

PCT

Industria y
SUPERINTE

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712-00000-0000

Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR. NUEVASCR
Tra. 11 PATENTE IYII Eve: 378 FASE NACIONAL I
Act. 411 PRESENTACION Folios: 227

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712-00000-0000

Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR. NUEVASCR
Tra. 11 PATENTE IYII Eve: 378 FASE NACIONAL I
Act. 411 PRESENTACION Folios: 257



DELEGATURA PROPIEDAD INDUSTRIAL

SOLICITUD
REGISTRO DE PATENTE DE INVENCION

PCT

EXPEDIENTE No

11-55712

TITULO ESQUEMA DE
CODIFICACIÓN/DECODIFICACIÓN DE AUDIO
CONMUTADO DE RESOLUCIÓN MÚLTIPLE

CLASIFICACION INTERNACIONAL _____

SOLICITANTE 1. FRAUNHOFER-GESELLSCHAFT
ZUR FORDERUNG DER ANGEWANDTEN
FORSCHUNG E.V.
2. VOICEAGE CORPORATION

DOMICILIO 1. MUNICH, ALEMANIA
2. QUEBEC, ALEMANIA

APODERADO DARIO CARDENAS NAVAS

BOGOTA D.C. _____

08-05-2011



No. 11-055712-00000-0000

Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 227

**FORMULARIO UNICO DE
PATENTE DE INVENCIÓN
ENTRADA EN FASE**

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712-00000-0000

Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 257

①

SOLICITUD

☒ Patente de Invención

☒ Capítulo I

☐ Capítulo II

②

SOLICITANTE

Nombre: 1. Fraunhofer-Gesellschaft zur Förderung
der angewandten Forschung e.V.
2. Voiceage Corporation
Dirección: 1. Hansastraße 27c 80686 Munich,
Alemania
2. 750 Lucerne Road, Suite 250, Montreal,
Québec H3R 2H6, Canada
Nacionalidad o Domicilio: 1. Munich, Alemania
2. Quebec, Canadá
Lugar de Constitución: 1. Alemania
2. Canadá
Teléfono: 3137800 Fax: 3122410
E-mail: general@cardenasycardenas.com

IDENTIFICACION

☐ NIT ☐ NIT

☐ Otro ☒ Otro

Cuál
Número

③

**REPRESENTANTE
O APODERADO**

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

IDENTIFICACION

☐ NIT ☐ NIT

☒ CC ☐ Otro

Cual 17066629 BTA
Número 1.250 C.S.J.

④

**INVENTOR(ES)
(72)**

Nombre: VER ANEXO
Dirección: VER ANEXO
Nacionalidad o Domicilio: VER ANEXO

⑤ **PCT (86)**

Solicitud Internacional No. WO 2010/040522 A2 Fecha 15.04.2010
Publicación Internacional No. PCT/EP2009/007205 Fecha 07.10.2009

⑥ **Título(54):**

**ESQUEMA DE CODIFICACIÓN/DECODIFICACIÓN DE AUDIO CONMUTADO
DE RESOLUCIÓN MÚLTIPLE**

⑦ **Clasificación Internacional (51):**

⑧ **Prioridad**

SI ☒ NO ☐

(33) País de Origen

(32) Fecha

(31) N° de Solicitud

US

08.10.2008

61/103,825

EP

08.10.2008

08017663.9

EP

10.02.2009

09002271.6

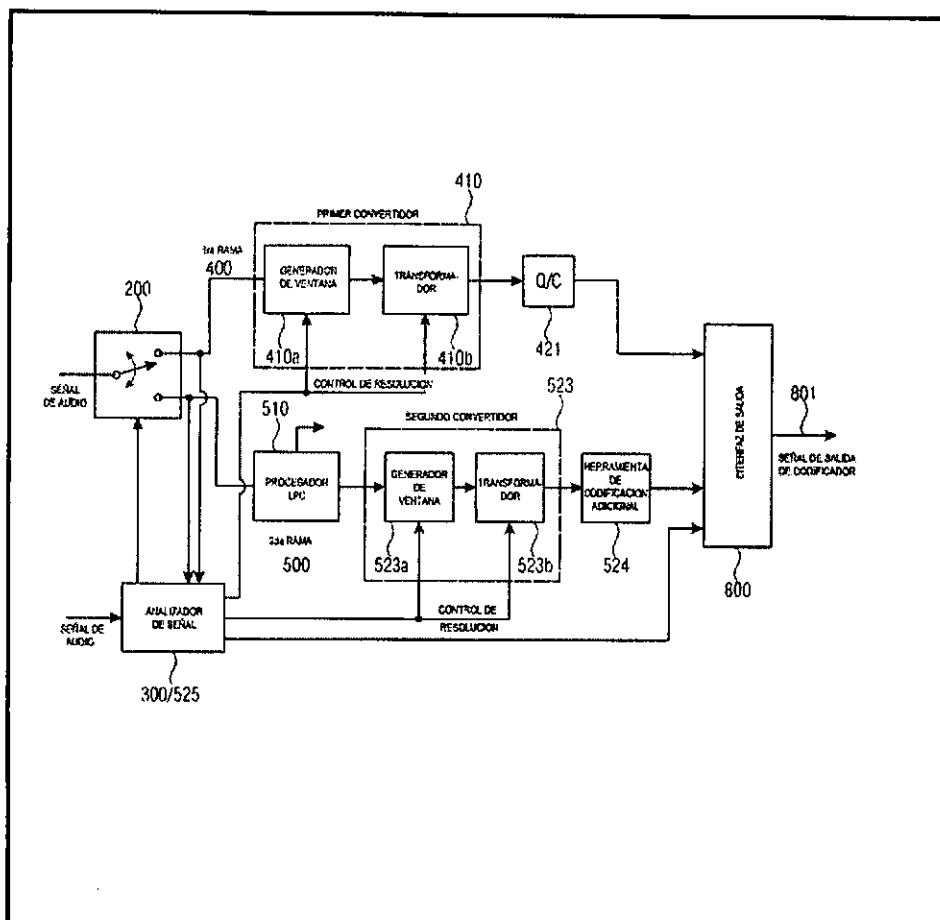
⑨ **Comprobante de Pago No.:** 11-0037528

Fecha 20.04.2011

10 ANEXOS

- ☒ Comprobante de pago de la tasa de presentación de la solicitud.
- ☒ Documento que acredite la existencia y representación legal cuando el solicitante sea persona jurídica.
- ☒ Poderes, si fuere el caso.
- ☐ Documento de cesión del inventor al solicitante o a su causante.
- ☒ Traducción de la solicitud internacional al castellano (Resumen, descripción, reivindicaciones, dibujos)
- ☐ 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.
- ☒ 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.
- ☒ Prioridad <http://www.wipo.int/pctdb/en/wo.jsp?WO=2010040522&1A=EP2009007205&DISPLAY=DOCS>
- ☒ Recibo de Reivindicaciones adicionales

11 FIGURA CARACTERISTICA



12 Solicito la Concesión de la patente

NOMBRE: **DARIO CARDENAS NAVAS**

FIRMA:

[Firma manuscrita]

C.C. 17066629 BTA TP. 1.250 C.S.J. **LHA**

④ INVENTOR(ES) (72)	Nombre:	1. Max Neuendorf / 2. Stefan Bayer / 3. Jérémie Lecomte / 4. Guillaume Fuchs / 5. Julien Robilliard / 6. Nikolaus Rettelbach / 7. Frederik Nagel / 8. Ralf Geiger / 9. Markus Multrus / 10. Bernhard Grill / 11. Philippe Gournay 12. Redwan Salami
	Dirección:	1. Kreilingstr.15, 90408 Nürnberg , Alemania 2. Johannisstr.148, 90419 Nürnberg, Alemania 3. Sulzbacher Strasse 39, 90489 Nürnberg, Alemania 4. Aussere Brucker Strasse Apt. West, 91052 Erlangen, Alemania 5. Innerer Kleinreuther Weg 25A, 90408 Nürnberg, Alemania 6. Amorbacher Str.2 a, 90427 Nürnberg, Alemania 7. Wilhelmshavener Strase 72, 90425 Nürnberg, Alemania 8. Münzstr. 8C, 98693 Ilmenau, Alemania 9. Etzlaubweg 7, 90469 Nürnberg, Alemania 10. Peter-Hentlein-Str.7, 91207 Lauf, Alemania 11. 3012 rue du Sauvignon, Sherbrooke, Québec J1L 0A2, Canadá 12. 4045 Albert – Dreux Place, Saint – Laurent, Québec H4R 2Y3, Canadá
	Nacionalidad o Domicilio:	1. Alemán 2. Alemán 3. Francés 4. Francés 5. Francés 6. Alemán 7. Alemán 8. Alemán 9. Alemán 10. Alemán 11. Francés 12. Libanés

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO**DIVISION DE NUEVAS CREACIONES****TARJETA ARCHIVO PROPIETARIOS**PATENTE DE INVENCION ____ PCT X 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.
-Voiceage Corporation**(72) Inventor(es) **Max Neuendorf, Stefan Bayer, Jérémie Lecomte, Guillaume Fuchs, Julien Robilliard, Nikolaus Rettelbach, Frederik Nagel...**(74) Apoderado: **DARIO CARDENAS NAVAS**

(30) Prioridad	(31) No. Prioridad	(32) Fecha	(33) Pais
	61/103,825	08.10.2008	US
	08017663.9	08.10.2008	EP
	09002271.6	18.02.2009	EP

(85) Fecha inicio fase nacional: 08.05.2011

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

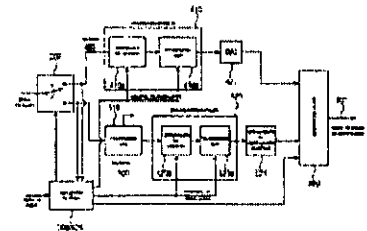
Fecha de presentación de la solicitud: 07.10.2009

No. de solicitud **PCT/EP2009/007205**(54) Título **ESQUEMA DE CODIFICACION/DECODIFICACIÓN DE AUDIO CONMUTADO DE RESOLUCIÓN MÚLTIPLE**

**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. -Voiceage Corporation			
(72) Inventor(es) Max Neuendorf, Stefan Bayer, Jérémie Lecomte, Guillaume Fuchs, Julien Robilliard, Nikolaus Rettelbach, Frederik Nagel...			
(74) Apoderado: DARIO CARDENAS NAVAS			
(30) Prioridad	(31) No. Prioridad	(32) Fecha	(33) País
	61/103,825	08.10.2008	US
	08017663.9	08.10.2008	EP
	09002271.6	18.02.2009	EP
(85) Fecha límite inicio fase nacional: 08.05.2011		(87) Publicación internacional:	
(86) Datos relativos a la presentación de la solíc. PCT		Fecha: 15.04.2010	
Fecha de presentación de la solicitud: 07.10.2009		No. Publicación WO 2010/040522	
No. de solicitud PCT/EP2009/007205			



(54) Título: **ESQUEMA DE CODIFICACION/DECODIFICACIÓN DE AUDIO CONMUTADO DE RESOLUCIÓN MÚLTIPLE**

(57) Resumen:

Un codificador de audio para codificar una señal de audio comprende una primera rama codificadora (400), donde la primera rama codificadora comprende un primer convertidor (410) para convertir una señal de un dominio del tiempo a un dominio de frecuencia. Más aun, el codificador de audio comprende una segunda rama codificadora (500) que comprende un segundo convertidor de tiempo/frecuencia (523). Además, se presenta un analizador de señales (300/525) para analizar la señal de audio. El analizador de señales, por un lado, determina si una porción de audio es eficaz en la señal de salida del codificador como primera señal codificada procedente de la primera rama codificadora o como segunda señal codificada de una segunda rama codificadora. Por otro lado, el analizador de señales determina una resolución en tiempo/frecuencia que han de aplicar los convertidores (410, 523) al generar las señales codificadas. Una interfaz de salida incluye, además de la primera señal codificada y la segunda señal codificada, una información de resolución que identifica la resolución empleada por el primer convertidor de tiempo/frecuencia y que es utilizada por el segundo convertidor de tiempo/frecuencia.



Industria y Comercio
SUPERINTENDENCIA

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

(21) N° de solicitud: (22) Fecha de solicitud:				(51) Int Cl: (71) Solicitante(s) -Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. -Voiceage Corporation	
(30) Prioridad	(31) No. Prioridad	(32) Fecha	(33) Pais	(72) Inventor(es) Max Neuendorf, Stefan Bayer, Jérémie Lecomte, Guillaume Fuchs, Julien Robilliard, Nikolaus Rettelbach, Fraderik Nagel...	
	61/103,825	08.10.2008	US		
	08017663.9	08.10.2008	EP		
	09002271.6	18.02.2009	EP	(74) Apoderado: DARIO CARDENAS NAVAS	
(85) Fecha limite inicio fase nacional: 08.05.2011				(87) Publicación internacional	
(86) Datos relativos a la presentación de la solicitud PCT				Fecha: 15.04.2010	
Fecha de presentación de la solicitud: 07.10.2009				N° publicación: WO 2010/040522 A2	
No. de solicitud PCT/EP2009/007205					
(54) Titulo: ESQUEMA DE CODIFICACION/DECODIFICACIÓN DE AUDIO CONMUTADO DE RESOLUCIÓN MÚLTIPLE					

Resumen:

Un codificador de audio para codificar una señal de audio comprende una primera rama codificadora (400), donde la primera rama codificadora comprende un primer convertidor (410) para convertir una señal de un dominio del tiempo a un dominio de frecuencia. Más aun, el codificador de audio comprende una segunda rama codificadora (500) que comprende un segundo convertidor de tiempo/frecuencia (523). Además, se presenta un analizador de señales (300/525) para analizar la señal de audio. El analizador de señales, por un lado, determina si una porción de audio es eficaz en la señal de salida del codificador como primera señal codificada procedente de la primera rama codificadora o como segunda señal codificada de una segunda rama codificadora. Por otro lado, el analizador de señales determina una resolución en tiempo/frecuencia que han de aplicar los convertidores (410, 523) al generar las señales codificadas. Una interfaz de salida incluye, además de la primera señal codificada y la segunda señal codificada, una información de resolución que identifica la resolución empleada por el primer convertidor de tiempo/frecuencia y que es utilizada por el segundo convertidor de tiempo/frecuencia.

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800.176.089-2

- / -

RECIBO DE CAJA

No. 11 - 0037528

Industria y Comercio

SUPERINTENDENCIA

Bogotá D.C., Abril 20 de 2011 - 09:22:32

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	430550444	19/04/2011	9.945.700.00

*** Conceptos Pagados ***

CANT. RENTISTICO	CONCEPTO	Vr. UNDITARIO	Vr. CONCEPTO
1 50005-01-01 SOLICITUDES	1 TRAMITES DE SOL. DE PATENTE DE INVENCIÓN	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. _____

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

No. 11-055712- -00000-0000
 Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR.NUEVASCR
 Tra. 11 PATENTEIYII Eva. 378 FASENACIONALI
 Act. 411 PRESENTACION Folios: 227

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

No. 11-055712- -00000-0000
 Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR.NUEVASCR
 Tra. 11 PATENTEIYII Eva. 378 FASENACIONALI
 Act. 411 PRESENTACION Folios: 257

GARCIA

BOGOTA

Sede Centro: Carrera 13 No. 27 - 00 Pisos 3,4,5 y 10 Bogotá, D.C.- Colombia

Web: www.sic.gov.co e-mail: info@sic.gov.co Conmutador: (571) 5870000 Fax: (571) 5870284 Línea: 018000-910165 Call Center: (571) 6513240

Abril 20 de 2011 - 09:22:33

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800.176.089-2

- / -

Industria y Comercio
SUPERINTENDENCIA

RECIBO DE CAJA

No. 11 - 0030365

Bogotá D.C., Abril 04 de 2011 - 15:08:14

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	431625784	04/04/2011	11.455.000.00

*** Conceptos Pagados ***

CANT.	RENTISTICO	CONCEPTO	Vr. UNDITARIO	Vr. CONCEPTO
1	50005-01-09 OTROS SERVICIOS DE PROPIEDAD INDUSTRIAL	45 CAUCIONES	242.000.00	242.000.00
				\$242.000.00

SON: **DOSCIENTOS CUARENTA Y DOS MIL PESOS MONEDA CORRIENTE***

Responsable: _____

Recibo de Caja Aplicado al Expediente No. _____

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712-00000-0000

Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eva: 378 FASENACIONALI
Act. 411 PRESENTACION

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712-00000-0000

Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eva: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 257CARDENAS & CARDENAS
ABOGADOS

Sede Centro: Carrera 13 No. 27 - 00 Pisos 3,4,5 y 10 Bogotá, D.C.- Colombia

Web: www.sic.gov.co e-mail: Info@sic.gov.co Conmutador: (571) 5870000 Fax: (571) 5870284 Línea: 018000-910165 Call Center: (571) 6513240

Abril 4 de 2011 - 15:08:17

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800.176.089-2

- / -



RECIBO DE CAJA

No. 11 - 0030362

Bogotá D.C., Abril 04 de 2011 - 15:08:14

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	431625784	04/04/2011	11.455.000.00

*** Conceptos Pagados ***

CANT.	RENTISTICO	CONCEPTO	VR. LUNDITARIO	VR. CONCEPTO
1	50005-01-08 PRORROGA DE TERMINOS	1609 S.DISTINTIVOS PRORROGA DE TERMINOS O PLAZO ADICIONAL	87.000.00	87.000.00
				\$87.000.00

SON: **OCHENTA Y SIETE MIL PESOS MONEDA CORRIENTE**

Responsable: _____

Recibo de Caja Aplicado al Expediente No. _____

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712- -00000-0000

Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 411 PRESENTACION Folios. 257

CARDENAS & CARDENAS
ABOGADOS

Sede Centro: Carrera 13 No. 27 - 00 Pisos 3,4,5 y 10 Bogotá, D.C.- Colombia

Web: www.sic.gov.co e-mail: info@sic.gov.co Conmutador: (571) 5870000 Fax: (571) 5870284 Línea: 018000-910165 Call Center: (571) 6513240

Abril 4 de 2011 - 15:08:17

000854

- Colombia -

POWER OF ATTORNEY

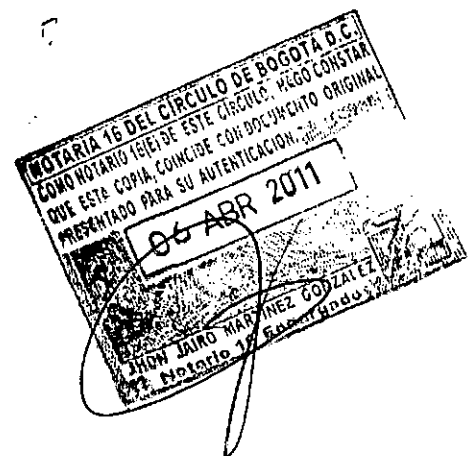
I, Werner Liebler, confirm:

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



File No. **H 2294** / 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,

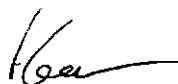
business address: 80686 München, Hansastr. 27 c,

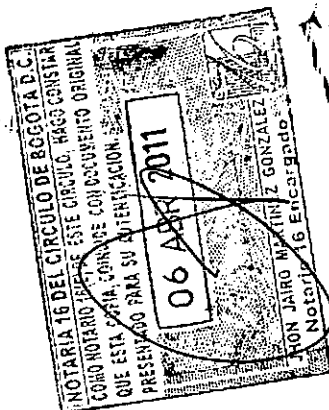
of power of attorney dated on July 13th, 2000, File No. 3076 K /2000, and

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



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

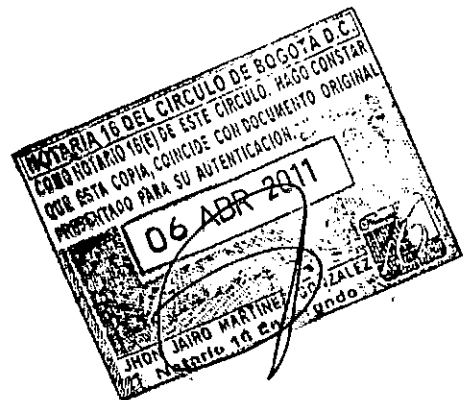
8. unter Nr. 910 a - 9274/2010

9. Siegel

10. Unterschrift
Im Auftrag:

Christine Köppl

Christine Köppl
Amtsinspektorin



Certified Copy



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

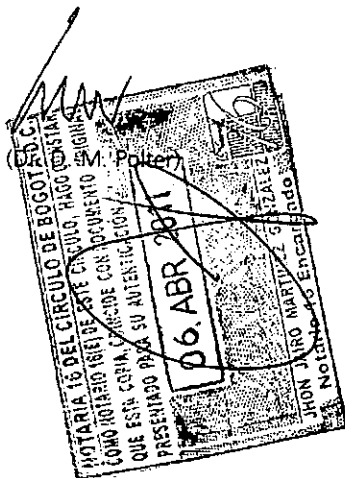
Hiermit bevollmächtigen wir

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. H.-U. Wiese)

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

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

Bankverbindung: Deutsche Bank, München
Konto 7521 933 812 700 700 10

URNr. 3076K

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

Herrn Dr. Dirk-Meints Polter,
Vorstandsmitglied in München,

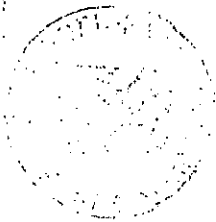
geschäftsansässig Leonrodstr. 54, 80636 München,
persönlich bekannt.

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



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,

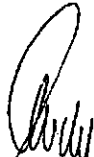
geschäftsansässig Leonrodstr. 54, 80636 München,
ausgewiesen durch amtlichen Lichtbildausweis.

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.

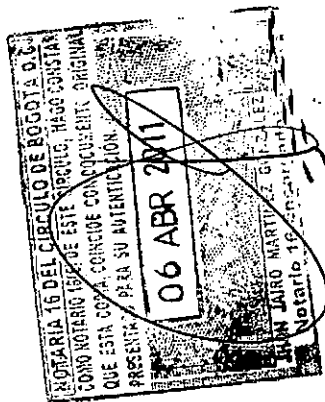
München, den 18.07.2000


Dr. Kerdel, Notar

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

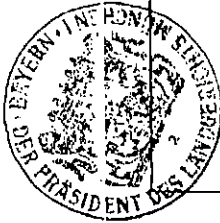
7. durch den Präsidenten des Landgerichts München I

8. unter Nr. 910 a - 9279/2010

9. Siegel

10. Unterschrift

Im Auftrag:



Christine Köppl

Christine Köppl
Amtsinspektorin



Martha S. Sarmiento S.
Martha S. Sarmiento S.

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 Sofia 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.

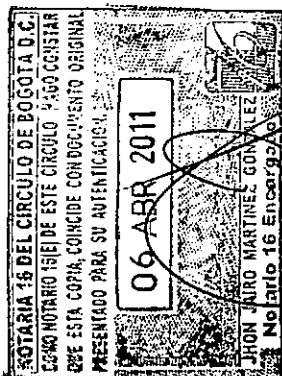
- COLOMBIA -

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



NO SE ACEPTA
RECURSOS JUDICIALES
DEL 1971

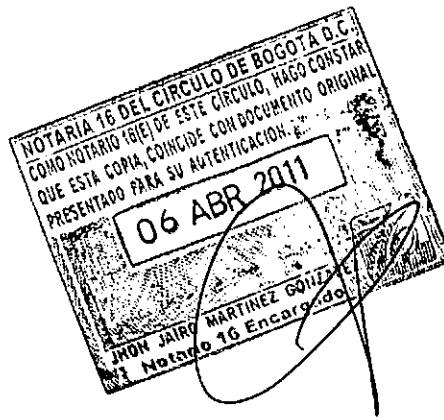
042633

Martha Sarmiento
COPIA DE LA FOTOCOPIA
DEL DOCUMENTO ORIGINAL
DE LA NOTARÍA DE BOGOTÁ

MINISTERIO DE
RELACIONES EXTERIORES

2000 ENE - 5/A 9 03

OFICINA DE RELACIONES
EXTERIORES
BOGOTÁ D.C.



MARTHA SOFÍA SARMIENTO SOLANO - TRADUCTORA OFICIAL
Carrera 11D # 124 - 51 Apto. 510 Int. 2 - Teléfono 2139034 Bogotá - Colombia
E-mail: msarmiento2@yahoo.com

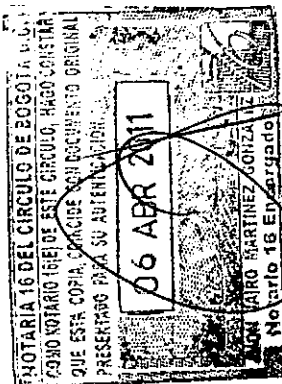
Martha S. Sarmiento S.
Martha S. Sarmiento S.

Traductora Oficial Juramentada
Escripción 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)



MARTHA SOFÍA SARMIENTO SOLANO - TRADUCTORA OFICIAL
Carrera 11D # 124 - C1 Apto. 610 Int. 2 - Teléfono 2130034 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

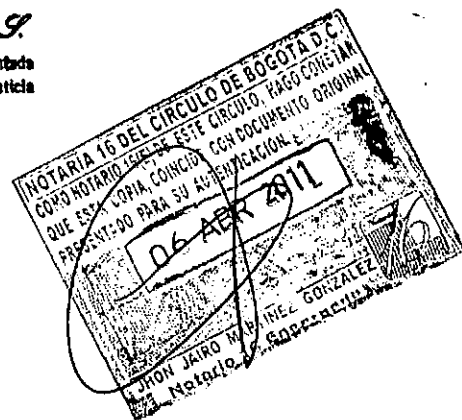
Sello del Notario:
THOMAS HAASEN
NOTARIO EN MUNICH

H:\Grunert\Unterschriftenbeglaubigungen\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 N° 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



Copia autenticada

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

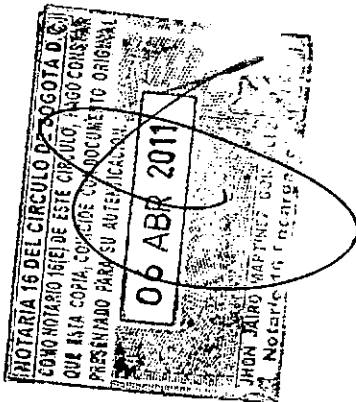
legales de protección.

Dr. D.-M. Polter

Dr. H.-U. Wiese

(firma)

(firma)



Miguel B. Fajardo V.
Traductor Oficial Juramentado
Escrit. 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)



Magno B. Fajardo V.
Traductor Oficial Juramentado
Prot. 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

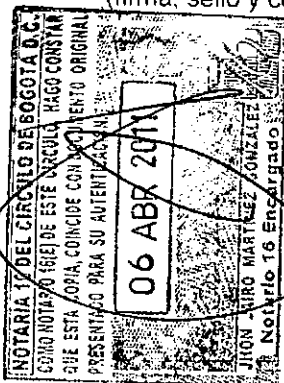
(firma y sello)

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

Munich, 6 de diciembre de 2010

El Notario, en Munich

(firma, sello y cordón)



María B. Fajardo V.
Traductor Oficial Juramentado
Reg. 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)



Magno R. 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

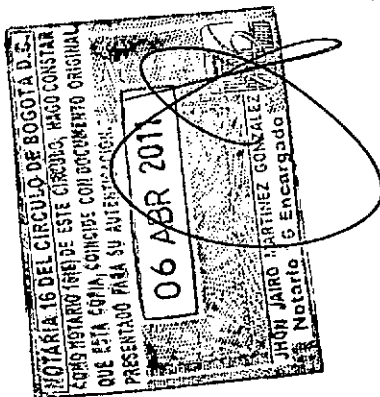
(firma y sello)

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

Munich, 6 de diciembre de 2010

El Notario, en Munich

(firma, sello y cordón)



Marco B. Fajardo V.
Traductor Oficial Juramentado
R.M. Justitia 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 Munich

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)



NO SE ACEPTA
RESPONSE
DEL 1999
04 24 50
Mace Tassende
BOGOTA D.C. 1999

MINISTERIO DE
RELACIONES EXTERIORES
2010 FEB - 1 A 8:16
BOGOTA D.C.
OFICE DE REGALIZACIONES

Margo B. Espinoza V.
Inspector Oficial Juramentado
Rut. Min. Justicia No. 839/86



Amtsgericht München -Registergericht-

VR 4461

Amtlicher aktueller Ausdruck aus dem Registerblatt

Datum der letzten Eintragung: 03.08.2010

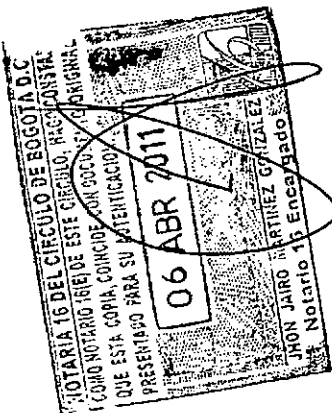
Datum des Abrufs: 19.10.2010

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.



Vereinsregister des Amtsgerichts München	Wiedergabe des aktuellen Registerinhalts Abruf vom 19.10.2010 08:33	Nummer des Vereins: VR 4461
-Ausdruck-	Seite 1 von 1	

Anzahl der bisherigen Eintragungen:

16

a) Name:

Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung eingetragener Verein

b) Sitz:

München

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

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 Mitgliederversammlung 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 Mitgliederversammlung 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

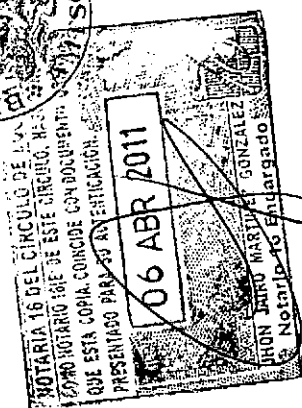
Bestätigt

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)



Miguel B. Fajardo V.
Inspector Oficial Juramentado
Rev. 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

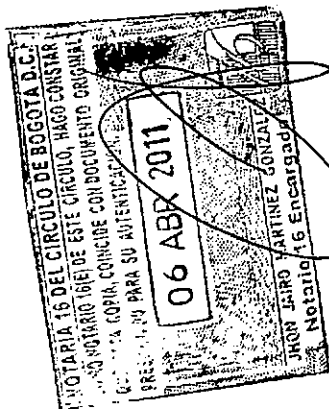
3. **a) Regulación general sobre representación:**

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.
Inspector Oficial Juramentado
Not. Min. Justicia No. 839/86

Junta directiva: Prof. (Universidad de Stellenbosch) Dr. Gossner, Alfred, de Erharting, nacido el 15 de octubre de 1950

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

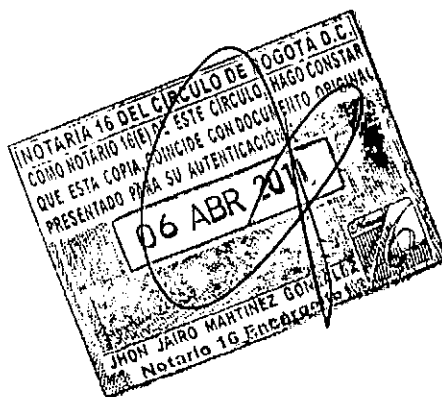
Estatutos del 26 de marzo de 1949

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



Margo R. Fajardo V.
Indagador Oficial Juramentado
Rol. Min. Justicia No. 939/86

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 autenticó 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



Alonso R. Fajardo V.
Procurador Oficial Interino
Resol. Min. Justicia No. 339/86

APOSTILLE

(Convention de La Haye du 5 octobre 1961)

1. Pais: **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
BOGOTÁ D.C.
2010 NOV 18 P 4 50
JEFE DE LEGALIZACIONES

NO SE ASUME
RESPONSABILIDAD
DEL TEXTO

038276
Recibido
UNA FIRMA SEPARADA EN EL
DOCUMENTO DE EMPEÑA
AS FUNCIONES INDICADAS

MINISTERIO DE RELACIONES EXTERIORES
RECIBO LEGALIZACION

No. 13843



Mayra B. Fajardo V.
Procuradora General del Estado
No. 839/86

**ESQUEMA DE CODIFICACIÓN/DECODIFICACIÓN DE AUDIO CONMUTADO DE RESOLUCIÓN
MÚLTIPLE**

MEMORIA DESCRIPTIVA

5

La presente invención se relaciona con la codificación de audio y, en particular, con esquemas de codificación de audio para baja tasa de transmisión de bits.

- 10 En la técnica se conocen los esquemas de codificación en el dominio de la frecuencia tales como MP3 o AAC. Estos codificadores en el dominio de la frecuencia se basan en la conversión de dominio del tiempo/dominio de la frecuencia, una etapa posterior de cuantificación, en la cual se controla el error de cuantificación utilizando información procedente de un módulo perceptivo y una etapa de
- 15 codificación, en la cual se realiza la codificación de entropía de los coeficientes espectrales cuantificados y la información secundaria correspondiente utilizando tablas de códigos.

- Por otra parte, hay codificadores que se adaptan muy favorablemente al procesamiento de voz tal como el AMR-WB+ descrito en 3GPP TS 26.290. Dichos
- 20 esquemas de codificación de voz ejecutan un filtrado Predictivo Lineal de una señal de dominio del tiempo. Ese filtrado LP deriva de un análisis de Predicción Lineal de la señal de entrada de dominio del tiempo. A continuación se cuantifican/codifican los coeficientes de filtrado LP y se los transmite en forma de información secundaria. El
- 25 proceso se conoce como Codificación de Predicción Lineal (LPC). A la salida del filtro, se codifica la señal residual de predicción o señal de error de predicción, la que también se conoce como señal de excitación, utilizando las etapas de análisis por síntesis del codificador ACELP o, por otro lado, se la codifica utilizando un codificador por transformada, que utiliza la transformada de Fourier con un
- 30 solapamiento. La opción entre la codificación ACELP y la codificación de la excitación Codificada por Transformada que también se denomina codificación TCX se decide utilizando un algoritmo de bucle cerrado o un algoritmo de bucle abierto.

- Los esquemas de codificación en el dominio de la frecuencia, como ser el esquema
- 35 de codificación AAC de Alta Eficiencia (HE-ACC), que combina un esquema de codificación ACC y una técnica de replicación de la banda espectral, también se puede combinar con una herramienta codificadora de estéreo conjunto o de canales múltiples que se conoce bajo la denominación "entorno MPEG".

220833
MB

- 40 Por otro lado, los codificadores de voz tales como AMR-WB+ también tienen una etapa de extensión de alta frecuencia y funcionalidad estéreo.

Los esquemas de codificación en el dominio de la frecuencia son ventajosos por el hecho de que exhiben calidad elevada a bajas tasas de transmisión de bits para señales musicales. Sin embargo, es problemática la calidad de las señales de voz a bajas tasas de transmisión de bits.

5

Los esquemas de codificación de voz exhiben calidad elevada en el caso de señales de voz, incluso a bajas tasas de transmisión de bits, que demuestran baja calidad con otras señales a bajas tasas de transmisión de bits.

- 10 Es un objetivo del presente invento el suministrar un concepto mejorado de codificación/decodificación.

Este objetivo se logra en virtud de un codificador de audio de acuerdo con la reivindicación 1, un método de codificación de audio de acuerdo con la reivindicación 9, un decodificador de acuerdo con la reivindicación 10, un método de decodificación de acuerdo con la reivindicación 19, una señal codificada de acuerdo con la reivindicación 20 o un programa de computación de acuerdo con la reivindicación 21.

La presente invención se basa en el hallazgo de que un esquema de codificación/decodificación conmutado híbrido o de dos modalidades es ventajoso por el hecho de que siempre se puede seleccionar el mejor algoritmo de codificación para una determinada característica de la señal. Dicho de otro modo, la presente invención no busca un algoritmo de codificación de señales que coincida perfectamente con todas las características de las señales. Ese tipo de esquema siempre representaría un menoscabo, como se puede apreciar por las enormes diferencias entre los codificadores de audio del estado actual de la técnica por un lado, y los codificadores de voz por el otro. Por el contrario, la presente invención combina diferentes algoritmos de codificación tales como el algoritmo de codificación de voz por un lado y un algoritmo de codificación de audio por el otro, dentro de un esquema conmutado de manera que, por cada porción de señal de audio, se seleccione el algoritmo de codificación que coincida de manera óptima. Más aun, otra característica de la presente invención es que ambas ramas codificadoras comprenden un convertidor de tiempo/frecuencia, aunque en una rama de codificación se presenta otro convertidor de dominio tal como un procesador LPC. Este convertidor de dominio se asegura de que la segunda rama codificadora se adecue mejor a una determinada característica de la señal que la primera rama codificadora. Sin embargo, otra de las características de la presente invención es que la señal producida por el procesador de dominio también se transforma en una representación espectral.

40

Ambos convertidores, es decir, el primer convertidor de la primera rama codificadora y el segundo convertidor de la segunda rama codificadora, están configurados para

5 aplicar una codificación por transformada de resolución múltiple, donde se ajusta la resolución del convertidor correspondiente dependiendo de la señal de audio y, en particular, dependiendo de la señal de audio codificada en realidad en la rama codificadora correspondiente, de manera que se obtenga una buena relación entre la calidad por un lado y la tasa de transmisión de bits por el otro, o en vista de una determinada calidad fija, de la tasa de transmisión de bits más baja o en vista de una tasa de transmisión de bits fija, la calidad más alta.

10 De acuerdo con la presente invención, la resolución de tiempo/frecuencia de los dos convertidores se puede ajustar preferentemente de manera independiente entre sí de manera que cada transformador de tiempo/frecuencia pueda ajustarse de manera óptima a los requerimientos de resolución de tiempo/frecuencia de la señal correspondiente. La eficiencia binaria, es decir, la relación entre los bits útiles por un lado y los bits de información secundaria por el otro, es más elevada el caso de
15 tamaños mayores de bloques/longitudes de ventana. Por lo tanto, es preferible que ambos convertidores estén más polarizados a una longitud de ventana más grande, puesto que básicamente la misma cantidad de información secundaria se refiere a una porción más larga de la señal de audio en comparación con la aplicación de menores tamaños de bloque/longitudes de ventana/ longitudes de transformada. De
20 preferencia, la resolución de tiempo/frecuencia en las ramas codificadoras también puede recibir la influencia de otras herramientas de codificación/decodificación situadas en estas ramas. De preferencia, la segunda rama codificadora que comprende el convertidor de dominio, como por ejemplo un procesador LPC, comprende otro esquema híbrido tal como una rama ACELP por un lado y un
25 esquema TCX por el otro, donde el segundo convertidor está incluido en el esquema TCX. De preferencia, el convertidor de resolución de tiempo/frecuencia situado en la rama TCX también está influenciado por la decisión de codificación, de manera que una porción de la señal de la segunda rama codificadora sea procesada en la rama TCX que tiene el segundo convertidor, o en la rama ACELP que no cuenta con
30 convertidor de tiempo/frecuencia.

Básicamente, ni el convertidor de dominio ni la segunda rama codificadora, y en especial la primera rama de procesamiento de la segunda rama codificadora y la segunda rama de procesamiento de la segunda rama codificadora deben ser
35 elementos relacionados con la voz tales como un analizador LPC en el caso del convertidor de dominio, un codificador TCX en el caso de la segunda rama de procesamiento y un codificador ACELP en el caso de la primera rama de procesamiento. También son de utilidad otras aplicaciones cuando se evalúan otras características de señal de una señal de audio diferente de la voz (voz) por un lado y
40 música por el otro. Se puede utilizar cualquier convertidor de dominio e implementaciones de ramas codificadoras y se puede encontrar el algoritmo con mayor coincidencia mediante un esquema de análisis por síntesis de manera que,

del lado del codificador, por cada porción de la señal de audio, se ejecuten todas las alternativas de codificación y se seleccione el mejor resultado, donde se puede encontrar el mejor resultado aplicando una función objetivo a los resultados de la codificación. A continuación, la información secundaria que identifica, a un
5 decodificador, el algoritmo de codificación para una determinada porción de la señal de audio codificada es anexada a la señal de audio codificada mediante una interfaz de salida del codificador, de manera que el decodificador no tenga necesidad de tomar decisiones del lado del codificador ni con respecto a alguna característica de la señal, sino que simplemente selecciona su rama codificadora dependiendo de la
10 información secundaria transmitida. Más aun, el codificador no sólo selecciona la rama decodificadora correcta sino que también selecciona, sobre la base de la información secundaria codificada en la señal codificada, qué resolución de tiempo/frecuencia se ha de aplicar en una primera rama decodificadora correspondiente y una segunda rama decodificadora correspondiente.

15 Por consiguiente, la presente invención da a conocer un esquema de codificación/decodificación que combina las ventajas de todos los algoritmos de codificación diferentes y evita las desventajas de estos algoritmos de codificación que se presentan, cuando tendría que codificarse la porción de señal, por un
20 algoritmo que no se ajusta a un determinado algoritmo de codificación. Más aun, la presente invención evita toda desventaja que se pudiera presentar si no se hubieran tomado en cuenta los requerimientos de resolución en el tiempo/frecuencia generados por dos porciones de señal de audio diferentes en las diferentes ramas codificadoras. Por el contrario, debido a la resolución variable de tiempo/frecuencia
25 de los convertidores de tiempo/frecuencia de ambas ramas, por lo menos se reduce, o incluso se evita por completo toda aberración que se pueda producir en el caso en que se aplicara la misma resolución de tiempo/frecuencia a ambas ramas codificadoras, o en el cual sólo fuera posible una resolución fija de tiempo/frecuencia para alguna de las ramas codificadoras.

30 El segundo conmutador decide otra vez entre las dos ramas de procesamiento, aunque en un dominio diferente del dominio de la primera rama "externa". Una vez más, la rama "interna" está motivada principalmente por un modelo de origen o por cálculos de SNR y la otra rama "interna" puede estar motivada por un modelo de
35 destino y/o un modelo psicoacústico, es decir, que enmascara, o por lo menos incluye aspectos de codificación en el dominio de la frecuencia/espectral. A manera de ejemplo, una rama "interna" tiene un codificador de dominio de la frecuencia/convertidor espectral y la otra rama tiene un codificador que codifica en el otro dominio tal como el dominio LPC, donde este codificador es, por ejemplo, un
40 cuantificador/escalador CELP o ACELP que procesa una señal de entrada sin conversión espectral.

Otra realización preferida es un codificador de audio que comprende una primera rama codificadora orientada a un primer destino de la información tal como una rama codificadora en el dominio espectral, una segunda rama codificadora orientada a una fuente de información o SNR tal como una rama codificadora en el dominio LPC y un conmutador para conmutar entre la primera rama codificadora y la segunda rama codificadora, donde la segunda rama codificadora comprende un convertidor a un dominio específico diferente del dominio del tiempo tal como una etapa de análisis LPC que genera una señal de excitación, y donde la segunda rama codificadora comprende además una rama de procesamiento en un dominio específico tal como un dominio LPC y una rama de procesamiento en un dominio espectral específico tal como un dominio LPC y un conmutador adicional para conmutar entre la rama codificadora de dominio específico y la rama codificadora de dominio espectral específica.

Una realización más de la invención es un decodificador de audio que comprende una rama codificadora de un primer dominio tal como un dominio espectral, una rama decodificadora de un segundo dominio tal como un dominio LPC para decodificar una señal tal como una señal de excitación en el Segundo dominio y una rama decodificadora de un tercer dominio tal como LPC espectral, donde el tercer dominio se obtiene ejecutando una conversión de frecuencia del segundo dominio, donde se presenta un primer conmutador para la señal del segundo dominio y la señal del tercer dominio y donde se presenta un segundo conmutador para conmutar entre el decodificador del primer dominio y el decodificador del segundo dominio o del tercer dominio.

A continuación se describen las realizaciones preferidas de la presente invención con respecto a los dibujos adjuntos, en los cuales:

La Fig. 1a es un diagrama de bloques de un esquema de codificación de acuerdo con un primer aspecto de la presente invención;

La Fig. 1b es un diagrama de bloques de un esquema de decodificación de acuerdo con el primer aspecto de la presente invención;

La Fig. 1c es un diagrama de bloques de un esquema de codificación de acuerdo con otro aspecto de la presente invención;

La Fig. 2a es un diagrama de bloques de un esquema de codificación de acuerdo con un segundo aspecto de la presente invención;

La Fig. 2b es un diagrama esquemático de un esquema de decodificación de acuerdo con el segundo aspecto de la presente invención.

- La Fig. 2c es un diagrama de bloques de un esquema de codificación de acuerdo con otro aspecto de la presente invención
- 5 La Fig. 3a ilustra un diagrama de bloques de un esquema de codificación de acuerdo con otro aspecto de la presente invención;
- La Fig. 3b ilustra un diagrama de bloques de un esquema de decodificación de acuerdo con el otro aspecto de la presente invención;
- 10 La Fig. 3c ilustra una representación esquemática del aparato/método de codificación con conmutadores en cascada;
- La Fig. 3d ilustra un diagrama esquemático de un aparato o método para decodificar, en el cual se utilizan combinadotes en cascada;
- 15 La Fig. 3e muestra una ilustración de una señal en el dominio del tiempo y una representación correspondiente de la señal codificada que ilustra regiones cortas de desvanecimiento cruzado que están incluidas en ambas señales codificadas;
- 20 La Fig. 4a ilustra un diagrama de bloques con un conmutador situado antes de las ramas codificadoras;
- 25 La Fig. 4b ilustra un diagrama de bloques de un esquema de codificación con el un conmutador situado después de las ramas codificadoras;
- La Fig. 5a ilustra una forma de onda de un segmento de voz en el dominio del tiempo en forma de señal de segmentos casi periódicos o por impulsos;
- 30 La Fig. 5b ilustra un espectro del segmento de la Fig. 5a;
- La Fig. 5c ilustra un segmento de voz en el dominio del tiempo de voz sin voz como ejemplo de segmento tipo ruido;
- 35 La Fig. 5d ilustra un espectro de la forma de onda en el dominio del tiempo de la Fig. 5c;
- La Fig. 6 ilustra un diagrama de bloques de un codificador CELP de análisis por síntesis;
- 40

Las Figs. 7a a 7d ilustran señales de excitación con voz/sin voz como ejemplo de señales de tipo por impulsos;

5 La Fig. 7e ilustra una etapa de LPC del lado del codificador que presenta información predictiva a corto plazo y la señal de error de predicción (excitación);

10 La Fig. 7f ilustra otra realización de un dispositivo LPC para generar una señal ponderada;

15 La Fig. 7g ilustra una implementación para transformar una señal ponderada en una señal de excitación mediante la aplicación de una operación de ponderación inversa y un posterior análisis de excitación necesario en el convertidor 537 de la Fig. 2b;

La Fig. 8 ilustra un diagrama de bloques de un algoritmo de múltiples canales de acuerdo con una realización de la presente invención;

20 La Fig. 9 ilustra una realización preferida de un algoritmo de extensión de ancho de banda;

La Fig. 10a ilustra una descripción detallada del conmutador al ejecutar una decisión de bucle abierto y

25 La Fig. 10b ilustra una representación del conmutador funcionando en un modo de decisión en bucle cerrado;

30 La Fig. 11A ilustra un diagrama de bloques de un codificador de audio de acuerdo con otro aspecto de la presente invención;

La Fig. 11B ilustra un diagrama de bloques de otra realización de un decodificador de audio de la invención;

35 La Fig. 12A ilustra otra realización de un codificador de la invención;

La Fig. 12B ilustra otra realización de un decodificador de la invención;

40 La Fig. 13A ilustra la interrelación entre longitudes de resolución ventana/transformación;

La Fig. 13B ilustra una vista general de una serie de ventanas de transformación correspondiente a la primera rama codificadora y una transición de la

primera a la segunda rama codificadora;

5 La Fig. 13C ilustra una pluralidad de diferentes secuencias de ventana que incluyen secuencias de ventana correspondientes a la primera rama codificadora y secuencias correspondientes a una transición a la segunda rama;

10 La Fig. 14A ilustra el marco de una realización preferida de la segunda rama codificadora;

La Fig. 14B ilustra ventanas cortas aplicadas en la segunda rama codificadora;

15 La Fig. 14C ilustra ventanas de tamaño mediado aplicadas en la segunda rama codificadora;

La Fig. 14D ilustra ventanas largas aplicadas por la segunda rama codificadora;

20 La Fig. 14E ilustra una secuencia ilustrativa de marcos ACELP y marcos TCX dentro de una división de super marco;

La Fig. 14F ilustra diferentes longitudes de transformadas correspondientes a diferentes resoluciones de tiempo/frecuencia correspondientes a la segunda rama codificadora y

25 La Fig. 14G ilustra una construcción de una ventana que utiliza las definiciones de la Fig. 14F

30 La Fig. 11A ilustra una realización de un codificador de audio para codificar una señal de audio. El codificador comprende una primera rama codificadora 400 para codificar una señal de audio utilizando un primer algoritmo de codificación para obtener una primera señal codificada.

35 El codificador de audio comprende además una segunda rama codificadora 500 para codificar una señal de audio utilizando un segundo algoritmo de codificación para obtener una segunda señal codificada. El primer algoritmo de codificación es diferente del segundo algoritmo de codificación. Además, se incluye un primer conmutador 200 para conmutar entre la primera rama codificadora y la segunda rama codificadora de manera que, para una porción de la señal de audio, esté la primera señal codificada o la segunda señal codificada en una señal de salida del codificador
40 801.

El codificador de audio ilustrado en la Fig. 11A comprende además un analizador de señales 300/525, que está configurado para analizar una porción de la señal de audio para determinar si la porción de la señal de audio está representada como primera señal codificada o segunda señal codificada en la señal de salida del codificador 801.

El analizador de señales 300/525 está configurado además para determinar de manera variable una resolución respectiva de tiempo/frecuencia de un primer convertidor 410 de la primera rama codificadora 400 o un segundo convertidor 523 de la segunda rama codificadora 500. Esta resolución de tiempo/frecuencia se aplica cuando se genera la primera señal codificada o la segunda señal codificada que representa la porción de la señal de audio.

El codificador de audio comprende además una interfaz de salida 800 para generar la señal de salida del codificador 801 que comprende una representación codificada de la porción de la señal de audio y una información que indica si la representación de la señal de audio es la primera señal codificada o la segunda señal codificada e indica la resolución de tiempo/frecuencia utilizada para decodificar la primera señal codificada y la segunda señal codificada.

La segunda rama codificadora es preferentemente diferente de la primera rama codificadora por el hecho de que la segunda rama codificadora comprende además un convertidor de dominio para convertir la señal de audio del dominio, en el cual se procesa la señal de audio en la primera rama codificadora, a un dominio diferente. De preferencia, el convertidor de dominio es un procesador LPC 510, aunque el convertidor de dominio puede ser implementado de cualquiera otra manera siempre que el convertidor de dominio sea diferente del primer convertidor 410 y del segundo convertidor 523.

El primer convertidor 410 es un convertidor de tiempo/frecuencia que comprende preferentemente un generador de ventanas ("windower") 410a y un transformador 410b. El generador de ventanas 410a aplica una ventana de análisis a la señal de audio de entrada y el transformador 410b ejecuta una conversión de la señal en ventana a la representación espectral.

De manera análoga, el segundo convertidor 523 comprende preferentemente un generador de ventanas 523a y un transformador conectado subsiguientemente 523b. El generador de ventana 523a recibe la señal emitida por el convertidor de dominio 510 y produce la representación en ventanas de la misma. El resultado de una ventana de análisis aplicada por el generador de ventanas 523a es enviado al transformador 523b para formar una representación espectral. El transformador puede ser un procesador FFT o preferentemente MDCT que implementa un

algoritmo correspondiente en software o hardware o en una implementación mixta de software/hardware. Por otro lado, el transformador puede consistir en una implementación de banco de filtros tal como un banco de filtros QMF que se puede basar en una modulación de valor real o complejo de un filtro prototípico. En el caso
5 de implementaciones específicas de banco de filtros, se aplica una ventana. Sin embargo, para otras implementaciones de banco de filtros, no es necesaria la visualización en ventanas indispensable para un algoritmo de transformada basado en un FFT de MDCT. Cuando se utiliza una implementación de banco de filtros, luego el banco de filtros es un banco de filtros de resolución variable y la resolución
10 controla la resolución en frecuencia del banco de filtros y, además, la resolución en tiempo o sólo la resolución en frecuencia y no la resolución en tiempo. No obstante, cuando se implementa el convertidor en forma de FFT o MDCT o cualquier otro transformador correspondiente, luego se conecta la resolución en frecuencia a la resolución en tiempo por el hecho de que un aumento de la resolución en frecuencia
15 obtenida por una mayor longitud de bloques en tiempo corresponde automáticamente a una resolución menor en tiempo y viceversa.

Además, la primera rama codificadora puede comprender una etapa de cuantificación/codificación 421 y la segunda rama codificadora puede comprender
20 además una o más herramientas codificadoras adicionales 524.

Lo importante es que el analizador de señales está configurado para generar una señal de control de resolución para el primer convertidor 510 y para el segundo convertidor 523. Por consiguiente, se implementa un control independiente de la
25 resolución en ambas ramas codificadoras para tener un esquema de codificación que, por un lado produce una baja tasa de transmisión de bits y, por el otro, otorga una calidad máxima en vista de la baja tasa de transmisión de bits. Para obtener la meta de la baja tasa de transmisión de bits, son preferibles mayores longitudes de ventana o mayores longitudes de transformación, aunque en situaciones en que
30 estas grandes longitudes dan lugar a una aberración debido a la baja resolución en tiempo, se aplican longitudes de ventana más cortas y longitudes de transformación más cortas, lo que da lugar a una resolución en frecuencia más baja. De preferencia, el analizador de señales aplica un análisis estadístico o cualquier otro análisis que se adecue a los algoritmos correspondientes en las ramas codificadoras. En un modo
35 de implementación, en el cual la primera rama codificadora es una rama codificadora en el dominio de la frecuencia tal como un codificador basado en AAC, y en el cual la segunda rama codificadora comprende, como convertidor de dominio, un procesador LPC 510, el analizador de señales ejecuta una discriminación de voz/música de manera que la porción de voz de la señal de audio sea alimentada a la segunda
40 rama codificadora controlando de manera correspondiente al conmutador 200. Se alimenta una porción de música de la señal de audio a la primera rama codificadora 400 controlando de manera correspondiente al conmutador 200 de acuerdo con lo

indicado por las líneas de control del conmutador. Por otro lado, como se describe más adelante con respecto a la Fig. 1C o la Fig. 4B, el conmutador también puede estar situado antes de la interfaz de salida 800.

- 5 Más aun, el analizador de señales puede recibir la señal de audio ingresada al conmutador 200, o la señal de audio emitida por el conmutador 200. Asimismo, el analizador de señales ejecuta un análisis para no solo alimentar la señal de audio a la correspondiente rama codificadora, sino también para determinar la resolución en tiempo/frecuencia apropiada del convertidor respectivo en la correspondiente rama
- 10 codificadora, como por ejemplo el primer convertidor 410 y el segundo convertidor 523 como lo indican las líneas controladas de resolución que conectan al analizador de señales y al convertidor

- La Fig. 11B comprende una realización preferida de un decodificador de audio que
- 15 corresponde al codificador de audio in Fig. 11A.

- El decodificador de audio de la Fig. 11B está configurado para decodificar una señal de audio codificada tal como la señal de salida del codificador 801 que sale de la interfaz de salida 800 de la Fig. 11A. La señal codificada comprende una primera
- 20 señal de audio codificada de acuerdo con un primer algoritmo de codificación, una segunda señal de audio codificada de acuerdo con un segundo algoritmo de codificación, donde el segundo algoritmo de codificación es diferente del primer algoritmo de codificación, e información que indica si se utiliza el primer algoritmo de codificación o el segundo algoritmo de codificación para decodificar la primera señal
- 25 codificada y la segunda señal codificada y una información de resolución en tiempo/frecuencia para la primera señal de audio codificada y la segunda señal de audio codificada.

- El decodificador de audio comprende una primera rama decodificadora 431, 440 para
- 30 decodificar la primera señal codificada basándose en el primer algoritmo de codificación. Más aun, el decodificador de audio comprende una segunda rama decodificadora para decodificar la segunda señal codificada usando el segundo algoritmo de codificación.

- La primera rama decodificadora comprende un primer convertidor controlable 440
- 35 para convertir de un dominio espectral a un dominio del tiempo. El convertidor controlable está configurado para ser controlado utilizando la información de resolución en tiempo/frecuencia de la primera señal codificada para obtener la primera señal decodificada.

- 40 La segunda rama decodificadora comprende un segundo convertidor controlable para convertir de una representación espectral a una representación en el tiempo,

donde el segundo convertidor controlable 534 está configurado para ser controlado utilizando la información de resolución en tiempo/frecuencia 991 para la segunda señal codificada.

- 5 El decodificador comprende además un controlador 990 para controlar el primer convertidor 540 y el segundo convertidor 534 de acuerdo con la información de resolución en tiempo/frecuencia 991.

Además, el decodificador comprende un convertidor de dominio para generar una
10 señal de síntesis utilizando la segunda señal decodificada para cancelar la conversión de dominio aplicada por el convertidor de dominio 510 en el codificador de la Fig. 11A.

De preferencia, el convertidor de dominio 540 es un procesador de síntesis LPC,
15 que se controla utilizando la información de filtro LPC incluida en la señal codificada, donde esta información de filtro LPC ha sido generada por el procesador LPC 510 de la Fig. 11A y ha sido ingresada a la señal de salida del codificador en forma de información secundaria. Por último, el decodificador de audio comprende un
20 combinador 600 para combinar la primera señal decodificada producida por el primer convertidor de dominio 440 y la señal de síntesis para obtener una señal de audio decodificada 609.

En la implementación preferida, la primera rama decodificadora comprende además una etapa descuantificador/decodificador 431 para revertir o, por lo menos revertir
25 parcialmente las operaciones ejecutadas por el correspondiente paso de codificación 421. Sin embargo, resulta claro que la cuantificación no se puede revertir, ya que es una operación con pérdida. Sin embargo, un descuantificador revierte una cierta falta de uniformidad en una cuantificación tal como una cuantificación logarítmica o por compresión/expansión.

30 En la segunda rama decodificadora, se aplica la etapa correspondiente 533 para deshacer ciertas operaciones de codificación aplicadas por la etapa 524. De preferencia, la etapa 524 comprende una cuantificación uniforme. Por lo tanto, la etapa correspondiente 533 no ha de tener una etapa de descuantificación específico
35 para deshacer una determinada cuantificación uniforme.

El primer convertidor 440, como así también el segundo convertidor 534, puede comprender un correspondiente paso de transformación inversa 440a, 534a, una
40 etapa de ventana de síntesis 440b, 534b, y la etapa conectada subsiguientemente de traslape/suma 440c, 534c. Las etapas de traslape/suma son indispensables cuando los convertidores y, más específicamente, las etapas transformadoras 440a, 534a aplican el solapamiento introduciendo transformadas tales como una transformada

de coseno discreta. A continuación, la operación de traslape/suma ejecuta una cancelación de solapamiento en el dominio del tiempo (TDAC). Cuando sin embargo, los transformadores aplican una transformada que no introduce solapamientos tales como un FFT inverso, luego no es necesaria una etapa de traslape/suma 440c. En
5 ese tipo de implementación, se puede aplicar una operación de desvanecimiento cruzado para evitar aberraciones de bloqueo.

De manera análoga, el combinador 600 puede ser un combinador conmutado o un combinador de desvanecimiento cruzado o, cuando se utiliza el solapamiento para
10 evitar aberraciones de bloqueo, el combinador implementa una operación de ventanas de transición similar a una etapa de traslape/suma dentro de una rama en sí.

La Fig. 1a ilustra una realización de la invención que presenta dos conmutadores en cascada. Se ingresa una señal mono, una señal estéreo o una señal de multicanal al conmutador 200. El conmutador 200 es controlado por la etapa de decisión 300. La
15 etapa de decisión recibe, en forma de entrada, una señal ingresada en el bloque 200. Por otro lado, la etapa de decisión 300 también puede recibir una información secundaria que está incluida en la señal mono, la señal estéreo o la señal de multicanal o que por lo menos está asociada a dicha señal, donde hay información
20 existente que fuera generada, por ejemplo, al producir en un principio la señal mono, la señal estéreo o la señal de multicanal.

La etapa de decisión 300 acciona el conmutador 200 a fin de alimentar una señal ya sea en la porción de codificación de frecuencia 400 ilustrada en una rama superior de la Fig. 1a o en la porción codificadora de dominio LPC 400 ilustrada en una rama inferior en la Fig. 1a. Un elemento clave de la rama codificadora en el dominio de la frecuencia es el bloque de conversión espectral 410 que cumple la función de
25 convertir una señal de salida de la etapa de procesamiento común (que se describe más adelante) en el dominio espectral. El bloque de conversión espectral puede incluir un algoritmo MDCT, un algoritmo QMF, FFT, un análisis de Wavelet o un banco de filtros tal como un banco de filtros con muestreo crítico que tiene un número de canales de banco de filtros, donde las señales de subbanda de este banco de filtro pueden ser señales de valor real o señales de valor complejo. La
30 salida del bloque de conversión espectral 410 se codifica empleando un codificador de audio espectral 421, que puede incluir bloques de procesamiento como se diera a conocer en el esquema de codificación AAC.

En general, la rama de procesamiento 400 es un procesamiento en un modelo
40 basado en la percepción o un modelo de destino de la información. Por consiguiente, esta rama modela el sistema auditivo humano que recibe sonido. A diferencia de éste, el procesamiento que tiene lugar en la rama 500 es para generar una señal en

el dominio de la excitación, residual o LPC. Por lo general, el procesamiento que tiene lugar en la rama 500 es un procesamiento en un modelo de voz o un modelo de generación de información. En el caso de las señales de voz, este modelo del sistema de generación de voz humana/sonido que genera sonido. Sin embargo, si se debe codificar un sonido procedente de un origen diferente que requiere un modelo de generación de sonido diferente, luego el procesamiento que se produce en la rama 500 puede ser diferente.

En la rama codificadora inferior 500, un elemento clave es un dispositivo LPC 510, que emite una información LPC que se utiliza para controlar las características de un filtro LPC. Esta información LPC es transmitida a un decodificador. La señal de salida de la etapa LPC 510 es una señal de dominio LPC que consiste en una señal de excitación y/o una señal ponderada.

El dispositivo LPC generalmente emite una señal en dominio LPC, que puede ser una señal en el dominio LPC tal como la señal de excitación de la Fig. 7e o una señal ponderada de la Fig. 7f o cualquier otra señal, que se haya generado mediante la aplicación de coeficientes de filtrado LPC a una señal de audio. Más aun, un dispositivo LPC también puede determinar estos coeficientes y también puede cuantificar/codificar estos coeficientes.

La decisión que se toma en la etapa de decisión puede ser adaptativa a la señal de manera que la etapa de decisión ejecuta una discriminación música/voz y controla al conmutador 200 de manera tal que las señales de música sean ingresadas a la rama superior 400 y las señales de voz sean ingresadas en la rama inferior 500. En una realización, la etapa de decisión alimenta esta información sobre la decisión a un flujo de bits de salida de manera que un decodificador pueda emplear esta información de decisión para ejecutar las operaciones de decodificación correctas.

Dicho decodificador está ilustrado en la Fig. 1b. La señal de salida del codificador de audio espectral 421 es ingresada, después de la transmisión, a un decodificador de audio espectral 431. La salida del decodificador de audio espectral 431 es ingresada a un convertidor en el dominio del tiempo 440. De manera análoga, la salida de la rama codificadora en el dominio LPC 500 de la Fig. 1a es recibida en el lado del decodificador y procesada por los elementos 531, 533, 534 y 532 para obtener una señal de excitación LPC. La señal de excitación LPC es ingresada a una etapa de síntesis LPC 540 que recibe, como entrada adicional, la información LPC generada por la correspondiente etapa de análisis LPC 510. La salida del convertidor en dominio del tiempo 440 y/o la salida de la etapa de síntesis LPC 540 son ingresadas a un conmutador 600. El conmutador 600 se controla por medio de una señal de control de conmutador que fuera generada, por ejemplo, por la etapa de decisión 300, o que se produjera en forma externa, como por ejemplo por un generador de la

señal mono, señal estéreo o señal de multicanal original. La salida del conmutador 600 es una señal mono, señal estéreo o señal de multicanal completa.

- La señal de entrada al conmutador 200 y la etapa de decisión 300 pueden ser una
5 señal mono, señal estéreo o señal de multicanal o, en términos generales, una señal de audio. Dependiendo de la decisión que se puede derivar de la señal de entrada al conmutador 200 o de cualquier fuente externa tal como un productor de una señal de audio original que subyace a la señal ingresada a la etapa 200, el conmutador conmuta entre la rama codificadora en frecuencia 400 y la rama codificadora en LPC
10 500. La rama codificadora en frecuencia 400 comprende una etapa de conversión espectral 410 y una etapa de cuantificación/codificación conectada subsiguientemente. La etapa de cuantificación/codificación puede incluir cualquiera de las funcionalidades conocidas de los codificadores en el dominio de la frecuencia conocidos tales como el codificador AAC. Más aun, la operación de cuantificación en
15 la etapa de cuantificación/codificación 421 puede ser controlada por medio de un módulo psicoacústico que genera información psicoacústica tal como un umbral de enmascaramiento psicoacústico sobre la frecuencia, donde esta información es ingresada en la etapa 421.
- 20 En la rama codificadora LPC, se procesa la señal de salida del conmutador por medio de una etapa de análisis LPC 510 que genera información secundaria LPC y una señal en dominio LPC. El codificador de excitación comprende, de acuerdo con la invención, un conmutador adicional para conmutar el posterior procesamiento de la señal en dominio LPC entre una operación de cuantificación/codificación 524, que
25 procesa valores en el dominio LPC–espectral. Para este fin, se incluye un convertidor espectral 523 a la entrada de la etapa de cuantificación/codificación 524. El conmutador 521 se controla en forma de bucle abierto o en forma de bucle cerrado, dependiendo de la configuración específica, de acuerdo con lo descrito, por ejemplo, en la especificación técnica AMR-WB+.
- 30 Con respecto a la modalidad de control por bucle cerrado, el codificador incluye además un cuantificador/codificador inverso 531 para la señal en el dominio LPC, un cuantificador/codificador inverso 533 para la señal en el dominio espectral LPC y un convertidor inverso 534 para la salida del elemento 533. Tanto las señales
35 codificadas como las decodificadas una vez más en las ramas de procesamiento de la segunda rama codificadora son ingresadas al dispositivo de control del conmutador 525. En el dispositivo de control del conmutador 525, se comparan entre sí y/o con una función objetivo estas dos señales de salida o se calcula una función objetivo que se puede basar en una comparación de la distorsión de ambas señales
40 de manera de utilizar la señal con menos distorsión para decidir qué posición debe tomar el conmutador 521. Por otro lado, en caso de que ambas ramas produzcan tasas de transferencia de bits no constantes, se podría seleccionar la rama que

produce la tasa de transferencia de bits más baja incluso cuando la relación señal a ruido de esta rama es más baja que la relación señal a ruido de la otra rama. Por otro lado, la función objetivo podría utilizar, como entrada, la relación señal a ruido de cada señal y una tasa de transferencia de bits de cada señal y/o criterios adicionales a fin de encontrar la mejor decisión para una meta específica. Si, por ejemplo, la meta es tal que la tasa de transferencia de bits debe ser lo más baja posible, luego la función objetivo debería confiar en gran medida en la tasa de transferencia de bits de las dos señales producidas por los elementos 531, 534. Sin embargo, cuando la meta principal es contar con la mejor calidad para determinada tasa de transferencia de bits, luego el control del conmutador 525 podría descartar, por ejemplo, cada señal que esté por encima de la tasa de transferencia de bits admitida y cuando ambas señales están por debajo de la tasa de transferencia de bits admitida, el control del conmutador seleccionaría la señal con la mejor relación señal a ruido, es decir, las menores distorsiones de cuantificación/codificación.

El esquema de decodificación de acuerdo con la presente invención está ilustrado, como se señalara anteriormente, en la Fig. 1b. Por cada uno de los tres tipos de señal de salida posibles, existe una etapa específica de decodificación/recuantificación 431, 531 o 533. Mientras la etapa 431 produce un espectro en el tiempo que se convierte a dominio en tiempo utilizando el convertidor de frecuencia/tiempo 440, la etapa 531 da salida a una señal en dominio LPC y el elemento 533 produce la salida de un espectro LPC. Para asegurarse de que las señales de entrada al conmutador 532 estén, en ambos casos, en el dominio LPC, se incluye el espectro LPC/convertidor LPC 534. Los datos de salida del conmutador 532 se transforman de regreso al dominio del tiempo utilizando una etapa de síntesis LPC 540, que se controla por medio de información LC generada y transmitida del lado del codificador. A continuación, con posterioridad al bloque 540, ambas ramas tienen información en dominio del tiempo que se conmuta de acuerdo con una señal de control del conmutador para obtener en última instancia una señal de audio tal como una señal mono, una señal estéreo o una señal de multicanal, lo que depende de la señal ingresada en el esquema de codificación de la Fig. 1a.

Fig. 1c ilustra otra realización con una disposición diferente del conmutador 521 similar a la del principio de la Fig. 4b.

Fig. 2a ilustra un esquema de codificación preferido de acuerdo con un Segundo aspecto de la presente invención. Un esquema de procesamiento común conectado a la entrada del conmutador 200 puede comprender un bloque estéreo envolvente/conjunto 101 que genera, como salida, parámetros conjuntos estéreo y una señal de salida mono, que se generan mediante el "downmix" de la señal de entrada, que es una señal que tiene dos o más canales. En general, la señal en el bloque de salida 101 también puede ser una señal con más canales, aunque debido

a la funcionalidad de downmixing del bloque 101, el número de canales a la salida del bloque 101 suele ser menor que el número de canales ingresados al bloque 101.

5 El esquema de procesamiento común puede comprender, como alternativa del bloque 101 o además del bloque 101, una etapa de extensión de ancho de banda 102. En la realización de la Fig. 2a, el bloque de salida 101 es ingresado al bloque de extensión de ancho de banda 102 que, en el codificador de la Fig. 2a, da salida a una señal limitada por la banda tal como la señal de banda baja o la señal paso bajo en su salida. De preferencia, esta señal también es submuestreada (por ejemplo, en un
10 factor de dos). Más aun, para la banda alta de la señal ingresada al bloque 102, los parámetros de extensión de ancho de banda tales como los parámetros de envoltura espectral, parámetros de filtrado inverso, parámetros de piso de ruido, etc., como se diera a conocer en el perfil HE-AAC del MPEG-4 se generan y envían a un multiplexor de secuencias de bits 800.

15 De preferencia, la etapa de decisión 300 recibe la señal ingresada al bloque 101 o ingresada al bloque 102 para decidir entre, por ejemplo, un modo de música o un modo de voz. En el modo de música, se selecciona la rama codificadora superior 400 en tanto que, en el modo de voz, se selecciona la rama codificadora inferior 500. De
20 preferencia, la etapa de decisión controla además el bloque de estéreo conjunto 101 y/o el bloque de extensión de ancho de banda 102 para adaptar la funcionalidad de estos bloques a la señal específica. Es así como, cuando la etapa de decisión determina que una determinada porción de tiempo de la señal de entrada es del primer modo tal como el modo música, luego las características específicas del
25 bloque 101 y/o del bloque 102 pueden ser controladas por la etapa de decisión 300. Por otro lado, cuando la etapa de decisión 300 determina que la señal está en modo voz o, en general, en un segundo modo en dominio LPC, luego las características específicas de los bloques 101 y 102 pueden ser controladas de acuerdo con la salida de la etapa de decisión.

30 De preferencia, la conversión espectral de la rama codificadora 400 se realiza utilizando una operación MDCT que, aun más preferentemente, es una operación MDCT con distorsión temporal, donde se puede controlar la potencia o, en general la potencia de distorsión entre una potencia de distorsión cero y una elevada. En una
35 potencia de distorsión cero, la operación MDCT del bloque 411 es una operación de MDCT directa conocida en la técnica. La potencia de distorsión temporal junto con la información secundaria de distorsión se pueden transmitir/ingresar al multiplexor de secuencias de bits 800 en forma de información secundaria.

40 En la reama codificadora LPC, el codificador en dominio LPC puede incluir un núcleo ACELP 526 que calcula una ganancia de tono, un retardo de tono y/o información de libro de codificación tal como un índice y ganancia de libro de codificación. El modo

TCX que se diera a conocer en 3GPP TS 26.290 incurre en el procesamiento de una señal perceptivamente ponderada en el dominio de la transformada. Se cuantifica una señal ponderada por transformada de Fourier utilizando una red de cuantificación multitasa dividida (VQ algebraica) con cuantificación del factor ruido. Se calcula una transformada en las ventanas de muestreo 1024, 512 o 256. Se recupera la señal de excitación mediante filtrado inverso de la señal ponderada cuantificada a través de un filtro de ponderación inversa.

En la primera rama codificadora 400, un convertidor espectral comprende preferentemente una operación MDCT específicamente adaptada que consta de ciertas funciones de ventana seguidas por una etapa de cuantificación/codificación de entropía que puede consistir en una etapa de cuantificación de vector único, aunque preferentemente es una combinación de cuantificador de escalas/codificador de entropía similar al cuantificador/codificador de la rama de codificación en el dominio de la frecuencia, es decir, del elemento 421 de la Fig. 2a.

En la segunda rama codificadora, está el bloque LPC 510 seguida por un conmutador 521, seguido a su vez por un bloque ACELP 526 o un bloque TCX 527. ACELP ha sido descrito en 3GPP TS 26.190 y TCX fue descrito en 3GPP TS 26.290. En general, el bloque ACELP 526 recibe una señal de excitación LPC calculada por un procedimiento descrito en la Fig. 7e. El bloque TCX 527 recibe una señal ponderada generada por la Fig. 7f.

En TCX, se aplica la transformada a la señal ponderada que se computa filtrando la señal de entrada a través de un filtro de ponderación basado en LPC. El filtro de ponderación utilizado en las realizaciones preferidas de la presente invención está dado por $(1 - A(z/\gamma))/(1 - \mu z^{-1})$. Por consiguiente, la señal ponderada es una señal en el dominio LPC y su transformada es un dominio espectral LPC. La señal procesada por el bloque ACELP 526 es la señal de excitación y es diferente de la señal procesada por el bloque 527, aunque ambas señales están en el dominio LPC.

