



(51) International Patent Classification:

H04S 3/00 (2006.01) *G01S 3/80* (2006.01)
G10L 19/008 (2013.01)

(21) International Application Number:

PCT/EP2014/052479

(22) International Filing Date:

7 February 2014 (07.02.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

13305156.5 8 February 2013 (08.02.2013) EP

(71) Applicant: THOMSON LICENSING [FR/FR]; 1-5 rue
Jeanne d'Arc, F-92130 Issy-les-Moulineaux (FR).(72) Inventors: KRUEGER, Alexander; Feuerbachstrasse 16,
30655 Hannover (DE). KORDON, Sven; Mühlenkamp-
straße 50 A, 31515 Wunstorf (DE).

(74) Agent: HARTNACK, Wolfgang; Deutsche Thomson
OHG, European Patent Operations, Karl-Wiechert-Allee
74, 30625 Hannover (DE).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, 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, LT, 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, 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,

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR DETERMINING DIRECTIONS OF UNCORRELATED SOUND SOURCES IN A
HIGHER ORDER AMBISONICS REPRESENTATION OF A SOUND FIELD

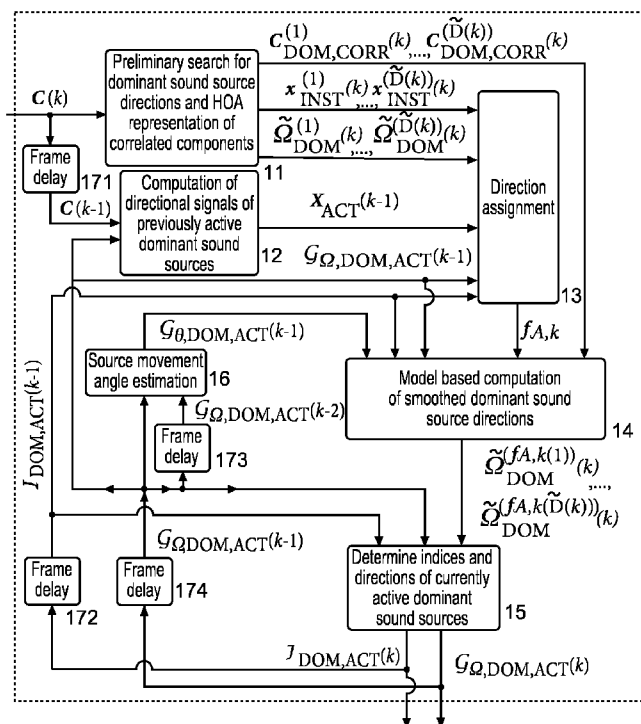


Fig. 1

(57) Abstract: Higher Order Ambisonics (HOA) represents three-dimensional sound. HOA provides high spatial resolution and facilitates analysing of the sound field with respect to dominant sound sources. The invention aims to identify independent dominant sound sources constituting the sound field, and to track their temporal trajectories. Known applications are searching for all potential candidates for dominant sound source directions by looking at the directional power distribution of the original HOA representation, whereas in the invention all components which are correlated with the signals of previously found sound sources are removed. By such operation the problem of erroneously detecting many instead of only one correct sound source can be avoided in case its contributions to the sound field are highly directionally dispersed.



MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, **Published:**
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, — *with international search report (Art. 21(3))*
GW, KM, ML, MR, NE, SN, TD, TG).

Method and Apparatus for determining directions of uncorrelated sound sources in a Higher Order Ambisonics representation of a sound field

5 The invention relates to a method and to an apparatus for determining directions of uncorrelated sound sources in a Higher Order Ambisonics representation of a sound field.

10 Background

Higher Order Ambisonics (HOA) offers one possibility to represent three-dimensional sound among other techniques like wave field synthesis (WFS) or channel based approaches like
15 22.2. In contrast to channel based methods, however, the HOA representation offers the advantage of being independent of a specific loudspeaker set-up. This flexibility, however, 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 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
20 binaural rendering to headphones.

HOA is based on a 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
30 represented by a time domain function. Hence, without loss of generality, the complete HOA sound field representation actually can be assumed to consist of O time domain functions, where O denotes the number of expansion coefficients. In the following, these time domain functions are referred

to as HOA coefficient sequences or as HOA channels.
HOA has the potential to provide a high spatial resolution,
which improves with a growing maximum order N of the expansion. It offers the possibility of analysing the sound field
5 with respect to dominant sound sources.

Invention

10 An application could be how to identify from a given HOA representation independent dominant sound sources constituting the sound field, and how to track their temporal trajectories. Such operations are required e.g. for the compression of HOA representations by decomposition of the sound
15 field into dominant directional signals and a remaining ambient component as described in patent application EP 12305537.8 . A further application for such direction tracking method would be a coarse preliminary source separation. It could also be possible to use the estimated direction
20 trajectories for the post-production of HOA sound field recordings in order to amplify or to attenuate the signals of particular sound sources.

In EP 12305537.8 it is proposed to successively perform the following three operations:

- 25 - The number of currently present dominant sound sources within a time frame is identified and the corresponding directions are searched for. The number of dominant sound sources is determined from the eigenvalues of the HOA channel cross-correlation matrix. For the search of the
30 dominant sound source directions the directional power distribution corresponding to a frame of HOA coefficients for a fixed high number of predefined test directions is evaluated. The first direction estimate is obtained by looking for the maximum in the directional power distribu-

tion. Then, the remaining identified directions are found by consecutively repeating the following two operations: the test directions in the spatial neighbourhood are eliminated from the remaining set of test directions and the
5 resulting set is considered for the search of the maximum of the directional power distribution.

- The estimated directions are assigned to the sound sources deemed to be active in the last time frame.
- Following the assignment, an appropriate smoothing of the
10 direction estimates is performed in order to obtain a temporally smooth direction trajectory.

However, although with such processing the temporal smoothing of the direction estimates is accomplished in principle by computing the exponentially-weighted moving average, this
15 technique has the disadvantage of not being able to accurately capture abrupt direction changes or onsets of new dominant sounds.

To overcome this problem, it was suggested in patent application EP 12306485.9 to introduce a simple statistical
20 source movement prediction model, which is employed for a statistically motivated smoothing implemented by the Bayesian learning rule. However, EP 12306485.9 and EP 12305537.8 compute the likelihood function for the sound source directions only from the directional power distribution. This
25 distribution represents the power of a high number of general plane waves from directions specified by nearly uniformly distributed sampling points on the unit sphere. It does not provide any information about the mutual correlation between general plane waves from different directions.
30 In practice, the order N of the HOA representation is usually limited, resulting in a spatially band-limited sound field. In particular, this means that the contribution of a directional sound source to the directional power distribution is smeared around the true direction of incidence to

directions in the neighbourhood. This smearing effect is mathematically described by a 'dispersion function', see below section *Spatial resolution of Higher Order Ambisonics*. Its extent grows with a decreasing order of the HOA representation. The EP 12306485.9 and EP 12305537.8 direction tracking methods, are considering this effect to a certain degree by constraining the search of directions to areas outside the neighbourhood of previously found directions. However, the specification of the neighbourhood assumes that all sound sources are encoded with the full order N of the HOA representation. This assumption is violated for HOA representations of order N which contain general plane waves encoded in a lower order than N . Such general plane waves of lower order than N may be the result of artistic creation in order to make sound sources appearing wider. However, they also occur with the recording of HOA sound field representations by spherical microphones.

The EP 12306485.9 and EP 12305537.8 direction tracking methods would identify more than a single sound source in case the sound field consists of a single general plane wave of lower order than N , which is an undesired property.

A problem to be solved by the invention is to improve the determination of dominant sound sources in an HOA sound field, such that their temporal trajectories can be tracked. This problem is solved by the methods disclosed in claims 1, 2 and 6. An apparatus that utilises the method of claim 6 is disclosed in claim 7.

The invention improves the EP 12306485.9 processing. The inventive processing looks for independent dominant sound sources and tracks their directions over time. The expression 'independent dominant sound sources' means that the signals of the respective sound sources are uncorrelated.

While the state-of-the-art methods EP 12305537.8 and EP 12306485.9 are searching for all potential candidates for dominant sound source directions by looking at the directional power distribution of the *original* HOA representation only, the inventive processing described below removes for the search of each direction candidate from the original HOA representation all the components which are correlated with the signals of previously found sound sources. By such operation the problem of erroneously detecting many instead of only one correct sound source can be avoided in case its contributions to the sound field are highly directionally dispersed. As mentioned above, such an effect would occur for HOA representations of order N which contain general plane waves encoded in an order lower than N .

Like in EP 12306485.9, the candidates found for the dominant sound source directions are then assigned to previously found dominant sound sources and are finally smoothed according to a statistical source movement model. Hence, like in EP 12306485.9 the inventive processing provides temporally smooth direction estimates, and is able to capture abrupt direction changes or onsets of new dominant sounds.

The inventive processing determines estimates of dominant sound source directions for successive frames of an HOA representation in two subsequent processings:

From a current time frame k of an HOA representation, candidates or estimates for dominant sound source directions are successively searched, and the components of the HOA representation, which are supposed to be created by the respective sound sources, are determined. In each iteration of this search process each further direction candidate is computed from a residual HOA representation which represents the original HOA representation from which all the components correlated with the signals of previously found sound

sources have been removed. The current direction candidate is selected out of a number of predefined test directions, such that the power of the related general plane wave of the residual HOA representation, impinging from the chosen direction on the listener position, is maximum compared to that of all other test directions.

Next, the selected direction candidates for the current time frame are assigned to dominant sound sources found in the previous time frame $k-1$ of HOA coefficients. Thereafter the final direction estimates, which are smoothed with respect to the resulting time trajectory, are computed by carrying out a Bayesian inference process, wherein this Bayesian inference process exploits on one hand a statistical a priori sound source movement model and, on the other hand, the directional power distributions of the dominant sound source components of the original HOA representation. That a priori sound source movement model statistically predicts the current movement of individual sound sources from their direction in the previous time frame $k-1$ and movement between the previous time frame $k-1$ and the penultimate time frame $k-2$.

The assignment of direction estimates to dominant sound sources found in the previous time frame $(k-1)$ of HOA coefficients is accomplished by a joint minimisation of the angles between pairs of a direction estimate and the direction of a previously found sound source, and maximisation of the absolute value of the correlation coefficient between the pairs of the directional signals related to a direction estimate and to a dominant sound source found in the previous time frame.

In principle, the inventive method is suited for determining directions of uncorrelated sound sources in a Higher Order

Ambisonics representation denoted HOA of a sound field, said method including the steps:

- in a current time frame of HOA coefficients, searching successively preliminary direction estimates of dominant sound sources, and computing HOA sound field components which are created by the corresponding dominant sound sources, and computing the corresponding directional signals;
 - assigning said computed dominant sound sources to corresponding sound sources active in the previous time frame of said HOA coefficients by comparing said preliminary direction estimates of said current time frame and smoothed directions of sound sources active in said previous time frame, and by correlating said directional signals of said current time frame and directional signals of sound sources active in said previous time frame, resulting in an assignment function;
 - computing smoothed dominant source directions using said assignment function, said set of smoothed directions in said previous time frame, a set of indices of active dominant sound sources in said previous time frame, a set of respective source movement angles between the penultimate time frame and said previous time frame, and said HOA sound field components created by the corresponding dominant sound sources;
 - determining indices and directions of the active dominant sound sources of said current time frame, using said smoothed dominant source directions, the frame delayed version of directions of the active dominant sound sources of said previous time frame and the frame delayed version of indices of the active dominant sound sources of said previous time frame,
- wherein said directional signals of sound sources active in said previous time frame are computed from said frame de-

layed version of directions of the active dominant sound sources of said previous time frame and the HOA coefficients of said previous time frame using mode matching, and wherein said set of source movement angles between said penultimate time frame and said previous time frame is computed from said frame delayed version of directions of the active dominant sound sources of said previous time frame and a further frame delayed version thereof.

10 In principle the inventive apparatus is suited for determining directions of uncorrelated sound sources in a Higher Order Ambisonics representation denoted HOA of a sound field, said apparatus including:

- 15 - means being adapted for searching successively in a current time frame of HOA coefficients preliminary direction estimates of dominant sound sources, and for computing HOA sound field components which are created by the corresponding dominant sound sources, and for computing the corresponding directional signals;
- 20 - means being adapted for assigning said computed dominant sound sources to corresponding sound sources active in the previous time frame of said HOA coefficients by comparing said preliminary direction estimates of said current time frame and smoothed directions of sound sources active in
25 said previous time frame, and by correlating said directional signals of said current time frame and directional signals of sound sources active in said previous time frame, resulting in an assignment function;
- 30 - means being adapted for computing smoothed dominant source directions using said assignment function, said set of smoothed directions in said previous time frame, a set of indices of active dominant sound sources in said previous time frame, a set of respective source movement angles between the penultimate time frame and said previous time

frame, and said HOA sound field components created by the corresponding dominant sound sources;

- means being adapted for determining indices and directions of the active dominant sound sources of said current

5 time frame, using said smoothed dominant source directions, the frame delayed version of directions of the active dominant sound sources of said previous time frame and the frame delayed version of indices of the active dominant sound sources of said previous time frame,

10 wherein said directional signals of sound sources active in said previous time frame are computed from said frame delayed version of directions of the active dominant sound sources of said previous time frame and the HOA coefficients of said previous time frame using mode matching,

15 and wherein said set of source movement angles between said penultimate time frame and said previous time frame is computed from said frame delayed version of directions of the active dominant sound sources of said previous time frame and a further frame delayed version thereof.

