

(12) SOLICITUD INTERNACIONAL PUBLICADA EN VIRTUD DEL TRATADO DE COOPERACIÓN EN MATERIA DE PATENTES (PCT)

(19) Organización Mundial de la Propiedad
Intelectual
Oficina internacional



(10) Número de Publicación Internacional
WO 2011/144775 A1

(43) Fecha de publicación internacional
24 de noviembre de 2011 (24.11.2011)

PCT

(51) Clasificación Internacional de Patentes:
H04N 7/24 (2011.01)

(21) Número de la solicitud internacional:
PCT/ES2011/000166

(22) Fecha de presentación internacional:
20 de mayo de 2011 (20.05.2011)

(25) Idioma de presentación: español

(26) Idioma de publicación: español

(30) Datos relativos a la prioridad:
P201030758 20 de mayo de 2010 (20.05.2010) ES

(71) Solicitante (para todos los Estados designados salvo US): **UNIVERSIDAD CARLOS III DE MADRID** [ES/ES]; Av. Gregorio Peces Barba, 1, Parque Científico, E-28918 Leganés (ES).

(72) Inventores; e

(75) Inventores/Solicitantes (para US solamente): **DE CASTRO ÁLVAREZ, Mercedes** [ES/ES]; Av. Gregorio Peces Barba, 1, Parque Científico, E-28918 Leganés (ES). **DE PEDRO SASTRE, Manuel** [ES/ES]; Av. Gregorio Peces Barba, 1, Parque Científico, E-28918 Leganés (ES). **RUIZ MEZCUA, María Belén** [ES/ES]; Av. Gregorio Peces Barba, 1, Parque Científico, E-28918 Leganés (ES).

JIMÉNEZ DORADO, Javier [ES/ES]; Av. Gregorio Peces Barba, 1, Parque Científico, E-28918 Leganés (ES).

(81) Estados designados (a menos que se indique otra cosa, para toda clase de protección nacional admisible): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Estados designados (a menos que se indique otra cosa, para toda clase de protección regional admisible): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), euroasiática (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europea (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continúa en la página siguiente]

(54) Title: METHOD AND DEVICE FOR SYNCHRONISING SUBTITLES WITH AUDIO FOR LIVE SUBTITLING

(54) Título : PROCEDIMIENTO Y DISPOSITIVO PARA SINCRONIZAR SUBTÍTULOS CON AUDIO EN SUBTITULACIÓN EN DIRECTO

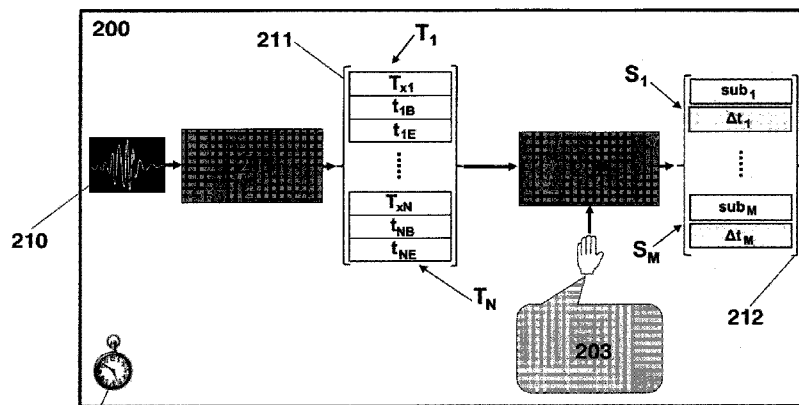
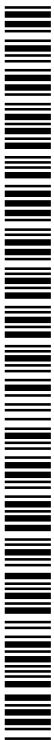


FIG. 2

(57) Abstract: The invention relates to live audio and subtitle synchronisation which determines an individual subtitle delay Δt_i for each subtitle i and creates a subtitle unit S_i with the subtitle (i) and the individual subtitle delay (Δt_i). Each subtitle unit S_i is introduced into one or more subtitle packets PES with a presentation time stamp PTS_i and an associated configurable movement delay REST. Said packets form part of an extended transport flux that includes audio packets containing at least one audio fragment j and a presentation time stamp PTS_j . The subtitle packets are synchronised with audio packets, creating a synchronised transport flux comprising all of the packets of the extended transport flux, delayed by a time dependant on an associated configurable movement delay REST.

(57) Resumen:

[Continúa en la página siguiente]



WO 2011/144775 A1



Publicada:

— con informe de búsqueda internacional (Art. 21(3))

— antes de la expiración del plazo para modificar las reivindicaciones y para ser republicada si se reciben modificaciones (Regla 48.2(h))

La sincronización de subtítulos y audio en directo que se propone determina un retardo de subtítulo individual Δt_i de cada subtítulo i y crea una unidad de subtítulo S_i con el subtítulo i y el retardo de subtítulo individual Δt_i . Cada unidad de subtítulo S_i se introduce en uno o más paquetes PES de subtítulo con una marca de tiempo de presentación PTS_i y junto con un retardo de desplazamiento asociado configurable REST. Dichos paquetes forman, parte de un flujo de transporte extendido que incluye paquetes de audio que contiene al menos un fragmento de audio j y una marca de tiempo de presentación PTS_j . Se realiza una sincronización de los paquetes de subtítulo con paquetes de audio creando un flujo de transporte sincronizado que comprende todos los paquetes del flujo de transporte extendido retardados un tiempo que depende de un I retardo de desplazamiento asociado configurable REST.

PROCEDIMIENTO Y DISPOSITIVO PARA SINCRONIZAR SUBTÍTULOS CON AUDIO EN SUBTITULACIÓN EN DIRECTO

CAMPO TÉCNICO DE LA INVENCION

5 La presente invención tiene su aplicación en el campo de la generación de subtítulos para eventos multimedia en directo, por ejemplo programas de televisión o radio en directo, para su visualización en tiempo real y sincronizados tanto con el audio como con el vídeo.

ANTECEDENTES DE LA INVENCION

10 Los subtítulos, como modo de acceder al contenido de audio de eventos multimedia suministrado a través de gran diversidad de dispositivos (equipos de televisión, teléfonos móviles 3G, ordenadores,...) son necesarios para personas con dificultades auditivas o personas que necesitan cierta ayuda para entender el idioma. Los subtítulos también son necesarios cuando se experimentan contenidos multimedia en entornos ruidosos o lugares en
15 los que debe desactivarse el sonido.

Aunque la subtitulación de eventos que no son en directo es una práctica normal hoy en día, la subtitulación de eventos en directo es menos frecuente debido a la complejidad inherente de la tecnología actual. La subtitulación en tiempo real de eventos audiovisuales es una investigación pluridisciplinar que abarca tecnologías de los campos de reconocimiento
20 automático del habla, informática y transmisión y difusión de paquetes en red. En eventos en directo en los que los subtítulos se presentan en una pantalla auxiliar, pero el vídeo y el audio se experimentan directamente, no existe la oportunidad de gestionar el retardo de los subtítulos. Lleva tiempo crear subtítulos a partir del habla, independientemente del procedimiento usado para la transcripción de audio a texto. En las mejores condiciones, el
25 retardo en la creación de subtítulos de texto puede ser de varios segundos. Cualquier procedimiento que intente tratar el problema del retardo entre audio y subtítulos debe proporcionar un procedimiento para determinar la correspondencia temporal entre un fragmento de audio y los subtítulos generados a partir del mismo.

En el actual estado de la técnica para la generación de subtítulos en programas de
30 televisión en directo, siempre hay un retardo significativo y variable entre el audio y el momento en el tiempo en el que están listos los subtítulos correspondientes. Como resultado, los subtítulos se muestran en la pantalla del usuario un número variable de segundos después y por tanto de manera no sincronizada con el audio/vídeo. Para entender mejor los procesos implicados en la subtitulación en directo (10), la figura 1 muestra las etapas relevantes:

35 Etapa 1)

Transcripción de audio (1): la transcripción inmediata de habla a partir de uno o varios hablantes a texto.

Etapa 2)

Generación de subtítulos (2): la generación de subtítulos a partir de texto.

Etapas 3 y 4)

Segmentación en paquetes (3) y transmisión/difusión (4) a través de los sistemas multimedia, por ejemplo, televisión digital terrestre (41), Internet (42), televisión por protocolo de Internet (43), etc.

Etapa 5)

Recepción y presentación (5) en la pantalla del usuario en, por ejemplo, un receptor de televisión (51), un ordenador (52), un teléfono móvil 3G o una televisión convencional con un codificador de IPTV (53), por mencionar sólo unos cuantos.

En entornos de generación de subtitulación en directo, el proceso de transcripción de audio (1) suministra el texto correspondiente varios segundos después de haber recibido el fragmento de habla. La transcripción de habla en texto en tiempo real es la causa originaria de los retardos de los subtítulos y es muy difícil de minimizar. En el pasado reciente y todavía hoy, el principal enfoque para la transcripción de audio (1) a texto a partir del habla ha sido el uso de estenotipia (11) en la que el texto transcrito a partir del audio se produce manualmente por estenotipistas. Aunque la calidad y velocidad de la estenotipia (11) son buenas, el coste de este proceso y la baja disponibilidad de estenotipistas hacen difícil que se convierta en una práctica extendida en la subtitulación masiva de eventos en tiempo real. Otra alternativa es el uso de ASR o motores de reconocimiento automático del habla (12) aplicados al audio del hablante, o bien directamente o a través de un rehablador intermedio (13), para reconocer automáticamente las palabras y frases con una mínima intervención humana. Los principales inconvenientes de la subtitulación con la tecnología de ASR actual son los retardos introducidos en los subtítulos con respecto a los flujos de vídeo y audio originales, las tasas de error de reconocimiento de habla y la necesidad de entrenar el sistema para la voz y el vocabulario del hablante. En la práctica, puede usarse ASR con la ayuda de un operador humano que manipula el texto de salida para mejorar o corregir los subtítulos. De hecho, para la subtitulación en directo de televisión, el uso de rehablado y manipulación es una práctica normal. El uso de ASR directamente sin un rehablador intermedio implica o bien aplicar un ASR independiente del hablante a cualquier voz en la señal de audio, o aplicar un ASR dependiente del hablante entrenado previamente para las diferentes voces en la señal de audio. El ASR basado en rehablado, en el que una persona habla a un ASR dependiente del hablante, tratando de repetir lo que está escuchando, es de los más usados en difusión de televisión en tiempo real en todo el mundo en la actualidad, por ejemplo, en el Reino Unido o en España. La estenografía se ha usado y todavía se usa ampliamente para subtitulación en tiempo real, por ejemplo en EE. UU., en conexión con un ordenador en el que una aplicación de software traduce los símbolos estenográficos en texto. Hay otros procedimientos, no muy comunes pero que se usan en ocasiones para subtitulación en tiempo real, tales como teclear directamente mediante un teclado o usar un teclado silábico tal como Veyboard™.

Para la generación de subtítulos (2), los subtítulos se crean a partir de los fragmentos de texto transcritos recibidos desde el subsistema de transcripción de audio (1). El subsistema de generación de subtítulos (2) permite la corrección y el formateo manuales (21), por ejemplo, asignando colores a los fragmentos de texto dependiendo de la información de hablante o de contexto, etc. Los subtítulos finales se suministran mediante el subsistema de subtítulos en directo (22) junto con otra información importante para su presentación final, tal como la duración (tiempo de permanencia en pantalla del subtítulo), el color, la posición en la pantalla, etc.

El posterior proceso de segmentación en paquetes (3) de los flujos de vídeo, audio y subtítulos genera paquetes de flujo de transporte (31) que se envían a través de sistemas de transmisión (4): por ejemplo por televisión digital o por IP. Cada paquete transmitido contiene, además de la carga útil e, independientemente de su tipo (vídeo, audio o datos), una indicación de un tiempo para la presentación en pantalla (`presentation_time_stamp`), que se usa en el lado de recepción para una presentación sincronizada de paquetes pertenecientes a los diferentes flujos de un canal de televisión. Dentro de los paquetes de flujo de transporte (31), hay también una referencia de reloj global (100) para la sincronización temporal, de modo que todas las marcas de tiempo de presentación hacen referencia a la misma referencia de reloj común.

Lo que es importante entender es que estas marcas de tiempo de presentación se calculan en el lado de emisión según el reloj en tiempo real (100) del sistema y corresponden al momento en el tiempo en el que se crean los paquetes de flujo de transporte (31). El resultado es que se mantienen los retardos entre audio/vídeo y subtítulos, es decir, los paquetes de audio y vídeo tienen marcas de tiempo de presentación que preceden en el tiempo a las marcas de tiempo de presentación de sus subtítulos correspondientes. En el lado de recepción, el usuario sólo detecta esta falta de alineación entre los flujos cuando activa los subtítulos.. Mientras que los flujos de audio y vídeo están sincronizados entre sí, los subtítulos se presentan varios segundos después y nunca se corresponden con el audio y vídeo en la pantalla.

SUMARIO DE LA INVENCION

La presente invención sirve para resolver el problema mencionado anteriormente sobre la desincronización entre audio y subtítulos en subtítulos en tiempo real, compensando los retardos individuales de los subtítulos con el fin de conseguir una resincronización eficaz de flujos de audio y de subtítulo antes de que sean presentados al usuario. La presente propuesta describe un sistema de extremo a extremo que permite sincronizar subtítulos y audio/vídeo en directo.

En escenarios de subtítulos en directo, cuando un evento audiovisual (por ejemplo, un programa de televisión en directo, un programa de radio, una obra de teatro o un discurso

en una conferencia) se digitaliza y transmite a dispositivos de recepción de usuario, existe la posibilidad de compensar el retardo introducido, principalmente por el proceso de transcripción de audio, entre el audio original y los correspondientes subtítulos. El hecho de que haya una fase de transmisión (a través de una red de comunicación) y una fase de presentación (en el receptor) en la transmisión en tiempo real de eventos en directo, como en la difusión de televisión por protocolo de Internet (IPTV) o de televisión digital terrestre (TDT), ofrece la oportunidad de manejar el proceso de extremo a extremo, lo que permite finalmente la presentación de vídeo/audio/subtítulos sincronizados en un dispositivo de presentación (por ejemplo, un dispositivo de visualización electrónico). Esto implica retardar el audio y el vídeo con respecto a los subtítulos en algún punto del proceso.

En el contexto de la invención, el flujo de transporte utiliza un protocolo de comunicaciones para audio, vídeo y datos (por ejemplo, subtítulos, teletexto,...), que constituye un tipo de formato de contenedor digital que encapsula flujos elementales segmentados en paquetes para dichos audio/vídeo y datos, y otra información adicional. Un ejemplo de flujo de transporte se especifica en MPEG-2 Parte 1, Sistemas (norma ISO/IEC 13818-1). El flujo de transporte se usa en aplicaciones de difusión tales como DVB (difusión de vídeo digital) para televisión digital terrestre (DVB-T), sistemas por cable (DVB-C), satélite (DVB-S), o en redes basadas en IP (DVB-IPTV) entre otros. El flujo de transporte MPEG es similar al flujo de programa MPEG aunque el primero es para transmisión de datos en la que es probable la pérdida de datos, mientras que el último está diseñado para medios más fiables tales como DVD. El flujo de transporte combina uno o más programas con una o más bases de tiempo independientes en un único flujo.

Los datos de audio y vídeo pueden codificarse como se describe en ITU-T Rec. H.262 | ISO/IEC 13818-2 e ISO/IEC 13818-3. Los flujos elementales de audio y vídeo comprimido resultantes se segmentan en paquetes para producir paquetes de flujo elemental segmentado en paquetes (PES), según se especifica en MPEG-2 Parte 1, Sistemas (norma ISO/IEC 13818-1). Los flujos elementales que forman un programa están constituidos por paquetes PES y comparten una base de tiempo común.

En general, para cualquier escenario aplicable de difusión digital (multimedia, televisión o radio), se construyen lógicamente flujos de transporte a partir de paquetes de flujo elemental (por ejemplo, paquetes PES especificados en MPEG-2 Parte 1), que son las estructuras de datos usadas para transportar datos de flujo elemental. Esto incluye los flujos elementales del flujo de transporte que contienen paquetes de subtítulo para transportar subtítulos.

