



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
15.04.2015 Bulletin 2015/16

(51) Int Cl.:
G10L 19/008 (2013.01) H04S 3/00 (2006.01)

(21) Application number: **13306391.7**

(22) Date of filing: **09.10.2013**

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME

• **Kordon, Sven**
31515 Wunstorf (DE)
• **Wuebbolt, Oliver**
30559 Hannover (DE)

(71) Applicant: **Thomson Licensing**
92130 Issy-les-Moulineaux (FR)

(74) Representative: **Hartnack, Wolfgang**
Deutsche Thomson OHG
European Patent Operations
Karl-Wiechert-Allee 74
30625 Hannover (DE)

(72) Inventors:
• **Krüger, Alexander**
30655 Hannover (DE)

(54) **Method and apparatus for encoding and for decoding directional side information**

(57) A Higher Order Ambisonics (HOA) representation causes a high data rate. Thus, compression of HOA representations is desirable by decomposing the HOA representation into a directional component and a residual ambient component. The directional component requires directional side information. The overall compression can be improved by coding (18) that directional side information using a specific quantisation of dominant signal direction values $\tilde{\alpha}(k)$, and by establishing a vector $(a(k))$ that defines which directions from a set of pre-defined directions are present in a current audio signal frame $(C(k))$.

nal direction values $\tilde{\alpha}(k)$, and by establishing a vector $(a(k))$ that defines which directions from a set of pre-defined directions are present in a current audio signal frame $(C(k))$.

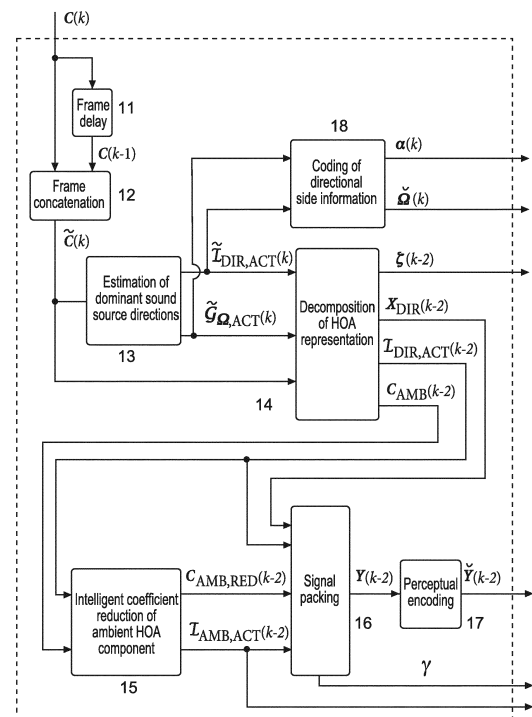


Fig. 1

DescriptionTechnical field

5 **[0001]** The invention relates to a method and to an apparatus for encoding and for decoding directional side information for a 3D audio signal.

Background

10 **[0002]** Higher Order Ambisonics (HOA) represents three-dimensional sound. Other techniques are wave field synthesis (WFS) or channel based approaches like 22.2. In contrast to channel based methods, however, the HOA representation offers the advantage of being independent of a specific loudspeaker set-up. But this flexibility is at the expense of a decoding process which is required for the playback of the HOA representation on a particular loudspeaker set-up. Compared to the WFS approach, where the number of required loudspeakers is usually very large, HOA may also be
15 rendered to set-ups consisting of only few loudspeakers. A further advantage of HOA is that the same representation can also be employed without any modification for binaural rendering to headphones.

[0003] HOA is based on the representation of the spatial density of complex harmonic plane wave amplitudes by a truncated Spherical Harmonics (SH) expansion. Each expansion coefficient is a function of angular frequency, which can be equivalently represented by a time domain function. Hence, without loss of generality, the complete HOA sound
20 field representation actually can be assumed to consist of O time domain functions, where O denotes the number of expansion coefficients. These time domain functions will be equivalently referred to as HOA coefficient sequences or as HOA channels in the following. An HOA representation can be expressed as a temporal sequence of HOA data frames containing HOA coefficients.

[0004] The spatial resolution of the HOA representation improves with a growing maximum order N of the expansion. Unfortunately, the number of expansion coefficients O grows quad-ratically with the order N , in particular $O=(N+1)^2$. For
25 example, typical HOA representations using order $N=4$ require $O=25$ HOA (expansion) coefficients. Accordingly, a total bit rate for the transmission of an HOA representation, given a desired single-channel sampling rate f_s and the number of bits N_b per sample, is determined by $O \cdot f_s \cdot N_b$. Consequently, transmitting an HOA representation of order $N=4$ with a sampling rate of $f_s=48\text{kHz}$ employing $N_b=16$ bits per sample results in a bit rate of 19.2Mbits/s, which is very high for
30 many practical applications like e.g. streaming.

Summary of invention

[0005] Thus, compression of HOA representations is highly desirable. The compression of HOA sound field representations is proposed in patent applications EP 12305537.8, EP 12306569.0 and EP 13305558.2: these approaches
35 have in common that they perform a sound field analysis and decompose the given HOA representation into a directional component and a residual ambient component. On one hand, the resulting compressed representation comprises of a number of quantised signals, resulting from the perceptual coding of the directional signals and relevant coefficient sequences of the ambient HOA component. On the other hand the resulting compressed representation comprises
40 additional side information related to the quantised signals, which side information is necessary for the reconstruction of the HOA representation from its compressed version.

[0006] A problem to be solved by the invention is to further improve the compression of HOA representations. This problem is solved by the methods disclosed in claims 1 and 10. Apparatuses utilising these methods are disclosed in claims 2 and 11.

[0007] The invention deals with the coding of the side information related to the directional component, which additional compression is not addressed in the above-mentioned patent applications EP 12305537.8, EP 12306569.0 and EP 13305558.2. In this prior art, in order to efficiently code or compress a given HOA representation, it is analysed on a
45 frame-by-frame basis and is decomposed into a directional component and a residual ambient component, whereby at compressor side the direction values are estimated based on a pre-defined grid of directions, and these direction values
50 are used for the extraction of directional signals from the given HOA representation in the HOA compressor. According to the invention, the resulting indices of directional signals as well as the direction values are encoded in a particular manner.

