



(11) **EP 3 691 298 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
05.08.2020 Bulletin 2020/32

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

(21) Application number: **19155151.4**

(22) Date of filing: **01.02.2019**

(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

(72) Inventors:
• **MATE, Sujeet Shyamsundar**
33720 Tampere (FI)
• **LEPPÄNEN, Jussi Artturi**
33580 Tampere (FI)
• **VILERMO, Miikka Tapani**
37200 Siuro (FI)
• **LEHTINIEMI, Arto**
33880 Lempäälä (FI)

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

(74) Representative: **Swindell & Pearson Limited**
48 Friar Gate
Derby DE1 1GY (GB)

(54) **APPARATUS, METHOD OR COMPUTER PROGRAM FOR ENABLING REAL-TIME AUDIO COMMUNICATION BETWEEN USERS EXPERIENCING IMMERSIVE AUDIO**

(57) An apparatus comprising means for:
causing rendering of a portion of an audio content to a user, wherein the portion of the audio content is selected by a current point-of-view of the user;
causing real-time communication between the first user and a second user by causing transmission, for rendering to the second user, of audio generated by the user and by causing rendering, to the user, of audio generated by the second user for rendering to the user;
causing adaptation of the portion of the audio content to create an adapted portion of the audio content, by replacing a sound source of the portion of the audio content with a different sound source; and
causing rendering, to the user, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

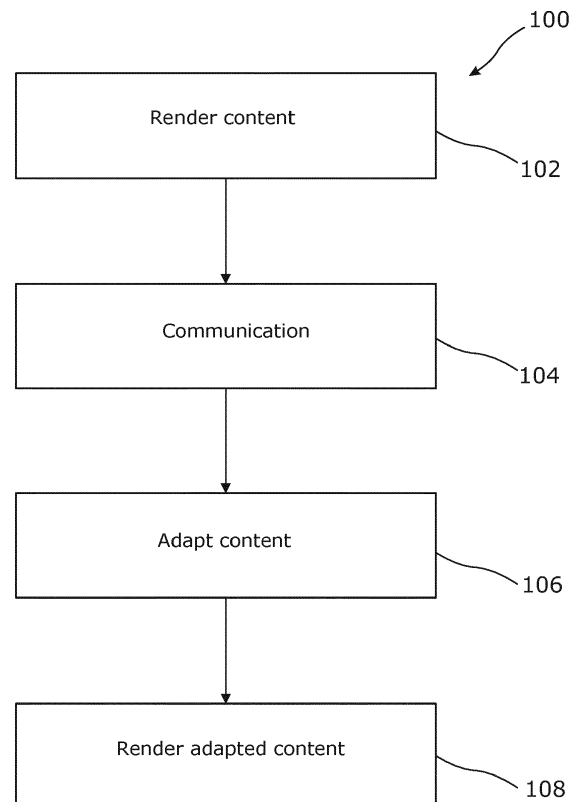


FIG. 4

Description

TECHNOLOGICAL FIELD

[0001] Embodiments of the present invention relate to apparatuses, methods and computer programs for enabling real-time audio communication between users experiencing immersive audio.

BACKGROUND

[0002] Immersive audio describes the rendering to a user of audio content selected by a current point-of-view of the user. The user therefore has the experience that they are immersed within a three-dimensional audio field that changes as their point-of-view changes.

BRIEF SUMMARY

[0003] It would be desirable to enable real-time audio communication between users experiencing immersive audio of the same content from different point of view.

[0004] According to claim 1 there is provided an apparatus. The apparatus comprises means for causing rendering of a portion of an audio content to a first user; causing real-time communication between the first user and a second user; causing adaptation of the portion of the audio content to create an adapted portion of the audio content; and causing rendering, to the first user, of the adapted portion of the audio content instead of the portion of the audio content.

[0005] The portion of the audio content is selected based at least in part on a current point-of-view of the first user;

Causing real-time communication between the first user and a second user comprises causing transmission, for rendering to the second user, of audio from the first user and causing rendering, to the first user, of audio from the second user for rendering to the first user.

[0006] Causing adaptation of the portion of the audio content to create an adapted portion of the audio content comprises replacing a sound source of the portion of the audio content with a different sound source.

[0007] The different sound source is rendered instead of the sound source.

[0008] According to various, but not necessarily all, embodiments there is provided an apparatus comprising means for:

causing rendering of a portion of an audio content to a first user, wherein the portion of the audio content is selected by a current point-of-view of the first user; causing real-time communication between the first user and a second user by causing transmission, for rendering to the second user, of audio generated by the first user and by causing rendering, to the first user, of audio generated by the second user for rendering to the first user;

causing adaptation of the portion of the audio content to create an adapted portion of the audio content, by replacing a sound source of the portion of the audio content with a different sound source; and

causing rendering, to the first user, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

According to various, but not necessarily all, embodiments there is provided an apparatus comprising means for:

causing rendering of a portion of an audio content to a first user, wherein the portion of the audio content is selected by a current point-of-view of the first user;

causing real-time communication between the first user and a second user by causing transmission, for rendering to the second user, of audio generated by the first user and by causing rendering, to the first user, of audio generated by the second user for rendering to the first user; causing adaptation of the portion of the audio content to create an adapted portion of the audio content, by replacing a sound source of the portion of the audio content with a different sound source; and

causing rendering, to the first user, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

[0009] In some but not necessarily all examples, the different sound source originates from a second portion of the audio content, different to the portion of the audio content, wherein the second portion of the audio content is selected by a current point-of-view of the second user.

[0010] In some but not necessarily all examples causing adaptation of the portion of the audio content to create an adapted portion of the audio content comprises replacing multiple sound sources with different sound sources.

[0011] In some but not necessarily all examples, the multiple sound sources are replaced by the different sound sources one-at-a-time and wherein the adapted portion of the audio content is rendered to the user while the one-at-a-time adaptation is on-going, wherein progressively more of the different sound sources are rendered instead of the multiple sound sources.

[0012] In some but not necessarily all examples, when the first user is in a first zone of a plurality of zones and the second user is in a second, different zone of the plurality of zones, wherein the portion of the audio content rendered to the user depends on the point-of-view of the user in the first zone and includes sound sources associated with the first zone and does not include any sound source associated with the second zone, wherein the

content rendered to the second user depends on the point-of-view of the second user in the second zone and includes sound sources associated with the second zone,

and wherein the adapted portion of the audio content rendered to the first user depends on the point-of-view of the first user in the first zone and includes at least one sound source associated with the second zone.

[0013] In some but not necessarily all examples, the apparatus comprises means for causing an undoing of some or all of the adaptation performed on the portion of the audio content to create the adapted portion of the audio content and/or causing rendering, to the first user, of the portion of the audio content instead of the portion of the adapted content, wherein the sound source is rendered instead of the different sound source.

[0014] In some but not necessarily all examples, the un-doing of some or all of the adaptation performed on the portion of the audio content to create the adapted portion of the audio content is performed as a consequence of a change in point-of-view of the first user.

[0015] In some but not necessarily all examples, the portion of audio content rendered to the first user, is defined by a point-of-view of a virtual user in a virtual space, which is determined by a point-of-view of the first user in a real space.

[0016] In some but not necessarily all examples, the point-of-view of the first user is determined by an orientation of the first user or wherein the point-of-view of the first user is determined by an orientation and a location of the user.

[0017] In some but not necessarily all examples, the apparatus is configured as a head mounted apparatus.

[0018] In some but not necessarily all examples, the apparatus comprises means for causing adaptation of the portion of the audio content to create adapted content as a consequence of an initiation of the real-time communication and an additional criterion or criteria.

[0019] In some but not necessarily all examples, the criterion or criteria include a condition based upon determining who will be the target of the adaptation.

[0020] In some but not necessarily all examples, the criterion or criteria are based upon an assessment of a difference between the content portions rendered to the first user and the second user.

[0021] According to various, but not necessarily all, embodiments there is provided a method comprising:

rendering a portion of an audio content to a first user, wherein the portion of the audio content is selected by a current point-of-view of the first user; enabling real-time communication between the first user and a second user by transmitting, to the second user audio generated by the first user and rendering, to the first user, audio received from the second user; adapting the portion of the audio content to create an adapted portion of the audio content, by replacing a sound source of the portion of the audio content

with a different sound source; and rendering, to the first user, the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

[0022] According to various, but not necessarily all, embodiments there is provided a computer program that, when run on a computer, performs:

causing rendering of a portion of an audio content to a first user, wherein the portion of the audio content is selected by a current point-of-view of the first user; causing real-time communication between the first user and a second user by causing transmission, for rendering to the second user, of audio generated by the first user and by causing rendering, to the first user, of audio generated by the second user for rendering to the first user; causing adaptation of the portion of the audio content to create an adapted portion of the audio content, by replacing a sound source of the portion of the audio content with a different sound source; and causing rendering, to the first user, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

[0023] According to various, but not necessarily all, embodiments there is provided an apparatus comprising:

at least one processor; and at least one memory including computer program code the at least one memory and the computer program code configured to, with the at least one processor, cause the apparatus at least to perform:

causing rendering of a portion of an audio content to a first user, wherein the portion of the audio content is selected by a current point-of-view of the first user; causing real-time communication between the first user and a second user by causing transmission, for rendering to the second user, of audio generated by the first user and by causing rendering, to the first user, of audio generated by the second user for rendering to the first user; causing adaptation of the portion of the audio content to create an adapted portion of the audio content, by replacing a sound source of the portion of the audio content with a different sound source; and causing rendering, to the first user, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

[0024] According to various, but not necessarily all, embodiments there is provided examples as claimed in the appended claims.

BRIEF DESCRIPTION

[0025] Some example embodiments will now be described with reference to the accompanying drawings in which:

FIG. 1A, 1B, 1C, 1D show example embodiments of the subject matter described herein;
 FIG. 2 shows another example embodiment of the subject matter described herein;
 FIG. 3 shows another example embodiment of the subject matter described herein;
 FIG. 4 shows another example embodiment of the subject matter described herein;
 FIG. 5A, 5B, 6A, 6B show other example embodiments of the subject matter described herein;
 FIG. 7A, 7B, 7C show other example embodiments of the subject matter described herein;
 FIG. 8 shows another example embodiment of the subject matter described herein;
 FIG. 9 shows another example embodiment of the subject matter described herein;
 FIG. 10 shows another example embodiment of the subject matter described herein;
 FIG. 11 shows another example embodiment of the subject matter described herein;

DEFINITIONS

[0026] "*artificial environment*" may be something that has been recorded or generated.

[0027] "*virtual visual space*" refers to fully or partially *artificial environment* that may be viewed, which may be three dimensional.

[0028] "*virtual visual scene*" refers to a representation of the *virtual visual space* viewed from a particular *point of view* (position) within the *virtual visual space*.

[0029] '*virtual visual object*' is a visible virtual object within a *virtual visual scene*.

[0030] "*sound space*" (or "*virtual sound space*") refers to an arrangement of sound sources in a three-dimensional space. A *sound space* may be defined in relation to recording sounds (a recorded sound space) and in relation to rendering sounds (a rendered sound space).

[0031] "*sound scene*" (or "*virtual sound scene*") refers to a representation of the *sound space* listened to from a particular *point of view* (*position*) within the *sound space*.

[0032] "*sound object*" refers to a sound source that may be located within the *sound space*. A source sound object represents a sound source within the sound space, in contrast to a sound source associated with an object in the virtual visual space. A recorded sound object represents sounds recorded at a particular microphone or

location. A rendered sound object represents sounds rendered from a particular location.

[0033] "*virtual space*" may mean a *virtual visual space*, mean a *sound space* or mean a combination of a *virtual visual space* and corresponding *sound space*. In some examples, the *virtual space* may extend horizontally up to 360° and may extend vertically up to 180°.

[0034] "*virtual scene*" may mean a *virtual visual scene*, mean a *sound scene* or mean a combination of a *virtual visual scene* and corresponding *sound scene*.

[0035] '*virtual object*' is an object within a *virtual scene*, it may be an augmented virtual object (e.g. a computer-generated virtual object) or it may be an image of a real object in a *real space* that is live or recorded. It may be a *sound object* and/or a *virtual visual object*.

[0036] "*Virtual position*" is a position within a virtual space. It may be defined using a *virtual location* and/or a *virtual orientation*. It may be considered to be a movable '*point of view*'.

[0037] "*Correspondence*" or "*corresponding*" when used in relation to a *sound space* and a *virtual visual space* means that the *sound space* and *virtual visual space* are time and space aligned, that is they are the same space at the same time.