20

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

25 Drawings

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

Fig. 1 Block diagram of the inventive processing for estimation of the directions of dominant and uncorrelated directional signals of a Higher Order Ambisonics signal;

30

Fig. 2 Detail of preliminary direction estimation;

Fig. 3 Computation of dominant directional signal and HOA

representation of sound field produced by the dominant sound source;

Fig. 4 Model based computation of smoothed dominant sound source directions;

5 Fig. 5 Spherical coordinate system;

Fig. 6 Normalised dispersion function $v_N(\theta)$ for different Ambisonics orders N and for angles $\theta \in [0, \pi]$.

10 Exemplary embodiments

The principle of the inventive direction tracking processing is illustrated in Fig. 1 and is explained in the following. It is assumed that the direction tracking is based on the successive processing of input frames $\mathbf{C}(k)$ of HOA coefficient sequences of length L , where k denotes the frame index. The frames are defined with respect to the HOA coefficient sequences specified in equation (45) in section *Basics of Higher Order Ambisonics* as

$$20 \quad f\mathbf{C}(k) := [\mathbf{c}((kB + 1)T_S) \quad \mathbf{c}((kB + 2)T_S) \quad \dots \quad \mathbf{c}((kB + L)T_S)] , \quad (1)$$

where T_S denotes the sampling period and $B \leq L$ indicates the frame shift. It is reasonable, but not necessary, to assume that successive frames are overlapping, i.e. $B < L$.

25 In a first step or stage 11, the k -th frame $\mathbf{C}(k)$ of the HOA representation is preliminary analysed for dominant sound sources. A detailed description of this processing is provided in below section *Preliminary direction search*. In particular, the number $\tilde{D}(k)$ of detected dominant directional signals is determined as well as the corresponding $\tilde{D}(k)$ preliminary direction estimates $\tilde{\mathbf{Q}}_{\text{DOM}}^{(1)}(k), \dots, \tilde{\mathbf{Q}}_{\text{DOM}}^{(\tilde{D}(k))}(k)$. Additionally, the HOA sound field components $\mathbf{C}_{\text{DOM,CORR}}^{(d)}(k)$, $d = 1, \dots, \tilde{D}(k)$, which are (supposed to be) created by the corresponding individual

dominant sound sources as well as the corresponding instantaneous directional signals $\mathbf{x}_{\text{INST}}^{(d)}(k)$, $d = 1, \dots, \tilde{D}(k)$ (i.e. general plane wave functions) are computed.

The individual preliminary direction estimates and related quantities are computed in a sequential manner, i.e. first for $d = 1$, then for $d = 2$ and so on. In the first step the directional power distribution of the original HOA representation $\mathbf{C}(k)$ is computed as proposed in EP 12305537.8 and successively analysed for the presence of dominant sound sources. In the case that a dominant sound source is detected, the respective preliminary direction estimate $\tilde{\mathbf{n}}_{\text{DOM}}^{(1)}(k)$ is computed. Additionally, the corresponding directional signal $\mathbf{x}_{\text{INST}}^{(1)}(k)$ is estimated, together with that component $\mathbf{C}_{\text{DOM,CORR}}^{(1)}(k)$ of current frame $\mathbf{C}(k)$ which is assumed to be created by this sound source. It is assumed that $\mathbf{C}_{\text{DOM,CORR}}^{(1)}(k)$ represents that component of $\mathbf{C}(k)$ which is correlated with the directional signal $\mathbf{x}_{\text{INST}}^{(1)}(k)$. Finally, the HOA component $\mathbf{C}_{\text{DOM,CORR}}^{(1)}(k)$ is subtracted from $\mathbf{C}(k)$ in order to obtain the residual HOA representation $\mathbf{C}_{\text{REM}}^{(2)}(k)$. The estimation of the d -th ($d \geq 2$) preliminary direction is performed in a completely analogous way as that of the first one, with the only exception of using the residual HOA representation $\mathbf{C}_{\text{REM}}^{(d)}(k)$ instead of $\mathbf{C}(k)$. It is thereby explicitly assured that sound field components created by the found d -th sound source are excluded for the further direction search.

In direction assignment step or stage 13, the dominant sound sources found in step/stage 11 in the k -th frame are assigned to the corresponding sound sources (assumed to be) active in the $(k-1)$ -th frame. On one hand, the assignment is accomplished by comparing the preliminary direction esti-

mates $\tilde{\mathbf{n}}_{\text{DOM}}^{(1)}(k), \dots, \tilde{\mathbf{n}}_{\text{DOM}}^{(\tilde{D}(k))}(k)$ for the current frame (k) and the smoothed directions of sound sources (assumed to be) active in the $(k-1)$ -th frame, which are contained in the set $\mathcal{G}_{\mathbf{n},\text{DOM},\text{ACT}}(k-1)$ and whose indices are contained in the set $\mathcal{I}_{\text{DOM},\text{ACT}}(k-1)$. On the other hand, for the assignment the correlation between the instantaneous directional signals $\mathbf{x}_{\text{INST}}^{(d)}(k)$, $d = 1, \dots, \tilde{D}(k)$ of the detected dominant sound sources at frame k and the directional signals $\mathbf{X}_{\text{ACT}}(k-1)$ of sound sources (assumed to be) active in the $(k-1)$ -th frame is exploited. The result of the assignment is formulated by an assignment function $f_{\mathcal{A},k}: \{1, \dots, \tilde{D}(k)\} \rightarrow \{1, \dots, D\}$, where D denotes the maximum number of expected sound sources to be tracked, meaning that the d -th newly found sound source is assigned to the previously active sound source with index $f_{\mathcal{A},k}(d)$.

15

In a model based computation of smoothed dominant sound source directions step or stage 14 the smoothed dominant source directions $\hat{\mathbf{n}}_{\text{DOM}}^{(f_{\mathcal{A},k}(d))}(k)$, $d = 1, \dots, \tilde{D}(k)$ are computed, based on the statistical sound source movement model proposed in EP 12306485.9 by using the set $\mathcal{I}_{\text{DOM},\text{ACT}}(k-1)$ of the indices of active dominant sound sources at frame $(k-1)$, the set $\mathcal{G}_{\mathbf{n},\text{DOM},\text{ACT}}(k-1)$ of the corresponding dominant source direction estimates at frame $(k-1)$, the set $\mathcal{G}_{\theta,\text{DOM},\text{ACT}}(k-1)$ of the respective source movement angles between the frames $(k-2)$ and $(k-1)$, the HOA sound field components $\mathbf{c}_{\text{DOM,CORR}}^{(d)}(k)$, $d = 1, \dots, \tilde{D}(k)$ which are supposed to be created by the the found dominant sound sources, and the assignment function $f_{\mathcal{A},k}$. A detailed description of this model based smoothing procedure is provided in below section *Model based computation of smoothed dominant sound source directions*.

30

In a last step or stage 15, the indices and the directions of the currently active dominant sound sources are determined, which are supposed to be contained in the sets $\mathcal{I}_{\text{DOM,ACT}}(k)$ and $\mathcal{G}_{\Omega,\text{DOM,ACT}}(k)$, respectively, using the smoothed dominant source directions $\hat{\Omega}_{\text{DOM}}^{(f_{\mathcal{A},k(d)})}(k)$, $d = 1, \dots, \tilde{D}(k)$ from step /stage 14 and the sets $\mathcal{G}_{\Omega,\text{DOM,ACT}}(k-1)$ and $\mathcal{I}_{\text{DOM,ACT}}(k-1)$ containing the smoothed directions and respective indices of sound sources assumed to be active in the $(k-1)$ -th frame. This operation has the purpose to not spuriously deactivate sound sources which have not been detected for a small number of successive frames.

Step or stage 12 performs the computation of the directional signals of sound sources supposed to be active in the $(k-1)$ -th frame using the HOA representation $\mathbf{C}(k-1)$ of frame $k-1$ and the set $\mathcal{G}_{\Omega,\text{DOM,ACT}}(k-1)$ of smoothed directions of sound sources supposed to be active in the $(k-1)$ -th frame. The computation is based on the principle of mode matching as described in M.A. Poletti, "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics", J. Audio Eng. Soc., vol.53(11), pp.1004-1025, 2005.

In a source movement angle estimation step or stage 16, the set $\mathcal{G}_{\Theta,\text{DOM,ACT}}(k-1)$ of movement angles of the dominant active sound sources at frame $k-1$ is computed from the two sets $\mathcal{G}_{\Omega,\text{DOM,ACT}}(k-1)$ and $\mathcal{G}_{\Omega,\text{DOM,ACT}}(k-2)$ of smoothed direction estimates of sound sources supposed to be active in the $(k-1)$ -th and $(k-2)$ -th frame, respectively. The movement is understood to happen between frames $k-2$ and $k-1$. The movement angle of an active dominant sound source is the arc between its smoothed direction estimate at frame $k-2$ and that at frame $k-1$.

Remarks: if no direction estimate for frame $k-2$ is available for a dominant sound source which is assumed to be active in frame $k-1$, the respective movement angle can be set to a maximum value of ' π '. In general, when initialising the processing for a first frame k and frame $k-1$ values are not yet available, the corresponding sets or values to be input in the steps or stages of Fig. 1 are empty or set to zero, respectively.

10 This operation causes the a-priori probability for the next direction of this sound source to become nearly uniform over all possible directions, cf. below section *Determine indices and directions of currently active dominant sound sources*.

15 Frame delays 171 to 174 are delaying the respective signals by one frame.

In the following, the above-mentioned steps and stages are explained in more detail.

20 *Preliminary direction search*

In the preliminary direction search step/stage 11, the current number $\tilde{D}(k)$ of present dominant sound sources (in frame k) and the respective directions $\tilde{\Omega}_{\text{DOM}}^{(d)}(k)$, $d = 1, \dots, \tilde{D}(k)$, are estimated. Additionally, the HOA sound field components $\mathbf{c}_{\text{DOM,CORR}}^{(d)}(k)$, $d = 1, \dots, \tilde{D}(k)$ which are supposed to be created by the individual sound sources, as well as the corresponding directional signals $\mathbf{x}_{\text{INST}}^{(d)}(k)$, $d = 1, \dots, \tilde{D}(k)$ (i.e. general plane wave functions) are computed. All the previously enumerated quantities are computed first for direction index $d=1$, then
 25 for $d=2$ and so on until $d = \tilde{D}(k)$.
 30

The computation procedure for a single direction d index is illustrated in Fig. 2. The remaining HOA representation

$\mathbf{C}_{\text{REM}}^{(d)}(k)$ produced after the estimation of the $(d-1)$ -th direction (related to the estimation of the d -th direction for the k -th time frame) is input to this stage. It is thereby understood that in the beginning of the loop $\mathbf{C}_{\text{REM}}^{(1)}(k)$ corresponds to the original HOA frame $\mathbf{C}(k)$. In a first step or stage 21, the directional power distribution $\mathbf{p}^{(d)}(k)$ of the remaining HOA representation $\mathbf{C}_{\text{REM}}^{(d)}(k)$ is computed for a predefined number of Q discrete test directions $\boldsymbol{\Omega}_q$, $q = 1, \dots, Q$, which are nearly uniformly distributed on the unit sphere. To be more specific, each test direction $\boldsymbol{\Omega}_q$ is defined as a vector containing an inclination angle $\theta_q \in [0, \pi]$ and azimuth angle $\phi_q \in [0, 2\pi[$ according to $\boldsymbol{\Omega}_q := (\theta_q, \phi_q)^T$, (2) where $(\cdot)^T$ denotes transposition. The directional power distribution is represented by the vector

$$\mathbf{p}^{(d)}(k) := \left(p_1^{(d)}(k), \dots, p_Q^{(d)}(k) \right)^T, \quad (3)$$

whose components $p_q^{(d)}(k)$ denote the joint power of all dominant sound sources remaining in the representation $\mathbf{C}_{\text{REM}}^{(d)}(k)$ related to the direction $\boldsymbol{\Omega}_q$ for the k -th time frame. The actual computation of the directional power distribution $\mathbf{p}^{(d)}(k)$ from $\mathbf{C}_{\text{REM}}^{(d)}(k)$ may be performed as proposed in EP 12305537.8. In step or stage 22, the directional power distribution $\mathbf{p}^{(d)}(k)$ is analysed for the presence of a dominant sound source. One way of detecting a dominant source is described in below section *Analysis for dominant sound source presence*. If the absence of a dominant sound source is detected, then the direction search is stopped and the total number of found dominant directions is set to $\tilde{D}(k) = d - 1$. Otherwise, if a dominant source is detected, a preliminary estimate of its direction $\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k)$ with respect to the coordinate origin is computed in step or stage 23, see below section *Search*

for dominant sound source direction for details.

Successively, the respective directional signal $\mathbf{x}_{\text{INST}}^{(d)}(k)$ and the HOA representation $\mathbf{C}_{\text{DOM,CORR}}^{(d)}(k)$ of the sound field component assumed to be created by the d -th dominant sound source are computed in step or stage 24 as described in more detail in below section *Computation of dominant directional signal and HOA representation of sound field produced by the dominant sound source*.

Finally, in step or stage 25 the HOA component $\mathbf{C}_{\text{DOM,CORR}}^{(d)}(k)$ is subtracted from $\mathbf{C}_{\text{REM}}^{(d)}(k)$ in order to obtain the residual HOA representation $\mathbf{C}_{\text{REM}}^{(d+1)}(k)$, which is used for the search of the next (i.e. $(d+1)$ -th) directional sound source. It is thereby explicitly assured that sound field components created by the d -th sound source found are excluded for the further direction search.

