

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 July 2008 (17.07.2008)

PCT

(10) International Publication Number
WO 2008/084436 A1

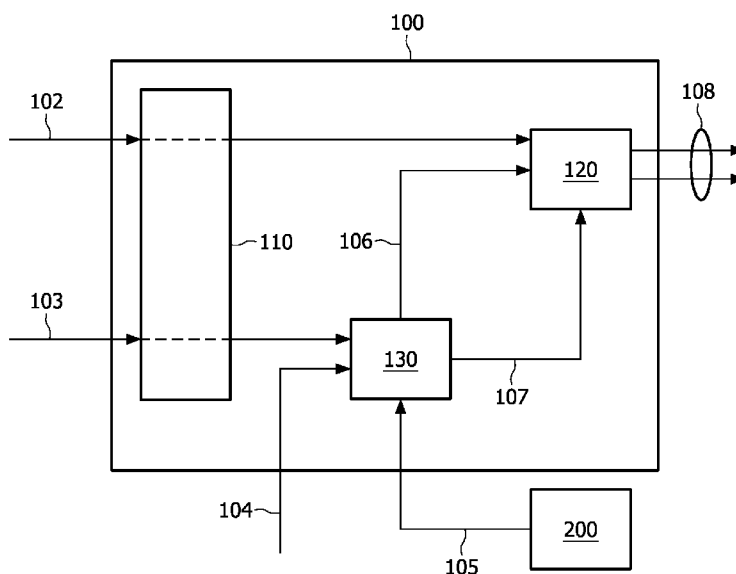
- (51) International Patent Classification:
G10L 19/00 (2006.01) *H04S 7/00* (2006.01)
- (21) International Application Number:
PCT/IB2008/050041
- (22) International Filing Date: 8 January 2008 (08.01.2008)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
07100343.8 10 January 2007 (10.01.2007) EP
- (71) Applicant (for all designated States except US): **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL];
Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **BREEBAART, Dirk, J.** [NL/NL]; c/o High Tech Campus 44, NL-5656 AE Eindhoven (NL).
- (74) Agents: **GROENENDAAL, Antonius, W., M.** et al.;
High Tech Campus 44, NL-5656 AE Eindhoven (NL).

- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, IS, JP, KE, KG, KM, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

Published:
— with international search report

(54) Title: AN OBJECT-ORIENTED AUDIO DECODER



(57) Abstract: An object-oriented audio decoder (100) comprising: conversion means and decoding means. The conversion means (130) convert received parametric data and received head-related transfer function parameters into spatial parameters and dynamic head-related transfer function parameters. Said parametric data comprises a plurality of object parameters for each of the plurality of audio objects. The decoding means (120) in accordance with the binaural MPEG Surround standard decode and render the audio objects from the received down-mix audio signals based on the spatial parameters and the dynamic head-related transfer function parameters. Said spatial parameters and said dynamic head-related transfer function parameters are provided from the conversion means. Said received down-mix 10 signals comprise a down-mix of a plurality of audio objects.

WO 2008/084436 A1

An object-oriented audio decoder

TECHNICAL FIELD

The invention relates to an object-oriented audio decoder comprising a binaural MPEG Surround decoder.

5

TECHNICAL BACKGROUND

In (parametric) spatial audio (en)coders, parameters are extracted from the original audio signals so as to produce a reduced number of down-mix audio signals (for example only a single down-mix signal corresponding to a mono, or two down-mix signals for a stereo down-mix signal), and a corresponding set of parameters describing the spatial properties of the original multi-channel audio signal. In (parametric) spatial audio decoders, the spatial properties described by the transmitted spatial parameters are used to recreate the original spatial multi-channel signal, which closely resembles the original audio signal.

Recently, techniques for processing and manipulating of individual audio objects at the decoding side have attracted significant interest. For example, within the MPEG framework, a workgroup has been started on object-based spatial audio coding. The aim of this workgroup is to “explore new technology and reuse of current MPEG Surround components and technologies for the bit rate efficient coding of multiple sound sources or objects into a number of down-mix channels and corresponding spatial parameters”. In other words, the aim is to encode multiple audio objects in a limited set of down-mix channels with corresponding parameters. At the decoder side, users interact with the content for example by repositioning the individual objects.

Such interaction with the content is easily realized in object-oriented decoders. It is then realized by including a rendering step that follows the decoding process. Said rendering is combined with the decoding as a single processing step to prevent the need of determining individual objects. For loudspeaker playback, such combination is described in Faller, C., “Parametric joint-coding of audio sources”, Proc. 120th AES Convention, Paris, France, May 2006. For headphone playback, an efficient combination of decoding and head-related transfer function processing is described in Breebaart, J., Herre, J., Villemoes, L., Jin,

C., Kjörling, K., Plogsties, J., Koppens, J. (2006), "Multi-channel goes mobile: MPEG Surround binaural rendering", Proc. 29th AES conference, Seoul, Korea. From the point of view of reuse it is preferable to use an existing MPEG Surround decoder as rendering engine for the object-oriented audio decoder. For headphone playback, MPEG Surround features a dedicated binaural decoding mode that generates a three-dimensional sound scene over the conventional headphones. However, the binaural MPEG Surround decoder has a disadvantage that the decoding process is defined specifically for only five virtual loudspeakers. This means that all objects from the down-mix audio signal can be mapped to only these positions where those five virtual loudspeakers are located.