[0038] "*Correspondence*" or "*corresponding*" when used in relation to a *sound scene* and a *virtual visual scene* (or visual scene) means that the *sound space* and *virtual visual space* (or visual scene) are corresponding and a *notional* (*virtual*) *listener* whose *point of view* defines the *sound scene* and a *notional* (*virtual*) *viewer* whose *point of view* defines the *virtual visual scene* (or visual scene) are at the same location and orientation, that is they have the same *point of view* (same virtual position).

[0039] "*real space*" (or "*physical space*") refers to a real environment, which may be three dimensional.

[0040] "*real scene*" refers to a representation of the *real space* from a particular *point of view* (position) within the real space.

[0041] "*real visual scene*" refers to a visual representation of the real space viewed from a particular real *point of view* (position) within the real space.

[0042] "*mediated reality*" in this document refers to a user experiencing, for example visually and/or aurally, a fully or partially artificial environment (a *virtual space*) as a *virtual scene* at least partially rendered by an apparatus to a *user*. The *virtual scene* is determined by a *point of view* (virtual position) within the *virtual space*. *Displaying the virtual scene* means providing a virtual visual scene in a form that can be perceived by the *user*.

[0043] "*augmented reality*" in this document refers to a form of *mediated reality* in which a *user* experiences a partially artificial environment (a *virtual space*) as a *virtual scene* comprising a *real scene*, for example a *real visual scene*, of a physical real environment (*real space*) supplemented by one or more visual or audio *elements* rendered by an apparatus to a *user*. The term augmented reality implies a mixed reality or hybrid reality and does

not necessarily imply the degree of virtuality (vs reality) or the degree of mediality;

"*virtual reality*" in this document refers to a form of *mediated reality* in which a user experiences a fully artificial environment (a *virtual visual space*) as a *virtual scene* displayed by an *apparatus* to a *user*;

"*virtual content*" is content, additional to *real content* from a real scene, if any, that enables *mediated reality* by, for example, providing one or more augmented *virtual objects*.

[0044] "*mediated reality content*" is virtual content which enables a *user* to experience, for example visually and/or aurally, a fully or partially artificial environment (a *virtual space*) as a *virtual scene*. *Mediated reality content* could include interactive content such as a video game or non-interactive content such as motion video.

[0045] "*augmented reality content*" is a form of mediated reality content which enables a *user* to experience, for example visually and/or aurally, a partially artificial environment (a *virtual space*) as a *virtual scene*. *Augmented reality content* could include interactive content such as a video game or non-interactive content such as motion video.

[0046] "*virtual reality content*" is a form of mediated reality content which enables a *user* to experience, for example visually and/or aurally, a fully artificial environment (a *virtual space*) as a *virtual scene*. *Virtual reality content* could include interactive content such as a video game or non-interactive content such as motion video.

[0047] "*perspective-mediated*" as applied to *mediated reality*, *augmented reality* or *virtual reality* means that *user* actions determine the *point of view* (virtual position) within the virtual space, changing the *virtual scene*;

"*first person perspective-mediated*" as applied to *mediated reality*, *augmented reality* or *virtual reality* means *perspective mediated* with the additional constraint that the *user's* real *point of view* (location and/or orientation) determines the *point of view* (virtual position) within the *virtual space of a virtual user*;

"*third person perspective-mediated*" as applied to *mediated reality*, *augmented reality* or *virtual reality* means *perspective mediated* with the additional constraint that the *user's* real *point of view* does not determine the *point of view* (virtual position) within the *virtual space*;

"*user interactive*" as applied to *mediated reality*, *augmented reality* or *virtual reality* means that *user* actions at least partially determine what happens within the *virtual space*;

"*displaying*" means providing in a form that is perceived visually (viewed) by the *user*.

"*rendering*" means providing in a form that is perceived by the *user*

[0048] "*virtual user*" defines the *point of view* (virtual position- location and/or orientation) in *virtual space* used to generate a *perspective-mediated sound scene* and/or *visual scene*. A virtual user may be a notional listener and/or a notional viewer.

[0049] "*notional listener*" defines the *point of view* (vir-

tual position- location and/or orientation) in *virtual space* used to generate a *perspective-mediated sound scene*, irrespective of whether or not a *user* is actually listening

[0050] "*notional viewer*" defines the *point of view* (virtual position- location and/or orientation) in *virtual space* used to generate a *perspective-mediated visual scene*, irrespective of whether or not a *user* is actually viewing.

[0051] *Three degrees of freedom (3DoF)* describes *mediated reality* where the *virtual position* is determined by orientation only (e.g. the three degrees of three-dimensional orientation). An example of three degrees of three-dimensional orientation is pitch, roll and yaw. In relation to *first person perspective-mediated reality 3DoF*, only the *user's* orientation determines the *virtual position*.

[0052] *Six degrees of freedom (6DoF)* describes mediated reality where the *virtual position* is determined by both orientation (e.g. the three degrees of three-dimensional orientation) and location (e.g. the three degrees of three-dimensional location). An example of three degrees of three-dimensional orientation is pitch, roll and yaw. An example of three degrees of three-dimensional location is a three-dimensional coordinate in a Euclidian space spanned by orthogonal axes such as left -to-right (**x**), front to back (**y**) and down to up (**z**) axes. In relation to *first person perspective-mediated reality 6DoF*, both the *user's* orientation and the *user's* location in the *real space* determine the *virtual position*. In relation to *third person perspective-mediated reality 6DoF*, the *user's* location in the *real space* does not determine the *virtual position*. The *user's* orientation in the *real space* may or may not determine the *virtual position*.

[0053] *Three degrees of freedom 'plus' (3DoF+)* describes an example of six degrees of freedom where a change in location (e.g. the three degrees of three-dimensional location) is a change in location relative to the user that can arise from a postural change of a user's head and/or body and does not involve a translation of the user through real space by, for example, walking.

[0054] "*spatial audio*" is the rendering of a sound scene. "*First person perspective spatial audio*" or "*immersive audio*" is spatial audio where the *user's* point of view determines the sound scene so that audio content selected by a current point-of-view of the user is rendered to the user.

DETAILED DESCRIPTION

[0055] FIGS. 1A, 1B, 1C, 1D, illustrate first person perspective mediated reality. In this context, mediated reality means the rendering of mediated reality for the purposes of achieving mediated reality for a remote user, for example augmented reality or virtual reality. It may or may not be user interactive. The mediated reality may support one or more of: 3DoF, 3DoF+ or 6DoF.

[0056] FIGS. 1A, 1C illustrate at a first time a real space 50 and a sound space 60. A user 40 in the real space 50 has a point of view (a position) 42 defined by a location

46 and an orientation 44. The location is a three-dimensional location and the orientation is a three-dimensional orientation.

[0057] In an example of 3DoF mediated reality, the user's real point-of-view 42 (orientation) determines the point-of-view 72 (virtual position) within the virtual space (e.g. sound space 60) of a virtual user 70. An orientation 44 of the user 40 controls a virtual orientation 74 of a virtual user 70. There is a correspondence between the orientation 44 and the virtual orientation 74 such that a change in the orientation 44 produces the same change in the virtual orientation 74.

[0058] The virtual orientation 74 of the virtual user 70 in combination with a virtual field of hearing defines a virtual sound scene 78.

[0059] A virtual sound scene 78 is that part of the sound space 60 that is rendered to a user. In 3DoF mediated reality, a change in the location 46 of the user 40 does not change the virtual location 76 or virtual orientation 74 of the virtual user 70.

[0060] In the example of 6DoF mediated reality, the user's real point-of-view 42 (location 46 and/or orientation 44) determines the point-of-view 72 (virtual position) within the virtual space (e.g. sound space 60) of a virtual user 70. The situation is as described for 3DoF and in addition it is possible to change the rendered virtual sound scene 78 by movement of a location 46 of the user 40. For example, there may be a mapping between the location 46 of the user 40 and the virtual location 76 of the virtual user 70.

[0061] A change in the location 46 of the user 40 produces a corresponding change in the virtual location 76 of the virtual user 70. A change in the virtual location 76 of the virtual user 70 changes the rendered virtual sound scene 78.

[0062] This may be appreciated from FIGS. 1B, 1D which illustrate the consequences of a change in location 46 and orientation 44 of the user 40 on the rendered virtual sound scene 78 (FIG. 1D). The change in location may arise from a postural change of the user and/or a translation of the user by walking or otherwise.

[0063] First person perspective mediated reality may control only a virtual sound scene 78, a virtual visual scene and both a virtual sound scene 78 and virtual visual scene, depending upon implementation.

[0064] In some situations, for example when the virtual sound scene 78 is rendered to a listener through a head-mounted audio output device, for example headphones using binaural audio coding, it may be desirable for the rendered sound space 60 to remain fixed in real space when the listener turns their head in space. This means that the rendered sound space 60 needs to be rotated relative to the audio output device by the same amount in the opposite sense to the head rotation. The orientation of the portion of the rendered sound space tracks with the rotation of the listener's head so that the orientation of the rendered sound space remains fixed in space and does not move with the listener's head.

[0065] A sound 'locked' to the real world may be referred to as a diegetic sound.

A sound 'locked' to the user's head may be referred to as a non-diegetic sound.

[0066] The rendering of a virtual sound scene 78 may also be described as providing spatial audio or providing immersive audio.

[0067] As illustrated in FIG 2, in at least some examples, the sound space 60 defined by audio content 10 comprises one or more sound sources 20 at different positions in the sound space 60. The audio rendered to the user depends upon the relative position of the virtual user 70 from the positions of the sound sources 20. Perspective mediated virtual reality, for example first person perspective mediated reality enables the user 40 to change the position of the virtual user 70 within the sound space 60 thereby changing the positions of the sound sources 20 relative to the virtual user which changes the virtual sound scene 78 rendered to the user 40.

[0068] Channel-based audio, for example, n,m surround sound (e.g. 5.1, 7.1 or 22.2 surround sound) or binaural audio, can be used or scene-based audio, including spatial information about a sound field and sound sources, can be used.

[0069] Audio content may encode spatial audio as audio objects. Examples include but are not limited to MPEG-4 and MPEG SAOC. MPEG SAOC is an example of metadata-assisted spatial audio.

[0070] Audio content may encode spatial audio as audio objects in the form of moving virtual loudspeakers.

[0071] Audio content may encode spatial audio as audio signals with parametric side information or metadata. The audio signals can be, for example, First Order Ambisonics (FOA) or its special case B-format, Higher Order Ambisonics (HOA) signals or mid-side stereo. For such audio signals, synthesis which utilizes the audio signals and the parametric metadata is used to synthesize the audio scene so that a desired spatial perception is created.

[0072] The parametric metadata may be produced by different techniques. For example, Nokia's spatial audio capture (OZO Audio) or Directional Audio Coding (DirAC) can be used. Both capture a sound field and represent it using parametric metadata. The parametric metadata may for example comprise: direction parameters that indicate direction per frequency band; distance parameters that indicate distance per frequency band; energy-split parameters that indicate diffuse-to-total energy ratio per frequency band. Each time-frequency tile may be treated as a sound source with the direction parameter controlling vector based amplitude panning for a direct version and the energy-split parameter controlling differential gain for an indirect (decorrelated) version.

[0073] The audio content encoded may be speech and/or music and/or generic audio.

[0074] 3GPP IVAS (3GPP, Immersive Voice and Audio services), which currently under development, is expected to support new immersive voice and audio services,

for example, mediated reality.

[0075] In some but not necessarily all examples amplitude panning techniques may be used to create or position a sound object. For example, the known method of vector-based amplitude panning (VBAP) can be used to position a sound source.

[0076] A sound object may be re-positioned by mixing a direct form of the object (an attenuated and directionally-filtered direct sound) with an indirect form of the object (e.g. positioned directional early reflections and/or diffuse reverberant).

[0077] FIG. 2 illustrates an example of a sound space 60 comprising a plurality of sound sources 20 at different locations within the sound space 60. Each sound source 20 has associated with it a sound field 22, which may be a bearing, an area or a volume. When the virtual user 70 is aligned with or is within the sound field 22, then the user 40 has a different experience of the sound source 20 than if they are outside the sound field 22. In some examples, the user 40 may only hear the sound source 20 when the virtual user 70 is within the sound field 22 and cannot hear the sound source 20 outside the sound field 22. In other examples, the sound source 20 can be best heard within the sound field 22 and the sound source 20 is attenuated outside of the sound field 22 and in some examples it is more attenuated the greater the deviation or distance from the sound field 22.

[0078] The sound sources 20 and their locations and other characteristics of the sound space 60 are defined by the audio content 10 which is spatial audio content because sound sources 20 are controllably located within the sound space 60 by the audio content 10. A reference to 'audio content' in this document can also therefore be a reference to 'spatial audio content'. It will therefore be understood that the user 40, who is represented by the virtual user 70 in the sound space 60, experiences immersive audio. A portion of the audio content 10 is selected by a current point-of-view 42 of the user 40 (point-of-view 72 of the virtual user 70). That portion of the audio content 10 is rendered to the user 40.

