



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
26.08.2009 Bulletin 2009/35

(51) Int Cl.:
H04S 3/02 (2006.01)

(21) Application number: **08101732.9**

(22) Date of filing: **19.02.2008**

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

• **Eilts-Grimm, Klaus**
21339 Lüneburg (DE)
• **Schmidt, Jürgen**
31515 Wunstorf (DE)

(71) Applicant: **Deutsche Thomson OHG**
30625 Hannover (DE)

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

(72) Inventors:
• **Batke, Johann-Markus**
30161 Hannover (DE)

(54) **Audio signal, method and apparatus for encoding or transmitting the same and method and apparatus for processing the same**

(57) Commonly used audio file formats relate directly to the audio channels, i.e. the audio information is stored like it is fed to the loudspeaker. However, for obtaining a general description of audio information, a paradigm shift is necessary. Instead of storing the loudspeaker channels, a new audio format may use the description of the spatial sound field. A method for encoding or transmitting an audio signal comprises the steps of providing one or

more audio source signals (X_{src}), determining for each of said audio source signals a specific position to which it relates, generating data sets containing the determined positions (r_{src}) of the audio sources, and encoding or transmitting (62) the data sets together with said audio source signals. Advantageously, this allows conversion into Higher Order Ambisonics (HOA) format and subsequent re-conversion into conventional audio data that can be adapted to a given loudspeaker configuration.

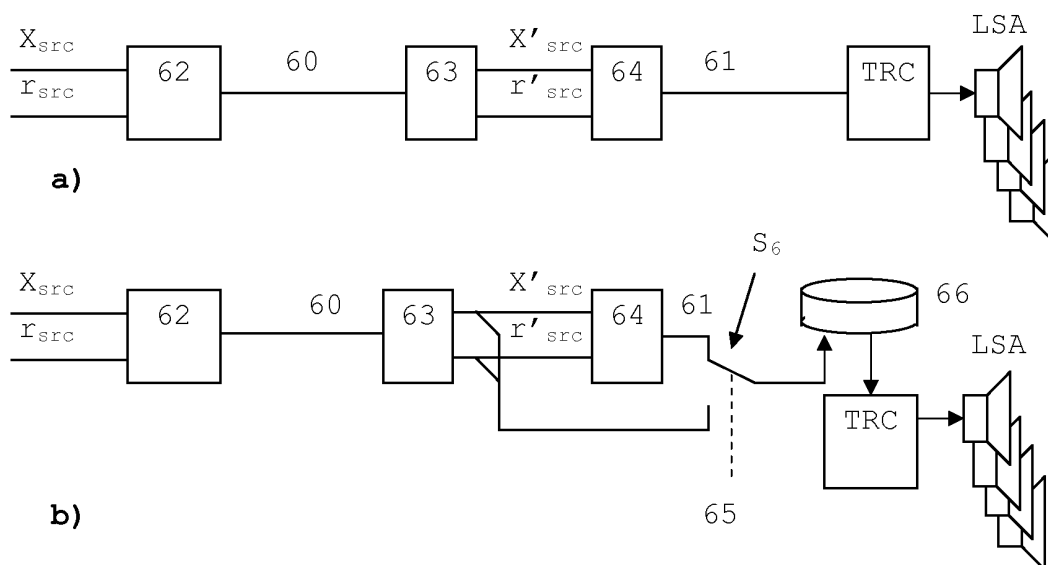


Fig. 6

DescriptionField of the invention

5 **[0001]** This invention relates to a generic audio signal format, a method and an apparatus for encoding or transmitting and a method and an apparatus for processing the same.

Background

10 **[0002]** Commonly used audio file formats relate directly to the audio channels, i.e. the audio information is stored like it is fed to the loudspeaker. The playback of such files is strictly bound to correct positioning of the loudspeakers. Audio formats like 2.0 (stereo) and 5.1 (surround sound) are able to reproduce a spatial impression of the audio content. However, this spatial impression is strictly two-dimensional (2D). Also extensions of these audio formats with higher number of audio channels like 7.1 or 9.1 stick to 2D sound field representation. The newly developed 22.2 format¹ is a
15 format capable of representing audio content with height information.

¹ K.Hamasaki, T. Nishiguchi, R.Okumaura, Y.Nakayama: "Wide listening area with exceptional spatial sound quality of a 22.2 multichannel sound system" (Audio Engineering Society Preprints, Vienna, Austria, May 2007)

[0003] For obtaining a general description of audio information, a paradigm shift is necessary. Instead of storing the loudspeaker channels, a new audio format may use the description of the spatial sound field. A known solution for spatial
20 sound field description is based on Higher Order Ambisonics (HOA), a technology that describes spatial sounds fields using the coefficients of the FOURIER-BESSEL series (also known under different names²). The possible spatial resolution using this description is determined by the order N of the series. This representation is very flexible and can hold any type of audio information, e.g. traditional stereo signals or surround sound. Loudspeaker channels are treated as a source at a distinct position (e.g. the loudspeaker's position). It is generally known in the art how to convert conventional
25 audio signals into HOA representations and vice versa, whereby audio signals' positioning information is required. However, HOA files containing traditional audio formats may result in a higher number of audio channels than the original file. Therefore traditional sound representations instead of HOA are usually used.

² e.g. Jerome Daniel: Spatial Sound Encoding Including Near Field Effect: Introducing distance Coding Filters and a Viable, New Ambisonic Format. AES 23rd International Conference, Copenhagen, Denmark, 2003

30

Summary of the Invention

[0004] It would be desirable to reproduce sound as close as possible to the original sound source, using a given loudspeaker configuration, and optimizing the possibilities of the given loudspeaker configuration. Further, it would be
35 desirable to have a sound representation that can be adapted to different actual loudspeaker configurations, so that sound can be reproduced optimally in any case. According to one aspect of the invention, an Ambisonics representation and particularly a Higher Order Ambisonics representation would enable such optimized playback of sound. It would also be desirable to minimize the effort for encoding, decoding and transcoding to and from Ambisonics representations, which also would minimize a loss of quality.