En el lado del decodificador ilustrado en la Fig. 2b, después de la transformación espectral inversa efectuada en el bloque 537, se aplica la inversa del filtro de ponderación, es decir $(1 - \mu z^{-1})/(1 - A(z/\gamma))$. A continuación, se filtra la señal a través de $(1 - A(z))$ para pasar el dominio de excitación LPC. De esa manera, la conversión al bloque de dominio LPC 534 y el bloque TCX⁻¹ 537 incluyen transformación inversa y luego filtración por medio de $\frac{(1 - \mu z^{-1})}{(1 - A(z/\gamma))}(1 - A(z))$ para convertir del dominio ponderado al dominio de la excitación.

Si bien el elemento 510 de las Figs. 1a, 1c, 2a, 2c ilustra un solo bloque, el bloque 510 puede emitir diferentes señales, siempre que estas señales estén en el dominio

LPC. El modo real del bloque 510, como por ejemplo el modo de señales de excitación o el modo de señales ponderadas, puede depender del estado real de conmutación. Por otro lado, el bloque 510 puede tener dos dispositivos de procesamiento paralelos, donde un dispositivo está implementado de manera similar a la Fig. 7e y el otro dispositivo está implementado como en la Fig. 7f. Por ende, el dominio LPC a la salida de 510 puede representar la señal de excitación LPC o la señal ponderada LPC o cualquier otra señal de dominio LPC.

En la segunda rama codificadora (ACELP/TCX) de la Fig. 2a o 2c, preferentemente se preenfatisa la señal por medio de un filtro $1 - 0.68z^{-1}$ antes de la codificación. En el decodificador ACELP/TCX de la Fig. 2b se desenfatisa la señal acentuada con el filtro $1/(1 - 0.68z^{-1})$. La preénfasis puede ser parte del bloque LPC 510 en que se preenfatisa la señal antes del análisis LPC y cuantificación. De modo similar, la desénfasis puede ser parte del bloque de síntesis LPC LPC^{-1} 540.

Fig. 2c ilustra otra realización de la implementación de la Fig. 2a, aunque con una disposición diferente del conmutador 521 similar al concepto de la Fig. 4b.

En una realización preferida, el primer conmutador 200 (ver la Fig. 1a o 2a) es controlado por una decisión de bucle abierto (como en la Fig. 4a) y el segundo conmutador es controlado por medio de una decisión de bucle cerrado (como en la figura 4b).

Por ejemplo, la Fig. 2c presenta el Segundo conmutador colocado después de las ramas ACELP y TCX como en la Fig. 4b. Luego, en la primera rama de procesamiento, el primer dominio LPC representa la excitación LPC y en la segunda rama de procesamiento, el Segundo dominio LPC representa la señal ponderada LPC. Es decir que la primera señal en dominio LPC se obtiene mediante filtrado por $(1 - A(z))$ para convertirla al dominio residual LPC, en tanto que la segunda señal en dominio LPC se obtiene mediante filtrado por el filtro $(1 - A(z/\gamma))/(1 - \mu z^{-1})$ para la conversión al dominio ponderado LPC.

La Fig. 2b ilustra un esquema de decodificación correspondiente al esquema de codificación de la Fig. 2a. La secuencia de bits generada por el multiplexor de secuencias de bits 800 de la Fig. 2a es ingresada al desmultiplexor de secuencias de bits 900. Dependiendo de una información derivada, por ejemplo, secuencia de bits a través de un bloque de detección de modos 601, se controla un conmutador del lado del decodificador 600 para retransmitir señales procedentes de la rama superior o señales procedentes de la rama inferior al bloque de extensión de ancho de banda 701. El bloque de extensión de ancho de banda 701 recibe, desde el desmultiplexor de secuencias de bits 900, información secundaria y, basándose en esta información

secundaria y en la salida de la decisión de modo 601, reconstruye la banda alta sobre la base de la banda baja emitida por el conmutador 600.

5 La señal de banda completa generada por el bloque 701 es ingresada a la etapa de procesamiento de estéreo conjunto/envolvente 702, que reconstruye dos canales estéreo o varios multicanales. En términos generales, el bloque 702 produce como salida más canales de los ingresados a este bloque. Dependiendo de la aplicación, la entrada al bloque 702 puede incluir incluso dos canales tales como en un modo estéreo y puede incluir aun más canales, siempre que la salida producida por este
10 bloque tenga más canales que la entrada a este bloque.

Se ha demostrado que el conmutador 200 conmuta entre ambas ramas de manera que solo una rama reciba una señal para procesar y la otra rama no reciba la señal para procesar. Sin embargo, en una realización alternativa, el conmutador también
15 puede estar dispuesto con posterioridad, por ejemplo, al codificador de audio 421 y al codificador de excitación 522, 523, 524, lo que significa que ambas ramas 400, 500 procesan la misma señal en paralelo. Para no duplicar la tasa de transferencia de bits, sin embargo, sólo se selecciona la señal de salida de una de esas ramas codificadoras 400 o 500 para escribirla en la secuencia de bits de salida. La etapa de
20 decisión funciona luego de manera que la señal escrita en la secuencia de bits minimice cierta función de costo, donde la función de costo puede ser la tasa de transferencia de bits generada o la distorsión perceptiva generada o una función de costo combinada de tasa/distorsión. Por lo tanto, ya sea en este modo o en el modo ilustrado en las Figuras, la etapa de decisión también puede funcionar en un modo
25 de bucle cerrado para asegurarse de que, por último, sólo se escriba la salida de la rama codificadora en la secuencia de bits, que tiene la menor tasa de transferencia de bits para la distorsión perceptiva o, para una determinada tasa de transferencia de bits, tiene la menor distorsión perceptiva. En el modo de bucle cerrado, se puede derivar la realimentación de entrada de las salidas de los tres bloques de
30 cuantificador/escalador 421, 522 y 424 de la Fig. 1a.

En la implementación con dos conmutadores, es decir el primer conmutador 200 y el Segundo conmutador 521, es preferible que la resolución en el tiempo correspondiente al primer conmutador sea menor que la resolución en tiempo
35 correspondiente al segundo conmutador. En otras palabras, los bloques de la señal de entrada al primer conmutador, que puede ser conmutado por medio de una operación de conmutación, son más grandes que los bloques conmutados por el segundo conmutador que opera en el dominio LPC. A título de ejemplo, el conmutador de dominio de la frecuencia/dominio LPC 200 puede conmutar bloques
40 de una longitud de 1024 muestras y el segundo conmutador puede conmutar bloques con 256 muestras cada uno.

Si bien algunas de las Figs. 1a a 10b inclusive se ilustran en forma de diagramas de bloques de un aparato; estas figuras son simultáneamente una ilustración de un método, donde las funcionalidades de los bloques corresponden a los pasos del método.

5

La Fig. 3a ilustra un codificador de audio para generar una señal de audio codificada como salida de la primera rama codificadora 400 y una segunda rama codificadora 500. Más aun, la señal de audio codificada incluye preferentemente información secundaria tal como parámetros de pre-procesamiento de la etapa de pre-procesamiento común o, como se describe con respecto a las anteriores Figs., información de control del conmutador.

De preferencia, la primera rama codificadora es operativa para codificar una señal de audio intermedia 195 de acuerdo con un primer algoritmo de codificación, donde el primer algoritmo de codificación tiene un modelo de destino de la información. La primera rama codificadora 400 genera la primera señal de salida del codificador que es una representación de información espectral codificada de la señal de audio intermedia 195.

Más aun, la segunda rama codificadora 500 está adaptada para codificar la señal de audio intermedia 195 de acuerdo con un segundo algoritmo de codificación, donde el segundo algoritmo de codificación tiene un modelo de origen de la información y genera, en una segunda señal de salida del codificador, parámetros codificados para el modelo de origen de la información que representa la señal de audio intermedia.

25

El codificador de audio comprende además la etapa de pre-procesamiento común para el pre-procesamiento de una señal de entrada de audio 99 para obtener la señal intermedia de audio 195. Específicamente, la etapa de pre-procesamiento común cumple la función de procesar la señal de entrada de audio 99 de manera que la señal intermedia de audio 195, es decir, la salida del algoritmo de pre-procesamiento común sea una versión comprimida de la señal de entrada de audio.

Un método preferido de codificación de audio para generar una señal de audio codificada comprende un paso de codificar 400 una señal de audio intermedia 195 de acuerdo con un primer algoritmo de codificación, donde el primer algoritmo de codificación tiene un modelo de destino de la información y genera, en una primera señal de salida, información espectral codificada que representa la señal de audio, un paso de codificación 500 de una señal de audio intermedia 195 de acuerdo con un segundo algoritmo de codificación, donde el segundo algoritmo de codificación tiene un modelo de origen de la información y genera, en una segunda señal de salida, parámetros codificados para el modelo de origen de la información que representa la señal intermedia 195, y un paso de pre-procesamiento común 100 de una señal de

entrada de audio 99 para obtener la señal de audio intermedia 195, donde, en el paso de pre-procesamiento común se procesa la señal de entrada de audio 99 de manera que la señal de audio intermedia 195 sea una versión comprimida de la señal de entrada de audio 99, donde la señal de audio codificada incluye, para una
5 determinada porción de la señal de audio, la primera señal de salida o la segunda señal de salida. El método incluye preferentemente el paso adicional de codificar una cierta porción de la señal de audio intermedia utilizando el primer algoritmo de codificación o utilizando el segundo algoritmo de codificación, o bien codificando la señal utilizando ambos algoritmos y dando salida, en una señal codificada, el
10 resultado del primer algoritmo de codificación o el resultado del segundo algoritmo de codificación.

En general, el algoritmo de codificación de audio empleado en la primera rama codificadora 400 refleja y modela la situación en un destino de audio. El destino de
15 una información de audio es normalmente el oído humano. El oído humano puede ser modelado como analizador de frecuencia. Por lo tanto, la primera rama codificadora emite información espectral codificada. De preferencia, la primera rama codificadora incluye además un modelo psicoacústico para aplicar adicionalmente un umbral de enmascaramiento psicoacústico. Este umbral de enmascaramiento
20 psicoacústico se utiliza cuando se cuantifican valores espectrales de audio en que, preferentemente, la cuantificación se realiza de tal manera que se introduce un ruido de cuantificación por la cuantificación de valores espectrales de audio, que se ocultan debajo del umbral de enmascaramiento psicoacústico.

25 La segunda rama codificadora representa un modelo de origen de la información, que refleja la generación de un sonido de audio. Por lo tanto, los modelos de origen de la información pueden incluir un modelo de voz humana que está reflejado por una etapa de análisis LPC, es decir transformando una señal en el dominio del tiempo en un dominio LPC y seguidamente procesando la señal residual LPC, es
30 decir, la señal de excitación. Otros modelos alternativos de origen del sonido son, sin embargo, modelos origen del sonido para representar un determinado instrumento o cualquier otro generador de sonido tal como una fuente de sonido específica existente en el mundo real. Se puede ejecutar una selección entre diferentes modelos de origen del sonido cuando se dispone de varios modelos de origen del
35 sonido, por ejemplo, basándose en un cálculo SNR, es decir basado en un cálculo de cuál de los modelos de origen es el más adecuado para codificar una determinada porción de tiempo y/o porción de frecuencia de una señal de audio. Sin embargo, es preferible que la conmutación entre ramas codificadoras se ejecute en el dominio del tiempo, es decir que cierta porción de tiempo se codifica utilizando un
40 modelo y una determinada porción diferente de la señal intermedia se codifica utilizando la otra rama codificadora.

Los modelos de origen de la información están representados por ciertos parámetros. Con respecto al modelo de voz, los parámetros son parámetros LPC y parámetros de excitación codificada cuando se considera un codificador de voz moderno tal como AMR-WB+. El AMR-WB+ comprende un codificador ACELP y un
5 codificador TCX. En este caso, los parámetros de excitación codificada pueden ser la ganancia global, piso de ruido y códigos de longitud variable.

La Fig. 3b ilustra un decodificador que corresponde al codificador ilustrado en la Fig. 3a. En general, Fig. 3b ilustra un decodificador de audio para decodificar una
10 de audio codificada para señal de audio decodificada 799. El decodificador incluye la primera rama decodificadora 450 para decodificar una señal de audio codificada de acuerdo con un primer algoritmo de codificación que tiene un modelo de destino de la información. El decodificador de audio incluye además una segunda rama decodificadora 550 para decodificar una señal de información codificada de acuerdo
15 con un segundo algoritmo de codificación que tiene un modelo de origen de la información. El decodificador de audio incluye asimismo un combinador para combinar señales de salida de la primera rama decodificadora 450 y la segunda rama decodificadora 550 para obtener una señal combinada. Se ingresa la señal combinada que se ilustra en la Fig. 3b como señal de audio intermedia decodificada
20 699 a una etapa común de post-procesamiento para postprocesar la señal de audio intermedia dedodificada 699, que es la señal de salida combinada por el combinador 600, por lo que una señal de salida de la etapa común de pre-procesamiento es una versión expandida de la señal combinada. Por consiguiente, la señal de audio decodificada 799 tiene un contenido de información aumentado en
25 comparación con la señal de audio intermedia decodificada 699. Esta expansión de la información es generada por la etapa de post procesamiento común con la ayuda de parámetros de pre/post procesamiento que se pueden transmitir de un codificador a un decodificador, o que se pueden obtener de la señal de audio intermedia decodificada en sí. Sin embargo, los parámetros de pre/post procesamiento se
30 transmiten preferentemente de un codificador a un decodificador, puesto que este procedimiento da lugar a una calidad mejorada de la señal de audio decodificada.

La Fig. 3c ilustra un codificador de audio para codificar una señal de entrada de audio 195, que puede ser igual a una señal de audio intermedia 195 de la Fig. 3a de
35 acuerdo con la realización preferida de la presente invención. La señal de audio de entrada 195 está presente en un primer dominio que puede ser, por ejemplo, el dominio del tiempo, aunque también puede ser cualquier otro dominio tal como un dominio de frecuencia, un dominio LPC, un dominio espectral LPC o cualquier otro dominio. En general, la conversión de un dominio a otro dominio se ejecuta mediante
40 un algoritmo de conversión tal como cualquier algoritmo de conversión de tiempo/frecuencia o algoritmo de conversión frecuencia/tiempo muy conocido.

Una transformación alternativa del dominio del tiempo, por ejemplo en el dominio LPC, es el resultado del filtrado LPC de una señal en dominio de tiempo que da lugar a una señal residual LPC o señal de excitación. Cualquier otra operación de filtrado que produzca una señal filtrada tiene impacto sobre un número sustancial de muestras de señales antes de usar la transformación como algoritmo de transformación según el caso. Por lo tanto, la ponderación de una señal de audio utilizando un filtro de ponderación basado en LPC es una transformación adicional, que genera una señal en el dominio LPC. En una transformación de tiempo/frecuencia, la modificación de un solo valor espectral influye sobre todos los valores en dominio del tiempo antes de la transformación. De manera análoga, una modificación de cualquier muestra en dominio del tiempo tiene efecto sobre cada muestra en dominio de frecuencia. Del mismo modo, una modificación de una muestra de la señal de excitación en una situación en dominio LPC tiene efecto, debido a la longitud del filtro LPC, sobre un número sustancial de muestras antes del filtrado LPC. Del mismo modo, una modificación de una muestra antes de una transformación LPC tiene impacto sobre muchas muestras obtenidas mediante esta transformación LPC debido al efecto de memoria inherente del filtro LPC.

El codificador de audio de Fig. 3c incluye una primera rama codificadora 400 que genera a primera señal codificada. Esta primera señal codificada puede estar en un cuarto dominio que es, en la realización preferida, el dominio espectral en tiempo, es decir, el dominio que se obtiene de una señal en dominio del tiempo mediante una conversión tiempo/frecuencia.

Por lo tanto, la primera rama codificadora 400 para codificar una señal de audio utiliza un primer algoritmo de codificación para obtener una primera señal codificada, donde este primer algoritmo de codificación puede o no incluir un algoritmo de conversión tiempo/frecuencia.

El codificador de audio incluye asimismo una segunda rama codificadora 500 para codificar una señal de audio. La segunda rama codificadora 500 utiliza un segundo algoritmo de codificación para obtener una segunda señal codificadora, que es diferente del primer algoritmo de codificación.

El codificador de audio incluye además un primer conmutador 200 para conmutar entre la primera rama codificadora 400 y la segunda rama codificadora 500 de manera que para una porción de la señal de audio de entrada, la primera señal codificada a la salida del bloque 400 o la segunda señal codificada a la salida de la segunda rama codificadora esté incluida en una señal de salida del codificador. Por consiguiente, cuando para una determinada porción de la señal de audio de entrada 195, la primera señal codificada en el cuarto dominio está incluida en la señal de salida del codificador, la segunda señal codificada que es la primera señal

procesada en el segundo dominio o la segunda señal procesada en el tercer dominio, no está incluida en la señal de salida del codificador. Esto garantiza que este codificador sea eficiente en cuanto a la tasa de transferencia de bits. En algunas realizaciones, cualquier porción de tiempo de la señal de audio que esté
5 incluida en dos señales codificadas diferentes es pequeña en comparación con una longitud de marco de un marco, como se describe con respecto a la Fig. 3e. Estas pequeñas porciones son útiles para un desvanecimiento cruzado de una señal codificada a la otra señal codificada en el caso de un evento de conmutación para reducir las aberraciones que se podrían producir sin desvanecimiento cruzado. Por
10 lo tanto, aparte de la región de desvanecimiento cruzado, cada bloque de dominio del tiempo está representado por una señal codificada de sólo un único dominio.

Como se ilustra en la Fig. 3c, la segunda rama codificadora 500 comprende un convertidor 510 para convertir la señal de audio en el primer dominio, es decir, la
15 señal 195, en un segundo dominio. Más aun, la segunda rama codificadora 500 comprende una primera rama de procesamiento 522 para procesar una señal de audio en el segundo dominio a fin de obtener una primera señal procesada que está también, preferentemente, en el segundo dominio de manera que la primera rama de procesamiento 522 no ejecute un cambio de dominio.

20 La segunda rama codificadora 500 comprende además una segunda rama de procesamiento 523, 524 que convierte la señal de audio en el segundo dominio a un tercer dominio, que es diferente del primer dominio y que también es diferente del segundo dominio y que procesa la señal de audio en el tercer dominio para obtener
25 una segunda señal procesada a la salida de la segunda rama de procesamiento 523, 524.

Más aun, la segunda rama codificadora comprende un segundo conmutador 521 para conmutar entre la primera rama de procesamiento 522 y la segunda rama de
30 procesamiento 523, 524 de manera que, respecto de una porción de la señal de audio ingresada a la segunda rama codificadora, la primera señal procesada en el segundo dominio o la segunda señal procesada en el tercer dominio esté en la segunda señal codificada.

35 La Fig. 3d ilustra un correspondiente decodificador para decodificar una señal de audio codificada generada por el codificador de la Fig. 3c. En general, cada bloque de la señal de audio en el primer dominio está representado por una señal en el segundo dominio, una señal en el tercer dominio o una señal codificada en el cuarto dominio aparte de una región opcional de desvanecimiento cruzado que
40 preferentemente es corta en comparación con la longitud de un marco para obtener un sistema que está, en lo posible, en el límite de muestreo crítico. La señal de audio codificada incluye la primera señal codificada, una segunda señal codificada en un

segundo dominio y una tercera señal codificada en un tercer dominio, donde la primera señal codificada, la segunda señal codificada en el segundo dominio y una tercera señal codificada en un tercer dominio, donde la primera señal codificada, la segunda señal codificada y la tercera señal codificada se relacionan, en todos los casos, con diferentes porciones de tiempo de la señal de audio decodificada y donde el segundo dominio, el tercer dominio y el primer dominio de una señal de audio decodificada son diferentes entre sí.

El decodificador comprende una primera rama decodificadora para decodificar sobre la base del primer algoritmo de codificación. La primera rama decodificadora está indicada en 431, 440 en la Fig. 3d y preferentemente comprende un convertidor de frecuencia/tiempo. primera señal codificada está preferentemente en un cuarto dominio y se convierte al primer dominio que es el dominio correspondiente a la señal de salida decodificada.

El decodificador de la Fig. 3d comprende además una segunda rama decodificadora que comprende varios elementos. Estos elementos son una primera rama de procesamiento inverso 531 para el procesamiento inverso de la segunda señal codificada a fin de obtener una primera señal procesada inversa en el segundo dominio a la salida del bloque 531. La segunda rama decodificadora comprende además una segunda rama de procesamiento inverso 533, 534 para el procesamiento inverso de una tercera señal codificada para obtener una segunda señal procesada inversa en el segundo dominio, donde la segunda rama de procesamiento inverso comprende un convertidor para la conversión del tercer dominio al segundo dominio.

La segunda rama decodificadora comprende además un primer combinador 532 para combinar la primera señal procesada inversa y la segunda señal procesada inversa a fin de obtener una señal en el segundo dominio, donde esta señal combinada sólo está influenciada, en el primer instante de tiempo, por la primera señal con procesamiento inverso y, en un instante de tiempo posterior, sólo está influenciada por la segunda señal procesada inversa.

La segunda rama decodificadora comprende además un convertidor 540 para convertir la señal combinada al primer dominio.

Por último, el decodificador ilustrado en la Fig. 3d comprende un segundo combinador 600 para combinar la primera señal decodificada procedente del bloque 431, 440 y la señal de salida del convertidor 540 para obtener una señal de salida decodificada en el primer dominio. Una vez más, la señal de salida decodificada en el primer dominio sólo está influenciada, en un primer instante, por la señal emitida

por el convertidor 540 y, en un instante posterior, sólo está influenciada por la primera señal decodificada saliente del bloque 431, 440.

Esta situación está representada, desde el punto de vista del codificador, en la Fig. 3e. La porción superior de la Fig. 3e ilustra en la representación esquemática, una señal de audio en un primer dominio, como por ejemplo una señal de audio en el dominio del tiempo, donde el índice de tiempo aumenta de izquierda a derecha y el elemento 3 se podría considerar una secuencia de muestras de audio que representan la señal 195 de la Fig. 3c. La Fig. 3e ilustra los marcos 3a, 3b, 3c, 3d que se pueden generar conmutando entre la primera señal codificada y la primera señal procesada y la segunda señal procesada ilustrada en el ítem 4 de la Fig. 3e. La primera señal codificada, la primera señal procesada y la segunda señal procesada están en todos los casos en dominios diferentes y, para garantizar que la conmutación entre los diferentes dominios no de origen a una aberración en el lado del decodificador, los marcos 3a, 3b de la señal en dominio del tiempo tienen un rango de traslape que está indicado como región de desvanecimiento cruzado y dicha región de desvanecimiento cruzado está en el marco 3d y en el 3c. Sin embargo, no existe ninguna región de desvanecimiento cruzado entre el marco 3d, 3c, lo que significa que el marco 3d también está representado por una segunda señal procesada, es decir, una señal en el tercer dominio, y no hay cambio de dominio alguno entre el marco 3c y el 3d. Por lo tanto, en general es preferible no incluir una región de desvanecimiento cruzado cuando no hay cambio de dominio e incluir una región de desvanecimiento cruzado, es decir, una porción de la señal de audio que está codificada por dos señales codificadas/procesadas subsiguientes cuando hay cambio de dominio, es decir, una acción de conmutación de uno de los dos conmutadores. De preferencia, se ejecutan desvanecimientos cruzados para otros cambios de dominio.

En la realización en la cual la primera señal codificada o la segunda señal procesada ha sido generada por un procesamiento MDCT que presenta, por ejemplo, 50 por ciento de traslape, cada muestra en el dominio del tiempo está incluida en dos marcos subsiguientes. Sin embargo, debido a las características del MDCT, esto no da origen a una sobrecarga, puesto que el MDCT es un sistema con muestreo crítico. En este contexto, muestreo crítico significa que el número de valores espectrales es igual al número de valores en dominio del tiempo. El MDCT es ventajosa porque el efecto de cruce de frecuencias tiene lugar sin una región de cruce de frecuencias específica por lo que el cruce de un bloque MDCT al siguiente bloque MDCT se realiza sin sobrecarga alguna que violaría el requisito de muestreo crítico.

De preferencia, el primer algoritmo de codificación en la primera rama codificadora se basa en un molo de destino de la información y el segundo algoritmo de

codificación de la segunda rama codificadora se basa en una fuente de información o un modelo SNR. Un modelo SNR es un modelo que no se relaciona específicamente con un mecanismo específico de generación de sonido pero que es un modo de codificación que se puede seleccionar entre una pluralidad de modos de codificación basándose, por ejemplo, en una decisión en bucle cerrado. Por consiguiente, un modelo SNR es cualquier modelo de codificación adecuado pero que no necesariamente tiene que estar relacionado con la constitución física del generador de sonido, sino que es cualquier modelo de codificación parametrizado diferente del modelo de destino de la información, que se puede seleccionar en virtud de una decisión de bucle cerrado y, específicamente, comparando diferentes resultados SNR de diferentes modelos.

Como se ilustra en la Fig. 3c, se incluye un controlador 300, 525. Este controlador puede incluir las funcionalidades de la etapa de decisión 300 de la Fig. 1a y, además, puede incluir la funcionalidad del dispositivo de control del conmutador 525 de la Fig. 1a. En términos generales, el controlador sirve para controlar el primer conmutador y el segundo conmutador de manera adaptativa a la señal. El controlador cumple la función de analizar una señal ingresada al primer conmutador o saliente de la primera o la segunda rama codificadora o señales obtenidas codificando y decodificando de la primera y la segunda ramas codificadoras con respecto a una función objetivo. Por otro lado o además, el controlador cumple la función de analizar la señal ingresada al segundo conmutador o salida de la primera rama de procesamiento o la segunda rama de procesamiento u obtenida mediante el procesamiento y procesamiento inverso de la primera rama de procesamiento y la segunda rama de procesamiento, una vez más con respecto a una función objetivo.

En una realización, la primera rama codificadora o la segunda rama codificadora comprende un solapamiento que introduce un algoritmo de conversión de tiempo/frecuencia tal como un algoritmo MDCT o MDST, que es diferente de la transformación FFT directa, que no introduce efecto de solapamiento. Más aun, una o ambas ramas comprenden un bloque cuantificador/codificador de entropía. Específicamente, sólo la segunda rama de procesamiento de la segunda rama codificadora incluye el convertidor de tiempo/frecuencia que introduce una operación de solapamiento y la primera rama de procesamiento de la segunda rama codificadora comprende un cuantificador/codificador de entropía que no introduce efectos de solapamiento. El convertidor tiempo/frecuencia que introduce solapamiento comprende preferentemente un generador de ventanas para aplicar una ventana de análisis y un algoritmo de transformación MDCT. Específicamente, el generador de ventanas cumple la función de aplicar la función ventana a los marcos subsiguientes en forma traslapada de manera que una muestra de la señal en ventana tenga lugar por lo menos en dos marcos de ventana subsiguientes.

En una realización, la primera rama de procesamiento comprende un codificador ACELP y una segunda rama de procesamiento comprende un convertidor espectral MDCT y el cuantificador para cuantificar los componentes espectrales a fin de obtener componentes espectrales cuantificados, donde cada componente espectral
5 cuantificado es cero o está definido por un índice cuantificador de la pluralidad de índices cuantificadores diferentes posibles.

Más aun, es preferible que el conmutador 200 funcione en bucle abierto y el segundo conmutador funcione en forma de bucle cerrado.

10

Como se indicara anteriormente, ambas ramas codificadoras cumplen la función de codificar la señal de audio en forma de bloques, donde el primer conmutador o el segundo conmutador conmuta en forma de bloques de manera que la acción de conmutación tenga lugar, como mínimo, después de un bloque de un número
15 predeterminado de muestras de una señal, donde el número predeterminado forma una longitud de marco para el correspondiente conmutador. De esa manera, el gránulo para conmutar correspondiente al primer conmutador puede ser, por ejemplo, un bloque de 2048 o 1028 muestras y la longitud de marco, sobre la base de la cual conmuta el primer conmutador 200, puede ser variable, aunque
20 preferentemente es fijo de ese período bastante prolongado.

A diferencia de esto, la longitud del bloque correspondiente al segundo conmutador 521, es decir, cuando el segundo conmutador 521 cambia de un modo al otro, es sustancialmente menor que la longitud del bloque correspondiente al primer
25 conmutador. De preferencia, ambas longitudes de bloque son seleccionadas de tal manera que la longitud del bloque más largo sea un múltiplo entero de la longitud del bloque más corto. En la realización preferida, la longitud del bloque del primer conmutador es de 2048 o 1024 y la longitud del bloque del segundo conmutador es de 1024 o, más preferentemente, 512 y aun más preferentemente 256, e incluso
30 más preferentemente 128 muestras, de manera que, como máximo, el segundo conmutador puede conmutar 16 veces cuando el primer conmutador conmuta sólo una vez. Sin embargo, una relación máxima preferida de longitudes de bloque es de 4:1.

En otra realización, el controlador 300, 525 cumple la función de ejecutar una discriminación de voz o música correspondiente al primer conmutador de manera tal que se favorezca una decisión por voz con respecto a una decisión por música. En esta realización, se toma incluso una decisión por voz aun cuando una porción inferior a 50% de un marco correspondiente al primer conmutador sea de voz y la
40 porción de más de 50% del marco sea de música.

Más aun, el controlador cumple la función para conmutar inmediatamente al modo voz, cuando una pequeña porción del primer marco sea de voz y, específicamente, cuando una porción del primer marco sea de voz, que es el 50% de la longitud del segundo marco de menor tamaño. Por consiguiente, una dedición preferida de
5 favorecimiento de la voz ya conmuta a voz cuando, por ejemplo, sólo 6% o 12% de un bloque que corresponde a la longitud de marco del primer conmutador es de voz.

Este procedimiento se realiza preferentemente para explotar por completo la capacidad de ahorro de tasa de transferencia de bits de la primera rama de
10 procesamiento, que tiene un núcleo de habla de voz en una realización, y para no perder calidad incluso en el resto de un primer marco de gran tamaño, que no es de voz debido al hecho de que la segunda rama de procesamiento incluye un convertidor y, por lo tanto, sirve para señales de audio que también tienen señales que no son de voz. De preferencia, esta segunda rama de procesamiento incluye un
15 MDCT traslapado, que se muestrea en forma crítica y que, incluso en tamaños menores de ventana otorga una operación sumamente eficiente y sin solapamientos debido al procesamiento de cancelación de solapamiento en dominio del tiempo tal como traslape y a la suma de un lado de decodificador. Más aun, es ventajosa una gran longitud de bloque correspondiente a la primera rama codificadora que es
20 preferentemente una rama codificadora MDCT similar a AAC, puesto que las señales que no son de voz normalmente son bastante fijas y una gran ventana de transformación produce una resolución de alta frecuencia y, por lo tanto, gran calidad y, además, otorga eficiencia de tasa de transferencia de bits debido a un módulo de cuantificación controlado por psicoacústica, que también se puede aplicar
25 al modo de codificación basado en transformada en la segunda rama de procesamiento de la segunda rama codificadora.

Con respecto a la ilustración del decodificador de la Fig. 3d, es preferible que la señal transmitida incluya un indicador explícito como información secundaria 4a de
30 acuerdo con lo ilustrado en la Fig. 3e. Esta información secundaria 4a es extraída por un analizador de tasa de transferencia de bits que no se ilustra en la Fig. 3d para reenviar la correspondiente primera señal codificada, primera señal procesada o segunda señal procesada al procesador apropiado tal como la primera rama decodificadora, la primera rama de procesamiento inverso o la segunda rama de
35 procesamiento inverso en la Fig. 3d. Por lo tanto, una señal codificada no sólo tiene las señales codificadas/procesadas sino que también incluye información secundaria relativa a estas señales. En otras realizaciones, sin embargo, puede haber una señalización implícita que permita al analizador de tasa de transferencia de bits del lado del decodificador distinguir entre las señales determinadas. En cuanto a la Fig.
40 3e, se indica que la primera señal procesada o la segunda señal procesada es la salida de la segunda rama codificadora y, por lo menos, la segunda señal codificada.

De preferencia, la primera rama decodificadora y/o la segunda rama de procesamiento inverso incluyen una transformación MDCT para convertir del dominio espectral al dominio del tiempo. Para este fin, se presenta un sumador de traslape para ejecutar una funcionalidad de cancelación de solapamiento en dominio del tiempo que, al mismo tiempo, produce un efecto de desvanecimiento cruzado para evitar aberraciones de bloqueo. En general, la primera rama decodificadora convierte una señal codificada en el cuarto dominio al primer dominio, en tanto que la segunda rama de procesamiento inverso ejecuta una conversión del tercer dominio al segundo dominio y seguidamente el controlador conectado al primer combinador produce la conversión del segundo dominio al primer dominio de manera que, a la entrada del combinador 600, sólo estén presentes las señales del primer dominio que representan, en la realización de la Fig. 3d, la señal de salida decodificada.

Las Fig. 4a y 4b ilustran dos realizaciones diferentes, que difieren en cuanto a la ubicación del conmutador 200. En la Fig. 4a, el conmutador 200 está situado entre una salida de la etapa de pre-procesamiento común 100 y la entrada de las dos ramas codificadas 400, 500. La realización de la Fig. 4a se asegura de que la señal de audio sea ingresada sólo a una rama codificadora y que la otra rama codificadora, que no está conectada a la salida de la etapa de pre-procesamiento común no esté en funcionamiento y que, por lo tanto, esté apagada o en modo de hibernación. Esta realización es preferible por el hecho de que la rama codificadora no activa no consume energía ni recursos informáticos, lo que es ventajoso en el caso de aplicaciones móviles especialmente, que funcionan con baterías y que, por lo tanto, tienen la limitación general del consumo de energía.

Sin embargo, por otro lado, la realización de la Fig. 4b puede ser preferible cuando el consumo de energía no constituye una preocupación. En esta realización, ambas ramas codificadoras 400, 500 permanecen activas en todo momento y sólo la salida de la rama codificadora seleccionada para una determinada porción de tiempo y/o una determinada porción de frecuencia es enviada al formateador de secuencias de bits que puede estar implementado en forma de multiplexor de secuencias de bits 800. Por lo tanto, en la realización de la Fig. 4b, ambas ramas codificadoras están activas en todo momento y la salida de una rama codificadora seleccionada por la etapa de decisión 300 es ingresada a la secuencia de bits de salida, en tanto que la salida de la otra rama codificadora no seleccionada 400 es descartada, es decir que no se la ingresa a la secuencia de bits de salida, es decir la señal de audio codificada.

De preferencia, la segunda norma de codificación/norma de decodificación es un algoritmo de codificación basado en LPC. En la codificación de voz basada en LPC, se realiza una diferenciación entre segmentos de señal o porciones de señal de

excitación del tipo de impulsos casi periódicos y segmentos de señal o porciones de señal de excitación del tipo ruido. Esto es ejecutado para los vocodificadores (codificadores de señales vocales) LPC de baja tasa de transferencia de bits (2.4 kbps) como el de la Fig 7b. Sin embargo, en los codificadores CELP de mediana
5 tasa de transferencia de bits, se obtiene la excitación para la suma de vectores escalados de un libro de códigos adaptativo y un libro de códigos fijo.

Los segmentos de señal de excitación del tipo de impulsos casi periódicos, es decir los segmentos de señal que tienen un tono específico son codificados con
10 mecanismos diferentes de los que codifican las señales de excitación tipo ruido. Si bien las señales de excitación del tipo de impulsos casi periódicos están conectadas al lenguaje con voz, las señales de tipo ruido están relacionadas con el lenguaje sin voz.

15 A título de ilustración, se hace referencia a las Figs. 5a a 5d. En ese caso, se describen a manera de ejemplo los segmentos de señal o porciones de señal de excitación del tipo de impulsos casi periódicos y los segmentos de señal o porciones de señal de excitación del tipo ruido. Específicamente, se describe un lenguaje con voz ilustrado en la Fig. 5a en el dominio del tiempo y en la Fig. 5b en el dominio de la
20 frecuencia como ejemplo de porción de señal del tipo de impulsos casi periódicos y un segmento de lenguaje no hablado como ejemplo de una porción de señal del tipo ruido en conexión con las Figs. 5c y 5d. El habla se puede clasificar, en general, como con voz, sin voz o mixta. En la Fig. 5a a 5d se ilustra un trazado de los dominios de tiempo y frecuencia correspondientes a segmentos con voz y sin voz
25 muestreados. El lenguaje hablado es casi periódico en el dominio del tiempo y estructurado armónicamente en el dominio de la frecuencia, en tanto que el lenguaje sin voz es aleatorio y de banda ancha. El corto espectro de tiempo del lenguaje con voz se caracteriza por su estructura formante de armonía fina. La estructura armónica fina es una consecuencia de la casi periodicidad del habla y se la puede
30 atribuir a la vibración de las cuerdas vocales. La estructura formante (envoltura espectral) se debe a la interacción de la fuente y los tractos vocales. Los tractos vocales consisten en la faringe y la cavidad oral. La forma de la envoltura espectral que "se ajusta" al corto espectro de tiempo del habla con voz está asociada a las características de transferencia del tracto vocal y la inclinación espectral (6 dB /Octava) debido al pulso de la glotis. La envoltura espectral se caracteriza por una
35 serie de picos que se denominan formantes. Los formantes son los modos resonantes del tracto vocal. Para el tracto vocal medio hay tres a cinco formantes por debajo de 5 kHz. Las amplitudes y ubicaciones de los tres primeros formantes, habitualmente inferiores a 3 kHz, son bastante importantes tanto en la síntesis como
40 en la percepción del habla. Los formantes superiores también son importantes para las representaciones de banda ancha y habla sin voz. Las propiedades del habla están relacionadas con el siguiente sistema físico de producción de habla. El

lenguaje con voz se produce mediante la excitación del tracto vocal con pulsos casi periódicos de aire de la glotis generados por las cuerdas vocales que vibran. La frecuencia de los pulsos periódicos se denomina frecuencia fundamental o tono. El habla sin voz se produce forzando aire a través de un estrechamiento del tracto vocal. Los sonidos nasales se deben al acoplamiento acústico del tracto nasal al tracto vocal y los sonidos explosivos se producen liberando abruptamente la presión de aire que se acumulara detrás del cierre del tracto.

Por consiguiente, una porción de tipo ruido de la señal de audio no exhibe ni estructura en el dominio del tiempo del tipo impulsos ni estructura armónica en el dominio de la frecuencia como se ilustra en la Fig. 5c y en la Fig. 5d, que es diferente de la porción del tipo de impulsos casi periódicos como se ilustra en la Fig. 5a y en la Fig. 5b. Sin embargo, como se esboza más adelante, la diferenciación entre las porciones del tipo ruido y las porciones del tipo de impulsos casi periódicos también se pueden observar después de un LPC correspondiente a la señal de excitación. El LPC es un método que modela el tracto vocal y extrae de la señal la excitación de los tractos vocales.

Más aun, las porciones del tipo de impulsos casi periódicos y las porciones del tipo ruido pueden tener lugar en forma oportuna, es decir, que significa que una porción de la señal de audio en el tiempo es ruidosa y otra porción de la señal de audio es casi periódica, es decir, tonal. Por otro lado, o además, la característica de una señal puede ser diferente en diferentes bandas de frecuencia. Por consiguiente, la determinación de si la señal de audio es de ruido o tonal, también se puede realizar por selección de frecuencia de manera que se considera que una determinada banda de frecuencia o varias bandas de frecuencia determinadas son de ruido y otras bandas de frecuencia se consideran tonales. En este caso, una determinada porción de tiempo de la señal de audio podría incluir componentes tonales y componentes de ruido.

La Fig. 7a ilustra un modelo lineal de un sistema de producción de habla. Este sistema supone una excitación en dos etapas, es decir, un tren de impulsos correspondiente al habla con voz como se indica en la Fig. 7c, y un ruido aleatorio correspondiente al habla sin voz indicada en la Fig. 7d. El tracto vocal está modelado en forma de filtro todo-polos que procesa pulsos de la Fig. 7c o la Fig. 7d, generados por el modelo de glotis. Por ende, la disposición de la Fig. 7a se puede reducir un modelo de filtro todo polo de la Fig. 7b que presenta una etapa de ganancia, un trayecto hacia delante, un trayecto de realimentación, y una etapa de suma. En el trayecto de realimentación hay un filtro de predicción, y todo el sistema de síntesis del modelo de origen ilustrado en la Fig. 7b se puede representar utilizando funciones de dominio z de la siguiente manera:

$$S(z)=g/(1-A(z))\cdot X(z),$$

donde g representa la ganancia, $A(z)$ es el filtro de predicción determinado por un análisis LPC, $X(z)$ es la señal de excitación y $S(z)$ es la salida de síntesis de habla.

5

Las Figs. 7c y 7d exponen una descripción gráfica del dominio del tiempo de la síntesis de habla con voz y sin voz utilizando el modelo del sistema de origen lineal. Este modelo y los parámetros de excitación de la ecuación anterior son desconocidos y se los debe determinar a partir de una serie finita de muestras de habla. Los coeficientes de $A(z)$ se obtienen utilizando una predicción lineal de la
10 señal de entrada y una cuantificación de los coeficientes de filtro. En un dispositivo de predicción lineal delantera de orden p , la presente muestra de la secuencia de habla se predica a partir de una combinación lineal de p muestras analizadas. Los coeficientes de predicción se pueden determinar mediante algoritmos muy conocidos
15 tales como el algoritmo de Levinson–Durbin o, en general, un método de autocorrelación o un método de reflexión.

La Fig. 7e ilustra una implementación más detallada del bloque de análisis LPC 510. La señal de audio es ingresada en un bloque de determinación de filtro que
20 determinar la información de filtro $A(z)$. Esta información es emitida en forma de información de predicción a corto plazo necesaria para un decodificador. La información de predicción a corto plazo es requerida por el filtro real de predicción 85. En un subtractor 86, se ingresa una muestra en curso de la señal de audio y se substrahe un valor predicho correspondiente a la muestra en curso de manera que
25 para esta muestra, la señal de error de predicción se genera en la línea 84. En la Fig. 7c o 7d se ilustra muy esquemáticamente una secuencia de dichas muestras de señal de error de predicción. Por lo tanto, la Fig. 7a, 7b se puede considerar un tipo de señal rectificada del tipo de impulsos.

Si bien la Fig. 7e ilustra una manera preferida de calcular la señal de excitación, la Fig. 7f ilustra una manera preferida de calcular la señal ponderada. A diferencia de la Fig. 7e, el filtro 85 es diferencia cuando γ es diferente de 1. Un valor inferior a 1 es preferible en el caso γ . Más aun, el bloque 87 está presente y μ es preferentemente un número inferior a 1. Por lo general, los elementos de las Fig. 7e y 7f se pueden
35 implementar como en 3GPP TS 26.190 o 3GPP TS 26.290.

La Fig. 7g ilustra un procesamiento inverso que se puede aplicar del lado del decodificador, como por ejemplo en el elemento 537 de Fig. 2b. Específicamente, el bloque 88 genera una señal no ponderada a partir de la señal ponderada y el bloque
40 89 calcula una excitación de la señal no ponderada. En general, todas las señales menos la señal no ponderada de la Fig. 7g están en el dominio LPC, aunque la señal de excitación y la señal ponderada son señales diferentes en el mismo dominio. El

bloque 89 emite una señal de excitación que se puede utilizar junto con la salida del bloque 536. A continuación, se puede ejecutar la transformación LPC inversa común en el bloque 540 de la Fig. 2b.

- 5 Seguidamente, se describe un codificador CELP de análisis por síntesis en conexión con la Fig. 6 para ilustrar las modificaciones aplicadas a este algoritmo. Este codificador CELP ha sido descrito en forma detallada en detalle en "Speech Coding: A Tutorial Review", Andreas Spanias, Proceedings of the IEEE, Vol. 82, No. 10, Octubre de 1994, páginas 1541–1582. El codificador CELP ilustrado en la Fig. 6
- 10 incluye un componente de predicción a largo plazo 60 y un componente de predicción a corto plazo 62. Además, se utiliza un libro de códigos que está indicado en 64. Se implementa un filtro de ponderación perceptiva $W(z)$ en 66, y en 68 se indica un controlador de minimización de errores. $s(n)$ es la señal de entrada en dominio del tiempo. Después de haber sido perceptivamente ponderada, la señal
- 15 ponderada es ingresada a un substractor 69, que calcula el error entre la señal de síntesis ponderada a la salida del bloque 66 y la señal ponderada original. $s_w(n)$. En general, los coeficientes de filtro de predicción a corto plazo $A(z)$ se calculan mediante una etapa de análisis LPC y sus coeficientes se cuantifican en $\hat{A}(z)$ como se indica en la Fig. 7e. La información de predicción a largo plazo $A_L(z)$ que incluye la
- 20 ganancia de predicción a largo plazo g y el índice de cuantificación de vectores, es decir, las referencias del libro de códigos, se calcula sobre la señal de error de predicción a la salida de la etapa de análisis LPC a la que se hace referencia como 10a en la Fig. 7e. Los parámetros LTP son el retardo de tono y la ganancia. En CELP, esto se implementa habitualmente en forma de libro de códigos adaptativos
- 25 que contiene la señal de excitación pasada (no la residual). El retardo CB y la ganancia se encuentran minimizando la media cuadrática del error ponderado (búsqueda de tono en bucle cerrado).

- 30 El algoritmo CELP codifica, a continuación, la señal residual de acuerdo con las predicciones a corto plazo y a largo plazo utilizando un libro de códigos de secuencias Gaussianas, por ejemplo. El algoritmo ACELP, en el que "A" significa "algebraico" tiene un libro de códigos específico de diseño algebraico.

- 35 Un libro de códigos puede contener más o menos vectores, donde cada vector tiene una longitud de algunas muestras. Un factor de ganancia g escala el vector de códigos y el código de ganancia es filtrado por el filtro de síntesis de predicción a largo plazo y el filtro de síntesis de predicción a corto plazo. El vector de códigos "óptimo" se selecciona de tal manera que se minimice la media cuadrática de error perceptivamente ponderado a la salida del substractor 60. El proceso de búsqueda
- 40 en CELP se realiza mediante una optimización de análisis por síntesis como se ilustra en la Fig. 6.

En casos específicos, cuando un marco es una mezcla de habla sin voz y con voz o cuando tiene lugar habla sobre música, la codificación TCX puede ser más apropiada para codificar la excitación en el dominio LPC. La codificación TCX procesa la señal ponderada en el dominio de la frecuencia sin presumir la
5 producción de excitación alguna. Por ello la TCX es más genérica que la codificación CELP y no se limita a un modelo de excitación de origen con voz o sin voz. La TCX es, de todas maneras, una codificación orientada a la fuente que utiliza un filtro de predicción lineal para modelar los formantes de las señales de habla.

10 En la codificación tipo AMR-WB+, tiene lugar una selección entre los modos TCX y ACELP, como se diera a conocer en la descripción de AMR-WB+. Los modos TCX son diferentes por el hecho de que la longitud de la Transformada Discreta de Fourier en bloques es diferente para diferentes modos y se puede seleccionar la mejor modalidad mediante una estrategia de análisis por síntesis o mediante un
15 modo de "proalimentación".

Como se mencionara con relación a las Fig. 2a y 2b, la etapa de pre-procesamiento común 100 incluye preferentemente un multicanal conjunto (dispositivo estéreo envolvente/conjunto) 101 y, además, una etapa de extensión del ancho de banda
20 102. De modo correspondiente, el decodificador incluye una etapa de extensión del ancho de banda 701 y una etapa de multicanal conjunto conectada a continuación 702. De preferencia, la etapa multicanal conjunto 101 está conectada, con respecto al codificador, antes de la etapa de extensión del ancho de banda 102, y, del lado del decodificador, la etapa de extensión del ancho de banda 701 está conectada
25 antes de la etapa de multicanal conjunta 702 con respecto a la dirección de procesamiento de la señal. Sin embargo, por otra parte la etapa de pre-procesamiento común puede incluir una etapa de multicanal conjunto sin la etapa de extensión del ancho de banda conectada a continuación o una etapa de extensión del ancho de banda sin etapa de multicanal conjunto conectada.

30 Se ilustra un ejemplo preferido de etapa de multicanal conjunto en el lado del codificador 101a, 101b y del lado del decodificador 702a y 702b en el contexto de la Fig. 8. Se ingresa al mezclador descendente 101a de manera que el mezclador descendente genere un número de K canales transmitidos, donde el número K es
35 mayor o igual que uno y es menor o igual a E.

De preferencia, los E canales de entrada son ingresados a un analizador de parámetros de multicanal conjunto 101b que genera información paramétrica. Esta información paramétrica es sometida preferentemente a codificación de entropía,
40 como por ejemplo codificando por diferencia y posterior codificación de Huffman o, por otro lado, la codificación aritmética subsiguiente. La información paramétrica codificada producida por el bloque 101b es transmitida a un decodificador de

parámetros 702b que puede ser parte del elemento 702 de la Fig. 2b. El decodificador de parámetros 702b decodifica la información paramétrica transmitida y reenvía la información paramétrica decodificada al upmixer 702a. El upmixer 702a recibe los K canales transmitidos y genera un número L de canales de salida, donde
5 el número de L es superior o igual a K y menor o igual a E.

La información paramétrica puede incluir diferentes de nivel entre canales, diferentes de tiempo entre canales, diferencias de fase entre canales y/o medidas de coherencia entre canales de acuerdo con lo conocido en la técnica BCC o como se
10 ha dado a conocer y descrito en forma detallada en la norma de entorno MPEG. El número de canales transmitidos puede ser un solo canal mono en el caso de aplicaciones de tasa de transferencia de bit ultra bajas o puede incluir una aplicación estéreo compatible o puede incluir una señal estéreo compatible, es decir dos canales. Por lo general, el número de E canales de entrada puede ser cinco o tal vez
15 incluso mayor. Por otro lado, el número de E canales de entrada puede consistir en E objetos de audio, como se conoce en el contexto de la codificación de objetos espaciales de audio.

En una implementación, el mezclador descendente ejecuta una suma ponderada o
20 no ponderada de los E canales de entrada originales o una suma de los E objetos de audio de entrada. En el caso de objetos de audio como canales de entrada, el analizador de parámetros de multicanal conjunto 101b calcula parámetros de objetos de audio tales como una matriz de correlaciones entre los objetos de audio preferentemente por cada porción de tiempo y aun más preferentemente por cada
25 banda de frecuencia. Para este fin, se puede dividir todo el rango de frecuencias en por lo menos 10 y, preferentemente, 32 o 64 bandas de frecuencia.

La Fig. 9 ilustra una realización preferida de la implementación de la etapa de extensión de ancho de banda 102 de la Fig. 2a y la correspondiente etapa de
30 extensión de ancho de banda 701 de la Fig. 2b. Del lado del codificador, el bloque de extensión de ancho de banda 102 incluye preferentemente un bloque de filtrado de paso bajo 102b, un bloque de submuestreo que sigue al paso bajo o que es parte del QMF inverso, que actúa sólo en la mitad de las bandas QMF y un analizador de banda alta 102a. La señal de audio original ingresada al bloque de extensión de
35 ancho de banda 102 es filtrado en paso bajo para generar la señal de banda baja que a continuación es ingresada a las ramas codificadoras y/o al conmutador. El filtro de paso bajo tiene una frecuencia de corte que puede estar en el rango de 3kHz a 10kHz. Más aun, el bloque de extensión de ancho de banda 102 incluye además un analizador de banda alta para calcular parámetros de extensión de
40 ancho de banda tales como una información de parámetros de envoltura espectral, una información de parámetros de piso de ruido, una información de parámetros de filtrado inverso, más información paramétrica relacionada con ciertas líneas

armónicas en la banda alta y otros parámetros descritos en detalle en la norma MPEG-4 en el capítulo relacionado con la replicación de la banda espectral.

5 Del lado del decodificador, el bloque de extensión de ancho de banda 701 incluye un corrector 701a, un dispositivo de ajuste 701b y un combinador 701c. El combinador 701c combina la señal de banda baja decodificada y la señal de banda alta reconstruida y ajustada por el dispositivo de ajuste 701b. La entrada al dispositivo de ajuste 701b es provista por un corrector que se pone en funcionamiento para derivar la señal de banda alta de la señal de banda baja, por ejemplo mediante replicación de la banda espectral o, generalmente, mediante extensión del ancho de banda. La corrección ejecutada por el corrector 701a puede ser una corrección realizada de manera armónica o de manera no armónica. A continuación, la señal generada por el corrector 701a es ajustada por el dispositivo de ajuste 701b empleando la información paramétrica de ancho de banda transmitida.

15 Como se indica en la Fig. 8 y en la Fig. 9, los bloques descritos pueden tener una entrada de control de modo en una realización preferida. Esta entrada de control de modo se deriva de la señal de salida de la etapa de decisión 300. En esa realización preferida, se puede adaptar una característica del bloque correspondiente a la salida de la etapa de decisión, es decir si, en una realización preferida, se realiza una decisión por habla o una decisión por música correspondiente a una determinada porción de tiempo de la señal de audio. De preferencia, el control de modo sólo se relaciona con una o más de las funcionalidades de estos bloques, pero no todas las funcionalidades de los bloques. Por ejemplo, la decisión puede afectar sólo al corrector 701a y puede no influir sobre los demás bloques de la Fig. 9, o puede incluir, por ejemplo, sólo en el analizador de parámetros de multicanal 101b de la Fig. 8 pero no en los otros bloques de la Fig. 8. Esta implementación es preferentemente tal que se obtenga una señal de salida con mayor flexibilidad y calidad más elevada y menor tasa de transferencia de bits otorgando flexibilidad en la etapa de pre-procesamiento común. Sin embargo, por otra parte, el uso de algoritmos en la etapa de pre-procesamiento común para ambos tipos de señales permite la implementación de un esquema de codificación/decodificación eficiente.

35 La Fig. 10a y la Fig. 10b ilustran dos implementaciones diferentes de la etapa de decisión 300. En la Fig. 10a, se indica una decisión en bucle abierto. En este caso, el analizador de señal 300a de la etapa de decisión tiene ciertas reglas para decidir si la determinada porción de tiempo o la determinada porción de frecuencia de la señal de entrada tiene una característica que requiera que esta porción de señal sea codificada por la primera rama codificadora 400 o por la segunda rama codificadora 500. Para este fin, el analizador de señales 300a puede analizar la señal de audio de entrada a la etapa de pre-procesamiento común o puede analizar la señal de audio producida por la etapa de pre-procesamiento común, es decir, la señal de audio

intermedia, o puede analizar una señal intermedia dentro de la etapa de pre-procesamiento común tal como la salida de la señal de downmix que puede ser una señal mono o que puede ser una señal con k canales indicados en la Fig. 8. Del lado de la salida, el analizador de señales 300a genera la decisión de conmutación para
5 controlar el conmutador 200 del lado del codificador y el correspondiente conmutador 600 o el combinador 600 del lado del decodificador.

Aunque no se describe en detalle con respecto al segundo conmutador 521, se debe hacer hincapié en que el segundo conmutador 521 puede estar colocado de manera
10 similar al primer conmutador 200, como se describe con relación a la Fig. 4a y la Fig. 4b. Por consiguiente, una posición alternativa del conmutador 521 de la Fig. 3c es a la salida de ambas ramas de procesamiento 522, 523, 524 de manera que ambas ramas de procesamiento operen en paralelo y sólo se escriba la salida de una rama de procesamiento en una secuencia de bits por medio de un formador de secuencias
15 de bits que no se ilustra en la Fig. 3c.

Más aun, el segundo combinador 600 puede contar con una funcionalidad de desvanecimiento cruzado específica como se describiera en la Fig. 4c. Por otro lado, o además, el primer combinador 532 puede tener la misma funcionalidad de
20 desvanecimiento cruzado. Asimismo, ambos combinadores pueden tener la misma funcionalidad de desvanecimiento cruzado o pueden tener diferentes funcionalidades de desvanecimiento cruzado o no tener ninguna funcionalidad de desvanecimiento cruzado en absoluto por lo que ambos combinadores son conmutadores sin ninguna funcionalidad adicional de desvanecimiento cruzado.

25 Como se describiera anteriormente, ambos conmutadores se pueden controlar por medio de una decisión en bucle abierto o una decisión en bucle cerrado como se describe con respecto a la Fig. 10a y la Fig. 10b, donde el controlador 300, 525 de la Fig. 3c puede tener diferentes o las mismas funcionalidades en ambos
30 conmutadores.

Más aun, puede existir una funcionalidad de distorsión temporal que es adaptativa a la señal no sólo en la primera rama codificadora ni en la primera rama decodificadora sino que también puede existir en la segunda rama de procesamiento
35 de la segunda rama codificadora del lado del codificador y también del lado del decodificador. Dependiendo de la señal procesada, ambas funcionalidades de distorsión pueden constar de la misma información de distorsión temporal de manera que se aplique la misma distorsión temporal a las señales en el primer dominio y en el segundo dominio. Esto ahorra carga de procesamiento y podría ser útil en
40 algunos casos, en situaciones en que los bloques subsiguientes tienen una misma característica temporal de distorsión temporal. Sin embargo, en realizaciones alternativas, es preferible contar con estimadores independientes de distorsión

temporal para la primera rama codificadora y la segunda rama de procesamiento de la segunda rama codificadora.

5 La señal de audio codificada de la presente invención puede ser almacenada en un medio de almacenamiento digital o puede ser transmitida en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión cableado tal como la internet.

10 En una realización diferente, el conmutador 200 de la Fig. 1a o 2a conmuta entre las dos ramas codificadoras 400, 500. En otra realización, puede haber ramas codificadoras adicionales tales como una tercera rama codificadora o incluso una cuarta rama codificadora o aun más ramas codificadoras. Del lado del decodificador, el conmutador 600 de la Fig. 1b o 2b conmuta entre las dos ramas decodificadoras 431, 440 y 531, 532, 533, 534, 540. En otra realización, puede haber otras ramas
15 decodificadoras tales como una tercera rama decodificadora o incluso una cuarta rama de codificadora o aun más ramas decodificadoras. Del mismo modo, los otros conmutadores 521 o 532 pueden conmutar entre más de dos algoritmos de codificación diferentes, cuando se presentan dichas ramas de codificación/decodificación adicionales.

20 La Fig. 12A ilustra una realización preferida de una implementación de codificador y la Fig. 12B ilustra una realización preferida de la implementación de decodificador correspondiente. Además de los elementos antes mencionados con respecto a los correspondientes números de referencia, la realización de la Fig. 12A ilustra un
25 módulo psicoacústico separado 1200, y además ilustra una implementación preferida de las herramientas codificadoras adicionales ilustradas en el bloque 421 de la Fig. 11A. Estas herramientas adicionales son una herramienta de perfilado temporal de ruido (TNS) 1201 y una herramienta de codificación intermedia/secundaria (M/S) 1202. Más aun, se ilustran otras funcionalidades de los elementos 421 y 524 en el
30 bloque 421/542 como implementación combinada de escalamiento, análisis de carga de ruido, cuantificación, codificación aritmética de valores espectrales.

En la implementación de decodificador correspondiente de la Fig. 12B se ilustran elementos adicionales, que son una herramienta de decodificación M/S 1203 y una
35 herramienta de decodificador TNS 1204. Más aun, un postfiltro de graves que no está ilustrado en las figuras anteriores está indicado en 1205. El bloque de transición de visualización en ventanas 532 corresponde al elemento 532 en la Fig. 2B, que aparece en forma de conmutador, pero que ejecuta una suerte de desvanecimiento cruzado que puede ser un desvanecimiento cruzado sobremuestreado o un
40 desvanecimiento cruzado con muestreo crítico. Este último se implementa como operación MDCT, donde dos porciones con solapamiento temporal se traslapan y se suman. Este procesamiento de transición con muestreo crítico se utiliza

preferentemente, donde sea apropiado, ya que se puede reducir la tasa total de transferencia de bits sin pérdida alguna de calidad. El bloque adicional de ventanas de transición 600 corresponde al combinador 600 de la Fig. 2B, que una vez más aparece ilustrado en forma de conmutador, aunque está claro que este elemento ejecuta una suerte de desvanecimiento cruzado ya sea con muestreo crítico o sin muestreo crítico para evitar las aberraciones de bloqueo y, específicamente, las aberraciones de conmutación, donde un bloque ha sido procesado en la primera rama y el otro bloque ha sido procesado en la segunda rama. Cuando sin embargo, el procesamiento en ambas ramas coincide perfectamente, luego la operación de desvanecimiento cruzado se puede "degradar" a un conmutador duro, en tanto que se entiende que una operación de desvanecimiento cruzado es una conmutación "blanda" entre ambas ramas.

El concepto de la Fig. 12A y 12B permite la codificación de señales con una mezcla arbitraria de contenido de habla y audio, y este concepto se ejecuta de manera comparable o mejor que la mejor tecnología de codificación se podría adaptar específicamente a la codificación del contenido de habla o de audio en general. La estructura general del codificador y el decodificador se puede describir en términos de que hay un pre y postprocesamiento común de una unidad funcional de entorno MPEG (MPEGs) para manejar el procesamiento estéreo o multicanal y una unidad SBR mejorada (eSBR), que se encarga de la representación paramétrica de las frecuencias de audio más elevadas en la señal de entrada. A continuación hay dos ramas: una que consiste en un trayecto de herramienta de codificación de audio avanzada modificada (AAC) y la otra que consiste en un trayecto basado en codificación de predicción lineal (dominio LP o LPC), que a su vez presenta una representación en dominio de la frecuencia o una representación en dominio del tiempo de la LPC residual. Todos los espectros transmitidos correspondientes a ambas, AAC y LPC, se representan en el dominio MDCT después de la cuantificación y codificación aritmética. La representación en dominio del tiempo utiliza un esquema de codificación de excitación ACELP. La estructura básica está ilustrada en la Fig. 12A correspondiente al codificador y en la Fig. 12B para el decodificador. El flujo de datos en este diagrama es de izquierda a derecha, de arriba hacia abajo. Las funciones del decodificador consisten en encontrar la descripción de la representación espectral de audio cuantificada o en dominio del tiempo en la carga útil de la secuencia de bits y decodificar los valores cuantificados y otra información de reconstrucción.

En el caso de la información espectral transmitida, el decodificador ha de reconstruir los espectros cuantificados, procesar los espectros reconstruidos por medio de las herramientas que estén activas en la carga útil de la secuencia de bits para poder arribar a los espectros de señal reales descritos por la carga útil de secuencia de bits de entrada y por último convertir los espectros en dominio de la frecuencia al

dominio del tiempo. Después de la reconstrucción inicial y el escalado de la reconstrucción de espectros, hay herramientas adicionales que modifican uno o más de los espectros a fin de ofrecer una codificación más eficiente.

- 5 En el caso de una representación de señal transmitida en dominio del tiempo, el decodificador ha de reconstruir la señal en tiempo cuantificada, procesar la señal temporal reconstruida por medio de las herramientas que estén activas en la carga útil de la secuencia de bits a fin de llegar a la señal en dominio del tiempo real descrita por la carga útil de la secuencia de bits de entrada.

10

Para cada una de las herramientas opcionales que operan en los datos de señal, la opción de "paso o transferencia" queda retenida y en todos los casos en que se omite el procesamiento, se transfieren los espectros o muestras temporales a su entrada directamente a través de la herramienta sin modificación.

15

En lugares en que la secuencia de bits cambia su representación de señal del dominio del tiempo a la representación en dominio de la frecuencia o del dominio LP al dominio no LP o viceversa, el decodificador facilita la transición de un dominio al otro por medio de una generación en ventanas de una transición adecuada de traslape-suma.

20

El procesamiento eSBR y MPEG-S se aplica de manera similar a ambos trayectos de codificación después de operar la transición.

- 25 La entrada a la herramienta desmultiplexora de carga útil de secuencia de bits es una carga útil de secuencias de bits. El desmultiplexor separa la carga útil de la secuencia de bits en partes por cada herramienta y dota a cada una de las herramientas de una información de carga útil de la secuencia de bits relacionada con esa herramienta.

30

Las salidas de la herramienta desmultiplexor de carga útil de la secuencia de bits son:

Dependiendo del tipo de codificación central en el marco actual una de:

- 35 los espectros cuantificados y codificados sin ruido representados por
información de factor de escala
líneas espectrales codificadas aritméticamente
o: parámetros de predicción lineal (LP) junto con una señal de excitación representada por:
líneas espectrales cuantificadas y codificadas aritméticamente (excitación
40 codificada por transformada, TCX) o
excitación en dominio del tiempo codificada por ACELP

- La información de carga de ruido espectral (opcional)
- La información sobre la decisión M/S (opcional)
- La información de perfilado temporal de ruido (TNS) (opcional)
- La información de control de bancos de filtros
- 5 La información de control de corrección de distorsión temporal (TW) (opcional)
- La información de control de replicación del ancho de la banda espectral (eSBR)
- La información de control del Entorno MPEG (MPEGS)

- 10 La herramienta decodificadora sin ruido del factor de escala toma información del desmultiplexor de carga útil de la secuencia de bits, analiza esa información y decodifica los factores de escala codificados de Huffman y DPCM.
- La entrada a la herramienta decodificadora sin ruido del factor de escala es:
 - La información de factor de escala para los espectros codificados sin ruido
- 15 La salida de la herramienta decodificadora sin ruido del factor de escala es:
 - La representación en enteros decodificada de los factores de escala:

- 20 La herramienta decodificadora espectral sin ruido toma información del desmultiplexor de carga útil de la secuencia de bits, analiza esa información y decodifica los datos codificados aritméticamente y reconstruye los espectros cuantificados. La entrada a esta herramienta decodificadora sin ruido es:
 - Los espectros codificados sin ruido
- 25 La salida de esta herramienta decodificadora sin ruido es:
 - Los valores cuantificados de los espectros

- 30 La herramienta de cuantificación inversa toma los valores cuantificados correspondientes a los espectros y convierte los valores en números enteros a espectros reconstruidos no escalados. Este cuantificador es un cuantificador de compresión/expansión cuyo factor de compresión expansión depende del modo de codificación elegido.
- 35 La entrada a la herramienta de Cuantificación Inversa es:
 - Los valores cuantificados correspondientes a los espectros

- La salida de la herramienta de Cuantificación Inversa es:
 - Los espectros no escalados con cuantificación inversa

La herramienta de carga de ruido se utiliza para llenar las brechas espectrales en los espectros decodificados que tienen lugar cuando se cuantifica a cero el nivel espectral, por ejemplo debido a una fuerte restricción de la demanda de bits en el codificador. El uso de la herramienta de carga de ruido es opcional.

Las entradas a la herramienta de carga de ruido son:

Los espectros no escalados con cuantificación inversa

Parámetros de carga de ruido

La representación en enteros decodificada de los factores de escala

Las salidas a la herramienta de carga de ruido son:

Los valores espectrales no escalados con cuantificación inversa correspondientes a líneas espectrales que fueran previamente cuantificadas en cero.

Representación modificada en números enteros de los factores de escala

La herramienta de reescalado convierte la representación en números enteros de los factores de escala a los valores reales y multiplica los espectros no escalados con cuantificación inversa por los factores de escala pertinentes.

Las entradas a la herramienta de factores de escala son:

La representación en enteros decodificada de los factores de escala

Los espectros no escalados con cuantificación inversa

La salida de la herramienta de factores de escala es:

Los espectros escalados con cuantificación inversa

Para obtener una reseña de la herramienta M/S, remitirse por favor a ISO/IEC 14496-3, subparte 4.1.1.2.

Para obtener una reseña sobre la herramienta de perfilado temporal de ruido (TNS) remitirse por favor a ISO/IEC 14496-3, subparte 4.1.1.2.

La herramienta de banco de filtros /conmutación de bloques aplica la inversa del mapeo de la frecuencia que se llevara a cabo en el codificador. Se utiliza una transformada de coseno discreta modificada inversa (IMDCT) para la herramienta de banco de filtros. La IMDCT puede estar configurada para soportar 120, 128, 240, 256, 320, 480, 512, 576, 960, 1024 o 1152 coeficientes espectrales.

Las entradas a la herramienta de banco de filtros son:

Los espectros (con cuantificación inversa)

La información de control de banco de filtros

La salida (salidas) de la herramienta del banco de filtros es (son):
 La señal (o señales) de audio reconstruida en el dominio del tiempo.

5 La herramienta con distorsión temporal de banco de filtros / conmutación de bloques reemplaza a la herramienta normal de banco de filtro/conmutación de bloques cuando se habilita el modo de distorsión. El banco de filtros es el mismo (IMDCT) que para el banco de filtros normal, y se mapean además las muestras en dominio del tiempo en ventana contra el dominio del tiempo lineal mediante remuestreo de tiempo variable.

10 Las entradas a las herramientas de banco de filtros con distorsión temporal son:

- Los espectros con cuantificación inversa
- La información de control del banco de filtros
- La información de control de distorsión temporal

15 La salida (salidas) de la herramienta del banco de filtros es (son):

- La señal (o señales) de audio lineal reconstruida en el dominio del tiempo .

20 La Herramienta potenciada SBR (eSBR) regenera la banda alta de la señal de audio. Se basa en la replicación de las secuencias de armonías, truncadas durante la codificación. Ajusta la envoltura espectral de la banda alta generada y aplica filtrado inverso y agrega ruido y componentes sinusoides a fin de recrear las características espectrales de la señal original.

25 La entrada a la herramienta eSBR es:

- Los datos de envoltura cuantificados
- Datos de control varios
- una señal en dominio del tiempo procedente del decodificador de núcleo AAC

30 La salida de la herramienta eSBR es una de las siguientes:

- una señal en dominio del tiempo o
- una representación en dominio QMF de una señal, por ejemplo en el caso de utilizarse la herramienta de Entorno MPEG.

35 La herramienta de Entorno MPEG (MPEGS) produce múltiples señales de una o más señales de entrada mediante la aplicación de un procedimiento de upmix sofisticado a la señal o señales de entrada controladas por parámetros espaciales. En el contexto USAC se utiliza MPEGS para codificar una señal de multicanal transmitiendo información secundaria junto con una señal transmitida con

40 downmixing.

La entrada a la herramienta MPEGS es:

una señal en dominio del tiempo con downmixing o

una representación en dominio QMF de una señal con downmixing procedente de la herramienta eSBR

- 5 La salida de la herramienta MPEGS es:
una señal de multicanal en dominio del tiempo

La Herramienta de Clasificación de Señales analiza la señal de entrada original y genera a partir de la misma información que inicia la selección de los diferentes
10 modos de codificación. El análisis de la señal de entrada depende de la implementación e intenta elegir el modo de codificación núcleo óptimo para un marco de señal de entrada dado. La salida del clasificador de señales puede utilizarse además (opcionalmente) para incluir en el comportamiento de otras herramientas, por ejemplo el Entorno MPEG, SBR potenciado, banco de filtros con distorsión
15 temporal y otras.

La entrada a la herramienta de Clasificación de Señales es:
la señal original de entrada sin modificaciones
otros parámetros dependientes de la implementación

20

La salida de herramienta de Clasificación de Señales es:
una señal de control para controlar la selección del códec núcleo (codificación en dominio de la frecuencia filtrado no LP, dominio de la frecuencia filtrado LP o codificación en dominio del tiempo filtrado LP)

25

De acuerdo con la presente invención, la resolución en tiempo/frecuencia que tiene lugar en el bloque 410 de la Fig. 12A y en el convertidor 523 de la Fig. 12A se controla dependiendo de la señal de audio. La interrelación entre longitud de ventana, longitud de transformada, resolución en tiempo y resolución en frecuencia
30 está ilustrada en la Fig. 13A, donde se pone de manifiesto que, para una gran longitud de ventana, la resolución en tiempo baja, aunque la resolución en frecuencia sube y, para una corta longitud de ventana, la resolución en tiempo es elevada aunque la resolución en frecuencia es baja.

35 En la primera rama codificadora, que es preferentemente la rama codificadora AAC indicada por los elementos 410, 1201, 1202, 4021 de la Fig. 12A, se pueden utilizar diferentes ventanas, donde la forma de la ventana está determinada por un analizador de señales que está codificado preferentemente en el bloque de clasificación de señales 300, aunque también puede ser un módulo separado. El
40 codificador selecciona una de las ventanas ilustradas en la Fig. 13B, que tienen diferentes resoluciones en tiempo/frecuencia. La resolución en tiempo/frecuencia de la primera ventana larga, la segunda ventana, la cuarta ventana, la quinta ventana y

la sexta ventana es igual a 2.048 valores de muestreo a una longitud de transformada de 1024. La ventana corta ilustrada en la tercera línea de la Fig. 13B tiene una resolución en tiempo de 256 valores de muestreo correspondientes al tamaño de la ventana. Esto corresponde a una longitud de transformada de 128.

5

De manera análoga, las dos últimas ventanas tienen una longitud de ventana igual a 2.304, que es una resolución en frecuencia mejor que la ventana de la primera línea aunque una resolución en tiempo más baja. La longitud de transformada de las ventanas de las dos últimas líneas es igual a 1.152.

10

En la primera rama codificadora, se pueden construir secuencias de ventanas que se construyen a partir de las ventanas de transformación de la Fig. 13B. Si bien en la Fig. 13C sólo se ilustra una secuencia corta, en tanto que las otras "secuencias" consisten sólo en una ventana única, también se pueden construir secuencias más

15 largas que consisten en más ventanas. Cabe señalar que de acuerdo con la Fig. 13B, para los números de coeficientes menores, es decir, 960 en lugar de 1.024, la resolución en tiempo también es menor que para el número más elevado de coeficientes correspondiente tal como 1024.

20 La Fig. 14A – 14G ilustra diferentes resoluciones/tamaños de ventana de la segunda rama codificadora. En una realización preferida de la presente invención, la segunda rama codificadora tiene una primera rama de procesamiento que es un codificador en domino del tiempo ACELP 526, y la segunda rama de procesamiento comprende el banco de filtros 523. En esta rama, se subdivide un super marco de, por ejemplo,

25 2048 muestras, en marcos de 256 muestras. Se pueden utilizar por separado marcos individuales de 256 muestras de manera que se pueda aplicar una secuencia de cuatro ventanas, donde cada ventana cubre dos marcos, cuando se aplica un MDCT con 50 por ciento de traslape. A continuación, se utiliza una elevada resolución en tiempo como se ilustra en la Fig. 14D. Por otro lado, cuando la señal permite

30 ventanas más largas, se puede aplicar la secuencia ilustrada en la Fig. 14C, donde se aplica un tamaño de ventana doble con 1.024 muestras por cada ventana (ventanas medianas), de manera que una ventana cubre cuatro marcos y hay un traslape de 50 por ciento.

