

PCT

re

Industria

SUPERINTENDENCIA

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 10-131388-00000-0000

Fecha: 2010-10-22 16:04:02 Dep. 2020 DIR.NUEVASCR  
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI  
Act. 411 PRESENTACION Folios: 206

DELEGATURA PROPIEDAD INDUSTRIAL

SOLICITUD  
REGISTRO DE PATENTE DE INVENCION

PCT

10-131388

EXPEDIENTE No

TITULO APARATO, METODO Y PROGRAMA DE  
COMPUTACIÓN PARA GENERAR UNA  
REPRESENTACIÓN DE UNA SEÑAL DE ANCHO DE  
BANDA AMPLIADO EN BASE A LA REPRESENTACIÓN DE  
UNA SEÑAL DE ENTRADA UTILIZANDO UNA  
COMBINACIÓN DE UNA AMPLIACIÓN DE ANCHO DE  
BANDA ARMÓNICO Y UNA AMPLIACIÓN DE ANCHO DE  
BANDA NO ARMÓNICO

CLASIFICACION INTERNACIONAL \_\_\_\_\_

SOLICITANTE Fraunhofer-Gesellschaft zur  
Förderung der angewandten Forschung e.V.

DOMICILIO Munich, Alemania

APODERADO DARIO CARDENAS NAVAS

BOGOTA D.C. \_\_\_\_\_

3.1.1  
02-11-2010

# PETITORIO



SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 10-131388- -00000-0000

Fecha: 2010-10-22 16:04:02 Dep. 2020 DIR.NUEVASCR  
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI  
Act. 411 PRESENTACION Folios: 206

## FORMULARIO UNICO

### ENTRADA EN FASE NACIONAL

①

#### SOLICITUD DE:

☒ Patente de Invención

☐ Patente de Modelo de Utilidad

☒ Capítulo I

☐ Capítulo II

②

#### SOLICITANTE

Nombre: **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.**

Dirección: Hansastrasse 27c 80686 Munich, Alemania

Nacionalidad o Domicilio: Munich, Alemania

Lugar de Constitución: Alemania

Teléfono: 3137800 Fax: 3122410

E-mail: [general@cardenasycardenas.com](mailto: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](mailto:general@cardenasycardenas.com)

#### IDENTIFICACION

☐ NIT ☐ NIT

☒ CC ☐ Otro

Cual 17066629 BTA  
Número 1250 C.S.J.

④

#### INVENTOR(ES) (72)

Nombre:

- Frederik Nagel
- Max Neuendorf
- Nikolaus Rettelbach
- Jérémie Lecomte
- Markus Multrus
- Berhard Grill

Dirección:

- Wilhelmshavener Strasse 72 90425 Nuernberg, Alemania
- Theatergasse 17 90402 Nuernberg, Alemania
- Spessartstr 38 90427 Nuernberg, Alemania
- Sulzbacher Strasse 39 90489 Nuernberg, Alemania
- Etzlaubweg 7 90469 Nuernberg, Alemania
- Peter-Henlein-Str. 7 91207 Lauf, Alemania

Nacionalidad o Domicilio:

- Alemán
- Alemán
- Alemán
- Francés
- Alemán
- Alemán

⑤ PCT (86)

Solicitud Internacional No. WO 2010/112587 Fecha 07.10.2010

Publicación Internacional No. PCT/EP2010/054422 Fecha 01.04.2010

⑥

**Título(54): APARATO, MÉTODO Y PROGRAMA DE COMPUTACIÓN PARA GENERAR UNA REPRESENTACIÓN DE UNA SEÑAL DE ANCHO DE BANDA AMPLIADO EN BASE A LA REPRESENTACIÓN DE UNA SEÑAL DE ENTRADA UTILIZANDO UNA COMBINACIÓN DE UNA AMPLIACIÓN DE ANCHO DE BANDA ARMÓNICO Y UNA AMPLIACIÓN DE ANCHO DE BANDA NO ARMÓNICO**

⑦

#### Clasificación Internacional (51):

⑧

#### Prioridad

SI ☒ NO ☐

(33)País de Origen

US

US

EP

(32)Fecha

02.04.2009

09.04.2009

30.12.2009

(31)N° de Solicitud

61/166,125

61/168,068

09181008.5

⑨

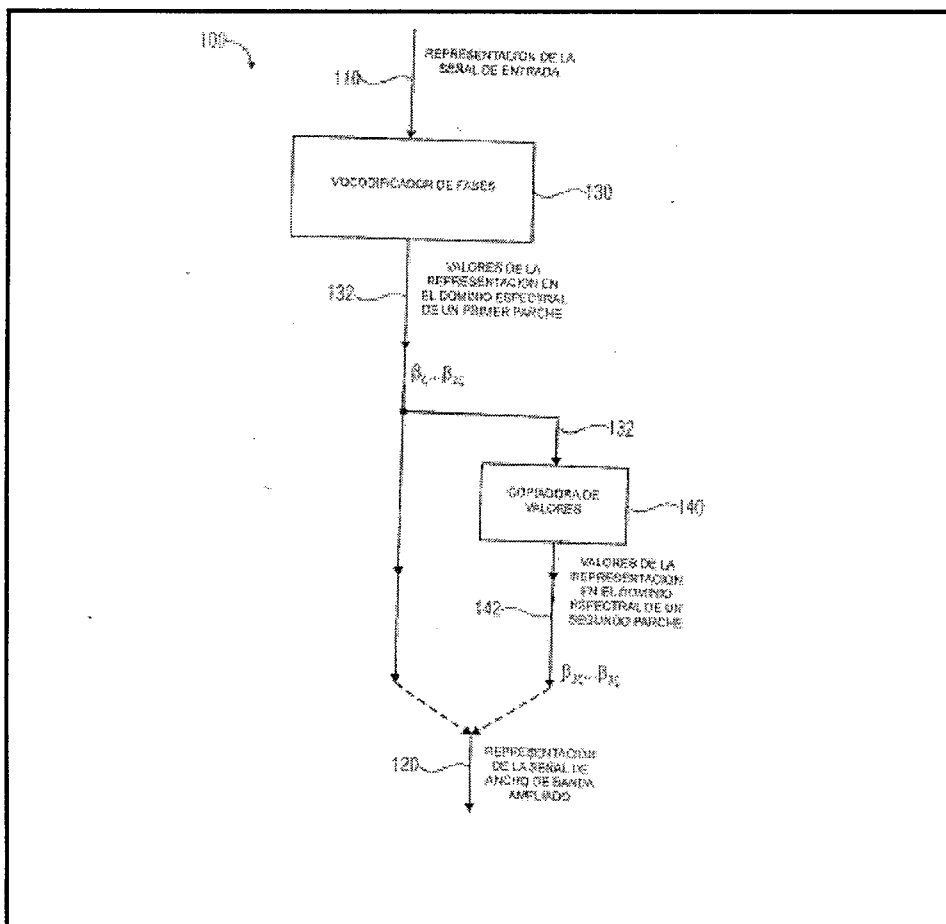
Comprobante de Pago No.: 10-84 203

Fecha 07.10.2010

## 10 ANEXOS

- ☒ Comprobante de pago de la tasa de presentación de la solicitud.
- ☐ Documento que acredita la existencia y representación legal cuando el solicitante sea persona jurídica.
- ☐ Poderes, si fuere el caso.
- ☐ Documento de cesión del inventor al solicitante o a su causante.
- ☒ Traducción de la solicitud internacional al castellano (Resumen, 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=2010112587&IA=EP2010054422&DISPLAY=DOCS>
- ☒ Recibo de Reivindicaciones adicionales

## 11 FIGURA CARACTERISTICA



## 12 Solicito la Concesión de la patente

NOMBRE: **DARIO CARDENAS NAVAS**

FIRMA:

CC. 17066629 BTA TP. 1.250 C.S.J.

LHA  
2

## SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 10-131388-00000-0000

Fecha: 2010-10-22 16:04:02 Dep. 2020 DIR.NUEVASCR  
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI  
Act. 411 PRESENTACION Folios: 206

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800176089-2

CERTIFICACION DE EXCEDENTE : 10 - 86,263

FECHA : OCTUBRE 13 DE 2010

## \*\*\*\*\* CONSIGNACION \*\*\*\*\*

| DEPOSITANTE              | TIPO PAGO    | BANCO           | CUENTA    | No. PAGO  | FECHA PGO  | Vr. PAGO     |
|--------------------------|--------------|-----------------|-----------|-----------|------------|--------------|
| CARDENAS Y CARDENAS P.I. | CONSIGNACION | BANCO DE BOGOTA | 062754387 | 138994973 | 13/10/2010 | 3,533,000.00 |

## \*\*\*\*\* CONCEPTO \*\*\*\*\*

| CANT. | RENTISTICO  | CONCEPTO                           | TOTAL CONCEPTO |
|-------|-------------|------------------------------------|----------------|
| 1     | 50005-01-09 | OTROS SERVICIOS DE PROPIEDAD I     |                |
|       |             | 99 OTROS SERVICIOS ADMINISTRATIVOS | 182,000.00     |
|       |             | TOTAL :                            | \$ 182,000.00  |

SON: CIENTO OCHENTA Y DOS MIL PESOS

RESPONSABLE :

CARDENAS &amp; CARDENAS

ABOGADOS

RECIBO DE CAJA APLICADO AL EXPEDIENTE No.

ESTA CERTIFICACION DE EXCEDENTE DE PAGO SOLO PODRA SER UTILIZADA  
COMO PARTE DE LA CANCELACION DE UNA TARIFA VIGENTE Y NO PODRA SER  
SUSTITUIDA POR NINGUN RECIBO.

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800176089-2

RECIBO OFICIAL DE CAJA : 10 - 84,203  
FECHA : OCTUBRE 7 DE 2010

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 10-131388- -00000-0000

Fecha: 2010-10-22 16:04:02 Dep. 2020 DIR.NUEVASCR  
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI  
Act. 411 PRESENTACION Folios: 206

\*\*\*\*\* CONSIGNACION \*\*\*\*\*

| DEPOSITANTE              | TIPO PAGO    | BANCO           | CUENTA    | No. PAGO  | FECHA PGO  | Vr. PAGO     |
|--------------------------|--------------|-----------------|-----------|-----------|------------|--------------|
| CARDENAS Y CARDENAS P.I. | CONSIGNACION | BANCO DE BOGOTA | 062754387 | 138994951 | 07/10/2010 | 8,104,000.00 |

\*\*\*\*\* CONCEPTO \*\*\*\*\*

| CANT.   | RENTISTICO  | CONCEPTO    | TOTAL CONCEPTO                                 |
|---------|-------------|-------------|------------------------------------------------|
| 1       | 50005-01-01 | SOLICITUDES | 1 TRAMITES DE SOL. DE PATENTE DE IN 554,000.00 |
| TOTAL : |             |             | \$ 554,000.00                                  |

SDN: QUINIENTOS CINCUENTA Y CUATRO MIL PESOS

RESPONSABLE :

RECIBO DE CAJA APLICADO AL EXPEDIENTE No.



**Industria y Comercio**

**SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO**

**DIVISION DE NUEVAS CREACIONES**

**TARJETA ARCHIVO PROPIETARIOS**

PATENTE DE INVENCION \_\_\_\_ PCT X MODELO DE UTILIDAD \_\_\_\_ DISEÑO INDUSTRIAL \_\_\_\_

**SUPERINTENDENCIA**

(21) N° de solicitud:

(22) Fecha de solicitud:

(51) Clasificación Internacional :

(71) Solicitante(s) **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.**

(72) Inventor(es) **Frederik Nagel, Max Neuendorf, Nikolaus Rettelbach, Jeremie Lecomte, Markus Multrus, Berhard Grill**

(74) Apoderado: **DARIO CARDENAS NAVAS**

| (30) Prioridad | (31) No. Prioridad | (32) Fecha | (33) País |
|----------------|--------------------|------------|-----------|
|                | 61/166,125         | 02.04.2009 | US        |
|                | 61/168,068         | 09.04.2009 | US        |
|                | 09181008.5         | 30.12.2009 | EP        |

(85) Fecha inicio fase nacional:

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

Fecha de presentación de la solicitud: 01.04.2010

No. de solicitud **PCT/EP2010/054422**

(54) Título **APARATO, MÉTODO Y PROGRAMA DE COMPUTACIÓN PARA GENERAR UNA REPRESENTACIÓN DE UNA SEÑAL DE ANCHO DE BANDA AMPLIADO EN BASE A LA REPRESENTACIÓN DE UNA SEÑAL DE ENTRADA UTILIZANDO UNA COMBINACIÓN DE UNA AMPLIACIÓN DE ANCHO DE BANDA ARMÓNICO Y UNA AMPLIACIÓN DE ANCHO DE BANDA NO ARMÓNICO.**



**Industria y Comercio**

**SUPERINTENDENCIA**

**SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO**

**DIVISION DE NUEVAS CREACIONES**

**TARJETA ARCHIVO TEMATICO**

PATENTE DE INVENCION

PCT X

MODELO DE UTILIDAD

DISEÑO INDUSTRIAL

(21) N° de solicitud: (22) Fecha de solicitud:  
(51) Clasificación Internacional:  
(71) Solicitante(s) **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.**

(72) Inventor(es) **Frederik Nagel, Max Neuendorf, Nikolaus Rettelbach, Jeremie Lecomte, Markus Multrus, Berhard Grill**

(74) Apoderado: **DARIO CARDENAS NAVAS**

| (30) Prioridad | (31) No. Prioridad | (32) Fecha | (33) País |
|----------------|--------------------|------------|-----------|
|                | 61/166,125         | 02.04.2009 | US        |
|                | 61/168,068         | 09.04.2009 | US        |
|                | 09181008.5         | 30.12.2009 | EP        |

(85) Fecha límite inicio fase nacional:

(86) Datos relativos a la presentación de la solíc. PCT

Fecha de presentación de la solicitud: 01.04.2010

No. de solicitud PCT/EP2010/054422

(87) Publicación internacional:

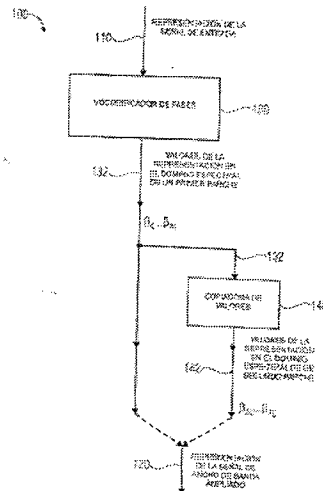
Fecha: **07.10.2010**

No. Publicación **WO 2010/112587**

(54) Título: **APARATO, MÉTODO Y PROGRAMA DE COMPUTACIÓN PARA GENERAR UNA REPRESENTACIÓN DE UNA SEÑAL DE ANCHO DE BANDA AMPLIADO EN BASE A LA REPRESENTACIÓN DE UNA SEÑAL DE ENTRADA UTILIZANDO UNA COMBINACIÓN DE UNA AMPLIACIÓN DE ANCHO DE BANDA ARMÓNICO Y UNA AMPLIACIÓN DE ANCHO DE BANDA NO ARMÓNICO.**

(57) Resumen:

Un aparato para generar una representación de una señal de ancho de banda ampliado en base a la representación de una señal de entrada que comprende un vocodificador de fase configurado para obtener valores de una representación en el dominio espectral de un primer parche de ancho de banda ampliado sobre la base de la representación de la señal de entrada. El aparato también comprende una copiadora de valores configurada para copiar una serie de valores de la representación en el dominio espectral del primer parche, valores son provistos por el vocodificador de fase, para obtener valores de una representación del dominio espectral de un segundo parche, donde el segundo parche está asociado a frecuencias más elevadas que el primer parche. El aparato está configurado para obtener la representación de la señal de ancho de banda ampliado empleando los valores de la representación en el dominio espectral del primer parche y los valores de la representación en el dominio espectral del segundo parche.



## **ASSIGNMENT**

WHEREAS, I/We, Frederik NAGEL, Markus MULTRUS, Jeremie LECOMTE, Stefan BAYER, Guillaume FUCHS, Johannes HILPERT, and Julien ROBILLIARD, hereinafter referred to as "ASSIGNOR(S)", have invented certain new and useful improvements, as described and set forth in the below-identified application for United States Letters Patent entitled:

### **SWITCHING BETWEEN DIFFERENT PATCHING ALGORITHMS IN FREQUENCY DOMAIN**

Filing Date: April 09, 2009 Serial No. 61/168,068

WHEREAS, Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e.V., a corporation duly organized under and pursuant to the laws of the country of Germany, having its principal place of business at Hansastrasse 27c, 80686 Muenchen, Germany, hereinafter referred to as "ASSIGNEE", is desirous of acquiring the entire right, title, and interest in the said invention and application and in any Letters Patent which may be granted with regard to the same;

NOW, THEREFORE, TO ALL WHOM IT MAY CONCERN: Be it known that, for One Dollar (\$1.00) and other good and valuable consideration, ASSIGNOR(S) have sold, assigned, and transferred, and by these presents does sell, assign, and transfer unto the said ASSIGNEE, and ASSIGNEE'S successors and assigns, all right, title, and interest in and to said invention, said application for United States Letters Patent and any Letters Patent which may be hereafter granted on the same in the United States and all countries throughout the world, including any divisions, renewals, continuations in whole or part, substitutions, conversions, reissues, revivals, prolongation, or extensions thereof, said interest to be held and enjoyed by said ASSIGNEE as fully and exclusively as it would have been held and enjoyed by said ASSIGNOR(S) had this assignment and transfer not been made, for all time.

ASSIGNOR(S) further agrees that they will, without charge to said ASSIGNEE, but at ASSIGNEE'S expense, cooperate with ASSIGNEE in the prosecution of said application and/or applications, execute, verify, acknowledge, and deliver all such further papers, including applications for Letters Patent and for the reissue thereof, and instruments of assignment and transfer thereof, and will perform such other acts as ASSIGNEE may lawfully request, to obtain or maintain Letters Patent for said invention and improvement in any and all countries, and to vest title thereto in said ASSIGNEE, or ASSIGNEE'S successors and assigns.

US Patent Application Serial No: 61/168,068

Attorney Docket No: SCH0548PR

ASSIGNOR(S) grants the attorney of record the power to insert on this Assignment any further identification that may be necessary or desirable in order to comply with the rules of the United States Patent and Trademark Office for recordation of this document.

This Assignment may be executed in one or more counterparts, each of which shall be deemed an original and all of which may be taken together as one and the same Assignment.

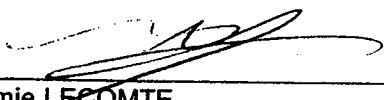
IN TESTIMONY WHEREOF, ASSIGNOR(S) has hereunto signed his name to the assignment on the date indicated below.

  
Frederik NAGEL

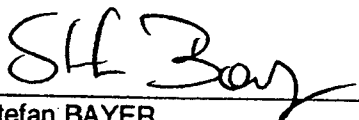
28.4.09  
Date

  
Markus MULTRUS

28.4.2009  
Date

  
Jeremie LECOMTE

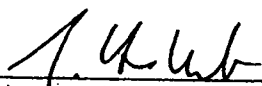
04.05.09  
Date

  
Stefan BAYER

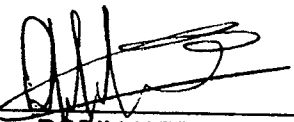
06.05.09  
Date

  
Guillaume FUCHS

28.04.09  
Date

  
Johannes HILPERT

29.4.09  
Date

  
Julien ROBILLIARD

29.04.09  
Date

# Memoria Descriptiva

de la

# Patente de Invención

## Sobre

*"APARATO, MÉTODO Y PROGRAMA DE COMPUTACIÓN PARA  
GENERAR UNA REPRESENTACIÓN DE UNA SEÑAL DE ANCHO DE  
BANDA AMPLIADO EN BASE A LA REPRESENTACIÓN DE UNA SEÑAL  
DE ENTRADA UTILIZANDO UNA COMBINACIÓN DE UNA AMPLIACIÓN  
DE ANCHO DE BANDA ARMÓNICO Y UNA AMPLIACIÓN DE ANCHO DE  
BANDA NO ARMÓNICO"*

## Solicitada por:

FRAUNHOFER-GESELLSCHAFT ZUR  
FÖRDERUNG DER ANGEWANDTEN  
FORSCHUNG E.V., residente en Hansastrasse  
27c, 80686 MUNICH,  
ALEMANIA



**APARATO, MÉTODO Y PROGRAMA DE COMPUTACIÓN PARA  
GENERAR UNA REPRESENTACIÓN DE UNA SEÑAL DE ANCHO DE  
BANDA AMPLIADO EN BASE A LA REPRESENTACIÓN DE UNA SEÑAL DE  
ENTRADA UTILIZANDO UNA COMBINACIÓN DE UNA AMPLIACIÓN DE  
ANCHO DE BANDA ARMÓNICO Y UNA AMPLIACIÓN DE ANCHO DE  
BANDA NO ARMÓNICO**

**Descripción**

Campo Técnico

10 Las realizaciones de acuerdo con la invención se relacionan con un  
aparato para generar una representación de una señal de ancho de banda  
ampliado sobre la base de la representación de una señal de entrada. Otras  
realizaciones de acuerdo con la invención se relacionan con un método para  
generar una representación de una señal de ancho de banda ampliado que se  
15 basa en la representación de una señal de entrada. Otras realizaciones de  
acuerdo con la invención se relacionan con un programa de computación para  
poner en práctica dicho método.

Algunas realizaciones de acuerdo con la invención se relacionan con  
novedosos métodos de introducción de parches dentro de la replicación de la  
20 banda espectral.

Antecedentes de la invención

El almacenamiento o transmisión de señales de audio con frecuencia es  
objeto de rigurosas restricciones de velocidad de transmisión de bits. Estas  
restricciones con frecuencia se superan mediante la codificación de la señal. En  
25 el pasado, se forzaba a los codificadores a reducir drásticamente el ancho de  
banda del audio transmitido cuando sólo se disponía de una velocidad de  
transmisión de bits muy baja. Los códecs (codificador-decodificador) de audio  
modernos pueden codificar las señales de banda ancha utilizando métodos de  
ampliación del ancho de banda (BWE). Esos métodos han sido descriptos, por



ejemplo, en las referencias [1] a [12]. Estos algoritmos se basan en una representación paramétrica del contenido de alta frecuencia (HF), que se genera desde la parte de baja frecuencia (LF) con codificación de forma de onda de la señal decodificada por medio de la transposición a la región espectral de HF  
5 ("introducción de parche") y la aplicación de un post procesamiento promovido por los parámetros.

En la técnica, se utilizan métodos de ampliación del ancho de banda tales como replicación de la banda espectral (SBR) como método eficiente para generar señales de alta frecuencia en los códecs basados en HFR  
10 (reconstrucción de alta frecuencia).

La replicación de la banda espectral (SBR) descrita en la referencia [1], a la que también se hace referencia en forma breve como "SBR", utiliza un banco de filtros espejo en cuadratura (QMF) para generar la información de HF (alta frecuencia). Con la ayuda del denominado proceso de "patching"  
15 (introducción de parches), las señales de banda QMF más bajas se copian a una posición (frecuencia) más elevada, lo que lleva a la replicación de la información de la parte de LF en la parte de HF. La parte de HF generada se adapta posteriormente a la parte de HF original con la ayuda de parámetros que adaptan (o ajustan) la envolvente espectral y la tonalidad (por ejemplo,  
20 utilizando el formateo de la envolvente).

En la SBR standard, la introducción de parches siempre se lleva a cabo mediante una operación de copia dentro del dominio de QMF. Se ha descubierto que esto puede llevar, en ocasiones, a artefactos auditivos, especialmente si se copian sinusoides cerca de otras en el borde de la LF y la parte de HF  
25 generada. Por consiguiente, se puede afirmar que la SBR standard tiene el problema de generar artefactos auditivos. Asimismo, algunas implementaciones convencionales del concepto de ampliación de ancho de banda traen aparejada



una complejidad comparativamente elevada. Más aun, en algunas implementaciones de la invención de los conceptos de ampliación del ancho de banda, el espectro se torna muy ralo en el caso de los parches altos (factores de extensión elevados), lo que puede dar lugar a distorsiones de audio incompatibles (audibles).

En vista de la explicación que antecede, un objetivo de la presente invención consiste en crear un concepto para generar la representación de una señal de ampliación de ancho de banda sobre la base de la representación de una señal de entrada, lo que trae aparejada una relación mejorada entre complejidad y calidad de audio.

#### Síntesis de la invención

Las realizaciones de acuerdo con la invención crean un aparato para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada. El aparato comprende un vocodificador de fases configurado para obtener valores de una representación del dominio espectral de un primer parche de la señal de ancho de banda ampliado sobre la base de la representación de la señal de entrada. El aparato comprende además una copiadora de valores para copiar una serie de valores de la representación del dominio espectral del primer parche, valores que son aportados por el vocodificador de fase, para obtener una serie de valores de una representación del dominio espectral de un segundo parche. El segundo parche está asociado a frecuencias más elevadas que el primer parche. El aparato está configurado para obtener la representación de la señal de ancho de banda ampliado usando los valores de la representación del dominio espectral del primer parche y los valores de la representación del dominio espectral del segundo parche.



El concepto clave de la presente invención es que se obtiene una relación particularmente favorable entre la complejidad informática y la calidad de audio de una señal de ancho de banda ampliado combinando un vocodificador en fase con una copiadora de valores, de tal manera que el primer parche de la señal de ancho de banda ampliado sea obtenida por el vocodificador en fase y que el  
5 segundo parche de la señal de ancho de banda ampliado se obtenga sobre la base del uso de la copiadora de valores por el primer parche. En consecuencia, el contenido del primer parche es una versión transpuesta armónicamente del contenido de la parte de baja frecuencia (LF) de la señal de entrada  
10 (representada por la representación de la señal de entrada), y el segundo parche es (o representa) una versión con desplazamiento de frecuencia (de manera no armónica) del contenido de la señal del primer parche. En consecuencia, se puede obtener el segundo parche con una complejidad informática relativamente baja, puesto que la copia de los valores es más sencilla, desde le punto de vista informático, que una operación de  
15 vocodificación de fases. Además, se evita que haya grandes agujeros espectrales en el segundo parche, puesto que los valores espectrales del primer parche por lo general están suficientemente cargados de datos (es decir, comprenden valores que no son cero), de tal manera que se reduzcan o eviten las distorsiones audibles, que se producirían en algunos casos, si el segundo  
20 parche estuviera sólo escasamente cargado de datos.

Para resumir, el concepto de la invención trae aparejadas considerables ventajas con respecto a los métodos de introducción convencionales de  
25 parches, porque la ampliación armónica del ancho de banda, se aplica sólo para la obtención de valores de representación del dominio espectral del primer parche, es decir, para la parte inferior del espectro, en tanto que la ampliación



no armónica del ancho de banda, que se basa en la copia de valores de la representación del dominio espectral del primer parche para obtener valores de la representación del dominio espectral del primer parche, es utilizada para frecuencias superiores. En consecuencia, el rango inferior (que también se  
5 denomina "primer parche" de la porción de frecuencia de la ampliación (que es una porción de frecuencia por encima de la frecuencia de cruce) se presenta como ampliación armónica del rango de la frecuencia fundamental (es decir, en el rango de frecuencia de la señal de entrada, que cubre frecuencias inferiores a las frecuencias de la porción de frecuencia de ampliación, por ejemplo  
10 frecuencias por debajo de la frecuencia de cruce), lo que da origen a una buena impresión auditiva de la señal de ancho de banda ampliado. Además, se ha encontrado que la simple generación de valores de la representación en el dominio espectral del rango más elevado de la porción de frecuencias de ampliación (lo que también se denomina "segundo parche"), que se realiza  
15 utilizando la copiadora, no trae aparejadas distorsiones auditivas, puesto que el oído humano no es particularmente sensible a los detalles espectrales del rango más elevado de la porción de frecuencias de ampliación (segundo parche).

En resumen, el concepto de la presente invención da origen a una buena  
20 impresión auditiva con una complejidad informática comparativamente baja.

En una realización preferida el vocodificador de fases está configurado para copiar una serie de valores de magnitud asociados con una pluralidad de subrangos de frecuencia dados de la representación espectral de entrada, para  
25 obtener una serie de valores de magnitud asociados con los subrangos de frecuencias correspondientes del primer parche, donde un par de un subrango de frecuencia dado de la representación espectral de entrada y un



correspondiente subrango de frecuencias del primer parche abarca (o comprende) un par integrado por una frecuencia fundamental y u armónico de la frecuencia fundamental (por ejemplo un primer armónico de la frecuencia fundamental). El vocodificador de fases también está configurado, preferentemente, para multiplicar los valores de fases asociados con la pluralidad de subrangos de frecuencia dados de la representación espectral de entrada por un factor predeterminado (por ejemplo 2), para obtener valores de fases asociados con los correspondientes subrangos de frecuencias del primer parche. De preferencia, la copiadora de valores está configurada para copiar una serie de valores asociados a una pluralidad de subrangos de frecuencia dados del primer parche, para obtener una serie de valores asociados a los correspondientes subrangos de frecuencia del segundo parche. La copiadora de valores está configurada preferentemente para dejar los valores de fases inalterados en el copiado. En consecuencia, el vocodificador de fases ejecuta, por lo menos aproximadamente, una transposición armónica, en tanto que la copiadora de valores ejecuta un desplazamiento no armónico de frecuencia. Los subrangos de frecuencias pueden ser, por ejemplo, rangos de frecuencias asociados a los coeficientes de una Transformada Rápida de Fourier (o cualquier transformada comparable). Por otro lado, los subrangos de frecuencias pueden ser rangos de frecuencias asociados a señales individuales de un banco de filtros QMF. Por lo general, el ancho de los subrangos de frecuencias es comparativamente pequeño en comparación con la frecuencia central, por lo que los subrangos de frecuencias cubren un intervalo de frecuencias que tiene una relación de frecuencias entre una frecuencia de partida, que es significativamente inferior a 2:1. En otras palabras, aunque los subrangos de frecuencias de la representación espectral de entrada (que puede tomar la forma, por ejemplo, de coeficientes FFT, o la forma de señales de bancos de



filtros QMF) y no es indispensable que los subrangos de frecuencias del primer parche sean exactamente armónicos entre sí, por lo general es posible identificar una asociación entre un subrango de frecuencias (por ej., con un índice de frecuencia  $k$ ) de la representación espectral de entrada y un  
5 correspondiente subrango de frecuencias (por ej., con un índice de frecuencia  $2k$ ) del primer parche, por lo que el subrango de frecuencias ( $2k$ ) del primer parche representa, por lo menos aproximadamente, una frecuencia armónica del subrango de frecuencias correspondiente ( $k$ ) de la representación espectral de entrada.

10

En consecuencia, el vocodificador de fases ejecuta una transposición armónica tomando en cuenta los valores de fases, que se procesan utilizando un escalamiento de fase. Por el contrario, la copiadora de valores meramente realiza (por lo menos de manera aproximada) una operación de desplazamiento  
15 no armónico de frecuencia.

En una realización preferida, la copiadora de valores está configurada para copiar los valores, de manera tal que se obtiene un desplazamiento espectral común (o desplazamiento de frecuencia) de los valores del primer  
20 parche sobre los valores del segundo parche.

En una realización preferida, el vocodificador de fases está configurado para obtener los valores de la representación del dominio espectral del primer parche por lo que los valores de la representación del dominio espectral del  
25 primer parche representan una versión armónicamente convertida con elevación de frecuencia de un rango de la frecuencia fundamental de la representación de la señal de entrada (por ejemplo, un rango de la frecuencia fundamental por



debajo de una frecuencia denominada de cruce). La copiadora de valores está configurada preferentemente para obtener los valores de la representación del dominio espectral del segundo parche de manera tal que los valores de la representación del dominio espectral del segundo parche representan una versión con desplazamiento de frecuencia del primer parche. En consecuencia, se obtienen las ventajas antes explicadas. En particular, la implementación es sencilla y se obtiene a la vez una buena impresión auditiva.

En una realización preferida, el aparato está configurado para recibir datos de audio de entrada modulados por codificación de pulsos (PCM), para obtener datos de audio modulados por codificación de pulsos con reducción del número de muestras. Por añadidura, el aparato está configurado para mostrar en ventana los datos de audio modulado por codificación de pulsos con reducción del número de copias y para convertir o transformar los datos de entrada colocados en ventana en un dominio de la frecuencia, a fin de obtener la representación de la señal de entrada. El aparato también está configurado preferentemente para computar valores de magnitud  $a_k$  (los que también se designan con la denominación  $\alpha_k$ ) y valores de fase  $\varphi_k$ , que representan un colector de frecuencia  $k$  (donde  $k$  es un índice de colector de frecuencia) de la representación de la señal de entrada, y para copiar los valores de magnitud valores de magnitud  $a_k$ , a fin de obtener valores de magnitud copiados  $a_{sk}$  (también designados con la notación  $\alpha_{sk}$ ) que representan un colector de frecuencia con un índice de colector de frecuencia  $sk$  del primer parche, donde  $s$  es un factor de extensión con  $s=2$ . Más aun, el aparato está configurado preferentemente para copiar y escalar los valores de fase  $\varphi_k$  asociados a un colector de frecuencia con un índice de colector de frecuencia  $k$  de la representación de la señal de entrada, para obtener valores de fase copiados y escalados  $\varphi_{sk}$  asociados con un colector de frecuencia que tiene un índice de



frecuencia  $s_k$  del primer parche. Además, el aparato está configurado preferentemente para copiar los valores  $\beta_{k-\zeta}$  asociados con un colector de frecuencia  $k-i\zeta$  de la representación del dominio espectral del primer parche, para obtener los valores  $\beta_k$  de la representación del dominio espectral del  
5 segundo parche. Por añadidura, el aparato está configurado preferentemente para convertir la representación de la señal de ancho de banda ampliado (que comprende la representación del dominio espectral del primer parche y la representación del dominio espectral del segundo parche) al dominio del tiempo, para obtener una representación en el dominio del tiempo, y para aplicar una  
10 ventana de síntesis a la representación en el dominio del tiempo. Haciendo uso del concepto antes descrito, es posible obtener una señal de ancho de banda ampliado con moderada complejidad informática. La ampliación del ancho de banda se realiza en el dominio de la frecuencia, en el cual se puede ejecutar una transformada en el dominio espectral, por ejemplo en un dominio FFT o un  
15 dominio QMF.

En una realización preferida, el aparato comprende un convertidor del dominio del tiempo al dominio espectral (por ejemplo, un medio de Transformada Rápida de Fourier o un banco de filtros QMF) configurado para producir, como representación de la señal de entrada, valores de una  
20 representación en el dominio espectral (por ejemplo, coeficientes de Transformada Rápida de Fourier o señales de subbandas QMF) de una señal de audio de entrada, o de una versión previamente procesada (por ej. con número de muestras reducida y/o exhibida en ventana) de la señal de audio de entrada (por ejemplo una señal modulada por codificación de pulsos provista por el  
25 núcleo de un decodificador de audio). El aparato comprende preferentemente un convertidor del dominio espectral al dominio del tiempo (por ejemplo, u medio de Transformada Rápida de Fourier inversa o un medio de síntesis de QMF)



configurado para producir una representación en el dominio del tiempo de la señal de ancho de banda ampliado empleando los valores de la representación del dominio espectral (por ej. coeficientes de FFT o señales de subbandas de QMF) del primer parche y los valores de la representación en el dominio espectral (por ej. coeficientes de FFT o señales de subbandas de QMF) del segundo parche. El convertidor del dominio espectral al dominio del tiempo está configurado preferentemente de tal manera que un número de valores espectrales diferentes (por ej. colectores FFT o bandas de QMF) recibido por el by el convertidor del dominio espectral al dominio del tiempo sea mayor que un número de valores espectrales diferentes (por ej. un número de colectores de frecuencia FFT o un número de bandas de QMF) provisto por el dominio del convertidor del dominio del tiempo al espectral (por ej. Medios de Transformada Rápida de Fourier o banco de filtros QMF), por lo que el convertidor del dominio espectral al dominio del tiempo está configurado para procesar un mayor número de colectores de frecuencia (por ej. colectores de frecuencia por Transformada Rápida de Fourier o bandas de frecuencia QMF) que el convertidor del dominio del tiempo al dominio de la frecuencia. En consecuencia, se obtiene una ampliación del ancho de banda por el hecho de que el convertidor del dominio espectral al dominio del tiempo comprende un número mayor de colectores de frecuencia que el convertidor del dominio del tiempo al dominio de la frecuencia.

En una realización preferida, el aparato comprende un generador de ventanas de análisis configurado para colocar en una ventana una señal de audio en el dominio del tiempo, para obtener una versión en ventanas de la señal de audio de entrada en el dominio del tiempo, lo que constituye la base para obtener la representación de la señal de entrada. Además, el aparato comprende un generador de ventanas de síntesis configurado para exhibir en



ventana una porción de una representación en el dominio del tiempo de la señal de ancho de banda ampliado, para obtener una porción exhibida en ventana de la representación en el dominio del tiempo de la señal de ancho de banda ampliado. En consecuencia, se reducen, o incluso se evitan, las distorsiones de la señal de ancho de banda ampliado.

En una realización preferida, el aparato está configurado para procesar una pluralidad de porciones con desplazamiento en el tiempo que se superponen transitoriamente de la señal de audio de entrada en el dominio del tiempo, para obtener una pluralidad de porciones con desplazamiento en el tiempo que se superponen transitoriamente de la representación en el dominio del tiempo de la señal de ancho de banda ampliado. La desviación temporal entre las porciones con desplazamiento en el tiempo que se superponen transitoriamente de la señal de audio de entrada en el dominio del tiempo es menor o igual a un cuarto de la longitud de ventana de una ventana de análisis.

Se ha descubierto que una superposición temporal comparativamente grande entre porciones adyacentes con desplazamiento en el tiempo de la señal de audio de entrada en el dominio del tiempo (y/o una superposición temporal comparativamente grande entre las porciones con desplazamiento temporal adyacentes de la representación en el dominio del tiempo de la señal de ancho de banda ampliado) da lugar a una ampliación del ancho de banda junto con una impresión auditiva favorable, puesto que se toman en cuenta las faltas de estacionaridad de la señal en virtud de la superposición temporal comparativamente grande.

En una realización preferida, el aparato comprende un productor de información transitoria configurado para producir una información que indica la presencia de un transitorio en la señal de entrada (representado por la representación de la señal de entrada). El aparato comprende asimismo una



primera rama de procesamiento para producir una representación de una porción de señal de ancho de banda ampliado sobre la base de una porción no transitoria de la representación de la señal de entrada y una segunda rama de procesamiento para producir una representación de una porción de la señal de ancho de banda ampliado sobre la base de una porción transitoria de la representación de la señal de entrada. La segunda rama de procesamiento está configurada para procesar una representación del dominio espectral de la señal de entrada que tiene una resolución espectral superior a la de una representación en el dominio espectral de la señal de entrada procesada por la primera rama de procesamiento. En consecuencia, las porciones de la señal que comprenden un transitorio pueden ser tratadas con una resolución espectral mayor, lo que evita los artefactos audibles en presencia de transitorios. Por otra parte, se puede utilizar una resolución espectral reducida para las porciones de señal sin transitorios (es decir, para porciones de la señal en las cuales el productor de información sobre transitorios no identifica un transitorio). De esa manera, se mantiene elevada la eficiencia informática y sólo se utiliza la resolución espectral aumentada cuando trae consigo ventajas (por ejemplo, si da lugar a una mejor impresión auditiva en la proximidad de transitorios).

En una realización preferida, el aparato comprende un atenuador a cero (dispositivo de relleno con ceros) configurado para atenuar a cero una porción transitoria de la señal de entrada, a fin de obtener una porción transitoria temporalmente ampliada de la señal de entrada. En este caso, la primera rama de procesamiento comprende un (primer) convertidor del dominio del tiempo al dominio de la frecuencia configurado para producir un primer número de valores del dominio espectral asociados con una primera porción no transitoria de la señal de entrada, y la segunda rama de procesamiento comprende un (segundo) convertidor del dominio del tiempo al dominio de la frecuencia



configurado para producir un segundo número de valores en el dominio espectral asociados con la porción transitoria temporalmente ampliada de la señal de entrada. El segundo número de valores en el dominio espectral es más elevado, por lo menos en un factor de 1,5, que el primer número de valores en el dominio espectral. En consecuencia, se obtiene un buen manejo de los transitorios.

En una realización preferida, la segunda rama de procesamiento comprende un eliminador de ceros configurado para eliminar una pluralidad de valores cero de una porción de señal de ancho de banda ampliado obtenidos sobre la base de la porción transitoria temporalmente ampliada de la señal de entrada. De esa manera se invierte la ampliación temporal de la señal de entrada, que se obtiene en virtud de la pérdida uniforme ajustada a cero.

En una realización preferida, el aparato comprende un reductor del número de muestras configurado para reducir el número de muestras de una representación en el dominio del tiempo de la señal de entrada. Mediante la reducción del número de muestras de la señal de entrada, se puede mejorar la eficiencia informática si la señal de entrada no cubre la totalidad del ancho de banda Nyquist de un flujo de entrada de muestras moduladas por codificación de pulsos.

Otra realización de acuerdo con la presente invención da origen a un aparato, en el cual se invierte el orden de procesamiento del procesamiento realizado por la copiadora de valores y el vocodificador de fases. Ese tipo de aparato para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada (110; 383) comprende una copiadora de valores configurada para copiar una serie de valores de la representación de la señal de entrada, a fin de obtener una serie de valores de una representación en el dominio espectral de un primer



parche, donde el primer parche está asociado a frecuencias superiores a las de la representación de la señal de entrada. El aparato comprende además un vocodificador de fases (130; 406) configurado para obtener valores ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ ) de una representación en el dominio espectral de un segundo parche de la señal de ancho de banda ampliado sobre la base de los valores ( $\beta_{4/3\zeta} \dots \beta_{2\zeta}$  de la representación en el dominio espectral del primer parche, donde el segundo parche está asociados a frecuencia superiores a las del primer parche. El aparato está configurado para obtener la representación (120;426) de la señal de ancho de banda ampliado empleando los valores de la representación en el dominio espectral del primer parche y los valores de la representación en el dominio espectral del segundo parche.

Este aparato tiene la capacidad de obtener una señal de ancho de banda ampliado con una complejidad informática comparativamente baja, y obtener a la vez una buena impresión auditiva de la señal de ancho de banda ampliado. Al realizar la vocodificación de fases después de la operación de copiado, el vocodificador de fases puede ser operado con una relación de frecuencia comparativamente pequeña (la relación entre la frecuencia de salida del vocodificador y la frecuencia de entrada del vocodificado), lo que da lugar a un buen llenado espectral y evita la presencia de grandes agujeros espectrales. Además, se ha encontrado que la impresión auditiva, utilizando este concepto, es aun mejor que en el caso de un concepto que se basa meramente en operaciones de copiado sin la acción de un vocodificador de fases, aunque el primer parche (parche de frecuencia más baja) se obtenga utilizando la operación de copiado y sólo se obtenga el segundo parche (parche de frecuencia más elevada) utilizando la operación de vocodificación de fases. Más aun, la complejidad informática es menor que en los sistemas en que todos los



parches se generan empleando vocodificadores de fases, y los agujeros espectrales se reducen en comparación con esos conceptos.

Naturalmente, esta realización puede ser complementada por cualquiera de las funcionalidades descriptas en este documento.

5        Otras realizaciones de acuerdo con la invención introducen métodos para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada. Dicho método se basa en las mismas ideas que el aparato precedentemente descrito.

10       Otra realización de acuerdo con la presente invención produce un programa de computación para implementar el método.

#### Breve Descripción de las Figuras

15       La Fig. 1       ilustra un diagrama esquemático de bloques de un aparato para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada, de acuerdo con una realización de la invención;

La Fig. 2       ilustra una representación esquemática del concepto de ampliación del ancho de banda de acuerdo con la presente invención;

20       La Fig. 3       ilustra un diagrama esquemático de bloques detallado de un decodificador de audio que comprende un aparato para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada, de acuerdo con una realización de la invención;

25       La Fig. 4       ilustra un gráfico de flujo de un método para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada, de acuerdo con una realización de la invención;



La Fig. 5 ilustra un diagrama esquemático de bloques de un decodificador de audio, de acuerdo con un primer ejemplo comparativo y

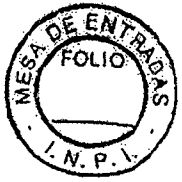
La Fig. 6 ilustra un diagrama esquemático de bloques de un decodificador de audio, de acuerdo con un segundo ejemplo comparativo.

5

### Descripción Detallada de las Realizaciones

#### 1. Aparato de acuerdo con la Fig. 1

La Fig. 1 ilustra un diagrama esquemático de bloques de un aparato 100 para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada. El aparato 100 está configurado para recibir la representación de una señal de entrada 110 y producir, sobre la base de ésta, una señal de ancho de banda ampliado 120. El aparato 100 comprende un vocodificador de fases configurado para obtener los valores de una representación del dominio espectral 130 de un primer parche de la señal de ancho de banda ampliado 120 sobre la base de la representación de la señal de entrada 110. Los valores de la representación en el dominio espectral del primer parche se designan, por ejemplo, con las letras  $\beta_{\zeta}$  a  $\beta_{2\zeta}$ . El aparato 100 comprende además una copiadora de valores 140 configurada para copiar una serie de valores de la representación del dominio espectral 132 del primer parche, que son aportados por el vocodificador de fases 130, para obtener una serie de valores de una representación en el dominio espectral 142 de un segundo parche, donde el segundo parche está asociado a frecuencias mayores que el primer parche. Los valores de la representación en el dominio espectral 142 del segundo parche se designan, por ejemplo, con las letras  $\beta_{2\zeta}$  a  $\beta_{3\zeta}$ . El aparato 100 está configurado para obtener la representación 120 de la señal de ancho de banda ampliado utilizando los valores  $\beta_{\zeta}$  a  $\beta_{2\zeta}$  de la representación en el dominio espectral 132 del primer parche y los valores  $\beta_{2\zeta}$  a



$\beta_{3z}$  de la representación en el dominio espectral 142 del segundo parche. Por ejemplo, la representación 120 de la señal de ancho de banda ampliado puede comprender tanto los valores de la representación en el dominio espectral 132 del primer parche como la representación en el dominio espectral 142 del segundo parche. Más aun, la representación 120 de la señal de ancho de banda ampliado puede comprender, por ejemplo, los valores de una representación en el dominio espectral de la señal de entrada (representados, por ejemplo, por la representación de la señal de entrada 110). Sin embargo, la representación 120 de la señal de ancho de banda ampliado también puede ser una representación en el dominio del tiempo, que se puede basar en los valores de la representación en el dominio espectral 132 del primer parche y los valores de la representación en el dominio espectral 142 del segundo parche (y, opcionalmente, otros valores, como por ejemplo los valores de la representación en el dominio espectral 116 de la señal de entrada, y/o los valores de una representación en el dominio espectral de parches adicionales).

A continuación se describe la funcionalidad y funcionamiento del aparato 100 en forma detallada, tomando como referencia la Fig. 2, que ilustra una representación esquemática del concepto de la invención para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada.

Una primera representación gráfica 200 ilustra una transposición armónica de la señal de entrada (representada por la representación de la señal de entrada 110), que es ejecutada por el vocodificador de fases 130. Como se puede apreciar, la señal de entrada está representada, por ejemplo, por una serie de valores de magnitud  $\alpha_k$ . El índice  $k$  designa un colector espectral (por ejemplo un colector que tiene un índice  $k$  de transformada rápida de Fourier, o una banda de frecuencias que tiene un índice  $k$  de una conversión por QMF). La



representación de la señal de entrada 110 puede comprender, por ejemplo, los valores de magnitud  $\alpha_k$  correspondientes a  $k = 1$  a  $k = \zeta$ , donde  $\zeta$  may designa un colector de frecuencia de la denominada frecuencia de cruce y describe una frecuencia inicial de la ampliación de ancho de banda. También se describe un

5 rango de la frecuencia fundamental, por ejemplo, mediante los valores de fases  $\varphi_k$ , donde  $k$  es un índice de colector de frecuencia, como se señalara anteriormente.

De manera similar, el primer parche está indicado por una serie de valores de una representación en el dominio espectral, por ejemplo, valores  $\beta_k$

10 con  $k$  de entre  $\zeta$  y  $2\zeta$ . Por otro lado, el primer parche puede estar representado por valores de magnitud  $\alpha_k$  y valores de fases  $\varphi_k$ , con el índice de colector de frecuencia  $k$  de entre  $\zeta$  y  $2\zeta$ .

Como ya se mencionara, el vocodificador de fases 130 está configurado para ejecutar una transposición armónica sobre la base de la representación de

15 la señal de entrada 110 para obtener los valores de la representación del dominio espectral 132 del primer parche. Para lograr este fin, el vocodificador de fases 130 puede establecer un valor de magnitud  $\alpha_{2k}$  de un colector de frecuencia con un índice de colector de frecuencia (colector de frecuencia)  $2k$  que es igual al valor de magnitud  $\alpha_k$  de un colector de frecuencia que tiene el

20 índice de colector de frecuencia (colector de frecuencia)  $k$ . Además, el vocodificador de fases 130 puede estar configurado para establecer el valor de fase  $\varphi_{2k}$  de un colector de frecuencia que tiene un índice  $2k$  en un valor que es igual al doble del valor de fase  $\varphi_k$  asociado con el colector de frecuencia que tiene el índice  $k$ . En este caso, el colector de frecuencia que tiene el índice  $k$

25 puede ser un colector de frecuencia de la representación de la señal de entrada 110, y el colector de frecuencia que tiene el índice  $2k$  puede ser un colector de frecuencia de la representación del dominio espectral 132 del primer parche.



Además, un colector de frecuencia que tiene el índice  $2k$  puede comprender una frecuencia que es un primer armónico de una frecuencia incluida en el colector de frecuencia que tiene el índice  $k$ . De esa manera se pueden obtener valores de magnitud  $\alpha_{2k}$  y valores de  $\varphi_{2k}$ , que son los valores de la representación en el dominio espectral 132 del primer parche, correspondientes a  $2k$  en el rango de  $\zeta$  a  $2\zeta$ , tal como  $\alpha_{2k} = \alpha_k$  y  $\varphi_{2k} = 2\varphi_k$ . Por otro lado, y de manera equivalente, se pueden obtener valores  $\beta_{2k}$ , que son los valores de la representación del dominio espectral 132 del primer parche, correspondientes a  $2k$  de entre  $\zeta$  y  $2\zeta$ , como por ejemplo  $\beta_{2k} = \alpha_k e^{j2\varphi_k}$ .

Para resumir, suponiendo que el colector de frecuencias que tiene índices  $k$  (o, de manera equivalente,  $2k$  y así sucesivamente), que son, por ejemplo, colectores de frecuencia de una representación de Transformada Rápida de Fourier, están espaciados en dirección lineal en la frecuencia (de tal manera que el índice del colector de frecuencia, por ej.  $k$  o  $2k$ , sea por lo menos aproximadamente proporcional a una frecuencia comprendida en el respectivo colector de frecuencia, por ejemplo una frecuencia central de un  $k^\circ$  colector de frecuencia por Transformada Rápida de Fourier o una frecuencia central en la  $k^\circ$  banda QMF), el vocodificador de fases 130 obtiene una transposición armónica.

No obstante, los valores de la representación del dominio espectral 142 del segundo parche son obtenidos por la copiadora de valores 140, que lleva a cabo un copiado no armónico de los valores de la representación del dominio espectral 132 del primer parche.

Tomando ahora como referencia la representación gráfica 250, se describe en forma breve el copiado no armónico. Como se puede ver, el primer parche está representado por los valores  $\beta_\zeta$  a  $\beta_{2\zeta}$  (o, de manera equivalente, por los valores de magnitud  $\alpha_\zeta$  a  $\alpha_{2\zeta}$  y los valores de fase  $\varphi_\zeta$  a  $\varphi_{2\zeta}$ . En consecuencia, los valores  $\beta_{2\zeta}$  a  $\beta_{3\zeta}$  (o, de manera equivalente, los valores de



magnitud  $\alpha_{2\zeta}$  a  $\alpha_{3\zeta}$  y los valores de fase  $\varphi_{2\zeta}$  a  $\varphi_{3\zeta}$ ) de la representación del dominio espectral 142 del segundo parche se obtienen mediante un copiado no armónico, que es ejecutado por la copiadora de valores 140. Por ejemplo, se pueden obtener los valores espectrales valorados de manera compleja  $\beta_{2\zeta}$  a  $\beta_{3\zeta}$  de la representación del dominio espectral 142 del segundo parche sobre la base de los correspondientes valores  $\beta_{\zeta}$  a  $\beta_{2\zeta}$  de la representación del dominio espectral 132 del primer parche de acuerdo con  $\beta_k = \beta_{k-\zeta}$  en el caso de  $k$  entre  $2\zeta$  y  $3\zeta$ . Del mismo modo, se pueden obtener valores de magnitud  $\alpha_{2\zeta}$  a  $\alpha_{3\zeta}$  de la representación del dominio espectral 142 del segundo parche sobre la base de los valores de magnitud de la representación en el dominio espectral 132 del primer parche de acuerdo con  $\alpha_k = \alpha_{k-\zeta}$  en el caso de  $k$  entre  $2\zeta$  y  $3\zeta$ . En este caso, se pueden obtener valores de fase  $\varphi_{2\zeta}$  a  $\varphi_{3\zeta}$  de la representación del dominio espectral 142 del segundo parche sobre la base de los valores de fase  $\varphi_{\zeta}$  a  $\varphi_{2\zeta}$  de la representación del dominio espectral 132 del primer parche de acuerdo con  $\varphi_k = \varphi_{k-\zeta}$  en el caso de  $k$  entre  $2\zeta$  y  $3\zeta$ .

En consecuencia, los valores de la representación del dominio espectral 142 del segundo parche representan una señal, que tiene un desplazamiento de frecuencia de modo no armónico (es decir, lineal) con respecto a una señal representada por los valores de la representación del dominio espectral 132 del primer parche.

Se pueden utilizar los valores  $\beta_{\zeta}$  a  $\beta_{2\zeta}$  de la representación del dominio espectral 132 del primer parche y los valores  $\beta_{2\zeta}$  a  $\beta_{3\zeta}$  de la representación del dominio espectral 142 del segundo parche para obtener la representación 120 de la señal de ancho de banda ampliado. Dependiendo de los requisitos, la representación 120 de la señal de ancho de banda ampliado puede ser una representación del dominio espectral o una representación en el dominio del tiempo. Si se pretende obtener una representación en el dominio del tiempo, se



puede emplear un convertidor del dominio de la frecuencia al dominio del tiempo para derivar la representación en el dominio del tiempo sobre la base de los valores  $\beta_z$  a  $\beta_{2z}$  de la representación del dominio espectral 132 del primer parche y los valores  $\beta_{2z}$  a  $\beta_{3z}$  de la representación del dominio espectral 142 del  
5 segundo parche. Por otro lado (y de manera equivalente), se pueden utilizar los valores  $\alpha_z$  a  $\alpha_{2z}$ ,  $\varphi_z$  a  $\varphi_{2z}$ ,  $\alpha_{2z}$  a  $\alpha_{3z}$  y  $\varphi_{2z}$  a  $\varphi_{3z}$  para derivar la representación 120 de la señal de ancho de banda ampliado (ya sea en el dominio espectral o en el dominio del tiempo).

Como se señalara anteriormente, el concepto descrito con respecto a  
10 las Figs. 1 y 2 trae aparejado una buena impresión auditiva y complejidad informática relativamente baja. La vocodificación de fases sólo es necesaria una vez, aunque se utilice una pluralidad de parches (por ejemplo el primer parche y el segundo parche). Además, se evita la existencia de grandes agujeros espectrales en el segundo parche, lo que ocurriría si se utilizar otro vocodificador  
15 de fase para obtener el segundo parche. Por consiguiente, el concepto de la presente invención trae consigo un compromiso favorable entre complejidad informática y una impresión auditiva.

Más aun, se debe notar que, en algunas realizaciones, se pueden obtener más parches sobre la base de los valores de la representación del dominio  
20 espectral 132 del primer parche. Por ejemplo, en una ampliación opcional del concepto de la invención, se pueden obtener los valores de una representación del dominio espectral de un tercer parche sobre la base de los valores de la representación en el dominio espectral 132 del primer parche empleando otra copiadora de valores, como se describe en forma más detallada tomando como  
25 referencia la Fig. 3.

Las realizaciones de acuerdo con las Figs 1 y 2 (y también las demás realizaciones) se pueden modificar de una amplia variedad de maneras. Por



ejemplo, se puede obtener un primer parche empleando un vocodificador de fases, y se puede obtener un segundo, tercer y cuarto parches mediante una operación de copiado de los valores espectrales. Por otro lado, se puede obtener un primer y segundo parches utilizando vocodificadores de fases, y se  
5 puede obtener un tercero y un cuarto parches utilizando el copiado de valores espectrales. Como es natural, se pueden aplicar diferentes combinaciones de la operación de vocodificación de fases y la operación de copiado.

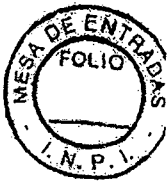
Por otro lado, sin embargo, se puede obtener un primer parche utilizando una operación de copiado (copiadora de valores) de los valores espectrales de  
10 la representación de la señal de entrada, y se puede obtener un segundo parche utilizando un vocodificador de fases (sobre la base de los valores copiados del primer parche, obtenidos de la copiadora de valores).

## 2. Realización de Acuerdo con la Fig. 3

A continuación se describe un decodificador de audio 300 haciendo  
15 referencia a la Fig. 3, donde la Fig. 3 ilustra un diagrama esquemático de bloques detallado de ese tipo de decodificador de audio 300 que comprende un aparato para la generación de una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada.

### 2.1. Reseña General del Decodificador de Audio

20 El decodificador de audio 300 está configurado para recibir un flujo de datos 310 y para producir, basándose en los mismos, una forma de onda de audio 312. El decodificador de audio 300 comprende un decodificador del núcleo 320, que está configurado para aportar, por ejemplo, datos modulados por codificación de pulsos ("datos PCM") 322 sobre la base del flujo de datos 310. El  
25 decodificador del núcleo 320 puede ser, por ejemplo, un decodificador de audio de acuerdo con lo descrito en la norma internacional ISO/IEC 14496-3:2005(e), parte 3: audio, subparte 4: codificación general de audio (GA)-AAC,



Twin VQ, BSAC. Por ejemplo, el decodificador del núcleo 320 puede ser un denominado decodificador del núcleo de codificación de audio avanzado (AAC), al que se describe más detalladamente en dicha normal y que es muy conocido por la persona con capacitación en la técnica. Por consiguiente, los datos de audio modulados por codificación de pulsos 322 pueden ser producidos por el decodificador del núcleo 220 sobre la base del flujo de datos 310. Por ejemplo, los datos de audio modulados por codificación de pulsos 322 pueden comprender una longitud de cuadro de 1024 muestras.

El decodificador de audio 300 comprende además una ampliación del ancho de banda (o amplificador del ancho de banda) 330, que está configurado para recibir los datos de audio modulados por codificación de pulsos 322 (por ejemplo, una longitud de cuadro de 1024 muestras) y para otorgar, sobre la base de las mismas, la forma de onda 312. La ampliación del ancho de banda (o amplificador del ancho de banda) 330 también recibe ciertos datos de control 332 del flujo de datos 310. La ampliación del ancho de banda 330 comprende una provisión de datos de QMF con parches (proveedor de datos QMF con parche) 340, que recibe los datos de audio modulados por codificación de pulsos 322 y que produce, sobre la base de éstos, datos QMF con parches 342. La ampliación del ancho de banda 330 comprende asimismo un formateo de la envolvente (o formateador de la envolvente) 344, que recibe los datos QMF con parches 342 y datos de control de formateo de la envolvente 346 y produce, sobre la base de estos, datos QMF con envolvente formateada 348. La ampliación del ancho de banda 330 comprende además una síntesis QMF (o sintetizador QMF) 350, que recibe los datos QMF con parches y envolvente formateada 348 y produce, sobre la base de estos, la forma de onda 312 mediante la ejecución de una síntesis QMF.

## 2.2. Provisión de datos QMF con Parche 340



### 2.2.1. Provisión de datos QMF con Parche – Reseña general

La provisión de datos QMF con parche 340 (que puede ser ejecutado por un proveedor de datos QMF con parche 340 en una implementación de hardware) puede ser conmutable entre dos modos, es decir un primer modo en el cual se ejecuta la introducción de parches por replicación de la banda espectral (SBR) y un segundo modo en el cual se ejecuta la introducción de parches por ampliación armónica del ancho de banda (HBE). Por ejemplo, los datos de audio modulados por codificación de pulsos 322 pueden ser retardados por un retardador 360, para obtener datos retardados de audio modulados por codificación de pulsos 362, y los datos retardados de audio modulados por codificación de pulsos 362 se pueden convertir al dominio QMF utilizando un analizador QMF de 32 bandas 364. El resultado del analizador QMF de 32 bandas 364, por ejemplo, se puede enviar una representación del dominio QMF de 32 bandas es decir, el dominio espectral) 365 de los datos retardados de audio modulados por codificación de pulsos 362, a un introductor de parches por SBR 366 y a un introductor de parches por ampliación armónica del ancho de banda 368.

El introductor de parches por replicación de la banda espectral 366 puede realizar, por ejemplo, la introducción de un parche por replicación de la banda espectral, que se describe, por ejemplo, en la sección 4.6.18 "SBR tool" de la norma internacional ISO/IEC 14496-3:2005(e), parte 3, subparte 4. En consecuencia, el introductor de parches por replicación de la banda espectral 366 puede otorgar una representación del dominio QMF de 64 bandas.

Por otro lado, o además, el introductor de parches por ampliación armónica del ancho de banda 368 puede aportar una representación del dominio QMF de 64 bandas 372, que es una representación de ancho de banda ampliado de los datos de audio PCM 322. Se puede utilizar un conmutador 374,



al que se controla dependiendo de los datos de control de ampliación del ancho de banda 332 extraídos del flujo de datos 310, para decidir si se aplica la introducción de parches por replicación de la banda espectral 366 o la introducción de parches por ampliación armónica del ancho de banda 368 para  
5 obtener los datos QMF con parche 342 (que pueden ser iguales a la representación del dominio QMF de 64 bandas 370 o iguales a la representación del dominio QMF de 64 bandas 372 dependiendo del estado del conmutador 374).

10 2.2.2. Provisión de datos QMF con parche – Ampliación armónica del ancho de banda 368

En lo sucesivo se describe en forma más detallada la introducción de parches (por lo menos parcialmente) por ampliación armónica del ancho de banda 368. La introducción de parches por ampliación armónica del ancho de banda 368 comprende un trayecto de señales, en la cual los datos de audio  
15 modulados por codificación de pulsos 322, o una versión previamente procesada de los mismos, se convierten a un dominio espectral (por ejemplo a un dominio del coeficiente de Transformada Rápida de Fourier o a un dominio QMF), en el cual se realiza una ampliación armónica del ancho de banda en el dominio espectral, y en el cual se utiliza la representación del dominio espectral  
20 de la señal de ancho de banda ampliado obtenida, o una representación derivada de la misma, para la introducción de parches por ampliación armónica del ancho de banda.

En la realización de la Fig. 3, se reduce el número de muestras de los datos de audio modulados por codificación de pulsos 322 por medio de un  
25 reductor del número de muestras 380, por ejemplo, en un factor de 2, para obtener datos de audio modulados por codificación de pulsos con número de muestras reducido 381. A continuación, los datos de audio modulados por



codificación de pulsos con número de muestras reducido 381 son colocados en ventana por un generador de ventanas 382, que puede comprender, por ejemplo, una longitud de ventana de 512. Se debe tener en cuenta que la ventana tiene un desplazamiento, por ejemplo, de 64 muestras de los datos de audio modulados por codificación de pulsos con número de muestras reducido 381 en los pasos subsiguientes de procesamiento, por lo que se obtiene una superposición comparativamente grande de las porciones encuadradas en ventanas 383 de los datos de audio modulados por codificación de pulsos con número de muestras reducido.

El decodificador de audio 300 comprende además un detector de transitorios 384, que está configurado para detectar un transitorio dentro de los datos de audio modulados por codificación de pulsos 322. El detector de transitorios 384 puede detectar la presencia de un transitorio, ya sea sobre la base de los datos de audio PCM 322 en sí, o sobre la base de una información secundaria, que está incluida en el flujo de datos 310.

Las porciones encuadradas en ventanas 383 de los datos de audio PCM con número de muestras reducido 381 pueden ser procesadas selectivamente empleando una primera rama de procesamiento 386 o una segunda rama de procesamiento 388. La primera rama 386 se puede utilizar para procesar una porción no transitoria encuadrada en ventana 383 de los datos de audio PCM con número de muestras reducido (para lo cual el detector de transitorios 384 niega la presencia de un transitorio) y se puede utilizar una segunda rama 388 para el procesamiento de una porción transitoria encuadrada en ventana 383 de los datos de audio PCM con número de muestras reducido (para lo cual el detector de transitorios 384 indica la presencia de un transitorio).

La primera rama 386 recibe una porción no transitoria encuadrada en ventana 383 y produce, sobre la base de ésta, una representación de ancho de



banda ampliado 387, 434 de la porción encuadrada en ventana 383. De manera similar, la segunda rama 388 recibe una porción transitoria encuadrada en ventana 383 de los datos de audio PCM con número de muestras reducido 381 y produce, sobre la base de estos, una representación de ancho de banda  
5 ampliado 389 de la porción encuadrada en ventana (transitoria) 383. Como se indicara anteriormente, el detector de transitorios 384 decide si la porción encuadrada en ventana actual 383 es una porción no transitoria encuadrada en ventana o una porción transitoria encuadrada en ventana, por lo que el procesamiento de la porción encuadrada en ventana en el momento 383 se  
10 realiza utilizando la primera rama 386 o la segunda rama 388. De esa manera, diferentes porciones encuadradas en ventanas 383 pueden ser procesadas por diferentes ramas 386, donde hay una superposición temporal considerable entre las representaciones sucesivas de ancho de banda ampliado 387, 389 de las sucesivas porciones encuadradas en ventanas 383 (puesto que hay una  
15 superposición temporal considerable de las sucesivas porciones temporariamente encuadradas en ventanas 383).

La ampliación armónica del ancho de banda 368 comprende asimismo un dispositivo de superposición y suma 390, que está configurado para superponer y sumar las diferentes representaciones de ancho de banda ampliado 387, 389  
20 asociadas con diferentes porciones encuadradas en ventana (temporalmente sucesivas) 383. Un incremento de superposición y suma puede ser ajustado, por ejemplo, a 256 muestras. De esa manera se obtiene una señal superpuesta y sumada 392.

La ampliación armónica del ancho de banda 368 comprende además un  
25 analizador QMF de 64 bandas 394, que está configurado para recibir la señal superpuesta y sumada 392 y producir, sobre la base de ésta, una señal en dominio QMF de 64 bandas 396. La señal en dominio QMF de 64 bandas 396



puede representar, por ejemplo, un rango de frecuencias más amplio que la señal en dominio QMF de 32 bandas 365 provista por el analizador QMF de 32 bandas 364.

La ampliación armónica del ancho de banda 368 comprende además un  
5 combinador 398, que está configurado para recibir tanto la señal en dominio QMF de 32 bandas provista por el analizador QMF de 32 bandas 364 como la señal en dominio QMF de 64 bandas 396 y para combinar esas señales. Por ejemplo, los componentes del rango de baja frecuencia (o rango de la frecuencia fundamental) de la señal en dominio QMF de 64 bandas 396 pueden ser  
10 reemplazados, o combinados, con la señal en dominio QMF de 32 bandas 365 provista por el analizador QMF de 32 bandas 364, de tal manera que, por ejemplo, los 32 componentes del rango de frecuencia más baja (o el rango de la frecuencia fundamental) de la señal en dominio QMF de 64 bandas 372 están determinados por la salida del analizador QMF de 32 bandas 364, y de tal  
15 manera que los 32 componentes del rango de frecuencia superior de la señal en dominio QMF de 64 bandas 372 están determinados por los 32 componentes del rango de mayor frecuencia de la señal en dominio QMF de 64 bandas 396.

Naturalmente, el número de componentes de las señales en el dominio QMF puede variar, dependiendo de los requerimientos específicos.  
20 Naturalmente, una posición de frecuencia de una transición entre un rango de la frecuencia fundamental (también denominado rango de frecuencia más baja) y un rango de frecuencias de ancho de banda ampliado (también denominado rango de frecuencias más elevadas) puede depender de la frecuencia de cruce o, de manera equivalente, del ancho de banda de la señal de audio  
25 representada por los datos de audio modulados por codificación de pulsos 322.

En los párrafos siguientes se describen detalles con respecto a la primera rama de procesamiento 386. La primera rama 386 comprende un convertidor del



- dominio del tiempo al dominio de la frecuencia 400, que está implementado, por ejemplo, en la forma de un medio de Transformada Rápida de Fourier configurado para producir 512 coeficientes de Transformada Rápida de Fourier sobre la base de una porción encuadrada en ventana 383 de 512 muestras en el
- 5 dominio del tiempo de los datos de audio modulados por codificación de pulsos con número de muestras reducido 381. En consecuencia, los colectores de frecuencia por Transformada Rápida de Fourier se designan con índices de números enteros  $k$  de colectores de frecuencia en el rango comprendido entre 1 y  $N=512$ .
- 10 La primera rama 386 comprende además un dispositivo para producir valores de magnitud 402, que está configurado para producir valores de magnitud  $\alpha_k$  de los coeficientes de Transformada Rápida de Fourier. Además, la primera rama 386 comprende un dispositivo para producir valores de fase 404 configurado para proporcionar valores de fase  $\varphi_k$  de los coeficientes de
- 15 Transformada Rápida de Fourier.
- La primera rama 386 comprende además un vocodificador de fases 406, que puede recibir los valores de magnitud  $\alpha_k$  y los valores de fase  $\varphi_k$  como representación de una señal de entrada, y que puede comprender la funcionalidad del vocodificador de fases 130 antes descripto. En consecuencia,
- 20 el vocodificador de fases 406 puede dar salida a valores  $\beta_{2k}$ , en un rango comprendido entre  $\beta_\xi$  y  $\beta_{2\xi}$ , de una representación en el dominio espectral de un primer parche. Los valores  $\beta_{2k}$  se designan con 408, y pueden ser equivalentes a los valores de la representación del dominio espectral 132 de un primer parche. La primera rama 386 comprende además una copiadora de valores 410,
- 25 que puede asumir la funcionalidad de la copiadora de valores 140, y que puede recibir, como información de entrada, los valores  $\beta_{2k}$  (por ej. en un rango comprendido entre  $\beta_\xi$  y  $\beta_{2\xi}$ ). En consecuencia, la primera copiadora de valores



410 puede proporcionar valores  $\beta_k$  en un rango de  $\beta_{2\xi}$  a  $\beta_{3\xi}$ , los que se designan con 412 y que pueden ser equivalentes a los valores  $\beta_{2\xi}$  a  $\beta_{3\xi}$  de la representación del dominio espectral 142 del segundo parche. Asimismo, la primera rama 386 puede comprender (opcionalmente) una segunda copiadora de valores 414, que está configurada para recibir los valores  $\beta_\xi$  y  $\beta_{2\xi}$  (que también se designan con 408) provistos por el vocodificador de fases 406 y para proporcionar, sobre la base de estos, valores espectrales  $\beta_{3\xi}$  a  $\beta_{4\xi}$  usando una operación de copiado (que da lugar con eficacia a un desplazamiento de frecuencia no armónico del espectro descrito por los valores  $\beta_\xi$  a  $\beta_{2\xi}$  (408)). En consecuencia, la segunda copiadora de valores 414 aporta valores espectrales  $\beta_{3\xi}$  a  $\beta_{4\xi}$  de una representación del dominio espectral de un tercer parche, la que también se designa 416.

La primera rama 386 puede comprender un interpolador opcional 420, que puede estar configurado para recibir los valores 412, 416 de las representaciones en el dominio espectral del segundo parche y del tercer parche (y, opcionalmente, también los valores 408 de la representación en el dominio espectral del primer parche) y para proporcionar valores interpolados 422 de la representación en el dominio espectral del segundo y tercer parches (y, opcionalmente, también del primer parche).

La primera rama 386 puede comprender además un atenuador a cero 424, que está configurado para recibir los valores interpolados 422 (o, de lo contrario, los valores originales 412, 416) de las representaciones en el dominio espectral del segundo y tercer parches (y, opcionalmente, también del primer parche) y para obtener, sobre la base de estas, una versión atenuada a cero de los valores de una representación del dominio espectral, que está atenuada a cero a fin de adaptarla a una dimensión del convertidor del dominio espectral al dominio del tiempo 428.



El convertidor del dominio espectral al dominio del tiempo 428 puede ser implementado, por ejemplo, en forma de Transformador Rápido de Fourier inverso. Por ejemplo, el Transformador Rápido de Fourier inverso 428 puede estar configurado para recibir una serie de 2048 valores espectrales (opcionalmente interpolados y reducidos a cero) y para producir, sobre la base de estos, una representación en el dominio del tiempo 430 de la porción de señal de ancho de banda ampliado. El primer trayecto 386 comprende además un generador de ventanas de síntesis 432, que está configurado para recibir la representación en el dominio del tiempo 430 de la porción de señal de ancho de banda ampliado y para aplicar una generación de ventanas de síntesis, a fin de obtener una representación en el dominio del tiempo encuadrada en ventana de síntesis de la porción de señal de ancho de banda ampliado 430.

El decodificador de audio 300 comprende además un segundo trayecto de procesamiento 388, que ejecuta un procesamiento muy similar en comparación con el primer trayecto 386. Sin embargo, el segundo trayecto 388 comprende un atenuador a cero en el dominio del tiempo 438, que está configurado para recibir la porción transitoria encuadrada en ventana 383 de los datos de audio modulados por codificación de pulsos con número de pulsos reducido 381 y para derivar una versión atenuada a cero 439 de la porción encuadrada en ventana 383, de tal manera que el comienzo de la porción atenuada a cero 439 y el final de la porción atenuada a cero 439 estén atenuadas con ceros, y de tal manera que el transitorio esté dispuesto en una región central (entre las muestras iniciales atenuadas a cero y las muestras finales atenuadas a cero) de la porción atenuada a cero 439.

El segundo trayecto 388 comprende además un transformador del dominio del tiempo al dominio espectral 440, por ejemplo, un Transformador Rápido de Fourier o un QMF (banco de filtros espejo en cuadratura). El



transformador del dominio del tiempo al dominio espectral 440 comprende por lo general un número mayor de colectores de frecuencia (por ejemplo, colectores de frecuencia por Transformada Rápida de Fourier o bandas QMF) que el transformador del dominio del tiempo al dominio espectral 400 de la primera  
5 rama. Por ejemplo, el Transformador Rápido de Fourier 440 puede estar configurado para derivar 1024 coeficientes de Transformada Rápida de Fourier a partir de una porción atenuada a cero 439 de 1024 muestras en el dominio del tiempo.

La segunda rama 388 comprende además un determinador de valores de  
10 magnitud 442 y un determinador de valores de fase 444, que pueden comprender la misma funcionalidad que los medios correspondientes 402, 404 de la primera rama 386, aunque con dimensión incrementada  $N=1024$ . De modo similar, la segunda rama 388 comprende además un vocodificador de fases 446, una primera copiadora de valores 450, una segunda copiadora de valores 454,  
15 un interpolador opcional 460 y un atenuador a cero opcional 464, que puede comprender las mismas funcionalidades que los medios correspondientes de la primera rama 386, aunque con dimensiones incrementadas. En particular, el índice  $\xi$  de la banda de cruce puede ser más elevado en la segunda rama 388 que en la primera rama 386, por ejemplo, en un factor de 2.

20 En consecuencia, se puede enviar una representación del dominio espectral que comprende, por ejemplo, 4096 coeficientes de Transformada Rápida de Fourier a un Transformador Rápido de Fourier inverso 468, que a su vez produce una señal en el dominio del tiempo 470 que consta de 4096 muestras.

25 La segunda rama 388 comprende además un generador de ventanas de síntesis 472, que está configurado para otorgar una versión encuadrada en



ventanas de la representación en el dominio del tiempo 470 de la porción de señal de ancho de banda ampliado.

La segunda rama 388 comprende además un eliminador de ceros configurado para producir una representación abreviada y encuadrada en  
5 ventana en el dominio del tiempo 478 de la porción de señal de ancho de banda ampliado, representación abreviada y encuadrada en ventana en el dominio del tiempo 478 que puede comprender, por ejemplo, 2048 muestras.

En consecuencia, la representación en el dominio del tiempo 387 se utiliza para porciones no transitorias (por ej. cuadros de audio) de los datos de  
10 audio modulados por codificación de pulsos 322, y la representación en el dominio del tiempo 478 se utiliza para porciones transitorias de los datos de audio modulados por codificación de pulsos 322. En consecuencia, las porciones transitorias son procesadas con una resolución más elevada en el dominio espectral en la segunda rama de procesamiento 388, en tanto que las  
15 porciones no transitorias son procesadas con una resolución espectral más baja en la primera rama de procesamiento 386.

### 2.3. Formateo de Envolvente 344

A continuación se resume brevemente el formateo de envolvente 344. Además, se hace referencia a los respectivos comentarios presentados en la  
20 sección introductoria, que también se aplican al concepto de la invención.

Los datos QMF con parche 342, que se obtienen sobre la base de la señal en el dominio QMF de 64 bandas 396, son procesados por el formateo de envolvente 344, para obtener la representación de señal 348, que es transferido como entrada al sintetizador QMF 350. El formateo de envolvente puede  
25 adaptar, por ejemplo, las señales de banda en el dominio QMF de los datos QMF con parche 342 para ejecutar el llenado de ruidos, a fin de reconstruir los armónicos faltantes y/o para obtener un filtrado inverso. Las variaciones de



relleno de ruido, inserción de armónicos faltantes y filtrado inverso pueden ser controladas, por ejemplo, por una información secundaria 346, que puede ser extraída del flujo de datos 310. Para más detalles, se hace referencia, por ejemplo, a la explicación sobre la herramienta SBR presentada en la sección 4.6.18 de la Norma Internacional ISC/IEC 14496-3:2005(e), parte 3, subparte 4. Sin embargo, también se pueden aplicar diferentes conceptos de formateo de envolvente de acuerdo con los requisitos.

### 3. Explicación y comparación de las diferentes soluciones

A continuación se presenta una breve explicación y resumen de la solución de la invención.

Las realizaciones de acuerdo con la presente invención, por ejemplo el aparato 100 de acuerdo con la Fig. 1 y el decodificador de audio 300 de acuerdo con la Fig. 3, son (o comprenden) nuevos algoritmos de parches dentro de la replicación de la banda espectral (SBR). Se puede utilizar la introducción de parches en el dominio espectral de diferentes maneras para dar cuenta de las diferentes características o restricciones de señal dictadas por los requerimientos de software o hardware.

En la SBR normal, la introducción de parches se lleva a cabo siempre mediante una operación de copia dentro del dominio QMF. Esto puede llevar, en ocasiones, a artefactos auditivos, especialmente si se copian sinusoides en las proximidades unas de otras en el límite de la parte de LF y HF generada. Por lo tanto, se ha introducido un nuevo algoritmo de parches que evita algunos problemas mediante el uso de un vocodificador de fases (ver, por ejemplo, la Referencia [13]). Este algoritmo está ilustrado en la Fig. 5 como ejemplo comparativo.

La SBR standard tiene el problema de los artefactos auditivos. El enfoque de vocodificador de fases presentado en la Referencia [13] tiene una



complejidad, especialmente debido al alto número de Transformadas Rápidas de Fourier que se deben calcular. Además, el espectro se torna muy ralo en el caso de los parches altos (factores de extensión elevados), que puede dar lugar a artefactos de audio desagradables.

5        Dos realizaciones evitan el alto número de Transformadas Rápidas de Fourier mediante el traslado de la generación de diferentes parches del dominio del tiempo al dominio de la frecuencia. En la Fig. 6, se presenta un ejemplo en el cual se obtiene la transformación al dominio de la frecuencia con la ayuda de una Transformada Rápida de Fourier. En lugar de la Transformación de Fourier,  
10       se pueden utilizar, sin embargo, otras transformaciones de tiempo-frecuencia.

La Fig. 3 ilustra una solución híbrida del algoritmo de la Fig. 6 para la introducción de parches SBR. Sólo el primer parche es generado por el algoritmo del vocodificador de fases (por ejemplo, el bloque 406 de la primera rama 386 y el bloque 446 de la segunda rama 388) en tanto que los parches  
15       superiores (por ejemplo, el segundo parche y el tercer parche) se generan sólo copiando el primer parche (por ejemplo, utilizando las copiadoras de valores 410, 414 de la primera rama 386, y/o las copiadoras de valores 450, 454 de la segunda rama 388). Esto da un espectro menos ralo.

A continuación se explica brevemente el algoritmo de comparación, que  
20       se implementa en el decodificador de audio ilustrado en la Fig. 6 y el algoritmo de la invención, que está implementado en el decodificador de audio ilustrado en la Fig. 3:

El algoritmo comparativo o algoritmo de referencia, que está implementado en el decodificador de audio ilustrado en la Fig. 6, comprende los  
25       siguientes pasos:

1. Reducción del número de muestras de la señal (si no va en desmedro del criterio de Nyquist)



2. Se encuadra la señal en ventanas (Se proponen las ventanas de "Hann" aunque se pueden utilizar otras formas de ventana) y se toman los denominados granos (por ejemplo, las porciones de señal encuadradas en ventana 383) de longitudes  $N$  de la señal. Las ventanas se desplazan por la señal con un tamaño de salto  $H$ . Se propone una superposición  $N/H=8$  veces.
3. Si el grano (por ejemplo, una porción de señal encuadrada en ventana 383) contiene un evento transitorio en los bordes, está rellena (por ejemplo por el atenuador a cero 438) con ceros, lo que lleva a un sobremuestreo en el dominio de la frecuencia.
4. Los granos se transforman al dominio de la frecuencia (por ejemplo, usando los transformadores del dominio del tiempo al dominio espectral 400, 440).
5. Los granos en el dominio de la frecuencia son rellenos (opcionalmente) a una longitud de salida conveniente del algoritmo de introducción de parches.
6. Se calcula la magnitud y la fase (por ejemplo, utilizando los medios 402, 404, 442, 444).
7. Se copia el contenido del colector de frecuencia  $n$  en la posición  $sn$  para el factor de extensión  $s$ . Se multiplica la fase por el factor de extensión  $s$ . Esto se hace con todos los factores de extensión  $s$  (sólo en el caso de las regiones en el espectro que cubre los parches convenientes). (a)  $\zeta \cdot (s-1)/s \leq n \leq \zeta$  o (b)  $\zeta/s \leq n \leq \zeta$ ; (b) da un espectro más denso que (a) ya que los parches se superponen. El  $\zeta$  denota la frecuencia más alta de la parte LF, la denominada frecuencia de cruce. En términos generales, se corrige la fase para una nueva posición de muestreo (por ej., una posición de frecuencia) que se puede obtener



utilizando el algoritmo descrito en este documento o cualquier algoritmo alternativo.

5 8. Se pueden llenar colectores de frecuencia que no adquieren datos por el copiado mediante la aplicación de una función de interpolación (por ejemplo, utilizando los interpoladores 420, 460).

