



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
22.07.2020 Bulletin 2020/30

(51) Int Cl.:
H04S 7/00 (2006.01)

(21) Application number: **19152619.3**

(22) Date of filing: **18.01.2019**

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

- **LAAKSONEN, Lasse Juhani**
33210 Tampere (FI)
- **LEPPÄNEN, Jussi Artturi**
33580 Tampere (FI)
- **LEHTINIEMI, Arto Juhani**
33880 Lempäälä (FI)

(71) Applicant: **Nokia Technologies Oy**
02610 Espoo (FI)

(74) Representative: **Whiting, Gary et al**
Venner Shipley LLP
5 Stirling House
Stirling Road
The Surrey Research Park
Guildford GU2 7RF (GB)

(72) Inventors:
• **ERONEN, Antti Johannes**
33820 Tampere (FI)

(54) **PROCESSING AUDIO SIGNALS**

(57) An apparatus, method and computer program is described comprising: receiving audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene and one or more virtual positions within the virtual audio scene; processing audio data of said first sound sources using an audio gain based on a distance function for the respective first sound source; and processing audio data of said second sound sources using an audio gain based on a distance function for the respective second sound source based on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

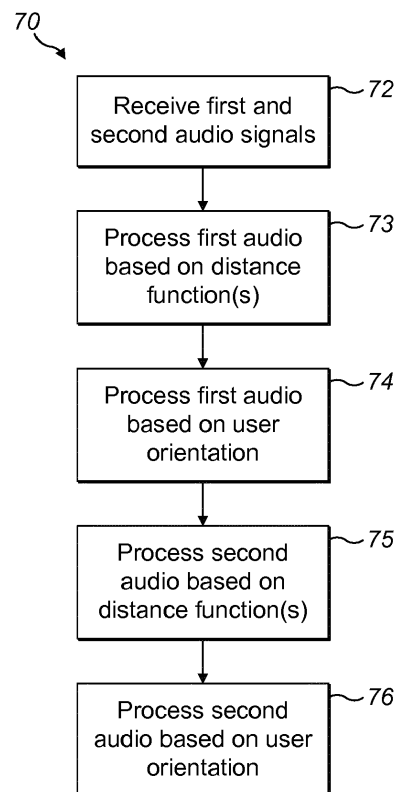


FIG. 7

Description

Field

[0001] This specification relates to processing audio signals and, more specifically, to processing audio signals in a virtual audio scene.

Background

[0002] Combining audio signals from multiple sound sources of a virtual audio scene is known. However, there remains a need for alternative arrangements for combining audio signals in a virtual sound source in which a user can move relative to the sound sources.

Summary

[0003] In a first aspect, this specification describes an apparatus comprising: means for receiving audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene; means for processing audio data of said first sound sources, comprising means for modifying audio data of said first sound sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and means for processing audio data of said second sound sources, comprising means for modifying audio data of said second sound sources by an audio gain that is based on a distance function for the respective second sound source, wherein the distance function for each second sound source is based, at least in part, on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

[0004] In some embodiments, the distance function for each first sound source is based on the distance from the virtual location of the user within the virtual audio scene to the location of the respective first sound source within the virtual audio scene.

[0005] Some embodiments comprise means for selecting said selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene by determining the closest of said initial and said one or more virtual positions to said virtual location of the user. Alternatively, or in addition, some embodiments comprise means for selecting said selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene by determining which of said initial and said one or more virtual positions is in a same sector

of said virtual audio scene as said virtual location of said user.

[0006] The means for processing audio data of said first sound sources may further comprise processing audio data of one or more of the first sound sources based on an orientation of the user relative to the position of the respective audio sound source within the virtual audio scene.

[0007] The means for processing audio data of said second sound sources may further comprise processing audio data of one or more of the second sound sources based on an orientation of the user relative to the initial position of the respective second sound source within the virtual audio scene.

[0008] The distance function for the respective first sound source and/or the distance function for the respective second sound source may be definable parameter(s) (e.g. user-definable parameters).

[0009] The distance function for the respective second sound source may be selected from a plurality of distance functions based, at least in part, on the selected one of the initial and one or more virtual positions of the respective second sound source.

[0010] The initial and/or one or more of the virtual positions of said second sound sources may be provided as metadata of said audio data.

[0011] Some embodiments further comprise means for determining the virtual location and/or an orientation of the user within the virtual audio scene.

[0012] An input may be provided for receiving the virtual location and/or an orientation of the user within the virtual audio scene.

[0013] Some embodiments further comprise means for generating an audio output, including means for combining the processed audio data of said first and second sound sources.

[0014] The said means may comprise: at least one processor; and at least one memory including computer program code, the at least one memory and the computer program code configured, with the at least one processor, to cause the performance of the apparatus.

[0015] In a second aspect, this specification describes a method comprising: receiving audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene; processing audio data of said first sound sources, comprising modifying audio data of said first sound sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and processing audio data of said second sound sources, comprising modifying audio data of one or more of said second sound sources by an audio gain that is based, at least in part, on a distance function for

the respective second sound source, wherein the distance function for each second sound source is based on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

[0016] The distance function for each first sound source maybe based on the distance from the virtual location of the user within the virtual audio scene to the location of the respective first sound source within the virtual audio scene.

[0017] The method may comprise selecting said selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene by determining the closest of said initial and said one or more virtual positions to said virtual location of the user. Alternatively, or in addition, the method may comprise selecting said selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene by determining which of said initial and said one or more virtual positions is in a same sector of said virtual audio scene as said virtual location of said user.

[0018] Processing audio data of said first sound sources may further comprise processing audio data of one or more of the first sound sources based on an orientation of the user relative to the position of the respective audio sound source within the virtual audio scene.

[0019] Processing audio data of said second sound sources may further comprise processing audio data of one or more of the second sound sources based on an orientation of the user relative to the initial position of the respective second sound source within the virtual audio scene.

[0020] The distance function for the respective first sound source and/or the distance function for the respective second sound source may be definable parameter(s) (e.g. user-definable parameters).

[0021] The distance function for the respective second sound source maybe selected from a plurality of distance functions based, at least in part, on the selected one of the initial and one or more virtual positions of the respective second sound source.

[0022] The initial and/or one or more of the virtual positions of said second sound sources maybe provided as metadata of said audio data.

[0023] Some embodiments further comprise determining the virtual location and/or an orientation of the user within the virtual audio scene.

[0024] Some embodiments further comprise generating an audio output, including means for combining the processed audio data of said first and second sound sources.

[0025] In a third aspect, this specification describes any apparatus configured to perform any method as described with reference to the second aspect.

[0026] In a fourth aspect, this specification describes

computer-readable instructions which, when executed by computing apparatus, cause the computing apparatus to perform any method as described with reference to the second aspect.

[0027] In a fifth aspect, this specification describes a computer program comprising instructions for causing an apparatus to perform at least the following: receive audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene and one or more virtual positions within the virtual audio scene; process audio data of said first sound sources, comprising modifying audio data of said first sound sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and process audio data of said second sound sources, comprising modifying audio data of one or more of said second sound sources by an audio gain that is based, at least in part, on a distance function for the respective second sound source, wherein the distance function for each second sound source is based on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

[0028] In a sixth aspect, this specification describes a computer-readable medium (such as a non-transitory computer readable medium) comprising program instructions stored thereon for performing at least the following: receiving audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene and one or more virtual positions within the virtual audio scene; processing audio data of said first sound sources, comprising modifying audio data of said first sound sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and processing audio data of said second sound sources, comprising modifying audio data of one or more of said second sound sources by an audio gain that is based, at least in part, on a distance function for the respective second sound source, wherein the distance function for each second sound source is based on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

[0029] In a seventh aspect, this specification describes an apparatus comprising: at least one processor; and at least one memory including computer program code

which, when executed by the at least one processor, causes the apparatus to: receive audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene; process audio data of said first sound sources, comprising modifying audio data of said first sound sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and process audio data of said second sound sources, comprising modifying audio data of one or more of said second sound sources by an audio gain that is based, at least in part, on a distance function for the respective second sound source, wherein the distance function for each second sound source is based on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

[0030] In an eighth aspect, this specification describes an apparatus comprising: a first input for receiving audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene and one or more virtual positions within the virtual audio scene; a first processor for processing audio data of said first sound sources, comprising means for modifying audio data of said first sound sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and a second processor for processing audio data of said second sound sources, comprising means for modifying audio data of said second sound sources by an audio gain that is based on a distance function for the respective second sound source, wherein the distance function for each second sound source is based, at least in part, on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene. The first and second processor may be implemented using the same processor.

Brief description of the drawings

[0031] So that the invention may be fully understood, embodiments thereof will now be described with reference to the accompanying drawings, in which:

Fig. 1 is a block diagram of a virtual reality display system in which example embodiments may be im-

plemented;

Figs. 2 to 4 show virtual environments demonstrating example uses of the system of Fig. 1;

Fig. 5 shows a virtual environment demonstrating an aspect of an example embodiment;

Fig. 6 shows a virtual environment demonstrating an aspect of an example embodiment;

Fig. 7 is a flow chart showing an algorithm in accordance with an example embodiment;

Fig. 8 shows a virtual environment demonstrating an aspect of an example embodiment;

Fig. 9 shows a virtual environment demonstrating an aspect of an example embodiment;

Fig. 10 is a flow chart showing an algorithm in accordance with an example embodiment;

Fig. 11 is a flow chart showing an algorithm in accordance with an example embodiment;

Fig. 12 shows a virtual environment demonstrating an aspect of an example embodiment;

Fig. 13 is a plot showing distance functions in accordance with example embodiments;

Figs. 14 to 17 are block diagrams of systems in accordance with example embodiments; and

FIGS. 18A and 18B show tangible media, respectively a removable memory unit and a compact disc (CD) storing computer-readable code which when run by a computer perform operations according to example embodiments.

Detailed description

[0032] In the description and drawings, like reference numerals refer to like elements throughout.

[0033] Virtual reality (VR) is a rapidly developing area of technology in which one or both of video and audio content is provided to a user device. The user device may be provided with a live or stored feed from a content source, the feed representing a virtual reality space or world for immersive output through the user device. In VR systems including audio (with or without visual signals), the audio which may be spatial audio representing captured or composed audio from multiple audio objects. A virtual space or virtual world is any computer-generated version of a space, for example a captured real world space, in which a user can be immersed through a user device such as a virtual reality headset. The virtual reality headset may be configured to provide one or more of virtual reality video and spatial audio content to the user, e.g. through the use of a pair of video screens and/or headphones.

[0034] Position and/or movement of a user device within a virtual environment can enhance an immersive experience. Some virtual reality user devices provide a so-called three degrees of freedom (3DoF) system, in which head movement in the yaw, pitch and roll axes are measured and determine what the user sees and/or hears. This facilitates the scene remaining largely static in a single location as the user rotates their head. A next stage

may be referred to as 3DoF+ which may facilitate limited translational movement in Euclidean space in the range of, e.g. tens of centimetres, around a location. A yet further stage is a six degrees-of-freedom (6DoF) system, where the user is able to freely move in the Euclidean space and rotate their head in the yaw, pitch and roll axes.

