(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 682 073**

(51) Int. CI.:

G10L 19/008 (2013.01)
H04S 3/00 (2006.01)
H04S 7/00 (2006.01)
G10L 19/02 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **13.02.2006 E 10179108 (5)**

(97) Fecha y número de publicación de la concesión europea: **02.05.2018 EP 2320414**

(54) Título: **Codificación conjunta paramétrica de fuentes de audio**

(30) Prioridad:

14.02.2005 EP 05101055

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
18.09.2018

(73) Titular/es:

FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (100.0%)
Hansastraße 27C
80686 München, DE

(72) Inventor/es:

FALLER, CHRISTOF

(74) Agente/Representante:

ARIZTI ACHA, Monica

Codificación conjunta paramétrica de fuentes de audio

DESCRIPCIÓN

1. Introducción

En un problema de codificación general, tenemos un número de(mono) señales de fuente $s_i(n)$ ($1 \leq i \leq M$ y un vector de descripción de escena $\mathbf{S}(n)$, donde n es el índice de tiempo. El vector de descripción de escena contiene parámetros, tal como posiciones de fuente (virtuales), anchuras de fuente y parámetros acústicos, tales como los parámetros de la sala (virtuales). La descripción de escena puede ser invariante del tiempo o puede cambiar con el tiempo. Las señales de fuente y la descripción de escena se codifican y transmiten a un descodificador. Las señales de fuente codificadas, $\hat{s}_i(n)$ se mezclan sucesivamente como una función de la descripción de escena, $\hat{S}(n)$, para generar síntesis de campo de onda, múltiples canales o señales estéreo, como una función del vector de descripción de escena. Las señales de salida del descodificador se indican como $\hat{x}_i(n)$ ($0 \leq i \leq N$). Obsérvese que el vector de descripción de escena S(n) puede no transmitirse, pero puede determinarse en el descodificador. En este documento, la expresión “señal de audio de estéreo” siempre se refiere a señales de audio de estéreo de dos canales.

El ISOMEC MPEG-4 trata el escenario de codificación descrito. Define la descripción de escena y usa para cada señal de 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 tasa de bits llega a ser alta, es decir, la tasa de bits escala de manera ascendente con el número de fuentes. La codificación de una señal de fuente con alta calidad requiere aproximadamente de 60 a 90 kb/s.

Previamente, tratamos un caso especial del problema de codificación descrito [1][2] con un esquema indicado Codificación de Indicador Binaural (BCC) para Representación Flexible. Transmitiendo solamente la suma de las señales de fuente dadas, más la información secundaria de baja tasa de bits, se consigue esta baja tasa de bits. Sin embargo, las señales de fuente no pueden recuperarse en el descodificador y el esquema se limita a la generación de señales envolventes estéreo y de múltiples canales. También, únicamente se usó una mezcla simplista, basándose en la panorámica de amplitud y retardo. Así, la dirección de fuentes puede controlarse, 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 a medida que aumenta el número de señales de fuente.

El documento [1] (Codificación de Indicador Binaural, Estéreo Paramétrico, MP3 Envolvente, MPEG Envolvente) cubre el caso donde se codifican N canales de audio y N canales de audio con **indicadores similares**, a continuación se descodifican los canales de audio originales. La información secundaria transmitida incluye parámetros de indicador de inter-canal relativos a las diferencias entre los canales de entrada.

Los canales de las señales de audio estéreo y de múltiples canales contienen mezclas de señales de fuente de audio y por lo tanto son diferentes en naturaleza que las señales de fuente de audio puras. Las señales de audio estéreo y de múltiples canales se mezclan de modo que cuando se reproducen en un sistema de reproducción apropiad, el oyente percibirá una imagen espacial auditiva (“etapa de sonido”) según se captura por el ajuste de grabación o se diseña por el ingeniero de grabación durante la mezcla. Se ha propuesto previamente un número de esquemas para la codificación de conjunta para los canales de una señal de audio de estéreo o de múltiples canales.

El documento US 2004/0049379 A1 desvela una tecnología de codificación y decodificación de audio de múltiples canales. Un codificador de audio realiza una transformación de múltiples canales de pre-procesamiento en datos de audio de múltiples canales, variando la transformación para controlar la calidad. El codificador agrupa múltiples ventanas de diferentes canales en una o más piezas y emite la información de configuración de pieza, que permite que el codificador aísle transitorios.

El documento US 2004/0101048 A1 desvela un procesamiento de señal de datos de múltiples canales. Los datos de múltiples canales se recopilan y representan cuaterniones. Estos datos se emiten a continuación a un predictor lineal. Se calcula una matriz de autocorrelación y, posteriormente, se generan pseudo-inversas y se emiten a coeficientes de predicción lineal y residual.

La Tesis N.º 3062 (2004), “Parametric coding of spatial audio”, Christof Faller, Lausanne, EPFL, documento XP-002343263 desvela varias tecnologías de codificación paramétrica tal como codificación de indicador binarual. La diferencia de tiempo inter-canal de indicadores espaciales, la diferencia de nivel inter-canal y la correlación inter-canal se estiman para señales estéreo y señales de audio de múltiples canales. Esto se realiza en una manera a nivel de subbanda. Un decodificador de BCC genera una señal de audio dada la señal suma transmitida más los indicadores espaciales.

La publicación “Estimation of auditory spatial cues for binaural cue coding”, Frank Baumgarte y Christof Faller, páginas 1801-1804, IEEE International Conference On Acoustics, Speech, And Signal Processing (ICASSP), Nueva York, 13 de mayo de 2002, documento XP010804245 resume la codificación de indicador binaural. La extracción de los indicadores espaciales de la señal estereofónica se realiza con un analizador de BCC. El analizador de BCC comprende un bloque de estimación de coherencia, bloques de estimación de potencia, bloques de compensación de retardo y un bloque de detección máxima. Las señales de entrada son una señal de audio desde un primer canal A y una señal de audio desde un segundo canal B y los canales de audio se someten a un banco de filtros coclear (CFB) y un modelo de célula ciliada interna (IHC).

Es un objeto de la invención proporcionar un concepto mejorado para codificación. Esto se consigue mediante el método de la reivindicación 1 o el aparato de la reivindicación 2. El objeto de la invención es proporcionar un método para transmitir una pluralidad de señales de fuente, mientras se usa un ancho de banda mínimo. En la mayoría de los métodos conocidos, el formato de reproducción (por ejemplo estéreo, 5.1) está 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, uniendo al usuario a un escenario de reproducción predefinido (por ejemplo, estéreo).

La invención propuesta codifica N señales de fuente de audio, típicamente sin canales de señales de estéreo o de múltiples canales, pero señales independientes, tal como diferentes señales de voces o instrumentos. La información de lado transmitido incluye parámetros estadísticos relacionados con las señales de fuente de audio de entrada.

La presente invención descodifica M canales de audio con **diferentes indicadores** que las señales de fuente de audio originales. Estos diferentes indicadores se sintetizan implícitamente aplicando un mezclador a la señal de suma recibida. Este mezclador se controla como una función de la información de fuente estadística recibida y los parámetros de formato de audio recibidos (o localmente determinados), y los parámetros de mezcla. Como alternativa, estos diferentes indicadores se calculan explícitamente como una función de la información de fuente estadística recibida, y los parámetros de formato de audio (o determinados localmente) recibidos y los parámetros de mezcla. Estos indicadores calculados se usan para controlar un descodificador de la técnica anterior (Codificación de Indicador Binaural, Estéreo Paramétrico, MPEG Envolvente) para sintetizar los canales de salida dada la señal de suma recibida.

El esquema propuesto para la codificación conjunta de las señales de fuente de audio es el primero de su clase, se diseñó para la codificación conjunta de las señales de fuente de audio. Las señales de fuente de audio son normalmente señales de audio mono que no son adecuadas para la reproducción sobre un sistema de audio de estéreo o de múltiples canales. Por brevedad, a continuación, las señales de fuente de audio se indican a menudo como señales de fuente.