[0079] The user 40, by changing their own point-of-view 42, can change the point-of-view 72 of the virtual user 70 to appreciate different aspects of the sound space 60. In some examples, the change in the point-of-view 42 of the user 40 is achieved by varying only the user's orientation 44 and in other examples it is achieved by changing the user's orientation 44 and/or the user's location 46. The audio content 10 can therefore support 3DoF, 3DoF+, and 6DoF.

[0080] In this example, the sound space 60 comprises a number of distinct zones 30. Each of the zones 30 is fully or partially isolated from the other zones or at least some of the other zones. Isolation in this context means that if the user is located within a particular zone 30, then the immersive audio that they experience is dominated by the sound sources of that zone. In some examples they may only hear the sound sources of that zone. In other examples they may not hear the sound sources of

some or all of the other zones. Even in the circumstances where the virtual user 70 is within a zone 30 it is likely that the sound sources of that zone will be dominant compared to the sound sources of any other zone 30.

[0081] The user 40 can change their point-of-view 42, to cause a consequent change in the point-of-view 72 of the virtual user 70 within a zone 30. This allows the user 40 to appreciate different aspects of the composition formed by the different sound sources 20 within the zone 30. As previously described, the change of point-of-view 72 within a zone may be achieved by 3DoF, 3DoF+, 6DoF. In at least some examples, there are one or more sweet spots in a zone 30. A sweet spot is a particular point-of-view 72 for a virtual user 70 at which a better composition of the sound sources 20 in the zone 30 is rendered. The composition is a mixed balance of the sound sources 20 of the zone 30.

[0082] The virtual user 70 can emphasize a sound source 20 in the rendering of the sound scene by, for example:

- (i) moving towards the sound source 20
- (ii) turning towards the sound source 20;
- (iii) moving out of the sound field 22 of a sound source 20

[0083] The virtual user 70 can de-emphasize a sound source 20 in the rendering of the sound scene by, for example:

- (i) moving away from the sound source 20
- (ii) turning away from the sound source 20;
- (iii) moving into the sound field 22 of a sound source 20

[0084] It is also possible for the virtual user 70 to move between the different zones 30. The user 40 is able to control the location of the virtual user 70 within the sound space 60.

[0085] It will therefore be appreciated that, in general, the virtual user 70 by changing their location and/or orientation with respect to the sound source 20 can control how the sound source 20 is rendered to the user 40. The point-of-view of the user 40 controls the point-of-view of the virtual user 70.

[0086] In the particular example illustrated, but not necessarily all examples, the sound sources 20 of the sound space 60 are musical instruments. Each of the zones 30 has a main instrument and none, or one or more complementing instruments. The main instrument is represented by a sound source 20. Each of the complementing instruments, if present, is represented by a distinct sound source 20. The secondary instruments of a zone 30 complement the primary instrument of the zone 30. However, the instruments of one zone do not necessarily complement the instruments of another zone. It may therefore be desirable to prevent a user 40 hearing a mix of instruments from different zones. It may, for example, be de-

sirable to prevent the simultaneous rendering to a user 40 of particular combinations of sound sources 20 from different zones.

[0087] FIG. 3 is an example of zonal spatial audio content 10 similar to the audio content 10 illustrated in FIG. 2 but at a higher obstruction level, highlighting the delineation of the different zones 30. In this example, zone 1 is isolated from zones 2, 3 and 4 but not from the zone associated with the baseline instruments. Likewise, zone 2 is isolated from zones 1, 3 and 4 but not from the zone associated with the baseline instruments. Likewise, zone 4 is isolated from zones 1, 2 and 3 but not from the zone associated with the baseline instruments. Likewise, zone 3 is isolated from zones 1, 2 and 4 but not from the zone associated with the baseline instruments. As a consequence, when the virtual user 70 is in zone 1 the sound scene rendered to the user 40 is primarily dependent upon the sound sources 20 of zone 1 and the point-of-view of the virtual user 70 within zone 1 but may also include at a secondary level sound sources from the zone associated with the baseline instruments. When the virtual user 70 is in zone 2 the sound scene rendered to the user 40 is primarily dependent upon the sound sources 20 of zone 2 and the point-of-view of the virtual user 70 within zone 2 but may also include at a secondary level sound sources from the zone 2 the zone associated with the baseline instruments.

[0088] Thus, the baseline instruments may be heard in all zones. The sound sources of the other zones can only be heard if the virtual user 70 is within that particular zone 30.

[0089] It will be seen that there are gaps between the various different zones 30. In some examples, in these gaps only the baseline instruments can be heard. In other examples no sound sources can be heard.

[0090] In the description thus far, it has been assumed that there is a single user 40 and a single virtual user 70. However, it may be possible for multiple users 40 to share the same audio content 10 (common audio content 10). Each of the users 40 is associated with a different virtual user 70. The point-of-view of each of the virtual users 70 can be independently controlled by each of the respective users 40. As a consequence, each of the users 40 can independently experience the immersive spatial audio defined by the audio content 10.

[0091] Thus, a first portion of the audio content 10 is selected by a current point-of-view 42 of a first user 40. This is rendered to the first user 40. The first user can change the current point-of-view 42 and change the portion of the audio content 10 that is selected and rendered to the first user 40.

[0092] Also, a second portion of the audio content 10 is selected by a current point-of-view 42 of a second user 40, different to the first user 40. This second portion of the audio content 10 is rendered to the second user 40. The second user 40 can change the current point-of-view 42 and change the portion of the audio content 10 that is selected and rendered to the second user 40.

[0093] Where there are multiple users 40 it is desirable to enable real-time communication between the users, for example between the first user 40 and the second user 40. For example, audio from by the first user, for example the voice of the first user, is transmitted to the second user and is then rendered to the second user. For example, audio from the second user, for example the voice of the second user, is transmitted to the first user and is then rendered to the first user

[0094] The audio from the first user may be audio that is generated by the first user, that originates from the first user and/or that is uploaded by the first user. In some examples, it can be a real-time recorded voice of the first user or a real-time recorded environment of the first user. In other examples it can be or include uploaded audio content.

[0095] The audio from the second user may be audio that is generated by the second user, that originates from the second user and/or that is uploaded by the second user. In some examples, it can be a real-time recorded voice of the second user or a real-time recorded environment of the second user. In other examples it can be or include uploaded audio content.

[0096] The first user would consequently hear the audio generated by the second user and the first portion of the audio content selected by the current point-of-view 42 of the first user 40.

[0097] The second user would consequently hear the audio generated by the first user and the second portion of the audio content 10 selected by a current point-of-view 42 of the second user 40.

[0098] The audio context (the rendered sound scene) in which the second user generates the audio transmitted to the first user (the second portion of the audio content 10) and the audio context (the rendered sound scene) in which the first user hears that audio (the first portion of the audio content 10) are different. The audio context in which the first user generates the audio transmitted to the second user (the first portion of the audio content 10) and the audio context in which the second user hears that audio (the second portion of the audio content 10) are different.

[0099] In some examples, this contextual difference may be undesirable.

[0100] In some examples, it is desirable to address this contextual difference by forcing a common context on one or other or both of the first user and the second user. For example, the second portion of the audio content 10 could additionally be rendered to the first user and/or the first portion of the audio content 10 could additionally be rendered to the second user. However, in either of these circumstances then a single user may simultaneously hear both the first portion of the audio content 10 and the second portion of the audio content 10. In some circumstances this may be undesirable because the first portion of the audio content 10 and the second portion of the audio content 10 should be excluded from simultaneous rendering.

[0101] FIG. 4 illustrates an example of a method 100 that is capable of addressing some or all of these problems and other problems.

[0102] In the method 100, at block 102, a portion of audio content is rendered to a user 40. The portion of audio content is selected by a current point-of-view 42 of the first user 40_A.

[0103] At block 104, real-time communication between the first user 40_A and a second user 40_B is enabled. Audio generated by the first user 40_A for rendering to the second user 40_B is transmitted to the second user 40_B. Audio generated by the second user 40_B for rendering to the first user 40_A is rendered to the first user 40_A.

[0104] At block 106, the method 100 enables adaptation of the portion of the audio content 10 to create an adapted portion of the audio content 10, by replacing a sound source 20 of the portion of the audio content with a different sound source.

[0105] At block 108, the method 100 enables rendering, to the first user 40_A, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

[0106] FIG. 5A illustrates an example of a first zone 30_A of the sound space 60 defined by the common audio content 10 at a first time. The first zone 30_A comprises sound sources 20* at different positions within the first zone 30_A. In this example, the first zone 30_A overlies the real space occupied by a first user 40_A. The first user 40_A has a point-of-view 42_A defined by an orientation 44_A and/or a location 46_A as previously described.

[0107] The point-of-view 42_A of the first user 40_A defines the point-of-view 72_A of a first virtual user in the first zone 30_A. In this example, the virtual user is co-located with the first user 40_A and is not illustrated in FIG. 5A for clarity. The first user 40_A is experiencing immersive audio defined by the first zone 30_A as previously described.

[0108] FIG. 5B illustrates an example of a second zone 30_B of the sound space 60 defined by the common audio content 10 at the first time. The second zone 30_B comprises sound sources 20# at different positions within the second zone 30_B. In this example, the second zone 30_B overlies the real space occupied by a second user 40_B. The second user 40_B has a point-of-view 42_B defined by an orientation 44_B and/or a location 46_B as previously described. The point-of-view 42_B of the second user 40_B defines the point-of-view 72_B of a second virtual user in the second zone 30_B. In this example, the virtual user is co-located with the second user 40_B and is not illustrated in FIG. 5B for clarity. The second user 40_B is experiencing immersive audio defined by the second zone 30_B as previously described.

[0109] FIG. 6A illustrates an example of a first zone 30_A of the sound space 60 defined by the common audio content 10 at a second time and is similar to FIG. 5A.

[0110] FIG. 6B illustrates an example of a second zone 30_B of the sound space 60 defined by the common audio content 10 at a second time and is similar to FIG. 5B. The

example illustrated in FIG. 6B is different to the example illustrated in FIG. 5B in that there is real-time communication 150 between the first user 40_A and the second user 40_B and the method 100 has been applied for the second user 40_B. There is transmission of first audio 150_A generated by the first user 40_A to the second user 40_B, and rendering of the first audio 150_A to the second user 40_B. There is also transmission of second audio 150_B generated by the second user 40_B to the first user 40_A and rendering of the second audio 150_B to the first user 40_A. In some, but not necessarily all examples, the first audio 150_A may be a sound source rendered at a particular location relative to the second user 40_B. In some, but not necessarily all examples, the second audio 150_B may be a sound source rendered at a particular location relative to the first user 40_A.

[0111] In the example of FIG. 6B, the second portion 10_B of the audio content 10 illustrated in FIG. 5B has been adapted to create an adapted second portion 10_B* of the audio content 10. Comparing FIG. 5B and 6B, a sound source 20# of the second portion 10_B of the audio content 10 is replaced with a different sound source 20*. The different sound source 20* is a sound source defined by the first portion 10_A of the common audio content 10. The adapted second portion 10_B* of the common audio content 10 is rendered to the second user 40_B instead of the second portion 10_B of the common audio content 10. The different sound source 20* is rendered instead of the original sound source 20#.

[0112] In this example, the different sound source 20* originates from the first portion 10_A of the same audio content 10. The first portion 10_A of the audio content 10 is selected by a current point-of-view 42_A of the first user 40_A. The original sound source 20# has a position in the second portion 10_B of the common audio content 10 relative to the second user 40_B (and the virtual user 70). That is, it has a particular position in the sound space 60. The different replacement sound source 20* has the same position in the adapted second portion 10_B* of the content relative to second user 40_B (and the virtual user 70). That is, the different sound source 20* has replaced the original sound source 20# at the same position within the sound space 60.

