



(51) International Patent Classification:
H04S 7/00 (2006.01)

(21) International Application Number:
PCT/EP2019/060996

(22) International Filing Date:
30 April 2019 (30.04.2019)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: HUAWEI TECHNOLOGIES CO., LTD.
[CN/CN]; Huawei Administration Building Bantian Long-gang District, Shenzhen, Guangdong 518129 (CN).

(72) Inventor; and

(71) Applicant (for US only): POLLOW, Martin [DE/DE];
c/o Huawei Technologies Duesseldorf GmbH Riesstr. 25,
80992 Munich (DE).

(72) Inventors: FALLER, Christof; Illusonic GmbH, Bahnstrasse 23, 8610 Uster (CH). FAVROT, Alexis; Illusonic GmbH, Bahnstrasse 23, 8610 Uster (CH). TAGHIZADEH, Mohammad; c/o Huawei Technologies Duesseldorf GmbH, Riesstr. 25, 80992 Munich (DE).

(74) Agent: KREUZ, Georg; Huawei Technologies Duesseldorf GmbH, Riesstr. 25, 80992 Munich (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(54) Title: DEVICE AND METHOD FOR RENDERING A BINAURAL AUDIO SIGNAL

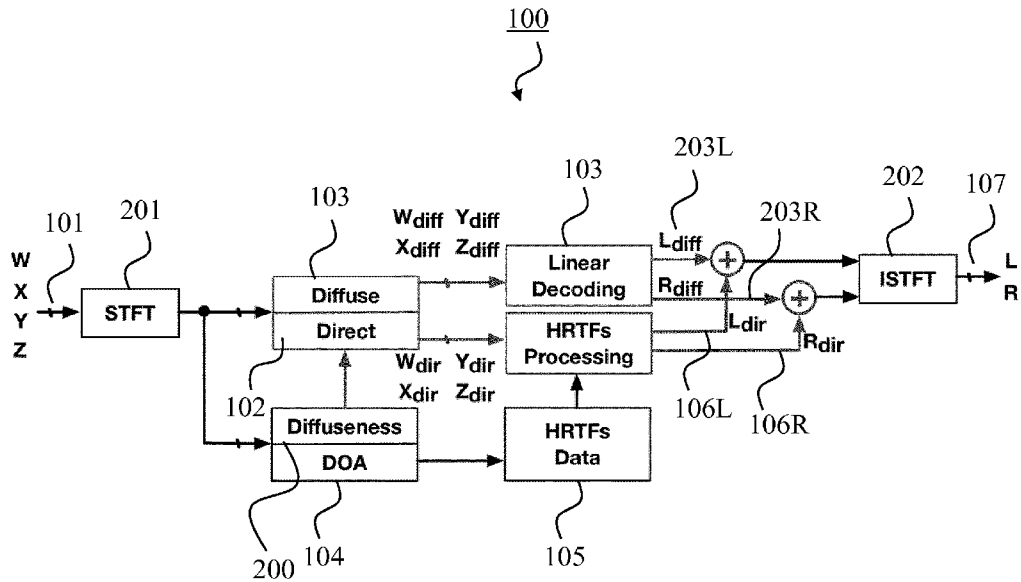


FIG. 2

(57) Abstract: The invention relates to the technical field of 3D sound, for instance, for virtual reality (VR) applications or surround sound. The invention proposes in particular a device and a method for rendering a binaural audio signal. The device is configured to obtain a direct component of the audio signal and a diffuse component and a Direction of Arrival (DoA) from a plurality of audio channels of the audio signal. Further, the device is configured to determine a head-related transfer function (HRTF) according to the DoA. Finally, the device is configured to filter the direct component based on the HRTF to obtain a modified direct component, and to generate the binaural audio signal based on the diffuse component and the modified direct component.



(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

DEVICE AND METHOD FOR RENDERING A BINAURAL AUDIO SIGNAL

TECHNICAL FIELD

5 The present invention relates to the technical field of three-dimensionality (3D) sound, for instance, for virtual reality (VR) applications or surround sound. The invention also relates to VR compatible audio formats, e.g. First Order Ambisonic (FOA) signals (also referred to as B-format). The invention relates specifically to generating binaural sounds/signals from such audio formats. The invention proposes to this end a device and a method for rendering a binaural
10 signal.

BACKGROUND

3D VR sounds are typically recorded and stored as FOA signals. The rendering of these FOA signals over headphones is then done by converting them into binaural sounds. The binaural
15 sounds are obtained based on Head Related Transfer Functions (HRTFs), which model the filters from a point source emitting to the ears, such that the impressions of immersion and externalisation are improved.

In particular, the binaural sounds are usually rendered by applying the HRTFs to decoded
20 virtual loudspeaker signals. For instance, a straightforward approach for obtaining the binaural sounds is, to decode the FOA signals into specific loudspeaker setups with pre-defined positions, and then applying the HRTFs relatively to these positions. However, direct and linear decoding of the FOA signals does not provide enough spatial resolution to cover the entire 3D space. Moreover, the performance is often restricted by a trade-off between the computational
25 complexity of the system and the precision (length) of the HRTFs models or measurements.