- Analysis for dominant sound source presence

For detecting the presence of a dominant sound source within the sound field represented by $\mathbf{C}_{\text{REM}}^{(d)}(k)$, the directional power distributions $\mathbf{p}^{(1)}(k), \dots, \mathbf{p}^{(d)}(k)$ of the remaining HOA representations $\mathbf{C}_{\text{REM}}^{(1)}(k), \dots, \mathbf{C}_{\text{REM}}^{(d)}(k)$ are considered. On one hand, it has been experimentally found that it is reasonable to monitor

$$\text{the variance ratio } \delta_{\mathbf{p}}^{(d)}(k) := \frac{\text{var}(\mathbf{p}^{(d)}(k))}{\text{var}(\mathbf{p}^{(1)}(k))} , \quad (4)$$

which can be regarded as a measure for the importance of the sound field represented by the remaining HOA representation $\mathbf{C}_{\text{REM}}^{(d)}(k)$ compared to the sound field represented by the initial HOA representation $\mathbf{C}(k)$. A small ratio $\delta_{\mathbf{p}}^{(d)}(k)$ indicates that none of the sound sources represented by the HOA representation $\mathbf{C}_{\text{REM}}^{(d)}(k)$ should be considered as being dominant.

On the other hand, it is also reasonable to watch the ratio

$$\delta_{\mathbf{p},\text{NORM}}^{(d)}(k) := \frac{\text{var}(\mathbf{p}_{\text{NORM}}^{(d)}(k))}{\text{var}(\mathbf{p}_{\text{NORM}}^{(d-1)}(k))}, \quad \text{for } d \geq 2, \quad (5)$$

of the variances of the normalised directional power distributions $\mathbf{p}_{\text{NORM}}^{(d)}(k)$ and $\mathbf{p}_{\text{NORM}}^{(d-1)}(k)$. The elements $p_{q,\text{NORM}}^{(d)}(k)$, $q = 1, \dots, Q$, of the normalised directional power distribution

$$\mathbf{p}_{\text{NORM}}^{(d)}(k) := \left(p_{1,\text{NORM}}^{(d)}(k), p_{2,\text{NORM}}^{(d)}(k), \dots, p_{Q,\text{NORM}}^{(d)}(k) \right)^T \quad (6)$$

are defined in dependence of those of $\mathbf{p}^{(d)}(k)$ by

$$p_{q,\text{NORM}}^{(d)}(k) := \frac{p_q^{(d)}(k)}{\sum_{q'=1}^Q p_{q'}^{(d)}(k)}. \quad (7)$$

The variance $\text{var}(\mathbf{p}_{\text{NORM}}^{(d)}(k))$ can be regarded as a measure of the uniformity of the directional power distribution $\mathbf{p}^{(d)}(k)$. In particular, the variance is the smaller the more uniform the power is distributed over all directions of incidence. In the limiting case of a spatially diffuse noise, the variance $\text{var}(\mathbf{p}_{\text{NORM}}^{(d)}(k))$ should approach a value of zero. Based on these considerations, the variance ratio $\delta_{\mathbf{p},\text{NORM}}^{(d)}(k)$ indicates whether the directional power of the HOA representation $\mathbf{C}_{\text{REM}}^{(d)}(k)$ is distributed more uniformly than that of $\mathbf{C}_{\text{REM}}^{(d-1)}(k)$.

To summarise the above considerations, it can be assumed that there is always at least a single dominant sound source present in the sound field represented by $\mathbf{C}(k)$, i.e. $\tilde{D}(k) \geq 1$. Further dominant sources are detected (for $d \geq 2$) if the value of the variance ratio $\delta_{\mathbf{p}}^{(d)}(k)$ remains above a certain predefined threshold $\varepsilon_p < 1$ and the value of the variance ratio is smaller than one, i.e. Dominant sound source is detected (for $d \geq 2$) if $\delta_{\mathbf{p}}^{(d)}(k) \geq \varepsilon_p$ and $\delta_{\mathbf{p},\text{NORM}}^{(d)}(k) < 1$. (8)

The value for ε_p is to be set with respect to the interpretation of what 'dominant' means. The inventors have found that a reasonable choice is given by $\varepsilon_p = 10^{-3}$.

- Search for dominant sound source direction

After the d -th sound source has been detected, a preliminary estimate of its direction $\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k)$ is searched for by employing the directional power distribution $\mathbf{p}^{(d)}(k)$. The search is accomplished by taking that test direction $\boldsymbol{\Omega}_q$ for which the directional power is the largest, i.e.

$$\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k) = \boldsymbol{\Omega}_{q_{\text{MAX}}^{(k,d)}}, \text{ where } q_{\text{MAX}}^{(k,d)} := \operatorname{argmax}_{1 \leq q \leq Q} p_q^{(d)}(k) . \quad (9)$$

- Computation of dominant directional signal and HOA representation of sound field produced by the dominant sound source

Subsequently, after having determined a preliminary estimate $\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k)$ of the dominant source direction, the respective directional signal $\mathbf{x}_{\text{INST}}^{(d)}(k)$, as well as the HOA representation $\mathbf{c}_{\text{DOM,CORR}}^{(d)}(k)$ of the sound field components assumed to be created by the same sound source, are computed according to Fig. 3. In step or stage 31, a fixed predefined spherical grid $\mathcal{G}_{\boldsymbol{\Omega},\text{INIT}}$ consisting of O sampling positions $\boldsymbol{\Omega}_{\text{INIT},o}$, $o = 1, \dots, O$, which are assumed to be nearly uniformly distributed on the unit sphere, is rotated to provide the grid $\mathcal{G}_{\boldsymbol{\Omega},\text{ROT}}^{(d)}(k)$ consisting of the rotated sampling positions $\boldsymbol{\Omega}_{\text{ROT},o}^{(d)}(k)$, $o = 1, \dots, O$. The rotation is performed such that the first rotated sampling position $\boldsymbol{\Omega}_{\text{ROT},1}^{(d)}(k)$ corresponds to the preliminary direction estimate $\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k)$.

In step or stage 32, the HOA representation $\mathbf{c}_{\text{REM}}^{(d)}(k)$ is transformed to the so-called spatial domain, where it is equivalently represented by O plane wave functions (also referred to as grid directional signals) $\mathbf{x}_{o,\text{INST}}^{(d)}(k)$, $o = 1, \dots, O$, which are assumed to imping on the observer position (i.e. the coordinate origin) from the rotated grid directions $\boldsymbol{\Omega}_{\text{ROT},o}^{(d)}(k)$, $o = 1, \dots, O$. To compute the plane wave functions $\mathbf{x}_{o,\text{INST}}^{(d)}(k)$, $o = 1, \dots, O$, the

mode matrix $\mathbf{E}_{\text{GRID}}^{(d)}(k)$ with respect to the rotated grid directions is computed as

$$\mathbf{E}_{\text{GRID}}^{(d)}(k) := [\mathbf{S}_{\text{GRID},1}^{(d)}(k) \quad \mathbf{S}_{\text{GRID},2}^{(d)}(k) \quad \dots \quad \mathbf{S}_{\text{GRID},O}^{(d)}(k)] \in \mathbb{R}^{O \times O} \quad (10)$$

with $\mathbf{S}_{\text{GRID},o}^{(d)}(k) :=$

$$5 \quad \left[S_0^0(\boldsymbol{\Omega}_{\text{ROT},o}^{(d)}(k)), S_1^{-1}(\boldsymbol{\Omega}_{\text{ROT},o}^{(d)}(k)), S_1^0(\boldsymbol{\Omega}_{\text{ROT},o}^{(d)}(k)), \dots, S_N^N(\boldsymbol{\Omega}_{\text{ROT},o}^{(d)}(k)) \right]^T \in \mathbb{R}^O. \quad (11)$$

Assuming each grid directional signal $\mathbf{x}_{o,\text{INST}}^{(d)}(k)$ to be a row vector composed of the individual samples of the k -th time frame as

$$\mathbf{x}_{o,\text{INST}}^{(d)}(k) = (x_{o,\text{INST}}^{(d)}(k, 1), x_{o,\text{INST}}^{(d)}(k, 2), \dots, x_{o,\text{INST}}^{(d)}(k, L)), \quad (12)$$

10 where L denotes the length (in samples) of the analysed HOA representation, the computation of all grid directional signals is accomplished by a Spherical Harmonics Transform (see below section *Spherical Harmonic Transform* for an explanation) as

$$15 \quad \begin{bmatrix} \mathbf{x}_{1,\text{INST}}^{(d)}(k) \\ \mathbf{x}_{2,\text{INST}}^{(d)}(k) \\ \vdots \\ \mathbf{x}_{O,\text{INST}}^{(d)}(k) \end{bmatrix} = \left(\mathbf{E}_{\text{GRID}}^{(d)}(k) \right)^{-1} \mathbf{C}(k). \quad (13)$$

Since the preliminary estimate $\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k)$ of the dominant sound source direction corresponds to the rotated sampling position $\boldsymbol{\Omega}_{\text{ROT},1}^{(d)}(k)$, the general plane wave function $\mathbf{x}_{1,\text{INST}}^{(d)}(k)$ can be regarded as the desired dominant directional signal $\mathbf{x}_{\text{INST}}^{(d)}(k)$,

$$20 \quad \text{i.e. } \mathbf{x}_{\text{INST}}^{(d)}(k) = \mathbf{x}_{1,\text{INST}}^{(d)}(k). \quad (14)$$

To determine that component of $\mathbf{C}_{\text{REM}}^{(d)}(k)$ which is produced by the d -th sound source, it is postulated that this component is equivalently represented by plane wave functions that can be predicted from $\mathbf{x}_{\text{INST}}^{(d)}(k)$ in step or stage 33. Hence, the

25 grid directional signals $\mathbf{x}_{o,\text{INST}}^{(d)}(k)$, $o = 2, \dots, O$ are attempted to be predicted from $\mathbf{x}_{\text{INST}}^{(d)}(k)$. The predicted signals are denoted by $\hat{\mathbf{x}}_{o,\text{INST}}^{(d)}(k)$, $o = 2, \dots, O$.

One way of accomplishing such prediction is to assume the predicted signals $\hat{\mathbf{x}}_{o,\text{INST}}^{(d)}(k)$, $o = 2, \dots, O$, to be created from $\mathbf{x}_{\text{INST}}^{(d)}(k)$ by linear filtering where the filters are determined so as to minimise the prediction error. If the filters are assumed to be finite impulse response (FIR) filters of a very short duration (compared to that of the analysis frame), the minimisation of the prediction error can be achieved by using state-of-the-art least squares techniques. Finally, the HOA representation of the dominant sound source signal $\mathbf{x}_{\text{INST}}^{(d)}(k)$ and all predicted correlated components is obtained in step or stage 34 by an inverse Spherical Harmonics Transform (see below section *Spherical Harmonic Transform* for an explanation) as

$$\mathbf{C}_{\text{DOM,CORR}}^{(d)}(k) = \mathbf{E}_{\text{GRID}}^{(d)}(k) \begin{bmatrix} \mathbf{x}_{\text{INST}}^{(d)}(k) \\ \hat{\mathbf{x}}_{2,\text{INST}}^{(d)}(k) \\ \hat{\mathbf{x}}_{3,\text{INST}}^{(d)}(k) \\ \hat{\mathbf{x}}_{O,\text{INST}}^{(d)}(k) \end{bmatrix}. \quad (15)$$

15

Computation of directional signals of previously active dominant sound sources

The directional signals $\mathbf{x}_{\text{ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1)$ of sound sources supposed to be active in the $(k-1)$ -th frame are contained within matrix $\mathbf{X}_{\text{ACT}}(k-1)$ according to equation (20). This matrix is computed using the principle of mode matching (see the above-mentioned Poletti article) by

$$\mathbf{X}_{\text{ACT}}(k-1) = (\mathbf{E}_{\text{ACT}}(k-1))^{-1} \mathbf{C}(k-1), \quad (16)$$

where $\mathbf{C}(k-1)$ denotes the $(k-1)$ -th frame of the original HOA sound field representation and $\mathbf{E}_{\text{ACT}}(k-1)$ denotes the mode matrix with respect to the directions $\bar{\mathbf{n}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1)$, $d' = 1, \dots, D_{\text{ACT}}(k-1)$, of sound sources supposed to be active in the $(k-1)$ -th frame. The mode matrix $\mathbf{E}_{\text{ACT}}(k-1)$ is computed by

$$\Xi_{\text{ACT}}(k-1) :=$$

$$[\mathbf{S}_{\text{ACT},1}(k-1), \mathbf{S}_{\text{ACT},2}(k-1), \dots, \mathbf{S}_{\text{ACT},D_{\text{ACT}}(k-1)}(k-1)] \in \mathbb{R}^{O \times D_{\text{ACT}}(k-1)} \quad (17)$$

$$\text{with } \mathbf{S}_{\text{ACT},d'}(k) :=$$

$$\left[S_0^0 \left(\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1) \right), S_1^{-1} \left(\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1) \right), S_1^1 \left(\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1) \right), \right. \\ \left. S_N^{N-1} \left(\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1) \right), S_N^N \left(\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1) \right) \right]^T \in \mathbb{R}^O \quad (18)$$

5

Direction assignment

As previously mentioned, on one hand the assignment in step/stage 13 of Fig. 1 is accomplished by comparing the

preliminary direction estimates $\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(1)}(k), \dots, \tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(\tilde{D}(k))}(k)$ and the
10 smoothed directions of sound sources supposed to be active in the $(k-1)$ -th frame, which are contained in the set

$$\mathcal{G}_{\boldsymbol{\Omega}, \text{DOM,ACT}}(k-1) := \left\{ \bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(1))}(k-1), \dots, \bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(D_{\text{ACT}}(k-1)))}(k-1) \right\}, \quad (19)$$

where $i_{\text{ACT},k-1}(d')$ denotes the index of the d' -th sound source assumed to be active in the $(k-1)$ -th frame. In particular,

15 it is assumed that the smaller the angle

$$\angle \left(\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k), \bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1) \right)$$

between a pair of a preliminary direction estimate $\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k)$

and a smoothed direction $\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1)$, the more likely the d -th newly found dominant sound source direction will correspond to the previously active sound source with index

20 $i_{\text{ACT},k-1}(d')$.

