



- (51) **International Patent Classification:**  
G10L 19/008 (2013.01)
- (21) **International Application Number:**  
PCT/EP2014/072570
- (22) **International Filing Date:**  
21 October 2014 (21.10.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**

61/893,770	21 October 2013 (21.10.2013)	US
61/974,544	3 April 2014 (03.04.2014)	US
62/037,693	15 August 2014 (15.08.2014)	US
- (71) **Applicant:** DOLBY INTERNATIONAL AB [NL/SE]; Apollo Building, 3E, Herikerbergweg 1-35, NL-1101 CN Amsterdam (NL).
- (72) **Inventors:** VILLEMOES, Lars; c/o Dolby Sweden AB, Gävlegatan 12A, S-113 30 Stockholm (SE). LEHTONEN, Heidi-Maria; c/o Dolby Sweden AB, Gävlegatan 12A, S-113 30 Stockholm (SE). PURNHAGEN, Heiko; c/o Dolby Sweden AB, Gävlegatan 12A, S-113 30 Stockholm (SE). HIRVONEN, Toni; c/o Dolby Sweden AB, Gävlegatan 12A, S-113 30 Stockholm (SE).
- (74) **Agent:** DOLBY INTERNATIONAL AB PATENT GROUP EUROPE; Apollo Building, 3E, Herikerbergweg 1-35, NL-1101 CN Amsterdam (NL).

- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

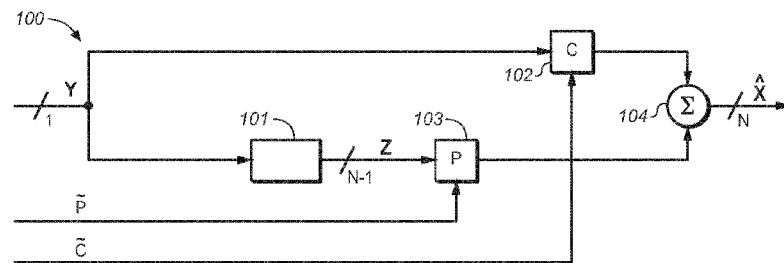
**Declarations under Rule 4.17:**

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

**Published:**

- with international search report (Art. 21(3))

(54) **Title:** PARAMETRIC RECONSTRUCTION OF AUDIO SIGNALS



**FIG. 1**

(57) **Abstract:** An encoding system (400) encodes an  $N$ -channel audio signal ( $X$ ), wherein  $N \geq 3$ , as a single-channel downmix signal ( $Y$ ) together with dry and wet upmix parameters ( $C$ ,  $P$ ). In a decoding system (200), a decorrelating section (101) outputs, based on the downmix signal, an  $(N-1)$ -channel decorrelated signal ( $Z$ ); a dry upmix section (102) maps the downmix signal linearly in accordance with dry upmix coefficients ( $C$ ) determined based on the dry upmix parameters; a wet upmix section (103) populates an intermediate matrix based on the wet upmix parameters and knowing that the intermediate matrix belongs to a predefined matrix class, obtains wet upmix coefficients ( $P$ ) by multiplying the intermediate matrix by a predefined matrix, and maps the decorrelated signal linearly in accordance with the wet upmix coefficients; and a combining section (104) combines outputs from the upmix sections to obtain a reconstructed signal ( $X$ ) corresponding to the signal to be reconstructed.

WO 2015/059153 A1

## PARAMETRIC RECONSTRUCTION OF AUDIO SIGNALS

### Cross-Reference to Related Applications

This application claims the benefit of priority to United States Provisional  
5 Patent Application No. 61/893,770 filed 21 October 2013; United States Provisional  
Patent Application No. 61/974,544 filed 3 April 2014; and United States Provisional  
Patent Application No. 62/037,693 filed 15 August 2014, each of which is hereby  
incorporated by reference in its entirety.

### 10 Technical Field of the Invention

The invention disclosed herein generally relates to encoding and decoding of  
audio signals, and in particular to parametric reconstruction of a multichannel audio  
signal from a downmix signal and associated metadata.

### 15 Background of the Invention

Audio playback systems comprising multiple loudspeakers are frequently used  
to reproduce an audio scene represented by a multichannel audio signal, wherein the  
respective channels of the multichannel audio signal are played back on respective  
loudspeakers. The multichannel audio signal may for example have been recorded  
20 via a plurality of acoustic transducers or may have been generated by audio  
authoring equipment. In many situations, there are bandwidth limitations for  
transmitting the audio signal to the playback equipment and/or limited space for  
storing the audio signal in a computer memory or on a portable storage device. There  
exist audio coding systems for parametric coding of audio signals, so as to reduce  
25 the bandwidth or storage size needed. On an encoder side, these systems typically  
downmix the multichannel audio signal into a downmix signal, which typically is a  
mono (one channel) or a stereo (two channels) downmix, and extract side information  
describing the properties of the channels by means of parameters like level  
differences and cross-correlation. The downmix and the side information are then  
30 encoded and sent to a decoder side. On the decoder side, the multichannel audio  
signal is reconstructed, i.e. approximated, from the downmix under control of the  
parameters of the side information.

In view of the wide range of different types of devices and systems available  
for playback of multichannel audio content, including an emerging segment aimed at

end-users in their homes, there is a need for new and alternative ways to efficiently encode multichannel audio content, so as to reduce bandwidth requirements and/or the required memory size for storage, and/or to facilitate reconstruction of the multichannel audio signal at a decoder side.

5

#### Brief Description of the Drawings

In what follows, example embodiments will be described in greater detail and with reference to the accompanying drawings, on which:

Fig. 1 is a generalized block diagram of a parametric reconstruction section for reconstructing a multichannel audio signal based on a single-channel downmix signal and associated dry and wet upmix parameters, according to an example embodiment;

Fig. 2 is a generalized block diagram of an audio decoding system comprising the parametric reconstruction section depicted in Fig. 1, according to an example embodiment;

Fig. 3 is a generalized block diagram of a parametric encoding section for encoding a multichannel audio signal as a single-channel downmix signal and associated metadata, according to an example embodiment;

Fig. 4 is a generalized block diagram of an audio encoding system comprising the parametric encoding section depicted in Fig. 3, according to an example embodiment;

Figures 5–11 illustrate alternative ways to represent an 11.1 channel audio signal by means of downmix channels, according to example embodiments;

Figures 12–13 illustrate alternative ways to represent a 13.1 channel audio signal by means of downmix channels, according to example embodiments; and

Figures 14–16 illustrate alternative ways to represent a 22.2 channel audio signal by means of downmix signals, according to example embodiments.

All the figures are schematic and generally only show parts which are necessary in order to elucidate the invention, whereas other parts may be omitted or merely suggested.

### Description of Example Embodiments

As used herein, an *audio signal* may be a pure audio signal, an audio part of an audiovisual signal or multimedia signal or any of these in combination with metadata.

5 As used herein, a *channel* is an audio signal associated with a predefined/fixed spatial position/orientation or an undefined spatial position such as "left" or "right".

#### I. Overview

10 According to a first aspect, example embodiments propose audio decoding systems as well as methods and computer program products for reconstructing an audio signal. The proposed decoding systems, methods and computer program products, according to the first aspect, may generally share the same features and advantages.

15 According to example embodiments, there is provided a method for reconstructing an  $N$ -channel audio signal, wherein  $N \geq 3$ . The method comprises receiving a single-channel downmix signal, or a channel of a multichannel downmix signal carrying data for reconstruction of more audio signals, together with associated dry and wet upmix parameters; computing a first signal with a plurality of ( $N$ ) channels, referred to as a dry upmix signal, as a linear mapping of the downmix  
20 signal, wherein a set of dry upmix coefficients is applied to the downmix signal as part of computing the dry upmix signal; generating an  $(N - 1)$ -channel decorrelated signal based on the downmix signal; computing a further signal with a plurality of ( $N$ ) channels, referred to as a wet upmix signal, as a linear mapping of the decorrelated signal, wherein a set of wet upmix coefficients is applied to the channels of the  
25 decorrelated signal as part of computing the wet upmix signal; and combining the dry and wet upmix signals to obtain a multidimensional reconstructed signal corresponding to the  $N$ -channel audio signal to be reconstructed. The method further comprises determining the set of dry upmix coefficients based on the received dry upmix parameters; populating an intermediate matrix having more elements than the  
30 number of received wet upmix parameters, based on the received wet upmix parameters and knowing that the intermediate matrix belongs to a predefined matrix class; and obtaining the set of wet upmix coefficients by multiplying the intermediate matrix by a predefined matrix, wherein the set of wet upmix coefficients corresponds

to the matrix resulting from the multiplication and includes more coefficients than the number of elements in the intermediate matrix.

In this example embodiment, the number of wet upmix coefficients employed for reconstructing the  $N$ -channel audio signal is larger than the number of received wet upmix parameters. By exploiting knowledge of the predefined matrix and the predefined matrix class to obtain the wet upmix coefficients from the received wet upmix parameters, the amount of information needed to enable reconstruction of the  $N$ -channel audio signal may be reduced, allowing for a reduction of the amount of metadata transmitted together with the downmix signal from an encoder side. By reducing the amount of data needed for parametric reconstruction, the required bandwidth for transmission of a parametric representation of the  $N$ -channel audio signal, and/or the required memory size for storing such a representation, may be reduced.

The  $(N - 1)$ -channel decorrelated signal serves to increase the dimensionality of the content of the reconstructed  $N$ -channel audio signal, as perceived by a listener. The channels of the  $(N - 1)$ -channel decorrelated signal may have at least approximately the same spectrum as the single-channel downmix signal, or may have spectra corresponding to rescaled/normalized versions of the spectrum of the single-channel downmix signal, and may form, together with the single-channel downmix signal,  $N$  at least approximately mutually uncorrelated channels. In order to provide a faithful reconstruction of the channels of the  $N$ -channel audio signal, each of the channels of the decorrelated signal preferably has such properties that it is perceived by a listener as similar to the downmix signal. Hence, although it is possible to synthesize mutually uncorrelated signals with a given spectrum from e.g. white noise, the channels of the decorrelated signal are preferably derived by processing the downmix signal, e.g. including applying respective all-pass filters to the downmix signal or recombining portions of the downmix signal, so as to preserve as many properties as possible, especially locally stationary properties, of the downmix signal, including relatively more subtle, psycho-acoustically conditioned properties of the downmix signal, such as timbre.

Combining the wet and dry upmix signals may include adding audio content from respective channels of the wet upmix signal to audio content of the respective

corresponding channels of the dry upmix signal, such as additive mixing on a per-sample or per-transform-coefficient basis.

The predefined matrix class may be associated with known properties of at least some matrix elements which are valid for all matrices in the class, such as  
5 certain relationships between some of the matrix elements, or some matrix elements being zero. Knowledge of these properties allows for populating the intermediate matrix based on fewer wet upmix parameters than the full number of matrix elements in the intermediate matrix. The decoder side has knowledge at least of the properties of, and relationships between, the elements it needs to compute all matrix elements  
10 on the basis of the fewer wet upmix parameters.

By the dry upmix signal being a linear mapping of the downmix signal is meant that the dry upmix signal is obtained by applying a first linear transformation to the downmix signal. This first transformation takes one channel as input and provides  $N$  channels as output, and the dry upmix coefficients are coefficients defining the  
15 quantitative properties of this first linear transformation.

By the wet upmix signal being a linear mapping of the decorrelated signal is meant that the wet upmix signal is obtained by applying a second linear transformation to the decorrelated signal. This second transformation takes  $N - 1$  channels as input and provides  $N$  channels as output, and the wet upmix coefficients  
20 are coefficients defining the quantitative properties of this second linear transformation.

In an example embodiment, receiving the wet upmix parameters may include receiving  $N(N - 1)/2$  wet upmix parameters. In the present example embodiment, populating the intermediate matrix may include obtaining values for  $(N - 1)^2$  matrix  
25 elements based on the received  $N(N - 1)/2$  wet upmix parameters and knowing that the intermediate matrix belongs to the predefined matrix class. This may include inserting the values of the wet upmix parameters immediately as matrix elements, or processing the wet upmix parameters in a suitable manner for deriving values for the matrix elements. In the present example embodiment, the predefined matrix may  
30 include  $N(N - 1)$  elements, and the set of wet upmix coefficients may include  $N(N - 1)$  coefficients. For example, receiving the wet upmix parameters may include receiving no more than  $N(N - 1)/2$  independently assignable wet upmix parameters and/or the number of received wet upmix parameters may be no more than half the

number of wet upmix coefficients employed for reconstructing the  $N$ -channel audio signal.