Non-linear decoding provides better localisation and spatialization, or better discrimination between direct and diffuse sounds for improved spaciousness. In particular, parametric approaches in the time-frequency domain, based on non-linear decoding into N virtual
30 loudspeaker signals, provide such better results. However, it is still necessary to apply the HRTFs to virtual loudspeaker signals by convolutions. In particular, they are applied to the N parametrically generated virtual loudspeaker signals.

Other approaches directly synthesise a two channel binaural signal using target binaural cues, which are computed either based on an analysis or based on MPEG surround spatial cues.

SUMMARY

5 In view of the above, embodiments of the invention aim to improve the current approaches for rendering binaural sounds. An objective is to obtain a binaural audio signal with improved spatial resolution and by less computational complexity. In particular, the use of virtual loudspeaker signals may be avoided. Further, binaural cues should not be necessary for the HRTFs. In addition, full compatibility with existing FOA signals is desired.

10

The objective is achieved by the embodiments of the invention as described in the enclosed independent claims. Advantageous implementations of the embodiments of the invention are further defined in the dependent claims.

15 In particular, embodiments of the invention propose separating a direct component from a diffuse component of an audio signal, then modifying the direct component based on a HRTF, which is determined based on a Direction of Arrival (DoA) related to the audio signal, and then rendering the binaural audio signal from the diffuse component and the modified direct component.

20

A first aspect of the invention provides a device for rendering a binaural audio signal, wherein the device is configured to: obtain a direct component of the audio signal and a diffuse component and a DoA, from a plurality of audio channels of the audio signal, determine a HRTF according to the DoA, filter the direct component based on the HRTF to obtain a modified
25 direct component, and generate the binaural audio signal based on the diffuse component and the modified direct component.

The device of the first aspect is able to render the binaural audio signal with high spatial resolution and low computational complexity. Thereby, the use of virtual loudspeaker signals
30 is not necessary, and also not binaural cues need to be used. The HRTF can be applied directly to the audio signal (direct component). Moreover, the device of the first aspect is full compatibility with existing audio signals, particularly FOA signals.

In an implementation form of the first aspect, the plurality of audio channels are B-format channels.

5 That is, the audio channels may be the channels of a FOA signal (i.e. W , X , Y and Z), and the audio signal may be the FOA signal.

In an implementation form of the first aspect, the device is configured to: generate the binaural audio signal by combining the diffuse component with the modified component.

10 In an implementation form of the first aspect, the device is configured to: generate a left diffuse component and a right diffuse component based on the diffuse component, generate a left modified direct component and a right modified direct component based on the modified direct component, and combine the left diffuse component with the left modified direct component, and combine the right diffuse component with the right modified direct component, wherein
15 the binaural audio signal is generated based on the result of the combining.

In this way the binaural audio signal can be rendered optimally, for instance, for a headphone device or VR device.

20 In an implementation form of the first aspect, the device is configured to: generate the left modified direct component and the right modified direct component by applying the HRTF to the direct component.

25 Applying the HRTF only to the direct component, in order to modify the direct component, ensures low computational complexity.

In an implementation form of the first aspect, the device is configured to: generate the left diffuse component and the right diffuse component by linear decoding of the diffuse component.

30 This provides a simple way to obtain the left/right diffuse components, with low computational complexity.

In an implementation form of the first aspect, the device is configured to: estimate a diffuseness of the audio signal, and obtain the direct component and optionally the diffuse component based on the estimated diffuseness.

- 5 In an implementation form of the first aspect, the device is configured to: estimate the diffuseness based on a short-time Fourier transform of the audio signal.

In an implementation form of the first aspect, the device is configured to: estimate the DoA based on a short-time Fourier transform of the audio channels.

10

In an implementation form of the first aspect, the HRTF includes a gain and a phase to be applied to each time-frequency tile of the direct component.

- 15 In an implementation form of the first aspect, the device is configured to: smooth the HRTF over time to obtain a smoothed HRTF, and modify the direct component based on the smoothed HRTF.

In this way, audible audio artefacts can be reduced.

- 20 In an implementation form of the first aspect, the audio signal is an Ambisonic signal.

In particular, the audio signal may be a FOA signal.

- 25 A second aspect of the invention provides a headphone device, wherein the headphone device comprises a device for rendering a binaural audio signal according to the first aspect or any of its implementation forms.

- The headphone device of the second aspect may particularly be for a 3D VR system. For instance, it may be included in such a system. The headphone device of the second aspect enjoys
30 all advantages of the device of the first aspect.

A third aspect of the invention provides a method for rendering a binaural audio signal, wherein the method comprises: obtaining a direct component of the audio signal and a diffuse component and a DoA from a plurality of audio channels of the audio signal, determining a

HRTF according to the DoA, filtering the direct component based on the HRTF to obtain a modified direct component, and generating the binaural audio signal based on the diffuse component and the modified direct component.

- 5 In an implementation form of the third aspect, the plurality of audio channels are B-format channels.

In an implementation form of the third aspect, the method comprises: generating the binaural audio signal by combining the diffuse component with the modified component.