35 Por último, la señal es tal que se puede utilizar una ventana larga, ventana larga que se extiende a lo largo de 4.096 muestras, una vez más con un traslape de 50 por ciento.

En la realización preferida, en la cual hay dos ramas, donde una rama cuenta con un

40 codificador ACELP, la posición del marco ACELP indicado por "A" en el super marco puede determinar el tamaño de ventana aplicado para dos marcos TCX adyacentes indicados por "T" en la Fig. 14E. Básicamente, uno está interesado en el uso de

ventanas largas siempre que sea posible. No obstante, se deben aplicar ventanas cortas cuando un marco T único está entre dos marcos A. Se pueden aplicar ventanas medianas cuando hay dos marcos T adyacentes. Sin embargo, cuando hay tres marcos T adyacentes, una ventana más larga correspondiente podría no ser eficiente debido a la complejidad adicional. Por lo tanto, el tercer marco T, aunque no está precedido por un marco A, puede ser procesado por una ventana corta. Cuando la totalidad del super marco tiene sólo marcos T, luego se puede aplicar una ventana larga.

- La Fig. 14F ilustra varias alternativas de ventanas, donde el tamaño de la ventana es siempre 2x el número lg de coeficientes espectrales debido al traslape preferido de 50 por ciento. Sin embargo, se pueden aplicar otros porcentajes de traslape para todas las ramas codificadoras de manera que la relación entre el tamaño de la ventana y la longitud de transformación también sea diferente de dos e incluso se acerque a uno, cuando no se aplica solapamiento en el domino del tiempo.

- La Fig. 14G ilustra reglas para construir una ventana sobre la base de reglas expuestas en la Fig. 14F. El valor ZL ilustra ceros al comienzo de la ventana. El valor L ilustra un número de coeficientes de ventana en una zona de solapamiento. Los valores presentes en la porción M son valores "1" que no introducen solapamiento alguno debido a un traslape con una ventana adyacente que tiene cero valores en la porción correspondiente a M. La porción M está seguida por una zona de traslape derecho R, a la que le sigue una zona de ceros Z\$, que correspondería a una porción M de una ventana subsiguiente.

- Se hace referencia al anexo adjunto a continuación, que describe una implementación preferida y detallada de un esquema de codificación/decodificación de audio, especialmente con respecto al lado del decodificador.

Anexo

1. Ventanas y secuencias de ventanas

- La cuantificación y codificación se realiza en el dominio de la frecuencia. Para este fin, se mapea la señal temporal en el domino de la frecuencia en el codificador. El codificador ejecuta el mapeo inverso de acuerdo con lo descrito en la subcláusula 2. Dependiendo de la señal, el codificador puede cambiar la resolución en tiempo/frecuencia utilizando tres tamaños diferentes de ventana: 2304, 2048 y 256. Para conmutar entre ventanas, se utilizan las ventanas de transición LONG_START_WINDOW (ventana inicio larga), LONG_STOP_WINDOW (ventana terminación larga), START_WINDOW_LPD (ventana inicio LPD),

STOP_WINDOW_1152 (ventana terminación 1152), STOP_START_WINDOW (ventana terminación inicio) y STOP_START_WINDOW_1152 (ventana terminación inicio 1152). La Tabla 5.11 enumera las ventanas, especifica la correspondiente longitud de transformada y muestra esquemáticamente la forma de las ventanas. Se utilizan tres coeficientes de longitud de transformada: 1152, 1024 (o 960) (a la que se hace referencia como transformada larga) y 128 (o 120) (a la que se hace referencia como transformada corta).

Las secuencias de ventanas están compuestas por ventanas de manera tal que un bloque de datos sin procesar (raw_data_block) siempre contenga datos que representan 1024 (o 960) muestras de salida. El elemento de datos **window_sequence** (secuencia de ventanas) indica la secuencia de ventanas que se utiliza en realidad. La Fig. 13C indica cómo las secuencias de ventanas están compuestas por ventanas individuales. Remitirse a la subcláusula 2 para obtener información más detallada sobre la transformación y las ventanas.

1.2 Bandas y agrupamientos de factores de escala

Ver ISO/IEC 14496-3, subparte 4, subcláusula 4.5.2.3.4

Como se explica en ISO/IEC 14496-3, subparte 4, subcláusula 4.5.2.3.4, el ancho de las bandas de factor de escala se construye imitando las bandas críticas del aparato auditivo humano. Por esa razón el número de bandas de factores de escala de un espectro y su ancho dependen de la longitud de transformada y la frecuencia de muestreo. La Tabla 4.110 a la Tabla 4.128, de ISO/IEC 14496-3, subparte 4, sección 4.5.4, enumeran la compensación al comienzo de cada banda de factor de escala en las longitudes de transformación 1024 (960) y 128 (120) y en las frecuencias de muestreo. Las tablas destinadas en un principio a LONG_WINDOW, LONG_START_WINDOW y LONG_STOP_WINDOW se utilizan también para START_WINDOW_LPD y STOP_START_WINDOW. Las tablas de compensación correspondientes a STOP_WINDOW_1152 y STOP_START_WINDOW_1152 están consignadas en la Tabla 4 a la Tabla 10.

1.3 Decodificación de lpd_channel_stream()

El elemento de flujo de bits lpd_channel_stream() contiene toda la información para decodificar un marco de señal codificada "en el dominio de predicción lineal". Contiene la carga útil para un marco de la señal codificada que fuera codificada en el dominio LPC, es decir, que incluye un paso de filtrado LPC. El residual de este filtro (la denominada "excitación") se representa luego con la ayuda de un módulo ACELP o en el dominio de transformación MDCT ("excitación codificada por transformada" TCX). Para permitir una adaptación estrecha a las características de la señal, se

descompone un marco en cuatro unidades más pequeñas de igual tamaño, cada una de las cuales es codificada con un esquema de codificación ACELP o TCX.

Este proceso es similar al esquema de codificación descrito en 3GPP TS 26.290. A partir de este documento hay una terminología ligeramente diferente, en que un "supermarco" significa un segmento de señal de 1024 muestras, en tanto que un "marco" es exactamente un cuarto de eso, es decir, 256 muestras. Cada uno de estos marcos se subdivide a su vez en cuatro "submarcos" de igual longitud. Sírvese notar que este subcapítulo adopta esta terminología

1.4 Definiciones, Elementos de Datos

acelp_core_mode Este campo de bits indica el esquema de asignación exacto de bits en caso de utilizar ACELP como modo de codificación lpd.

lpd_mode El modo de campo de bits define los modos de codificación por cada uno de los cuatro marcos dentro de un supermarco del lpd_channel_stream() (corresponde a un marco AAC). Los modos de codificación están almacenados en la serie mod[] y pueden tomar valores de 0 a 3. El mapeo de lpd_mode a mod[] se puede determinar tomando en cuenta la siguiente Tabla 1.

Tabla 1 – Mapeo de los modos de codificación correspondientes a lpd_channel_stream()

	significado de los bits en el modo de campos de bits					entradas mod[] restantes
lpd_mode	bit 4	bit 3	bit 2	bit 1	bit 0	
0..15	0	mod[3]	mod[2]	mod[1]	mod[0]	
16..19	1	0	0	mod[3]	mod[2]	mod[1]=2 mod[0]=2
20..23	1	0	1	mod[1]	mod[0]	mod[3]=2 mod[2]=2
24	1	1	0	0	0	mod[3]=2 mod[2]=2 mod[1]=2 mod[0]=2
25	1	1	0	0	1	mod[3]=3 mod[2]=3 mod[1]=3 mod[0]=3
26..31						reservado

mod[0..3] Los valores de la matriz mod[] indican los modos de codificación respectivos en cada marco:

Tabla 2 – Modos de codificación indicados por mod[]

valor de mod[x]	modo de codificación en marco	elemento de flujo de bits
0	ACELP	acelp_coding()
1	un marco de TCX	tcx_coding()
2	TCX que cubre la mitad de un supermarco	tcx_coding()
3	TCX que cubre todo el supermarco	tcx_coding()

- acelp_coding()
- Elemento de sintaxis que contiene todos los datos para decodificar un marco de excitación ACELP.
- 5 tcx_coding()
- Elemento de sintaxis que contiene todos los datos para decodificar un marco de excitación codificada por transformada (TCX) basada en MDCT.
- first_tcx_flag
- Bandera que indica si el marco TCX procesado en curso es el primero del supermarco.
- 10 lpc_data()
- Elemento de sintaxis que contiene todos los datos para decodificar todas las series de parámetros de filtro LPC necesarios para decodificar el supermarco actual.
- first_lpd_flag
- Bandera que indica si el supermarco actual es el primero de una secuencia de supermarcos que están codificados en el dominio LPC. Esta bandera también se puede determinar de la historia del elemento de flujo de bits core_mode (core_mode0 y core_mode1 en el caso de un channel_pair_element) de acuerdo con Tabla 3.
- 15

Tabla 3 – Definición de first_lpd_flag

core_mode del marco anterior (supermarco)	core_mode del marco actual (supermarco)	first_lpd_flag
0	1	1
1	1	0

- 20 last_lpd_mode
- Indica el lpd_mode del marco decodificado anteriormente.

1.5 Proceso de decodificación

- 25 En la lpd_channel_stream el orden de decodificación es:
- Obtener acelp_core_mode
- Obtener lpd_mode y determinar del mismo el contenido del mod[] variable auxiliar

Obtener datos de acelp_coding o tcx_coding data, dependiendo del contenido del mod[] variable auxiliar

Obtener lpc_data

5 1.6 Combinaciones de modos de codificación **ACELP/TCX**

De manera análoga a [8], sección 5.2.2, hay 26 combinaciones permitidas de ACELP o TCX dentro de un supermarco de una carga útil de lpd_channel_stream. Una de estas 26 combinaciones de modos está señalizada en el elemento de flujo de bits lpd_mode. En la Tabla 1 y la Tabla 2 se exhibe el mapeo de lpd_mode a los modos de codificación reales de cada marco en un submarco.

15 **Tabla 4 – bandas de factores de escala correspondientes a una longitud de ventana de 2304 para STOP_START_1152_WINDOW y STOP_1152_WINDOW a 44.1 y 48 kHz**

fs [kHz]	44.1,48
num_swb_long_window	49
swb	swb_offset_long_window
0	0
1	4
2	8
3	12
4	16
5	20
6	24
7	28
8	32
9	36
10	40
11	48
12	56
13	64
14	72
15	80
16	88
17	96
18	108
19	120
20	132
21	144
22	160
23	176
24	196

swb	swb_offset_long_window
25	216
26	240
27	264
28	292
29	320
30	352
31	384
32	416
33	448
34	480
35	512
36	544
37	576
38	608
39	640
40	672
41	704
42	736
43	768
44	800
45	832
46	864
47	896
48	928
	1152

**Tabla 5 – bandas de factores de escala correspondientes a una longitud de ventana de 2304 para
STOP_START_1152_WINDOW y STOP_1152_WINDOW a 32 kHz**

fs [kHz]	32
num_swb_long_window	51
swb	swb_offset_long_window
0	0
1	4
2	8
3	12
4	16
5	20
6	24
7	28
8	32
9	36
10	40
11	48
12	56
13	64
14	72
15	80
16	88
17	96
18	108
19	120
20	132
21	144
22	160
23	176
24	196
25	216

swb	swb_offset_long_window
26	240
27	264
28	292
29	320
30	352
31	384
32	416
33	448
34	480
35	512
36	544
37	576
38	608
39	640
40	672
41	704
42	736
43	768
44	800
45	832
46	864
47	896
48	928
49	960
50	992
	1152

Tabla 6 – bandas de factores de escala correspondientes a una longitud de ventana de 2304 para
STOP_START_1152_WINDOW y STOP_1152_WINDOW a 8 kHz

fs [kHz]	8		
num_swb_long_window	40		
swb	swb_offset_long_window	swb	Swb_offset_long_window
0	0	21	288
1	12	22	308
2	24	23	328
3	36	24	348
4	48	25	372
5	60	26	396
6	72	27	420
7	84	28	448
8	96	29	476
9	108	30	508
10	120	31	544
11	132	32	580
12	144	33	620
13	156	34	664
14	172	35	712
15	188	36	764
16	204	37	820
17	220	38	880
18	236	39	944
19	252		1152
20	268		

Tabla 7 – bandas de factores de escala correspondientes a una longitud de ventana de 2304 para STOP_START_1152_WINDOW y STOP_1152_WINDOW a 11,025, 12 y 16 kHz

fs [kHz]	11.025, 12, 16
num_swb_long_window	43
swb	swb_offset_long_window
0	0
1	8
2	16
3	24
4	32
5	40
6	48
7	56
8	64
9	72
10	80
11	88
12	100
13	112
14	124
15	136
16	148
17	160
18	172
19	184
20	196
21	212

swb	swb_offset_long_window
22	228
23	244
24	260
25	280
26	300
27	320
28	344
29	368
30	396
31	424
32	456
33	492
34	532
35	572
36	616
37	664
38	716
39	772
40	832
41	896
42	960
	1152

**Tabla 8 – bandas de factores de escala correspondientes a una longitud de ventana de 2304 para
STOP_START_1152_WINDOW y STOP_1152_WINDOW a 22.05 y 24 kHz**

fs [kHz]	22.05 and 24
num_swb_long_window	47
swb	swb_offset_long_window
0	0
1	4
2	8
3	12
4	16
5	20
6	24
7	28
8	32
9	36
10	40
11	44
12	52
13	60
14	68
15	76
16	84
17	92
18	100
19	108
20	116
21	124
22	136
23	148

swb	swb_offset_long_window
24	160
25	172
26	188
27	204
28	220
29	240
30	260
31	284
32	308
33	336
34	364
35	396
36	432
37	468
38	508
39	552
40	600
41	652
42	704
43	768
44	832
45	896
46	960
	1152

**Tabla 9 – bandas de factores de escala correspondientes a una longitud de ventana de 2304 para
STOP_START_1152_WINDOW y STOP_1152_WINDOW a 64 kHz**

fs [kHz]	64
num_swb_long_window	47 (46)
swb	swb_offset_long_window
0	0
1	4
2	8
3	12
4	16
5	20
6	24
7	28
8	32
9	36
10	40
11	44
12	48
13	52
14	56
15	64
16	72
17	80
18	88
19	100
20	112
21	124
22	140
23	156

swb	swb_offset_long_window
24	172
25	192
26	216
27	240
28	268
29	304
30	344
31	384
32	424
33	464
34	504
35	544
36	584
37	624
38	664
39	704
40	744
41	784
42	824
43	864
44	904
45	944
46	984
	1152

Tabla 10 – bandas de factores de escala correspondientes a una longitud de ventana de 2304 para STOP_START_1152_WINDOW y STOP_1152_WINDOW a 88,2 y 96 kHz

fs [kHz]	88.2 y 96		
num_swb_long_window	41		
swb	swb_offset_long_window	swb	swb_offset_long_window
0	0	21	120
1	4	22	132
2	8	23	144
3	12	24	156
4	16	25	172
5	20	26	188
6	24	27	212
7	28	28	240
8	32	29	276
9	36	30	320
10	40	31	384
11	44	32	448
12	48	33	512
13	52	34	576
14	56	35	640
15	64	36	704
16	72	37	768
17	80	38	832
18	88	39	896
19	96	40	960
20	108		1152

5 1.7 Referencias de las tablas de bandas de factores de escala

Por todas las demás tablas de bandas de factores de escala remitirse por favor a ISO/IEC 14496-3, subparte 4, sección 4.5.4 Tabla 4.129 a Tabla 4.147.

1.8 Cuantificación

10

Para la cuantificación de los coeficientes espectrales AAC en el codificador se utiliza un cuantificador no uniforme. Por lo tanto, el decodificador debe realizar una cuantificación inversa no uniforme después de la decodificación de Huffman de los factores de escala (ver la subcláusula 6.3) y la decodificación sin ruido de los datos espectrales (ver la subcláusula 6.1).

15

Para la cuantificación de los coeficientes espectrales TCX, se utiliza un cuantificador uniforme. No es necesaria la cuantificación inversa en el decodificador después de la decodificación sin ruido de los datos espectrales.

20

2. Banco de Filtros y conmutación de bloques

2.1 Descripción de la herramienta

La representación de tiempo/frecuencia de la señal se mapea sobre el dominio del tiempo alimentándola al módulo de banco de filtros. Este módulo consiste en una transformada discreta de coseno modificada (MDCT) y una ventana y una función de traslape-suma. Para adaptar la resolución de tiempo/frecuencia del banco de filtro a las características de la señal de entrada, también se adopta una herramienta de conmutación de bloques. N representa la longitud de ventana, donde N es una función de la secuencia de ventanas (**window_sequence**) (ver la subcláusula 1.1). Por cada canal, los valores de tiempo frecuencia $N/2 X_{i,k}$ se transforman en los valores en dominio del tiempo $N x_{i,n}$ por medio de la IMDCT. Después de aplicar la función de ventana, por cada canal, se agrega la primera mitad de la secuencia $z_{i,n}$ a la segunda mitad de la secuencia visualizada en ventana del bloque anterior $z_{(i-1),n}$ para reconstruir las muestras de salida por cada canal $out_{i,n}$.

2.2 Definiciones

window_sequence (secuencia_ventana) 2 bits que indican qué secuencia de ventana (es decir, tamaño de bloque) se utiliza.

window_shape (forma_ventana) 1 bit que indica qué función de ventana se ha seleccionado.

La Fig. 13C ilustra las ocho secuencias de ventana (**window_sequences**) (ONLY_LONG_SEQUENCE, LONG_START_SEQUENCE, EIGHT_SHORT_SEQUENCE, LONG_STOP_SEQUENCE, STOP_START_SEQUENCE, STOP_1152_SEQUENCE, LPD_START_SEQUENCE, STOP_START_1152_SEQUENCE).

(Sólo secuencia larga, secuencia de inicio larga, Ocho secuencias Cortas, secuencia de terminación larga, secuencia de terminación_iniciación, Secuencia de terminación_1152_ secuencia de iniciación LPD, secuencia de terminación, iniciación_1152)

A continuación, LPD_SEQUENCE se refiere a todas las combinaciones de modos de ventanas/modos de codificación admitidas dentro del denominado códec de dominio de predicción lineal (ver la sección 1.3). En el contexto de la decodificación de un marco en el dominio de la frecuencia es importante saber sólo si un marco siguiente está codificado con los modos de codificación de dominio LP, lo que está representado por una LPD_SEQUENCE. Sin embargo, se aborda la estructura exacta dentro de la LPD_SEQUENCE (secuencia LPD) al decodificar el marco codificado en el dominio LP.

2.3 Proceso de decodificación

2.3.1 IMDCT

La expresión analítica del IMDCT es:

$$x_{i,n} = \frac{2}{N} \sum_{k=0}^{\frac{N}{2}-1} spec[i][k] \cos\left(\frac{2\pi}{N}\left(n+n_0\right)\left(k+\frac{1}{2}\right)\right) \quad \text{correspondiente a } 0 \leq n < N$$

5 donde:

n = índice de muestreo

i = índice de ventana

k = índice de coeficientes espectrales

10 N = longitud de ventana basada en el valor window_sequence (secuencia de ventana)

$$n_0 = (N / 2 + 1) / 2$$

La longitud de ventana de síntesis N correspondiente a la transformada inversa es una función del elemento de sintaxis **window_sequence** y del contexto algorítmico.
15 Se define de la siguiente manera:

Longitud de ventana 2304:

$$N = \begin{cases} 2304, & \text{si STOP_1152_SEQUENCE} \\ 2304, & \text{si STOP_START_1152_SEQUENCE} \end{cases}$$

20

Longitud de ventana 2048:

$$N = \begin{cases} 2048, & \text{si ONLY_LONG_SEQUENCE} \\ 2048, & \text{si LONG_START_SEQUENCE} \\ 256, & \text{si EIGHT_SHORT_SEQUENCE} \\ 2048, & \text{si LONG_STOP_SEQUENCE} \\ 2048, & \text{si STOP_START_SEQUENCE} \\ 2048, & \text{si LPD_START_SEQUENCE} \end{cases}$$

25

Las transiciones de bloques significativas son las siguientes:

De ONLY_LONG_SEQUENCE a $\begin{cases} \text{ONLY_LONG_SEQUENCE} \\ \text{LONG_START_SEQUENCE} \\ \text{LPD_START_SEQUENCE} \end{cases}$

30 de LONG_START_SEQUENCE a $\begin{cases} \text{EIGHT_SHORT_SEQUENCE} \\ \text{LONG_STOP_SEQUENCE} \end{cases}$

- de LONG_STOP_SEQUENCE a $\begin{cases} \text{ONLY_LONG_SEQUENCE} \\ \text{LONG_START_SEQUENCE} \\ \text{LPD_START_SEQUENCE} \end{cases}$
- de EIGHT_SHORT_SEQUENCE a $\begin{cases} \text{EIGHT_SHORT_SEQUENCE} \\ \text{LONG_STOP_SEQUENCE} \\ \text{STOP_START_SEQUENCE} \end{cases}$
- 5 de LPD_SEQUENCE a $\begin{cases} \text{LPD_SEQUENCE} \\ \text{STOP_1152_SEQUENCE} \\ \text{STOP_START_1152_SEQUENCE} \end{cases}$
- de STOP_START_SEQUENCE a $\begin{cases} \text{EIGHT_SHORT_SEQUENCE} \\ \text{LONG_STOP_SEQUENCE} \end{cases}$
- 10 de LPD_START_SEQUENCE a $\{ \text{LPD_SEQUENCE} \}$
- de STOP_1152_SEQUENCE a $\begin{cases} \text{ONLY_LONG_SEQUENCE} \\ \text{LONG_START_SEQUENCE} \end{cases}$
- de STOP_START_1152_SEQUENCE a $\begin{cases} \text{EIGHT_SHORT_SEQUENCE} \\ \text{LONG_STOP_SEQUENCE} \end{cases}$
- 15

2.3.2 Visualización en ventanas y conmutación de bloques

Dependiendo del elemento **window_sequence** (secuencia de ventanas) **window_shape** (forma de ventana) se utilizan diferentes ventanas de transformación. Una combinación de las mitades de ventanas descritas a continuación ofrece todas las secuencias de ventanas posibles.

20

Para **window_shape** == 1, los coeficientes de ventanas están dados por la ventana derivada de Kaiser - Bessel (KBD, a saber:

$$W_{KBD_1152, N}(n) = \begin{cases} \frac{\sum_{p=0}^n [W'(p, \alpha)]}{\sum_{p=0}^{N/2} [W'(p, \alpha)]} & \text{para } 0 \leq n < \frac{N}{2} \end{cases}$$

25

$$W_{KBD_DER, N}(n) = \sqrt{\frac{\sum_{p=0}^{N-n-1} [W'(p, \alpha)]}{\sum_{p=0}^{N/2} [W'(p, \alpha)]}} \quad \text{para } \frac{N}{2} \leq n < N$$

donde:

W' , la función de ventana de modulo central de Kaiser – Bessel, ver también [5], se define de la siguiente manera:

5

$$W'(n, \alpha) = \frac{I_0 \left[\pi \alpha \sqrt{1.0 - \left(\frac{n - N/4}{N/4} \right)^2} \right]}{I_0[\pi \alpha]} \quad \text{en el caso } 0 \leq n \leq \frac{N}{2}$$

$$I_0[x] = \sum_{k=0}^{\infty} \frac{\left(\frac{x}{2} \right)^k}{k!}$$

$$\alpha = \text{factor alfa ventana princ} \quad ; \quad \alpha = \begin{cases} 4 \text{ para } N = 2048 (1920) \\ 6 \text{ para } N = 256 (240) \end{cases}$$

- 10 De lo contrario, en el caso de **window_shape == 0**, se emplea una ventana de seno de acuerdo con lo siguiente:

$$W_{SEN_IZQ, N}(n) = \sin\left(\frac{\pi}{N} \left(n + \frac{1}{2}\right)\right) \quad \text{para } 0 \leq n < \frac{N}{2}$$

$$W_{SEN_DER, N}(n) = \sin\left(\frac{\pi}{N} \left(n + \frac{1}{2}\right)\right) \quad \text{para } \frac{N}{2} \leq n < N$$

15

La longitud de ventana N puede ser 2048 (1920) o 256 (240) en el caso de la KBD y la ventana de seno. En el caso de STOP_1152_SEQUENCE y STOP_START_1152_SEQUENCE, N puede ser aún 2048 o 256, las pendientes de ventana son similares pero las regiones planas superiores son más largas.

20

Sólo en el caso de LPD_START_SEQUENCE la derecha de la ventana es una ventana de seno de 64 muestras.

- La manera de obtener las secuencias de ventanas posibles se explica en los puntos a)–h) de esta subcláusula.

25

En el caso de todos los tipos de secuencias de ventana, la forma de ventana de la mitad izquierda de la primera ventana de transformación está determinada por la forma de ventana del bloque anterior. La siguiente fórmula expresa este hecho:

30

$$W_{LEFT,N}(n) = \begin{cases} W_{KBD_LEFT,N}(n), & \text{si } window_shape_previous_block == 1 \\ W_{SEN_LEFT,N}(n), & \text{si } window_shape_previous_block == 0 \end{cases}$$

en la cual:

- 5 *window_shape_previous_block*: forma de ventana del bloque anterior (i-1).
para que se decodifique el primer bloque de datos sin procesar ("raw_data_block()")
las formas de ventana ("**window_shape**") de la mitad izquierda y derecha de la
ventana son idénticas.

10

a) ONLY_LONG_SEQUENCE (SÓLO SECUENCIA LARGA):

- 15 La secuencia de ventana **window_sequence** == ONLY_LONG_SEQUENCE es igual
a una LONG_WINDOW (ventana larga) con una longitud total de ventana N_l de
2048 (1920).

En el caso de **window_shape** == 1 la ventana correspondiente a
ONLY_LONG_SEQUENCE se expresa de la siguiente manera:

20
$$W(n) = \begin{cases} W_{LEFT,N_l}(n), & \text{para } 0 \leq n < N_l/2 \\ W_{KBD_RIGHT,N_l}(n), & \text{para } N_l/2 \leq n < N_l \end{cases}$$

Si **window_shape** == 0 la ventana correspondiente a ONLY_LONG_SEQUENCE se
puede describir de la siguiente manera:

25
$$W(n) = \begin{cases} W_{LEFT,N_l}(n), & \text{para } 0 \leq n < N_l/2 \\ W_{SEN_RIGHT,N_l}(n), & \text{para } N_l/2 \leq n < N_l \end{cases}$$

Después de la visualización en ventanas, los valores en dominio del tiempo ($z_{i,n}$) se
pueden expresar de la siguiente manera:

30
$$z_{i,n} = w(n) \cdot x_{i,n};$$

b) LONG_START_SEQUENCE (SECUENCIA DE INICIO LARGA):

- 35 La LONG_START_SEQUENCE es necesaria para obtener un traslape correcto y
agregar un bloque de transición de una ONLY_LONG_SEQUENCE
(SÓLO_SECUENCIA_LARGA) a una EIGHT_SHORT_SEQUENCE
(OCHO_SECUENCIA_CORTA) .

LA Longitud de ventana N_l y N_s se ajusta en 2048 (1920) y 256 (240) respectivamente.

Si **window_shape** == 1 la ventana correspondiente a LONG_START_SEQUENCE se presenta de la siguiente manera:

$$5 \quad W(n) = \begin{cases} W_{LEFT, N_l}(n), & \text{para } 0 \leq n < N_l / 2 \\ 10, & \text{para } N_l / 2 \leq n < \frac{3N_l - N_s}{4} \\ W_{KBD_RIGHT, N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{para } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{para } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

Si **window_shape** == 0 la ventana correspondiente a LONG_START_SEQUENCE se asemeja a:

$$10 \quad W(n) = \begin{cases} W_{LEFT, N_l}(n), & \text{para } 0 \leq n < N_l / 2 \\ 10, & \text{para } N_l / 2 \leq n < \frac{3N_l - N_s}{4} \\ W_{SEN_RIGHT, N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{para } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{para } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

Los valores en ventana en dominio del tiempo se pueden calcular con la fórmula explicada en a).

15

c) EIGHT_SHORT (OCHO CORTAS)

La secuencia de ventanas (**window_sequence**) == EIGHT_SHORT comprende ocho ventanas cortas traslapadas y sumadas SHORT_WINDOWS con una longitud de N_s de 256 (240) cada una. La longitud total de la secuencia de ventanas **window_sequence** junto con los ceros anteriores y posteriores es de 2048 (1920). En primer lugar se coloca cada uno de los ocho bloques cortos en ventanas por separado. El número de bloques cortos se indexa con la variable $j = 0, \dots, M - 1$ ($M = N_l / N_s$).

25

La forma de ventana (**window_shape**) del bloque anterior influye sobre el primero de los ocho bloques cortos ($W_0(n)$) solamente. Si **window_shape** == 1 las funciones de la ventana se pueden expresar de la siguiente manera:

$$W_0(n) = \begin{cases} W_{IZQ, N_s}(n), & \text{para } 0 \leq n < N_s / 2 \\ W_{KBD_DER, N_s}(n), & \text{para } N_s / 2 \leq n < N_s \end{cases}$$

$$W_{1-(M-1)}(n) = \begin{cases} W_{KBD_IZQ, N_s}(n) & \text{para } 0 \leq n < N_s / 2 \\ W_{KBD_DER, N_s}(n), & \text{para } N_s / 2 \leq n < N_s \end{cases}$$

30

De lo contrario, si **window_shape** == 0, las funciones de la ventana se pueden describir de la siguiente manera:

$$w_0(n) = \begin{cases} W_{LEFT, N_s}(n), & \text{for } 0 \leq n < N_s/2 \\ W_{SIN_RIGHT, N_s}(n), & \text{for } N_s/2 \leq n < N_s \end{cases}$$

$$w_0(n) = \begin{cases} W_{IZQ, N_s}(n), & \text{para } 0 \leq n < N_s/2 \\ W_{SEN_DER, N_s}(n), & \text{para } N_s/2 \leq n < N_s \end{cases}$$

$$w_{1-(M-1)}(n) = \begin{cases} W_{SEN_IZQ, N_s}(n), & \text{para } 0 \leq n < N_s/2 \\ W_{SEN_DER, N_s}(n), & \text{para } N_s/2 \leq n < N_s \end{cases}$$

El traslape y suma entre la secuencia de ventanas de ocho cortas (EIGHT_SHORT) que da lugar a los valores en domino de tiempo en ventana $z_{i,n}$ se describe de la siguiente manera:

10

$$z_{i,n} = \begin{cases} 0, & \text{para } 0 \leq n < \frac{N-l-N_s}{4} \\ x_{0,n-\frac{N-l-N_s}{4}} \cdot W_0(n - \frac{N-l-N_s}{4}), & \text{para } \frac{N-l-N_s}{4} \leq n < \frac{N-l+N_s}{4} \\ x_{j-1,n-\frac{N-l+(2j-3)N_s}{4}} \cdot W_{j-1}(n - \frac{N-l+(2j-3)N_s}{4}) + x_{j,n-\frac{N-l+(2j-1)N_s}{4}} \cdot W_j(n - \frac{N-l+(2j-1)N_s}{4}), & \text{para } 1 \leq j < M, \frac{N-l+(2j-1)N_s}{4} \leq n < \frac{N-l+(2j+1)N_s}{4} \\ x_{M-1,n-\frac{N-l+(2M-3)N_s}{4}} \cdot W_{M-1}(n - \frac{N-l+(2M-3)N_s}{4}), & \text{para } \frac{N-l+(2M-1)N_s}{4} \leq n < \frac{N-l+(2M+1)N_s}{4} \\ 0, & \text{para } \frac{N-l+(2M+1)N_s}{4} \leq n < N-l \end{cases}$$

d) LONG_STOP_SEQUENCE (SECUENCIA LARGA DE TERMINACIÓN)

15

Esta secuencia de ventanas es necesaria para conmutar de EIGHT_SHORT_SEQUENCE de regreso a ONLY_LONG_SEQUENCE.

Si **window_shape** == 1 la ventana correspondiente a LONG_STOP_SEQUENCE se calcula de la siguiente manera:

20

$$w(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N-l-N_s}{4} \\ W_{IZQ, N_s}(n - \frac{N-l-N_s}{4}), & \text{para } \frac{N-l-N_s}{4} \leq n < \frac{N-l+N_s}{4} \\ 1.0, & \text{para } \frac{N-l+N_s}{4} \leq n < N-l/2 \\ W_{KBD_DER, N_l}(n), & \text{para } N-l/2 \leq n < N-l \end{cases}$$

Si **window_shape** == 0 la ventana correspondiente a LONG_START_SEQUENCE (SECUENCIA LARGA DE INICIO) se determina por:

$$w(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < \frac{N_l - N_s}{4} \\ W_{IZQ_N_s}(n - \frac{N_l - N_s}{4}), & \text{para } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ 1.0, & \text{para } \frac{N_l + N_s}{4} \leq n < N_l/2 \\ W_{SEN_DER_N_l}(n), & \text{para } N_l/2 \leq n < N_l \end{cases}$$

Los valores en dominio del tiempo visualizados en ventana se pueden calcular con la fórmula explicada en a).

5

e) STOP_START_SEQUENCE (SECUENCIA DE TERMINACIÓN_INICIO:

La secuencia de terminación_inicio (STOP_START_SEQUENCE) es necesaria para obtener un traslape y suma correctos correspondientes a una transición de bloques de una secuencia de ocho cortas (EIGHT_SHORT_SEQUENCE) a una secuencia de ocho cortas (EIGHT_SHORT_SEQUENCE) cuando sólo se necesita una secuencia larga (ONLY_LONG_SEQUENCE).

Las longitudes de ventana N_l y N_s se ajustan a 2048 (1920) y 256 (240) respectivamente.

15

Si **window_shape** == 1 la ventana correspondiente a STOP_START_SEQUENCE (SECUENCIA TERMINACIÓN_INICIO) se da de la siguiente manera:

$$W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l - N_s}{4} \\ W_{IZQ_N_s}(n - \frac{N_l - N_s}{4}), & \text{para } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ 1.0, & \text{para } \frac{N_l + N_s}{4} \leq n < \frac{3N_l - N_s}{4} \\ W_{KBD_DER_N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{para } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{para } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

20

Si **window_shape** == 0 la ventana correspondiente a la secuencia terminación_inicio STOP_START_SEQUENCE tiene el siguiente aspecto:

$$W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l - N_s}{4} \\ W_{IZQ_N_s}(n - \frac{N_l - N_s}{4}), & \text{para } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ 1.0, & \text{para } \frac{N_l + N_s}{4} \leq n < \frac{3N_l - N_s}{4} \\ W_{SEN_DER_N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{para } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{para } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

Los valores en dominio de tiempo visualizados en ventana se pueden calcular con la fórmula explicada en a).

5

f) LPD_START_SEQUENCE (SECUENCIA DE INICIO LPD):

La secuencia de inicio LPD (LPD_START_SEQUENCE) es necesaria para obtener un traslape y suma correctos para una transición de bloques de ONLY_LONG_SEQUENCE (sólo secuencia larga) a una LPD_SEQUENCE (secuencia LPD).

Las longitudes de ventana N_l y N_s se ajustan a 2048 (1920) y 256 (240) respectivamente.

15

Si **window_shape** == 1 la ventana correspondiente a LPD_START_SEQUENCE se da de la siguiente manera:

$$W(n) = \begin{cases} W_{IZQ_N_l}(n), & \text{para } 0 \leq n < \frac{N_l}{2} \\ 1.0, & \text{para } \frac{N_l}{2} \leq n < \frac{3N_l - N_s}{4} \\ W_{KBD_DER_N_s/2}(n + \frac{N_s}{4} - \frac{3N_l - N_s}{4}), & \text{para } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l}{4} \\ 0.0, & \text{para } \frac{3N_l}{4} \leq n < N_l \end{cases}$$

20

Si **window_shape** == 0 la ventana correspondiente a LPD_START_SEQUENCE tiene el siguiente aspecto:

$$W(n) = \begin{cases} W_{IZQ_N_l}(n), & \text{para } 0 \leq n < \frac{N_l}{2} \\ 1.0, & \text{para } \frac{N_l}{2} \leq n < \frac{3N_l - N_s}{4} \\ W_{SEN_DER_N_s/2}(n + \frac{N_s}{4} - \frac{3N_l - N_s}{4}), & \text{para } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l}{4} \\ 0.0, & \text{para } \frac{3N_l}{4} \leq n < N_l \end{cases}$$

Los valores en dominio de tiempo visualizados en ventana se pueden calcular con la fórmula explicada en a).

5 **g) STOP_1152_SEQUENCE (SECUENCIA TERMINACIÓN_1152) :**

La secuencia STOP_1152_SEQUENCE es necesaria para obtener un traslape y suma correctos para una transición de bloques de una secuencia LPD LPD_SEQUENCE a sólo secuencia larga ONLY_LONG_SEQUENCE.

10

Las longitudes de ventana N_l y N_s se ajustan a 2048 (1920) y 256 (240) respectivamente.

Si **window_shape** == 1 la ventana correspondiente a STOP_1152_SEQUENCE se da de la siguiente manera:

$$W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l}{4} \\ W_{IZQ_N_s}(n - \frac{N_l}{4}), & \text{para } \frac{N_l}{4} \leq n < \frac{N_l + 2N_s}{4} \\ 1.0, & \text{para } \frac{N_l + 2N_s}{4} \leq n < \frac{2N_l + 3N_s}{4} \\ W_{KBD_DER_N_l}(n + \frac{N_l}{2} - \frac{2N_l + 3N_s}{4}), & \text{para } \frac{2N_l + 3N_s}{4} \leq n < N_l + \frac{3N_s}{4} \\ 0.0, & \text{para } N_l + \frac{3N_s}{4} \leq n < N_l + N_s \end{cases}$$

15

Si **window_shape** == 0 la ventana correspondiente a STOP_1152_SEQUENCE tiene el siguiente aspecto:

$$W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l}{4} \\ W_{IZQ_N_s}(n - \frac{N_l}{4}), & \text{para } \frac{N_l}{4} \leq n < \frac{N_l + 2N_s}{4} \\ 1.0, & \text{para } \frac{N_l + 2N_s}{4} \leq n < \frac{2N_l + 3N_s}{4} \\ W_{SEN_DER_N_l}(n + \frac{N_l}{2} - \frac{2N_l + 3N_s}{4}), & \text{para } \frac{2N_l + 3N_s}{4} \leq n < N_l + \frac{3N_s}{4} \\ 0.0, & \text{para } N_l + \frac{3N_s}{4} \leq n < N_l + N_s \end{cases}$$

20

Los valores en dominio de tiempo visualizados en ventana se pueden calcular con la fórmula explicada en a).

h) STOP_START_1152_SEQUENCE

25 **(SECUENCIA TERMINACIÓN_INICIO_1152):**

La secuencia STOP_START_1152_SEQUENCE es necesaria para obtener un traslape y suma correctos para una transición de bloques de una secuencia LPD LPD_SEQUENCE a una secuencia de ocho cortas EIGHT_SHORT_SEQUENCE cuando sólo se necesita ONLY_LONG_SEQUENCE (sólo secuencia larga).

5

Las longitudes de ventana N_l y N_s se ajustan a 2048 (1920) y 256 (240) respectivamente.

Si **window_shape** == 1 la ventana correspondiente a la secuencia

10 STOP_START_SEQUENCE se ve de la siguiente manera:

$$15 \quad W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l}{4} \\ W_{IZQ, N_s}(n - \frac{N_l}{4}), & \text{para } \frac{N_l}{4} \leq n < \frac{N_l + 2N_s}{4} \\ 1.0, & \text{para } \frac{N_l + 2N_s}{4} \leq n < \frac{3N_l}{4} + \frac{N_s}{2} \\ W_{KBD_DER, N_s}(n + \frac{N_s}{2} - \frac{3N_l}{4} + \frac{N_s}{2}), & \text{para } \frac{3N_l}{4} + \frac{N_s}{2} \leq n < \frac{3N_l}{4} + N_s \\ 0.0, & \text{para } \frac{3N_l}{4} + N_s \leq n < N_l + N_s \end{cases}$$

Si **window_shape** == 0 la ventana correspondiente a la secuencia terminación_inicio

15 STOP_START_SEQUENCE tiene el siguiente aspecto:

$$W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l}{4} \\ W_{IZQ, N_s}(n - \frac{N_l}{4}), & \text{para } \frac{N_l}{4} \leq n < \frac{N_l + 2N_s}{4} \\ 1.0, & \text{para } \frac{N_l + 2N_s}{4} \leq n < \frac{3N_l}{4} + \frac{N_s}{2} \\ W_{SEN_DER, N_s}(n + \frac{N_s}{2} - \frac{3N_l}{4} + \frac{N_s}{2}), & \text{para } \frac{3N_l}{4} + \frac{N_s}{2} \leq n < \frac{3N_l}{4} + N_s \\ 0.0, & \text{para } \frac{3N_l}{4} + N_s \leq n < N_l + N_s \end{cases}$$

20 Los valores en dominio de tiempo visualizados en ventana se pueden calcular con la fórmula explicada en a).

2.3.3 Traslape y suma con secuencia de ventana previa

Aparte del traslape y suma dentro de la secuencia de ventana EIGHT_SHORT, la primera parte (izquierda) de cada secuencia de ventana (**window_sequence**) se

traslapa y suma con la segunda parte (derecha) de la secuencia de ventana anterior, dando lugar a los valores finales en dominio del tiempo $out_{i,n}$. La expresión matemática correspondiente a esta operación se puede describir de la siguiente manera.

- 5 En el caso de ONLY_LONG_SEQUENCE, LONG_START_SEQUENCE, EIGHT_SHORT_SEQUENCE, LONG_STOP_SEQUENCE, STOP_START_SEQUENCE, LPD_START_SEQUENCE:

$$out_{i,n} = z_{i,n} + z_{i-1, n + \frac{N}{2}} ; \text{ para } 0 \leq n < \frac{N}{2}, \quad N = 2048 \text{ (1920)}$$

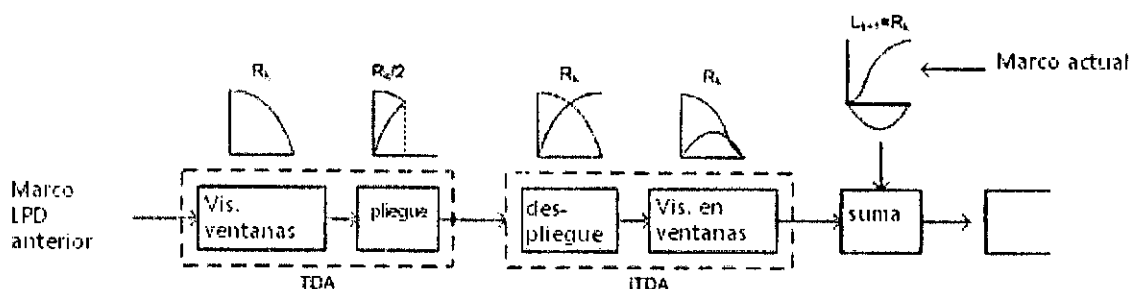
Y en el caso de STOP_1152_SEQUENCE, STOP_START_1152_SEQUENCE:

$$out_{i,n} = z_{i,n} + z_{i-1, n + \frac{N_l}{2} + \frac{3N_s}{4}} ; \text{ para } 0 \leq n < \frac{N_l}{2}, \quad N_l = 2048, N_s = 256$$

- 15 En el caso de LPD_START_SEQUENCE, la siguiente secuencia es una secuencia LPD_SEQUENCE. Se aplica una ventana SEN o KBD a la parte izquierda de la LPD_SEQUENCE para obtener un traslape y suma apropiados.

$$W_{SEN_IZQ, N}(n) = \text{seno}\left(\frac{\pi}{N}\left(n + \frac{1}{2}\right)\right) \text{ para } 0 \leq n < \frac{N}{2} \text{ Con } N = 128$$

- 20 En el caso de STOP_1152_SEQUENCE, STOP_START_1152_SEQUENCE la secuencia anterior es una secuencia LPD LPD_SEQUENCE. Se aplica una ventana TDAC a la parte derecha de la LPD_SEQUENCE obtener un traslape y suma apropiados.



25 3. IMDCT

Ver la subcláusula 2.3.1

3.1 Selección de Ventanas y conmutación de bloques

Dependiendo del elemento **window_shape** se utilizan diferentes prototipos de ventanas por transformadas con sobremuestreo y la longitud de las ventanas sobremuestreadas es

$$N_{os} = 2 \cdot n_long \cdot os_factor_gan$$

En el caso de **window_shape** == 1, los coeficientes de ventana están dados por la ventana de derivada de Kaiser - Bessel (KBD), a saber:

$$W_{KBD}\left(n - \frac{N_{os}}{2}\right) = \sqrt{\frac{\sum_{p=0}^{N_{os}/2} [W(p, \alpha)]}{\sum_{p=0}^{N_{os}/2} [W(p, \alpha)]}} \quad \text{para} \quad \frac{N_{os}}{2} \leq n < N_{os}$$

donde:

W' , función de ventana en modo Kaiser - Bessel, ver también [5], se define de la siguiente manera

$$W'(n, \alpha) = \frac{I_0\left[\pi\alpha\sqrt{1.0 - \left(\frac{n - N_{os}/4}{N_{os}/4}\right)^2}\right]}{I_0[\pi\alpha]} \quad \text{para} \quad 0 \leq n \leq \frac{N_{os}}{2}$$

$$I_0[x] = \sum_{k=0}^{\infty} \frac{\left(\frac{x}{2}\right)^k}{k!}$$

α = factor alfa de ventana de modo principal, $\alpha = 4$

De lo contrario, para **window_shape** == 0, se emplea una ventana de seno, a saber:

$$W_{SEN}\left(n - \frac{N_{os}}{2}\right) = \sin\left(\frac{\pi}{N_{os}}\left(n + \frac{1}{2}\right)\right) \quad \text{para} \quad \frac{N_{os}}{2} \leq n < N_{os}$$

En todos los tipos de secuencias de ventanas, el prototipo utilizado para la parte izquierda es el determinado por la forma de ventana del bloque anterior. La siguiente fórmula expresa este hecho:

$$forma_ventana_izquierda[n] = \begin{cases} W_{KBD}[n] & \text{si } window_shape_bloque \text{ anterior} = 1 \\ W_{SEN}[n] & \text{si } window_shape_bloque \text{ anterior} = 0 \end{cases}$$

Del mismo modo, el prototipo para la forma de la ventana derecha se determina mediante la siguiente fórmula:

$$\text{forma_ventaña_derecha}[n] = \begin{cases} W_{KBD}[n] & \text{si } \text{window_shape} == 1 \\ W_{SEN}[n] & \text{si } \text{window_shape} == 0 \end{cases}$$

Dado que las longitudes de transición ya están determinadas, sólo tiene que diferenciarse entre EIGHT_SHORT_SEQUENCES y todas las demás:

5

a) EIGHT SHORT SEQUENCE:

La siguiente porción en código c describe la visualización en ventanas y el traslape-suma interno de una EIGHT_SHORT_SEQUENCE:

10

```
tw_windowing_short(X[],z[],first_pos,last_pos,warpe_trans_len_left,warped_trans_le
n_right,left_window_shape[],right_window_shape[]){
```

```
    compensación = n_long - 4*n_short - n_short/2;
```

15

```
    tr_scale_1 = 0.5*n_long/warped_trans_len_left*os_factor_win;
```

```
    tr_pos_l = warped_trans_len_left+(first_pos-n_long/2)+0.5)*tr_scale_l;
```

```
    tr_scale_r = 8*os_factor_win;
```

```
    tr_pos_r = tr_scale_r/2;
```

20

```
    para ( i = 0 ; i < n_short ; i++ ) {
```

```
        z[i] = X[0][i];
```

```
    }
```

25

```
    para (i=0;i<first_pos;i++)
```

```
        z[i] = 0.;
```

```
    para (i=n_long-1-first_pos;i>=first_pos;i--) {
```

```
        z[i] *= left_window_shape[floor(tr_pos_l)];
```

30

```
    tr_pos_l += tr_scale_l;
```

```
    }
```

```
    para (i=0;i<n_short;i++) {
```

```
        z[compensación+i+n_short]=
```

35

```
        X[0][i+n_short]*right_window_shape[floor(tr_pos_r)];
```

```
        tr_pos_r += tr_scale_r;
```

```
    }
```

```
    compensación += n_short;
```

40

```

para ( k = 1 ; k < 7 ; k++ ) {
    tr_scale_l = n_short*os_factor_win;
    tr_pos_l = tr_scale_l/2;
    tr_pos_r = os_factor_win*n_long-tr_pos_l;
5   para ( i = 0 ; i < n_short ; i++ ) {
        z[i + compensación] += X[k][i]*right_window_shape[floor(tr_pos_r)];
        z[compensación + n_short + i] =
            X[k][n_short + i]*right_window_shape[floor(tr_pos_l)];
        tr_pos_l += tr_scale_l;
10    tr_pos_r -= tr_scale_l;
    }
    compensación += n_short;
}

15   tr_scale_l = n_short*os_factor_win;
    tr_pos_l = tr_scale_l/2;

    para ( i = n_short - 1 ; i >= 0 ; i-- ) {
        z[i + offset] += X[7][i]*right_window_shape[(int) floor(tr_pos_l)];
20    tr_pos_l += tr_scale_l;
    }

    para ( i = 0 ; i < n_short ; i++ ) {
        z[compensación + n_short + i] = X[7][n_short + i];
25 }

    tr_scale_r = 0.5*n_long/warpedTransLenRight*os_factor_win;
    tr_pos_r = 0.5*tr_scale_r+.5;

30   tr_pos_r = (1.5*n_long-(float)wEnd-0.5+warpedTransLenRight)*tr_scale_r;
    para (i=3*n_long-1-last_pos ; i<=wEnd;i++) {
        z[i] *= right_window_shape[floor(tr_pos_r)];
        tr_pos_r += tr_scale_r;
    }

35   para(i=lsat_pos+1;i<2*n_long;i++)
        z[i] = 0.;

b) todos los demás:

40   tw_windowing_long(X[],z[],first_pos,last_pos,warpe_trans_len_left,warped_trans_len
    _right,left_window_shape[],right_window_shape[]){

```

```

    para (i=0;i<first_pos;i++)
        z[i] = 0.;
    para (i=last_pos+1;i<N;i++)
5      z[i] = 0.;

    tr_scale = 0.5*n_long/warped_trans_len_left*os_factor_win;
    tr_pos = (warped_trans_len_left+first_pos-N/4)+0.5)*tr_scale;

10   para (i=N/2-1-first_pos;i>=first_pos;i--) {
        z[i] = X[0][i]*left_window_shape[floor(tr_pos)];
        tr_pos += tr_scale;
    }

15   tr_scale = 0.5*n_long/warped_trans_len_right*os_factor_win;
    tr_pos = (3*N/4-last_pos-0.5+warped_trans_len_right)*tr_scale;

    para (i=3*N/2-1-last_pos;i<=last_pos;i++) {
        z[i] = X[0][i]*right_window_shape[floor(tr_pos)];
20     tr_pos += tr_scale;
    }
}
```

4. **TCX basada en MDCT**

25 **4.1 Descripción de la Herramienta**

 Cuando el **core_mode** es igual a 1 y cuando se seleccionan uno o más de tres modos TCX como codificación en “dominio de predicción lineal”, es decir, una de las cuatro entradas de serie de mod[] es superior a 0, se utiliza la herramienta TCVX basada en MDCT. La TCX basada en MDCT recibe los coeficientes espectrales

30 cuantificados del decodificador aritmético. En primer lugar se completan los coeficientes cuantificados mediante un ruido de confort (“comfort noise”) antes de la aplicación de una transformación inversa MDCT para obtener una síntesis ponderada en dominio del tiempo que luego es alimentada al filtro LPC de síntesis de

 ponderación

35

4.2 Definiciones

40	lg	Número de coeficientes espectrales cuantificados emitido por el decodificador aritmético
	factor_ruido	Índice de cuantificación de nivel de ruido
	nivel de ruido	Nivel de ruido inyectado en el espectro reconstruido

- noise[]
- Vector de ruido generado
- global_gain
- Índice de cuantificación de ganancia de reescalado
- g
- Ganancia de reescalado
- rms
- Media de la raíz cuadrada de la señal sintetizada en dominio del tiempo, x[],
- 5
- x[]
- Señal sintetizada en dominio del tiempo

4.3 Proceso de decodificación

- 10
- El TCX basado en MDCT solicita al decodificador aritmético un número de coeficientes espectrales cuantificados, lg, que está determinado por los valores mod[] y last_lpd_mode. Estos dos valores también definen la longitud y forma de las ventanas que se han de aplicar en la MDCT inversa. La ventana está compuesta por tres partes, un traslape del lado izquierdo de L muestras, una parte central de una de
- 15
- M muestras y una parte de traslape derecho de R muestras. Para obtener una ventana MDCT de longitud 2*lg, se agregan ZL ceros a la izquierda y ZR ceros del lado derecho, como se indica en la Fig. 14G correspondiente a la Tabla 3/Fig. 14F.

20

Tabla 3 – Número de Coeficientes Espectrales en Función de last_lpd_mode y mod[]

Valor de last_lpd_mode	valor de mod[x]	Número lg de coeficientes espectrales	ZL	L	M	R	ZR
0	1	320	160	0	256	128	96
0	2	576	288	0	512	128	224
0	3	1152	512	128	1024	128	512
1..3	1	256	64	128	128	128	64
1..3	2	512	192	128	384	128	192
1..3	3	1024	448	128	896	128	448

La ventana MDCT está dada por

$$W(n) = \begin{cases} 0 & \text{para } 0 \leq n < ZL \\ W_{SEN_IZQ_L}(n - ZL) & \text{para } ZL \leq n < ZL + L \\ 1 & \text{para } ZL + L \leq n < ZL + L + M \\ W_{SEN_DER_R}(n - ZL - L - M) & \text{para } ZL + L + M \leq n < ZL + L + M + R \\ 0 & \text{para } ZL + L + M + R \leq n < 2lg \end{cases}$$

Los coeficientes espectrales cuantificados, $quant[]$, aportados por el decodificador aritmético se completan con un ruido de confort. El nivel de ruido inyectado se determina por el factor ruido decodificado de la siguiente manera:

$$nivel_ruido = 0,0625 \cdot (8 - factor_ruido)$$

5

A continuación se computa un vector de ruido, $noise[]$, utilizando una función aleatoria, $random_sign()$, que produce al azar el valor -1 o +1.

$$noise[i] = random_sign() \cdot noise_level;$$

- 10 Los vectores $quant[]$ y $noise[]$ se combinan para formar el vector de coeficientes espectrales reconstruidos, $r[]$, en forma tal que las series de 8 ceros consecutivos de $quant[]$ son reemplazadas por los componentes de $noise[]$. Una serie de 8 no ceros se detecta de acuerdo con la fórmula:

$$\begin{cases} r[i] = 1 & \text{para } i \in [0, lg/6[\\ r[lg/6 + i] = \sum_{k=0}^7 |quant[lg/6 + 8 \cdot \lfloor i/8 \rfloor + k]| & \text{para } i \in [0, 7 \cdot lg/6[\end{cases}$$

- 15 Se obtiene el siguiente espectro reconstruido:

$$r[i] = \begin{cases} quant[i] & \text{si } r[i] = 1 \\ noise[i] & \text{de lo contrario} \end{cases}$$

Con anterioridad a la aplicación de MDCT inversa se aplica una deformación del espectro de acuerdo con los siguientes pasos:

- 20 1. se calcula la energía E_m del bloque de 8 dimensiones en el índice m por cada bloque en 8 dimensiones del primer cuarto del espectro
2. se computa la relación $R_m = \sqrt{E_m/E_l}$, donde l es el índice de bloques con un valor máximo de todos E_m
3. si $R_m < 0,1$, luego la serie $R_m = 0,1$
- 25 4. si $R_m < R_{m-1}$, luego la serie $R_m = R_{m-1}$

Cada bloque de 8 dimensiones que pertenece al primer cuarto del espectro se multiplica a continuación por el factor R_m .

- 30 Se alimenta el espectro reconstruido en un MDCT inverso. La señal de salida no visualizada en ventana, $x[]$, es reescalada por la ganancia, g , obtenida por la cuantificación inversa del índice de ganancia global $global_gain$:

$$g = 10^{global_gain/28 \cdot (2 \cdot rms)}$$

- 35 donde rms se calcula en términos de:

$$rms = \sqrt{\frac{\sum_{i=lg/2}^{3 \cdot lg/2 - 1} x^2[i]}{L + M + R}}$$

Entonces la señal sintetizada y reescalada en dominio del tiempo es igual a:

$$x_w[i] = x[i] \cdot g$$

Después se aplica el reescalado de la visualización en ventanas y la suma de traslape.

- 10 A continuación se filtra el objetivo TCX reconstruido $x(n)$ por medio de un filtro de síntesis ponderada inversa en estado cero $\hat{A}(z)(1 - \alpha z^{-1})/(\hat{A}(z/\lambda))$ para hallar la señal de excitación que se ha de aplicar al filtro de síntesis. Nótese que se utiliza el filtro LP interpolado por cada submarco en el filtrado. Una vez determinada la excitación, se reconstruye la señal filtrando la excitación a través del filtro de síntesis $1/\hat{A}(z)$ y
- 15 luego desenfazando mediante filtrado a través del filtro $1/(1 - 0.68z^{-1})$ como se describiera anteriormente.

- Nótese que la excitación también es necesaria para actualizar el libro de códigos adaptativos ACELP y permitir la conmutación de TCX a ACELP en un marco subsiguiente. Nótese asimismo que la longitud de la síntesis TCX está dada por la longitud de los marcos TCX (sin el traslape): 256, 512 o 1024 muestras correspondientes a mod[] de 1, 2 o 3 respectivamente.

Referencias de Normativas

- 25 [1] ISO/IEC 11172-3:1993, Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s, Parte 3: Audio.
- [2] ITU-T Rec.H.222.0(1995) | ISO/IEC 13818-1:2000, Information technology -
- 30 Generic coding of moving pictures and associated audio information: - Parte 1: Systems.
- [3] ISO/IEC 13818-3:1998, Information technology - Generic coding of moving pictures and associated audio information: - Parte 3: Audio.
- 35 [4] ISO/IEC 13818-7:2004, Information technology - Generic coding of moving pictures and associated audio information: - Parte 7: Advanced Audio Coding (AAC).
- [5] ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects -
- 40 Parte 1: Systems

- [6] ISO/IEC 14496-3:2005, Information technology – Coding of audio-visual objects – Parte 3: Audio
- 5 [7] ISO/IEC 23003-1:2007, Information technology — MPEG audio technologies — Parte 1: MPEG Surround
- [8] 3GPP TS 26.290 V6.3.0, Extended Adaptive Multi-Rate – Wideband (AMR-WB+) codec; Transcoding functions
- 10 [9] 3GPP TS 26.190, Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; Transcoding functions
- [10] 3GPP TS 26.090, Adaptive Multi-Rate (AMR) speech codec; Transcoding functions
- 15

Definiciones

- Se pueden encontrar definiciones en ISO/IEC 14496-3, subparte 1, subcláusula 1.3 (Términos y definiciones) y en 3GPP TS 26.290, sección 3 (Definiciones y abreviaturas).
- 20

- Si bien se han descrito algunos aspectos en el contexto de un aparato, está claro que estos aspectos sólo representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a un paso del método o a una característica de un paso del método. De manera análoga, los aspectos descritos en el contexto de un paso del método también representan una descripción de un bloque o elemento correspondiente o una característica de un aparato correspondiente.
- 25
- 30

- La señal de audio codificada de la invención puede ser almacenada en un medio de almacenamiento digital o puede ser transmitida en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión cableado tal como la Internet.
- 35

- Dependiendo de ciertos requisitos de implementación, las realizaciones de la presente invención se pueden implementar en hardware o en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo un disco blando, un DVD, un CD, una memoria ROM, PROM, EPROM, EEPROM o FLASH, que consta de señales de control legibles electrónicamente
- 40

almacenadas en los mismos, que cooperan (o tienen capacidad para cooperar) con un sistema de computación programable de tal manera que se ejecute el método respectivo.

- 5 Algunas realizaciones de acuerdo con la presente invención comprenden un portador de datos que consta de señales de control legibles electrónicamente con capacidad para cooperar con un sistema de computación programable de tal manera que se ejecute uno de los métodos descritos en la presente.
- 10 En general, las realizaciones de la presente invención se pueden implementar en forma de producto programa de computación con un código de programa que cumple la función de ejecutar uno de los métodos cuando el programa de computación se ejecuta en una computadora. El código de programa puede estar almacenado, por ejemplo, en un portador legible en una máquina.
- 15 Otras realizaciones comprenden el programa de computación para ejecutar uno de los métodos descritos en la presente, almacenados en un portador legible en una máquina.
- 20 En otras palabras, una realización del método de la presente invención consiste, por lo tanto, en un programa de computación que consta de un código de programa para ejecutar uno de los métodos descritos en la presente, cuando el programa de computación se ejecuta en una computadora.
- 25 Por lo tanto, otra realización de los métodos de la invención consiste 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 descritos en la presente.
- 30 Por lo tanto, otra realización del método de la invención consiste en un flujo de datos o una secuencia de señales que representan el programa de computación para ejecutar uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales puede estar configurada, por ejemplo, para ser transferido por medio de una conexión de comunicación de datos, por ejemplo por la Internet.
- 35 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 descritos en la presente.
- 40 Otra realización comprende una computadora que tiene, instalado en la misma, el programa de computación para ejecutar uno de los métodos descritos en la presente.

- En algunas realizaciones, se puede emplear un dispositivo lógico programable (por ejemplo una matriz de puertas programables) para ejecutar algunas o todas las funcionalidades de los métodos descritos en este documento. En algunas
- 5 realizaciones, una matriz de puertas programables en campo puede cooperar con un microprocesador para ejecutar uno de los métodos descritos en la presente. En general, los métodos se ejecutan preferentemente mediante cualquier aparato de hardware.
- 10 Las realizaciones precedentemente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles aquí descritos han de ser evidentes para otras personas con capacitación normal en la técnica. Por lo tanto, se pretende que sólo
- 15 detalles específicos presentados a manera de descripción y explicación de las realizaciones aquí presentadas.

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 codificador de audio para codificar una señal de audio que comprende:

una primera rama codificadora (400) para codificar una señal de audio utilizando un primer algoritmo de codificación para obtener una primera señal codificada, donde la primera rama codificadora comprende el primer convertidor (410) para convertir una señal de entrada en un dominio espectral;

una segunda rama codificadora (500) para codificar una señal de audio utilizando un segundo algoritmo de codificación para obtener una segunda señal codificada, donde el primer algoritmo de codificación es diferente del segundo algoritmo de codificación, donde la segunda rama codificadora comprende un convertidor de dominio para convertir una señal de entrada de un dominio de entrada en un dominio de salida y un segundo convertidor (523) para convertir una señal de entrada en un dominio espectral;

un conmutador (200) para conmutar entre la primera rama codificadora y la segunda rama codificadora de manera que, para una porción de la señal de audio de entrada, la primera señal codificada o la segunda señal codificada esté en una señal de salida del codificador;

un analizador de señales (300/525), que está configurado para analizar una porción de la señal de audio para determinar si la porción de la señal de audio está representada como primera señal codificada o segunda señal codificada en la señal de salida del codificador, donde el analizador de señales está configurado además para determinar en forma variable una resolución respectiva en tiempo/frecuencia del primer convertidor y el segundo convertidor, cuando se genera la primera señal codificada o la segunda señal codificada que representa la porción de la señal de audio; y

una interfaz de salida (800) para generar una señal de salida del codificador que comprende la primera señal codificada y la segunda señal codificada y una información que indica la primera señal codificada y la segunda señal codificada, y una información que indica la resolución en tiempo/frecuencia aplicada para codificar la primera señal codificada y para codificar la segunda señal codificada.

2. Un codificador de audio de acuerdo con la reivindicación 1, en el cual el analizador de señales (300, 525) está configurado para clasificar la porción de la señal de audio en forma de señal de audio de habla o señal de audio de música y para ejecutar la detección transitoria en el caso de una señal de música para determinar la resolución en tiempo/frecuencia del primer convertidor (410) o para ejecutar el procesamiento de análisis por síntesis para determinar la resolución en tiempo/frecuencia del segundo convertidor (523).
3. Un codificador de audio de acuerdo con la reivindicación 1 o 2, en el cual el primer convertidor (410) y el segundo convertidor (523) comprenden un procesador por transformada en ventanas variables que comprende una función de ventana con tamaño de ventana variable y una función de transformada con una longitud de transformada variable y
- donde el analizador de señales (300/525) está configurado para controlar, sobre la base del análisis de señales, el tamaño de ventana y/o la longitud de las transformadas.
4. Un codificador de audio de acuerdo con una de las reivindicaciones precedentes, en el cual la segunda rama codificadora comprende una primera rama de procesamiento (522) para procesar una señal de audio en el dominio determinada por el convertidor de dominio (510), y una segunda rama de procesamiento (523, 524) que comprende el segundo convertidor (523),
- donde el analizador de señales está configurado para subdividir la porción de la señal de audio en una secuencia de subporciones y donde el analizador de señales está configurado para determinar la resolución en tiempo/frecuencia del segundo convertidor (523) dependiendo de la posición de la subporción procesada por la primera rama de procesamiento con respecto a una subporción de la porción procesada por la segunda rama de procesamiento.
5. Un codificador de audio de acuerdo con la reivindicación 4, en el cual la primera rama de procesamiento comprende un codificador ACELP (526),
- en el cual la segunda rama de procesamiento comprende un dispositivo de procesamiento MDCT-TCX (527),
- en el cual el analizador de señales (300/525) está configurado para establecer la resolución en tiempo del segundo convertidor en un valor elevado determinado por la longitud de una subporción o un valor comparativamente menor determinado por la longitud de una subporción multiplicada por un valor entero superior a uno.

6. Un codificador de audio de acuerdo con una de las reivindicaciones precedentes, en el cual el analizador de señales (300, 525) está configurado para determinar una clasificación de señal en un mapa de bits constante que cubre una pluralidad de bloques de igual tamaño de muestras de audio, y para subdividir un bloque en un número variable de bloques dependiendo de la señal de audio, donde una longitud del subbloque determina la primera resolución en tiempo/frecuencia o la segunda resolución en tiempo/frecuencia.
7. Un aparato de acuerdo con una de las reivindicaciones anteriores, en el cual el analizador de señales (300, 525) está configurado para determinar la resolución en tiempo/frecuencia que se ha de seleccionar de una pluralidad de longitudes de ventanas diferentes, donde la pluralidad de longitudes de ventana diferentes es de por lo menos dos de 2304, 2048, 256, 1920, 2160, 240 muestras, o
- que utiliza la pluralidad de diferentes longitudes de transformada, donde las diferentes longitudes de transformada comprenden por lo menos dos del grupo que consiste en 1152, 1024, 1080, 960, 128, 120 coeficientes por bloque de transformada o
- en el cual el analizador de señales (300, 525) está configurado para determinar la resolución en tiempo/frecuencia del segundo convertidor en términos de una de una pluralidad de longitudes de ventanas diferentes, donde la pluralidad de longitudes de ventana diferentes es de por lo menos dos de 640, 1152, 2304, 512, 1024 o 2048 muestras, o
- que utiliza una pluralidad de longitudes de transformada diferentes, donde las diferentes longitudes de transformadas comprenden por lo menos dos del grupo que consiste en 320, 576, 1152, 256, 512, 1024 coeficientes espectrales por bloque de transformadas.
8. Un codificador de audio de acuerdo con una de las reivindicaciones precedentes, en el cual la segunda rama codificadora comprende:
- una primera rama de procesamiento (522) para procesar una señal de audio;
- una segunda rama de procesamiento, donde la segunda rama de procesamiento comprende el segundo convertidor; y

otro conmutador (521) para conmutar entre primera rama de procesamiento (522) y la segunda rama de procesamiento (523, 524) de manera que, para una porción de la señal de audio ingresada en la segunda rama codificadora, una primera señal procesada o una segunda señal procesada esté en la
5 segunda señal codificada.

9. Un método de codificación de audio de una señal de audio, que comprende:

la codificación, en una primera rama codificadora (400), de una señal de audio
10 utilizando un primer algoritmo de codificación para obtener una primera señal codificada, donde la primera rama codificadora comprende el primer convertidor (410) para convertir una señal de entrada en un dominio espectral;

la codificación, en una segunda rama codificadora (500), de una señal de
15 audio utilizando un segundo algoritmo de codificación para obtener una segunda señal codificada, donde el primer algoritmo de codificación es diferente del segundo algoritmo de codificación, donde la segunda rama codificadora comprende un convertidor de dominio para convertir una señal de entrada de un dominio de entrada en un dominio de salida, y un segundo
20 convertidor (523) para convertir una señal de entrada en un dominio espectral;

conmutar (200) entre la primera rama codificadora y la segunda rama
codificadora de manera que, para una porción de la señal de audio de entrada, la primera señal codificada o la segunda señal codificada esté en una señal de
25 salida del codificador;

analizar (300, 525) la porción de la señal de audio para determinar si la
porción de la señal de audio está representada como primera señal codificada
o segunda señal codificada en la señal de salida del codificador,
30

determinar en forma variable una resolución respectiva en tiempo/frecuencia
del primer convertidor y el segundo convertidor, cuando se genera la primera
señal codificada o la segunda señal codificada que representa la porción de la
señal de audio y
35

generar (800) una señal de salida del codificador que comprende la primera
señal codificada y la segunda señal codificada y una información que indica la
primera señal codificada y la segunda señal codificada, y una información que
indica la resolución en tiempo/frecuencia aplicada para codificar la primera
40 señal codificada y para codificar la segunda señal codificada.

10. Un decodificador de audio para decodificar una señal codificada, donde la señal codificada comprende una primera señal codificada, una segunda señal codificada, una indicación que indica la primera señal codificada y la segunda señal codificada, y una información de resolución en tiempo/frecuencia a utilizar para decodificar la primera señal codificada y la segunda señal de audio codificada, que comprende:
- una primera rama decodificadora (431, 440) para decodificar la primera señal codificada utilizando un primer convertidor controlable de frecuencia/tiempo (440), donde el primer convertidor controlable de frecuencia/tiempo está configurado para ser controlado utilizando la información de resolución en tiempo/frecuencia correspondiente a la primera señal codificada para obtener una primera señal decodificada;
- una segunda rama decodificadora para decodificar la segunda señal codificada utilizando un segundo convertidor controlable de frecuencia/tiempo (534), donde el segundo convertidor controlable de frecuencia/tiempo (534) está configurado para ser controlado utilizando la información de resolución en tiempo/frecuencia correspondiente a la segunda señal codificada;
- un controlador (990) para controlar el primer convertidor de frecuencia/tiempo (440) y el segundo convertidor de frecuencia/tiempo (534) utilizando la información de resolución en tiempo/frecuencia;
- un convertidor de dominio (540) para generar una señal de síntesis utilizando la segunda señal decodificada y
- un combinador (604) para combinar la primera señal decodificada y la señal de síntesis a fin de obtener una señal de audio decodificada.
11. Un decodificador de audio de acuerdo con la reivindicación 10, en el cual el controlador (990) está configurado para controlar el primer convertidor de frecuencia/tiempo (440) y el segundo convertidor de frecuencia/tiempo (534) de manera que,
- en el caso del primer convertidor de frecuencia/tiempo (440) la resolución en tiempo/frecuencia sea seleccionada de una pluralidad de longitudes de ventanas diferentes, donde las longitudes de ventanas diferentes son por lo menos dos de 2304, 2048, 256, 1920, 2160, 240 muestras, o

- 5 sea seleccionada de una pluralidad de longitudes de transformada diferentes, donde las longitudes de transformada diferentes comprenden por lo menos dos del grupo que consiste en 1152, 1024, 1080, 960, 128, 120 coeficientes por bloque de transformadas o
- 10 en el caso del segundo convertidor de frecuencia/tiempo (534) la resolución en tiempo/frecuencia es seleccionada como una de una pluralidad de longitudes de ventanas diferentes, donde la pluralidad de longitudes de ventanas diferentes son por lo menos dos de 640, 1152, 2304, 512, 1024 o 2048 muestras, o
- 15 sea seleccionada de una pluralidad de longitudes de transformada diferentes, donde las longitudes de transformada diferentes comprenden por lo menos dos del grupo que consiste en 320, 576, 1152, 256, 512, 1024 coeficientes espectrales por bloque de transformadas.
12. Un decodificador de audio de acuerdo con la reivindicación 10 o 11, en el cual la segunda rama decodificadora comprende una primera rama de procesamiento inverso (531) para el procesamiento inverso de una primera señal procesada que está incluida además en la señal codificada para obtener una primera señal procesada inversa;
- 20 donde el segundo convertidor controlable de frecuencia tiempo (534) está situado en una segunda rama de procesamiento inverso configurada para el procesamiento inverso de la segunda señal codificada en un dominio idéntico al dominio de la primera señal procesada inversa para obtener una segunda señal procesada inversa;
- 25 otro combinador (532) para combinar la primera señal procesada inversa y la segunda señal procesada inversa para obtener una señal combinada y
- 30 donde la señal combinada es enviada al combinador (600).
- 35 13. Un decodificador de audio de acuerdo con cualquiera de las reivindicaciones 10 - 12, en el cual el primer convertidor frecuencia/tiempo (440) y el segundo convertidor frecuencia/tiempo con convertidores de cancelación de solapamiento en el dominio del tiempo que constan de una unidad de traslape/suma (440c) para cancelar un solapamiento en el dominio del tiempo
- 40 incluido en la primera señal codificada y la segunda señal codificada.