[0008] In principle, the inventive method is suited for encoding directional side information for a 3D audio signal, and includes the steps:

- 55
- receiving a data set of dominant signal direction values for a current audio signal frame and a data set of indices of corresponding directional signals, wherein the dominant signal directions were estimated from candidates determined from a pre-defined grid of directions, and the determined dominant signal direction values were used for an

extraction of said directional signals from said 3D audio signal;

- encoding said directional side information for said current audio signal frame by quantising, using said pre-defined grid, the direction values in said received data set of dominant signal directions, and by establishing a vector that defines which directions from a set of pre-defined directions are present in said current audio signal frame.

[0009] In principle the inventive apparatus is suited for encoding directional side information for a 3D audio signal, said apparatus including:

means being adapted for encoding said directional side information for a current audio signal frame, which means receive a data set of dominant signal direction values for said current audio signal frame and a data set of indices of corresponding directional signals, wherein the dominant signal directions were estimated from candidates determined from a pre-defined grid of directions, and the determined dominant signal direction values were used for an extraction of said directional signals from said 3D audio signal, and which means quantise, using said pre-defined grid, the direction values in said received data set of dominant signal directions, and establish a vector that defines which directions from a set of pre-defined directions are present in said current audio signal frame.

[0010] In principle the inventive method is suited for decoding directional side information for a 3D audio signal which directional side information was encoded according to the above encoding method, and includes the steps:

- receiving for a current audio signal frame direction values quantised according to said pre-defined grid and a vector that comprises encoded indices about which directions from said set of pre-defined directions are present in said current audio signal frame;
- re-quantising according to said pre-defined grid said quantised direction values, and decoding said vector;
- providing from said re-quantised direction values and said decoded vector a data set of dominant signal direction values for said current audio signal frame and a data set of indices of corresponding directional signals.

[0011] In principle the inventive apparatus is suited for decoding directional side information for a 3D audio signal, which directional side information was encoded according to the above encoding method, said apparatus including means being adapted for:

- receiving for a current audio signal frame direction values quantised according to said pre-defined grid and a vector that comprises encoded indices about which directions from said set of pre-defined directions are present in said current audio signal frame;
- re-quantising according to said pre-defined grid said quantised direction values, and decoding said vector;
- providing from said re-quantised direction values and said decoded vector a data set of dominant signal direction values for said current audio signal frame and a data set of indices of corresponding directional signals.

[0012] Advantageous additional embodiments of the invention are disclosed in the respective dependent claims.

Brief description of drawings

[0013] Exemplary embodiments of the invention are described with reference to the accompanying drawings, which show in:

- Fig. 1 Block diagram for an HOA compression including the encoding of directional side information;
- Fig. 2 Block diagram for an HOA decompression including the decoding of directional side information;
- Fig. 3 Spherical coordinate system;
- Fig. 4 Exemplary illustration of direction estimation;
- Fig. 5 Activity diagram related to Fig. 4.

Description of embodiments

[0014] In order to compress a given HOA representation, it is analysed on a frame-by-frame basis and decomposed into a directional component and a residual ambient component, for example as described in patent applications EP 12305537.8, EP 12306569.0 and EP 13305558.2.

[0015] As an example for embedding directional side information coding according to the invention in an HOA processing that uses splitting into directional and residual ambient components, Fig. 1 shows an HOA compression processing as described in patent application EP 13305558.2, in which - following estimation of dominant sound source directions

- a coding of directional side information is carried out. For the HOA representation compression a frame-wise processing with non-overlapping input frames $\mathbf{C}(k)$ of HOA coefficient sequences of length L is used, where k denotes the frame index. The first step or stage 11/12 in Fig. 1 is optional and consists of concatenating the non-overlapping k -th and $(k-1)$ -th frames of HOA coefficient sequences into a long frame $\tilde{\mathbf{C}}(k)$ as $\tilde{\mathbf{C}}(k) := [\mathbf{C}(k-1) \mathbf{C}(k)]$ (the tilde symbol indicates long overlapping frames). This long frame is 50% overlapped with an adjacent long frame and is successively used for the estimation of dominant sound source directions. If step/stage 11/12 is not present, the tilde symbol has no specific meaning.

[0016] In step or stage 13 dominant sound sources are estimated. The estimation provides a data set $\tilde{\mathcal{I}}_{\text{DIR,ACT}}(k)$

$\subseteq \{1, \dots, D\}$ of indices of directional signals that have been detected as well as the set $\tilde{\mathcal{G}}_{\Omega, \text{ACT}}(k)$ of corresponding direction estimates. D denotes the maximum number of directional signals that has to be set before starting the HOA compression. In step or stage 14, the current frame $\tilde{\mathbf{C}}(k)$ of HOA coefficient sequences is decomposed into a number

of directional signals $\mathbf{X}_{\text{DIR}}(k-2)$ belonging to the directions contained in the set $\tilde{\mathcal{G}}_{\Omega, \text{ACT}}(k)$, and a residual ambient HOA component $\mathbf{C}_{\text{AMB}}(k-2)$. The delay of two frames is introduced as a result of overlap-add processing in order to obtain smooth signals. It is assumed that $\mathbf{X}_{\text{DIR}}(k-2)$ is containing a total of D channels, of which however only those corresponding to the active directional signals are non-zero. The indices specifying these channels are assumed to be output in the data set $\mathcal{I}_{\text{DIR,ACT}}(k-2)$. Additionally, the decomposition in step/stage 14 provides some parameters $\zeta(k-2)$ which are used at decompression side for predicting portions of the original HOA representation from the directional signals. In step or stage 15, the number of coefficients of the ambient HOA component $\mathbf{C}_{\text{AMB}}(k-2)$ is reduced so as to contain only $O_{\text{RED}} + D - N_{\text{DIR,ACT}}(k-2)$ non-zero HOA coefficient sequences, where $N_{\text{DIR,ACT}}(k-2) = |\mathcal{I}_{\text{DIR,ACT}}(k-2)|$ indicates the cardinality of the data set $\mathcal{I}_{\text{DIR,ACT}}(k-2)$, i.e. the number of active directional signals in frame $k-2$. Since the ambient HOA component is assumed to be always represented by a minimum number O_{RED} of HOA coefficient sequences, this problem can be actually reduced to the selection of the remaining $D - N_{\text{DIR,ACT}}(k-2)$ HOA coefficient sequences out of the possible $O - O_{\text{RED}}$ ones. In order to obtain a smooth reduced ambient HOA representation, this choice is accomplished such that, compared to the choice taken at the previous frame $k-3$, as few changes as possible will occur.