On the other hand, for the assignment the correlation between

the instantaneous directional signals $\mathbf{x}_{\text{INST}}^{(d)}(k)$, $d =$

$1, \dots, \tilde{D}(k)$ of the detected dominant sound sources at frame k

and the directional signals $\mathbf{X}_{\text{ACT}}(k-1)$ of sound sources supposed

25 to be active in the $(k-1)$ -th frame is exploited. It

is here assumed that the frame $\mathbf{X}_{\text{ACT}}(k-1)$ is composed of the individual directional signals $\mathbf{x}_{\text{ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1)$ of sound sources supposed to be active in the $(k-1)$ -th frame as

$$\mathbf{X}_{\text{ACT}}(k-1) = \begin{bmatrix} \mathbf{x}_{\text{ACT}}^{(i_{\text{ACT},k-1}(1))}(k-1) \\ \mathbf{x}_{\text{ACT}}^{(i_{\text{ACT},k-1}(2))}(k-1) \\ \vdots \\ \mathbf{x}_{\text{ACT}}^{(i_{\text{ACT},k-1}(D_{\text{ACT}}(k-1)))}(k-1) \end{bmatrix}. \quad (20)$$

- 5 Using this definition, it is postulated that the higher the absolute value of the correlation coefficient

$$\rho_{\text{CORR}} \left(\mathbf{x}_{\text{INST}}^{(d)}(k), \mathbf{x}_{\text{ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1) \right)$$

- between the two signals $\mathbf{x}_{\text{INST}}^{(d)}(k)$ and $\mathbf{x}_{\text{ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1)$ is, the more likely the d -th newly found dominant sound source direction will correspond to the previously active sound source with index $i_{\text{ACT},k-1}(d')$. Such postulation is justified by the fact that the correlation coefficient provides a measure for the linear dependency between two signals.

Based on these considerations, an assignment function

$$f_{\mathcal{A},k}: \{1, \dots, \tilde{D}(k)\} \rightarrow \{1, \dots, D\}$$

- specifying the assignment is computed such as to minimise the following cost function (21)

$$\sum_{d=1}^{\tilde{D}(k)} \left[\angle \left(\tilde{\mathbf{n}}_{\text{DOM}}^{(d)}(k), \bar{\mathbf{n}}_{\text{DOM,ACT}}^{(f_{\mathcal{A},k}(d))}(k-1) \right) \right] \cdot \left[1 - \left| \rho_{\text{CORR}} \left(\mathbf{x}_{\text{INST}}^{(d)}(k), \mathbf{x}_{\text{ACT}}^{(f_{\mathcal{A},k}(d))}(k-1) \right) \right| \right].$$

It is implicitly assumed that for the direction indices $d'' \in \{1, \dots, D\} \setminus \mathcal{J}_{\text{DOM,ACT}}(k-1)$, which do not belong to any active sound source in the $(k-1)$ -th frame, the angles

$$\angle \left(\tilde{\mathbf{n}}_{\text{DOM}}^{(d)}(k), \bar{\mathbf{n}}_{\text{DOM,ACT}}^{(d'')}(k-1) \right)$$

- 20 are virtually set to a minimum angle of θ_{MIN} , where e.g. $\theta_{\text{MIN}} = 2\pi/N$. Further, the correlation coefficients

$$\rho_{\text{CORR}} \left(\mathbf{x}_{\text{INST}}^{(d)}(k), \mathbf{x}_{\text{ACT}}^{(d'')}(k-1) \right)$$

for the direction indices $d'' \in \{1, \dots, D\} \setminus \mathcal{I}_{\text{DOM,ACT}}(k-1)$ are virtually set to zero. The first operation has the effect that, if the angles between the d -th newly found direction $\tilde{\Omega}_{\text{DOM}}^{(d)}(k)$ and the directions of all previously active dominant sound sources are greater than θ_{MIN} , this newly found direction is favoured to belong to a new sound source.

The assignment problem can be solved by using the well-known Hungarian algorithm described in H.W. Kuhn, "The Hungarian method for the assignment problem", Naval research logistics quarterly, vol.2(1-2), pp.83-97, 1955.

Model based computation of smoothed dominant sound source directions

This section addresses the computation of the smoothed dominant sound source directions in step/stage 14 of Fig. 1 according to a statistical sound source movement model. The individual steps for this computation are illustrated in Fig. 4 and are explained in detail in the following.

20 - Computation of directional a priori probability functions for dominant sound source directions

The directional a priori probability functions $P_{\text{PRIO}}^{(f_{\mathcal{A},k}(d))}(k)$, $d = 1, \dots, \tilde{D}(k)$, for the newly found dominant sound source directions are computed in step or stage 42 using:

- 25 - the set $\mathcal{I}_{\text{DOM,ACT}}(k-1)$ of the indices $i_{\text{ACT},k-1}(d')$, $d' = 1, \dots, D_{\text{ACT}}(k-1)$, of active dominant sound sources at frame $(k-1)$,
- the set $\mathcal{G}_{\Omega,\text{DOM,ACT}}(k-1)$ of the corresponding dominant source direction estimates $\tilde{\Omega}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1)$, $d' = 1, \dots, D_{\text{ACT}}(k-1)$, at frame $(k-1)$,
- 30 - the set $\mathcal{G}_{\hat{\theta},\text{DOM,ACT}}(k-1)$ of the respective source movement angles $\hat{\theta}_{i_{\text{ACT},k-1}(d')}(k-1)$, $d' = 1, \dots, D_{\text{ACT}}(k-1)$, between the frame $(k-2)$ and $(k-1)$,

- and the assignment function $f_{\mathcal{A},k}$.

The computation is based on a simple sound source movement prediction model introduced in EP 12306485.9. In particular,

the directional a priori probability function $\mathbf{P}_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k)$ for
 5 the d -th newly found dominant sound source is assumed to be a discrete version of the von Mises-Fisher distribution on the unit sphere in the three-dimensional space.
 In the following it is assumed that the directional a priori probability function $\mathbf{P}_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k)$ is given by a vector composed
 10 of the probabilities $P_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q)$ for the individual test directions $\boldsymbol{\Omega}_q$, $q = 1, \dots, Q$, as $\mathbf{P}_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k) :=$

$$\left[P_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_1) \quad P_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_2) \quad \dots \quad P_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_Q) \right]^T \in \mathbb{R}^Q . \quad (22)$$

To compute the a priori probabilities for the individual
 15 test directions $\boldsymbol{\Omega}_q$ two cases are to be distinguished:

a) If the source index $f_{\mathcal{A},k(d)}$ assigned to the d -th newly found dominant sound source is contained within the set $\mathcal{I}_{\text{DOM,ACT}}(k-1)$, the a priori probabilities are computed according to

$$20 \quad P_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q) = \frac{\kappa_d(k)}{Q \sinh(\kappa_d(k))} \exp \left\{ \kappa_d(k) \cos(\theta_{q,d}(k)) \right\} \text{ for } q = 1, \dots, Q , \quad (23)$$

where $\theta_{q,d}(k)$ denotes the angle between the estimated direction $\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(f_{\mathcal{A},k(d)})}(k-1)$ and the test direction $\boldsymbol{\Omega}_q$, i.e.

$$\theta_{q,d}(k) := \angle \left(\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(f_{\mathcal{A},k(d)})}(k-1), \boldsymbol{\Omega}_q \right) . \quad (24)$$

Further, $\kappa_d(k)$ denotes a concentration parameter that is computed using the source movement angle estimate $\hat{\theta}_{f_{\mathcal{A},k(d)}}(k-1)$
 25

$$\text{according to } \kappa_d(k) = \frac{\ln(C_R)}{\cos(\hat{\theta}_{f_{\mathcal{A},k(d)}}(k-1)) - 1 - C_D} , \quad (25)$$

where C_D may be set to $C_D = \frac{\ln(C_R)}{-\kappa_{MAX}}$. (26)

Reasonable values for the parameters κ_{MAX} and C_R have been found to be (see EP 12306485.9) $\kappa_{MAX} = 8$, $C_R = 0.5$. (27)

The principle behind this computation is to increase the concentration of the a priori probability function the less the sound source has moved before. If the sound source has moved a lot before, the uncertainty about its successive direction is high and thus the concentration parameter has to achieve a small value.

- 10 b) If the source index $f_{\mathcal{A},k}(d)$ assigned to the d -th newly found dominant sound source is *not* contained within the set $\mathcal{I}_{DOM,ACT}(k-1)$, then the respective sound source is considered to not having been active before. Consequently, no a priori knowledge about the direction of this source is actually
- 15 available. Hence, the a priori probability function $P_{PRIO}^{(f_{\mathcal{A},k}(d))}(k)$ is assumed to be uniform on the unit sphere, where the individual probabilities are equal for all test positions $\boldsymbol{\Omega}_q$, i.e. $P_{PRIO}^{(f_{\mathcal{A},k}(d))}(k, \boldsymbol{\Omega}_q) = \frac{1}{Q}$ for $q = 1, \dots, Q$. (28)

- 20 - *Computation of directional likelihood functions for dominant sound source directions*

The directional likelihood functions $L^{(f_{\mathcal{A},k}(d))}(k)$, $d = 1, \dots, \tilde{D}(k)$, are computed in step or stage 41 using the HOA sound field components $C_{DOM,CORR}^{(d)}(k)$, $d = 1, \dots, \tilde{D}(k)$, which are supposed to be

25 created by the individual newly detected dominant sound sources, as well as the assignment function $f_{\mathcal{A},k}$. The directional likelihood function $L^{(f_{\mathcal{A},k}(d))}(k)$ is assumed to be a vector composed of the likelihoods $L^{(f_{\mathcal{A},k}(d))}(k, \boldsymbol{\Omega}_q)$ for the individual test directions $\boldsymbol{\Omega}_q$, $q = 1, \dots, Q$, as (29)

30 $L^{(f_{\mathcal{A},k}(d))}(k) := \begin{bmatrix} L^{(f_{\mathcal{A},k}(d))}(k, \boldsymbol{\Omega}_1) & L^{(f_{\mathcal{A},k}(d))}(k, \boldsymbol{\Omega}_2) & \dots & L^{(f_{\mathcal{A},k}(d))}(k, \boldsymbol{\Omega}_Q) \end{bmatrix}^T \in \mathbb{R}^Q$.

The individual likelihoods $L^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q)$ are computed to be approximations of the powers of general plane waves impinging from the test direction $\boldsymbol{\Omega}_q$, as described in EP 12305537.8. In particular,

$$L^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q) = (\mathbf{S}_{\text{TEST},q})^T \boldsymbol{\Sigma}_{\text{DOM,CORR}}^{(d)}(k) \mathbf{S}_{\text{TEST},q} \quad \text{for } q = 1, \dots, Q, \quad (30)$$

where $\mathbf{S}_{\text{TEST},q} :=$

$$[S_0^0(\boldsymbol{\Omega}_q), S_1^{-1}(\boldsymbol{\Omega}_q), S_1^0(\boldsymbol{\Omega}_q), S_1^1(\boldsymbol{\Omega}_q), \dots, S_N^{N-1}(\boldsymbol{\Omega}_q), S_N^N(\boldsymbol{\Omega}_q)]^T \in \mathbb{R}^O \quad (31)$$

denotes the mode vector with respect to the test direction $\boldsymbol{\Omega}_q$ (with $S_n^m(\cdot)$ representing the real valued Spherical Harmonics defined in below section *Definition of real valued Spherical Harmonics*) and where

$$\boldsymbol{\Sigma}_{\text{DOM,CORR}}^{(d)}(k) := \mathbf{C}_{\text{DOM,CORR}}^{(d)}(k) \left(\mathbf{C}_{\text{DOM,CORR}}^{(d)}(k) \right)^T \quad (32)$$

indicates the HOA inter-coefficients correlation matrix with respect to the HOA representation $\mathbf{C}_{\text{DOM,CORR}}^{(d)}(k)$.

15

- *Computation of directional a posteriori probability functions for dominant sound source directions*

The directional a posteriori probability functions

$\mathbf{P}_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k)$, $d = 1, \dots, \tilde{D}(k)$, are computed in step or stage 43 using the directional a priori probability functions

$\mathbf{P}_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k)$, $d = 1, \dots, \tilde{D}(k)$ and the directional likelihood functions $\mathbf{L}^{(f_{\mathcal{A},k(d)})}(k)$, $d = 1, \dots, \tilde{D}(k)$. Here, once again, the direc-

tional a posteriori probability function $\mathbf{P}_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k)$ is assumed to be a vector composed of the a posteriori probabili-

ties $P_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q)$ for the individual test directions $\boldsymbol{\Omega}_q$,

$q = 1, \dots, Q$, as $\mathbf{P}_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k) :=$

$$\left[P_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_1) \ P_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_2) \ \dots \ P_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_Q) \right]^T \in \mathbb{R}^Q. \quad (33)$$

The individual a posteriori probabilities $P_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q)$ are

computed according to the Bayesian rule (see EP 12306485.9) as

$$P_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q) = \frac{P_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q) L^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q)}{\sum_{q'=1}^Q P_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_{q'}) L^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_{q'})} \quad \text{for } q = 1, \dots, Q \quad . \quad (34)$$

Assuming a fixed direction index d the denominator of equation (37) is constant for each test direction $\boldsymbol{\Omega}_q$. For the purpose of the following direction search, where only the maximum of the a posteriori probability functions is of interest, such a global scaling is irrelevant. Hence, it is noted that the computation of the denominator of equation (37) may be completely waived to save computational power.