10

SUMMARY OF THE INVENTION

It is an object of the invention to provide an enhanced object-oriented decoder for headphones playback that allows an arbitrary virtual positioning of objects in a space.

This object is achieved by an object-oriented audio decoder according to the invention. It is assumed that a set of objects, each with its corresponding waveform, has previously been encoded in an object-oriented encoder, which generates a down-mix audio signal (a single signal in case of a single channel), said down-mix audio signal being a down-mix of a plurality of audio objects characterized by corresponding parametric data. The parametric data comprises a set of object parameters for each of the different audio objects.

15

The receiver receives said down-mix audio signal and said parametric data. This down-mix audio signal is further fed into a decoding means, said decoding means being in accordance with the binaural MPEG Surround standard. These decoding means perform both decoding and rendering of the audio objects from the down-mix audio signals based on the spatial parameters and dynamic head-related transfer function parameters. The dynamic head-related

20

transfer function parameters are provided from an outside of the decoding means. The conversion means convert the parametric data and head-related transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters. The head-related transfer function parameters are provided from the outside of the object-oriented decoder. The decoding means generate a spatial output audio signal from the audio objects to be played back over e.g. headphones.

25

The advantage of the object-oriented audio decoder according to the invention is that the head-related transfer function parameters are moved out of the object-oriented audio decoder comprising decoding means in accordance with the binaural MPEG Surround standard, therefore freeing the object-oriented audio decoder from the inherent limitation of a

predetermined maximum of the virtual locations of objects, when the head-related transfer function parameters are hard-wired in the binaural MPEG Surround decoder. The corresponding conversion of the parametric data and head-related transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters results in freedom in placing the decoded objects in the space. Furthermore, the advantage of the object-oriented audio decoder according to the invention is that no explicit object decoding is required, and the rendering emerged in the MPEG Surround decoder comprised in the object-oriented audio decoder is preserved.

In an embodiment, the objects can be virtually placed at any position in a space by manipulating of the head-related transfer function parameters.

In an embodiment, converting the parametric data and head-related transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters is combined with generation of binaural parameters as used by decoding means in one step. Said decoding means perform decoding in accordance with the binaural MPEG Surround standard. Said decoding means use the binaural parameters for decoding. Said binaural parameters being derived based on the spatial parameters and dynamic head-related transfer function parameters. Making such a combination simplifies the implementation of the object-oriented audio decoder, as no intermediate spatial parameters or dynamic head-related transfer function parameters are needed. Instead, the binaural parameters are directly derived from the parametric data and the head-related transfer function parameters.

The invention further provides a receiver and a communication system, as well as corresponding methods.

In an embodiment, the head-related transfer function parameters are set in response to user input. This allows a user to position the objects at any position in the virtual space according to user preferences.

The invention further provides a computer program product enabling a programmable device to perform the method according to the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments shown in the drawings, in which:

Fig 1 schematically shows an object-oriented decoder according to the invention;

Fig 2 shows an example set-up of virtual loudspeakers;

Fig 3 shows a method of decoding according in accordance with some embodiments of the invention;

Fig 4 shows a transmission system for communication of an audio signal in accordance with some embodiments of the invention.

Throughout the figures, same reference numerals indicate similar or corresponding features. Some of the features indicated in the drawings are typically implemented in software, and as such represent software entities, such as software modules or objects.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Fig 1 schematically shows an object-oriented decoder 100 according to the invention. It is assumed that a set of objects, each with its corresponding waveform, has previously been encoded in an object-oriented encoder, which generates a down-mix audio signal 102 (a single signal in case of a single channel, or two signals in case of two channels (= stereo)), said down-mix audio signal 102 being a down-mix of a plurality of audio objects characterized by corresponding parametric data 103. The parametric data 103 comprises a set of object parameters for each of the different audio objects. The receiver 110 receives said down-mix audio signal 102 and said parametric data 103.

Although the down-mix audio signal and parametric data are indicated as a separate signals/data paths, they could be multiplexed into one signal/data stream comprising concatenated down-mix audio data that corresponds to the down-mix audio signal and the parametric data. The function of the receiver is then demultiplexing of the two data streams. If the down-mix audio signal 102 is provided in a compressed form (such as MPEG-1 layer 3), the receiver 110 also performs decompression or decoding of the compressed audio signal into a time-domain audio down-mix signal.