9. Los granos se transforman de nuevo al dominio del tiempo (por ejemplo, utilizando los Transformadores Rápidos de Fourier inversos 428, 468).

10 10. Los granos en el dominio del tiempo se multiplican por una ventana de síntesis (una vez más, se proponen las ventanas de Hann) (por ejemplo utilizando los generadores de ventanas de síntesis 432, 472).

15 11. En caso de haberse llevado a cabo la atenuación a cero en el paso 3, se vuelven a eliminar los ceros (por ejemplo, utilizando el eliminador de ceros 476).

12. Se genera una señal o cuadro de ancho de banda ampliado (por ejemplo, la señal 392), respectivamente, utilizando superposición y suma (OLA) (por ejemplo, utilizando la superposición y suma 390).

20 Sin embargo, también se puede intercambiar el orden de los pasos individuales en algunas realizaciones alternativas, y se pueden fusionar algunos de los pasos en un solo paso en algunas realizaciones alternativas.

El algoritmo de la invención, que está implementado en el decodificador de audio ilustrado en la Fig. 3, comprende los siguientes pasos:

- 25 1. La reducción del número de muestras de la señal (si no va en desmedro del criterio de Nyquist)
2. Se encuadra en ventana la señal (Se proponen las ventanas de "Hann" aunque se pueden utilizar otras formas de ventana) y se toman los denominados



granos (por ejemplo, las porciones de señal encuadradas en ventana 383) de longitudes  $N$  de la señal. Las ventanas se desplazan por la señal con un tamaño de salto  $H$ . Se propone una superposición  $N/H=8$  veces.

3. Si el grano (por ejemplo, una porción de señal encuadrada en ventana 383) contiene un evento transitorio en los bordes, está rellena (por ejemplo por el atenuador a cero 438) con ceros, lo que lleva a un sobremuestreo en el dominio de la frecuencia.
4. Los granos se transforman al dominio de la frecuencia (por ejemplo, usando los transformadores del dominio del tiempo al dominio espectral 400, 440).
5. Los granos en el dominio de la frecuencia son rellenos (opcionalmente) a una longitud de salida conveniente del algoritmo de introducción de parches.
7. Se calcula la magnitud y la fase (por ejemplo, utilizando los medios 402, 404, 442, 444).
- 15 a) Se copia el contenido del colector de frecuencia a la posición  $2n$ . Se multiplica la fase por el 2.  
(a)  $\zeta \cdot (s-1)/s \leq n \leq \zeta$  o (b)  $\zeta/s \leq n \leq \zeta$  (ver lo anterior).
7. b) Se copia el contenido del colector de frecuencia  $2n$  a la posición  $sn$  para todos los factores de extensión  $s > 2$  en los rangos  $1 \leq n \leq \zeta$ .
- 20 8. Se pueden llenar colectores de frecuencia que no adquieren datos por el copiado mediante la aplicación de una función de interpolación (por ejemplo, utilizando los interpoladores 420, 460).
9. Los granos se vuelven a transformar al dominio del tiempo (por ejemplo, utilizando los Transformadores Rápidos de Fourier inversos 428, 468).
- 25 10. Los granos en el dominio del tiempo se multiplican por una ventana de síntesis (una vez más, se proponen las ventanas de Hann) (por ejemplo utilizando los generadores de ventanas de síntesis 432, 472).



11. En caso de haberse llevado a cabo la atenuación a cero en el paso 3, se vuelven a eliminar los ceros (por ejemplo, utilizando el eliminador de ceros 476).

12. Se genera una señal o cuadro de ancho de banda ampliado (por ejemplo, la señal 392), respectivamente, utilizando superposición y suma (OLA) (por ejemplo, utilizando la superposición y suma 390).

Sin embargo, también se puede intercambiar el orden de los pasos individuales en algunas realizaciones alternativas, y se pueden fusionar algunos de los pasos en un solo paso en algunas realizaciones alternativas.

Por consiguiente, todos los pasos son idénticos en el algoritmo de referencia (que está implementado en el decodificador de audio ilustrado en la Fig. 6) y el algoritmo de la invención (que está implementado en el decodificador de audio ilustrado en la Fig. 3), excepto por el paso 7, que ha sido reemplazado por los siguientes pasos:

7.a) Se copia el contenido del colector de frecuencia a la posición  $2n$ . Se multiplica la fase por el 2.

(a)  $\zeta \cdot (s-1)/s \leq n \leq \zeta$  o (b)  $\zeta/s \leq n \leq \zeta$  (ver lo anterior).

7.b) Se copia el contenido del colector de frecuencia  $2n$  a la posición  $sn$  en el caso de todos los factores de extensión  $s > 2$  en los rangos  $1 \leq n \leq \zeta$ .

En resumen, en primer lugar las realizaciones de acuerdo con las Figs. 1, 2, 3 y 4 (como así también el decodificador de audio ilustrado en la Fig. 6) reducen drásticamente la complejidad en comparación con las mencionadas soluciones convencionales. En segundo lugar, dan lugar a modificaciones diferentes del espectro en el plano de la SBR o la presentada en la Fig. 5 (ver, por ejemplo, la Referencia [13]).

Por ejemplo, las señales de voz se podrían beneficiar con el algoritmo, que es ejecutado por el aparato, el decodificador de audio y el método de acuerdo con las Figs. 1, 2, 3 y 4, ya que se mantiene mejor la estructura de tren



de pulsos, que es típica de las señales de voz, que con la estrategia presentada en la Referencia [13].

Las aplicaciones más prominentes de las realizaciones de acuerdo con la invención son los decodificadores de audio, que con frecuencia se implementan en dispositivos manuales y, por consiguiente, funcionan con abastecimiento de energía por baterías.

#### 4. Método de acuerdo con la Fig. 4.

A continuación se describe un método 400 para generar una representación de la señal de ampliación del ancho de banda sobre la base de la representación de una señal de entrada haciendo referencia a la Fig. 4, que ilustra un gráfico de flujo de ese método. El método 400 comprende un paso 410 que consiste en obtener los valores de una representación en el dominio espectral de un primer parche de la señal de ancho de banda ampliado sobre la base de la representación de la señal de entrada utilizando una vocodificación de fases. El método 400 comprende además un paso 420 de copiado de una serie de valores de la representación en el dominio espectral del primer parche, valores que se obtienen empleando la vocodificación de fases, para obtener una serie de valores de una representación en el dominio espectral de un segundo parche, donde el segundo parche está asociado a frecuencias más elevadas que el primer parche. El método 400 comprende además un paso 430 de obtención de una representación de la señal de ancho de banda ampliado utilizando los valores de la representación en el dominio espectral del primer parche y los valores de la representación en el dominio espectral del segundo parche.

El método 400 puede ser complementado por cualquiera de los medios y funcionalidades aquí descriptos con respecto al aparato de la invención.

#### 5. Alternativas de implementación



Si bien se han descripto algunos aspectos en el contexto de un aparato, queda claro que estos aspectos también representan una descripción de un método correspondiente, en el que un bloque o dispositivo corresponde a un paso del método o a una característica de un paso del método. De manera  
5 análoga, los aspectos descriptos en el contexto de un paso del método representan una descripción de un correspondiente bloque o elemento o característica de un correspondiente aparato. Algunos o la totalidad de los pasos del método pueden ser ejecutados por medio de (o utilizando) un aparato de hardware, como por ejemplo un microprocesador, una computadora  
10 programable o un circuito electrónico. En algunas realizaciones, uno o más cualesquiera de los pasos más importantes del método pueden ser ejecutados por ese aparato.

Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La  
15 implementación se puede realizar utilizando un medio de almacenamiento digital, como por ejemplo un disco blando, un DVD, un Blue-Ray, un CD, una memoria ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, con señales de control aptas para leer electrónicamente almacenadas en la misma, que cooperan (o tiene la capacidad de cooperar) con un sistema de  
20 computación programable de tal manera que se ejecute el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por computadora.

Algunas realizaciones de acuerdo con la invención comprenden un portador de datos con señales de control aptas para leer electrónicamente, que  
25 tienen la capacidad de cooperar con un sistema de computación programable de tal manera que se ejecute uno de los métodos aquí descriptos.



En general, las realizaciones de la presente invención se pueden implementar en forma de producto programa de computación con un código de programa, donde el código de programa cumple la función de desarrollar uno de los métodos al ejecutarse el programa de computación en una computadora. El  
5 código del programa puede estar almacenado, por ejemplo, en un portador legible por una máquina.

Otras realizaciones comprenden el programa de computación para ejecutar uno de los métodos descritos en la presente, almacenado en un portador legible por una máquina.

10 En otras palabras, una realización del método de la invención consiste, por lo tanto, en un programa de computación que consta de un código de programa para desarrollar uno de los métodos aquí descritos al ejecutarse un programa en una computadora.

Por lo tanto, otra realización de los métodos de la invención consiste en  
15 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 este documento.

Otra realización del método de la invención consiste, por consiguiente, en un flujo de datos o una secuencia de señales que representan el programa de  
20 computación para ejecutar uno de los métodos aquí descritos. El flujo de datos o la secuencia de señales puede estar configurada, por ejemplo, para ser transferida a través de una conexión de comunicaciones, por ejemplo por internet.

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



Otra realización comprende una computadora que tiene instalado en la misma el programa de computación para ejecutar uno de los métodos aquí descritos.

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

Las realizaciones antes descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las personas con capacitación en la técnica considerarán evidentes las modificaciones y variaciones de las disposiciones y detalles descritos en este documento. Por lo tanto, se pretende que el alcance sólo esté limitado por el alcance de las reivindicaciones de patente siguientes y no por los detalles específicos presentados a manera de descripción y explicación de las realizaciones aquí expuestas.

#### 6. Ejemplo comparativo de acuerdo con la Fig. 5

A continuación, se describe en forma breve un ejemplo comparativo haciendo referencia a la Fig. 5. La funcionalidad del ejemplo comparativo de acuerdo con la Fig. 5 es similar a la función del decodificador de audio de acuerdo con la Fig. 3, por lo que no se vuelven a explicar los medios y funcionalidades. Sin embargo, el ejemplo comparativo de acuerdo con la Fig. 5 se basa en el uso de tres vocodificadores de fases 590, 592, 594, o 596, 597, 598 por rama. Los Transformadores Rápidos de Fourier inversos, los generadores de ventanas de síntesis, los dispositivos de superposición y suma



individuales, están asociados a los vocodificadores de fases individuales, como se puede apreciar en la Fig. 5. Además, en algunas de las subramas, se utiliza la reducción individual del número de muestras (factor  $\downarrow$ ) y el retardo individual ( $z^{-\text{muestras}}$ ). En consecuencia, el aparato 500 de acuerdo con la Fig. 5 no es tan eficiente desde el punto de vista informático como el aparato 300 de acuerdo con la Fig. 3. De todas maneras, el aparato 500 trae aparejadas significativas mejoras con respecto a algunos decodificadores de audio convencionales.

#### 7. Ejemplo comparativo de acuerdo con la Fig. 6

La Fig. 6 ilustra otro decodificador de audio 600, de acuerdo con un ejemplo comparativo. El decodificador de audio 600 de acuerdo con la Fig. 6 es similar a los decodificadores de audio 300, 500 de acuerdo con las Figs. 3 y 5. Sin embargo, el decodificador de audio 600 también se basa en el uso de una pluralidad de vocodificadores de fases individuales 690, 692, 694 o 696, 697, 698 por cada rama, lo que hace que el aparato 600 demande más elementos de computación que el aparato 300, además de traer consigo artefactos auditivos en algunos casos. De todas maneras, el aparato 500 trae consigo significativas mejoras con respecto a algunos decodificadores de audio convencionales..

#### 8. Conclusión

En vista de la explicación que antecede, se puede apreciar que el aparato 100 de acuerdo con la Fig. 1, el decodificador de audio 300 de acuerdo con la Fig. 3 y el método 400 de acuerdo con la Fig. 4 presentan un número de ventajas con respecto a los ejemplos comparativos, que han sido descriptos en forma breve con referencia a las Figs. 5 y 6.

El concepto de la invención se puede aplicar a una amplia variedad de aplicaciones y puede ser modificado en un gran número de formas. En particular, Los Transformadores Rápidos de Fourier pueden ser reemplazados



por bancos de filtros QMF y los Transformadores Rápidos de Fourier inversos pueden ser reemplazados por sintetizadores QMF.

- Además, en algunas realizaciones, se pueden resumir algunos o todos los pasos de procesamiento en un solo paso. Por ejemplo, se puede simplificar
- 5 una secuencia de procesamiento que comprende una síntesis QMF y un posterior Análisis QMF omitiendo las transformadas repetidas.

**Referencias:**

- [1] M. Dietz, L. Liljeryd, K. Kjörling y O. Kunz, "Spectral Band Replication, a novel approach in audio coding," en la 112a. Convención de AES, Munich, mayo de 2002.
- 5 [2] S. Meltzer, R. Böhm y F. Henn, "SBR enhanced audio codecs for digital broadcasting such as "Digital Radio Mondiale" (DRM)," en la 112a. Convención de AES, Munich, mayo de 2002.
- [3] T. Ziegler, A. Ehret, P. Ekstrand y M. Lutzky, "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm," en la 112a.
- 10 Convención de AES, Munich, mayo de 2002.
- [4] International Standard ISO/IEC 14496-3:2001/FPDAM 1, "Bandwidth Extension," ISO/IEC, 2002. Speech bandwidth extension method and apparatus Vasu lyengar et al.
- [5] E. Larsen, R. M. Aarts, y M. Danessis. Efficient high-frequency
- 15 bandwidth extension of music and speech. En la 112a. Convención de AES, Munich, Alemania, mayo de 2002.
- [6] R. M. Aarts, E. Larsen, y O. Ouweltjes. A unified approach to low- and high frequency bandwidth extension. En la 115a. Convención de AES, New York, USA, octubre de 2003.
- 20 [7] K. Käyhkö. A Robust Wideband Enhancement for Narrowband Speech Signal. Research Report, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 2001.
- [8] E. Larsen and R. M. Aarts. Audio Bandwidth Extension – Application to psychoacoustics, Signal Processing and Loudspeaker Design.
- 25 John Wiley & Sons, Ltd, 2004.



- [9] E. Larsen, R. M. Aarts, and M. Danessis. Efficient high-frequency bandwidth extension of music and speech. En la 112a. Convención de AES, Munich, Alemania, mayo de 2002.
- [10] J. Makhoul. Spectral Analysis of Speech by Linear Prediction. IEEE Transactions on Audio and Electroacoustics, AU-21(3), junio de 1973.
- [11] Solicitud de patente de los Estados Unidos 08/951,029, Ohmori , et al. Audio band width extending system and method.
- [12] Patente de los Estados Unidos 6895375, Malah, D & Cox, R. V.: System for bandwidth extension of Narrow-band speech.
- [13] Frederik Nagel, Sascha Disch, "A harmonic bandwidth extension method for audio codecs," ICASSP International Conference on Acoustics, Speech and Signal Processing, IEEE CNF, Taipei, Taiwan, abril de 2009.



## REIVINDICACIONES

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

- 5           1.       Un aparato (100;386) para generar una representación (120; 426) de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada (110; 383), donde el aparato comprende:

                  un vocodificador de fases (130; 406) configurado para obtener valores ( $\beta_z \dots \beta_{2z}$ , 408) de una representación en el dominio espectral de un  
10   primer parche de la señal de ancho de banda ampliado sobre la base de la representación de la señal de entrada y

                  una copiadora de valores (140; 410,416) configurada para copiar una serie de valores ( $\beta_z \dots \beta_{2z}$ , 408) de la representación en el dominio espectral del primer parche, valores que son provistos por el vocodificador de fases, para  
15   obtener una serie de valores ( $\beta_{2z} \dots \beta_{3z}$ , 408) de una representación en el dominio espectral de un segundo parche, donde el segundo parche está asociado a frecuencias más elevadas que el primer parche;

                  donde el aparato está configurado para obtener la representación (120; 426) de la señal de ancho de banda ampliado empleando los valores de la  
20   representación en el dominio espectral del primer parche y los valores de la representación en el dominio espectral del segundo parche.

2.       El aparato (100;386) de acuerdo con la reivindicación 1, en el cual el vocodificador de fases (130; 406) está configurado para copiar una serie de valores de magnitud ( $\alpha_{z/2} \dots \alpha_z$ ) asociados a una pluralidad de subrangos de  
25   frecuencias dados de la representación de la señal de entrada (110; 383), para obtener una serie de valores de magnitud ( $\alpha_z \dots \alpha_{2z}$ ) asociados con los correspondientes subrangos de frecuencias del primer parche,



donde un par de un subrango de frecuencias dado de la representación de la señal de entrada y de un subrango de frecuencias correspondiente del primer parche cubre un par de una frecuencia fundamental y un armónico de la frecuencia fundamental,

5        en el cual el vocodificador de fases (130; 406) está configurado para multiplicar valores de fase ( $\varphi_{\zeta/2} \dots \varphi_{\zeta}$ ) asociados a la pluralidad de subrangos de frecuencias dados de la representación de la señal de entrada por un factor predeterminado, para obtener una serie de valores de fase ( $\varphi_{\zeta} \dots \varphi_{2\zeta}$ ) asociados a los correspondientes subrangos de frecuencias del primer parche, y

10        donde la copiadora de valores (140; 410) está configurada para copiar una serie de valores ( $\beta_{\zeta} \dots \beta_{2\zeta}$ ) asociados a una pluralidad de subrangos dados de frecuencias del primer parche, para obtener una serie de valores ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ ) asociados a los correspondientes subrangos de frecuencias del segundo parche, donde la copiadora de valores está configurada para dejar inalterados  
15        los valores de fase en el copiado.

3.        El aparato (100;386) de acuerdo con la reivindicación 2, en el cual la copiadora de valores (140; 410) está configurada para copiar los valores de tal manera que se obtenga un desplazamiento espectral común entre los valores ( $\beta_{\zeta} \dots \beta_{2\zeta}$ ) del primer parche y los correspondientes valores ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ ) del  
20        segundo parche.

4.        El aparato (100;386) de acuerdo con una de las reivindicaciones 1 a 3, en el cual el vocodificador de fases (130; 410) está configurado para obtener los valores ( $\beta_{\zeta} \dots \beta_{2\zeta}$ ) de la representación en el dominio espectral (132; 408) del primer parche por lo que los valores de la representación en el dominio  
25        espectral del primer parche representan una versión armónicamente convertida en aumento de un rango de la frecuencia fundamental de la representación de la señal de entrada (110; 383) y



donde la copiadora de valores (140;410) está configurada para obtener los valores ( $\beta_{2z} \dots \beta_{3z}$ ) de la representación en el dominio espectral (142; 412) del segundo parche por lo que los valores de la representación en el dominio espectral del segundo parche representan una versión con desplazamiento en la frecuencia del contenido de audio del primer parche.

5        5. El aparato (100; 380,382, 386) de acuerdo con una de las reivindicaciones 1 a 4, donde el aparato está configurado para recibir datos de audio de entrada (322),

para reducir el número de muestras (380) de los datos de audio de entrada (322), a fin de obtener datos de audio con número reducido de muestras (381),

para encuadrar en ventanas (382) los datos de audio con número de muestras reducido (381), para obtener datos de entrada encuadrados en ventanas (383),

15        para convertir (400) o transformar los datos de entrada encuadrados en ventanas (383) al dominio espectral, para obtener la representación de la señal de entrada (383) en forma de una representación en el dominio espectral (401),

para computar (402, 404) los valores de magnitud  $\alpha_k$  y los valores de fase  $\varphi_k$  que representan un colector de frecuencia que tiene el índice  $k$  de la representación de la señal de entrada (383),

20        para usar (130; 406) una pluralidad de valores de magnitud  $\alpha_k$  que representan colectores de frecuencia con índices de colector de frecuencia  $k$  de la representación de la señal de entrada (383), para obtener valores de magnitud  $\alpha_{2k}$  que representan colectores de frecuencia con índices de colector de frecuencia  $s_k$  del primer parche,

cuando  $s$  es un factor de extensión con  $s$  de entre 1,5 y 2,5 y



para copiar y escalar (130; 406) los valores de fase  $\varphi_k$  asociados a colectores de frecuencia que tienen índices de colector de frecuencia  $k$  de la representación de la señal de entrada (383), para obtener valores de fase  $\varphi_{2k} = s\varphi_k$  copiados y escalados asociados con colectores de frecuencia que

5 tienen índices de colector de frecuencia  $2k$  del primer parche,

para copiar (140; 410) los valores  $\beta_{k-i\zeta}$  asociados a colectores de frecuencia que tienen índices de colector de frecuencia  $k-i\zeta$  de la representación en el dominio espectral (132;408) del primer parche, para obtener los valores  $\beta_k$  de la representación en el dominio espectral (142; 412) del segundo parche,

10 para convertir (428) la representación (426) de la señal de ancho de banda ampliado al dominio del tiempo, para obtener una representación en el dominio del tiempo (430), y

para aplicar (432) una ventana de síntesis a la representación en el dominio del tiempo.

15 6. El aparato (100; 386) de acuerdo con una de las reivindicaciones 1 a 5, en el cual el aparato comprende un convertidor del dominio del tiempo al dominio espectral (400) configurado para proporcionar, como representación de la señal de entrada (401), los valores de una representación del dominio espectral de una señal de audio de entrada (322), o de una versión previamente

20 procesada (383) de la señal de audio de entrada (322) y

donde el aparato comprende un convertidor del dominio espectral al dominio del tiempo (428) configurado para producir una representación en el dominio del tiempo (430) de la señal de ancho de banda ampliado utilizando los valores  $(\beta_\zeta \dots \beta_{2\zeta}, 408)$  de la representación del dominio espectral del primer

25 parche y los valores  $(\beta_{2\zeta} \dots \beta_{3\zeta}, 412)$  de la representación del dominio espectral del segundo parche;



donde el convertidor del dominio espectral al dominio del tiempo (428) está configurado de tal manera que un número ( $N=2048$ ) de valores espectrales diferentes (426) recibidos por el convertidor del dominio espectral al dominio del tiempo (428) sea mayor que un número ( $N=512$ ) de valores espectrales diferentes (401) provistos por el dominio del convertidor del dominio del tiempo al espectral (400), por lo que el convertidor del dominio espectral al dominio del tiempo (428) está configurado para procesar un número mayor de colectores de frecuencia que el convertidor del convertidor del dominio del tiempo al dominio espectral (400).

7. El aparato (100; 382, 386) de acuerdo con una de las reivindicaciones 1 a 6, donde el aparato comprende un generador de ventanas de análisis (382) configurado para encuadrar en ventana una señal de audio de entrada en el dominio del tiempo (322), para obtener una versión encuadrada en ventana (383) de la señal de audio de entrada en el dominio del tiempo, lo que forma la base para la obtención de la representación de la señal de entrada en forma de una representación en el dominio espectral (401) y

donde el aparato comprende un generador de ventanas de síntesis (432) configurado para encuadrar en ventana una porción de una representación en el dominio del tiempo (430) de la señal de ancho de banda ampliado, para obtener una porción encuadrada en ventana (434) de la representación en el dominio del tiempo de la señal de ancho de banda ampliado.

8. El aparato (100; 382,386) de acuerdo con la reivindicación 7, donde el aparato está configurado para procesar una pluralidad de porciones con desplazamiento temporal temporalmente superpuestas de la señal de audio de entrada en el dominio del tiempo (322), para obtener una pluralidad de porciones con desplazamiento temporal temporalmente superpuestas



encuadradas en ventanas (434) de la representación en el dominio del tiempo de la señal de ancho de banda ampliado,

donde un desvío en el tiempo ( $\text{Inc}=64$ ) entre porciones con desplazamiento en el tiempo temporalmente adyacentes de la señal de audio de entrada en el dominio del tiempo (322) es menor o igual a un cuarto de la longitud de una ventana (512) del generador de ventana de análisis (382).

9. El aparato (100; 382,386) de acuerdo con una de las reivindicaciones 1 a 8, donde el aparato comprende un productor de información transitoria (384) configurado para proporcionar una información que indica la presencia de un transitorio en la señal de entrada (322) y

donde el aparato comprende una primera rama de procesamiento (386) para producir una representación (434) de una porción de señal de ancho de banda ampliado sobre la base de una porción no transitoria de la representación de la señal de entrada (383) y una segunda rama de procesamiento (388) para producir una representación (478) de una porción de la señal de ancho de banda ampliado sobre la base de una porción transitoria de la representación de la señal de entrada (383);

donde la segunda rama de procesamiento (388) está configurada para procesar una representación del dominio espectral (441) de la señal de entrada que tiene una resolución espectral más elevada ( $N=1024$ ) que una representación del dominio espectral (401) de la señal de entrada procesada por la primera rama de procesamiento (386).

10. El aparato (100; 382,386) de acuerdo con la reivindicación 9, en el cual la segunda rama de procesamiento (388) comprende un atenuador a cero en el dominio del tiempo (438) configurado para atenuar a cero una porción con contenido de transitorios (383) de la señal de entrada, a fin de obtener una



porción con contenido de transitorios temporalmente ampliada (439) de la señal de entrada y

en el cual la primera rama de procesamiento (386) comprende un convertidor del dominio del tiempo al dominio de la frecuencia (400) configurado para producir un primer número ( $N=512$ ) de valores en el dominio espectral (401) asociados a la porción no transitoria (383) de la señal de entrada y

en el cual la segunda rama de procesamiento (388) comprende un convertidor del dominio del tiempo al dominio de la frecuencia (440) configurado para producir un segundo número ( $N=1024$ ) de valores en el dominio espectral (441) asociados a la porción con contenido de transitorios temporalmente ampliada (439) de la señal de entrada,

donde el segundo número ( $N=1024$ ) de valores en el dominio espectral es mayor, por lo menos en un factor de 1,5, que el primer número ( $N=512$ ) de valores en el dominio espectral.

11. El aparato (100; 382,386) de acuerdo con la reivindicación 10, en el cual la segunda rama de procesamiento comprende un eliminador de ceros (476) configurado para eliminar una pluralidad de valores cero de una porción de señal de ancho de banda ampliado (474) obtenida sobre la base de la porción con contenido de transitorios temporalmente ampliada (439) de la señal de entrada.

12. El aparato (100; 380, 382,386) de acuerdo con una de las reivindicaciones 1 a 11, donde el aparato comprende un reductor del número de muestras (380) configurado para reducir el número de muestras de una representación en el dominio del tiempo (322) de la señal de entrada.

13. Un decodificador de audio (300) que comprende un aparato (100; 386) de acuerdo con una de las reivindicaciones 1 a 12.



14. Un método (400) para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada, método que comprende:

5 obtener (410), utilizando una vocodificación de fases, los valores de una representación del dominio espectral de un primer parche de la señal de ancho de banda ampliado sobre la base de la representación de la señal de entrada y

copiar (420) una serie de valores de la representación del dominio espectral del primer parche, valores que son provistos por la vocodificación de fases, para obtener una serie de valores de una representación del dominio espectral de un segundo parche, donde el segundo parche está asociado a frecuencias más elevadas que el primer parche y

10 obtener (430) la representación de la señal de ancho de banda ampliado utilizando los valores de la representación del dominio espectral del primer parche y los valores de la representación del dominio espectral del segundo parche.

15. Un aparato (100;386) para generar una representación (120; 426) de una señal de ancho de banda ampliado sobre la base de la representación de una señal de entrada (110; 383), aparato que comprende:

20 una copiadora de valores configurada para copiar una serie de valores ( $\beta_1 \dots \beta_{\zeta}$ ) de la representación de la señal de entrada, para obtener una serie de valores ( $\beta_{\zeta} \dots \beta_{2\zeta}$ ) de una representación en el dominio espectral de un primer parche, donde el primer parche está asociado a frecuencias más elevadas que la representación de la señal de entrada y

25 un vocodificador de fases (130; 406) configurado para obtener valores ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ ) de una representación en el dominio espectral de un segundo parche de la señal de ancho de banda ampliado sobre la base de los valores



( $\beta_{4/3\zeta} \dots \beta_{2\zeta}$  de la representación en el dominio espectral del primer parche, donde el segundo parche está asociado a frecuencias más elevadas que el primer parche y

donde el aparato está configurado para obtener la representación  
5 (120;426) de la señal de ancho de banda ampliado utilizando los valores de la representación en el dominio espectral del primer parche y los valores de la representación en el dominio espectral del segundo parche.

16. Un método (400) para generar una representación de una señal de ancho de banda ampliado sobre la base de la representación de una señal de  
10 entrada, método que comprende:

copiar los valores de la representación de la señal de entrada, para obtener los valores de una representación del dominio espectral de un primer parche de la señal de ancho de banda ampliado sobre la base de la representación de la señal de entrada, donde el primer parche está asociado a  
15 frecuencias más elevadas que la representación de la señal de entrada y

obtener, empleando una vocodificación de fases, una serie de valores de la representación del dominio espectral del segundo parche sobre la base de una serie de valores de la representación del dominio espectral del primer parche, valores de la representación en el dominio espectral del primer  
20 parche que son obtenidos del copiado, donde el segundo parche está asociado a frecuencias más elevadas que el primer parche y

obtener (430) la representación de la señal de ancho de banda ampliado empleando los valores de la representación del dominio espectral del primer parche y los valores de la representación del dominio espectral del  
25 segundo parche.



17. Un programa de computación para desarrollar el método de acuerdo con la reivindicación 14 o la reivindicación 16, al ejecutarse el programa de computación en una computadora.

5

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

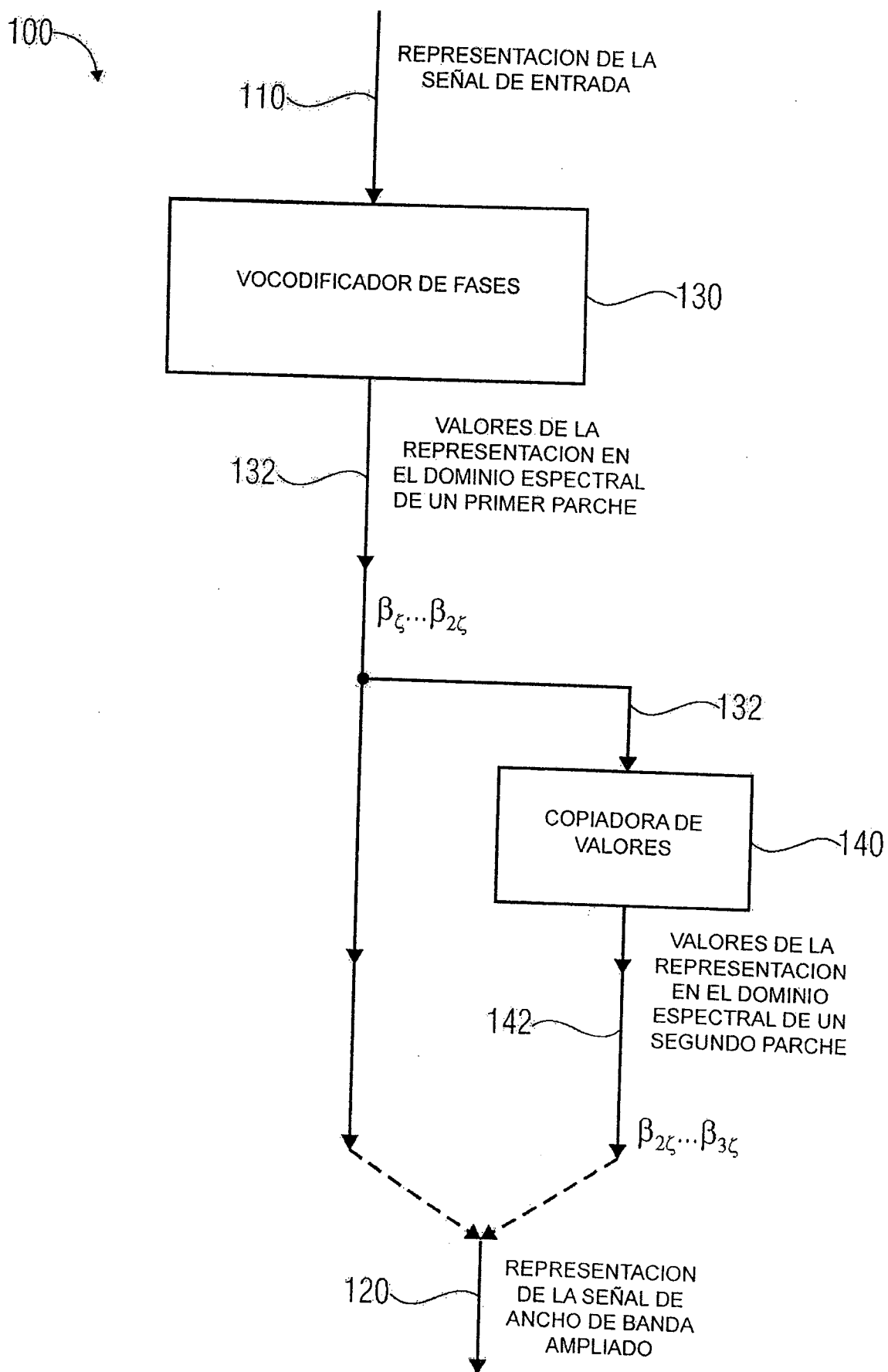


FIG 1

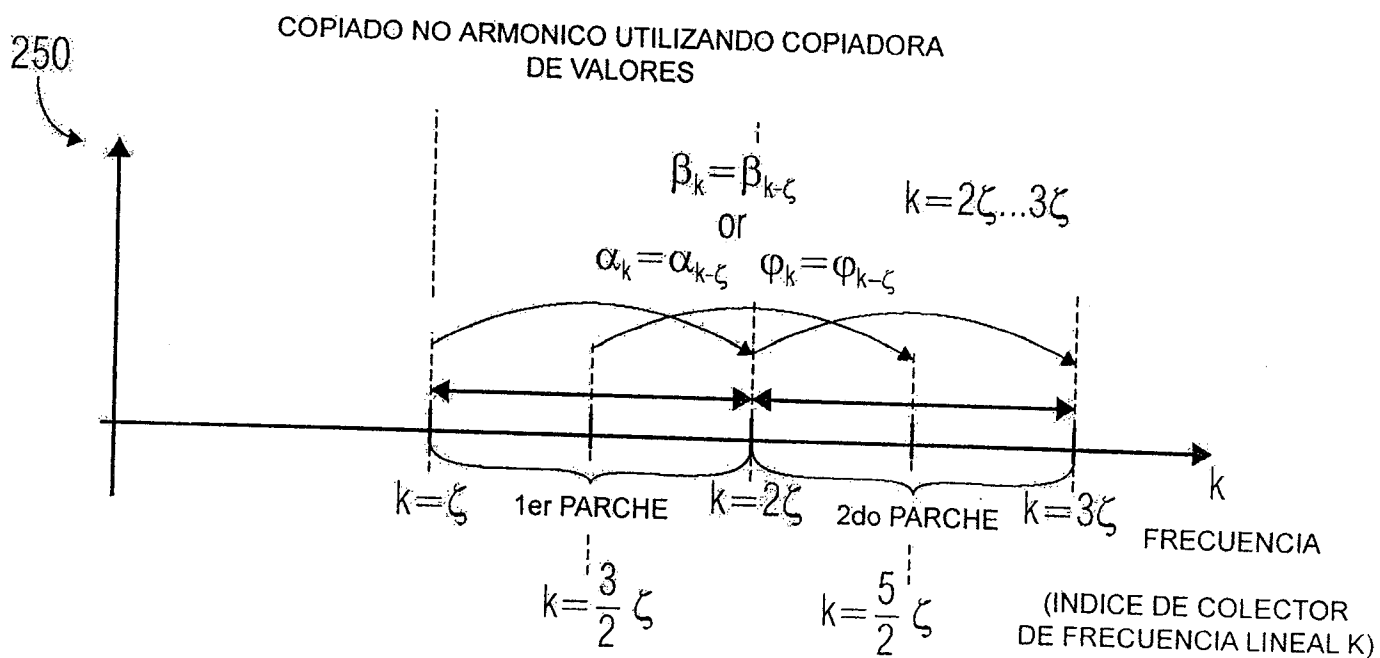
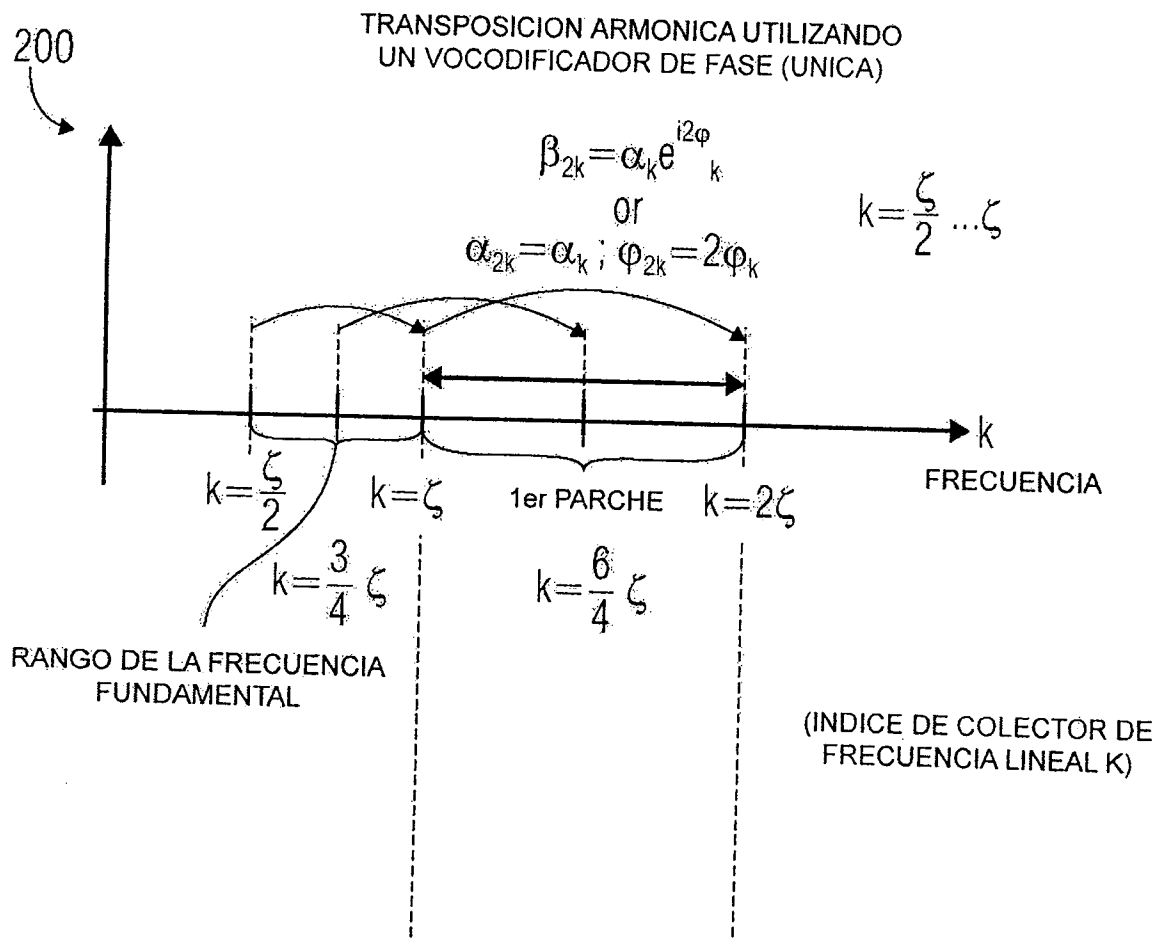


FIG 2

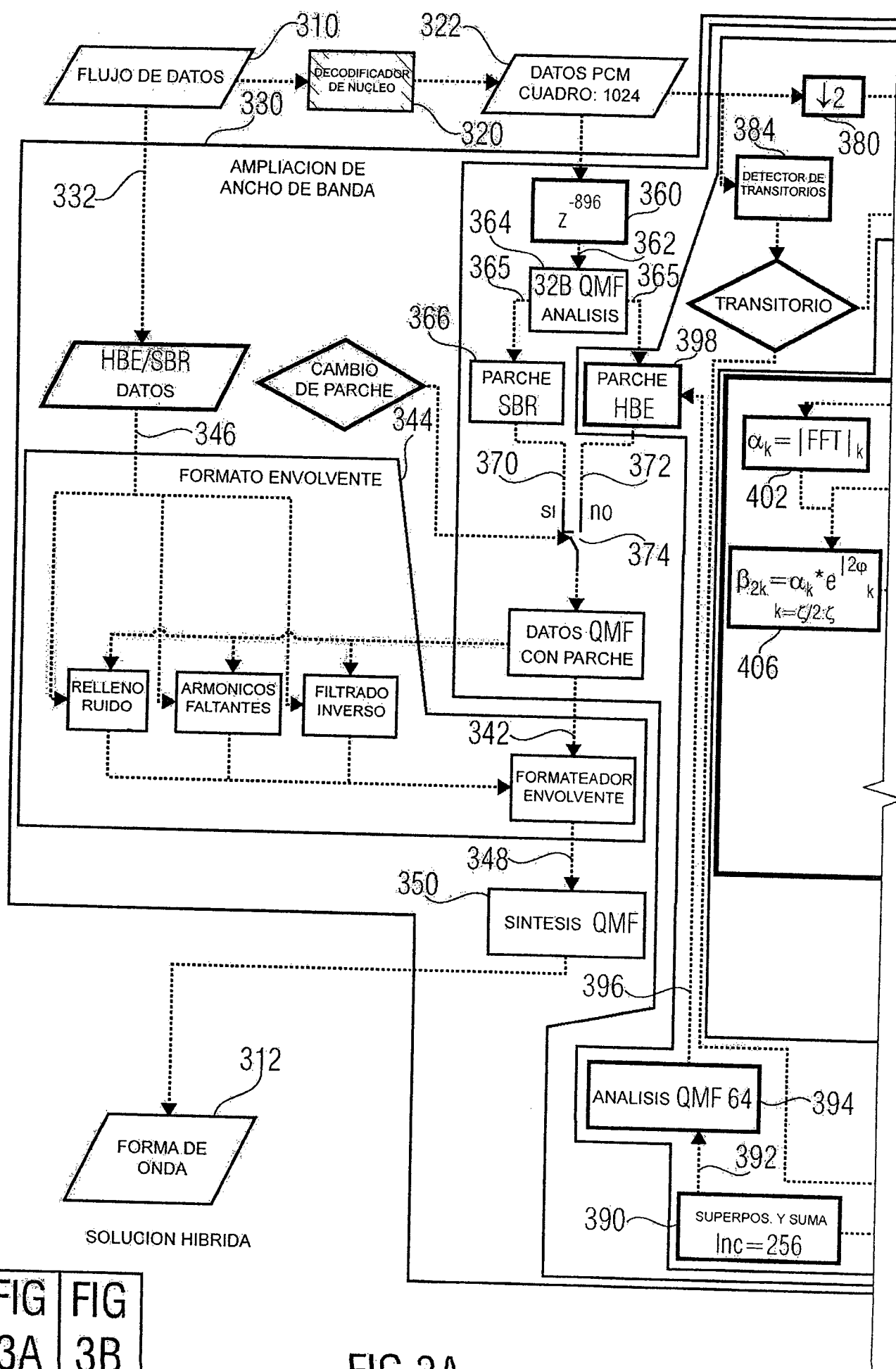
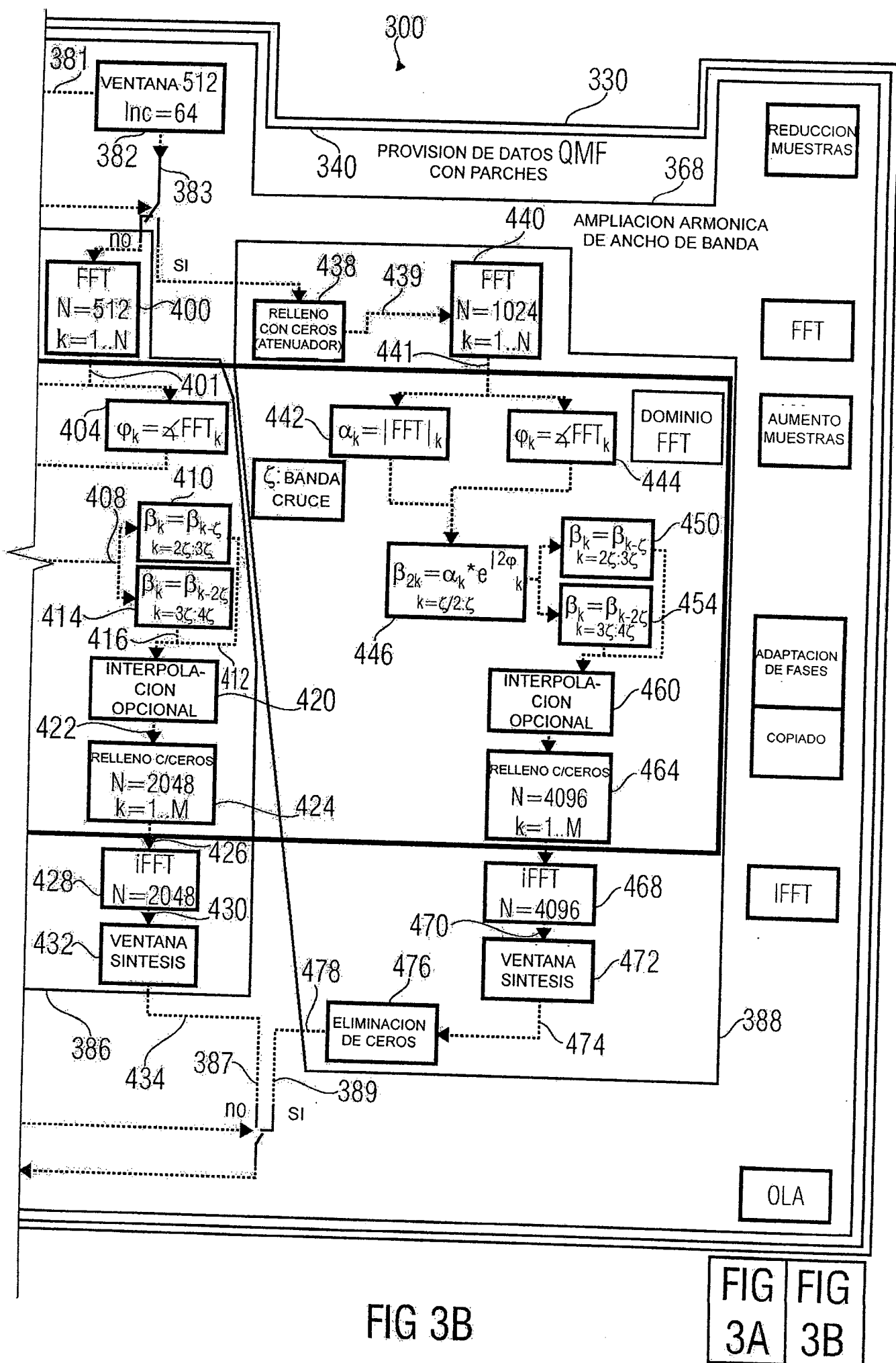


FIG 3A FIG 3B

FIG 3A



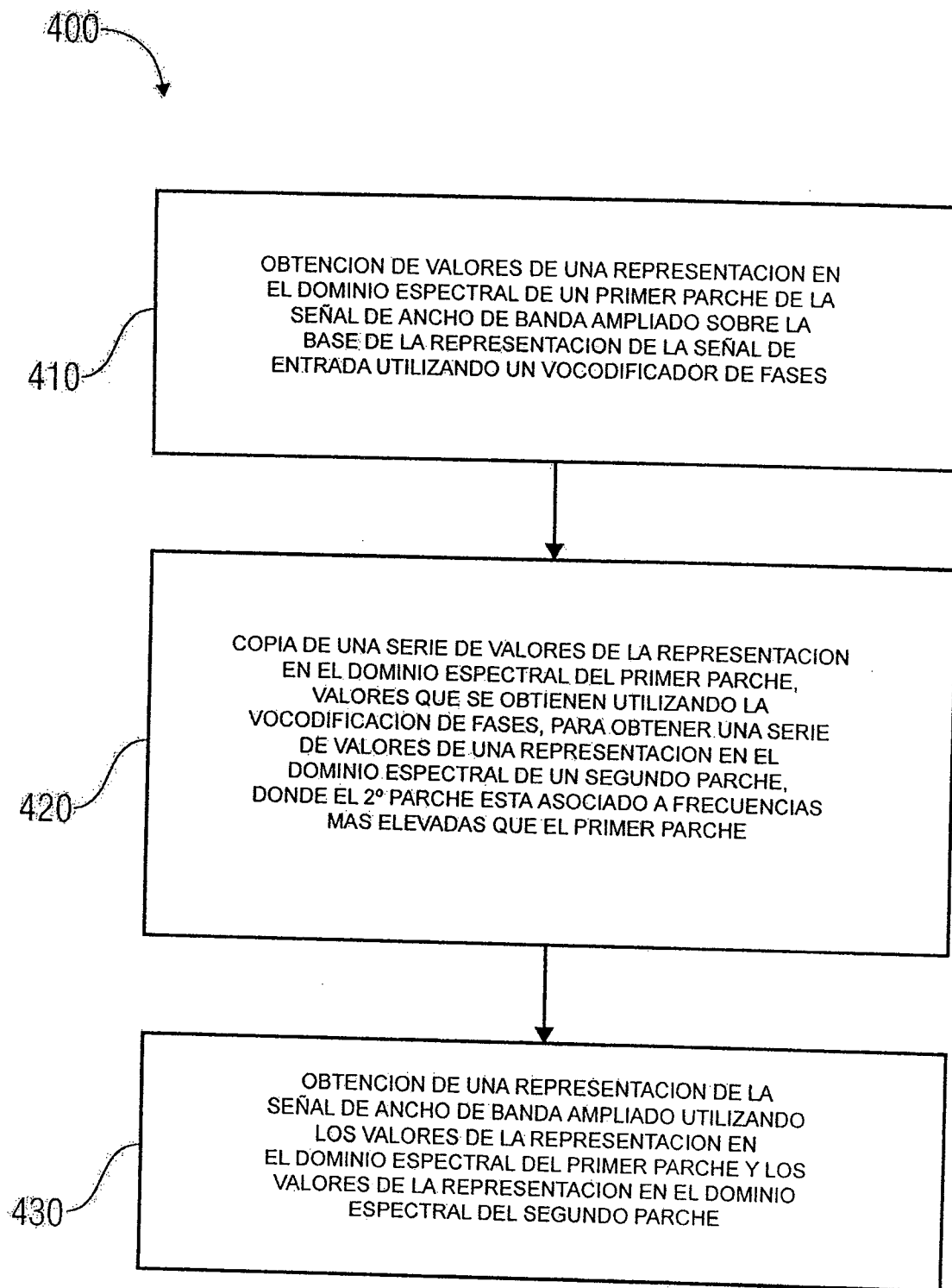


FIG 4

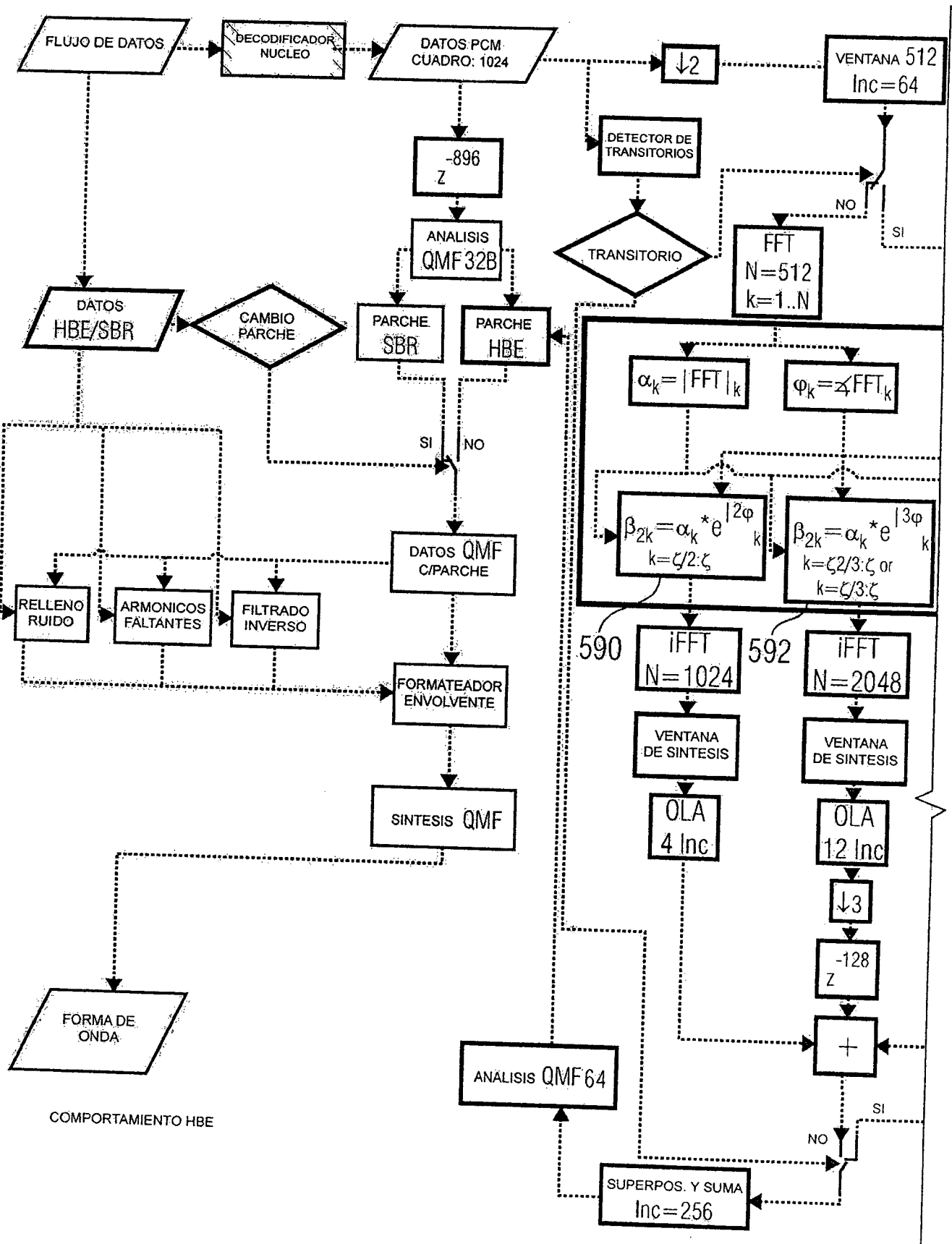


FIG 5A

FIG 5A FIG 5B

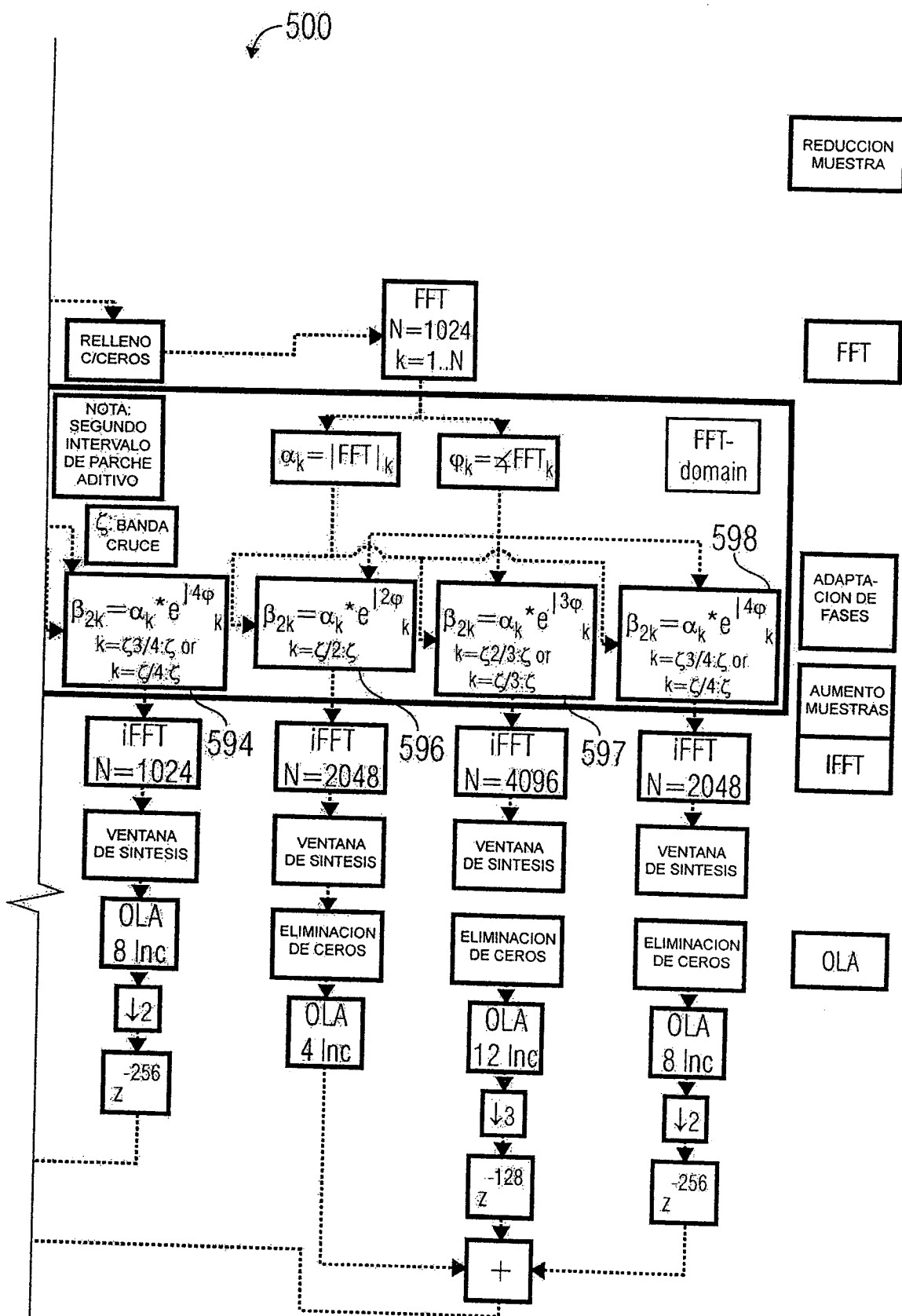


FIG 5B

FIG 5A

FIG 5B

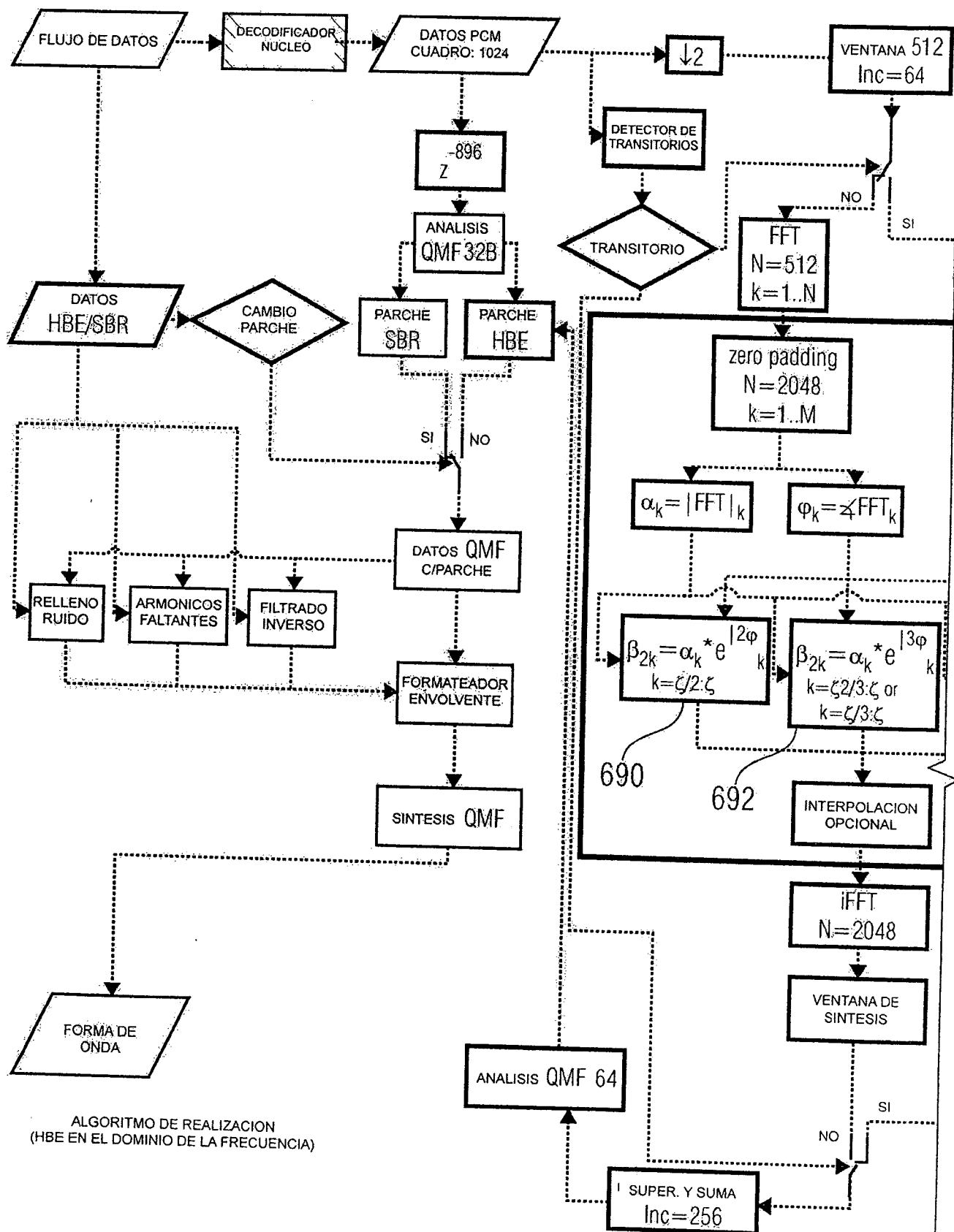


FIG 6A

FIG 6A FIG 6B

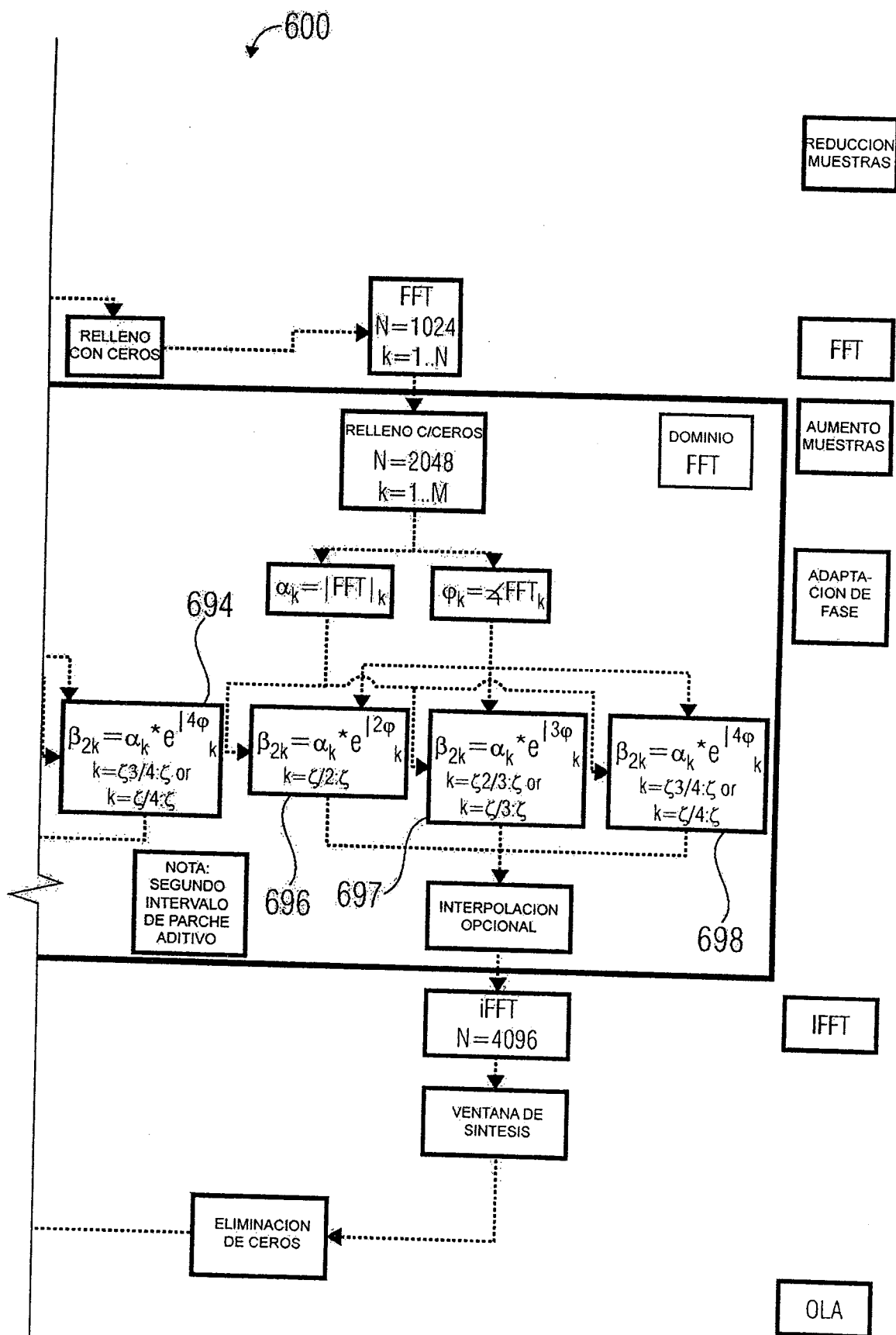


FIG 6B

FIG 6A

FIG 6B

**Apparatus, Method and Computer Program for Generating a Representation of a Bandwidth-Extended Signal on the Basis of an Input Signal Representation Using a Combination of a Harmonic Bandwidth-Extension and a Non-Harmonic Bandwidth-Extension**

5

Description

Technical Field

10 Embodiments according to the invention are related to an apparatus for generating a representation of a bandwidth-extended signal on the basis of an input signal representation. Other embodiments according to the invention are related to a method for generating a representation of a bandwidth-extended signal on the basis of an input signal representation. Further embodiments according to the invention are related to a computer  
15 program for performing such method.

Some embodiments according to the invention are related to novel patching methods inside spectral band replication.

20 Background of the Invention

Storage or transmission of audio signals is often subject to strict bitrate constraints. These constraints are usually overcome by a coding of the signal. In the past, coders were forced to drastically reduce the transmitted audio bandwidth when only a very low bitrate was  
25 available. Modern audio codecs are nowadays able to preserve the audible bandwidth by using bandwidth extension (BWE) methods. Such methods are described, for example, in references [1] to [12]. These algorithms rely on a parametric representation of the high-frequency content (HF), which is generated from the waveform-coded low-frequency part (LF) of the decoded signal by means of transposition into the HF spectral region  
30 (“patching”) and the application of a parameter driven post processing.

In the art, methods of bandwidth extension, such as spectral band replication (SBR) are used as an efficient method to generate high-frequency signals in HFR (high-frequency reconstruction) based codecs.

35

The spectral band replication described in reference [1], which is also briefly designated as “SBR”, uses a quadrature mirror filterbank (QMF) for generating the HF information. With the help of the so-called “patching” process, lower QMF-bands are copied to higher

(frequency) position yielding in a replication of the information of the LF part in the HF part. The generated HF is afterwards adapted to the original HF part with the help of parameters that adopt (or adjust) the spectral envelope and the tonality (for example using an envelope formatting).

5

In standard SBR, patching is always carried out by a copy operation inside the QMF-domain. It has been found that this can sometimes lead to auditory artifacts, particularly if sinusoids are copied into the vicinity of each other at the border of LF and the generated HF part. Thus, it can be stated that the standard SBR has the problem of auditory artifacts.

10

Also, some conventional implementations of bandwidth extension concept bring along a comparatively high complexity. Additionally, in some invention implementations of bandwidth extension concepts, the spectrum becomes very sparse for high patches (high stretching factors), which may result in undesired (audible) audio artifacts.

15

In view of the above discussion, it is an objective of the present invention to create a concept for generating a representation of a bandwidth-extended signal on the basis of an input signal representation, which brings along an improved tradeoff between complexity and audio quality.

20

#### Summary of the Invention

Embodiments according to the invention create an apparatus for generating a representation of a bandwidth-extended signal on the basis of an input signal representation. The apparatus comprises a phase vocoder configured to obtain values of a spectral-domain representation of a first patch of the bandwidth-extended signal on the basis of the input signal representation. The apparatus also comprises a value copier configured to copy a set of values of the spectral-domain representation of the first patch, which values are provided by the phase vocoder, to obtain a set of values of a spectral-domain representation of a second patch. The second patch is associated with higher frequencies than the first patch. The apparatus is configured to obtain the representation of the bandwidth-extended signal using the values of the spectral-domain representation of the first patch and the values of the spectral-domain representation of the second patch.

35

It is the key idea of the present invention that a particularly good tradeoff between computational complexity and audio quality of a bandwidth-extended signal is obtained by combining a phase vocoder with a value copier, such that the first patch of the bandwidth-extended signal is obtained by the phase vocoder, and such that the second patch of the bandwidth-extended signal is obtained on the basis of the first patch using the value copier.

Accordingly, the content of the first patch is a harmonically transposed version of the content of the low-frequency part (LF) of the input signal (represented by the input signal representation), and the second patch is (or represents) a (non-harmonically) frequency-shifted version of the signal content of the first patch. Accordingly, the second patch can  
 5 be obtained with relatively low computational complexity because the copying of the values is computationally simpler than a phase vocoding operation. Also, it is avoided that there are large spectral holes in the second patch, because the spectral values of the first patch are typically populated (i.e. comprise non-zero values) sufficiently, such that audible artifacts, which would be caused, in some cases, if the second patch was only sparsely  
 10 populated, are reduced or avoided.

To summarize, the inventive concept brings along significant advantages over conventional patching methods, because the harmonic bandwidth-extension, using the phase vocoder, is applied only for obtaining values of the spectral-domain representation  
 15 of the first patch, i.e. for the lower part of the spectrum, while a non-harmonic bandwidth extension, which relies on a copying of values of the spectral-domain representation of the first patch to obtain values of the spectral-domain representation of the first patch, is used for higher frequencies. Accordingly, the lower range (which is also designated as "first patch") of the extension-frequency portion (which is a frequency portion above the  
 20 crossover frequency) is provided as a harmonic extension of the fundamental frequency range (i.e. in the frequency range of the input signal, which covers frequencies lower than the frequencies of the extension frequency portion, for example frequencies below the crossover frequency), which brings along a good hearing impression of the bandwidth-extended signal. Also, it has been found that the simple generation of the values of the  
 25 spectral domain representation of the higher range of the extension-frequency portion (which is also designated as "second patch"), which is performed using the copier, does not bring along significant auditory artifacts because the human hearing is not particularly sensitive to spectral details of the higher range of the extension-frequency portion (second patch).

30 To summarize, the inventive concept brings along a good hearing impression at a comparatively small computational complexity.

In a preferred embodiment the phase vocoder is configured to copy a set of magnitude  
 35 values associated with a plurality of given frequency subranges of the input spectral representation, to obtain a set of magnitude values associated with corresponding frequency subranges of the first patch, wherein a pair of a given frequency subrange of the input spectral representation and a corresponding frequency subrange of the first patch

covers (or comprises) a pair of a fundamental frequency and a harmonic of the fundamental frequency (for example a first harmonic of the fundamental frequency). The phase vocoder is also preferably configured to multiply phase values associated with the plurality of given frequency subranges of the input spectral representation with a predetermined factor (for example 2), to obtain phase values associated with corresponding frequency subranges of the first patch. Preferably, the value copier is configured to copy a set of values associated with a plurality of given frequency subranges of the first patch, to obtain a set of values associated with corresponding frequency subranges of the second patch. The value copier is preferably configured to leave phase values unchanged in the copying. Accordingly, the phase vocoder performs, at least approximately, a harmonic transposition, while the value copier performs a non-harmonic frequency shift. The frequency subranges may for example be frequency ranges associated with coefficients of a Fast Fourier Transform (or any comparable transform). Alternatively, the frequency subranges may be frequency ranges associated with individual signals of a QMF filterbank. Typically, a width of the frequency subranges is comparatively small compared to the center frequency, such that frequency subranges cover a frequency span having a frequency ratio between an end frequency and a starting frequency, which is significantly smaller than 2:1. In other words, even though the frequency subranges of the input spectral representation (which may, for example, take the form of FFT coefficients, or the form of QMF filterbank signals) and the frequency subranges of the first patch do not need to be exactly harmonic with respect to each other, it is typically possible to identify an association between a frequency subrange (e.g., having frequency index  $k$ ) of the input spectral representation and a corresponding frequency subrange (e.g., having frequency index  $2k$ ) of the first patch, such that the frequency subrange ( $2k$ ) of the first patch represents, at least approximately, a harmonic frequency of the corresponding frequency subrange ( $k$ ) of the input spectral representation.

Accordingly, a harmonic transposition is performed by the phase vocoder, taking into account the phase values, which are processed using a phase scaling. In contrast, the value copier merely performs (at least approximately), a non-harmonic frequency-shift operation.

In a preferred embodiment, the value copier is configured to copy the values such that a common spectral shift (or frequency shift) of values of the first patch onto values of the second patch is obtained.

In a preferred embodiment, the phase vocoder is configured to obtain the values of the spectral-domain representation of the first patch such that the values of the spectral-domain representation of the first patch represent a harmonically upconverted version of a

fundamental frequency range of the input signal representation (for example, a fundamental frequency range below a so-called crossover frequency). The value copier is preferably configured to obtain the values of the spectral-domain representation of the second patch such that the values of the spectral-domain representation of the second patch represent a frequency-shifted version of the first patch. Accordingly, the above described advantages are obtained. In particular, the implementation is simple while obtaining a good auditory impression.

In a preferred embodiment, the apparatus is configured to receive pulse-code-modulated (PCM) input audio data, to down-sample the pulse-code-modulated input audio data in order to obtain down-sampled pulse-code-modulated audio data. Also, the apparatus is configured to window the down-sampled pulse-code-modulated audio data, in order to obtain windowed input data, and to convert or transform the windowed input data into a frequency-domain, in order to obtain the input signal representation. The apparatus is also preferably configured to compute magnitude values  $a_k$  (also designated with  $\alpha_k$ ) and phase values  $\varphi_k$ , representing a frequency bin  $k$  (wherein  $k$  is a frequency bin index) of the input signal representation, and to copy the magnitude values  $a_k$ , to obtain copied magnitude values  $a_{sk}$  (also designated with  $\alpha_{sk}$ ) representing a frequency bin having a frequency bin index  $sk$  of the first patch, wherein  $s$  is a stretching factor with  $s=2$ . Also, the apparatus is preferably configured to copy and scale phase values  $\varphi_k$  associated with a frequency bin having frequency bin index  $k$  of the input signal representation, to obtain copied and scaled phase values  $\varphi_{sk}$  associated with a frequency bin having a frequency index  $sk$  of the first patch. Also, the apparatus is preferably configured to copy values  $\beta_{k-i\zeta}$  associated with a frequency bin  $k-i\zeta$  of the spectral-domain representation of the first patch, to obtain values  $\beta_k$  of the spectral-domain representation of the second patch. Also, the apparatus is preferably configured to convert the representation of the bandwidth-extended signal (which comprises the spectral-domain representation of the first patch and the spectral-domain representation of the second patch) into the time-domain, to obtain a time-domain representation, and to apply a synthesis window to the time-domain representation. Using the above-described concept, it is possible to obtain a bandwidth-extended signal with moderate computational complexity. The bandwidth-extension is performed in the frequency-domain, wherein a transform may be performed into a spectral domain, for example, into a FFT domain or a QMF domain.

In a preferred embodiment, the apparatus comprises a time-domain to spectral-domain converter (for example, a Fast-Fourier-Transform means or a QMF filterbank) configured to provide, as the input signal representation, values of a spectral domain representation

(for example, Fast-Fourier-Transform coefficients or QMF subband signals) of an input audio signal, or of a preprocessed (e.g. down-sampled and/or windowed) version of the input audio signal (for example a pulse-code-modulated signal provided by an audio decoder core). The apparatus preferably comprises a spectral-domain to time-domain converter (for example, an inverse Fast-Fourier-Transform means or a QMF synthesis means) configured to provide a time-domain representation of the bandwidth-extended signal using values of the spectral-domain representation (e.g. FFT coefficients, or QMF subband signals) of the first patch and values of the spectral domain representation (e.g. FFT coefficients, or QMF subband signals) of the second patch. The spectral-domain to time-domain converter is preferably configured such that a number of different spectral values (e.g. FFT bins or QMF bands) received by the spectral-domain-to-time-domain converter is larger than a number of different spectral values (e.g. a number of FFT frequency bins, or a number of QMF bands) provided by the time-domain-to-spectral-domain converter (e.g. Fast-Fourier-Transform means or QMF filterbank), such that the spectral-domain-to-time-domain converter is configured to process a larger number of frequency bins (e.g. Fast-Fourier-Transform frequency bins or QMF frequency bands) than the time-domain-to-frequency-domain converter. Accordingly, a bandwidth-extension is reached by the fact that the spectral-domain-to-time-domain converter comprises a larger number of frequency bins than the time-domain-to-frequency-domain converter.

In a preferred embodiment, the apparatus comprises an analysis windower configured to window a time-domain input audio signal, to obtain a windowed version of the time-domain input audio signal, which forms the basis for obtaining the input signal representation. Also, the apparatus comprises a synthesis windower configured to window a portion of a time-domain representation of the bandwidth-extended signal, to obtain a windowed portion of the time-domain representation of the bandwidth-extended signal. Accordingly, artifacts in the bandwidth-extended signal are reduced or even avoided.

In a preferred embodiment, the apparatus is configured to process a plurality of temporally overlapping time-shifted portions of the time-domain input audio signal, to obtain a plurality of temporally overlapping time-shifted windowed portions of the time-domain representation of the bandwidth-extended signal. A time-offset between temporally adjacent time-shifted portions of the time-domain input audio signal is smaller than or equal to one fourth of a window length of the analysis window. It has been found that a comparatively large temporal overlap between adjacent time-shifted portions of the time-domain input audio signal (and/or a comparatively large temporal overlap between temporally adjacent time-shifted portions of the time-domain representation of the bandwidth-extended signal) results in a bandwidth-extension bringing along a good

hearing impression, because non-stationarities of the signal are taken into account because of the comparatively large temporal overlap.

In a preferred embodiment, the apparatus comprises a transient information provider  
 5 configured to provide an information indicating the presence of a transient in the input  
 signal (represented by the input signal representation). The apparatus also comprises a first  
 processing branch for providing a representation of a bandwidth-extended signal portion  
 on the basis of a non-transient portion of the input signal representation and a second  
 10 processing branch for providing a representation of a bandwidth-extended signal portion  
 on the basis of a transient portion of the input signal representation. The second processing  
 branch is configured to process a spectral-domain representation of the input signal having  
 a higher spectral resolution than a spectral domain representation of the input signal  
 processed by the first processing branch. Accordingly, signal portions comprising a  
 15 transient can be treated with higher spectral resolution, which avoids audible artifacts in  
 the presence of transients. On the other hand, a reduced spectral resolution can be used for  
 non-transient signal portions (i.e. for signal portions in which the transient information  
 provider does not identify a transient). Thus, a computational efficiency is kept high, and  
 the increased spectral resolution is used only when it brings along advantages (for  
 example, in that it results in a better hearing impression in the proximity of transients).

20 In a preferred embodiment, the apparatus comprises a time-domain zero-padder configured  
 to a zero-pad a transient portion of the input signal, in order to obtain a temporally  
 extended transient portion of the input signal. In this case, the first processing branch  
 comprises a (first) time-domain-to-frequency-domain converter configured to provide a  
 25 first number of spectral domain values associated with a non-transient portion of the input  
 signal, and the second processing branch comprises a (second) time-domain-to-frequency-  
 domain converter configured to provide a second number of spectral domain values  
 associated with the temporally extended transient portion of the input signal. The second  
 number of spectral-domain values is larger, at least by a factor of 1.5, than the first number  
 30 of spectral domain values. Accordingly, a good transient handling is obtained.

In a preferred embodiment, the second processing branch comprises a zero-stripper  
 configured to remove a plurality of zero values from a bandwidth-extended signal portion  
 obtained on the basis of the temporally extended transient portion of the input signal.  
 35 Accordingly, the temporal extension of the input signal, which is obtained by the zero-  
 padding, is reversed.

In a preferred embodiment, the apparatus comprises a down-sampler configured to down-sample a time-domain representation of the input signal. By down-sampling the input signal, a computational efficiency can be improved if the input signal does not cover the full Nyquist bandwidth of a pulse-code-modulated sample input stream.

5

Another embodiment according to the invention creates an apparatus, in which the processing order of the processing by the value copier and the phase vocoder is inversed.

15. Such an apparatus for generating a representation of a bandwidth-extended signal on the basis of an input signal representation (110; 383) comprises a value copier configured to copy a set of values of the input signal representation, to obtain a set of values of a spectral domain representation of a first patch, wherein the first patch is associated with higher frequencies than the input signal representation. The apparatus also comprises a phase vocoder (130; 406) configured to obtain values ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ ) of a spectral domain representation of a second patch of the bandwidth-extended signal on the basis of the values ( $\beta_{4/3\zeta} \dots \beta_{2\zeta}$ ) of the spectral domain representation of the first patch, wherein the second patch is associated with higher frequencies than the first patch. The apparatus is configured to obtain the representation (120; 426) of the bandwidth-extended signal using the values of the spectral domain representation of the first patch and the values of the spectral domain representation of the second patch.

20

This apparatus is capable of obtaining a bandwidth-extended signal with comparatively low computational complexity while still achieving a good hearing impression of the bandwidth-extended signal. By performing the phase vocoding after the copying operation, the phase vocoder can be operated with a comparatively small frequency ratio (ratio between vocoder output frequency and vocoder input frequency), which results in a good spectral filling and avoids the presence of large spectral holes. Also, it has been found that The hearing impression using this concept is still better than for a concept which merely relies on copying operations, without a phase vocoder action, even though the first patch (lower frequency patch) is obtained using the copying operation, and only the second patch (higher frequency patch) is obtained using the phase vocoding operation. Also, computational complexity is smaller than in systems in which all of the patches are generated using phase vocoders, and spectral holes are reduced when compared to such concepts.

35

Naturally, this embodiment can be supplemented by any of the functionalities discussed herein.

Other embodiments according to the invention create methods for generating a representation of a bandwidth-extended signal on the basis of an input signal representation. Said method is based on the same ideas as the above-discussed apparatus.

- 5 Another embodiment according to the invention creates a computer program for implementing the method.