- *Computation of smoothed dominant sound source directions*

The smoothed dominant sound source directions $\hat{\boldsymbol{\Omega}}_{\text{DOM}}^{(f_{\mathcal{A},k(d)})}(k)$, $d = 1, \dots, \tilde{D}(k)$, are computed in step or stage 44 using the a posteriori probability functions $P_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k)$, $d = 1, \dots, \tilde{D}(k)$. In particular, the smoothed direction $\hat{\boldsymbol{\Omega}}_{\text{DOM}}^{(f_{\mathcal{A},k(d)})}(k)$ of the d -th sound source found for frame k is obtained by searching for the maximum in the a posteriori probability function

$$P_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k), \quad \text{i.e.} \quad \hat{\boldsymbol{\Omega}}_{\text{DOM}}^{(f_{\mathcal{A},k(d)})}(k) = \operatorname{argmax}_{\boldsymbol{\Omega}_q} P_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k, \boldsymbol{\Omega}_q) \quad . \quad (35)$$

Determine indices and directions of currently active dominant sound sources

The set $\mathcal{I}_{\text{DOM,ACT}}(k)$ of the indices $i_{\text{ACT},k}(d')$, $d' = 1, \dots, D_{\text{ACT}}(k)$ of all $D_{\text{ACT}}(k)$ active dominant sound sources at frame k and the set $\mathcal{G}_{\boldsymbol{\Omega},\text{DOM,ACT}}(k)$ of the corresponding dominant source direction estimates $\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k}(d'))}(k)$, $d' = 1, \dots, D_{\text{ACT}}(k)$, at frame k are computed in step or stage 15 of Fig. 1 using the set

$\mathcal{G}_{\boldsymbol{\Omega},\text{DOM,ACT}}(k-1)$ of the smoothed estimates $\bar{\boldsymbol{\Omega}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k-1}(d'))}(k-1)$,

$d' = 1, \dots, D_{\text{ACT}}(k-1)$, of all active dominant sound source directions at frame $(k-1)$, the set $\mathcal{I}_{\text{DOM,ACT}}(k-1)$ of the corresponding indices $i_{\text{ACT},k-1}(d')$, $d' = 1, \dots, D_{\text{ACT}}(k-1)$, and the smoothed dominant sound source direction estimates

5 $\hat{\Omega}_{\text{DOM}}^{(f_{\mathcal{A},k}(d))}(k)$, $d = 1, \dots, \tilde{D}(k)$ obtained for frame k . This operation has the purpose of not spuriously deactivating sound sources which have not been detected for a small number of successive frames, which might happen for sources like e.g. castanets producing impulse-like sounds with short pauses between
 10 the individual impulses. Thus, it is reasonable to deactivate sound sources which were assumed to be active in the last (i.e. the $(k-1)$ -th) frame, only if they have not been detected for a predefined number K_{INACT} of successive frames. According to the previous considerations, in a first step
 15 the joined set $\mathcal{I}_{\text{JOINED}}(k)$ of the set $\mathcal{I}_{\text{DOM,ACT}}(k-1)$ of the indices $i_{\text{ACT},k-1}(d')$, $d' = 1, \dots, D_{\text{ACT}}(k-1)$ of all $D_{\text{ACT}}(k-1)$ active dominant sound sources at frame $(k-1)$ and the set

$$\mathcal{I}_{\text{NEW}}(k) := \{f_{\mathcal{A},k}(d) | 1 \leq d \leq \tilde{D}(k)\} \quad (36)$$

of the indices of all newly detected sound sources are computed:
 20 $\mathcal{I}_{\text{JOINED}}(k) := \mathcal{I}_{\text{NEW}}(k) \cup \mathcal{I}_{\text{DOM,ACT}}(k-1)$. (37)

From this set the desired set $\mathcal{I}_{\text{DOM,ACT}}(k)$ is obtained by removing from $\mathcal{I}_{\text{JOINED}}(k)$ the indices of such sources which have not been detected for a number of K_{INACT} previous successive
 25 frames. The number $D_{\text{ACT}}(k)$ of active dominant sound sources at frame k is set to the number of elements of $\mathcal{I}_{\text{DOM,ACT}}(k)$.

Finally, the dominant source direction estimates $\bar{\Omega}_{\text{DOM,ACT}}^{(i_{\text{ACT},k}(d'))}(k)$, $d' = 1, \dots, D_{\text{ACT}}(k)$, where $i_{\text{ACT},k}(d')$ indicate the elements of $\mathcal{I}_{\text{DOM,ACT}}(k)$, are determined by

$$\bar{\mathbf{Q}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k}(d'))}(k) = \begin{cases} \hat{\mathbf{Q}}_{\text{DOM}}^{(i_{\text{ACT},k}(d'))}(k) & \text{if } i_{\text{ACT},k}(d') \in \mathcal{I}_{\text{NEW}}(k) \\ \bar{\mathbf{Q}}_{\text{DOM,ACT}}^{(i_{\text{ACT},k}(d'))}(k-1) & \text{else} \end{cases} . \quad (38)$$

This means that the directions of previously active dominant sound sources are held fixed if the respective sound source is not newly detected at frame k .

5

Basics of Higher Order Ambisonics

Higher Order Ambisonics (HOA) is based on the description of a sound field within a compact area of interest, which is assumed to be free of sound sources. In that case the spatio-temporal behaviour of the sound pressure $p(t, \mathbf{x})$ at time t and position $\mathbf{x}$ within the area of interest is physically fully determined by the homogeneous wave equation. In the following a spherical coordinate system as shown in Fig. 5 is assumed. In the used coordinate system the x axis points to the frontal position, the y axis points to the left, and the z axis points to the top. A position in space $\mathbf{x} = (r, \theta, \phi)^T$ is represented by a radius $r > 0$ (i.e. the distance to the coordinate origin), an inclination angle $\theta \in [0, \pi]$ measured from the polar axis z and an azimuth angle $\phi \in [0, 2\pi[$ measured counter-clockwise in the $x-y$ plane from the x axis. $(\cdot)^T$ denotes the transposition.

Then, it can be shown (cf. E.G. Williams, "Fourier Acoustics", vol.93 of Applied Mathematical Sciences, Academic Press, 1999) that the Fourier transform of the sound pressure with respect to time denoted by $\mathcal{F}_t(\cdot)$, i.e.

$$P(\omega, \mathbf{x}) = \mathcal{F}_t(p(t, \mathbf{x})) = \int_{-\infty}^{\infty} p(t, \mathbf{x}) e^{-i\omega t} dt \quad (39)$$

with ω denoting the angular frequency and i indicating the imaginary unit, can be expanded into a series of Spherical Harmonics according to

$$P(\omega = kc_s, r, \theta, \phi) = \sum_{n=0}^N \sum_{m=-n}^n A_n^m(k) j_n(kr) S_n^m(\theta, \phi) . \quad (40)$$

In equation (40), c_s denotes the speed of sound and k denotes

30

the angular wave number, which is related to the angular frequency ω by $k = \frac{\omega}{c_s}$, $j_n(\cdot)$ denotes the spherical Bessel functions of the first kind and $S_n^m(\theta, \phi)$ denotes the real-valued Spherical Harmonics of order n and degree m , which are defined in below section *Definition of real-valued Spherical Harmonics*. The expansion coefficients $A_n^m(k)$ are depending only on the angular wave number k . It is implicitly assumed that the sound pressure is spatially band-limited. Thus the series is truncated with respect to the order index n at an upper limit N , which is called the order of the HOA representation.

If the sound field is represented by a superposition of an infinite number of harmonic plane waves of different angular frequencies ω arriving from all possible directions specified by the angle tuple (θ, ϕ) , it can be shown (see B. Rafaely, "Plane-wave Decomposition of the Sound Field on a Sphere by Spherical Convolution", J. Acoust. Soc. Am., vol.4(116), pp.2149-2157, 2004) that the respective plane wave complex amplitude function $C(\omega, \theta, \phi)$ can be expressed by the following Spherical Harmonics expansion:

$$C(\omega = kc_s, \theta, \phi) = \sum_{n=0}^N \sum_{m=-n}^n C_n^m(k) S_n^m(\theta, \phi) , \quad (41)$$

where the expansion coefficients $C_n^m(k)$ are related to the expansion coefficients $A_n^m(k)$ by $A_n^m(k) = 4\pi i^n C_n^m(k)$.

When assuming that the individual coefficients $C_n^m(k = \omega/c_s)$ are functions of the angular frequency ω , the application of the inverse Fourier transform (denoted by $\mathcal{F}^{-1}(\cdot)$) provides time domain functions

$$c_n^m(t) = \mathcal{F}_t^{-1}(C_n^m(\omega/c_s)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C_n^m\left(\frac{\omega}{c_s}\right) e^{i\omega t} d\omega \quad (43)$$

for each order n and degree m , which can be collected in a single vector $\mathbf{c}(t)$ by $\mathbf{c}(t) =$

$$[c_0^0(t), c_1^{-1}(t), c_1^0(t), c_1^1(t), c_2^{-2}(t), c_2^{-1}(t), c_2^0(t), c_2^1(t), c_2^2(t), \dots, c_N^{N-1}(t), c_N^N(t)]^T .$$

The position index of a time domain function $c_n^m(t)$ within the

vector $\mathbf{c}(t)$ is given by $n(n+1)+1+m$. The overall number of elements in the vector $\mathbf{c}(t)$ is given by $O = (N+1)^2$.

The final Ambisonics format provides the sampled version of $\mathbf{c}(t)$ using a sampling frequency f_s as

$$\{\mathbf{c}(lT_s)\}_{l \in \mathbb{N}} = \{\mathbf{c}(T_s), \mathbf{c}(2T_s), \mathbf{c}(3T_s), \mathbf{c}(4T_s), \dots\} \quad (45)$$

where $T_s = 1/f_s$ denotes the sampling period. The elements of $\mathbf{c}(lT_s)$ are referred to as Ambisonics coefficients. The time domain signals $c_n^m(t)$ and hence the Ambisonics coefficients are real-valued.

10

- *Definition of real-valued Spherical Harmonics*

The real-valued Spherical Harmonics $S_n^m(\theta, \phi)$ are expressed by

$$S_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} P_{n,|m|}(\cos\theta) \text{trg}_m(\phi) \quad (46)$$

$$\text{with } \text{trg}_m(\phi) = \begin{cases} \sqrt{2}\cos(m\phi) & \text{for } m > 0 \\ 1 & \text{for } m = 0 \\ -\sqrt{2}\sin(m\phi) & \text{for } m < 0 \end{cases} . \quad (47)$$

15 The associated Legendre functions $P_{n,m}(x)$ are defined as

$$P_{n,m}(x) = (1-x^2)^{\frac{m}{2}} \frac{d^m}{dx^m} P_n(x), \quad m \geq 0 \quad (48)$$

with the Legendre polynomial $P_n(x)$ and, unlike in the above-mentioned E.G. Williams textbook, without the Condon-Shortley phase term $(-1)^m$.

20

- *Spatial resolution of Higher Order Ambisonics*

A general plane wave function $\mathbf{x}(t)$ arriving from a direction $\boldsymbol{\Omega}_0 = (\theta_0, \phi_0)^T$ is represented in HOA by

$$c_n^m(t) = x(t) S_n^m(\boldsymbol{\Omega}_0), \quad 0 \leq n \leq N, |m| \leq n . \quad (49)$$

25 The corresponding spatial density of plane wave amplitudes

$c(t, \boldsymbol{\Omega}) := \mathcal{F}_t^{-1}(C(\omega, \boldsymbol{\Omega}))$ is given by

$$c(t, \boldsymbol{\Omega}) = \sum_{n=0}^N \sum_{m=-n}^n c_n^m(t) S_n^m(\boldsymbol{\Omega}) \quad (50)$$

$$= x(t) \underbrace{\left[\sum_{n=0}^N \sum_{m=-n}^n S_n^m(\boldsymbol{\Omega}_0) S_n^m(\boldsymbol{\Omega}) \right]}_{v_N(\theta)} . \quad (51)$$

It can be seen from equation (51) that it is a product of

the general plane wave function $x(t)$ and a spatial dispersion function $v_N(\theta)$, which can be shown as depending only on the angle θ between $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ having the property

$$\cos\theta = \cos\theta\cos\theta_0 + \cos(\phi - \phi_0)\sin\theta\sin\theta_0 . \quad (52)$$

5 As expected, in the limit of an infinite order, i.e. $N \rightarrow \infty$, the spatial dispersion function turns into a Dirac delta

$$\delta(\cdot), \text{ i.e. } \lim_{N \rightarrow \infty} v_N(\theta) = \frac{\delta(\theta)}{2\pi} . \quad (53)$$

However, in the case of a finite order N , the contribution of the general plane wave from direction $\boldsymbol{\Omega}_0$ is smeared to
10 neighbouring directions, where the extent of the blurring decreases with an increasing order. A plot of the normalised function $v_N(\theta)$ for different values of N is provided in Fig. 6.

For any direction $\boldsymbol{\Omega}$ the time domain behaviour of the spatial
15 density of plane wave amplitudes is a multiple of its behaviour at any other direction. In particular, the functions $c(t, \boldsymbol{\Omega}_1)$ and $c(t, \boldsymbol{\Omega}_2)$ for some fixed directions $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are highly correlated with each other with respect to time t .