[0035] Volumetric virtual reality content comprises data representing spaces and/or objects in three-dimensions from all angles, enabling the user to move fully around the spaces and/or objects to view and/or hear them from any angle.

[0036] For the avoidance of doubt, references to virtual reality (VR) are also intended to cover related technologies such as mixed reality (MR) and augmented reality (AR) that refers to a real-world view that is augmented by computer-generated sensory input.

[0037] Fig. 1 is a schematic illustration of a virtual reality display system 10 which represents example user-end equipment. The virtual reality display system 10 includes a user device in the form of a virtual reality headset 14, for displaying visual data and presenting audio data for a virtual reality space, and a virtual reality media player 12 for rendering visual and audio data on the virtual reality headset 14. In some example embodiments, a separate user control (not shown) may be associated with the virtual reality display system, e.g. a hand-held controller.

[0038] In the context of this specification, a virtual space, world or environment is a computer-generated version of a space, for example a captured real world space, in which a user can be immersed. In some example embodiments, the virtual space may be entirely computer-generated. The virtual reality headset 14 may be of any suitable type. The virtual reality headset 14 may be configured to provide virtual reality video and/or audio content data to a user. As such, the user may be immersed in virtual space.

[0039] In the example virtual reality display system 10, the virtual reality headset 14 receives the virtual reality content data from a virtual reality media player 12. The virtual reality media player 12 may be part of a separate device that is connected to the virtual reality headset 14 by a wired or wireless connection. For example, the virtual reality media player 12 may include a games console, or a PC (Personal Computer) configured to communicate visual data to the virtual reality headset 14.

[0040] Alternatively, the virtual reality media player 12 may form part of the virtual reality headset 14.

[0041] The virtual reality media player 12 may comprise a mobile phone, smartphone or tablet computer configured to play content through its display. For example, the virtual reality media player 12 may be a touch-screen device having a large display over a major surface of the device, through which video content can be displayed. The virtual reality media player 12 may be inserted into a holder of a virtual reality headset 14. With such virtual reality headsets 14, a smart phone or tablet computer may display visual data which is provided to a user's eyes via respective lenses in the virtual reality headset

14. The virtual reality audio may be presented, e.g., by loudspeakers that are integrated into the virtual reality headset 14 or headphones that are connected to it. The virtual reality display system 10 may also include hardware configured to convert the device to operate as part of virtual reality display system 10. Alternatively, the virtual reality media player 12 may be integrated into the virtual reality headset 14. The virtual reality media player 12 may be implemented in software. In some example embodiments, a device comprising virtual reality media player software is referred to as the virtual reality media player 12.

[0042] The virtual reality display system 10 may include means for determining the spatial position of the user and/or orientation of the user's head. This may be by means of determining the spatial position and/or orientation of the virtual reality headset 14. Over successive time frames, a measure of movement may therefore be calculated and stored. Such means may comprise part of the virtual reality media player 12. Alternatively, the means may comprise part of the virtual reality headset 14. For example, the virtual reality headset 14 may incorporate motion tracking sensors which may include one or more of gyroscopes, accelerometers and structured light systems. These sensors generate position data from which a current visual field-of-view (FOV) is determined and updated as the user, and so the virtual reality headset 14, changes position and/or orientation. The virtual reality headset 14 may comprise two digital screens for displaying stereoscopic video images of the virtual world in front of respective eyes of the user, and also two headphones, earphone or speakers for delivering audio. The example embodiments herein are not limited to a particular type of virtual reality headset 14.

[0043] In some example embodiments, the virtual reality display system 10 may determine the spatial position and/or orientation of the user's head using the above-mentioned six degrees-of-freedom method. These may include measurements of pitch, roll and yaw and also translational movement in Euclidean space along side-to-side, front-to-back and up-and-down axes.

[0044] The virtual reality display system 10 may be configured to display virtual reality content data to the virtual reality headset 14 based on spatial position and/or the orientation of the virtual reality headset. A detected change in spatial position and/or orientation, i.e. a form of movement, may result in a corresponding change in the visual and/or audio data to reflect a position or orientation transformation of the user with reference to the space into which the visual data is projected. This allows virtual reality content data to be consumed with the user experiencing a 3D virtual reality environment.

[0045] In the context of volumetric virtual reality spaces or worlds, a user's position may be detected relative to content provided within the volumetric virtual reality content, e.g. so that the user can move freely within a given virtual reality space or world, around individual objects or groups of objects, and can view and/or listen to the

objects from different angles depending on the rotation of their head.

[0046] Audio data may be provided to headphones provided as part of the virtual reality headset 14. The audio data may represent spatial audio source content. Spatial audio may refer to directional rendering of audio in the virtual reality space or world such that a detected change in the user's spatial position or in the orientation of their head may result in a corresponding change in the spatial audio rendering to reflect a transformation with reference to the space in which the spatial audio data is rendered.

[0047] The angular extent of the environment observable or hearable through the virtual reality headset 14 is called the visual or audible field of view (FOV). The actual FOV observed by a user in terms of visuals depends on the inter-pupillary distance and on the distance between the lenses of the virtual reality headset 14 and the user's eyes, but the FOV can be considered to be approximately the same for all users of a given display device when the virtual reality headset is being worn by the user. The audible FOV can be omnidirectional and independent of the visual FOV in terms of the angular direction (yaw, pitch, roll), or it can relate to the visual FOV.

[0048] Fig. 2 shows a virtual environment, indicated generally by the reference numeral 20, that may be implemented using the virtual reality display system 10. The virtual environment 20 shows a user 22 and first to fourth sound sources 24 to 27. The user 22 may be wearing the virtual reality headset 14 described above.

[0049] The virtual environment 20 is a virtual audio scene and the user 22 has a position and an orientation within the scene. The audio presented to the user 22 (e.g. using the virtual reality headset 14) is dependent on the position and orientation of the user 22, such that a 6DoF audio scene is provided.

[0050] Fig. 3 shows a virtual environment, indicated generally by the reference numeral 30, in which the orientation of the user 22 has changed relative to the orientation shown in Fig. 2. The user position is unchanged. By changing the presentation to the user 22 of the audio from the sound sources 22 to 27 on the basis of the orientation of the user, an immersive experience can be enhanced.

[0051] Fig. 4 shows a virtual environment, indicating generally by the reference numeral 40, in which the position of the user 22 has changed relative to the position shown in Fig. 2 (indicated by the translation arrow 42), but the orientation of the user is unchanged relative to the orientation shown in Fig. 2. By changing the presentation to the user 22 of the audio from the sound sources 22 to 27 on the basis of the position of the user (e.g. by making audio sources louder and less reverberant and the user approaches the audio source in the virtual environment), an immersive experience can be enhanced.

[0052] Clearly, both the position and the orientation of the user 22 could be changed at the same time. It is also noted that a virtual audio environment can include both diegetic and non-diegetic audio elements, i.e., audio el-

ements that are presented to the user from a static direction/position of the virtual environment during change in user orientation and audio elements that are presented from an unchanged direction/position of the virtual environment regardless of any user rotation.

[0053] Fig. 5 shows a virtual environment, indicated generally by the reference numeral 50, demonstrating aspect of an example embodiment. The virtual environment 50 includes a first user position 52a and a plurality of sound sources. The sound sources are provided from different sound source zones. In the example virtual environment 50, a first zone 54, a second zone 55 and a third zone 56 are shown. Two sound sources are shown in each sound source zone (such that the first zone 54 includes sound sources 54a and 54b, the second zone 55 includes sound sources 55a and 55b, and the third zone 56 includes sound sources 56a and 56b). Of course, the number of zones and the number of sound sources within each zone can vary in different example embodiments and it is not essential that each zone includes the same number of sound sources. The virtual environment 50 also includes an anchor sound zone 57 comprising a first anchor sound source 57a and a second anchor sound source 57b. Again, the number and location of anchor sound zones within the virtual environment 50 and the number and location of sound sources within an anchor sound zone can be varied. Further details regarding anchor sound sources are provided below. The sound sources in the first, second and third zones 54 to 56 are sometimes collectively referred to below as "first sound sources" and sound sources in the anchor source zone are sometimes collectively referred to below as "second sound sources".

[0054] In one example embodiment, the virtual environment 50 represents an audio scene with different subsets of instruments being provided in the different zones 54, 55 and 56. For example, the first zone 54 may include guitar sounds, the second zone 55 may include keyboard and backing vocal sounds, and the third zone 56 may include lead vocal sounds. As the user moves around the virtual environment 50, the relative volumes of the different sound sources changes, thereby providing an immersive effect. For example, with the user at the first user position 52a, each of the sound sources may have a similar volume, but when the user is at a second user position 52b, the audio of the third zone 56 (i.e. lead vocals) maybe louder such that user appears to move towards the lead singer within the scene. (Other effects, such as reverberation, may also be adjusted.) The audio provided to the user may also be adjusted based on the orientation of the user within the virtual environment.

[0055] The sounds of anchor sound zone 57 may define sounds (e.g. important sounds) that are intended to be heard throughout the virtual environment 50. The sounds of anchor sound zone 57 may be provided close to the centre of the virtual environment 50, although this is not essential to all example embodiments. In the example audio output described above, the first anchor

sound source 57a may provide drum sounds and the second anchor sound source 57b may provide bass sounds, such that the anchor sound zone 57 provides drum and bass audio for the virtual environment 50.

[0056] Thus, as the user moves from the first user position 52a to the second user position 52b, the sounds from the third zone 56 are accentuated, but the sounds from the anchor sound zone 57 remain strong.

[0057] Fig. 6 shows a virtual environment, indicated generally by the reference numeral 60, demonstrating aspect of an example embodiment. The virtual environment 60 includes the first zone 54, second zone 55, third zone 56 and anchor sound zone 57 described above with reference to the virtual environment 50. As with the virtual environment 50, two sound sources are shown in each sound source zone (such that the first zone 54 includes sound sources 54a and 54b, the second zone 55 includes sound sources 55a and 55b, the third zone 56 includes sound sources 56a and 56b, and the anchor sound zone 57 includes anchor sound sources 57a and 57b).

[0058] The virtual environment 60 is consumed by a user. The user has a first position 62a close to the anchor zone (and similar to the first user position 52a described above). The user moves (as indicated by arrow 64) to a second position 62b that is within the third zone 56. Thus, as the user moves from the first user position 62a to the second user position 62b, the audio of the third zone 56 (e.g. lead vocals) may be increased and audio from other zone (including the anchor zone) may be reduced. Indeed, in some embodiments, the sounds from the anchor sound zone 57 may be too quiet when the user is at the second user position 62b, unless the anchor sounds are made so loud that they are too loud with the user in other locations (such as the first position 62a).

[0059] In some embodiments, suitable distance attenuation (or for example, no distance attenuation) differing from distance attenuation used for the first sound sources may be used for the second/ anchor sound sources. In addition to gain adjustment, the rendering of the second sound sources may be altered using, e.g., spatial extent processing depending on the user distance. For example, even if the gain of the second sound sources is not reduced greatly due to user distance, increasing the spatial extent of the sound sources can still convey to the user a feeling of immersion and the effect of their action (movement in the virtual environment).