14. Un decodificador de audio de acuerdo con cualquiera de las reivindicaciones 10 – 13, en el cual la señal codificada comprende información de modo de codificación que identifica si la señal codificada es la primera señal codificada y la segunda señal codificada y
- 5 donde el decodificador comprende además una interfaz de entrada (900) para interpretar la información de modo de codificación para determinar si se debe alimentar la señal codificada a la primera rama decodificadora o a la segunda rama decodificadora.
- 10 15. Un decodificador de audio de acuerdo con una de las reivindicaciones anteriores, en el cual la primera señal codificada es codificada aritméticamente y donde la primera rama codificadora comprende un decodificador aritmético.
- 15 16. Un decodificador de audio de acuerdo con una de las reivindicaciones anteriores, en el cual la primera rama codificadora comprende un descuantificador que tiene una característica de descuantificación no uniforme para cancelar un resultado de una cuantificación no uniforme aplicada al generar la primera señal codificada,
- 20 donde la segunda rama codificadora comprende un descuantificador que utiliza la característica de descuantificación diferente, donde la segunda rama codificadora no comprende un descuantificador.
- 25 17. Un decodificador de audio de acuerdo con una de las reivindicaciones anteriores, en el cual el controlador (990) está configurado para controlar el primer convertidor frecuencia/tiempo y el segundo convertidor frecuencia/tiempo mediante la aplicación, por cada convertidor, de una resolución discreta de frecuencia/tiempo de un número de resoluciones
- 30 discretas de frecuencia/tiempo diferentes posibles, donde el número de resoluciones de frecuencia/tiempo diferentes posibles es mayor para el segundo convertidor en comparación con el número de resoluciones de frecuencia/tiempo diferentes posibles para el primer convertidor.
- 35 18. Un decodificador de audio de acuerdo con cualquiera de las reivindicaciones 10 - 17, en el cual el convertidor de dominio es un procesador de síntesis LPC (544) que genera la señal de síntesis empleando una información de filtro PC, donde la información de filtro LPC está incluida en la señal codificada.
- 40 19. Un método de decodificación de audio de una señal codificada, donde la señal codificada comprende una primera señal codificada, una segunda señal codificada, una indicación que indica la primera señal codificada y la segunda

señal codificada, e información de resolución en tiempo/frecuencia a utilizar para decodificar la primera señal codificada la segunda señal de audio codificada, que comprende:

5 decodificar, mediante una primera rama decodificadora (431, 440), la primera señal codificada utilizando un primer convertidor controlable frecuencia/tiempo (440), donde el primer convertidor controlable frecuencia/tiempo está configurado para ser controlado utilizando la información de resolución en tiempo/frecuencia correspondiente a la primera señal codificada para obtener
10 una primera señal decodificada;

decodificar, mediante una segunda rama decodificadora, la segunda señal codificada utilizando un segundo convertidor controlable frecuencia/tiempo (534), donde el segundo convertidor controlable frecuencia/tiempo está
15 configurado para ser controlado utilizando la información de resolución en tiempo/frecuencia correspondiente a la segunda señal codificada;

controlar (990) el primer convertidor frecuencia/tiempo (440) y el segundo convertidor frecuencia/tiempo (534) utilizando la información de resolución en
20 tiempo/frecuencia;

generar (540), por medio de un convertidor de dominio, una señal de síntesis utilizando la segunda señal decodificada y

25 combinar (604) la primera señal decodificada y la señal de síntesis para obtener una señal de audio decodificada.

20. Una señal de audio codificada que comprende:

30 una primera señal codificada;

una segunda señal codificada, donde una porción de una señal de audio está representada por la primera señal codificada o por la segunda señal codificada;

35 una indicación que indica la primera señal codificada y la segunda señal codificada;

una indicación de una primera información de resolución en tiempo/frecuencia a utilizar para decodificar la primera señal codificada, y
40

una indicación de una segunda información de resolución en tiempo/frecuencia a utilizar para decodificar la segunda señal codificada.

21. Un programa de computación para ejecutar, cuando se ejecuta un procesador,
5 un método de acuerdo con la reivindicación 9 o la reivindicación 19.

p.p.de: FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER
ANGEWANDTEN FORSCHUNG E.V. * VOICEAGE CORPORATION

10

RESUMEN

Un codificador de audio para codificar una señal de audio comprende una
5 primera rama codificadora (400), donde la primera rama codificadora comprende un
primer convertidor (410) para convertir una señal de un dominio del tiempo a un
dominio de frecuencia. Más aun, el codificador de audio comprende una segunda
rama codificadora (500) que comprende un segundo convertidor de
10 tiempo/frecuencia (523). Además, se presenta un analizador de señales (300/525)
para analizar la señal de audio. El analizador de señales, por un lado, determina si
una porción de audio es eficaz en la señal de salida del codificador como primera
señal codificada procedente de la primera rama codificadora o como segunda señal
codificada de una segunda rama codificadora. Por otro lado, el analizador de señales
determina una resolución en tiempo/frecuencia que han de aplicar los convertidores
15 (410, 523) al generar las señales codificadas. Una interfaz de salida incluye, además
de la primera señal codificada y la segunda señal codificada, una información de
resolución que identifica la resolución empleada por el primer convertidor de
tiempo/frecuencia y que es utilizada por el segundo convertidor de tiempo/frecuencia.

20

25

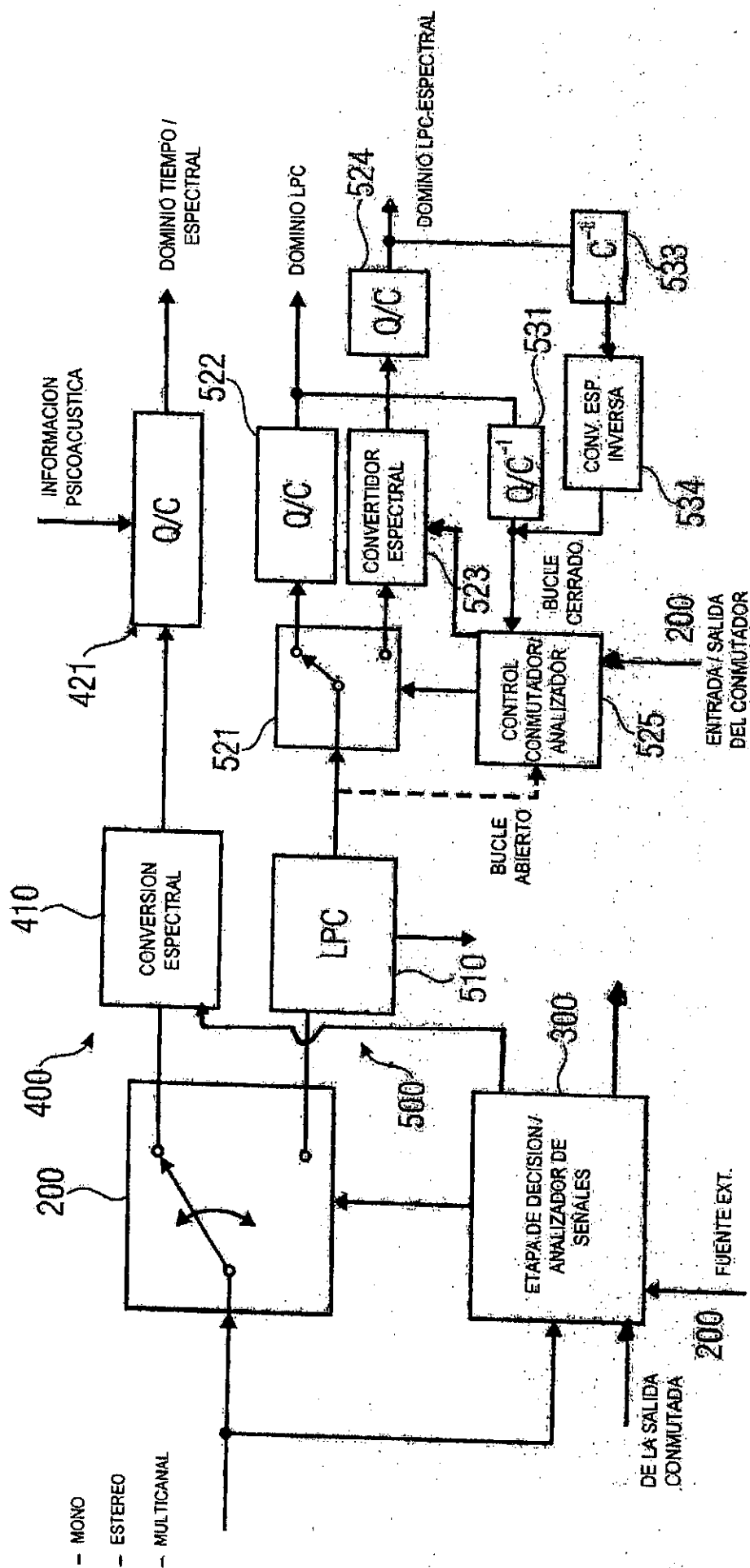


FIG 1A
(CODIFICADOR)

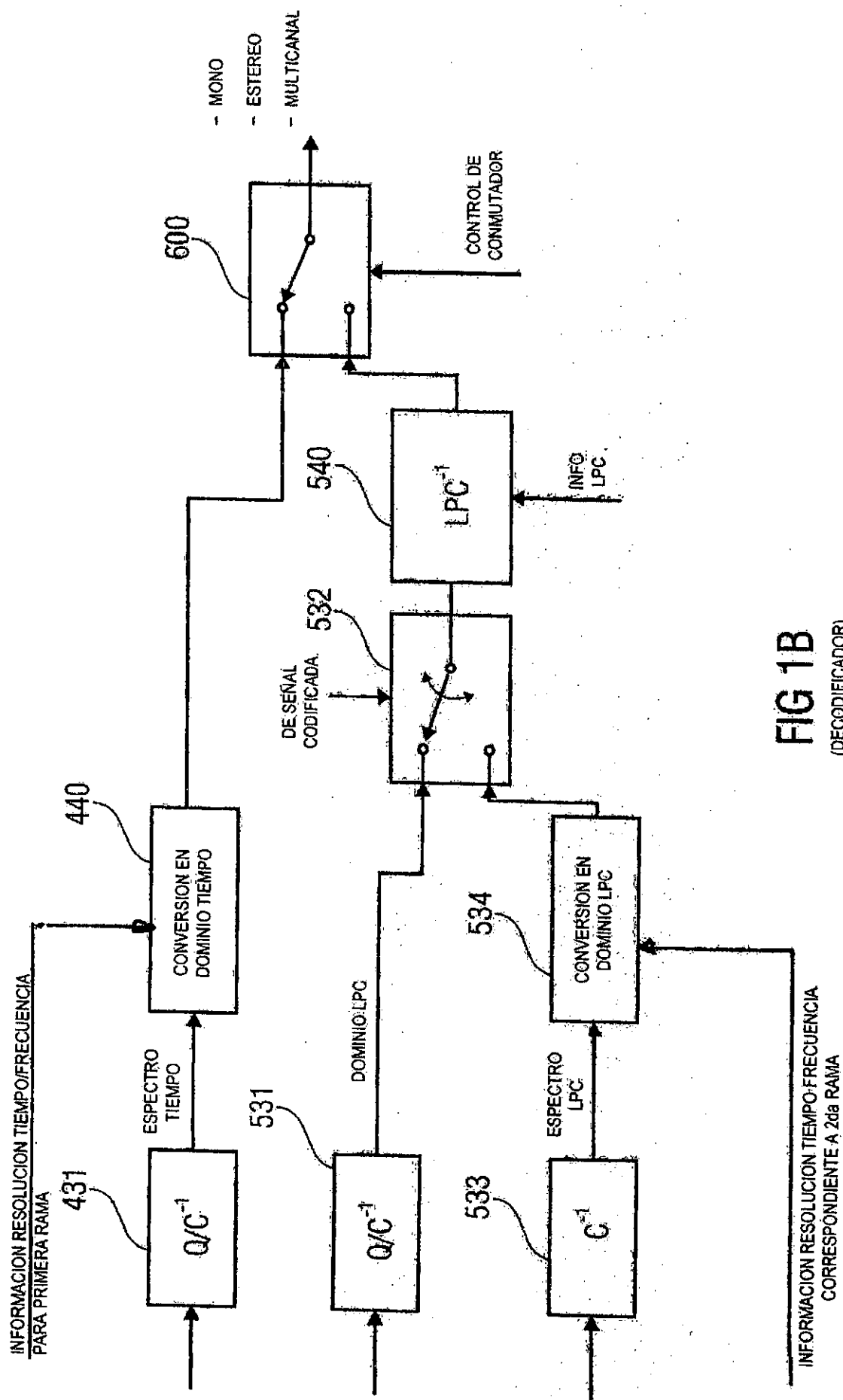


FIG 1B
(DECODIFICADOR)

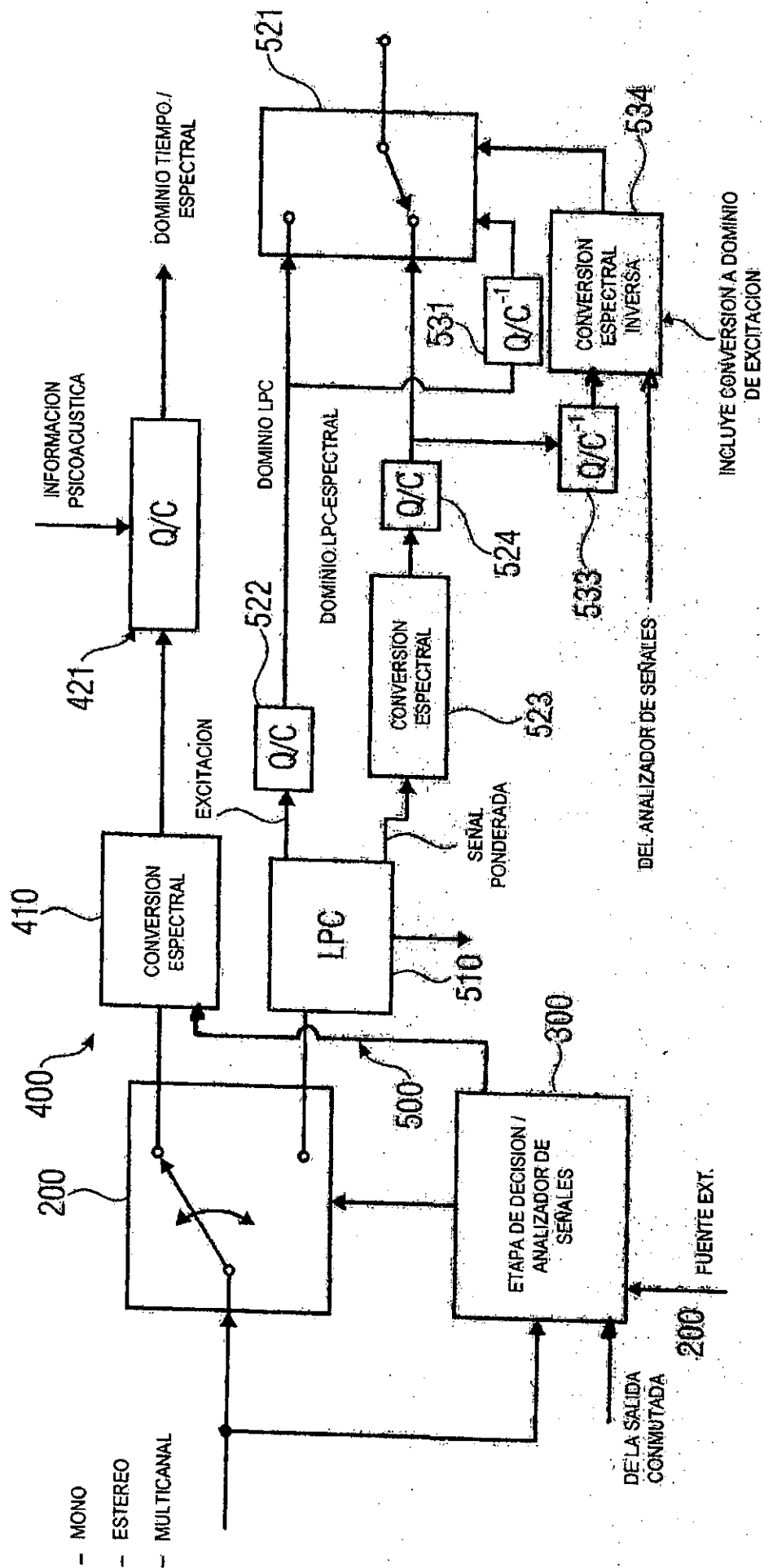


FIG 1C
(CODIFICADOR)

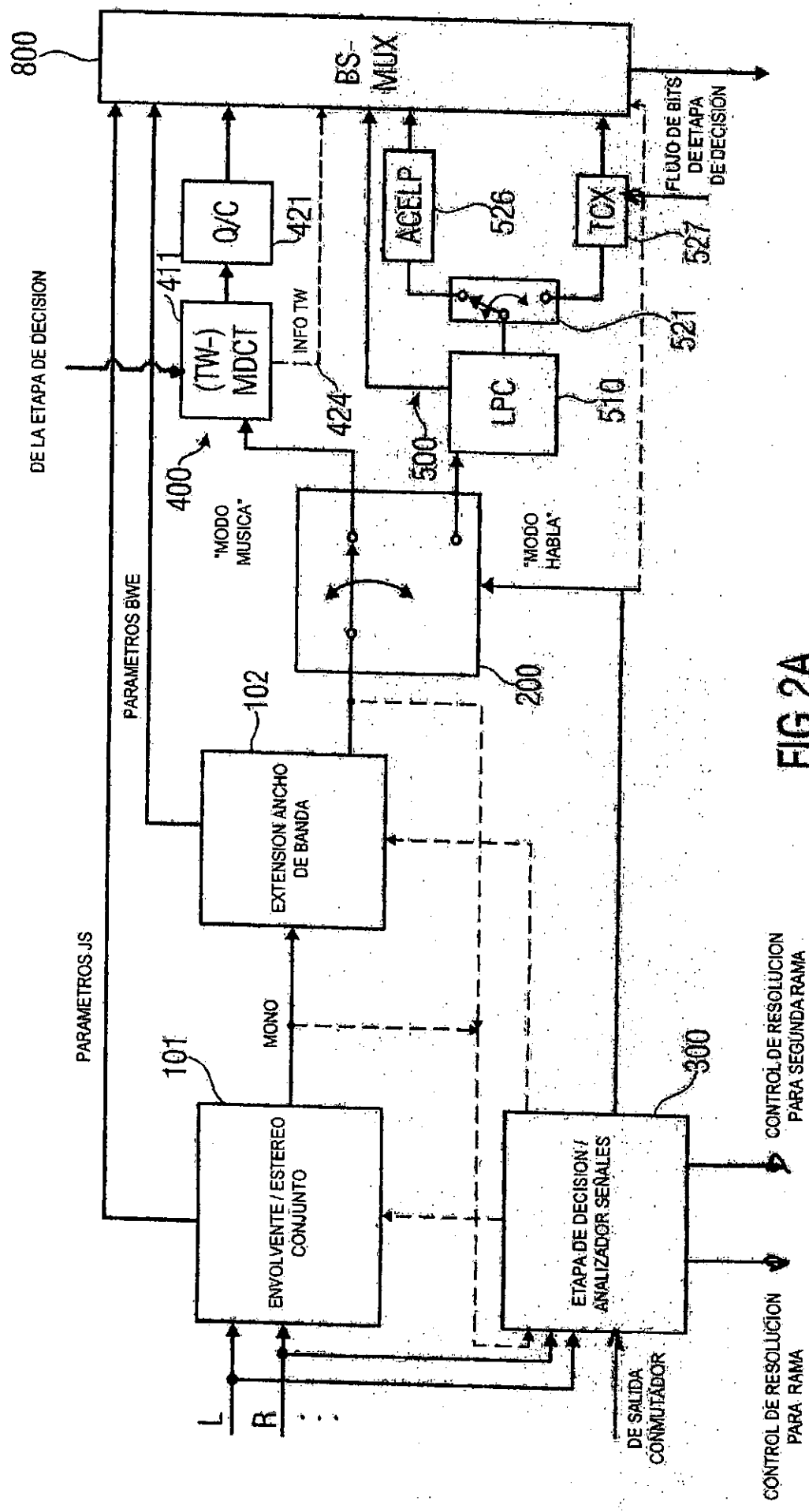


FIG 2A
(CODIFICADOR)

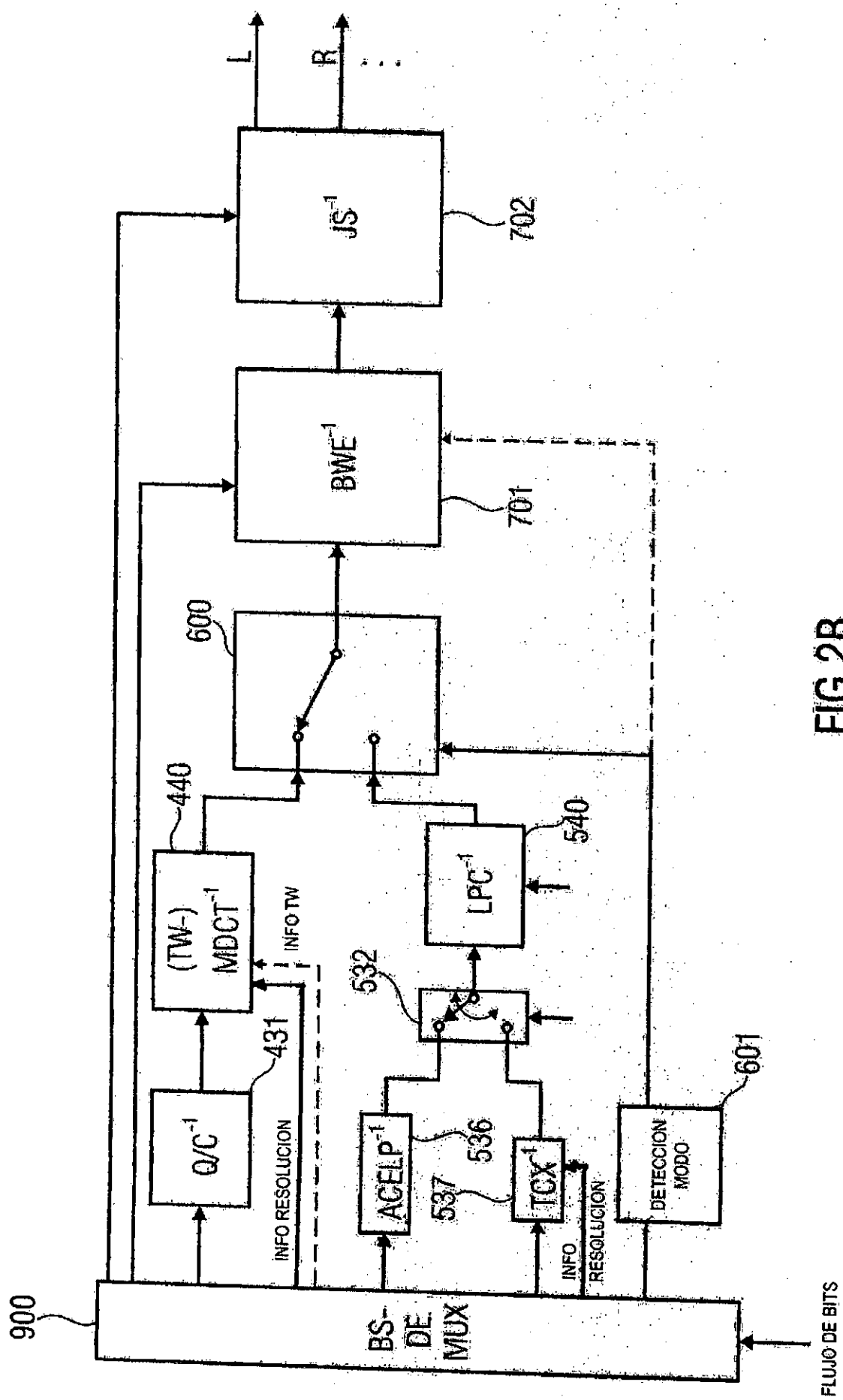


FIG 2B
(DECODIFICADOR)

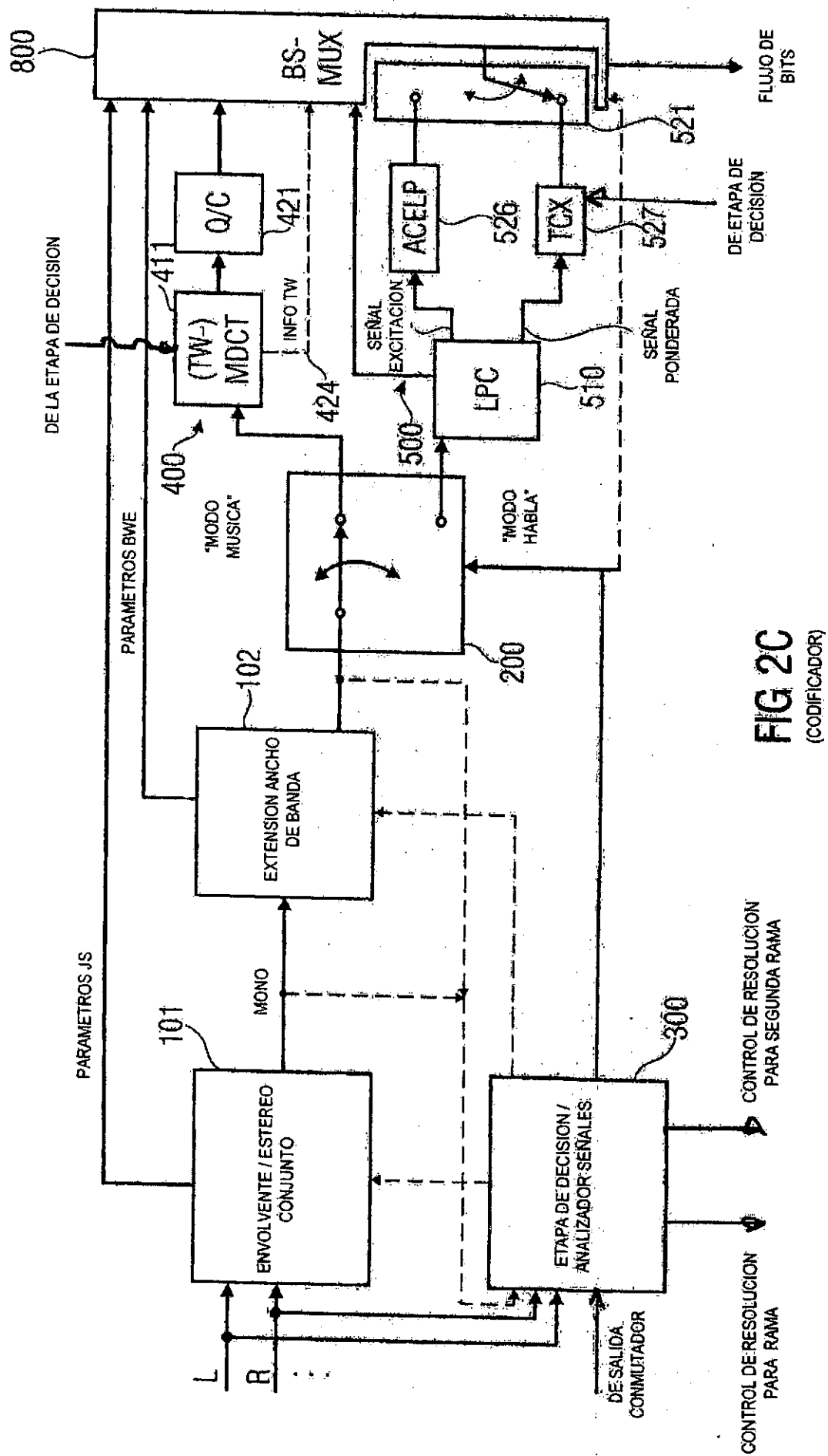


FIG 2C
(CODIFICADOR)

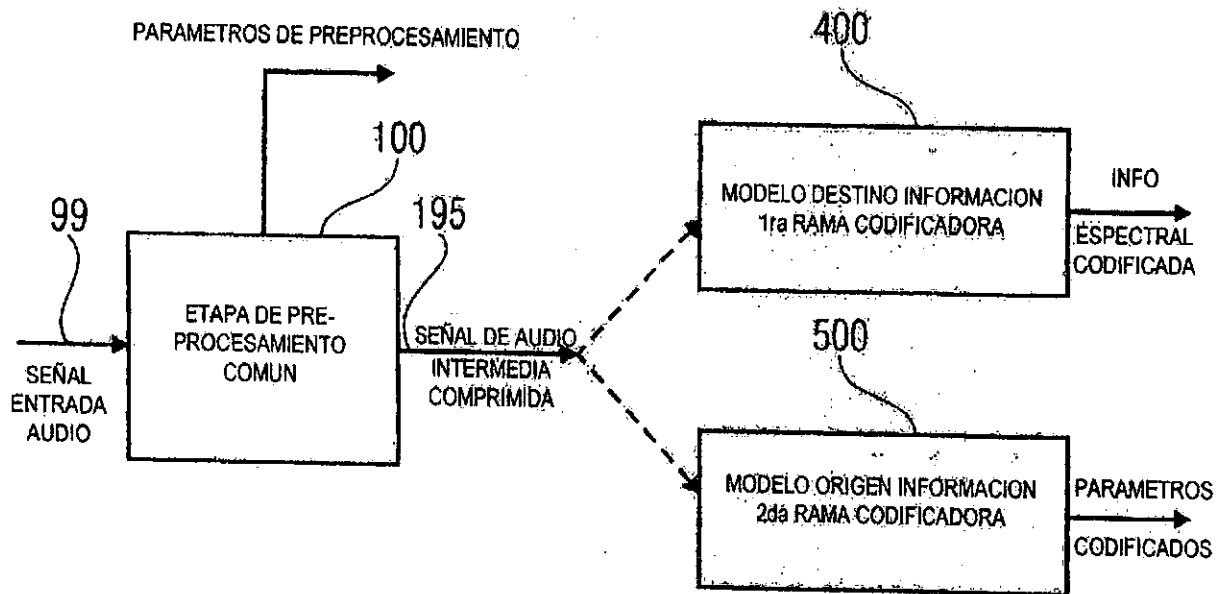


FIG 3A

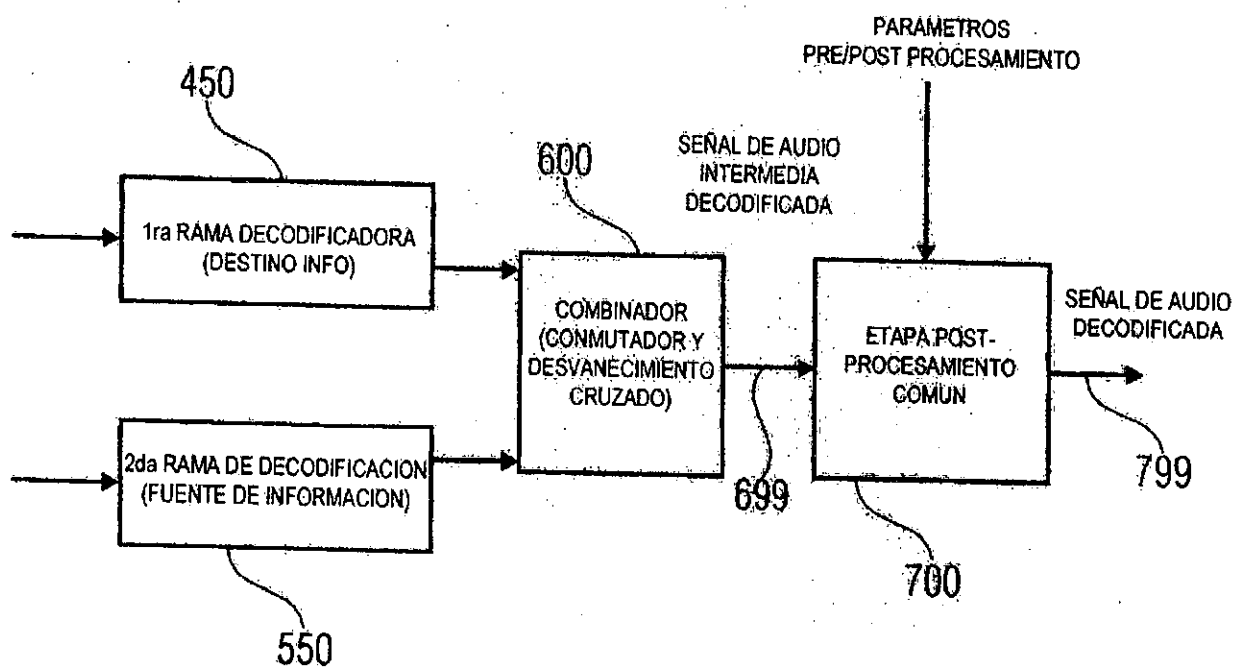


FIG 3B

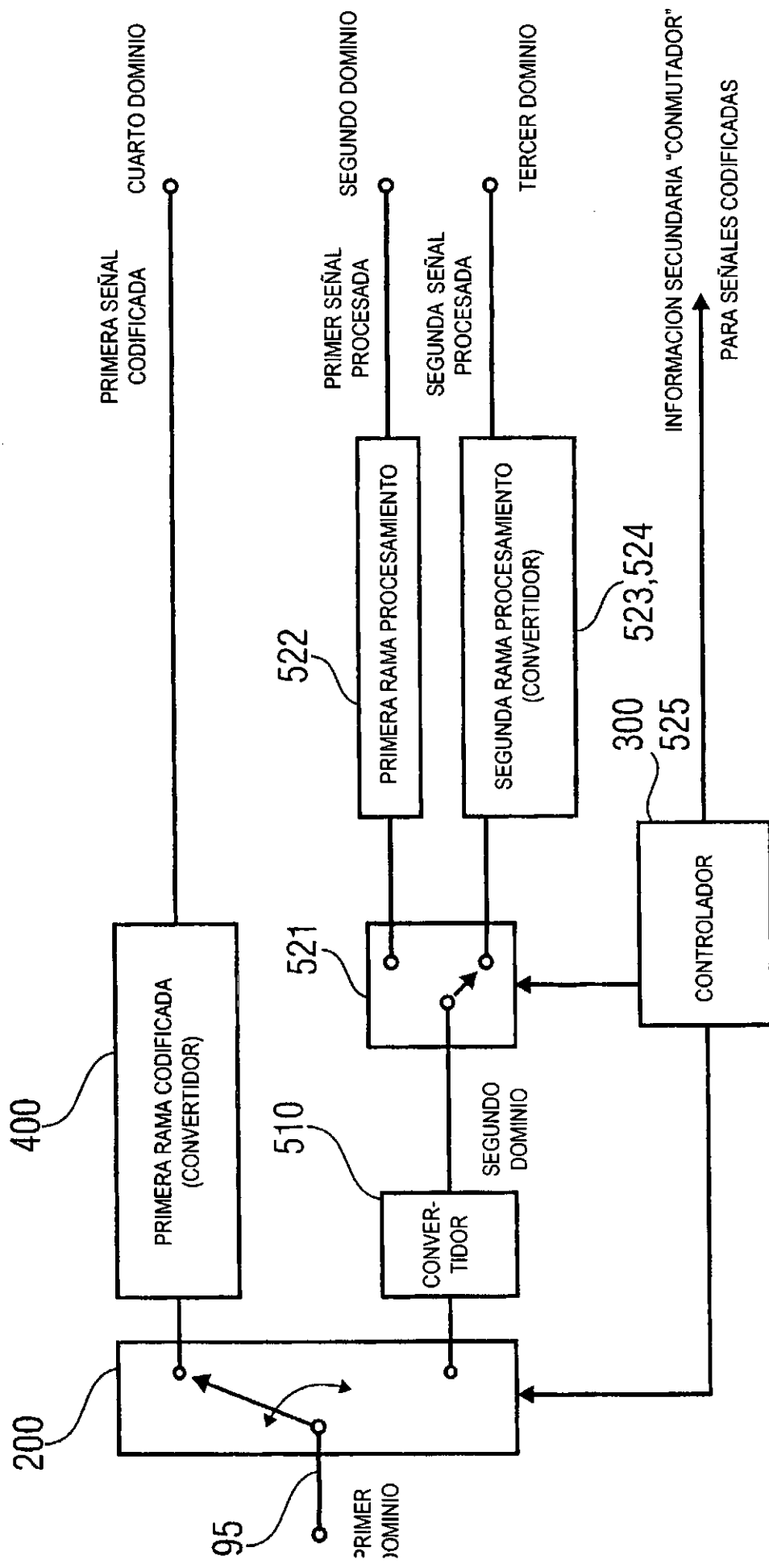
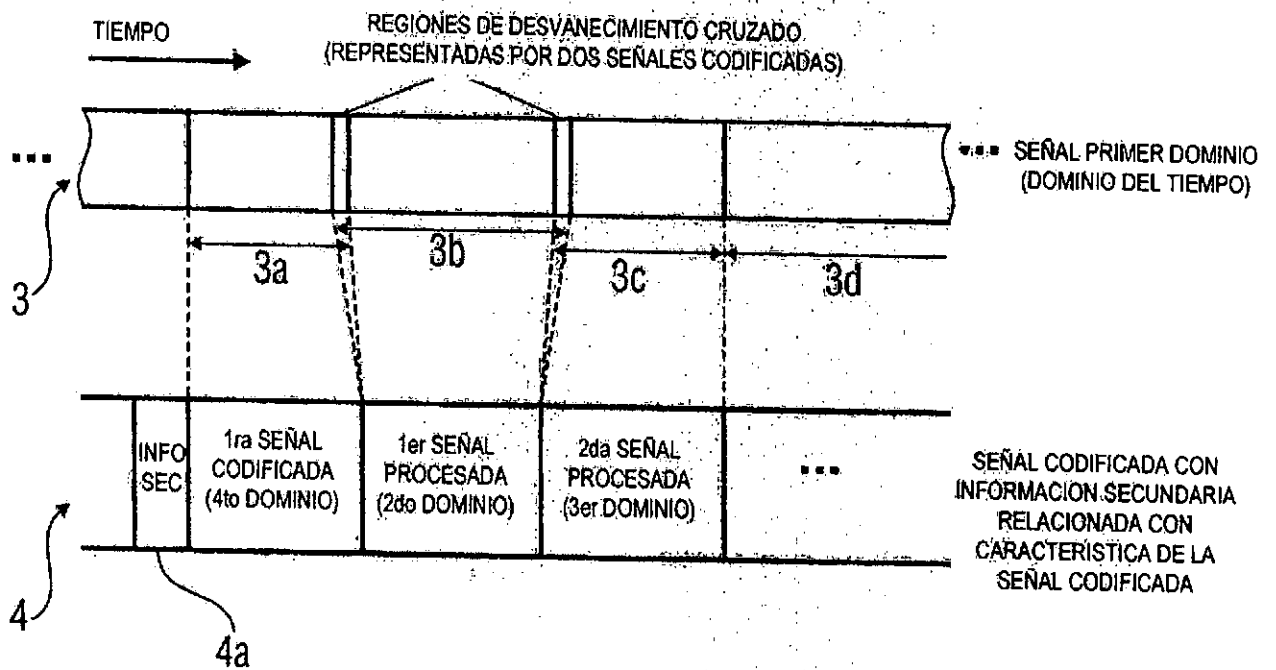
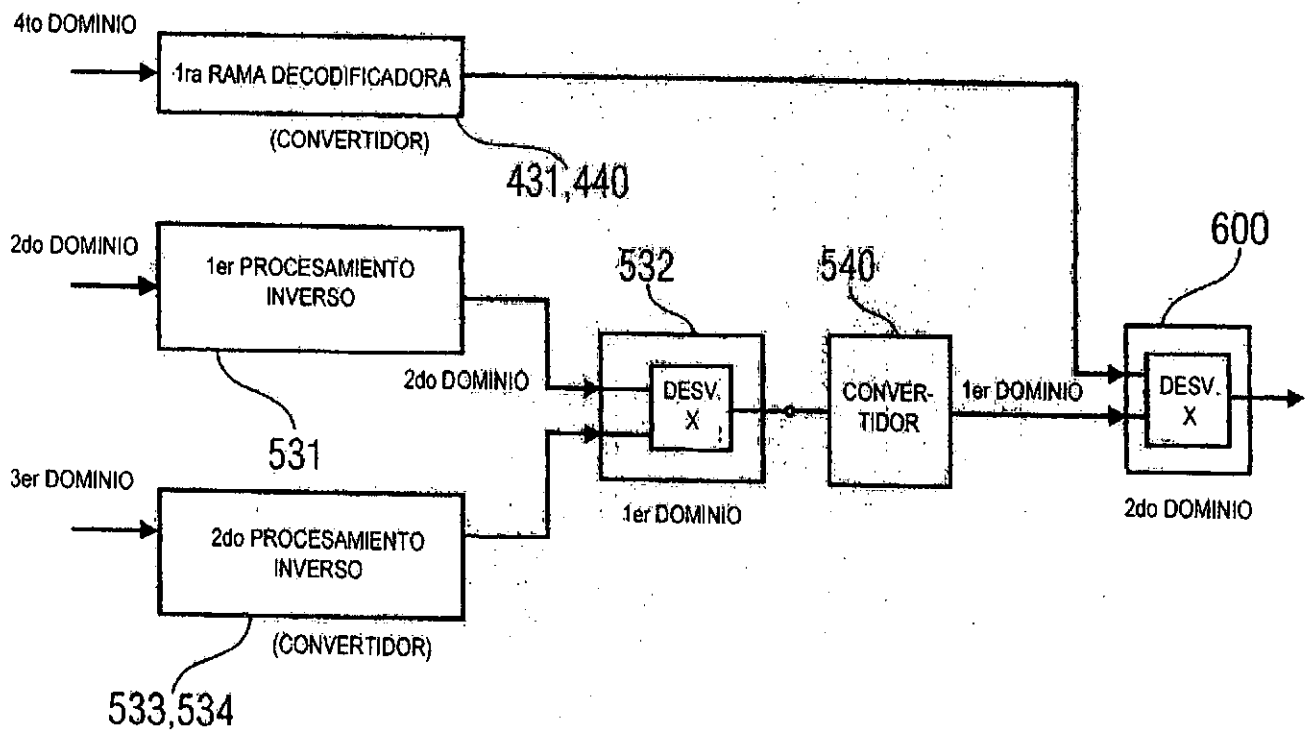


FIG 3C

- CADA BLOQUE DE LA SEÑAL DE AUDIO DEL PRIMER DOMINIO ESTA REPRESENTADA POR UNA SEÑAL CODIFICADA EN EL SEGUNDO DOMINIO, UN TERCER DOMINIO, O UN CUARTO DOMINIO APARTE DE UNA REGION DE CRUCE OPCIONAL



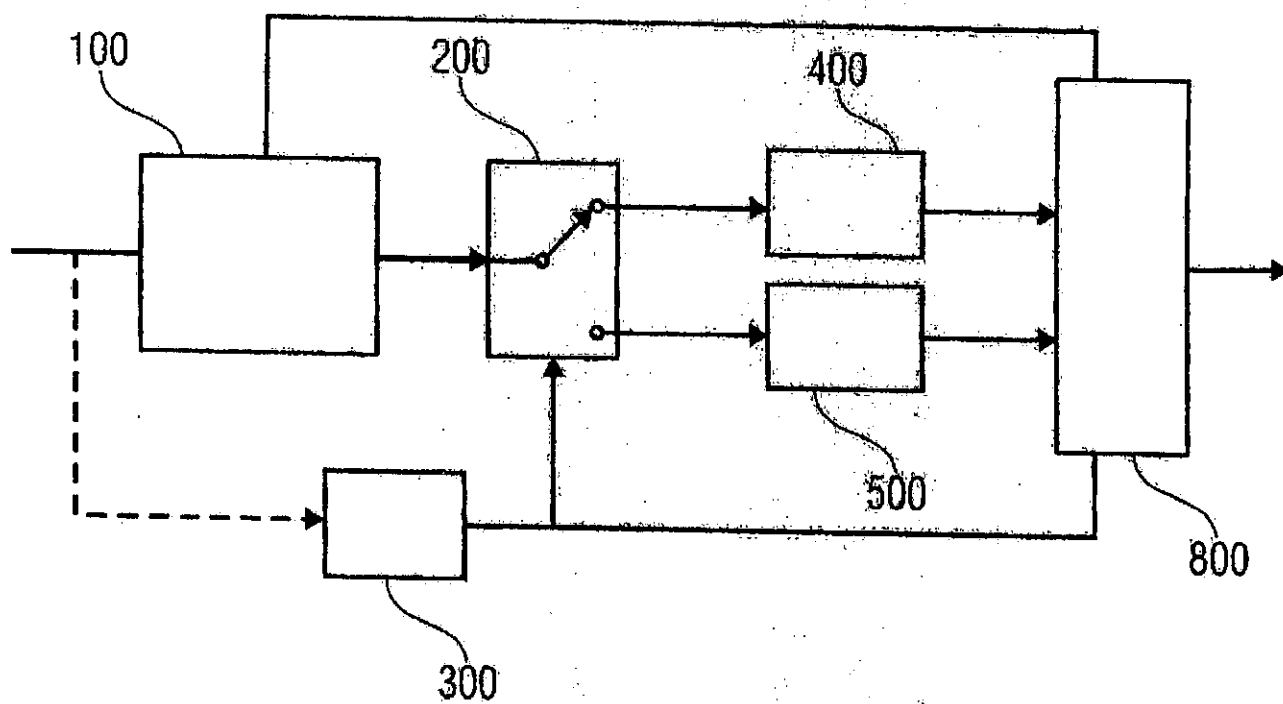


FIG 4A

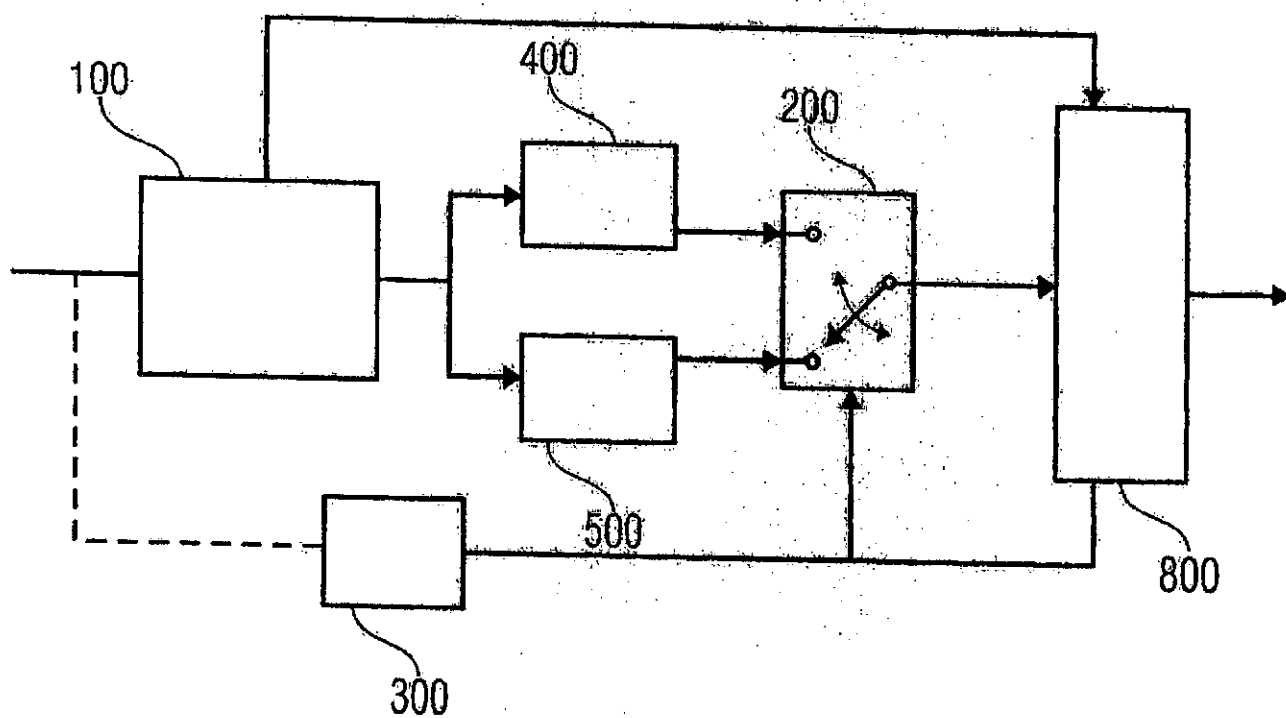


FIG 4B

SEGMENTO SEÑAL TIPO IMPULSOS (POR EJ. HABLA CON VOZ)

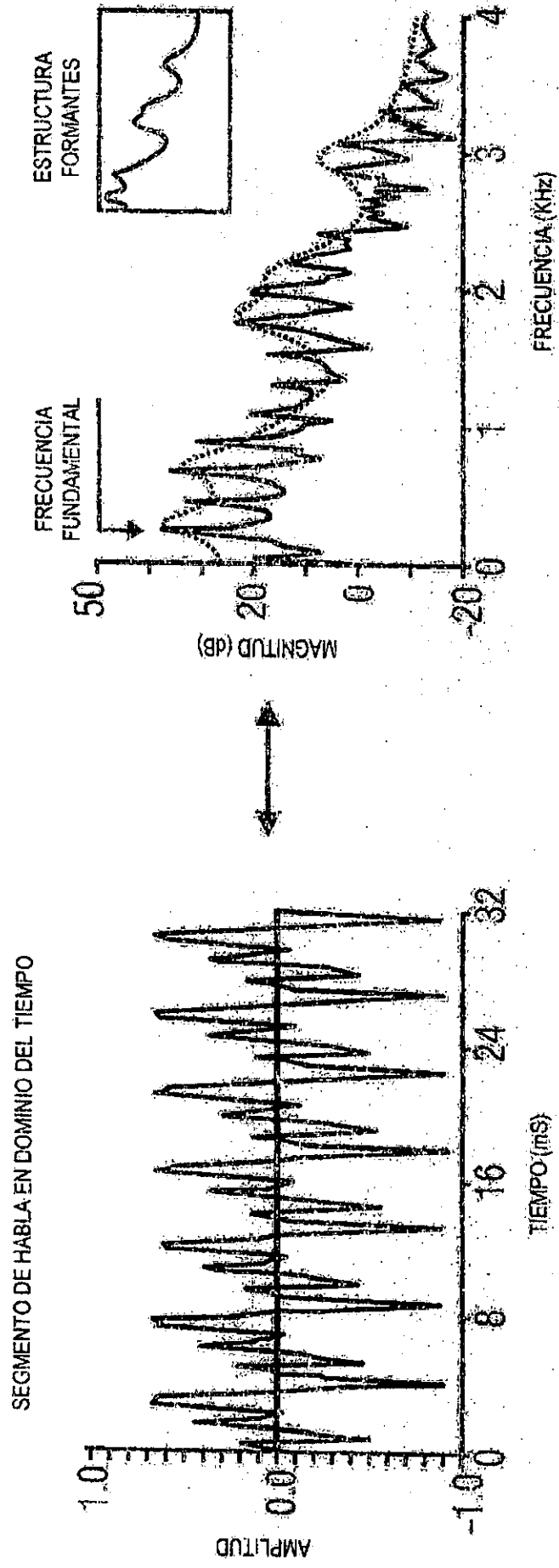


FIG 5A

FIG 5B

SEGMENTO FIJO (POR EJ. HABLA SIN VOZ)

SEGMENTO DE HABLA EN DOMINIO DEL TIEMPO

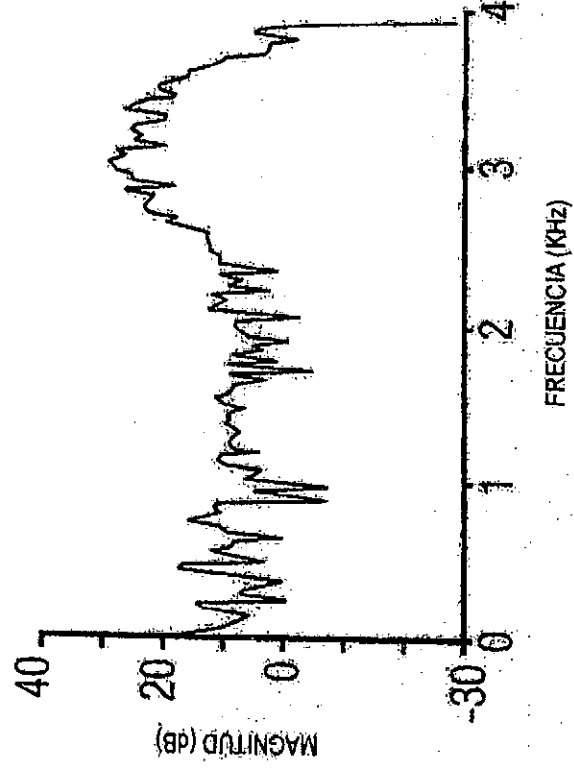
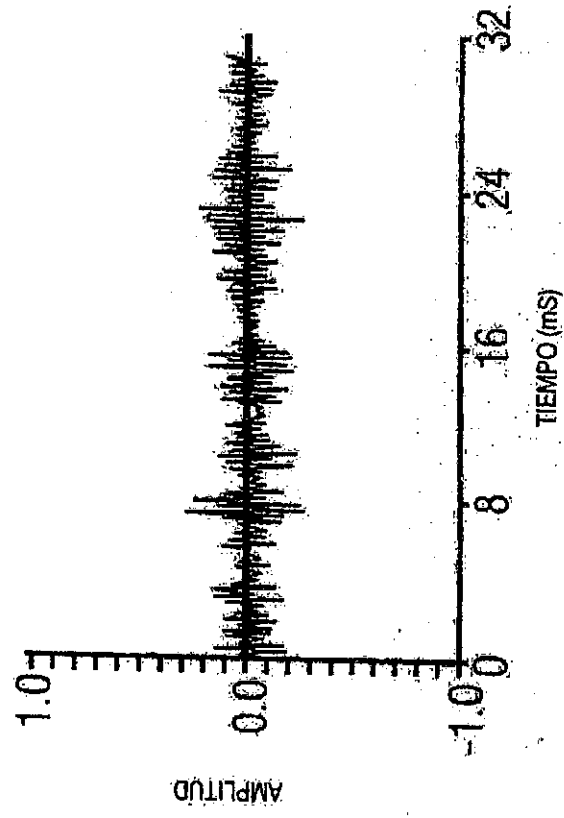
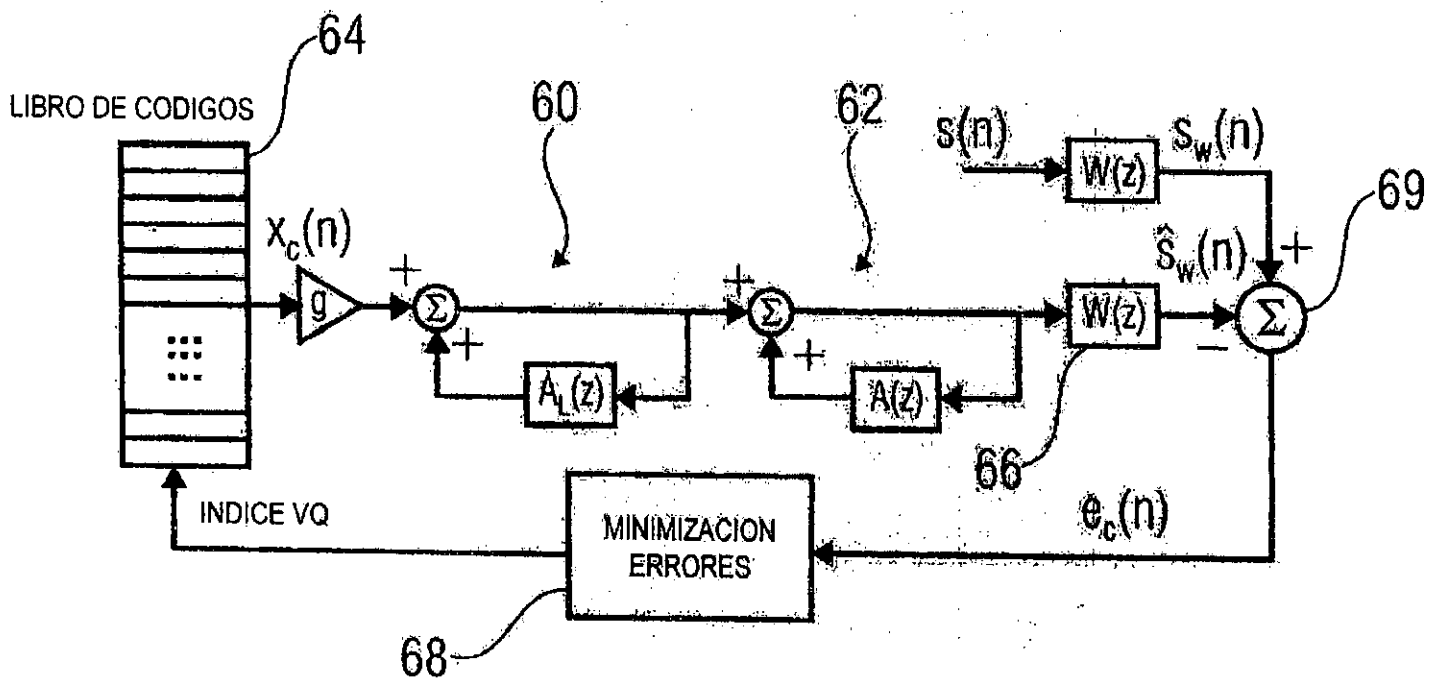


FIG 5C

FIG 5D

CELP ANALISIS POR SINTESIS



$A_L(z)$: PREDICCIÓN A LARGO PLAZO
 $\hat{=}$ TONO ESTRUCTURA (FINA)

$A(z)$: PREDICCIÓN A CORTO PLAZO
 $\hat{=}$ ESTRUCTURA FORMANTES / ENVOLTURA ESPECTRAL

FIG 6

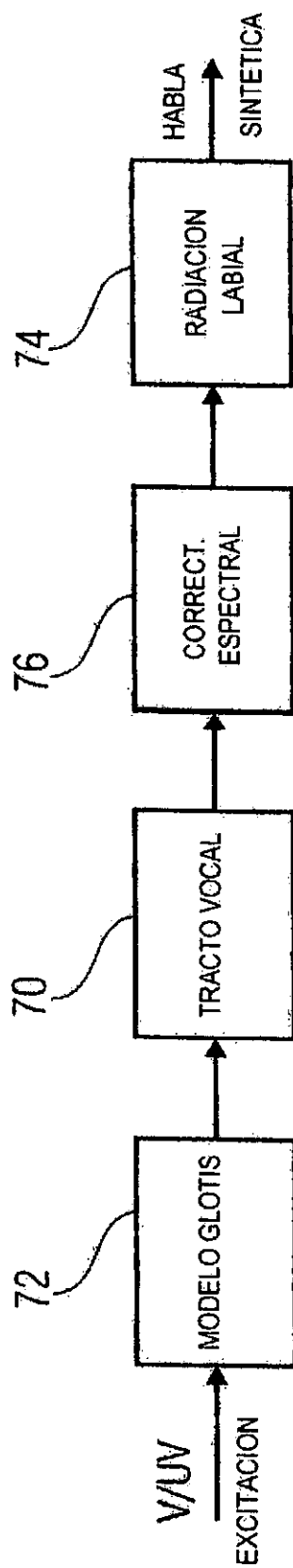


FIG 7A

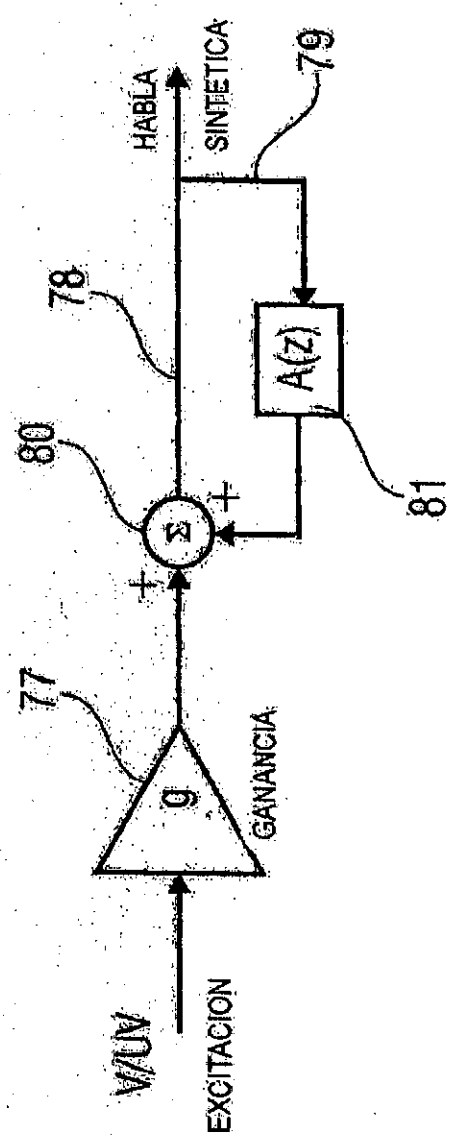


FIG 7B

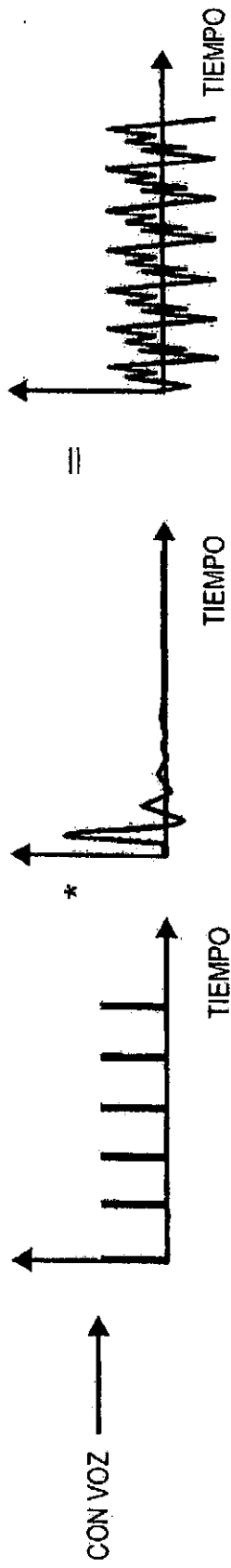


FIG 7C



FIG 7D

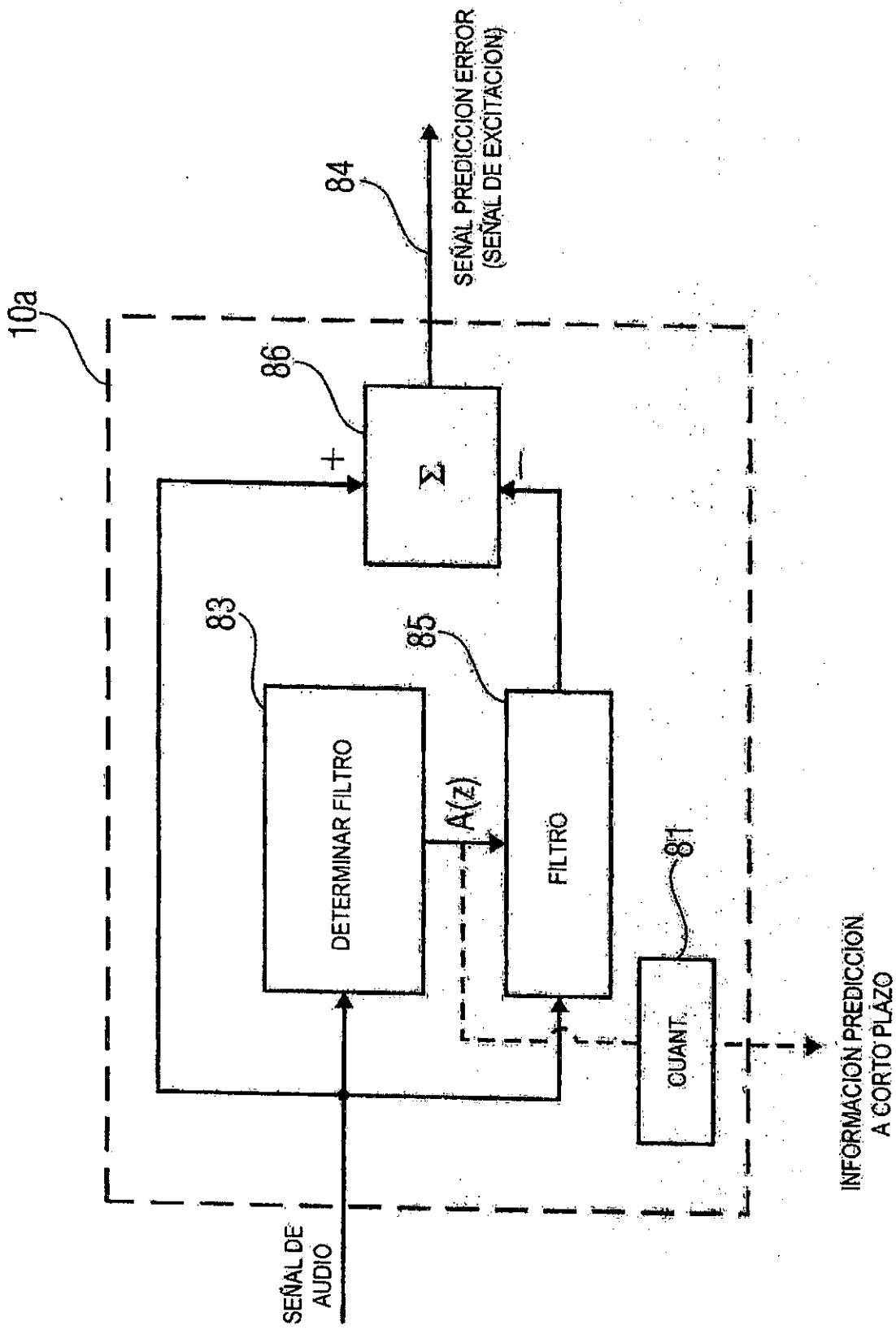


FIG 7E

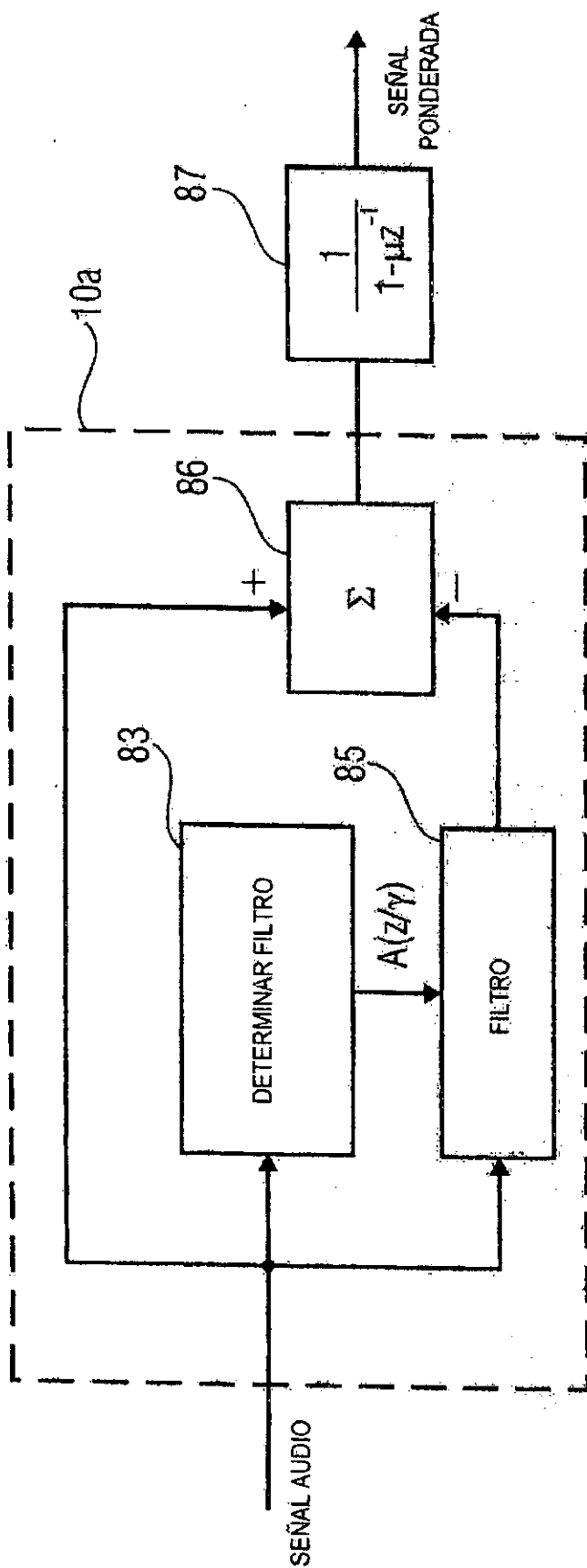


FIG 7F

(LADO CODIFICADOR)

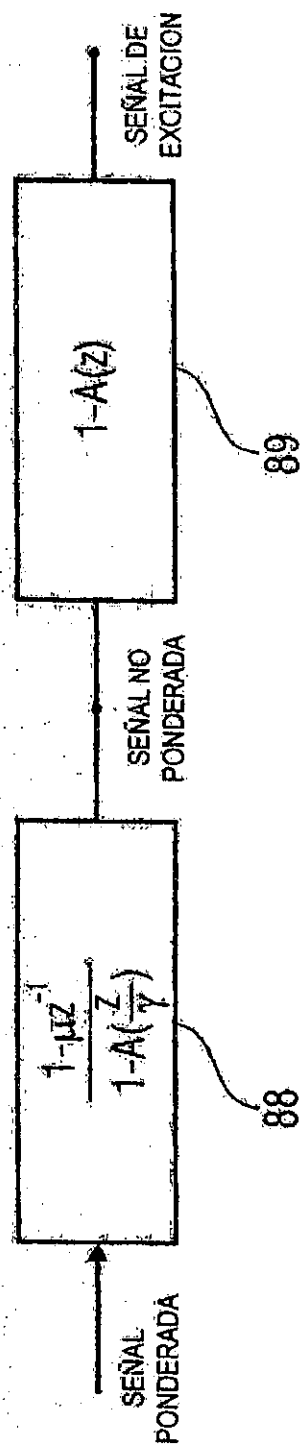


FIG 7G

(LADO DEL DECODIFICADOR)