Las señales de fuente de audio en primer lugar necesitan mezclarse a señales de audio estéreo, de múltiples canales o de síntesis de campo de onda, antes de la reproducción. Una señal de fuente de audio puede ser un único instrumento o hablante, o la suma de un número de instrumentos y hablantes. Otro tipo de señal de fuente de audio es una señal de audio mono capturada con un micrófono puntual durante un concierto. A menudo las señales de 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 las señales de fuente de audio, está basado únicamente en transmitir la suma de las señales de fuente de audio,

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

o una suma ponderada de las señales de fuente. Opcionalmente, la suma ponderada puede llevarse a cabo con diferentes pesos en diferentes subbandas y los pesos pueden adaptarse en el tiempo. La suma con ecualización, como se describe en el Capítulo 3.3.2 en [1], puede aplicarse también. A continuación, cuando nos referimos a la suma o señal de suma, siempre se entiende una señal generada por (1) o generada como se describe. Además de la señal de suma, se transmite la información secundaria. La suma y la información secundaria representan el flujo de audio producido. Opcionalmente, la señal de suma se codifica usando un codificador convencional de audio mono. Este flujo puede almacenarse en un archivo (CD, DVD, Disco Duro) o difundirse al receptor. La información secundaria representa las propiedades estadísticas de las señales de fuente que son los factores más importantes que determinan los indicadores espaciales percibidos de las señales de salida del mezclador. Se mostrará que estas propiedades son envolventes espectrales desarrollados temporalmente y funciones de auto-correlación. Se transmite aproximadamente 3 kb/s de información secundaria por la señal de fuente. En el receptor, las señales de fuente $\hat{s}_i(n)$ ($1 \leq u \leq M$) se recuperan con las propiedades estadísticas anteriormente mencionadas, que se aproximan a las propiedades correspondientes de las señales de fuente originales y la señal de suma.

Breve descripción de los dibujos

La invención se entenderá mejor gracias a las Figuras adjuntas en las cuales:

- la figura 1 muestra un esquema en el cual la transmisión de cada señal de fuente se hace independientemente para el proceso adicional;
- la figura 2 muestra un número de fuentes transmitidas como la señal de suma más la información secundaria;
- la figura 3 muestra un diagrama de bloques de un esquema de Codificación de Indicador Binaural (BCC);
- la figura 4 muestra un mezclador para generar señales estéreo, basándose en las varias señales de fuente,
- la figura 5 muestra la dependencia entre ICTD, ICLD e ICC y la potencia de subbanda de la señal de fuente;
- la Figura 6 muestra el proceso de la generación de información secundaria;
- la figura 7 muestra el proceso de estimación de los parámetros de LPC de cada señal de fuente;
- la figura 8 muestra el proceso de recrear las señales de fuente desde una señal de suma;
- la figura 9 muestra un esquema alternativo para la generación de cada señal desde la señal de suma;
- la figura 10 muestra un mezclador para generar señales estéreo basándose en la señal de suma;
- la figura 11 muestra un algoritmo panorámico de la amplitud que evita que los niveles de 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 de fuente en el receptor procesando la mezcla descendente de los canales transmitidos; y
- la figura 14 muestra cómo recuperar una estimación de las señales de fuente en el receptor procesando 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 de fuente;
d	índice de retardo;
M	número de señales de fuente de entrada del codificador;
N	número de canales de salida del descodificador;
$x_i(n)$	señales de fuente originales mezcladas;
$\hat{x}_i(n)$	señales de salida del descodificador mezcladas;
$s_i(n)$	señales de fuente de entrada del codificador;
$\hat{s}_i(n)$	señales de fuente transmitidas, también llamadas señales de pseudo-fuente;
$s(n)$	señal de suma transmitida;
$y_i(n)$	señal de audio del canal L; (señal de audio que se va a re-mezclar);
$s_i(k)$	una señal de subbanda de $s_i(n)$ (definida de manera similar para otras señales);
$E(s_i^2(n)]$	estimación de tiempo corto de $s_i^2(n)$ (definida de manera similar para otras señales);
ICLD	diferencia de nivel inter-canal;
ICTD	diferencia de tiempo inter-canal;
ICC	coherencia inter-canal;
$\Delta L(n)$	ICLD de subbanda estimada;
$\tau(n)$	ICTD de subbanda estimado;
$c(n)$	ICC de subbanda estimado;
$p_i(n)$	potencia de subbanda de fuente relativa;
a_i, b_i	factores de escala del mezclador;
c_i, d_i	retardos del mezclador;

AL_j, $\tau(n)$ nivel del mezclador y diferencia de tiempo;
G_i ganancia de fuente del mezclador;

III. Codificación conjunta de señales de fuente de audio

En primer lugar, se describe la Codificación de Indicador Binaural (BCC), una técnica de codificación de audio de múltiples canales paramétrica. A continuación se muestra que con la misma idea en la que está basada la BCC se puede concebir un algoritmo para la codificación conjunta de las señales de fuente para un escenario de codificación.

A. Codificación de Indicador Binaural (BCC)

Un esquema de BCC [1][2] para una codificación de audio de múltiples canales se muestra en la siguiente figura. La señal de audio de múltiples canales de entrada se mezcla de manera descendente a un solo canal. En oposición a la información de codificación y transmisión de aproximadamente todas las formas de onda del canal, solamente se codifica (con el codificador de audio mono convencional) y transmite la señal mezclada de manera descendente. Adicionalmente, las “diferencias de canal de audio” motivadas percibidas se estiman entre los canales de audio originales y también se transmiten al descodificador. El descodificador genera sus canales de salida, de modo que las diferencias del 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 del canal de audio perceptualmente relevantes para una pareja de canales de la señal de altavoces son la diferencia de tiempo inter-canal (ICTD) y la diferencia de nivel inter-canal (ICLD), ICTD e ICLD pueden estar relacionadas con la dirección percibida de los eventos auditivos. Otros atributos de imagen espacial auditiva, tal como una anchura de fuente aparente y la envolvente del oyente, pueden estar relacionados con la *coherencia interaural* (IC). Para parejas de altavoces en la parte delantera o trasera de un oyente, la *coherencia interaural* está a menudo relacionada directamente con la coherencia inter-canal (ICC), que es así considerada como la tercera medida de la diferencia de canal de audio por BCC. ICTD, ICLD e ICC se estiman en las subbandas como una función del tiempo. Tanto la resolución espectral como la temporal que se usan, se ven motivadas por la percepción.

B. Codificación conjunta paramétrica de fuentes de audio

Un descodificador de BCC es capaz de generar una señal de audio de múltiples canales, con cualquier imagen espacial auditiva, tomando una mono-señal y sintetizando a intervalos de tiempo regulares un solo indicador de ICTD, ICLD e ICC específico por subbanda y pareja de canales. El buen rendimiento de los esquemas de BCC para una amplia gama de material de audio [véase 1] implica que la imagen espacial auditiva percibida se ve enormemente determinada por ICTD, ICLD e ICC. Por lo tanto, en oposición a las señales de fuente “limpias” requeridas, $\hat{s}_i(n)$ como entrada del mezclador en la Figura 1, requerimos justamente señales de pseudo-fuente $\hat{s}_i(n)$, con la propiedad que ellas den como resultado ICTD, ICLD e ICC similares, en la salida del mezclador, como para el caso de suministro de las señales de fuente reales al mezclador. Hay tres objetivos para la generación de $\hat{s}_i(n)$:

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

Para derivar el esquema propuesto, consideramos un mezclador de estéreo (M = 2). Una simplificación adicional sobre el caso general es que sólo la panorámica de amplitud y retardo se aplican para la mezcla. Si las señales de fuente discretas están disponibles en el descodificador, una señal de estéreo se mezclaría como se muestra en la Figura 4, es decir,

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

En este caso, el vector de descripción de escena $\mathbf{S}(n)$ contiene solamente direcciones de 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. Obsérvese que para los parámetros de mezcla, ignoramos el índice de tiempo por conveniencia de notación.

Los parámetros más convenientes para controlar el mezclador son el tiempo y la diferencia de nivel, T_i y ΔL_i, que están relacionados con a_i, b_i, c_i y d_i por

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

donde G_i es un factor de ganancia de fuente en dB.

A continuación, calculamos ICTD, ICLD e ICC de la salida del mezclador de estéreo como una función de las señales de fuente de entrada $s_i(n)$. Las expresiones obtenidas proporcionarán indicación de qué propiedades de señales de fuente determinan, ICTD, ICLD e ICC (junto con los parámetros de mezcla). $\hat{s}_i(n)$ se generan a continuación de modo que las propiedades de la señal de fuente identificadas se aproximen a las propiedades correspondientes de las señales de fuente originales.

