

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 396 461**

51 Int. Cl.:

G11B 19/02 (2006.01)

G06F 17/30 (2006.01)

G11B 27/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.02.2011 E 11153536 (5)**

97 Fecha y número de publicación de la concesión europea: **26.09.2012 EP 2362392**

54 Título: **Procedimiento de navegación en un contenido sonoro**

30 Prioridad:

15.02.2010 FR 1051030

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

21.02.2013

73 Titular/es:

**FRANCE TÉLÉCOM (100.0%)
78 rue Olivier de Serres
75015 Paris, FR**

72 Inventor/es:

**LE MER, PASCAL;
CHARLET, DELPHINE;
DENJEAN, MARC y
GONOT, ANTOINE**

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 396 461 T3

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

DESCRIPCIÓN

Procedimiento de navegación en un contenido sonoro

5 La presente invención se refiere a un procedimiento y un sistema de navegación en un contenido sonoro de tipo vocal.

10 La adquisición del conocimiento de las informaciones contenidas en el cuerpo de una palabra necesita la escucha completa de la señal sonora correspondiente. En el caso de un cuerpo de tamaño grande, esta operación puede llegar a necesitar bastante tiempo.

15 Las técnicas de compresión temporal de un fichero sonoro, tales como la aceleración, la supresión de las porciones de señal inútiles (por ejemplo, las pausas), etc., no permiten ganar mucho tiempo en la medida en que un contenido no es inteligible más que después de que la compresión alcance un factor "dos".

20 Técnicas conocidas permiten transcribir una señal de audio en un texto. El texto así obtenido permite a continuación ser visualizado por ejemplo en una pantalla de ordenador y leído por un usuario. Al ser la lectura de un texto más rápida que la escucha, el usuario puede obtener así más rápidamente las informaciones que considere pertinente. Sin embargo, el sonido es asimismo portador de informaciones que son difíciles de cuantificar y de representar por medio de la imagen. Tales informaciones son, por ejemplo, la expresividad, el género o la personalidad del locutor. El fichero texto obtenido mediante este método no contiene estas informaciones. Además, la transcripción automática de lenguaje natural engendra numerosos errores de transcripción y el texto obtenido puede ser difícil de comprender por el lector.

25 La solicitud de patente FR 0854340 depositada el 27 de Junio de 2008, divulga un procedimiento de visualización de informaciones relativas a un mensaje sonoro en el que una parte del mensaje sonoro se presenta en forma de representación visual cronológica, y otra parte de las palabras claves son presentadas en forma textual en función de su posición cronológica. Las palabras claves presentadas informan al usuario sobre el contenido del mensaje.

30 Este procedimiento permite conocer lo esencial de un mensaje, mediante inspección visual, ofreciendo la posibilidad de escuchar la totalidad, o una parte, del mensaje.

35 Este procedimiento no está adaptado a los cuerpos sonoros cuyo tamaño sea importante. En efecto, el número de palabras presentadas está limitado, principalmente por el tamaño de la pantalla. La aplicación de este método a un cuerpo de gran tamaño no permite presentar más que un número de palabras restringido y no representativo del conjunto contenido. En consecuencia, no proporciona una apreciación real del contenido del cuerpo.

40 Una función "zoom" permite obtener más detalles, especialmente más palabras clave, sobre una porción de mensaje reducida. Para conocer lo esencial del documento, el usuario debe barrer el conjunto del documento, es decir, "hacer zoom" sobre diferentes partes del contenido.

También se conocen por medio de la técnica anterior técnicas de navegación en un contenido sonoro:

45 El documento US 5.031.113 describe un sistema de tratamiento de texto que permite la visualización de una representación del contenido sonoro y la puesta de relieve de una palabra reproducida.

El documento EP 0877378 A2 describe un sistema para la edición de registros vocales, que presenta una representación gráfica del contenido sonoro y un índice textual que permite un acceso directo al contenido sonoro.

50 El documento US 2002/0178002 A1 describe una solución que permite filtrar una transcripción y no representar más que la aparición de una palabra clave buscada por el usuario.

55 La aplicación de esta función "zoom" a un gran número de porciones del contenido es larga y molesta, puesto que necesita muchas manipulaciones por parte del usuario.

Además, si el usuario desea visualizar una porción visualizada con anterioridad, debe repetir al menos un cierto número de las operaciones de "zoom" efectuadas con anterioridad.

60 La navegación en un contenido vocal de gran tamaño no es por lo tanto fácil.

65 Existe por ello una necesidad de acceder de manera simple y rápida a las informaciones pertinentes de un contenido vocal de gran tamaño.

La invención pretende mejorar la situación.

A este efecto, la invención se refiere a un procedimiento de navegación en un contenido sonoro definido en la

reivindicación 1.

Según la invención, al menos una palabra clave se almacena en memoria en asociación con al menos dos posiciones representativas de al menos dos apariciones de dicha palabra clave en el contenido sonoro, y el procedimiento incluye:

- una etapa de presentación de una representación del contenido sonoro;
- durante la reproducción sonora del contenido, una etapa de detección de un fragmento actual representativo de una palabra clave almacenada en una primera posición;
- una etapa de determinación de al menos un segundo fragmento representativo de dicha palabra clave y de una segunda posición en función de las posiciones almacenadas;
- una etapa de revelación de la posición de los fragmentos en la representación del contenido sonoro.

De ese modo, durante la escucha de un contenido vocal, una interfaz gráfica revela los diferentes emplazamientos de una palabra clave pronunciada en el contenido o representativa del fragmento. El usuario puede así identificar fácilmente las porciones de contenidos que le interesen, y solicitar la escucha de porciones específicas.

Según un modo particular de realización, el procedimiento de navegación incluye además, a continuación de una recepción de una petición del usuario, una etapa de detención de la reproducción sonora en curso, seguido de una etapa de reproducción sonora del contenido a partir de la segunda posición.

El usuario puede así navegar rápidamente en el contenido en función de sus puntos de interés sin tener que escuchar el conjunto del contenido.

Según una característica particular del procedimiento de navegación, la reproducción del fragmento actual va seguida de una etapa de reproducción de al menos un fragmento determinado.

El usuario tiene así la posibilidad de entender sucesivamente los diferentes fragmentos representativos de una misma palabra clave. Una navegación objetivada de esa manera, permite una ganancia de tiempo para el usuario.

