



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
27.03.2019 Bulletin 2019/13

(51) Int Cl.:
H04S 7/00 (2006.01) G06F 3/01 (2006.01)

(21) Application number: **17192113.3**

(22) Date of filing: **20.09.2017**

(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:
MA MD

- **LAAKSONEN, Lasse**
33210 Tampere (FI)
- **LEHTINIEMI, Arto**
33880 Lempäälä (FI)
- **CRICRI, Francesco**
33100 Tampere (FI)

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

(74) Representative: **Nokia EPO representatives**
Nokia Technologies Oy
Karaportti 3
02610 Espoo (FI)

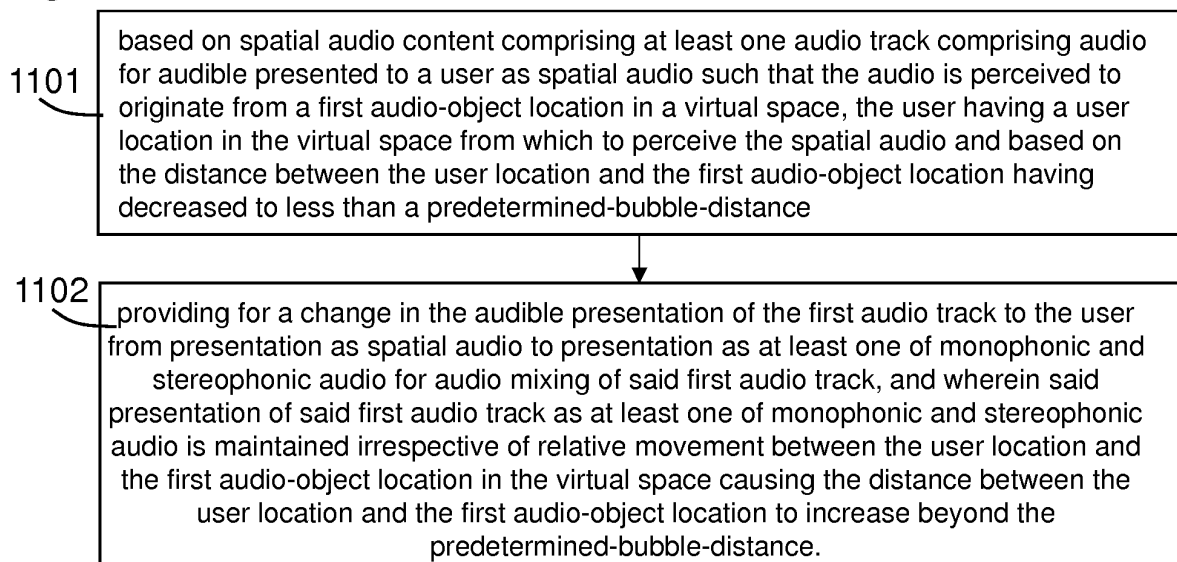
(72) Inventors:
• **MATE, Sujeet Shyamsundar**
33720 Tampere (FI)

(54) **AN APPARATUS AND ASSOCIATED METHODS FOR AUDIO PRESENTED AS SPATIAL AUDIO**

(57) An apparatus which, based on spatial audio content comprising at least one audio track comprising audio for audible presentation as spatial audio perceived to originate from a first audio-object location in a virtual space relative to a user location, and based on the distance between the user location and the first audio-object location having decreased to less than a predetermined-bubble-distance; is configured to change the au-

dible presentation of the first audio track to the user from spatial audio to monophonic/stereophonic audio for audio mixing of said first audio track, and wherein said presentation as monophonic/stereophonic audio is maintained irrespective of relative movement causing the distance between the user location and the first audio-object location to increase beyond the predetermined-bubble-distance.

Figure 11



Description

Technical Field

[0001] The present disclosure relates to the field of spatial audio and, in particular, to the field of providing for audio mixing of spatial audio in a virtual space, associated methods, computer programs and apparatus.

Background

[0002] The augmentation of real-world environments with graphics and audio is becoming common, with augmented/virtual reality content creators providing more and more content for augmentation of the real-world as well as for virtual environments. The presentation of audio as spatial audio, which is such that the audio is perceived to originate from a particular location in space, is useful for creating realistic augmented reality environments and virtual reality environments. The effective and efficient control of the presentation of spatial audio may be challenging.

[0003] The listing or discussion of a prior-published document or any background in this specification should not necessarily be taken as an acknowledgement that the document or background is part of the state of the art or is common general knowledge. One or more aspects/examples of the present disclosure may or may not address one or more of the background issues.

Summary

[0004] In a first example aspect 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 to perform at least the following:

based on spatial audio content comprising at least one audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a first audio-object location in a virtual space relative to a user location of the user in the virtual space, and based on the distance between the user location and the first audio-object location having decreased to less than a predetermined-bubble-distance;

provide for a change in the audible presentation of the first audio track to the user from presentation as spatial audio to presentation as at least one of monophonic and stereophonic audio for audio mixing of said first audio track, and wherein said presentation of said first audio track as

at least one of monophonic and stereophonic audio is maintained irrespective of relative movement between the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond the predetermined-bubble-distance.

[0005] In one or more examples, based on the relative movement between the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond a predetermined-stretched-bubble-distance which is greater than the predetermined-bubble-distance;

provide for a change in the audible presentation of the first audio track to the user from presentation as at least one of monophonic and stereophonic audio to presentation as spatial audio such that the audio of the first audio track is perceived from a direction based on the first audio-object location relative to the user location.

[0006] Accordingly, in one or more examples, relative movement between the audio-object location and the user location while the distance remains below the predetermined-stretched-bubble-distance does not affect the audible presentation of the first audio track in terms of one or more of the direction from which the audio of the first audio track is heard, the volume of the audio of the first audio track, application of a Room Impulse Response function to the first audio track, and the degree to which the audio of the first audio track is presented to left and right channels of an audio presentation device configured to provide the audio presented on the channels to left and right ears of the user.

[0007] In one or more examples, one or both of:

- i) user initiated movement of their user location in the virtual space; and
- ii) movement of the audio-object in the virtual space;

provides for the relative movement.

[0008] In one or more examples, the spatial audio content comprises a plurality of audio tracks including at least the first audio track and a second audio track, the second audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a second audio-object location in the virtual space; and

based on the presentation of said first audio track as at least one of monophonic and stereophonic audio, provide for audible presentation of the second audio track as spatial audio such that the audio of the second audio track is perceived from a direction based on the second audio-object location relative to the

user location.

[0009] In one or more examples, the apparatus is caused to provide for audible presentation of the spatial audio content and simultaneous audible presentation of ambience audio content, the ambience audio content comprising audio for audible presentation that is not perceived to originate from a particular direction in the virtual space. In one or more examples, the ambience audio content presented to the user is a function of location in the virtual space, the apparatus caused to provide for one or more of:

- i) said simultaneous audible presentation of ambience audio content based on the user-location;
- ii) said simultaneous audible presentation of ambience audio content based on the first audio object location when providing for presentation of the audio the first audio object monophonically or stereophonically;
- iii) ceasing of presentation of ambience audio content. when providing for presentation of the audio the first audio object monophonically or stereophonically.

[0010] In one or more examples, the spatial audio content comprises part of virtual reality content for viewing in virtual reality, the virtual reality content comprising visual content for display in the three-dimensional virtual space and the one or more audio tracks of the spatial audio content configured for presentation from one or more respective audio-object locations, at least a subset of the audio-object locations corresponding to features in the visual content.

[0011] In one or more examples, the spatial audio content comprises augmented reality content, the virtual space corresponding to a real-world space in which the user is located such that a location of the user in the real-world space corresponds to the user location in the virtual space and the audio-object location in the virtual space corresponds to a real-world-audio-object location in the real-world space.

[0012] In one or more examples, the spatial audio content comprises mixed reality content, the virtual space corresponding to a real-world space in which the user is located such that a location of the user in the real-world space corresponds to the user location in the virtual space and the audio-object location in the virtual space corresponds to a real-world-audio-object location in the real-world space.

[0013] In one or more examples, with the first audio track audibly presented as one of monophonic and stereophonic audio, based on user input indicative of a desire to return to spatial audio presentation of the first audio track, provide for a change in the audible presentation of the first audio track to the user from presentation as at least one of monophonic and stereophonic audio to presentation as spatial audio such that the audio of the first

audio track is perceived from a direction based on the first audio-object location relative to the user location.

[0014] In one or more examples, the direction is based on the first audio-object location at the time of said user input or a current first audio-object location.

[0015] In one or more examples, in addition to the apparatus being caused to provide for a change in the audible presentation of the first audio track, the apparatus is caused to provide an audio mixing user interface for modification of one or more audio parameters of the first audio track.

[0016] In one or more examples, the one or more audio parameters comprise at least one or more of: volume, bass level, mid-tone level, treble-level, reverberation level and echo.

[0017] In one or more examples, based on receipt of user input to the audio mixing user interface causing changes to one or more audio parameters of the first audio track and, subsequently, at least one of:

- i) relative movement between the user location and the audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond a predetermined-stretched-bubble-distance which is greater than the predetermined-bubble-distance;
- ii) user input indicative of a desire to return to spatial audio presentation of the first audio track;

provide for one of:

- a) discarding of the changes to one or more audio parameters of the first audio track unless a user initiated save input is received; and
- b) a change in the audible presentation of the first audio track to presentation as spatial audio such that the audio of the first audio track is perceived from a direction based on the first audio-object location relative to the user location with the changes to the one or more audio parameters applied.