[0017] The final ambient HOA representation with the reduced number of $O_{\text{RED}} + N_{\text{DIR,ACT}}(k-2)$ non-zero coefficient sequences is denoted by $\mathbf{C}_{\text{AMS,RED}}(k-2)$. The indices of the chosen ambient HOA coefficient sequences are output in the data set $\mathcal{I}_{\text{AMB,ACT}}(k-2)$. In step/stage 16, the active directional signals contained in $\mathbf{X}_{\text{DIR}}(k-2)$ and the HOA coefficient sequences contained in $\mathbf{C}_{\text{AMS,RED}}(k-2)$ are assigned to the frame $\mathbf{Y}(k-2)$ of I channels for individual perceptual encoding.

[0018] According to the present invention, the data set $\tilde{\mathcal{I}}_{\text{DIR,ACT}}(k) \subseteq \{1, \dots, D\}$ of indices of directional signals and

the data set $\tilde{\mathcal{G}}_{\Omega, \text{ACT}}(k)$ of corresponding direction value estimates from the estimation step/stage 13 are fed to a step or stage 18 that encodes the directional side information as described in the following. Step/stage 18 outputs a vector $\mathbf{a}(k)$ denoting which directional signals are active in frame k , as well as a coded representation $\tilde{\mathcal{N}}(k)$ of all directions.

The values of $\tilde{\mathcal{N}}(k)$ can be entropy encoded.

[0019] The HOA decompression processing described in patent application EP 13305558.2, together with an additional step or stage 34 for decoding the received encoded directional side information, is depicted in Fig. 2. Step/stage 34

receives vector $\mathbf{a}(k)$ denoting which directional signals are active in frame k , and the coded representation $\tilde{\mathcal{N}}(k)$ of all directions. Step/stage 34 decodes as described below this directional side information and outputs the data set

$\tilde{\mathcal{I}}_{\text{DIR,ACT}}(k) \subseteq \{1, \dots, D\}$ of indices of directional signals and the set $\tilde{\mathcal{G}}_{\Omega, \text{ACT}}(k)$ of corresponding direction estimates.

In step or stage 31 a perceptual decoding of the I signals contained in $\tilde{\mathbf{Y}}(k-2)$ is performed in order to obtain the I decoded signals in $\hat{\mathbf{Y}}(k-2)$. In signal re-distributing step or stage 32, the perceptually decoded signals in $\hat{\mathbf{Y}}(k-2)$ are redistributed in order to recreate the frame $\hat{\mathbf{X}}_{\text{DIR}}(k-2)$ of directional signals and the frame $\hat{\mathbf{C}}_{\text{AMB,RED}}(k-2)$ of the ambient HOA component. The information about how to re-distribute the signals is obtained by reproducing the assigning operation performed for the HOA compression, using the index data sets $\mathcal{I}_{\text{DIR,ACT}}(k)$ and $\mathcal{I}_{\text{AMB,ACT}}(k-2)$.

[0020] In composition step or stage 33, a current frame $\hat{\mathbf{C}}(k-3)$ of the desired total HOA representation is re-composed

using the frame $\hat{\mathbf{X}}_{\text{DIR}}(k-2)$ of the directional signals, the set $\tilde{\mathcal{I}}_{\text{DIR,ACT}}(k)$ of the active directional signal indices together with the set $\tilde{\mathcal{G}}_{\Omega, \text{ACT}}(k)$ of the corresponding directions, the parameters $\zeta(k-2)$ for predicting portions of the HOA

representation from the directional signals, and the frame $\hat{C}_{\text{AMS,RED}}(k-2)$ of HOA coefficient sequences of the reduced ambient HOA component. I.e., directional signals with respect to uniformly distributed directions are predicted from the directional signals ($\hat{X}_{\text{DIR}}(k-2)$) using the received parameters ($\zeta(k-2)$) for such prediction, and thereafter the current decompressed frame ($\hat{C}(k-3)$) is re-composed from the frame of directional signals ($\hat{X}_{\text{DIR}}(k-2)$), the predicted portions and the reduced ambient HOA component ($\hat{C}_{\text{AMB,RED}}(k-2)$).

[0021] As mentioned above, in patent application EP 13305558.2 the directional HOA component for the k -th frame is represented by a number $D_{\text{ACT}}(k)$ of directional signals and additional side information. This side information comprises on one hand the set

$$\tilde{I}_{\text{DIR,ACT}}(k) := \{i_{\text{ACT},d}(k) | d = 1, \dots, D_{\text{ACT}}(k)\} \quad (1)$$

of indices $i_{\text{ACT},d}(k)$ of directional signals that have been detected. On the other hand, the side information consists of the set

$$\tilde{G}_{\Omega,\text{ACT}}(k) := \{\Omega_{\text{ACT},d}(k) | d = 1, \dots, D_{\text{ACT}}(k)\} \quad (2)$$

of the corresponding directions $\Omega_{\text{ACT},d}(k)$.

[0022] To illustrate the meaning of the side information by way of an example, the case is considered where the maximum number D of directional signals is equal to two. Fig. 4 illustrates an exemplary result of the direction estimation for the first 7 frames (cf. Fig. 3 and the corresponding description of the representation of a direction in a spherical coordinate system). The dots in Fig. 4 represent a grid of possible directions. The direction estimates related to the directional signal with index 1 are marked by diamonds and the direction estimates related to the directional signal with index 2 are marked by crosses. The directional signal with index 1 representing a first trajectory is supposed to be active from frame $k = 1$ to $k = 4$, whereas the directional signal with index 2 representing a second trajectory is supposed to be active from frame $k = 3$ to $k = 7$.

[0023] This activity information is additionally illustrated in Fig. 5, which shows for each frame index k if the direction with the respective index is active (indicated by white) or not (indicated by black). The resulting index sets $\tilde{G}_{\Omega,\text{ACT}}(k)$, $k = 1, 2, \dots, 7$ corresponding to Fig. 4 are summarised in Table 1:

k	1	2	3	4	5	6	7
$\tilde{I}_{\text{DIR,ACT}}(k)$	{1}	{1}	{1,2}	{1,2}	{2}	{2}	{2}

Coding of the indices of directional signals

[0024] Because the indices of active directional signals correspond to the indices of D channels to which these directional signals are assigned, for coding the indices of the directional signals a bit array of length D is used that is represented by the vector

$$\mathbf{a}(k) := [a_1(k) \ a_2(k) \ \dots \ a_D(k)]^T, \quad (3)$$

where $(\cdot)^T$ denotes transposition and the elements $a_i(k)$, $i \in \{1, \dots, D\}$, are defined as follows:

$$a_i(k) = \begin{cases} 1 & \text{if the } i\text{-th directional signal is active in frame } k \\ 0 & \text{else} \end{cases} \quad (4)$$

[0025] There are exactly $D_{\text{ACT}}(k)$ non-zero elements in vector $\mathbf{a}(k)$, which corresponds to the number of active directional signals.