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

Los indicadores se estiman en subbandas y como una función del tiempo. A continuación, se supone que las señales de fuente $s_i(n)$ son de media cero y mutuamente independientes. Una pareja de señales de subbanda de la salida (2) del mezclador se indica $\hat{x}_1(n)$ y $\hat{x}_2(n)$. Obsérvese que por simplicidad de notación usamos el mismo índice de tiempo n, para las señales del dominio del tiempo y del dominio de subbanda. Igualmente no se usa un índice de subbanda y el análisis/procesamiento descrito se aplica a cada subbanda independientemente. La potencia de la subbanda de las dos señales de salida del mezclador es:

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

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

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

donde K determina la longitud de la media móvil. Obsérvese que los valores de potencia de subbanda $E\{\tilde{s}_2^2(n)\}$ representan para cada señal de fuente, la envolvente espectral como una función de tiempo. El ICLD $\Delta L(n)$, es

$$\Delta L(n) = 10 \log_{10} \frac{\sum_{i=1}^{M} b_1^2 E\{\tilde{s}_i^2(n)\}}{\sum_{i=1}^{M} a_i^2 E\{\tilde{s}_i^2(n)\}} \quad (7)$$

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

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

La ICC, $c(n)$ se calcula de acuerdo con

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

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

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

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

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

que es equivalente a

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

donde la función de auto-correlación normalizada $\Phi(n,e)$ es

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

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

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

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

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

Un ejemplo numérico para dos señales de fuente, que ilustra la dependencia entre ICTD, ICLD e ICC y la potencia de subbanda de fuente, se muestra en la Figura 5. La parte superior, media e inferior de la Figura 5 muestran $\Delta L(n)$, $T(n)$ y $c(n)$, respectivamente, como una función de la relación de la potencia de subbanda de las dos señales de fuente $E\{\tilde{s}_1^2(n)\}/\{E\{\tilde{s}_1^2(n)\} + E\{\tilde{s}_2^2(n)\}\}$, para diferentes parámetros de mezcla (4) ΔL_1, ΔL_2, T_1 y T_2. Obsérvese que cuando sólo una fuente tiene potencia en la subbanda ($a = 0$ o $a = 1$) entonces $\Delta L(n)$ y $T(n)$ calculados son iguales a los parámetros de mezcla (ΔL_1, ΔL_2, T_1 y 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 de la potencia de subbanda de tiempo corto de las fuentes, $E\{s_i^2(n)\}$ (6). La función de correlación cruzada de subbanda normalizada $\Phi(n,d)$ (12), que es necesaria para el cálculo de la ICTD (10) e ICC (9) depende de $E\{\tilde{s}_i^2(n)\}$ y adicionalmente de la función de auto-correlación de subbanda normalizada $\Phi_i(n,e)$ (13) para cada señal de fuente. El máximo de $\Phi(n,d)$ radica dentro del intervalo $\min_i\{T_i\} \le d \le \max_i\{T_i\}$. Para fuente i con parámetro de mezclador $T_i = d_i - c_i$, el intervalo correspondiente para el cual la propiedad de la subbanda de señal de fuente $\Phi_i(n,e)$ es necesaria, es

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

Puesto que los indicadores de ICTD, ICLD e ICC dependen de las propiedades de subbanda de la señal de fuente $E\{\tilde{s}_i^2(n)\}$ y $\Phi_i(n, e)$, en el intervalo (14), en principio estas propiedades de la subbanda de la señal de fuente necesitan transmitirse como información secundaria. Suponemos que cualquier otra clase de mezclador (por ejemplo mezclador con efectos, mezclador de síntesis de campo de onda/convolucionador, etc.) tiene propiedades similares y así esta información secundaria es útil también cuando se usan otros mezcladores distintos del descrito.

Para reducir la cantidad de la información secundaria, se puede almacenar un conjunto de funciones de auto-correlación predefinidas en el descodificador y solamente transmitir índices de selección de aquellas que corresponden más estrechamente con las propiedades de la señal de fuente. Una primera versión de nuestro algoritmo supone que dentro del intervalo (14) $\Phi_i(n, e) = 1$ y así se calcula (12) usando sólo los valores (6) de potencia de subbanda como la información secundaria. Los datos mostrados en la Figura 5 se han calculado suponiendo $\Phi_i(n, e) = 1$.

Con el fin de reducir la cantidad de la información secundaria, se limita el intervalo dinámico relativo de las señales de fuente. En cada momento, para cada subbanda, se selecciona la potencia de la fuente más intensa. Encontramos que es suficiente disminuir el límite de la potencia de la subbanda correspondiente de todas las otras fuentes a un valor de 24 dB menor que la potencia de la subbanda más intensa. Así, el intervalo dinámico del cuantificador puede limitarse a 24 dB.

Suponiendo que las señales de 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 se puede calcular localmente. Dada esta idea, la tasa de información secundaria puede reducirse ligeramente transmitiendo la potencia de subbanda de fuentes con índice $2 \leq i \leq M$ con relación a la potencia de la primera fuente,

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

Obsérvese que el intervalo dinámico que limita como se ha descrito previamente, 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 de suma, en oposición a la normalización con relación a una potencia (15) de subbanda de fuente. Para una frecuencia de muestreo de 44,1 kHz, usamos 20 subbandas y se transmite para cada subbanda $\Delta\tilde{p}_i(n)$ $(2 \leq i \leq M)$ aproximadamente 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 subbandas de 20, por ejemplo 40 subbandas. El número de subbandas y los anchos de banda de estas subbandas se escogen de acuerdo con el tiempo y la resolución de frecuencia del sistema auditivo. Una implementación de baja calidad del esquema requiere al menos tres subbandas (frecuencias baja, media y alta).

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

Los valores de potencia relativos se cuantifican con un esquema similar al cuantificador de ICLD descrito en [2], dando como resultado una tasa 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 tasa de información secundaria puede reducirse adicionalmente analizando la actividad para cada señal de fuente y sólo transmitir la información secundaria asociada con la fuente si está activa.

En oposición a transmitir los valores de potencia de subbanda $E\{\tilde{s}_i^2(n)\}$ como información estadística, puede transmitirse otra información que representa las envolventes espectrales de las señales de fuente. Por ejemplo, pueden transmitirse parámetros de la codificación predictiva lineal (LPC), u otros parámetros correspondientes, tal como los parámetros del filtro de malla o parámetros del par espectral de línea (LP). El proceso de estimación de los parámetros de LPC de cada señal de fuente se ilustra en la Figura 7,

B.3 Calcular $\hat{s}_i(n)$.

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

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

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

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

Como filtros de descorrelación $h_i(n)$, pueden usarse filtros peine complementarios, filtros paso todo, filtros de retardo o filtros con respuestas de impulso aleatorio. El objetivo para el procesamiento de descorrelación es reducir la correlación entre las señales, mientras no modifican cómo se perciben las formas de onda individuales. Diferentes técnicas de descorrelación provocan diferentes artefactos. Los filtros peine complementarios provocan coloración. Todas las técnicas descritas dispersan la energía de transitorios en el tiempo que provocan artefactos tales como los "ecos previos". Dado su potencial para los artefactos, las técnicas de descorrelación deben aplicarse tan poco como sea posible. La siguiente sección describe técnicas y estrategias que requieren menos procesos de descorrelación que la simple generación de señales independientes $\hat{s}_i(n)$.

Un esquema alternativo para la generación de las señales $\hat{s}_i(n)$ se muestra en la Figura 9. En primer lugar se aplana el espectro de $s(n)$ por medio del cálculo del error de predicción lineal $e(n)$. A continuación, dados los filtros de LPC estimados en el codificador, f_i, se calculan filtros todo polo correspondientes como la transformación z inversa de

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

Los filtros resultantes todo polo resultantes, $\overline{f}_i$, representan la envolvente espectral de las señales de fuente. Si se transmite información secundaria distinta de los parámetros LPC, los parámetros de LPC en primer lugar necesitan calcularse como una función de la información secundaria. Como en el otro esquema, los filtros de descorrelación h_i se usan para obtener las señales de fuente independientes.

IV. Implementaciones que consideran restricciones prácticas

En la primera parte de esta sección, se proporciona un ejemplo de implementación, usando un esquema de síntesis de BCC, como un mezclador de estéreo o de múltiples canales. Esto es particularmente interesante, puesto que un esquema de síntesis de tipo BCC es parte de una norma de la ISO/IEC MPEG próxima, indicada "codificación de audio espacial". Las señales de fuente $\hat{s}i(n)$ no se calculan explícitamente en este caso, dando como resultado complejidad computacional reducida. Igualmente, este esquema ofrece el potencial de mejor calidad de audio, puesto que de manera efectiva es necesaria menos descorrelación que en el caso donde se calculan explícitamente las señales de fuente $\hat{s}_i(n)$.