#### Brief Description of the Figures

- 10 Fig. 1 shows a block-schematic diagram of an apparatus for generating a representation of a bandwidth-extended signal on the basis of an input signal representation, according to an embodiment of the invention;
- 15 Fig. 2 shows a schematic representation of the bandwidth extension concept, according to the present invention;
- 20 Fig. 3 shows a detailed block-schematic diagram of an audio decoder comprising an apparatus for generating a representation of a bandwidth-extended signal on the basis of an input signal representation, according to an embodiment of the invention;
- 25 Fig. 4 shows a flowchart of a method for generating a representation of a bandwidth-extended signal on the basis of an input signal representation, according to an embodiment of the invention;
- Fig. 5 shows a block-schematic diagram of an audio decoder, according to a first comparison example; and
- 30 Fig. 6 shows a block-schematic diagram of an audio decoder, according to a second comparison example.

#### Detailed description of the embodiments

##### 1. Apparatus according to Fig. 1

- 35 Fig. 1 shows a block-schematic diagram of an apparatus 100 for generating a representation of a bandwidth-extended signal on the basis of an input signal

representation. The apparatus 100 is configured to receive an input signal representation 110 and provide, on the basis thereof, a bandwidth-extended signal 120. The apparatus 100 comprises a phase vocoder configured to obtain values of a spectral-domain representation 130 of a first patch of the bandwidth-extended signal 120 on the basis of the input signal representation 110. The values of the spectral domain representation of the first patch are designated, for example, with  $\beta_{\zeta}$  to  $\beta_{2\zeta}$ . The apparatus 100 also comprises a value copier 140 configured to copy a set of values of the spectral-domain representation 132 of the first patch, which are provided by the phase vocoder 130, to obtain a set of values of a spectral domain representation 142 of a second patch, wherein the second patch is associated with higher frequencies than the first patch. The values of the spectral domain representation 142 of the second patch are designated, for example, with  $\beta_{2\zeta}$  to  $\beta_{3\zeta}$ . The apparatus 100 is configured to obtain the representation 120 of the bandwidth-extended signal using the values  $\beta_{\zeta}$  to  $\beta_{2\zeta}$  of the spectral domain representation 132 of the first patch and the values  $\beta_{2\zeta}$  to  $\beta_{3\zeta}$  of the spectral domain representation 142 of the second patch. For example, the representation 120 of the bandwidth-extended signal may comprise both the values of the spectral domain representation 132 of the first patch and the spectral domain representation 142 of the second patch. In addition, the representation 120 of the bandwidth-extended signal may, for example, comprise values of a spectral domain representation of the input signal (represented, for example, by the input signal representation 110). However, the representation 120 of the bandwidth-extended signal may also be a time-domain representation, which may be based on the values of the spectral domain representation 132 of the first patch and the values of the spectral domain representation 142 of the second patch (and, optionally, additional values, for example values of the spectral domain representation 116 of the input signal, and/or values of a spectral domain representation of additional patches).

In the following, the functionality and operation of the apparatus 100 will be described in detail taking reference to Fig. 2, which shows a schematic representation of the inventive concept for generating a representation of a bandwidth-extended signal on the basis of an input signal representation.

A first graphic representation 200 shows a harmonic transposition of the input signal (represented by the input signal representation 110), which is performed by the phase vocoder 130. As can be seen, the input signal is represented, for example, by a set of magnitude values  $\alpha_k$ . The index  $k$  designates a spectral bin (for example a bin having index  $k$  of a fast Fourier transform, or a frequency band having index  $k$  of a QMF conversion). The input signal representation 110 may, for example, comprise magnitude values  $\alpha_k$  for  $k = 1$  to  $k = \zeta$ , wherein  $\zeta$  may designate a so-called cross-over frequency bin and describes a

frequency onset of the bandwidth-extension. A fundamental frequency range is further described, for example, by phase values  $\varphi_k$ , wherein  $k$  is a frequency bin index, as discussed before.

5 Similarly, the first patch is described by a set of values of a spectral domain representation, for example, values  $\beta_k$  with  $k$  between  $\zeta$  and  $2\zeta$ . Alternatively, the first patch may be represented by magnitude values  $\alpha_k$  and phase values  $\varphi_k$ , with the frequency bin index  $k$  between  $\zeta$  and  $2\zeta$ .

10 As mentioned, the phase vocoder 130 is configured to perform a harmonic transposition on the basis of the input signal representation 110 to obtain values of the spectral-domain representation 132 of the first patch. For this purpose, the phase vocoder 130 may set a magnitude value  $\alpha_{2k}$  of a frequency bin having (frequency bin) index  $2k$  to be equal to the magnitude value  $\alpha_k$  of a frequency bin having (frequency bin) index  $k$ . Also, the phase  
15 vocoder 130 may be configured to set the phase value  $\varphi_{2k}$  of a frequency bin having index  $2k$  to a value which is equal to 2 times the phase value  $\varphi_k$  associated with the frequency bin having index  $k$ . In this case, the frequency bin having index  $k$  may be a frequency bin of the input signal representation 110, and the frequency bin having index  $2k$  may be a frequency bin of the spectral-domain representation 132 of the first patch. Also, a  
20 frequency bin having index  $2k$  may comprise a frequency, which is a first harmonic of a frequency included in the frequency bin having index  $k$ . Accordingly, magnitude values  $\alpha_{2k}$  and phase values  $\varphi_{2k}$  may be obtained, which are values of the spectral domain representation 132 of the first patch, for  $2k$  ranging from  $\zeta$  to  $2\zeta$ , such that  $\alpha_{2k} = \alpha_k$  and  $\varphi_{2k} = 2\varphi_k$ . Alternatively, and equivalently, values  $\beta_{2k}$ , which are values of the spectral-  
25 domain representation 132 of the first patch, may be obtained for  $2k$  between  $\zeta$  and  $2\zeta$ , such that  $\beta_{2k} = \alpha_k e^{j2\varphi_k}$ .

To summarize, assuming that the frequency bins having indices  $k$  (or equivalently,  $2k$ , and so on), which are, for example, frequency bins of a Fast Fourier Transform representation  
30 or frequency bands of a QMF domain representation, are spaced linearly in frequency (such that the frequency bin index, e.g.  $k$  or  $2k$ , is at least approximately proportional to a frequency comprised in the respective frequency bin, for example, a center frequency of a  $k$ -th Fast Fourier Transform frequency bin or a center frequency of a  $k$ -th QMF band), a harmonic transposition is obtained by the phase vocoder 130.

35 However, the values of the spectral-domain representation 142 of the second patch are obtained by the value copier 140, which performs a non-harmonic copying up of values of the spectral-domain representation 132 of the first patch.

Taking reference now to the graphical representation 250, the non-harmonic copying up will be briefly discussed. As can be seen, the first patch is represented by values  $\beta_{\zeta}$  to  $\beta_{2\zeta}$  (or, equivalently, by magnitude values  $\alpha_{\zeta}$  to  $\alpha_{2\zeta}$  and phase values  $\varphi_{\zeta}$  to  $\varphi_{2\zeta}$ .  
 5 Accordingly, the values  $\beta_{2\zeta}$  to  $\beta_{3\zeta}$  (or, equivalently, magnitude values  $\alpha_{2\zeta}$  to  $\alpha_{3\zeta}$  and phase values  $\varphi_{2\zeta}$  to  $\varphi_{3\zeta}$ ) of the spectral-domain representation 142 of the second patch are obtained by a non-harmonic copying, which is performed by the value copier 140. For example, complex-valued spectral values  $\beta_{2\zeta}$  to  $\beta_{3\zeta}$  of the spectral-domain representation 142 of the second patch may be obtained on the basis of corresponding values  $\beta_{\zeta}$  to  $\beta_{2\zeta}$  of  
 10 the spectral-domain representation 132 of the first patch according to  $\beta_k = \beta_{k-\zeta}$  for  $k$  between  $2\zeta$  and  $3\zeta$ . Equivalently, magnitude values  $\alpha_{2\zeta}$  to  $\alpha_{3\zeta}$  of the spectral-domain representation 142 of the second patch may be obtained on the basis of magnitude values of the spectral domain representation 132 of the first patch according to  $\alpha_k = \alpha_{k-\zeta}$  for  $k$  between  $2\zeta$  and  $3\zeta$ . In this case, phase values  $\varphi_{2\zeta}$  to  $\varphi_{3\zeta}$  of the spectral-domain  
 15 representation 142 of the second patch may be obtained on the basis of phase values  $\varphi_{\zeta}$  to  $\varphi_{2\zeta}$  of the spectral-domain representation 132 of the first patch according to  $\varphi_k = \varphi_{k-\zeta}$  for  $k$  between  $2\zeta$  and  $3\zeta$ .

Accordingly, the values of the spectral-domain representation 142 of the second patch represent a signal, which is non-harmonically (i.e. linearly) frequency-shifted with respect  
 20 to a signal represented by the values of the spectral-domain representation 132 of the first patch.

The values  $\beta_{\zeta}$  to  $\beta_{2\zeta}$  of the spectral-domain representation 132 of the first patch and the values  $\beta_{2\zeta}$  to  $\beta_{3\zeta}$  of the spectral-domain representation 142 of the second patch may be used to obtain the representation 120 of the bandwidth-extended signal. Depending on the  
 25 requirements, the representation 120 of the bandwidth-extended signal may be a spectral-domain representation or a time-domain representation. If it is desired to obtain a time-domain representation, a frequency-domain-to-time-domain converter may be used to derive the time-domain representation on the basis of the values  $\beta_{\zeta}$  to  $\beta_{2\zeta}$  of the spectral-domain representation 132 of the first patch and the values  $\beta_{2\zeta}$  to  $\beta_{3\zeta}$  of the spectral-domain representation 142 of the second patch. Alternatively (and equivalently) the values  
 30  $\alpha_{\zeta}$  to  $\alpha_{2\zeta}$ ,  $\varphi_{\zeta}$  to  $\varphi_{2\zeta}$ ,  $\alpha_{2\zeta}$  to  $\alpha_{3\zeta}$  and  $\varphi_{2\zeta}$  to  $\varphi_{3\zeta}$  may be used in order to derive the representation 120 of the bandwidth-extended signal (either in the spectral-domain or in the time-domain).

As discussed above, the concept described with respect to Figs. 1 and 2 brings along a  
 35 good hearing impression and comparatively low computational complexity. Phase

vocoding is only required once, even though a plurality of patches (for example the first patch and the second patch) are used. Also, it is avoided that there are large spectral holes in the second patch, which would occur if another phase vocoder was used to obtain the second patch. Thus, the inventive concept brings along a very good tradeoff between computational complexity and an achievable hearing impression.

Moreover, it should be noted that additional patches may be obtained on the basis of the values of the spectral-domain representation 132 of the first patch in some embodiments. For example, in an optional extension of the inventive concept, values of a spectral-domain representation of a third patch may be obtained on the basis of the values of the spectral domain representation 132 of the first patch using another value copier, as will be described in more detail taking reference to Fig. 3.

The embodiments according to Figs 1 and 2 (and also the other embodiments) can be modified in a wide variety of ways. For example A first patch can be obtained using a phase vocoder, and second, third and fourth patches can be obtained by a copying-up operation of spectral values. Alternatively, a first and a second patch can be obtained using phase vocoders, and a third and a fourth patch can be obtained using a copying-up of spectral values. Naturally, different combinations of the phase vocoding operation and the copying-up operation can be applied.

Alternatively, however, a first patch can be obtained using a copying-up operation (value copier) of spectral values off the input signal representation, and a second patch can be obtained using a phase vocoder (on the basis of the copied values of the first patch, obtained using the value copier).

## 2. Embodiment According to Fig. 3

In the following, an audio decoder 300 will be described taking reference to Fig. 3, wherein Fig. 3 shows a detailed block-schematic diagram of such an audio decoder 300 comprising an apparatus for a generating a representation of a bandwidth-extended signal on the basis of an input signal representation.

### 2.1. Audio decoder overview

The audio decoder 300 is configured to receive a data stream 310 and to provide, on the basis thereof, an audio waveform 312. The audio decoder 300 comprises a core decoder 320, which is configured to provide, for example, pulse-code-modulated data ("PCM data") 322 on the basis of the data stream 310. The core decoder 320 may for example be an audio decoder as described in the international standard ISO/IEC 14496-3:2005(e), part 3: audio, subpart 4: general audio coding (GA)-AAC, Twin VQ, BSAC. For example, the core decoder 320 may be a so-called advanced-audio-coding (AAC) core decoder, which is described in said standard, and which is well-known to the man skilled in the art. Thus, the pulse-code-modulated audio data 322 may be provided by the core decoder 220 on the basis of the data stream 310. For example, the pulse-code-modulated audio data 322 may comprise the frame length of 1024 samples.

The audio decoder 300 also comprises a bandwidth-extension (or bandwidth extender) 330, which is configured to receive the pulse-code-modulated audio data 322 (for example, a frame length of 1024 samples) and to provide, on the basis thereof, the waveform 312. The bandwidth-extension (or bandwidth extender) 330 also receives some control data 332 from the data stream 310. The bandwidth-extension 330 comprises a patched QMF data provision (or patched QMF data provider) 340, which receives the pulse-code-modulated audio data 322 and which provides, on the basis thereof, patched QMF data 342. The bandwidth-extension 330 also comprises an envelope formatting (or envelope formatter) 344, which receives the patched QMF data 342 and envelope formatting control data 346 and provides, on the basis thereof, patched and envelope-formatted QMF data 348. The bandwidth-extension 330 also comprises a QMF synthesis (or QMF synthesizer) 350, which receives the patched and envelope-formatted QMF data 348 and provides, on the basis thereof, the waveform 312 by performing a QMF synthesis.

## 2.2. Patched QMF data provision 340

### 2.2.1. Patched QMF data provision - Overview

The patched QMF data provision 340 (which may be performed by a patched QMF data provider 340 in a hardware implementation) may be switchable between two modes, namely a first mode, in which a spectral band replication (SBR) patching is performed, and a second mode in which a harmonic bandwidth-extension (HBE) patching is performed. For example, the pulse-code-modulated audio data 322 may be delayed by a delayer 360, to obtain delayed pulse-code-modulated audio data 362, and the delayed pulse-code-modulated audio data 362 may be converted into a QMF domain using a 32 band QMF analyzer 364. The result of the 32 band QMF analyzer 364, for example, a 32 band QMF domain (i.e. spectral-domain) representation 365 of the delayed pulse-code-modulated

audio data 362, may be provided to a SBR patcher 366 and to a harmonic bandwidth-extension patcher 368.

The spectral band replication patcher 366 may, for example, perform a spectral band replication patching, which is described, for example, in section 4.6.18 “SBR tool” of the international standard ISO/IEC 14496-3:2005(e), part 3, subpart 4. Accordingly, a 64 band QMF domain representation 370 may be provided by the spectral-band-replication patcher 366.

Alternatively, or in addition, the harmonic-bandwidth-extension patcher 368 may provide a 64 band QMF domain representation 372, which is a bandwidth-extended representation of the PCM audio data 322. A switch 374, which is controlled in dependence on bandwidth-extension control data 332 extracted from the data stream 310, may be used to decide whether the spectral band replication patching 366 or the harmonic bandwidth-extension patching 368 is applied in order to obtain the patched QMF data 342 (which may be equal to the a 64 band QMF domain representation 370 or equal to the 64 band QMF domain representation 372 depending on the state of the switch 374) .

### 2.2.2. Patched QMF data provision - Harmonic Bandwidth-Extension 368

In the following, the (at least partially) harmonic bandwidth-extension patching 368 will be described in more detail. The harmonic bandwidth-extension patching 368 comprises a signal path, in which pulse-code-modulated audio data 322, or a pre-processed version thereof, are converted into a spectral-domain (for example into a Fast-Fourier-Transform coefficient domain or a QMF domain), in which a harmonic bandwidth-extension is performed in the spectral-domain, and in which the obtained spectral domain representation of the bandwidth-extended signal, or a representation derived therefrom, is used for the harmonic bandwidth-extension patching.

In the embodiment of Fig. 3, the pulse-code-modulated audio data 322 are down-sampled in a down-sampler 380, for example, by a factor of 2, to obtain down-sampled pulse-code-modulated audio data 381. The down-sampled pulse-code-modulated audio data 381 are subsequently windowed by a windower 382, which may, for example, comprise a window length of 512 samples. It should be noted that the window is, for example, shifted by 64 samples of the down-sampled pulse-code-modulated audio data 381 in subsequent processing steps, such that a comparatively large overlap of the windowed portions 383 of the down-sampled pulse-code-modulated audio data is obtained.

The audio decoder 300 also comprises a transient detector 384, which is configured to detect a transient within the pulse-code-modulated audio data 322. The transient detector 384 may detect the presence of a transient either on the basis of the PCM audio data 322  
 5 itself, or on the basis of a side information, which is included in the data stream 310.

The windowed portions 383 of the down-sampled PCM audio data 381 can be selectively processed using a first processing branch 386 or a second processing branch 388. The first branch 386 may be used for processing a non-transient windowed portion 383 of the down-sampled PCM audio data (for which the transient detector 384 denies the presence of a  
 10 transient), and a second branch 388 may be used for a processing of a transient windowed portion 383 of the down-sampled PCM audio data (for which the transient detector 384 indicates the presence of a transient).

The first branch 386 receives a non-transient windowed portion 383 and provides, on the basis thereof, a bandwidth-extended representation 387,434 of the windowed portion 383. Similarly, the second branch 388 receives a transient windowed portion 383 of the down-sampled PCM audio data 381 and provides, on the basis thereof, a bandwidth-extended representation 389 of the (transient) windowed portion 383. As discussed above, the  
 15 transient detector 384 decides whether the current windowed portion 383 is a non-transient windowed portion or a transient windowed portion, such that the processing of the current windowed portion 383 is performed either using the first branch 386 or the second branch 388. Thus, different windowed portions 383 may be processed by different branches 386, wherein there is a significant temporal overlap between the subsequent bandwidth-extended representations 387, 389 of the subsequent windowed portions 383 (because there  
 20 is a significant temporal overlap of temporally subsequent windowed portions 383).

The harmonic bandwidth-extension 368 further comprises an overlapper-and-adder 390, which is configured to overlap-and-add the different bandwidth-extended representations 387, 389 associated with different (temporally subsequent) windowed portions 383. An  
 30 overlap-and-add increment may, for example, be set to 256 samples. Accordingly, an overlapped-and-added signal 392 is obtained.

The harmonic bandwidth-extension 368 also comprises a 64-band QMF analyzer 394, which is configured to receive the overlapped-and-added signal 392 and to provide, on the basis thereof, a 64-band QMF domain signal 396. The 64 band QMF-domain signal 396 may for example represent a broader frequency range than the 32-band QMF domain  
 35 signal 365 provided by the 32-band QMF analyzer 364.

The harmonic bandwidth-extension 368 also comprises a combiner 398, which is configured to receive both the 32-band QMF-domain signal provided by the 32-band QMF analyzer 364 and the 64-band QMF domain signal 396 and to combine those signals. For example, the low-frequency-range (or fundamental frequency range) components of the 64-band QMF domain signal 396 may be replaced by, or combined with, the 32-band QMF-domain signal 365 provided by the 32-band QMF analyzer 364, such that, for example, the 32 lower-frequency-range (or fundamental frequency range) components of the 64-band QMF domain signal 372 are determined by the output of the 32-band QMF analyzer 364, and such that the 32 higher-frequency-range components of the 64-band QMF-domain signal 372 are determined by the 32 higher-frequency-range components of the 64-band QMF domain signal 396.

Naturally, the number of components of the QMF-domain signals may vary, depending on the specific requirements. Naturally, a frequency position of a transition between a fundamental frequency range (also designated as lower-frequency-range) and a bandwidth-extended frequency range (also designated as higher-frequency-range) may depend on the cross-over frequency, or, equivalently, the bandwidth of the audio signal represented by the pulse-code-modulated audio data 322.

In the following, details regarding the first processing branch 386 will be described. The first branch 386 comprises a time-domain-to-frequency-domain converter 400, which is implemented, for example, in the form of a Fast-Fourier-Transform-means configured to provide 512 Fast-Fourier-Transform coefficients on the basis of a windowed portion 383 of 512 time-domain samples of the down-sampled pulse-code-modulated audio data 381. Accordingly, the Fast-Fourier-Transform frequency bins are designated with subsequent integer frequency bin indices  $k$  in a range between 1 and  $N=512$ .

The first branch 386 also comprises a magnitude value provider 402, which is configured to provide magnitude values  $\alpha_k$  of the Fast-Fourier-Transform coefficients. Also, the first branch 386 comprises a phase value provider 404 configured to provide phase values  $\phi_k$  of the Fast-Fourier-Transform coefficients.

The first branch 386 also comprises a phase vocoder 406, which may receive the magnitude values  $\alpha_k$  and the phase values  $\phi_k$  as an input signal representation, and which may comprise the functionality of the phase vocoder 130 discussed above. Accordingly, the phase vocoder 406 may output values  $\beta_{2k}$ , in a range between  $\beta_{\xi}$  and  $\beta_{2\xi}$ , of a spectral

domain representation of a first patch. The values  $\beta_{2k}$  are designated with 408, and may be equivalent to the values of the spectral-domain representation 132 of a first patch. The first branch 386 also comprises a value copier 410, which may take over the functionality of the value copier 140, and which may receive, as an input information, the values  $\beta_{2k}$  (e.g. in a range between  $\beta_{\xi}$  and  $\beta_{2\xi}$ ). Accordingly, the first value copier 410 may provide values  $\beta_k$  in a range between  $\beta_{2\xi}$  and  $\beta_{3\xi}$ , which are designated with 412 and which may be equivalent to the values  $\beta_{2\xi}$  to  $\beta_{3\xi}$  of the spectral-domain representation 142 of the second patch. Also, the first branch 386 may (optionally) comprise a second value copier 414, which is configured to receive the values  $\beta_{\xi}$  and  $\beta_{2\xi}$  (also designated with 408) provided by the phase vocoder 406 and to provide, on the basis thereof, spectral values  $\beta_{3\xi}$  to  $\beta_{4\xi}$  using a copy-operation (which effectively results in a non-harmonic frequency-shift of the spectrum described by the values  $\beta_{\xi}$  to  $\beta_{2\xi}$  (408)). Accordingly, the second value copier 414 provides spectral values  $\beta_{3\xi}$  to  $\beta_{4\xi}$  of a spectral-domain representation of a third patch, which are also designated 416.

The first branch 386 may comprise an optional interpolator 420, which may be configured to receive the values 412, 416 of the spectral-domain representations of the second patch and of the third patch (and, optionally, also the values 408 of the spectral domain representation of the first patch) and to provide interpolated values 422 of the spectral-domain representation of the second and third patch (and, optionally, also of the first patch).

The first branch 386 may additionally comprise a zero padder 424, which is configured to receive the interpolated values 422 (or, alternatively, the original values 412, 416) of the spectral-domain representations of the second and third patch (and, optionally also of the first patch) and to obtain, on the basis thereof, a zero-padded version of values of a spectral-domain representation, which is zero-padded in order to be adapted to a dimension of a spectral-domain-to-time-domain converter 428.

The spectral-domain-to-time-domain converter 428 may be implemented, for example, as an inverse Fast-Fourier-Transformer. For example, the inverse Fast-Fourier-Transformer 428 may be configured to receive a set of 2048 (optionally interpolated and zero-padded) spectral values, and to provide, on the basis thereof, a time-domain representation 430 of the bandwidth-extended signal portion. The first path 386 also comprises a synthesis windower 432, which is configured to receive the time-domain representation 430 of the bandwidth-extended signal portion and to apply a synthesis windowing, in order to obtain a synthesis-windowed time-domain representation of the bandwidth-extended signal portion 430.

The audio decoder 300 also comprises a second processing path 388, which performs a very similar processing when compared to the first path 386. However, the second path 388 comprises a time-domain zero-padder 438, which is configured to receive the windowed transient portion 383 of the down-sampled pulse-code-modulated audio data 381 and to derive a zero-padded version 439 from the windowed portion 383, such that a beginning of the zero-padded portion 439 and an end of the zero-padded portion 439 are padded with zeros, and such that the transient is arranged in a central region (between the zero padded beginning samples and the zero-padded end samples) of the zero-padded portion 439.

The second path 388 also comprises a time-domain-to-spectral-domain transformer 440, for example, a Fast-Fourier-Transformer or a QMF (quadrature-mirror-filterbank). The time-domain-to-spectral-domain transformer 440 typically comprises a larger number of frequency bins (for example, Fast-Fourier-Transform frequency bins, or QMF bands) than the time-domain-to-spectral-domain transformer 400 of the first branch. For example, the Fast-Fourier-Transformer 440 may be configured to derive 1024 Fast-Fourier-Transform coefficients from a zero-padded portion 439 of 1024 time domain samples.

The second branch 388 also comprises a magnitude value determinator 442 and a phase value determinator 444, which may comprise the same functionality as the corresponding means 402, 404 of the first branch 386, though with increased dimension  $N=1024$ .

Similarly, the second branch 388 also comprises a phase vocoder 446, a first value copier 450, a second value copier 454, an optional interpolator 460, and an optional zero padder 464, which may comprise the same functionalities as the corresponding means of the first branch 386, though with increased dimensions. In particular, the index  $\xi$  of the cross-over band may be higher in the second branch 388 than the first branch 386, for example, by a factor of 2.

Accordingly, a spectral-domain representation comprising, for example, 4096 Fast-Fourier-Transform coefficients may be provided to an inverse Fast-Fourier-Transformer 468, which in turn provides a time-domain signal 470 having 4096 samples.

The second branch 388 also comprises a synthesis windower 472, which is configured to provide a windowed version of the time-domain-representation 470 of the bandwidth-extended signal portion.

The second branch 388 also comprises a zero stripper configured to provide a shortened, windowed time-domain representation 478 of the bandwidth-extended signal portion, which shortened, windowed time-domain representation 478 may, for example, comprise 2048 samples.

5

Accordingly, the time-domain representation 387 is used for non-transient portions (e.g. audio frames) of the pulse-code-modulated audio data 322, and the time-domain representation 478 is used for transient portions of the pulse-code-modulated audio data 322. Accordingly, transient portions are processed with higher spectral-domain resolution in the second processing branch 388, while non-transient portions are processed with lower spectral resolution in the first processing branch 386.

10

### 2.3. Envelope Formatting 344

In the following the envelope formatting 344 will be briefly summarized. In addition, reference is made to the respective remarks in the introductory section, which also apply to the inventive concept.

15

The patched QMF data 342, which are obtained on the basis of the 64 band QMF domain signal 396, are processed by the envelope formatting 344, to obtain the signal representation 348, which is input into the QMF synthesizer 350. The envelope formatting may for example adapt the QMF domain band signals of the patched QMF data 342 in order to perform a noise filling, in order to reconstruct missing harmonics, and/or in order to obtain an inverse filtering. Variations of noise filling, missing harmonics insertion and inverse filtering may for example be controlled by a side information 346, which may be extracted from the data stream 310. For further details, reference is made, for example, to the discussion of the SBR tool in section 4.6.18 of the International Standard ISO/IEC 14496-3:2005(e), part 3, subpart 4. However, different concepts of envelope formatting may also be applied in accordance with the requirements.

20

25

30

### 3. Discussion and comparison of different Solutions

In the following, a brief discussion and summary of the inventive solution will be provided.

35

Embodiments according to the present invention, for example the apparatus 100 according to Fig. 1 and the audio decoder 300 according to Fig. 3, are (or comprise) new patching

algorithms inside spectral band replication (SBR). Spectral domain patching in different manners can be used in order to account for different signal characteristics or restrictions dictated by soft- or hardware requirements.

- 5 In standard SBR, patching is always carried out by a copy operation inside the QMF domain. This can sometimes lead to auditory artifacts, particularly if sinusoids are copied into vicinity of each other at the border of LF and generated HF part. Therefore, a new patching algorithm has been introduced that avoids some problems by using a phase vocoder (see, for example, Reference [13]). This algorithm is illustrated in Fig. 5 as a  
10 comparison example.

The standard SBR has the problem of auditory artifacts. The phase vocoder approach presented in Reference [13] has a complexity, particularly because of the high number of Fast Fourier Transforms that need to be calculated. Additionally, the spectrum becomes  
15 very sparse for high patches (high stretching factors), which may result in undesired audio artifacts.

Two embodiments avoid the high number of Fast Fourier Transforms by moving the generation of different patches from the time domain to the frequency domain. In Fig. 6, an  
20 example is given in which the transformation to the frequency-domain is achieved with the help of a Fast Fourier Transform. Instead of the Fourier Transformation, other time-frequency transformations are, however, useable.

Fig. 3 shows a hybrid solution of the algorithm of Fig. 6 for SBR patching. Only the first  
25 patch is generated by the phase vocoder algorithm (for example, block 406 of the first branch 386, and block 446 of the second branch 388) while higher patches (for example, the second patch and the third patch) are created just by copying the first patch (for example, using the value copiers 410, 414 of the first branch 386, and/or the value copiers 450, 454 of the second branch 388). This yields a less sparse spectrum.  
30

In the following the comparison algorithm, which is implemented in the audio decoder shown in Fig. 6, and the inventive algorithm, which is implemented in the audio decoder shown in Fig. 3, will be shortly explained:

- 35 The comparison algorithm or reference algorithm, which is implemented in the audio decoder shown in Fig. 6, comprises the following steps:

1. Signal downsampling (if Nyquist criterion is not harmed)

2. Signal is windowed ("Hann" windows are proposed but other window shapes may be used) and so called grains (for example, windowed signal portions 383) of lengths  $N$  are taken from the signal. The windows are shifted over the signal with a hop size  $H$ . A  $N/H=8$  times overlap is proposed.
- 5 3. If the grain (for example, a windowed signal portion 383) contains a transient event at the edges, it is padded (for example, by the zero padder 438) with zeros which leads to an oversampling in frequency domain.
4. Grains are transformed to frequency domain (for example, using the time-domain-to-spectral-domain transformers 400,440).
- 10 5. Frequency domain grains are (optionally) padded to a desired output length of the patching algorithm.
6. Magnitude and phase are calculated (for example, using the means 402, 404, 442, 444).
- 15 7. Frequency bin content  $n$  is copied to position  $sn$  for stretching factor  $s$ . The phase is multiplied with the stretching factor  $s$ . This is done for all stretching factors  $s$  (only for the regions in the spectrum that cover the desired patches). (a)  $\zeta \cdot (s-1)/s \leq n \leq \zeta$  or (b)  $\zeta/s \leq n \leq \zeta$ ; (b) yields a more dense spectrum than (a) as the patches overlap. The  $\zeta$  denotes the highest frequency of the LF part, the so called cross over frequency. Generally speaking, the phase is corrected for a new sample position (e.g.,
- 20 frequency position), which can be achieved using the algorithm discussed here or any appropriate alternative algorithm.
8. Frequency domain bins that get no data by the copying can be filled by applying an interpolation function (for example, using the interpolators 420,460).
- 25 9. Grains are transformed back to time domain (for example, using the inverse Fast Fourier Transformers 428,468).
10. Time domain grains are multiplied with a synthesis window (again Hann windows are proposed) (for example using the synthesis windowers 432,472).
11. If zero padding in step 3 was carried out, zeros are stripped again (for example, using the zero stripper 476).
- 30 12. Bandwidth extended signal or frame (for example, signal 392), respectively, is created using overlap and add (OLA) (for example, using overlap-and-add 390).

However, the order of the individual steps can also be exchanged in some alternative embodiments, and some of the steps can be merged into a single step in some alternative

35 embodiments.

The inventive algorithm, which is implemented in the audio decoder shown in Fig. 3, comprises the following steps:

1. Signal downsampling (if Nyquist criterion is not harmed)
2. Signal is windowed ("Hann" windows are proposed but other window shapes may be used) and so called grains (for example, windowed signal portions 383) of lengths  $N$  are taken from the signal. The windows are shifted over the signal with a hop size  $H$ . A  $N/H=8$  times overlap is proposed.
3. If the grain (for example, a windowed signal portion 383) contains a transient event at the edges, it is padded (for example, by the zero padder 438) with zeros which leads to an oversampling in frequency domain.
4. Grains are transformed to frequency domain (for example, using the time-domain-to-spectral-domain transformers 400,440).
5. Frequency domain grains are (optionally) padded to a desired output length of the patching algorithm.
6. Magnitude and phase are calculated (for example, using the means 402, 404, 442, 444).
7. a) Frequency bin content  $n$  is copied to position  $2n$ .  
The phase is multiplied with the 2.  
(a)  $\zeta \cdot (s-1)/s \leq n \leq \zeta$  or (b)  $\zeta/s \leq n \leq \zeta$  (see above).
7. b) Frequency bin content  $2n$  is copied to position  $sn$  for all stretching factors  $s > 2$  in the ranges  $1 \leq n \leq \zeta$ .
8. Frequency domain bins that get no data by the copying can be filled by applying an interpolation function (for example, using the interpolators 420,460).
9. Grains are transformed back to time domain (for example, using the inverse Fast Fourier Transformers 428,468).
10. Time domain grains are multiplied with a synthesis window (again Hann windows are proposed) (for example using the synthesis windowers 432,472).
11. If zero padding in step 3 was carried out, zeros are stripped again (for example, using the zero stripper 476).
12. Bandwidth extended signal or frame (for example, signal 392), respectively, is created using overlap and add (OLA) (for example, using overlap-and-add 390).

However, the order of the individual steps can also be exchanged in some alternative embodiments, and some of the steps can be merged into a single step in some alternative embodiments.

Thus, all steps are identical in the reference algorithm (which is implemented in the audio decoder shown in Fig. 6) and the inventive algorithm (which is implemented in the audio

decoder shown in Fig. 3), except for step 7, which has been replaced by the following steps:

7.a) Frequency bin content  $n$  is copied to position  $2n$ . The phase is multiplied with the 2.

5 (a)  $\zeta \cdot (s-1)/s \leq n \leq \zeta$  or (b)  $\zeta/s \leq n \leq \zeta$  (see above).

7.b) Frequency bin content  $2n$  is copied to position  $sn$  for all stretching factors  $s > 2$  in the ranges  $1 \leq n \leq \zeta$ .

10 To summarize, the embodiments according to Figs. 1, 2, 3 and 4 (and also the audio decoder shown in Fig. 6) firstly reduce complexity dramatically when compared to the mentioned conventional solutions. Secondly, they allow for different spectrum modifications different to either plane SBR or as presented in Fig. 5 (see, for example, Reference [13]).

15 For example, speech signals might benefit from the algorithm, which is performed by the apparatus, audio decoder and method according to Figs. 1, 2, 3 and 4, as the pulse train structure, which is typical for speech signals, is better maintained than with the approach presented in Reference [13].

20 Most prominent applications of embodiments according to the invention are audio decoders, which are often implemented on hand-held devices and thus operate on a battery power supply.

#### 25 4. Method according to Fig. 4.

In the following, a method 400 for generating a representation of a bandwidth-extended signal on the basis of an input signal representation will be described taking reference to  
 30 Fig. 4, which shows a flow chart of such a method. The method 400 comprises a step 410 of obtaining values of a spectral domain representation of a first patch of the bandwidth-extended signal on the basis of the input signal representation using a phase vocoding. The method 400 also comprises a step 420 of copying a set of values of the spectral domain representation of the first patch, which values are obtained using the phase vocoding, to  
 35 obtain a set of values of a spectral domain representation of a second patch, wherein the second patch is associated with higher frequencies than the first patch. The method 400 also comprises a step 430 of obtaining a representation of the bandwidth-extended signal

using the values of the spectral domain representation of the first patch and the values of the spectral domain representation of the second patch.

5 The method 400 can be supplemented by any of the means and functionalities discussed here with respect to the inventive apparatus.

#### 5. Implementation alternatives

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. Some or all of the method steps may be executed by (or using) a hardware apparatus, like for example, a microprocessor, a  
15 programmable computer or an electronic circuit. In some embodiments, some one or more of the most important method steps may be executed by such an apparatus.

Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a  
20 digital storage medium, for example a floppy disk, a DVD, a Blue-Ray, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed. Therefore, the digital storage medium may be computer readable.

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.

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.

#### 35 6. Comparison example according to Fig. 5

In the following, a comparison example will be briefly discussed taking reference to Fig. 5. The functionality of the comparison example according to Fig. 5 is similar to the function of the audio decoder according to Fig. 3, such that the means and functionalities will not be

explained again. However, the comparison example according to Fig. 5 relies on the usage of three phase vocoders 590, 592, 594, or 596, 597, 598 per branch. Individual inverse Fast Fourier Transformers, synthesis windowers, overlappers-and-adders are associated to the individual phase vocoders, as can be seen in Fig. 5. Also, in some of the sub-branches,  
 5 individual down-sampling ( $\downarrow$  factor) and individual delay ( $z^{-\text{samples}}$ ) is used. Accordingly, the apparatus 500 according to Fig. 5 is not as computationally efficient as the apparatus 300 according to Fig. 3. Nevertheless, the apparatus 500 brings along significant improvements over some conventional audio decoders.

## 10 7. Comparison example according to Fig. 6

Fig. 6 shows another audio decoder 600, according to a comparison example. The audio decoder 600 according to Fig. 6 is similar to the audio decoders 300, 500 according to Figs. 3 and 5. However, the audio decoder 600 is also based on the usage of a plurality of  
 15 individual phase vocoders 690, 692, 694 or 696, 697, 698 per branch, which renders the apparatus 600 computationally more demanding than the apparatus 300, and which brings along audible artifacts in some cases. Nevertheless, the apparatus 500 brings along significant improvements over some conventional audio decoders.

## 20 8. Conclusion

In view of the above discussion, it can be seen that the apparatus 100 according to Fig. 1, the audio decoder 300 according to Fig. 3 and the method 400 according to Fig. 4 bring along a number of advantages over the comparison examples, which have been briefly  
 25 discussed with reference to Figs. 5 and 6.

The inventive concept is applicable in a wide variety of applications and can be modified in a wide number of ways. In particular, the Fast Fourier Transformers can be replaced by QMF filterbanks, and the inverse Fast Fourier Transformers can be replaced by QMF  
 30 synthesizers.

Also, in some embodiments some or all of the processing steps can be summarized into a single step. For example, a processing sequence comprising a QMF synthesis and a subsequent QMF Analysis may be simplified by omitting the repeated transforms.

## References:

- 5 [1] M. Dietz, L. Liljeryd, K. Kjörling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in 112th AES Convention, Munich, May 2002.
- [2] S. Meltzer, R. Böhm and F. Henn, "SBR enhanced audio codecs for digital broadcasting such as "Digital Radio Mondiale" (DRM)," in 112th AES Convention, Munich, May 2002.
- 10 [3] T. Ziegler, A. Ehret, P. Ekstrand and M. Lutzky, "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm," in 112th AES Convention, Munich, May 2002.
- [4] International Standard ISO/IEC 14496-3:2001/FPDAM 1, "Bandwidth Extension,"  
15 ISO/IEC, 2002. Speech bandwidth extension method and apparatus Vasu Iyengar et al.
- [5] E. Larsen, R. M. Aarts, and M. Danessis. Efficient high-frequency bandwidth extension of music and speech. In AES 112th Convention, Munich, Germany, May 2002.
- 20 [6] R. M. Aarts, E. Larsen, and O. Ouweltjes. A unified approach to low- and highfrequency bandwidth extension. In AES 115th Convention, New York, USA, October 2003.
- [7] K. Käyhkö. A Robust Wideband Enhancement for Narrowband Speech Signal.  
25 Research Report, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 2001.
- [8] E. Larsen and R. M. Aarts. Audio Bandwidth Extension - Application to psychoacoustics, Signal Processing and Loudspeaker Design. John Wiley & Sons, Ltd,  
30 2004.
- [9] E. Larsen, R. M. Aarts, and M. Danessis. Efficient high-frequency bandwidth extension of music and speech. In AES 112th Convention, Munich, Germany, May 2002.
- 35 [10] J. Makhoul. Spectral Analysis of Speech by Linear Prediction. IEEE Transactions on Audio and Electroacoustics, AU-21(3), June 1973.

- [11] United States Patent Application 08/951,029, Ohmori , et al. Audio band width extending system and method.
- [12] United States Patent 6895375, Malah, D & Cox, R. V.: System for bandwidth  
5 extension of Narrow-band speech.
- [13] Frederik Nagel, Sascha Disch, "A harmonic bandwidth extension method for audio codecs," ICASSP International Conference on Acoustics, Speech and Signal Processing, IEEE CNF, Taipei, Taiwan, April 2009.

## Claims

1. An apparatus (100;386) for generating a representation (120; 426) of a bandwidth-extended signal on the basis of an input signal representation (110; 383), the apparatus comprising:

a phase vocoder (130; 406) configured to obtain values ( $\beta_{\zeta} \dots \beta_{2\zeta}$ , 408) of a spectral domain representation of a first patch of the bandwidth-extended signal on the basis of the input signal representation; and

a value copier (140; 410,416) configured to copy a set of values ( $\beta_{\zeta} \dots \beta_{2\zeta}$ , 408) of the spectral domain representation of the first patch, which values are provided by the phase vocoder, to obtain a set of values ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ , 408) of a spectral domain representation of a second patch, wherein the second patch is associated with higher frequencies than the first patch;

wherein the apparatus is configured to obtain the representation (120;426) of the bandwidth-extended signal using the values of the spectral domain representation of the first patch and the values of the spectral domain representation of the second patch.

2. The apparatus (100;386) according to claim 1, wherein the phase vocoder (130; 406) is configured to copy a set of magnitude values ( $\alpha_{\zeta/2} \dots \alpha_{\zeta}$ ) associated with a plurality of given frequency subranges of the input signal representation (110; 383), to obtain a set of magnitude values ( $\alpha_{\zeta} \dots \alpha_{2\zeta}$ ) associated with corresponding frequency subranges of the first patch,

wherein a pair of a given frequency subrange of the input signal representation and of a corresponding frequency subrange of the first patch cover a pair of a fundamental frequency and a harmonic of the fundamental frequency,

wherein the phase vocoder (130; 406) is configured to multiply phase values ( $\varphi_{\zeta/2} \dots \varphi_{\zeta}$ ) associated with the plurality of given frequency subranges of the input signal representation with a predetermined factor, to obtain a set of phase values ( $\varphi_{\zeta} \dots \varphi_{2\zeta}$ ) associated with the corresponding frequency subranges of the first patch, and

wherein the value copier (140; 410) is configured to copy a set of values ( $\beta_{\zeta} \dots \beta_{2\zeta}$ ) associated with a plurality of given frequency subranges of the first patch, to obtain

a set of values ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ ) associated with corresponding frequency subranges of the second patch, wherein the value copier is configured to leave phase values unchanged in the copying.

5     3.     The apparatus (100;386) according to claim 2, wherein the value copier (140; 410) is configured to copy the values such that a common spectral shift between values ( $\beta_{\zeta} \dots \beta_{2\zeta}$ ) of the first patch and corresponding values ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ ) of the second patch is obtained.

10     4.     The apparatus (100;386) according to one of claims 1 to 3, wherein the phase vocoder (130; 410) is configured to obtain the values ( $\beta_{\zeta} \dots \beta_{2\zeta}$ ) of the spectral domain representation (132; 408) of the first patch such that the values of the spectral domain representation of the first patch represent a harmonically up-converted version of a fundamental frequency range of the input signal representation (110; 383); and

15

wherein the value copier (140;410) is configured to obtain the values ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ ) of the spectral domain representation (142; 412) of the second patch such that the values of the spectral domain representation of the second patch represent a frequency-shifted version of the audio content of the first patch.

20

5.     The apparatus (100; 380,382, 386) according to one of claims 1 to 4, wherein the apparatus is configured to receive input audio data (322),

25     to down-sample (380) the input audio data (322), in order to obtain down-sampled audio data (381),

to window (382) the down-sampled audio data (381), in order to obtain windowed input data (383),

30

to convert (400) or transform the windowed input data (383) into a spectral domain, in order to obtain the input signal representation (383) in the form of a spectral domain representation (401),

35     to compute (402, 404) magnitude values  $\alpha_k$  and phase values  $\phi_k$  representing a frequency bin having index  $k$  of the input signal representation (383),

to use (130; 406) a plurality of magnitude values  $\alpha_k$  representing frequency bins having frequency bin indices  $k$  of the input signal representation (383), to obtain magnitude values  $\alpha_{2k}$  representing frequency bins having frequency bin indices  $sk$  of the first patch,

5

when  $s$  is a stretching factor with  $s$  between 1.5 and 2.5, and

10

to copy and scale (130; 406) phase values  $\phi_k$  associated to frequency bins having frequency bin indices  $k$  of the input signal representation (383), to obtain copied and scaled phase values  $\phi_{2k}=s\phi_k$  associated with frequency bins having frequency bin indices  $2k$  of the first patch,

15

to copy (140; 410) values  $\beta_{k-i\zeta}$  associated with frequency bins having frequency bin indices  $k-i\zeta$  of the spectral domain representation (132;408) of the first patch, to obtain values  $\beta_k$  of the spectral domain representation (142; 412) of the second patch,

20

to convert (428) the representation (426) of the bandwidth-extended signal into the time-domain, to obtain a time-domain representation (430), and

to apply (432) a synthesis window to the time-domain representation.

25

6. The apparatus (100; 386) according to one of claims 1 to 5, wherein the apparatus comprises a time-domain to spectral-domain converter (400) configured to provide, as the input signal representation (401), values of a spectral-domain representation of an input audio signal (322), or of a pre-processed version (383) of the input audio signal (322); and

30

wherein the apparatus comprises a spectral-domain-to-time-domain converter (428) configured to provide a time-domain representation (430) of the bandwidth-extended signal using values ( $\beta_\zeta \dots \beta_{2\zeta}$ , 408) of the spectral-domain representation of the first patch and values ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ , 412) of the spectral-domain representation of the second patch;

35

wherein the spectral-domain-to-time-domain converter (428) is configured such that a number ( $N=2048$ ) of different spectral values (426) received by the spectral-domain-to-time-domain converter (428) is larger than a number ( $N=512$ ) of different spectral values (401) provided by the time-domain-to-spectral-domain

converter (400), such that the spectral-domain-to-time-domain converter (428) is configured to process a larger number of frequency bins than the time-domain-to-spectral-domain converter (400).

- 5     7.     The apparatus (100; 382, 386) according to one of claims 1 to 6, wherein the apparatus comprises an analysis windower (382) configured to window a time-domain input audio signal (322), to obtain a windowed version (383) of the time-domain input audio signal, which forms the basis for obtaining the input signal representation in the form of a spectral domain representation (401); and

10

wherein the apparatus comprises a synthesis windower (432) configured to window a portion of a time-domain representation (430) of the bandwidth-extended signal, to obtain a windowed portion (434) of the time-domain representation of the bandwidth-extended signal.

15

8.     The apparatus (100; 382,386) according to claim 7, wherein the apparatus is configured to process a plurality of temporally overlapping time-shifted portions of the time-domain input audio signal (322), to obtain a plurality of temporally overlapping time-shifted windowed portions (434) of the time-domain representation of the bandwidth-extended signal,

20

wherein a time offset ( $\text{Inc}=64$ ) between temporally adjacent time-shifted portions of the time-domain input audio signal (322) is smaller than or equal to one fourth of a window length (512) of the analysis windower (382).

25

9.     The apparatus (100; 382,386) according to one claims 1 to 8, wherein the apparatus comprises a transient information provider (384) configured to provide an information indicating the presence of a transient in the input signal (322); and

30

wherein the apparatus comprises a first processing branch (386) for providing a representation (434) of a bandwidth-extended signal portion on the basis of a non-transient portion of the input signal representation (383) and a second processing branch (388) for providing a representation (478) of a bandwidth-extended signal portion on the basis of a transient portion of the input signal representation (383);

35

wherein the second processing branch (388) is configured to process a spectral-domain representation (441) of the input signal having a higher spectral resolution

( $N=1024$ ) than a spectral-domain representation (401) of the input signal processed by the first processing branch (386).

5 10. The apparatus (100; 382,386) according to claim 9, wherein the second processing branch (388) comprises a time-domain zero-padder (438) configured to zero-pad a transient-containing portion (383) of the input signal, in order to obtain a temporally extended transient-containing portion (439) of the input signal; and

10 wherein the first processing branch (386) comprises a time-domain-to-frequency-domain converter (400) configured to provide a first number ( $N=512$ ) of spectral-domain values (401) associated with the non-transient portion (383) of the input signal; and

15 wherein the second processing branch (388) comprises a time-domain-to-frequency-domain converter (440) configured to provide a second number ( $N=1024$ ) of spectral-domain values (441) associated with the temporally extended transient-containing portion (439) of the input signal,

20 wherein the second number ( $N=1024$ ) of spectral domain values is larger, at least by a factor of 1.5, than the first number ( $N=512$ ) of spectral-domain values.

25 11. The apparatus (100; 382,386) according to claim 10, wherein the second processing branch comprises a zero stripper (476) configured to remove a plurality of zero values from a bandwidth-extended signal portion (474) obtained on the basis of the temporally extended transient-containing portion (439) of the input signal.

30 12. The apparatus (100; 380, 382,386) according to one of claims 1 to 11, wherein the apparatus comprises a down-sampler (380) configured to down-sample a time-domain representation (322) of the input signal.

13. An audio decoder (300) comprising an apparatus (100; 386) according to one of claims 1 to 12.

35 14. A method (400) for generating a representation of a bandwidth-extended signal on the basis of an input signal representation, the method comprising:

Obtaining (410), using a phase vocoding, values of a spectral-domain representation of a first patch of the bandwidth-extended signal on the basis of the input signal representation; and

5 copying (420) a set of values of the spectral-domain representation of the first patch, which values are provided by the phase vocoding, to obtain a set of values of a spectral-domain representation of a second patch, wherein the second patch is associated with higher frequencies than the first patch; and

10 obtaining (430) the representation of the bandwidth-extended signal using the values of the spectral-domain representation of the first patch and the values of the spectral-domain representation of the second patch.

15 15. An apparatus (100;386) for generating a representation (120; 426) of a bandwidth-extended signal on the basis of an input signal representation (110; 383), the apparatus comprising:

20 a value copier configured to copy a set of values ( $\beta_1 \dots \beta_\zeta$ ) of the input signal representation, to obtain a set of values ( $\beta_\zeta \dots \beta_{2\zeta}$ ) of a spectral domain representation of a first patch, wherein the first patch is associated with higher frequencies than the input signal representation; and

25 a phase vocoder (130; 406) configured to obtain values ( $\beta_{2\zeta} \dots \beta_{3\zeta}$ ) of a spectral domain representation of a second patch of the bandwidth-extended signal on the basis of the values ( $\beta_{4/3\zeta} \dots \beta_{2\zeta}$ ) of the spectral domain representation of the first patch, wherein the second patch is associated with higher frequencies than the first patch; and

30 wherein the apparatus is configured to obtain the representation (120;426) of the bandwidth-extended signal using the values of the spectral domain representation of the first patch and the values of the spectral domain representation of the second patch.

35 16. A method (400) for generating a representation of a bandwidth-extended signal on the basis of an input signal representation, the method comprising:

copying values of the input signal representation, to obtain values of a spectral-domain representation of a first patch of the bandwidth-extended signal on the basis

of the input signal representation, wherein the first patch is associated with higher frequencies than the input signal representation; and

5 obtaining, using a phase vocoding, a set of values of the spectral-domain representation of the second patch on the basis of a set of values of the spectral-domain representation of the first patch, which values of the spectral domain representation of the first patch are obtained by the copying, wherein the second patch is associated with higher frequencies than the first patch; and

10 obtaining (430) the representation of the bandwidth-extended signal using the values of the spectral-domain representation of the first patch and the values of the spectral-domain representation of the second patch.

15 17. A computer program for performing the method according to claim 14 or claim 16, when the computer program runs on a computer.

**Apparatus, Method and Computer Program for Generating a Representation of a Bandwidth-Extended Signal on the Basis of an Input Signal Representation Using a Combination of a Harmonic Bandwidth-Extension and a Non-Harmonic Bandwidth-Extension**

5

Abstract

10 An apparatus for generating a representation of a bandwidth-extended signal on the basis of an input signal representation comprises a phase vocoder configured to obtain values of a spectral domain representation of a first patch of the bandwidth-extended signal on the basis of the input signal representation. The apparatus also comprises a value copier configured to copy a set of values of the spectral domain representation of the first patch, which values are provided by the phase vocoder, to obtain a set of values of a spectral  
15 domain representation of a second patch, wherein the second patch is associated with higher frequencies than the first patch. The apparatus is configured to obtain the representation of the bandwidth-extended signal using the values of the spectral domain representation of the first patch and the values of the spectral domain representation of the second patch.

20

Figure 1

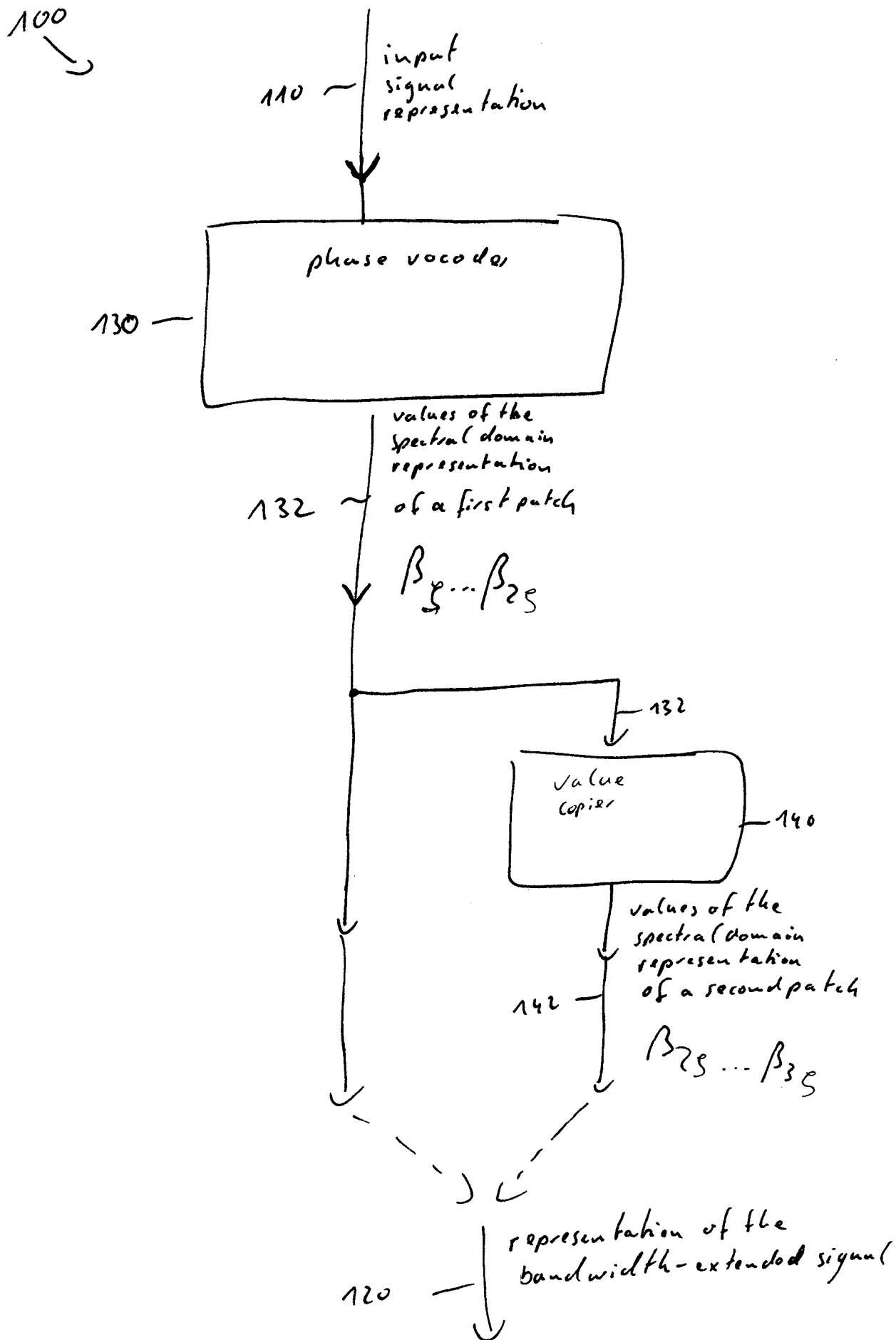
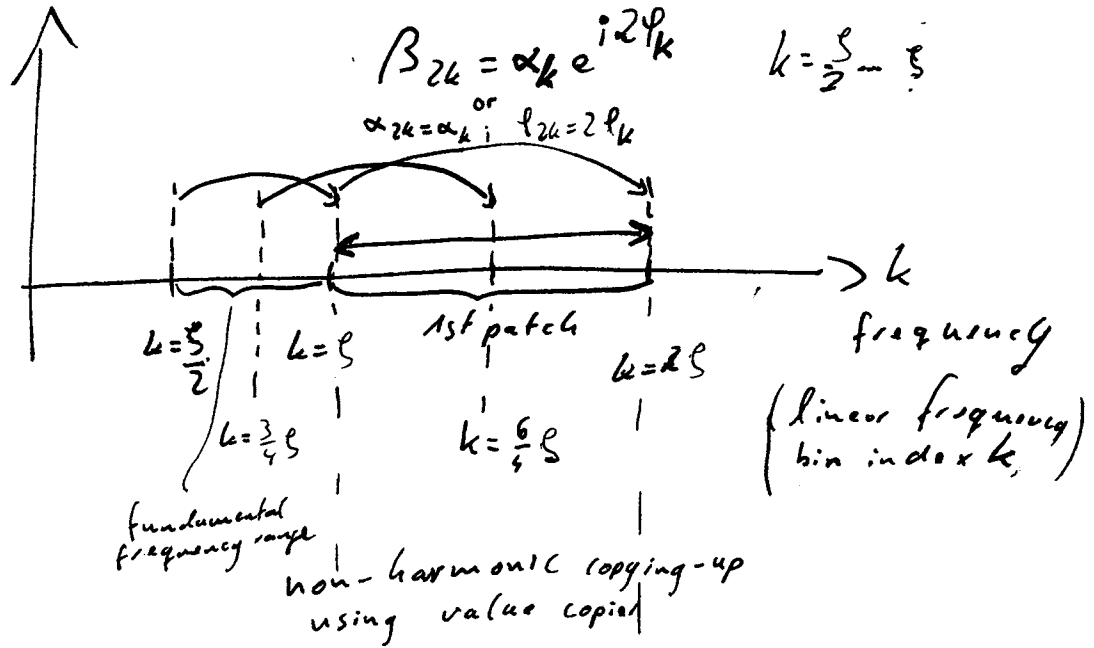


Fig. 1

harmonic transposition  
using (single) phase vocoder

200



250

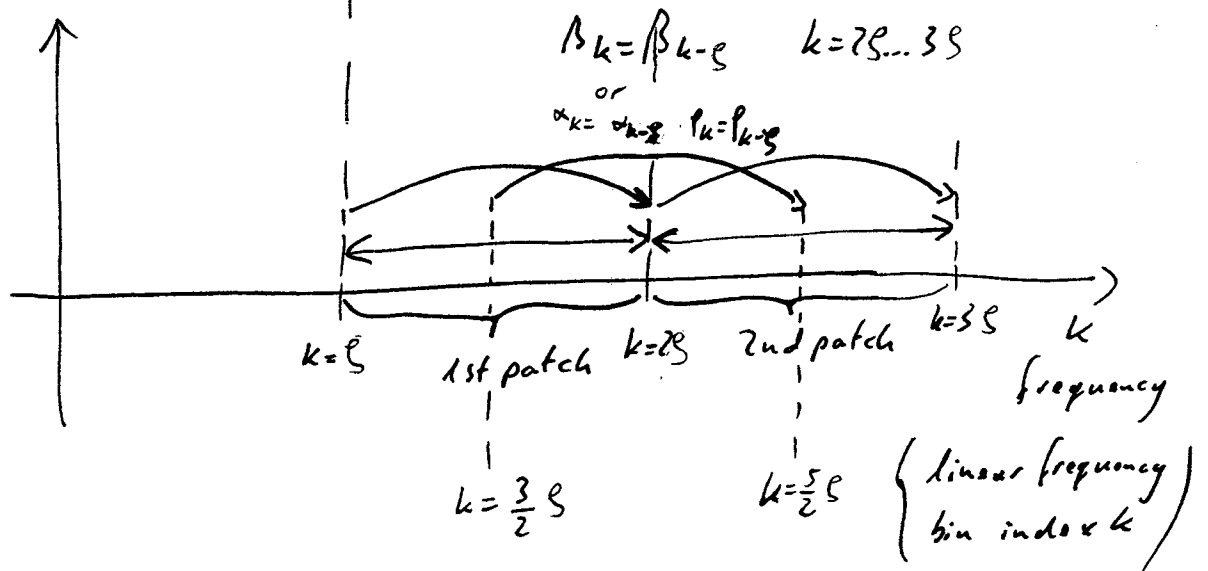


Fig. 2

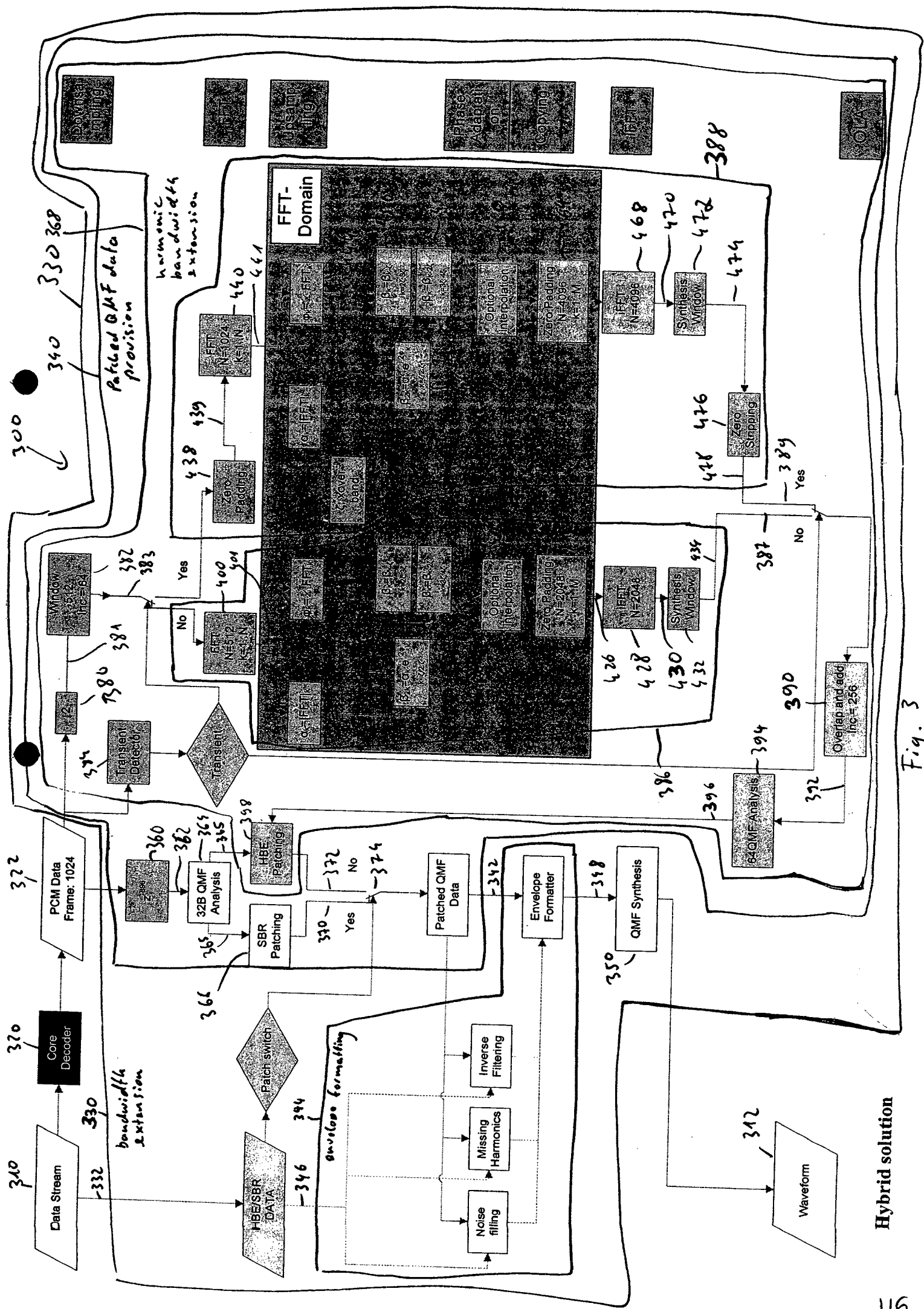


Fig. 3

400

410 ~

obtaining values of a spectral domain representation of a first patch of the bandwidth-extended signal on the basis of the input signal representation using a phase vocoding

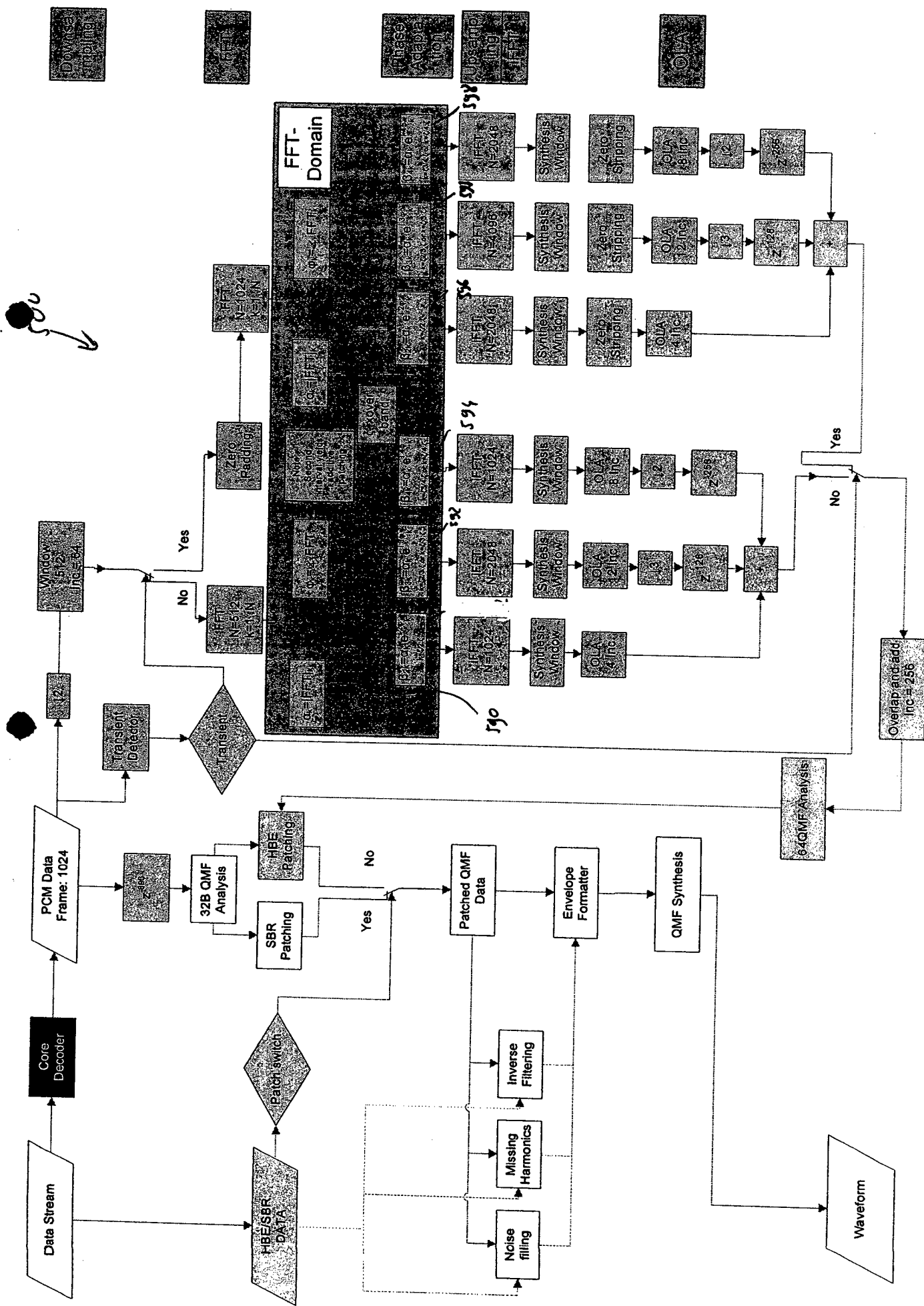
420 ~

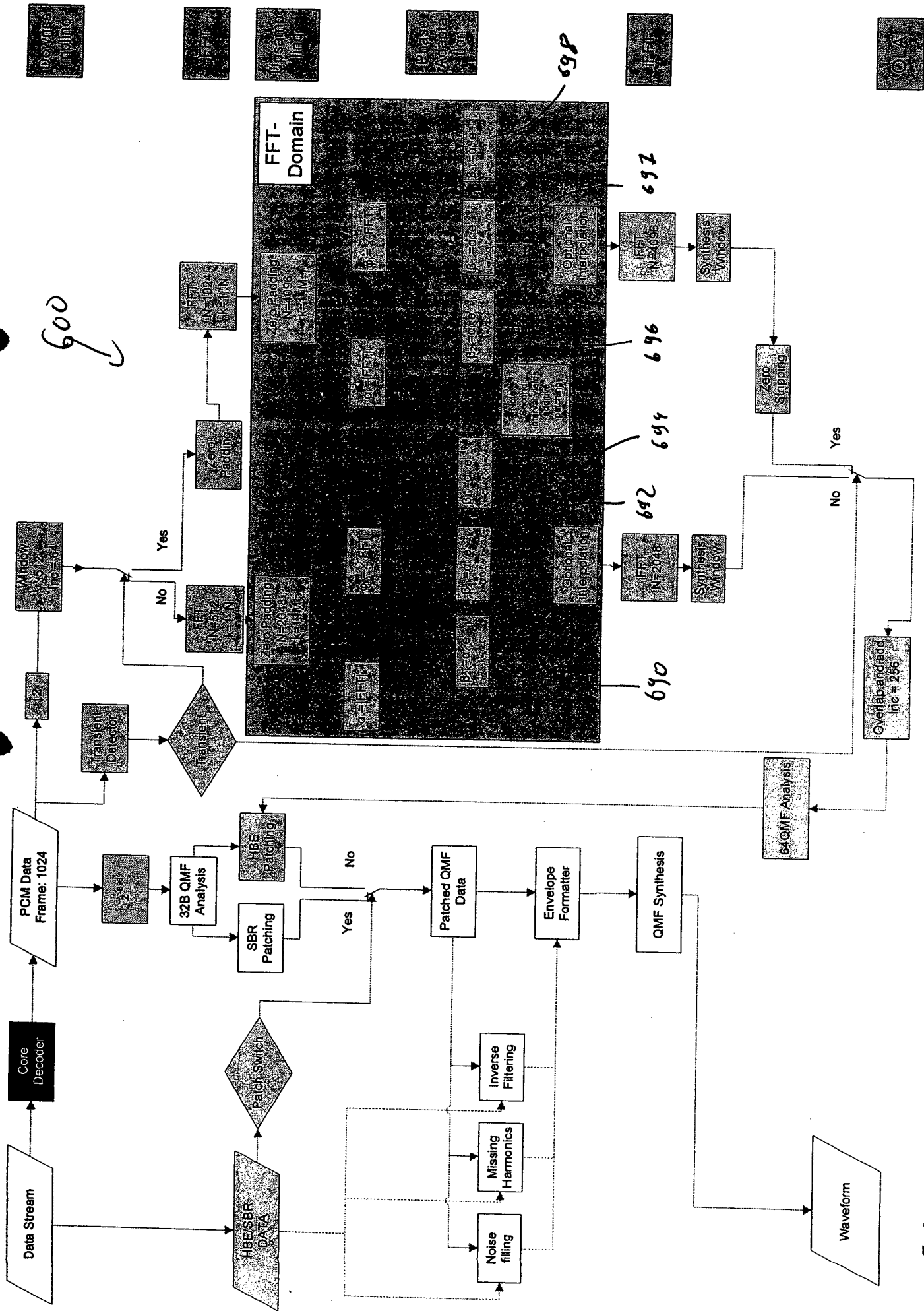
copying a set of values of the spectral domain representation of the first patch, which values are obtained using the phase vocoding, to obtain a set of values of a spectral domain representation of a 2nd patch, wherein the 2nd patch is associated with higher frequencies than the 1st patch

430 ~

obtaining a representation of the bandwidth-extended signal using the values of the spectral domain representation of the 1st patch and the values of the spectral domain representation of the 2nd patch

Fig. 4





Embodiment Algorithm (HBE in frequency domain)

Fig. 6

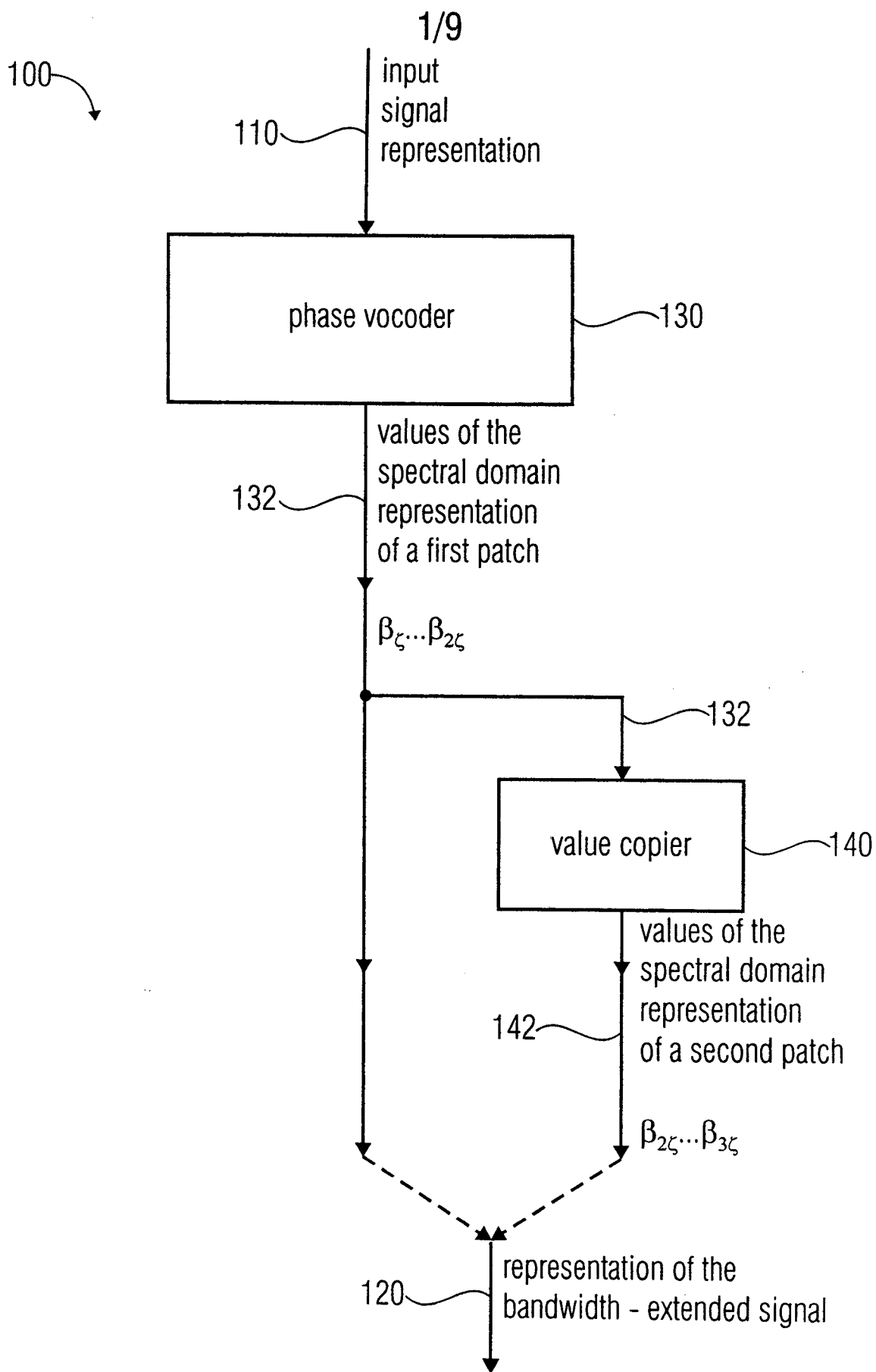


FIG 1

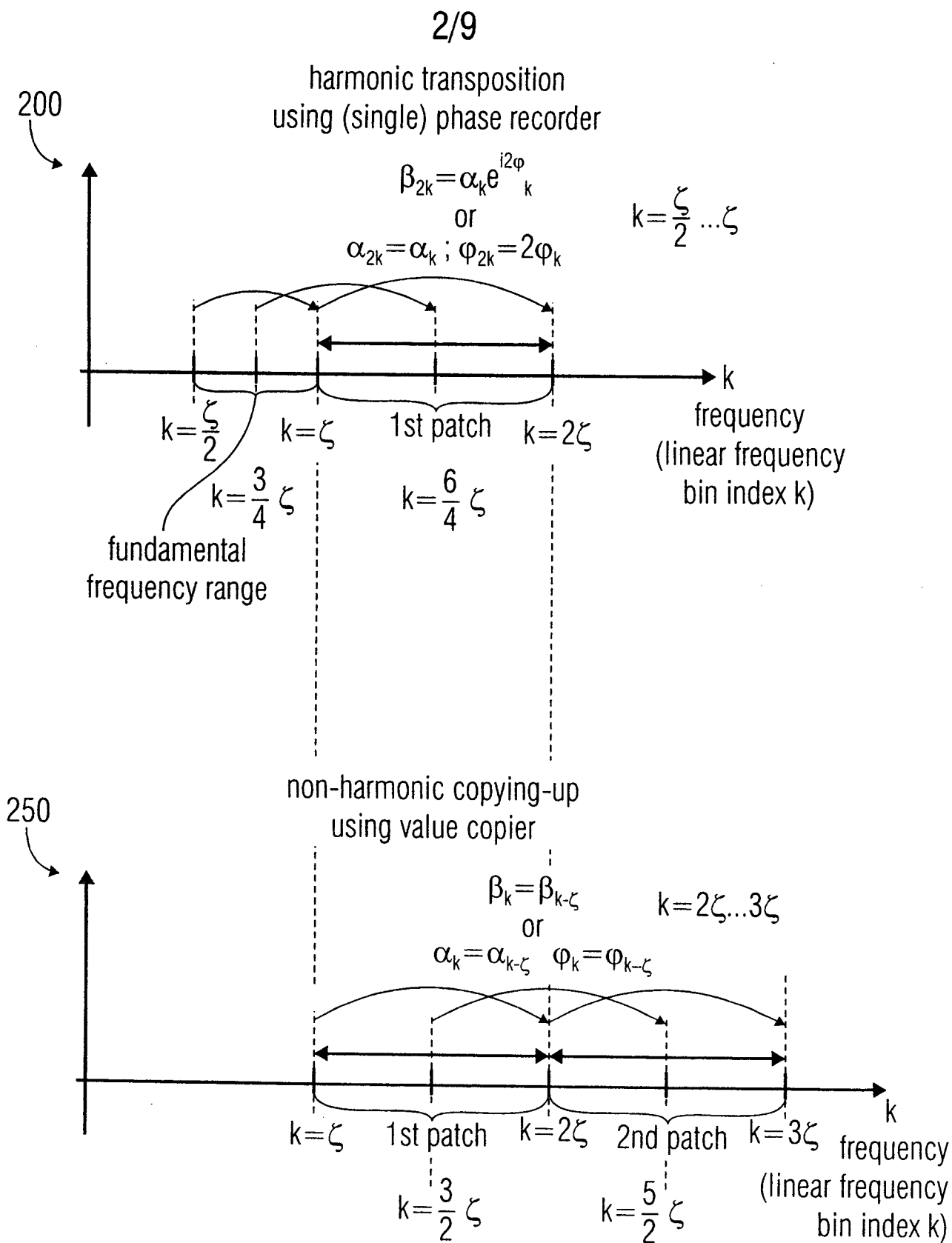
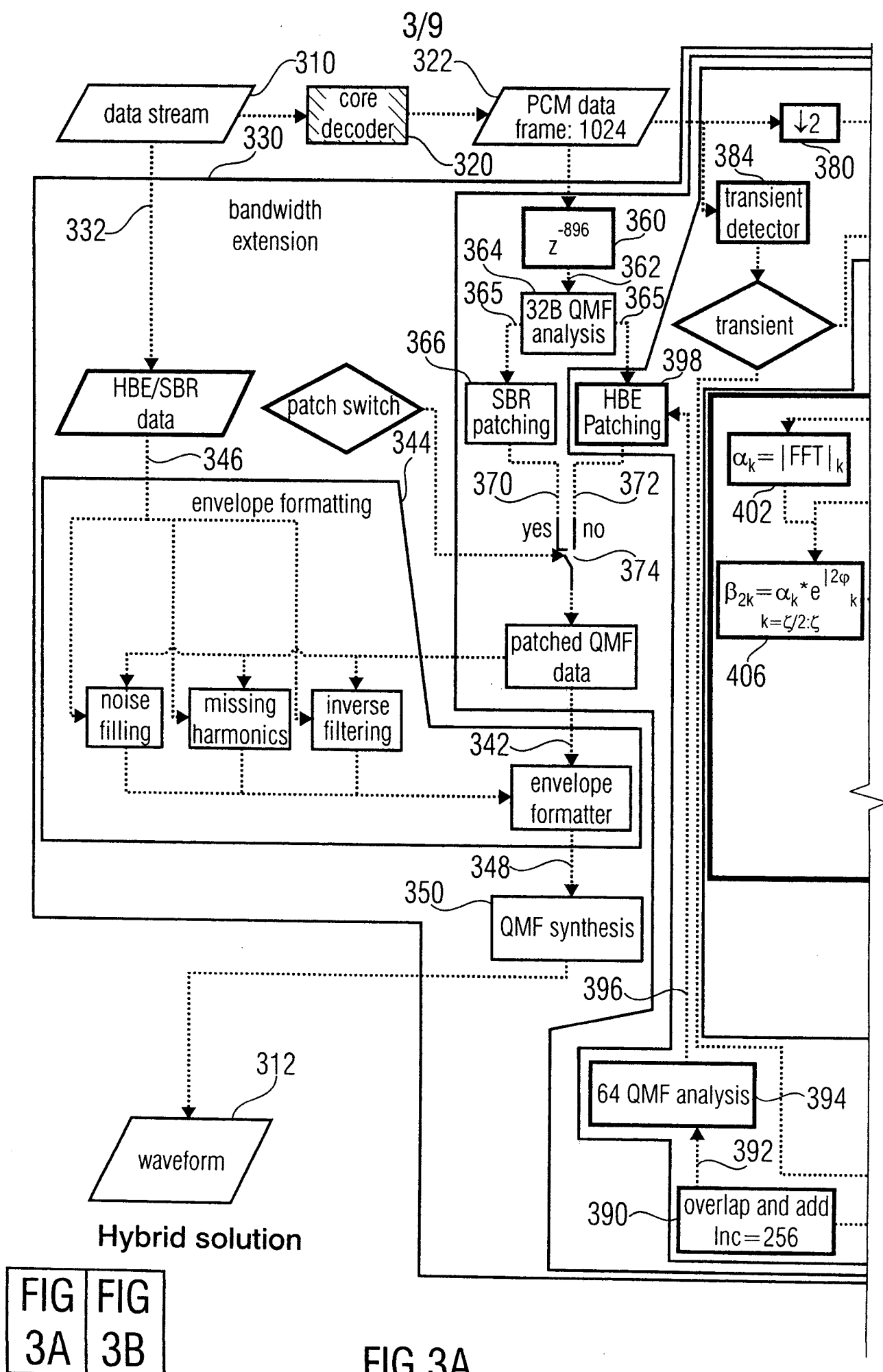