[0026] For the compression of 3D audio signals it is reasonable to assume not more than four to eight directional signals in a frame, and therefore $\mathbf{a}(k)$ will contain e.g. 4 to 8 bits per frame. A current frame k will contain none, part or all of this set of pre-determined directional signals. For a current frame k the current vector $\mathbf{a}(k)$ is transferred to the decoder or decompression side.

[0027] To illustrate the coding of indices of directional signals, in Table 2 the values of the vector $\mathbf{a}(k)$, $k=1, \dots, 7$, for

the index sets $\tilde{l}_{\text{DIR,ACT}}(k)$ given in Table 1 are provided as an example:

k	1	2	3	4	5	6	7
$\mathbf{a}(k)$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Coding of the directions

[0028] The direction values used for a frame may vary from frame to frame. Assuming that the indices $i_{\text{ACT},d}(k)$, $d = 1, \dots, D_{\text{ACT}}(k)$ are ordered in an ascending order, it is sufficient to code the direction values $\Omega_{\text{ACT},d}(k)$, $d = 1, \dots, D_{\text{ACT}}(k)$ one after the other in order to be able to unambiguously link them to the indices. In other words, given the vector $\mathbf{a}(k)$ and the sequence of coded directions, it can be assumed that the coded direction value $\Omega_{\text{ACT},1}(k)$ corresponds to the index indicated by the first non-zero element in $\mathbf{a}(k)$, and the coded direction value $\Omega_{\text{ACT},2}(k)$ is assumed to correspond to the index indicated by the second non-zero element in $\mathbf{a}(k)$, etc. As mentioned above, for frame k the coded representation of all direction values in the set $\tilde{\Omega}_{\text{ACT}}(k)$ are denoted by $\tilde{\Omega}(k)$, and $\tilde{\Omega}(k)$ is transferred to the decoder or decompression side.

[0029] In the following, the problem of how to efficiently encode the direction values $\Omega_{\text{ACT},d}(k)$, $d = 1, \dots, D_{\text{ACT}}(k)$ for generating $\tilde{\Omega}(k)$ is addressed. In principle, assuming a spherical coordinate system as shown in Fig. 3, each direction $\Omega_{\text{ACT},d}(k)$ can be unambiguously represented by the tuple

$$\Omega_{\text{ACT},d}(k) = (\theta_d(k), \phi_d(k)) , \quad (5)$$

where $\theta_d(k) \in [0, \pi]$ denotes an inclination angle measured from the polar axis \mathbf{z} and $\phi_d(k) \in [0, 2\pi[$ indicates an azimuth angle measured counter-clockwise in the $x - y$ plane from the x axis.

[0030] On one hand, the inclination and azimuth angles could be quantised individually, in particular by assuming $M_\theta = 2^{Q_\theta}$ possible discrete values for the inclination angle and $M_\phi = 2^{Q_\phi}$ possible discrete values for the azimuth angle, resulting in a total number of $Q_\theta + Q_\phi$ bits required for the coding of a single direction. On the other hand, the disadvantage of such individual or specific quantisation processing is that likely such specifically quantised direction values will not exactly match with the pre-defined grid of directions: in order to not introduce errors when carrying out in the HOA decompressor the re-synthesis of the HOA representation of the directional signals due to direction quantisation errors, the extraction of directional signals from the given HOA representation in the HOA compressor in step/stage 13 in Fig. 1 is based on that pre-defined grid of directions. Patent applications EP 12306569.0 and EP 13305558.2 describe how directional signals can be extracted from an HOA representation.

[0031] The problem that the directions of the quantised direction values do not exactly match with the estimated directions can be solved in a first embodiment by exploiting the fact that the splitting into directional and ambient residual components and the direction estimation described in patent applications EP 12306569.0 and EP 13305558.2 is based on a direction search which is carried out on a fixed grid of directions (cf. patent application EP 13305156.5 for a description of direction estimation as an example). Such fixed grid represents the above-mentioned re-quantisation. In particular, the estimated direction $\Omega_{\text{ACT},d}(k)$ is an element of a set $\{\Omega_q | q = 1, \dots, Q\}$ of Q predefined directions. Exploiting this knowledge, the direction values in step/stage 18 can be quantised according to this pre-defined grid, by representing a direction by the index $q \in \{1, \dots, Q\}$. Then, a quantised representation of a single direction will require $\lceil \log_2(Q) \rceil$ bits. For instance, using a grid consisting of $Q=900$ predefined directions would require 10 bits for a corresponding quantisation.

[0032] Such coding of the directions offers the further advantage that it is not recursive, meaning that no knowledge of the direction estimates from previous frames is required for the decoding of the directions. However, a disadvantage of such processing is that in general it does not achieve the related minimum possible average bit rate.

[0033] At decompressor side, for a current audio signal frame $\mathbf{C}(k)$, the possibly entropy encoded representation $\tilde{\Omega}(k)$ of all directions is received, wherein these direction values $\tilde{\Omega}(k)$ were quantised according to said pre-defined grid, and vector $\mathbf{a}(k)$ is received that comprises the encoded indices about which directions from the set of pre-defined directions are present in the current audio signal frame $\mathbf{C}(k)$. If necessary, an entropy decoding takes place. The quantised

direction values $\tilde{\Omega}(k)$ are re-quantised according to the pre-defined grid said, and vector $\mathbf{a}(k)$ is decoded. From the re-quantised direction values and the decoded vector a data set $\tilde{\mathcal{G}}_{\Omega, \text{ACT}}(k)$ of dominant signal direction values for said current audio signal frame $\mathbf{C}(k)$ and a data set $\tilde{\mathcal{I}}_{\text{DIR}, \text{ACT}}(k)$ of indices of corresponding directional signals is provided.

[0034] In a second embodiment, the average bit rate for the coding of the directions for successive frames is further reduced by exploiting the relation between the direction estimates of successive frames. In particular, the direction estimation as proposed in patent application EP 13305558.2 is based on an sound source movement model, which predicts the direction of a sound source in the k -th frame based on its movement between the $(k-2)$ -th and $(k-1)$ -th frame.