20 - Spherical Harmonic Transform

If the spatial density of plane wave amplitudes is discretised at a number of O spatial directions $\boldsymbol{\Omega}_o$, $1 \leq o \leq O$, which are nearly uniformly distributed on the unit sphere, O directional signals $c(t, \boldsymbol{\Omega}_o)$ are obtained. Collecting these signals
25 into a vector as $\mathbf{c}_{\text{SPAT}}(t) := [c(t, \boldsymbol{\Omega}_1) \ \dots \ c(t, \boldsymbol{\Omega}_O)]^T$, (54)

it can be verified by using equation (50) that this vector can be computed from the continuous Ambisonics representation $\mathbf{d}(t)$ defined in equation (44) by a simple matrix multiplication as $\mathbf{c}_{\text{SPAT}}(t) = \boldsymbol{\Psi}^H \mathbf{c}(t)$, (55)

30 where $(\cdot)^H$ indicates the joint transposition and conjugation, and $\boldsymbol{\Psi}$ denotes a mode-matrix defined by $\boldsymbol{\Psi} := [\mathbf{S}_1 \ \dots \ \mathbf{S}_O]$ (56)
with $\mathbf{S}_o :=$

$$[S_0^0(\boldsymbol{\Omega}_o) \ S_1^{-1}(\boldsymbol{\Omega}_o) \ S_1^0(\boldsymbol{\Omega}_o) \ S_1^1(\boldsymbol{\Omega}_o) \ \dots \ S_N^{N-1}(\boldsymbol{\Omega}_o) \ S_N^N(\boldsymbol{\Omega}_o)] \ . \quad (57)$$

Because the directions $\boldsymbol{\Omega}_o$ are nearly uniformly distributed on the unit sphere, the mode matrix is invertible in general. Hence, the continuous Ambisonics representation can be
 5 computed from the directional signals $c(t, \boldsymbol{\Omega}_o)$ by

$$\mathbf{c}(t) = \boldsymbol{\Psi}^{-H} \mathbf{c}_{\text{SPAT}}(t) \ . \quad (58)$$

Both equations constitute a transform and an inverse transform between the Ambisonics representation and the 'spatial domain'. These transforms are denoted the Spherical Harmonic
 10 Transform and the inverse Spherical Harmonic Transform, respectively. Because the directions $\boldsymbol{\Omega}_o$ are nearly uniformly distributed on the unit sphere, there is the approximation

$$\boldsymbol{\Psi}^H \approx \boldsymbol{\Psi}^{-1} \ , \quad (59)$$

which justifies the use of $\boldsymbol{\Psi}^{-1}$ instead of $\boldsymbol{\Psi}^H$ in equation
 15 (55). All mentioned relations are valid for the discrete-time domain, too.

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

Claims

1. Method for determining directions ($\mathcal{G}_{\Omega, \text{DOM}, \text{ACT}}(k)$) of uncorrelated sound sources in a Higher Order Ambisonics representation denoted HOA of a sound field, said method including the step:
 - in a current time frame (k) of HOA coefficients ($\mathcal{C}(k)$), searching (11) successively preliminary direction estimates ($\tilde{\Omega}_{\text{DOM}}^{(d)}(k)$) of dominant sound sources, and computing (11) HOA sound field components ($\mathcal{C}_{\text{DOM}, \text{CORR}}^{(d)}(k)$) created by the corresponding dominant sound sources, wherein in each iteration of said searching each further direction estimate is computed from a residual HOA representation ($\mathcal{C}_{\text{REM}}^d(k)$) which represents the original HOA representation from which all the components correlated with the signals of previously found sound sources have been removed, wherein a current direction estimate is selected out of a number of predefined test directions, such that the power of the related general plane wave of the residual HOA representation ($\mathcal{C}_{\text{REM}}^d(k)$), impinging from the chosen direction on a listener position, is maximum compared to that of all other test directions.
2. Method according to claim 1, wherein said selected direction estimates for said current time frame (k) of HOA coefficients ($\mathcal{C}(k)$) are assigned (13) to dominant sound sources found in the previous time frame ($k-1$) of HOA coefficients ($\mathcal{C}(k-1)$) and the final direction estimates are smoothed with respect to the resulting time trajectory.
3. Method according to claim 2, wherein said smoothing is performed by carrying out a Bayesian inference process, wherein this Bayesian inference process exploits a sta-

tistical a priori sound source movement model and the directional power distributions of the dominant sound source components of the original HOA representation.

- 5 4. Method according to claim 3, wherein said statistical a priori model statistically predicts the movement of individual sound sources from the knowledge of their direction in said previous time frame $(k-1)$ and the knowledge of the movement between said previous time frame $(k-1)$ and the penultimate time frame $(k-2)$.
10
5. Method according to claim 3 or 4, wherein said assignment of direction estimates to dominant sound sources found in said previous time frame $(k-1)$ of HOA coefficients is accomplished by a joint minimisation of the angles between
15 pairs of a direction estimate and the direction of a previously found sound source, and maximisation of the absolute value of the correlation coefficient between the pairs of the directional signals related to a direction estimate and to a dominant sound source found in said
20 previous time frame $(k-1)$ of HOA coefficients.
6. Method for determining directions $(\mathcal{G}_{\mathbf{D},\mathbf{DOM},\mathbf{ACT}}(k))$ of uncorrelated sound sources in a Higher Order Ambisonics representation denoted HOA of a sound field, said method including the steps:
25
 - in a current time frame (k) of HOA coefficients $(\mathbf{C}(k))$, searching (11) successively preliminary direction estimates $(\tilde{\mathbf{D}}_{\mathbf{DOM}}^{(d)}(k))$ of dominant sound sources, and computing
30 (11) HOA sound field components $(\mathbf{C}_{\mathbf{DOM},\mathbf{CORR}}^{(d)}(k))$ which are created by the corresponding dominant sound sources, and computing (11) the corresponding directional signals $(\mathbf{x}_{\mathbf{INST}}^{(d)}(k))$;

- assigning (13) said computed dominant sound sources to corresponding sound sources active in the previous time frame $(k-1)$ of said HOA coefficients by comparing said preliminary direction estimates $(\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k))$ of said current time frame (k) and smoothed directions $(\mathcal{G}_{\boldsymbol{\Omega},\text{DOM},\text{ACT}}(k-1))$ of sound sources active in said previous time frame $(k-1)$, and by correlating said directional signals $(\mathbf{x}_{\text{INST}}^{(d)}(k))$ of said current time frame (k) and directional signals $(\mathbf{X}_{\text{ACT}}(k-1))$ of sound sources active in said previous time frame $(k-1)$, resulting in an assignment function $(f_{\mathcal{A},k})$;
 - computing (14) smoothed dominant source directions $(\hat{\boldsymbol{\Omega}}_{\text{DOM}}^{(f_{\mathcal{A},k}(d)})(k))$ using said assignment function $(f_{\mathcal{A},k})$, said set $(\mathcal{G}_{\boldsymbol{\Omega},\text{DOM},\text{ACT}}(k-1))$ of smoothed directions in said previous time frame, a set $(\mathcal{I}_{\text{DOM},\text{ACT}}(k-1))$ of indices of active dominant sound sources in said previous time frame $(k-1)$, a set $(\mathcal{G}_{\boldsymbol{\theta},\text{DOM},\text{ACT}}(k-1))$ of respective source movement angles between the penultimate time frame $(k-2)$ and said previous time frame $(k-1)$, and said HOA sound field components $(\mathbf{C}_{\text{DOM,CORR}}^{(d)}(k))$ created by the corresponding dominant sound sources;
 - determining (15) indices $(\mathcal{I}_{\text{DOM},\text{ACT}}(k))$ and directions $(\mathcal{G}_{\boldsymbol{\Omega},\text{DOM},\text{ACT}}(k))$ of the active dominant sound sources of said current time frame (k) , using said smoothed dominant source directions $(\hat{\boldsymbol{\Omega}}_{\text{DOM}}^{(f_{\mathcal{A},k}(d)})(k))$, the frame delayed (174) version of directions $(\mathcal{G}_{\boldsymbol{\Omega},\text{DOM},\text{ACT}}(k-1))$ of the active dominant sound sources of said previous time frame $(k-1)$ and the frame delayed (172) version of indices $(\mathcal{I}_{\text{DOM},\text{ACT}}(k-1))$ of the active dominant sound sources of said previous time frame $(k-1)$,
- wherein said directional signals $(\mathbf{X}_{\text{ACT}}(k-1))$ of sound

sources active in said previous time frame ($k-1$) are computed (12) from said frame delayed (174) version of directions ($\mathcal{G}_{\Omega, \text{DOM}, \text{ACT}}(k-1)$) of the active dominant sound sources of said previous time frame ($k-1$) and the HOA coefficients ($\mathcal{C}(k-1)$) of said previous time frame using mode matching,

and wherein said set ($\mathcal{G}_{\hat{\theta}, \text{DOM}, \text{ACT}}(k-1)$) of source movement angles between said penultimate time frame ($k-2$) and said previous time frame ($k-1$) is computed from said frame delayed (174) version of directions ($\mathcal{G}_{\Omega, \text{DOM}, \text{ACT}}(k-1)$) of the active dominant sound sources of said previous time frame ($k-1$) and a further frame delayed (173) version ($\mathcal{G}_{\Omega, \text{DOM}, \text{ACT}}(k-2)$) thereof.

7. Apparatus for determining directions ($\mathcal{G}_{\Omega, \text{DOM}, \text{ACT}}(k)$) of uncorrelated sound sources in a Higher Order Ambisonics representation denoted HOA of a sound field, said apparatus including:
 - means (11) being adapted for searching successively in a current time frame (k) of HOA coefficients ($\mathcal{C}(k)$) preliminary direction estimates ($\tilde{\mathcal{D}}_{\text{DOM}}^{(d)}(k)$) of dominant sound sources, and for computing HOA sound field components ($\mathcal{C}_{\text{DOM}, \text{CORR}}^{(d)}(k)$) which are created by the corresponding dominant sound sources, and for computing the corresponding directional signals ($\mathbf{x}_{\text{INST}}^{(d)}(k)$);
 - means (13) being adapted for assigning said computed dominant sound sources to corresponding sound sources active in the previous time frame ($k-1$) of said HOA coefficients by comparing said preliminary direction estimates ($\tilde{\mathcal{D}}_{\text{DOM}}^{(d)}(k)$) of said current time frame (k) and smoothed directions ($\mathcal{G}_{\Omega, \text{DOM}, \text{ACT}}(k-1)$) of sound sources active in said previous time frame ($k-1$), and by correlating said direc-

- tional signals ($\mathbf{x}_{\text{INST}}^{(d)}(k)$) of said current time frame (k) and directional signals ($\mathbf{X}_{\text{ACT}}(k-1)$) of sound sources active in said previous time frame ($k-1$), resulting in an assignment function ($f_{\mathcal{A},k}$);
- 5 - means (14) being adapted for computing smoothed dominant source directions ($\hat{\boldsymbol{\Omega}}_{\text{DOM}}^{(f_{\mathcal{A},k(d)})}(k)$) using said assignment function ($f_{\mathcal{A},k}$), said set ($\mathcal{G}_{\boldsymbol{\Omega},\text{DOM},\text{ACT}}(k-1)$) of smoothed directions in said previous time frame, a set ($\mathcal{I}_{\text{DOM},\text{ACT}}(k-1)$) of indices of active dominant sound sources in said previous
- 10 time frame ($k-1$), a set ($\mathcal{G}_{\hat{\boldsymbol{\Omega}},\text{DOM},\text{ACT}}(k-1)$) of respective source movement angles between the penultimate time frame ($k-2$) and said previous time frame ($k-1$), and said HOA sound field components ($\mathbf{C}_{\text{DOM},\text{CORR}}^{(d)}(k)$) created by the corresponding dominant sound sources;
- 15 - means (15) being adapted for determining indices ($\mathcal{I}_{\text{DOM},\text{ACT}}(k)$) and directions ($\mathcal{G}_{\boldsymbol{\Omega},\text{DOM},\text{ACT}}(k)$) of the active dominant sound sources of said current time frame (k), using said smoothed dominant source directions ($\hat{\boldsymbol{\Omega}}_{\text{DOM}}^{(f_{\mathcal{A},k(d)})}(k)$), the frame delayed (174) version of directions ($\mathcal{G}_{\boldsymbol{\Omega},\text{DOM},\text{ACT}}(k-1)$)
- 20 of the active dominant sound sources of said previous time frame ($k-1$) and the frame delayed (172) version of indices ($\mathcal{I}_{\text{DOM},\text{ACT}}(k-1)$) of the active dominant sound sources of said previous time frame ($k-1$),
- wherein said directional signals ($\mathbf{X}_{\text{ACT}}(k-1)$) of sound
- 25 sources active in said previous time frame ($k-1$) are computed (12) from said frame delayed (174) version of directions ($\mathcal{G}_{\boldsymbol{\Omega},\text{DOM},\text{ACT}}(k-1)$) of the active dominant sound sources of said previous time frame ($k-1$) and the HOA coefficients ($\mathbf{C}(k-1)$) of said previous time frame using
- 30 mode matching,
- and wherein said set ($\mathcal{G}_{\hat{\boldsymbol{\Omega}},\text{DOM},\text{ACT}}(k-1)$) of source movement

angles between said penultimate time frame $(k-2)$ and said previous time frame $(k-1)$ is computed from said frame delayed (174) version of directions $(\mathcal{G}_{\Omega, \text{DOM}, \text{ACT}}(k-1))$ of the active dominant sound sources of said previous
 5 time frame $(k-1)$ and a further frame delayed (173) version $(\mathcal{G}_{\Omega, \text{DOM}, \text{ACT}}(k-2))$ thereof.

