaje. Por este motivo, en la visualización de los mensajes en los que se reconozcan varios rostros, se extrae la representación de dichos rostros y se verifican en comparación con otros rostros conocidos del interlocutor, identificado en el mensaje mediante su dirección, en un proceso de reconocimiento facial (80) o en procesos paralelos. La dirección del remitente se utiliza para obtener los rostros de referencia / modelos de referencia / patrones de referencia necesarios para la verificación.

55 Asimismo, únicamente para la zona del rostro de imagen/vídeo se genera una representación tridimensional, cuya verificación en comparación con el modelo de referencia obtenido mediante la dirección, en 50, da como resultado un valor mínimo de coincidencia (certeza) con el modelo de referencia del remitente, de modo que puede establecerse si para una verificación satisfactoria es imprescindible sobrepasar por exceso o por defecto un umbral de certeza predeterminado (preferentemente alrededor del 70-80%). Es decir, la generación de una imagen tridimensional o de una secuencia de imágenes tridimensionales (60) se considera completada únicamente en el caso de que la verificación del rostro en el mensaje de imagen/vídeo sea satisfactoria (véase la figura 1). Las

imágenes sin rostros detectados o que carezcan de por lo menos un rostro verificado satisfactoriamente del remitente no se modifican en el marco del presente proceso.

5 Para una certeza adicional de la verificación del interlocutor de un mensaje de vídeo, tanto a partir de las imágenes del interlocutor (vídeo = secuencia de imágenes individuales y preferentemente señales acústicas), como de las secuencias de la voz extraídas, es posible combinar la verificación del rostro del remitente (conforme a la dirección de remitente) con la verificación de la voz del interlocutor; a este respecto, se determina adicionalmente mediante modelos de comparación implementados por reconocimiento facial, si un rostro con sus labios moviéndose se corresponde con el del remitente. Si ello es el caso, se verifica el rostro del interlocutor mediante por lo menos un modelo de referencia existente en 50 y/o 90 del interlocutor identificado con ayuda de la dirección del remitente. El resultado de la verificación se guarda en forma de una puntuación de fiabilidad.

15 Asimismo se verifica la voz registrada del interlocutor mediante por lo menos un modelo de referencia existente en 50 y/o 90 del interlocutor identificado con ayuda de la dirección del remitente.

El resultado de la verificación del interlocutor se guarda asimismo en forma de una puntuación de fiabilidad. El orden de la secuencia de ambas verificaciones carece de importancia para el procesado combinado.

20 En combinación con el reconocimiento facial o a la inversa, a continuación puede evaluarse adicionalmente la voz del interlocutor localizado, con ayuda de procedimientos conocidos de verificación del interlocutor / reconocimiento de voz.

Al combinar los resultados de la verificación del interlocutor, del reconocimiento facial y en caso necesario de la verificación de otros medios, bajo ciertas circunstancias pueden surgir resultados contradictorios con valores de fiabilidad distintos. Así, con la condición de ejemplo de disponer de una zona de resultados unitaria comprendida entre 0 y 100, para la verificación del interlocutor (SV) y reconocimiento facial (GE) y la suposición de que los valores comprendidos entre 0 y 80 significan el rechazo y valores comprendidos entre 81 y 100 la aceptación de las señales, resulta totalmente posible que por ejemplo el valor de SV sea de 79, mientras que el valor de GE sea de 81. A fin de poder generar una señal de decisión, la decisión todavía resulta más ambigua, cuando además de la evaluación de más de dos señales de medios, en los valores de un resultado es preciso partir de distintos umbrales de fiabilidad para los resultados individuales.

Dicho problema puede resolverse, por ejemplo, con una normalización de los valores individuales y una suma ponderada de los valores normalizados para generar los valores de decisión que debe compararse, por ejemplo:

$$E_w = W_{sv} \frac{S_{sv} - S_{sv_{\min}}}{S_{sv_{\max}} - S_{sv_{\min}}} + W_{ge} \frac{S_{ge} - S_{ge_{\min}}}{S_{ge_{\max}} - S_{ge_{\min}}} + W_{v3} \frac{S_{v3} - S_{v3_{\min}}}{S_{v3_{\max}} - S_{v3_{\min}}} + \dots + W_{vn} \frac{S_{vn} - S_{vn_{\min}}}{S_{vn_{\max}} - S_{vn_{\min}}}$$

- donde
- E<sub>w</sub> - valor de decisión normalizado
  - S<sub>sv</sub> - puntuación de fiabilidad de la verificación del interlocutor
  - S<sub>svmin</sub> - valor mínimo de la puntuación de fiabilidad de la verificación del interlocutor
  - S<sub>svmax</sub> - valor máximo de la puntuación de fiabilidad de la verificación del interlocutor
  - W<sub>sv</sub> - factor de ponderación para la puntuación de fiabilidad normalizada de la verificación del interlocutor
  - S<sub>ge</sub> - puntuación de fiabilidad del reconocimiento facial
  - S<sub>gemin</sub> - valor mínimo de la puntuación de fiabilidad de reconocimiento facial
  - S<sub>gemax</sub> - valor máximo de la puntuación de fiabilidad de reconocimiento facial
  - W<sub>v3</sub> - factor de ponderación para la puntuación de fiabilidad normalizada de reconocimiento facial
  - W<sub>v3</sub> - factor de ponderación para la puntuación de fiabilidad normalizada de la posible tercera Verificación (por ejemplo, reconocimiento de los gestos)
  - S<sub>v3</sub> - puntuación de fiabilidad del reconocimiento facial
  - S<sub>v3min</sub> - valor mínimo de la puntuación de fiabilidad de la tercera Verificación (por ejemplo, reconocimiento de los gestos)
  - S<sub>v3max</sub> - valor máximo de la puntuación de fiabilidad de la tercera Verificación (por ejemplo, reconocimiento de los gestos)
  - W<sub>v3</sub> - factor de ponderación para la puntuación de fiabilidad normalizada de la tercera Verificación (por ejemplo, reconocimiento de los gestos)
  - S<sub>vn</sub> - puntuación de fiabilidad del enésimo medio verificado
  - S<sub>vnmin</sub> - valor mínimo de la puntuación de fiabilidad del enésimo medio verificado
  - S<sub>vnmax</sub> - valor máximo de la puntuación de fiabilidad del enésimo medio verificado
  - W<sub>vn</sub> - factor de ponderación para la puntuación de fiabilidad normalizada del enésimo medio verificado