[0018] In one or more examples, between the audible presentation of the first audio track to the user as at least one of monophonic and stereophonic audio and the presentation of the first audio track as spatial audio such that the audio of the first audio track is perceived from a direction based on the first audio-object location relative to the user location, the apparatus is caused to provide for audible presentation of the first audio track with a transitional spatial audio effect comprising the perceived origin of the audio of first audio track progressively moving away from the user location to the current first audio-object location.

[0019] In one or more examples, based on the relative movement between the user location and the audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase to within a threshold of the predetermined-

stretched-bubble-distance, provide for audible presentation of the first audio track to the user as at least one of monophonic and stereophonic audio with an audio effect to thereby audibly indicate that the user is approaching the predetermined-stretched-bubble-distance.

[0020] In one or more examples, when presenting one of a plurality of audio tracks of the spatial audio content monophonically or stereophonically with the other audio tracks presented as spatial audio, the apparatus is caused to apply a room impulse response function to at least one of said other audio tracks, the room impulse response function configured to modify the audio of the at least one other audio tracks to sound as if it is heard in a predetermined room with a particular location in said predetermined room, the particular location in said predetermined room based on either:

- i) the user location; and
- ii) a current location of the first audio object.

[0021] In one or more examples, the apparatus is caused to provide for user-selection of one of:

- a) the Room Impulse Response function based on the user location; and
- b) the Room Impulse Response function based on the current location of the first audio object.

[0022] In one or more examples, the spatial audio content is audibly presented as spatial audio by processing the first audio track using one or more of:

- i) a head-related-transfer-function filtering technique;
- ii) a vector-base-amplitude panning technique; and
- iii) binaural audio presentation.

[0023] In one or more examples, when the first audio track is presented as spatial audio, signalling indicative of movement of the user provides for modification of one or more of the direction from which the audio track is perceived to originate relative to the user's head and its volume; and

when the first audio track is presented as at least one of monophonic and stereophonic audio, signalling indicative of movement of the user does not provide for modification of one or more of the direction from which the audio track is perceived to originate relative to the user's head and its volume.

[0024] In a further aspect there is provided a method, the method comprising:

based on spatial audio content comprising at least one audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a first audio-object lo-

cation in a virtual space relative to a user location of the user in the virtual space, and based on the distance between the user location and the first audio-object location having decreased to less than a predetermined-bubble-distance;

providing for a change in the audible presentation of the first audio track to the user from presentation as spatial audio to presentation as at least one of monophonic and stereophonic audio for audio mixing of said first audio track, and wherein said presentation of said first audio track as at least one of monophonic and stereophonic audio is maintained irrespective of relative movement between the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond the predetermined-bubble-distance.

[0025] In a further aspect there is provided a computer readable medium comprising computer program code stored thereon, the computer readable medium and computer program code being configured to, when run on at least one processor, perform the method of:

based on spatial audio content comprising at least one audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a first audio-object location in a virtual space relative to a user location of the user in the virtual space, and based on the distance between the user location and the first audio-object location having decreased to less than a predetermined-bubble-distance;

providing for a change in the audible presentation of the first audio track to the user from presentation as spatial audio to presentation as at least one of monophonic and stereophonic audio for audio mixing of said first audio track, and wherein said presentation of said first audio track as at least one of monophonic and stereophonic audio is maintained irrespective of relative movement between the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond the predetermined-bubble-distance.

[0026] In a further aspect there is provided an apparatus, the apparatus comprising means configured to;

based on spatial audio content comprising at least one audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a first audio-object location in a virtual space relative to a user location of the user in the virtual space, and based on the distance between the user location and the first audio-object location having decreased to less than a predetermined-bubble-distance;

provide for a change in the audible presentation of the first audio track to the user from presentation as spatial audio to presentation as at least one of monophonic and stereophonic audio for audio mixing of said first audio track, and wherein said presentation of said first audio track as at least one of monophonic and stereophonic audio is maintained irrespective of relative movement between the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond the predetermined-bubble-distance.

[0027] The present disclosure includes one or more corresponding aspects, examples or features in isolation or in various combinations whether or not specifically stated (including claimed) in that combination or in isolation. Corresponding means and corresponding functional units (e.g., function enabler, AR/VR graphic renderer, display device) for performing one or more of the discussed functions are also within the present disclosure.

[0028] Corresponding computer programs for implementing one or more of the methods disclosed are also within the present disclosure and encompassed by one or more of the described examples.

[0029] The above summary is intended to be merely exemplary and non-limiting.

Brief Description of the Figures

[0030] A description is now given, by way of example only, with reference to the accompanying drawings, in which:

figure 1 illustrates an example apparatus for providing for a change in the audible presentation of an audio track;

figure 2 shows an example virtual space showing the location of a user and five audio objects each having an audio track associated therewith and the location relative to the user illustrating where in the virtual space the user perceives the audio tracks as originating;

figure 3 shows the example virtual space as figure 2 with a bubble around each audio object illustrating the predetermined-bubble-distance;

figure 4 shows the user approaching the predetermined-bubble-distance of one of the audio objects; figure 5 shows the user having moved such that the distance between the user's location and the one of the audio-object locations has decreased to less than the predetermined-bubble-distance and the apparatus has provided for a change in the audible presentation of the first audio track to the user;

figure 6 shows the situation illustrated in figure 5 except the audio object has moved in the virtual space relative to the user;

figure 7 shows the situation illustrated in figure 5 except the user has moved relative to the audio object in the virtual space;

figure 8 shows an example application of room impulse response processing to the audio track;

figure 9 shows an example in which relative movement of the user and the audio object away from one another has caused the distance between them to approach a predetermined-stretched-bubble-distance;

figure 10 shows an example in which the user and audio object have moved apart beyond the stretched-bubble-distance and the apparatus provides for a change in presentation of the first audio track back to externalized (relative to the user's head) spatial audio rather than non-externalized monophonic/stereophonic audio;

figure 11 shows a flowchart illustrating an example method; and

figure 12 shows a computer readable medium.

Description of Example Aspects

[0031] Virtual reality (VR) may use a VR display comprising a headset, such as glasses or goggles or virtual retinal display, or one or more display screens that surround a user to provide the user with an immersive virtual experience. A virtual reality apparatus, which may or may not include the VR display, may provide for presentation of multimedia VR content representative of a virtual reality scene to a user to simulate the user being present within the virtual reality scene. Accordingly, in one or more examples, the VR apparatus may provide signalling to a VR display for display of the VR content to a user while in one or more other examples, the VR apparatus may be part of the VR display, e.g. part of the headset. The virtual reality scene may therefore comprise the VR content displayed within a three-dimensional virtual reality space so that the user feels immersed in the scene, as if they were there, and may look around the VR space at the VR content displayed around them. The virtual reality scene may replicate a real world scene to simulate the user being physically present at a real world location or the virtual reality scene may be computer generated or a combination of computer generated and real world multimedia content. Thus, the VR content may be considered to comprise the imagery (e.g. static or video imagery), audio and/or accompanying data from which a virtual reality scene may be generated for display. The VR apparatus may therefore provide the VR scene by generating the virtual, three-dimensional, VR space in which to display the VR content. The virtual reality scene may be provided by a panoramic video (such as a panoramic live broadcast), comprising a video having a wide or 360° field of view (or more, such as above and/or below a horizontally oriented field of view). A panoramic video may have a wide field of view in that it has a spatial extent greater than a field of view of a user or greater than a

field of view with which the panoramic video is intended to be displayed.

[0032] The VR content provided to the user may comprise live or recorded images of the real world, captured by a VR content capture device, for example. An example VR content capture device comprises a Nokia Technologies OZO device. As the VR scene is typically larger than a portion a user can view with the VR display, the VR apparatus may provide, for display on the VR display, a virtual reality view of the VR scene to a user, the VR view showing only a spatial portion of the VR content that is viewable at any one time. The VR apparatus may provide for panning around of the VR view in the VR scene based on movement of a user's head and/or eyes. A VR content capture device may be configured to capture VR content for display to one or more users. A VR content capture device may comprise one or more cameras and, optionally, one or more (e.g. directional) microphones configured to capture the surrounding visual and aural scene from a capture point of view. In some examples, the VR content capture device comprises multiple, physically separate cameras and/or microphones. Thus, a musical performance may be captured (and recorded) using a VR content capture device, which may be placed on stage, with the performers moving around it or from the point of view of an audience member. In each case a consumer of the VR content may be able to look around using the VR display of the VR apparatus to experience the performance at the capture location as if they were present.

[0033] Augmented reality (AR) may use an AR display, such as glasses or goggles or a virtual retinal display, to augment a view of the real world (such as seen through the glasses or goggles) with computer generated content. An augmented reality apparatus, which may or may not include an AR display, may provide for presentation of multimedia AR content configured to be overlaid over the user's view of the real-world. Thus, a user of augmented reality may be able to view the real world environment around them, which is augmented or supplemented with content provided by the augmented reality apparatus, which may be overlaid on their view of the real world and/or aurally overlaid over an aural real world scene they can hear. The content may comprise multimedia content such as pictures, photographs, video, diagrams, textual information, aural content among others. Thus, while augmented reality may provide for direct viewing of the real world with the addition of computer generated graphics and/or audio content, a user of virtual reality may only be able to see content presented on the VR display of the virtual reality apparatus substantially without direct viewing of the real world.