[0113] FIGS 7A, 7B and 7C illustrate an extension of the example illustrated in FIGS 5A, 5B, 6A and 6B. In this example FIG. 7A is equivalent to FIG. 5B and FIG. 7B is equivalent to FIG. 6B. FIG. 7C illustrates that the second portion 10_B of the common audio content 10 that is adapted to create the adapted second portion 10_B* of the audio content 10 is adapted by replacing multiple sound sources by different sound sources. In this example, all of the sound sources 20# associated with the second portion 10_B of the common audio content 10 have been replaced with sound sources 20* associated with the first portion 10_A of the common audio content 10. It can be seen by the time progression from FIGS 7A to 7B to 7C, that the second portion 10_B of the common audio content 10 is adapted one sound source at a time. Thus

the multiple sound sources are replaced by the different sound sources one-at-a-time. For example, the shaker sound source is replaced by the electric guitar with sound source in FIG. 7B, then the ukulele sound source is replaced by the electric guitar sound source in FIG. 7C. The adapted second portion 10_B^* of the common audio content 10 is rendered to the second user 40_B instead of the second portion 10_B of the common content while the adaptation is on-going. Consequently, progressively more of the different sound sources 20^* are rendered instead of the original sound sources $20\#$. Thus the adapted portion of the audio content is rendered to the user while the one-at-a-time adaptation is on-going, wherein progressively more of the different sound sources are rendered instead of the multiple sound sources. In this example, but not necessarily all examples, the multiple replacement sound sources 20^* have the same positions in the sound space 60 as the original sound sources $20\#$. Thus, the position of the multiple original sound sources $20\#$ relative to the second user 40_B is the same as the positions of the multiple different replacement sound sources 20^* relative to the second user 40_B .

[0114] In the examples illustrated in FIG. 5A, 5B, 6A, 6B, 7A, 7B, 7C, the first user 40_A (and the corresponding first virtual user) is in a first zone 30_A of a plurality of zones and the second user 40_B (and the corresponding first virtual user) is in a second, different second zone 30_B . The first portion 10_A of the audio content 10 rendered to the first user 40_A depends on the point-of-view 42_A of the first user 40_A in the first zone 30_A and only includes sound sources 20 associated with the first zone 30_A (FIG. 5A, 6A).

[0115] The second portion 10_B of the audio content 10 rendered to the second user 40_B depends on the point-of-view 42_B of the second user 40_B in the second zone 30_B and only includes sound sources 20 associated with the second zone 30_B (FIG. 5B, FIG. 7A).

[0116] The adapted second portion 10_B^* of the common audio content 10 rendered to the second user 10_B depends on the point-of-view 42_B of the second user 40_B in the second zone 30_B and includes at least one sound source 20^* associated with the first zone 30_A .

[0117] In the above examples, the sound sources $20\#$ that were originally rendered to the second user 40_B have been replaced with one or more sound sources 20^* that are rendered to the first user 40_A .

[0118] It should be appreciated, however, that the sound source swap may occur in the reverse direction either additionally to or as an alternative to the above-described sound source swap thus, in some examples, the sound sources 20^* rendered to the first user 40_A may be replaced by one or more different sound sources $20\#$ that are being rendered to the second user 40_B .

[0119] In some, but not necessarily all, examples, it may be desirable to wholly or partially undo the replacement of one sound source by another sound source or indeed undo the replacement of all of the sound sources that have been replaced. It is therefore desirable to pro-

vide a user with control over undoing some or all of the adaptation performed on the portion of the audio content to create the adapted portion of the audio content. After this undoing, the portion of the audio content can be rendered to the user instead of the adapted portion of the audio content and then the original sound source is rendered instead over the different, replacement sound source.

[0120] In some, but not necessarily all, examples, this undoing may be as a consequence of a change in a point-of-view 42 of a user 40 (or changing point-of-view 72 of a virtual user 70). In other examples, or in the same examples, the undoing may be as a consequence of a user 40 (and virtual user 70) changing from one zone to another zone.

[0121] There is therefore some flexibility in the method 100. Decisions may need to be made as to whether or not a sound source 20 is to be replaced and, if so, what it is to be replaced with. Likewise, a decision may need to be taken in relation to the undoing of any replacements.

[0122] In some, but not necessarily all, examples, it is desirable to define mutually exclusive subsets of audio, for example sound sources. Thus, if as a consequence of real-time communication between the first user and the second user, as illustrated in FIGS 6A and 6B, then the users receive new audio 150_A , 150_B . It is possible that the received sound audio 150_A , 150_B that are to be rendered are mutually exclusive in relation to the sound sources already being rendered to the users. Thus, in the examples of FIG. 6A and 6B, a sound source 150_B that is transmitted to the second user 40_B for rendering to the second user 40_B is incompatible with the one or more of the original sound sources $20\#$ currently being rendered to the second user 40_B . As a consequence, the mutual exclusion is resolved by replacing the one or more of the original sound sources $20\#$ with a different sound source 20^* . In this example, the replacement sound source 20^* is a sound source from the first zone 30_A from which the new sound source 150_B originates. The replacement of the one or more sound sources $20\#$ with different sound sources 20^* removes the mutual exclusion between the newly received sound source 150_B and the sound sources $20\#$ rendered in the second zone 30_B . Rules may be defined that specify how such mutual exclusions are resolved, which sound sources are to be replaced, with what timing and in what order. These rules and decisions may be based on one or more different criteria. The replacement of the original sound sources $20\#$ with different sound sources 20^* may be done in a manner that smoothens the replacement for example by performing a fade-in and fade-out or by performing the replacement sequentially or otherwise controlling replacement of one sound source by another.

[0123] In some, but not necessarily all, examples, whether or not the sound sources of the first zone 30_A are replaced by sound sources of the second zone 30_B or the sound sources of the second zone 30_B are replaced by sound sources of the first zone 30_A is dependent upon

whether the first user 40_A or the second user 40_B initiated the real-time communication between the first user 40_A and the second user 40_B. For example, it may be determined that the first user that initiated the call has precedence and therefore the sound sources for that user should not be replaced whereas the sound sources of the second user should be replaced. The reverse is also a possibility.

[0124] In some examples alternative or an additional criterion or criteria may be used. In some but not necessarily all examples, the criterion or criteria include a condition based upon determining who will be the target of the adaptation. The criterion or criteria can in some examples be based upon an assessment of a difference between the content portions rendered to the first user and the a second user, for example, to determine if the they should be mutually exclusive.

[0125] The criterion or criteria can in some examples, additionally or alternatively, be based on a user request. For example, a first user may be able to make a request as to whether or not to share their context with a second user and therefore replace that second user's sound sources with the sound sources that are being rendered to them or to request that they share the context of the second user and have the sound sources that are currently rendered to them replaced by one or more sound sources being rendered to the second user.

[0126] There may be communication between the apparatus as used by the first user 40_A and the second user 40_B to determine whether and what sound sources should be replaced by different sound sources. In some examples, metadata associated with the common audio content 10 may define exclusive subsets of sound sources that should not be rendered simultaneously. The metadata may, in addition, define which sound sources should be dominant and should not be replaced and which sound sources should be replaced. The metadata may even define different conflicts between different sound sources and how each of those conflicts should be resolved, that is which sound source should be replaced and which sound source should not be replaced when there is mutual exclusion.

[0127] In some examples, the metadata may be defined by a content creator and associated with the common audio content 10. In other examples, the metadata may be generated using artificial intelligence or machine learning using a learning algorithm. The sound sources may, for example, be parameterized in relation to key analysis, instrumentation or other audio aspects or characteristics and these may be provided as input parameters to a machine learning algorithm such as a neural network. The neural network may be trained in advance or by the user so that it is able to detect the simultaneous rendering of sound sources that should be mutually exclusive before they are rendered. It is therefore possible to use this method not only to detect the possibility that mutually exclusive sound sources will be rendered but also to determine which sound source should be re-

placed.

[0128] It will be appreciated from the foregoing that it is not always necessary to replace sound sources. In circumstances in which there is no conflict, that is no mutual exclusion, the original sound sources may be maintained during the real-time communication between the first user and the second user.

[0129] In some, but not necessarily all, examples, the sound sources 150A, 150B that are communicated in real time between the first user 40_A and the second user 40_B are voice communications that are recorded at microphones. In some circumstances, the voice recordings captured by the microphones may be augmented with metadata that defines the context of the user whose voice is recorded at that time or which provides the context of the user at that time. In this example, context refers to the audio scene that is rendered to the user at that time and includes for example the point-of-view of the user (which then defines the sound scene) or somehow otherwise defines the arrangement of sound sources relative to the user. It would of course be possible to transmit the sound scene as rendered to the user in the transmitted sound source 150, however, as the audio content 10 is shared content it is more efficient to instead transmit the point-of-view 42 of the user 40 which defines the audio context.

[0130] The real-time duplex communication between the first user 40_A and the second user 40_B therefore enables a conversation between the first user and the second user which has voice only or which also includes the ambient context of the user, that is, the sound scene currently being rendered to the user while their voice is recorded.

[0131] FIG. 8 illustrates an example of the method 100 that illustrates a number of the concepts described above. The FIG illustrates the processes that occur at the first apparatus 200_A used by the first user 40_A and the processes that occur at a second apparatus 200_B used by the second user 40_B. It also illustrates the exchange of information between the first apparatus 200_A and the second apparatus 200_B. In this example, the method 100 is performed with respect to the first user 40_A and the first apparatus 200_A.

[0132] At block 102_A a first portion 10_A of a common audio content 10 is rendered to a first user 40_A by the first apparatus 200_A. The first portion 10_A of the common audio content 10 is selected by a current point-of-view 42_A of the first user 40_A.

[0133] At block 102_B, a second portion 10_B of the common audio content 10 is rendered to a second user 40_B by the second apparatus 200_B. The second portion 10_B of the audio content 10 is selected by a current point-of-view 42_B of the second user 40_B.

[0134] Next there is communication between the first apparatus 200_A and the second apparatus 200_B to set up a duplex real-time communication between the first user 40_A and the second user 40_B. The first apparatus 200_A informs 120 the second apparatus 200_B of the cur-

rent point-of-view 42_A of the first user 40_A and the second apparatus 200_B informs 122 the first apparatus 200_A of the point-of-view 42_B of the second user 40_B. The first apparatus 200_A therefore has knowledge of the current point-of-view 42_B of the second user 40_B and as a consequence what second adapted portion 10_B of the common audio content 10 is being rendered to the second user 40_B. Also, the second apparatus 200_B has knowledge of the current point-of-view 42_A of the first user 40_A and therefore also of what first portion 10_A of the common audio content 10 is being rendered to the first user 40_A.

[0135] The communications 120, 122 may also establish whether or not there is mutual exclusivity between the first portion 10_A and the second portion 10_B of the common audio content 10, and if there is, how this should be handled. At block 106_A it is determined whether or not to adapt the first portion 10_A of the common content. At block 106_B it is determined whether or not to adapt the second portion 10_B of the common audio content 10. In the example illustrated, the first apparatus 200_A, at block 106_A, determines that there is mutual exclusivity between the first portion 10_A of the common audio content 10 and the second portion 10_B of the common audio content 10. It determines that it should adapt the first portion 10_A of the common audio content 10 to produce, at block 106_A(2), an adapted first portion 10_A* of the common audio content 10 by replacing one or more sound sources of the first portion 10_A of the common audio content 10 with one or more different sound sources. The different sound sources may, as described above, originate from the second portion 10_B of the common audio content 10.

[0136] At block 106_B, the second apparatus 200_B determines that it is not necessary to adapt the second portion 10_B of the common audio content 10.

[0137] At block 104, real-time duplex communication between the first user 40_A and the second user 40_B is established. There is transmission 102, for rendering to the second user 40_B, of first audio 150_A generated by the first user 40_A. There is transmission 122, for rendering to the first user 40_A, of second audio 150_B generated by the second user 40_B.

[0138] The audio 150_A, 150_B that is exchanged may be a selectable common audio element. For example, it may be vocals that are recorded by microphones. For example, the first audio 150_A may be audio that has been recorded by a microphone at the first apparatus 200_A and likewise the second audio 150_B may be audio that has been recorded at a microphone of the second apparatus 200_B. The audio 150_A, 150_B is therefore in this example referred to as local user content (LUC).

[0139] In some, but not necessarily all, examples, it may be desirable to mix a respective user's voice recorded as audio 150_A, 150_B with the ambient audio content experienced by the respective user 40_A, 40_B at transmission, that is, with the audio scene rendered to the user. Alternatively, it may be sufficient to indicate that this mix should occur at rendering.

[0140] In this way, whole or part of the immersive audio

scene that is rendered to the first user 40_A can be provided to the second user 40_B and/or whole or part of the immersive audio scene rendered to the second user 40_B can be delivered to the first user 40_A for rendering to the first user 40_A. In the example illustrated, the immersive audio scene of the first user 40_A and the immersive audio scene of the second user 40_B are incompatible and mutually exclusive. In this example, the immersive audio scene of the second user is sent to the first user 40_A and in whole or in part replaces the immersive audio scene rendered to the first user 40_A.

[0141] Next at block 109_A the audio 150_B generated by the second user 40_B is rendered to the first user 40_A. The adapted first portion 10_A* of the common audio content 10 is also rendered 108 to the first user 40_A instead of the first portion 10_A of the common audio content 10. In this example the one or more different sound sources of the adapted first portion 10_A* of the common audio content 10 are rendered instead of one or more of the sound sources of the first portion 10_A of the common audio content 10.

