(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 December 2024 (12.12.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/253691 A1

(51) International Patent Classification:
G10K 15/08 (2006.01) *H04S 7/00* (2006.01)

(21) International Application Number:
PCT/US2023/068146

(22) International Filing Date:
08 June 2023 (08.06.2023)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **GOOGLE LLC** [US/US]; 1600 Amphitheatre Parkway, Mountain View, California 94043 (US).

(72) Inventors: **SHIN, Dongeek**; GOOGLE LLC, 1600 Amphitheatre Parkway, Mountain View, California 94043 (US). **HERSEK, Sinan**; GOOGLE LLC, 1600 Amphitheatre Parkway, Mountain View, California 94043 (US).

(74) Agent: **GORDON, Ronald L.** et al.; BRAKE HUGHES BELLERMANN LLP, P.O. Box 1077, Middletown, Maryland 21769 (US).

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ,

(54) **Title:** PERFORMING COMMON AUDIO CONVOLUTIONS USING ISO-TRAJECTORIES IN THREE DIMENSIONAL (3D) SPACE

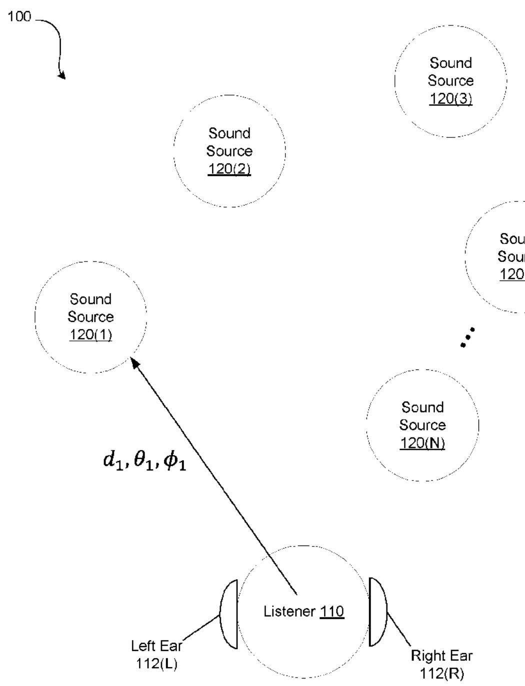

FIG. 1A

(57) **Abstract:** Techniques include using the fact that the HRTF depends on a sound source position on a unit sphere relative to a listener while the reverb transfer function depends on a range relative to the listener. In this way, sound source positions along a line in three-dimensional space have an identical HRTF while speaker positions along a circular arc in three-dimensional space have an identical reverb transfer function. Accordingly, identify those sound source positions along a line and use a common HRTF in a single convolution for each ear, and those speaker positions along an arc and use a common reverb transfer function in a single convolution for each ear.

WO 2024/253691 A1

[Continued on next page]

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, ME, 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).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

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

PERFORMING COMMON AUDIO CONVOLUTIONS USING ISO-TRAJECTORIES IN THREE DIMENSIONAL (3D) SPACE

TECHNICAL FIELD

[0001] This description relates in general to audio devices used in virtual reality and augmented reality (VR/AR) devices.

BACKGROUND

[0039] A sound field generated at an array of virtual sound sources can reproduce the effect of a sound source from any vantage point relative to a listener. Such a sound field may be decoded and used in the delivery of audio through headphone speakers in, e.g., Virtual Reality (VR) and/or augmented reality (AR) systems.

SUMMARY

[0003] This application is directed to binaural rendering of audio signals in a three-dimensional space for a listener. Binaural rendering involves computing convolutions of audio signals with head-related transfer functions (HRTFs) for each ear of the listener. In addition, the room in which the listener hears the audio signals may have its own reverb transfer function that represents the effects of reflections off the walls, ceilings, floors, and other obstacles in the room on the audio signals. For a binaural rendering, there are four convolutions (two HRTFs and two reverbs, one for each ear) for each audio source. Accordingly, in a situation in which there are many sound sources (e.g., audio sources), the convolution computation may become a bottleneck in processing the binaural rendering. It turns out, however, that the HRTFs depend only on the angle of a sound source with respect to the listener (e.g., point on a unit sphere with the listener at center) and the reverb TFs depend only on the distances of the sound sources from the listener. Accordingly, when the sound sources are arranged along a line at an angle with respect to the listener, the HRTFs are the same for each of those sound sources and the respective convolutions can be reduced to a single convolution with the sum of the sound fields from the sound sources along the line. Moreover, when the sound sources are arranged along a circular arc with he listener at center, the reverb TFs are the same for each of those sound sources and the respective

convolutions can be reduced to a single convolution with the sum of the sound fields from the sound sources along the arc. Thus, by finding a set of sound sources that lie along a curve (e.g., line or arc), the number of convolutions to compute in the binaural rendering of the audio field may be reduced, thus significantly reducing the computational burden of binaural rendering.

[0004] In one general aspect, a method includes identifying a set of sound sources in a three-dimensional space having positions that, within a threshold, intersect a curve, the curve being defined by a parameter value that is constant along the curve. The method also includes summing sound fields from the set of sound sources to produce a net sound field. The method further includes performing a convolution operation on the net sound field and a transfer function to produce a binaurally rendered sound field, the transfer function depending on the parameter value.

[0005] In another general aspect, computer program product comprising a nontransitory storage medium, the computer program product including code that, when executed by at least one processor, causes the at least one processor to perform a method. The method includes identifying a set of sound sources in a three-dimensional space that, within a threshold, intersect a curve, the curve being defined by a parameter value that is constant along the curve. The method also includes summing sound fields from the set of sound sources to produce a net sound field. The method further includes performing a convolution operation on the net sound field and a transfer function to produce a binaurally rendered sound field, the transfer function depending on the parameter value.

[0006] In another general aspect, an apparatus includes memory and processing circuitry coupled to the memory. The processing circuitry is configured to identify a set of sound sources in a three-dimensional space that, within a threshold, intersect a curve, the curve being defined by a parameter value that is constant along the curve. The processing circuitry is also configured to sum sound fields from the set of sound sources to produce a net sound field. The processing circuitry is further configured to perform a convolution operation on the net sound field and a transfer function to produce a binaurally rendered sound field, the transfer function depending on the parameter value.

[0007] The details of one or more implementations are set forth in the accompanying drawings and the description below. Other features will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] FIG. 1A is a diagram that illustrates an example audio arrangement with a set of sound sources and a listener.

[0009] FIG. 1B is a diagram that illustrates an example computation involved in binaural rendering of the audio from the sound sources in FIG. 1A for the listener.

[0010] FIG. 2A is a diagram that illustrates an example audio arrangement with a set of sound sources arranged along an arc and a listener.

[0011] FIG. 2B is a diagram that illustrates an example computation involved in binaural rendering of the audio from the sound sources in FIG. 2A for the listener.

[0012] FIG. 3A is a diagram that illustrates an example audio arrangement with a set of sound sources arranged along a line and a listener.

[0013] FIG. 3B is a diagram that illustrates an example computation involved in binaural rendering of the audio from the sound sources in FIG. 3A for the listener.

[0014] FIG. 4 is a flow chart that illustrates an example process of determining sound sources that lie along a curve of constant parameter.

[0015] FIG. 5 is a diagram that illustrates an example electronic environment for performing the improved techniques described herein.

[0016] FIG. 6 is a flow chart that illustrates an example method of performing the binaural rendering according to the improved techniques described herein.

DETAILED DESCRIPTION

[0017] A sound field generated at an array of virtual sound sources can reproduce the effect of a sound source from any vantage point relative to a listener. Such a sound field may be decoded and used in the delivery of audio through headphone speakers in, e.g., Virtual Reality (VR) and/or augmented reality (AR) systems. Frequently, such rendering takes into account the effect of a human auditory system using a set of Head-Related Transfer Functions (HRTFs). Performing convolutions on signals from each sound source with the set of HRTFs provides the listener with a faithful reproduction of the sound source. In addition, an audio signal may reflect from obstacles and the boundaries of a room; this is represented in a reverb transfer function.

[0018] Approaches to performing binaural rendering involve performing four convolutions per sound source signal, i.e., a convolution of a HRTF with a decoded signal for that sound source for each ear (L,R) as well as a convolution of a reverb transfer function for each ear.

[0019] It is noted that the HRTFs and the reverb transfer functions (TFs) are examples of transfer functions used in the computation of the sound field at the ears of a listener. Generally speaking, a a transfer function herein is a function in the time domain that, when convolved with an audio field from a sound source, describes a physical effect on the audio field from the sound source. For example, a HRTF that provides binaural rendering of the audio field into the sound field at the ear describes the physical effect of the human head anatomy on the audio field. Also, the reverb TF describes the effect of room reverberation on the audio field.

[0020] A technical problem with the above-described conventional approaches is that the number of convolutions of HRTFs and reverb transfer functions is linearly proportional to the number of spatial sound sources that are to be used in a three-dimensional space. In cases with a larger number of sound sources, the linear proportional relationship may cause increased usage of computational resources and computational bottlenecks when there are many sound sources in the three-dimensional space. In addition, there may be increased memory usage, increased requirements for computational capacity, increased power usage, and increased latency due to the number of convolutions that must be computed.

[0021] A technical solution to some or all of the above technical problems includes using the fact that the HRTF depends on a sound source position on a unit sphere relative to a listener while the reverb transfer function depends on a range relative to the listener. In this way, sound source positions along a line in three-dimensional space have an identical HRTF while speaker positions along a circular arc in three-dimensional space have an identical reverb transfer function. Accordingly, those sound source positions along a line use a common HRTF in a single convolution for each ear, and those speaker positions along an arc use a common reverb transfer function in a single convolution for each ear. As a result, the number of convolutions that need to be computed will be decreased, thus reducing some or all of the necessary computational capacity, used computation, used memory resources, and power usage. The latency of calculating the transfer functions may also be reduced.