It is to be understood that omitting a contribution from a channel of the decorrelated signal when forming a channel of the wet upmix signal as a linear  
5 mapping of the channels of the decorrelated signal corresponds to applying a coefficient with the value zero to that channel, i.e. omitting a contribution from a channel does not affect the number of coefficients applied as part of the linear mapping.

In an example embodiment, populating the intermediate matrix may include  
10 employing the received wet upmix parameters as elements in the intermediate matrix. Since the received wet upmix parameters are employed as elements in the intermediate matrix without being processed any further, the complexity of the computations required for populating the intermediate matrix, and to obtain the upmix coefficients may be reduced, allowing for a computationally more efficient  
15 reconstruction of the  $N$ -channel audio signal.

In an example embodiment, receiving the dry upmix parameters may include receiving  $(N - 1)$  dry upmix parameters. In the present example embodiment, the set of dry upmix coefficients may include  $N$  coefficients, and the set of dry upmix coefficients is determined based on the received  $(N - 1)$  dry upmix parameters and  
20 based on a predefined relation between the coefficients in the set of dry upmix coefficients. For example, receiving the dry upmix parameters may include receiving no more than  $(N - 1)$  independently assignable dry upmix parameters. For example, the downmix signal may be obtainable, according to a predefined rule, as a linear mapping of the  $N$ -channel audio signal to be reconstructed, and the predefined  
25 relation between the dry upmix coefficients may be based on the predefined rule.

In an example embodiment, the predefined matrix class may be one of: lower or upper triangular matrices, wherein known properties of all matrices in the class include predefined matrix elements being zero; symmetric matrices, wherein known properties of all matrices in the class include predefined matrix elements (on either  
30 side of the main diagonal) being equal; and products of an orthogonal matrix and a diagonal matrix, wherein known properties of all matrices in the class include known relations between predefined matrix elements. In other words, the predefined matrix class may be the class of lower triangular matrices, the class of upper triangular

matrices, the class of symmetric matrices or the class of products of an orthogonal matrix and a diagonal matrix. A common property of each of the above classes is that its dimensionality is less than the full number of matrix elements.

In an example embodiment, the downmix signal may be obtainable, according to a predefined rule, as a linear mapping of the  $N$ -channel audio signal to be reconstructed. In the present example embodiment, the predefined rule may define a predefined downmix operation, and the predefined matrix may be based on vectors spanning the kernel space of the predefined downmix operation. For example, the rows or columns of the predefined matrix may be vectors forming a basis, e.g. an orthonormal basis, for the kernel space of the predefined downmix operation.

In an example embodiment, receiving the single-channel downmix signal together with associated dry and wet upmix parameters may include receiving a time segment or time/frequency tile of the downmix signal together with dry and wet upmix parameters associated with that time segment or time/frequency tile. In the present example embodiment, the multidimensional reconstructed signal may correspond to a time segment or time/frequency tile of the  $N$ -channel audio signal to be reconstructed. In other words, the reconstruction of the  $N$ -channel audio signal may in at least some example embodiments be performed one time segment or time/frequency tile at a time. Audio encoding/decoding systems typically divide the time-frequency space into time/frequency tiles, e.g. by applying suitable filter banks to the input audio signals. By a time/frequency tile is generally meant a portion of the time-frequency space corresponding to a time interval/segment and a frequency sub-band.

According to example embodiments, there is provided an audio decoding system comprising a first parametric reconstruction section configured to reconstruct an  $N$ -channel audio signal based on a first single-channel downmix signal and associated dry and wet upmix parameters, wherein  $N \geq 3$ . The first parametric reconstruction section comprises a first decorrelating section configured to receive the first downmix signal and to output, based thereon, a first  $N - 1$ -channel decorrelated signal. The first parametric reconstruction section also comprises a first dry upmix section configured to: receive the dry upmix parameters and the downmix signal; determine a first set of dry upmix coefficients based on the dry upmix parameters; and output a first dry upmix signal computed by mapping the first

downmix signal linearly in accordance with the first set of dry upmix coefficients. In other words, the channels of the first dry upmix signal are obtained by multiplying the single-channel downmix signal by respective coefficients, which may be the dry upmix coefficients themselves, or which may be coefficients controllable via the dry upmix coefficients. The first parametric reconstruction section further comprises a first wet upmix section configured to: receive the wet upmix parameters and the first decorrelated signal; populate a first intermediate matrix having more elements than the number of received wet upmix parameters, based on the received wet upmix parameters and knowing that the first intermediate matrix belongs to a first predefined matrix class, i.e. by employing properties of certain matrix elements known to hold for all matrices in the predefined matrix class; obtain a first set of wet upmix coefficients by multiplying the first intermediate matrix by a first predefined matrix, wherein the first set of wet upmix coefficients corresponds to the matrix resulting from the multiplication and includes more coefficients than the number of elements in the first intermediate matrix; and output a first wet upmix signal computed by mapping the first decorrelated signal linearly in accordance with the first set of wet upmix coefficients, i.e. by forming linear combinations of the channels of the decorrelated signal employing the wet upmix coefficients. The first parametric reconstruction section also comprises a first combining section configured to receive the first dry upmix signal and the first wet upmix signal and to combine these signals to obtain a first multidimensional reconstructed signal corresponding to the  $N$ -dimensional audio signal to be reconstructed.

In an example embodiment, the audio decoding system may further comprise a second parametric reconstruction section operable independently of the first parametric reconstruction section and configured to reconstruct an  $N_2$ -channel audio signal based on a second single-channel downmix signal and associated dry and wet upmix parameters, wherein  $N_2 \geq 2$ . It may for example hold that  $N_2 = 2$  or that  $N_2 \geq 3$ . In the present example embodiment, the second parametric reconstruction section may comprise a second decorrelating section, a second dry upmix section, a second wet upmix section and a second combining section, and the sections of the second parametric reconstruction section may be configured analogously to the corresponding sections of the first parametric reconstruction section. In the present example embodiment, the second wet upmix section may be configured to employ a

second intermediate matrix belonging to a second predefined matrix class and a second predefined matrix. The second predefined matrix class and the second predefined matrix may be different than, or equal to, the first predefined matrix class and the first predefined matrix, respectively.

5           In an example embodiment, the audio decoding system may be adapted to reconstruct a multichannel audio signal based on a plurality of downmix channels and associated dry and wet upmix parameters. In the present example embodiment, the audio decoding system may comprise: a plurality of reconstruction sections, including parametric reconstruction sections operable to independently reconstruct  
10   respective sets of audio signal channels based on respective downmix channels and respective associated dry and wet upmix parameters; and a control section configured to receive signaling indicating a coding format of the multichannel audio signal corresponding to a partition of the channels of the multichannel audio signal into sets of channels represented by the respective downmix channels and, for at  
15   least some of the downmix channels, by respective associated dry and wet upmix parameters. In the present example embodiment, the coding format may further correspond to a set of predefined matrices for obtaining wet upmix coefficients associated with at least some of the respective sets of channels based on the respective wet upmix parameters. Optionally, the coding format may further  
20   correspond to a set of predefined matrix classes indicating how respective intermediate matrices are to be populated based on the respective sets of wet upmix parameters.

          In the present example embodiment, the decoding system may be configured to reconstruct the multichannel audio signal using a first subset of the plurality of  
25   reconstruction sections, in response to the received signaling indicating a first coding format. In the present example embodiment, the decoding system may be configured to reconstruct the multichannel audio signal using a second subset of the plurality of reconstruction sections, in response to the received signaling indicating a second coding format, and at least one of the first and second subsets of the reconstruction  
30   sections may comprise the first parametric reconstruction section.

          Depending on the composition of the audio content of the multichannel audio signal, the available bandwidth for transmission from an encoder side to a decoder side, the required playback quality as perceived by a listener and/or the required fidelity of the audio signal as reconstructed on a decoder side, the most appropriate

coding format may differ between different applications and/or time periods. By supporting multiple coding formats for the multichannel audio signal, the audio decoding system in the present example embodiment allows an encoder side to employ a coding format more specifically suited for the current circumstances.

5           In an example embodiment, the plurality of reconstruction sections may include a single-channel reconstruction section operable to independently reconstruct a single audio channel based on a downmix channel in which no more than a single audio channel has been encoded. In the present example embodiment, at least one of the first and second subsets of the reconstruction sections may comprise the  
10 single-channel reconstruction section. Some channels of the multichannel audio signal may be particularly important for the overall impression of the multichannel audio signal, as perceived by a listener. By employing the single-channel reconstruction section to encode e.g. such a channel separately in its own downmix channel, while other channels are parametrically encoded together in other downmix  
15 channels, the fidelity of the multichannel audio signal as reconstructed may be increased. In some example embodiments, the audio content of one channel of the multichannel audio signal may be of a different type than the audio content of the other channels of the multichannel audio signal, and the fidelity of the multichannel audio signal as reconstructed may be increased by employing a coding format in  
20 which that channel is encoded separately in a downmix channel of its own.

          In an example embodiment, the first coding format may correspond to reconstruction of the multichannel audio signal from a lower number of downmix channels than the second coding format. By employing a lower number of downmix channels, the required bandwidth for transmission from an encoder side to a decoder  
25 side may be reduced. By employing a higher number of downmix channels, the fidelity and/or the perceived audio quality of the multichannel audio signal as reconstructed may be increased.

          According to a second aspect, example embodiments propose audio encoding systems as well as methods and computer program products for encoding a  
30 multichannel audio signal. The proposed encoding systems, methods and computer program products, according to the second aspect, may generally share the same features and advantages. Moreover, advantages presented above for features of decoding systems, methods and computer program products, according to the first

aspect, may generally be valid for the corresponding features of encoding systems, methods and computer program products according to the second aspect.

According to example embodiments, there is provided a method for encoding an  $N$ -channel audio signal as a single-channel downmix signal and metadata suitable for parametric reconstruction of the audio signal from the downmix signal and an  
5  $(N-1)$ -channel decorrelated signal determined based on the downmix signal, wherein  $N \geq 3$ . The method comprises: receiving the audio signal; computing, according to a predefined rule, the single-channel downmix signal as a linear mapping of the audio signal; and determining a set of dry upmix coefficients in order to define a linear  
10 mapping of the downmix signal approximating the audio signal, e.g. via a minimum mean square error approximation under the assumption that only the downmix signal is available for the reconstruction. The method further comprises determining an intermediate matrix based on a difference between a covariance of the audio signal as received and a covariance of the audio signal as approximated by the linear  
15 mapping of the downmix signal, wherein the intermediate matrix when multiplied by a predefined matrix corresponds to a set of wet upmix coefficients defining a linear mapping of the decorrelated signal as part of parametric reconstruction of the audio signal, and wherein the set of wet upmix coefficients includes more coefficients than the number of elements in the intermediate matrix. The method further comprises  
20 outputting the downmix signal together with dry upmix parameters, from which the set of dry upmix coefficients is derivable, and wet upmix parameters, wherein the intermediate matrix has more elements than the number of output wet upmix parameters, and wherein the intermediate matrix is uniquely defined by the output wet upmix parameters provided that the intermediate matrix belongs to a predefined  
25 matrix class.

A parametric reconstruction copy of the audio signal at a decoder side includes, as one contribution, a dry upmix signal formed by the linear mapping of the downmix signal and, as a further contribution, a wet upmix signal formed by the linear mapping of the decorrelated signal. The set of dry upmix coefficients defines the  
30 linear mapping of the downmix signal and the set of wet upmix coefficients defines the linear mapping of the decorrelated signals. By outputting wet upmix parameters which are fewer than the number of wet upmix coefficients, and from which the wet upmix coefficients are derivable based on the predefined matrix and the predefined

matrix class, the amount of information sent to a decoder side to enable reconstruction of the  $N$ -channel audio signal may be reduced. By reducing the amount of data needed for parametric reconstruction, the required bandwidth for transmission of a parametric representation of the  $N$ -channel audio signal, and/or the  
5 required memory size for storing such a representation, may be reduced.