8. Method according to claim 6, or apparatus according to claim 7, wherein in said determination of the number
 10 $(\tilde{D}(k))$ of detected dominant directional signals and the corresponding preliminary direction estimates $(\tilde{\Omega}_{\text{DOM}}^{(d)}(k))$, an HOA sound field component $(\mathcal{C}_{\text{DOM}, \text{CORR}}^{(d)}(k))$ which is created by the corresponding dominant sound sources is subtracted from said current time frame (k) of HOA coefficients
 15 $(\mathcal{C}(k))$ in order to obtain a corresponding residual HOA representation $(\mathcal{C}_{\text{REM}}^{(2)}(k))$, and this subtraction processing is repeatedly performed based on the in each case remaining residual HOA representation $(\mathcal{C}_{\text{REM}}^{(d)}(k))$ for further such sound field components, such that sound field
 20 components found are excluded for the further direction search.

9. Method according to the method of claim 8, or apparatus according to the apparatus of claim 8, wherein for a single direction index (d) the directional power distribution
 25 $(p^{(d)}(k))$ of the remaining residual HOA representation $(\mathcal{C}_{\text{REM}}^{(d)}(k))$ is computed for a predefined number of discrete test directions (Ω_q) which are nearly uniformly distributed on the unit sphere and said directional power distribution is analysed for the presence of a dominant
 30 sound source, and if the absence of a dominant sound source is detected the direction search is stopped and if

a dominant source is detected a preliminary estimate of its direction ($\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k)$) with respect to the coordinate origin is computed.

- 5 10. Method according to the method of claims 8 and 9, or apparatus according to the apparatus of claims 8 and 9, wherein, after having determined a preliminary estimate ($\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k)$) of a dominant source direction, the respective directional signal ($\boldsymbol{x}_{\text{INST}}^{(d)}(k)$) and the HOA representation ($\boldsymbol{C}_{\text{DOM,CORR}}^{(d)}(k)$) of the sound field components which are assumed to be created by the same sound source are computed as follows:
- rotating (31) a fixed predefined spherical grid ($\mathcal{G}_{\boldsymbol{\Omega},\text{INIT}}$) consisting of sampling positions ($\boldsymbol{\Omega}_{\text{INIT},o}$), which are targeted to be uniformly distributed on the unit sphere, to provide the grid ($\mathcal{G}_{\boldsymbol{\Omega},\text{ROT}}^{(d)}(k)$) of rotated sampling positions ($\boldsymbol{\Omega}_{\text{ROT},o}^{(d)}(k)$), wherein said rotation is performed such that a first rotated sampling position ($\boldsymbol{\Omega}_{\text{ROT},1}^{(d)}(k)$) corresponds to said preliminary direction estimate ($\tilde{\boldsymbol{\Omega}}_{\text{DOM}}^{(d)}(k)$);
 - 20 - transforming (32) said remaining residual HOA representation ($\boldsymbol{C}_{\text{REM}}^{(d)}(k)$) to a spatial domain where it is equivalently represented by corresponding plane wave functions ($\boldsymbol{x}_{\text{INST}}^{(d)}(k)$) which are assumed to impinge on the coordinate origin from the rotated grid directions, and computing
 - 25 dominant sound source signals and grid direction signals;
 - performing (33) a prediction of said grid direction signals from dominant sound source signals;
 - computing (34) the HOA representation ($\boldsymbol{C}_{\text{DOM,CORR}}^{(d)}(k)$) of
 - 30 the predicted grid directional signals, representing the contribution of the dominant sound source to the sound field represented by said remaining residual HOA repre-

sentation $(\mathbf{C}_{\text{REM}}^{(d)}(k))$, by an inverse Spherical Harmonics Transform.

11. Method according to the method of one of claims 6 and 8 to 10, or apparatus according to the apparatus of one of claims 7 to 10, wherein said computing (14) of smoothed dominant source directions $(\hat{\mathbf{\Omega}}_{\text{DOM}}^{(f_{\mathcal{A},k(d)})}(k))$ is carried out as follows:
 - computing (42) a directional a priori probability functions $(\mathbf{P}_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k))$ for dominant sound source directions using said assignment function $(f_{\mathcal{A},k})$, said set $(\mathcal{G}_{\mathbf{\Omega},\text{DOM},\text{ACT}}(k-1))$ of smoothed directions in said previous time frame, said set $(\mathcal{I}_{\text{DOM},\text{ACT}}(k-1))$ of indices of active dominant sound sources in said previous time frame, and said set $(\mathcal{G}_{\hat{\mathbf{\Omega}},\text{DOM},\text{ACT}}(k-1))$ of source movement angles;
 - computing (41) directional likelihood functions $(L^{(f_{\mathcal{A},k(d)})}(k))$ for dominant sound source directions using said assignment function $(f_{\mathcal{A},k})$ and using said HOA sound field components $(\mathbf{C}_{\text{DOM,CORR}}^{(d)}(k))$ created by dominant sound sources;
 - computing (43) directional a posteriori probability functions $(\mathbf{P}_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k))$ for dominant sound source directions using said directional likelihood functions $(L^{(f_{\mathcal{A},k(d)})}(k))$ and using said directional a priori probability functions $(\mathbf{P}_{\text{PRIO}}^{(f_{\mathcal{A},k(d)})}(k))$;
 - determining (44) smoothed dominant sound source directions $(\hat{\mathbf{\Omega}}_{\text{DOM}}^{(f_{\mathcal{A},k(d)})}(k))$ using said directional a posteriori probability functions $(\mathbf{P}_{\text{POST}}^{(f_{\mathcal{A},k(d)})}(k))$ for dominant sound source directions.