[0060] Fig. 7 is a flow chart showing an algorithm, indicated generally by the reference numeral 70, in accordance with an example embodiment. The algorithm 70 is described below with reference to the virtual environment 80 shown in Fig. 8.

[0061] Fig. 8 shows a virtual environment, indicated generally by the reference numeral 80, demonstrating an aspect of an example embodiment. The virtual environment 80 includes the first, second and third zones 54 to 56 described above, and an anchor zone 85, wherein two sound sources are shown in each sound source zone (such that the first zone 54 includes sound sources 54a

and 54b, the second zone 55 includes sound sources 55a and 55b, the third zone 56 includes sound sources 56a and 56b, and the anchor zone 85 includes anchor sound sources 85a and 85b). (Again, the number of sound sources per zone could be varied. Alternatively, or in addition, the number of zones could be varied.) The virtual environment 80 is consumed by a user. The user has a first position 82a close to the anchor zone. The user moves (as indicated by arrow 84) to a second position 82b that is close to the third zone 56.

[0062] The algorithm 70 starts at operation 72 where audio data is received relating to a plurality of audio sound sources of a virtual audio scene (such as the virtual environment 80). The plurality of audio sound sources comprises one or more first sound sources (such as one or more of the sound sources 54a, 54b, 55a, 55b, 56a and 56b of the first to third zones) and one or more second sound sources (e.g. one or more of the anchor sound sources 85a and 85b of the anchor zone 85). Each first sound source has a position within the virtual audio scene. Moreover, each second sound source has an initial position (such as the positions 85a and/or 85b) within the virtual audio scene and one or more virtual positions (such as the positions 86, 87 and/or 88 shown in Fig. 8) within the virtual audio scene.

[0063] At operation 73, the first audio sound sources are processed, including modifying audio data of one or more of said first sound sources by an audio gain (e.g. an attenuation) that is based on a distance function for the respective first sound source. For example, a degree of attenuation of each of the first audio sound sources may increase as the user moves away from the respective sound source. As discussed further below, the attenuation may be based on a $1/\text{distance}$ function, although many alternative arrangements (including user-definable arrangements) are possible, for example to allow for artistic intent by the creator of the virtual environment to be implemented.

[0064] At operation 74, the first audio sound sources are processed, including processing audio data of at least some of the first sound sources based on an orientation of the user relative to the respective audio sound source within the virtual audio scene.

[0065] At operation 75, the second sound sources are processed, including modifying audio data of one or more of said second sound sources by an audio gain (e.g. an attenuation) that is based on a distance function for the respective second sound source. For example, a degree of attenuation of each of the second audio sound sources may increase as the user moves away from the respective sound source. As discussed further below, the attenuation may be based on a $1/\text{distance}$ function, although many alternative arrangements (including user-definable arrangements) are possible.

[0066] As described further below, the distance function for each first sound source (used in the operation 73 described above) may be based on the distance from the virtual location of the user within the virtual audio scene

to the location of the respective first sound source within the virtual audio scene. The distance function for each second sound source (referred to in the operation 75 described above) may be based on a distance from a virtual location of the user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

[0067] At operation 76, the second sound sources are processed, including processing audio data of said second sound sources based on an orientation of the user relative to a position of the respective second sound source within the virtual audio scene. In one embodiment, the operation 76 processes the second sound source based on the orientation of the user relative to the initial position of that second sound source, regardless of whether the initial or the virtual anchor sound source is used in the operation 75. Thus, directionality (operation 76) may be dependent on the position of the relevant initial second/ anchor sound source and the attenuation (operation 75) may be based on the position of a selected one of the second/ anchor sound source and a virtual second/ anchor sound source.

[0068] Many variants of the algorithms 70 are possible. For example, the operations 73 to 76 may be performed in a different order and/or some of the operations may be combined. Thus, for example, the operations 73 and 74 may be merged into a single operation and/or the operations 75 and 76 may be merged into a single operation. In some embodiments, some operations of the algorithm 70 may be omitted. For example, the operations 74 and 76 may be omitted in the event that orientation processing is not provided. For example, a mono or stereo mix of a multi-track audio content dependent on user location or distance from at least one reference point can be thus be implemented without explicit head orientation tracking or processing.

[0069] One such application is described as follows, where a distance tracking that corresponds to a route is provided, rather than a volumetric virtual environment. In this example application, a user is provided a multi-track audio soundtrack for inspiring background music during a jogging exercise. The audio is presented to the user using a mobile device application and a traditional Bluetooth headphone device that provides no headtracking capability. A GPS tracking of the user along the jogging route is utilized to control the balance of the mixing of the multi-track audio content. In the beginning of the jog, for example, a relaxed acoustic instrumentation (first sound sources) over a drum and base anchor content (second sound sources) is provided. Towards the middle of the jogging route, the acoustic instrumentation (first sound sources) fades out and it is faded in a more aggressive electric instrumentation (first sound sources) while maintaining the audibility of the second sound sources. As the user approaches the end of the jogging route, vocal tracks (first sound sources) pushing the user towards the finish are added.

[0070] In the example virtual environment 80 described above, the one or more second sound sources are mapped onto a single virtual position within each zone of the virtual audio scene (the positions 86, 87 and 88). This is not essential to all embodiments. For example, at least some of the one or more second sound sources may be mapped to a subset of said zones. Moreover, multiple virtual positions may be provided within each audio zone, with different second sound sources being mapped to different virtual positions within each audio zone.

[0071] By way of example, Fig. 9 shows a virtual environment, indicated generally by the reference numeral 90, demonstrating an aspect of an example embodiment. The virtual environment 90 includes the first zone 54, second zone 55 and third zone 56 described above, and an anchor zone 93, wherein two sound sources are shown in each sound source zone (such that the first zone 54 includes sound sources 54a and 54b, the second zone 55 includes sound sources 55a and 55b, the third zone 56 includes sound sources 56a and 56b, and the anchor zone 93 includes anchor sound sources 93a and 93b). The virtual environment 90 has a user 92. Two virtual anchor sound sources are provided in each of said zones (one for each of the anchor sound sources 93a and 93b). Thus, the first zone 54 includes first and second virtual anchor sound sources 94a and 94b, the second zone 55 includes first and second virtual anchor sound sources 95a and 95b, and the third zone 56 includes first and second virtual anchor sound sources 96a and 96b.

[0072] Fig. 10 is a flow chart showing an algorithm, indicated generally by the reference numeral 100, in accordance with an example embodiment. The algorithm 100 shows an example implementation of the operation 75 described above in which the second audio is processed based on a selected distance function.

[0073] Consider the virtual environment 80 described above in which the user is in the first position 82a. As discussed above, the audio data of the anchor sound sources (the anchor sound sources 85a and 85b) are modified based on a distance function for the respective anchor sound source, wherein the distance function is based on a distance from the first position 82a of the user within the virtual environment 80 to the location of a selected one of the initial and one or more virtual positions of the anchor sound sources. Consider, by way of example, the first anchor sound source 85a.

[0074] The algorithm 100 starts at operation 102, where a distance from the user (the first position 82a) to the initial position of the second sound source (the position 85a) is determined.

[0075] At operation 104, the distances from the user (the first position 82a) to the virtual positions of the second sound source (such as the positions 86, 87 and 88) are determined.

[0076] At operation 106, the minimum of the distances determined in the operations 102 and 104 above is selected and, at operation 108, the distance function for the

operation 75 is set based on the distance selected in operation 106.

[0077] Thus, in operation 106, said selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene is selected by determining the closest of said initial and said one or more virtual positions to said user position (in this case, the initial position 85a is closest to the first position 82a). The distance function is therefore based on the distance from the first position 82a to the initial position of the first anchor sound source 85a.

[0078] Now, consider the virtual environment 80 described above in which the user 82 has moved to the second position 82b. As discussed above, the audio data of the anchor sound sources (the anchor sound sources 85a and 85b) are modified based on a distance function for the respective anchor sound source, wherein the distance function is based on a distance from the second position 82b of the user within the virtual environment 80 to the location of a selected one of the initial and one or more virtual positions of the anchor sound sources. Consider, again, the first anchor sound source 85a.

[0079] The algorithm 100 starts at operation 102, where a distance from the user (the second position 82b) to the initial position of the second sound source (the position 85a) is determined.

[0080] At operation 104, the distances from the user (the second position 82b) to the virtual positions of the second sound source (such as the positions 86, 87 and 88) are determined.

[0081] At operation 106, the minimum of the distances determined in the operations 102 and 104 above is selected and, at operation 108, the distance function for the operation 75 is set based on the distance selected in operation 106.

[0082] Thus, in operation 106, said selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene is selected by determining the closest of said initial and said one or more virtual positions to said user position (in this case, the third virtual position 88 is closest to the second position 82b). The distance function in this instance is therefore based on the distance from the second position 82b to the third virtual position 88 of the first anchor sound source 85a.

[0083] The algorithm 100 is not the only mechanism by which the distance function may be set. An alternative arrangement is described below, although yet further alternatives are possible.

[0084] Fig. 11 is a flow chart showing an algorithm, indicated generally by the reference numeral 110, in accordance with an example embodiment. The algorithm 110 is described with reference to the virtual environment 120 shown in Fig. 12.

[0085] Fig. 12 shows a virtual environment, indicated generally by the reference numeral 120, demonstrating an aspect of an example embodiment. The virtual environment 120 includes the first, second and third zones

54 to 56 described above, and an anchor zone 125, wherein the first zone 54 includes first sound sources 54a and 54b and a first virtual second (or anchor) sound source 126, the second zone 55 includes first sound sources 55a and 55b and a second virtual second (or anchor) sound source 127, the third zone 56 includes first sound sources 56a and 56b and a third virtual second (or anchor) sound source 128, and the anchor zone 125 includes second (or anchor) sound sources 125a and 125b. (Again, the number of sound sources per zone could be varied. Alternatively, or in addition, the number of zones could be varied.) The virtual environment 120 is consumed by a user. The user has a first position 122a close to the anchor zone. The user moves (as indicated by arrow 124) to a second position 122b that is close to the third zone 56.

[0086] The virtual environment 120 is divided into sectors. A first sector 129a includes the first zone 54, a second sector 129b includes the second zone 55, a third sector 129c includes the third zone 56 and a fourth sector 129d includes the anchor zone 125.

[0087] Assume that the user is initially in the position 122a.

[0088] The algorithm 110 starts at operation 112, where the user position (the position 122a) is determined. At operation 114, the sector in which the user position is located is determined (the sector 129d). At operation 116, the distance function is set based on which of the initial and virtual positions of the second sound source fall within the same sector. In the virtual environment 120, it can be seen that the initial position of the second sound source in the same sector (the sector 129d) as the first position 122a and so the distance function is determined based on the initial position.

[0089] Now, assume that the user moves to the position 122b.

[0090] The algorithm 110 starts at operation 112, where the user position (the position 122b) is determined. At operation 114, the sector in which the user position is located is determined (the sector 129d). At operation 116, the distance function is set based on which of the initial and virtual positions of the second sound source fall within the same sector. In the virtual environment 120, it can be seen that the initial position of the second sound source in the same sector (the sector 129d) as the second position 122b and so the distance function is determined based on the initial position.