The intermediate matrix may be determined based on the difference between the covariance of the audio signal as received and the covariance of the audio signal as approximated by the linear mapping of the downmix signal, e.g. for a covariance of the signal obtained by the linear mapping of the decorrelated signal to supplement  
10 the covariance of the audio signal as approximated by the linear mapping of the downmix signal.

In an example embodiment, determining the intermediate matrix may include determining the intermediate matrix such that a covariance of the signal obtained by the linear mapping of the decorrelated signal, defined by the set of wet upmix  
15 coefficients, approximates, or substantially coincides with, the difference between the covariance of the audio signal as received and the covariance of the audio signal as approximated by the linear mapping of the downmix signal. In other words, the intermediate matrix may be determined such that a reconstruction copy of the audio signal, obtained as a sum of a dry upmix signal formed by the linear mapping of the  
20 downmix signal and a wet upmix signal formed by the linear mapping of the decorrelated signal completely, or at least approximately, reinstates the covariance of the audio signal as received.

In an example embodiment, outputting the wet upmix parameters may include outputting no more than  $N(N - 1)/2$  independently assignable wet upmix  
25 parameters. In the present example embodiment, the intermediate matrix may have  $(N - 1)^2$  matrix elements and may be uniquely defined by the output wet upmix parameters provided that the intermediate matrix belongs to the predefined matrix class. In the present example embodiment, the set of wet upmix coefficients may include  $N(N - 1)$  coefficients.

30 In an example embodiment, the set of dry upmix coefficients may include  $N$  coefficients. In the present example embodiments, outputting the dry upmix parameters may include outputting no more than  $N - 1$  dry upmix parameters, and

the set of dry upmix coefficients may be derivable from the  $N - 1$  dry upmix parameters using the predefined rule.

In an example embodiment, the determined set of dry upmix coefficients may define a linear mapping of the downmix signal corresponding to a minimum mean square error approximation of the audio signal, i.e. among the set of linear mappings of the downmix signal, the determined set of dry upmix coefficients may define the linear mapping which best approximates the audio signal in a minimum mean square sense.

According to example embodiments, there is provided an audio encoding system comprising a parametric encoding section configured to encode an  $N$ -channel audio signal as a single-channel downmix signal and metadata suitable for parametric reconstruction of the audio signal from the downmix signal and an  $(N - 1)$ -channel decorrelated signal determined based on the downmix signal, wherein  $N \geq 3$ . The parametric encoding section comprises: a downmix section configured to receive the audio signal and to compute, according to a predefined rule, the single-channel downmix signal as a linear mapping of the audio signal; and a first analyzing section configured to determine a set of dry upmix coefficients in order to define a linear mapping of the downmix signal approximating the audio signal. The parametric encoding section further comprises a second analyzing section configured to determine an intermediate matrix based on a difference between a covariance of the audio signal as received and a covariance of the audio signal as approximated by the linear mapping of the downmix signal, wherein the intermediate matrix when multiplied by a predefined matrix corresponds to a set of wet upmix coefficients defining a linear mapping of the decorrelated signal as part of parametric reconstruction of the audio signal, wherein the set of wet upmix coefficients includes more coefficients than the number of elements in the intermediate matrix. The parametric encoding section is further configured to output the downmix signal together with dry upmix parameters, from which the set of dry upmix coefficients is derivable, and wet upmix parameters, wherein the intermediate matrix has more elements than the number of output wet upmix parameters, and wherein the intermediate matrix is uniquely defined by the output wet upmix parameters provided that the intermediate matrix belongs to a predefined matrix class.

In an example embodiment, the audio encoding system may be configured to provide a representation of a multichannel audio signal in the form of a plurality of downmix channels and associated dry and wet upmix parameters. In the present example embodiment, the audio encoding system may comprise: a plurality of  
5 encoding sections, including parametric encoding sections operable to independently compute respective downmix channels and respective associated upmix parameters based on respective sets of audio signal channels. In the present example embodiment, the audio encoding system may further comprise a control section configured to determine a coding format for the multichannel audio signal  
10 corresponding to a partition of the channels of the multichannel audio signal into sets of channels to be represented by the respective downmix channels and, for at least some of the downmix channels, by respective associated dry and wet upmix parameters. In the present example embodiment, the coding format may further correspond to a set of predefined rules for computing at least some of the respective  
15 downmix channels. In the present example embodiment, the audio encoding system may be configured to encode the multichannel audio signal using a first subset of the plurality of encoding sections, in response to the determined coding format being a first coding format. In the present example embodiment, the audio encoding system may be configured to encode the multichannel audio signal using a second subset of  
20 the plurality of encoding sections, in response to the determined coding format being a second coding format, and at least one of the first and second subsets of the encoding sections may comprise the first parametric encoding section. In the present example embodiment, the control section may for example determine the coding format based on an available bandwidth for transmitting an encoded version of the  
25 multichannel audio signal to a decoder side, based on the audio content of the channels of the multichannel audio signal and/or based on an input signal indicating a desired coding format.

In an example embodiment, the plurality of encoding sections may include a single-channel encoding section operable to independently encode no more than a  
30 single audio channel in a downmix channel, and at least one of the first and second subsets of the encoding sections may comprise the single-channel encoding section.

According to example embodiments, there is provided a computer program product comprising a computer-readable medium with instructions for performing any of the methods of the first and second aspects.

According to example embodiments, it may hold that  $N = 3$  or  $N = 4$  in any of the methods, encoding systems, decoding systems and computer program products of the first and second aspects.

Further example embodiments are defined in the dependent claims. It is noted  
5 that example embodiments include all combinations of features, even if recited in mutually different claims.

## II. Example embodiments

On an encoder side, which will be described with reference to Figs. 3 and 4, a  
10 single-channel downmix signal  $Y$  is computed as a linear mapping of an  $N$ -channel audio signal  $X = [x_1 \dots x_N]^T$  according to

$$Y = [d_1 \dots d_N] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \sum_{n=1}^N d_n x_n = DX, \quad (1)$$

15 where  $d_n$ ,  $n = 1, \dots, N$ , are downmix coefficients represented by a downmix matrix  $D$ . On a decoder side, which will be described with reference to Figs. 1 and 2, parametric reconstruction of the  $N$ -channel audio signal  $X$  is performed according to

$$\hat{X} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} Y + \begin{bmatrix} p_{11} & \dots & p_{1,N-1} \\ p_{21} & \dots & p_{2,N-1} \\ \vdots & \ddots & \vdots \\ p_{N,1} & \dots & p_{N,N-1} \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_{N-1} \end{bmatrix} = CY + PZ, \quad (2)$$

20

where  $c_n$ ,  $n = 1, \dots, N$ , are dry upmix coefficients represented by a matrix dry upmix matrix  $C$ ,  $p_{n,k}$ ,  $n = 1, \dots, N$ ,  $k = 1, \dots, N - 1$ , are wet upmix coefficients represented by a wet upmix matrix  $P$ , and  $z_k$ ,  $k = 1, \dots, N - 1$  are the channels of an  $(N - 1)$ -channel decorrelated signal  $Z$  generated based on the downmix signal  $Y$ . If the channels of  
25 each audio signal are represented as rows, the covariance matrix of the original audio signal  $X$  may be expressed as  $R = XX^T$ , and the covariance matrix of the audio signal as reconstructed  $\hat{X}$  may be expressed as  $\hat{R} = \hat{X}\hat{X}^T$ . It is to be noted that if for example the audio signals are represented as rows comprising complex-valued

transform coefficients, the real part of  $XX^*$ , where  $X^*$  is the complex conjugate transpose of the matrix  $X$ , may for example be considered instead of  $XX^T$ .

In order to provide a faithful reconstruction of the original audio signal  $X$ , it may be advantageous for the reconstruction given by equation (2) to reinstate full  
 5 covariance, i.e., it may be advantageous to employ dry and wet upmix matrices  $C$  and  $P$  such that

$$R = \widehat{R}. \quad (3)$$

10 One approach is to first find a dry upmix matrix  $C$  giving the best possible “dry” upmix  $\widehat{X}_0 = CY$  in the least squares sense, by solving the normal equations

$$CYY^T = XY^T. \quad (4)$$

15 For  $\widehat{X}_0 = CY$ , with a matrix  $C$  solving equation (4), it holds that

$$R = \widehat{X}_0 \widehat{X}_0^T + (\widehat{X}_0 - X)(\widehat{X}_0 - X)^T = R_0 + \Delta R. \quad (5)$$

Assuming that the channels of the decorrelated signal  $Z$  are mutually uncorrelated  
 20 and all have the same energy  $\|Y\|^2$  equal to that of the single-channel downmix signal  $Y$ , the positive definite missing covariance  $\Delta R$  can be factorized according to

$$\Delta R = PP^T \|Y\|^2. \quad (6)$$

25 Full covariance may be reinstated according to equation (3) by employing a dry upmix matrix  $C$  solving equation (4) and a wet upmix matrix  $P$  solving equation (6). Equations (1) and (4) imply that  $DCYY^T = YY^T$ , and thereby that

$$\sum_{n=1}^N d_n c_n = DC = 1, \quad (7)$$

30

for non-degenerate downmix matrices  $D$ . Equations (5) and (7) imply that  $D(X_0 - X) = DCY - Y = 0$  and

$$D\Delta R = 0. \quad (8)$$

Hence, the missing covariance  $\Delta R$  has rank  $N - 1$ , and may indeed be provided by employing a decorrelated signal  $Z$  with  $N - 1$  mutually uncorrelated channels.

- 5 Equation (6) and (8) imply that  $DP = 0$ , so that the columns of the wet upmix matrix  $P$  solving equation (6) can be constructed from vectors spanning the kernel space of the downmix matrix  $D$ . The computations for finding a suitable wet upmix matrix  $P$  may therefore be moved to that lower-dimensional space.

Let  $V$  be a matrix of size  $N(N - 1)$  containing an orthonormal basis for the  
 10 kernel space of the downmix matrix  $D$ , i.e. a linear space of vectors  $v$  with  $Dv = 0$ . Examples of such predefined matrixes  $V$  for  $N = 2$ ,  $N = 3$ , and  $N = 4$ , respectively, are

$$\frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} \\ 0 & -2/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} \text{ and } \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}. \quad (9)$$

15

In the basis given by  $V$ , the missing covariance can be expressed as  $R_v = V^T(\Delta R)V$ . To find a wet upmix matrix  $P$  solving equation (6) one may therefore first find a matrix  $H$  by solving  $R_v = HH^T$ , and then obtain  $P$  as  $P = VH/\|Y\|$ , where  $\|Y\|$  is the square root of the energy of the single-channel downmix signal  $Y$ . Other suitable upmix  
 20 matrices  $P$  may be obtained as  $P = VHO/\|Y\|$ , where  $O$  is an orthogonal matrix. Alternatively, one may rescale the missing covariance  $R_v$  by the energy  $\|Y\|^2$  of the single-channel downmix signal  $Y$  and instead solve the equation

$$\frac{R_v}{\|Y\|^2} = H_R H_R^T, \quad (10)$$

25

where  $H = H_R\|Y\|$ , and obtain  $P$  as

$$P = VH_R. \quad (11)$$

When the entries of  $H_R$  are quantized and the desired output has a silent channel, the properties of the predefined matrix  $V$  as stated above may be inconvenient. As an example, for  $N = 3$ , a better choice for the second matrix of (9) would be

$$5 \quad \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} \\ -1/\sqrt{2} & 0 \end{bmatrix}. \quad (12)$$

Fortunately, the requirement that the columns of the matrix  $V$  are pairwise orthogonal can be dropped as long as these columns are linearly independent. The desired solution  $R_v$  to  $\Delta R = VR_vV^T$  is then obtained by  $R_v = W^T(\Delta R)W$  with  $W = V(V^TV)^{-1}$ , the pseudoinverse of  $V$ .

The matrix  $R_v$  is a positive semi-definite matrix of size  $(N - 1)^2$  and there are several approaches to finding solutions to equation (10), leading to solutions within respective matrix classes of dimension  $N(N - 1)/2$ , i.e. in which the matrices are uniquely defined by  $N(N - 1)/2$  matrix elements. Solutions may for example be obtained by employing:

- a. Cholesky factorization, leading to a lower triangular  $H_R$ ;
- b. positive square root, leading to a symmetric positive semi-definite  $H_R$ ; or
- 20 c. polar, leading to  $H_N$  of the form  $H_R = O\Lambda$ , where  $O$  is orthogonal and  $\Lambda$  is diagonal.