[0022] A technical advantage of the above-described technical solution is that it reduces the computational burden of the binaural rendering of the audio fields from the sound sources, especially when there are a large number of such sound sources.

[0023] FIG. 1A is a diagram that illustrates an example top-down view of an audio arrangement 100 with a group of sound sources and a listener. As shown in FIG. 1A, there is a listener 110 and a group of sound sources (audio sources) 120(1), 120(2), 120(3), 120(4),..., 120(N) of interest arranged in three-dimensional space in a vicinity of the listener. The listener 110 and the sound sources 120(1..N) are arranged in a room with walls (not shown).

[0024] The listener 110 is a human with a left ear 112(L) and a right ear 112(R). Herein, the listener is stationary with respect to the sound sources 120(1..N), but this is not a requirement.

[0025] Each sound source 120(k), $1 \leq k \leq N$, is a source of an audio signal that propagates to the listener 110. Each sound source 120(k) is at a respective distance d_k from the center of the head of the listener 110. With respect to the center of the head of the listener 110, each sound source 120(k) is at a polar angle θ_k and azimuthal angle ϕ_k (e.g., at a point on the surface of a unit sphere (θ_k, ϕ_k) with the center of the listener's head at center). Each sound source 120(k) may be a speaker, a person, or any audio source.

[0026] FIG. 1B is a diagram that illustrates an example computation 150 involved in binaural rendering of the audio from sound sources 120(1..N) in FIG. 1A for the listener 110. The audio field y rendered at the listener's left ear 112(L) from all of the sound sources 120(1..N) is derived as follows. Consider the audio field received at the listener from the kth sound source $x^{(k)}$. To render this field binaurally, the field is convolved with a left HRTF $H^{(L,k)}$ from the kth sound source to produce an intermediate field, and this intermediate field is convolved with the reverb TF from the kth sound source $U^{(k)}$. The result is as follows for the temporal signal.

$$y^{(L)} = \sum_{k=1}^{N} x^{(k)} \otimes H^{(L,k)} \otimes U^{(k)}, \tag{1}$$

where $\otimes$ denotes convolution. The result for the right ear is similar but using the right HRTF $H^{(R,k)}$.

$$y^{(R)} = \sum_{k=1}^{N} x^{(k)} \otimes H^{(R,k)} \otimes U^{(k)}. \tag{2}$$

[0026] The computation of the above equations is illustrated in FIG. 1B. The kth sound source is associated with a left and right HRTF 160(k). A computer (e.g., processing circuitry) performs a convolution 165(k) of the signal from the sound source at the listener with the left HRTF from that sound source to produce an intermediate left field. The computer also performs a convolution 170(k) of the signal from the sound source at the listener with the right HRTF from that sound source to produce an intermediate right field. The kth sound source is also associated with a reverb TF 175(k). The computer performs a convolution 180(k) of the intermediate left field with the reverb TF 175(k) to produce a rendered audio field in the left ear and a convolution 185(k) of the intermediate right field with the reverb TF 175(k) to produce a rendered audio field in the right ear. The computer sums the rendered audio field in the left and right ear over the sound sources to produce the left binaurally rendered audio field 190 and the right binaurally rendered audio field 195.

[0027] It is noted that Eqs. (1) and (2) represent summations over sound sources of two convolution operations. Accordingly, there are four convolutions for each sound source, and the results are summed. Nevertheless, in some situations the HRTFs and reverb TFs may be the same for certain sound sources. In that case, because of the linearity of the convolution operation, those HRTFs and reverb TFs may be factored, which reduces the number of convolutions to perform. For example, for sound sources that are arranged along a curve having a parameter value that is constant along that curve, either or both of the HRTFs and reverb TFs may be identical along that curve. In some implementations, the curve is a circular arc and the parameter is distance from the listener: because the reverb TF depends on radius, all sound sources arranged on that circular arc may have the same reverb TF. In some implementations, the curve is a line and the parameter is angle; because the HRTF depends on the angle, all sound sources arranged on that line may have the same left and right HRTFs. This phenomenon is illustrated with respect to FIGs. 2A, 2B, 3A,and 3B.

[0028] It is also noted that each sound source (sound source) is represented in FIG. 1A as a circle of finite radius. This reflects the fact that a sound source has a position (e.g., the center of the circle) but that the center may lie within a threshold from a curve. Graphically, the curve intersecting any part of the circle is equivalent to the sound source being a threshold from the curve.

[0029] It is further noted that a curve having a parameter value constant along the curve is a member of a family of curves that are defined by that parameter value. For example, a circle is a member of a family of curves, each being a curve of constant distance from a listener at its center. A line is a family of curves, each being oriented at a constant angle with respect to the listener.

[0030] FIG. 2A is a diagram that illustrates an example audio arrangement 200 with a set of sound sources 220(1), 220(2), and 220(3) arranged along a circular arc 230 and a listener 210 at the center of the circular arc. It is noted that the circular arc 230 does not go through the center of each sound source 220(1), 220(2), 220(3), but rather the sound sources 220(1), 220(2), 220(3) are positioned substantially along the circular arc 230 with their respective centers within a threshold from the circular arc 230. In some implementations, the threshold is about 1 cm, 5 cm, 10 cm, 20 cm, 50 cm, 1 meter, 2 m, 5 m, and so on.

[0031] FIG. 2B is a diagram that illustrates an example computation 250 involved in binaural rendering of the audio from sound sources for the listener 210. In this case, however, sound sources 255(1..N) are arranged along a circular arc with the listener at its center. That is, each of the sound sources 255(1..N) are at the same distance from the listener. Accordingly, each of the sound sources 255(1..N) have the same reverb TF U.

[0032] Thus, Eqs. (1) and (2) take the following form.

$$y^{(L)} = \left(\sum_{k=1}^{N} x^{(k)} \otimes H^{(L,k)}\right) \otimes U, \tag{3}$$

$$y^{(R)} = \left(\sum_{k=1}^{N} x^{(k)} \otimes H^{(R,k)}\right) \otimes U. \tag{4}$$

That is, the number of convolutions has been reduced from $4N$ to $2N + 2$.

[0033] FIG. 2B illustrates the computation in Eqs. (3) and (4). For each sound source 255(k), there is a corresponding left and right HRTF 260(k). A computer (e.g., processing circuitry) performs a convolution 265(k) on the left HRTF and a sound field at the listener from that sound source to produce a left intermediate field corresponding to that sound source. The computer also performs a convolution 270(k) on the right HRTF and the sound field at the listener from that sound source to produce a right intermediate field corresponding to that sound source. The computer then sums the left intermediate fields over the sound sources to produce a left net sound field 275 and sums the right intermediate fields over the sound sources to produce a right net sound field 280. The computer then performs a single convolution 290 of the left net sound field 275 and the reverb TF 285 to produce a left binaurally rendered audio field and a single convolution 295 of the right net sound field 280 and the reverb TF 285 to produce a right binaurally rendered audio field.

[0034] To summarize, the fact that the reverb TFs of the sound sources are essentially the same when the sound sources lie along a circular arc allows the reverb TF to be factored out of the sums in Eqs. (3) and (4) above. This factoring out of the reverb TF from the sums in Eqs. (3) and (4) significantly reduces the number of convolutions that need to be performed to binaurally render the audio fields in the listener’s ears.

[0035] FIG. 3A is a diagram that illustrates an example audio arrangement 300 with a set of sound sources 320(1) and 320(2) arranged along a line 330 and a listener 310 at the origin of the line 330. It is noted that the line 330 does not go through the center of each sound source 320(1), 320(2), but rather the sound sources 320(1), 320(2) are positioned substantially along the line 330 with their respective centers within a threshold from the line 330. In some implementations, the threshold is about 1 degree, 2 degrees, 5 degrees, 10 degrees, 20 degrees, and so on.

[0036] FIG. 3B is a diagram that illustrates an example computation 350 involved in binaural rendering of the audio from sound sources for the listener 310. In this case, however, sound sources 355(1..N) are arranged along a line with the listener at the origin. That is, each of the sound sources 355(1..N) are along the same angle with respect to the listener. Accordingly, each of the sound sources 355(1..N) have the same HRTFs $H^{(L)}$ and $H^{(R)}$.

[0037] Thus, Eqs. (1) and (2) take the following form.

$$y^{(L)} = \left(\sum_{k=1}^{N} x^{(k)} \otimes U^{(k)}\right) \otimes H^{(L)}, \tag{5}$$

$$y^{(R)} = \left(\sum_{k=1}^{N} x^{(k)} \otimes U^{(k)}\right) \otimes H^{(R)}. \tag{6}$$

That is, the number of convolutions has been reduced from $4N$ to $N + 2$.

[0038] FIG. 3B illustrates the computation in Eqs. (5) and (6). For each sound source 355(k), there is a corresponding reverb TF 360(k). A computer (e.g., processing circuitry) performs a convolution 365(k) on the reverb TF and a sound field at the listener from that sound source to produce an intermediate field corresponding to that sound source. The computer then sums the intermediate fields over the sound sources to produce a net sound field 375. The computer then performs a single convolution 390 of the net sound field 375 and the left HRTF 385 to produce a left binaurally rendered audio field and a single convolution 395 of the net sound field 375 and the right HRTF 385 to produce a right binaurally rendered audio field.

