SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

No. 11-177816- -00000-0000

Fecha: 2011-12-23 14:58:29 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 379 FASENACIONALII
Act. 411 PRESENTACION Folios: 349

Industria y Comercio

SUPERINTENDENCIA

DELEGATURA PROPIEDAD INDUSTRIAL

SOLICITUD
REGISTRO DE PATENTE DE INVENCION
PCT

EXPEDIENTE No _____

TITULO DECODIFICADOR DE SEÑALES DE AUDIO, MÉTODO PARA DECODIFICAR UNA SEÑAL DE AUDIO Y PROGRAMA DE COMPUTACIÓN QUE UTILIZA ETAPAS EN CASCADA DE PROCESAMIENTO DE OBJETOS DE AUDIO

CLASIFICACION INTERNACIONAL _____

SOLICITANTE FRAUNHOFER-GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG E.V.

DOMICILIO MUNICH, ALEMANIA

APODERADO JULY GALINDO QUINTERO

PETITORIO

Industria y Comercio
SUPERINTENDENCIA

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
No. 11-177816- -00000-0000
Fecha: 2011-12-23 14:58:29 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 379 FASENACIONALII
Act. 411 PRESENTACION Folios: 349

FORMULARIO UNICO DE
PCT
ENTRADA EN FASE NACIONAL

① **SOLICITUD DE:**

[X] Patente de Invención [ ] Patente de Modelo de Utilidad

[ ] Capítulo I [X] Capítulo II

② **SOLICITANTE**

Nombre: **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.**
Dirección: Hansastrasse 27c 80686 Munich, Alemania
Nacionalidad o Domicilio: Munich, Alemania
Lugar de Constitución: Alemania
Teléfono: 3137800 Fax: 3122410
E-mail: general@cardenasycardenas.com

IDENTIFICACION
[ ] NIT [ ] NIT
[ ] Otro [X] Otro
Cuál
Número

③ **REPRESENTANTE O APODERADO**

Nombre: **JULY GALINDO QUNTERO**
Dirección: Carrera 7 No 71-52 Torre B Piso 9
Teléfono: 312 3600 Fax: 312 2410
E-Mail: **general@cardenasycardenas.com**

IDENTIFICACION
[ ] NIT [ ] NIT
[X] CC [ ] Otro
Cual 52996295 BTA
Número 158806 C.S.J.

④ **INVENTOR(ES) (72)**

Nombre:
1. Oliver Hellmuth
2. Cornelia Falch
3. Jürgen Herre
4. Johannes Hilpert
5. Leonid Terentiev
6. Falko Ridderbusch

Dirección:
1. Geschwister – Voemel- Weg 60 91052 Erlangen, Alemania
2. Kasseler Str. 12 90491 Nürnberg, Alemania
3. Hallerstr. 24 91054 Buckenhof, Alemania
4. Herrnhüttestraße 46 90411 Nürnberg, Alemania
5. Am Europakanal 26, App. 11, 91056 Erlangen, Alemania
6. Adam – Kraft- Str. 57, 90419 Nürnberg , Alemania

Nacionalidad o Domicilio:
1. Alemán
2. Alemán
3. Alemán
4. Alemán
5. Alemán
6. Alemán

⑤ **PCT (86)**

Solicitud Internacional No. PCT/EP2010/058906 Fecha 23.06.2010
Publicación Internacional No. WO 2010/149700 Fecha 29.12.2010

⑥ **Título(54):**

DECODIFICADOR DE SEÑALES DE AUDIO, MÉTODO PARA DECODIFICAR UNA SEÑAL DE AUDIO Y PROGRAMA DE COMPUTACIÓN QUE UTILIZA ETAPAS EN CASCADA DE PROCESAMIENTO DE OBJETOS DE AUDIO

⑦ **Clasificación Internacional (51):**

⑧ **Prioridad** SI [X] NO [ ]

(33)País de Origen	(32)Fecha	(31)N° de Solicitud
US	24.06.2009	61/220,042

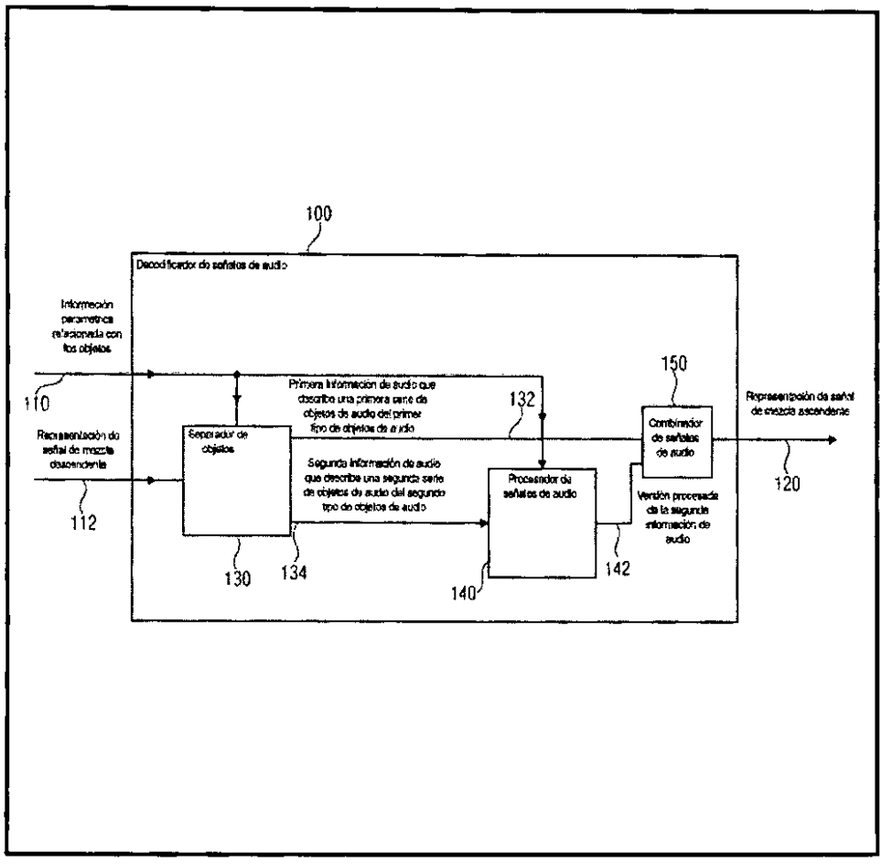
⑨ **Comprobante de Pago No.:** 11-124657 **Fecha** 22.12.2011

P-619

⑩ ANEXOS

- [X] Comprobante de pago de la tasa de presentación de la solicitud.
- [ ] Documento que acredita la existencia y representación legal cuando el solicitante sea persona jurídica.
- [X] Poderes, si fuere el caso.
- [ ] Documento de cesión del inventor al solicitante o a su causante.
- [X] Traducción de la solicitud internacional al castellano (Resumen, descripción, reivindicaciones, dibujos)
- [X] Copia del examen preliminar internacional efectuado por la Autoridad Internacional de Examen Preliminar (IPEA)
- [ ] Traducción del examen preliminar internacional efectuado por la IPEA.
- [ ] De ser el caso, copia del contrato de acceso.
- [ ] De ser el caso, documento que acredite la licencia o autorización de uso de conocimientos tradicionales de las comunidades indígenas.
- [ ] De ser el caso, certificado de depósito del material biológico.
- [X] Arte final 12x12.
- [ ] De ser el caso, información sobre otras solicitudes de patente o títulos obtenidos en el extranjero por el mismo titular o su causante, relacionadas parcial o totalmente con la invención de esta solicitud.
- [ ] De ser el caso, modificaciones efectuadas a la solicitud internacional.
- [X] Prioridad http://www.wipo.int/patentscope/search/fr/detail.jsf;jsessionid=035BD85B1B1C7ED41B6997E86F10484D.wapp1?docId=WO2010149700&recNum=1&tab=PCTDocuments&maxRec=&office=&prevFilter=&sortOption=&queryString=
- [X] Recibo de Reivindicaciones adicionales (27) 11-124658

11 FIGURA CARACTERISTICA



12 Solicito la Concesión de la patente

NOMBRE: JULY GALINDO QUINTERO

FIRMA:

C.C. 52996295 BTA TP. 158806 C.S.J.

P-583

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
NIT : 800.176.089-2
- / -

Industria y Comercio
SUPERINTENDENCIA

RECIBO DE CAJA

No. 11 - 124657

Bogotá D.C., Diciembre 22 de 2011 - 14:16:46

RECIBIDO DE : CARDENAS & CARDENAS ABOGADOS PI LTDA NI 830.083.491

*** Soporte del Pago ***

TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PAGO	VR.PAGO
CONSIGNACION	BANCO DE BOGOTA	062754387	456036700	22/12/2011	9.474.000.00

*** Conceptos Pagados ***

CANT.	RENTISTICO	CONCEPTO	Vr.UNDITARIO	Vr.CONCEPTO
1	50005-01-01 SOLICITUDES	1 TRAMITES DE SOL. DE PATENTE DE INVENCION	571.000.00	571.000.00
				$571.000.00

SON: **QUINIENTOS SETENTA Y UN MIL PESOS MONEDA CORRIENTE***

Responsable: _____
*****

Recibo de Caja Aplicado al Expediente No. _____

CARDENAS & CARDENAS
ABOGADOS

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
No. 11-177816- -00000-0000
Fecha: 2011-12-23 14:58:29 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 379 FASENACIONALII
Act. 411 PRESENTACION Folios: 349

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
NIT : 800.176.089-2
- / -

Industria y Comercio
SUPERINTENDENCIA

RECIBO DE CAJA

No. 11 - 124658

Bogotá D.C., Diciembre 22 de 2011 - 14:16:46

RECIBIDO DE : CARDENAS & CARDENAS ABOGADOS PI LTDA NI 830.083.491

*** Soporte del Pago ***

TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PAGO	VR.PAGO
CONSIGNACION	BANCO DE BOGOTA	062754387	456036700	22/12/2011	9.474.000.00

*** Conceptos Pagados ***

CANT.	RENTISTICO	CONCEPTO	Vr.UNDITARIO	Vr.CONCEPTO
1	50005-01-09 OTROS SERVICIOS DE PROPIEDAD INDUSTRIAL	99 OTROS SERVCIOS ADMINISTRATIVOS	729.000.00	729.000.00
				$729.000.00

SON: **SETECIENTOS VEINTINUEVE MIL PESOS MONEDA CORRIENTE***

Responsable: ******

Recibo de Caja Aplicado al Expediente No. _____

CARDENAS & CARDENAS
ABOGADOS

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
No. 11-177816- -00000-0000
Fecha: 2011-12-23 14:58:29 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 379 FASENACIONALII
Act. 411 PRESENTACION Folios: 349

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO S

DIVISION DE NUEVAS CREACIONES

TARJETA ARCHIVO PROPIETARIOS

PATENTE DE INVENCION _X___ PCT ___ MODELO DE UTILIDAD ____ DISEÑO INDUSTRIAL ___

Industria y Comercio

SUPERINTENDENCIA

(21) N° de solicitud: (22) Fecha de solicitud:

(51) Clasificación Internacional :

(71)Solicitante(s) **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.**

(72) Inventor(es) **OLIVER HELLMUTH, CORNELIA FALCH, JÜRGEN HERRE, JOHANNES HILPERT, LEONID TERENTIEV, FALKO RIDDERBUSCH**

(74) Apoderado: JULY GALINDO QUINTERO

(30) Prioridad	(31) ***No. Prioridad***	32) ***Fecha***	(33) ***País***
	61/220,042	**24.06.2009**	**US**

(85) Fecha inicio fase nacional: 24.01.2012

(86) Datos relativos a la presentación de la solicitud PCT

Fecha de presentación de la solicitud: 23.06.2010

No. de solicitud *PCT/EP2010/058906*

(54) Titulo **DECODIFICADOR DE SEÑALES DE AUDIO, MÉTODO PARA DECODIFICAR UNA SEÑAL DE AUDIO Y PROGRAMA DE COMPUTACIÓN QUE UTILIZA ETAPAS EN CASCADA DE PROCESAMIENTO DE OBJETOS DE AUDIO**

Industria y Comercio
SUPERINTENDENCIA

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
DIVISION DE NUEVAS CREACIONES
TARJETA ARCHIVO TEMATICO

PATENTE DE INVENCION ___ PCT ____ MODELO DE UTILIDAD ___ DISEÑO INDUSTRIAL

(21) N° de solicitud: (22) Fecha de solicitud:
(51) Clasificación Internacional :

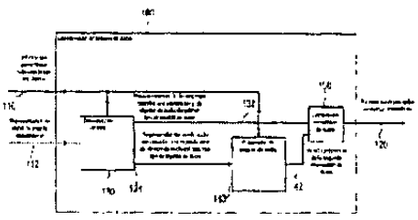
(71)Solicitante(s) **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.**

(72) Inventor(es) **OLIVER HELLMUTH, CORNELIA FALCH, JÜRGEN HERRE, JOHANNES HILPERT, LEONID TERENTIEV, FALKO RIDDERBUSCH**

(74) Apoderado: JULY GALINDO QUINTERO

(30) Prioridad

(31) ***No. Prioridad***	32) ***Fecha***	(33) ***País***
61/220,042	**24.06.209**	**US**



(85) Fecha límite inicio fase nacional: 24.01.2012
(86) Datos relativos a la presentación de la solic. PCT
Fecha de presentación de la solicitud: 23.06.2010
No. de solicitud *PCT/EP2010/058906*

(87) Publicación internacional:
Fecha: ***29.12.2010***
No. Publicación WO 2010/149700

(54) Titulo: **DECODIFICADOR DE SEÑALES DE AUDIO, MÉTODO PARA DECODIFICAR UNA SEÑAL DE AUDIO Y PROGRAMA DE COMPUTACIÓN QUE UTILIZA ETAPAS EN CASCADA DE PROCESAMIENTO DE OBJETOS DE AUDIO**

(57) Resumen:

Un decodificador de señales de audio para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente e información paramétrica relacionada con los objetos comprende un separador de objetos configurado para descomponer la representación de señal de mezcla descendente, a fin de proporcionar una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio, que depende de la representación de la señal de mezcla descendente y usar por lo menos parte de la información paramétrica relacionada con los objetos. El decodificador de señales de audio comprende asimismo un procesador de señales de audio configurado para recibir la segunda información de audio y para procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, para obtener una versión procesada de la segunda información de audio. El decodificador de señales de audio comprende asimismo un combinador de señales de audio configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente.

Industria y Comercio
SUPERINTENDENCIA

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
DIVISION DE NUEVAS CREACIONES
EXTRACTO PARA PUBLICACIÓN
SOLICITUD DE PATENTE DE INVENCION

(21) N° de solicitud:

22) Fecha de solicitud:

(30) Prioridad (31) ***No. Priodidad*** 32) ***Fecha*** (33) ***País***
61/220,042 **24.06.2009** **US**

(51) Int Cl:

(71)Solicitante(s) **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.**

(72) Inventor(es) **OLIVER HELLMUTH, CORNELIA FALCH, JÜRGEN HERRE, JOHANNES HILPERT, LEONID TERENTIEV, FALKO RIDDERBUSCH**

(74) Apoderado: **JULY GALINDO QUINTERO**

(54) Titulo: **DECODIFICADOR DE SEÑALES DE AUDIO, MÉTODO PARA DECODIFICAR UNA SEÑAL DE AUDIO Y PROGRAMA DE COMPUTACIÓN QUE UTILIZA ETAPAS EN CASCADA DE PROCESAMIENTO DE OBJETOS DE AUDIO**

Reivindicación(es):

Un decodificador de señales de audio para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente e información paramétrica relacionada con los objetos comprende un separador de objetos configurado para descomponer la representación de señal de mezcla descendente, a fin de proporcionar una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio, que depende de la representación de la señal de mezcla descendente y usar por lo menos parte de la información paramétrica relacionada con los objetos. El decodificador de señales de audio comprende asimismo un procesador de señales de audio configurado para recibir la segunda información de audio y para procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, para obtener una versión procesada de la segunda información de audio. El decodificador de señales de audio comprende asimismo un combinador de señales de audio configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente.

- Colombia -

POWER OF ATTORNEY

I, Werner Liebler, confirm:

FIRST: That I act in my capacity of deputy chief of the Patent Department of **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.**, legally constituted and in existence in accordance with the laws of Germany, and domiciled at Hansastrasse 27c, 80686 Muenchen, Germany.

SECOND: That in such capacity I grant hereby special, ample and sufficient Power of Attorney to DARIO CARDENAS-NAVAS and/or LUZ HELENA ADARVE-GOMEZ and/or EDUARDO CARDENAS-CABALLERO and/or BERNARDO P. CARDENAS-MARTINEZ and/or JUANITA ACOSTA GOMEZ and/or JULY GALINDO QUINTERO, registered lawyers of the firm CARDENAS & CARDENAS-LAWYERS, domiciled at Carrera 7 No. 71-52, Torre B, Piso 9 Bogotá, D.C., Colombia, to carry before the corresponding Colombian national Offices and authorities the proceedings to obtain trademarks, patents, utility models, industrial drawings and designs, deposits of commercial names and ensigns, as well as processing of renewals, assignments, change of name or domicile of the owner, waivers and licenses of the same for which purpose they are empowered to take before said authorities all necessary actions for the indicated purposes; to file applications, statements, claims, formulate descriptions, amendments, oppositions and appeals; pay all taxes and installments and make any other payments determined by law; to receive every document and values issuing the respective discharge; to fulfill any other requirements and take, in general, every measure that they shall deem opportune for the defense and protection of the Industrial Property of the company in the event of infringements, oppositions, proceedings for cancellation or annulment, imposition of fines, passing the antecedents to the courts; they are empowered to appear as plaintiff or defendant before the competent judges, being entitled to transact, submit to arbitration, waive, receive, resign, withdraw, appeal and all other legal powers which may be necessary, and it is hereby declared ever since now valid and good whatever such attorneys do for its benefit, granting them also power to substitute this proxy and if necessary to revoke such substitutions.

SIGNED AT Dezember 6th, 2010 ON Munich / Germany

BY Werner Liebler

NOTARIA 16 DEL CIRCULO DE BOGOTA D.C.
COMO NOTARIO (E) DE ESTE CIRCULO, HAGO CONSTAR
QUE ESTA COPIA, CONCIDE CON DOCUMENTO ORIGINAL
PRESENTADO PARA SU AUTENTICACION
22 DIC 2011
JAIRO FREDY SATIZABAL HURTADO
Notaria 16 Suplente

File No. H2294 **/ 2010**

ag

I hereby certify that above is the true signature, subscribed in my presence of

Mr. Dr. Werner Liebler,
born on May 6th, 1961,
business address: Hansastr. 27 c, 80686 Munich,
personally known to me.

Dr. Werner Liebler acting by virtue of
Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.
with its registered office in Munich,
business address: 80686 München, Hansastr. 27 c,
of power of attorney dated on July 13th, 2000, File No. 3076 K /2000, and July 18th, 2000, File No. 3170 K /2000, by Notary Public, Dr. Helmut Keidel, Munich, presented to me today in the original and which is attached as certified copy.

Munich, Dezember 6th, 2010

Thomas Haasen,
Notary Public, Munich

22 DIC 2011

H:\Grunert\Unterschriftsbeglaubigungen\Dr. Liebler (Fraunhofer) engl 10-12-06.doc

APOSTILLE

(Convention de La Haye du 5 octobre 1961)

1. Land: Bundesrepublik Deutschland

Diese öffentliche Urkunde
2. ist unterschrieben von **Thomas Haasen**

3. in seiner Eigenschaft als **Notar**

4. Sie ist versehen mit dem Siegel des **Notars Thomas Haasen in München**

Bestätigt

5. in München 6. am 7. Dezember 2010

7. durch den Präsidenten des Landgerichts München I

8. unter Nr. 910 a - 9274/2010

9. Siegel 10. Unterschrift
Im Auftrag:

Christine Köppl
Amtsinspektorin

NOTARIA 16 DEL CIRCULO DE BOGOTA D.C.
COMO NOTARIO 16 (E) DE ESTE CIRCULO, HAGO CONSTAR
QUE ESTA COPIA, COINCIDE CON DOCUMENTO ORIGINAL
PRESENTADO PARA SU AUTENTICACION
22 DIC 2011
JAIRO FREDY SATIZABAL HURTADO
Notario 16 Encargado

Fraunhofer Gesellschaft

Fraunhofer-Gesellschaft Postfach 19 03 39 D-80603 München

Vorstand

Leonrodstraße 54
D-80636 München

Telefon +49 (0) 89/12 05-01
Telefax +49 (0) 89/12 05-700

Vollmacht

Hiermit bevollmächtigen wir

Herrn Dr. Werner Liebler
geb. 06.05.1961
geschäftsansässig: Patentstelle für
die Deutsche Forschung
Leonrodstraße 68
80636 München

die

Fraunhofer-Gesellschaft zur Förderung
der angewandten Forschung e.V.,
Leonrodstraße 54, 80636 München

bis auf Widerruf in allen Angelegenheiten des Gesetzes über Arbeitnehmererfindungen rechtsverbindlich zu vertreten. Darüberhinaus wird Herr Dr. Liebler bevollmächtigt, bei Schutzrechtsverfahren im Ausland lokale Patentanwälte zu beauftragen.

(Dr. D.-M. Polter)

(Dr. H.-U. Wiese)

Vorstand der Fraunhofer-Gesellschaft:
Prof. Dr.-Ing. Dr. h. c. mult.
Hans-Jürgen Warnecke, Präsident
Dr. jur. Dirk-Meints Polter
Dr. rer. pol. Hans-Ulrich Wiese

Fraunhofer-Gesellschaft zur Förderung
der angewandten Forschung e. V., München

Bankverbindung: Deutsche Bank, München
Konto 7521 933 BLZ 700 700 10

URNr. 3076 K

Ich beglaubige hiermit die Echtheit der vorstehenden, vor mir vollzogenen Unterschrift von

Herrn Dr. Dirk-Meints P o l t e r ,
Vorstandsmitglied in München,

geschäftsansässig Leonrodstr. 54, 80636 München,
persönlich bekannt.

Hierzu bescheinige ich aufgrund Einsicht in das Vereinsregister des Amtsgerichts München (VR-Nr. 4461) vom 12.07.2000, daß die Herren Dr. Dirk-Meints Polter und Dr. Hans-Ulrich Wiese zur Vertretung des Vereins

Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung eingetragener Verein mit dem Sitz in München

als Vorstandsmitglieder gemeinsam berechtigt sind.

München, den 13.07.2000

Dr. Keidel, Notar

NOTARIA 16 DEL CIRCULO DE BOGOTA D.C.
COMO NOTARIO 16 (E) DE ESTE CIRCULO, HAGO CONSTAR
QUE ESTA COPIA, COINCIDE CON DOCUMENTO ORIGINAL
PRESENTADO PARA SU AUTENTICACION.
22 DIC 2011
JAIRO FREDY SARIZABAL HURTADO
Notario 16 Encargado

URNr. 3170 K

Ich beglaubige hiermit die Echtheit der vorstehenden, vor mir vollzogenen Unterschrift von

Herrn Dr. Hans-Ulrich W i e s e ,
Vorstandsmitglied in München,

geschäftsansässig Leonrodstr. 54, 80636 München,
ausgewiesen durch amtlichen Lichtbildausweis.

Hierzu bescheinige ich aufgrund Einsicht in das Vereinsregister des Amtsgerichts München –VR-Nr. 4461- vom 12.07.2000, dass die Herren Dr. Dirk-Meints Polter und Dr. Hans-Ulrich Wiese zur Vertretung des Vereins

Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung eingetragener Verein mit dem Sitz in München

als Vorstandsmitglieder gemeinsam berechtigt sind.

München, den 18.07.2000

Dr. Keidel, Notar

22 DIC 2011

I hereby certify that this is a true copy of the original document.

Munich, 06. Dez. 2010

Notary Public in Munich

APOSTILLE

(Convention de La Haye du 5 octobre 1961)

1. Land: Bundesrepublik Deutschland

   Diese öffentliche Urkunde
2. ist unterschrieben von **Thomas Haasen**

3. in seiner Eigenschaft als **Notar**

4. Sie ist versehen mit dem Siegel des **Notars Thomas Haasen in München**

Bestätigt

5. in München 6. am 7. Dezember 2010

7. durch den Präsidenten des Landgerichts München I

8. unter Nr. 910 a – 9279/2010

9. Siegel 10. Unterschrift
   Im Auftrag:

   Christine Köppl
   Amtsinspektorin

FREISTAAT BAYERN · INNERHALB SEINER GRENZEN · DER PRÄSIDENT DES LANDGERICHTS

NOTARIA 10 DEL CIRCULO DE BOGOTÁ D.C.
COMO NOTARIO 10 (E) DE ESTE CIRCULO, HAGO CONSTAR
QUE ESTA COPIA, COINCIDE CON DOCUMENTO ORIGINAL
PRESENTADO PARA SU AUTENTICACION.
22 DIC 2011

Carrera 11D # 124 - 51 Apto. 510 Int. 2 - Teléfono 2139034 Bogotá - Colombia
E-mail: mssarmiento2@yahoo.com

Traductora Oficial Juramentada
Resolución No. 835 Minjusticia

TRADUCCIÓN OFICIAL DE LOS SIGUIENTES DOCUMENTOS: PODER DE ABOGADO otorgado por la Compañía Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., a la Firma Cárdenas y Cárdenas Abogados de Bogotá - Colombia, firmado por el señor Werner Liebler, Jefe Encargado del Departamento de Patentes de dicha Compañía, el 6 de diciembre de 2010. CERTIFICACIÓN NOTARIAL, Archivo No. H2294/2010, firmada por Thomas Haasen, Notario Público, en Munich, Alemania, con fecha del 6 de diciembre de 2010.

Estos documentos vienen acompañados por la APOSTILLA DE LA CONVENCIÓN DE LA HAYA No. 910 a - 9274/2010, en idioma alemán, con fecha del 7 de diciembre de 2010, firmada por Christine Köppl.

Traducción realizada por Martha Sofía Sarmiento Solano, Traductora e Intérprete Oficial Español – Inglés - Español, nombrada para el cargo mediante Resolución del Ministerio de Justicia, N° 835, del 3 de mayo de 1988.

- COLOMBIA -

PODER DE ABOGADO

Yo, Werner Liebler, declaro lo siguiente:

PRIMERO: Que **yo** actúo en mi calidad de **Jefe Encargado del Departamento de Patentes** de **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.,** una Compañía legalmente constituida y en existencia de acuerdo con las leyes de Alemania y cuyo domicilio es Hansastrasse 27c, 80686 Muenchen Germany/Alemania.

SEGUNDO: Que haciendo uso de las facultades del cargo arriba mencionado, yo otorgo Poder de Abogado, especial, amplio y suficiente a **DARÍO CÁRDENAS NAVAS**, y/o a **LUZ HELENA ADARVE GÓMEZ**, y/o a **EDUARDO CÁRDENAS CABALLERO**, y/o a **BERNARDO P. CÁRDENAS MARTÍNEZ**, y/o a **JUANITA ACOSTA GÓMEZ**, y/o a **JULY GALINDO QUINTERO**, abogados registrados quienes trabajan para la firma **CÁRDENAS & CÁRDENAS - ABOGADOS,** con sede en la Carrera 7 No. 71-52, Torre B, Piso 9, en Bogotá, D.C., Colombia, para que gestionen en Colombia ante las Oficinas y Autoridades correspondientes la obtención de registros de marca comercial, patentes, modelos de utilidad, diseños y dibujos industriales, depósitos de nombres y enseñas comerciales, como también para el procesamiento de renovaciones, cesiones, cambios de nombre o

MINISTERIO DE
RELACIONES EXTERIORES

2010 ENE -6 P 9 03

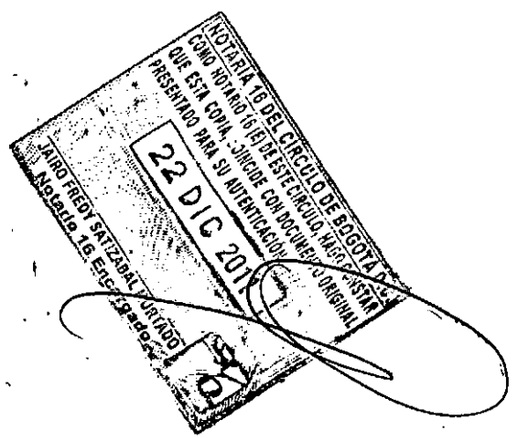
JEFE DE LEGALIZACIONES
BOGOTA D.C.

NO SE ASUME
RESPONSABILIDAD
DEL TEXTO

042633

Martha Sarmiento

COPIA FIEL QUE APARECE EN EL
DOCUMENTO DESEMPEÑA
LAS FUNCIONES INDICADAS



12

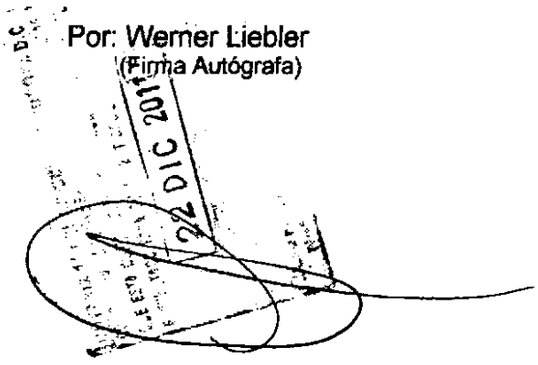
Carrera 11D # 124 - 61 Apto. 510 Int. 2 - Teléfono 2139034 Bogotá - Colombia
E-mail: mssarmiento2@yahoo.com

Traductora Oficial Juramentada
Resolución No. 835 Minjusticia

domicilio del titular, renuncias y licencias de lo arriba mencionado, propósitos para los cuales se les otorga Poder para que lleven a cabo ante dichas autoridades todas las acciones necesarias para los propósitos indicados, como radicar solicitudes, hacer declaraciones, reclamaciones, formular descripciones, realizar enmiendas o modificaciones, interponer recursos de oposición y apelaciones; pagar todos los impuestos y cuotas y hacer cualquier otro pago establecido por ley; recibir todos los documentos y dineros y emitir el descargo correspondiente; cumplir con cualquier otro requerimiento y tomar, en general, todas las medidas que consideren apropiadas y oportunas para defender y proteger la Propiedad Industrial de la Compañía en el evento de infracciones, oposiciones, procesos para cancelación o anulación, imposición de sanciones, presentar antecedentes ante las Cortes y Tribunales. Además, quedan facultados para comparecer como demandantes o demandados ante Jueces competentes, también se les da Poder para negociar, solicitar arbitraje, ceder, recibir, dimitir, rescindir o desistir, apelar, ratificar y todos los poderes legales adicionales que sean necesarios. También se declara que a partir de ahora todas las acciones que realicen dichos abogados respaldando este Poder de Representación y para beneficio de la Compañía son buenas y válidas. Adicionalmente se les otorga facultades para sustituir este Poder y si es necesario revocar dichas sustituciones.

FIRMADO EL: **6 de diciembre de 2010** EN: **Munich, Alemania**

Por: Werner Liebler
(Firma Autógrafa)



Carrera 11D # 124 - 51 Apto. 510 Int. 2 - Teléfono 2139034 Bogotá - Colombia
E-mail: mssarmiento2@yahoo.com

Archivo No. H2294/2010
ag

Yo, por la presente certifico que la anterior firma es verdadera, que fue suscrita en mi presencia por

El Sr. Dr. Werner Liebler, Nacido el 6 de mayo de 1961,
Dirección comercial: Hansastr. 27c, 80686 Munich,
a quien conozco personalmente.

El **Dr. Werner Liebler** actuando en representación de **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.**, con sus oficinas registradas en Munich,
Dirección comercial: 80686 München, Hansastr. 27c,
Hoy presentó ante mí en original los siguientes documentos: Poder de Abogado fechado el 13 de julio del 2000, Archivo No. 3076 K/2000, Poder de Abogado fechado el 18 de julio del 2000, Archivo No. 3170 K/2000, documentos expedidos y firmados por el Dr. Helmut Keidel, Notario Público de Munich, y como documento adjunto la Copia Certificada.

Munich, 6 de diciembre de 2010

Firma Autógrafa
Thomas Haasen
Notario Público, Munich

Sello del Notario:
THOMAS HAASEN
NOTARIO EN MUNICH

H:\Grunert\Unterschriftsbeglaubigungen\Dr. Liebler (Fraunhofer) engl 10-12-06.doc

CERTIFICO QUE ESTA ES UNA TRADUCCIÓN FIEL DEL INGLÉS AL ESPAÑOL DE LOS DOCUMENTOS QUE HE TENIDO A LA VISTA; QUE SOY TRADUCTORA E INTERPRETE OFICIAL ACREDITADA POR RESOLUCIÓN Nº 835 DEL 3 DE MAYO DE 1988 DEL MINISTERIO DE JUSTICIA Y QUE ESTOY REGISTRADA ANTE EL MINISTERIO DE RELACIONES EXTERIORES DE COLOMBIA.

Bogotá, D.C. 2 de enero de 2011

Martha S. Sarmiento S.
Martha S. Sarmiento S.
Traductora Oficial Juramentada
Resolución No. 835 Minjusticia

NOTARÍA 16 DEL CÍRCULO DE BOGOTÁ D.C.
COTEJADA(O) DE ESTE CÍRCULO, ALGO CONSTA
QUE ESTA COPIA, JUNTO CON DOCUMENTO ORIGINAL
PRESENTADO PARA SU AUTENTIFICACIÓN.
22 DIC 2011

FhG – Fraunhofer Gesellschaft

Fraunhofer Gesellschaft Postfach 190339 D-80603 München

Junta Directiva Leonrodstrasse 54, D-80636 München

Tel.: +49 (0) 89 / 12 05 - 01 Telefax: +49 (0) 89 / 12 – 05 - 700

PODER

Por este medio, damos poder al

Señor Doctor Werner Liebler, nacido el 6 de mayo de 1961, con residencia comercial en Patentstelle für die Deutsche Forschung, Leonrodstrasse 68, 80636 München

para representar, hasta nuevo aviso, con carácter obligatorio, a la firma

Fraunhofer-Gesellschaft zur Förderung der angewandten

Forschung e.V., Leonrodstrasse 54, 80636 München

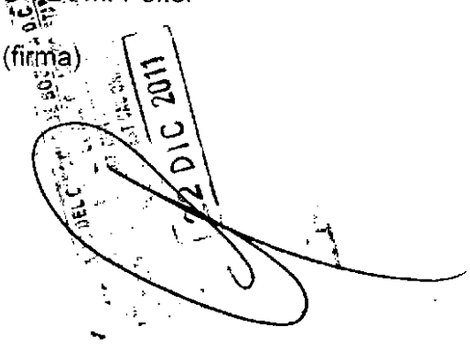
en todos los asuntos de ley sobre invenciones de los asalariados. Además, el señor Dr. Liebler está autorizado a delegar abogados de patentes locales en el caso de procesos legales de protección.

Dr. D. M. Polter

(firma)

Dr. H.-U. Wiese

(firma)



Marco R. Fajardo V.
Traductor Oficial Juramentado
Resol. Min. Justicia No. 839/86

URNo. 307 CK

Por este medio, autentico la veracidad de la firma anterior, trazada ante mí por el Señor Dr. Dirk-Meints P o l t e r, miembro de la Junta directiva, en Munich, con residencia comercial en Leonrodstrasse 68, 80636 München, conocido en persona.

En ese sentido, certifico, con base en la inspección ocular en el registro de asociaciones del Tribunal municipal de Munich (No. VR 4461), del 12 de julio de 2000, que los señores Dr. Dirk-Meints Polter y Dr. Hans-Ulrich Wiese, como miembros de la Junta directiva, están autorizados, en conjunto, para la representación de la asociación

Fraunhofer-Gesellschaft zur Förderung der

Angewandten Forschung asociación registrada,

con domicilio en Munich.

Munich, 13 de julio de 2000

Dr. Keidel, Notario

(firma y sello)

NOTARÍA 16 DEL CÍRCULO DE BOGOTÁ D.C.
COMO NOTARIO 16 (E) DE ESTE CÍRCULO, HAGO CONSTAR
QUE ESTA COPIA COINCIDE CON DOCUMENTO ORIGINAL
PRESENTADO PARA SU AUTENTICACIÓN.
22 DIC 2011
JAIRO FREDY SARIZA
Notario 16

Margo B. Fajardo V.
Traductor Oficial Juramentado
Resol. Min. Justicia No. 839/86

URNo. 3070 K

Por este medio, autentico la veracidad de la firma anterior, trazada ante mí por el

Señor Dr. Hans-Ulrich W i e s e,

miembro de la Junta directiva, en Munich,

con residencia comercial en Leonrodstrasse 68, 80636 München,

identificado con documento oficial

conocido en persona.

En ese sentido, certifico, con base en la inspección ocular en el registro de asociaciones del Tribunal municipal de Munich (No. VR 4461), del 12 de julio de 2000, que los señores Dr. Dirk-Meints Polter y Dr. Hans-Ulrich Wiese, como miembros de la Junta directiva, están autorizados, en conjunto, para la representación de la asociación

Fraunhofer-Gesellschaft zur Förderung der

Angewandten Forschung asociación registrada,

con domicilio en Munich.

Munich, 18 de julio de 2000

Dr. Keidel, Notario

(firma y sello)

Por este emdio certifico que el presente documento es una copia auténtica del original.

Munich, 6 de diciembre de 2010

El Notario, en Muncih

(firma, sello y cordón)

Marco B. Tijerdo V.
Traductor Oficial Juramentado
Resol. Min. Justicia No. 839/86

APOSTILLE

(Convention de La Haye du 5 octobre 1961)

1. País: **República Federal de Alemania**

   el presente documento público

2. está firmado por Thomas Haasen

3. en su calidad de Notario

4. el documento está provisto del sello del Notario Thomas Haasen, en Munich

confirmado

5. en Muncih

6. el 7 de diciembre de 2010

7. por el Presidente del Tribunal regional de Munich I

8. bajo el No. 910 a - 9279 / 2010

9. Sello:

10. Firma

en representación

Christine Köppl

Inspectora oficial

(firma, cordón y sello)

NOTARÍA 16 DEL CÍRCULO DE BOGOTÁ D.C.
NOTARIO 16 (E) DE ESTE CÍRCULO, HAGO CONSTAR
QUE ESTA COPIA, COINCIDE CON DOCUMENTO ORIGINAL
PRESENTADO PARA SU AUTENTICACIÓN.
22 DIC 2010

MINISTERIO DE RELACIONES EXTERIORES
OFICINA DE LEGALIZACIONES
BOGOTÁ D.C.
2010 FEB -4 A 8:19

Marco B. Tejada V.
Traductor Oficial Juramentado
Resol. Min. Justicia No. 839/86

Amtsgericht München -Registergericht-

VR 4461

Amtlicher aktueller Ausdruck aus dem Registerblatt

Datum der letzten Eintragung: 03.08.2010

Datum des Abrufs: 19.10.2010

Ort und Tag der Ausstellung: München, den 19. Oktober 2010

Der Ausdruck bezeugt den Inhalt des Vereinsregisters.

Ersteller: Werdich, Justizangestellter,
Urkundsbeamter/Urkundsbeamtin der Geschäftsstelle

Dieser Ausdruck wird nicht unterschrieben und gilt als beglaubigte Abschrift.





16

1. **Anzahl der bisherigen Eintragungen:**

16

2. **a) Name:**

Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung eingetragener Verein

b) Sitz:

München

3. **a) Allgemeine Vertretungsregelung:**

Jeweils zwei Vorstandsmitglieder vertreten den Verein gemeinsam.

b) Vertretungsberechtigte und besondere Vertretungsbefugnis:

Vorstand: Prof. Dr. Buller, Ulrich, Golm, *01.06.1946
Vorstand: Prof.Dr.-Ing. Bullinger, Hans-Jörg, Stuttgart-Botnang, *13.04.1944
Vorstand: Prof. (Univ. Stellenbosch) Dr. Gossner, Alfred, Erharting, *15.10.1950

Vertretungsberechtigt in Gemeinschaft mit einem Vorstandsmitglied:
Besonderer Vertreter gemäß § 30 BGB für den Bereich Finanzen und Rechnungswesen: Meuer, Andreas, München, *12.06.1961



4. **a) Satzung:**

Eingetragener Verein

Satzung vom 26.03.1949
Zuletzt geändert durch Beschluss vom 20.05.2010

b) Sonstige Rechtsverhältnisse:

Der Verein hat im Wege der Abspaltung gemäß Spaltungs- und Übernahmevertrag vom 23.06.2009 sowie Beschluß seiner Mitgliederversammlung vom 24.06.2009 und Beschluß der Mitgliederversammlung des übertragenden Vereins vom 08.05.2009 Teile des Vermögens vom Forschungsinstitut für Pigmente und Lacke mit dem Sitz in Stuttgart (Amtsgericht Stuttgart VR 537) übernommen.

Die Forschungsgesellschaft für angewandte Naturwissenschaften e.V. mit dem Sitz in Bonn (Amtsgericht Bonn VR 2530) ist auf Grund des Verschmelzungsvertrages vom 23.06.2009 sowie des Beschlusses der Mitgliederversammmlung vom 24.06.2009 und des Beschlusses der Mitgliederversammlung des übertragenden Vereins vom 04.06.2009 mit dem Verein verschmolzen.
Das Institut für Solare Energieversorgungstechnik Verein an der Universität Kassel mit dem Sitz in Kassel (Amtsgericht Kassel VR 2029) ist auf Grund des Angebots auf Abschluss eines Verschmelzungsvertrages vom 07.05.2009 und der Annahme des Angebots vom 23.06.2009 sowie des Beschlusses der Mitgliederversammmlung vom 24.06.2009 und des Beschlusses der Mitgliederversammlung des übertragenden Vereins vom 07.05.2009 mit dem Verein verschmolzen.

5. **a) Tag der letzten Eintragung:**

03.08.2010

Die Übereinstimmung dieses Ausdrucks mit dem Handelsregister wird beglaubigt.

München, den 19. Oktober 2010

Amtsgericht München -Registergericht-

Jörg Werdich, Justizangestellter
als Urkundsbeamter der Geschäftsstelle

APOSTILLE
(Convention de La Haye du 5 octobre 1961)

1. Land: Bundesrepublik Deutschland
   Diese öffentliche Urkunde
2. ist unterschrieben von
   Herrn Justizangestellter Jörg Werdich
3. in der Eigenschaft als
   Urkundsbeamter der Geschäftsstelle
4. sie ist versehen mit dem Siegel des
   Amtsgerichts München

Bestätigt

5. in München 6. am 19. Oktober 2010
7. durch den Richter am Amtsgericht
   Dr. Ulrich Kühn als weiterem Vertreter des Präsidenten des Amtsgerichts München
8. unter Nr. GenA 9101 a Nr. 636/10
9. Stempel 10. Unterschrift

Dr. Ulrich Kühn

Tribunal municipal de Munich – Tribunal de registro

VR 4461

Extracto actual oficial de la hoja de registro

Fecha de la última anotación: 3 de agosto de 2010

Fecha de la petición de entrega: 19 de agosto de 2010

Lugar y fecha de expedición: Munich, 19 de octubre de 2010

El extracto confirma el contenido del registro de la asociación

Proveedor: Werdich, funcionario judicial

Fedatario/funcionaria fedataria del despacho

El presente extracto no va firmado y rige como copia autenticada

(sello y cordón)



Hugo B. Fajardo V.
Traductor Oficial Juramentado
Resol. Min. Justicia No. 839/86

Registro de asociación del Tribunal municipal de Munich	Reproducción del contenido actual del Registro Petición del 19 de octubre de 2010 8.33 a.m.	Número de la asociación: **VR 4461**
- Extracto -	Página 1 de 1	

1. **Número de las anotaciones hechas hasta ahora:**

   16

2. **a) Nombre:**

   Fraunhofer-Gesellschaft zur Förderung de angewandten Forschung

   eingetragener Verein

   b) Domicilio

   Munich

3. **a) Regulación general sobre representación:**

   Cada vez dos miembros de la junta directiva representarán la asociación juntos

   b) Autorizados a representar y permisos especiales de representación

   Junta directiva: Prof. Dr. Buller, Ulrich, de Golm, nacido el 1 de junio de 1946

   Junta directiva: Prof. Dr.-Ing. Bullinger, Hans-Jörg, de Stuttgart-Botnang, nacido el 13 de abril de 1944

Marco B. Fajardo V.
Traductor Oficial Juramentado
Res. Min. Justicia No. 839/86

Junta directiva: Prof. (Universidad de Stellenbosch) Dr. Gossner, Alfred, de Erharting, nacido el 15 de octubre de 150

Autorizado a representar junto con un miembro de la junta directiva:

Representante especial, según parágrafo 30 del código civil para el área de las finanzas y la contabilidad: Meuer, Andreas, de Munich, nacido el 12 de junio de 1961

4. **a) Estatutos:**

Sociedad registrada

Estatutos del 26 de marzo de 1949

Última modificación según acuerdo del 20 de mayo de 2010

b) Otras relaciones jurídicas:

En el transcurso de la desmembración de la asociación cesionista, según contrato de desmembración y recepción, del 23 de junio de 2009, así como del acuerdo de su asamblea de socios, del 24 de junio de 2009, y del acuerdo de la asamblea de socios de la asociación cesionista, del 8 de mayo de 2009, la asociación tomó parte del patrimonio del Instituto para la investigación de pigmentos y lacas (Forschungsinstitut für Pigmente und Lacke), con domicilio en Stuttgart (Tribunal municipal de Stuttgart, referencia VR 537).

La Forschungsgesellschaft für angewandte Naturwissenschaften e.V. (Sociedad de investigación para ciencias naturales aplicadas, asociación registrada), con

NOTARIA 16 DEL CIRCULO DE BOGOTÁ D.C.
COPIA QUE HAGO CONSTAR
COMO NOTARIO, HAGO CONSTAR
QUE ESTA COPIA ES FIEL DE ESTE ORIGINAL
QUE CONCUERDA CON DOCUMENTO
QUE ESTA COPIA SE EXPIDIO PARA SU AUTENTICACION
FECHA DE EXPEDICION PARA SU AUTENTICACION
2 DIC 2011

Marco B. Tejada V.
Traductor Oficial Juramentado
Res. Min. Justicia No. 839/86

domicilio en Bonn (Tribunal municipal de Bonn VR 2530) se fusionó con la asociación, con base en el contrato de fusión, del 23 de junio de 2009, así como del acuerdo de la asamblea de socios , del 24 de junio de 2009 y del acuerdo de la asamblea de socios de la asociación cesionista, del 4 de junio de 2009.

El Instituto para técnicas de aprovisionamiento de energía solar (Institut für Solare Energieversorgungstechnik) asociación de la Universidad de Kassel, con domicilio en Kassel (Tribunal municipal de Kassel VR 2029), se fusionó con la asociación, con base en la oferta de firma de un contrato de fusión, del 7 de mayo de 2009, y de la aceptación de tal oferta, del 23 de junio de 2009, así como del acuerdo de la asamblea de socios, del 24 de junio de 2009 y del acuerdo de la asamblea de socios de la sociedad cesionista, del 7 de mayo de 2009.

5. **a) Fecha de la última anotación:**

3 de agosto de 2010

Se autentica la coincidencia del extracto con el registro comercial.

Munich, 19 de octubre de 2010

Tribunal municipal de Munich – Tribunal de registro

Jörg Werdich, funcionario judicial

como funcionario fedatario del despacho

(firma, cordón y sello)

Marco B. Fajardo V.
Traductor Oficial Juramentado
Resol. Min. Justicia No. 839/86

APOSTILLE

(Convention de La Haye du 5 octobre 1961)

1. País: **República Federal de Alemania**

   el presente documento público

2. está firmado por el funcionario judicial Jörg Werdich

3. en su calidad de funcionario fedatario del despacho

4. el documento está provisto del sello del Tribunal municipal de Munich

confirmado

5. en Munich 6. 19 de octubre de 2010

7. por el Juez del tribunal municipal, Dr. Ulrich Kühn, como uno de los representantes del Presidente del tribunal municipal de Munich

8. bajo el No. GenA 9101 a No. 636/10

9. Sello: 10. Firma

Dr. Ulrich Kühn

(firma y sello)

MINISTERIO DE RELACIONES EXTERIORES
2010 NOV 18 P 4:50
JEFE DE LEGALIZACIONES
BOGOTA D.C.

NOTARIA 16 DEL CIRCULO DE BOGOTA D.C.
COMO NOTARIO 16 (E) DE ESTE CIRCULO, HAGO CONSTAR
QUE ESTA COPIA, CONCUERDA CON DOCUMENTO ORIGINAL
PRESENTADO PARA SU AUTENTICACION.
22 DIC 2011

MINISTERIO DE RELACIONES EXTERIORES
RECIBO LEGALIZACION

038276
COPIA FIRMA APARECE EN EL
DOCUMENTO DESEMPEÑA
ASFUNCIONES INDICADAS

NO SE ASUME
RESPONSABILIDAD
DEL TEXTO

Mayor B. Fajardo V.
Traductor Oficial Juramentado
Res. Min. Justicia No. 839/86

DECODIFICADOR DE SEÑALES DE AUDIO, MÉTODO PARA DECODIFICAR UNA SEÑAL DE AUDIO Y PROGRAMA DE COMPUTACIÓN QUE UTILIZA ETAPAS EN CASCADA DE PROCESAMIENTO DE OBJETOS DE AUDIO

Descripción

Campo Técnico

Las realizaciones de acuerdo con la invención se relacionan con un decodificador de señales de audio para producir una representación de señal de mezcla ascendente que depende de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto. Otras realizaciones de acuerdo con la invención se relacionan con un método para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto. Otras realizaciones de acuerdo con la invención se relacionan con un programa de computación.

Algunas realizaciones de acuerdo con la invención se relacionan con un sistema SAOC de Karaoke/Solos mejorado.

Antecedentes de la Invención

En las disposiciones de audio modernas, es conveniente transferir y almacenar información de audio de manera eficiente con respecto a la velocidad de transmisión de bits. Además, con frecuencia es conveniente reproducir un contenido de audio empleando una pluralidad de dos o incluso más parlantes, que están espacialmente distribuidos en un recinto. En esos casos, es conveniente aprovechar las capacidades de dicha disposición de múltiples parlantes para permitir que un usuario identifique espacialmente diferentes contenidos de audio o diferentes elementos de un único contenido de audio. Esto se puede lograr distribuyendo individualmente los diferentes contenidos de audio en los diferentes parlantes.

222.108
MB

Dicho de otro modo, en la técnica del procesamiento de audio, la transmisión de audio y el almacenamiento de audio, existe un creciente interés por manejar contenidos de canales múltiples para mejorar la impresión auditiva. El uso de contenido de audio de múltiples canales trae aparejadas significativas mejoras para el usuario. Por ejemplo, se puede obtener una impresión auditiva tridimensional, lo que trae consigo una mayor satisfacción del usuario en

aplicaciones de entretenimiento. Sin embargo, los contenidos de audio en múltiples canales también son de utilidad en entornos profesionales, por ejemplo en aplicaciones de conferencias telefónicas, puesto que se puede mejorar la inteligibilidad de los parlantes utilizando una reproducción de audio en canales múltiples.

Sin embargo, también es conveniente contar con un buen equilibrio entre buena calidad de audio y requerimientos de velocidad de transferencia de bits, a fin de evitar una excesiva carga de recursos causada por aplicaciones de canales múltiples.

Últimamente se han propuesto técnicas paramétricas para la transmisión eficiente de bits y/o para el almacenamiento de escenas de audio que contienen múltiples objetos de audio, por ejemplo, Binaural Cue Coding (Type I) (Ver, por ejemplo la referencia [BCC]), Joint Source Coding (ver, por ejemplo, la referencia [JSCJ), y MPEG Spatial Audio Object Coding (SAOC) (ver, por ejemplo, las referencias [SAOCl], [SAOC2]).

Estas técnicas apuntan a reconstruir perceptualmente la escena de salida de audio ventajosa en lugar de hacerlo por coincidencia de formas de onda.

La Fig. 8 ilustra una reseña general de la disposición correspondiente a ese sistema (en este caso: MPEG SAOC). El sistema MPEG SAOC 800 ilustrado en la Fig. 8 comprende un codificador SAOC 810 y un decodificador SAOC 820.

El codificador SAOC 810 recibe una pluralidad de señales de objeto X_1 a X_N, que pueden estar representadas, por ejemplo, en forma de señales en el dominio del tiempo o en forma señales en el dominio del tiempo y la frecuencia (por ejemplo, en forma de una serie de coeficientes de transformación del tipo de transformadas de Fourier, o en forma de señales de subbanda QMF). El codificador SAOC 810 también recibe, por lo general, coeficientes de mezcla descendente d_1 a d_N, que están asociados con las señales de objeto x_1 a x_N. Se puede disponer de series independientes de de coeficientes de mezcla descendente por cada canal de de la señal de mezcla descendente. El codificador SAOC 810 está configurado típicamente para obtener un canal de la señal de mezcla descendente combinando las señales de objeto X_1 a X_N de acuerdo con los coeficientes de mezcla descendente asociados d_1 a d_N. Por lo general, hay menos canales de mezcla descendente que señales de objeto X_1 a X_N. Para dar

lugar (por lo menos aproximadamente) a una separación (o tratamiento separado) de las señales de objeto del lado del decodificador SAOC 820, el codificador SAOC 8 10 suministra tanto la una o más señales de mezcla descendente (designados canales de mezcla descendente) 812 como una información complementaria 814. La información complementaria 814 describe características de las señales de objeto X_1 a X_N, para dar lugar a un procesamiento específico de objetos del lado del decodificador.

El decodificador SAOC 820 está configurado para recibir tanto dicha una o más señales de mezcla descendente 812 como la información complementaria 814. Además, el decodificador SAOC 820 está configurado por lo general para recibir una información sobre interacción con el usuario y/o una información de interacción con el usuario y/o una información de control del usuario 822, que describe una configuración de renderización. Por ejemplo, la información de interacción con el usuario/información de control del usuario 822 puede describir una configuración de parlantes y la colocación espacial pretendida de los objetos provistos por las señales de objeto X_1 a X_N, El decodificador SAOC 820 está configurado para producir, por ejemplo, una pluralidad de señales decodificadas de canales de mezcla ascendente El decodificador SAOC 820 está configurado para producir, por ejemplo, una pluralidad de señales de canales decodificadas de mezcla descendente $\hat{y}_1$ a $\hat{y}_M$. Las señales de canales de mezcla ascendente pueden estar asociadas, por ejemplo a parlantes individuales de una disposición de renderización de múltiples parlantes. El decodificador SAOC 820 puede comprender, por ejemplo, un separador de objetos 820a, que está configurado para reconstruir, por lo menos aproximadamente, las señales de objeto x_1 a x_N sobre la base de dichas una o más señales de mezcla descendente 812 e información complementaria 814, a fin de obtener así las señales de objetos reconstruidas 820b. Sin embargo, las señales de objetos reconstruidas 820b se pueden desviar hasta cierto punto de las señales de objetos originales x_1 a x_N, por ejemplo, porque la información complementaria 814 no basta totalmente para una reconstrucción perfecta debido a restricciones de velocidad de transmisión de bits. El decodificador SAOC 820 puede comprender además un mezclador 820c, que puede estar configurado para recibir las señales de objetos reconstruidas 820b y la información de interacción con el usuario/información de control de usuarios

822, y para suministrar, sobre la base de éstas, las señales de canales de mezcla ascendente $\hat{y}_1$ a $\hat{y}_M$. El mezclador 820 puede estar configurado para usar la información de interacción con el usuario /información de control de usuarios 822 para determinar la contribución de las señales de objetos reconstruidas 820b individuales a las señales de canales de mezcla ascendente $\hat{y}_1$ a $\hat{y}_M$. La información de interacción con el usuario/información de control de usuarios 822 puede comprender, por ejemplo, parámetros de renderización (que también se denominan coeficientes de renderización), que determinan la contribución de las señales de objetos reconstruidas 822 individuales a las señales de canales de mezcla ascendente $\hat{y}_1$ a $\hat{y}_M$.

Sin embargo, cabe señalar que, en muchas realizaciones, la separación de los objetos, que está indicada por el separador de objetos 820a de la Fig. 8, y la mezcla, que está indicada por el mezclador 820c de la Fig. 8, se llevan a cabo en un solo paso. Para este fin, se pueden computar los parámetros generales que describen un mapeo directo de dichas una o más señales de mezcla descendente 812 sobre las señales de canales de mezcla ascendente $\hat{y}_1$ a $\hat{y}_M$. Estos parámetros se pueden computar sobre la base de la información complementaria y la información de interacción con el usuario/ información de control de usuarios 822.

Tomando, ahora, como referencia las Figs. 9a, 9b y 9c, se describen diferentes aparatos para obtener una representación de la señal de mezcla ascendente sobre la base de una representación de señal de mezcla descendente e información complementaria relacionada con objetos. La Fig. 9a ilustra un diagrama esquemático de bloques de un sistema MPEG SAOC 900 que comprende un decodificador SAOC 920. El decodificador SAOC 920 comprende, como bloques funcionales separados, un decodificador de objetos 922 y un mezclador/renderizador 926. El decodificador de objetos 922 produce una pluralidad de señales de objetos reconstruidas 924 que depende de la representación de la señal de mezcla descendente (por ejemplo, en forma de una o más señales de mezcla descendente representadas en el dominio del tiempo o en el dominio del tiempo–frecuencia) e información complementaria relacionada con objetos (por ejemplo, en forma de metadatos de objeto). El mezclador/renderizador 926 recibe las señales de objetos reconstruidas 924

asociadas a una pluralidad de objetos N y produce, sobre la base de éstas, una o más señales de canales de mezcla ascendente 928. En el decodificador SAOC 920, la extracción de las señales de objeto 924 se realiza en forma independiente de la mezcla/renderización, lo que da lugar a una separación de la funcionalidad de decodificación de objetos de la funcionalidad de mezcla/renderización, aunque trae aparejada una complejidad informática relativamente elevada.

Haciendo referencia, ahora, a la Fig. 9b, se describe brevemente otro sistema MPEG SAOC 930, que comprende un decodificador SAOC 950. El decodificador SAOC 950 produce una pluralidad de señales de canales de mezcla ascendente 958 que depende de una representación de señal de mezcla descendente (por ejemplo, en forma de una o más señales de mezcla descendente) e información complementaria relacionada con un objeto (por ejemplo, en forma de metadatos de objeto). El decodificador SAOC 950 comprende una combinación de decodificador de objeto y mezclador/renderizador, que está configurado para obtener las señales de canales de mezcla ascendente 958 en un proceso de mezclado conjunto sin separación de la decodificación de objetos y la mezcla/renderización, donde los parámetros para dicho proceso conjunto de mezcla ascendente dependen tanto de la información complementaria relacionada con objetos y la información sobre la renderización. El proceso conjunto de mezcla ascendente depende asimismo de la información sobre mezcla descendente, que se considera parte de la información complementaria relacionada con objetos.

Para resumir lo anterior, la provisión de las señales de canales de mezcla ascendente 928, 958 se puede realizar en un proceso de un solo paso o un proceso de dos pasos.

Haciendo referencia, ahora, a la Fig. 9c, se describe un sistema MPEG SAOC 960. El sistema SAOC 960 comprende un transcodificador de SAOC a MPEG Envolvente 980, en lugar de un decodificador SAOC.

El transcodificador de SAOC a MPEG Envolvente comprende un transcodificador de información complementaria 982, que está configurado para recibir la información complementaria relacionada con objetos (por ejemplo, en forma de metadatos de objeto) u, opcionalmente, información de dicha una o más señales de mezcla descendente y la información sobre la renderización. El

transcodificador de información complementaria también está configurado para proporcionar una información complementaria sobre MPEG Envolvente 984 (por ejemplo, en forma de flujo de bits de MPEG Envolvente) sobre la base de ciertos datos recibidos. En consecuencia, el transcodificador de información complementaria 982 está configurado para transformar una información complementaria relacionada con objetos (paramétrica), que es emitida por un codificador de objetos, en una información complementaria (paramétrica) relacionada con los canales 984, tomando en cuenta la información sobre la renderización y, opcionalmente, la información sobre el contenido de dichas una o más señales de mezcla descendente.

Opcionalmente, el transcodificador de SAOC a MPEG Envolvente 980 puede estar configurado para manipular dichas una o más señales de mezcla descendente, descriptas, por ejemplo, por la representación de la señal de mezcla descendente, para obtener una representación de señal de mezcla descendente manipulada 988. Sin embargo, se puede omitir el manipulador de señal de mezcla descendente 986, de manera que la representación de señal de mezcla descendente 988 de salida del transcodificador de SAOC a MPEG Envolvente 980 sea idéntica a la representación de señal de mezcla descendente de entrada del transcodificador de SAOC a MPEG Envolvente. El manipulador de señales de mezcla descendente 986 se puede utilizar, por ejemplo, en caso de que la información complementaria de MPEG Envolvente relacionada con los canales 984 no permita la producción de una impresión auditiva adecuada sobre la base de la representación de señal de mezcla descendente de entrada del transcodificador de SAOC a MPEG Envolvente 980, lo que puede ocurrir en algunas constelaciones de renderización.

En consecuencia, el transcodificador de SAOC a MPEG Envolvente 980 da origen a la representación de la señal de mezcla descendente 988 y el flujo de bits de MPEG Envolvente 984 razón por la cual se puede generar una pluralidad de señales de canales de mezcla ascendente, que representan los objetos de audio de acuerdo con la información sobre la renderización ingresada en el transcodificador de SAOC a MPEG Envolvente 980 utilizando un decodificador de MPEG Envolvente que recibe el flujo de bits de MPEG Envolvente 984 y la representación de la señal de mezcla descendente 988.

Para resumir lo anterior, se pueden emplear diferentes conceptos para decodificar señales de audio codificadas por SAOC. En algunos casos, se utiliza un decodificador SAOC, que produce señales de canales de mezcla ascendente (por ejemplo, las señales de canales de mezcla ascendente 928, 958) que dependen de la representación de la señal de mezcla descendente y la información complementaria paramétrica relacionada con los objetos. Se pueden ver ejemplos de este concepto en las Figs. 9a y 9b. Por otro lado, se puede transcodificar la información de audio codificada por SAOC para obtener una representación de señal de mezcla descendente (por ejemplo, una representación de señal de mezcla descendente 988) y una información complementaria relacionada con los canales (por ejemplo, el flujo de bits de MPEG Envolvente relacionado con los canales 984), que puede ser utilizada por un decodificador de MPEG Envolvente para producir las pretendidas señales de canales de mezcla ascendente.

En el sistema MPEG SAOC 800, una reseña general de la cual se presenta en la Fig. 8, se lleva a cabo el procesamiento general en forma selectiva de la frecuencia y se puede describir de la siguiente manera dentro de cada banda de frecuencias:

– Se realiza la mezcla descendente de N señales de audio de entrada de objeto x_1 a x_N como parte del procesamiento del codificador SAOC. Para una mezcla descendente mono, los coeficientes están indicados por d_1 a d_N. Además, el codificador SAOC 810 extrae información complementaria 814 que describe las características de los objetos de audio de entrada. En el caso de la MPEG SAOC, las relaciones de las potencias de objeto entre sí son la forma más básica de dicha información complementaria.

– La señal (o señales) de mezcla descendente 812 y la información complementaria 814 son transmitidas y/o almacenadas. Para este fin, se puede comprimir la señal de audio de mezcla descendente empleando codificadores de audio perceptuales muy conocidos tales como MPEG–1 Capa II o III (también conocido como ".mp3"), Codificación de Audio Avanzada MPEG (AAC), o cualquier otro codificador de audio

– Del lado del receptor, el decodificador SAOC 820 intenta conceptualmente restablecer la señal de objeto original ("separación de objetos") utilizando la

información complementaria transmitida 814 (y, naturalmente, la una o más señales de mezcla descendente 812). Estas señales de objeto aproximadas (que también se denominan señales de objetos reconstruidas 820b) se mezclan a continuación en una escena objetivo representada por M canales de salida de audio (que pueden estar representados, por ejemplo, por las señales de canales de mezcla ascendente $\hat{y}_1$ a $\hat{y}_M$) utilizando una matriz de renderización. En el caso de una salida mono, los coeficientes de la matriz de renderización están expresados por r_1 a r_N.

– En efecto, raramente se ejecuta la separación de las señales de objeto (o incluso nunca se las ejecuta), puesto que tanto el paso de separación (indicado por el separador de objetos 820a) como el paso de mezcla (indicado por el mezclador 820c) se combinan para obtener un solo paso de transcodificación, que con frecuencia da lugar a una enorme reducción de la complejidad informática.

Se ha encontrado que ese tipo de esquema es tremendamente eficiente, tanto en términos de velocidad de transmisión de bits (sólo es necesario transmitir unos pocos canales de mezcla descendente más cierta información complementaria en lugar de N señales discretas de audio de objeto o un sistema discreto) y complejidad informática (la complejidad de procesamiento se relaciona principalmente con el número de canales de salida en lugar del número de objetos de audio). Otras ventajas para el usuario del lado de la recepción incluyen la libertad de elegir una configuración de renderización de su elección (mono, estéreo, envolvente, reproducción virtualizada con auriculares y demás) y la característica de interactividad con el usuario: se puede ajustar la matriz de renderización y, por consiguiente, la escena de salida, puede ser ajustada y cambiada interactivamente por el usuario según su voluntad, preferencias personales u otros criterios. Por ejemplo, es posible localizar los interlocutores de un grupo juntos en un área espacial para maximizar la discriminación de las demás personas que conversan. Esta interactividad se obtiene produciendo una interfaz de usuario del decodificador.

Por cada objeto de sonido transmitido, se puede ajustar su nivel relativo y (en el caso de la renderización no mono) la posición espacial de la renderización. Esto puede ocurrir en tiempo real al cambiar el usuario la posición de los

dispositivos deslizantes de interfaz gráfica asociada (GUI) (por ejemplo: nivel de objeto = +5dB, posición de objeto = –30grados).

Sin embargo, se ha encontrado que es difícil manejar objetos de audio de diferentes tipos de objetos de audio en ese tipo de sistema. En particular, se ha encontrado que es difícil procesar objetos de audio de diferentes tipos de objetos de audio, por ejemplo, objetos de audio con los cuales se asocia diferente información complementaria, si el número total de objetos de audio a procesar no ha sido predeterminado.

En vista de esta situación, el objetivo de la presente invención consiste en crear un concepto que da lugar a la decodificación informáticamente eficiente y flexible de una señal de audio que comprende una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto, donde la información paramétrica relacionada con los objetos describe objetos de audio de dos o más tipos de objetos de audio diferentes.

Síntesis de la Invención

Este objetivo se logra mediante un decodificador de señales de audio para producir una representación de señal de mezcla ascendente que depende de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto, un método para producir una representación de señal de mezcla ascendente que depende de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto y un programa de computación, de acuerdo con lo definido por las reivindicaciones independientes.

Una realización de acuerdo con la invención crea un decodificador de señales de audio para producir una representación de señal de mezcla ascendente que depende de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto. El decodificador de señales de audio comprende un separador de objetos configurado para descomponer representación de señal de mezcla descendente, para producir una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio que depende de la representación de la señal de mezcla

descendente y usar por lo menos parte de la información paramétrica relacionada con los objetos. El decodificador de señales de audio comprende además comprende un procesador de señales de audio configurado para recibir la segunda información de audio y para procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, para obtener una versión procesada de la segunda información de audio. El decodificador de señales de audio comprende asimismo un combinador de señales de audio configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio para obtener la representación de señal de mezcla ascendente.

Es una idea clave de la presente invención que se puede obtener un procesamiento eficiente de diferentes tipos de objetos de audio en una estructura de etapas en cascada, que permite la separación de los diferentes tipos de objetos de audio utilizando por lo menos parte de la información paramétrica relacionada con los objetos en un primer paso de procesamiento ejecutado por el separador de objetos, y que da lugar a un procesamiento espacial adicional en un segundo paso de procesamiento ejecutado dependiendo de por lo menos parte de la información paramétrica relacionada con los objetos por el procesador de señales de audio.

Se ha descubierto que la extracción de una segunda información de audio, que comprende objetos de audio del segundo tipo de objetos de audio, de una representación de señal de mezcla descendente se puede ejecutar con moderada complejidad incluso si hay un gran número de objetos del segundo tipo de objetos de audio. Además, se ha encontrado que se puede ejecutar un procesamiento espacial de los objetos de audio del segundo tipo de audio de manera eficiente, una vez separada la segunda información de audio de la primera información de audio que describe los objetos de audio del primer tipo de objetos de audio.

Por añadidura, se ha encontrado que el algoritmo de procesamiento ejecutado por el separador de objetos para separar la primera información de audio y la segunda información de audio puede ser ejecutado con relativamente poca complejidad si se posterga el procesamiento individual por objetos de los objetos de audio del segundo tipo de objetos de audio al procesador de señales de

audio y no se ejecuta al mismo tiempo que la separación de la primera información de audio y la segunda información de audio.

En una realización preferida, el decodificador de señales de audio está configurado para producir la representación de señal de mezcla ascendente dependiendo de la representación de señal de mezcla descendente, la información paramétrica relacionada con los objetos y una información residual asociada con una subserie de objetos de audio representada por la representación de señal de mezcla descendente. En este caso, el separador de objetos está configurado para decomponer la representación de señal de mezcla descendente para producir la primera información de audio que describe la primera serie de uno o más objetos de audio (por ejemplo, los objetos en primer plano FGO) del primer tipo de objetos de audio con el cual está asociada la información residual y la segunda información de audio que describe la segunda serie de uno o más objetos de audio (por ejemplo, los objetos de fondo BGO) del segundo tipo de objetos de audio con el cual no hay ninguna información residual asociada que depende de la representación de la señal de mezcla descendente para usar por lo menos parte de la información paramétrica relacionada con los objetos y la información residual.

Esta realización se basa en el hallazgo de que se puede obtener una separación particularmente precisa entre la primera información de audio que describe la primera serie de objetos de audio del primer tipo de objetos de audio y la segunda información de audio que describe la segunda serie de objetos de audio del segundo tipo de objetos de audio utilizando una información residual además de la información paramétrica relacionada con los objetos. Se ha encontrado que el mero uso de la información paramétrica relacionada con los objetos daría lugar, en muchos casos, a distorsiones, que se pueden reducir significativamente, o incluso eliminar por completo, por el uso de la información residual. La información residual describe, por ejemplo, una distorsión residual, que se estima que ha de quedar si se aísla un objeto de audio del primer tipo de objetos de audio meramente utilizando la información paramétrica relacionada con los objetos. La información residual es calculada, por lo general, por un codificador de señales de audio. Aplicando la información residual, se puede mejorar la

separación entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio.

Esto permite obtener la primera información de audio y la segunda información de audio con una separación particularmente buena entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio, lo que, a su vez, permite obtener un procesamiento altamente espacial de los objetos de audio del segundo tipo de objetos de audio al procesar la segunda información de audio en el procesador de señales de audio.

En una realización preferida, el separador de objetos está configurado, por lo tanto, para proporcionar la primera información de audio de tal manera que se realcen los objetos de audio del primer tipo de objetos de audio con respecto a los objetos de audio del segundo tipo de objetos de audio de la primera información de audio. El separador de objetos también está configurado para transmitir la segunda información de audio de tal manera que se realcen los objetos de audio del segundo tipo de objetos de audio con respecto a los objetos de audio del primer tipo de objetos de audio de la segunda información de audio.

En una realización preferida, el decodificador de señales de audio está configurado para ejecutar un procesamiento en dos etapas, de tal manera que el procesamiento de la segunda información de audio del procesador de señales de audio se ejecute después de la separación entre la primera información de audio que describe la primera serie de uno o más objetos de audio del primer tipo de objetos de audio y la segunda información de audio que describe la segunda serie de uno o más objetos de audio del segundo tipo de objetos de audio.

En una realización preferida, el procesador de señales de audio está configurado para procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos asociada a los objetos de audio del segundo tipo de objetos de audio e independiente de la información paramétrica relacionada con los objetos asociada a los objetos de audio del primer tipo de objetos de audio. En consecuencia, se puede obtener un procesamiento separado de los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio.

En una realización preferida, el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio

utilizando una combinación lineal de uno o más canales de mezcla descendente y uno o más canales residuales. En este caso, el separador de objetos está configurado para obtener parámetros de combinación para ejecutar la combinación lineal dependiendo de los parámetros de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio y dependiendo de los coeficientes de predicción de canales de los objetos de audio del primer tipo de objetos de audio. La computación de los coeficientes de predicción de canales de los objetos de audio del primer tipo de objetos de audio puede tomar en cuenta, por ejemplo, los objetos de audio del segundo tipo de objetos de audio como único objeto de audio común. En consecuencia, se puede ejecutar un proceso de separación con una complejidad informática suficientemente baja que puede ser, por ejemplo, casi independiente del número de objetos de audio del segundo tipo de objetos de audio.

En una realización preferida, el separador de objetos está configurado para aplicar una matriz de renderización a la primera información de audio para mapear las señales de objeto de la primera información de audio sobre los canales de audio de la representación de señal de audio de mezcla ascendente. Esto se puede hacer porque el separador de objetos puede tener la capacidad de extraer señales de audio separadas que individualmente representan los objetos de audio del primer tipo de objetos de audio. En consecuencia, es posible mapear las señales de objeto de la primera información de audio directamente sobre los canales de audio de la representación de señal de audio de mezcla ascendente.

En una realización preferida, el procesador de audio está configurado para ejecutar un procesamiento en estéreo de la segunda información de audio que depende de una información de renderización, una información de covarianza relacionada con los objetos y una información de mezcla descendente, para obtener los canales de audio de la representación de señal de audio de mezcla ascendente.

En consecuencia, el procesamiento en estéreo de los objetos de audio del segundo tipo de objetos de audio es independiente de la separación entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio. De esa manera, no resulta afectada (ni se degrada) la separación eficiente entre los objetos de audio del primer tipo de

objetos de audio y los objetos de audio del segundo tipo de objetos de audio por el procesamiento en estéreo, lo que por lo general da lugar a una distribución de los objetos de audio en toda una pluralidad de canales de audio sin producir un alto grado de separación de objetos, que se puede obtener en el separador de objetos, por ejemplo, empleando la información residual.

En otra realización preferida, el procesador de audio está configurado para ejecutar un postprocesamiento de la segunda información de audio dependiendo de una información de renderización, una información de covarianza relacionada con los objetos y una información de mezcla descendente. Esta forma de postprocesamiento da lugar a una ubicación espacial de los objetos de audio del segundo tipo de objetos de audio dentro de una escena de audio. No obstante, debido al concepto de procesamiento en cascada, la complejidad informática del procesador de audio se puede mantener suficientemente baja, puesto que no es necesario que el procesador de audio tome en cuenta la información paramétrica relacionada con los objetos asociada a los objetos de audio del primer tipo de objetos de audio.

Además, el procesador de audio puede ejecutar diferentes tipos de procesamiento, como por ejemplo un procesamiento mono a binaural, un procesamiento mono a estéreo, un procesamiento estéreo a binaural o un procesamiento estéreo a estéreo.

En una realización preferida, el separador de objetos está configurado para tratar objetos de audio del segundo tipo de objetos de audio, a los cuales no está asociada ninguna información residual, en forma de objeto de audio único. Asimismo, el procesador de señales de audio está configurado para tomar en cuenta parámetros de renderización de objetos específicos para ajustar las contribuciones de los objetos del segundo tipo de objetos de audio a la representación de señal de mezcla ascendente. Por consiguiente, el separador de objetos considera los objetos de audio del segundo tipo de objetos de audio como único objeto de audio, lo que reduce significativamente la complejidad del separador de objetos y también permite contar con una información residual distintiva, que es independiente de los parámetros de renderización asociados con los objetos de audio del segundo tipo de objetos de audio.