10

In an implementation form of the third aspect, the method comprises: generating a left diffuse component and a right diffuse component based on the diffuse component, generating a left modified direct component and a right modified direct component based on the modified direct component, and combining the left diffuse component with the left modified direct component, and combining the right diffuse component with the right modified direct component, wherein the binaural audio signal is generated based on the result of the combining.

15

In an implementation form of the third aspect, the method comprises: generating the left modified direct component and the right modified direct component by applying the HRTF to the direct component.

20

In an implementation form of the third aspect, the method comprises: generating the left diffuse component and the right diffuse component by linear decoding of the diffuse component.

- 25 In an implementation form of the third aspect, the method comprises: estimating a diffuseness of the audio signal, and obtaining the direct component and optionally the diffuse component based on the estimated diffuseness.

In an implementation form of the third aspect, the method comprises: estimating the diffuseness based on a short-time Fourier transform of the audio signal.

30

In an implementation form of the third aspect, the method comprises: estimating the DoA based on a short-time Fourier transform of the audio channels.

In an implementation form of the third aspect, the HRTF includes a gain and a phase to be applied to each time-frequency tile of the direct component.

5 In an implementation form of the third aspect, the method comprises: smoothing the HRTF over time to obtain a smoothed HRTF, and modifying the direct component based on the smoothed HRTF.

In an implementation form of the third aspect, the audio signal is an Ambisonic signal.

10 The method of the third aspect and its implementation forms achieve all advantages of the device of the first aspect and its respective implementation forms.

A fourth aspect of the invention provides a computer program product comprising a program code for controlling a device according to the first aspect or any of its implementation forms, 15 or for controlling a headphone device according to the second aspect, or for carrying out, when executed by a processor, the method according to the third aspect.

It has to be noted that all devices, elements, units and means described in the present application could be implemented in the software or hardware elements or any kind of combination thereof. 20 All steps which are performed by the various entities described in the present application as well as the functionalities described to be performed by the various entities are intended to mean that the respective entity is adapted to or configured to perform the respective steps and functionalities. Even if, in the following description of specific embodiments, a specific functionality or step to be performed by external entities is not reflected in the description of a 25 specific detailed element of that entity which performs that specific step or functionality, it should be clear for a skilled person that these methods and functionalities can be implemented in respective software or hardware elements, or any kind of combination thereof.

BRIEF DESCRIPTION OF DRAWINGS

30 The above described aspects and implementation forms of the invention will be explained in the following description of specific embodiments in relation to the enclosed drawings, in which

FIG. 1 shows a device according to an embodiment of the invention.

FIG. 2 shows a device according to an embodiment of the invention.

FIG. 3 shows a method according to an embodiment of the invention.

5 DETAILED DESCRIPTION OF EMBODIMENTS

FIG. 1 shows a device 100 according to an embodiment of the invention. The device 100 is configured to render a binaural audio signal 107, particularly from an audio signal 101 comprising multiple audio channels, e.g. an FOA signal. The device 100 may be used in a headphone or VR device.

10

The device 100 may comprise processing circuitry (not shown) configured to perform, conduct or initiate the various operations of the device 100 described herein. The processing circuitry may comprise hardware and software. The hardware may comprise analog circuitry or digital circuitry, or both analog and digital circuitry. The digital circuitry may comprise components
15 such as application-specific integrated circuits (ASICs), field-programmable arrays (FPGAs), digital signal processors (DSPs), or multi-purpose processors. In one embodiment, the processing circuitry comprises one or more processors and a non-transitory memory connected to the one or more processors. The non-transitory memory may carry executable program code which, when executed by the one or more processors, causes the device 100 to perform, conduct
20 or initiate the operations or methods described herein.

20

The device 100 is specifically configured to obtain a direct component 102 of the audio signal 101 and a diffuse component 103 and a DoA 104 from a plurality of audio channels of the audio signal 101. That is, the device 100 may be configured to separate the audio signal 101 into the
25 direct component 102 and the diffuse component 103, and to derive the DoA 104 related to the audio signal 101. The audio signal 101 be a FOA signal or B-format signal. The plurality of audio channels may be B-format channels.

25

The device 100 is further configured to determine a HRTF 105 according to the DoA 104. To
30 this end, it may be configured to first estimate the DoA 104, for instance, based on a short-time Fourier transform of the audio channels of the audio signal 101. Then, it may use the DoA 104 to generate the HRTF 105.

30

The device 100 is further configured to filter the direct component 102 based on the HRTF 105 (e.g. applying the HRTF 105 to the direct component 102), in order to obtain a modified direct component 106, and then to generate the binaural audio signal 107 based on the diffuse component 103 and the modified direct component 106. In particular, the device 100 may
5 generate the binaural audio signal 107 by combining the diffuse component 103 with the modified direct component 106.

FIG. 2 shows the device 100 of FIG. 1 according to an embodiment of the invention in more detail, particularly with further optional features. The device 100 shown in FIG. 2 can generate
10 a left diffuse component 203L and a right diffuse component 203R based on the diffuse component 103, and can generate a left modified direct component 106L and a right modified direct component 106R based on the modified direct component 106. Then, in order to generate the binaural audio signal 107, the device 100 can combine in particular the left diffuse component 203L with the left modified direct component 106L, and also combine the right
15 diffuse component 203R with the right modified direct component 106R. The binaural audio signal 107 is thus generated based on the result of the combining.