40 **[0005]** According to one aspect of the invention, a conventional audio signal is enhanced by additional data or metadata, wherein the additional data comprise sound source position information that enables conversion of the conventional audio signal into a Higher Order Ambisonics (HOA) representation of the sound field. Advantageously, this allows subsequent re-conversion into conventional audio data that can be adapted to a given loudspeaker configuration (e.g. during said re-conversion).

45 **[0006]** According to another aspect of the invention, a method for encoding or transmitting an audio signal comprises steps of providing one or more audio source signals, determining respective position information and encoding or transmitting the audio source signals together with metadata that comprise the determined audio source position information.

[0007] According to yet another aspect of the invention, the method for encoding or transmitting an audio signal further comprises steps of determining the number of source channels that are required for Ambisonics encoding, said number
50 being $(2N+1)$ for the 2D case and $(N+1)^2$ for the 3D case (N is the order of Ambisonics encoding), determining the number of available transmission or storage channels, and comparing the number of source channels required for Ambisonics encoding of the order N with the number of available transmission or storage channels, depending on said comparison, generating a mode decision information having either a first value if the number of source channels required for Ambisonics encoding of the order N is not less than the number of available transmission or storage channels, or
55 having a different second value otherwise, and generating, and storing or transmitting, the Ambisonics encoded version of the audio signal if the mode decision information has said first value, or otherwise storing or transmitting the received or retrieved audio signal. In one embodiment where the Ambisonics encoded version of the audio signal is transmitted, the mode decision information will also be transmitted. The order N may be determined as a function of a target number

of reproduction channels, or from the number of available transmission or storage channels.

[0008] According to a further aspect of the invention, a method for processing an audio signal comprises steps of receiving or retrieving from storage an encoded audio signal, extracting first audio source signals and additional information from the received or retrieved signal, wherein the first audio source signals relate to first audio source positions provided by the additional information, transforming the first audio source signals relating to first audio source positions into second audio source signals relating to different second audio source positions, and supplying said second audio source signals for storage or playback.

[0009] According to one aspect of the invention, in the method for processing an audio signal, the step of transforming the first audio source signals into second audio source signals comprises generating an Ambisonics representation of the sound field from a conventional audio source signal and said additional information describing the positions of sound sources, wherein the Ambisonics signal can be of higher order (HOA).

[0010] Corresponding apparatuses that utilize the methods are disclosed in the following detailed description.

[0011] Advantageous embodiments of the invention are disclosed in the dependent claims, the following description and the figures.

Brief description of the drawings

[0012] Exemplary embodiments of the invention are described with reference to the accompanying drawings, which show in

Fig.1 a general audio production chain;

Fig.2 conventional loudspeaker setup for playback in stereo and surround sound;

Fig.3 the principle of Ambisonics encoding;

Fig.4 an encoder according to one embodiment of the invention;

Fig.5 a decoder according to one embodiment of the invention;

Fig.6a an audio processing system according to one embodiment;

Fig.6b an audio processing system according to another embodiment; and

Fig.7 an audio transmission system according to one embodiment.

Detailed description of the invention

[0013] In a general audio production chain, as shown in Fig.1, acquisition of audio signals is achieved by one or more microphones M. The audio signals are encoded E, stored S and later decoded D for reproduction via one or more loudspeakers LS. Conventionally, each of the audio signals in the decoded signal relates to a particular loudspeaker. E.g. in a stereo setup, also denominated as 2.0 (since it has two direction-related audio channels and audio no channel that is not direction-related), the audio signals relate to the left and right microphones and loudspeakers.

[0014] Fig.2 a) shows a conventional loudspeaker setup for playback in stereo, and Fig.2 b) a conventional loudspeaker setup for surround sound, also known as 5.0 format. It is a convention that an angle of 60° must be between the two stereo loudspeaker boxes in order to reproduce the audio signal in the best possible manner. Similarly, in the 5.0 format the optimal angles between loudspeakers are subject to convention. Thus, the respective audio signals relate to specific relative positions that cannot be changed for a given reproduction system. That is, loudspeaker boxes need to be positioned according to these fixed relative positions in order to optimize the sound reproduction.

[0015] Using Ambisonics, and particularly HOA, as general representation of audio content has the following advantages. First, the audio content is independent from the loudspeaker setup. Thus, it has to be processed to match a given setup, wherein it will be optimized to match this setup. Second, 3D representation and high spatial resolution of audio content is fully supported.

[0016] A general Ambisonics based system is shown in Fig.3. A microphone array MA acquires the signals in a spatial manner. Position information P describing the microphone positions is added in an encoder E, which generates an Ambisonics representation 30 of given order N of the signals. This signal can be transcoded TR into a conventional audio signal having a desired number of channels that relate to desired positions of the loudspeakers. The localization of the different loudspeaker channels is the better, the higher the order N of the Ambisonics representation was. However,

the order N and the spatial positions (2-dimensional or 3-dimensional) have also an impact on the number of channels that the Ambisonics signal 30 requires, as described below. The conventional audio signal can be reproduced LSA on a loudspeaker array that may but need not correspond to the microphone array. However, positions of the loudspeakers must be known for transcoding the signal.

[0017] There are various aspects of the invention. In one aspect, a receiver performs a conversion from a given audio source arrangement to a required audio target arrangement, such as an individual loudspeaker arrangement. An encoder or transmitter provides a conventional audio signal with one or more microphone/loudspeaker related channels (such as 5.0) and attached position information that defines the positions of the microphone/loudspeaker of each channel. At the receiver side this signal can be converted to an Ambisonics representation, and in particular to a HOA (Higher Order Ambisonics) representation. This signal can be stored or transcoded/re-mapped for a desired channel and position configuration (e.g. according to an actual loudspeaker configuration or a particular configuration desired for other reasons). It is possible to select at the receiver side whether the Ambisonics representation or the conventional representation with additional position information shall be stored or further processed.