Moreover, there are normalized version of the options a) and b) in which  $H_R$  may be expressed as  $H_R = \Lambda H_0$ , where  $\Lambda$  is diagonal and  $H_0$  has all diagonal elements equal to one. The alternatives a, b and c, above, provide solutions  $H_R$  in different matrix classes, i.e. lower triangular matrices, symmetric matrices and products of diagonal and orthogonal matrices. If the matrix class to which  $H_R$  belongs is known at a decoder side, i.e. if it is known that  $H_R$  belongs to a predefined matrix class, e.g. according to any the above alternatives a, b and c,  $H_R$  may be populated based on only  $N(N - 1)/2$  of its elements. If also the matrix  $V$  is known at the decoder side, e.g. if it is known that  $V$  is one of the matrices given in (9), the wet upmix matrix  $P$ ,

needed for reconstruction according to equation (2), may then be obtained via equation (11).

Fig. 3 is a generalized block diagram of a parametric encoding section 300 according to an example embodiment. The parametric encoding section 300 is configured to encode an  $N$ -channel audio signal  $X$  as a single-channel downmix signal  $Y$  and metadata suitable for parametric reconstruction of the audio signal  $X$  according to equation (2). The parametric encoding section 300 comprises a downmix section 301, which receives the audio signal  $X$  and computes, according to a predefined rule, the single-channel downmix signal  $Y$  as a linear mapping of the audio signal  $X$ . In the present example embodiment, the downmix section 301 computes the downmix signal  $Y$  according to equation (1), wherein the downmix matrix  $D$  is predefined and corresponds to the predefined rule. A first analyzing section 302 determines a set of dry upmix coefficients, represented by the dry upmix matrix  $C$ , in order to define a linear mapping of the downmix signal  $Y$  approximating the audio signal  $X$ . This linear mapping of the downmix signal  $Y$  is denoted by  $CY$  in equation (2). In the present example embodiment,  $N$  dry upmix coefficients  $C$  are determined according to equation (4) such that the linear mapping  $CY$  of the downmix signal  $Y$  corresponds to a minimum mean square approximation of the audio signal  $X$ . A second analyzing section 303 determines an intermediate matrix  $H_R$  based on a difference between the covariance matrix of the audio signal  $X$  as received and the covariance matrix of the audio signal as approximated by the linear mapping  $CY$  of the downmix signal  $Y$ . In the present example embodiment, the covariance matrices are computed by first and second processing sections 304, 305, respectively, and are then provided to the second analyzing section 303. In the present example embodiment, the intermediate matrix  $H_R$  is determined according to above described approach b to solving equation (10), leading to an intermediate matrix  $H_R$  which is symmetric. As indicated in equations (1) and (11), the intermediate matrix  $H_R$ , when multiplied by a predefined matrix  $V$ , defines, via a set of wet upmix parameters  $P$ , a linear mapping  $PZ$  of a decorrelated signal  $Z$  as part of parametric reconstruction of the audio signal  $X$  at a decoder side. In the present example embodiment, the intermediate matrix  $V$  is the second matrix in (9) for the case  $N = 3$ , and the third matrix in (9) for the case  $N = 4$ . The parametric encoding section 300 outputs the downmix signal  $Y$  together with dry upmix parameters  $\tilde{C}$  and wet upmix parameters

$\tilde{P}$ . In the present example embodiment,  $N - 1$  of the  $N$  dry upmix coefficients  $C$  are the dry upmix parameters  $\tilde{C}$ , and the remaining one dry upmix coefficient is derivable from the dry upmix parameters  $\tilde{C}$  via equation (7) if the predefined downmix matrix  $D$  is known. Since the intermediate matrix  $H_R$  belongs to the class of symmetric

5 matrices, it is uniquely defined by  $N(N - 1)/2$  of its  $(N - 1)^2$  elements. In the present example embodiment,  $N(N - 1)/2$  of the elements of the intermediate matrix  $H_R$  are therefore wet upmix parameters  $\tilde{P}$  from which the rest of the intermediate matrix  $H_R$  is derivable knowing that it is symmetric.

Fig. 4 is a generalized block diagram of an audio encoding system 400

10 according to an example embodiment, comprising the parametric encoding section 300 described with reference to Fig. 3. In the present example embodiment, audio content, e.g. recorded by one or more acoustic transducers 401, or generated by audio authoring equipment 401, is provided in the form of the  $N$ -channel audio signal  $X$ . A quadrature mirror filter (QMF) analysis section 402 transforms the audio signal

15  $X$ , time segment by time segment, into a QMF domain for processing by the parametric encoding section 300 of the audio signal  $X$  in the form of time/frequency tiles. The downmix signal  $Y$  output by the parametric encoding section 300 is transformed back from the QMF domain by a QMF synthesis section 403 and is transformed into a modified discrete cosine transform (MDCT) domain by a transform

20 section 404. Quantization sections 405 and 406 quantize the dry upmix parameters  $\tilde{C}$  and wet upmix parameters  $\tilde{P}$ , respectively. For example, uniform quantization with a step size of 0.1 or 0.2 (dimensionless) may be employed, followed by entropy coding in the form of Huffman coding. A coarser quantization with step size 0.2 may for example be employed to save transmission bandwidth, and a finer quantization with

25 step size 0.1 may for example be employed to improve fidelity of the reconstruction on a decoder side. The MDCT-transformed downmix signal  $Y$  and the quantized dry upmix parameters  $\tilde{C}$  and wet upmix parameters  $\tilde{P}$  are then combined into a bitstream  $B$  by a multiplexer 407, for transmission to a decoder side. The audio encoding system 400 may also comprise a core encoder (not shown in Fig. 4) configured to

30 encode the downmix signal  $Y$  using a perceptual audio codec, such as Dolby Digital or MPEG AAC, before the downmix signal  $Y$  is provided to the multiplexer 407.

Fig. 1 is a generalized block diagram of a parametric reconstruction section 100, according to an example embodiment, configured to reconstruct the  $N$ -channel

audio signal  $X$  based on a single-channel downmix signal  $Y$  and associated dry upmix parameters  $\tilde{C}$  and wet upmix parameters  $\tilde{P}$ . The parametric reconstruction section 100 is adapted to perform reconstruction according to equation (2), i.e. using dry upmix parameters  $C$  and wet upmix parameters  $P$ . However, instead of receiving the dry upmix parameters  $C$  and the wet upmix parameters  $P$  themselves, dry upmix parameters  $\tilde{C}$  and wet upmix parameters  $\tilde{P}$  are received from which the dry upmix parameters  $C$  and wet upmix parameters  $P$  are derivable. A decorrelating section 101 receives the downmix signal  $Y$  and outputs, based thereon, a  $(N - 1)$ -channel decorrelated signal  $Z = [z_1 \cdots z_{N-1}]^T$ . In the present example embodiment, the channels of the decorrelated signal  $Z$  are derived by processing the downmix signal  $Y$ , including applying respective all-pass filters to the downmix signal  $Y$ , so as to provide channels that are uncorrelated to the downmix signal  $Y$ , and with audio content which is spectrally similar to and also perceived as similar to that of the downmix signal  $Y$  by a listener. The  $(N - 1)$ -channel decorrelated signal  $Z$  serves to increase the dimensionality of the reconstructed version  $\hat{X}$  of  $N$ -channel audio signal  $X$ , as perceived by a listener. In the present example embodiment, the channels of the decorrelated signal  $Z$  have at least approximately the same spectra as that of the single-channel downmix signal  $Y$  and form, together with the single-channel downmix signal  $Y$ ,  $N$  at least approximately mutually uncorrelated channels. A dry upmix section 102 receives the dry upmix parameters  $\tilde{C}$  and the downmix signal  $Y$ . In the present example embodiment, the dry upmix parameters  $\tilde{C}$  coincide with the first  $N - 1$  of the  $N$  dry upmix coefficients  $C$ , and the remaining dry upmix coefficient is determined based on a predefined relation between the dry upmix coefficients  $C$  given by equation (7). The dry upmix section 102 outputs a dry upmix signal computed by mapping the downmix signal  $Y$  linearly in accordance with the set of dry upmix coefficients  $C$ , and denoted by  $CY$  in equation (2). A wet upmix section 103 receives the wet upmix parameters  $\tilde{P}$  and the decorrelated signal  $Z$ . In the present example embodiment, the wet upmix parameters  $\tilde{P}$  are  $N(N - 1)/2$  elements of the intermediate matrix  $H_R$  determined at the encoder side according to equation (10). In the present example embodiment, the wet upmix section 103 populates the remaining elements of the intermediate matrix  $H_R$  knowing that the intermediate matrix  $H_R$  belongs to a predefined matrix class, i.e. that it is symmetric, and exploiting the corresponding relationships between the elements of the matrix. The wet upmix

section 103 then obtains a set of wet upmix coefficients  $P$  by employing equation (11), i.e. by multiplying the intermediate matrix  $H_R$  by the predefined matrix  $V$ , i.e. the second matrix in (9) for the case  $N = 3$ , and the third matrix in (9) for the case  $N = 4$ . Hence, the  $N(N - 1)$  wet upmix coefficients  $P$  are derived from the received  $N(N -$   
5  $1)/2$  independently assignable wet upmix parameters  $\tilde{P}$ . The wet upmix section 103 outputs a wet upmix signal computed by mapping the decorrelated signal  $Z$  linearly in accordance with the set of wet upmix coefficients  $P$ , and denoted by  $PZ$  in equation (2). A combining section 104 receives the dry upmix signal  $CY$  and the wet upmix signal  $PZ$  and combines these signals to obtain a first multidimensional reconstructed  
10 signal  $\hat{X}$  corresponding to the  $N$ -channel audio signal  $X$  to be reconstructed. In the present example embodiment, the combining section 104 obtains the respective channels of the reconstructed signal  $\hat{X}$  by combining the audio content of the respective channels of the dry upmix signal  $CY$  with the respective channels of the wet upmix signal  $PZ$ , according to equation (2).

15 Fig. 2 is a generalized block diagram of an audio decoding system 200 according to an example embodiment. The audio decoding system 200 comprises the parametric reconstruction section 100 described with reference to Fig. 1. A receiving section 201, e.g. including a demultiplexer, receives the bitstream  $B$  transmitted from the audio encoding system 400 described with reference to Fig. 4,  
20 and extracts the downmix signal  $Y$  and the associated dry upmix parameters  $\tilde{C}$  and wet upmix parameters  $\tilde{P}$  from the bitstream  $B$ . In case the downmix signal  $Y$  is encoded in the bitstream  $B$  using a perceptual audio codec such as Dolby Digital or MPEG AAC, the audio decoding system 200 may comprise a core decoder (not shown in Fig. 2) configured to decode the downmix signal  $Y$  when extracted from the  
25 bitstream  $B$ . A transform section 202 transforms the downmix signal  $Y$  by performing inverse MDCT and a QMF analysis section 203 transforms the downmix signal  $Y$  into a QMF domain for processing by the parametric reconstruction section 100 of the downmix signal  $Y$  in the form of time/frequency tiles. Dequantization sections 204 and 205 dequantize the dry upmix parameters  $\tilde{C}$  and wet upmix parameters  $\tilde{P}$ , e.g.,  
30 from an entropy coded format, before supplying them to the parametric reconstruction section 100. As described with reference to Fig. 4, quantization may have been performed with one of two different step sizes, e.g. 0.1 or 0.2. The actual step size employed may be predefined, or may be signaled to the audio decoding

system 200 from the encoder side, e.g. via the bitstream  $B$ . In some example embodiments, the dry upmix coefficients  $C$  and the wet upmix coefficients  $P$  may be derived from the dry upmix parameters  $\tilde{C}$  and wet upmix parameters  $\tilde{P}$ , respectively, already in the respective dequantization sections 204 and 205, which may optionally  
5 be regarded as part of the dry upmix section 102 and the wet upmix section 103, respectively. In the present example embodiment, the reconstructed audio signal  $\hat{X}$  output by the parametric reconstruction section 100 is transformed back from the QMF domain by a QMF synthesis section 206 before being provided as output of the audio decoding system 200 for playback on a multispeaker system 207.