Considering specifically, but as an example only, a FOA signal as the audio signal 101, which is to be parametrically decoded into binaural sound (i.e. into the binaural audio signal 107), the
20 device 100 may first estimate the DoA 104 of the FOA signal 101 (e.g. azimuth and elevation angles), as well as the directness, respectively, the diffuseness 200 of the FOA signal 101.

Then, the direct component 102 and the diffuse component 103 of the FOA signal 101 may be separated. The diffuse component 103 may be decoded linearly into the left and right diffuse
25 components 203L and 203R. Meanwhile, HRTF processing may be applied to the direct component 102, in order to obtain the left and right direct components 106L and 106R, respectively. Eventually both the direct and diffuse components may be combined to obtain the binaural audio signal 107 (including binaural audio channel, e.g. left and right).

30 The device 100 does not rely on numerous decoded virtual signals, but the HRTF 105 can be applied directly to the FOA signal 101. Moreover, the HRTF 105 can be integrated in the parametrical model and can take advantage of FOA analysis, i.e. can be fully adaptively computed from the DOA estimate. Thus, the device 100 and the algorithm it performs are compact and computational efficient, since all operations may be done independently for each

time-frequency tile. An exemplary specific algorithm, which the device 100 may perform, is described in the following.

5 Considering the complex spectra W , X , Y and Z of the FOA signal 101, which may be obtained running a N_{STFT} points short-time Fourier transform (STFT 201 in FIG. 2), a DOA analysis can be performed according to:

$$\theta_{FOA}[k,i] = \arctan \frac{\text{Re}\{W^* Y\}}{\text{Re}\{W^* X\}}, \quad (1)$$

10 and

$$\varphi_{FOA}[k,i] = \arctan \frac{\text{Re}\{W^* Z\}}{\sqrt{|\text{Re}\{W^* X\}|^2 + |\text{Re}\{W^* Y\}|^2}}. \quad (2)$$

15 $\theta_{FOA}[k,i]$ and $\varphi_{FOA}[k,i]$ are, respectively, the corresponding azimuth and elevation angles estimated at frequency bin i and time frame k , “Re{...}” means “The real part of ...”. Simultaneously, a directness estimation can be performed based on the same spectra of the FOA signal 101, according to:

$$\delta_{FOA}[k,i] = \sqrt{2} \frac{\sqrt{|\text{Re}\{W^* X\}|^2 + |\text{Re}\{W^* Y\}|^2 + |\text{Re}\{W^* Z\}|^2}}{|W|^2 + 1/2(|X|^2 + |Y|^2 + |Z|^2)}. \quad (3)$$

20

The directness estimate (3) can then be used to separate the direct component 102 from the diffuse component of the FOA signal 101, according to:

$$\begin{aligned} W_{dir} &= \delta_{FOA}^y W, & X_{dir} &= \delta_{FOA}^y X, \\ Y_{dir} &= \delta_{FOA}^y Y, & Z_{dir} &= \delta_{FOA}^y Z, \end{aligned} \quad (4)$$

25

and

$$\begin{aligned} W_{diff} &= (1 - \delta_{FOA})^y W, & X_{diff} &= (1 - \delta_{FOA})^y X, \\ Y_{diff} &= (1 - \delta_{FOA})^y Y, & Z_{diff} &= (1 - \delta_{FOA})^y Z. \end{aligned} \quad (5)$$

Thereby, γ stands for an exponent design parameter, which is typically chosen such that $\gamma=0.5$.

5 The left and right diffuse components 203L and 203R (L_{diff} and R_{diff} in FIG. 2) may then be obtained by linear decoding of the FOA diffuse component 102 obtained in (5). For instance, left and right channels can be rendered through two decoded cardioid signals with maximum angle separation, i.e. according to:

$$10 \quad L_{\text{diff}} = W_{\text{diff}} + Y_{\text{diff}} \quad \text{and} \quad R_{\text{diff}} = W_{\text{diff}} - Y_{\text{diff}}. \quad (6)$$

In this way, the left and right rendered diffuse components 203L and 203R will benefit from the best possible de-correlation. Moreover, the decoding can possibly be made frequency dependent, in order to follow the physical shape of the cross-correlation coefficient between
15 both ear signals.

