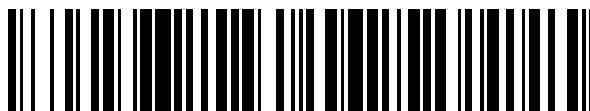


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 374 434**

51 Int. Cl.:
G10L 19/00 (2006.01)
H04S 3/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **06708241 .2**
96 Fecha de presentación: **13.02.2006**
97 Número de publicación de la solicitud: **1844465**
97 Fecha de publicación de la solicitud: **17.10.2007**

54 Título: **CODIFICACIÓN PARAMÉTRICA CONJUNTA DE FUENTES DE AUDIO.**

30 Prioridad:
14.02.2005 EP 05101055

45 Fecha de publicación de la mención BOPI:
16.02.2012

45 Fecha de la publicación del folleto de la patente:
16.02.2012

73 Titular/es:
**Fraunhofer-Gesellschaft zur Förderung der
angewandten Forschung e.V.
Hansastraße 27c
80686 München, DE**

72 Inventor/es:
FALLER, Christof

74 Agente: **Arizti Acha, Monica**

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).

ES 2 374 434 T3

DESCRIPCION

Codificación paramétrica conjunta de fuentes de audio.

1. INTRODUCCIÓN

En un problema de codificación general, existe un número de señales (mono) fuente $s_i(n)$ ($1 \leq i \leq M$) y un vector de descripción de escena $S(n)$, donde n es el índice de tiempo. El vector de descripción de escena contiene parámetros, tal como posiciones de fuente (virtuales), anchos de fuente y parámetros acústicos, tal como los parámetros del área (virtuales). La descripción de escena puede ser invariante con el tiempo o puede cambiar con el tiempo. Las señales fuente y la descripción de escena se codifican y transmiten a un descodificador. Las señales fuente codificadas, $\hat{s}_i(n)$ son mezcladas sucesivamente como una función de la descripción de escena, $\hat{S}(n)$, para generar síntesis de campo de onda, señales multicanal o estéreo, como una función del vector de descripción de escena. Las señales de salida del descodificador se denotan como $(\hat{x}_i(n)) (0 \leq i \leq N)$.

Cabe mencionar que el vector de descripción de escena $S(n)$ puede no ser transmitido, pero puede ser determinado en el descodificador. En este documento, el término de "señal de audio estéreo" siempre se refiere a señales de audio estéreo de dos canales.

Se trata ISOMEC MPEG-4 del escenario de codificación descrito. Define la descripción de escena y usa para cada señal fuente ("natural") un codificador de audio mono separado, por ejemplo, un codificador de audio de AAC. Sin embargo, cuando una escena compleja con muchas fuentes se va a mezclar, la velocidad de bits llega a ser alto, es decir, se escala en forma ascendente la velocidad de bits con el número de fuentes. La codificación de una señal fuente con alta calidad requiere de aproximadamente 60 a 90 kb/seg.

Anteriormente, se refirió a un caso especial del problema de codificación descrito [1][2] con un esquema denominado Codificación de Indicación Binaural (BCC, *Binaural Cue Coding*) para la Presentación Flexible. Al transmitir sólo la suma de las señales fuente dadas, más la información secundaria de la velocidad de bits baja, se logra esta velocidad de bits baja. Sin embargo, no pueden recuperarse las señales fuente en el descodificador y el esquema se limitó a la generación de señales circundantes estéreo y multicanales. Igualmente, sólo se utilizó la mezcla sencilla, con base en la panorámica de amplitud y retardo. Así, podía controlarse la dirección de fuentes, pero no otros atributos de imagen espacial auditivos. Otra limitación de este esquema es su calidad de audio limitada. Especialmente, hay una disminución en la calidad de audio conforme se aumenta el número de señales fuente.

El documento [1] (Codificación de Indicación Binaural, Estéreo Paramétrico, MP3 Circundante, MPEG Circundante) cubre el caso donde N canales de audio se codifican y N canales de audio, con indicaciones similares y a continuación los canales de audio originales se descodifican. La información secundaria transmitida incluye parámetros de indicación entre canales relativos a las diferencias entre los canales de entrada.

Los canales de las señales de audio estéreo y multicanal contienen mezclas de señales fuente de audio y son así diferentes en naturaleza con respecto a las señales fuente de audio puras. Las señales de audio estéreo y multicanal se mezclan de modo que cuando se reproducen en un sistema de reproducción apropiado, el oyente percibirá una imagen espacial auditivo ("etapa de sonido") tal y como se captura por el ajuste de grabación o se diseña por el ingeniero de grabación durante la mezcla. Se ha propuesto anteriormente un número de esquemas para la codificación conjunta para los canales de una señal de audio estéreo o multicanal.

Se da a conocer en la tesis doctoral titulada "Parametric Coding of Spatial Audio", de Christof Faller, Lausanne ELF, 2004, la codificación paramétrica de audio espacial utilizando indicaciones de percepción. Se introduce en un banco de filtros una señal suma transmitida y una etapa de síntesis realiza una síntesis en el dominio del banco de filtros.

Es objeto de la invención proporcionar un concepto mejorado para sintetizar una pluralidad de canales de audio.

Se alcanza este objeto mediante un método para sintetizar una pluralidad de canales de audio según la reivindicación 1, o un aparato para sintetizar una pluralidad de canales de audio según la reivindicación 13.

Preferiblemente, la invención proporciona un método para transmitir una pluralidad de señales fuente utilizando un ancho de banda mínimo. En la mayoría de los métodos conocidos, el formato de reproducción (por ejemplo estéreo, 5.1) es predefinido y tiene una influencia directa en el escenario de codificación. El flujo de audio en el lado del descodificador debe usar solamente este formato de reproducción predefinido, por lo tanto, obligando al usuario a utilizar un escenario de reproducción predefinido (por ejemplo, estéreo).

La invención propuesta codifica N señales fuente de audio, típicamente no canales de señales estéreo o multicanal, sino señales independientes, tal como diferentes señales de voz o de instrumentos. La información secundaria transmitida incluye parámetros estadísticos que se relacionan con las señales fuente de audio de entrada.

La invención propuesta descodifica M canales de audio con indicaciones diferentes con respecto a las señales fuente de audio originales. Estas indicaciones diferentes se sintetizan implícitamente mediante la aplicación de un mezclador a la señal suma recibida. El mezclador se controla como una función de la información fuente estadística recibida y los parámetros de formato de audio recibidos (o determinados de manera local) y los parámetros de mezcla. Alternativamente, estas diferentes indicaciones se calculan explícitamente como una función de la información fuente estadística recibida y los parámetros de formato de audio recibidos (o determinados de manera local) y los parámetros de mezcla. Estas indicaciones calculadas se usan para controlar un descodificador de la técnica anterior (Codificación de Indicación Binaural, Estéreo Paramétrico, MPEG Circundante) para sintetizar los canales de salida dada la señal suma recibida.

El esquema propuesto para la codificación conjunta de señales fuente de audio es el primero de su clase. Se ha diseñado para la codificación conjunta de señales fuente de audio. Las señales fuente de audio habitualmente son señales de audio mono que no son adecuadas para la reproducción mediante un sistema de audio estéreo o multicanal. Para brevedad, en lo siguiente, las señales fuente de audio son a menudo denominadas como señales fuente.

Las señales fuente de audio primero necesitan ser mezcladas a las señales de audio estéreo, multicanal o de síntesis de campo de onda anteriormente a la reproducción. Una señal fuente de audio puede ser un único instrumento o conversador, o la suma de un número de instrumentos y conversadores. Otro tipo de señal fuente de audio es una señal de audio mono capturada con un micrófono de zona durante un concierto. A menudo las señales fuente de audio se almacenan en grabadoras de múltiples pistas o en sistemas de grabación de disco duro.

El esquema reivindicado para la codificación conjunta de señales fuente de audio se basa en transmitir solamente la suma de las señales fuente de audio,

$$s(n) = \sum_{i=1}^M s_i(n) \quad , \quad (1)$$

o una suma ponderada de las señales fuente. Opcionalmente, la suma ponderada puede realizarse con diferentes pesos en diferentes subbandas y pueden adaptarse los pesos en el tiempo. Además puede aplicarse la suma con ecualización, tal y como se describe en el Capítulo 3.3.2 en [1]. En lo siguiente, al referirse a la suma o señal suma, siempre se entiende una señal generada por (1) o generada según lo descrito. Además de la señal suma, se transmite la información secundaria. La suma y la información secundaria representan el flujo de audio producido. Opcionalmente, se codifica la señal suma usando un codificador convencional de audio mono. Puede almacenarse este flujo en un archivo (CD, DVD, disco duro) o transmitirse al receptor. La información secundaria representa las propiedades estadísticas de las señales fuente que son los factores más importantes que determinan las indicaciones espaciales de percepción de las señales de salida del mezclador. Se mostrará que estas propiedades son envolventes espectrales que se desarrollan en el tiempo y funciones de autocorrelación. Se transmiten alrededor de 3 kb/seg de información secundaria por señal fuente. En el receptor, se recuperan las señales fuente $\hat{s}_i(n)$ ($1 \leq i \leq M$) con las propiedades estadísticas antes mencionadas, que se aproximan a las propiedades correspondientes de las señales fuente originales y la señal suma.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Se comprenderá mejor la invención gracias a las figuras adjuntas, en las que:

- la figura 1 muestra un esquema en el que la transmisión de cada señal fuente se hace independientemente para un procesamiento ulterior;
- la figura 2 muestra un número de fuentes transmitidas como la señal suma más la información secundaria;
- la figura 3 muestra un diagrama de bloques del esquema de Codificación de Indicación Binaural (BCC);
- la figura 4 muestra un mezclador para generar señales estéreo con base en varias señales fuente,
- la figura 5 muestra la dependencia entre ICTD, ICLD e ICC y la potencia de subbanda de la señal fuente;
- la figura 6 muestra el proceso de la generación de información secundaria;
- la figura 7 muestra el proceso de estimar los parámetros de LPC de cada señal fuente;
- la figura 8 muestra el proceso de recrear las señales fuente desde una señal suma;
- la figura 9 muestra un esquema alternativo para la generación de cada señal desde la señal suma;
- la figura 10 muestra un mezclador para generar señales estéreo basándose en la señal suma;

- la figura 11 muestra un algoritmo panorámico de amplitud que evita que los niveles fuente dependan de los parámetros de mezcla;
- la figura 12 muestra un conjunto de altavoces de un sistema de reproducción de síntesis de campo de onda;
- la figura 13 muestra cómo recuperar una estimación de las señales fuente en el receptor mediante el procesamiento de la mezcla descendente de los canales transmitidos;
- la figura 14 muestra cómo recuperar una estimación de las señales fuente en el receptor mediante el procesamiento de los canales transmitidos.