La segunda parte de esta sección analiza los problemas cuando el esquema propuesto se aplica con cualquier mezclador y no se aplica en absoluto procesamiento de descorrelación. Un esquema de este tipo tiene una menor complejidad que un esquema con procesamiento de descorrelación, pero puede tener otros inconvenientes, como se analizará.

Idealmente, se podría 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 introducción de artefactos, se podría aplicar el procesamiento de descorrelación tan poco como sea posible. La tercera parte de esta sección analiza cómo la cantidad del procesamiento de descorrelación problemático puede reducirse, mientras se consiguen beneficios como si la $\hat{s}_i(n)$ generada fuera independiente.

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

La mezcla se aplica directamente a la señal de suma transmitida (1), sin el cálculo explícito de $\hat{s}_i(n)$. Se usa un esquema de síntesis de BCC para este fin. A continuación, consideramos el caso de estéreo, pero todos los principios descritos pueden aplicarse para la generación de señales de audio de múltiples canales igualmente.

Un esquema de síntesis de BCC de estéreo (o un esquema de "estéreo paramétrico), aplicado por el procesamiento de la señal de suma (1), se muestra en la Figura 10. Sería conveniente que el esquema de síntesis de BCC generara una señal que se perciba de manera similar como la señal de salida de un mezclador, como se muestra en la Figura 4. Esto es así, cuando ICTD, ICLD e ICC entre los canales de salida del esquema de síntesis de BCC son similares como los indicadores correspondientes que aparecen entre los canales de la señal de salida del mezclador

(4).

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

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

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

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

La ICC $c(n)$ se calculan de acuerdo con (9), determinando el procesamiento de descorrelación en la Figura 10. Este procesamiento de descorrelación (síntesis de ICC) se describe en [1]. Las ventajas de aplicar el procesamiento de descorrelación a los canales de salida del mezclador, en comparación con aplicarlo para la generación independiente de $\hat{s}_i(n)$ son:

- Normalmente, el número de señales M de fuente es mayor que el número de canales N de salida de audio. Así, el número de canales de audio independiente que se necesitan generar es menor cuando se descorrelacionan N canales de salida a diferencia de descorrelacionar M señales de fuente.
- A menudo, los N canales de salida de audio se correlacionan (ICC > 0) y pueden aplicarse menos procesamientos de descorrelación que serían necesarios para generar M o N canales independientes.

Debido a los menores procesamientos de descorrelación se espera mejor calidad de audio.

Se espera mejor calidad de audio cuando se restringen 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 de suma transmitida (1) es la misma que la potencia de la misma fuente en la señal de salida del descodificador mixto. La señal de salida del descodificador (Figura 10) es la misma que si la señal de salida del mezclador (Figura 4) se codificara y descodificara por un codificador/ descodificador de BCC en este caso. Así, se puede esperar una calidad igualmente similar.

El descodificador puede no sólo determinar la dirección en la que ha de aparecer cada fuente, sino también puede variarse la ganancia de cada fuente. La ganancia se aumenta seleccionando $a_i^2 + b_i^2 > 1$, (G_i > 0 dB) y se reduce seleccionando $a_i^2 + b_i^2 < 1$ (G_i < 0 dB).

B. Uso de procesamiento sin de descorrelación

La restricción de la técnica previamente descrita es que la mezcla se lleva a cabo con un esquema de síntesis de BCC. Se puede imaginar la implementación no sólo de la síntesis de ICTD, ICLD e ICC sino adicionalmente el procesamiento de efectos en la síntesis de BCC.

Sin embargo, puede ser deseable que puedan usarse los mezcladores y procesadores de efectos existentes. Esto también incluye los mezcladores de síntesis del campo de onda (a menudo indicados como “convolucionadores”). Para usar mezcladores y procesadores de efectos existentes, se calculan las $\hat{s}_i(n)$ explícitamente y se usan como si fueran las señales de fuente originales.

Cuando no se aplica el proceso de descorrelación ($h_i(n) = \delta(n)$ en (16) puede conseguirse también buena calidad de audio. Es un compromiso entre los artefactos introducidos debido al procesamiento de descorrelación y los artefactos debido al hecho de que las señales de fuente $\hat{s}_i(n)$ están correlacionadas. Cuando no se usa el procesamiento de descorrelación, la imagen espacial auditiva resultante puede sufrir inestabilidad [1]. Pero el mezclador puede introducir por sí mismo alguna descorrelación cando se usan reverberadores u otros efectos y de esta manera hay menos necesidad del procesamiento de descorrelación.

Si se generan $\hat{s}_i(n)$ sin el procesamiento de descorrelación, el nivel de las fuentes depende de la dirección en la que se mezclaran en relación con otras fuentes. Reemplazando los 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. Un algoritmo de amplitud de compensación de nivel se muestra en la Figura 11 que tiene como objetivo compensar la dependencia del nivel de fuente en los parámetros de mezcla. Dados los factores de ganancia de un algoritmo panorámico de amplitud convencional (por ejemplo, Figura 4), a_i y bi los pesos de la Figura 11, $\bar{a}_i$ y $\bar{b}_i$ se calculan por

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

Obsérvese que $\overline{\alpha_i}$ y $\bar{b}_i$ se calculan de modo que la potencia de subbanda de salida sea la misma que si $\hat{s}_i(n)$ fuera independiente en cada subbanda.

C. Reducir la cantidad de procesamiento de descorrelación

Como se mencionó previamente, la generación de $\hat{s}_i(n)$ independiente es problemática. Aquí, se describen estrategias para aplicar menor procesamiento de descorrelación, mientras se logra efectivamente un efecto similar como si $\hat{s}_i(n)$ fuera independiente.

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

1. Generar grupos de índices de fuentes que corresponden a las fuentes cercanas entre sí. Por ejemplo, en la Figura 8 estas pueden ser {1}, {2, 5}, {3}, y {4, 6}.

2. En cada momento en cada subbanda seleccionar el índice de fuente de la fuente más intensa,

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

Aplicar el procesamiento no de descorrelación para los índices de fuente en parte del grupo que contiene $i_{\max}$, es decir, $h_i(n) = \delta(n)$.

3. Para cada otro grupo, seleccionar la misma $h_i(n)$ dentro del grupo.

El algoritmo descrito modifica los mínimos componentes de señal más intensos. Adicionalmente, se reduce el número de diferentes $h_i(n)$ que se usa. Esto es una ventaja debido a que la descorrelación es más fácil que se necesiten generar menos canales independientes. La técnica descrita también es aplicable cuando se mezclan señales de audio estéreo o de múltiples canales.

V. Escalabilidad en términos de calidad y tasa de bits

El esquema propuesto transmite solamente la suma de todas las señales de fuente, que puede codificarse con un codificador de audio mono convencional. Cuando no es necesaria compatibilidad hacia atrás y la capacidad está disponible para la transmisión/almacenamiento de más de una forma de onda de audio, el esquema propuesto puede escalarse para el uso con más de un canal de transmisión. Esto se implementa generando varias señales de suma con diferentes subconjuntos de las señales de fuente dadas, es decir, a cada subconjunto de las señales de fuente se aplica individualmente el esquema de codificación propuesto. La calidad de audio se espera mejore conforme se aumenta el número de canales de audio transmitidos, debido a que han de generarse menos canales independientes por la descorrelación de cada canal transmitido (en comparación con el caso de un canal transmitido).

VI. Compatibilidad hacia atrás a formatos de audio estéreo y envolvente existentes

Consideremos el siguiente escenario de entrega de audio. Un consumidor obtiene una señal estéreo o envolvente de múltiples canales de máxima calidad (por ejemplo, por medio de un CD, DVD de audio o tienda de música en línea, 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 estéreo/envolvente convencional.

Esto se implementa entregando al consumidor (por ejemplo, como una opción de compra opcional en una tienda de

música en línea) un flujo de bits de información secundaria que permitir el cálculo de $\hat{s}_i(n)$ como una función de la señal de audio de estéreo o de múltiples canales dada. El algoritmo de mezcla del consumidor se aplica a continuación a $\hat{s}_i(n)$. A continuación, se describen dos posibilidades para calcular $\hat{s}_i(n)$, dadas las señales de audio de estéreo y de múltiples canales.

