(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 507 550**

(51) Int. Cl.:

H04M 3/56 (2006.01)
H04M 3/53 (2006.01)
H04M 3/42 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **08.05.2006** **E 06380104 (7)**

(97) Fecha y número de publicación de la concesión europea: **20.08.2014** **EP 1727340**

(54) Título: **Sistema y método de transcripción de conversación telefónica en tiempo real**

(30) Prioridad:

24.05.2005 ES 200501257

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
15.10.2014

(73) Titular/es:

VODAFONE GROUP PLC (50.0%)
Vodafone House The Connection
Newbury Berkshire RG14 2FN, GB y
VODAFONE ESPANA, S.A. (50.0%)

(72) Inventor/es:

TOUSET RÍOS, MIGUEL ÁNGEL

(74) Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 507 550 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

Sistema y método de transcripción de conversación telefónica en tiempo real

Campo de la invención

La presente invención se enmarca dentro del campo de las telecomunicaciones, y de manera más específica en el envío de datos correspondientes a una llamada de voz entre dos o más teléfonos móviles.

Antecedentes de la invención

El estado de la técnica actual proporciona dispositivos que son capaces de transcribir en texto un flujo de conversación en tiempo casi real. Existen también sistemas que capturan un texto “dedicado”, transformando la conversación en caracteres alfanuméricos.

Existen asimismo sistemas capaces de separar la voz de los diferentes hablantes participantes en una conversación o diálogo.

En otros sistemas propuestos, tales como la solicitud de patente de US US-2005/0013419- A1 por ejemplo, se propone almacenar un texto dictado por un hablante para enviarlo más tarde a través de diferentes servicios de intercambio de mensajes, tales como SMS o correo electrónico. En la solicitud de patente Australiana WO 03/071774- A1, se propone un método para llevar a cabo una llamada de conferencia entre dos o más participantes. Es un sistema para redes de paquetes en el cual se requiere un nodo de central en el que se establece la conferencia. FR2846826 es un documento que describe una técnica de transcripción entre hablantes en terminales móviles.

Es conocido que se utilizan frecuentemente abreviaturas y acrónimos en el campo de la telefonía móvil. A continuación se presenta un glosario de acrónimos / términos que son utilizados en esta memoria:

DTMF:	Multi – Frecuencia de Tono Dual	Dual Tone Multi – Frequency, en inglés
HLR:	Registro de Ubicación de abonados Locales	Home Location Register, en inglés
MSC:	Centro de Conmutación para Móviles	Mobile Switching Centre, en inglés
MSISDN:	Número de ISDN Internacional de Estación de Telefonía Móvil	Mobile Station International ISDN Number, en inglés
SCP:	Punto de Control de Servicio	Service Control Point, en inglés
SIP:	Protocolo de Iniciación de Sesión	Session Initiation Protocol, en inglés
SMS:	Servicio de Mensajes Cortos	Short Message Service, en inglés
SMSC:	Centro de Servicio de Mensajes Cortos	Short Message Service Centre, en inglés
URL:	Localizador de Recurso Universal	Universal Resource Locator, en inglés
USSD:	Datos de Servicio Suplementario No estructurado	Unstructured Supplementary Service Data, en inglés

Descripción de la invención

La invención se refiere a un sistema de transcripción de conversación telefónica en tiempo real de acuerdo con la reivindicación 1, y a un método de acuerdo con la reivindicación 11. Realizaciones preferidas del sistema y del método están definidas en las reivindicaciones dependientes.

La presente invención proporciona un sistema, a diferencia de los conocidos en la tecnología actual, que permite la transcripción en tiempo real de una conversación que ocurre entre dos terminales móviles y que mantiene la comunicación de la conversación activa. De acuerdo con las características del sistema de la invención, este servicio de transcripción o subtitulado de llamadas bajo demanda (puesto que el abonado activa y desactiva el servicio a voluntad) puede ser solicitado al principio de o durante una llamada de voz, sin requerir que todas las llamadas que el abonado está realizando pasen a través de un nodo de central. Las muestras de voz de cada abonado son obtenidas del procesamiento de la voz incluida en la conversación.