FIG 2



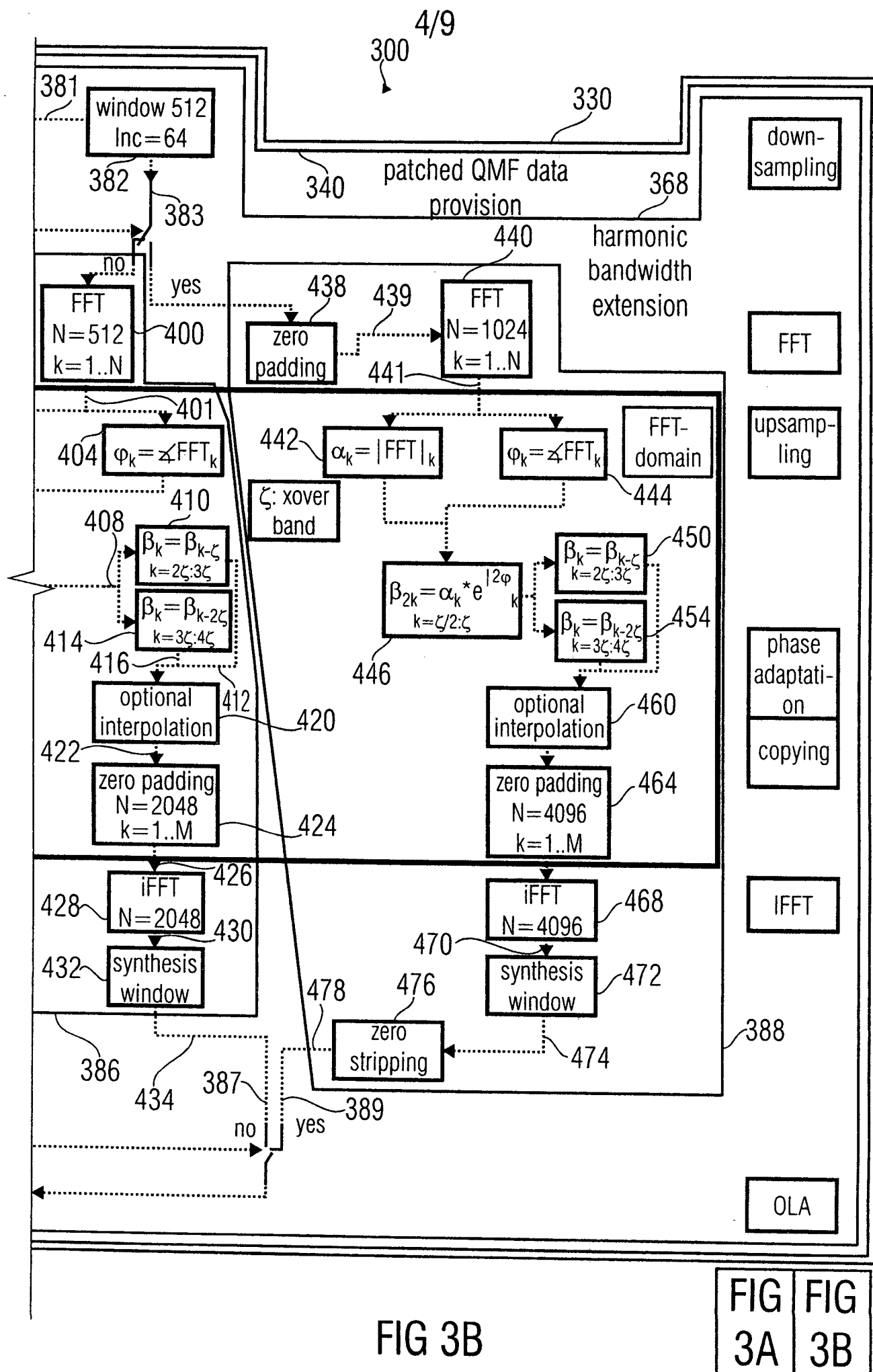


FIG 3B

FIG 3A

FIG 3B

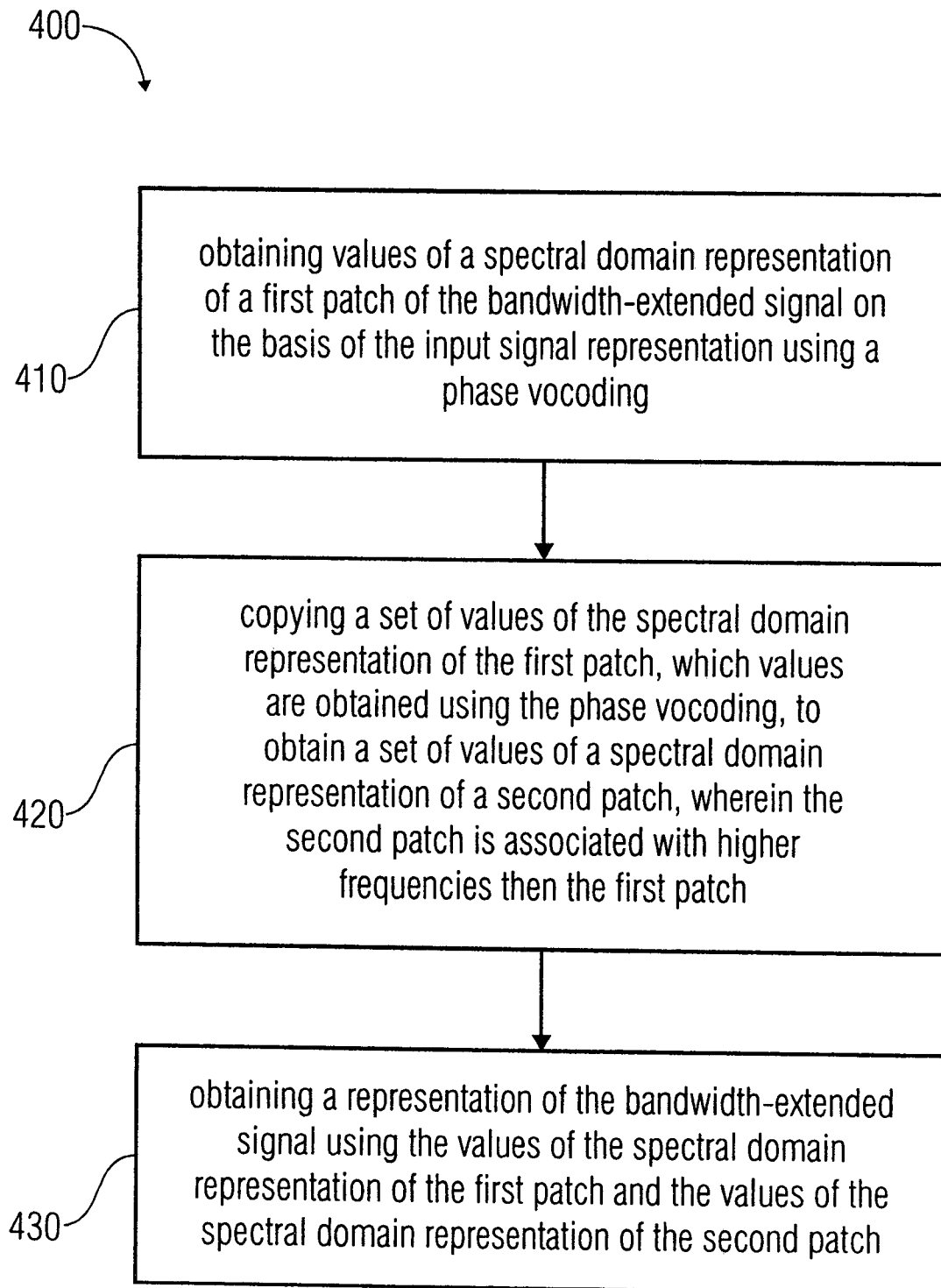


FIG 4

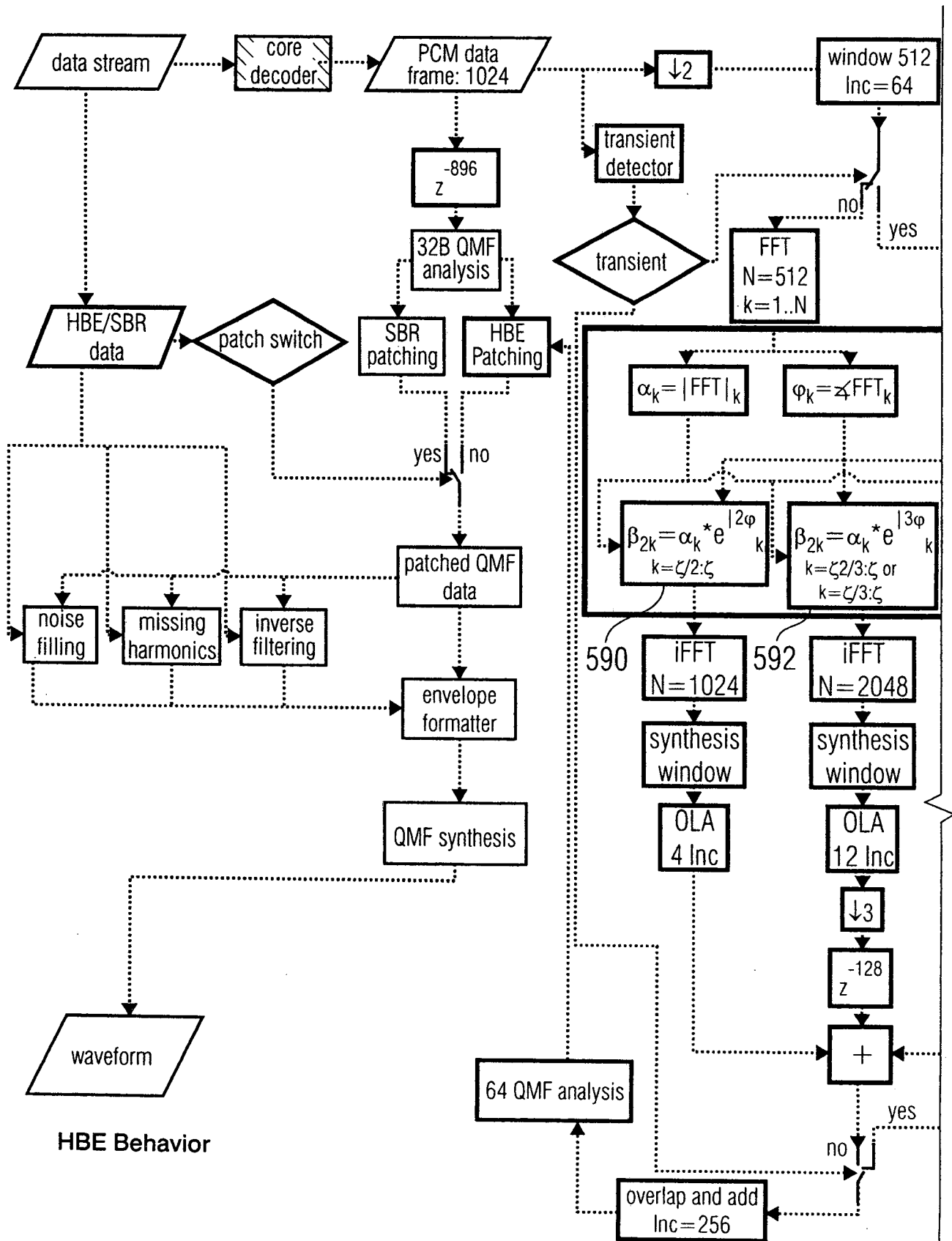


FIG 5A

|           |           |
|-----------|-----------|
| FIG<br>5A | FIG<br>5B |
|-----------|-----------|

7/9

500

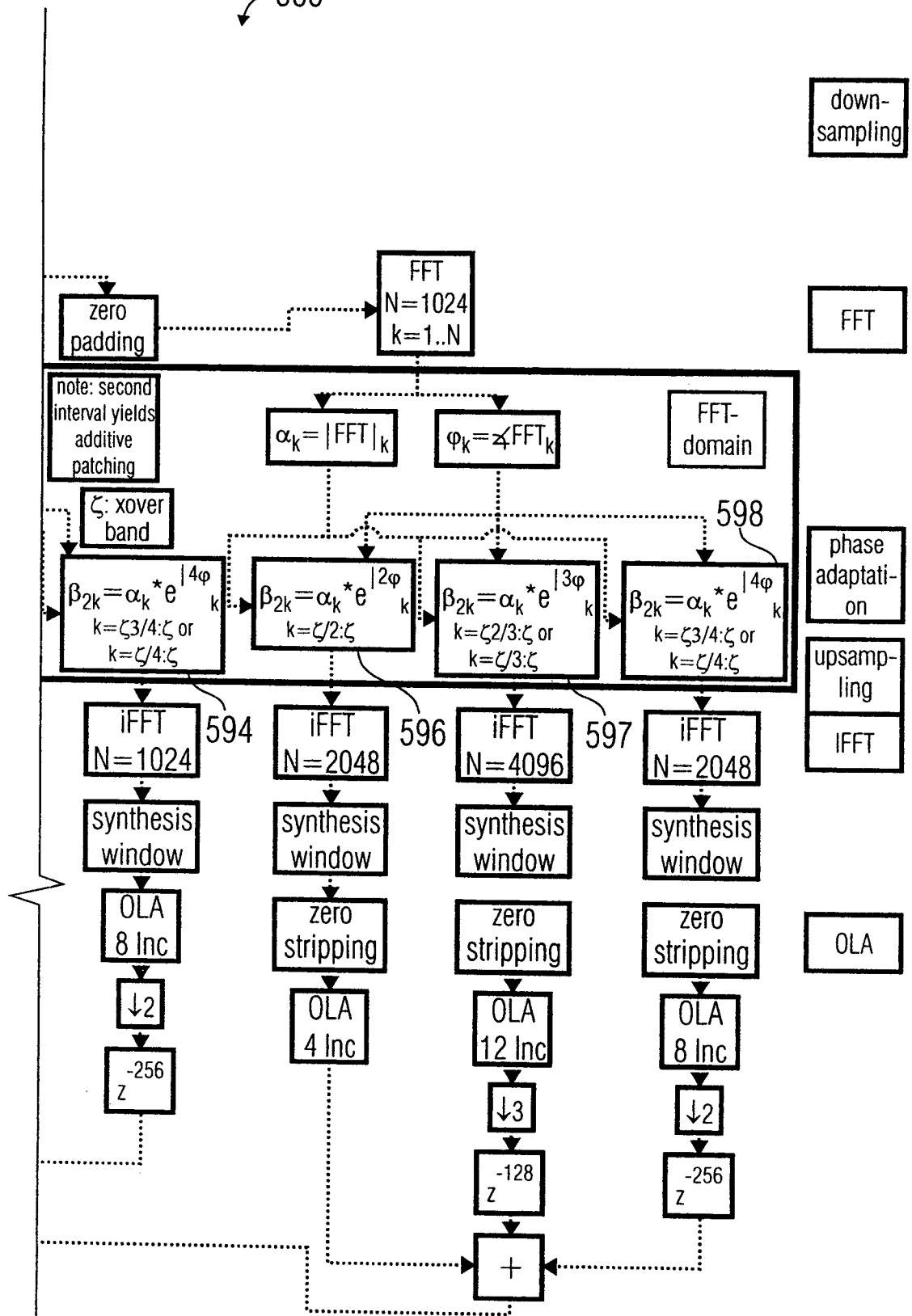
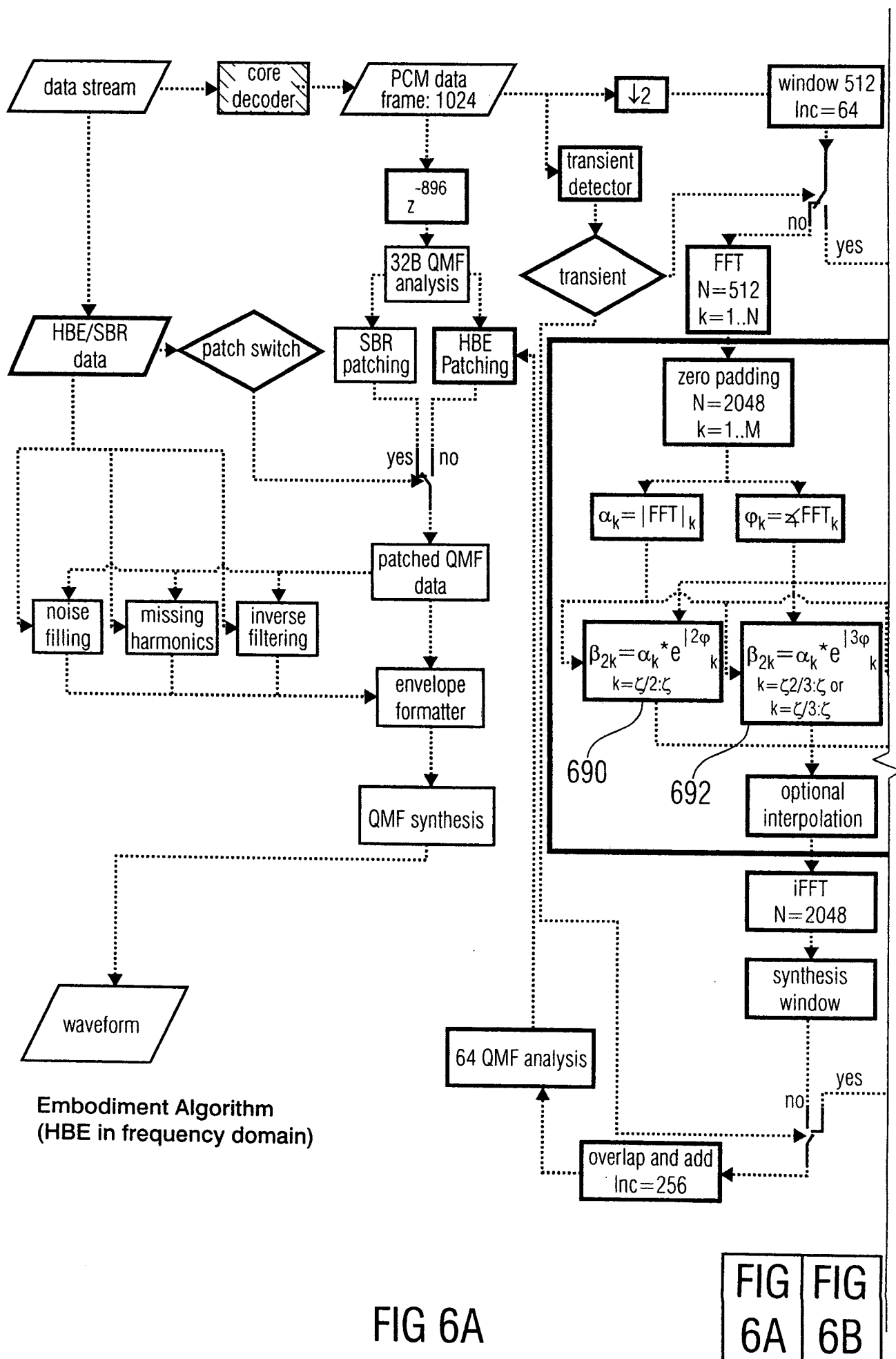


FIG 5B

FIG 5A

FIG 5B



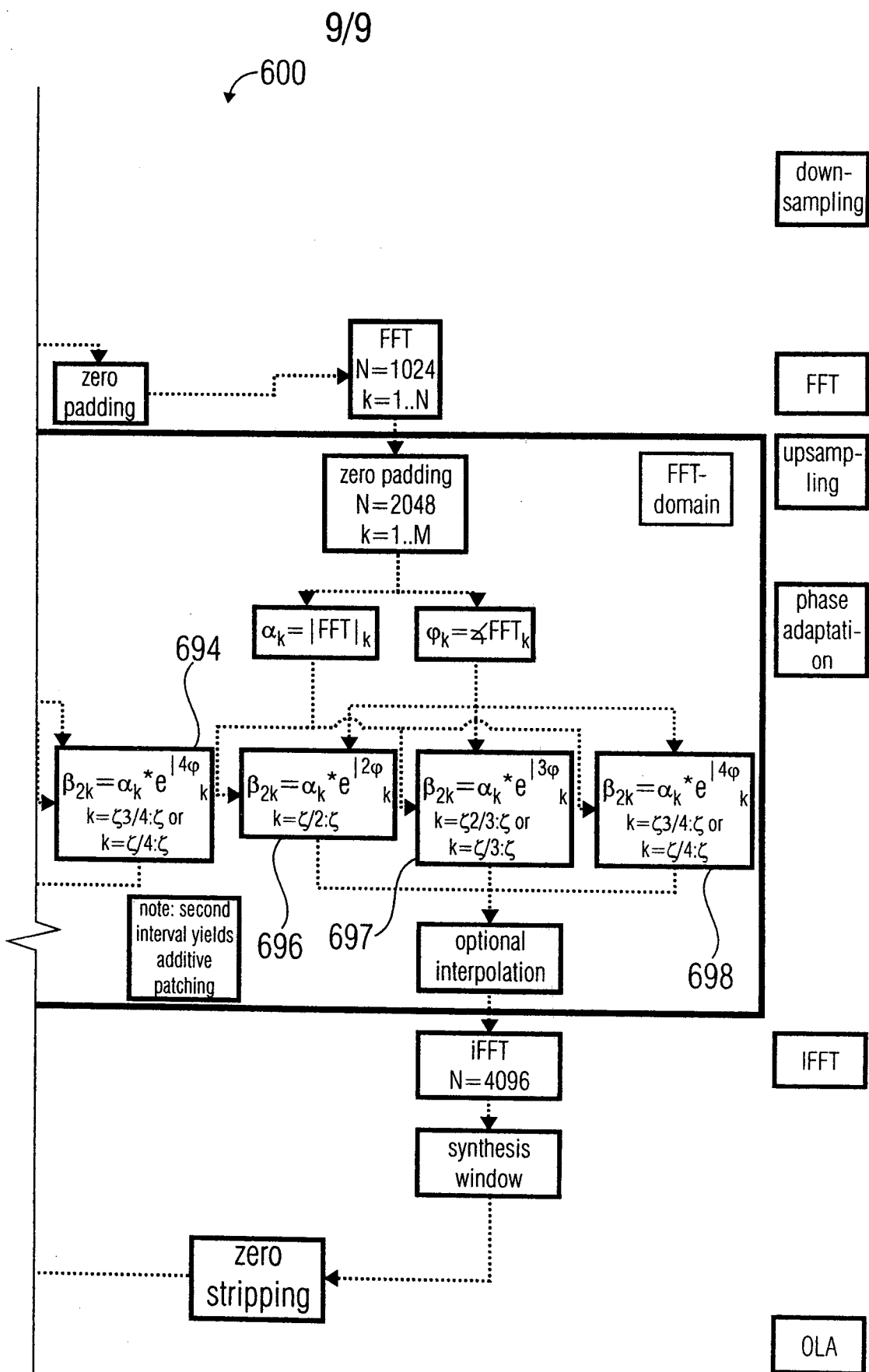


FIGURE FOR ABSTRACT

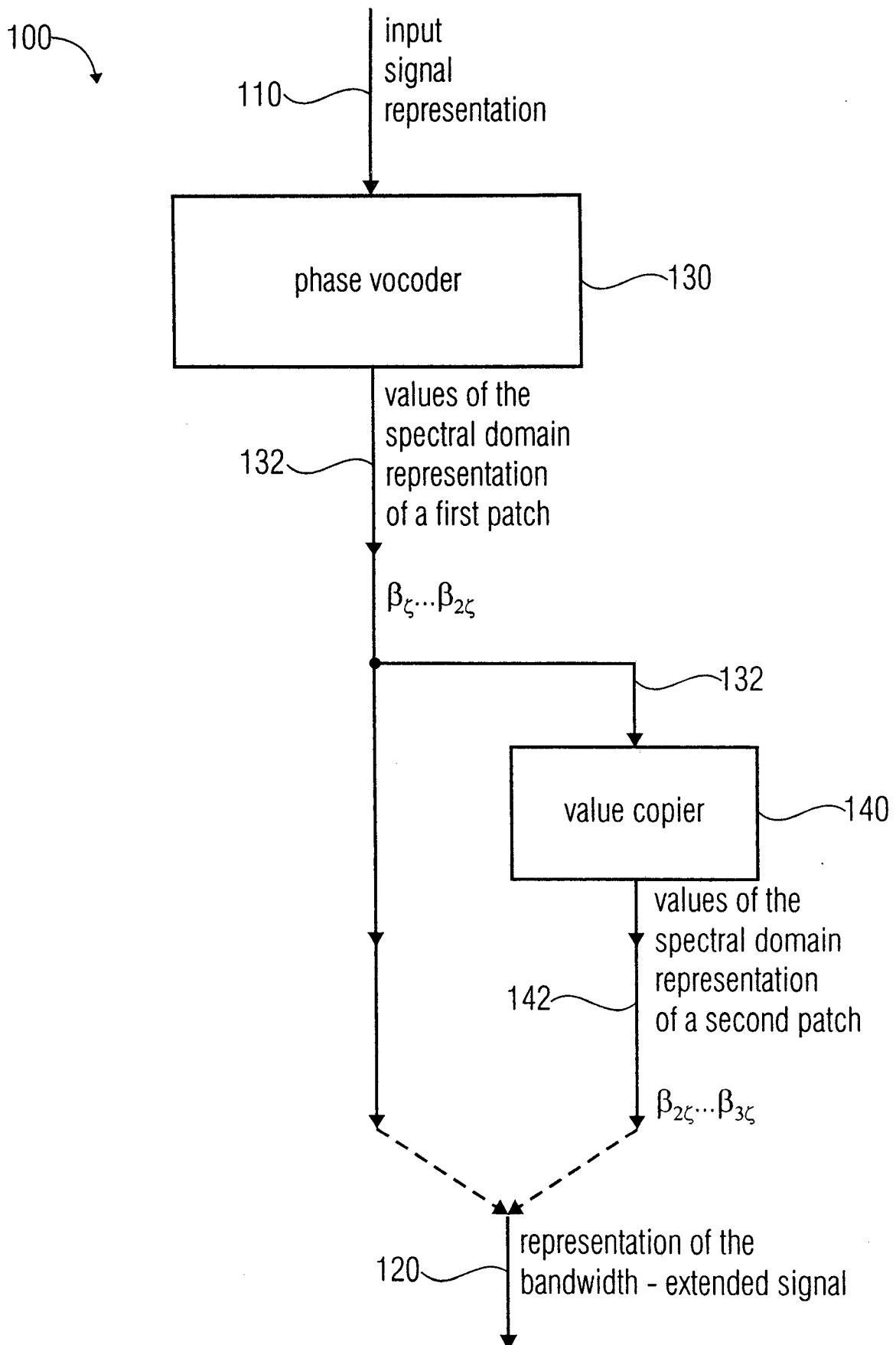


FIG 1

# PATENT COOPERATION TREATY

# PCT

## INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

|                                                            |                                                                                               |                                                         |
|------------------------------------------------------------|-----------------------------------------------------------------------------------------------|---------------------------------------------------------|
| Applicant's or agent's file reference<br>FH100402PCT       | <b>FOR FURTHER ACTION</b><br>see Form PCT/ISA/220 as well as, where applicable, item 5 below. |                                                         |
| International application No.<br>PCT/EP2010/054422         | International filing date (day/month/year)<br>01/04/2010                                      | (Earliest) Priority Date (day/month/year)<br>02/04/2009 |
| Applicant<br>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 43.6bis(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. 1  
☒ as suggested by the applicant  
☐ as selected by this Authority, because the applicant failed to suggest a figure  
☐ as selected by this Authority, because this figure better characterizes the invention  
 b. ☐ none of the figures is to be published with the abstract

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2010/054422

## A. CLASSIFICATION OF SUBJECT MATTER

INV. G10L21/02  
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, INSPEC

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                               | Relevant to claim No. |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| A         | WO 98/57436 A2 (LILJERYD LARS GUSTAF [SE]; EKSTRAND PER RUNE ALBIN [SE]; HENN LARS FRE) 17 December 1998 (1998-12-17)<br>figure 2<br>page 6, line 20 - line 34<br>figure 8                                                                                                       | 1,15-17               |
| A         | DIETZ M ET AL: "SPECTRAL BAND REPLICATION, A NOVEL APPROACH IN AUDIO CODING"<br>AUDIO ENGINEERING SOCIETY CONVENTION PAPER, NEW YORK, NY, US,<br>vol. 112, no. 5553,<br>10 May 2002 (2002-05-10), pages 1-08,<br>XP009020921<br>cited in the application<br>abstract<br>figure 6 | 1,15-17               |

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

### \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

Date of the actual completion of the international search

21 May 2010

Date of mailing of the international search report

07/06/2010

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

Krembel, Luc

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2010/054422

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                              | Relevant to claim No. |
|-----------|---------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| A         | <p>US 2007/238415 A1 (SINHA DEEPEN [US] ET AL) 11 October 2007 (2007-10-11) paragraphs [0089], [0090] figure 8</p> <p>-----</p> | 1,15-17               |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2010/054422

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|-------------------------------------------|---------------------|----------------------------|---------------------|
| WO 9857436                                | A2                  | 17-12-1998                 |                     |
|                                           |                     | AT 303679 T                | 15-09-2005          |
|                                           |                     | AT 257987 T                | 15-01-2004          |
|                                           |                     | AU 7446598 A               | 30-12-1998          |
|                                           |                     | BR 9805989 A               | 31-08-1999          |
|                                           |                     | CN 1272259 A               | 01-11-2000          |
|                                           |                     | CN 1629937 A               | 22-06-2005          |
|                                           |                     | DE 69821089 D1             | 19-02-2004          |
|                                           |                     | DE 69821089 T2             | 11-11-2004          |
|                                           |                     | DE 69831435 D1             | 06-10-2005          |
|                                           |                     | DE 69831435 T2             | 14-06-2006          |
|                                           |                     | DK 1367566 T3              | 10-10-2005          |
|                                           |                     | DK 940015 T3               | 26-04-2004          |
|                                           |                     | EP 0940015 A1              | 08-09-1999          |
|                                           |                     | ES 2247466 T3              | 01-03-2006          |
|                                           |                     | ES 2213901 T3              | 01-09-2004          |
|                                           |                     | HK 1030843 A1              | 23-09-2005          |
|                                           |                     | HK 1057815 A1              | 11-11-2005          |
|                                           |                     | JP 3871347 B2              | 24-01-2007          |
|                                           |                     | JP 2001521648 T            | 06-11-2001          |
|                                           |                     | JP 4220461 B2              | 04-02-2009          |
|                                           |                     | JP 2005173607 A            | 30-06-2005          |
|                                           |                     | PT 1367566 E               | 30-11-2005          |
|                                           |                     | PT 940015 E                | 30-06-2004          |
|                                           |                     | SE 512719 C2               | 02-05-2000          |
|                                           |                     | SE 9800268 A               | 11-12-1998          |
|                                           |                     | US 6680972 B1              | 20-01-2004          |
| US 2007238415                             | A1                  | 11-10-2007                 | NONE                |

# PATENT COOPERATION TREATY

From the  
INTERNATIONAL SEARCHING AUTHORITY

# PCT

To:

see form PCT/ISA/220

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/EP2010/054422

International filing date (day/month/year)  
01.04.2010

Priority date (day/month/year)  
02.04.2009

International Patent Classification (IPC) or both national classification and IPC  
INV. G10L21/02

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

Krembel, Luc

Telephone No. +49 89 2399-6005



WRITTEN OPINION OF THE  
INTERNATIONAL SEARCHING AUTHORITY

International application No.  
PCT/EP2010/054422

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:
  - ☒ the international application in the language in which it was filed
  - ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).
2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43bis.1(a))
3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing filed or furnished:
  - a. (means)
    - ☐ on paper
    - ☐ in electronic form
  - b. (time)
    - ☐ in the international application as filed
    - ☐ together with the international application in electronic form
    - ☐ subsequently to this Authority for the purposes of search
4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
5. Additional comments:

Box No. V Reasoned statement under Rule 43bis.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>1-17</u> |
|                               | No: Claims  |             |
| Inventive step (IS)           | Yes: Claims | <u>1-17</u> |
|                               | No: Claims  |             |
| Industrial applicability (IA) | Yes: Claims | <u>1-17</u> |
|                               | No: Claims  |             |

2. Citations and explanations

see separate sheet

**WRITTEN OPINION OF THE  
INTERNATIONAL SEARCHING AUTHORITY**

International application No.  
PCT/EP2010/054422

---

**Box No. VIII    Certain observations on the international application**

---

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

see separate sheet

**Re Item V**

**Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1 Reference is made to documents:

D1: WO 98/57436 A2 (LILJERYD LARS GUSTAF [SE]; EKSTRAND PER RUNE ALBIN [SE]; HENN LARS FRE) 17 December 1998 (1998-12-17)

D2: DIETZ M ET AL: "SPECTRAL BAND REPLICATION, A NOVEL APPROACH IN AUDIO CODING" AUDIO ENGINEERING SOCIETY CONVENTION PAPER, NEW YORK, NY, US, vol. 112, no. 5553, 10 May 2002 (2002-05-10) , pages 1-08, XP009020921 cited in the application

D3: US 2007/238415 A1 (SINHA DEEPEN [US] ET AL) 11 October 2007 (2007-10-11)

2 Claim 1

2.1 The closest prior art is disclosed by document D1 which discloses the following features:

- Apparatus for generating a representation of a bandwidth-extended signal on the basis of an input signal representation the apparatus comprising a plurality of frequency transposers (see D1, page 6 lines 20-34, fig.8 ref.807,809,811).

2.2 D1 does not disclose the features:

f1: a phase vocoder configured to obtain values of a spectral domain representation of a first patch of the bandwidth-extended signal on the basis of the input signal representation

f2: a value copier configured to copy a set of values of the spectral domain representation of the first patch, which values are provided by the phase vocoder, to obtain a set of values of a spectral domain representation of a second patch, wherein the second patch is associated with higher frequencies

than the first patch;

f3: wherein the apparatus is configured to obtain the representation of the bandwidth-extended signal using the values of the spectral domain representation of the first patch and the value of the spectral domain representation of the second patch.

- 2.3 The subject-matter of claim 1 is novel. Claim 1 meets the requirements of Article 33(2) PCT.
- 2.4 The problems solved by features f1,f2,f3 appears that of the computation efficient extension of a bandwidth limited signal.
- 2.5 Although document D1 discloses a frequency transposition method having effects similar to phase vocoding, in that a first patch is generated by stretching the input signal into the frequency domain, D1 does not disclose nor indicate copying up the patch obtained by transposition.
- 2.6 Document D2 discloses state-of-the art bandwidth extension using copy-up operation, but does not disclose not hint that said copy-up should be based on a first spectral patch.
- 2.7 Document D3 discloses an iterative bandwidth extension, wherein spectral patches are obtained and used for subsequent patch generation. D3 however does not disclose either phase vocoding, or copy-up operation. In particular, D3 does not hint at using copy-up operation for the next iteration of the bandwidth extension.
- 2.8 The prior art does not teach or indicate in an obvious manner how to solve the problem of efficient bandwidth extension. The subject-matter defined by claim 1 involves an inventive step and fulfils the requirements of Article 33(3) PCT.
- 3 Claim 14

3.1 The same conclusion as for claim 1 applies mutatis mutandis. Claim 14 meets the requirements of articles 33(2) and 33(3) PCT.

4 Claim 15

4.1 Claim 15 defines an alternative solution to the solution of claim 1, whereas the order of application of features f1,f2 is inverted. D2 is the closest prior art in this case.

4.2 Claim 15 is novel because D2 does not disclose applying phase vocoding on a spectral patch obtained by copy-up operation. Claim 15 meets the requirements of Article 33(2) PCT.

4.3 For reasons analog to those presented under items 2.4-2.8, claim 15 is considered to solve the problem of efficient bandwidth extension, and to define a non obvious solution with regard to the available prior art. Claim 15 meets the requirements of Article 33(3) PCT.

5 Claim 16

The same conclusion as for claim 15 applies mutatis mutandis. Claim 16 meets the requirements of articles 33(2) and 33(3) PCT.

6 Claim 17

6.1 The same conclusion as for claims 14,16 applies mutatis mutandis. Claim 17 meets the requirements of articles 33(2) and 33(3) PCT.

### **Re Item VIII**

#### **Certain observations on the international application**

1 The drawings do not meet the requirements of Rule 11.13 PCT.

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 | <p>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).</p> <p>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).</p> |
| Filing informal comments                                  | After receipt of the ISR/WO-ISA the applicant may file informal comments on the WO-ISA directly with the International Bureau of WIPO. These will be communicated to the designated Offices together with the IPRP (International Preliminary Report on Patentability) at 30 months from the priority date. Please also refer to the next box.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| End of the international phase                            | At the end of the international phase the International Bureau of WIPO will transform the WO-ISA or, if a demand was filed, the written opinion of the IPEA into the IPRP, which will then be transmitted together with possible informal comments to the designated Offices. The IPRP replaces the former IPER (international preliminary examination report).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| Relevant PCT Rules and more information                   | Rule 43 PCT, Rule 43bis PCT, Rule 44 PCT, Rule 44bis PCT, PCT Newsletter 12/2003, OJ 11/2003, OJ 12/2003                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

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.

# PATENT COOPERATION TREATY

PCT/EP2010/054422

## ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

### PCT

#### NOTIFICATION CONCERNING SUBMISSION OR TRANSMITTAL OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

To:

BURGER, Markus  
P.O. Box 246  
82043 Pullach  
ALLEMAGNE

|                                                                                         |                                                                                                                                              |
|-----------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| Date of mailing (day/month/year)<br>08 June 2010 (08.06.2010)                           | IMPORTANT NOTIFICATION                                                                                                                       |
| Applicant's or agent's file reference<br>FH100402PCT                                    |                                                                                                                                              |
| International application No.<br>PCT/EP2010/054422                                      |                                                                                                                                              |
| International publication date (day/month/year)<br>Not yet published                    |                                                                                                                                              |
| Applicant<br>Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. et al | International filing date (day/month/year)<br>01 April 2010 (01.04.2010)<br><br>Priority date (day/month/year)<br>02 April 2009 (02.04.2009) |

- By means of this Form, which replaces any previously issued notification concerning submission or transmittal of priority documents, the applicant is hereby notified of the date of receipt by the International Bureau of the priority document(s) relating to all earlier application(s) whose priority is claimed. Unless otherwise indicated by the letters "NR", in the right-hand column or by an asterisk appearing next to a date of receipt, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).
- (If applicable)* The letters "NR" appearing in the right-hand column denote a priority document which, on the date of mailing of this Form, had not yet been received by the International Bureau under Rule 17.1(a) or (b). Where, under Rule 17.1(a), the priority document must be submitted by the applicant to the receiving Office or the International Bureau, but the applicant fails to submit the priority document within the applicable time limit under that Rule, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.
- (If applicable)* An asterisk (\*) appearing next to a date of receipt, in the right-hand column, denotes a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b) (the priority document was received after the time limit prescribed in Rule 17.1(a) or the request to prepare and transmit the priority document was submitted to the receiving Office after the applicable time limit under Rule 17.1(b)). Even though the priority document was not furnished in compliance with Rule 17.1(a) or (b), the International Bureau will nevertheless transmit a copy of the document to the designated Offices, for their consideration. In case such a copy is not accepted by the designated Office as the priority document, Rule 17.1(c) provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

| Priority date                 | Priority application No. | Country or regional Office<br>or PCT receiving Office | Date of receipt<br>of priority document |
|-------------------------------|--------------------------|-------------------------------------------------------|-----------------------------------------|
| 02 April 2009 (02.04.2009)    | 61/166,125               | US                                                    | 31 May 2010 (31.05.2010)                |
| 09 April 2009 (09.04.2009)    | 61/168,068               | US                                                    | 31 May 2010 (31.05.2010)                |
| 30 December 2009 (30.12.2009) | 09181008.5               | EP                                                    | 31 May 2010 (31.05.2010)                |

|                                                                                               |                                                                                                            |
|-----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| The International Bureau of WIPO<br>34, chemin des Colombettes<br>1211 Geneva 20, Switzerland | Authorized officer<br><br>Yolaine Cussac<br><br>e-mail PT05.PCT@WIPO.INT<br>Telephone No. +41 22 338 74 05 |
|-----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|

Facsimile No. +41 22 338 82 70

Form PCT/IB/304 (October 2005)

1/EHJ63P7M0

192

# PATENT COOPERATION TREATY

From the RECEIVING OFFICE

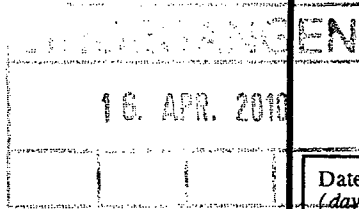
# PCT

To:

Burger, Markus  
SCHOPPE, ZIMMERMANN, STÖCKELER & ZINKLER  
Postfach 246  
D-82043 Pullach  
ALLEMAGNE

NOTIFICATION OF THE INTERNATIONAL  
APPLICATION NUMBER AND OF THE  
INTERNATIONAL FILING DATE

(PCT Rule 20.2(c))



Date of mailing  
(day/month/year)

14-04-2010

Applicant's or agent's file reference  
FH100402PCT

IMPORTANT NOTIFICATION

International application No.

PCT/EP2010/054422

International filing date (day/month/year)

1 April 2010 (01-04-2010)

Priority date (day/month/year)

2 April 2009 (02-04-2009)

Applicant

FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ...

Title of the invention

1. The applicant is hereby notified that the international application has been accorded the international application number and the international filing date indicated above.

2. The applicant is further notified that the record copy of the international application:



was transmitted to the International Bureau on see above date of mailing



has not yet been transmitted to the International Bureau for the reason indicated below and a copy of this notification has been sent to the International Bureau\*:



because the necessary national security clearance has not yet been obtained



because (reason to be specified)

Please see additional information on enclosed form PCT/RO/106

\* The International Bureau monitors the transmittal of the record copy by the receiving Office and will notify the applicant (with Form PCT/IB/301) of its receipt. Should the record copy not have been received by the expiration of 14 months from the priority date, the International Bureau will notify the applicant (Rule 22.1(c)).

Name and mailing address of the Receiving Office



European Patent Office, P.B. 5818 Patentlaan 2  
NL-2280 HV Rijswijk  
Tel. (+31-70) 340-2040  
Fax: (+31-70) 340-3016

Authorized officer

Carr, Juliette



Acknowledgement of receipt

We hereby acknowledge receipt of your request for the processing of an international application according to the Patent Cooperation Treaty as follows:

|                        |                                                                                                                                                                                                                                                   |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Submission number      | 793017                                                                                                                                                                                                                                            |
| PCT application number | PCT/EP2010/054422                                                                                                                                                                                                                                 |
| Date of receipt        | 01 April 2010                                                                                                                                                                                                                                     |
| Receiving Office       | European Patent Office, The Hague                                                                                                                                                                                                                 |
| Your reference         | FH100402PCT                                                                                                                                                                                                                                       |
| Applicant              | Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.                                                                                                                                                                              |
| Number of applicants   | 7                                                                                                                                                                                                                                                 |
| Country                | DE                                                                                                                                                                                                                                                |
| Title                  | APPARATUS, METHOD AND COMPUTER PROGRAM FOR GENERATING A REPRESENTATION OF A BANDWIDTH-EXTENDED SIGNAL ON THE BASIS OF AN INPUT SIGNAL REPRESENTATION USING A COMBINATION OF A HARMONIC BANDWIDTH-EXTENSION AND A NON-HARMONIC BANDWIDTH-EXTENSION |

|                     |                                                                                                                                                            |                                                                                                                        |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| Documents submitted | <div>eolf-pkda.xml</div> <div>eolf-appb.xml</div> <div>eolf-vlog.xml</div> <div>eolf-othd-000001.pdf (37 p.)</div> <div>eolf-appb-P000001.pdf (6 p.)</div> | <div>eolf-requ.xml</div> <div>eolf-fees.xml</div> <div>pct101.GML</div> <div>eolf-abst.txt</div> <div>pct101.1WO</div> |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|

|                                 |                                                                    |              |
|---------------------------------|--------------------------------------------------------------------|--------------|
|                                 | pct101u.gml                                                        | referenc.inf |
| Submitted by                    | CN=M. Burger 21130,O=Schoppe, Zimmermann, Stöckeler & Zinkler,C=FR |              |
| Method of submission            | Online                                                             |              |
| Date and time receipt generated | 01 April 2010, 14:00 (CEST)                                        |              |
| Message Digest                  | AC:75:06:24:10:DC:36:4D:FC:1A:B8:87:A0:13:54:A3:F3:0F:3B:DB        |              |

**Correction by the EPO of errors in debit instructions filed by eOLF**

Errors in debit instructions filed by eOLF that are caused by the editing of Form 1038E entries or the continued use of outdated software (all forms) may be corrected automatically by the EPO, leaving the payment date unchanged (see decision T 152/82, OJ EPO 1984, 301 and point 6.3 ff ADA, Supplement to OJ EPO 10/2007).

/European Patent Office/

## PCT REQUEST

Print Out (Original in Electronic Form)

|                |                                                                                                                                                |                                                                                                                                                                                                                                                   |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>0</b>       | <b>For receiving Office use only</b>                                                                                                           |                                                                                                                                                                                                                                                   |
| <b>0-1</b>     | International Application No.                                                                                                                  |                                                                                                                                                                                                                                                   |
| <b>0-2</b>     | International Filing Date                                                                                                                      |                                                                                                                                                                                                                                                   |
| <b>0-3</b>     | Name of receiving Office and "PCT International Application"                                                                                   |                                                                                                                                                                                                                                                   |
| <b>0-4</b>     | <b>Form PCT/RO/101 PCT Request</b>                                                                                                             |                                                                                                                                                                                                                                                   |
| <b>0-4-1</b>   | Prepared Using                                                                                                                                 | PCT Online Filing<br>Version 3.5.000.219 MT/FOP<br>20020701/0.20.5.9                                                                                                                                                                              |
| <b>0-5</b>     | <b>Petition</b><br>The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty |                                                                                                                                                                                                                                                   |
| <b>0-6</b>     | <b>Receiving Office (specified by the applicant)</b>                                                                                           | European Patent Office (EPO) (RO/EP)                                                                                                                                                                                                              |
| <b>0-7</b>     | <b>Applicant's or agent's file reference</b>                                                                                                   | FH100402PCT                                                                                                                                                                                                                                       |
| <b>I</b>       | <b>Title of Invention</b>                                                                                                                      | APPARATUS, METHOD AND COMPUTER PROGRAM FOR GENERATING A REPRESENTATION OF A BANDWIDTH-EXTENDED SIGNAL ON THE BASIS OF AN INPUT SIGNAL REPRESENTATION USING A COMBINATION OF A HARMONIC BANDWIDTH-EXTENSION AND A NON-HARMONIC BANDWIDTH-EXTENSION |
| <b>II</b>      | <b>Applicant</b>                                                                                                                               |                                                                                                                                                                                                                                                   |
| <b>II-1</b>    | This person is                                                                                                                                 | Applicant only                                                                                                                                                                                                                                    |
| <b>II-2</b>    | Applicant for                                                                                                                                  | All designated States except US                                                                                                                                                                                                                   |
| <b>II-4</b>    | Name                                                                                                                                           | Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.                                                                                                                                                                              |
| <b>II-5</b>    | Address                                                                                                                                        | Hansastraße 27c<br>80686 München<br>Germany                                                                                                                                                                                                       |
| <b>II-6</b>    | State of nationality                                                                                                                           | DE                                                                                                                                                                                                                                                |
| <b>II-7</b>    | State of residence                                                                                                                             | DE                                                                                                                                                                                                                                                |
| <b>III-1</b>   | <b>Applicant and/or inventor</b>                                                                                                               |                                                                                                                                                                                                                                                   |
| <b>III-1-1</b> | This person is                                                                                                                                 | Applicant and inventor                                                                                                                                                                                                                            |
| <b>III-1-2</b> | Applicant for                                                                                                                                  | US only                                                                                                                                                                                                                                           |
| <b>III-1-4</b> | Name (LAST, First)                                                                                                                             | NAGEL, Frederik                                                                                                                                                                                                                                   |
| <b>III-1-5</b> | Address                                                                                                                                        | Wilhelmshavener Strasse 72<br>90425 Nürnberg<br>Germany                                                                                                                                                                                           |
| <b>III-1-6</b> | State of nationality                                                                                                                           | DE                                                                                                                                                                                                                                                |
| <b>III-1-7</b> | State of residence                                                                                                                             | DE                                                                                                                                                                                                                                                |

## PCT REQUEST

Print Out (Original in Electronic Form)

|         |                           |                                                    |
|---------|---------------------------|----------------------------------------------------|
| III-2   | Applicant and/or inventor |                                                    |
| III-2-1 | This person is            | Applicant and inventor                             |
| III-2-2 | Applicant for             | US only                                            |
| III-2-4 | Name (LAST, First)        | NEUENDORF, Max                                     |
| III-2-5 | Address                   | Theatergasse 17<br>90402 Nuernberg<br>Germany      |
| III-2-6 | State of nationality      | DE                                                 |
| III-2-7 | State of residence        | DE                                                 |
| III-3   | Applicant and/or inventor |                                                    |
| III-3-1 | This person is            | Applicant and inventor                             |
| III-3-2 | Applicant for             | US only                                            |
| III-3-4 | Name (LAST, First)        | RETTELBACH, Nikolaus                               |
| III-3-5 | Address                   | Spessartstr. 38<br>90427 Nürnberg<br>Germany       |
| III-3-6 | State of nationality      | DE                                                 |
| III-3-7 | State of residence        | DE                                                 |
| III-4   | Applicant and/or inventor |                                                    |
| III-4-1 | This person is            | Applicant and inventor                             |
| III-4-2 | Applicant for             | US only                                            |
| III-4-4 | Name (LAST, First)        | LECOMTE, Jérémie                                   |
| III-4-5 | Address                   | Sulzbacher Strasse 39<br>90489 Nürnberg<br>Germany |
| III-4-6 | State of nationality      | FR                                                 |
| III-4-7 | State of residence        | DE                                                 |
| III-5   | Applicant and/or inventor |                                                    |
| III-5-1 | This person is            | Applicant and inventor                             |
| III-5-2 | Applicant for             | US only                                            |
| III-5-4 | Name (LAST, First)        | MULTRUS, Markus                                    |
| III-5-5 | Address                   | Etzlaubweg 7<br>90469 Nürnberg<br>Germany          |
| III-5-6 | State of nationality      | DE                                                 |
| III-5-7 | State of residence        | DE                                                 |

## PCT REQUEST

Print Out (Original in Electronic Form)

|                |                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                       |
|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| III-6          | <b>Applicant and/or inventor</b>                                                                                                                                                                                                                                                                                                           |                                                                                                                                       |
| III-6-1        | This person is                                                                                                                                                                                                                                                                                                                             | Applicant and inventor                                                                                                                |
| III-6-2        | Applicant for                                                                                                                                                                                                                                                                                                                              | US only                                                                                                                               |
| III-6-4        | Name (LAST, First)                                                                                                                                                                                                                                                                                                                         | GRILL, Bernhard                                                                                                                       |
| III-6-5        | Address                                                                                                                                                                                                                                                                                                                                    | Peter-Henlein-Str. 7<br>91207 Lauf<br>Germany                                                                                         |
| III-6-6        | State of nationality                                                                                                                                                                                                                                                                                                                       | DE                                                                                                                                    |
| III-6-7        | State of residence                                                                                                                                                                                                                                                                                                                         | DE                                                                                                                                    |
| IV-1           | <b>Agent or common representative; or address for correspondence</b><br>The person identified below is hereby/ has been appointed to act on behalf of the applicant(s) before the competent International Authorities as:                                                                                                                  | Agent                                                                                                                                 |
| IV-1-1         | Name (LAST, First)                                                                                                                                                                                                                                                                                                                         | BURGER, Markus                                                                                                                        |
| IV-1-2         | Address                                                                                                                                                                                                                                                                                                                                    | P.O. Box 246<br>82043 Pullach<br>Germany                                                                                              |
| IV-1-3         | Telephone No.                                                                                                                                                                                                                                                                                                                              | 089-790 445-0                                                                                                                         |
| IV-1-4         | Facsimile No.                                                                                                                                                                                                                                                                                                                              | 089-790 22 15                                                                                                                         |
| IV-1-5         | e-mail                                                                                                                                                                                                                                                                                                                                     | mail@schoppe-zimmermann.com                                                                                                           |
| IV-1-5(a)<br>) | E-mail authorization<br>The receiving Office, the International Searching Authority, the International Bureau and the International Preliminary Examining Authority are authorized to use this e-mail address to send, if the Office or Authority so wishes, advance copies of notifications in respect of this international application. | Yes                                                                                                                                   |
| IV-2           | <b>Additional agent(s)</b>                                                                                                                                                                                                                                                                                                                 | additional agent(s) with same address as first named agent                                                                            |
| IV-2-1         | Name(s)                                                                                                                                                                                                                                                                                                                                    | SCHOPPE, Fritz; ZIMMERMANN, Tankred;<br>STÖCKELER, Ferdinand; SCHENK, Markus;<br>HERSINA, Günter; GÖTTLING, Stefan;<br>ZINKLER, Franz |
| V              | <b>DESIGNATIONS</b>                                                                                                                                                                                                                                                                                                                        |                                                                                                                                       |
| V-1            | The filing of this request constitutes under Rule 4.9(a), the designation of all Contracting States bound by the PCT on the international filing date, for the grant of every kind of protection available and, where applicable, for the grant of both regional and national patents.                                                     |                                                                                                                                       |
| VI-1           | <b>Priority claim of earlier national application</b>                                                                                                                                                                                                                                                                                      |                                                                                                                                       |
| VI-1-1         | Filing date                                                                                                                                                                                                                                                                                                                                | 02 April 2009 (02.04.2009)                                                                                                            |
| VI-1-2         | Number                                                                                                                                                                                                                                                                                                                                     | 61/166,125                                                                                                                            |
| VI-1-3         | Country                                                                                                                                                                                                                                                                                                                                    | US                                                                                                                                    |

## PCT REQUEST

Print Out (Original in Electronic Form)

|              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                       |                             |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|-----------------------------|
| <b>VI-2</b>  | <b>Priority claim of earlier national application</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                       |                             |
| VI-2-1       | Filing date                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 09 April 2009 (09.04.2009)            |                             |
| VI-2-2       | Number                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 61/168,068                            |                             |
| VI-2-3       | Country                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | US                                    |                             |
| <b>VI-3</b>  | <b>Priority claim of earlier regional application</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                       |                             |
| VI-3-1       | Filing date                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 30 December 2009 (30.12.2009)         |                             |
| VI-3-2       | Number                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 09181008.5                            |                             |
| VI-3-3       | Regional Office                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | EP                                    |                             |
| <b>VI-4</b>  | <b>Incorporation by reference :</b><br>where an element of the international application referred to in Article 11(1)(iii)(d) or (e) or a part of the description, claims or drawings referred to in Rule 20.5(a) is not otherwise contained in this international application but is completely contained in an earlier application whose priority is claimed on the date on which one or more elements referred to in Article 11(1)(iii) were first received by the receiving Office, that element or part is, subject to confirmation under Rule 20.6, incorporated by reference in this international application for the purposes of Rule 20.6. |                                       |                             |
| <b>VII-1</b> | <b>International Searching Authority Chosen</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | European Patent Office (EPO) (ISA/EP) |                             |
| <b>VIII</b>  | <b>Declarations</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Number of declarations                |                             |
| VIII-1       | Declaration as to the identity of the inventor                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | -                                     |                             |
| VIII-2       | Declaration as to the applicant's entitlement, as at the international filing date, to apply for and be granted a patent                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | -                                     |                             |
| VIII-3       | Declaration as to the applicant's entitlement, as at the international filing date, to claim the priority of the earlier application                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | -                                     |                             |
| VIII-4       | Declaration of inventorship (only for the purposes of the designation of the United States of America)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | -                                     |                             |
| VIII-5       | Declaration as to non-prejudicial disclosures or exceptions to lack of novelty                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | -                                     |                             |
| <b>IX</b>    | <b>Check list</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | Number of sheets                      | Electronic file(s) attached |
| IX-1         | Request (including declaration sheets)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 5                                     | ✓                           |
| IX-2         | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 29                                    | ✓                           |
| IX-3         | Claims                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 7                                     | ✓                           |
| IX-4         | Abstract                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 1                                     | ✓                           |
| IX-5         | Drawings                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 6                                     | ✓                           |
| IX-7         | TOTAL                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 48                                    |                             |

PCT REQUEST

Print Out (Original in Electronic Form)

|       | Accompanying Items                                         | Paper document(s) attached                                    | Electronic file(s) attached |
|-------|------------------------------------------------------------|---------------------------------------------------------------|-----------------------------|
| IX-8  | Fee calculation sheet                                      | -                                                             | ✓                           |
| IX-18 | PCT-SAFE physical media                                    | -                                                             | -                           |
| IX-20 | Figure of the drawings which should accompany the abstract |                                                               |                             |
| IX-21 | Language of filing of the international application        | English                                                       |                             |
| X-1   | Signature of applicant, agent or common representative     | (PKCS7 Digital Signature)                                     |                             |
| X-1-1 | Name (LAST, First)                                         | Burger, Markus                                                |                             |
| X-1-2 | Name of signatory                                          | FR, Schoppe, Zimmermann, Stöckeler & Zinkler, M. Burger 21130 |                             |
| X-1-3 | Capacity                                                   | (Representative)                                              |                             |

FOR RECEIVING OFFICE USE ONLY

|        |                                                                                                                                         |        |
|--------|-----------------------------------------------------------------------------------------------------------------------------------------|--------|
| 10-1   | Date of actual receipt of the purported international application                                                                       |        |
| 10-2   | Drawings:                                                                                                                               |        |
| 10-2-1 | Received                                                                                                                                |        |
| 10-2-2 | Not received                                                                                                                            |        |
| 10-3   | Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application |        |
| 10-4   | Date of timely receipt of the required corrections under PCT Article 11(2)                                                              |        |
| 10-5   | International Searching Authority                                                                                                       | ISA/EP |
| 10-6   | Transmittal of search copy delayed until search fee is paid                                                                             |        |

FOR INTERNATIONAL BUREAU USE ONLY

|      |                                                                |  |
|------|----------------------------------------------------------------|--|
| 11-1 | Date of receipt of the record copy by the International Bureau |  |
|------|----------------------------------------------------------------|--|

**PCT (ANNEX - FEE CALCULATION SHEET)**

Print Out (Original in Electronic Form)

(This sheet is not part of and does not count as a sheet of the international application)

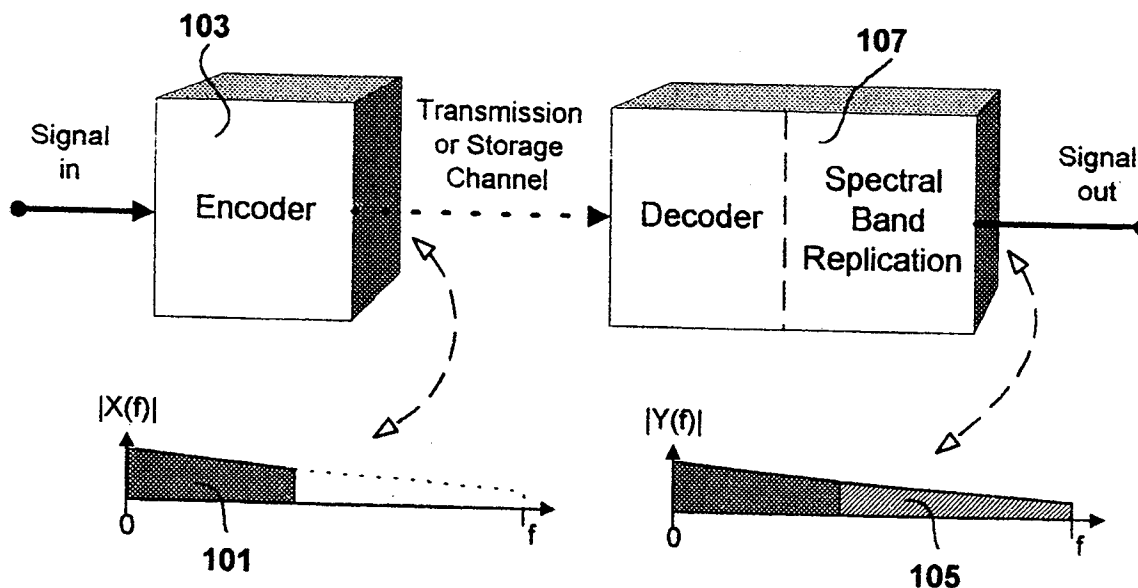
|                |                                                                                                    |                                                                                           |                                                                             |
|----------------|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>0</b>       | <b>For receiving Office use only</b>                                                               |                                                                                           |                                                                             |
| <b>0-1</b>     | International Application No.                                                                      |                                                                                           |                                                                             |
| <b>0-2</b>     | Date stamp of the receiving Office                                                                 |                                                                                           |                                                                             |
| <b>0-4</b>     | <b>Form PCT/RO/101 (Annex)</b>                                                                     |                                                                                           |                                                                             |
| <b>0-4-1</b>   | PCT Fee Calculation Sheet Prepared Using                                                           | <b>PCT Online Filing</b><br><b>Version 3.5.000.219 MT/FOP</b><br><b>20020701/0.20.5.9</b> |                                                                             |
| <b>0-9</b>     | Applicant's or agent's file reference                                                              |                                                                                           | <b>FH100402PCT</b>                                                          |
| <b>2</b>       | Applicant                                                                                          |                                                                                           | <b>FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.</b> |
| <b>12</b>      | <b>Calculation of prescribed fees</b>                                                              | Fee amount/multiplier                                                                     | Total amounts (EUR)                                                         |
| <b>12-1</b>    | Transmittal fee <b>T</b>                                                                           | ⇒                                                                                         | <b>115</b>                                                                  |
| <b>12-2-1</b>  | Search fee <b>S</b>                                                                                | ⇒                                                                                         | <b>1785</b>                                                                 |
| <b>12-2-2</b>  | International search to be carried out by                                                          | <b>EP</b>                                                                                 |                                                                             |
| <b>12-3</b>    | International filing fee (first 30 sheets) <b>i1</b>                                               | <b>878</b>                                                                                |                                                                             |
| <b>12-4</b>    | Remaining sheets                                                                                   | <b>18</b>                                                                                 |                                                                             |
| <b>12-5</b>    | Additional amount <b>(X)</b>                                                                       | <b>10</b>                                                                                 |                                                                             |
| <b>12-6</b>    | Total additional amount <b>i2</b>                                                                  | <b>180</b>                                                                                |                                                                             |
| <b>12-7</b>    | <b>i1 + i2 =</b> <b>i</b>                                                                          | <b>1058</b>                                                                               |                                                                             |
| <b>12-12</b>   | Electronic Filing reduction (Image) <b>R</b>                                                       | <b>-132</b>                                                                               |                                                                             |
| <b>12-13</b>   | Total International filing fee (i-R) <b>i</b>                                                      | ⇒                                                                                         | <b>926</b>                                                                  |
| <b>12-14</b>   | Fee for priority document                                                                          |                                                                                           |                                                                             |
|                | Number of priority documents requested                                                             | <b>0</b>                                                                                  |                                                                             |
| <b>12-15</b>   | Fee per document <b>(X)</b>                                                                        | <b>45</b>                                                                                 |                                                                             |
| <b>12-16</b>   | Total priority document fee: <b>P</b>                                                              | ⇒                                                                                         |                                                                             |
| <b>12-17</b>   | Fee for restoration of priority rights <b>RP</b>                                                   |                                                                                           |                                                                             |
|                | Number of requests for restoration of priority rights                                              | <b>0</b>                                                                                  |                                                                             |
|                | Total amount of fees for restoration of priority rights                                            |                                                                                           |                                                                             |
| <b>12-19</b>   | <b>TOTAL FEES PAYABLE (T+S+I+P+RP)</b>                                                             | ⇒                                                                                         | <b>2826</b>                                                                 |
| <b>12-21</b>   | Mode of payment <b>Authorization to charge current account</b>                                     |                                                                                           |                                                                             |
| <b>12-22</b>   | Current account instructions                                                                       |                                                                                           |                                                                             |
|                | The receiving Office                                                                               | <b>European Patent Office (EPO) (RO/EP)</b>                                               |                                                                             |
| <b>12-22-1</b> | Authorization to charge the total fees indicated above                                             | ✓                                                                                         |                                                                             |
| <b>12-22-2</b> | Authorization to charge any deficiency or credit any overpayment in the total fees indicated above | ✓                                                                                         |                                                                             |
| <b>12-23</b>   | Current account No.                                                                                |                                                                                           | <b>28000601</b>                                                             |
| <b>12-24</b>   | Date                                                                                               |                                                                                           | <b>01 April 2010 (01.04.2010)</b>                                           |
| <b>12-25</b>   | Name and signature                                                                                 |                                                                                           |                                                                             |



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (51) International Patent Classification <sup>6</sup> :<br><b>H04B 1/66</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <b>A3</b> | (11) International Publication Number: <b>WO 98/57436</b><br>(43) International Publication Date: 17 December 1998 (17.12.98)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| (21) International Application Number: PCT/IB98/00893<br>(22) International Filing Date: 9 June 1998 (09.06.98)<br>(30) Priority Data:<br>9702213-1 10 June 1997 (10.06.97) SE<br>9704634-6 12 December 1997 (12.12.97) SE<br>9800268-6 30 January 1998 (30.01.98) SE<br>(71)(72) Applicant and Inventor: LILJERYD, Lars, Gustaf [SE/SE]; Vintervagen 19, S-171 34 Solna (SE).<br>(72) Inventors; and<br>(75) Inventors/Applicants (for US only): EKSTRAND, Per, Rune, Albin [SE/SE]; Renstiernas Gata 29, S-116 31 Stockholm (SE). HENN, Lars, Fredrik [CH/SE]; Ritarvagen 14, S-168 31 Bromma (SE). KJÖRLING, Hans, Magnus, Kristofer [SE/SE]; Vindhemsgatan 19C, S-752 27 Uppsala (SE).<br>(74) Common Representative: LILJERYD, Lars, Gustaf; Vintervagen 19, S-171 34 Solna (SE). |           | (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).<br><br>Published<br>With international search report.<br>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.<br><br>(88) Date of publication of the international search report: 10 February 2000 (10.02.00) |

(54) Title: SOURCE CODING ENHANCEMENT USING SPECTRAL-BAND REPLICATION



## (57) Abstract

The present invention proposes a new method and apparatus for the enhancement of source coding systems. The invention employs bandwidth reduction (101) prior to or in the encoder (103), followed by spectral-band replication (105) at the decoder (107). This is accomplished by the use of new transposition methods, in combination with spectral envelope adjustments. Reduced bitrate at a given perceptual quality or an improved perceptual quality at a given bitrate is offered. The invention is preferably integrated in a hardware or software codec, but can also be implemented as a separate processor in combination with a codec. The invention offers substantial improvements practically independent of codec type and technological progress.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |                                          |    |                                              |    |                          |
|----|--------------------------|----|------------------------------------------|----|----------------------------------------------|----|--------------------------|
| AL | Albania                  | ES | Spain                                    | LS | Lesotho                                      | SI | Slovenia                 |
| AM | Armenia                  | FI | Finland                                  | LT | Lithuania                                    | SK | Slovakia                 |
| AT | Austria                  | FR | France                                   | LU | Luxembourg                                   | SN | Senegal                  |
| AU | Australia                | GA | Gabon                                    | LV | Latvia                                       | SZ | Swaziland                |
| AZ | Azerbaijan               | GB | United Kingdom                           | MC | Monaco                                       | TD | Chad                     |
| BA | Bosnia and Herzegovina   | GE | Georgia                                  | MD | Republic of Moldova                          | TG | Togo                     |
| BB | Barbados                 | GH | Ghana                                    | MG | Madagascar                                   | TJ | Tajikistan               |
| BE | Belgium                  | GN | Guinea                                   | MK | The former Yugoslav<br>Republic of Macedonia | TM | Turkmenistan             |
| BF | Burkina Faso             | GR | Greece                                   |    |                                              | TR | Turkey                   |
| BG | Bulgaria                 | HU | Hungary                                  | ML | Mali                                         | TT | Trinidad and Tobago      |
| BJ | Benin                    | IE | Ireland                                  | MN | Mongolia                                     | UA | Ukraine                  |
| BR | Brazil                   | IL | Israel                                   | MR | Mauritania                                   | UG | Uganda                   |
| BY | Belarus                  | IS | Iceland                                  | MW | Malawi                                       | US | United States of America |
| CA | Canada                   | IT | Italy                                    | MX | Mexico                                       | UZ | Uzbekistan               |
| CF | Central African Republic | JP | Japan                                    | NE | Niger                                        | VN | Viet Nam                 |
| CG | Congo                    | KE | Kenya                                    | NL | Netherlands                                  | YU | Yugoslavia               |
| CH | Switzerland              | KG | Kyrgyzstan                               | NO | Norway                                       | ZW | Zimbabwe                 |
| CI | Côte d'Ivoire            | KP | Democratic People's<br>Republic of Korea | NZ | New Zealand                                  |    |                          |
| CM | Cameroon                 |    |                                          | PL | Poland                                       |    |                          |
| CN | China                    | KR | Republic of Korea                        | PT | Portugal                                     |    |                          |
| CU | Cuba                     | KZ | Kazakistan                               | RO | Romania                                      |    |                          |
| CZ | Czech Republic           | LC | Saint Lucia                              | RU | Russian Federation                           |    |                          |
| DE | Germany                  | LI | Liechtenstein                            | SD | Sudan                                        |    |                          |
| DK | Denmark                  | LK | Sri Lanka                                | SE | Sweden                                       |    |                          |
| EE | Estonia                  | LR | Liberia                                  | SG | Singapore                                    |    |                          |

# INTERNATIONAL SEARCH REPORT

International Application No

/IB 98/00893

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 H04B1/66

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)  
IPC 6 H04B

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)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                        | Relevant to claim No.        |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|
| X          | US 5 127 054 A (HONG DAEHYOUNG ET AL) 30<br>June 1992<br>cited in the application<br>see figure FIGURE<br>see column 1, line 41 - line 50<br>see column 1, line 65 - column 2, line 2<br>see column 2, line 11 - column 3, line 33<br>--- | 1-4, 9,<br>11, 24,<br>26, 28 |
| X<br>A     | US 4 667 340 A (ARJMAND MASUD ET AL) 19<br>May 1987<br><br>see abstract; figures 4, 6<br>see column 3, line 8 - line 43<br>see column 7, line 48 - column 8, line 22<br>---<br>-/--                                                       | 1-4, 24,<br>26<br>11-14      |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in 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

26 April 1999

Date of mailing of the international search report

21.12.1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Sieben, S

INTERNATIONAL SEARCH REPORT

International Application No  
F /IB 98/00893

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                     |                           |
|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|
| Category *                                           | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                  | Relevant to claim No.     |
| A                                                    | US 5 068 899 A (ELLIS JOHN G ET AL) 26<br>November 1991<br>see abstract; figures 1-7<br>see column 1, line 39 - line 52<br>see column 2, line 1 - line 36<br>see column 2, line 64 - column 3, line 5<br>see column 5, line 62 - column 6, line 14<br>---           | 1-4,6,<br>24,26           |
| A                                                    | A J S FERREIRA: "Audio Spectral Coder"<br>AES REPRINT 4201, 100TH CONVENTION,<br>COPENHAGEN,<br>11 - 14 May 1996, XP002100534<br>cited in the application<br>see figure 3<br>see page 5, paragraph 4.2; figures 2,4,5<br>see page 7, paragraph 5; figure 4<br>----- | 1-4,7,<br>17,24,<br>26,30 |

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/IB 98/00893

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-20, 24-30

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-20,24-30

Enhancement of a source coding system comprising an encoder discarding frequency bands, a decoder performing a transposition of frequency bands from a 1st signal to a 2nd signal and a combiner to generate an output signal from 1st and 2nd signal

2. Claims: 21-23,31,32

Method and apparatus for computationally efficient transposition by expanding or compressing a time signal

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

/IB 98/00893

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|-------------------------------------------|---------------------|----------------------------|---------------------|
| US 5127054 A                              | 30-06-1992          | NONE                       |                     |
| US 4667340 A                              | 19-05-1987          | EP 0124728 A               | 14-11-1984          |
|                                           |                     | JP 1807228 C               | 10-12-1993          |
|                                           |                     | JP 5016599 B               | 04-03-1993          |
|                                           |                     | JP 60035799 A              | 23-02-1985          |
| US 5068899 A                              | 26-11-1991          | CA 1220282 A               | 07-04-1987          |



US 20070238415A1

(19) **United States**

(12) **Patent Application Publication**  
Sinha et al.

(10) **Pub. No.: US 2007/0238415 A1**

(43) **Pub. Date: Oct. 11, 2007**

(54) **METHOD AND APPARATUS FOR ENCODING AND DECODING**

(76) Inventors: **Deepen Sinha**, Chatham, NJ (US);  
**Anibal J. S. Ferreira**, Sao Mamede Infesta (PT); **Erumbi Vallabhan Harinarayanan**, Noida (IN)

**Publication Classification**

(51) **Int. Cl.**  
**H04B 7/00** (2006.01)

(52) **U.S. Cl.** ..... **455/66.1**

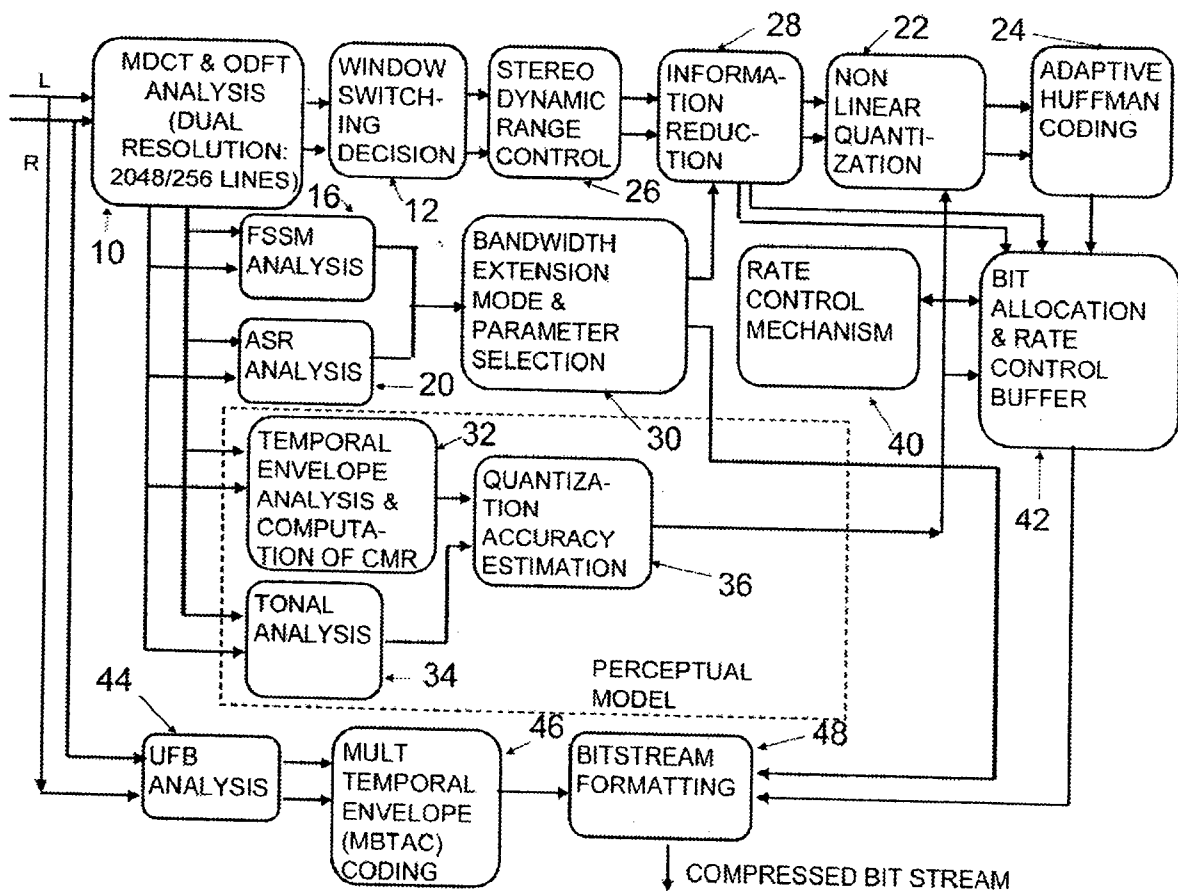
Correspondence Address:  
**THE LAW OFFICES OF THOMAS L. ADAMS**  
**120 EAGLE ROCK AVENUE**  
**P.O. BOX 340**  
**EAST HANOVER, NJ 07936 (US)**

(21) Appl. No.: **11/544,901**  
(22) Filed: **Oct. 6, 2006**

**Related U.S. Application Data**  
(60) Provisional application No. 60/724,856, filed on Oct. 7, 2005.

(57) **ABSTRACT**

A novel bandwidth extension technique allows information to be encoded and decoded using a fractal self similarity model or an accurate spectral replacement model, or both. Also a multi-band temporal amplitude coding technique, useful as an enhancement to any coding/decoding technique, helps with accurate reconstruction of the temporal envelope and employs a utility filterbank. A perceptual coder using a comodulation masking release model, operating typically with more conventional perceptual coders, makes the perceptual model more accurate and hence increases the efficiency of the overall perceptual coder.