[0035] Therefore the quantised direction values (e.g. the direction index as proposed in the first embodiment) of the k -th frame are in the second embodiment coded using entropy coding like e.g. Huffman coding. The individual code words for the direction values have a variable bit size depending on the frame adaptively determined probability of the individual directions. In particular, direction values with a high probability are coded using small-size code words and direction values with a low probability are coded using large-size code words.

[0036] Such a coding strategy requires computation of the probability for the individual directions during HOA decomposition in the same way as for the HOA compression. At decompression side, the received entropy encoded quantised direction values are entropy decoded wherein frame adaptively an probability of the individual directions is determined. However, this requires a high computational complexity in the HOA decompressor for computing the probabilities of the Q possible directions in each frame. Further, the processing is recursive, meaning that the decoding of a direction at decompression side is based on the knowledge of the directions from the previous two frames.

[0037] In a third embodiment, in order to reduce the computational complexity introduced by evaluating frame-by-frame the a-priori probabilities for all Q possible directions in the HOA decompressor as described for the second embodiment, the number of possible probabilities is constrained to Q by making the probabilities dependent on the corresponding direction estimate in the last frame.

[0038] One possibility to define such conditional probabilities is to set them inversely proportional to the angular distance between a direction estimate in the current frame and the corresponding direction estimate in the last frame. Another possibility is to measure the conditional a-priori probabilities of the direction estimates from some HOA representations instead of setting them.

[0039] At decompression side, the received entropy encoded quantised direction values $\tilde{\Omega}(k)$ are frame adaptively entropy decoded depending on an probability of the individual directions, whereby the number of possible probabilities is constrained to the number of directions in the pre-defined grid and the probabilities are dependent on the corresponding direction estimate in the last frame. Such a technique requires that the HOA decompressor holds for each of the Q possible test directions an entropy decoding table (e.g. a Huffman table) containing Q code words and respective indices of the temporally following directions.

[0040] In a fourth embodiment, instead of considering the conditioned probabilities for the entropy coding, the non-conditional probabilities of the quantised directions are employed. Such probabilities can be measured from some test HOA representations, or can be assigned according to expectations about typical HOA sound field representation. For example, high probabilities are assigned for directions in the front and low probabilities for directions in the back.

[0041] Such a processing has the advantage of not being recursive, i.e. the decoding of a direction value is not based on the knowledge of any directions from previous frames. However, due to the use of non-conditional probabilities, in general the efficiency of this kind of processing is likely lower than that of the third embodiment.

[0042] In a fifth embodiment, in order to alleviate computational load and extensive storage requirements for the HOA decompressor, for each frame it is decided which one of the above embodiments is used, resulting in a set of four (or less) modes:

- processing according to the first embodiment;
- processing according to a combination of the first embodiment and the second embodiment;
- processing according to a combination of the first embodiment and the third embodiment;
- processing according to a combination of the first embodiment and the fourth embodiment.

[0043] The mode decision can be indicated by a Boolean variable which is prepended to the coded representation of a direction. Such a mode decision will in most cases minimise the bit amount of the corresponding code.

[0044] The inventive processing can be carried out by a single processor or electronic circuit, or by several processors or electronic circuits operating in parallel and/or operating on different parts of the inventive processing.

[0045] The invention can be applied in any application where some directional information has to be efficiently coded, e.g. object based 3D audio where directional signals and object based side information have to be coded.

Claims

1. Method for encoding directional side information for a 3D audio signal, **characterised by** the steps:

- receiving a data set $(\tilde{G}_{\Omega,ACT}(k))$ of dominant signal direction values for a current audio signal frame $(\mathbf{C}(k))$ and a data set $(\tilde{I}_{DIR,ACT}(k))$ of indices of corresponding directional signals, wherein the dominant signal directions were estimated (13) from candidates determined from a pre-defined grid of directions, and the determined dominant signal direction values were used for an extraction (13) of said directional signals from said 3D audio signal;
 - encoding (18) said directional side information for said current audio signal frame $(\mathbf{C}(k))$ by quantising, using said pre-defined grid, the direction values in said received data set $(\tilde{G}_{\Omega,ACT}(k))$ of dominant signal directions, and by establishing a vector $(\mathbf{a}(k))$ that defines which directions from a set of pre-defined directions are present in said current audio signal frame $(\mathbf{C}(k))$.

2. Apparatus for encoding directional side information for a 3D audio signal, said apparatus including:

means (18) being adapted for encoding (18) said directional side information for a current audio signal frame $(\mathbf{C}(k))$, which means receive a data set $(\tilde{G}_{\Omega,ACT}(k))$ of dominant signal direction values for said current audio signal frame $(\mathbf{C}(k))$ and a data set $(\tilde{I}_{DIR,ACT}(k))$ of indices of corresponding directional signals, wherein the dominant signal directions were estimated (13) from candidates determined from a pre-defined grid of directions, and the determined dominant signal direction values were used for an extraction (13) of said directional signals from said 3D audio signal,

and which means (18) quantise, using said pre-defined grid, the direction values in said received data set $(\tilde{G}_{\Omega,ACT}(k))$ of dominant signal directions, and establish a vector $(\mathbf{a}(k))$ that defines which directions from a set of pre-defined directions are present in said current audio signal frame $(\mathbf{C}(k))$.

3. Method according to claim 1, or apparatus according to claim 2, wherein frame adaptively an probability of the individual directions is determined based on the knowledge of the directions from the previous two frames, and wherein said quantised direction values are coded using entropy coding with variable bit size code words for the direction values depending on said probability.

4. Method according to claim 1, or apparatus according to claim 2, wherein frame adaptively an probability of the individual directions is determined whereby the number of possible probabilities is constrained to the number of directions in said pre-defined grid and said probabilities are dependent on the corresponding direction estimate in the last frame, and wherein said quantised direction values are coded using entropy coding with variable bit size code words for the direction values depending on said probability.

5. Method according to the method of claim 4, or apparatus according to the apparatus of claim 4, wherein said conditional probabilities are inversely proportional to the angular distance between a direction estimate in the current frame and the corresponding direction estimate in the last frame.

6. Method according to the method of claim 4, or apparatus according to the apparatus of claim 4, wherein said conditional a-priori probabilities are measured from several HOA representations.

7. Method according to claim 1, or apparatus according to claim 2, wherein said quantised direction values are coded using entropy coding with variable bit size code words for the direction values depending on non-conditional a-priori probabilities.