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

La manera más sencilla de usar el esquema de codificación propuesto con una transmisión de audio estéreo o de múltiples canales se ilustra en la Figura 13, donde $y_i(n)$ $(1 \leq i \leq L)$ son los canales L de la señal de audio estéreo o de múltiples canales. La señal de suma de las fuentes se estima por la mezcla descendente de los canales transmitidos a un canal de audio sencillo. La mezcla descendente se lleva a cabo por medio del cálculo de la suma de los canales $y_i(n)$ $(1 \leq u \leq L)$ o pueden aplicarse técnicas más sofisticadas.
Para el mejor rendimiento, se recomienda que el nivel de las señales de fuente se adapta antes de la estimación de $E\{\tilde{s}_i^2(n)\}$, (6) de modo que la relación de potencia entre las señales de fuente se aproxime a la relación de potencia con la que están contenidas las fuentes en la señal de estéreo o de múltiples canales dada. En este caso, la mezcla descendente de los canales transmitidos es una estimación relativamente buena de la suma de las fuentes (1) (o su versión escalada).

Puede usarse un proceso automatizado para ajustar el nivel de las entradas de la señal de fuente del codificador $s_i(n)$ antes del cálculo de la información secundaria. Este proceso adaptado en el tiempo estima el nivel en el que está contenida cada señal de fuente en la señal de estéreo o de múltiples canales dada. Antes del cálculo de la información secundaria, el nivel de cada señal de fuente se adapta a continuación en el tiempo ajustado de modo que sea igual al nivel en el que la fuente está contenida en la señal de audio estéreo o de múltiples canales.

B. Usar los canales transmitidos individualmente

La Figura 14 muestra una implementación diferente del esquema propuesto con la transmisión de señal envolvente estéreo o de múltiples canales. Aquí, los canales transmitidos no se mezclan de manera descenderte, pero se usan individuamente para la generación de $\hat{s}_i(n)$. Más generalmente, las señales de subbanda $\hat{s}_i(n)$ se calculan por

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

donde $w_l(n)$ son pesos que determinan las combinaciones lineales específicas de las subbandas de los canales transmitidos. Las combinaciones lineales se escogen de modo que $\hat{s}_i(n)$ ya esté tan descorrelacionada como sea posible. Así, ninguna o sólo una pequeña cantidad del procesamiento de descorrelación necesita aplicarse, lo cual es favorable, como se analizó anteriormente.

VII. Aplicaciones

Previamente mencionamos un número de aplicaciones para los esquemas de codificación propuestos. Aquí, resumimos estas y mencionamos unas cuantas aplicaciones más.

A. Codificación de audio para mezcla

Cada vez que las señales de fuente de audio necesiten almacenarse o transmitirse antes de mezclarlas a las señales de audio estéreo, de múltiples canales o de síntesis de campo de onda, puede aplicarse el esquema propuesto. Con la técnica anterior, un codificador de audio mono podría aplicarse a cada señal de fuente independientemente, dando como resultado una tasa de bits que escala con el número de fuentes. El esquema del código propuesto puede codificar un alto número de señales de fuente de audio con un codificador de audio mono único más la información secundaria de tasa de bits relativamente baja. Como se describe en la Sección V, la calidad de audio puede mejorarse usando más de un canal transmitido, si la memoria/capacidad para hacerlo está disponible.

B. Remezcla con metadatos

Como se describió en la Sección VI, las señales de audio estero y de múltiples canales existentes, pueden remezclarse con la ayuda de información secundaria adicional (es decir “metadatos”). En oposición a comercializar únicamente contenido de audio mezclado de estéreo y de múltiples canales optimizado, los meta-datos pueden comercializarse permitiendo a un usuario la remezcla de su música estéreo y de múltiples canales. Esto puede usarse también, por ejemplo, para atenuar las voces en una canción para karaoke o para atenuar instrumentos

específicos para tocar un instrumento junto con la música.

Aún si el almacenamiento no fuera un problema, el esquema descrito sería muy atractivo para posibilitar la mezcla de música personalizada. Es decir, debido a que probable que la industria de la música nunca estará dispuesta a suministrar parte de las grabaciones de múltiples pistas. Existe demasiado peligro para el abuso. El esquema propuesto posibilita la capacidad de la remezcla sin apartar las grabaciones de múltiples pistas.

Asimismo, tan pronto como se remezclan las señales de estéreo o de múltiples canales, tiene lugar un cierto grado de reducción de calidad, haciendo la distribución ilegal de la remezcla menos atractiva.

c. Estéreo/múltiples canales a la conversión de síntesis de campo de ondas

Otra aplicación para el esquema descrito en la Sección VI se describe a continuación. Las películas en movimiento que acompañan el audio de estéreo y de múltiples canales (por ejemplo 5.1 envolvente), pueden extenderse para representación de síntesis de campo de ondas, agregando información secundaria. Por ejemplo Dolby AC-3 (audio para DVD) puede extenderse para el audio de codificación de compatibilidad hacia atrás 5.1 para los sistemas de síntesis del campo de ondas, es decir los DVD reproducen sonido envolvente 5.1 en reproductores heredados convencionales y sonido de síntesis de campo de ondas en una nueva generación de reproductores que soportan el procesamiento de la información secundaria.

VIII: evaluaciones subjetivas

Implementamos un descodificador de tiempo real de los algoritmos propuestos en la Sección IV-A y IV-B. Se usa un banco de filtros STFT basado en FFT. Se usa una FFT de 1024 puntos y un tamaño de ventana de STFT de 768 (con relleno de cero). Los coeficientes espectrales se agrupan juntos de modo que cada grupo represente la señal con un ancho de banda de dos veces el ancho de anda rectangular equivalente (ERB). El oyente informal reveló que la calidad de audio no mejora notablemente cuando se escoge la resolución de frecuencia mayor. Una resolución de frecuencia menor es favorable puesto que da como resultado que han de transmitirse menos parámetros.

Para cada fuente, la panorámica de amplitud/retardo y ganancia pueden ajustarse individualmente. El algoritmo se usó para codificar varias grabaciones de audio de múltiples pistas de 12 a 14 pistas.

El descodificador permite mezcla envolvente de 5.1 usando un mezclador de panorámica de amplitud de base vectorial (VBPAP). La dirección y ganancia de cada señal de fuente puede ajustarse. El software permite la conmutación al vuelo entre la mezcla de la señal de fuente codificada y la mezcla de las señales de fuente discretas originales.

El oyente casual normalmente revela nada o poca diferencia entre la mezcla de las señales de fuente originales o codificadas si para cada fuente se usa una ganancia G_i de cero dB. Cuanto más ganancias de fuente se varían, más cantidad de artefactos ocurren. La amplificación y atenuación leves de las fuentes (por ejemplo, hasta ± 6 dB) sonarán aún bien. Un escenario crítico es cuando todas la fuentes se mezclan en un lado y solamente una fuente al otro lado opuesto. En este caso, la calidad de audio puede reducirse, dependiendo de la mezcla específica y las señales de fuente.

IX. conclusiones

Se propuso un esquema de codificación para la codificación de conjunta de las señales de 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 de fuente con alta calidad, caso en el que la codificación conjunta proporcionaría mínima ganancia de codificación, puesto que las fuentes de audio son normalmente independientes. El objetivo es que cuando las señales de fuente codificadas se mezclan se obtenga una señal de audio de alta calidad. Considerando las propiedades estadísticas de las señales de fuente, las propiedades de los esquemas de mezcla, y la escucha espacial, muestran que se logra una mejora en la ganancia de codificación significativa por la codificación conjunta de las señales de fuente.

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

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

La tasa de información secundaria es de alrededor de 3 kbs por señal de fuente. Cualquier mezclador puede aplicarse con las señales de fuente codificadas, por ejemplo mezcladores de síntesis de estéreo, de múltiples canales o de campo de onda.

Es fácil escalar el esquema propuesto para la tasa de bits mayor y la calidad por medio de transmitir más de un

canal de audio. Asimismo, se propuso una variación del esquema, que permite la remezcla de la señal de audio de estéreo o de múltiples canales dada (e incluso cambiando el formato de audio, por ejemplo la síntesis de estéreo a múltiples canales o al campo de ondas).

Las aplicaciones del esquema propuesto son múltiples. Por ejemplo MPEG-4 puede ampliarse con el esquema propuesto para reducir la tasa de bits cuando más de un “objeto de audio natural” (señal de fuente) necesita transmitirse. Igualmente, el esquema propuesto ofrece la representación compacta del contenido de los sistemas de síntesis de campo de ondas. Como se mencionó, las señales etéreo o de múltiples canales existentes pueden complementarse con la información secundaria para permitir que el usuario remezcle las señales a su gusto.