II. DEFINICIONES, NOTACIÓN Y VARIABLES

Se usan en este documento la siguiente notación y variables:

n	índice de tiempo;
i	canal de audio o índice fuente;
d	índice de retardo;
M	número de señales fuente de entrada del codificador;
N	número de canales de salida del decodificador;
$x_i(n)$	señales fuente originales mezcladas;
$\hat{x}_i(n)$	señales de salida del decodificador mezcladas;
$s_i(n)$	señales fuente de entrada del codificador;
$\hat{s}_i(n)$	señales fuente transmitidas, también llamadas señales pseudo fuente
$s(n)$	señal suma transmitida;
$y_i(n)$	señal de audio del canal L; (señal de audio a remezclar);
$\tilde{s}_i(k)$	una señal de subbanda de $s_i(n)$ (definida de manera análoga para otras señales);
$E\{\tilde{s}_i^2(n)\}$	estimación de tiempo corto de $\tilde{s}_i^2(n)$ (definida de manera análoga para otras señales);
ICLD	diferencia de nivel entre canales;
ICTD	diferencia de tiempo entre canales;
ICC	coherencia entre canales
$\Delta L(n)$	ICLD de subbanda estimada;
$\tau(n)$	ICTD de subbanda estimada;
$c(n)$	ICC de subbanda estimada;
$\tilde{p}_i(n)$	potencia de subbanda de fuente relativa;
a_i, b_i	factores de escala del mezclador;
c_i, d_i	retardos del mezclador;
$\Delta L_i, \tau(n)$	nivel del mezclador y diferencia de tiempo;

G_i ganancia fuente del mezclador

III. CODIFICACIÓN CONJUNTA DE SEÑALES FUENTE DE AUDIO

Primero, se describe la Codificación de Indicación Binaural (BCC), una técnica de codificación de audio multicanal paramétrica. A continuación se muestra que con la misma perspicacia en la que se basa la BCC, es posible concebir un algoritmo para la codificación conjunta de las señales fuente para un escenario de codificación.

A. Codificación de indicación binaural (BCC)

Se muestra en la figura más adelante un esquema de BCC [1][2] para una codificación de audio multicanal. Se mezcla en forma descendente la señal de audio multicanal de entrada a un solo canal. A diferencia de la información de codificación y transmisión sobre todas las formas de onda del canal, se codifica y transmite solamente la señal mezclada en forma descendente (con el codificador de audio mono convencional). Adicionalmente, se estiman las "diferencias de canal de audio" con motivos de percepción entre los canales de audio originales y también se transmiten al descodificador. El descodificador genera sus canales de salida de modo que las diferencias de canal de audio se aproximen a las diferencias de canal de audio correspondientes de la señal de audio original.

La localización de suma implica que las diferencias de canal de audio relevantes con percepción para una pareja de canales de señal de altavoz son la diferencia de tiempo entre canales (ICTD) y la diferencia de nivel entre canales (ICLD). ICTD e ICLD pueden estar relacionadas a la dirección percibida de eventos auditivos. Otros atributos de imagen espacial auditiva, tal como un ancho de fuente aparente y el envolvimiento del oyente, pueden estar relacionados con la coherencia interaural (IC). Para parejas de altavoces delante o detrás de un oyente, la coherencia interaural está a menudo relacionada directamente a la coherencia entre canales (ICC), que es así considerada como la tercera medida de la diferencia de canal de audio por BCC. Se estiman la ICTD, ICLD e ICC en las subbandas como una función del tiempo. Tanto la resolución espectral como temporal que se usan son motivadas por percepción.

B. Codificación conjunta paramétrica de fuentes de audio

Un descodificador BCC es capaz de generar una señal de audio multicanal con cualquier imagen espacial auditiva tomando una señal mono y sintetizando en intervalos de tiempo regulares una sola indicación de ICTD, ICLD e ICC específica por subbanda y pareja de canales. El buen desempeño de los esquemas de BCC para un amplio intervalo de material de audio [véase 1] implica que la imagen espacial auditiva percibida se determina en gran medida por la ICTD, ICLD e ICC. Por lo tanto, a diferencia de exigir las señales fuente "limpias" $s_i(n)$ como entrada del mezclador en la figura 1, sólo se exigen señales pseudo fuente $\hat{s}_i(n)$ con la particularidad de que ellas resulten en ICTD, ICLD e ICC en la salida del mezclador similares al caso de suministro de señales fuente reales al mezclador. Hay tres objetivos para la generación de $\hat{s}_i(n)$:

- Si se suministran $\hat{s}_i(n)$ a un mezclador, los canales de salida del mezclador tendrán aproximadamente las mismas indicaciones espaciales (ICLD, ICTD, ICC) como si se suministraran $s_i(n)$ al mezclador.
- Se van a generar $\hat{s}_i(n)$ con tan poca información como sea posible acerca de las señales fuente originales $s(n)$ (debido a el objetivo es tener la información secundaria de la velocidad de bits baja).
- Las $\hat{s}_i(n)$ se generan desde la señal suma transmitida $s(n)$ de modo que sea introducida una cantidad mínima de distorsión de señal.

Para proporcionar el esquema propuesto, se considera un mezclador estéreo ($M=2$). Una simplificación más sobre el caso general es que sólo se aplican la panorámica de amplitud y retardo para la mezcla. Si las señales fuente discretas estuvieran disponibles en el descodificador, se mezclaría una señal estéreo tal como se muestra en la figura 4, es decir,

$$x_1(n) = \sum_{i=1}^M a_i s_i(n - c_i) \quad x_2(n) = \sum_{i=1}^M b_i s_i(n - d_i) \quad (2)$$

En este caso, el vector de descripción de escena $S(n)$ contiene solamente direcciones fuente que determinan los parámetros de mezcla,

$$\mathbf{M}(n) = (a_1, a_2, \dots, a_M, b_1, b_2, \dots, b_M, c_1, c_2, \dots, c_M, d_1, d_2, \dots, d_M)^T \quad (3)$$

donde T es la transposición de un vector. Cabe mencionar que para los parámetros de mezcla, se hizo caso omiso al índice de tiempo por conveniencia de notación.

- 5 Los parámetros más convenientes para controlar el mezclador son la diferencia de tiempo y de nivel, T_i y ΔL_i , que se relacionan a a_i , b_i , c_i y d_i por

$$a_i = \frac{10^{G_i/20}}{\sqrt{1+10^{\Delta L_i/10}}} \quad b_i = 10^{(G_i+\Delta L_i)/20} a_i \quad c_i = \max \{-T_i, 0\} \quad d_i = \max \{T_i, 0\} \quad (4)$$

donde G_i es un factor de ganancia fuente en dB.

- 10 En lo siguiente, se calcularán ICTD, ICLD e ICC de la salida del mezclador estéreo como una función de las señales fuente de entrada $s_i(n)$. Las expresiones obtenidas indicarán las propiedades de señales fuentes que determinan ICTD, ICLD e ICC (junto con los parámetros de mezcla). Entonces se generan $\hat{s}_i(n)$ de modo que las propiedades de la señal fuente identificadas se aproximen a las propiedades correspondientes de las señales fuente originales.

B.1 ICTD, ICLD e ICC de la salida del mezclador.