1/5

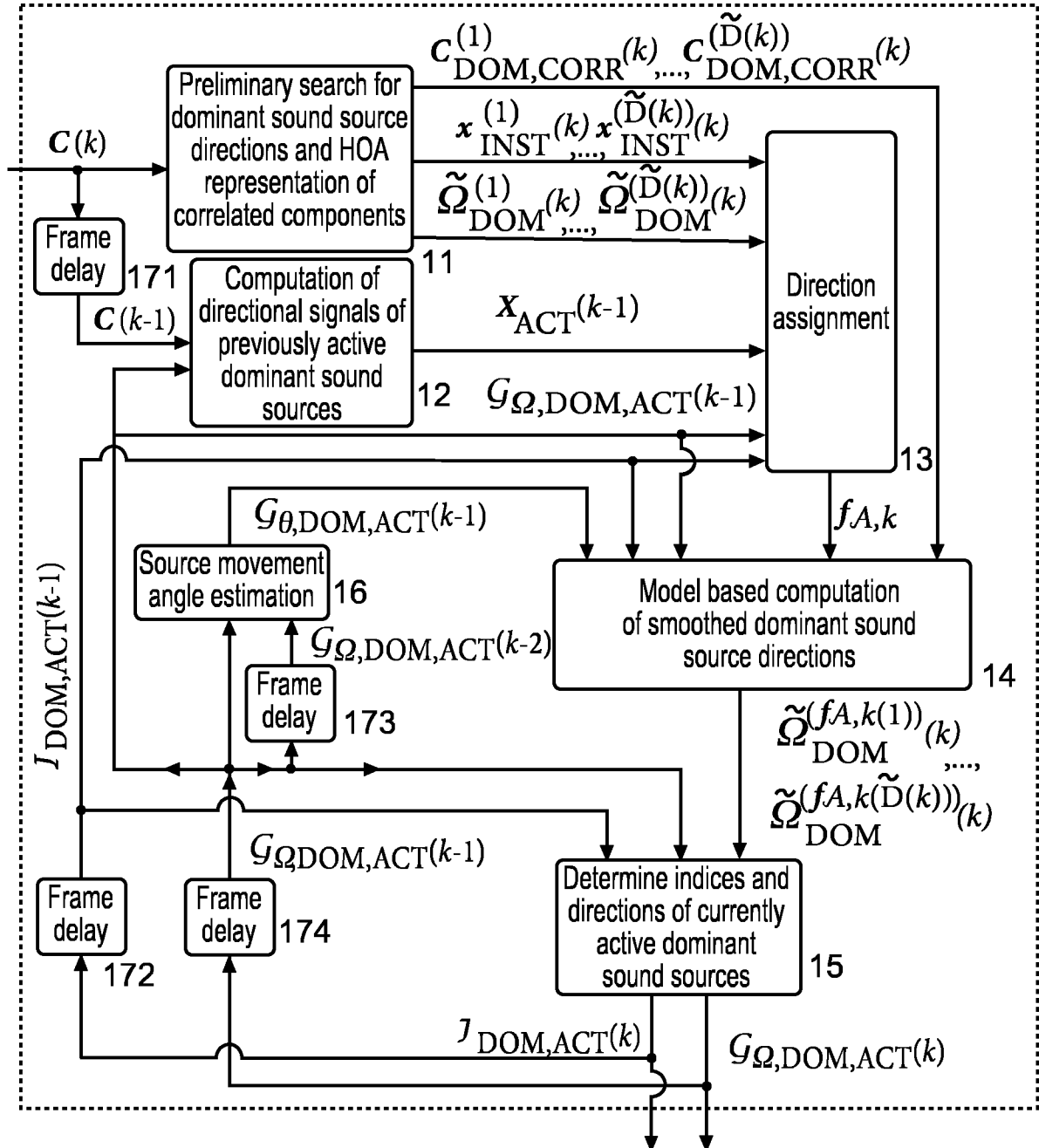


Fig. 1

2/5

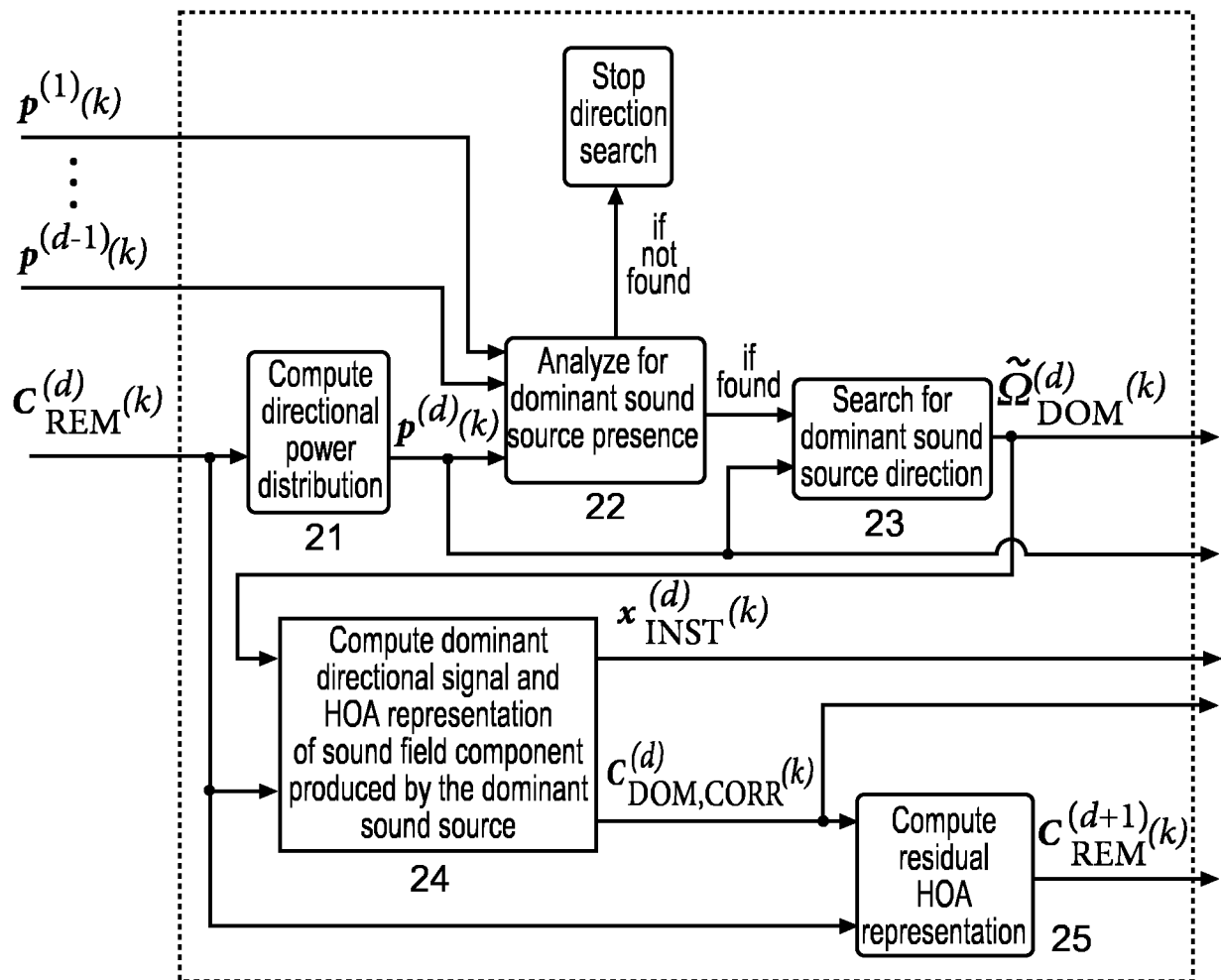


Fig. 2

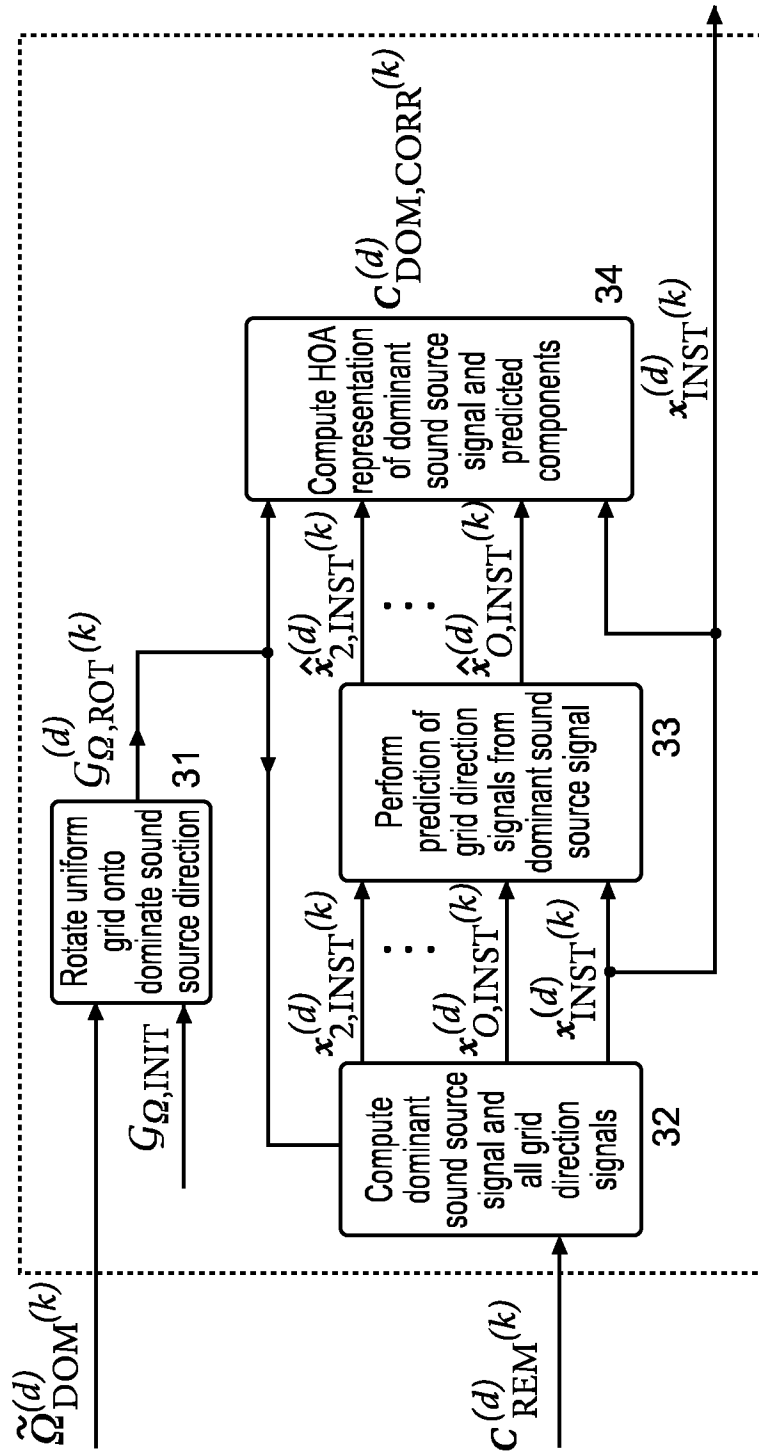


Fig. 3

4/5

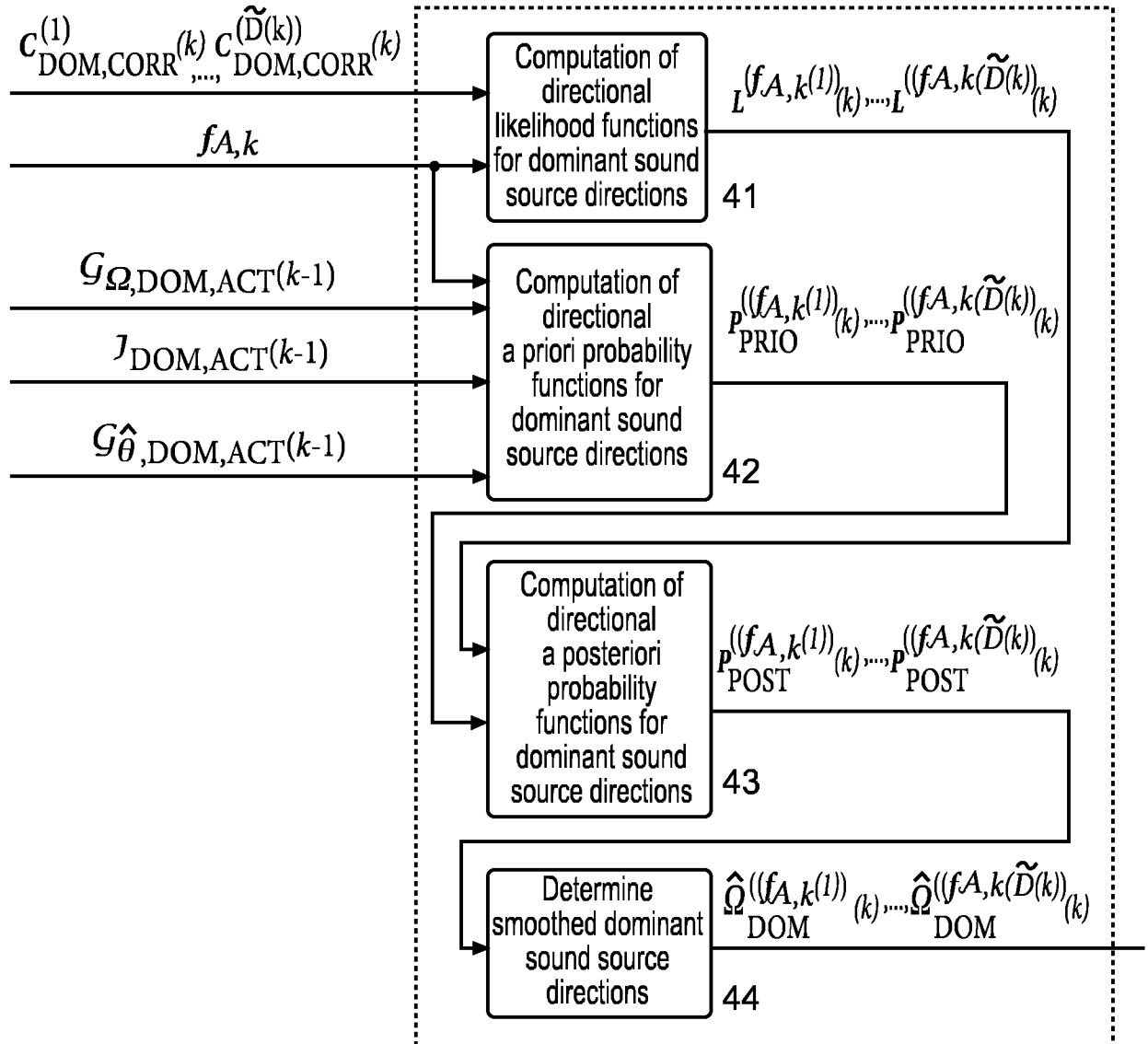


Fig. 4

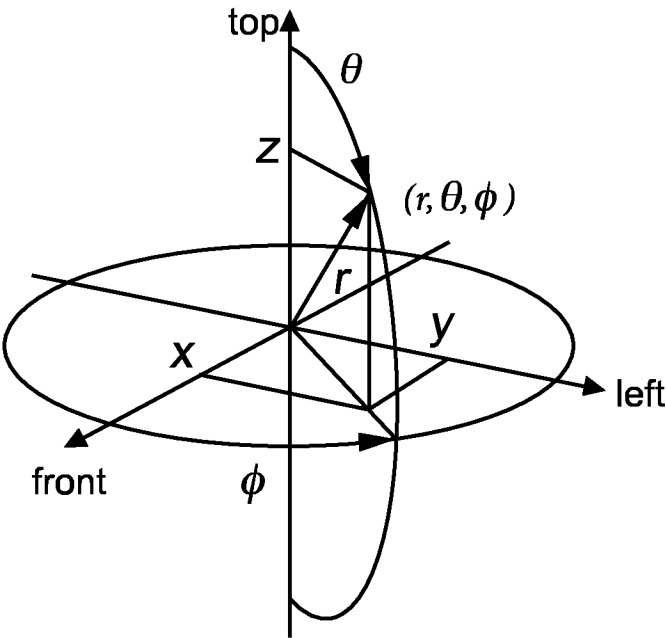


Fig. 5

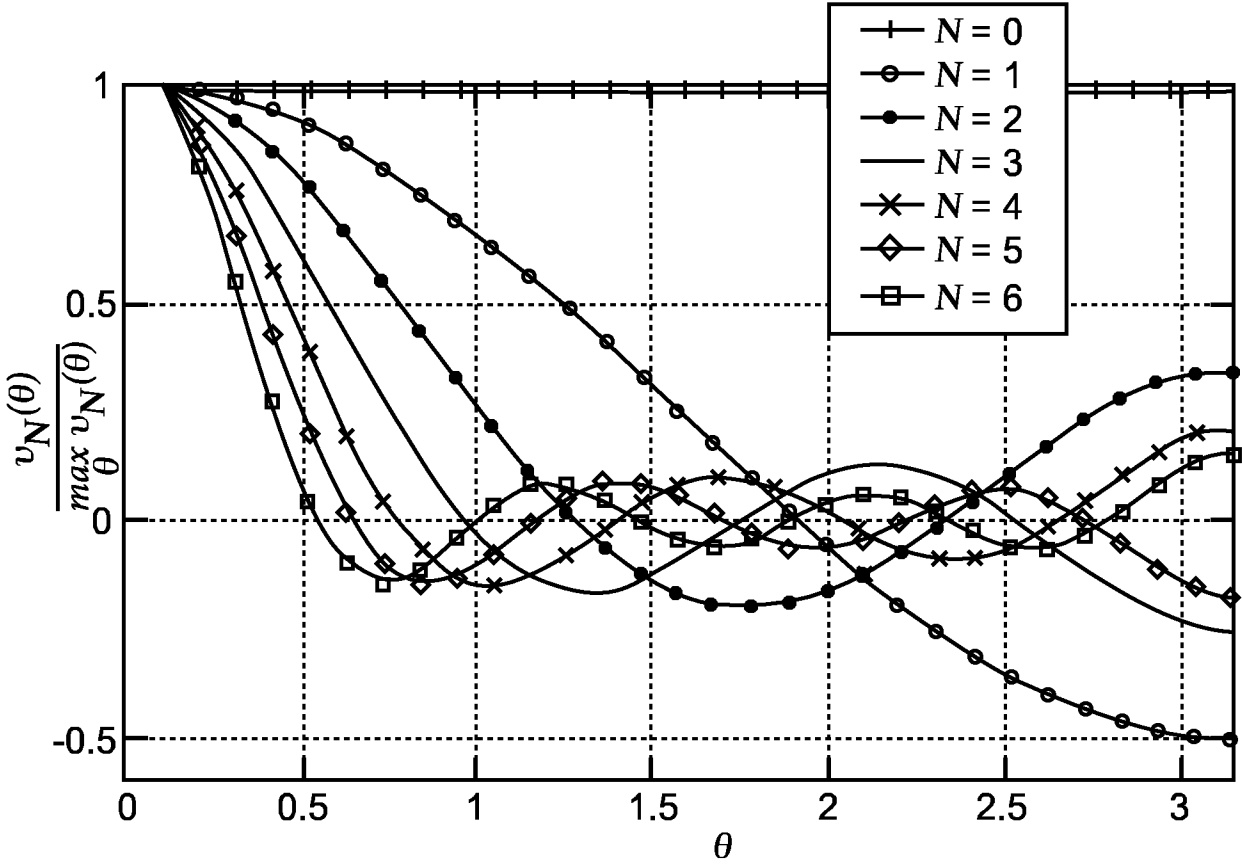


Fig. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/052479

A. CLASSIFICATION OF SUBJECT MATTER

INV. H04S3/00 G10L19/008 G01S3/80
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)

H04S G10L G01S

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.
A	HAOHAI SUN ET AL: "Optimal 3-D ho encoding with applications in improving close-spaced source localization", APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS (WASPAA), 2011 IEEE WORKSHOP ON, IEEE, 16 October 2011 (2011-10-16), pages 249-252, XP032011472, DOI: 10.1109/ASPAA.2011.6082263 ISBN: 978-1-4577-0692-9 the whole document ----- -/--	1-11



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

7 May 2014

Date of mailing of the international search report

14/05/2014

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

Schneider, Daniel

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2014/052479

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ERIK HELLERUD ET AL: "Spatial redundancy in Higher Order Ambisonics and its use for lowdelay lossless compression", ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2009. ICASSP 2009. IEEE INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 19 April 2009 (2009-04-19), pages 269-272, XP031459218, ISBN: 978-1-4244-2353-8 the whole document -----	1-11
A	EP 2 469 741 A1 (THOMSON LICENSING [FR]) 27 June 2012 (2012-06-27) the whole document -----	1-11
A	JÉRÔME DANIEL ET AL: "Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging", PREPRINTS OF PAPERS PRESENTED AT THE AES CONVENTION, XX, XX, 22 March 2003 (2003-03-22), pages 1-18, XP007904475, the whole document -----	1-11

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2014/052479

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
EP 2469741	A1	27-06-2012	CN 102547549 A	04-07-2012
			EP 2469741 A1	27-06-2012
			EP 2469742 A2	27-06-2012
			JP 2012133366 A	12-07-2012
			KR 20120070521 A	29-06-2012
			US 2012155653 A1	21-06-2012