De acuerdo con un primer aspecto de la invención, lo último se refiere a un sistema de transcripción en tiempo real para transcribir una llamada telefónica entre terminales móviles primero y segundo (A, B), que comprende:

- medios para recibir una señal de activación del servicio de transcripción de uno de los dos terminales móviles,

- y en respuesta a la citada activación, medios para recibir una señal de conversación conjunta que incluye las voces de los citados terminales móviles primero y segundo,

- medios para separar la citada conversación en tiempo real y obtener señales de conversación primera y segunda independientes de los citados terminales móviles primero y segundo,

- medios para transcribir las citadas señales de conversación primera y segunda en un único formato preestablecido,

- medios de envío de las citadas señales de conversación primera y segunda transcritas en un formato preestablecido único al terminal móvil (A o B) que solicitó el servicio.

Los citados medios para recibir la señal de conversación conjunta son gestionados directamente por el terminal móvil que solicita el servicio, o pueden ser gestionados por un nodo de control de la red de telefonía móvil.

El citado formato preestablecido único puede ser determinado mediante las características técnicas del terminal que solicita el servicio de transcripción, o por el abonado que solicita el servicio o por la red.

El citado formato preestablecido único puede ser texto, imagen u otros.

Los citados medios de envío de las citadas señales de conversación primera y segunda transcritas preferiblemente incluyen medios de conexión con un Centro de Servicio de Mensajes Cortos (SMSC – Short Message Service Center, en inglés) para enviar por medio de SMS. O pueden preferiblemente incluir medios de conexión con una Puerta de Enlace de USSD (USSD GW – USSD GateWay, en inglés) y un Registro de Ubicación de Abonados Locales (HLR – Home Location Register, en inglés) del abonado. O preferiblemente las citadas señales de conversación primera y segunda transcritas están incluidas en una sesión de datos.

El sistema preferiblemente comprende medios para recibir una señal de desactivación del servicio de transcripción.

De acuerdo con un segundo aspecto de la invención, lo último se refiere a un método de transcripción en tiempo real para transcribir una conversación telefónica entre terminales móviles primero y segundo (A, B), que comprende:

- el que uno de los dos terminales móviles envíe una señal de activación del servicio de transcripción,

- y en respuesta a la citada activación, que reciba una señal de conversación conjunta que incluye las voces y los citados terminales móviles primero y segundo,

- separar la citada conversación en tiempo real y obtener señales de conversación primera y segunda independientes desde los citados terminales móviles primero y segundo,

- transcribir las citadas señales de conversación primera y segunda a un único formato preestablecido,

- enviar las citadas señales de conversación primera y segunda transcritas a un único formato preestablecido a un terminal móvil (A o B) que solicitó el servicio.

La citada señal de conversación conjunta puede estar gestionada directamente por el terminal móvil que solicita el servicio, o por un nodo de control de la red de telefónica móvil.

El envío de las señales de conversación primera y segunda transcritas en un único formato preestablecido puede ser realizado por medio de SMS, o por medio de una sesión de USSD generada por la red, o por medio de una sesión de datos (como es el caso cuando un terminal móvil está implicado y es un terminal de clase A).

La activación del servicio puede ser solicitada mediante el envío de un tono de DTMF predeterminado, o enviando un código de USSD predeterminado, o por medio de SMS, o por medio de una sesión de datos. Por ejemplo, en CAMEL de fase 3, los SMSs pueden activar un servicio de red inteligente. O bien la activación del servicio puede ser realizada por medio de la solicitud de la identidad del sistema de transcripción de la invención (por ejemplo, MSISDN o SIP URL).

La activación del servicio puede ser también solicitada mediante voz. En el citado caso la activación de voz, además de obtener señales de conversación primera y segunda independientes de los citados terminales móviles primero y segundo, el sistema reconoce cuál corresponde a cada uno, y por lo tanto sólo la señal de conversación primera o segunda correspondiente al otro terminal móvil que no solicitó el servicio, transcrita a un único formato o formatos preestablecido o preestablecidos, es enviada al terminal móvil que solicitó el servicio.