Referencias

[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 y F. Baumgarte, “Binaural Cue Coding - Part II: Schemes and applications”, IEEE Trans. on Speech and Audio Proc., vol. 11, n.º 6, noviembre de 2003

REIVINDICACIONES

1. Método de codificación de una pluralidad de señales de fuente ($s_1(n)$, $s_2(n)$,⋯, $s_M(n)$), que comprende:

calcular, para la pluralidad de señales de fuente ($s_1(n)$,$s_2(n)$,⋯,$s_M(n)$), información estadística que representa envolventes espectrales de las señales de fuente ($s_1(n)$,$s_2(n)$,⋯,$s_M(n)$) de la pluralidad de señales de fuente, en el que la información estadística comprende adicionalmente, para cada señal de fuente de la pluralidad de señales de fuente y para cada subbanda de una pluralidad de subbandas, información sobre una función de autocorrelación de subbanda normalizada (($\Phi_i(n,e)$)) de una señal de fuente específica; y

transmitir la información estadística que representa envolventes espectrales de las señales de fuente ($s_1(n)$,$s_2(n)$,⋯,$s_M(n)$) y la información sobre la función de autocorrelación de subbanda normalizada para cada señal de fuente de la pluralidad de señales de fuente y para cada subbanda de la pluralidad de subbandas como metadatos para una señal de audio derivada de la pluralidad de señales de fuente ($s_1(n)$,$s_2(n)$,⋯,$s_M(n)$).

2. Aparato para codificar una pluralidad de señales de fuente ($s_1(n)$,$s_2(n)$,⋯,$s_M(n)$), en el que el aparato es operativo para:

calcular, para la pluralidad de señales de fuente ($s_1(n)$,$s_2(n)$,⋯,$s_M(n)$), información estadística que representa envolventes espectrales de las señales de fuente ($s_1(n)$,$s_2(n)$,⋯,$s_M(n)$), en el que la información estadística comprende adicionalmente, para cada señal de fuente de la pluralidad de señales de fuente y para cada subbanda de una pluralidad de subbandas, información sobre una función de autocorrelación de subbanda normalizada (($\Phi_i(n,e)$)) de una señal de fuente específica; y

transmitir la información estadística que representa envolventes espectrales de las señales de fuente ($s_1(n)$,$s_2(n)$,⋯,$s_M(n)$) y la información sobre la función de autocorrelación de subbanda normalizada para cada señal de fuente de la pluralidad de señales de fuente y para cada subbanda de la pluralidad de subbandas como metadatos para una señal de audio derivada de la pluralidad de señales de fuente ($s_1(n)$,$s_2(n)$,⋯,$s_M(n)$).

3. El aparato de la reivindicación 2,
en el que el cálculo comprende calcular la información sobre la función de autocorrelación de subbanda normalizada de la señal de fuente específica durante un intervalo de tiempo determinado mezclando parámetros para mezclar la pluralidad de señales de fuente, y en el que el transmisor está configurado para transmitir la información sobre la función de autocorrelación de subbanda normalizada de la señal de fuente específica para el intervalo de tiempo como metadatos para una señal de audio derivada de la pluralidad de señales de fuente.

4. El aparato de la reivindicación 2,
en el que la transmisión comprende transmitir, como la información transmitida, un índice que identifica una función de autocorrelación predefinida en un conjunto almacenado de funciones de autocorrelación predefinidas.

5. El aparato de la reivindicación 2,
en el que el cálculo comprende calcular la función de autocorrelación de subbanda normalizada de la señal de fuente específica para un intervalo de tiempo determinado basándose en la siguiente ecuación:

$$\min_i \{T_i\} - T_i \leq e \leq \max_i \{T_i\} - T_i$$

en la que e es el intervalo de tiempo, $\min_i \{T_i\}$ es un parámetro de mezcla más pequeño entre los parámetros de mezcla para las señales de fuente de la pluralidad de señales de fuente, $\max_i \{T_i\}$ es un parámetro de mezcla mayor entre los parámetros de mezcla para las señales de fuente de la pluralidad de señales de fuente, y T_i es un parámetro de mezcla para la señal de fuente específica bajo consideración.

6. El aparato de la reivindicación 2,
en el que el cálculo comprende calcular la función de autocorrelación de subbanda normalizada de la señal de fuente específica para un intervalo de tiempo basándose en la siguiente ecuación:

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

en el que $\Phi(n,e)$ es la función de autocorrelación de subbanda normalizada de la señal de fuente específica, n es un índice de tiempo, e es el intervalo de tiempo, E es un operador de expectativa, i es un índice que identifica una señal de fuente, y s_i es la señal de fuente específica bajo consideración.

7. El aparato de la reivindicación 2,
en el que el cálculo comprende calcular, para una fuente, como la información estadística que representa envolventes espectrales de las señales de fuente ($s_1(n),s_2(n),\cdots,s_M(n)$), una potencia de subbanda para cada subbanda de la pluralidad de subbandas, o parámetros de filtro de malla o parámetros de LPC o parámetros de par espectral de línea, y
en el que la transmisión comprende transmitir, como la información estadística que representa envolventes espectrales de las señales de fuente ($s_1(n),s_2(n),\cdots,s_M(n)$), las potencias de subbanda para cada subbanda de la pluralidad de subbandas para cada señal de fuente o los parámetros de filtro de malla o los parámetros de LPC o los parámetros del par espectral de línea como los metadatos.

8. El aparato de la reivindicación 2,
en el que el calculador está configurado para seleccionar, en cada tiempo y para cada subbanda, la potencia de la fuente más intensa y para reducir el límite de la potencia de subbanda correspondiente de todas las otras fuentes a un valor de 24 dB inferior a la fuente de subbanda más intensa para limitar un rango dinámico de un cuantificador.