Un paquete de flujo elemental consiste en una cabecera de paquete, que incluye una referencia temporal, seguida de un número de bytes contiguos pertenecientes al flujo de datos elemental. Dicha referencia temporal es relativa a la referencia de sincronismo común transportada en el flujo de transporte. Según esta referencia de tiempo, el retardo de extremo a extremo desde la entrada de señal a un codificador hasta la salida de señal desde un

decodificador es constante.

Un dispositivo de recepción digital puede ser un dispositivo de recepción de radio o televisión digital. Las tecnologías de radio digital para difusión de audio mediante el transporte de señales moduladas digitales se basan en normas que incluyen la difusión de televisión digital. Un dispositivo de recepción de televisión digital puede ser un módulo decodificador de televisión digital terrestre o un terminal de recepción integrado TDT, un equipo de usuario adaptado para recibir contenido de televisión por IP o difusión multimedia digital, o un receptor de medios digital conectado a una red doméstica que soporta control de televisión en directo (por ejemplo, desde un ordenador).

Un soporte de almacenamiento multimedia puede ser un medio de almacenamiento de disco o un servidor de medios. Una red de distribución multimedia puede ser una red de difusión de televisión digital terrestre, una red de transmisión de televisión por IP o, en general, una red IP conmutada que proporciona sitios web con canales de televisión.

Según un aspecto de la invención, se proporciona un procedimiento para sincronizar subtítulos con audio en sistemas de subtitulación en directo, que genera al menos un subtítulo i a partir de al menos un bloque de transcripción T_j correspondiente a un fragmento de audio j y que comprende los siguientes pasos:

- determinar un retardo de subtítulo individual Δt_i de cada subtítulo i ;
- crear una unidad de subtítulo S_i que comprende el subtítulo i y el retardo de subtítulo individual Δt_i ;
- poner cada unidad de subtítulo S_i junto con un retardo de desplazamiento asociado configurable REST en al menos un paquete de subtítulo PES, el cual tiene una marca de tiempo de presentación PTS_i , formando dichos paquetes de subtítulo PES parte de un flujo de transporte extendido que incluye al menos un paquete de audio que contiene el fragmento de audio j y una marca de tiempo de presentación PTS_j .
- sincronizar paquetes de subtítulo PES con paquetes de audio creando un flujo de transporte sincronizado que comprende todos los paquetes del flujo de transporte extendido retardados un tiempo que depende del retardo de desplazamiento asociado configurable REST.

En una posible realización de la invención, cuando la sincronización de subtítulos se hace para ser visualizados seguidamente en un dispositivo de presentación, en la sincronización de paquetes de subtítulo PES con paquetes de audio se retarda cada paquete de subtítulo PES asociándole una marca de tiempo de presentación en el flujo de transporte sincronizado que es igual a $PTS_i - \Delta t_i + REST$, y retardar cada paquete de audio asociándole una marca de tiempo de presentación en el flujo de transporte sincronizado que es igual a $PTS_j + REST$, siendo PTS_i y PTS_j las marcas de tiempo de presentación en el flujo de transporte extendido y REST el retardo de desplazamiento asociado del flujo de transporte extendido.

Según un aspecto adicional de la invención, se proporciona un dispositivo sincronizador de subtítulos, que puede implementarse en un subsistema de subtitulación en directo antes de la transmisión y/o almacenamiento de subtítulos generados o en un receptor, que comprende un controlador con medios de procesamiento para realizar el procedimiento descrito anteriormente. La generación de subtítulos para el almacenamiento y la transmisión y su sincronización por el dispositivo antes de dichas etapas de almacenamiento o transmisión se refieren a transmisión multimedia, de televisión digital o de radio digital. El receptor puede ser un receptor de radio digital, un receptor multimedia digital, un receptor de medios digital, o un receptor de televisión digital tal como un módulo decodificador de TDT o de IPTV.

El dispositivo sincronizador comprende medios de procesamiento configurados para:

- crear un flujo de transporte extendido a partir de un flujo de transporte inicial que comprende todos los paquetes del flujo de transporte inicial, los cuales tienen una marca de tiempo de presentación asociada, comprendiendo el flujo de transporte extendido:

- al menos un paquete de audio que contiene un fragmento de audio j ,
- al menos una unidad de subtítulo S_i que comprende un subtítulo i generado a partir de al menos un bloque de transcripción T_j que corresponde al fragmento de audio j y un determinado retardo de subtítulo individual Δt_i asociado a cada subtítulo i ,
- y un retardo de desplazamiento asociado REST que es configurable e indica un retardo fijo para sincronizar paquetes de subtítulo con paquetes de audio,

- ajustar la marca de tiempo de presentación asociado de cada paquete de subtítulo del flujo de transporte extendido de manera que su asociándole una marca de tiempo de presentación asociada en el flujo de transporte sincronizado sea que es igual a $PTS_i - \Delta t_i + REST$, y retardar cada paquete de audio asociándole una de manera que su marca de tiempo de presentación asociada en el flujo de transporte sincronizado que es sea igual a $PTS_j + REST$, siendo PTS_i y PTS_j unas marcas de tiempo de presentación asociadas en el flujo de transporte extendido al paquete de subtítulo y el paquete de audio respectivamente y REST el retardo de desplazamiento asociado en el flujo de transporte extendido.

Según un último aspecto de la invención, ésta trata de un programa informático que comprende medios de código de programa que ejecutan el procedimiento descrito anteriormente, cuando se cargan en un controlador implementado por un procesador de propósito general, uno o más procesadores de señal digital (DSP), o por una pluralidad de circuitos integrados programables dedicados independientes tales como circuitos integrados de aplicación específica (ASIC) y dispositivos lógicos programables (PLD) tales como PLA (matriz

de lógica programable), FPGA (matriz de puertas programables en campo), etc.

Las principales ventajas de la solución descrita son:

- 5 a) Proporciona un enfoque práctico para mejorar la comprensión de eventos subtítulados en directo mediante la compensación de los efectos molestos de la recepción simultánea de entradas no sincronizadas (por ejemplo, audio y subtítulos). Esto puede ser crítico para personas con dificultades auditivas y personas con habilidades limitadas en un idioma no nativo así como en entornos en los que debe desactivarse el audio.
- 10 b) La solución descrita permite su aplicación como opción seleccionable por el usuario. Esto permite no afectar a las personas que no están interesadas en subtítulos sincronizados.
- c) La implementación de la solución antes de la fase de transmisión no afecta a los receptores actuales.
- 15 d) No se requiere ancho de banda adicional en la implementación de la solución en el lado de recepción ni en la opción de difusión "casi-directo".
- e) El sincronizador de subtítulos propuesto puede incorporarse en las tecnologías actuales con modificaciones menores.
- 20 f) Proporciona un modo de sincronizar subtítulos cuando vuelve a emitirse un programa subtulado en directo sin coste adicional, tanto si esta sincronización se realiza en el momento de la redifusión como si se realiza en el momento del almacenamiento.
- g) El uso de un formato común que soporta un modo flexible de adaptarse a múltiples variantes para su implementación en difusión de televisión, cable, radio digital, IPTV, etc.
- 25 h) La etapa descrita para calcular retardos de subtítulo individuales para el procedimiento propuesto, permite su configuración y ajuste para adaptarlo a las herramientas, entornos, tipos de programas de televisión o radio y operadores de subtitulación en tiempo real.

30 DESCRIPCIÓN DE LOS DIBUJOS

Para completar la descripción que está realizándose y con el objeto de ayudar a un mejor entendimiento de las características de la invención, según un ejemplo preferido de realización práctica de ésta, adjunto a dicha descripción como parte integrante de la misma, hay un juego de dibujos en los que, a modo de ilustración y de manera no restrictiva, se ha representado lo siguiente:

35

Figura 1. – Muestra un sistema de extremo a extremo de subtitulación en directo según se conoce en el estado de la técnica.

Figura 2. – Muestra un diagrama de bloques esquemático de un sistema de

subtitulación en directo, según una posible realización de la invención.

Figura 3. – Muestra un diagrama de tiempo de un esquema de subtitulación en directo, según otra posible realización de la invención, para hacer un seguimiento de los retardos de subtítulo individuales para algunos casos particulares: (A) relación 1-1 entre bloques de transcripción y subtítulos, (B) relación 1-M entre bloques de transcripción y subtítulos, (C) relación N-1 entre bloques de transcripción y subtítulos.

Figura 4. – Muestra un diagrama de tiempo de un esquema de subtitulación en directo, según otra posible realización de la invención, para hacer un seguimiento de los retardos de subtítulo individuales para algunos casos particulares: (D) relación N-M, $N < M$, entre bloques de transcripción y subtítulos y (E) relación N-M, $N > M$, entre bloques de transcripción y subtítulos.

Figura 5. – Muestra un diagrama de tiempo de un subsistema de transcripción de audio para hacer un seguimiento de los retardos de subtítulo individuales cuando se realiza transcripción de audio sin intervención humana, según una posible realización de la invención.

Figura 6. – Muestra un diagrama de tiempo de un subsistema de transcripción de audio para hacer un seguimiento de retardos de subtítulo individuales cuando se realiza transcripción de audio con intervención humana, según otra posible realización de la invención.

Figura 7. – Muestra una representación esquemática de las etapas y flujos de datos para sincronización de subtítulos en recepción, según una posible realización de la invención.

Figura 8. – Muestra una representación esquemática de las etapas y flujos de datos para sincronización de subtítulos antes de la transmisión, según otra posible realización de la invención.

Figura 9. – Muestra una representación esquemática de las etapas y flujos de datos para la sincronización de subtítulos antes de su almacenamiento para una posterior difusión/transmisión/distribución de canales de televisión, según una posible realización adicional de la invención.

Figura 10. – Muestra un diagrama de estados con los posibles modos de reproducción para reproducción sincronizada de subtítulos en el lado del receptor, según la realización de la invención mostrada en la figura 7.

Figura 11. – Muestra una representación esquemática de la utilización de eventos subtitulados en directo y almacenados previamente, para su redifusión o distribución fuera de línea en donde los subtítulos ya están sincronizados, según otra posible realización de la invención.

DESCRIPCIÓN DETALLADA DE LA INVENCION

La figura 2 muestra un sistema de subtitulación en directo 200, que comprende un subsistema de transcripción de audio 201 y un subsistema de generación de subtítulos 202. El subsistema de transcripción de audio 201 obtiene fragmentos de texto a partir de una fuente de audio (que puede ser la señal de audio asociada a la señal de vídeo u otra fuente de audio tal

como el reablado) y asigna referencias de tiempo a esos fragmentos de texto según sus fragmentos de audio originales correspondientes, según una posible realización de la invención. El subsistema de generación de subtítulos 202 genera unidades de subtítulo a partir de uno o muchos fragmentos de texto obtenidos en el subsistema de transcripción de audio 201, y asigna un retardo de subtítulo individual Δt_i a cada subtítulo i , $i = 1..M$. Estos retardos Δt_i pueden calcularse haciendo un seguimiento del tiempo empleado en cada uno de estos dos subsistemas y estimando el tiempo en el que se produjo el correspondiente audio siempre que no sea posible medirlo con precisión.

El subsistema de transcripción de audio 201 se encarga de obtener una transcripción de textos de una señal de audio 210 y de poner una marca de tiempo en la misma para hacer un seguimiento del comienzo de su fragmento de audio correspondiente, transcrito dentro de un bloque de transcripción 211. La salida del subsistema de transcripción de audio 201 es un grupo de $N \geq 1$ bloques de transcripción 211, en el que cada bloque de transcripción generado T_j , $j = 1..N$, contiene un fragmento de texto T_{xj} que es el texto de transcripción correspondiente a un fragmento de audio j y que está asociado con las marcas de tiempo correspondientes t_{jB} y t_{jE} donde:

t_{jB} es el tiempo en el que comienza el fragmento de audio j de la señal de audio 210;

t_{jE} es el tiempo en el que finaliza el fragmento de audio j .

Cuando el subsistema de transcripción de audio 201 no puede proporcionar valores exactos para t_{jB} y t_{jE} , los estima como se explica a continuación. La notación usada para los valores estimados es ts_{jB} y ts_{jE} . Para mayor simplicidad, la figura 2 se refiere tanto a t_{jB} como a ts_{jB} y como t_{jB} .

El subsistema de generación de subtítulos 202 se encarga de crear subtítulos a partir de los bloques de transcripción 211 obtenidos en el subsistema de transcripción de audio 201 previo. El subsistema de generación de subtítulos 202 permite la entrada de información de formateo 203 generada mediante corrección manual y/o NLP (herramientas de procesamiento de lenguaje natural) de los bloques de transcripción 211, edición, formateo y definición de duración (si no se proporcionan automáticamente), signos de puntuación, etc. Así, el subsistema de generación de subtítulos 202 también debe evaluar el tiempo total empleado en estas tareas para calcular el retardo de subtítulo individual Δt_i final entre el momento en el tiempo en el que se suministra el subtítulo i , presente, y el tiempo, en el pasado, en el que se produjo el fragmento de audio j correspondiente. Asimismo, este subsistema calcula la duración (longitud de tiempo que un subtítulo tiene que estar presente en la pantalla en el dispositivo terminal del usuario) de cada subtítulo creado. Finalmente, el subsistema de generación de subtítulos 202 suministra cada subtítulo i , por ejemplo, imagen DVB o texto para subtítulo de Teletexto, junto con otra información adicional de subtítulo tal como tiempo de duración, color, posición en la pantalla, etc., formando una unidad de subtítulo S_i asociada con su retardo de subtítulo individual Δt_i , $i = 1..M$. Por tanto, la salida desde el subsistema de generación de

subtítulos 202, y por consiguiente del sistema de subtitulación en directo 200 en su conjunto, es un número $M \geq 1$ de unidades de subtítulo 212 en el que cada cual comprende su respectivo retardo individual Δt_i e información de formateo 203.

5 El número de unidades de subtítulo 212 creadas a partir de un grupo de bloques de transcripción 211 depende de la longitud de los fragmentos de texto T_{xj} y las marcas de tiempo t_{jB} , t_{jE} relacionados con cada bloque de transcripción T_j . Es necesario considerar qué tipo de relación existe entre los bloques de transcripción y los subtítulos con el fin de hacer un seguimiento del tiempo:

10 Relación 1-1: Un bloque de transcripción T_j obtenido en el subsistema de transcripción de audio 201 puede corresponder sólo a una unidad de subtítulo S_i en la salida del subsistema de generación de subtítulos 202.

Relación 1-M: si un bloque de transcripción T_j es muy largo, puede corresponder también a dos o más subtítulos.

15 Relación N-1: si los bloques de transcripción son muy cortos y las marcas de tiempo están muy próximas en tiempo, es posible crear una unidad de subtítulo S_i a partir de más de un bloque de transcripción.

Relación N-M: en general, dependiendo de la longitud y las marcas de tiempo, un conjunto de M unidades de subtítulo 212 puede generarse a partir de un grupo de N bloques de transcripción 211, siendo o bien $N \geq M$ o bien $N < M$.

20 El subsistema de generación de subtítulos 202 puede ser completamente automático, semiautomático o completamente manual. Esto significa que este subsistema puede crear automáticamente el texto del subtítulo y toda la información incluida en unidades de subtítulo 212 sin intervención humana en la edición o formateo (usando NLP, por ejemplo). En las otras implementaciones posibles, el subsistema de generación de subtítulos 202 crea
25 automáticamente las unidades de subtítulo 212 mientras que una persona las monitoriza y eventualmente realiza cambios, edición, formateo, etc. O bien la persona puede tener toda la responsabilidad en la creación del conjunto de unidades de subtítulo 212 a partir de los bloques de transcripción 211, dividir los subtítulos, definir la duración, etc. En cualquier caso, con el fin de calcular cada retardo de subtítulo individual Δt_i , es necesario medir el tiempo empleado en
30 este subsistema, desde el momento en el que el primer bloque de transcripción T_1 de un grupo de N bloques de transcripción (usado para crear un conjunto de M unidades de subtítulo) está disponible hasta que la primera unidad de subtítulo S_1 de las M unidades de subtítulo (creadas a partir de estos N bloques de transcripción) se suministra a las fases de segmentación en paquetes y transmisión/difusión. Este tiempo se muestra en las figuras 3 y 4, indicado como
35 Δt_{SGi} , que indica el retardo de generación de subtítulos de la unidad de subtítulo S_i . El tiempo total empleado en la creación de la unidad de subtítulo S_i , dentro de un conjunto de unidades de subtítulos, a partir de su(s) fragmento(s) de audio j correspondiente(s) es Δt_i , calculado como la diferencia entre el tiempo de entrega t_{sub_1} de la primera unidad de subtítulo S_1 y el tiempo en

el que se produjo el primer fragmento de audio j correspondiente t_{1B} : $\Delta t_i = t_{sub_1} - t_{1B}$. Este retardo de subtítulo Δt_i puede usarse en combinación con la marca de tiempo de presentación MPEG para permitir la sincronización de audio y subtítulo.