8. Method according to the method of one of claims 1, 3, 4 or 7, or apparatus according to the apparatus of one of claims 2, 3, 4 or 7, wherein for each 3D audio signal frame $(\mathbf{C}(k))$ it is decided which one of the processings according to claims 1, 3, 4 or 7 is carried out, and a corresponding mode code word representing said selection is provided.

9. Method according to the method of one of claims 1 and 3 to 8, or apparatus according to the apparatus of one of claims 2 to 8, wherein said 3D audio signal is an HOA audio signal.

10. Computer program product comprising instructions which, when carried out on a computer, perform the method according to one of claims 1 and 3 to 9.

11. Method for decoding directional side information for a 3D audio signal, which directional side information was encoded according to claim 1, **characterised by** the steps:

- receiving for a current audio signal frame ($\mathbf{C}(k)$) direction values ($\tilde{\mathbf{D}}(k)$) quantised according to said pre-defined grid and a vector ($\mathbf{a}(k)$) that comprises encoded indices about which directions from said set of pre-defined directions are present in said current audio signal frame ($\mathbf{C}(k)$);

- re-quantising (34) according to said pre-defined grid said quantised direction values ($\tilde{\mathbf{D}}(k)$), and decoding (34) said vector ($\mathbf{a}(k)$);

- providing (34) from said re-quantised direction values and said decoded vector a data set ($\tilde{\mathbf{G}}_{\Omega, \text{ACT}}(k)$) of dominant signal direction values for said current audio signal frame ($\mathbf{C}(k)$) and a data set ($\tilde{\mathbf{I}}_{\text{DIR}, \text{ACT}}(k)$) of indices of corresponding directional signals.

12. Apparatus for decoding directional side information for a 3D audio signal, which directional side information was encoded according to claim 1, said apparatus including means (34) being adapted for:

- receiving for a current audio signal frame ($\mathbf{C}(k)$) direction values ($\tilde{\mathbf{D}}(k)$) quantised according to said pre-defined grid and a vector ($\mathbf{a}(k)$) that comprises encoded indices about which directions from said set of pre-defined directions are present in said current audio signal frame ($\mathbf{C}(k)$);

- re-quantising according to said pre-defined grid said quantised direction values ($\tilde{\mathbf{D}}(k)$), and decoding said vector ($\mathbf{a}(k)$);

- providing from said re-quantised direction values and said decoded vector a data set ($\tilde{\mathbf{G}}_{\Omega, \text{ACT}}(k)$) of dominant signal direction values for said current audio signal frame ($\mathbf{C}(k)$) and a data set ($\tilde{\mathbf{I}}_{\text{DIR}, \text{ACT}}(k)$) of indices of corresponding directional signals.

13. Method according to claim 11, or apparatus according to claim 12, wherein said received quantised direction values ($\tilde{\mathbf{D}}(k)$) are entropy encoded quantised direction values and frame adaptively an a-priori probability of the individual directions is determined based on the knowledge of the directions from the previous two frames, and wherein said entropy encoded quantised direction values are entropy decoded depending on said a-priori probability.

14. Method according to claim 11, or apparatus according to claim 12, wherein said received quantised direction values ($\tilde{\mathbf{D}}(k)$) are entropy encoded quantised direction values and frame adaptively an a-priori probability of the individual directions is determined whereby the number of possible a-priori probabilities is constrained to the number of directions in said pre-defined grid and said a-priori probabilities are dependent on the corresponding direction estimate in the last frame, and wherein said entropy encoded quantised direction values are entropy decoded with variable bit size code words depending on said a-priori probability.

15. Method according to the method of one of claims 11, 13 and 14, or apparatus according to the apparatus of one of claims 12 to 14, wherein said 3D audio signal is an HOA audio signal.

16. Method according to the method of claims 11 and 14, or apparatus according to the apparatus of claims 12 and 14, wherein said decoding of directional side information is carried out in an HOA decompressor and this HOA decompressor holds for each possible direction of said pre-defined grid an entropy decoding table containing a corresponding code word and respective indices of the temporally following directions.