[0034] In addition to the audio received from the microphone(s) of the VR content capture device further microphones each associated with a distinct audio source may be provided. In one or more examples, the VR content capture device may not have microphones and the aural scene may be captured by microphones remote

from the VR content capture device. Thus, microphones may be provided at one or more locations within the real world scene captured by the VR content capture device, each configured to capture audio from a distinct audio source. For example, using the musical performance example, a musical performer or a presenter may have a personal microphone. Knowledge of the location of each distinct audio source may be obtained by using transmitters/receivers or identification tags to track the position of the audio sources, such as relative to the VR content capture device, in the scene captured by the VR content capture device. Thus, the VR content may comprise the visual imagery captured by one or more VR content capture devices and the audio captured by the one or more VR content capture devices and, optionally/alternatively, one or more further microphones. The location of the further microphones may be provided for providing spatial audio.

[0035] The virtual reality content may comprise, and a VR apparatus presenting said VR content may provide, predefined-viewing-location VR or free-viewing-location VR. In predefined-viewing-location VR, the location of the user in the virtual reality space may be fixed or follow a predefined path. Accordingly, a user may be free to change their viewing direction with respect to the virtual reality imagery provided for display around them in the virtual reality space, but they may not be free to arbitrarily change their viewing location in the VR space to explore the VR space. Thus, the user may experience such VR content from a fixed point of view or viewing location (or a limited number of locations based on where the VR content capture devices were located in the scene). In some examples of predefined-viewing-location VR the imagery may be considered to move past them. In predefined-viewing-location VR content captured of the real world, the user may be provided with the point of view of the VR content capture device. Predefined-viewing-location VR content may provide the user with three degrees of freedom in the VR space comprising rotation of the viewing direction around any one of x, y and z axes and may therefore be known as three degrees of freedom VR (3DoF VR).

[0036] In free-viewing-location VR, the VR content and VR apparatus presenting said VR content may enable a user to be free to explore the virtual reality space. Thus, the user may be provided with a free point of view or viewing location in the virtual reality space. Free-viewing-location VR is also known as six degrees of freedom (6DoF) VR or volumetric VR to those skilled in the art. Thus, in 6DoF VR the user may be free to look in different directions around the VR space by modification of their viewing direction and also free to change their viewing location (their virtual location) in the VR space by translation along any one of orthogonal x, y and z axes. The movement available in a 6DoF virtual reality space may be divided into two categories: rotational and translational movement (with three degrees of freedom each). Rotational movement enables a user to turn their head to

change their viewing direction. The three rotational movements are around x-axis (roll), around y-axis (pitch), and around z-axis (yaw). Translational movement means that the user may also change their point of view in the space to view the VR space from a different virtual location, i.e., move along the x, y, and z axes according to their wishes. The translational movements may be referred to as surge (x), sway (y), and heave (z) using the terms derived from ship motions.

[0037] Mixed reality comprises a combination of augmented and virtual reality in which a three-dimensional model of the real-world environment is used to enable virtual objects to appear to interact with real-world objects in terms of one or more of their movement and appearance.

[0038] One or more examples described herein relate to 6DoF virtual reality content in which the user is at least substantially free to move in the virtual space either by user-input through physically moving or, for example, via a dedicated user interface (UI).

[0039] Spatial audio comprises audio presented in such a way to a user that it is perceived to originate from a particular location, as if the source of the audio was located at that particular location. Thus, virtual reality content may be provided with spatial audio having directional properties, such that the audio is perceived to originate from a point in the VR space, which may be linked to the imagery of the VR content. Augmented reality may be provided with spatial audio, such that the spatial audio is perceived as originating from real world objects visible to the user and/or from augmented reality graphics overlaid over the user's view.

[0040] Spatial audio may be presented independently of visual virtual reality or visual augmented reality content. Nevertheless, spatial audio, in some examples, may be considered to be augmented reality content because it augments the aural scene perceived by a user. As an example of independent presentation of spatial audio, a user may wear headphones and, as they explore the real world, they may be presented with spatial audio such that the audio appears to originate at particular locations associated with real world objects or locations. For example, a city tour could be provided by a device that tracks the location of the user in the city and presents audio describing points of interest as spatial audio such that the audio appears to originate from the point of interest around the user's location.

[0041] The spatial positioning of the spatial audio may be provided by 3D audio effects, such as those that utilise a head related transfer function to create a spatial audio space in which audio can be positioned for presentation to a user. Spatial audio may be presented by headphones by using head-related-transfer-function (HRTF) filtering techniques or, for loudspeakers, by using vector-based-amplitude panning techniques to position the perceived aural origin of the audio content. In other embodiments ambisonic audio presentation may be used to present spatial audio. Spatial audio may use one or more of vol-

ume differences, timing differences and pitch differences between audible presentation to each of a user's ears to create the perception that the origin of the audio is at a particular location in space.

[0042] In some examples, an audio track, which comprises audio content for presentation to a user, may be provided for presentation as spatial audio. Accordingly, the audio track may be associated with a particular location which defines where the user should perceive the audio of the audio track as originating. The particular location may be defined relative to a virtual space or a real-world space. The virtual space may comprise a three-dimensional environment that at least partially surrounds the user and may be explorable by the user. The virtual space may be explorable in terms of the user being able to move about the virtual space by at least translational movement based on user input. If the spatial audio is provided with virtual reality content, virtual reality imagery may be displayed in the virtual space along with spatial audio to create a virtual reality experience. If the spatial audio is provided with visual augmented reality content or independently of augmented or virtual reality content, the particular location may be defined relative to a location in the real world, such as in a real-world room or city.

In one or more examples, the virtual space may be configured to correspond to the real world space in which the user is located. Accordingly, the virtual space may be used to determine the interaction between real world objects and locations and virtual objects and locations.

[0043] In one or more examples, the audio track and location information may be considered to define an audio object in the virtual space comprising a source location for the audio of an associated audio track. The audio objects may be moveable or non-moveable. The audio objects may or may not have a corresponding visual object.

[0044] Figure 1 shows an example system 100 for presentation of spatial audio content to a user. The system 100 includes an example apparatus 101 for controlling the presentation of audio tracks of the spatial audio content based on the user's location relative to the audio objects and movement of one or both of the user and the audio object relative to the other. The apparatus 101 may comprise or be connected to a processor 101A and a memory 101B and may be configured to execute computer program code. The apparatus 101 may have only one processor 101 A and one memory 101B but it will be appreciated that other embodiments may utilise more than one processor and/or more than one memory (e.g. same or different processor/memory types). Further, the apparatus 101 may be an Application Specific Integrated Circuit (ASIC).

[0045] The processor may be a general purpose processor dedicated to executing/processing information received from other components, such as from a location tracker 102 and a content store 103, in accordance with instructions stored in the form of computer program code in the memory. The output signalling generated by such

operations of the processor is provided onwards to further components, such as to audio presentation equipment, such as headphones 108.

[0046] The memory 101B (not necessarily a single memory unit) is a computer readable medium (solid state memory in this example, but may be other types of memory such as a hard drive, ROM, RAM, Flash or the like) that stores computer program code. This computer program code stores instructions that are executable by the processor, when the program code is run on the processor. The internal connections between the memory and the processor can be understood to, in one or more example embodiments, provide an active coupling between the processor and the memory to allow the processor to access the computer program code stored on the memory.

[0047] In this example the respective processors and memories are electrically connected to one another internally to allow for electrical communication between the respective components. In this example the components are all located proximate to one another so as to be formed together as an ASIC, in other words, so as to be integrated together as a single chip/circuit that can be installed into an electronic device. In some examples one or more or all of the components may be located separately from one another.

[0048] The apparatus 101, in this example, forms part of a virtual reality apparatus 104 for presenting visual imagery in virtual reality. In one or more other examples, the apparatus 101 may form part of an AR apparatus. In one or more examples, the apparatus 100 may be independent of an AR or VR apparatus and may provide signalling to audio presentation equipment 108 (such as speakers, which may be incorporated in headphones) for presenting the audio to the user. In this example, the processor 101A and memory 101B is shared by the VR apparatus 104 and the apparatus 101, but in other examples, they may have their own processors and/or memory.

[0049] The VR apparatus 104 may provide for display of virtual reality content comprising visual imagery displayed in a virtual space that is viewable by a user using the VR headset 107. In one or more examples in which the apparatus 100 is independent of an AR or VR apparatus, the VR headset 107 may not be required and instead only the audio presentation equipment 108 may be provided.

[0050] The apparatus 101 or the VR apparatus 104 under the control of the apparatus 101 may provide for aural presentation of the audio tracks to the user using the headphones 108. The apparatus 101 may be configured to process the audio such that, at any one time, it is presented as one of spatial, monophonic and stereophonic audio or, alternatively or in addition, the apparatus 101 may provide signalling to control the processing and/or presentation of the audio tracks. Accordingly, an audio processor (not shown) may perform the audio processing in order to present the audio in the ways men-

tioned above under the control of the apparatus 101.