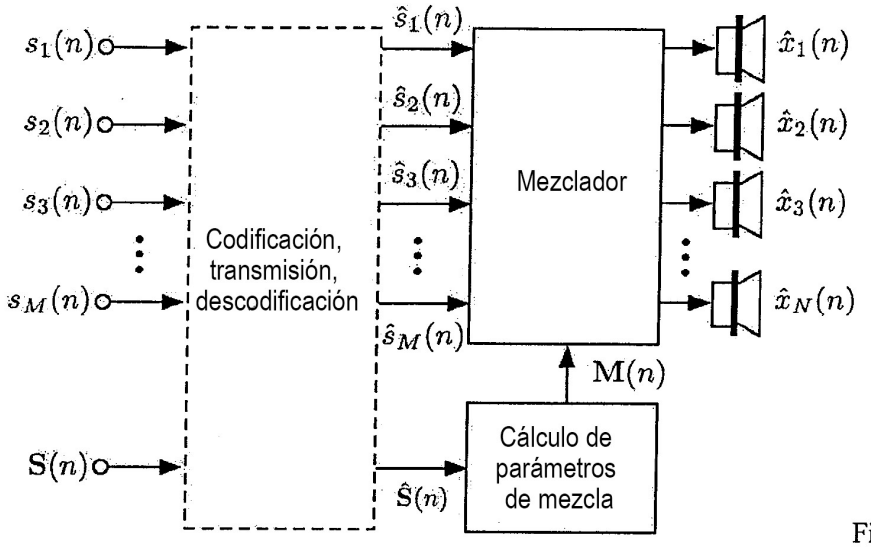

Fig. 1

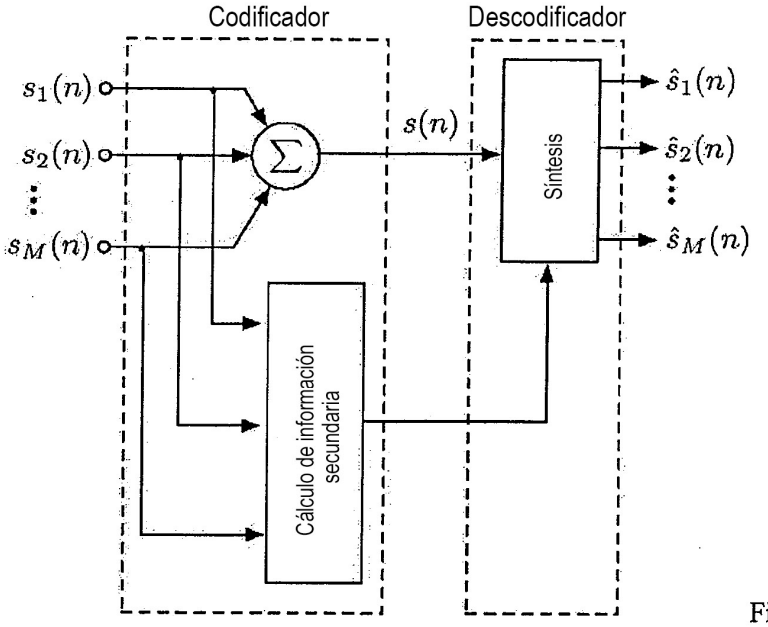

Fig. 2

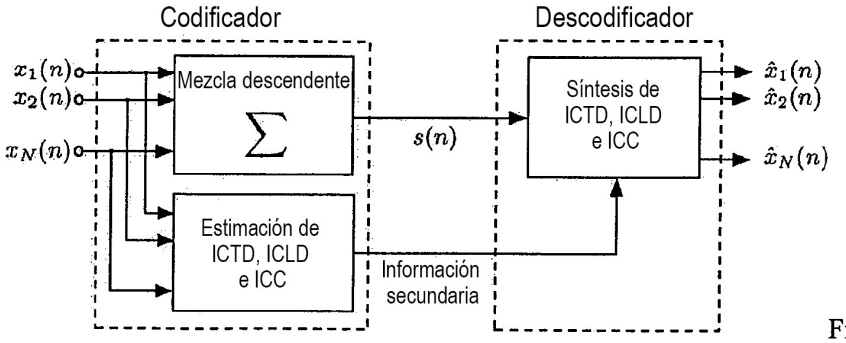

Fig. 3

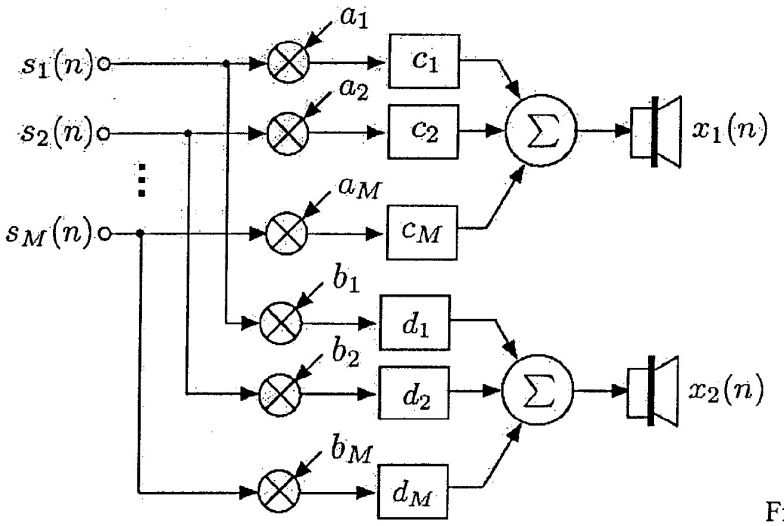

Fig. 4

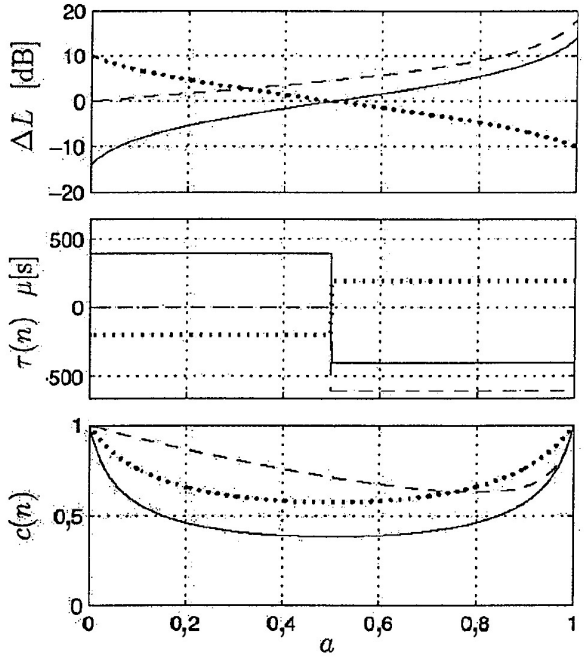

Fig. 5

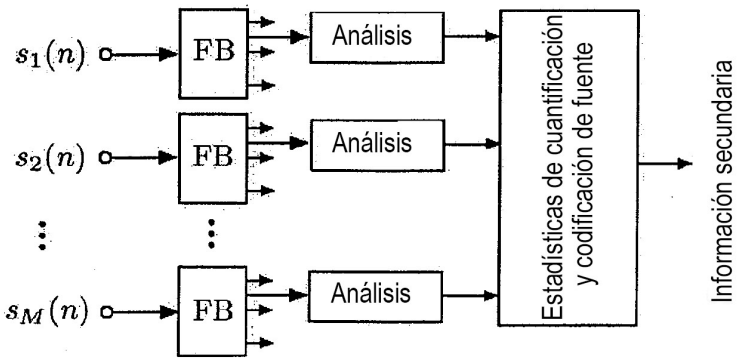

Fig. 6

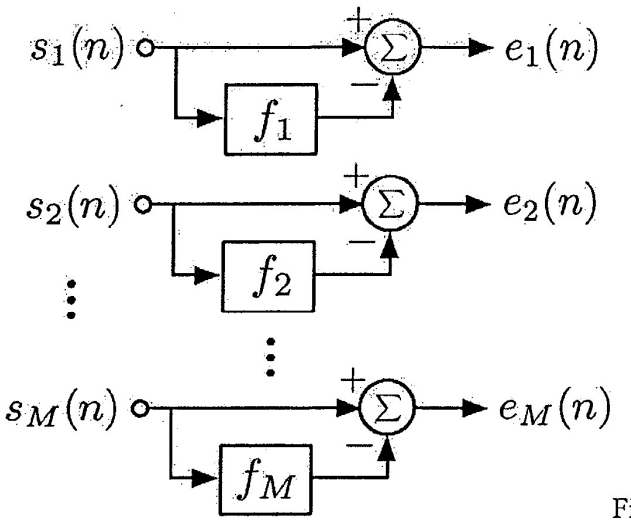

Fig. 7

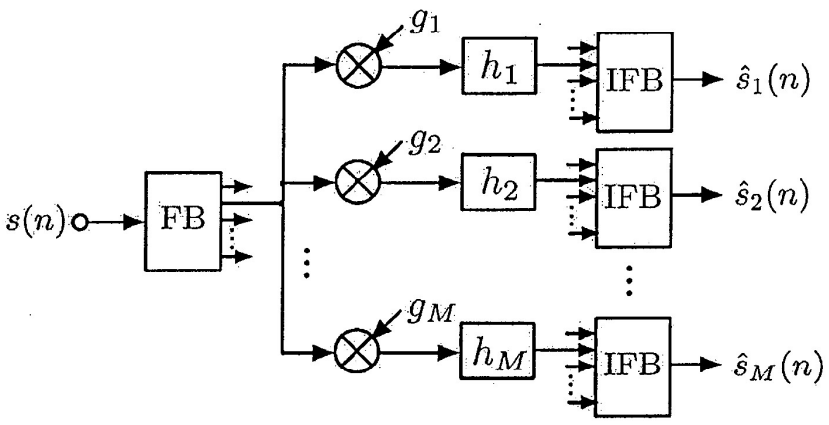

Fig. 8

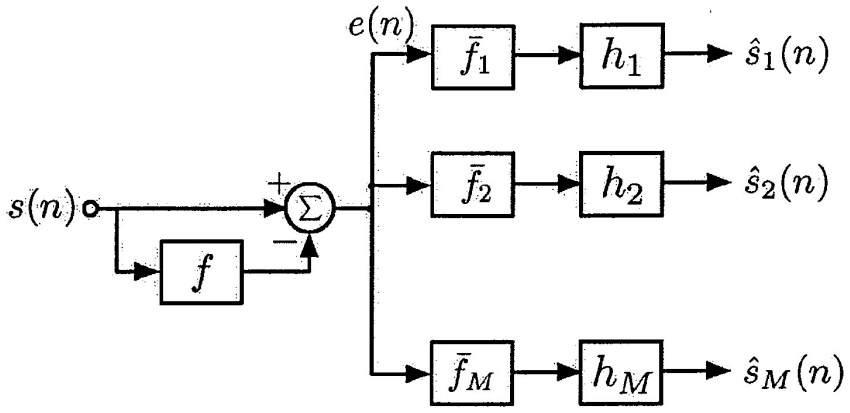

Fig. 9

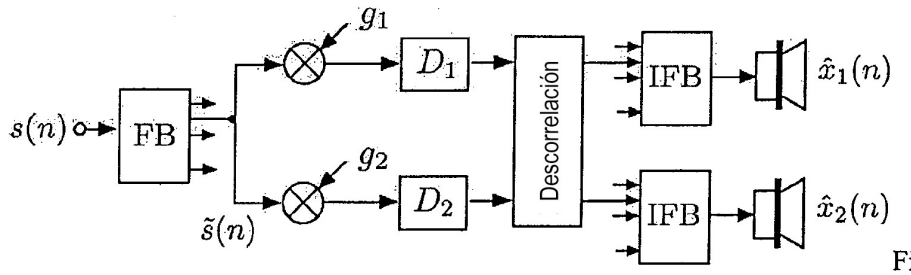

Fig. 10

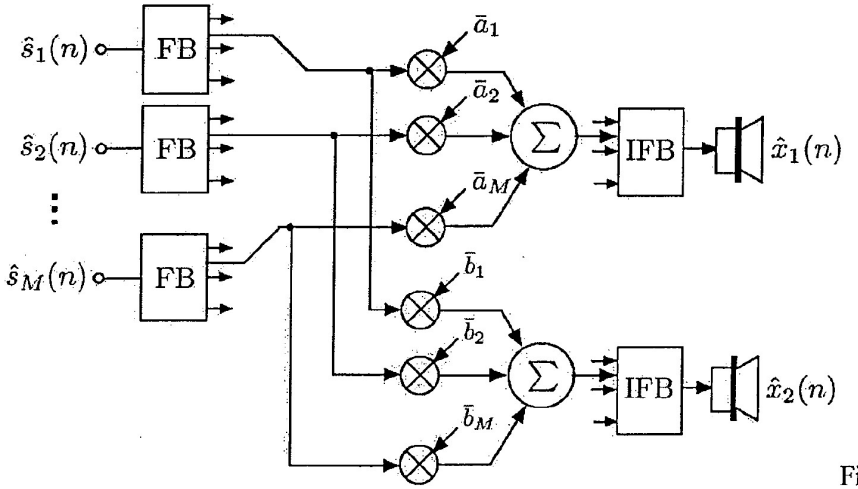