10 Figs. 5–11 illustrate alternative ways to represent an 11.1 channel audio signal by means of downmix channels, according to example embodiments. In the present example embodiments, the 11.1 channel audio signal comprises the channels: left (L), right (R), center (C), low-frequency effects (LFE), left side (LS), right side (RS), left back (LB), right back (RB), top front left (TFL), top front right (TFR), top back left  
15 (TBL) and top back right (TBR), which are indicated in Figures 5–11 by uppercase letters. The alternative ways to represent the 11.1 channel audio signal correspond to alternative partitions of the channels into sets of channels, each set being represented by a single downmix signal, and optionally by associated wet and dry upmix parameters. Encoding of each of the sets of channels into its respective single-  
20 channel downmix signal (and metadata) may be performed independently and in parallel. Similarly, reconstruction of the respective sets of channels from their respective single-channel downmix signals may be performed independently and in parallel.

It is to be understood that, in the example embodiments described with  
25 reference to Figs. 5–11 (and also below with reference to Figs. 13–16), none of the reconstructed channels may comprise contributions from more than one downmix channel and any decorrelated signals derived from that single downmix signal, i.e. contributions from multiple downmix channels are not combined/mixed during parametric reconstruction.

30 In Fig. 5, the channels LS, TBL and LB form a group 501 of channels represented by the single downmix channel  $ls$  (and its associated metadata). The parametric encoding section 300 described with reference to Fig. 3 may be employed with  $N = 3$  to represent the three audio channels LS, TBL and LB by the single

downmix channels and associated dry and wet upmix parameters. Given that a predefined matrix  $V$  and predefined matrix class of an intermediate matrix  $H_R$ , both associated with the encoding performed in the parametric encoding section 300, are known on a decoder side, the parametric reconstruction section 100, described with reference to Fig. 1, may be employed to reconstruct the three channels LS, TBL and LB from the downmix signal ls and the associated dry and wet upmix parameters. Similarly, the channels RS, TBR and RB form a group 502 of channels represented by the single downmix channel rs, and another instance of the parametric encoding section 300 may be employed in parallel with the first encoding section to represent the three channels RS, TBR and RB by the single downmix channel rs and associated dry and wet upmix parameters. Moreover, given that a predefined matrix  $V$  and a predefined matrix class to which an intermediate matrix  $H_R$  belongs, both associated with the second instance of the parametric encoding section 300, are known at a decoder side, another instance of the parametric reconstruction section 100 may be employed in parallel with the first parametric reconstruction section to reconstruct the three channels RS, TBR and RB from the downmix signal rs and the associated dry and wet upmix parameters. Another group 503 of channels includes only two channels L and TFL represented by a downmix channel l. Encoding of these two channels into the downmix channel l and associated wet and dry upmix parameters may be performed by encoding sections and reconstruction section analogous to those described with reference to Figs. 3 and 1, respectively, but for  $N = 2$ . Another group 504 of channels comprises only a single channel LFE represented by a downmix channel lfe. In this case, no downmixing is required and the downmix channel lfe may be the channel LFE itself, optionally transformed into an MDCT domain and/or encoded using a perceptual audio codec.

The total number of downmix channels employed in Figs. 5–11 to represent the 11.1 channel audio signal varies. For example, the example illustrated in Fig. 5 employs 6 downmix channels while the example in Fig. 7 employs 10 downmix channels. Different downmix configurations may be suitable for different situations, e.g. depending on available bandwidth for transmission of the downmix signals and associated upmix parameter, and/or requirements on how faithful the reconstruction of the 11.1 channel audio signal should be.

According to example embodiments, the audio encoding system 400 described with reference to Fig. 4 may comprise a plurality of parametric encoding sections, including the parametric encoding section 300 described with reference to Fig. 3. The audio encoding system 400 may comprise a control section (not shown in Fig. 4) configured to determine/select a coding format for the 11.1-channel audio signal, from a collection for coding formats corresponding to the respective partitions of the 11.1 channel audio signal illustrated in Figs. 5–11. The coding format further corresponds to a set of predefined rules (at least some of which may coincide) for computing the respective downmix channels, a set of predefined matrix classes (at least some of which may coincide) for intermediate matrices  $H_R$  and a set of predefined matrices  $V$  (at least some of which may coincide) for obtaining wet upmix coefficients associated with at least some of the respective sets of channels based on respective associated wet upmix parameters. According to the present example embodiments, the audio encoding system is configured to encode the 11.1 channel audio signal using a subset of the plurality of encoding sections appropriate to the determined coding format. If, for example, the determined coding format corresponds to the partition of the 11.1 channels illustrated in Fig. 1, the encoding system may employ 2 encoding sections configured for representing respective sets of 3 channels by respective single downmix channels, 2 encoding sections configured for representing respective sets of 2 channels by respective single downmix channels, and 2 encoding sections configured for representing respective single channel as respective single downmix channels. All the downmix signals and the associated wet and dry upmix parameters may be encoded in the same bitstream  $B$ , for transmittal to a decoder side. It is to be noted that the compact format of the metadata accompanying the downmix channels, i.e. the wet upmix parameters and the wet upmix parameters, may be employed by some of the encoding sections, while in at least some example embodiments, other metadata formats may be employed. For example, some of the encoding sections may output the full number of the wet and dry upmix coefficients instead of the wet and dry upmix parameters. Embodiments are also envisaged in which some channels are encoded for reconstruction employing fewer than  $N - 1$  decorrelated channels (or even no decorrelation at all), and where metadata for parametric reconstruction may therefore take a different form.

According to example embodiments, the audio decoding system 200 described with reference to Fig. 2 may comprise a corresponding plurality of reconstruction sections, including the parametric reconstruction section 100 described with reference to Fig. 1, for reconstructing the respective sets of channels  
5 of the 11.1 channel audio signal represented by the respective downmix signals. The audio decoding system 200 may comprise a control section (not shown in Fig. 2) configured to receive signaling from the encoder side indicating the determined coding format, and the audio decoding system 200 may employ an appropriate subset of the plurality of reconstruction sections for reconstructing the 11.1 channel  
10 audio signal from the received downmix signals and associated dry and wet upmix parameters.

Figs. 12–13 illustrate alternative ways to represent a 13.1 channel audio signal by means of downmix channels, according to example embodiments. The 13.1 channel audio signal includes the channels: left screen (LSCRN), left wide (LW), right  
15 screen (RSCRN), right wide (RW), center (C), low-frequency effects (LFE), left side (LS), right side (RS), left back (LB), right back (RB), top front left (TFL), top front right (TFR), top back left (TBL) and top back right (TBR). Encoding of the respective groups of channels as the respective downmix channels may be performed by respective encoding sections operating independently in parallel, as described above  
20 with reference to Figs. 5–11. Similarly, reconstruction of the respective groups of channels based on the respective downmix channels and associated upmix parameters may be performed by respective reconstruction sections operating independently in parallel.

Figs. 14–16 illustrate alternative ways to represent a 22.2 channel audio signal  
25 by means of downmix signals, according to example embodiments. The 22.2 channel audio signal includes the channels: low-frequency effects 1 (LFE1), low-frequency effects 2 (LFE2), bottom front center (BFC), center (C), top front center (TFC), left wide (LW), bottom front left (BFL), left (L), top front left (TFL), top side left (TSL), top back left (TBL), left side (LS), left back (LB), top center (TC), top back  
30 center (TBC), center back (CB), bottom front right (BFR), right (R), right wide (RW), top front right (TFR), top side right (TSR), top back right (TBR), right side (RS), and right back (RB). The partition of the 22.2 channel audio signal illustrated in Fig. 16 includes a group 1601 of channels including four channels. The parametric encoding section 300 described with reference to Fig. 3, but implemented with  $N = 4$ , may be

employed to encode these channels as a downmix signal and associated wet and dry upmix parameters. Analogously, the parametric reconstruction section 100 described with reference to Fig. 1, but implemented with  $N = 4$ , may be employed to reconstruct these channels from the downmix signal and associated wet and dry upmix parameters.

### III. Equivalents, extensions, alternatives and miscellaneous

Further embodiments of the present disclosure will become apparent to a person skilled in the art after studying the description above. Even though the present description and drawings disclose embodiments and examples, the disclosure is not restricted to these specific examples. Numerous modifications and variations can be made without departing from the scope of the present disclosure, which is defined by the accompanying claims. Any reference signs appearing in the claims are not to be understood as limiting their scope.

Additionally, variations to the disclosed embodiments can be understood and effected by the skilled person in practicing the disclosure, from a study of the drawings, the disclosure, and the appended claims. In the claims, the word "comprising" does not exclude other elements or steps, and the indefinite article "a" or "an" does not exclude a plurality. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage.

The devices and methods disclosed hereinabove may be implemented as software, firmware, hardware or a combination thereof. In a hardware implementation, the division of tasks between functional units referred to in the above description does not necessarily correspond to the division into physical units; to the contrary, one physical component may have multiple functionalities, and one task may be carried out by several physical components in cooperation. Certain components or all components may be implemented as software executed by a digital signal processor or microprocessor, or be implemented as hardware or as an application-specific integrated circuit. Such software may be distributed on computer readable media, which may comprise computer storage media (or non-transitory media) and communication media (or transitory media). As is well known to a person skilled in the art, the term computer storage media includes both volatile and nonvolatile, removable and non-removable media implemented in any method or

technology for storage of information such as computer readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by a computer. Further, it is well known to the skilled person that communication media typically embodies computer readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media.

## CLAIMS

1. A method for reconstructing an  $N$ -channel audio signal ( $X$ ), wherein  $N \geq 3$ , the method comprising:
  - 5 receiving a single-channel downmix signal ( $Y$ ) together with associated dry and wet upmix parameters ( $\tilde{C}, \tilde{P}$ );  
computing a dry upmix signal as a linear mapping of the downmix signal, wherein a set of dry upmix coefficients ( $C$ ) is applied to the downmix signal;  
generating an  $(N - 1)$ -channel decorrelated signal ( $Z$ ) based on the downmix  
10 signal;  
computing a wet upmix signal as a linear mapping of the decorrelated signal, wherein a set of wet upmix coefficients ( $P$ ) is applied to the channels of the decorrelated signal; and  
combining the dry and wet upmix signals to obtain a multidimensional  
15 reconstructed signal ( $\hat{X}$ ) corresponding to the  $N$ -channel audio signal to be reconstructed,  
wherein the method further comprises:  
determining the set of dry upmix coefficients based on the received dry upmix parameters;  
20 populating an intermediate matrix having more elements than the number of received wet upmix parameters, based on the received wet upmix parameters and knowing that the intermediate matrix belongs to a predefined matrix class; and  
obtaining the set of wet upmix coefficients by multiplying the intermediate  
matrix by a predefined matrix, wherein the set of wet upmix coefficients corresponds  
25 to the matrix resulting from the multiplication and includes more coefficients than the number of elements in the intermediate matrix.
2. The method of claim 1, wherein receiving the wet upmix parameters includes receiving  $N(N - 1)/2$  wet upmix parameters, wherein populating the intermediate  
30 matrix includes obtaining values for  $(N - 1)^2$  matrix elements based on the received  $N(N - 1)/2$  wet upmix parameters and knowing that the intermediate matrix belongs to the predefined matrix class, wherein the predefined matrix includes  $N(N - 1)$

elements, and wherein the set of wet upmix coefficients includes  $N(N - 1)$  coefficients.

3. The method of claim 1 or 2, wherein populating the intermediate matrix  
5 includes employing the received wet upmix parameters as elements in the intermediate matrix.

4. The method of any of the preceding claims, wherein receiving the dry upmix  
10 parameters includes receiving  $(N - 1)$  dry upmix parameters, wherein the set of dry upmix coefficients includes  $N$  coefficients, and wherein the set of dry upmix coefficients is determined based on the received  $(N - 1)$  dry upmix parameters and based on a predefined relation between the coefficients in the set of dry upmix coefficients.

15 5. The method of any of the preceding claims, wherein the predefined matrix class is one of:

lower or upper triangular matrices, wherein known properties of all matrices in the class include predefined matrix elements being zero;

20 symmetric matrices, wherein known properties of all matrices in the class include predefined matrix elements being equal; and

products of an orthogonal matrix and a diagonal matrix, wherein known properties of all matrices in the class include known relations between predefined matrix elements.