En una realización preferida, el separador de objetos está configurado para obtener un valor de diferencia de nivel común de los objetos correspondiente a una pluralidad de objetos de audio del segundo tipo de objetos de audio. El separador de objetos está configurado para usar el valor de diferencia de nivel común de los objetos para la computación de los coeficientes de predicción de canales. Además, el separador de objetos está configurado para usar los coeficientes de predicción de canales para obtener uno o dos canales de audio que representan la segunda información de audio. Para obtener un valor común de diferencia de nivel de los objetos, los objetos de audio del segundo tipo de objetos de audio pueden ser tratados de manera eficiente como único objeto de audio por el separador de objetos.

En una realización preferida, el separador de objetos está configurado para obtener un valor común de diferencia de nivel de los objetos correspondiente a una pluralidad de objetos de audio del segundo tipo de objetos de audio y el separador de objetos está configurado para utilizar el valor común de diferencia de nivel de los objetos correspondiente a la computación de las anotaciones de una matriz de mapeo en modo de energía. El separador de objetos está configurado para usar la matriz de mapeo en modo de energía para obtener dicho uno o más canales de audio que representan la segunda información de audio. Una vez más, el valor común de diferencia de nivel de los objetos permite un tratamiento común informáticamente eficiente de los objetos de audio del segundo tipo de objetos de audio por el separador de objetos.

En una realización preferida, el separador de objetos está configurado para obtener selectivamente un valor común de correlación entre objetos asociado a los objetos de audio del segundo tipo de objetos de audio dependiendo de la información paramétrica relacionada con los objetos si se encuentra que hay dos objetos de audio del segundo tipo de objetos de audio y para ajustar el valor de correlación entre objetos asociado a los objetos de audio del segundo tipo de objetos de audio a cero si se halla que hay más o menos de dos objetos de audio del segundo tipo de objetos de audio.

El separador de objetos está configurado para usar el valor común de correlación entre objetos asociado a los objetos de audio del segundo tipo de objetos de audio para obtener el uno o más canales de audio que representan la

segunda información de audio. Empleando esta estrategia, se aprovecha el valor de correlación entre objetos si se lo puede obtener con gran eficiencia de computación, es decir, si hay dos objetos de audio del segundo tipo de objetos de audio. De lo contrario, sería una gran exigencia de computación la obtención de valores de correlación entre objetos. En consecuencia, se ha encontrado que se obtendría un buen equilibrio en términos de impresión auditiva y complejidad informática la fijación del valor de correlación entre objetos asociado a los objetos de audio del segundo tipo de objetos de audio a cero si hay más o menos de dos objetos de audio del segundo tipo de objetos.

En una realización preferida, el procesador de señales de audio está configurado para renderizar la segunda información de audio dependiendo de (por lo menos parte de) la información paramétrica relacionada con los objetos, para obtener una representación renderizada de los objetos de audio del segundo tipo de objetos de audio en forma de versión procesada de la segunda información de audio. En este caso, la renderización se puede realizar independientemente de los objetos de audio del primer tipo de objetos de audio.

En una realización preferida, el separador de objetos está configurado para proporcionar la segunda información de audio de tal manera que la segunda información de audio describa más de dos objetos de audio del segundo tipo de objetos de audio. Las realizaciones de acuerdo con la invención permiten un ajuste flexible del número de objetos de audio del segundo tipo de objetos de audio, lo que se ve considerablemente facilitado por la estructura de etapas en cascada del procesamiento.

En una realización preferida, el separador de objetos está configurado para obtener, como segunda información de audio, una representación de señal de audio de un canal único o una representación de señal de audio de dos canales que representan más de dos objetos de audio del segundo tipo de objetos de audio.

La extracción de uno o dos canales de señales de audio puede ser ejecutada por el separador de objetos con baja complejidad informática. En particular, la complejidad del separador de objetos se puede mantener significativamente más baja en comparación con un caso en el cual el separador

de objetos necesitaría tratar más de dos objetos de audio del segundo tipo de objetos de audio.

Sin embargo, se ha encontrado que es una representación informáticamente eficiente de los objetos de audio del segundo tipo de objetos de audio el uso de uno o dos canales de una señal de audio.

En una realización preferida, el procesador de señales de audio está configurado para recibir la segunda información de audio y para procesar la segunda información de audio dependiendo de (por lo menos parte de) la información paramétrica relacionada con los objetos, tomando en cuenta información paramétrica relacionada con los objetos asociada a más de dos objetos de audio del segundo tipo de objetos de audio. En consecuencia, el procesador de audio ejecuta un procesamiento de objetos individuales, en tanto que el separador de objetos no lleva a cabo un procesamiento de objetos individuales respecto de los objetos de audio del segundo tipo de objetos de audio.

En una realización preferida, el decodificador de audio está configurado para extraer una información de número total de objetos y una información de número de objetos en primer plano a partir de una información de configuración relacionada con la información paramétrica relacionada con los objetos. El decodificador de audio también está configurado para determinar un número de objetos de audio del segundo tipo de objetos de audio formando una diferencia entre la información del número total de objetos y la información del número de objetos en primer plano. En consecuencia, se obtiene una señalización eficiente del número de objetos de audio del segundo tipo de objetos de audio. Más aun, este concepto ofrece un alto grao de flexibilidad con respecto al número de objetos de audio del segundo tipo de objetos de audio.

En una realización preferida, el separador de objetos está configurado para usar información paramétrica relacionada con los objetos asociada a N_{eao} objetos de audio del primer tipo de objetos de audio para obtener, como primera información de audio, N_{eao} señales de audio que representan (preferentemente en forma individual), los N_{eao} objetos de audio del primer tipo de objetos de audio, y para obtener, como segunda información de audio, una o dos señales de audio que representan los $N-N_{eao}$ objetos de audio del segundo tipo de objetos de audio, tratar los $N-N_{eao}$ objetos de audio del segundo tipo de objetos de audio en forma

de un único objeto de audio monocanal o de dos canales. El procesador de señales de audio está configurado para renderizar individualmente los $N-N_{eao}$ objetos de audio representados por la una o dos señales de audio de la segunda información de audio usando la información paramétrica relacionada con los objetos asociada a los $N-N_{eao}$ objetos de audio del segundo tipo de objetos de audio. En consecuencia, la separación de objetos de audio entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio es independiente del procesamiento posterior de los objetos de audio del segundo tipo de objetos de audio.

Una realización de acuerdo con la invención da origen a un método para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto.

Otra realización de acuerdo con la invención genera un programa de computación para poner en práctica dicho método.

Breve Descripción de las Figuras.

A continuación se describen las realizaciones de acuerdo con la invención con referencia a las Figs. adjuntas, en las cuales:

La Fig. 1 ilustra un diagrama esquemático de bloques de un decodificador de señales de audio, de acuerdo con una realización de la invención;

La Fig. 2 ilustra un diagrama esquemático de bloques de otro decodificador de señales de audio de acuerdo con una realización de la invención;

Las Figs. 3a y 3b ilustran diagramas esquemáticos de bloques de un procesador residual que se puede utilizar como separador de objetos en una realización de la invención;

Las Figs. 4a a 4e ilustran diagramas esquemáticos de bloques de procesadores de señales de audio que se pueden utilizar en un decodificador de señales de audio de acuerdo con una realización de la invención:

La Fig. 4f ilustra un diagrama de bloques de un modo de procesamiento del transcodificador SAOC;

La Fig. 4g ilustra un diagrama de bloques de un modo de procesamiento del decodificador SAOC;

La Fig. 5a ilustra un diagrama esquemático de bloques de un decodificador de señales de audio de acuerdo con una realización de la invención;

La Fig. 5b ilustra un diagrama esquemático de bloques de otro decodificador de señales de audio de acuerdo con una realización de la invención;

La Fig. 6a ilustra una Tabla que representa una descripción del diseño de una prueba de audición;

La Fig. 6b ilustra una Tabla que representa los sistemas en estudio;

La Fig. 6c ilustra una Tabla que representa los elementos de la prueba de audición y las matrices de renderización;

La Fig. 6d ilustra una representación gráfica de puntuaciones MUSHRA correspondientes a la prueba de audición de la renderización del tipo Karaoke/Solo;

La Fig. 6e ilustra una representación gráfica de puntuaciones MUSHRA correspondientes a una prueba de audición de renderización clásica;

La Fig. 7 ilustra un gráfico de flujo de un método para producir una representación de señal de mezcla ascendente de acuerdo con una realización de la invención;

La Fig. 8 ilustra un diagrama esquemático de bloques de un sistema MPEG SAOC de referencia;

La Fig. 9a ilustra un diagrama esquemático de bloques de un sistema SAOC de referencia que utiliza decodificador y mezclador separados;

La Fig. 9b ilustra un diagrama esquemático de bloques de un sistema SAOC de referencia que utiliza un decodificador y mezclador integrados y

La Fig. 9c ilustra un diagrama esquemático de bloques de un sistema SAOC de referencia que utiliza un transcodificador SAOC a MPEG.

Descripción Detallada de las Realizaciones

1. Decodificador de señales de audio de acuerdo con la Fig. 1

La Fig. 1 ilustra un diagrama esquemático de bloques de un decodificador de señales de audio 100 de acuerdo con una realización de la invención.

El decodificador de señales de audio 100 está configurado para recibir una información paramétrica relacionada con los objetos 110 y una representación de señal de mezcla descendente 112. El decodificador de señales de audio 100 está configurado para transmitir una representación de señal de mezcla ascendente

120 dependiendo de la representación de la señal de mezcla descendente y la información paramétrica relacionada con los objetos 110. El decodificador de señales de audio 100 comprende un separador de objetos 130, que está configurado para descomponer la representación de señal de mezcla descendente 112 para ofrecer una primera información de audio 132 que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio 134 que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio que depende de la representación de señal de mezcla descendente 112 y utilizar por lo menos parte de la información paramétrica relacionada con los objetos 110. El decodificador de señales de audio 100 comprende asimismo un procesador de señales de audio 140, que está configurado para recibir la segunda información de audio 134 y para procesar la segunda información de audio que depende de por lo menos parte de la información paramétrica relacionada con los objetos 112, para obtener una versión procesada 142 de la segunda información de audio 134. El decodificador de señales de audio 100 comprende asimismo un combinador de señales de audio 150 configurado para combinar la primera información de audio 132 con la versión procesada 142 de la segunda información de audio 134, para obtener la representación de señal de mezcla ascendente 120.

El decodificador de señales de audio 100 implementa un procesamiento en etapas en cascada de la representación de señal de mezcla descendente, que representa objetos de audio del primer tipo de objetos de audio y objetos de audio del segundo tipo de objetos de audio de manera combinada.

En una primera etapa de procesamiento, que es ejecutado por el separador de objetos 130, la segunda información de audio que describe una segunda serie de objetos de audio del segundo tipo de objetos de audio se separa de la primera información de audio 132 que describe una primera serie de objetos de audio de un primer tipo de objetos de audio empleando la información paramétrica relacionada con los objetos 110. Sin embargo, la segunda información de audio 134 es por lo general una información de audio (por ejemplo, una señal de audio de un solo canal o una señal de audio de dos canales) que describe los objetos de audio del segundo tipo de objetos de audio de manera combinada.

En la segunda etapa de procesamiento, el procesador de señales de audio 140 procesa la segunda información de audio 134 que depende de la información paramétrica relacionada con los objetos. En consecuencia, el procesador de señales de audio 140 tiene la capacidad de ejecutar un procesamiento de objetos individuales o de renderizar los objetos de audio del segundo tipo de objetos de audio, lo que se describe en la segunda información de audio 134, y que por lo general no es ejecutado por el separador de objetos 130.

Por consiguiente, si bien es preferible que el separador de objetos 130 no procese los objetos de audio del segundo tipo de objetos de audio de manera individual, los objetos de audio del segundo tipo de objetos de audio son procesados, por cierto, de manera individual por objeto (por ejemplo, renderizados de manera individual por objeto) en la segunda etapa de procesamiento, que es ejecutada por el procesador de señales de audio 140. Por consiguiente, la separación entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio, que es ejecutada por el separador de objetos 130, es independiente del procesamiento individual por objetos de los objetos de audio del segundo tipo de objetos de audio, que es ejecutada posteriormente por el procesador de señales de audio 140. En consecuencia, el procesamiento ejecutado por el separador de objetos 130 es sustancialmente independiente de un número de objetos de audio del segundo tipo de objetos de audio. Además, el formato (por ejemplo, señal de audio de un solo canal o señal de audio de dos canales) de la segunda información de audio 134 depende, por lo general, del número de objetos de audio del segundo tipo de objetos de audio. Por consiguiente, se puede variar el número de objetos de audio del segundo tipo de objetos de audio sin necesidad de modificar la estructura del separador de objetos 130. Dicho de otro modo, los objetos de audio del segundo tipo de objetos de audio son tratados como único objeto de audio (por ejemplo, de un canal o de dos canales) con respecto al cual el separador de objetos 140 obtiene una información paramétrica común relacionada con los objetos (por ejemplo, un valor común de diferencia de nivel de los objetos asociado con uno o dos canales de audio).

En consecuencia, el decodificador de señales de audio 100 de acuerdo con la Fig. 1 es apto para manejar un número variable de objetos de audio del

segundo tipo de objetos de audio sin una modificación estructural del separador de objetos 130. Además, se pueden aplicar diferentes algoritmos de procesamiento de objetos mediante el separador de objetos 130 y el procesador de señales de audio 140.

En consecuencia, por ejemplo, es posible realizar una separación de objetos de audio mediante el uso por el separador de objetos 130 de información residual, lo que da lugar a una separación particularmente eficaz de diferentes objetos de audio, haciendo uso de la información residual, que constituye una información complementaria para mejorar la calidad de una separación de objetos. Por el contrario, el procesador de señales de audio 140 puede ejecutar un procesamiento individual por objeto sin emplear una información residual. Por ejemplo, el procesador de señales de audio 140 puede estar configurado para ejecutar un procesamiento de señales de audio del tipo convencional de codificación espacial de objetos de audio (SAOC) para renderizar los diferentes objetos de audio.

2. Decodificador de señales de audio de acuerdo con la Fig. 2

A continuación se describe un decodificador de señales de audio 200 de acuerdo con una realización de la invención. En la Fig. 2 se ilustra un diagrama esquemático de bloques de este decodificador de señales de audio 200.

El decodificador de audio 200 está configurado para recibir una señal de mezcla descendente 210, un denominado flujo de bits SAOC 212, información de la matriz de renderización 2 14 y, opcionalmente, parámetros de transferencia de cabeza (HRTF) 216. El decodificador de señales de audio 200 también está configurado para emitir una señal de mezcla descendente de salida/MPS 220 y (opcionalmente) un flujo de bits MPS 222.

2.1. Señales de entrada y señales de salida del decodificador de señales de audio 200

A continuación, se describen diversos detalles con respecto a las señales de entrada y las señales de salida del decodificador de audio 200.

La señal de mezcla descendente 200 puede ser, por ejemplo, una señal de audio de un solo canal o una señal de audio de dos canales. La señal de mezcla descendente 210 puede derivar, por ejemplo, de una representación codificada de la señal de mezcla descendente.

El flujo de bits por codificación espacial de objetos de audio (Flujo de bits SAOC) 212 puede comprender, por ejemplo, información paramétrica relacionada con los objetos. Por ejemplo, el flujo de bits SAOC 212 puede comprender información de diferencia de nivel de los objetos, por ejemplo, en forma de parámetros de diferencia de nivel de los objetos OLD, una información de correlación entre objetos, por ejemplo, en forma de parámetros de correlación entre objetos IOC.

Por otra parte, el flujo de bits SAOC 212 puede comprender una información de mezcla descendente que describe de que manera se han producido las señales de mezcla descendente sobre la base de una pluralidad de señales de objeto de audio utilizando un proceso de mezcla descendente. Por ejemplo, el flujo de bits SAOC puede comprender un parámetro de ganancia de mezcla descendente DMG y (opcionalmente) parámetros de diferencia de nivel de canales de mezcla descendente DCLD.

La información de matrices de renderización 214 puede describir, por ejemplo, de qué manera se los diferentes objetos de audio han de ser renderizados por el decodificador de audio. Por ejemplo, la información de matrices de renderización 214 puede describir una asignación de un objeto de audio a uno o más canales de la señal de mezcla descendente de salida/MPS 220.

La información opcional de parámetros de función de transferencia de cabeza (HRTF) 216 puede describir además una función de transferencia para derivar una señal binaural para auriculares.

La señal de mezcla descendente de salida/MPEG Envolvente (también abreviada como "señal de mezcla descendente de salida/MPS") 220 representa uno o más canales de audio, por ejemplo, en forma de una representación de una señal de audio en el dominio de tiempo o una representación de una señal de audio en el dominio de la frecuencia. Sola o en combinación con el flujo de bits opcional MPEG–Envolvente (flujo de bits MPS) 222, que comprende parámetros de MPEG–Envolvente que describen un mapeo de la señal de mezcla descendente de salida/MPS 220 sobre una pluralidad de canales de audio, se forma una representación de señal de mezcla ascendente.

2.2. Estructura y funcionalidad del decodificador de señales de audio 200

A continuación se describe en forma más detallada la estructura del decodificador de señales de audio 200, que puede cumplir la función de un transcodificador SAOC o la función de un decodificador SAOC.

El decodificador de señales de audio 200 comprende un procesador de mezcla descendente 230, que está configurado para recibir la señal de mezcla descendente 210 y para producir, sobre la base de ésta, la señal de mezcla descendente de salida/MPS 220. El procesador de mezcla descendente 230 también está configurado para recibir por lo menos parte de la información de flujo de bits SAOC 212 y por lo menos parte de la información de la matriz de renderización 214. Además, el procesador de mezcla descendente 230 puede recibir asimismo una información sobre parámetros SAOC procesados 240 procedente de un procesador un procesador de parámetros 250.

El procesador de parámetros 250 está configurado para recibir la información sobre flujo de bits SAOC 212, información sobre matrices de renderización 214 y, opcionalmente, la información de parámetros de función de transferencia de cabeza 260, y para producir, sobre la base de ésta, el flujo de bits MPEG Envolvente 222 que acarrea los parámetros de MPEG envolvente (en caso de que sean necesarios los parámetros de MPEG envolvente, como ocurre, por ejemplo, en el modo de operación de transcodificación).

Además, el procesador de parámetros 250 suministra la información SAOC procesada (en caso de ser necesaria esta información de SAOC procesada).

A continuación se describe en forma más detallada la estructura y funcionalidad del procesador de mezcla descendente 230.

El procesador de mezcla descendente 230 comprende un procesador residual 260, que está configurado para recibir la señal de mezcla descendente 210 y para producir, sobre la base de ésta, una primera señal de objetos de audio 262 que describe los denominados objetos de audio realzados (EAOs), que se pueden considerar objetos de audio de un primer tipo de objetos de audio. La primera señal de objetos de audio puede comprender uno o más canales de audio y se la puede considerar como primera información de audio. El procesador residual 260 también está configurado para producir una segunda señal de objetos de audio 264, que describe objetos de audio de un segundo tipo de objetos de audio y puede ser considerada segunda información de audio. La

segunda señal de objetos de audio 264 puede comprender uno o más canales y por lo general puede comprender uno o dos canales de audio que describen una pluralidad de objetos de audio. Generalmente, la segunda señal de objetos de audio puede describir aun más de dos objetos de audio del segundo tipo de objetos de audio.

El procesador de mezcla descendente 230 comprende asimismo un preprocesador de mezcla descendente SAOC 270, que está configurado para recibir la segunda señal de objetos de audio 264 y para producir, sobre la base de ésta, una versión procesada 272 de la segunda señal de objetos de audio 264, que se puede considerar una versión procesada de la segunda información de audio.

El procesador de mezcla descendente 230 comprende asimismo un combinador de señales de audio 280, que está configurado para recibir la primera señal de objetos de audio 262 y la versión procesada 272 de la segunda señal de objetos de audio 264 y para producir, sobre la base de ésta, la señal de mezcla descendente de salida/MPS 220, que puede ser considerada, sola o junto con el correspondiente flujo de bits MPEG–Envolvente (opcional) 222, una representación de señal de mezcla ascendente.

A continuación se describe en forma más detallada la funcionalidad de las unidades individuales del procesador de mezcla descendente 230.

El procesador residual 260 está configurado para aportar, por separado, la primera señal de objetos de audio 262 y la segunda señal de objetos de audio 264. Para este fin, el procesador residual 260 puede estar configurado para aplicar por lo menos parte de la información de flujo de bits SAOC 212. Por ejemplo, el procesador residual 260 puede estar configurado para evaluar una información paramétrica relacionada con los objetos asociada a los objetos de audio del primer tipo de objetos de audio, es decir, los denominados "objetos de audio realzados" EAO. Además, el procesador residual 260 puede estar configurado para obtener una información general que describe comúnmente los objetos de audio del segundo tipo de objetos de audio, por ejemplo, los denominados "objetos de audio no realzados".

El procesador residual 260 puede estar configurado asimismo para evaluar una información residual, que está incluida en la información de flujo de bits SAOC

212, para la separación entre objetos de audio realzados (objetos de audio del primer tipo de objetos de audio) y objetos de audio no realzados (objetos de audio del segundo tipo de objetos de audio). La información residual puede codificar, por ejemplo, una señal residual en el dominio del tiempo, que se aplica para obtener una separación particularmente limpia entre los objetos de audio realzados y los objetos de audio no realzados. Además, el procesador residual 260 puede evaluar, opcionalmente, por lo menos una parte de la información de matrices de renderización 214, por ejemplo, a fin de determinar la distribución de los objetos de audio realzados por los canales de audio de la primera señal de objetos de audio 262.

El preprocesador de mezcla descendente SAOC 270 comprende un redistribuidor de canales 274, que está configurado para recibir el uno o más canales de audio de la segunda señal de objetos de audio 264 y para producir, sobre la base de ésta, uno o más canales de audio (típicamente dos) de la segunda señal de objetos de audio procesada 272. Además, el preprocesador de mezcla descendente SAOC 270 comprende un productor de señales descorrelacionadas 276, que está configurado para recibir el uno o más canales de audio de la segunda señal de objetos de audio 264 y para producir, sobre la base de ésta, uno o más señales descorrelacionadas 278a, 278b, que se agregan a las señales provistas por el re–distribuidor de canales 274 a fin de obtener la versión procesada 272 de la segunda señal de objetos de audio 264. Más adelante se explican más detalles con respecto al procesador SAOC de mezcla descendente.

El combinador de señales de audio 280 combina la primera señal de objetos de audio 262 con la versión procesada 272 de la segunda señal de objetos de audio. Para este fin, se puede realizar una combinación por canales. De manera acorde se obtiene la señal de mezcla descendente de salida/MPS 220.

El procesador de parámetros 250 está configurado para obtener los parámetros (opcionales) de MPEG–Envolvente, que conforman el flujo de bits MPEG–Envolvente 222 de la representación de señal de mezcla ascendente, sobre la base del flujo de bits SAOC, tomando en cuenta la información de matrices de renderización 214 y, opcionalmente, la información de parámetros HRTF 21 6. Dicho de otro modo, el procesador de parámetros SAOC 252 está

configurado para traducir la información de parámetros relacionados con los objetos, que es la descripta por la información de flujo de bits SAOC 212, en una información paramétrica relacionada con los canales, descripta por el flujo de bits MPEG Envolvente 222.

A continuación se presenta una breve reseña general de la estructura de la arquitectura del transcodificador/decodificador SAOC expuesta en la Fig. 2. La codificación espacial de objetos de audio (SAOC) es una técnica paramétrica de codificación de múltiples objetos. Está destinada a transmitir un número de objetos de audio en una señal de audio (por ejemplo la señal de audio de mezcla descendente 210) que comprende M canales. Junto con esta señal de mezcla descendente inversa compatible, se transmiten parámetros de objetos (por ejemplo, utilizando la información de flujo de bits SAOC 212) que dan lugar a la recreación y manipulación de las señales de objeto originales. Un codificador SAOC (que no se ilustra aquí) produce una mezcla descendente de las señales de objeto en su entrada y extrae estos parámetros de objeto. El número de objetos que se puede manejar no está limitado en principio. Los parámetros de objeto son cuantificados y codificados eficientemente para integrar el flujo de bits SAOC 212. La señal de mezcla descendente 210 puede ser comprimida y transmitida sin necesidad de actualizar los codificadores e infraestructuras existentes. Los parámetros de objetos, o información complementaria SAOC, se transmiten por un canal secundario de baja velocidad de transmisión de bits, por ejemplo, la porción de datos auxiliares del flujo de bits de mezcla descendente.

Del lado del decodificador, los objetos de entrada son reconstruidos y renderizados a un número de canales de reproducción. La información de renderización que contiene el nivel de reproducción y la posición de paneo por cada objeto es provista por el usuario o se la puede extraer del flujo de bits SAOC (por ejemplo, como información preestablecida). La información de renderización puede variar temporalmente.

Las configuraciones de salida pueden variar de mono a multicanal (por ejemplo, 5.1) y son independientemente tanto del número de objetos de entrada como del número de canales de mezcla descendente. La renderización binaural de objetos es posible, incluyendo el azimut y la elevación de las posiciones virtuales de los objetos. Una interfaz de efectos opcionales da lugar a la

manipulación avanzada de señales de objeto, aparte de la modificación de nivel y paneo. Los objetos en sí pueden ser señales mono, señales estereofónicas, como así también señales multicanales (por ejemplo 5.1 canales). Las configuraciones típicas de mezcla descendente son mono y estéreo.

A continuación, se pasa a explicar la estructura básica del transcodificador/decodificador SAOC, que se ilustra en la Fig. 2. El módulo de transcodificador/decodificador SAOC aquí descripto puede actuar como decodificador autoportante o como transcodificador de un flujo de bits SAOC a MPEG–envolvente, dependiendo de la configuración pretendida de los canales de salida. En un primer modo de funcionamiento, la configuración de la señal de salida es mono, estéreo o binaural, y se utilizan dos canales de salida. En este primer caso, el módulo SAOC puede operar en un modo de decodificador y la salida del módulo SAOC es una salida de modulación por pulsos codificados (salida PCM). En el primer caso, no es necesario un decodificador MPEG envolvente. Por el contrario, la representación de señal de mezcla ascendente puede comprender sólo la señal de salida 220, y la vez se puede omitir la provisión del flujo de bits MPEG envolvente 222. En un segundo caso, la configuración de la señal de salida es una configuración multicanal con más de dos canales de salida. El módulo SAOC puede ser operativo en un modo de transcodificador. La salida del módulo SAOC puede comprender tanto una señal de mezcla descendente 220 como un flujo de bits MPEG envolvente 222 en este caso, como se ilustra en la Fig. 2. En consecuencia, se necesita un decodificador MPEG envolvente para obtener una representación final de la señal de audio para la salida de los parlantes.

La Fig. 2 ilustra la estructura básica de la arquitectura del transcodificador/decodificador SAOC. El procesador residual 216 extrae el objeto de audio realzado de la señal de mezcla descendente de entrada 210 usando la información residual contenida en el flujo de bits SAOC 212. El preprocesador de mezcla descendente 270 procesa los objetos de audio normales (que son, por ejemplo, objetos de audio no realzados, es decir, objetos de audio con respecto a los cuales no se transmite información en el flujo de bits SAOC 212). Los objetos de audio realzados (representados por la primera señal de objetos de audio 262) y los objetos de audio normales procesados (representados, por ejemplo, por la

versión procesada 272 de la segunda señal de objetos de audio 264) se combinan en la señal de salida 220 correspondiente al modo de decodificador SAOC o a la señal de mezcla descendente MPEG envolvente 220 correspondiente al modo de transcodificador SAOC. A continuación se presentan descripciones detalladas de los bloques de procesamiento.

3. Arquitectura y funcionalidad del Procesador Residual y del Procesador de Modo de Energía

A continuación, se describen detalles referentes a un procesador residual que puede asumir, por ejemplo, la funcionalidad del separador de objetos 130 del decodificador de señales de audio 100 o del procesador residual 260 del decodificador de señales de audio 200. Para este fin, las Figs. 3a y 3b ilustran diagramas esquemáticos de boques de ese tipo de procesador residual 300, que puede tomar el lugar del separador de objetos 130 o del procesador residual 260. La Fig. 3a presenta menos detalles que la Fig. 3b. Sin embargo, la siguiente descripción se aplica al procesador residual 300 de acuerdo con la Fig. 3a y también al procesador residual 380 de acuerdo con la Fig. 3b. El procesador residual 300 está configurado para recibir una señal de mezcla descendente SAOC 3 10, que puede ser equivalente a la representación de señal de mezcla descendente 112 de la Fig. 1 o la representación de señal de mezcla descendente 210 de la Fig. 2. El procesador residual 300 está configurado para proporcionar, sobre la base de ésta, una primera información de audio 320 que describe uno o más objetos de audio realzados, que pueden ser equivalentes, por ejemplo, a la primera información de audio 132 o a la primera señal de objetos de audio 262. Además, el procesador residual 300 puede suministrar una segunda información de audio 322 que describe uno o más objetos de audio adicionales (por ejemplo, objetos de audio no realzados, con respecto a los cuales no se dispone de información residual), donde la segunda información de audio 322 puede ser equivalente a la segunda información de audio 134 o a la segunda señal de objetos de audio 264.

El procesador residual 300 comprende una unidad 1–a–N/2–a–N (unidad OTN/TTN) 330, que recibe la señal de mezcla descendente SAOC 310 y que también recibe datos SAOC y residuales 332. La unidad 1–a–N/2–a–N 330 también emite una señal de objetos de audio realzados 334, que describe los

objetos de audio realzados (EAO) contenidos en la señal de mezcla descendente SAOC 310.

Asimismo, la unidad 1–a–N/2–a–N 330 proporciona la segunda información de audio 322. El procesador residual 300 comprende asimismo una unidad de renderización 340, que recibe la señal de objetos de audio realzados 334 e información de matrices de renderización 342 y produce, sobre la base de ésta, la primera información de audio 320.

A continuación, se describe en forma más detallada el procesamiento de objetos de audio realzados (procesamiento de EAO), que es ejecutado por el procesador residual 300.

3.1. Introducción a la Operación del procesador residual 300

Con respecto a la funcionalidad del procesador residual 300, cabe señalar que la tecnología SAOC permite la manipulación individual de un número de objetos de audio en términos de su amplificación/atenuación de nivel sin desmedro significativo de la calidad del sonido obtenido como resultado, sólo de manera limitada. Una configuración especial de la aplicación del tipo "karaoke " requiere una supresión total (o casi total) de los objetos específicos, por lo general la primera voz, manteniendo indemne la calidad perceptual del sonido de fondo. Un caso típico de aplicación contiene hasta cuatro señales de objetos de audio realzados (EAO), que pueden representar, por ejemplo, dos objetos estéreo independientes (por ejemplo, dos objetos estéreo independientes preparados para ser retirados del lado del decodificador).

Se debe tener en cuenta que los (uno o más) objetos de audio de calidad realzada (o, más precisamente, las contribuciones a la señal de audio asociadas a los objetos de audio realzados) están incluidos en la señal de mezcla descendente SAOC 310. Por lo general, las contribuciones de las señales de audio asociadas a los (uno o más) objetos de audio realzados son mezcladas por el procesamiento de mezcla descendente ejecutado por el codificador de señales de audio, con las contribuciones de señales de audio de otros objetos de audio, que no son objetos de audio realzados. Además, también se debe notar que las contribuciones de señales de audio de una pluralidad de objetos de audio realzados también se superponen o mezclan, por lo general, en el procesamiento de mezcla descendente ejecutado por el codificador de señales de audio.

3.2 Arquitectura SOAC que sirve como soporte para los Objetos de audio realzados

A continuación, se describen detalles con respecto al procesador residual 300. El procesamiento de objetos de audio realzados incorpora las unidades 1–a–N o 2–a–N, dependiendo del modo de mezcla descendente SAOC. La unidad de procesamiento 1–a–N está dedicada a la señal de mezcla descendente mono y la unidad de procesamiento 2–a–N está dedicada a una señal de mezcla descendente estéreo 310. Estas dos unidades representan una modificación generalizada y realzada de la caja 2–a–2 (caja TTT) que fuera dada a conocer por ISO/IEC 23003–1:2007. En el codificador, se combinan señales regulares y EAO en la mezcla descendente. Las unidades de procesamiento OTN^{-1}–$/TTN^{-1}$ (que son unidades de procesamiento uno a N inversas o unidades de procesamiento 2 a N inversas) son utilizadas para producir y codificar las correspondientes señales residuales.

Las señales EAO y regulares son recuperadas de la mezcla descendente 310 por las unidades OTN/TTN 330 empleando la información complementaria SAOC y las señales residuales incorporadas. Los EAOs recuperados (descriptos por las señales de objetos de audio realzados 334) son alimentados a la unidad de renderización 340 que representa (o crea el producto de la matriz de renderización correspondiente (descripta por la información sobre matrices de renderización 342) y la salida obtenida como resultado de la unidad OTN/TTN. Los objetos de audio normales (descriptos por la segunda información de audio 322) son transmitidos al preprocesador de mezcla descendente SAOC, por ejemplo, el preprocesador de mezcla descendente SAOC 270, para continuar su procesamiento. Las Figs. 3a y 3b ilustran la estructura general del procesador residual, es decir, la arquitectura del procesador residual.

Las señales de salida del procesador residual 320, 322 se computan de la siguiente manera:

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}\mathbf{X}_{res},$$

$$\mathbf{X}_{EAO} = \mathbf{A}_{EAO}\mathbf{M}_{EAO}\mathbf{X}_{res},$$

donde $\mathbf{X}_{obj}$ representa la señal de mezcla descendente de los objetos de audio normales (es decir, los no–EAOs) y $\mathbf{X}_{EAO}$ es la señal de salida EAO

renderizada para el modo de decodificador SAOC o la correspondiente señal de mezcla descendente de EAO para el modo de transcodificador SAOC. El procesador residual puede operar en el modo de predicción (utilizando la información residual) o en el modo de energía (sin información residual). La señal de entrada ampliada $\mathbf{X}_{res}$ se define de manera consiguiente:

$$\mathbf{X}_{res} = \begin{cases} \begin{pmatrix} \mathbf{X} \\ \hline \mathbf{res} \end{pmatrix}, & \text{para el modo de predicción} \\ \mathbf{X}, & \text{para el modo de energía} \end{cases}$$

En este caso, **X** puede representar, por ejemplo, el uno o más canales de la representación de señal de mezcla descendente 310, que pueden ser transportados en el flujo de bits que representa el contenido de audio multicanal. **res** puede designar una más señales residuales, que pueden ser descriptas por el flujo de bits que representan el contenido de audio de múltiples canales.

El procesamiento OTN/TTN está representado por la matriz M y el procesador EAO por la matriz $\mathbf{A}_{EAO}$.

La matriz de procesamiento OTN/TTN M se define de acuerdo con el modo de operación de EAO (es decir, el de predicción o energía) de la siguiente manera:

$$\mathbf{M} = \begin{cases} \mathbf{M}_{\text{predicción}}, & \text{en el caso del modo de predicción} \\ \mathbf{M}_{\text{energía}}, & \text{en el caso del modo de energía} \end{cases}$$

La matriz de procesamiento OTN/TTN **M** está representada por

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{OBJ} \\ \hline \mathbf{M}_{EAO} \end{pmatrix},$$

donde la matriz N se relaciona con los objetos de audio normales (es decir, los no EAOs) y $\mathbf{M}_{EAO}$ con los objetos de audio realzados (EAOs).

En algunas realizaciones, uno o más objetos de fondo multicanales (MBO) pueden ser tratados de igual manera por el procesador residual 300.

Un Objeto de Fondo Multicanal (MBO) es una mezcla descendente mono o estéreo MPS que forma parte de la mezcla descendente SAOC. A diferencia del uso de objetos SAOC individuales para cada canal en una señal multicanal, se puede utilizar un MBO que habilite a SAOC para manejar con más eficiencia un objeto multicanal. En el caso del MBO, la sobrecarga de SAOC se reduce a medida que los parámetros SAOC de los MBOs sólo se relacionan con los canales

de mezcla descendente en lugar de hacerlo con todos los canales de mezcla ascendente.

3.3 Otras Definiciones

3.3.1 Dimensionalidad de las Señales y Parámetros

15 A continuación, se describe brevemente la dimensionalidad de las señales y parámetros a fin de arrojar más claridad sobre la frecuencia con que se realizan los diferentes cálculos.

Las señales de audio se definen por cada franja horaria n y cada subbanda híbrida (que puede ser una subbanda de frecuencia) k. Los correspondientes parámetros SAOC se definen por cada franja horaria de parámetros 1 y la banda de procesamiento m. Un mapeo subsiguiente entre el dominio híbrido y de los parámetros está especificado por la tabla A. 31 ISO/IEC 23003–1:2007. Por ende, todos los cálculos se realizan con respecto a ciertos índices de banda horaria y las correspondientes dimensionalidades están implícitas por cada variable introducida.

Sin embargo, a continuación se omiten, en ocasiones, los índices de bandas de tiempo y frecuencia para mantener la notación concisa.

3.3.2 Cálculo de la matriz A_{EAO},

La matriz de pre–renderización de EAO A_{EAO} se define de acuerdo con el número de canales de salida (es decir, mono, estéreo o binaural) de la siguiente manera:

$$\mathbf{A}_{EAO} = \begin{cases} \mathbf{A}_1^{EAO}, & \text{en el caso de mono} \\ \mathbf{A}_2^{EAO}, & \text{para otros casos} \end{cases}.$$

Las matrices A_1^{EAO} del tamaño 1 x N_{EAO} y A_2^{EAO} del tamaño 2x N_{EAO} se definen de la siguiente manera:

$$\mathbf{A}_1^{EAO} = \mathbf{D}_{16}^{EAO}\mathbf{M}_{ren}^{EAO}, \qquad \mathbf{D}_{16}^{EAO} = \left(w_1^{EAO} \quad w_2^{EAO} \quad w_3^{EAO} \quad w_3^{EAO} \quad w_1^{EAO} \quad w_2^{EAO}\right),$$

$$\mathbf{A}_2^{EAO} = \mathbf{D}_{26}^{EAO}\mathbf{M}_{ren}^{EAO}, \qquad \mathbf{D}_{26}^{EAO} = \begin{pmatrix} w_1^{EAO} & 0 & \frac{w_3^{EAO}}{\sqrt{2}} & \frac{w_3^{EAO}}{\sqrt{2}} & w_1^{EAO} & 0 \\ 0 & w_2^{EAO} & \frac{w_3^{EAO}}{\sqrt{2}} & \frac{w_3^{EAO}}{\sqrt{2}} & 0 & w_2^{EAO} \end{pmatrix},$$

25 donde la submatriz de renderización M_{res}^{EAO} corresponde a la renderización EAO (y describe un mapeo conveniente de los objetos de audio realzados sobre los canales de la representación de señal de mezcla ascendente).

Los valores W_1^{EAO} se computan dependiendo de la información de renderización asociada a los objetos de audio realzados usando los correspondientes elementos de EAO y utilizando las ecuaciones de la sección 4.2.2.1.

En el caso de la renderización binaural la matriz A_2^{EAO} está definida por las ecuaciones presentadas en la sección 4.1.2, para la cual la matriz binaural objetivo correspondiente contiene sólo elementos relacionados con los EAO.

3.4 Cálculo de los Elementos OTN/TTN en el Modo Residual

A continuación, se describe cómo la señal de mezcla descendente SAOC 310, que por lo general comprende uno o dos canales de audio, se mapea sobre la señal de objetos de audio realzados 334, que típicamente comprende uno o más canales de objetos de audio realzados, y la segunda información de audio 322, que por lo general comprende uno o dos canales de objetos de audio normales.

La funcionalidad de la unidad 1–a–N o la unidad 2–a–N 330 se puede implementar, por ejemplo, utilizando una multiplicación de vectores de matriz, de tal manera que se obtiene un vector que describe tanto los canales de la señal de objetos de audio realzados 334 y los canales de la segunda información de audio 322 multiplicando un vector que describe los canales de la señal de mezcla descendente SAOC 3 y (opcionalmente) una o más señales residuales con una matriz $M_{Predicción}$ o $M_{Energía}$.

En consecuencia, la determinación de la matriz $M_{Predicción}$ o $M_{Energía}$ es un paso importante en la derivación de la primera información de audio 320 y la segunda información de audio 322 de la mezcla descendente SAOC 310.

Para resumir, el proceso de mezcla descendente OTN/TTN se presenta en una matriz en el modo $M_{Predicción}$ en el caso del modo de predicción o $M_{Energía}$ en el caso del modo de energía.

El procedimiento de codificación/decodificación basado en la energía está destinado a la codificación de preservación no de forma de onda de la señal de mezcla descendente. Por consiguiente, la matriz de mezcla ascendente OTN/TTN para el correspondiente modo de energía no se basa en formas de onda específicas, pero sólo describen la distribución relativa de energía de los objetos de audio de entrada, como se describe más adelante en forma más detallada.

3.4.1 Modo de predicción

En el caso del modo de predicción, la matriz $M_{predicción}$ se define utilizando la información de mezcla descendente contenida en la matriz D^{-1} y los datos y los datos de CPC de la matriz C:

$M_{predicción}$ D^{1}

Con respecto a los varios modos SAOC, la matriz de mezcla descendente ampliada $\tilde{\mathbf{D}}$ y la matriz CPC **C** exhiben las siguientes dimensiones y estructuras:

3.4.1.1 Los modos de mezcla descendente estéreo (TTN): En el caso de los modos de mezcla descendente estéreo (TTN) (por ejemplo, en el caso de una mezcla descendente estéreo sobre la base de dos canales de objetos de audio normales y N_{EAO} canales de objetos de audio realzados), la matriz de mezcla descendente (ampliada) $\tilde{\mathbf{D}}$ y la matriz de CPC **C** se puede obtener de la siguiente manera:

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right),$$

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right).$$

En el caso de una mezcla descendente estéreo, cada EAO *j* tiene dos CPCs, $c_{j,0}$ y $c_{j,1}$, para dar la matriz C.

Las señales de salida del procesador residual se computan de la siguiente manera:

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Predicción}} \left(\begin{array}{c} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{array}\right),$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Predicción}} \left(\begin{array}{c} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{array}\right).$$

En consecuencia, se obtienen dos señales, y_L e y_R (que están representadas por X_{obj}), que representan uno o dos, o incluso más de dos objetos de audio normales (también denominados objetos de audio no ampliados)). Asimismo, las señales, N_{EAO} (representadas por X_{EAO}) que representan los N_{EAO} objetos de audio realzados. Estas señales se obtienen sobre la base de dos señales de mezcla descendente SAOC señal de mezcla descendentes $l_0 r_0$ y N_{EAO} o de señales residuales res_0 a $resN_{EAO-1}$, que son codificadas en la Información complementaria SAOC, por ejemplo, como parte de la información paramétrica relacionada con los objetos.

Se debe tener en cuenta que las señales y_L e y_R pueden ser equivalentes a la señal 322, y que las señales y_0 e yN_{EAO-1}, EAO (que están representadas por X_{EAO}) pueden ser equivalentes a las señales 320.

La matriz de A^{EAO} es una matriz de renderización. Los asientos de la matriz pueden describir, por ejemplo, un mapeo de objetos de audio realzados con los canales de las señales de objetos de audio realzados 334 (X_{EAO}).

En consecuencia, una elección apropiada de la matriz de A^{EAO} puede dar lugar a una integración opcional de la funcionalidad de la unidad de renderización 340, de tal manera que la multiplicación del vector que describe los canales (l_0,r_o) de la señal de mezcla descendente SAOC 310 y una o más señales residuales (res_0 y $resN_{EAO-1}$) con la matriz $\mathbf{A}^{EAO}\mathbf{M}_{EAO}^{Predicción}$ ~puede dar origen a una representación X_{EAO} de la primera información de audio 320.

3.4.1.2 Modos de mezcla descendente mono (OTN):

A continuación, se describe la derivación de las señales de objetos de audio realzados 320 (o, de lo contrario, de las señales de objetos de audio realzados 334) y de la señal de objetos de audio normales 322 con respecto al caso en que la señal de mezcla descendente SAOC 310 comprende sólo un canal de señal.

En el caso de los modos de mezcla descendente mono (OTN) (por ej., una mezcla descendente mono basada en un canal de objetos de audio normales y canales de objetos de audio realzados N_{EAO}), se puede obtener la matriz de mezcla descendente (ampliada) $\tilde{\mathbf{D}}$ y la matriz de CPC $\boldsymbol{C}$ de la siguiente manera:

$$\tilde{\mathbf{D}} = \left(\begin{array}{c|ccc} 1 & m_0 & \cdots & m_{N_{EAO}-1} \\ \hline m_0 & -1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right),$$

$$\mathbf{C} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1,0} & 0 & \cdots & 1 \end{array}\right).$$

Con una mezcla descendente mono mezcla descendente, un EAO *j* es previsto sólo por un coeficiente c_j para dar la matriz C. Todos los elementos de matriz c_j se obtienen, por ejemplo, de los parámetros SAOC (por ejemplo, de los datos SAOC 322 de acuerdo con las relaciones provistas más adelante en la sección 3.4.1.4

Las señales de salida del procesador residual se computan de la siguiente manera:

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Predicción}} \left(\begin{array}{c} d_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{array}\right),$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Predicción}} \left(\begin{array}{c} d_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{array}\right).$$

La señal de salida X_{OBJ} comprende, por ejemplo, un canal que describe los objetos de audio normales (objetos de audio no realzados) . La señal de salida $\mathbf{X}_{EAO}$ comprende, por ejemplo, uno, dos o incluso más canales que describen los objetos de audio realzados (preferentemente canales N_{EAO} que describen los objetos de audio realzados). Una vez más, dichas señales son equivalentes a las señales 320, 322.

3.4.1.3 Cálculo de la matriz de mezcla descendente ampliada inversa

La matriz $\tilde{\mathbf{D}}^{-1}$ es la inversa de la matriz de mezcla descendente ampliada $\tilde{\mathbf{D}}$ y C implica los CPCs.

La matriz $\tilde{\mathbf{D}}^{-1}$ es la inversa de la matriz de mezcla descendente ampliada $\tilde{\mathbf{D}}$ y se la puede calcular de la siguiente manera:

$$\tilde{\mathbf{D}}^{-1} = \frac{\tilde{d}_{i,j}}{den}.$$

Los elementos $\tilde{d}_{i,j}$ (por ejemplo, de la $\tilde{\mathbf{D}}^{-1}$ inversa de la matriz de mezcla descendente ampliada $\tilde{\mathbf{D}}$ de un tamaño de 6 x 6, se derivan utilizando los siguientes valores:

$$\tilde{d}_{1,1} = 1 + \sum_{j=1}^{4} n_j^2,$$

$$\tilde{d}_{1,2} = -\left(\sum_{j=1}^{4} m_j n_j \right),$$

$$\tilde{d}_{1,3} = m_1 + m_1 n_2^2 + m_1 n_3^2 + m_1 n_4^2 - m_2 n_1 n_2 - m_3 n_1 n_3 - m_4 n_1 n_4,$$

$$\tilde{d}_{1,4} = m_2 + m_2 n_1^2 + m_2 n_3^2 + m_2 n_4^2 - m_1 n_2 n_1 - m_3 n_2 n_3 - m_4 n_2 n_4,$$

$$\tilde{d}_{1,5} = m_3 + m_3 n_1^2 + m_3 n_2^2 + m_3 n_4^2 - m_1 n_3 n_1 - m_2 n_3 n_2 - m_4 n_3 n_4,$$

$$\tilde{d}_{1,6} = m_4 + m_4 n_1^2 + m_4 n_2^2 + m_4 n_3^2 - m_1 n_4 n_1 - m_2 n_4 n_2 - m_3 n_4 n_3,$$

$$\tilde{d}_{2,2} = 1 + \sum_{j=1}^{4} m_j^2,$$

$$\tilde{d}_{2,3} = n_1 + n_1 m_2^2 + n_1 m_3^2 + n_1 m_4^2 - m_1 m_2 n_2 - m_1 m_3 n_3 - m_1 m_4 n_4,$$

$$\tilde{d}_{2,4} = n_2 + n_2 m_1^2 + n_2 m_3^2 + n_2 m_4^2 - m_2 m_1 n_1 - m_2 m_3 n_3 - m_2 m_4 n_4,$$

$$\tilde{d}_{2,5} = n_3 + n_3 m_1^2 + n_3 m_2^2 + n_3 m_4^2 - m_3 m_1 n_1 - m_3 m_2 n_2 - m_3 m_4 n_4,$$

$$\tilde{d}_{2,6} = n_4 + n_4 m_1^2 + n_4 m_2^2 + n_4 m_3^2 - m_4 m_1 n_1 - m_4 m_2 n_2 - m_4 m_3 n_3,$$

$$\tilde{d}_{3,3} = -1 - \sum_{j=2}^{4} m_j^2 - \sum_{j=2}^{4} n_j^2 - m_2^2 n_2^2 - m_4^2 n_2^2 - m_2^2 n_3^2 - m_4^2 n_3^2 - m_2^2 n_4^2 - m_3^2 n_4^2 + 2m_2 m_3 n_2 n_3 + 2m_2 m_4 n_2 n_4 + 2m_3 m_4 n_3 n_4$$

$$,\tilde{d}_{3,4} = m_1 m_2 + n_1 n_2 + m_3^2 n_1 n_2 + m_4^2 n_1 n_2 + m_1 m_2 n_3^2 + m_1 m_2 n_4^2 - m_2 m_3 n_1 n_3 - m_1 m_3 n_2 n_3 - m_2 m_4 n_1 n_4 - m_1 m_4 n_2 n_4,$$

$$\tilde{d}_{3,5} = m_1 m_3 + n_1 n_3 + m_2^2 n_1 n_3 + m_4^2 n_1 n_3 + m_1 m_3 n_2^2 + m_1 m_3 n_4^2 - m_2 m_3 n_1 n_2 - m_1 m_2 n_2 n_3 - m_3 m_4 n_1 n_4 - m_1 m_4 n_3 n_4,$$

$$\tilde{d}_{3,6} = m_1 m_4 + n_1 n_4 + m_2^2 n_1 n_4 + m_3^2 n_1 n_4 + m_1 m_4 n_2^2 + m_1 m_4 n_3^2 - m_2 m_4 n_1 n_2 - m_3 m_4 n_1 n_3 - m_1 m_2 n_2 n_4 - m_1 m_3 n_3 n_4,$$

$$\tilde{d}_{4,4} = -1 - \sum_{\substack{j=1 \\ j\neq 2}}^{4} m_j^2 - \sum_{\substack{j=1 \\ j\neq 2}}^{4} n_j^2 - m_2^2 n_1^2 - m_4^2 n_1^2 - m_1^2 n_2^2 - m_4^2 n_2^2 - m_1^2 n_4^2 - m_2^2 n_4^2 + 2m_1 m_3 n_1 n_3 + 2m_1 m_4 n_1 n_4 + 2m_3 m_4 n_3 n_4,$$

$$\tilde{d}_{4,5} = m_2 m_3 + n_2 n_3 + m_1^2 n_2 n_3 + m_4^2 n_2 n_3 + m_2 m_3 n_1^2 + m_2 m_3 n_4^2 - m_1 m_3 n_1 n_2 - m_1 m_2 n_1 n_3 - m_3 m_4 n_2 n_4 - m_2 m_4 n_3 n_4,$$

$$\tilde{d}_{4,6} = m_2 m_4 + n_2 n_4 + m_1^2 n_2 n_4 + m_3^2 n_2 n_4 + m_2 m_4 n_1^2 + m_2 m_4 n_3^2 - m_1 m_4 n_1 n_2 - m_3 m_4 n_2 n_3 - m_1 m_2 n_1 n_4 - m_2 m_3 n_3 n_4,$$

$$\tilde{d}_{5,5} = -1 - \sum_{\substack{j=1 \\ j\neq 3}}^{4} m_j^2 - \sum_{\substack{j=1 \\ j\neq 3}}^{4} n_j^2 - m_2^2 n_1^2 - m_4^2 n_1^2 - m_1^2 n_2^2 - m_4^2 n_2^2 - m_1^2 n_4^2 - m_2^2 n_4^2 + 2m_1 m_2 n_1 n_2 + 2m_1 m_4 n_1 n_4 + 2m_2 m_4 n_2 n_4,$$

$$\tilde{d}_{5,6} = m_3 m_4 + n_3 n_4 + m_1^2 n_3 n_4 + m_2^2 n_3 n_4 + m_3 m_4 n_1^2 + m_3 m_4 n_2^2 - m_1 m_4 n_1 n_3 - m_2 m_4 n_2 n_3 - m_1 m_3 n_1 n_4 - m_2 m_3 n_2 n_4,$$

$$\tilde{d}_{6,6} = -1 - \sum_{j=1}^{3} m_j^2 - \sum_{j=1}^{3} n_j^2 - m_2^2 n_1^2 - m_3^2 n_1^2 - m_1^2 n_2^2 - m_3^2 n_2^2 - m_1^2 n_3^2 - m_2^2 n_3^2 + 2m_1 m_2 n_1 n_2 + 2m_1 m_3 n_1 n_3 + 2m_2 m_3 n_2 n_3,$$

$$den = 1 + \sum_{j=1}^{4} m_j^2 + \sum_{j=1}^{4} n_j^2 + m_2^2 n_1^2 + m_3^2 n_1^2 + m_4^2 n_1^2 + m_1^2 n_2^2 + m_3^2 n_2^2 + m_4^2 n_2^2 + m_1^2 n_3^2 + m_2^2 n_3^2 + m_4^2 n_3^2 + m_1^2 n_4^2 + m_2^2 n_4^2 +$$
$$+ m_3^2 n_4^2 - 2m_1 m_2 n_1 n_2 - 2m_1 m_3 n_1 n_3 - 2m_2 m_3 n_2 n_3 - 2m_1 m_4 n_1 n_4 - 2m_2 m_4 n_2 n_4 - 2m_3 m_4 n_3 n_4.$$

Los coeficientes m, y *n*, de la matriz de mezcla descendente ampliada $\tilde{\mathbf{D}}$ denotan los valores de mezcla descendente por cada EAO *j* correspondiente a los canales de mezcla descendente derecho e izquierdo según

$$m_j = d_{0,EAO(j)},\ n_j = d_{1,EAO(j)}.$$

Los elementos d_{ij} de la matriz de mezcla descendente $\tilde{\mathbf{D}}$ se obtienen utilizando la información de ganancia de mezcla descendente DMG y la información (opcional) de diferencias de nivel de los canales de mezcla descendente DCLD, que está incluida en la información de SAOC 332, que está representada, por ejemplo, por la información paramétrica relacionada con los objetos 110 o por la información de flujo de bits SAOC 212.

En el caso de la mezcla descendente estéreo la matriz de mezcla descendente **D** del tamaño 2 x N con los elementos $d_{i,j}$ (i = 0,1; j = 0,..., N – 1) se obtiene de los parámetros DMG y DCLD, a saber

$$d_{0,j} = 10^{0.05DMG_j}\sqrt{\frac{10^{0.1DCLD_j}}{1+10^{0.1DCLD_j}}}, \qquad d_{1,j} = 10^{0.05DMG_j}\sqrt{\frac{1}{1+10^{0.1DCLD_j}}}.$$

En el caso de la mezcla descendente mono mezcla descendente la matriz de mezcla descendente **D** del tamaño 1 x N con los elementos $d_{i,j}$ (i = 0,1; j = 0,..., N – 1) se obtiene de los parámetros DMG de la siguiente manera:

$d_{0,j} = 10^{0,05DMG\,j}$

En este caso, se obtienen los parámetros de mezcla descendente DMGj y DCLDj descuantificados, por ejemplo, de la información parametrica complementaria 110 o del flujo de bits SAOC 2 12.

La función EAO(j) determina el mapeo entre los índices de canales de objeto de audio de entrada y las señales de EAO:

$EAO_j = N - 1 - j \qquad j = 0,\ldots N_{EAO} - 1$

3.4.1.4 Cálculo de la matriz C

La matriz **C** incluye los CPCs y se deriva de los parámetros SAOC transmitidos (es decir, los OLDs, IOCs, DMGs y DCLDs) de la siguiente manera:

$$c_{j,0} = (1-\lambda)\tilde{c}_{j,0} + \lambda\gamma_{j,0}, \qquad c_{j,1} = (1-\lambda)\tilde{c}_{j,1} + \lambda\gamma_{j,1}.$$

Dicho de otro modo, se obtienen los CPCs de restricción de acuerdo con las ecuaciones anteriores, que se pueden considerar algoritmos de constricción. Sin embargo, los CPCs de restricción también se pueden derivar de los valores $\tilde{c}_{j,0}, \tilde{c}_{j,1}$ utilizando una estrategia de limitación diferente (algoritmo de restricción) o pueden ser fijados en valores iguales a $\tilde{c}_{j,0}, \tilde{c}_{j,1}$.

Se ha de notar que las anotaciones de la matriz $c_{j,1}$ (y las cantidades intermedias sobre la base de las cuales se computan las anotaciones de la matriz $c_{j,1}$), sólo son necesarias por lo general si la señal de mezcla descendente es una señal de mezcla descendente estéreo.

Los CPCs son restringidos por las siguientes funciones limitantes:

$$\gamma_{j,1} = \frac{m_j OLD_L + n_j e_{L,R} - \sum_{i=0}^{N_{EAO}-1} m_i e_{i,j}}{2\left(OLD_L + \sum_{i=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_i m_k e_{i,k}\right)}, \qquad \gamma_{j,2} = \frac{n_j OLD_R + m_j e_{L,R} - \sum_{i=0}^{N_{EAO}-1} n_i e_{i,j}}{2\left(OLD_R + \sum_{i=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} n_i n_k e_{i,k}\right)},$$

donde el factor λ se determina de la siguiente manera:

$$\lambda = \left(\frac{P_{LoRo}^2}{P_{Lo}P_{Ro}}\right)^{\delta}.$$

En el caso de un canal EAO específico $j = 0 \ldots N_{EAO} - 1$ los CPCs sin restricción son calculados por

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}, \qquad \tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}.$$

Las cantidades de energía P_{lo}, P_{ro}, P_{LoRo} , P_{LoCo} , $P_{LoRo,\,j}$, $P_{LoCo,\,j}$, se computan como

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k},$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k},$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k},$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i\neq j}}^{N_{EAO}-1} m_i e_{i,j},$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i\neq j}}^{N_{EAO}-1} n_i e_{i,j}.$$

La matriz de covarianza $e_{i,j}$, se define de la siguiente manera: La matriz de covarianza E del tamaño N x N con los elementos $e_{i,j}$, representa una aproximación de la matriz de covarianza de señales original $E \approx \mathbf{SS}^*$ y se la obtiene de los parámetros OLD y LOC de la siguiente manera

$$e_{i,j} = \sqrt{OLD_i OLD_j}\, IOC_{i,j}.$$

En este caso se obtienen los parámetros de objetos descuantificados OLD_i, $LOC_{i,j}$, por ejemplo de la información paramétrica complementaria 110 o del flujo de bits SAOC 212.

Además, se puede obtener $e_{L,R}$ de la siguiente manera

$$e_{L,R} = \sqrt{OLD_L OLD_R}\, IOC_{L,R}.$$

Los parámetros OLD_L, OLD_R y $IOC_{L,R}$ corresponden a los objetos (de audio) normales y se los puede derivar utilizando la información de mezcla descendente:

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & \text{de lo contrario} \end{cases}.$$

Como se puede ver, se computan dos valores diferentes de niveles de objeto OLD_L y OLD_R correspondientes a los objetos de audio normales en el caso de una señal de mezcla descendente estéreo (que implica preferentemente una

señal de objetos de audio normales de dos canales). Por el contrario, sólo se computa un valor de nivel de objeto diferente OLD_L para los objetos de audio normales en el caso de una señal de mezcla descendente (mono) monocanal (que preferentemente conlleva una señal de objetos de audio normales de de un canal).

Como se puede apreciar, se obtiene el primer (en el caso de una señal de mezcla descendente de dos canales o el único (en el caso de una señal de mezcla descendente de un solo canal) valor común de diferencia de niveles de los objetos OLD_L sumando las contribuciones de los objetos de audio normales que tienen un índice (o índices) de objetos de audio i correspondiente al canal izquierdo (o el único canal) de la señal de mezcla descendente SAOC 310.

El segundo valor común de diferencia de niveles de los objetos OLD_R (que se utiliza en el caso de una señal de mezcla descendente de dos canales) se obtiene sumando las contribuciones de los objetos de audio normales que tienen un índice (o índices) de objetos de audio i correspondiente al canal derecho de la señal de mezcla descendente SAOC 310.

La contribución OLD_L de los objetos de audio normales (que tienen los índices de objetos de audio $i = N-N_{EAO}-1$) sobre la señal del canal izquierdo (o la señal del único canal) de la señal de mezcla descendente SAOC 710 se computa, por ejemplo, tomando en cuenta la ganancia de mezcla descendente ganancia de mezcla descendente $d_{0,i}$, que describe la ganancia de mezcla descendente aplicada al objeto de audio normal que tiene el índice de objeto de audio i al obtener la señal del canal izquierdo de la señal de mezcla descendente SAOC 310, y también el nivel de objeto del objeto de audio normal que tiene el objeto de audio i, que está representado por el valor OLD_i.

De manera similar, el valor común de diferencia de nivel de los objetos OLD_R se obtiene usando los coeficientes de mezcla descendente $d_{1,i}$, que describen la ganancia de mezcla descendente que se aplica al objeto de audio normal que tiene el índice de objeto de audio i al formar la señal del canal izquierdo de la señal de mezcla descendente SAOC 310, y la información de nivel OLD_i asociada al objeto de audio normal que tiene el índice de objeto de audio i.

Como se puede ver, las ecuaciones para el cálculo de las cantidades P_{lo}, P_{ro}, P_{LoRo} , P_{LoCo} , $P_{LoRo,\ j}$ y $P_{LoCo,\ j}$, no distinguen entre los objetos de audio

normales individuales, sino meramente hacen uso de los valores comunes de diferencia de nivel de los objetos OLD_L, OLD_R, considerando así los objetos de audio normales (que tienen los índices de objeto de audio i) como un solo objeto de audio.

Además, se establece el valor de correlación entre objetos $IOC_{L,R}$ que está asociado a los objetos de audio normales, de 0 a menos que haya dos objetos de audio normales.

La matriz de covarianza $e_{i,j}$ (y $e_{L,R}$) se define de la siguiente manera:

La matriz de covarianza **E** del tamaño N x *N* con los elementos $e_{i,j}$ representan una aproximación de la matriz de covarianza de señal original $\mathbf{E} \approx \mathbf{SS}^*$ es obtenida de los parámetros OLD y IOC según

$$e_{i,j} = \sqrt{OLD_i OLD_j}\, IOC_{i,j}.$$

Por ejemplo

$$e_{L,R} = \sqrt{OLD_L OLD_R}\, IOC_{L,R}.$$

donde OLD_L y OLD_R y $IOC_{L,R}$ se computan de la manera antes descripta.

En este caso, los parámetros de objeto descuantificados se obtienen como

$OLD_i = \mathbf{D}_{OLD}(i,l,m)$, $IOC_{i,j,} = \mathbf{D}_{IOC}(i, j,l,m)$,

donde $\mathbf{D}_{OLD}$ y $\mathbf{D}_{IOC}$ son matrices que comprenden parámetros de diferencias de nivel de los objetos y parámetros de correlación entre objetos.

20 3.4.2. Modo de Energía

A continuación se describe otro concepto que se puede utilizar para separar las señales de objeto de audio ampliadas 320 y las señales de objetos de audio normales (objeto de audio no ampliado) 322 y que se puede utilizar en combinación con una codificación de audio sin conservación de la forma de onda de los canales SAOC de mezcla descendente 3 10.

Dicho de otro modo, el procedimiento de codificación/decodificación basado en la energía está destinado a la codificación de audio sin conservación de la forma de onda de la señal de mezcla descendente. Por consiguiente, la matriz de mezcla ascendente OTN/TTN para el correspondiente modo de energía no se basa en formas de onda específicas, sino que sólo describe la distribución relativa de energía de los objetos de audio de entrada.

Además, el concepto aquí explicado, que se designa concepto de "modo de energía", se puede utilizar sin transmitir una información de señal residual. Una

vez más, los objetos de audio normales (objetos de audio no realzados) son tratados en forma de objetos de audio de uno o dos canales que tienen uno o dos valores comunes de diferencia de nivel de los objetos, OLD_L, OLD_R.

En el caso del modo de energía, la matriz $M_{energía}$ se define utilizando la información de mezcla descendente y los OLDs, como se describe a continuación.

3.4.2.1. Modo de energía correspondiente a los Modos de Mezcla Descendente Estéreo (TTN)

En el caso de un estéreo (por ejemplo, una mezcla descendente estéreo basado en dos canales de objetos de audio normales y N_{EAO} canales de objetos de audio realzados), las matrices $\mathbf{M}_{OBJ}^{Energía}$ y $\mathbf{M}_{EAO}^{Energía}$ se obtienen de los OLDs correspondientes de acuerdo con

$$\mathbf{M}_{OBJ}^{Energía} = \begin{pmatrix} \sqrt{\dfrac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\dfrac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{Energía} = \begin{pmatrix} \sqrt{\dfrac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\dfrac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\dfrac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\dfrac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

El procesador residual señales de salida se computa de la siguiente manera:

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energía} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix},$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Energía} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}.$$

Lasa señales y_L e y_R, que están representadas por la señal X_{OBJ}, describen los objetos de audio normales (y pueden ser equivalentes a la señal 322), y las señales $y_{0,EAO}$ a $y_{NEAO-1,EAO}$, que están descriptas por la señal X_{EAO}, describen los

objetos de audio realzados (y pueden ser equivalentes a la señal 334 o a la señal 320).

Si se pretende una señal de mezcla ascendente mono para el caso de una señal de mezcla descendente estéreo, se puede efectuar un procesamiento de 2a 1, por ejemplo lo puede realizar el preprocesador 270 basándose en la señal bicanal X_{OBJ}.

3.4.2.2. Modo de energía correspondiente a Modos de Mezcla Descendente Mono (OTN)

En el caso mono (por ejemplo, una mezcla descendente mono sobre la base de un canal de objetos de audio normales y N_{EAO} canales de objetos de audio realzados), las matrices $\mathbf{M}_{OBJ}^{Energía}$ y $\mathbf{M}_{EAO}^{Energía}$ se obtienen a partir de los correspondientes OLDs de acuerdo con

$$\mathbf{M}_{OBJ}^{Energía} = \left(\sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right),$$

$$\mathbf{M}_{EAO}^{Energía} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}.$$

El procesador residual señales de salida se computa de la siguiente manera

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energía}\left(d_0\right),$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{Energía}\left(d_0\right).$$

Se puede obtener un canal único de objetos de audio normales 322 (representado por X_{OBJ}) y N_{EAO} canales de objetos de audio realzados 320 (representados por X_{EAO}) aplicando las matrices $\mathbf{M}_{OBJ}^{Energía}$ y $\mathbf{M}_{EAO}^{Energía}$ a una representación de una señal de mezcla descendente SAOC de canal único 310 (representado aquí por d_0).

Si se desea una señal de mezcla ascendente de dos canales (estéreo) para el caso de una señal de mezcla descendente de un canal (mono), se puede ejecutar un procesamiento de 1 a 2, por ejemplo mediante el preprocesador 270, sobre la base de una señal monocanal X_{OBJ}.

4. Arquitectura y operación del preprocesador de mezcla descendente SAOC

A continuación se describe el funcionamiento del preprocesador de mezcla descendente SAOC 270 tanto con respecto a algunos modos de decodificación como en el caso de los modos de operación de transcodificación.

4.1 Operación en los Modos de Decodificación

4.1.1 Introducción

A continuación se describe un método para obtener una señal de salida utilizando parámetros SAOC e información de paneo (o información de renderización) asociada con cada objeto de audio. El decodificador SAOC 495 está ilustrado en la Fig. 4g y consiste en el procesador de parámetros SAOC 496 y el procesador de mezcla descendente 497.

Se debe notar que se puede utilizar el decodificador SAOC 494 para procesar los objetos de audio normales, y por lo tanto recibir, como señal de mezcla descendente 497a, la segunda señal de objetos de audio 264 o la señal de objetos de audio normales 322 o la segunda información de audio 134.

En consecuencia, el procesador de mezcla descendente 497 puede producir, como sus señales de salida 497b, la versión procesada 272 de la segunda señal de objetos de audio 264 o la versión procesada 142 de la segunda información de audio 134. En consecuencia, el procesador de mezcla descendente 497 puede asumir el rol del preprocesador de mezcla descendente SAOC 270, o el rol del procesador de señales de audio 140.

El procesador de parámetros SAOC 496 puede asumir la función del procesador de parámetros SAOC 252 y, en consecuencia, proporciona la información de mezcla descendente 496a.

4.1.2 Procesador de mezcla descendente

A continuación se describe en forma más detallada el procesador de mezcla descendente, que es parte del procesador de señales de audio 140, y que ha sido

designado "preprocesador de mezcla descendente SAOC" 270 en la realización de la Fig. 2, y que ha sido designada 497 en el decodificador SAOC 495.

En el caso del modo de decodificador del sistema SAOC, la señal de salida 142, 272, 497b del procesador de mezcla descendente (representado en el dominio híbrido QMF) es alimentada al correspondiente banco de filtros de síntesis (no se ilustra en las Figs. 1 y 2) de acuerdo con lo descripto en ISO/IEC 23003–1: 2007 para dar la señal PCM de salida final. Sin embargo, la señal de salida 142, 272, 497b del procesador de mezcla descendente se combina por lo general con una o más señales de audio 132, 262 que representan los objetos de audio realzados. Esta combinación se puede realizar antes del banco de filtro de síntesis correspondiente (de tal manera que se transmita una señal combinada que combina la salida del procesador de mezcla descendente y dichas una o más señales que representan los objetos de audio realzados como entrada al banco de filtros de síntesis). Por otro lado, se puede combinar la salida del procesador de mezcla descendente con una o más señales de audio que representan los objetos de audio realzados sólo después del procesamiento en el banco de filtros de. En consecuencia, la representación de señal de mezcla ascendente 120, 220 puede ser una representación en el dominio QMF o una representación en el dominio PCM (o cualquier otra representación apropiada). El procesamiento de mezcla descendente incorpora, por ejemplo, el procesamiento mono, el procesamiento en estéreo y, si se necesita, el posterior procesamiento binaural.

La señal de salida $\hat{\mathbf{X}}$ del procesador de mezcla descendente 270, 497 (también designada con los números 142, 272, 497b) se computa a partir de la señal de mezcla descendente mono X (también indicada con 134, 264, 497a) y la señal de mezcla descendente mono descorrelacionada $\mathbf{X}_d$ según

$$\hat{\mathbf{X}} = \mathbf{G}\mathbf{X} + \mathbf{P}_2\mathbf{X}_d.$$

La señal de mezcla descendente mono descorrelacionada $\mathbf{X}_d$ se computa de la siguiente manera

X_d = Func. descor (**X**).

Las señales descorrelacionadas X_d se generan a partir de descorrelacionador descripto en ISO/IEC 23003–1:2007, subcláusula 6.6.2. Siguiendo este esquema, se debe emplear la configuración bsDecorrConfig == 0 con un índice de descorrelación, X = 8, de acuerdo con la Tabla A.26 a la Tabla

A.29 de ISO/IEC 23003–1:2007. Por ende, la *decorrFunc( )* denota el proceso de descorrelación:

$$\mathbf{X}_d = \begin{pmatrix} x_{1d} \\ x_{2d} \end{pmatrix} = \begin{pmatrix} decorrFunc\left((1 \quad 0)\mathbf{P}_1\mathbf{X}\right) \\ decorrFunc\left((0 \quad 1)\mathbf{P}_1\mathbf{X}\right) \end{pmatrix}.$$

En el caso de la salida binaural se aplican los parámetros de mezcla ascendente **G** y $\mathbf{P}_2$ derivados de los datos de SAOC, la información de renderización $\mathbf{M}_{ren}^{l,m}$ y los parámetros HRTF a la señal de mezcla descendente **X** (y $\mathbf{X}_d$) para dar origen a la salida binaural $\hat{\mathbf{X}}$, ver la Fig. 2, número de referencia 270, donde se ilustra la estructura básica del procesador de mezcla descendente.

La matriz de renderización binaural objetivo $A^{l,m}$ del tamaño 2xN consiste en los elementos $a_{x,y}^{l,m}$. Cada elemento $a_{x,y}^{l,m}$ deriva de los parámetros HRTF y la matriz de renderización $\mathbf{M}_{ren}^{l,m}$ con los elementos $m_{y,i}^{l,m}$, por ejemplo, por el procesador de parámetros SAOC. La matriz de renderización binaural objetivo $A^{l,m}$ representa la relación entre todos los objetos de audio de entrada y la salida binaural pretendida.

$$a_{y,1}^{l,m} = \sum_{i=0}^{N_{input}-1} m_{y,i}^{l,m} H_{i,L}^{m} \exp\left(j\frac{\phi_i^m}{2}\right), \qquad a_{y,2}^{l,m} = \sum_{i=0}^{N_{input}-1} m_{y,i}^{l,m} H_{i,R}^{m} \exp\left(-j\frac{\phi_i^m}{2}\right).$$

Los parámetros HRTF están dados por $H_{i,L}^m$, $H_{i,R}^m$ y ϕ_i^m por cada banda de procesamiento *m*. Las posiciones espaciales para las cuales se dispone de parámetros HRTF se caracterizan por el índice i. Estos parámetros han sido descriptos en ISO/IEC 23003–1:2007.

4.1.2.1 Reseña general

A continuación, se presenta una reseña general con respecto al procesamiento de mezcla descendente con referencia a las Figs. 4a y 4b, que ilustran una representación de bloques del procesamiento de mezcla descendente, que puede ser ejecutado por el procesador de señales de audio 140 o por la combinación de procesador de parámetros SAOC 252 y el preprocesador de mezcla descendente SAOC 270, o por la combinación del procesador de parámetros SAOC 496 y el procesador de mezcla descendente 497.

Haciendo referencia, ahora, a la Fig. 4a, el procesamiento de mezcla descendente recibe una matriz de renderización M, una información sobre diferencias de nivel de los objetos OLD, una información sobre correlación entre objetos IOC, una información de ganancia de mezcla descendente DMG y

(opcionalmente) una información sobre diferencias de nivel de canales de mezcla descendente DCLD. El procesamiento de mezcla descendente 400 de acuerdo con la Fig. 4a obtiene una matriz de renderización A sobre la base de la matriz de renderización M, por ejemplo, usando un ajustador de parámetros y un mapeo de M con A. Además, las anotaciones de una matriz de covarianza E se obtienen dependiendo de la información sobre diferencias de nivel de los objetos OLD y la información sobre correlación entre objetos IOC, por ejemplo, de acuerdo conde acuerdo con lo descripto anteriormente. De manera similar, las anotaciones de una matriz de mezcla descendente D se obtienen dependiendo de la información sobre ganancia de mezcla descendente DMG y la información sobre diferencias de nivel de los canales de mezcla descendente DCLD.

Las anotaciones f de una matriz de covarianza conveniente F se obtienen dependiendo de la matriz de renderización A y la matriz de covarianza E. Asimismo, se obtiene un valor escalar v que depende de la matriz de covarianza E y la matriz de mezcla descendente D (o dependiendo de las anotaciones de las mismas).