[0039] To summarize, the fact that the HRTFs of the sound sources are essentially the same when the sound sources lie along a line allows the HRTFs to be factored out of the sums in Eqs. (5) and (6) above. This factoring out of the HRTFs from the sums in Eqs. (5) and (6) significantly reduces the number of convolutions that need to be performed to binaurally render the audio fields in the listener's ears.

[0040] One issue that has not yet been discussed in detail is finding sound sources from a set of sound sources that lie along a curve such as a circular arc or a line. In a situation with a set of sound sources in a room with a listener, it is usually not clear which sound sources may lie along such a curve. An algorithm for finding such sound sources is illustrated with respect to FIG. 4.

[0041] FIG. 4 is a flow chart that illustrates an example process 400 of determining sound sources that lie along a curve of constant parameter, such as a circular arc with the listener at the center or a line through the listener. Consider a curve $\gamma(p)$ where p is a parameter. For example, $\gamma(r)$ may be a circle of radius r or $\gamma(\theta,\phi)$ is a line through a point (θ,ϕ) on the unit sphere. The problem is to evaluate the following quantity:

$$\underset{p}{\operatorname{argmax}} \sum_{k=1}^{N} \mathcal{I}[\|S_k - \gamma(p)\|_2^2 < \varepsilon] \tag{7}$$

where $\mathcal{I}$ is an indicator function (e.g., 1 if the condition in the argument is true, 0 if false), S_k is the position of the kth sound source in three-dimensional space, and ε is a distance threshold. That is, the goal is to find the curve (e.g., circular arc, line) along which there are as many sound sources within a threshold ε as possible.

[0042] The approach described herein is a grid search over the parameter p. That is, the parameter p takes on the values p_i, $i \in [1, i_{\max}]$. Nevertheless, other methods can also be used to determine the optimal parameter value, e.g., simulated annealing.

[0043] Referring to the flow chart in FIG. 4, at 405, the values used in the grid search are initialized: $i = 1, n = 0$. At 410, the value of the parameter is set to p_i. For example, if the curve is a circle, then the parameter is the radius, and the grid search may take the form $p_i = 0.5\ i$, where i varies from 1 to 20.

[0044] At 415, the sound sources that satisfy the criterion $\|S_k - \gamma(p_i)\|_2^2 < \varepsilon$ are determined. This criterion is equivalent to a distance metric between each sound source and the curve being less than a threshold. In some implementations, in which the curve is a circular arc. the threshold is about 1 cm, 5 cm, 10 cm, 20 cm, 50 cm, 1 meter,, 1.5 m 2 m, 5 m, and so on, or values in between those thresholds. In some implementations, in which the curve is a line, the threshold is 1 degree, 2 degrees, 5 degrees, 10 degrees, 15 degrees, 20 degrees, and so on, or values in between those thresholds.

[0045] At 420, the number of sound sources satisfying the criterion is counted. If the number of sound sources is greater than the current value of n, then at 425 the current value of n is set to the number of sound sources satisfying the criterion and the current value of i is noted. In either case, at 430, the value of i is incremented at 430.

[0046] At 435, if the value of $i > i_{\max}$ then the parameter value at the current value of i is output and the sound sources along the curve corresponding to that parameter value are identified. If not, then the algorithm returns to step 410. In this way, the curve that includes the maximum number of sound sources may be found.

[0047] FIG. 5 is a diagram that illustrates an example electronic environment for performing binaural rendering of sound fields. The processing circuitry 520 includes a network interface 522, one or more processing units 524, and nontransitory memory (storage medium) 526.

[0048] In some implementations, one or more of the components of the processing circuitry 520 can be, or can include processors (e.g., processing units 524) configured to process instructions stored in the memory 526 as a computer program product. Examples of such instructions as depicted in FIG. 5 include sound source identification manager 530, sound field manager 540, and convolution manager 550. Further, as illustrated in FIG. 5, the memory 526 is configured to store various data, which is described with respect to the respective services and managers that use such data.

[0049] The sound source identification manager 530 is configured to identify sound sources that are positioned along a curve by which a transfer function (e.g., a reverb TF, a HRTF) may have a constant value, as for example in the algorithm described with respect to FIG. 4. The sound source identification manager 530 as a consequence is configured to output sound source position data 532, which includes identifiers for sound sources along the curve having the parameter value determined from Eq. (7).

[0050] The sound field manager 540 is configured to produce sound field data 542 by summing the sound fields at the listener from the sound sources identified in the sound source position data 532.

[0051] The convolution manager 550 is configured to perform the convolutions with the reverb TF and HRTFs for binaural rendering of the sound fields in the listener's ears, according to Eqs. (3) and (4), or (5) and (6). To perform the convolutions, the convolution manager 550 uses convolution data 552, including the sound field data 553, HRTF data 554, and reverb TF data 555.

[0052] In some implementations, the sound field data 553 may be derived from the sound field data 542 by propagating the sound field represented by sound field data 542 to each of the listener's ears.

[0053] The components (e.g., modules, processing units 524) of processing circuitry 520 can be configured to operate based on one or more platforms (e.g., one or more similar or different platforms) that can include one or more types of hardware, software, firmware, operating systems, runtime libraries, and/or so forth. In some implementations, the components of the processing circuitry 520 can be configured to operate within a cluster of devices (e.g., a server farm). In such an implementation, the functionality and processing of the components of the processing circuitry 520 can be distributed to several devices of the cluster of devices.

[0054] The components of the processing circuitry 520 can be, or can include, any type of hardware and/or software configured to process private data from a wearable device in a split-compute architecture. In some implementations, one or more portions of the components shown in the components of the processing circuitry 520 in FIG. 5 can be, or can include, a hardware-based module (e.g., a digital signal processor (DSP), a field programmable gate array (FPGA), a memory), a firmware module, and/or a software-based module (e.g., a module of computer code, a set of computer-readable instructions that can be executed at a computer). For example, in some implementations, one or more portions of the components of the processing circuitry 520 can be, or can include, a software module configured for execution by at least one processor (not shown). In some implementations, the functionality of the components can be included in different modules and/or different components than those shown in FIG. 5, including combining functionality illustrated as two components into a single component.

[0055] The network interface 522 includes, for example, wireless adaptors, and the like, for converting electronic and/or optical signals received from the network to electronic form for use by the processing circuitry 520. The set of processing units 524 include one or more processing chips and/or assemblies. The memory 526 includes both volatile memory (e.g., RAM) and non-volatile memory, such as one or more ROMs, disk drives, solid state drives, and the like. The set of processing units 524 and the memory 526 together form processing circuitry, which is configured and arranged to carry out various methods and functions as described herein.

[0056] Although not shown, in some implementations, the components of the processing circuitry 520 (or portions thereof) can be configured to operate within, for example, a data center (e.g., a cloud computing environment), a computer system, one or more server/host devices, and/or so forth. In some implementations, the components of the processing circuitry 520 (or portions thereof) can be configured to operate within a network. Thus, the components of the processing circuitry 520 (or portions thereof) can be configured to function within various types of network environments that can include one or more devices and/or one or more server devices. For example, the network can be, or can include, a local area network (LAN), a wide area network (WAN), and/or so forth. The network can be, or can include, a wireless network and/or wireless network implemented using, for example, gateway devices, bridges, switches, and/or so forth. The network can include one or more segments and/or can have portions based on various protocols such as Internet Protocol (IP) and/or a proprietary protocol. The network can include at least a portion of the Internet.

[0057] In some implementations, one or more of the components of the processing circuitry 520 can be, or can include, processors configured to process instructions stored in a memory. For example, sound source identification manager 530 (and/or a portion thereof), sound field manager 540 (and/or a portion thereof), and convolution manager 550 (and/or a portion thereof) are examples of such instructions.

[0058] In some implementations, the memory 526 can be any type of memory such as a random-access memory, a disk drive memory, flash memory, and/or so forth. In some implementations, the memory 526 can be implemented as more than one memory component (e.g., more than one RAM component or disk drive memory) associated with the components of the processing circuitry 520. In some implementations, the memory 526 can be a database memory. In some implementations, the memory 526 can be, or can include, a non-local memory. For example, the memory 526 can be, or can include, a memory shared by multiple devices (not shown). In some implementations, the memory 526 can be associated with a server device (not shown) within a network and configured to serve the components of the processing circuitry 520. As illustrated in FIG. 5, the memory 526 is configured to store various data, including sound source position data 532, sound field data 542, and convolution data 552.

[0059] FIG. 6 is a flow chart that illustrates an example method 600 of performing binaural rendering of sound fields. The method 600 may be performed using the processing circuitry 520 of FIG. 5.

[0060] At 602, the sound source identification manager 530 identifies a set of sound sources (e.g., sound sources 220(1..3)) in a three-dimensional space that are positioned substantially along a curve (e.g., circular arc 230) having a parameter value that is constant along the curve.

[0061] At 604, the sound field manager 540 sums sound fields from the set of sound sources to produce a net sound field (e.g., net sound field 275).

[0062] At 606, the convolution manager 550 performs a convolution operation on the net sound field and a transfer function to produce a binaurally rendered sound field, the transfer function depending on the parameter value so that the transfer function has identical values for the set of sound sources positioned substantially along the curve.

[0063] Specific structural and functional details disclosed herein are merely representative for purposes of describing example embodiments. Example embodiments, however, may be embodied in many alternate forms and should not be construed as limited to only the embodiments set forth herein.

[0064] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the embodiments. As used herein, the singular forms "a," "an," and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms "comprises," "comprising," "includes," and/or "including," when used in this specification, specify the presence of the stated features, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, steps, operations, elements, components, and/or groups thereof.