## ES 2 464 341 T3

En el caso más sencillo, en este caso para valores umbrales individuales normalizados y  $n=3$ ,  $W_{sv}=W_{ge}=W_{v3}=1/n=1/3$

$$S_n = \frac{S_{vn} - S_{n_{\min}}}{S_{n_{\max}} - S_{n_{\min}}}$$

5 En el caso de umbrales individuales  $S_{gn}$  de valores distintos, las puntuaciones de fiabilidad normalizadas se adaptan a un mismo umbral normalizado común (GNG), de modo que de nuevo es posible trabajar con los mismos factores de ponderación.

10 En este tipo de normalización,  $S_{n_{\min}}=0$  y  $S_{n_{\max}}=1$  puede servir para la resolución del polinomio de segundo orden

$$GNG = a_n S_{gn}^2 + (1 - a_n) S_{gn}$$

15 para el cálculo del valor  $a_n$  mediante los valores conocidos de GNG y  $S_{gn}$  (umbrales de puntuación normalizada de la  $n$ -ésima verificación).

Para GNG puede tomarse por ejemplo un valor medio de los umbrales individuales normalizados de valor distinto.

20 De este modo, el valor de decisión se calcula a continuación:

$$E_w = W_{v1} * (a_1 S_{g1}^2 + (1 - a_1) S_{g1}) + W_{v2} * (a_2 S_{g2}^2 + (1 - a_2) S_{g2}) + W_{v3} * (a_3 S_{g3}^2 + (1 - a_3) S_{g3}) + \dots + W_{vn} * (a_n S_{gn}^2 + (1 - a_n) S_{gn})$$

25 Los factores de ponderación pueden ajustarse de nuevo al valor  $1/n$ , en el caso de la misma importancia de los distintos procedimientos de verificación.

En el caso de que  $E_w$  sobrepase por exceso o por defecto el umbral predeterminado para GNG, se transmiten o se bloquean las órdenes de control para resaltar la zona (por ejemplo, mediante la generación tridimensional de una imagen del interlocutor) en 60. Puede suceder una acción parecida si únicamente se deben transmitir señales de activación o desactivación.

30 En la memoria de archivo permanente 90 para vídeos/imágenes de remitente, a partir del mensaje recibido se almacenan únicamente las imágenes del remitente asignadas a las direcciones correspondientes que se hayan verificado satisfactoriamente, de modo que en un caso necesario posterior se utilicen para la generación de la imagen tridimensional (véase a, b, c) únicamente las imágenes/vídeos, cuya verificación previa en la comparación con el modelo de referencia de 50 o 90, obtenido mediante la dirección, proporcione un valor mínimo de coincidencia (certeza) con dicho modelo de referencia.

40 1) En una grabación bidimensional de un mensaje de vídeo de un interlocutor individual, en el procedimiento anterior se parte de la base que para cada imagen contenida en la secuencia del vídeo, ya exista una imagen para el ojo derecho (imagen derecha) o para el ojo izquierdo (imagen izquierda), necesaria para una visualización tridimensional

En este caso, un vídeo se considera como una secuencia de imágenes individuales.

45 A fin de obtener una representación tridimensional completa, es preciso generar asimismo una segunda imagen de una persona o de su rostro para el otro ojo. Para la descripción de ejemplo simplificada del procedimiento de generación tridimensional, a continuación se partirá de la base de que la imagen existente esté calificada como imagen derecha existente. Asimismo, todavía es imprescindible generar la imagen izquierda para cada imagen de la secuencia de vídeo.

50 Sin embargo, antes de la preparación de la imagen izquierda necesaria para la generación tridimensional, resulta práctico verificar si el interlocutor visualizado en por lo menos una imagen de la secuencia de vídeo o por lo menos uno de los interlocutores visualizados, coincida con el de una imagen identificable mediante la dirección de remitente y en caso necesario obtenida anteriormente y guardada en 50 o 90.

55 En el caso de una verificación satisfactoria del interlocutor individual, se guardan el contorno y las coordenadas de la posición de la imagen correspondiente al remitente, y en una primera fase, la imagen derecha se copia totalmente en la imagen izquierda, aunque ello todavía no provoca ninguna percepción tridimensional. En los pasos siguientes, se recorta la imagen del interlocutor a partir de la nueva imagen izquierda conforme al contorno y a las coordenadas de la posición, y se sustituye por una imagen de tamaño adaptado del mismo interlocutor del mensaje desde otra

perspectiva situada más a la izquierda. A este respecto, se presupone el conocimiento de los procedimientos para la determinación automática del contorno, el recorte, el aumento, la reducción y la inserción de objetos en imágenes.