The down-mix audio signal 102 is further fed into decoding means 120, said decoding means 120 being in accordance with the binaural MPEG Surround standard. These decoding means 120 perform both, decoding and rendering of the audio objects from the down-mix audio signals 102 based on the spatial parameters 106 and dynamic head-related transfer function parameters 107. The dynamic head-related transfer function (HRTF) parameters 107 are being provided from an outside of the decoding means 120. The conversion means 130 perform conversion of the parametric data 103 and HRTF parameters 105 into the spatial parameters 106 and the dynamic HRTF parameters 107. The HRTF

parameters 105 are provided from the outside of the object-oriented decoder by the HRTF parameter database 200.

Since decoding means 120 comprise essentially an MPEG Surround decoder, the spatial parameters 106 are preferably provided in MPEG Surround format; furthermore
5 the dynamic head-related transfer functions 107 are preferably provided in the corresponding MPEG Surround format for HRTF parameters.

Optionally, the conversion means can use a user-control data 104 in order to generate the spatial parameters 106 and the dynamic head-related transfer function parameters 107. Said user-control data 104 comprises data concerned with rendering of the
10 audio objects. For example, the user-control data 104 may indicate the desired spatial position (for example in terms of elevation and azimuth) of one or more audio objects. In that case, conversion means 130 select HRTF parameters 105 that correspond to the desired position from HRTF database 200. If the desired position is not directly available in HRTF parameter database 200, interpolation of parameters may be required that correspond to
15 HRTF database positions surrounding the desired position.

The decoding means 120 generate a spatial output audio signal 108 from the audio objects to be played back at e.g. the headphones.

For a person skilled in art, it is known that for headphones playback the MPEG Surround standard comprises a dedicated binaural decoding mode that generates a
20 three-dimensional (3D) sound scene over conventional stereo headphones. The conventional 3D synthesis as known from MPEG Surround requires audio configured up to five loudspeakers, as currently standardized, which is further convolved with HRTFs followed by summation of the convolved signals to result in a binaural output signal pair. The advantage of MPEG Surround is that it provides means to perform multi-channel HRTF processing in
25 the down-mix domain, without multi-channel audio as intermediate step. This is realized by converting HRTF parameters to the parameter domain, and subsequently computing binaural parameters from the combined spatial parameters and HRTF parameters. Detailed information about the said parameter conversion is provided in Breebaart, J., Herre, J., Vиллемoes, L., Jin, C., Kjörling, K., Plogsties, J., Koppens, J. (2006). Multi-channel goes
30 mobile: MPEG Surround binaural rendering. Proc. 29th AES conference, Seoul, Korea. The HRTF parameters are provided to the MPEG Surround decoder, however the end user can change the actual parameter values. This is the case when the user wants to use his/her personalization options.

The HRTF parameters comprise: a level of the left-ear channel output, a level of the right-ear channel output, the phase difference between the left and right-ear binaural output, and optionally a coherence between the left and right-ear channel output. The level parameters are defined as a change in a level with respect to the original input signal level.

5 The HRTF parameters are defined as a function of spatial position of a sound source, either by a mathematical model or by using a database 200.

The resulting binaural parameters are: a level of a left-channel binaural output, a level of a right-channel binaural output, the phase difference between the left and right-channel outputs, and the coherence between the left and right-channel output. Preferably, the
10 level parameters are expressed relatively to the level(s) of the down-mix signal(s).

The proposed object-oriented decoder overcomes however the limitation of audio up to five loudspeakers only. In the MPEG Surround decoder, HRTF parameters represent the binaural properties of each virtual object separately, i.e., one HRTF parameter set for each virtual loudspeaker. The MPEG Surround decoder combines the binaural
15 properties of each virtual loudspeaker into binaural parameters of the (up to) five virtual loudspeakers simultaneously, with the help of spatial parameters 106 that describe the spatial relations between the virtual loudspeaker signals. This process is only defined for (up to) five virtual loudspeakers. The HRTF parameters and binaural parameters, however, have very similar representations. It is therefore possible to estimate binaural parameters of a complex
20 auditory scene comprising more than five virtual loudspeakers by conversion means 130, and providing the resulting binaural parameters as dynamic HRTF parameters to MPEG Surround decoder 120. In the proposed object-oriented decoder, the parametric data 103 and user control data 104 are combined together in the process of computing the binaural parameters, which are fed into the HRTF parameter input of decoding means 120 in a dynamic fashion.
25 At the same time the spatial parameters are computed and fed into the decoding means to be used to map each down-mix channel to one of the possible configurations for five virtual loudspeakers.

Fig 2 shows an example set-up of virtual loudspeakers. The object-oriented audio decoder is adapted to use virtual loudspeakers at a center 320, left-front 310, right-front
30 330, left-surround 340, and right-surround 350 positions, whereby the user is positioned at the location 400. In such a set-up, in case of a mono down-mix signal, this down-mix signal is mapped e.g. to the center 320 loudspeaker of the virtual loudspeaker set-up. For the stereo down-mix signal, the signal corresponding to the left channel can be mapped to the left-front

310 loudspeaker, while the signal corresponding to the right channel can be mapped to the right-front 330 loudspeaker.