[0065] It will be understood that when an element is referred to as being "coupled," "connected," or "responsive" to, or "on," another element, it can be directly coupled, connected, or responsive to, or on, the other element, or intervening elements may also be present. In contrast, when an element is referred to as being "directly coupled," "directly connected," or "directly responsive" to, or "directly on," another element, there are no intervening elements present. As used herein the term "and/or" includes any and all combinations of one or more of the associated listed items.

[0066] Spatially relative terms, such as "beneath," "below," "lower," "above," "upper," and the like, may be used herein for ease of description to describe one element or feature in relationship to another element(s) or feature(s) as illustrated in the figures. It will be understood that the spatially relative terms are intended to encompass different orientations of the device in use or operation in addition to the orientation depicted in the figures. For example, if the device in the figures is turned over, elements described as "below" or "beneath" other elements or features would then be oriented "above" the other elements or features. Thus, the term "below" can encompass both an orientation of above and below. The device may be otherwise oriented (rotated 70 degrees or at other orientations) and the spatially relative descriptors used herein may be interpreted accordingly.

[0067] Example embodiments of the concepts are described herein with reference to cross-sectional illustrations that are schematic illustrations of idealized embodiments (and intermediate structures) of example embodiments. As such, variations from the shapes of the illustrations as a result, for example, of manufacturing techniques and/or tolerances, are to be expected. Thus, example embodiments of the described concepts should not be construed as limited to the particular shapes of regions illustrated herein but are to include deviations in shapes that result, for example, from manufacturing. Accordingly, the regions illustrated in the figures are schematic in nature and their shapes are not intended to illustrate the actual shape of a region of a device and are not intended to limit the scope of example embodiments.

[0068] It will be understood that although the terms "first," "second," etc. may be used herein to describe various elements, these elements should not be limited by these terms. These terms are only used to distinguish one element from another. Thus, a "first" element could be termed a "second" element without departing from the teachings of the present embodiments.

[0069] Unless otherwise defined, the terms (including technical and scientific terms) used herein have the same meaning as commonly understood by one of ordinary skill in the art to which these concepts belong. It will be further understood that terms, such as those defined in commonly used dictionaries, should be interpreted as having a meaning that is consistent with their meaning in the context of the relevant art and/or the present specification and will not be interpreted in an idealized or overly formal sense unless expressly so defined herein.

[0070] While certain features of the described implementations have been illustrated as described herein, many modifications, substitutions, changes, and equivalents will now occur to those skilled in the art. It is, therefore, to be understood that the appended claims are intended to cover such modifications and changes as fall within the scope of the implementations. It should be understood that they have been presented by way of example only, not limitation, and various changes in form and details may be made. Any portion of the apparatus and/or methods described herein may be combined in any combination, except mutually exclusive combinations. The implementations described herein can include various combinations and/or sub-combinations of the functions, components, and/or features of the different implementations described.

WHAT IS CLAIMED IS:

1. A method, comprising:

identifying a set of sound sources in a three-dimensional space having positions that, within a threshold, intersect a curve, the curve being defined by a parameter value that is constant along the curve;

summing sound fields from the set of sound sources to produce a net sound field; and

performing a convolution operation on a transfer function and the net sound field to produce a binaurally rendered sound field, the transfer function depending on the parameter value.

2. The method as in claim 1, wherein a parameter having the parameter value defines a family of curves in which the parameter value constant along the curve defines the curve in particular.

3. The method as in claim 1 or 2, wherein the parameter value is distance from a listener, the curve is a circular arc, and the transfer function is a reverb transfer function.

4. The method as in any one of claims 1-3, wherein summing the sound fields from the set of sound sources includes:

for each sound source of the set of sound sources, performing a convolution operation of a sound field at the listener from that sound source and a head-related transfer function (HRTF) corresponding to that sound source.

5. The method as in any one of claims 3-4, wherein identifying the set of sound sources includes:

for a circular arc of a specified radius, determining positions of sound sources having distances from the circular arc that are less than the threshold.

6. The method as in any one of claims 3-5, wherein identifying the set of sound sources further includes:

determining a circular arc having, among a group of sound sources of interest, a maximum number of positions of sound sources having distances from the circular arc that are less than the threshold.

7. The method as in claim 6, wherein the threshold is between 0.5 meters and 1.5 meters.

8. The method as in any one of the preceding claims, wherein the parameter value is position on a unit sphere with a listener at an origin of the unit sphere, and the curve is a line through the origin and the position on the unit sphere.

9. The method as in claim 8, wherein summing the sound fields from the set of sound sources includes:

for each sound source of the set of sound sources, performing a convolution operation of a sound field at the listener from that sound source and a reverb transfer function corresponding to that sound source.

10. The method as in claim 8 or 9, wherein identifying the set of sound sources includes:

for a line through a specified point on the unit sphere, determining positions of sound sources having distances from the line that are less than the threshold.

11. The method as in any one of claims 8-10, wherein identifying the set of sound sources further includes:

determining a line having, among a group of sound sources of interest, a maximum number of positions of sound sources having distances from the line that are less than the threshold.

12. The method as in claim 11, wherein the threshold is between 5 degrees and 15 degrees.

13. A computer program product comprising a nontransitory storage medium, the computer program product including code that, when executed by processing circuitry, causes the processing circuitry to perform a method, the method comprising:

identifying a set of sound sources in a three-dimensional space having positions that, within a threshold, intersect a curve, the curve being defined by a parameter value that is constant along the curve;

summing sound fields from the set of sound sources to produce a net sound field; and

performing a convolution operation on the net sound field and a transfer function to produce a binaurally rendered sound field, the transfer function depending on the parameter value.

14. The computer program product as in claim 13, wherein the parameter value is distance from a listener, the curve is a circular arc, and the transfer function is a reverb transfer function.

15. The computer program product as in claim 14, wherein summing the sound fields from the set of sound sources includes:

for each sound source of the set of sound sources, performing a convolution operation of a sound field at the listener from that sound source and a head-related transfer function (HRTF) corresponding to that sound source.

16. The computer program product as in any one of claims 14-15, wherein identifying the set of sound sources includes:

for a circular arc of a specified radius, determining positions of sound sources having distances from the circular arc that are less than the threshold.

17. The computer program product as in any one of claims 14-16, wherein identifying the set of sound sources further includes:

determining a circular arc having, among a group of sound sources of interest, a maximum number of positions of sound sources having distances from the circular arc that are less than the threshold.

18. The computer program product as in claim 17, wherein the threshold is between 0.5 meters and 1.5 meters.

19. The computer program product as in any one of claims 13-18, wherein the parameter value is position on a unit sphere with a listener at an origin of the unit sphere, and the curve is a line through the origin and the position on the unit sphere.

20. The computer program product as in claim 19, wherein summing the sound fields from the set of sound sources includes:

for each sound source of the set of sound sources, performing a convolution operation of a sound field at the listener from that sound source and a reverb transfer function corresponding to that sound source.

21. The computer program product as in claim 19 or 20, wherein identifying the set of sound sources includes:

for a line through a specified point on the unit sphere, determining positions of sound sources having distances from the line that are less than the threshold.

22. The computer program product as in any one of claims 19-21, wherein identifying the set of sound sources further includes:

determining a line having a maximum number of positions of sound sources having distances from the line that are less than the threshold.

23. The computer program product as in claim 22, wherein the threshold is between 5 degrees and 15 degrees.

24. An electronic apparatus configured to render sound fields in ears of a listener, the electronic apparatus comprising:

memory; and

processing circuitry coupled to the memory, the processing circuitry being configured to:

identify a set of sound sources in a three-dimensional space having positions that, within a threshold, intersect a curve, the curve being defined by a parameter value that is constant along the curve;

sum sound fields from the set of sound sources to produce a net sound field; and

perform a convolution operation on the net sound field and a transfer function to produce a binaurally rendered sound field, the transfer function depending on the parameter value.