- 15 Se estiman las indicaciones en subbandas y como una función del tiempo. En lo siguiente, se supone que las señales fuente $s_i(n)$ son cero y mutuamente independientes. Una pareja de señales de subbanda de la salida (2) del mezclador se denota $\hat{x}_1(n)$ y $\hat{x}_2(n)$. Cabe mencionar que para simplicidad de notación se utilizan el mismo índice de tiempo n para señales de dominio de tiempo y de dominio de subbanda. Igualmente no se utiliza un índice de subbanda y se aplica el análisis/procesamiento descrito a cada subbanda independientemente. La potencia de la subbanda de las dos señales de salida del mezclador es:

$$E\{\tilde{x}_1^2(n)\} = \sum_{i=1}^M a_i^2 E\{\tilde{s}_i^2(n)\} \quad E\{\tilde{x}_2^2(n)\} = \sum_{i=1}^M b_i^2 E\{\tilde{s}_i^2(n)\} \quad (5)$$

donde $\tilde{s}_i(n)$ es una señal de subbanda de la fuente $s_i(n)$ y $E\{\cdot\}$ denota la expectación de tiempo corto, por ejemplo,

$$E\{\tilde{s}_2^2(n)\} = \frac{1}{K} \sum_{n=K/2}^{n+K/2-1} \tilde{s}_2^2(n) \quad (6)$$

- 25 donde K determina la longitud del promedio móvil. Cabe mencionar que los valores de potencia de subbanda $E\{\tilde{s}_2^2(n)\}$ representan para cada señal fuente el envolvente espectral como una función de tiempo. La ICLD, $\Delta L(n)$, es

$$\Delta L(n) = 10 \log_{10} \frac{\sum_{i=1}^M b_i^2 E\{\tilde{s}_i^2(n)\}}{\sum_{i=1}^M a_i^2 E\{\tilde{s}_i^2(n)\}} \quad (7)$$

Para estimar la ICTD e ICC, se estima la función de correlación cruzada normalizada

$$\Phi(n,d) = \frac{E\{\tilde{x}_1(n)\tilde{x}_2(n+d)\}}{\sqrt{E\{\tilde{x}_1^2(n)\}E\{\tilde{x}_2^2(n+d)\}}}$$
(8)

Se calcula la IC, $c(n)$, de acuerdo con

$$c(n) = \max_d \Phi(n,d)$$
(9)

Para calcular la ICTD, $T(n)$, se calcula la ubicación del pico más alto en el eje de retardo,

$$T(n) = \arg \max_d \Phi(n,d)$$
(10)

Ahora la cuestión es cómo puede calcularse la función de correlación cruzada normalizada como una función de los parámetros de mezcla. Junto con (2), (8) puede escribirse como

$$\Phi(n,d) = \frac{\sum_{i=1}^M E\{a_i b_i \tilde{s}_i(n - c_i) \tilde{s}_i(n - d_i + d)\}}{\sqrt{E\{\sum_{i=1}^M a_i^2 s_i^2(n - c_i)\} E\{\sum_{i=1}^M b_i^2 s_i^2(n - d_i)\}}}$$
(11)

que es equivalente a

$$\Phi(n,d) = \frac{\sum_{i=1}^M a_i b_i E\{\tilde{s}_i^2(n)\} \Phi_i(n, d_i - T_i)}{\sqrt{(\sum_{i=1}^M a_i^2 E\{\tilde{s}_i^2(n)\}) (\sum_{i=1}^M b_i^2 E\{s_i^2(n)\})}}$$
(12)

donde la función de autocorrelación normalizada $\varphi(n,e)$ es

$$\Phi(n,\theta) = \frac{E\{s_i(n)s_i(n+e)\}}{E\{s_i^2(n)\}}$$
(13)

y $T_i = d_i - c_i$. Cabe mencionar que para calcular (12) dada (11) se ha supuesto que las señales son estacionarias de sentido amplio dentro del intervalo considerado de retardos, es decir,

$$E \{ \tilde{s}_i^2(n) \} = E \{ \tilde{s}_i^2(n - c_i) \}$$

$$E \{ \tilde{s}_i^2(n) \} = E \{ \tilde{s}_i^2(n - d_i) \}$$

$$E \{ \tilde{s}_i(n) \tilde{s}_i(n + c_i - d_i + d) \} = E \{ \tilde{s}_i(n - c_i) \tilde{s}_i(n - d_i + d) \}$$

Se muestra en la figura 5 un ejemplo numérico para dos señales fuente, que ilustra la dependencia entre ICTD, ICLD e ICC y la potencia de subbanda de fuente. Se muestra en las partes superior, intermedia e inferior de la figura 5 $\Delta L(n)$, $T(n)$ y $c(n)$, respectivamente, como una función de la relación de la potencia de subbanda de las dos señales fuente $a = E \{ \tilde{s}_1^2(n) \} / E \{ \tilde{s}_1^2(n) \} + E \{ \tilde{s}_2^2(n) \}$ para diferentes parámetros de mezcla (4) ΔL_1 , ΔL_2 , T_1 y T_2 . Cabe mencionar que cuando sólo una fuente tiene potencia en la subbanda ($a = 0$ ó $a = 1$), entonces $\Delta L(n)$ y $T(n)$ calculados son iguales a los parámetros de mezcla ($\Delta L_1, \Delta L_2, T_1, T_2$).

B.2. Información secundaria necesaria

La ICLD (7) depende de los parámetros de mezcla (a_i, b_i, c_i, d_i) y en la potencia de subbanda de corto tiempo de las fuentes, $E \{ \tilde{s}_i^2(n) \}$ (6). La función de correlación cruzada de subbanda normalizada $\Phi(n, d)$ (12) que es necesaria para calcular la ICTD (10) e ICC (9) depende de $E \{ \tilde{s}_i^2(n) \}$ y adicionalmente de la función de autocorrelación de subbanda normalizada, $\Phi_i(n, e)$ (13), para cada señal fuente. El máximo de $\Phi(n, d)$ está dentro del intervalo $\min_i \{T_i\} \leq d \leq \max_i \{T_i\}$. Para fuente i con parámetro de mezclador $T_i = c_i - d_i$, el intervalo correspondiente para el cual se necesita la propiedad de la subbanda de señal fuente $\Phi_i(n, e)$ (13) es

$$\min_i \{T_i\} - T_i \leq e \leq \max_i \{T_i\} - T_i \quad (14)$$

Puesto que las indicaciones de ICTD, ICLD e ICC dependen de las propiedades de subbanda de la señal fuente $E \{ \tilde{s}_i^2(n) \}$ y $\Phi_i(n, e)$ en el intervalo (14), en principio es necesario transmitir esas propiedades de subbanda de la señal fuente como información secundaria. Se supone que cualquier otra clase de mezclador (por ejemplo, mezclador con efectos, mezclador/dispositivo de convolución de síntesis de campo de onda, etc.) tiene propiedades similares y así esta información secundaria es útil también cuando se utilizan otros mezcladores además del descrito. Para reducir la cantidad de la información secundaria, puede almacenarse un conjunto de funciones de autocorrelación predefinidas en el descodificador y solamente se transmiten índices de selección de aquellas que corresponden más estrechamente con las propiedades de la señal fuente. Una primera versión del presente algoritmo supone que dentro del intervalo (14) $\Phi(n, e) = 1$ y así (12) se calcula utilizando sólo los valores de potencia de subbanda (6) como información secundaria. Se han calculado los datos mostrados en la figura 5 suponiendo $\Phi(n, e) = 1$.

Con el fin de reducir la cantidad de información secundaria, se limita el intervalo dinámico relativo de las señales fuente. En cada momento, para cada subbanda, se selecciona la potencia de la fuente más fuerte. Es suficiente limitar inferiormente la potencia de la subbanda correspondiente de todas las otras fuentes a un valor de 24 dB más bajo que la potencia de subbanda más fuerte. Así, el intervalo dinámico del cuantificador puede limitarse a 24 dB.

Suponiendo que las señales fuente son independientes, el descodificador puede calcular la suma de la potencia de subbanda de todas las fuentes como $E \{ \tilde{s}^2(n) \}$. Así, en principio es suficiente transmitir al descodificador sólo los valores de potencia de subbanda de $M-1$ fuentes, mientras la potencia de subbanda de la fuente restante puede calcularse de manera local. Dada esta idea, la velocidad de información secundaria puede reducirse levemente mediante la transmisión de la potencia de subbanda de fuentes con índices $2 \leq i \leq M$ con relación a la potencia de la primera fuente.

$$\Delta \tilde{p}_i(n) = 10 \log_{10} \frac{E\{\tilde{s}_i^2(n)\}}{E\{\tilde{s}_1^2(n)\}}. \quad (15)$$

5 Cabe mencionar que el intervalo dinámico que limita como se describió anteriormente se lleva a cabo antes de (15). Como una alternativa, los valores de potencia de subbanda pueden normalizarse con relación a la potencia de subbanda de la señal suma, a diferencia de la normalización con relación a una potencia de subbanda de una fuente (15). Para una frecuencia de muestreo de 44.1 kHz, se utilizan 20 subbandas y se transmiten para cada subbanda $\Delta \tilde{p}_i(n)$ ($2 \leq i \leq M$) alrededor de cada 12 ms. 20 subbandas corresponden a la mitad de la resolución espectral del sistema auditivo (una subbanda es de una amplitud de dos “anchos de banda críticos”). Los experimentos informales indican que sólo se logra una leve mejora usando más de 20 subbandas, por ejemplo 40 subbandas. El número de subbandas y los anchos de banda de estas subbandas se escogen de acuerdo con la resolución de frecuencia y tiempo del sistema auditivo. Una implementación de baja calidad del esquema exige al menos tres subbandas (frecuencias baja, media y alta).

De acuerdo con una realización particular, las subbandas tienen anchos de banda diferentes, las subbandas a frecuencias más bajas tienen anchos de banda menores que las subbandas a frecuencias más altas.

15 Los valores de potencia relativos se cuantifican con un esquema similar al cuantificador de ICLD descrito en [2], que resulta en una velocidad de bits de aproximadamente $3(M-1)$ kb/s. La figura 6 ilustra el proceso de la generación de información secundaria (que corresponde al bloque de “generación de información secundaria” en la figura 2).