[0051] The apparatus 101 may receive signalling indicative of the location of the user from a location tracker 102. The location tracker 102 may determine the user's head orientation and/or the user's location in the real world so that the spatial audio may be presented to take account of head rotation and movement so that the audio is perceived to originate from a direction relative to the user irrespective of user's head movement. If the spatial audio is provided in a virtual reality environment, the location tracker 102 may provide signalling indicative of user movement so that corresponding changes in the user's virtual location in the virtual space can be made.

[0052] In the examples that follow, spatial audio content comprising one or more audio tracks, which may be provided from content store 103, may be processed such that they are presented to the user as spatial audio or stereophonic or monophonic audio. Accordingly, in a first instance, the audio track may be presented as spatial audio and as such may undergo audio processing such that it is perceived to originate from a particular location. In a second instance, the same audio track may be presented as monophonic audio and as such may undergo audio processing (if required) such that the audio is presented monophonically to one or both of a left and right speaker associated with the left and right ears of the user. In a third instance, the same audio track may be presented as stereophonic audio (if required) and as such may undergo audio processing such that the audio of the audio track is presented to one or both of a left and right speaker associated with the left and right ear of the user respectively (or even in between the two ears). Monophonic audio, when presented to two speakers provides the same audio to both ears. Stereophonic audio may define two (left and right) or three (left, right, centre) stereo audio channels and the audio of the audio track may be presented to one or more of those channels. In some examples, the difference between stereophonic presentation and spatial audio presentation may be, for spatial audio, the use of a time delay between corresponding audio being presented to speakers associated with a respective left and right ear of the user and, for stereophonic presentation, the non-use of said time delay. It will be appreciated that the presentation of spatial audio may additionally use other presentation effects in addition to differences in the time that corresponding portions of the audio is presented to the user's ears to create the perception of a direction or location from which the audio is heard, such as volume differences amongst others.

[0053] While the same audio track may undergo audio processing in order to provide for its presentation as spatial audio or stereophonic or monophonic audio, as described above, in one or more other examples, the audio tracks may be pre-processed and may thus include different versions for presentation as spatial audio or stereophonic or monophonic audio. In one or more examples, the presentation of an audio track as spatial audio may decrease its fidelity and thus, presentation as mono-

phonic or stereophonic audio may provide for an increase in audio quality.

[0054] Figure 2 shows an example virtual space 200. The virtual space may comprise a three-dimensional virtual environment in which the location of a user, comprising user-location 201, is defined as well as the location of one or more audio objects 202-206. In this example, the spatial audio content that defines the audio objects is part of virtual reality content, which includes visual content to accompany the spatial audio content, although in other examples the audio objects may not have corresponding visual objects. In this example, the location of the user is shown diagrammatically by an image of a person but it will be appreciated that in one or more examples, the user location 201 designates the location in the virtual space at which the user perceives the audio presented to them, i.e. a "point-of-hearing location" similar to a "point-of-view location". The first through fifth audio objects 202-206 are illustrated diagrammatically by their corresponding visual representations. Thus, the first audio object 202 represents the audio from a first drummer who appears in the visual content. The second audio object 203 represents the audio from a second drummer who appears in the visual content. The third audio object 204 represents the audio from a guitarist who appears in the visual content. The fourth audio object 205 represents the audio from a ballerina who appears in the visual content. The fifth audio object 206 represents the audio from a singer who appears in the visual content. In this example, at least the ballerina may move about the virtual space 200 and accordingly the visual imagery of the ballerina and the audio object may, correspondingly, move with elapsed time through the virtual reality content.

[0055] Each of the audio objects 202-206 may be associated with an audio track which may be presented as spatial audio to the user. Accordingly, the first audio object 202 defines the location of the perceived origin of an associated first audio track as perceived by the user from their user location 201. Likewise, the audio of the audio tracks associated with the second to fifth audio objects is presented to the user such that the user perceives the origin of the second to fifth audio tracks as originating at the location of the respective audio objects relative to the user location 201.

[0056] As will be appreciated when the audio is presented as spatial audio, when the user changes their location in the virtual space 200, there is a corresponding change in the presentation of the audio track as spatial audio. For example, the volume of the audio track presented to the user may be a function of the distance of the user location 201 from the corresponding audio object location. Thus, in one or more examples, as the user moves towards the audio object location the spatial audio presented audio track is presented louder and as the user moves away the audio track is presented more quietly. Also, as the user moves their head, the direction (relative to the user's head) from which the spatial audio

is perceived to originate changes in accordance with the direction to the audio object location relative to the user's direction of view.

[0057] In this examples, the virtual reality content is free-viewing-location VR or "6DoF" and therefore the apparatus 101 or VR apparatus 104 may be configured to visually and aurally present the visual imagery and spatial audio content in accordance with user-input to move their user location in the virtual space 200 as well as their viewing direction.

[0058] While in these examples the spatial audio content may be associated with virtual reality content, in one or more examples, the spatial audio content comprises augmented reality content and the virtual space 200 corresponding to a real-world space in which the user is located such that a location of the user in the real-world space corresponds to the user location in the virtual space and the audio-object location in the virtual space corresponds to a real-world-audio-object location in the real-world space. Thus, in general, the apparatus 101 may be configured to provide control of the presentation of spatial audio in a virtual or augmented reality environment.

[0059] In one or more examples, the user may wish to control the way in which the spatial audio is presented. Accordingly, the user may wish to control audio parameters of the audio tracks associated with the audio objects 202-206, such as in terms of their relative volume, frequency dependent gain applied or other audio presentation parameter, for example. The term "audio mixing" may be used to refer to the control of one or more audio parameters associated with the audible presentation of the audio track. The change in audio parameters may then be applied when the spatial audio content is later consumed by the same or a different user or may be applied to live content that is then provided for sending to multiple consumers of VR content.

[0060] In one or more examples, the user 201 may comprise a content producer and the apparatus 101 may provide a spatial audio mixing apparatus for spatial audio content production. In one or more examples, the user 201 may comprise a consumer of spatial audio content, or VR/AR content that includes spatial audio content, and the apparatus 101 may comprise at least part of a spatial audio presentation apparatus.

[0061] In one or more examples, the user may wish to select one of the audio objects 202-206 on which to perform audio mixing by moving about the virtual space such that the user-location is within a threshold distance of the audio object. With the audio object selected in this way the user may be provided with an audio mixing interface to provide for control of one or more audio parameters of the audio track associated with the selected audio object 202-206. However, as mentioned above, if the user can move about the virtual space 200 and, alternatively or in addition, the audio object 205 may move about the virtual space 200, maintaining selection of the audio object with such relative movement may be problematic.

[0062] In one or more examples, if the user wishes to maintain selection of a moving audio object, the user may have to inconveniently change their location to keep track of the audio object. If the control of user location in the virtual space 200 is controlled by tracking physical user movement, this may be physically tiring. In other examples, not comprising a function of the apparatus 101, there may be a simplistic latching or locking onto the audio object. However, in such a situation, the user's point of view in the virtual space is limited to that controlled by the movements of the object of interest, which may be confusing.

[0063] The example apparatus 101, as explained below, provides a way of maintaining selection of an audio object irrespective of at least some relative movement between the user and the audio object. With said selection, the audio parameters that affect how the audio of the audio track associated with the selected audio object is presented may be more readily controlled.

[0064] The example figure 3 shows the same virtual space 200 as figure 2 alongside a plan view 300 of the virtual space 200. The audio objects 202-206 and the user location 201 are also shown.

[0065] In this example, the audio objects are shown having a bubble 302-306 surrounding them. The bubbles may, in one or more examples, represent a predetermined-bubble-distance 307 that extends around each audio object 202-206. The predetermined-bubble-distance 307 may comprise a predetermined threshold distance which may be used as a means for selecting the audio objects when the user 201 location is within said predetermined-bubble-distance 307 of a particular audio object 202-206. In one or more embodiments, the predetermined-bubble-distance may define a spherical volume around the audio objects to be used for selection thereof. In other embodiments, the predetermined-bubble-distance may be different depending on the angle of approach towards said audio object. In one or more examples, the predetermined bubble-distance may extend from a centre of the audio object or a visual object associated with the audio object to define the bubble. In other examples, the predetermined bubble-distance may extend from a surface or a bounding geometrical volume (e.g. bounding cuboid or bounding ellipsoid) of the audio object or a visual object associated with the audio object. Accordingly, the bubble 302-306 may be non-spherical and may generally follow the shape and extent of the visual object or the shape and extent of the audio object (if the audio object is not a point source for the audio). The bubbles 302-306 may or may not be visible. In one or more examples, only a subset of the audio objects may be selectable and thus only a subset may have an associated bubble. Whether or not an audio object is selectable or not may be defined in information associated with the spatial audio content.

[0066] The distance between the user 201 location and the fourth audio-object 205 location is shown by arrow 308 and it is greater than the predetermined-bubble-dis-

tance 307.