[0018] Advantages of this aspect are that the transmitted/received signal is backward compatible, since it can be decoded by conventional receivers that ignore the additional metadata information, and that the transmitted/received signal uses practically the same bandwidth than a conventional audio signal, since the additional metadata information is very little compared to the audio information (although it may be transmitted frequently, e.g. in fixed time intervals such as once every second, or every k audio frames).

[0019] In another aspect of the invention, the conversion from a given audio source arrangement into a HOA representation can be performed before transmitting, so that either the conventional audio signal plus position information or the generic HOA representation of the audio signal is transmitted. The latter is preferred if the required number of transmission channels is equal for both formats. The transmission signal comprises a mode indication showing whether the audio format is HOA or conventional, because two different formats are possible. The receiver extracts and evaluates the indication and performs the further processing according to the mode indication.

[0020] An audio processing system according to one embodiment of the invention is shown in Fig.6 a). A signal as described above, having one or more audio source signals X_{src} and position information r_{src} giving the positions of the audio source signals is received, and multiplexed 62 into a common signal 60. This signal can be stored (not shown) or transmitted, e.g. between different devices in a network. The signal is demultiplexed 63, wherein the audio source signals X'_{src} and position information r'_{src} are regained, and these are input into an Ambisonics encoder 64 that generates an Ambisonics signal 61. The order of this signal may be determined according to the number of available positions r'_{src} , but can also be influenced by available storage area and/or processing bandwidth. In particular, it is advantageous that the Ambisonics encoder 64 can be a HOA encoder, i.e. $N > 1$. The Ambisonics signal is fed into a transcoder TR and there re-mapped to a given loudspeaker configuration, and output to conventional multi-channel audio processing and loudspeakers LSA.

[0021] One advantage of this processing system is that the step of re-mapping can easily be adapted to the actual loudspeaker configuration, so that after a change in this configuration the optimization can also be changed according to the new loudspeaker number and/or positions. E.g. when a new loudspeaker is added to the reproduction system, its position information is provided to the transcoder and the Ambisonics signal can be re-mapped to match the new configuration.

[0022] The position information can be provided by user input (e.g. using a GUI), or by automatic loudspeaker position measuring systems. E.g. relative loudspeaker positions can be determined by reproducing a reference signal at a known position and measuring the different signal run-times. For the 3D case, reproduction of three reference signals at three distinct known positions can be used for automatic loudspeaker location.

[0023] An audio processing system according to another embodiment of the invention is shown in Fig.6 b). The signal 60 being composed of one or more audio source signals X_{src} and position information r_{src} giving the positions of the audio source signals is received and demultiplexed 62 into its components. It is possible to select S_6 whether these components or the HOA encoded signal representation 61 shall be used for the further processing, such as optional storing 66, transcoding and multi-channel audio processing. As described above, it may be advantageous to store the generic HOA representation, depending on the application. The selection signal 65 may depend on the parameters mentioned below, such as required order N , number of source positions or number of target (loudspeaker) positions.

[0024] The following section gives a brief overview on encoding and decoding HOA signals.

[0025] In the following, the spatial positions refer to a spherical coordinate system. The distance of sources (audio signals on the encoding side, loudspeakers on the playback side) is not taken into account for the sake of clarity. However, it is easily integrated to this encoding scheme using a known distance coding scheme, e.g. that of Jerome Daniel (reference cited above).

[0026] HOA encoding of audio signals is be done using

$$\Psi \mathbf{w} = \mathbf{A} \quad (\text{eq.1})$$

where Ψ is the mode matrix, $\mathbf{w}$ holds the speaker signals and $\mathbf{A}$ are the resulting HOA coefficients. The HOA coefficients in $\mathbf{A}$ are arranged in this order:

$$\mathbf{A} = [A_0^0 A_1^{-1} A_1^0 A_1^1 \dots]^T \quad (\text{eq.2})$$

Vector $\mathbf{A}$ holds

$$O = \begin{cases} (N+1)^2 & \text{for 3D representation} \\ 2N+1 & \text{for 2D representation} \end{cases} \quad (\text{eq.3})$$

elements. The speaker signals $\mathbf{w}$ are arranged as

$$\mathbf{w} = [w_1(t) \ w_2(t) \ \dots \ w_L(t)]^T \quad (\text{eq.4})$$

where L is the number of loudspeakers. As an example, a stereo signal is simply described as $\mathbf{w} = [W_1(t), W_2(t)]^T$ with left and right channel respectively. The mode matrix Ψ finally contains

$$\Psi = [\Psi_1 \ \Psi_2 \ \dots \ \Psi_L]^T \quad (\text{eq.5})$$

where Ψ_i with $i = 1 \dots L$ are the mode vectors for the individual speaker positions containing

$$\psi_i = [Y_0^0(\theta_i, \phi_i), Y_1^{-1}(\theta_i, \phi_i), Y_1^0(\theta_i, \phi_i), Y_1^1(\theta_i, \phi_i), \dots]^T \quad (\text{eq.6})$$

[0027] The directional position of the individual speakers is given by θ_i, ϕ_i in spherical coordinates, $Y_n^m(\theta_i, \phi_i)$ is the spherical harmonic function. The position of the speaker is referred as $\mathbf{r}_i = (r_i, \theta_i, \phi_i)$. As an example, the stereo setup of two loudspeakers is described by $\mathbf{r}_{\text{left}} = (r, 90^\circ, -30^\circ)$ and $\mathbf{r}_{\text{right}} = (r, 90^\circ, 30^\circ)$, where r denotes the speaker distance in meter, 90° is the declination angle and 30° is the azimuth angle.