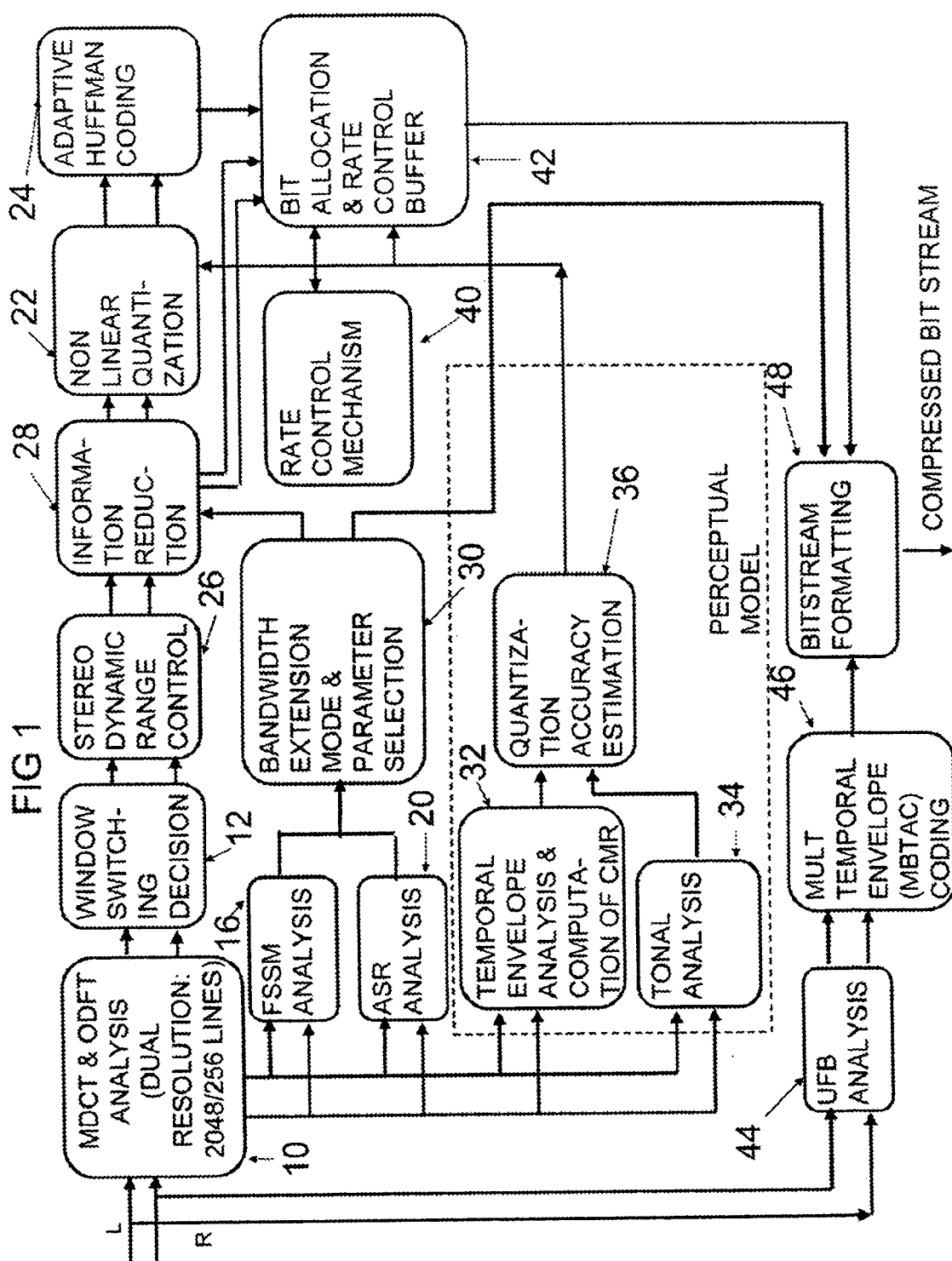


FIG 2

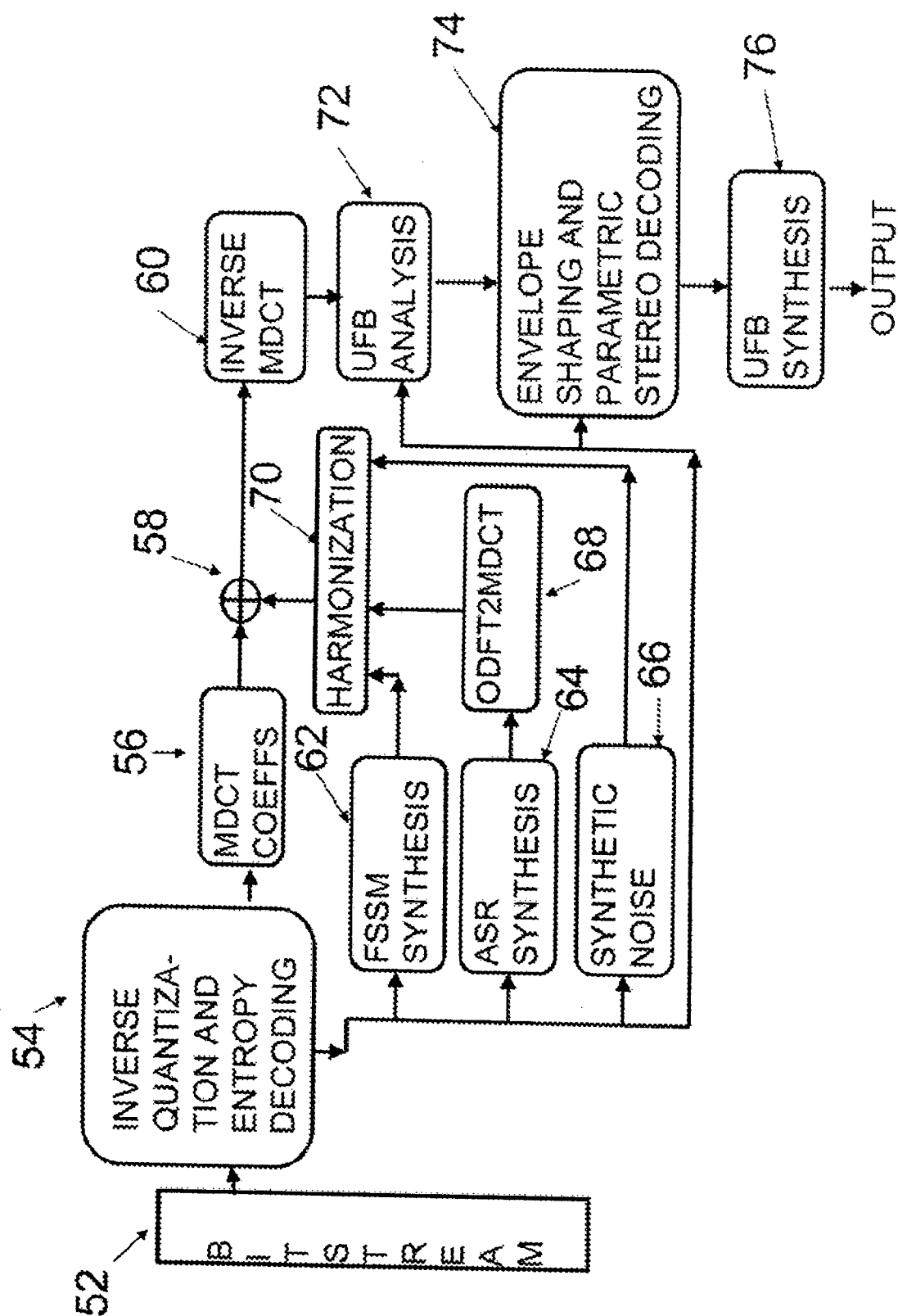


FIG 3

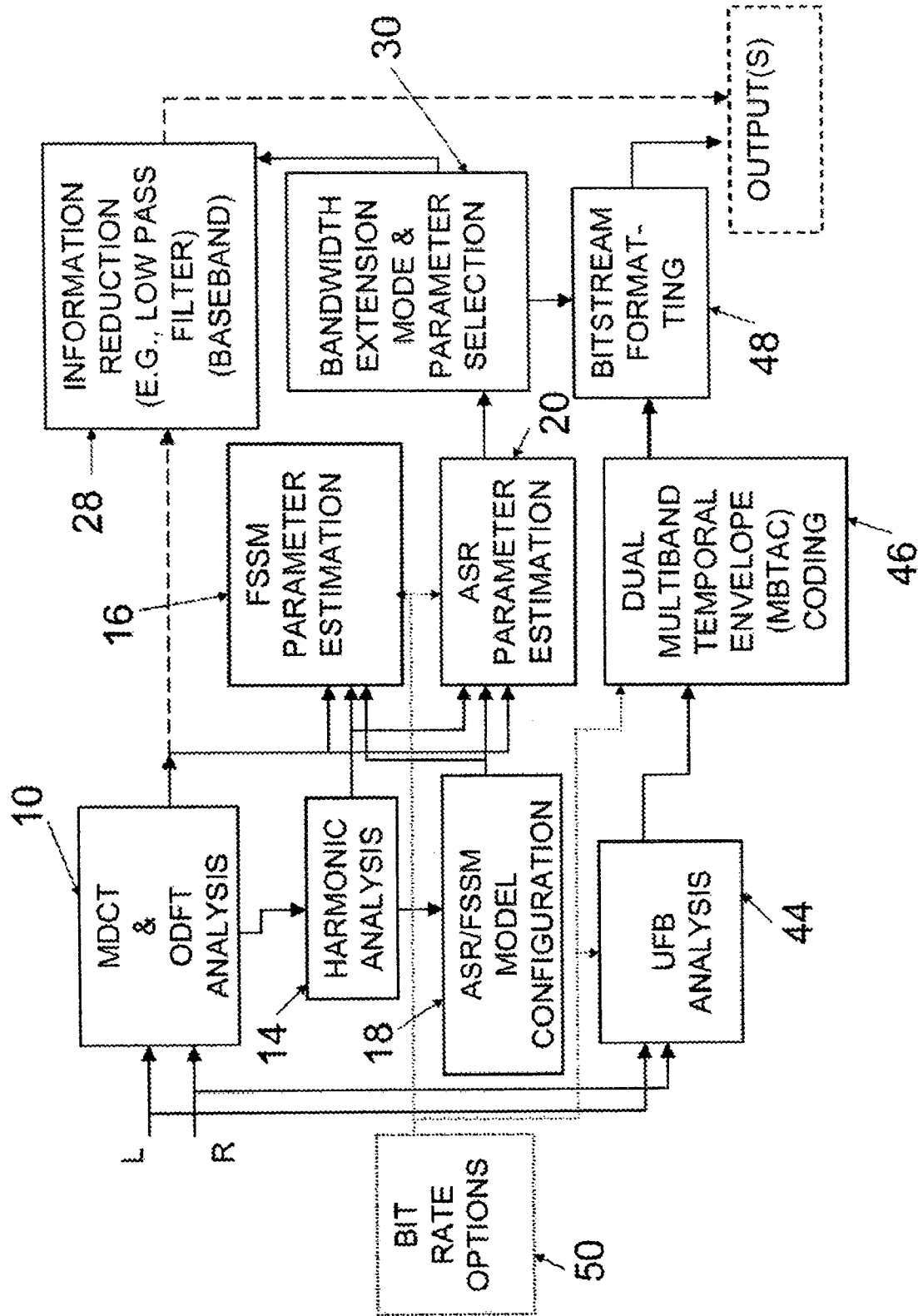


FIG 4

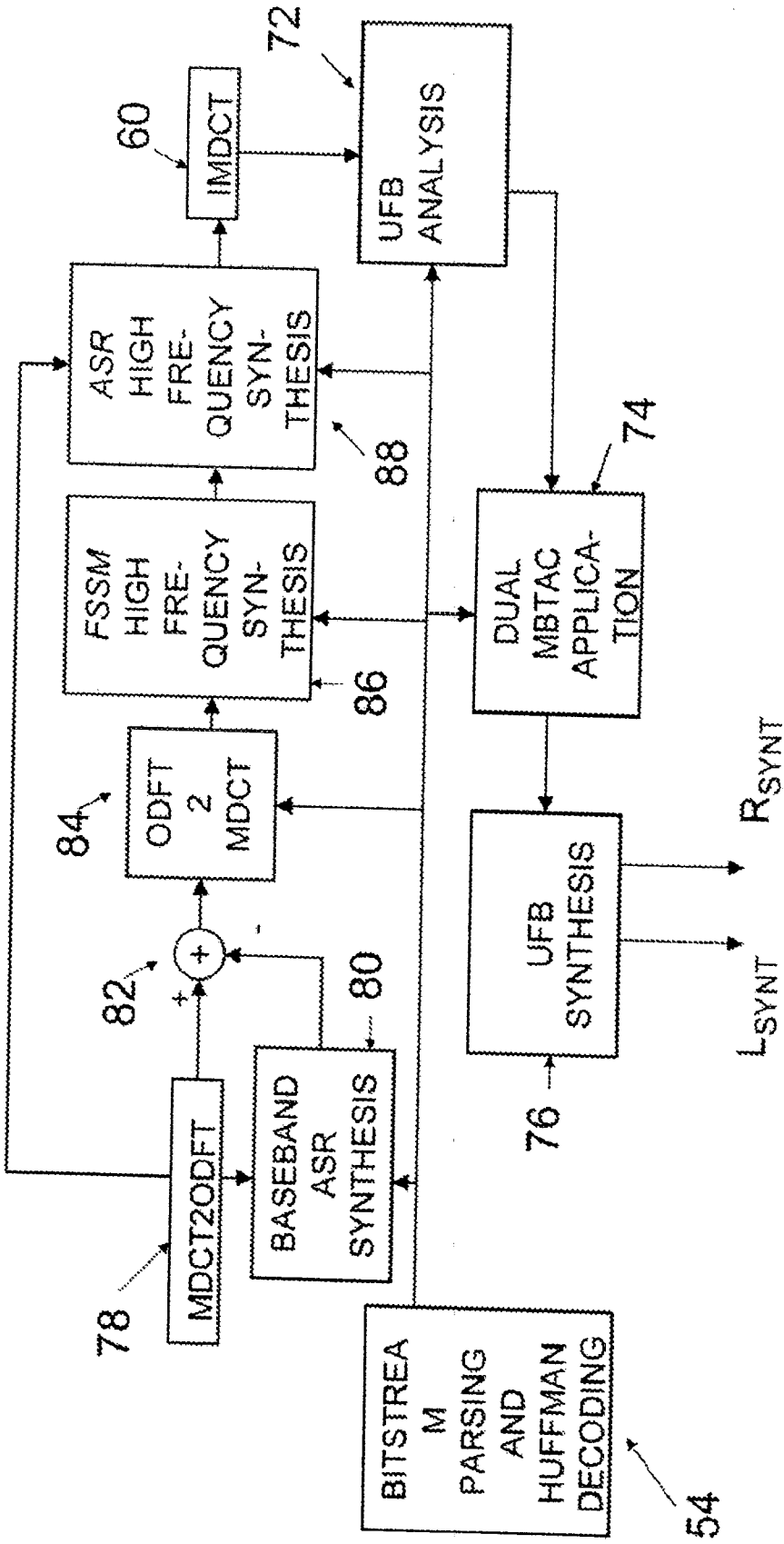


FIG 5

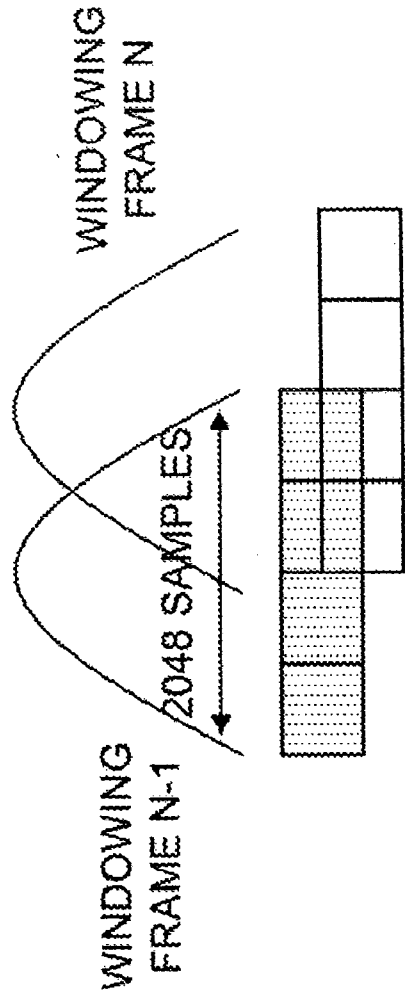


FIG 6

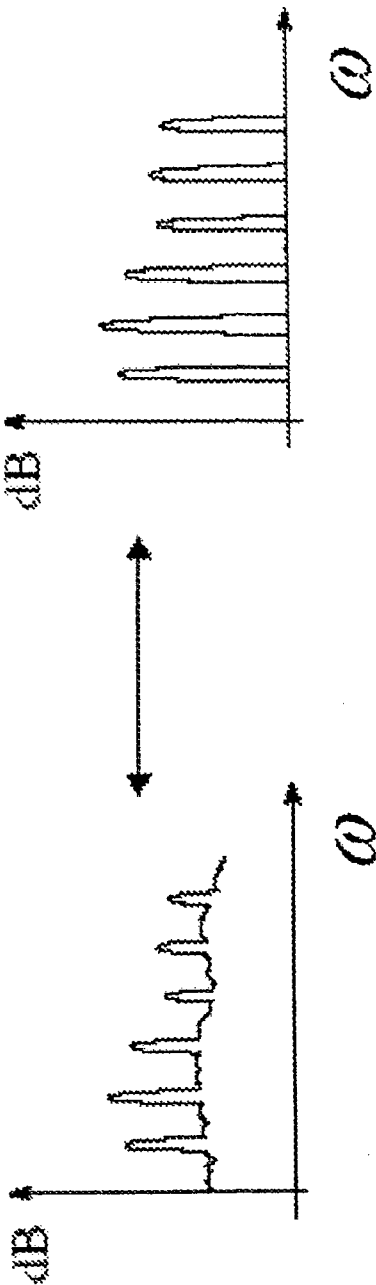


FIG 7

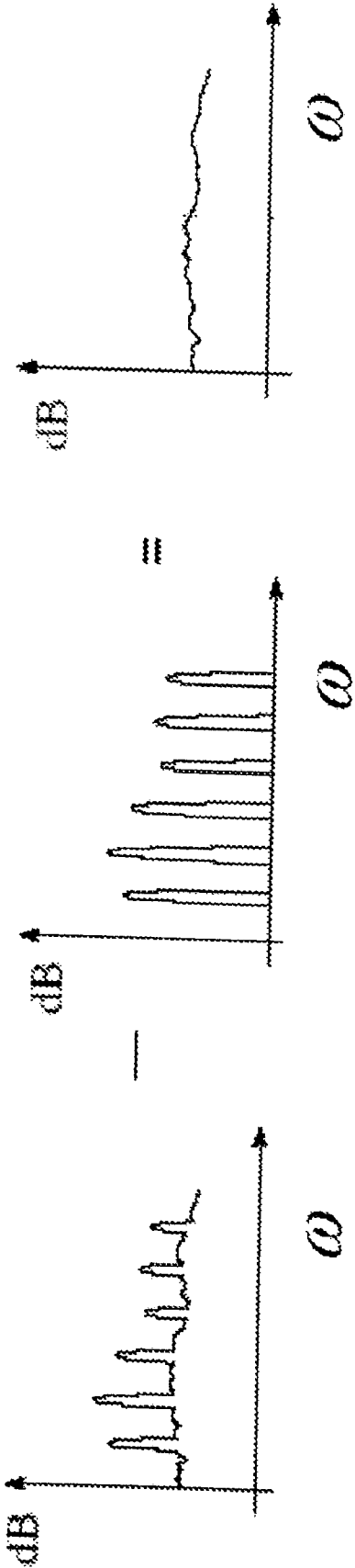
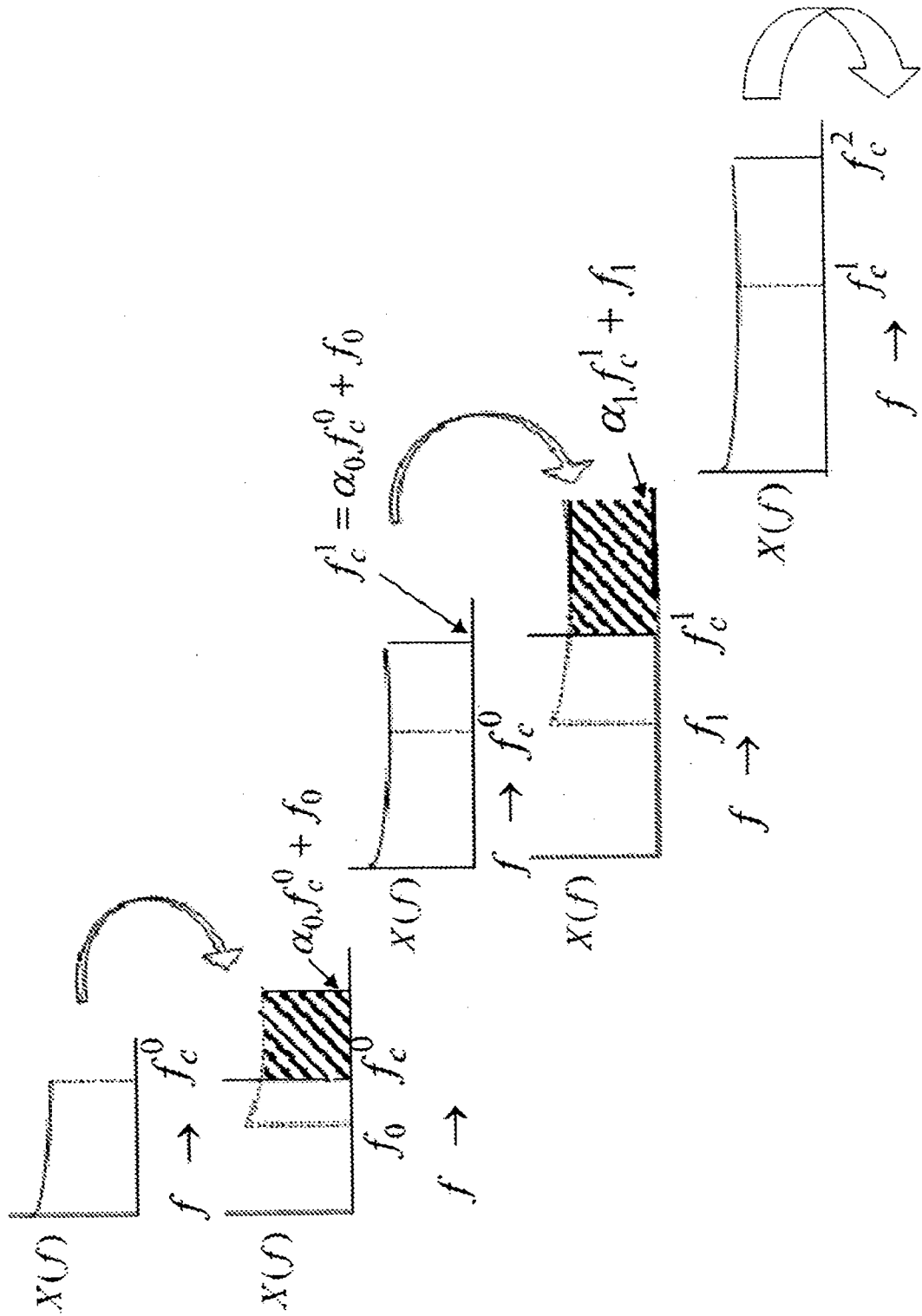


FIG 8



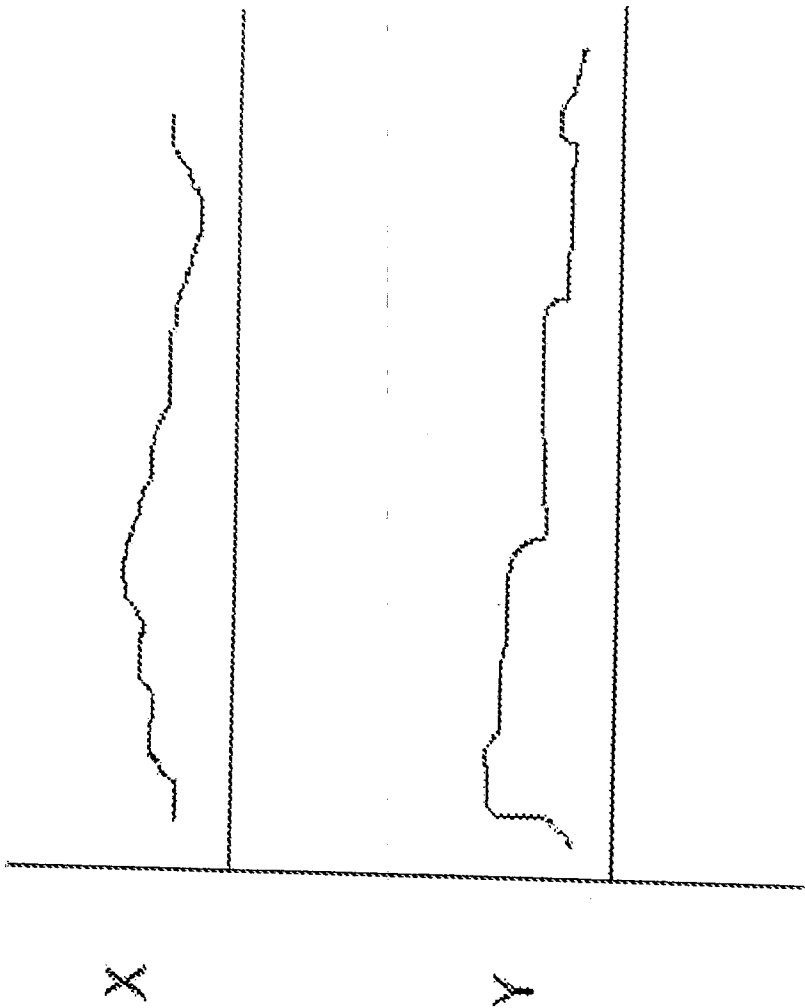


FIG 9

$$Y = \begin{bmatrix} y_1 & y_2 & y_3 & \cdot & \cdot & \cdot & y_n \end{bmatrix}$$
$$X = \begin{bmatrix} x_1 & x_2 & x_3 & \cdot & \cdot & \cdot & x_n \end{bmatrix}$$

FIG 10

$$\begin{matrix} \begin{bmatrix} x_{R1} & x_{R2} & x_{R3} & \cdot & \cdot & \cdot & x_{Rn} \end{bmatrix} \\ X_R = \end{matrix} \begin{matrix} \begin{bmatrix} x_{L1} & x_{L2} & x_{L3} & \cdot & \cdot & \cdot & x_{Ln} \end{bmatrix} \\ X_L = \end{matrix}$$

FIG 11

FIG 12

$$\bar{X}_L = \begin{bmatrix} (x_{L1} - x_{R1})/2 \\ (x_{L2} - x_{R2})/2 \\ (x_{L3} - x_{R3})/2 \\ \vdots \\ (x_{Ln} - x_{Rn})/2 \end{bmatrix}$$
$$X_R = \begin{bmatrix} (x_{L1} + x_{R1})/2 \\ (x_{L2} + x_{R2})/2 \\ (x_{L3} + x_{R3})/2 \\ \vdots \\ (x_{Ln} + x_{Rn})/2 \end{bmatrix}$$

FIG 13

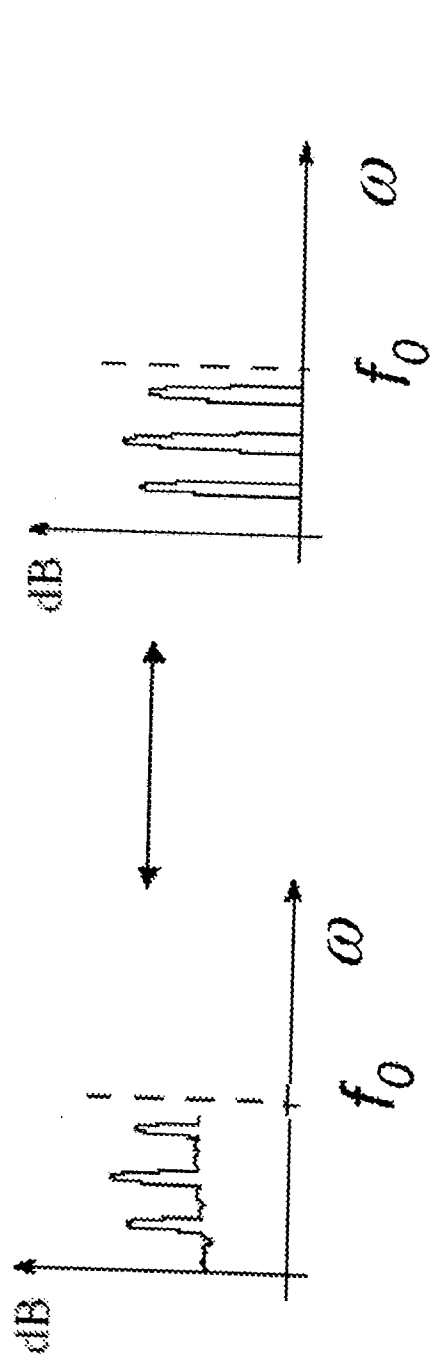


FIG 14

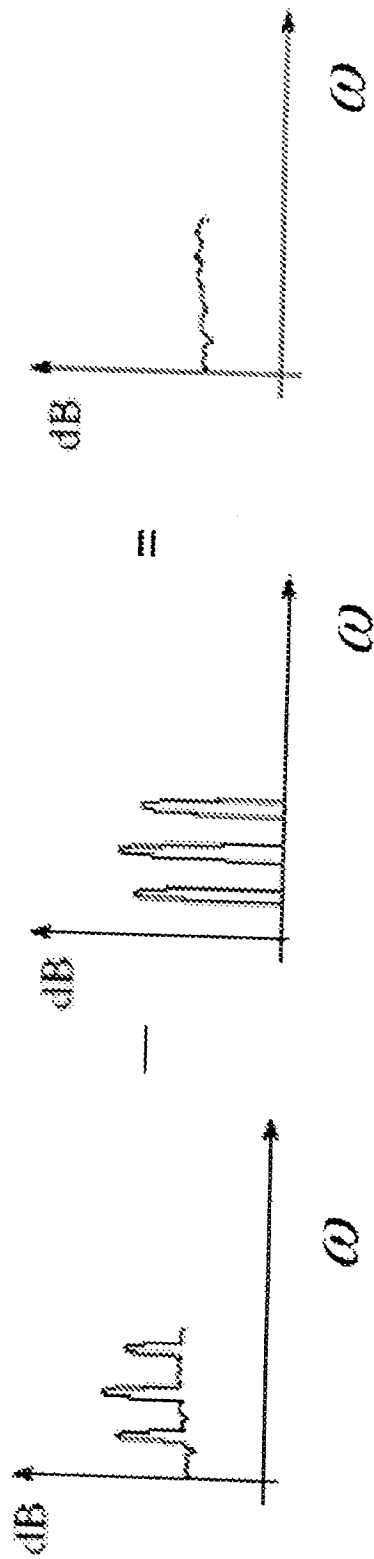
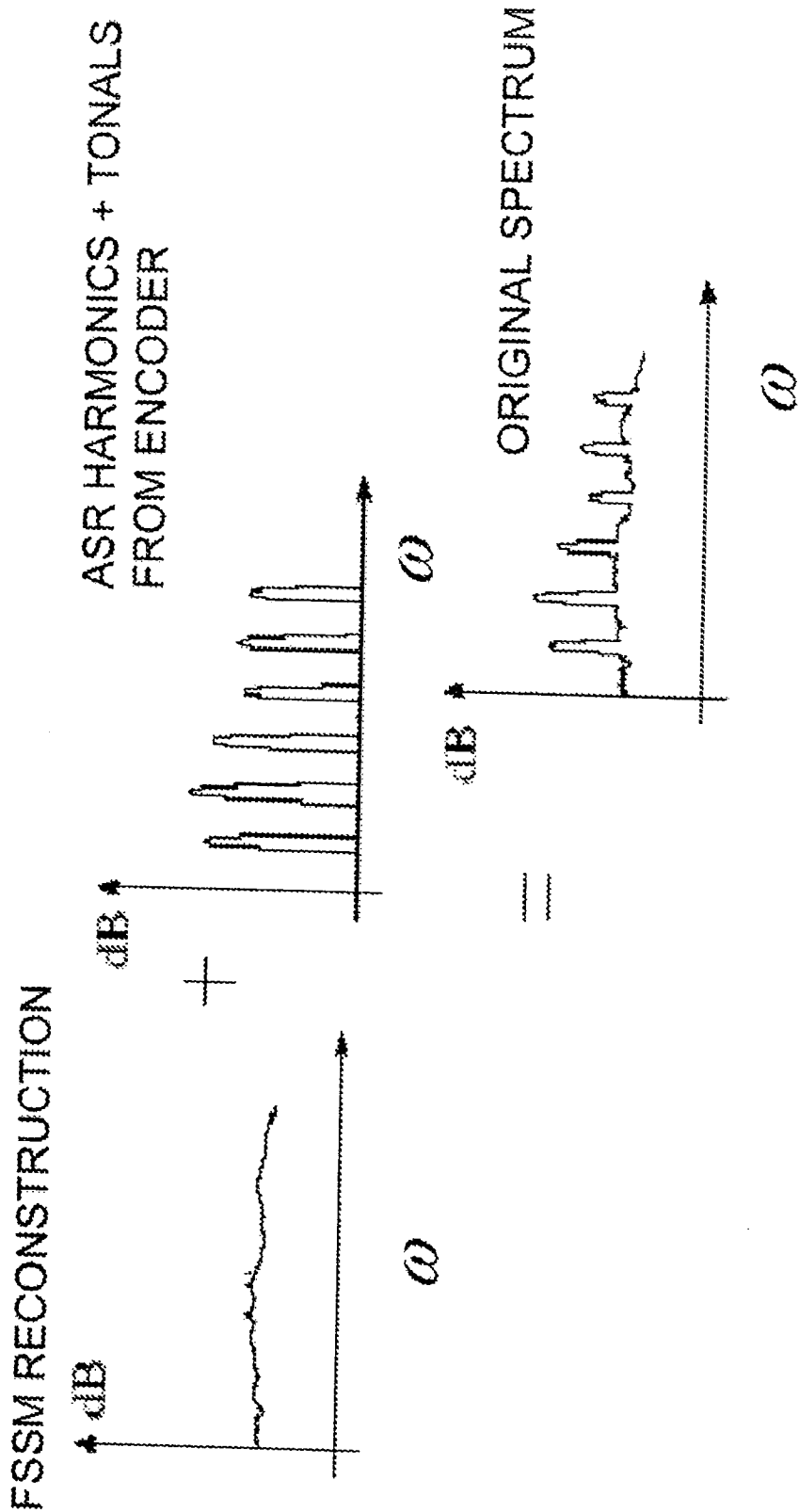


FIG 16



## METHOD AND APPARATUS FOR ENCODING AND DECODING

### CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application, Ser. No. 60/724,856, filed 7 Oct. 2005, the contents of which are hereby incorporated by reference herein.

### BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to coding and decoding of audio signals to reduce transmission bandwidth without unacceptably degrading the quality of the reconstructed signal.

[0004] 2. Description of Related Art

[0005] Many techniques exist in the field of audio compression for encoding a signal that can later be decoded without significant loss of quality. A common scheme is to sample a signal and use these samples to produce a discrete frequency transform. Varieties of transforms exist such as Discrete Fourier Transform (DFT), Odd-frequency Discrete Fourier Transform (ODFT), and Modified Discrete Cosine Transform (MDCT).

[0006] Also, transmission bandwidth can be conserved by sending only lower frequency (base band) spectral components. To restore the higher frequency components on the decoding side, various bandwidth extension techniques have been proposed. A simple technique is to take the base band components and scale them up in frequency.

[0007] Also, certain frequency components are difficult to perceive by the human ear when they are close in frequency to a dominant, high energy component. Accordingly, such dominant components can have associated with them a masking function to attenuate nearby frequency components, the attenuation being greater the closer a component is to the dominant masking component. Techniques of this type are part of the field of perceptual coding.

[0008] The field of perceptual coding for audio coding has been an active one over the past two decades. Typical configuration for the perceptual model used in audio codecs such as PAC, AAC, MPEG-LayerIII etc. may be found in [1-5]. The centerpiece of perceptual modeling is the concept of auditory masking [11-15, 27]. The goal is to quantize the audio signal in such a way that the quantization noise is either fully masked or rendered less annoying due to masking by the audio signal. Building of a perception model in audio codec typically involves the utilization of following four key concepts: simultaneous masking, temporal masking, frequency spread of masking, and, tone vs. noise like nature of the masker. Simultaneous masking is a phenomenon whereby a masker is found to mask the perception of a maskee occurring at the same time. Temporal masking refers to a phenomenon in which a masker masks a maskee occurring either prior to or after its occurrence. Frequency spread of masking refers to the phenomenon that a masker at a certain frequency has a masking potential not only at that frequency but also at neighboring frequencies. Finally, the masking potential of a narrow band masker is strongly

dependent on the tone vs. noise like nature of the masker. These factors are utilized to estimate desired quantization accuracy, or Signal to Mask Ratio (SMR) for each band of frequency.

[0009] In many audio codecs the masking model for wideband audio signals is constructed using a two step procedure. First the (short-term) signal spectrum is analyzed in multiple partitions (which are narrower than a critical band). The masking potential of each narrow-band masker is estimated by convolving it with a spreading function which models the frequency spread of masking. The masked threshold of the wide band audio signal is then estimated by considering it to be the superposition of multiple narrow band maskers. Recent studies suggest that this assumption of superposition may not always be a valid one. In particular a phenomenon called Comodulation Release of Masking (CMR) has implication towards the extension of narrow band model to a wide band model. B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th Ed., Academic Press, San Diego (2003). See Hall J W, Grose J H, Mendoza L (1995) Across-channel processes in masking. In: Hearing (Moore B C J, ed), pp 243-266. San Diego: Academic.

### SUMMARY OF THE INVENTION

[0010] In accordance with the illustrative embodiments demonstrating features and advantages of the present invention, there is provided a method for encoding an audio signal. The method includes the step of transforming the audio signal into a discrete plurality of (a) basic transform coefficients corresponding to basic spectral components located in a base band and (b) extended transform coefficients corresponding to components located beyond the base band. Another step is correlating that is (i) based on at least some of the basic transform coefficients and at least some of the extended transform components and (ii) performed by programmatically determining and applying a primary frequency scaling parameter and a primary frequency translation parameter to form a revised relation between the basic transform coefficients and extended transform coefficients that increases their correlation. The method also includes the step of forming an encoded signal based on the basic transform coefficients, the primary frequency scaling parameter and the primary frequency translation parameter.

[0011] In accordance with another aspect of the present invention, there is provided an encoder for encoding an audio signal that includes a processor, which has a transform, a correlator and a former. The transform can transform the audio signal into a discrete plurality of (a) basic transform coefficients corresponding to basic spectral components located in a base band and (b) extended transform coefficients corresponding to components located beyond the base band. The correlator can provide a correlation that is (i) based on at least some of the basic transform coefficients and at least some of the extended transform components and (ii) performed by programmatically determining and applying a primary frequency scaling parameter and a primary frequency translation parameter to form a revised relation between the basic transform coefficients and extended transform coefficients that increases their correlation. The former can form an encoded signal based on the basic transform coefficients, the primary frequency scaling parameter and the primary frequency translation parameter.

Oct. 11, 2007

[0012] In accordance with yet another aspect of the present invention, a method is provided for decoding a compressed audio signal signifying (a) basic transform coefficients of basic spectral components derived from a base band, (b) one or more frequency scaling parameters, and (c) one or more frequency translation parameters. The method includes the step of applying the one or more frequency scaling parameters and the one or more frequency translation parameters to the basic transform coefficients to provide a plurality of altered primary coefficients having altered spectral significance. Another step is inverting the basic transform coefficients and the altered primary coefficients to form a time-domain signal.

[0013] In accordance with still yet another aspect of the present invention, there is provided a decoder for decoding a compressed audio signal signifying (a) basic transform coefficients of basic spectral components derived from a base band, (b) one or more frequency scaling parameters, and (c) one or more frequency translation parameters. The decoder has a relocater for applying the one or more frequency scaling parameters and the one or more frequency translation parameters to the basic transform coefficients to provide a plurality of altered primary coefficients having altered spectral significance. The decoder also has an inverter for inverting the basic transform coefficients and the altered primary coefficients to form a time-domain signal.

[0014] In accordance with a further aspect of the present invention, a method is provided for encoding an audio signal. The method includes the step of transforming the audio signal into a discrete plurality of primary transform coefficients corresponding to spectral components located in a designated band. Another step is correlating based on a correspondence between at least some of the primary transform coefficients and programmatically synthesized data corresponding to a synthetic harmonic or individual sinusoids spectrum comprising any combination of one or more harmonic patterns and one or more individual sinusoids. The method also includes the step of forming an encoded signal based on at least some of the primary transform coefficients, and one or more harmonic parameters signifying one or more characteristics of the synthetic harmonic or individual sinusoids spectrum.

[0015] In accordance with another further aspect of the present invention, there is provided an encoder for encoding an audio signal. The encoder has a transform for transforming the audio signal into a discrete plurality of primary transform coefficients corresponding to spectral components located in a designated band. Also included is a correlation device for correlating based on a correspondence between at least some of the primary transform coefficients and programmatically synthesized data corresponding to a synthetic harmonic spectrum. The encoder also has a former for forming an encoded signal based on at least some of the primary transform coefficients, and one or more harmonic parameters signifying one or more characteristics of the synthetic harmonic spectrum.

[0016] In accordance with yet another further aspect of the present invention, a method is provided for decoding a compressed audio signal signifying (a) a plurality of basic transform coefficients corresponding to basic spectral components located in a base band, and (b) one or more harmonic parameters signifying one or more characteristics

of a synthetic harmonic or individual sinusoids spectrum comprising any combination of one or more harmonic patterns and one or more individual sinusoids. The method includes the step of synthesizing one or more harmonically related transform coefficients based on the one or more harmonic parameters. Another step is inverting the basic transform coefficients and the one or more harmonically related transform coefficients into a time-domain signal.

[0017] In accordance with still yet another further aspect of the present invention, there is provided a decoder for decoding a compressed audio signal signifying (a) a plurality of basic transform coefficients corresponding to basic spectral components located in a base band, and (b) one or more harmonic parameters signifying one or more characteristics of a synthetic harmonic or individual sinusoids spectrum comprising any combination of one or more harmonic patterns and one or more individual sinusoids. The decoder has a synthesizer for synthesizing one or more harmonically related transform coefficients based on the one or more harmonic parameters. Also included is an inverter for inverting the basic transform coefficients and the one or more harmonically related transform coefficients into a time-domain signal.

[0018] In accordance with still yet another aspect of the present invention, a method is provided for encoding an audio signal. The method includes the step of transforming the audio signal into a discrete plurality of transform coefficients corresponding to spectral components located in a designated band, some of the transform coefficients corresponding to one or more standard time intervals and others individually corresponding to one of a plurality of subintervals within the one or more standard time intervals.

[0019] Another step is forming an encoded signal based on (a) the plurality of transform coefficients associated with the one or more standard time intervals, and (b) magnitude information based on the plurality of transform coefficients associated with the plurality of subintervals.

[0020] In accordance with yet another aspect of the present invention, there is provided an encoder for encoding an audio signal. The encoder has a transform for transforming the audio signal into a discrete plurality of transform coefficients corresponding to spectral components located in a designated band, some of the transform coefficients corresponding to one or more standard time intervals and others individually corresponding to one of a plurality of subintervals within the one or more standard time intervals.

[0021] The encoder also has a former for forming an encoded signal based on (a) the plurality of transform coefficients associated with the one or more standard time intervals, and (b) magnitude information based on the plurality of transform coefficients associated with the plurality of subintervals.

[0022] In accordance with yet another aspect of the present invention, a method is provided for processing a decompressed audio signal obtained from a discrete plurality of transform coefficients corresponding to one or more standard time intervals, using magnitude information based on a plurality of transform coefficients corresponding to one of a plurality of subintervals of the one or more standard time intervals. The method includes the step of inverting the discrete plurality of transform coefficients associated with

the one or more standard time intervals into a first time-domain signal. Another step is successively transforming the first time-domain signal into a frequency domain to obtain a discrete plurality of local coefficients individually assigned to a plurality of successive time slots corresponding in duration to the plurality of subintervals. The method also includes the step of resealing the plurality of local coefficients using from the compressed audio signal the transform coefficients associated with the plurality of subintervals. Another step is inverting the discrete plurality of local coefficients into a corrected time-domain signal.

[0023] In accordance with yet another aspect of the present invention, there is provided a decoding accessory for processing a decompressed audio signal obtained from a discrete plurality of transform coefficients corresponding to one or more standard time intervals, using magnitude information based on a plurality of transform coefficients corresponding to one of a plurality of subintervals of the one or more standard time intervals. The accessory has a first inverter for inverting the discrete plurality of transform coefficients associated with the one or more standard time intervals into a first time-domain signal. Also included is a transform for successively transforming the first time-domain signal into a frequency domain to obtain a discrete plurality of local coefficients individually assigned to a plurality of successive time slots corresponding in duration to the plurality of subintervals. The accessory also has a rescaler for resealing the plurality of local coefficients using from the compressed audio signal the transform coefficients associated with the plurality of subintervals. Also included is a second inverter for inverting the discrete plurality of local coefficients into a corrected time-domain signal.

[0024] In accordance with another aspect of the present invention, a method is provided for encoding an audio signal. The method includes the step of transforming the audio signal into at least a discrete plurality of transform coefficients corresponding to spectral components located in a designated band, the transform coefficients including a standard grouping and a substandard grouping, the standard grouping being associated with one or more standard time intervals, the substandard grouping being dividable into a plurality of isofrequency sequences, each of the plurality of isofrequency sequences encompassing the one or more standard time intervals and being associated with a corresponding one of the transform coefficients in the standard grouping, the transform coefficients of the standard grouping each being assigned a masking characteristic for perceptually attenuating spectrally nearby ones of the standard grouping according to a predefined masking function having a predefined domain. Also included is the step of weakening the masking characteristic of each of the transform coefficients in the standard grouping based on the extent its corresponding one of the isofrequency sequences varies and correlates with spectrally nearby ones of the isofrequency sequences.

[0025] In accordance with another aspect of the present invention, there is provided an encoder for encoding an audio signal. The encoder has a transform for transforming the audio signal into at least a discrete plurality of transform coefficients corresponding to spectral components located in a designated band, the transform coefficients including a standard grouping and a substandard grouping, the standard grouping being associated with one or more standard time

intervals, the substandard grouping being dividable into a plurality of isofrequency sequences, each of the plurality of isofrequency sequences encompassing the one or more standard time intervals and being associated with a corresponding one of the transform coefficients in the standard grouping, the transform coefficients of the standard grouping each being assigned a masking characteristic for perceptually attenuating spectrally nearby ones of the standard grouping according to a predefined masking function having a predefined domain. Also included is a weakener for weakening the masking characteristic of each of the transform coefficients in the standard grouping based on the extent its corresponding one of the isofrequency sequences varies and correlates with spectrally nearby ones of the isofrequency sequences.

[0026] The present audio bandwidth extension (BWE) technique is based upon two algorithms, namely Accurate Spectral Replacement (ASR) and Fractal Self-Similarity Model (FSSM). The ASR technique is described in a paper by Anibal J. S. Ferreira and Deepen Sinha, "Accurate Spectral Replacement," *118th Convention of the Audio Engineering Society*, May 2005, Paper 6383, which paper is incorporated herein by reference. The FSSM and ASR technique are described in a paper by Deepen Sinha, Anibal Ferreira, and, Deep Sen "A Fractal Self-Similarity Model for the Spectral Representation of Audio Signals," *118th Convention of the Audio Engineering Society*, May 2005, Paper 6467; and Deepen Sinha, and Anibal Ferreira, "A New Broadcast Quality Low Bit Rate Audio Coding Scheme Utilizing Novel Bandwidth Extension Tools," *119th Convention of the Audio Engineering Society*, October 2005, Paper 6588 of the which papers are incorporated herein by reference.

[0027] The ASR and FSSM techniques work directly in the frequency domain with a high frequency resolution representation of the signal. These representations are supplemented by a third tool "Multi Band Temporal Amplitude Coding" (MBTAC), which ensures accurate reconstruction of the time-varying envelope of the signal representation in the frequency domain. The MBTAC tool utilizes a Utility Filterbank (UFB) that generates a frequency representation of the signal that varies in time with a relatively high time resolution to provide a time-frequency representation of the signal.

[0028] With the ASR technique the spectrum is segmented into sinusoids and residual (or noise), this residual results by removing (i.e., by subtracting) sinusoids directly from the complex discrete frequency representation of the audio signals from block 10. Coefficients for the sinusoids are coded and transmitted to the decoder.

[0029] The FSSM technique implements a bandwidth extension model employing the basic principle of creating a high frequency bandwidth from a low frequency spectrum. The model involves identifying dilation (frequency scaling) and frequency translation parameters which when applied on a low frequency band, efficiently represents the high frequency signal. Maximizing intra spectral-cross correlation is the basic criterion in choosing dilation and translation parameters. A brief functional description of FSSM's operation is as follows:

[0030] 1) The dilation and translation parameters are estimated and applied to the low frequency base band

to allow synthesis of a replica of the originally detected high frequency components.

[0031] 2) To determine the fit of the FSSM model, the frequency spectrum may be split into multiple slices and for each slice a determination is made to either apply the model or replace it by an independent signal such as synthetic noise. The FSSM model therefore, in general, is a FSSM+Noise model.

[0032] 3) The shape of the temporal and frequency envelope of the signal is an important consideration. The FSSM model may not accurately reconstruct the coarse frequency envelope and so this may coded separately.

[0033] In parallel to the above sequence of processes (which emanate from a high resolution frequency analysis), a second time-frequency analysis may be optionally performed and used to encode the time frequency envelope of the signal as well as the inter-aural phase cues. This sequence of parallel functional blocks is as follows:

[0034] A Utility Filterbank (UFB) is a complex modulated filterbank with several-times oversampling. It allows for a time resolution as high as  $16/F_s$  (where  $F_s$  is the sampling frequency) and frequency resolution as high  $F_s/256$ . It also optionally supports a non-uniform time-frequency resolution.

[0035] Multi Band Temporal Amplitude Coding (MBTAC) involves efficient coding of two channel (stereo) time-frequency envelopes in multiple frequency bands. The resolution of MBTAC frequency bands is user selectable. The envelope information is grouped in time and frequency and jointly coded (across two channels) for coding efficiency. Various noiseless coding tools are used to reduce bit demand.

[0036] The present disclosure also has a perceptual model employing psychometric data and results related to comodulation release of masking.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0037] The above brief description as well as other objects, features and advantages of the present invention will be more fully appreciated by reference to the following detailed description of illustrative embodiments in accordance with the present invention when taken in conjunction with the accompanying drawings, wherein:

[0038] FIG. 1 is a block diagram of an encoder implementing an encoding method in accordance with principles of the present invention;

[0039] FIG. 2 is a block diagram of a decoder implementing a decoding method in accordance with principles of the present invention;

[0040] FIG. 3 is a more detailed block diagram of portions of the diagram of FIG. 1;

[0041] FIG. 4 detailed block diagram showing enhancements to the diagram of FIG. 2;

[0042] FIG. 5 is a diagram showing the process of applying long windows on a stationary frame with the transform of FIG. 1;

[0043] FIG. 6 diagrammatically illustrates the detection of harmonic components in a base band as performed by the encoder of FIG. 1;

[0044] FIG. 7 diagrammatically illustrates on the left an original spectrum, detection of harmonics and tonals (central diagram) to produce a residual spectrum, using the encoder of FIG. 1;

[0045] FIG. 8 diagrammatically illustrates the FSSM process of extending/reconstructing bandwidth using nested iterations with the encoder of FIG. 1, as well as the decoder of FIG. 3;

[0046] FIG. 9 diagrammatically shows two waveforms X and Y that are to be correlated in accordance with a third level of frequency grouping performed by the encoder of FIG. 1;

[0047] FIG. 10 is a vector representation of the waveforms of FIG. 9;

[0048] FIG. 11 is a vector representation of right and left channels as employed in the encoder of FIG. 1 and a decoder of FIG. 2;

[0049] FIG. 12 is a vector representation of right and left channels employing sum and difference factors, which is an alternate to that shown in FIG. 11;

[0050] FIG. 13 shows detection of harmonic content from a base band as performed in the decoder of FIG. 2;

[0051] FIG. 14 diagrammatically illustrates on the-left an original low pass spectrum, harmonics detected therein (central diagram), to produce a flattened spectrum, using the decoder of FIG. 2;

[0052] FIG. 15 diagrammatically illustrates FSSM reconstruction as performed in the decoder of FIG. 2; and

[0053] FIG. 16 diagrammatically illustrates in the first two diagrams the spectra reconstructed by FSSM and by ASR, which are combined to produce an original spectrum (last diagram) using the decoder of FIG. 2.

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0054] Referring to FIGS. 1 and 3, an encoder apparatus and method is described by means of a block diagrams indicating certain algorithms performed on a processor, such as a microprocessor, dedicated computerized controller, personal computer, or more general purpose computer. Input to the system consists of sampled digital audio, specifically 16 or 24 bit PCM Stereo (Left/Right) with a sampling frequency  $F_s$  (illustrative values for  $F_s$  are 44100 Hz and 32000 Hz). High resolution frequency analysis (MDCT/ODFT) is performed in the first block 10 of the encoder/BWE encoder for one frame of audio (illustratively a frame of audio consists of 1024 samples). It simultaneously computes the MDCT and ODFT representation (for two channels L and R (left and right)). The MDCT/ODFT analysis is computed for two frequency resolutions: (i) a Long window which is typically 2048 samples long (with 1024 sample overlap between two consecutive windows), (ii) a Short window which is typically 256 samples long (with 128 sample overlap between two consecutive windows).

[0055] Block 10 is herein referred to as a transform for producing a plurality of transform coefficients (sometimes referred to as primary transform coefficients located in a designated band) indicating the magnitude or entity all discrete spectral components. These transform coefficients may be segregated into basic transform coefficients corresponding to basic spectral components located in a base band and extended transform coefficients that may not be directly encoded but may be simulated by the herein disclosed bandwidth extension method. The basic transform coefficients may be encoded and individually transmitted.

[0056] A window type detector 12 is applied to decide the window structure (Long/Short window) to be used to establish an input frame appropriate to avoid pre-echo condition; in other words, a trade off on time-frequency resolution is done based on the stationarity of the input frame. Specifically, detector 12 selects an increased time resolution (short window) for a non-stationary frame and an increased frequency resolution (long window) for a stationary frame. In case of a window state transition a well-known Start or Stop window is suitably inserted.

[0057] The present codec utilizes an algorithm for the detection and accurate parameter estimation of sinusoidal components in the signal. The algorithm may be based on the work by Anibal J. S. Ferreira and Deepen Sinha, "Accurate and Robust Frequency Estimation in ODFT Domain," in 2005 *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 16-19, 2005; and Anibal J. S. Ferreira, "Accurate Estimation in the ODFT Domain of the Frequency, Phase and Magnitude of Stationary Sinusoids," in 2001 *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 21-24 2001, pp. 47-50.

[0058] The detected sinusoids may be further analyzed for the presence of harmonic patterns using techniques similar to that described by Anibal J. S. Ferreira, "Combined Spectral Envelope Normalization and Subtraction of Sinusoidal Components in the ODFT and MDCT Frequency Domains," in 2001 *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 21-24 2001, pp. 51-54.

[0059] Depending on the chosen window (Long window=1024/Short window=128) MDCT and ODFT coefficients are calculated as graphically indicated in FIG. 5. The MDCT filter bank takes advantage of ODFT coefficients in that the results of the MDCT analysis filter bank can be decomposed from the results of the complex ODFT filter bank. The ODFT provides magnitude and phase information. The MDCT, ODFT and ODFT to MDCT transformation is as given below.

$$X_M(K) = \sum_{n=0}^{N-1} h(n)x(n) \cos \left[ \frac{2\pi}{N} (k + 1/2)(n + n_0) \right]$$

where  $X_M(K)$  is MDCT of the input sequence  $x(n)$  and  $h(n)$  the windowing function and  $n_0 = 1/2 + N/4$ .

[0060] Taking  $0 \leq K \leq N/2 - 1$  it can be shown that,

$$X_M(K) = \Re(X_0(K)) \cos \Theta(K) + \Im(X_0(K)) \sin \Theta(K)$$

Where,  $X_0(K)$  is ODFT of  $x(n)$  and

$$\Theta(K) = \frac{\pi}{N} (K + 1/2)(1 + N/2).$$

[0061] ODFT of a sequence  $x(n)$  is defined as:

$$X_0(K) = \sum_{n=0}^{N-1} h(n)x(n) e^{-j \frac{2\pi}{N} (K+1/2)n}$$

ODFT of two channels is computed using an efficient algorithm described in the TechOnline Paper "A Fast Algorithm for Computing Two Channel Odd Frequency Transforms with Application to Audio Coding" Sinha, N. and Ferreira J. S. TechOnline October 2005. The default window shape used in ODFT analysis is the sine window. Higher order smooth windows as described in the Sinha Ferreira Paper in AES 120<sup>th</sup> convention, NY may also be used for this analysis. In case of a Long to Short (Short to Long) transition the Long window immediately preceding (following) the Short window has a special non-symmetrical shape characterized as a Start (Stop) window. In such a case the ODFT/MDCT analysis is recomputed using the appropriate transition window shape.

[0062] The MDCT components thus produced are processed using a conventional stereo dynamic range control in block 26 before being bandwidth limited in block 28 for purposes to be described presently.

[0063] Thereafter, the magnitudes associated with the bandwidth limited components of the baseband are quantized in block 22. The quantizing steps can be adjusted dynamically in a manner to be described hereinafter.

[0064] Thereafter, entropy coding can be performed in block 24, which implements the well-known Huffman coding technique. Since the entropy coding can produce a time varying bit rate, a buffer is used in block 42, which is controlled by a rate control mechanism in block 40 in a conventional manner.

[0065] The final results of the processing in this main channel are forwarded to bitstream formatting block 48, which combines data from this channel with other data to form a bitstream having an appropriate transport protocol.

#### Psychoacoustic Model

[0066] The present codec includes a perceptual coding scheme whereby a sophisticated psychoacoustic model is employed to quantize the output of an analysis filter bank.

[0067] Two key aspects of the present psychoacoustic model pertains respectively to the extension of a narrow band masking model to wide band audio signals and to the accurate detection of tonal components in the signal.

[0068] In block 34 a conventional tonal analysis is performed and its results are forwarded to the quantizing control block 36, which connects to the quantizer 22 in the main channel.

[0069] Unlike conventional perceptual models, analysis is performed in block 32 taking into account comodulation release (CMR). Comodulation release is a phenomenon that suggests reducing conventional masking in the presence of a wide band (bandwidth greater than a critical band) noise-like signal which is coherently amplitude modulated (comodulated) over the entire spectrum range covered by a masking function. The reduction in masking has been variously reported to be between 4.0 dB to as high as 18 dB. See Jesko L. Verhey, Torsten Dau, and Birger Kollmeier "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation filter bank model" *Journal of the Acoustical Society of America*, 106(5), p. 2733-2745.

[0070] The exact physiological phenomenon responsible for CMR is still being investigated by various researchers. However, there is some evidence that CMR occurs due to a combination of multiple factors. It has been hypothesized that the masking release results from cues available within a critical band and from cues generated by comparisons across critical bands.

[0071] In audio codecs this implies that superposition of masking does not hold in the presence of strong temporal envelope and masking of wide band signals can be significantly lower than the sum of masking due to individual narrow (sub-critical) band components depending upon the coherence of their temporal envelopes. It is tempting to think that CMR can be accounted for through adequate temporal shaping of the quantization noise (since the masking threshold during the dips in envelope is very likely to be lower), but experiments indicate that (the lack of) temporal shaping of maskee does not explain all (or most) of the CMR phenomenon. In particular masking release of about 4-8 dB should be accounted for directly in the psychoacoustic model.

[0072] The present psychoacoustic model works with the short windows (substandard grouping) produced by block 10 so that some finer time variation is obtained about the temporal envelope for the frequency components of the critical bands (one or more isofrequency sequences formed from the substandard grouping). In this specification, the long windows are considered part of a standard grouping and are associated with one or more standard time intervals, where the isofrequency sequences encompass one or more standard time intervals.

[0073] A CMR model is incorporated which takes into account: (i) the effective bandwidth of the  $i^{\text{th}}$  critical band masker (masking value), EBM defined as

$$EBM_i = \frac{1}{2N} \sum_{j=i-N}^{j=i+N} \langle \phi_i, \phi_j \rangle \quad (1)$$

where  $\phi_i$  and  $\phi_j$  are respectively the normalized temporal envelopes of  $i^{\text{th}}$  and  $j^{\text{th}}$  critical band maskers (a suitable value for  $N$  is about 5); and, (ii) dip in the temporal envelope of the masker,  $P$  (having an individual value defined for each critical band as the peak to valley ratio between the minimum and maximum of the temporal envelope of the masker in a 20-30 msec window). Estimation for the reduced

masking potential of the narrowband masker,  $i$ , (CMR-COMP) is then made as below

$$CMRCOMP_i = -10 \log_{10} [p_i N(EBM_i)] \quad (2)$$

where  $N(\alpha)$  is a non-linearity and the CMRCOMP<sub>*i*</sub> value in (2) is saturated to a minimum of 0 dB (a piecewise linear function with a linear rise for a below 0.7 and above 0.8 and rapid rise angle of over 80° for a between 0.7 and 0.8 was found suitable in our experiments). Therefore, each narrowband masker is reduced in accordance with CMRCOMP<sub>*i*</sub>. Partial support for this model is based on data in Verhey et al., supra, and is supported by listening data based on expert listeners. The estimated CMR compensation is utilized when combining the masking effect of multiple bands.

[0074] Basically, the masking characteristic (a predefined masking function with a predefined domain) ordinarily assigned to transform coefficients of the standard grouping are weakened (with a weakener in block 32) based on the comodulation value, CMRCOMP<sub>*i*</sub>.

#### Bandwidth Extension

[0075] The transform coefficients from block 10 are to be segregated into basic transform coefficients in a low-frequency base band and extended transform coefficients located above the base band. The basic transform coefficients will be processed in a main channel as MDCT coefficients capable of representing a signal with relatively high fidelity (these are directly coded using either a conventional perceptual coding technique or its extensions described herein). Other parameters indicating qualities of the extended transform coefficients located beyond the base band.

[0076] Harmonic analysis block 14 (shown in FIG. 3 with an input coupled to block 10 and outputs coupled to blocks 16, 18 and 20) can detect all significant tonal components (magnitude above average) in the ODFT representation produced by block 10. These tonal components are further analyzed in block 14 to determine if these fit into a harmonic structure (the possibility of missing harmonics is allowed). Accordingly, block 14 acts as a correlation device when finding such harmonic structure. Accurate spectral replacement (ASR) model parameter estimation block 20 and fractal self-similarity model (FSSM) block 16 each has inputs coupled to blocks 10, 14 and 18.

[0077] ASR/FSSM Model Configuration block 18 has an input coupled to block 14. Block 18 can be configured (either permanently or based on user selected parameters) to issue control signal for specifying processing issues, such as processing order (ASR or FSSM first), components to be handled by ASR and FSSM, allowed number of harmonic patterns to be coded, bandwidth extension range, etc. See Table 1, which is discussed further hereinafter. Accordingly, FSSM block 16 and ASR block 20 will respond to this control signal and code accordingly the specified frequency structures (harmonics and tones).

[0078] While the present embodiment employs both an FSSM block 16 and ASR block 20, other embodiments may employ only one of them.

[0079] In ASR block 20 the spectrum is segmented into sinusoids and residual (or noise-like frequency components). This residual is created by removing (i.e., by subtracting) sinusoids directly from the complex discrete fre-

quency representation of the audio signals from block 10. Coefficients for the sinusoids are coded with sub-bin accuracy and transmitted to the decoder.

[0080] Referring again to accurate harmonic analysis block 14, this block identifies sinusoidal components from the input spectrum by identifying peaks in the fine structure of the spectral-envelope, harmonic structures present, if any, and strong high frequency (HF) tonals from the input spectrum produced by block 10. Identifying peaks in the fine structure of the spectral-envelope and strong HF tonals is a simple peak picking process. Detecting harmonic structure is a more complex process involving identification of relevant structures of sinusoids harmonically related in a way that is tolerant to local harmonic discontinuities. A condition for a harmonic structure to be recognized as such is that it contains at least four sinusoids. Alternatively, strong sinusoids not harmonically related may also be coded individually in case their spectral power exceeds a fraction of the total power of the audio signal. The results of detecting and separating harmonics and strong tonals is graphically illustrated in FIG. 6. Specifically, the algorithm of block 14 identifies the envelope of the spectrum. Spectral peaks and HF tonals are identified and a rough estimate of pitch is predicted from the envelope. Based on the rough estimate of pitch, harmonics with a maximum of 7 missing partials can be identified. In the process of identifying harmonic structure, pitch value is constantly updated on a per frame basis to match the original pitch of the spectrum.

[0081] Harmonic analysis block 14 models the input spectrum as a sum of harmonics plus noise-like components. The analysis involves identifying the harmonics to be removed from the spectrum. The analysis can be understood by considering the underlying signal as a spectrally spaced plurality of time-domain signals  $x(n)$  that can be viewed as sharply distinct harmonics among other components that fit within a smoother, almost noise-like spectrum, as follows:

$$x(n) = \sum_{n_1} \sum_k A_{n_1,k} \sin(n_1 \cdot f_k \pi + \phi_k) + \sum \text{noisy\_sinusoids}$$

where  $f_k$  is the fundamental frequency and  $\phi_k$  is the phase of the  $k^{\text{th}}$  harmonic; and  $n_1$  are the partials corresponding to a harmonic sequence (for a non-harmonic tone only one partial will be present). Harmonic analysis results in identification of values of  $A_{n_1,k}$ ,  $f_k$  and  $\phi_k$ . The spectrum remaining after removing the harmonics (and in some cases the tonal peaks) may be relatively flat and can be adequately represented by a flat (white noise) spectrum represented by a limited number of noise parameters indicating the envelope of a flattened noise spectrum. In other cases, the flattened spectrum will be subjected to analysis with the FSSM model.

[0082] While the harmonic components found among the coefficients produced by block 10 will normally be most efficiently handled by the ASR model of block 20, the ultimate choice of coding the harmonic structures using either ASR or FSSM is left to the user as a configuration parameter. If the user configures a flag in block 18 indicating a predominant FSSM mode, the strongest of existing harmonics structures is modeled by FSSM in a manner to be

described presently. In the absence of this flag both harmonics are modeled by the ASR algorithm of block 20.

[0083] Block 18 also assigns harmonics to the ASR block 20 and FSSM block 16 based on maximum allowable number of harmonics to be coded through ASR block 20, which is established either as a hard coded limit or as one modified by a user-defined parameter. Block 18 also resolves any overlap of frequencies between the tonals and harmonics for both the channels (Left/Right) and also resolves any overlap of frequencies between the channels' HF and harmonic structures.

#### ASR Parameter Estimation

[0084] ASR parameter estimation is performed in block 20, which generates parameters indicating the structure for certain harmonic and tonal values that are assigned to ASR processing by the model configuration block 18. These synthetically generated sinusoids are removed (subtracted) from the input spectrum of from block 10 to give a flattened spectrum that is graphically illustrated in FIG. 7. The high frequency tonals are also removed to further flatten the noise/residual of the input spectrum. Parameters indicating the foregoing harmonics and tonals (to be eventually used in reconstructive synthesis) are quantized and coded in blocks 22 and 24.

[0085] The foregoing assumed long windows. For short windows the tonal removal is done using a different approach. Until the transition frame, for a short window, tonals are removed using the parameters computed from the previous long window frame and after the transition frame, the tonal parameters from the from the future long window frame are used for synthesis.

[0086] For the purpose of ASR parameter estimation, the time-domain representation may be modeled as:

$$x(n) = \sum_{n_1} \sum_k A_{n_1,k} \sin(n_1 \cdot f_k \pi + \phi_k) + \sum \text{noisy\_sinusoids}$$

where  $x(n)$  is the time-domain representation of the original signal that was analyzed during harmonic analysis;  $f_k$  is the one or more fundamental frequencies and  $\phi_k$  is the phase of the  $k^{\text{th}}$  harmonic; and  $n_1$  are the partials corresponding to a harmonic sequence. Also, continuing with the time-domain representation yields

$$x_1(n) = \sum_{n_1} \sum_k A_{n_1,k} \sin(n_1 \cdot f_k \pi + \phi_k)$$

where  $x_1(n)$  is the proposed combination of synthetically generated harmonics which uses parameters identified by the harmonic analysis block 14.

[0087] Depending upon the bit rate and configuration of the codec, the phase parameter  $\phi_k$  may either not be used or used only at the "birth" of a harmonic sequence and then computed for the subsequent frames (e.g. long windows) using a "harmonic continuation algorithm". In addition,

$$y(n) = x(n) - x_1(n) = \sum \text{noisy\_sinusoids}$$

where,  $y(n)$  is the residual after the ASR parameter estimation block removes harmonics to yield a noise-like spectrum (note, for missing partials no removal is necessary and therefore the indicated subtraction will not actually occur). Removal of such harmonics or strong tonals is herein referred to as elimination of dominant ones of the basic transform coefficients in the base band. The coefficients to be removed are selected by determining whether their magnitude exceeds to give an extent the magnitudes in predefined neighborhoods (e.g., a predetermined number of dB greater than the average in a predefined guard band, such as  $\pm 4$  kHz).

[0088] Accordingly, the ASR technique results in an abbreviated list of parameters signifying one or more characteristics of a synthetic harmonic spectrum. In order to allow later reconstruction, each harmonic structure will be represented by (a) a fundamental frequency existing in the base band, the other harmonics being assumed to be integer multiples of that fundamental frequency, (b) an optional phase parameter related to either the fundamental or one of the harmonics in either the base band or the extended band, and (c) optional magnitude information. The magnitude information can be explicitly sent as a shape parameter indicating the declination of the harmonics from one harmonic to the next. Such shape is efficiently coded using signal normalization using a smooth spectral envelope model that can be estimated using conventional (Linear Predictive Coding) LPC-based techniques, cepstrum-based techniques or other appropriate modeling techniques; and is described by a compact set of parameters. In some embodiments no explicit magnitude information will be sent as part of the ASR process, but some magnitude tailoring will be accomplished as part of the MBTAC process described below.

#### FSSM Parameter Estimation:

[0089] The FSSM algorithm executed in block 16 includes a correlator categorizer and developer and is used for extension of bandwidth for higher frequencies based on low frequency spectrum values using the following programmatically determined and applied estimates of dilation and translation parameters. An introduction to the concept of FSSM is given followed by the functional implementation of FSSM in a BWE decoder.

[0090] The working of FSSM, described in detail, can be mathematically represented as a summation of terms with each having an iterative form, as indicated below:

$$R_{HP}(f) = \dots EO_i(\dots (EO_1(EO_0 X_{LP}(f)) \dots))$$

Where each expansion operator EO, is assumed to have the form:

$$EO_i X_{LP}(f) = H_i X_{LP}(\alpha_i f - f_i)$$

Where,  $\alpha_i$  is a dilation parameter ( $\alpha_i \leq 1$ ) and  $f_i$  is a frequency translation parameter (although in some embodiments dilation parameters greater than one may be employed).  $H_i$  is a high pass filter with a cut-off frequency

$$f_c^i = \alpha_i f_c^{(i-1)} + f_i$$

with  $f_c^0 = f_c$ . This sequence of nested expansion is graphically illustrated in FIG. 8 showing how an  $n^{\text{th}}$  composite band can be composed by adding beyond the prior  $((n-1)^{\text{th}})$  composite band, relocated coefficients lying in another higher band. Specifically, these relocated coefficients were relocated

using an  $n^{\text{th}}$  frequency scaling parameter  $n$  and an  $n^{\text{th}}$  frequency translation parameter  $f_n$  (i.e., an  $n^{\text{th}}$  adjusted pair). Note, the first composite band will be placed after frequency  $f_c$  and will proceed through  $M$  iterations (i.e.,  $M$  adjusted pairs and  $M$  composite bands).

[0091] Using the correlator in block 16, the values of  $\alpha_i$  and  $f_i$  are chosen to maximize the cross correlation between FSSM-representative spectrum and the original spectrum. Mathematically,  $\alpha_i$  and  $f_i$  are chosen such that,

$$\phi(\alpha_i, f_i) = \langle X(f), X(\alpha_i f - f_i) \rangle$$

with these two discrete spectra being correlated through, for example, a dot product. The correlation is maximized by programmatically adjusting the dilation and translation parameters:

$$\phi(\tilde{\alpha}_i, \tilde{f}_i) = \max \phi(\alpha_i, f_i), \forall \alpha_i \in A, f_i \in F$$

Where,  $A$  is a set of possible values for dilation parameter  $\alpha_i$  and  $F$  is the set of possible values for the translation frequency  $f_i$ . For the model to be meaningful for bandwidth extension, the range of  $A$  and  $F$  should be restricted such that  $\alpha_i f_c + f_i > f_c + C$ ,  $\forall \alpha_i \in A, f_i \in F$  for some suitably chosen minimum extension band  $C$ .

[0092] The foregoing self similarity model coherence maximization criterion works well in many cases. However, in certain instances special considerations need to be taken into account as listed below:

[0093] 1) In signals containing prominent harmonic structures the maximization criterion is not the best suited from a perceptual point of view. For such signals the presence of a harmonic structure as well as the fundamental frequency of the dominant harmonic can be accurately estimated. In most cases the translation parameter  $f_i$  best chosen as a value that ensures the continuity of the harmonic structure and the best value for the dilation parameter is close to unity.

[0094] 2) Because of the nature of the MDCT filterbank fluctuation in translation parameter  $f_i$  from one MDCT frame to the next can cause aliasing distortion, an "unsteady" perception for the high frequency harmonics may result. This is particularly true for signals for which a strong and steady smoothing or locking mechanism is necessary to avoid this problem.

[0095] After performing FSSM on the spectrum, the cross-correlation between spectral frequencies of the original spectrum from block 10 and the FSSM coded spectrum from block 10 is expected to be over a pre-defined threshold; if not, the FSSM parameters/results for the particular frames are discarded and the decoder generates instead synthetic noise with its envelopes following the RMS of the coded values. For a valid structure having valid dilation and translation parameters, the RMS values of the spectrum may be quantized and coded; or the magnitude shaping task may be left for the MBTAC processor-described below.

[0096] Accordingly, the output of block 16 is a sequence of ordered, adjusted pairs of frequency scaling parameters  $\alpha_i$  and frequency translation parameters  $f_i$  (the members 1,  $f_1$  of the first pair being referred to herein as a primary frequency scaling parameter and a primary frequency translation parameter). In most cases no magnitude information is included with this FSSM data because magnitude adjustment can be accomplished through the MBTAC process

described below. However, in some instances the MBTAC process is disabled in which case limited magnitude information may be sent with the FSSM data, although this magnitude information may be a coarse grouping of the relocated upper frequency bands created by the pairs  $\alpha_i, f_i$ .

[0097] These FSSM parameters are processed through selection block 30 together with the parameters produced by the ASR block 20 before being forwarded to block 48 (herein referred to as a former 48) where they are formatted into an appropriate transport protocol. It will be noted that the selection block 30 transmits the size of the extended band with two low pass filter block 28, which eliminates any high frequency components that are to be modeled by FSSM or ASR.

#### UFB and MBTA

[0098] To perform the task of shaping the temporal envelope of the reconstructed higher frequency components (in those cases when it is needed) we need to examine time trajectories of the spectral energy in multiple frequency bands. Furthermore, these time trajectories need to be examined at a time resolution that is substantially higher than that afforded by the high frequency resolution MDCT filterbank. For accurate temporal shaping for voiced speech and dynamic musical instruments a time resolution of 4-5 msec (or lower) is desirable. The desired temporal shaping can be computed by utilizing a separate higher time resolution "Utility Filter Bank" (UFB). It is desirable for the UFB to be a complex, over-sampled modulated filterbank because of several desirable characteristics of such filterbanks such as very low aliasing distortion. The magnitude of the complex output of the filterbank provides an estimate of the instantaneous spectral magnitude in the corresponding frequency band. Since UFB is not the primary coding filterbank its output may be suitably oversampled at the desired time resolution. Several options exist for the choice of the UFB. These include:

[0099] (a) Discrete Fourier Transform (DFT) with a higher time resolution (compared to MDCT): A DFT with 64-256 size power complementary window may be used in a sequence of overlapping blocks (with a 50% overlap between 2 consecutive windows)

[0100] (b) A complex modulated filterbank with sub-band filters of the form

$$h_i = h_0 \cdot e^{j\frac{\pi}{N}(i-1)n}$$

where  $h_0$  is a suitably optimized prototype filter. The DFT is a sub-class of this type of filterbanks. The more general framework allows for selection of longer windows (compared to the down-sampling factor).

[0101] (c) A complex non-Uniform filterbank; e.g., one with two or more uniform sections and transition filters to link the 2 adjacent uniform sections.

[0102] The exact choice of the UFB is application dependent. The complex-modulated filterbanks with a higher over-sampling ratio offer superior performance when compared to the DFT but at a cost of higher computational complexity. The non-uniform filterbank with higher fre-

quency resolution at lower frequencies is useful if envelope shaping at very low frequencies (1.2 kHz and lower) is desirable.

#### MBTAC

[0103] The functional requirement of MBTAC is to extract and code the temporal envelope (or time-frequency envelope) of the signal. Specifically, the signal envelope is analyzed in multiple frequency bands using a complex filterbank called a UFB. In a particular implementation of UFB shown herein as block 44, the signal is filtered in 128 uniform frequency sub-bands and each sub-band analysis is down sampled by a factor of 16.

[0104] In block 46 (which contains a categorizer and developer, as described further hereinafter) the over sampled signal, corresponding to a frame of input data (1024 samples) is arranged in a 2-D matrix of size 128x64 (128 frequency bands vs. 64 time samples). These 64 time samples are subintervals of the standard time interval for an MBCT frame (i.e., the MDCT timeframe is 64 times greater). Additional details regarding UFB may be obtained from the above noted reference, Deepen Sinha, Anibal Ferreira, and, Deep Sen "A Fractal Self-Similarity Model for the Spectral Representation of Audio Signals," 118th Convention of the Audio Engineering Society, May 2005, Paper 6467. It may also be noticed that due to the complex nature of the UFB output only the first 64 of the 128 frequency bands need to be analyzed.

[0105] The detailed time-frequency envelope generated by this process is grouped using a combination of one or more of the techniques described below, which constitute the categorizer of block 46. The bit rate requirement for coding and transmitting the (grouped) time-frequency envelope is further reduced using the techniques described immediately thereafter.

#### First Level Time-Frequency Envelope Grouping

[0106] The initial, finely partitioned, time-frequency envelope is first grouped by assigning UFB sub-bands to N critical ordered frequencies so-bands (each critical band may be a partition using the well-known concept of Bark bands, each containing one or more of the UFB bands). Furthermore, several adjacent time samples are grouped into a single time slot. For the purpose of this time grouping, the system uses either 8 or 16 adjacent UFB time samples. Therefore, the 64 time samples in a frame get arranged into M ordered time slots, here either 8 or 4 time slots. As an illustrative example, assuming there are 17 critical bands between 0 and  $F_s/2$  ( $F_s$  being the sampling frequency) after this level of frequency/time grouping, the result is a still relatively fine  $N \times M$  matrix of 17x8 or 17x4 RMS envelope values (instead of a 128x64 finely detailed envelope). This  $N \times M$  matrix has a corresponding frequency index and subinterval index and forms an  $N \times M$  group index. A "base band" envelope is also computed by averaging across the critical bands between 1kHz and 3.5kHz. This base band envelope may be used in a subsequent, optional grouping technique described below (third level frequency grouping).

[0107] If no higher level of grouping is performed (i.e. Second Level or Third Level Grouping as described below) coefficients having the same index (from the  $N \times M$  group index) will be merged using the developer of block 46 to

form indexed proxies signifying, for example, the average magnitude of members of the group (an effective recoding with a recoder).

### Second Level Time-Frequency Envelope Grouping

[0108] The RMS coded time-frequency envelope after the first level of grouping may optionally be grouped through a second level into consolidated collections that combine adjacent envelopes (adjacent in both time and frequency).

[0109] Time grouping is first done on each of the M time indices, with successive time slots being grouped if the difference between maximum-minimum RMS values in each frequency sub-band are within a predetermined limitation on magnitude variation (although sub-band to sub-band differences may be rather large). This grouping is performed over the time slots iteratively until reaching that index where, the latest RMS values cause the calculated difference between the maximum and minimum RMS values in the growing collection of time-grouped values to exceed a threshold in at least one frequency sub-band, in which case this latest time slot is not added to the growing collection. Once closed, all the time-grouped values within this collection are replaced with a single RMS averaged value, one for each frequency sub-band.

[0110] As the time grouping above and below transition bands might differ in the first level of grouping, based on the preset values, the second level of grouping is done separately above and below the transition band.

[0111] The above mentioned time grouping technique is followed with frequency grouping. In particular all of the time groups are evaluated to determine if all time groups can be partitioned with the same frequency breaks to form, two or more common frequency groups where in each frequency group (and in all time groups) the difference between the greatest and the smallest RMS value falls within a pre-specified frequency grouping limit. As before, the averaged RMS value of frequency groups is calculated to replace the grouped values, which then become indexed proxies replacing those of the first grouping.

[0112] This grouping is performed so that each of the consolidated collections do not exclude any one of the indexed proxies that intervene by aligning on a common row or common column (of the N×M group index) contained in the collection. For each of the consolidated collections the encoded signal will include information based on the gross characteristics of the consolidated collection.

### Third level Frequency Grouping

[0113] Unlike the other two grouping techniques this is done only on frequency envelope. The technique exploits the correlation between the frequency grouped values. The second level of grouping encompasses only those waveforms which are closer in RMS value to their neighbors; this grouping is done depending on the correlation of grouped frequency values. In this technique the time envelopes in each of the higher frequency bands (critical bands or grouped critical bands constituting higher temporal sequences) is analyzed for closeness to the baseband envelope (a pilot sequence having M temporally sequential values developed from one or more of the lower ones of the N ordered frequency sub-bands) computed in the first grouping. If the "shape" of the envelope is close to the shape in

the baseband envelope, only a scaling factor is transmitted (instead of the detailed envelope shape).

[0114] The following gives an algorithmic description of this grouping technique and computation of the scaling factor.

[0115] FIG. 9 illustrates two monochromatic waveforms X and Y (two vectors, each representing a plurality of time slots in only one respective frequency sub-band). The criterion of this coding technique is to code X and a correlation parameter 'a' (a scalar) into the bit-stream if and only if the reconstructed waveform, say  $Z=aX$ , at the decoder and Y are highly correlated else X and Y are coded separately. Reconstruction of Y, from X and 'a' is done at the decoder. In this process, the criterion of importance is a suitable value for 'a' such that the distance/error in the process of reconstruction is as small as possible.

[0116] To find a value 'a' such that the "distance" between 'aX' and Y is as small as possible, following procedure is used. With distance as the criterion consider X and Y monochromatic vectors as shown in FIG. 10. Distance between Y and 'aX' is given by the following separation function:

$$D=(y_1-ax_1)^2+(y_2-ax_2)^2+\dots+(y_n-ax_n)^2$$

To minimize the value of distance (D) with respect to 'a', begin by differentiating D with respect to 'a' and equate it to zero.

$$\begin{aligned}\frac{dD}{da} &= 2(y_1-ax_1)*(-x_1)+2(y_2-ax_2)* \\ &\quad (-x_2)+\dots+2(y_n-ax_n)*(-x_n) \\ &= 0\end{aligned}$$

Realigning the above equation,

$$a = \frac{x_1 \cdot y_1 + x_2 \cdot y_2 + \dots + x_n \cdot y_n}{x_1^2 + x_2^2 + \dots + x_n^2}$$

From the above calculated value of 'a' maximum dB difference between the original (Y) waveform and the reconstructed waveform ( $Z=aX$ , that is, programmatic changing of scale) is compared with a predetermined threshold and a decision either to code and transmit as part of encoder bitstream X and Y individually or to code only X and the distance parameter 'a' is made. Coding of grouped values:

[0117] The above Time-Frequency grouped values are efficiently coded based on a comparative analysis based on bit demand. There are four different ways of differential coding (recoding) the above grouped Time-Frequency envelope, based on the adjacency along the ordered frequency sub-bands and ordered time slots, defined as follows:

(a) Time-Frequency Differential Coding

[0118] In this method, every element of the two dimensional matrix say,  $N_{ij}$  is time, frequency and time-frequency differentially coded (i.e)

$$N_{ij}=N_{i,j}-(N_{i-1,j}+N_{i,j-1}-N_{i-1,j-1})$$

where  $N_{i,j}$  represents the value in the Time-Frequency matrix at  $i^{th}$  frequency and  $j^{th}$  time instant.

(b) Time Differential Coding

[0119] In this method, every element of the two dimensional matrix say,  $N_{i,j}$  is only time differentially coded (i.e)

$$N_{i,j}=N_{i,j}-N_{i,j-1}$$

(c) Frequency Differential Coding

[0120] In this method, every element of the two dimensional matrix say,  $N_{i,j}$  is only time differentially coded (i.e)

$$N_{i,j}=N_{i,j}-N_{i-1,j}$$

(d) No Differential Coding

[0121] As the name suggests, no differential coding is done and the individual values are quantized and Huffman coded.

[0122] All the above schemes are compared based on their bit demand and the one with the least bit demand is chosen to code the Time-Frequency envelope. This coding produces at plurality of utility coefficients signifying the magnitude for a specific time-frequency coordinate.

[0123] The above coding scheme applies equally both for a stereo and a mono file. the above coding schemes are applied to individual images on a stereo file. In addition to the above coding method stereo files are R-L diff coded, to lower the bit demand. In a stereo file R-L diff coding is performed first followed by any of the above coding schemes.

[0124] R-L differential coding exploits the temporal similarity of the left and right image of a stereo waveform. In this coding technique Left and Right images are differenced and halved and is stored as the new Left image of the stereo audio and the Left and Right images (from the original audio) are averaged and stored onto the Right image. See

FIG. 10 where,  $X_L$  and  $X_R$  are Left and Right images of the stereo audio. The new values of Left and Right vectors are given in FIG. 11 where,  $X_L$  and  $X_R$  are new images of Left and Right channel after applying R-L differential coding.

[0125] Table 1 shows five default configurations (modes) controlling the assignment of tasks between the FSSM and ASR model as well as a corresponding adjustment in the role of the MBTAC process. It will be noted that the modes are listed in descending transmission bit rate (second column). Also, the top three modes (ST1 through ST3) use a bandwidth expansion range that is 50% of the overall bandwidth (half the sampling frequency  $f_s$ ) produced by the analysis block 10 (FIG. 1), while the last two modes may operate with a larger expansion range.

[0126] In mode ST1 the ASR model handles secondary harmonics and isolated tones. In mode ST2 the ASR model handles tonal components. In modes ST3 and ST4, the ASR model handles isolated tones. In mode M1 there is no ASR model functioning. In each of these modes, components that are not handled by the ASR model are handled by the FSSM model.

[0127] In modes, ST1, ST3, and ST4 full MBTAC compensation is provided down to the indicated frequency and will handle both the right and left channel (full stereo). In modes ST3 and ST4 MBTAC compensation will be provided to even lower frequencies but this compensation will only correct the ratio (dB difference) between the right and left channels (or equivalently the sum and difference channels often used to represent two stereo channels). In mode ST2 essentially MBTAC compensation is absent and instead some magnitude information is sent along with the ASR and FSSM data indicating magnitudes at least for certain frequency bands. Finally, in monaural mode M1 the MBTAC operates down to 2 kHz. It will be noted that bit rate options are provided to blocks 16, 20, 44 and 46 by a user-controlled options block 50.