Los valores de ganancia P_L y P_R correspondientes a dos canales se obtienen dependiendo de las anotaciones de la matriz de covarianza pretendida F y el valor escalar v. Además, se obtiene un valor de diferencias de fases entre canales φ_C dependiendo de las anotaciones f de la matriz de covarianza pretendida F. También se obtiene un ángulo de rotación α que depende de las anotaciones f de la matriz de covarianza pretendida F, tomando en cuenta, por ejemplo, una constante c. Por añadidura, se obtiene un segundo ángulo de rotación β, dependiendo de las ganancias de canales P_L y P_R c y del primer ángulo de rotación α. Las anotaciones de una matriz **G** se obtienen, por ejemplo, dependiendo de los dos valores de ganancia de los canales P_L, P_R y dependiendo asimismo de la diferencia de fases entre canales φ_C y, opcionalmente, de los ángulos de rotación α, β. De manera similar, las anotaciones de una matriz $\mathbf{P_2}$ se determinan dependiendo de algunos o todos dichos valores P_L, P_R, φ_C, α, β.

A continuación se describe la manera en que se puede obtener la matriz **G** y/o **P2** (o las anotaciones de las mismas), que se puede aplicar con los diferentes modos de procesamiento.

4.1.2.2 Modo de Procesamiento Mono a Binaural "x–1–b"

A continuación se describe un modo de procesamiento en el cual los objetos de audio normales están representados por las señales de mezcla descendente monocanales 134, 264, 322, 497a y en las cuales se busca la renderización binaural.

Los parámetros de mezcla ascendente $G^{l,m}$ y $P_2^{l,m}$ se computan de la siguiente manera

$$\mathbf{G}^{l,m} = \begin{pmatrix} P_L^{l,m} \exp\left(j\frac{\phi_C^{l,m}}{2}\right)\cos\left(\beta^{l,m}+\alpha^{l,m}\right) \\ P_R^{l,m} \exp\left(-j\frac{\phi_C^{l,m}}{2}\right)\cos\left(\beta^{l,m}-\alpha^{l,m}\right) \end{pmatrix},$$

$$\mathbf{P}_2^{l,m} = \begin{pmatrix} P_L^{l,m} \exp\left(j\frac{\phi_C^{l,m}}{2}\right)\sin\left(\beta^{l,m}+\alpha^{l,m}\right) \\ P_R^{l,m} \exp\left(-j\frac{\phi_C^{l,m}}{2}\right)\sin\left(\beta^{l,m}-\alpha^{l,m}\right) \end{pmatrix}.$$

Las ganancias $P_L^{l,m}$ y $P_R^{l,m}$ correspondientes a los canales de salida derecho e izquierdo son

$$P_L^{l,m} = \sqrt{\max\left(\frac{f_{1,1}^{l,m}}{v^{l,m}}, \varepsilon^2\right)}, \qquad P_R^{l,m} = \sqrt{\max\left(\frac{f_{2,2}^{l,m}}{v^{l,m}}, \varepsilon^2\right)}.$$

La matriz de covarianza pretendida $F^{l,m}$ del tamaño 2x2 con los elementos $f_{i,j}^{l,m}$ se expresan como

$$\mathbf{F}^{l,m} = \mathbf{A}^{l,m}\mathbf{E}^{l,m}\left(\mathbf{A}^{l,m}\right)^*.$$

El escalar $v^{l,m}$ se computa de la siguiente manera

$$v^{l,m} = \mathbf{D}^l\mathbf{E}^{l,m}\left(\mathbf{D}^l\right)^* + \varepsilon^2.$$

La diferencia de fases entre canales $\phi_C^{l,m}$ se expresa como

$$\phi_C^{l,m} = \begin{cases} \arg\left(f_{1,2}^{l,m}\right), & 0 \le m \le 11, \quad \rho_C^{l,m} \ge 0.6, \\ 0, & \text{de otro modo}. \end{cases}$$

La coherencia entre canales $\rho_C^{l,m}$ se computa según

$$\rho_C^{l,m} = \min\left(\frac{\left|f_{1,2}^{l,m}\right|}{\sqrt{\max\left(f_{1,1}^{l,m}f_{2,2}^{l,m}, \varepsilon^2\right)}}, 1\right).$$

Los ángulos de rotación $\alpha^{l,m}$ y $\beta^{l,m}$ se expresan como

$$\alpha^{l,m} = \begin{cases} \frac{1}{2}\arccos\left(\rho_C^{l,m}\cos\left(\arg\left(f_{1,2}^{l,m}\right)\right)\right), & 0 \le m \le 11, \quad \rho_C^{l,m} < 0.6, \\ \frac{1}{2}\arccos\left(\rho_C^{l,m}\right), & \text{de otro modo.} \end{cases}$$

$$\beta^{l,m} = \arctan\left(\tan\left(\alpha^{l,m}\right)\frac{P_R^{l,m} - P_L^{l,m}}{P_L^{l,m} + P_R^{l,m} + \varepsilon}\right).$$

4.1.2.3 Modo de Procesamiento Mono a Estéreo "x–1–2"

A continuación se describe un modo de procesamiento en el cual los objetos de audio normales están representados por la señal monocanal 134, 264, 222, y en los cuales se desea una renderización estéreo.

En el caso de la salida estéreo se puede aplicar el modo de procesamiento "x–1–b" sin utilizar la información HRTF. Esto se puede efectuar derivando todos los elementos $a_{x,y}^{l,m}$ de la matriz de renderización **A**, para dar:

$$a_{1,y}^{l,m} = m_{Lf,y}^{l,m}, \qquad a_{2,y}^{l,m} = m_{Rf,y}^{l,m}.$$

4.1.2.4 Modo de Procesamiento mono a mono "x–1– 1"

A continuación se describe un modo de procesamiento en el cual los objetos de audio normales están representados por un canal de señal 134, 264, 322, 497a y en el cual se busca una renderización de dos canales de los objetos de audio normales.

En el caso de la salida mono, se puede aplicar el modo de procesamiento "x–1–2" con las siguientes anotaciones:

$$a_{1,y}^{l,m} = m_{C,y}^{l,m}, \quad a_{2,y}^{l,m} = 0$$

4.1.2.5 Modo de procesamiento de Estéreo a binaural "x–2–b"

A continuación, se describe un modo de procesamiento en el cual los objetos de audio normales están representados por una señal de canales 134, 264, 322, 497a, y en el cual se pretende una renderización binaural de los objetos de audio normales.

Los parámetros de mezcla ascendente $\mathbf{G}^{l,m}$ y $\mathbf{P}_2^{l,m}$ se computan de la siguiente manera:

$$\mathbf{G}^{l,m}=\begin{pmatrix} P_L^{l,m,1}\exp\left(j\frac{\phi^{l,m,1}}{2}\right)\cos\left(\beta^{l,m}+\alpha^{l,m}\right) & P_L^{l,m,2}\exp\left(j\frac{\phi^{l,m,2}}{2}\right)\cos\left(\beta^{l,m}+\alpha^{l,m}\right) \\ P_R^{l,m,1}\exp\left(-j\frac{\phi^{l,m,1}}{2}\right)\cos\left(\beta^{l,m}-\alpha^{l,m}\right) & P_R^{l,m,2}\exp\left(-j\frac{\phi^{l,m,2}}{2}\right)\cos\left(\beta^{l,m}-\alpha^{l,m}\right) \end{pmatrix},$$

$$\mathbf{P}_2^{l,m}=\begin{pmatrix} P_L^{l,m}\exp\left(j\frac{\arg\left(c_{1,2}^{l,m}\right)}{2}\right)\sin\left(\beta^{l,m}+\alpha^{l,m}\right) \\ P_R^{l,m}\exp\left(-j\frac{\arg\left(c_{1,2}^{l,m}\right)}{2}\right)\sin\left(\beta^{l,m}-\alpha^{l,m}\right) \end{pmatrix}.$$

$P_L^{l,m,x}$, $P_R^{l,m,x}$ y $P_L^{l,m}$, $P_R^{l,}$

Las ganancias correspondientes, a los canales de salida izquierdo y derecho son

$$P_L^{l,m,x}=\sqrt{\max\left(\frac{f_{1,1}^{l,m,x}}{v^{l,m,x}},\varepsilon^2\right)},\quad P_R^{l,m,x}=\sqrt{\max\left(\frac{f_{2,2}^{l,m,x}}{v^{l,m,x}},\varepsilon^2\right)}$$

$$P_L^{l,m}=\sqrt{\max\left(\frac{c_{1,1}^{l,m}}{v^{l,m}},\varepsilon^2\right)},\qquad P_R^{l,m}=\sqrt{\max\left(\frac{c_{2,2}^{l,m}}{v^{l,m}},\varepsilon^2\right)}.$$

La matriz de covarianza pretendida $F^{1,m,x}$ del tamaño 2x2 con los elementos $f_{u,v}^{l,m,x}$ se expresan de la siguiente manera

$$\mathbf{F}^{l,m,x}=\mathbf{A}^{l,m}\mathbf{E}^{l,m,x}\left(\mathbf{A}^{l,m}\right)^*.$$

La matriz de covarianza $\boldsymbol{C}^{1,m}$ del tamaño 2 x 2 con los elementos $c_{u,v}^{l,m}$ de la señal binaural "seca" se estima de la siguiente manera

$$\mathbf{C}^{l,m}=\tilde{\mathbf{G}}^{l,m}\mathbf{D}^{l}\mathbf{E}^{l,m}\left(\mathbf{D}^{l}\right)^*\left(\tilde{\mathbf{G}}^{l,m}\right)^*,$$

donde

$$\tilde{\mathbf{G}}^{l,m}=\begin{pmatrix} P_L^{l,m,1}\exp\left(j\frac{\phi^{l,m,1}}{2}\right) & P_L^{l,m,2}\exp\left(j\frac{\phi^{l,m,2}}{2}\right) \\ P_R^{l,m,1}\exp\left(-j\frac{\phi^{l,m,1}}{2}\right) & P_R^{l,m,2}\exp\left(-j\frac{\phi^{l,m,2}}{2}\right) \end{pmatrix}.$$

Los correspondientes escalares $v^{1,m,x}$ y $v^{1,m}$ se computan según

$$v^{l,m,x}=\mathbf{D}^{l,x}\mathbf{E}^{l,m}\left(\mathbf{D}^{l,x}\right)^*+\varepsilon^2,\quad v^{l,m}=\left(\mathbf{D}^{l,1}+\mathbf{D}^{l,2}\right)\mathbf{E}^{l,m}\left(\mathbf{D}^{l,1}+\mathbf{D}^{l,2}\right)^*+\varepsilon^2.$$

La matriz de mezcla descendente $D^{l,x}$ del tamaño 1 x *N* con los elementos $d_1^{l,m}$ se puede encontrar como

$$d_i^{l,1} = 10^{0.05\,DMG_i^l}\sqrt{\frac{10^{0.1\,DCLD_i^l}}{1+10^{0.1\,DCLD_i^l}}}, \qquad d_i^{l,2} = 10^{0.05\,DMG_i^l}\sqrt{\frac{1}{1+10^{0.1\,DCLD_i^l}}}.$$

La matriz de mezcla descendente estéreo $\mathbf{D}^1$ del tamaño 2 x N con los elementos $d^1_{x,j}$ se pueden encontrar como $d^l_{x,i} = d_i^{l,x}$.

La matriz $E^{l,m,,x}$ con los elementos $e_i^{l,m,x}$ se pueden obtener según

$$e_{i,j}^{l,m,x} = e_{i,j}^{l,m}\left(\frac{d_i^{l,x}}{d_i^{l,1}+d_i^{l,2}}\right)\left(\frac{d_j^{l,x}}{d_j^{l,1}+d_j^{l,2}}\right).$$

Las diferencias de fase entre canales $\phi_C^{l,m}$ se expresan de la siguiente manera:

$$\phi^{l,m,x} = \begin{cases} \arg\left(f_{1,2}^{l,m,x}\right), & 0 \le m \le 11, \quad \rho_C^{l,m} > 0.6, \\ 0, & \text{de otro modo.} \end{cases}$$

Los ICCs $\rho_C^{l,m}$ y $\rho_T^{l,m}$ se computan como

$$\rho_T^{l,m} = \min\left(\frac{\left|f_{1,2}^{l,m}\right|}{\sqrt{\max\left(f_{1,1}^{l,m}f_{2,2}^{l,m},\varepsilon^2\right)}},1\right), \qquad \rho_C^{l,m} = \min\left(\frac{\left|c_{1,2}^{l,m}\right|}{\sqrt{\max\left(c_{1,1}^{l,m}c_{2,2}^{l,m},\varepsilon^2\right)}},1\right).$$

Los ángulos de rotación $\alpha^{l,m}$ y $\beta^{l,m}$ se expresan como

$$\alpha^{l,m} = \frac{1}{2}\left(\arccos\left(\rho_T^{l,m}\right) - \arccos\left(\rho_C^{l,m}\right)\right), \qquad \beta^{l,m} = \arctan\left(\tan(\alpha^{l,m})\frac{P_R^{l,m}-P_L^{l,m}}{P_L^{l,m}+P_R^{l,m}}\right).$$

4.1.2.6 Modo de procesamiento Estéreo a estéreo "x–2–2"

A continuación se describe un modo de procesamiento en el cual los objetos de audio normales son descriptos por una señal de dos canales (estéreo) 134, 264, 322, 497a y en el cual se pretende una renderización de 2 canales (estéreo).

En el caso de la salida estéreo, se aplica directamente el preprocesamiento estéreo, que se describe más adelante en la Sección 4.2.2.3.

4.1.2.7 Modo de procesamiento Estéreo a mono "x–2–1"

A continuación se describe un modo de procesamiento en el cual los objetos de audio normales están representados por una señal de dos canales (estéreo) 134, 264, 322, 497a y en el cual se pretende una renderización de un canal (mono).

En el caso de la salida mono, se aplica el preprocesamiento estéreo con una sola anotación activa de la matriz de renderización, como se describe más adelante en la Sección 4.2.2.3.

4.1.2.8 Conclusión

Tomando como referencia, una vez más, las Figs. 4a y 4b, se describe un procesamiento que se puede aplicar a una señal de 1 canal o una señal de dos canales 134, 264, 322,497a que representa los objetos de audio normales con posterioridad a la separación entre los objetos de audio ampliados y los objetos de audio normales. Las Figs. 4a y 4b ilustran el procesamiento, donde el procesamiento de las Figs. 4a y 4b difiere en que se introduce un ajuste opcional de los parámetros en diferentes etapas del procesamiento.

4.2. Operación en los modos de transcodificación

4.2.1 Introducción

A continuación se explica un método para combinar parámetros SAOC e información de paneo (o información de renderización) asociada a cada objeto de audio (o, preferentemente, a cada objeto de audio normal) en un flujo de bits MPEG envolvente compatible standard (flujo de bits MPS).

El transcodificador SAOC 490 está ilustrado en la Fig. 4f y consiste en un procesador de parámetros SAOC 491 y un procesador de mezcla descendente 492 aplicados a una mezcla descendente estéreo.

El transcodificador SAOC 490 puede asumir, por ejemplo, la funcionalidad del procesador de señales de audio 140. Por otro lado, el transcodificador SAOC 490 puede asumir la funcionalidad del preprocesador de mezcla descendente SAOC 270 tomado en combinación con el procesador de parámetros SAOC 252.

Por ejemplo, el procesador de parámetros SAOC 491 puede recibir un flujo de bits SAOC 491a, que es equivalente a la información paramétrica relacionada con los objetos 110 o al flujo de bits SAOC 212. Además, el procesador de parámetros SAOC 491 puede recibir una información de matriz de renderización 491 que puede estar incluida en la información paramétrica relacionada con los objetos 110, o que puede ser equivalente a la información de matriz de renderización 214. El procesador de parámetros SAOC 491 puede producir asimismo la información de procesamiento de mezcla descendente 491c al procesador de mezcla descendente 492, que puede ser equivalente a la

información 240. Más aun, el procesador de parámetros SAOC 491 puede producir un flujo de bits MPEG envolvente (o un flujo de bits de parámetros MPEG envolvente) 491d, que comprende una información paramétrica de envolvente que es compatible con la norma sobre MPEG envolvente. El flujo de bits MPEG envolvente 491d puede ser parte, por ejemplo, de la versión procesada 142 de la segunda información de audio, o puede ser parte o tomar el lugar, por ejemplo, del flujo de bits MPS 222.

El procesador de mezcla descendente 492 está configurado para recibir una señal de mezcla descendente 492a, que es preferentemente una señal de mezcla descendente de un canal o una señal de mezcla descendente de dos canales, y que preferentemente es equivalente a la segunda información de audio 134, o a la segunda señal de objetos de audio 264, 322. El procesador de mezcla descendente 492 puede producir asimismo una señal de mezcla descendente MPEG envolvente 492b, que es equivalente (o forma parte de) la versión procesada 142 de la segunda información de audio 134, o es equivalente (o es parte de) la versión procesada 272 de la segunda señal de objetos de audio 264.

Sin embargo, hay diferentes maneras de combinar la señal de mezcla descendente MPEG envolvente 492b con la señal de objetos de audio realzados 132, 262. La combinación se puede realizar en el dominio MPEG envolvente.

Por otro lado, no obstante, la representación MPEG envolvente, que comprende el flujo de bits de parámetros MPEG envolvente 491d y la señal de mezcla descendente MPEG envolvente 492b, de los objetos de audio normales se puede convertir de vuelta a una representación en el domino del tiempo multicanal o una representación multicanal en el dominio de la frecuencia (que individualmente representan diferentes canales de audio) por un decodificador MPEG envolvente y seguidamente se la puede combinar con las señales de objetos de audio realzados.

Se ha de notar que los modos de transcodificación comprenden tanto uno o más modos de procesamiento de mezcla descendente mono como uno o más modos de procesamiento de mezcla descendente estéreo. Sin embargo, a continuación sólo se describe el modo de procesamiento de mezcla descendente estéreo, puesto que el procesamiento de las señales de objetos de audio normales es más elaborado en el modo de procesamiento de mezcla descendente estéreo.

4.2.2 Procesamiento de mezcla descendente; en el modo de procesamiento de mezcla descendente estéreo ("x–2–5")

4.2.2.1 Introducción

En la siguiente sección se presenta una descripción del modo de transcodificación SAOC correspondiente al caso de la mezcla descendente estéreo.

Los parámetros de objetos (diferencia de nivel de los objetos OLD, correlación entre objetos IOC, la ganancia de mezcla descendente DMG y la diferencias de nivel de canales de mezcla descendente DCMD) del flujo de bits SAOC se transcodifican en parámetros espaciales (preferentemente relacionados con los canales) (diferencias de nivel de los canales CLD, correlación entre canales ICC, coeficiente de predicción de canales CPC) correspondientes al flujo de bits MPEG envolvente de acuerdo con la información de renderización. La mezcla descendente se modifica de acuerdo con los parámetros de objetos y una matriz de renderización.

Haciendo referencia, ahora, a las Figs. 4c, 4d y 4e, se presenta una reseña general del procesamiento, y en particular de la modificación de la mezcla descendente.

La Fig. 4c ilustra una representación en bloques de un procesamiento que se ejecuta para modificar la señal de mezcla descendente, por ejemplo la señal de mezcla descendente 134, 264, 322,492a que describe uno o, preferentemente, más objetos de audio normales. Como se puede apreciar en las Figs. 4c, 4d y 4e, el procesamiento recibe una matriz de renderización M_{ren}, una información de ganancia de mezcla descendente DMG, una información de diferencias de nivel de los canales de mezcla descendente DCLD, una información de diferencias de nivel de los objetos OLD, y una información de correlación entre objetos IOC. La matriz de renderización puede ser modificada opcionalmente mediante un ajuste de parámetros, como se ilustra en la Fig. 4c. Las anotaciones de una matriz de mezcla descendente D se obtienen dependiendo de la información de ganancia de mezcla descendente DMG y la información de diferencias de nivel de los canales de mezcla descendente DCLD. Las anotaciones de una matriz de coherencia E se obtienen dependiendo de la información de diferencias de nivel de los objetos OLD y la información de correlación entre objetos IOC. Además, se puede obtener una

matriz J dependiendo de la matriz de mezcla descendente D y de la matriz de coherencia E, o supeditada a las anotaciones de las mismas. Seguidamente se puede obtener una matriz C_3 dependiendo de la matriz de renderización M_{ren}, la matriz de mezcla descendente D, la matriz de coherencia E y la matriz J. Se puede obtener una matriz **G** dependiendo de una matriz D_{TTT}, que puede ser una matriz que consta de anotaciones predeterminadas, y que también depende de la matriz **C3.** La matriz **G** puede ser modificada, opcionalmente, para obtener una matriz modificada $\mathbf{G}_{mod}$. Se puede utilizar la matriz **G** o la versión modificada $\mathbf{G}_{mod}$ de la misma para derivar la versión procesada 142, 272,492b de la segunda información de audio 134, 264 a partir de la segunda información de audio 134, 264,492a (donde la segunda información de audio 134, 264 se designa X y donde la versión procesada 142,272 está indicada con $\hat{\mathbf{X}}$.

A continuación se describe la renderización de la energía de los objetos, que es ejecutada a fin de obtener los parámetros MPEG envolvente. Además, el preprocesamiento estéreo, que es ejecutado a fin de obtener la versión procesada 142, 272,492b de la segunda información de audio 134,264,492a que representa los objetos de audio normales.

4.2.2.2 Renderización de energías de los objetos

El transcodficador determina los parámetros para el decodificador MPS de acuerdo con la renderización objetivo según lo descripto por la matriz de renderización $\mathbf{M}_{ren}$. La covarianza objetivo de seis canales está indicada con **F** y dada por

$$\mathbf{F} = \mathbf{YY}^* = \mathbf{M}_{ren}\mathbf{S}(\mathbf{M}_{ren}\mathbf{S})^* =^* \mathbf{M}_{ren}(\mathbf{SS}^*)\mathbf{M}^+_{ren} = \mathbf{M}_{ren}\mathbf{EM}^+_{ren}.$$

El proceso de transcodificación se puede dividir conceptualmente en dos partes. En una parte, se ejecuta una renderización de tres canales al canal izquierdo, derecho y central. En esta etapa, se obtienen los parámetros correspondientes a la modificación de la mezcla descendente, como así también los parámetros de predicción correspondientes a la caja TTT para el decodificador MPS. En la otra parte se determinan los parámetros CLD e ICC para la renderización entre los canales anterior y envolvente (parámetros OTT, anterior izquierdo –envolvente izquierdo, anterior derecho –envolvente derecho).

4.2.2.2.1 Renderización al canal izquierdo, derecho y central

En esta etapa se determinan los parámetros espaciales que controlan la renderización a los canales izquierdo y derecho, que consisten en las señales anterior y envolvente. Estos parámetros describen la matriz de predicción de la caja TTT para la decodificación MPS $\mathbf{C}_{TTT}$ (parámetros CPC para el decodificador MPS) y la matriz del convertidor de mezcla descendente **G**.

$\mathbf{C}_{TTT}$ es la matriz de predicción para obtener la renderización objetivo a partir de la mezcla descendente modificada.

$\hat{\mathbf{X}} = \mathbf{GX}$:

$$\mathbf{C}_{TTT}\hat{\mathbf{X}} = \mathbf{C}_{TTT}\mathbf{GX} \approx \mathbf{A}_3\mathbf{S}.$$

$\mathbf{A}_3$ es una matriz de renderización reducida del tamaño 3 x *N*, que describe la renderización al canal izquierdo, derecho y central, respectivamente. Se la obtiene en términos de $\mathbf{A}_3 = \mathbf{D}_{36}\mathbf{M}_{ren}$, donde la matriz de mezcla descendente parcial 6 a 3 $\mathbf{D}_{36}$ se define según

$$\mathbf{D}_{36} = \begin{pmatrix} w_1 & 0 & 0 & 0 & w_1 & 0 \\ 0 & w_2 & 0 & 0 & 0 & w_2 \\ 0 & 0 & w_3 & w_3 & 0 & 0 \end{pmatrix}.$$

Los pesos parciales de mezcla descendente $\mathbf{w}_p$ p = 1,2,3 se ajustan de tal manera que la energía de $w_p(y_{2p-1} + y_{2p})$ es igual a la suma de las energías $\|y_{2p-1}\|^2 + \|y_{2p}\|$ hasta un factor límite.

donde

$$w_1 = \frac{f_{1,1} + f_{5,5}}{f_{1,1} + f_{5,5} + 2f_{1,5}}, \qquad w_2 = \frac{f_{2,2} + f_{6,6}}{f_{2,2} + f_{6,6} + 2f_{2,6}}, \qquad w_3 = 0.5$$

donde $f_{i,j}$ indica los elementos de **F**

Para la estimación de la matriz de predicción pretendida $\mathbf{C}_{TTT}$ y la matriz de procesamiento de mezcla descendente **G** definimos una matriz de predicción $\mathbf{C}_3$ del tamaño 3 x 2, que conduce a la renderización objetivo.

$\mathbf{C}_3\mathbf{X} \approx \mathbf{A}_3\mathbf{S}$

Dicha matriz se deriva teniendo en cuenta las ecuaciones normales

$\mathbf{C}_3(\mathbf{DED}^*) \approx \mathbf{A}_3\mathbf{ED}^*$.

La solución a las ecuaciones normales da la mejor coincidencia posible de la forma de onda para la salida objetivo dado el modelo de covarianza de objetos. Ahora se obtiene **G** y $\mathbf{C}_{TTT}$ resolviendo el sistema de ecuaciones

$$\mathbf{C}_{TTT}\mathbf{G} = \mathbf{C}_3$$

Para evitar problemas numéricos al calcular el término $\mathbf{J} = (\mathbf{DED}^*)^{-1}$, se modifica **J**. En primer lugar se calculan los eigenvalues (término en otro idioma) $\lambda_{1,2}$ de **J**, resolviendo $\mathbf{det}(\mathbf{J} - \lambda_{1,2}\mathbf{I}) = 0$.

Los eigenvalues se clasifican en orden descendente ($\lambda_1 \geq \lambda_2$) y el eigenvector correspondiente al eigenvalue más alto se calcula de acuerdo con la ecuación anterior. Con seguridad ha de yacer en el plano x positivo (el primer elemento tiene que ser positivo). El segundo eigenvector se obtiene a partir del primero por una rotación de –90 grados:

$$\mathbf{J} = (\mathbf{v}_1\mathbf{v}_2)\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}(\mathbf{v}_1\mathbf{v}_2)^*.$$

Se computa una matriz de ponderación a partir de la matriz de mezcla descendente **D** y la matriz de predicción $\mathbf{C}_3$,

$$\mathbf{W} = (\mathbf{D}\ \mathrm{diag}(\mathbf{C}_3)).$$

Dado que $\mathbf{C}_{TTT}$ es una función de los parámetros de predicción c_1 y c_2 (definidos en ISO/IEC 23003–1:2007), $\mathbf{C}_{TTT}\mathbf{G}=\mathbf{C}_3$ se reformula de la siguiente manera, para encontrar el punto estacionario o los puntos de la función,

$$\mathbf{\Gamma}\begin{pmatrix}\tilde{c}_1\\ \tilde{c}_2\end{pmatrix} = \mathbf{b},$$

donde $\mathbf{r} = \mathbf{D}_{TTT}\ \mathbf{C}_3)\ \mathbf{W}\ (\mathbf{D}_{TTT}\ \mathbf{C}_3)^*$ y $\mathbf{b} = \mathbf{GWC}_3\mathbf{v}$,

donde $\mathbf{D}_{TTT} = \begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 1\end{pmatrix}$ y $\mathbf{v} = \begin{pmatrix}1 & 1 & -1\end{pmatrix}$.

Si r no ofrece una solución única ($\det(r) < 10^{-3}$), se elige el punto que yace más cerca del punto que da lugar a un pasaje TTT. Como primer paso, se elige la fila *i* de r $\gamma=[\gamma_{i,1}\ \gamma_{i,2}]$ donde los elementos contienen la mayor cantidad de energía, por consiguiente, $\gamma_{i,1}^{\ 2} + \gamma_{i,2}^{\ 2} \geq \gamma_{j,1}^{\ 2} + \gamma_{j,2}^{\ 2},\quad j = 1,2.$

Luego se determina una solución de tal manera que

$$\begin{pmatrix}\tilde{c}_1\\ \tilde{c}_2\end{pmatrix} = \begin{pmatrix}1\\ 1\end{pmatrix} - 3\mathbf{y} \text{ donde } \mathbf{y} = \frac{b_{i,3}}{\left(\sum_{j=1,2}(\gamma_{i,j})^2\right) + \varepsilon}\gamma^T.$$

Si la solución obtenida para $\tilde{c}_1$ y $\tilde{c}_2$ está fuera del rango de coeficientes de predicción permitidos que se define como $-2 \le \tilde{c}_j \le 3$ (según definición en ISO/IEC 23003–1:2007), se calculan de acuerdo con lo siguiente.

En primer lugar se define la serie de puntos, x_p, de acuerdo con:

$$\mathbf{x}_p \in \begin{bmatrix} \begin{pmatrix} \min\left(3, \max\left(-2, -\frac{-2\gamma_{1,2} - b_1}{\gamma_{1,1} + \varepsilon}\right)\right) \\ -2 \end{pmatrix}, \begin{pmatrix} \min\left(3, \max\left(-2, -\frac{3\gamma_{1,2} - b_1}{\gamma_{1,1} + \varepsilon}\right)\right) \\ 3 \end{pmatrix} \\ \begin{pmatrix} -2 \\ \min\left(3, \max\left(-2, -\frac{-2\gamma_{2,1} - b_2}{\gamma_{2,2} + \varepsilon}\right)\right) \end{pmatrix}, \begin{pmatrix} 3 \\ \min\left(3, \max\left(-2, -\frac{3\gamma_{2,1} - b_2}{\gamma_{2,2} + \varepsilon}\right)\right) \end{pmatrix} \end{bmatrix},$$

y la función de distancia,

$$FuncDist\left(\mathbf{x}_p\right) = \mathbf{x}_p^* \mathbf{\Gamma} \mathbf{x}_{p1} - 2\mathbf{b}\mathbf{x}_p .$$

Luego se definen los parámetros de predicción de acuerdo con

$$\begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_2 \end{pmatrix} = \arg\min_{\mathbf{x} \in \mathbf{x}_p} \left(FuncDist\left(\mathbf{x}\right) \right).$$

Se restringen los parámetros de predicción de acuerdo con:

$$c_1 = (1-\lambda)\tilde{c}_1 + \lambda\gamma_1, \quad c_2 = (1-\lambda)\tilde{c}_2 + \lambda\gamma_2$$

donde λ, γ_1 y γ_2 se definen de la siguiente manera

$$\gamma_1 = \frac{2f_{1,1} + 2f_{5,5} - f_{3,3} + f_{1,3} + f_{5,3}}{2f_{1,1} + 2f_{5,5} + 2f_{3,3} + 4f_{1,3} + 4f_{5,3}},$$

$$\gamma_2 = \frac{2f_{2,2} + 2f_{6,6} - f_{3,3} + f_{2,3} + f_{6,3}}{2f_{2,2} + 2f_{6,6} + 2f_{3,3} + 4f_{2,3} + 4f_{6,3}},$$

$$\lambda = \left(\frac{\left(f_{1,2} + f_{1,6} + f_{5,2} + f_{5,6} + f_{1,3} + f_{5,3} + f_{2,3} + f_{6,3} + f_{3,3}\right)^2}{\left(f_{1,1} + f_{5,5} + f_{3,3} + 2f_{1,3} + 2f_{5,3}\right)\left(f_{2,2} + f_{6,6} + f_{3,3} + 2f_{2,3} + 2f_{6,3}\right)} \right)^8.$$

En el caso del decodificador MPS, se obtienen los CPCs y los correspondiente ICC_{TTT} de la siguiente manera

$D_{CPC\_1} = c_1(l,m)$, $D_{CPC\_2} = c_2(l,m)$ y $D_{ICC,TTT} = 1$

4.2.2.2.2 Renderización entre canales anteriores y envolventes

Los parámetros que determinan la renderización entre canales anteriores y envolventes se puede estimar directamente de la matriz de covarianza objetivo **F**

$$CLD_{a,b} = 10\log_{10}\left(\frac{\max\left(f_{a,a},\varepsilon^2\right)}{\max\left(f_{b,b},\varepsilon^2\right)}\right), \qquad ICC_{a,b} = \frac{\max\left(f_{a,b},\varepsilon^2\right)}{\sqrt{\max\left(f_{a,a},\varepsilon^2\right)\max\left(f_{b,b},\varepsilon^2\right)}},$$

donde *(a,b)* = (1,2) y (3,4).

Los parámetros MPS se producen en la forma

$$CLD_h^{l,m} = \mathbf{D}_{\mathrm{CLD}}(h,l,m) \quad \text{y} \quad ICC_h^{l,m} = \mathbf{D}_{\mathrm{ICC}}(h,l,m),$$

por cada caja OTT *h*.

4.2.2.3 Procesamiento estéreo

A continuación, se describe el procesamiento estéreo de las señales de objetos de audio normales 134 a 64, 322. El procesamiento en estéreo se utiliza para derivar un proceso a la representación general 142, 272 sobre la base de una representación de dos canales de los objetos de audio normales.

La mezcla descendente estéreo **X**, que está representada por las señales de objetos de audio normales 134, 264, 492a se procesa para obtener la señal de mezcla descendente modificada $\hat{\mathbf{X}}$, que está representada por las señales de objeto de audio normales 142, 272:

$$\hat{\mathbf{X}} = \mathbf{GX},$$

donde

$$\mathbf{G} = \mathbf{D}_{\mathrm{TTT}}\mathbf{C}_3 = \mathbf{D}_{\mathrm{TTT}}\,\mathbf{M}_{\mathrm{ren}}\mathbf{E}\mathbf{D}^*\mathbf{J}.$$

La salida estéreo final del transcodificador SAOC $\hat{\mathbf{X}}$ se produce mezclando **X** con un componente de señal descorrelacionada de acuerdo con:

$$\hat{\mathbf{X}} = \mathbf{G}_{\mathrm{Mod}}\mathbf{X} + \mathbf{P}_2\mathbf{X}_d.$$

donde la señal descorrelacionada $\mathbf{X}_d$ se calcula de acuerdo con lo descripto anteriormente, y las matrices de mezcla $\mathbf{G}_{\mathrm{Mod}}$ y $\mathbf{P}_2$ de acuerdo con lo siguiente:

En primer lugar, se define la matriz de error de renderización de mezcla ascendente de la siguiente manera:

$$R = A_{\mathrm{dif}}EA_{\mathrm{dif}}^{*},$$

donde

$$\mathbf{A}_{\mathrm{dif}} = \mathbf{D}_{\mathrm{TTT}}\mathbf{A}_3 - \mathbf{GD},$$

y, además, se define la matriz de covarianza de la señal predicha $\hat{\mathbf{R}}$ de la siguiente manera:

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{r}_{1,1} & \hat{r}_{1,2} \\ \hat{r}_{2,1} & \hat{r}_{2,2} \end{pmatrix} = \mathbf{GDED}^*\mathbf{G}^*.$$

A continuación se puede calcular el vector de ganancia g_{vec} de la siguiente manera:

$$\mathbf{g}_{vec} = \left(\min\left(\sqrt{\max\left(\frac{\hat{r}_{1,1} + r_{1,1} + \varepsilon^2}{r_{1,1} + \varepsilon^2}, 0 \right)}, 1.5 \right) \quad \min\left(\sqrt{\max\left(\frac{\hat{r}_{2,2} + r_{2,2} + \varepsilon^2}{r_{2,2} + \varepsilon^2}, 0 \right)}, 1.5 \right) \right),$$

la matriz de mezclado se expresa como

$$\mathbf{G}_{Mod} = \begin{cases} diag(\mathbf{g}_{vec})\mathbf{G}, & r_{1,2} > 0, \\ \mathbf{G}, & \text{de otro modo} \end{cases}$$

De manera similar, la matriz de mezclado $\mathbf{P}_2$ se expresa en los siguientes términos:

$$\mathbf{P}_2 = \begin{cases} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & , r_{1,2} > 0, \\ \mathbf{v}_R diag(\mathbf{W}_d) & , \text{de otro modo.} \end{cases}$$

Para derivar v_R y W_d, se debe resolver la ecuación característica de **R**:

$\det(R - \lambda_{1,2}I)v_{R1,R2} = 0$

Los eigenvalues se clasifican en orden descendente ($\lambda_1 \geq \lambda_2$) y el eigenvector correspondiente al eigenvalue más elevado se calcula de acuerdo con la ecuación anterior. Con seguridad ha de yacer en el plano x positivo (el primer elemento tiene que ser positivo). El segundo eigenvector se obtiene a partir del primero mediante una rotación de –90 grados:

$$\mathbf{R} = (\mathbf{v}_{R1}\mathbf{v}_{R2}) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} (\mathbf{v}_{R1}\mathbf{v}_{R2})^{*}.$$

La incorporación de P_1 = 1 1)G, R_d se puede calcular de acuerdo con:

$$\mathbf{R}_d = \begin{pmatrix} \mathbf{r}_{d11} & \mathbf{r}_{d12} \\ \mathbf{r}_{d21} & \mathbf{r}_{d22} \end{pmatrix} = diag\left(\mathbf{P}_1 \left(\mathbf{DED}^{*} \right) \mathbf{P}_1^{*} \right),$$

lo que da

$$\mathbf{w}_{d1} = \min\left(\sqrt{\frac{\lambda_1}{r_{d1} + \varepsilon}}, 2 \right), \qquad \mathbf{w}_{d2} = \min\left(\sqrt{\frac{\lambda_2}{r_{d2} + \varepsilon}}, 2 \right),$$

y por último la matriz de mezcla,

$$\mathbf{P}_2 = (\mathbf{v}_{R1} \quad \mathbf{v}_{R2}) \begin{pmatrix} \mathbf{w}_{d1} & 0 \\ 0 & \mathbf{w}_{d2} \end{pmatrix}$$

4.2.2.4 Modo Dual

El transcodificador SAOC puede dar lugar al cálculo de las matrices de mezcla $\mathbf{P}_1$, $\mathbf{P}_2$ y la matriz de predicción C_3 de acuerdo con un esquema alternativo para el rango de frecuencias superiores. Este esquema alternativo es especialmente ventajoso para las señales de mezcla descendente, donde el rango de frecuencias superiores es codificado por un algoritmo de codificación sin conservación de la forma de onda, por ejemplo SBR en AAC de Alta Eficiencia.

En el caso de las bandas de parámetros superiores, definidas por **bsTttBandsLow** ≤ pb < *numBands* , $\mathbf{P}_1$, $\mathbf{P}_2$ y $\mathbf{C}_3$ se deben calcular de acuerdo con el esquema alternativo descripto a continuación:

$$\begin{cases} \mathbf{P}_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \mathbf{P}_2 = \mathbf{G}. \end{cases}$$

Se define la mezcla descendente de energía y los vectores objetivos de energía, respectivamente:

$$\begin{cases} \mathbf{e}_{dmx} = \begin{pmatrix} e_{dmx1} \\ e_{dmx2} \end{pmatrix} = diag\left(\mathbf{D}\mathbf{E}\mathbf{D}^*\right) + \varepsilon\mathbf{I}, \\ \mathbf{e}_{tar} = \begin{pmatrix} e_{tar1} \\ e_{tar2} \\ e_{tar3} \end{pmatrix} = diag\left(\mathbf{A}_3\mathbf{E}\mathbf{A}_3^*\right), \end{cases}$$

y la matriz de ayuda

$$\mathbf{T} = \begin{pmatrix} t_{1,1} & t_{1,2} \\ t_{2,1} & t_{2,2} \\ t_{3,1} & t_{3,2} \end{pmatrix} = \mathbf{A}_3\mathbf{D}^* + \varepsilon\mathbf{I}.$$

Luego se calcula el vector de ganancia

$$\mathbf{g} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} \sqrt{\dfrac{e_{tar1}}{t_{1,1}^2 e_{dmx1} + t_{1,2}^2 e_{dmx2}}} \\ \sqrt{\dfrac{e_{tar2}}{t_{2,1}^2 e_{dmx1} + t_{2,2}^2 e_{dmx2}}} \\ \sqrt{\dfrac{e_{tar3}}{t_{3,1}^2 e_{dmx1} + t_{3,2}^2 e_{dmx2}}} \end{pmatrix},$$

que, por último, da la nueva matriz de predicción

$$\mathbf{C}_3 = \begin{pmatrix} g_1 t_{1,1} & g_1 t_{1,2} \\ g_2 t_{2,1} & g_2 t_{2,2} \\ g_3 t_{3,1} & g_3 t_{3,2} \end{pmatrix}.$$

5. Modo combinado de decodificación/transcodificación EKS SAOC, codificador da la Fig. 10 y sistemas de acuerdo con las Figs. 5a, 5b

A continuación, se presenta una breve descripción del esquema de procesamiento combinado EKS SAOC. Se propone un esquema de procesamiento combinado "EKS SAOC", donde el procesamiento EKS se integra a la cadena de decodificación/transcodificación SAOC normal mediante un esquema de etapas en cascada.

5.1. Codificador de señales de audio de acuerdo con la Fig. 5

En un primer paso, se identifican los objetos especiales para el procesamiento EKS (procesamiento Karaoke/solo realzado) como objetos en primer plano (FGO) y su número N_{FGO} (también designado N_{EAO}) se determina mediante una variable de flujo de bits "bsNumGroupsFG0". Dicha variable de flujo de bits puede estar incluida, por ejemplo, en un flujo de bits SAOC, de acuerdo con lo descripto anteriormente.

Para la generación del flujo de bits (en un codificador de señales de audio), los parámetros de todos los objetos de entrada N_{obj} se reordenan de tal manera que los objetos en primer plano FGO comprendan los últimos parámetros N_{FGO} (o, por otro lado, N_{EAO}) en cada caso, por ejemplo, OLD_1 para $[N_{obj} - N_{FGO} \leq i \leq N_{obj} - 1]$.

Del resto de los objetos que son, por ejemplo, objetos BGO de fondo u objetos de audio no realzados, se genera una señal de mezcla descendente a "estilo SAOC normal " que, al mismo tiempo, sirve como objeto BGO de fondo. A continuación, se ejecuta la mezcla descendente del objeto de fondo y los objetos en primer plano al "estilo de procesamiento EKS" y se extrae la información residual de cada objeto de fondo. De esta manera, no es necesario introducir ninguna etapa adicional. Por consiguiente, no se requiere ningún cambio de sintaxis del flujo de bits.

Dicho de otro modo, del lado del codificador, los objetos de audio no realzados se distinguen de los objetos de audio realzados. Se produce una señal de audio de mezcla descendente de objetos de audio de un canal o de dos canales (objetos de audio no realzados), en la cual puede haber uno, dos o incluso más objetos de audio normales (objetos de audio no realzados). A continuación se combina la señal de mezcla descendente de objetos de audio de un canal o de

dos canales con una o más señales de objeto de audio realzados (que pueden ser, por ejemplo, señales de un canal o señales de dos canales) para obtener una señal de mezcla descendente común (que puede ser, por ejemplo, una señal de mezcla descendente de un canal o una señal de mezcla descendente de dos canales) combinando las señales de audio de los objetos de audio realzados y la señal de mezcla descendente de objetos de audio normales.

A continuación se describe en forma sucinta la estructura básica de dicho codificador de etapas en cascada tomando como referencia la Fig. 10, que ilustra una representación esquemática de bloques de un codificador SAOC 1000, de acuerdo con una realización de la invención. El codificador SAOC 1000 comprende un primer dispositivo de mezcla descendente SAOC 1010, que por lo general es un dispositivo de mezcla descendente que no produce información residual. El dispositivo de mezcla descendente SAOC 1010 está configurado para recibir una pluralidad de señales de objetos de audio N_{BGO} 1012 de objetos de audio normales (no realzados). Además, el dispositivo de mezcla descendente SAOC 1010 está configurado para proporcionar una señal de mezcla descendente de objetos de audio normales 1014 sobre la base de los objetos de audio normales 1012, de tal manera que la señal de mezcla descendente de objetos de audio normales 1014 combine las señales de objetos de audio normales 1012 de acuerdo con los parámetros de mezcla descendente. El dispositivo de mezcla descendente SAOC 1010 también produce una información SAOC de objetos de audio normales 1016, que describe las señales de objetos de audio normales y la mezcla descendente. Por ejemplo, la información SAOC de objetos de audio normales 1016 puede comprender una información de ganancia de mezcla descendente DMG y una información de diferencias de nivel de mezcla descendente de los canales DCLD que describe la mezcla descendente ejecutada por el dispositivo de mezcla descendente SAOC 1010. Además, la información SAOC de objetos de audio normales 1016puede comprender una información de diferencias de nivel de los objetos y una información de correlación entre objetos que describe una relación entre los objetos de audio normales descripta por la señal de objetos de audio normales 1012.

El codificador 1000 comprende asimismo un segundo dispositivo de mezcla descendente SAOC 1020, que por lo general está configurado para producir una

información residual. El segundo dispositivo de mezcla descendente SAOC 1020 está configurado preferentemente para recibir una o más señales de objetos de audio realzados 1022 y también para recibirla señal de mezcla descendente de objetos de audio normales 1014.

El segundo dispositivo de mezcla descendente SAOC 1020 está configurado asimismo para producir una señal de mezcla descendente SAOC común 1024 sobre la base de las señales de objetos de audio 1022 y la señal de mezcla descendente de objetos de audio normales 1014. Al producir señal de mezcla descendente SAOC común, el segundo dispositivo de mezcla descendente SAOC 1020 trata típicamente la señal de mezcla descendente de objetos de audio normales 1014 como señal única de de objetos de un canal o de dos canales.

El segundo dispositivo de mezcla descendente SAOC 1020 está configurado además para producir una información SAOC sobre objetos de audio realzados que describe, por ejemplo, diferencias de nivel de mezcla descendente de los valores de los canales DCLD asociadas a los objetos de audio realzados, valores OLD de diferencia de nivel de los objetos asociados a los objetos de audio realzados y valores de correlación entre objetos IOC asociados a los objetos de audio realzados. Además, el segundo SAOC 1020 está configurado preferentemente para proporcionar información residual asociada a cada uno de los objetos de audio realzados, por lo que la información residual asociada a los objetos de audio realzados describe la diferencia entre una señal de objetos de audio realzados individual original y una señal de objetos de audio realzados individual estimada que se puede extraer de la señal de mezcla descendente usando la información de mezcla descendente DMG, DCLD y la información de objetos OLD, IOC.

El codificador de audio 1000 es muy adecuado para cooperar con el decodificador de audio aquí descripto.

5.2. Decodificador de señales de audio de acuerdo con la Fig. 5a

A continuación, se describe la estructura básica de un decodificador combinado EKS SAOC 500, un diagrama esquemático del cual está expuesto en la Fig. 5a.

El decodificador de audio 500 de acuerdo con la Fig. 5a está configurado para recibir una señal de mezcla descendente 510, información de flujo de bits SAOC 512 y una información sobre matrices de renderización 514. El decodificador de audio 500 comprende un procesamiento realzado Karaoke/Solo y una renderización de objetos en primer plano 520, que está configurado para producir una primera señal de objetos de audio 562, que describe objetos renderizados en primer plano, y una segunda señal de objetos de audio 564, que describe los objetos de fondo. Los objetos en primer plano pueden ser, por ejemplo, los denominados "objetos de audio realzados" y los objetos de fondo pueden ser, por ejemplo, los denominados "objetos de audio normales" u "objetos de audio no realzados". El decodificador de audio 500 comprende asimismo decodificación SAOC normal 570, que está configurado para recibir la segunda señal de objetos de audio 562 y para producir, sobre la base de ésta, una versión procesada 572 de la segunda señal de objetos de audio 564. El decodificador de audio 500 comprende asimismo un combinador 580, que está configurado para combinar la primera señal de objetos de audio 562 y la versión procesada 572 de la segunda señal de objetos de audio 564, para obtener una señal de salida 520.

A continuación, se describe la funcionalidad del decodificador de audio 500 en forma algo más detallada. Del lado de la decodificación/transcodificación SAOC, el proceso de mezcla ascendente da lugar a un esquema de etapas en cascada que comprende, en primer lugar, un procesamiento realzado Karaoke–Solo (procesamiento EKS) para descomponer la señal de mezcla descendente para obtener el objeto de fondo (BGO) y los objetos en primer plano (FGOs). Las diferencias de nivel requeridas de los objetos (OLDs) y las correlaciones entre objetos (IOCs) correspondientes a los objetos de fondo se derivan de la información sobre objetos y sobre mezcla descendente (que es, en ambos casos, información paramétrica relacionada con los objetos, y que, en ambos casos, está típicamente incluida en el flujo de bits SAOC):

$$OLD_L = \sum_{i=0}^{N-N_{FGO}-1} d_{0,i}^2 OLD_i$$

$$OLD_R = \sum_{i=0}^{N-N_{FGO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{LR} = \begin{cases} IOC_{0,1}, & N - N_{FGO} = 2, \\ 0, & \text{de otro modo.} \end{cases}$$

Además, esta etapa (que por lo general es ejecutada por el procesamiento EKS y la renderización de objetos en primer plano 520) incluye el mapeo de los objetos en primer plano contra los canales de salida finales 25 (por lo que, por ejemplo, la primera señal de objetos de audio 562 es una señal multicanal en la cual se mapean los objetos en primer plano contra uno o más canales cada uno). El objeto de fondo (que por lo general comprende una pluralidad de los denominados "objetos de audio normales") se renderiza a los correspondientes canales de salida mediante un proceso de decodificación SAOC normal (o, de lo contrario, en algunos casos por un proceso de transcodificación SAOC). Este proceso puede ser ejecutado, por ejemplo, por la decodificación SAOC normal 570. La etapa de mezcla final (por ejemplo, el combinador 580) produce una combinación ventajosa de los objetos renderizados en primer plano y las señales de objetos de fondo a la salida.

Este sistema combinado EKS SAOC representa una combinación de todas las propiedades ventajosas del sistema SAOC normal y s modo EKS. Esta estrategia permite obtener la correspondiente eficiencia utilizando el sistema propuesto con el mismo flujo de bits para entornos de reproducción tanto clásicos (renderización moderada) como similares a Karaoke/Solo (renderización extrema).

5.3. Estructura generalizada de acuerdo con la Fig. 5b

A continuación, se describe una estructura generalizada de un sistema combinado EKS SAOC 590 haciendo referencia a la Fig. 5b, que ilustra un diagrama esquemático de bloques de ese sistema combinado EKS SAOC. El sistema combinado EKS SAOC 590 de la Fig. 5b puede ser considerado asimismo como decodificador de audio.

El sistema combinado EKS SAOC 590 está configurado para recibir una señal de mezcla descendente 510a, una información de flujo de bits SAOC 512a e información sobre matrices de renderización 514a. Además, el sistema combinado EKS SAOC 590 está configurado para producir una señal de salida 520a sobre la base de ésta.

El sistema combinado EKS SAOC 590 comprende una etapa de procesamiento I del tipo SAOC 520a, que recibe la señal de mezcla descendente 510a, la información de flujo de bits SAOC 512a (o por lo menos parte de la misma) y la información de matrices de renderización 514a (o por lo menos parte de la misma). En particular, la etapa de procesamiento I del tipo SAOC 520a recibe valores de diferencias de nivel de los objetos de la primera etapa (OLD,). La etapa de procesamiento I del tipo SAOC 520a suministra una o más señales 562a que describe una primera serie de objetos (por ejemplo, objetos de audio de un primer tipo de objetos de audio). La etapa de procesamiento I del tipo SAOC 520a proporciona asimismo una o más señales 564a que describen una segunda serie de objetos.

El sistema combinado EKS SAOC comprende asimismo una etapa de procesamiento II del tipo SAOC 570a, que está configurada para recibir dichas una o más señales 564a que describe la segunda serie de objetos y para producir, sobre la base de ésta, una o más señales 572a que describen una tercera serie de objetos usando las diferencias de nivel de los objetos de la segunda etapa, que están incluidos en la información de flujo de bits SAOC 512a, y también por lo menos parte de la información de matrices de renderización 514. El sistema combinado EKS SAOC comprende asimismo un combinador 580a, que puede ser, por ejemplo, un sumador, para producir las señales de salida 520a combinando dichas una o más señales 562a que describen la primera serie de objetos y dichas una o más señales 570a que describen la tercera serie de objetos (donde la tercera serie de objetos puede ser una versión procesada de la segunda serie de objetos).

Para resumir lo antedicho, la Fig. 5b ilustra una forma generalizada de la estructura básica descripta con referencia a la Fig. 5a anterior de otra realización de la invención.

6. Evaluación Perceptual del Esquema de Procesamiento Combinado EKS SAOC

6.1 Metodología de Ensayo, Diseño y Elementos

Estas pruebas de audición subjetivas se llevaron a cabo en un recinto de audición acústicamente aislado que está diseñado para permitir la audición de alta calidad. La reproducción se realizó utilizando auriculares (STAX SR Lambda Pro con Convertidor Lake–People DIA y Monitor STAX SRM). El método de ensayo siguió los procedimientos standard empleados en las pruebas de verificación de audio espacial, basadas en el método de "estímulo múltiple con referencia y anclas ocultos" (MUSHRA) para la evaluación subjetiva de la calidad de audio intermedia (ver la referencia [7]).

Un total de ocho oyentes participaron en la prueba realizada. Todos los sujetos pueden ser considerados oyentes experimentados. De acuerdo con la metodología MUSHRA, los oyentes fueron instruidos para comparar todas las condiciones de ensayo contra la referencia. Las condiciones de ensayo fueron distribuidas al azar por cada elemento de ensayo y por cada oyente. Las respuestas subjetivas fueron registradas por un programa MUSHRA de computación en una escala en el rango de 0 a 100.

Se permitió un desplazamiento instantáneo entre los elementos en estudio. La prueba MUSHRA se llevó a cabo para evaluar la eficiencia perceptual de los modos SAOC considerados y el sistema propuesto descripto en la tabla de la Fig. 6a, que presenta una descripción del diseño de ensayo de audición. Las correspondientes señales de mezcla descendente fueron codificadas utilizando un codificador de núcleo AAC con una velocidad de transmisión de bits de 128 kbps. Para evaluar la calidad perceptual del sistema combinado EKS SAOC, se lo compara contra el sistema SAOC RM normal (sistema de modelos de referencia SAOC) y el modo EKS actual (modo realzado de Karaoke–Solo) correspondientes a dos entornos de ensayo de renderización diferentes descriptos en la tabla de la Fig. 6b, que describe los sistemas en estudio.

Se aplicó la codificación residual con una velocidad de transmisión de bits de 20 kbps para el modo EKS actual y un sistema combinado EKS SAOC propuesto. Se debe notar que en el caso del modo EKS actual es necesario generar un objeto de fondo estéreo (BGO) con anterioridad al procedimiento actual

de codificación/decodificación, ya que este modo tiene limitaciones con respecto al número y tipo de objetos de entrada.

El material de ensayo de audición y los correspondientes parámetros de mezcla descendente y renderización utilizados en las pruebas realizadas fueron seleccionados de la serie de elementos de audio propuestos (CfP) descriptos en el documento [2]. Los correspondientes datos sobre configuraciones de aplicaciones de renderización para "Karaoke" y "Clásico" se pueden encontrar en la tabla de la Fig. 6c, que describe los elementos de la prueba de audición y las matrices de renderización.

6.2 Resultados de la Prueba de Audición

Se puede encontrar una breve reseña general en términos de diagramas que demuestran los resultados obtenidos de la prueba de audición en las Figs. 6d y 6e, donde la Fig. 6d ilustra las puntuaciones MUSHRA promediadas correspondientes a la prueba de audición de renderización del tipo Karaoke/Solo y la Fig. 6e ilustra las puntuaciones MUSHRA promediadas correspondientes a la prueba de audición de renderización clásica. Los trazados ilustran la calificación MUSHRA promedio por elemento asignada por todos los oyentes y el valor medio estadístico de todos los elementos evaluados junto con los intervalos de confianza asociados de 95%.

Se pueden plantear las siguientes conclusiones basándose en los resultados de las pruebas de audición realizadas:

La Fig. 6d representa la comparación del modo actual EKS con el sistema combinado EKS SAOC para aplicaciones del tipo Karaoke. Con respecto a todos los elementos analizados, no se puede observar diferencia significativa alguna (en el sentido estadístico) en el rendimiento entre estos dos sistemas. De esta observación se puede concluir que el sistema combinado EKS SAOC puede explotar eficientemente la información residual que alcanza la eficiencia del modo EKS. También se puede notar que la eficiencia del sistema SAOC normal (sin residual) es inferior a los otros dos sistemas.

La Fig. 6e representa la comparación del SAOC normal actual con el sistema combinado EKS SAOC para entornos de renderización clásica. En el caso de todos los elementos analizados, la eficiencia de estos dos sistemas es

estadísticamente igual. Esto demuestra la funcionalidad correcta del sistema combinado EKS SAOC para una situación de renderización clásica.

Por lo tanto, se puede concluir que el sistema unificado propuesto que combina el modo EKS con el SAOC normal mantiene las ventajas de la calidad auditiva subjetiva respecto de los tipos de renderización correspondientes.

Tomando en cuenta el hecho de que el sistema combinado EKS SAOC propuesto ya no tiene restricciones con respecto a los objetos BGO, sino que tiene una capacidad de renderización totalmente flexible del modo SAOC normal y puede utilizar el mismo flujo de bits para todos los tipos de renderización, parece ser ventajoso incorporarlo a la norma MPEG SAOC.

7. Método de acuerdo con la Fig. 7

A continuación se describe un método para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente e información paramétrica relacionada con los objetos con referencia a la Fig. 7, que ilustra un gráfico de flujo de dicho método.

El método 700 comprende una etapa 710 de descomposición de una representación de señal de mezcla descendente, para dar origen a una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio que depende de la representación de señal de mezcla descendente y por lo menos parte de la información paramétrica relacionada con los objetos. El método 700 comprende asimismo una etapa 720 de procesamiento de la segunda información de audio que depende de la información paramétrica relacionada con los objetos, para obtener una versión procesada de la segunda información de audio. El método 700 comprende asimismo una etapa 730 de combinación de la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente.

El método 700 de acuerdo con la Fig. 7 puede ser complementado con cualquiera de las características y funcionalidades aquí descriptas con respecto al aparato de la invención. Además, el método 700 trae aparejadas las ventajas mencionadas con respecto al aparato de la invención.

8. Alternativas de Implementación

Si bien se han descripto algunos aspectos en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a una etapa del método o a una característica de una etapa del método. De manera análoga, los aspectos descriptos en el contexto de una etapa del método también representan una descripción de un correspondiente bloque o elemento o característica de un aparato correspondiente. Algunas o todas las etapas del método pueden ser ejecutadas por (o utilizando) un aparato de hardware, como por ejemplo un microprocesador, una computadora programable o un circuito electrónico. En algunas realizaciones, una o más de las etapas más importantes del método pueden ser ejecutadas por ese tipo de aparato.

La señal de audio codificada de la invención puede ser almacenada en un medio de almacenamiento digital o puede ser transmitida por un medio de transmisión tal como un medio de transmisión inalámbrica o un medio de transmisión cableado tal como la Internet.

Dependiendo de ciertos requerimientos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software. La implementación se puede ejecutar utilizando un medio de almacenamiento digital, por ejemplo un disco blando, un DVD, un Blue–Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, con señales de control legibles electrónicamente almacenados en el mismo, que cooperan (o pueden cooperar) con un sistema de computación programable a fin de ejecutar el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por computación.

Algunas realizaciones de acuerdo con la invención comprenden un portador de datos que cuenta con señales de control legibles electrónicamente, con capacidad para cooperar con un sistema de computación programable para la ejecución de los métodos aquí descriptos.

En general, las realizaciones de la presente invención pueden ser implementadas en forma de producto programa de computación con un código de programa, donde el código de programa es operativo para realizar uno de los métodos al ejecutarse el programa de computación en una computadora. El

código de programa puede ser almacenado, por ejemplo, en un portador legible por una máquina.

Otras realizaciones comprenden el programa de computación para ejecutar uno de los métodos aquí descriptos, almacenados en un portador legible por una máquina. En otras palabras, una realización del método de la invención consiste, por lo tanto, en un programa de computación que consta de un código de programa para realizar uno de los métodos aquí descriptos al ejecutarse el programa de computación en una computadora.

Otra realización del método de la invención consiste, por lo tanto, en un portador de datos (o un medio de almacenamiento digital, o un medio legible por computadora) que comprende, grabado en el mismo, el programa de computación para ejecutar uno de los métodos descriptos en la presente. El portador de datos, el medio de almacenamiento digital o el medio registrado son por lo general tangibles y/no no de transmisión.

Otra realización del método de la invención consiste, por lo tanto, en un flujo de datos o una secuencia de señales que representan el programa de computación para ejecutar uno de los métodos aquí descriptos. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para ser transferido a través de una conexión de comunicación de datos, por ejemplo a través de la Internet.

Otra realización comprende un medio de procesamiento, por ejemplo una computadora o un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los métodos aquí descriptos.

Otra realización comprende una computadora en la cual se ha instalado el programa de computación para ejecutar uno de los métodos aquí descriptos.

En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo, una matriz de puertas programable en el campo) para ejecutar algunas o todas las funcionalidades de los métodos aquí descriptos. En algunas realizaciones, una matriz de puertas programable en el campo puede cooperar con un microprocesador para ejecutar uno de los métodos aquí descriptos. En general, los métodos se ejecutan preferentemente mediante cualquier aparato de hardware.

Las realizaciones precedentemente descriptas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y detalles descriptos en este documento han de ser evidentes para otras personas con capacitación en la técnica. Por lo tanto, sólo se pretende limitarse al alcance de las reivindicaciones de patente siguientes y no a los detalles específicos presentados a título de descripción y explicaciones de las presentes realizaciones.

9. Conclusiones

A continuación se resumen brevemente algunos aspectos y ventajas del sistema combinado EKS SAOC de acuerdo con la presente invención. En el caso de los entornos de reproducción de Karaoke y Solo, el modo de procesamiento SAOC EKS soporta tanto la reproducción de los objetos de fondo/objetos en primer plano) de estos grupos de objetos.

Asimismo, el primer modo se considera el objetivo principal del procesamiento EKS, el último aporta mayor flexibilidad.

Se ha encontrado que una generalización de la funcionalidad EKS, en consecuencia, conlleva el esfuerzo de combinar el EKS con el modo de procesamiento SAOC normal para obtener un sistema unificado. Los potenciales de dicho sistema unificado son:

Una única estructura clara de decodificación/transcodificación SAOC;

Un flujo de bits tanto para el modo EKS como para el SAOC normal; no hay limitación con respecto al número de objetos de entrada que comprenden el objeto de fondo (BGO), por lo que no hay necesidad de generar el objeto de fondo antes de la etapa de codificación SAOC y

Soporte de una codificación residual correspondiente a objetos en primer plano que da una calidad perceptual mejorada en situaciones demandantes de reproducción para Karaoke/Solo.

Estas ventajas se hacen posibles merced al sistema unificado aquí descripto.

Referencias

[1] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N8853, "Call for Proposals on Spatial Audio Object Coding", 79th MPEG Meeting, Marrakech, January 2007.

[2] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N9099, "Final Spatial Audio Object Coding Evaluation Procedures and Criterion", 80th MPEG Meeting, San Jose, April 2007.

[3] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N9250, "Report on Spatial Audio Object Coding RM0 Selection", 81st MPEG Meeting, Lausanne, July 2007.

[4] ISO/IEC JTC1/SC29/WG11 (MPEG), Document M15123, "Information and Verification Results for CE on Karaoke/Solo system improving the performance of MPEG SAOC RM0", 83rd MPEG Meeting, Antalya, Turkey, January 2008.

[5] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N10659, "Study on ISO/IEC 23003-2:200x Spatial Audio Object Coding (SAOC)", 88th MPEG Meeting, Maui, USA, April 2009.

[6] ISO/IEC JTC1/SC29/WG11 (MPEG), Document M10660, "Status and Workplan on SAOC Core Experiments", 88th MPEG Meeting, Maui, USA, April 2009.

[7] EBU Technical recommendation: "MUSHRA-EBU Method for Subjective Listening Tests of Intermediate Audio Quality", Doc. B/AIM022, October 1999.

[8] ISO/IEC 23003-1:2007, Information technology – MPEG audio technologies – Part 1: MPEG Surround.

REIVINDICACIONES

Habiendo así especialmente descripto y determinado la naturaleza de la presente invención y la forma como la misma ha de ser llevada a la práctica, se declara reivindicar como de propiedad y derecho exclusivo:

1. Un decodificador de señales de audio (100; 200; 500; 590) para producir una representación de señal de mezcla ascendente dependiendo de una señal de mezcla descendente y representación (1 12; 210; 510; 510a), una información paramétrica relacionada con los objetos (1 10; 2 12; 5 12; 5 12a) donde el decodificador de señales de audio comprende:

un separador de objetos (130; 260; 520; 520a) configurado para descomponer una representación de señal de mezcla descendente, para producir una primera información de audio (132; 262; 562; 562a) que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio, y una segunda información de audio (134; 264; 564; 564a) que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de señal de mezcla descendente;

un procesador de señales de audio configurado para recibir la segunda información de audio (134; 264; 564; 564a) y para procesar la segunda información de audio que depende de la información paramétrica relacionada con los objetos, para obtener una versión procesada (142; 272; 572; 572a) de la segunda información de audio y

un combinador de señales de audio (150; 280; 580; 580a) configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente.

2. El decodificador de señales de audio de acuerdo con la reivindicación 1, donde el decodificador de señales de audio está configurado para producir la representación de señal de mezcla ascendente dependiendo de una información residual asociada a una subserie de objetos de audio representados por la representación de señal de mezcla descendente, donde el separador de objetos está configurado para descomponer la representación de señal de mezcla descendente para producir la primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio

a los cuales está asociada la información residual, y la segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio, a los cuales no se asocia ninguna información residual, dependiendo de la representación de la señal de mezcla descendente y usando la información residual.

3. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con la reivindicación 1 o 2, donde el separador de objetos está configurado para producir la primera información de audio de tal manera que uno o más objetos de audio del primer tipo de objetos de audio se realcen con respecto a los objetos de audio del segundo tipo de objetos de audio de la primera información de audio, y donde el separador de objetos está configurado para producir la segunda información de audio de tal manera que los objetos de audio del segundo tipo de objetos de audio sobresalgan con respecto a los objetos de audio del primer tipo de objetos de audio de la segunda información de audio.

4. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 3, donde el decodificador de señales de audio está configurado para ejecutar un proceso en 2 etapas, de tal manera que un procesamiento de la segunda información de audio en el procesador de señales de audio (140; 270; 570; 570a) se ejecute después de una separación entre la primera información de audio, que describe la primera serie de uno o más objetos de audio del primer tipo de objetos de audio, y la segunda información de audio que describe la segunda serie de uno o más objetos de audio del segundo tipo de objetos de audio.

5. El decodificador de señales de audio (100; 200; 500; 570) de acuerdo con una de las reivindicaciones 1 a 4, donde el procesador de señales de audio está configurado para procesar la segunda información de audio (134; 264; 564; 564a) dependiendo de la información paramétrica relacionada con los objetos (110; 212; 512; 512a) asociados con los objetos de audio del segundo tipo de objetos de audio e independiente de la información paramétrica relacionada con los objetos (110; 212; 512; 512a) asociado a los objetos de audio del primer tipo de objetos de audio.

6. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 5, donde el separador de objetos está configurado

para obtener la primera información de audio (132; 262; 562; 562a, X_{EAO}) y la segunda información de audio (134; 264; 564; 564a, X_{OBJ}) usando una combinación lineal de uno o más canales de señal de mezcla descendente de la representación de señal de mezcla descendente y uno o más canales residuales, donde el separador de objetos está configurado para obtener parámetros de combinación para ejecutar la combinación lineal dependiendo de los parámetros de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio (m_0 ...$m_{NEAO0-1}$) y dependiendo de coeficientes de predicción de canales ($c_{j,0}$, $c_{i,1}$) de los objetos de audio del primer tipo de objetos de audio.

7. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 6, en el cual el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Predicción}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Predicción}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{M}_{\text{Predicción}} = \tilde{\mathbf{D}}^{-1}\mathbf{C}$$

donde $\mathbf{X}_{OBJ}$ representa los canales de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objeto de la primera información de audio;

donde $\tilde{D}^{-1}$ representa una matriz que es inversa de una matriz de mezcla descendente ampliada;

donde C describes una matriz que representa una pluralidad de coeficientes de predicción de canales, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;; donde l_0 y r_0 representan los canales de la representación de señal de mezcla descendente;

donde $\mathbf{res}_0$ a $\mathbf{res}_{NEAO-1}$ representan canales residuales y

donde $\mathbf{A}^{EAO}$ es una matriz de pre–renderización de EAO; y

8. El decodificador de señales de audio de acuerdo con la reivindicación 7, en el cual el separador de objetos está configurado para obtener la matriz de mezcla descendente inversa $\tilde{D}^{-1}$ en forma inversa a una matriz de mezcla descendente ampliada $\tilde{D}$ que se define de la siguiente manera:

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

donde el separador objetos está configurado para obtener la matriz C de la siguiente manera:

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

donde m_0 a m_{NEAO-1} son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio;

donde n_0 a n_{NEAO-1} son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio.

9. El decodificador de señales de audio de acuerdo con la reivindicación 8, en el cual el separador de objetos está configurado para computar los coeficientes de predicción $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$; de la siguiente manera:

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ y}$$

donde el separador de objetos está configurado para derivar coeficientes de predicción restringidos $c_{j,0}$ y $c_{j,1}$ a partir de los coeficientes de predicción $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$; utilizando un algoritmo de restricción, o para usar los coeficientes de restricción $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$; como coeficientes de predicción $c_{j,0}$ y $c_{j,1}$;

donde las cantidades de energía P_{Lo}, P_{Ro}, P_{LoRo}, P_{LoCoj}, y $P_{RoCo,j}$, se definen como

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k},$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k},$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k},$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j},$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}.$$

donde los parámetros OLD_L, OLDR y $IOC_{L,R}$ corresponden a objetos de audio del segundo tipo de objetos de audio y se definen de acuerdo con

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & \text{de otro modo.} \end{cases}$$

donde $d_{0,i}$ y $d_{1,i}$ son valores de mezcla descendente asociados a los objetos de audio del segundo tipo de objetos de audio;

donde OLD_i son valores de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio;

donde N es un número total de objetos de audio;

donde N_{EAO} es un número de objetos de audio del primer tipo de objetos de audio;

donde $IOC_{0,1}$ es un valor de correlación entre objetos con un par de objetos de audio del segundo tipo de objetos de audio;

donde $e_{i,j}$ y $e_{L,R}$ son valores de covarianza derivados de los parámetros de diferencia de nivel de los objetos y parámetros de correlación entre objetos y donde $e_{i,j}$ está asociado a un par de objetos de audio del primer tipo de objetos de audio y $e_{L,R}$ está asociado a un par de objetos de audio del segundo tipo de objetos de audio.

10. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 6, donde el separador de objetos está configurado

para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Predicción}} \begin{pmatrix} \frac{d_0}{\overline{res_0}} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Predicción}} \begin{pmatrix} \frac{d_0}{\overline{res_0}} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

donde

$$\mathbf{M}_{\text{Predicción}} = \tilde{\mathbf{D}}^{-1}\mathbf{C}$$

donde $\mathbf{X}_{OBJ}$ representa un canal de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objeto de la primera información de audio;

donde $\tilde{\mathbf{D}}^{-1}$ representa una matriz que es inversa a una matriz de mezcla descendente ampliada;

donde $\mathbf{C}$ describe una matriz que representa una pluralidad de coeficientes de predicción de canales $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

donde d_0 representa un canal de la representación de señal de mezcla descendente; y

donde $\mathbf{res_0}$ **a** $\mathbf{res_{NEAO-1}}$ representan canales residuales y

donde $\mathbf{A}^{\mathbf{EAO}}$ es una matriz de pre–renderización de EAO.

11. El decodificador de señales de audio de acuerdo con la reivindicación 10, donde el separador de objetos está configurado para obtener la matriz de mezcla descendente inversa $\tilde{\mathbf{D}}^{-1}$ que es inversa a una matriz de mezcla descendente ampliada $\tilde{D}$ que se define como

$$\tilde{\mathbf{D}} = \left(\begin{array}{c|ccc} 1 & m_0 & \cdots & m_{N_{EAO}-1} \\ \hline m_0 & -1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

donde el separador de objetos está configurado para obtener la matriz C de la siguiente manera

$$\mathbf{C}=\left(\begin{array}{c|ccc} 1 & 0 & \ldots & 0 \\ \hline c_0 & 1 & \ldots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1} & 0 & \ldots & 1 \end{array}\right);$$

donde m_0 a m_{NEAO-1}, son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio.

12. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 6, donde el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ}=\mathbf{M}_{OBJ}^{Energia}\begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO}=\mathbf{A}^{EAO}\mathbf{M}_{EAO}^{Energia}\begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

donde $\mathbf{X}_{OBJ}$ representa los canales de la segunda información de audio;

donde $\boldsymbol{X}_{EAO}$ representa las señales de objeto de la primera información de audio;

donde

$$\mathbf{M}_{OBJ}^{Energia}=\begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L+\sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R+\sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{Energia}=\begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L+\sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R+\sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L+\sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R+\sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

donde m_0 a m_{NEAO-1}, son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio;

donde n_0 a n_{NEAO-1} son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio

donde OLD_i son valores de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio;

donde OLD_L y OLD_R son valores comunes de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio

y donde A^{EAO} es una matriz de pre–renderización.

13. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 5, donde el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energia}\left(d_0\right)$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{Energia}\left(d_0\right)$$

donde $\mathbf{X}_{OBJ}$ representa un canal de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objeto de la primera información de audio;

donde **d0** representa un canal de la representación de señal de mezcla descendente;

donde

$$\mathbf{M}_{OBJ}^{Energia} = \left(\sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}}\right)$$

$$\mathbf{M}_{EAO}^{Energia} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

donde m_0 a m_{NEAO-1},son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio;

donde OLD_i son valores de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio;

donde OLD_L es un valor común de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio y

donde $\mathbf{A}^{EAO}$ es una matriz de pre–renderización.

14. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 13, en el cual el separador de objetos está configurado para aplicar una matriz de renderización a la primera información de audio (132; 262; 562; 562a) para mapear las señales de objeto de la primera información de audio sobre canales de audio de la representación de señal de audio de mezcla ascendente (120; 220,222; 562; 562a).

15. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 14, en el cual el procesador de señales de audio (140; 270; 570; 570a) está configurado para ejecutar un preprocesamiento estéreo de la segunda información de audio (134; 264; 564; 564a) dependiendo de una información de renderización (M_{ren}), una información de covarianza relacionada con los objetos (E), una información de mezcla descendente (D), para obtener canales de audio de acuerdo con la versión procesada de la segunda información de audio.

16. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con la reivindicación 15, donde el procesador de señales de audio (140; 270; 570; 570a) está configurado para ejecutar el procesamiento estéreo a fin de mapear una contribución estimada de los objetos de audio (ED*JX) de la segunda información de audio (134; 264; 564; 564a) sobre una pluralidad de canales de la representación de señal de mezcla ascendente, dependiendo de una información de renderización y una información de covarianza.

17. El decodificador de señales de audio de acuerdo con la reivindicación 15 o la reivindicación 16, en el cual el procesador de señales de audio está configurado para agregar una contribución de señales de audio descorrelacionadas (P_2X_d) a la segunda información de audio, o una información derivada de la segunda información de audio, que depende de una información de error de renderización (R) y uno o más valores de escalamiento de intensidad de señal descorrelacionados (w_{d1}, w_{d2}).

18. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 14, en el cual el procesador de señales de audio (140; 270;

570; 570a) está configurado para ejecutar un postprocesamiento de la segunda información de audio (134; 264; 564; 564a) dependiendo de una información de renderización (A), una información de covarianza relacionada con los objetos (E) y una información de mezcla descendente (D).