In an embodiment, the objects can be virtually placed at any position in a space through manipulating of the head-related transfer function parameters. The HRTF parameters corresponding to e.g. the center loudspeaker are set to represent a combination of a multiple objects placed differently in the space. In other words, despite using the five virtual speakers the sound can be generated that can be perceived to originate from objects placed arbitrarily in space and not located at the loudspeaker locations. This advantage of the proposed solution is realized by the HRTF parameters being taken out of control of the decoding means comprising the MPEG Surround decoder, therefore allowing an arbitrary modification of the HRTF parameters to influence the placing of object arbitrarily in the space.

In an embodiment, the head-related transfer function parameters comprise at least left-ear magnitude, right-ear magnitude, and phase difference, respectively, $p_{l,i}$, $p_{r,i}$, ϕ_i . Said parameters correspond to the perceived position of an object. Optionally the coherence parameter ρ_i can also be comprised in the head-related transfer function parameters. If this parameter is not specified, it can be assumed to be equal to +1, or derived from the default-parameter value table as specified in the MPEG Surround standard.

Fig 3 shows a method of decoding according in accordance with some embodiments of the invention. The step 510 comprises receiving at least one down-mix audio signal and parametric data. Each down-mix audio signal comprises a down-mix of a plurality of audio objects. The parametric data comprises a plurality of object parameters for each of the plurality of audio objects.

The step 520 comprises decoding and rendering the audio objects from the down-mix audio signals based on the spatial parameters and dynamic head-related transfer function parameters. The dynamic head-related transfer function parameters being provided from an outside of decoding means. The decoding means are performing decoding in accordance with the binaural MPEG Surround standard.

The step 530 comprises converting the parametric data and head-related transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters. The steps 520 and 530 can be performed in the reversed sequence or simultaneously. The synchronization in time of the said parameters with the down-mix signal should be taken care off.

In an embodiment, the converting the parametric data and head-related transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters is combined with generation of binaural parameters as used by decoding means in one step. Said decoding means perform decoding in accordance with the binaural
 5 MPEG Surround standard. Further, said decoding means use the binaural parameters for decoding, where said binaural parameters are based on the spatial parameters and dynamic head-related transfer function parameters.

In an embodiment, the binaural parameters as used by the decoding means have the same format as the head-related transfer function parameters. The binaural
 10 parameters comprise the left-ear magnitude, the right-ear magnitude, the phase difference, and coherence parameter, respectively, $p_{l,b}$, $p_{r,b}$, ϕ_b , and ρ_b .

The multiple objects are comprised as object (subband – i.e., a set of band-limited and possibly down-sampled) signals x_i in a down-mix signal s :

$$s = \sum_i x_i .$$

15 The parametric data represents the power σ_i^2 of each object signal x_i within the down-mix signal s :

$$\sigma_i^2 = \langle x_i x_i^* \rangle ,$$

where $\langle \cdot \rangle$ is the expected value operator and x^* is the complex conjugate of x .

In an embodiment, the binaural parameters are derived from the parametric
 20 data and head-related transfer function parameters according to:

$$p_{l,b} = \sqrt{\sum_i p_{l,i}^2 \sigma_i^2} ,$$

$$p_{r,b} = \sqrt{\sum_i p_{r,i}^2 \sigma_i^2} ,$$

$$\phi_b = \angle \left(\sum_i e^{j\phi_i} \rho_i p_{l,i} p_{r,i} \sigma_i^2 \right) ,$$

$$\rho_b = \frac{\sum_i e^{j\phi_i} \rho_i p_{l,i} p_{r,i} \sigma_i^2}{p_{l,b} p_{r,b}} ,$$

25 wherein $p_{l,i}$, $p_{r,i}$, ϕ_i and ρ_i , are the head-related transfer function parameters corresponding respectively to a left-ear magnitude, a right-ear magnitude, a phase difference and a coherence parameter corresponding to a perceived position of object i , and $p_{l,b}$, $p_{r,b}$, ϕ_b and ρ_b , are the binaural parameters corresponding respectively to a left-ear magnitude, a

right-ear magnitude, a phase difference and a coherence parameter, said binaural parameters being representative for an object or a plurality of objects, and $\angle(\cdot)$ represents the complex phase angle operator:

$$5 \quad \angle(X) = \arctan\left(\frac{\text{Im}(X)}{\text{Re}(X)}\right).$$

The above formulas comprise summation over i , where i is the index of the objects comprised in the down-mix signal. In case of mono down-mix signal, these binaural parameters representative for a multitude of objects are fed into the decoding means
 10 comprising the MPEG Surround decoder as parameters corresponding to a single audio channel.

In case of stereo down-mix signal, each down-mix signal, each corresponding to one of the two channels, is mapped to a single multi-channel output, e.g. to the left-front and right-front virtual loudspeakers. Subsequently, the binaural parameters for both the left-
 15 front and right-front virtual loudspeakers as indicated by the above formulas. The difference with the mono down-mix signal case is that the parametric data σ is different for the two down-mix channels because one or more objects may be present predominantly in only one of the down-mix signals.