TABLE 1

| FSSM/ASR/MBTAC DEFAULT CONFIGURATIONS |                                            |                            |                                                                                                         |                                                                                  |
|---------------------------------------|--------------------------------------------|----------------------------|---------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Mode Name                             | Intended Bit Rate Range & Application Type | Bandwidth Extension Range  | FSSM/ASR Configuration                                                                                  | MBTAC/Parametric Stereo Configuration                                            |
| ST1                                   | 45-56 kbps (or higher) - Broadcast         | 50% of bandwidth           | FSSM for dominant harmonic and non-harmonic components<br>ASR for secondary harmonic and isolated tones | Full Stereo MBTAC from 6 kHz                                                     |
| ST2                                   | 40-72 kbps - Low Complexity                | 50% of bandwidth (or less) | ASR for all tonal components<br>FSSM for non-tonal components                                           | Frequency shape only                                                             |
| ST3                                   | 36-42 kbps                                 | 50% of bandwidth           | FSSM for harmonic and non-harmonic components<br>ASR for isolated tones                                 | Full Stereo MBTAC from 8 kHz<br>Differential (stereo) MBTAC from 2 kHz           |
| ST4                                   | 24-36 kbps                                 | 50-75% of bandwidth        | FSSM for harmonic and non-harmonic components<br>ASR for isolated tones                                 | Full MBTAC starting from 4-8 kHz<br>Differential MBTAC starting from 250-2000 Hz |
| M1                                    | 12-24 kbps (Mono) or lower                 | 75% of bandwidth           | FSSM for harmonic and non-harmonic components                                                           | Mono MBTAC from 2 kHz                                                            |

## BWE Decoder

[0128] Referring to FIG. 2, this block diagram indicates an algorithm that may be executed by various types of processors and computers. The bitstream 52 is subjected to inverse quantization and entropy decoding in block 54 to recover MDCT coefficients 56. These MDCT coefficients are also forwarded to blocks 62, 64, and 66 together with recovered ODFT coefficients. In some embodiments, the low pass spectrum 56 is analyzed for harmonic structure that ought to be removed before being passed to the FSSM process (because a band extension model such as ASR will be responsible for harmonic replication in the extended band).

[0129] FSSM reconstruction in block 62 is applied on a spectrum that was flattened at the encoder (FIG. 1) by the removal of certain harmonic components, as noted above. On applying dilation and translation parameters with spectral norm values, a high frequency flattened spectrum is approximately reconstructed by the FSSM technique.

[0130] ASR reconstruction at the decoder in block 64 involves synthesizing (with a synthesizer in block 64) the harmonic structure and high frequency tonals contained in the encoded information from block 54. The synthesized sinusoids are processed in block 68 (being converted from ODFT to MDCT) and combined in harmonization block 70 with the FSSM full band spectrum from block 62 before being sent to summation node 58. Also, information from decoder block 54 indicating the desired shape of a synthetic noise spectrum is also combined in node 58 with the FSSM and ASR components from block 70 to reconstruct the original spectrum. In block 60 the MDCT coefficients are inverted into a time-domain signal.

[0131] In addition, MBTAC parameters passed from block 54 to compensation blocks 72 and 74 (having a inserter and restorer) ensure that the temporal envelope of the original signal is maintained after the reconstruction from the bandwidth extension technique. Adjustment of this temporal envelope is performed in blocks 72, 74, and 76.

## MDCT to ODFT Transformation

[0132] Returning again to block 54, an MDCT to ODFT transformation proceeds as follows:

[0133] The coefficients of an MDCT filter bank can be decomposed as complex ODFT filter bank. The ODFT representation provides magnitude and phase information. MDCT to ODFT and ODFT to MDCT transformation is as given below:

$$X_M(K) = \sum_{n=0}^{N-1} h(n)x(n) \cos \left[ \frac{2\pi}{N} (k + 1/2)(n + n_0) \right]$$

where  $X_M(K)$  is MDCT of the input sequence  $x(n)$  and  $h(n)$  the windowing function and  $n_0 = 1/2 + N/4$ .

[0134] Taking  $0 < K < N/2-1$  it can be shown that,  $X_M(K) = 9(X_0(k)) \cos E(K) + 3(X_0(K)) \sin O(K)$  where,  $X_0(K)$  is ODFT of  $x(n)$  and

$$\theta(K) = \frac{\pi}{N} (K + 1/2)(1 + N/2).$$

[0135] ODFT of a sequence  $x(n)$  is defined as,

$$X_0(K) = \sum_{n=0}^{N-1} h(n)x(n) e^{-j \frac{2\pi}{N} (K+1/2)n}$$

Similarly, a transformation of the MDCT domain to the aliased ODFT domain can be obtained by computing:

[0136]  $X_0(k) = 2 [X_M(K) * \cos O(K) + j X_M(K) \sin O(K)]$  where  $j = \sqrt{-1}$  and

$$\theta(K) = \frac{\pi}{N} (K + 1/2)(1 + N/2).$$

Aliasing iss cancelled in the overlap-add operation following inverse ODFT computation:

$$x(n) = \sum_{k=0}^{N-1} X_0(K) e^{j \frac{2\pi}{N} (K+1/2)n}$$

with  $0 < n < N-1$  and  $X_0(K) = X^*(N-1-K)$  with  $0 < K < N/2-1$ . ASR Analysis:

[0137] The purpose of this ASR analysis at the decoder is to create a cleaner baseband from which FSSM synthesis described below can proceed. This aids in avoiding interference between FSSM synthesized components and ASR synthesized components when both the models are in use. Referring to FIG. 4, the foregoing FSSM and ASR synthesis (blocks 62 and 64 of FIG. 2) is revised to provide more accurate reconstruction. (Although components in FIGS. 2 and 4 having the same reference numerals have the same structure and perform the same functions.) The incoming MDCT components are converted in block 78 to ODFT components and applied to the summing input of combining node 82. Also, the ASR data from the coding block 54 is analyzed in block 80 to determine a pattern of ODFT components that are to be ultimately used in the ASR model of reconstruction. The thus identified ODFT components are applied to the subtracting input of combining node 82 to remove from the base band harmonics and tonals that will be processed by the ASR model. See FIG. 13. In fact this harmonic structure is removed from the incoming spectrum to create a flattened spectrum suitable for FSSM reconstruction. See FIG. 14.

[0138] The content of the ODFT spectrum lock 78 may be thought of as a signal, which if converted to the time-domain, would be represented as follows:

$$x_{\text{lowpass}} = \sum_{n_1} \sum_{k=1,2} A_{n_1,k} \sin(n_1 \cdot f_k \pi + \phi_k) + \sum \text{noisy\_sinusoids}$$

where,  $x_{\text{lowpass}}$  is the lowpass, time-domain signal of interest and  $n_1 K_1/2 < f_0 \leq n_1 K_1$ . ASR processing in block 80 involves identifying the values of  $A_{n_1,k}$ ,  $n_1$  and  $K_1$ . Also,  $f_0$  in the above inequality is the cut-off frequency of the spectrum. Upon identifying the harmonics, node 82 eliminates the harmonics in order to smooth the spectrum to one suitable for FSSM processing. In the time-domain, this smoothing process can be considered

$$\begin{aligned} y_{\text{lowpass}}(n) &= x_{\text{lowpass}}(n) - \sum_{n_1} \sum_{k=1,2} A_{n_1,k} \sin(n_1 \cdot f_k \pi + \phi_k) \\ &= \sum \text{noisy\_sinusoids} \end{aligned}$$

After this smoothing process, the ODFT components are converted back to MDCT components in block 84. FSSM Reconstruction:

[0139] The flattened low pass spectrum is now extended using FSSM's adjusted pairs of dilation and translation parameters,  $1, f_1$ , which were extracted from the bitstream in decoder block 54 and sent to FSSM synthesizer block 86, which includes a relocater. The concept of reconstruction of FSSM from a low band signal is illustrated in FIG. 8.

[0140] Specifically, the spectral components in the MDCT base band are multiplied by a first dilation (frequency scaling) parameter 1 and then shifted by a first frequency translation parameter  $f_1$ . All relocated components (such relocated components being referred to as altered coefficients were altered primary coefficients) that fall beyond the base band are used to create a first FSSM reconstructed sub-band, which is added to the base band to form a first composite band. This first composite band is then subjected to a second dilation parameter 2 before being shifted by a second frequency translation parameter  $f_2$ . All components relocated thereby (by the relocater) that fall beyond the first composite band are used to create a second FSSM reconstructed sub-band; which is added to the first composite band to form a second composite band. This process is repeated iteratively for all remaining adjusted pairs of dilation parameter and frequency translation parameters to create the FSSM extended band through a growing sequence of composite bands.

[0141] After FSSM reconstruction, high band frequencies are normalized coded to maintain the temporal envelope of the original flattened spectrum.

[0142] ASR Synthesis:

[0143] To reconstruct the original spectrum, the flattened full band signal from block 86 must be supplemented with harmonics and HF tonals, which were ASR coded at the encoder. ASR synthesis proceeds by using the information in the incoming encoded signal that signifies one or more fundamental frequencies and, where applicable, a phase

signal. Specifically, fundamentals are identified by ASR information that is sent from block 54 to block 88, with the actual ODFT representation of that fundamental being sent from block 78 to block 88.

[0144] Each such fundamental frequency is multiplied in frequency by all the integers between a start and a stop integer to construct harmonics in the extended band (that is, synthesize harmonically related transform coefficients based on the harmonic parameters relayed from block 54). Since ASR works with ODFT components, phasing information is included to maintain proper phasing from harmonic to harmonic. In some cases the incoming encoded signal also includes information about a single tonal (essentially a single sinusoid without harmonics).

[0145] In some embodiments the incoming encoded signal includes magnitude information that is used to adjust the magnitude of the synthesized harmonics. In other embodiments, however, no magnitude adjustment is performed except for such adjustment that may be performed in the MBTAC process described hereinafter.

[0146] The phase continuity of the tonals/partial is ensured by maintaining the phase of the tonal in co-ordination with previous frame's phase, if any were present, else, a null value is assigned to that particular phase value of the tonal. Using a time-domain representation, the signal may be deemed:

$$x(n) = \sum_{n_1} \sum_k A_{n_1,k} \sin(n_1 \cdot f_k \pi + \phi_{\text{prev},n_1,k,\text{tonals}})$$

(in case of non-harmonic tonals only the first partial corresponding to  $n_1=1$  will be synthesized where,  $\phi_{\text{prev},n_1,k,\text{tonals}}$  is expressed as a function of the tonal's frequency value to take care of the phase continuity from the previous tonal values.

[0147] All the ODFT components produced by block 88 are converted in that block to MDCT components which are then combined with the FSSM model components from block 86 before being forwarded to block 60 where they are converted from MDCT components to the time domain.

MBTAC Decoder:

[0148] Essentially, the MDCT components from block 88 may be considered to have high frequency resolution but its frequencies correspond to a relatively long standard time interval. For the application of MBTAC a higher time resolution is necessary. Therefore, the time domain signal from block 60 is processed by the UFB of block 72 into a number of local coefficients in the time-frequency plane to create a time-frequency matrix that is as fine as the matrix that was created by the encoder UFB analysis.

[0149] Desired RMS values of the time-frequency grouped UFB output samples are calculated from the log quantized MBTAC RMS parameters in the incoming encoded signal. Inverse differential coding based on the method chosen at the encoder is done. Inverse R-L differential coding is applied for a stereo signal to recover the R and L RMS values.

[0150] Inverse correlation coding is then performed at the decoder to reverse the third level of frequency grouping (in

case this was done at the encoder). This is performed by first computing the pilot sequence envelope information from the UFB sub-bands which correspond to the baseband and then determining the corresponding higher frequency envelope by scaling the pilot sequence envelope with the transmitted distance parameters as described above (employing the above noted inserter and restorer). After this is done an inversion of the second level of Time-Frequency grouping, described above is done to fill all Time-Frequency bands. The purpose of this inversion is generate a set of  $N \times M$  target RMS values for the UFB samples. The partitioning  $N \times M$  is identical to the partitioning used by the encoder MBATC processor after first level of grouping. Since due to the second-level of grouping only a reduced number of RMS values were coded and transmitted to the decoder (and made available to block 74 by block 54), these values are then mapped to the original  $N \times M$  grid to determine the desired RMS value at each of these grid point.

[0151] The ratio of the desired block RMS computed above and that of the reconstructed spectrum for every time-frequency block (i.e. each point of the  $N \times M$  grid) is then computed in block 74 and used to scale the complex reconstructed time-frequency UFB samples for that time-frequency block. This ensures, that the envelope of the original spectrum is restored (using the above mentioned restorer) to desired accuracy. The above spectrum is then UFB synthesized in block 76 to regain the time domain signals.

[0152] After these components are adjusted by the MBTAC process, the components of the base band and the extended band are now inverted in block 76 to produce the final corrected time-domain signal.

[0153] Obviously, many modifications and variations of the present invention are possible in light of the above teachings. It is therefore to be understood that within the scope of the appended claims, the invention may be practiced otherwise than as specifically described.

1. A method for encoding an audio signal, the method comprising the steps of:

transforming the audio signal into a discrete plurality of (a) basic transform coefficients corresponding to basic spectral components located in a base band and (b) extended transform coefficients corresponding to components located beyond the base band;

correlating that is (i) based on at least some of the basic transform coefficients and at least some of the extended transform components and (ii) performed by programmatically determining and applying a primary frequency scaling parameter and a primary frequency translation parameter to form a revised relation between the basic transform coefficients and extended transform coefficients that increases their correlation; and

forming an encoded signal based on the basic transform coefficients, the primary frequency scaling parameter and the primary frequency translation parameter.

2. A method according to claim 1 wherein the step of transforming the audio signal employs MDCT.

3. A method according to claim 1 wherein the step of transforming the audio signal employs MDCT and DFT.

4. A method according to claim 4 wherein the step of correlating is performed by:

composing a 1<sup>st</sup> composite band by combining the basic transform coefficients with relocated coefficients formed by mapping with the 1<sup>st</sup> adjusted pair from the base band into another band located between the base band's upper limit and its image, said image formed using the primary adjusted pair; and

starting with  $n=2$ , iteratively:

(a) sequentially adjusting an  $n^{\text{th}}$  frequency scaling parameter and an  $n^{\text{th}}$  frequency translation parameter in a predetermined manner and selecting an  $n^{\text{th}}$  adjusted pair of them that causes the highest correlation, the  $(n-1)^{\text{th}}$  frequency translation parameter exceeding the  $n^{\text{th}}$  frequency translation parameter; and

(b) composing an  $n^{\text{th}}$  composite band by combining the  $(n-1)^{\text{th}}$  composite band with relocated coefficients formed by mapping with the  $n^{\text{th}}$  adjusted pair from the  $(n-1)^{\text{th}}$  composite band into another band located between the  $(n-1)^{\text{th}}$  composite band's upper limit and its image, formed using the  $n^{\text{th}}$  adjusted pair.

5. A method according to claim 4 wherein the iterative steps of adjusting and composing are terminated after composing the  $M^{\text{th}}$  composite band, the step of forming an encoded signal is performed by including the 1<sup>st</sup> through  $M^{\text{th}}$  adjusted pairs.

6. A method according to claim 1 wherein the step of correlating is performed after eliminating from the correlation dominant ones of the basic transform coefficients having a magnitude exceeding to a given extent magnitudes in neighborhoods that are predefined for each of said dominant ones.

7. An encoder for encoding an audio signal including a processor comprising:

a transform for transforming the audio signal into a discrete plurality of (a) basic transform coefficients corresponding to basic spectral components located in a base band and (b) extended transform coefficients corresponding to components located beyond the base band;

a correlator for providing a correlation that is (i) based on at least some of the basic transform coefficients and at least some of the extended transform components and (ii) performed by programmatically determining and applying a primary frequency scaling parameter and a primary frequency translation parameter to form a revised relation between the basic transform coefficients and extended transform coefficients that increases their correlation; and

a former for forming an encoded signal based on the basic transform coefficients, the primary frequency scaling parameter and the primary frequency translation parameter.

8. An encoder according to claim 7 wherein the basic transform coefficients are grouped into a plurality of sub-bands with members of each sub-band being assigned a corresponding representative coefficient that is included as a group substitute in said encoded signal to reduce its coefficient count.

9. An encoder according to claim 7 wherein the transform is operable to transform the audio signal with MDCT.

10. An encoder according to claim 7 wherein the transform is operable to transform the audio signal with MDCT and DFT.

11. An encoder according to claim 7 wherein the correlator is operable to sequentially adjusting the primary frequency scaling parameter and the primary frequency translation parameter in a predetermined manner and select a 1<sup>st</sup> adjusted pair of them that causes the highest correlation.

12. An encoder according to claim 11 wherein the correlator is operable to compose a 1<sup>st</sup> composite band by combining the basic transform coefficients with relocated coefficients formed by mapping with the 1<sup>st</sup> adjusted pair from the base band-into another band located between the base band's upper limit and its image, said image formed using the primary adjusted pair, the correlator being further operable, starting, with  $n=2$ , to iteratively:

(a) sequentially adjust an  $n^{\text{th}}$  frequency scaling parameter and an  $n^{\text{th}}$  frequency translation parameter in a predetermined manner and select an  $n^{\text{th}}$  adjusted pair of them that causes the highest correlation, the  $(n-1)^{\text{th}}$  frequency translation parameter exceeding the  $n^{\text{th}}$  frequency translation parameter; and

(b) compose an  $n^{\text{th}}$  composite band by combining the  $(n-1)^{\text{th}}$  composite band with relocated coefficients formed by mapping with the  $n^{\text{th}}$  adjusted pair from the  $(n-1)^{\text{th}}$  composite band into another band located between the  $(n-1)^{\text{th}}$  composite band's upper limit and its image, formed using the  $n^{\text{th}}$  adjusted pair.

13. An encoder according to claim 7 wherein the correlator is operable to correlate after eliminating dominant ones of the basic transform coefficients having a magnitude exceeding to a given extent magnitudes in neighborhoods that are predefined for each of said dominant ones.

14. An encoder according to claim 7 wherein the transform is operable to provide the basic and extended transform coefficients with some corresponding to one or more standard time intervals and others individually corresponding to one of a plurality of subintervals within said one or more standard time intervals, the encoded signal including a plurality of utility coefficients associated with the plurality of subintervals.

15. An encoder according to claim 14 wherein said utility coefficients are considered a fine matrix whose rows and columns are finely indexed by a frequency index and a subinterval index, the encoder comprising:

a categorizer for categorizing each element of said fine matrix into one of  $N$  ordered frequency sub-bands and one-of  $M$  ordered time slots to non-exclusively form an  $N \times M$  group index for each element of said fine matrix; and

a developer for developing a plurality of indexed proxies by merging those elements of said fine matrix that match under the  $N \times M$  group index, said encoded signal including information based on said indexed plurality of proxies.

16. A method for decoding a compressed audio signal signifying (a) basic transform coefficients of basic spectral components derived from a base band, (b) one or more frequency scaling parameters, and (c) one or more frequency translation parameters, the method comprising the steps of:

applying the one or more frequency scaling parameters and the one or more frequency translation parameters to

the basic transform coefficients to provide a plurality of altered primary coefficients having altered spectral significance; and

inverting the basic transform coefficients and the altered primary coefficients to form a time-domain signal.

17. A method according to claim 16 wherein the one or more frequency scaling parameters, and the one or more frequency translation parameters form  $M$  adjusted pairs that are ordered, the step of applying parameters being performed by:

applying the 15 of the  $M$  adjusted pairs to the basic transform coefficients to produce the altered primary coefficients, and combining the basic transform coefficients with the altered primary coefficients to produce a 1<sup>st</sup> composite band; and

starting with  $n=2$ , iteratively applying an  $n^{\text{th}}$  adjusted pair to the  $(n-1)^{\text{th}}$  composite band and combining the results lying above the  $(n-1)^{\text{th}}$  composite band with the  $(n-1)^{\text{th}}$  composite band to form an  $n^{\text{th}}$  composite band.

18. A method according to claim 16 wherein the basic transform coefficients correspond to one or more standard time intervals, said compressed audio signal comprising a plurality of utility coefficients individually corresponding to one of a plurality of subintervals of said one or more standard time intervals, the method comprising the steps of:

transforming the time-domain signal into a frequency domain to obtain a discrete plurality of local coefficients individually assigned to a plurality of successive time slots corresponding in duration to the plurality of subintervals;

resealing the plurality of local coefficients using the utility coefficients from the compressed audio signal; and

inverting the rescaled, discrete plurality of local coefficients into a corrected audio signal in the time-domain.

19. A decoder for decoding a compressed audio signal signifying (a) basic transform coefficients of basic spectral components derived from a base band, (b) one or more frequency scaling parameters, and (c) one or more frequency translation parameters, the decoder comprising:

a relocater for applying the one or more frequency scaling parameters and the one or more frequency translation parameters to the basic transform coefficients to provide a plurality of altered primary coefficients having altered spectral significance; and

an inverter for inverting the basic transform coefficients and the altered primary coefficients to form a time-domain signal.

20. A decoder according to claim 19 wherein the one or more frequency scaling parameters, and the one or more frequency translation parameters form  $M$  adjusted pairs that are ordered, the relocater being operable to applying the 1<sup>st</sup> of the  $M$  adjusted pairs to the basic transform coefficients to produce the altered primary coefficients, and to combine the basic transform coefficients with the primary altered coefficients to produce a 1<sup>st</sup> composite band, the relocater being operable, starting with  $n=2$ , to iteratively apply an  $n^{\text{th}}$  adjusted pair to the  $(n-1)^{\text{th}}$  composite band and combine the results lying above the  $(n-1)^{\text{th}}$  composite band with the  $(n-1)^{\text{th}}$  composite band to form an  $n^{\text{th}}$  composite band.

21. A decoder according to claim 19 wherein the basic transform coefficients correspond to one or more standard time intervals, said compressed audio signal comprising a plurality of utility coefficients individually corresponding to one of a plurality of subintervals of said one or more standard time intervals, the decoder comprising:

a transform for transforming the time-domain signal into a frequency domain to obtain a discrete plurality of local coefficients individually assigned to a plurality of successive time slots corresponding in duration to the plurality of subintervals;

a rescaler for rescaling the plurality of local coefficients using the utility coefficients from the compressed audio signal, the inverter being operable to invert the rescaled, discrete plurality of local coefficients into a corrected audio signal in the time-domain.

22. A decoder according to claim 21 wherein said plurality of subintervals are indexed under an  $N \times M$  group index signifying indexing according to  $N$  ordered frequency subbands and  $M$  ordered time slots.

23. A method for encoding an audio signal, the method comprising the steps of:

transforming the audio signal into a discrete plurality of primary transform coefficients corresponding to spectral components located in a designated band;

correlating based on a correspondence between at least some of the primary transform coefficients and programmatically synthesized data corresponding to a synthetic harmonic or individual sinusoids spectrum comprising any combination of one or more harmonic patterns and one or more individual sinusoids; and

forming an encoded signal based on at least some of the primary transform coefficients, and one or more harmonic parameters signifying one or more characteristics of the synthetic harmonic or individual sinusoids spectrum.

24. A method according to claim 23 wherein said encoded signal does not include those ones of the primary transform coefficients that correspond to components of the synthetic harmonic spectrum.

25. A method according to claim 24 wherein said encoded signal includes one or more noise parameters signifying a flattened spectrum produced by eliminating from the encoded signal those ones of the primary transform coefficients that correspond to components of the synthetic harmonic spectrum.

26. A method according to claim 23 wherein the step of transforming is performed by

transforming the audio signal into (a) a discrete plurality of basic transform coefficients corresponding to basic spectral components located in a base band, and (b) extended transform coefficients located beyond the base band, the step of correlating primary coefficients being performed by

correlating the extended transform coefficients to programmatically synthesized data corresponding to a synthetic harmonic spectrum, the encoded signal including at least some of the basic transform coefficients.

27. A method according to claim 26 comprising the step of:

removing those ones of the extended transform coefficients that correspond to components of a synthetic harmonic or individual sinusoids spectrum comprising any combination of one or more harmonic patterns and one or more individual sinusoids to establish a flattened spectrum.

28. A method according to claim 27 wherein said encoded signal includes one or more noise parameters signifying the flattened spectrum.

29. A method according to claim 27 comprising the step of:

correlating at least some of the basic transform coefficients to at least some of the extended transform coefficients by programmatically determining and applying a primary frequency scaling parameter and a primary frequency translation parameter to recast the relation between basic transform coefficients and extended transform coefficients and increase their correlation, the encoded signal including the primary frequency scaling parameter and the primary frequency translation parameter.

30. A method according to claim 29 wherein the step of correlating basic transform coefficients is performed after eliminating dominant ones of the basic transform coefficients having a magnitude exceeding to a given extent magnitudes in neighborhoods that are predefined of each of said dominant ones.

31. A method according to claim 29 wherein the step of correlating basic components is performed by:

composing a 1<sup>st</sup> composite band by combining the basic transform coefficients with relocated coefficients formed by mapping with the 1<sup>st</sup> adjusted pair from the base band into another band located between the base band's upper limit and its image, said image formed using the primary adjusted pair; and

starting with  $n=2$ , iteratively:

(a) sequentially adjusting an  $n^{\text{th}}$  frequency scaling parameter and an  $n^{\text{th}}$  frequency translation parameter in a predetermined manner and selecting an  $n^{\text{th}}$  adjusted pair of them that causes the highest correlation, the  $(n-1)^{\text{th}}$  frequency translation parameter exceeding the  $n^{\text{th}}$  frequency translation parameter; and

(b) composing an  $n^{\text{th}}$  composite band by combining the  $(n-1)^{\text{th}}$  composite band with relocated coefficients formed by mapping with the  $n^{\text{th}}$  adjusted pair from the  $(n-1)^{\text{th}}$  composite band into another band located between the  $(n-1)^{\text{th}}$  composite band's upper limit and its image, formed using the  $n^{\text{th}}$  adjusted pair.

32. An encoder for encoding an audio signal comprising:

a transform for transforming the audio signal into a discrete plurality of primary transform coefficients corresponding to spectral components located in a designated band;

a correlation device for correlating based on a correspondence between at least some of the primary transform coefficients and programmatically synthesized data corresponding to a synthetic harmonic spectrum; and

a former for forming an encoded signal based on at least some of the primary transform coefficients, and one or more harmonic parameters signifying one or more characteristics of the synthetic harmonic spectrum.

33. An encoder according to claim 32 wherein the primary transform coefficients are grouped into a plurality of sub-bands with members of each sub-band being assigned a corresponding representative coefficient that is included as a group substitute in said encoded signal to reduce its coefficient count.

34. An encoder according to claim 32 wherein said synthetic harmonic spectrum comprises at least two distinct harmonic patterns.

35. An encoder according to claim 32 wherein said encoded signal does not include those ones of the primary transform coefficients that correspond to components of the synthetic harmonic spectrum.

36. An encoder according to claim 35 wherein said form is operable to form said encoded signal to include one or more noise parameters signifying a flattened spectrum produced by eliminating from the encoded signal those ones of the primary transform coefficients that correspond to components of the synthetic harmonic spectrum.

37. An encoder according to claim 32 wherein the transform is operable to transform the audio signal into (a) a discrete plurality of basic transform coefficients corresponding to basic spectral components located in a base band, and (b) extended transform coefficients located beyond the base band, the correlator being operable to correlate the extended transform coefficients to programmatically synthesized data corresponding to a synthetic harmonic spectrum, former being operable to include in the encoded signal at least some of the basic transform coefficients.

38. An encoder according to claim 37 wherein said synthetic harmonic spectrum comprises at least two distinct harmonic patterns.

39. An encoder according to claim 37 wherein the former is operable to remove those ones of the extended transform coefficients that correspond to components of the synthetic harmonic spectrum to establish a flattened spectrum.

40. An encoder according to claim 39 wherein said former is operable to include in the encoded signal one or more noise parameters signifying the flattened spectrum.

41. An encoder according to claim 39 comprising:

a correlator for correlating at least some of the basic transform coefficients to at least some of the extended transform coefficients by programmatically determining and applying a primary frequency scaling parameter and a primary frequency translation parameter to recast the relation between basic transform coefficients and extended transform coefficients and increase their correlation, said former being operable to include in the encoded signal the primary frequency scaling parameter and the primary frequency translation parameter.

42. An encoder according to claim 41 wherein the correlation device is operable to correlate after eliminating dominant ones of the basic transform coefficients having a magnitude exceeding to a given extent magnitudes in neighborhoods that are predefined for each of said dominant ones.

43. An encoder according to claim 41 wherein the correlation device is operable to correlate by sequentially adjusting the primary frequency scaling parameter and the primary

frequency translation parameter in a predetermined manner and selecting a 1<sup>st</sup> adjusted pair of them that causes the highest correlation.

44. An encoder according to claim 43 wherein the correlation device is operable to compose a 1<sup>st</sup> composite band by combining the basic transform coefficients with relocated coefficients formed by mapping with the 1<sup>st</sup> adjusted pair from the base band into another band located between the base band's upper limit and its image, said image formed using the primary adjusted pair, the correlation device being operable, starting with  $n=2$ , to iteratively:

(a) sequentially adjust an  $n^{\text{th}}$  frequency scaling parameter and an  $n^{\text{th}}$  frequency translation parameter in a predetermined manner and select an  $n^{\text{th}}$  adjusted pair of them that causes the highest correlation, the  $(n-1)^{\text{th}}$  frequency translation parameter exceeding the  $n^{\text{th}}$  frequency translation parameter; and

(b) compose an  $n^{\text{th}}$  composite band by combining the  $(n-1)^{\text{th}}$  composite band with relocated coefficients formed by mapping with the  $n^{\text{th}}$  adjusted pair from the  $(n-1)^{\text{th}}$  composite band into another band located between the  $(n-1)^{\text{th}}$  composite band's upper limit and its image, formed using the  $n^{\text{th}}$  adjusted pair.

45. An encoder according to claim 32 wherein the transform is operable to provide the primary transform coefficients with some corresponding to one or more standard time intervals and others individually corresponding to one of a plurality of subintervals within said one or more standard time intervals, the former being operable to include in the encoded signal a plurality of utility coefficients associated with the plurality of subintervals.

46. An encoder according to claim 45 wherein said utility coefficients are considered a fine matrix whose rows and columns are finely indexed by a frequency index and a subinterval index, the encoder comprising:

a categorizer for categorizing each element of said fine matrix into one of  $N$  ordered frequency sub-bands and one of  $M$  ordered time slots to non-exclusively form an  $N \times M$  group index for each element of said fine matrix; and

a developer for developing a plurality of indexed proxies by merging those elements of said fine matrix that match under the  $N \times M$  group index, said encoded signal including information based on said indexed plurality of proxies.

47. A method for decoding a compressed audio signal signifying (a) a plurality of basic transform coefficients corresponding to basic spectral components located in a base band, and (b) one or more harmonic parameters signifying one or more characteristics of a synthetic harmonic or individual sinusoids spectrum comprising any combination of one or more harmonic patterns and one or more individual sinusoids, the method comprising the steps of:

synthesizing one or more harmonically related transform coefficients based on the one or more harmonic parameters; and

inverting the basic transform coefficients and the one or more harmonically-related transform coefficients into a time-domain signal.

48. A method according to claim 47, wherein the compressed audio signal includes one or more frequency scaling

parameters, and one or more frequency translation parameters, the method comprising the step of:

applying the one or more frequency scaling parameters and the one or more frequency translation parameters to the basic transform coefficients to provide a plurality of altered primary coefficients having altered spectral significance, the step of inverting being performed by including the altered primary coefficients when forming the time-domain signal.

49. A method according to claim 48 wherein the one or more frequency scaling parameters, and the one or more frequency translation parameters form M adjusted pairs that are ordered, the step of applying parameters being performed by:

applying a 1<sup>st</sup> adjusted pair to the basic transform coefficients to provide the primary altered coefficients, and combining the basic transform coefficients with the primary altered coefficients to produce a 1<sup>st</sup> composite band; and

starting with n=2, iteratively applying an n<sup>th</sup> adjusted pair to the (n-1)<sup>th</sup> composite band and combining the results lying above the (n-1)<sup>th</sup> composite band with the (n-i)<sup>th</sup> composite band to form an n<sup>th</sup> composite band.

50. A method according to claim 47 wherein the basic transform coefficients correspond to one or more standard time intervals, said compressed signal comprising a plurality of utility coefficients individually corresponding to one of a plurality of subintervals of said one or more standard time intervals, the method comprising the steps of:

transforming the time-domain signal into a frequency domain to obtain a discrete plurality of local coefficients individually assigned to a plurality of successive time slots corresponding in duration to the plurality of subintervals;

resealing the plurality of local coefficients using the utility coefficients from the compressed audio signal; and

inverting the rescaled, discrete plurality of local coefficients into a corrected audio signal in the time-domain.

51. A decoder for decoding a compressed audio signal signifying (a) a plurality of basic transform coefficients corresponding to basic spectral components located in a base band, and (b) one or more harmonic parameters signifying one or more characteristics of a synthetic harmonic or individual sinusoids spectrum comprising any combination of one or more harmonic patterns and one or more individual sinusoids, the decoder comprising:

a synthesizer for synthesizing one or more harmonically related transform coefficients based on the one or more harmonic parameters; and

an inverter for inverting the basic transform coefficients and the one or more harmonically related transform coefficients into a time-domain signal.

52. A method for encoding an audio signal, the method comprising the steps of:

transforming the audio signal into a discrete plurality of transform coefficients corresponding to spectral components located in a designated band, some of the transform coefficients corresponding to one or more standard time intervals and others individually corre-

sponding to one of a plurality of subintervals within said one or more standard time intervals;

forming an encoded signal based on (a) the plurality of transform coefficients associated with the one or more standard time intervals, and (b) magnitude information based on the plurality of transform coefficients associated with the plurality of subintervals.

53. A method according to claim 51 wherein said transform coefficients corresponding to one of a plurality of subintervals are considered a fine matrix whose rows and columns are finely indexed by a frequency index and a subinterval index, the method including the step of:

categorizing each element of said fine matrix into one of N ordered frequency sub-bands and one of M ordered time slots to non-exclusively form an N×M group index for each element of said fine matrix; and

developing a plurality of indexed proxies by merging those elements of said fine matrix that match under the N×M group index, said encoded signal including information based on said indexed plurality of proxies.

54. A method according to claim 52 comprising the step of:

recoding one or more selections from said plurality of indexed proxies by substituting a value corresponding to a difference between said one or more selections and one or more corresponding adjacent ones of said indexed proxies, adjacency occurring when a pair of indexed proxies separately occupy either (a) an immediately succeeding pair of the N ordered frequency sub-bands or (b) an immediately succeeding pair of said M ordered time slots.

55. A method according to claim 52 comprising the step of:

recoding a selection from said plurality of indexed proxies by substituting a value corresponding to a difference between said selection and a corresponding adjacent pair of said indexed proxies, said adjacent pair separately occupying relative to said selection (a) an immediately preceding one of the N ordered frequency sub-bands, and (b) an immediately preceding one of said M ordered time slots.

56. A method according to claim 52 comprising the step of:

forming one or more consolidated collections from said plurality of indexed proxies, each of the consolidated collections being populated with selected ones of the indexed proxies that together satisfy a predetermined limitation on magnitude variation, each consolidated collection that includes a distinct pair of the indexed proxies will not exclude any intervening one of the indexed proxies that intervene by aligning between the distinct pair by lying on either a common row or common column of the N×M group index, said encoded signal including information based on gross characteristics of the one or more consolidated collections.

57. A method according to claim 52 comprising the step of:

developing from a predetermined number of the lowest ones of the N ordered frequency sub-bands a pilot sequence having M temporally sequential values rep-

representative of the M ordered time slots among the predetermined number; and

correlating the pilot sequence with higher temporal sequences presented by the M ordered time slots for each of the N ordered frequency sub-bands that are beyond the predetermined number, said encoded signal including information based on results of the step of correlating the pilot sequence.

58. A method according to claim 56 wherein the step of correlating the pilot sequence is performed by

pairing the pilot sequence and each of the higher temporal sequences and for each pair: (a) programmatically changing scaling between them, and (b) evaluating them with a separation function to determine whether pair correlation reaches a predetermined threshold before including information on the pair correlation in the encoded signal.

59. An encoder for encoding an audio signal, comprising:

a transform for transforming the audio signal into a discrete plurality of transform coefficients corresponding to spectral components located in a designated band, some of the transform coefficients corresponding to one or more standard time intervals and others individually corresponding to one of a plurality of subintervals within said one or more standard time intervals;

a former for forming an encoded signal based on (a) the plurality of transform coefficients associated with the one or more standard time intervals, and (b) magnitude information based on the plurality of transform coefficients associated with the plurality of subintervals.

60. An encoder according to claim 58 wherein said transform coefficients corresponding to one of a plurality of subintervals are considered a fine matrix whose rows and columns are finely indexed by a frequency index and a subinterval index, the encoder comprising:

a categorizer for categorizing each element of said fine matrix into one of N ordered frequency sub-bands and one of M ordered time slots to non-exclusively form an N×M group index for each element of said fine matrix; and

a developer for developing a plurality of indexed proxies by merging those elements of said fine matrix that match under the N×M group index, said encoded signal including information based on said indexed plurality of proxies.

61. An encoder according to claim 59 comprising:

a recoder for recoding one or more selections from said plurality of indexed proxies by substituting a value corresponding to a difference between said one or more selections and one or more corresponding adjacent ones of said indexed proxies, adjacency occurring when a pair of indexed proxies separately occupy either (a) an immediately succeeding pair of the N ordered frequency sub-bands or (b) an immediately succeeding pair of said M ordered time slots.

62. An encoder according to claim 59 comprising:

a recoder for recoding a selection from said plurality of indexed proxies by substituting a value corresponding to a difference between said selection and a correspond-

ing adjacent pair of said indexed proxies, said adjacent pair separately occupying relative to said selection (a) an immediately preceding one of the N ordered frequency sub-bands, and (b) an immediately preceding one of said M ordered time slots.

63. An encoder according to claim 59 comprising:

a former for forming one or more consolidated collections from said plurality of indexed proxies, each of the consolidated collections being populated with selected ones of the indexed proxies that together satisfy a predetermined limitation on magnitude variation, each consolidated collection that includes a distinct pair of the indexed proxies will not exclude any intervening one of the indexed proxies that intervene by aligning between the distinct pair by lying on either a common row or common column of the N x M group index, said encoded signal including information based on gross characteristics of the one or more consolidated collections.

64. An encoder according to claim 59 comprising:

a developer for developing from a predetermined number of the lowest ones of the N ordered frequency sub-bands a pilot sequence having M temporally sequential values representative of the M ordered time slots among the predetermined number; and

a correlator for correlating the pilot sequence with higher temporal sequences presented by the M ordered time slots for each of the N ordered frequency sub-bands that are beyond the predetermined number, said encoded signal including information based on results of the step of correlating the pilot sequence.

65. An encoder according to claim 63 wherein the correlator is operable to pair the pilot sequence and each of the higher temporal sequences and for each pair: (a) programmatically change scaling between them, and (b) evaluate them with a separation function to determine whether pair correlation reaches a predetermined threshold before including information on the pair correlation in the encoded signal.

66. A method for processing a decompressed audio signal obtained from a discrete plurality of transform coefficients corresponding to one or more standard time intervals, using magnitude information based on a plurality of transform coefficients corresponding to one of a plurality of subintervals of said one or more standard time intervals, the method comprising the steps of:

inverting the discrete plurality of transform coefficients associated with the one or more standard time intervals into a first time-domain signal;

successively transforming the first time-domain signal into a frequency domain to obtain a discrete plurality of local coefficients individually assigned to a plurality of successive time slots corresponding in duration to the plurality of subintervals;

resealing the plurality of local coefficients using from the compressed audio signal the transform coefficients associated with the plurality of subintervals; and

inverting the discrete plurality of local coefficients into a corrected time-domain signal.

67. A method according to claim 65 wherein said plurality of subintervals are indexed under an N×M group index

signifying indexing according to N ordered frequency sub-bands and M ordered time slots.

68. A method according to claim 65 wherein the encoded signal includes a pilot sequence having M temporal sequential values that are representative of M ordered time slots, the method comprising the step of:

populating positions of said  $N \times M$  group index by inserting in each of a plurality of its N ordered frequency sub-bands a corresponding replica of said pilot sequence.

69. A method according to claim 66 wherein one or more of said plurality of subintervals are designated as recoded, the method comprising the step of:

restoring recoded ones of said subintervals by substituting a value corresponding to a summation between each of the recoded ones and one or more adjacent ones of subintervals, adjacency occurring when a pair of subintervals separately occupy either (a) an immediately succeeding pair of the N ordered frequency sub-bands or (b) an immediately succeeding pair of said M ordered time slots.

70. A method according to claim 65 wherein one or more of said plurality of subintervals are designated as recoded, the method comprising the step of:

restoring recoded ones of said subintervals by substituting a value corresponding to a summation between each of the recoded ones and a corresponding adjacent pair of subintervals, said adjacent pair separately occupying relative to each recoded one (a) an immediately preceding one of the N ordered frequency sub-bands, and (b) an immediately preceding one of said M ordered time slots.

71. A decoding accessory for processing a decompressed audio signal obtained from a discrete plurality of transform coefficients corresponding to one or more standard time intervals, using magnitude information based on a plurality of transform coefficients corresponding to one of a plurality of subintervals of said one or more standard time intervals, the accessory comprising:

a first inverter for inverting the discrete plurality of transform coefficients associated with the one or more standard time intervals into a first time-domain signal;

a transform for successively transforming the first time-domain signal into a frequency domain to obtain a discrete plurality of local coefficients individually assigned to a plurality of successive time slots corresponding in duration to the plurality of subintervals;

a rescaler for resealing the plurality of local coefficients using from the compressed audio signal the transform coefficients associated with the plurality of subintervals; and

a second inverter for inverting the discrete plurality of local coefficients into a corrected time-domain signal.

72. A decoding accessory according to claim 70 wherein said plurality of subintervals are indexed under an  $N \times M$  group index signifying indexing according to N ordered frequency sub-bands and M ordered time slots.

73. A decoding accessory according to claim 70 wherein the encoded signal includes a pilot sequence having M temporal sequential values that are representative of M ordered time slots, the accessory comprising:

an inserter for populating positions of said  $N \times M$  group index by inserting in each of a plurality of its N ordered frequency sub-bands a corresponding replica of said pilot sequence.

74. A decoding accessory according to claim 71 wherein one or more of said plurality of subintervals are designated as recoded, the accessory comprising:

a restorer for restoring recoded ones of said subintervals by substituting a value corresponding to a summation between each of the recoded ones and one or more adjacent ones of subintervals, adjacency occurring when a pair of subintervals separately occupy either (a) an immediately succeeding pair of the N ordered frequency sub-bands or (b) an immediately succeeding pair of said M ordered time slots.

75. A decoding accessory according to claim 70 wherein one or more of said plurality of subintervals are designated as recoded, the accessory comprising:

a restorer for restoring recoded ones of said subintervals by substituting a value corresponding to a summation between each of the recoded ones and a corresponding adjacent pair of subintervals, said adjacent pair separately occupying relative to each recoded one (a) an immediately preceding one of the N ordered frequency sub-bands, and (b) an immediately preceding one of said M ordered time slots.

76. A method for encoding an audio signal, the method comprising the steps of:

transforming the audio signal into at least a discrete plurality of transform coefficients corresponding to spectral components located in a designated band, said transform coefficients including a standard grouping and a substandard grouping, the standard grouping being associated with one or more standard time intervals, the substandard grouping being dividable into a plurality of isofrequency sequences, each of the plurality of isofrequency sequences encompassing said one or more standard time intervals and being associated with a corresponding one of the transform coefficients in the standard grouping, said transform coefficients of said standard grouping each being assigned a masking characteristic for perceptually attenuating spectrally nearby ones of said standard grouping according to a predefined masking function having a predefined domain, and weakening the masking characteristic of each of the transform coefficients in the standard grouping based on the extent its corresponding one of the isofrequency sequences varies and correlates with spectrally nearby ones of the isofrequency sequences.

77. A method according to claim 75 wherein the step of weakening based on sequence variation is performed by evaluating a peak to valley ratio in the corresponding one of the isofrequency sequences.

78. A method according to claim 76 wherein the step of weakening includes the steps of:

calculating a correlation value; and

multiplicatively combining the peak to valley ratio and the correlation value to form a comodulation masking release value.

79. An encoder for encoding an audio signal comprising: a transform for transforming the audio signal into at least a discrete plurality of transform coefficients corresponding to spectral components located in a designated band, said transform coefficients including a standard grouping and a substandard grouping, the standard grouping being associated with one or more standard time intervals, the substandard grouping being dividable into a plurality of isofrequency sequences, each of the plurality of isofrequency sequences encompassing said one or more standard time intervals and being associated with a corresponding one of the transform coefficients in the standard grouping, said transform coefficients of said standard grouping each being assigned a masking characteristic for perceptually attenuating spectrally nearby ones of said standard grouping according to a predefined masking function

having a predefined domain, and a weakener for weakening the masking characteristic of each of the transform coefficients in the standard grouping based on the extent its corresponding one of the isofrequency sequences varies and correlates with spectrally nearby ones of the isofrequency sequences.

80. A encoder according to claim 78 wherein the weakener is operable to evaluate a peak to valley ratio in the corresponding one of the isofrequency sequences.

81. A encoder according to claim 79 wherein the weakener is operable to calculating a correlation value; and multiplicatively combining the peak to valley ratio and the correlation value to form a comodulation masking release value.

\* \* \* \* \*



# Audio Engineering Society Convention Paper 5553

Presented at the 112th Convention  
2002 May 10–13 Munich, Germany

*This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42<sup>nd</sup> Street, New York, New York 10165-2520, USA; also see [www.aes.org](http://www.aes.org). All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

## Spectral Band Replication, a novel approach in audio coding

Martin Dietz<sup>1</sup>, Lars Liljeryd<sup>2</sup>, Kristofer Kjörling<sup>2</sup> and Oliver Kunz<sup>1</sup>

<sup>1</sup> Coding Technologies, Nuremberg, Germany

<sup>2</sup> Coding Technologies, Stockholm, Sweden

Correspondence should be addressed to Martin Dietz ([diz@codingtechnologies.com](mailto:diz@codingtechnologies.com))

### ABSTRACT

Spectral Band Replication (SBR) is a novel technology which significantly improves the compression efficiency of perceptual audio codecs. SBR reconstructs the high frequency components of an audio signal on the receiver side. Thus, it takes the burden of encoding and transmitting high frequency components off the encoder, allowing for a much higher audio quality at low data rates. In December 2001, SBR has been chosen as Reference Model for the MPEG standardization process for bandwidth extension.

The paper will highlight the underlying technical ideas and the achievable efficiency improvements. A second focus will be a description of current and future applications of SBR.

### INTRODUCTION

Back in 1997, Coding Technologies started research on a novel approach to improve the efficiency of digital audio coding algorithms. A team of Swedish researchers led by Lars Liljeryd had the idea to recreate (replicate) the high-frequencies of a decoded audio signal in an accurate way, mainly by the re-use of signal information obtained from the decoded base-band signal. Thus a new concept of redundancy coding in the frequency domain was developed. This concept led to what is today known as Spectral Band Replication technology (SBR). SBR is able to significantly improve the compression efficiency of audio codecs over a wide range of bitrates.

As SBR is an enhancement technology, it always needs an audio codec to hook upon. Two obvious candidates were identified: mp3

[1,2,6], as the world's most popular coding format, and MPEG AAC [3,7], so far the most powerful waveform audio coding algorithm. After a year of successful cooperation with the Fraunhofer Institute for Integrated Circuits (FhG-IIS), several mp3 and AAC experts from FhG joined Coding Technologies in autumn 2000 to work on the SBR enhanced audio coding systems. Since then, several applications have adopted SBR. mp3PRO is the result of combining mp3 with SBR in a back- and forwards compatible way. mp3PRO is available since June 2001 and has in the meantime gained significant market acceptance. In January 2001, AAC+SBR has been adopted as audio codec for the Digital Radio Mondiale System. In December 2001, SBR has been successfully submitted to MPEG and became reference model of the MPEG-4 version 3 audio standardization process. The following chapters will describe in more detail the

background of SBR, the technology behind SBR and applications for this novel technology.

BACKGROUND OF THE SBR DEVELOPMENT

The use of means to recreate missing or reduced high-frequencies in audio is nothing new, and has been used more or less frequent over the years in the recording industry. One of the most famous products was the "Aphex Aural Exciter" invented by Kurt Knoppel in the early 70's. Although Knoppel was not the first (or last) in this field, similar technologies can be found in some audio editing software tools today as well as in some audio processing hardware units. The main aim with such technologies is to add some "sparkle", "warmth" or high-end "enrichment" of the recorded vocals, speech, instruments, CD-master or broadcast feed, using typically non-linearities or "controlled" forms of high-frequency distortion. However, such technologies are performance limited by the way the high-frequencies are recreated. This limits their scope of use to just "touching-up" existing programme material, and making them unable to replicate major frequency ranges of a high quality programme material.

A related technology, "Harmonizing", was first successfully pioneered by the Eventide Clockworks "Harmonizer" in the early 70's, where programme material can be transposed in a "musically correct" manner. This is quite opposite to the "Bode Frequency Shifter" developed for electronic music applications by Harald Bode where harmonic relations are "non-musically" shifted (scrambled). Finally, the Ring-Modulator, one of the oldest "weird" sound effects units, is sometimes used in electronic music. Spin-offs from Harmonizer-like algorithms are the popular pitch-shift and time-stretch algorithms used in many audio workstation sound-editing software tools today.

The above-mentioned technologies or methods are sufficient for their intended use, but unsuitable for high quality SBR-like High-Frequency-Reconstruction (HFR) methods as introduced by Coding Technologies for audio coding. Therefore the development of new (patented) and more suitable algorithms was a prerequisite for the successful deployment of SBR. Strict demand was set on basic algorithm performance due to the unusual need to transpose signal components from the available (coded) baseband, by at least a factor three, without creating annoying artefacts.

Particular emphasis was put on recreating psychoacoustically accurate representations of the original high-band signal components without violating the well-known consonance-dissonance criteria. However compared to the original signal, the recreated high-band is not necessarily signal-coherent in a technical sense, but coherent in a mere psychoacoustic sense, a most important characteristic.

WHY SBR IMPROVES AUDIO CODING

Research on perceptual audio coders started about 20 years ago. Research on the human auditory system revealed that hearing is mainly based on a short-term spectral analysis of the audio signal. The so-called masking effect was observed: the human auditory system is not able to perceive distortions that are masked by a stronger signal in the spectral neighborhood. Thus, when looking at the short-term spectrum, the so-called masking threshold can be calculated. Distortions below this threshold are – in the ideal case – inaudible. Figure 1 illustrates in a very simplified way how the masking threshold can look like. Figure 1 also reveals that the distortion generated by the PCM format used on CD's is for all frequencies way below this threshold. However, since the PCM signal, representing the time domain, does not allow frequency dependent shaping of the quantization noise, a high SNR needs to be

applied in order to achieve high quality sound across a large dynamic range. This also reveals the basic idea of perceptual audio coding: if it would be possible to shape the quantization noise in the frequency domain, compression of audio signals beyond the CD format has been achievable.

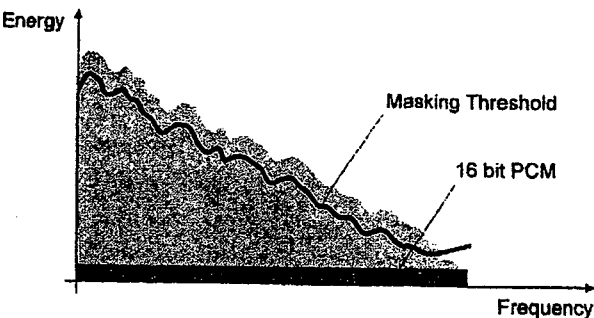


Figure 1: Spectrum and Masking Threshold

Research started on how to calculate the masking threshold ("psychoacoustic model") and how to process the audio signal in a way that only audible information resides in the signal. The ideal audio codec introduces quantization distortions that are exactly below the masking threshold (figure 2).

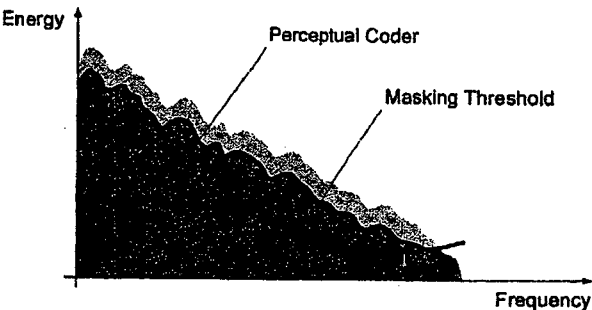


Figure 2: Ideal Perceptual Coding

This research led to today's waveform based perceptual audio codecs, such as MPEG Layer 2, AC-3, mp3, Atrac, PAC and MPEG AAC. All these codecs are based on the same, well-known principle, as shown in figure 3. The audio signal is transformed into the frequency domain by means of a filterbank or transform on a block-by-block basis. The resulting short-time spectrum is quantized in a way that the masking threshold calculated by the psychoacoustic model is not violated. The quantized spectrum gets coded and packed into a bitstream. The decoder performs the reverse steps.

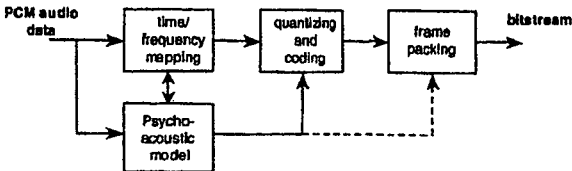


Figure 3: Perceptual waveform encoder

Although all the codecs mentioned follow the same principle, they still differ significantly in compression efficiency. The performance of such codecs depends mainly on two aspects: how precisely the masking threshold can be approached and how efficiently the quantized spectrum can be coded. AAC, for example, is, according to several tests, twice as efficient as Layer 2. Nevertheless, traditional waveform coding has its limits. Having reached a performance like AAC it gets very hard to further increase

compression. Figure 4 illustrates what happens if the bitrate is lowered significantly from the codecs reasonable operating point. The quantization noise significantly exceeds the masking threshold, thus generating audible artefacts.

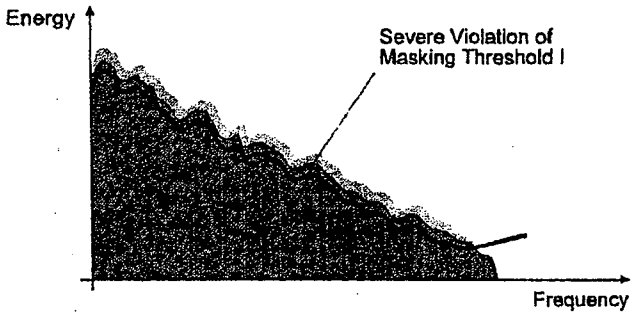


Figure 4: Waveform coding beyond its limits

Two main methods have been used so far to overcome this problem in perceptual waveform codecs. The most important one is to limit the audio bandwidth of the signal in or prior to the coding process. As there is no high frequency energy to be coded, more information is available for the remainder of the spectrum, resulting in a clean, but hollow sounding signal (figure 5). The other method, called intensity stereo, can only be used for stereo signals. In intensity stereo, only one channel and a panning information is transmitted instead of a left and a right channel. However, this is only of rather limited use for increasing compression efficiency, as in many cases the stereo image of the audio signal gets destroyed. A third method used is insertion of artificial noise. This method, however, works only with a rather limited class of audio signals.

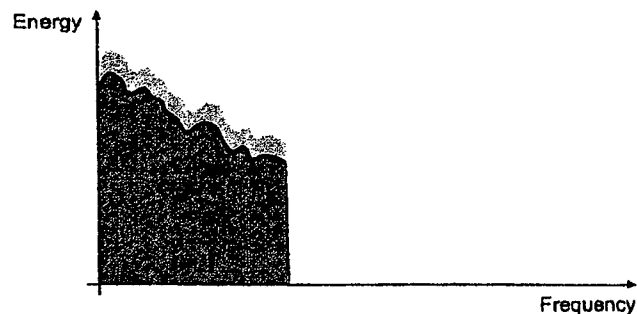


Figure 5: Limiting the audio bandwidth

At this point, the SBR technology comes on the scene. SBR deviates from the waveform coding principle towards a hybrid waveform/parametric method. It is based on the fact that in most cases there are large dependencies between the lower and higher frequency parts of an audio signal. Therefore, the high frequency part of an audio signal can be efficiently reconstructed from the low frequency part. Transmission of the high frequency part is therefore not necessary - only a small amount of SBR control data needs to be carried in the bitstream to guarantee an optimal reconstruction of the high frequencies. Figure 6 illustrates the first step in performing the SBR enhancement. The low frequency part is still coded by an ordinary waveform codec such as mp3 or AAC. The high frequency part, however, is generated by a high quality transposition algorithm.

As can be seen in figure 6, the mere generation of high frequency content is not at all sufficient for accurate high frequency reconstruction, since the reconstructed part does not reflect the spectral envelope of the original. Therefore careful adjustment of the spectrum is essential for the performance of the system. The adjustment is controlled by the SBR information carried in the

bitstream and results in a correctly shaped high frequency part (figure 7).

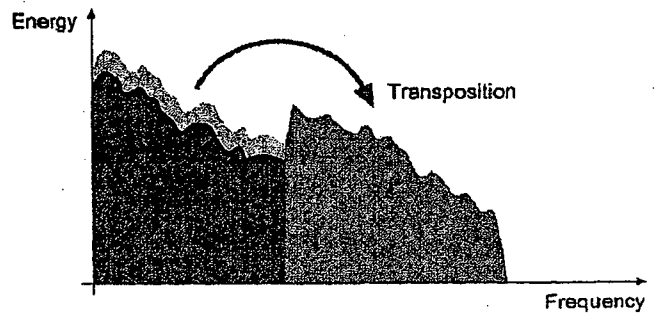


Figure 6: High frequency generation based on the waveform coded low frequency part

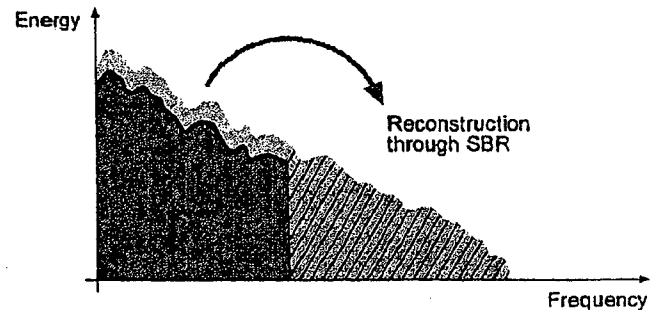


Figure 7: Spectrum after high frequency adjustment

To summarize, SBR enhanced codecs perform significantly better mainly because

- a) SBR allows the recreation of the high frequencies using only a very small amount of transmitted side information. The high frequencies, which normally consume a significant amount of bits, do not need to be waveform coded anymore, resulting in a significant coding gain
- b) the underlying waveform coder can run with a comparatively high SNR, as it is only responsible for the lower frequencies. It can even operate at the optimum sampling rate, which is usually different from the desired output sampling rate. SBR converts the waveform codec sampling rate into the desired output sampling rate.

Obviously there are signals where the reconstruction method does not deliver the desired results, e.g. when there is little relationship between the low and high frequency part. Care has been taken to equip SBR with additional tools so that such situations can be handled very well without losing compression efficiency.

#### HOW SBR WORKS

The SBR system is preferably used as a dual-rate system, with the underlying codec operating at half the original sampling-rate, while SBR operates at the original sampling rate. The following description will briefly explain the different parts in the encoder and decoder of the SBR system.

The SBR encoder works in parallel with the underlying core codec, albeit at a higher sampling-rate. Although SBR is mainly a post process in the decoder, important parameters are extracted in the encoder in order to ensure the most accurate high frequency reconstruction in the decoder. The encoder process is depicted in figure 8. The encoder estimates the spectral envelope of the SBR range for a time and frequency range/resolution suitable for the current input signal segments characteristics. The spectral envelope

is estimated by a complex QMF analysis and subsequent energy calculation. The time and frequency resolutions of the spectral envelopes can be chosen with a high level of freedom, in order to ensure the best suited time frequency resolution for the given input segment. The envelope estimation needs to consider that a transient in the original, mainly situated in the high frequency region (for instance a high-hat), will be present to a minor extent in the SBR generated highband prior to envelope adjustment, since the highband in the decoder is based on the low band where the transient is much less pronounced compared to the highband. This aspect imposes different requirements for the time frequency resolution of the spectral envelope data, compared to ordinary spectral envelope estimation as used in other audio coding algorithms.

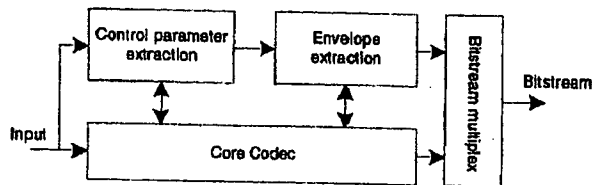


Figure 8: Basic block-diagram of the SBR encoder.

Apart from the spectral envelope, several additional parameters are extracted representing spectral characteristics of the input signal for different time and frequency regions. Since the encoder naturally has access to the original signal as well as information on how the SBR-unit in the decoder will create the high-band, given the specific set of control parameters, it is possible for the system to handle situations where the lowband constitutes a strong harmonic series and the high band, to be recreated, mainly constitutes random signal components, as well as situations where strong tonal components are present in the original highband without counterparts in the lowband, upon which the highband region is based. Furthermore, the SBR encoder works in close relation to the underlying core codec to assess which frequency range should be covered by SBR at a given time. The SBR data is efficiently coded prior to transmission by exploiting entropy coding as well as channel dependencies of the control data, in the case of stereo signals.

The control parameter extraction algorithms need to be carefully tuned to the underlying codec at a given bitrate and a given sampling rate. This is due to the fact that a lower bitrate, usually implies a larger SBR range compared to a high bitrate, and different sampling rates correspond to different time resolutions of the SBR frames. Different core codecs may also display different characteristics during bit constraints, i.e. more or less severe spectral band shut-downs etc.

The decoder is constituted of several different parts, as outlined in figure 9. It comprises a bitstream decoding module, a high frequency reconstruction (HFR) module, an additional high frequency components module, and an envelope adjuster module. The system is based around a complex valued QMF filterbank [4].

In the bitstream extraction module the control data is read from the bitstream and decoded. The time frequency grid is obtained for the current frame, prior to reading the envelope data from the bitstream, and decoding of the same. If a stereo signal is processed, suitable stereo decoding is performed according to the control data. Appropriate CRC-checks and error resilience algorithms are also included.

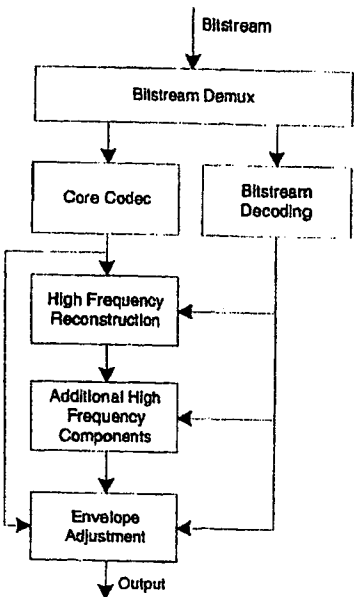


Figure 9: Basic block-diagram of the SBR decoder.

The underlying core decoder decodes the audio signal of the current frame (albeit at the lower sampling rate). The resulting frame of audio data is used for high frequency reconstruction by the HFR module. A possible output from the core decoder is displayed in figure 10. Here it is evident that the output is of low-pass character with a bandwidth of approx. 6kHz.

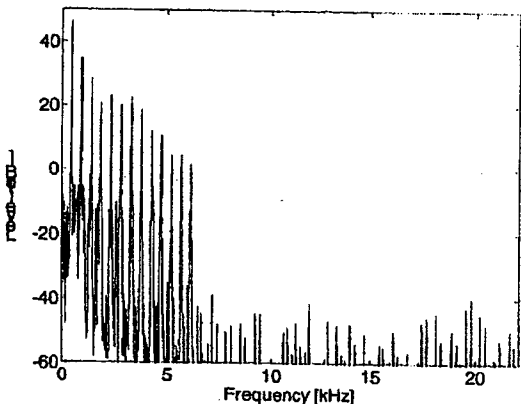


Figure 10: Spectrum of the output signal from the core decoder, for an input signal coded with AAC+SBR at 24kbps.

The decoded low-band signal is analyzed using a QMF filterbank. The high frequency reconstruction and envelope adjustment is subsequently performed on the subband samples of the QMF filterbank. The high frequencies are reconstructed from the low-band in a flexible way, based on the given control parameters. Furthermore, the reconstructed highband is adaptively filtered on a subband channel basis according to the control data to ensure the appropriate spectral characteristics of the given time/frequency region.

The HFR can be done in several ways dependent on the characteristics of the signal being processed. The output from such a transposition based high frequency reconstruction is depicted in figure 11. Here it is evident that the frequency range of the signal has been extended to approx. 15kHz. There are no algorithmic

constraints preventing the HFR module to extend the frequency range even higher. However, for this example 15kHz was deemed sufficient.

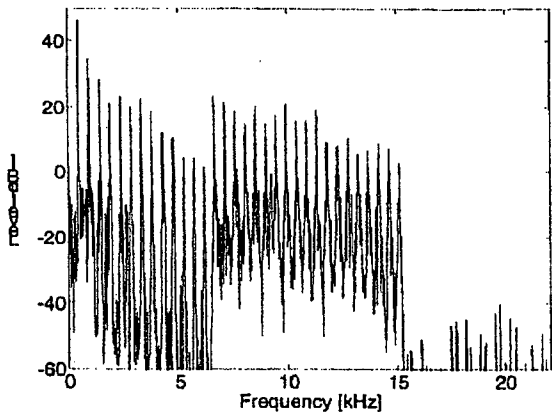


Figure 11: Spectrum of the output signal from a transposition based SBR module, given the input signal displayed in figure 11

The reconstructed highband is subsequently envelope adjusted according to the time/frequency resolution obtained from the encoder. The output signal is enhanced with additional signal components in order to ensure the correct spectral characteristics of the final output.

The low band decoded by the underlying decoder is added to the reconstructed and envelope adjusted highband, giving the output signal. Since the lowband has a sampling rate of half of the highband it is up-sampled during the course of the SBR process. The final output, after envelope adjustment, is depicted in figure 12. It is evident that the SBR module successfully has recreated a high band very similar to that of the original, fig 13.

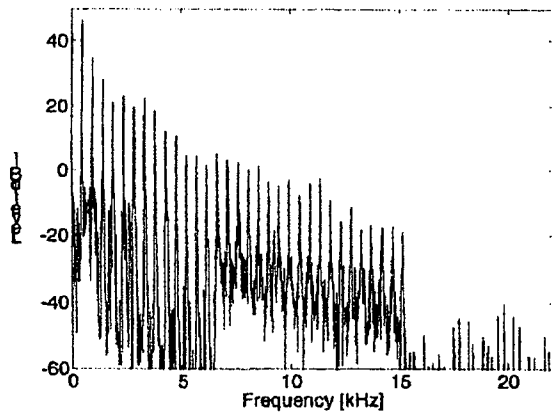


Figure 12: Spectrum of the output signal after envelope adjustment, coded with AAC+SBR at 24kbps.

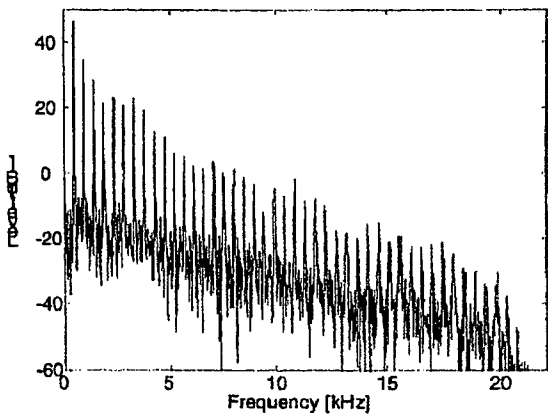


Figure 13: Spectrum of the original signal.

SBR APPLICATIONS

SBR enhanced audio codecs are already in use in several applications, some of which are listed hereafter.

mp3PRO

In January 2001, Coding Technologies and its licensing partner Thomson Multimedia presented mp3PRO, the combination of mp3 and SBR, for the first time. In June 2001, a free demo software has been made available. Since then several software applications integrating mp3PRO have been released (e.g. from Ahead, Steinberg, Magix, ...).

First hardware products are expected in Q2/2002. mp3PRO offers significantly improved performance compared to mp3 and outperforms competitive codecs according to independent tests.

mp3PRO is back- and forward compatible with mp3. Not only will any mp3PRO decoder decode mp3 content, all mp3 players will be able to decode mp3PRO bitstreams, although without the quality improvement achieved through SBR.

Figure 14 shows the integration of mp3 and SBR. Basically, SBR can be seen as pre/postprocessing around the existing mp3 modules. Even the bitstream format remains unchanged, as the SBR data can be carried in the ancillary data field.

More detailed information about mp3PRO can be found in [4].

Digital Radio Mondiale (DRM)

The DRM consortium has been founded in 1998 with the goal to standardize a new digital system for medium, long- and shortwave transmissions. The system should deliver significantly better audio and reception quality than ordinary analog AM. At the same time the existing AM channel spacing should be used. In addition, reuse of existing transmitter equipment should be possible to guarantee a cost efficient introduction of the new system.

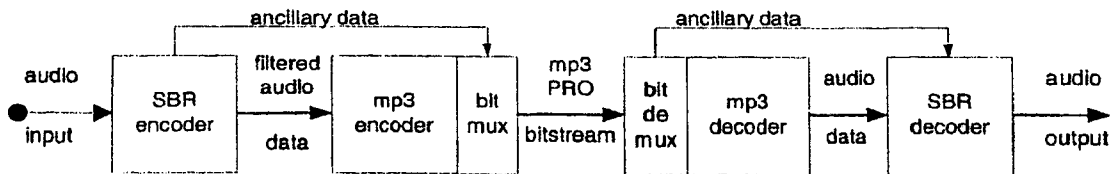


Figure 14: mp3PRO block diagram

Depending on the wavelength and the region, the channel spacing in the AM bands is either 9 kHz or 10 kHz. The propagation conditions, which depend very much on the wavelength, determine the spectral efficiency achievable. Extensive tests have resulted in an efficiency in the range of 1.5 to 3 bit/Hz. As a result, the bitrate available in one transmission channel is in the range of 13...30 kbit/sec. Normal operation will use bitrates between 20 and 25 kbit/sec. Delivering high quality audio at such a low bitrate asks for the most powerful source coding scheme available. Consequently, AAC+SBR, the combination of MPEG-AAC and SBR has been chosen as audio coding method. Tests done within the DRM consortium clearly show the superiority of this coding system compared to previous state of the art. More details about DRM and the listening test can be found in [5,8,9] and at the DRM web site [www.drm.org](http://www.drm.org). DRM finalised the specification of the system in January 2001. It is recommended by the ITU [11] and standardised by ETSI [10]. First regular transmissions and receiver products are expected in 2003.

MPEG

Because of its outstanding compression performance, the combination of AAC and SBR is of high interest not only for broadcasting, but also for audio and audio/video applications like streaming over the internet, streaming and delivery over mobile networks and storage in portable devices. For a lot of these applications, open standards like MPEG play an important role. They usually provide state-of-the-art algorithms and guarantee interoperability and accessibility of the technology. As a consequence, MPEG issued a call for proposals in January 2001, asking for technologies that could further enhance the compression efficiency of the MPEG-4 audio coding algorithms (in particular AAC). SBR was submitted as a proposal; a competitive test was performed by MPEG. AAC+SBR did not only show an excellent performance, but already met the acceptance criteria set up for final acceptance of the technology in the standard. As a result, SBR became Reference Model 0 for version 3 of the MPEG-4 standard. Figure 15 shows the result of the test at 24 kbps mono at the T-Nova test site. The test was done with experienced listeners and the well proven MUSHRA test method. As test items, the "usual" MPEG test items were used (e.g. harpsichord, glockenspiel, pitchpipe, male german speech, castagnettes, a.s.o). Till today those items are known to be very critical for audio coding. AAC+SBR shows a vast improvement compared to AAC, even when compared to the higher bitrate AAC (30 kbps). Moreover AAC+SBR shows a very good absolute performance at this very low bitrate.

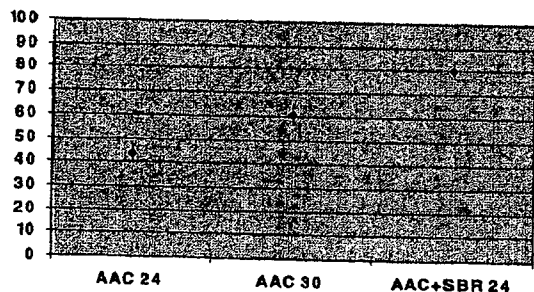


Figure 15: MPEG test result at 24 kbps mono (test site: T-Nova)

More detailed results are given in figure 17. Table 1 explains the acronyms for the excerpts.

|      |                       |
|------|-----------------------|
| Es01 | Suzanne Vega          |
| es02 | Male German Speech    |
| es03 | Female English Speech |
| sc01 | Haydn Trumpet Concert |

|      |                 |
|------|-----------------|
| Sc02 | Meistersinger   |
| Sc03 | Spot            |
| Si01 | Harpsichord     |
| Si02 | Castagnettes    |
| Si03 | Pitchpipe       |
| Sm01 | Bagpipe         |
| Sm02 | Glockenspiel    |
| Sm03 | Plucked Strings |

Table 1: The MPEG excerpts

The stereo test results are shown in figure 16 and Figure 18. Again, AAC+SBR shows a significant gain over AAC and a very good absolute grading of more than 85 at a bitrate of 48 kbit/sec. For both tests it should be noted that the comparison was done against MPEG-4 AAC. A comparison to the yet much more popular MPEG-2 AAC, which is the AAC version available on the market today, would show an even higher improvement through SBR (which operates based on MPEG-2 AAC). Since AAC was so far deemed to be state-of-the-art audio compression, it can be concluded from the this test that AAC+SBR is the new performance leader in audio coding.

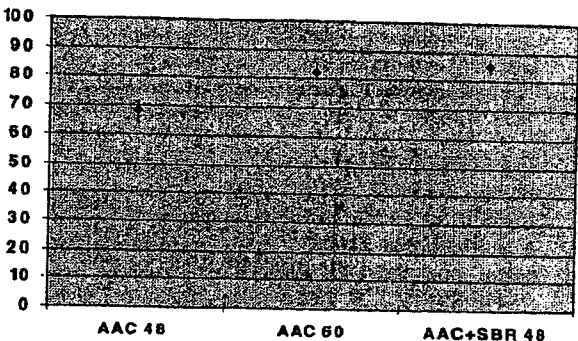


Figure 16: MPEG test result at 48 kbps stereo (test site: T-Nova)

The MPEG-4 version 3 standard will reach Committee Draft (CD) status in May 2002 and the Final Draft International Standard (FDIS) status in March 2003. The improved audio coding standard is expected to be adopted for several of the above mentioned applications.

CONCLUSIONS

Spectral Band Replication is a novel technology that combines traditional audio coding with the capabilities of high quality high frequency reconstruction methods. Through the use of SBR in mp3PRO, the compression efficiency of mp3 could be significantly improved while remaining compatible with the widespread mp3 format. mp3PRO will succeed mp3 in all application areas. The combination of AAC and SBR offers the most powerful audio compression available today. As such it is most suited for digital broadcasting and streaming/delivery over networks with limited resources. AAC+SBR is already used in the market place and is the Reference Model in the MPEG-4 standardization. Both mp3PRO and AAC+SBR are available today in soft- and hardware, several applications have already been launched. It is expected that the number of applications using SBR will increase significantly in the near future.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC IS 11172-3 "Information technology – Coding of moving pictures and associated audio for digital storage media up to about 1,5 Mbit/s", Part 3: Audio, 1991.

- [2] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC IS 13818-3 "Information technology – Generic Coding of Moving Pictures and Associated Audio:", Part 3: Audio, 1994.
- [3] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC IS 13818-7 "Information technology – Generic Coding of Moving Pictures and Associated Audio:", Part 7: Advanced Audio Coding (AAC), 1997.
- [4] Per Ekstrand, Andreas Ehret, Manfred Lutzky and Thomas Ziegler. Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm. In 112th AES Convention, Munich, May 10-May 13, 2002.
- [5] Stefan Meltzner, Fredrik Henn and Reinhold Böhm. SBR-enhanced audio codecs for digital broadcasting such as "Digital Radio Mondial" (DRM). In 112th AES Convention, Munich, May 10-May 13, 2002.
- [6] K. Brandenburg, G. Stoll, Y.F. Doherty, J.D. Johnston, L.v.d. Kerkhof, E.F. Schröder: "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio", JAES 42, 780-792
- [7] Bosi, Brandenburg, Quackenbush, Fielder, Akagiri, Fuchs, Dietz, Herre, Davidson, Oikawa. "ISO/IEC MPEG-2 Advanced Audio Coding", 101st AES Convention, Los Angeles, November 1996
- [8] DRM Source Coding Group, 2001: Report on Subjective Listening Tests of SBR-LC, an AAC-based Audio Bandwidth Widening Tool
- [9] Jonathan Stott, 2001. DRM – Key Technical Features, Reprint from EBU Technical Review No. 286 (March 2001)
- [10] ETSI TS 101980 v1.1.1 (2001-09), Digital Radio Mondiale (DRM): System Specification
- [11] Recommendation ITU-R BS.1348-1 (02/01): Service requirements for digital sound broadcasting at frequencies below 30 MHz

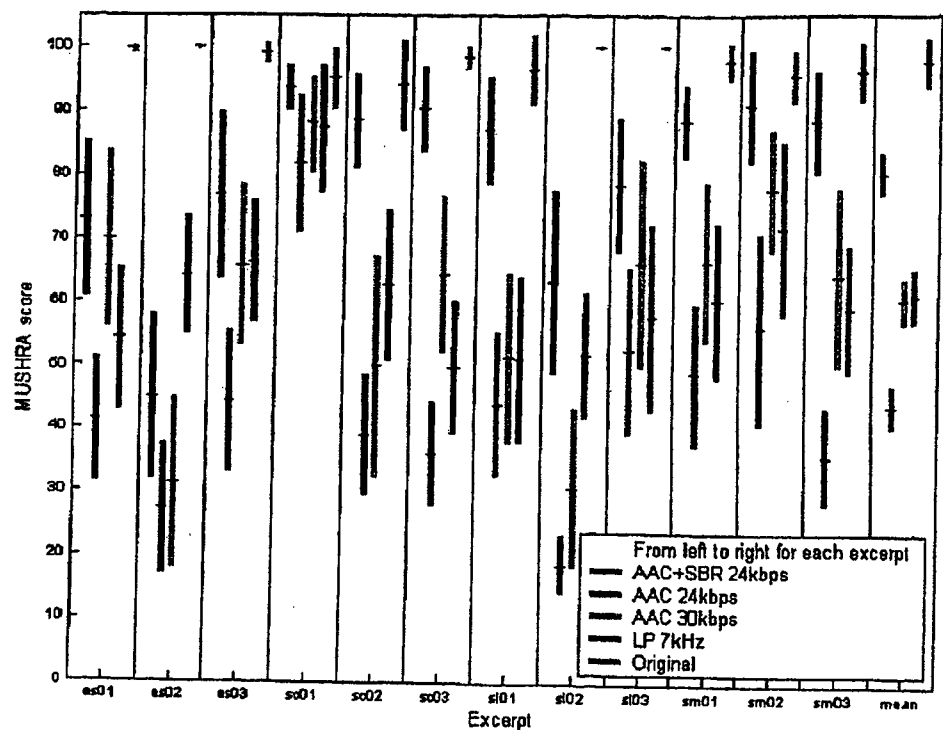


Figure 17: MPEG test results per item, 24 kbps mono (test site: T-Nova)

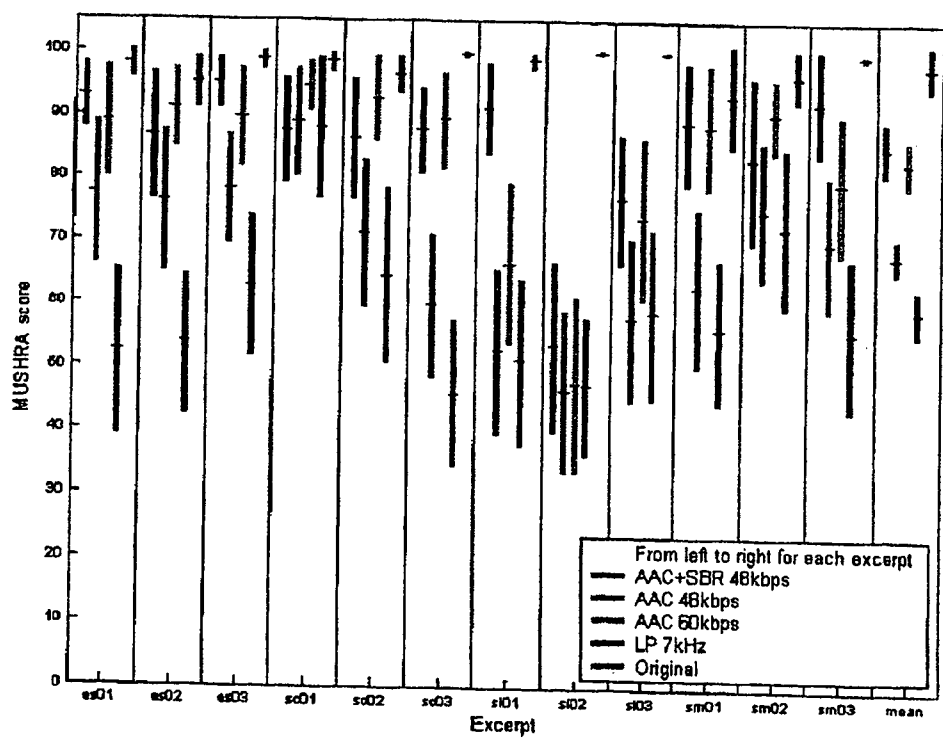


Figure 18: MPEG test results per item, 48 kbps (test site: T-Nova)

## **ASSIGNMENT**

WHEREAS, I/We, Frederik NAGEL, Max NEUENDORF, Nikolaus RETTELBACH, Jeremie LECOMTE, Markus MULTRUS, and Bernhard GRILL, hereinafter referred to as "ASSIGNOR(S)", have invented certain new and useful improvements, as described and set forth in the below-identified application for United States Letters Patent entitled:

### **Novel Patching Methods Inside Spectral Band Replication**

Filing Date: April 2, 2009 Serial No. 61/166,125

WHEREAS, Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e.V., a corporation duly organized under and pursuant to the laws of the country of Germany, having its principal place of business at Hansastrasse 27c, 80686 Munich, Germany, hereinafter referred to as "ASSIGNEE", is desirous of acquiring the entire right, title, and interest in the said invention and application and in any Letters Patent which may be granted with regard to the same;

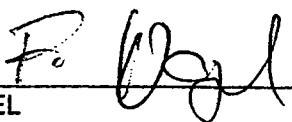
NOW, THEREFORE, TO ALL WHOM IT MAY CONCERN: Be it known that, for One Dollar (\$1.00) and other good and valuable consideration, ASSIGNOR(S) have sold, assigned, and transferred, and by these presents does sell, assign, and transfer unto the said ASSIGNEE, and ASSIGNEE'S successors and assigns, all right, title, and interest in and to said invention, said application for United States Letters Patent and any Letters Patent which may be hereafter granted on the same in the United States and all countries throughout the world, including any divisions, renewals, continuations in whole or part, substitutions, conversions, reissues, revivals, prolongation, or extensions thereof, said interest to be held and enjoyed by said ASSIGNEE as fully and exclusively as it would have been held and enjoyed by said ASSIGNOR(S) had this assignment and transfer not been made, for all time.