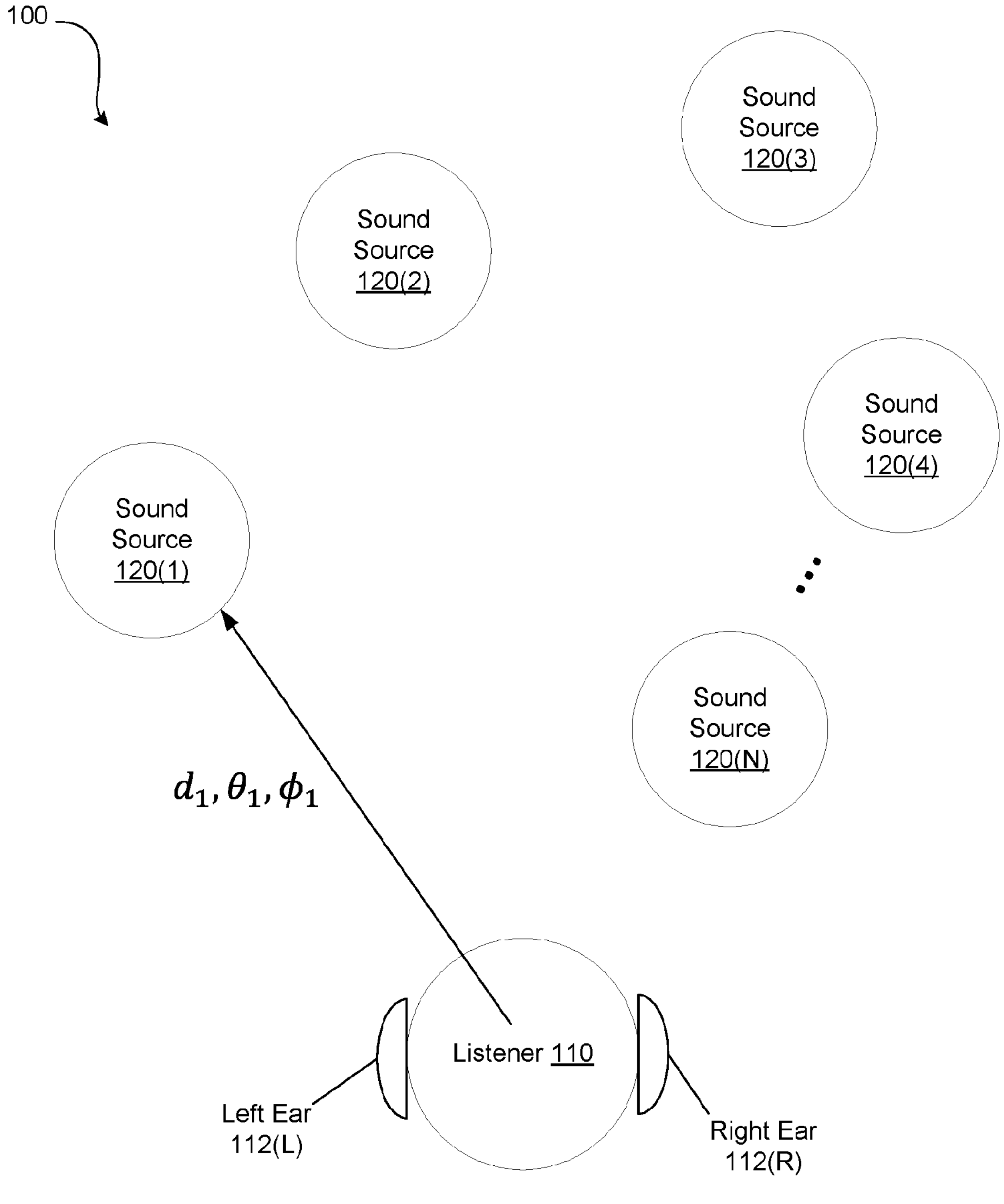
100
Sound Source 120(3)
Sound Source 120(2)
Sound Source 120(4)
Sound Source 120(1)
Sound Source 120(N)
d_1, θ_1, ϕ_1
Listener 110
Left Ear 112(L)
Right Ear 112(R)

FIG. 1A

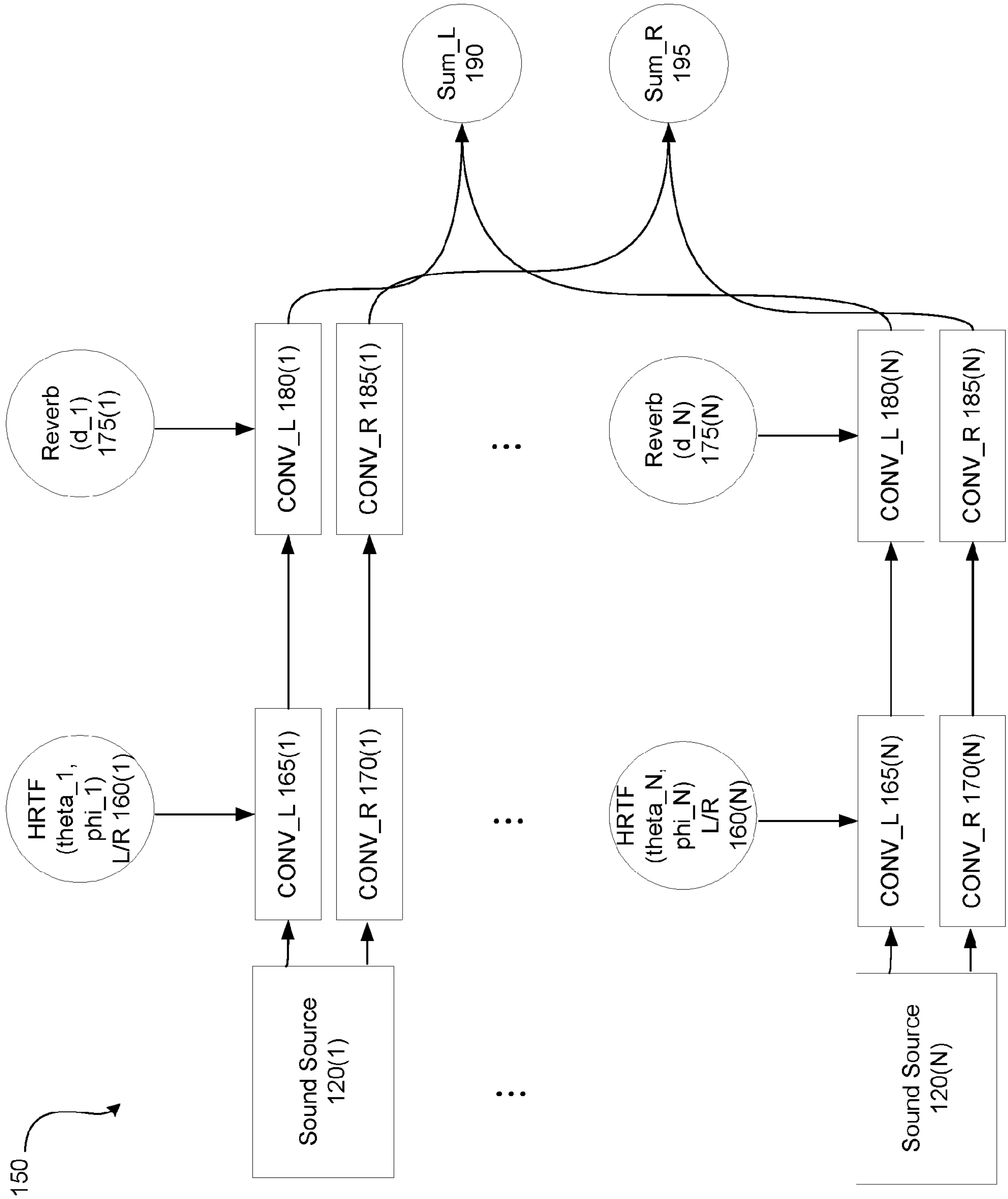
150
Sound Source 120(1)
HRTF (theta_1, phi_1) L/R 160(1)
CONV_L 165(1)
CONV_R 170(1)
Reverb (d_1) 175(1)
CONV_L 180(1)
CONV_R 185(1)
Sum_L 190
Sum_R 195
...
Sound Source 120(N)
HRTF (theta_N, phi_N) L/R 160(N)
CONV_L 165(N)
CONV_R 170(N)
Reverb (d_N) 175(N)
CONV_L 180(N)
CONV_R 185(N)

FIG. 1B

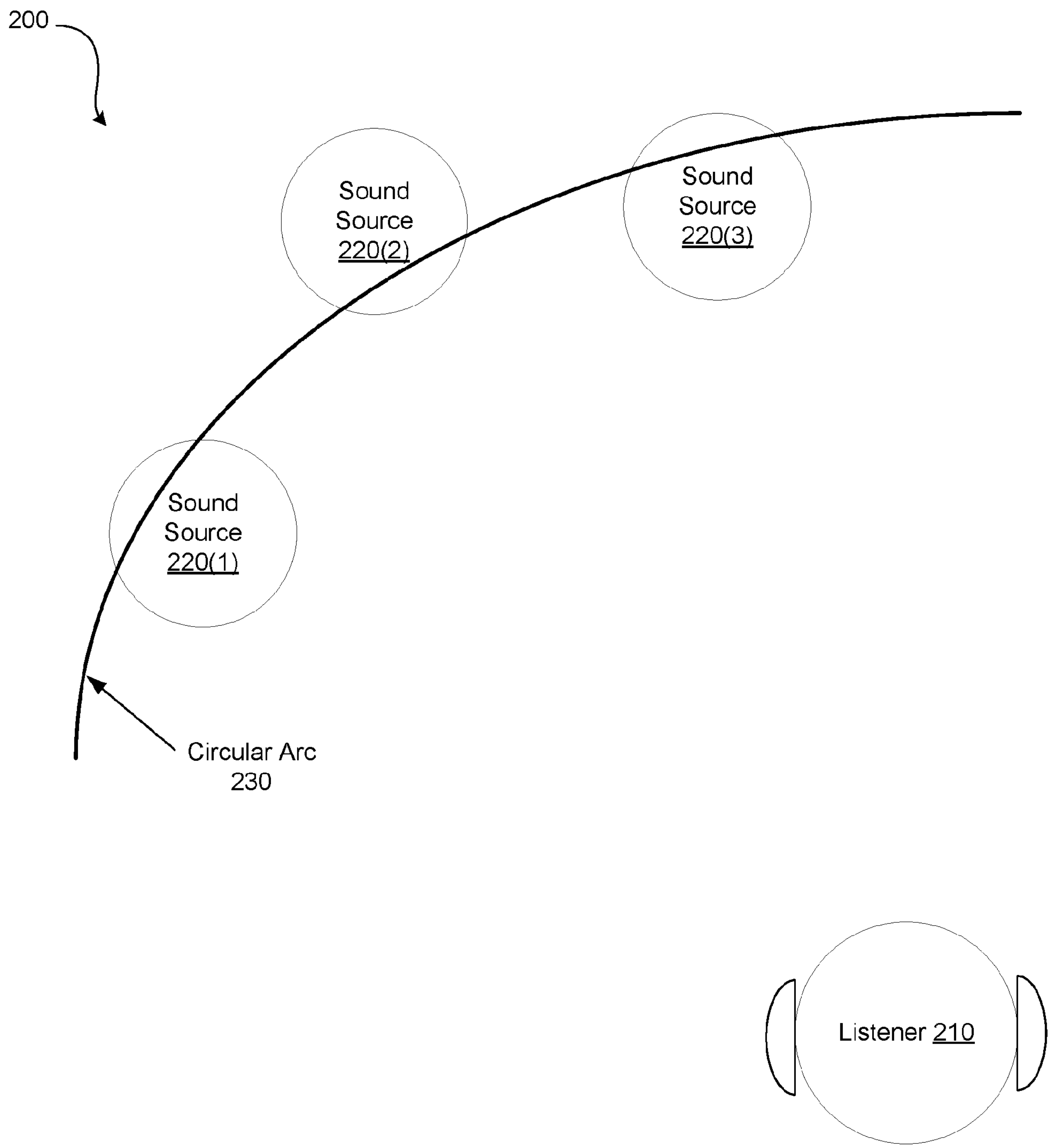
200
Sound Source 220(2)
Sound Source 220(3)
Sound Source 220(1)
Circular Arc 230
Listener 210