El método preferiblemente incluye una señal de desactivación del servicio de transcripción. Esta desactivación del servicio se realizaría de una manera similar a la indicada para la activación del servicio.

Puede ocurrir también que el terminal que solicitó el servicio reciba en una primera aproximación la transcripción completa de la conversación, y seleccione a continuación la parte que desea que sea enviada. Si la activación se realiza mediante voz, la propia voz puede ser utilizada para saber quién activó el servicio para poder enviarle sólo la transcripción del otro.

Puede considerarse también que una vez que el sistema de la invención está recibiendo la señal de conversación conjunta incluyendo las voces de los citados terminales móviles primero y segundo, el abonado que solicita el servicio puede ser distinguido con un identificador preestablecido (tono de DTMF, código de USSD, SMS, muestra de voz u otros) con el fin de enviarle así sólo la transcripción del otro terminal.

El citado formato preestablecido único puede ser determinado por el abonado que solicita el servicio de transcripción, o mediante las características técnicas del terminal, o por la red.

El sistema propuesto por la invención puede ser escalado al caso en el cual existen N hablantes (N >= 2) y los N hablantes solicitaron el servicio, puesto que mientras se mantiene la conexión de voz establecida, partes de la conversación de los N hablantes podrían ser pasadas a texto, lo que se mostraría a continuación en la pantalla del abonado que solicita el servicio.

El sistema puede ser aplicado a cualquier tipo de terminales, ya sean terminales móviles o fijos, PCs, Palms, etc., que tenga un elemento de visualización al que se pueda acceder durante una conversación telefónica.

Por otro lado, el sistema puede ser extendido al caso de transcripción en tiempo real, simplemente traduciendo el texto obtenido a partir de la conversación a otro idioma, que puede ser seleccionado cuando se solicita el servicio. Y de manera más general, debido al hecho de que un canal de voz está abierto con el abonado del terminal móvil que solicitó el servicio de transcripción, el audio traducido podría ser incluido por el citado canal. Por ejemplo, si el abonado desea conocer cómo pronunciar y escribir una frase dada en otro idioma, simplemente debe solicitar el servicio (llamando a un número B predeterminado por el operador, por ejemplo, y en este caso el sistema de la invención tiene medios para que sea traducido directamente), decir la frase y ver su traducción en texto en la pantalla de su terminal y oír la pronunciación de la misma en el receptor. En este caso, N = 1.

Además de estar en formato de texto, puede ocurrir también que la información mostrada al abonado sea en formato multimedios (en cuyo caso se utiliza un enlace de datos para mostrar la información procesada, o se utilizan MMS). Por ejemplo, las frases pronunciadas por B, o imágenes relativas a la conversación (lenguaje de signos) u otros, pueden ser añadidas a una imagen de la otra persona almacenada en el terminal de A, como en las tiras de comics. Estas últimas opciones requieren una función adicional en los elementos de la red con el fin de hacer que lleguen al terminal, típicamente a través de un canal de datos, así como en el terminal que debe procesar estos datos y mostrarlos en la pantalla de acuerdo con las especificaciones del servicio.

Breve descripción de los dibujos

Una serie de dibujos que ayudan en la mejor comprensión de la invención y que están expresamente relacionados con una realización de la citada invención, presentada como un ejemplo no limitativo de la misma, están brevemente descritos en lo que sigue.

La Figura 1 muestra una posible implementación del sistema de la invención por medio de una red inteligente.

La Figura 2 muestra otra posible implementación del sistema de la invención por medio de lógica en el terminal móvil.

Descripción de una realización preferida de la invención

Tanto en la implementación mostrada en la Figura 1 como en la mostrada en la Figura 2, se encuentra una llamada de voz entre dos abonados de terminales móviles A y B. Durante el curso de la citada llamada, uno de ellos, por ejemplo el hablante A, puede decidir (porque requiere una traducción o transcripción de la conversación o simplemente porque es sordo) que prefiere que la manera de interactuar con el otro hablante sea enriquecida por medio de la visualización de la transcripción de la conversación en la pantalla de su terminal.