In an embodiment a receiver for receiving audio signals comprises: a receiver
 20 element, decoding means, and conversion means. The receiver element is receiving from a transmitter at least one down-mix audio signal and parametric data. Each down-mix audio signal comprises a down-mix of a plurality of audio objects. Said parametric data comprises a plurality of object parameters for each of the plurality of audio objects.

The decoding means are in accordance with the binaural MPEG Surround
 25 standard and are decoding and rendering the audio objects from the down-mix audio signals based on the spatial parameters and dynamic head-related transfer function parameters. Said dynamic head-related transfer function parameters are provided from an outside of the decoding means.

The conversion means are converting the parametric data and head-related
 30 transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters.

Fig 4 shows a transmission system for communication of an audio signal in accordance with some embodiments of the invention. The transmission system comprises a

transmitter 600, which is coupled with a receiver 800 through a network 700. The network 700 could be e.g. Internet.

The transmitter 600 is for example a signal recording device and the receiver 800 is for example a signal player device.

5 In the specific example when a signal recording function is supported, the transmitter 600 comprises means 610 for receiving a plurality of audio objects. Consequently, these objects are encoded by encoding means 620 for encoding the plurality of audio objects in at least one down-mix audio signal and parametric data. An example of such encoding means 620 is given in Faller, C., "Parametric joint-coding of audio sources", Proc. 120th AES
10 Convention, Paris, France, May 2006. Each down-mix audio signal comprises a down-mix of a plurality of audio objects. Said parametric data comprises a plurality of object parameters for each of the plurality of audio objects. The encoded audio objects are transmitted to the receiver 800 by means 630 for transmitting down-mix audio signals and the parametric data. Said means 630 have an interface with the network 700, and may transmit the down-mix
15 signals through the network 700.

The receiver comprises a receiver element 810 for receiving from the transmitter 600 at least one down-mix audio signal and parametric data. Each down-mix audio signal comprises a down-mix of a plurality of audio objects. Said parametric data comprises a plurality of object parameters for each of the plurality of audio objects.

20 The received down-mix audio signal is decoded by decoding means 830 in accordance with the binaural MPEG Surround standard. The decoding means perform decoding and rendering of the audio objects from the down-mix audio signals based on the spatial parameters and dynamic head-related transfer function parameters. Said dynamic head-related transfer function parameters are provided from an outside of the decoding means
25 830.

The conversion means 820 perform converting the parametric data and head-related transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters.

30 In an embodiment, the head-related transfer function parameters are set in response to user input. The user can by means of e.g. button, slider, knob, or graphical user interface, set the HRTF parameters according to own preferences.

It should be noted that the above-mentioned embodiments illustrate rather than limit the invention, and those skilled in the art will be able to design many alternative embodiments without departing from the scope of the appended claims.

In the accompanying claims, any reference signs placed between parentheses shall not be construed as limiting the claim. The word “comprising” does not exclude the presence of elements or steps other than those listed in a claim. The word “a” or “an” preceding an element does not exclude the presence of a plurality of such elements. The
5 invention can be implemented by means of hardware comprising several distinct elements, and by means of a suitably programmed computer.

CLAIMS:

1. An object-oriented audio decoder (100) comprising:
conversion means (130) for converting received parametric data and received head-related transfer function parameters into spatial parameters and dynamic head-related transfer function parameters, said parametric data comprising a plurality of object parameters
5 for each of the plurality of audio objects;
decoding means (120) in accordance with the binaural MPEG Surround standard for decoding and rendering the audio objects from the received down-mix audio signals based on the spatial parameters and the dynamic head-related transfer function parameters, said spatial parameters and said dynamic head-related transfer function
10 parameters being provided from the conversion means, said received down-mix signals comprising a down-mix of a plurality of audio objects.
2. An object-oriented audio decoder as claimed in claim 1, wherein said object-oriented audio decoder is adapted to use virtual loudspeakers at a center (320), left-front
15 (310), right-front (330), left-surround (340), and right-surround (350) positions.
3. An object-oriented audio decoder as claimed in claim 1, wherein the objects can be virtually placed at any position in a space through manipulating of the head-related transfer function parameters.
20
4. An object-oriented audio decoder as claimed in claim 1, wherein the head-related transfer function parameters comprise at least left-ear magnitude, right-ear magnitude, and phase difference, said parameters corresponding to the perceived position of an object.
- 25 5. A method of decoding audio signals comprising:
receiving (510) at least one down-mix audio signal and parametric data, each down-mix audio signal comprising a down-mix of a plurality of audio objects, said parametric data comprising a plurality of object parameters for each of the plurality of audio objects;

decoding and rendering (520) the audio objects from the down-mix audio signals based on the spatial parameters and dynamic head-related transfer function parameters, said dynamic head-related transfer function parameters being provided from an outside of decoding means, said decoding in accordance with the binaural MPEG Surround standard;

converting (530) the parametric data and head-related transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters.