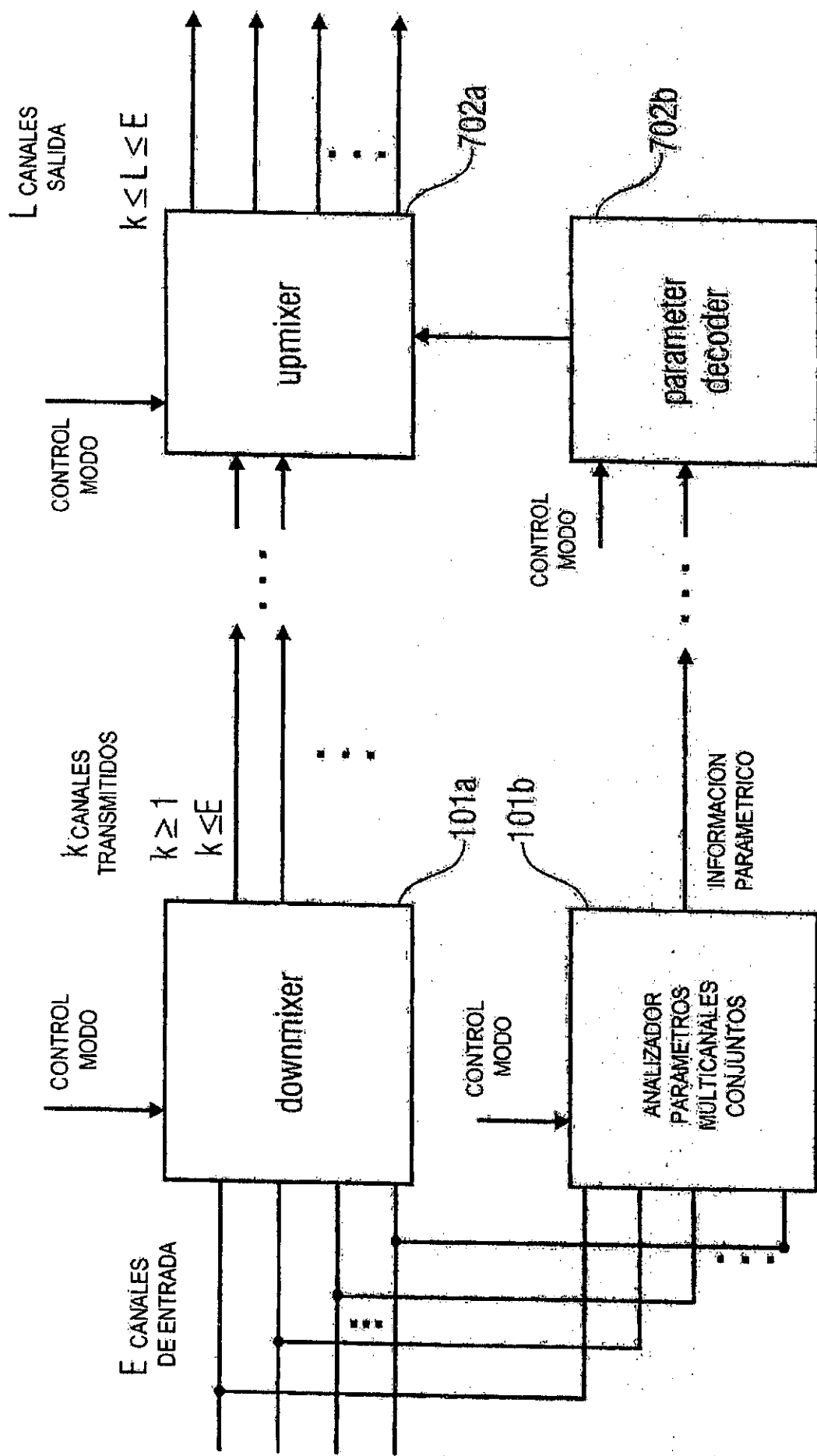


FIG 8

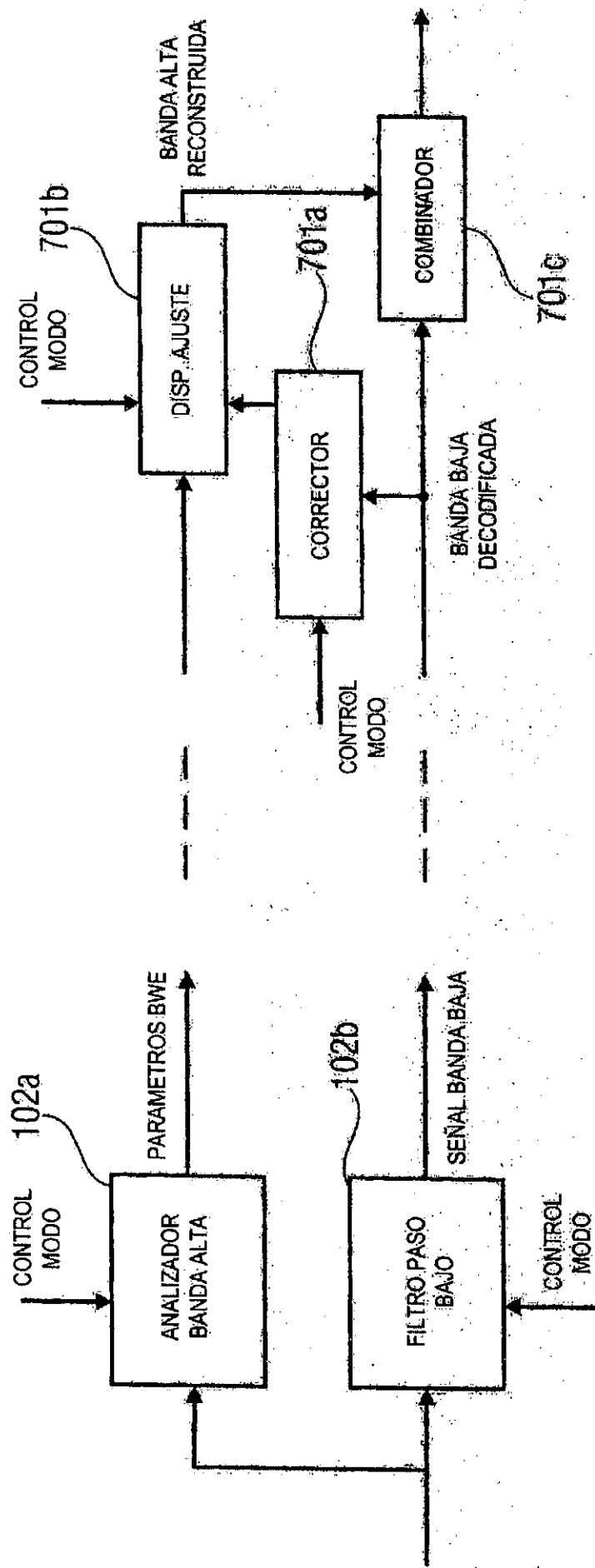
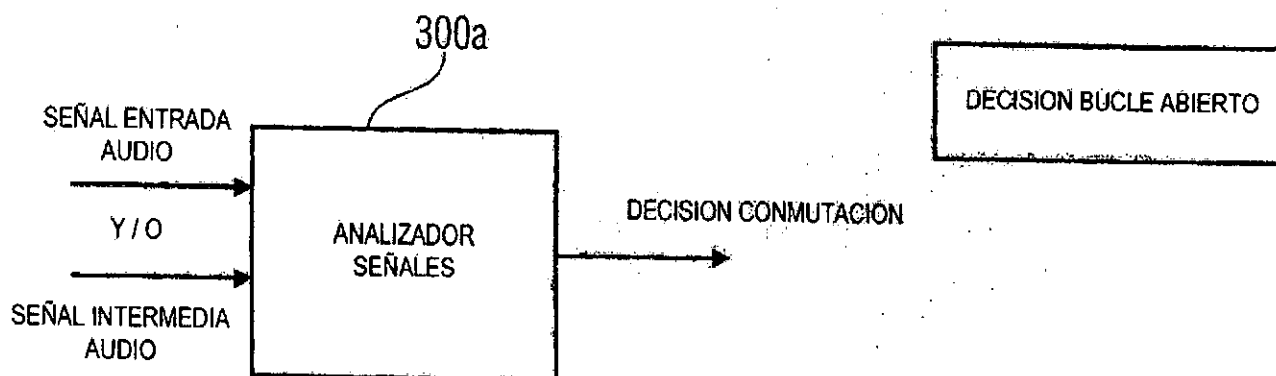


FIG 9



SEÑAL INTERMEDIA AUDIO:

- SEÑAL BANDA BAJA,
- SEÑAL DOWNMIX O
- PORCION BANDA BAJA DE SEÑAL DOWNMIX

FIG 10A

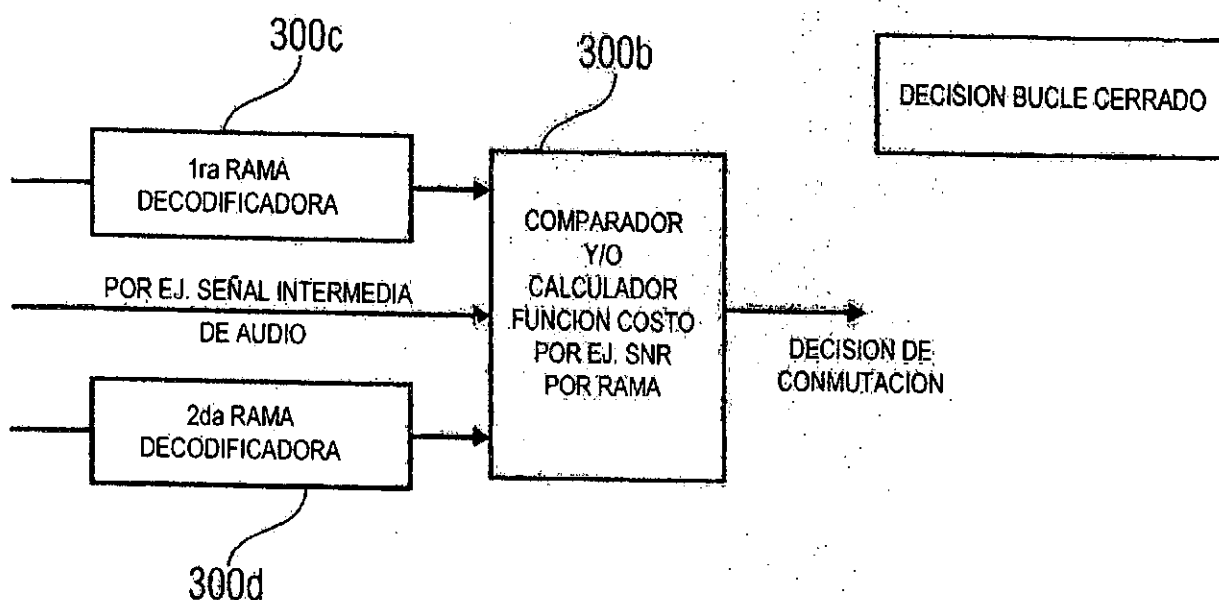


FIG 10B

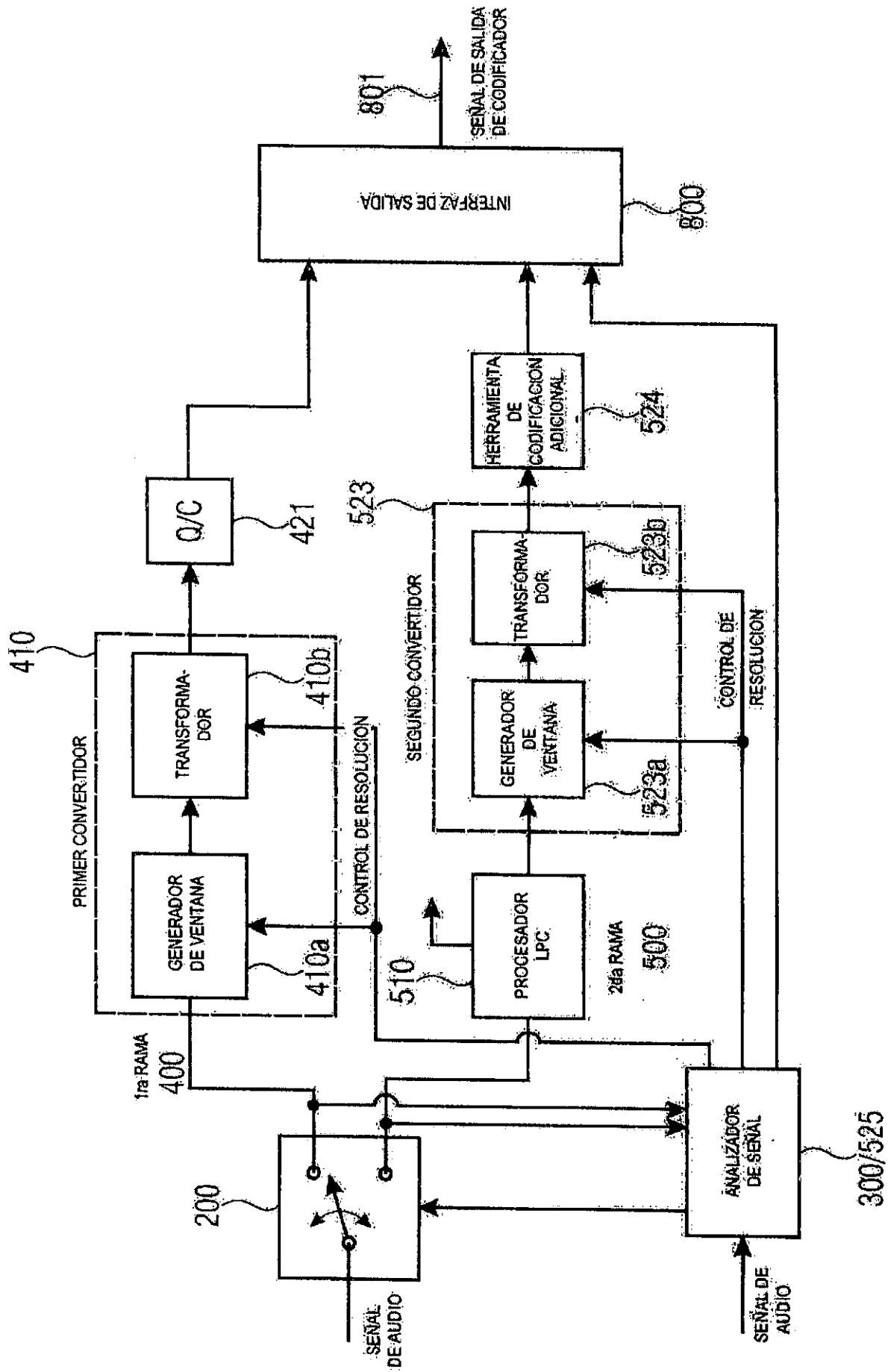


FIG. 11A

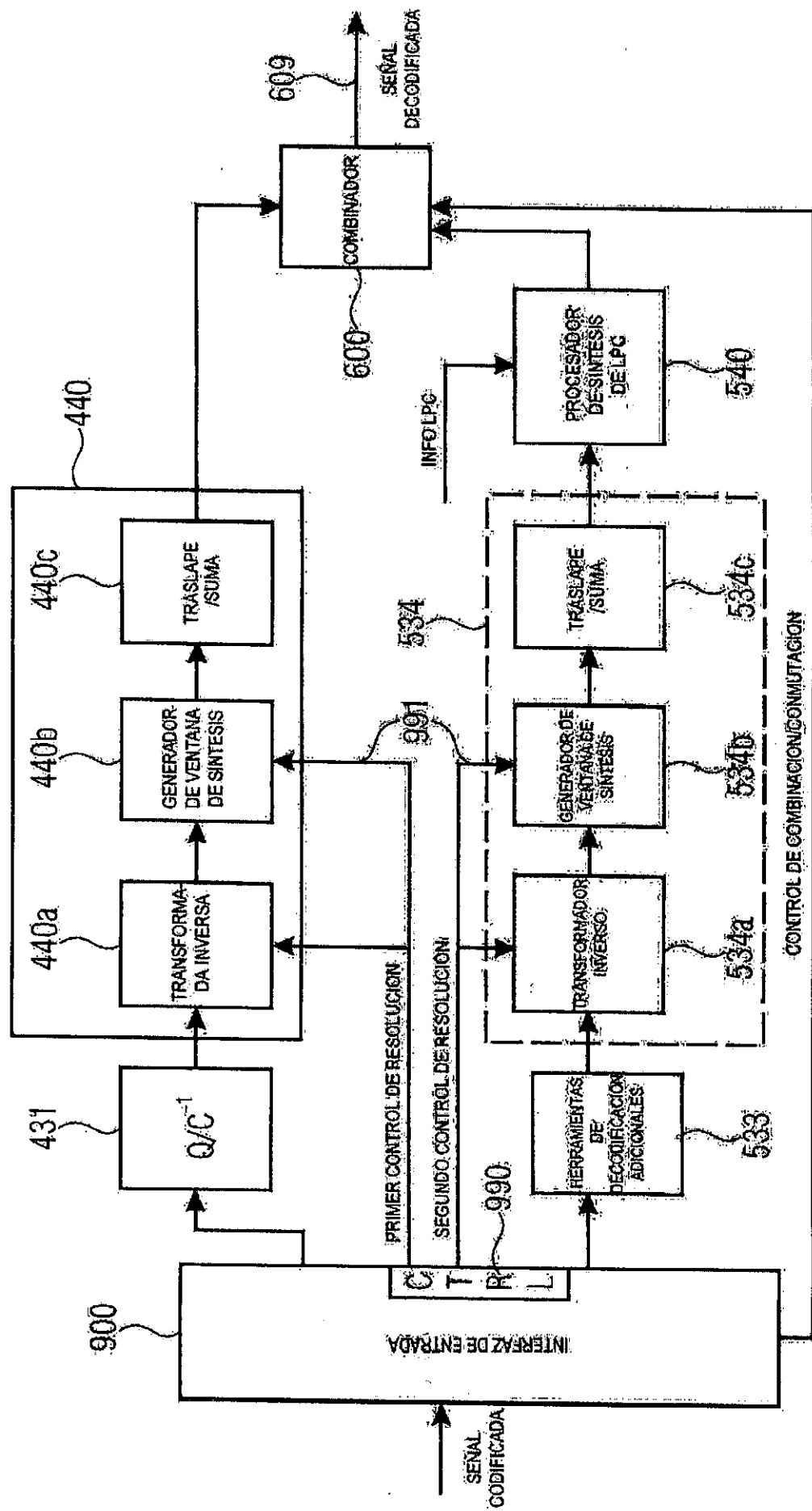


FIG. 11B

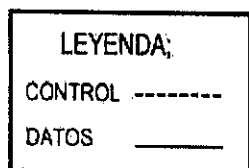
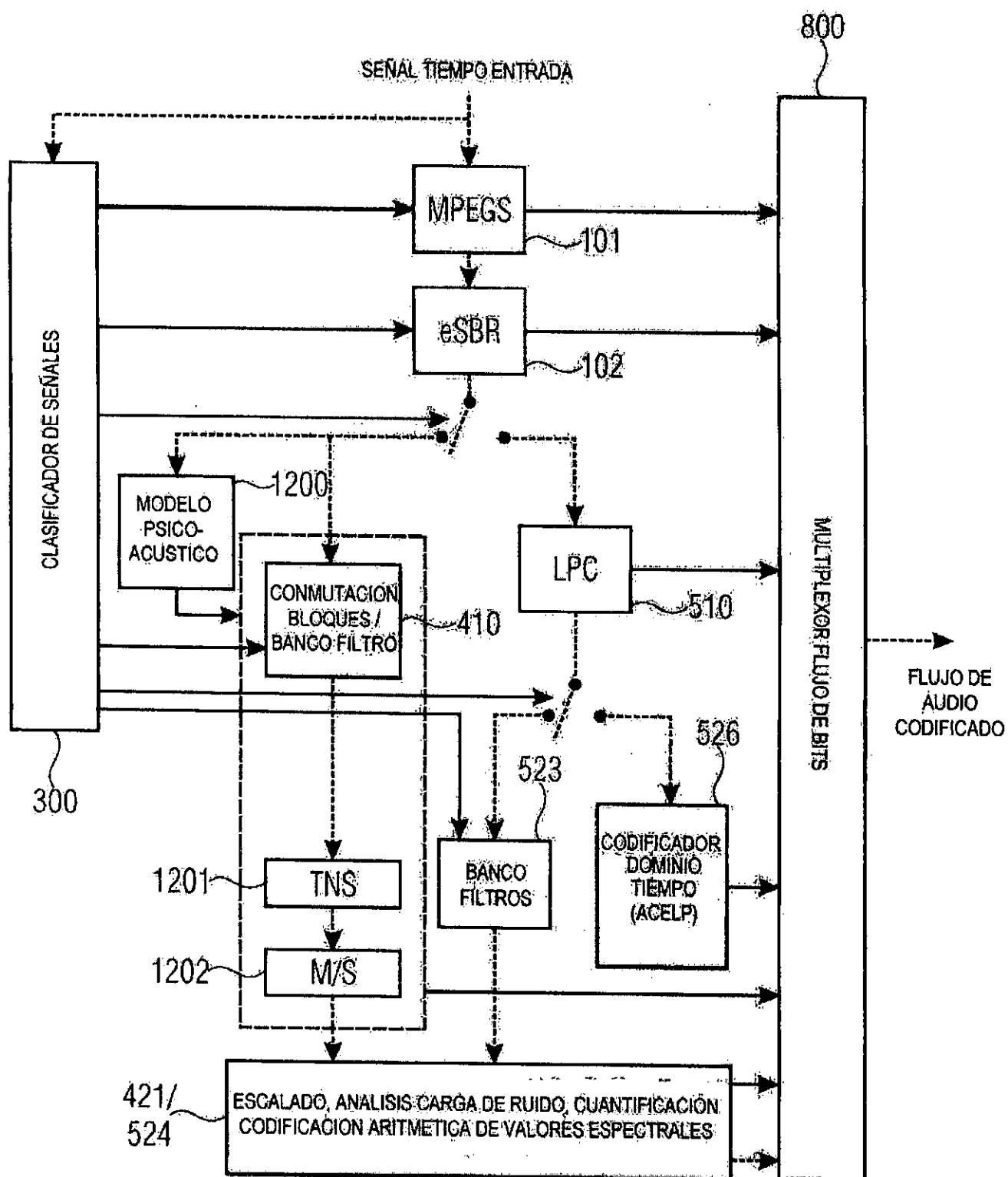


FIG. 12A

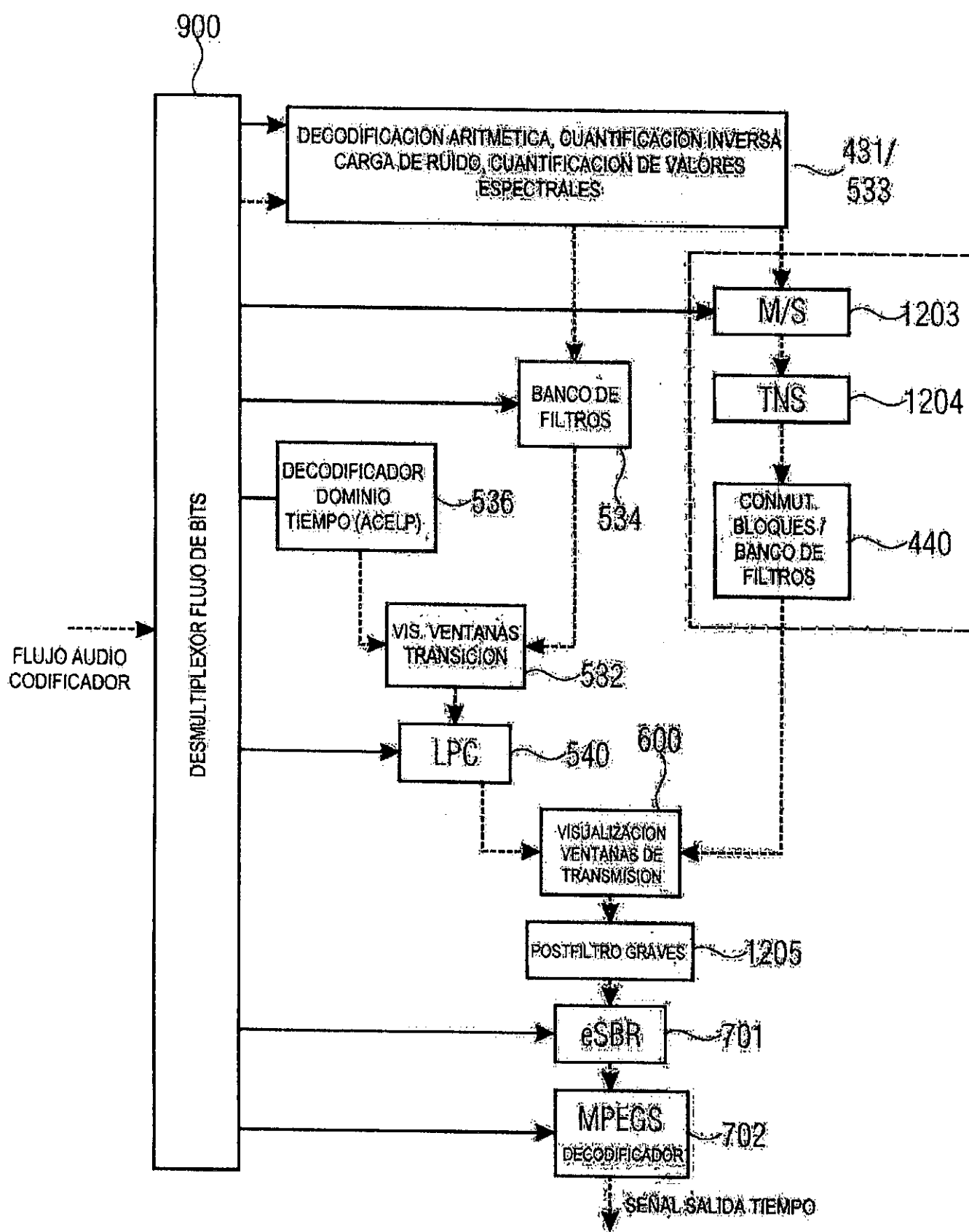
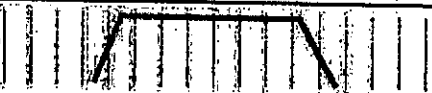
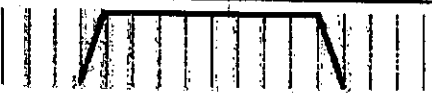


FIG. 12B

LONGITUD VENTANA	LONGITUD TRANSFORMADA	RESOLUCION TIEMPO	RESOLUCION FRECUENCIA
CORTA	PEQUEÑA	ALTA	BAJA
LARGA	GRANDE	BAJA	ALTA

FIG. 13A

VENTANA	Nº COEF.	APARECE COMO
LONG_WINDOW	1024/ 960	
SHORT_WINDOW	128/ 120	
LONG_START_WINDOW	1024/ 960	
LONG_STOP_WINDOW	1024/ 960	
STOP_START_WINDOW	1024/ 960	
START_WINDOW_LPD	1024/ 960	
STOP_WINDOW_1152	1152/ 1080	
STOP_START_WINDOW_1152	1152/ 1080	

VENTANAS DE TRANSFORMADAS

FIG. 13B

(RAMAS ACC Y TRANSICION)

VALOR	VENTANA DE SECUENCIAS	NUMERO DE VENTANAS	SE PARECE A
0	ONLY_LONG_SEQUENCE = LONG_WINDOW	1	
1	LONG_START_SEQUENCE = LONG_START_WINDOW	1	
2	EIGHT_SHORT_SEQUENCE = 8 * SHORT_WINDOW	8	
3	LONG_STOP_SEQUENCE = LONG_STOP_WINDOW	1	
1	STOP_START_SEQUENCE = STOP_START_WINDOW	1	
3	LPD_START_SEQUENCE = START_WINDOW_LPD	1	
3	STOP_1152_SEQUENCE = STOP_WINDOW_1152	1	
1	STOP_START_1152_SEQUENCE = STOP_START_WINDOW_1152	1	

SECUENCIA DE VENTANAS

FIG. 13C
(RAMAS ACC Y TRANSICION)

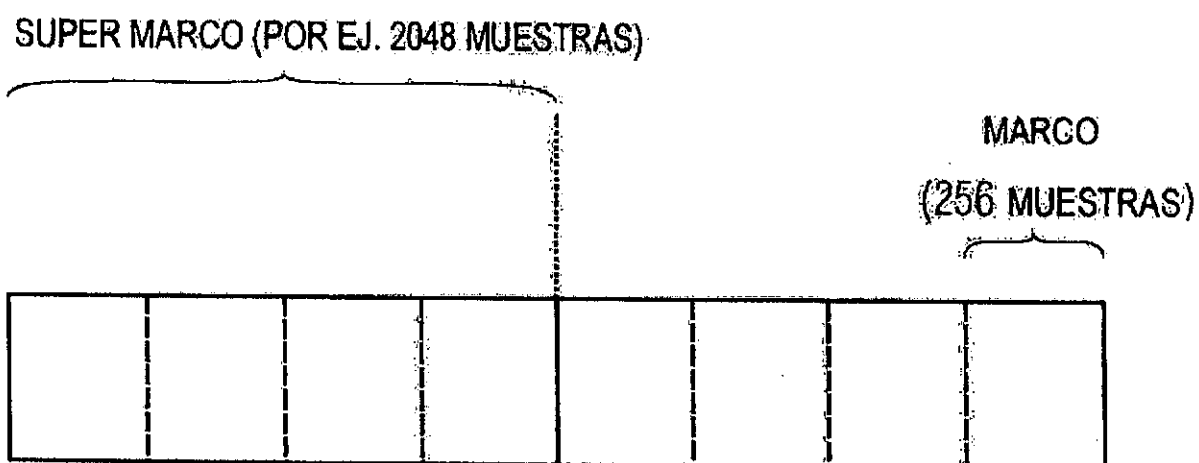


FIG. 14A

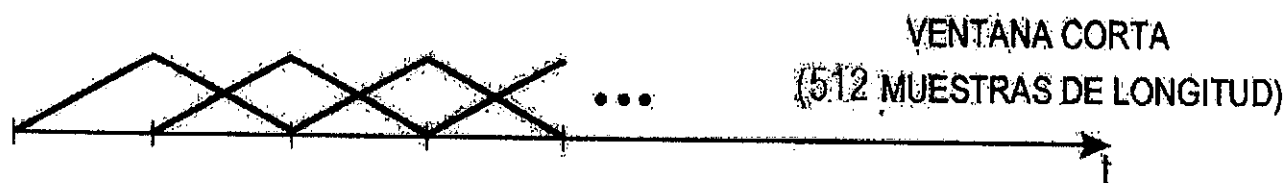


FIG. 14B

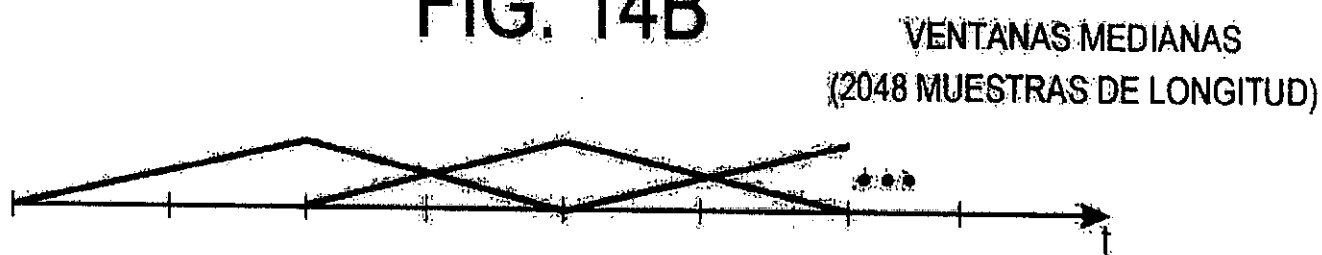


FIG. 14C

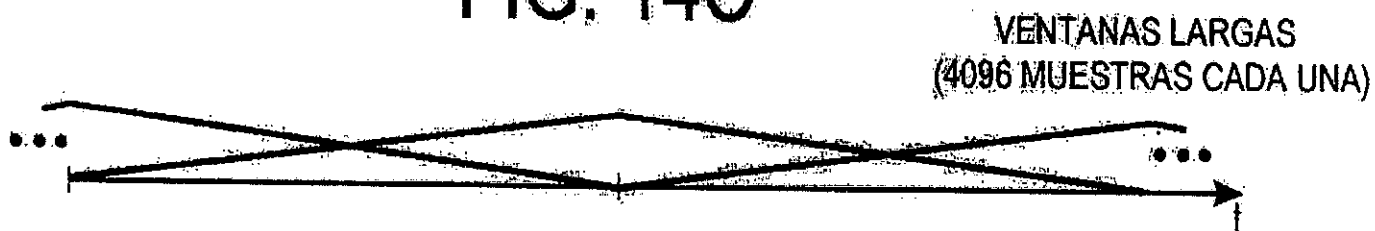


FIG. 14D

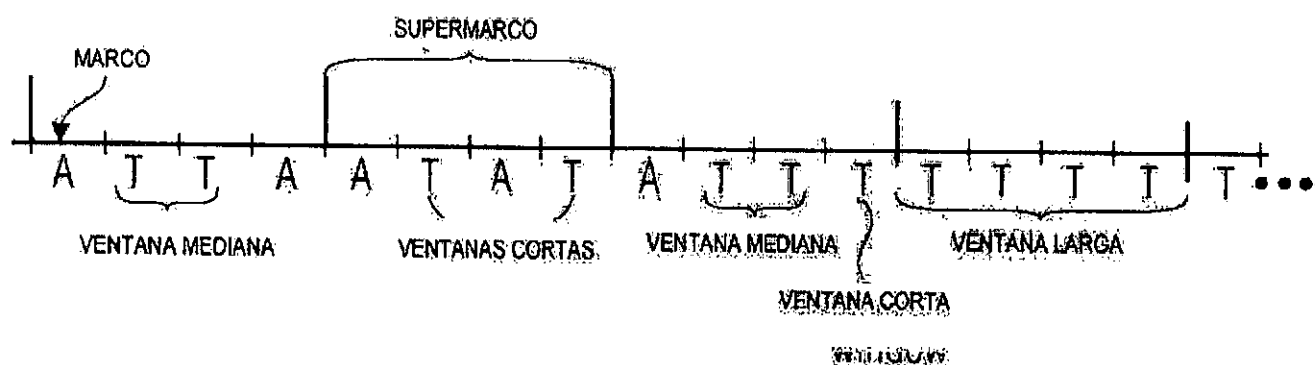
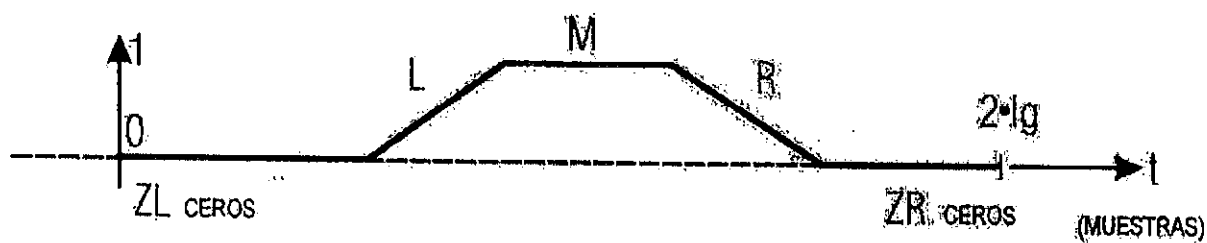


FIG. 14E

VALOR DE last_lpd_mode	VALOR DE mod[x]	NÚMERO DE COEFICIENTES ESPECTRALES l_0	ZL	L	M	R	ZR
0	1	320	160	0	256	128	96
0	2	576	288	0	512	128	224
0	3	1152	512	128	1024	128	512
1..3	1	256	64	128	128	128	64
1..3	2	512	192	128	384	128	192
1..3	3	1024	448	128	896	128	448

FIG. 14F



DEFINICIÓN DE VENTANAS DE LA FIG. 14F

FIG. 14G

Multi-Resolution Switched Audio Encoding/Decoding Scheme

5

Specification

The present invention is related to audio coding and, particularly, to low bit rate audio coding schemes.

10 In the art, frequency domain coding schemes such as MP3 or AAC are known. These frequency-domain encoders are based on a time-domain/frequency-domain conversion, a subsequent quantization stage, in which the quantization error is controlled using information from a perceptual module, and an encoding stage, in which the quantized spectral coefficients and corresponding side information are entropy-encoded using code tables.

15

On the other hand there are encoders that are very well suited to speech processing such as the AMR-WB+ as described in 3GPP TS 26.290. Such speech coding schemes perform a Linear Predictive filtering of a time-domain signal. Such a LP filtering is derived from a Linear Prediction analysis of the input time-domain signal. The resulting LP filter coefficients are then quantized/coded and transmitted as side information. The process is known as Linear Prediction Coding (LPC). At the output of the filter, the prediction residual signal or prediction error signal which is also known as the excitation signal is encoded using the analysis-by-synthesis stages of the ACELP encoder or, alternatively, is encoded using a transform encoder, which uses a Fourier transform with an overlap. The decision between the ACELP coding and the Transform Coded eXcitation coding which is also called TCX coding is done using a closed loop or an open loop algorithm.

Frequency-domain audio coding schemes such as the High Efficiency AAC (HE-AAC) encoding scheme, which combines an AAC coding scheme and a spectral band replication (SBR) technique can also be combined with a joint stereo or a multi-channel coding tool which is known under the term "MPEG surround".

On the other hand, speech encoders such as the AMR-WB+ also have a high frequency extension stage and a stereo functionality.

35

Frequency-domain coding schemes are advantageous in that they show a high quality at low bitrates for music signals. Problematic, however, is the quality of speech signals at low bitrates.

40 Speech coding schemes show a high quality for speech signals even at low bitrates, but show a poor quality for other signals at low bitrates.

It is an object of the present invention to provide an improved encoding/decoding concept.

This object is achieved by an audio encoder in accordance with claim 1, a method of audio encoding in accordance with claim 9, a decoder in accordance with claim 10, a method of decoding in accordance with claim 19, an encoded signal in accordance with claim 20 or a
 5 computer program in accordance with claim 21.

The present invention is based on the finding that a hybrid or dual-mode switched coding/encoding scheme is advantageous in that the best coding algorithm can always be selected for a certain signal characteristic. Stated differently, the present invention does not
 10 look for a signal coding algorithm which is perfectly matched to all signal characteristics. Such scheme would always be a compromise as can be seen from the huge differences between state of the art audio encoders on the one hand, and speech encoders on the other hand. Instead, the present invention combines different coding algorithms such as a speech coding algorithm on the one hand, and an audio coding algorithm on the other hand within a
 15 switched scheme so that, for each audio signal portion, the optimally matching coding algorithm is selected. Furthermore, it is also a feature of the present invention that both coding branches comprise a time/frequency converter, but in one coding branch, a further domain converter such an LPC processor is provided. This domain converter makes sure that the second coding branch is better suited for a certain signal characteristic than the first coding
 20 branch. However, it is also a feature of the present invention that the signal output by the domain processor is also transformed into a spectral representation.

Both converters, i.e., the first converter in the first coding branch and the second converter in the second coding branch are configured for applying a multi-resolution transform coding,
 25 where the resolution of the corresponding converter is set dependent on the audio signal, and particularly dependent on the audio signal actually coded in the corresponding coding branch so that a good compromise between quality on the one hand, and bitrate on the other hand, or in view of a certain fixed quality, the lowest bitrate, or in view of a fixed bitrate, the highest quality is obtained.

30 In accordance with the present invention, the time/frequency resolution of the two converters can preferably be set independent from each other so that each time/frequency transformer can be optimally matched to the time/frequency resolution requirements of the corresponding signal. The bit efficiency, i.e., the relation between useful bits on the one hand, and side
 35 information bits on the other hand is higher for longer block sizes/window lengths. Therefore, it is preferred that both converters are more biased to a longer window length, since, basically the same amount of side information refers to a longer time portion of the audio signal compared to applying shorter block sizes/window lengths/ transform lengths. Preferably, the time/frequency resolution in the encoding branches can also be influenced by other
 40 encoding/decoding tools located in these branches. Preferably, the second coding branch comprising the domain converter such as an LPC processor comprises another hybrid scheme such as an ACELP branch on the one hand, and an TCX scheme on the other hand, where the second converter is included in the TCX scheme. Preferably, the resolution of the

time/frequency converter located in the TCX branch is also influenced by the encoding decision, so that a portion of the signal in the second encoding branch is processed in the TCX branch having the second converter or in the ACELP branch not having a time/frequency converter.

5

Basically, neither the domain converter nor the second coding branch, and particularly the first processing branch in the second encoding branch and the second processing branch in the second coding branch, must be speech-related elements such as an LPC analyzer for the domain converter, a TCX encoder for the second processing branch and an ACELP encoder for the first processing branch. Other applications are also useful when other signal characteristics of an audio signal different from speech on the one hand, and music on the other hand are evaluated. Any domain converters and encoding branch implementations can be used and the best matching algorithm can be found by an analysis-by-synthesis scheme so that, on the encoder side, for each portion of the audio signal, all encoding alternatives are conducted and the best result is selected, where the best result can be found applying a target function to the encoding results. Then, side information identifying, to a decoder, the underlying encoding algorithm for a certain portion of the encoded audio signal is attached to the encoded audio signal by an encoder output interface so that the decoder does not have to care for any decisions on the encoder side or on any signal characteristics, but simply selects its coding branch depending on the transmitted side information. Furthermore, the decoder will not only select the correct decoding branch, but will also select, based on side information encoded in the encoded signal, which time/frequency resolution is to be applied in a corresponding first decoding branch and a corresponding second decoding branch.

Thus, the present invention provides an encoding/decoding scheme, which combines the advantages of all different coding algorithms and avoids the disadvantages of these coding algorithms which come up, when the signal portion would have to be encoded, by an algorithm that does not fit to a certain coding algorithm. Furthermore, the present invention avoids any disadvantages, which would come up, if the different time/frequency resolution requirements raised by different audio signal portions in different encoding branches had not been accounted for. Instead, due to the variable time/frequency resolution of time/frequency converters in both branches, any artifacts are at least reduced or even completely avoided, which would come up in the scenario where the same time/frequency resolution would be applied for both coding branches, or in which only a fixed time/frequency resolution would be possible for any coding branches.

The second switch again decides between two processing branches, but in a domain different from the "outer" first branch domain. Again one "inner" branch is mainly motivated by a source model or by SNR calculations, and the other "inner" branch can be motivated by a sink model and/or a psycho acoustic model, i.e. by masking or at least includes frequency/spectral domain coding aspects. Exemplarily, one "inner" branch has a frequency domain encoder/spectral converter and the other branch has an encoder coding on the other domain

such as the LPC domain, wherein this encoder is for example an CELP or ACELP quantizer/scaler processing an input signal without a spectral conversion.

5 A further preferred embodiment is an audio encoder comprising a first information sink oriented encoding branch such as a spectral domain encoding branch, a second information source or SNR oriented encoding branch such as an LPC-domain encoding branch, and a switch for switching between the first encoding branch and the second encoding branch, wherein the second encoding branch comprises a converter into a specific domain different from the time domain such as an LPC analysis stage generating an excitation signal, and
10 wherein the second encoding branch furthermore comprises a specific domain such as LPC domain processing branch and a specific spectral domain such as LPC spectral domain processing branch, and an additional switch for switching between the specific domain coding branch and the specific spectral domain coding branch.

15 A further embodiment of the invention is an audio decoder comprising a first domain such as a spectral domain decoding branch, a second domain such as an LPC domain decoding branch for decoding a signal such as an excitation signal in the second domain, and a third domain such as an LPC-spectral decoder branch for decoding a signal such as an excitation signal in a third domain such as an LPC spectral domain, wherein the third domain is obtained by
20 performing a frequency conversion from the second domain wherein a first switch for the second domain signal and the third domain signal is provided, and wherein a second switch for switching between the first domain decoder and the decoder for the second domain or the third domain is provided.

25 Preferred embodiments of the present invention are subsequently described with respect to the attached drawings, in which:

Fig. 1a is a block diagram of an encoding scheme in accordance with a first aspect of the present invention;
30

Fig. 1b is a block diagram of a decoding scheme in accordance with the first aspect of the present invention;

Fig. 1c is a block diagram of an encoding scheme in accordance with a further aspect of the present invention;
35

Fig. 2a is a block diagram of an encoding scheme in accordance with a second aspect of the present invention;

40 Fig. 2b is a schematic diagram of a decoding scheme in accordance with the second aspect of the present invention.

- Fig. 2c is a block diagram of an encoding scheme in accordance with a further aspect of the present invention
- 5 Fig. 3a illustrates a block diagram of an encoding scheme in accordance with a further aspect of the present invention;
- Fig. 3b illustrates a block diagram of a decoding scheme in accordance with the further aspect of the present invention;
- 10 Fig. 3c illustrates a schematic representation of the encoding apparatus/method with cascaded switches;
- Fig. 3d illustrates a schematic diagram of an apparatus or method for decoding, in which cascaded combiners are used;
- 15 Fig. 3e illustrates an illustration of a time domain signal and a corresponding representation of the encoded signal illustrating short cross fade regions which are included in both encoded signals;
- 20 Fig. 4a illustrates a block diagram with a switch positioned before the encoding branches;
- Fig. 4b illustrates a block diagram of an encoding scheme with the switch positioned subsequent to encoding the branches;
- 25 Fig. 5a illustrates a wave form of a time domain speech segment as a quasi-periodic or impulse-like signal segment;
- Fig. 5b illustrates a spectrum of the segment of Fig. 5a;
- 30 Fig. 5c illustrates a time domain speech segment of unvoiced speech as an example for a noise-like segment;
- Fig. 5d illustrates a spectrum of the time domain wave form of Fig. 5c;
- 35 Fig. 6 illustrates a block diagram of an analysis by synthesis CELP encoder;
- Figs. 7a to 7d illustrate voiced/unvoiced excitation signals as an example for impulse-like signals;
- 40 Fig. 7e illustrates an encoder-side LPC stage providing short-term prediction information and the prediction error (excitation) signal;

- Fig. 7f illustrates a further embodiment of an LPC device for generating a weighted signal;
- 5 Fig. 7g illustrates an implementation for transforming a weighted signal into an excitation signal by applying an inverse weighting operation and a subsequent excitation analysis as required in the converter 537 of Fig. 2b;
- Fig. 8 illustrates a block diagram of a joint multi-channel algorithm in accordance with an embodiment of the present invention;
- 10 Fig. 9 illustrates a preferred embodiment of a bandwidth extension algorithm;
- Fig. 10a illustrates a detailed description of the switch when performing an open loop decision; and
- 15 Fig. 10b illustrates an illustration of the switch when operating in a closed loop decision mode;
- Fig. 11A illustrates a block diagram of an audio encoder in accordance with another aspect of the present invention;
- 20 Fig. 11B illustrates a block diagram of another embodiment of an inventive audio decoder;
- 25 Fig. 12A illustrates another embodiment of an inventive encoder;
- Fig. 12B illustrates another embodiment of an inventive decoder;
- Fig. 13A illustrates the interrelation between resolution and window/transform lengths;
- 30 Fig. 13B illustrates an overview of a set of transform windows for the first coding branch and a transition from the first to the second coding branch;
- 35 Fig. 13C illustrates a plurality of different window sequences including window sequences for the first coding branch and sequences for a transition to the second branch;
- Fig. 14A illustrates the framing of a preferred embodiment of the second coding branch;
- 40 Fig. 14B illustrates short windows as applied in the second coding branch;

Fig. 14C illustrates medium sized windows applied in the second coding branch;

Fig. 14D illustrates long windows applied by the second coding branch;

5 Fig. 14E illustrates an exemplary sequence of ACELP frames and TCX frames within a super frame division;

Fig. 14F illustrates different transform lengths corresponding to different time/frequency resolutions for the second encoding branch; and

10

Fig. 14G illustrates a construction of a window using the definitions of Fig. 14F

Fig. 11A illustrates an embodiment of an audio encoder for encoding an audio signal. The encoder comprise a first coding branch 400 for encoding an audio signal using a first coding algorithm to obtain a first encoded signal.

15

The audio encoder furthermore comprises a second coding branch 500 for encoding an audio signal using a second coding algorithm to obtain a second encoded signal. The first coding algorithm is different from the second coding algorithm. Additionally, a first switch 200 for switching between the first coding branch and the second coding branch is provided so that, for a portion of the audio signal, either the first encoded signal or the second encoded signal is in an encoder output signal 801.

20

The audio encoder illustrated in Fig. 11A additionally comprises a signal analyzer 300/525, which is configured for analyzing a portion of the audio signal to determine, whether the portion of the audio signal is represented as the first encoded signal or the second encoded signal in the encoder output signal 801.

25

The signal analyzer 300/525 is furthermore configured for variably determining a respective time/frequency resolution of a first converter 410 in the first coding branch 400 or a second converter 523 in the second encoding branch 500. This time/frequency resolution is applied, when the first encoded signal or the second encoded signal representing the portion of the audio signal is generated.

30

The audio encoder additionally comprises an output interface 800 for generating the encoder output signal 801 comprising an encoded representation of the portion of the audio signal and an information indicating whether the representation of the audio signal is the first encoded signal or the second encoded signal, and indicating the time/frequency resolution used for decoding the first encoded signal and the second encoded signal.

35

40

The second encoding branch is preferably different from the first encoding branch in that the second encoding branch additionally comprises a domain converter for converting the audio signal from the domain, in which the audio signal is processed in the first encoding branch

into a different domain. Preferably the domain converter is an LPC processor 510, but the domain converter can be implemented in any other way as long as the domain converter is different from the first converter 410 and the second converter 523.

- 5 The first converter 410 is a time/frequency converter preferably comprising a windower 410a and a transformer 410b. The windower 410a applies an analysis window to the input audio signal, and the transformer 410b performs a conversion of the windowed signal into a spectral representation.
- 10 Analogously, the second converter 523 preferably comprises a windower 523a and a subsequently connected transformer 523b. The windower 523a receives the signal output by the domain converter 510 and outputs the windowed representation thereof. The result of one analysis window applied by the windower 523a is input into the transformer 523b to form a spectral representation. The transformer can be an FFT or preferably MDCT processor
- 15 implementing a corresponding algorithm in software or hardware or in a mixed hardware/software implementation. Alternatively, the transformer can be a filterbank implementation such as a QMF filterbank which can be based on a real-valued or complex modulation of a prototype filter. For specific filterbank implementations, a window is applied. However, for other filterbank implementations, a windowing as required for a transform
- 20 algorithm based on a FFT or MDCT is not necessary. When a filterbank implementation is used, then the filterbank is a variable resolution filterbank and the resolution controls the frequency resolution of the filterbank, and additionally, the time resolution or only the frequency resolution and not the time resolution. When however, the converter is implemented as an FFT or MDCT or any other corresponding transformer, then the frequency resolution is
- 25 connected to the time resolution in that an increase of the frequency resolution obtained by a larger block length in time automatically corresponds to a lower time resolution and vice versa.

30 Additionally, the first coding branch may comprise a quantizer/coder stage 421, and the second encoding branch may also comprise one or more further coding tools 524.

Importantly, the signal analyzer is configured for generating a resolution control signal for the first converter 510 and for the second converter 523. Thus, an independent resolution control in both coding branches is implemented in order to have a coding scheme which, on the one

35 hand, provides a low bitrate, and on the other hand, provides a maximum quality in view of the low bitrate. In order to achieve the low bitrate goal, longer window lengths or longer transform lengths are preferred, but in situations where these long lengths will result in an artifact due to the low time resolution, shorter window lengths and shorter transform lengths are applied, which results in a lower frequency resolution. Preferably, the signal analyzer

40 applies a statistical analysis or any other analysis which is suited to the corresponding algorithms in the encoding branches. In one implementation mode, in which the first coding branch is a frequency domain coding branch such as an AAC-based encoder, and in which the second coding branch comprises, as a domain converter, an LPC processor 510, the signal

analyzer performs a speech/music discrimination so that the speech portion of the audio signal is fed into the second coding branch by correspondingly controlling the switch 200. A music portion of the audio signal is fed into the first coding branch 400 by correspondingly controlling the switch 200 as indicated by the switch control lines. Alternatively, as will be
 5 later discussed with respect to Fig. 1C or Fig. 4B, the switch can also be positioned before the output interface 800.

Furthermore, the signal analyzer can receive the audio signal input into the switch 200, or the audio signal output by the switch 200. Furthermore, the signal analyzer performs an analysis
 10 in order to not only feed the audio signal into the corresponding coding branch, but to also determine the appropriate time/frequency resolution of the respective converter in the corresponding coding branch, such as the first converter 410 and the second converter 523 as indicated by the resolution controlled lines connecting the signal analyzer and the converter.

15 Fig. 11B comprises a preferred embodiment of an audio decoder matching to the audio encoder in Fig. 11A.

The audio decoder in Fig. 11B is configured for decoding an encoded audio signal such as the encoder output signal 801 output by the output interface 800 in Fig. 11A. The encoded signal
 20 comprises a first encoded audio signal encoded in accordance with a first coding algorithm, a second encoded signal encoded in accordance with a second coding algorithm, the second coding algorithm being different from the first coding algorithm, and information, indicating whether the first coding algorithm or the second coding algorithm is used for decoding the first encoded signal and the second encoded signal, and a time/frequency resolution
 25 information for the first encoded audio signal and the second encoded audio signal.

The audio decoder comprises a first decoding branch 431, 440 for decoding the first encoded signal based on the first coding algorithm. Furthermore, the audio decoder comprises a second decoding branch for decoding the second encoded signal using the second coding algorithm.
 30

The first decoding branch comprises a first controllable converter 440 for converting from a spectral domain into the time domain. The controllable converter is configured for being controlled using the time/frequency resolution information from the first encoded signal to obtain the first decoded signal.
 35

The second decoding branch comprises a second controllable converter for converting from a spectral representation in a time representation, the second controllable converter 534 being configured for being controlled using the time/frequency resolution information 991 for the second encoded signal.
 40

The decoder additionally comprises a controller 990 for controlling the first converter 540 and the second converter 534 in accordance with the time/frequency resolution information 991.

Furthermore, the decoder comprises a domain converter for generating a synthesis signal using the second decoded signal in order to cancel the domain conversion applied by the domain converter 510 in the encoder of Fig. 11A.

5 Preferably, the domain converter 540 is an LPC synthesis processor, which is controlled using LPC filter information included in the encoded signal, where this LPC filter information has been generated by the LPC processor 510 in Fig. 11A and has been input into the encoder output signal as side information. The audio decoder finally comprises a combiner 600 for combining the first decoded signal output by the first domain converter 440 and the synthesis
10 signal to obtain a decoded audio signal 609.

In the preferred implementation, the first decoding branch additionally comprises a dequantizer/decoder stage 431 for reversing or at least for partly reversing the operations performed by the corresponding encoder stage 421. However, it is clear that quantization
15 cannot be reversed, since this is a lossy operation. However, a dequantizer will reverse a certain non-uniformity in a quantization such as a logarithmic or companding quantization.

In the second decoding branch, the corresponding stage 533 is applied for undoing certain encoding operations applied by the stage 524. Preferably, stage 524 comprises a uniform
20 quantization. Therefore, the corresponding stage 533 will not have a specific dequantization stage for undoing a certain uniform quantization.

The first converter 440 as well as the second converter 534 may comprise a corresponding inverse transformer stage 440a, 534a, a synthesis window stage 440b, 534b, and the
25 subsequently connected overlap/add stage 440c, 534c. The overlap/add stages are required, when the converters, and more specifically, the transformer stages 440a, 534a apply aliasing introducing transforms such as a modified discrete cosine transform. Then, the overlap/add operation will perform a time domain aliasing cancellation (TDAC). When however, the transformers apply a non-aliasing introducing transform such as an inverse FFT, then an
30 overlap/add stage 440c is not required. In such an implementation, a cross fading operation to avoid blocking artifacts may be applied.

Analogously, the combiner 600 may be a switched combiner or a cross fading combiner, or when aliasing is used for avoiding blocking artifacts, a transition windowing operation is
35 implemented by the combiner similar to an overlap/add stage within a branch itself.

Fig. 1a illustrates an embodiment of the invention having two cascaded switches. A mono signal, a stereo signal or a multi-channel signal is input into the switch 200. The switch 200 is controlled by the decision stage 300. The decision stage receives, as an input, a signal input
40 into block 200. Alternatively, the decision stage 300 may also receive a side information which is included in the mono signal, the stereo signal or the multi-channel signal or is at least associated to such a signal, where information is existing, which was, for example, generated when originally producing the mono signal, the stereo signal or the multi-channel signal.

The decision stage 300 actuates the switch 200 in order to feed a signal either in the frequency encoding portion 400 illustrated at an upper branch of Fig. 1a or the LPC-domain encoding portion 500 illustrated at a lower branch in Fig. 1a. A key element of the frequency domain encoding branch is the spectral conversion block 410 which is operative to convert a common preprocessing stage output signal (as discussed later on) into a spectral domain. The spectral conversion block may include an MDCT algorithm, a QMF, an FFT algorithm, a Wavelet analysis or a filterbank such as a critically sampled filterbank having a certain number of filterbank channels, where the subband signals in this filterbank may be real valued signals or complex valued signals. The output of the spectral conversion block 410 is encoded using a spectral audio encoder 421, which may include processing blocks as known from the AAC coding scheme.

Generally, the processing in branch 400 is a processing in a perception based model or information sink model. Thus, this branch models the human auditory system receiving sound. Contrary thereto, the processing in branch 500 is to generate a signal in the excitation, residual or LPC domain. Generally, the processing in branch 500 is a processing in a speech model or an information generation model. For speech signals, this model is a model of the human speech/sound generation system generating sound. If, however, a sound from a different source requiring a different sound generation model is to be encoded, then the processing in branch 500 may be different.

In the lower encoding branch 500, a key element is an LPC device 510, which outputs an LPC information which is used for controlling the characteristics of an LPC filter. This LPC information is transmitted to a decoder. The LPC stage 510 output signal is an LPC-domain signal which consists of an excitation signal and/or a weighted signal.

The LPC device generally outputs an LPC domain signal, which can be any signal in the LPC domain such as the excitation signal in Fig. 7e or a weighted signal in Fig. 7f or any other signal, which has been generated by applying LPC filter coefficients to an audio signal. Furthermore, an LPC device can also determine these coefficients and can also quantize/encode these coefficients.

The decision in the decision stage can be signal-adaptive so that the decision stage performs a music/speech discrimination and controls the switch 200 in such a way that music signals are input into the upper branch 400, and speech signals are input into the lower branch 500. In one embodiment, the decision stage is feeding its decision information into an output bit stream so that a decoder can use this decision information in order to perform the correct decoding operations.

Such a decoder is illustrated in Fig. 1b. The signal output by the spectral audio encoder 421 is, after transmission, input into a spectral audio decoder 431. The output of the spectral audio decoder 431 is input into a time-domain converter 440. Analogously, the output of the LPC

domain encoding branch 500 of Fig. 1a is received on the decoder side and processed by elements 531, 533, 534, and 532 for obtaining an LPC excitation signal. The LPC excitation signal is input into an LPC synthesis stage 540, which receives, as a further input, the LPC information generated by the corresponding LPC analysis stage 510. The output of the time-domain converter 440 and/or the output of the LPC synthesis stage 540 are input into a switch 600. The switch 600 is controlled via a switch control signal which was, for example, generated by the decision stage 300, or which was externally provided such as by a creator of the original mono signal, stereo signal or multi-channel signal. The output of the switch 600 is a complete mono signal, stereo signal or multichannel signal.

The input signal into the switch 200 and the decision stage 300 can be a mono signal, a stereo signal, a multi-channel signal or generally an audio signal. Depending on the decision which can be derived from the switch 200 input signal or from any external source such as a producer of the original audio signal underlying the signal input into stage 200, the switch switches between the frequency encoding branch 400 and the LPC encoding branch 500. The frequency encoding branch 400 comprises a spectral conversion stage 410 and a subsequently connected quantizing/coding stage 421. The quantizing/coding stage can include any of the functionalities as known from modern frequency-domain encoders such as the AAC encoder. Furthermore, the quantization operation in the quantizing/coding stage 421 can be controlled via a psychoacoustic module which generates psychoacoustic information such as a psychoacoustic masking threshold over the frequency, where this information is input into the stage 421.

In the LPC encoding branch, the switch output signal is processed via an LPC analysis stage 510 generating LPC side info and an LPC-domain signal. The excitation encoder inventively comprises an additional switch for switching the further processing of the LPC-domain signal between a quantization/coding operation 522 in the LPC-domain or a quantization/coding stage 524, which is processing values in the LPC-spectral domain. To this end, a spectral converter 523 is provided at the input of the quantizing/coding stage 524. The switch 521 is controlled in an open loop fashion or a closed loop fashion depending on specific settings as, for example, described in the AMR-WB+ technical specification.

For the closed loop control mode, the encoder additionally includes an inverse quantizer/coder 531 for the LPC domain signal, an inverse quantizer/coder 533 for the LPC spectral domain signal and an inverse spectral converter 534 for the output of item 533. Both encoded and again decoded signals in the processing branches of the second encoding branch are input into the switch control device 525. In the switch control device 525, these two output signals are compared to each other and/or to a target function or a target function is calculated which may be based on a comparison of the distortion in both signals so that the signal having the lower distortion is used for deciding, which position the switch 521 should take. Alternatively, in case both branches provide non-constant bit rates, the branch providing the lower bit rate might be selected even when the signal to noise ratio of this branch is lower than the signal to noise ratio of the other branch. Alternatively, the target function could use, as an input, the

signal to noise ratio of each signal and a bit rate of each signal and/or additional criteria in order to find the best decision for a specific goal. If, for example, the goal is such that the bit rate should be as low as possible, then the target function would heavily rely on the bit rate of the two signals output by the elements 531, 534. However, when the main goal is to have the best quality for a certain bit rate, then the switch control 525 might, for example, discard each signal which is above the allowed bit rate and when both signals are below the allowed bit rate, the switch control would select the signal having the better signal to noise ratio, i.e., having the smaller quantization/coding distortions.

The decoding scheme in accordance with the present invention is, as stated before, illustrated in Fig. 1b. For each of the three possible output signal kinds, a specific decoding/re-quantizing stage 431, 531 or 533 exists. While stage 431 outputs a time-spectrum which is converted into the time-domain using the frequency/time converter 440, stage 531 outputs an LPC-domain signal, and item 533 outputs an LPC-spectrum. In order to make sure that the input signals into switch 532 are both in the LPC-domain, the LPC-spectrum/LPC-converter 534 is provided. The output data of the switch 532 is transformed back into the time-domain using an LPC synthesis stage 540, which is controlled via encoder-side generated and transmitted LPC information. Then, subsequent to block 540, both branches have time-domain information which is switched in accordance with a switch control signal in order to finally obtain an audio signal such as a mono signal, a stereo signal or a multi-channel signal, which depends on the signal input into the encoding scheme of Fig. 1a.

Fig. 1c illustrates a further embodiment with a different arrangement of the switch 521 similar to the principle of Fig. 4b.

Fig. 2a illustrates a preferred encoding scheme in accordance with a second aspect of the invention. A common preprocessing scheme connected to the switch 200 input may comprise a surround/joint stereo block 101 which generates, as an output, joint stereo parameters and a mono output signal, which is generated by downmixing the input signal which is a signal having two or more channels. Generally, the signal at the output of block 101 can also be a signal having more channels, but due to the downmixing functionality of block 101, the number of channels at the output of block 101 will be smaller than the number of channels input into block 101.

The common preprocessing scheme may comprise alternatively to the block 101 or in addition to the block 101 a bandwidth extension stage 102. In the Fig. 2a embodiment, the output of block 101 is input into the bandwidth extension block 102 which, in the encoder of Fig. 2a, outputs a band-limited signal such as the low band signal or the low pass signal at its output. Preferably, this signal is downsampled (e.g. by a factor of two) as well. Furthermore, for the high band of the signal input into block 102, bandwidth extension parameters such as spectral envelope parameters, inverse filtering parameters, noise floor parameters etc. as known from HE-AAC profile of MPEG-4 are generated and forwarded to a bitstream multiplexer 800.

Preferably, the decision stage 300 receives the signal input into block 101 or input into block 102 in order to decide between, for example, a music mode or a speech mode. In the music mode, the upper encoding branch 400 is selected, while, in the speech mode, the lower encoding branch 500 is selected. Preferably, the decision stage additionally controls the joint stereo block 101 and/or the bandwidth extension block 102 to adapt the functionality of these blocks to the specific signal. Thus, when the decision stage determines that a certain time portion of the input signal is of the first mode such as the music mode, then specific features of block 101 and/or block 102 can be controlled by the decision stage 300. Alternatively, when the decision stage 300 determines that the signal is in a speech mode or, generally, in a second LPC-domain mode, then specific features of blocks 101 and 102 can be controlled in accordance with the decision stage output.

Preferably, the spectral conversion of the coding branch 400 is done using an MDCT operation which, even more preferably, is the time-warped MDCT operation, where the strength or, generally, the warping strength can be controlled between zero and a high warping strength. In a zero warping strength, the MDCT operation in block 411 is a straight-forward MDCT operation known in the art. The time warping strength together with time warping side information can be transmitted/input into the bitstream multiplexer 800 as side information.

In the LPC encoding branch, the LPC-domain encoder may include an ACELP core 526 calculating a pitch gain, a pitch lag and/or codebook information such as a codebook index and gain. The TCX mode as known from 3GPP TS 26.290 incurs a processing of a perceptually weighted signal in the transform domain. A Fourier transformed weighted signal is quantized using a split multi-rate lattice quantization (algebraic VQ) with noise factor quantization. A transform is calculated in 1024, 512, or 256 sample windows. The excitation signal is recovered by inverse filtering the quantized weighted signal through an inverse weighting filter.

In the first coding branch 400, a spectral converter preferably comprises a specifically adapted MDCT operation having certain window functions followed by a quantization/entropy encoding stage which may consist of a single vector quantization stage, but preferably is a combined scalar quantizer/entropy coder similar to the quantizer/coder in the frequency domain coding branch, i.e., in item 421 of Fig. 2a.

In the second coding branch, there is the LPC block 510 followed by a switch 521, again followed by an ACELP block 526 or an TCX block 527. ACELP is described in 3GPP TS 26.190 and TCX is described in 3GPP TS 26.290. Generally, the ACELP block 526 receives an LPC excitation signal as calculated by a procedure as described in Fig. 7e. The TCX block 527 receives a weighted signal as generated by Fig. 7f.

In TCX, the transform is applied to the weighted signal computed by filtering the input signal through an LPC-based weighting filter. The weighting filter used preferred embodiments of the invention is given by $(1 - A(z/\gamma))/(1 - \mu z^{-1})$. Thus, the weighted signal is an LPC domain

signal and its transform is an LPC-spectral domain. The signal processed by ACELP block 526 is the excitation signal and is different from the signal processed by the block 527, but both signals are in the LPC domain.

- 5 At the decoder side illustrated in Fig. 2b, after the inverse spectral transform in block 537, the inverse of the weighting filter is applied, that is $(1 - \mu z^{-1}) / (1 - A(z/\gamma))$. Then, the signal is filtered through $(1 - A(z))$ to go to the LPC excitation domain. Thus, the conversion to LPC domain block 534 and the TCX⁻¹ block 537 include inverse transform and then filtering through $\frac{(1 - \mu z^{-1})}{(1 - A(z/\gamma))} (1 - A(z))$ to convert from the weighted domain to the excitation domain.

10

- Although item 510 in Figs. 1a, 1c, 2a, 2c illustrates a single block, block 510 can output different signals as long as these signals are in the LPC domain. The actual mode of block 510 such as the excitation signal mode or the weighted signal mode can depend on the actual switch state. Alternatively, the block 510 can have two parallel processing devices, where one device is implemented similar to Fig. 7e and the other device is implemented as Fig. 7f. Hence, the LPC domain at the output of 510 can represent either the LPC excitation signal or the LPC weighted signal or any other LPC domain signal.

- 20 In the second encoding branch (ACELP/TCX) of Fig. 2a or 2c, the signal is preferably pre-emphasized through a filter $1 - 0.68z^{-1}$ before encoding. At the ACELP/TCX decoder in Fig. 2b the synthesized signal is deemphasized with the filter $1/(1 - 0.68z^{-1})$. The preemphasis can be part of the LPC block 510 where the signal is preemphasized before LPC analysis and quantization. Similarly, deemphasis can be part of the LPC synthesis block LPC⁻¹ 540.

- 25 Fig. 2c illustrates a further embodiment for the implementation of Fig. 2a, but with a different arrangement of the switch 521 similar to the principle of Fig. 4b.

- In a preferred embodiment, the first switch 200 (see Fig. 1a or 2a) is controlled through an open-loop decision (as in Fig. 4a) and the second switch is controlled through a closed-loop decision (as in figure 4b).

- 30 For example, Fig. 2c, has the second switch placed after the ACELP and TCX branches as in Fig. 4b. Then, in the first processing branch, the first LPC domain represents the LPC excitation, and in the second processing branch, the second LPC domain represents the LPC weighted signal. That is, the first LPC domain signal is obtained by filtering through $(1 - A(z))$ to convert to the LPC residual domain, while the second LPC domain signal is obtained by filtering through the filter $(1 - A(z/\gamma)) / (1 - \mu z^{-1})$ to convert to the LPC weighted domain.

- 40 Fig. 2b illustrates a decoding scheme corresponding to the encoding scheme of Fig. 2a. The bitstream generated by bitstream multiplexer 800 of Fig. 2a is input into a bitstream demultiplexer 900. Depending on an information derived for example from the bitstream via a mode detection block 601, a decoder-side switch 600 is controlled to either forward signals

from the upper branch or signals from the lower branch to the bandwidth extension block 701. The bandwidth extension block 701 receives, from the bitstream demultiplexer 900, side information and, based on this side information and the output of the mode decision 601, reconstructs the high band based on the low band output by switch 600.

5

The full band signal generated by block 701 is input into the joint stereo/surround processing stage 702, which reconstructs two stereo channels or several multi-channels. Generally, block 702 will output more channels than were input into this block. Depending on the application, the input into block 702 may even include two channels such as in a stereo mode and may
10 even include more channels as long as the output by this block has more channels than the input into this block.

The switch 200 has been shown to switch between both branches so that only one branch receives a signal to process and the other branch does not receive a signal to process. In an
15 alternative embodiment, however, the switch may also be arranged subsequent to for example the audio encoder 421 and the excitation encoder 522, 523, 524, which means that both branches 400, 500 process the same signal in parallel. In order to not double the bitrate, however, only the signal output by one of those encoding branches 400 or 500 is selected to be written into the output bitstream. The decision stage will then operate so that the signal written
20 into the bitstream minimizes a certain cost function, where the cost function can be the generated bitrate or the generated perceptual distortion or a combined rate/distortion cost function. Therefore, either in this mode or in the mode illustrated in the Figures, the decision stage can also operate in a closed loop mode in order to make sure that, finally, only the encoding branch output is written into the bitstream which has for a given perceptual
25 distortion the lowest bitrate or, for a given bitrate, has the lowest perceptual distortion. In the closed loop mode, the feedback input may be derived from outputs of the three quantizer/scaler blocks 421, 522 and 524 in Fig. 1a.

In the implementation having two switches, i.e., the first switch 200 and the second switch
30 521, it is preferred that the time resolution for the first switch is lower than the time resolution for the second switch. Stated differently, the blocks of the input signal into the first switch, which can be switched via a switch operation are larger than the blocks switched by the second switch operating in the LPC-domain. Exemplarily, the frequency domain/LPC-domain switch 200 may switch blocks of a length of 1024 samples, and the second switch 521 can
35 switch blocks having 256 samples each.

Although some of the Figs. 1a through 10b are illustrated as block diagrams of an apparatus, these figures simultaneously are an illustration of a method, where the block functionalities correspond to the method steps.

40

Fig. 3a illustrates an audio encoder for generating an encoded audio signal as an output of the first encoding branch 400 and a second encoding branch 500. Furthermore, the encoded audio signal preferably includes side information such as pre-processing parameters from the

common pre-processing stage or, as discussed in connection with preceding Figs., switch control information.

5 Preferably, the first encoding branch is operative in order to encode an audio intermediate signal 195 in accordance with a first coding algorithm, wherein the first coding algorithm has an information sink model. The first encoding branch 400 generates the first encoder output signal which is an encoded spectral information representation of the audio intermediate signal 195.

10 Furthermore, the second encoding branch 500 is adapted for encoding the audio intermediate signal 195 in accordance with a second encoding algorithm, the second coding algorithm having an information source model and generating, in a second encoder output signal, encoded parameters for the information source model representing the intermediate audio signal.

15 The audio encoder furthermore comprises the common pre-processing stage for pre-processing an audio input signal 99 to obtain the audio intermediate signal 195. Specifically, the common pre-processing stage is operative to process the audio input signal 99 so that the audio intermediate signal 195, i.e., the output of the common pre-processing algorithm is a
20 compressed version of the audio input signal.

A preferred method of audio encoding for generating an encoded audio signal, comprises a step of encoding 400 an audio intermediate signal 195 in accordance with a first coding algorithm, the first coding algorithm having an information sink model and generating, in a
25 first output signal, encoded spectral information representing the audio signal; a step of encoding 500 an audio intermediate signal 195 in accordance with a second coding algorithm, the second coding algorithm having an information source model and generating, in a second output signal, encoded parameters for the information source model representing the intermediate signal 195, and a step of commonly pre-processing 100 an audio input signal 99
30 to obtain the audio intermediate signal 195, wherein, in the step of commonly pre-processing the audio input signal 99 is processed so that the audio intermediate signal 195 is a compressed version of the audio input signal 99, wherein the encoded audio signal includes, for a certain portion of the audio signal either the first output signal or the second output signal. The method preferably includes the further step encoding a certain portion of the audio
35 intermediate signal either using the first coding algorithm or using the second coding algorithm or encoding the signal using both algorithms and outputting in an encoded signal either the result of the first coding algorithm or the result of the second coding algorithm.

40 Generally, the audio encoding algorithm used in the first encoding branch 400 reflects and models the situation in an audio sink. The sink of an audio information is normally the human ear. The human ear can be modeled as a frequency analyzer. Therefore, the first encoding branch outputs encoded spectral information. Preferably, the first encoding branch furthermore includes a psychoacoustic model for additionally applying a psychoacoustic

masking threshold. This psychoacoustic masking threshold is used when quantizing audio spectral values where, preferably, the quantization is performed such that a quantization noise is introduced by quantizing the spectral audio values, which are hidden below the psychoacoustic masking threshold.

5

The second encoding branch represents an information source model, which reflects the generation of audio sound. Therefore, information source models may include a speech model which is reflected by an LPC analysis stage, i.e., by transforming a time domain signal into an LPC domain and by subsequently processing the LPC residual signal, i.e., the excitation signal. Alternative sound source models, however, are sound source models for representing a certain instrument or any other sound generators such as a specific sound source existing in real world. A selection between different sound source models can be performed when several sound source models are available, for example based on an SNR calculation, i.e., based on a calculation, which of the source models is the best one suitable for encoding a certain time portion and/or frequency portion of an audio signal. Preferably, however, the switch between encoding branches is performed in the time domain, i.e., that a certain time portion is encoded using one model and a certain different time portion of the intermediate signal is encoded using the other encoding branch.

10

15

Information source models are represented by certain parameters. Regarding the speech model, the parameters are LPC parameters and coded excitation parameters, when a modern speech coder such as AMR-WB+ is considered. The AMR-WB+ comprises an ACELP encoder and a TCX encoder. In this case, the coded excitation parameters can be global gain, noise floor, and variable length codes.

25

Fig. 3b illustrates a decoder corresponding to the encoder illustrated in Fig. 3a. Generally, Fig. 3b illustrates an audio decoder for decoding an encoded audio signal to obtain a decoded audio signal 799. The decoder includes the first decoding branch 450 for decoding an encoded signal encoded in accordance with a first coding algorithm having an information sink model.

30

The audio decoder furthermore includes a second decoding branch 550 for decoding an encoded information signal encoded in accordance with a second coding algorithm having an information source model. The audio decoder furthermore includes a combiner for combining output signals from the first decoding branch 450 and the second decoding branch 550 to obtain a combined signal. The combined signal which is illustrated in Fig. 3b as the decoded

35

audio intermediate signal 699 is input into a common post processing stage for post processing the decoded audio intermediate signal 699, which is the combined signal output by the combiner 600 so that an output signal of the common pre-processing stage is an expanded version of the combined signal. Thus, the decoded audio signal 799 has an enhanced information content compared to the decoded audio intermediate signal 699. This information

40

expansion is provided by the common post processing stage with the help of pre/post processing parameters which can be transmitted from an encoder to a decoder, or which can be derived from the decoded audio intermediate signal itself. Preferably, however, pre/post

processing parameters are transmitted from an encoder to a decoder, since this procedure allows an improved quality of the decoded audio signal.

Fig. 3c illustrates an audio encoder for encoding an audio input signal 195, which may be equal to the intermediate audio signal 195 of Fig. 3a in accordance with the preferred embodiment of the present invention. The audio input signal 195 is present in a first domain which can, for example, be the time domain but which can also be any other domain such as a frequency domain, an LPC domain, an LPC spectral domain or any other domain. Generally, the conversion from one domain to the other domain is performed by a conversion algorithm such as any of the well-known time/frequency conversion algorithms or frequency/time conversion algorithms.

An alternative transform from the time domain, for example in the LPC domain is the result of LPC filtering a time domain signal which results in an LPC residual signal or excitation signal. Any other filtering operations producing a filtered signal which has an impact on a substantial number of signal samples before the transform can be used as a transform algorithm as the case may be. Therefore, weighting an audio signal using an LPC based weighting filter is a further transform, which generates a signal in the LPC domain. In a time/frequency transform, the modification of a single spectral value will have an impact on all time domain values before the transform. Analogously, a modification of any time domain sample will have an impact on each frequency domain sample. Similarly, a modification of a sample of the excitation signal in an LPC domain situation will have, due to the length of the LPC filter, an impact on a substantial number of samples before the LPC filtering. Similarly, a modification of a sample before an LPC transformation will have an impact on many samples obtained by this LPC transformation due to the inherent memory effect of the LPC filter.