[0067] The example of figure 4 happens to show the selection of the fourth audio object 205 by virtue of the user moving to a new user 201 location in the virtual space shown by arrow 400. It will be understood that the user 201 could have approached any other of the audio objects. The old user location 401 is shown for understanding. In such a location, the distance between the user 201 location and the fourth audio-object 205 location, shown by arrow 308, has decreased from the distance shown in figure 3 to less than the predetermined-bubble-distance 307. Accordingly, the apparatus 101 may, based on this condition being met, provide for selection of the fourth audio object 205. In this configuration, the user may be considered to be within the bubble 305.

[0068] Thus, the apparatus 101, based on spatial audio content comprising at least one audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from an fourth audio-object 205 location in the virtual space 200, the user having a user 201 location in the virtual space 200 from which to perceive the spatial audio and based on the distance 308 between the user 201 location and the fourth audio-object 205 location having decreased to less than the predetermined-bubble-distance 307 has provided for selection of the audio object. Selection of the fourth audio object 205 is configured to provide for a change in the audible presentation of the fourth audio track associated with the fourth audio object to the user from presentation as spatial audio to presentation as one of monophonic and stereophonic audio.

[0069] In one or more examples, the apparatus 101 may require additional conditions to be met before selecting one of the audio objects. For example, the user may be required to approach the fourth audio object 205 below the predetermined-bubble-distance 307 and provide a user input before the apparatus provides for selection of the fourth audio object 205.

[0070] The example of figure 5 shows the change in audible presentation of the audio tracks associated with the first to fifth audio objects 202-206. In the plan view of figure 5 the first, second, third and fifth audio objects are shown with arrows 502, 503, 504, 506 to indicate the direction from which the user perceives the spatial audio of the audio track associated with those audio objects 202, 203, 204, 206. Accordingly, the audio of the audio tracks of the first, second, third and fifth audio objects is heard externalized, that is from a particular direction or location in the virtual space. The audio track 507 of the selected fourth audio object 205 is shown positioned within the user's head and is shown as a circle in figure 5. This signifies that the audio of the fourth audio track is rendered as non-externalized in-head audio, which comprises one of monophonic audio or stereophonic audio. The fourth audio object 205 is not shown in the plan view 300 as it would be overlapping with the user's 201 head of the plan view designating the user location.

[0071] The selection of the fourth audio object 205 may

provide for audio mixing of audio track associated with that audio object. Figure 5 shows an example audio mixing interface 500 that may be visually presented for audio mixing. The apparatus 101 may be configured to receive user-input via said audio mixing interface for modification of audio parameters associated with the selected fourth audio track. In other examples, no visual interface is provided and instead other user input, such as predetermined gestures or voice commands may provide for control of one or more audio parameters.

[0072] In one or more examples, the apparatus 101 may be caused to provide the audio mixing user interface 500 for modification of one or more audio parameters of only the selected audio track. The one or more audio parameters may be selected from at least one or more of: volume, bass level, mid-tone level, treble-level, reverberation level and echo among others. The audio mixing user interface 500 may be visually presented and include sliders or other control interfaces for each audio parameter. In other examples, the audio mixing user interface may not be visualized and predetermined user gestures may control predetermined audio parameters. For example, rotation of the user's left hand may control volume level while rotation of the user's right hand may control reverberation. It will be appreciated that other gesture/audio parameter combinations are possible.

[0073] Example figures 6 and 7 illustrate relative movement between the user 201 location and the location of the fourth audio object 205 in the virtual space 200 causing the distance 308 between the user 201 location and the location of the fourth audio object 205 to increase. Relative movement of the audio objects and/or the user may make maintaining selection of one of the audio objects difficult.

[0074] Figure 6 shows the fourth audio object 205 moving away from the user 201 location. The apparatus 101 is configured to maintain the presentation of the audio of the fourth audio track 507 as monophonic or stereophonic audio, which may be advantageous for audio mixing, irrespective of movement of the fourth audio object 205 away from the user 201 location in the virtual space causing the distance 308 between the user location and the fourth audio-object 205 location to increase beyond the predetermined-bubble-distance 307.

[0075] As shown in figure 6, this may be compared to the bubble 305 stretching beyond its original size, the original size comprising the size and/or shape prior to the user "entering the bubble". Accordingly, in one or more examples, the user may continue to benefit from higher audio quality provided by the monophonic/stereophonic presentation of the fourth audio track and thus avoid distractions caused by changes in spatial audio rendering due to relative movement of the audio objects and user location.

[0076] The apparatus 101 may be configured to provide for audible presentation of the audio tracks associated with the other, unselected audio objects 202, 203, 204, 206 as spatial audio and therefore they are per-

ceived as originating from their respective locations in the virtual space.

[0077] It will be appreciated that in one or more examples, ambience audio content may be provided, in addition, and the apparatus may provide for presentation of the ambience audio content.

[0078] The ambience audio content may comprise audio that does not have location information such that it may not be presented as originating from a particular point in space and is presented as ambience.

[0079] Figure 7 shows the user moving in the virtual space such that the user 201 location is moving away from the fourth audio object 205. The apparatus 101 is configured to maintain selection and thus presentation of the audio of the fourth audio track 507 as monophonic or stereophonic audio, which may be advantageous for audio mixing, irrespective of movement of the user 201 location away from the fourth audio object 205 in the virtual space causing the distance 308 between the user location and the fourth audio-object 205 location to increase beyond the predetermined-bubble-distance 307.

[0080] As shown in figure 7 and similar to figure 6, this may be compared to the bubble 305 stretching beyond its original size. Accordingly, in one or more examples, the user may continue to benefit from higher audio quality provided by the monophonic/stereophonic presentation of the fourth audio track and thus avoid distractions caused by changes in spatial audio rendering due to relative movement of the audio objects and user location.

[0081] As described above in relation to figure 6, the apparatus 101 may be configured to provide for audible presentation of the audio tracks associated with the other, unselected audio objects 202, 203, 204, 206 as spatial audio and therefore they are presented such that they are perceived as originating from their respective locations in the virtual space relative to the user 201 location that the user has moved to. As in the previous example, the apparatus may additionally provide for presentation of the ambience audio content.

[0082] The provision of selection of an audio object based on a distance between the user and the audio object decreasing to within the predetermined-bubble-distance, i.e. the user moving near to or bumping into the audio object, and maintaining selection even if subsequent movement of the user and/or audio object increases said distance above the predetermined-bubble-distance may have technical advantages. The maintenance of selection may allow for less exertion by the user to "track" the selected, moving, audio object. Further, selection of the audio objects in this way may allow for presentation of a more stable audio scene containing other audio objects because the user does not have to move to maintain selection and thus the direction from which the user perceives other spatial audio in the scene does not have to be modified based on the moving user location. As described in these examples, the selection of the audio object provide for presentation of the audio of the audio object from its default presentation as spatial audio

to one of stereophonic/monophonic audio, which may be technically advantageous for audio mixing. Further, given that the selected audio object is not presented as spatial audio, movement of the user and/or audio object in the virtual space does not lead to an audible distraction or modification which would occur with presentation as spatial audio.

[0083] Example figures 8 and 9 show the bubble 305 stretching and breaking. The breaking of the bubble 305 is symbolic of the apparatus 101 being configured to provide for a change in the audible presentation of the fourth audio track 507 to the user from presentation as at least one of monophonic and stereophonic audio to presentation as spatial audio.

[0084] When the bubble 305 breaks may be determined based on a further threshold distance between the user 201 location and the selected audio object 205. The threshold distance may be termed a predetermined-stretched-bubble-distance which is greater than the predetermined-bubble-distance.

[0085] Figure 8 shows the relative movement between the user 201 location and the current location of the fourth audio object 205 in the virtual space 200 causing the distance 308 between the user location and the first audio-object location to approach the predetermined-stretched-bubble-distance 800. In one or more examples an audio effect may be applied to the monophonically or stereophonically presented fourth audio track to audible indicate that the predetermined-stretched-bubble-distance 800 is almost reached. Thus, within a threshold below the predetermined-stretched-bubble-distance 800 one or more of an audible, visual and haptic feedback may be provided to the user. The visual feedback may comprise a message or graphic. The audio feedback, may comprise a spoken message or an audio effect which may comprise an echo effect or reverberation effect or an underwater effect or a distinguishable audio tone.

[0086] Figure 9 shows the relative movement between the user 201 location and the location of the fourth audio-object 205 having caused the distance 308 between the user location and the first audio-object location to have increased beyond the predetermined-stretched-bubble-distance 800. Accordingly, with reference to the plan view 900, the apparatus 101 may provide for a change in the audible presentation of the fourth audio track to the user from presentation as at least one of monophonic and stereophonic audio to presentation as spatial audio such that the audio of the fourth audio track is perceived from a direction 901 based on the current location of the fourth audio object 205 relative to the current user location 201. The other audio objects 202, 203, 204, 206 may continue to be presented as spatial audio.

[0087] Accordingly, relative movement between the audio object location and the user 201 location while the distance remains below the predetermined-stretched-bubble-distance may not affect the audible presentation of the fourth audio track because it is presented as mono-

phonic or stereophonic audio. However, once the bubble breaks and the apparatus provides for presentation of the fourth audio track as spatial audio, then relative movement between the audio object location and the user location 201 will affect the audible presentation because spatial audio is rendered based on the location of the audio object relative to the user 201 location.