6. A method as claimed in claim 5, wherein the converting the parametric data and head-related transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters is combined with generation of binaural parameters as used by decoding means in one step; said decoding means performing decoding in accordance with the binaural MPEG Surround standard; said decoding means using the binaural parameters for decoding; said binaural parameters being based on the spatial parameters and dynamic head-related transfer function parameters.

7. A method as claimed in claim 6, wherein the binaural parameters as used by the decoding means have the same format as the head-related transfer function parameters.

8. A method as claimed in claim 7, wherein the binaural parameters are derived from the parametric data and head-related transfer function parameters according to:

$$p_{l,b} = \sqrt{\sum_i p_{l,i}^2 \sigma_i^2} ,$$

$$p_{r,b} = \sqrt{\sum_i p_{r,i}^2 \sigma_i^2} ,$$

$$\phi_b = \angle \left(\sum_i e^{j\phi_i} \rho_i p_{l,i} p_{r,i} \sigma_i^2 \right) ,$$

$$\rho_b = \frac{\sum_i e^{j\phi_i} \rho_i p_{l,i} p_{r,i} \sigma_i^2}{p_{l,b} p_{r,b}} ,$$

wherein $p_{l,i}$, $p_{r,i}$, ϕ_i and ρ_i are the head-related transfer function parameters corresponding respectively to a left-ear magnitude, a right-ear magnitude, a phase difference and a coherence parameter corresponding to a perceived position of object i , and $p_{l,b}$, $p_{r,b}$, ϕ_b ,

and p_b , are the binaural parameters corresponding respectively to a left-ear magnitude, a right-ear magnitude, a phase difference and a coherence parameter, said binaural parameters being representative for an object or a plurality of objects.

5 9. A receiver for receiving audio signals, the receiver comprising the object-oriented audio decoder of claim 1 and a receiver element (110) for receiving from a transmitter at least one down-mix audio signal and parametric data, each down-mix audio signal comprising a down-mix of a plurality of audio objects, said parametric data comprising a plurality of object parameters for each of the plurality of audio objects, the receiver element
10 being coupled to the conversion means (130) and the decoding means (120) of the object-oriented decoder (100).

10. A communication system for communicating audio signals, the communication system comprising:
15 a transmitter (600) comprising:
means (610) for receiving a plurality of audio objects,
encoding means (620) for encoding the plurality of audio objects in at least one down-mix audio signal and parametric data, each down-mix audio signal comprising a down-mix of a plurality of audio objects, said parametric data comprising a plurality of
20 object parameters for each of the plurality of audio objects, and
means (630) for transmitting down-mix audio signals and the parametric data to the receiver as claimed in claim 9.

11. A method of receiving audio signals, the method comprising:
25 receiving from a transmitter at least one down-mix audio signal and parametric data, each down-mix audio signal comprising a down-mix of a plurality of audio objects, said parametric data comprising a plurality of object parameters for each of the plurality of audio objects;
decoding and rendering the audio objects from the down-mix audio signals
30 based on the spatial parameters and dynamic head-related transfer function parameters, said dynamic head-related transfer function parameters being provided from an outside of the decoding means, said decoding in accordance with the binaural MPEG Surround standard;
converting the parametric data and head-related transfer function parameters into the spatial parameters and the dynamic head-related transfer function parameters.

12. A method of transmitting and receiving audio signals, the method comprising:
at a transmitter performing the steps of:
receiving a plurality of audio objects,
5 encoding the plurality of audio objects in at least one down-mix audio signal
and parametric data, each down-mix audio signal comprising a down-mix of a plurality of
audio objects, said parametric data comprising a plurality of object parameters for each of the
plurality of audio objects, and
transmitting down-mix audio signals and the parametric data to a receiver; and
10 at the receiver performing the steps of:
receiving from a transmitter at least one down-mix audio signal and parametric
data, each down-mix audio signal comprising a down-mix of a plurality of audio objects, said
parametric data comprising a plurality of object parameters for each of the plurality of audio
objects;
15 decoding and rendering the audio objects from the down-mix audio signals
based on the spatial parameters and dynamic head-related transfer function parameters, said
dynamic head-related transfer function parameters being provided from an outside of the
decoding means, said decoding in accordance with the binaural MPEG Surround standard;
converting the parametric data and head-related transfer function parameters
20 into the spatial parameters and the dynamic head-related transfer function parameters.
13. A method claimed in any of claims 6-8 , 11, and 12, wherein the head-related
transfer function parameters are set in response to user input.
- 25 14. A computer program product for executing the method of any claims 6-8, 11,
and 12.
15. An audio playing device comprising an object-oriented audio decoder
according to claim 1.