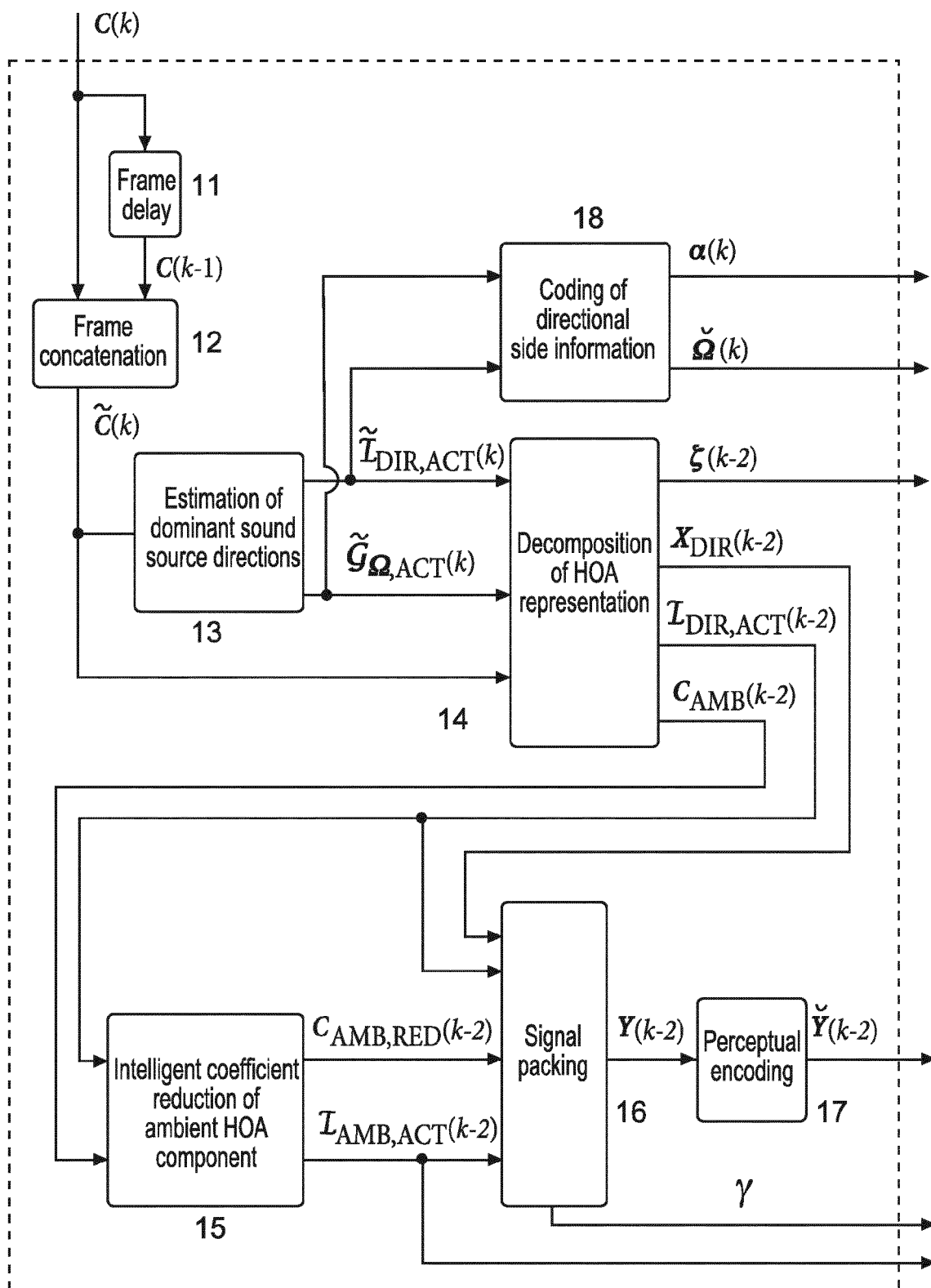


Fig. 1

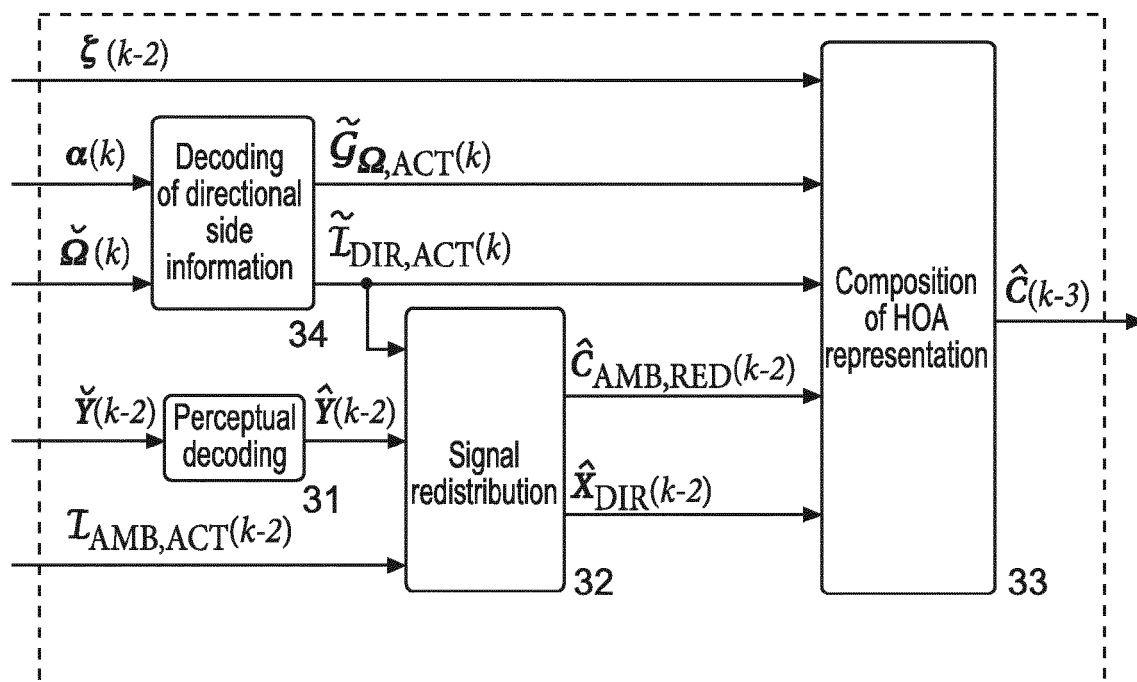


Fig. 2

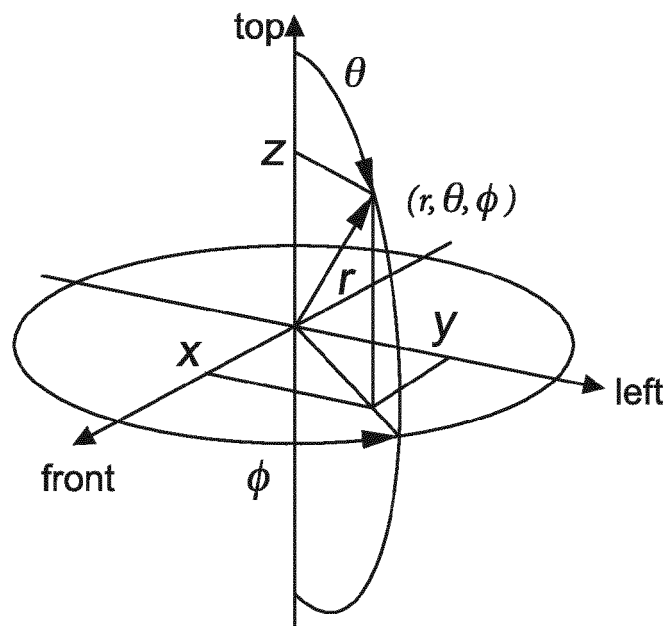


Fig. 3

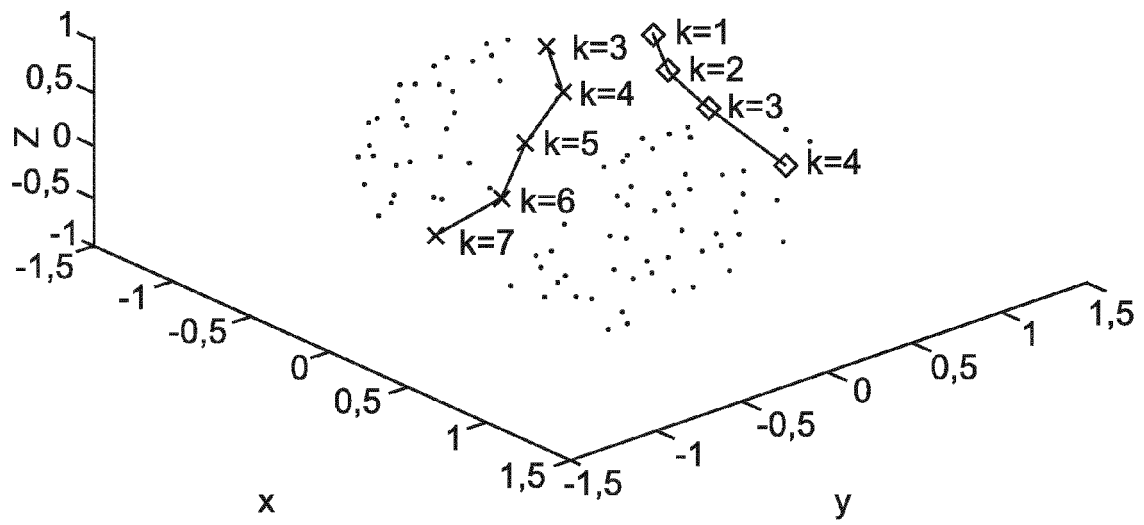


Fig. 4

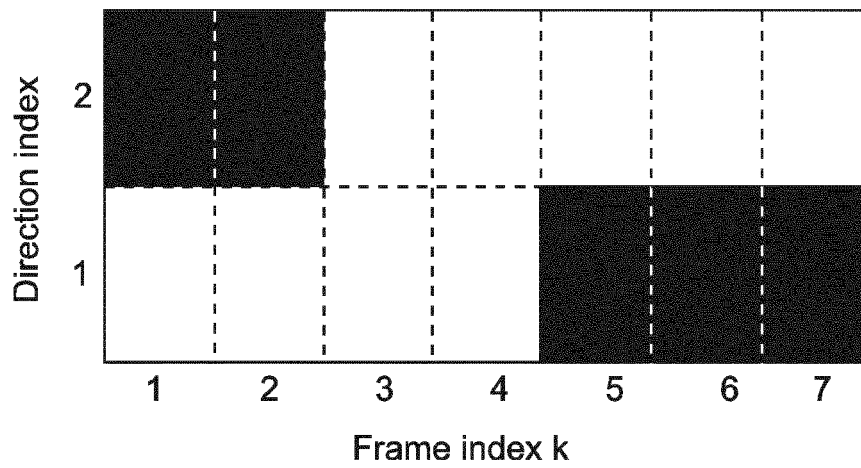


Fig. 5



EUROPEAN SEARCH REPORT

 Application Number
EP 13 30 6391

5

10

15

20

25

30

35

40

45

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2011/249821 A1 (JAILLET FLORENT [FR] ET AL) 13 October 2011 (2011-10-13)	1-3, 7-13, 15, 16	INV. G10L19/008 H04S3/00
A	* page 1, paragraph [0014] - page 2, paragraph [0054] * * page 3, paragraph [0081] - page 5, paragraph [0134]; figures 2, 5 * * page 7, paragraph [0165] - paragraph [0174] *	4-6, 14	
X	----- MANUEL BRIAND: "Codage paramétrique basé sur l'Analyse en Composante Principale", INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE 20 March 2007 (2007-03-20), pages 133-147, XP002661237, Retrieved from the Internet: URL: http://tel.archives-ouvertes.fr/docs/00/14/18/62/PDF/2007_Briand_ParametricAudioCoding.pdf [retrieved on 2010-10-13]	1-3, 7-13, 15, 16	
A	* page 133, paragraph 5. - page 135, paragraph 5.1.1 * * page 139, paragraph 5.1.2 - page 145, paragraph 5.1.2.2 *	4-6, 14	TECHNICAL FIELDS SEARCHED (IPC) G10L H04S
X	----- US 2007/269063 A1 (GOODWIN MICHAEL [US] ET AL) 22 November 2007 (2007-11-22)	1-3, 7-13, 15, 16	
A	* page 2, paragraph [0029] - page 4, paragraph [0046] * * page 5, paragraph [0064] - page 7, paragraph [0083] * * page 9, paragraph [0097] - paragraph [0100] * * page 10, paragraph [0110] - page 11, paragraph [0115] *	4-6, 14	
	----- -/-		
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 6 March 2014	Examiner Virette, David
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons ----- & : member of the same patent family, corresponding document	

2

50

55

EPO FORM 1503 03.82 (P04C01)



EUROPEAN SEARCH REPORT

Application Number
EP 13 30 6391