La velocidad de información secundaria puede reducirse adicionalmente mediante el análisis de actividad para cada señal fuente y la transmisión únicamente de la información secundaria asociada con la fuente si es activa.

20 A diferencia de transmitir los valores de potencia de subbanda $E\{\tilde{s}_i^2(n)\}$ como información estadística, puede transmitirse otra información que representa los envolventes espectrales de las señales fuente. Por ejemplo, pueden transmitirse los parámetros de la codificación predictiva lineal (LPC) u otros parámetros correspondientes, tal como los parámetros del filtro de red o los parámetros de la pareja espectral de línea (LSP). El proceso de estimar los parámetros LPC de cada señal fuente se ilustra en la figura 7.

B.3 Cálculo de $\hat{s}_i(n)$

25 La figura 8 ilustra el proceso que se usa para recrear las señales fuente, dada la señal suma (1). Este proceso forma parte del bloque de “Síntesis” en la figura 2. Las señales fuente individuales se recuperan escalando cada subbanda de la señal suma con $g_i(n)$ y aplicando un filtro de descorrelación con respuesta de impulso $h_i(n)$,

$$\hat{s}_i(n) = h_i(n) * (g_i(n) \tilde{s}(n)) = h_i(n) * \left(\sqrt{\frac{E\{\tilde{s}_i^2(n)\}}{E\{\tilde{s}_1^2(n)\}}} \tilde{s}(n) \right) \quad (16)$$

donde * es el operador de convolución lineal y se calcula $E\{\tilde{s}_i^2(n)\}$ con la información secundaria mediante

$$30 \quad E\{\tilde{s}_i^2(n)\} = 1 / \sqrt{1 + \sum_{i=2}^M 10^{\frac{\Delta \tilde{p}_i(n)}{10}}} \text{ para } i = 1 \text{ ó } 10^{\frac{\Delta \tilde{p}_i(n)}{10}} E\{\tilde{s}_1^2(n)\} \text{ de otra forma} \quad (17)$$

35 Como filtros de descorrelación $h_i(n)$, pueden utilizarse filtros de peine complementarios, filtros de todo paso, retardos o filtros con respuestas de impulso aleatorio. El objetivo para el procesamiento de descorrelación es reducir la correlación entre las señales sin modificar cómo se perciben las formas de onda individuales. Diferentes técnicas de descorrelación causan diferentes artefactos. Los filtros de peine complementarios causan coloración. Todas las técnicas descritas diseminan la energía de transitorios en el tiempo que causan artefactos, tal como los “ecos previos”. Dado su potencial para artefactos, las técnicas de descorrelación deben aplicarse tan poco como sea posible. La siguiente sección describe técnicas y estrategias que requieren menos procesamiento de descorrelación que la simple generación de señales independientes $\hat{s}_i(n)$.

Un esquema alternativo para la generación de señales $\hat{s}_i(n)$ se muestra en la figura 9. Primero se aplanan el espectro de $s(n)$ por medio del cálculo del error de predicción lineal $e(n)$. Entonces, dados los filtros de LPC estimados en el codificador, $\bar{f}_i$, se calculan los filtros de todos los polos correspondientes como la transformada z inversa de

$$\bar{F}_i(z) = \frac{1}{1 - z^{-1}F_i(z)}$$

- 5 Los filtros resultantes de todos los polos, $\bar{f}_i$, representan el envolvente espectral de las señales fuente. Si se transmite otra información secundaria distinta a los parámetros LPC, primero es necesario calcular los parámetros LPC como una función de la información secundaria. Igual que en el otro esquema, se utilizan los filtros de descorrelación h_i para hacer que las señales fuente sean independientes.

IV. IMPLEMENTACIONES CONSIDERANDO RESTRICCIONES PRÁCTICAS

- 10 En la primera parte de esta sección, se proporciona un ejemplo de implementación usando un esquema de síntesis BCC como un mezclador estéreo o multicanal. Esto es particularmente interesante puesto que tal esquema de síntesis de tipo BCC forma parte de un estándar ISO/IEC MPEG próximo, denominado "codificación de audio espacial". No se calculan explícitamente las señales fuentes $\hat{s}_i(n)$ en este caso, dando como resultado una complejidad de cálculo reducida. Igualmente, este esquema ofrece el potencial de mejor calidad de audio, puesto que efectivamente se
- 15 necesita menos descorrelación que en el caso en el que se calculan explícitamente las señales fuente $\hat{s}_i(n)$.

La segunda parte de esta sección se dirige a los problemas cuando se aplica el esquema propuesto con cualquier mezclador y no se aplica ningún procesamiento de descorrelación. Tal esquema tiene una menor complejidad que un esquema con procesamiento de descorrelación, pero puede tener otros inconvenientes, tal como se describirán más adelante.

- 20 Idealmente, sería deseable aplicar el procesamiento de descorrelación de modo que la $\hat{s}_i(n)$ generada pueda considerarse independiente. Sin embargo, puesto que el procesamiento de descorrelación es problemático en términos de introducir artefactos, sería deseable aplicar tan poco procesamiento de descorrelación como sea posible. La tercera parte de esta sección se dirige a cómo puede reducirse la cantidad de procesamiento de descorrelación problemático mientras otorga beneficios como si la $\hat{s}_i(n)$ generada fuera independiente.

- 25 A. Implementación sin cálculo explícito de $\hat{s}_i(n)$

Se aplica la mezcla directamente a la señal suma transmitida (1), sin explicitarse cálculo explícito de $\hat{s}_i(n)$. Se utilizó un esquema de síntesis de BCC para este propósito. En lo siguiente, se considera el caso de estéreo, pero todos los principios descritos pueden aplicarse también para la generación de señales de audio multicanal.

- 30 Se muestra en la figura 10 un esquema de síntesis de BCC de estéreo (o un esquema "estéreo paramétrico"), aplicado para procesar la señal suma (1). Sería deseable que el esquema de síntesis de BCC genere una señal que se percibe de manera similar a la señal de salida de un mezclador, tal como se muestra en la figura 4. Esto es así cuando ICTD, ICLD e ICC en los canales de salida del esquema de síntesis son similares, como las indicaciones correspondientes que aparecen entre los canales de la señal de salida (4) del mezclador.

- Se utiliza la misma información secundaria que en el esquema general anteriormente descrito, permitiendo que
- 35 el descodificador calcule los valores de potencia de subbanda de tiempo corto $E\{\tilde{s}_i^2(n)\}$ de las fuentes. Dada $E\{\tilde{s}_i^2(n)\}$, se calculan los factores de ganancia g_1 y g_2 en la figura 10 como

$$g_1(n) = \sqrt{\frac{\sum_{i=1}^M a_i^2 E\{\tilde{s}_i^2(n)\}}{E\{\tilde{s}^2(n)\}}} \quad g_2(n) = \sqrt{\frac{\sum_{i=1}^M b_i^2 E\{\tilde{s}_i^2(n)\}}{E\{\tilde{s}^2(n)\}}} \quad (18)$$

de modo que la potencia de subbanda de salida e ICLD (7) sean los mismos como para el mezclador en la figura 4. Se calcula la ICTD $T(n)$ de acuerdo con (10), determinando los retardos D_1 y D_2 en la figura 10,

$$D_1(n) = \max\{-T(n), 0\} \quad D_2(n) = \max\{T(n), 0\} \quad (19)$$

Se calcula la ICC $c(n)$ de acuerdo con (9), determinando el procesamiento de descorrelación en la figura 10.

5 Se describe el procesamiento de descorrelación (síntesis de ICC) en [1]. Las ventajas de aplicar el procesamiento de descorrelación a los canales de salida del mezclador frente a aplicarlo para generar independientes $\hat{S}_i(n)$ son:

- Normalmente, el número de señales fuente M es mayor que el número de canales de salida de audio N . Así, el número de canales de audio independientes que se necesitan generar es menor cuando se descorrelacionan N canales de salida a diferencia de la descorrelación de las M señales fuente.

10 • A menudo, se correlacionan los N canales de salida de audio ($ICC > 0$) y puede aplicarse menos procesamiento de descorrelación que sería necesario para generar M o N canales independientes.

Se espera mejor calidad de audio debido a menos procesamiento de descorrelación.

15 Se espera la mejor calidad de audio cuando se limitan los parámetros del mezclador de modo que $a_i^2 + b_i^2 = 1$, es decir, $G_i = 0$ dB. En este caso, la potencia de cada fuente en la señal suma transmitida (1) es idéntica a la potencia de la misma fuente en la señal de salida del decodificador mixta. La señal de salida del decodificador (figura 10) es la misma como si la señal de salida del mezclador (figura 4) se codificara y decodificara por un codificador/decodificador de BCC en este caso. Así, puede esperarse una calidad igualmente similar.

20 El decodificador puede no sólo determinar la dirección en la que cada fuente va a aparecer, sino también puede variarse la ganancia de cada fuente. Se aumenta la ganancia eligiendo $a_i^2 + b_i^2 > 1$ ($G_i > 0$ dB) y se disminuye eligiendo $a_i^2 + b_i^2 < 1$ ($G_i < 0$ dB).

B. Uso de ningún procesamiento de descorrelación

La restricción de la técnica anteriormente descrita es que la mezcla se lleva a cabo sin el esquema de síntesis de BCC. Puede imaginarse la implementación no sólo de las síntesis de ICTD, ICLD e ICC, sino adicionalmente el procesamiento de efectos dentro de la síntesis de BCC.