5 A continuación, para cada imagen individual de la secuencia de vídeo, se obtiene la imagen (recorte) del mismo interlocutor del mensaje desde una perspectiva situada más a la izquierda, a partir de otras imágenes de la secuencia de vídeo conforme a la figura 2b o la figura 3b, a partir de las que conforme al dibujo mediante las distancias reducidas B y/o B' y las distancias aumentadas A y/o A', se derive que el ángulo de grabación se encuentra más a la izquierda de la posición de grabación de la imagen derecha original. A este respecto, conforme al tamaño del rostro visualizado en la imagen, se puede ajustar un posible valor máximo y/o un valor mínimo para la relación entre B, B', A, A' de la figura 2a y correspondientemente entre B, B', A, A' de la figura 2b, o entre B, A de la figura 3a y correspondientemente entre B,A de la figura 3b, de modo que la base estereoscópica tridimensional virtual no se diferencie esencialmente de la distancia promedio entre los ojos.

15 Si para una imagen derecha individual de la secuencia de vídeo, no se encuentra ninguna imagen izquierda adicional, es decir ninguna imagen adicional procedente de una posición que se encuentre más a la izquierda, en este caso extremo se puede añadir a la imagen izquierda existente (la imagen de la derecha se ha copiado en la imagen izquierda) una imagen de la derecha adecuada procedente de otra imagen de la secuencia de vídeo, teniendo en cuenta las relaciones entre B, B', A, A' de la figura 2a y correspondientemente entre B, B', A, A' de la figura 2b, o entre B, A de la figura 3a y correspondientemente entre B,A de la figura 3b. En este caso, para la selección es preciso asimismo observar los valores mínimo y/o máximo mencionados anteriormente.

En el caso de que la base estereoscópica virtual establecida por los valores mínimo y/o máximo no se encuentre en el interior de unos valores límite determinados, la percepción tridimensional podría "vibrar".

25 Las imágenes de la izquierda o de la derecha que faltan y que deben añadirse, se pueden asimismo obtener a partir de imágenes / vídeos guardados y verificados en 90 procedentes de mensajes del historial del mismo remitente.

30 2) si en el mensaje recibido existe únicamente una imagen individual (por ejemplo, valorada como imagen de la derecha) del posible interlocutor (fotografía), conforme a los procedimientos descritos anteriormente se verifica dicha imagen, y en caso necesario sus resultados se combinan tal como se ha descrito anteriormente. Tras la verificación satisfactoria, asimismo en este caso se almacenan el contorno y las coordenadas de la posición de la imagen correspondiente al remitente, y en una primera fase se copia completamente la imagen derecha en la imagen izquierda, que todavía debe generarse para el formato tridimensional, y se recorta la imagen del interlocutor a partir de la nueva imagen izquierda conforme al contorno y a las coordenadas de la posición almacenadas, y se sustituye por una imagen de tamaño adaptado del mismo interlocutor del mensaje desde otra perspectiva situada más a la izquierda.

40 A continuación, se obtiene la imagen (recorte) del mismo interlocutor del mensaje desde una perspectiva situada más a la izquierda, a partir de imágenes y/o vídeos, que provienen de mensajes anteriores y que se han guardado en 90, juntamente con la referencia a la dirección del remitente y datos, que representan las proporciones de la derecha/izquierda de la imagen del remitente (por ejemplo, cociente entre A' y B'). Para el control de las proporciones de la derecha/izquierda, se puede seleccionar la relación de las distancias de distintas zonas prominentes del rostro, por ejemplo, los ojos, la boca, las orejas, la nariz, la mejilla, etc., se pueden marcar y tras el cálculo de las características determinantes de la proporción, se archivan en 90 haciendo referencia a la imagen correspondiente.

50 En la selección de la imagen que debe añadirse, se tiene en cuenta que conforme al tamaño del rostro visualizado en la imagen, no se sobrepase por exceso o por defecto un posible valor máximo y/o un valor mínimo para la relación entre B, B', A, A' de la figura 2a y correspondientemente entre B, B', A, A' de la figura 2b, o entre B, A de la figura 3a y correspondientemente entre B, A de la figura 3b, de modo que asimismo en este caso la base estereoscópica tridimensional virtual no se diferencie esencialmente de la distancia promedio entre los ojos.

55 En la sistemática descrita, naturalmente puede copiarse asimismo la imagen derecha, como imagen derecha existente, en la imagen izquierda, de modo que a continuación se modifica dicha última imagen como anteriormente se ha descrito.

60 Para las imágenes tridimensionales generadas del interlocutor, procedentes de por lo menos dos imágenes bidimensionales con distintas perspectivas, en ambos casos (vídeo y foto) su tamaño se adapta de nuevo mediante procedimientos de normalización al tamaño del interlocutor en el mensaje bidimensional, y conforme a la orientación del interlocutor en el mensaje original, se insertan en el mensaje original ahora modificado en su visualización de imagen digital o de vídeo, (43b, 44b), entendiéndose por el concepto de inserción esencialmente añadir por lo menos una segunda imagen desde otra perspectiva adicional

La orientación del interlocutor en el mensaje bidimensional se obtiene mediante la evaluación de la posición de características esenciales del rostro y/o del cuerpo, como la posición de la parte de la frente donde comienza la cabellera, los ojos, la nariz, los labios y/o las orejas en relación con la distancia al fin de la cabeza, presuponiéndose una forma del rostro prácticamente simétrica en el caso de una orientación directa del rostro a por lo menos una cámara d grabación.

Puesto que la orientación del interlocutor en la visualización tridimensional se adapta dinámicamente a la orientación original en la visualización bidimensional, se puede incorporar directamente asimismo la representación del movimiento de los labios para las imágenes individuales de un mensaje de vídeo, que entonces presentan la misma orientación, aceptando una pérdida reducida de la calidad.