Según un modo de realización particular del procedimiento de navegación, el contenido sonoro está representado en forma de un círculo, en el que una posición sobre el círculo representa una posición cronológica en el contenido, y la etapa de puesta de manifiesto incluye una etapa de presentación de un enlace entre la primera y la segunda posiciones.

El círculo es una forma particularmente adaptada para visualizar la correspondencia entre varios fragmentos de contenidos. Esta forma permite una representación simple, concisa y clara de los enlaces entre las diferentes posiciones de los fragmentos.

Según un modo de realización particular del procedimiento de navegación en el que el contenido está dividido en segmentos, la representación del contenido es una representación segmentada y la etapa de puesta de manifiesto comprende la revelación del segmento que contiene el fragmento actual y la revelación de al menos un segmento que contiene al menos un segundo fragmento.

El corte en segmentos permite un corte natural del contenido, por ejemplo un corte en párrafos, en capítulos o en objetos tratados.

Según una característica particular, un segmento representado es seleccionable por medio de una interfaz de usuario.

El corte en segmentos permite así una navegación fácil del usuario por el contenido sonoro.

Según un modo de realización particular del procedimiento de navegación, la reproducción del contenido se efectúa sobre un conjunto de altavoces situados espacialmente sobre el círculo, el fragmento actual es reproducido en un primer altavoz del conjunto y un fragmento determinado es reproducido simultáneamente en un segundo altavoz del conjunto, siendo los altavoces seleccionados en función de la posición del fragmento sobre el círculo que representa el contenido sonoro.

Este efecto de separación espacial del sonido permite al usuario entender, además de la reproducción en curso del contenido sonoro, un fragmento del contenido relativo a la misma palabra clave. De ese modo se tiene una idea general del contenido del fragmento determinado y puede, en función de los elementos entendidos, decidir si solicita o no el fragmento por un solo altavoz. El efecto obtenido es comparable a lo que entiende una persona en presencia de dos interlocutores que hablan al mismo tiempo.

La separación espacial hace que sea posible la lectura simultánea de varias señales de audio, todo ello conservando un cierto grado de inteligibilidad respecto a cada una de ellas.

- 5 Las diferencias de posición espacial de un sonido facilitan la escucha selectiva. Cuando varios sonidos son audibles simultáneamente, el oyente está capacitado para dirigir su atención sobre uno solo.

10 Por otra parte, si sólo son audibles dos fragmentos sonoros simultáneamente, las diferencias de posición espacial pueden facilitar de ese modo la escucha compartida. Se espera en consecuencia un mayor confort, una disminución de la carga de trabajo y una mejor inteligibilidad del contenido durante la escucha simultánea.

15 El oyente identifica más fácilmente un elemento en cada pantalla si un sonido le indica dónde se encuentra ese elemento. El hecho de que los fragmentos sean reproducidos en altavoces situados espacialmente en correspondencia con la representación del contenido sonoro, facilita la inteligibilidad de los contenidos difundidos.

20 La separación espacial del sonido se utiliza en la presente memoria para difundir simultáneamente dos partes de un contenido sonoro a dos azimuts distintos, por ejemplo una primera parte del contenido en una dirección del espacio, por un primer altavoz, y otra parte del contenido en otra dirección del espacio, por ejemplo por un segundo altavoz, lo que permite al usuario tener una idea rápida de la segunda parte del contenido. Se puede a continuación navegar en el contenido y escuchar más atentamente los pasajes que hayan mantenido su atención.

Según una característica particular, los altavoces son altavoces virtuales de un sistema de reproducción binaural.

25 El efecto de separación espacial se obtiene así a partir de un casco de audio conectado a un dispositivo de reproducción.

Según una característica particular, el nivel sonoro de la señal emitida por el primer altavoz es superior al nivel sonoro de la señal emitida por el segundo altavoz.

- 30 La reproducción en curso se mantiene así como la principal fuente sonora.

La invención se refiere igualmente a un dispositivo de navegación en un contenido sonoro según la reivindicación 10, que incluye:

- 35 - medios de almacenamiento de al menos una palabra clave en asociación con al menos dos posiciones representativas de al menos dos apariciones de la citada palabra clave en el contenido sonoro;

- medios de presentación de una representación del contenido sonoro;

- 40 - medios de detección, durante la reproducción sonora del contenido, de un fragmento actual representativo de una palabra clave almacenada en una primera posición;

45 - medios de determinación de al menos un segundo fragmento representativo de la citada palabra clave y de una segunda posición en función de las posiciones almacenadas;

- medios de puesta de relieve de la posición de los fragmentos en la representación del contenido sonoro.

50 La invención se refiere además a un sistema de navegación que comprende un dispositivo de reproducción tal como el que se ha descrito con anterioridad, y al menos dos altavoces.

La invención se refiere, por último, a un producto de programa de ordenador que comprende las instrucciones para llevar a cabo las etapas del procedimiento de navegación tal como se ha descrito en lo que antecede, cuando éste se carga y se ejecuta mediante un procesador.

- 55 Otras particularidades y ventajas de la presente invención se pondrán de relieve en la descripción que sigue de modos de realización dados a título de ejemplos no limitativos, con referencia a los dibujos anexos, en los que:

- la figura 1 ilustra un sistema de navegación según un modo de realización de la invención;

- 60 - la figura 2 es un organigrama que ilustra las diferentes etapas de un procedimiento de navegación puesto en práctica en un sistema de navegación según un modo de realización;

- la figura 3 es un ejemplo de representación de un contenido vocal obtenido mediante la puesta en práctica de un procedimiento de navegación según un modo de realización;

- 65 - la figura 4 ilustra un ejemplo de metadatos asociados a un contenido sonoro según un primer modo de realización;

- las figuras 5a y 5b son ejemplos de representación de un contenido vocal obtenido mediante la puesta en práctica de un procedimiento de navegación según un primer modo de realización particular;

5 - la figura 6 ilustra un ejemplo de metadatos asociados a un contenido sonoro según un segundo modo de realización;

- la figura 7 ilustra una representación de un contenido vocal obtenido mediante la puesta en práctica de un procedimiento de navegación según un segundo modo de realización; y

10

- la figura 8 ilustra un ejemplo de interfaz gráfica.

Ahora se va a describir un modo de realización de la invención con referencia a las figuras 1 a 3.

15 La figura 1 representa un sistema de navegación SYS según un modo de realización de la invención.

El sistema SYS comprende un dispositivo de navegación NAV y un sistema de reproducción sonora HP.