La figura 3 muestra ejemplos del marcado de tiempo en el subsistema de transcripción de audio 201 y en el subsistema de generación de subtítulos 202 para los siguientes casos de relación entre el número N de bloques de transcripción 211 y el número M de unidades de subtítulo 212:

(A) relación 1-1, hay una unidad de subtítulo S_1 que corresponde a un único bloque de transcripción T_1 ;

(B) relación N-1, en el ejemplo de la figura 3 (B) $N=2$;

(C) relación 1-M, en el ejemplo de la figura 3 (C) $M=2$.

La figura 4 muestra ejemplos del marcado de tiempo en el subsistema de transcripción de audio 201 y en el subsistema de generación de subtítulos 202 para los siguientes casos de relación entre el número N de bloques de transcripción 211 y el número M de unidades de subtítulo 212:

(D) relación N-M, $N < M$, en el ejemplo de la figura 4 (D) $N=2$, $M=3$;

(E) relación N-M, $N > M$, en el ejemplo de la figura 4 (E) $N=3$, $M=2$.

En todos los casos, el retardo para cada unidad de subtítulo individual generada a partir de un conjunto de bloques de transcripción es el mismo. Su valor Δt_1 se calcula como la diferencia entre el tiempo en el que se crea la primera unidad de subtítulo en 202 y el tiempo en el que comenzó el fragmento de audio correspondiente al primer bloque de transcripción. La fórmula para el primer subtítulo es: $\Delta t_1 = t_{sub_1} - t_{1B}$. En el caso, como el de las Figuras 3C, 4D y 4E, en el que se genera más de un subtítulo, el retardo para el segundo y sucesivos subtítulos es también el mismo: $\Delta t_2 = \Delta t_1 = t_{sub_1} - t_{1B}$ y $\Delta t_2 = \Delta t_3$ en el ejemplo de la Figura 4E. Las duraciones de subtítulo pueden calcularse según las normas de subtitulación o según otros criterios. En cualquier caso, las duraciones no se ven afectadas por el procedimiento de sincronización descrito en el presente documento y viceversa.

La estimación de las marcas de tiempo de comienzo y de finalización t_{jB} y t_{jE} de cada bloque de transcripción T_j es crítica para el posterior cálculo del retardo de tiempo de subtítulo Δt_i de una unidad de subtítulo S_i con respecto al (a los) fragmento(s) de audio j correspondiente(s). La figura 5 y la figura 6 muestran estas marcas de tiempo t_{jB} y t_{jE} para cada bloque de transcripción T_j creado por el subsistema de transcripción de audio 201.

En un subsistema de transcripción de audio 201 completamente automático, en el que el proceso de transcripción se basa en ASR sin reahablado, el software de ASR que realiza todo el proceso de transcripción (sin intervención humana) puede proporcionar información acerca del tiempo exacto que corresponde a la transcripción de audio (además de la propia transcripción); por tanto, resulta innecesario estimar el tiempo empleado en esta parte del proceso puesto que se dispone de valores exactos de t_{jB} y t_{jE} como se muestra en la figura 5.

En este caso, las marcas de tiempo t_{jB} y t_{jE} no son medidas de retardo sino referencias de tiempo t_j que indican cuándo el fragmento de audio j transcrito se produjo con respecto al reloj de referencia 100, por tanto: $t_{jB} = t_j$ y $t_{jE} = t_j + d_j$, donde d_j es la duración del fragmento de audio j .

5 Cuando el proceso de transcripción es manual o semiautomático como en un subsistema de transcripción de audio 201, basado en una persona pulsando teclas en una máquina de estenografía, en un teclado convencional tipo ordenador o en un teclado silábico, o basado en rehablado usando un software de ASR (usado en RTVE en España, por ejemplo), hay una fase de procesamiento de escucha y mental 600 realizada por una persona seguida de un proceso de transcripción (que puede implicar incluso una traducción idiomática) realizada
10 también por una persona aunque usando dispositivos o software especiales (ASR en rehablado, estenografía en estenotipia, etc.), completando así la generación de los bloques de transcripción 211, que son la salida del subsistema de transcripción de audio 201. El bloque de transcripción T_j contiene el texto de los fragmentos de audio transcritos y las marcas de tiempo, que representan el tiempo t_{jB} y t_{jE} de comienzo y de final respectivamente del bloque de
15 transcripción T_j , que se estiman según lo siguiente:

El hecho de que haya retardos variables desde la escucha a la transcripción que dependen de la persona (retardo de procesamiento mental) y el fragmento de audio particular, y que sean difíciles de medir con precisión, requiere usar un algoritmo que proporcione una estimación de retardos en la mayoría de los escenarios. En otros escenarios, puede usarse una
20 entrada adicional proporcionada por el operador, para hacer un seguimiento del inicio de cada fragmento de audio por medio de una herramienta auxiliar (pedal, instrucción de voz...) para registrar digitalmente y numerar el fragmento. Por otro lado, el proceso de transcripción (ASR, estenografía u otros) también introduce a su vez un retardo aunque puede obtenerse a partir de las marcas de tiempo th_{jB} y th_{jE} exactos dados por el subsistema de transcripción de audio 201,
25 mostrado en la figura 6, como se explica posteriormente.

En el escenario representado en la figura 6, el resultado de la etapa de "escucha y procesamiento mental" 600 sobre el audio original 610 es la transcripción humana (por ejemplo audio del rehablado, tecleo por estenografía, etc.) 611 que se introduce al "proceso de transcripción" 501 (por ejemplo ASR para el rehablado, software de estenotipia para el tecleo
30 de estenografía, etc.) con el fin de generar un conjunto de bloques de transcripción aunque sin tener en cuenta todavía el retardo de procesamiento mental Δt_{mpj} . El proceso de transcripción 501 de la transcripción humana proporciona un texto transcrito y dos marcas de tiempo th_{jB} y th_{jE} . La siguiente etapa 601 es la estimación del retardo de procesamiento mental Δt_{mpj} que va a añadirse a las dos marcas de tiempo th_{jB} y th_{jE} exactos asociados con los bloques 612
35 transcritos con el fin de obtener una estimación de las marcas de tiempo de comienzo y finalización, t_{jB} y t_{jE} , de los bloques de transcripción 211. La estimación de t_{jB} y t_{jE} se denomina ts_{jB} y ts_{jE} respectivamente. De ahí:

$$ts_{jB} = th_{jB} + \Delta t_{mpj}$$

$$ts_{jE} = th_{jE} + \Delta t_{mpj}$$

Es necesario resaltar que en el caso ilustrado por la figura 6 las marcas de tiempo de los bloques de transcripción 211 no son retardos sino referencias de tiempo para el comienzo y la finalización de los fragmentos de audio transcritos con respecto al reloj de referencia de tiempo real 100. Como los valores exactos de th_{jB} y th_{jE} pueden obtenerse mediante el software o dispositivo correspondiente usado por el rehablador (o estenógrafo, etc.), la parte crítica es calcular el retardo de procesamiento mental Δt_{mpj} introducido por el rehablador (o estenógrafo, etc.). Aunque no se considerase el retardo de procesamiento mental Δt_{mpj} , hacer sólo un seguimiento de th_{jB} y th_{jE} daría como resultado una mejora importante en la sincronización de subtítulos. Sin embargo, para una mejor sincronización, es necesario diseñar un algoritmo para estimar el tiempo empleado en la etapa de "escucha y procesamiento mental" 600. Cuanto mejor sea dicha estimación, mejor será la aproximación de los tiempos de comienzo y finalización, t_{jB} y t_{jE} , a las referencias de tiempo reales, es decir: $ts_{jB} \approx t_{jB}$ y $ts_{jE} \approx t_{jE}$

El algoritmo para estimar el retardo de procesamiento mental Δt_{mpj} depende de qué tipo de transcripción mental se lleva a cabo. Se producen dos escenarios principales en la etapa de "escucha y procesamiento mental" 600: (a) una transcripción literal o casi literal en la que hay una transcripción continua tal como ocurre en estenografía. (b) una transcripción con periodos de escucha más largos aunque menos transcripción literal (transcripción discreta tal como ocurre en un sistema de subtitulación de rehablado). Para transcripciones mentales continuas, el retardo de procesamiento mental Δt_{mpj} es corto y casi constante; por tanto, puede considerarse un parámetro configurable, asignado a la persona según sus capacidades. En transcripciones discretas, el retardo de procesamiento mental Δt_{mpj} es más largo y con una gran variación que depende del grado de interpretación/resumen. Sin embargo, en cualquier escenario, el retardo de procesamiento mental Δt_{mpj} puede estimarse usando un algoritmo general: Δt_{mpj} es directamente proporcional a la longitud del texto transcrito e inversamente proporcional al grado de literalidad de la transcripción. Matemáticamente este algoritmo de estimación de retardo de procesamiento mental puede expresarse como:

$$\Delta t_{mpj} = \text{TextLength} \cdot (K/\mu) + C$$

donde dicho retardo Δt_{mpj} se estima en segundos; TextLength es la longitud de transcripción del fragmento de audio j (en unidades de longitud, u.l.); K es la duración de una unidad de longitud (en segundos / u.l.); C es una constante de tiempo (en segundos); μ es una variable de grado de literalidad ($0 \leq \mu \leq 1$) que representa el grado de parecido a la literalidad de la transcripción (por ejemplo, 1 significa el 100% de literalidad; 0,5 significa el 50%, de modo que de 2 palabras o letras pronunciadas en el audio, la persona que transcribe las resume en 1; 0,25 significa que de 4 palabras/letras, transcribe sólo 1, y así sucesivamente).

El parámetro μ debe ajustarse para reflejar las capacidades de la persona que transcribe y también depende del tipo de programa de televisión o escenario en directo, etc. Por ejemplo, la longitud de la transcripción TextLength puede medirse en palabras o en

caracteres, y K en segundos/palabra o segundos/carácter. La multiplicación de TextLength y K da como resultado la duración estimada del fragmento de audio. La constante C representa el tiempo empleado en escuchar y procesar que no está relacionado con la duración del audio y se usa como tiempo fijo que también depende de las capacidades humanas.

5 Para un escenario de transcripción mental continua, pueden considerarse las siguientes suposiciones:

$K \cdot \text{TextLength}$, es un tiempo constante fijado como parámetro (por ejemplo, 2 segundos)

μ es igual a 1 (transcripción literal)

10 C es una constante de tiempo (por ejemplo, 1 segundo).

Por tanto, el retardo de procesamiento mental Δt_{mpj} es una constante (en este ejemplo, 3 segundos).

Para un escenario de transcripción mental discreta, los valores son diferentes, por ejemplo:

15 K puede ser 0,33 a 0,5 segundos/palabra ó 0,8 segundos/carácter aproximadamente.

TextLength_j es una variable que depende del bloque de transcripción.

μ , puede ser 0,5 como promedio.

C puede ser 1 segundo.

20 Un flujo de transporte MPEG contiene flujos elementales para audio, vídeo y subtítulos. Los flujos elementales (PES) de audio, vídeo y datos se codifican en paquetes de flujo elemental (paquetes PES) que son las unidades de transmisión dentro de un flujo de transporte. Cuando se convierte una señal de audio en un PES cada fragmento de audio j mencionado anteriormente se codifica según la norma ISO 13818-3 y se empaqueta en al
25 menos un paquete PES del flujo elemental de audio correspondiente, junto con la información de sincronismo conocida como marca de tiempo de presentación (PTS). Cada paquete PES de audio contiene una PTS que se refiere al tiempo de presentación del paquete de audio cuando se reproduce en un receptor de usuario final. Un fragmento de audio j puede transportarse en varios paquetes PES de audio. De manera similar la señal de vídeo, cuando está presente, se
30 transporta en paquetes PES dentro de un flujo elemental de vídeo que también incluye información de sincronismo de PTS. Es importante resaltar que PTS se refiere al tiempo en el que se produce la segmentación en paquetes.

Los subtítulos también se codifican como paquetes dentro de un flujo elemental de subtítulos donde se codifica cada unidad de subtítulo i . Se prevén diferentes procedimientos de
35 codificación para los subtítulos en un flujo de transporte MPEG/DVB, por ejemplo subtitulación DVB-Sub y de teletexto. Como resultado del procedimiento de codificación, una unidad de subtítulo i puede dividirse en uno o más paquetes PES del flujo elemental de subtítulos, conteniendo cada uno, además de los datos de la unidad de subtítulo (texto, color, posición en

pantalla,...), una PTS como en los casos de audio o vídeo.

El procedimiento de codificación usado para subtítulos DVB-Sub, definido en ETSI EN 300 743, establece que una unidad de subtítulo se codifica en un "conjunto de visualización" y que un conjunto de visualización está compuesto de un conjunto de "segmentos de subtítulo".

5 Todos los segmentos de subtítulo de un conjunto de visualización se segmentan y transportan en paquetes PES que comparten la misma PTS. Esto implica que una unidad de subtítulo se transporta en uno o más paquetes PES del flujo de subtítulos.

10 Las realizaciones de la invención que se describen usan una extensión de la norma DVB/MPEG en la generación del flujo MPEG en el lado de emisión, para incluir la información de retardo variable calculada para cada subtítulo, como se explicó anteriormente, en los flujos elementales de DVB correspondientes de un canal de televisión: DVB-Sub para subtitulación y DVB-TXT para teletexto. Esta ampliación también soporta la información de control necesaria.

15 En las diferentes realizaciones de la invención ilustradas en las figuras 7-9, se usa un generador G1 de MPEG/DVB que soporta las ampliaciones requeridas que acepta como entradas:

a) Una o más señales de vídeo 61, si se requiere vídeo (por ejemplo, en un escenario de aplicación de canal de televisión),

b) Una o más señales de audio 62 (por ejemplo, de un canal de televisión o canal de radio), y

20 c) una secuencia de datos de subtítulado que incluye, para cada subtítulo que va a visualizarse en la pantalla, la información de retardo individual 64 de ese subtítulo (obtenida o estimada según el objetivo de la invención) con respecto al audio/vídeo correspondiente, junto con información de subtítulo típica tal como su texto 63, su tiempo de presentación, parámetros 65 tales como duración o tiempo de finalización, posición en la pantalla, color, etc.

25 El generador G1 de MPEG/DVB propuesto, según diferentes implementaciones de sincronización de subtitulación mostradas en las figuras 7-9, comprende medios de generación G2 de MPEG/DVB extendidos para considerar los retardos individuales de cada subtítulo, que puede componer, a partir de las entradas mencionadas anteriormente, un canal de televisión completo en formato MPEG/DVB cuya característica distintiva es que incluye los valores del retardo asociados a cada subtítulo. Para resaltar que este formato incluye las ampliaciones necesarias, se denomina en lo sucesivo en el presente documento MPEG/DVB++.