19. El decodificador de señales de audio de acuerdo con la reivindicación 18, en el cual el procesador de señales de audio está configurado para ejecutar un procesamiento de mono a binaural de la segunda información de audio, para mapear un solo canal de la segunda información de audio sobre dos canales de la representación de señal de mezcla ascendente, tomando en cuenta una función de transferencia de cabeza.

20. El decodificador de señales de audio de acuerdo con la reivindicación 18, donde el procesador de señales de audio está configurado para ejecutar un procesamiento de mono a estéreo de la segunda información de audio, para mapear un solo canal de la segunda información de audio sobre dos canales de de la representación de señal de mezcla ascendente.

21. El decodificador de señales de audio de acuerdo con la reivindicación 18, donde el procesador de señales de audio está configurado para ejecutar un procesamiento de estéreo a binaural de la segunda información de audio, para mapear dos canales de la segunda información de audio sobre dos canales de la representación de señal de mezcla ascendente, tomando en cuenta una función de transferencia de cabeza.

22. El decodificador de señales de audio de acuerdo con la reivindicación 18, donde el procesador de señales de audio está configurado para ejecutar un procesamiento de estéreo a estéreo de la segunda información de audio sobre dos canales de la segunda información de audio sobre dos canales de la representación de señal de mezcla ascendente.

23. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 22, donde el separador de objetos está configurado para tratar objetos de audio del segundo tipo de objetos de audio, con los cuales no está asociada ninguna información residual, en forma de objeto de audio único y donde el procesador de señales de audio (140; 270; 570; 570a) está configurado para considerar parámetros de renderización de objetos específicos asociados a los objetos de audio del segundo tipo de objetos de audio a fin de ajustar las

contribuciones de los objetos de audio del segundo tipo de objetos de audio a la representación de señal de mezcla ascendente.

24. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 23, donde el separador de objetos está configurado para obtener uno o dos valores comunes de diferencia de nivel de los objetos (OLD_L, OLD_R) correspondientes a una pluralidad de objetos de audio del segundo tipo de objetos de audio; y

donde el separador de objetos está configurado para usar el valor común de diferencia de nivel de los objetos para el cómputo de los coeficientes de predicción de canales (CPC); y

donde el separador de objetos está configurado para usar los coeficientes de predicción de canales a fin de obtener uno o dos canales de audio que representan la segunda información de audio.

25. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 24, donde el separador de objetos está configurado para obtener uno o dos o uno o dos valores comunes de diferencia de nivel de los objetos (OLD_L, OLD_R) correspondientes a una pluralidad de objetos de audio del segundo tipo de objetos de audio; y

donde el separador de objetos está configurado para usar el valor común de diferencia de nivel de los objetos para el cómputo de las anotaciones de una matriz (M); y

donde el separador de objetos está configurado para usar la matriz (M) a fin de obtener uno o dos canales de audio que representan la segunda información de audio.

26. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 25, donde el separador de objetos está configurado para usar la matriz (M) a fin de obtener selectivamente un valor común de correlación entre objetos ($IOC_{L,R}$) asociado al objeto de audio del segundo tipo de objetos de audio dependiendo la información paramétrica relacionada con los objetos si se encuentra que hay dos objetos de audio del segundo tipo de objetos de audio, y para fijar el valor común de correlación entre objetos de los objetos de audio del segundo tipo de objetos de audio en cero si se encuentra que hay más o menos de dos objetos de audio del segundo tipo de objetos y audio; y

donde el separador de objetos está configurado para usar el valor común de correlación entre objetos para el cómputo de las anotaciones de una matriz (M); y donde el separador de objetos está configurado para usar el valor común de correlación entre objetos asociado a los objetos de audio del segundo tipo de objetos de audio para obtener dichos uno o más canales de audio que representan la segunda información de audio.

27. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 26, donde el procesador de señales de audio está configurado para renderizar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, para obtener una representación renderizada de los objetos de audio del segundo tipo de objetos de audio como versión procesada de la segunda información de audio.

28. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 27, donde el separador de objetos está configurado para producir la segunda información de audio de manera tal que la segunda información de audio describa más de dos objetos de audio del segundo tipo de objetos de audio.

29. El decodificador de señales de audio de acuerdo con la reivindicación 28, donde el separador de objetos está configurado para obtener, como segunda información de audio, una representación de una señal de audio de un solo canal o una representación de una señal de audio de dos canales que representa más de dos objetos de audio del segundo tipo de objetos de audio.

30. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 29, donde el procesador de señales de audio está configurado para recibir la segunda información de audio y para procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, tomando en cuenta la información paramétrica relacionada con los objetos asociada a más de dos objetos de audio del segundo tipo de objetos de audio.

31. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 30, donde el decodificador de señales de audio está configurado para extraer una información sobre número total de objetos (bsNumObjects) y una información del número de objetos en primer plano

(bsNumGroupsFGO) a partir de una información de configuración (SAOCSpecificConfig) de la información paramétrica relacionada con los objetos y para determinar el número de objetos de audio del segundo tipo de objetos de audio formando una diferencia entre la información del número total de objetos y la información del número de objetos en primer plano.

32. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 31, donde el separador de objetos está configurado usar la información paramétrica relacionada con los objetos asociados a los objetos N_{EAO} del primer tipo de objetos de audio para obtener, como primera información de audio, N_{EAO} señales de audio (X_{EAO}) que representan los $N–N_{EAO}$ objetos de audio del primer tipo de objetos de audio y para obtener, como segunda información de audio, una o dos señales de audio (X_{OBJ}) que representan los $N–N_{EAO}$ objetos de audio del segundo tipo de objetos de audio, para tratar los N–NEAO objetos de audio del segundo tipo de objetos de audio como único objeto de audio de un canal o de dos canales; y

donde el procesador de señales de audio está configurado para renderizar individualmente los $N–N_{EAO}$ objetos de audio representados por dichas una o dos señales de audio de la segunda información de audio usando la información paramétrica relacionada con los objetos asociada a los $N–N_{EAO}$ objetos de audio del segundo tipo de objetos de audio.

33. Un método para producir una representación de señal de mezcla ascendente que depende de una representación de la señal de mezcla descendente e información paramétrica relacionada con los objetos, método que comprende:

descomponer la representación de señal de mezcla descendente, para producir una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de la señal de a descendente y utilizando por lo menos parte de la información paramétrica relacionada con los objetos;

y

procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, a fin de obtener una versión procesada de la segunda información de audio y

combinar la primera información de audio con la versión procesada de la segunda información de audio para obtener la representación de señal de mezcla ascendente.

34. Un programa de computación para poner en práctica el método de acuerdo con la reivindicación 33 al ejecutarse el programa de computación en una computadora.

p.de: FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.

RESUMEN

Un decodificador de señales de audio para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente e información paramétrica relacionada con los objetos comprende un separador de objetos configurado para descomponer la representación de señal de mezcla descendente, a fin de proporcionar una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio, que depende de la representación de la señal de mezcla descendente y usar por lo menos parte de la información paramétrica relacionada con los objetos. El decodificador de señales de audio comprende asimismo un procesador de señales de audio configurado para recibir la segunda información de audio y para procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, para obtener una versión procesada de la segunda información de audio. El decodificador de señales de audio comprende asimismo un combinador de señales de audio configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente.

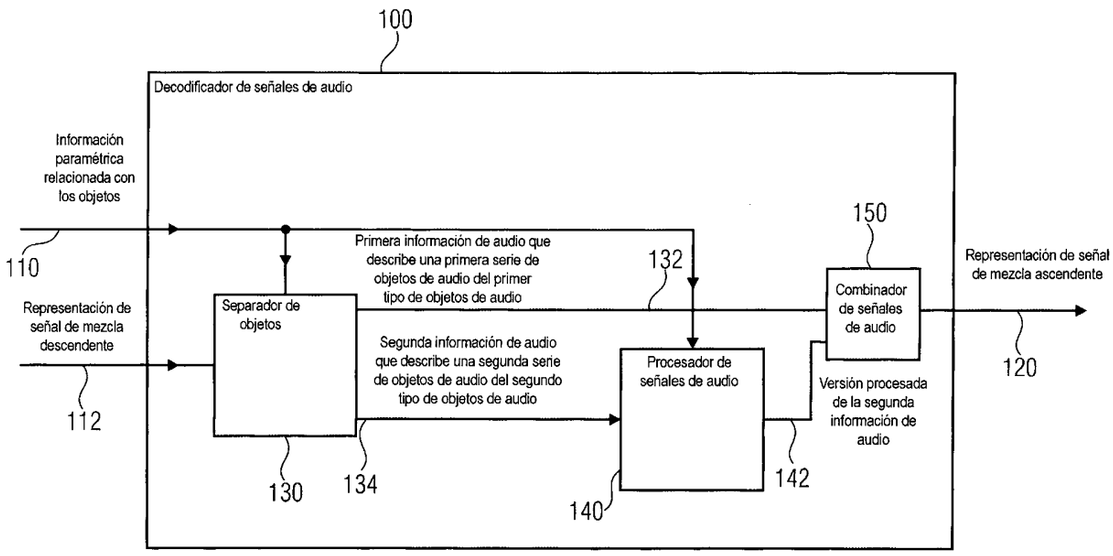
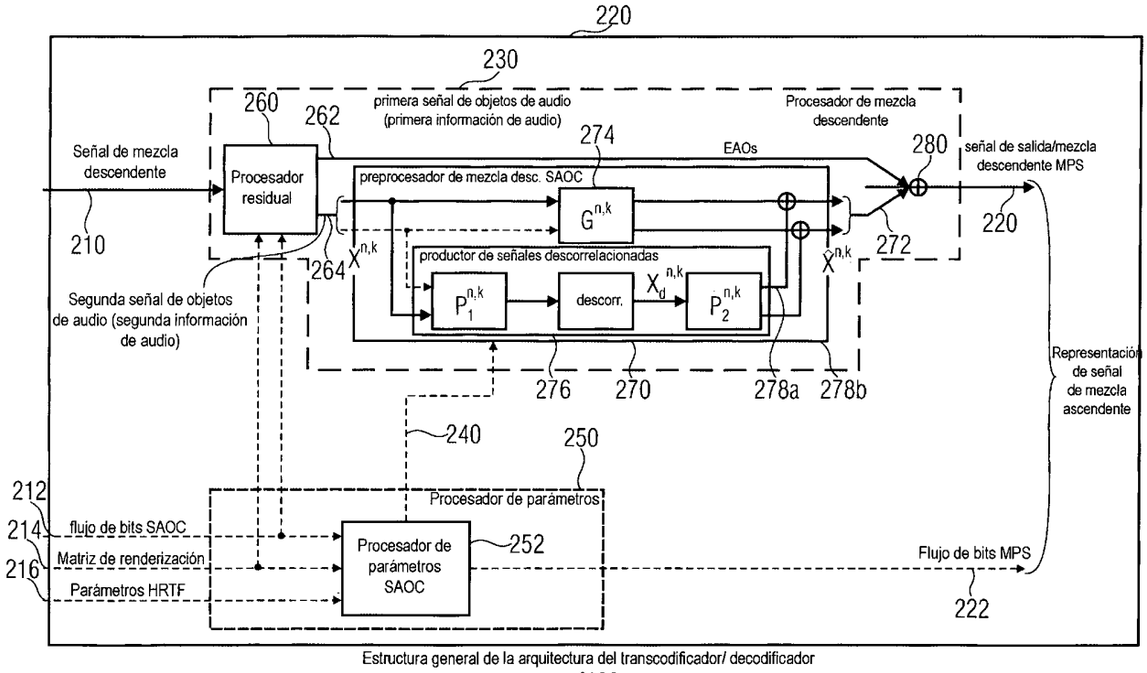
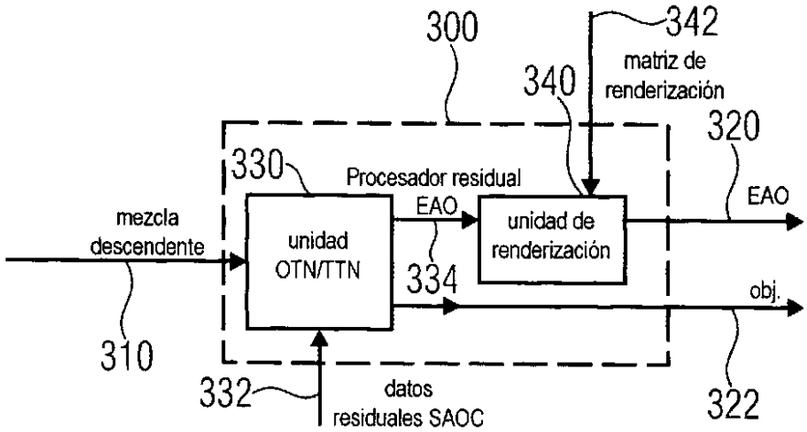

100
Decodificador de señales de audio
Información paramétrica relacionada con los objetos
110
Representación de señal de mezcla descendente
112
Separador de objetos
130
Primera información de audio que describe una primera serie de objetos de audio del primer tipo de objetos de audio
132
Segunda información de audio que describe una segunda serie de objetos de audio del segundo tipo de objetos de audio
134
Procesador de señales de audio
140
142
Versión procesada de la segunda información de audio
150
Combinador de señales de audio
Representación de señal de mezcla ascendente
120

FIG. 1



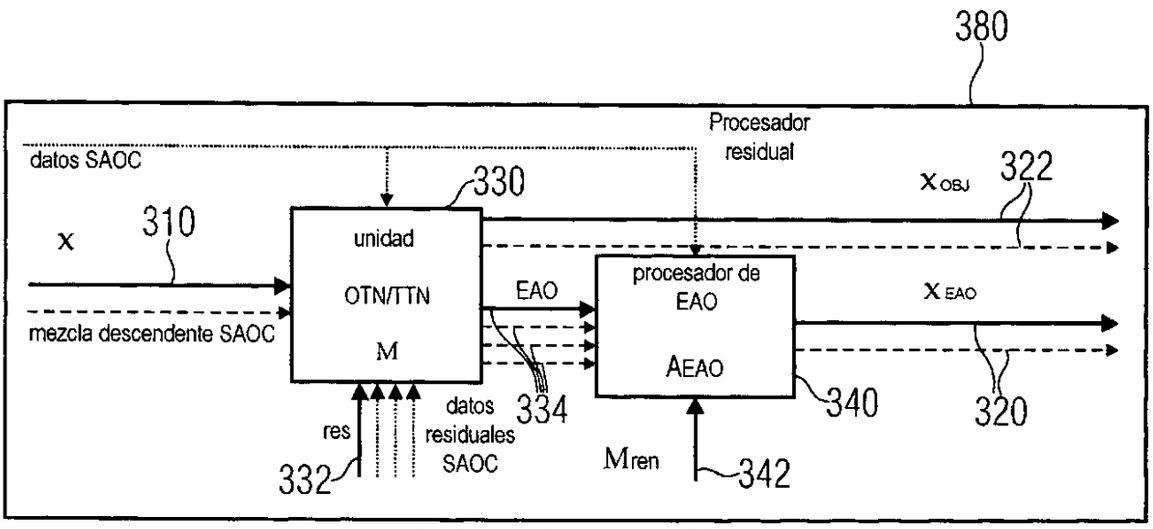
Estructura general de la arquitectura del transcodificador/ decodificador SAOC

FIG. 2



Arquitectura del procesador residual

FIG. 3 A



Arquitectura del procesador residual

FIG. 3 B

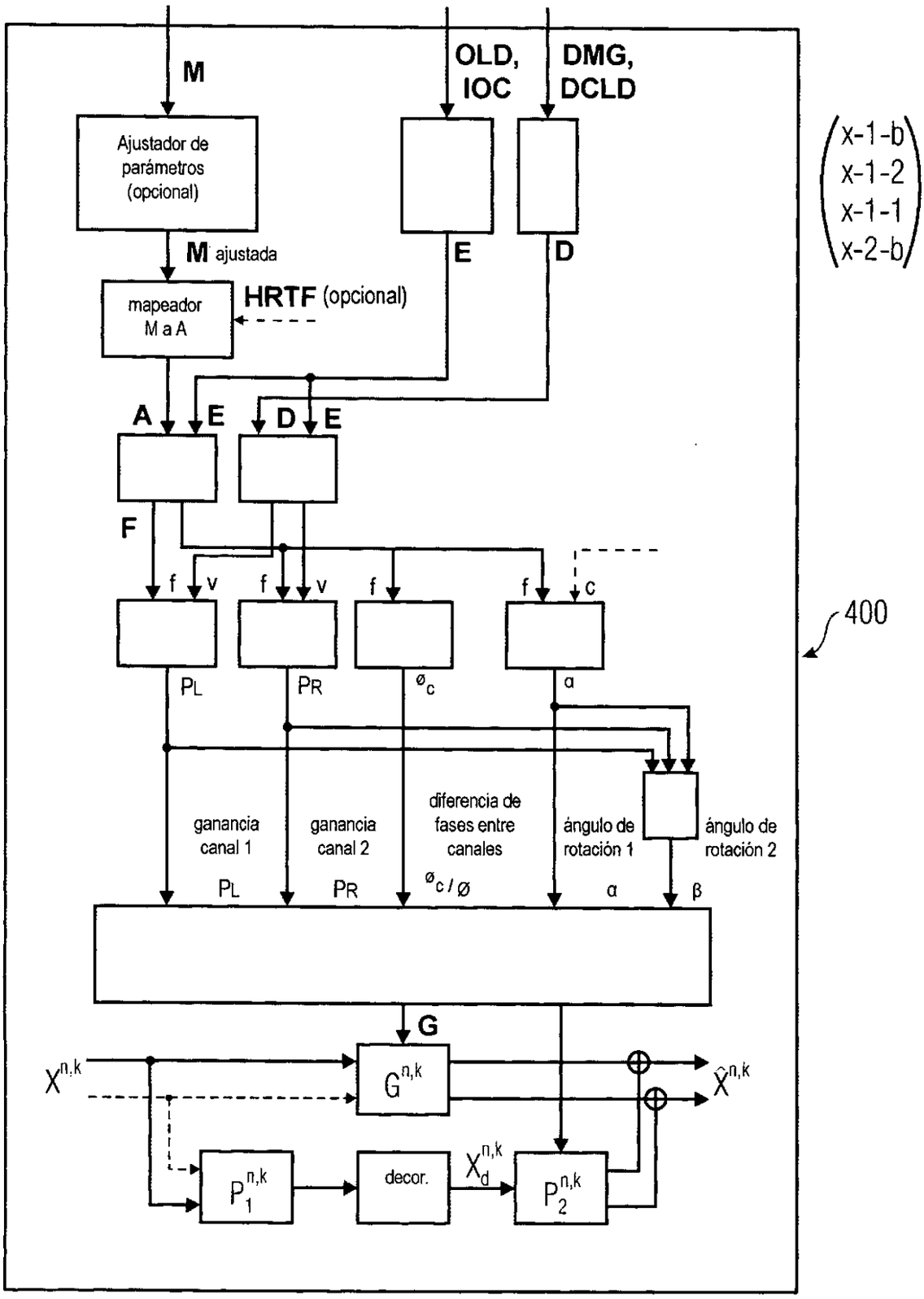

M
OLD, IOC
DMG, DCLD
Ajustador de parámetros (opcional)
M ajustada
E
D
mapeador M a A
HRTF (opcional)
A
E
D
E
F
f
v
f
v
f
f
c
400
PL
PR
øc
α
ganancia canal 1
ganancia canal 2
diferencia de fases entre canales
ángulo de rotación 1
ángulo de rotación 2
PL
PR
øc / ø
α
β
G
Xn,k
Gn,k
X̂n,k
P1n,k
decor.
Xdn,k
P2n,k
x-1-b
x-1-2
x-1-1
x-2-b

FIG. 4 A

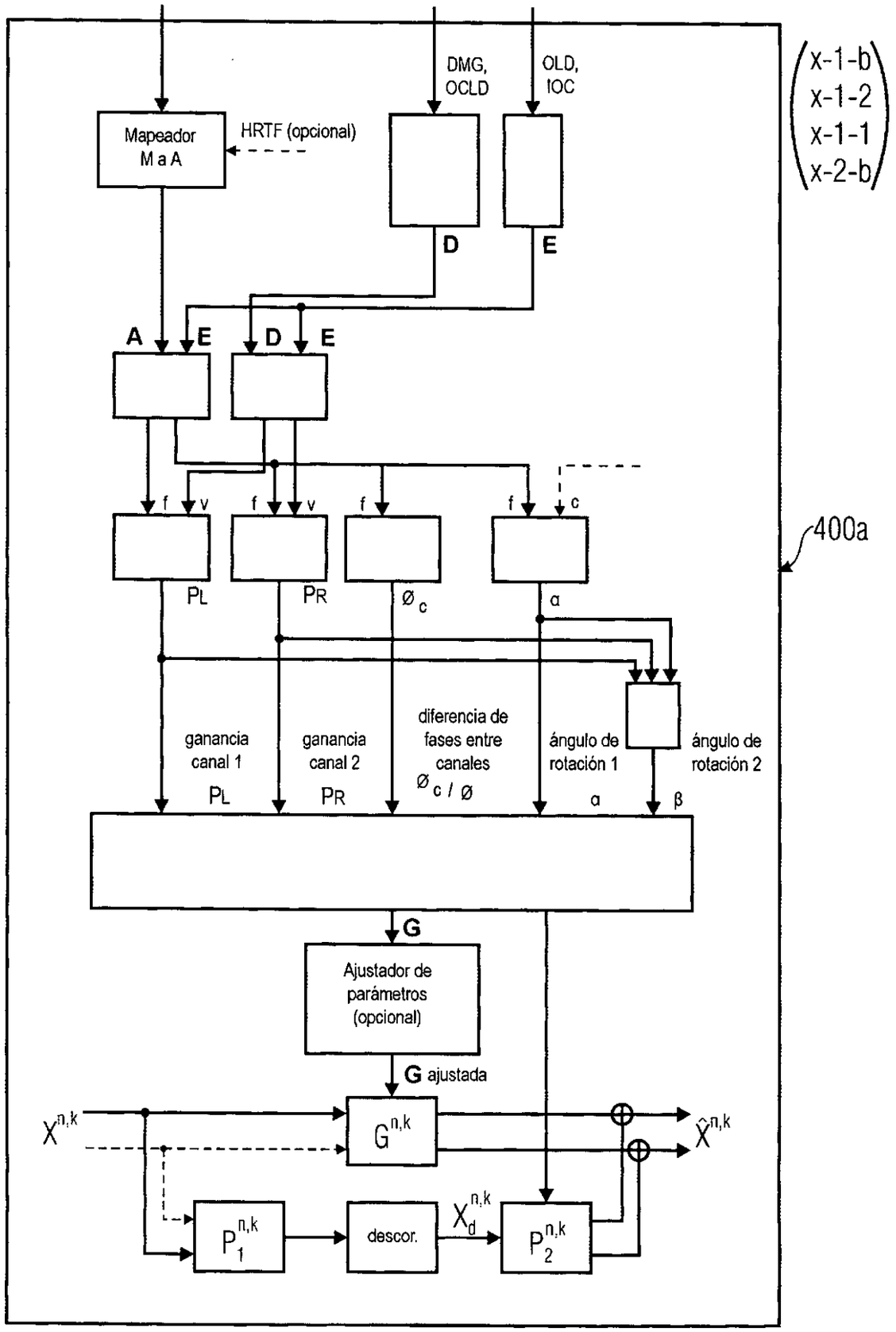
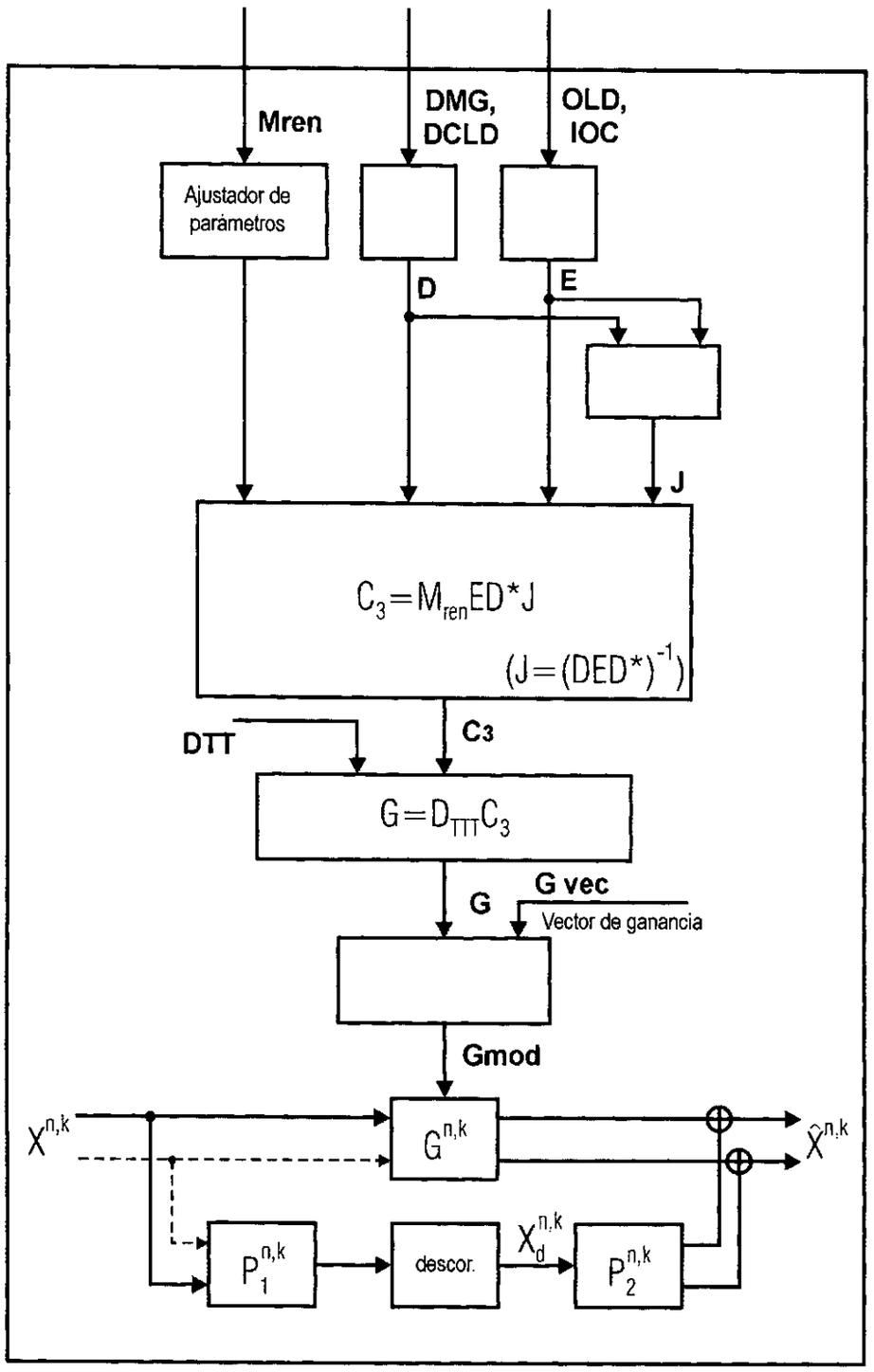
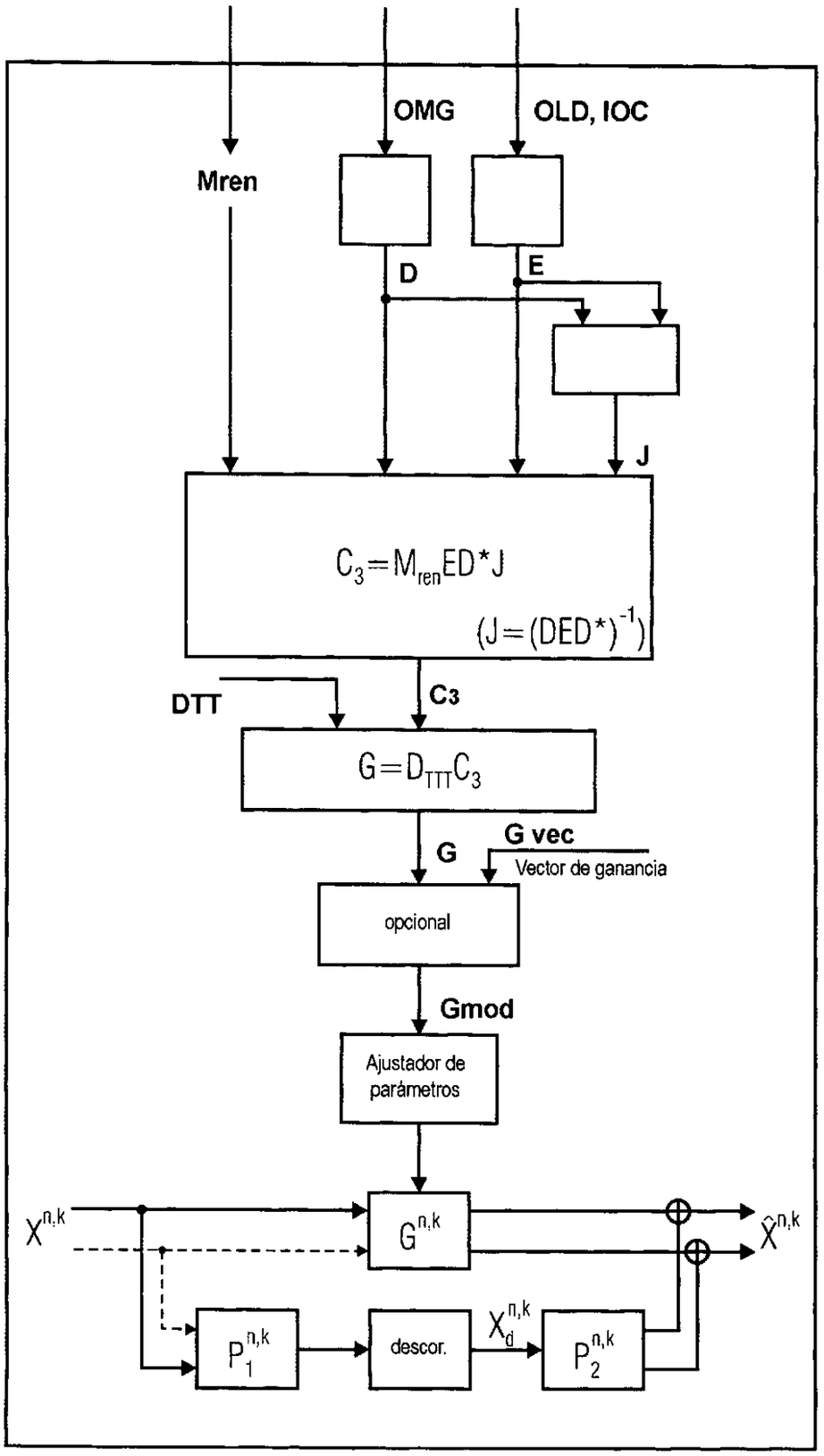

DMG, OCLD
OLD, IOC
Mapeador M a A
HRTF (opcional)
D
E
A
E
D
E
f
v
f
v
f
f
c
PL
PR
Ø c
α
ganancia canal 1
PL
ganancia canal 2
PR
diferencia de fases entre canales
Ø c / Ø
ángulo de rotación 1
α
ángulo de rotación 2
β
G
Ajustador de parámetros (opcional)
G ajustada
X n,k
G n,k
X̂ n,k
P 1 n,k
descor.
X d n,k
P 2 n,k
x-1-b
x-1-2
x-1-1
x-2-b
400a

FIG. 4 B



(x-2-5, x-2-2, x-2-1)

FIG. 4 C



(X-2-2)
(X-2-1)

FIG. 4 D

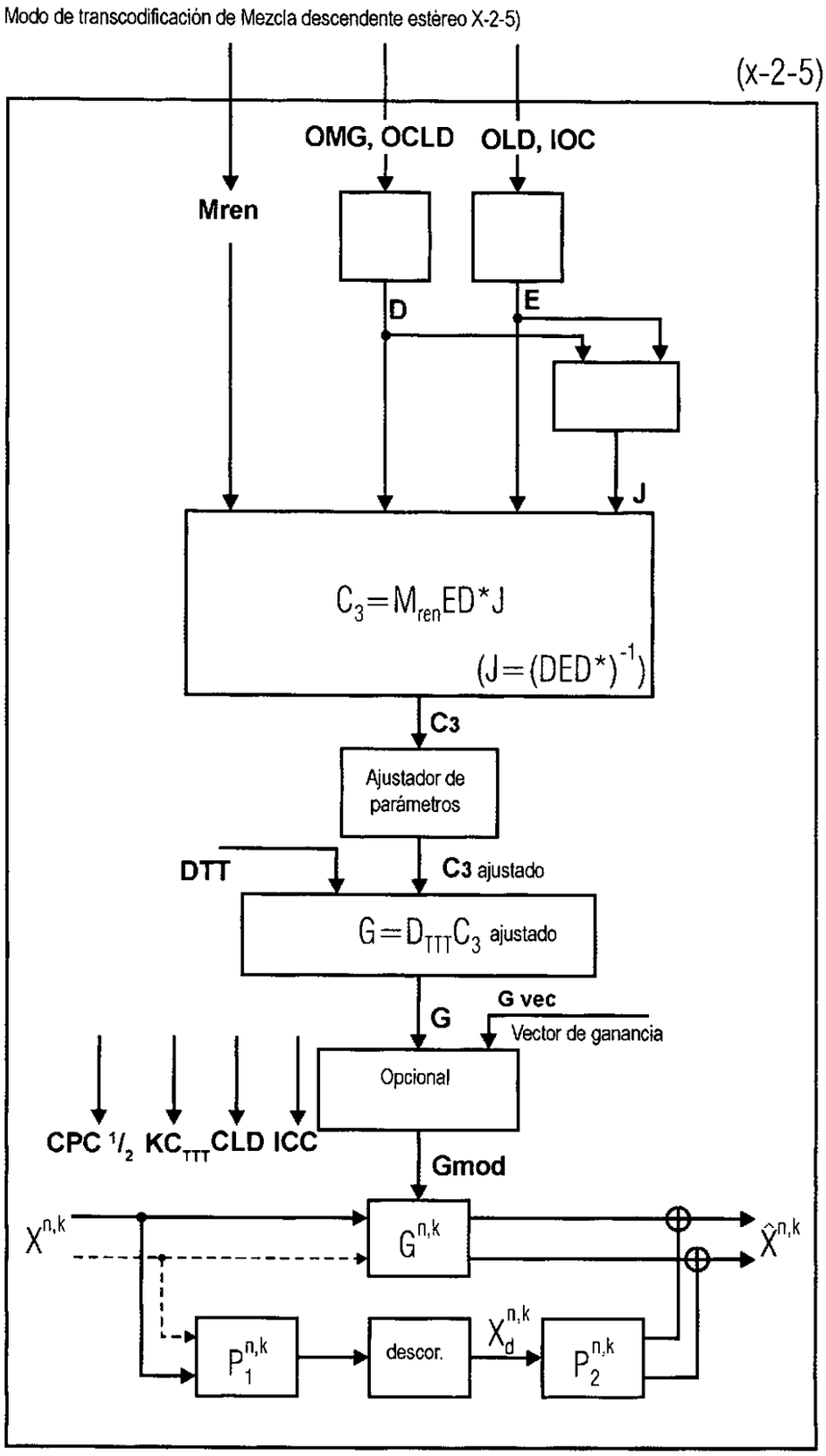

Modo de transcodificación de Mezcla descendente estéreo X-2-5)
(X-2-5)
OMG, OCLD
OLD, IOC
Mren
D
E
J
$C_3 = M_{ren} ED^* J$
$(J = (DED^*)^{-1})$
C3
Ajustador de parámetros
DTT
C3 ajustado
$G = D_{TTT} C_3$ ajustado
G
G vec
Vector de ganancia
Opcional
CPC 1/2 KC_{TTT} CLD ICC
Gmod
$X^{n,k}$
$G^{n,k}$
$\hat{X}^{n,k}$
$P_1^{n,k}$
descor.
$X_d^{n,k}$
$P_2^{n,k}$

FIG. 4 E

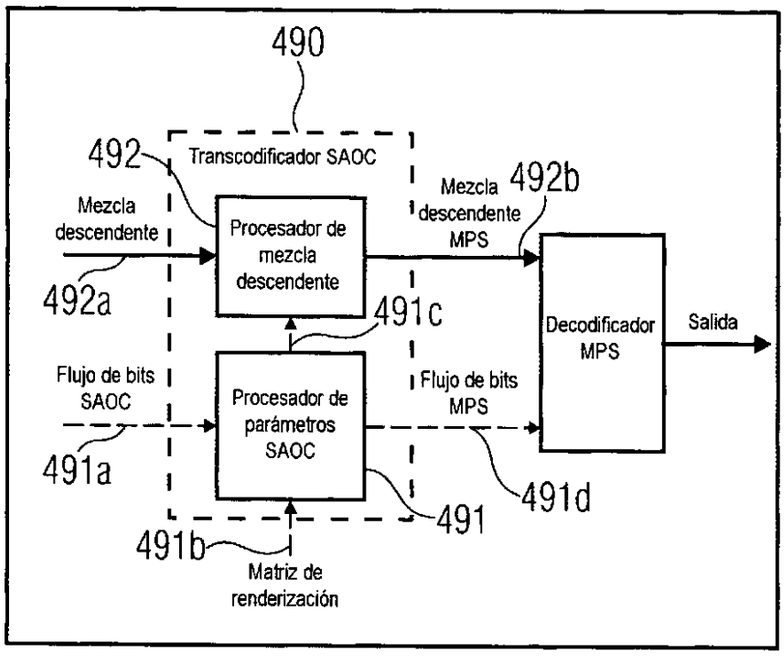

490
492
Transcodificador SAOC
Mezcla descendente
Procesador de mezcla descendente
Mezcla descendente MPS
492b
492a
491c
Decodificador MPS
Salida
Flujo de bits SAOC
Procesador de parámetros SAOC
Flujo de bits MPS
491a
491d
491
491b
Matriz de renderización

FIG. 4 F

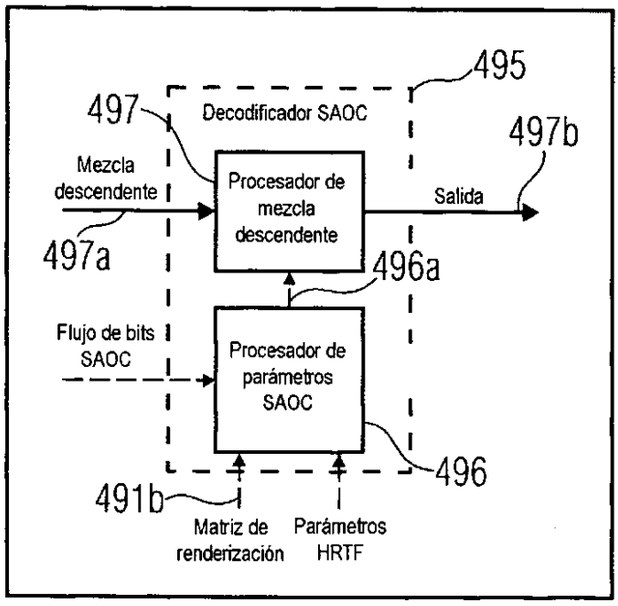

495
497
Decodificador SAOC
497b
Mezcla descendente
Procesador de mezcla descendente
Salida
497a
496a
Flujo de bits SAOC
Procesador de parámetros SAOC
496
491b
Matriz de renderización
Parámetros HRTF

FIG. 4 G

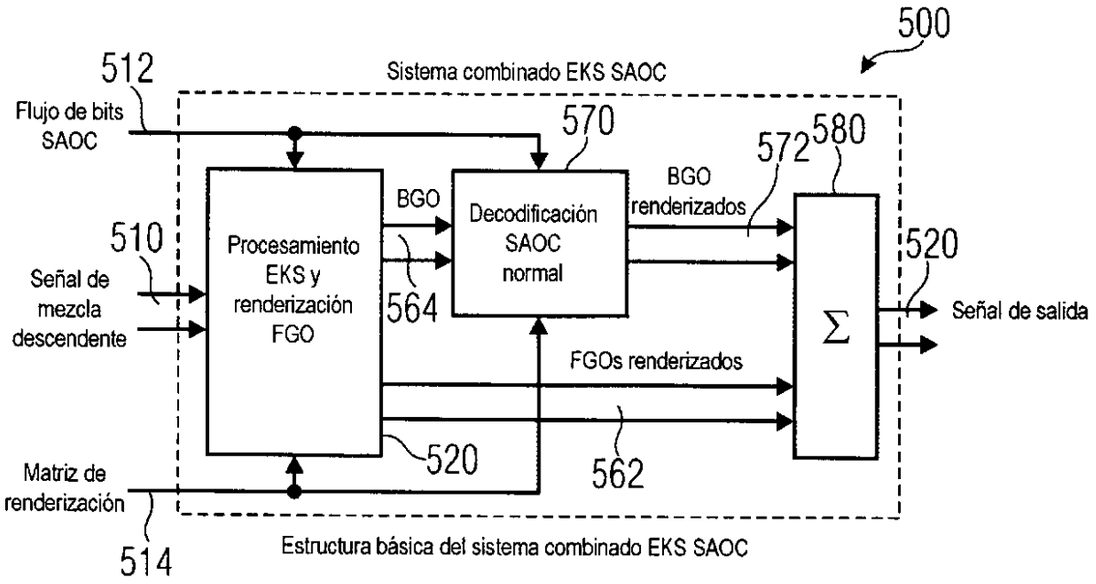

500
512
Sistema combinado EKS SAOC
Flujo de bits SAOC
570
572
580
BGO
BGO renderizados
510
Procesamiento EKS y renderización FGO
Decodificación SAOC normal
520
Señal de mezcla descendente
564
Σ
Señal de salida
FGOs renderizados
520
562
Matriz de renderización
514
Estructura básica del sistema combinado EKS SAOC

FIG. 5 A

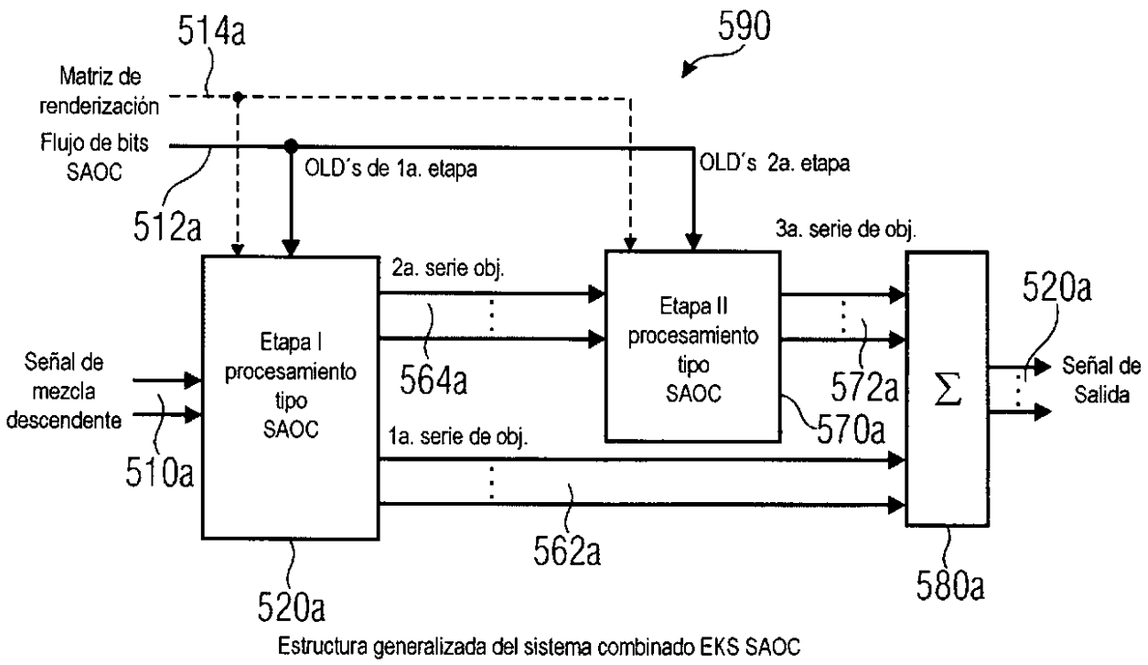

514a
590
Matriz de renderización
Flujo de bits SAOC
OLD´s de 1a. etapa
OLD´s 2a. etapa
512a
3a. serie de obj.
2a. serie obj.
520a
Etapa I procesamiento tipo SAOC
Etapa II procesamiento tipo SAOC
Señal de mezcla descendente
564a
572a
Σ
Señal de Salida
570a
1a. serie de obj.
510a
562a
520a
580a
Estructura generalizada del sistema combinado EKS SAOC

FIG. 5 B

Nombre codificador	Descripción
Referencia oculta	escena de audio c/renderiz. ideal (referencia oculta)
Anclaje inferior	anclaje inferior @ 3.5 kHz
SAOC normal	Modo de decodificación SAOC normal
EKS actual	Modo actual SAOC EKS
Sistema combinado	Sistema combinado (EKS + SAOC) (sistema propuesto)

FIG. 6 A

Sistema	Flujo de bits	Renderización Karaoke	Renderización Clásica
Sistema Combinado	SAOC A + datos res. A datos SAOC	X	X
SAOC normal		(X)	X
EKS actual	SAOC B + datos res. B	X	-

FIG. 6 B

Elemento	Tipo	Matriz de renderización	Matriz de mezcla descendente
Pop	Karaoke	[0.6 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.038 0.0 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.038]	[0.0 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.6 0.6 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.6]
	Clásico	[0.6 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.6 0.0 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.6]	
Rock	Karaoke	[1.0 0.0 0.707 0.707 0.000 0.038 0.0 1.0 0.707 0.000 0.707 0.038]	[1.0 0.0 0.707 0.707 0.000 0.707 0.0 1.0 0.707 0.000 0.707 0.707]
	Clásico	[1.0 0.0 0.6 0.0 0.707 0.000 0.4 0.0 1.0 0.4 0.0 0.000 0.707 0.6]	

FIG. 6 C

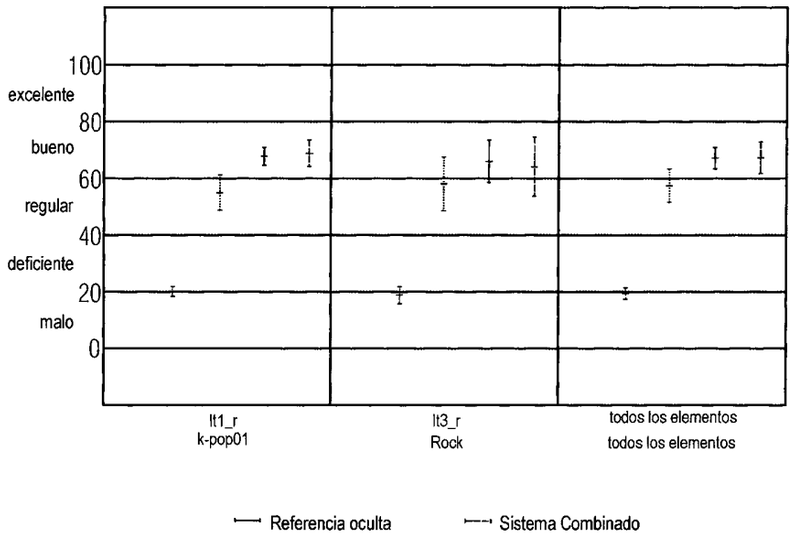

100
excelente
80
bueno
60
regular
40
deficiente
20
malo
0
lt1_r
k-pop01
lt3_r
Rock
todos los elementos
todos los elementos
Referencia oculta
Sistema Combinado

FIG. 6 D

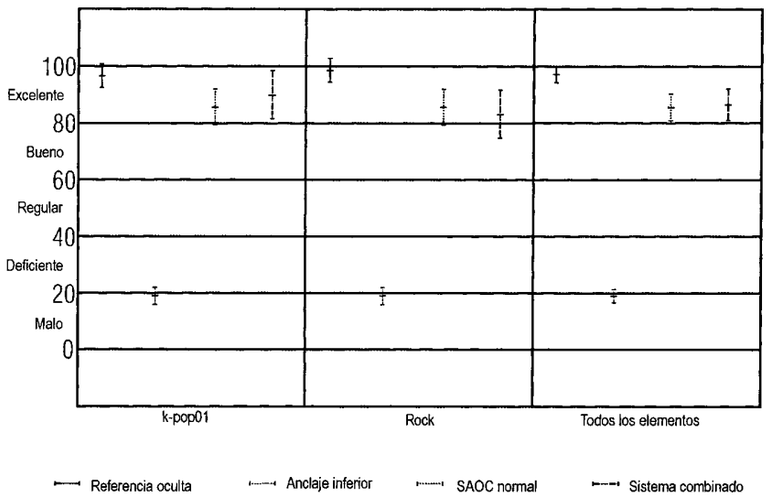

100
Excelente
80
Bueno
60
Regular
40
Deficiente
20
Malo
0
k-pop01
Rock
Todos los elementos
Referencia oculta
Anclaje inferior
SAOC normal
Sistema combinado

FIG. 6 E

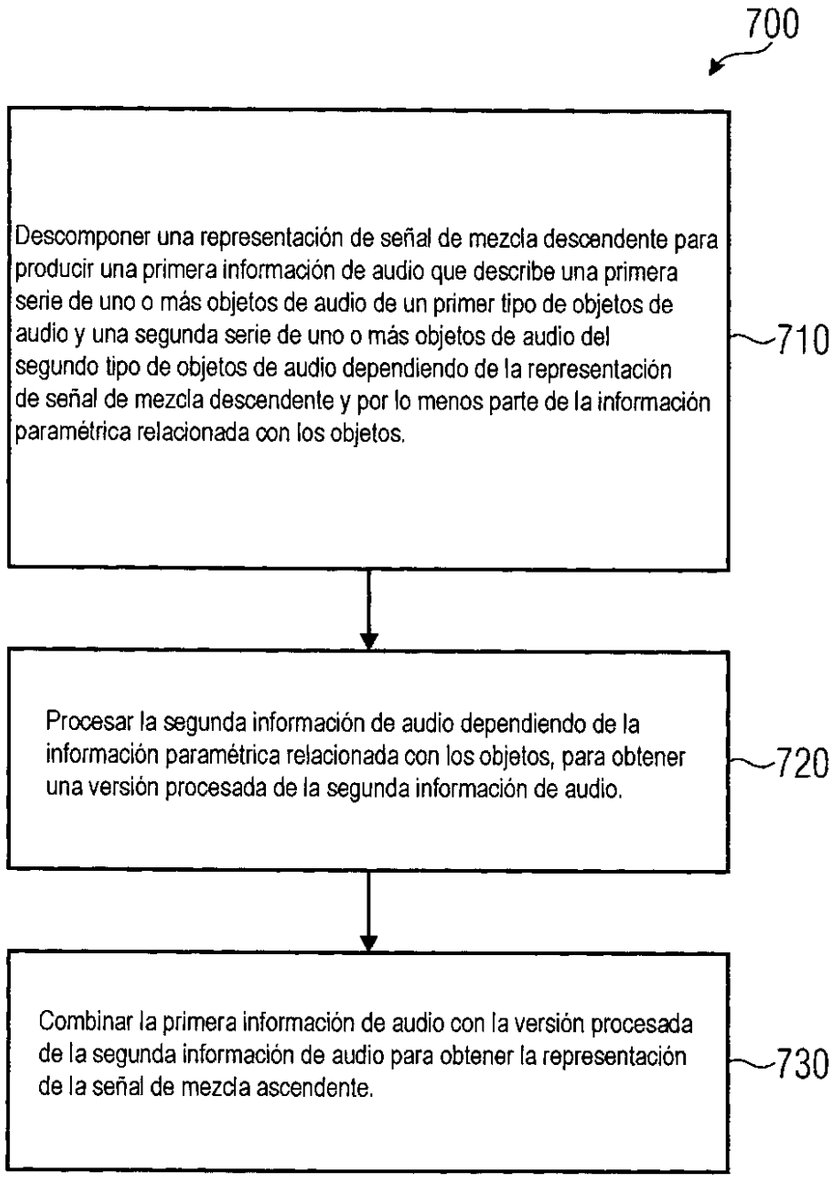

700
Descomponer una representación de señal de mezcla descendente para producir una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda serie de uno o más objetos de audio del segundo tipo de objetos de audio dependiendo de la representación de señal de mezcla descendente y por lo menos parte de la información paramétrica relacionada con los objetos.
710
Procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, para obtener una versión procesada de la segunda información de audio.
720
Combinar la primera información de audio con la versión procesada de la segunda información de audio para obtener la representación de la señal de mezcla ascendente.
730

FIG. 7

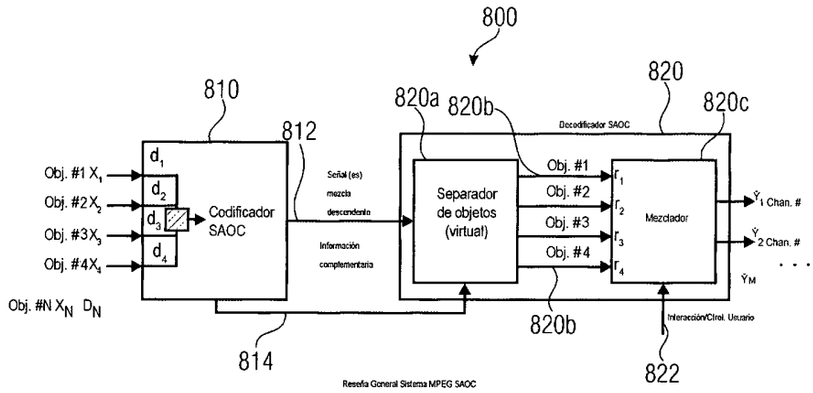

800
810
812
820
820a
820b
820c
Decodificador SAOC
Obj. #1 X1
Obj. #2 X2
Obj. #3 X3
Obj. #4 X4
Obj. #N XN DN
d1
d2
d3
d4
Codificador SAOC
Señal (es) mezcla descendente
Información complementaria
Separador de objetos (virtual)
Obj. #1
Obj. #2
Obj. #3
Obj. #4
r1
r2
r3
r4
Mezclador
Ŷ1 Chan. #
Ŷ2 Chan. #
ŶM
Interacción/Ctrol. Usuario
814
820b
822
Reseña General Sistema MPEG SAOC

FIG. 8

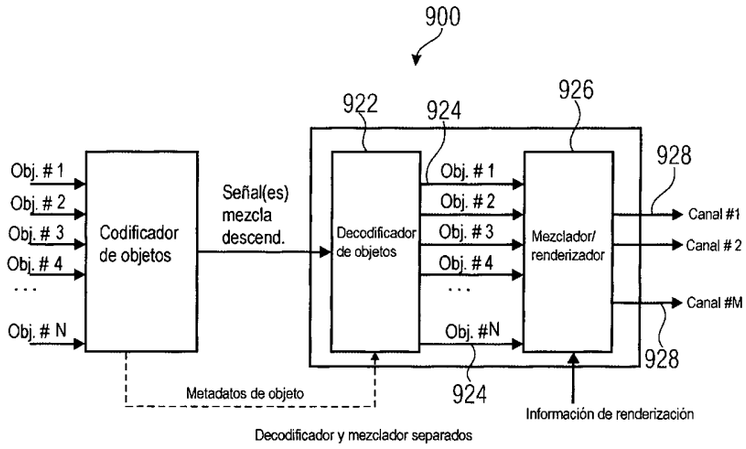
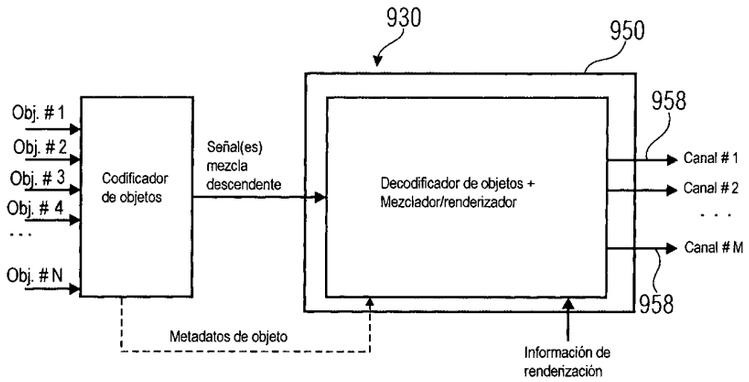

900
922
924
926
928
Obj. # 1
Obj. # 2
Obj. # 3
Obj. # 4
...
Obj. # N
Codificador de objetos
Señal(es) mezcla descend.
Decodificador de objetos
Obj. # 1
Obj. # 2
Obj. # 3
Obj. # 4
...
Obj. #N
Mezclador/ renderizador
Canal #1
Canal # 2
Canal #M
928
924
Metadatos de objeto
Información de renderización
Decodificador y mezclador separados

FIG. 9 A



Mezclador y Decodificador integrados

FIG. 9 B

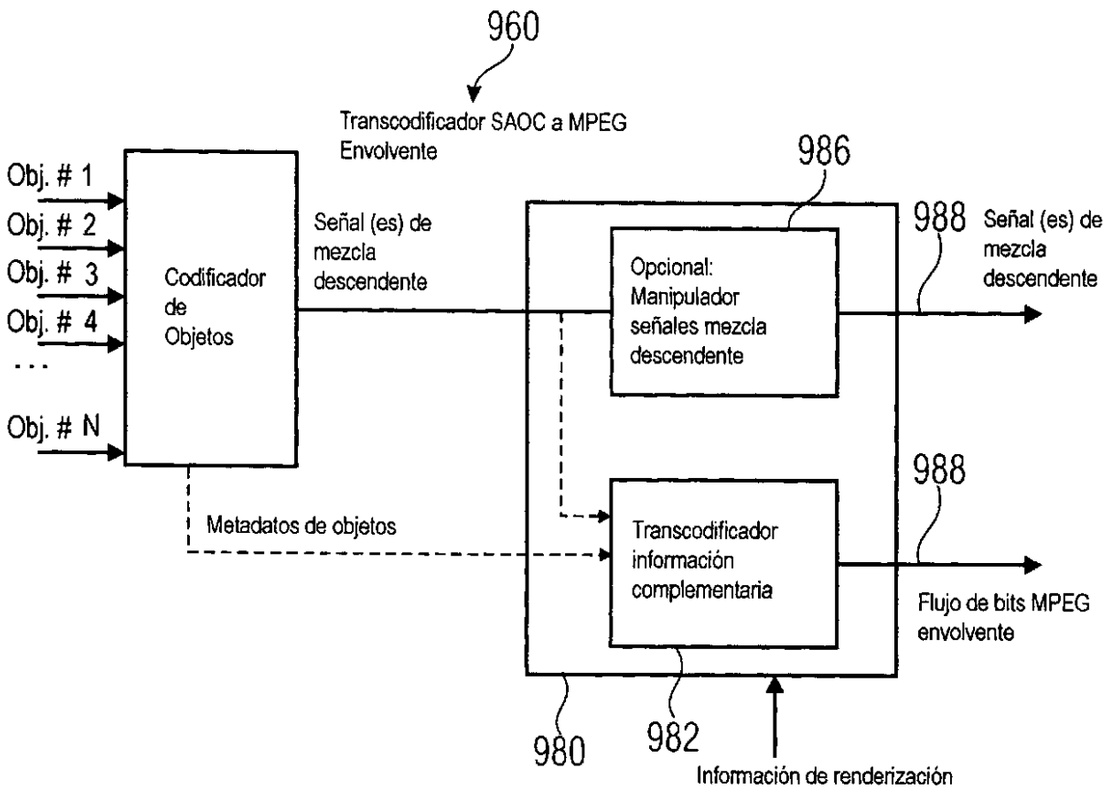

960
Transcodificador SAOC a MPEG
Envolvente
Obj. # 1
Obj. # 2
Obj. # 3
Obj. # 4
...
Obj. # N
Codificador
de
Objetos
Señal (es) de
mezcla
descendente
986
Opcional:
Manipulador
señales mezcla
descendente
988
Señal (es) de
mezcla
descendente
Metadatos de objetos
Transcodificador
información
complementaria
988
Flujo de bits MPEG
envolvente
980
982
Información de renderización

FIG. 9 C

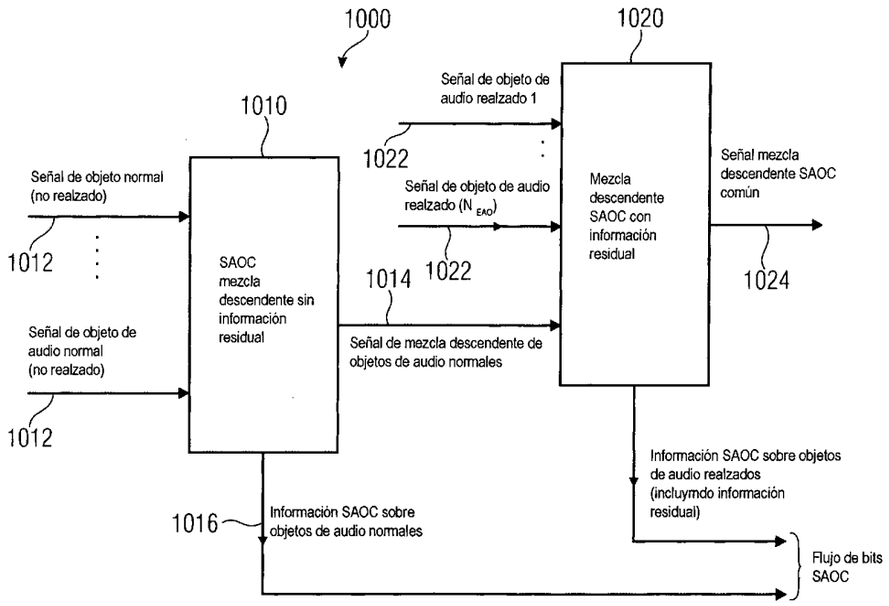

1000
1020
1010
Señal de objeto de audio realzado 1
1022
Señal de objeto de audio realzado (N_{EAO})
1022
Señal de objeto normal (no realzado)
1012
SAOC mezcla descendente sin información residual
1014
Señal de mezcla descendente de objetos de audio normales
Mezcla descendente SAOC con información residual
Señal mezcla descendente SAOC común
1024
Señal de objeto de audio normal (no realzado)
1012
Información SAOC sobre objetos de audio realzados (incluymdo información residual)
1016
Información SAOC sobre objetos de audio normales
Flujo de bits SAOC

FIG. 10

Audio Signal Decoder, Method for Decoding an Audio Signal and Computer Program using Cascaded Audio Object Processing Stages

Description

Technical Field

Embodiments according to the invention are related to an audio signal decoder for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information.

Further embodiments according to the invention are related to a method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information.

Further embodiments according to the invention are related to a computer program.

Some embodiments according to the invention are related to an enhanced Karaoke/Solo SAOC system.

Background of the Invention

In modern audio systems, it is desired to transfer and store audio information in a bitrate-efficient way. In addition, it is often desired to reproduce an audio content using a plurality of two or even more speakers, which are spatially distributed in a room. In such cases, it is desired to exploit the capabilities of such a multi-speaker arrangement to allow for a user to spatially identify different audio contents or different items of a single audio content. This may be achieved by individually distributing the different audio contents to the different speakers.

In other words, in the art of audio processing, audio transmission and audio storage, there is an increasing desire to handle multi-channel contents in order to improve the hearing impression. Usage of multi-channel audio content brings along significant improvements for the user. For example, a 3-dimensional hearing impression can be obtained, which brings along an improved user satisfaction in entertainment applications. However, multi-channel audio contents are also useful in professional environments, for example in telephone conferencing applications, because the speaker intelligibility can be improved by using a multi-channel audio playback.

However, it is also desirable to have a good tradeoff between audio quality and bitrate requirements in order to avoid an excessive resource load caused by multi-channel applications.

Recently, parametric techniques for the bitrate-efficient transmission and/or storage of audio scenes containing multiple audio objects has been proposed, for example, Binaural Cue Coding (Type I) (see, for example reference [BCC]), Joint Source Coding (see, for example, reference [JSC]), and MPEG Spatial Audio Object Coding (SAOC) (see, for example, references [SAOC1], [SAOC2]).

These techniques aim at perceptually reconstructing the desired output audio scene rather than by a waveform match.

Fig. 8 shows a system overview of such a system (here: MPEG SAOC). The MPEG SAOC system 800 shown in Fig. 8 comprises an SAOC encoder 810 and an SAOC decoder 820. The SAOC encoder 810 receives a plurality of object signals x_1 to x_N, which may be represented, for example, as time-domain signals or as time-frequency-domain signals (for example, in the form of a set of transform coefficients of a Fourier-type transform, or in the form of QMF subband signals). The SAOC encoder 810 typically also receives downmix coefficients d_1 to d_N, which are associated with the object signals x_1 to x_N. Separate sets of downmix coefficients may be available for each channel of the downmix signal. The SAOC encoder 810 is typically configured to obtain a channel of the downmix signal by combining the object signals x_1 to x_N in accordance with the associated downmix coefficients d_1 to d_N. Typically, there are less downmix channels than object signals x_1 to x_N. In order to allow (at least approximately) for a separation (or separate treatment) of the object signals at the side of the SAOC decoder 820, the SAOC encoder 810 provides both the one or more downmix signals (designated as downmix channels) 812 and a side information 814. The side information 814 describes characteristics of the object signals x_1 to x_N, in order to allow for a decoder-sided object-specific processing.

The SAOC decoder 820 is configured to receive both the one or more downmix signals 812 and the side information 814. Also, the SAOC decoder 820 is typically configured to receive a user interaction information and/or a user control information 822, which describes a desired rendering setup. For example, the user interaction information/user control information 822 may describe a speaker setup and the desired spatial placement of the objects provided by the object signals x_1 to x_N.

The SAOC decoder 820 is configured to provide, for example, a plurality of decoded upmix channel signals $\hat{y}_1$ to $\hat{y}_M$. The upmix channel signals may for example be associated with individual speakers of a multi-speaker rendering arrangement. The SAOC decoder 820 may, for example, comprise an object separator 820a, which is configured to reconstruct, at least approximately, the object signals x_1 to x_N on the basis of the one or more downmix signals 812 and the side information 814, thereby obtaining reconstructed object signals 820b. However, the reconstructed object signals 820b may deviate somewhat from the original object signals x_1 to x_N, for example, because the side information 814 is not quite sufficient for a perfect reconstruction due to the bitrate constraints. The SAOC decoder 820 may further comprise a mixer 820c, which may be configured to receive the reconstructed object signals 820b and the user interaction information/user control information 822, and to provide, on the basis thereof, the upmix channel signals $\hat{y}_1$ to $\hat{y}_M$. The mixer 820c may be configured to use the user interaction information /user control information 822 to determine the contribution of the individual reconstructed object signals 820b to the upmix channel signals $\hat{y}_1$ to $\hat{y}_M$. The user interaction information/user control information 822 may, for example, comprise rendering parameters (also designated as rendering coefficients), which determine the contribution of the individual reconstructed object signals 820b to the upmix channel signals $\hat{y}_1$ to $\hat{y}_M$.

However, it should be noted that in many embodiments, the object separation, which is indicated by the object separator 820a in Fig. 8, and the mixing, which is indicated by the mixer 820c in Fig. 8, are performed in one single step. For this purpose, overall parameters may be computed which describe a direct mapping of the one or more downmix signals 812 onto the upmix channel signals $\hat{y}_1$ to $\hat{y}_M$. These parameters may be computed on the basis of the side information 814 and the user interaction information/user control information 822.

Taking reference now to Figs. 9a, 9b and 9c, different apparatus for obtaining an upmix signal representation on the basis of a downmix signal representation and object-related side information will be described. Fig. 9a shows a block schematic diagram of an MPEG SAOC system 900 comprising an SAOC decoder 920. The SAOC decoder 920 comprises, as separate functional blocks, an object decoder 922 and a mixer/renderer 926. The object decoder 922 provides a plurality of reconstructed object signals 924 in dependence on the downmix signal representation (for example, in the form of one or more downmix signals represented in the time domain or in the time-frequency-domain) and object-related side information (for example, in the form of object meta data). The mixer/renderer 926 receives the reconstructed object signals 924 associated with a plurality of N objects and provides, on the basis thereof, one or more upmix channel signals 928. In the SAOC

decoder 920, the extraction of the object signals 924 is performed separately from the mixing/rendering which allows for a separation of the object decoding functionality from the mixing/rendering functionality but brings along a relatively high computational complexity.

Taking reference now to Fig. 9b, another MPEG SAOC system 930 will be briefly discussed, which comprises an SAOC decoder 950. The SAOC decoder 950 provides a plurality of upmix channel signals 958 in dependence on a downmix signal representation (for example, in the form of one or more downmix signals) and an object-related side information (for example, in the form of object meta data). The SAOC decoder 950 comprises a combined object decoder and mixer/renderer, which is configured to obtain the upmix channel signals 958 in a joint mixing process without a separation of the object decoding and the mixing/rendering, wherein the parameters for said joint upmix process are dependent on both, the object-related side information and the rendering information. The joint upmix process also depends on the downmix information, which is considered to be part of the object-related side information.

To summarize the above, the provision of the upmix channel signals 928, 958 can be performed in a one step process or a two-step process.

Taking reference now to Fig. 9c, an MPEG SAOC system 960 will be described. The SAOC system 960 comprises an SAOC to MPEG Surround transcoder 980, rather than an SAOC decoder.

The SAOC to MPEG Surround transcoder comprises a side information transcoder 982, which is configured to receive the object-related side information (for example, in the form of object meta data) and, optionally, information on the one or more downmix signals and the rendering information. The side information transcoder is also configured to provide an MPEG Surround side information 984 (for example, in the form of an MPEG Surround bitstream) on the basis of a received data. Accordingly, the side information transcoder 982 is configured to transform an object-related (parametric) side information, which is relieved from the object encoder, into a channel-related (parametric) side information 984, taking into consideration the rendering information and, optionally, the information about the content of the one or more downmix signals.

Optionally, the SAOC to MPEG Surround transcoder 980 may be configured to manipulate the one or more downmix signals, described, for example, by the downmix signal representation, to obtain a manipulated downmix signal representation 988. However, the

downmix signal manipulator 986 may be omitted, such that the output downmix signal representation 988 of the SAOC to MPEG Surround transcoder 980 is identical to the input downmix signal representation of the SAOC to MPEG Surround transcoder. The downmix signal manipulator 986 may, for example, be used if the channel-related MPEG Surround side information 984 would not allow to provide a desired hearing impression on the basis of the input downmix signal representation of the SAOC to MPEG Surround transcoder 980, which may be the case in some rendering constellations.

Accordingly, the SAOC to MPEG Surround transcoder 980 provides the downmix signal representation 988 and the MPEG Surround bitstream 984 such that a plurality of upmix channel signals, which represent the audio objects in accordance with the rendering information input to the SAOC to MPEG Surround transcoder 980 can be generated using an MPEG Surround decoder which receives the MPEG Surround bitstream 984 and the downmix signal representation 988.

To summarize the above, different concepts for decoding SAOC-encoded audio signals can be used. In some cases, an SAOC decoder is used, which provides upmix channel signals (for example, upmix channel signals 928, 958) in dependence on the downmix signal representation and the object-related parametric side information. Examples for this concept can be seen in Figs. 9a and 9b. Alternatively, the SAOC-encoded audio information may be transcoded to obtain a downmix signal representation (for example, a downmix signal representation 988) and a channel-related side information (for example, the channel-related MPEG Surround bitstream 984), which can be used by an MPEG Surround decoder to provide the desired upmix channel signals.

In the MPEG SAOC system 800, a system overview of which is given in Fig. 8, the general processing is carried out in a frequency selective way and can be described as follows within each frequency band:

- N input audio object signals x_1 to x_N are downmixed as part of the SAOC encoder processing. For a mono downmix, the downmix coefficients are denoted by d_1 to d_N. In addition, the SAOC encoder 810 extracts side information 814 describing the characteristics of the input audio objects. For MPEG SAOC, the relations of the object powers with respect to each other are the most basic form of such a side information.

- Downmix signal (or signals) 812 and side information 814 are transmitted and/or stored. To this end, the downmix audio signal may be compressed using well-known

perceptual audio coders such as MPEG-1 Layer II or III (also known as ".mp3"), MPEG Advanced Audio Coding (AAC), or any other audio coder.

- On the receiving end, the SAOC decoder 820 conceptually tries to restore the original object signal ("object separation") using the transmitted side information 814 (and, naturally, the one or more downmix signals 812). These approximated object signals (also designated as reconstructed object signals 820b) are then mixed into a target scene represented by M audio output channels (which may, for example, be represented by the upmix channel signals $\hat{y}_1$ to $\hat{y}_M$) using a rendering matrix. For a mono output, the rendering matrix coefficients are given by r_1 to r_N.

- Effectively, the separation of the object signals is rarely executed (or even never executed), since both the separation step (indicated by the object separator 820a) and the mixing step (indicated by the mixer 820c) are combined into a single transcoding step, which often results in an enormous reduction in computational complexity.

It has been found that such a scheme is tremendously efficient, both in terms of transmission bitrate (it is only necessary to transmit a few downmix channels plus some side information instead of N discrete object audio signals or a discrete system) and computational complexity (the processing complexity relates mainly to the number of output channels rather than the number of audio objects). Further advantages for the user on the receiving end include the freedom of choosing a rendering setup of his/her choice (mono, stereo, surround, virtualized headphone playback, and so on) and the feature of user interactivity: the rendering matrix, and thus the output scene, can be set and changed interactively by the user according to will, personal preference or other criteria. For example, it is possible to locate the talkers from one group together in one spatial area to maximize discrimination from other remaining talkers. This interactivity is achieved by providing a decoder user interface.

For each transmitted sound object, its relative level and (for non-mono rendering) spatial position of rendering can be adjusted. This may happen in real-time as the user changes the position of the associated graphical user interface (GUI) sliders (for example: object level = +5dB, object position = -30deg).

However, it has been found that it is difficult to handle audio objects of different audio object types in such a system. In particular, it has been found that it is difficult to process audio objects of different audio object types, for example, audio objects to which different

side information is associated, if the total number of audio objects to be processed is not predetermined.

In view of this situation, it is an objective of the present invention to create a concept, which allows for a computationally-efficient and flexible decoding of an audio signal comprising a downmix signal representation and an object-related parametric information, wherein the object-related parametric information describes audio objects of two or more different audio object types.

Summary of the Invention

This objective is achieved by an audio signal decoder for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, a method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, and a computer program, as defined by the independent claims.

An embodiment according to the invention creates an audio signal decoder for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information. The audio signal decoder comprises an object separator configured to decompose the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information. The audio signal decoder also comprises an audio signal processor configured to receive the second audio information and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information. The audio signal decoder also comprises an audio signal combiner configured to combine the first audio information with the processed version of the second audio information to obtain the upmix signal representation.

It is a key idea of the present invention that an efficient processing of different types of audio objects can be obtained in a cascaded structure, which allows for a separation of the different types of audio objects using at least a part of the object-related parametric information in a first processing step performed by the object separator, and which allows for an additional spatial processing in a second processing step performed in dependence on at least a part of the object-related parametric information by the audio signal processor.

It has been found that extracting a second audio information, which comprises audio objects of the second audio object type, from a downmix signal representation can be performed with a moderate complexity even if there is a larger number of audio objects of the second audio object type. In addition, it has been found that a spatial processing of the audio objects of the second audio type can be performed efficiently once the second audio information is separated from the first audio information describing the audio objects of the first audio object type.

Additionally, it has been found that the processing algorithm performed by the object separator for separating the first audio information and the second audio information can be performed with comparatively small complexity if the object-individual processing of the audio objects of the second audio object type is postponed to the audio signal processor and not performed at the same time as the separation of the first audio information and the second audio information.