ASSIGNOR(S) further agrees that they will, without charge to said ASSIGNEE, but at ASSIGNEE'S expense, cooperate with ASSIGNEE in the prosecution of said application and/or applications, execute, verify, acknowledge, and deliver all such further papers, including applications for Letters Patent and for the reissue thereof, and instruments of assignment and transfer thereof, and will perform such other acts as ASSIGNEE may lawfully request, to obtain or maintain Letters Patent for said invention and improvement in any and all countries, and to vest title thereto in said ASSIGNEE, or ASSIGNEE'S successors and assigns.

ASSIGNOR(S) grants the attorney of record the power to insert on this Assignment any further identification that may be necessary or desirable in order to comply with the rules of the United States Patent and Trademark Office for recordation of this document.

This Assignment may be executed in one or more counterparts, each of which shall be deemed an original and all of which may be taken together as one and the same Assignment.

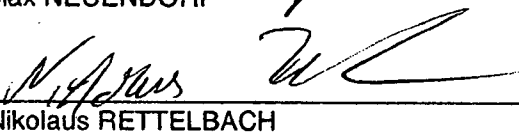
IN TESTIMONY WHEREOF, ASSIGNOR(S) has hereunto signed his name to the assignment on the date indicated below.

  
Frederik NAGEL

16.4.09  
DATE

  
Max NEUENDORF

16.4.09  
DATE

  
Nikolaus RETTELBACH

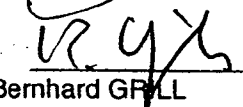
16.4.09  
DATE

  
Jeremie LECOMTE

16.06.08  
DATE

  
Markus MULTRUS

16.06.2008  
DATE

  
Bernhard GRILL

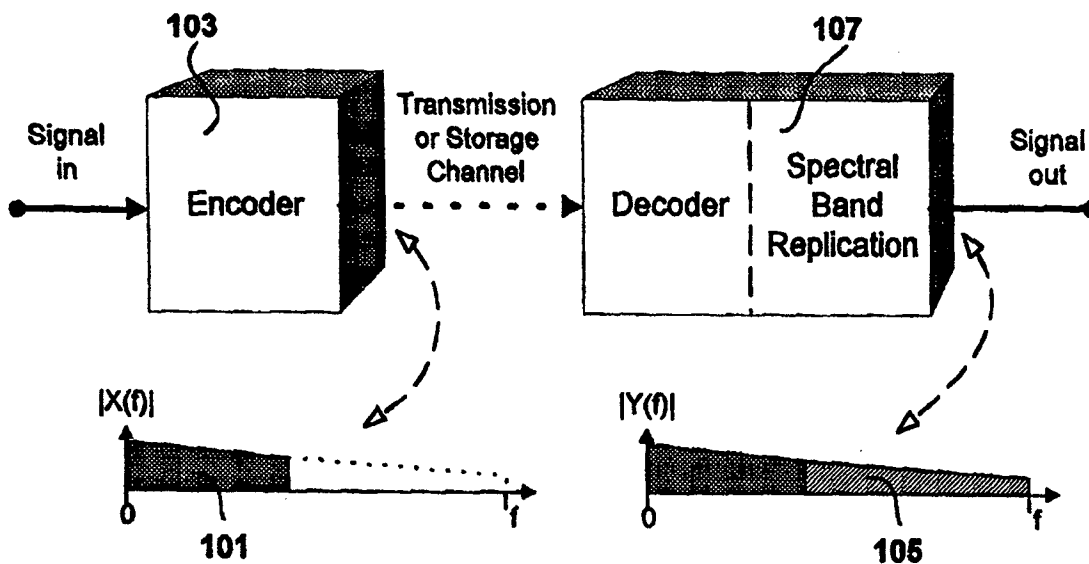
28.04.08  
DATE



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|                                                                                                                                                                                                                                                                                   |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| (51) International Patent Classification <sup>6</sup> :<br><b>H04B 1/66</b>                                                                                                                                                                                                       |  | A2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | (11) International Publication Number: <b>WO 98/57436</b>        |
|                                                                                                                                                                                                                                                                                   |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | (43) International Publication Date: 17 December 1998 (17.12.98) |
| (21) International Application Number: PCT/IB98/00893                                                                                                                                                                                                                             |  | (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). |                                                                  |
| (22) International Filing Date: 9 June 1998 (09.06.98)                                                                                                                                                                                                                            |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                  |
| (30) Priority Data:<br>9702213-1 10 June 1997 (10.06.97) SE<br>9704634-6 12 December 1997 (12.12.97) SE<br>9800268-6 30 January 1998 (30.01.98) SE                                                                                                                                |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                  |
| (71)(72) Applicant and Inventor: LILJERYD, Lars, Gustaf [SE/SE]; Vintervagen 19, S-171 34 Solna (SE).                                                                                                                                                                             |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                  |
| (72) Inventors; and                                                                                                                                                                                                                                                               |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                  |
| (75) Inventors/Applicants (for US only): EKSTRAND, Per, Rune, Albin [SE/SE]; Renstiernas Gata 29, S-116 31 Stockholm (SE). HENN, Lars, Fredrik [CH/SE]; Ritarvagen 14, S-168 31 Bromma (SE). KJÖRLING, Hans, Magnus, Kristofer [SE/SE]; Vindhemsgatan 19C, S-752 27 Uppsala (SE). |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                  |
| (74) Common Representative: LILJERYD, Lars, Gustaf; Vintervagen 19, S-171 34 Solna (SE).                                                                                                                                                                                          |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                  |
| Published<br>Without international search report and to be republished upon receipt of that report.                                                                                                                                                                               |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                  |

(54) Title: SOURCE CODING ENHANCEMENT USING SPECTRAL-BAND REPLICATION



## (57) Abstract

The present invention proposes a new method and apparatus for the enhancement of source coding systems. The invention employs bandwidth reduction (101) prior to or in the encoder (103), followed by spectral-band replication (105) at the decoder (107). This is accomplished by the use of new transposition methods, in combination with spectral envelope adjustments. Reduced bitrate at a given perceptual quality or an improved perceptual quality at a given bitrate is offered. The invention is preferably integrated in a hardware or software codec, but can also be implemented as a separate processor in combination with a codec. The invention offers substantial improvements practically independent of codec type and technological progress.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |                                          |    |                                              |    |                          |
|----|--------------------------|----|------------------------------------------|----|----------------------------------------------|----|--------------------------|
| AL | Albania                  | ES | Spain                                    | LS | Lesotho                                      | SI | Slovenia                 |
| AM | Armenia                  | FI | Finland                                  | LT | Lithuania                                    | SK | Slovakia                 |
| AT | Austria                  | FR | France                                   | LU | Luxembourg                                   | SN | Senegal                  |
| AU | Australia                | GA | Gabon                                    | LV | Latvia                                       | SZ | Swaziland                |
| AZ | Azerbaijan               | GB | United Kingdom                           | MC | Monaco                                       | TD | Chad                     |
| BA | Bosnia and Herzegovina   | GE | Georgia                                  | MD | Republic of Moldova                          | TG | Togo                     |
| BB | Barbados                 | GH | Ghana                                    | MG | Madagascar                                   | TJ | Tajikistan               |
| BE | Belgium                  | GN | Guinea                                   | MK | The former Yugoslav<br>Republic of Macedonia | TM | Turkmenistan             |
| BF | Burkina Faso             | GR | Greece                                   |    |                                              | TR | Turkey                   |
| BG | Bulgaria                 | HU | Hungary                                  | ML | Mali                                         | TT | Trinidad and Tobago      |
| BJ | Benin                    | IE | Ireland                                  | MN | Mongolia                                     | UA | Ukraine                  |
| BR | Brazil                   | IL | Israel                                   | MR | Mauritania                                   | UG | Uganda                   |
| BY | Belarus                  | IS | Iceland                                  | MW | Malawi                                       | US | United States of America |
| CA | Canada                   | IT | Italy                                    | MX | Mexico                                       | UZ | Uzbekistan               |
| CF | Central African Republic | JP | Japan                                    | NE | Niger                                        | VN | Viet Nam                 |
| CG | Congo                    | KE | Kenya                                    | NL | Netherlands                                  | YU | Yugoslavia               |
| CH | Switzerland              | KG | Kyrgyzstan                               | NO | Norway                                       | ZW | Zimbabwe                 |
| CI | Côte d'Ivoire            | KP | Democratic People's<br>Republic of Korea | NZ | New Zealand                                  |    |                          |
| CM | Cameroon                 |    |                                          | PL | Poland                                       |    |                          |
| CN | China                    | KR | Republic of Korea                        | PT | Portugal                                     |    |                          |
| CU | Cuba                     | KZ | Kazakstan                                | RO | Romania                                      |    |                          |
| CZ | Czech Republic           | LC | Saint Lucia                              | RU | Russian Federation                           |    |                          |
| DE | Germany                  | LI | Liechtenstein                            | SD | Sudan                                        |    |                          |
| DK | Denmark                  | LK | Sri Lanka                                | SE | Sweden                                       |    |                          |
| EE | Estonia                  | LR | Liberia                                  | SG | Singapore                                    |    |                          |

## SOURCE CODING ENHANCEMENT USING SPECTRAL-BAND REPLICATION

### TECHNICAL FIELD

In source coding systems, digital data is compressed before transmission or storage to reduce the required bitrate or storing capacity. The present invention relates to a new method and apparatus for the improvement of source coding systems by means of Spectral Band Replication (SBR). Substantial bitrate reduction is achieved while maintaining the same perceptual quality or conversely, an improvement in perceptual quality is achieved at a given bitrate. This is accomplished by means of spectral bandwidth reduction at the encoder side and subsequent spectral band replication at the decoder, whereby the invention exploits new concepts of signal redundancy in the spectral domain.

### BACKGROUND OF THE INVENTION

Audio source coding techniques can be divided into two classes: natural audio coding and speech coding. Natural audio coding is commonly used for music or arbitrary signals at medium bitrates, and generally offers wide audio bandwidth. Speech coders are basically limited to speech reproduction but can on the other hand be used at very low bitrates, albeit with low audio bandwidth. Wideband speech offers a major subjective quality improvement over narrow band speech. Increasing the bandwidth not only improves intelligibility and naturalness of speech, but also facilitates speaker recognition. Wideband speech coding is thus an important issue in next generation telephone systems. Further, due to the tremendous growth of the multimedia field, transmission of music and other non-speech signals at high quality over telephone systems is a desirable feature.

A high-fidelity linear PCM signal is very inefficient in terms of bitrate versus the perceptual entropy. The CD standard dictates 44.1 kHz sampling frequency, 16 bits per sample resolution and stereo. This equals a bitrate of 1411 kbit/s. To drastically reduce the bitrate, source coding can be performed using split-band perceptual audio codecs. These natural audio codecs exploit perceptual irrelevancy and statistical redundancy in the signal. Using the best codec technology, approximately 90% data reduction can be achieved for a standard CD-format signal with practically no perceptible degradation. Very high sound quality in stereo is thus possible at around 96 kbit/s, i.e. a compression factor of approximately 15:1. Some perceptual codecs offer even higher compression ratios. To achieve this, it is common to reduce the sample-rate and thus the audio bandwidth. It is also common to decrease the number of quantization levels, allowing occasionally audible quantization distortion, and to employ degradation of the stereo field, through intensity coding. Excessive use of such methods results in annoying perceptual degradation. Current codec technology is near saturation and further progress in coding gain is not expected. In order to improve the coding performance further, a new approach is necessary.

The human voice and most musical instruments generate quasistationary signals that emerge from oscillating systems. According to Fourier theory, any periodic signal may be expressed as a sum of sinusoids with the frequencies  $f$ ,  $2f$ ,  $3f$ ,  $4f$ ,  $5f$  etc. where  $f$  is the fundamental frequency. The frequencies form a harmonic series. A bandwidth limitation of such a signal is equivalent to a truncation of the harmonic series. Such a truncation alters the perceived timbre, tone colour, of a musical instrument or voice, and yields an audio signal that will sound "muffled" or "dull", and intelligibility may be reduced. The high frequencies are thus important for the subjective impression of sound quality.

Prior art methods are mainly intended for improvement of speech codec performance and in particular intended for High Frequency Regeneration (HFR), an issue in speech coding. Such methods employ broadband linear frequency shifts, non-linearities or aliasing [U.S. Pat. 5,127,054] generating intermodulation products or other non-harmonic frequency components which cause severe dissonance when applied to music signals. Such dissonance is referred to in the speech coding literature as "harsh" and "rough" sounding. Other synthetic speech HFR methods generate sinusoidal harmonics that are based on fundamental pitch estimation and are thus limited to tonal stationary sounds [U.S. Pat. 4,771,465]. Such prior art methods, although useful for low-quality speech applications, do not work for high quality speech or music signals. A few methods attempt to improve the performance of high quality audio source codecs. One uses synthetic noise signals generated at the decoder to substitute noise-like signals in speech or music previously discarded by the encoder ["Improving Audio Codecs by Noise Substitution" D. Schultz, JAES, Vol. 44, No. 7/8, 1996]. This is performed within an otherwise normally transmitted highband at an intermittent basis when noise signals are present. Another method recreates some missing highband harmonics that were lost in the coding process ["Audio Spectral Coder" A.J.S. Ferreira, AES Preprint 4201, 100<sup>th</sup> Convention, May 11-14 1996, Copenhagen] and is again dependent on tonal signals and pitch detection. Both methods operate at a low duty-cycle basis offering comparatively limited coding or performance gain.

#### SUMMARY OF THE INVENTION

The present invention provides a new method and an apparatus for substantial improvements of digital source coding systems and more specifically for the improvements of audio codecs. The objective includes bitrate reduction or improved perceptual quality or a combination thereof. The invention is based on new methods exploiting harmonic redundancy, offering the possibility to discard passbands of a signal prior to transmission or storage. No perceptual degradation is perceived if the decoder performs high quality spectral replication according to the invention. The discarded bits represent the coding gain at a fixed perceptual quality. Alternatively, more bits can be allocated for encoding of the lowband information at a fixed bitrate, thereby achieving a higher perceptual quality.

The present invention postulates that a truncated harmonic series can be extended based on the direct relation between lowband and highband spectral components. This extended series resembles the original in a perceptual sense if certain rules are followed: First, the extrapolated spectral components must be harmonically related to the truncated harmonic series, in order to avoid dissonance-related artifacts. The present invention uses transposition as a means for the spectral replication process, which ensures that this criterion is met. It is however not necessary that the lowband spectral components form a harmonic series for successful operation, since new replicated components, harmonically related to those of the lowband, will not alter the noise-like or transient nature of the signal. A transposition is defined as a transfer of partials from one position to another on the musical scale while maintaining the frequency ratios of the partials. Second, the spectral envelope, i.e. the coarse spectral distribution, of the replicated highband, must reasonably well resemble that of the original signal. The present invention offers two modes of operation, SBR-1 and SBR-2, that differ in the way the spectral envelope is adjusted.

SBR-1, intended for the improvement of intermediate quality codec applications, is a single-ended process which relies exclusively on the information contained in a received lowband or lowpass signal at the decoder. The spectral envelope of this signal is determined and extrapolated, for instance using polynomials together with a set of rules or

a codebook. This information is used to continuously adjust and equalise the replicated highband. The present SBR-1 method offers the advantage of post-processing, i.e. no modifications are needed at the encoder side. A broadcaster will gain in channel utilisation or will be able to offer improved perceptual quality or a combination of both. Existing bitstream syntax and standards can be used without modification.

SBR-2, intended for the improvement of high quality codec applications, is a double-ended process where, in addition to the transmitted lowband signal according to SBR-1, the spectral envelope of the highband is encoded and transmitted. Since the variations of the spectral envelope has a much lower rate than the highband signal components, only a limited amount of information needs to be transmitted in order to successfully represent the spectral envelope. SBR-2 can be used to improve the performance of current codec technologies with no or minor modifications of existing syntax or protocols, and as a valuable tool for future codec development.

Both SBR-1 and SBR-2 can be used to replicate smaller passbands of the lowband when such bands are shut down by the encoder as stipulated by the psychoacoustic model under bit-starved conditions. This results in improvement of the perceptual quality by spectral replication within the lowband in addition to spectral replication outside the lowband. Further, SBR-1 and SBR-2 can also be used in codecs employing bitrate scalability, where the perceptual quality of the signal at the receiver varies depending on transmission channel conditions. This usually implies annoying variations of the audio bandwidth at the receiver. Under such conditions, the SBR methods can be used successfully in order to maintain a constantly high bandwidth, again improving the perceptual quality.

The present invention operates on a continuous basis, replicating any type of signal contents, i.e. tonal or non-tonal (noise-like and transient signals). In addition, the present spectral replication method creates a perceptually accurate replica of the discarded bands from available frequency bands at the decoder. Hence, the SBR method offers a substantially higher level of coding gain or perceptual quality improvement compared to prior art methods. The invention can be combined with such prior art codec improvement methods; however, no performance gain is expected due to such combinations.

The SBR-method comprises the following steps:

- encoding of a signal derived from an original signal, where frequency bands of the signal are discarded and the discarding is performed prior to or during encoding, forming a first signal,
- during or after decoding of the first signal, transposing frequency bands of the first signal, forming a second signal,
- performing spectral envelope adjustment, and
- combining the decoded signal and the second signal, forming an output signal.

The passbands of the second signal may be set not to overlap or partly overlap the passbands of the first signal, and may be set in dependence of the temporal characteristics of the original signal and/or the first signal, or transmission channel conditions. The spectral envelope adjustment is performed based on estimation of the original spectral envelope from said first signal or on transmitted envelope information of the original signal.

The present invention includes to basic types of transposers: multiband transposers and time-variant pattern search prediction transposers, having different properties. A basic multiband transposition may be performed according to the present invention by the following:

- filtering the signal to be transposed through a set of  $N \geq 2$  bandpass filters with passbands comprising the frequencies  $[f_1, \dots, f_N]$  respectively, forming  $N$  bandpass signals,
- shifting the bandpass signals in frequency to regions comprising the frequencies  $M [f_1, \dots, f_N]$  where  $M \neq 1$  is the transposition factor, and
- combining the shifted bandpass signals, forming the transposed signal.

Alternatively, this basic multiband transposition may be performed according to the invention by the following:

- bandpass filtering the signal to be transposed signal using an analysis filterbank or transform of such a nature that real- or complex-valued subband signals of lowpass type are generated,
- an arbitrary number of channels  $k$  of said analysis filterbank or transform are connected to channels  $Mk$ ,  $M \neq 1$ , in a synthesis filterbank or transform, and
- the transposed signal is formed using the synthesis filterbank or transform.

An improved multiband transposition according to the invention incorporates phase adjustments, enhancing the performance of the basic multiband transposition.

The time-variant pattern search prediction transposition according to the present invention may be performed by the following:

- performing transient detection on the first signal,
- determining which segment of the first signal to be used when duplicating/discarding parts of the first signal depending on the outcome of the transient detection,
- adjusting statevector and codebook properties depending on the outcome of the transient detection, and
- searching for synchronisation points in chosen segment of the first signal, based on the synchronisation point found in the previous synchronisation point search.

The SBR methods and apparatuses according to the present invention offer the following features:

1. The methods and apparatuses exploit new concepts of signal redundancy in the spectral domain.
2. The methods and apparatuses are applicable on arbitrary signals.
3. Each harmonic set is individually created and controlled.
4. All replicated harmonics are generated in such a manner as to form a continuation of the existing harmonic series.
5. The spectral replication process is based on transposition and creates no or imperceptible artifacts.
6. The spectral replication can cover multiple smaller bands and/or a wide frequency range.
7. In the SBR-1 method, the processing is performed at the decoder side only, i.e. all standards and protocols can be used without modification.
8. The SBR-2 method can be implemented in accordance with most standards and protocols with no or minor modifications.
9. The SBR-2 method offers the codec designer a new powerful compression tool.
10. The coding gain is significant.

The most attractive application relates to the improvement of various types of low bitrate codecs, such as MPEG 1/2 Layer I/II/III [U.S. Pat. 5,040,217], MPEG 2/4 AAC, Dolby AC-2/3, NTT TwinVQ [U.S. Pat. 5,684,920], AT&T/Lucent PAC etc. The invention is also useful in high-quality speech codecs such as wide-band CELP and SB-ADPCM G.722 etc. to improve perceived quality. The above codecs are widely used in multimedia, in the telephone industry, on the Internet as well as in professional applications. T-DAB (Terrestrial Digital Audio Broadcasting) systems use low bitrate protocols that will gain in channel utilisation by using the present method, or improve quality in FM and AM DAB. Satellite S-DAB will gain considerably, due to the excessive system costs involved, by using the present method to increase the number of programme channels in the DAB multiplex. Furthermore, for the first time, full bandwidth audio real-time streaming over the Internet is achievable using low bitrate telephone modems.

## BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described by way of illustrative examples, not limiting the scope or spirit of the invention, with reference to the accompanying drawings, in which:

- Fig. 1 illustrates SBR incorporated in a coding system according to the present invention;
- Fig. 2 illustrates spectral replication of upper harmonics according to the present invention;
- Fig. 3 illustrates spectral replication of inband harmonics according to the present invention;
- Fig. 4 is a block diagram for a time-domain implementation of a transposer according to the present invention;
- Fig. 5 is a flow-chart representing a cycle of operation for the pattern-search prediction transposer according to the present invention;
- Fig. 6 is a flow-chart representing the search for synchronisation point according to the present invention;
- Fig. 7a – 7b illustrates the codebook positioning during transients according to the present invention;
- Fig. 8 is a block diagram for an implementation of several time-domain transposers in connection with a suitable filterbank, for SBR operation according to the present invention;
- Fig. 9a – 9c are block diagrams representing a device for STFT analysis and synthesis configured for generation of 2<sup>nd</sup> order harmonics according to the present invention;
- Fig. 10a – 10b are block diagrams of one sub-band with a linear frequency shift in the STFT device according to the present invention;
- Fig. 11 shows one sub-band using a phase-multiplier according to the present invention;
- Fig. 12 illustrates how 3<sup>rd</sup> order harmonics are generated according to the present invention;
- Fig. 13 illustrates how 2<sup>nd</sup> and 3<sup>rd</sup> order harmonics are generated simultaneously according to the present invention;
- Fig. 14 illustrates generation of a non-overlapping combination of several harmonic orders according to the present invention;
- Fig. 15 illustrates generation of an interleaved combination of several harmonic orders according to the present invention;
- Fig. 16 illustrates generation of broadband linear frequency shifts;
- Fig. 17 illustrates how sub-harmonics are generated according to the present invention;
- Fig. 18a – 18b are block diagrams of a perceptual codec;
- Fig. 19 shows a basic structure of a maximally decimated filterbank;

Fig. 20 illustrates generation of 2<sup>nd</sup> order harmonics in a maximally decimated filterbank according to the present invention;

Fig. 21 is a block diagram for the improved multiband transposition in a maximally decimated filterbank operating on subband signals according to the present invention;

5 Fig. 22 is a flow-chart representing the improved multiband transposition in a maximally decimated filterbank operating on subband signals according to the present invention;

Fig. 23 illustrates subband samples and scalefactors of a typical codec;

Fig. 24 illustrates subband samples and envelope information for SBR-2 according to the present invention;

Fig. 25 illustrates hidden transmission of envelope information in SBR-2 according to the present invention;

10 Fig. 26 illustrates redundancy coding in SBR-2 according to the present invention;

Fig. 27 illustrates an implementation of a codec using the SBR-1 method according to the present invention; and

Fig. 28 illustrates an implementation of a coded using the SBR-2 method according to the present invention;

Fig. 29 is a block diagram of a "pseudo-stereo" generator according to the present invention.

## 15 DESCRIPTION OF PREFERRED EMBODIMENTS

Throughout the explanation of the embodiments herein, emphasis is given to natural audio source coding applications. However, it should be understood that the present invention is applicable on a range of source coding applications other than that of encoding and decoding audio signals.

### 20 Transposition basics

Transposition as defined according to the present invention, is the ideal method for spectral replication, and has several major advantages over prior art, such as: no pitch detection is required, equally high performance for single-pitched and polyphonic programme material is obtained, and the transposition works equally well for tonal and non-tonal signals. Contrary to other methods, the transposition according to the invention can be used in arbitrary audio source coding systems for arbitrary signal types.

An exact transposition a factor  $M$  of a discrete time signal  $x(n)$  in the form of a sum of cosines with time varying amplitudes, is defined by the relation

$$x(n) = \sum_{i=0}^{N-1} e_i(n) \cos(2\pi f_i n / f_s + \alpha_i) \rightarrow \quad (1)$$

$$30 \quad y(n) = \sum_{i=0}^{N-1} e_i(n) \cos(2\pi M f_i n / f_s + \beta_i) \quad (2)$$

where  $N$  is the number of sinusoids, hereafter referred to as partials,  $f_i$ ,  $e_i(n)$ ,  $\alpha_i$  are the individual input frequencies, time envelopes and phase constants respectively,  $\beta_i$  are the arbitrary output phase constants and  $f_s$  is the sampling frequency, and  $0 \leq M f_i \leq f_s/2$ .

35 In Fig. 2, the generation of  $M^{\text{th}}$  order harmonics, where  $M$  is an integer  $\geq 2$ , is shown. The term  $M^{\text{th}}$  order harmonics is used for simplicity, albeit the process generates  $M^{\text{th}}$  order harmonics to all signals in a certain frequency region, which in most cases are themselves harmonics of unknown order. The input signal with the frequency domain

representation  $X(f)$  is bandlimited to the range 0 to  $f_{\max}$ , 201. The signal contents in the range  $f_{\max}/M$  to  $Qf_{\max}/M$ , where  $Q$  is the desired bandwidth expansion factor  $1 < Q \leq M$ , is extracted by means of a bandpass filter, forming a bandpass signal with spectrum  $X_{BP}(f)$  203. The bandpass signal is transposed a factor  $M$ , forming a second bandpass signal with spectrum  $X_T(f)$  covering the range  $f_{\max}$  to  $Qf_{\max}$ , 205. The spectral envelope of this signal is adjusted by means of a programme-controlled equaliser, forming a signal with spectrum  $X_E(f)$  207. This signal is then combined with a delayed version of the input signal in order to compensate for the delay imposed by the bandpass filter and transposer, whereby an output signal with spectrum  $Y(f)$  covering the range 0 to  $Qf_{\max}$  is formed 209. Alternatively, bandpass filtering may be performed after the transposition  $M$ , using cut-off frequencies  $f_{\max}$  and  $Qf_{\max}$ . By using multiple transposers, simultaneous generation of different harmonic orders is of course possible.

The above scheme may also be used to "fill in" stopbands within the input signal, as shown in Fig. 3, where the input signal has a stopband extending from  $f_0$  to  $Qf_0$  301. A passband  $[f_0/M, Qf_0/M]$  is then extracted 303, transposed a factor  $M$  to  $[f_0, Qf_0]$  305, envelope adjusted 307 and combined with the delayed input signal forming the output signal with spectrum  $Y(f)$  309.

An approximation of an exact transposition may be used. According to the present invention, the quality of such approximations is determined using dissonance theory. A criterion for dissonance is presented by Plomp ["Tonal Consonance and Critical Bandwidth" R. Plomp, W. J. M. Levelt JASA, Vol 38, 1965], and states that two partials are considered dissonant if the frequency difference is within approximately 5 to 50% of the bandwidth of the critical band in which the partials are situated. For reference, the critical bandwidth for a given frequency can be approximated by

$$cb(f) = 25 + 75 \left( 1 + 1.4 \left( \frac{f}{1000} \right)^2 \right)^{0.69} \quad (3)$$

with  $f$  and  $cb$  in Hz. Further, Plomp states that the human auditory system can not discriminate two partials if they differ in frequency by approximately less than five percent of the critical bandwidth in which they are situated. The exact transposition in Eq. 2 is approximated by

$$y_{approx}(n) = \sum_{i=0}^{N-1} e_i(n) \cos(2\pi(Mf_i \pm \Delta f_i)n / f_s + \beta_i), \quad (4)$$

where  $\Delta f_i$  is the deviation from the exact transposition. If the input partials form a harmonic series, a hypothesis of the invention states that the deviations from the harmonic series of the transposed partials must not exceed five percent of the critical bandwidth in which they are situated. This would explain why prior art methods give unsatisfactory "harsh" and "rough" results, since broad band linear frequency shifts yields a much larger deviation than acceptable. When prior art methods produce more than one partial for only one input partial, the partials must nevertheless be within the above stated deviation limit, as to be perceived as one partial. This again explains the poor results obtained with prior art methods using nonlinearities etc, since they produce intermodulation partials not within the limit of deviation.

When using the above transposition based method of spectral replication according to the present invention, the following important properties are achieved:

- Normally, no frequency domain overlap occur between replicated harmonics and existing partials.

- The replicated partials are harmonically related to the partials of the input signal and will not give rise to any annoying dissonance or artifacts.
- The spectral envelope of the replicated harmonics forms a smooth continuation of the input signal spectral envelope, perceptually matching the original envelope.

5

#### Transposition based on time-variant pattern search prediction

Various ways to design the required transposers exist. Typical time-domain implementations expand the signal in time by duplicating signal segments based on the pitch-period. This signal is subsequently read out at a different rate. Unfortunately such methods are strictly dependent on pitch-detection for accurate time splicing of the signal segments. Furthermore, the constraint to work on pitch-period based signal segments makes them sensitive to transients. Since the detected pitch-period can be much longer than the actual transient, the risk of duplicating the entire transient rather than just expanding it in time is obvious. Another type of time domain algorithms obtains time expansion/compression of speech signals by utilising pattern search prediction of the output signal ["Pattern Search Prediction of Speech" R. Bogner, T. Li, Proc. ICASSP '89, Vol. 1, May 1989, "Time-Scale Modification of Speech based on a nonlinear Oscillator Model" G. Kubin, W. B. Kleijn, IEEE, 1994]. This is a form of granular synthesis, where the input signal is divided into small parts, granules, used to synthesise the output signal. This synthesis is usually done by performing correlation of signal segments in order to determine the best splicing points. This means that the segments used to form the output signal are not dependent on the pitch period, and thus the non-trivial task of pitch detection is not required. Nevertheless, problems with rapidly changing signal amplitudes remain in these methods, and high quality transposition tends to raise high computational demands. However, an improved time-domain pitch shifter/transposer is now presented, where the use of transient detection and dynamic system parameters produces a more accurate transposition for high transposition factors during both stationary (tonal or non-tonal) and transient sounds, at a low computational cost.

Referring to the drawings wherein like numerals indicate like elements, there is shown in Fig. 4 nine separate modules: a transient-detector 401, a window position adjuster 403, a codebook generator 405, a synchronisation signal selector 407, a synchronisation position memory 409, a minimum difference estimator 411, an output segment memory 413, a mix unit 415, and a down sampler 417. The input signal is fed to both the codebook generator 405 and the transient-detector 401. If a transient is detected, its position is sent to the window position module 403. This module stipulates the size and position of the window that is multiplied with the input signal when creating the codebook. The codebook generator 405 receives a synchronisation position from the sync select module 407, provided it has been connected to another transposer. If this synchronisation position is within the codebook, it is used and an output segment is produced. Otherwise the codebook is sent to the minimum difference estimator 411 which returns a new synchronisation position. The new output segment is windowed together with the previous output segment in the mix module 415 and subsequently down sampled in module 417.

In order to clarify the explanation, a state space representation is introduced. Here, the state vectors, or granules, represent the input and output signals. The input signal is represented by a statevector  $\mathbf{x}(n)$ :

$$\mathbf{x}(n) = [x(n), x(n-D), x(n-2D), \dots, x(n-(N-1)D)] \quad (5)$$

which is obtained from  $N$  delayed samples of the input signal, where  $N$  is the dimension of the state vector and  $D$  is the delay between the input samples used to build the vector. The granular mapping yields the sample  $x(n)$  following each statevector  $\mathbf{x}(n-1)$ . This gives Eq. 6, where  $a(\cdot)$  is the mapping:

$$x(n) = a(\mathbf{x}(n-1)). \quad (6)$$

- 5 In the present method the granular mapping is used to determine the next output based on the former output, using a state transition codebook. The codebook of length  $L$  is continuously rebuilt containing the statevectors and the next sample following each statevector. Each statevector is separated from its neighbour by  $K$  samples; this enables the system to adjust the time resolution depending on the characteristics of the currently processed signal, where  $K$  equal to one represents the finest resolution. The input signal segment used to build the codebook is chosen based on  
10 the position of a possible transient and the synchronisation position in the previous codebook.

This means that the mapping  $a(\cdot)$ , theoretically, is evaluated for all transitions included in the codebook:

$$a \left( \begin{bmatrix} \mathbf{x}(n-L) \\ \mathbf{x}(n-L+K) \\ \vdots \\ \mathbf{x}(n-1) \end{bmatrix} \right) = \begin{bmatrix} x(n-L+1) \\ x(n-L+K+1) \\ \vdots \\ x(n) \end{bmatrix}. \quad (7)$$

- 15 With this transition codebook, the new output  $y(n)$  is calculated by searching for the statevector in the codebook most similar to the current statevector  $\mathbf{y}(n-1)$ . This nearest-neighbour search is done by calculating the minimum difference and gives the new output sample:

$$y(n) = a(\mathbf{y}(n-1)). \quad (8)$$

- 20 However, the system is not limited to work on a sample by sample basis, but is preferably operated on a segment by segment basis. The new output segment is windowed and added, mixed, with the previous output segment, and subsequently down sampled. The pitch transposition factor is determined by the ratio of the input segment length represented by the codebook and the output segment length read out of the codebook.

- Returning to the drawings, in Fig. 5 and Fig. 6 flowcharts are presented, displaying the cycle of operation of the transposer. In 501 the input data is represented, a transient detection 503 is performed on a segment of the input  
25 signal; the search for transients is performed on a segment length equal to the output segment length. If a transient is found 505, the position of the transient is stored 507 and the parameters  $L$  (representing the codebook length),  $K$  (representing the distance in samples between each statevector), and  $D$  (representing the delay between samples in each statevector) are adjusted 509. The position of the transient is compared to the position of the previous output segment 511, in order to determine whether the transient has been processed. If so 513, the position of the codebook  
30 (window  $L$ ), and the parameters  $K$ ,  $L$ , and  $D$  are adjusted 515. After the necessary parameter adjustments, based on the outcome of the transient detection, the search for a new synchronisation, or splicing point takes place 517. This procedure is displayed in Fig. 6. First a new synchronisation point is calculated based on the previous 601, according to:

$$Sync\_pos = Sync\_pos\_old + S \cdot M - S, \quad (9)$$

- 35 where  $Sync\_pos$  and  $Sync\_pos\_old$  are the new and old synchronisation positions respectively,  $S$  is the length of the input segment being processed, and  $M$  is the transposition factor. This synchronisation point is used to compare the

accuracy of the new splicing point with the accuracy of the old splicing point 603. If the match is as good as or better than the previous 605, this new synchronisation point is returned 607 provided it is within the codebook. If not, a new synchronisation point is searched for in the loop 609. This is performed with a similarity measure, in this case a minimum difference function 611, however, it is also possible to use correlation in the time- or frequency-domain. If the position yields a better match than that of the previous position found 613 the synchronisation position is stored 615. When all positions are tried 617 the system returns 619 to the flowchart in Fig. 5. The new synchronisation point obtained is stored 519 and a new segment is read out from the codebook 521 starting at the given synchronisation point. This segment is windowed and added to the previous 523, down sampled by the transposition factor 525, and stored in the output buffer 527.

In Fig. 7 the behaviour of the system under transient conditions regarding the position of the codebook is illustrated. Prior to the transient, the codebook 1 representing the input segment 1 is positioned "to the left" of segment 1. Correlation segment 1 represents a part of the previous output and is used to find synchronisation point 1 in codebook 1. When the transient is detected, and the point of the transient is processed, the codebook is moved according to Fig. 7a and is stationary until the input segment currently being processed is once again "to the right" of the codebook. This makes it impossible to duplicate the transient since the system is not allowed to search for synchronisation points prior to the transient.

Most pitch transposers, or time expanders, based on pattern search prediction give satisfactory results for speech and single-pitched material. However, their performance deteriorates rapidly for high complexity signals, like music, in particular at large transposition factors. The present invention offers several solutions for improved performance, therefore producing excellent results for any type of signal. Contrary to other designs, the system is time-variant and the system parameters are based on the properties of the input signal, and the parameters used during the previous operation cycle. The use of a transient detector controlling not only the codebook size and position, but also the properties of the statevectors included, is a very robust and computationally efficient method to avoid audible degradation during rapidly changing signal segments. Furthermore, alteration of the length of the signal segment being processed, which would raise higher computational demands, is not required. Also, the present invention utilises a refined codebook search based on the results from the preceding search. This means that contrary to an ordinary correlation of two signal segments, as is usually done in time-domain systems based on pattern search prediction, the most likely synchronisation positions are tried first instead of trying all positions consecutively. This new method for reducing the codebook search drastically reduces the computational complexity of the system. Further, when using several transposers, synchronisation position information can be shared among the transposers for further reduction of the computational complexity, as shown in the following implementation.

The time-domain transposers as explained above are used to implement the SBR-1 and SBR-2 systems according to the following, illustrative but not limiting, example. In Fig. 8 three time expansion modules are used in order to generate second, third and fourth order harmonics. Since, in this example, each time domain expansion /transposer works on a wideband signal, it is beneficial to adjust the spectral envelope of the source frequency range prior to transposition, considering that there will be no means to do so after the transpositions, without adding a separate equaliser system. The spectral envelope adjusters, 801, 803 and 805, each work on several filterbank channels. The gain of each channel in the envelope adjusters must be set so that the sum, 813, 815, 817, at the output, after

transposition, yields the desired spectral envelope. The transposers 807, 809 and 811 are interconnected in order to share synchronisation position information. This is based on the fact that under certain conditions, a high correlation will occur between the synchronisation positions found in the codebook during correlation in the separate transposing units. Assume, as an example and again not limiting the scope of the invention, the fourth order harmonic transposer works on a time frame basis half of that of the second order harmonic transposer but at twice the duty cycle. Assume further, that the codebooks used for the two expanders are the same and that the synchronisation positions of the two time-domain expanders are labelled *sync\_pos4* and *sync\_pos2*, respectively. This yields the following relation:

$$\text{sync\_pos2} = \text{sync\_pos4} - n \cdot 4 \cdot S - \text{sync\_offset}, \text{ for } n=1,2,3,4,\dots, \quad (10)$$

10 where

$$\text{sync\_offset} = \text{sync\_pos4} - \text{sync\_pos2}, \text{ for } n=0, \quad (11)$$

and *S* is the length of the input segment represented by the codebook. This is valid as long as neither of the synchronisation position pointers reaches the end of the codebook. During normal operation *n* is increased by one for each time-frame processed by the second order harmonic transposer, and when the codebook end inevitably is reached, by either of the pointers, the counter *n* is set to *n*=0, and *sync\_pos2* and *sync\_pos4* are computed individually. Similar results are obtained for the third order harmonic transposer when connected to the fourth order harmonic transposer.

20 The above-presented use of several interconnected time-domain transposers, for the creation of higher order harmonics, introduces substantial computational reduction. Furthermore, the proposed use of time-domain transposers in connection with a suitable filterbank, presents the opportunity to adjust the envelope of the created spectrum while maintaining the simplicity and low computational cost of a time domain transposer, since these, more or less, may be implemented using fixed point arithmetic and solely additive/subtractive-operations.

25 Other, illustrative but not limiting, examples of the present invention are:

- the use of a time domain transposer within each subband in a subband filter bank, thus reducing the signal complexity for each transposer.
- the use of a time domain transposer in combination with a frequency domain transposer, thus enabling the system to use different methods for transposition depending on the characteristics of the input signal being processed.
- the use of a time domain transposer in a wideband speech codec, operating on for instance the residual signal obtained after linear prediction.

35 It should be recognised that the method outlined above may be advantageously used for timescale modification only, by simply omitting the sample rate conversion. Further it is understood, that although the outlined method focuses on pitch transposing to a higher pitch, i.e. time expansion, the same principles apply when transposing to a lower pitch, i.e. time compression, as is obvious to those skilled in the art.

### Filter bank based transposition

Various new and innovative filter bank based transposition techniques will now be described. The signal to be transposed is divided into a series of BP- or subband signals. The subband signals are then transposed, exact or approximately, which is advantageously accomplished by a reconnection of analysis- and synthesis subbands, hereinafter referred to as a "patch". The method is first demonstrated using a Short Time Fourier Transform, STFT.

The N-point STFT of a discrete-time signal  $x(n)$  is defined by

$$X_k(n) = \sum_{p=-\infty}^{\infty} x(p) h(n-p) e^{-j\omega_k p}, \quad (12)$$

where  $k = 0, 1, \dots, N-1$  and  $\omega_k = 2\pi k/N$  and  $h(n)$  is a window. If the window satisfies the following conditions

$$\begin{cases} h(0) = 1 \\ h(n) = 0 \quad \text{for } n = \pm N, \pm 2N, \pm 3N, \dots \end{cases} \quad (13)$$

an inverse transform exists and is given by

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_k(n) e^{j\omega_k n}. \quad (14)$$

The direct transform may be interpreted as an analyser, see Fig. 9a, consisting of a bank of  $N$  BP-filters with impulse responses  $h(n)\exp(j\omega_k n)$  followed by a bank of  $N$  multipliers with carriers  $\exp(-j\omega_k n)$  which shift the BP-signals down to regions around 0 Hz, forming the  $N$  analysis signals  $X_k(n)$ . The window acts as a prototype LP-filter.  $X_k(n)$  have small bandwidths and are normally downsampled. Eq. 12 need thus only be evaluated at  $n = rR$ , where  $R$  is the decimation factor and  $r$  is the new time variable.  $X_k(n)$  can be recovered from  $X_k(rR)$  by upsampling, see Fig 9b, i.e. insertion of zeros followed by LP-filtering. The inverse transform may be interpreted as a synthesiser consisting of a bank of  $N$  multipliers with carriers  $(1/N)\exp(j\omega_k n)$  that shift the signals  $X_k(n)$  up to their original frequencies, followed by stages that add the contributions  $y_k(n)$  from all channels. The STFT and ISTFT may be rearranged in order to use the DFT and IDFT, which makes the use of FFT algorithms possible ["Implementation of the Phase Vocoder using the Fast Fourier Transform" M. R. Portnoff, IEEE ASSP, Vol. 24, No. 3, 1976].

Fig. 9c shows a patch for generation of second harmonics,  $M = 2$ , with  $N = 32$ . For the sake of simplicity, only channels 0 through 16 are shown. The centre frequency of BP 16 equals the Nyquist frequency, channels 17 through 31 correspond to negative frequencies. The blocks denoted P and the gain blocks will be described later and should presently be considered shorted out. The input signal is in this example bandlimited so that only channels 0 through 7 contain signals. Analyser channels 8 through 16 are thus empty and need not be mapped to the synthesiser. Analyser channels 0 through 7 are connected to synthesiser channels 0 through 7, corresponding to an input signal delay path. Analysis channels  $k$  where  $4 \leq k \leq 7$  are also connected to synthesis channels  $Mk$ ,  $M = 2$ , which shift the signals to frequency regions at two times the centre-frequencies of BP filters  $k$ . Hence, the signals are upshifted to their original ranges as well as transposed one octave up. To explore the harmonic generation in terms of real-valued filter responses and modulators the negative frequencies must also be considered, see the lower branch of Fig. 10a. Hence, the combined output of the remapping  $k \rightarrow Mk$  and  $N-k \rightarrow N-Mk$  where  $4 \leq k \leq 7$  must be evaluated.

This yields

$$y(n) = \frac{2}{N} [x(n) * h(n) \cos(\omega_k n)] \cos((M-1)\omega_k n) + \frac{2}{N} [x(n) * h(n) \sin(\omega_k n)] \sin((M-1)\omega_k n) \quad (15)$$

where  $M=2$ . Eq. 15 may be interpreted as a BP-filtering of the input signal, followed by a linear frequency shift or Upper Side Band (USB) modulation, i.e. single side band modulation using the upper side band, see Fig. 10b, where 1005 and 1007 form a Hilbert transformer, 1009 and 1011 are multipliers with cosine and sine carriers and 1013 is a difference stage which selects the upper sideband. Clearly, such a multiband BP and SSB method may be implemented explicitly, i.e. without filterbank patching, in the time or frequency domain, allowing arbitrary selection of individual passbands and oscillator frequencies.

According to Eq. 15, a sinusoid with the frequency  $\omega_i$  within the passband of analysis channel  $k$  yields a harmonic at the frequency  $M\omega_k + (\omega_i - \omega_k)$ . Hence the method, referred to as basic multiband transposition, only generates exact harmonics for input signals with frequencies  $\omega_i = \omega_k$ , where  $4 \leq k \leq 7$ . However, if the number of filters is sufficiently large, the deviation from an exact transposition is negligible, see Eq. 4. Further, the transposition is made exact for quasi-stationary tonal signals of arbitrary frequencies by inserting the blocks denoted P 917 (Fig. 9c), provided every analysis channel contains maximum one partial. In this case  $X_k(rR)$  are complex exponentials with frequencies equal to the differences between the partial frequencies  $\omega_i$  and the centre frequencies  $\omega_k$  of the analysis filters. To obtain the exact transposition  $M$ , these frequencies must be increased by a factor  $M$ , modifying the above frequency relationship to  $\omega_i \rightarrow M\omega_k + M(\omega_i - \omega_k) = M\omega_i$ . The frequencies of  $X_k(rR)$  are equal to the time derivatives of their respective unwrapped phase angles and may be estimated using first order differences of successive phase angles. The frequency estimates are multiplied by  $M$  and synthesis phase angles are calculated using those new frequencies. However, the same result, aside from a phase constant, is obtained in an simplified way by multiplying the analysis arguments by  $M$  directly, eliminating the need for frequency estimation. This is described in Fig. 11, representing the blocks 917. Thus  $X_k(rR)$ , where  $4 \leq k \leq 7$  in this example, are converted from rectangular to polar coordinates, illustrated by the blocks  $R \rightarrow P$ , 1101. The arguments are multiplied by  $M=2$  1103 and the magnitudes are unaltered. The signals are then converted back to rectangular coordinates ( $P \rightarrow R$ ) 1105 forming the signals  $Y_{Mk}(rR)$  and fed to synthesiser channels according to Fig. 9c. This improved multiband transposition method thus has two stages: The patch provides a coarse transposition, as in the basic method, and the phase-multipliers provide fine frequency corrections. The above multiband transposition methods differ from traditional pitch shifting techniques using the STFT, where lookup-table oscillators are used for the synthesis or, when the ISTFT is used for the synthesis the signal is time-stretched and decimated, i.e. no patch is used.

The harmonic patch of Fig. 9c is easily modified for other transposition factors than two. Fig. 12 shows a patch 1203 for generation of 3<sup>rd</sup> order harmonics, where 1201 are the analysis channels and 1205 are the synthesis channels. Different harmonic orders may be created simultaneously as shown in Fig. 13, where 2<sup>nd</sup> and 3<sup>rd</sup> order harmonics are used. Fig. 14 illustrates a non-overlapping combination of 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> order harmonics. The lowest possible harmonic number is used as high in frequency as possible. Above the upper limit of the destination range of harmonic  $M$ , harmonic  $M+1$  is used. Fig. 15 demonstrates a method of mapping all synthesiser channels ( $N=64$ , channels 0 – 32 shown). All highband channels with non prime-number indices are mapped according to the

following relation between source and destination channel number:  $k_{dest} = Mk_{source}$  where  $M$  is the smallest integer  $\geq 2$  that satisfies the condition that  $k_{source}$  lies in the lowband and  $k_{dest}$  in the highband. Hence, no synthesiser channel receives signal from more than one analysis channel. Prime-number highband channels may be mapped to  $k_{source} = 1$  or lowband channels  $k_{source} > 1$  that yield good approximations of the above relation (Only non-prime number connections with  $M = 2, 3, 4, 5$  are shown in Fig. 15).

It is also possible to combine amplitude and phase information from different analyser channels. The amplitude signals  $|X_k(rR)|$  may be connected according to Fig. 16, whereas the phase signals  $\arg\{X_k(rR)\}$  are connected according to the principle of Fig. 16. In this way the lowband frequencies will still be transposed, whereby a periodic repetition of the source region envelope is generated instead of the stretched envelope that results from a transposition according to Eq. 2. Gating or other means may be incorporated in order to avoid amplification of "empty" source channels. Fig. 17 illustrates another application, the generation of sub-harmonics to a highpass filtered or bass limited signal by using connections from higher to lower subbands. When using the above transpositions it may be beneficial to employ adaptive switching of patch based on the characteristics of the signal.

In the above description it was assumed that the highest frequency contained in the input signal was significantly lower than the Nyquist frequency. Thus, it was possible to perform a bandwidth expansion without an increase in sample rate. This is however not always the case, why a preceding upsampling may be necessary. When using filter bank methods for transposition, it is possible to integrate upsampling in the process.

Most perceptual codecs employ maximally decimated filter banks in the time to frequency mapping ["Introduction to Perceptual Coding" K. Brandenburg, AES, Collected Papers on Digital Audio Bitrate Reduction, 1996]. Fig. 18a shows the basic structure of a perceptual encoder system. The analysis filter bank 1801 splits the input signal into several subband signals. The subband samples are individually quantised 1803, using a reduced number of bits, where the number of quantization levels are determined from a perceptual model 1807 which estimates the minimum masking threshold. The subband samples are normalised, coded with optional redundancy coding methods and combined with side information consisting of the normalisation factors, bit-allocation information and other codec specific data 1805, to form the serial bit stream. The bit stream is then stored or transmitted. In the decoder, Fig. 18b, the coded bitstream is demultiplexed 1809, decoded and the subband samples are re-quantised to the equal number of bits 1811. A synthesis filter bank combines the subband samples in order to recreate the original signal 1813. Implementations using maximally decimated filter banks will drastically reduce computational costs. In the following descriptions, there is a focus on cosine modulated filter banks. It should be appreciated however, that the invention can be implemented using other types of filter banks or transforms, including filter bank interpretations of the wavelet transform, other non-equal bandwidth filter banks or transforms and multi-dimensional filter banks or transforms.

In the illustrative, but not limiting, descriptions below it is assumed that an  $L$ -channel cosine modulated filter bank splits the input signal  $x(n)$  into  $L$  subband signals. The generic structure of a maximally decimated filter bank is shown in Fig. 19. The analysis filters are denoted  $H_k(z)$  1901, where  $k = 0, 1, \dots, L-1$ . The subband signals  $v_k(n)$  are maximally decimated 1903, each of sampling frequency  $f_s/L$ , where  $f_s$  is the sampling frequency of  $x(n)$ . The synthesis section reassembles the subband signals after interpolation 1905 and filtering 1907 to produce  $\hat{x}(n)$ . The

synthesis filters are denoted  $F_k(z)$ . In addition, the present invention performs a spectral replication on  $\hat{x}(n)$ , giving an enhanced signal  $y(n)$ .

Synthesising the subband signals with a  $QL$ -channel filter bank, where only the  $L$  lowband channels are used and the bandwidth expansion factor  $Q$  is chosen so that  $QL$  is an integer value, will result in an output bit stream with sampling frequency  $Qf_s$ . Hence, the extended filter bank will act as if it is an  $L$ -channel filter bank followed by an upsampler. Since, in this case, the  $L(Q-1)$  highband filters are unused (fed with zeros), the audio bandwidth will not change – the filter bank will merely reconstruct an upsampled version of  $\hat{x}(n)$ . If, however, the  $L$  subband signals are patched to the highband filters, the bandwidth of  $\hat{x}(n)$  will be increased by a factor  $Q$ , producing  $y(n)$ . This is the maximally decimated filter bank version of the basic multiband transposer, according to the invention. Using this scheme, the upsampling process is integrated in the synthesis filtering as explained earlier. It should be noted that any size of the synthesis filter bank may be used, resulting in different sample-rates of the output signal, and hence different bandwidth expansion factors. Performing spectral replication on  $\hat{x}(n)$  according to the present invention of the basic multiband transposition method with an integer transposition factor  $M$ , is accomplished by patching the subband signals as

$$v_{Mk}(n) = e_{Mk}(n) (-1)^{(M-1)kn} v_k(n), \quad (16)$$

where  $k \in [0, L-1]$  and chosen so that  $Mk \in [L, QL-1]$ ,  $e_{Mk}(n)$  is the envelope correction and  $(-1)^{(M-1)kn}$  is a correction factor for spectral inverted subbands. Spectral inversion results from decimation of subband signals, and the inverted signals may be reinverted by changing sign on every second sample in those channels. Referring to Fig. 20, consider an 16-channel synthesis filter bank, patched 2009 for a transposition factor  $M = 2$ , with  $Q = 2$ . The blocks 2001 and 2003 denote the analysis filters  $H_k(z)$  and the decimators of Fig. 19 respectively. Similarly, 2005 and 2007 are the interpolators and synthesis filters  $F_k(z)$ . Eq. 16 then simplifies to patching of the four upper frequency subband signals of the received data into every second of the eight uppermost channels in the synthesis filter bank. Due to spectral inversion, every second patched subband signal must be frequency inverted before the synthesis. Additionally, the magnitudes of the patched signals must be adjusted 2011 according to the principles of SBR-1 or SBR-2.

Using the basic multiband transposition method according to the present invention, the generated harmonics are in general not exact multiples of the fundamentals. All frequencies but the lowest in every subband differs in some extent from an exact transposition. Further, the replicated spectrum contains zeros since the target interval covers a wider frequency range than the source interval. Moreover, the alias cancellation properties of the cosine modulated filter bank vanishes, since the subband signals are separated in frequency in the target interval. That is, neighbouring subband signals do not overlap in the high-band area. However, aliasing reduction methods, known by those skilled in the art, may be used to reduce this type of artifacts. Advantages of this transposition method are ease of implementation, and the very low computational cost.

To achieve perfect transposition of sinusoids, an effective maximally decimated filter bank solution of the improved multiband transposition method is now presented. The system uses an additional modified analysis filter bank, while the synthesis filter bank is cosine modulated as described by Vaidyanathan ["Multirate Systems and Filter Banks" P. P. Vaidyanathan, Prentice Hall, Englewood Cliffs, New Jersey, 1993, ISBN 0-13-605718-7]. The steps for

operation, using the improved multiband transposition method according to the present invention, based on maximally decimated filter banks, are shown schematically in Fig. 21 and in the flow-chart of Fig. 22 and are as follows:

1. The  $L$  received subband signals are synthesised with a  $QL$ -channel filter bank 2101, 2201, 2203, where the  $L(Q-1)$  upper channels are fed with zeros, to form signal  $x_1(n)$ , which is thus oversampled by the bandwidth expansion factor  $Q$ .
2.  $x_1(n)$  is downsampled by a factor  $Q$ , to form signal  $x_2(n')$  2103, 2205, i. e.  $x_2(n') = x_1(Qn')$ .
3. An integer-value  $K$  is chosen as the size of a synthesis filter bank, constrained so that  $T = KM/Q$  is an integer, where  $T$  is the size of the modified analysis filter bank and  $M$  is the transposition factor 2207, 2209, 2211.  $K$  should preferably be chosen large for stationary (tonal) signals, and smaller for dynamic (transient) signals.
4.  $x_2(n')$  is filtered through a  $T$ -channel modified analysis filter bank 2107, 2213, where the  $T$  analysis filters are exponentially modulated, producing a set of complex-valued subband signals. The subband signals are downsampled by a factor  $T/M$ , giving subband signals  $v_k^{(M)}(n'')$ ,  $k = 0, 1, \dots, T-1$ . Hence, the filter bank will be oversampled by a factor  $M$ .
5. The samples  $v_k^{(M)}(n'')$  are converted to a polar representation (magnitude and phase-angle). The phase-angles are multiplied by the factor  $M$ , and the samples are converted back to a rectangular representation according to the scheme of Fig. 11. The real parts of the complex-valued samples are taken, giving the signals  $s_k^{(M)}(n'')$  2109, 2215. After this operation, the signals  $s_k^{(M)}(n'')$  are critically sampled.
6. The gains of the signals  $s_k^{(M)}(n'')$  are adjusted according to the principles of SBR-1 or SBR-2 2111, 2217.
7. The subband signals  $s_k^{(M)}(n'')$ , where  $k \in [T/M, \min(K, T)-1]$ , are synthesised with an ordinary cosine modulated  $K$ -channel filter bank, where the channels 0 through  $T/M-1$  are fed with zeros 2105, 2221. This produces the signal  $x_3^{(M)}(n)$ .
8.  $x_3^{(M)}(n)$  is finally added to  $x_1(n)$  to give  $y(n)$  2223, which is the desired spectral replicated signal.

Steps 3 to 6 may be repeated for different values of the transposition factor  $M$ , thus adding multiple harmonics to  $x_1(n)$ . This mode of operation is illustrated by the dotted figures of Fig. 21, and in Fig. 22, by iterating the loop over boxes 2211 – 2219. In this case,  $K$  is chosen as to make  $T$  integer-valued for all choices of  $M$  – for integer valued  $M$ 's; preferably select  $K$  as to make  $K/Q$  a positive integer. All subband signals  $s_k^{(M_i)}(n'')$ , where  $i = 1, 2, \dots, m$ , and  $m$  is the number of transposition factors, are added according to

$$s_k(n'') = \sum_{i=1}^m s_k^{(M_i)}(n'') \quad (17)$$

for every applicable  $k$ . In the first iteration of the loop of Fig. 22, the signals  $s_k(n'')$  may be considered to be subband samples of zeros only, where  $k = 0, 1, \dots, K-1$ . In every loop, the new samples are added 2219 to  $s_k(n'')$  as

$$s_k(n'') = s_k(n'') + s_k^{(M_i)}(n''), \quad (18)$$

- where  $k = K/Q, K/Q+1, \dots, \min(K, T)-1$ . The subband signals  $s_k(n'')$  are synthesised once with a  $K$ -channel filter bank according to step 7.

The modified analysis filter bank of step 4, is derived through the theory of cosine modulated filter banks, where the modulated lapped transform (MLT) ["Lapped Transforms for Efficient Transform/Subband Coding" H. S. Malvar, IEEE Trans ASSP, vol. 38, no. 6, 1990] is a special case. The impulse responses  $h_k(n)$  of the filters in a  $T$ -channel cosine modulated filter bank may be written

$$h_k(n) = C p_0(n) \cos \left[ \frac{\pi}{2T} (2k+1)(n - \frac{N-1}{2}) + \Phi_k \right], \quad (19)$$

where  $k = 0, 1, \dots, T-1$ ,  $N$  is the length of the lowpass prototype filter  $p_0(n)$ ,  $C$  is a constant and  $\Phi_k$  is a phase-angle that ensures alias cancellation between adjacent channels. The constraints on  $\Phi_k$  is

$$\Phi_0 = \pm \frac{\pi}{4}, \quad \Phi_{T-1} = \pm \frac{\pi}{4} \quad \text{and} \quad \Phi_k = \Phi_{k-1} \pm \frac{\pi}{2} \quad (20a-c)$$

which may be simplified to the closed form expression

$$\Phi_k = \pm (-1)^k \frac{\pi}{4}. \quad (21)$$

With this choice of  $\Phi_k$ , perfect reconstruction systems or approximate reconstruction systems (pseudo QMF systems) may be obtained using synthesis filter banks with impulse responses as

$$f_k(n) = C p_0(n) \cos \left[ \frac{\pi}{2T} (2k+1)(n - \frac{N-1}{2}) - \Phi_k \right]. \quad (22)$$

Consider the filters

$$h'_k(n) = C p_0(n) \sin \left[ \frac{\pi}{2T} (2k+1)(n - \frac{N-1}{2}) + \Phi_k \right], \quad (23)$$

where  $h'_k(n)$  are sine-modulated versions of the prototype filter  $p_0(n)$ . The filters  $H'_k(z)$  and  $H_k(z)$  have identical passband supports, but the phase responses differ. The passbands of the filters are actually Hilbert transforms of each other (this is not valid for frequencies close to  $\omega = 0$  and  $\omega = \pi$ ). Combining Eq. 19 and Eq. 23 according to

$$h_k^a(n) = h_k(n) + j h'_k(n) = C p_0(n) \exp \left[ \frac{j\pi}{2T} (2k+1)(n - \frac{N-1}{2}) + j\Phi_k \right] \quad (24)$$

yields filters that have the same shape of the magnitude responses as  $H_k(z)$  for positive frequencies but are zero for negative frequencies. Using a filter bank with impulse responses as in Eq. 24 gives a set of subband signals that may be interpreted as the analytic (complex) signals corresponding to the subband signals obtained from a filter bank with impulse responses as in Eq. 19. Analytic signals are suitable for manipulation, since the complex-valued samples may be written in a polar form, that is  $z(n) = r(n) + j i(n) = |z(n)| \exp\{j \arg(z(n))\}$ . However, when using the complex filter bank for transposition, the constraint on  $\Phi_k$  has to be generalised to retain the alias cancellation property. The new constraint on  $\Phi_k$ , to ensure alias cancellation in combination with a synthesis filter bank with impulse responses as in Eq. 22 is

$$\Phi_k = \pm (-1)^k \frac{\pi}{4M} \quad (25)$$

which simplifies to Eq. 21 when  $M = 1$ . With this choice, transposed partials will have the same relative phases as they would have when  $M = 1$  (no transposition).

Combining Eq. 24 and Eq. 25 results in

$$h_k^2(n) = C p_0(n) \exp \left\{ j\pi \left[ \frac{(2k+1)}{2T} \left( n - \frac{N-1}{2} \right) \pm \frac{(-1)^k}{4M} \right] \right\} \quad (26)$$

which are the filters used in the modified filter bank of step 4, according to the present invention.

- 5 Some clarifications concerning step 5: downsampling the complex-valued subband signals by a factor  $T/M$  makes them oversampled by  $M$ , which is an essential criterion when the phase-angles subsequently are multiplied by the transposition factor  $M$ . The oversampling forces the number of subband samples per bandwidth, after transposition to the target range, to equal that of the source range. The individual bandwidths of the transposed subband signals are  $M$  times greater than those in the source range, due to the phase-multiplier. This makes the subband signals
- 10 critically sampled after step 5, and additionally, there will be no zeros in the spectrum when transposing tonal signals.

In order to avoid trigonometric calculations, that is, having to compute the new subband signals as

$$\begin{aligned} s_k^{(M)}(n'') &= \text{real} \left\{ \left| v_k^{(M)}(n'') \right| \exp \left\{ jM \arctan \left( \frac{\text{imag}\{v_k^{(M)}(n'')\}}{\text{real}\{v_k^{(M)}(n'')\}} \right) \right\} \right\} = \\ &= \left| v_k^{(M)}(n'') \right| \cos \left\{ M \arctan \left( \frac{\text{imag}\{v_k^{(M)}(n'')\}}{\text{real}\{v_k^{(M)}(n'')\}} \right) \right\}, \end{aligned} \quad (27)$$

where  $|v_k^{(M)}(n'')|$  is the absolute value of  $v_k^{(M)}(n'')$ , the following trigonometric relationship is used:

$$\cos(M\alpha) = \cos^M(\alpha) - \binom{M}{2} \sin^2(\alpha) \cos^{M-2}(\alpha) + \binom{M}{4} \sin^4(\alpha) \cos^{M-4}(\alpha) - \dots \quad (28)$$

Letting

$$\alpha = \arctan \left( \frac{\text{imag}\{v_k^{(M)}(n'')\}}{\text{real}\{v_k^{(M)}(n'')\}} \right), \quad (29)$$

- 20 and noting that

$$\cos(\alpha) = \cos \left( \arctan \left( \frac{\text{imag}\{v_k^{(M)}(n'')\}}{\text{real}\{v_k^{(M)}(n'')\}} \right) \right) = \frac{\text{real}\{v_k^{(M)}(n'')\}}{|v_k^{(M)}(n'')|}, \quad (30)$$

and

$$\sin(\alpha) = \sin \left( \arctan \left( \frac{\text{imag}\{v_k^{(M)}(n'')\}}{\text{real}\{v_k^{(M)}(n'')\}} \right) \right) = \frac{\text{imag}\{v_k^{(M)}(n'')\}}{|v_k^{(M)}(n'')|}, \quad (31)$$

- the computations of step 5 may be accomplished without trigonometric calculations, reducing computational
- 25 complexity.

When using transpositions where  $M$  is even, obstacles with the phase-multiplier may arise, depending on the characteristics of the lowpass prototype filter  $p_0(n)$ . All applicable prototype filters have zeros on the unit circle in the  $z$ -plane. A zero on the unit circle imposes a  $180^\circ$  shift in the phase response of the filter. For  $M$  even, the phase-multiplier translates these shifts to  $360^\circ$  shifts; i.e. the phase-shifts vanish. The partials so located in frequency that such phase-shifts vanish will give rise to aliasing in the synthesised signal. The worst case scenario is when a partial is located at a point in frequency corresponding to the top of the first side lobe of an analysis filter. Depending on the rejection of this lobe in the magnitude response, the aliasing will be more or less audible. As an example, the first side lobe of the prototype filter used in the ISO/MPEG layer 1 and 2 standard is rejected 96 dB, while the rejection is only 23 dB for the first side lobe of the sine-window used in the MDCT scheme of the ISO/MPEG layer 3 standard. It is clear, that this type of aliasing, using the sine-window, will be audible. A solution to this problem will be presented, and is referred to as relative phase locking.

The filters  $h_k^n(n)$  all have linear phase responses. The phase-angles  $\Phi_k$  introduce relative phase differences between adjacent channels, and the zeros on the unit circle introduce  $180^\circ$  phase-shifts at locations in frequency that may differ between channels. By monitoring the phase-difference between neighbouring subband signals, before the phase-multiplier is activated, it is easy to detect the channels that contain phase-inverted information. Considering tonal signals, the phase-difference is approximately  $\pi/2M$ , according to Eq. 25, for non-inverted signals, and consequently approximately  $\pi(1-1/2M)$  for signals, where either of the signals is inverted. The detection of inverted signals may be accomplished simply by computing the dot product of samples in adjacent subbands as

$$v_k^{(M)}(n) \circ v_{k+1}^{(M)}(n) = \text{real}\{v_k^{(M)}(n)\}\text{real}\{v_{k+1}^{(M)}(n)\} + \text{imag}\{v_k^{(M)}(n)\}\text{imag}\{v_{k+1}^{(M)}(n)\}. \quad (32)$$

If the product in Eq. 32 is negative, the phase-difference is greater than  $90^\circ$ , and a phase-inversion condition is present. The phase-angles of the complex-valued subband signals are multiplied by  $M$ , according to the scheme of step 5, and finally, the inversion-tagged signals are negated. The relative phase locking method thus forces the  $180^\circ$  shifted subband signals to retain this shift after the phase-multiplication, and hence maintain the aliasing cancellation properties.

#### Spectral envelope adjustment

Most sounds, like speech and music, are characterised as products of slowly varying envelopes and rapidly varying carriers with constant amplitude, as described by Stockham ["The Application of Generalized Linearity to Automatic Gain Control" T.G. Stockham, Jr, IEEE Trans. on Audio and Electroacoustics, Vol. AU-16, No. 2, June 1968] and Eq. 1.

In split-band perceptual audio coders, the audio signal is segmented into frames and split into multiple frequency bands using subband filters or a time-to-frequency domain transform. In most codec types, the signal is subsequently separated into two major signal components for transmission or storage, the spectral envelope representation and the normalised subband samples or coefficients. Throughout the following description, the term "subband samples" or "coefficients" refers to sample values obtained from subband filters as well as coefficients obtained from a time-to-frequency transform. The term "spectral envelope" or "scale factors" represent values of the subbands on a time-frame basis, such as the average or maximum magnitude in each subband, used for normalisation of the subband samples. However, the spectral envelope may also be obtained using linear prediction

LPC, [U.S. Pat. 5,684,920]. In a typical codec, the normalised subband samples require coding at a high bitrate (using approximately 90% of the available bitrate), compared to the slowly varying temporal envelopes, and thus the spectral envelopes, that may be coded at a much-reduced rate (using approximately 10% of the available bitrate).

5 Accurate spectral envelope of the replicated bandwidth is important if the timbral qualities of the original signal are to be preserved. The perceived timbre of a musical instrument, or voice, is mainly determined by the spectral distribution below a frequency  $f_{lim}$ , located in the highest octaves of hearing. The spectral details above  $f_{lim}$  are thus of less importance, and consequently the highband fine structures obtained by the above transposition methods require no adjustment, while the coarse structures generally do. In order to enable such adjustment, it is useful to  
10 filter the spectral representation of the signal to separate the envelope coarse structure from the fine structure.

In the SBR-1 implementation according to the present invention, the highband coarse spectral envelope is estimated from the lowband information available at the decoder. This estimation is performed by continuously monitoring the envelope of the lowband and adjusting the highband spectral envelope according to specific rules. A novel method  
15 to accomplish the envelope estimation uses asymptotes in a logarithmic frequency-magnitude space, which is equivalent to curve fitting with polynomials of varying order in the linear space. The level and slope of an upper portion of the lowband spectrum are estimated, and the estimates are used to define the level and slope of one or several segments representing the new highband envelope. The asymptote intersections are fixed in frequency and act as pivot points. However not always necessary, it is beneficial to stipulate constraints to keep the highband  
20 envelope excursions within realistic boundaries. An alternative approach to estimation of the spectral envelope is to use vector quantization, VQ, of a large number of representative spectral envelopes, and store these in a lookup-table or codebook. Vector quantization is performed by training the desired number of vectors on a vast amount of training data, in this case audio spectral envelopes. The training is usually done with the Generalised Lloyd Algorithm ["Vector Quantization and Signal Compression" A. Gersho, R. M. Gray, Kluwer Academic Publishers, USA 1992, ISBN 0-7923-9181-0], and yields vectors that optimally cover the contents of the training data.  
25 Considering a VQ codebook consisting of A spectral envelopes trained by B envelopes ( $B \gg A$ ), then the A envelopes represent the A most likely transitions from the lowband envelope to the highband envelope, based on B observations of a wide variety of sounds. This is, theoretically, the A optimum rules for predicting the envelope based on the B observations. When estimating a new highband spectral envelope, the original lowband envelope is  
30 used to search the codebook and the highband part of the best matching codebook entry is applied to create the new highband spectrum.

In Fig. 23, the normalised subband samples are represented by 2301 and the spectral envelopes are represented by the scalefactors 2305. For illustrative purposes the transmission to decoder 2303 is shown in parallel form. In the  
35 SBR-2 method Fig. 24, the spectral envelope information is generated and transmitted according to Fig. 23, whereby only the lowband subband samples are transmitted. Transmitted scalefactors thus span the full frequency range while the subband samples only span a restricted frequency range, excluding the highband. At the decoder the lowband subband samples 2401 are transposed 2403 and combined with the received highband spectral envelope information 2405. In this way the synthetic highband spectral envelope is identical to that of the original, while  
40 maintaining a significant bit rate reduction.

In some codecs, it is possible to transmit the scalefactors for the entire spectral envelope while omitting the highband subband samples, as shown in Fig. 24. Other codec standards stipulate that scalefactors and subband samples must cover the same frequency range, i.e. scale-factors cannot be transmitted if the subband samples are omitted. In such cases, there are several solutions; the highband spectral envelope information can be transmitted in separate frames, where the frames have their own headers and optional error protection, followed by the data. Regular decoders, not taking advantage of the present invention, will not recognise the headers and therefore discard the extra frames. In a second solution, the highband spectral envelope information is transmitted as auxiliary data within the encoded bitstream. However, the available auxiliary data field must be large enough to hold the envelope information. In cases where none of the first two solutions are adaptable, a third solution, where the highband spectral envelope information is hidden as subband samples, may be applied. Subband scalefactors cover a large dynamic range, typically exceeding 100 dB. It is thus possible to set an arbitrary number of subband scalefactors, 2505 in Fig. 25, to very low values, and to transmit the highband scalefactors "camouflaged" as subband samples, 2501. This way of transmitting the highband scale factors to the decoder 2503 ensures compatibility with the bitstream syntax. Hence, arbitrary data may be transmitted in this fashion. A related method exists where information is coded into the subband sample stream [U. S. Pat. 5,687,191]. A fourth solution, Fig. 26, can be applied when a coding system uses Huffman- or other redundancy coding 2603. The subband samples for the highband is then set to zero 2601 or a constant value as to achieve a high redundancy.

#### Transient response improvements

Transient related artifacts are common problems in audio codecs, and similar artifacts occur in the present invention. In general, patching generates spectral "zeros" or notches, corresponding to time domain pre- and post-echoes, i.e. spurious transients before and after "true" transients. Albeit the P-blocks "fill in the zeros" for slowly varying tonal signals, the pre- and post-echoes remain. The improved multiband method is intended to work on discrete sinusoids, where the number of sinusoids is restricted to one per subband. Transients or noise in a subband can be viewed as a large number of discrete sinusoids within that subband. This generates intermodulation distortion. These artifacts are considered as additional quantization-noise sources connected to the replicated highband channels during transient intervals. Traditional methods to avoid pre- and post-echo artifacts in perceptual audio coders, for example adaptive window switching, may hence be used to enhance the subjective quality of the improved multiband method. By using the transient detection provided by the codec or a separate detector and reducing the number of channels under transient conditions the "quantization noise" is forced not to exceed the time-dependent masking threshold. A smaller number of channels is used during transient passages whereas a larger is used during tonal passages. Such adaptive window switching is commonly used in codecs in order to trade frequency resolution for time resolution. Different methods may be used in applications where the filterbank size is fixed. One approach is to shape the "quantization noise" in time via linear prediction in the spectral domain. The transposition is then performed on the residual signal, which is the output of the linear prediction filter. Subsequently, an inverse prediction filter is applied to the original- and spectral replicated channels simultaneously. Another approach employs a compander system i.e. dynamic amplitude compression of the transient signal prior to transposition or coding, and a complementary expansion after transposition. It is also possible to switch between transposition methods in a signal dependent manner, for example, a high resolution filterbank transposition method is used for stationary signals, and a time-variant pattern search prediction method is employed for transient signals.

### Practical implementations

Using a standard signal-processor or a powerful PC, real-time operation of a SBR-enhanced codec is possible. The SBR enhanced codec may also be hard-coded on a custom chip. It may also be implemented in various kinds of systems for storage or transmission of signals, analogue or digital, using arbitrary codecs, Fig. 27 and Fig. 28. The SBR-1 method may be integrated in a decoder or supplied as an add-on hardware or software post-processing module. The SBR-2 method needs additional modification of the encoder. In Fig. 27 the analogue input signal is fed to the A/D-converter 2701, forming a digital signal which is fed to the an arbitrary encoder 2703, where source coding is performed. The signal fed into the system may be of such a low-pass type that spectral bands within the auditory range already have been discarded, or spectral bands are discarded in the arbitrary encoder. The resulting lowband signals are fed to the multiplexer 2705, forming a serial bitstream which is transmitted or stored 2707. The de-multiplexer 2709 restores the signals and feeds them to an arbitrary decoder 2711. The spectral envelope information 2715 is estimated at the decoder 2713 and fed to the SBR-1 unit 2713 which transposes the lowband signal to a highband signal and creates an envelope adjusted wideband signal. Finally, the digital wideband signal is converted 2717 to an analogue output signal.

The SBR-2 method needs additional modification of the encoder. In Fig. 28 the analogue input signal is fed to the A/D-converter 2801, forming a digital signal which is fed to the an arbitrary encoder 2803, where source coding is performed. The spectral envelope information is extracted 2805. The resulting signals, lowband subband samples or coefficients and wideband envelope information, are fed to the multiplexer 2807, forming a serial bitstream which is transmitted or stored 2809. The de-multiplexer 2811 restores the signals, lowband subband samples or coefficients and wideband envelope information, and feeds them to an arbitrary decoder 2815. The spectral envelope information 2813 is fed from the de-multiplexer 2811 to the SBR-2 unit 2817 which transposes the lowband signal to a highband signal and creates an envelope adjusted wideband signal. Finally, the digital wideband signal is converted 2819 to an analogue output signal.

When only very low bitrates are available, (Internet and slow telephone modems, AM-broadcasting etc.) mono coding of the audio program material is unavoidable. In order to improve the perceived quality and make the programme more pleasant sounding, a simple "pseudo-stereo" generator, Fig. 29, is obtained by the introduction of a tapped delayline 2901. This may feed 10ms and 15ms delayed signals at approximately -6dB 2903 to each output channel in addition to the original mono signal 2905. The pseudo-stereo generator offers a valuable perceptual improvement at a low computational cost.

The above-described embodiments are merely illustrative for the principles of the present invention for audio source coding improvement. 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. A method for enhancement of a source coding system where said source coding system comprises an encoder representing all operations performed prior to storage or transmission, and a decoder representing all operations performed after storage or transmission, **characterised by:**
- 5       at said encoder, discarding frequency band(s) of an original signal, forming a first signal;  
      at said decoder, by means of transposition performed on said first signal, replicating frequency band(s) of said original signal, forming a second signal; and  
      combining said first and said second signal, forming an output signal; whereby reduced bitrate at a given  
10       perceptual quality or improved perceptual quality at a given bitrate is obtained.
2. A method according to claim 1, **characterised in** that passband(s) of said second signal are set not to overlap or only partly overlap passband(s) of said first signal.
- 15   3. A method according to claims 1 - 2, **characterised in** that spectral envelope adjustment is performed based on estimation of the spectral envelope of said discarded frequency band(s) of said original signal using said first signal.
4. A method according to claims 1 - 2, **characterised in** that spectral envelope adjustment is performed based on transmitted envelope information of said discarded frequency band(s) of said original signal.
- 20   5. A method according to claim 4, **characterised in** that said spectral envelope information is transmitted as subband samples in an arbitrary number of subband channels where the gains of said subband channels are set to a low level; whereby compatibility with standardised decoders is ensured.
- 5   6. A method according to claim 4, **characterised in** that said envelope information is transmitted as scale factors without transmission of the corresponding subband samples.
7. A method according to claim 4, **characterised in** that said envelope information is transmitted as scale factors and the corresponding subband samples are set to zero or a constant value; whereby the entropy of the subband  
30       samples is reduced.
8. A method according to claims 1 - 7, **characterised in** that said output signal, when monophonic audio, is split into two signals each comprised of said output signal and delayed versions of the same to obtain a pseudo-stereo signal.
- 35   9. A method according to claims 1 - 7, where said transposition is **characterised by:**  
      filtering a signal through a set of  $N \geq 2$  bandpass filters with passbands comprising the frequencies  $[f_1, \dots, f_N]$  respectively, forming  $N$  bandpass signals;  
      shifting said bandpass signals in frequency to regions comprising the frequencies  $M [f_1, \dots, f_N]$  where  $M \neq 1$  is  
40       the transposition factor; and  
      forming a transposed signal by combining said shifted bandpass signals.

10. A method according to claim 9, **characterised in** that said frequency shifting is obtained through upper side band (USB) modulation.

11. A method for transposition by a factor  $M$ , **characterised by**:

5 bandpass filtering a signal using an analysis filter bank or transform of such a nature that real- or complex-valued subband signals of lowpass type are generated;

patching an arbitrary number of channels  $k$  of said analysis filter bank or transform to channels  $Mk$ ,  $M \neq 1$ , in a synthesis filter bank or transform; and

forming a transposed signal using said synthesis filter bank or transform.

10 12. A method according to claim 11, **characterised in** that said filter bank is maximally decimated and said patching is performed according to the relation

$$v_{Mk}(n) = (-1)^{(M-1)kn} v_k(n),$$

15 where  $(-1)^{(M-1)kn}$  is a correction factor,  $v_k(n)$  the subband signal of channel  $k$ , and  $v_{Mk}(n)$  the subband signal of channel  $Mk$ ; whereby compensation of spectral inverted subband signals is obtained.

13. A method according to claims 11 - 12, **characterised by**:

patching the phases of the subband signals from channels  $k$  of said analysis filter bank or transform as the phases of the subband signals associated with synthesis channels  $Mk$ ,  $M \neq 1$ ; and

20 patching the magnitudes of the subband signals from consecutive channels  $l$  of said analysis filter bank or transform as the magnitudes of the subband signals associated with consecutive synthesis channels  $l+S$ , where  $S$  is an integer  $\neq 1$ .

25 14. A method according to claims 11 - 13, **characterised in** that the phases of said subband signals of said channels  $k$  are multiplied by said factor  $M$  before using said synthesis filter bank or transform.

15. A method according to claims 11 - 14, **characterised in** that  $M = K^{\pm 1}$ , where  $K$  is an integer  $> 1$ .

30 16. A method according to claims 11 - 15, **characterised in** that said patching employs multiple values of said transposition factor  $M$ .

17. A method for transposition by a factor  $M$ , characterised by:

filtering a signal through a parallel bank of  $L$  filters with impulse responses as

$$h_k(n) = K p_0(n) \exp \left[ j \frac{\pi}{2L} (2k+1) \left( n - \frac{N-1}{2} \right) + j (-1)^k \frac{\pi}{4M} \right],$$

where  $k = 0, 1, \dots, L-1$ ,  $K$  is a constant, and  $p_0(n)$  is a lowpass prototype filter of length  $N$ , producing a set of  $L$  complex-valued signals;

downsampling said  $L$  signals with a factor  $L/M$ , producing a set of  $L$  complex-valued subband signals;

multiplying the phase-angles of said complex-valued subband signals by  $M$ , giving a new set of subband signals;

selecting the real part of said new set of subband signals, resulting in a set of  $L$  real-valued subband signals;

upsampling a subset of said real-valued subband signals with a factor  $L'$ , producing a set of real-valued signals;

filtering said real-valued signals through a parallel bank of  $L'$  filters with impulse responses as

$$f_k(n) = K' p'_0(n) \cos \left[ \frac{\pi}{2L'} (2k+1) \left( n - \frac{N'-1}{2} \right) - (-1)^k \frac{\pi}{4} \right],$$

where  $k = 0, 1, \dots, L'-1$ ,  $K'$  is a constant and  $p'_0(n)$  is a lowpass prototype filter of length  $N'$ , forming a set of  $L'$  filtered signals; and

adding said  $L'$  filtered signals to produce a transposed signal.

18. A method according to claim 17, characterised in that said multiplication of said phase-angles and said selecting of the real part, is computed by:

writing said complex-valued subband signals as

$$Z_k(n) = R_k(n) + j I_k(n),$$

where  $R_k(n)$  and  $I_k(n)$  are the real and imaginary parts of  $Z_k(n)$  respectively;

calculating said real-valued subband signals  $W_k(n)$  as

$$W_k(n) = |Z_k(n)| \cos \left\{ M \arctan \left( \frac{I_k(n)}{R_k(n)} \right) \right\},$$

where  $|Z_k(n)| = \sqrt{R_k(n)^2 + I_k(n)^2}$  and  $M$  is a positive integer transposition factor, using the trigonometric identity  $\cos(M\alpha) = \cos^M(\alpha) - \binom{M}{2} \sin^2(\alpha) \cos^{M-2}(\alpha) + \binom{M}{4} \sin^4(\alpha) \cos^{M-4}(\alpha) - \dots$ ,

where  $\alpha = \arctan\{I_k(n)/R_k(n)\}$ , and the relations

$$\cos(\alpha) = \frac{R_k(n)}{|Z_k(n)|} \text{ and } \sin(\alpha) = \frac{I_k(n)}{|Z_k(n)|};$$

whereby reducing computational complexity by elimination of all trigonometric calculations.