Asimismo, si el entorno en el que se encuentra uno de los hablantes, por ejemplo A, excede un umbral de ruido que puede ser configurado en el terminal del hablante, el terminal solicitará el servicio de transcripción de la conversación.

De acuerdo con la invención y en la manera que se describe en lo que sigue (implementaciones en redes inteligentes o por medio de lógica en el terminal móvil), el hablante A solicitante recibirá la voz y la transcripción de la misma en el terminal, mientras que el otro hablante B mantendrá la interfaz de voz. La llamada existente al principio se mantiene y el hablante A que solicita el servicio también recibirá en tiempo real la transcripción de la conversación como si fuera un subtítulo.

La Figura 1 muestra la implementación basada en red inteligente, es decir, el caso en el cual el establecimiento del nodo responsable de recibir el audio de la conversación para la transcripción de la misma en modo de escucha es gestionado por la red de comunicaciones.

El proceso por el que pasará el servicio es como sigue:

P1. El abonado A en este caso indica a la red que desea establecer una llamadas de voz a su homologo B.

P2. La llegada de la solicitud al nodo de conmutación (MSC) 10 hace que la red inteligente active un nodo de control (SCP) 20 en el mismo de manera que lo último indica cómo continuar con el establecimiento de la llamada.

P3. El nodo de control SCP 20 devuelve el control de la llamada al nodo de conmutación MSC 10, indicando que debe establecer la llamada con el abonado B.

P4. El nodo de conmutación MSC 10 pone a los abonados A y B en comunicación. Por lo tanto, la llamada de voz entre ellos es establecida. A y B reciben voz.

P5. El abonado A (o su terminal cuando se detecta un nivel de ruido que está por encima de un umbral) desea solicitar el servicio de transcripción de conversación para la conversación en curso. Con este fin envía un cierto tono de DTMF o código de USSD a la red, o lo solicita, por ejemplo, por medio de una orden de voz o por medio de un SMS. O bien por medio de una solicitud de la identidad del nodo 30 (por ejemplo MSISDN).

P6. Esto provoca que el nodo de control SCP 20 sea “despertado” dado uno de los eventos activados previamente.

P7. La ejecución de la lógica en el nodo de control SCP 20 devuelve que para este tono de DTMF o código de USSD el nodo de conmutación MSC 10 establece una llamada entre el hablante A y el nodo de procesamiento de la señal de conversación que pertenece a la llamada entre A y B (nodo STT 30). Esto hace que la llamada entre A y B sea momentáneamente puesta en espera.

P8. El nodo STT 30 es puesto en conferencia con A y B por indicación del nodo de control SCP.

P9. En este momento, el nodo STT 30 escucha la conversación entre A y B. La primera tarea del nodo STT 30 es separar las muestras de voz de cada hablante para a continuación procesar cada una de ellas y convertirlas en texto. El texto resultante será enviado al hablante que solicita el servicio, A, por medio de un SMS de tipo 0 por ejemplo (es decir, no son almacenados en la memoria del terminal), una sesión de USSD generada por la red, o una sesión de datos (como es el caso cuando el terminal móvil del hablante A es un terminal de clase A). En el caso de utilizar la sesión de USSD, el nodo STT 30 pasa el texto transcrito de cada uno de los hablantes al HLR 50 tras pasar a través de una puerta de enlace de USSD (USSD GW – USSD GateWay, en inglés) 40.

P10. El HLR 50 se utiliza en la emisión del mensaje de USSD (o de SMS o creación de sesión de datos).

P11. El abonado A recibe la transcripción de la conversación por medio de la sesión de USSD generada por la red. Por lo tanto, A recibirá una transcripción y B sólo recibirá el audio.

El abonado puede solicitar qué parte de la conversación desea recibir transcrita, interactuando con el nodo STT 30 en la sesión de USSD establecida o solicitándolo por ejemplo por medio de una orden de voz; o por medio de tonos de DTMF procesados por el nodo de control SCP 20 e interactuando entre el nodo de control SCP 20 y el nodo STT 30 por medio de un enlace externo. De la misma manera, el abonado A o su terminal pueden indicar a la red que desea abandonar el servicio de transcripción.