30 Las normas MPEG, DVB-Sub y DVB-TXT proporcionan un marco que puede ampliarse fácilmente para soportar la transmisión de estas extensiones. A continuación se dan detalles del formato y procedimiento para crear el flujo MPEG/DVB++. El flujo de salida de MPEG/DVB++ se compone de uno o más canales de televisión, consistiendo cada uno en:

- Flujos elementales para vídeo (normalmente un único flujo aunque pueden ser más flujos de vídeo)

- Flujos elementales para audio (uno o más flujos de audio elementales)

- Flujo(s) elemental(es) para DVB-Sub con información adicional según se detalla a continuación:

Según la norma ETSI EN 300 743 (Sistemas de subtítulos de difusión de vídeo digital), el elemento sintáctico básico de los flujos de subtítulos es el `subtitling_segment`. A efectos de la invención, se define un nuevo segmento de subtítulos de DVB que se incluye justo antes del segmento denominado `end_of_display_set_segment` en la norma DVB-Sub y que denominaremos `subtitle_delay_segment`, al que se le asigna uno de los valores `segment_type` que todavía no se haya asignado en la norma ETSI EN 300 743 v1.3.1 (los valores previstos se encuentran en el intervalo de 0x81 - 0xEF reservado para datos privados; otros valores son posibles usando el intervalo reservado para un uso futuro en la norma mencionada). El `subtitle_delay_segment` es un segmento adicional que contiene el tiempo de retardo asociado con un `display_set` creado (según la norma). Un `display_set` es el conjunto de segmentos de subtítulo de un servicio de subtítulo específico al que está asociado el mismo valor de marca de tiempo de presentación (PTS, según se define en la norma ISO/IEC 13818-1). Cada `display_set` contiene un `subtitle_delay_segment` compuesto por dos campos según se muestra en la tabla 1:

Tabla 1

Sintaxis	Tamaño	Tipo
<code>subtitle_delay_segment () {</code>		
<code>variable_subtitle_delay_time</code>	16 bits	uimsbf
<code>recommended_event_shift_time</code>	8 bits	uimsbf
<code>}</code>		

`variable_subtitle_delay_time`: es el retardo individual aplicable al `display_set` medido en unidades de 100 milisegundos.

`recommended_event_shift_time`: es el tiempo mínimo en segundos que debe retardarse el evento para hacer posible la compensación de los retardos de subtítulo individuales; se usa durante el proceso posterior sincronización de subtítulos; es configurable y normalmente permanece sin cambios durante un evento de televisión dado o incluso para un canal de televisión dado.

- Flujo(s) elemental(es) para subtítulos de teletexto con información incrustada de retardo variable e insertada para cada subtítulo, que puede incluirse, de manera similar al caso DVB-Sub, en el flujo de salida de MPEG/DVB++.

- Una tabla de mapa de programa (`Program_Map_Table` o PMT, según se define en la norma ISO/IEC 13818-1) para el canal de televisión que contiene un descriptor adicional, que denominaremos `synchronizable_subtitling_descriptor`, para la entrada correspondiente al flujo elemental DVB-Sub. Este descriptor se define a continuación según se muestra en la tabla 2:

Tabla 2

Sintaxis	T	
	amaño	ipo
<code>synchronizable_subtitling_descriptor () {</code>		
<code>descriptor_tag</code>	8 bits	uimsbf
<code>descriptor_length</code>	8 bits	uimsbf
<code>synchronizable_subtitle_stream_type</code>	8 bits	uimsbf
<code>recommended_event_shift_time</code>	8 bits	uimsbf
<code>}</code>		

5 `descriptor_tag` adopta uno de los valores en el intervalo de 0x80 a 0xFE especificado como definido por el usuario en la norma ETSI EN 300 468; otros valores son posibles usando el mecanismo de descriptor de extensión definido en la norma.

`descriptor_length` es otro campo de la norma anterior que especifica el número de bytes incluidos en los campos de datos inmediatamente siguientes.

10 `synchronizable_subtitle_stream_type`: es el nuevo campo definido para indicar el tipo de flujo de subtítulo sincronizable; especifica información adicional acerca de cómo usar y acceder a los retardos de subtítulo variables individuales transportados en el flujo de subtítulo.

`recommended_event_shift_time`: es el nuevo campo definido para indicar es el tiempo mínimo en segundos que el evento debe retardarse con el fin de adaptarse a los retardos de subtítulo individuales; este tiempo se aplica durante el proceso de sincronización de subtítulos.

15 El `synchronizable_subtitling_descriptor` mostrado en la Tabla 2 puede seguir normalmente, dentro de la PMT, al `subtitling_descriptor` existente. Además, el resto de tablas de información específica de programa (PSI), tales como la tabla de asociación de programa (PAT), la tabla de fecha y hora (TDT), la tabla de desplazamiento de tiempo (TOT), etc., así como las tablas de información de servicio (SI) tales como la tabla de descriptor de servicio (SDT), y la tabla de información de evento (EIT), etc., pueden incluirse en el flujo de salida de
20 MPEG/DVB++ en la forma habitual.

El flujo de salida del generador G1 de MPEG/DVB++ con la información de retardo variable incluida en los flujos de DVB-Sub y en los flujos elementales de Teletexto, está listo para usarse como entrada en un módulo de sincronización de subtítulos (701, 801, 901).

25 Existen diferentes alternativas de implementación y uso con diversas implicaciones: la figura 7 muestra el módulo de sincronización de subtítulos 701 implementado en el lado de recepción, mientras que las figuras 8 y 9 muestran el módulo de sincronización de subtítulos (801, 901) implementado en el lado de transmisión. La elección de la alternativa más adecuada debe tener en cuenta los requisitos en cuanto a la asignación de ancho de banda, limitaciones

regulatorias y restricciones de uso. En cualquier caso, el módulo de sincronización de subtítulos (701, 801, 901) proporciona un flujo de MPEG/DVB en el que se ha retardado el evento audiovisual una cantidad de tiempo igual al `recommended_event_shift_time` y en el que los flujos elementales de audio, vídeo y subtítulo están sincronizados (un flujo de este tipo lo denominaremos MPEG/DVBsync).

Puesto que los flujos elementales segmentados en paquetes (PES), especificados en la norma ISO/IEC 13818-1, correspondientes a flujos elementales de vídeo, audio y subtítulo, se someten a un proceso de puesta en cola/retardo combinado con una manipulación de PTS para alinear los paquetes de subtítulos a su tiempo objetivo en las señales de vídeo y audio (61, 62) originales, el proceso de sincronización realizado por el módulo de sincronización de subtítulos (701, 801, 901) propuesto puede describirse como un proceso de puesta en cola y recálculo que implementa lo siguiente:

```

Si (paquete_vídeo)      { esperar (REST) segundos; PTSi = PTSi + REST}
Si (paquete_audio)     { esperar (REST) segundos; PTSi = PTSi + REST }
Si (paquete_subtítulo) { si (REST - Δti) ≥ 0
                        {esperar (REST- Δti) segundos;
                         PTSi = (PTSi - Δti) + REST}
                        si no
                        {esperar (0) segundos;
                         PTSi = PTSi }
                        }

```

donde

Δt_i es igual al `variable_subtitle_delay_time`, que es el retardo variable individual para el subtítulo *i*.

REST es un tiempo de desplazamiento igual o mayor al `recommended_event_shift_time` que, en este ejemplo de implementación, es el valor del retardo mínimo requerido para compensar el retardo de subtitulación más desfavorable considerado. El retardo aplicado para sincronizar individualmente el paquete de subtítulo *i* con el audio (y vídeo y/o cualquier otra información sincronizable con el audio como, por ejemplo, un vídeo con lenguaje de signos) es $REST - \Delta t_i$ y, por tanto, $PTS_i = PTS_i - \Delta t_i + REST$. Si se reciben subtítulos con Δt_i mayor que REST, su retardo con respecto al audio y vídeo se compensa sólo parcialmente (REST segundos).

PTS_i es la marca de tiempo de presentación de cada paquete PES en MPEG-2/DVB; la marca de tiempo de presentación PTS_i de cada paquete *i* de vídeo, audio y subtítulo se lleva al presente aumentándolo en REST, con el fin de garantizar que cada PTS_i representa los pulsos de reloj en tiempo real actuales.

Todos estos valores (las marcas de tiempo PTS_i, los retardos variables Δt_i , y el valor

para REST), requeridos por el módulo de sincronización de subtítulos (701, 801, 901), se reciben desde el generador G1 de MPEG-2/DVB extendido como valores de datos dentro del flujo de salida de MPEG-2/DVB++.

5 Para generar un servicio sincronizado, es necesaria una memoria intermedia que pueda contener REST segundos de paquetes (un REST de 20 segundos, por ejemplo, requiere un tamaño de memoria intermedia de aproximadamente 12,5 Mbytes para un canal del flujo de transporte con Audio, Video y subtítulos).

10 Todos los parámetros del formato de flujo salida de MPEG-2/DVB++ propuesto se incluyen de modo que se garantiza plena compatibilidad con las normas MPEG/DVB y los receptores TDT actuales, lo que es indispensable para las alternativas de uso descritas a continuación.

15 Una posible realización de la invención se refiere a sincronización de subtítulos seleccionable por el usuario en el lado de recepción, según se muestra en la figura 7. A partir de las señales de vídeo y audio (61, 62) originales, el sistema de subtítulos en directo mejorado 200 suministra el retardo de subtítulo individual 64 junto con el subtítulo 63 y el generador G1 de MPEG/DVB extendido genera un flujo de MPEG/DVB++ 70 que incluye dichos retardos variables y datos de señalización relacionados con los subtítulos para una difusión o transmisión de canal de televisión. Los flujos de MPEG/DVB++ 70 pasan a través de una red N1, (por ejemplo, red de difusión TDT o red de transmisión de IPTV) hacia un receptor 702 (por ejemplo, TDT o IPTV) que soporta la sincronización de subtítulos como una opción de usuario (el usuario puede seleccionar si prefiere ver la versión con subtítulos sincronizados con el vídeo) por medio de un módulo de sincronización de subtítulos 701 según se explicó anteriormente.

25 A partir de los flujos de MPEG/DVB++ 70 recibidos, el módulo de sincronización de subtítulos 701 obtiene los MPEG/DVBsync 72 mencionados anteriormente en los que los flujos elementales de audio y vídeo del canal de televisión se sincronizan con los subtítulos, proporcionando entonces una reproducción ligeramente retardada (casi-directo) del evento audiovisual.

30 A partir del flujo de MPEG/DVB++ 70 transmitido, uno de los dos flujos de salida posibles (71, 72) en el receptor de TDT/IPTV es enviado al módulo de reproducción de televisión 704, según la selección del usuario 703:

i) el flujo de salida de MPEG/DVB++ 71 es una opción en directo real, plenamente compatible con los receptores convencionales, que consiste en audio/vídeo y subtítulos que no están sincronizados mediante el módulo de sincronización 701;

35 ii) el flujo MPEG/DVBsync 72 es una versión casi-directo generada opcionalmente en el lado de recepción si se selecciona el módulo de sincronización 701.

En el lado de recepción tienen lugar las siguientes acciones:

a) El receptor 702 (módulo decodificador de IPTV, decodificador TDT,...) reconoce en el

flujo de MPEG/DVB++ 70 entrante la existencia de la información adicional que permite al usuario ver un canal de televisión con subtítulos activados y sincronizados con los flujos de audio y vídeo;

5 b) El receptor 702 muestra al usuario la disponibilidad de este servicio de sincronización de subtítulos;

c) Tras la activación por parte del usuario, el receptor 702 empieza a funcionar para proporcionar una reproducción sincronizada y retardada para el canal de televisión o programa de televisión seleccionado. Dependiendo del receptor, existen diferentes alternativas para la implementación de los menús de control y usuario en el receptor.

10 Por tanto, con el fin de proporcionar al usuario las dos opciones i) e ii) en la recepción, los nuevos módulos decodificadores de TDT o IPTV tendrían que implementar el algoritmo de sincronización descrito, manejar el nuevo conjunto de descriptores y actuar en consecuencia. El comportamiento de los antiguos módulos decodificadores no se ve afectado porque ignoran los nuevos campos, descriptores, o tipos de flujo.

15 El comportamiento del receptor 702 para implementar sincronización de subtítulos se representa mediante los estados mostrados en la figura 10, en donde se representa la transición del estado normal 1000 al estado sincronizado y retardado 1002, pasando por un estado intermedio 1001 de transición entre ambos. Es decir, cuando el usuario selecciona ver un programa de televisión en un modo sincronizado y retardado está solicitando al receptor 702 exactamente lo siguiente: "detén la reproducción del vídeo y el audio en mi aparato de televisión hasta que puedas mostrarlos sincronizados con sus subtítulos".

20 Después de que el usuario habilita la sincronización de subtítulos y si la entrada de receptor consiste en un MPEG/DVB++ con $REST \neq 0$ (transición 1010 entre los estados 1000 y 1001), el receptor 702 entra en el estado de transición 1001 en el que el módulo decodificador de televisión deja de reproducir el programa de televisión (es decir, la imagen está congelada en la pantalla y el audio está detenido) aunque sigue recibiendo y almacenando en la memoria intermedia paquetes de MPEG a partir del flujo de transporte durante el tiempo de espera necesario, siendo el tiempo de espera $\geq REST$, para garantizar que las memorias intermedias de recepción contienen todos los paquetes para audio, vídeo y subtítulos (llegando éstos varios segundos más tarde) necesarios para una presentación sincronizada de subtítulos. Después de que el tiempo de espera ha transcurrido, el sistema entra en el estado sincronizado y retardado 1002.

30 Durante la transición 1011 del estado 1001 al 1002, todos los paquetes de subtítulos, para los que el recálculo de la marca de tiempo de presentación (PTS) da como resultado una PTS en el pasado, se muestran en una pantalla con la imagen congelada (y sin sonido) ya que estos paquetes contenían subtítulos pertenecientes a muestras de audio que ya han sido reproducidas cuando el usuario conmuta del estado normal 1000 al estado sincronizado y retardado 1002. Además, volver al modo normal (en directo real con subtítulos no

sincronizados) da como resultado una pérdida de varios segundos (el tiempo de espera dado) del programa que está difundiéndose. Esto se representa en la figura 10 mediante el estado de transición 1003 de sincronizado y retardado a normal entre las transiciones 1012 y 1013.

5 Ambos estados de transición 1001 y 1003 pueden implementarse en una única etapa (por ejemplo, esperando todo el tiempo de espera de una vez) o en etapas incrementales (por ejemplo, esperando varios segundos cada vez) hasta que el retardo requerido (de manera ideal REST) se almacene en la memoria intermedia. Tales implementaciones proporcionan una transición continua (por ejemplo, múltiples aunque más cortos periodos de congelación de imagen) aunque tardan más en alcanzar un estado final completamente sincronizado 1002 o en
10 volver al estado normal 1000.

El tiempo de espera puede configurarse con un valor similar a REST que se recibe como un parámetro para hacer frente al tiempo transcurrido más desfavorable en la preparación de subtítulos. Valores típicos conservadores son de aproximadamente 15 segundos para retardos de generación de subtítulos de reablado de ASR y de
15 aproximadamente 5 segundos para estenotipia. Cuando ha transcurrido el tiempo de espera, las memorias intermedias para los flujos de vídeo, audio y subtítulo implicados contienen suficientes paquetes para iniciar una reproducción sincronizada que tiene en cuenta los retardos de subtítulos individuales.

Siempre que esté disponible y se seleccione el modo de reproducción sincronizado y retardado 1002, todo el programa de televisión se reproduce con un pequeño retardo, y el receptor 702 recalcula la PTS de cada subtítulo para presentarlo en la pantalla en el mismo momento que el fragmento de audio original. El modo de reproducción sincronizado y retardado
20 92 finaliza o bien automáticamente, cuando la entrada al reproductor de televisión ya no contiene datos de sincronización (por ejemplo REST=0, o cuando la entrada ya no es MPEG-2/DVB++), o bien manualmente cuando se desactivan los subtítulos o tras una selección por parte del usuario que deshabilita explícitamente esta opción. Son posibles diferentes implementaciones según preferencias definidas por el usuario. El reproductor de televisión, o bien finaliza la presentación de todos los subtítulos almacenados en la memoria intermedia sincronizados con audio/vídeo antes de que el receptor 702 cambie al estado normal 1000, o
25 30 bien pasa al estado normal inmediatamente cuando es el usuario quien solicita volver al modo de reproducción normal. Ambas son opciones de implementación/configuración en el receptor.