In a preferred embodiment, the audio signal decoder is configured to provide the upmix signal representation in dependence on the downmix signal representation, the object-related parametric information and a residual information associated to a sub-set of audio objects represented by the downmix signal representation. In this case, the object separator is configured to decompose the downmix signal representation to provide the first audio information describing the first set of one or more audio objects (for example, foreground objects FGO) of the first audio object type to which residual information is associated and the second audio information describing the second set of one or more audio objects (for example, background objects BGO) of the second audio object type to which no residual information is associated in dependence on the downmix signal representation and using at least part of the object-related parametric information and the residual information.

This embodiment is based on the finding that a particularly accurate separation between the first audio information describing the first set of audio objects of the first audio object type and the second audio information describing the second set of audio objects of the second audio object type can be obtained by using a residual information in addition to the object-related parametric information. It has been found that the mere use of the object-related parametric information would result in distortions in many cases, which can be reduced significantly or even entirely eliminated by the use of residual information. The residual information describes, for example, a residual distortion, which is expected to remain if an audio object of the first audio object type is isolated merely using the object-related parametric information. The residual information is typically estimated by an audio signal encoder. By applying the residual information, the separation between the audio

objects of the first audio object type and the audio objects of the second audio object type can be improved.

This allows to obtain the first audio information and the second audio information with particularly good separation between the audio objects of the first audio object type and the audio objects of the second audio object type, which, in turn, allows to achieve a high-quality spatial processing of the audio objects of the second audio object type when processing the second audio information in the audio signal processor.

In a preferred embodiment, the object separator is therefore configured to provide the first audio information such that audio objects of the first audio object type are emphasized over audio objects of the second audio object type in the first audio information. The object separator is also configured to provide the second audio information such that audio objects of the second audio object type are emphasized over audio objects of the first audio object type in the second audio information.

In a preferred embodiment, the audio signal decoder is configured to perform a two-step processing, such that a processing of the second audio information in the audio signal processor is performed subsequently to a separation between the first audio information describing the first set of one or more audio objects of the first audio object type and the second audio information describing the second set of one or more audio objects of the second audio object type.

In a preferred embodiment, the audio signal processor is configured to process the second audio information in dependence on the object-related parametric information associated with the audio objects of the second audio object type and independent from the object-related parametric information associated with the audio objects of the first audio object type. Accordingly, a separate processing of the audio objects of the first audio object type and the audio objects of the second audio object type can be obtained.

In a preferred embodiment, the object separator is configured to obtain the first audio information and the second audio information using a linear combination of one or more downmix channels and one or more residual channels. In this case, the object separator is configured to obtain combination parameters for performing the linear combination in dependence on downmix parameters associated with the audio objects of the first audio object type and in dependence on channel prediction coefficients of the audio objects of the first audio object type. The computation of the channel prediction coefficients of the audio objects of the first audio object type may, for example, take into consideration the audio

objects of the second audio object type as a single, common audio object. Accordingly, a separation process can be performed with sufficiently small computational complexity, which may, for example, be almost independent from the number of audio objects of the second audio object type.

In a preferred embodiment, the object separator is configured to apply a rendering matrix to the first audio information to map object signals of the first audio information onto audio channels of the upmix audio signal representation. This can be done, because the object separator may be capable of extracting separate audio signals individually representing the audio objects of the first audio object type. Accordingly, it is possible to map the object signals of the first audio information directly onto the audio channels of the upmix audio signal representation.

In a preferred embodiment, the audio processor is configured to perform a stereo processing of the second audio information in dependence on a rendering information, an object-related covariance information and a downmix information, to obtain audio channels of the upmix audio signal representation.

Accordingly, the stereo processing of the audio objects of the second audio object type is separated from the separation between the audio objects of the first audio object type and the audio objects of the second audio object type. Thus, the efficient separation between audio objects of the first audio object type and audio objects of the second audio object type is not affected (or degraded) by the stereo processing, which typically leads to a distribution of audio objects over a plurality of audio channels without providing the high degree of object separation, which can be obtained in the object separator, for example, using the residual information.

In another preferred embodiment, the audio processor is configured to perform a post-processing of the second audio information in dependence on a rendering information, an object-related covariance information and a downmix information. This form of post-processing allows for a spatial placement of the audio objects of the second audio object type within an audio scene. Nevertheless, due to the cascaded concept, the computational complexity of the audio processor can be kept sufficiently small, because the audio processor does not need to consider the object-related parametric information associated with the audio objects of the first audio object type.

In addition, different types of processing can be performed by the audio processor, like, for example, a mono-to-binaural processing, a mono-to-stereo processing, a stereo-to-binaural processing or a stereo-to-stereo processing.

In a preferred embodiment, the object separator is configured to treat audio objects of the second audio object type, to which no residual information is associated, as a single audio object. In addition, the audio signal processor is configured to consider object-specific rendering parameters to adjust contributions of the objects of the second audio object type to the upmix signal representation. Thus, the audio objects of the second audio object type are considered as a single audio object by the object separator, which significantly reduces the complexity of the object separator and also allows to have a unique residual information, which is independent from the rendering parameters associated with the audio objects of the second audio object type.

In a preferred embodiment, the object separator is configured to obtain a common object-level difference value for a plurality of audio objects of the second audio object type. The object separator is configured to use the common object-level difference value for a computation of channel prediction coefficients. In addition, the object separator is configured to use the channel prediction coefficients to obtain one or two audio channels representing the second audio information. For obtaining a common object-level difference value, the audio objects of the second audio object type can be handled efficiently as a single audio object by the object separator.

In a preferred embodiment, the object separator is configured to obtain a common object level difference value for a plurality of audio objects of the second audio object type and the object separator is configured to use the common object-level difference value for a computation of entries of an energy-mode mapping matrix. The object separator is configured to use the energy-mode mapping matrix to obtain the one or more audio channels representing the second audio information. Again, the common object level difference value allows for a computationally efficient common treating of the audio objects of the second audio object type by the object separator.

In a preferred embodiment, the object separator is configured to selectively obtain a common inter-object correlation value associated to the audio objects of the second audio object type in dependence on the object-related parametric information if it is found that there are two audio objects of the second audio object type and to set the inter-object correlation value associated to the audio objects of the second audio object type to zero if it is found that there are more or less than two audio objects of the second audio object type.

The object separator is configured to use the common inter-object correlation value associated to the audio objects of the second audio object type to obtain the one or more audio channels representing the second audio information. Using this approach, the inter-object correlation value is exploited if it is obtainable with high computational efficiency, i.e. if there are two audio objects of the second audio object type. Otherwise, it would be computationally demanding to obtain inter-object correlation values. Accordingly, it has been found to be a good compromise in terms of hearing impression and computational complexity to set the inter-object correlation value associated to the audio objects of the second audio object type to zero if there are more or less than two audio objects of the second object type.

In a preferred embodiment, the audio signal processor is configured to render the second audio information in dependence on (at least a part of) the object-related parametric information, to obtain a rendered representation of the audio objects of the second audio object type as a processed version of the second audio information. In this case, the rendering can be made independent from the audio objects of the first audio object type.

In a preferred embodiment, the object separator is configured to provide the second audio information such that the second audio information describes more than two audio objects of the second audio object type. Embodiments according to the invention allow for a flexible adjustment of the number of audio objects of the second audio object type, which is significantly facilitated by the cascaded structure of the processing.

In a preferred embodiment, the object separator is configured to obtain, as the second audio information, a one-channel audio signal representation or a two-channel audio signal representation representing more than two audio objects of the second audio object type. Extracting one or two audio signal channels can be performed by the object separator with low computational complexity. In particular, the complexity of the object separator can be kept significantly smaller when compared to a case in which the object separator would need to deal with more than two audio objects of the second audio object type. Nevertheless, it has been found that it is a computationally efficient representation of the audio objects of the second audio object type to use one or two channels of an audio signal.

In a preferred embodiment, the audio signal processor is configured to receive the second audio information and to process the second audio information in dependence on (at least a part of) the object-related parametric information, taking into consideration object-related parametric information associated with more than two audio objects of the second audio object type. Accordingly, an object-individual processing is performed by the audio

processor, while such an object-individual processing is not performed for audio objects of the second audio object type by the object separator.

In a preferred embodiment, the audio decoder is configured to extract a total object number information and a foreground object number information from a configuration information related to the object-related parametric information. The audio decoder is also configured to determine a number of audio objects of the second audio object type by forming a difference between the total object number information and the foreground object number information. Accordingly, efficient signalling of the number of audio objects of the second audio object type is achieved. In addition, this concept provides for a high degree of flexibility regarding the number of audio objects of the second audio object type.

In a preferred embodiment, the object separator is configured to use object-related parametric information associated with N_{eao} audio objects of the first audio object type to obtain, as the first audio information, N_{eao}, audio signals representing (preferably, individually) the N_{eao} audio objects of the first audio object type, and to obtain, as the second audio information, one or two audio signals representing the $N\text{-}N_{eao}$ audio objects of the second audio object type, treating the $N\text{-}N_{eao}$ audio objects of the second audio object type as a single one-channel or two-channel audio object. The audio signal processor is configured to individually render the $N\text{-}N_{eao}$ audio objects represented by the one or two audio signals of the second audio information using the object-related parametric information associated with the $N\text{-}N_{eao}$ audio objects of the second audio object type. Accordingly, the audio object separation between the audio objects of the first audio object type and the audio objects of the second audio object type is separated from the subsequent processing of the audio objects of the second audio object type.

An embodiment according to the invention creates a method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information.

Another embodiment according to the invention creates a computer program for performing said method.

Brief Description of the Figs.

Embodiments according to the invention will subsequently be described taking reference to the enclosed Figs., in which:

Fig. 1 shows a block schematic diagram of an audio signal decoder, according to an embodiment of the invention;

Fig. 2 shows a block schematic diagram of another audio signal decoder, according to an embodiment of the invention;

Figs. 3a and 3b show a block schematic diagrams of a residual processor, which can be used as an object separator in an embodiment of the invention;

Figs. 4a to 4e show block schematic diagrams of audio signal processors, which can be used in an audio signal decoder according to an embodiment of the invention:

Fig. 4f shows a block diagram of an SAOC transcoder processing mode;

Fig. 4g shows a block diagram of an SAOC decoder processing mode;

Fig. 5a shows a block schematic diagram of an audio signal decoder, according to an embodiment of the invention;

Fig. 5b shows a block schematic diagram of another audio signal decoder, according to an embodiment of the invention;

Fig. 6a shows a Table representing a listening test design description;

Fig. 6b shows a Table representing systems under test;

Fig. 6c shows a Table representing the listening test items and rendering matrices;

Fig. 6d shows a graphical representation of average MUSHRA scores for a Karaoke/Solo type rendering listening test;

Fig. 6e shows a graphical representation of average MUSHRA scores for a classic rendering listening test;

Fig. 7 shows a flow chart of a method for providing an upmix signal representation, according to an embodiment of the invention;

Fig. 8 shows a block schematic diagram of a reference MPEG SAOC system;

Fig. 9a shows a block schematic diagram of a reference SAOC system using a separate decoder and mixer;

Fig. 9b shows a block schematic diagram of a reference SAOC system using an integrated decoder and mixer; and

Fig. 9c shows a block schematic diagram of a reference SAOC system using an SAOC-to-MPEG transcoder.

Detailed Description of the Embodiments

1. Audio signal decoder according to Fig. 1

Fig. 1 shows a block schematic diagram of an audio signal decoder 100 according to an embodiment of the invention.

The audio signal decoder 100 is configured to receive an object-related parametric information 110 and a downmix signal representation 112. The audio signal decoder 100 is configured to provide an upmix signal representation 120 in dependence on the downmix signal representation and the object-related parametric information 110. The audio signal decoder 100 comprises an object separator 130, which is configured to decompose the downmix signal representation 112 to provide a first audio information 132 describing a first set of one or more audio objects of a first audio object type and a second audio information 134 describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation 112 and using at least a part of the object-related parametric information 110. The audio signal decoder 100 also comprises an audio signal processor 140, which is configured to receive the second audio information 134 and to process the second audio information in dependence on at least a part of the object-related parametric information 112, to obtain a processed version 142 of the second audio information 134. The audio signal decoder 100 also comprises an audio signal combiner 150 configured to combine the first audio information 132 with the processed version 142 of the second audio information 134, to obtain the upmix signal representation 120.

The audio signal decoder 100 implements a cascaded processing of the downmix signal representation, which represents audio objects of the first audio object type and audio objects of the second audio object type in a combined manner.

In a first processing step, which is performed by the object separator 130, the second audio information describing a second set of audio objects of the second audio object type is separated from the first audio information 132 describing a first set of audio objects of a first audio object type using the object-related parametric information 110. However, the second audio information 134 is typically an audio information (for example, a one-channel audio signal or a two-channel audio signal) describing the audio objects of the second audio object type in a combined manner.

In the second processing step, the audio signal processor 140 processes the second audio information 134 in dependence on the object-related parametric information. Accordingly, the audio signal processor 140 is capable of performing an object-individual processing or rendering of the audio objects of the second audio object type, which are described by the second audio information 134, and which is typically not performed by the object separator 130.

Thus, while the audio objects of the second audio object type are preferably not processed in an object-individual manner by the object separator 130, the audio objects of the second audio object type are, indeed, processed in an object-individual manner (for example, rendered in an object-individual manner) in the second processing step, which is performed by the audio signal processor 140. Thus, the separation between the audio objects of the first audio object type and the audio objects of the second audio object type, which is performed by the object separator 130, is separated from the object-individual processing of the audio objects of the second audio object type, which is performed afterwards by the audio signal processor 140. Accordingly, the processing which is performed by the object separator 130 is substantially independent from a number of audio objects of the second audio object type. In addition, the format (for example, one-channel audio signal or the two-channel audio signal) of the second audio information 134 is typically independent from the number of audio objects of the second audio object type. Thus, the number of audio objects of the second audio object type can be varied without having the need to modify the structure of the object separator 130. In other words, the audio objects of the second audio object type are treated as a single (for example, one-channel or two-channel) audio object for which a common object-related parametric information (for example, a common object-level-difference value associated with one or two audio channels) is obtained by the object separator 140.

Accordingly, the audio signal decoder 100 according to Fig. 1 is capable to handle a variable number of audio objects of the second audio object type without a structural modification of the object separator 130. In addition, different audio object processing algorithms can be applied by the object separator 130 and the audio signal processor 140. Accordingly, for example, it is possible to perform an audio object separation using a residual information by the object separator 130, which allows for a particularly good separation of different audio objects, making use of the residual information, which constitutes a side information for improving the quality of an object separation. In contrast, the audio signal processor 140 may perform an object-individual processing without using a residual information. For example, the audio signal processor 140 may be configured to perform a conventional spatial-audio-object-coding (SAOC) type audio signal processing to render the different audio objects.

2. Audio Signal Decoder according to Fig. 2

In the following, an audio signal decoder 200 according to an embodiment of the invention will be described. A block-schematic diagram of this audio signal decoder 200 shown in Fig. 2.

The audio decoder 200 is configured to receive a downmix signal 210, a so-called SAOC bitstream 212, rendering matrix information 214 and, optionally, head-related-transfer-function (HRTF) parameters 216. The audio signal decoder 200 is also configured to provide an output/MPS downmix signal 220 and (optionally) a MPS bitstream 222.

2.1. Input signals and output signals of the audio signal decoder 200

In the following, various details regarding input signals and output signals of the audio decoder 200 will be described.

The downmix signal 200 may, for example, be a one-channel audio signal or a two-channel audio signal. The downmix signal 210 may, for example, be derived from an encoded representation of the downmix signal.

The spatial-audio-object-coding bitstream (SAOC bitstream) 212 may, for example, comprise object-related parametric information. For example, the SAOC bitstream 212 may comprise object-level-difference information, for example, in the form of object-level-

difference parameters OLD, an inter-object-correlation information, for example, in the form of inter-object-correlation parameters IOC.

In addition, the SAOC bitstream 212 may comprise a downmix information describing how the downmix signals have been provided on the basis of a plurality of audio object signals using a downmix process. For example, the SAOC bitstream may comprise a downmix gain parameter DMG and (optionally) downmix-channel-level difference parameters DCLD.

The rendering matrix information 214 may, for example, describe how the different audio objects should be rendered by the audio decoder. For example, the rendering matrix information 214 may describe an allocation of an audio object to one or more channels of the output/MPS downmix signal 220.

The optional head-related-transfer-function (HRTF) parameter information 216 may further describe a transfer function for deriving a binaural headphone signal.

The output/MPEG-Surround downmix signal (also briefly designated with "output/MPS downmix signal") 220 represents one or more audio channels, for example, in the form of a time domain audio signal representation or a frequency-domain audio signal representation. Alone or in combination with the optional MPEG-Surround bitstream (MPS bitstream) 222, which comprises MPEG-Surround parameters describing a mapping of the output/MPS downmix signal 220 onto a plurality of audio channels, an upmix signal representation is formed.

2.2. Structure and functionality of the audio signal decoder 200

In the following, the structure of the audio signal decoder 200, which may fulfill the functionality of an SAOC transcoder or the functionality of a SAOC decoder, will be described in more detail.

The audio signal decoder 200 comprises a downmix processor 230, which is configured to receive the downmix signal 210 and to provide, on the basis thereof, the output/MPS downmix signal 220. The downmix processor 230 is also configured to receive at least a part of the SAOC bitstream information 212 and at least a part of the rendering matrix information 214. In addition, the downmix processor 230 may also receive a processed SAOC parameter information 240 from a parameter processor 250.

The parameter processor 250 is configured to receive the SAOC bitstream information 212, the rendering matrix information 214 and, optionally, the head-related-transfer-function parameter information 260, and to provide, on the basis thereof, the MPEG Surround bitstream 222 carrying the MPEG surround parameters (if the MPEG surround parameters are required, which is, for example, true in the transcoding mode of operation). In addition, the parameter processor 250 provides the processed SAOC information 240 (if this processed SAOC information is required).

In the following, the structure and functionality of the downmix processor 230 will be described in more detail.

The downmix processor 230 comprises a residual processor 260, which is configured to receive the downmix signal 210 and to provide, on the basis thereof, a first audio object signal 262 describing so-called enhanced audio objects (EAOs), which may be considered as audio objects of a first audio object type. The first audio object signal may comprise one or more audio channels and may be considered as a first audio information. The residual processor 260 is also configured to provide a second audio object signal 264, which describes audio objects of a second audio object type and may be considered as a second audio information. The second audio object signal 264 may comprise one or more channels and may typically comprise one or two audio channels describing a plurality of audio objects. Typically, the second audio object signal may describe even more than two audio objects of the second audio object type.

The downmix processor 230 also comprises an SAOC downmix pre-processor 270, which is configured to receive the second audio object signal 264 and to provide, on the basis thereof, a processed version 272 of the second audio object signal 264, which may be considered as a processed version of the second audio information.

The downmix processor 230 also comprises an audio signal combiner 280, which is configured to receive the first audio object signal 262 and the processed version 272 of the second audio object signal 264, and to provide, on the basis thereof, the output/MPS downmix signal 220, which may be considered, alone or together with the (optional) corresponding MPEG-Surround bitstream 222, as an upmix signal representation.

In the following, the functionality of the individual units of the downmix processor 230 will be discussed in more detail.

The residual processor 260 is configured to separately provide the first audio object signal 262 and the second audio object signal 264. For this purpose, the residual processor 260 may be configured to apply at least a part of the SAOC bitstream information 212. For example, the residual processor 260 may be configured to evaluate an object-related parametric information associated with the audio objects of the first audio object type, i.e. the so-called "enhanced audio objects" EAO. In addition, the residual processor 260 may be configured to obtain an overall information describing the audio objects of the second audio object type, for example, the so-called "non-enhanced audio objects", commonly. The residual processor 260 may also be configured to evaluate a residual information, which is provided in the SAOC bitstream information 212, for a separation between enhanced audio objects (audio objects of the first audio object type) and non-enhanced audio objects (audio objects of the second audio object type). The residual information may, for example, encode a time domain residual signal, which is applied to obtain a particularly clean separation between the enhanced audio objects and the non-enhanced audio objects. In addition, the residual processor 260 may, optionally, evaluate at least a part of the rendering matrix information 214, for example, in order to determine a distribution of the enhanced audio objects to the audio channels of the first audio object signal 262.

The SAOC downmix pre-processor 270 comprises a channel re-distributor 274, which is configured to receive the one or more audio channels of the second audio object signal 264 and to provide, on the basis thereof, one or more (typically two) audio channels of the processed second audio object signal 272. In addition, the SAOC downmix pre-processor 270 comprises a decorrelated-signal-provider 276, which is configured to receive the one or more audio channels of the second audio object signal 264 and to provide, on the basis thereof, one or more decorrelated signals 278a, 278b, which are added to the signals provided by the channel re-distributor 274 in order to obtain the processed version 272 of the second audio object signal 264.

Further details regarding the SAOC downmix processor will be discussed below.

The audio signal combiner 280 combines the first audio object signal 262 with the processed version 272 of the second audio object signal. For this purpose, a channel-wise combination may be performed. Accordingly, the output/MPS downmix signal 220 is obtained.

The parameter processor 250 is configured to obtain the (optional) MPEG-Surround parameters, which make up the MPEG-Surround bitstream 222 of the upmix signal representation, on the basis of the SAOC bitstream, taking onto consideration the rendering matrix information 214 and, optionally, the HRTF parameter information 216. In other words, the SAOC parameter processor 252 is configured to translate the object-related parameter information, which is described by the SAOC bitstream information 212, into a channel-related parametric information, which is described by the MPEG Surround bit stream 222.

In the following, a short overview of the structure of the SAOC transcoder/decoder architecture shown in Fig. 2 will be given. Spatial audio object coding (SAOC) is a parametric multiple object coding technique. It is designed to transmit a number of audio objects in an audio signal (for example the downmix audio signal 210) that comprises M channels. Together with this backward compatible downmix signal, object parameters are transmitted (for example, using the SAOC bitstream information 212) that allow for recreation and manipulation of the original object signals. An SAOC encoder (not shown here) produces a downmix of the object signals at its input and extracts these object parameters. The number of objects that can be handled is in principle not limited. The object parameters are quantized and coded efficiently into the SAOC bitstream 212. The downmix signal 210 can be compressed and transmitted without the need to update existing coders and infrastructures. The object parameters, or SAOC side information, are transmitted in a low bit rate side channel, for example, the ancillary data portion of the downmix bitstream.

On the decoder side, the input objects are reconstructed and rendered to a certain number of playback channels. The rendering information containing reproduction level and panning position for each object is user-supplied or can be extracted from the SAOC bitstream (for example, as a preset information). The rendering information can be time-variant. Output scenarios can range from mono to multi-channel (for example, 5.1) and are independent from both, the number of input objects and the number of downmix channels. Binaural rendering of objects is possible including azimuth and elevation of virtual object positions. An optional effect interface allows for advanced manipulation of object signals, besides level and panning modification.

The objects themselves can be mono signals, stereophonic signals, as well as a multi-channel signals (for example 5.1 channels). Typical downmix configurations are mono and stereo.

In the following, the basic structure of the SAOC transcoder/decoder, which is shown in Fig. 2, will be explained. The SAOC transcoder/decoder module described herein may act either as a stand-alone decoder or as a transcoder from an SAOC to an MPEG-surround bitstream, depending on the intended output channel configuration. In a first mode of operation, the output signal configuration is mono, stereo or binaural, and two output channels are used. In this first case, the SAOC module may operate in a decoder mode, and the SAOC module output is a pulse-code-modulated output (PCM output). In the first case, an MPEG surround decoder is not required. Rather, the upmix signal representation may only comprise the output signal 220, while the provision of the MPEG surround bit stream 222 may be omitted. In a second case, the output signal configuration is a multi-channel configuration with more than two output channels. The SAOC module may be operational in a transcoder mode. The SAOC module output may comprise both a downmix signal 220 and an MPEG surround bit stream 222 in this case, as shown in Fig. 2. Accordingly, an MPEG surround decoder is required in order to obtain a final audio signal representation for output by the speakers.

Fig. 2 shows the basic structure of the SAOC transcoder/decoder architecture. The residual processor 216 extracts the enhanced audio object from the incoming downmix signal 210 using the residual information contained in the SAOC bit stream 212. The downmix preprocessor 270 processes the regular audio objects (which are, for example, non-enhanced audio objects, i.e., audio objects for which no residual information is transmitted in the SAOC bit stream 212). The enhanced audio objects (represented by the first audio object signal 262) and the processed regular audio objects (represented, for example, by the processed version 272 of the second audio object signal 264) are combined to the output signal 220 for the SAOC decoder mode or to the MPEG surround downmix signal 220 for the SAOC transcoder mode. Detailed descriptions of the processing blocks are given below.

3. Architecture and functionality of Residual Processor and Energy Mode Processor

In the following, details regarding a residual processor will be described, which may, for example, take over the functionality of the object separator 130 of the audio signal decoder 100 or of the residual processor 260 of the audio signal decoder 200. For this purpose, Figs. 3a and 3b show block schematic diagrams of such a residual processor 300, which may take the place of the object separator 130 or of the residual processor 260. Fig. 3a shows less details than Fig. 3b. However, the following description applies to the residual processor 300 according to Fig. 3a and also to the residual processor 380 according to Fig. 3b.

The residual processor 300 is configured to receive an SAOC downmix signal 310, which may be equivalent to the downmix signal representation 112 of Fig. 1 or the downmix signal representation 210 of Fig. 2. The residual processor 300 is configured to provide, on the basis thereof, a first audio information 320 describing one or more enhanced audio objects, which may, for example, be equivalent to the first audio information 132 or to the first audio object signal 262. Also, the residual processor 300 may provide a second audio information 322 describing one or more other audio objects (for example, non-enhanced audio objects, for which no residual information is available), wherein the second audio information 322 may be equivalent to the second audio information 134 or to the second audio object signal 264.

The residual processor 300 comprises a 1-to-N/2-to-N unit (OTN/TTN unit) 330, which receives the SAOC downmix signal 310 and which also receives SAOC data and residuals 332. The 1-to-N/2-to-N unit 330 also provides an enhanced-audio-object signal 334, which describes the enhanced audio objects (EAO) contained in the SAOC downmix signal 310. Also, the 1-to-N/2-to-N unit 330 provides the second audio information 322. The residual processor 300 also comprises a rendering unit 340, which receives the enhanced-audio-object signal 334 and a rendering matrix information 342 and provides, on the basis thereof, the first audio information 320.

In the following, the enhanced audio object processing (EAO processing), which is performed by the residual processor 300, will be described in more detail.

3.1. Introduction into the Operation of the Residual Processor 300

Regarding the functionality of the residual processor 300, it should be noted that the SAOC technology allows for the individual manipulation of a number of audio objects in terms of their level amplification/attenuation without significant decrease in the resulting sound quality only in a very limited way. A special "karaoke-type" application scenario requires a total (or almost total) suppression of the specific objects, typically the lead vocal, keeping the perceptional quality of the background sound scene unharmed.

A typical application case contains up to four enhanced audio objects (EAO) signals, which can, for example, represent two independent stereo objects (for example, two independent stereo objects which are prepared to be removed at the side of the decoder).

It should be noted that the (one or more) quality enhanced audio objects (or, more precisely, the audio signal contributions associated with the enhanced audio objects) are included in the SAOC downmix signal 310. Typically, the audio signal contributions associated with the (one or more) enhanced audio objects are mixed, by the downmix processing performed by the audio signal encoder, with audio signal contributions of other audio objects, which are not enhanced audio objects. Also, it should be noted that audio signal contributions of a plurality of enhanced audio objects are also typically overlapped or mixed by the downmix processing performed by the audio signal encoder.

3.2 SOAC Architecture Supporting Enhanced Audio Objects

In the following, details regarding the residual processor 300 will be described. Enhanced audio object processing incorporates the 1-to-N or 2-to-N units, depending on the SAOC downmix mode. The 1-to-N processing unit is dedicated to a mono downmix signal and the 2-to-N processing unit is dedicated to a stereo downmix signal 310. Both these units represent a generalized and enhanced modification of the 2-to-2 box (TTT box) known from ISO/IEC 23003-1:2007. In the encoder, regular and EAO signals are combined into the downmix. The OTN^{-1}/TTN^{-1} processing units (which are inverse one-to-N processing units or inverse 2-to-N processing units) are employed to produce and encode the corresponding residual signals.

The EAO and regular signals are recovered from the downmix 310 by the OTN/TTN units 330 using the SAOC side information and incorporated residual signals. The recovered EAOs (which are described by the enhanced audio object signal 334) are fed into the rendering unit 340 which represents (or provides) the product of the corresponding rendering matrix (described by the rendering matrix information 342) and the resulting output of the OTN/TTN unit. The regular audio objects (which are described by the second audio information 322) are delivered to the SAOC downmix pre-processor, for example, the SAOC downmix preprocessor 270, for further processing. Figs. 3a and 3b depict the general structure of the residual processor, i.e., the architecture of the residual processor.

The residual processor output signals 320,322 are computed as

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}\mathbf{X}_{res},$$

$$\mathbf{X}_{EAO} = \mathbf{A}_{EAO}\mathbf{M}_{EAO}\mathbf{X}_{res},$$

where $\mathbf{X}_{OBJ}$ represents the downmix signal of the regular audio objects (i.e. non-EAOs) and $\mathbf{X}_{EAO}$ is the rendered EAO output signal for the SAOC decoding mode or the corresponding EAO downmix signal for the SAOC transcoding mode.

The residual processor can operate in prediction (using residual information) mode or energy (without residual information) mode. The extended input signal $\mathbf{X}_{res}$ is defined accordingly:

$$\mathbf{X}_{res} = \begin{cases} \begin{pmatrix} \mathbf{X} \\ \hline \mathbf{res} \end{pmatrix}, & \text{for pediction mode,} \\ \mathbf{X}, & \text{for energy mode.} \end{cases}$$

Here, $\mathbf{X}$ may, for example, represent the one or more channels of the downmix signal representation 310, which may be transported in the bitstream representing the multi-channel audio content. **res** may designate one or more residual signals, which may be described by the bitstream representing the multi-channel audio content.

The OTN/TTN processing is represented by matrix $\mathbf{M}$ and EAO processor by matrix $\mathbf{A}_{EAO}$.

The OTN/TTN processing matrix $\mathbf{M}$ is defined according to the EAO operation mode (i.e. prediction or energy) as

$$\mathbf{M} = \begin{cases} \mathbf{M}_{\text{Prediction}}, & \text{for pediction mode,} \\ \mathbf{M}_{\text{Energy}}, & \text{for energy mode.} \end{cases}$$

The OTN/TTN processing matrix $\mathbf{M}$ is represented as

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{OBJ} \\ \hline \mathbf{M}_{EAO} \end{pmatrix},$$

where the matrix $\mathbf{M}_{OBJ}$ relates to the regular audio objects (i.e. non-EAOs) and $\mathbf{M}_{EAO}$ to the enhanced audio objects (EAOs).

In some embodiments, one or more multichannel background objects (MBO) may be treated the same way by the residual processor 300.

A Multi-channel Background Object (MBO) is an MPS mono or stereo downmix that is part of the SAOC downmix. As opposed to using individual SAOC objects for each channel in a multi-channel signal, an MBO can be used enabling SAOC to more efficiently handle a multi-channel object. In the MBO case, the SAOC overhead gets lower as the MBO's SAOC parameters only are related to the downmix channels rather than all the upmix channels.

3.3 Further Definitions

3.3.1 Dimensionality of Signals and Parameters

In the following, the dimensionality of the signals and parameters will be briefly discussed in order to provide an understanding how often the different calculations are performed.

The audio signals are defined for every time slot n and every hybrid subband (which may be a frequency subband) k. The corresponding SAOC parameters are defined for each parameter time slot l and processing band m. A Subsequent mapping between the hybrid and parameter domain is specified by table A.31 ISO/IEC 23003-1:2007. Hence, all calculations are performed with respect to the certain time/band indices and the corresponding dimensionalities are implied for each introduced variable.

However, in the following, the time and frequency band indices will be omitted sometimes to keep the notation concise.

3.3.2 Calculation of the matrix $\mathbf{A}_{EAO}$

The EAO pre-rendering matrix $\mathbf{A}_{EAO}$ is defined according to the number of output channels (i.e. mono, stereo or binaural) as

$$\mathbf{A}_{EAO} = \begin{cases} \mathbf{A}_1^{EAO}, & \text{for mono case,} \\ \mathbf{A}_2^{EAO}, & \text{for other cases.} \end{cases}$$

The matrices $\mathbf{A}_1^{EAO}$ of size $1 \times N_{EAO}$ and $\mathbf{A}_2^{EAO}$ of size $2 \times N_{EAO}$ are defined as

$$\mathbf{A}_1^{EAO} = \mathbf{D}_{16}^{EAO}\mathbf{M}_{ren}^{EAO}, \qquad \mathbf{D}_{16}^{EAO} = \begin{pmatrix} w_1^{EAO} & w_2^{EAO} & w_3^{EAO} & w_3^{EAO} & w_1^{EAO} & w_2^{EAO} \end{pmatrix},$$

$$\mathbf{A}_2^{EAO} = \mathbf{D}_{26}^{EAO}\mathbf{M}_{ren}^{EAO}, \qquad \mathbf{D}_{26}^{EAO} = \begin{pmatrix} w_1^{EAO} & 0 & \frac{w_3^{EAO}}{\sqrt{2}} & \frac{w_3^{EAO}}{\sqrt{2}} & w_1^{EAO} & 0 \\ 0 & w_2^{EAO} & \frac{w_3^{EAO}}{\sqrt{2}} & \frac{w_3^{EAO}}{\sqrt{2}} & 0 & w_2^{EAO} \end{pmatrix},$$

where the rendering sub-matrix M_{ren}^{EAO} corresponds to the EAO rendering (and describes a desired mapping of enhanced audio objects onto channels of the upmix signal representation).

The values w_i^{EAO} are computed in dependence on rendering information associated with the enhanced audio objects using the corresponding EAO elements and using the equations of section 4.2.2.1.

In case of binaural rendering the matrix A_2^{EAO} is defined by equations given in section 4.1.2, for which the corresponding target binaural rendering matrix contains only EAO related elements.

3.4 Calculation of the OTN/TTN Elements in the Residual Mode

In the following, it will be discussed how the SAOC downmix signal 310, which typically comprises one or two audio channels, is mapped onto the enhanced audio object signal 334, which typically comprises one or more enhanced audio object channels, and the second audio information 322, which typically comprises one or two regular audio object channels.

The functionality of the 1-to-N unit or 2-to-N unit 330 may, for example, be implemented using a matrix vector multiplication, such that a vector describing both the channels of the enhanced audio object signal 334 and the channels of the second audio information 322 is obtained by multiplying a vector describing the channels of the SAOC downmix signal 310 and (optionally) one or more residual signals with a matrix $\mathbf{M}_{Prediction}$ or $\mathbf{M}_{Energy}$. Accordingly, the determination of the matrix $\mathbf{M}_{Prediction}$ or $\mathbf{M}_{Energy}$ is an important step in the derivation of the first audio information 320 and the second audio information 322 from the SAOC downmix 310.

To summarize, the OTN/TTN upmix process is presented by either a matrix $\mathbf{M}_{Prediction}$ for a prediction mode or $\mathbf{M}_{Energy}$ for an energy mode.

The energy based encoding/decoding procedure is designed for non-waveform preserving coding of the downmix signal. Thus the OTN/TTN upmix matrix for the corresponding energy mode does not rely on specific waveforms, but only describe the relative energy distribution of the input audio objects, as will be discussed in more detail below.

3.4.1 Prediction mode

For the prediction mode the matrix $\mathbf{M}_{Prediction}$ is defined exploiting the downmix information contained in the matrix $\tilde{\mathbf{D}}^{-1}$ and the CPC data from matrix $\mathbf{C}$:

$$\mathbf{M}_{Prediction} = \tilde{\mathbf{D}}^{-1}\mathbf{C}.$$

With respect to the several SAOC modes, the extended downmix matrix $\tilde{\mathbf{D}}$ and CPC matrix $\mathbf{C}$ exhibit the following dimensions and structures:

3.4.1.1 Stereo downmix modes (TTN):

For stereo downmix modes (TTN) (for example, for the case of a stereo downmix on the basis of two regular-audio-object channels and N_{EAO} enhanced-audio-object-channels), the (extended) downmix matrix $\tilde{\mathbf{D}}$ and the CPC matrix $\mathbf{C}$ can be obtained as follows:

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right),$$

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right).$$

With a stereo downmix, each EAO j holds two CPCs $c_{j,0}$ and $c_{j,1}$ yielding matrix $\mathbf{C}$.

The residual processor output signals are computed as

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \left(\begin{array}{c} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{array}\right),$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \left(\begin{array}{c} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{array}\right).$$

Accordingly, two signals y_L, y_R (which are represented by $\mathbf{X}_{OBJ}$) are obtained, which represent one or two or even more than two regular audio objects (also designated as non-extended audio objects). Also, N_{EAO} signals (represented by $\mathbf{X}_{EAO}$) representing N_{EAO} enhanced audio objects are obtained. These signals are obtained on the basis of two SAOC downmix signals l_0,r_0 and N_{EAO} residual signals res_0 to res_{NEAO-1}, which will be encoded in the SAOC side information, for example, as a part as the object-related parametric information.

It should be noted that the signals y_L and y_R may be equivalent to the signal 322, and that the signals $y_{0,EAO}$ to $y_{NEAO-1,\ EAO}$ (which are represented by $\mathbf{X}_{EAO}$) may equivalent to the signals 320.

The matrix $\mathbf{A}^{EAO}$ is a rendering matrix. Entries of the matrix $\mathbf{A}^{EAO}$ may describe, for example, a mapping of enhanced audio objects to the channels of the enhanced audio object signal 334 ($\mathbf{X}_{EAO}$).

Accordingly, an appropriate choice of the matrix $\mathbf{A}^{EAO}$ may allow for an optional integration of the functionality of the rendering unit 340, such that the multiplication of the vector describing the channels (l_0,r_0) of the SAOC downmix signal 310 and one or more residual signals $(res_0,\ldots,res_{NEAO-1})$ with the matrix $\mathbf{A}^{EAO}\mathbf{M}_{EAO}^{Prediction}$ may directly result in a representation $\mathbf{X}_{EAO}$ of the first audio information 320.

3.4.1.2 Mono downmix modes (OTN):

In the following, the derivation of the enhanced audio object signals 320 (or, alternatively, of the enhanced audio object signals 334) and of the regular audio object signal 322 will be described for the case in which the SAOC downmix signal 310 comprises a signal channel only.

For mono downmix modes (OTN) (e.g., a mono downmix on the basis of one regular-audio-object channel and N_{EAO} enhanced-audio-object channels), the (extended) downmix matrix $\tilde{\mathbf{D}}$ and the CPC matrix $\mathbf{C}$ can be obtained as follows:

$$\tilde{\mathbf{D}} = \left(\begin{array}{c|ccc} 1 & m_0 & \cdots & m_{N_{EAO}-1} \\ \hline m_0 & -1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right),$$

$$\mathbf{C} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1,0} & 0 & \cdots & 1 \end{array}\right).$$

With a mono downmix, one EAO j is predicted by only one coefficient c_j yielding the matrix $\mathbf{C}$. All matrix elements c_j are obtained, for example, from the SAOC parameters (for example, from the SAOC data 322) according to the relationships provided below (section 3.4.1.4).

The residual processor output signals are computed as

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} d_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix},$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} d_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}.$$

The output signal $\mathbf{X}_{OBJ}$ comprises, for example, one channel describing the regular audio objects (non-enhanced audio objects) . The output signal $\mathbf{X}_{EAO}$ comprises, for example, one, two, or even more channels describing the enhanced audio objects (preferably N_{EAO} channels describing the enhanced audio objects). Again, said signals are equivalent to the signals 320, 322.

3.4.1.3 Calculation of the inverse extended downmix matrix

The matrix $\tilde{\mathbf{D}}^{-1}$ is the inverse of the extended downmix matrix $\tilde{\mathbf{D}}$ and $\mathbf{C}$ implies the CPCs.

The matrix $\tilde{\mathbf{D}}^{-1}$ is the inverse of the extended downmix matrix $\tilde{\mathbf{D}}$ and can be calculated as

$$\tilde{\mathbf{D}}^{-1} = \frac{\tilde{d}_{i,j}}{den}.$$

The elements $\tilde{d}_{i,j}$ (for example, of the inverse $\tilde{\mathbf{D}}^{-1}$ of the extended downmix matrix $\tilde{\mathbf{D}}$ of size 6×6) are derived using the following values:

$$\tilde{d}_{1,1} = 1 + \sum_{j=1}^{4} n_j^2 ,$$

$$\tilde{d}_{1,2} = -\left(\sum_{j=1}^{4} m_j n_j \right),$$

$$\tilde{d}_{1,3} = m_1 + m_1 n_2^2 + m_1 n_3^2 + m_1 n_4^2 - m_2 n_1 n_2 - m_3 n_1 n_3 - m_4 n_1 n_4 ,$$

$$\tilde{d}_{1,4} = m_2 + m_2 n_1^2 + m_2 n_3^2 + m_2 n_4^2 - m_1 n_2 n_1 - m_3 n_2 n_3 - m_4 n_2 n_4 ,$$

$$\tilde{d}_{1,5} = m_3 + m_3 n_1^2 + m_3 n_2^2 + m_3 n_4^2 - m_1 n_3 n_1 - m_2 n_3 n_2 - m_4 n_3 n_4 ,$$

$$\tilde{d}_{1,6} = m_4 + m_4 n_1^2 + m_4 n_2^2 + m_4 n_3^2 - m_1 n_4 n_1 - m_2 n_4 n_2 - m_3 n_4 n_3 ,$$

$$\tilde{d}_{2,2} = 1 + \sum_{j=1}^{4} m_j^2 ,$$

$$\tilde{d}_{2,3} = n_1 + n_1 m_2^2 + n_1 m_3^2 + n_1 m_4^2 - m_1 m_2 n_2 - m_1 m_3 n_3 - m_1 m_4 n_4 ,$$

$$\tilde{d}_{2,4} = n_2 + n_2 m_1^2 + n_2 m_3^2 + n_2 m_4^2 - m_2 m_1 n_1 - m_2 m_3 n_3 - m_2 m_4 n_4 ,$$

$$\tilde{d}_{2,5} = n_3 + n_3 m_1^2 + n_3 m_2^2 + n_3 m_4^2 - m_3 m_1 n_1 - m_3 m_2 n_2 - m_3 m_4 n_4 ,$$

$$\tilde{d}_{2,6} = n_4 + n_4 m_1^2 + n_4 m_2^2 + n_4 m_3^2 - m_4 m_1 n_1 - m_4 m_2 n_2 - m_4 m_3 n_3 ,$$

$$\tilde{d}_{3,3} = -1 - \sum_{j=2}^{4} m_j^2 - \sum_{j=2}^{4} n_j^2 - m_3^2 n_2^2 - m_4^2 n_2^2 - m_2^2 n_3^2 - m_4^2 n_3^2 - m_2^2 n_4^2 - m_3^2 n_4^2 + 2m_2 m_3 n_2 n_3 + 2m_2 m_4 n_2 n_4 + 2m_3 m_4 n_3 n_4$$

$$, \tilde{d}_{3,4} = m_1 m_2 + n_1 n_2 + m_3^2 n_1 n_2 + m_4^2 n_1 n_2 + m_1 m_2 n_3^2 + m_1 m_2 n_4^2 - m_2 m_3 n_1 n_3 - m_1 m_3 n_2 n_3 - m_2 m_4 n_1 n_4 - m_1 m_4 n_2 n_4 ,$$

$$\tilde{d}_{3,5} = m_1 m_3 + n_1 n_3 + m_2^2 n_1 n_3 + m_4^2 n_1 n_3 + m_1 m_3 n_2^2 + m_1 m_3 n_4^2 - m_2 m_3 n_1 n_2 - m_1 m_2 n_2 n_3 - m_3 m_4 n_1 n_4 - m_1 m_4 n_3 n_4 ,$$

$$\tilde{d}_{3,6} = m_1 m_4 + n_1 n_4 + m_2^2 n_1 n_4 + m_3^2 n_1 n_4 + m_1 m_4 n_2^2 + m_1 m_4 n_3^2 - m_2 m_4 n_1 n_2 - m_3 m_4 n_1 n_3 - m_1 m_2 n_2 n_4 - m_1 m_3 n_4 n_3 ,$$

$$\tilde{d}_{4,4} = -1 - \sum_{\substack{j=1 \\ j\neq 2}}^{4} m_j^2 - \sum_{\substack{j=1 \\ j\neq 2}}^{4} n_j^2 - m_3^2 n_1^2 - m_4^2 n_1^2 - m_1^2 n_3^2 - m_4^2 n_3^2 - m_1^2 n_4^2 - m_3^2 n_4^2 + 2m_1 m_3 n_1 n_3 + 2m_1 m_4 n_1 n_4 + 2m_3 m_4 n_3 n_4 ,$$

$$\tilde{d}_{4,5} = m_2 m_3 + n_2 n_3 + m_1^2 n_2 n_3 + m_4^2 n_2 n_3 + m_2 m_3 n_1^2 + m_2 m_3 n_4^2 - m_1 m_3 n_1 n_2 - m_1 m_2 n_1 n_3 - m_3 m_4 n_2 n_4 - m_2 m_4 n_3 n_4 ,$$

$$\tilde{d}_{4,6} = m_2 m_4 + n_2 n_4 + m_1^2 n_2 n_4 + m_3^2 n_2 n_4 + m_2 m_4 n_1^2 + m_2 m_4 n_3^2 - m_1 m_4 n_1 n_2 - m_3 m_4 n_2 n_3 - m_1 m_2 n_1 n_4 - m_2 m_3 n_3 n_4 ,$$

$$\tilde{d}_{5,5} = -1 - \sum_{\substack{j=1 \\ j\neq 3}}^{4} m_j^2 - \sum_{\substack{j=1 \\ j\neq 3}}^{4} n_j^2 - m_2^2 n_1^2 - m_4^2 n_1^2 - m_1^2 n_2^2 - m_4^2 n_2^2 - m_1^2 n_4^2 - m_2^2 n_4^2 + 2m_1 m_2 n_1 n_2 + 2m_1 m_4 n_1 n_4 + 2m_2 m_4 n_2 n_4 ,$$

$$\tilde{d}_{5,6} = m_3 m_4 + n_3 n_4 + m_1^2 n_3 n_4 + m_2^2 n_3 n_4 + m_3 m_4 n_1^2 + m_3 m_4 n_2^2 - m_1 m_4 n_1 n_3 - m_2 m_4 n_2 n_3 - m_1 m_3 n_1 n_4 - m_2 m_3 n_2 n_4 ,$$

$$\tilde{d}_{6,6} = -1 - \sum_{j=1}^{3} m_j^2 - \sum_{j=1}^{3} n_j^2 - m_2^2 n_1^2 - m_3^2 n_1^2 - m_1^2 n_2^2 - m_3^2 n_2^2 - m_1^2 n_3^2 - m_2^2 n_3^2 + 2m_1 m_2 n_1 n_2 + 2m_1 m_3 n_1 n_3 + 2m_2 m_3 n_2 n_3 ,$$

$$den = 1 + \sum_{j=1}^{4} m_j^2 + \sum_{j=1}^{4} n_j^2 + m_2^2 n_1^2 + m_3^2 n_1^2 + m_4^2 n_1^2 + m_1^2 n_2^2 + m_3^2 n_2^2 + m_4^2 n_2^2 + m_1^2 n_3^2 + m_2^2 n_3^2 + m_4^2 n_3^2 + m_1^2 n_4^2 + m_2^2 n_4^2 +$$

$$+ m_3^2 n_4^2 - 2m_1 m_2 n_1 n_2 - 2m_1 m_3 n_1 n_3 - 2m_2 m_3 n_2 n_3 - 2m_1 m_4 n_1 n_4 - 2m_2 m_4 n_2 n_4 - 2m_3 m_4 n_3 n_4 .$$

The coefficients m_j and n_j of the extended downmix matrix $\tilde{\mathbf{D}}$ denote the downmix values for every EAO j for the right and left downmix channel as

$$m_j = d_{0,EAO(j)},\ n_j = d_{1,EAO(j)}.$$

The elements $d_{i,j}$ of the downmix matrix **D** are obtained using the downmix gain information DMG and the (optional) downmix channel level different information DCLD, which is included in the SAOC information 332, which is represented, for example, by the object-related parametric information 110 or the SAOC bitstream information 212.

For the stereo downmix case the downmix matrix **D** of size $2 \times N$ with elements $d_{i,j}$ $(i = 0,1;\ j = 0,\ldots,N-1)$ is obtained from the DMG and DCLD parameters as

$$d_{0,j} = 10^{0.05 DMG_j} \sqrt{\frac{10^{0.1 DCLD_j}}{1+10^{0.1 DCLD_j}}}, \qquad d_{1,j} = 10^{0.05 DMG_j} \sqrt{\frac{1}{1+10^{0.1 DCLD_j}}}.$$

For the mono downmix case the downmix matrix **D** of size $1 \times N$ with elements $d_{i,j}$ $(i = 0;\ j = 0,\ldots,N-1)$ is obtained from the DMG parameters as

$$d_{0,j} = 10^{0.05 DMG_j}.$$

Here, the dequantized downmix parameters DMG_j and $DCLD_j$ are obtained, for example, from the parametric side information 110 or from the SAOC bitstream 212.

The function $EAO(j)$ determines mapping between indices of input audio object channels and EAO signals:

$$EAO(j) = N-1-j, \qquad j = 0,\ldots,N_{EAO}-1.$$

3.4.1.4 Calculation of the matrix C

The matrix **C** implies the CPCs and is derived from the transmitted SAOC parameters (i.e. the OLDs, IOCs, DMGs and DCLDs) as

$$c_{j,0} = (1-\lambda)\tilde{c}_{j,0} + \lambda\gamma_{j,0}, \qquad c_{j,1} = (1-\lambda)\tilde{c}_{j,1} + \lambda\gamma_{j,1}.$$

In other words, the constrained CPCs are obtained in accordance with the above equations, which may be considered as a constraining algorithm. However, the constrained CPCs may also be derived from the values $\tilde{c}_{j,0}, \tilde{c}_{j,1}$ using a different limitation approach (constraining algorithm), or can be set to be equal to the values $\tilde{c}_{j,0}, \tilde{c}_{j,1}$.

It should be noted, that matrix entries $c_{j,1}$ (and the intermediate quantities on the basis of which the matrix entries $c_{j,1}$ are computed) are typically only required if the downmix signal is a stereo downmix signal.

The CPCs are constrained by the subsequent limiting functions:

$$\gamma_{j,1} = \frac{m_j OLD_L + n_j e_{L,R} - \sum_{i=0}^{N_{EAO}-1} m_i e_{i,j}}{2\left(OLD_L + \sum_{i=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_i m_k e_{i,k}\right)}, \qquad \gamma_{j,2} = \frac{n_j OLD_R + m_j e_{L,R} - \sum_{i=0}^{N_{EAO}-1} n_i e_{i,j}}{2\left(OLD_R + \sum_{i=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} n_i n_k e_{i,k}\right)},$$

with the weighting factor λ determined as

$$\lambda = \left(\frac{P_{LoRo}^2}{P_{Lo}P_{Ro}}\right)^8.$$

For one specific EAO channel $j = 0 \ldots N_{EAO} - 1$ the unconstrained CPCs are estimated by

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}, \qquad \tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}.$$

The energy quantities P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ and $P_{RoCo,j}$ are computed as

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k},$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k},$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k} ,$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j} ,$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j} .$$

The covariance matrix $e_{i,j}$ is defined in the following way: The covariance matrix $\mathbf{E}$ of size $N \times N$ with elements $e_{i,j}$ represents an approximation of the original signal covariance matrix $\mathbf{E} \approx \mathbf{SS}^*$ and is obtained from the OLD and IOC parameters as

$$e_{i,j} = \sqrt{OLD_i OLD_j}\, IOC_{i,j} .$$

Here, the dequantized object parameters OLD_i, $IOC_{i,j}$ are obtained, for example, from the parametric side information 110 or from the SAOC bitstream 212.

In addition, $e_{L,R}$ may, for example, be obtained as

$$e_{L,R} = \sqrt{OLD_L OLD_R}\, IOC_{L,R} .$$

The parameters OLD_L, OLD_R and $IOC_{L,R}$ correspond to the regular (audio) objects and can be derived using the downmix information:

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

As can be seen, two common object-level-different values OLD_L and OLD_R are computed for the regular audio objects in the case of a stereo downmix signal (which preferably

implies a two-channel regular audio object signal). In contrast, only one common object-level-different value OLD_L is computed for the regular audio objects in the case of a one-channel (mono) downmix signal (which preferably implies a one-channel regular audio object signal).

As can be seen, the first (in the case of a two-channel downmix signal) or sole (in the case of a one-channel downmix signal) common object-level-difference value OLD_L is obtained by summing contributions of the regular audio objects having audio object index (or indices) i to the left channel (or sole channel) of the SAOC downmix signal 310.

The second common object-level-difference value OLD_R (which is used in the case of a two-channel downmix signal) is obtained by summing the contributions of the regular audio objects having the audio object index (or indices) i to the right channel of the SAOC downmix signal 310.

The contribution OLD_L of the regular audio objects (having audio objects indices i=0 to $i=N-N_{EAO}-1$) onto the left channel signal (or sole channel signal) of the SAOC downmix signal 710 is computed, for example, taking into consideration the downmix gain $d_{0,i}$, describing the downmix gain applied to the regular audio object having audio object index i when obtaining the left channel signal of the SAOC downmix signal 310, and also the object level of the regular audio object having the audio object i, which is represented by the value OLD_i.

Similarly, the common object level difference value OLD_R is obtained using the downmix coefficients $d_{1,i}$, describing the downmix gain which is applied to the regular audio object having the audio object index i when forming the right channel signal of the SAOC downmix signal 310, and the level information OLD_i associated with the regular audio object having the audio object index i.

As can be seen, the equations for the calculation of the quantities P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ and $P_{RoCo,j}$ do not distinguish between the individual regular audio objects, but merely make use of the common object level difference values OLD_L, OLD_R, thereby considering the regular audio objects (having audio object indices i) as a single audio object.

Also, the inter-object-correlation value $IOC_{L,R}$, which is associated with the regular audio objects, is set to 0 unless there are two regular audio objects.

The covariance matrix $e_{i,j}$ (and $e_{L,R}$) is defined as follows:

The covariance matrix $\mathbf{E}$ of size $N \times N$ with elements $e_{i,j}$ represents an approximation of the original signal covariance matrix $\mathbf{E} \approx \mathbf{SS}^*$ and is obtained from the OLD and IOC parameters as

$$e_{i,j} = \sqrt{OLD_i OLD_j}\, IOC_{i,j}.$$

For example,

$$e_{L,R} = \sqrt{OLD_L OLD_R}\, IOC_{L,R},$$

wherein OLD_L and OLD_R and $IOC_{L,R}$ are computed as described above.

Here, the dequantized object parameters are obtained as

$$OLD_i = \mathbf{D}_{\mathrm{OLD}}(i,l,m), \qquad IOC_{i,j} = \mathbf{D}_{\mathrm{IOC}}(i,j,l,m),$$

wherein $\mathbf{D}_{\mathrm{OLD}}$ and $\mathbf{D}_{\mathrm{IOC}}$ are matrices comprising objects-level-difference parameters and inter-object-correlation parameters.

3.4.2. Energy Mode

In the following, another concept will be described, which can be used to separate the extended-audio-object signals 320 and the regular-audio-object (non-extended audio object) signals 322, and which can be used in combination with a non-waveform-preserving audio coding of the SAOC downmix channels 310.

In other words, the energy based encoding/decoding procedure is designed for non-waveform preserving coding of the downmix signal. Thus the OTN/TTN upmix matrix for the corresponding energy mode does not rely on specific waveforms, but only describe the relative energy distribution of the input audio objects.

Also, the concept discussed here, which is designated as an "energy mode" concept, can be used without transmitting a residual signal information. Again, the regular audio objects (non-enhanced audio objects) are treated as a single one-channel or two-channel audio object having one or two common object-level-difference values OLD_L, OLD_R.

For the energy mode the matrix $\mathbf{M}_{Energy}$ is defined exploiting the downmix information and the OLDs, as will be described in the following.

3.4.2.1. Energy Mode for Stereo Downmix Modes (TTN)

In case of a stereo (for example, a stereo downmix on the basis of two regular-audio-object channels and N_{EAO} enhanced-audio-object channels), the matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ are obtained from the corresponding OLDs according to

$$\mathbf{M}_{OBJ}^{Energy} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{Energy} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

The residual processor output signals are computed as

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix},$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}.$$

The signals y_L, y_R, which are represented by the signal $\mathbf{X}_{OBJ}$, describe the regular audio objects (and may be equivalent to the signal 322) , and the signals $\mathbf{y}_{0,EAO}$ to $\mathbf{y}_{NEAO-1,EAO}$, which are described by the signal $\mathbf{X}_{EAO}$, describe the enhanced audio objects (and may be equivalent to the signal 334 or to the signal 320).

If a mono upmix signal is desired for the case of a stereo downmix signal, a 2-to-1 processing may be performed, for example, by the pre-processor 270 on the basis of the two-channel signal $\mathbf{X}_{OBJ}$.

3.4.2.2. Energy Mode for Mono Downmix Modes (OTN)

For the mono case (for example, a mono downmix on the basis of one regular-audio-object channel and N_{EAO} enhanced-audio-object channels), the matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ are obtained from the corresponding OLDs according to

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \left(\sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right),$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}.$$

The residual processor output signals are computed as

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}}\left(d_0\right),$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Energy}}\left(d_0\right).$$

A single regular-audio-object channel 322 (represented by $\mathbf{X}_{OBJ}$) and N_{EAO} enhanced-audio-object channels 320 (represented by $\mathbf{X}_{EAO}$) can be obtained by applying the matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ to a representation of a single channel SAOC downmix signal 310 (represented here by d_0).

If a two-channel (stereo) upmix signal is desired for the case of a one-channel (mono) downmix signal, a 1-to-2 processing may be performed, for example, by the pre-processor 270 on the basis of the one-channel signal $\mathbf{X}_{OBJ}$.

4. Architecture and operation of the SAOC Downmix Pre-Processor

In the following, the operation of the SAOC downmix pre-processor 270 will be described both for some decoding modes of operation and for some transcoding modes of operation.

4.1 Operation in the Decoding Modes

4.1.1 Introduction

In the following, a method for obtaining an output signal using SAOC parameters and panning information (or rendering information) associated with each audio object is described. The SAOC decoder 495 is depicted in Fig. 4g and consists of the SAOC parameter processor 496 and the downmix processor 497.

It should be noted that the SAOC decoder 494 may be used to process the regular audio objects, and may therefore receive, as the downmix signal 497a, the second audio object signal 264 or the regular-audio-object signal 322 or the second audio information 134. Accordingly, the downmix processor 497 may provide, as its output signals 497b, the processed version 272 of the second audio object signal 264 or the processed version 142 of the second audio information 134. Accordingly, the downmix processor 497 may take

the role of the SAOC downmix pre-processor 270, or the role of the audio signal processor 140.

The SAOC parameter processor 496 may take the role of the SAOC parameter processor 252 and consequently provides downmix information 496a.

4.1.2 Downmix Processor

In the following, the downmix processor, which is part of the audio signal processor 140, and which is designated as a "SAOC downmix pre-processor" 270 in the embodiment of Fig. 2, and which is designated with 497 in the SAOC decoder 495, will be described in more detail.

For the decoder mode of the SAOC system, the output signal 142, 272, 497b of the downmix processor (represented in the hybrid QMF domain) is fed into the corresponding synthesis filterbank (not shown in Figs. 1 and 2) as described in ISO/IEC 23003-1: 2007 yielding the final output PCM signal. Nevertheless, the output signal 142, 272, 497b of the downmix processor is typically combined with one or more audio signals 132, 262 representing the enhanced audio objects. This combination may be performed before the corresponding synthesis filterbank (such that a combined signal combining the output of the downmix processor and the one or more signals representing the enhanced audio objects is input to the synthesis filterbank). Alternatively, the output signal of the downmix processor may be combined with one or more audio signals representing the enhanced audio objects only after the synthesis filterbank processing. Accordingly, the upmix signal representation 120, 220 may be either a QMF domain representation or a PCM domain representation (or any other appropriate representation). The downmix processing incorporates, for example, the mono processing, the stereo processing and, if required, the subsequent binaural processing.

The output signal $\hat{\mathbf{X}}$ of the downmix processor 270, 497 (also designated with 142, 272, 497b) is computed from the mono downmix signal $\mathbf{X}$ (also designated with 134, 264, 497a) and the decorrelated mono downmix signal $\mathbf{X}_d$ as

$$\hat{\mathbf{X}} = \mathbf{G}\mathbf{X} + \mathbf{P}_2\mathbf{X}_d.$$

The decorrelated mono downmix signal $\mathbf{X}_d$ is computed as

$$\mathbf{X}_d = decorrFunc(\mathbf{X}).$$

The decorrelated signals $\mathbf{X}_d$ are created from the decorrelator described in ISO/IEC 23003-1:2007, subclause 6.6.2. Following this scheme, the bsDecorrConfig == 0 configuration should be used with a decorrelator index, $X = 8$, according to Table A.26 to Table A.29 in ISO/IEC 23003-1:2007. Hence, the $decorrFunc(\,)$ denotes the decorrelation process:

$$\mathbf{X}_d = \begin{pmatrix} x_{1d} \\ x_{2d} \end{pmatrix} = \begin{pmatrix} decorrFunc\left((1 \quad 0)\mathbf{P}_1\mathbf{X}\right) \\ decorrFunc\left((0 \quad 1)\mathbf{P}_1\mathbf{X}\right) \end{pmatrix}.$$

In case of binaural output the upmix parameters $\mathbf{G}$ and $\mathbf{P}_2$ derived from the SAOC data, rendering information $\mathbf{M}_{ren}^{l,m}$ and HRTF parameters are applied to the downmix signal $\mathbf{X}$ (and $\mathbf{X}_d$) yielding the binaural output $\hat{\mathbf{X}}$, see Fig. 2, reference numeral 270, where the basic structure of the downmix processor is shown.

The target binaural rendering matrix $\mathbf{A}^{l,m}$ of size $2 \times N$ consists of the elements $a_{x,y}^{l,m}$. Each element $a_{x,y}^{l,m}$ is derived from HRTF parameters and rendering matrix $\mathbf{M}_{ren}^{l,m}$ with elements $m_{y,i}^{l,m}$, for example, by the SAOC parameter processor. The target binaural rendering matrix $\mathbf{A}^{l,m}$ represents the relation between all audio input objects y and the desired binaural output.

$$a_{y,1}^{l,m} = \sum_{i=0}^{N_{HRTF}-1} m_{y,i}^{l,m} H_{i,L}^{m} \exp\left(j\frac{\phi_i^m}{2}\right), \qquad a_{y,2}^{l,m} = \sum_{i=0}^{N_{HRTF}-1} m_{y,i}^{l,m} H_{i,R}^{m} \exp\left(-j\frac{\phi_i^m}{2}\right).$$

The HRTF parameters are given by $H_{i,L}^m$, $H_{i,R}^m$ and ϕ_i^m for each processing band m. The spatial positions for which HRTF parameters are available are characterized by the index i. These parameters are described in ISO/IEC 23003-1:2007.

4.1.2.1 Overview

In the following, an overview over the downmix processing will be given taking reference to Figs. 4a and 4b, which show a block representation of the downmix processing, which may be performed by the audio signal processor 140 or by the combination of the SAOC

parameter processor 252 and the SAOC downmix pre-processor 270, or by the combination of the SAOC parameter processor 496 and the downmix processor 497.

Taking reference now to Fig. 4a, the downmix processing receives a rendering matrix **M**, an object level difference information OLD, an inter-object-correlation information IOC, a downmix gain information DMG and (optionally) a downmix channel level difference information DCLD. The downmix processing 400 according to Fig. 4a obtains a rendering matrix **A** on the basis of the rendering matrix **M**, for example, using a parameter adjuster and a **M**-to-**A** mapping. Also, entries of a covariance matrix **E** are obtained in dependence on the object level difference information OLD and the inter-object correlation information IOC, for example, as discussed above. Similarly, entries of a downmix matrix **D** are obtained in dependence on the downmix gain information DMG and the downmix channel level difference information DCLD.

Entries f of a desired covariance matrix **F** are obtained in dependence on the rendering matrix **A** and the covariance matrix **E**. Also, a scalar value v is obtained in dependence on the covariance matrix **E** and the downmix matrix **D** (or in dependence on the entries thereof).

Gain values P_L, P_R for two channels are obtained in dependence on entries of the desired covariance matrix **F** and the scalar value v. Also, an inter-channel phase difference value φ_C is obtained in dependence entries f of the desired covariance matrix **F**. A rotation angle α is also obtained in dependence on entries f of the desired covariance matrix **F**, taking into consideration, for example, a constant c. In addition, a second rotation angle β is obtained, for example, in dependence on the channel gains P_L, P_R and the first rotation angle α. Entries of a matrix **G** are obtained, for example, in dependence on the two channel gain values P_L,P_R and also in dependence on the inter-channel phase difference φ_C and, optionally, the rotation angles α, β. Similarly, entries of a matrix $\mathbf{P}_2$ are determined in dependence on some or all of said values P_L, P_R, φ_c, α, β.

In the following, it will be described how the matrix **G** and/or $\mathbf{P}_2$ (or the entries thereof), which may be applied by the downmix processor as discussed above, can be obtained for different processing modes.

4.1.2.2 Mono to Binaural "x-1-b" Processing Mode

In the following, a processing mode will be discussed in which the regular audio objects are represented by a single channel downmix signal 134, 264, 322, 497a and in which a binaural rendering is desired.

The upmix parameters $\mathbf{G}^{l,m}$ and $\mathbf{P}_2^{l,m}$ are computed as

$$\mathbf{G}^{l,m} = \begin{pmatrix} P_L^{l,m} \exp\left(j\frac{\phi_C^{l,m}}{2} \right) \cos\left(\beta^{l,m} + \alpha^{l,m} \right) \\ P_R^{l,m} \exp\left(-j\frac{\phi_C^{l,m}}{2} \right) \cos\left(\beta^{l,m} - \alpha^{l,m} \right) \end{pmatrix},$$

$$\mathbf{P}_2^{l,m} = \begin{pmatrix} P_L^{l,m} \exp\left(j\frac{\phi_C^{l,m}}{2} \right) \sin\left(\beta^{l,m} + \alpha^{l,m} \right) \\ P_R^{l,m} \exp\left(-j\frac{\phi_C^{l,m}}{2} \right) \sin\left(\beta^{l,m} - \alpha^{l,m} \right) \end{pmatrix}.$$

The gains $P_L^{l,m}$ and $P_R^{l,m}$ for the left and right output channels are

$$P_L^{l,m} = \sqrt{\max\left(\frac{f_{1,1}^{l,m}}{v^{l,m}}, \varepsilon^2 \right)}, \qquad P_R^{l,m} = \sqrt{\max\left(\frac{f_{2,2}^{l,m}}{v^{l,m}}, \varepsilon^2 \right)}.$$

The desired covariance matrix $\mathbf{F}^{l,m}$ of size 2×2 with elements $f_{i,j}^{l,m}$ is given as

$$\mathbf{F}^{l,m} = \mathbf{A}^{l,m} \mathbf{E}^{l,m} \left(\mathbf{A}^{l,m} \right)^*.$$

The scalar $v^{l,m}$ is computed as

$$v^{l,m} = \mathbf{D}^l \mathbf{E}^{l,m} \left(\mathbf{D}^l \right)^* + \varepsilon^2.$$

The inter channel phase difference $\phi_C^{l,m}$ is given as

$$\phi_C^{l,m} = \begin{cases} \arg\left(f_{1,2}^{l,m} \right), & 0 \le m \le 11, \quad \rho_C^{l,m} \ge 0.6, \\ 0, & otherwise. \end{cases}$$

The inter channel coherence $\rho_C^{l,m}$ is computed as

$$\rho_C^{l,m} = \min\left(\frac{\left| f_{1,2}^{l,m} \right|}{\sqrt{\max\left(f_{1,1}^{l,m} f_{2,2}^{l,m}, \varepsilon^2 \right)}}, 1 \right).$$

The rotation angles $\alpha^{l,m}$ and $\beta^{l,m}$ are given as

$$\alpha^{l,m} = \begin{cases} \frac{1}{2}\arccos\left(\rho_C^{l,m}\cos\left(\arg\left(f_{1,2}^{l,m}\right)\right)\right), & 0 \le m \le 11, \quad \rho_C^{l,m} < 0.6, \\ \frac{1}{2}\arccos\left(\rho_C^{l,m}\right), & \textit{otherwise.} \end{cases}$$

$$\beta^{l,m} = \arctan\left(\tan\left(\alpha^{l,m}\right)\frac{P_R^{l,m} - P_L^{l,m}}{P_L^{l,m} + P_R^{l,m} + \varepsilon}\right).$$

4.1.2.3 Mono-to-Stereo "x-1-2" Processing Mode

In the following, a processing mode will be described in which the regular audio objects are represented by a single-channel signal 134, 264, 222, and in which a stereo rendering is desired.

In case of stereo output the "x-1-b" processing mode can be applied without using HRTF information. This can be done by deriving all elements $a_{x,y}^{l,m}$ of the rendering matrix $\mathbf{A}$, yielding:

$$a_{1,y}^{l,m} = m_{Lf,y}^{l,m}, \qquad a_{2,y}^{l,m} = m_{Rf,y}^{l,m}.$$

4.1.2.4 Mono-to-Mono "x-1-1" Processing Mode

In the following, a processing mode will be described in which the regular audio objects are represented by a signal channel 134, 264, 322, 497a and in which a two-channel rendering of the regular audio objects is desired.

In case of mono output the "x-1-2" processing mode can be applied with the following entries:

$$a_{1,y}^{l,m} = m_{C,y}^{l,m}, \quad a_{2,y}^{l,m} = 0$$

4.1.2.5 Stereo-to-binaural "x-2-b" processing mode

In the following, a processing mode will be described in which regular audio objects are represented by a two-channel signal 134, 264, 322, 497a, and in which a binaural rendering of the regular audio objects is desired.

The upmix parameters $\mathbf{G}^{l,m}$ and $\mathbf{P}_2^{l,m}$ are computed as

$$\mathbf{G}^{l,m} = \begin{pmatrix} P_L^{l,m,1} \exp\left(j\frac{\phi^{l,m,1}}{2} \right) \cos\left(\beta^{l,m} + \alpha^{l,m} \right) & P_L^{l,m,2} \exp\left(j\frac{\phi^{l,m,2}}{2} \right) \cos\left(\beta^{l,m} + \alpha^{l,m} \right) \\ P_R^{l,m,1} \exp\left(-j\frac{\phi^{l,m,1}}{2} \right) \cos\left(\beta^{l,m} - \alpha^{l,m} \right) & P_R^{l,m,2} \exp\left(-j\frac{\phi^{l,m,2}}{2} \right) \cos\left(\beta^{l,m} - \alpha^{l,m} \right) \end{pmatrix},$$

$$\mathbf{P}_2^{l,m} = \begin{pmatrix} P_L^{l,m} \exp\left(j\frac{\arg\left(c_{1,2}^{l,m} \right)}{2} \right) \sin\left(\beta^{l,m} + \alpha^{l,m} \right) \\ P_R^{l,m} \exp\left(-j\frac{\arg\left(c_{1,2}^{l,m} \right)}{2} \right) \sin\left(\beta^{l,m} - \alpha^{l,m} \right) \end{pmatrix}.$$

The corresponding gains $P_L^{l,m,x}$, $P_R^{l,m,x}$ and $P_L^{l,m}$, $P_R^{l,m}$ for the left and right output channels are

$$P_L^{l,m,x} = \sqrt{\max\left(\frac{f_{1,1}^{l,m,x}}{v^{l,m,x}}, \varepsilon^2 \right)}, \quad P_R^{l,m,x} = \sqrt{\max\left(\frac{f_{2,2}^{l,m,x}}{v^{l,m,x}}, \varepsilon^2 \right)},$$

$$P_L^{l,m} = \sqrt{\max\left(\frac{c_{1,1}^{l,m}}{v^{l,m}}, \varepsilon^2 \right)}, \qquad P_R^{l,m} = \sqrt{\max\left(\frac{c_{2,2}^{l,m}}{v^{l,m}}, \varepsilon^2 \right)}.$$

The desired covariance matrix $\mathbf{F}^{l,m,x}$ of size 2×2 with elements $f_{u,v}^{l,m,x}$ is given as

$$\mathbf{F}^{l,m,x} = \mathbf{A}^{l,m} \mathbf{E}^{l,m,x} \left(\mathbf{A}^{l,m} \right)^*.$$

The covariance matrix $\mathbf{C}^{l,m}$ of size 2×2 with elements $c_{u,v}^{l,m}$ of the "dry" binaural signal is estimated as

$$\mathbf{C}^{l,m} = \tilde{\mathbf{G}}^{l,m} \mathbf{D}^l \mathbf{E}^{l,m} \left(\mathbf{D}^l \right)^* \left(\tilde{\mathbf{G}}^{l,m} \right)^*,$$

where

$$\tilde{\mathbf{G}}^{l,m} = \begin{pmatrix} P_L^{l,m,1} \exp\left(j\frac{\phi^{l,m,1}}{2} \right) & P_L^{l,m,2} \exp\left(j\frac{\phi^{l,m,2}}{2} \right) \\ P_R^{l,m,1} \exp\left(-j\frac{\phi^{l,m,1}}{2} \right) & P_R^{l,m,2} \exp\left(-j\frac{\phi^{l,m,2}}{2} \right) \end{pmatrix}.$$

The corresponding scalars $v^{l,m,x}$ and $v^{l,m}$ are computed as

$$v^{l,m,x} = \mathbf{D}^{l,x}\mathbf{E}^{l,m}\left(\mathbf{D}^{l,x}\right)^{*} + \varepsilon^2, \quad v^{l,m} = \left(\mathbf{D}^{l,1} + \mathbf{D}^{l,2}\right)\mathbf{E}^{l,m}\left(\mathbf{D}^{l,1} + \mathbf{D}^{l,2}\right)^{*} + \varepsilon^2.$$

The downmix matrix $\mathbf{D}^{l,x}$ of size $1 \times N$ with elements $d_i^{l,x}$ can be found as

$$d_i^{l,1} = 10^{0.05DMG_i^l}\sqrt{\frac{10^{0.1DCLD_i^l}}{1+10^{0.1DCLD_i^l}}}, \qquad d_i^{l,2} = 10^{0.05DMG_i^l}\sqrt{\frac{1}{1+10^{0.1DCLD_i^l}}}.$$

The stereo downmix matrix $\mathbf{D}^l$ of size $2 \times N$ with elements $d_{x,i}^l$ can be found as

$$d_{x,i}^l = d_i^{l,x}.$$

The matrix $\mathbf{E}^{l,m,x}$ with elements $e_{i,j}^{l,m,x}$ are derived from the following relationship

$$e_{i,j}^{l,m,x} = e_{i,j}^{l,m}\left(\frac{d_i^{l,x}}{d_i^{l,1}+d_i^{l,2}}\right)\left(\frac{d_j^{l,x}}{d_j^{l,1}+d_j^{l,2}}\right).$$

The inter channel phase differences $\phi_C^{l,m}$ are given as

$$\phi^{l,m,x} = \begin{cases} \arg\left(f_{1,2}^{l,m,x}\right), & 0 \le m \le 11, \quad \rho_C^{l,m} > 0.6, \\ 0, & otherwise. \end{cases}$$

The ICCs $\rho_C^{l,m}$ and $\rho_T^{l,m}$ are computed as

$$\rho_T^{l,m} = \min\left(\frac{\left|f_{1,2}^{l,m}\right|}{\sqrt{\max\left(f_{1,1}^{l,m}f_{2,2}^{l,m},\varepsilon^2\right)}},1\right), \qquad \rho_C^{l,m} = \min\left(\frac{\left|c_{1,2}^{l,m}\right|}{\sqrt{\max\left(c_{1,1}^{l,m}c_{2,2}^{l,m},\varepsilon^2\right)}},1\right).$$

The rotation angles $\alpha^{l,m}$ and $\beta^{l,m}$ are given as

$$\alpha^{l,m} = \frac{1}{2}\left(\arccos\left(\rho_T^{l,m}\right) - \arccos\left(\rho_C^{l,m}\right)\right), \qquad \beta^{l,m} = \arctan\left(\tan(\alpha^{l,m})\frac{P_R^{l,m}-P_L^{l,m}}{P_L^{l,m}+P_R^{l,m}}\right).$$

4.1.2.6 Stereo-to-stereo "x-2-2" processing mode

In the following, a processing mode will be described in which the regular audio objects are described by a two-channel (stereo) signal 134, 264, 322, 497a and in which a 2-channel (stereo) rendering is desired.