25 Sin embargo, puede ser deseable que puedan utilizarse los mezcladores y procesadores de efectos existentes. Esto también incluye los mezcladores de síntesis de campo de onda (a menudo denominados como "dispositivos de convolución"). Para el uso de los mezcladores y procesadores de efectos existentes, se calculan las $\hat{S}_i(n)$ explícitamente y se utilizan como si fueran las señales fuente originales.

30 Cuando no se aplica ningún procesamiento de descorrelación ($h_i(n) = \delta(n)$ en (16) también puede lograrse buena calidad de audio. Es un compromiso entre los artefactos introducidos debidos al procesamiento de descorrelación y los artefactos debidos al hecho de que las señales fuente $\hat{S}_i(n)$ se correlacionan. Cuando no se utiliza ningún procesamiento de descorrelación, la imagen espacial auditiva resultante puede sufrir de inestabilidad [1]. Pero puede introducir el propio mezclador alguna descorrelación cuando se utilizan reverberadores u otros efectos y así hay menos necesidad de procesamiento de descorrelación.

35 Si se generan $\hat{S}_i(n)$ sin procesamiento de descorrelación, el nivel de las fuentes depende de la dirección en la que se mezclan en relación con otras fuentes. Mediante la sustitución de algoritmos panorámicos de amplitud en los mezcladores existentes con un algoritmo que compensa esta dependencia de nivel, puede evitarse el efecto negativo de la dependencia de la sonoridad en los parámetros de mezcla. Se muestra en la figura 11 un algoritmo de amplitud de compensación de nivel, el cual se dirige a compensar la dependencia del nivel de fuente en los parámetros de mezcla.

40 Dados los factores de ganancia de un algoritmo panorámico de amplitud (por ejemplo, la figura 4), a_i y b_i , se calculan los pesos de la figura 11, $\bar{a}_i$ y $\bar{b}_i$, mediante

$$\bar{a}_i(n) = \sqrt{\frac{\sum_{i=1}^M a_i^2 E\{\tilde{s}_i^2(n)\}}{E\{(\sum_{i=1}^M a_i \tilde{s}_i(n))^2\}}} \quad \text{y} \quad \bar{b}_i(n) = \sqrt{\frac{\sum_{i=1}^M b_i^2 E\{\tilde{s}_i^2(n)\}}{E\{(\sum_{i=1}^M b_i \tilde{s}_i(n))^2\}}} \quad (20)$$

Cabe mencionar que se calculan $\bar{a}_i$ y $\bar{b}_i$ de modo que la potencia de subbanda de salida sea la misma como si $\hat{s}_i(n)$ fueran independientes en cada subbanda.

c. Reducción de la cantidad de procesamiento de decorrelación

- 5 Tal como se mencionó anteriormente, la generación de las $\hat{s}_i(n)$ independientes es problemática. Se describen en el presente documento estrategias para aplicar menor procesamiento de decorrelación mientras se logra efectivamente un efecto similar como si $\hat{s}_i(n)$ fueran independientes.

10 Se contempla, por ejemplo, un sistema de síntesis de campo de ondas, tal como se muestra en la figura 12. Se indican las posiciones de fuente virtuales deseadas para $s_1, s_2, \dots, s_6$ ($M = 6$). Una estrategia para calcular $\hat{s}_i(n)$ (16) sin generar M señales completamente independientes es:

- 1) generar grupos de índices de fuentes que corresponden a las fuentes cercanas entre sí. Por ejemplo, en la figura 8 podrían ser $\{1\}, \{2,5\}, \{3\}$, and $\{4,6\}$.
- 2) En cada momento en cada subbandas, seleccionar el índice de fuente de la fuente más fuerte

$$i_{\max} = \max_i E\{\bar{s}(n)\} \quad (21)$$

- 15 No se aplica ningún procesamiento de decorrelación para la parte de los índices de fuentes que contiene $i_{\max}$, es decir, $h_i(n) = \delta(n)$.

- 3) Para cada otro grupo, seleccionar el mismo $h_i(n)$ dentro del grupo.

20 El algoritmo descrito modifica las componentes de señal más fuertes a menor medida. Adicionalmente, se reduce el número de diferentes $h_i(n)$ que se utilizan. Esto es una ventaja debido a que la decorrelación es más fácil cuanto menos canales independientes que se necesitan generar. También es aplicable la técnica descrita cuando se mezclan señales de audio estéreo o multicanal.

V. ESCALABILIDAD EN TÉRMINOS DE CALIDAD Y VELOCIDAD DE BITS.

25 El esquema propuesto transmite solamente la suma de todas las señales fuentes, la cual puede codificarse con un codificador de audio mono convencional. Cuando ninguna compatibilidad hacia atrás es necesaria y la capacidad está disponible para la transmisión/almacenamiento de más de una forma de onda de audio, el esquema propuesto puede ser escalado para el uso con más de un canal de transmisión. Esto se realiza generando varias señales suma con diferentes subconjuntos de las señales fuente dadas, es decir, a cada subconjunto de las señales fuente se aplica individualmente el esquema de codificación propuesto. Se espera que la calidad de audio mejore conforme se aumenta el número de canales de audio transmitidos debido a que tienen que generarse canales menos independientes mediante la decorrelación desde cada canal transmitido (comparado con el caso de un canal transmitido).

VI. COMPATIBILIDAD HACÍA ATRÁS CON FORMATOS DE AUDIO ESTÉREO Y CIRCUNDANTES EXISTENTES

35 Considérese el siguiente escenario de entrega de audio. Un consumidor obtiene una señal circundante estéreo o multicanal de máxima calidad (por ejemplo, por medio de un CD, DVD de audio o tienda de música en Internet, etc.) El objetivo es entregar opcionalmente al consumidor la flexibilidad de generar una mezcla personalizada del contenido de audio obtenido, sin comprometer la calidad de reproducción estándar en estéreo/circundante.

Esto se realiza entregando al consumidor (por ejemplo, como una opción de compra en una tienda de música en Internet) un flujo de bits de información secundaria que permite el cálculo de $\hat{s}_i(n)$ como una función de la señal de

audio dada estéreo o multicanal. Entonces se aplica el algoritmo de mezcla del consumidor a $\hat{S}_i(n)$. Se describen en lo siguiente dos posibilidades para calcular $\hat{S}_i(n)$ dadas las señales de audio estéreo o multicanal.

A. Estimación de la suma de las señales fuente en el receptor

- 5 Se ilustra en la figura 13 la manera más directa de usar el esquema de codificación propuesto con una transmisión de audio estéreo o multicanal, donde $y_i(n)$ ($1 \leq i \leq L$) son los canales L de la señal de audio estéreo o multicanal. Se estima la señal suma de las fuentes mediante la mezcla descendente de los canales transmitidos a un único canal de audio. Se lleva a cabo la mezcla descendente por medio del cálculo de la suma de los canales $y_i(n)$ ($1 \leq i \leq L$) o pueden aplicarse técnicas más sofisticadas.

- 10 Para el mejor rendimiento, se recomienda que el nivel de las señales fuentes se adapte antes de la estimación $E\{\hat{S}_i^2(n)\}$ (6) de modo que la relación de potencia entre las señales fuentes se aproxime a la relación de potencia con la cual las fuentes están contenidas en la señal estéreo o multicanal dada. En este caso, la mezcla descendente de los canales transmitidos es una estimación relativamente buena de la suma de las fuentes (1) (o una versión escalada de la misma).

- 15 Puede utilizarse un proceso automatizado para ajustar el nivel de las señales fuente de entrada del codificador $S_i(n)$ antes de calcular la información secundaria. Este proceso estima de manera adaptada en el tiempo el nivel al que cada señal fuente está contenida en la señal estéreo o multicanal dada. Antes de calcular la información secundaria, entonces se ajusta de manera adaptada en el tiempo el nivel de cada señal fuente de modo que sea igual al nivel en el que la fuente está contenida en la señal de audio estéreo o multicanal.

B. Uso de los canales transmitidos individualmente

- 20 La figura 14 muestra una implementación diferente del esquema propuesto con la transmisión de señal circundante estéreo o multicanal. Aquí, no se mezclan en forma descendente los canales transmitidos, pero se usan individualmente para la generación de $\hat{S}_i(n)$. Más generalmente, se calculan las señales de subbanda $\hat{S}_i(n)$ mediante

$$\hat{S}_i(n) = h_i(n) * (g_i(n) \sum_{l=1}^L w_l(n) \tilde{y}_l(n)) \quad (22)$$

- 25 donde $w_l(n)$ son pesos que determinan las combinaciones lineales específicas de las subbandas de los canales transmitidas. Se eligen las combinaciones lineales de modo que las $\hat{S}_i(n)$ estén siempre descorrelacionadas en lo posible. Así, es necesario aplicar ninguno o sólo una pequeña cantidad de procesamiento de descorrelación, lo cual es favorable, tal como se mencionó anteriormente.

VII. APLICACIONES

- 30 Anteriormente se ha mencionado un número de aplicaciones para los esquemas de codificación propuestos. Aquí, se resumen y se mencionan unas cuantas aplicaciones más.