Esta implementación de la opción de sincronización de subtítulos en el receptor es adecuada, por ejemplo, para la difusión de televisión terrestre o por satélite, ya que no requiere ancho de banda adicional, aunque limita la oferta de reproducción con subtítulos sincronizados
35 a usuarios que tienen los nuevos módulos decodificadores con el receptor 702 mejorado según se muestra en la figura 7. Los receptores que no implementan la opción son completamente compatibles con esta realización de la invención aunque sólo pueden presentar el evento como en la actualidad, es decir, en directo de manera no sincronizada.

Otra posible realización de la invención se refiere a la sincronización de subtítulos con el audio y vídeo correspondiente en el lado de emisión, antes de que tenga lugar la transmisión/difusión de televisión, según se muestra en la figura 8. Se genera un canal de televisión en casi-directo CH1 con subtítulos sincronizados 81, de modo que los eventos audiovisuales pueden transmitirse o difundirse varios segundos más tarde al receptor 802. La difusión/transmisión del canal de televisión en casi-directo CH1 con subtítulos sincronizados no requiere ancho de banda adicional y puede ser adecuado para difundir eventos para los que una difusión en directo real no implica una ganancia en instantaneidad (por ejemplo lectura de un informe redactado hace 2 horas). Sin embargo, esto puede ser ilegal para la difusión de TDT de algunos eventos (al menos en algunos países). Un canal adicional CH2 para ofrecer también las versiones no sincronizadas originales en directo real 82 es también una posibilidad según se muestra en la figura 8. Las dos alternativas básicas mostradas en la figura 8 tienen en común el formato MPEG/DVB++ 80 proporcionado por el Generador MPEG/DVB extendido G1, a partir del cual el módulo de sincronización de subtítulos 801 implementado antes de la red de transmisión/difusión N1 genera la versión sincronizada de subtítulos, a la que se hace referencia como MPEG/DVBsync 81. La ventaja principal de la sincronización de subtítulos antes de la difusión es su plena compatibilidad con receptores de TDT existentes o módulos decodificadores de IPTV. El uso simultáneo de dos canales, CH1 para la difusión/transmisión en casi-directo y CH2 para la reproducción del mismo evento en directo real, es adecuado cuando está disponible el ancho de banda adicional requerido para el canal de televisión adicional (una alternativa real para IPTV o televisión por Cable/Fibra). En este caso, puesto que el canal CH1 sólo es lógico seleccionarlo cuando se desean subtítulos sincronizados, el uso de subtítulos en abierto es también una alternativa de implementación práctica.

Una posible realización adicional de la invención se refiere al uso de la sincronización de subtítulos en la redifusión de eventos o programas que fueron subtitulados en directo. Esto es posible en cualquiera de las dos opciones que para el lado de emisión se han descrito en anteriormente, y que se muestran con líneas discontinuas en la figura 9. Por tanto, a partir de la generación del formato MPEG/DVB++ 90 mediante el Generador MPEG/DVB extendido G1, el almacenamiento de un evento audiovisual de televisión puede realizarse en paralelo a estas dos opciones: o bien almacenar subtítulos sincronizados (en un soporte de almacenamiento multimedia BD1 usando la salida del módulo de sincronización de subtítulos 901) proporcionados desde el MPEG/DVBsync 91 para el canal de televisión casi en directo CH1 o versiones no sincronizadas 92 de eventos audiovisuales subtitulados en directo (en otro soporte de almacenamiento multimedia BD2) del canal de televisión en directo real CH2. La difusión o distribución mediante cualquier medio (DVD, descarga de archivos, televisión por web, etc.) de un evento audiovisual de televisión almacenado previamente, con sus subtítulos sincronizados, puede llevarse a cabo, o bien directamente si se almacena el MPEG/DVBsync 91, o bien requerir una etapa de sincronización previa realizada por el módulo de

sincronización de subtítulos 901 implementado en el lado de emisión antes de la transmisión/difusión.

Una de las ventajas de esta realización es que, incluso si se toma la decisión de difundir un evento subtulado en directo sin aplicar ningún proceso de sincronización a sus subtítulos, la sincronización puede ser aplicada posteriormente de manera sencilla y sin costes adicionales, según se muestra en la figura 11, para su redifusión (o distribución de DVD, transmisión de televisión por web,...) con audio/vídeo y subtítulos sincronizados.

Para la redifusión o distribución diferida de eventos audiovisuales subtitulados en directo puede utilizarse una versión sincronizada 91 MPEG/DVBSync almacenada en un soporte de almacenamiento multimedia BD1, o una versión no sincronizada 92 MPEG/DVB++ que contiene la información necesaria para ser sincronizada antes de la redifusión mediante la aplicación del módulo de sincronización 901.

Un sistema de subtitulación en directo que implementa el procedimiento de sincronización descrito anteriormente puede proporcionar toda la información relacionada con subtítulos (duración, color, posición,...) junto con sus retardos individuales, a un módulo de generación de MPEG/DVB. Según se muestra en la figura 9, almacenar esta información en este punto en un soporte de almacenamiento multimedia BD3 también podría usarse para facilitar la edición de subtítulos antes de su posible redifusión) sin perder las capacidades de sincronización.

Todas las opciones anteriores, descritas para canales de televisión, pueden aplicarse directamente a implementaciones de canales de radio digital subtitulados. El subtítulo puede ofrecerse en aparatos de televisión u otros dispositivos.

Las opciones de implementación de sincronización de subtítulos mostradas en el presente documento resaltan el valor de usar el proceso y formato propuestos (es decir, MPEG/DVB con retardos de subtítulo individuales) para proporcionar una fuente común que soporta alternativas de difusión en directo real, casi- directo y redifusión.

Las funcionalidades proporcionadas por las realizaciones descritas de la invención pueden aplicarse por diferentes entidades u organismos para proporcionar nuevas prestaciones de la siguiente manera:

- Por operadores de televisión para difundir eventos o canales audiovisuales en los que la subtitulación en directo puede sincronizarse opcionalmente en el receptor.
- Por operadores de televisión para difundir eventos o canales audiovisuales en los que la subtitulación en directo se sincroniza para ofrecer una reproducción en casi-directo del evento.
- Por operadores de televisión para almacenar programas de televisión subtitulados en directo que van a usarse en futuras transmisiones sincronizadas de IPTV grabadas

previamente, redifusiones de TDT u otros medios de distribución.

- Por emisoras de radio digital para ofrecer subtítulos sincronizados

- Fabricantes de receptores de TDT externos o integrados y receptores de IPTV para incluir la sincronización de subtítulos como una opción seleccionable por parte del usuario en
5 futuros receptores.

- Empresas, instituciones y desarrolladores de capacidades de receptores de TDT para incluir la sincronización de subtítulos en cualquiera de sus productos.

- En general, organismos, instituciones y empresas que desean ofrecer un servicio que soporte la subtitulación en directo de eventos que pueden reproducirse con audio, vídeo y
10 subtítulos sincronizados.

Obsérvese que, en este texto, el término "comprende" y sus derivaciones (tales como "comprendiendo", etc.) no deben entenderse en un sentido exclusivo, es decir, estos términos no deben interpretarse como que excluyen la posibilidad de que lo que se describe y define pueda incluir elementos, etapas, etc., adicionales.

REIVINDICACIONES

1. Procedimiento para sincronizar subtítulos con audio en sistemas de subtitulación en directo, que comprende:
 - 5 - generar al menos un subtítulo i a partir de al menos un bloque de transcripción T_j que corresponde a un fragmento de audio j ;
caracterizado porque comprende además:
 - determinar un retardo de subtítulo individual Δt_i de cada subtítulo i ;
 - crear una unidad de subtítulo S_i que comprende el subtítulo i y el retardo de
10 subtítulo individual Δt_i ;
 - poner cada unidad de subtítulo S_i junto con un retardo de desplazamiento asociado configurable REST en al menos un paquete de subtítulo PES, el cual tiene una marca de tiempo de presentación PTS_i , formando dichos paquetes de subtítulo PES parte de un flujo de transporte extendido que incluye al menos un paquete de
15 audio que contiene el fragmento de audio j y una marca de tiempo de presentación PTS_j .
 - sincronizar paquetes de subtítulo PES con paquetes de audio creando un flujo de transporte sincronizado que comprende todos los paquetes del flujo de transporte extendido retardados un tiempo que depende del retardo de
20 desplazamiento asociado configurable REST.
2. Procedimiento según la reivindicación 1, en el que la sincronización de paquetes de subtítulo PES con paquetes de audio comprende retardar cada paquete de subtítulo PES asociándole una marca de tiempo de presentación en el flujo de transporte sincronizado que es igual a $PTS_i - \Delta t_i + REST$, y retardar cada paquete de audio asociándole una marca de tiempo de presentación en el flujo de transporte sincronizado que es igual a $PTS_j + REST$, siendo PTS_i y PTS_j las marcas de tiempo de presentación en el flujo de transporte extendido y REST el retardo de desplazamiento asociado del
25 flujo de transporte extendido.
3. Procedimiento según la reivindicación 2, en el que, si el flujo de transporte extendido
30 incluye al menos un paquete de vídeo o al menos un paquete de datos de cualquier clase de datos sincronizables con paquetes de audio, la sincronización de paquetes de subtítulo comprende además retardar todos los paquetes de vídeo y de datos un tiempo que depende del retardo de desplazamiento asociado configurable REST.
4. Procedimiento según cualquiera de las reivindicaciones 1 a 3, en el que la
35 sincronización de paquetes de subtítulo con paquetes de audio se realiza antes de transmitir el flujo de transporte sincronizado a un dispositivo de recepción digital.

5. Procedimiento según la reivindicación 4, que comprende recibir el flujo de transporte sincronizado en el dispositivo de recepción digital, que es un receptor convencional seleccionado de un receptor de radiodifusión digital, un módulo decodificador de televisión digital terrestre, un receptor de televisión por IP, un receptor de difusión multimedia digital y un receptor de medios digital con medios de control para televisión en directo.
5
6. Procedimiento según cualquiera de las reivindicaciones 1 a 3, en el que la sincronización de paquetes de subtítulo con paquetes de audio se realiza después de recibir el flujo de transporte extendido en un dispositivo de recepción digital.
- 10 7. Procedimiento según la reivindicación 6, que comprende además permitir a un usuario elegir si la sincronización de paquetes de subtítulo con paquetes de audio es realizada por el dispositivo de recepción digital o no.
8. Procedimiento según cualquiera de las reivindicaciones 6 a 7, que comprende recibir el flujo de transporte extendido en el dispositivo de recepción digital, que se selecciona de
15 un receptor de radiodifusión digital, un módulo decodificador de televisión digital terrestre, un receptor de televisión por IP, un receptor de difusión multimedia digital y un receptor de medios digital con medios de control para televisión en directo.
9. Procedimiento según cualquiera de las reivindicaciones 1 a 3, en el que la sincronización de paquetes de subtítulo con paquetes de audio se realiza antes de
20 almacenar el flujo de transporte sincronizado en un soporte de almacenamiento multimedia.
10. Procedimiento según cualquiera de las reivindicaciones 1 a 3, en el que la sincronización de paquetes de subtítulo con paquetes de audio se realiza después de almacenar el flujo de transporte extendido en un soporte de almacenamiento multimedia y antes de transmitir el flujo de transporte sincronizado a una red de distribución multimedia, que se selecciona de una red de IP, una red de difusión de televisión digital terrestre y una red de radiodifusión digital.
- 25 11. Procedimiento según cualquier reivindicación anterior, en el que la determinación de un retardo de subtítulo individual Δt_i de cada subtítulo i comprende:
30 - determinar una marca de tiempo de comienzo t_{jB} que es el momento en el que comienza el primer fragmento de audio correspondiente al primer bloque de transcripción T_j de un grupo de bloques de transcripción, el grupo formado por al menos un bloque de transcripción y creado por un subsistema de transcripción de audio 201, y a partir del cual se genera al menos una unidad de subtítulo S_i mediante un subsistema de generación de subtítulos 202, con respecto a un reloj de referencia 100, usándose el reloj de referencia 100 tanto por el subsistema de generación de subtítulos 202 como
35

por el subsistema de transcripción de audio 201;

- determinar un tiempo de entrega t_{sub_i} que es cuando se entrega el primer subtítulo generado a partir de dicho grupo de bloques de transcripción por el subsistema de generación de subtítulos 202;

5 - restar la marca de tiempo de comienzo t_{jB} de dicho primer fragmento de audio del tiempo de entrega t_{sub_i} para calcular el retardo de subtítulo individual $\Delta t_i = t_{sub_i} - t_{jB}$;

- asignar dicho retardo de subtítulo individual Δt_i a cada unidad de subtítulo S_i creada a partir de dicho grupo de bloques de transcripción.

12. Procedimiento según la reivindicación 11, que comprende además determinar una
10 marca de tiempo de finalización t_{jE} que es el momento en el que finaliza el primer fragmento de audio correspondiente al primer bloque de transcripción T_j del grupo de bloques de transcripción, con respecto al reloj de referencia 100.

13. Procedimiento según la reivindicación 12, en el que el subsistema de transcripción de
15 audio 201 usa una transcripción mental por parte de una persona y la marca de tiempo de comienzo t_{jB} y la marca de tiempo de finalización t_{jE} se estiman respectivamente como $t_{jB} = th_{jB} + \Delta t_{mpj}$ y $t_{jE} = th_{jE} + \Delta t_{mpj}$ calculando un retardo de procesamiento mental Δt_{mpj} que depende del tipo de transcripción mental, que se selecciona de transcripción mental continua y transcripción mental discreta, th_{jB} es un instante de tiempo marcado cuando la persona comienza la transcripción mental y th_{jE} es un instante de tiempo
20 marcado cuando la persona finaliza la transcripción mental, ambos tiempos dados automáticamente por el subsistema de transcripción de audio 201 con respecto al reloj de referencia 100.

14. Procedimiento según la reivindicación 13, en el que el retardo de procesamiento mental Δt_{mpj} es una constante fijada por el subsistema de transcripción de audio 201 cuando
25 usa transcripción mental continua.

15. Procedimiento según la reivindicación 13, en el que el retardo de procesamiento mental Δt_{mpj} se calcula por el subsistema de transcripción de audio 201 cuando usa
30 transcripción mental discreta como $\Delta t_{mpj} = \text{TextLength}_j \cdot (K/\mu) + C$, siendo TextLength_j una longitud en unidades de longitud del fragmento de audio j , K es la duración de tiempo de una unidad de longitud; C es una constante de tiempo configurable en el subsistema de transcripción de audio 201, y μ es una variable de grado de literalidad, $0 \leq \mu \leq 1$, configurable y que representa la proximidad de la transcripción mental a la literalidad.

16. Dispositivo sincronizador de subtítulos, **caracterizado porque** comprende medios de
35 procesamiento configurados para:

- crear un flujo de transporte extendido a partir de un flujo de transporte inicial que

comprende todos los paquetes del flujo de transporte inicial, los cuales tienen una marca de tiempo de presentación asociada, comprendiendo el flujo de transporte extendido:

al menos un paquete de audio que contiene un fragmento de audio j ,

5 al menos una unidad de subtítulo S_i que comprende un subtítulo i generado a partir de al menos un bloque de transcripción T_j que corresponde al fragmento de audio j y un determinado retardo de subtítulo individual Δt_i asociado a cada subtítulo i ,

10 y un retardo de desplazamiento asociado $REST$ que es configurable e indica un retardo fijo para sincronizar paquetes de subtítulo con paquetes de audio,

- ajustar la marca de tiempo de presentación asociado de cada paquete de subtítulo del flujo de transporte extendido asociándole una marca de tiempo de presentación en el flujo de transporte sincronizado que es igual a $PTS_i - \Delta t_i + REST$, y retardar cada paquete de audio asociándole una marca de tiempo de presentación en el flujo de transporte sincronizado que es igual a $PTS_j + REST$, siendo PTS_i y PTS_j unas marcas de tiempo de presentación asociadas en el flujo de transporte extendido al paquete de subtítulo y el paquete de audio respectivamente y $REST$ el retardo de desplazamiento asociado en el flujo de transporte extendido.

17. Dispositivo sincronizador de subtítulos según la reivindicación 16, en el que, si el flujo de transporte inicial incluye al menos un paquete de vídeo o al menos un paquete de datos de cualquier clase de datos sincronizables con paquetes de audio, los medios de procesamiento están configurados para retardar todos los paquetes de vídeo y de datos el tiempo de desplazamiento global.