The audio encoder of Fig. 3c includes a first coding branch 400 which generates a first encoded signal. This first encoded signal may be in a fourth domain which is, in the preferred embodiment, the time-spectral domain, i.e., the domain which is obtained when a time domain signal is processed via a time/frequency conversion.

Therefore, the first coding branch 400 for encoding an audio signal uses a first coding algorithm to obtain a first encoded signal, where this first coding algorithm may or may not include a time/frequency conversion algorithm.

The audio encoder furthermore includes a second coding branch 500 for encoding an audio signal. The second coding branch 500 uses a second coding algorithm to obtain a second encoded signal, which is different from the first coding algorithm.

The audio encoder furthermore includes a first switch 200 for switching between the first coding branch 400 and the second coding branch 500 so that for a portion of the audio input signal, either the first encoded signal at the output of block 400 or the second encoded signal at the output of the second encoding branch is included in an encoder output signal. Thus,

when for a certain portion of the audio input signal 195, the first encoded signal in the fourth domain is included in the encoder output signal, the second encoded signal which is either the first processed signal in the second domain or the second processed signal in the third domain is not included in the encoder output signal. This makes sure that this encoder is bit rate efficient. In embodiments, any time portions of the audio signal which are included in two different encoded signals are small compared to a frame length of a frame as will be discussed in connection with Fig. 3e. These small portions are useful for a cross fade from one encoded signal to the other encoded signal in the case of a switch event in order to reduce artifacts that might occur without any cross fade. Therefore, apart from the cross-fade region, each time domain block is represented by an encoded signal of only a single domain.

As illustrated in Fig. 3c, the second coding branch 500 comprises a converter 510 for converting the audio signal in the first domain, i.e., signal 195 into a second domain. Furthermore, the second coding branch 500 comprises a first processing branch 522 for processing an audio signal in the second domain to obtain a first processed signal which is, preferably, also in the second domain so that the first processing branch 522 does not perform a domain change.

The second encoding branch 500 furthermore comprises a second processing branch 523, 524 which converts the audio signal in the second domain into a third domain, which is different from the first domain and which is also different from the second domain and which processes the audio signal in the third domain to obtain a second processed signal at the output of the second processing branch 523, 524.

Furthermore, the second coding branch comprises a second switch 521 for switching between the first processing branch 522 and the second processing branch 523, 524 so that, for a portion of the audio signal input into the second coding branch, either the first processed signal in the second domain or the second processed signal in the third domain is in the second encoded signal.

Fig. 3d illustrates a corresponding decoder for decoding an encoded audio signal generated by the encoder of Fig. 3c. Generally, each block of the first domain audio signal is represented by either a second domain signal, a third domain signal or a fourth domain encoded signal apart from an optional cross fade region which is, preferably, short compared to the length of one frame in order to obtain a system which is as much as possible at the critical sampling limit. The encoded audio signal includes the first coded signal, a second coded signal in a second domain and a third coded signal in a third domain, wherein the first coded signal, the second coded signal and the third coded signal all relate to different time portions of the decoded audio signal and wherein the second domain, the third domain and the first domain for a decoded audio signal are different from each other.

The decoder comprises a first decoding branch for decoding based on the first coding algorithm. The first decoding branch is illustrated at 431, 440 in Fig. 3d and preferably

comprises a frequency/time converter. The first coded signal is preferably in a fourth domain and is converted into the first domain which is the domain for the decoded output signal.

5 The decoder of Fig. 3d furthermore comprises a second decoding branch which comprises several elements. These elements are a first inverse processing branch 531 for inverse processing the second coded signal to obtain a first inverse processed signal in the second domain at the output of block 531. The second decoding branch furthermore comprises a second inverse processing branch 533, 534 for inverse processing a third coded signal to obtain a second inverse processed signal in the second domain, where the second inverse
10 processing branch comprises a converter for converting from the third domain into the second domain.

15 The second decoding branch furthermore comprises a first combiner 532 for combining the first inverse processed signal and the second inverse processed signal to obtain a signal in the second domain, where this combined signal is, at the first time instant, only influenced by the first inverse processed signal and is, at a later time instant, only influenced by the second inverse processed signal.

20 The second decoding branch furthermore comprises a converter 540 for converting the combined signal to the first domain.

Finally, the decoder illustrated in Fig. 3d comprises a second combiner 600 for combining the decoded first signal from block 431, 440 and the converter 540 output signal to obtain a decoded output signal in the first domain. Again, the decoded output signal in the first domain
25 is, at the first time instant, only influenced by the signal output by the converter 540 and is, at a later time instant, only influenced by the first decoded signal output by block 431, 440.

This situation is illustrated, from an encoder perspective, in Fig. 3e. The upper portion in Fig. 3e illustrates in the schematic representation, a first domain audio signal such as a time domain audio signal, where the time index increases from left to right and item 3 might be considered as a stream of audio samples representing the signal 195 in Fig. 3c. Fig. 3e illustrates frames 3a, 3b, 3c, 3d which may be generated by switching between the first encoded signal and the first processed signal and the second processed signal as illustrated at item 4 in Fig. 3e. The first encoded signal, the first processed signal and the second processed
35 signals are all in different domains and in order to make sure that the switch between the different domains does not result in an artifact on the decoder-side, frames 3a, 3b of the time domain signal have an overlapping range which is indicated as a cross fade region, and such a cross fade region is there at frame 3b and 3c. However, no such cross fade region is existing between frame 3d, 3c which means that frame 3d is also represented by a second processed
40 signal, i.e., a signal in the third domain, and there is no domain change between frame 3c and 3d. Therefore, generally, it is preferred not to provide a cross fade region where there is no domain change and to provide a cross fade region, i.e., a portion of the audio signal which is encoded by two subsequent coded/processed signals when there is a domain change, i.e., a

switching action of either of the two switches. Preferably, crossfades are performed for other domain changes.

5 In the embodiment, in which the first encoded signal or the second processed signal has been generated by an MDCT processing having e.g. 50 percents overlap, each time domain sample is included in two subsequent frames. Due to the characteristics of the MDCT, however, this does not result in an overhead, since the MDCT is a critically sampled system. In this context, critically sampled means that the number of spectral values is the same as the number of time domain values. The MDCT is advantageous in that the crossover effect is provided without a
10 specific crossover region so that a crossover from an MDCT block to the next MDCT block is provided without any overhead which would violate the critical sampling requirement.

Preferably, the first coding algorithm in the first coding branch is based on an information sink model, and the second coding algorithm in the second coding branch is based on an
15 information source or an SNR model. An SNR model is a model which is not specifically related to a specific sound generation mechanism but which is one coding mode which can be selected among a plurality of coding modes based e.g. on a closed loop decision. Thus, an SNR model is any available coding model but which does not necessarily have to be related to the physical constitution of the sound generator but which is any parameterized coding model
20 different from the information sink model, which can be selected by a closed loop decision and, specifically, by comparing different SNR results from different models.

As illustrated in Fig. 3c, a controller 300, 525 is provided. This controller may include the functionalities of the decision stage 300 of Fig. 1a and, additionally, may include the
25 functionality of the switch control device 525 in Fig. 1a. Generally, the controller is for controlling the first switch and the second switch in a signal adaptive way. The controller is operative to analyze a signal input into the first switch or output by the first or the second coding branch or signals obtained by encoding and decoding from the first and the second encoding branch with respect to a target function. Alternatively, or additionally, the controller
30 is operative to analyze the signal input into the second switch or output by the first processing branch or the second processing branch or obtained by processing and inverse processing from the first processing branch and the second processing branch, again with respect to a target function.

35 In one embodiment, the first coding branch or the second coding branch comprises an aliasing introducing time/frequency conversion algorithm such as an MDCT or an MDST algorithm, which is different from a straightforward FFT transform, which does not introduce an aliasing effect. Furthermore, one or both branches comprise a quantizer/entropy coder block. Specifically, only the second processing branch of the second coding branch includes the
40 time/frequency converter introducing an aliasing operation and the first processing branch of the second coding branch comprises a quantizer and/or entropy coder and does not introduce any aliasing effects. The aliasing introducing time/frequency converter preferably comprises a windower for applying an analysis window and an MDCT transform algorithm. Specifically,

the windower is operative to apply the window function to subsequent frames in an overlapping way so that a sample of a windowed signal occurs in at least two subsequent windowed frames.

- 5 In one embodiment, the first processing branch comprises an ACELP coder and a second processing branch comprises an MDCT spectral converter and the quantizer for quantizing spectral components to obtain quantized spectral components, where each quantized spectral component is zero or is defined by one quantizer index of the plurality of different possible quantizer indices.

10

Furthermore, it is preferred that the first switch 200 operates in an open loop manner and the second switch operates in a closed loop manner.

- 15 As stated before, both coding branches are operative to encode the audio signal in a block wise manner, in which the first switch or the second switch switches in a block-wise manner so that a switching action takes place, at the minimum, after a block of a predefined number of samples of a signal, the predefined number forming a frame length for the corresponding switch. Thus, the granule for switching by the first switch may be, for example, a block of 2048 or 1028 samples, and the frame length, based on which the first switch 200 is switching
20 may be variable but is, preferably, fixed to such a quite long period.

- Contrary thereto, the block length for the second switch 521, i.e., when the second switch 521 switches from one mode to the other, is substantially smaller than the block length for the first switch. Preferably, both block lengths for the switches are selected such that the longer block
25 length is an integer multiple of the shorter block length. In the preferred embodiment, the block length of the first switch is 2048 or 1024 and the block length of the second switch is 1024 or more preferably, 512 and even more preferably, 256 and even more preferably 128 samples so that, at the maximum, the second switch can switch 16 times when the first switch switches only a single time. A preferred maximum block length ratio, however, is 4:1.

30

- In a further embodiment, the controller 300, 525 is operative to perform a speech music discrimination for the first switch in such a way that a decision to speech is favored with respect to a decision to music. In this embodiment, a decision to speech is taken even when a portion less than 50% of a frame for the first switch is speech and the portion of more than
35 50% of the frame is music.

- Furthermore, the controller is operative to already switch to the speech mode, when a quite small portion of the first frame is speech and, specifically, when a portion of the first frame is speech, which is 50% of the length of the smaller second frame. Thus, a preferred
40 speech/favouring switching decision already switches over to speech even when, for example, only 6% or 12% of a block corresponding to the frame length of the first switch is speech.

This procedure is preferably in order to fully exploit the bit rate saving capability of the first processing branch, which has a voiced speech core in one embodiment and to not lose any quality even for the rest of the large first frame, which is non-speech due to the fact that the second processing branch includes a converter and, therefore, is useful for audio signals which have non-speech signals as well. Preferably, this second processing branch includes an overlapping MDCT, which is critically sampled, and which even at small window sizes provides a highly efficient and aliasing free operation due to the time domain aliasing cancellation processing such as overlap and add on the decoder-side. Furthermore, a large block length for the first encoding branch which is preferably an AAC-like MDCT encoding branch is useful, since non-speech signals are normally quite stationary and a long transform window provides a high frequency resolution and, therefore, high quality and, additionally, provides a bit rate efficiency due to a psycho acoustically controlled quantization module, which can also be applied to the transform based coding mode in the second processing branch of the second coding branch.

Regarding the Fig. 3d decoder illustration, it is preferred that the transmitted signal includes an explicit indicator as side information 4a as illustrated in Fig. 3e. This side information 4a is extracted by a bit stream parser not illustrated in Fig. 3d in order to forward the corresponding first encoded signal, first processed signal or second processed signal to the correct processor such as the first decoding branch, the first inverse processing branch or the second inverse processing branch in Fig. 3d. Therefore, an encoded signal not only has the encoded/processed signals but also includes side information relating to these signals. In other embodiments, however, there can be an implicit signaling which allows a decoder-side bit stream parser to distinguish between the certain signals. Regarding Fig. 3e, it is outlined that the first processed signal or the second processed signal is the output of the second coding branch and, therefore, the second coded signal.

Preferably, the first decoding branch and/or the second inverse processing branch includes an MDCT transform for converting from the spectral domain to the time domain. To this end, an overlap-adder is provided to perform a time domain aliasing cancellation functionality which, at the same time, provides a cross fade effect in order to avoid blocking artifacts. Generally, the first decoding branch converts a signal encoded in the fourth domain into the first domain, while the second inverse processing branch performs a conversion from the third domain to the second domain and the converter subsequently connected to the first combiner provides a conversion from the second domain to the first domain so that, at the input of the combiner 600, only first domain signals are there, which represent, in the Fig. 3d embodiment, the decoded output signal.

Fig. 4a and 4b illustrate two different embodiments, which differ in the positioning of the switch 200. In Fig. 4a, the switch 200 is positioned between an output of the common pre-processing stage 100 and input of the two encoded branches 400, 500. The Fig. 4a embodiment makes sure that the audio signal is input into a single encoding branch only, and the other encoding branch, which is not connected to the output of the common pre-

processing stage does not operate and, therefore, is switched off or is in a sleep mode. This embodiment is preferable in that the non-active encoding branch does not consume power and computational resources which is useful for mobile applications in particular, which are battery-powered and, therefore, have the general limitation of power consumption.

5

On the other hand, however, the Fig. 4b embodiment may be preferable when power consumption is not an issue. In this embodiment, both encoding branches 400, 500 are active all the time, and only the output of the selected encoding branch for a certain time portion and/or a certain frequency portion is forwarded to the bit stream formatter which may be implemented as a bit stream multiplexer 800. Therefore, in the Fig. 4b embodiment, both
10 encoding branches are active all the time, and the output of an encoding branch which is selected by the decision stage 300 is entered into the output bit stream, while the output of the other non-selected encoding branch 400 is discarded, i.e., not entered into the output bit stream, i.e., the encoded audio signal.

15

Preferably, the second encoding rule/decoding rule is an LPC-based coding algorithm. In LPC-based speech coding, a differentiation between quasi-periodic impulse-like excitation signal segments or signal portions, and noise-like excitation signal segments or signal portions, is made. This is performed for very low bit rate LPC vocoders (2.4 kbps) as in Fig
20 7b. However, in medium rate CELP coders, the excitation is obtained for the addition of scaled vectors from an adaptive codebook and a fixed codebook.

Quasi-periodic impulse-like excitation signal segments, i.e., signal segments having a specific pitch are coded with different mechanisms than noise-like excitation signals. While quasi-
25 periodic impulse-like excitation signals are connected to voiced speech, noise-like signals are related to unvoiced speech.

Exemplarily, reference is made to Figs. 5a to 5d. Here, quasi-periodic impulse-like signal segments or signal portions and noise-like signal segments or signal portions are exemplarily
30 discussed. Specifically, a voiced speech as illustrated in Fig. 5a in the time domain and in Fig. 5b in the frequency domain is discussed as an example for a quasi-periodic impulse-like signal portion, and an unvoiced speech segment as an example for a noise-like signal portion is discussed in connection with Figs. 5c and 5d. Speech can generally be classified as voiced, unvoiced, or mixed. Time-and-frequency domain plots for sampled voiced and unvoiced
35 segments are shown in Fig. 5a to 5d. Voiced speech is quasi periodic in the time domain and harmonically structured in the frequency domain, while unvoiced speech is random-like and broadband. The short-time spectrum of voiced speech is characterized by its fine harmonic formant structure. The fine harmonic structure is a consequence of the quasi-periodicity of speech and may be attributed to the vibrating vocal chords. The formant structure (spectral
40 envelope) is due to the interaction of the source and the vocal tracts. The vocal tracts consist of the pharynx and the mouth cavity. The shape of the spectral envelope that "fits" the short time spectrum of voiced speech is associated with the transfer characteristics of the vocal tract and the spectral tilt (6 dB /Octave) due to the glottal pulse. The spectral envelope is characterized

by a set of peaks which are called formants. The formants are the resonant modes of the vocal tract. For the average vocal tract there are three to five formants below 5 kHz. The amplitudes and locations of the first three formants, usually occurring below 3 kHz are quite important both, in speech synthesis and perception. Higher formants are also important for wide band and unvoiced speech representations. The properties of speech are related to the physical speech production system as follows. Voiced speech is produced by exciting the vocal tract with quasi-periodic glottal air pulses generated by the vibrating vocal chords. The frequency of the periodic pulses is referred to as the fundamental frequency or pitch. Unvoiced speech is produced by forcing air through a constriction in the vocal tract. Nasal sounds are due to the acoustic coupling of the nasal tract to the vocal tract, and plosive sounds are produced by abruptly releasing the air pressure which was built up behind the closure in the tract.

Thus, a noise-like portion of the audio signal shows neither any impulse-like time-domain structure nor harmonic frequency-domain structure as illustrated in Fig. 5c and in Fig. 5d, which is different from the quasi-periodic impulse-like portion as illustrated for example in Fig. 5a and in Fig. 5b. As will be outlined later on, however, the differentiation between noise-like portions and quasi-periodic impulse-like portions can also be observed after a LPC for the excitation signal. The LPC is a method which models the vocal tract and extracts from the signal the excitation of the vocal tracts.

Furthermore, quasi-periodic impulse-like portions and noise-like portions can occur in a timely manner, i.e., which means that a portion of the audio signal in time is noisy and another portion of the audio signal in time is quasi-periodic, i.e. tonal. Alternatively, or additionally, the characteristic of a signal can be different in different frequency bands. Thus, the determination, whether the audio signal is noisy or tonal, can also be performed frequency-selective so that a certain frequency band or several certain frequency bands are considered to be noisy and other frequency bands are considered to be tonal. In this case, a certain time portion of the audio signal might include tonal components and noisy components.

Fig. 7a illustrates a linear model of a speech production system. This system assumes a two-stage excitation, i.e., an impulse-train for voiced speech as indicated in Fig. 7c, and a random-noise for unvoiced speech as indicated in Fig. 7d. The vocal tract is modelled as an all-pole filter 70 which processes pulses of Fig. 7c or Fig. 7d, generated by the glottal model 72. Hence, the system of Fig. 7a can be reduced to an all pole-filter model of Fig. 7b having a gain stage 77, a forward path 78, a feedback path 79, and an adding stage 80. In the feedback path 79, there is a prediction filter 81, and the whole source-model synthesis system illustrated in Fig. 7b can be represented using z-domain functions as follows:

$$S(z) = g / (1 - A(z)) \cdot X(z),$$

where g represents the gain, $A(z)$ is the prediction filter as determined by an LP analysis, $X(z)$ is the excitation signal, and $S(z)$ is the synthesis speech output.

Figs. 7c and 7d give a graphical time domain description of voiced and unvoiced speech synthesis using the linear source system model. This system and the excitation parameters in the above equation are unknown and must be determined from a finite set of speech samples. The coefficients of $A(z)$ are obtained using a linear prediction of the input signal and a quantization of the filter coefficients. In a p -th order forward linear predictor, the present sample of the speech sequence is predicted from a linear combination of p passed samples. The predictor coefficients can be determined by well-known algorithms such as the Levinson-Durbin algorithm, or generally an autocorrelation method or a reflection method.

Fig. 7e illustrates a more detailed implementation of the LPC analysis block 510. The audio signal is input into a filter determination block which determines the filter information $A(z)$. This information is output as the short-term prediction information required for a decoder. The short-term prediction information is required by the actual prediction filter 85. In a subtracter 86, a current sample of the audio signal is input and a predicted value for the current sample is subtracted so that for this sample, the prediction error signal is generated at line 84. A sequence of such prediction error signal samples is very schematically illustrated in Fig. 7c or 7d. Therefore, Fig. 7a, 7b can be considered as a kind of a rectified impulse-like signal.

While Fig. 7e illustrates a preferred way to calculate the excitation signal, Fig. 7f illustrates a preferred way to calculate the weighted signal. In contrast to Fig. 7e, the filter 85 is different, when γ is different from 1. A value smaller than 1 is preferred for γ . Furthermore, the block 87 is present, and μ is preferable a number smaller than 1. Generally, the elements in Fig. 7e and 7f can be implemented as in 3GPP TS 26.190 or 3GPP TS 26.290.

Fig. 7g illustrates an inverse processing, which can be applied on the decoder side such as in element 537 of Fig. 2b. Particularly, block 88 generates an unweighted signal from the weighted signal and block 89 calculates an excitation from the unweighted signal. Generally, all signals but the unweighted signal in Fig. 7g are in the LPC domain, but the excitation signal and the weighted signal are different signals in the same domain. Block 89 outputs an excitation signal which can then be used together with the output of block 536. Then, the common inverse LPC transform can be performed in block 540 of Fig. 2b.

Subsequently, an analysis-by-synthesis CELP encoder will be discussed in connection with Fig. 6 in order to illustrate the modifications applied to this algorithm. This CELP encoder is discussed in detail in "Speech Coding: A Tutorial Review", Andreas Spanias, Proceedings of the IEEE, Vol. 82, No. 10, October 1994, pages 1541-1582. The CELP encoder as illustrated in Fig. 6 includes a long-term prediction component 60 and a short-term prediction component 62. Furthermore, a codebook is used which is indicated at 64. A perceptual weighting filter $W(z)$ is implemented at 66, and an error minimization controller is provided at 68. $s(n)$ is the time-domain input signal. After having been perceptually weighted, the weighted signal is input into a subtracter 69, which calculates the error between the weighted synthesis signal at the output of block 66 and the original weighted signal $s_w(n)$. Generally, the short-term

prediction filter coefficients $A(z)$ are calculated by an LP analysis stage and its coefficients are quantized in $\hat{A}(z)$ as indicated in Fig. 7e. The long-term prediction information $A_L(z)$ including the long-term prediction gain g and the vector quantization index, i.e., codebook references are calculated on the prediction error signal at the output of the LPC analysis stage referred as 10a in Fig. 7e. The LTP parameters are the pitch delay and gain. In CELP this is usually implemented as an adaptive codebook containing the past excitation signal (not the residual). The adaptive CB delay and gain are found by minimizing the mean-squared weighted error (closed-loop pitch search).

10 The CELP algorithm encodes then the residual signal obtained after the short-term and long-term predictions using a codebook of for example Gaussian sequences. The ACELP algorithm, where the "A" stands for "Algebraic" has a specific algebraically designed codebook.

15 A codebook may contain more or less vectors where each vector is some samples long. A gain factor g scales the code vector and the gained code is filtered by the long-term prediction synthesis filter and the short-term prediction synthesis filter. The "optimum" code vector is selected such that the perceptually weighted mean square error at the output of the subtracter 69 is minimized. The search process in CELP is done by an analysis-by-synthesis optimization as illustrated in Fig. 6.

20

For specific cases, when a frame is a mixture of unvoiced and voiced speech or when speech over music occurs, a TCX coding can be more appropriate to code the excitation in the LPC domain. The TCX coding processes the weighted signal in the frequency domain without doing any assumption of excitation production. The TCX is then more generic than CELP coding and is not restricted to a voiced or a non-voiced source model of the excitation. TCX is still a source-oriented model coding using a linear predictive filter for modelling the formants of the speech-like signals.

30 In the AMR-WB+-like coding, a selection between different TCX modes and ACELP takes place as known from the AMR-WB+ description. The TCX modes are different in that the length of the block-wise Discrete Fourier Transform is different for different modes and the best mode can be selected by an analysis by synthesis approach or by a direct "feedforward" mode.

35 As discussed in connection with Fig. 2a and 2b, the common pre-processing stage 100 preferably includes a joint multi-channel (surround/joint stereo device) 101 and, additionally, a band width extension stage 102. Correspondingly, the decoder includes a band width extension stage 701 and a subsequently connected joint multichannel stage 702. Preferably, the joint multichannel stage 101 is, with respect to the encoder, connected before the band
40 width extension stage 102, and, on the decoder side, the band width extension stage 701 is connected before the joint multichannel stage 702 with respect to the signal processing direction. Alternatively, however, the common pre-processing stage can include a joint

multichannel stage without the subsequently connected bandwidth extension stage or a bandwidth extension stage without a connected joint multichannel stage.

5 A preferred example for a joint multichannel stage on the encoder side 101a, 101b and on the decoder side 702a and 702b is illustrated in the context of Fig. 8. A number of E original input channels is input into the downmixer 101a so that the downmixer generates a number of K transmitted channels, where the number K is greater than or equal to one and is smaller than or equal E.

10 Preferably, the E input channels are input into a joint multichannel parameter analyzer 101b which generates parametric information. This parametric information is preferably entropy-encoded such as by a difference encoding and subsequent Huffman encoding or, alternatively, subsequent arithmetic encoding. The encoded parametric information output by block 101b is transmitted to a parameter decoder 702b which may be part of item 702 in Fig. 2b. The
15 parameter decoder 702b decodes the transmitted parametric information and forwards the decoded parametric information into the upmixer 702a. The upmixer 702a receives the K transmitted channels and generates a number of L output channels, where the number of L is greater than or equal K and lower than or equal to E.

20 Parametric information may include inter channel level differences, inter channel time differences, inter channel phase differences and/or inter channel coherence measures as is known from the BCC technique or as is known and is described in detail in the MPEG surround standard. The number of transmitted channels may be a single mono channel for ultra-low bit rate applications or may include a compatible stereo application or may include a
25 compatible stereo signal, i.e., two channels. Typically, the number of E input channels may be five or maybe even higher. Alternatively, the number of E input channels may also be E audio objects as it is known in the context of spatial audio object coding (SAOC).

30 In one implementation, the downmixer performs a weighted or unweighted addition of the original E input channels or an addition of the E input audio objects. In case of audio objects as input channels, the joint multichannel parameter analyzer 101b will calculate audio object parameters such as a correlation matrix between the audio objects preferably for each time portion and even more preferably for each frequency band. To this end, the whole frequency range may be divided in at least 10 and preferable 32 or 64 frequency bands.

35 Fig. 9 illustrates a preferred embodiment for the implementation of the bandwidth extension stage 102 in Fig. 2a and the corresponding band width extension stage 701 in Fig. 2b. On the encoder-side, the bandwidth extension block 102 preferably includes a low pass filtering block 102b, a downsampler block, which follows the lowpass, or which is part of the inverse
40 QMF, which acts on only half of the QMF bands, and a high band analyzer 102a. The original audio signal input into the bandwidth extension block 102 is low-pass filtered to generate the low band signal which is then input into the encoding branches and/or the switch. The low pass filter has a cut off frequency which can be in a range of 3kHz to 10kHz. Furthermore, the

bandwidth extension block 102 furthermore includes a high band analyzer for calculating the bandwidth extension parameters such as a spectral envelope parameter information, a noise floor parameter information, an inverse filtering parameter information, further parametric information relating to certain harmonic lines in the high band and additional parameters as discussed in detail in the MPEG-4 standard in the chapter related to spectral band replication.

On the decoder-side, the bandwidth extension block 701 includes a patcher 701a, an adjuster 701b and a combiner 701c. The combiner 701c combines the decoded low band signal and the reconstructed and adjusted high band signal output by the adjuster 701b. The input into the adjuster 701b is provided by a patcher which is operated to derive the high band signal from the low band signal such as by spectral band replication or, generally, by bandwidth extension. The patching performed by the patcher 701a may be a patching performed in a harmonic way or in a non-harmonic way. The signal generated by the patcher 701a is, subsequently, adjusted by the adjuster 701b using the transmitted parametric bandwidth extension information.

As indicated in Fig. 8 and Fig. 9, the described blocks may have a mode control input in a preferred embodiment. This mode control input is derived from the decision stage 300 output signal. In such a preferred embodiment, a characteristic of a corresponding block may be adapted to the decision stage output, i.e., whether, in a preferred embodiment, a decision to speech or a decision to music is made for a certain time portion of the audio signal. Preferably, the mode control only relates to one or more of the functionalities of these blocks but not to all of the functionalities of blocks. For example, the decision may influence only the patcher 701a but may not influence the other blocks in Fig. 9, or may, for example, influence only the joint multichannel parameter analyzer 101b in Fig. 8 but not the other blocks in Fig. 8. This implementation is preferably such that a higher flexibility and higher quality and lower bit rate output signal is obtained by providing flexibility in the common pre-processing stage. On the other hand, however, the usage of algorithms in the common pre-processing stage for both kinds of signals allows to implement an efficient encoding/decoding scheme.

Fig. 10a and Fig. 10b illustrates two different implementations of the decision stage 300. In Fig. 10a, an open loop decision is indicated. Here, the signal analyzer 300a in the decision stage has certain rules in order to decide whether the certain time portion or a certain frequency portion of the input signal has a characteristic which requires that this signal portion is encoded by the first encoding branch 400 or by the second encoding branch 500. To this end, the signal analyzer 300a may analyze the audio input signal into the common pre-processing stage or may analyze the audio signal output by the common pre-processing stage, i.e., the audio intermediate signal or may analyze an intermediate signal within the common pre-processing stage such as the output of the downmix signal which may be a mono signal or which may be a signal having k channels indicated in Fig. 8. On the output-side, the signal analyzer 300a generates the switching decision for controlling the switch 200 on the encoder-side and the corresponding switch 600 or the combiner 600 on the decoder-side.

Although not discussed in detail for the second switch 521, it is to be emphasized that the second switch 521 can be positioned in a similar way as the first switch 200 as discussed in connection with Fig. 4a and Fig. 4b. Thus, an alternative position of switch 521 in Fig. 3c is at
 5 the output of both processing branches 522, 523, 524 so that, both processing branches operate in parallel and only the output of one processing branch is written into a bit stream via a bit stream former which is not illustrated in Fig. 3c.

Furthermore, the second combiner 600 may have a specific cross fading functionality as
 10 discussed in Fig. 4c. Alternatively or additionally, the first combiner 532 might have the same cross fading functionality. Furthermore, both combiners may have the same cross fading functionality or may have different cross fading functionalities or may have no cross fading functionalities at all so that both combiners are switches without any additional cross fading functionality.

15 As discussed before, both switches can be controlled via an open loop decision or a closed loop decision as discussed in connection with Fig. 10a and Fig. 10b, where the controller 300, 525 of Fig. 3c can have different or the same functionalities for both switches.

20 Furthermore, a time warping functionality which is signal-adaptive can exist not only in the first encoding branch or first decoding branch but can also exist in the second processing branch of the second coding branch on the encoder side as well as on the decoder side. Depending on a processed signal, both time warping functionalities can have the same time warping information so that the same time warp is applied to the signals in the first domain
 25 and in the second domain. This saves processing load and might be useful in some instances, in cases where subsequent blocks have a similar time warping time characteristic. In alternative embodiments, however, it is preferred to have independent time warp estimators for the first coding branch and the second processing branch in the second coding branch.

30 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.

In a different embodiment, the switch 200 of Fig. 1a or 2a switches between the two coding
 35 branches 400, 500. In a further embodiment, there can be additional encoding branches such as a third encoding branch or even a fourth encoding branch or even more encoding branches. On the decoder side, the switch 600 of Fig. 1b or 2b switches between the two decoding branches 431, 440 and 531, 532, 533, 534, 540. In a further embodiment, there can be additional decoding branches such as a third decoding branch or even a fourth decoding
 40 branch or even more decoding branches. Similarly, the other switches 521 or 532 may switch between more than two different coding algorithms, when such additional coding/decoding branches are provided.

Fig. 12A illustrates a preferred embodiment of an encoder implementation, and Fig. 12B illustrates a preferred embodiment of the corresponding decoder implementation. In addition to the elements discussed before with respect to corresponding reference numbers, the embodiment of Fig. 12A illustrates a separate psychoacoustic module 1200, and additionally,
 5 illustrates a preferred implementation of the further encoder tools illustrated at block 421 in Fig. 11A. These additional tools are a temporal noise shaping (TNS) tool 1201 and a mid/side coding tool (M/S) 1202. Furthermore, additional functionalities of the elements 421 and 524 are illustrated in block 421/542 as a combined implementation of scaling, noise filling analysis, quantization, arithmetic coding of spectral values.

10 In the corresponding decoder implementation Fig. 12B, additional elements are illustrated, which are an M/S decoding tool 1203 and a TNS-decoder tool 1204. Furthermore, a bass postfilter not illustrated in the preceding figures is indicated at 1205. The transition windowing block 532 corresponds to the element 532 in Fig. 2B, which is illustrated as a switch, but
 15 which performs a kind of a cross fading which can either be an over sampled cross fading or a critically sampled cross fading. The latter one is implemented as an MDCT operation, where two time aliased portions are overlapped and added. This critically sampled transition processing is preferably used where appropriate, since the overall bitrate can be reduced without any loss in quality. The additional transition windowing block 600 corresponds to the
 20 combiner 600 in Fig. 2B, which is again illustrated as a switch, but it is clear that this element performs a kind of cross fading either critically sampled or non-critically sampled in order to avoid blocking artifacts, and specifically switching artifacts, when one block has been processed in the first branch and the other block has been processed in the second branch. When however, the processing in both branches is perfectly matched to its other, then the
 25 cross fading operation can “degrade” to a hard switch, while a cross fading operation is understood to be a “soft” switching between both branches.

The concept in Fig. 12A and 12B permits coding of signals having an arbitrary mix of speech and audio content, and this concept performs comparable to or better than the best coding
 30 technology that might be tailored specifically to coding of either speech or general audio content. The general structure of the encoder and decoder can be described in that there is a common pre-post processing consisting of an MPEG surround (MPEGS) functional unit to handle stereo or multi-channel processing and an enhanced SBR (eSBR) unit, which handles the parametric representation of the higher audio frequencies in the input signal. Then, there
 35 are two branches, one consisting of a modified advanced audio coding (AAC) tool path and the other consisting of a linear prediction coding (LP or LPC domain) based path, which in turn features either a frequency domain representation or a time domain representation of the LPC residual. All transmitted spectra for both, AAC and LPC, are represented in MDCT domain following quantization and arithmetic coding. The time domain representation uses an
 40 ACELP excitation coding scheme. The basic structure is shown in Fig. 12A for the encoder and Fig. 12B for the decoder. The data flow in this diagram is from left to right, top to bottom. The functions of the decoder are to find the description of the quantized audio spectral or time

domain representation in the bitstream payload and decode the quantized values and other reconstruction information.

5 In case of transmitted spectral information the decoder shall reconstruct the quantized spectra, process the reconstructed spectra through whatever tools are active in the bitstream payload in order to arrive at the actual signal spectra as described by the input bitstream payload, and finally convert the frequency domain spectra to the time domain. Following the initial reconstruction and scaling of the spectrum reconstruction, there are optional tools that modify one or more of the spectra in order to provide more efficient coding.

10

In case of a transmitted time domain signal representation, the decoder shall reconstruct the quantized time signal, process the reconstructed time signal through whatever tools are active in the bitstream payload in order to arrive at the actual time domain signal as described by the input bitstream payload.

15

For each of the optional tools that operate on the signal data, the option to “pass through” is retained, and in all cases where the processing is omitted, the spectra or time samples at its input are passed directly through the tool without modification.

20 In places where the bitstream changes its signal representation from time domain to frequency domain representation or from LP domain to non-LP domain or vice versa, the decoder shall facilitate the transition from one domain to the other by means of an appropriate transition overlap-add windowing.

25 eSBR and MPEG-S processing is applied in the same manner to both coding paths after transition handling.

30 The input to the bitstream payload demultiplexer tool is a bitstream payload. The demultiplexer separates the bitstream payload into the parts for each tool, and provides each of the tools with the bitstream payload information related to that tool.

The outputs from the bitstream payload demultiplexer tool are:

- Depending on the core coding type in the current frame either:
 - the quantized and noiselessly coded spectra represented by
 - 35 • scalefactor information
 - arithmetically coded spectral lines
 - or: linear prediction (LP) parameters together with an excitation signal represented by either:
 - 40 • quantized and arithmetically coded spectral lines (transform coded excitation, TCX) or
 - ACELP coded time domain excitation

- The spectral noise filling information (optional)
- The M/S decision information (optional)
- The temporal noise shaping (TNS) information (optional)
- The filterbank control information
- 5 • The time unwarping (TW) control information (optional)
- The enhanced spectral bandwidth replication (eSBR) control information
- The MPEG Surround (MPEGS) control information

10 The scalefactor noiseless decoding tool takes information from the bitstream payload demultiplexer, parses that information, and decodes the Huffman and DPCM coded scalefactors.

The input to the scalefactor noiseless decoding tool is:

- The scalefactor information for the noiselessly coded spectra

15

The output of the scalefactor noiseless decoding tool is:

- The decoded integer representation of the scalefactors:

20 The spectral noiseless decoding tool takes information from the bitstream payload demultiplexer, parses that information, decodes the arithmetically coded data, and reconstructs the quantized spectra. The input to this noiseless decoding tool is:

- The noiselessly coded spectra

The output of this noiseless decoding tool is:

- 25 • The quantized values of the spectra

The inverse quantizer tool takes the quantized values for the spectra, and converts the integer values to the non-scaled, reconstructed spectra. This quantizer is a companding quantizer, whose companding factor depends on the chosen core coding mode.

30

The input to the Inverse Quantizer tool is:

- The quantized values for the spectra

The output of the inverse quantizer tool is:

- 35 • The un-scaled, inversely quantized spectra

The noise filling tool is used to fill spectral gaps in the decoded spectra, which occur when spectral value are quantized to zero e.g. due to a strong restriction on bit demand in the encoder. The use of the noise filling tool is optional.

5 The inputs to the noise filling tool are:

- The un-scaled, inversely quantized spectra
- Noise filling parameters
- The decoded integer representation of the scalefactors

10 The outputs to the noise filling tool are:

- The un-scaled, inversely quantized spectral values for spectral lines which were previously quantized to zero.
- Modified integer representation of the scalefactors

15 The rescaling tool converts the integer representation of the scalefactors to the actual values, and multiplies the un-scaled inversely quantized spectra by the relevant scalefactors.

The inputs to the scalefactors tool are:

- The decoded integer representation of the scalefactors
- 20 • The un-scaled, inversely quantized spectra

The output from the scalefactors tool is:

- The scaled, inversely quantized spectra

For an overview over the M/S tool, please refer to ISO/IEC 14496-3, subpart 4.1.1.2.

25

For an overview over the temporal noise shaping (TNS) tool, please refer to ISO/IEC 14496-3, subpart 4.1.1.2.

30 The filterbank / block switching tool applies the inverse of the frequency mapping that was carried out in the encoder. An inverse modified discrete cosine transform (IMDCT) is used for the filterbank tool. The IMDCT can be configured to support 120, 128, 240, 256, 320, 480, 512, 576, 960, 1024 or 1152 spectral coefficients.

The inputs to the filterbank tool are:

- 35 • The (inversely quantized) spectra
- The filterbank control information

The output(s) from the filterbank tool is (are):

- The time domain reconstructed audio signal(s).

40 The time-warped filterbank / block switching tool replaces the normal filterbank / block switching tool when the time warping mode is enabled. The filterbank is the same (IMDCT) as

for the normal filterbank, additionally the windowed time domain samples are mapped from the warped time domain to the linear time domain by time-varying resampling.

The inputs to the time-warped filterbank tools are:

- 5 • The inversely quantized spectra
- The filterbank control information
- The time-warping control information

The output(s) from the filterbank tool is (are):

- 10 • The linear time domain reconstructed audio signal(s).

The enhanced SBR (eSBR) tool regenerates the highband of the audio signal. It is based on replication of the sequences of harmonics, truncated during encoding. It adjusts the spectral envelope of the generated high-band and applies inverse filtering, and adds noise and sinusoidal components in order to recreate the spectral characteristics of the original signal.

The input to the eSBR tool is:

- The quantized envelope data
- Misc. control data
- 20 • a time domain signal from the AAC core decoder

The output of the eSBR tool is either:

- a time domain signal or
- a QMF-domain representation of a signal, e.g. in case the MPEG Surround tool is used.

The MPEG Surround (MPEGS) tool produces multiple signals from one or more input signals by applying a sophisticated upmix procedure to the input signal(s) controlled by appropriate spatial parameters. In the USAC context MPEGS is used for coding a multichannel signal, by transmitting parametric side information alongside a transmitted downmixed signal.

The input to the MPEGS tool is:

- a downmixed time domain signal or
- a QMF-domain representation of a downmixed signal from the eSBR tool

The output of the MPEGS tool is:

- a multi-channel time domain signal

The Signal Classifier tool analyses the original input signal and generates from it control information which triggers the selection of the different coding modes. The analysis of the input signal is implementation dependent and will try to choose the optimal core coding mode for a given input signal frame. The output of the signal classifier can (optionally) also be used

to influence the behaviour of other tools, for example MPEG Surround, enhanced SBR, time-warped filterbank and others.

The input to the Signal Classifier tool is:

- 5 • the original unmodified input signal
- additional implementation dependent parameters

The output of the Signal Classifier tool is:

- 10 • a control signal to control the selection of the core codec (non-LP filtered frequency domain coding, LP filtered frequency domain or LP filtered time domain coding)

In accordance with the present invention, the time/frequency resolution in block 410 in Fig. 12A and in the converter 523 in Fig. 12A is controlled dependent on the audio signal. The interrelation between window length, transform length, time resolution and frequency resolution is illustrated in Fig. 13A, where it becomes clear that, for a long window length, the time resolution gets low, but the frequency resolution gets high, and for a short window length, the time resolution is high, but the frequency resolution is low.

In the first encoding branch, which is preferably the AAC encoding branch indicated by elements 410, 1201, 1202, 4021 of Fig. 12A, different windows can be used, where the window shape is determined by a signal analyzer which is preferably encoded in the signal classifier block 300, but which can also be a separate module. The encoder selects one of the windows illustrated in Fig. 13B, which have different time/frequency resolutions. The time/frequency resolution of the first long window, the second window, the fourth window, the fifth window and the sixth window are equal to 2,048 sampling values to a transform length of 1,024. The short window illustrated in the third line in Fig. 13B has a time resolution of 256 sampling values corresponding to the window size. This corresponds to a transform length of 128.

Analogously, the last two windows have a window length equal to 2,304, which is a better frequency resolution than the window in the first line but a lower time resolution. The transform length of the windows in the last two lines is equal to 1,152.

In the first encoding branch, different window sequences which are built from the transform windows in the Fig. 13B can be constructed. Although in Fig. 13C only a short sequence is illustrated, while the other "sequences" consist of a single window only, larger sequences consisting of more windows can also be constructed. It is noted that according to Fig. 13B, for the smaller number of coefficients, i.e., 960 instead of 1,024, the time resolution is also lower than for the corresponding higher number of coefficients such as 1024.

Fig. 14A – 14G illustrates different resolutions/window sizes in the second encoding branch. In a preferred embodiment of the present invention, the second encoding branch has a first processing branch which is an ACELP time domain coder 526, and the second processing

branch comprises the filterbank 523. In this branch, a super frame of, for example 2048 samples, is sub-divided into frames of 256 samples. Individual frames of 256 samples can be separately used so that a sequence of four windows, each window covering two frames, can be applied when an MDCT with 50 percents overlap is applied. Then, a high time resolution is used as illustrated in Fig. 14D. Alternatively, when the signal allows longer windows, the sequence as in Fig. 14C can be applied, where a double window size having 1,024 samples for each window (medium windows) is applied, so that one window covers four frames and there is an overlap of 50 percent.

Finally, when the signal is such that a long window can be used, this long window extends over 4,096 samples again with a 50 percent overlap.

In the preferred embodiment, in which there are two branches, where one branch has an ACELP encoder, the position of the ACELP frame indicated by "A" in the super frame also may determine the window size applied for two adjacent TCX frames indicated by "T" in Fig. 14E. Basically, one is interested in using long windows whenever possible. Nevertheless, short windows have to be applied when a single T frame is between two A frames. Medium windows can be applied when there are two adjacent T frames. However, when there are three adjacent T frames, a corresponding larger window might not be efficient due to the additional complexity. Therefore, the third T frame, although not preceded by an A frame can be processed by a short window. When the whole super frame only has T frames then a long window can be applied.

Fig. 14F illustrates several alternatives for windows, where the window size is always 2x the number lg of spectral coefficients due to a preferred 50 percent overlap. However, other overlap percentages for all encoding branches can be applied so that the relation between window size and transform length can also be different from two and even approach one, when no time domain aliasing is applied.

Fig. 14G illustrates rules for constructing a window based on rules given in Fig. 14F. The value ZL illustrates zeroes at the beginning of the window. The value L illustrates a number of window coefficients in an aliasing zone. The values in portion M are "1" values not introducing any aliasing due to an overlap with an adjacent window which has zero values in the portion corresponding to M. The portion M is followed by a right overlap zone R, which is followed by a ZR zone of zeros, which would correspond to a portion M of a subsequent window.

Reference is made to the subsequently attached annex, which describes a preferred and detailed implementation of an inventive audio encoding/decoding scheme, particularly with respect to the decoder-side.

Annex

1. Windows and window sequences

Quantization and coding is done in the frequency domain. For this purpose, the time signal is mapped into the frequency domain in the encoder. The decoder performs the inverse mapping as described in subclause 2. Depending on the signal, the coder may change the time/frequency resolution by using three different windows size: 2304, 2048 and 256. To switch between windows, the transition windows LONG_START_WINDOW, LONG_STOP_WINDOW, START_WINDOW_LPD, STOP_WINDOW_1152, STOP_START_WINDOW and STOP_START_WINDOW_1152 are used. Table 5.11 lists the windows, specifies the corresponding transform length and shows the shape of the windows schematically. Three transform lengths are used: 1152, 1024 (or 960) (referred to as long transform) and 128 (or 120) coefficients (referred to as short transform).

Window sequences are composed of windows in a way that a raw_data_block always contains data representing 1024 (or 960) output samples. The data element **window_sequence** indicates the window sequence that is actually used. Fig. 13C lists how the window sequences are composed of individual windows. Refer to subclause 2 for more detailed information about the transform and the windows.

1.2 Scalefactor bands and grouping

See ISO/IEC 14496-3, subpart 4, subclause 4.5.2.3.4

As explain in ISO/IEC 14496-3, subpart 4, subclause 4.5.2.3.4, the width of the scalefactor bands is built in imitation of the critical bands of the human auditory system. For that reason the number of scalefactor bands in a spectrum and their width depend on the transform length and the sampling frequency. Table 4.110 to Table 4.128, in ISO/IEC 14496-3, subpart 4, section 4.5.4, list the offset to the beginning of each scalefactor band on the transform lengths 1024 (960) and 128 (120) and on the sampling frequencies. The tables originally designed for LONG_WINDOW, LONG_START_WINDOW and LONG_STOP_WINDOW are used also for START_WINDOW_LPD and STOP_START_WINDOW. The offset tables for STOP_WINDOW_1152 and STOP_START_WINDOW_1152 are Table 4 to Table 10.

1.3 Decoding of lpd_channel_stream()

The lpd_channel_stream() bitstream element contains all necessary information to decode one frame of "linear prediction domain" coded signal. It contains the payload for one frame of encoded signal which was coded in the LPC-domain, i.e. including an LPC filtering step. The residual of this filter (so-called "excitation") is then represented either with the help of an ACELP module or in the MDCT transform domain ("transform coded excitation", TCX). To allow close adaptation to the signal characteristics, one frame is broken down in to four smaller units of equal size, each of which is coded either with ACELP or TCX coding scheme.

This process is similar to the coding scheme described in 3GPP TS 26.290. Inherited from this document is a slightly different terminology, where one “superframe” signifies a signal segment of 1024 samples, whereas a “frame” is exactly one fourth of that, i.e. 256 samples. Each one of these frames is further subdivided into four “subframes” of equal length. Please note that this subchapter adopts this terminology

1.4 Definitions, Data Elements

acelp_core_mode This bitfield indicates the exact bit allocation scheme in case ACELP is used as a lpd coding mode.

lpd_mode The bit-field mode defines the coding modes for each of the four frames within one superframe of the lpd_channel_stream() (corresponds to one AAC frame). The coding modes are stored in the array mod[] and can take values from 0 to 3. The mapping from lpd_mode to mod[] can be determined from Table 1 below.

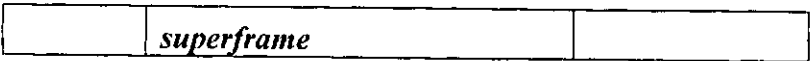
Table 1 – Mapping of coding modes for lpd_channel_stream()

	meaning of bits in bit-field mode					remaining mod[] entries
lpd_mode	bit 4	bit 3	bit 2	bit 1	bit 0	
0..15	0	mod[3]	mod[2]	mod[1]	mod[0]	
16..19	1	0	0	mod[3]	mod[2]	mod[1]=2 mod[0]=2
20..23	1	0	1	mod[1]	mod[0]	mod[3]=2 mod[2]=2
24	1	1	0	0	0	mod[3]=2 mod[2]=2 mod[1]=2 mod[0]=2
25	1	1	0	0	1	mod[3]=3 mod[2]=3 mod[1]=3 mod[0]=3
26..31						reserved

mod[0..3] The values in the array mod[] indicate the respective coding modes in each frame:

Table 2 – Coding modes indicated by mod[]

value of mod[x]	coding mode in frame	bitstream element
0	ACELP	acelp_coding()
1	one frame of TCX	tcx_coding()
2	TCX covering half a superframe	tcx_coding()
3	TCX covering entire	tcx_coding()



- acelp_coding() Syntax element which contains all data to decode one frame of ACELP excitation.
- 5 tcx_coding() Syntax element which contains all data to decode one frame of MDCT based transform coded excitation (TCX).
- first_tcx_flag Flag which indicates if the current processed TCX frame is the first in the superframe.
- lpc_data() Syntax element which contains all data to decode all LPC filter parameter sets required to decode the current superframe.
- 10 first_lpd_flag Flag which indicates whether the current superframe is the first of a sequence of superframes which are coded in LPC domain. This flag can also be determined from the history of the bitstream element core_mode (core_mode0 and core_mode1 in case of a channel_pair_element) according to Table 3.

15 **Table 3 – Definition of first_lpd_flag**

core_mode of previous frame (superframe)	core_mode of current frame (superframe)	first_lpd_flag
0	1	1
1	1	0

last_lpd_mode Indicates the lpd_mode of the previously decoded frame.

20 **1.5 Decoding Process**

- In the lpd_channel_stream the order of decoding is:
- Get acelp_core_mode
- Get lpd_mode and determine from it the content of the helper variable mod[]
- Get acelp_coding or tcx_coding data, depending on the content of the helper variable
- 25 mod[]
- Get lpc_data

1.6 ACELP/TCX coding mode combinations

30 In analogy to [8], section 5.2.2, there are 26 allowed combinations of ACELP or TCX within one superframe of an lpd_channel_stream payload. One of these 26 mode combinations is signaled in the bitstream element lpd_mode. The mapping of lpd_mode to actual coding modes of each frame in a subframe is shown in Table 1 and Table 2.

**Table 4 – scalefactor bands for a window length of 2304 for
STOP_START_1152_WINDOW and STOP_1152_WINDOW at 44.1 and 48 kHz**

fs [kHz]	44.1,48
num_swb_long_window	49
swb	swb_offset_long_window
0	0
1	4
2	8
3	12
4	16
5	20
6	24
7	28
8	32
9	36
10	40
11	48
12	56
13	64
14	72
15	80
16	88
17	96
18	108
19	120
20	132
21	144
22	160
23	176
24	196

swb	swb_offset_long_window
25	216
26	240
27	264
28	292
29	320
30	352
31	384
32	416
33	448
34	480
35	512
36	544
37	576
38	608
39	640
40	672
41	704
42	736
43	768
44	800
45	832
46	864
47	896
48	928
	1152

**Table 5 – scalefactor bands for a window length of 2304 for
STOP_START_1152_WINDOW and STOP_1152_WINDOW at 32 kHz**

fs [kHz]	32
num_swb_long_window	51
swb	swb_offset_long_window
0	0
1	4
2	8
3	12
4	16
5	20
6	24
7	28
8	32
9	36
10	40
11	48
12	56
13	64
14	72
15	80
16	88
17	96
18	108
19	120
20	132
21	144
22	160
23	176
24	196
25	216

swb	swb_offset_long_window
26	240
27	264
28	292
29	320
30	352
31	384
32	416
33	448
34	480
35	512
36	544
37	576
38	608
39	640
40	672
41	704
42	736
43	768
44	800
45	832
46	864
47	896
48	928
49	960
50	992
	1152

Table 6 – scalefactor bands for a window length of of 2304 for STOP_START_1152_WINDOW and STOP_1152_WINDOW at 8 kHz

fs [kHz]	8
num_swb_long_window	40
swb	swb_offset_long_window
0	0
1	12
2	24
3	36
4	48
5	60
6	72
7	84
8	96
9	108
10	120
11	132
12	144
13	156
14	172
15	188
16	204
17	220
18	236
19	252
20	268

swb	Swb_offset_long_window
21	288
22	308
23	328
24	348
25	372
26	396
27	420
28	448
29	476
30	508
31	544
32	580
33	620
34	664
35	712
36	764
37	820
38	880
39	944
	1152

**Table 7 – scalefactor bands for a window length of 2304 for
STOP_START_1152_WINDOW and STOP_1152_WINDOW at 11.025, 12 and 16 kHz**

fs [kHz]	11.025, 12, 16
num_swb_lo ng_window	43
swb	swb_offset _long_win dow
0	0
1	8
2	16
3	24
4	32
5	40
6	48
7	56
8	64
9	72
10	80
11	88
12	100
13	112
14	124
15	136
16	148
17	160
18	172
19	184
20	196
21	212

swb	swb_offset_l ong_window
22	228
23	244
24	260
25	280
26	300
27	320
28	344
29	368
30	396
31	424
32	456
33	492
34	532
35	572
36	616
37	664
38	716
39	772
40	832
41	896
42	960
	1152

**Table 8 – scalefactor bands for a window length of 2304 for
STOP_START_1152_WINDOW and STOP_1152_WINDOW at 22.05 and 24 kHz**

fs [kHz]	22.05 and 24		
num_swb_long_window	47		
swb	swb_offset_long_window	swb	swb_offset_long_window
0	0	24	160
1	4	25	172
2	8	26	188
3	12	27	204
4	16	28	220
5	20	29	240
6	24	30	260
7	28	31	284
8	32	32	308
9	36	33	336
10	40	34	364
11	44	35	396
12	52	36	432
13	60	37	468
14	68	38	508
15	76	39	552
16	84	40	600
17	92	41	652
18	100	42	704
19	108	43	768
20	116	44	832
21	124	45	896
22	136	46	960
23	148		1152

**Table 9 – scalefactor bands for a window length of 2304 for
STOP_START_1152_WINDOW and STOP_1152_WINDOW at 64 kHz**

fs [kHz]	64
num_swb_long_window	47 (46)
swb	swb_offset_long_window
0	0
1	4
2	8
3	12
4	16
5	20
6	24
7	28
8	32
9	36
10	40
11	44
12	48
13	52
14	56
15	64
16	72
17	80
18	88
19	100
20	112
21	124
22	140
23	156

swb	swb_offset_long_window
24	172
25	192
26	216
27	240
28	268
29	304
30	344
31	384
32	424
33	464
34	504
35	544
36	584
37	624
38	664
39	704
40	744
41	784
42	824
43	864
44	904
45	944
46	984
	1152

**Table 10 – scalefactor bands for a window length of 2304 for
STOP_START_1152_WINDOW and STOP_1152_WINDOW at 88.2 and 96 kHz**

fs [kHz]	88.2 and 96		
num_swb_long_window	41		
swb	swb_offset_long_window	swb	swb_offset_long_window
0	0	21	120
1	4	22	132
2	8	23	144
3	12	24	156
4	16	25	172
5	20	26	188
6	24	27	212
7	28	28	240
8	32	29	276
9	36	30	320
10	40	31	384
11	44	32	448
12	48	33	512
13	52	34	576
14	56	35	640
15	64	36	704
16	72	37	768
17	80	38	832
18	88	39	896
19	96	40	960
20	108		1152

1.7 Scale factor band tables references

- 5 For all other scalefactor band tables please refer to ISO/IEC 14496-3, subpart 4, section 4.5.4 Table 4.129 to Table 4.147.

1.8 Quantization

- 10 For quantization of the AAC spectral coefficients in the encoder a non uniform quantizer is used. Therefore the decoder must perform the inverse non uniform quantization after the Huffman decoding of the scalefactors (see subclause 6.3) and the noiseless decoding of the spectral data (see subclause 6.1).
- 15 For the quantization of the TCX spectral coefficients, a uniform quantizer is used. No inverse quantization is needed at the decoder after the noiseless decoding of the spectral data.

2. Filterbank and block switching

2.1 Tool description

20

The time/frequency representation of the signal is mapped onto the time domain by feeding it into the filterbank module. This module consists of an inverse modified discrete cosine transform (IMDCT), and a window and an overlap-add function. In order to adapt the

161

time/frequency resolution of the filterbank to the characteristics of the input signal, a block switching tool is also adopted. N represents the window length, where N is a function of the **window_sequence** (see subclause 1.1). For each channel, the N/2 time-frequency values $X_{i,k}$ are transformed into the N time domain values $x_{i,n}$ via the IMDCT. After applying the window function, for each channel, the first half of the $z_{i,n}$ sequence is added to the second half of the previous block windowed sequence $z_{(i-1),n}$ to reconstruct the output samples for each channel $out_{i,n}$.

2.2 Definitions

window_sequence 2 bit indicating which window sequence (i.e. block size) is used.
window_shape 1 bit indicating which window function is selected.

Fig. 13C shows the eight **window_sequences** (ONLY_LONG_SEQUENCE, LONG_START_SEQUENCE, EIGHT_SHORT_SEQUENCE, LONG_STOP_SEQUENCE, STOP_START_SEQUENCE, STOP_1152_SEQUENCE, LPD_START_SEQUENCE, STOP_START_1152_SEQUENCE).

In the following LPD_SEQUENCE refers to all allowed window/coding mode combinations inside the so called linear prediction domain codec (see section 1.3). In the context of decoding a frequency domain coded frame it is important to know only if a following frame is encoded with the LP domain coding modes, which is represented by an LPD_SEQUENCE. However, the exact structure within the LPD_SEQUENCE is taken care of when decoding the LP domain coded frame.

2.3 Decoding process

2.3.1 IMDCT

The analytical expression of the IMDCT is:

$$x_{i,n} = \frac{2}{N} \sum_{k=0}^{\frac{N}{2}-1} spec[i][k] \cos\left(\frac{2\pi}{N}\left(n+n_0\right)\left(k+\frac{1}{2}\right)\right) \quad \text{for } 0 \leq n < N$$

where:

n = sample index

i = window index

k = spectral coefficient index

N = window length based on the window_sequence value

$n_0 = (N / 2 + 1) / 2$

The synthesis window length N for the inverse transform is a function of the syntax element **window_sequence** and the algorithmic context. It is defined as follows:

Window length 2304:

5

$$N = \begin{cases} 2304, & \text{if STOP_1152_SEQUENCE} \\ 2304, & \text{if STOP_START_1152_SEQUENCE} \end{cases}$$

Window length 2048:

10

$$N = \begin{cases} 2048, & \text{if ONLY_LONG_SEQUENCE} \\ 2048, & \text{if LONG_START_SEQUENCE} \\ 256, & \text{if EIGHT_SHORT_SEQUENCE} \\ 2048, & \text{if LONG_STOP_SEQUENCE} \\ 2048, & \text{if STOP_START_SEQUENCE} \\ 2048, & \text{if LPD_START_SEQUENCE} \end{cases}$$

The meaningful block transitions are as follows:

15 From ONLY_LONG_SEQUENCE to $\begin{cases} \text{ONLY_LONG_SEQUENCE} \\ \text{LONG_START_SEQUENCE} \\ \text{LPD_START_SEQUENCE} \end{cases}$

from LONG_START_SEQUENCE to $\begin{cases} \text{EIGHT_SHORT_SEQUENCE} \\ \text{LONG_STOP_SEQUENCE} \end{cases}$

from LONG_STOP_SEQUENCE to $\begin{cases} \text{ONLY_LONG_SEQUENCE} \\ \text{LONG_START_SEQUENCE} \\ \text{LPD_START_SEQUENCE} \end{cases}$

20

from EIGHT_SHORT_SEQUENCE to $\begin{cases} \text{EIGHT_SHORT_SEQUENCE} \\ \text{LONG_STOP_SEQUENCE} \\ \text{STOP_START_SEQUENCE} \end{cases}$

from LPD_SEQUENCE to $\begin{cases} \text{LPD_SEQUENCE} \\ \text{STOP_1152_SEQUENCE} \\ \text{STOP_START_1152_SEQUENCE} \end{cases}$

25

from STOP_START_SEQUENCE to $\begin{cases} \text{EIGHT_SHORT_SEQUENCE} \\ \text{LONG_STOP_SEQUENCE} \end{cases}$

from LPD_START_SEQUENCE to $\{ \text{LPD_SEQUENCE} \}$

from STOP_1152_SEQUENCE to $\begin{cases} \text{ONLY_LONG_SEQUENCE} \\ \text{LONG_START_SEQUENCE} \end{cases}$

from STOP_START_1152_SEQUENCE to $\begin{cases} \text{EIGHT_SHORT_SEQUENCE} \\ \text{LONG_STOP_SEQUENCE} \end{cases}$

5 2.3.2 Windowing and block switching

Depending on the **window_sequence** and **window_shape** element different transform windows are used. A combination of the window halves described as follows offers all possible window_sequences.

- 10 For **window_shape** == 1, the window coefficients are given by the Kaiser - Bessel derived (KBD) window as follows:

$$W_{KBD_LEFT,N}(n) = \sqrt{\frac{\sum_{p=0}^n [W'(p, \alpha)]}{\sum_{p=0}^{N/2} [W'(p, \alpha)]}} \quad \text{for } 0 \leq n < \frac{N}{2}$$

$$15 \quad W_{KBD_RIGHT,N}(n) = \sqrt{\frac{\sum_{p=0}^{N-n-1} [W'(p, \alpha)]}{\sum_{p=0}^{N/2} [W'(p, \alpha)]}} \quad \text{for } \frac{N}{2} \leq n < N$$

where:

W' , Kaiser - Bessel kernel window function, see also [5], is defined as follows:

$$W'(n, \alpha) = \frac{I_0 \left[\pi \alpha \sqrt{1.0 - \left(\frac{n - N/4}{N/4} \right)^2} \right]}{I_0[\pi \alpha]} \quad \text{for } 0 \leq n \leq \frac{N}{2}$$

$$I_0[x] = \sum_{k=0}^{\infty} \left[\frac{\left(\frac{x}{2} \right)^k}{k!} \right]^2$$

α = kernel window alpha factor, $\alpha = \begin{cases} 4 \text{ for } N = 2048 \text{ (1920)} \\ 6 \text{ for } N = 256 \text{ (240)} \end{cases}$

20

Otherwise, for **window_shape** == 0, a sine window is employed as follows:

$$W_{SIN_LEFT,N}(n) = \sin\left(\frac{\pi}{N} \left(n + \frac{1}{2}\right)\right) \quad \text{for } 0 \leq n < \frac{N}{2}$$

$$W_{SIN_RIGHT,N}(n) = \sin\left(\frac{\pi}{N} \left(n + \frac{1}{2}\right)\right) \quad \text{for } \frac{N}{2} \leq n < N$$

The window length N can be 2048 (1920) or 256 (240) for the KBD and the sine window. In case of STOP_1152_SEQUENCE and STOP_START_1152_SEQUENCE, N can still be 2048 or 256, the window slopes are similar but the flat top regions are longer.

5

Only in the case of LPD_START_SEQUENCE the right part of the window is a sine window of 64 samples.

How to obtain the possible window sequences is explained in the parts a)-h) of this subclause.

10

For all kinds of window_sequences the window_shape of the left half of the first transform window is determined by the window shape of the previous block. The following formula expresses this fact:

$$15 \quad W_{LEFT,N}(n) = \begin{cases} W_{KBD_LEFT,N}(n), & \text{if } window_shape_previous_block == 1 \\ W_{SIN_LEFT,N}(n), & \text{if } window_shape_previous_block == 0 \end{cases}$$

where:

window_shape_previous_block: **window_shape** of the previous block (i-1).

20 For the first raw_data_block() to be decoded the **window_shape** of the left and right half of the window are identical.

25 a) ONLY_LONG_SEQUENCE:

The **window_sequence** == ONLY_LONG_SEQUENCE is equal to one LONG_WINDOW with a total window length N_l of 2048 (1920).

30 For **window_shape** == 1 the window for ONLY_LONG_SEQUENCE is given as follows:

$$W(n) = \begin{cases} W_{LEFT,N_l}(n), & \text{for } 0 \leq n < N_l / 2 \\ W_{KBD_RIGHT,N_l}(n), & \text{for } N_l / 2 \leq n < N_l \end{cases}$$

If **window_shape** == 0 the window for ONLY_LONG_SEQUENCE can be described as follows:

$$35 \quad W(n) = \begin{cases} W_{LEFT,N_l}(n), & \text{for } 0 \leq n < N_l / 2 \\ W_{SIN_RIGHT,N_l}(n), & \text{for } N_l / 2 \leq n < N_l \end{cases}$$

After windowing, the time domain values ($z_{i,n}$) can be expressed as:

$$z_{i,n} = W(n) \cdot x_{i,n}$$

40

b) LONG_START_SEQUENCE:

The LONG_START_SEQUENCE is needed to obtain a correct overlap and add for a block transition from a ONLY_LONG_SEQUENCE to a EIGHT_SHORT_SEQUENCE.

Window length N_l and N_s is set to 2048 (1920) and 256 (240) respectively.

If **window_shape** == 1 the window for LONG_START_SEQUENCE is given as follows:

$$W(n) = \begin{cases} W_{LEFT, N_l}(n), & \text{for } 0 \leq n < N_l / 2 \\ 1.0, & \text{for } N_l / 2 \leq n < \frac{3N_l - N_s}{4} \\ W_{KBD_RIGHT, N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{for } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{for } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

If **window_shape** == 0 the window for LONG_START_SEQUENCE looks like:

$$W(n) = \begin{cases} W_{LEFT, N_l}(n), & \text{for } 0 \leq n < N_l / 2 \\ 1.0, & \text{for } N_l / 2 \leq n < \frac{3N_l - N_s}{4} \\ W_{SIN_RIGHT, N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{for } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{for } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

The windowed time-domain values can be calculated with the formula explained in a).

c) EIGHT_SHORT

The **window_sequence** == EIGHT_SHORT comprises eight overlapped and added SHORT_WINDOWS with a length N_s of 256 (240) each. The total length of the window_sequence together with leading and following zeros is 2048 (1920). Each of the eight short blocks are windowed separately first. The short block number is indexed with the variable $j = 0, \dots, M-1$ ($M = N_l / N_s$).

The **window_shape** of the previous block influences the first of the eight short blocks ($W_0(n)$) only. If **window_shape** == 1 the window functions can be given as follows:

$$W_0(n) = \begin{cases} W_{LEFT, N_s}(n), & \text{for } 0 \leq n < N_s / 2 \\ W_{KBD_RIGHT, N_s}(n), & \text{for } N_s / 2 \leq n < N_s \end{cases}$$

$$W_{1-(M-1)}(n) = \begin{cases} W_{KBD_LEFT, N_s}(n) & \text{for } 0 \leq n < N_s / 2 \\ W_{KBD_RIGHT, N_s}(n), & \text{for } N_s / 2 \leq n < N_s \end{cases}$$

Otherwise, if **window_shape** == 0, the window functions can be described as:

$$W_0(n) = \begin{cases} W_{LEFT, N_s}(n), & \text{for } 0 \leq n < N_s / 2 \\ W_{SIN_RIGHT, N_s}(n), & \text{for } N_s / 2 \leq n < N_s \end{cases}$$

$$w_{1-(M-1)}(n) = \begin{cases} W_{\text{SIN_LEFT}, N_s}(n), & \text{for } 0 \leq n < N_s/2 \\ W_{\text{SIN_RIGHT}, N_s}(n), & \text{for } N_s/2 \leq n < N_s \end{cases}$$

- 5 The overlap and add between the EIGHT_SHORT **window_sequence** resulting in the windowed time domain values $z_{i,n}$ is described as follows:

$$z_{i,n} = \begin{cases} 0, & \text{for } 0 \leq n < \frac{N_l - N_s}{4} \\ x_{0, n - \frac{N_l - N_s}{4}} \cdot W_0\left(n - \frac{N_l - N_s}{4}\right), & \text{for } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ x_{j-1, n - \frac{N_l + (2j-3)N_s}{4}} \cdot W_{j-1}\left(n - \frac{N_l + (2j-3)N_s}{4}\right) + x_{j, n - \frac{N_l + (2j-1)N_s}{4}} \cdot W_j\left(n - \frac{N_l + (2j-1)N_s}{4}\right), & \text{for } 1 \leq j < M, \frac{N_l + (2j-1)N_s}{4} \leq n < \frac{N_l + (2j+1)N_s}{4} \\ x_{M-1, n - \frac{N_l + (2M-3)N_s}{4}} \cdot W_{M-1}\left(n - \frac{N_l + (2M-3)N_s}{4}\right), & \text{for } \frac{N_l + (2M-1)N_s}{4} \leq n < \frac{N_l + (2M+1)N_s}{4} \\ 0, & \text{for } \frac{N_l + (2M+1)N_s}{4} \leq n < N_l \end{cases}$$

10

d) LONG_STOP_SEQUENCE

This window_sequence is needed to switch from a EIGHT_SHORT_SEQUENCE back to a
15 ONLY_LONG_SEQUENCE.

If **window_shape** == 1 the window for LONG_STOP_SEQUENCE is given as follows:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < \frac{N_l - N_s}{4} \\ W_{\text{LEFT}, N_s}\left(n - \frac{N_l - N_s}{4}\right), & \text{for } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ 1.0, & \text{for } \frac{N_l + N_s}{4} \leq n < N_l/2 \\ W_{\text{KBD_RIGHT}, N_l}(n), & \text{for } N_l/2 \leq n < N_l \end{cases}$$

- 20 If **window_shape** == 0 the window for LONG_START_SEQUENCE is determined by:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < \frac{N_l - N_s}{4} \\ W_{\text{LEFT}, N_s}\left(n - \frac{N_l - N_s}{4}\right), & \text{for } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ 1.0, & \text{for } \frac{N_l + N_s}{4} \leq n < N_l/2 \\ W_{\text{SIN_RIGHT}, N_l}(n), & \text{for } N_l/2 \leq n < N_l \end{cases}$$

The windowed time domain values can be calculated with the formula explained in a).

25

e) STOP_START_SEQUENCE:

The STOP_START_SEQUENCE is needed to obtain a correct overlap and add for a block transition from a EIGHT_SHORT_SEQUENCE to a EIGHT_SHORT_SEQUENCE when just a ONLY_LONG_SEQUENCE is needed.

5

Window length N_l and N_s is set to 2048 (1920) and 256 (240) respectively.

If **window_shape** == 1 the window for STOP_START_SEQUENCE is given as follows:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < \frac{N_l - N_s}{4} \\ W_{LEFT, N_s}(n - \frac{N_l - N_s}{4}), & \text{for } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ 1.0, & \text{for } \frac{N_l + N_s}{4} \leq n < \frac{3N_l - N_s}{4} \\ W_{KBD_RIGHT, N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{for } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{for } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

10

If **window_shape** == 0 the window for STOP_START_SEQUENCE looks like:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < \frac{N_l - N_s}{4} \\ W_{LEFT, N_s}(n - \frac{N_l - N_s}{4}), & \text{for } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ 1.0, & \text{for } \frac{N_l + N_s}{4} \leq n < \frac{3N_l - N_s}{4} \\ W_{SIN_RIGHT, N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{for } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{for } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

15 The windowed time-domain values can be calculated with the formula explained in a).

f) LPD_START_SEQUENCE:

20

The LPD_START_SEQUENCE is needed to obtain a correct overlap and add for a block transition from a ONLY_LONG_SEQUENCE to a LPD_SEQUENCE.

Window length N_l and N_s is set to 2048 (1920) and 256 (240) respectively.

25

If **window_shape** == 1 the window for LPD_START_SEQUENCE is given as follows:

$$W(n) = \begin{cases} W_{LEFT_N_l}(n), & \text{for } 0 \leq n < \frac{N_l}{2} \\ 1.0, & \text{for } \frac{N_l}{2} \leq n < \frac{3N_l - N_s}{4} \\ W_{KBD_RIGHT_N_s/2}(n + \frac{N_s}{4} - \frac{3N_l - N_s}{4}), & \text{for } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l}{4} \\ 0.0, & \text{for } \frac{3N_l}{4} \leq n < N_l \end{cases}$$

If **window_shape == 0** the window for LPD_START_SEQUENCE looks like:

$$W(n) = \begin{cases} W_{LEFT_N_l}(n), & \text{for } 0 \leq n < \frac{N_l}{2} \\ 1.0, & \text{for } \frac{N_l}{2} \leq n < \frac{3N_l - N_s}{4} \\ W_{SIN_RIGHT_N_s/2}(n + \frac{N_s}{4} - \frac{3N_l - N_s}{4}), & \text{for } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l}{4} \\ 0.0, & \text{for } \frac{3N_l}{4} \leq n < N_l \end{cases}$$

5

The windowed time-domain values can be calculated with the formula explained in a).

10 g) STOP_1152_SEQUENCE:

The STOP_1152_SEQUENCE is needed to obtain a correct overlap and add for a block transition from a LPD_SEQUENCE to ONLY_LONG_SEQUENCE.