A. Codificación de audio para la mezcla

- 35 Siempre que las señales fuente de audio necesiten ser almacenadas o transmitidas antes de mezclarlas a las señales de audio estéreo, multicanal o de síntesis de campo de onda, se puede aplicar el esquema propuesto. Con la técnica anterior, podría aplicarse un codificador de audio mono a cada señal fuente independientemente, que resulta en una velocidad de bits que escala con el número de fuentes. El esquema de codificación propuesto puede codificar un alto número de señales fuente de audio con un único codificador de audio mono más la información secundaria de velocidad de bits relativamente baja. Tal como se describe en la Sección V, puede mejorarse la calidad de audio puede usando más de un canal transmitido si se dispone de memoria/capacidad para hacerlo.

B. Remezcla con metadatos

- 40 Tal como se describió en la Sección VI, pueden remezclarse las señales de audio estero y multicanal existentes con la ayuda de información secundaria adicional (es decir "metadatos"). A diferencia de vender solamente el contenido de audio mezclado estéreo y multicanal optimizado, también pueden venderse los metadatos permitiendo a un usuario la remezcla de su música estéreo y multicanal. Esto puede, por ejemplo, también ser usado para atenuar los sonidos vocales en una canción para karaoke o para atenuar instrumentos específicos para tocar un instrumento junto con la música.

Aunque no fuera un problema el almacenamiento, el esquema descrito sería muy atractivo para habilitar la mezcla personalizada de música. Es decir, debido a que es probable que la industria de la música nunca esté dispuesta a regalar las grabaciones de múltiples pistas. Existe demasiado peligro para el abuso. El esquema propuesto habilita la capacidad de la remezcla sin regalar las grabaciones de múltiples pistas.

- 5 Además, tan pronto como se remezclan las señales estéreo o multicanal, ocurre cierto grado de reducción de calidad, haciendo que la distribución ilegal de la remezcla sea menos atractiva.

c. Conversión de estéreo/multicanal a síntesis de campo de ondas

- 10 Otra aplicación para el esquema descrito en la Sección VI se describe en lo siguiente. Puede extenderse el audio estéreo y multicanal (por ejemplo 5.1 circundante) que acompaña a imágenes en movimiento para obtener una síntesis de campo de ondas agregando la información secundaria. Por ejemplo, puede extenderse el Dolby ACC-3 (audio para DVD) para el audio de codificación de compatibilidad hacia atrás de 5.1 para los sistemas de síntesis de campo de ondas, es decir, DVDs se reproducen en sonido circundante 5.1 en reproductores heredados convencionales y el sonido de síntesis de campo de ondas en una nueva generación de reproductores que soportan el procesamiento de la información secundaria.

15 VIII. EVALUACIONES SUBJETIVAS

- 20 Se ha realizado un descodificador en tiempo real de los algoritmos propuestos en las Secciones IV-A y IV-B. Se utiliza un banco de filtros STFT basado en FFT. Se utilizan una FFT de 1024 puntos y un tamaño de ventana STFT de 768 (con relleno con ceros). Se agrupan los coeficientes espectrales de modo que cada grupo represente la señal con un ancho de banda de dos veces el ancho de banda rectangular equivalente (ERB). El escucha informal mostró que la calidad de audio no mejora notablemente cuando escóglase escoger la resolución de frecuencia más alta. Una resolución de frecuencia más baja es favorable puesto que resulta en menos parámetros a transmitir.

Para cada fuente, pueden ajustarse individualmente la panorámica y ganancia de amplitud/retardo. Se utilizó el algoritmo para codificar varias grabaciones de audio de múltiples pistas con 12 a 14 pistas.

- 25 El descodificador permite la mezcla circundante 5.1 utilizando un mezclador de panorámica de amplitud de base (VBAP). Pueden ajustarse la dirección y ganancia de cada señal fuente. El software permite la interrupción al vuelo entre la mezcla de la señal fuente codificada y la mezcla de las señales fuente discretas originales.

- 30 Un escucha sencillo normalmente muestra ninguna o poca diferencia entre la mezcla de las señales fuente originales o codificadas si para cada fuente se utiliza una ganancia G_i de cero dB. Cuanto más se varían las ganancias fuente, más artefactos ocurren. La amplificación y atenuación leves de las fuentes (por ejemplo hasta ± 6 dB) darán aún buenos resultados. Un escenario crítico es cuando todas la fuentes se mezclan en un lado y solamente una única fuente en el otro lado opuesto. En este caso, puede reducirse la calidad de audio, dependiendo de la mezcla específica y las señales fuente

IX. CONCLUSIONES

- 35 Se ha propuesto un esquema de codificación para la codificación conjunta de las señales fuente de audio, por ejemplo, los canales de una grabación de múltiples pistas. El objetivo no es codificar las formas de onda de la señal fuente con alta calidad, en cuyo caso la codificación conjunta resultaría en una ganancia de codificación mínima, puesto que las fuentes de audio normalmente son independientes. El objetivo es que cuando se mezclan las señales fuente codificadas se obtenga una señal de audio de alta calidad. Considerando las propiedades estadísticas de las señales fuente, las propiedades de los esquemas de mezcla, y el sonido espacial, se ha mostrado que se logra una mejora en la ganancia de codificación significativa mediante la codificación conjunta de las señales fuente.

La mejora en la ganancia de codificación se debe al hecho de que solamente se transmite una forma de onda de audio.

Adicionalmente, se transmite la información secundaria, que representa las propiedades estadísticas de las señales fuente que son factores relevantes que determinan la percepción espacial de la señal mezclada final.

- 45 La velocidad de información secundaria es de alrededor de 3 kbs por señal fuente. Puede aplicarse cualquier mezclador con las señales fuente codificadas, por ejemplo, mezcladores estéreo, multicanal o de síntesis de campo de onda.

- 50 Una manera directa de escalar el esquema propuesto para la velocidad de bits y la calidad más altas es mediante la transmisión de más de un canal de audio. Además, se ha propuesto una variación del esquema, lo que permite la remezcla de la señal de audio estéreo o multicanal dada (y aún cambiando el formato de audio, por ejemplo, de estéreo a multicanal o a síntesis de campo de ondas).

Las aplicaciones del esquema propuesto son múltiples. Por ejemplo, podría extenderse MPEG-4 con el esquema propuesto para reducir la velocidad de bits cuando se necesita transmitir más de un "objeto de audio natural"

(señal fuente). Igualmente, el esquema propuesto ofrece la representación compacta del contenido de los sistemas de síntesis de campo de ondas. Tal como se mencionó, las señales etéreo o multicanal existentes pueden ser complementadas con la información secundaria para permitir que el usuario remezcle las señales a su gusto.

REFERENCIAS

- 5 [1] C. Faller, **Parametric Coding of Spatial Audio**, Ph.D. thesis, Swiss Federal Institute of Technology Lausanne (EPFL), 2004, Ph.D. Thesis No. 3062.
- [2] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," *IEEE Trans. on Speech and Audio Proc.*, vol. 11, no. 6, Nov. 2003.

REIVINDICACIONES

1. Método para sintetizar una pluralidad de canales de audio $(\hat{x}_1(n), \hat{x}_2(n))$, que comprende las etapas de:
 - recuperar de un flujo de audio al menos una señal suma $(s(n))$ que representa una suma de señales fuente $(s_1(n), s_2(n), \dots, s_M(n))$,
 - recuperar del flujo de audio la información estadística $(\Phi)(n, d)$, $(\Phi)(n, e)$, $E\{\tilde{s}_i(n)\}$ que representa el envolvente espectral de una o más señales fuente $(s_1(n), s_2(n), \dots, s_M(n))$,
 - recibir del flujo de audio, o determinar de manera local, parámetros (S_n) que describen un formato de audio de salida y unos parámetros de mezcla (a_i, b_i) de las señales fuente, y
 - sintetizar la pluralidad de canales de audio $(\hat{x}_1(n), \hat{x}_2(n))$ a partir de la al menos una señal suma basándose en unos parámetros de mezclador de salida (g_1, g_2, D_1, D_2) calculados,en el que se calculan los parámetros de mezclador de salida (g_1, g_2, D_1, D_2) a partir de la información estadística recibida, los parámetros que describen el formato de audio de salida, y dichos parámetros de mezcla (a_i, b_i) de las señales fuente, de manera que la pluralidad de canales de audio sintetizados tienen indicaciones (ICLD, ICTD, ICC) similares a las indicaciones (ICLD, ICTD, ICC) de los canales de audio obtenibles mediante la mezcla de las señales fuente utilizando los parámetros (S_n) que describen el formato de audio de salida y dichos parámetros de mezcla de las señales fuente.
2. Método de la reivindicación 1, en el que los envolventes espectrales estadísticas de la una o más señales fuente de audio comprenden parámetros de filtro de red o parámetros espectrales de línea.
3. Método de la reivindicación 1, en el que la información estadística representa una potencia relativa como una función de la frecuencia y del tiempo de la una o más señales fuente.
4. Método de la reivindicación 1, en el que se sintetiza la pluralidad de canales de audio en un dominio de subbanda de un banco de filtros.
5. Método de la reivindicación 4, en el que se determinan el número y anchos de banda del dominio de subbanda según una resolución espectral y temporal de un sistema auditivo humano.
6. Método de la reivindicación 4, en el que el número de subbandas está comprendido entre 3 y 40.
7. Método de la reivindicación 4, en el que las subbandas tienen diferentes anchos de bandas, en donde las subbandas de frecuencias más bajas tienen anchos de banda menores que las subbandas de frecuencias más altas.
8. Método de la reivindicación 4, en el que se utiliza un banco de filtros basado en una transformada de Fourier de tiempo corto (STFT) y se combinan los coeficientes espectrales de modo que cada grupo de coeficientes espectrales forman una subbanda.
9. Método de la reivindicación 1, en el que la información estadística también comprende funciones de autocorrelación.
10. Método de la reivindicación 2, en el que se representan los envolventes espectrales como parámetros de codificación predictiva lineal (LPC).
11. Método de la reivindicación 1, en el que la etapa de calcular los parámetros de mezclador de salida comprende el cálculo de las indicaciones de la pluralidad de canales de audio y el cálculo de los parámetros de mezclador de salida utilizando las indicaciones calculadas de la pluralidad de canales de audio.
12. Método de la reivindicación 1, en el que la etapa de calcular los parámetros de mezclador de salida comprende el cálculo de un primer factor de ganancia (g_1) y un segundo factor de ganancia (g_2), un primer retardo (D_1) y un segundo retardo (D_2), y un valor de determinación de procesamiento de descorrelación $c(n)$.