18. Dispositivo sincronizador de subtítulos según cualquiera de las reivindicaciones 16 a 17, en el que los medios de procesamiento están configurados para crear el flujo de transporte sincronizado antes de transmitir dicho flujo a un dispositivo de recepción digital.

19. Dispositivo sincronizador de subtítulos según cualquiera de las reivindicaciones 16 a 17, en el que los medios de procesamiento están configurados para permitir a un usuario elegir si la creación del flujo de transporte sincronizado se realiza después de recibir el flujo de transporte extendido en un dispositivo de recepción digital.

20. Dispositivo sincronizador de subtítulos según la reivindicación 19, que está integrado en el dispositivo de recepción digital seleccionado de un receptor de radiodifusión digital, un módulo decodificador de televisión digital terrestre, un receptor de televisión por IP, un receptor de difusión multimedia digital y un receptor de medios digital con medios de control para televisión en directo.

21. Dispositivo sincronizador de subtítulos según cualquiera de las reivindicaciones 16 a 17, en el que los medios de procesamiento están configurados para crear el flujo de transporte sincronizado antes de almacenar dicho flujo en un soporte de almacenamiento multimedia.
- 5 22. Dispositivo sincronizador de subtítulos según cualquiera de las reivindicaciones 16 a 17, en el que los medios de procesamiento están configurados para crear el flujo de transporte sincronizado a partir de un flujo de transporte extendido almacenado previamente en un soporte de almacenamiento multimedia y antes de transmitir el flujo de transporte sincronizado a una red de distribución multimedia, que se selecciona de
10 una red de IP, una red de difusión de televisión digital terrestre y una red de radiodifusión digital.
23. Un producto de programa informático que comprende medios de código de programa que, cuando se cargan en un procesador de propósito general, un procesador de señal digital, circuitos integrados de aplicación específica o cualquier dispositivo lógico
15 programable, hace que dichos medios de código de programa ejecuten el procedimiento según cualquiera de las reivindicaciones 1 a 15.