La Figura 2 muestra la implementación por medio de lógica en el terminal móvil, es decir, el caso en el que el poner el nodo responsable de recibir el audio de la conversación para la transcripción del mismo en el modo de escucha es gestionado por el propio terminal móvil.

En este caso, el proceso por el cual pasará el servicio es como sigue:

S1. El abonado A indica a la red que desea establecer una llamada con su homólogo B.

S2. La llamadas de voz entre A y B es establecida en el nodo de conmutación (MSC) 10. A y B reciben voz.

S3. El terminal decide solicitar el servicio de transcripción de llamada (por ejemplo porque detecta un nivel de ruido en el entorno que está por encima de un umbral). O bien el abonado A decide el mismo o ella misma solicitar el servicio (por ejemplo pulsando un botón o seleccionando una función en el menú del terminal o por medio de una orden de voz).

S4. El terminal establece una llamada de voz con el nodo STT 30. Así que la llamada entre A y B será momentáneamente puesta en ESPERA.

S5. El terminal pone al nodo STT 30 en conferencia. Por lo tanto, A, B y STT recibirán el audio de la conferencia entre A y B.

S6. Como en la implementación mostrada en la Figura 1, la primera tarea del nodo STT 30 es separar las muestras de voz de cada hablante para a continuación procesar una de ellas y convertirla en texto. El texto resultante es enviado a A, por ejemplo, por medio de un SMS de tipo 0 (es decir, no son almacenados en la memoria del terminal), de una sesión de USSD generada por la red, o de una sesión de datos (como en el caso en el cual el terminal móvil del abonado A era un móvil de clase A). En el caso de utilizar la sesión de USSD, el nodo STT pasará el texto transcrito de cada uno de los hablantes al HLR 50 tras pasar a través de una puerta de enlace de USSD (USSD GW – USSD GateWay, en inglés) 40.

S7. El HLR 50 se utiliza en emitir el mensaje de USSD (o SMS o creación de sesión de datos).

S8. Un abonado A recibe la transcripción de la conversación por medio de la sesión de USSD generada por la red. Por lo tanto, A recibirá el audio y la transcripción y B recibirá sólo el audio.

El abonado puede solicitar qué parte de la conversación desea recibir transcrita, interactuando con el nodo STT 30 en la sesión de USSD establecida o los tonos de DTMF procesados por el nodo STT o solicitándolo mediante voz o SMS. De la misma manera, el abonado A o su terminal pueden indicar a la red que desea abandonar el servicio de transcripción.

Aunque la presente invención ha sido descrita para redes de telefonía móvil, también sería aplicable a las redes de telefonía fija, tal como por ejemplo para conversaciones a través de Internet entre dos PCs.

REIVINDICACIONES

1. Un sistema de transcripción en tiempo real para transcribir una conversación telefónica entre terminales móviles primero y segundo (A, B) en una red de comunicaciones, que comprende:

- medios (10) para recibir una señal de activación del servicio de transcripción desde uno de los dos terminales móviles (A, B),

- y en respuesta a la citada señal de activación, los medios (30) para recibir una señal de conversación conjunta que incluye las voces de los citados terminales móviles primero y segundo,

- medios (30) para separar la citada conversación en tiempo real y obtener señales de conversación primera y segunda independientes de los citados terminales móviles primero y segundo (A, B),

- medios (30) para transcribir las citadas señales de activación primera y segunda en un formato preestablecido,

- medios (40; 50) de envío de las citadas señales de conversación primera y segunda transcritas al terminal móvil (A o B) que solicitó el servicio, de manera que las señales de conversación transcritas son enviadas por medio de una sesión de Datos de Servicio Suplementario No estructurado, USSD (Unstructured Supplementary Service Data, en inglés) establecida por la red (P11).

2. Un sistema de acuerdo con la reivindicación 1, caracterizado porque los citados medios (10) para recibir la señal de conversación conjunta están gestionados directamente por el terminal móvil (A) que solicita el servicio.

3. Un sistema de acuerdo con la reivindicación 1, caracterizado porque los citados medios (10) para recibir la señal de conversación conjunta están gestionados por un nodo de control (20) de la red de telefonía móvil.