[0091] Thus, the algorithm 110 (based on the virtual environment 120) comes to a different conclusion to the algorithm 100 (based on the virtual environment 80) when the user is in the second position 82b (in Fig. 8) and 122b (in Fig. 12).

[0092] As suggested above, the processing of the audio signals in the operations 73 and 75 may be based on a $1/\text{distance}$ function.

[0093] Fig. 13 is a plot, indicated generally by the reference numeral 130, showing distance functions in accordance with example embodiments. The plot 130 in-

cludes distance from a sound source plotted on the x-axis and gain plotted on the y-axis. An initial position 131 of a second (or anchor) sound source is plotted, together with a virtual position 132 of the second (or anchor) sound source. By way of example, the position 131 may indicate the position of the second sound source 85a described above and the position 132 may indicate the virtual position 88 of the second sound source.

[0094] A first curve 133 plots gain as a function of the distance of a user from the position 131 of the second sound source. When the user is close to the position 131, the gain is high (e.g. 4 in the example of plot 130), but when the user is far from the position 131, the gain is low (below 0.5 in the example of plot 130).

[0095] A second curve 134 plots gain as a function of the distance of a user from the virtual sound source 132. When the user is close to the position 132, the gain is high (e.g. 4 in the example of plot 130), but when the user is far from the position 132, the gain is low (below 0.5 in the example of plot 130).

[0096] In operation 73 of the algorithm 70 described above, a first audio signal is processed based on a distance function. The distance function may, for example, have the form of the first curve 133, such that gain is reduced with distance between the user and the first audio signal.

[0097] In operation 75 of the algorithm 70 described above, a second audio signal is processed based on a selected distance function. The distance function may, for example, be selected from the first and second curves 133 and 134.

[0098] In the plot 130, the curves 133 and 134 have the same maximum gain (4 in the example plot 130). This is not essential. For example, the second sound source at an initial position may have a higher gain than the second sound source at a virtual position. For example, Fig. 13 shows a virtual position 135 of the second sound source. A third curve 136 plots gain as a function of the distance of a user from the position 135. When the user is close to the position 135, the gain is high (e.g. 2.5 in the example of plot 130), but not as high as when the user is close to the position 132.

[0099] The distance functions described above are 1/distance function; this is not essential to all embodiments. Other distance functions may be provided. (Thus, for example, the slopes of some or all of the curves shown in Fig. 13 could be different.) Moreover, one or more of said distance functions for the respective first sound source and/or the distance function for the respective second sound source may be definable parameter(s). Thus, for example, at least some of the curves of the plot 130 may be user-definable (e.g. using a user interface).

[0100] Fig. 14 is a block diagram of a system, indicated generally by the reference numeral 140, in accordance with an example embodiment.

[0101] The system 140 comprises a parameter selection module 141, a delay line 142, a distance gain function module 143, a filtering module 144, a reverberator 145,

a first summing module 146 and a second summing module 147.

[0102] The parameter selection module 141 receives user position and orientation data and sound gain limits based on virtual positions and uses these data to provide control signals to the distance gain function module 143 and the filtering module 144.

[0103] The delay line 142 receives audio input to the system 140 and generates a plurality of outputs having successively greater time delay. The distance gain function module 143 comprises multiple instances of modules implementing a distance function (as discussed, for example, with reference to Fig. 13). Each instance of the gain function module 143 operates on a differently delayed version of the audio input signal.

[0104] The filtering module 144 may modifies the outputs of the gain function module 143 based on user head orientation. The filtering module 144 may, for example implement a head related transfer function (HRTF).

[0105] The reverberator 145 receives the output of the undelayed instance of the distance gain function module and generates a reverberated version of that signal based on one or more reverberation parameters. The reverberator 145 may, for example, seek to recreate different sound spaces.

[0106] Finally, the summing modules 146 and 147 sum the output of the reverberator and the outputs of the filtering module 144 to provide separate left and right audio outputs to a user.

[0107] The system 140 is provided by way of example only. The skilled person will be aware of many alternative arrangements. For example, at least some of the modules (such as the delay line 142 and/or the reverberator 145 may be omitted and one or more of the modules may be reconfigured).

[0108] Fig. 15 is a block diagram of a system, indicated generally by the reference numeral 150, in accordance with an example embodiment. The system 150 comprises a head mounted display 152 and a rendering module 154. The rendering module comprises one or more of: memory; one or more processors and/or application logic modules; a graphics rendering module; input and output modules, such as a camera, a display, one or more sensors, audio output and audio playback; an orientation and position sensing module; and a radio module. The rendering module 154 may be implemented using a mobile communication device, such as a mobile phone. The rendering module 154 is provided by way of example only; many modifications, such as the provision of different combinations of modules, could be made.

[0109] Fig. 16 is a block diagram of a system, indicated generally by the reference numeral 160, in accordance with an example embodiment. The system 160 comprises a decoder 172, a position and orientation module 174 and a rendering module 176. As indicated in Fig. 16, the decoder receives a bitstream. The decoder 172 converts the received bitstream into audio data and audio metadata. The metadata may, for example, include audio po-

sition information. The position and orientation module 174 provides information relating to a position and an orientation of a user within a virtual environment. The rendering module receives the user position and orientation information, the audio data and the audio metadata and renders the audio accordingly. The rendering module 176 may, for example, be implemented using the system 140 described above.

[0110] By way of example, the position information for initial and virtual positions of an audio sound source (e.g. as output by the decoder 172) may be provided in the following format:

```
<Audio_object_metadata>
<Position>

    <x>1.0</x>

    <y>2.3</y>

    <z>1/2</z>

</Position>
<Virtual_positions>

    <x>2.0</x>

    <y>2.3</y>

    <z>1.2</z>

    <x>8.0</x>

    <y>3.3</y>

    <z>1.2</z>

</Virtual_positions>
```

[0111] For completeness, FIG. 17 is a schematic diagram of components of one or more of the example embodiments described previously, which hereafter are referred to generically as processing systems 300. A processing system 300 may have a processor 302, a memory 304 closely coupled to the processor and comprised of a RAM 314 and ROM 312, and, optionally, user input 310 and a display 318. The processing system 300 may comprise one or more network/apparatus interfaces

308 for connection to a network/apparatus, e.g. a modem which maybe wired or wireless. Interface 308 may also operate as a connection to other apparatus such as device/apparatus which is not network side apparatus. Thus, direct connection between devices/apparatus without network participation is possible.

[0112] The processor 302 is connected to each of the other components in order to control operation thereof.

[0113] The memory 304 may comprise a non-volatile memory, such as a hard disk drive (HDD) or a solid-state drive (SSD). The ROM 312 of the memory 314 stores, amongst other things, an operating system 315 and may store software applications 316. The RAM 314 of the memory 304 is used by the processor 302 for the temporary storage of data. The operating system 315 may contain code which, when executed by the processor implements aspects of the algorithms 70, 100 and 110 described above. Note that in the case of small device/ apparatus the memory can be most suitable for small size usage i.e. not always hard disk drive (HDD) or solid-state drive (SSD) is used.

[0114] The processor 302 may take any suitable form. For instance, it may be a microcontroller, a plurality of microcontrollers, a processor, or a plurality of processors.

[0115] The processing system 300 may be a standalone computer, a server, a console, or a network thereof. The processing system 300 and needed structural parts may be all inside device/ apparatus such as IoT device/ apparatus i.e. embedded to very small size

[0116] In some example embodiments, the processing system 300 may also be associated with external software applications. These may be applications stored on a remote server device/ apparatus and may run partly or exclusively on the remote server device/apparatus. These applications maybe termed cloud-hosted applications. The processing system 300 may be in communication with the remote server device/ apparatus in order to utilize the software application stored there.

[0117] FIGS. 18A and 18B show tangible media, respectively a removable memory unit 365 and a compact disc (CD) 368, storing computer-readable code which when run by a computer may perform methods according to example embodiments described above. The removable memory unit 365 may be a memory stick, e.g. a USB memory stick, having internal memory 366 storing the computer-readable code. The memory 366 may be accessed by a computer system via a connector 367. The CD 368 may be a CD-ROM or a DVD or similar. Other forms of tangible storage media may be used. Tangible media can be any device/ apparatus capable of storing data/information which data/information can be exchanged between devices/apparatus/network.

[0118] Embodiments of the present invention may be implemented in software, hardware, application logic or a combination of software, hardware and application logic. The software, application logic and/or hardware may reside on memory, or any computer media. In an example embodiment, the application logic, software or an instruc-

tion set is maintained on any one of various conventional computer-readable media. In the context of this document, a "memory" or "computer-readable medium" may be any non-transitory media or means that can contain, store, communicate, propagate or transport the instructions for use by or in connection with an instruction execution system, apparatus, or device, such as a computer.

[0119] Reference to, where relevant, "computer-readable storage medium", "computer program product", "tangibly embodied computer program" etc., or a "processor" or "processing circuitry" etc. should be understood to encompass not only computers having differing architectures such as single/multi-processor architectures and sequencers/parallel architectures, but also specialised circuits such as field programmable gate arrays FPGA, application specific circuits ASIC, signal processing devices/ apparatus and other devices/ apparatus. References to computer program, instructions, code etc. should be understood to express software for a programmable processor firmware such as the programmable content of a hardware device/ apparatus as instructions for a processor or configured or configuration settings for a fixed function device/ apparatus, gate array, programmable logic device/ apparatus, etc.

[0120] As used in this application, the term "circuitry" refers to all of the following: (a) hardware-only circuit implementations (such as implementations in only analogue and/or digital circuitry) and (b) to combinations of circuits and software (and/or firmware), such as (as applicable): (i) to a combination of processor(s) or (ii) to portions of processor(s)/ software (including digital signal processor(s)), software, and memory(ies) that work together to cause an apparatus, such as a server, to perform various functions) and (c) to circuits, such as a microprocessor(s) or a portion of a microprocessor(s), that require software or firmware for operation, even if the software or firmware is not physically present.

[0121] If desired, the different functions discussed herein may be performed in a different order and/or concurrently with each other. Furthermore, if desired, one or more of the above-described functions may be optional or may be combined. Similarly, it will also be appreciated that the flow diagrams of Figures 7, 10 and 11 are examples only and that various operations depicted therein maybe omitted, reordered and/or combined.

[0122] It will be appreciated that the above described example embodiments are purely illustrative and are not limiting on the scope of the invention. Other variations and modifications will be apparent to persons skilled in the art upon reading the present specification.

[0123] Moreover, the disclosure of the present application should be understood to include any novel features or any novel combination of features either explicitly or implicitly disclosed herein or any generalization thereof and during the prosecution of the present application or of any application derived therefrom, new claims may be formulated to cover any such features and/or combination of such features.

Claims

1. An apparatus comprising:

5 means for receiving audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene and one or more virtual positions within the virtual audio scene;
10 means for processing audio data of said first sound sources, comprising means for modifying audio data of said first sound sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and
15 means for processing audio data of said second sound sources, comprising means for modifying audio data of said second sound sources by an audio gain that is based on a distance function for the respective second sound source, wherein the distance function for each second sound source is based, at least in part, on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

2. An apparatus as claimed in claim 1, wherein the distance function for each first sound source is based on the distance from the virtual location of the user within the virtual audio scene to the location of the respective first sound source within the virtual audio scene.