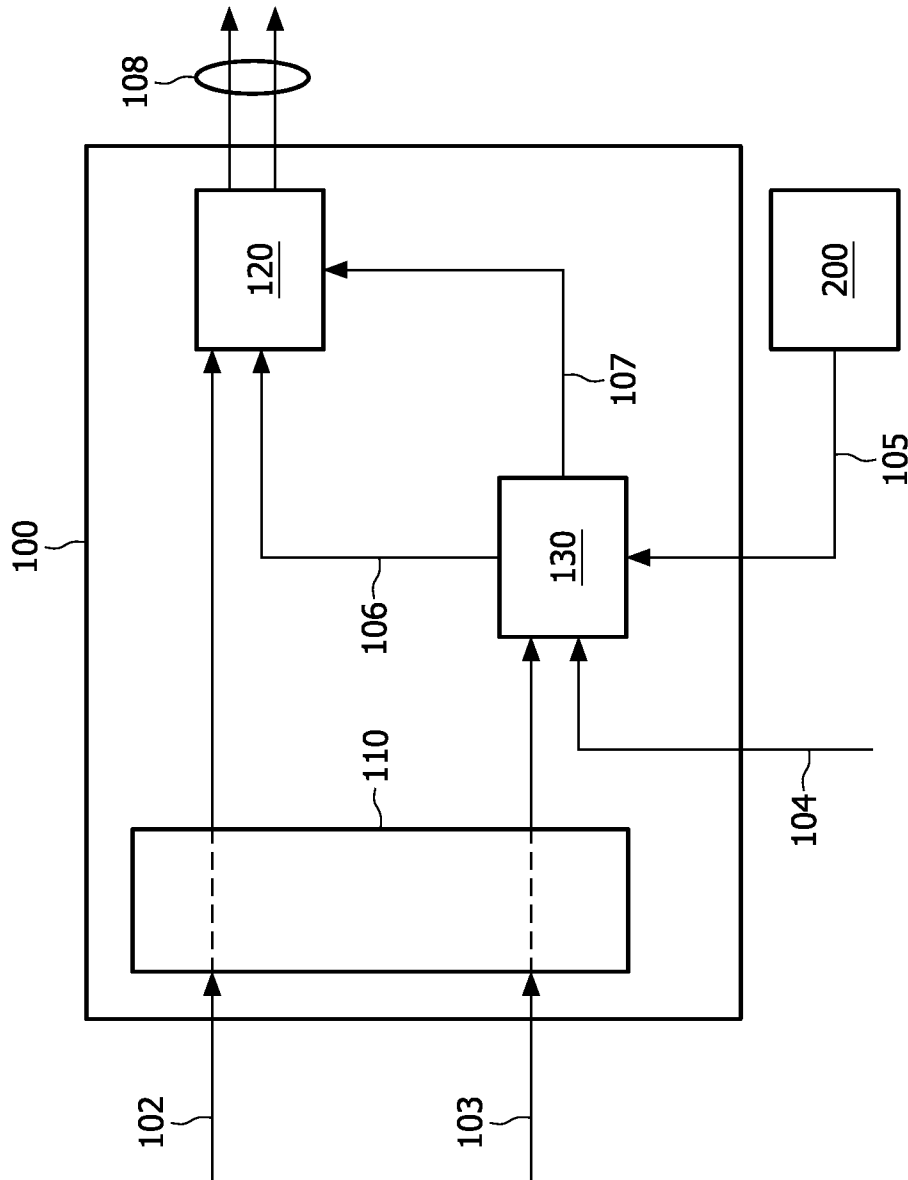


FIG. 1

2/3

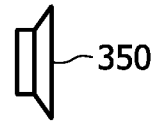
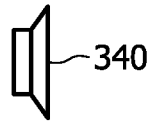
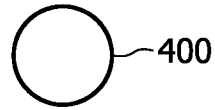
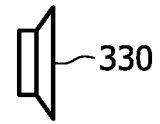
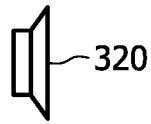
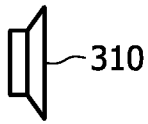