[0142] At block 109_B, the audio 150_A generated by the first user 40_A is rendered to the second user 40_B and the second portion 10_B of the common audio content 10 is also rendered to the second user 40_B.

[0143] In the foregoing examples, the communication between two users, a first user 40_A and a second user 40_B. However in other examples more than two users may simultaneously communicate and the examples described and illustrated can be extended to this scenario.

[0144] FIG. 9 illustrates an example of a controller 210. Implementation of a controller 210 may be as controller circuitry. The controller 210 may be implemented in hardware alone, have certain aspects in software including firmware alone or can be a combination of hardware and software (including firmware).

[0145] As illustrated in FIG. 9 the controller 210 may be implemented using instructions that enable hardware functionality, for example, by using executable instructions of a computer program 206 in a general-purpose or special-purpose processor 202 that may be stored on a computer readable storage medium (disk, memory etc) to be executed by such a processor 202.

[0146] The processor 202 is configured to read from and write to the memory 204. The processor 202 may also comprise an output interface via which data and/or commands are output by the processor 202 and an input interface via which data and/or commands are input to the processor 202.

[0147] The memory 204 stores a computer program 206 comprising computer program instructions (computer program code) that controls the operation of the apparatus 200 when loaded into the processor 202. The computer program instructions, of the computer program 206, provide the logic and routines that enables the apparatus to perform the methods illustrated in FIGS 1 to 8. The processor 202 by reading the memory 204 is able to load and execute the computer program 206.

[0148] The apparatus 200 therefore comprises:

at least one processor 202; and
 at least one memory 204 including computer program code
 the at least one memory 204 and the computer program code configured to, with the at least one processor 202, cause the apparatus 200 at least to perform:

causing rendering of a portion of an audio content to a user, wherein the portion of the audio content is selected by a current point-of-view of the first user;

causing real-time communication between the first user and a second user by causing transmission, for rendering to the second user, of audio generated by the first user and by causing rendering, to the first user, of audio generated by the second user for rendering to the first user; causing adaptation of the portion of the audio content to create an adapted portion of the audio content, by replacing a sound source of the portion of the audio content with a different sound source; and

causing rendering, to the first user, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

[0149] As illustrated in FIG. 10, the computer program 206 may arrive at the apparatus 200 via any suitable delivery mechanism 220. The delivery mechanism 220 may be, for example, a machine readable medium, a computer-readable medium, a non-transitory computer-readable storage medium, a computer program product, a memory device, a record medium such as a Compact Disc Read-Only Memory (CD-ROM) or a Digital Versatile Disc (DVD) or a solid state memory, an article of manufacture that comprises or tangibly embodies the computer program 206. The delivery mechanism may be a signal configured to reliably transfer the computer program 206. The apparatus 200 may propagate or transmit the computer program 206 as a computer data signal.

[0150] Computer program instructions for causing an apparatus to perform at least the following or for performing at least the following:

causing rendering of a portion of an audio content to a user, wherein the portion of the audio content is selected by a current point-of-view of the first user; causing real-time communication between the first user and a second user by causing transmission, for rendering to the second user, of audio generated by the first user and by causing rendering, to the first user, of audio generated by the second user for rendering to the first user;

causing adaptation of the portion of the audio content to create an adapted portion of the audio content, by replacing a sound source of the portion of the audio content with a different sound source; and causing rendering, to the first user, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

[0151] The computer program instructions may be comprised in a computer program, a non-transitory computer readable medium, a computer program product, a machine readable medium. In some but not necessarily all examples, the computer program instructions may be distributed over more than one computer program.

[0152] Although the memory 204 is illustrated as a single component/circuitry it may be implemented as one or more separate components/circuitry some or all of which may be integrated/removable and/or may provide permanent/semi-permanent/ dynamic/cached storage.

[0153] Although the processor 202 is illustrated as a single component/circuitry it may be implemented as one or more separate components/circuitry some or all of which may be integrated/removable. The processor 202 may be a single core or multi-core processor.

[0154] References to 'computer-readable storage medium', 'computer program product', 'tangibly embodied computer program' etc. or a 'controller', 'computer', 'processor' etc. should be understood to encompass not only computers having different architectures such as single /multi- processor architectures and sequential (Von Neumann)/parallel architectures but also specialized circuits such as field-programmable gate arrays (FPGA), application specific circuits (ASIC), signal processing devices and other processing circuitry. References to computer program, instructions, code etc. should be understood to encompass software for a programmable processor or firmware such as, for example, the programmable content of a hardware device whether instructions for a processor, or configuration settings for a fixed-function device, gate array or programmable logic device etc.

[0155] As used in this application, the term 'circuitry' may refer to one or more or all of the following:

(a) hardware-only circuitry implementations (such as implementations in only analog and/or digital circuitry) and
 (b) combinations of hardware circuits and software, such as (as applicable):

(i) a combination of analog and/or digital hardware circuit(s) with software/firmware and
 (ii) any portions of hardware processor(s) with software (including digital signal processor(s)), software, and memory(ies) that work together to cause an apparatus, such as a mobile phone or server, to perform various functions and

(c) hardware circuit(s) and or processor(s), such as a microprocessor(s) or a portion of a microprocessor(s), that requires software (e.g. firmware) for operation, but the software may not be present when it is not needed for operation.

[0156] This definition of circuitry applies to all uses of this term in this application, including in any claims. As a further example, as used in this application, the term circuitry also covers an implementation of merely a hardware circuit or processor and its (or their) accompanying software and/or firmware. The term circuitry also covers, for example and if applicable to the particular claim element, a baseband integrated circuit for a mobile device or a similar integrated circuit in a server, a cellular network device, or other computing or network device.

[0157] The blocks illustrated in the FIGS 1 to 8 may represent steps in a method and/or sections of code in the computer program 206. The illustration of a particular order to the blocks does not necessarily imply that there is a required or preferred order for the blocks and the order and arrangement of the block may be varied. Furthermore, it may be possible for some blocks to be omitted.

[0158] Fig 11 illustrates an example of an apparatus 200. The apparatus 200 is configured to enable first person perspective mediated reality. For example, the apparatus may include circuitry 250 that is capable of tracking a user's point-of-view 42, for example, by tracking movement of a user's head while they are wearing the apparatus 200, as a head mounted apparatus, or are wearing a head-mounted tracking device communicating with the apparatus 200.

[0159] The head mounted device or apparatus may, in some but not necessarily all examples, include a head-mounted display for one or both eyes of the user 40.

[0160] The apparatus 200 comprises a decoder 252 for decoding the audio content 10. The decoding produces the audio content 10 in a format that can be used to identify and separately process sound sources 20. The decoded audio content 10 (spatial audio content) is provided to rendering control block 254 that performs the method 100.

[0161] The rendering control block 254 determines the portion or adapted portion of the audio content the audio content 10 that will be rendered. The rendering control block 254 is configured to enable first person perspective mediated reality with respect to the audio content 10 and takes into account the point-of-view 42 of the user 40. The rendering control block 254 is configured to identify and control each sound source 20 separately if required. It is capable of removing and adding one or more sound sources to/from a rendered sound scene.

[0162] In this example the rendering control block 254 and the renderer 256 are housed within the same apparatus 200, in other examples, the rendering control block 254 and the renderer 256 may be housed in separate devices.

The rendering control block 254 provides a control output to the renderer 256 which may be one or more loudspeakers, for example. The loudspeakers may be arranged around a user or have be part of a headset worn by the user.

[0163] In some of the preceding examples, the audio content 10 and the sound sources have been music based. However, this is not always the case. Other content is possible. The method 100 is particularly suitable when there is a conflict or potential conflict between the current sound scene rendered to a user and a sound source received by that user for rendering. The removal or replacement of sound objects in the sound scene 20 obviates the conflict.

[0164] In one example, each of the different zones 30 represents a different language. In one example, each of the different zones 30 represents a different age-restricted audio content.

[0165] Where a structural feature has been described, it may be replaced by means for performing one or more of the functions of the structural feature whether that function or those functions are explicitly or implicitly described.

[0166] In some but not necessarily all examples, the apparatus 200 is configured to communicate data from the apparatus 200 with or without local storage of the data in a memory 204 at the apparatus 200 and with or without local processing of the data by circuitry or processors at the apparatus 200.

[0167] The data may be stored in processed or unprocessed format remotely at one or more devices. The data may be stored in the Cloud.

[0168] The data may be processed remotely at one or more devices. The data may be partially processed locally and partially processed remotely at one or more devices.

[0169] The data may be communicated to the remote devices wirelessly via short range radio communications such as Wi-Fi or Bluetooth, for example, or over long range cellular radio links. The apparatus may comprise a communications interface such as, for example, a radio transceiver for communication of data.

[0170] The apparatus 200 may be part of the Internet of Things forming part of a larger, distributed network.

[0171] The processing of the data, whether local or remote, may be for the purpose of health monitoring, data aggregation, patient monitoring, vital signs monitoring or other purposes.

[0172] The processing of the data, whether local or remote, may involve artificial intelligence or machine learning algorithms. The data may, for example, be used as learning input to train a machine learning network or may be used as a query input to a machine learning network, which provides a response. The machine learning network may for example use linear regression, logistic regression, vector support machines or an acyclic machine learning network such as a single or multi hidden layer neural network.

[0173] The processing of the data, whether local or remote, may produce an output. The output may be communicated to the apparatus 200 where it may produce an output sensible to the subject such as an audio output, visual output or haptic output.

[0174] The systems, apparatus, methods and computer programs may use machine learning which can include statistical learning. Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. The computer learns from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E. The computer can often learn from prior training data to make predictions on future data. Machine learning includes wholly or partially supervised learning and wholly or partially unsupervised learning. It may enable discrete outputs (for example classification, clustering) and continuous outputs (for example regression). Machine learning may for example be implemented using different approaches such as cost function minimization, artificial neural networks, support vector machines and Bayesian networks for example. Cost function minimization may, for example, be used in linear and polynomial regression and K-means clustering. Artificial neural networks, for example with one or more hidden layers, model complex relationship between input vectors and output vectors. Support vector machines may be used for supervised learning. A Bayesian network is a directed acyclic graph that represents the conditional independence of a number of random variables.

[0175] The above described examples find application as enabling components of: automotive systems; telecommunication systems; electronic systems including consumer electronic products; distributed computing systems; media systems for generating or rendering media content including audio, visual and audio visual content and mixed, mediated, virtual and/or augmented reality; personal systems including personal health systems or personal fitness systems; navigation systems; user interfaces also known as human machine interfaces; networks including cellular, non-cellular, and optical networks; ad-hoc networks; the internet; the internet of things; virtualized networks; and related software and services.

The term 'comprise' is used in this document with an inclusive not an exclusive meaning. That is any reference to X comprising Y indicates that X may comprise only one Y or may comprise more than one Y. If it is intended to use 'comprise' with an exclusive meaning then it will be made clear in the context by referring to "comprising only one.." or by using "consisting".

[0176] In this description, reference has been made to various examples. The description of features or functions in relation to an example indicates that those features or functions are present in that example. The use of the term 'example' or 'for example' or 'can' or 'may' in the text denotes, whether explicitly stated or not, that

such features or functions are present in at least the described example, whether described as an example or not, and that they can be, but are not necessarily, present in some of or all other examples. Thus 'example', 'for example', 'can' or 'may' refers to a particular instance in a class of examples. A property of the instance can be a property of only that instance or a property of the class or a property of a sub-class of the class that includes some but not all of the instances in the class. It is therefore implicitly disclosed that a feature described with reference to one example but not with reference to another example, can where possible be used in that other example as part of a working combination but does not necessarily have to be used in that other example.

[0177] Although embodiments have been described in the preceding paragraphs with reference to various examples, it should be appreciated that modifications to the examples given can be made without departing from the scope of the claims.

[0178] Features described in the preceding description may be used in combinations other than the combinations explicitly described above.

[0179] Although functions have been described with reference to certain features, those functions may be performable by other features whether described or not.

[0180] Although features have been described with reference to certain embodiments, those features may also be present in other embodiments whether described or not.

[0181] The term 'a' or 'the' is used in this document with an inclusive not an exclusive meaning. That is any reference to X comprising a/the Y indicates that X may comprise only one Y or may comprise more than one Y unless the context clearly indicates the contrary. If it is intended to use 'a' or 'the' with an exclusive meaning then it will be made clear in the context. In some circumstances the use of 'at least one' or 'one or more' may be used to emphasis an inclusive meaning but the absence of these terms should not be taken to infer and exclusive meaning.

[0182] The presence of a feature (or combination of features) in a claim is a reference to that feature or (combination of features) itself and also to features that achieve substantially the same technical effect (equivalent features). The equivalent features include, for example, features that are variants and achieve substantially the same result in substantially the same way. The equivalent features include, for example, features that perform substantially the same function, in substantially the same way to achieve substantially the same result.