In case of stereo output, the stereo preprocessing is directly applied, which will be described below in Section 4.2.2.3.

4.1.2.7 Stereo-to-mono "x-2-1" processing mode

In the following, a processing mode will be described in which the regular audio objects are represented by a two-channel (stereo) signal 134, 264, 322, 497a, and in which a one-channel (mono) rendering is desired.

In case of mono output, the stereo preprocessing is applied with a single active rendering matrix entry, as described below in Section 4.2.2.3.

4.1.2.8 Conclusion

Taking reference again to Figs. 4a and 4b, a processing has been described which can be applied to a 1-channel or a two-channel signal 134, 264, 322, 497a representing the regular audio objects subsequent to a separation between the extended audio objects and the regular audio objects. Figs. 4a and 4b illustrate the processing, wherein the processing of Figs. 4a and 4b differs in that an optional parameter adjustment is introduced in different stages of the processing.

4.2. Operation in the transcoding modes

4.2.1 Introduction

In the following, a method for combining SAOC parameters and panning information (or rendering information) associated with each audio object (or, preferably, with each regular audio object) in a standard compliant MPEG surround bitstream (MPS bitstream) is explained.

The SAOC transcoder 490 is depicted in Fig. 4f and consists of an SAOC parameter processor 491 and a downmix processor 492 applied for a stereo downmix.

The SAOC transcoder 490 may, for example, take over the functionality of the audio signal processor 140. Alternatively, the SAOC transcoder 490 may take over the functionality of the SAOC downmix pre-processor 270 when taken in combination with the SAOC parameter processor 252.

For example, the SAOC parameter processor 491 may receive an SAOC bitstream 491a, which is equivalent to the object-related parametric information 110 or the SAOC bitstream 212. Also, the SAOC parameter processor 491 may receive a rendering matrix

information 491b, which may be included in the object-related parametric information 110, or which may be equivalent to the rendering matrix information 214. The SAOC parameter processor 491 may also provide downmix processing information 491c to the downmix processor 492, which may be equivalent to the information 240. Moreover, the SAOC parameter processor 491 may provide an MPEG surround bitstream (or MPEG surround parameter bitstream) 491d, which comprises a parametric surround information which is compatible with the MPEG surround standard. The MPEG surround bitstream 491d may, for example, be part of the processed version 142 of the second audio information, or may, for example be part of or take the place of the MPS bitstream 222.

The downmix processor 492 is configured to receive a downmix signal 492a, which is preferably a one-channel downmix signal or a two-channel downmix signal, and which is preferably equivalent to the second audio information 134, or to the second audio object signal 264, 322. The downmix processor 492 may also provide an MPEG surround downmix signal 492b, which is equivalent to (or part of) the processed version 142 of the second audio information 134, or equivalent to (or part of) the processed version 272 of the second audio object signal 264.

However, there are different ways of combining the MPEG surround downmix signal 492b with the enhanced audio object signal 132, 262. The combination may be performed in the MPEG surround domain.

Alternatively, however, the MPEG surround representation, comprising the MPEG surround parameter bitstream 491d and the MPEG surround downmix signal 492b, of the regular audio objects may be converted back to a multi-channel time domain representation or a multi-channel frequency domain representation (individually representing different audio channels) by an MPEG surround decoder and may be subsequently combined with the enhanced audio object signals.

It should be noted that the transcoding modes comprise both one or more mono downmix processing modes and one or more stereo downmix processing modes. However, in the following only the stereo downmix processing mode will be described, because the processing of the regular audio object signals is more elaborate in the stereo downmix processing mode.

4.2.2 Downmix processing in the stereo downmix ("x-2-5") processing mode

4.2.2.1 Introduction

In the following section, a description of the SAOC transcoding mode for the stereo downmix case will be given.

The object parameters (object level difference OLD, inter-object correlation IOC, downmix gain DMG and downmix channel level difference DCMD) from the SAOC bitstream are transcoded into spatial (preferably channel-related) parameters (channel level difference CLD, inter-channel-correlation ICC, channel prediction coefficient CPC) for the MPEG surround bitstream according to the rendering information. The downmix is modified according to object parameters and a rendering matrix.

Taking reference now to Figs. 4c, 4d and 4e, an overview of the processing, and in particular of the downmix modification, will be given.

Fig. 4c shows a block representation of a processing which is performed for modifying the downmix signal, for example the downmix signal 134, 264, 322,492a describing the one or, preferably, more regular audio objects. As can be seen from Figs. 4c, 4d and 4e, the processing receives a rendering matrix $\mathbf{M}_{ren}$, a downmix gain information DMG, a downmix channel level difference information DCLD, an object level difference information OLD, and an inter-object-correlation information IOC. The rendering matrix may optionally be modified by a parameter adjustment, as it is shown in Fig. 4c. Entries of a downmix matrix D are obtained in dependence on the downmix gain information DMG and the downmix channel level difference information DCLD. Entries of a coherence matrix **E** are obtained in dependence on the object level difference information OLD and the inter-object correlation information IOC. In addition, a matrix **J** may be obtained in dependence on the downmix matrix **D** and the coherence matrix **E,** or in dependence on the entries thereof. Subsequently, a matrix C_3 may be obtained in dependence on the rendering matrix $\mathbf{M}_{ren}$, the downmix matrix **D**, the coherence matrix **E** and the matrix **J**. A matrix **G** may be obtained in dependence on a matrix $\mathbf{D}_{TTT}$, which may be a matrix having predetermined entries, and also in dependence on the matrix $\mathbf{C}_3$. The matrix **G** may, optionally, be modified, to obtain a modified matrix $\mathbf{G}_{mod}$. The matrix **G** or the modified version $\mathbf{G}_{mod}$ thereof may be used to derive the processed version 142, 272,492b of the second audio information 134, 264 from the second audio information 134, 264,492a (wherein the second audio information 134, 264 is designed with **X**, and wherein the processed version 142, 272 thereof is designated with $\hat{\mathbf{X}}$.

In the following, the rendering of the object energy, which is performed in order to obtain the MPEG surround parameters, will be discussed. Also, the stereo preprocessing, which is

performed in order to obtain the processed version 142, 272,492b of the second audio information 134, 264,492a representing the regular audio objects will be described.

4.2.2.2 Rendering of object energies

The transcoder determines the parameters for the MPS decoder according to the target rendering as described by the rendering matrix $\mathbf{M}_{ren}$. The six channel target covariance is denoted with $\mathbf{F}$ and given by

$$\mathbf{F} = \mathbf{YY}^* = \mathbf{M}_{ren}\mathbf{S}(\mathbf{M}_{ren}\mathbf{S})^* = \mathbf{M}_{ren}(\mathbf{SS}^*)\mathbf{M}_{ren}^* = \mathbf{M}_{ren}\mathbf{E}\mathbf{M}_{ren}^*.$$

The transcoding process can conceptually be divided into two parts. In one part a three channel rendering is performed to a left, right and center channel. In this stage the parameters for the downmix modification as well as the prediction parameters for the TTT box for the MPS decoder are obtained. In the other part the CLD and ICC parameters for the rendering between the front and surround channels (OTT parameters, left front – left surround, right front – right surround) are determined.

4.2.2.2.1 Rendering to left, right and center channel

In this stage the spatial parameters are determined that control the rendering to a left and right channel, consisting of front and surround signals. These parameters describe the prediction matrix of the TTT box for the MPS decoding $\mathbf{C}_{TTT}$ (CPC parameters for the MPS decoder) and the downmix converter matrix $\mathbf{G}$.

$\mathbf{C}_{TTT}$ is the prediction matrix to obtain the target rendering from the modified downmix $\hat{\mathbf{X}} = \mathbf{GX}$:

$$\mathbf{C}_{TTT}\hat{\mathbf{X}} = \mathbf{C}_{TTT}\mathbf{GX} \approx \mathbf{A}_3\mathbf{S}.$$

$\mathbf{A}_3$ is a reduced rendering matrix of size $3 \times N$, describing the rendering to the left, right and center channel respectively. It is obtained as $\mathbf{A}_3 = \mathbf{D}_{36}\mathbf{M}_{ren}$ with the 6 to 3 partial downmix matrix $\mathbf{D}_{36}$ defined by

$$\mathbf{D}_{36} = \begin{pmatrix} w_1 & 0 & 0 & 0 & w_1 & 0 \\ 0 & w_2 & 0 & 0 & 0 & w_2 \\ 0 & 0 & w_3 & w_3 & 0 & 0 \end{pmatrix}.$$

The partial downmix weights w_p, $p=1,2,3$ are adjusted such that the energy of $w_p(y_{2p-1}+y_{2p})$ is equal to the sum of energies $\|y_{2p-1}\|^2+\|y_{2p}\|^2$ up to a limit factor.

$$w_1=\frac{f_{1,1}+f_{5,5}}{f_{1,1}+f_{5,5}+2f_{1,5}},\qquad w_2=\frac{f_{2,2}+f_{6,6}}{f_{2,2}+f_{6,6}+2f_{2,6}},\qquad w_3=0.5,$$

where $f_{i,j}$ denote the elements of $\mathbf{F}$.

For the estimation of the desired prediction matrix $\mathbf{C}_{TTT}$ and the downmix preprocessing matrix $\mathbf{G}$ we define a prediction matrix $\mathbf{C}_3$ of size 3×2, that leads to the target rendering

$$\mathbf{C}_3\mathbf{X}\approx\mathbf{A}_3\mathbf{S}.$$

Such a matrix is derived by considering the normal equations

$$\mathbf{C}_3\left(\mathbf{D}\mathbf{E}\mathbf{D}^*\right)\approx\mathbf{A}_3\mathbf{E}\mathbf{D}^*.$$

The solution to the normal equations yields the best possible waveform match for the target output given the object covariance model. $\mathbf{G}$ and $\mathbf{C}_{TTT}$ are now obtained by solving the system of equations

$$\mathbf{C}_{TTT}\mathbf{G}=\mathbf{C}_3.$$

To avoid numerical problems when calculating the term $\mathbf{J}=\left(\mathbf{D}\mathbf{E}\mathbf{D}^*\right)^{-1}$, $\mathbf{J}$ is modified. First the eigenvalues $\lambda_{1,2}$ of $\mathbf{J}$ are calculated, solving $\det(\mathbf{J}-\lambda_{1,2}\mathbf{I})=0$.

Eigenvalues are sorted in descending ($\lambda_1\geq\lambda_2$) order and the eigenvector corresponding to the larger eigenvalue is calculated according to the equation above. It is assured to lie in the positive x-plane (first element has to be positive). The second eigenvector is obtained from the first by a – 90 degrees rotation:

$$\mathbf{J}=\left(\mathbf{v}_1\mathbf{v}_2\right)\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}\left(\mathbf{v}_1\mathbf{v}_2\right)^*.$$

A weighting matrix is computed from the downmix matrix $\mathbf{D}$ and the prediction matrix $\mathbf{C}_3$, $\mathbf{W}=\left(\mathbf{D}\ diag(\mathbf{C}_3)\right)$.

Since $\mathbf{C}_{TTT}$ is a function of the MPS prediction parameters c_1 and c_2 (as defined in ISO/IEC 23003-1:2007), $\mathbf{C}_{TTT}\mathbf{G}=\mathbf{C}_3$ is rewritten in the following way, to find the stationary point or points of the function,

$$\mathbf{\Gamma}\begin{pmatrix}\tilde{c}_1\\ \tilde{c}_2\end{pmatrix}=\mathbf{b},$$

with $\mathbf{\Gamma}=(\mathbf{D}_{TTT}\,\mathbf{C}_3)\mathbf{W}(\mathbf{D}_{TTT}\,\mathbf{C}_3)^*$ and $\mathbf{b}=\mathbf{GWC}_3\mathbf{v}$,

where $\mathbf{D}_{TTT}=\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 1\end{pmatrix}$ and $\mathbf{v}=\begin{pmatrix}1 & 1 & -1\end{pmatrix}$.

If $\mathbf{\Gamma}$ does not provide a unique solution ($\det(\mathbf{\Gamma})<10^{-3}$), the point is chosen that lies closest to the point resulting in a TTT pass through. As a first step, the row i of $\mathbf{\Gamma}$ is chosen $\boldsymbol{\gamma}=[\gamma_{i,1}\quad\gamma_{i,2}]$ where the elements contain most energy, thus $\gamma_{i,1}^{\,2}+\gamma_{i,2}^{\,2}\geq\gamma_{j,1}^{\,2}+\gamma_{j,2}^{\,2},\quad j=1,2$. Then a solution is determined such that

$$\begin{pmatrix}\tilde{c}_1\\ \tilde{c}_2\end{pmatrix}=\begin{pmatrix}1\\ 1\end{pmatrix}-3\mathbf{y}\ \text{with}\ \mathbf{y}=\frac{b_{i,3}}{\left(\sum_{j=1,2}\left(\gamma_{i,j}\right)^2\right)+\varepsilon}\boldsymbol{\gamma}^T.$$

If the obtained solution for $\tilde{c}_1$ and $\tilde{c}_2$ is outside the allowed range for prediction coefficients that is defined as $-2\leq\tilde{c}_j\leq 3$ (as defined in ISO/IEC 23003-1:2007), $\tilde{c}_j$ shall be calculated according to below.

First define the set of points, $\mathbf{x}_p$ as:

$$\mathbf{x}_p\in\begin{bmatrix}\begin{pmatrix}\min\left(3,\max\left(-2,-\frac{-2\gamma_{1,2}-b_1}{\gamma_{1,1}+\varepsilon}\right)\right)\\ -2\end{pmatrix},\begin{pmatrix}\min\left(3,\max\left(-2,-\frac{3\gamma_{1,2}-b_1}{\gamma_{1,1}+\varepsilon}\right)\right)\\ 3\end{pmatrix}\\ \begin{pmatrix}-2\\ \min\left(3,\max\left(-2,-\frac{-2\gamma_{2,1}-b_2}{\gamma_{2,2}+\varepsilon}\right)\right)\end{pmatrix},\begin{pmatrix}3\\ \min\left(3,\max\left(-2,-\frac{3\gamma_{2,1}-b_2}{\gamma_{2,2}+\varepsilon}\right)\right)\end{pmatrix}\end{bmatrix},$$

and the distance function,

$$distFunc(\mathbf{x}_p)=\mathbf{x}_p^*\mathbf{\Gamma}\mathbf{x}_{p1}-2\mathbf{b}\mathbf{x}_p.$$

Then the prediction parameters are defined according to:

$$\begin{pmatrix}\tilde{c}_1\\ \tilde{c}_2\end{pmatrix}=\arg\min_{\mathbf{x}\in\mathbf{x}_p}\left(distFunc(\mathbf{x})\right).$$

The prediction parameters are constrained according to:

$$c_1 = (1-\lambda)\tilde{c}_1 + \lambda\gamma_1, \quad c_2 = (1-\lambda)\tilde{c}_2 + \lambda\gamma_2,$$

where λ, γ_1 and γ_2 are defined as

$$\gamma_1 = \frac{2f_{1,1} + 2f_{5,5} - f_{3,3} + f_{1,3} + f_{5,3}}{2f_{1,1} + 2f_{5,5} + 2f_{3,3} + 4f_{1,3} + 4f_{5,3}},$$
$$\gamma_2 = \frac{2f_{2,2} + 2f_{6,6} - f_{3,3} + f_{2,3} + f_{6,3}}{2f_{2,2} + 2f_{6,6} + 2f_{3,3} + 4f_{2,3} + 4f_{6,3}},$$

$$\lambda = \left(\frac{\left(f_{1,2} + f_{1,6} + f_{5,2} + f_{5,6} + f_{1,3} + f_{5,3} + f_{2,3} + f_{6,3} + f_{3,3}\right)^2}{\left(f_{1,1} + f_{5,5} + f_{3,3} + 2f_{1,3} + 2f_{5,3}\right)\left(f_{2,2} + f_{6,6} + f_{3,3} + 2f_{2,3} + 2f_{6,3}\right)} \right)^8.$$

For the MPS decoder, the CPCs and corresponding ICC_{TTT} are provided as follows

$$\mathbf{D}_{CPC\_1} = c_1(l,m), \ \mathbf{D}_{CPC\_2} = c_2(l,m) \text{ and } \mathbf{D}_{ICC_{TTT}} = 1.$$

4.2.2.2.2 Rendering between front and surround channels

The parameters that determine the rendering between front and surround channels can be estimated directly from the target covariance matrix **F**

$$CLD_{a,b} = 10\log_{10}\left(\frac{\max\left(f_{a,a}, \varepsilon^2\right)}{\max\left(f_{b,b}, \varepsilon^2\right)}\right), \qquad ICC_{a,b} = \frac{\max\left(f_{a,b}, \varepsilon^2\right)}{\sqrt{\max\left(f_{a,a}, \varepsilon^2\right)\max\left(f_{b,b}, \varepsilon^2\right)}},$$

with (a,b) = (1,2) and (3,4).

The MPS parameters are provided in the form

$$CLD_h^{l,m} = \mathbf{D}_{CLD}(h,l,m) \text{ and } ICC_h^{l,m} = \mathbf{D}_{ICC}(h,l,m),$$

for every OTT box h.

4.2.2.3 Stereo processing

In the following, a stereo processing of the regular audio object signal 134 to 64, 322 will be described. The stereo processing is used to derive a process to general representation 142, 272 on the basis of a two-channel representation of the regular audio objects.

The stereo downmix $\mathbf{X}$, which is represented by the regular audio object signals 134, 264, 492a is processed into the modified downmix signal $\hat{\mathbf{X}}$, which is represented by the processed regular audio object signals 142, 272:

$$\hat{\mathbf{X}} = \mathbf{G}\mathbf{X},$$

where

$$\mathbf{G} = \mathbf{D}_{TTT}\mathbf{C}_3 = \mathbf{D}_{TTT}\mathbf{M}_{ren}\mathbf{E}\mathbf{D}^*\mathbf{J}.$$

The final stereo output from the SAOC transcoder $\hat{\boldsymbol{X}}$ is produced by mixing $\mathbf{X}$ with a decorrelated signal component according to:

$$\hat{\mathbf{X}} = \mathbf{G}_{Mod}\mathbf{X} + \mathbf{P}_2\mathbf{X}_d,$$

where the decorrelated signal $\mathbf{X}_d$ is calculated as described above, and the mix matrices $\mathbf{G}_{Mod}$ and $\mathbf{P}_2$ according to below.

First, define the render upmix error matrix as

$$\mathbf{R} = \mathbf{A}_{diff}\mathbf{E}\mathbf{A}_{diff}^*,$$

where

$$\mathbf{A}_{diff} = \mathbf{D}_{TTT}\mathbf{A}_3 - \mathbf{G}\mathbf{D},$$

and moreover define the covariance matrix of the predicted signal $\hat{\mathbf{R}}$ as

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{r}_{1,1} & \hat{r}_{1,2} \\ \hat{r}_{2,1} & \hat{r}_{2,2} \end{pmatrix} = \mathbf{G}\mathbf{D}\mathbf{E}\mathbf{D}^*\mathbf{G}^*.$$

The gain vector $\mathbf{g}_{vec}$ can subsequently be calculated as:

$$\mathbf{g}_{vec} = \left(\min\left(\sqrt{\max\left(\frac{\hat{r}_{1,1} + r_{1,1} + \varepsilon^2}{r_{1,1} + \varepsilon^2}, 0 \right)}, 1.5 \right) \quad \min\left(\sqrt{\max\left(\frac{\hat{r}_{2,2} + r_{2,2} + \varepsilon^2}{r_{2,2} + \varepsilon^2}, 0 \right)}, 1.5 \right) \right),$$

and the mix matrix $\mathbf{G}_{\text{Mod}}$ is given as:

$$\mathbf{G}_{\text{Mod}} = \begin{cases} diag(\mathbf{g}_{\text{vec}})\mathbf{G}, & r_{1,2} > 0, \\ \mathbf{G}, & otherwise. \end{cases}$$

Similarly, the mix matrix $\mathbf{P}_2$ is given as:

$$\mathbf{P}_2 = \begin{cases} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & , r_{1,2} > 0, \\ \mathbf{v}_R diag(\mathbf{W}_d) & , otherwise. \end{cases}$$

To derive $\mathbf{v}_R$ and $\mathbf{W}_d$, the characteristic equation of $\mathbf{R}$ needs to be solved:

$\det(\mathbf{R} - \lambda_{1,2}\mathbf{I}) = 0$, giving the eigenvalues, λ_1 and λ_2.

The corresponding eigenvectors $\mathbf{v}_{R1}$ and $\mathbf{v}_{R2}$ of $\mathbf{R}$ can be calculated solving the equation system:

$$(\mathbf{R} - \lambda_{1,2}\mathbf{I})\mathbf{v}_{R1,R2} = 0.$$

Eigenvalues are sorted in descending ($\lambda_1 \geq \lambda_2$) order and the eigenvector corresponding to the larger eigenvalue is calculated according to the equation above. It is assured to lie in the positive x-plane (first element has to be positive). The second eigenvector is obtained from the first by a – 90 degrees rotation:

$$\mathbf{R} = (\mathbf{v}_{R1}\mathbf{v}_{R2})\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}(\mathbf{v}_{R1}\mathbf{v}_{R2})^*.$$

Incorporating $\mathbf{P}_1 = (1 \quad 1)\mathbf{G}$, $\mathbf{R}_d$ can be calculated according to:

$$\mathbf{R}_d = \begin{pmatrix} \mathbf{r}_{d11} & \mathbf{r}_{d12} \\ \mathbf{r}_{d21} & \mathbf{r}_{d22} \end{pmatrix} = diag\left(\mathbf{P}_1(\mathbf{DED}^*)\mathbf{P}_1^*\right),$$

which gives

$$\mathbf{w}_{d1} = \min\left(\sqrt{\frac{\lambda_1}{r_{d1} + \varepsilon}}, 2\right), \qquad \mathbf{w}_{d2} = \min\left(\sqrt{\frac{\lambda_2}{r_{d2} + \varepsilon}}, 2\right),$$

and finally the mix matrix,

$$\mathbf{P}_2 = \begin{pmatrix} \mathbf{v}_{R1} & \mathbf{v}_{R2} \end{pmatrix} \begin{pmatrix} \mathbf{w}_{d1} & 0 \\ 0 & \mathbf{w}_{d2} \end{pmatrix}.$$

4.2.2.4 Dual mode

The SAOC transcoder can let the mix matrices $\mathbf{P}_1$, $\mathbf{P}_2$ and the prediction matrix $\mathbf{C}_3$ be calculated according to an alternative scheme for the upper frequency range. This alternative scheme is particularly useful for downmix signals where the upper frequency range is coded by a non-waveform preserving coding algorithm e.g. SBR in High Efficiency AAC.

For the upper parameter bands, defined by **bsTttBandsLow** $\leq pb < numBands$, $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{C}_3$ should be calculated according to the alternative scheme described below:

$$\begin{cases} \mathbf{P}_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \\ \mathbf{P}_2 = \mathbf{G}. \end{cases}$$

Define the energy downmix and energy target vectors, respectively:

$$\begin{cases} \mathbf{e}_{dmx} = \begin{pmatrix} e_{dmx1} \\ e_{dmx2} \end{pmatrix} = diag\left(\mathbf{DED}^*\right) + \varepsilon\mathbf{I}, \\ \mathbf{e}_{tar} = \begin{pmatrix} e_{tar1} \\ e_{tar2} \\ e_{tar3} \end{pmatrix} = diag\left(\mathbf{A}_3\mathbf{E}\mathbf{A}_3^*\right), \end{cases}$$

and the help matrix

$$\mathbf{T} = \begin{pmatrix} t_{1,1} & t_{1,2} \\ t_{2,1} & t_{2,2} \\ t_{3,1} & t_{3,2} \end{pmatrix} = \mathbf{A}_3\mathbf{D}^* + \varepsilon\mathbf{I}.$$

Then calculate the gain vector

$$\mathbf{g} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} \sqrt{\dfrac{e_{tar1}}{t_{1,1}^2 e_{dmx1} + t_{1,2}^2 e_{dmx2}}} \\ \sqrt{\dfrac{e_{tar2}}{t_{2,1}^2 e_{dmx1} + t_{2,2}^2 e_{dmx2}}} \\ \sqrt{\dfrac{e_{tar3}}{t_{3,1}^2 e_{dmx1} + t_{3,2}^2 e_{dmx2}}} \end{pmatrix},$$

which finally gives the new prediction matrix

$$\mathbf{C}_3 = \begin{pmatrix} g_1 t_{1,1} & g_1 t_{1,2} \\ g_2 t_{2,1} & g_2 t_{2,2} \\ g_3 t_{3,1} & g_3 t_{3,2} \end{pmatrix}.$$

5. Combined EKS SAOC decoding/transcoding mode, encoder according to Fig. 10 and systems according to Figs. 5a, 5b

In the following, a brief description of the combined EKS SAOC processing scheme will be given. A preferred "combined EKS SAOC" processing scheme is proposed, where the EKS processing is integrated into the regular SAOC decoding/transcoding chain by a cascaded scheme.

5.1. Audio signal Encoder according to Fig. 5

In a first step, objects dedicated to EKS processing (enhanced Karaoke/solo processing) are identified as foreground objects (FGO) and their number N_{FGO} (also designated as N_{EAO}) is determined by a bitstream variable "bsNumGroupsFGO". Said bitstream variable may, for example, be included in an SAOC bitstream, as described above.

For the generation of the bitstream (in an audio signal encoder), the parameters of all input objects N_{obj} are reordered such that the foreground objects FGO comprise the last N_{FGO} (or alternatively, N_{EAO}) parameters in each case, for example, OLD_i for $[N_{obj} - N_{FGO} \leq i \leq N_{obj} - 1]$.

From the remaining objects which are, for example, background objects BGO or non-enhanced audio objects, a downmix signal in the "regular SAOC style" is generated which

at the same time serves as a background object BGO. Next, the background object and the foreground objects are downmixed in the "EKS processing style" and residual information is extracted from each foreground object. This way, no extra processing steps need to be introduced. Thus, no change of the bitstream syntax is required.

In other words, at the encoder side, non-enhanced audio objects are distinguished from enhanced audio objects. A one-channel or two-channels regular audio objects downmix signal is provided which represents the regular audio objects (non-enhanced audio objects), wherein there may be one, two or even more regular audio objects (non-enhanced audio objects). The one-channel or two-channel regular audio object downmix signal is then combined with one or more enhanced audio object signals (which may, for example, be one-channel signals or two-channel signals), to obtain a common downmix signal (which may, for example, be a one-channel downmix signal or a two-channel downmix signal) combining the audio signals of the enhanced audio objects and the regular audio object downmix signal.

In the following, the basic structure of such a cascaded encoder will be briefly described taking reference to Fig. 10, which shows a block schematic representation of an SAOC encoder 1000, according to an embodiment of the invention. The SAOC encoder 1000 comprises a first SAOC downmixer 1010, which is typically an SAOC downmixer which does not provide a residual information. The SAOC downmixer 1010 is configured to receive a plurality of N_{BGO} audio object signals 1012 from regular (non-enhanced) audio objects. Also, the SAOC downmixer 1010 is configured to provide a regular audio object downmix signal 1014 on the basis of the regular audio objects 1012, such that the regular audio object downmix signal 1014 combines the regular audio objects signals 1012 in accordance with downmix parameters. The SAOC downmixer 1010 also provides a regular audio object SAOC information 1016, which describes the regular audio object signals and the downmix. For example, the regular audio object SAOC information 1016 may comprise a downmix gain information DMG and a downmix channel level difference information DCLD describing the downmix performed by the SAOC downmixer 1010. In addition, the regular audio object SAOC information 1016 may comprise an object level difference information and an inter-object correlation information describing a relationship between the regular audio objects described by the regular audio object signal 1012.

The encoder 1000 also comprises a second SAOC downmixer 1020, which is typically configured to provide a residual information. The second SAOC downmixer 1020 is

preferably configured to receive one or more enhanced audio object signals 1022 and also to receive the regular audio object downmix signal 1014.

The second SAOC downmixer 1020 is also configured to provide a common SAOC downmix signal 1024 on the basis of the enhanced audio object signals 1022 and the regular audio object downmix signal 1014. When providing the common SAOC downmix signal, the second SAOC downmixer 1020 typically treats the regular audio object downmix signal 1014 as a single one-channel or two-channel object signal.

The second SAOC downmixer 1020 is also configured to provide an enhanced audio object SAOC information which describes, for example, downmix channel level difference values DCLD associated with the enhanced audio objects, object level difference values OLD associated with the enhanced audio objects and inter-object correlation values IOC associated with the enhanced audio objects. In addition, the second SAOC 1020 is preferably configured to provide residual information associated with each of the enhanced audio objects, such that the residual information associated with the enhanced audio objects describes the difference between an original individual enhanced audio object signal and an expected individual enhanced audio object signal which can be extracted from the downmix signal using the downmix information DMG, DCLD and the object information OLD, IOC.

The audio encoder 1000 is well-suited for cooperation with the audio decoder described herein.

5.2. Audio signal decoder according to Fig. 5a

In the following, the basic structure of a combined EKS SAOC decoder 500, a block schematic diagram of which is shown in Fig. 5a will be described.

The audio decoder 500 according to Fig. 5a is configured to receive a downmix signal 510, an SAOC bitstream information 512 and a rendering matrix information 514. The audio decoder 500 comprises an enhanced Karaoke/Solo processing and a foreground object rendering 520, which is configured to provide a first audio object signal 562, which describes rendered foreground objects, and a second audio object signal 564, which describes the background objects. The foreground objects may, for example, be so-called "enhanced audio objects" and the background objects may, for example, be so-called "regular audio objects" or "non-enhanced audio objects". The audio decoder 500 also comprises regular SAOC decoding 570, which is configured to receive the second audio

object signal 562 and to provide, on the basis thereof, a processed version 572 of the second audio object signal 564. The audio decoder 500 also comprises a combiner 580, which is configured to combine the first audio object signal 562 and the processed version 572 of the second audio object signal 564, to obtain an output signal 520.

In the following, the functionality of the audio decoder 500 will be discussed in some more detail. At the SAOC decoding/transcoding side, the upmix process results in a cascaded scheme comprising firstly an enhanced Karaoke-Solo processing (EKS processing) to decompose the downmix signal into the background object (BGO) and foreground objects (FGOs). The required object level differences (OLDs) and inter-object correlations (IOCs) for the background object are derived from the object and downmix information (which is both object-related parametric information, and which is both typically included in the SAOC bitstream):

$$OLD_L = \sum_{i=0}^{N-N_{FGO}-1} d_{0,i}^2 OLD_i$$

$$OLD_R = \sum_{i=0}^{N-N_{FGO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{LR} = \begin{cases} IOC_{0,1}, & N - N_{FGO} = 2, \\ 0, & otherwise. \end{cases}$$

In addition, this step (which is typically executed by the EKS processing and foreground object rendering 520) includes mapping the foreground objects to the final output channels (such that, for example, the first audio object signal 562 is a multi-channel signal in which the foreground objects are mapped to one or more channels each). The background object (which typically comprises a plurality of so-called "regular audio objects") is rendered to the corresponding output channels by a regular SAOC decoding process (or, alternatively, in some cases by an SAOC transcoding process). This process may, for example, be performed by the regular SAOC decoding 570. The final mixing stage (for example, the combiner 580) provides a desired combination of rendered foreground objects and background object signals at the output.

This combined EKS SAOC system represents a combination of all beneficial properties of the regular SAOC system and its EKS mode. This approach allows to achieve the corresponding performance using the proposed system with the same bitstream for both classic (moderate rendering) and Karaoke/Solo-similar (extreme rendering) playback scenarios.

5.3. Generalized Structure according to Fig. 5b

In the following, a generalized structure of a combined EKS SAOC system 590 will be described taking reference to Fig. 5b, which shows a block schematic diagram of such a generalized combined EKS SAOC system. The combined EKS SAOC system 590 of Fig. 5b may also be considered as an audio decoder.

The combined EKS SAOC system 590 is configured to receive a downmix signal 510a, an SAOC bitstream information 512a and the rendering matrix information 514a. Also, the combined EKS SAOC system 590 is configured to provide an output signal 520a on the basis thereof.

The combined EKS SAOC system 590 comprises an SAOC type processing stage I 520a, which receives the downmix signal 510a, the SAOC bitstream information 512a (or at least a part thereof) and the rendering matrix information 514a (or at least a part thereof). In particular, the SAOC type processing stage I 520a receives first stage object level difference values (OLD_s). The SAOC type processing stage I 520a provides one or more signals 562a describing a first set of objects (for example, audio objects of a first audio object type). The SAOC type processing stage I 520a also provides one or more signal 564a describing a second set of objects.

The combined EKS SAOC system also comprises an SAOC type processing stage II 570a, which is configured to receive the one or more signals 564a describing the second set of objects and to provide, on the basis thereof, one or more signals 572a describing a third set of objects using second stage object level differences, which are included in the SAOC bitstream information 512a, and also at least a part of the rendering matrix information 514. The combined EKS SAOC system also comprises a combiner 580a, which may, for example, be a summer, to provide the output signals 520a by combining the one or more signals 562a describing the first set of objects and the one or more signals 570a describing the third set of objects (wherein the third set of objects may be a processed version of the second set of objects).

To summarize the above, Fig. 5b shows a generalized form of the basic structure described with reference to Fig. 5a above in a further embodiment of the invention.

6. Perceptual Evaluation of the Combined EKS SAOC Processing Scheme

6.1 Test Methodology , Design and Items

This subjective listening tests were conducted in an acoustically isolated listening room that is designed to permit high-quality listening. The playback was done using headphones (STAX SR Lambda Pro with Lake-People D/A-Converter and STAX SRM-Monitor). The test method followed the standard procedures used in the spatial audio verification tests, based on the "multiple stimulus with hidden reference and anchors" (MUSHRA) method for the subjective assessment of intermediate quality audio (see reference [7]).

A total of eight listeners participated in the performed test. All subjects can be considered experienced listeners. In accordance with the MUSHRA methodology, the listeners were instructed to compare all test conditions against the reference. The test conditions were randomized automatically for each test item and for each listener. The subjective responses were recorded by a computer-based MUSHRA program on a scale ranging from 0 to 100. An instantaneous switching between the items under test was allowed. The MUSHRA test has been conducted in order to assess the perceptual performance of the considered SAOC modes and the proposed system described in the table of Fig. 6a, which provides a listening test design description.

The corresponding downmix signals were coded using an AAC core-coder with a bitrate of 128 kbps. In order to assess the perceptual quality of the proposed combined EKS SAOC system, it is compared against the regular SAOC RM system (SAOC reference model system) and the current EKS mode (enhanced-Karaoke-Solo mode) for two different rendering test scenarios described in the table of Fig. 6b, which describes the systems under test.

Residual coding with a bit rate of 20 kbps was applied for the current EKS mode and a proposed combined EKS SAOC system. It should be noted that for the current EKS mode it is necessary to generate a stereo background object (BGO) prior to the actual encoding/decoding procedure, since this mode has limitations on the number and type of input objects.

The listening test material and the corresponding downmix and rendering parameters used in the performed tests have been selected from the set of the call-for-proposals (CfP) audio items described in the document [2]. The corresponding data for "Karaoke" and "Classic" rendering application scenarios can be found in the table of Fig. 6c, which describes listening test items and rendering matrices.

6.2 Listening Test Results

A short overview in terms of the diagrams demonstrating the obtained listening test results can be found in Figs. 6d and 6e, wherein Fig. 6d shows average MUSHRA scores for the Karaoke/Solo type rendering listening test, and Fig. 6e shows average MUSHRA scores for the classic rendering listening test. The plots show the average MUSHRA grading per item over all listeners and the statistical mean value over all evaluated items together with the associated 95% confidence intervals.

The following conclusions can be drawn based upon the results of the conducted listening tests:

- Fig. 6d represents the comparison for the current EKS mode with the combined EKS SAOC system for Karaoke-type of applications. For all tested items no significant difference (in the statistical sense) in performance between these two systems can be observed. From this observation it can be concluded that the combined EKS SAOC system is able to efficiently exploit the residual information reaching the performance of the EKS mode. One can also note that the performance of the regular SAOC system (without residual) is below both other systems.

- Fig. 6e represents the comparison of the current regular SAOC with the combined EKS SAOC system for classic rendering scenarios. For all tested items the performance of these two systems is statistically the same. This demonstrates the proper functionality of the combined EKS SAOC system for a classic rendering scenario.

Therefore, it can be concluded that the proposed unified system combining the EKS mode with the regular SAOC preserves the advantages in subjective audio quality for the corresponding types of a rendering.

Taking into account the fact that the proposed combined EKS SAOC system has no longer restrictions on the BGO object, but has entirely flexible rendering capability of the regular SAOC mode and can use the same bitstream for all types of rendering, it appears to be advantageous to incorporate it into the MPEG SAOC standard.

7. Method According to Fig. 7

In the following, a method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information will be described with reference to Fig. 7, which shows a flowchart of such a method.

The method 700 comprises a step 710 of decomposing a downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and at least a part of the object-related parametric information. The method 700 also comprises a step 720 of processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information.

The method 700 also comprises a step 730 of combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation.

The method 700 according to Fig. 7 may be supplemented by any of the features and functionalities which are discussed herein with respect to the inventive apparatus. Also, the method 700 brings along the advantages discussed with respect to the inventive apparatus.

8. Implementation Alternatives

Although some aspects have been described in the context of an apparatus, it is clear that these aspects also represent a description of the corresponding method, where a block or device corresponds to a method step or a feature of a method step. Analogously, aspects described in the context of a method step also represent a description of a corresponding block or item or feature of a corresponding apparatus. Some or all of the method steps may be executed by (or using) a hardware apparatus, like for example, a microprocessor, a

programmable computer or an electronic circuit. In some embodiments, some one or more of the most important method steps may be executed by such an apparatus.

The inventive encoded audio signal can be stored on a digital storage medium or can be transmitted on a transmission medium such as a wireless transmission medium or a wired transmission medium such as the Internet.

Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a digital storage medium, for example a floppy disk, a DVD, a Blue-Ray, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed. Therefore, the digital storage medium may be computer readable.

Some embodiments according to the invention comprise a data carrier having electronically readable control signals, which are capable of cooperating with a programmable computer system, such that one of the methods described herein is performed.

Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing one of the methods when the computer program product runs on a computer. The program code may for example be stored on a machine readable carrier.

Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine readable carrier.

In other words, an embodiment of the inventive method is, therefore, a computer program having a program code for performing one of the methods described herein, when the computer program runs on a computer.

A further embodiment of the inventive methods is, therefore, a data carrier (or a digital storage medium, or a computer-readable medium) comprising, recorded thereon, the computer program for performing one of the methods described herein. The data carrier, the digital storage medium or the recorded medium are typically tangible and/or non-transmitting.

A further embodiment of the inventive method is, therefore, a data stream or a sequence of signals representing the computer program for performing one of the methods described herein. The data stream or the sequence of signals may for example be configured to be transferred via a data communication connection, for example via the Internet.

A further embodiment comprises a processing means, for example a computer, or a programmable logic device, configured to or adapted to perform one of the methods described herein.

A further embodiment comprises a computer having installed thereon the computer program for performing one of the methods described herein.

In some embodiments, a programmable logic device (for example a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

The above described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

9. Conclusions

In the following, some aspects and advantages of the combined EKS SAOC system according to the present invention will be briefly summarized. For Karaoke and Solo playback scenarios, the SAOC EKS processing mode supports both reproduction of the background objects/foreground objects exclusively and an arbitrary mixture (defined by the rendering matrix) of these object groups.

Also, the first mode is considered to be the main objective of EKS processing, the latter provides additional flexibility.

It has been found that a generalization of the EKS functionality consequently involves the effort of combining EKS with the regular SAOC processing mode to obtain one unified system. The potentials of such a unified system are:

- One single clear SAOC decoding/transcoding structure;
- One bitstream for both EKS and regular SAOC mode;
- No limitation to the number of input objects comprising the background object (BGO), such that there is no need to generate the background object prior to the SAOC encoding stage ; and
- Support of a residual coding for foreground objects yielding enhanced perceptual quality in demanding Karaoke/Solo playback situations.

These advantages can be obtained by the unified system described herein.

References

[1] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N8853, "Call for Proposals on Spatial Audio Object Coding", 79th MPEG Meeting, Marrakech, January 2007.

[2] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N9099, "Final Spatial Audio Object Coding Evaluation Procedures and Criterion", 80th MPEG Meeting, San Jose, April 2007.

[3] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N9250, "Report on Spatial Audio Object Coding RM0 Selection", 81st MPEG Meeting, Lausanne, July 2007.

[4] ISO/IEC JTC1/SC29/WG11 (MPEG), Document M15123, "Information and Verification Results for CE on Karaoke/Solo system improving the performance of MPEG SAOC RM0", 83rd MPEG Meeting, Antalya, Turkey, January 2008.

[5] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N10659, "Study on ISO/IEC 23003-2:200x Spatial Audio Object Coding (SAOC)", 88th MPEG Meeting, Maui, USA, April 2009.

[6] ISO/IEC JTC1/SC29/WG11 (MPEG), Document M10660, "Status and Workplan on SAOC Core Experiments", 88th MPEG Meeting, Maui, USA, April 2009.

[7] EBU Technical recommendation: "MUSHRA-EBU Method for Subjective Listening Tests of Intermediate Audio Quality", Doc. B/AIM022, October 1999.

[8] ISO/IEC 23003-1:2007, Information technology – MPEG audio technologies – Part 1: MPEG Surround.

Claims

1. An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal and representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation;

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation.

2. The audio signal decoder according to claim 1, wherein the audio signal decoder is configured to provide the upmix signal representation in dependence on a residual information associated to a subset of audio objects represented by the downmix signal representation,

wherein the object separator is configured to decompose the downmix signal representation to provide the first audio information describing a first set of one or more audio objects of a first audio object type to which residual information is associated, and the second audio information describing a second set of one or more audio objects of a second audio object type, to which no residual information is associated, in dependence on the downmix signal representation and using the residual information.

3. The audio signal decoder (100; 200; 500; 590) according to claim 1 or 2, wherein the object separator is configured to provide the first audio information such that

one or more audio objects of the first audio object type are emphasized over audio objects of the second audio object type in the first audio information, and

wherein the object separator is configured to provide the second audio information such that audio objects of the second audio object type are emphasized over audio objects of the first audio object type in the second audio information.

4. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 3, wherein the audio signal decoder is configured to perform a 2-step processing, such that a processing of the second audio information in the audio signal processor (140; 270; 570; 570a) is performed subsequent to a separation between the first audio information, describing the first set of one or more audio objects of the first audio object type, and the second audio information describing the second set of one or more audio objects of the second audio object type.

5. The audio signal decoder (100; 200; 500; 570) according to one of claims 1 to 4, wherein the audio signal processor is configured to process the second audio information (134; 264; 564; 564a) in dependence on the object-related parametric information (110; 212; 512; 512a) associated with the audio objects of the second audio object type and independent from the object-related parametric information (110; 212; 512; 512a) associated with the audio objects of the first audio object type.

6. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 5, wherein the object separator is configured to obtain the first audio information (132; 262; 562; 562a, $\mathbf{X}_{EAO}$) and the second audio information (134; 264; 564; 564a, $\mathbf{X}_{OBJ}$) using a linear combination of one or more downmix signal channels of the downmix signal representation and one or more residual channels, wherein the object separator is configured to obtain combination parameters for performing the linear combination in dependence on downmix parameters associated with the audio objects of the first audio object type ($m_0 \ldots m_{NEAO-1}$; $n_0 \ldots n_{NEAO-1}$) and in dependence on channel prediction coefficients ($c_{j,0}$, $c_{j,1}$) of the audio objects of the first audio object type.

7. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 6, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\widetilde{D}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein C describes a matrix representing a plurality of channel prediction coefficients, $\widetilde{c}_{j,0}$, $\widetilde{c}_{j,1}$; wherein l_0 and r_0 represent channels of the downmix signal representation;

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

8. The audio signal decoder according to claim 7, wherein the object separator is configured to obtain the inverse downmix matrix $\widetilde{D}^{-1}$ as an inverse of an extended downmix matrix $\widetilde{D}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the object separator is configured to obtain the matrix C as

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type.

9. Audio signal decoder according to claim 8, wherein the object separator is configured to compute the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ and}$$

wherein the object separator is configured to derive constrained prediction coefficients $c_{j,0}$ and $c_{j,1}$ from the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ using a constraining algorithm, or to use the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as the prediction coefficients $c_{j,0}$ and $c_{j,1}$;

wherein energy quantities P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ and $P_{RoCo,j}$ are defined as

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

wherein parameters OLD_L, OLD_R and $IOC_{L,R}$ correspond to audio objects of the second audio object type and are defined according to

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

wherein $d_{0,i}$ and $d_{1,i}$ are downmix values associated with the audio objects of the second audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the second audio object type;

wherein N is a total number of audio objects;

wherein N_{EAO} is a number of audio objects of the first audio object type;

wherein $IOC_{0,1}$ is an inter-object-correlation value associated with a pair of audio objects of the second audio object type;

wherein $e_{i,j}$ and $e_{L,R}$ are covariance values derived from object-level-difference parameters and inter-object-correlation parameters; and

wherein $e_{i,j}$ are associated with a pair of audio objects of the 1st audio object type and $e_{L,R}$ is associated with a pair of audio objects of the second audio object type.

10. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 6, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} \frac{d_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} \frac{d_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C}$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\tilde{\mathbf{D}}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein $\mathbf{C}$ describes a matrix representing a plurality of channel prediction coefficients, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

wherein d_0 represents a channel of the downmix signal representation; and

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

11. The audio signal decoder according to claim 10, wherein the object separator is configured to obtain the inverse downmix matrix $\tilde{\mathbf{D}}^{-1}$ is an inverse of an extended downmix matrix $\tilde{\mathbf{D}}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{c|ccc} 1 & m_0 & \cdots & m_{N_{EAO}-1} \\ \hline m_0 & -1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the object separator is configured to obtain the matrix C as

$$\mathbf{C} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline c_0 & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1} & 0 & \cdots & 1 \end{array}\right);$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type.

12. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 6, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\dfrac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\dfrac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\dfrac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\dfrac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\dfrac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\dfrac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L and OLD_R are common object level difference values associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

13. The audio signal decoder according to one of claims 1 to 5, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}}\left(d_0\right)$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Energy}}\left(d_0\right)$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\mathbf{d}_0$ represents a channel of the downmix signal representation;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \left(\sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right)$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \left(\begin{array}{c} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{array} \right)$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L is a common object level difference value associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

14. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 13, wherein the object separator is configured to apply a rendering matrix to the first audio information (132; 262; 562; 562a) to map object signals of the first audio information onto audio channels of the upmix audio signal representation (120; 220, 222; 562; 562a).

15. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 14, wherein the audio signal processor (140; 270; 570; 570a) is configured to perform a stereo preprocessing of the second audio information (134; 264; 564; 564a) in dependence on a rendering information (M_{ren}), an object-related covariance information (**E**), a downmix information (**D**), to obtain audio channels of the processed version of the second audio information;

16. The audio signal decoder (100; 200; 500; 590) according to claim 15, wherein the audio signal processor (140; 270; 570; 570a) is configured to perform the stereo processing to map an estimated audio object contribution (ED^*JX) of the second audio information (134; 264; 564; 564a) onto a plurality of channels of the upmix audio signal representation in dependence on a rendering information and a covariance information.

17. The audio signal decoder according to claim 15 or claim 16, wherein the audio signal processor is configured to add a decorrelated audio signal contribution (P_2X_d) to the second audio information, or an information derived from the second audio information, in dependence on a render upmix error information (R) and one or more decorrelated-signal-intensity scaling values (w_{d1}, w_{d2}).

18. The audio signal decoder according to one of claims 1 to 14, wherein the audio signal processor (140; 270; 570; 570a) is configured to perform a postprocessing of the second audio information (134; 264; 564; 564a) in dependence on a rendering information (A), an object-related covariance information (E) and a downmix information (D).

19. The audio signal decoder according to claim 18, wherein the audio signal processor is configured to perform a mono-to-binaural processing of the second audio information, to map a single channel of the second audio information onto two channels of the upmix signal representation, taking into consideration a head-related transfer function.

20. The audio signal decoder according to claim 18, wherein the audio signal processor is configured to perform a mono-to-stereo processing of the second audio information, to map a single channel of the second audio information onto two channels of the upmix signal representation.

21. The audio signal decoder according to claim 18, wherein the audio signal processor is configured to perform a stereo-to-binaural processing of the second audio information, to map two channels of the second audio information onto two channels of the upmix signal representation, taking into consideration a head-related transfer function.

22. The audio signal decoder according to claim 18, wherein the audio signal processor is configured to perform a stereo-to-stereo processing of the second audio information, to map two channels of the second audio information onto two channels of the upmix signal representation.

23. The audio signal decoder according to one of claims 1 to 22, wherein the object separator is configured to treat audio objects of the second audio object type, to which no residual information is associated, as a single audio object, and

wherein the audio signal processor (140; 270; 570; 570a) is configured to consider object-specific rendering parameters associated to the audio objects of the second audio object type to adjust contributions of the audio objects of the second audio object type to the upmix signal representation.

24. The audio signal decoder according to one of claims 1 to 23, wherein the object separator is configured to obtain one or two common object level difference values (OLD_L, OLD_R) for a plurality of audio objects of the second audio object type; and

wherein the object separator is configured to use the common object level difference value for a computation of channel prediction coefficients (CPC); and

wherein the object separator is configured to use the channel prediction coefficients to obtain one or two audio channels representing the second audio information.

25. The audio signal decoder according to one of claims 1 to 24, wherein the object separator is configured to obtain one or two common object level difference values (OLD_L, OLD_R) for a plurality of audio objects of the second audio object type, and

wherein the object separator is configured to use the common object level difference value for a computation of entries of an matrix ($\mathbf{M}$); and

wherein the object separator is configured to use the matrix (**M**) to obtain one or more audio channels representing the second audio information.

26. The audio signal decoder according to one of claims 1 to 25, wherein the object separator is configured to selectively obtain a common inter-object correlation value ($IOC_{L,R}$) associated to the audio object of the second audio object type in dependence on the object-related parametric information if it is found that there are two audio objects of the second audio object type, and to set the inter-object correlation value associated to the audio objects of the second audio object type to zero if it is found that there are more or less than two audio objects of the second audio object type; and

wherein the object separator is configured to use the common inter-object correlation value for a computation of entries of an matrix (**M**); and

wherein the object separator is configured to use the common inter-object correlation value associated to the audio objects of the second audio object type to obtain the one or more audio channels representing the second audio information.

27. The audio signal decoder according to one of claims 1 to 26, wherein the audio signal processor is configured to render the second audio information in dependence on the object-related parametric information, to obtain a rendered representation of the audio objects of the second audio object type as the processed version of the second audio information.

28. The audio signal decoder according to one of claims 1 to 27, wherein the object separator is configured to provide the second audio information such that the second audio information describes more than two audio objects of the second audio object type.

29. The audio signal decoder according to claim 28, wherein the object separator is configured to obtain, as the second audio information, a one-channel audio signal representation or a two-channel audio signal representation representing more than two audio objects of the second audio object type.

30. The audio signal decoder according to one of claims 1 to 29, wherein the audio signal processor is configured to receive the second audio information and to process the second audio information in dependence of the object-related

parametric information, taking into consideration object-related parametric information associated with more than two audio objects of the second audio object type.

31. The audio signal decoder according to one of claims 1 to 30, wherein the audio signal decoder is configured to extract a total object number information (bsNumObjects) and a foreground object number information (bsNumGroupsFGO) from a configuration information (SAOCSpecificConfig) of the object-related parametric information, and to determine the number of audio objects of the second audio object type by forming a difference between the total object number information and the foreground object number information.

32. The audio signal decoder according to one of claims 1 to 31, wherein the object separator is configured to use object-related parametric information associated with N_{EAO} audio objects of the first audio object type to obtain, as the first audio information, N_{EAO} audio signals ($\mathbf{X}_{EAO}$) representing the N_{EAO} audio objects of the first audio object type and to obtain, as the second audio information, one or two audio signals ($\mathbf{X}_{OBJ}$) representing the N-N_{EAO} audio objects of the second audio object type, treating the N-N_{EAO} audio objects of the second audio object type as a single one-channel or a two-channel audio object; and

wherein the audio signal processor is configured to individually render the N-N_{EAO} audio objects represented by the one or two audio signals of the second audio information using the object-related parametric information associated with the N-N_{EAO} audio objects of the second audio object type.

33. A method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, the method comprising:

decomposing the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type, and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information; and

processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information; and

combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation.

34. A computer program for performing the method according to claim 33 when the computer program runs on a computer.

Audio Signal Decoder, Method for Decoding an Audio Signal and Computer Program using Cascaded Audio Object Processing Stages

An audio signal decoder for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information comprises an object separator configured to decompose the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type and a second audio information describing a second set of one or more audio objects of a second audio object type, in dependence on the downmix signal representation and using at least a part of the object-related parametric information. The audio signal decoder also comprises an audio signal processor configured to receive the second audio information and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information. The audio signal decoder also comprises an audio signal combiner configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation.

Fig. 1

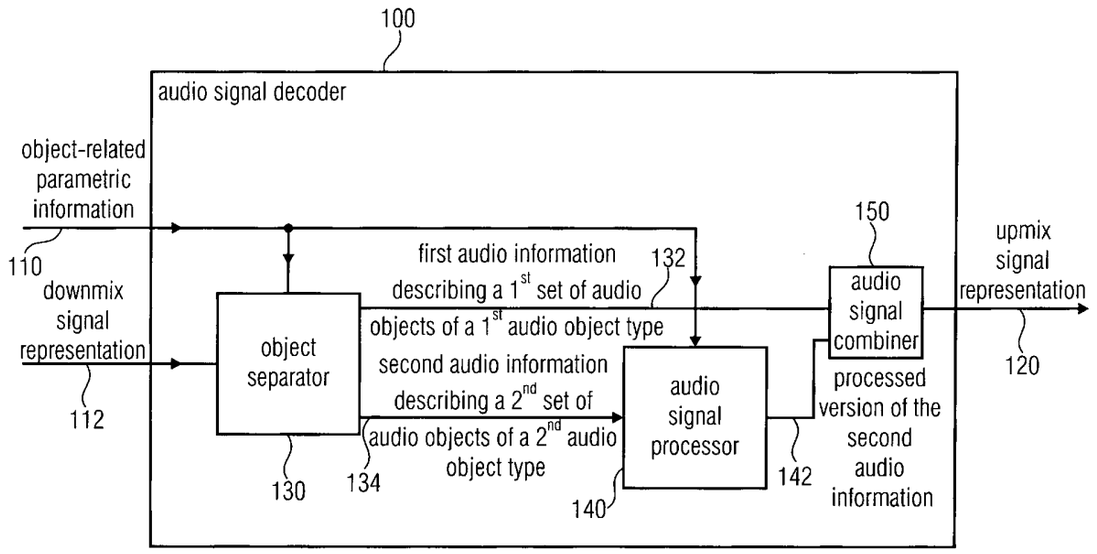
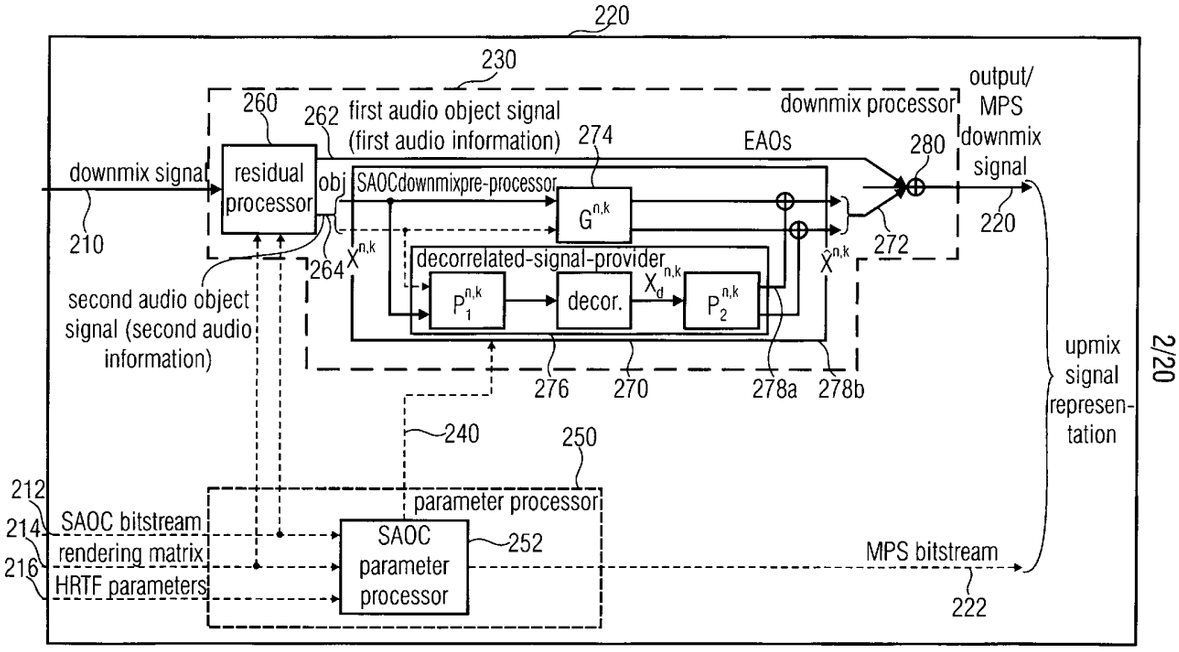
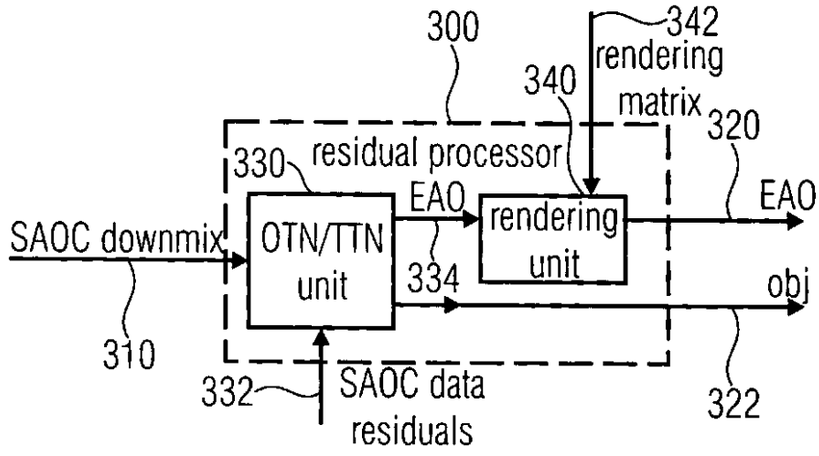

100
audio signal decoder
object-related parametric information
110
downmix signal representation
112
object separator
130
first audio information describing a 1st set of audio objects of a 1st audio object type
132
second audio information describing a 2nd set of audio objects of a 2nd audio object type
134
audio signal processor
140
142
150
audio signal combiner
processed version of the second audio information
upmix signal representation
120

FIG 1



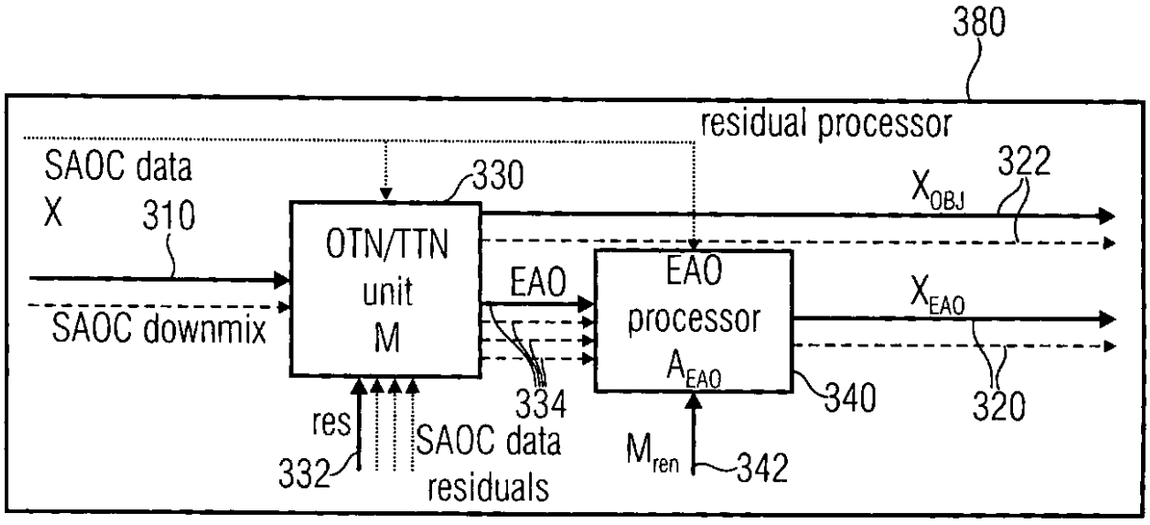
OVERALL STRUCTURE OF THE SAOC TRANSCODER/DECODER ARCHITECTURE

FIG 2



ARCHITECTURE OF THE RESIDUAL PROCESSOR

FIG 3A



ARCHITECTURE OF THE RESIDUAL PROCESSOR

FIG 3B

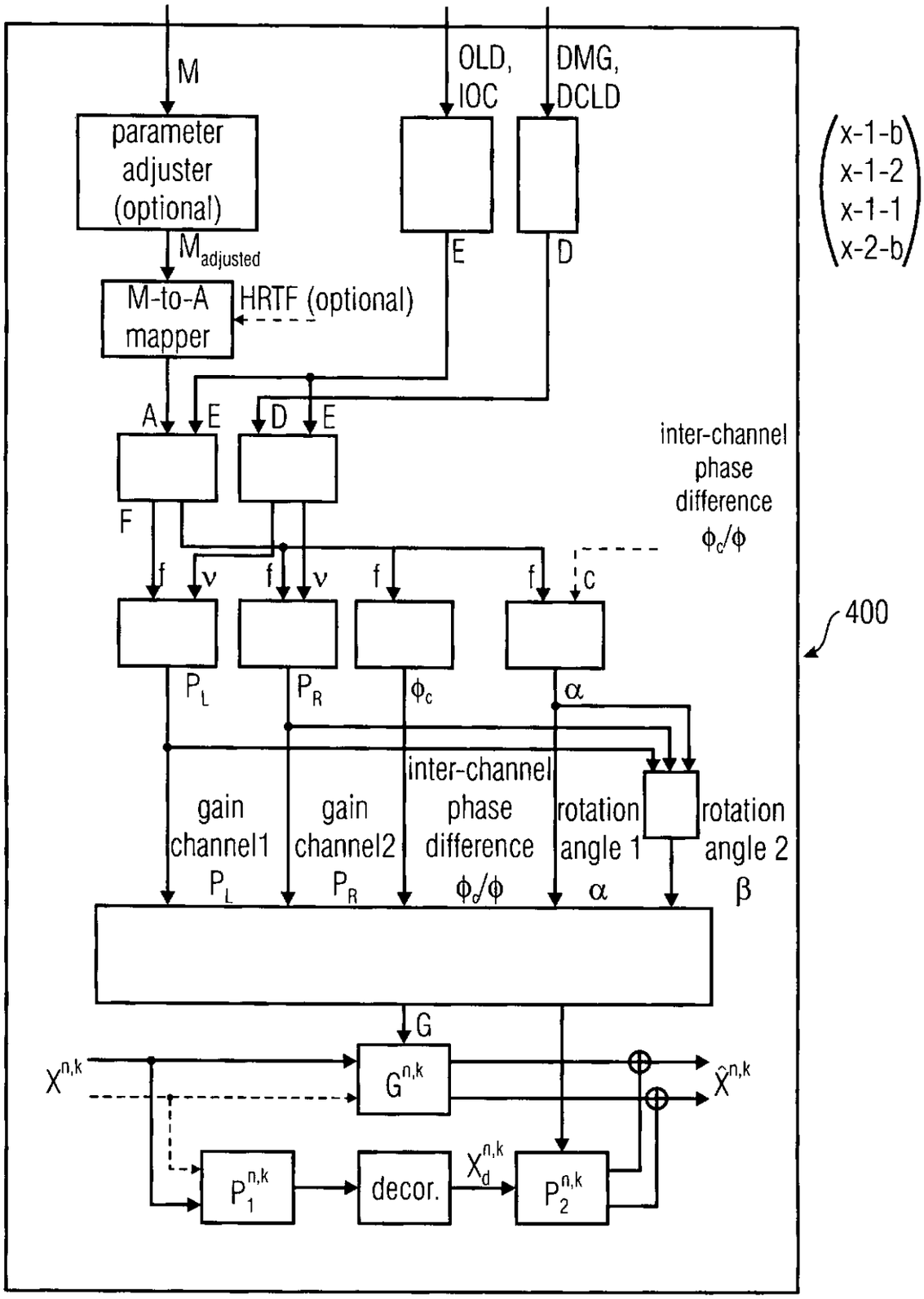

M
OLD, IOC
DMG, DCLD
parameter adjuster (optional)
$M_{adjusted}$
M-to-A mapper
HRTF (optional)
E
D
A
E
D
E
F
inter-channel phase difference ϕ_c/ϕ
f
ν
f
ν
f
f
c
P_L
P_R
ϕ_c
α
gain channel1 P_L
gain channel2 P_R
inter-channel phase difference ϕ_c/ϕ
rotation angle 1 α
rotation angle 2 β
G
$X^{n,k}$
$G^{n,k}$
$\hat{X}^{n,k}$
$P_1^{n,k}$
decor.
$X_d^{n,k}$
$P_2^{n,k}$
(x-1-b
x-1-2
x-1-1
x-2-b)
400

FIG 4A

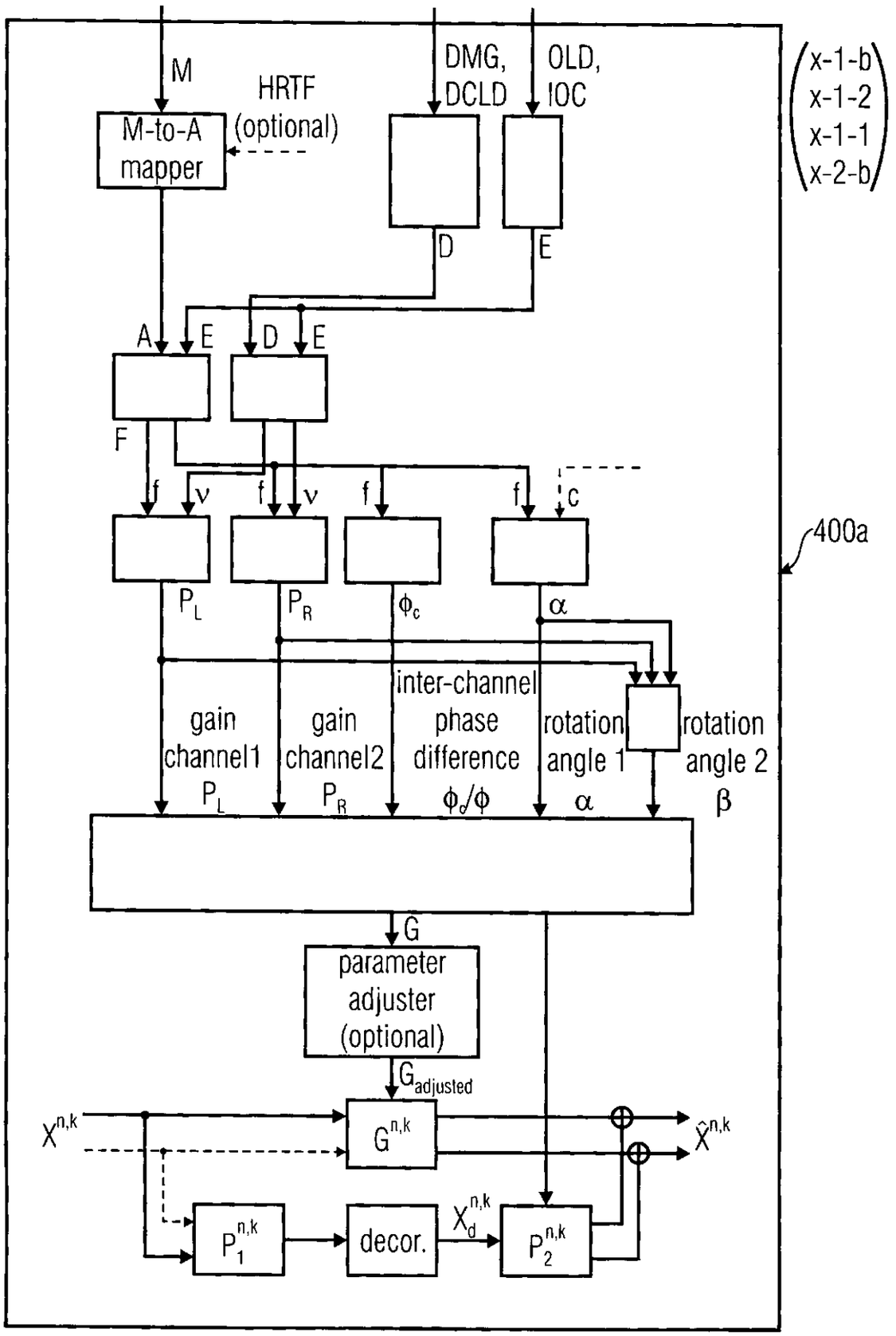

M
HRTF
(optional)
M-to-A
mapper
DMG,
DCLD
OLD,
IOC
x-1-b
x-1-2
x-1-1
x-2-b
D
E
A
E
D
E
F
f
ν
f
ν
f
f
c
400a
P_L
P_R
ϕ_c
α
gain
channel1
P_L
gain
channel2
P_R
inter-channel
phase
difference
ϕ_c/ϕ
rotation
angle 1
α
rotation
angle 2
β
G
parameter
adjuster
(optional)
$G_{adjusted}$
$X^{n,k}$
$G^{n,k}$
$\hat{X}^{n,k}$
$P_1^{n,k}$
decor.
$X_d^{n,k}$
$P_2^{n,k}$

FIG 4B

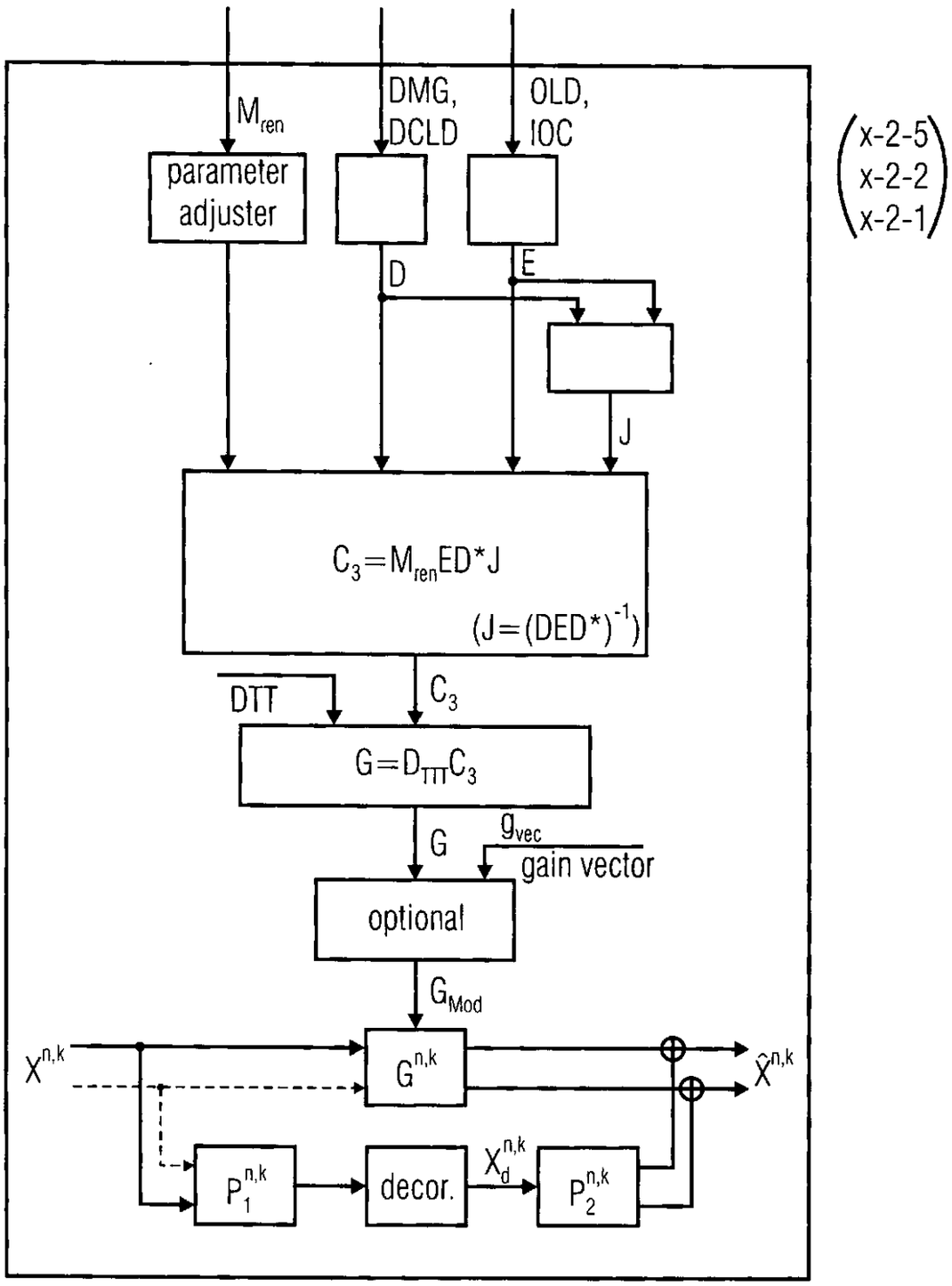

M_ren
DMG, DCLD
OLD, IOC
parameter adjuster
D
E
J
C3=MrenED*J
(J=(DED*)-1)
DTT
C3
G=DTTC3
G
gvec
gain vector
optional
GMod
Xn,k
Gn,k
X̂n,k
P1n,k
decor.
Xdn,k
P2n,k
(x-2-5
x-2-2
x-2-1)