25 6. The method of any of the preceding claims, wherein the downmix signal is obtainable, according to a predefined rule, as a linear mapping of the  $N$ -channel audio signal to be reconstructed, wherein the predefined rule defines a predefined downmix operation, and wherein said predefined matrix is based on vectors spanning the kernel space of said predefined downmix operation.

30

7: The method of any of the preceding claims, wherein receiving the single-channel downmix signal together with associated dry and wet upmix parameters includes receiving a time segment or time/frequency tile of the downmix signal

together with associated dry and wet upmix parameters, and wherein said multidimensional reconstructed signal corresponds to a time segment or time/frequency tile of the  $N$ -channel audio signal to be reconstructed.

- 5 8. An audio decoding system (200) comprising a first parametric reconstruction section (100) configured to reconstruct an  $N$ -channel audio signal ( $X$ ) based on a first single-channel downmix signal ( $Y$ ) and associated dry and wet upmix parameters ( $\tilde{C}, \tilde{P}$ ), wherein  $N \geq 3$ , the first parametric reconstruction section comprising:
- 10 a first decorrelating section (101) configured to receive the first downmix signal and to output, based thereon, a first  $(N - 1)$ -channel decorrelated signal ( $Z$ );
- a first dry upmix section (102) configured to
- receive the dry upmix parameters ( $\tilde{C}$ ) and the downmix signal, determine a first set of dry upmix coefficients ( $C$ ) based on the dry upmix parameters, and
- 15 output a first dry upmix signal computed by mapping the first downmix signal linearly in accordance with the first set of dry upmix coefficients;
- a first wet upmix section (103) configured to
- receive the wet upmix parameters ( $\tilde{P}$ ) and the first decorrelated signal, populate a first intermediate matrix having more elements than the
- 20 number of received wet upmix parameters, based on the received wet upmix parameters and knowing that the first intermediate matrix belongs to a first predefined matrix class,
- obtain a first set of wet upmix coefficients ( $P$ ) by multiplying the first intermediate matrix by a first predefined matrix, wherein the first set of wet upmix coefficients corresponds to the matrix resulting from the multiplication and includes more coefficients than the number of elements in the first
- 25 intermediate matrix, and
- output a first wet upmix signal computed by mapping the first decorrelated signal linearly in accordance with the first set of wet upmix
- 30 coefficients; and
- a first combining section (104) configured to receive the first dry upmix signal and the first wet upmix signal and to combine these signals to obtain a first

multidimensional reconstructed signal ( $\hat{X}$ ) corresponding to the  $N$ -channel audio signal to be reconstructed.

9. The audio decoding system of claim 8, further comprising a second parametric reconstruction section operable independently of the first parametric reconstruction section and configured to reconstruct an  $N_2$ -channel audio signal based on a second single-channel downmix signal and associated dry and wet upmix parameters, wherein  $N_2 \geq 2$ , the second parametric reconstruction section comprising a second decorrelating section, a second dry upmix section, a second wet upmix section and a second combining section, said sections of the second parametric reconstruction section being configured analogously to the corresponding sections of the first parametric reconstruction section, wherein the second wet upmix section is configured to employ a second intermediate matrix belonging to a second predefined matrix class and a second predefined matrix.

15

10. The audio decoding system of claim 8 or 9, wherein the audio decoding system is adapted to reconstruct a multichannel audio signal based on a plurality of downmix channels and associated dry and wet upmix parameters, wherein the audio decoding system comprises:

20 a plurality of reconstruction sections, including parametric reconstruction sections operable to independently reconstruct respective sets of audio signal channels based on respective downmix channels and respective associated dry and wet upmix parameters; and

a control section configured to receive signaling indicating a coding format of the multichannel audio signal corresponding to a partition of the channels of the multichannel audio signal into sets (501–504) of channels represented by the respective downmix channels and, for at least some of the downmix channels, by respective associated dry and wet upmix parameters, the coding format further corresponding to a set of predefined matrices for obtaining wet upmix coefficients associated with at least some of the respective sets of channels based on the respective associated wet upmix parameters,

30 wherein the decoding system is configured to reconstruct the multichannel audio signal using a first subset of the plurality of reconstruction sections, in response

to the received signaling indicating a first coding format, wherein the decoding system is configured to reconstruct the multichannel audio signal using a second subset of the plurality of reconstruction sections, in response to the received signaling indicating a second coding format, and wherein at least one of the first and second subsets of the reconstruction sections comprises said first parametric reconstruction section.

11. The audio decoding system of claim 10, wherein the plurality of reconstruction sections includes a single-channel reconstruction section operable to independently reconstruct a single audio channel based on a downmix channel in which no more than a single audio channel has been encoded, and wherein at least one of the first and second subsets of the reconstruction sections comprises the single-channel reconstruction section.

12. The audio decoding system of claim 10 or 11, wherein the first coding format corresponds to reconstruction of said multichannel audio signal from a lower number of downmix channels than the second coding format.

13. A method for encoding an  $N$ -channel audio signal ( $X$ ) as a single-channel downmix signal ( $Y$ ) and metadata suitable for parametric reconstruction of said audio signal from the downmix signal and an  $(N - 1)$ -channel decorrelated signal ( $Z$ ) determined based on the downmix signal, wherein  $N \geq 3$ , the method comprising:

- receiving said audio signal;
- computing, according to a predefined rule, the single-channel downmix signal as a linear mapping of said audio signal;
- determining a set of dry upmix coefficients ( $C$ ) in order to define a linear mapping of the downmix signal approximating said audio signal;
- determining an intermediate matrix based on a difference between a covariance of said audio signal as received and a covariance of said audio signal as approximated by the linear mapping of the downmix signal, wherein the intermediate matrix when multiplied by a predefined matrix corresponds to a set of wet upmix coefficients ( $P$ ) defining a linear mapping of said decorrelated signal as part of parametric reconstruction of said audio signal, wherein the set of wet upmix

coefficients includes more coefficients than the number of elements in the intermediate matrix; and

outputting the downmix signal together with dry upmix parameters ( $\tilde{C}$ ), from which the set of dry upmix coefficients is derivable, and wet upmix parameters ( $\tilde{P}$ ),  
5 wherein the intermediate matrix has more elements than the number of output wet upmix parameters, and wherein the intermediate matrix is uniquely defined by the output wet upmix parameters provided that the intermediate matrix belongs to a predefined matrix class.

10 14. The method of claim 13, wherein determining the intermediate matrix includes determining the intermediate matrix such that a covariance of the signal obtained by the linear mapping of said decorrelated signal, defined by the set of wet upmix coefficients, approximates the difference between the covariance of said audio signal as received and the covariance of said audio signal as approximated by the linear  
15 mapping of the downmix signal.

15. The method of claim 13 or 14, wherein outputting the wet upmix parameters includes outputting no more than  $N(N - 1)/2$  wet upmix parameters, wherein the intermediate matrix has  $(N - 1)^2$  matrix elements and is uniquely defined by the  
20 output wet upmix parameters provided that the intermediate matrix belongs to the predefined matrix class, and wherein the set of wet upmix coefficients includes  $N(N - 1)$  coefficients.

16. The method of any of claims 13 to 15, wherein the set of dry upmix coefficients  
25 includes  $N$  coefficients, and wherein outputting the dry upmix parameters includes outputting no more than  $N - 1$  dry upmix parameters, the set of dry upmix coefficients being derivable from the  $N - 1$  dry upmix parameters using said predefined rule.

30 17. The method of any of claims 13 to 16, wherein the determined set of dry upmix coefficients defines a linear mapping of the downmix signal corresponding to a minimum mean square error approximation of said audio signal.

18. An audio encoding system (400) comprising a parametric encoding section (300) configured to encode an  $N$ -channel audio signal ( $X$ ) as a single-channel downmix signal ( $Y$ ) and metadata suitable for parametric reconstruction of said audio signal from the downmix signal and an  $(N - 1)$ -channel decorrelated signal ( $Z$ )  
5 determined based on the downmix signal, wherein  $N \geq 3$ , the parametric encoding section comprising:

a downmix section (301) configured to receive said audio signal and to compute, according to a predefined rule, the single-channel downmix signal as a linear mapping of said audio signal;

10 a first analyzing section (302) configured to determine a set of dry upmix coefficients ( $C$ ) in order to define a linear mapping of the downmix signal approximating said audio signal; and

a second analyzing section (303) configured to determine an intermediate matrix based on a difference between a covariance of said audio signal as received  
15 and a covariance of said audio signal as approximated by the linear mapping of the downmix signal, wherein the intermediate matrix when multiplied by a predefined matrix corresponds to a set of wet upmix coefficients ( $P$ ) defining a linear mapping of said decorrelated signal as part of parametric reconstruction of said audio signal, wherein the set of wet upmix coefficients includes more coefficients than the number  
20 of elements in the intermediate matrix,

wherein the parametric encoding section is configured to output the downmix signal together with dry upmix parameters ( $\tilde{C}$ ), from which the set of dry upmix coefficients is derivable, and wet upmix parameters ( $\tilde{P}$ ), wherein the intermediate matrix has more elements than the number of output wet upmix parameters, and  
25 wherein the intermediate matrix is uniquely defined by the output wet upmix parameters provided that the intermediate matrix belongs to a predefined matrix class.

19. The audio encoding system of claim 18, wherein the audio encoding system is  
30 adapted to provide a representation of a multichannel audio signal in the form of a plurality of downmix channels and associated dry and wet upmix parameters, wherein the audio encoding system comprises:

a plurality of encoding sections, including parametric encoding sections operable to independently compute respective downmix channels and respective associated upmix parameters based on respective sets of audio signal channels;

5 a control section configured to determine a coding format for said multichannel audio signal corresponding to a partition of the channels of said multichannel audio signal into sets (501–504) of channels to be represented by the respective downmix channels and, for at least some of the downmix channels, by respective associated upmix parameters, the coding format further corresponding to a set of predefined rules for computing at least some of the respective downmix channels,

10 wherein the audio encoding system is configured to encode the multichannel audio signal using a first subset of the plurality of encoding sections, in response to the determined coding format being a first coding format, wherein the audio encoding system is configured to encode the multichannel audio signal using a second subset of the plurality of encoding sections, in response to the determined coding format  
15 being a second coding format, and wherein at least one of the first and second subsets of the encoding sections comprises said first parametric encoding section.

20. The audio encoding system of claim 19, wherein the plurality of encoding sections includes a single-channel encoding section operable to independently  
20 encode no more than a single audio channel in a downmix channel, and wherein at least one of the first and second subsets of the encoding sections comprises the single-channel encoding section.

21. A computer program product comprising a computer-readable medium with  
25 instructions for performing the method of any of claims 1 to 7 and 13 to 17.

22. The method of any of claims 1 to 7 and 13 to 17, the audio decoding system of any of claims 8 to 12, the audio encoding system of any of claim 18 to 20, or the computer program product of claim 21, wherein  $N = 3$  or  $N = 4$ .