3. An apparatus as claimed in claim 1 or claim 2, further comprising means for selecting said selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene by determining the closest of said initial and said one or more virtual positions to said virtual location of the user.

4. An apparatus as claimed in any one of claims 1 to 3, further comprising means for selecting said selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene by determining which of said initial and said one or more virtual positions is in a same sector of said virtual audio scene as said virtual location of said user.

5. An apparatus as claimed in any one of the preceding

claims, wherein the means for processing audio data of said first sound sources further comprises processing audio data of one or more of the first sound sources based on an orientation of the user relative to the position of the respective audio sound source within the virtual audio scene.

6. An apparatus as claimed in any one of the preceding claims, wherein the means for processing audio data of said second sound sources further comprises processing audio data of one or more of the second sound sources based on an orientation of the user relative to the initial position of the respective second sound source within the virtual audio scene.
7. An apparatus as claimed in any one of the preceding claims, wherein the distance function for the respective first sound source and/or the distance function for the respective second sound source is/are definable parameter(s).
8. An apparatus as claimed in any one of the preceding claims, wherein the distance function for the respective second sound source is selected from a plurality of distance functions based, at least in part, on the selected one of the initial and one or more virtual positions of the respective second sound source.
9. An apparatus as claimed in any one of the preceding claims, wherein the initial and/or one or more of the virtual positions of said second sound sources are provided as metadata of said audio data.
10. An apparatus as claimed in any one of the preceding claims, further comprising means for determining the virtual location and/or an orientation of the user within the virtual audio scene.
11. An apparatus as claimed in any one of the preceding claims, further comprising an input for receiving the virtual location and/or an orientation of the user within the virtual audio scene.
12. An apparatus as claimed in any one of the preceding claims, further comprising means for generating an audio output, including means for combining the processed audio data of said first and second sound sources.
13. An apparatus as claimed in any one of the preceding claims, wherein the means comprise:

at least one processor; and
at least one memory including computer program code, the at least one memory and the computer program code configured, with the at least one processor, to cause the performance of the apparatus.

14. A method comprising:

receiving audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene and one or more virtual positions within the virtual audio scene;
processing audio data of said first sound sources, comprising modifying audio data of said first sound sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and
processing audio data of said second sound sources, comprising modifying audio data of one or more of said second sound sources by an audio gain that is based, at least in part, on a distance function for the respective second sound source, wherein the distance function for each second sound source is based on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

15. Computer-readable instructions which, when executed by computing apparatus, cause the computing apparatus to perform a method of:

receiving audio data relating to a plurality of audio sound sources of a virtual audio scene, wherein the plurality of audio sound sources comprises one or more first sound sources and one or more second sound sources, wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene and one or more virtual positions within the virtual audio scene;
processing audio data of said first sound sources, comprising modifying audio data of said first sound sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and
processing audio data of said second sound sources, comprising modifying audio data of one or more of said second sound sources by an audio gain that is based on a distance function for the respective second sound source, wherein the distance function for each second sound source is based, at least in part, on a distance from a virtual location of a user within the virtual audio scene to the location of a selected one of

said initial and one or more virtual positions for the respective second sound source within the virtual audio scene.