FIG 4C

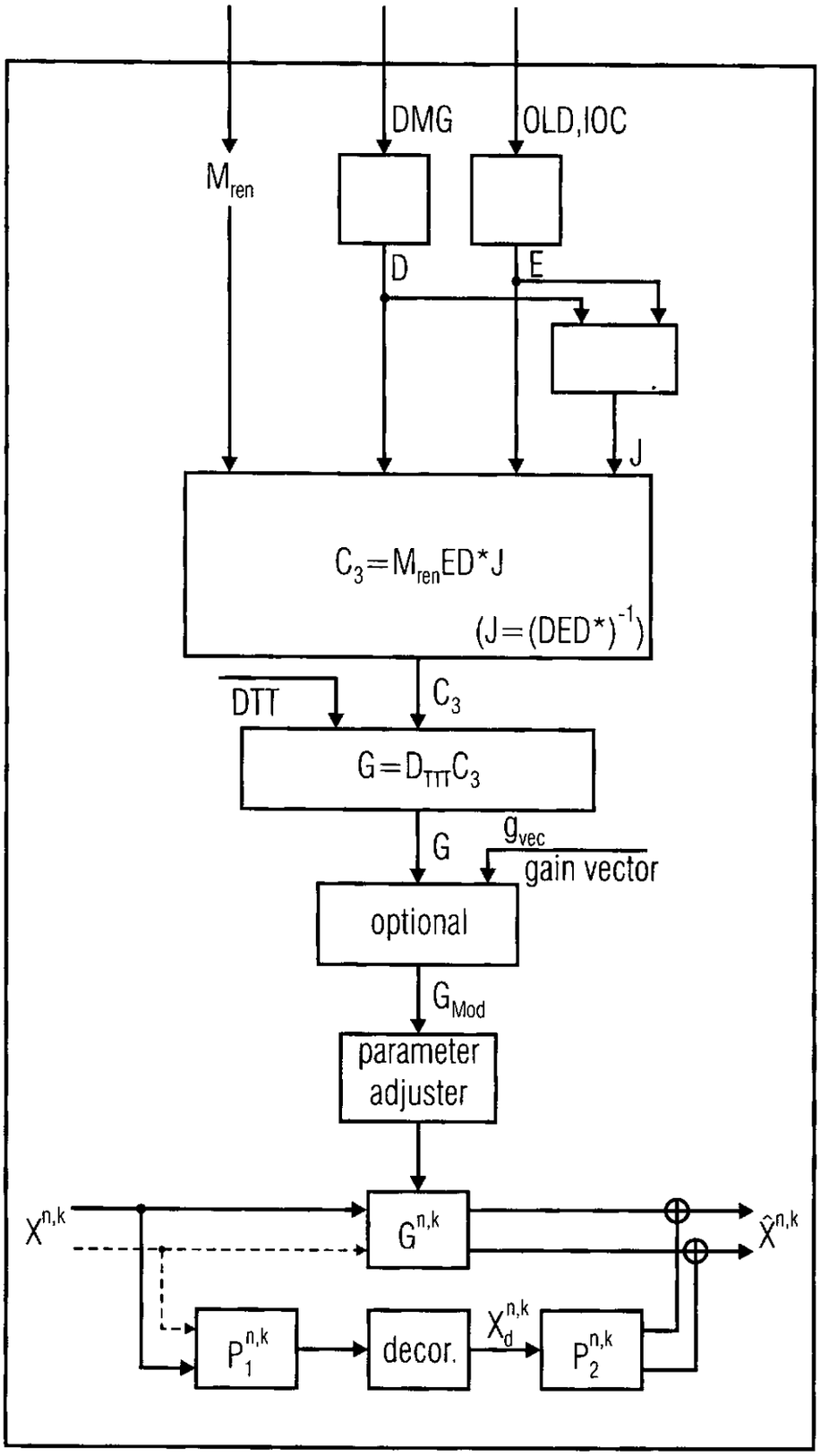

DMG
OLD,IOC
(x-2-2
x-2-1)
M_{ren}
D
E
J
$C_3 = M_{ren}ED^*J$
$(J=(DED^*)^{-1})$
DTT
C_3
$G=D_{TTT}C_3$
G
g_{vec}
gain vector
optional
G_{Mod}
parameter adjuster
$X^{n,k}$
$G^{n,k}$
$\hat{X}^{n,k}$
$P_1^{n,k}$
decor.
$X_d^{n,k}$
$P_2^{n,k}$

FIG 4D

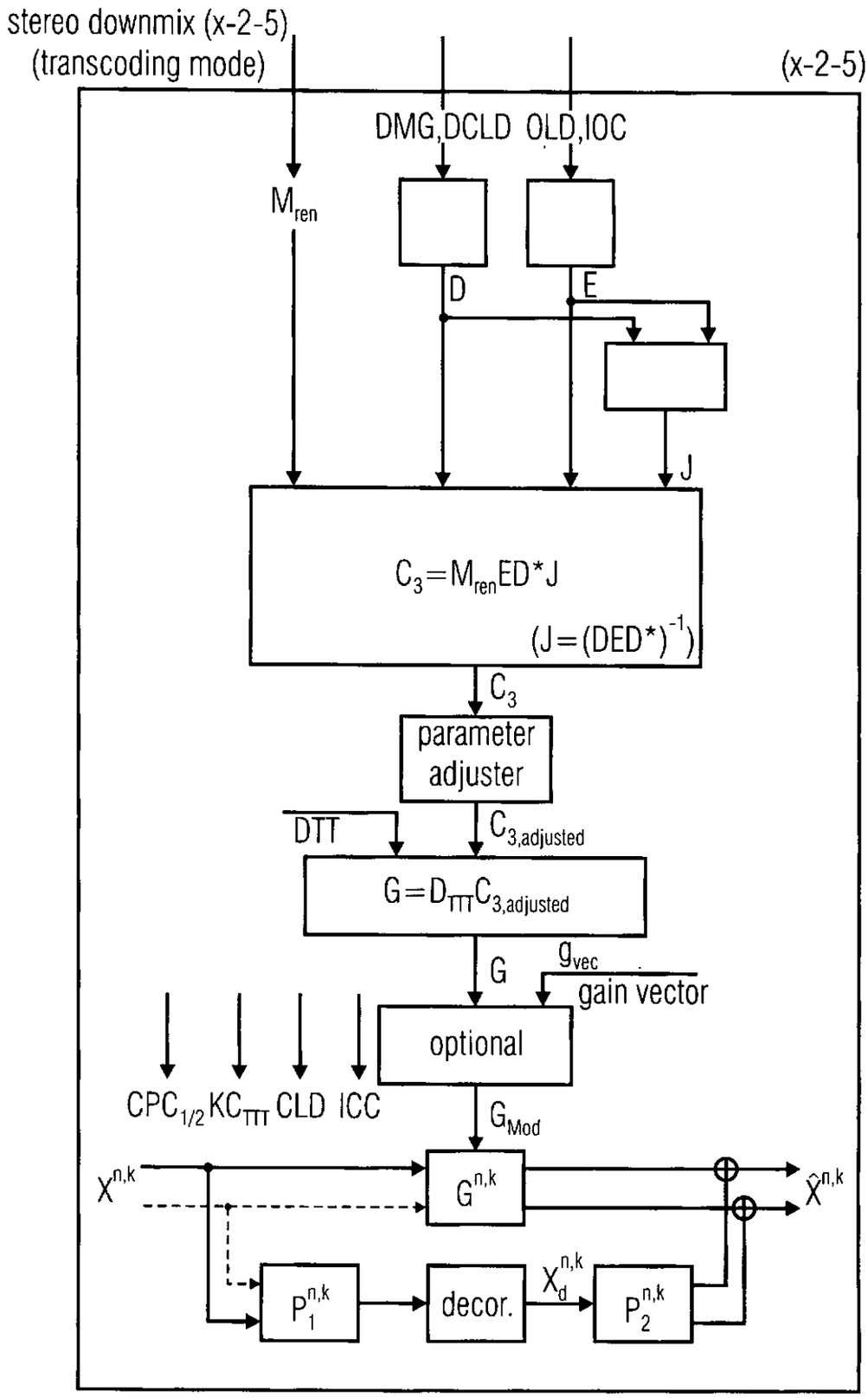

stereo downmix (x-2-5)
(transcoding mode)
(x-2-5)
DMG,DCLD OLD,IOC
M_{ren}
D
E
J
$C_3 = M_{ren} E D^* J$
$(J = (D E D^*)^{-1})$
C_3
parameter adjuster
DTT
$C_{3,adjusted}$
$G = D_{TTT} C_{3,adjusted}$
G
g_{vec}
gain vector
optional
$CPC_{1/2}$ KC_{TTT} CLD ICC
G_{Mod}
$X^{n,k}$
$G^{n,k}$
$\hat{X}^{n,k}$
$P_1^{n,k}$
decor.
$X_d^{n,k}$
$P_2^{n,k}$

FIG 4E

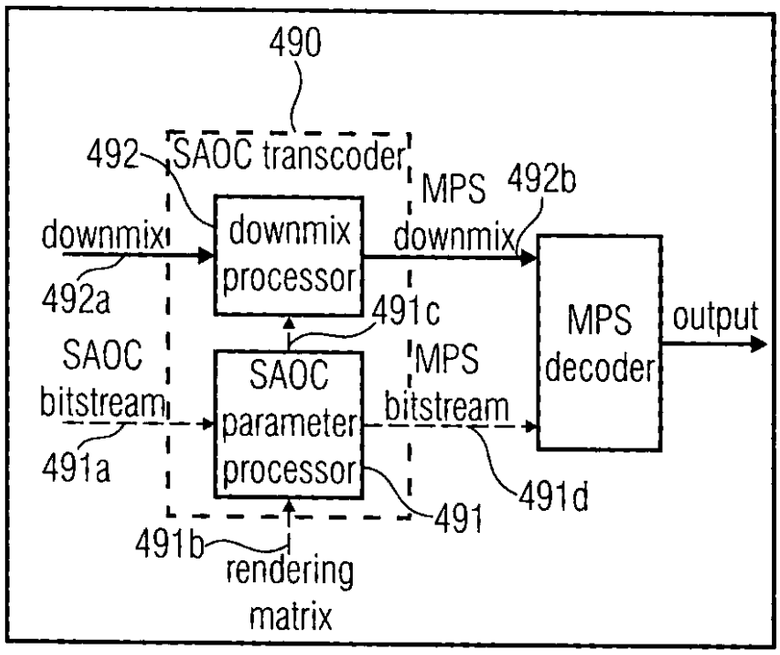

490
492
SAOC transcoder
MPS
492b
downmix
downmix
processor
downmix
492a
491c
MPS
decoder
output
SAOC
bitstream
SAOC
parameter
processor
MPS
bitstream
491a
491d
491
491b
rendering
matrix

FIG 4F

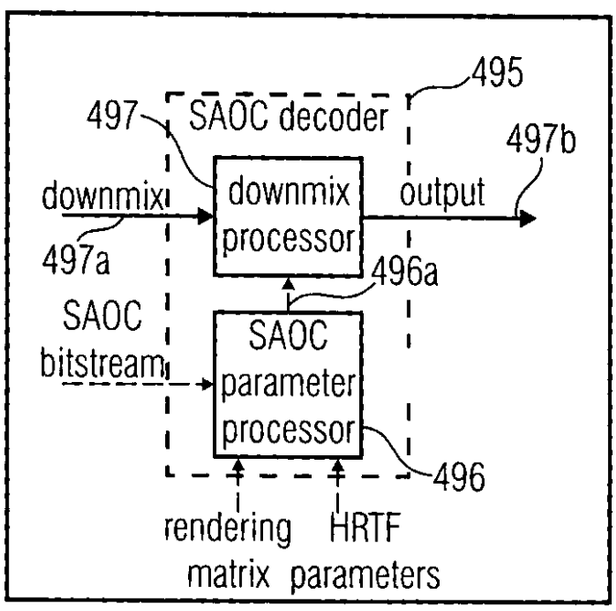
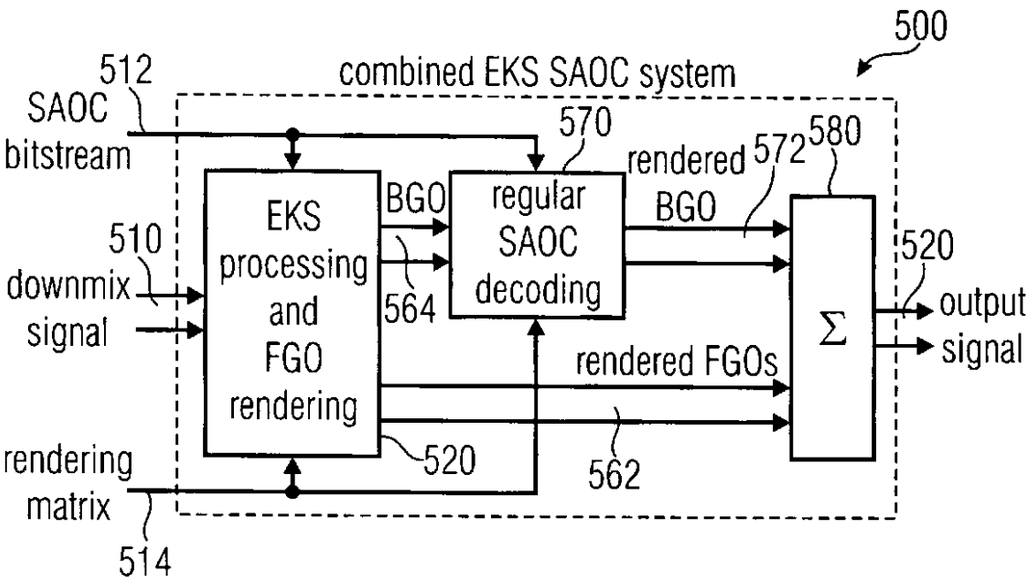
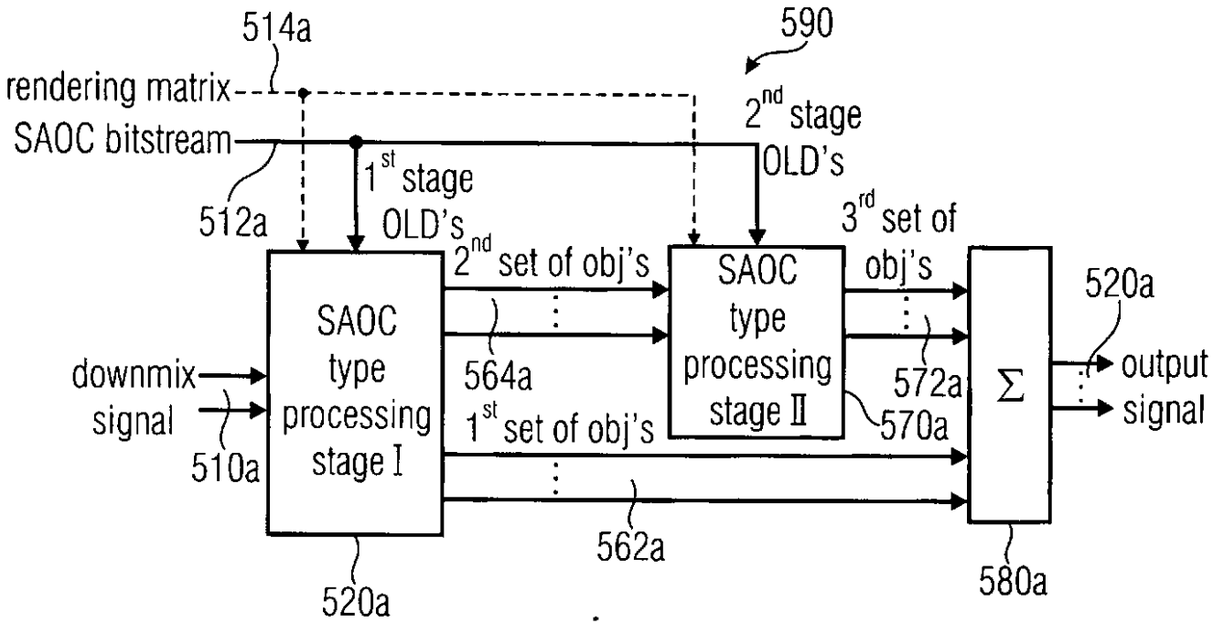

495
497
SAOC decoder
497b
downmix
downmix
processor
output
497a
496a
SAOC
bitstream
SAOC
parameter
processor
496
rendering
matrix
HRTF
parameters

FIG 4G



BASIC STRUCTURE OF
THE COMBINED EKS SAOC SYSTEM

FIG 5A



GENERALIZED STRUCTURE OF
THE COMBINED EKS SAOC SYSTEM

FIG 5B

Coder name	Description
hidden reference	ideally rendered audio scene (hidden reference)
lower anchor	lower anchor @ 3.5kHz
regular SAOC	regular SAOC decoding mode
current EKS	current SAOC EKS mode
combined system	combined (EKS+SAOC) system (proposed system)

FIG 6A

System	Bitstream	Karaoke rendering	Classic rendering
combined system	SAOC A + res. A data	X	X
regular SAOC	SAOC A data	(X)	X
current EKS	SAOC B + res. B data	X	-

FIG 6B

Item	Type	Rendering matrix	Downmix matrix
k-pop 1	Karaoke	[0.6 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.038 0.0 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.038]	[0.0 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.6 0.6 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.6]
	Classic	[0.6 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.6 0.0 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.6]	
rock	Karaoke	[1.0 0.0 0.707 0.707 0.000 0.038 0.0 1.0 0.707 0.000 0.707 0.038]	[1.0 0.0 0.707 0.707 0.000 0.707 0.0 1.0 0.707 0.000 0.707 0.707]
	Classic	[1.0 0.0 0.6 0.0 0.707 0.000 0.4 0.0 1.0 0.4 0.0 0.000 0.707 0.6]	

FIG 6C

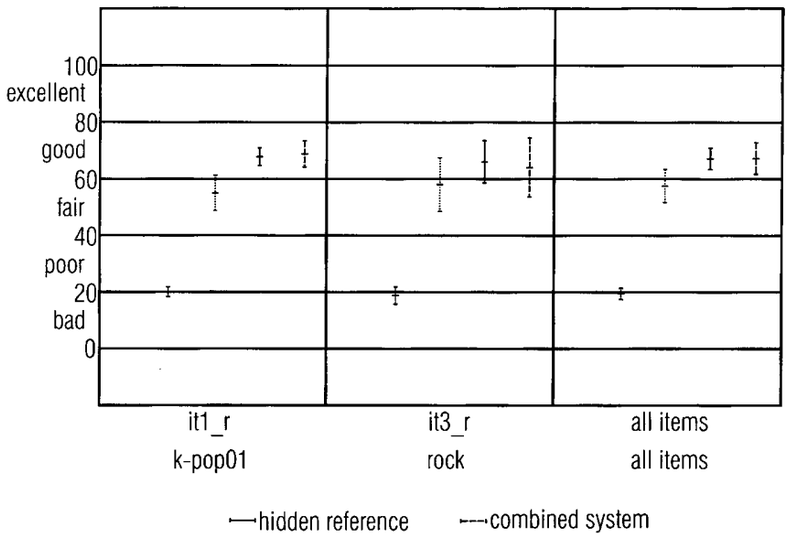

100
excellent
80
good
60
fair
40
poor
20
bad
0
it1_r
k-pop01
it3_r
rock
all items
all items
hidden reference
combined system

FIG 6D

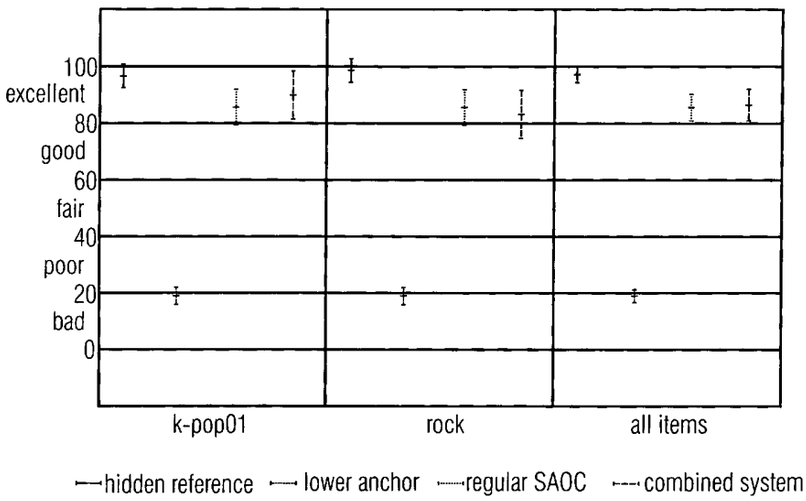

100
excellent
80
good
60
fair
40
poor
20
bad
0
k-pop01
rock
all items
hidden reference
lower anchor
regular SAOC
combined system

FIG 6E

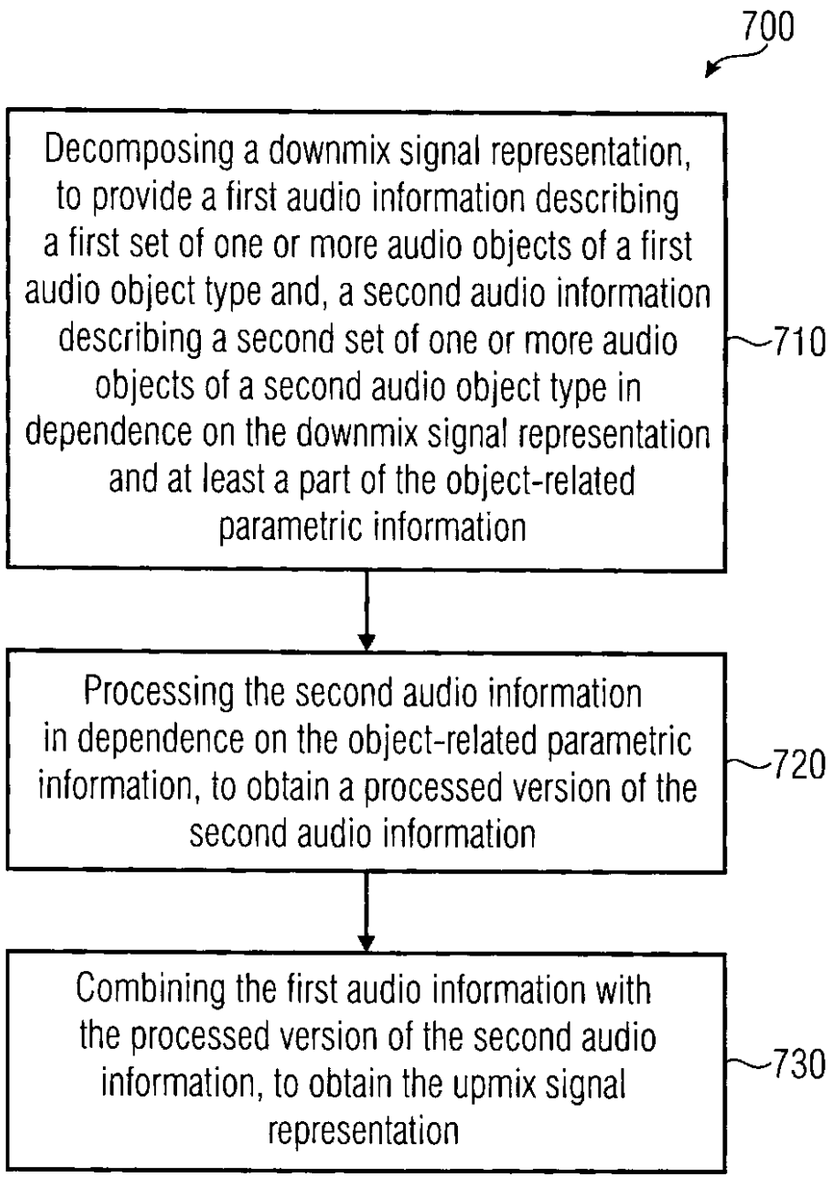
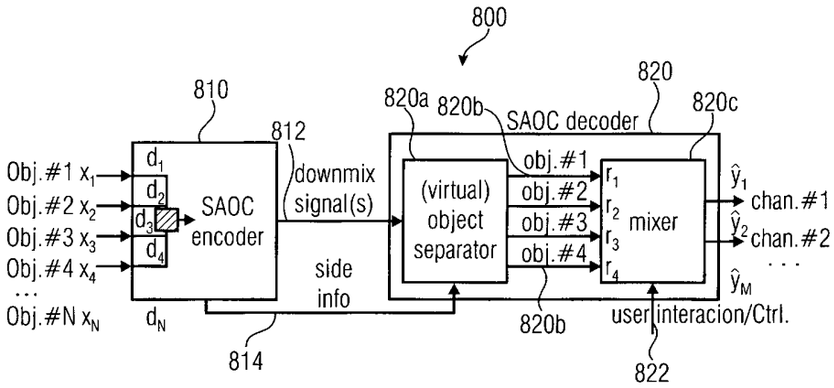
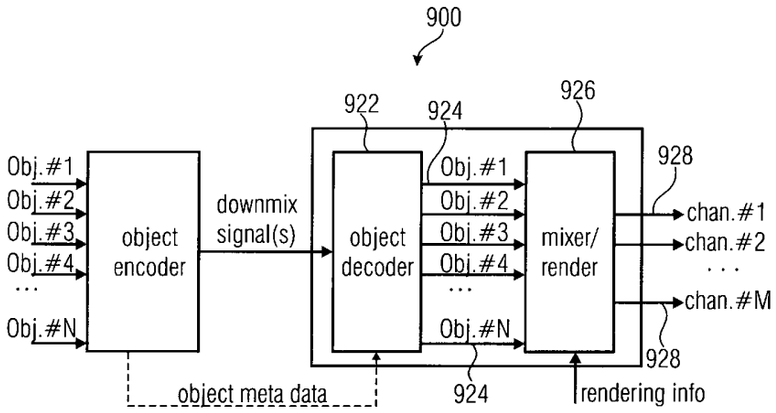

700
Decomposing a downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type and, a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and at least a part of the object-related parametric information
710
Processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information
720
Combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation
730

FIG 7



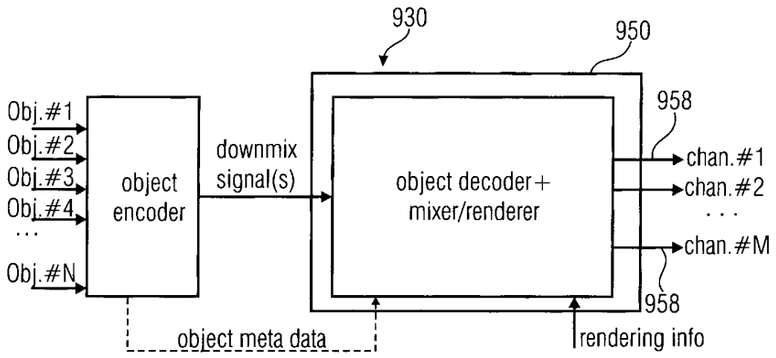
MPEG SAOC SYSTEM OVERVIEW

FIG 8



SEPARATE DECODER AND MIXER

FIG 9A



INTEGRATED DECODER AND MIXER

FIG 9B

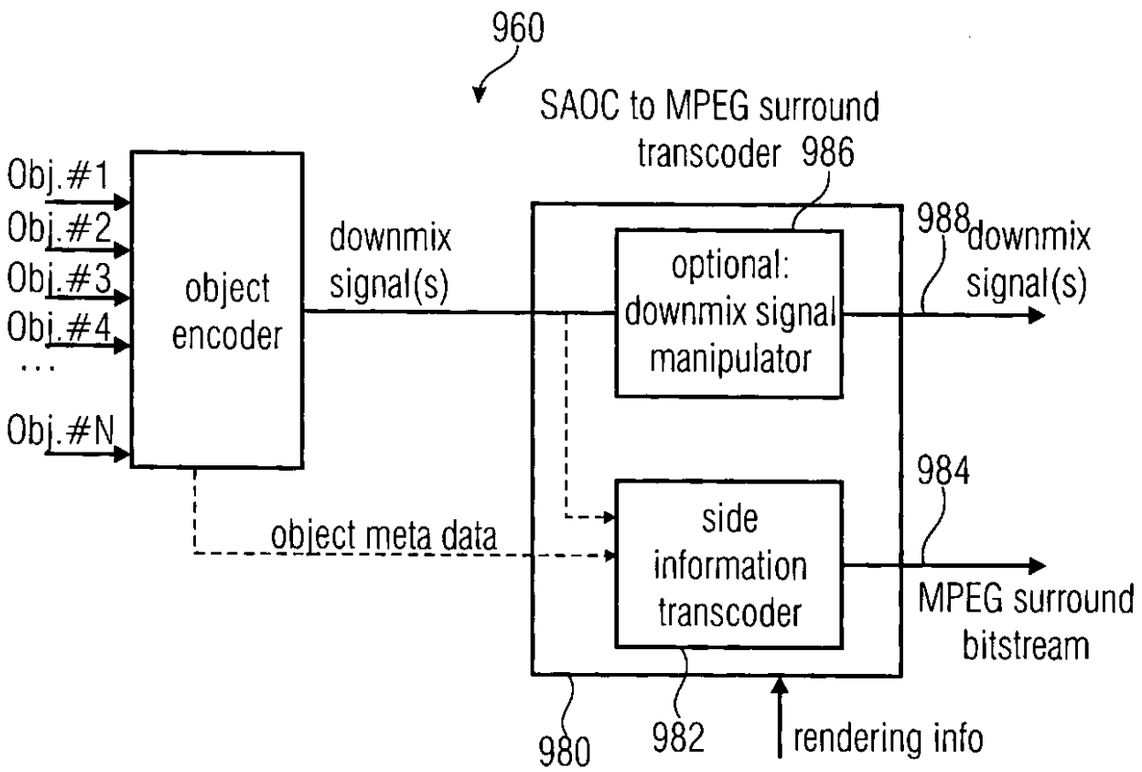

960
SAOC to MPEG surround transcoder 986
Obj.#1
Obj.#2
Obj.#3
Obj.#4
...
Obj.#N
object encoder
downmix signal(s)
optional: downmix signal manipulator
988
downmix signal(s)
object meta data
side information transcoder
984
MPEG surround bitstream
980
982
rendering info

FIG 9C

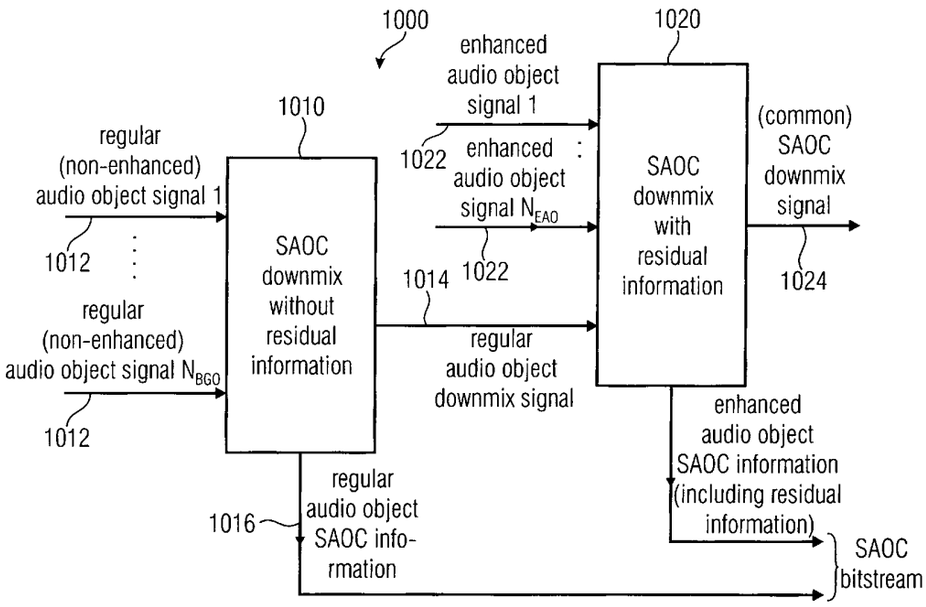

1000
1020
enhanced audio object signal 1
1010
regular (non-enhanced) audio object signal 1
1012
1022
enhanced audio object signal N_{EAO}
SAOC downmix with residual information
(common) SAOC downmix signal
1024
1022
1014
SAOC downmix without residual information
regular (non-enhanced) audio object signal N_{BGO}
1012
regular audio object downmix signal
enhanced audio object SAOC information (including residual information)
1016
regular audio object SAOC info-rmation
SAOC bitstream

FIG 10

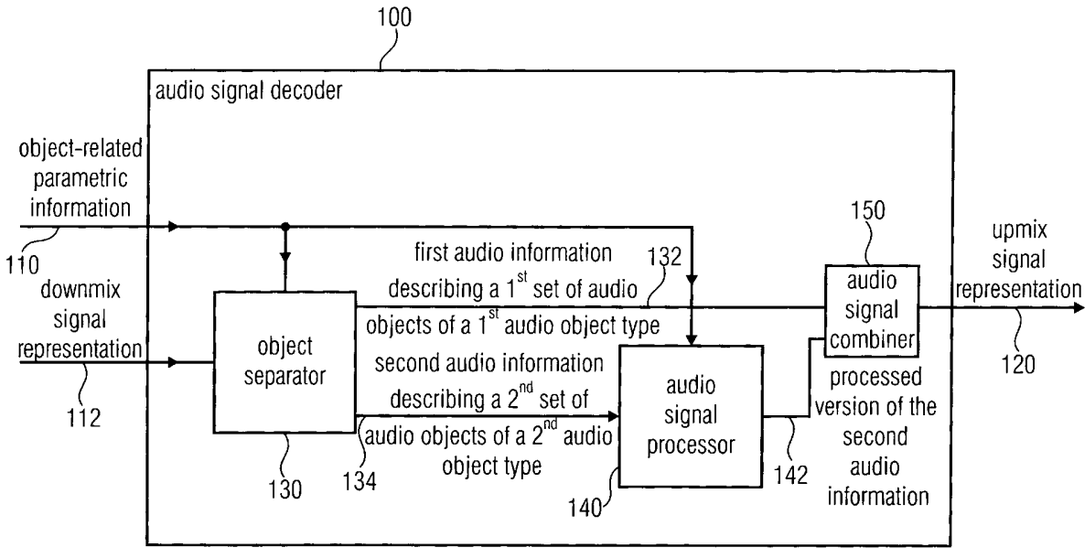


FIG 1

FIGURE FOR ABSTRACT

235

PATENT COOPERATION TREATY

EINGEGANGEN
- 6. SEP. 2011

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To:

SCHOPPE, ZIMMERMANN, STÖCKELER,
ZINKLER & PARTNER
Postfach 246
82043 Pullach
ALLEMAGNE

PCT

NOTIFICATION OF TRANSMITTAL OF
THE INTERNATIONAL PRELIMINARY
REPORT ON PATENTABILITY

(PCT Rule 71.1)

Date of mailing *(day/month/year)*	05.09.2011
Applicant's or agent's file reference FH100602PCT	**IMPORTANT NOTIFICATION**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/EP2010/058906	23.06.2010	24.06.2009

Applicant
FRAUNHOFER GESELLSCHAFT ZUR FÖRDERUNG

1. The applicant is hereby notified that this International Preliminary Examining Authority transmits herewith the international preliminary report on patentability and its annexes, if any, established on the international application.

2. A copy of the report and its annexes, if any, is being transmitted to the International Bureau for communication to all the elected Offices.

3. Where required by any of the elected Offices, the International Bureau will prepare an English translation of the report (but not of any annexes) and will transmit such translation to those Offices.

4. **REMINDER**

The applicant must enter the national phase before each elected Office by performing certain acts (filing translations and paying national fees) within 30 months from the priority date (or later in some Offices) (Article 39(1)) (see also the reminder sent by the International Bureau with Form PCT/IB/301).

Where a translation of the international application must be furnished to an elected Office, that translation must contain a translation of any annexes to the international preliminary report on patentability. It is the applicant's responsibility to prepare and furnish such translation directly to each elected Office concerned.

For further details on the applicable time limits and requirements of the elected Offices, see Volume II of the PCT Applicant's Guide.

The applicant's attention is drawn to Article 33(5), which provides that the criteria of novelty, inventive step and industrial applicability described in Article 33(2) to (4) merely serve the purposes of international preliminary examination and that "any Contracting State may apply additional or different criteria for the purposes of deciding whether, in that State, the claimed inventions is patentable or not" (see also Article 27(5)). Such additional criteria may relate, for example, to exemptions from patentability, requirements for enabling disclosure, clarity and support for the claims.

FRISTENDE
VOR
FRIS
ERLEDIGT

Name and mailing address of the international preliminary examining authority:

European Patent Office
D-80298 Munich
Tel. +49 89 2399 - 0
Fax: +49 89 2399 - 4465

Authorized Officer
Schalinatus, Dana
Tel. +49 89 2399-8242

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

(Chapter II of the Patent Cooperation Treaty)

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference FH100602PCT	**FOR FURTHER ACTION**	See Form PCT/IPEA/416
International application No. PCT/EP2010/058906	International filing date *(day/month/year)* 23.06.2010	Priority date *(day/month/year)* 24.06.2009
International Patent Classification (IPC) or national classification and IPC INV. G10L19/00		
Applicant FRAUNHOFER GESELLSCHAFT ZUR FÖRDERUNG		

1. This report is the international preliminary examination report, established by this International Preliminary Examining Authority under Article 35 and transmitted to the applicant according to Article 36.
2. This REPORT consists of a total of 8 sheets, including this cover sheet.
3. This report is also accompanied by ANNEXES, comprising:
   a. ☒ *(sent to the applicant and to the International Bureau)* a total of 32 sheets, as follows:
      ☒ sheets of the description, claims and/or drawings which have been amended and/or sheets containing rectifications authorized by this Authority, unless those sheets were superseded or cancelled, and any accompanying letters (see Rules 46.5, 66.8, 70.16, 91.2, and Section 607 of the Administrative Instructions).
      ☐ sheets containing rectifications, where the decision was made by this Authority not to take them into account because they were not authorized by or notified to this Authority at the time when this Authority began to draw up this report, and any accompanying letters (Rules 66.4bis, 70.2(e), 70.16 and 91.2).
      ☐ superseded sheets and any accompanying letters, where this Authority either considers that the superseding sheets contain an amendment that goes beyond the disclosure in the international application as filed, or the superseding sheets were not accompanied by a letter indicating the basis for the amendments in the application as filed, as indicated in item 4 of Box No. I and the Supplemental Box (see Rule 70.16(b)).
   b. ☐ *(sent to the International Bureau only)* a total of (indicate type and number of electronic carrier(s)) , containing a sequence listing, in electronic form only, as indicated in the Supplemental Box Relating to Sequence Listing (see paragraph 3*bis* of Annex C of the Administrative Instructions).

4. This report contains indications relating to the following items:
   - ☒ Box No. I Basis of the report
   - ☐ Box No. II Priority
   - ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - ☒ Box No. IV Lack of unity of invention
   - ☒ Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
   - ☐ Box No. VI Certain documents cited
   - ☐ Box No. VII Certain defects in the international application
   - ☒ Box No. VIII Certain observations on the international application

Date of submission of the demand	Date of completion of this report
26.04.2011	05.09.2011
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Fax: +49 89 2399 - 4465	Authorized officer Tilp, Jan Telephone No. +49 89 2399-4753

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/EP2010/058906

Box No. I Basis of the report

1. With regard to the **language**, this report is based on

   ☒ the international application in the language in which it was filed

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of:
   - ☐ international search (under Rules 12.3(a) and 23.1(b))
   - ☐ publication of the international application (under Rule 12.4(a))
   - ☐ international preliminary examination (under Rules 55.2(a) and/or 55.3(a) and (b))

2. With regard to the **elements*** of the international application, this report is based on *(replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report)*:

Description, Pages

1-69	as originally filed	

Claims, Numbers

1-37	filed with the letter of	24-06-2011

Drawings, Sheets

1/20-20/20	as originally filed	

☐ a sequence listing - see Supplemental Box Relating to Sequence Listing.

3. ☐ The amendments have resulted in the cancellation of:
   - ☐ the description, pages
   - ☐ the claims, Nos.
   - ☐ the drawings, sheets/figs
   - ☐ the sequence listing *(specify)*:
   - ☐ any table(s) related to sequence listing *(specify)*:

4. ☐ This report has been established as if (some of) the amendments annexed to this report and listed below had not been made, since either they are considered to go beyond the disclosure as filed, or they were not accompanied by a letter indicating the basis for the amendments in the application as filed, as indicated in the Supplemental Box (Rules 70.2(c) and (c-bis)):
   - ☐ the description, pages
   - ☐ the claims, Nos.
   - ☐ the drawings, sheets/figs
   - ☐ the sequence listing *(specify)*:

5. ☐ This report has been established:
   - ☐ taking into account the rectification of an obvious mistake authorized by or notified to this Authority under Rule 91 (Rules 66.1(d-bis) and 70.2(e)).
   - ☐ without taking into account the rectification of an obvious mistake authorized by or notified to this Authority under Rule 91(Rules 66.4bis and 70.2(e)).

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/EP2010/058906

6. ☐ Supplementary international search report(s) from Authority(ies) has/have been received and taken into account in establishing this report (Rule 45bis.8(b) and (c)).

Box No. IV Lack of unity of invention

1. ☒ In response to the invitation to restrict or pay additional fees, the applicant has, within the applicable time limit:

   ☐ restricted the claims.

   ☒ paid additional fees.

   ☐ paid additional fees under protest and, where applicable, the protest fee.

   ☐ paid additional fees under protest but the applicable protest fee was not paid.

   ☐ neither restricted the claims nor paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied with and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rules 13.1, 13.2 and 13.3 is:

   ☐ complied with.

   ☒ not complied with for the following reasons:

   see separate sheet

4. Consequently, this report has been established in respect of the following parts of the international application:

   ☒ all parts.

   ☐ the parts relating to claims Nos. .

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/EP2010/058906

Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-37
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-37
	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	1-37
	No:	Claims	

2. Citations and explanations (Rule 70.7):

see separate sheet

Box No. VIII Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

see separate sheet

Re Item IV

Lack of unity of invention

1 The application does not meet the requirements of unity of invention as formulated in Rule 13 PCT, the reasons therefor being as follows:

Present independent **claims 31-33** - and thus corresponding present independent **claims 34-37** - appear to meet the requirements of Article 33(2) and (3) PCT. In particular, the special technical features, which represent the contribution over D1 and the other prior art, are claimed in i) claim 31, page 17, line 9 - page 18, line 25, ii) claim 32, page 19, line 19 - page 20, line 16, and iii) claim 33, page 21, line 14 - page 22, line 15.

Said special technical features represent alternative implementation variants for solving the technical problem of efficiently separating the audio objects received by a spatial audio object decoder.

Contrary to the dispositions of Rule 13.2 PCT, neither the above-mentioned nor any other corresponding special technical features are present in independent **claims 1, 29 and 30** (which have not been the subject of the international search) newly-filed with the demand, so that the technical relationship between the subject-matter of claims 31-37, on the one hand, and claims 1, 29 and 30, on the other hand, is lacking.

Summarizing the above, the present application relates to the following two separate inventions which are not so linked as to form a single general inventive concept (Rule 13.1 PCT):

Invention 1

claims 31-37: efficiently separating the audio objects received by a spatial audio object decoder

Invention 2

claims 1-30: providing an audio decoder with enhanced object-individual rendering capability

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1 Reference is made to the following document:

D1 ENGDEGARD J ET AL: "Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding", 124TH AES CONVENTION, AUDIO ENGINEERING SOCIETY, PAPER 7377, 17 May 2008 (2008-05-17) , pages 1-15, XP002541458

2 **Invention 1 - claims 31-37:**

Present independent **claims 31-33** - and thus corresponding present independent **claims 34-37** - appear to meet the requirements of Article 33(2) and (3) PCT (for the reasoning, see section 1 under Re Item IV above).

3 **Invention 2 - claims 1-30:**

3.1 **Claim 1** meets the requirements of Article 33(2) and (3) PCT for the following reasons:

3.1.1 Document D1, which is considered to represent the most relevant state of the art, discloses (see corresponding passages cited in the International Search Report) - using the present wording of independent claim 1:

3.2 "An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal and representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information,

wherein the second audio information is an audio information describing the audio objects of the second audio object type in a combined manner;

an audio signal processor configured to receive the second audio information (134;264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the audio signal decoder is configured to provide the upmix signal representation in dependence on a residual information associated to a subset of audio objects represented by the downmix signal representation,

wherein the object separator is configured to decompose the downmix signal representation to provide the first audio information describing a first set of one or more audio objects of a first audio object type to which residual information is associated, and the second audio information describing a second set of one or more audio objects of a second audio object type, to which no residual information is associated, in dependence on the downmix signal representation and using the residual information; and

wherein the residual information describes a residual distortion, which is expected to remain if an audio object of the first audio object type is isolated merely using the object-related parametric information."

3.2.1 From this, the subject-matter of independent claim 1 differs in that "the audio signal processor is configured to process the second audio information, to perform an object-individual processing of the audio objects of the second audio object type, taking into consideration object-related parametric information associated with more than two audio objects of the second audio object type" (see claim 1, page 2, lines 1-6).

The subject-matter of claim 1 is therefore novel (Article 33(2) PCT).

3.2.2 The problem to be solved by the present invention may be regarded as providing an audio decoder with enhanced rendering capability for audio information describing audio objects for which no object-individual information (such as object-individual residual distortion information determined in an encoder) is known.

The solution to this problem proposed in claim 1 of the present application is considered as involving an inventive step (Article 33(3) PCT) even when the other prior art is taken into consideration.

3.3 **Claims 29 and 30** claim a method and a computer program which respectively correspond to the decoder of claim 1. Consequently, claims 29 and 30 are as well considered novel and inventive (Articles 33(2) and 33(3) PCT).

Re Item VIII

Certain observations on the international application

1 The application does not meet the requirements of Article 6 PCT, because present dependent **claim 2** is not clear.

In particular, the term "emphasized" claimed in claim 2, lines 14 and 18, is neither per se technical, nor does it define with respect to what type of technical quantity (e.g. object amplitude, object power, ...) it is to be understood.

Moreover, the reason for the first audio information comprising audio objects of the second audio object type and the second audio information comprising audio objects of the first audio object type cannot clearly be derived, thus rendering claim 2 unclear when compared to claim 1, which, in contrast to claim 2, exclusively claims the "first audio information describing [...] audio objects of a first audio object type, and a second audio information describing [...] audio objects of a second audio object type".

Finally, claim 2 attempts to define the subject-matter in terms of the result to be achieved (i.e. "audio objects of the first audio object type are emphasized over audio objects of the second audio object type" and vice versa), which merely amounts to a statement of the underlying problem, without providing the technical features necessary for achieving this result.

SCHOPPE · ZIMMERMANN · STÖCKELER ZINKLER & PARTNER

Patentanwälte European Patent Attorneys European Trademark Attorneys

Patentanwälte :: Postfach 246 :: 82043 Pullach bei München

Fritz Schoppe · Dipl.-Ing.
Tankred Zimmermann · Dipl.-Ing.
Ferdinand Stöckeler · Dipl.-Ing.
Franz Zinkler Dipl.-Ing.
Markus Schenk Dipl.-Phys.
Günter Hersina Dipl.-Ing.
Markus Burger Dipl.-Ing.

European Patent Office

80298 Munich

EPO - Munich 62
24. Juni 2011

June 24, 2011

International Patent application PCT/EP2010/058906
Applicant: Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.
Our Ref.: FH100602PCT / gm / es / kk

In response to the invitation to restrict or pay additional fees dated May 24, 2011, received May 25, 2011:

1. Payment of Additional Examination Fee:

The payment of one additional examination fee is effected with the enclosed voucher.

2. Amended Claims:

It is respectfully requested to consider the enclosed amended claims 1 to 37, which replace the pending claims, when examining the present application.

3. Original Disclosure:

Independent claim 1 has been clarified in that the term "downmix signal and representation" has been replaced by the term "downmix signal representation". Regarding this clarification, reference is made, for example, to Fig. 1 where the

Zur Kasse €1760,-(A)

Telefon/Telephone +49(0)89/79 04 45-0
Telefax/Facsimile +49(0)89/7 90 22 15
Telefax/Facsimile +49(0)89/74 99 69 77
mail@schoppe-zimmermann.com
www.schoppe-zimmermann.com

Postanschrift/Mail address:
Postfach/P.O. Box 246, 82043 Pullach bei München

Kanzleianschrift/Office address:
Hermann-Roth-Weg 1, 82049 Pullach bei München

Partnerschaftsgesellschaft,
eingetragen in das PR unter Nr. PR 974

USt-Nr. 148/233/00493
VAT Registration No. DE 24-06-2011

downmix signal representation 112 is shown. Also, reference is made to the method claims.

In dependent claim 9, the definition "wherein d_0 represents a channel of the downmix signal representation" has been deleted. It is clear from the dependent claim 9 that d_0 is a representation of a single SAOC downmix signal (cf. pending claim 9, last paragraph). Moreover, the equations for $\mathbf{X}_{OBJ}$ and $\mathbf{X}_{EAO}$ have been clarified in that the brackets around d_0 are removed. The fact that there is a multiplication between the matrices $\mathbf{M}_{OBJ}^{ENERGY}$ or $\mathbf{M}_{EAO}^{ENERGY}$ and the scalar d_0 is clear from the description of the present application, in particular from the definition in the last sentence of claim 9. Moreover, it should be noted that it is apparent for a man skilled in the art that said matrices $\mathbf{M}_{OBJ}^{ENERGY}$ and $\mathbf{M}_{EAO}^{ENERGY}$ are not a function of d_0.

Moreover, the preamble of independent claims 31, 32 and 33 has been adapted in the same manner as the preamble of the independent claim 1.

Also, independent claim 33 has been clarified in the same manner as the dependent claim 9 in that the equations for $\mathbf{X}_{OBJ}$ and $\mathbf{X}_{EAO}$ have been clarified. Moreover, the definition "wherein d_0 represents a channel of the downmix signal representation" has been deleted, as discussed above.

Independent claim 36 has been clarified in parallel with the independent claim 33.

4. Clarity:

It is believed that the clarity objections raised with respect to independent claims 31 to 36 have been overcome by the above-discussed clarifications.

In particular, the preambles of independent claims 31 to 33 are now clear. Also, in the independent claims 33 and 36, it is now clear from the chosen form of notation that there is a multiplication between the respective matrices and d_0. Also, the unclarity with respect to the definition of d_0 has been eliminated.

5. Requests:

In view of the amended documents and the above discussion, it is respectfully requested to issue a positive International Preliminary Examination Report for claims 1 to 37.

However, should there be any questions or objections, it is respectfully requested to telephone the undersigned representative of the applicant or to issue an Official Communication.

Respectfully submitted,

Franz Zinkler

Encls.:
Voucher (additional examination fee)
Amended claims 1-37 (clean and annotated copies)

Claims (clean version)

1. An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information,

wherein the second audio information is an audio information describing the audio objects of the second audio object type in a combined manner;

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the audio signal decoder is configured to provide the upmix signal representation in dependence on a residual information associated to a subset of audio objects represented by the downmix signal representation,

wherein the object separator is configured to decompose the downmix signal representation to provide the first audio information describing a first set of one or more audio objects of a first audio object type to which residual information is associated, and the second audio information describing a second set of one or more audio objects of a second audio object type, to which no residual information is associated, in dependence on the downmix signal representation and using the residual information; and

wherein the audio signal processor is configured to process the second audio information, to perform an object-individual processing of the audio objects of the second audio object type, taking into consideration object-related parametric information associated with more than two audio objects of the second audio object type; and

wherein the residual information describes a residual distortion, which is expected to remain if an audio object of the first audio object type is isolated merely using the object-related parametric information.

2. The audio signal decoder (100; 200; 500; 590) according to claim 1, wherein the object separator is configured to provide the first audio information such that one or more audio objects of the first audio object type are emphasized over audio objects of the second audio object type in the first audio information, and

wherein the object separator is configured to provide the second audio information such that audio objects of the second audio object type are emphasized over audio objects of the first audio object type in the second audio information.

3. The audio signal decoder (100; 200; 500; 570) according to one of claims 1 to 2, wherein the audio signal processor is configured to process the second audio information (134; 264; 564; 564a) in dependence on the object-related parametric information (110; 212; 512; 512a) associated with the audio objects of the second audio object type and independent from the object-related parametric information (110; 212; 512; 512a) associated with the audio objects of the first audio object type.

4. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 3, wherein the object separator is configured to obtain the first audio information (132; 262; 562; 562a, $\mathbf{X}_{EAO}$) and the second audio information (134; 264; 564; 564a, $\mathbf{X}_{OBJ}$) using a linear combination of one or more downmix signal channels of the downmix signal representation and one or more residual channels, wherein the object separator is configured to obtain combination parameters for performing the linear combination in dependence on downmix parameters associated with the audio objects of the first audio object type ($m_0 \ldots m_{NEAO-1}$; $n_0 \ldots n_{NEAO-1}$) and in dependence on channel prediction coefficients ($c_{j,0}$, $c_{j,1}$) of the audio objects of the first audio object type.

5. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 4, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C},$$

wherein

$$\mathbf{M}^{Prediction} = \begin{pmatrix} \mathbf{M}_{OBJ}^{Prediction} \\ \hline \mathbf{M}_{EAO}^{Prediction} \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\tilde{D}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein C describes a matrix representing a plurality of channel prediction coefficients, $\tilde{c}_{J,0}$, $\tilde{c}_{J,1}$;

wherein l_0 and r_0 represent channels of the downmix signal representation;

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix, entries of which describe a mapping of enhanced audio objects to channels of an enhanced audio object signal $\mathbf{X}_{EAO}$;

wherein the object separator is configured to obtain the inverse downmix matrix $\tilde{D}^{-1}$ as an inverse of an extended downmix matrix $\tilde{D}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the object separator is configured to obtain the matrix C as

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein the object separator is configured to compute the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ and}$$

wherein the object separator is configured to derive constrained prediction coefficients $c_{j,0}$ and $c_{j,1}$ from the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ using a constraining algorithm, or to use the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as the prediction coefficients $c_{j,0}$ and $c_{j,1}$;

wherein energy quantities P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ and $P_{RoCo,j}$ are defined as

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

wherein parameters OLD_L, OLD_R and $IOC_{L,R}$ correspond to audio objects of the second audio object type and are defined according to

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

wherein $d_{0,i}$ and $d_{1,i}$ are downmix values associated with the audio objects of the second audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the second audio object type;

wherein N is a total number of audio objects;

wherein N_{EAO} is a number of audio objects of the first audio object type;

wherein $IOC_{0,1}$ is an inter-object-correlation value associated with a pair of audio objects of the second audio object type;

wherein $e_{i,j}$ and $e_{L,R}$ are covariance values derived from object-level-difference parameters and inter-object-correlation parameters; and

wherein $e_{i,j}$ are associated with a pair of audio objects of the 1st audio object type and $e_{L,R}$ is associated with a pair of audio objects of the second audio object type.

6. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 4, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} \frac{d_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} \frac{d_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C}$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\tilde{\mathbf{D}}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein $\mathbf{C}$ describes a matrix representing a plurality of channel prediction coefficients, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

wherein d_0 represents a channel of the downmix signal representation; and

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

7. The audio signal decoder according to claim 6, wherein the object separator is configured to obtain the inverse downmix matrix $\tilde{\mathbf{D}}^{-1}$ is an inverse of an extended downmix matrix $\tilde{\mathbf{D}}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{c|ccc} 1 & m_0 & \cdots & m_{N_{EAO}-1} \\ \hline m_0 & -1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the object separator is configured to obtain the matrix C as

$$\mathbf{C} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline c_0 & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1} & 0 & \cdots & 1 \end{array}\right);$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type.

8. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 4, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{Energy} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{Energy} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L and OLD_R are common object level difference values associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

9. The audio signal decoder according to one of claims 1 to 3, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energy} d_0$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Energy} d_0$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{Energy} = \left(\sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right)$$

$$\mathbf{M}_{EAO}^{Energy} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L is a common object level difference value associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix;

wherein the matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ are applied to a representation d_0 of a single SAOC downmix signal.

10. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 9, wherein the object separator is configured to apply a rendering matrix to the first audio information (132; 262; 562; 562a) to map object signals of the first audio information onto audio channels of the upmix audio signal representation (120; 220, 222; 562; 562a).

11. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 10, wherein the audio signal processor (140; 270; 570; 570a) is configured to perform a stereo preprocessing of the second audio information (134; 264; 564; 564a) in dependence on a rendering information (M_{rcn}), an object-related covariance information (**E**), a downmix information (**D**), to obtain audio channels of the processed version of the second audio information;

12. The audio signal decoder (100; 200; 500; 590) according to claim 11, wherein the audio signal processor (140; 270; 570; 570a) is configured to perform the stereo processing to map an estimated audio object contribution (ED^*JX) of the second audio information (134; 264; 564; 564a) onto a plurality of channels of the upmix audio signal representation in dependence on a rendering information and a covariance information.

13. The audio signal decoder according to claim 11 or claim 12, wherein the audio signal processor is configured to add a decorrelated audio signal contribution (P_2X_d), obtained on the basis of one or more audio channels of the second audio information, to the second audio information, or an information derived from the second audio information, in dependence on a render upmix error information (R) and one or more decorrelated-signal-intensity scaling values (w_{d1}, w_{d2}).

14. The audio signal decoder according to one of claims 1 to 10, wherein the audio signal processor (140; 270; 570; 570a) is configured to perform a postprocessing of the second audio information (134; 264; 564; 564a) in dependence on a rendering information (A), an object-related covariance information (E) and a downmix information (D).

15. The audio signal decoder according to claim 14, wherein the audio signal processor is configured to perform a mono-to-binaural processing of the second audio information, to map a single channel of the second audio information onto two channels of the upmix signal representation, taking into consideration a head-related transfer function.

16. The audio signal decoder according to claim 14, wherein the audio signal processor is configured to perform a mono-to-stereo processing of the second audio information, to map a single channel of the second audio information onto two channels of the upmix signal representation.

17. The audio signal decoder according to claim 14, wherein the audio signal processor is configured to perform a stereo-to-binaural processing of the second audio information, to map two channels of the second audio information onto two channels of the upmix signal representation, taking into consideration a head-related transfer function.

18. The audio signal decoder according to claim 14, wherein the audio signal processor is configured to perform a stereo-to-stereo processing of the second audio information, to map two channels of the second audio information onto two channels of the upmix signal representation.

19. The audio signal decoder according to one of claims 1 to 18, wherein the object separator is configured to treat audio objects of the second audio object type, to which no residual information is associated, as a single audio object, and

wherein the audio signal processor (140; 270; 570; 570a) is configured to consider object-specific rendering parameters associated to the audio objects of the second audio object type to adjust contributions of the audio objects of the second audio object type to the upmix signal representation.

20. The audio signal decoder according to one of claims 1 to 19, wherein the object separator is configured to obtain one or two common object level difference values (OLD_L, OLD_R) for a plurality of audio objects of the second audio object type; and

wherein the object separator is configured to use the common object level difference value for a computation of channel prediction coefficients (CPC); and

wherein the object separator is configured to use the channel prediction coefficients to obtain one or two audio channels representing the second audio information.

21. The audio signal decoder according to one of claims 1 to 20, wherein the object separator is configured to obtain one or two common object level difference values (OLD_L, OLD_R) for a plurality of audio objects of the second audio object type, and

wherein the object separator is configured to use the common object level difference value for a computation of entries of an matrix (**M**); and

wherein the object separator is configured to use the matrix (**M**) to obtain one or more audio channels representing the second audio information.

22. The audio signal decoder according to one of claims 1 to 21, wherein the object separator is configured to selectively obtain a common inter-object correlation value ($IOC_{L,R}$) associated to the audio object of the second audio object type in dependence on the object-related parametric information if it is found that there are two audio objects of the second audio object type, and to set the inter-object correlation value associated to the audio objects of the second audio object type to zero if it is found that there are more or less than two audio objects of the second audio object type; and

wherein the object separator is configured to use the common inter-object correlation value for a computation of entries of an matrix (**M**); and

wherein the object separator is configured to use the common inter-object correlation value associated to the audio objects of the second audio object type to obtain the one or more audio channels representing the second audio information.

23. The audio signal decoder according to one of claims 1 to 22, wherein the audio signal processor is configured to render the second audio information in dependence on the object-related parametric information, to obtain a rendered representation of the audio objects of the second audio object type as the processed version of the second audio information.

24. The audio signal decoder according to one of claims 1 to 23, wherein the object separator is configured to provide the second audio information such that the second audio information describes more than two audio objects of the second audio object type.

25. The audio signal decoder according to claim 24, wherein the object separator is configured to obtain, as the second audio information, a one-channel audio signal representation or a two-channel audio signal representation representing more than two audio objects of the second audio object type.

26. The audio signal decoder according to one of claims 1 to 25, wherein the audio signal processor is configured to receive the second audio information and to process the second audio information in dependence of the object-related parametric information, taking into consideration object-related parametric information associated with more than two audio objects of the second audio object type.

27. The audio signal decoder according to one of claims 1 to 26, wherein the audio signal decoder is configured to extract a total object number information (bsNumObjects) and a foreground object number information (bsNumGroupsFGO) from a configuration information (SAOCSpecificConfig) of the object-related parametric information, and to determine the number of audio objects of the second audio object type by forming a difference between the total object number information and the foreground object number information.

28. The audio signal decoder according to one of claims 1 to 27, wherein the object separator is configured to use object-related parametric information associated with N_{EAO} audio objects of the first audio object type to obtain, as the first audio information, N_{EAO} audio signals ($\mathbf{X}_{EAO}$) representing the N_{EAO} audio objects of the first audio object type and to obtain, as the second audio information, one or two audio signals ($\mathbf{X}_{OBJ}$) representing the $N\text{-}N_{EAO}$ audio objects of the second audio object type, treating the $N\text{-}N_{EAO}$ audio objects of the second audio object type as a single one-channel or a two-channel audio object; and

wherein the audio signal processor is configured to individually render the $N\text{-}N_{EAO}$ audio objects represented by the one or two audio signals of the second audio information using the object-related parametric information associated with the $N\text{-}N_{EAO}$ audio objects of the second audio object type.

29. A method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, the method comprising:

decomposing the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type, and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information, wherein the second audio information is an audio information describing the audio objects of the second audio object type in a combined manner; and

processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information; and

combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the upmix signal representation is provided in dependence on a residual information associated to a subset of audio objects represented by the downmix signal representation,

wherein the downmix signal representation is decomposed, to provide the first audio information describing a first set of one or more audio objects of a first audio object type to which residual information is associated, and the second audio information describing a second set of one or more audio objects of a second audio object type, to which no residual information is associated, in dependence on the downmix signal representation and using the residual information;

wherein an object-individual processing of the audio objects of the second audio object type is performed, taking into consideration object-related parametric information associated with more than two audio objects of the second audio object type; and

wherein the residual information describes a residual distortion, which is expected to remain if an audio object of the first audio object type is isolated merely using the object-related parametric information.

30. A computer program for performing the method according to claim 33 when the computer program runs on a computer.

31. An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information;

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C},$$

wherein

$$\mathbf{M}^{Prediction} = \begin{pmatrix} \mathbf{M}_{OBJ}^{Prediction} \\ \hline \mathbf{M}_{EAO}^{Prediction} \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\tilde{D}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein C describes a matrix representing a plurality of channel prediction coefficients, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

wherein l_0 and r_0 represent channels of the downmix signal representation;

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix, entries of which describe a mapping of enhanced audio objects to channels of an enhanced audio object signal $\mathbf{X}_{EAO}$;

wherein the object separator is configured to obtain the inverse downmix matrix $\tilde{D}^{-1}$ as an inverse of an extended downmix matrix $\tilde{D}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the object separator is configured to obtain the matrix C as

$$C = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein the object separator is configured to compute the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ and}$$

wherein the object separator is configured to derive constrained prediction coefficients $c_{j,0}$ and $c_{j,1}$ from the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ using a constraining algorithm, or to use the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as the prediction coefficients $c_{j,0}$ and $c_{j,1}$;

wherein energy quantities P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ and $P_{RoCo,j}$ are defined as

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i\neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

wherein parameters OLD_L, OLD_R and $IOC_{L,R}$ correspond to audio objects of the second audio object type and are defined according to

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

wherein $d_{0,i}$ and $d_{1,i}$ are downmix values associated with the audio objects of the second audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the second audio object type;

wherein N is a total number of audio objects;

wherein N_{EAO} is a number of audio objects of the first audio object type;

wherein $IOC_{0,1}$ is an inter-object-correlation value associated with a pair of audio objects of the second audio object type;

wherein $e_{i,j}$ and $e_{L,R}$ are covariance values derived from object-level-difference parameters and inter-object-correlation parameters; and

wherein $e_{i,j}$ are associated with a pair of audio objects of the 1st audio object type and $e_{L,R}$ is associated with a pair of audio objects of the second audio object type.

32. An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information;

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L and OLD_R are common object level difference values associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

33. An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or

more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information;

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} d_0$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} d_0$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \left(\sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right)$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L is a common object level difference value associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix;

wherein the matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ are applied to a representation d_0 of a single SAOC downmix signal.

34. A method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, the method comprising:

decomposing the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type, and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information; and

processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information; and

combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the first audio information and the second audio information are obtained according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C},$$

wherein

$$\mathbf{M}^{Prediction} = \begin{pmatrix} \mathbf{M}_{OBJ}^{Prediction} \\ \hline \mathbf{M}_{EAO}^{Prediction} \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\tilde{D}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein C describes a matrix representing a plurality of channel prediction coefficients, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

wherein l_0 and r_0 represent channels of the downmix signal representation;

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix, entries of which describe a mapping of enhanced audio objects to channels of an enhanced audio object signal $\mathbf{X}_{EAO}$;

wherein the inverse downmix matrix $\tilde{D}^{-1}$ is obtained as an inverse of an extended downmix matrix $\tilde{D}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the matrix C is obtained as

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ are computed as

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ and}$$

wherein constrained prediction coefficients $c_{j,0}$ and $c_{j,1}$ are derived from the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ using a constraining algorithm, or wherein the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ are used as the prediction coefficients $c_{j,0}$ and $c_{j,1}$;

wherein energy quantities P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ and $P_{RoCo,j}$ are defined as

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

wherein parameters OLD_L, OLD_R and $IOC_{L,R}$ correspond to audio objects of the second audio object type and are defined according to

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

wherein $d_{0,i}$ and $d_{1,i}$ are downmix values associated with the audio objects of the second audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the second audio object type;

wherein N is a total number of audio objects;

wherein N_{EAO} is a number of audio objects of the first audio object type;

wherein $IOC_{0,1}$ is an inter-object-correlation value associated with a pair of audio objects of the second audio object type;

wherein $e_{i,j}$ and $e_{L,R}$ are covariance values derived from object-level-difference parameters and inter-object-correlation parameters; and

wherein $e_{i,j}$ are associated with a pair of audio objects of the 1st audio object type and $e_{L,R}$ is associated with a pair of audio objects of the second audio object type.

35. A method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, the method comprising:

decomposing the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type, and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information; and

processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information; and

combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the first audio information and the second audio information are obtained according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{Energy} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{Energy} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L and OLD_R are common object level difference values associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

36. A method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, the method comprising:

decomposing the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object

type, and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information; and

processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information; and

combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the first audio information and the second audio information are obtained according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} d_0$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} d_0$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L is a common object level difference value associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix;

wherein the matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ are applied to a representation d_0 of a single SAOC downmix signal.

37. A computer program for performing the method according to one of claims 34 to 36 when the computer program runs on a computer.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 December 2010 (29.12.2010)

PCT

(10) International Publication Number
WO 2010/149700 A1

(51) International Patent Classification:
G10L 19/00 (2006.01) *H04S 7/00* (2006.01)
G10L 19/14 (2006.01)

(21) International Application Number:
PCT/EP2010/058906

(22) International Filing Date:
23 June 2010 (23.06.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/220,042 24 June 2009 (24.06.2009) US

(71) Applicant *(for all designated States except US)*: **FRAUNHOFER GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Hansastraße 27c, 80686 München (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HELLMUTH, Oliver** [DE/DE]; Geschwister-Voemel-Weg 60, 91052 Erlangen (DE). **FALCH, Cornelia** [DE/DE]; Kasseler Str. 12, 90491 Nürnberg (DE). **HERRE, Jürgen** [DE/DE]; Hallerstr. 24, 91054 Buckenhof (DE). **HILPERT, Johannes** [DE/DE]; Herrnhüttestraße 46, 90411 Nürnberg (DE). **RIDDERBUSCH, Falko** [DE/DE]; Adam-Kraft-Str. 57, 90419 Nürnberg (DE). **TERENTIEV, Leonid** [DE/DE]; Am Europakanal 26, App. 11, 91056 Erlangen (DE).

(74) Agents: **BURGER, Markus** et al.; Postfach 246, 82043 Pullach (DE).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) Title: AUDIO SIGNAL DECODER, METHOD FOR DECODING AN AUDIO SIGNAL AND COMPUTER PROGRAM USING CASCADED AUDIO OBJECT PROCESSING STAGES

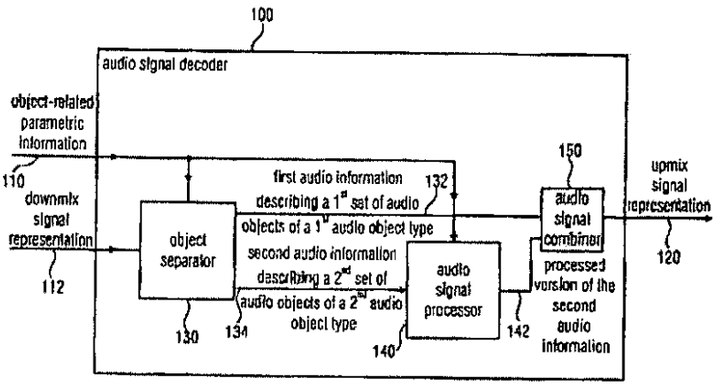


FIG 1

(57) Abstract: An audio signal decoder for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information comprises an object separator configured to decompose the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type and a second audio information describing a second set of one or more audio objects of a second audio object type, in dependence on the downmix signal representation and using at least a part of the object-related parametric information. The audio signal decoder also comprises an audio signal processor configured to receive the second audio information and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information. The audio signal decoder also comprises an audio signal combiner configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation.

WO 2010/149700 A1

277

PATENT COOPERATION TREATY PCT/EP2010/058906

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING SUBMISSION, OBTENTION OR TRANSMITTAL OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

To:

BURGER, Markus
Postfach 246
82043 Pullach
ALLEMAGNE

Date of mailing *(day/month/year)* 27 September 2010 (27.09.2010)	
Applicant's or agent's file reference FH100602PCT	**IMPORTANT NOTIFICATION**
International application No. PCT/EP2010/058906	International filing date *(day/month/year)* 23 June 2010 (23.06.2010)
ernational publication date *(day/month/year)* **Not yet published**	Priority date *(day/month/year)* 24 June 2009 (24.06.2009)
Applicant FRAUNHOFER GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V. et al	

The applicant is hereby notified of the date of receipt (or of obtaining by the International Bureau) of the priority document(s) relating to all earlier application(s) whose priority is claimed. Unless otherwise indicated by the letters "NR", in the right-hand column or by an asterisk appearing next to the date of receipt, the priority document concerned was submitted or transmitted to or obtained by the International Bureau in compliance with Rule 17.1(a), (b) or (b-*bis*). This Form replaces any previously issued notification concerning submission, transmittal or obtaining of priority documents.

Priority date	Priority application No.	Country or regional Office or PCT receiving Office	Date of receipt of priority document
24 June 2009 (24.06.2009)	61/220,042	US	20 September 2010 (20.09.2010)

The **letters "NR"** denote **a priority document which, on the date of mailing of this Form, had not yet been received or obtained by the International Bureau** in compliance with Rule 17.1(a), (b) or (b-*bis*). Where the applicant has failed to either submit, request to transmit or obtain and transmit the priority document within the applicable time limit under that Rule, **the attention of the applicant** is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

An asterisk "*" next to a date of receipt, denotes **a priority document submitted or transmitted to or obtained by the International Bureau but not in compliance with Rule 17.1(a), (b) or (b-*bis*)** (the priority document was received after the time limit prescribed in Rule 17.1(a); the request to prepare and transmit the priority document was submitted to the receiving Office after the applicable time limit under Rule 17.1(b) or the request to the receiving Office or the International Bureau to obtain the priority document was made after the applicable time limit under Rule 17.1(b-*bis*)). Even though the priority document was not furnished in compliance with Rule 17.1(a), (b) or (b-*bis*), the International Bureau will nevertheless transmit a copy of the document to the designated Offices, for their consideration. In case such a copy is not accepted by the designated Office as the priority document, Rule 17.1(c) provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Yolaine Cussac
e-mail pt05.pct@wipo.int
Telephone No. +41 22 338 74 05

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

To:
Burger, Markus
SCHOPPE, ZIMMERMANN, STÖCKELER & ZINKLER
Postfach 246
D-82043 Pullach
ALLEMAGNE

Applicant's or agent's file reference FH100602PCT	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/EP2010/058906	International filing date *(day/month/year)* 23 June 2010 (23-06-2010)
Applicant FRAUNHOFER GESELLSCHAFT ZUR FÖRDERUNG DER...	
	Date of mailing *(day/month/year)* 6 October 2010 (06-10-2010)

1. [X] The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the International Application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the International Search Report.

   Where? Directly to the International Bureau of WIPO, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Fascimile No.: (41-22) 338.82.70

   For more detailed instructions, see the notes on the accompanying sheet.

2. [ ] The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. [ ] **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

   [ ] the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

   [ ] no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   Shortly after the expiration of **18 months** from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90*bis*.1 and 90*bis*.3, respectively, before the completion of the technical preparations for international publication.

   The applicant may submit comments on an informal basis on the written opinion of the International Searching Authority to the International Bureau. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established. These comments would also be made available to the public but not before the expiration of 30 months from the priority date.

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices.

   In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months.

   See the Annex to Form PCT/IB/301 and, for details about the applicable time limits, Office by Office, see the *PCT Applicant's Guide*, National Chapters.

Name and mailing address of the International Searching Authority	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040 Fax: (+31-70) 340-3016	POQUET, Rosa Tel: +49 (0)89 2399-2911

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
FH100602PCT		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/EP2010/058906	23/06/2010	24/06/2009
Applicant		
FRAUNHOFER GESELLSCHAFT ZUR FÖRDERUNG DER...		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**
   a. With regard to the **language,** the international search was carried out on the basis of:
      [X] the international application in the language in which it was filed
      [ ] a translation of the international application into _____, which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b))
   b. [ ] This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).
   c. [ ] With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. [ ] **Certain claims were found unsearchable** (See Box No. II)

3. [ ] **Unity of invention is lacking** (see Box No III)

4. With regard to the **title,**
   [X] the text is approved as submitted by the applicant
   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**
   [X] the text is approved as submitted by the applicant
   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority

6. With regard to the **drawings,**
   a. the figure of the **drawings** to be published with the abstract is Figure No. 1
      [X] as suggested by the applicant
      [ ] as selected by this Authority, because the applicant failed to suggest a figure
      [ ] as selected by this Authority, because this figure better characterizes the invention
   b. [ ] none of the figures is to be published with the abstract

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2010/058906

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L19/00 G10L19/14 H04S7/00
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L H04S

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ENGDEGORD J ET AL: "Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding" 124TH AES CONVENTION, AUDIO ENGINEERING SOCIETY, PAPER 7377,, 17 May 2008 (2008-05-17), pages 1-15, XP002541458	1-8,10, 11,14-34
A	page 5, left-hand column, lines 14-20 sections 3.3.1-3.4.2, 3.7 figures 1,3,5,6	9,12,13
A	WO 2008/060111 A1 (LG ELECTRONICS INC [KR]; OH HYEN O [KR]; JUNG YANG WON [KR]) 22 May 2008 (2008-05-22) paragraphs [0005], [0006], [0008], [0009] figure 1	1-34

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents :

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

Date of the actual completion of the international search	Date of mailing of the international search report
28 September 2010	06/10/2010

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Tilp, Jan

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2010/058906

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008060111 A1	22-05-2008	AU 2007320218 A1	22-05-2008
		CA 2669091 A1	22-05-2008
		EP 2092516 A1	26-08-2009
		JP 2010509884 T	25-03-2010
		KR 20090082927 A	31-07-2009
		US 2009171676 A1	02-07-2009
		US 2008269929 A1	30-10-2008

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY
(PCT Rule 43*bis*.1)

Date of mailing
(day/month/year) see form PCT/ISA/210 (second sheet)

Applicant's or agent's file reference
see form PCT/ISA/220

FOR FURTHER ACTION
See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/EP2010/058906	23.06.2010	24.06.2009

International Patent Classification (IPC) or both national classification and IPC
INV. G10L19/00 G10L19/14 H04S7/00

Applicant
FRAUNHOFER GESELLSCHAFT ZUR FÖRDERUNG DER...