En la inserción automática, asistida por ordenador, de una representación tridimensional gráfica de una persona en un recorte de una imagen, ocupado originalmente por una representación bidimensional de la misma persona, existe el reducido riesgo de que en los márgenes de la representación gráfica insertada de la persona, aparezcan ciertas zonas que no queden cubiertas ni por la nueva representación gráfica de la persona, ni por el fondo original. En este caso, se puede efectuar un pequeño aumento de la inserción y en caso necesario de la primera imagen original en el recorte calculado, en un valor predeterminado, o bien efectuar un aumento dinámico en función del área a cubrir que se ha calculado, hasta que la zona quede totalmente cubierta. Y a la inversa es posible efectuar asimismo una compresión reducida del fondo, a fin de posibilitar la inserción de la imagen, tal como se describe por ejemplo en el documento DE69628662T2 .

Los mensajes puramente de texto (41) originales no disponen en general de grabaciones de imagen o de vídeo a sustituir o verificar. A fin de dotarlos adicionalmente de una imagen tridimensional o de una secuencia de imágenes tridimensionales (vídeo) del remitente, que se reproducen en la visualización/reproducción del mensaje, en este caso se puede utilizar directamente la dirección del remitente, para obtener las imágenes verificadas necesarias en 90 para la generación de una imagen tridimensional complementaria (41b). En la matriz de datos 90, se almacenan todas las imágenes ya recibidas y verificadas de un remitente vinculadas a su identificación / información de dirección, y si resulta necesario, puede accederse a las mismas en el caso de la generación de imágenes tridimensionales de dicho interlocutor o con el objetivo de la verificación de nuevos datos gráficos de un interlocutor mediante la identificación de la dirección asignada al mensaje original.

Tampoco los mensajes de voz (42) originales no disponen en general de grabaciones de imagen o de vídeo a sustituir o verificar. Sin embargo, en este caso puede emplearse para la verificación (70) del interlocutor, es decir para comprobar si el modelo de voz de referencia guardado en 90 y/o en 50 y asignado en caso necesario a la dirección del remitente/identificación del mensaje actual, se corresponde con la voz del interlocutor del mensaje de voz actual. En caso afirmativo, se utiliza de modo seguro la información de dirección asignada al interlocutor, a fin de acceder a imágenes existentes en caso necesario en 90 para la generación de una imagen tridimensional complementaria (60). Entonces, dicha imagen tridimensional generada se utiliza para generar en la imagen un avatar comparable del remitente, que reproduzca el mensaje de voz conforme a procedimientos conocidos de síntesis de voz (modelo fuente-filtro, síntesis de formantes, modelo acústico, síntesis articulatoria, método Overlap-Add, etc.) mediante la evaluación de las características fonéticas del mensaje de voz y adaptando el movimiento de la boca.

En la figura 1, se representa especialmente el proceso de generación de imágenes tridimensionales a partir de imágenes / secuencias de imágenes bidimensionales de un remitente.

De modo complementario, es posible utilizar fonemas / secuencias de fonemas/ sílabas / palabras / frases del interlocutor de mensajes de voz y/o de vídeo almacenados en 90, a fin de utilizar el procedimiento de síntesis de voz basado en *Unit Selection* en textos guardados en 41, y conforme a la identificación de dirección obtenida (en este caso, utilizada asimismo como reconocimiento de voz).

Habiendo obtenido la dirección del remitente del mensaje de texto, a continuación se buscan en 90 fragmentos de voz correspondientes del remitente verificado (véase 70) análogamente por su voz, a fin de generar un mensaje de voz a partir del texto entregado. Un avatar tridimensional generado asimismo con ayuda de la dirección del remitente a partir de imágenes bidimensionales (véase la descripción anterior), puede ahora reproducir acústicamente el mensaje de voz vocalizando el texto y adaptando el movimiento de la boca. De este modo, es posible generar un mensaje de voz y de vídeo tridimensional a partir de un mensaje de texto sencillo.

Así, del mismo modo en que a partir de fragmentos de voz recopilados del remitente puede generarse una salida vocal con su voz de mensajes de texto, en principio el receptor del mensaje puede ajustar asimismo otra voz para el mensaje (por ejemplo, asignando a otro valor la dirección del remitente obtenida originalmente, que se utiliza en principio como referencia a la voz deseada), y adicionalmente pueden expresarse variaciones emocionales de la salida vocal con ayuda de los identificadores insertados el texto.