13. Aparato para sintetizar una pluralidad de canales de audio ($\hat{x}_1(n), \hat{x}_2(n)$), en el que el aparato comprende medios operativos para:

- recuperar de un flujo de audio al menos una señal suma ($s(n)$) que representa una suma de señales fuente ($s_1(n), s_2(n), \dots, s_M(n)$),

5

- recuperar del flujo de audio la información estadística ($\Phi(n, d), \Phi(n, e), \Phi\{\tilde{s}(n)\}$) que representa el envolvente espectral de una o más señales fuente ($s_1(n), s_2(n), \dots, s_M(n)$),

- recibir del flujo de audio, o determinar de manera local, parámetros (S_n) que describen un formato de audio de salida y unos parámetros de mezcla (a_i, b_i) de las señales fuente, y

10

- sintetizar la pluralidad de canales de audio ($\hat{x}_1(n), \hat{x}_2(n)$) a partir de la al menos una señal suma basándose en unos parámetros de mezclador de salida (g_1, g_2, D_1, D_2) calculados,

en el que el aparato es adicionalmente operativo para calcular dichos parámetros de mezclador de salida (g_1, g_2, D_1, D_2) a partir de la información estadística recibida, los parámetros que describen el formato de audio de salida, y dichos parámetros de mezcla (a_i, b_i) de las señales fuente, de manera que la pluralidad de canales de audio sintetizados tienen indicaciones (ICLD, ICTD, ICC) similares a las indicaciones (ICLD, ICTD, ICC) de los canales de audio obtenibles mediante la mezcla de las señales fuente utilizando los parámetros (S_n) que describen el formato de audio de salida y los parámetros de mezcla de las señales fuente.