[0028] The decoding of the HOA coefficients $\mathbf{A}$ is done using

$$\mathbf{w}_{\text{dec}} = \mathbf{D}\mathbf{A} \quad (\text{eq.7})$$

where $\mathbf{D}$ is the decoding matrix. It is chosen to pseudo inverse matrix

$$\mathbf{D} = \Psi^\dagger [\Psi \Psi^\dagger]^{-1} \quad (\text{eq. 8})$$

where $\dagger$ denotes the conjugate complex matrix transform. The property of the pseudo inverse

$$\mathbf{D} \Psi = \mathbf{I} \quad (\text{eq. 9})$$

with $\mathbf{I}$ denoting the identity matrix ensures the proper reconstruction of $\mathbf{w}$.

[0029] The HOA coefficient transmission is usually done by transmission of the individual vector elements of $\mathbf{A}$ resulting from the Fourier-Bessel representation. This results in a possibly higher number of channels than formerly required and very high numbers of channels for high orders. Existing ideas for HOA usage inside audio formats are therefore limited to an order of $N = 1$.

[0030] The invention results in a new audio format that is backward compatible to existing audio content. It is capable of holding audio content with full 3D information and any high spatial resolution, and therefore it is forward compatible with any audio content.

[0031] Using a pure HOA signal representation to carry standard audio formats like 2.0 and 5.1 and some others has the disadvantage of a higher number of channels required for sound field representation. Therefore, in one embodiment of the invention, a parametric approach is suggested driven by the following: Stereo (2.0) is 2D and two audio channels are required, transport using HOA however requires $O = 3$ channels. Surround sound (5.1) is 2D and 6 audio channels are required, transport using HOA however requires $O = 7$ channels.

New formats are capable of representing 3D information. As an example, 22.2 format requires 24 audio channels. For 3D HOA representation a number of $O = 25$ is necessary. The invention aims to provide an efficient solution to this problem: for 2.0 and 5.1 audio content it is generally less expensive in terms of channel/storage capacity to transmit/store the original audio information and additionally the locations of the sources. The result is a parameterised HOA signal representation at lowest cost. The full HOA representation is calculated on the receiver side, if necessary. Thus, a smooth mode selection is proposed that allows selection of the best possible format. The HOA representation also provides the advantage of scalability. A format like 22.2 requires 24 audio channels, as stated above. The order N defines the spatial resolution. To reduce the number of channels, the spatial resolution could be diminished. E.g. a HOA representation of order $N = 3$ would require only $O = 16$ channels. Generally, the HOA representation is scalable in terms of the spatial resolution.

[0032] An audio channel of a traditional audio format like stereo is viewed upon here as an audio source with a distinct position. An exemplary audio file format for HOA coefficients carrying several audio sources can be generated as follows:

1. Determine whether the arrangement of signal positions is plain (2D) or spatial (3D). Also the listener's position can be taken into account. The signal comprises O_{src} channels, which is 2 in the stereo case. Depending on this information the necessary order N is calculated using

$$N = \begin{cases} \left\lceil \sqrt{O_{src}} - 1 \right\rceil & \text{for 3D} \\ \left\lceil (O_{src} - 1) / 2 \right\rceil & \text{for 2D} \end{cases} \quad (\text{eq. 12a})$$

Using this order N in turn yields

$$O_s = \begin{cases} (N+1)^2 & \text{for 3D representation} \\ 2N+1 & \text{for 2D representation} \end{cases} \quad (\text{eq. 13})$$

[0033] O_s source channels are required for HOA coefficient representation (see eq.3).

2. The minimum size of channels for transport or storage is achieved as follows:

[0034] If $O_s < O_c$, then store a file containing

- (a) position of sources (example see above)
 - (b) is the arrangement 2D or 3D (implicitly given by source positions)
 - (c) order N (implicitly given by number of sources)
 - (d) unprocessed audio signals
- otherwise do HOA encoding of the audio signals using order N as described above and store HOA coefficients.
- The result is an audio file with the minimum number of required channels. Depending on the playback device, the audio content is HOA encoded using the additionally stored parameters and then transcoded to the loudspeaker setup, or it is played back unprocessed (e.g. stereo content on a stereo device). In one embodiment, the file is converted into a signal for transmission using the following steps:

3. If $O_s < O_c$, the signal is multiplexed using a first multiplexer MX1, else the signal is HOA encoded.

4. The result of the former step is multiplexed with a mode indication indicating the result of condition $O_s < O_c$ (i.e. the encoding mode) using a multiplexer MX2.

[0035] In one embodiment, decoding of this new signal is done as follows:

- 1. To extract the encoding mode information (i.e. $O_s < O_c$), DMX2 is used.
- 2. If $O_s < O_c$ is true (i.e. HOA encoding mode), the signal is demultiplexed using DMX1, after that it is HOA encoded. In the other case it can be transcoded directly.

[0036] Another aspect of usage of the invention is described in the following. The goal is to use a given number of channels available for transport in an optimal way. It is assumed that the number of source channels O_{src} is higher than the number of available channels.

[0037] A vector r_{src} holds all positions of source channels, e.g. L sources are described using the positions

$$r_{src} = [r_1 \ r_2 \ \dots \ r_L \] \quad (\text{eq.10})$$

with positions $r_i = (r_i, \theta_i, \phi_i)$ of source number i. All positions are assumed to be different from each other (otherwise the situation is trivial, since two sources with same position can be added into one). Using spherical coordinates is not mandatory, though.