19. A method according to claim 17, characterised by:

on a block basis, extracting information conveyed by the phase-difference of an adjacent pair of said complex-valued subband signals;

performing said multiplication of said phase-angles by  $M$ , forming a pair of said new subband signals; and

negating one of said new subband signals, on a condition provided by said information; whereby  $180^\circ$  phase-shifts of the subband signals are retained when employing an even integer-valued transposition factor  $M$ .

20. A method according to claim 19, characterised in that said information is given by the dot-product of said complex-valued subband signals  $Z_k(n)$  and  $Z_{k+1}(n)$  according to

$$Z_k(n) \circ Z_{k+1}(n) = R_k(n)R_{k+1}(n) + I_k(n)I_{k+1}(n),$$

where  $R_i(n)$  and  $I_i(n)$  are the real and imaginary parts of  $Z_i(n)$  respectively,  $i = k, k+1$ , and one of said new subband signals is negated provided said dot-product is negative.

21. A method for transposition, by expanding or compressing a first signal in time and duplicating or discarding arbitrary long segments of said first signal, subsequently down- or up-sample said first signal, characterised by:

performing transient detection on said first signal;

determining which segment of said first signal to be used when duplicating or discarding parts of said first signal depending on the outcome of the transient detection;

adjusting the length  $L$  of said signal segment depending on the outcome of the transient detection;

adjusting the number of samples  $N$  used for each statevector depending on the outcome of the transient detection;

adjusting the delay  $D$  between samples in said statevector depending on the outcome of the transient detection;

adjusting the number of samples  $K$  between each statevector depending on the outcome of the transient detection; and

searching for synchronisation points in chosen segment of said first signal, based on the synchronisation point found in the previous synchronisation point search.

22. A method according to claim 21, characterised in that several transposers are interconnected in order to share synchronisation point information; whereby reduced computational complexity is achieved.

23. A method according to claims 21 - 22, characterised in that said transposers are connected to a suitable filter bank, and the signals fed into each of said transposers are filtered as to obtain an arbitrary spectral envelope of the new signal being the sum of said signals being processed by said transposers.

24. A system for enhanced decoding of a source coded signal derived from an original signal, **characterised by:**  
transposition means for transposing frequency band(s) of said source coded signal, forming a first signal;  
estimation means operating on said source coded signal, for estimation of the spectral envelope of said original signal;

5 adjusting means for adjusting the spectral envelope of said first signal, based on said estimation;  
combining means for combining said source coded signal and said adjusted first signal; whereby reduced  
bitrate at a given perceptual quality or improved perceptual quality at a given bitrate is achieved.

25. An apparatus according to claim 24, operating when said output signal is monophonic audio **characterised by:**  
10 delaying means for delaying and attenuating means for attenuating said output signal forming a first delayed  
signal;

delaying means for delaying and attenuating means for attenuating said output signal, using different  
parameters, forming a second delayed signal;

means for adding said output and said first delayed signal, forming a left-channel output signal; and

15 means for adding said output and said second delayed signal, forming a right-channel output signal; whereby  
obtaining a pseudo stereophonic signal.

26. A system for enhanced source coding where said system comprises an encoder representing all units preceding a  
storage media or transmission channel, and a decoder representing all units following said storage media or  
20 transmission channel, **characterised by:**

discarding means at said encoder for discarding frequency band(s) of an original signal, forming a first  
signal;

extracting means at said encoder for extracting spectral envelope information of said original signal, forming  
a second signal;

25 means at said encoder for coding of said first and second signals;

transposition means at said decoder for transposing frequency band(s) of said first signal, forming a third  
signal;

adjusting means at said decoder for spectral envelope adjustment of said third signal, based on said second  
signal; and

30 combining means at said decoder for combining said first and said adjusted third signal; whereby reduced  
bitrate at a given perceptual quality or improved perceptual quality at a given bitrate is achieved.

27. An apparatus according to claim 26, operating when said output signal is monophonic audio **characterised by:**

35 delaying means for delaying and attenuating means for attenuating said output signal forming a first delayed  
signal;

delaying means for delaying and attenuating means for attenuating said output signal, using different  
parameters, forming a second delayed signal;

means for adding said output and said first delayed signal, forming a left-channel output signal; and

40 means for adding said output and said second delayed signal, forming a right-channel output signal; whereby  
obtaining a pseudo stereophonic signal.

28. An apparatus for transposition a factor  $M$ , characterised by:

filtering a signal by means of an analysis filter bank or transform of such a nature that real- or complex-valued subband signals of lowpass type are generated;

means for patching an arbitrary number of channels  $k$  of said analysis filter bank or transform to channels

5  $Mk$ ,  $M \neq 1$ , in a synthesis filter bank or transform; and

forming a transposed signal by means of said synthesis filter bank or transform.

29. An apparatus according to claim 28, characterised in that when  $M = K^{\pm 1}$ , where  $K$  is an integer  $> 1$ , the phases of the subband signals of said channels  $k$  are multiplied by  $M$  before using said synthesis filter bank or transform.

10

30. An apparatus for transposition by a factor  $M$ , characterised by:

filtering means for filtering a signal through a parallel bank of  $L$  filters with impulse responses as

$$h_k(n) = K p_0(n) \exp \left[ j \frac{\pi}{2L} (2k+1) \left( n - \frac{N-1}{2} \right) + j (-1)^k \frac{\pi}{4M} \right],$$

where  $k = 0, 1, \dots, L-1$ ,  $K$  is a constant,  $p_0(n)$  is a lowpass prototype filter of length  $N$ , and  $M$  is the transposition factor, producing a set of  $L$  complex-valued signals;

15

means for downsampling said  $L$  signals with a factor  $L/M$ , producing a set of  $L$  complex-valued subband signals;

means for multiplying the phase-angles of said complex-valued subband signals by  $M$ , giving a new set of subband signals;

20

means for selecting the real part of said new set of subband signals, resulting in a set of  $L$  real-valued subband signals;

means for upsampling a subset of said real-valued subband signals with a factor  $L'$ , producing a set of real-valued signals;

filtering means for filtering said real-valued signals through a parallel bank of  $L'$  filters with impulse

25

responses as

$$f_k(n) = K' p'_0(n) \cos \left[ \frac{\pi}{2L'} (2k+1) \left( n - \frac{N'-1}{2} \right) - (-1)^k \frac{\pi}{4} \right],$$

where  $k = 0, 1, \dots, L'-1$ ,  $K'$  is a constant and  $p'_0(n)$  is a lowpass prototype filter of length  $N'$ , forming a set of  $L'$  filtered signals; and

means for adding said  $L'$  filtered signals to produce the transposed signal.

31. An apparatus for transposition, by expanding or compressing a first signal in time by duplicating or discarding arbitrary long segments of said first signal, and subsequently down- or up-sample said first signal, characterised by:

detection means for performing transient detection on said first signal;

5 means for using the position of a possible transient when determining which segment of said first signal to be used when duplicating or discarding parts of said first signal, in order to obtain said transposition;

adjusting means for adjusting the length ( $L$ ) of said signal segment depending on the output from the transient detector;

10 adjusting means for adjusting the number of samples ( $N$ ) used for each statevector depending on the output from the transient detector;

adjusting means for adjusting the delay ( $D$ ) between samples in said statevector depending on the output from the transient detector;

adjusting means for adjusting the number of samples ( $K$ ) between each statevector depending on the output from the transient detector; and

15 searching means for searching for synchronisation points in chosen segment of said first signal, based on the synchronisation point found in the previous synchronisation point search.

32. An apparatus according to claim 31, operating on subband signals characterised by:

means for sharing synchronisation information between multiple instances of said transposer;

20 means for forming subsets of said subband signals;

means for amplitude adjustment of every channel within each of said subsets;

synthesis filter bank means for forming, from each of said subsets, an input signal to each instance of said transposers;

processing said input signals by said transposers; and

25 summation means for acquiring a new signal by summation of said processed signals; whereby an arbitrary spectral envelope can be obtained.

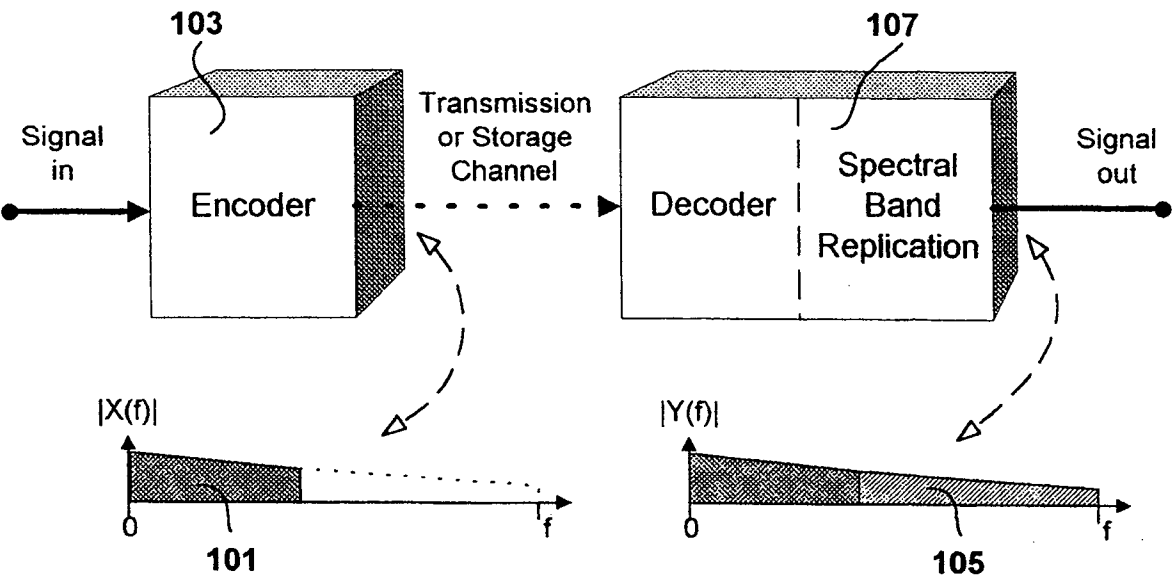


Fig. 1

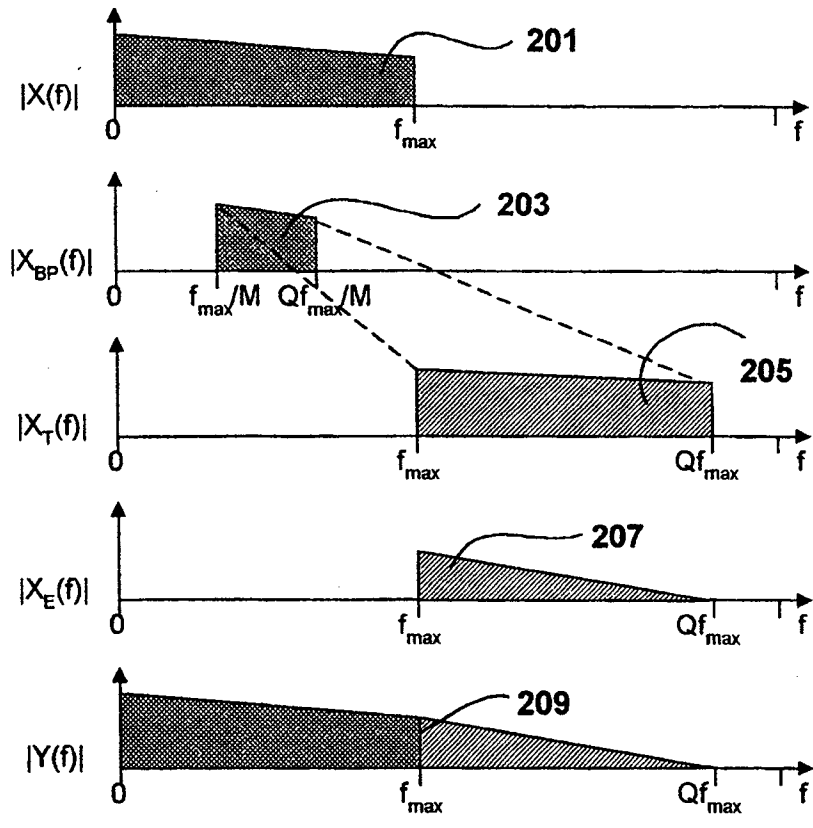


Fig. 2

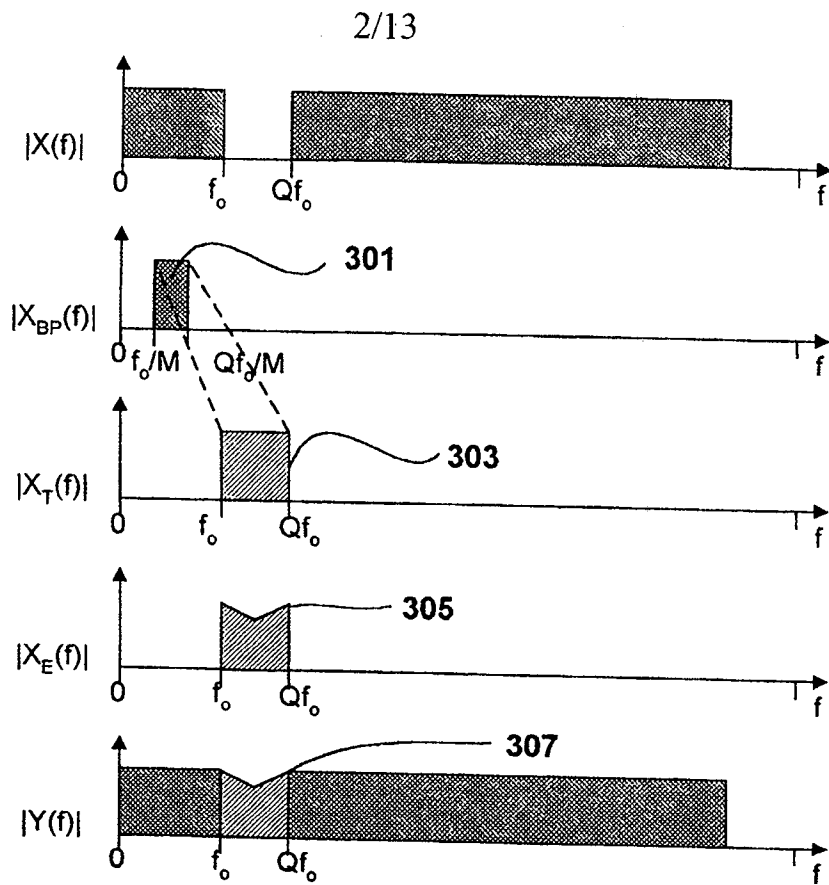


Fig. 3

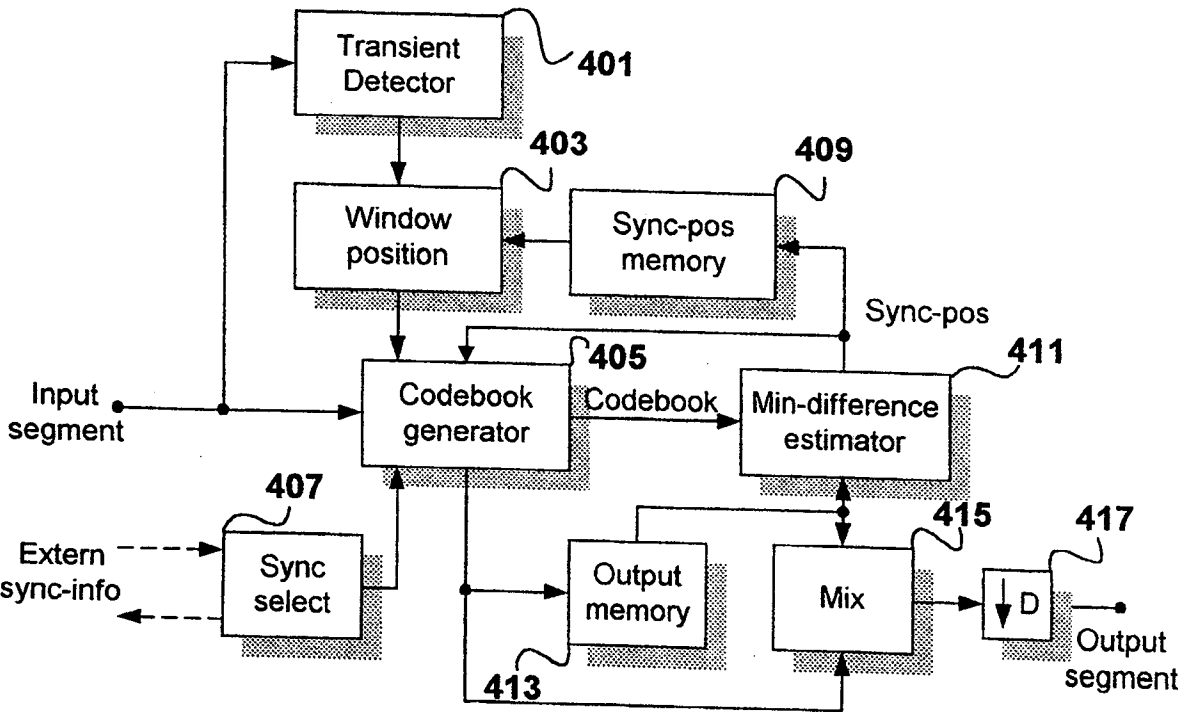
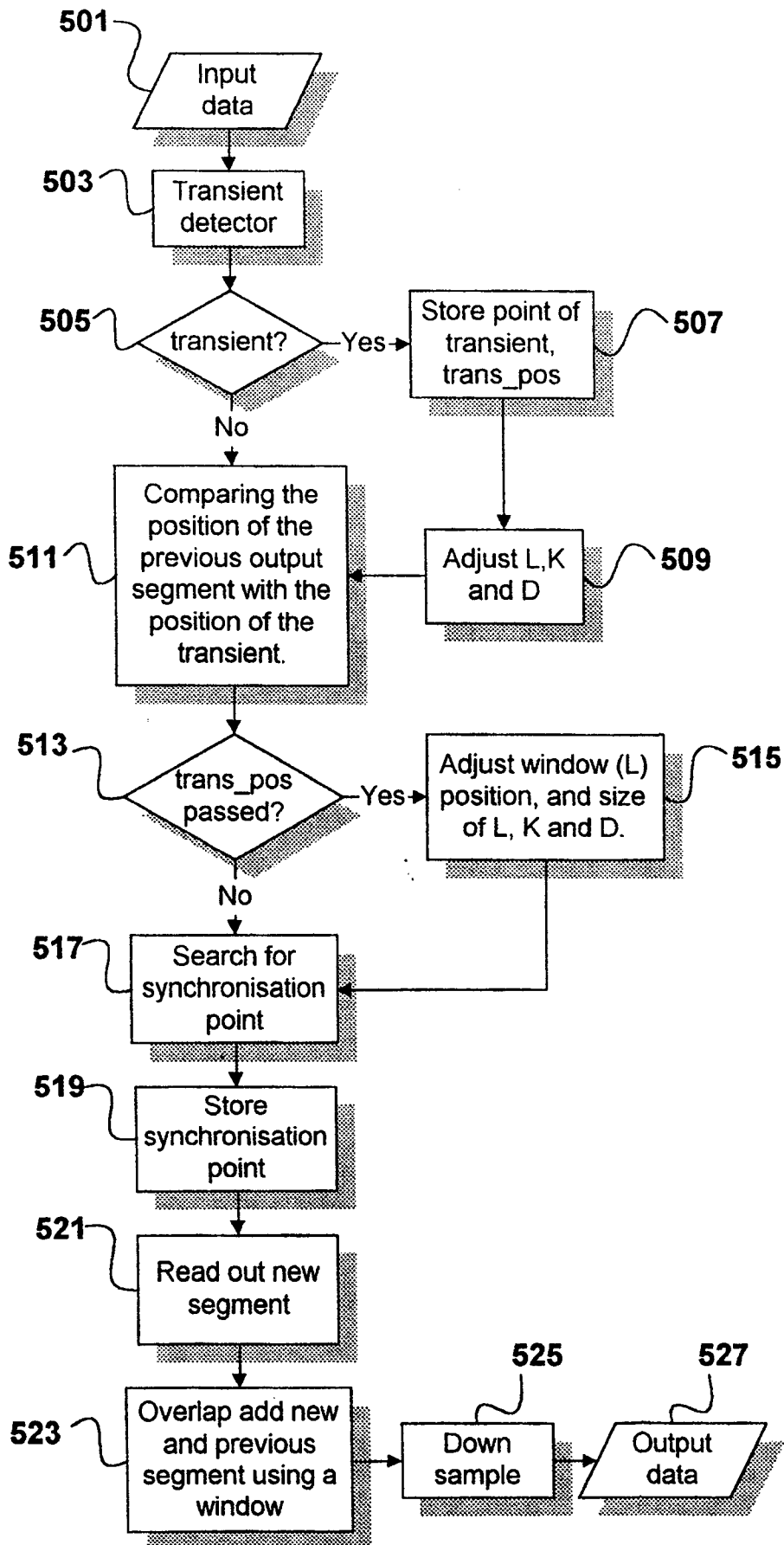


Fig. 4

3/13

*Fig. 5*

4/13

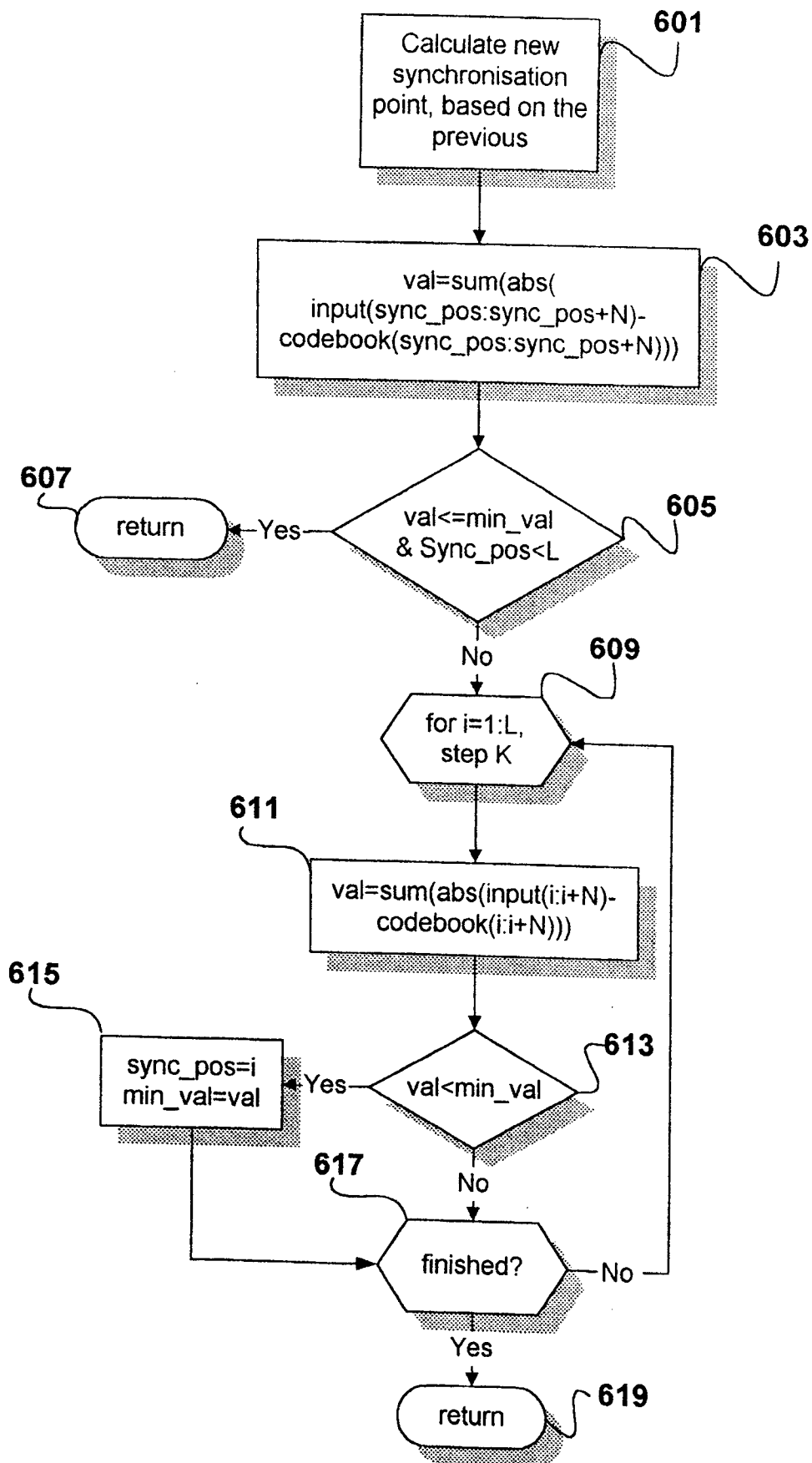


Fig. 6

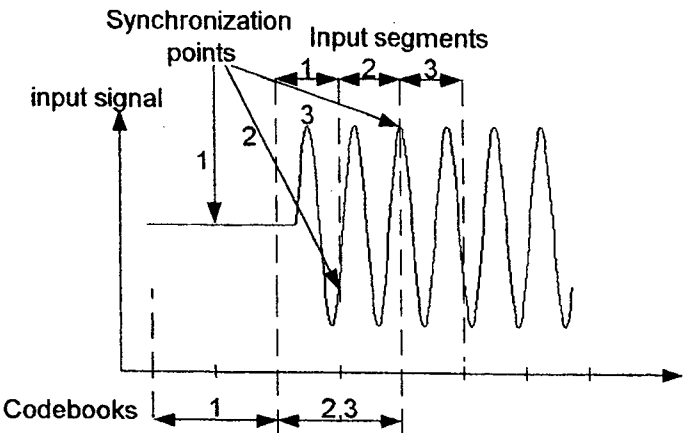


Fig. 7a

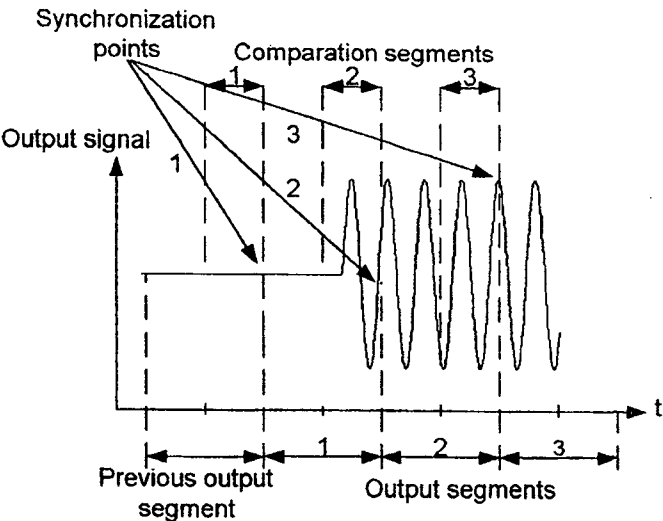


Fig. 7b

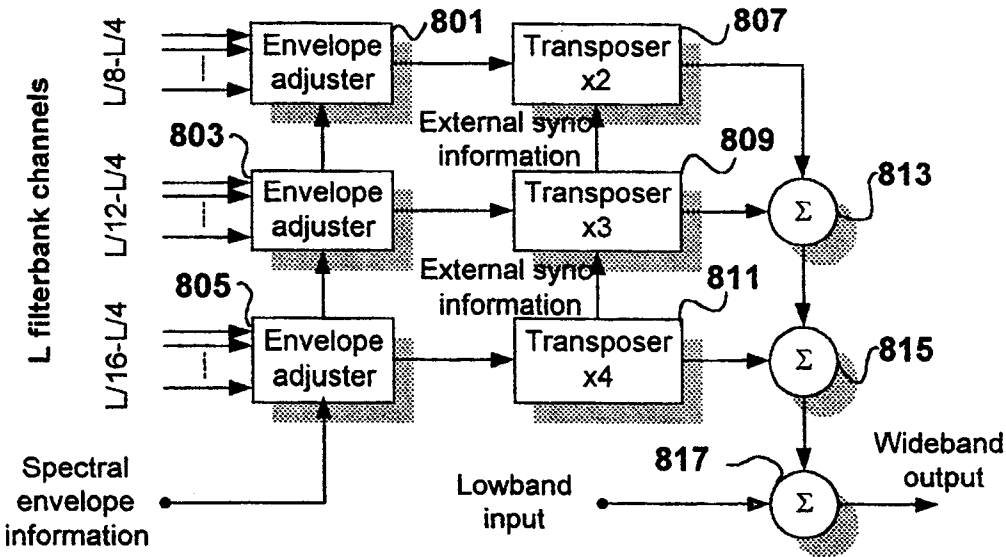


Fig. 8

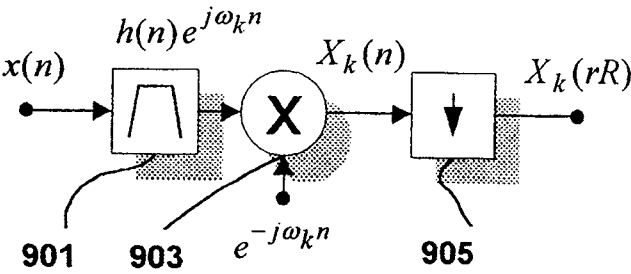


Fig. 9a

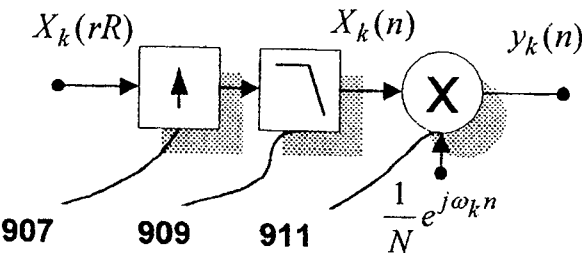


Fig. 9b

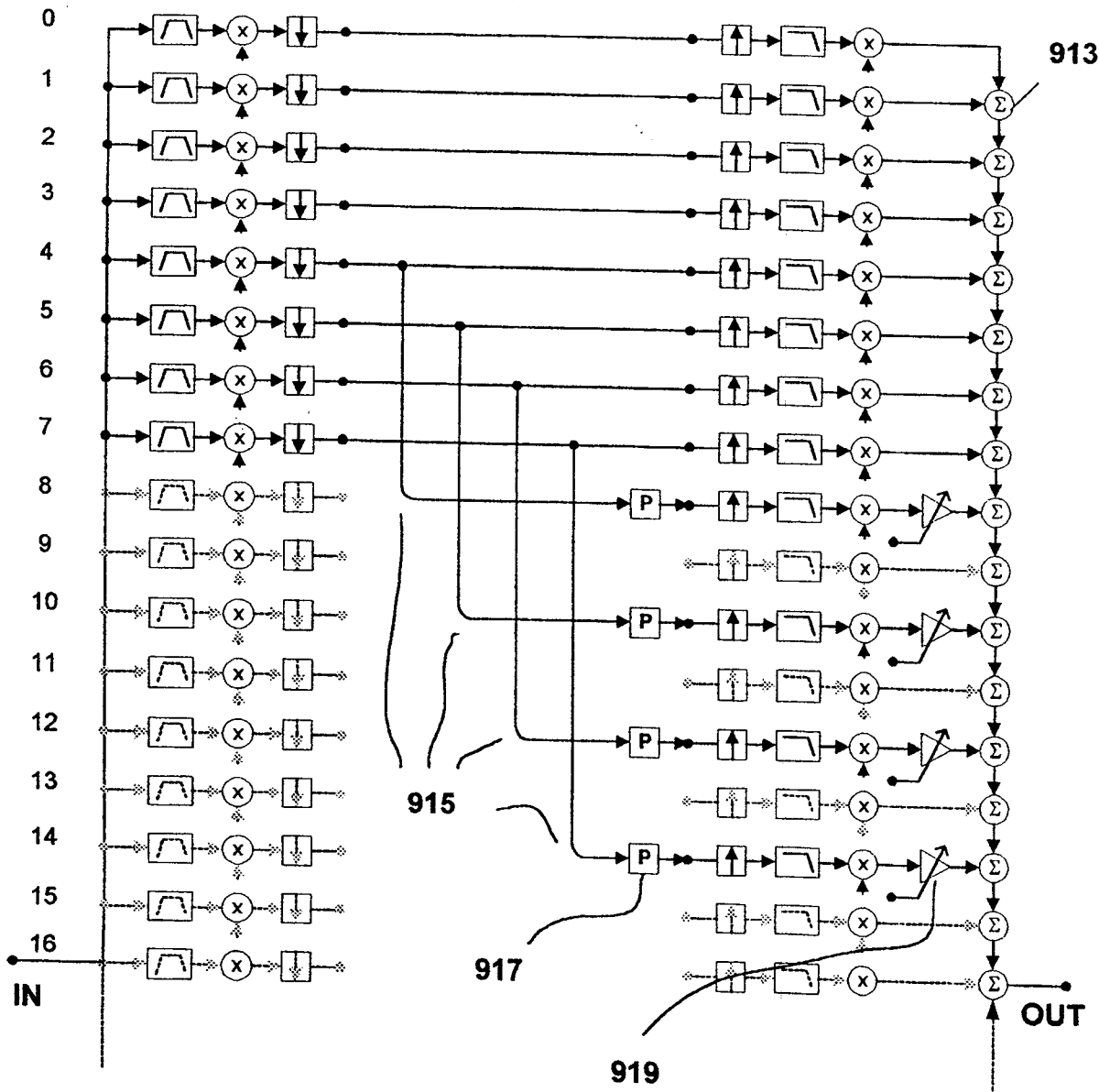


Fig. 9c

7/13

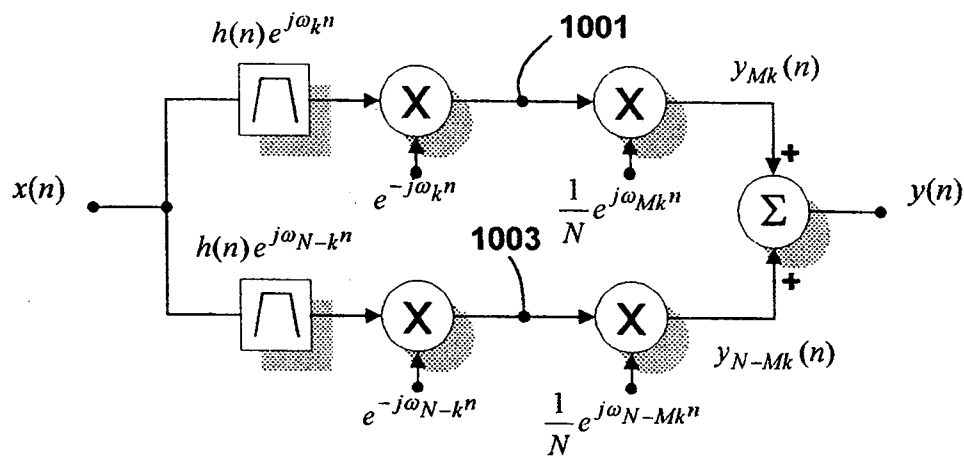


Fig. 10a

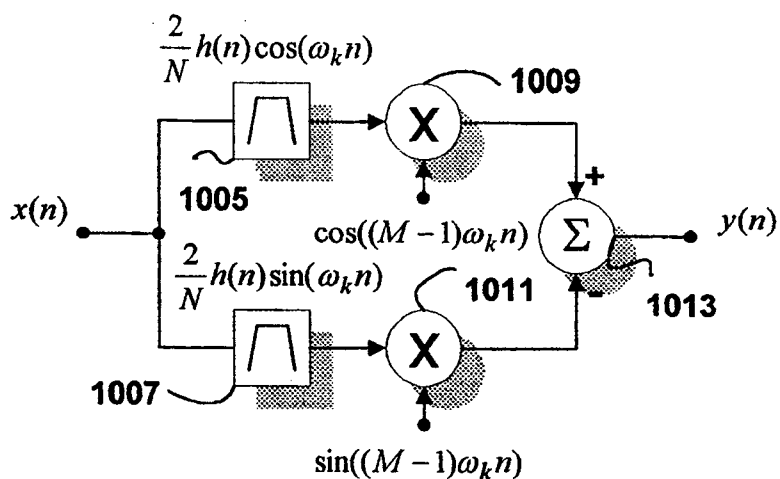


Fig. 10b

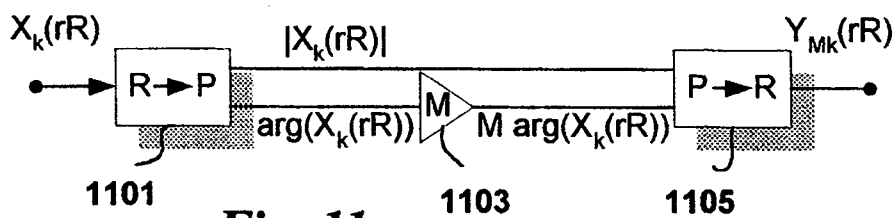


Fig. 11

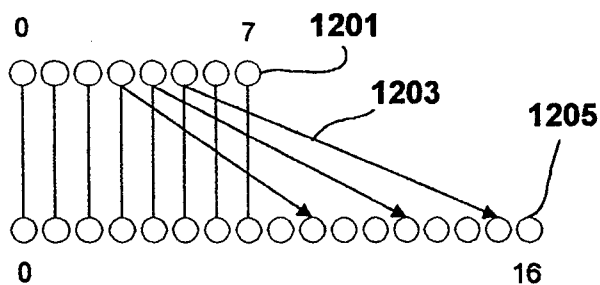


Fig. 12

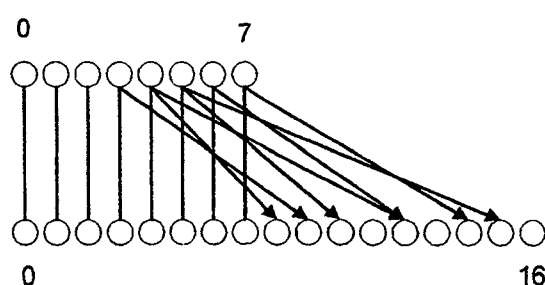


Fig. 13

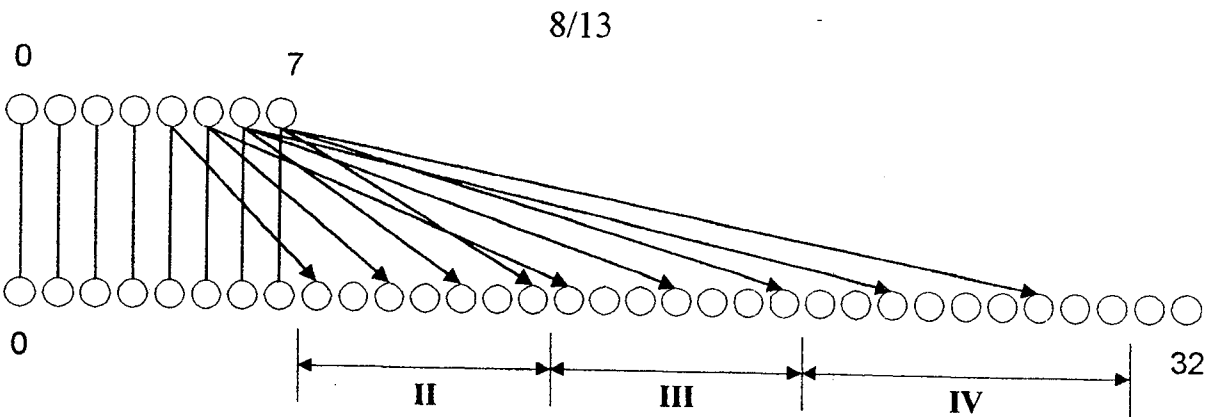


Fig. 14

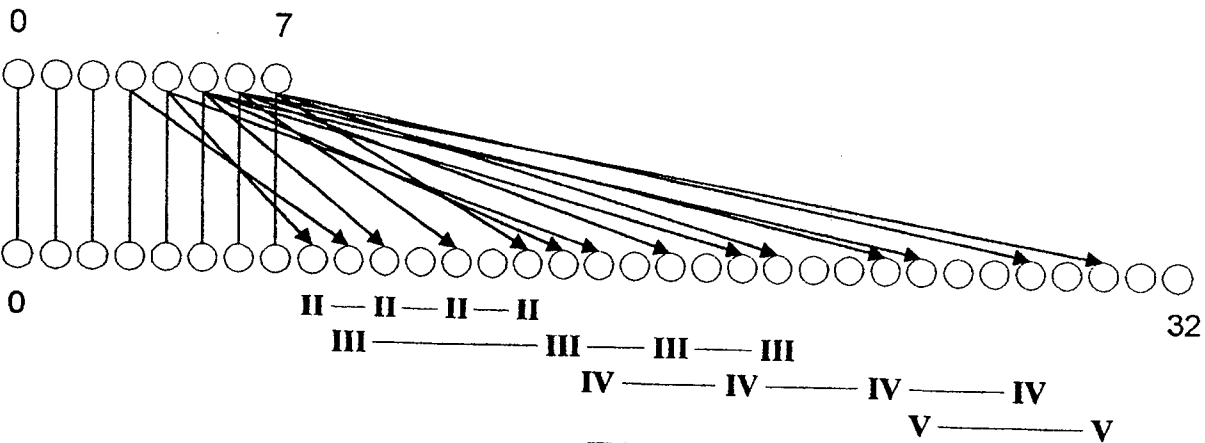


Fig. 15

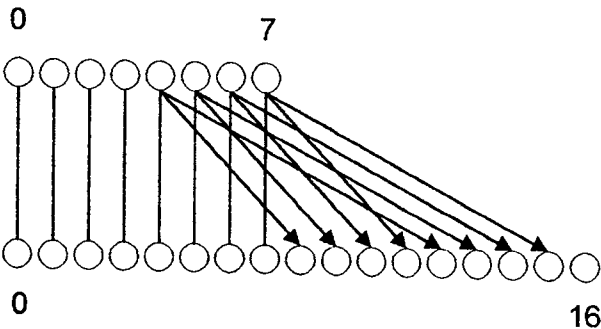


Fig. 16

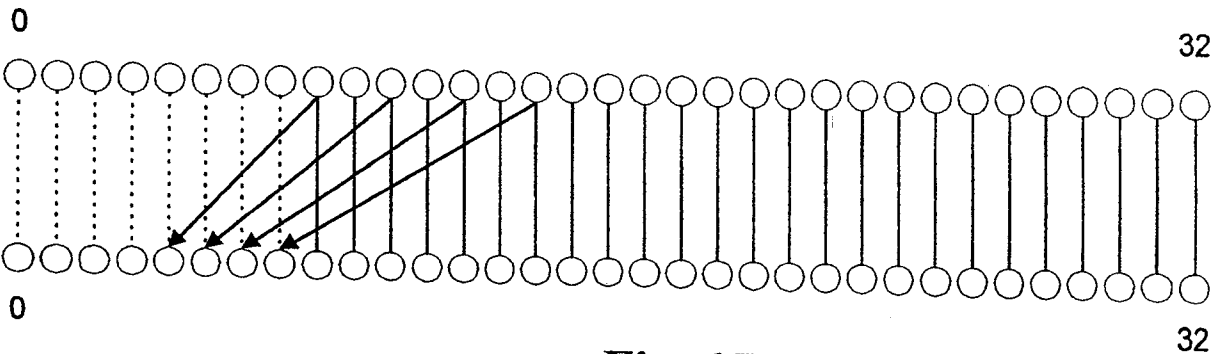


Fig. 17

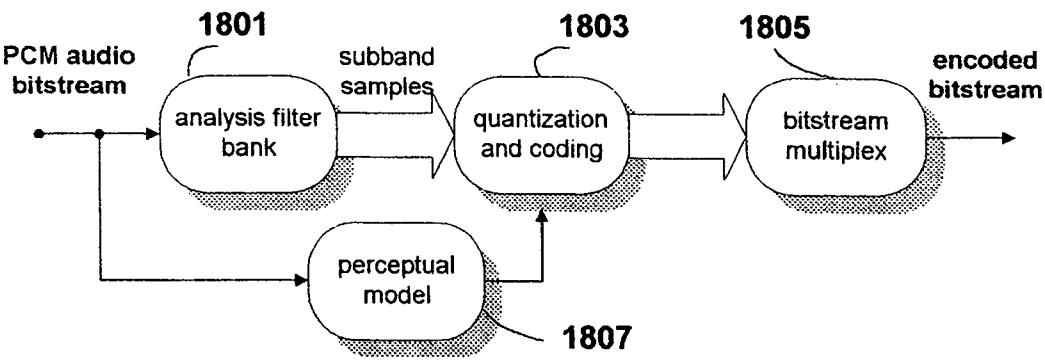


Fig. 18a

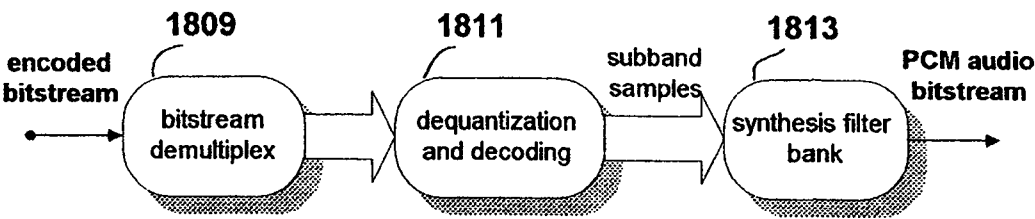


Fig. 18b

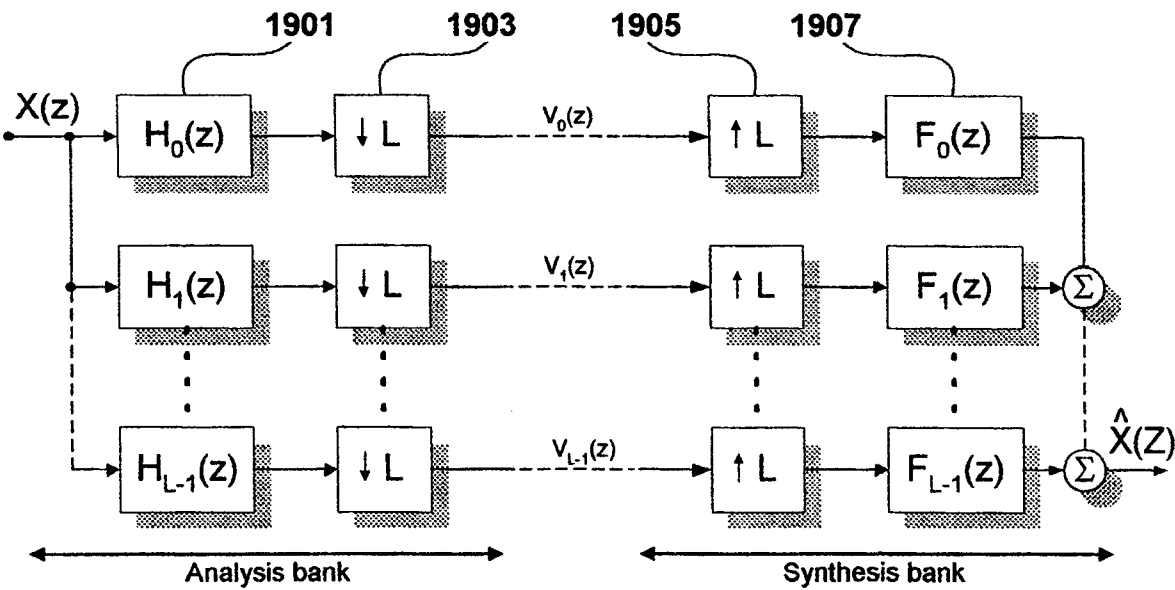


Fig. 19

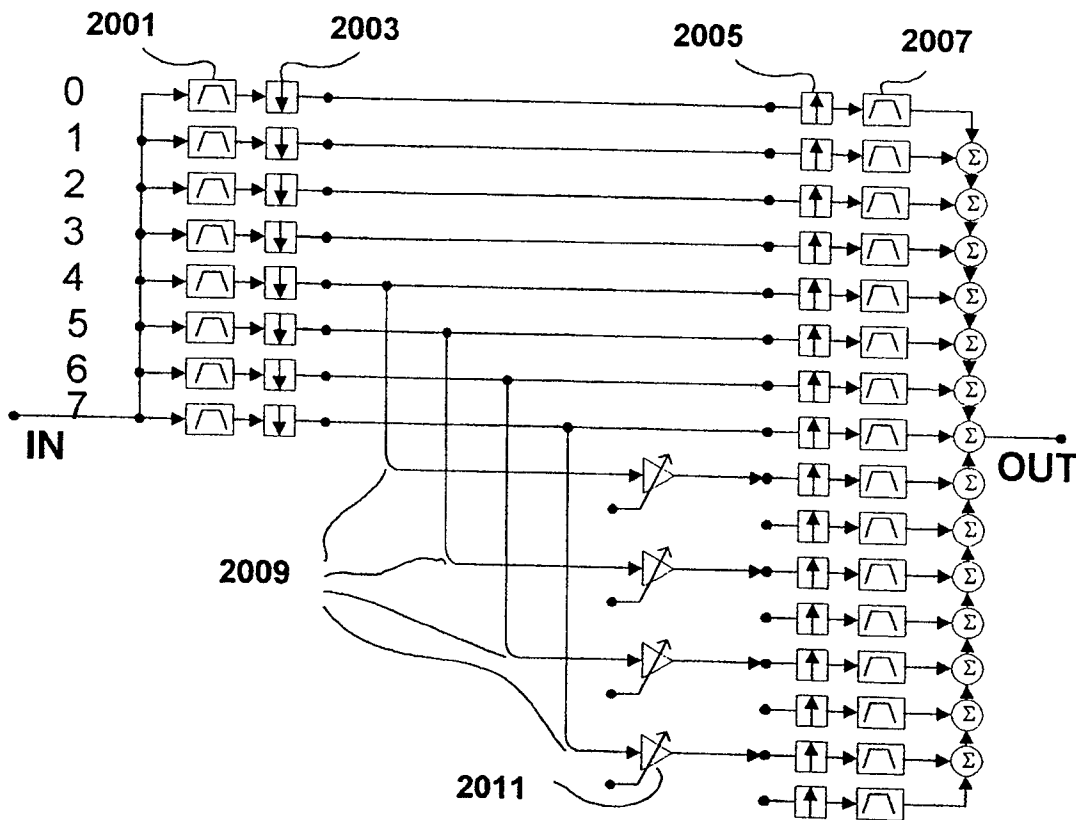


Fig. 20

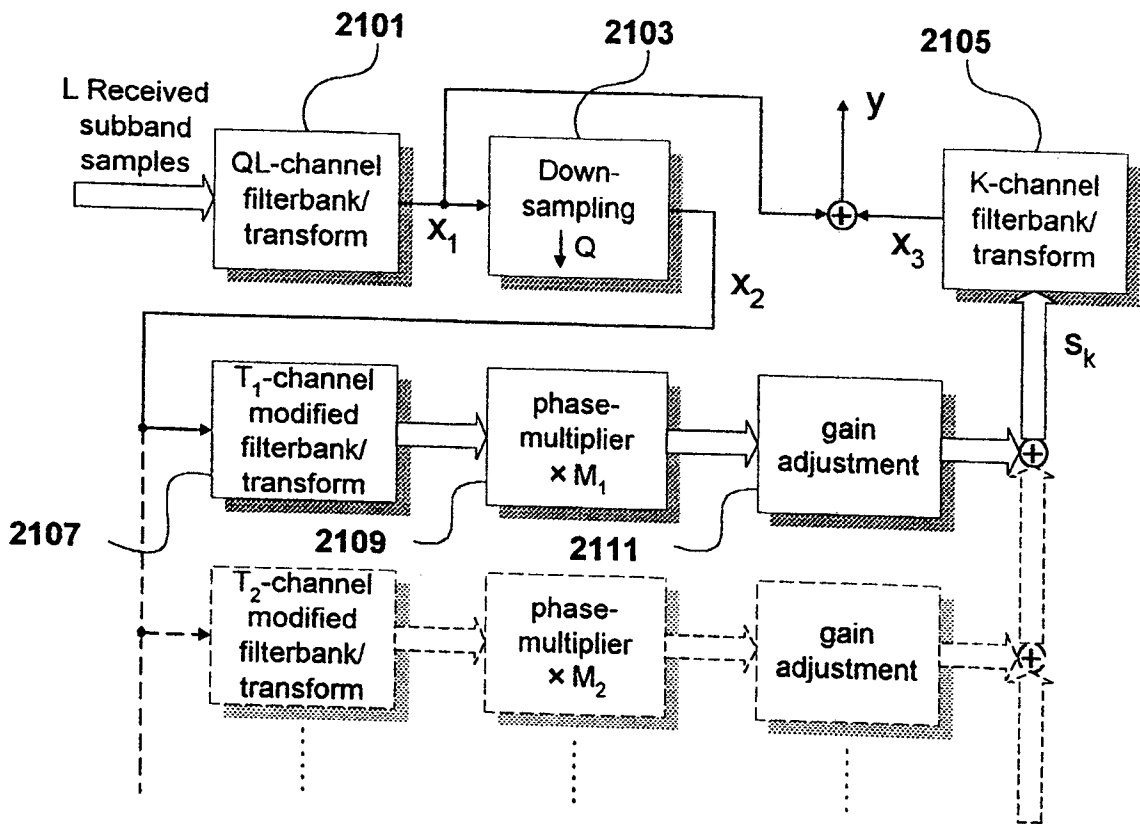


Fig. 21

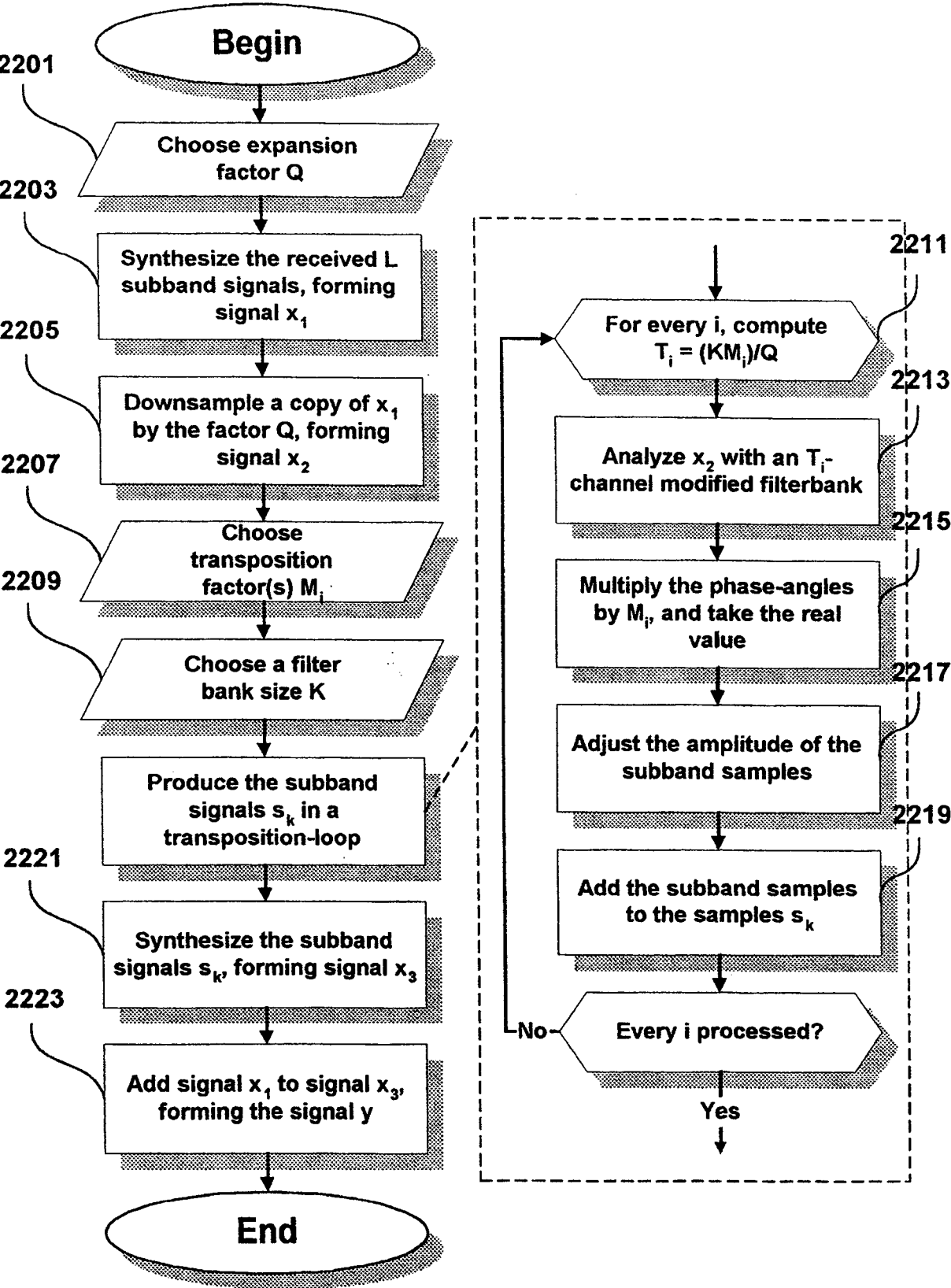
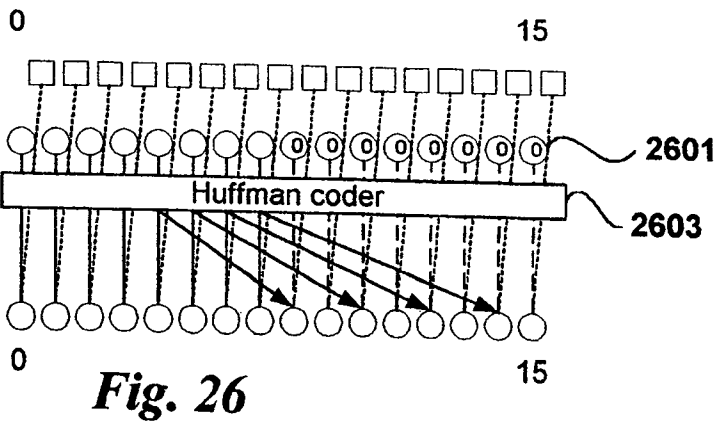
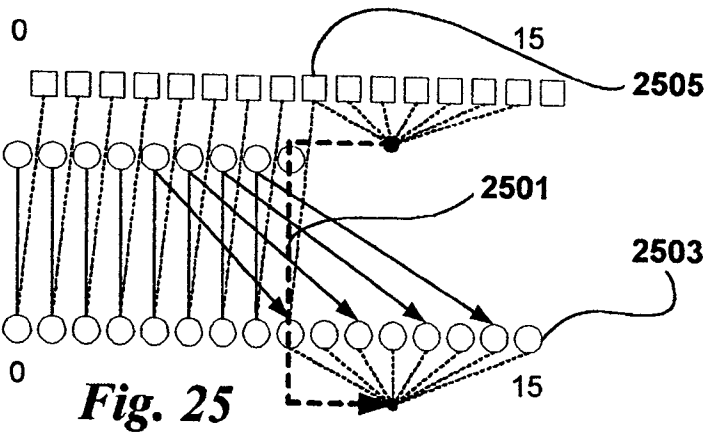
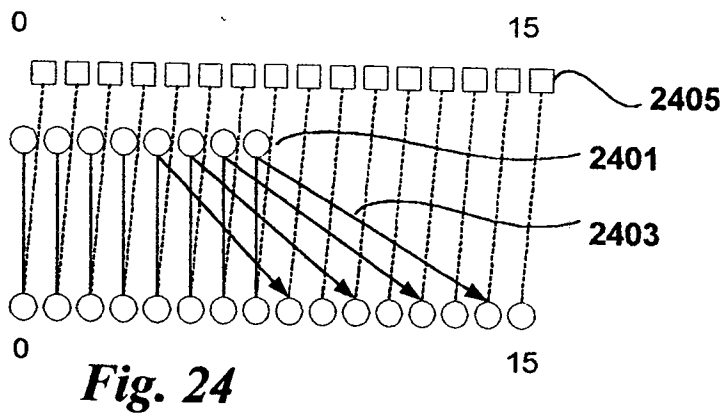
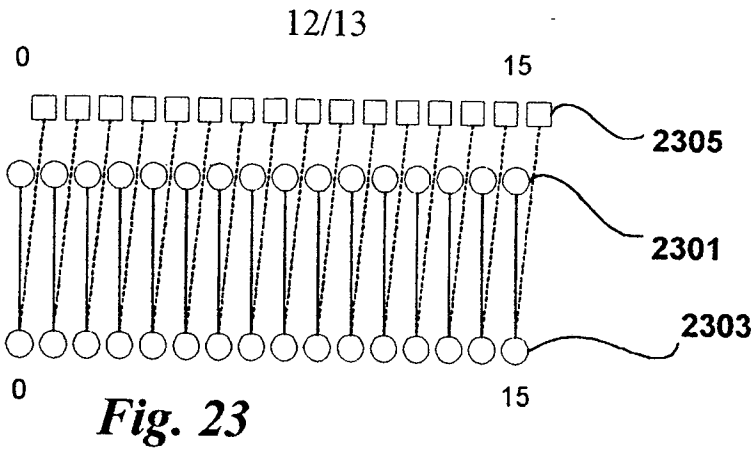


Fig. 22



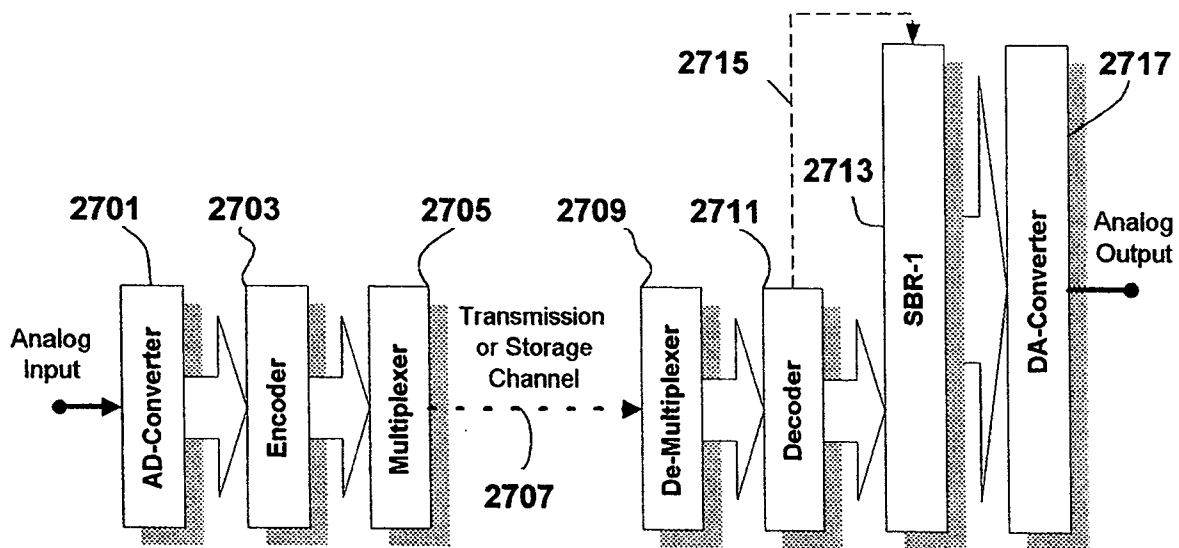


Fig. 27

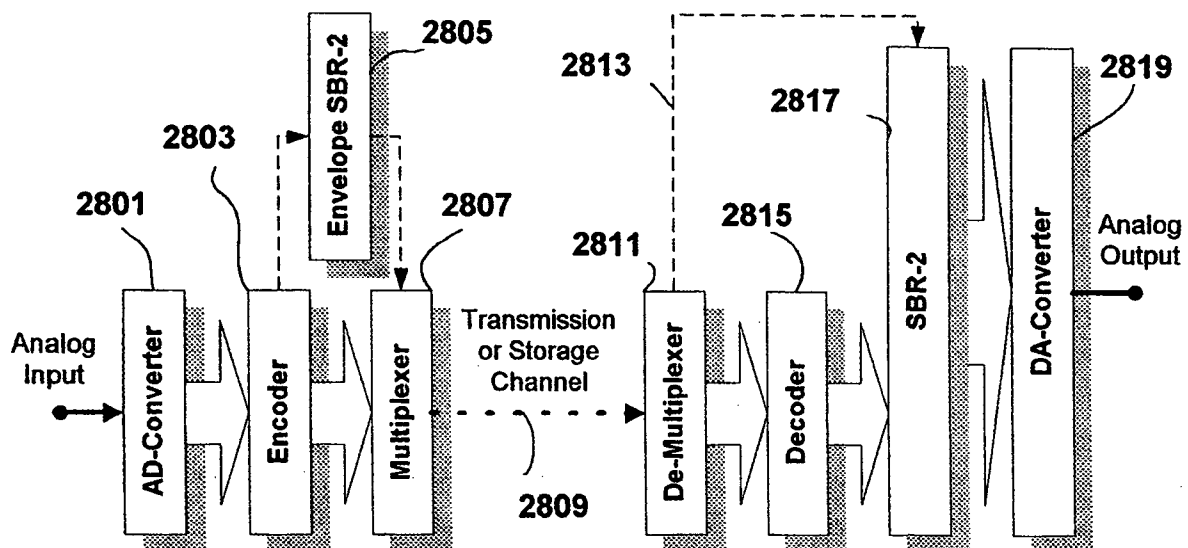


Fig. 28

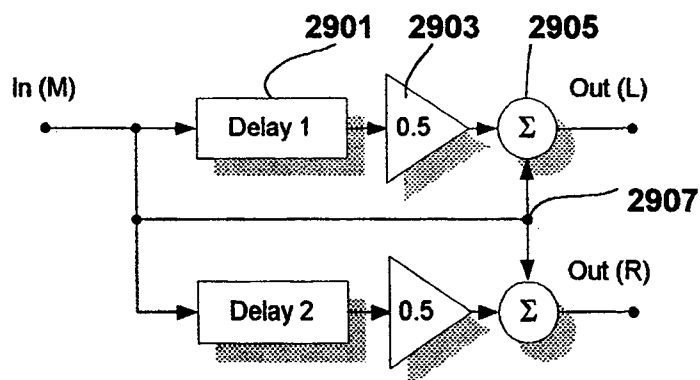


Fig. 29

## COPIA

SE ACOMPAÑAN NUEVA HOJA TECNICA POR DUPLICADO Y CARATULA, JUNTO CON LA MEMORIA DESCRIPTIVA, REIVINDICACIONES Y TRADUCCION CERTIFICADA CORRESPONDIENTE EN EL EXPEDIENTE DE LA SOLICITUD DE PATENTE ACTA N° P 10 01 01129 Y SE ACLARA QUE LA PRESENTE ACTUACION NO DEBE TOMARSE COMO CONTESTACION A VISTA ALGUNA.

Buenos Aires, 16 ABR 2010

Sr. Comisario de la Administración Nacional de Patentes:

EDUARDO A. RUSSOMANNO  
Agente 871

, con domicilio legal constituido en Av. Leandro N. Alem 928, 7° piso, por poder de FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V., en el expediente de la solicitud de patente acta N° P 10 01 01129, al Sr. Comisario com

De acuerdo con lo manifestado en el formulario de presentación y dando cumplimiento a lo oportunamente solicitado por esa Administración, viene por el presente acompañando dentro del plazo previsto por el art. 1, inciso 'E' apartado 4 - Ley N° 19.549, una traducción al castellano, certificada por traductor público matriculado, de la versión en idioma inglés originalmente presentada, de modo tal de satisfacer los requerimientos del art. 28 del Decreto N° 1759/72 T.O. 1991.

Se acompañan además, nueva hoja técnica por duplicado y carátula como así también la memoria descriptiva y las reivindicaciones para el trámite regular de la solicitud.

Finalmente, se acompaña la copia complementaria de la primera reivindicación.

Se indica que el apoderado del solicitante acredita personería según inscripción en el Registro de Poderes de ese Instituto bajo los números 18998, 27066, 56015, 56073, 56074 y 56075.

Sin otro particular, saludo a Ud. muy atentamente.

INPI

Trámite: 10072019 Importe: \$0.-

Fecha/Hora: 16/04/2010 13:42:3.770

Ant.: 20100101129 Patente  
CARRELL JORGE EDUARDO



202

|                                                                                                                                        |                                                                                                                                                        |                              |
|----------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|
| <p>REPUBLICA ARGENTINA<br/>(Asi)</p>  <p>I.N.P.I.</p> | <p>SOLICITUD DE</p> <p>PATENTE DE INVENCION: <input checked="" type="checkbox"/></p> <p>CERTIFICADO DE MODELO DE UTILIDAD <input type="checkbox"/></p> | <p>Fecha de Presentación</p> |
|----------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|

**I. Solicitante**

Acta N°

1) Apellido y Nombre/Denominación o Razón Social: FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.  
(TIENE N° EN EL REGISTRO DE PODERES)

2) Documento de Identidad:

INPI Exp.: 20100101129

Estado Civil:



Trámite: 10062926 PATENTES Importe: \$1280.-

Nupcias:

Fecha/Hora: 05/04/2010 13:47:6.840

Nombre del Cónyuge:

Agente: O' FARRELL JORGE EDUARDO

3) Caja de Jubilación o AFJP:

N° de CUIL O CUIT:

IVA:

4) Inscripto en el Registro Industrial de la Nación (Decreto-Ley 19.971/72) N°

5) Domicilio Real: HANSASTRASSE 27C, 80686 MUNICH, ALEMANIA (DE)

Legal: AVDA. LEANDRO N. ALEM 928, 7° PISO, CAPITAL FEDERAL

**II Objeto**

6) Título de la Invención: "APARATO, MÉTODO Y PROGRAMA DE COMPUTACIÓN PARA GENERAR UNA REPRESENTACIÓN DE UNA SEÑAL DE ANCHO DE BANDA AMPLIADO EN BASE A LA REPRESENTACIÓN DE UNA SEÑAL DE ENTRADA UTILIZANDO UNA COMBINACIÓN DE UNA AMPLIACIÓN DE ANCHO DE BANDA ARMÓNICO Y UNA AMPLIACIÓN DE ANCHO DE BANDA NO ARMÓNICO" ✓

7) Carácter de la Patente:

a) Definitiva, por el término de V E I N T E años

b) Adicional a la Patente N°

8) Ley 17.011. Fecha prioridad: 02/04/2009 \* 09/04/2009 \* 30/12/2009

País: US \* US \* EP

N°: 61/166125 \* 61/168068 \* 09181008

**III. Documentación acompañada**

9) Se acompaña:

a) Comprobante pago servicio requerido

b) Formulario anexo en duplicado

c) Carátula en duplicado

☒  
☒  
☒

- d) Memoria descriptiva en duplicado ☒
- e) Reivindicaciones en duplicado firmadas ☒
- f) 2 copias de la 1ª reivindicación ☒
- g) Dibujos en triplicado ☒
- h) Número de planchas ☐
- i) Reducciones ☐
- j) Copia certificada (Ley 17.011) y traducción certif. por traductor público + CD traducc. ☐
- k) Documento de cesión ☐
- l) Dibujos informales ☒

#### V. Sociedades

10) Sociedad, representada por: EDUARDO A. RUSSOMANNO  
Agente 871

quien declara bajo juramento que inviste el carácter de APODERADO

que su mandato se encuentra vigente y que la Sociedad se halla inscripta en el

Fecha:

Nº

Fº

Lib.

Tº

#### V. Mandato

11) Poder inscripto en:

Registrado en el INPI bajo Nº: 61.894

Otros Registros: PODERES SUSTITUTOS Nº: 4069, 4969, 11227, 18998, 27066, 56015, 56073, 56074 y 56075

12) En este acto, autoriza a MIGUEL B.O'FARRELL AG.415, PEDRO E.BURG AG.741, SERGIO M.ELLMANN AG.837, SANTIAGO R.O'CONOR AG.845, JULIO FERNANDEZ MOUJAN AG.924, MARTIN BENSADON AG.1196, GUSTAVO P.GIAY AG.1483, EDUARDO A.RUSSOMANNO AG.871, IRIS V.QUADRIO AG.864 Y/O IVÁN A.POLI AG.563

Para tramitar este asunto hasta su terminación con facultades para firmar documentos, desistir, si fuere menester y solicitar testimonios y percibir.

13) Se acompaña copia del poder ☐

14) Caja Jubilación o AFJP CO. Y ACT. CIVILES

Nº CUIL o CUIT: 30-52927856-6

15) Agente Nº: E-195

#### VI. Declaración

16) A los efectos del Decreto sin número del 7 de Junio de 1901 (sobre patentabilidad en el extranjero) manifiesta que el presente invento NO ha sido patentado en el extranjero.

VI. Observaciones: PIDE PLAZO DE 90 DIAS PARA ACOMPAÑAR COPIAS CERTIFICADAS DE LOS DOCUMENTOS DE PRIORIDAD Y DIBUJOS FORMALES//SE ACOMPAÑA VERSION COMPLETA EN IDIOMA INGLES, MANIFESTANDO QUE SE ACOMPAÑARA A LA BREVEDAD TRADUCCION CERTIFICADA DEL TEXTO ADJUNTADO//SE PRESENTA ESTA SOLICITUD DE ACUERDO AL ART. 4C3 DEL TEXTO DE LISBOA (CONVENIO DE PARIS)//SE ABONA TASA DE EXAMEN DE FONDO EN ESTA PRESENTACION

(Firma del autorizado)

EDUARDO A. RUSSOMANNO  
Agente 871  
(Firma del solicitante)

"2010 – Año del Bicentenario de la Revolución de Mayo"



*Instituto Nacional de la Propiedad Industrial*

INPI Exp.: **20100101129**



Trámite: 10062926 PATENTES Importe: \$1280.-

Fecha/Hora: 05/04/2010 13:47:6.840

Agente: O' FARRELL JORGE EDUARDO

BUENOS AIRES,

Advirtiéndose que se efectuó la presentación de la solicitud referida, en idioma extranjero, sin acompañar la correspondiente traducción, y que, asimismo, por tal razón no puede verificarse en esta instancia si se encuentran reunidos los extremos dispuestos por el Art. 18 de la Ley 24.481 (t.o. 1996) para otorgar fecha de presentación a la solicitud de la referencia, queda el solicitante:

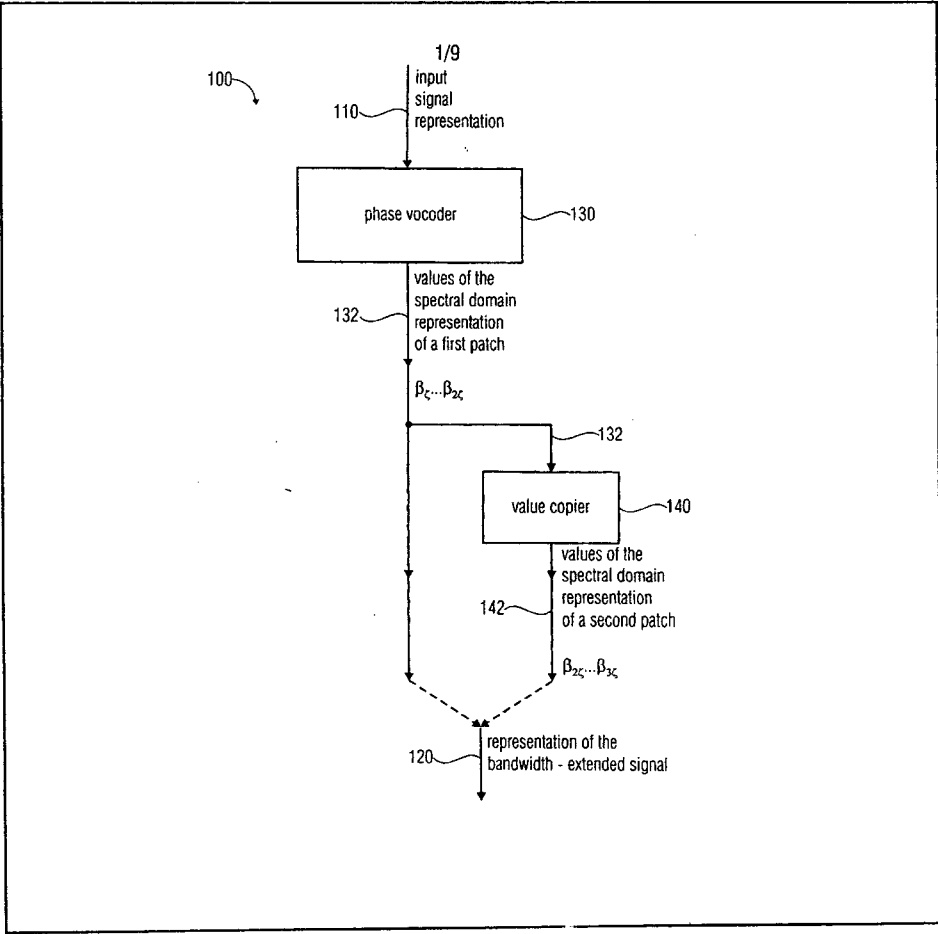
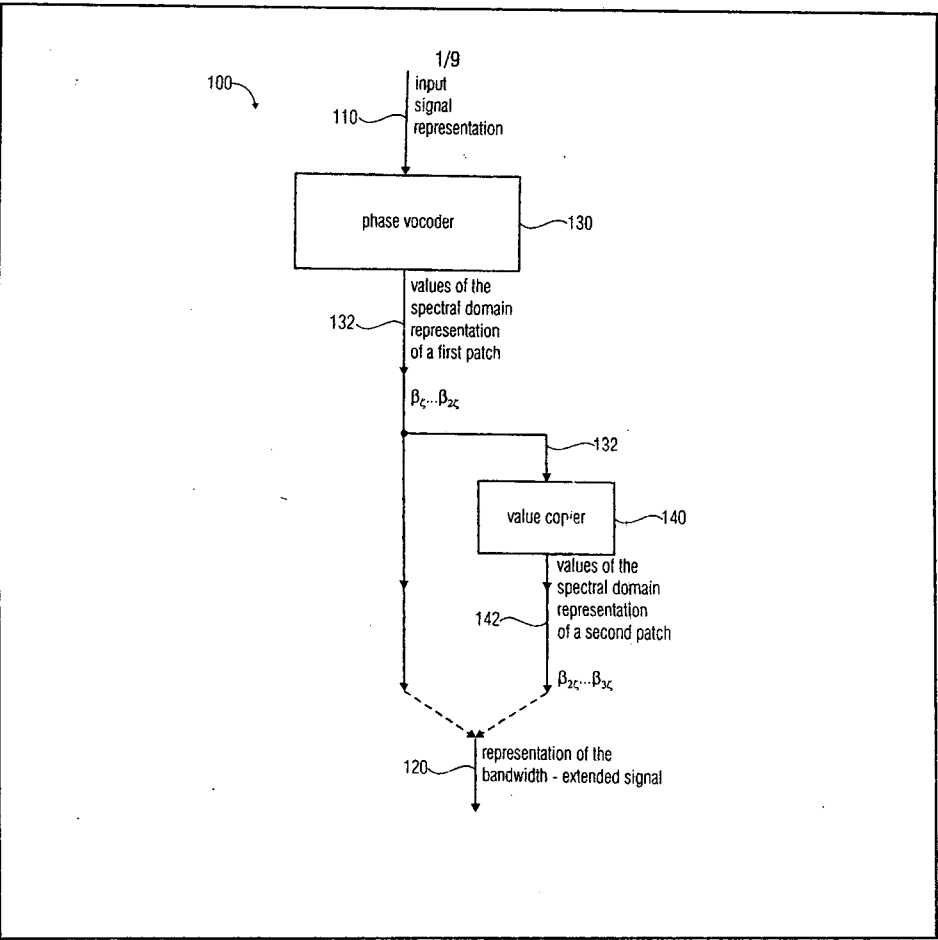
- 1.- Intimado, en este acto, para que dentro del plazo de diez días hábiles (Art. 1 inciso "E" apartado 4 - Ley 19.549) acompañe las traducciones certificadas por traductor matriculado (Art. 28 del Decreto 1759/72 - t.o. 1.991) de los documentos agregados.
- 2.- Apercebido que, de no cumplimentar la intimación en el término fijado, se tendrá precluida toda ulterior oportunidad para hacerlo y consecuentemente no tendrá fecha de presentación la solicitud de referencia, conforme lo dispone el Art. 18 inc. "C" de la Ley 24.481 (T.O. 1996).
- 3.- Intimado, para que en el mismo plazo señalado en el punto 1, dé cumplimiento con lo dispuesto en el Art. 18 de la Ley 24.481, y en tal supuesto se le hace saber que la fecha de presentación de la solicitud de referencia, será la del efectivo cumplimiento del citado artículo dentro del plazo otorgado, bajo apercibimiento que, de no cumplimentar la intimación en el plazo fijado, se tendrá por precluida toda ulterior oportunidad para hacerlo y consecuentemente se procederá al dictado de la disposición denegatoria sin más trámite.

A los efectos de darle carácter fehaciente a la presente interpelación, se labran y fechan dos ejemplares que son suscriptos por el funcionario interviniente y el solicitante, recibiendo este último el suyo e integrándose el restante al expediente. -

CONSTE.-

NOTIFIQUESE,-

CARLOS M. CALVO  
DEL DEPARTAMENTO ADMINISTRATIVO  
DE LA NACIONAL DE PATENTES







Bogotá D.C.,  
2020

Doctor(a)  
CARDENAS NAVAS DARIO  
Apoderado(a) y/o Representante de  
FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV

|            |                            |                   |
|------------|----------------------------|-------------------|
| REFERENCIA | Radicación No.             | 10 131388         |
|            | Solicitud internacional    | PCT/EP2010/054422 |
|            | Fecha inicio fase nacional | 02/11/2012        |
|            | Trámite                    | 11                |
|            | Evento                     | 378               |
|            | Actuación                  | 430               |
|            | Oficio No.                 | <b>15773</b>      |

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 el poder debidamente otorgado, con el lleno de los requisitos exigidos por los artículos 63 y subsiguientes del Código de Procedimiento Civil. Adicionalmente, deberá presentar el documento que acredita la existencia y representación legal de la sociedad solicitante, cuando ésta fuere persona jurídica.
2. 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.

Continúa en la página siguiente...



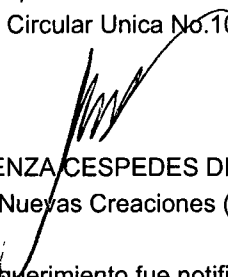
Continuación radicado No: 10 131388

208



**Industria y Comercio**  
SUPERINTENDENCIA

Notifíquese el presente oficio conforme a lo dispuesto en el numeral 5.2, literal e), del capítulo quinto del título primero de la Circular Unica No.10 de 2001.

  
ALIX CARMENZA CESPEDES DE VERGEL  
Directora de Nuevas Creaciones (c)

El anterior requerimiento fue notificado por fijación en lista de fecha \_\_\_\_\_  
adpa11

03 DIC. 2010