The left and right modified direct components 106L and 106R (L_{dir} and R_{dir} in FIG. 2) may be obtained by applying the HRTF processing based on the HRTF 105 to the FOA direct components 102 obtained in (4). A simple HRTF 105 model can be directly derived from the
20 DOA estimation given in (1) and (2), wherein the inter-aural level differences (ILDs) between both ears may be derived given the first order filter according to:

$$H(s) = \frac{a_+(\theta, \phi)s + b}{s + b}, \quad (7)$$

25 Thereby, the coefficient a_+ and b may be derived from:

$$a_+(\theta, \phi) = 1 + \sin \theta \cos \phi, \quad (8)$$

and

$$30 \quad b = \frac{2c}{K}, \quad (9)$$

K is here the diameter of the head of the HRTF model and c being the speed of sound. Considering the magnitude of the filter H at a given frequency f the left and right HRTF gains are of the HRTF model can be obtained as:

$$5 \quad gL = \frac{\sqrt{(2a_+f)^2 + b^2}}{\sqrt{(2f)^2 + b^2}}, \quad (10)$$

and

$$gR = \frac{\sqrt{(2a_-f)^2 + b^2}}{\sqrt{(2f)^2 + b^2}}, \quad (11)$$

10 Thereby, $a_-(\theta, \phi) = 1 - \sin\theta \cos\phi$. The inter-aural time differences (ITDs) between both ears can be derived assuming plane waves propagation:

$$dL = \frac{Ka_+}{2c}, \quad (12)$$

and

$$dR = \frac{Ka_-}{2c}, \quad (13)$$

15

Given the ILDs and ITDs derived in (10)-(13), the HRTF model may simply comprise or consists of a gain and phase to be apply on each time-frequency tiles of the direct sound contributions of the FOA signal 101:

$$20 \quad hL[k, i] = \hat{g}L[k, i] \exp\{j2\pi f \hat{d}L[k, i]\} \quad (14)$$

and

$$hR[k, i] = \hat{g}R[k, i] \exp\{j2\pi f \hat{d}R[k, i]\} \quad (15)$$

25 with, by construction,

$$f = \frac{i\pi}{\frac{N_{SFFT}}{2} + 1}. \quad (16)$$

The derived HRTF 105 parameters may be smoothed over time to reduce audible audio artefacts, e.g. according to:

$$\hat{g}L[k, i] = \alpha_{HRTF} gL[k, i] + (1 - \alpha_{HRTF}) \hat{g}L[k, i - 1], \quad (17)$$

5

Thereby, α_{HRTF} may be determined by:

$$\alpha_{HRTF} = \frac{N_{SIFT}}{T_{HRTF} f_s}, \quad (18)$$

10 T_{HRTF} is the averaging time-constant in seconds and f_s is the spectrum sampling frequency. \hat{g}_R , \hat{d}_L and \hat{d}_R may be obtained analogously.

The left and right direct components 106L and 106R, respectively, i.e. the direct sound contributions to the binaural audio signal 107, may be obtained by applying the previously

15 derived filters to the omnidirectional direct signal W_{dir} according to:

$$\begin{aligned} L_{dir} &= \hat{g}_L[k, i] W_{dir}[k, i] \\ R_{dir} &= \hat{g}_R[k, i] W_{dir}[k, i]. \end{aligned} \quad (19)$$

20 The direct components 106L and 106R may also be obtained by applying the previously derived filters to combinations of B-Format channels, e.g. according to:

$$\begin{aligned} L_{dir} &= \hat{g}_L[k, i] \left(\frac{2}{3} W_{dir}[k, i] + \frac{1}{3} Y_{dir}[k, i] \right) \\ R_{dir} &= \hat{g}_R[k, i] \left(\frac{2}{3} W_{dir}[k, i] - \frac{1}{3} Y_{dir}[k, i] \right), \end{aligned} \quad (20)$$

In this, case left and right oriented sub-cardioid B-Format linear combinations may be used.

25 \hat{g}_L and \hat{g}_R may be adjusted to compensate for a gain difference between W_{dir} and the left and right sub-cardioid, respectively.

Finally, the final binaural signal 107 may be reconstructed by adding the direct and diffuses

sound contributions.

$$L = L_{dir} + L_{diff}, \quad (21)$$

and

$$R = R_{dir} + R_{diff}. \quad (22)$$

FIG. 3 shows a method 300 according to an embodiment of the invention. The method 300 is for rendering a binaural signal 107, and can be performed by the device 100 as shown in FIG. 1 or FIG. 2.

10

The method 300 comprises: a step 301 of obtaining a direct component 102 of the audio signal 101 and a diffuse component 103 and a DoA 104 from a plurality of audio channels of the audio signal 101; a step 302 of determining a HRTF 104 according to the DoA 105; a step 303 of filtering the direct component 102 based on the HRTF 104 to obtain a modified direct component 106; and a step 304 of generating the binaural audio signal 107 based on the diffuse component 103 and the modified direct component 106.

15

The device 100, method 300, and the detailed algorithm can have the following advantages:

20

- There may be no need for virtual loudspeaker signals. Instead, the direct component 102 and the diffuse component 103 (particularly the left/right diffuse components 203L/203R) may be separately obtained from the single FOA channels (i.e. the FOA signal 101).
- The HRTF 105 may not require target binaural cues. Instead, the two binaural channels can be synthesized based on:
 - The DOA for the direct sound.
 - Purely (frequency dependent) linear rendering of the diffuse sound.
- The device 100, method 300, and algorithm may be compatible with existing FOA signals.