[0183] In this description, reference has been made to various examples using adjectives or adjectival phrases to describe characteristics of the examples. Such a description of a characteristic in relation to an example indicates that the characteristic is present in some examples exactly as described and is present in other examples substantially as described.

[0184] Whilst endeavoring in the foregoing specifica-

tion to draw attention to those features believed to be of importance it should be understood that the Applicant may seek protection via the claims in respect of any patentable feature or combination of features hereinbefore referred to and/or shown in the drawings whether or not emphasis has been placed thereon.

Claims

1. An apparatus comprising means for:

causing rendering of a portion of an audio content to a first user, wherein the portion of the audio content is selected based at least in part on a current point-of-view of the first user;
causing real-time communication between the first user and a second user comprising causing transmission, for rendering to the second user, of audio and causing rendering, to the first user, of audio from the second user for rendering to the first user;
causing adaptation of the portion of the audio content to create an adapted portion of the audio content, comprising replacing a sound source of the portion of the audio content with a different sound source; and
causing rendering, to the first user, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source.

2. An apparatus as claimed in claim 1, wherein the different sound source originates from a second portion of the audio content, different to the portion of the audio content, wherein the second portion of the audio content is selected based at least in part on a current point-of-view of the second user.

3. An apparatus as claimed in claim 1 or 2, wherein causing adaptation of the portion of the audio content to create an adapted portion of the audio content comprises replacing multiple sound sources with different sound sources.

4. An apparatus as claimed in claim 3 wherein the multiple sound sources are replaced with the different sound sources one-at-a-time and wherein the adapted portion of the audio content is rendered to the user while the one-at-a-time adaptation is on-going, wherein progressively more of the different sound sources are rendered instead of the multiple sound sources.

5. An apparatus as claimed in any preceding claim, wherein when the first user is in a first zone of a plurality of zones and the second user is in a second,

different zone of the plurality of zones, wherein the portion of the audio content rendered to the user depends on the point-of-view of the user in the first zone and includes sound sources associated with the first zone and does not include any sound source associated with the second zone,
, wherein the content rendered to the second user depends on the point-of-view of the second user in the second zone and includes sound sources associated with the second zone,
and wherein the adapted portion of the audio content rendered to the first user depends on the point-of-view of the first user in the first zone and includes at least one sound source associated with the second zone.

6. An apparatus as claimed in any preceding claim comprising means for causing an undoing of some or all of the adaptation performed on the portion of the audio content to create the adapted portion of the audio content and/or causing rendering, to the first user, of the portion of the audio content instead of the portion of the adapted content, wherein the sound source is rendered instead of the different sound source.

7. An apparatus as claimed in claim 6, wherein the undoing of some or all of the adaptation performed on the portion of the audio content to create the adapted portion of the audio content is performed based at least in part on a change in point-of-view of the first user.

8. An apparatus as claimed in any preceding claim, wherein the portion of audio content rendered to the first user, is defined by a point-of-view of a virtual user in a virtual space, which is determined based at least in part on a point-of-view of the first user in a real space.

9. An apparatus as claimed in any preceding claim, wherein the point-of-view of the first user is determined based at least in part on an orientation of the first user or wherein the point-of-view of the first user is determined based at least in part on an orientation and a location of the user.

10. An apparatus as claimed in any preceding claim configured as a head mounted apparatus.

11. An apparatus as claimed in any preceding claim comprising means for causing adaptation of the portion of the audio content to create adapted content based at least in part on an initiation of the real-time communication and an additional criterion or criteria.

12. An apparatus as claimed in claim 11, wherein the criterion or criteria include a condition based upon

determining who will be the target of the adaptation.

13. An apparatus as claimed in claim 11 or 12, wherein the criterion or criteria are based upon an assessment of a difference between the content portions rendered to the first user and the second user. 5

14. A method comprising:

rendering a portion of an audio content to a first user, wherein the portion of the audio content is selected based at least in part on a current point-of-view of the first user; 10
 enabling real-time communication between the first user and a second user comprising transmitting, to the second user audio and rendering, to the first user, audio received from the second user; 15
 adapting the portion of the audio content to create an adapted portion of the audio content, comprising replacing a sound source of the portion of the audio content with a different sound source; and 20
 rendering, to the first user, the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source. 25

15. A computer program that, when run on a computer, performs: 30

causing rendering of a portion of an audio content to a first user, wherein the portion of the audio content is selected based at least in part on a current point-of-view of the first user; 35
 causing real-time communication between the first user and a second user comprising causing transmission, for rendering to the second user, of audio generated by the first user and causing rendering, to the first user, of audio generated by the second user for rendering to the first user; 40
 causing adaptation of the portion of the audio content to create an adapted portion of the audio content, comprising replacing a sound source of the portion of the audio content with a different sound source; and 45
 causing rendering, to the first user, of the adapted portion of the audio content instead of the portion of the audio content, wherein the different sound source is rendered instead of the sound source. 50