[0038] A vector x_{src} holds all channels belonging to the source, e.g. L time signals

$$x_{src} = [x_1(t) \ x_2(t) \ \dots \ x_L(t) \] \quad (\text{eq.11})$$

[0039] The integer number O_{chan} defines the number of available channels. The order N available for a HOA description of signal vector x_{src} is calculated using

$$N = \begin{cases} \lfloor \sqrt{O_{chan}} - 1 \rfloor & \text{for 3D} \\ \left\lfloor \frac{O_{chan} - 1}{2} \right\rfloor & \text{for 2D} \end{cases} \quad (\text{eq.12})$$

[0040] Using this order N in turn yields

$$O_c = \begin{cases} (N+1)^2 & \text{for 3D representation} \\ 2N+1 & \text{for 2 D representation} \end{cases} \quad (\text{eq.13})$$

[0041] This is the number of channels to use to describe a HOA signal with order N as calculated above. Encoding of a signal x_{src} according to eq.11 with positional description r_{src} , as described by eq.10, is done as follows. In this case the signal is adapted to the channel properties.

1. Use all positions in r_{src} to determine if the arrangement is 2D or 3D.

2. Using this result and the given number O_{chan} of transport channels, the HOA representation of the source with maximum spatial resolution requires O_c signals following eq.13.

If $O_s > O_c$, this encoding ensures usage of given channel with maximum possible spatial resolution of the audio sources.

[0042] Mixing different HOA representations with different orders is possible. This is useful if different audio contents are encoded with different effort regarding spatial resolution. E.g. for a computer game, environment noise needs only low spatial resolution, whereas the audio information of the actor in the game should be encoded with high resolution.

[0043] An encoder according to one embodiment of the invention is shown in Fig.4. Audio source signals X_{src} and position information r_{src} are provided, as described above, and can be encoded in two different modes: either they are multiplexed MX1 into a common data stream, so that a receiver is enabled to generate a HOA representation (since all the data necessary for a HOA representation are included), or a HOA representation is generated HOA_{e1} before transmission. Depending on a mode selection signal MD it is possible to select S_1, S_2, S_3 one of the two modes. This mode decision signal is obtained by the above-described comparison CMP between the required number of channels O_s and the available number of channels O_c . The latter may be fixed or given. The required number of channels O_s is determined in a block eq4 according to equation 4 above, using the result of the block eq3 that performs equation 3 above, and a spatial arrangement information 2D3D indicating whether the spatial arrangement is 2-dimensional or 3-dimensional. The spatial arrangement information 2D3D is also an input to the block eq3 that performs equation 3. Finally, the mode decision information MD is multiplexed MX2 into the output data stream A_{enc} so that all necessary information for proper decoding is contained.

[0044] A decoder according to one embodiment of the invention is shown in Fig.5. A signal A_{enc} as encoded by the encoder of Fig.4 is demultiplexed DMX2 so that the mode decision information MD' is obtained. Depending on this information MD' the remaining signal is either demultiplexed into its audio and position components X'_{src}, r'_{src} and then HOA encoded HOA_e (if it was not HOA encoded), or it is directly used if it is already HOA encoded. Switching means S_4, S_5 controlled by the mode decision information MD' switch between these modes. The HOA encoded signal is provided to a transcoder, as described above.

[0045] In one embodiment, a device for encoding or transmitting an audio signal, comprises means for providing one or more audio source signals, means for determining for each of said audio source signals a specific position to which it relates, means for generating data sets containing the determined positions of the audio sources, and means for encoding or transmitting the data sets together with said audio source signals.

[0046] In one embodiment, said data sets are suitable for calculating a generic audio field representation based on Ambisonics representation.

[0047] In one embodiment, the device further comprises means for determining the number (O_s) of source channels that are required for Ambisonics encoding, said number being $(2N+1)$ for the 2D case and $(N+1)^2$ for the 3D case, where N is the order of Ambisonics encoding, means for determining the number (O_c) of available transmission or storage channels,

means for comparing the number (O_s) of source channels required for Ambisonics encoding of the order N with the number (O_c) of available transmission or storage channels; means for generating, depending on said comparison, a mode decision information (MD), having a first value if the number (O_s) of source channels required for Ambisonics encoding of the order N is not less than the number of available transmission or storage channels (O_c), or having a different second value otherwise, and means HOA_e for generating, and means for storing or transmitting, the Ambisonics encoded version of the audio signal if the mode decision information has said first value, or otherwise storing or transmitting the received or retrieved audio signal.

[0048] In another embodiment, a device for processing audio signals comprises means for receiving or retrieving from storage encoded audio signals, means for extracting first audio source signals and additional information from the

received or retrieved signals, wherein the first audio source signals relate to first audio source positions provided by the additional information, means (TRC) for transforming the first audio source signals relating to first audio source positions into second audio source signals relating to different second audio source positions, and means (LSA) for supplying said second audio source signals for storage or playback.

[0049] The invention can be used for all kinds of audio processing devices. These may be targeting music reproduction, but also voice reproduction, such as multi-channel teleconferencing systems. Advantageously, spatial information can be added to conventional multi-channel audio signals, and scalability in terms of spatial resolution can be provided.

[0050] It will be understood that the present invention has been described purely by way of example, and modifications of detail can be made without departing from the scope of the invention.

Each feature disclosed in the description and (where appropriate) the claims and drawings may be provided independently or in any appropriate combination. Features may, where appropriate be implemented in hardware, software, or a combination of the two. Connections may, where applicable, be implemented as wireless connections or wired, not necessarily direct or dedicated, connections. Reference numerals appearing in the claims are by way of illustration only and shall have no limiting effect on the scope of the claims.

Claims

1. Audio signal (60) comprising one or more audio source signals (X_{src}) relating to specific positions and additional data (r_{src}), **characterized in that** the additional data define said specific positions to which the audio source signals relate.

2. Audio signal according to claim 1, wherein said additional data (r_{src}) are suitable for calculating a generic audio field representation based on Ambisonics representation.

3. Method for encoding or transmitting an audio signal, comprising the steps of

- providing one or more audio source signals (X_{src});

characterized in the further steps of

- determining for each of said audio source signals a specific position to which it relates;
 - generating data sets containing the determined positions (r_{src}) of the audio sources; and
 - encoding or transmitting (62) the data sets together with said audio source signals.