5

10

15

20

25

30

35

40

45

50

55

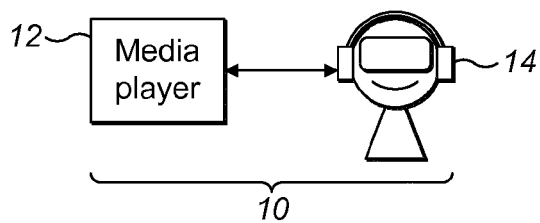


FIG. 1

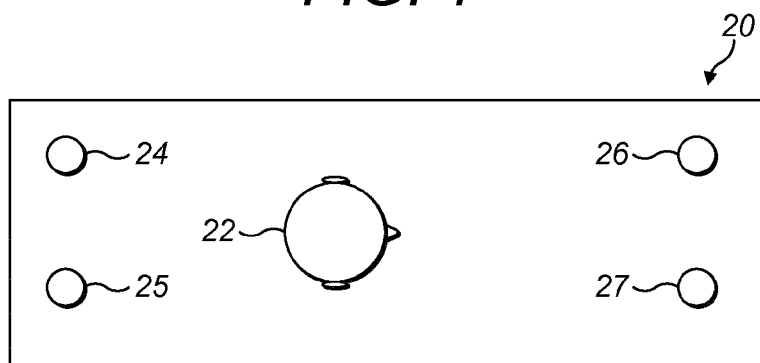


FIG. 2

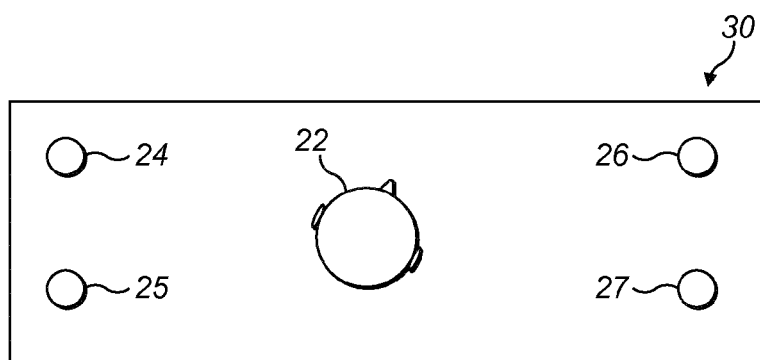


FIG. 3

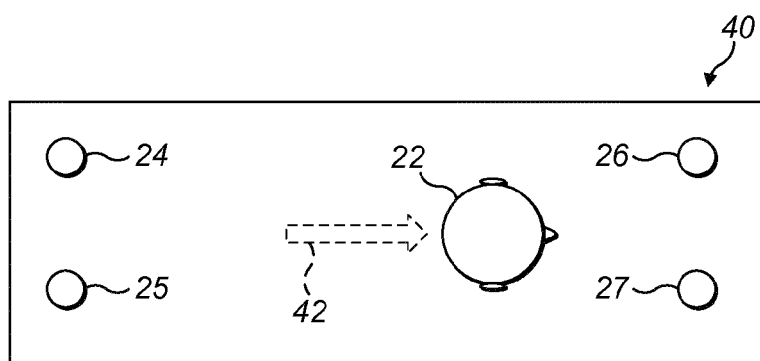


FIG. 4

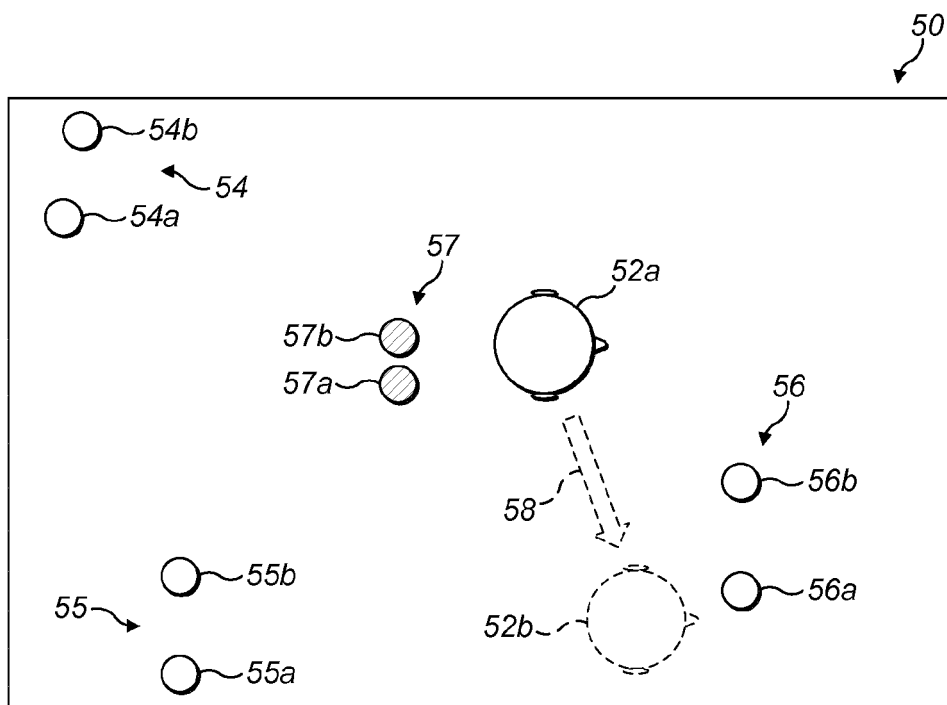


FIG. 5

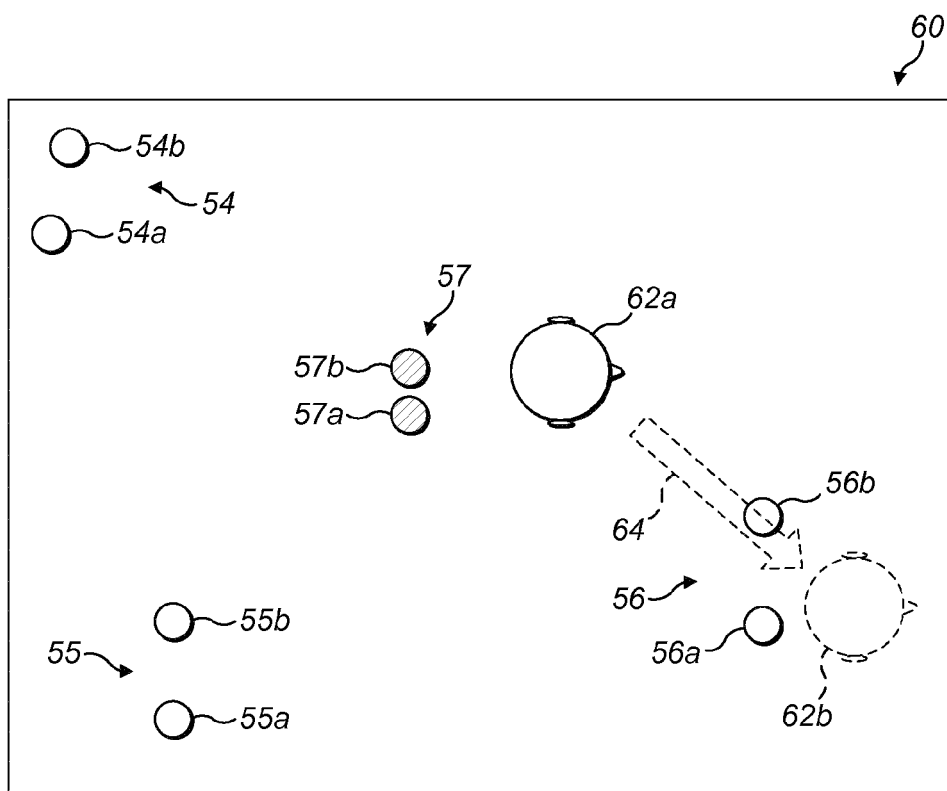


FIG. 6

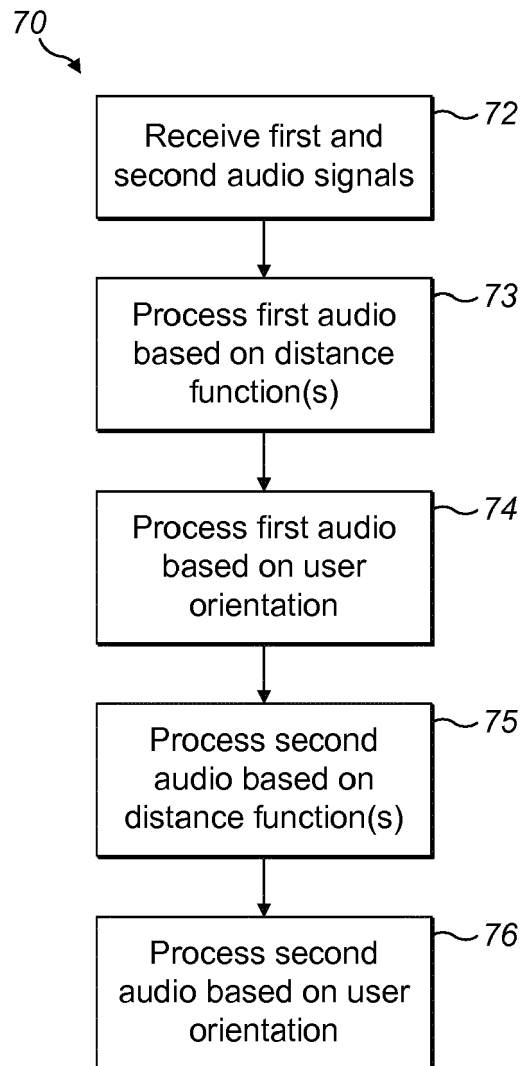


FIG. 7

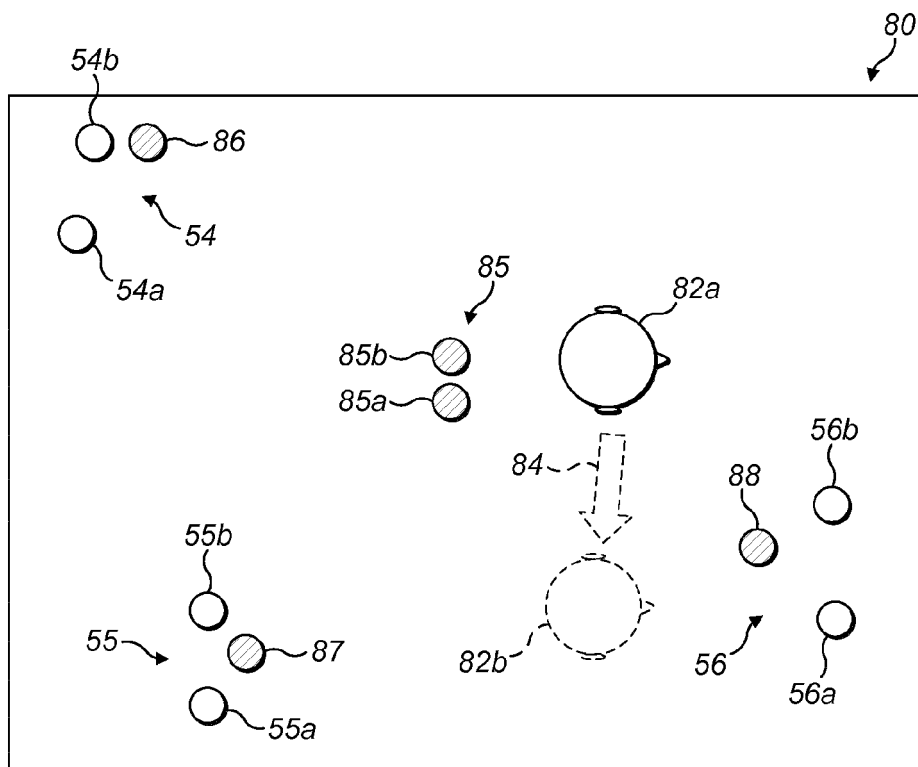


FIG. 8

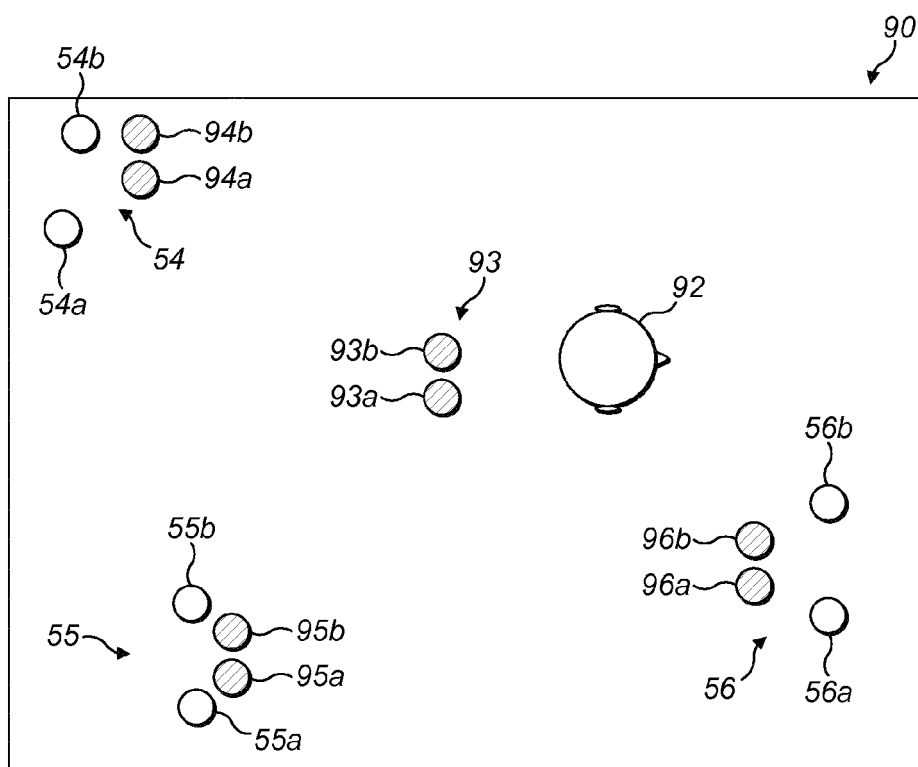


FIG. 9

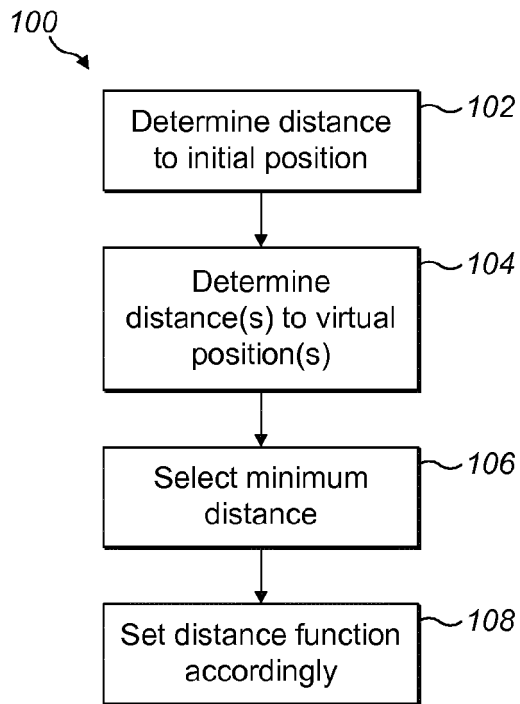


FIG. 10

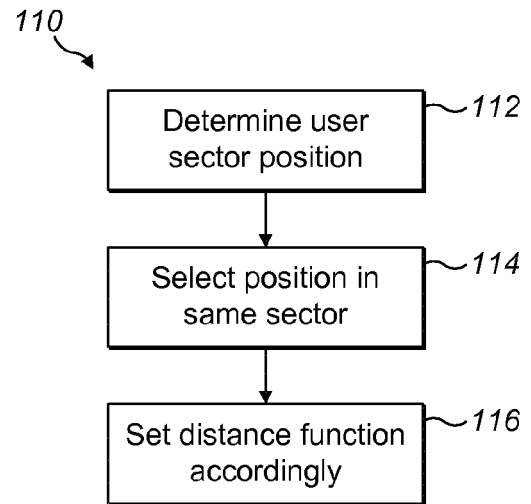


FIG. 11

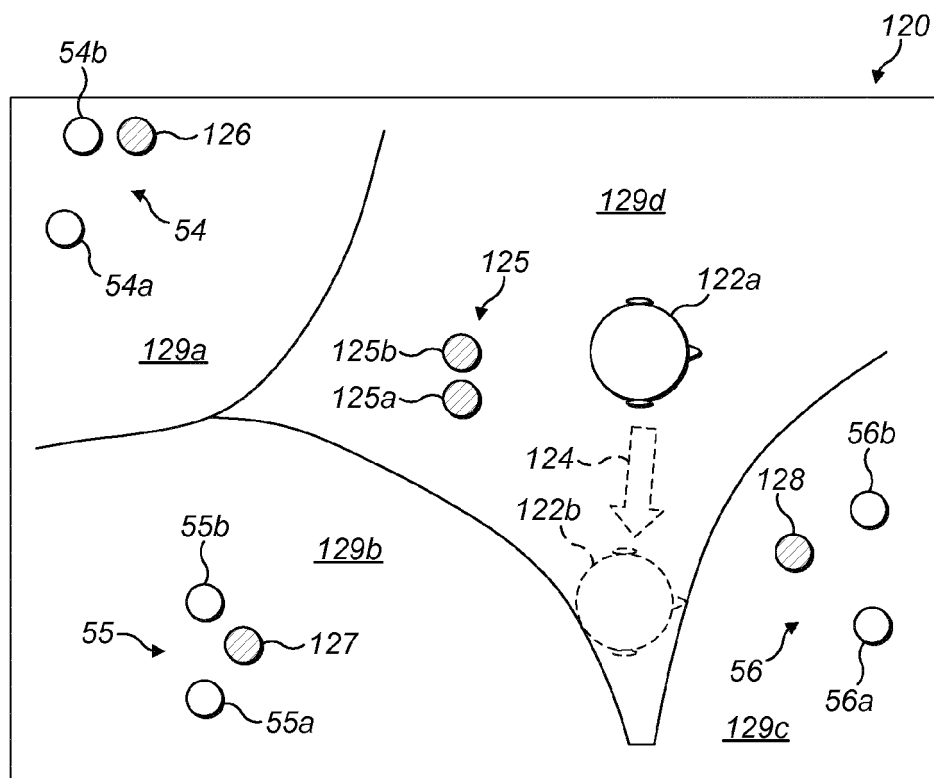


FIG. 12

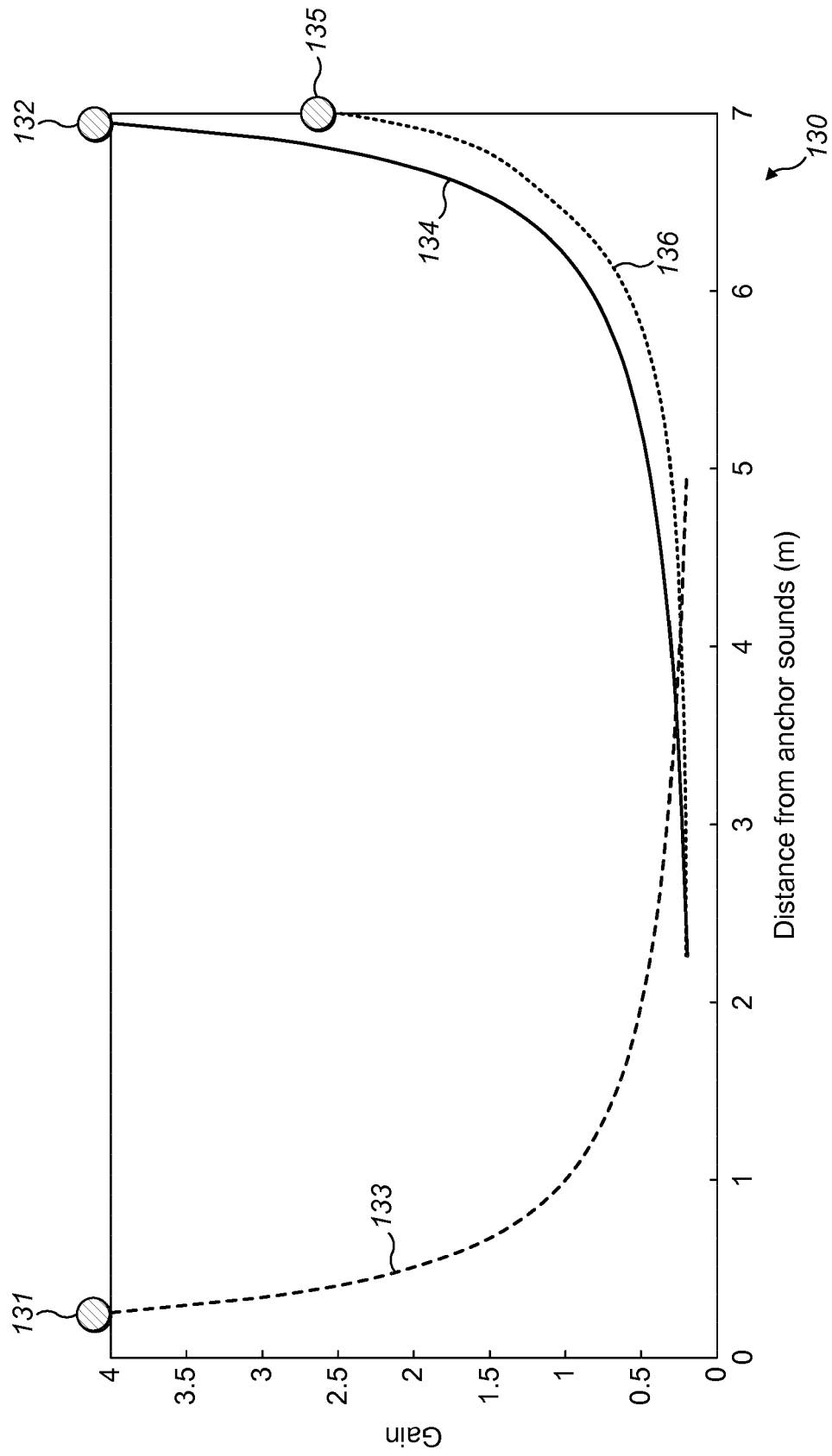


FIG. 13

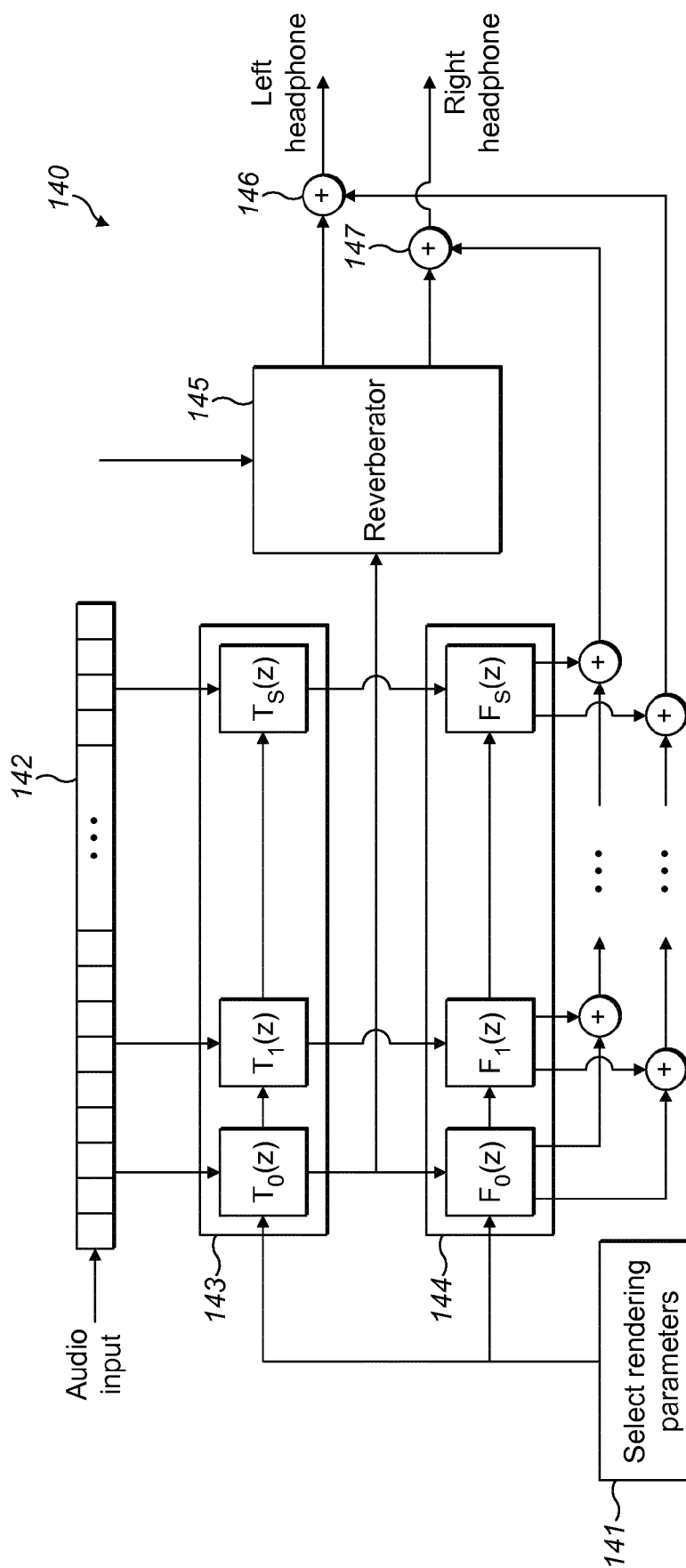


FIG. 14

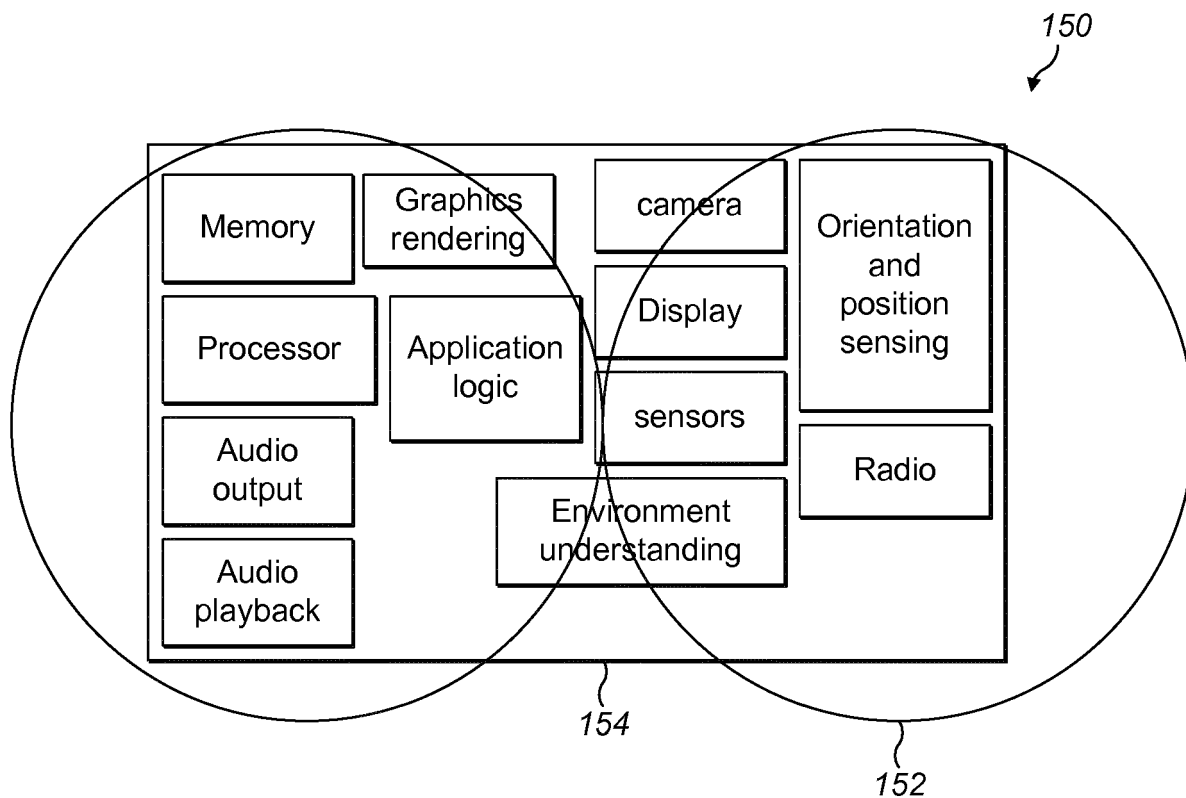


FIG. 15

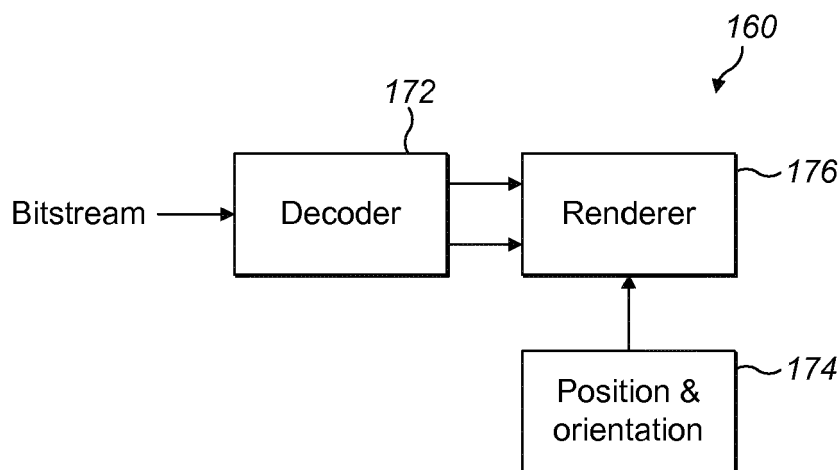


FIG. 16

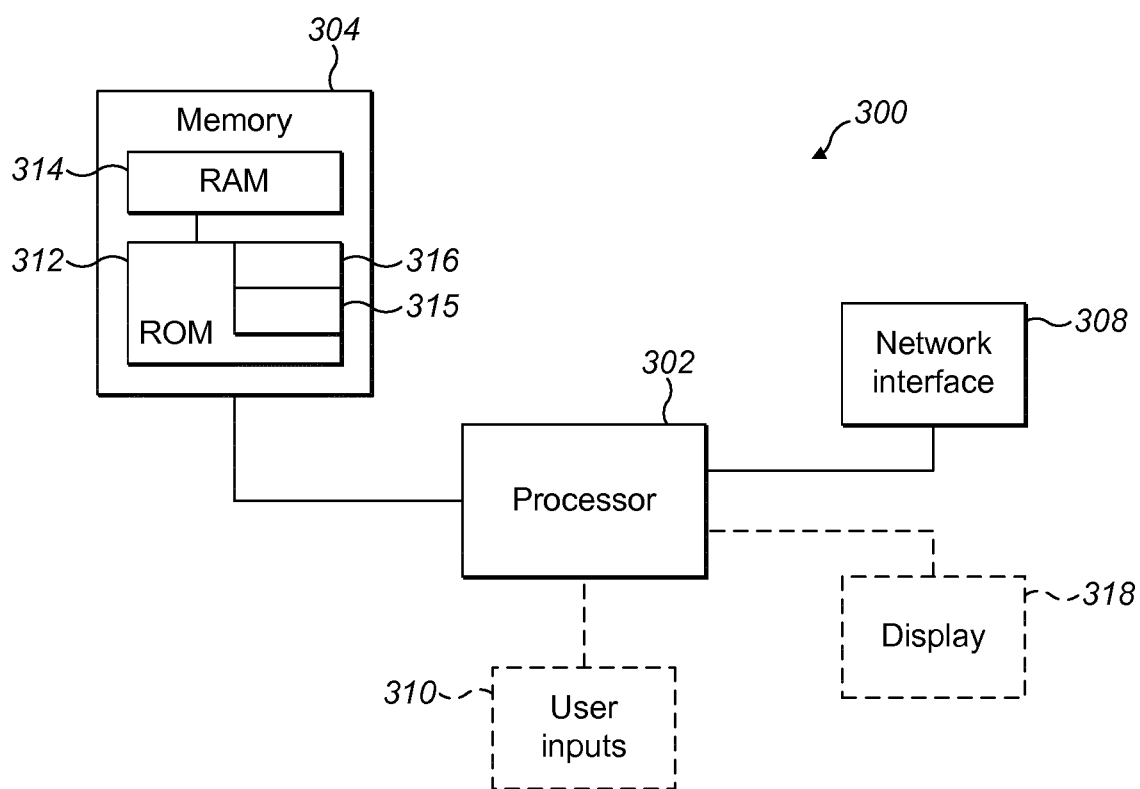


FIG. 17

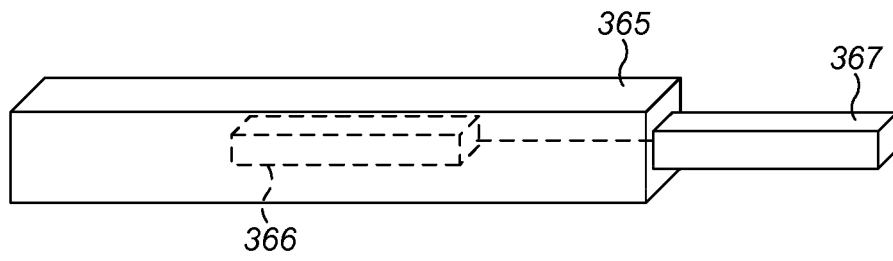


FIG. 18A

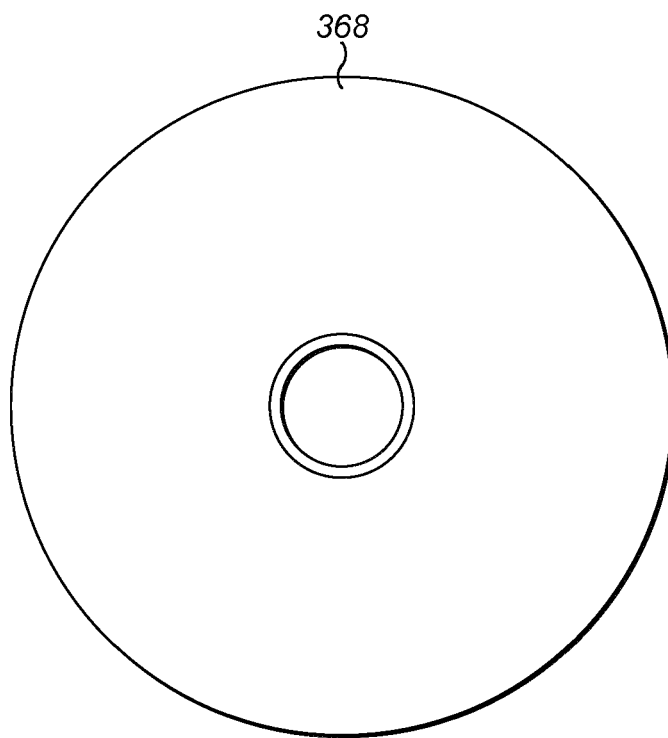


FIG. 18B

**DECLARATION**

Application Number

which under Rule 63 of the European Patent Convention EP 19 15 2619 shall be considered, for the purposes of subsequent proceedings, as the European search report

The Search Division considers that the present application, does not comply with the provisions of the EPC to such an extent that it is not possible to carry out a meaningful search into the state of the art on the basis of all claims

Reason:

1. The application does not meet the requirements of Article 83 EPC, according to which the application must disclose the invention in a manner sufficiently clear and complete for it to be carried out by a person skilled in the art.

1.1 Claim 1 reads:

"An apparatus comprising: means for receiving audio data [...] one or more second sound sources".

From this first passage the skilled person derives that a plurality of audio sound sources are present in a virtual audio scene and are split in two categories: first and second sound sources. How those sound sources are categorized in the two categories is not mentioned in the claims nor in the description. The description only mentions that the second source may be also called "anchor", without defining what an anchor is in this context.

1.2 Claim 1 further reads:

"wherein each first sound source has a position within the virtual audio scene and each second sound source has an initial position within the virtual audio scene and one or more virtual positions within the virtual audio scene".

From this passage the skilled person derives that the first sound source has a position within the virtual audio scene. And that the second sound source has an initial position within the virtual audio scene and one or more virtual positions within the virtual audio scene. However, the features "an initial position within the virtual audio scene" and "one