El sistema de reproducción sonora HP comprende, por ejemplo, medios de reproducción sonora tipo altavoz.

20

En el modo de realización ilustrado en la presente memoria, el sistema de reproducción HP está separado del, y conectado al, dispositivo NAV.

A modo de alternativa, el sistema de reproducción sonora HP está integrado en el dispositivo NAV.

25

El dispositivo NAV es por ejemplo un ordenador de tipo PC.

Típicamente, el dispositivo NAV puede estar integrado en un ordenador, un terminal de comunicación tal como un teléfono móvil, un decodificador de TV conectado a un televisor o de manera más general a un equipo multimedia cualquiera.

30

Este dispositivo NAV incluye una unidad de tratamiento MT dotada de un microprocesador, conectada a una memoria MEM. La unidad de tratamiento MT está pilotada por un programa informático PG. El programa informático PG incluye instrucciones adaptadas para llevar a cabo, en especial, un procedimiento de navegación según un modo de realización de la invención que se describirá posteriormente con referencia a la figura 2.

35

La unidad de tratamiento MT está capacitada para recibir, a través de un módulo de entrada ME, instrucciones de una interfaz de usuario INT, por ejemplo un ratón de ordenador o cualquier otro medio de selección por el usuario sobre la pantalla de visualización.

40

Este dispositivo NAV incluye asimismo una pantalla de visualización E y un módulo de presentación MA para la presentación sobre la pantalla de visualización E.

El mismo incluye igualmente un módulo de reproducción del sonido MS, para la difusión de un contenido vocal en el sistema de reproducción sonora HP.

45

El dispositivo NAV incluye igualmente un módulo de detección de palabras clave REC y un módulo de determinación DET de fragmentos sonoros representativos de palabras claves.

50 Un contenido sonoro CV y metadatos D asociados, han sido registrados en la memoria MEM.

Los metadatos están registrados, por ejemplo, en forma de un fichero de metadatos en la memoria MEM.

El contenido sonoro CV es un contenido de audio que contiene al menos una señal de palabra.

55

Un modo de realización del procedimiento de navegación en el sistema SYS va a ser descrito ahora con referencia a la figura 2.

Durante una etapa previa de indexación E0, se determinan las palabras claves MC_i para el contenido vocal CV. Una palabra clave representa en la presente memoria una palabra o un conjunto de palabras representativas de al menos dos fragmentos de contenido. Una palabra clave es un identificador de contenido.

60

Cada palabra clave determinada MC_i es registrada a continuación en un fichero de metadatos D de la memoria MEM, en asociación con al menos dos posiciones P1_i, P2_j de la palabra-clave MC_i del contenido sonoro CV.

65

Por ejemplo, el fichero de metadatos D contiene una palabra clave MC₁ en asociación con una primera posición P1₁

y con una segunda posición P12, y una palabra clave MC2 en asociación con una primera posición P21 y con una segunda posición P22.

5 La determinación de las palabras claves se realiza, por ejemplo, según un proceso que consiste en convertir el contenido sonoro en una transcripción textual con la ayuda de un algoritmo convencional de "Speech to Text", en efectuar una segmentación temática, y en extraer palabras claves de los segmentos.

10 La segmentación temática se efectúa, por ejemplo, mediante detección en la curva que representa la señal de audio del contenido, de picos que representan la similitud entre dos palabras o grupo de palabras. Un método de ese tipo, de medición de grado de similitud, se encuentra descrito por ejemplo en un documento de G. Salton, J. Allan, C. Buckley y A. Singhal, titulado "Automatic analysis, theme generation and summarization of machine-readable texts", de 1994.

15 La extracción de las palabras claves en los segmentos se fundamenta, por ejemplo, en la frecuencia relativa de las palabras o del grupo de palabras en el segmento.

La etapa previa E0 se realiza una sola vez para el contenido sonoro CV.

20 A modo de alternativa, la etapa previa E0 se determina por medio de un dispositivo de indexación (no representado) y el fichero de metadatos D asociado al fichero CV es transmitido al dispositivo NAV por el dispositivo de indexación a través de, por ejemplo, una red de telecomunicación.

25 Aún a modo de alternativa, los metadatos D son determinados por medio de un análisis editorial manual del contenido sonoro CV efectuado por un operador.

30 Durante una etapa E2, el módulo de presentación MA presenta una representación del contenido vocal CV sobre la pantalla E.

La figuras 3 y 5a ilustran ejemplos de representación del contenido vocal CV.

35 En el modo de realización descrito en la figura 3, el contenido vocal CV está representado por un eje temporal cronológico horizontal. El comienzo del contenido está representado por la marca O, y el final del contenido está representado por la marca F. Marcas intermedias indican el retardo calculado desde el comienzo O.

40 Durante una etapa E4, el módulo MS comanda la reproducción sonora del contenido vocal CV a partir de la marca O, en el sistema de reproducción sonora HP.

A modo de alternativa, la reproducción sonora puede ser iniciada a partir de otro punto del contenido vocal.

45 Un cursor C indica la posición actual de reproducción.

50 Durante una etapa E6, el módulo de detección REC detecta si se ha pronunciado una palabra clave MCi presente en el fichero de metadatos D, o si el fragmento actual Zc pronunciado es representativo de la palabra clave MCi. La posición Pc del fragmento actual Zc corresponde a una de las posiciones registradas en asociación con la palabra clave CMi en el fichero de metadatos D.

Por ejemplo, el fragmento actual Zc, situado en la posición P11, es representativo de la palabra clave MC1.

55 Esta etapa de detección E6 consiste por ejemplo en comparar la posición actual Pc del cursor C con las posiciones registradas en el fichero de metadatos D.

60 Un fragmento es representativo de una palabra clave si contiene esa palabra clave o si el fragmento y la palabra clave son derivados de un mismo lema, siendo el lema la forma de la palabra que se encuentra en la entrada de un diccionario (verbo en infinitivo, nombre singular, adjetivo en masculino singular).

65 Por ejemplo, las palabras "comido", "comerán", son derivados del lema "comer".

A modo de alternativa, un fragmento es representativo de una palabra clave si tiene la misma significación en el contexto.

Por ejemplo, el término "hexágono" es significativo de la palabra "Francia" si el contexto se refiere al país.

Durante una etapa E8, el módulo de determinación DET determina, por lectura en el fichero de metadatos D, al menos un segundo fragmento Zd, representativo de la palabra clave MCi, y una posición Pd para cada segundo fragmento Zd determinado. Una posición Pd de un fragmento Zd representa una segunda posición.