4. Method according to claim 3, wherein said data sets are suitable for calculating a generic audio field representation based on Ambisonics representation.

5. Method according to claim 3 or 4, further comprising the steps of

- determining the number (O_s) of source channels that are required for Ambisonics encoding, said number being $(2N+1)$ for the 2D case and $(N+1)^2$ for the 3D case, where N is the order of Ambisonics encoding;
 - determining the number (O_c) of available transmission or storage channels; and
 - comparing the number (O_s) of source channels required for Ambisonics encoding of the order N with the number (O_c) of available transmission or storage channels;
 - depending on said comparison, generating a mode decision information (MD), having a first value if the number (O_s) of source channels required for Ambisonics encoding of the order N is not less than the number of available transmission or storage channels (O_c), or having a different second value otherwise; and
 - generating (HOA_e, 64), and storing or transmitting, the Ambisonics encoded version (61) of the audio signal if the mode decision information (MD, 65) has said first value, or otherwise storing or transmitting the received or retrieved audio signal (60).

6. Method according to claim 5, wherein the Ambisonics encoded version (61) of the audio signal is transmitted, further comprising the step of transmitting said mode decision information (MD).

7. Method according to claim 5, further comprising the steps of

- determining a number of sources, such as microphones, or target reproduction channels, such as loudspeakers;

- selecting N being the order of HOA encoding as a function of the determined number of sources or reproduction channels.

8. Method according to claim 5, wherein N is determined from the number (O_c) of available transmission or storage channels.

9. Method for processing an audio signal, comprising the steps of

- receiving or retrieving from storage an encoded audio signal (60);

characterized in the further steps of

- extracting (63) first audio source signals (X'_{src}) and additional information (r'_{src}) from the received or retrieved signal, wherein the first audio source signals (X'_{src}) relate to first audio source positions provided by the additional information (r'_{src});

- transforming (TRC) the first audio source signals relating to first audio source positions into second audio source signals relating to different second audio source positions; and

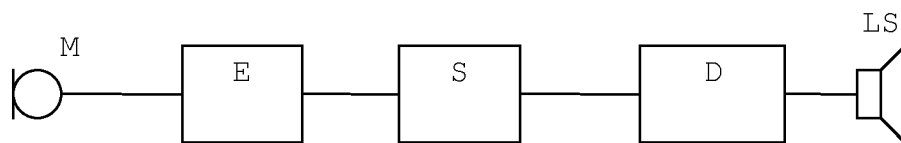
- supplying (LSA) said second audio source signals for storage or playback.

10. Method according to claim 9, wherein a generic audio field representation is calculated.

11. Method according to claim 10, wherein the generic audio field representation is an Ambisonics representation (HOA) of an order higher than one.