-/--

CLASSIFICATION OF THE APPLICATION (IPC)

INV.
H04S7/00

EPO FORM 1504 (P04F37)

Place of search

The Hague

Date

22 May 2019

Examiner

Bensa, Julien

**DECLARATION**

Application Number

which under Rule 63 of the European Patent Convention EP 19 15 2619 shall be considered, for the purposes of subsequent proceedings, as the European search report

The Search Division considers that the present application, does not comply with the provisions of the EPC to such an extent that it is not possible to carry out a meaningful search into the state of the art on the basis of all claims

CLASSIFICATION OF THE APPLICATION (IPC)

Reason:

or more virtual positions" for the second sound source are unclear. Those features are not further defined in the description nor in the claims.

The skilled person could guess that an "initial position" is the position of a sound source at the beginning of a movement of said sound source. However, any sound source has an initial position. It is therefore not clear why the initial position is mentioned for the second sound source and not for the first. It is further not clear what are the "one or more virtual positions" of the second sound source. Do they correspond to the final position after a possible movement? But it is then not clear why there may be several of them for a given source. Further, figures 8 and 9 seem to suggest that initial and virtual positions are defined simultaneously, which would be contradictory with a source actually moving. It is finally not clear why two different expressions are used to designate what seems to be the same feature: the first sound source has a "position within the virtual audio scene" while the second sound source has "one or more virtual positions within the virtual audio scene". The difference between a position and a virtual position, both within a virtual audio scene, is not clear. The description does not give further details.

1.3 Claim 1 further reads:

"means for processing audio data of said first sound sources, comprising means for modifying audio data of said first sound

-/--

EPO FORM 1504 (P04F37)

Place of search

The Hague

Date

22 May 2019

Examiner

Bensa, Julien

**DECLARATION**

Application Number

which under Rule 63 of the European Patent Convention EP 19 15 2619 shall be considered, for the purposes of subsequent proceedings, as the European search report

The Search Division considers that the present application, does not comply with the provisions of the EPC to such an extent that it is not possible to carry out a meaningful search into the state of the art on the basis of all claims

Reason:

sources by an audio gain that is based, at least in part, on a distance function for the respective first sound source; and". From this passage the skilled person derives that audio data of the first sound sources are modified using a gain based in part on a distance function. Claim 1 does not define further this distance function. However, claim 2 mentions "the distance function [...] within the virtual audio scene."

In claim 1 the first sound sources had "a position within the virtual audio scene". In claim 2 the first sound sources now have a "virtual location" within the virtual audio scene. It is unclear if these are the same features in view of the fact that "position" and "virtual position" seem already to refer to two different features, as seen above. The description does not give further details. It is therefore not possible, from the application as a whole, to derive how this distance is actually defined.

1.4 Claim 1 further reads:
"means for processing audio data [...] within the virtual audio scene."

From this passage the skilled person can derive that audio data of the second sources are modified based on a distance function. This distance function seems to be different than the previous one. It is now the distance from a virtual location of a user within the virtual audio scene to the location of a selected one of said initial and one or more virtual positions. As seen earlier, those initial and virtual

-/--

CLASSIFICATION OF THE APPLICATION (IPC)

EPO FORM 1504 (P04F37)

Place of search

The Hague

Date

22 May 2019

Examiner

Bensa, Julien

**DECLARATION**

Application Number

which under Rule 63 of the European Patent Convention EP 19 15 2619 shall be considered, for the purposes of subsequent proceedings, as the European search report

The Search Division considers that the present application, does not comply with the provisions of the EPC to such an extent that it is not possible to carry out a meaningful search into the state of the art on the basis of all claims Reason: positions are not defined in the application. It is therefore not possible, from the application as a whole, to clearly derive how this other distance is defined. 1.5 The same objections apply, mutatis mutandis, to the subject-matter of claims 14 and 15. 2. In view of the fact that, considering the claims but also the application as a whole, the first and second sources are not clearly defined, that the "position", "initial position", "virtual positions", "virtual location" are not defined, that the distance functions are consequently not clearly defined either, it is not possible to derive to which subject-matter the claims are related. The applicant's attention is drawn to the fact that a search may be carried out during examination following a declaration of no search under Rule 63 EPC, should the problems which led to the declaration being issued be overcome (see EPC Guideline C-IV, 7.2).	CLASSIFICATION OF THE APPLICATION (IPC)	
Place of search The Hague	Date 22 May 2019	Examiner Bensa, Julien

EPO FORM 1504 (P04F37)