15 Window length N_l and N_s is set to 2048 (1920) and 256 (240) respectively.

If **window_shape == 1** the window for STOP_1152_SEQUENCE is given as follows:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < \frac{N_l}{4} \\ W_{LEFT_N_s}(n - \frac{N_l}{4}), & \text{for } \frac{N_l}{4} \leq n < \frac{N_l + 2N_s}{4} \\ 1.0, & \text{for } \frac{N_l + 2N_s}{4} \leq n < \frac{2N_l + 3N_s}{4} \\ W_{KBD_RIGHT_N_l}(n + \frac{N_l}{2} - \frac{2N_l + 3N_s}{4}), & \text{for } \frac{2N_l + 3N_s}{4} \leq n < N_l + \frac{3N_s}{4} \\ 0.0, & \text{for } N_l + \frac{3N_s}{4} \leq n < N_l + N_s \end{cases}$$

20

If **window_shape == 0** the window for STOP_1152_SEQUENCE looks like:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < \frac{N-l}{4} \\ W_{LEFT.N-s}(n - \frac{N-l}{4}), & \text{for } \frac{N-l}{4} \leq n < \frac{N-l+2N-s}{4} \\ 1.0, & \text{for } \frac{N-l+2N-s}{4} \leq n < \frac{2N-l+3N-s}{4} \\ W_{SIN_RIGHT.N-l}(n + \frac{N-l}{2} - \frac{2N-l+3N-s}{4}), & \text{for } \frac{2N-l+3N-s}{4} \leq n < N-l + \frac{3N-s}{4} \\ 0.0, & \text{for } N-l + \frac{3N-s}{4} \leq n < N-l + N-s \end{cases}$$

The windowed time-domain values can be calculated with the formula explained in a).

5

h) STOP_START_1152_SEQUENCE:

The STOP_START_1152_SEQUENCE is needed to obtain a correct overlap and add for a block transition from a LPD_SEQUENCE to a EIGHT_SHORT_SEQUENCE when just a ONLY_LONG_SEQUENCE is needed.

Window length N_l and N_s is set to 2048 (1920) and 256 (240) respectively.

15

If **window_shape == 1** the window for STOP_START_SEQUENCE is given as follows:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < \frac{N-l}{4} \\ W_{LEFT.N-s}(n - \frac{N-l}{4}), & \text{for } \frac{N-l}{4} \leq n < \frac{N-l+2N-s}{4} \\ 1.0, & \text{for } \frac{N-l+2N-s}{4} \leq n < \frac{3N-l}{4} + \frac{N-s}{2} \\ W_{KBD_RIGHT.N-s}(n + \frac{N-s}{2} - \frac{3N-l}{4} + \frac{N-s}{2}), & \text{for } \frac{3N-l}{4} + \frac{N-s}{2} \leq n < \frac{3N-l}{4} + N-s \\ 0.0, & \text{for } \frac{3N-l}{4} + N-s \leq n < N-l + N-s \end{cases}$$

20

If **window_shape == 0** the window for STOP_START_SEQUENCE looks like:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < \frac{N-l}{4} \\ W_{LEFT, N-s}(n - \frac{N-l}{4}), & \text{for } \frac{N-l}{4} \leq n < \frac{N-l+2N-s}{4} \\ 1.0, & \text{for } \frac{N-l+2N-s}{4} \leq n < \frac{3N-l}{4} + \frac{N-s}{2} \\ W_{SIN_RIGHT, N-s}(n + \frac{N-s}{2} - \frac{3N-l}{4} + \frac{N-s}{2}), & \text{for } \frac{3N-l}{4} + \frac{N-s}{2} \leq n < \frac{3N-l}{4} + N-s \\ 0.0, & \text{for } \frac{3N-l}{4} + N-s \leq n < N-l + N-s \end{cases}$$

The windowed time-domain values can be calculated with the formula explained in a).

5 2.3.3 Overlapping and adding with previous window sequence

Besides the overlap and add within the EIGHT_SHORT **window_sequence** the first (left) part of every **window_sequence** is overlapped and added with the second (right) part of the previous **window_sequence** resulting in the final time domain values $out_{i,n}$. The mathematic expression for this operation can be described as follows.

- 10 In case of ONLY_LONG_SEQUENCE, LONG_START_SEQUENCE, EIGHT_SHORT_SEQUENCE, LONG_STOP_SEQUENCE, STOP_START_SEQUENCE, LPD_START_SEQUENCE:

$$out_{i,n} = z_{i,n} + z_{i-1, n + \frac{N}{2}}; \quad \text{for } 0 \leq n < \frac{N}{2}, \quad N = 2048 \text{ (1920)}$$

15

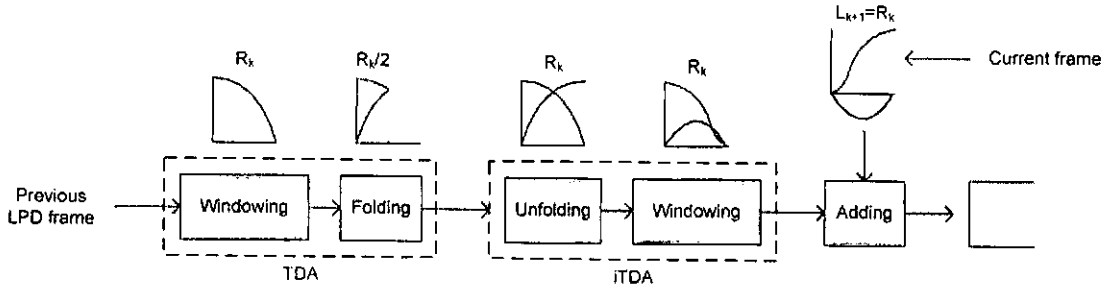
And in case of STOP_1152_SEQUENCE, STOP_START_1152_SEQUENCE:

$$out_{i,n} = z_{i,n} + z_{i-1, n + \frac{N-l}{2} + \frac{3N-s}{4}}; \quad \text{for } 0 \leq n < \frac{N-l}{2}, \quad N-l = 2048, N-s = 256$$

- 20 In case of LPD_START_SEQUENCE, the next sequence is a LPD_SEQUENCE. A SIN or KBD window is apply on the left part of the LPD_SEQUENCE to have a good overlap and add.

$$W_{SIN_LEFT, N}(n) = \sin(\frac{\pi}{N}(n + \frac{1}{2})) \quad \text{for } 0 \leq n < \frac{N}{2} \quad \text{With } N = 128$$

- 25 In case of STOP_1152_SEQUENCE, STOP_START_1152_SEQUENCE the previous sequence is a LPD_SEQUENCE. A TDAC window is apply on the right part of the LPD_SEQUENCE to have a good overlap and add.



3. IMDCT

See subclause 2.3.1

5 3.1 Windowing and block switching

Depending on the **window_shape** element different oversampled transform window prototypes are used, the length of the oversampled windows is

$$N_{os} = 2 \cdot n_long \cdot os_factor_win$$

10

For **window_shape** == 1, the window coefficients are given by the Kaiser - Bessel derived (KBD) window as follows:

$$W_{KBD}\left(n - \frac{N_{os}}{2}\right) = \frac{\sum_{p=0}^{N_{os}-n-1} [W(p, \alpha)]}{\sum_{p=0}^{N_{os}/2} [W(p, \alpha)]} \quad \text{for } \frac{N_{os}}{2} \leq n < N_{os}$$

where:

15 W' , Kaiser - Bessel kernel window function, see also [5], is defined as follows:

$$W'(n, \alpha) = \frac{I_0\left[\pi\alpha\sqrt{1.0 - \left(\frac{n - N_{os}/4}{N_{os}/4}\right)^2}\right]}{I_0[\pi\alpha]} \quad \text{for } 0 \leq n \leq \frac{N_{os}}{2}$$

$$I_0[x] = \sum_{k=0}^{\infty} \left[\frac{\left(\frac{x}{2}\right)^k}{k!} \right]^2$$

α = kernel window alpha factor, $\alpha = 4$

20

Otherwise, for **window_shape** == 0, a sine window is employed as follows:

$$W_{sin}\left(n - \frac{N_{os}}{2}\right) = \sin\left(\frac{\pi}{N_{os}}\left(n + \frac{1}{2}\right)\right) \quad \text{for } \frac{N_{os}}{2} \leq n < N_{os}$$

For all kinds of window_sequences the used prototype for the left window part is the determined by the window shape of the previous block. The following formula expresses this fact:

$$5 \quad \text{left_window_shape}[n] = \begin{cases} W_{KBD}[n] & \text{if } \text{window_shape_previous_block} == 1 \\ W_{SIN}[n] & \text{if } \text{window_shape_previous_block} == 0 \end{cases}$$

Likewise the prototype for the right window shape is determined by the following formula:

$$\text{right_window_shape}[n] = \begin{cases} W_{KBD}[n] & \text{if } \text{window_shape} == 1 \\ W_{SIN}[n] & \text{if } \text{window_shape} == 0 \end{cases}$$

10

Since the transition lengths are already determined, it only has to be differentiated between EIGHT_SHORT_SEQUENCES and all other:

a)EIGHT SHORT SEQUENCE:

15

The following c-code like portion describes the windowing and internal overlap-add of a EIGHT_SHORT_SEQUENCE:

```
20 tw_windowing_short(X[],z[],first_pos,last_pos,warpe_trans_len_left,warped_trans_len_right
, left_window_shape[],right_window_shape[]){
```

```
    offset = n_long - 4*n_short - n_short/2;
```

```
25 tr_scale_l = 0.5*n_long/warped_trans_len_left*os_factor_win;
tr_pos_l = warped_trans_len_left+(first_pos-n_long/2)+0.5)*tr_scale_l;
tr_scale_r = 8*os_factor_win;
tr_pos_r = tr_scale_r/2;
```

```
30 for ( i = 0 ; i < n_short ; i++ ) {
    z[i] = X[0][i];
}
```

```
for(i=0;i<first_pos;i++)
    z[i] = 0.;
```

35

```
for(i=n_long-1-first_pos;i>=first_pos;i--) {
    z[i] *= left_window_shape[floor(tr_pos_l)];
    tr_pos_l += tr_scale_l;
}
```

40

```
for(i=0;i<n_short;i++) {
```

```

    z[offset+i+n_short]=
        X[0][i+n_short]*right_window_shape[floor(tr_pos_r)];
    tr_pos_r += tr_scale_r;
}
5
offset += n_short;

for ( k = 1 ; k < 7 ; k++ ) {
    tr_scale_l = n_short*os_factor_win;
10    tr_pos_l = tr_scale_l/2;
    tr_pos_r = os_factor_win*n_long-tr_pos_l;
    for ( i = 0 ; i < n_short ; i++ ) {
        z[i + offset] += X[k][i]*right_window_shape[floor(tr_pos_r)];
        z[offset + n_short + i] =
15        X[k][n_short + i]*right_window_shape[floor(tr_pos_l)];
        tr_pos_l += tr_scale_l;
        tr_pos_r -= tr_scale_l;
    }
    offset += n_short;
20 }

tr_scale_l = n_short*os_factor_win;
tr_pos_l = tr_scale_l/2;

25 for ( i = n_short - 1 ; i >= 0 ; i-- ) {
    z[i + offset] += X[7][i]*right_window_shape[(int) floor(tr_pos_l)];
    tr_pos_l += tr_scale_l;
}

30 for ( i = 0 ; i < n_short ; i++ ) {
    z[offset + n_short + i] = X[7][n_short + i];
}

tr_scale_r = 0.5*n_long/warpedTransLenRight*os_factor_win;
35 tr_pos_r = 0.5*tr_scale_r+.5;

tr_pos_r = (1.5*n_long-(float)wEnd-0.5+warpedTransLenRight)*tr_scale_r;
for(i=3*n_long-1-last_pos;i<=wEnd;i++) {
    z[i] *= right_window_shape[floor(tr_pos_r)];
40    tr_pos_r += tr_scale_r;
}

for(i=lsat_pos+1;i<2*n_long;i++)

```

$z[i] = 0.;$

b) all others:

5 $\text{tw_windowing_long}(X[], z[], \text{first_pos}, \text{last_pos}, \text{warpe_trans_len_left}, \text{warped_trans_len_right},$
 $\text{left_window_shape}[], \text{right_window_shape}[])\{$

$\text{for}(i=0; i < \text{first_pos}; i++)$

$z[i] = 0.;$

10 $\text{for}(i=\text{last_pos}+1; i < N; i++)$

$z[i] = 0.;$

$\text{tr_scale} = 0.5 * n_long / \text{warped_trans_len_left} * \text{os_factor_win};$

$\text{tr_pos} = (\text{warped_trans_len_left} + \text{first_pos} - N/4 + 0.5) * \text{tr_scale};$

15

$\text{for}(i=N/2-1-\text{first_pos}; i \geq \text{first_pos}; i--) \{$

$z[i] = X[0][i] * \text{left_window_shape}[\text{floor}(\text{tr_pos})];$

$\text{tr_pos} += \text{tr_scale};$

$\}$

20

$\text{tr_scale} = 0.5 * n_long / \text{warped_trans_len_right} * \text{os_factor_win};$

$\text{tr_pos} = (3 * N/4 - \text{last_pos} - 0.5 + \text{warped_trans_len_right}) * \text{tr_scale};$

$\text{for}(i=3 * N/2 - 1 - \text{last_pos}; i \leq \text{last_pos}; i++) \{$

25 $z[i] = X[0][i] * \text{right_window_shape}[\text{floor}(\text{tr_pos})];$

$\text{tr_pos} += \text{tr_scale};$

$\}$

$\}$

30 **4. MDCT based TCX**

4.1 Tool Description

When the **core_mode** is equal to 1 and when one or more of the three TCX modes is selected as the “linear prediction-domain” coding, i.e. one of the 4 array entries of **mod[]** is greater than 0, the MDCT based TCX tool is used. The MDCT based TCX receives the quantized spectral coefficients from the arithmetic decoder. The quantized coefficients are first completed by a comfort noise before applying an inverse MDCT transformation to get a time-domain weighted synthesis which is then fed to the weighting synthesis LPC-filter

40 **4.2 Definitions**

lg Number of quantized spectral coefficients output by the arithmetic decoder

	noise_factor	Noise level quantization index
	noise level	Level of noise injected in reconstructed spectrum
	noise[]	Vector of generated noise
	global_gain	Re-scaling gain quantization index
5	g	Re-scaling gain
	rms	Root mean square of the synthesized time-domain signal, x[],
	x[]	Synthesized time-domain signal

4.3 **Decoding Process**

10

15

The MDCT-based TCX requests from the arithmetic decoder a number of quantized spectral coefficients, lg, which is determined by the mod[] and last_lpd_mode values. These two values also define the window length and shape which will be applied in the inverse MDCT. The window is composed of three parts, a left side overlap of L samples, a middle part of ones of M samples and a right overlap part of R samples. To obtain an MDCT window of length 2*lg, ZL zeros are added on the left and ZR zeros on the right side as indicated in Fig. 14G for Table 3/ Fig. 14F.

Table 3 – Number of Spectral Coefficients as a Function of last_lpd_mode and mod[]

Value of last_lpd_mode	value of mod[x]	Number lg of spectral coefficients	ZL	L	M	R	ZR
0	1	320	160	0	256	128	96
0	2	576	288	0	512	128	224
0	3	1152	512	128	1024	128	512
1..3	1	256	64	128	128	128	64
1..3	2	512	192	128	384	128	192
1..3	3	1024	448	128	896	128	448

20

The MDCT window is given by

$$W(n) = \begin{cases} 0 & \text{for } 0 \leq n < ZL \\ W_{SIN_LEFT,L}(n - ZL) & \text{for } ZL \leq n < ZL + L \\ 1 & \text{for } ZL + L \leq n < ZL + L + M \\ W_{SIN_RIGHT,R}(n - ZL - L - M) & \text{for } ZL + L + M \leq n < ZL + L + M + R \\ 0 & \text{for } ZL + L + M + R \leq n < 2lg \end{cases}$$

The quantized spectral coefficients, $\text{quant}[]$, delivered by the arithmetic decoder are completed by a comfort noise. The level of the injected noise is determined by the decoded noise_factor as follows:

$$\text{noise_level} = 0.0625 * (8 - \text{noise_factor})$$

5

A noise vector, $\text{noise}[]$, is then computed using a random function, $\text{random_sign}()$, delivering randomly the value -1 or +1.

$$\text{noise}[i] = \text{random_sign}() * \text{noise_level};$$

- 10 The $\text{quant}[]$ and $\text{noise}[]$ vectors are combined to form the reconstructed spectral coefficients vector, $r[]$, in a way that the runs of 8 consecutive zeros in $\text{quant}[]$ are replaced by the components of $\text{noise}[]$. A run of 8 non-zeros are detected according to the formula:

$$\begin{cases} r[i] = 1 & \text{for } i \in [0, \lg/6[\\ r[\lg/6 + i] = \sum_{k=0}^7 \lfloor \text{quant}[\lg/6 + 8 \lfloor i/8 \rfloor + k] \rfloor & \text{for } i \in [0, 7 \cdot \lg/6[\end{cases}$$

One obtains the reconstructed spectrum as follows:

15

$$r[i] = \begin{cases} \text{quant}[i] & \text{if } r[i] = 1 \\ \text{noise}[i] & \text{otherwise} \end{cases}$$

Prior to applying the inverse MDCT a spectrum de-shaping is applied according to the following steps:

1. calculate the energy E_m of the 8-dimensional block at index m for each 8-dimensional block of the first quarter of the spectrum
- 20 2. compute the ratio $R_m = \sqrt{E_m/E_l}$, where l is the block index with the maximum value of all E_m
3. if $R_m < 0.1$, then set $R_m = 0.1$
4. if $R_m < R_{m-1}$, then set $R_m = R_{m-1}$

25

Each 8-dimensional block belonging to the first quarter of spectrum are then multiplying by the factor R_m .

- 30 The reconstructed spectrum is fed in an inverse MDCT. The non-windowed output signal, $x[]$, is re-scaled by the gain, g , obtained by an inverse quantization of the decoded global_gain index:

$$g = 10^{\text{global_gain} / 28 / (2 \cdot \text{rms})}$$

Where rms is calculated as:

35

$$\text{rms} = \sqrt{\frac{\sum_{i=\lg/2}^{3 \cdot \lg/2 - 1} x^2[i]}{L + M + R}}$$

The rescaled synthesized time-domain signal is then equal to:

$$x_w[i] = x[i] \cdot g$$

After rescaling the windowing and overlap add is applied.

5

The reconstructed TCX target $x(n)$ is then filtered through the zero-state inverse weighted synthesis filter $\hat{A}(z)(1 - \alpha z^{-1})/(\hat{A}(z/\lambda))$ to find the excitation signal which will be applied to the synthesis filter. Note that the interpolated LP filter per subframe is used in the filtering. Once the excitation is determined, the signal is reconstructed by filtering the excitation through synthesis filter $1/\hat{A}(z)$ and then de-emphasizing by filtering through the filter $1/(1 - 0.68z^{-1})$ as described above.

10

Note that the excitation is also needed to update the ACELP adaptive codebook and allow to switch from TCX to ACELP in a subsequent frame. Note also that the length of the TCX synthesis is given by the TCX frame length (without the overlap): 256, 512 or 1024 samples for the mod[] of 1,2 or 3 respectively.

15

Normative References

[1] ISO/IEC 11172-3:1993, Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s, Part 3: Audio.

20

[2] ITU-T Rec.H.222.0(1995) | ISO/IEC 13818-1:2000, Information technology - Generic coding of moving pictures and associated audio information: - Part 1: Systems.

25

[3] ISO/IEC 13818-3:1998, Information technology - Generic coding of moving pictures and associated audio information: - Part 3: Audio.

30

[4] ISO/IEC 13818-7:2004, Information technology - Generic coding of moving pictures and associated audio information: - Part 7: Advanced Audio Coding (AAC).

30

[5] ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects - Part 1: Systems

35

[6] ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects - Part 3: Audio

[7] ISO/IEC 23003-1:2007, Information technology — MPEG audio technologies — Part 1: MPEG Surround

40

[8] 3GPP TS 26.290 V6.3.0, Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions

[9] 3GPP TS 26.190, Adaptive Multi-Rate -- Wideband (AMR-WB) speech codec; Transcoding functions

[10] 3GPP TS 26.090, Adaptive Multi-Rate (AMR) speech codec; Transcoding functions

5

Definitions

Definitions can be found in ISO/IEC 14496-3, subpart 1, subclause 1.3 (Terms and definitions) and in 3GPP TS 26.290, section 3 (Definitions and abbreviations).

10

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.

15

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.

20

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 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.

25

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.

30

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.

35

Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine readable carrier.

40

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.

- 5 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.

- 10 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.

- 15 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.

- 20 A further embodiment comprises a computer having installed thereon the computer program for performing one of the methods described herein.

- 25 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.

- 30 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.

Claims

1. Audio encoder for encoding an audio signal, comprising:
 - 5 a first coding branch (400) for encoding an audio signal using a first coding algorithm to obtain a first encoded signal, the first coding branch comprising the first converter (410) for converting an input signal into a spectral domain;
 - 10 a second coding branch (500) for encoding an audio signal using a second coding algorithm to obtain a second encoded signal, wherein the first coding algorithm is different from the second coding algorithm, the second coding branch comprising a domain converter for converting an input signal from an input domain into an output domain, and a second converter (523) for converting an input signal into a spectral domain;
 - 15 a switch (200) for switching between the first coding branch and the second coding branch so that, for a portion of the audio input signal, either the first encoded signal or the second encoded signal is in an encoder output signal;
 - 20 a signal analyzer (300, 525) for analyzing the portion of the audio signal to determine, whether the portion of the audio signal is represented as the first encoded signal or the second encoded signal in the encoder output signal, wherein the signal analyzer is furthermore configured for variably determining a respective time/frequency resolution of the first converter and the second converter, when the first encoded signal or the
 - 25 second encoded signal representing the portion of the audio signal is generated; and
 - an output interface (800) for generating an encoder output signal comprising the first encoded signal and the second encoded signal and an information indicating the first encoded signal and the second encoded signal, and an information indicating the
 - 30 time/frequency resolution applied for encoding the first encoded signal and for encoding the second encoded signal.
2. Audio encoder in accordance with claim 1, in which the signal analyzer (300, 525) is
 - 35 configured for classifying the portion of the audio signal as a speech-like audio signal or a music-like audio signal and for performing a transient detection in case of a music signal for determining the time/frequency resolution of the first converter (410) or for performing an analysis-by-synthesis processing for determining the time/frequency resolution of the second converter (523).
- 40 3. Audio encoder in accordance with claim 1 or 2, in which the first converter (410) and the second converter (523) comprise a variable windowed transform processor comprising a window function with a variable window size and a transform function with a variable transform length, and

wherein the signal analyzer (300/525) is configured for controlling, based on the signal analysis, the window size and /or the transform length.

- 5 4. Audio encoder in accordance with one of the preceding claims, in which the second encoder branch comprises a first processing branch (522) for processing an audio signal in the domain determined by the domain converter (510), and a second processing branch (523, 524) comprising the second converter (523),

10 wherein the signal analyzer is configured for sub-dividing the portion of the audio signal into a sequence of sub-portions, and wherein the signal analyzer is configured for determining the time/frequency resolution of the second converter (523) depending on the position of the sub-portion processed by the first processing branch with respect to a sub-portion of the portion processed by the second processing branch.

- 15 5. Audio encoder in accordance with claim 4, in which the first processing branch comprises an ACELP encoder (526),

20 in which the second processing branch comprises an MDCT-TCX processing device (527),

25 in which the signal analyzer (300/525) is configured for setting the time resolution of the second converter to a high value determined by a length of a sub-portion or a comparatively lower value determined by a length of the sub-portion multiplied by an integer value greater than one.

- 30 6. Audio encoder in accordance with one of the preceding claims, in which the signal analyzer (300, 525) is configured for determining a signal classification in a constant raster covering a plurality of equally sized blocks of audio samples, and for sub-dividing a block into a variable number of blocks depending on the audio signal, wherein a length of the sub-block determines the first time/frequency resolution or the second time/frequency resolution.

- 35 7. Apparatus in accordance with one of the preceding claims, in which the signal analyzer (300, 525) is configured for determining the time/frequency resolution to be selected from a plurality of different window lengths, the different window lengths being at least two of 2304, 2048, 256, 1920, 2160, 240 samples, or

40 using the plurality of different transform lengths, the different transform lengths comprising at least two of the group consisting of 1152, 1024, 1080, 960, 128, 120 coefficients per transform block, or

in which the signal analyzer (300, 525) is configured for determining the time/frequency resolution of the second converter as one of a plurality of different window lengths, the plurality of different window lengths being at least two of 640, 1152, 2304, 512, 1024 or 2048 samples, or

5

using a plurality of different transform lengths, the different transform lengths comprising at least two of the group consisting of 320, 576, 1152, 256, 512, 1024 spectral coefficients per transform block.

10 8. Audio encoder in accordance with one of the preceding claims, in which the second coding branch comprises:

a first processing branch (522) for processing an audio signal;

15 a second processing branch, the second processing branch comprising the second converter; and

a further switch (521) for switching between the first processing branch (522) and the second processing branch (523, 524) so that, for a portion of the audio signal input into the second coding branch, either a first processed signal or a second processed signal is in the second encoded signal.

20

9. Method of audio encoding an audio signal, comprising:

25 encoding, in a first coding branch (400), an audio signal using a first coding algorithm to obtain a first encoded signal, the first coding branch comprising the first converter (410) for converting an input signal into a spectral domain;

30 encoding, in a second coding branch (500), an audio signal using a second coding algorithm to obtain a second encoded signal, wherein the first coding algorithm is different from the second coding algorithm, the second coding branch comprising a domain converter for converting an input signal from an input domain into an output domain, and a second converter (523) for converting an input signal into a spectral domain;

35

switching (200) between the first coding branch and the second coding branch so that, for a portion of the audio input signal, either the first encoded signal or the second encoded signal is in an encoder output signal;

40 analyzing (300, 525) the portion of the audio signal to determine, whether the portion of the audio signal is represented as the first encoded signal or the second encoded signal in the encoder output signal,

variably determining a respective time/frequency resolution of the first converter and the second converter, when the first encoded signal or the second encoded signal representing the portion of the audio signal is generated; and

5 generating (800) an encoder output signal comprising the first encoded signal and the second encoded signal and an information indicating the first encoded signal and the second encoded signal, and an information indicating the time/frequency resolution applied for encoding the first encoded signal and for encoding the second encoded signal.

10

10. Audio decoder for decoding an encoded signal, the encoded signal comprising a first encoded signal, a second encoded signal, an indication indicating the first encoded signal and the second encoded signal, and a time/frequency resolution information to be used for decoding the first encoded signal and the second encoded audio signal, comprising:

15

a first decoding branch (431, 440) for decoding the first encoded signal using a first controllable frequency/time converter (440), the first controllable frequency/time converter being configured for being controlled using the time/frequency resolution information for the first encoded signal to obtain a first decoded signal;

20

a second decoding branch for decoding the second encoded signal using a second controllable frequency/time converter (534), the second controllable frequency/time converter (534) being configured for being controlled using the time/frequency resolution information for the second encoded signal;

25

a controller (990) for controlling the first frequency/time converter (440) and the second frequency/time converter (534) using the time/frequency resolution information;

30

a domain converter (540) for generating a synthesis signal using the second decoded signal; and

a combiner (604) for combining the first decoded signal and the synthesis signal to obtain a decoded audio signal.

35

11. Audio decoder in accordance with claim 10, in which the controller (990) is configured for controlling the first frequency/time converter (440) and the second frequency/time converter (534) so that,

40

for the first frequency/time converter (440) the time/frequency resolution is selected from a plurality of different window lengths, the different window lengths being at least two of 2304, 2048, 256, 1920, 2160, 240 samples, or

is selected from a plurality of different transform lengths, the different transform lengths comprising at least two of the group consisting of 1152, 1024, 1080, 960, 128, 120 coefficients per transform block, or

5

for the second frequency/time converter (534) the time/frequency resolution is selected as one of a plurality of different window lengths, the plurality of different window lengths being at least two of 640, 1152, 2304, 512, 1024 or 2048 samples, or

10

is selected from a plurality of different transform lengths, the different transform lengths comprising at least two of the group consisting of 320, 576, 1152, 256, 512, 1024 spectral coefficients per transform block.

15

12. Audio decoder in accordance with claim 10 or 11, in which the second decoding branch comprises a first inverse processing branch (531) for inverse processing a first processed signal being additionally included in the encoded signal to obtain a first inverse processed signal;

20

wherein the second controllable frequency/time converter (534) is located in a second inverse processing branch configured for inverse processing the second encoded signal in a domain identical to the domain of the first inverse processed signal to obtain a second inverse processed signal;

25

a further combiner (532) for combining the first inverse processed signal and the second inverse processed signal to obtain a combined signal; and

wherein the combined signal is input into the combiner (600).

30

13. Audio decoder in accordance with any one of claims 10 - 12, in which the first frequency/time converter (440) and the second frequency/time converter are time domain aliasing cancellation converters having an overlap/add unit (440c) for canceling a time-domain aliasing included in the first encoded signal and the second encoded signal.

35

14. Audio decoder in accordance with any one of claims 10 - 13, in which the encoded signal comprises coding mode information identifying, whether an encoded signal is the first encoded signal and the second encoded signal, and

40

wherein the decoder further comprises an input interface (900) for interpreting the coding mode information to determine, whether the encoded signal is to be fed either into the first decoding branch or into the second decoding branch.

15. Audio decoder in accordance with one of the preceding claims, in which the first encoded signal is arithmetically encoded, and wherein the first coding branch comprises an arithmetic decoder.

5 16. Audio decoder in accordance with one of the preceding claims, in which the first coding branch comprises a dequantizer having a non-uniform dequantization characteristic for canceling a result of a non-uniform quantization applied when generating the first encoded signal,

10 wherein the second coding branch comprises a dequantizer using the different dequantization characteristic, where the second coding branch does not comprise a dequantizer.

15 17. Audio decoder in accordance with one of the preceding claims, in which the controller (990) is configured for controlling the first frequency/time converter and the second frequency/time converter by applying, for each converter, a discrete frequency/time resolution of a number of possible different discrete frequency/time resolutions, the number of possible different frequency/time resolutions being higher for the second converter compared to the number of possible different frequency/time resolutions for the first converter.

20 18. Audio decoder in accordance with any one claims 10 - 17, in which the domain converter is an LPC synthesis processor (544) generating the synthesis signal using a PC filter information, the LPC filter information being included in the encoded signal.

25 19. Method of audio decoding an encoded signal, the encoded signal comprising a first encoded signal, a second encoded signal, an indication indicating the first encoded signal and the second encoded signal, and a time/frequency resolution information to be used for decoding the first encoded signal and the second encoded audio signal, comprising:

30 decoding, by a first decoding branch (431, 440), the first encoded signal using a first controllable frequency/time converter (440), the first controllable frequency/time converter being configured for being controlled using the time/frequency resolution information for the first encoded signal to obtain a first decoded signal;

35 decoding, by a second decoding branch, the second encoded signal using a second controllable frequency/time converter (534), the second controllable frequency/time converter (534) being configured for being controlled using the time/frequency resolution information for the second encoded signal;

40 controlling (990) the first frequency/time converter (440) and the second frequency/time converter (534) using the time/frequency resolution information;

generating (540), by a domain converter, a synthesis signal using the second decoded signal; and

5 combining (604) the first decoded signal and the synthesis signal to obtain a decoded audio signal.

20. Encoded audio signal comprising:

10 a first encoded signal;

a second encoded signal, wherein a portion of an audio signal is either represented by the first encoded signal or the second encoded signal;

15 an indication indicating the first encoded signal and the second encoded signal;

an indication of a first time/frequency resolution information to be used for decoding the first encoded signal, and

20 an indication of a second time/frequency resolution information to be used for decoding the second encoded signal.

21. Computer program for performing, when running on a processor, a method of claim 9 or claim 19.

25

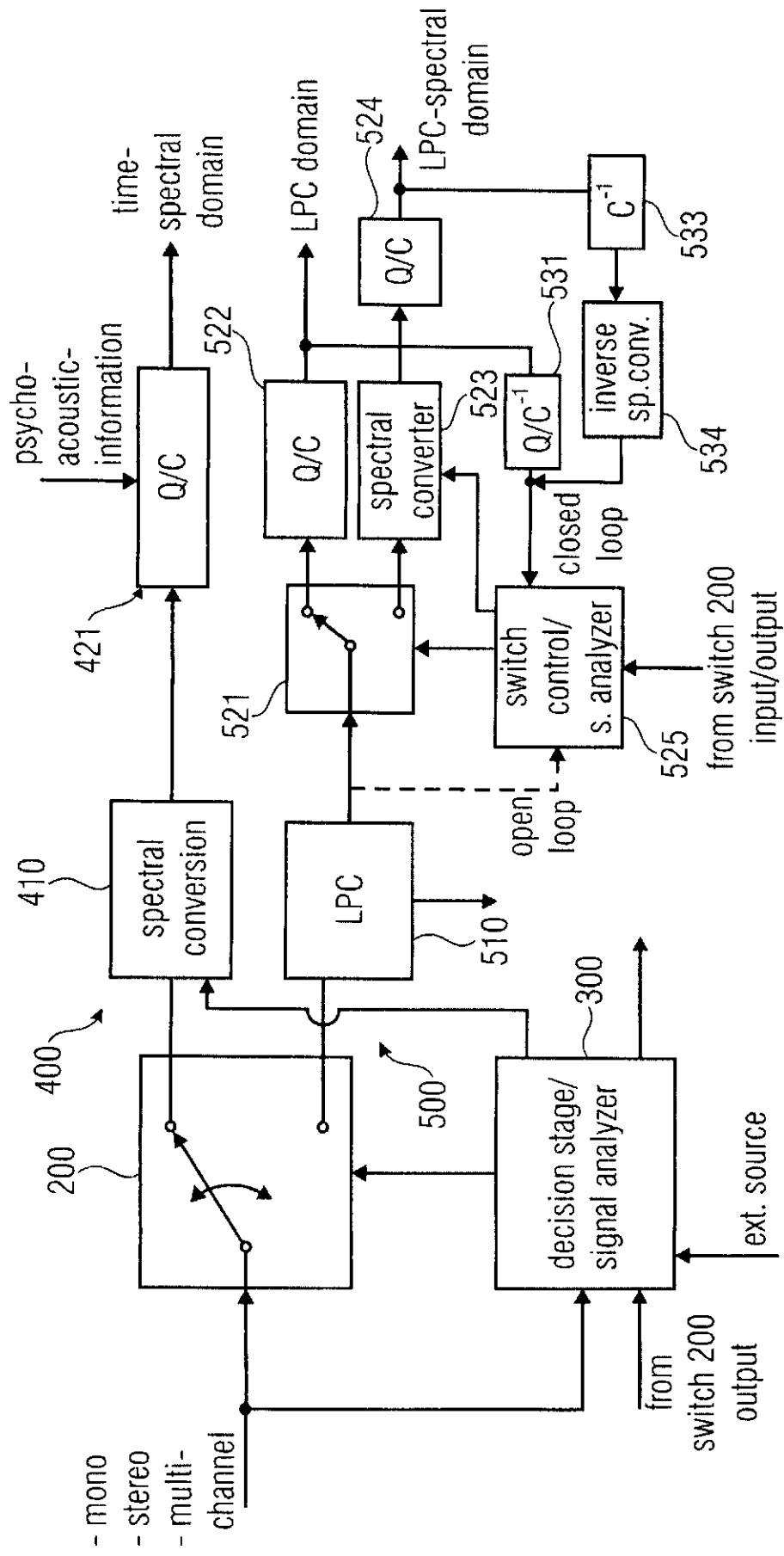
Multi-Resolution Switched Audio Encoding/Decoding Scheme

Abstract

5

An audio encoder for encoding an audio signal comprises a first coding branch (400), the first coding branch comprising a first converter (410) for converting a signal from a time domain into a frequency domain. Furthermore, the audio encoder comprises a second coding branch (500) comprising a second time/frequency converter (523). Additionally, a signal analyzer (300/525) for analyzing the audio signal is provided. The signal analyzer, on the hand, determines whether an audio portion is effective in the encoder output signal as a first encoded signal from the first encoding branch or as a second encoded signal from a second encoding branch. On the other hand, the signal analyzer determines a time/frequency resolution to be applied by the converters (410, 523) when generating the encoded signals. An output interface includes, in addition to the first encoded signal and the second encoded signal, a resolution information identifying the resolution used by the first time/frequency converter and used by the second time/frequency converter.

20



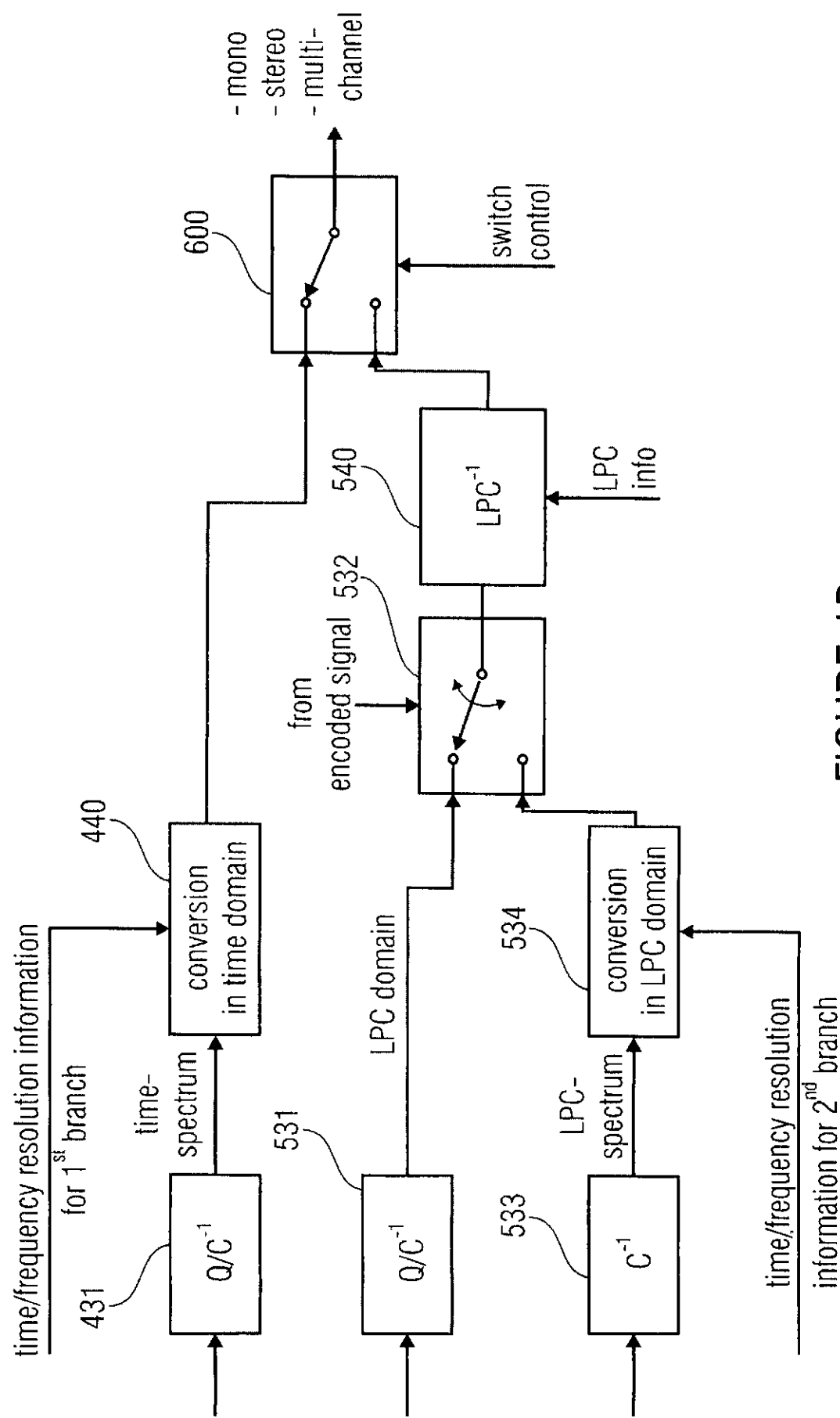


FIGURE 1B
(DECODER)

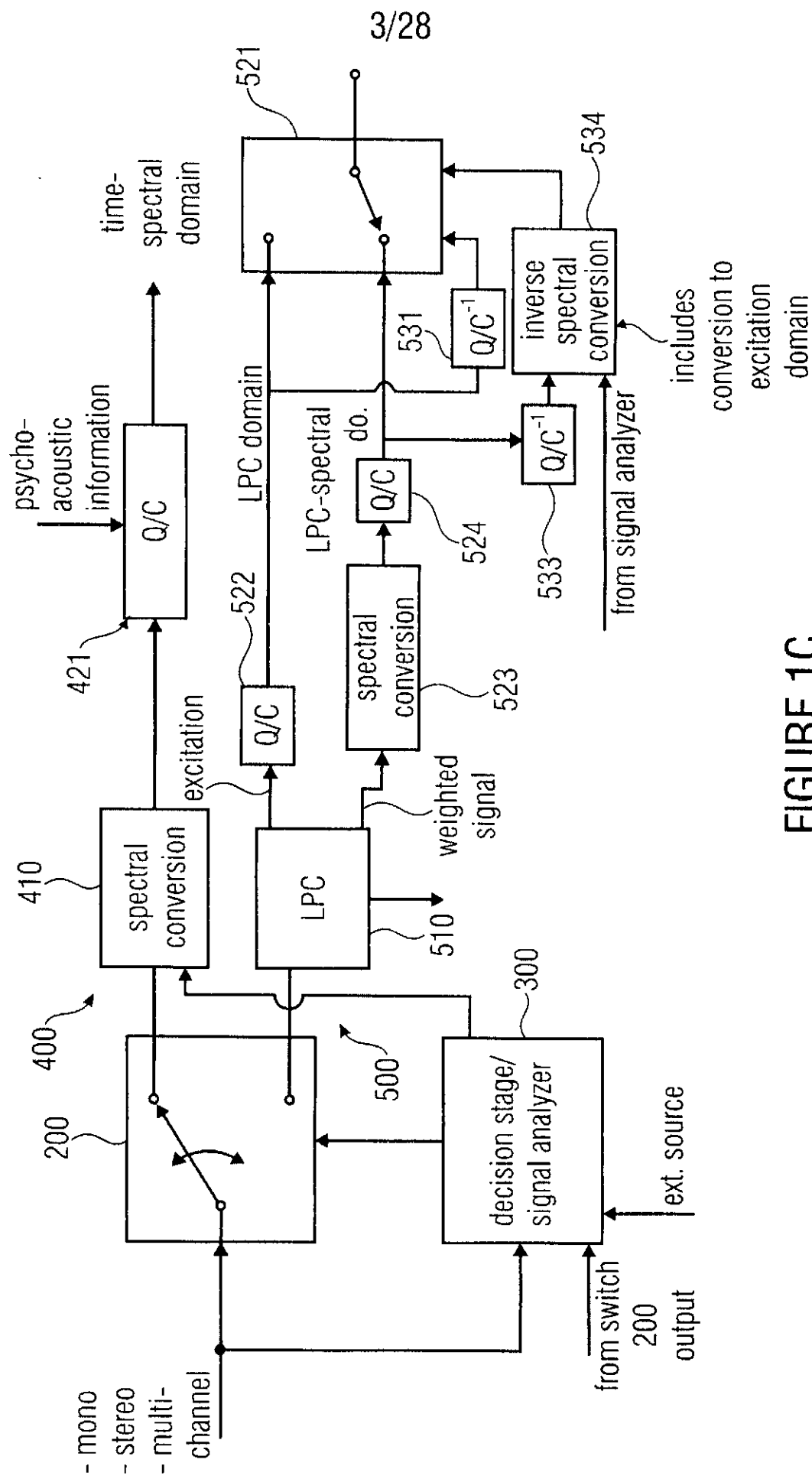


FIGURE 1C
(ENCODER)

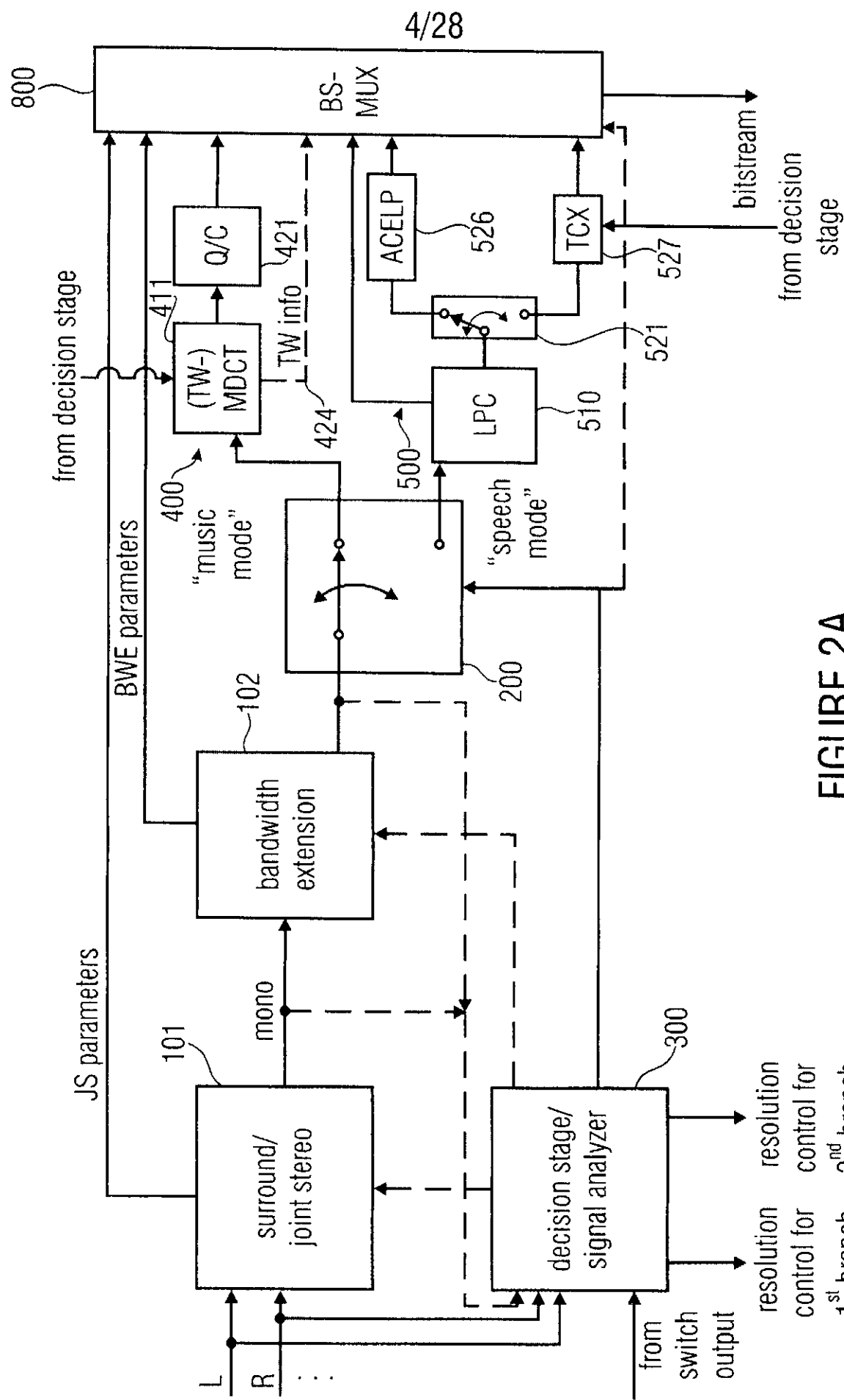


FIGURE 2A
(ENCODER)

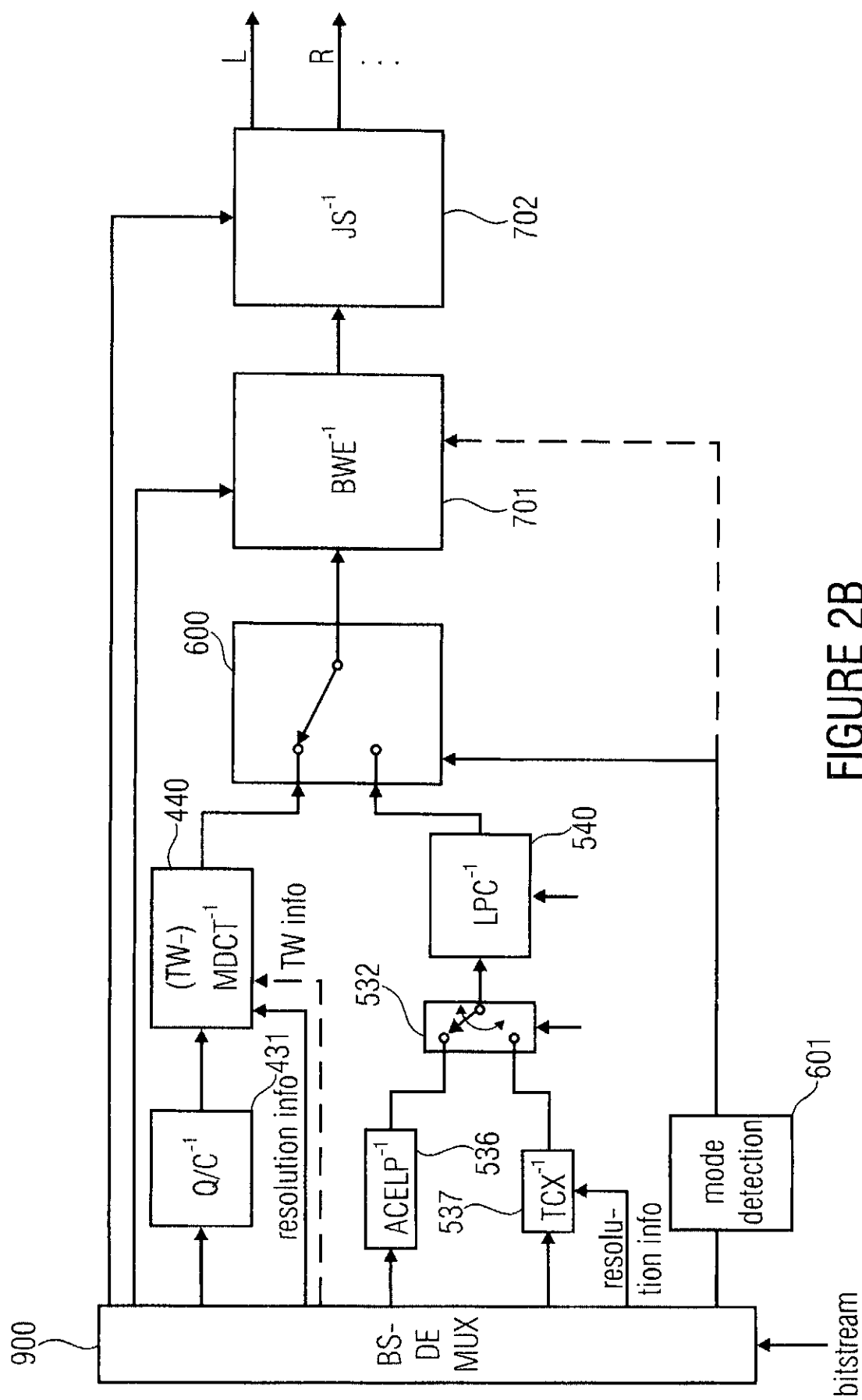


FIGURE 2B
(DECODER)

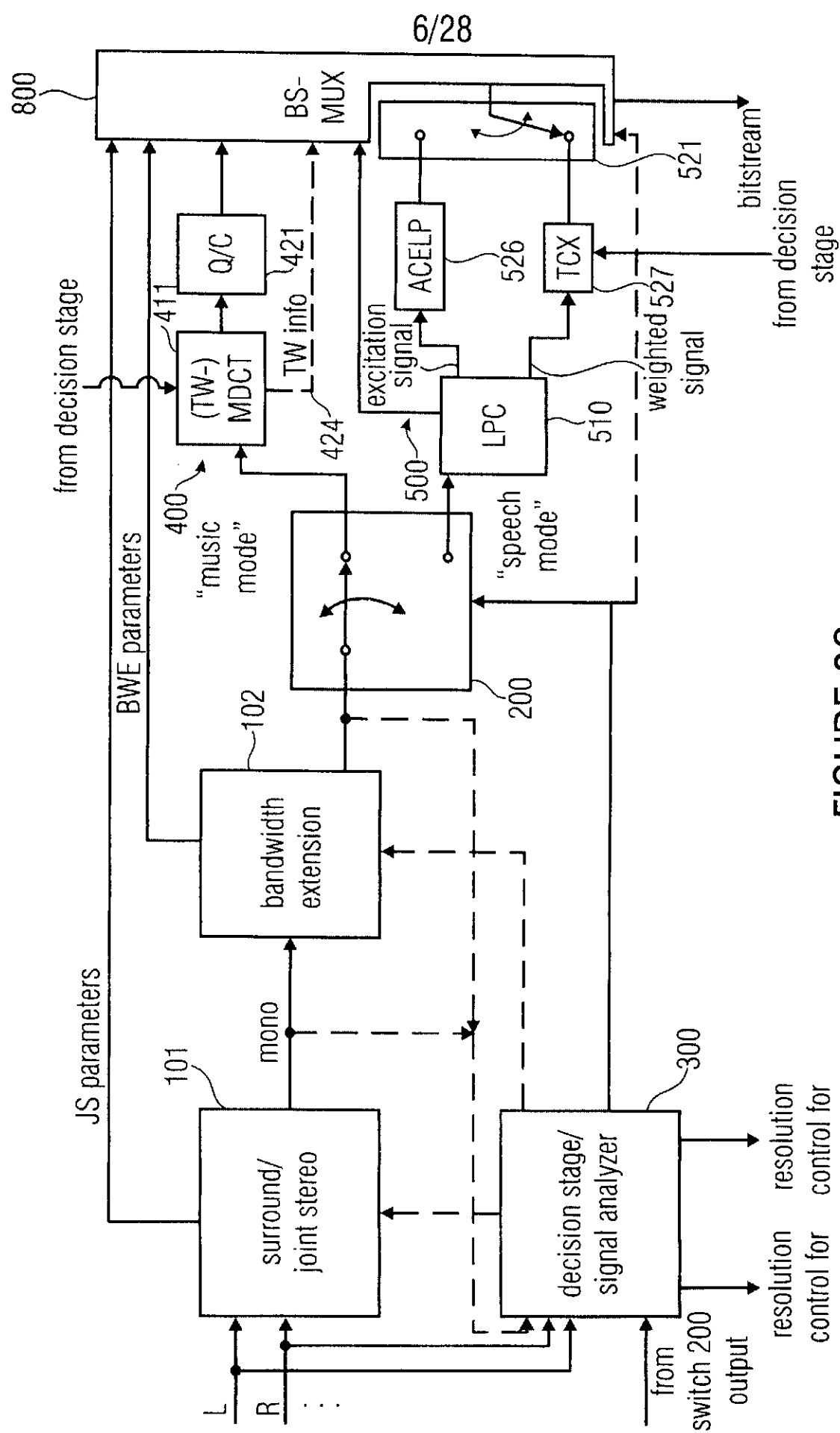


FIGURE 2C
(ENCODER)

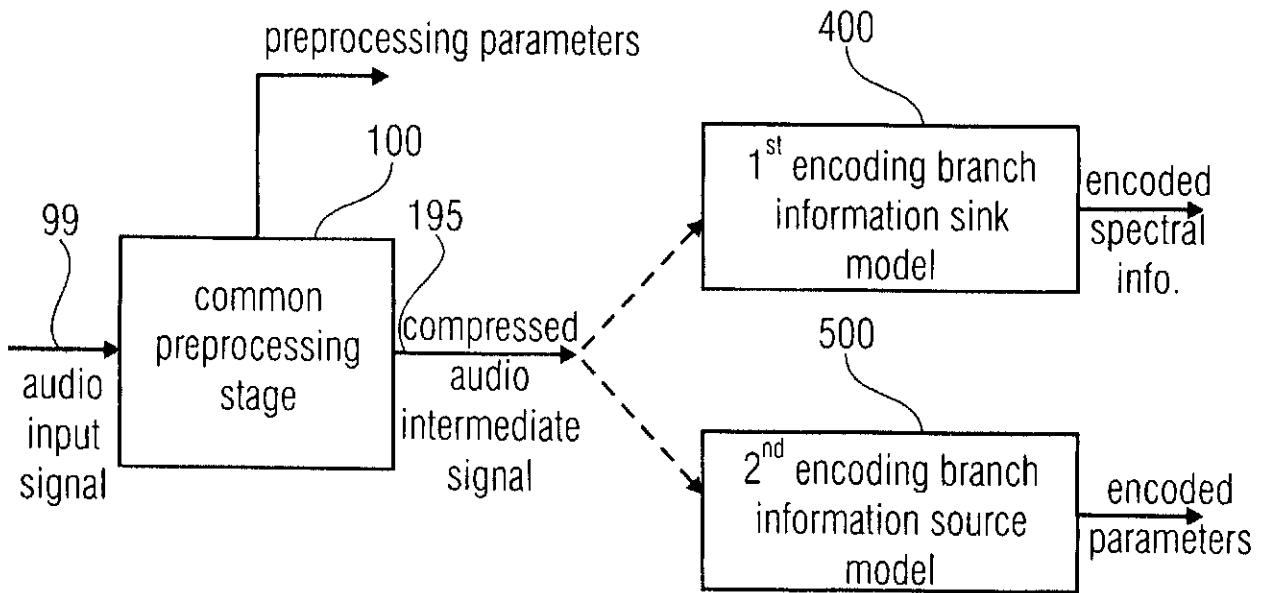


FIGURE 3A

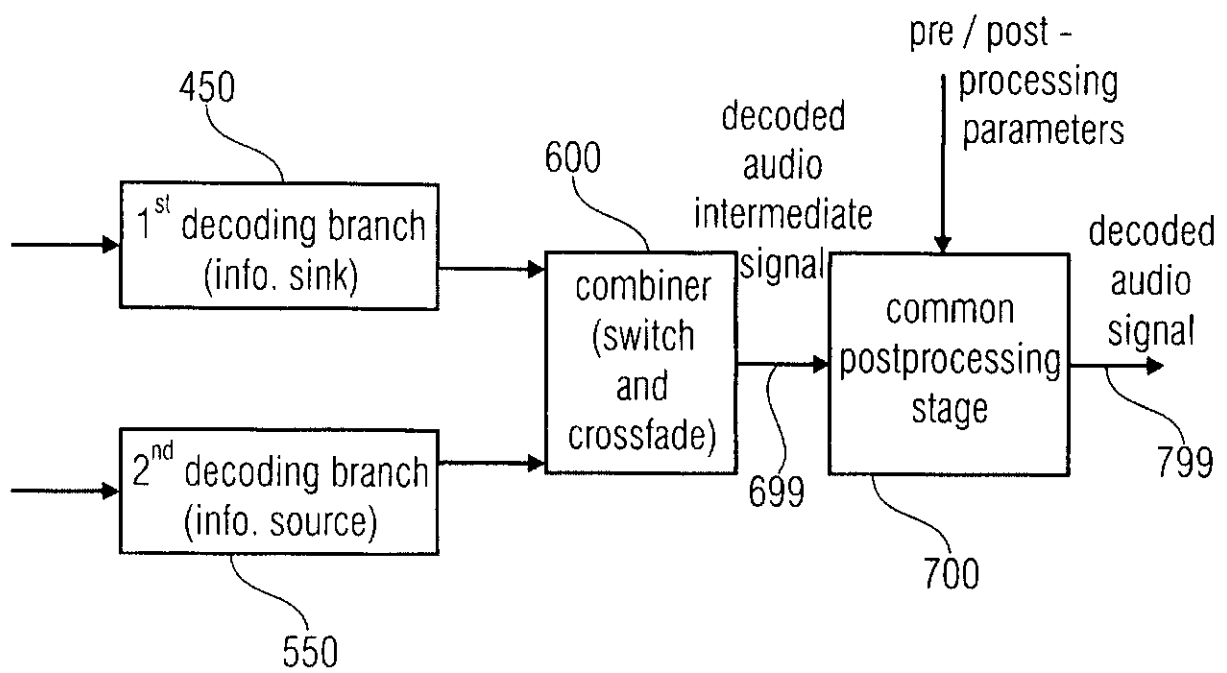
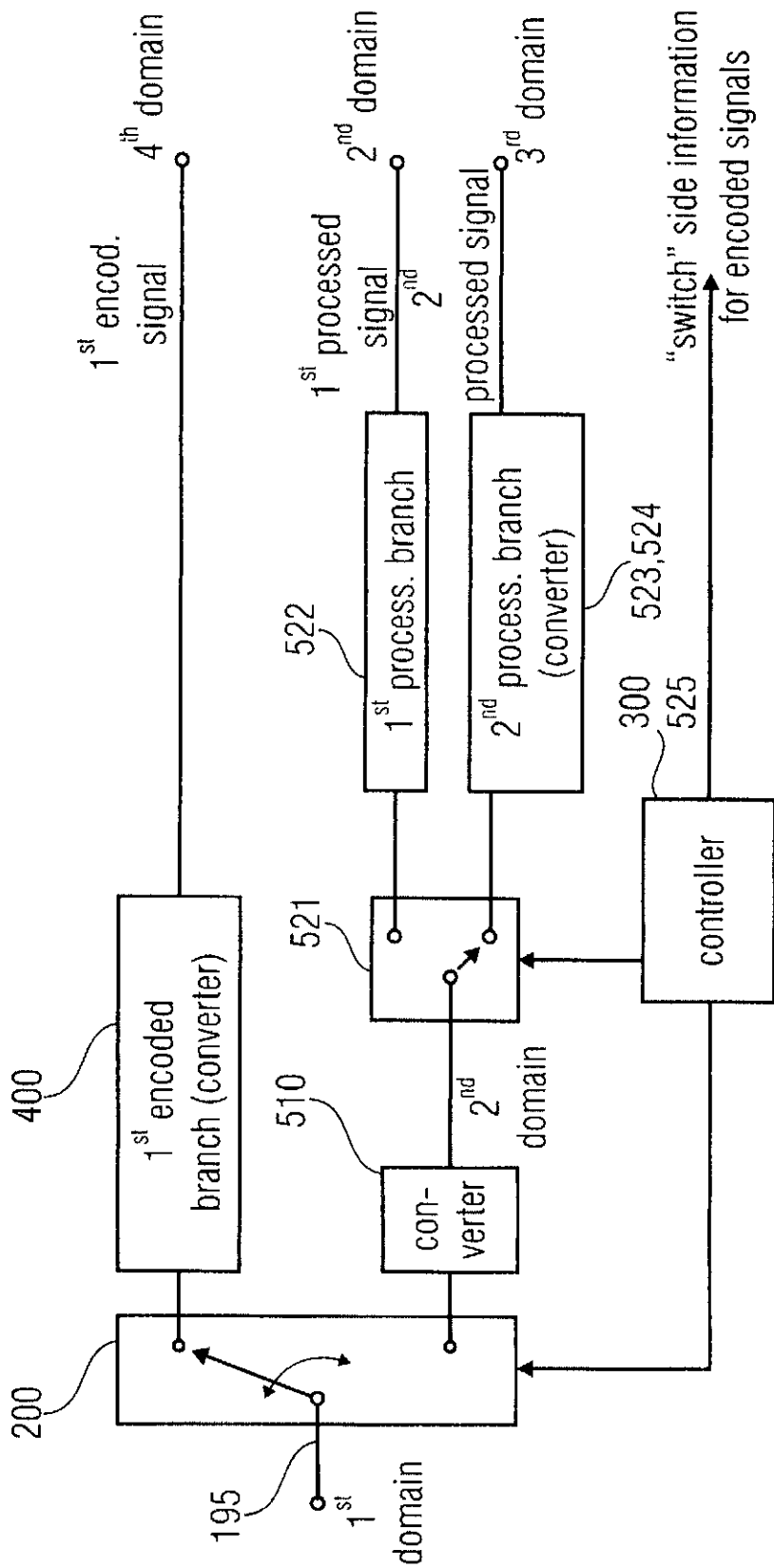


FIGURE 3B



- each block of the 1st domain audio signal is represented by either a 2nd domain, a 3rd domain or a 4th domain encoded signal, apart from an optional crossover region

FIGURE 3C

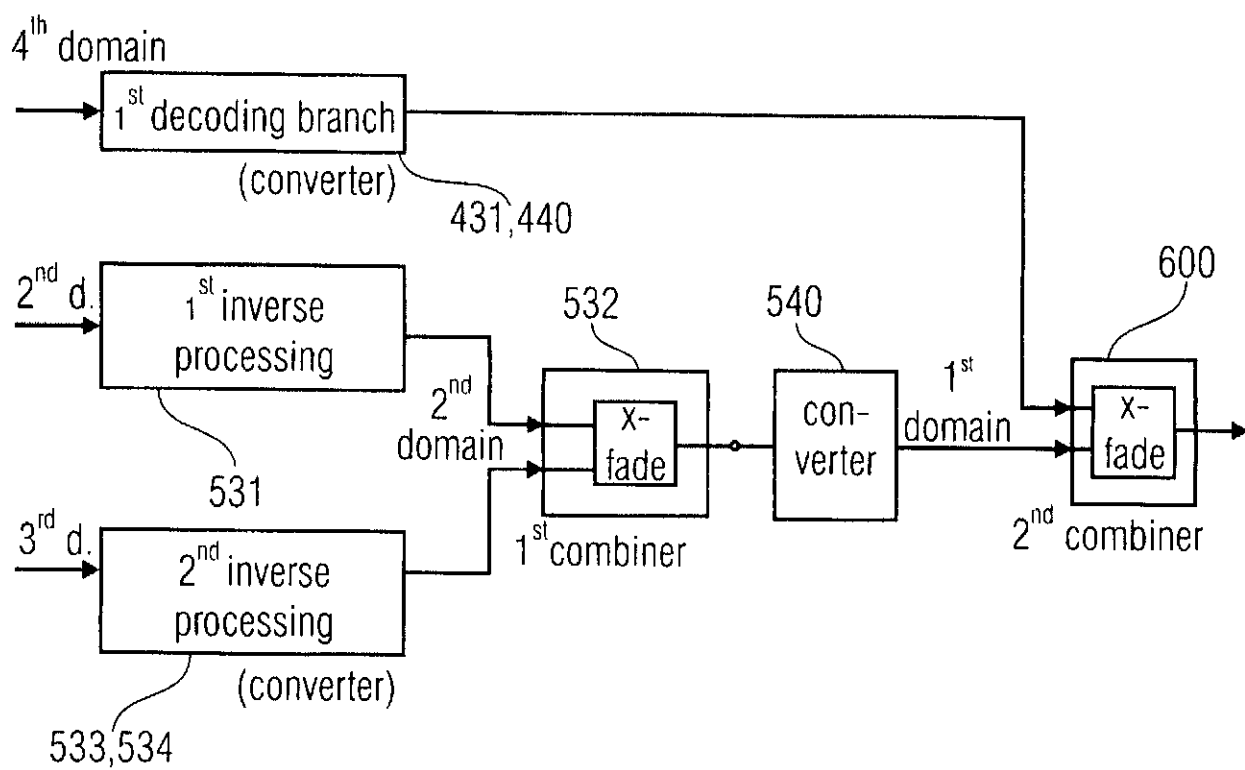


FIGURE 3D

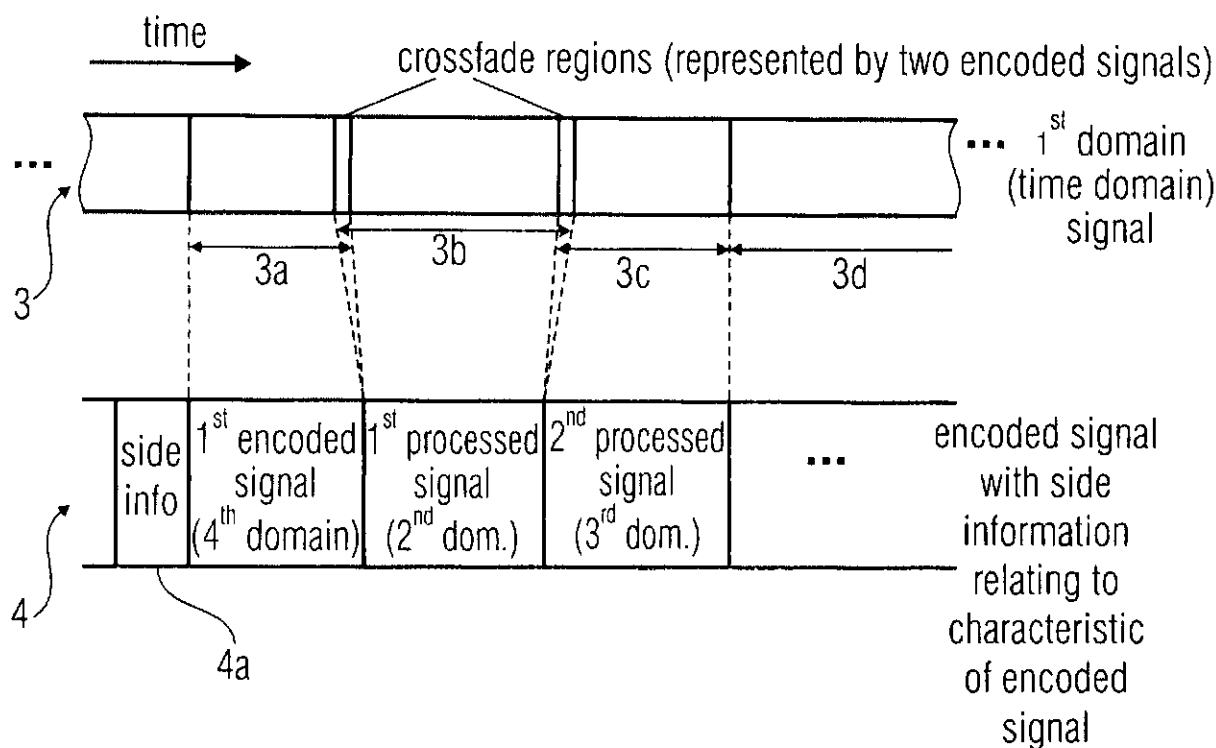


FIGURE 3E

10/28

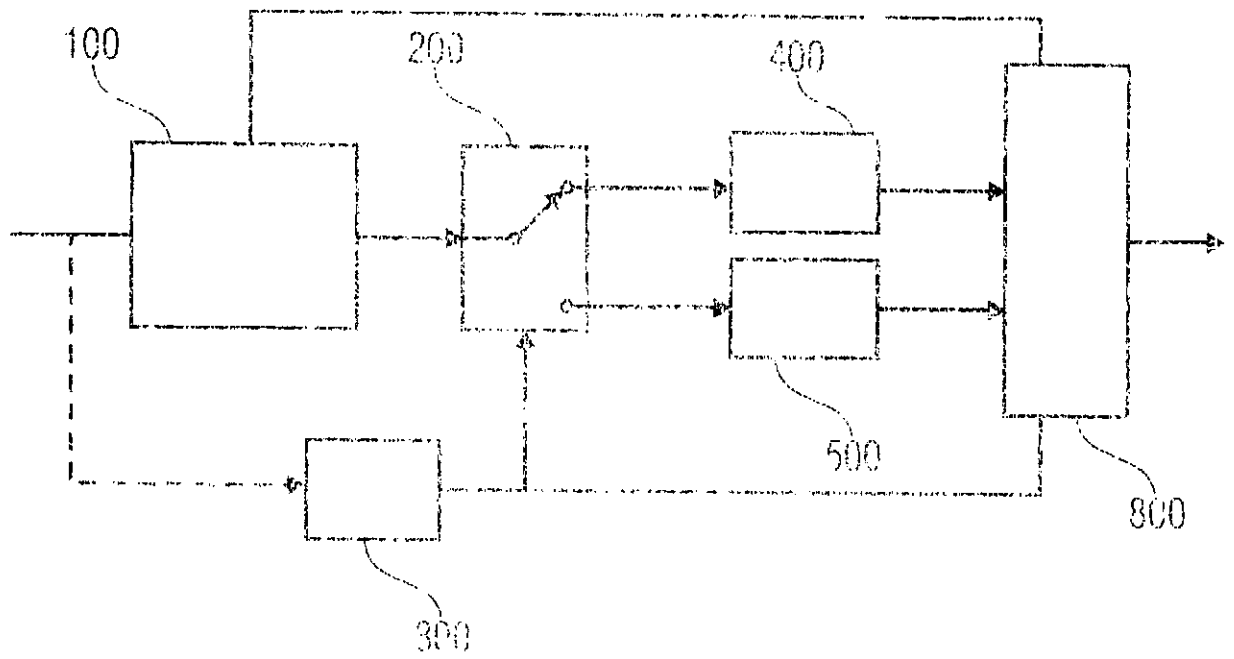


FIGURE 4A

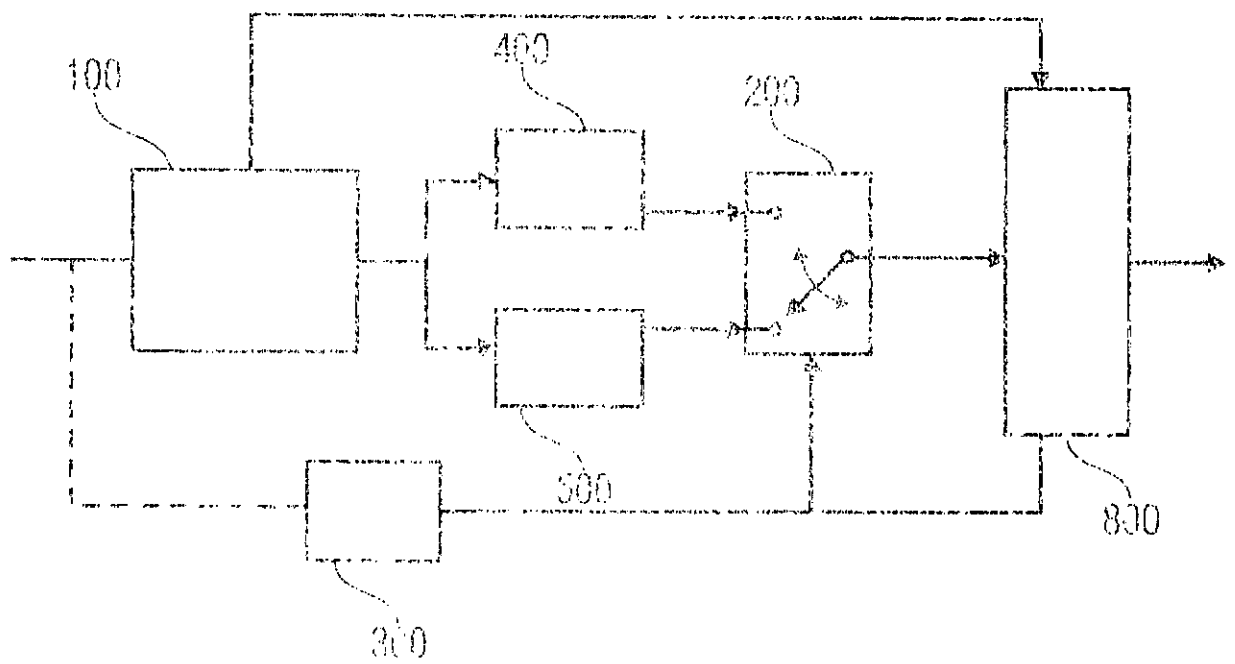


FIGURE 4B

impulse-like signal segment (e.g. voiced speech)