FIG. 2A

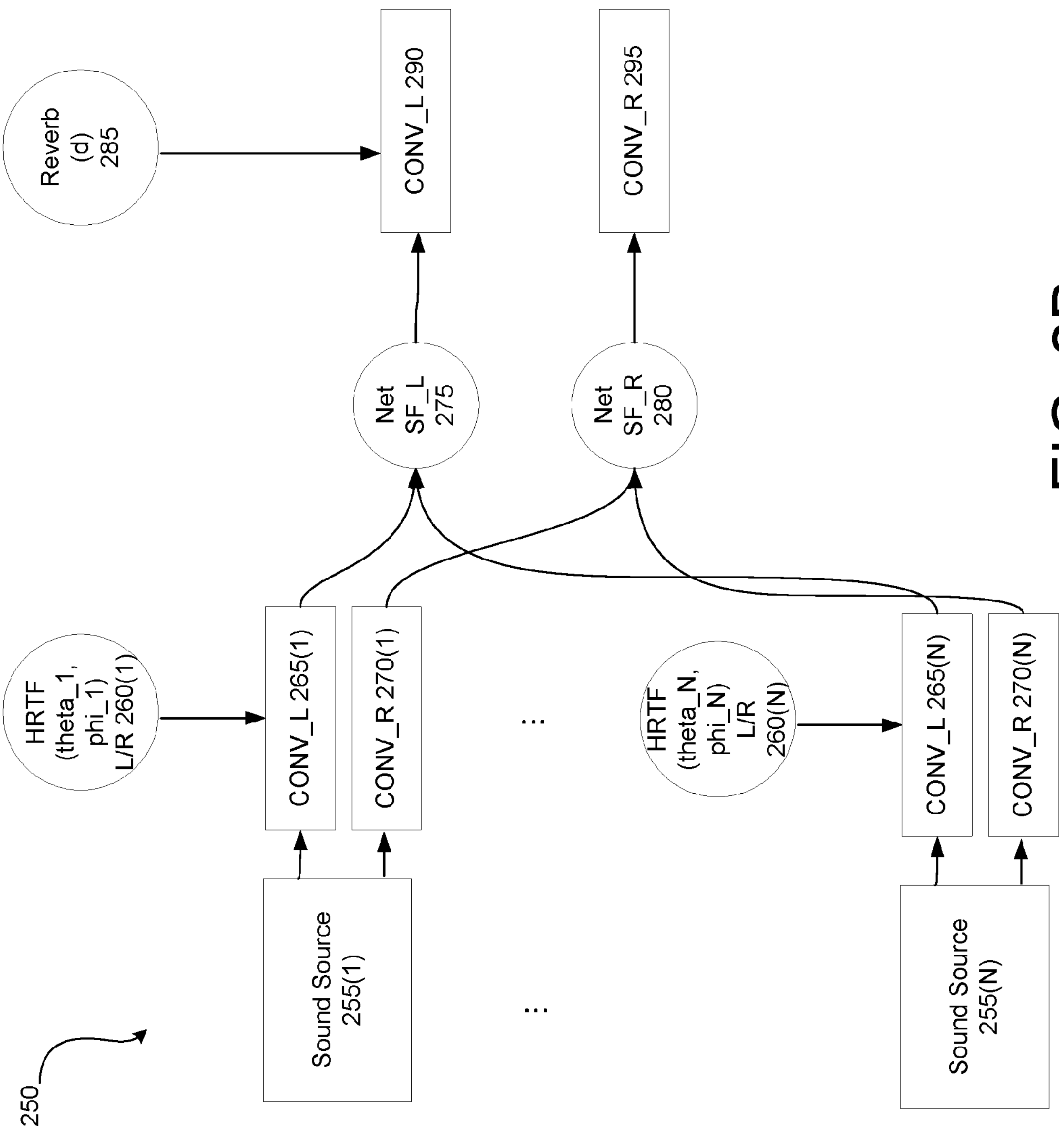
250
Sound Source 255(1)
CONV_L 265(1)
CONV_R 270(1)
HRTF (theta_1, phi_1) L/R 260(1)
...
...
Sound Source 255(N)
CONV_L 265(N)
CONV_R 270(N)
HRTF (theta_N, phi_N) L/R 260(N)
Net SF_L 275
Net SF_R 280
Reverb (d) 285
CONV_L 290
CONV_R 295

FIG. 2B

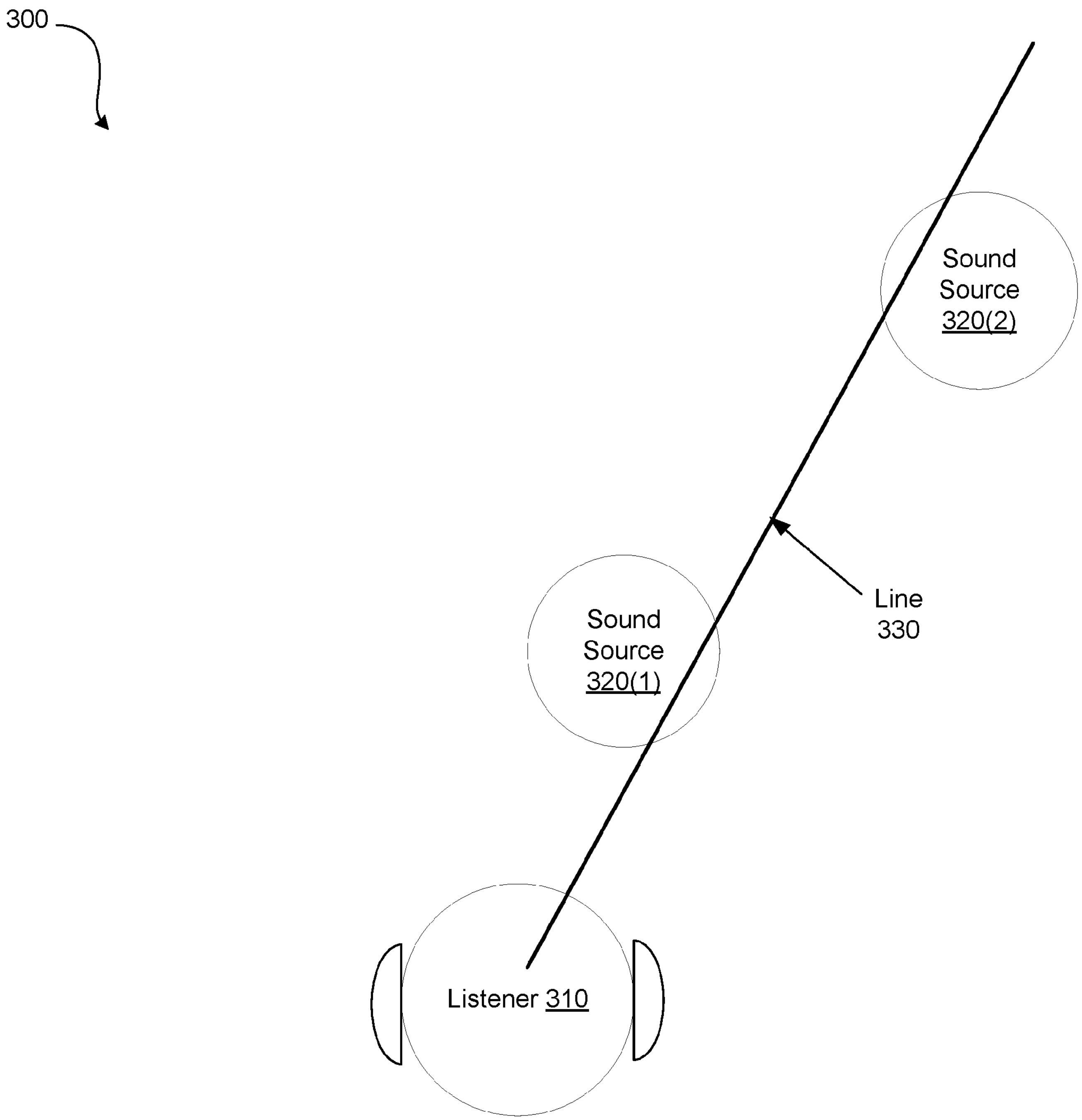
300
Sound Source 320(2)
Line 330
Sound Source 320(1)
Listener 310