[0088] In the above example, the predetermined-stretched-bubble-distance was used by the apparatus 101 to determine when to deselect and thus switch back to spatial audio presentation of an audio track. In one or more examples, the apparatus 101 may additionally or alternatively provide said change in audible presentation based on a user request. Thus, with the fourth audio track selected and audibly presented as one of monophonic and stereophonic audio and based on user input indicative of a desire to return to spatial audio presentation of the fourth audio track, the apparatus 101 may provide for a change in the audible presentation of the fourth audio track to the user from presentation as at least one of monophonic and stereophonic audio to presentation as spatial audio.

[0089] It will be appreciated that the relative locations of the user and/or the audio object may have changed since the audio object was selected and thus, when changing back to presentation as spatial audio the audio of the fourth audio track may be perceived from a direction based on the current audio-object location relative to the current user 201 location at the time of said user input or the time of the predetermined-stretched-bubble-distance being exceeded.

[0090] The apparatus 101 may provide for changes requested by way of the audio mixing user interface 500 to be audibly presented in real time to the user. The changes may be automatically saved or may require a user input to be saved for future presentation of the spatial audio content or for application to live spatial audio content for onward transmission.

[0091] In one or more examples, based on receipt of user input to the audio mixing user interface 500 causing changes to one or more audio parameters of the audio track and, subsequently, the bubble 305 is broken due to the distance between the user location and the audio-object location increasing beyond the predetermined-stretched-bubble-distance or user input indicative of a desire to return to spatial audio presentation of the fourth audio track, the apparatus may be configured to provide for discarding of the changes to one or more audio parameters. Thus, to save the changes the user must provide a save command prior to said bubble 305 breaking. In other examples, on breaking of the bubble 305 the current changes to the audio parameters are saved.

[0092] Although the examples above illustrate the operation of the apparatus 101 in relation to selection of the fourth audio object, any of the audio objects may be selected.

[0093] With reference to figures 8 and 9, on breaking of the bubble 305, the audible presentation of the fourth

audio track may abruptly transition from monophonic/stereophonic presentation to spatial audio presentation from a particular location in the virtual space and thus a particular direction. This may be confusing.

[0094] In one or more examples, the presentation of audio monophonically or stereophonically may be considered to locate the source of the audio within a user's head. Accordingly, on transition to presentation as spatial audio, the apparatus 101 may provide for rendering of the spatial audio of the now deselected fourth audio object 205 from at least one or more intermediate positions between the user 201 location and the current location of the fourth audio object (as shown in figure 9) over a transition period. This may be perceived as the audio of the fourth audio track gradually moving away from its in-head presentation (monophonically/stereophonically) to its true, current position in the virtual space 200. Thus, between the audible presentation of the fourth audio track to the user as at least one of monophonic and stereophonic audio and the presentation of the fourth audio track as spatial audio, the apparatus 101 may be caused to provide for audible presentation of the fourth audio track with a transitional effect comprising the perceived origin of the audio of fourth audio track progressively moving away from the user 201 location to the current fourth audio-object 205 location over a transition time. The transition time may comprise less than a second or less than one, two, three, four or five seconds or, in general, the transition time may range from almost instantaneous to multiple seconds.

[0095] When presenting spatial audio, it may be advantageous to apply effects to the audio to replicate how the audio would be heard in a particular room. In one or more examples, the audio tracks may comprise audio captured from close up microphones. Audio captured of an audio source by close up microphones may sound different to audio heard by a user in a room with the audio source because the user would hear the same audio but it would typically include component echos and reverberations caused by the sound waves from the audio source interacting with the surfaces in the room. Thus, in one or more examples, an audio effect termed a Room Impulse Response may be applied to the audio tracks which may make them sound as if heard in a particular room. The Room Impulse Response may comprise an audio processing function that simulates the effect of the surfaces of a particular room. The Room Impulse Response may also comprise a function of the user's location in the particular room relative to the audio object. The particular room itself may be based on the virtual reality content presented to the user. Thus, if the visual content of the VR content shows a hard walled, and therefore echo-prone, room, the apparatus may apply a Room Impulse Response function to replicate such an audio environment.

[0096] With application to the present apparatus 101, when presenting one of a plurality of audio tracks of the spatial audio content monophonically or stereophonically

with the other audio tracks presented as spatial audio, the apparatus may be configured to apply a room impulse response function to at least one of said other audio tracks, the room impulse response function configured to modify the audio of the at least one other audio tracks to sound as if it is heard in a predetermined room with a particular location in said predetermined room.

[0097] With reference to figure 10, the particular location in said predetermined room may, in a first example, be determined based on the user 201 location and, in a second example, be determined based on the current location of the selected audio object. Figure 10 shows the location of the room impulse response function being based on the user location at 1001. Figure 10 shows the location of the room impulse response function being based on the location of the selected audio object at 1002. Accordingly, the different room impulse response functions may be applied to the audio tracks of the first, second, third and fifth audio objects 202, 203, 204, 206.

[0098] Accordingly, the apparatus 101 may provide for presentation of the other audio in the virtual space with a room impulse response function from either the user's 201 listening position or the position corresponding to the fourth audio object 205.

[0099] In one or more examples, the apparatus 101 may provide for user-selection of the Room Impulse Response function in terms of being based on the user location or based on the current location of the selected, fourth audio object.

[0100] In one or more examples, the Room Impulse Response function may be applied to the selected audio track to simulate the audio being heard within a room defined by the bubble 305.

[0101] In one or more examples, the Room Impulse Response function may be continually or periodically updated based on the current position in the room of the user 201 and/or the one or more audio objects.

[0102] Figure 11 shows a flow diagram illustrating the steps of,

based on 1101 spatial audio content comprising at least one audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a first audio-object location in a virtual space, the user having a user location in the virtual space from which to perceive the spatial audio and based on the distance between the user location and the first audio-object location having decreased to less than a predetermined-bubble-distance;

providing for 1102 a change in the audible presentation of the first audio track to the user from presentation as spatial audio to presentation as at least one of monophonic and stereophonic audio for audio mixing of said first audio track, and wherein said presentation of said first audio track as at least one of monophonic and stereophonic audio is maintained irrespective of relative movement between

the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond the predetermined-bubble-distance.

[0103] In one or more examples, the spatial audio content may include a second audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a second audio-object location in the virtual space. At least when the distance between the user location and the second audio-object location is greater than the predetermined-bubble-distance, the method may provide for audible presentation of the second audio track to the user as spatial audio based on the relative location of the user location and the second audio-object location while providing the above-mentioned change in the audible presentation of the first audio track to the user from presentation as spatial audio to presentation as at least one of monophonic and stereophonic audio for audio mixing of said first audio track.

[0104] Figure 12 illustrates schematically a computer/processor readable medium 1200 providing a program according to an example. In this example, the computer/processor readable medium is a disc such as a digital versatile disc (DVD) or a compact disc (CD). In some examples, the computer readable medium may be any medium that has been programmed in such a way as to carry out an inventive function. The computer program code may be distributed between the multiple memories of the same type, or multiple memories of a different type, such as ROM, RAM, flash, hard disk, solid state, etc.

[0105] User inputs may be gestures which comprise one or more of a tap, a swipe, a slide, a press, a hold, a rotate gesture, a static hover gesture proximal to the user interface of the device, a moving hover gesture proximal to the device, bending at least part of the device, squeezing at least part of the device, a multi-finger gesture, tilting the device, or flipping a control device. Further the gestures may be any free space user gesture using the user's body, such as their arms, or a stylus or other element suitable for performing free space user gestures.

[0106] The apparatus shown in the above examples may be a portable electronic device, a laptop computer, a mobile phone, a Smartphone, a tablet computer, a personal digital assistant, a digital camera, a smartwatch, smart eyewear, a pen based computer, a non-portable electronic device, a desktop computer, a monitor, a smart TV, a server, a wearable apparatus, a virtual reality apparatus, or a module/circuitry for one or more of the same.

[0107] Any mentioned apparatus and/or other features of particular mentioned apparatus may be provided by apparatus arranged such that they become configured to carry out the desired operations only when enabled, e.g. switched on, or the like. In such cases, they may not necessarily have the appropriate software loaded into the active memory in the non-enabled (e.g. switched off

state) and only load the appropriate software in the enabled (e.g. on state). The apparatus may comprise hardware circuitry and/or firmware. The apparatus may comprise software loaded onto memory. Such software/computer programs may be recorded on the same memory/processor/functional units and/or on one or more memories/processors/ functional units.

[0108] In some examples, a particular mentioned apparatus may be pre-programmed with the appropriate software to carry out desired operations, and wherein the appropriate software can be enabled for use by a user downloading a "key", for example, to unlock/enable the software and its associated functionality. Advantages associated with such examples can include a reduced requirement to download data when further functionality is required for a device, and this can be useful in examples where a device is perceived to have sufficient capacity to store such pre-programmed software for functionality that may not be enabled by a user.

[0109] Any mentioned apparatus/circuitry/elements/processor may have other functions in addition to the mentioned functions, and that these functions may be performed by the same apparatus/circuitry/elements/processor. One or more disclosed aspects may encompass the electronic distribution of associated computer programs and computer programs (which may be source/transport encoded) recorded on an appropriate carrier (e.g. memory, signal).

[0110] Any "computer" described herein can comprise a collection of one or more individual processors/processing elements that may or may not be located on the same circuit board, or the same region/position of a circuit board or even the same device. In some examples one or more of any mentioned processors may be distributed over a plurality of devices. The same or different processor/processing elements may perform one or more functions described herein.