- Si así lo desea, el receptor del mensaje o el usuario del sistema de buzón de correo electrónico puede, por ejemplo, modificar la imagen y o la voz del remitente. Para ello, la identificación / dirección del remitente obtenida a partir del mensaje recibido se sustituye por una dirección / identificación (100) ajustable por parte del receptor del mensaje, y para la generación tridimensional de las imágenes individuales o para la generación de la voz a reproducir, conforme a la nueva dirección / identificación predeterminada, se buscan en 90 fragmentos de imagen o de voz a fin de generar las imágenes tridimensionales o las salidas vocales de un nuevo interlocutor.
- Además de la selección directa de la imagen y/o de la voz durante la reproducción, en los parámetros de selección pueden incluirse las modificaciones del perfil de usuario (100) gestionado por el receptor.
- Aquí se encuentra asimismo la localización de almacenamiento para datos que definen formas adicionales de la generación tridimensional.
- En el caso más sencillo, tal como se ha descrito anteriormente, se genera un vídeo tridimensional del remitente a partir de un mensaje de vídeo bidimensional de dicho remitente, evaluando sus datos (por ejemplo, mediante la dirección entregada), o se genera una única imagen tridimensional del remitente, que comprende dos imágenes bidimensionales, a partir de una única imagen bidimensional del remitente adjunta al mensaje, evaluando sus datos (por ejemplo, mediante la dirección entregada).
- Una configuración distinta del perfil de usuario puede implicar que en lugar de una imagen individual bidimensional vinculada con un mensaje de voz recibido, o de un mensaje de voz sin imagen, se deba reproducir un vídeo tridimensional preferente del remitente, ya generado y almacenado en 90, de nuevo con el mensaje de voz recibido actualmente. Dado que el movimiento de los labios del remitente grabado deja de estar sincronizado con el texto actual del mensaje que se vocaliza, es imprescindible efectuar una adaptación de dicho movimiento. A partir de la generación y aplicación de avatares, se conocen diversas asignaciones de fonemas a visemas, de modo que para la salida vocal de texto escrito, unos caracteres virtuales y artificiales comunican dicho texto modelándose convenientemente el movimiento de los labios. Asimismo, se conocen procedimientos de reconocimiento de la voz, en los que a partir de un cierto flujo oral, pueden extraerse no únicamente palabras, sino asimismo series de fonemas (series de sonidos). Para todos los idiomas / todos los dialectos, existe una cantidad limitada de fonemas, y por lo tanto una cantidad limitada de transiciones entre fonemas. En el idioma alemán, en función del dialecto se parte de unos 40 fonemas. En el caso existente, a partir del mensaje de vídeo archivado (histórico) seleccionado, se extrae la serie de fonemas de un remitente o de otro interlocutor con ayuda de un sistema de reconocimiento de voz, y se almacena sincronizadamente con la mímica correspondiente como combinación de fonemas y visemas. Lo mismo se efectúa para el almacenamiento de las transiciones entre fonemas y de los correspondientes visemas. Como visemas de un remitente o de otro, en caso necesario, asimismo interlocutor artificial, se pueden considerar por ejemplo secuencias de imágenes completas, inclusive el fondo del interlocutor, o bien únicamente el rostro tridimensional recortado o la boca recortada. Las colecciones de visemas asignadas distintos interlocutores quedan vinculadas asimismo a la dirección de dichos interlocutores.
- Entonces, para el mensaje de voz que debe convertirse en un mensaje de vídeo, se analiza asimismo su serie de fonemas y las transiciones entre fonemas. En el caso de que la serie de fonemas y las transiciones estén contenidas por lo menos una vez en el mensaje de vídeo archivado, conforme a la serie de fonemas y las transiciones del mensaje de voz actual, la secuencia de imágenes / visemas correspondientes se añade sincronizadamente con la secuencia de fonemas del mensaje actual.
- En el caso de que en un mensaje de vídeo preferente archivado no existan todas las series de fonemas y/o transiciones entre fonemas del mensaje actual, evaluando la serie de fonemas y las transiciones, se buscan mensajes archivados adicionales (por ejemplo, bidimensionales, tridimensionales, audio, etc.) del mismo remitente, que contengan adicionalmente la serie de fonemas y transiciones necesaria. Debido a la existencia de fondos de imágenes distintos y de distintas tonalidades cromáticas, conforme a los ejemplos anteriores, a partir de la serie de visemas correspondientes, se recorta únicamente por ejemplo la visualización de rostro o de la cabeza del remitente, adaptándose el tamaño, color, y/o los márgenes, y se inserta en la secuencia de imágenes del mensaje de vídeo preferente.
- Si se insertan imágenes tridimensionales de otras personas distintas a las del interlocutor en la zona gráfica prefijada delante del fondo original, tal como se ha descrito, o bien se puede efectuar un aumento de la inserción en un valor predeterminado, o bien efectuar un aumento dinámico en función del área a cubrir que se ha calculado, hasta que la zona quede totalmente cubierta.
- Del mismo modo que en este caso se ha descrito para el interlocutor de un mensaje, utilizando procedimientos de reconocimiento de imagen, es posible dotar objetos estáticos identificados en el mensaje de imagen y/o de vídeo de identificadores, y modificarlos en el mensaje para que sean imágenes tridimensionales (por ejemplo, edificios como objetos del fondo).

Un complemento adicional constituye la inserción de otro fondo en función del paradero predeterminado del remitente del mensaje. Si, por ejemplo al dejar y/o enviar un mensaje de imagen o de vídeo a través de un terminal móvil, evaluando los datos de itinerancia se constata que el remitente del mensaje se encuentra en un país, provincia o lugar determinados, se puede insertar una nueva imagen de fondo en función de dicha ubicación. Para ello, por ejemplo en el sistema de buzón de correo electrónico de texto, voz y/o vídeo, al dejar un mensaje de dicho tipo se puede guardar adicionalmente la información que determina la ubicación.

Dicha información, vinculada de este modo con el mensaje que se ha dejado, y mediante la que se determina la ubicación,

1. al dejarse el mensaje por parte de un terminal móvil o de un terminal, mediante el que es posible la detección de los datos siguientes, puede comprender

a) el ID de la célula de origen Y/O

b) los ID de las células del entorno con el correspondiente tiempo de propagación de la señal a los mástiles de radio cercanos de las estaciones base. Y/O

c) la posición o el rango de posiciones calculado a partir de dichas señales.

El tiempo de propagación de las señales hacia los mástiles de radio del entorno se utiliza asimismo en los denominados procedimientos A-GPS, resultando útil para la localización mediante procedimientos de GPS.

2. al dejarse el mensaje por parte de un terminal que no reciba señales de radio para la determinación de la ubicación, puede comprender una identificación del país y/o prefijo del lugar, a partir del cual en caso necesario pueda derivarse la ubicación del terminal.

3. al dejarse el mensaje por parte de un sistema de voz sobre IP, puede comprender por lo menos una dirección IP y el punto temporal en que se dejó un mensaje y/o direcciones de puertos DSLAM, a partir de las cuales pueda asimismo derivarse la ubicación. (<http://www.voip-info.de/wissen/ Artikel Allgemein 2011.php?page=3>).