FIG. 2

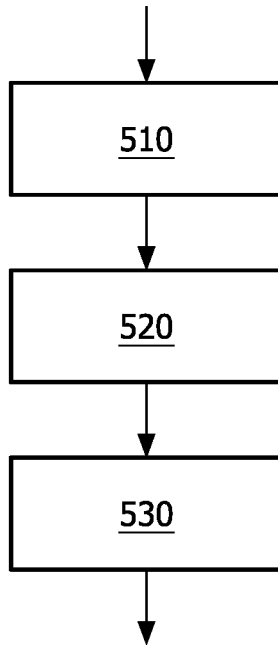


FIG. 3

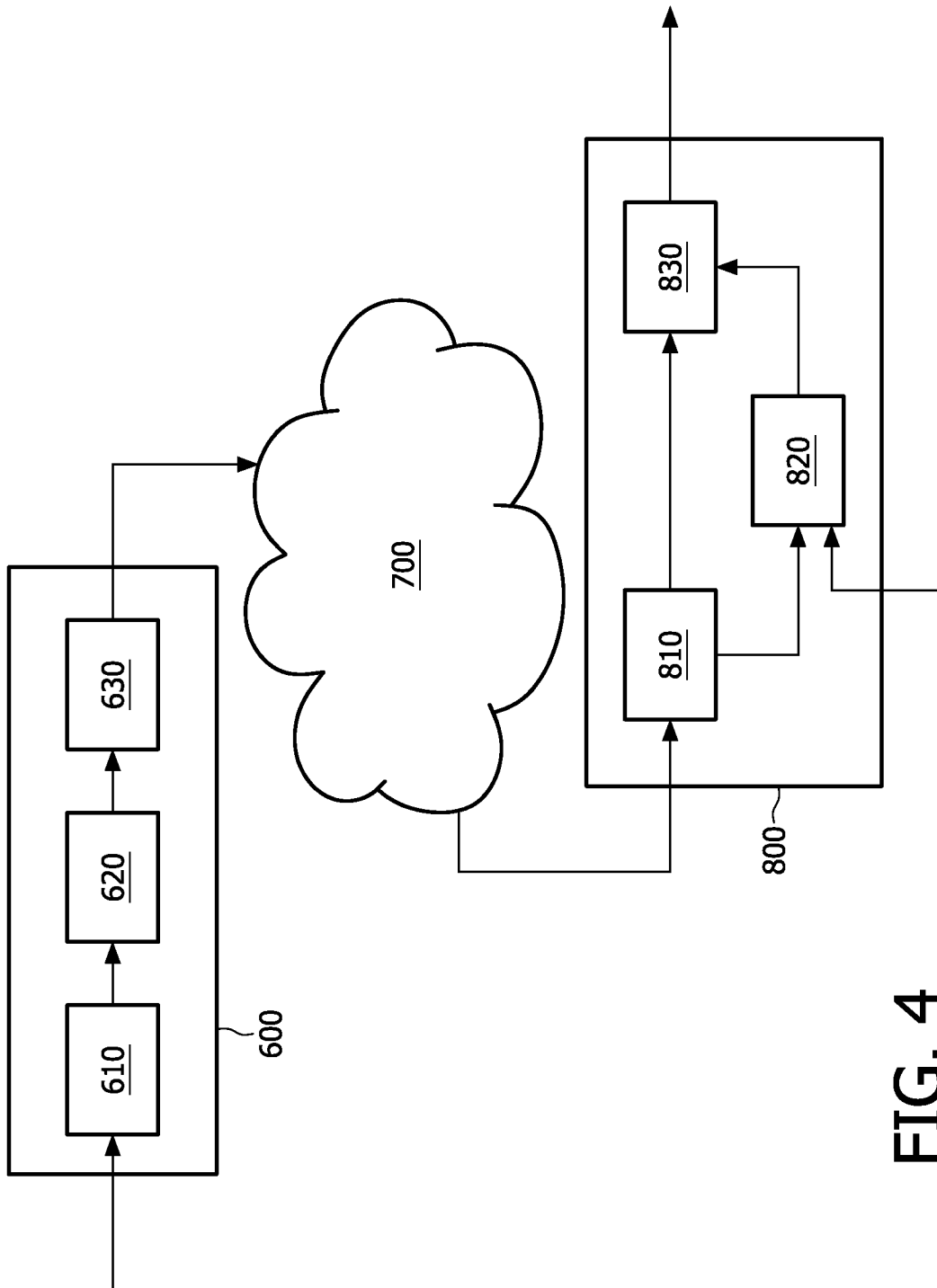


FIG. 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2008/050041

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L19/00 H04S7/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L H04S

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
X	BREEBAART J ET AL: "Multi-channel goes mobile: MPEG surround binaural rendering" AES INTERNATIONAL CONFERENCE. AUDIO FOR MOBILE AND HANDHELD DEVICES, XX, XX, 2 September 2006 (2006-09-02), pages 1-13, XP007902577 Section 2.1 figure 2	10
Y	-----	6-8, 12, 14, 15
Y	PASI OJALA ET AL: "Further information on binaural decoder functionality" VIDEO STANDARDS AND DRAFTS, XX, XX, no. M13233, 29 March 2006 (2006-03-29), XP030041902 Section 3 figure 2 -----	1-9, 11-15
	-/--	

Further documents are listed in the continuation of Box C See patent family annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed
- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search 17 April 2008	Date of mailing of the international search report 28/04/2008
---	---

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Fruhmann, Markus
---	---

INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2008/050041

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	"SAOC use cases, draft requirements and architecture" VIDEO STANDARDS AND DRAFTS, XX, XX, no. W8638, 27 October 2006 (2006-10-27), XP030015132 Section 3.2 figure 5	1-9, 11-15
A	----- PASI OJALA ET AL: "Description of Binaural Audio Image Control to FCD" VIDEO STANDARDS AND DRAFTS, XX, XX, no. M13545, 12 July 2006 (2006-07-12), XP030042214 the whole document -----	1-15