En el ejemplo particular en que el fragmento actual Zc, en la posición P11, representa la palabra clave MC1, se determina un solo segundo fragmento Zd y su posición es P12.

5 Después, durante una etapa E10, se ponen de manifiesto la posición Pc del fragmento actual Zc y la posición Pd del, o de los, segundo(s) fragmento(s) Zd sobre la pantalla E a través de los medios de presentación MA, presentándolos por ejemplo con un color diferente o en resaltado.

En el ejemplo particular descrito, se ponen de manifiesto las posiciones P11 y P12.

10 A modo de alternativa, la palabra clave MC1 es asimismo presentada sobre la pantalla E.

De ese modo, el usuario que entienda un fragmento representativo de la palabra clave MC1, está advertido de que otro fragmento representativo de la misma palabra clave MC1 está presente en otra posición del contenido sonoro CV.

15 Éste puede entonces continuar con la escucha del contenido sonoro o demandar la reproducción del contenido a partir de la posición Pd o a partir de una posición P anterior a la posición Pd para entender el contexto asociado al segundo fragmento Zd.

20 Si sigue con la escucha, las etapas E2 a E10 son repetidas con la palabra clave siguiente, por ejemplo la palabra clave MC2.

25 El procedimiento puede prever de ese modo una reproducción sucesiva de fragmentos determinados, de forma automática o a petición del usuario, o incluso una reproducción simultánea de fragmentos determinados por ejemplo en un sistema de reproducción separada en el espacio como se va a describir posteriormente con referencia a la figura 5a.

Un primer modo de realización particular va a ser descrito ahora con referencia a las figuras 4 y 5a.

30 En este modo de realización, el sistema de reproducción sonora HP comprende un conjunto de 7 altavoces HP1, HP2, ..., HP7 conectados al dispositivo NAV y repartidos espacialmente en torno a este dispositivo.

Por ejemplo, los altavoces están repartidos sobre un círculo virtual alrededor del dispositivo.

35 A modo de alternativa, el sistema SYS comprende un número diferente de altavoces.

En este modo de realización, el contenido sonoro CV se divide en segmentos, por ejemplo 7 segmentos S1, S2, ..., S7. Un segmento es una sub-parte del contenido, por ejemplo, una sub-parte temática.

40 En el modo de realización descrito, un fichero de metadatos D1 ha sido registrado en la memoria. En este fichero de metadatos D1, una posición corresponde a un identificador de segmento y a una posición cronológica con relación al comienzo del segmento.

45 La figura 4 representa un ejemplo de fichero de metadatos D1 para el contenido vocal CV.

La palabra clave MC1 está registrada, por ejemplo, en asociación con una posición P11 en el segmento S1 y con dos posiciones P12 y P13 en el segmento S5. La palabra clave MC3 está registrada, por ejemplo, en asociación con dos posiciones P31 y P32 en el segmento S2 y con una posición P33 en el segmento S4.

50 Una misma palabra clave puede estar representada una sola vez o varias veces en un mismo segmento.

La figura 5a ilustra un ejemplo de representación del contenido vocal CV.

55 El contenido vocal CV está representado por un eje temporal cronológico que tiene forma de círculo. Los segmentos S1 a S7 están repartidos por el círculo. El comienzo del contenido está representado por la marca O situada sobre el segmento S1.

La longitud de los segmentos es, por ejemplo, proporcional a su duración de reproducción sonora.

60 Durante la reproducción sonora, el altavoz por el que se difunde el contenido sonoro CV es el altavoz asociado al segmento reproducido.

Un altavoz está asociado a un segmento en función de la posición del segmento sobre el círculo visualizado en la pantalla, de la posición del altavoz, y de la repartición espacial de los altavoces del conjunto de altavoces.

65 En el modo de realización descrito en la presente memoria, en el que el número de segmentos es igual al número de

altavoces, cada altavoz está asociado a un solo segmento.

A modo de alternativa, un altavoz está asociado a varios segmentos

- 5 Aún a modo de alternativa, un segmento está asociado a varios altavoces. Esta alternativa está particularmente adaptada al caso de que uno de los segmentos sea más largo que los otros.

Durante la reproducción de un segmento, las palabras claves de un segmento son leídas en memoria en el fichero de metadatos D1, y presentadas en la pantalla E.

- 10 El segmento actual, es decir, el segmento en curso de reproducción, es puesto de relieve sobre la pantalla E, por ejemplo mediante la utilización de un color determinado, siendo los otros segmentos de otro color.

Un cursor C indica la posición actual de reproducción sobre el círculo.

- 15 La figura 5a ilustra un ejemplo de representación en el que el cursor C está en la posición P11 del segmento S1.

El fragmento actual pronunciado Zc es entonces representativo de la palabra clave MC1.

- 20 Se determinan dos fragmentos Z2 y Z3 que se sitúan respectivamente en las posiciones P12 y P13 del segmento S5. Los fragmentos Z2 y Z3, representativos de la palabra clave MC1, representan los segundos fragmentos.

Las posiciones P12 y P13 en el segmento S5 representan en la presente memoria las segundas posiciones.

- 25 Según se ha ilustrado en la figura 5a, el segmento S5 asociado al fragmento Z2 es puesto de relieve con la utilización de un color determinado o mediante resaltado, parpadeo, etc., y se visualiza un trazo entre la posición P11 del segmento S1 del fragmento actual Zc y la posición P12 del segmento S5 del segundo fragmento Z2.

- 30 De ese modo, el usuario que entienda la palabra clave MC1, es advertido visualmente de que una palabra clave similar está presente en la posición P12 del segmento S5.

En el modo de realización descrito, cuando se representan varios fragmentos que corresponden a la palabra clave en un mismo segmento diferente, sólo se conserva la posición del primer fragmento determinado.

- 35 En paralelo con la reproducción del contenido sonoro, se reproduce el segundo fragmento Z2 por el altavoz HP5 asociado al segmento S5 según una técnica de separación espacial del sonido.

De ese modo, los dos fragmentos Zc y Z2 estarán localizados en dos posiciones diferentes en el espacio sonoro.

- 40 El usuario puede entonces continuar con la escucha del contenido sonoro CV o demandar la reproducción del contenido CV a partir de la posición P12 del segmento S5 o a partir de alguna otra posición del fichero, por ejemplo una posición P anterior a la posición P12 para entender el contexto asociado al segundo fragmento Z2.