Fig. 11

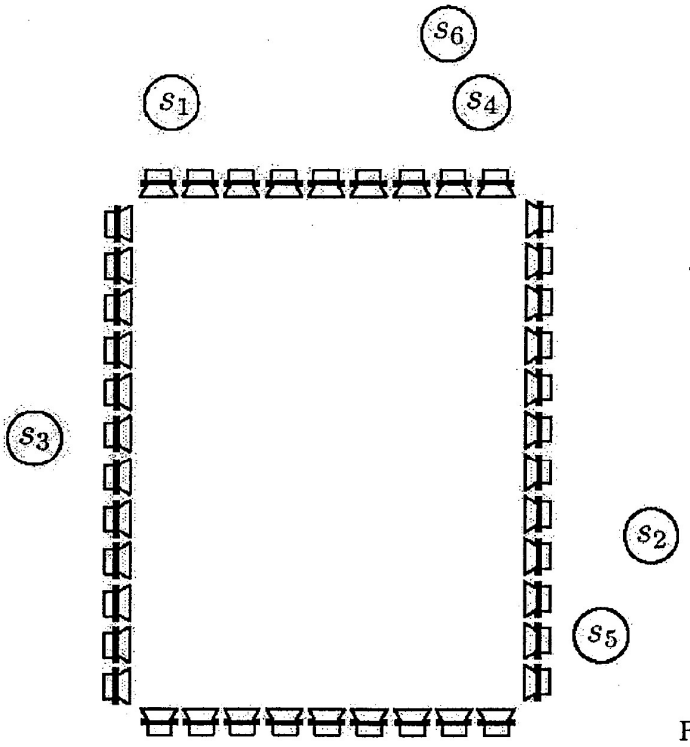

Fig. 12

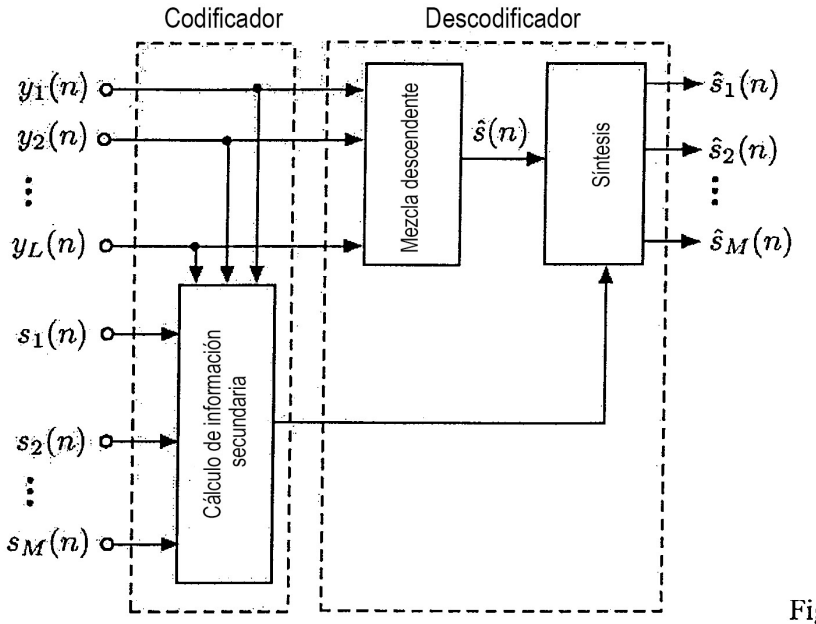
Codificador
Descodificador
$y_1(n)$
$y_2(n)$
$y_L(n)$
$s_1(n)$
$s_2(n)$
$s_M(n)$
Cálculo de información secundaria
Mezcla descendente
$\hat{s}(n)$
Síntesis
$\hat{s}_1(n)$
$\hat{s}_2(n)$
$\hat{s}_M(n)$

Fig. 13

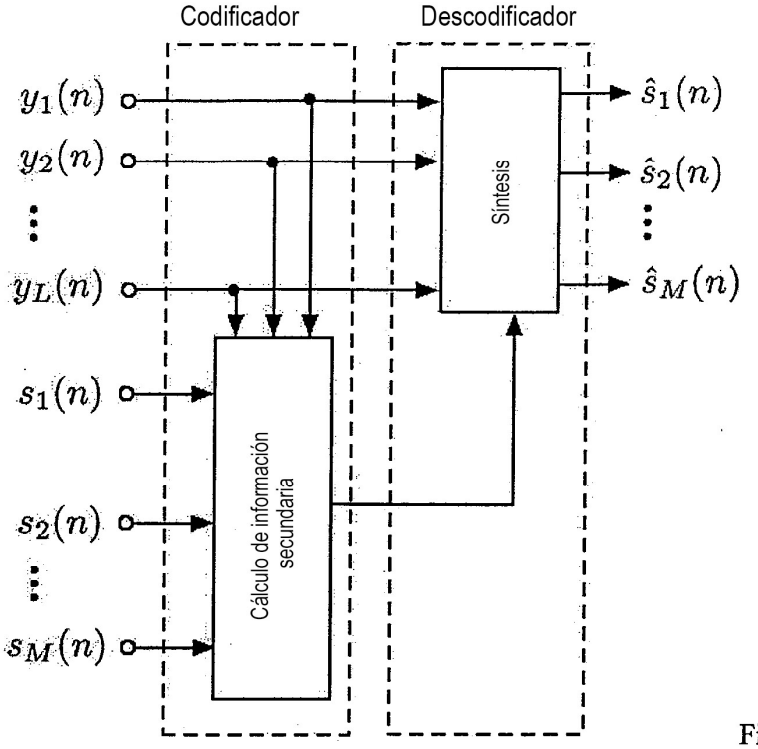
Codificador
Descodificador
$y_1(n)$
$y_2(n)$
$y_L(n)$
$s_1(n)$
$s_2(n)$
$s_M(n)$
Cálculo de información secundaria
Síntesis
$\hat{s}_1(n)$
$\hat{s}_2(n)$
$\hat{s}_M(n)$

Fig. 14