[0111] The term "signalling" may refer to one or more signals transmitted as a series of transmitted and/or received electrical/optical signals. The series of signals may comprise one, two, three, four or even more individual signal components or distinct signals to make up said signalling. Some or all of these individual signals may be transmitted/received by wireless or wired communication simultaneously, in sequence, and/or such that they temporally overlap one another.

[0112] With reference to any discussion of any mentioned computer and/or processor and memory (e.g. including ROM, CD-ROM etc), these may comprise a computer processor, Application Specific Integrated Circuit (ASIC), field-programmable gate array (FPGA), and/or other hardware components that have been programmed in such a way to carry out the inventive function.

[0113] The applicant hereby discloses in isolation each individual feature described herein and any combination of two or more such features, to the extent that such features or combinations are capable of being carried out based on the present specification as a whole, in the

light of the common general knowledge of a person skilled in the art, irrespective of whether such features or combinations of features solve any problems disclosed herein, and without limitation to the scope of the claims. The applicant indicates that the disclosed aspects/examples may consist of any such individual feature or combination of features. In view of the foregoing description it will be evident to a person skilled in the art that various modifications may be made within the scope of the disclosure.

[0114] While there have been shown and described and pointed out fundamental novel features as applied to examples thereof, it will be understood that various omissions and substitutions and changes in the form and details of the devices and methods described may be made by those skilled in the art without departing from the scope of the disclosure. For example, it is expressly intended that all combinations of those elements and/or method steps which perform substantially the same function in substantially the same way to achieve the same results are within the scope of the disclosure. Moreover, it should be recognized that structures and/or elements and/or method steps shown and/or described in connection with any disclosed form or examples may be incorporated in any other disclosed or described or suggested form or example as a general matter of design choice. Furthermore, in the claims means-plus-function clauses are intended to cover the structures described herein as performing the recited function and not only structural equivalents, but also equivalent structures. Thus although a nail and a screw may not be structural equivalents in that a nail employs a cylindrical surface to secure wooden parts together, whereas a screw employs a helical surface, in the environment of fastening wooden parts, a nail and a screw may be equivalent structures.

Claims

1. 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 to perform at least the following:

based on spatial audio content comprising at least one audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a first audio-object location in a virtual space relative to a user location of the user in the virtual space, and based on the distance between the user location and the first audio-object location having

decreased to less than a predetermined-bubble-distance;

provide for a change in the audible presentation of the first audio track to the user from presentation as spatial audio to presentation as at least one of monophonic and stereophonic audio for audio mixing of said first audio track, and wherein said presentation of said first audio track as at least one of monophonic and stereophonic audio is maintained irrespective of relative movement between the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond the predetermined-bubble-distance.

2. The apparatus of claim 1, wherein based on the relative movement between the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond a predetermined-stretched-bubble-distance which is greater than the predetermined-bubble-distance; provide for a change in the audible presentation of the first audio track to the user from presentation as at least one of monophonic and stereophonic audio to presentation as spatial audio such that the audio of the first audio track is perceived from a direction based on the first audio-object location relative to the user location.

3. The apparatus of claim 1 or claim 2, wherein one or both of:

- i) user initiated movement of their user location in the virtual space; and
- ii) movement of the audio-object in the virtual space;

provides for the relative movement.

4. The apparatus of any preceding claim, wherein the spatial audio content comprises a plurality of audio tracks including at least the first audio track and a second audio track, the second audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a second audio-object location in the virtual space; and based on the presentation of said first audio track as at least one of monophonic and stereophonic audio, provide for audible presentation of the second audio track as spatial audio such that the audio of the second audio track is perceived from a direction based on the second audio-object location relative to the user location.

5. The apparatus of any preceding claim, wherein, with the first audio track audibly presented as one of monophonic and stereophonic audio, based on user input indicative of a desire to return to spatial audio presentation of the first audio track, provide for a change in the audible presentation of the first audio track to the user from presentation as at least one of monophonic and stereophonic audio to presentation as spatial audio such that the audio of the first audio track is perceived from a direction based on the first audio-object location relative to the user location.
6. The apparatus of any preceding claim wherein in addition to the apparatus being caused to provide for a change in the audible presentation of the first audio track, the apparatus is caused to provide an audio mixing user interface for modification of one or more audio parameters of the first audio track.
7. The apparatus of claim 6, wherein the one or more audio parameters comprise at least one or more of: volume, bass level, mid-tone level, treble-level, reverbation level and echo.
8. The apparatus of claim 6 or 7, wherein based on receipt of user input to the audio mixing user interface causing changes to one or more audio parameters of the first audio track and, subsequently, at least one of:
- i) relative movement between the user location and the audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond a predetermined-stretched-bubble-distance which is greater than the predetermined-bubble-distance;
 - ii) user input indicative of a desire to return to spatial audio presentation of the first audio track;
- provide for one of:
- c) discarding of the changes to one or more audio parameters of the first audio track unless a user initiated save input is received; and
 - d) a change in the audible presentation of the first audio track to presentation as spatial audio such that the audio of the first audio track is perceived from a direction based on the first audio-object location relative to the user location with the changes to the one or more audio parameters applied.
9. The apparatus of claim 2 or claim 5, wherein between the audible presentation of the first audio track to the user as at least one of monophonic and stereophonic audio and the presentation of the first audio track as spatial audio such that the audio of the first audio track is perceived from a direction based on the first audio-object location relative to the user location, the apparatus is caused to provide for audible presentation of the first audio track with a transitional spatial audio effect comprising the perceived origin of the audio of first audio track progressively moving away from the user location to the current first audio-object location.
10. The apparatus of claim 2, wherein based on the relative movement between the user location and the audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase to within a threshold of the predetermined-stretched-bubble-distance, provide for audible presentation of the first audio track to the user as at least one of monophonic and stereophonic audio with an audio effect to thereby audibly indicate that the user is approaching the predetermined-stretched-bubble-distance.
11. The apparatus of any preceding claim wherein when presenting one of a plurality of audio tracks of the spatial audio content monophonically or stereophonically with the other audio tracks presented as spatial audio, the apparatus is caused to apply a room impulse response function to at least one of said other audio tracks, the room impulse response function configured to modify the audio of the at least one other audio tracks to sound as if it is heard in a predetermined room with a particular location in said predetermined room, the particular location in said predetermined room based on either:
- i) the user location; and
 - ii) a current location of the first audio object.
12. The apparatus of claim 11, wherein the apparatus is caused to provide for user-selection of one of:
- a) the Room Impulse Response function based on the user location; and
 - b) the Room Impulse Response function based on the current location of the first audio object.
13. The apparatus of claim 1, wherein the spatial audio content is audibly presented as spatial audio by processing the first audio track using one or more of:
- i) a head-related-transfer-function filtering technique;
 - ii) a vector-base-amplitude panning technique; and
 - iii) binaural audio presentation.
14. A method, the method comprising based on spatial audio content comprising at least

one audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a first audio-object location in a virtual space relative to a user location of the user in the virtual space, and based on the distance between the user location and the first audio-object location having decreased to less than a predetermined-bubble-distance; 5

providing for a change in the audible presentation of the first audio track to the user from presentation as spatial audio to presentation as at least one of monophonic and stereophonic audio for audio mixing of said first audio track, and wherein said presentation of said first audio track as at least one of monophonic and stereophonic audio is maintained irrespective of relative movement between the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond the predetermined-bubble-distance. 10 15 20

15. A computer readable medium comprising computer program code stored thereon, the computer readable medium and computer program code being configured to, when run on at least one processor, perform the method of: 25

based on spatial audio content comprising at least one audio track comprising audio for audible presented to a user as spatial audio such that the audio is perceived to originate from a first audio-object location in a virtual space relative to a user location of the user in the virtual space, and based on the distance between the user location and the first audio-object location having decreased to less than a predetermined-bubble-distance; 30 35

providing for a change in the audible presentation of the first audio track to the user from presentation as spatial audio to presentation as at least one of monophonic and stereophonic audio for audio mixing of said first audio track, and wherein said presentation of said first audio track as at least one of monophonic and stereophonic audio is maintained irrespective of relative movement between the user location and the first audio-object location in the virtual space causing the distance between the user location and the first audio-object location to increase beyond the predetermined-bubble-distance. 40 45 50