Un identificador que determine la ubicación y/o un identificador de la dirección del remitente sirven adicionalmente para añadir al mensaje un fondo dependiente de la localización o del remitente, que refuerce la integración del interlocutor en un entorno tridimensional, para ello intercalándose diversos ruidos de fondo multicanal grabados del mensaje modificado dependientes de la ubicación y/o del remitente. Así, es posible intercalar en la imagen un fondo guardado previamente, para ello empleando procedimientos conocidos para reemplazar dicho fondo.

En el caso de que la ubicación de la grabación resulte de especial importancia para el remitente del mensaje, se puede adjuntar al mensaje enviado una señal complementaria, que solicite al sistema de buzón de correo electrónico que la generación de la imagen tridimensional se realice únicamente a partir del fondo, análogamente al procedimiento descrito anteriormente. A este respecto, se dobla asimismo el original bidimensional y en la copia se recorta el fondo. A continuación, dicho fondo recortado se reemplaza por el mismo fondo, inclusive los objetos secundarios, grabado desde otra perspectiva. Es asimismo posible la inclusión de fondos tridimensionales previamente generados y dependientes de la ubicación. En este caso, el interlocutor aparecería en formato bidimensional.

Asimismo, el receptor puede iniciar la solicitud para la generación del fondo tridimensional mediante una señal directa o bien ajustando el perfil del receptor (100).

**REIVINDICACIONES**

- 5 1. Procedimiento para resaltar objetos seleccionados en datos de imagen y/o de vídeo almacenados en mensajes, estando guardados dichos mensajes en un sistema digital de mensajería, en el que se almacenan múltiples mensajes vinculados con personas en una memoria, estando provistos dichos mensajes de la dirección del remitente,
- 10 - identificación (80) de zonas de la imagen con rostros y/o personas en los datos digitales de la imagen o en los datos de vídeo en los mensajes del sistema digital de mensajería, que presenten la misma dirección del remitente;
- 15 - determinación de zonas de la imagen (90) relevantes del remitente, que representan la persona del remitente, a partir de zonas identificadas de la imagen, basándose en imágenes de referencia que están almacenadas en una memoria, y/o
- 20 basándose en procedimientos de reconocimiento de voz, procedimientos de verificación del interlocutor o procedimientos de identificación del interlocutor (70), teniendo en cuenta los datos de voz y los datos de imagen del mensaje, los datos de voz de referencia y/o los datos de voz en los datos de vídeo, que se comparan entre sí, y/o
- 25 determinando la frecuencia de aparición de las zonas de la imagen en mensajes con la misma dirección de remitente;
- Modificación de los datos de imagen y/o de vídeo del mensaje recibido, resaltando la zona de la imagen relevante del remitente.
3. Procedimiento según la reivindicación anterior, de modo que el receptor pueda cancelar la modificación y volver al estado anterior, para ello preferentemente proporcionando un diálogo de selección.
3. Procedimiento según una o varias de las reivindicaciones anteriores, empleándose las zonas de la imagen identificadas a fin de generar un modelo tridimensional
4. Procedimiento según la reivindicación anterior, caracterizado porque en el caso de identificar una zona de la imagen como relevante del remitente 40, se realizan los pasos siguientes:
- 35 - Almacenamiento de la zona de la imagen bidimensional relevante del remitente que se haya identificado, junto con la dirección de remitente y parámetros que caractericen la perspectiva de la imagen de dicho remitente,
- 40 - Transmisión de la zona de la imagen bidimensional relevante del remitente, así como de zonas adicionales relevantes del remitente procedentes de antiguos mensajes almacenados en la memoria, a un módulo a fin de generar datos de la imagen tridimensionales,
- 45 - Determinación de la perspectiva / orientación de la imagen bidimensional, a partir de proporciones corporales representadas,
- Búsqueda de una segunda imagen bidimensional para la percepción tridimensional; con una segunda orientación necesaria y evaluando las proporciones corporales, estando dicha orientación caracterizada por parámetros que describen una perspectiva adicional en el mismo plano horizontal.
- 50 - Adaptación del tamaño y del color de la imagen encontrada y que se debe insertar,
- Generación de por lo menos una representación tridimensional del interlocutor del mensaje, complementando la imagen bidimensional original con una segunda imagen copiada de dicha imagen, en la que se haya recortado la imagen del interlocutor y se haya reemplazado por una imagen desde otra perspectiva.
5. Procedimiento según la reivindicación anterior, de modo que en la búsqueda de una segunda imagen bidimensional, en primer lugar se busque en el mensaje actual las imágenes desde otra perspectiva, preferentemente empleando procedimientos de reconocimiento facial en el caso de la presencia de diversas personas, y en el caso de que en el mensaje actual no existan imágenes adicionales aprovechables desde otra perspectiva, la búsqueda se efectúa en la memoria, y a este respecto se realiza preferentemente una verificación de las proporciones de características de la imagen del remitente y/o si las imágenes encontradas adicionalmente desde otra perspectiva cumplen con los límites prefijados, en particular, unos valores límite que describan un plano de grabación horizontal, especialmente un ángulo de grabación en la vertical, y un plano de grabación vertical, especialmente un ángulo de grabación en la horizontal.
6. Procedimiento según una o varias de las reivindicaciones anteriores, comparándose los datos de voz del mensaje y los datos de voz del mensaje de vídeo, y/o comparándose entre sí múltiples datos de voz de los mensajes de

vídeo, a fin de encontrar la zona de la imagen relevante o las imágenes relevantes del mensaje de vídeo, si los mensajes de voz están superpuestos.

5 7. Procedimiento según una o varias de las reivindicaciones anteriores, de modo que la generación tridimensional, el reconocimiento de la voz y el reconocimiento facial se ejecutan totalmente o en parte en un terminal o en un servidor accesible a través de una red, quedando almacenados los datos que caracterizan las distintas perspectivas de las imágenes de remitente de los mensajes en la memoria de un servidor o localmente en una memoria de un terminal móvil.