FIG. 3A

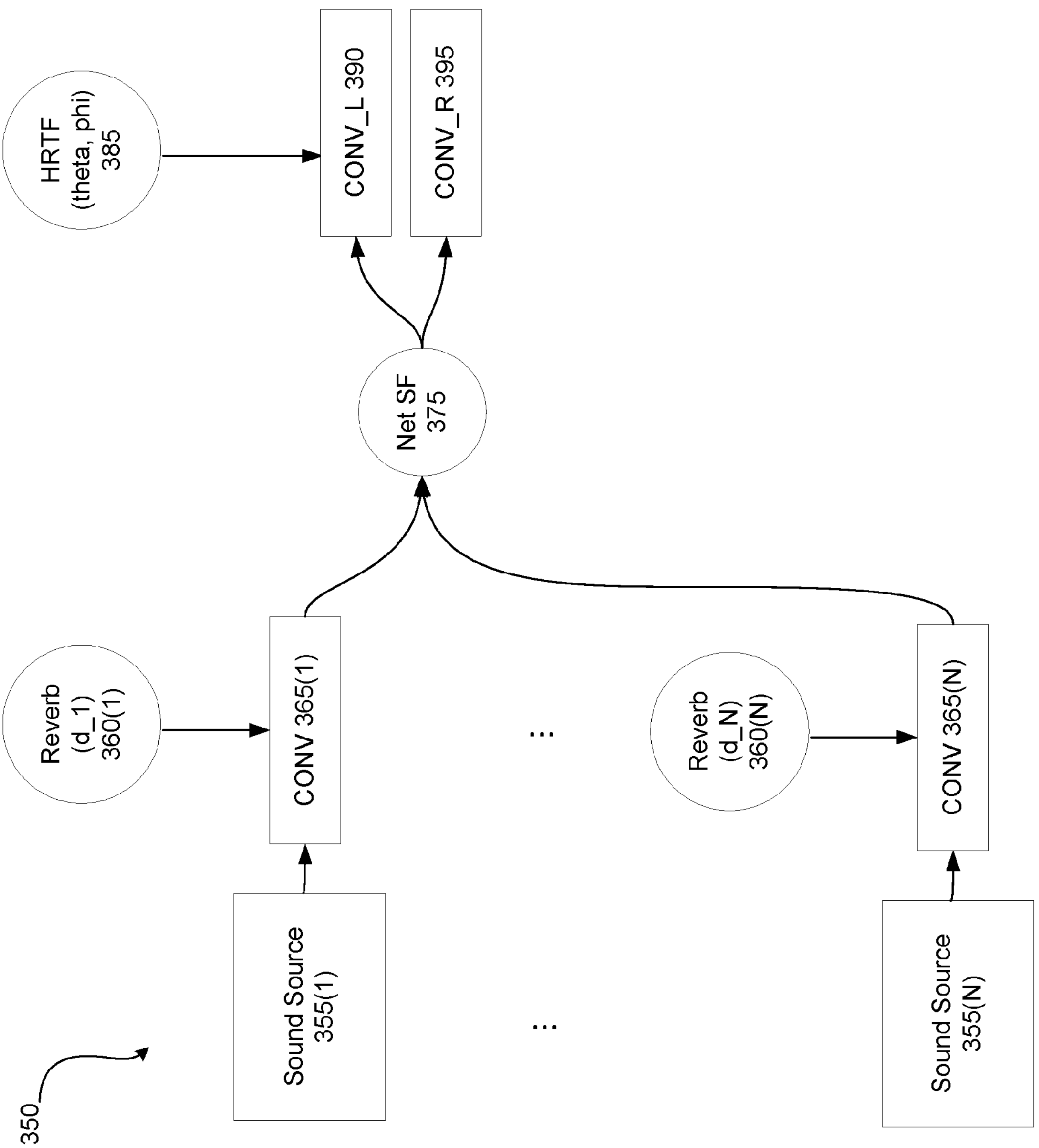
HRTF
(theta, phi)
385
CONV_L 390
CONV_R 395
Net SF
375
Reverb
(d_1)
360(1)
CONV 365(1)
...
Reverb
(d_N)
360(N)
CONV 365(N)
Sound Source
355(1)
...
Sound Source
355(N)
350

FIG. 3B

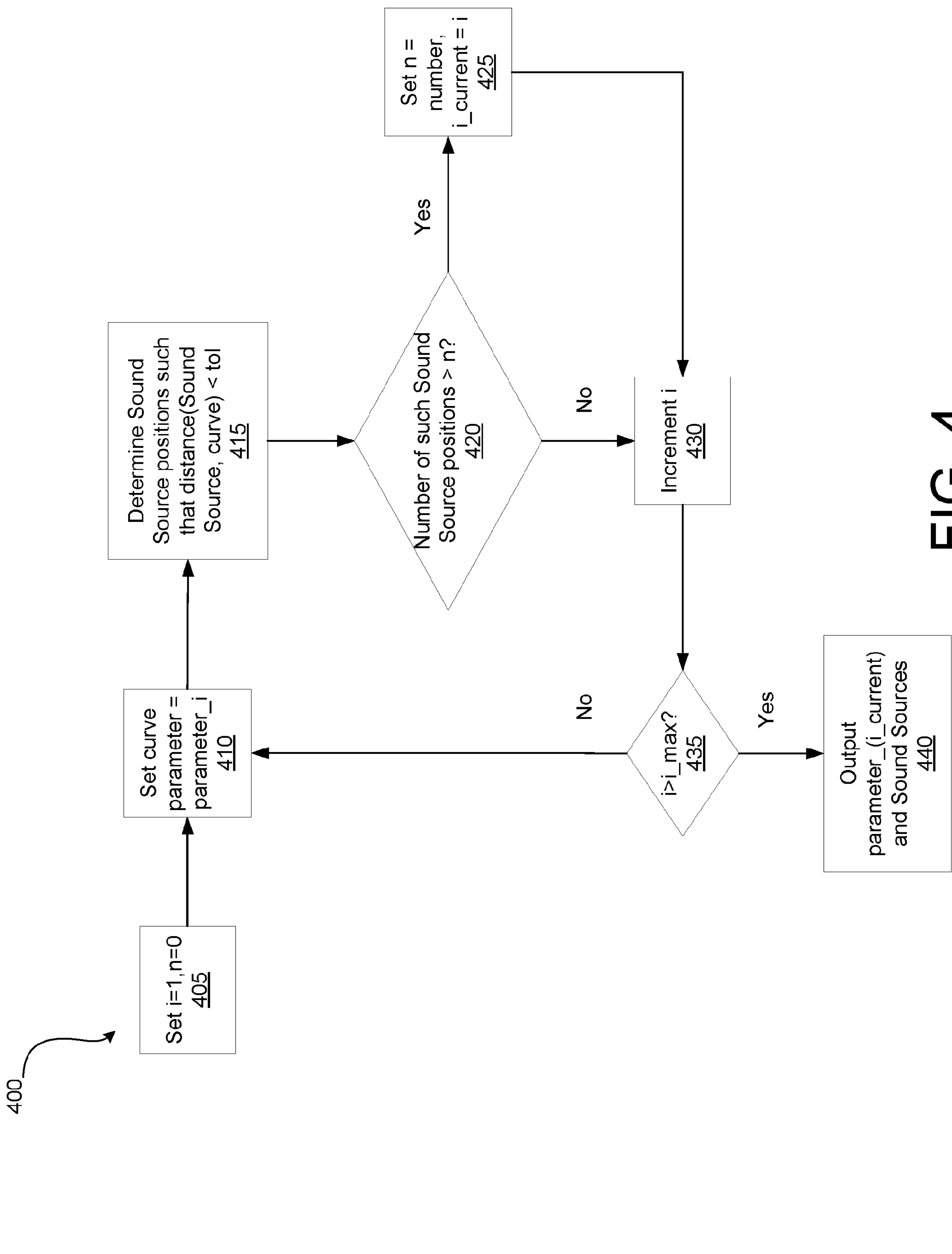
400
Set i=1,n=0 405
Set curve parameter = parameter_i 410
Determine Sound Source positions such that distance(Sound Source, curve) < tol 415
Number of such Sound Source positions > n? 420
Yes
Set n = number, i_current = i 425
No
Increment i 430
No
i>i_max? 435
Yes
Output parameter_(i_current) and Sound Sources 440