- 45 Una navegación de ese tipo en el contenido, puede ser llevada a cabo mediante posicionamiento del cursor, por desplazamiento circular del dedo o apuntando directamente sobre un segmento o incluso con la utilización de funciones de salto por medio de botones específicos.

El cursor permite al usuario determinar de forma precisa la posición a partir de la cual desea entender el contenido vocal.

- 50 El desplazamiento circular del dedo permite comandar un avance o un retorno rápido por el contenido.

En lo que sigue de la descripción, se va a describir un ejemplo de interfaz gráfica que ofrece una navegación fácil, con referencia a la figura 8.

- 55 El usuario puede comandar igualmente la aceleración de la escucha mediante escucha selectiva. Por ejemplo, puede obtener la reproducción de las frases que contengan únicamente palabras claves, es decir únicamente las palabras claves.

- 60 La figura 5b ilustra un ejemplo de representación del contenido sonoro CV cuando el fragmento actual Zc pronunciado es representativo de la palabra clave MC4. Después se determinan fragmentos Z3 y Z4, respectivamente, en la posición P42 del segmento S3 y en la posición P43 del segmento S6.

- 65 El segmento actual S1, así como los segmentos S3 y S6, son puestos de relieve, por ejemplo, mediante resaltado. Tres trazos Tr1, Tr2 y Tr3 ponen en evidencia la posición de los fragmentos Zc, Z3 y Z4.

Un segundo modo de realización particular del procedimiento de navegación puesto en práctica en el sistema SYS va a ser descrito ahora con referencia a las figuras 6 y 7.

La figura 6 representa un segundo ejemplo de fichero de metadatos D2 para el contenido vocal CV.

5

En este ejemplo, el fichero de metadatos D2 comprende 3 partes D21, D22 y D23.

La primera parte D21 contiene, para cada segmento del contenido vocal CV, las palabras clave representativas del segmento y, en asociación con cada palabra clave, un identificador de la palabra clave en el segmento y al menos dos posiciones. Una posición es, en la presente memoria, la posición de una palabra o de un grupo de palabras que representan la palabra clave asociada.

10

En este modo de realización, una palabra clave es representativa de un segmento si está representada al menos dos veces en el segmento.

15

Por ejemplo, en el segmento S1, la palabra clave MC1 que tiene como identificador Id1, está asociada a las posiciones V1 y W1. En el segmento S3, la palabra clave MC2 que tiene como identificador Id2, está asociada a las posiciones V2 y W2. En el segmento S7, la palabra clave MC3 que tiene como identificador Id3, está asociada a la posición V3.

20

La segunda parte contiene las correspondencias entre palabras claves, y de forma más precisa las correspondencias entre identificadores de palabras clave.

Por ejemplo, los identificadores Id2 e Id3 se encuentran puestos en correspondencia en la parte D22 lo que significa que las palabras MC2 y MC3 son representativas de una misma palabra clave.

25

La tercera parte D23 contiene el conjunto de fragmentos del contenido sonoro CV representados en forma de cuádrupletes. Cada cuádruplete contiene una posición de inicio de fragmento en el contenido sonoro, su duración, un identificador de segmento y un identificador de palabra clave. Los valores de "identificador de segmento" e "identificador de palabra clave" están vacíos o son nulos si el fragmento no es representativo de una palabra clave.

30

Por ejemplo, el fragmento que se inicia en la posición 11600 ms y es de una duración de 656 ms, es un fragmento representativo de la palabra clave que tenga como identificador a Id2 en el segmento S2.

35

Durante la reproducción de un segmento del contenido sonoro CV, por uno de los altavoces del conjunto, se visualizan las palabras claves representativas del segmento reproducido.

Durante la reproducción de un segmento del contenido sonoro CV por uno de los altavoces del conjunto de altavoces, las palabras clave de dicho segmento, por ejemplo MC2 para el segmento S3, son leídas en memoria en el fichero de metadatos D2 y visualizadas sobre la pantalla E.

40

El segmento actual se pone de relieve.

Cuando el fragmento actual Zc pronunciado es representativo de la palabra clave MC2, se determinan fragmentos Z2 y Z3, representativos de la palabra clave MC2. El fragmento Z2 se determina por lectura en la parte D21 del fichero de metadatos D2, y el fragmento Z3 del segmento S7 se determina por lectura de la segunda parte D22 del fichero de metadatos D2.

45

Las posiciones respectivas W2 y V3 de los fragmentos Z2 y Z3, son leídas en el fichero de metadatos D2.

50

La posición V2 del fragmento actual Zc y las posiciones W2 y V3 de los fragmentos Z2 y Z3, son puestos de relieve en la pantalla E.

Por ejemplo, según se ilustra en la figura 7, la puesta de relieve consiste en unir las posiciones de los fragmentos Zc y Z2 a un punto X representativo de la palabra clave MC2, en unir la posición V3 del fragmento Z3 a un punto Y representativo de la palabra clave MC3 (equivalente a MC2), y en unir los puntos X e Y.

55

En paralelo con la reproducción del contenido sonoro, el fragmento Z3 es reproducido por el altavoz asociado al segmento S7 según una técnica de separación espacial del sonido.

60

En el modo de realización descrito, el sistema comprende un conjunto de 7 altavoces repartidos geográficamente por un círculo alrededor de la pantalla.

A modo de alternativa, el sistema SYS comprende un dispositivo D2 tal como el descrito en lo que antecede, y un casco para una reproducción binaural. Los altavoces del sistema de reproducción son entonces altavoces virtuales de esta reproducción binaural.

65

Ahora se va a describir un modo de realización de la etapa de separación espacial del sonido.

Se supone que dos palabras claves han sido detectadas para la palabra clave pronunciada.

5 Se tienen así tres señales de palabra a reproducir por tres altavoces. La señal en curso de reproducción es la señal principal. Las otras señales son señales secundarias.

10 La señal principal se reproduce por un altavoz de los altavoces del conjunto, siendo seleccionado el altavoz en función de la posición actual en la reproducción del contenido. Esta señal principal se entiende a un azimut correspondiente a su posición sobre el círculo de representación sonora.

15 Las señales secundarias son reproducidas sucesivamente, según el orden de aparición en el contenido vocal, al mismo tiempo que la señal principal. Una reproducción sucesiva de las señales secundarias de ese tipo, permite que las señales reproducidas, es decir, la señal principal y una de las señales secundarias, sean mucho mejor comprensibles por el usuario. El número de señales reproducidas simultáneamente se limita así a dos.