12. Apparatus for encoding or transmitting an audio signal, using a method according to claims 3-8.

13. Apparatus for processing an audio signal, using a method according to any of the claims 9-11.



M=Micro E=Encoding S=Storage D=Decoding LS=Loudspeaker

Fig.1

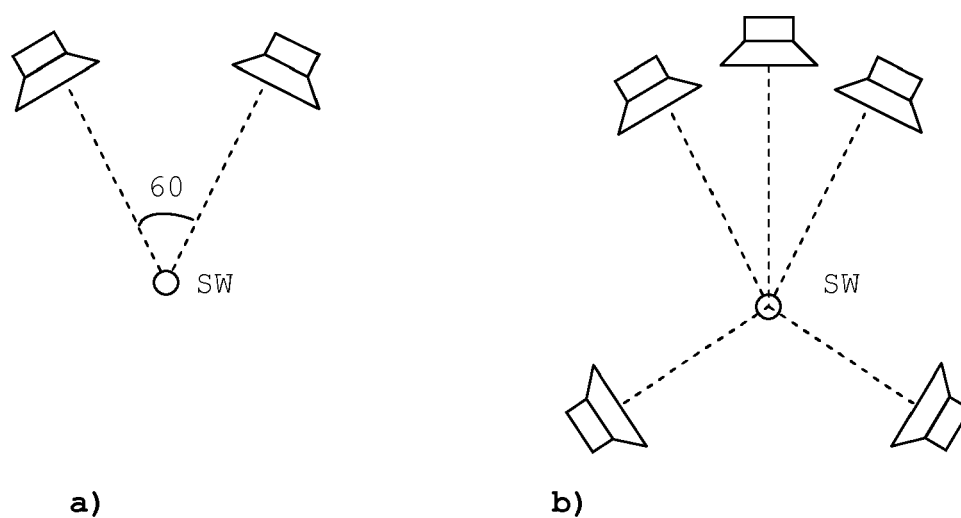


Fig.2

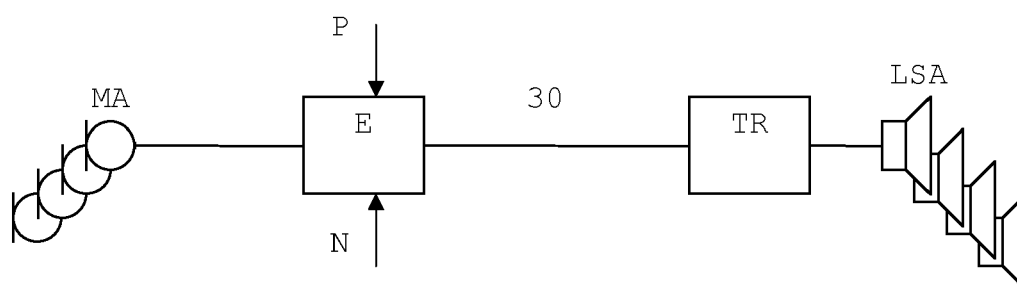


Fig.3

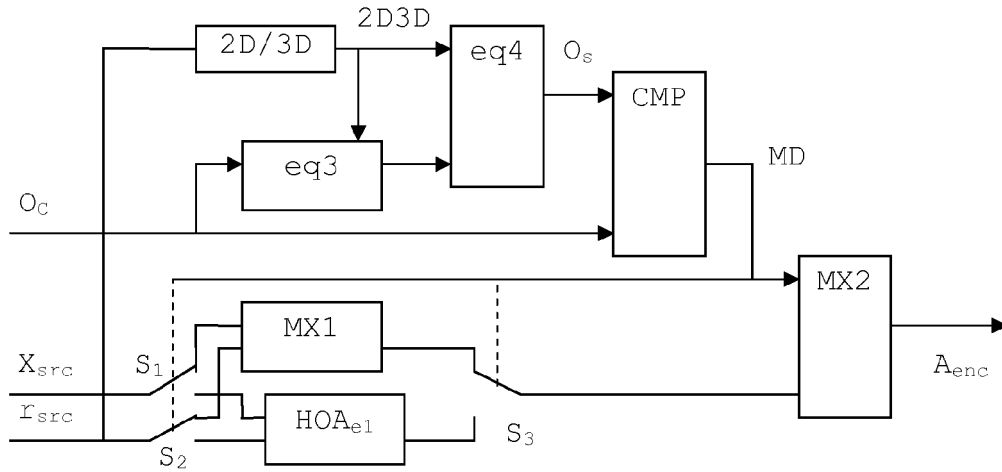


Fig. 4

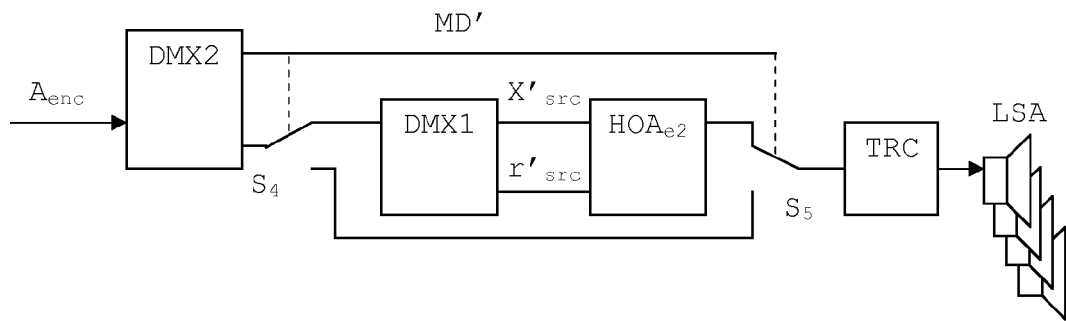


Fig. 5

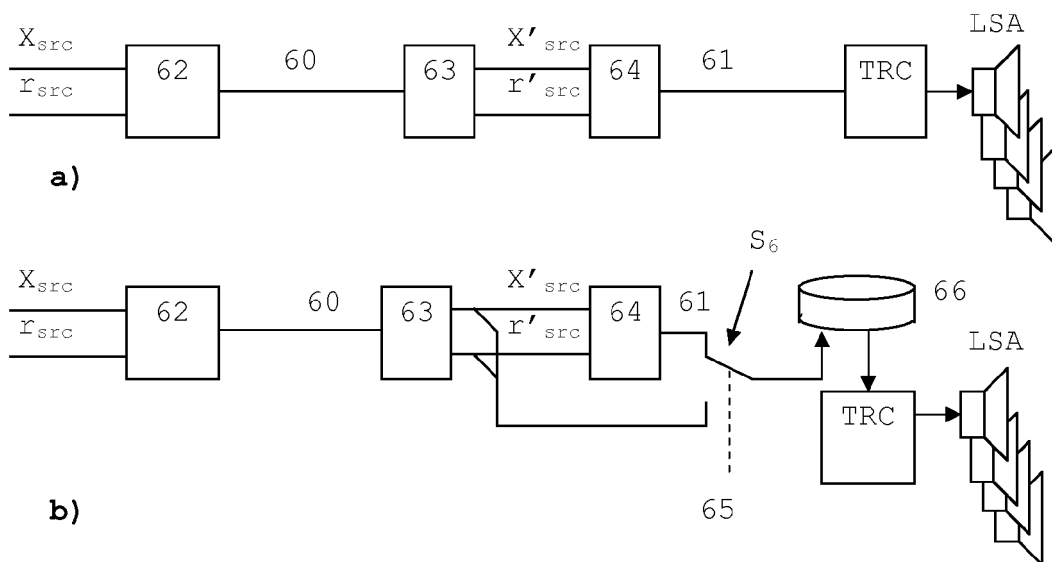


Fig. 6

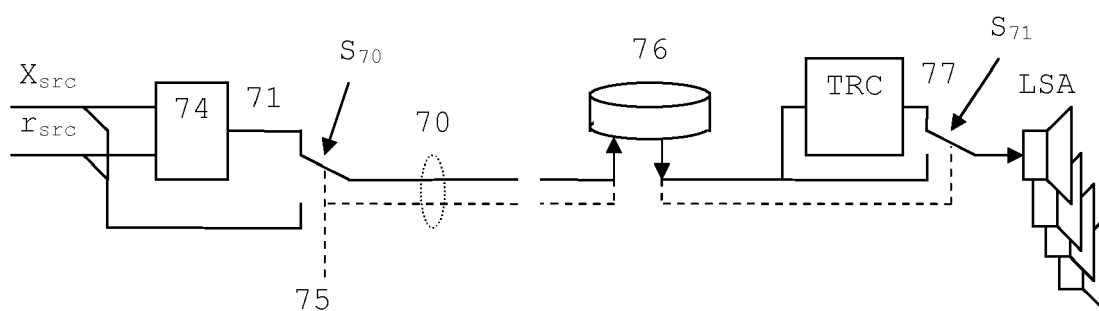


Fig.7



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 08 10 1732

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2005/141728 A1 (MOORER JAMES A [US]) 30 June 2005 (2005-06-30)	1-4,9-13	INV. H04S3/02
Y	* abstract; claims 10-12; figures 1-7 * * paragraphs [0002], [0006] - [0014], [0028], [0031] - [0036], [0045], [0046], [0048], [0050], [0055] - [0058], [0065], [0077], [0087], [0091], [0110] *	5-8	
X	FR 2 847 376 A (FRANCE TELECOM [FR]) 21 May 2004 (2004-05-21)	1-4,9-13	
A	* abstract; figures 1,2,7,8 * * page 2, line 12 - page 3, line 19 * * page 8, line 5 - line 19 * * page 9, line 28 - page 10, line 10 * * page 11, line 25 - page 12, line 9 * * page 14, line 16 - line 30 * * page 17, line 20 - page 18, line 24 * * page 19, line 23 - page 20, line 5 * * page 24, line 1 - line 14 * * page 28, line 18 - page 29, line 15 * * page 30, line 21 - line 26 * * page 38, line 6 - page 39, line 21 * * page 54, line 1 - line 16 *	5-8	TECHNICAL FIELDS SEARCHED (IPC) H04S G10L
X	US 2007/269063 A1 (GOODWIN MICHAEL [US] ET AL) 22 November 2007 (2007-11-22)	1-4,9-13	
Y	* abstract; figures 1-3,11,13,15 * * paragraphs [0005], [0026], [0027], [0029], [0047] - [0050], [0056] - [0061], [0097], [0098], [0112] *		
Y	EP 1 515 308 A (NOKIA CORP [FI]) 16 March 2005 (2005-03-16)	5-8	
	* abstract; figures 3-5 * * paragraphs [0008], [0024], [0042] - [0044], [0052] - [0054] *		
The present search report has been drawn up for all claims			
Place of search The Hague		Date of completion of the search 3 July 2008	Examiner Scappazzoni, E
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

2
EPO FORM 1503 03.82 (P04C01)



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 08 10 1732

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	WO 2007/136187 A (KOREA ELECTRONICS TELECOMM [KR]; LEE YONG-JU [KR]; LEE TAE-JIN [KR]; Y) 29 November 2007 (2007-11-29)	1-4,9, 10,12,13	TECHNICAL FIELDS SEARCHED (IPC)