10 8. Procedimiento según una o varias de las reivindicaciones anteriores, de modo que el modelo tridimensional comprende múltiples imágenes bidimensionales, que se almacenan junto con ciertos datos, que describen la perspectiva/orientación de una imagen bidimensional individual.

15 9. Procedimiento según una o varias de las reivindicaciones anteriores, de modo que se selecciona la inserción de otro fondo en función del paradero predeterminado del remitente del mensaje.

20 10. Sistema digital de mensajería para resaltar objetos seleccionados en datos de imagen y/o de vídeo, en el que se almacenan múltiples mensajes vinculados con personas en una memoria, estando provistos dichos mensajes de la dirección del remitente, de modo que una parte de los mensajes comprende asimismo datos digitales de imagen o datos de vídeo, con los componentes:

- unidad para la identificación (80) de las zonas de la imagen con rostros y/o personas en los datos digitales de imagen o de vídeo de los mensajes, que presentan la misma dirección del remitente.
- unidad para la determinación de zonas de la imagen relevantes del remitente (90), que representan la persona del remitente, a partir de zonas identificadas de la imagen, basándose en imágenes de referencia que están almacenadas en una memoria, y/o basándose en procedimientos de reconocimiento de voz, teniendo en cuenta los datos de voz del mensaje, en datos de voz de referencia y/o los datos de voz en los datos de vídeo, que se comparan entre sí, y/o determinando la frecuencia de aparición de las zonas de la imagen en mensajes con la misma dirección de remitente;
- unidad para la modificación de los datos de imagen y/o de vídeo del mensaje recibido, resaltando la zona de la imagen relevante del remitente.

35 11. Sistema digital de mensajería según las reivindicaciones del sistema anteriores, de modo que existen unos medios para que el receptor pueda cancelar la modificación y volver al estado anterior, para ello preferentemente proporcionando un diálogo de selección.

40 12. Sistema digital de mensajería según cualquiera de las reivindicaciones del sistema anteriores, caracterizado por

- una interfaz para el almacenamiento de la zona de la imagen bidimensional relevante del remitente que se haya identificado, junto con la dirección de remitente y parámetros que caractericen la perspectiva de la imagen de dicho remitente,
- una interfaz para la transmisión de la zona de la imagen bidimensional relevante del remitente, así como de zonas adicionales relevantes del remitente procedentes de antiguos mensajes almacenados en la memoria, a un módulo a fin de generar datos de la imagen tridimensionales,
- unos medios para la determinación de la perspectiva / orientación de la imagen bidimensional, a partir de proporciones corporales representadas,
- unos medios para la búsqueda de una segunda imagen bidimensional para la percepción tridimensional; con una segunda orientación necesaria y evaluando las proporciones corporales, estando dicha orientación caracterizada por parámetros que describen una perspectiva adicional en el mismo plano horizontal,
- unos medios para la generación de por lo menos una representación tridimensional del interlocutor del mensaje, complementando la imagen bidimensional original con una segunda imagen copiada de dicha imagen, en la que se haya recortado la imagen del interlocutor y se haya reemplazado por una imagen desde otra perspectiva.

60 13. Sistema digital de mensajería según las reivindicaciones del sistema anteriores, estando diseñado el sistema de búsqueda, de modo que en la búsqueda de una segunda imagen bidimensional, en primer lugar se busque en el mensaje actual las imágenes desde otra perspectiva, preferentemente empleando procedimientos de reconocimiento facial en el caso de la presencia de diversas personas, y en el caso de que en el mensaje actual no existan imágenes adicionales aprovechables desde otra perspectiva, la búsqueda se efectúa en la memoria de archivo permanente del sistema de correo electrónico, y a este respecto se realiza preferentemente una verificación de las

proporciones de características de la imagen del remitente y/o si las imágenes encontradas adicionalmente desde otra perspectiva cumplen con los límites prefijados, en particular, unos valores límite que describan un plano de grabación horizontal, especialmente un ángulo de grabación en la vertical, y un plano de grabación vertical, especialmente un ángulo de grabación en la horizontal.

5  
14. Sistema digital de mensajería según una o varias de las reivindicaciones del sistema anteriores, que comprende unos medios para la comparación de los datos de voz del mensaje y los datos de voz del mensaje de vídeo, y/o la comparación entre sí de múltiples datos de voz de los mensajes de vídeo, a fin de encontrar la zona de la imagen relevante o las imágenes relevantes del mensaje de vídeo, si los mensajes de voz están superpuestos.

10  
15. Sistema digital de mensajería según una o varias de las reivindicaciones del sistema anteriores, disponiéndose los medios para la generación tridimensional, el reconocimiento de la voz y el reconocimiento facial totalmente o en parte en un terminal o en un servidor accesible a través de una red, quedando almacenados los datos que caracterizan las distintas perspectivas de las imágenes de remitente de los mensajes en la memoria de un servidor o en una memoria de un terminal móvil.

15  
16. Sistema digital de mensajería según una o varias de las reivindicaciones del sistema anteriores, comprendiendo el modelo tridimensional múltiples imágenes bidimensionales, que se almacenan en la memoria del sistema junto con ciertos datos, que describen la perspectiva/orientación de una imagen bidimensional individual.

20  
17. Sistema digital de mensajería según una o varias de las reivindicaciones del sistema anteriores, proporcionándose un sistema de inserción que inserta otro fondo en función del paradero predeterminado del remitente del mensaje.