20 Las señales secundarias son entendidas a un azimut correspondiente a su posición sobre el círculo de representación sonora.

A modo de alternativa, la señal principal y las señales secundarias son reproducidas simultáneamente.

25 La relación de campo directo/campo reverberado de las señales secundarias, se reduce en algunos decibelios con respecto a la relación de campo directo/campo reverberado de la señal principal. Otro efecto de la creación de una diferencia de nivel entre las dos fuentes consiste en que este tratamiento permite una sensación más estable de distancia, y en consecuencia, un retroceso plano de las señales secundarias más perceptible.

30 La sensación de retroceso plano así dada a las señales secundarias, permite igualmente mantener una mayor inteligibilidad de la señal principal, y reducir la dispersión de la audición auditiva.

Un filtro pasa alto de segundo orden, se aplica a las señales secundarias. La frecuencia de corte de este filtro está comprendida, por ejemplo, en el intervalo [1 kHz : 2 kHz]. La limitación de la "imbricación" de frecuencia de la señal principal y de la señal secundaria así obtenida, permite una diferenciación de timbre entre las señales reproducidas.

35 Se aplica igualmente un proceso de desincronización de las señales. Este proceso consiste en determinar si el inicio de la reproducción de una palabra de la señal principal coincide con una palabra en curso de reproducción de la señal secundaria. En caso de que ese momento coincida con una ausencia de señal secundaria (espacio entre palabras), se determina el intervalo temporal entre el inicio de la palabra reproducida de la señal principal y el comienzo de la palabra a reproducir de la señal secundaria. Si el espacio temporal es inferior a un umbral predeterminado, la reproducción de la palabra a reproducir de la señal secundaria se retarda. El umbral predeterminado está comprendido preferentemente entre 100 ms y 300 ms.

A modo de alternativa, el proceso de desincronización de las señales no se aplica.

45 La puesta de relieve de la relación entre las posiciones de una misma palabra clave durante la reproducción sonora, permite una navegación fácil en el contenido sonoro por parte de un usuario en función de sus centros de interés.

La figura 8 ilustra un ejemplo de interfaz gráfica G que permite una navegación en el contenido sonoro CV.

50 En una primera ventana A, se presenta un círculo Cc que representa el contenido vocal CV. El círculo Cc está cortado en diferentes segmentos S1, S2, ... Las palabras clave o lemas asociados al contenido vocal CV, se presentan en forma de lista desenrollable L en el centro del círculo Cc. Cada palabra clave de un segmento está representada por un punto Y situado a lo largo del círculo Cc en función de la posición de la palabra clave en un contenido sonoro.

55 Por ejemplo, la palabra clave MC1 presentada en el centro del círculo está igualmente representada por los puntos Y1, Y2 e Y3.

Una segunda ventana B contiene los botones de comando para la lectura y la navegación por el contenido sonoro.

60 Un botón B1 permite comandar la lectura.

Un botón B2 permite desplazar el cursor a la palabra clave siguiente.

65 El botón B3 permite desplazar el cursor a la palabra clave anterior.

Un botón B4 permite desplazar el cursor al comienzo del segmento siguiente o del segmento precedente.

Un botón B5 permite seleccionar un modo en el que solamente se leen las frases que contengan las palabras claves.

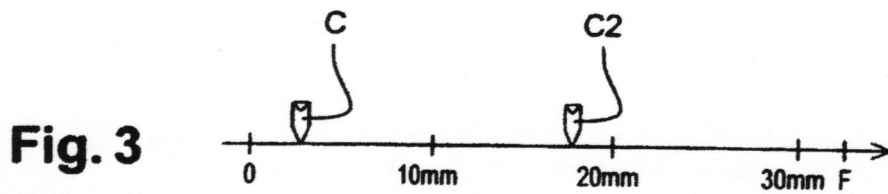
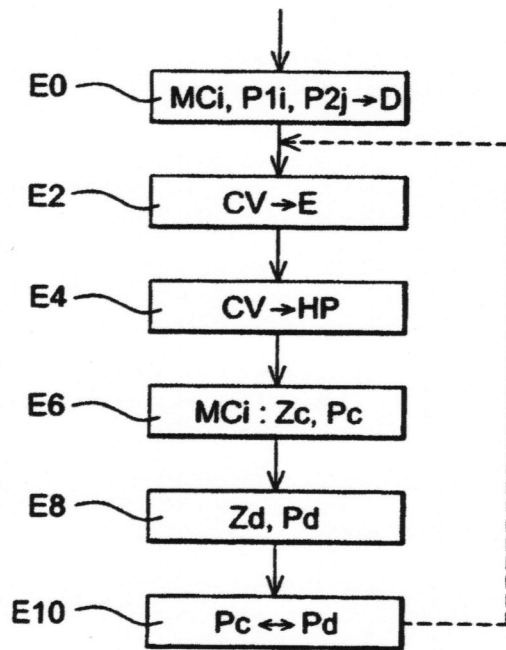
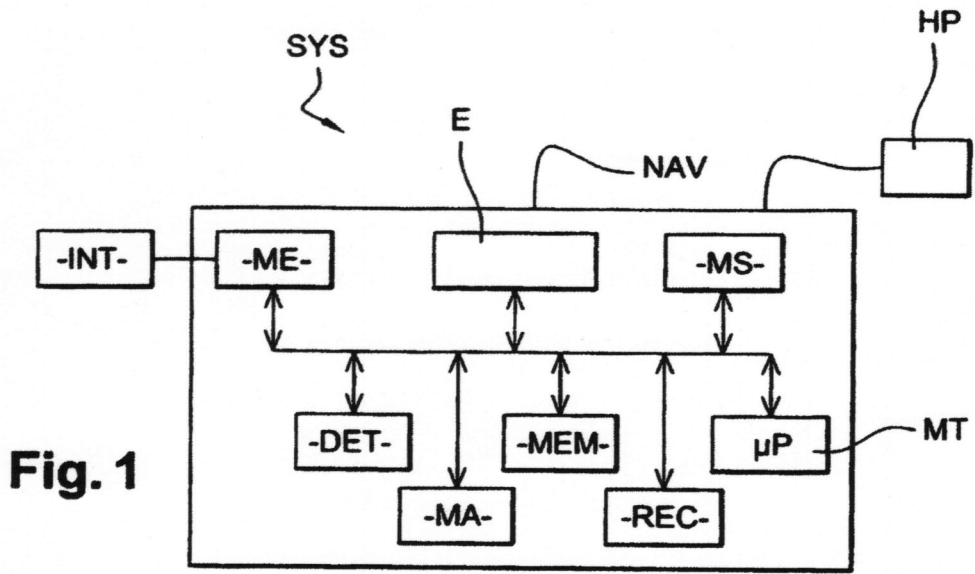
- 5 Un botón B6 permite desplazar el cursor a la instancia siguiente de una palabra clave.
- Un botón B7 permite desplazar el cursor a la instancia anterior de una palabra clave.
- 10 El desplazamiento del cursor a una instancia de una palabra clave puede hacerse igualmente seleccionando una palabra clave en la lista desenrollable con el ratón o con la ayuda de una pantalla táctil.
- Una interfaz gráfica W permite igualmente seleccionar la velocidad de lectura del contenido.
- 15 Una tercera ventana puede contener informaciones relativas al contenido sonoro, por ejemplo una fecha o una duración del contenido.