30



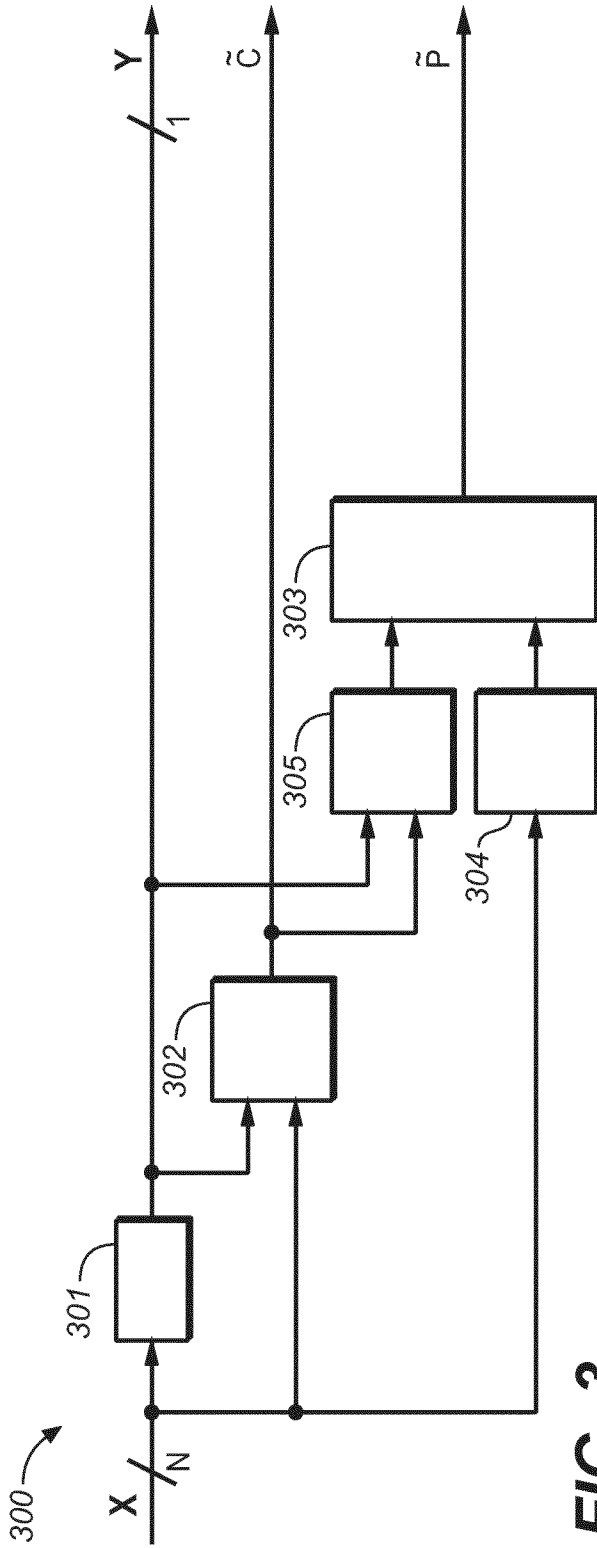


FIG. 3

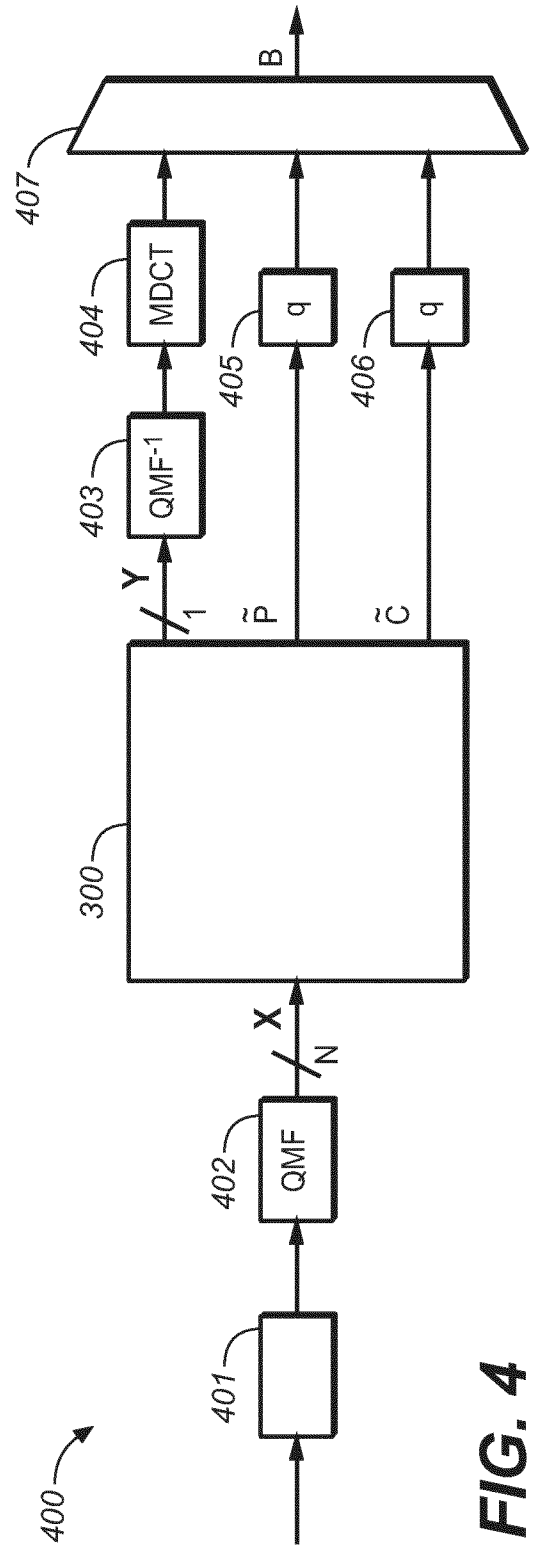
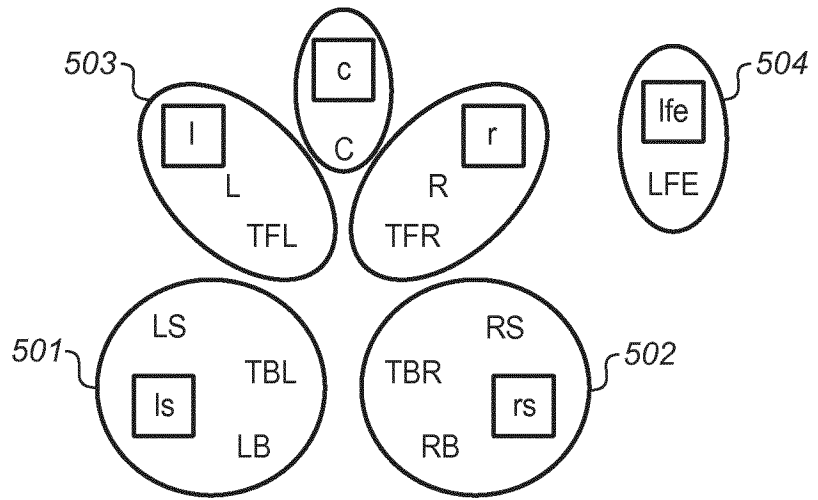