4. Un sistema de acuerdo con cualquiera de las reivindicaciones previas, caracterizado porque el citado formato preestablecido está determinado por las características técnicas del terminal (A) que solicita el servicio de transcripción.

5. Un sistema de acuerdo con cualquiera de las reivindicaciones 1 – 3, caracterizado porque el citado formato preestablecido está determinado por el abonado que solicita el servicio.

6. Un sistema de acuerdo con cualquiera de las reivindicaciones 1 – 3, caracterizado porque el citado formato preestablecido está determinado por la red.

7. Un sistema de acuerdo con cualquiera de las reivindicaciones precedentes, en el que los medios (10) para recibir una señal de activación del servicio de transcripción están configurados para recibir un código de USSD predeterminado como la señal de activación del servicio de transcripción.

8. Un sistema de acuerdo con cualquiera de las reivindicaciones 1 – 7, caracterizado porque los citados medios (40; 50) de envío de las citadas señales de conversación primera y segunda transcritas incluyen medios para conectarse con una puerta de enlace de USSD, USSD GW (40) y un Registro de Ubicación de Abonados Locales, HLR, (50).

9. Un sistema de acuerdo con cualquiera de las reivindicaciones previas, caracterizado porque comprende además medios (10) para recibir una señal de desactivación del servicio de transcripción.

10. Un método de transcripción en tiempo real para transcribir una conversación telefónica entre los terminales móviles primero y segundo (A, B) en una red de comunicaciones, que comprende:

- el que uno de los dos terminales móviles envíe una señal de activación del servicio de transcripción (P1; S1),

- y en respuesta a la citada señal de activación, recibir una señal de conversación conjunta que incluye las voces de los citados terminales móviles primero y segundo (P8; S4),

- separar la citada conversación en tiempo real y obtener señales de conversación primera y segunda independientes de los citados terminales móviles primero y segundo,

- transcribir las citadas señales de conversación primera y segunda a un formato preestablecido,

- enviar las citadas señales de conversación primera y segunda transcritas al terminal móvil (A o B) que solicitó el servicio, de manera que las señales de conversación transcritas son enviadas por medio de una sesión de Datos de Servicio Suplementario No estructurado, USSD, establecida por la red (P10, P11; S7, S8).

11. Un método de acuerdo con la reivindicación 10, caracterizado porque la citada señal de conversación conjunta es gestionada directamente por el terminal móvil que solicitó el servicio.

12. Un método de acuerdo con la reivindicación 10, caracterizado porque la citada señal de conversación conjunta es gestionada por un nodo de control (20) de la red de telefonía móvil.

13. Un método de acuerdo con una cualquiera de las reivindicaciones 10 a 12, en el que la señal de activación del servicio de transcripción (P1; S1) es un código de USSD predeterminado.

14. Un método de acuerdo con cualquiera de las reivindicaciones 10 - 13, caracterizado porque una vez que se recibe la señal de conversación conjunta que incluye las voces de los citados terminales móviles primero y segundo, el abonado que solicita el servicio es distinguido por medio de un identificador preestablecido, y sólo la señal de conversación primera o segunda correspondiente al otro terminal móvil (B o A), transcrita y enviada al abonado.

15. Un método de acuerdo con cualquiera de las reivindicaciones 10 – 14, caracterizado porque el abonado que solicita el servicio es quién determina el citado formato preestablecido.

16. Un método de acuerdo con cualquiera de las reivindicaciones 10 – 14, caracterizado porque las características del terminal móvil que solicita el servicio de transcripción determinan el citado formato preestablecido.

17. Un método de acuerdo con cualquiera de las reivindicaciones 10 – 14, caracterizado porque la red determina el citado formato preestablecido.

18. Un método de acuerdo con cualquiera de las reivindicaciones 10 – 17, caracterizado porque incluye el envío de una señal de desactivación del servicio de transcripción.