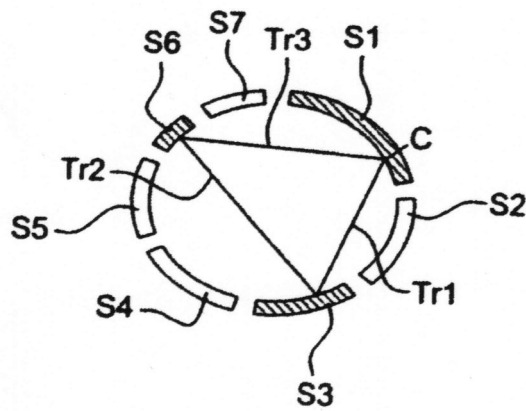
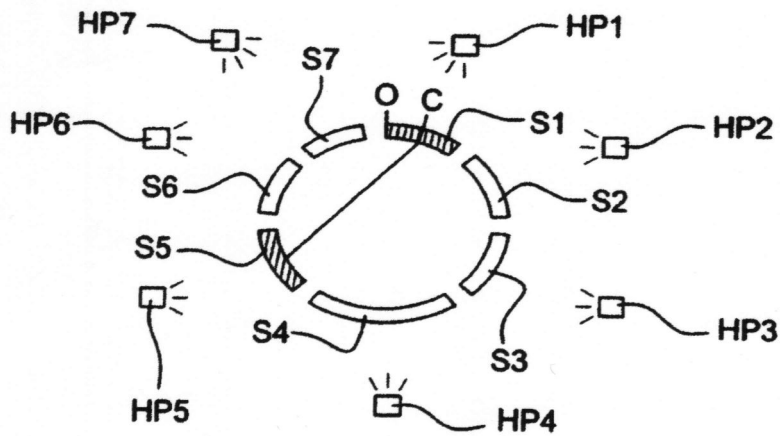
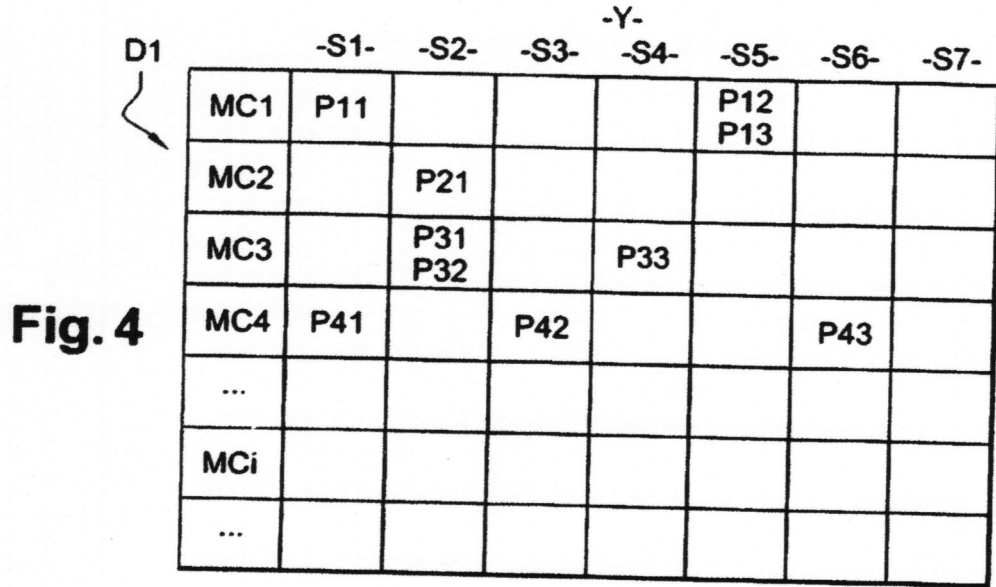
La navegación por el contenido sonoro puede ser efectuada igualmente por introducción de un movimiento, por ejemplo un arco de círculo en un sentido o en el otro, efectuado por el usuario sobre la pantalla de visualización E.