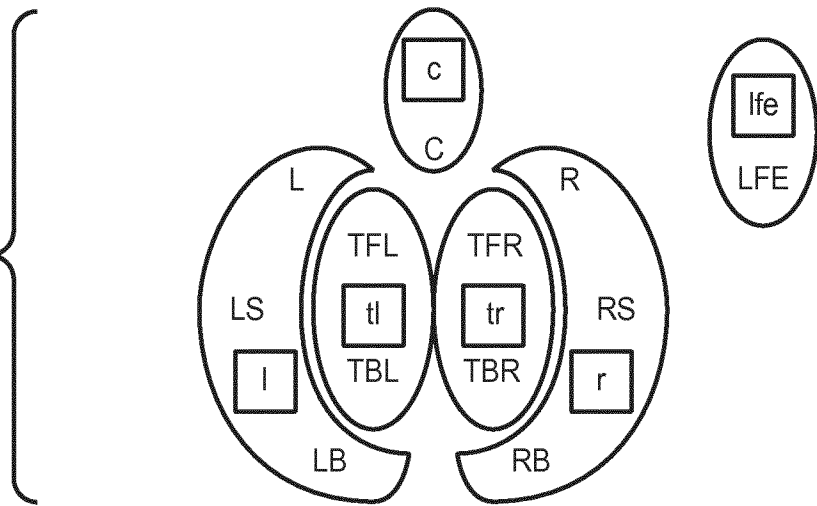
FIG. 4

3 / 7

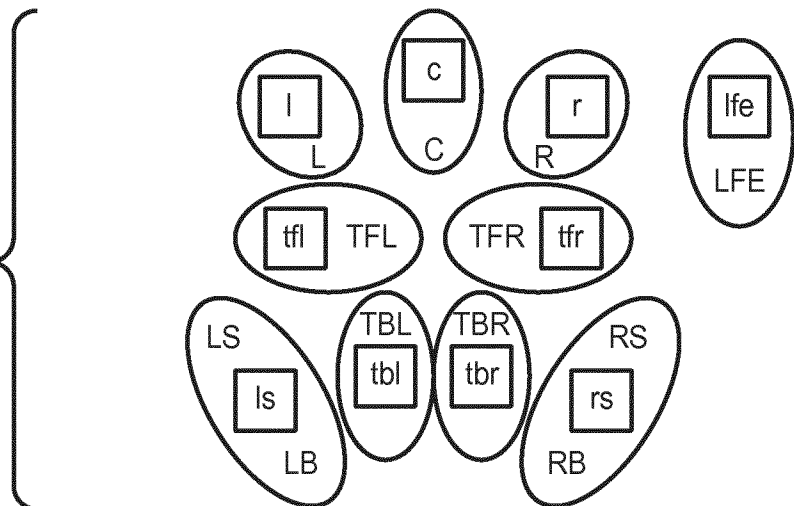
**FIG. 5**



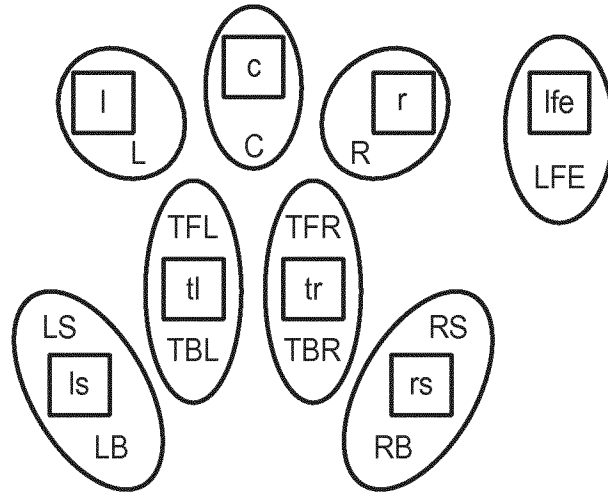
**FIG. 6**



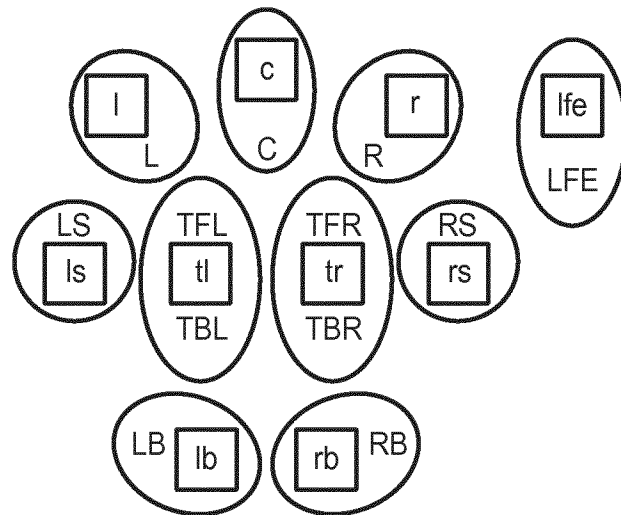
**FIG. 7**



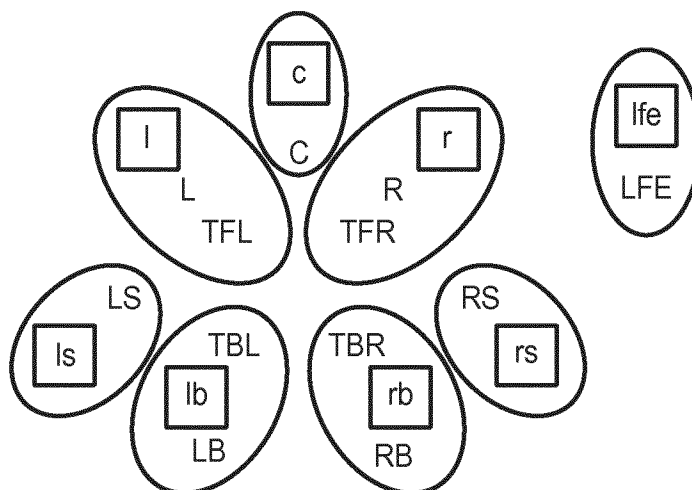
**FIG. 8**



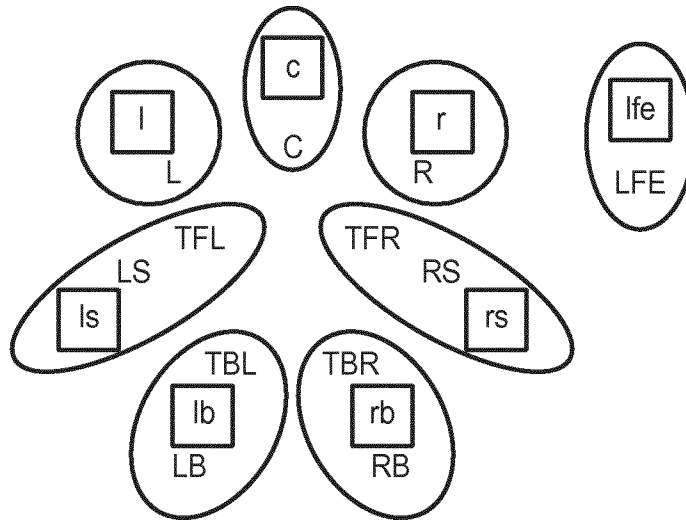
**FIG. 9**



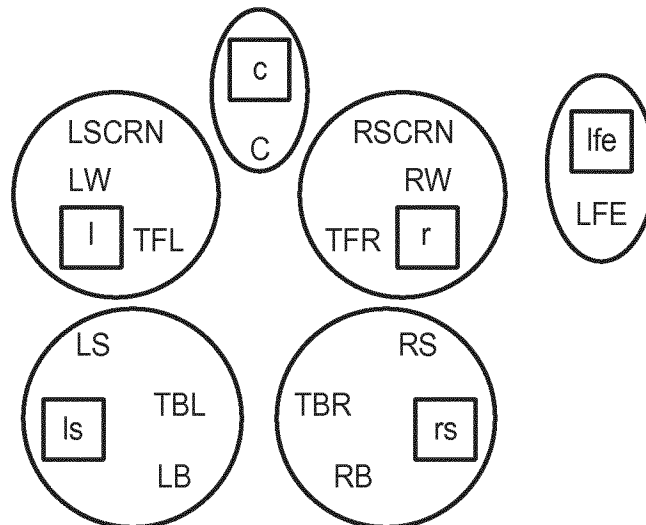
**FIG. 10**



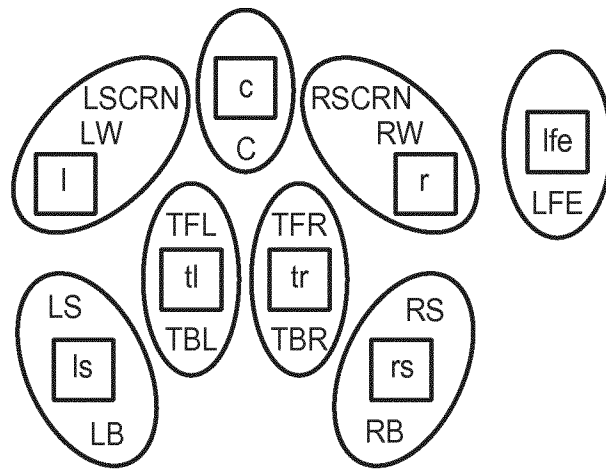
**FIG. 11**



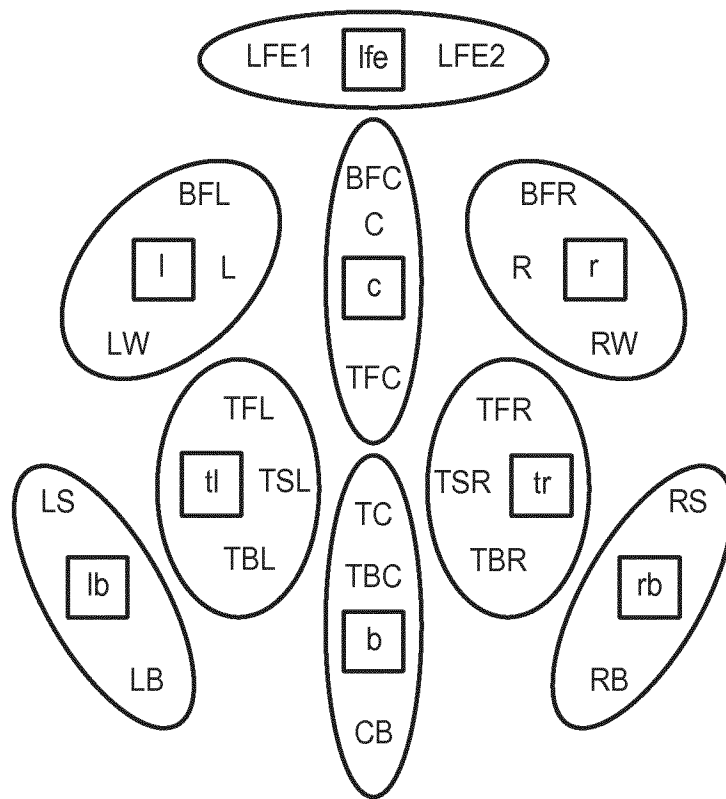
**FIG. 12**



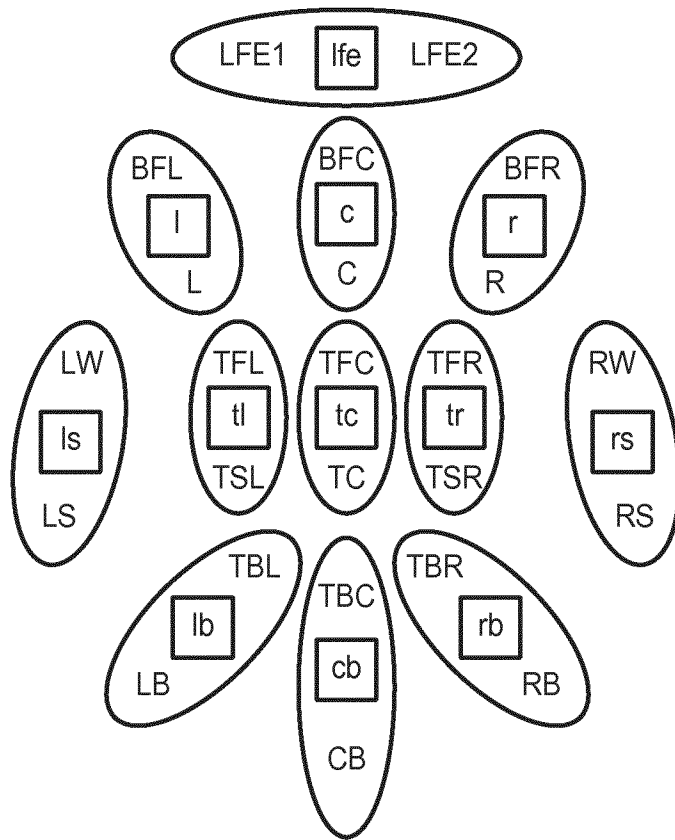
**FIG. 13**



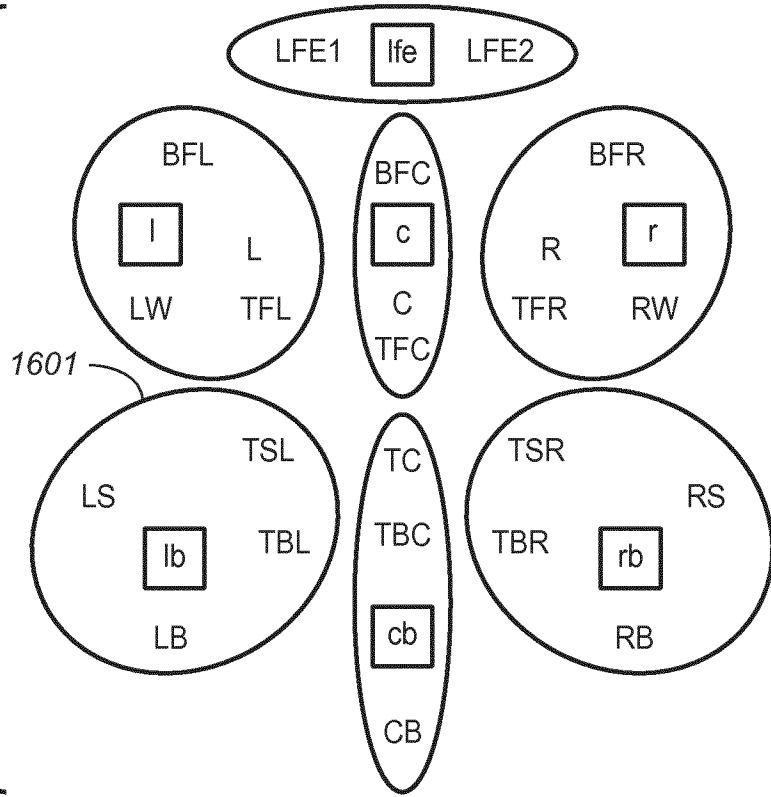
**FIG. 14**



**FIG. 15**



**FIG. 16**



INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2014/072570

A. CLASSIFICATION OF SUBJECT MATTER  
INV. G10L19/008  
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 practicable, search terms used)  
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2008/131903 A1 (DOLBY SWEDEN AB [SE]; FRAUNHOFER GES FORSCHUNG [DE]; ENGDEGARD JONAS []) 6 November 2008 (2008-11-06)	1,8,13, 18,21,22
A	page 12, line 9 - line 16; figures 3B, 4A-F, 14 page 19, line 1 - line 7 eq. (15) on page 29; page 20, line 20 - line 37 -----	2-7, 9-12, 14-17, 19,20
A	WO 2007/007263 A2 (KONINKL PHILIPS ELECTRONICS NV [NL]; CODING TECH AB [SE]; SCHUIJERS ER) 18 January 2007 (2007-01-18) figure 12 -----	1-22
A	EP 2 214 162 A1 (FRAUNHOFER GES FORSCHUNG [DE]) 4 August 2010 (2010-08-04) figures 2A, 2B -----	1-22

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 application or patent 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 January 2015	Date of mailing of the international search report  29/01/2015
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  Müller, Achim

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2014/072570

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO 2008131903	A1	06-11-2008	AU 2008243406 A1	06-11-2008
			BR PI0809760 A2	07-10-2014
			CA 2684975 A1	06-11-2008
			CN 101809654 A	18-08-2010
			EP 2137725 A1	30-12-2009
			ES 2452348 T3	01-04-2014
			HK 1142712 A1	30-05-2014
			JP 5133401 B2	30-01-2013
			JP 2010525403 A	22-07-2010
			KR 20100003352 A	08-01-2010
			KR 20120048045 A	14-05-2012
			MY 148040 A	28-02-2013
			RU 2009141391 A	10-06-2011
			TW 200910328 A	01-03-2009
			US 2010094631 A1	15-04-2010
			WO 2008131903 A1	06-11-2008
WO 2007007263	A2	18-01-2007	AT 433182 T	15-06-2009
			AT 523877 T	15-09-2011
			BR PI0613469 A2	06-11-2012
			CN 101223575 A	16-07-2008
			CN 102013256 A	13-04-2011
			EP 1902443 A2	26-03-2008
			EP 2088580 A2	12-08-2009
			ES 2327158 T3	26-10-2009
			ES 2374309 T3	15-02-2012
			HK 1154984 A1	23-05-2014
			JP 5097702 B2	12-12-2012
			JP 5269039 B2	21-08-2013
			JP 2009501354 A	15-01-2009
			JP 2011059711 A	24-03-2011
			KR 20080037672 A	30-04-2008
			KR 20100134084 A	22-12-2010
PL 2088580 T3	31-07-2012			
RU 2010137467 A	20-03-2012			
US 2008255856 A1	16-10-2008			
WO 2007007263 A2	18-01-2007			
EP 2214162	A1	04-08-2010	AR 075162 A1	16-03-2011
			AU 2010209872 A1	25-08-2011
			CA 2750451 A1	05-08-2010
			CN 102334158 A	25-01-2012
			EP 2214162 A1	04-08-2010
			EP 2382624 A1	02-11-2011
			JP 5490143 B2	14-05-2014
			JP 2012516596 A	19-07-2012
			KR 20110114605 A	19-10-2011
			RU 2011133697 A	10-03-2013
			TW 201032217 A	01-09-2010
			US 2012020499 A1	26-01-2012
WO 2010086218 A1	05-08-2010			