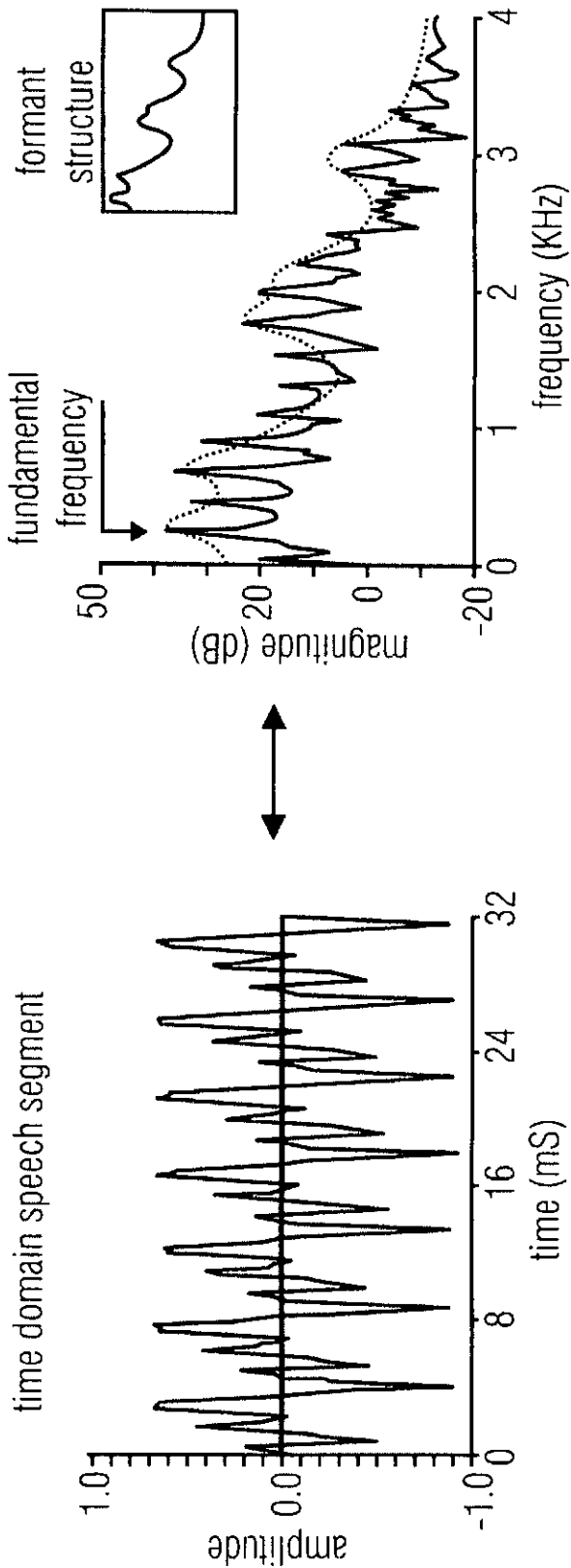


FIGURE 5A

FIGURE 5B

stationary segment (e.g. unvoiced speech)

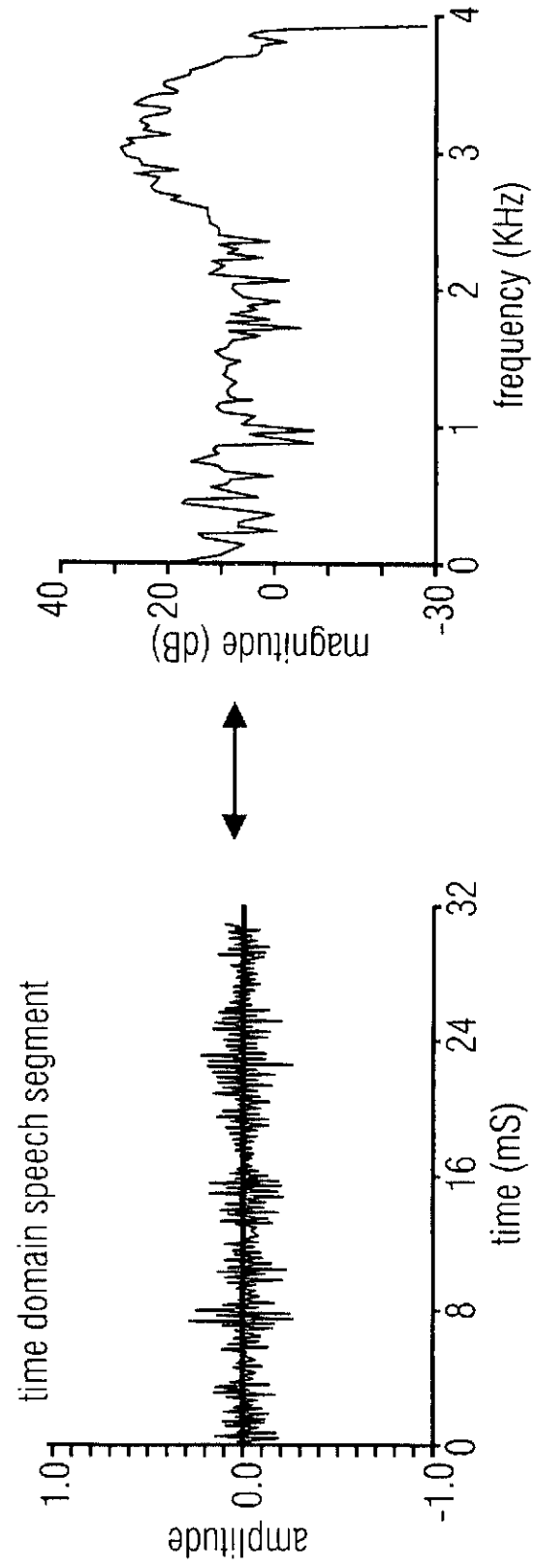
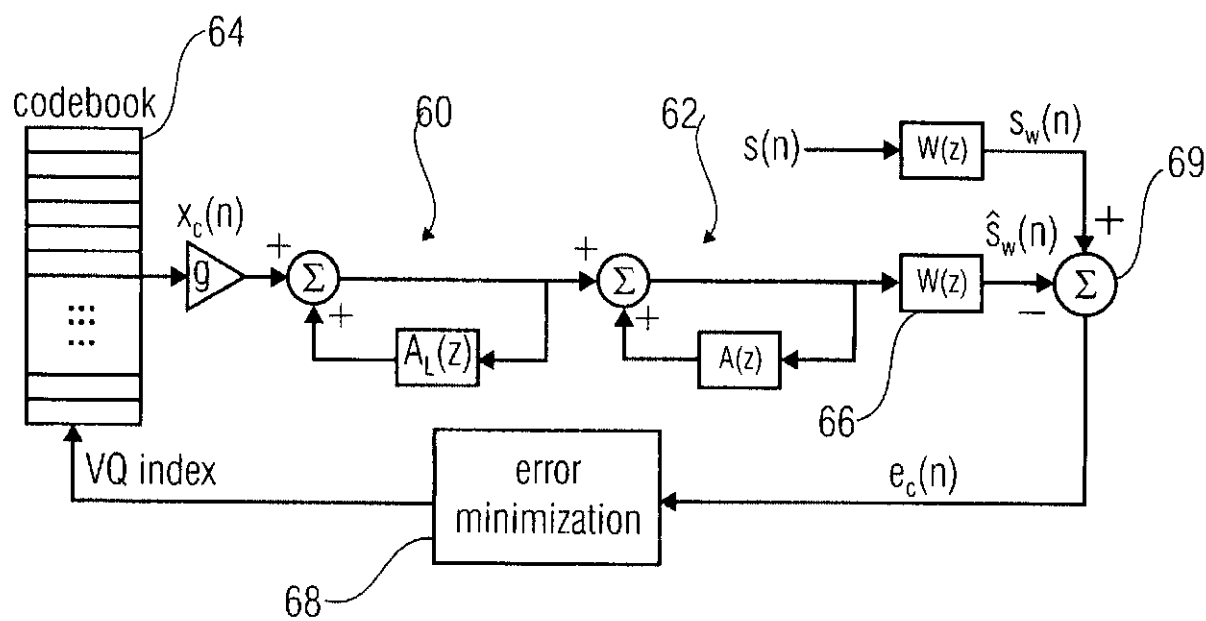


FIGURE 5C

FIGURE 5D

analysis-by-synthesis CELP



$A_L(z)$: long term prediction
 $\hat{=}$ pitch (fine) structure

$A(z)$: short term prediction
 $\hat{=}$ format structure / spectral envelope

FIGURE 6

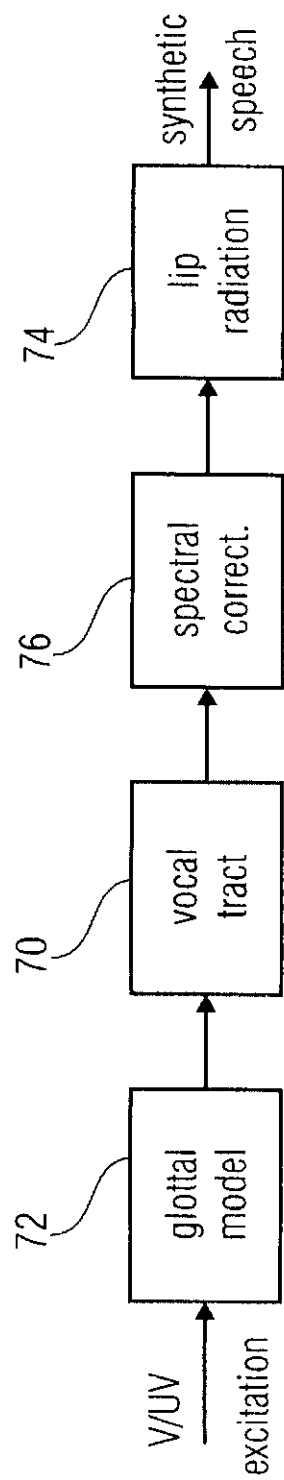


FIGURE 7A

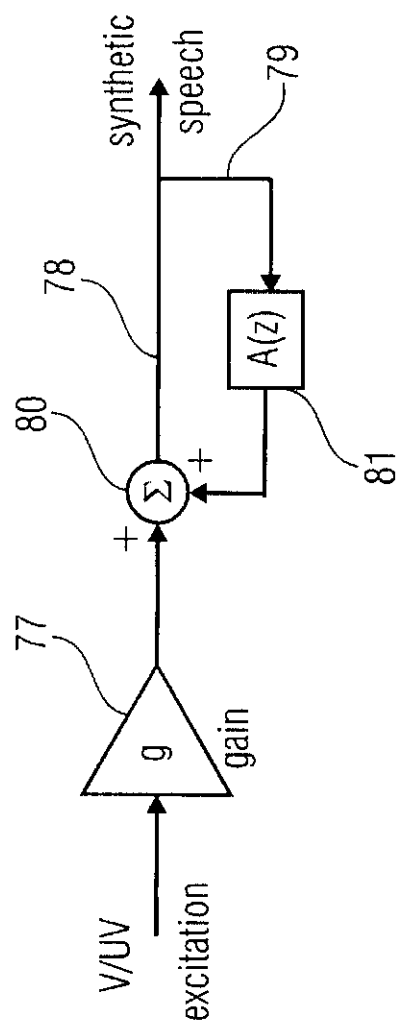


FIGURE 7B

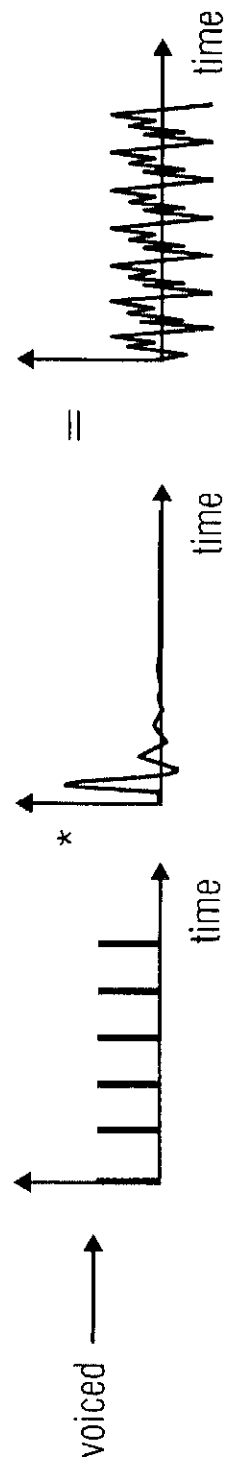


FIGURE 7C

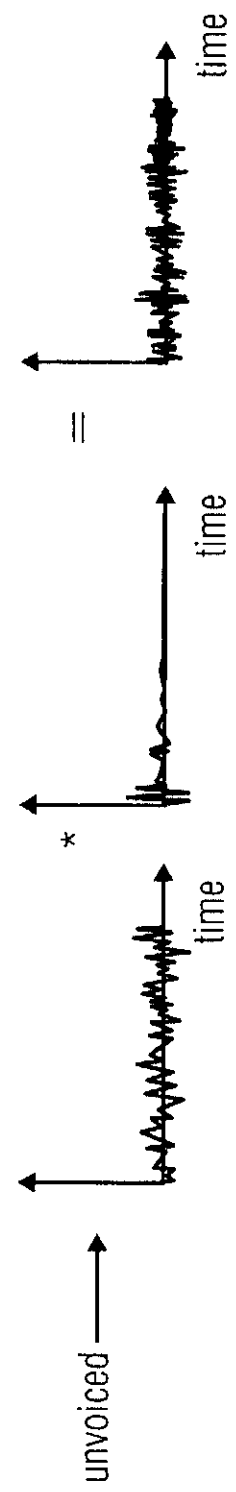


FIGURE 7D

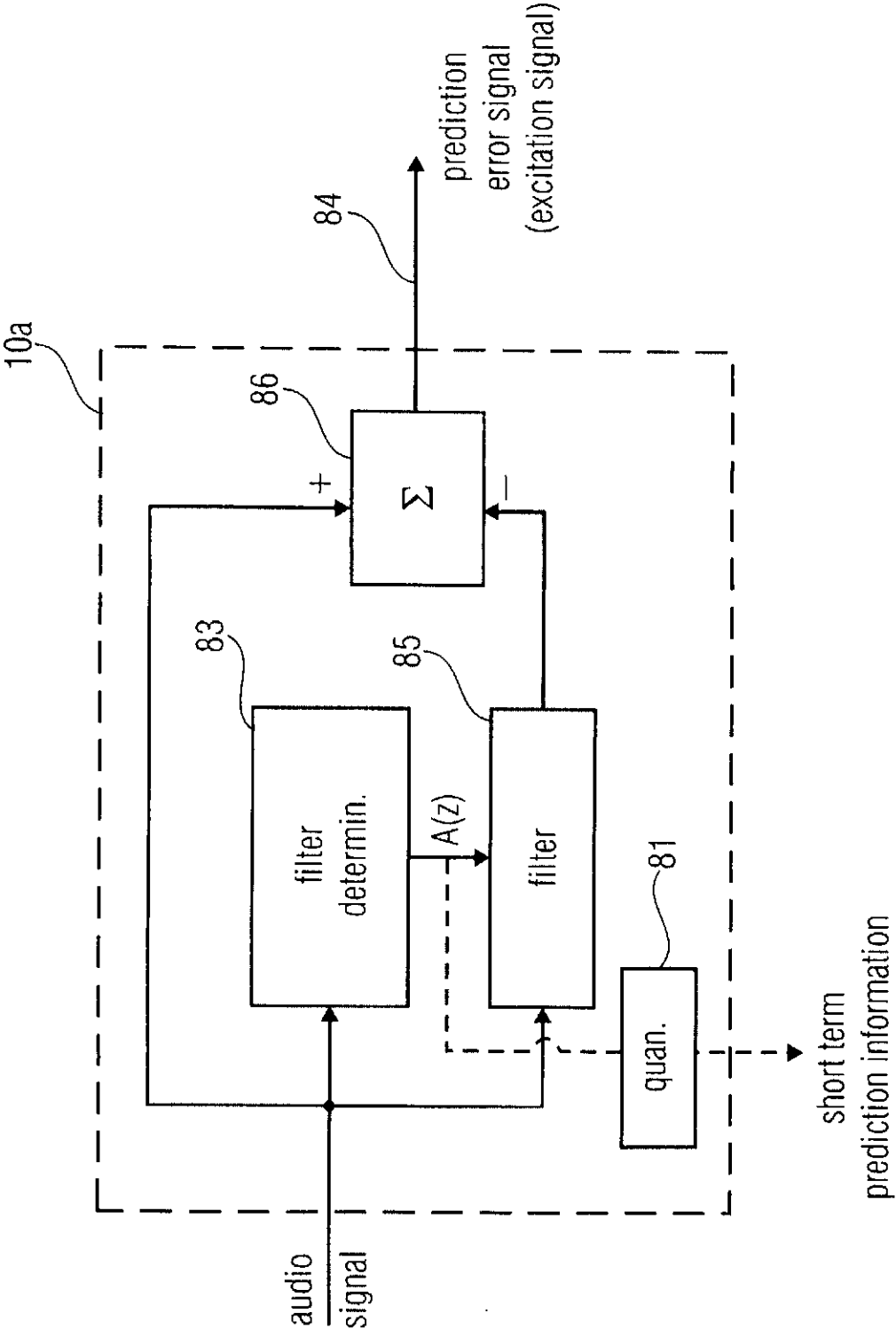


FIGURE 7E

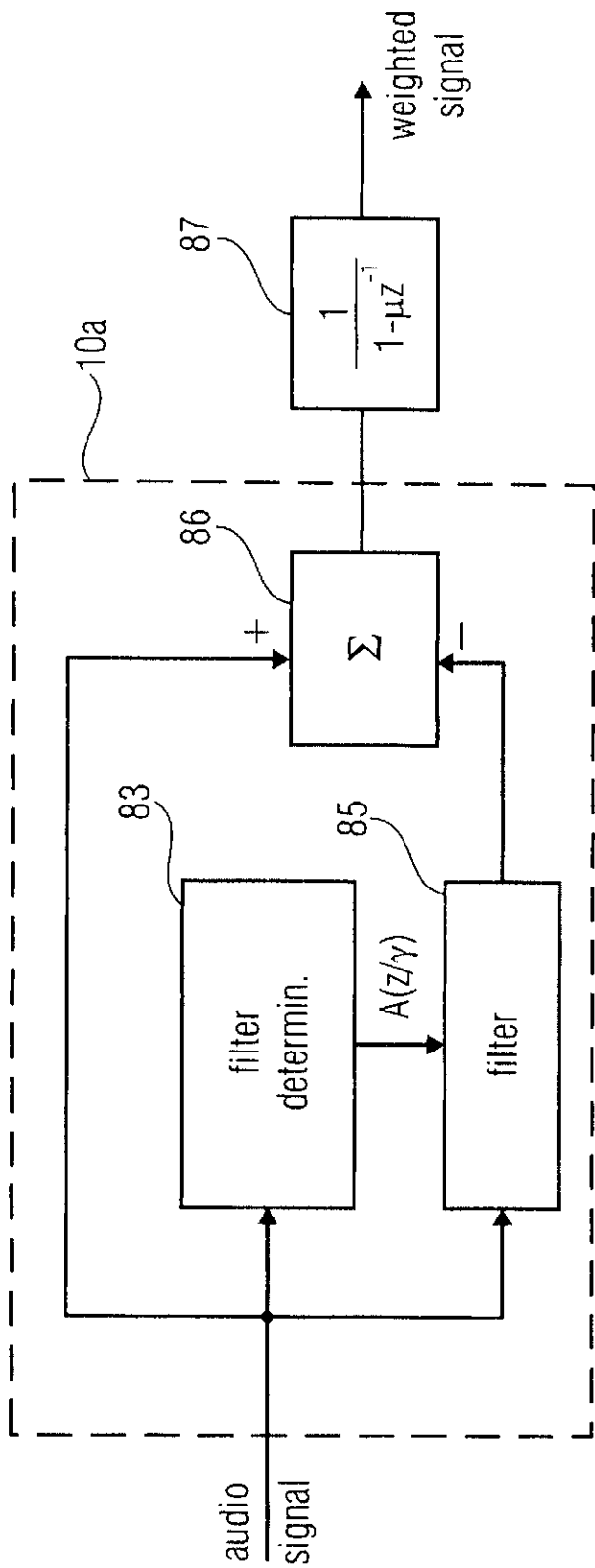


FIGURE 7F
(ENCODER SIDE)

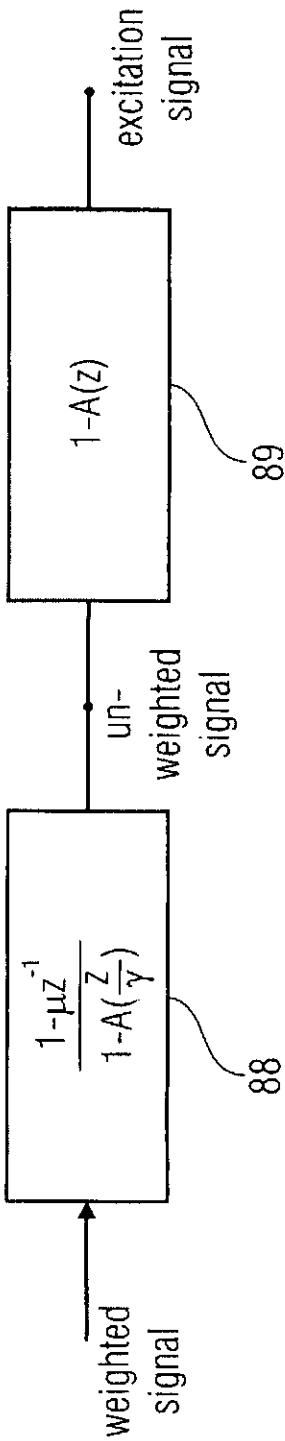


FIGURE 7G
(DECODER SIDE)

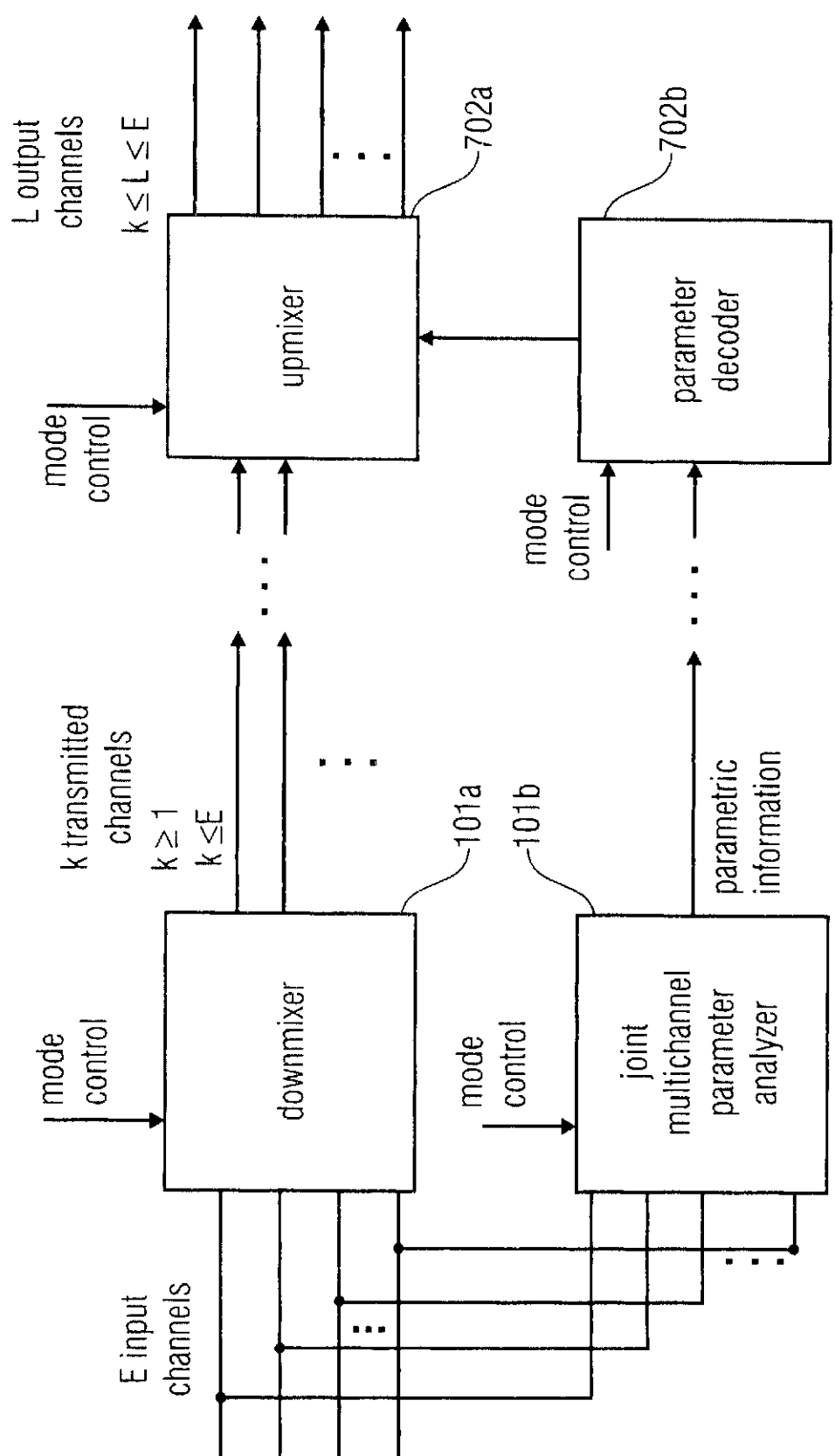


FIGURE 8

206

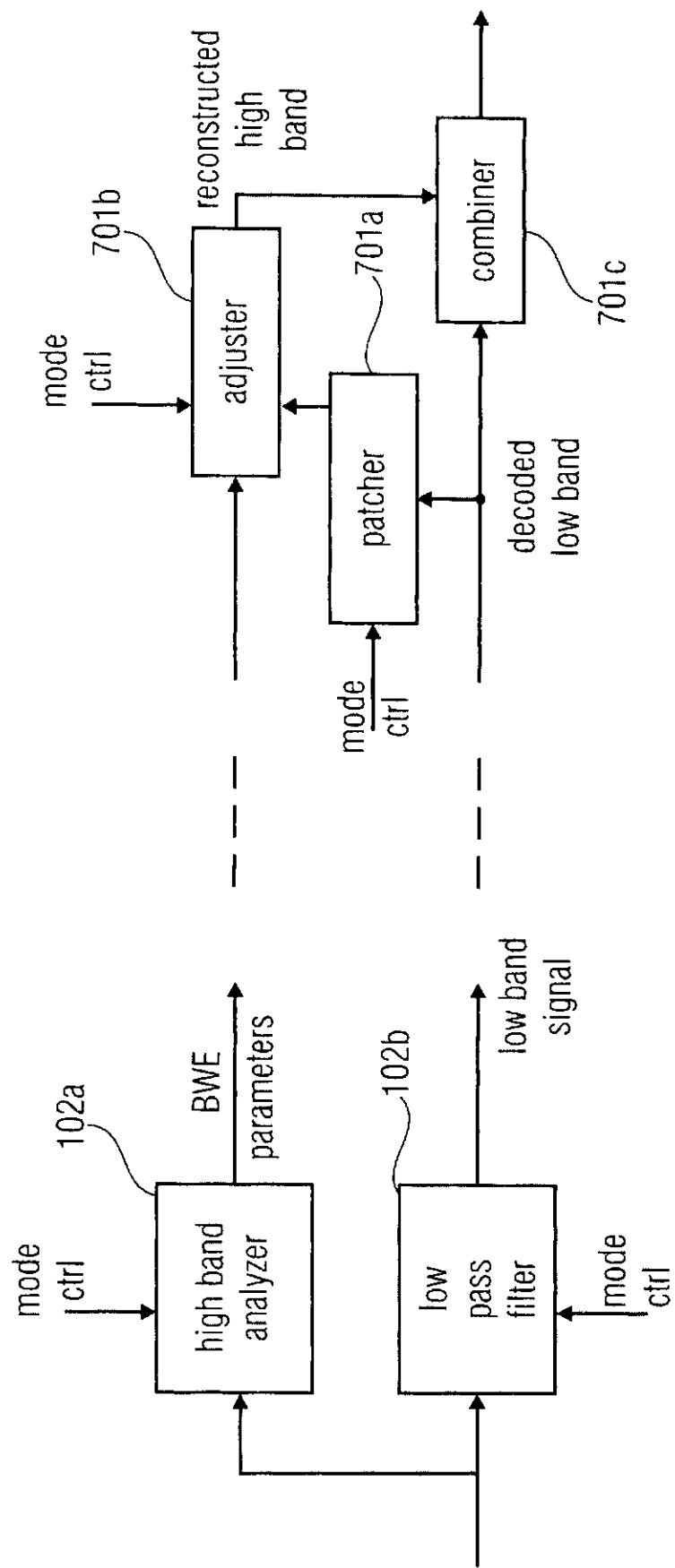
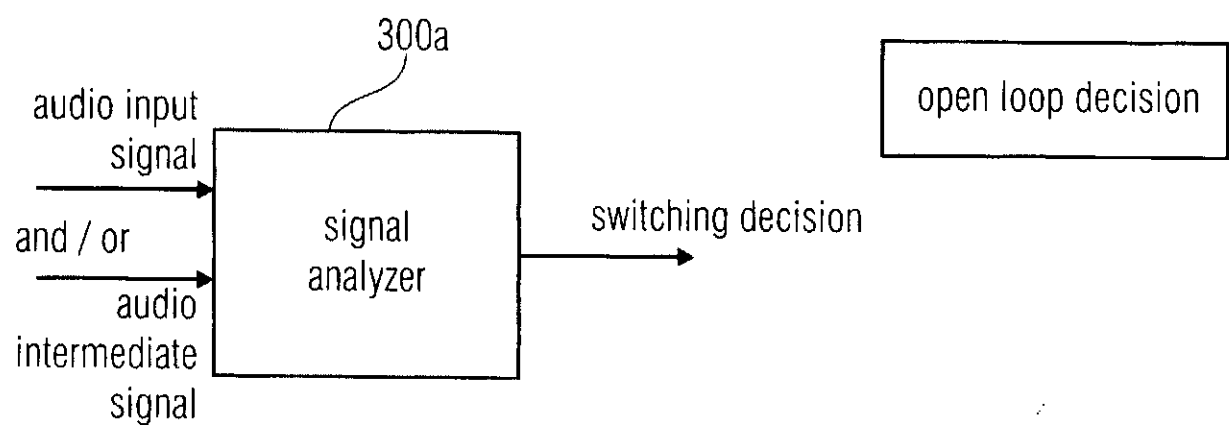


FIGURE 9



audio intermediate signal:
- low band signal;
- downmix signal; or
- low band portion of downmix
signal

FIGURE 10A

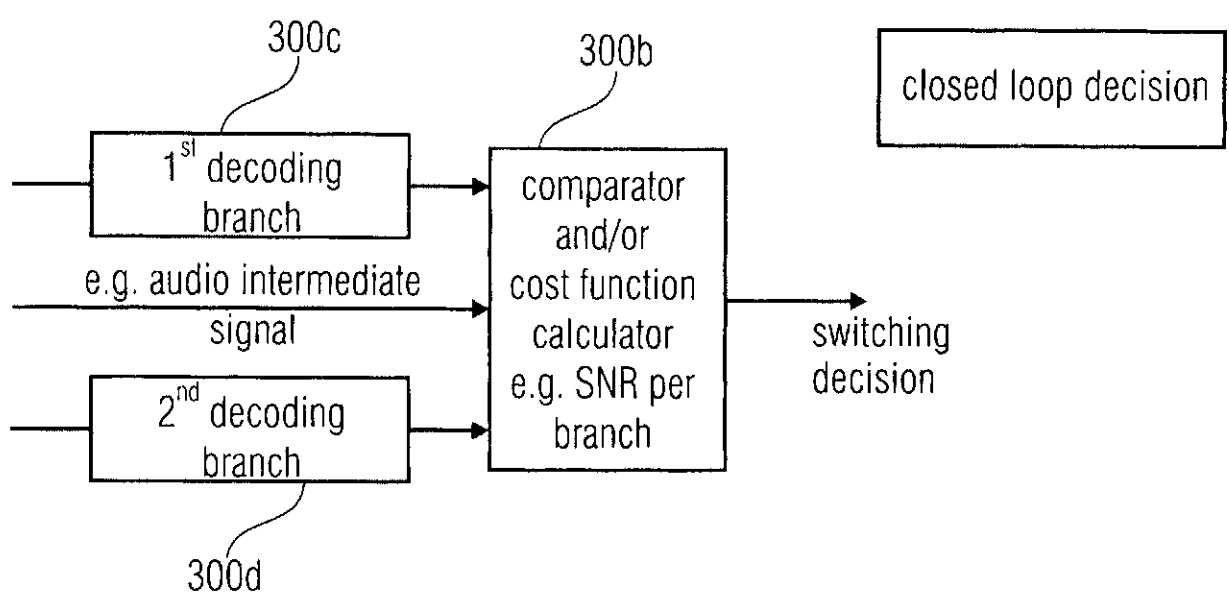


FIGURE 10B

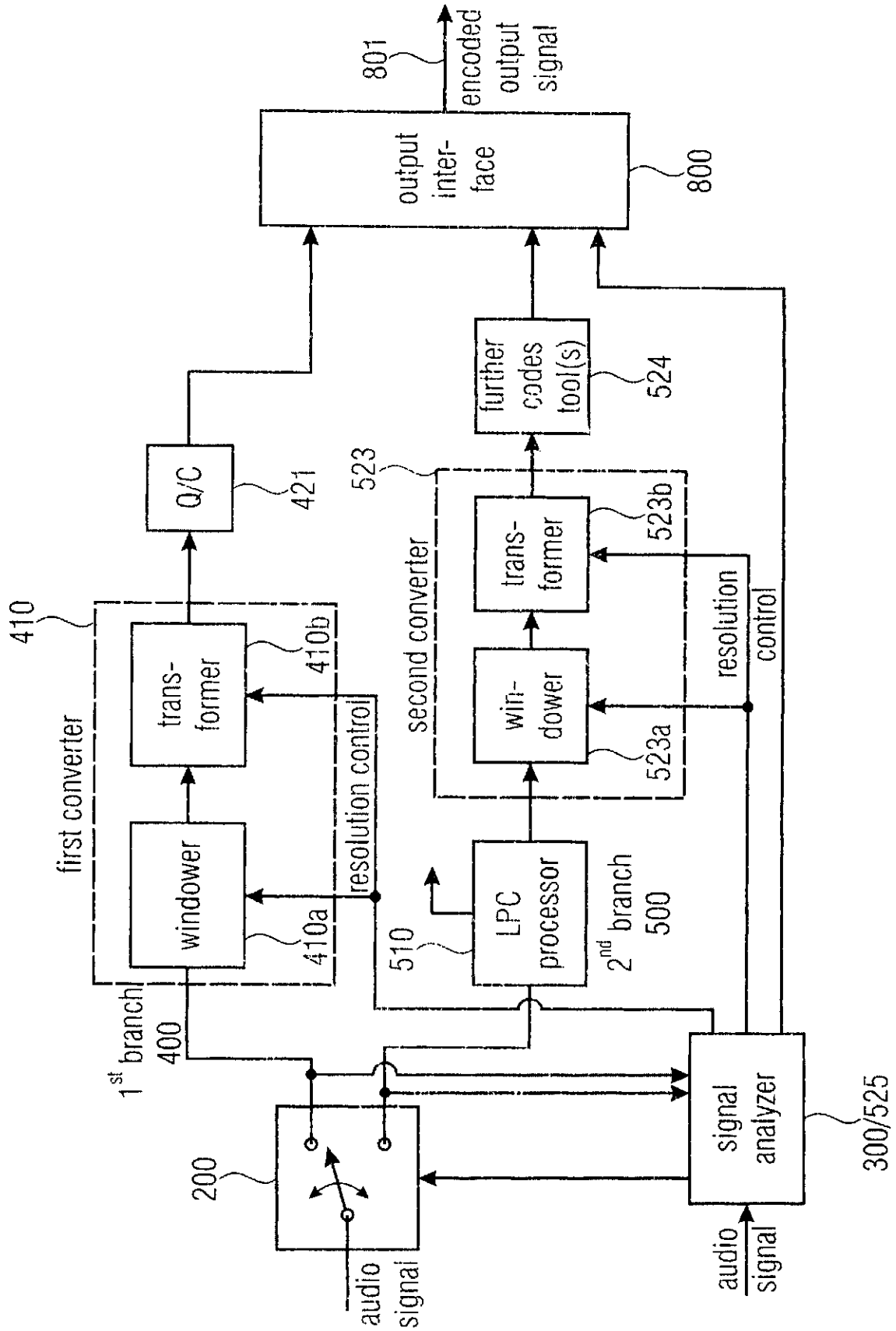


FIGURE 11A

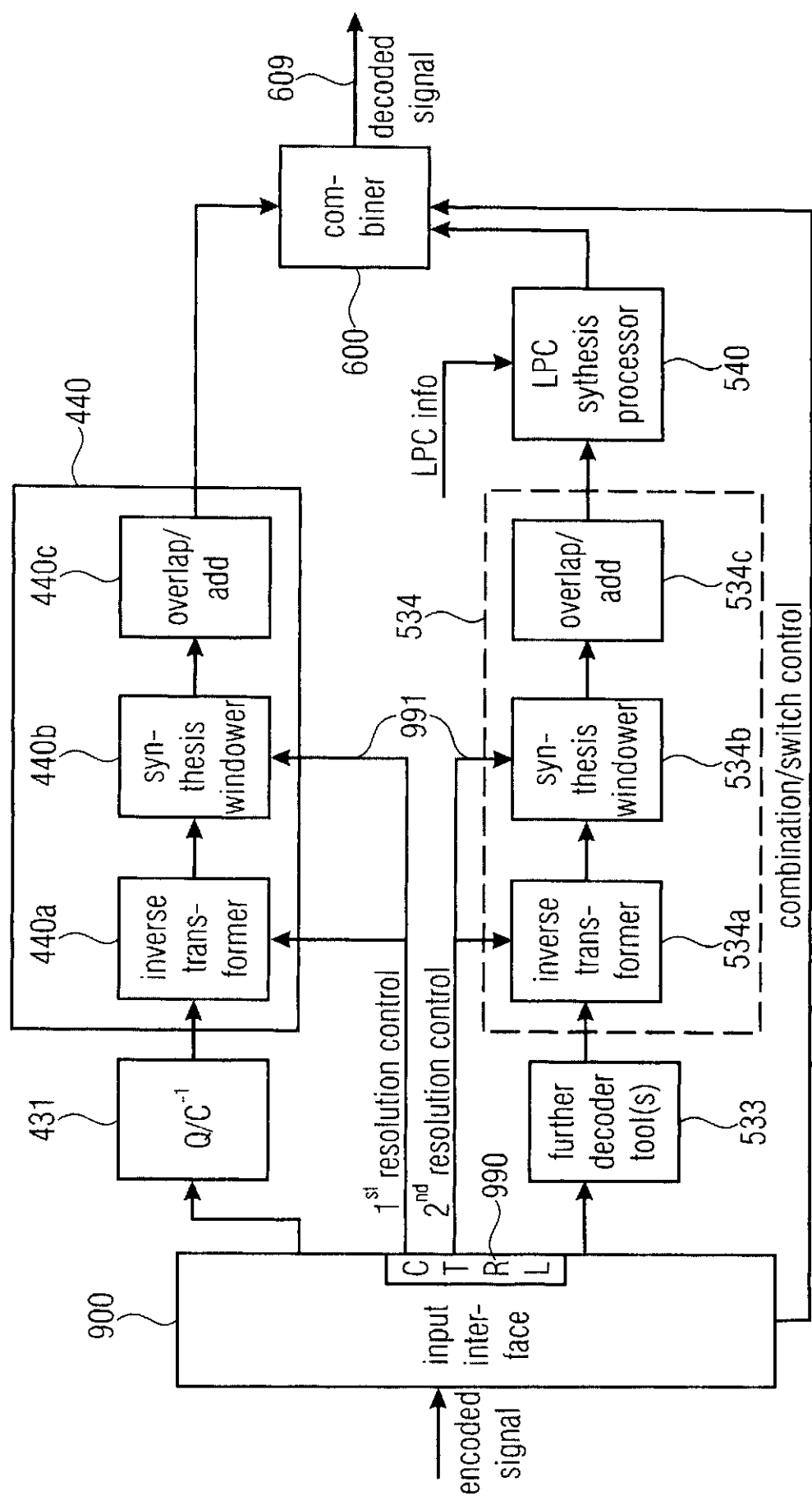


FIGURE 11B

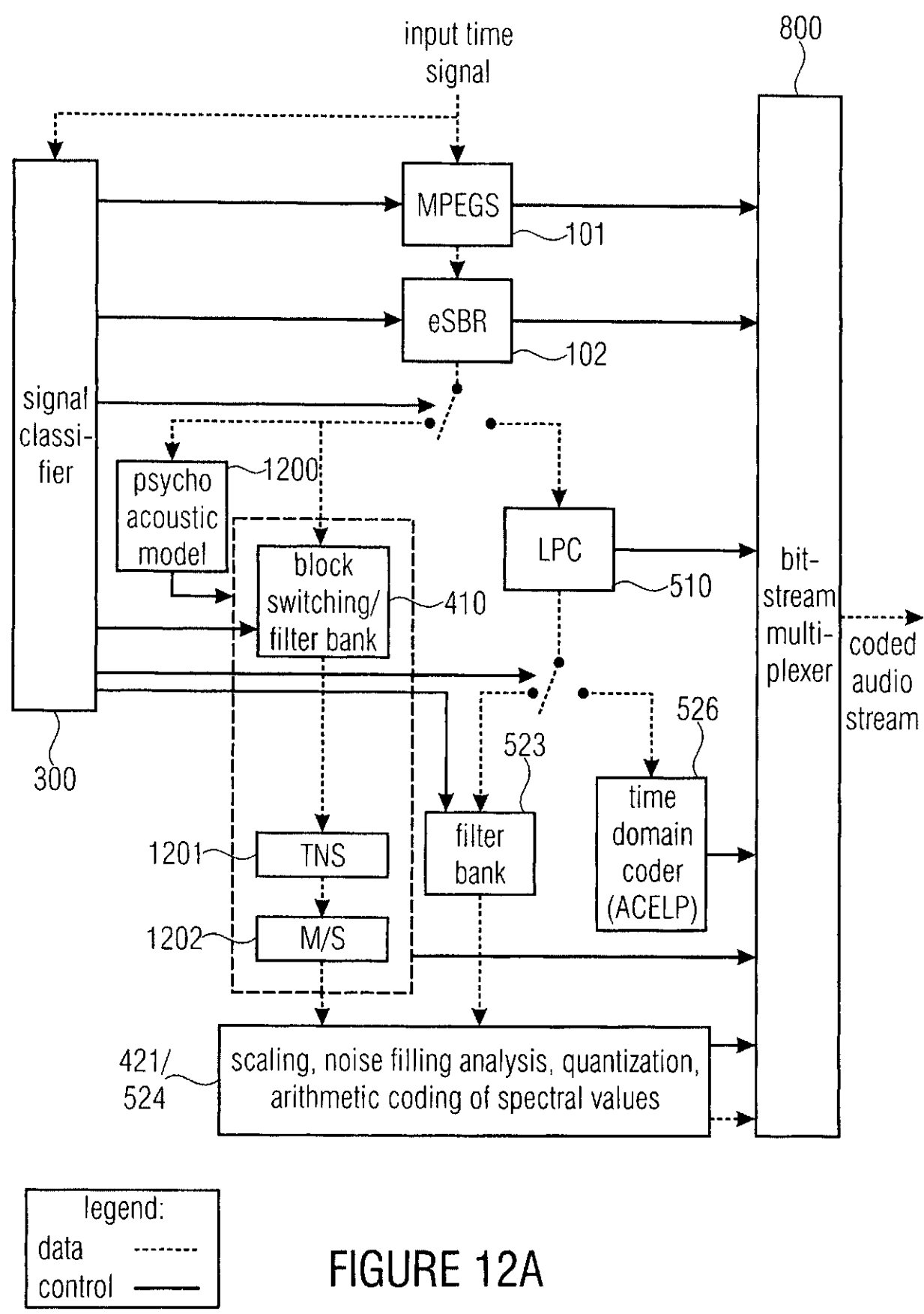


FIGURE 12A

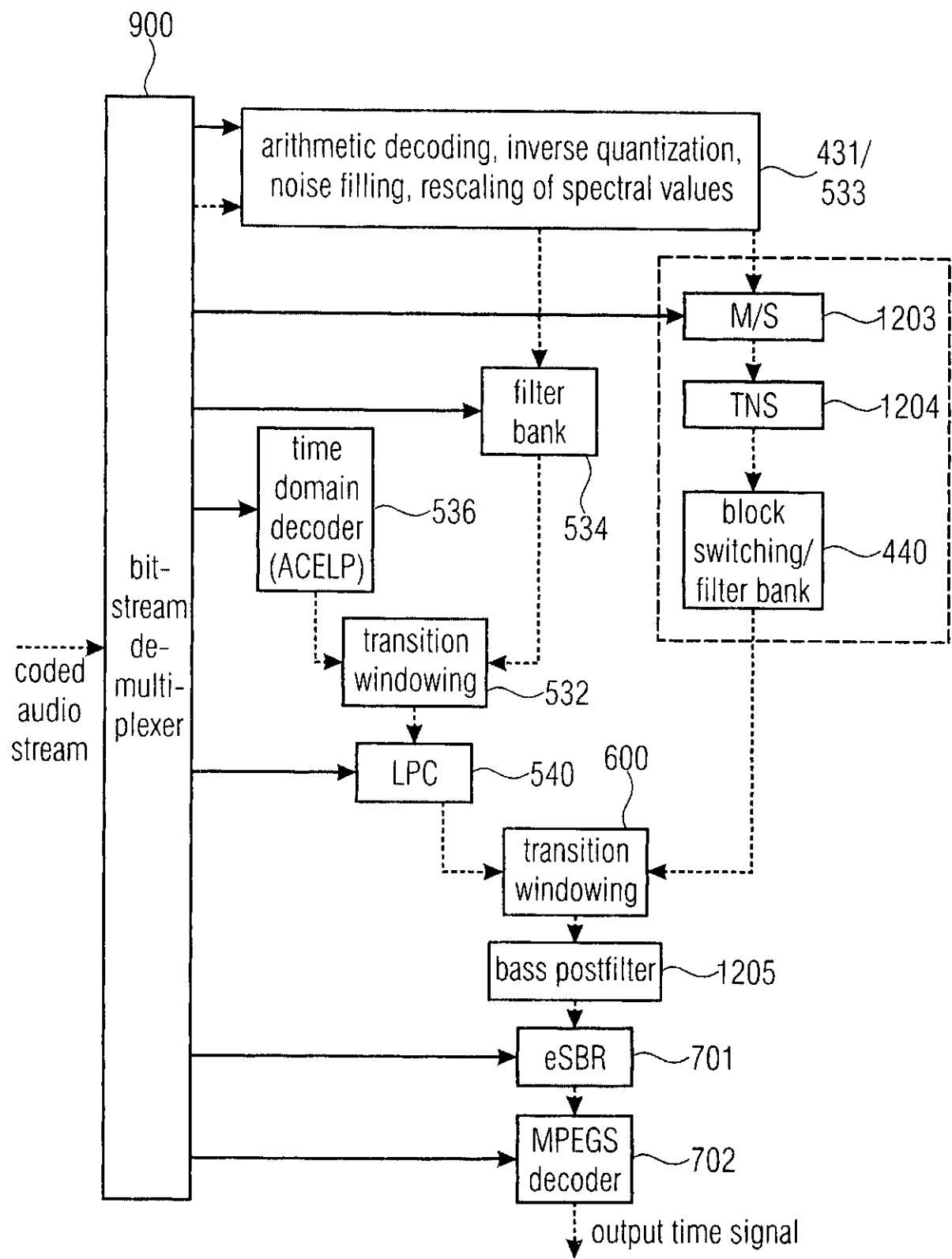
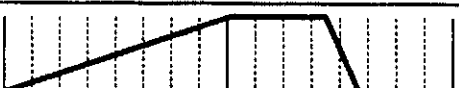
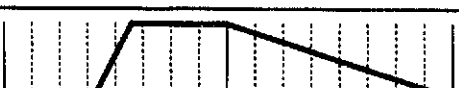
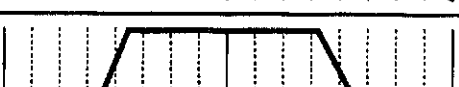
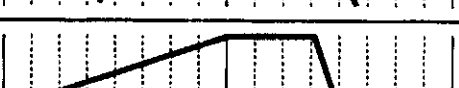
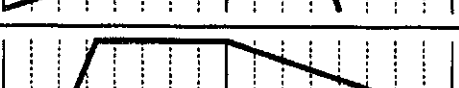
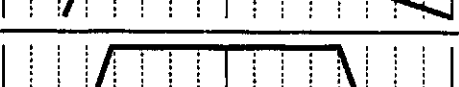


FIGURE 12B

window length	transform length	time resolution	frequency resolution
short	small	high	low
long	large	low	high

FIGURE 13A

window	#coeffs	looks like
LONG_WINDOW	1024/ 960	
SHORT_WINDOW	128/ 120	
LONG_START_WINDOW	1024/ 960	
LONG_STOP_WINDOW	1024/ 960	
STOP_START_WINDOW	1024/ 960	
START_WINDOW_LPD	1024/ 960	
STOP_WINDOW_1152	1152/ 1080	
STOP_START_WINDOW_1152	1152/ 1080	

transform windows

FIGURE 13B
(AAC BRANCH AND TRANSITION)

value	window_sequence	num_windows	looks like
0	ONLY_LONG_SEQUENCE = LONG_WINDOW	1	
1	LONG_START_SEQUENCE = LONG_START_WINDOW	1	
2	EIGHT_SHORT_SEQUENCE = 8 * SHORT_WINDOW	8	
3	LONG_STOP_SEQUENCE = LONG_STOP_WINDOW	1	
1	STOP_START_SEQUENCE = STOP_START_WINDOW	1	
3	LPD_START_SEQUENCE = START_WINDOW_LPD	1	
3	STOP_1152_SEQUENCE = STOP_WINDOW_1152	1	
1	STOP_START_1152_SEQUENCE = STOP_START_WINDOW_1152	1	

window sequences

FIGURE 13C
(AAC BRANCH AND TRANSITION)

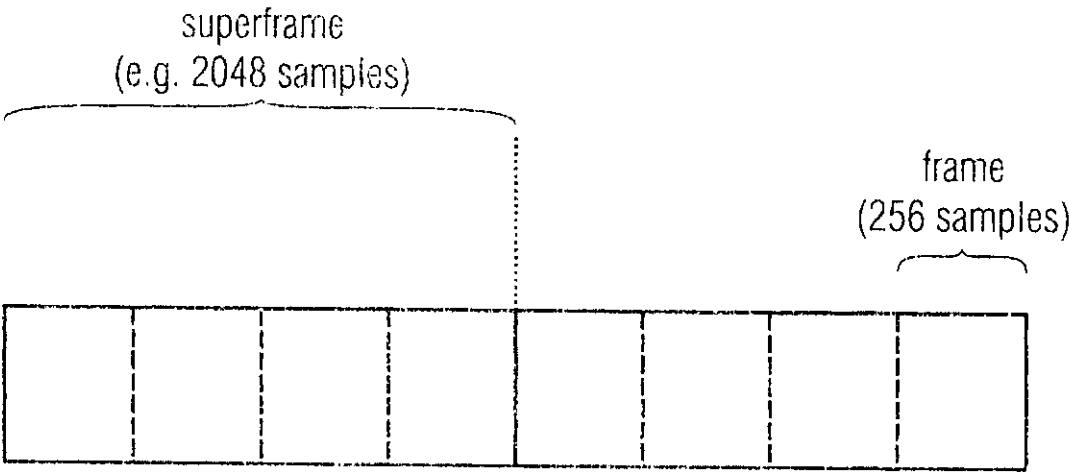


FIGURE 14A

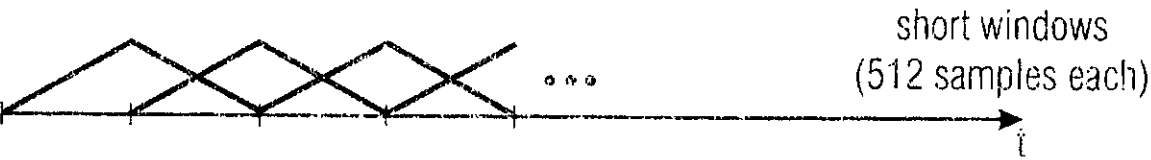


FIGURE 14B

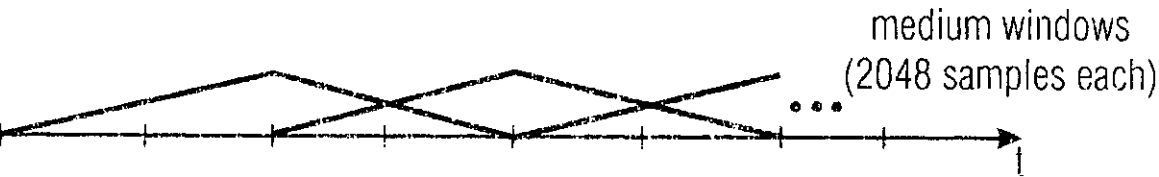


FIGURE 14C

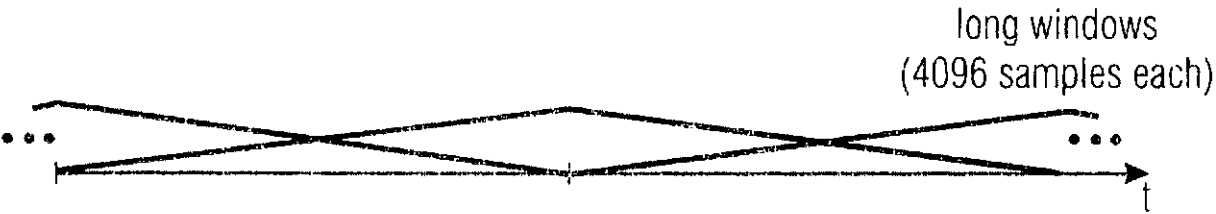


FIGURE 14D

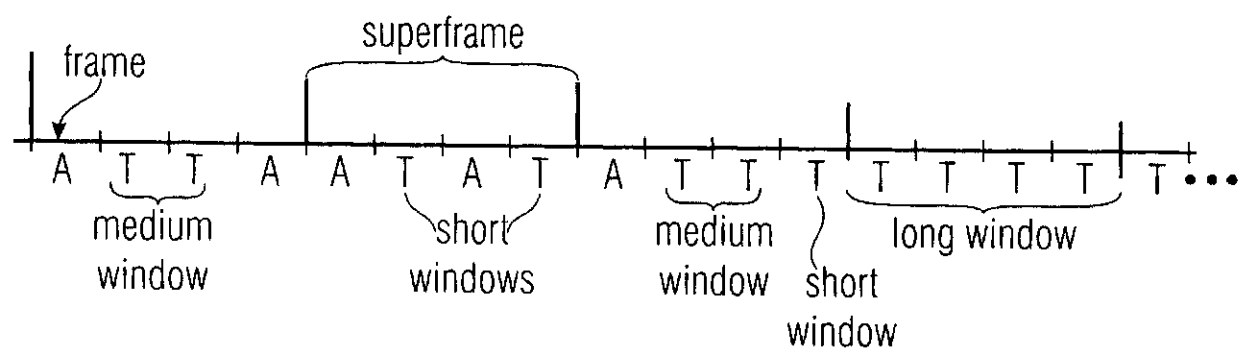
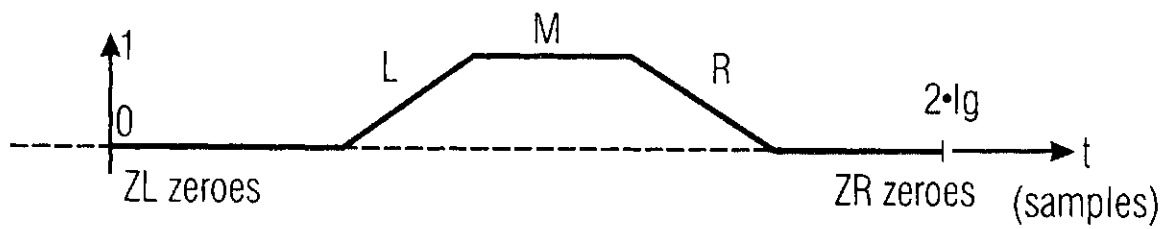


FIGURE 14E

value of last_lpd_mode	value of mod[x]	number lg of spectral coefficients	ZL	L	M	R	ZR
0	1	320	160	0	256	128	96
0	2	576	288	0	512	128	224
0	3	1152	512	128	1024	128	512
1..3	1	256	64	128	128	128	64
1..3	2	512	192	128	384	128	192
1..3	3	1024	448	128	896	128	448

FIGURE 14F



window definition of FIGURE 14F

FIGURE 14G



- (51) International Patent Classification:
G10L 19/14 (2006.01)
- (21) International Application Number:
PCT/EP2009/007205
- (22) International Filing Date:
7 October 2009 (07.10.2009)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/103,825 8 October 2008 (08.10.2008) US
08017663.9 8 October 2008 (08.10.2008) EP
09002271.6 18 February 2009 (18.02.2009) EP
- (71) Applicants (for all designated States except US):
FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E. V. [DE/DE]; Hansastr. 27c, 80686 München (DE). **VOICEAGE CORPORATION** [CA/CA]; 750 Lucerne Road, Suite 250, Montreal, Québec H3R 2H6 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **NEUENDORF, Max** [DE/DE]; Kreilingstr.15, 90408 Nürnberg (DE). **BAYER, Stefan** [DE/DE]; Johannisstr.148, 90419 Nürnberg (DE). **LECOMTE, Jérémie** [FR/DE]; Sulzbacher Strasse 39, 90489 Nürnberg (DE). **FUCHS, Guillaume** [FR/DE]; Äussere Brucker Strasse Apt. West, 91052 Er-

langen (DE). **ROBILLIARD, Julien** [FR/DE]; Innerer Kleinreuther Weg 25A, 90408 Nürnberg (DE). **REIT-TEL-BACH, Nikolaus** [DE/DE]; Amorbacher Str.2 a, 90427 Nürnberg (DE). **NAGEL, Frederik** [DE/DE]; Wilhelmshavener Strasse 72, 90425 Nürnberg (DE). **GEIGER, Ralf** [DE/DE]; Münzstr. 8c, 98693 Ilmenau (DE). **MULTRUS, Markus** [DE/DE]; Etzlaubweg 7, 90469 Nürnberg (DE). **GRILL, Bernhard** [DE/DE]; Peter-Henlein-Str.7, 91207 Lauf (DE). **GOURNAY, Philippe** [FR/CA]; 3012 rue du Sauvignon, Sherbrooke, Québec J1L 0A2 (CA). **SALAMI, Redwan** [LB/CA]; 4045 Albert -Dreux Place, Saint - Laurent, Québec H4R 2Y3 (CA).

(74) Agents: **ZINKLER, Franz** et al.; Schoppe, Zimmermann, Stöckeler & Zinkler, Postfach 246, 82043 Pullach bei München (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, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[Continued on next page]

(54) Title: MULTI-RESOLUTION SWITCHED AUDIO ENCODING/DECODING SCHEME

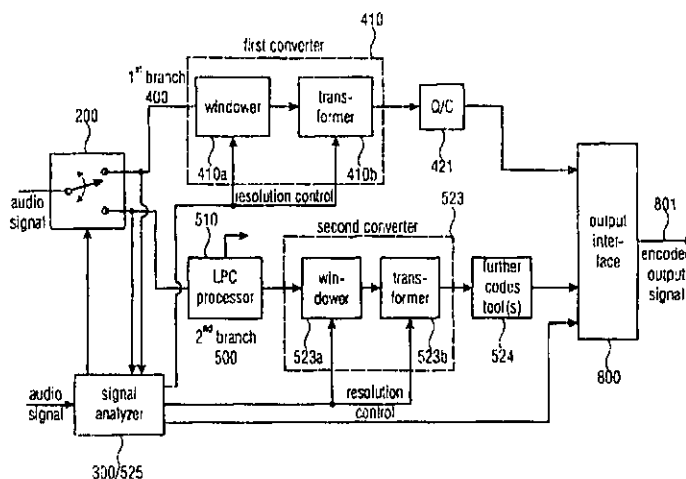


FIGURE 11A

(57) Abstract: An audio encoder for encoding an audio signal comprises a first coding branch (400), the first coding branch comprising a first converter (410) for converting a signal from a time domain into a frequency domain. Furthermore, the audio encoder comprises a second coding branch (500) comprising a second time/frequency converter (523). Additionally, a signal analyzer (300/525) for analyzing the audio signal is provided. The signal analyzer, on the hand, determines whether an audio portion is effective in the encoder output signal as a first encoded signal from the first encoding branch or as a second encoded signal from a second encoding branch. On the other hand, the signal analyzer determines a time/frequency resolution to be applied by the converters (410, 523) when generating the encoded signals. An output interface includes, in addition to the first encoded signal and the second encoded signal, a resolution information identifying the resolution used by the first time/frequency converter and used by the second time/frequency converter.



(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference FH091002PCT	FOR FURTHER ACTION see Form PCT/ISA/220 as well as, where applicable, item 5 below.	
International application No. PCT/EP2009/007205	International filing date (day/month/year) 07/10/2009	(Earliest) Priority Date (day/month/year) 08/10/2008
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 4 sheets.

☒ 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:

- ☒ 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 43.6b/s(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**,

- ☒ 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**,

- ☒ 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. 1a

- ☐ 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/EP2009/007205

A. CLASSIFICATION OF SUBJECT MATTER

INV. G10L19/02 G10L19/14
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

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.
A	WO 2008/071353 A (FRAUNHOFER GES FORSCHUNG [DE]; GEIGER RALF [DE]; NEUENDORF MAX [DE]; Y) 19 June 2008 (2008-06-19) page 6, line 20 - line 26 page 19, line 4 - line 6 page 20, line 1 - page 22, line 31 figures 1A-1F, 3A-3D ----- -/-	1-21

☒ 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.
- *Z* document member of the same patent family

Date of the actual completion of the international search

13 July 2010

Date of mailing of the international search report

21/07/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel: (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Tilp, Jan

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2009/007205

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SEAN A RAMPRASHAD: "The Multimode Transform Predictive Coding Paradigm" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 11, no. 2, 1 March 2003 (2003-03-01), XP011079700 ISSN: 1063-6676 page 117, right-hand column, lines 14-48 page 118, paragraph II, left-hand column, lines 11-13 page 118, paragraph II, right-hand column, lines 10-50 page 119, left-hand column, line 1- right-hand column, line 35 page 121, left-hand column, lines 1-39 figures 2,3 -----	1-21
A	US 2003/009325 A1 (KIRCHHERR RAIF [DE] ET AL) 9 January 2003 (2003-01-09) paragraphs [0048] - [0050] figures 2,2A,2B -----	1-21
A	WO 98/02971 A (FRAUNHOFER GES FORSCHUNG [DE]; DOLBY LABORATORIES INC [US]) 22 January 1998 (1998-01-22) page 7, line 2 - line 16 page 8, line 26 - page 9, line 7 page 10, line 26 - page 11, line 3 figures 1-7 -----	1-4,6,7, 9-11, 19-21

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2009/007205

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008071353 A	19-06-2008	AU 2007331763 A1 CA 2672165 A1 CN 101589623 A EP 2052548 A2 JP 2010512550 T KR 20090085655 A US 2010138218 A1	19-06-2008 19-06-2008 25-11-2009 29-04-2009 22-04-2010 07-08-2009 03-06-2010
US 2003009325 A1	09-01-2003	NONE	
WO 9802971 A	22-01-1998	AT 226375 T AU 710868 B2 AU 7696196 A CA 2260033 A1 DE 69624383 D1 DE 69624383 T2 EP 0910900 A1 JP 3171598 B2 KR 20000023674 A US 5848391 A	15-11-2002 30-09-1999 09-02-1998 22-01-1998 21-11-2002 12-06-2003 28-04-1999 28-05-2001 25-04-2000 08-12-1998

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY
(PCT Rule 43bis.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.
PCT/EP2009/007205

International filing date (day/month/year)
07.10.2009

Priority date (day/month/year)
08.10.2008

International Patent Classification (IPC) or both national classification and IPC
INV. G10L19/02 G10L19/14

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 43bis.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 notified the International Bureau under Rule 66.1bis(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/EP2009/007205

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. II Priority

1. ☐ The validity of the priority claim has not been considered because the International Searching Authority does not have in its possession a copy of the earlier application whose priority has been claimed or, where required, a translation of that earlier application. This opinion has nevertheless been established on the assumption that the relevant date (Rules 43bis.1 and 64.1) is the claimed priority date.
2. ☒ This opinion has been established as if no priority had been claimed due to the fact that the priority claim has been found invalid (Rules 43bis.1 and 64.1). Thus for the purposes of this opinion, the international filing date indicated above is considered to be the relevant date.
3. Additional observations, if necessary:

see separate sheet

**WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY**

International application No.
PCT/EP2009/007205

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	<u>1-21</u>
	No: Claims	
Inventive step (IS)	Yes: Claims	<u>1-21</u>
	No: Claims	
Industrial applicability (IA)	Yes: Claims	<u>1-21</u>
	No: Claims	

2. Citations and explanations

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 II

Priority

The priority claims have been found invalid (Rules 43bis.1 and 64.1). In particular, none of the claimed priority documents discloses

- a) the functionality / combination of features of the signal analyzer and of the output interface as claimed in present claims 1, 9 and 21, and
- b) the functionality / combination of features of the controllable frequency/time converters as claimed in present claims 10, 19 and 21.

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

Notwithstanding the invalidity of the priority claims mentioned above under Re Item II and the lack of clarity mentioned below under Re Item VIII, claims 1-21 as presently on file appear to fulfill the requirements of the PCT with regard to novelty (Article 33(2) PCT) and inventive step (Article 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 certain claims are not clear.
- 1.1 **Claim 1** is unclear in that it claims the expression "the first converter" (see lines 6-7) for which no antecedent basis can be identified.
- 1.2 In light of line 24 of **claim 1** it is obvious that the claimed "first converter" and "second converter" should - more clearly and specifically - be claimed as "the first time-frequency converter" and "the second time-frequency converter". Accordingly, the claimed "converting an input signal into a spectral domain" (see claim 1, line 7) should be replaced by "converting an input signal from the time domain into a spectral domain".
- 1.3 It appears to be intended that **claim 1** be interrelated to claim 10. Therefore, to truly interrelate to said latter claim, the wording of unclear claim 1 should be adapted so as to comprise "a first/second controllable time-frequency converter" instead of "a first/second converter".

1.4 The terms

- i) "first encoded signal" (see claim 1, lines 6, 17, 21-22, 25 and 29-32),
- ii) "input signal" (see claim 1, line 7),
- iii) "input signal from an input domain" (see claim 1, line 12),
- iv) "input signal" (see claim 1, line 13),
- v) "second encoded signal" (see claim 1, lines 10, 18, 22, 25 and 29-32),
- vi) "an input domain" (see claim 1, line 12),
- vii) "encoder output signal" (see claim 1, lines 18, 22),
- viii) "an encoder output signal" (see claim 1, line 28),
- ix) "a/the signal analyzer" (see claim 1, lines 20 and 23), and
- x) "the audio signal" (see claim 1, lines 20, 21 and 25-26)

used in **claim 1** are vague and unclear and leave 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 the following, said terms are interpreted so as to refer to

- 1) "first encoded audio signal",
- 2) "audio input signal from the time domain",
- 3) "audio input signal",
- 4) "input signal from the output domain",
- 5) "second encoded audio signal",
- 6) "the time domain",
- 7) "a portion of an output signal of the audio encoder",
- 8) "the output signal of the audio encoder",
- 9) "two/the signal analyzers", and
- 10) "audio input signal".

- 1.5 The objections raised against claim 1 also, apply, mutatis mutandis, to all corresponding and/or interrelated claims on file in case of said latter claims containing the terms / expressions objected against under sections 1.1 - 1.4 above.

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 IPEA (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

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



SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712- -00000-0000

Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR.NUEVASCR
Fve: 378 FASENACIONALI

PATENTE DE INVENCION ☐

MODELO DE U

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712- -00000-0000

Fecha: 2011-05-06 12:33:26 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYI
Act. 411 PRESENTACION Ene: 378 FASENACIONALI
Folios: 257

☐	Indicación que se solicita una patente.
☐	Datos de identificación del solicitante o de
☐	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 INVENCION 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 ☐



Bogotá D.C.,
2020

Doctor(a)
CARDENAS NAVAS DARIO
Apoderado(a) y/o Representante de
VOICEAGE CORPORATION

REFERENCIA	Radicación No.	11 055712
	Solicitud internacional	PCT/EP2009/7205
	Fecha inicio fase nacional	08/05/2011
	Trámite	11
	Evento	378
	Actuación	430
	Oficio No.	08900

Como resultado del examen de forma, realizado con fundamento en el artículo 38 de la Decisión 486 de la Comisión de la Comunidad Andina, artículo 27 del Tratado de Cooperación en Materia de Patentes (PCT), regla 51 bis del Reglamento y Circular Única de la Superintendencia de Industria y Comercio, se le comunica que:

1. La solicitud no contiene la copia del documento en que consta la cesión del derecho a la patente del inventor al solicitante o a su causante, conforme a lo establecido por el artículo 26 literal k) de la Decisión 486 de la Comisión de la Comunidad Andina. Cuando se trate de documentos otorgados en el extranjero, estos deberán legalizarse de conformidad con lo dispuesto por los artículos 259 y 260 del Código de Procedimiento Civil. Cuando se trate de las declaraciones a que hace referencia la Regla 4.17 ii) del Tratado de Cooperación en Materia de Patentes PCT., el alcance de las tales declaraciones se determinará según indiquen claramente una cesión de derechos del inventor al solicitante o su causante, de conformidad con lo dispuesto en el numeral 5.2.15 del Capítulo Quinto, Título X de la Circular Única de Superintendencia de Industria y Comercio.

En cumplimiento de lo dispuesto por el artículo 39 de la Decisión 486, se le requiere para que dentro del término de dos meses siguientes a la fecha de notificación del presente oficio, complete los requisitos anteriormente enunciados. En caso de no dar cumplimiento al presente requerimiento, la solicitud se considerará abandonada y perderá su prelación.

Notifíquese el presente oficio conforme a lo dispuesto en el numeral 5.2, literal e), del capítulo quinto del título primero de la Circular Única No.10 de 2001.

JOSE LUIS SALAZAR LOPEZ
Director de Nuevas Creaciones (c)

El anterior requerimiento fue notificado por fijación en lista de fecha 03 JUL 2011
adpa11

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712- -00002-0000

Fecha: 2011-07-29 18:17:01 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 467 SOLPRORROGASOL Folios: 2

Señores

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

DIVISIÓN DE NUEVAS CREACIONES

E. S. D.

Ref.: Expediente No. 11055712

Solicitud de registro Patente de invención PCT denominada
ESQUEMA DE CODIFICACIÓN / DECODIFICACIÓN DE AUDIO
CONMUTADO DE RESOLUCIÓN MÚLTIPLE

Solicitante: **FRAUNHOFER-GESELLSCHAFT** **ZUR**
FORDERING DER ANGEWANDTEN FORSCHUNG e.V.

Darío Cárdenas Navas, identificado como aparece al pie de mi firma, atentamente me dirijo a ese Despacho dando respuesta al Oficio No. 08900 notificado por fijación en lista de fecha 3 de Junio de 2011.

En virtud de lo dispuesto en Artículo 39 de la Decisión 486 de la Comunidad Andina, atentamente solicito se me conceda la prórroga a que se refiere esta disposición.

Para tal efecto, adjunto el recibo oficial correspondiente a tal solicitud de prórroga.

Atentamente,

Darío Cárdenas Navas
c.c. 17.066.629 de Bogotá
t.p. 1.250 C.S.J. *AN*

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800.176.089-2

- / -

Industria y Comercio
SUPERINTENDENCIA

RECIBO DE CAJA

No. 11 - 0072500

Bogotá D.C., Julio 28 de 2011 - 16:11:58

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	391965011	28/07/2011	1.611.000.00

*** Conceptos Pagados ***

CANT.	RENTISTICO	CONCEPTO	VR. UNIDITARIO	VR. CONCEPTO
1	50005-01-08	PRORROGA DE TERMINOS	86.000.00	86.000.00
		1608 N CREACION PRORROGA DE TERMINOS O PLAZO ADICIONAL		\$86.000.00
			=====	

SON: **OCHENTA Y SEIS MIL PESOS MONEDA CORRIENTE**

Responsable: _____

Recibo de Caja Aplicado al Expediente No. _____

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-055712- -00002-0000

Fecha: 2011-07-29 16:17:01 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 467 SOLPRORROGASOL Folios: 2

Sede Centro: Carrera 13 No. 27 - 00 Pisos 3,4,5 y 10 Bogotá, D.C.- Colombia

Web: www.sic.gov.co e-mail: info@sic.gov.co Conmutador: (571) 5870000 Fax: (571) 5870284 Línea: 018000-910165 Call Center: (571) 6513240

Julio 28 de 2011 - 16:11:59

2