25



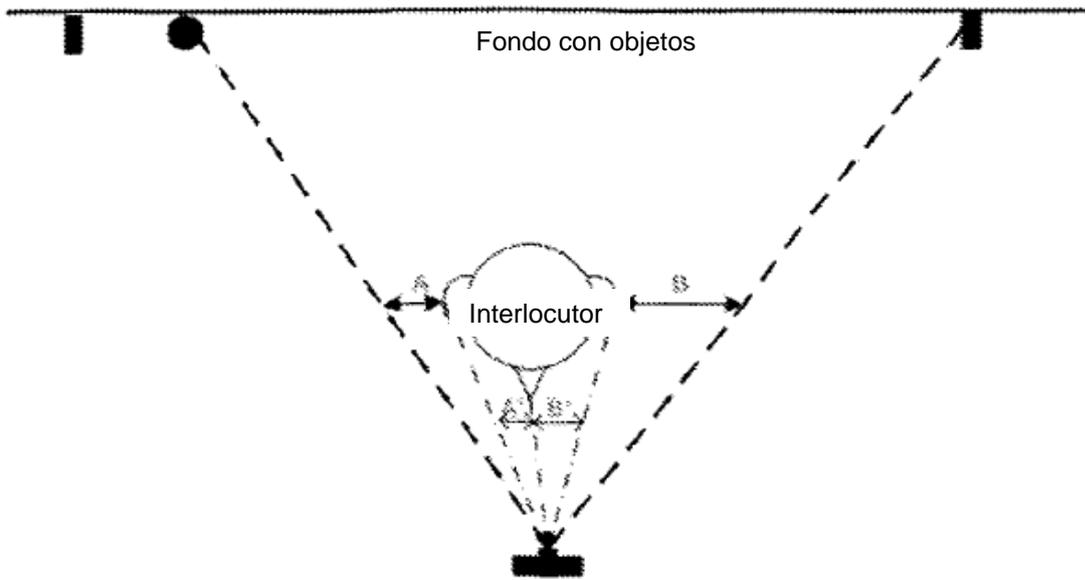


Fig. 2a

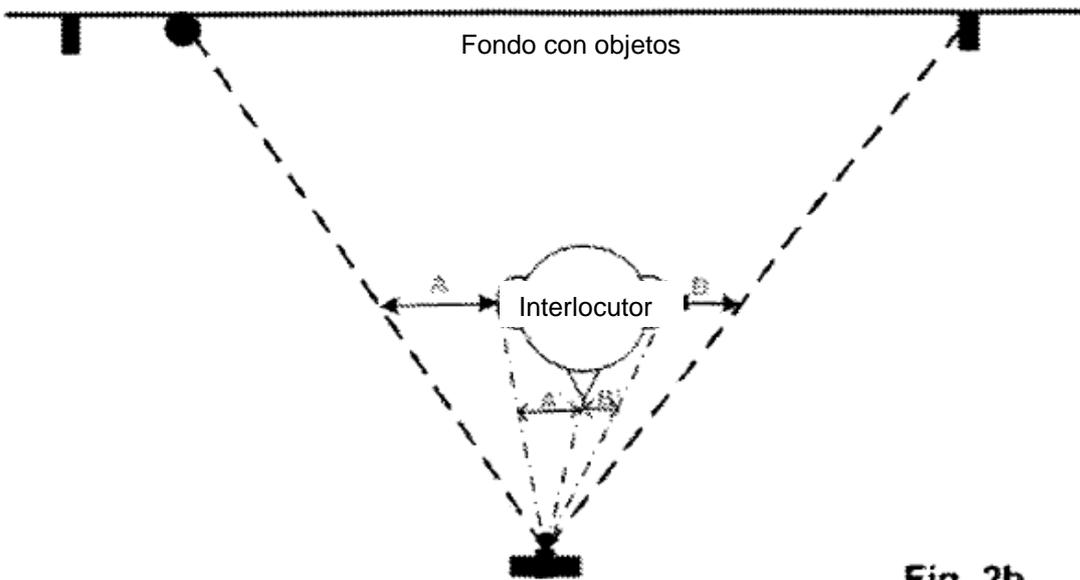
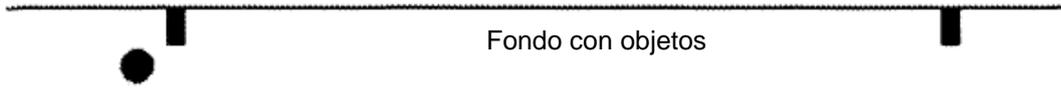


Fig. 2b

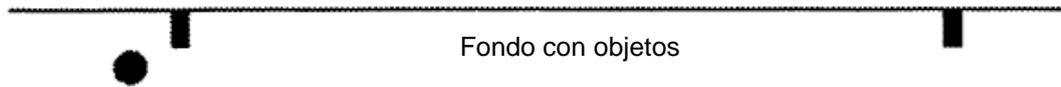


Fondo con objetos

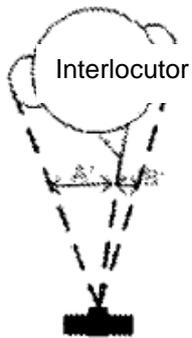


Cámara de grabación

**Fig. 3a**



Fondo con objetos



Cámara de grabación

**Fig. 3b**

**REFERENCIAS CITADAS EN LA MEMORIA DESCRIPTIVA**

5 La lista siguiente de los documentos mencionados por parte del solicitante ha sido realizada exclusivamente a fin de informar al lector y no forma parte del documento de patente europeo. Ha sido elaborada con mucho esmero; sin embargo, la Oficina Europea de Patentes no asume ninguna responsabilidad en el caso de errores u omisiones eventuales.

**Documentos de patente citados en la memoria descriptiva**

10

- EP 1648151 A1
- DE 102004014189 A1
- WO 2006047347 A
- EP 1412917 B1
- EP 0991023 B1
- US 4925294 A1
- DE 69628662 T2