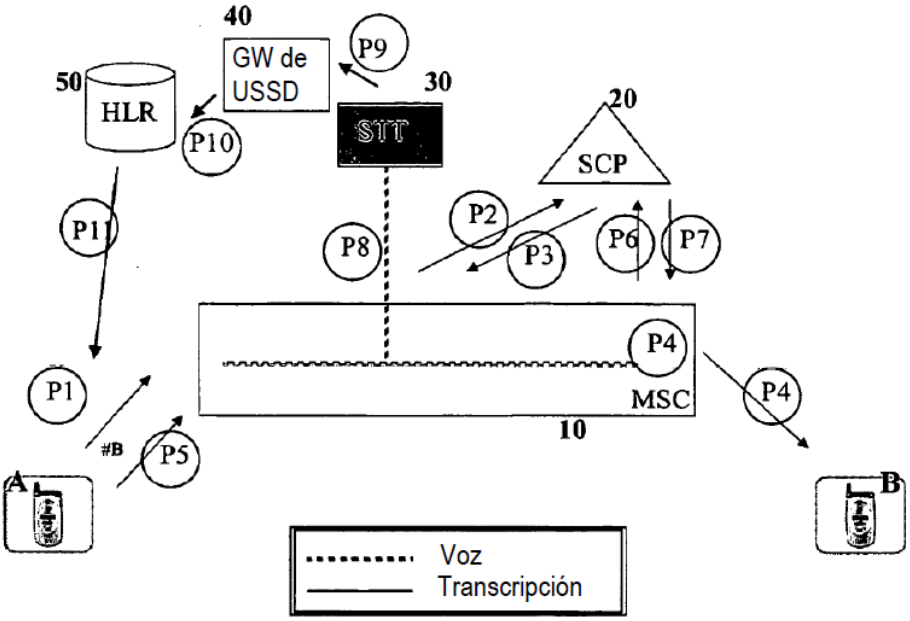
40
GW de USSD
P9
50
HLR
30
STT
20
SCP
P10
P11
P2
P3
P6
P7
P8
P4
MSC
10
P1
#B
P5
A
B
Voz
Transcripción

FIG. 1

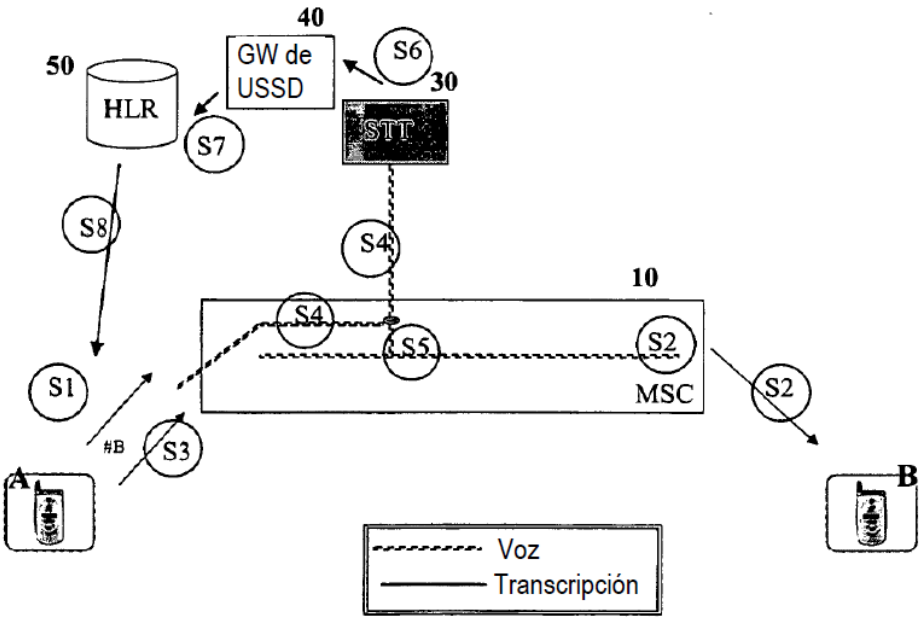
40
GW de USSD
S6
50
HLR
30
STT
S7
S8
S4
10
S4
S5
S2
MSC
S1
#B
S3
S2
A
B
Voz
Transcripción

FIG. 2