55

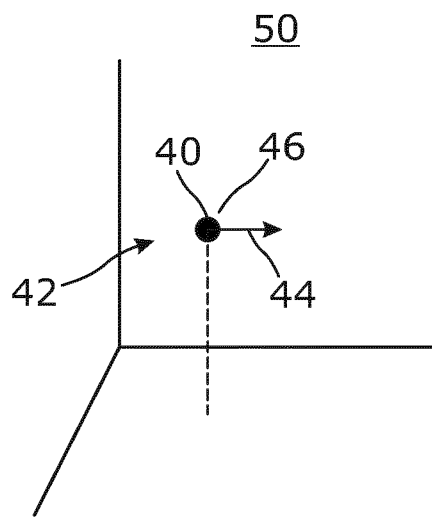


FIG. 1A

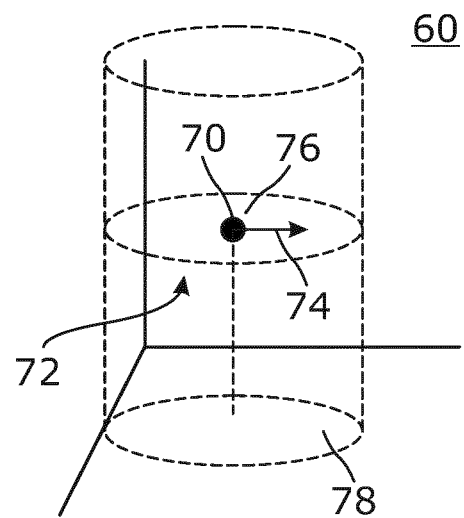


FIG. 1C

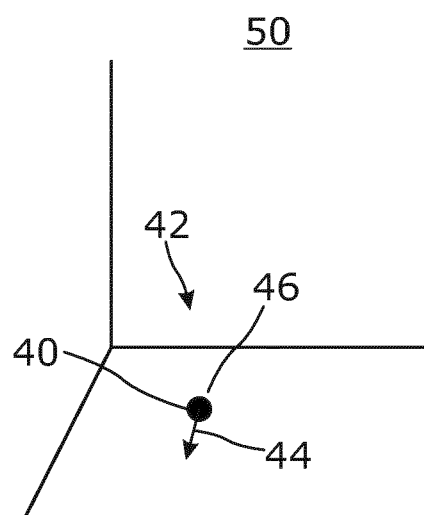


FIG. 1B

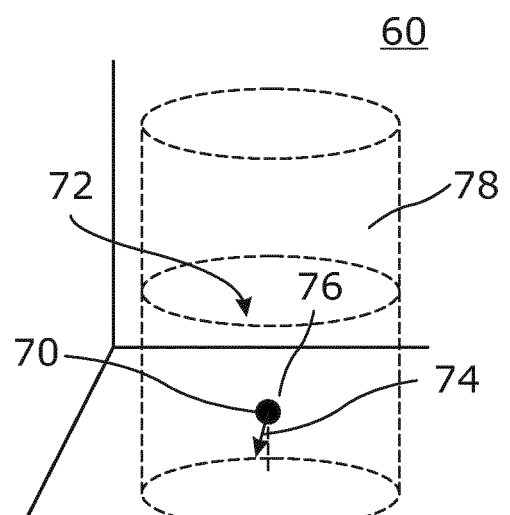


FIG. 1D

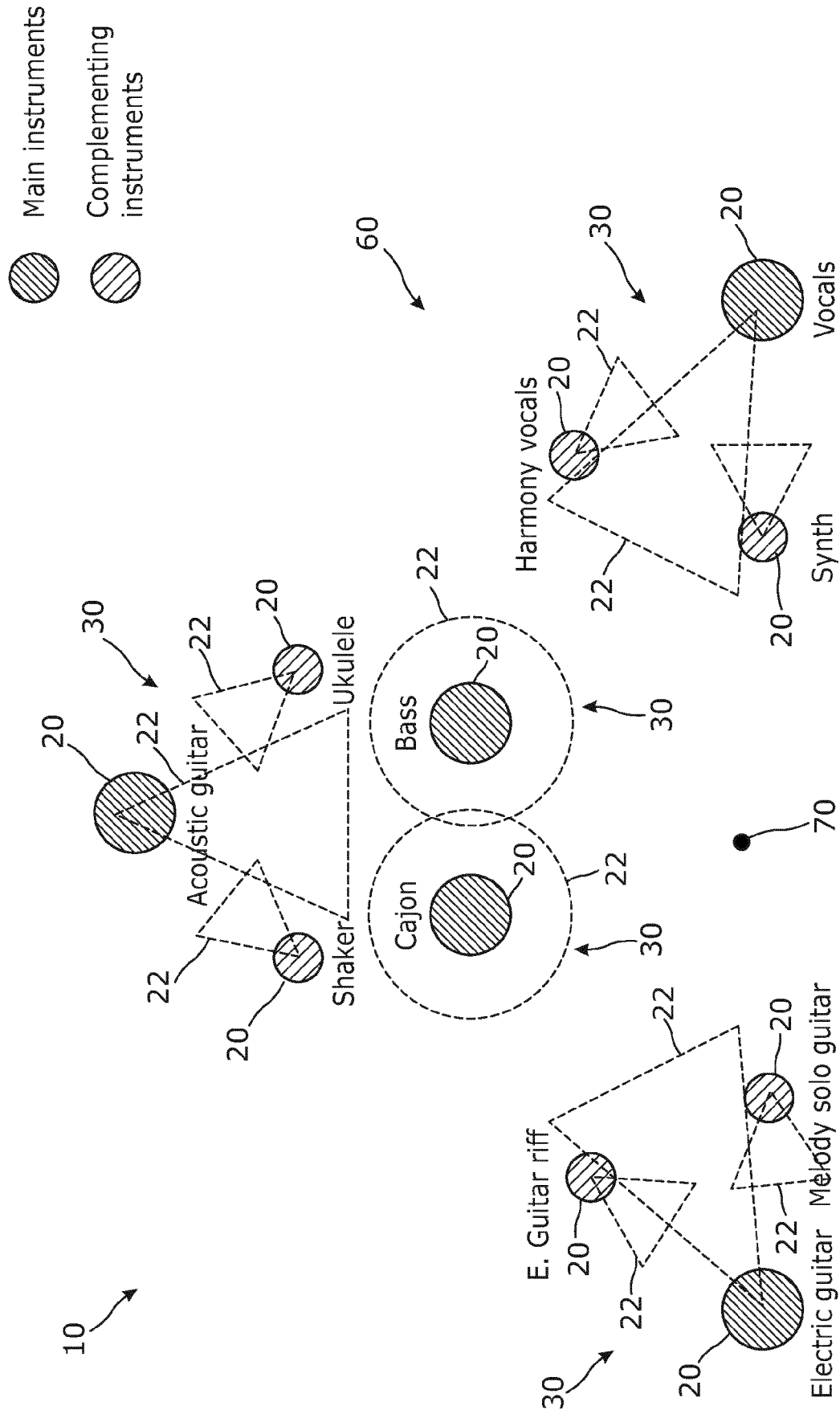


FIG. 2

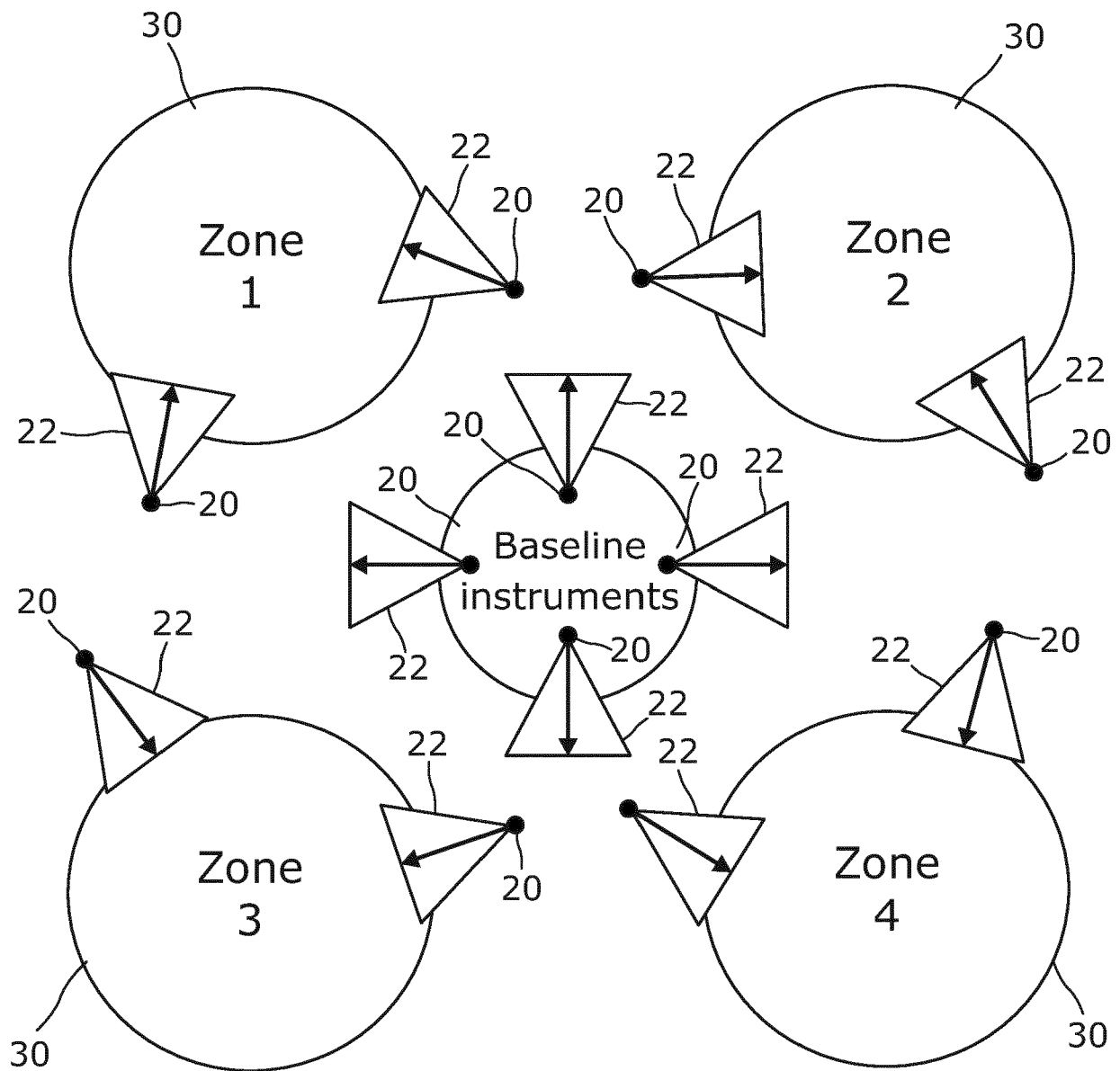


FIG. 3

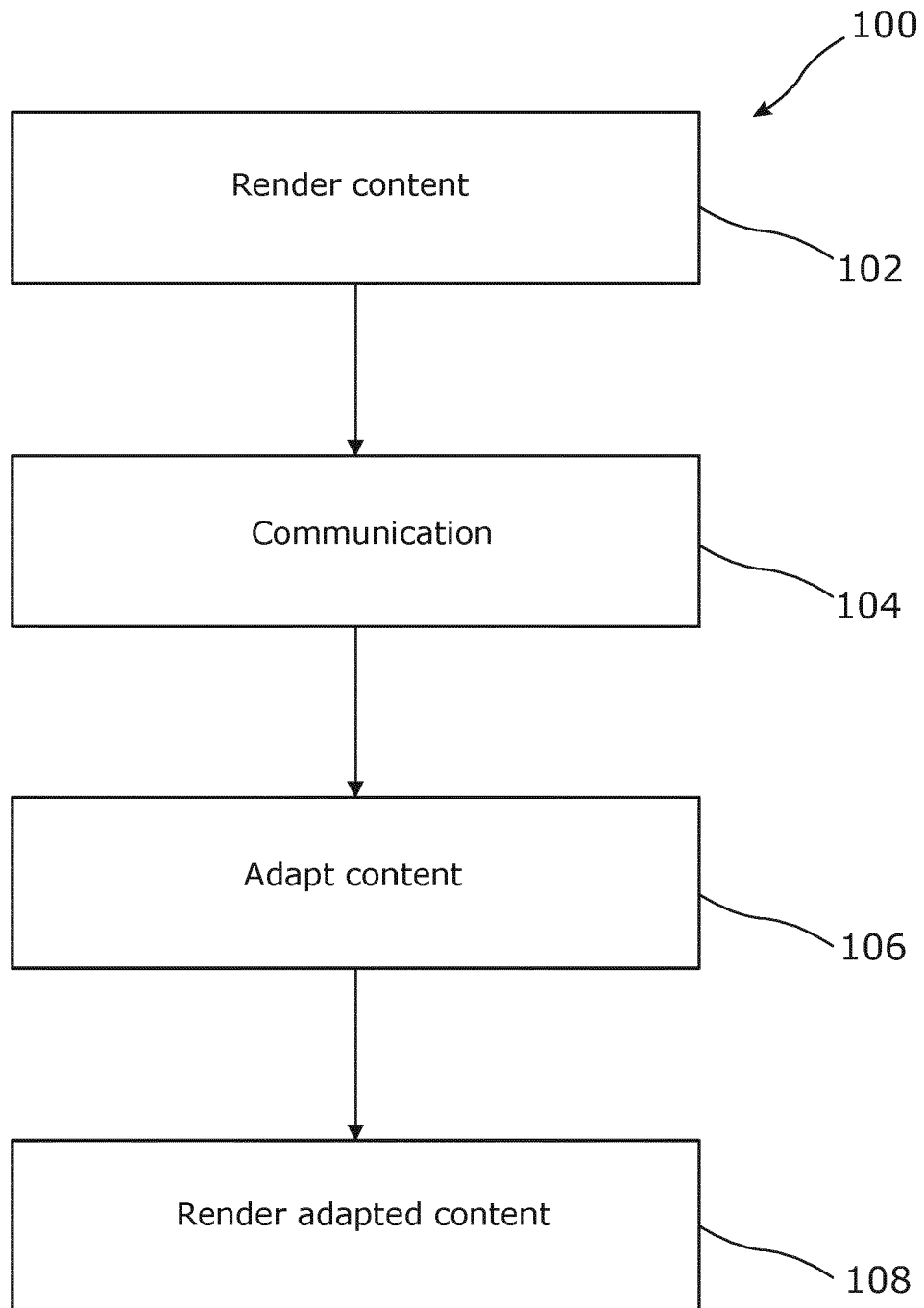


FIG. 4

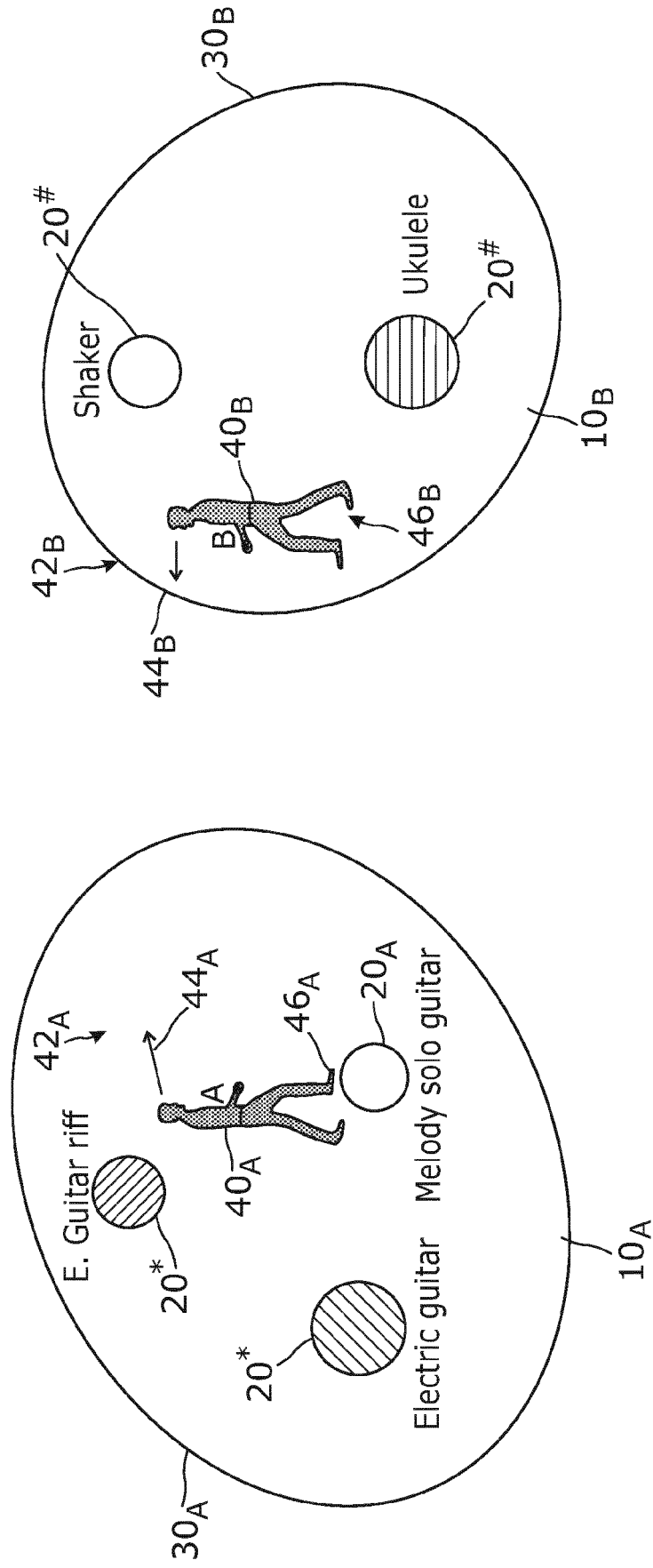


FIG. 5B

FIG. 5A

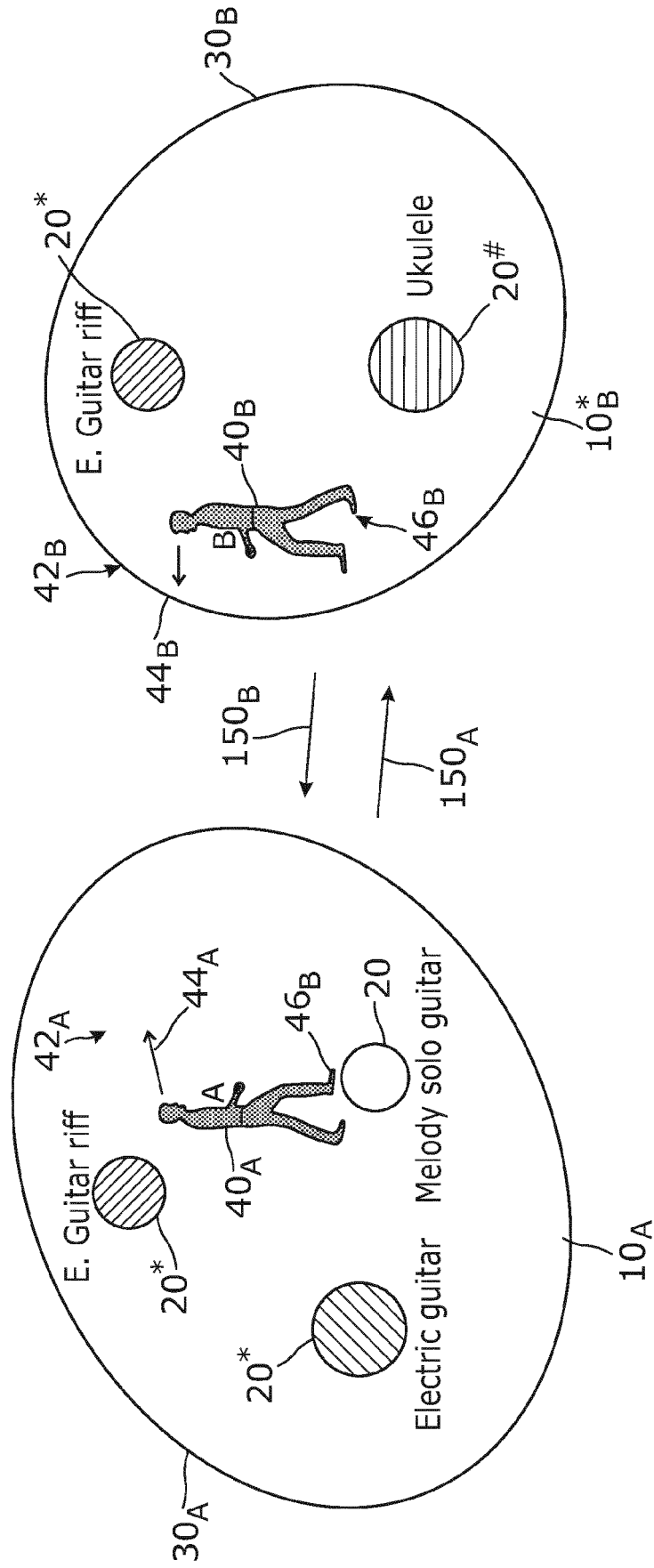
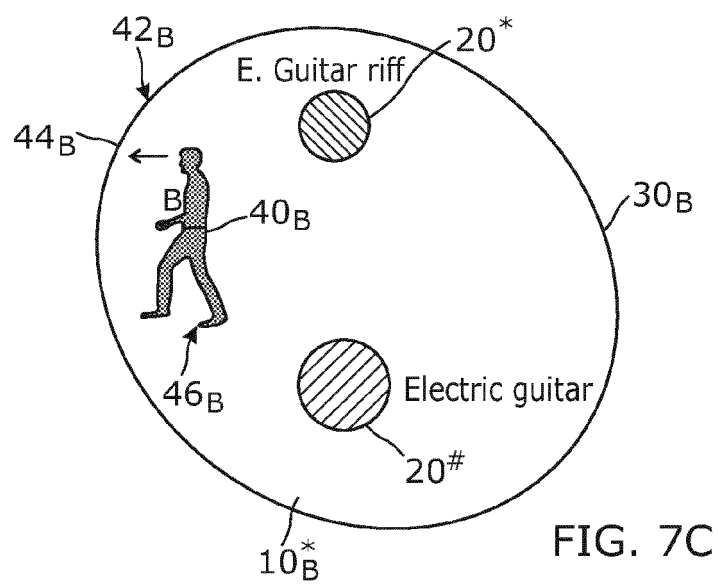
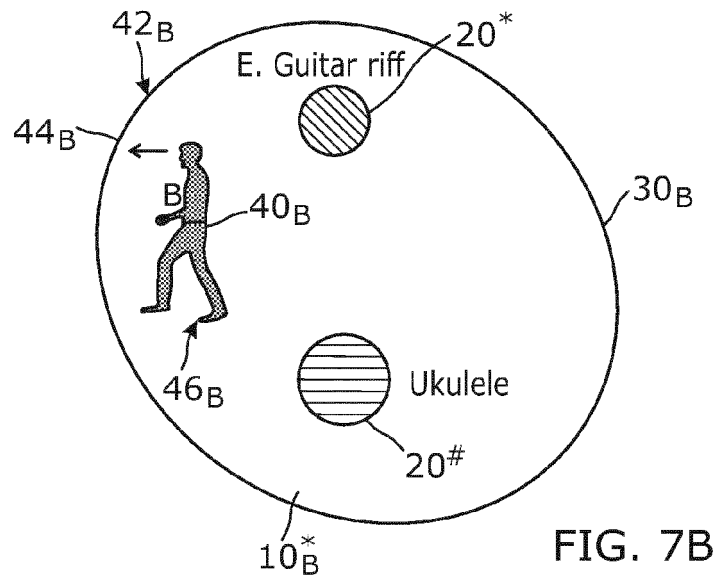
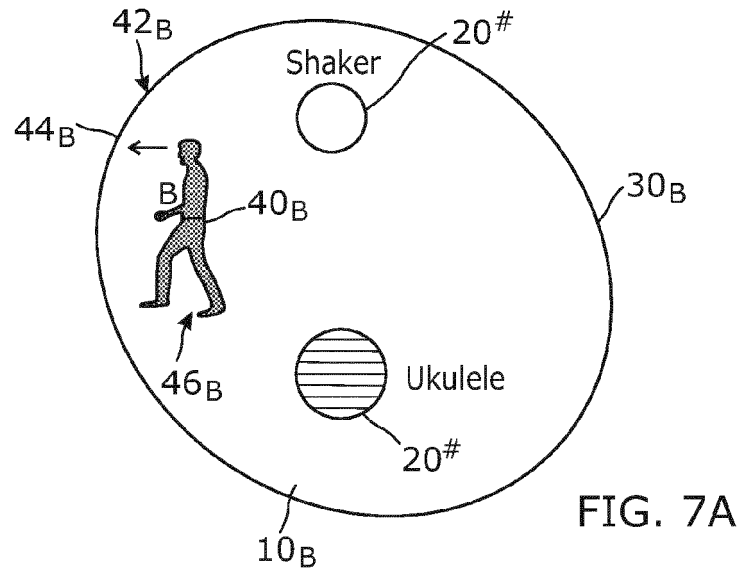


FIG. 6B

FIG. 6A



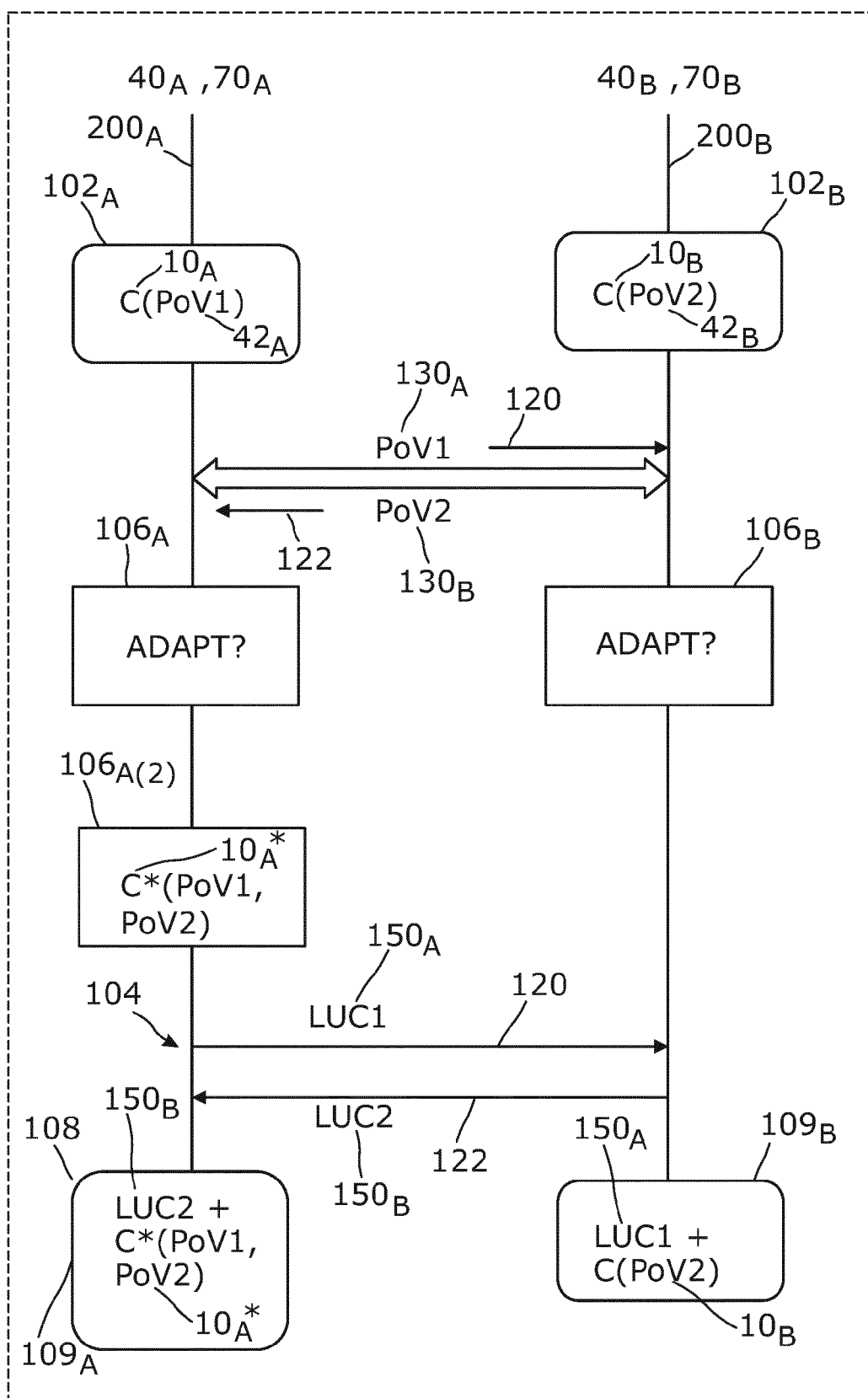


FIG. 8

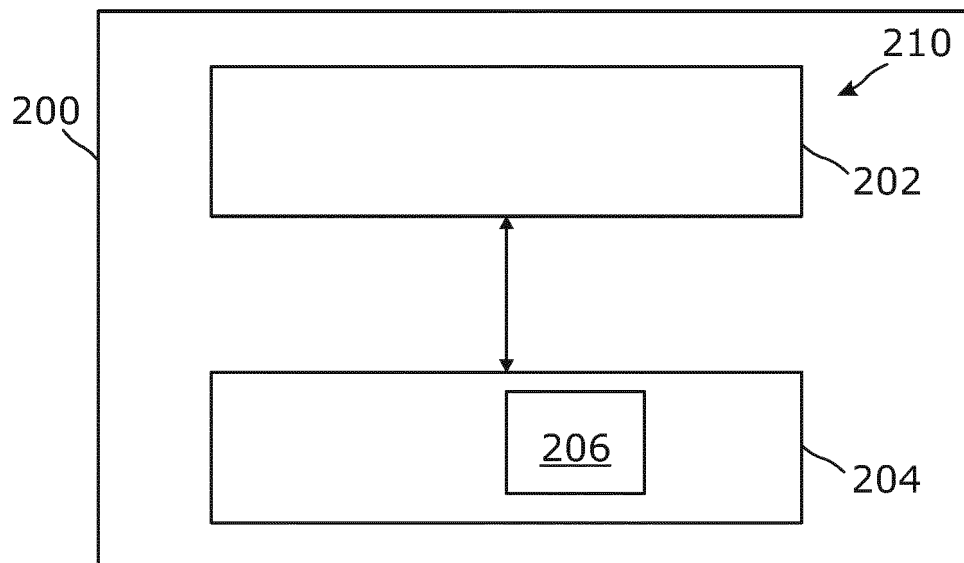


FIG. 9

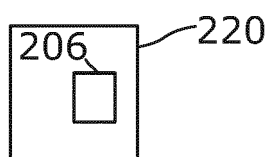


FIG. 10

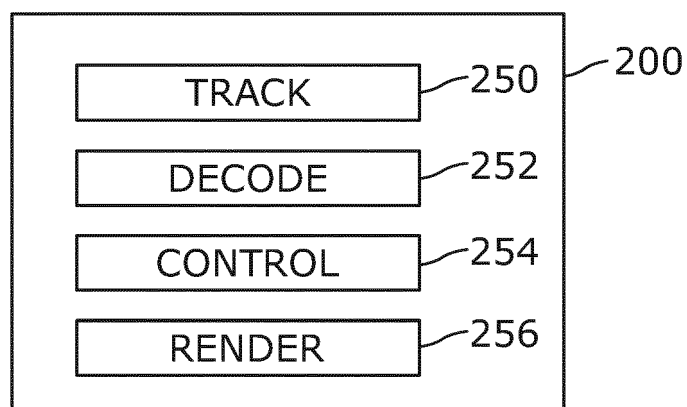


FIG. 11



EUROPEAN SEARCH REPORT

Application Number
EP 19 15 5151

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2018/270601 A1 (LAAKSONEN LASSE JUHANI [FI]) 20 September 2018 (2018-09-20) * paragraph [0024] - paragraph [0048] * -----	1-15	INV. H04S7/00
			TECHNICAL FIELDS SEARCHED (IPC)
			H04S
The present search report has been drawn up for all claims			
Place of search The Hague		Date of completion of the search 14 March 2019	Examiner Pigniez, Thierry
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

 2
EPO FORM 1503 03/82 (P04C01)

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

EP 19 15 5151

5

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

14-03-2019

10

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2018270601 A1	20-09-2018	US 2018270601 A1	20-09-2018
		WO 2018167363 A1	20-09-2018

15

20

25

30

35

40

45

50

55

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82