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	Jing Wang ET AL: "Context-based adaptive arithmetic coding in time and frequency domain for the lossless compression of audio coding parameters at variable rate", EURASIP Journal on Audio, Speech, and Music Processing, 21 May 2013 (2013-05-21), page 1, XP055104567, Retrieved from the Internet: URL:http://asmp.eurasipjournals.com/content/pdf/1687-4722-2013-9.pdf [retrieved on 2014-02-26] * page 3, paragraph 2. - page 5, paragraph 2.3.2 *	1-16	TECHNICAL FIELDS SEARCHED (IPC)
A	US 2009/083044 A1 (BRIAND MANUEL [FR] ET AL BRIAND MANUEL [SE] ET AL) 26 March 2009 (2009-03-26) * page 5, paragraph [0122] - page 7, paragraph [0178]; figures 7,9,13 * * page 8, paragraph [0210] *	1-16	
A	HIRVONEN TONI ET AL: "Perceptual Compression Methods for Metadata in Directional Audio Coding Applied to Audiovisual Teleconference", AES CONVENTION 126; MAY 2009, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, 7 May 2009 (2009-05-07), XP040508988, * page 2, paragraph 2. - page 5, paragraph 3.3; figure 1 *	1-16	
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 6 March 2014	Examiner Virette, David
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

EPO FORM 1503 03.82 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 13 30 6391

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

06-03-2014

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011249821 A1	13-10-2011	EP 2374123 A1	12-10-2011
		US 2011249821 A1	13-10-2011
		WO 2010070225 A1	24-06-2010

US 2007269063 A1	22-11-2007	NONE	

US 2009083044 A1	26-03-2009	AT 531036 T	15-11-2011
		CN 101401152 A	01-04-2009
		EP 2005420 A1	24-12-2008
		JP 5166292 B2	21-03-2013
		JP 2009530651 A	27-08-2009
		KR 20080104065 A	28-11-2008
		US 2009083044 A1	26-03-2009
		WO 2007104882 A1	20-09-2007

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- EP 12305537 A [0005] [0007] [0014]
- EP 12306569 A [0005] [0007] [0014] [0030] [0031]
- EP 13305558 A [0005] [0007] [0014] [0015] [0019]
[0021] [0030] [0031] [0034]
- EP 13305156 A [0031]