FIG. 4

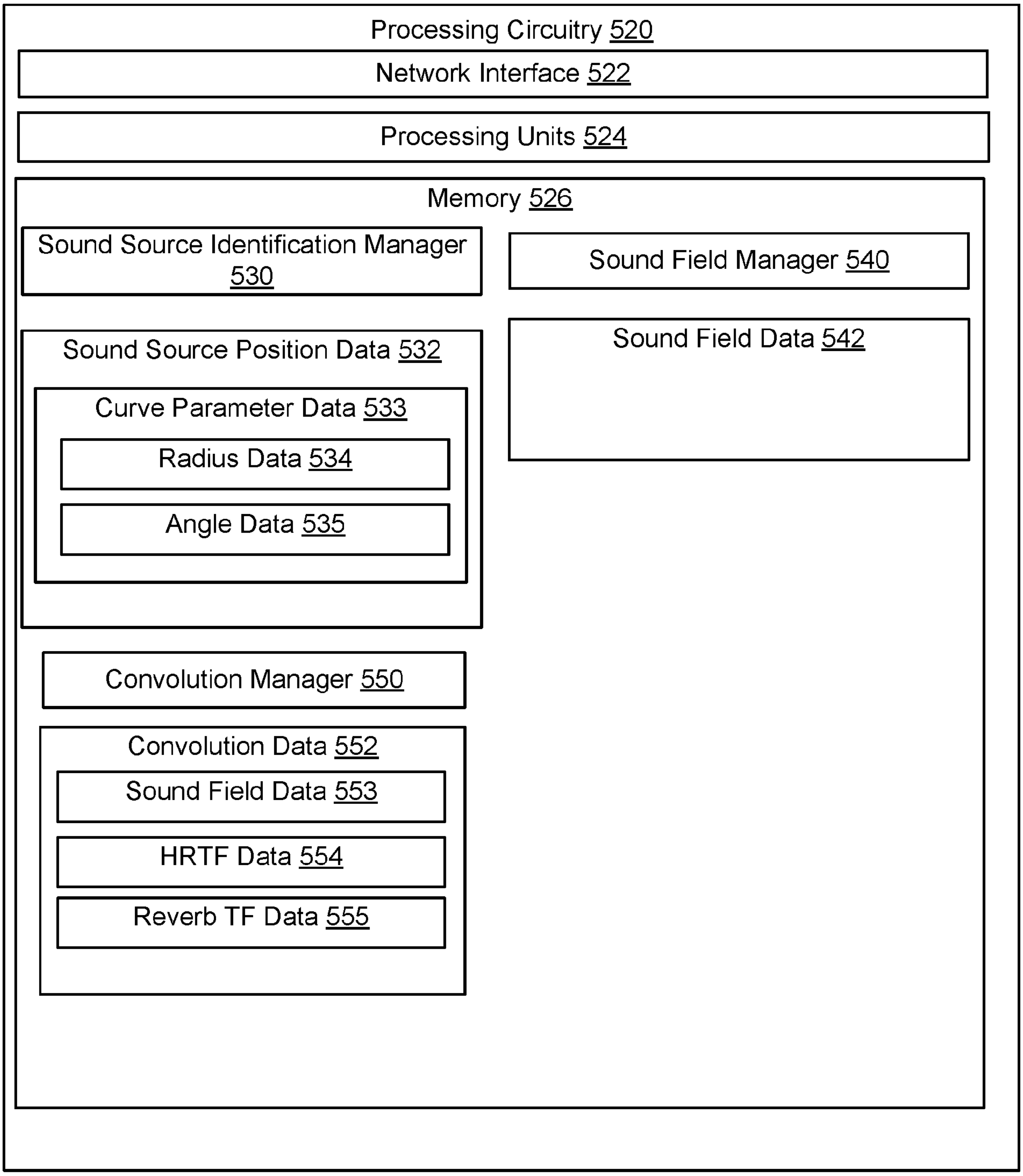
Processing Circuitry 520
Network Interface 522
Processing Units 524
Memory 526
Sound Source Identification Manager 530
Sound Field Manager 540
Sound Source Position Data 532
Sound Field Data 542
Curve Parameter Data 533
Radius Data 534
Angle Data 535
Convolution Manager 550
Convolution Data 552
Sound Field Data 553
HRTF Data 554
Reverb TF Data 555

FIG. 5

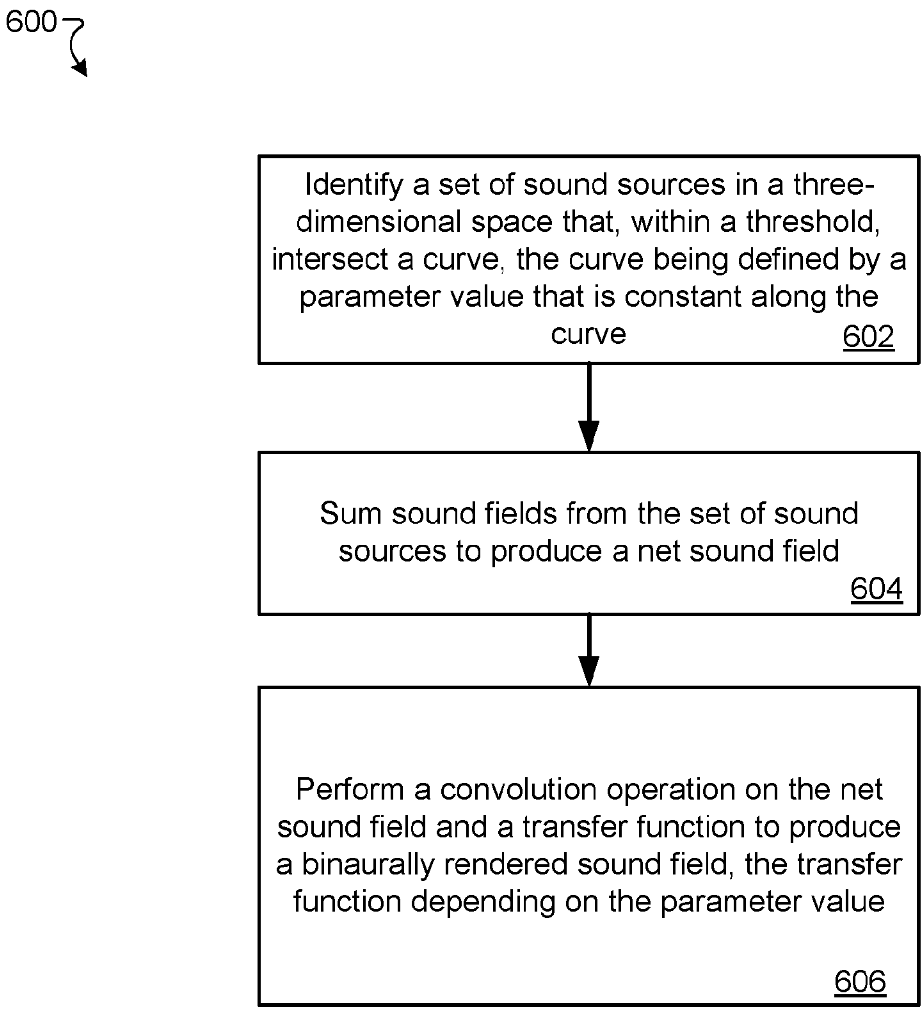
600
Identify a set of sound sources in a three-dimensional space that, within a threshold, intersect a curve, the curve being defined by a parameter value that is constant along the curve
602
Sum sound fields from the set of sound sources to produce a net sound field
604
Perform a convolution operation on the net sound field and a transfer function to produce a binaurally rendered sound field, the transfer function depending on the parameter value
606

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2023/068146

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10K15/08 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)
G10K 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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HERDER J: "Optimization of sound spatialization resource management through clustering", 3D EIZO - JOURNAL OF THREE DIMENSIONAL IMAGES, SAN JIGEN EIZO NO FORAMU, TOKYO, JP, vol. 13, no. 3, 1 September 1999 (1999-09-01), pages 59-63, XP002453153, ISSN: 1342-2189 the whole document -----	1,2, 8-13, 19-24
X	US 2020/128351 A1 (EHARA HIROYUKI [JP] ET AL) 23 April 2020 (2020-04-23) paragraphs [0009], [0039] - [0041], [0051], [0056] - [0058]; figures 4, 7 -----	1,2, 8-13, 19-24

-/--

[x] Further documents are listed in the continuation of Box C.

[x] 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	Date of mailing of the international search report
8 December 2023	18/12/2023

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

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2021/051435 A1 (YEN KUAN-CHIEH [US] ET AL) 18 February 2021 (2021-02-18) paragraphs [0006], [0083], [0088], [0092], [0097] - [0100], [0148]; figures 2,3 -----	1-7, 13-18,24

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2023/068146

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2020128351 A1	23-04-2020	CN 109792582 A	21-05-2019
		CN 114025301 A	08-02-2022
		EP 3533242 A1	04-09-2019
		EP 3822968 A1	19-05-2021
		JP 6977030 B2	08-12-2021
		JP 7222054 B2	14-02-2023
		JP 2019532579 A	07-11-2019
		JP 2022010174 A	14-01-2022
		US 2019246236 A1	08-08-2019
		US 2020128351 A1	23-04-2020
		US 2020329332 A1	15-10-2020
		US 2021067897 A1	04-03-2021
		US 2022248163 A1	04-08-2022
		WO 2018079254 A1	03-05-2018
US 2021051435 A1	18-02-2021	AU 2014374182 A1	30-06-2016
		AU 2018203746 A1	21-06-2018
		AU 2020203222 A1	04-06-2020
		AU 2022202513 A1	12-05-2022
		AU 2023203442 A1	29-06-2023
		BR 112016014949 A2	08-08-2017
		BR 122020013590 B1	06-09-2022
		BR 122020013603 B1	06-09-2022
		CA 2935339 A1	09-07-2015
		CA 3043057 A1	09-07-2015
		CA 3148563 A1	09-07-2015
		CA 3170723 A1	09-07-2015
		CN 104768121 A	08-07-2015
		CN 111065041 A	24-04-2020
		CN 114401481 A	26-04-2022
		EP 3402222 A1	14-11-2018
		EP 3806499 A1	14-04-2021
		EP 4270386 A2	01-11-2023
		JP 6215478 B2	18-10-2017
		JP 7183467 B2	05-12-2022
		JP 2017507525 A	16-03-2017
		JP 2022172314 A	15-11-2022
		JP 2023018067 A	07-02-2023
		KR 20160095042 A	10-08-2016
		KR 20180071395 A	27-06-2018
		KR 20210037748 A	06-04-2021
		KR 20220043242 A	05-04-2022
		KR 20220141925 A	20-10-2022
		MX 352134 B	10-11-2017
		RU 2637990 C1	08-12-2017
		US 2021051435 A1	18-02-2021
		US 2022182779 A1	09-06-2022
		US 2023199427 A1	22-06-2023