15

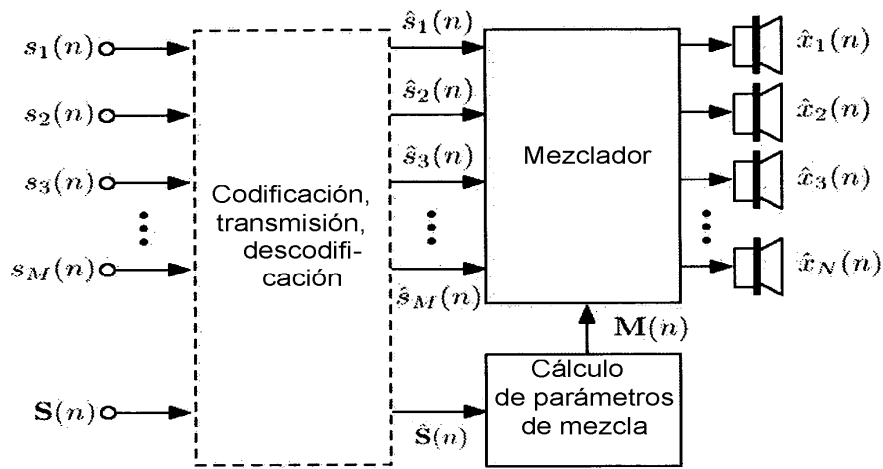


Fig. 1

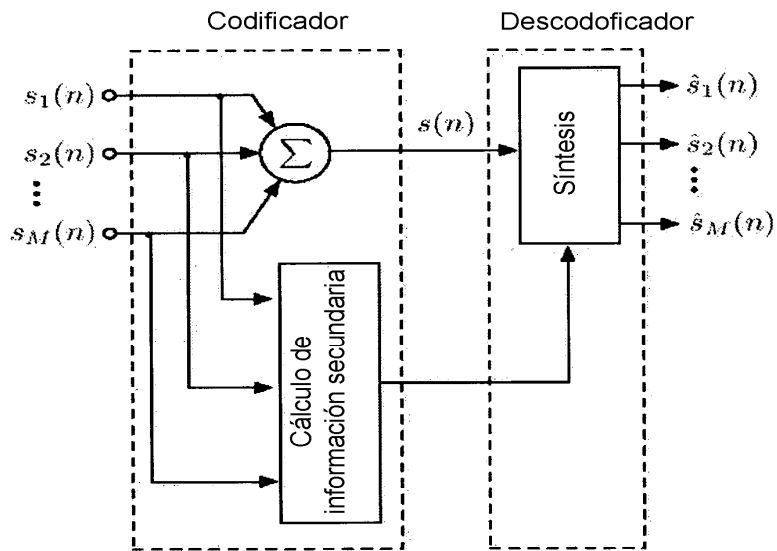


Fig. 2

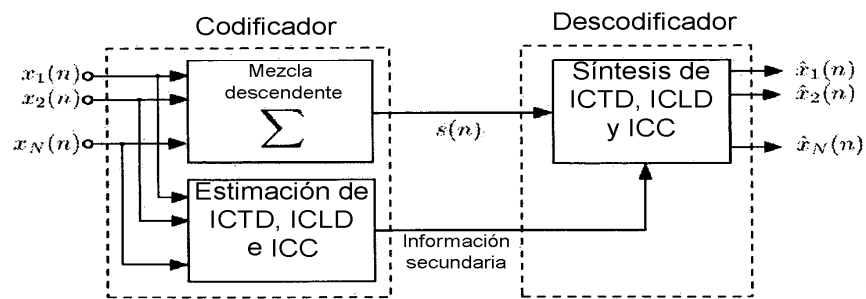


Fig. 3

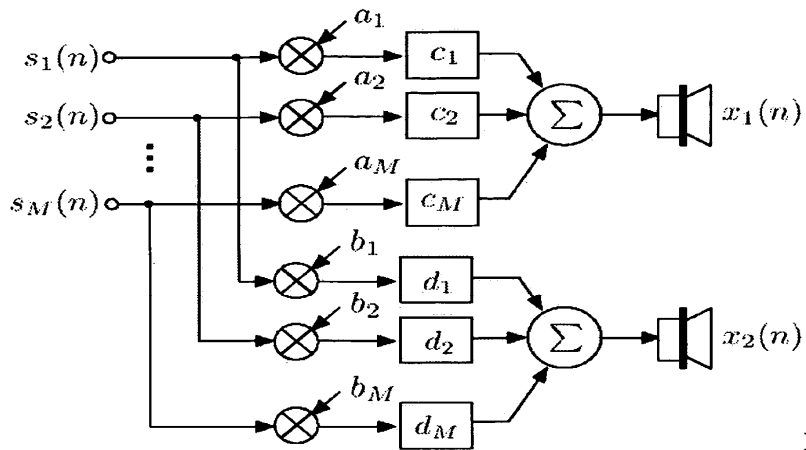


Fig. 4

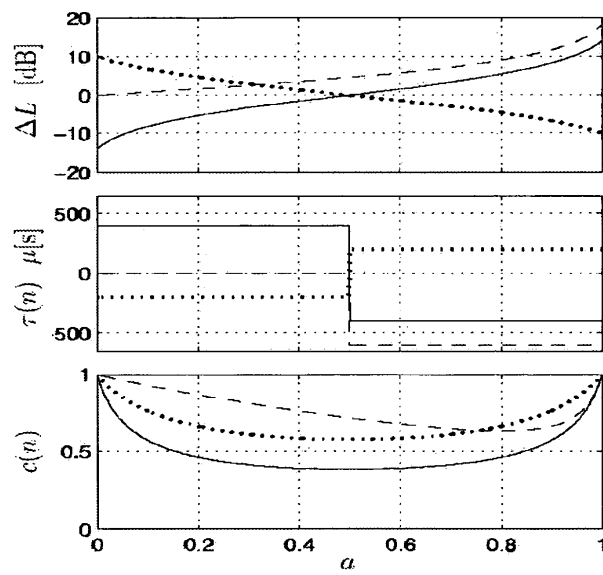


Fig. 5

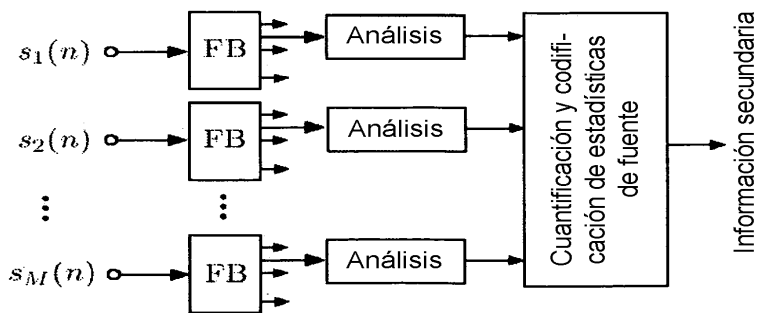


Fig. 6

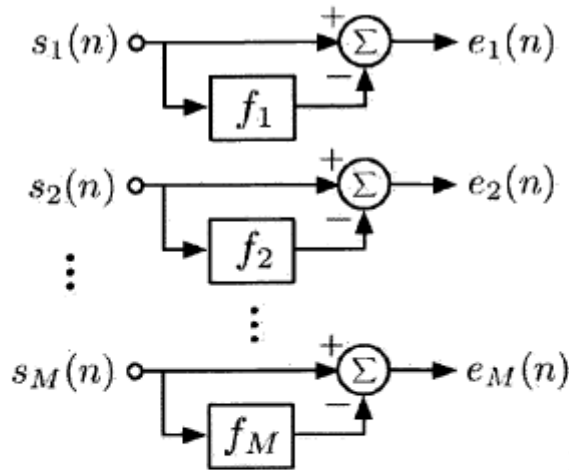


Fig. 7

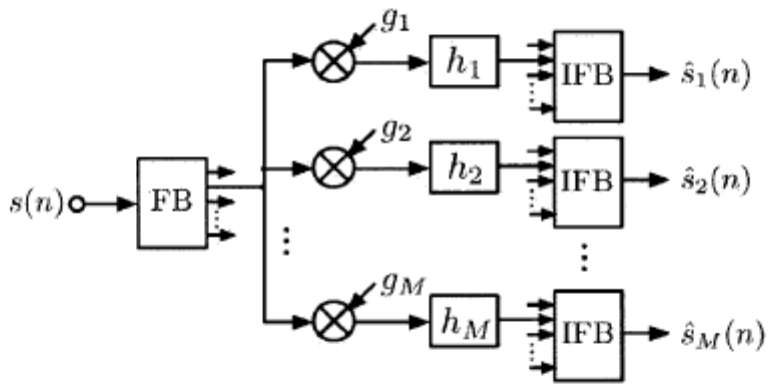


Fig. 8

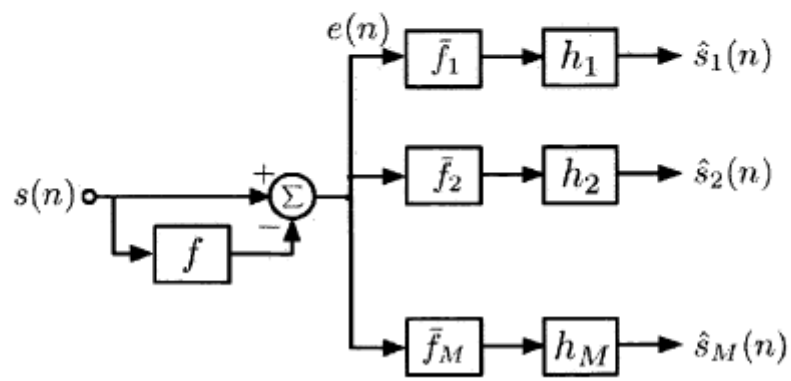


Fig. 9

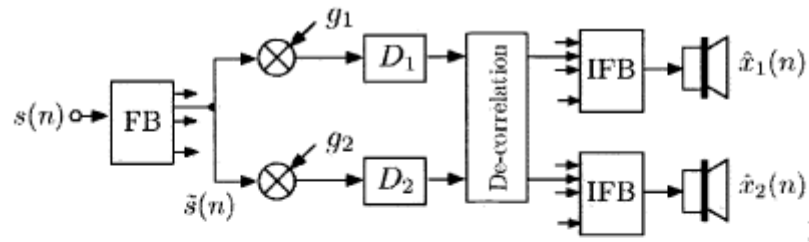


Fig. 10

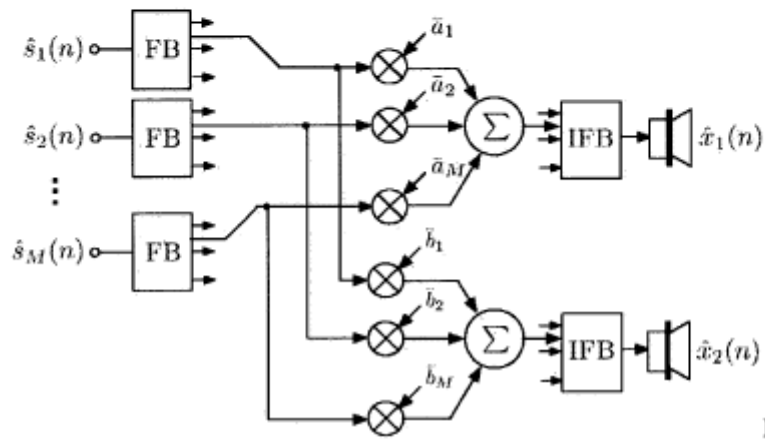


Fig. 11

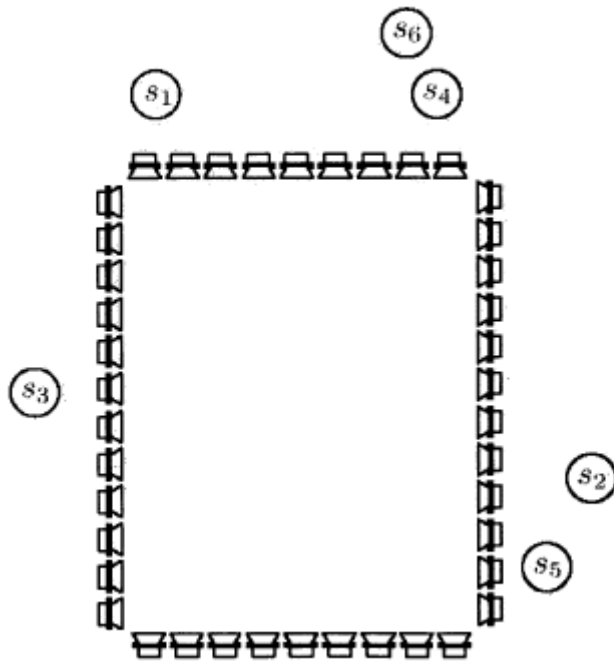


Fig. 12

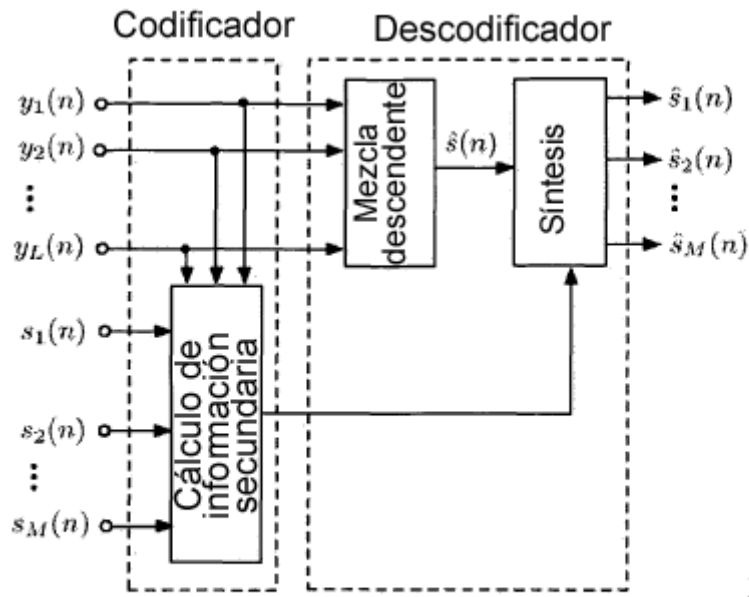


Fig. 13

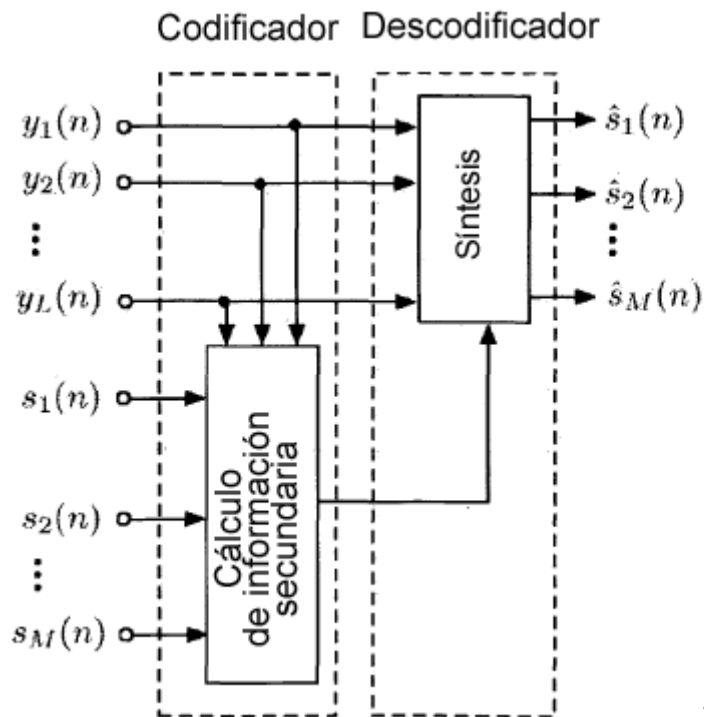


Fig. 14