25

The invention has been described in conjunction with various embodiments as examples as well as implementations. However, other variations can be understood and effected by those persons skilled in the art and practicing the claimed invention, from the studies of the drawings, this disclosure and the independent claims. In the claims as well as in the description the word “comprising” does not exclude other elements or steps and the indefinite article “a” or “an” does not exclude a plurality. A single element or other unit may fulfill the functions of several

30

entities or items recited in the claims. The mere fact that certain measures are recited in the mutual different dependent claims does not indicate that a combination of these measures cannot be used in an advantageous implementation.

CLAIMS

1. A device (100) for rendering a binaural audio signal (107), wherein the device (100) is configured to:
 - 5 obtain a direct component (102) of the audio signal (101) and a diffuse component (103) and a Direction of Arrival, DoA, (104) from a plurality of audio channels of the audio signal (101),
determine a head-related transfer function, HRTF, (105) according to the DoA (104),
filter the direct component (102) based on the HRTF (105) to obtain a modified direct
10 component (106), and
generate the binaural audio signal (107) based on the diffuse component (103) and the modified direct component (106).
2. The device (100) according to claim 1, wherein:
15 the plurality of audio channels are B-format channels.
3. The device (100) according to claim 1 or 2, configured to:
generate the binaural audio signal (107) by combining the diffuse component (103) with
the modified direct component (106).
20
4. The device (100) according to one of the claims 1 to 3 configured to:
generate a left diffuse component (203L) and a right diffuse component (203R) based
on the diffuse component (103),
generate a left modified direct component (106L) and a right modified direct component
25 (106R) based on the modified direct component (106), and
combine the left diffuse component (203L) with the left modified direct component
(106L), and combine the right diffuse component (203R) with the right modified direct
component (106R), wherein the binaural audio signal (107) is generated based on the result of
the combining.
30
5. The device (100) according to claim 4, configured to:
generate the left modified direct component (106L) and the right modified direct
component (106R) by applying the HRTF (105) to the direct component (102).

6. The device (100) according to claim 4 or 5, configured to:
generate the left diffuse component (203L) and the right diffuse component (203R) by linear decoding of the diffuse component (103).
- 5 7. The device (100) according to one of the claims 1 to 6, configured to:
estimate a diffuseness (200) of the audio signal (101), and
obtain the direct component (102) and optionally the diffuse component (103) based on the estimated diffuseness (200).
- 10 8. The device (100) according to claim 7, configured to:
estimate the diffuseness based on a short-time Fourier transform (201) of the audio signal (101).
9. The device (100) according to claim 8, configured to:
15 estimate the DoA (104) based on a short-time Fourier transform of the audio channels.
10. The device (100) according to one of the claims 1 to 9, wherein:
the HRTF (105) includes a gain and a phase to be applied to each time-frequency tile of the direct component (102).
- 20 11. The device (100) according to one of the claims 1 to 10, configured to:
smooth the HRTF (104) over time to obtain a smoothed HRTF, and
modify the direct component (102) based on the smoothed HRTF.
- 25 12. The device (100) according to one of the claims 1 to 11, wherein:
the audio signal (101) is an Ambisonic signal.
13. A headphone device, wherein the headphone device comprises a device (100) for rendering a binaural audio signal (107) according to one of the claims 1 to 12.
- 30 14. A method (300) for rendering a binaural audio signal (107), wherein the method (300) comprises:

obtaining (301) a direct component (102) of the audio signal (101) and a diffuse component (103) and a Direction of Arrival, DoA, (104) from a plurality of audio channels of the audio signal (101),

5 determining (302) a head-related transfer function, HRTF, (104) according to the DoA (105),

filtering (303) the direct component (102) based on the HRTF (104) to obtain a modified direct component (106), and

generating (304) the binaural audio signal (107) based on the diffuse component (103) and the modified direct component (106).

10

15. A computer program product comprising a program code for controlling a device (100) according to any one of the claims 1 to 12, or for controlling a headphone device according to claim 13, or for carrying out, when executed by a processor, the method (300) according to claim 14.