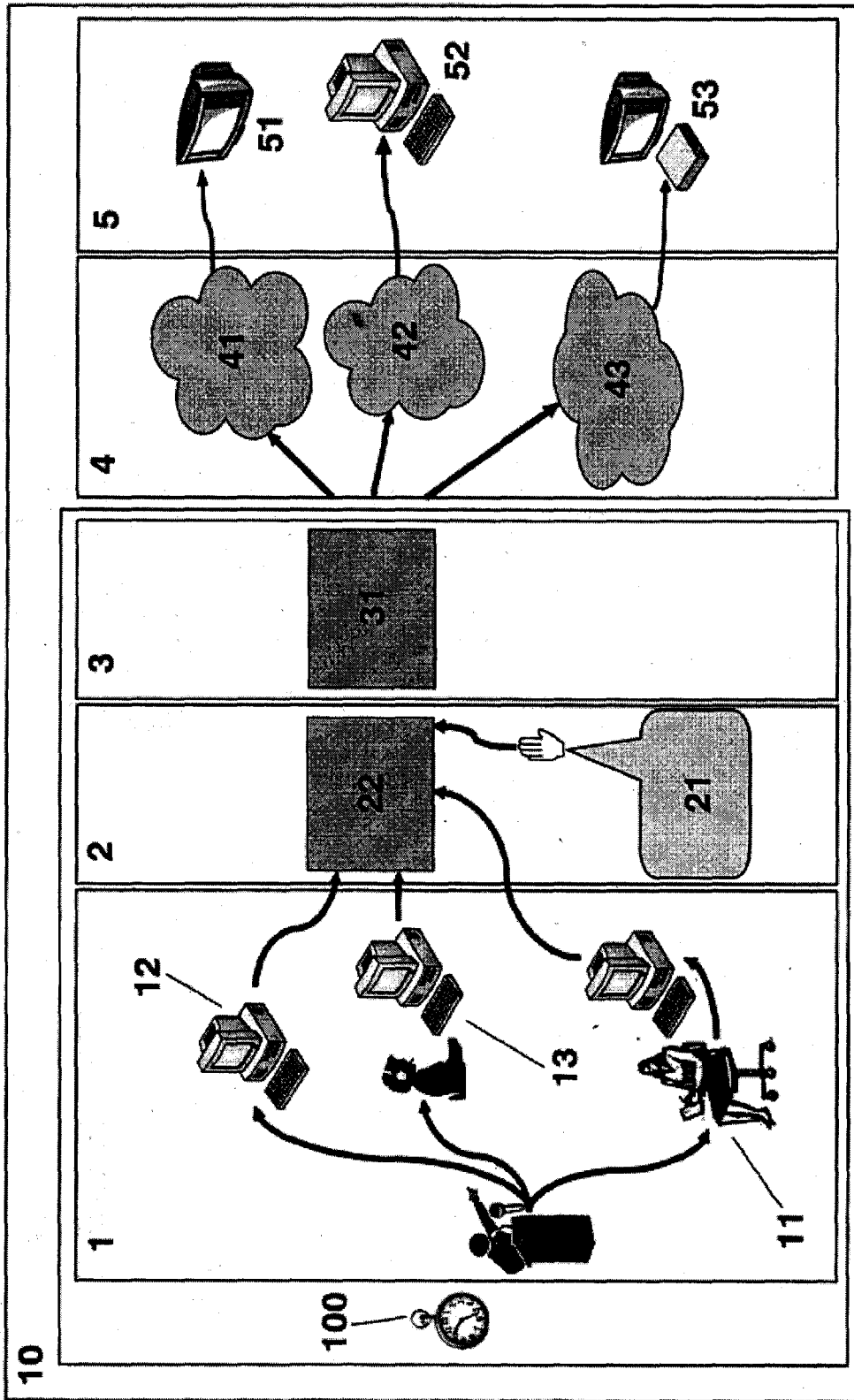


FIG. 1

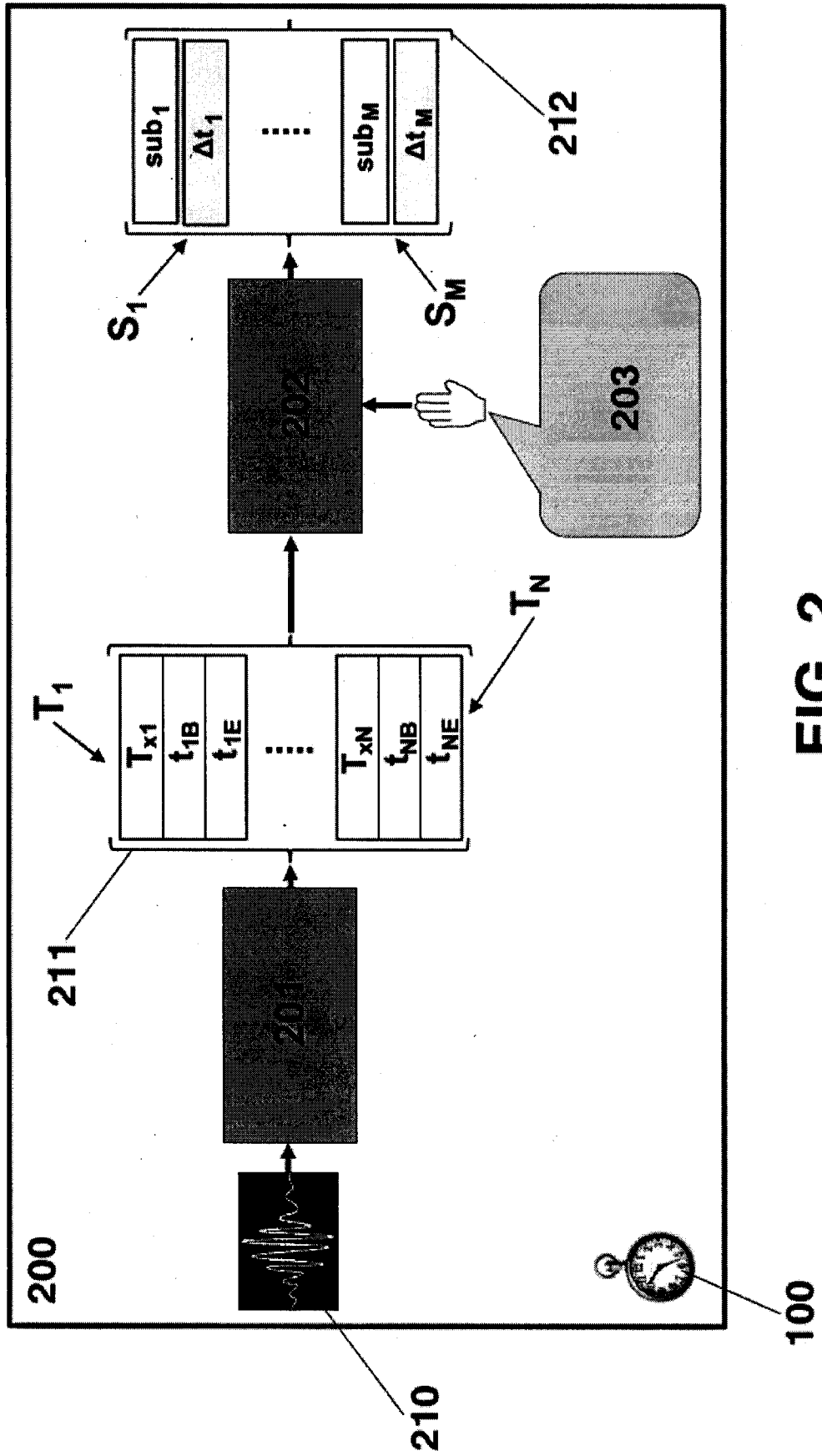


FIG. 2

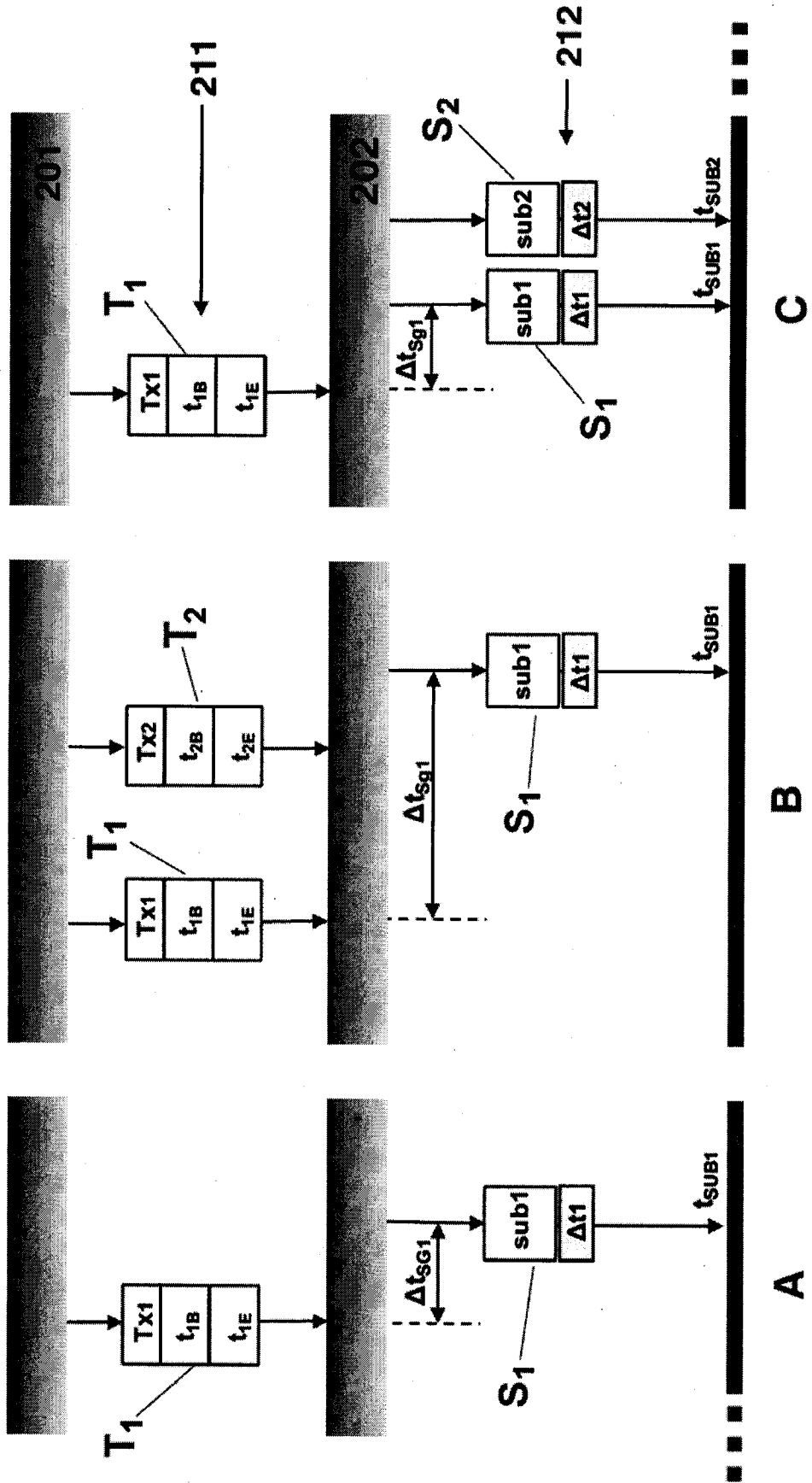


FIG. 3

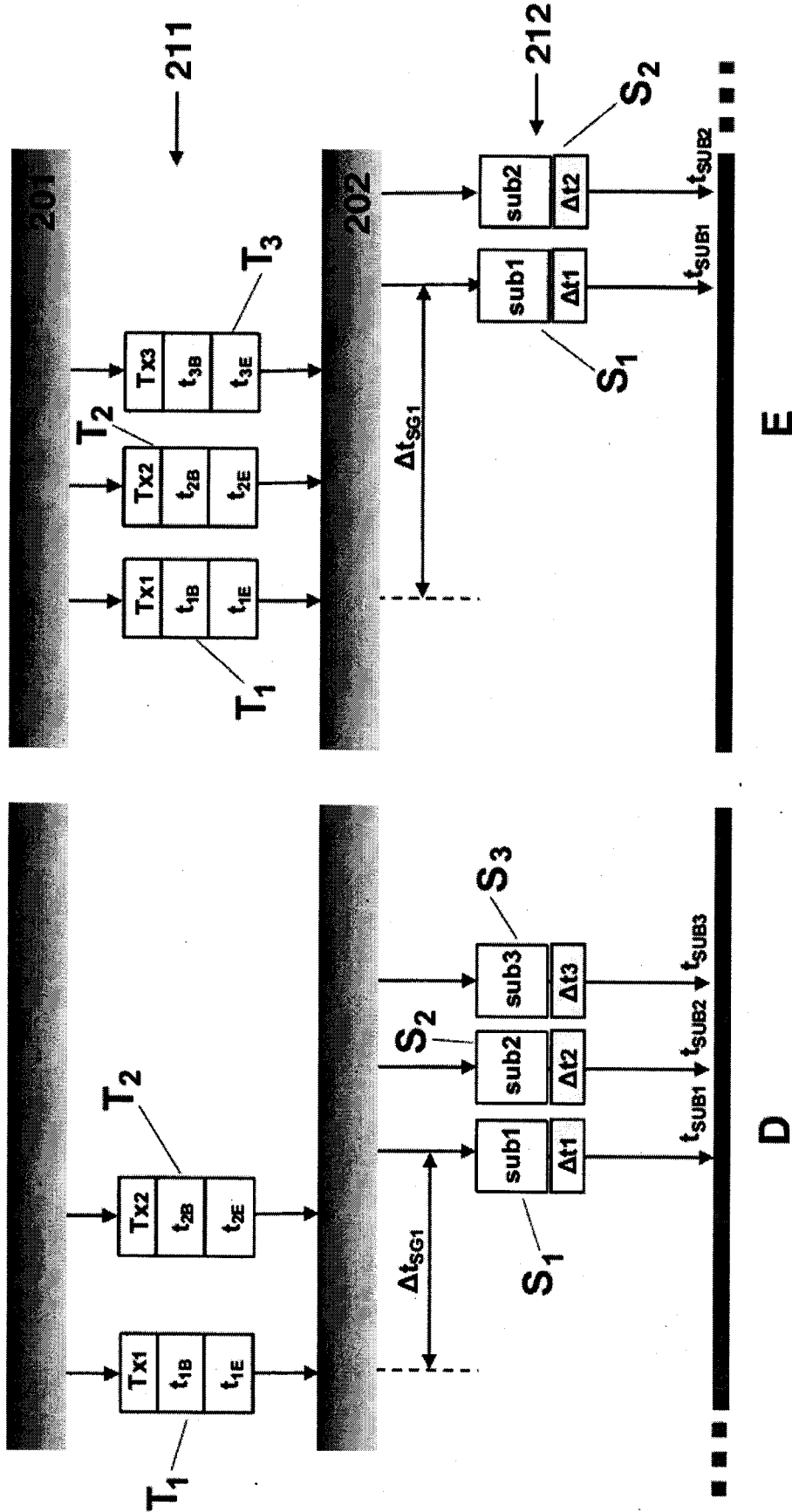


FIG. 4

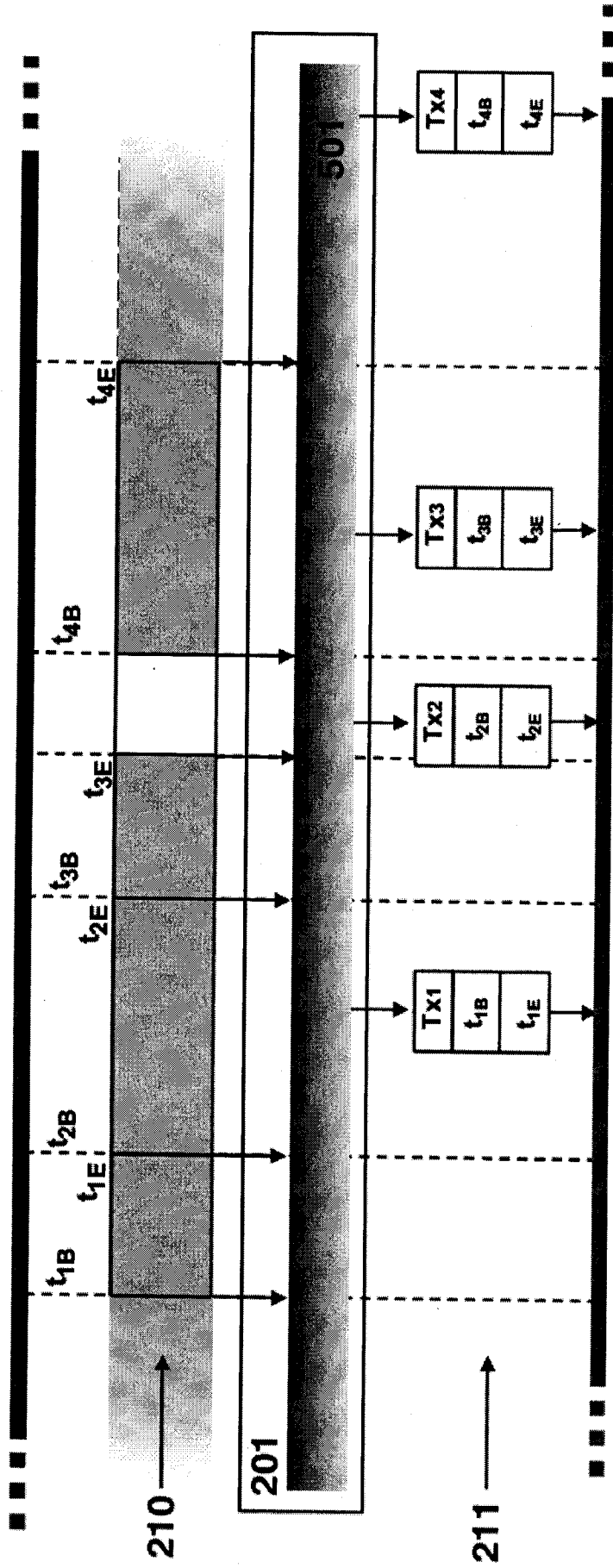


FIG. 5

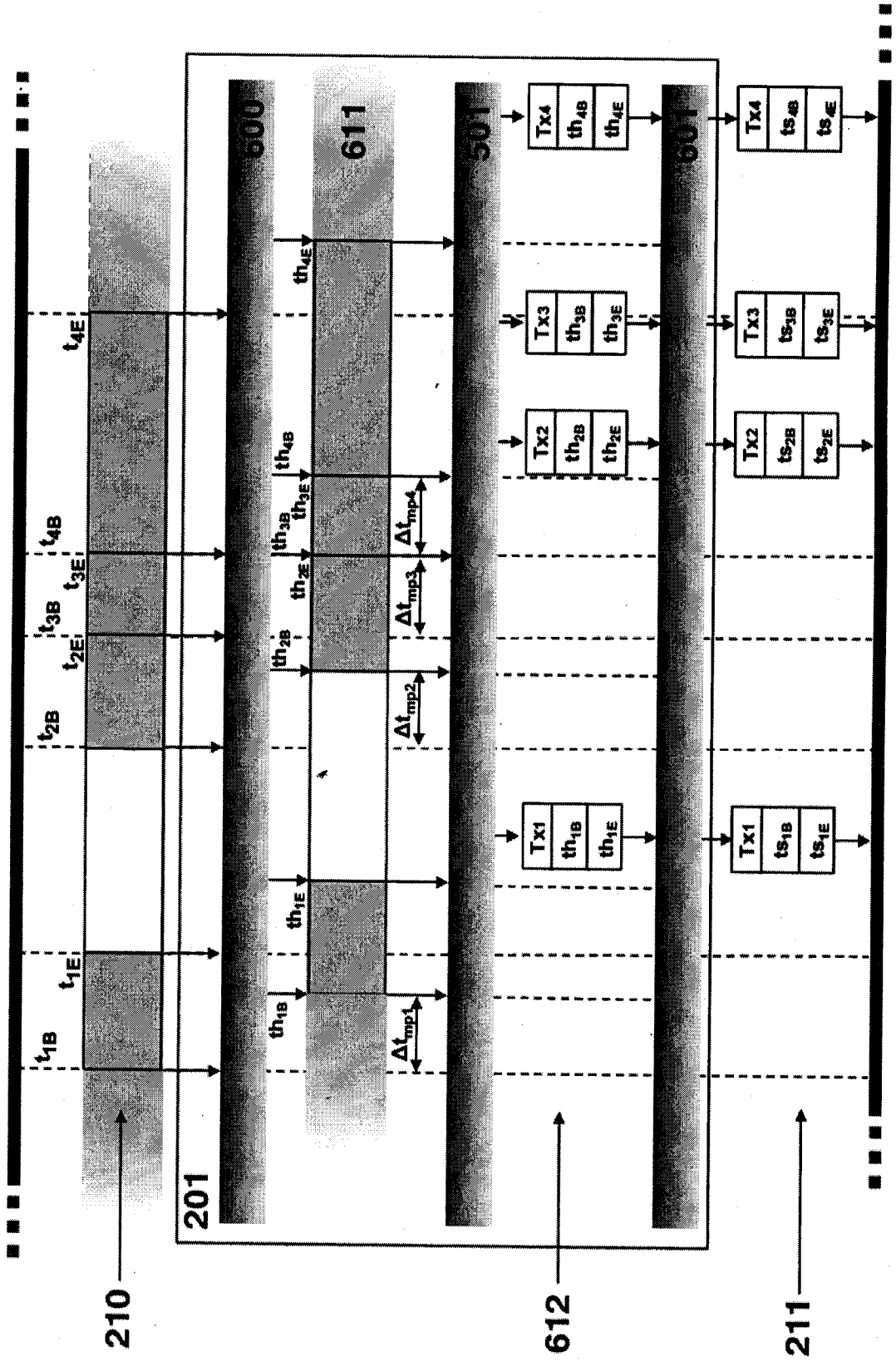


FIG. 6

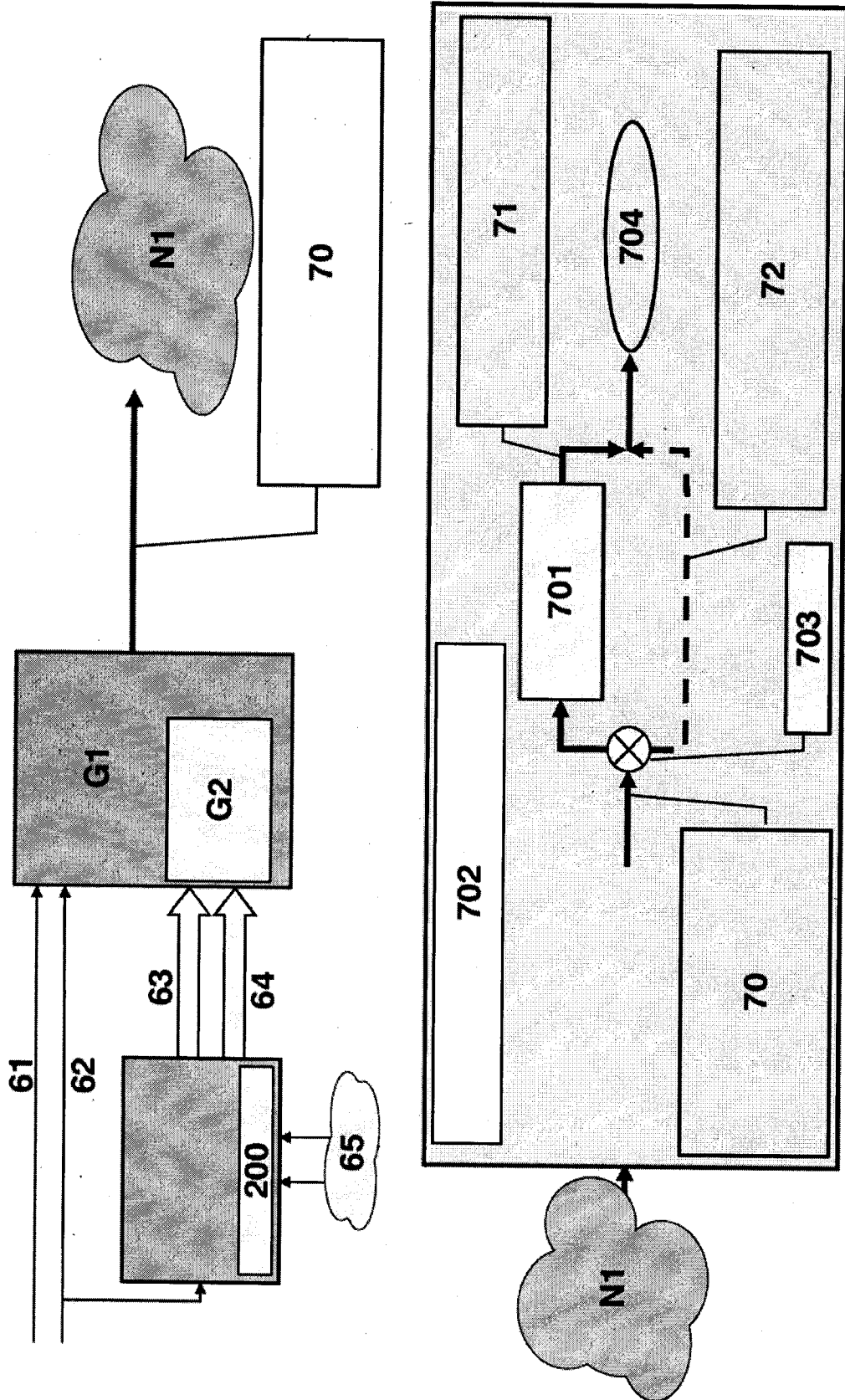


FIG. 7

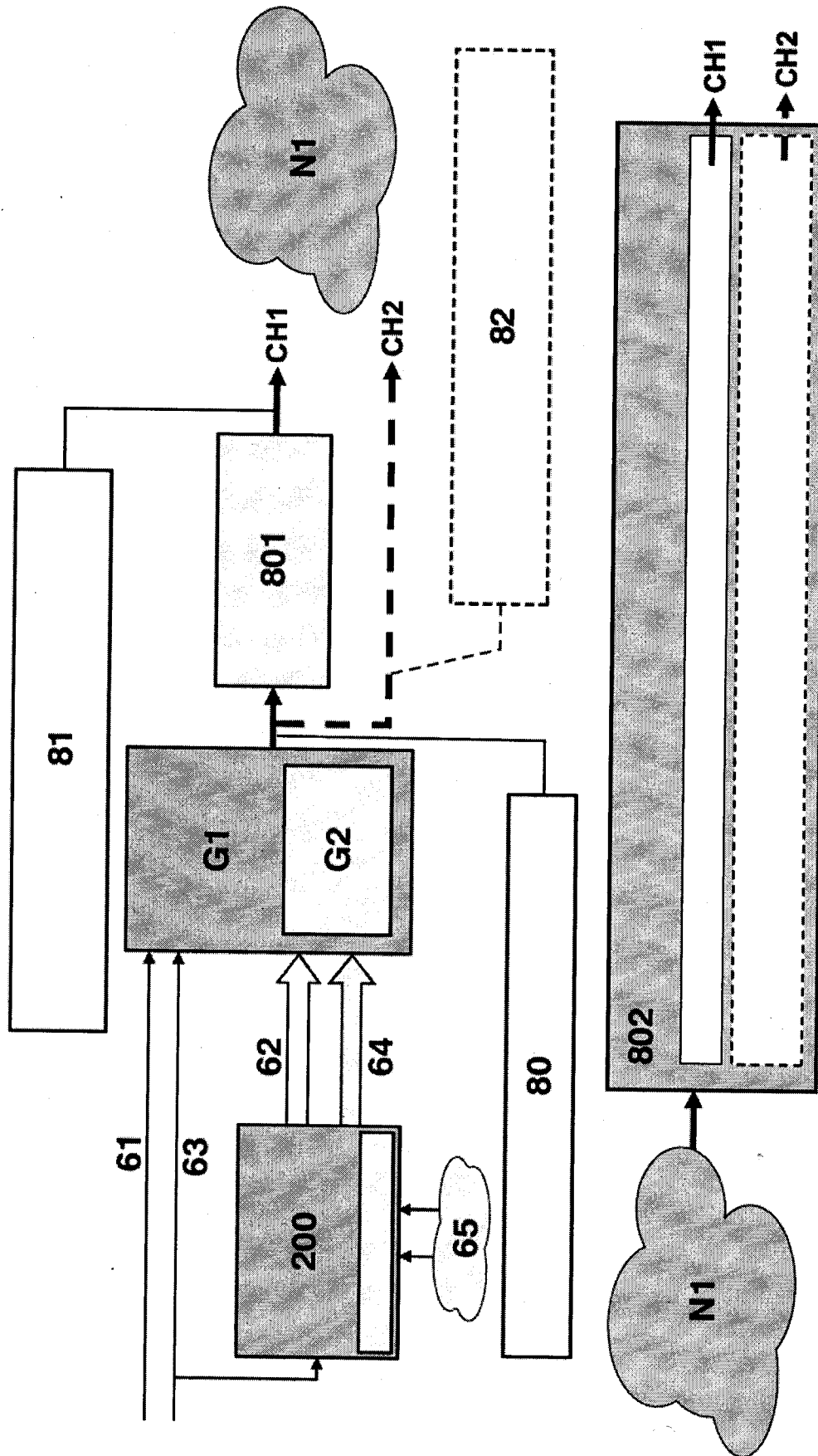


FIG. 8

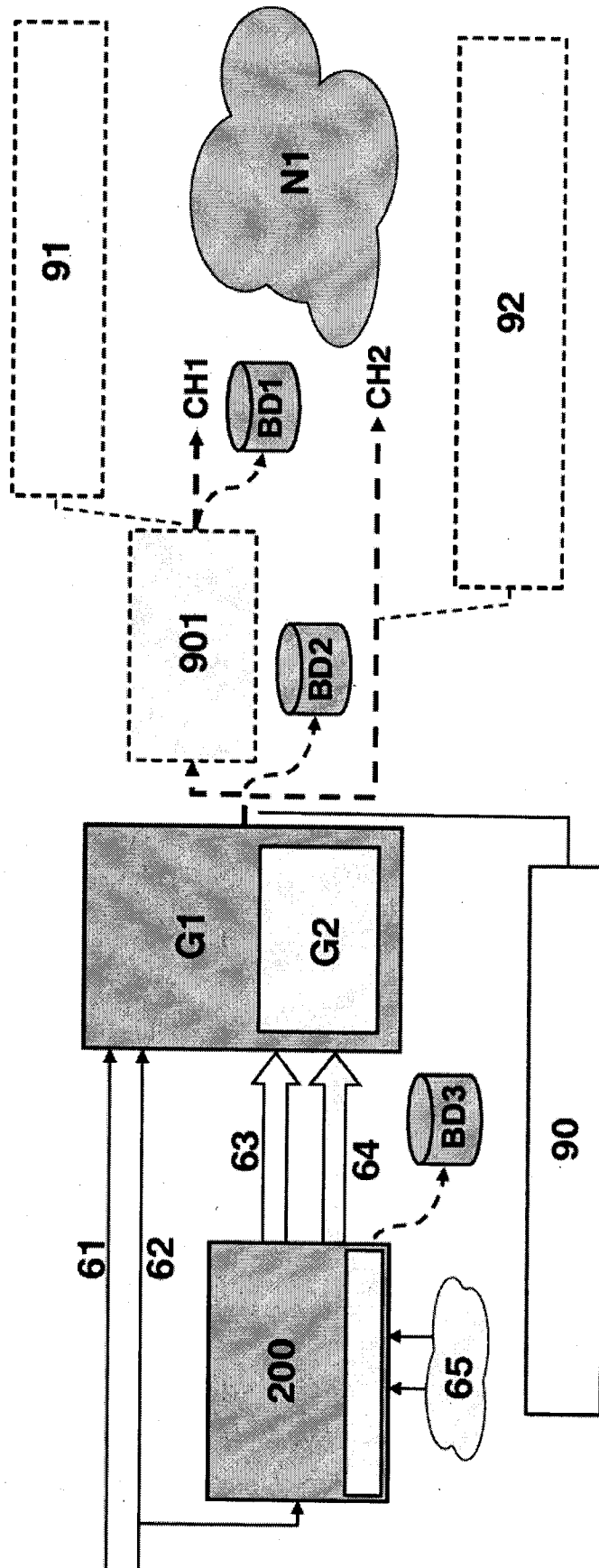


FIG. 9

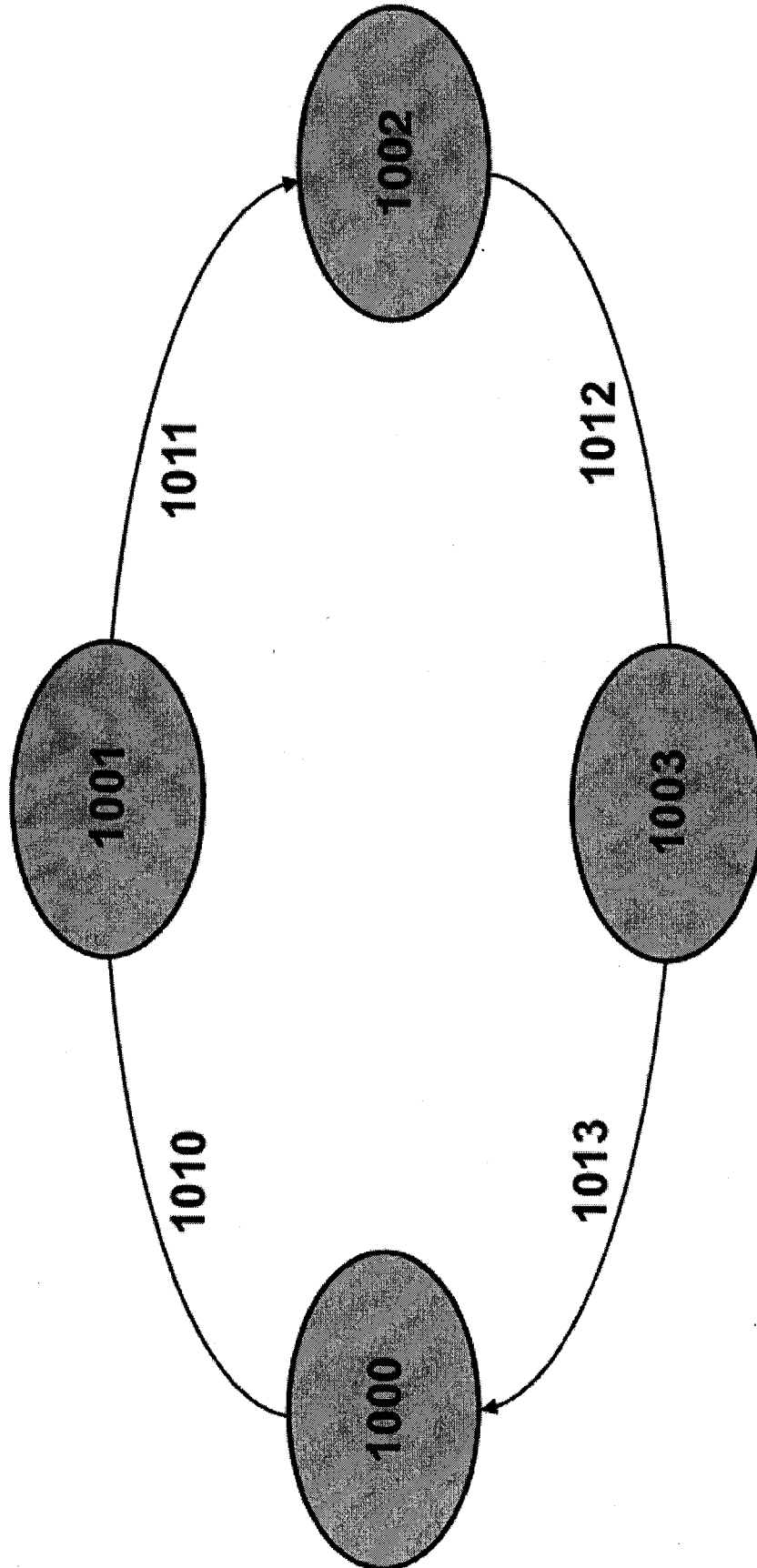


FIG. 10

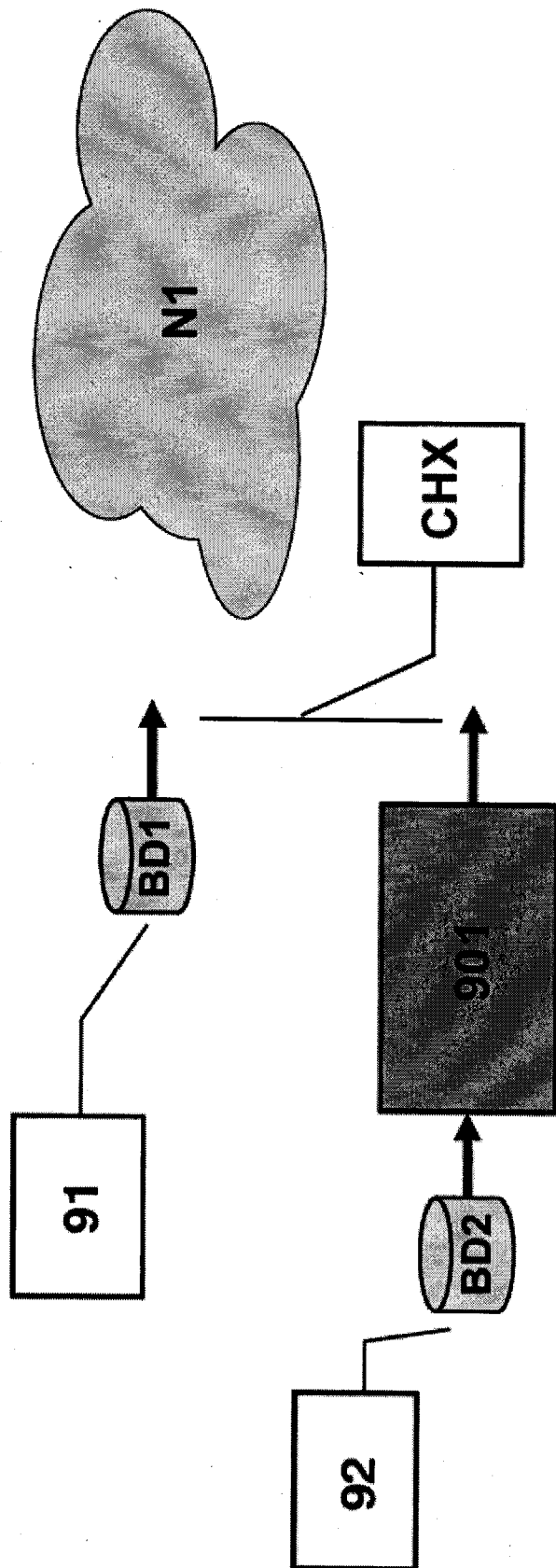


FIG. 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/ES2011/000166

A. CLASSIFICATION OF SUBJECT MATTER

H04N7/24 (2011.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N, G11B, G09G

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPODOC, INVENES, WPI

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2004168203 A1 (SEO KANG SOO ET AL.) 26/08/2004, figures 1 - 3. paragraphs[6 - 16, 25 - 33];	1-23
A	EP 1909278 A1 (SONY COMP ENTERTAINMENT INC) 09/04/2008, paragraphs[9, 17 - 28, 40 - 55, 72 - 82]; figures 1, 3, 4, 6.	1-23
A	US 2005185929 A1 (KANG MAN-SEOK ET AL.) 25/08/2005, figure 1, paragraphs[28 - 35];	1-23

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance.</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure use, exhibition, or other means.</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p>
--	--

Date of the actual completion of the international search
30/09/2011

Date of mailing of the international search report
(24/10/2011)

Name and mailing address of the ISA/

OFICINA ESPAÑOLA DE PATENTES Y MARCAS
Paseo de la Castellana, 75 - 28071 Madrid (España)
Facsimile No.: 91 349 53 04

Authorized officer
B. Pérez García

Telephone No. 91 3493408

INTERNATIONAL SEARCH REPORT

International application No.

Information on patent family members

PCT/ES2011/000166

Patent document cited in the search report	Publication date	Patent family member(s)	Publication date
US2004168203 A	26.08.2004	KR20040051446 A KR100939711 B	18.06.2004 01.02.2010

EP1909278 A	09.04.2008	WO2007004395 A JP2007012218 A JP4311570B2 B EP20060766690 US2009214178 A US7844166 B	11.01.2007 18.01.2007 12.08.2009 14.06.2006 27.08.2009 30.11.2010

US2005185929 A	25.08.2005	KR20050083258 A KR100716973 B WO2005079172 A CA2556526 A EP1716569 A EP20050726921 CN1922681 A CN1922681 B JP2007523438 A CN101068329 A CN100546362 C EP1968068 A EP20080157600 RU2337415 C US2008267587 A SG147413 A BRPI0507879 A CN101568002 A RU2008119131 A	26.08.2005 10.05.2007 01.09.2005 01.09.2005 02.11.2006 19.02.2005 28.02.2007 27.07.2011 16.08.2007 07.11.2007 30.09.2009 10.09.2008 19.02.2005 27.10.2008 30.10.2008 28.11.2008 10.03.2009 28.10.2009 20.11.2009

INFORME DE BÚSQUEDA INTERNACIONAL

Solicitud internacional nº
PCT/ES2011/000166

A. CLASIFICACIÓN DEL OBJETO DE LA SOLICITUD
H04N7/24 (2011.01)

De acuerdo con la Clasificación Internacional de Patentes (CIP) o según la clasificación nacional y CIP.

B. SECTORES COMPRENDIDOS POR LA BÚSQUEDA

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)
H04N, G11B, G09G

Otra documentación consultada, además de la documentación mínima, en la medida en que tales documentos formen parte de los sectores comprendidos por la búsqueda

Bases de datos electrónicas consultadas durante la búsqueda internacional (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

EPODOC, INVENES, WPI

C. DOCUMENTOS CONSIDERADOS RELEVANTES