A	* abstract; figures 2-4 * * page 1, line 5 - line 14 * * page 4, line 7 - page 5, line 23 * * page 6, line 6 - line 14 * * page 7, line 1 - page 8, line 26 * * page 9, line 7 - line 28 *	5	
X	DE 10 2005 008366 A1 (FRAUNHOFER GES FORSCHUNG [DE]) 24 August 2006 (2006-08-24)	1-4,9, 10,12,13	
A	* abstract; figures 1-5 * * paragraphs [0001], [0036], [0037], [0041], [0054], [0055], [0062], [0069], [0070], [0076] *	11	
X	EP 1 416 769 A (KOREA ELECTRONICS TELECOMM [KR]) 6 May 2004 (2004-05-06) * abstract; figures 1-9 * * paragraphs [0002], [0009], [0015] - [0021], [0026], [0033] - [0035], [0044] - [0046], [0050] - [0059] *	1-4,9, 10,12,13	
X	US 2005/117753 A1 (MIURA MASAYOSHI [JP] ET AL) 2 June 2005 (2005-06-02)	1-4,12	
A	* abstract; claim 1; figures 1,2A,2B,2C,6 * * paragraphs [0012], [0031], [0032] *	9,10,13	
The present search report has been drawn up for all claims			
Place of search The Hague		Date of completion of the search 3 July 2008	Examiner Scappazzoni, E
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

2
EPO FORM 1503 03.82 (P04C01)

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

EP 08 10 1732

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

03-07-2008

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2005141728 A1	30-06-2005	NONE	
FR 2847376 A	21-05-2004	AT 322065 T	15-04-2006
		AU 2003290190 A1	18-06-2004
		BR 0316718 A	18-10-2005
		CN 1735922 A	15-02-2006
		DE 60304358 T2	07-12-2006
		EP 1563485 A1	17-08-2005
		ES 2261994 T3	16-11-2006
		WO 2004049299 A1	10-06-2004
		JP 2006506918 T	23-02-2006
		KR 20050083928 A	26-08-2005
		US 2006045275 A1	02-03-2006
		ZA 200503969 A	27-09-2006
US 2007269063 A1	22-11-2007	NONE	
EP 1515308 A	16-03-2005	DE 602004004219 T2	11-10-2007
		US 2005055203 A1	10-03-2005
WO 2007136187 A	29-11-2007	NONE	
DE 102005008366 A1	24-08-2006	CN 101129089 A	20-02-2008
		EP 1844628 A1	17-10-2007
		WO 2006089684 A1	31-08-2006
		US 2008123864 A1	29-05-2008
EP 1416769 A	06-05-2004	KR 20040037437 A	07-05-2004
		US 2004111171 A1	10-06-2004
US 2005117753 A1	02-06-2005	CN 1625302 A	08-06-2005
		JP 2005167612 A	23-06-2005
		KR 20050053313 A	08-06-2005

REFERENCES CITED IN THE DESCRIPTION

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

Non-patent literature cited in the description

- **K.Hamasaki ; T. Nishiguchi ; R.Okumaura ; Y.Nakayama.** Wide listening area with exceptional spatial sound quality of a 22.2 multichannel sound system. *Audio Engineering Society Preprints*, May 2007 **[0002]**
- **Jerome Daniel.** *Spatial Sound Encoding Including Near Field Effect: Introducing distance Coding Filters and a Viable*, 2003 **[0003]**