REIVINDICACIONES

- 5 1.- Procedimiento de navegación en un contenido sonoro (CV), caracterizado porque al menos una palabra clave (MC1) se encuentra almacenada en memoria en asociación con al menos dos posiciones (P11, P12) representativas de al menos dos apariciones de dicha palabra clave en el contenido sonoro, y porque comprende:
- una etapa de visualización (E2) de una representación del contenido sonoro;
 - 10 - durante la reproducción sonora del contenido, una etapa de detección (E6) de un fragmento actual (Zc) representativo de una palabra clave (MC1) almacenada en una primera posición (P11);
 - una etapa de determinación (E8) de al menos un segundo fragmento (Zd) representativo de dicha palabra clave y de una segunda posición (P12) en función de las posiciones almacenadas;
 - 15 - una etapa de puesta de relieve (E10) de la posición de los fragmentos en la representación del contenido sonoro.
- 20 2.- Procedimiento de navegación según la reivindicación 1, caracterizado porque, a continuación de la recepción de una petición del usuario, incluye además una etapa de detención de la reproducción sonora en curso, seguida de una etapa de reproducción sonora del contenido a partir de la segunda posición.
- 25 3.- Procedimiento de navegación según la reivindicación 1, caracterizado porque la reproducción del fragmento actual va seguida de una etapa de reproducción de al menos un fragmento determinado.
- 4.- Procedimiento de navegación según la reivindicación 1, caracterizado porque el contenido sonoro está representado en forma de un círculo, representando una posición sobre el círculo una posición cronológica en el contenido, y porque la puesta de relieve incluye una etapa de presentación de un enlace entre la primera y la segunda posiciones.
- 30 5.- Procedimiento de navegación según la reivindicación 1, caracterizado porque al estar el contenido dividido en segmentos, la representación del contenido es una representación segmentada y la etapa de puesta de relieve comprende la puesta de relieve del segmento que contiene el fragmento actual y la puesta de relieve de al menos un segmento que contiene al menos un segundo fragmento.
- 35 6.- Procedimiento de navegación según la reivindicación 5, caracterizado porque un segmento representado es seleccionable a través de una interfaz de usuario.
- 40 7.- Procedimiento de navegación según la reivindicación 4, en el que la reproducción del contenido se efectúa en un conjunto de altavoces situados espacialmente sobre un círculo, caracterizado porque el fragmento actual es reproducido por un primer altavoz del conjunto, un fragmento determinado es reproducido simultáneamente por un segundo altavoz del conjunto, siendo los altavoces seleccionados en función de la posición del fragmento sobre el círculo que representa el contenido sonoro.
- 45 8.- Procedimiento de navegación según la reivindicación 7, caracterizado porque los altavoces son altavoces virtuales de un sistema de reproducción binaural.
- 9.- Procedimiento de navegación según la reivindicación 7, caracterizado porque el nivel sonoro de la señal emitida por el primer altavoz es superior al nivel sonoro de la señal emitida por el segundo altavoz.
- 50 10.- Dispositivo de navegación (NAV) en un contenido sonoro (CV), caracterizado porque incluye:
- medios de almacenamiento (MEM) de al menos una palabra clave (MC1) en asociación con al menos dos posiciones (P11, P12) representativas de al menos dos apariciones de dicha palabra clave en el contenido sonoro;
 - 55 - medios de presentación (MA) de una representación del contenido sonoro;
 - medios de detección (REC), durante la reproducción sonora del contenido, de un fragmento actual (Zc) representativo de una palabra clave almacenada en una primera posición (P11);
 - 60 - medios de determinación (DET) de al menos un segundo fragmento (Zd) representativo de dicha palabra clave y de una segunda posición (P12) en función de las posiciones almacenadas;
 - medios de puesta de relieve (MA) de la posición de los fragmentos en la representación del contenido sonoro.
- 65 11.- Sistema de navegación (SYS) que comprende un dispositivo de reproducción conforme a la reivindicación 10, y al menos dos altavoces.

12.- Producto de programa de ordenador que comprende instrucciones para llevar a cabo las etapas del procedimiento de navegación según la reivindicación 1, cuando se carga y se ejecuta por medio de un ordenador.





D2

D21

S1	MC1	ld1	V1	W1		
S2						
S3	MC2	ld2	V2	W2		
S4						
S5						
S6						
S7	MC3	ld3	V3			
MC2		MC3				
9680, 96,,						
9776, 368,,						
11600, 656, 2, ld2						

D22

D23

Fig. 6

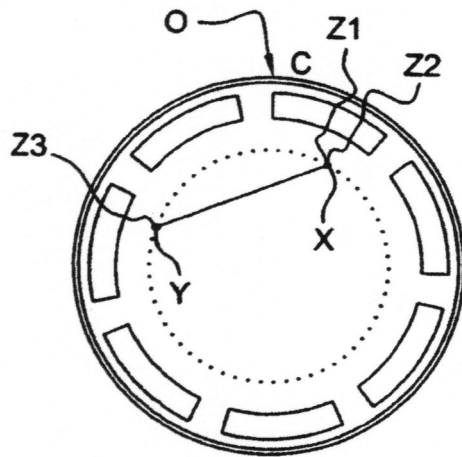


Fig. 7

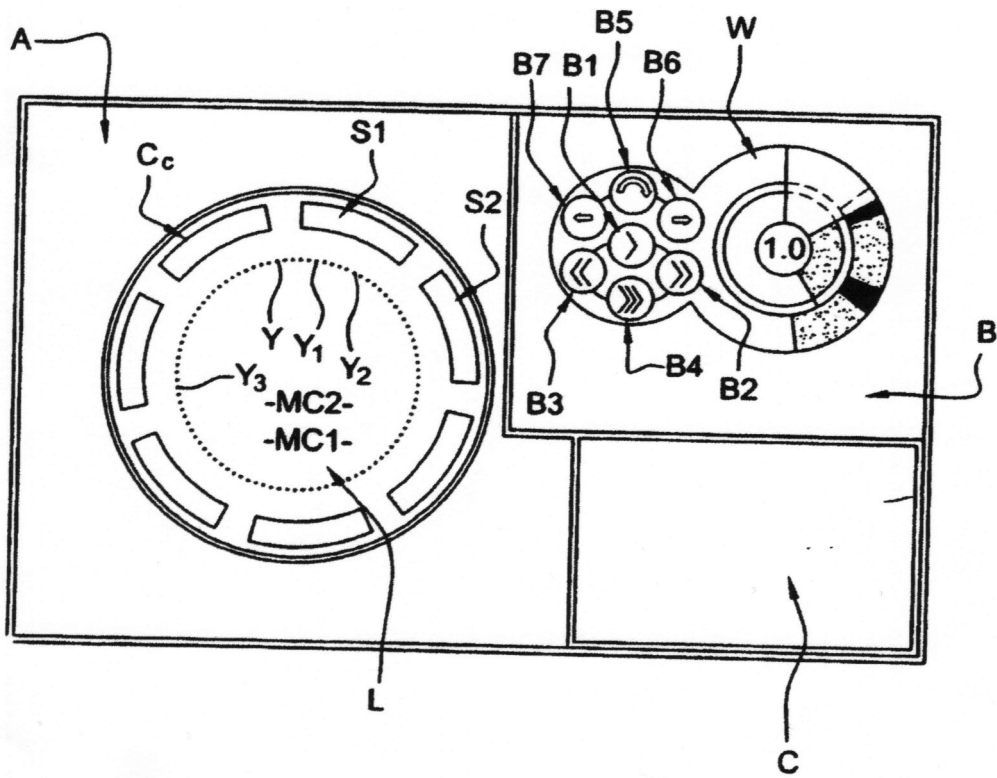


Fig. 8