Categoría*	Documentos citados, con indicación, si procede, de las partes relevantes	Relevante para las reivindicaciones nº
A	US 2004168203 A1 (SEO KANG SOO ET AL.) 26/08/2004, figuras 1 - 3. párrafos[6 - 16, 25 - 33];	1-23
A	EP 1909278 A1 (SONY COMP ENTERTAINMENT INC) 09/04/2008, párrafos[9, 17 - 28, 40 - 55, 72 - 82]; figuras 1, 3, 4, 6.	1-23
A	US 2005185929 A1 (KANG MAN-SEOK ET AL.) 25/08/2005, figura 1, párrafos[28 - 35];	1-23

En la continuación del recuadro C se relacionan otros documentos Los documentos de familias de patentes se indican en el anexo

* Categorías especiales de documentos citados:	"T"	documento ulterior publicado con posterioridad a la fecha de presentación internacional o de prioridad que no pertenece al estado de la técnica pertinente pero que se cita por permitir la comprensión del principio o teoría que constituye la base de la invención.
"A" documento que define el estado general de la técnica no considerado como particularmente relevante.	"X"	documento particularmente relevante; la invención reivindicada no puede considerarse nueva o que implique una actividad inventiva por referencia al documento aisladamente considerado.
"E" solicitud de patente o patente anterior pero publicada en la fecha de presentación internacional o en fecha posterior.	"Y"	documento particularmente relevante; la invención reivindicada no puede considerarse que implique una actividad inventiva cuando el documento se asocia a otro u otros documentos de la misma naturaleza, cuya combinación resulta evidente para un experto en la materia.
"L" documento que puede plantear dudas sobre una reivindicación de prioridad o que se cita para determinar la fecha de publicación de otra cita o por una razón especial (como la indicada).	"&"	documento que forma parte de la misma familia de patentes.
"O" documento que se refiere a una divulgación oral, a una utilización, a una exposición o a cualquier otro medio.		
"P" documento publicado antes de la fecha de presentación internacional pero con posterioridad a la fecha de prioridad reivindicada.		

Fecha en que se ha concluido efectivamente la búsqueda internacional.
30/09/2011

Fecha de expedición del informe de búsqueda internacional.
24 de octubre de 2011 (24/10/2011)

Nombre y dirección postal de la Administración encargada de la búsqueda internacional
OFICINA ESPAÑOLA DE PATENTES Y MARCAS
Paseo de la Castellana, 75 - 28071 Madrid (España)
Nº de fax: 91 349 53 04

Funcionario autorizado
B. Pérez García
Nº de teléfono 91 3493408

INFORME DE BÚSQUEDA INTERNACIONAL

Solicitud internacional nº

Informaciones relativas a los miembros de familias de patentes

PCT/ES2011/000166

Documento de patente citado en el informe de búsqueda	Fecha de Publicación	Miembro(s) de la familia de patentes	Fecha de Publicación
US2004168203 A	26.08.2004	KR20040051446 A KR100939711 B	18.06.2004 01.02.2010

EP1909278 A	09.04.2008	WO2007004395 A JP2007012218 A JP4311570B2 B EP20060766690 US2009214178 A US7844166 B	11.01.2007 18.01.2007 12.08.2009 14.06.2006 27.08.2009 30.11.2010

US2005185929 A	25.08.2005	KR20050083258 A KR100716973 B WO2005079172 A CA2556526 A EP1716569 A EP20050726921 CN1922681 A CN1922681 B JP2007523438 A CN101068329 A CN100546362 C EP1968068 A EP20080157600 RU2337415 C US2008267587 A SG147413 A BRPI0507879 A CN101568002 A RU2008119131 A	26.08.2005 10.05.2007 01.09.2005 01.09.2005 02.11.2006 19.02.2005 28.02.2007 27.07.2011 16.08.2007 07.11.2007 30.09.2009 10.09.2008 19.02.2005 27.10.2008 30.10.2008 28.11.2008 10.03.2009 28.10.2009 20.11.2009