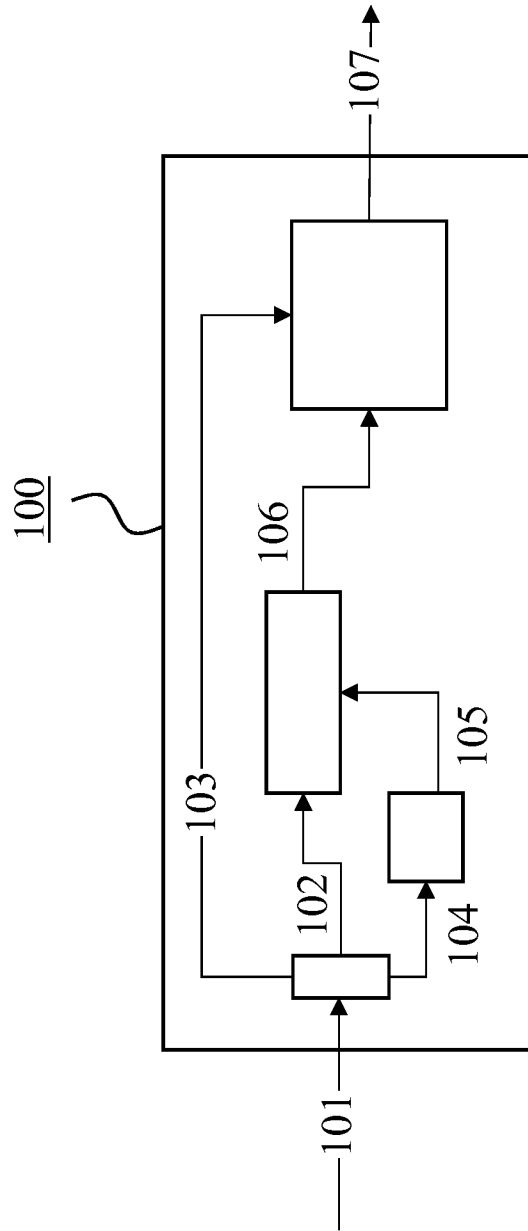


FIG. 1

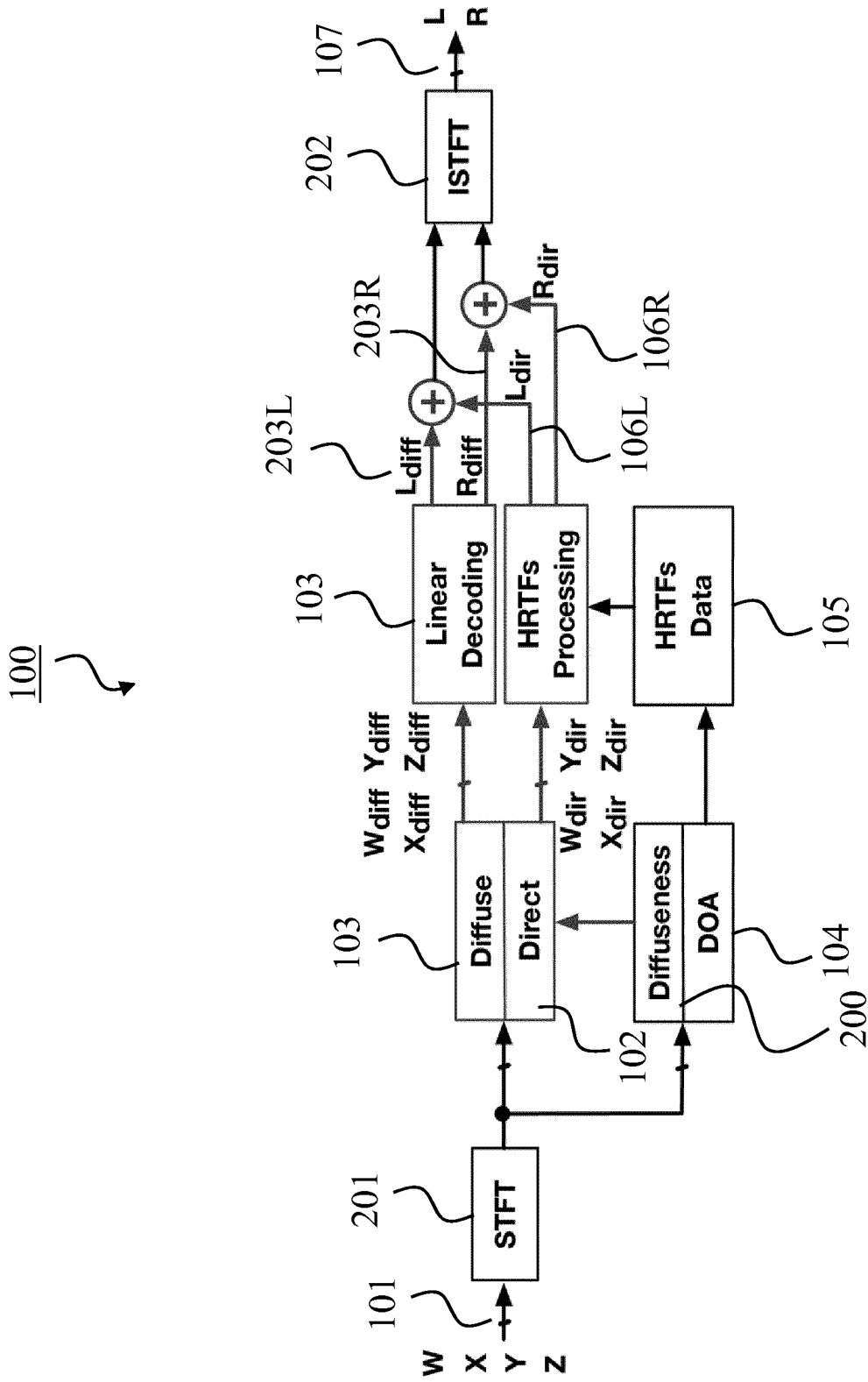


FIG. 2

300

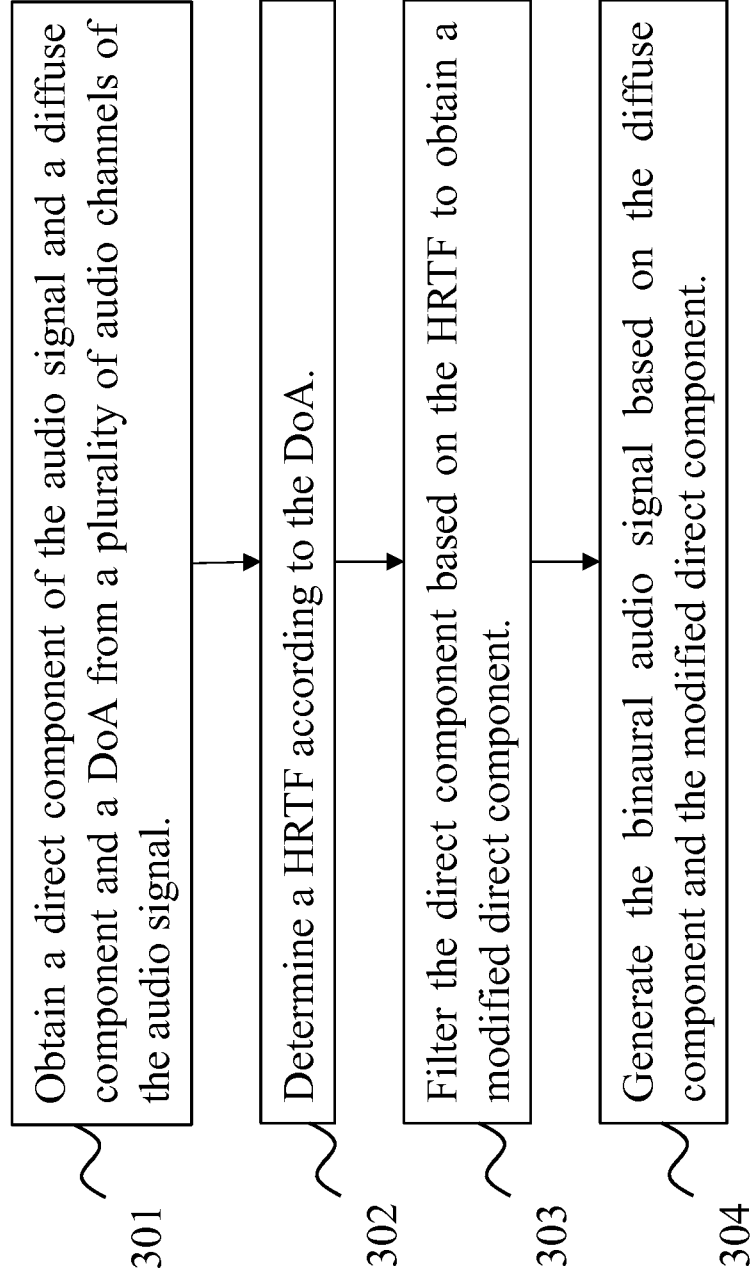


FIG. 3

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/060996

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04S7/00
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

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 2 942 981 A1 (FRAUNHOFER GES ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E V [DE] ET AL.) 11 November 2015 (2015-11-11)	1,3-10, 13-15
Y	paragraphs [0084] - [0152], [0252]; figures 2,3	2,11,12
Y	----- EP 2 249 334 A1 (FRAUNHOFER GES FORSCHUNG [DE]) 10 November 2010 (2010-11-10)	2,12
A	paragraphs [0024] - [0038]; figures 5-7	1,3-11, 13-15
Y	----- US 2003/035553 A1 (BAUMGARTE FRANK [US] ET AL) 20 February 2003 (2003-02-20)	11
A	paragraphs [0069] - [0070] -----	1-10, 12-15

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 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

9 January 2020

Date of mailing of the international search report

21/01/2020