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☒ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will usually be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notifed the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

3. For further details, see notes to Form PCT/ISA/220.

Name and mailing address of the ISA:

European Patent Office
D-80298 Munich
Tel. +49 89 2399 - 0
Fax: +49 89 2399 - 4465

Date of completion of this opinion
see form PCT/ISA/210

Authorized Officer
Tilp, Jan
Telephone No. +49 89 2399-4753

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/EP2010/058906

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43bis.1(a))

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing filed or furnished:

   a. (means)

      ☐ on paper

      ☐ in electronic form

   b. (time)

      ☐ in the international application as filed

      ☐ together with the international application in electronic form

      ☐ subsequently to this Authority for the purposes of search

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	9-13, 15-22, 26
	No:	Claims	1-8, 14, 23-25, 27-34
Inventive step (IS)	Yes:	Claims	9, 12, 13
	No:	Claims	1-8, 10, 11, 14-34
Industrial applicability (IA)	Yes:	Claims	1-34
	No:	Claims	

2. Citations and explanations

see separate sheet

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/EP2010/058906

Box No. VIII Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

see separate sheet

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1 Reference is made to the following document:

D1 ENGDEGORD J ET AL: "Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding"124TH AES CONVENTION, AUDIO ENGINEERING SOCIETY, PAPER 7377,, 17 May 2008 (2008-05-17) , pages 1-15, XP002541458

2 INDEPENDENT CLAIMS

The lack of clarity mentioned below under Re Item VIII notwithstanding, the subject-matter of **claims 1, 33 and 34** is not new in the sense of Article 33(2) PCT, and the criteria of Article 33(1) PCT are therefore not met.

2.1 Document D1 discloses - applying the wording of present **claim 1**:

"An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal and representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) (see D1, e.g. figure 1a, wherein the decoder and mixer/renderer disclosed therein represent the claimed decoder and the output of the mixer/ renderer represents the claimed upmix signal representation) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation (see D1, e.g. figures 1a and 6 (the latter representing a special application scenario of the SAOC conceptual overview of figure 1a), wherein the TTN box disclosed in figure 6 represents the claimed separator, and wherein - according to D1, page 10, left-hand column, section 3.4, lines 4-14, and page 10, right-hand column, lines 8-11 - the output of said TTN box (consisting of audio signals corresponding to foreground and background objects) represents the claimed first and second audio information); and

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information (see D1, e.g. figure 6 , wherein - in accordance with D1, page 10, right-hand column, lines 11-14 - the mixing stage disclosed therein implies the claimed audio signal processor and is - as claimed - configured to receive and process the audio information corresponding to background object(s) (e.g. suppress background object(s) in the disclosed "solo" mode) in dependence on object-related parametric information represented by the enhanced karaoke/solo mode information input to the mixing stage); and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation (see D1, e.g. figure 6, wherein the mixing stage disclosed therein implies the claimed combination of first and processed second audio information)."

Consequently, claim 1 fails to fulfill the requirements of the PCT with respect to novelty (Article 33(2) PCT).

2.2 The subject-matter of **claims 33 and 34** corresponds, in terms of method and computer program features, to that of device claim 1. The objections raised against said latter claim therefore also apply, mutatis mutandis, to claims 33 and 34, which thus also fail to fulfill the requirements of the PCT with respect to novelty (Article 33(2) PCT).

3 DEPENDENT CLAIMS

3.1 The lack of clarity and/or conciseness mentioned below under Re Item VIII notwithstanding, dependent **claims 2-8, 10, 11 and 14-32** do not contain any features which, in combination with the features of any claim to which they refer, meet the requirements of the PCT in respect of novelty and/or inventive step, the reasons therefor being as follows:

3.1.1 The features additionally claimed by **claim 2** are already known from D1 - see e.g. figure 6 and page 11, left-hand column, lines 10-11, wherein residual information is associated to foreground objects (i.e. objects of first audio object type).

3.1.2 The features additionally claimed by **claim 3** are also already known from D1 - see e.g. page 11, left-hand column, lines 2-10 and lines 35-38 as well as equations (14)-(16), wherein the matrix M_{TTN} contains the claimed weights.

3.1.3 According to what is mentioned below under Re Item VIII, the features additionally claimed by **claim 4** appear to be already contained in what is claimed by claim 1.

3.1.4 The features additionally claimed in **claim 5** are already disclosed by D1 - see e.g. page 10, left-hand column, lines 8-12, where isolated processing of background objects (i.e. objects of second audio object type) is disclosed.

3.1.5 The features additionally claimed in **claim 6** are already disclosed by D1 - see e.g. equations (14), (16) and (17) and the corresponding textual passages on page 11, left-hand column.

3.1.6 The features additionally claimed in **claims 7 and 8** are already disclosed by D1 - see e.g. equations (14), (16) and (17), wherein the claimed $\mathbf{A}^{EAO}$ is assumed to be the identity matrix.

3.1.7 Taking into account the multi-channel object separation defined by claim 7 as well as what is disclosed by D1, section 3.3.1, the combination of features of **claims 10 and 11** referring to single-channel object separation is obvious for those skilled in the art.

3.1.8 The feature additionally claimed in **claim 14** is already disclosed by D1, figure 1b, which shows the claimed integrated decoder/separator and mixer/ renderer.

3.1.9 In light of D1, sections 3.3.1-3.3.3 and figure 3, the combinations of features of **claims 15-22** are obvious for those skilled in the art.

3.1.10 The features additionally claimed in **claims 23-25** are already disclosed by D1, section 3.3.4 (claim 23), page 5, left-hand column, lines 14-16 (claims 24-25), and sections 3.3.2 and 3.4.2 (claims 24-25).

3.1.11 In light of D1, page 5, left-hand column, lines 18-20, the combination of features of **claim 26** is obvious for those skilled in the art.

3.1.12 The objections with regard novelty raised against claim 1 also apply to non-concise dependent **claim 27**.

3.1.13 In light of D1, section 3.3.4, the combinations of features of **claims 28-30** is obvious for those skilled in the art.

3.1.14 In light of D1, section 3.7, lines 11-14, the combination of features of **claim 31** is obvious for those skilled in the art.

3.1.15 In light of D1, figure 5 and section 3.3.4, the combination of features of **claim 32** is obvious for those skilled in the art.

3.2 The combination of the features of dependent **claims 9, 12 and 13** is neither known from, nor rendered obvious by, the available prior art. In particular, the claimed specific equations for deriving prediction taking into account object-type dependent parameters, solves the problem of efficiently separating the received audio objects.

Re Item VIII

Certain observations on the international application

1 The application does not meet the requirements of Article 6 PCT, because certain claims are not clear; not concise or not supported by the description.

1.1 **Claim 1** is not supported by the description as required by Article 6 PCT, as its scope is broader than justified by the description and drawings.
In particular, the description discloses (see e.g. page 7, lines 33-36) "using at least part of the object-related parametric information in a first processing step performed by the object separator", whereas the object separator as of claim 1 is not claimed to use the object-related parametric information.

1.2 The term "residual information" used in **claim 2** is vague and unclear and leaves the reader in doubt as to the meaning of the technical feature to which it refers, thereby rendering the definition of the subject-matter of said claim unclear, Article 6 PCT. In light of the description (see e.g. page 8, lines 34-37), it appears possible to define said term in more concrete words.

1.3 The term "emphasized" used in **claim 3** is vague and unclear and leaves the reader in doubt as to the meaning of the technical feature to which it refers, thereby rendering the definition of the subject-matter of said claim unclear, Article 6 PCT.

1.4 **Claim 3** is also unclear in that it claims "audio objects of the second audio object type in the first audio information" and "audio objects of the first audio object type in the second audio information", whilst claim 1, which claim 3 is claimed to depend on, claims said "first audio information describing [...] audio objects of a first audio object type, and a second audio information describing [...] audio objects of a second audio object type".

1.5 **Claim 4** appears not to claim any additional features which are not already claimed in claim 1, which claim 4 is claimed to depend on. Claim 4 therefore lacks conciseness and as such does not meet the requirements of Article 6 PCT.

1.6 The quantities "$\mathbf{M}_{OBJ}^{Prediction}$" and "$\mathbf{M}_{EAO}^{Prediction}$" as well as the terms "extended downmix matrix", "residual channels" and "EAO pre-rendering matrix" used in **claim 7** are vague and unclear and leave the reader in doubt as to the meaning of the technical feature to which they refer, thereby rendering the definition of the subject-matter of said claim unclear, Article 6 PCT.

1.7 It is not clear whether the quantities "$c_{0,0}$, $c_{1,0}$, ..., $c_{NEAO-1,0}$" and "$c_{0,1}$, $c_{1,1}$, ..., $c_{NEAO-1,1}$" claimed in **claim 8** are equal to the channel prediction coefficients claimed in claim 7, which claim 8 is claimed to depend on.

1.8 **Claim 13** is unclear in that it claims the expressions "$\mathbf{M}_{OBJ}^{Energy}(d_0)$" and "$\mathbf{M}_{EAO}^{Energy}(d_0)$", which are ambiguous in that they can be interpreted either i) as matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ being dependent on d_0 or ii) as products of said matrices and with d_0.

1.9 **Claim 17** claims "a decorrelated audio signal contribution" and "decorrelated-signal-intensity scaling values". It is not clear in respect of what further quantities said contribution and scaling values are decorrelated, thus rendering the subject-matter of claim 17 unclear.

1.10 The distinction between the expressions "to render" as claimed in **claim 27** and "to process" as claimed in claim 1 is not clear. Consequently, claim 27 appears not to claim any additional features which are not already claimed in claim 1, which claim 27 is claimed to depend on. Claim 27 therefore lacks conciseness and as such does not meet the requirements of Article 6 PCT.

1.11 The above objections concerning clarity, conciseness and/or support by the description also apply, mutatis mutandis, to all corresponding, inter-related, and further dependent claims.

1.12 In the following, all unclear claims are interpreted in the broadest possible sense.

Possible steps after receipt of the international search report (ISR) and written opinion of the International Searching Authority (WO-ISA)

General information	For all international applications filed on or after 01/01/2004 the competent ISA will establish an ISR. It is accompanied by the WO-ISA. Unlike the former written opinion of the IPEA (Rule 66.2 PCT), the WO-ISA is not meant to be responded to, but to be taken into consideration for further procedural steps. This document explains about the possibilities.
Amending claims under Art. 19 PCT	Within 2 months after the date of mailing of the ISR and the WO-ISA the applicant may file amended claims under Art. 19 PCT directly with the International Bureau of WIPO. The PCT reform of 2004 did not change this procedure. For further information please see Rule 46 PCT as well as form PCT/ISA/220 and the corresponding Notes to form PCT/ISA/220.
Filing a demand for international preliminary examination	In principle, the WO-ISA will be considered as the written opinion of the IPEA. This should, in many cases, make it unnecessary to file a demand for international preliminary examination. If the applicant nevertheless wishes to file a demand this must be done before expiry of 3 months after the date of mailing of the ISR/WO-ISA or 22 months after priority date, whichever expires later (Rule 54bis PCT). Amendments under Art. 34 PCT can be filed with the IPEA as before, normally at the same time as filing the demand (Rule 66.1 (b) PCT). If a demand for international preliminary examination is filed and no comments/amendments have been received the WO-ISA will be transformed by the IPEA into an IPRP (International Preliminary Report on Patentability) which would merely reflect the content of the WO-ISA. The demand can still be withdrawn (Art. 37 PCT).
Filing informal comments	After receipt of the ISR/WO-ISA the applicant may file informal comments on the WO-ISA directly with the International Bureau of WIPO. These will be communicated to the designated Offices together with the IPRP (International Preliminary Report on Patentability) at 30 months from the priority date. Please also refer to the next box.
End of the international phase	At the end of the international phase the International Bureau of WIPO will transform the WO-ISA or, if a demand was filed, the written opinion of the IPEA into the IPRP, which will then be transmitted together with possible informal comments to the designated Offices. The IPRP replaces the former IPER (international preliminary examination report).
Relevant PCT Rules and more information	Rule 43 PCT, Rule 43bis PCT, Rule 44 PCT, Rule 44bis PCT, PCT Newsletter 12/2003, OJ 11/2003, OJ 12/2003

BNSDOCID: <XS_ _20070401OCK_L>

Bitte beachten Sie, dass angeführte Nichtpatentliteratur (wie z. B. wissenschaftliche oder technische Dokumente) je nach geltendem Recht dem Urheberrechtsschutz und/oder anderen Schutzarten für schriftliche Werke unterliegen könnte. Die Vervielfältigung urheberrechtlich geschützter Texte, ihre Verwendung in anderen elektronischen oder gedruckten Publikationen und ihre Weitergabe an Dritte ist ohne ausdrückliche Zustimmung des Rechtsinhabers nicht gestattet.

Veuillez noter que les ouvrages de la littérature non-brevets qui sont cités, par exemple les documents scientifiques ou techniques, etc., peuvent être protégés par des droits d'auteur et/ou toute autre protection des écrits prévue par les législations applicables. Les textes ainsi protégés ne peuvent être reproduits ni utilisés dans d'autres publications électroniques ou imprimées, ni rediffusés sans l'autorisation expresse du titulaire du droit d'auteur.

Please be aware that cited works of non-patent literature such as scientific or technical documents or the like may be subject to copyright protection and/or any other protection of written works as appropriate based on applicable laws. Copyrighted texts may not be copied or used in other electronic or printed publications or re-distributed without the express permission of the copyright holder.

XS CPRTENFRDE

BNSDOCID: <XS_ 2006100103CF_1 >

REIVINDICACIONES

1. Un decodificador de señales de audio (100; 200; 500; 590) para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente (1 12; 210; 510; 510a), una información paramétrica relacionada con los objetos (1 10; 2 12; 5 12; 5 12a) donde el decodificador de señales de audio comprende:

un separador de objetos (130; 260; 520; 520a) configurado para descomponer una representación de señal de mezcla descendente, para producir una primera información de audio (132; 262; 562; 562a) que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio, y una segunda información de audio (134; 264; 564; 564a) que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de señal de mezcla descendente y usar al menos una parte de la información paramétrica relacionada con los objetos, donde la segunda información de audio es una información de audio que describe los objetos de audio del segundo tipo de objetos de audio en una forma combinada;

un procesador de señales de audio configurado para recibir la segunda información de audio (134; 264; 564; 564a) y para procesar la segunda información de audio que depende de la información paramétrica relacionada con los objetos, para obtener una versión procesada (142; 272; 572; 572a) de la segunda información de audio y

un combinador de señales de audio (150; 280; 580; 580a) configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente; donde el decodificador de señales de audio está configurado para producir la representación de señal de mezcla ascendente dependiendo de una información residual asociada a una subserie de objetos de audio representados por la representación de señal de mezcla descendente, donde el separador de objetos está configurado para descomponer la representación de señal de mezcla descendente para producir la primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio a los cuales está asociada la información residual, y la segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio, a los cuales no se asocia ninguna información

residual, dependiendo de la representación de la señal de mezcla descendente y usando la información residual; y

el procesador de las señales de audio está configurado para procesar la segunda información de audio, para llevar a cabo un proceso individual de objeto de los objetos de audio del segundo tipo de objetos de audio, teniendo en consideración la información paramétrica relacionada con los objetos asociada con más de dos objetos de audio del segundo tipo de objetos de audio; y

donde la información residual describe una distorsión residual, la cual se espera que permanezca si un objeto de audio del primer tipo de objetos de audio es aislado solamente usando la información paramétrica relacionada con los objetos.

2. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con la reivindicación 1, donde el separador de objetos está configurado para producir la primera información de audio de tal manera que uno o más objetos de audio del primer tipo de objetos de audio se realcen con respecto a los objetos de audio del segundo tipo de objetos de audio de la primera información de audio, y donde el separador de objetos está configurado para producir la segunda información de audio de tal manera que los objetos de audio del segundo tipo de objetos de audio sobresalgan con respecto a los objetos de audio del primer tipo de objetos de audio de la segunda información de audio.

3. El decodificador de señales de audio (100; 200; 500; 570) de acuerdo con una de las reivindicaciones 1 o 2, donde el procesador de señales de audio está configurado para procesar la segunda información de audio (134; 264; 564; 564a) dependiendo de la información paramétrica relacionada con los objetos (110; 212; 512; 512a) asociados con los objetos de audio del segundo tipo de objetos de audio e independiente de la información paramétrica relacionada con los objetos (110; 212; 512; 512a) asociado a los objetos de audio del primer tipo de objetos de audio.

4. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 3, donde el separador de objetos está configurado para obtener la primera información de audio (132; 262; 562; 562a, X_{EAO}) y la segunda información de audio (134; 264; 564; 564a, X_{OBJ}) usando una combinación lineal de uno o más canales de señal de mezcla descendente de la

representación de señal de mezcla descendente y uno o más canales residuales, donde el separador de objetos está configurado para obtener parámetros de combinación para ejecutar la combinación lineal dependiendo de los parámetros de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio (m_0 ...$m_{NEAO0-1}$) y dependiendo de coeficientes de predicción de canales ($c_{j,0}$, $c_{j,1}$) de los objetos de audio del primer tipo de objetos de audio.

5. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 4, en el cual el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Predicción}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Predicción}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{M}_{\text{Predicción}} = \tilde{\mathbf{D}}^{-1}\mathbf{C}$$

En donde

$$\mathbf{M}^{Prediction} = \left(\frac{\mathbf{M}_{OBJ}^{Prediction}}{\mathbf{M}_{EAO}^{Prediction}} \right)$$

donde $\mathbf{X}_{OBJ}$ representa los canales de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objeto de la primera información de audio;

donde $\tilde{D}^{-1}$ representa una matriz que es inversa de una matriz de mezcla descendente ampliada;

donde C describes una matriz que representa una pluralidad de coeficientes de predicción de canales, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

donde l_0 y r_0 representan los canales de la representación de señal de mezcla descendente;

donde $\mathbf{res}_0$ a $\mathbf{res}_{NEAO-1}$ representan canales residuales y

donde $\mathbf{A}^{EAO}$ es una matriz de pre–renderización de EAO, las entradas de la cual describen un mapeo de objetos de audio mejorados en canales de una señal de audio $\mathbf{X}_{EAO}$ mejorada;

en el cual el separador de obietos está configurado para obtener la matriz de mezcla descendente inversa $\tilde{D}^{-1}$ en forma inversa a una matriz de mezcla descendente ampliada $\tilde{D}$ que se define de la siguiente manera:

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

donde el separador objetos está configurado para obtener la matriz C de la siguiente manera:

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

donde m_0 a m_{NEAO-1} son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio;

donde n_0 a n_{NEAO-1} son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio.

en el cual el separador de objetos está configurado para computar los coeficientes de predicción $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$; de la siguiente manera:

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ y}$$

donde el separador de objetos está configurado para derivar coeficientes de predicción restringidos $c_{j,0}$ y $c_{j,1}$ a partir de los coeficientes de predicción $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

utilizando un algoritmo de restricción, o para usar los coeficientes de restricción $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$; como coeficientes de predicción $c_{j,0}$ y $c_{j,1}$;

donde las cantidades de energía P_{Lo}, P_{Ro}, P_{LoRo}, P_{LoCoj}, y $P_{RoCo,j}$, se definen como

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k},$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k},$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k},$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j},$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}.$$

donde los parámetros OLD_L, OLDR y $IOC_{L,R}$ corresponden a objetos de audio del segundo tipo de objetos de audio y se definen de acuerdo con

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & \text{de otro modo.} \end{cases}$$

.

donde $d_{0,i}$ y $d_{1,i}$ son valores de mezcla descendente asociados a los objetos de audio del segundo tipo de objetos de audio;

donde OLD_i son valores de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio;

donde N es un número total de objetos de audio;

donde N_{EAO} es un número de objetos de audio del primer tipo de objetos de audio;

donde $IOC_{0,1}$ es un valor de correlación entre objetos con un par de objetos de audio del segundo tipo de objetos de audio;

donde $e_{i,j}$ y $e_{L,R}$ son valores de covarianza derivados de los parámetros de diferencia de nivel de los objetos y parámetros de correlación entre objetos y donde $e_{i,j}$ está asociado a un par de objetos de audio del primer tipo de objetos de audio y $e_{L,R}$ está asociado a un par de objetos de audio del segundo tipo de objetos de audio.

6. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 4, donde el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Predicción}} \begin{pmatrix} \frac{d_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Predicción}} \begin{pmatrix} \frac{d_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

donde

$$\mathbf{M}_{\text{Predicción}} = \tilde{\mathbf{D}}^{-1}\mathbf{C}$$

donde $\mathbf{X}_{OBJ}$ representa un canal de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objeto de la primera información de audio;

donde $\tilde{\mathbf{D}}^{-1}$ representa una matriz que es inversa a una matriz de mezcla descendente ampliada;

donde $\boldsymbol{C}$ describe una matriz que representa una pluralidad de coeficientes de predicción de canales $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

donde d_0 representa un canal de la representación de señal de mezcla descendente; y

donde $\mathbf{res_0}$ **a** $\mathbf{res_{NEAO-1}}$ representan canales residuales y

donde $\mathbf{A}^{EAO}$ es una matriz de pre–renderización de EAO.

7. El decodificador de señales de audio de acuerdo con la reivindicación 6, donde el separador de objetos está configurado para obtener la matriz de mezcla descendente inversa $\tilde{\mathbf{D}}^{-1}$ que es inversa a una matriz de mezcla descendente ampliada $\tilde{D}$ que se define como

$$\tilde{\mathbf{D}} = \left(\begin{array}{c|ccc} 1 & m_0 & \cdots & m_{N_{EAO}-1} \\ \hline m_0 & -1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

donde el separador de objetos está configurado para obtener la matriz C de la siguiente manera

$$\mathbf{C}=\left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline c_0 & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1} & 0 & \cdots & 1 \end{array}\right);$$

donde m_0 a m_{NEAO-1},son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio.

8. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 4, donde el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ}=\mathbf{M}_{OBJ}^{\text{Energía}}\begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO}=\mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Energía}}\begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

donde $\mathbf{X}_{OBJ}$ representa los canales de la segunda información de audio;

donde $\boldsymbol{X}_{EAO}$ representa las señales de objeto de la primera información de audio;

donde

$$\mathbf{M}_{OBJ}^{\text{Energía}}=\begin{pmatrix} \sqrt{\dfrac{OLD_L}{OLD_L+\sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\dfrac{OLD_R}{OLD_R+\sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energía}}=\begin{pmatrix} \sqrt{\dfrac{m_0^2 OLD_0}{OLD_L+\sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\dfrac{n_0^2 OLD_0}{OLD_R+\sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\dfrac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L+\sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\dfrac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R+\sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

donde m_0 a m_{NEAO-1},son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio;

donde n_0 a n_{NEAO-1} son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio

donde OLD_i son valores de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio;

donde OLD_L y OLD_R son valores comunes de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio

y donde A^{EAO} es una matriz de pre–renderización.

9. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 3, donde el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energía}\left(d_0\right)$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{Energía}\left(d_0\right)$$

donde $\mathbf{X}_{OBJ}$ representa un canal de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objeto de la primera información de audio;

donde

$$\mathbf{M}_{OBJ}^{Energía} = \left(\sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}}\right)$$

$$\mathbf{M}_{EAO}^{Energía} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

donde m_0 a m_{NEAO-1},son valores de mezcla descendente asociados a los objetos de audio del primer tipo de objetos de audio;

donde OLD_i son valores de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio;

donde OLD_L es un valor común de diferencia de nivel de los objetos asociados a los objetos de audio del segundo tipo de objetos de audio y

donde $\mathbf{A}^{EAO}$ es una matriz de pre–renderización;

donde las matrices $\mathbf{M}_{OBJ}^{Energía}$ y $\mathbf{M}_{EAO}^{Energía}$ son aplicadas a una representación d_0 de una sola señal de mezcla descendente SAOC.

10. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 9, en el cual el separador de objetos está configurado para aplicar una matriz de renderización a la primera información de audio (132; 262; 562; 562a) para mapear las señales de objeto de la primera información de audio sobre canales de audio de la representación de señal de audio de mezcla ascendente (120; 220,222; 562; 562a).

11. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con una de las reivindicaciones 1 a 10, en el cual el procesador de señales de audio (140; 270; 570; 570a) está configurado para ejecutar un preprocesamiento estéreo de la segunda información de audio (134; 264; 564; 564a) dependiendo de una información de renderización (M_{ren}), una información de covarianza relacionada con los objetos (E), una información de mezcla descendente (D), para obtener canales de audio de acuerdo con la versión procesada de la segunda información de audio.

12. El decodificador de señales de audio (100; 200; 500; 590) de acuerdo con la reivindicación 11, donde el procesador de señales de audio (140; 270; 570; 570a) está configurado para ejecutar el procesamiento estéreo a fin de mapear una contribución estimada de los objetos de audio (ED*JX) de la segunda información de audio (134; 264; 564; 564a) sobre una pluralidad de canales de la representación de señal de mezcla ascendente, dependiendo de una información de renderización y una información de covarianza.

13. El decodificador de señales de audio de acuerdo con la reivindicación 11 o la reivindicación 12, en el cual el procesador de señales de audio está configurado para agregar una contribución de señales de audio descorrelacionadas (P_2X_d), obtenida sobre la base de uno o más canales de audio de la segunda información de audio, a la segunda información de audio, o una información derivada de la segunda información de audio, que depende de una información de error de renderización (R) y uno o más valores de escalamiento de intensidad de señal descorrelacionados (w_{d1}, w_{d2}).

14. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 10, en el cual el procesador de señales de audio (140; 270;

570; 570a) está configurado para ejecutar un postprocesamiento de la segunda información de audio (134; 264; 564; 564a) dependiendo de una información de renderización (A), una información de covarianza relacionada con los objetos (E) y una información de mezcla descendente (D).

15. El decodificador de señales de audio de acuerdo con la reivindicación 14, en el cual el procesador de señales de audio está configurado para ejecutar un procesamiento de mono a binaural de la segunda información de audio, para mapear un solo canal de la segunda información de audio sobre dos canales de la representación de señal de mezcla ascendente, tomando en cuenta una función de transferencia de cabeza.

16. El decodificador de señales de audio de acuerdo con la reivindicación 14, donde el procesador de señales de audio está configurado para ejecutar un procesamiento de mono a estéreo de la segunda información de audio, para mapear un solo canal de la segunda información de audio sobre dos canales de de la representación de señal de mezcla ascendente.

17. El decodificador de señales de audio de acuerdo con la reivindicación 14, donde el procesador de señales de audio está configurado para ejecutar un procesamiento de estéreo a binaural de la segunda información de audio, para mapear dos canales de la segunda información de audio sobre dos canales de la representación de señal de mezcla ascendente, tomando en cuenta una función de transferencia de cabeza.

18. El decodificador de señales de audio de acuerdo con la reivindicación 14, donde el procesador de señales de audio está configurado para ejecutar un procesamiento de estéreo a estéreo de la segunda información de audio sobre dos canales de la segunda información de audio sobre dos canales de la representación de señal de mezcla ascendente.

19. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 18, donde el separador de objetos está configurado para tratar objetos de audio del segundo tipo de objetos de audio, con los cuales no está asociada ninguna información residual, en forma de objeto de audio único y donde el procesador de señales de audio (140; 270; 570; 570a) está configurado para considerar parámetros de renderización de objetos específicos asociados a los objetos de audio del segundo tipo de objetos de audio a fin de ajustar las

contribuciones de los objetos de audio del segundo tipo de objetos de audio a la representación de señal de mezcla ascendente.

20. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 19, donde el separador de objetos está configurado para obtener uno o dos valores comunes de diferencia de nivel de los objetos (OLD_L, OLD_R) correspondientes a una pluralidad de objetos de audio del segundo tipo de objetos de audio; y

donde el separador de objetos está configurado para usar el valor común de diferencia de nivel de los objetos para el cómputo de los coeficientes de predicción de canales (CPC); y

donde el separador de objetos está configurado para usar los coeficientes de predicción de canales a fin de obtener uno o dos canales de audio que representan la segunda información de audio.

21. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 20, donde el separador de objetos está configurado para obtener uno o dos o uno o dos valores comunes de diferencia de nivel de los objetos (OLD_L, OLD_R) correspondientes a una pluralidad de objetos de audio del segundo tipo de objetos de audio; y

donde el separador de objetos está configurado para usar el valor común de diferencia de nivel de los objetos para el cómputo de las anotaciones de una matriz (M); y

donde el separador de objetos está configurado para usar la matriz (M) a fin de obtener uno o dos canales de audio que representan la segunda información de audio.

22. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 21, donde el separador de objetos está configurado para usar la matriz (M) a fin de obtener selectivamente un valor común de correlación entre objetos ($IOC_{L,R}$) asociado al objeto de audio del segundo tipo de objetos de audio dependiendo la información paramétrica relacionada con los objetos si se encuentra que hay dos objetos de audio del segundo tipo de objetos de audio, y para fijar el valor común de correlación entre objetos de los objetos de audio del segundo tipo de objetos de audio en cero si se encuentra que hay más o menos de dos objetos de audio del segundo tipo de objetos y audio; y

donde el separador de objetos está configurado para usar el valor común de correlación entre objetos para el cómputo de las anotaciones de una matriz (M); y donde el separador de objetos está configurado para usar el valor común de correlación entre objetos asociado a los objetos de audio del segundo tipo de objetos de audio para obtener dichos uno o más canales de audio que representan la segunda información de audio.

23. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 22, donde el procesador de señales de audio está configurado para renderizar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, para obtener una representación renderizada de los objetos de audio del segundo tipo de objetos de audio como versión procesada de la segunda información de audio.

24. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 23, donde el separador de objetos está configurado para producir la segunda información de audio de manera tal que la segunda información de audio describa más de dos objetos de audio del segundo tipo de objetos de audio.

25. El decodificador de señales de audio de acuerdo con la reivindicación 24, donde el separador de objetos está configurado para obtener, como segunda información de audio, una representación de una señal de audio de un solo canal o una representación de una señal de audio de dos canales que representa más de dos objetos de audio del segundo tipo de objetos de audio.

26. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 25, donde el procesador de señales de audio está configurado para recibir la segunda información de audio y para procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, tomando en cuenta la información paramétrica relacionada con los objetos asociada a más de dos objetos de audio del segundo tipo de objetos de audio.

27. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 26, donde el decodificador de señales de audio está configurado para extraer una información sobre número total de objetos (bsNumObjects) y una información del número de objetos en primer plano

(bsNumGroupsFGO) a partir de una información de configuración (SAOCSpecificConfig) de la información paramétrica relacionada con los objetos y para determinar el número de objetos de audio del segundo tipo de objetos de audio formando una diferencia entre la información del número total de objetos y la información del número de objetos en primer plano.

28. El decodificador de señales de audio de acuerdo con una de las reivindicaciones 1 a 27, donde el separador de objetos está configurado usar la información paramétrica relacionada con los objetos asociados a los objetos N_{EAO} del primer tipo de objetos de audio para obtener, como primera información de audio, N_{EAO} señales de audio (X_{EAO}) que representan los $N–N_{EAO}$ objetos de audio del primer tipo de objetos de audio y para obtener, como segunda información de audio, una o dos señales de audio (X_{OBJ}) que representan los $N–N_{EAO}$ objetos de audio del segundo tipo de objetos de audio, para tratar los N–NEAO objetos de audio del segundo tipo de objetos de audio como único objeto de audio de un canal o de dos canales; y

donde el procesador de señales de audio está configurado para renderizar individualmente los $N–N_{EAO}$ objetos de audio representados por dichas una o dos señales de audio de la segunda información de audio usando la información paramétrica relacionada con los objetos asociada a los $N–N_{EAO}$ objetos de audio del segundo tipo de objetos de audio.

29. Un método para producir una representación de señal de mezcla ascendente que depende de una representación de la señal de mezcla descendente e información paramétrica relacionada con los objetos, método que comprende:

descomponer la representación de señal de mezcla descendente, para producir una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de la señal de a descendente y utilizando por lo menos parte de la información paramétrica relacionada con los objetos, en donde la segunda información de audio es una información de audio que describe los objetos de audio del segundo tipo de objetos de audio en una forma combinada; y

procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, a fin de obtener una versión procesada de la segunda información de audio y

combinar la primera información de audio con la versión procesada de la segunda información de audio para obtener la representación de señal de mezcla ascendente;

en donde la representación de señal de mezclado ascendente se proporciona en dependencia de una información residual asociada a un subgrupo de objetos de audio representados por la representación de la señal de mezclado descendente,

en donde la representación de la señal de mezclado descendente está descompuesta, para proporcionar la primera información de audio que describe un primer grupo de uno o más objetos de audio de un primer tipo de objetos de audio al cual está asociada información residual, y la segunda información de audio describe un segundo grupo uno o más objetos de audio de un segundo tipo de objetos de audio, al cual no está asociada información residual, en dependencia de la representación de señal de mezcla descendente y usa la información residual;

donde un proceso individual de un objeto de los objetos de audio del segundo tipo de objetos de audio es llevado a cabo, teniendo en consideración información paramétrica relacionada con los objetos asociada con más de dos objetos de audio del segundo tipo de objetos de audio; y

donde la información residual describe una distorsión residual, que se espera que permanezca si un objeto de audio del primer tipo de objetos de audio es aislado simplemente usando la información paramétrica relacionada con los objetos.

30. Un programa de computación para poner en práctica el método de acuerdo con la reivindicación 33 al ejecutarse el programa de computación en una computadora.

31. Un decodificador de señales de audio (100; 200; 500; 590) para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente (1 12; 210; 510; 510a), una información paramétrica relacionada con los objetos (1 10; 2 12; 5 12; 5 12a) donde el decodificador de señales de audio comprende:

un separador de objetos (130; 260; 520; 520a) configurado para descomponer una representación de señal de mezcla descendente, para producir una primera información de audio (132; 262; 562; 562a) que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio, y una segunda información de audio (134; 264; 564; 564a) que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de señal de mezcla descendente y usando al menos una parte de la información paramétrica relacionada con los objetos,

un procesador de señales de audio configurado para recibir la segunda información de audio (134; 264; 564; 564a) y para procesar la segunda información de audio que depende de la información paramétrica relacionada con los objetos, para obtener una versión procesada (142; 272; 572; 572a) de la segunda información de audio y

un combinador de señales de audio (150; 280; 580; 580a) configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente;

donde el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

donde

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C},$$

donde

$$\mathbf{M}^{Prediction} = \begin{pmatrix} \mathbf{M}_{OBJ}^{Prediction} \\ \hline \mathbf{M}_{EAO}^{Prediction} \end{pmatrix}$$

donde $\mathbf{X}_{OBJ}$ representa canales de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objetos de la primera información de audio;

donde $\tilde{D}^{-1}$ representa una matriz que es un inverso de una matriz de mezcla descendente extendida;

donde C describe una matriz que representa una pluralidad de coeficientes del canal de predicción, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

donde l_0 y r_0 representan canales de la representación de señal de mezcla descendente;

donde $\mathbf{res}_0$ a $\mathbf{res}_{N_{EAO}-1}$ representan canales residuals; y

donde $\mathbf{A}^{EAO}$ es una matriz EAO de pre-renderización, las entradas de la cual describen un mapeo de objetos de audio mejorados para canales de una señal de audio $\mathbf{X}_{EAO}$ mejorada;

donde el separador de objetos está configurado para obtener la matriz inversa de mezcla descendente $\tilde{D}^{-1}$ como una inversa de una matriz de mezcla descendente $\tilde{D}$ la cual se define como:

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

donde el separador de objetos está configurado para obtener la matriz C como

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

donde m_0 a $m_{N_{EAO}-1}$ son valores de mezcla descendente Asociados con los objetos de audio del primer tipo de objetos de audio;

donde n_0 a $n_{N_{EAO}-1}$ son valores de mezcla descendente Asociados con los objetos de audio del primer tipo de objetos de audio;

donde el separador de objetos está configurado para computar los coeficientes de predicción $\tilde{c}_{j,0}$ y $\tilde{c}_{j,1}$ como:

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ y}$$

donde el separador de objetos está configurado para derivar coeficientes de predicción restringidos $c_{j,0}$ y $c_{j,1}$ a partir de los coeficientes de predicción $\tilde{c}_{j,0}$ y $\tilde{c}_{j,1}$ usando un algoritmo de restricción, o el uso de los coeficientes de predicción $\tilde{c}_{j,0}$ y $\tilde{c}_{j,1}$ como los coeficientes de predicción $c_{j,0}$ y $c_{j,1}$;

donde las cantidades de energía P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ y $P_{RoCo,j}$ se definen como

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i\neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i\neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

Donde los parámetros OLD_L, OLD_R y $IOC_{L,R}$ corresponden a objetos de audio del segundo tipo de objetos de audio y son definidos de acuerdo con

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i \,,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i \,,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

Donde $d_{0,i}$ y $d_{1,i}$ son valores de mezcla descendente Asociados con los objetos de audio del segundo tipo de objetos de audio;

donde OLD_i son objetos de diferente nivel de valores Asociados con los objetos del Segundo tipo de objetos de audio;

donde N es un número total de objetos de audio;

donde N_{EAO} es un numero de objetos de audio del primer tipo de objetos de audio;

donde $IOC_{0,1}$ es un valor de correlación inter objetos asociado con un par de objetos de audio del segundo tipo de objetos de audio;

donde $e_{i,j}$ y $e_{L,R}$ son valores de covarianza derivados de parámetros de objetos de diferente nivel y parámetros de correlación inter objetos; y

donde $e_{i,j}$ están asociados con un par de objetos de audio del primer tipo de objetos de audio y $e_{L,R}$ está asociado con un par de objetos de audio del Segundo tipo de objetos de audio.

32. Un decodificador de señales de audio (100; 200; 500; 590) para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente (1 12; 210; 510; 510a), una información paramétrica relacionada con los objetos (1 10; 2 12; 5 12; 5 12a) donde el decodificador de señales de audio comprende:

un separador de objetos (130; 260; 520; 520a) configurado para descomponer una representación de señal de mezcla descendente, para producir una primera información de audio (132; 262; 562; 562a) que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio, y una segunda información de audio (134; 264; 564; 564a) que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de señal de mezcla descendente y usa al menos un parte de la información paramétrica relacionada con los objetos.

un procesador de señales de audio configurado para recibir la segunda información de audio (134; 264; 564; 564a) y para procesar la segunda información de audio que depende de la información paramétrica relacionada con los objetos, para obtener una versión procesada (142; 272; 572; 572a) de la segunda información de audio y

un combinador de señales de audio (150; 280; 580; 580a) configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente;

donde el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

donde $\mathbf{X}_{OBJ}$ representa canales de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objetos de la primera información de audio;

donde

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

donde m_0 a m_{NEAO-1} son valores de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio;

donde n_0 a $n_{N_{EAO}-1}$ son valores de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio;

donde OLD_i son objetos de diferentes niveles de valor asociados con los objetos de audio del primer tipo de objetos de audio;

donde OLD_L y OLD_R son objetos comunes de valores de diferentes niveles Asociados con los objetos de audio del segundo tipo de objetos de audio; y

donde $\mathbf{A}^{EAO}$ es una matriz EAO de pre-renderización.

33. Un decodificador de señales de audio (100; 200; 500; 590) para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente (1 12; 210; 510; 510a), una

información paramétrica relacionada con los objetos (1 10; 2 12; 5 12; 5 12a) donde el decodificador de señales de audio comprende:

un separador de objetos (130; 260; 520; 520a) configurado para descomponer una representación de señal de mezcla descendente, para producir una primera información de audio (132; 262; 562; 562a) que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio, y una segunda información de audio (134; 264; 564; 564a) que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de señal de mezcla descendente y usa al menos una parte de la información paramétrica relacionada con los objetos,

un procesador de señales de audio configurado para recibir la segunda información de audio (134; 264; 564; 564a) y para procesar la segunda información de audio que depende de la información paramétrica relacionada con los objetos, para obtener una versión procesada (142; 272; 572; 572a) de la segunda información de audio y

un combinador de señales de audio (150; 280; 580; 580a) configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente;

donde el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}}\left(d_0\right)$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Energy}}\left(d_0\right)$$

donde $\mathbf{X}_{OBJ}$ representa un canal de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objetos de la primera información de audio;

donde

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \left(\sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}}\right)$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\dfrac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\dfrac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

donde m_0 a m_{NEAO-1} son valores de mezcla descendente Asociados con los objetos de audio del primer tipo de objetos de audio;

donde OLD_i son objetos de valores de diferente nivel asociados con los objetos de audio del primer tipo de objetos de audio;

donde OLD_L es un objeto común de valor de diferente nivel asociado con los objetos de audio del segundo tipo de objetos de audio; y

donde $\mathbf{A}^{EAO}$ es una matriz EAO de pre-renderización;

donde las matrices $\mathbf{M}_{OBJ}^{Energy}$ y $\mathbf{M}_{EAO}^{Energy}$ son aplicadas a una representación d_0 de una señal simple de mezcla descendente SAOC.

34. Un método para producir una representación de señal de mezcla ascendente que depende de una representación de la señal de mezcla descendente e información paramétrica relacionada con los objetos, método que comprende:

descomponer la representación de señal de mezcla descendente, para producir una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de la señal de a descendente y utilizando por lo menos parte de la información paramétrica relacionada con los objetos; y

procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, a fin de obtener una versión procesada de la segunda información de audio y

combinar la primera información de audio con la versión procesada de la segunda información de audio para obtener la representación de señal de mezcla ascendente;

donde la primera información de audio y la segunda información de audio son obtenidas de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

donde

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C},$$

donde

$$\mathbf{M}^{Prediction} = \begin{pmatrix} \mathbf{M}_{OBJ}^{Prediction} \\ \hline \mathbf{M}_{EAO}^{Prediction} \end{pmatrix}$$

donde $\mathbf{X}_{OBJ}$ representa canales de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objetos de la primera información de audio;

donde $\tilde{D}^{-1}$ representa una matriz que es la inversa de una matriz de mezcla descendente extendida;

donde C describe a matriz que representa una pluralidad de coeficientes de predicción de canal, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

donde l_0 y r_0 representa canales de la representación de señal de mezcla descendente;

donde $\mathbf{res}_0$ a $\mathbf{res}_{N_{EAO}-1}$ representan canales residuales; y

donde $\mathbf{A}^{EAO}$ es una matriz EAO de pre-renderización, de la cual las entradas describen un mapeo de objetos de audio mejorados para canales de señal de audio $\mathbf{X}_{EAO}$ mejorada;

donde la inversa de la matriz de mezcla descendente $\tilde{D}^{-1}$ es obtenida como una inversa de una matriz de mezcla descendente $\tilde{D}$ la cual es definida como:

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

Donde la matriz C es obtenida como

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

donde m_0 a $m_{N_{EAO}-1}$ son valores de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio;

donde n_0 a $n_{N_{EAO}-1}$ son valores de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio;

donde los coeficientes de predicción $\tilde{c}_{j,0}$ y $\tilde{c}_{j,1}$ son computados como

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ y}$$

Donde los coeficientes de predicción restringidos $c_{j,0}$ y $c_{j,1}$ son derivados de los coeficientes de predicción $\tilde{c}_{j,0}$ y $\tilde{c}_{j,1}$ usando un algoritmo de restricción, o donde los coeficientes de predicción $\tilde{c}_{j,0}$ y $\tilde{c}_{j,1}$ son usados como los coeficientes de predicción $c_{j,0}$ y $c_{j,1}$;

Donde las cantidades de energía P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ y $P_{RoCo,j}$ son definidas como

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

donde los parámetros OLD_L, OLD_R y $IOC_{L,R}$ corresponden a objetos de audio del Segundo tipo de objetos de audio y son definidos de acuerdo con

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

donde $d_{0,i}$ y $d_{1,i}$ son valores de mezcla descendente Asociados con los objetos de audio del Segundo tipo de objetos de audio;

donde OLD_i son objetos de valores de diferente nivel Asociados con los objetos de de audio del Segundo tipo de objetos de audio;

donde N es un número total de objetos de audio;

donde N_{EAO} es un numero de objetos de audio del primer tipo de objetos de audio;

donde $IOC_{0,1}$ es un valor de correlación inter asociado con un par de objetos de audio del segundo tipo de objetos de audio;

donde $e_{i,j}$ y $e_{L,R}$ son valores de covarianza derivados de los parámetros del objeto de diferente nivel; y

donde $e_{i,j}$ está asociado con un par de objetos de audio del primer tipo de objetos de audio y $e_{L,R}$ esta asociado con un par de objetos de audio del segundo tipo de objetos de audio.

35. Un método para producir una representación de señal de mezcla ascendente que depende de una representación de la señal de mezcla descendente e información paramétrica relacionada con los objetos, método que comprende:

descomponer la representación de señal de mezcla descendente, para producir una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de la señal de a descendente y utilizando por lo menos parte de la información paramétrica relacionada con los objetos; y

procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, a fin de obtener una versión procesada de la segunda información de audio y

combinar la primera información de audio con la versión procesada de la segunda información de audio para obtener la representación de señal de mezcla ascendente;

donde la primera información de audio y la segunda información de audio son obtenidas de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

donde $\mathbf{X}_{OBJ}$ representa canales de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objetos de la primera información de audio;

donde

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\dfrac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\dfrac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

donde m_0 a $m_{N_{EAO}-1}$ son valores de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio;

donde n_0 a $n_{N_{EAO}-1}$ son valores de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio;

donde OLD_i son objetos de valores de diferente nivel asociados con los objetos de audio del primer tipo de objetos de audio;

donde OLD_L es un objeto común de valor de diferente nivel asociado con los objetos de audio del segundo tipo de objetos de audio; y

donde $\mathbf{A}^{EAO}$ es una matriz EAO de pre-renderización.

36. Un método para producir una representación de señal de mezcla ascendente que depende de una representación de la señal de mezcla descendente e información paramétrica relacionada con los objetos, método que comprende:

descomponer la representación de señal de mezcla descendente, para producir una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio dependiendo de la representación de la señal de a descendente y utilizando por lo menos parte de la información paramétrica relacionada con los objetos; y

procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, a fin de obtener una versión procesada de la segunda información de audio y

combinar la primera información de audio con la versión procesada de la segunda información de audio para obtener la representación de señal de mezcla ascendente;

donde la primera información de audio y la segunda información de audio son obtenidas de acuerdo con

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}}\left(d_0\right)$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Energy}}\left(d_0\right)$$

donde $\mathbf{X}_{OBJ}$ representa canales de la segunda información de audio;

donde $\mathbf{X}_{EAO}$ representa señales de objetos de la primera información de audio;

donde

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\dfrac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\dfrac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\dfrac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

donde m_0 a m_{NEAO-1} son mezcla de valores descendentes Asociados con los objetos de audio del primer tipo de objetos de audio;

donde OLD_i son objetos de valores de diferente nivel asociados con los objetos de audio del primer tipo de objetos de audio;

donde OLD_L es un nivel de objeto común de diferente valor asociado con los objetos de audio del Segundo tipo de objetos de audio; y

donde $\mathbf{A}^{EAO}$ es una matriz EAO pre-renderización;

donde las matrices $\mathbf{M}_{OBJ}^{Energia}$ y $\mathbf{M}_{EAO}^{Energia}$ son aplicadas a la representación d_0 de una sola señal de mezclado descendente SAOC.

37. Un programa de computación para poner en práctica el método de acuerdo con una de las reivindicaciones 34 a 36 al ejecutarse el programa de computación en una computadora.

Claims (clean version)

1. An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information,

wherein the second audio information is an audio information describing the audio objects of the second audio object type in a combined manner;

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the audio signal decoder is configured to provide the upmix signal representation in dependence on a residual information associated to a subset of audio objects represented by the downmix signal representation,

wherein the object separator is configured to decompose the downmix signal representation to provide the first audio information describing a first set of one or more audio objects of a first audio object type to which residual information is associated, and the second audio information describing a second set of one or more audio objects of a second audio object type, to which no residual information is associated, in dependence on the downmix signal representation and using the residual information; and

wherein the audio signal processor is configured to process the second audio information, to perform an object-individual processing of the audio objects of the second audio object type, taking into consideration object-related parametric information associated with more than two audio objects of the second audio object type; and

wherein the residual information describes a residual distortion, which is expected to remain if an audio object of the first audio object type is isolated merely using the object-related parametric information.

2. The audio signal decoder (100; 200; 500; 590) according to claim 1, wherein the object separator is configured to provide the first audio information such that one or more audio objects of the first audio object type are emphasized over audio objects of the second audio object type in the first audio information, and

wherein the object separator is configured to provide the second audio information such that audio objects of the second audio object type are emphasized over audio objects of the first audio object type in the second audio information.

3. The audio signal decoder (100; 200; 500; 570) according to one of claims 1 to 2, wherein the audio signal processor is configured to process the second audio information (134; 264; 564; 564a) in dependence on the object-related parametric information (110; 212; 512; 512a) associated with the audio objects of the second audio object type and independent from the object-related parametric information (110; 212; 512; 512a) associated with the audio objects of the first audio object type.

4. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 3, wherein the object separator is configured to obtain the first audio information (132; 262; 562; 562a, $\mathbf{X}_{EAO}$) and the second audio information (134; 264; 564; 564a, $\mathbf{X}_{OBJ}$) using a linear combination of one or more downmix signal channels of the downmix signal representation and one or more residual channels, wherein the object separator is configured to obtain combination parameters for performing the linear combination in dependence on downmix parameters associated with the audio objects of the first audio object type ($m_0 \ldots m_{NEAO-1}$; $n_0 \ldots n_{NEAO-1}$) and in dependence on channel prediction coefficients ($c_{j,0}$, $c_{j,1}$) of the audio objects of the first audio object type.

5. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 4, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C},$$

wherein

$$\mathbf{M}^{Prediction} = \begin{pmatrix} \mathbf{M}_{OBJ}^{Prediction} \\ \hline \mathbf{M}_{EAO}^{Prediction} \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\tilde{D}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein C describes a matrix representing a plurality of channel prediction coefficients, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

wherein l_0 and r_0 represent channels of the downmix signal representation;

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix, entries of which describe a mapping of enhanced audio objects to channels of an enhanced audio object signal $\mathbf{X}_{EAO}$;

wherein the object separator is configured to obtain the inverse downmix matrix $\tilde{D}^{-1}$ as an inverse of an extended downmix matrix $\tilde{D}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the object separator is configured to obtain the matrix C as

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein the object separator is configured to compute the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ and}$$

wherein the object separator is configured to derive constrained prediction coefficients $c_{j,0}$ and $c_{j,1}$ from the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ using a constraining algorithm, or to use the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as the prediction coefficients $c_{j,0}$ and $c_{j,1}$;

wherein energy quantities P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ and $P_{RoCo,j}$ are defined as

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

wherein parameters OLD_L, OLD_R and $IOC_{L,R}$ correspond to audio objects of the second audio object type and are defined according to

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

wherein $d_{0,i}$ and $d_{1,i}$ are downmix values associated with the audio objects of the second audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the second audio object type;

wherein N is a total number of audio objects;

wherein N_{EAO} is a number of audio objects of the first audio object type;

wherein $IOC_{0,1}$ is an inter-object-correlation value associated with a pair of audio objects of the second audio object type;

wherein $e_{i,j}$ and $e_{L,R}$ are covariance values derived from object-level-difference parameters and inter-object-correlation parameters; and

wherein $e_{i,j}$ are associated with a pair of audio objects of the 1st audio object type and $e_{L,R}$ is associated with a pair of audio objects of the second audio object type.

6. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 4, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} \frac{d_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} \frac{d_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C}$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\tilde{\mathbf{D}}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein $\mathbf{C}$ describes a matrix representing a plurality of channel prediction coefficients, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

wherein d_0 represents a channel of the downmix signal representation; and

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

7. The audio signal decoder according to claim 6, wherein the object separator is configured to obtain the inverse downmix matrix $\tilde{\mathbf{D}}^{-1}$ is an inverse of an extended downmix matrix $\tilde{\mathbf{D}}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{c|ccc} 1 & m_0 & \cdots & m_{N_{EAO}-1} \\ \hline m_0 & -1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the object separator is configured to obtain the matrix C as

$$\mathbf{C} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline c_0 & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1} & 0 & \cdots & 1 \end{array}\right);$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type.

8. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 4, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L and OLD_R are common object level difference values associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

9. The audio signal decoder according to one of claims 1 to 3, wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} d_0$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} d_0$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \left(\sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right)$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L is a common object level difference value associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix;

wherein the matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ are applied to a representation d_0 of a single SAOC downmix signal.

10. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 9, wherein the object separator is configured to apply a rendering matrix to the first audio information (132; 262; 562; 562a) to map object signals of the first audio information onto audio channels of the upmix audio signal representation (120; 220, 222; 562; 562a).

11. The audio signal decoder (100; 200; 500; 590) according to one of claims 1 to 10, wherein the audio signal processor (140; 270; 570; 570a) is configured to perform a stereo preprocessing of the second audio information (134; 264; 564; 564a) in dependence on a rendering information (M_{ren}), an object-related covariance information (**E**), a downmix information (**D**), to obtain audio channels of the processed version of the second audio information;

12. The audio signal decoder (100; 200; 500; 590) according to claim 11, wherein the audio signal processor (140; 270; 570; 570a) is configured to perform the stereo processing to map an estimated audio object contribution (ED^*JX) of the second audio information (134; 264; 564; 564a) onto a plurality of channels of the upmix audio signal representation in dependence on a rendering information and a covariance information.

13. The audio signal decoder according to claim 11 or claim 12, wherein the audio signal processor is configured to add a decorrelated audio signal contribution (P_2X_d), obtained on the basis of one or more audio channels of the second audio information, to the second audio information, or an information derived from the second audio information, in dependence on a render upmix error information (R) and one or more decorrelated-signal-intensity scaling values (w_{d1}, w_{d2}).

14. The audio signal decoder according to one of claims 1 to 10, wherein the audio signal processor (140; 270; 570; 570a) is configured to perform a postprocessing of the second audio information (134; 264; 564; 564a) in dependence on a rendering information (A), an object-related covariance information (E) and a downmix information (D).

15. The audio signal decoder according to claim 14, wherein the audio signal processor is configured to perform a mono-to-binaural processing of the second audio information, to map a single channel of the second audio information onto two channels of the upmix signal representation, taking into consideration a head-related transfer function.

16. The audio signal decoder according to claim 14, wherein the audio signal processor is configured to perform a mono-to-stereo processing of the second audio information, to map a single channel of the second audio information onto two channels of the upmix signal representation.

17. The audio signal decoder according to claim 14, wherein the audio signal processor is configured to perform a stereo-to-binaural processing of the second audio information, to map two channels of the second audio information onto two channels of the upmix signal representation, taking into consideration a head-related transfer function.

18. The audio signal decoder according to claim 14, wherein the audio signal processor is configured to perform a stereo-to-stereo processing of the second audio information, to map two channels of the second audio information onto two channels of the upmix signal representation.

19. The audio signal decoder according to one of claims 1 to 18, wherein the object separator is configured to treat audio objects of the second audio object type, to which no residual information is associated, as a single audio object, and

wherein the audio signal processor (140; 270; 570; 570a) is configured to consider object-specific rendering parameters associated to the audio objects of the second audio object type to adjust contributions of the audio objects of the second audio object type to the upmix signal representation.

20. The audio signal decoder according to one of claims 1 to 19, wherein the object separator is configured to obtain one or two common object level difference values (OLD_L, OLD_R) for a plurality of audio objects of the second audio object type; and

wherein the object separator is configured to use the common object level difference value for a computation of channel prediction coefficients (CPC); and

wherein the object separator is configured to use the channel prediction coefficients to obtain one or two audio channels representing the second audio information.

21. The audio signal decoder according to one of claims 1 to 20, wherein the object separator is configured to obtain one or two common object level difference values (OLD_L, OLD_R) for a plurality of audio objects of the second audio object type, and

wherein the object separator is configured to use the common object level difference value for a computation of entries of an matrix (**M**); and

wherein the object separator is configured to use the matrix (**M**) to obtain one or more audio channels representing the second audio information.

22. The audio signal decoder according to one of claims 1 to 21, wherein the object separator is configured to selectively obtain a common inter-object correlation value ($IOC_{L,R}$) associated to the audio object of the second audio object type in dependence on the object-related parametric information if it is found that there are two audio objects of the second audio object type, and to set the inter-object correlation value associated to the audio objects of the second audio object type to zero if it is found that there are more or less than two audio objects of the second audio object type; and

wherein the object separator is configured to use the common inter-object correlation value for a computation of entries of an matrix (**M**); and

wherein the object separator is configured to use the common inter-object correlation value associated to the audio objects of the second audio object type to obtain the one or more audio channels representing the second audio information.

23. The audio signal decoder according to one of claims 1 to 22, wherein the audio signal processor is configured to render the second audio information in dependence on the object-related parametric information, to obtain a rendered representation of the audio objects of the second audio object type as the processed version of the second audio information.

24. The audio signal decoder according to one of claims 1 to 23, wherein the object separator is configured to provide the second audio information such that the second audio information describes more than two audio objects of the second audio object type.

25. The audio signal decoder according to claim 24, wherein the object separator is configured to obtain, as the second audio information, a one-channel audio signal representation or a two-channel audio signal representation representing more than two audio objects of the second audio object type.

26. The audio signal decoder according to one of claims 1 to 25, wherein the audio signal processor is configured to receive the second audio information and to process the second audio information in dependence of the object-related parametric information, taking into consideration object-related parametric information associated with more than two audio objects of the second audio object type.

27. The audio signal decoder according to one of claims 1 to 26, wherein the audio signal decoder is configured to extract a total object number information (bsNumObjects) and a foreground object number information (bsNumGroupsFGO) from a configuration information (SAOCSpecificConfig) of the object-related parametric information, and to determine the number of audio objects of the second audio object type by forming a difference between the total object number information and the foreground object number information.

28. The audio signal decoder according to one of claims 1 to 27, wherein the object separator is configured to use object-related parametric information associated with N_{EAO} audio objects of the first audio object type to obtain, as the first audio information, N_{EAO} audio signals ($\mathbf{X}_{EAO}$) representing the N_{EAO} audio objects of the first audio object type and to obtain, as the second audio information, one or two audio signals ($\mathbf{X}_{ORI}$) representing the N-N_{EAO} audio objects of the second audio object type, treating the N-N_{EAO} audio objects of the second audio object type as a single one-channel or a two-channel audio object; and

wherein the audio signal processor is configured to individually render the N-N_{EAO} audio objects represented by the one or two audio signals of the second audio information using the object-related parametric information associated with the N-N_{EAO} audio objects of the second audio object type.

29. A method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, the method comprising:

decomposing the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type, and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information, wherein the second audio information is an audio information describing the audio objects of the second audio object type in a combined manner; and

processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information; and

combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the upmix signal representation is provided in dependence on a residual information associated to a subset of audio objects represented by the downmix signal representation,

wherein the downmix signal representation is decomposed, to provide the first audio information describing a first set of one or more audio objects of a first audio object type to which residual information is associated, and the second audio information describing a second set of one or more audio objects of a second audio object type, to which no residual information is associated, in dependence on the downmix signal representation and using the residual information;

wherein an object-individual processing of the audio objects of the second audio object type is performed, taking into consideration object-related parametric information associated with more than two audio objects of the second audio object type; and

wherein the residual information describes a residual distortion, which is expected to remain if an audio object of the first audio object type is isolated merely using the object-related parametric information.

30. A computer program for performing the method according to claim 33 when the computer program runs on a computer.

31. An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information;

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C},$$

wherein

$$\mathbf{M}^{Prediction} = \left(\frac{\mathbf{M}_{OBJ}^{Prediction}}{\mathbf{M}_{EAO}^{Prediction}}\right)$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\tilde{D}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein C describes a matrix representing a plurality of channel prediction coefficients, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

wherein l_0 and r_0 represent channels of the downmix signal representation;

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix, entries of which describe a mapping of enhanced audio objects to channels of an enhanced audio object signal $\mathbf{X}_{EAO}$;

wherein the object separator is configured to obtain the inverse downmix matrix $\tilde{D}^{-1}$ as an inverse of an extended downmix matrix $\tilde{D}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the object separator is configured to obtain the matrix C as

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein the object separator is configured to compute the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ and}$$

wherein the object separator is configured to derive constrained prediction coefficients $c_{j,0}$ and $c_{j,1}$ from the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ using a constraining algorithm, or to use the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ as the prediction coefficients $c_{j,0}$ and $c_{j,1}$;

wherein energy quantities P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ and $P_{RoCo,j}$ are defined as

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1}\sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i\neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

wherein parameters OLD_L, OLD_R and $IOC_{L,R}$ correspond to audio objects of the second audio object type and are defined according to

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

wherein $d_{0,i}$ and $d_{1,i}$ are downmix values associated with the audio objects of the second audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the second audio object type;

wherein N is a total number of audio objects;

wherein N_{EAO} is a number of audio objects of the first audio object type;

wherein $IOC_{0,1}$ is an inter-object-correlation value associated with a pair of audio objects of the second audio object type;

wherein $e_{i,j}$ and $e_{L,R}$ are covariance values derived from object-level-difference parameters and inter-object-correlation parameters; and

wherein $e_{i,j}$ are associated with a pair of audio objects of the 1st audio object type and $e_{L,R}$ is associated with a pair of audio objects of the second audio object type.

32. An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information;

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L and OLD_R are common object level difference values associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

33. An audio signal decoder (100; 200; 500; 590) for providing an upmix signal representation in dependence on a downmix signal representation (112; 210; 510; 510a), an object-related parametric information (110; 212; 512; 512a) the audio signal decoder comprising:

an object separator (130; 260; 520; 520a) configured to decompose the downmix signal representation, to provide a first audio information (132; 262; 562; 562a) describing a first set of one or more audio objects of a first audio object type, and a second audio information (134; 264; 564; 564a) describing a second set of one or

more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information;

an audio signal processor configured to receive the second audio information (134; 264; 564; 564a) and to process the second audio information in dependence on the object-related parametric information, to obtain a processed version (142; 272; 572; 572a) of the second audio information; and

an audio signal combiner (150; 280; 580; 580a) configured to combine the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the object separator is configured to obtain the first audio information and the second audio information according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} d_0$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} d_0$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\dfrac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\dfrac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\dfrac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L is a common object level difference value associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix;

wherein the matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ are applied to a representation d_0 of a single SAOC downmix signal.

34. A method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, the method comprising:

decomposing the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type, and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information; and

processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information; and

combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the first audio information and the second audio information are obtained according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO}\mathbf{M}_{EAO}^{\text{Prediction}} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

wherein

$$\mathbf{M}_{\text{Prediction}} = \tilde{\mathbf{D}}^{-1}\mathbf{C},$$

wherein

$$\mathbf{M}^{Prediction} = \begin{pmatrix} \mathbf{M}_{OBJ}^{Prediction} \\ \hline \mathbf{M}_{EAO}^{Prediction} \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein $\tilde{D}^{-1}$ represents a matrix which is an inverse of an extended downmix matrix;

wherein C describes a matrix representing a plurality of channel prediction coefficients, $\tilde{c}_{j,0}$, $\tilde{c}_{j,1}$;

wherein l_0 and r_0 represent channels of the downmix signal representation;

wherein $\mathbf{res}_0$ to $\mathbf{res}_{N_{EAO}-1}$ represent residual channels; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix, entries of which describe a mapping of enhanced audio objects to channels of an enhanced audio object signal $\mathbf{X}_{EAO}$;

wherein the inverse downmix matrix $\tilde{D}^{-1}$ is obtained as an inverse of an extended downmix matrix $\tilde{D}$ which is defined as

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \cdots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \cdots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \cdots & -1 \end{array}\right)$$

wherein the matrix C is obtained as

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{array}\right)$$

wherein m_0 to $m_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ are computed as

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2}; \text{ and}$$

wherein constrained prediction coefficients $c_{j,0}$ and $c_{j,1}$ are derived from the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ using a constraining algorithm, or wherein the prediction coefficients $\tilde{c}_{j,0}$ and $\tilde{c}_{j,1}$ are used as the prediction coefficients $c_{j,0}$ and $c_{j,1}$;

wherein energy quantities P_{Lo}, P_{Ro}, P_{LoRo}, $P_{LoCo,j}$ and $P_{RoCo,j}$ are defined as

$$P_{L,L} = OLD_L + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{R,R} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{L,R} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{L,Co,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{R,Co,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

wherein parameters OLD_L, OLD_R and $IOC_{L,R}$ correspond to audio objects of the second audio object type and are defined according to

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & otherwise. \end{cases}$$

wherein $d_{0,i}$ and $d_{1,i}$ are downmix values associated with the audio objects of the second audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the second audio object type;

wherein N is a total number of audio objects;

wherein N_{EAO} is a number of audio objects of the first audio object type;

wherein $IOC_{0,1}$ is an inter-object-correlation value associated with a pair of audio objects of the second audio object type;

wherein $e_{i,j}$ and $e_{L,R}$ are covariance values derived from object-level-difference parameters and inter-object-correlation parameters; and

wherein $e_{i,j}$ are associated with a pair of audio objects of the 1st audio object type and $e_{L,R}$ is associated with a pair of audio objects of the second audio object type.

35. A method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, the method comprising:

decomposing the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object type, and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information; and

processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information; and

combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the first audio information and the second audio information are obtained according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

wherein $\mathbf{X}_{OBJ}$ represent channels of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

wherein n_0 to $n_{N_{EAO}-1}$ are downmix values associated with the audio objects of the first audio object type;

wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

wherein OLD_L and OLD_R are common object level difference values associated with the audio objects of the second audio object type; and

wherein $\mathbf{A}^{EAO}$ is a EAO pre-rendering matrix.

36. A method for providing an upmix signal representation in dependence on a downmix signal representation and an object-related parametric information, the method comprising:

decomposing the downmix signal representation, to provide a first audio information describing a first set of one or more audio objects of a first audio object

type, and a second audio information describing a second set of one or more audio objects of a second audio object type in dependence on the downmix signal representation and using at least a part of the object-related parametric information; and

processing the second audio information in dependence on the object-related parametric information, to obtain a processed version of the second audio information; and

combining the first audio information with the processed version of the second audio information, to obtain the upmix signal representation;

wherein the first audio information and the second audio information are obtained according to

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energy}} d_0$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energy}} d_0$$

wherein $\mathbf{X}_{OBJ}$ represents a channel of the second audio information;

wherein $\mathbf{X}_{EAO}$ represent object signals of the first audio information;

wherein

$$\mathbf{M}_{OBJ}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{\text{Energy}} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

wherein m_0 to m_{NEAO-1} are downmix values associated with the audio objects of the first audio object type;

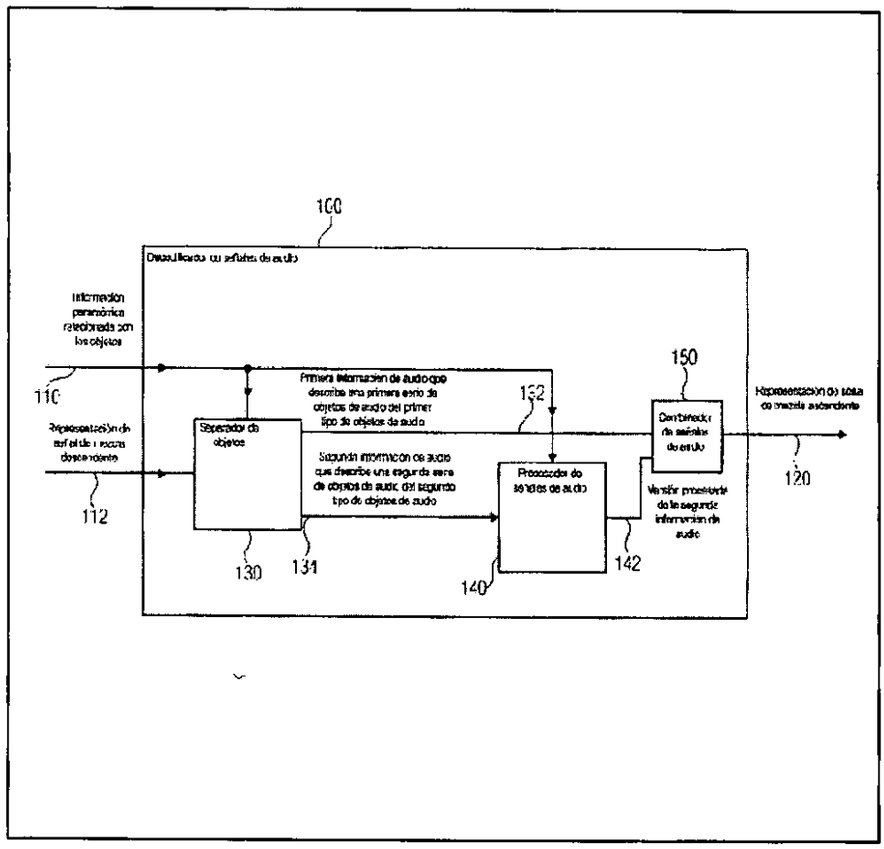
wherein OLD_i are object level difference values associated with the audio objects of the first audio object type;

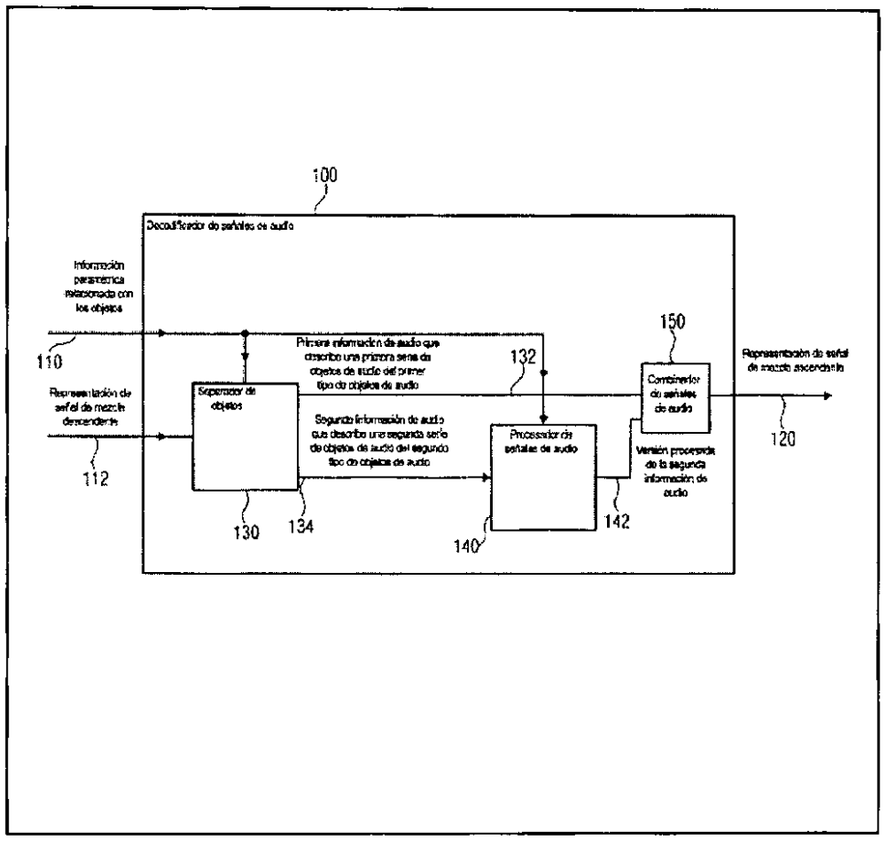
wherein OLD_L is a common object level difference value associated with the audio objects of the second audio object type; and

wherein A^{EAO} is a EAO pre-rendering matrix;

wherein the matrices $\mathbf{M}_{OBJ}^{Energy}$ and $\mathbf{M}_{EAO}^{Energy}$ are applied to a representation d_0 of a single SAOC downmix signal.

37. A computer program for performing the method according to one of claims 34 to 36 when the computer program runs on a computer.


100
Información paramétrica relacionada con los objetos
110
Representación de señal de mezcla descendente
112
Separador de objetos
Primera información de audio que describe una primera serie de objetos de audio del primer tipo de objetos de audio
132
Segunda información de audio que describe una segunda serie de objetos de audio del segundo tipo de objetos de audio
130
134
Procesador de señales de audio
140
142
150
Combinador de señales de audio
Versión procesada de la segunda información de audio
120
Representación de señal de mezcla ascendente


100
Información paramétrica relacionada con los objetos
110
Representación de señal de mezcla descendente
112
Separador de objetos
Primera información de audio que describe una primera serie de objetos de audio del primer tipo de objetos de audio
132
Segunda información de audio que describe una segunda serie de objetos de audio del segundo tipo de objetos de audio
130
134
Procesador de señales de audio
140
142
150
Combinador de señales de audio
Versión procesada de la segunda información de audio
120
Representación de señal de mezcla ascendente

Industria y Comercio
SUPERINTENDENCIA

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
No. 11-177816- -00000-0000
Fecha: 2011-12-23 14:58:29 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 379 FASENACIONALII
Act. 411 PRESENTACION Folios: 349

PATENTE DE INVENCIÓN ☐ MODELO DE UTILIDAD ☐ Art 33 Decisión 486/00

- ☐ Indicación que se solicita una patente.
- ☐ Datos de identificación del solicitante o de la persona que presenta la solicitud
- ☐ Descripción de la invención
- ☐ Dibujos de ser estos pertinentes
- ☐ Comprobante de pago de las tasas establecidas (De ser el caso formato de descuento)

Completa ☐ Incompleta ☐

PATENTE DE INVENCIÓN PCT ☑ MODELO DE UTILIDAD PCT ☐ Art.33 Decisión 486/00, Circular Única

- ☑ Indicación que se solicita una PCT
- ☑ Copia de la solicitud en español, tal como fue presentada inicialmente (capítulo descriptivo, reivindicatorio, resumen)
- ☐ Dibujos de ser estos pertinentes
- ☑ Comprobante de pago de las tasas establecidas (de ser el caso formato de descuento)

Completa ☐ Incompleta ☐

DISEÑO INDUSTRIAL ☐ (Art. 119 Decisión 486/00)

- ☐ Indicación que se solicita Diseño industrial
- ☐ Datos de identificación del solicitante o de la persona que presenta la solicitud
- ☐ Representación gráfica y fotográfica del Diseño industrial o muestra del material que incorpora el diseño
- ☐ Comprobante de pago de las tasas establecidas

Completa ☐ Incompleta ☐

ESQUEMA DE TRAZADO ☐ (Art. 92 Decisión 486/00)

- ☐ Indicación que se solicita un esquema de trazado
- ☐ Datos de identificación del solicitante o de la persona que presenta la solicitud
- ☐ Representación gráfica de un esquema de trazado
- ☐ Comprobante de pago de las tasas establecidas

Completa ☐ Incompleta ☐