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

Fruhmann, Markus

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No PCT/EP2019/060996

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
EP 2942981	A1	11-11-2015	CN 106664485 A	10-05-2017
			CN 106664501 A	10-05-2017
			EP 2942981 A1	11-11-2015
			EP 2942982 A1	11-11-2015
			EP 3141000 A1	15-03-2017
			EP 3141001 A1	15-03-2017
			JP 6466968 B2	06-02-2019
			JP 6466969 B2	06-02-2019
			JP 2017517947 A	29-06-2017
			JP 2017517948 A	29-06-2017
			US 2017078818 A1	16-03-2017
			US 2017078819 A1	16-03-2017
			WO 2015169617 A1	12-11-2015
			WO 2015169618 A1	12-11-2015
EP 2249334	A1	10-11-2010	AU 2010244393 A1	24-11-2011
			BR PI1007730 A2	06-03-2018
			CA 2761439 A1	11-11-2010
			CN 102422348 A	18-04-2012
			EP 2249334 A1	10-11-2010
			EP 2427880 A1	14-03-2012
			ES 2426136 T3	21-10-2013
			JP 5400954 B2	29-01-2014
			JP 2012526296 A	25-10-2012
			KR 20120013986 A	15-02-2012
			PL 2427880 T3	31-01-2014
			RU 2011145865 A	27-05-2013
			US 2012114126 A1	10-05-2012
			WO 2010128136 A1	11-11-2010
US 2003035553	A1	20-02-2003	NONE	