55

Figure 1

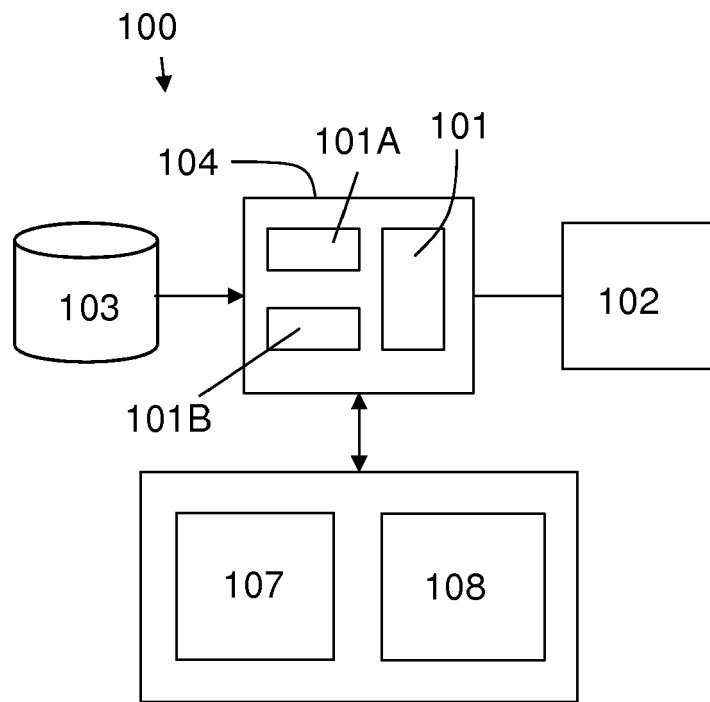


Figure 2

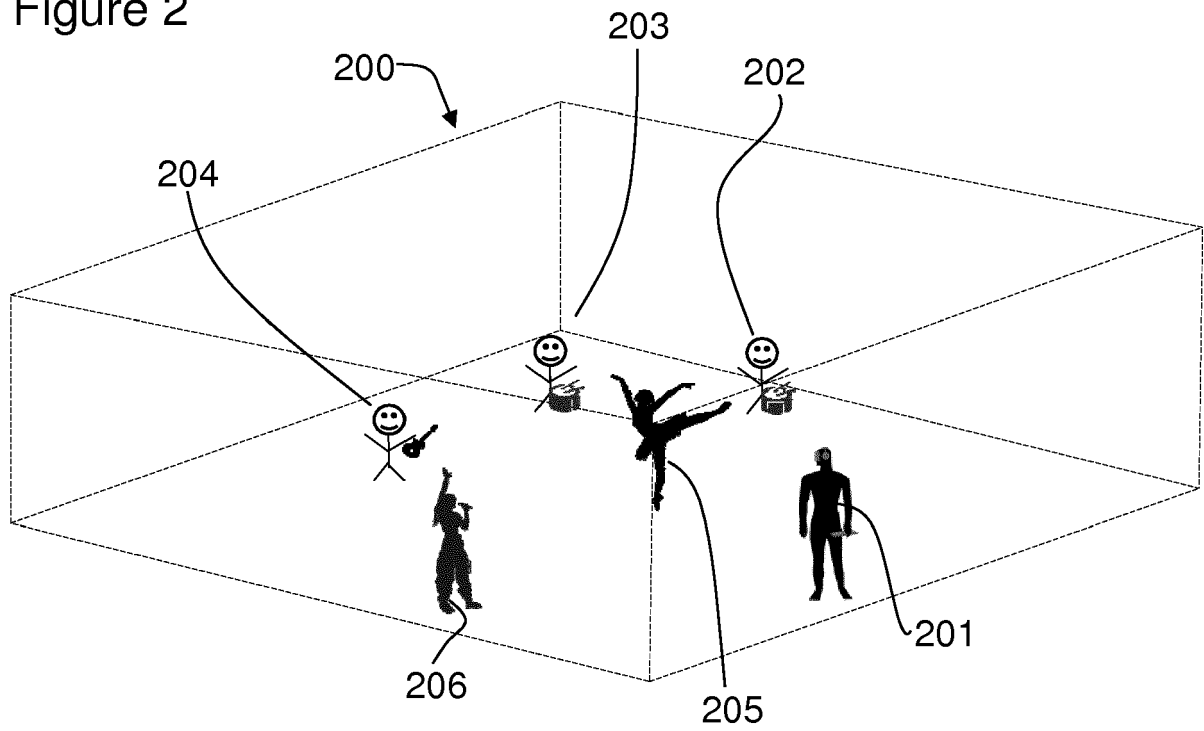


Figure 3

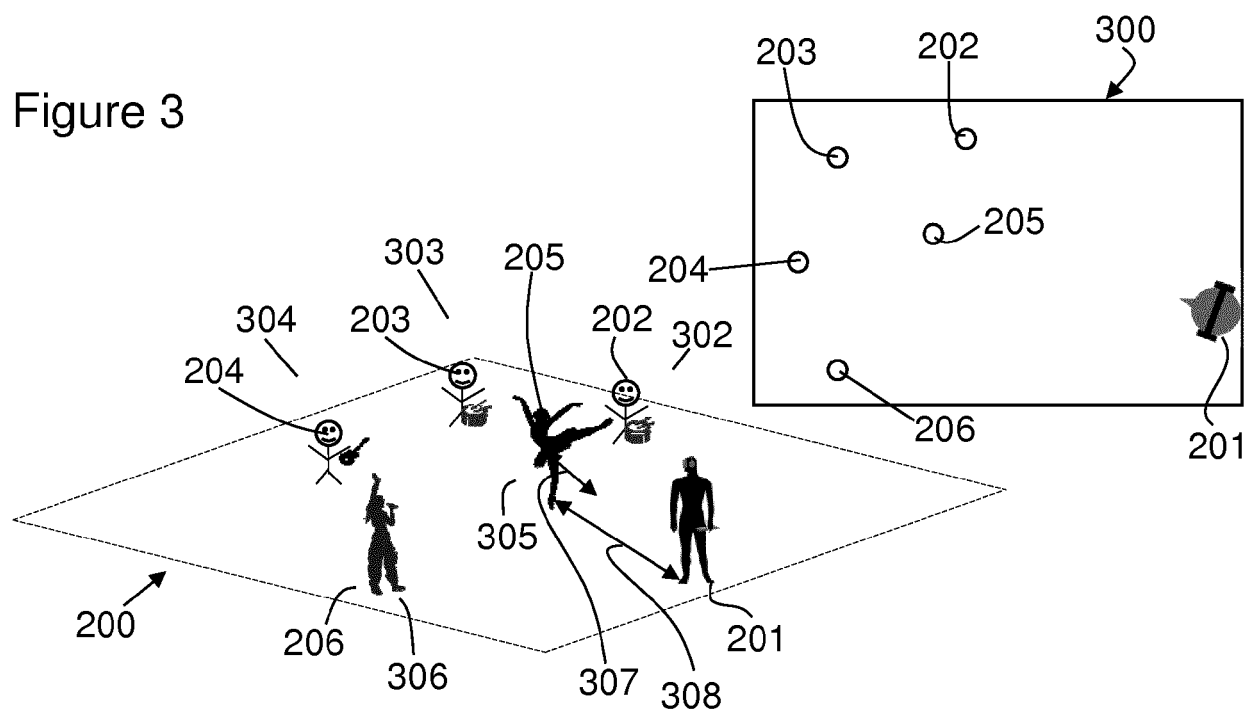


Figure 4

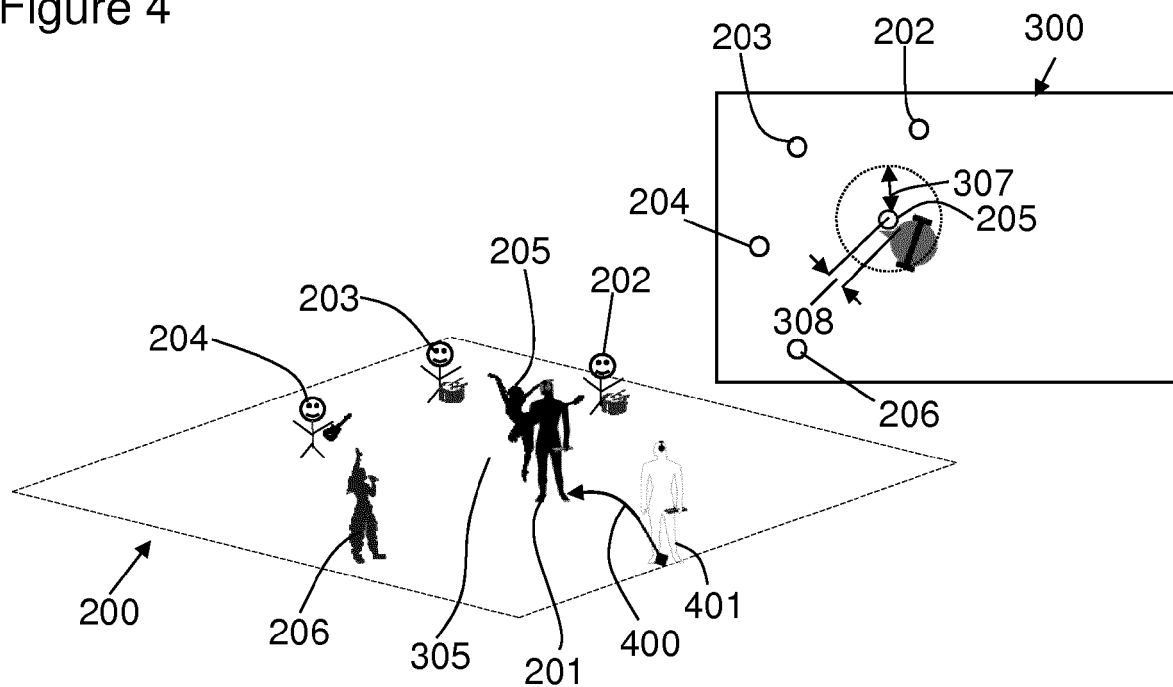


Figure 5

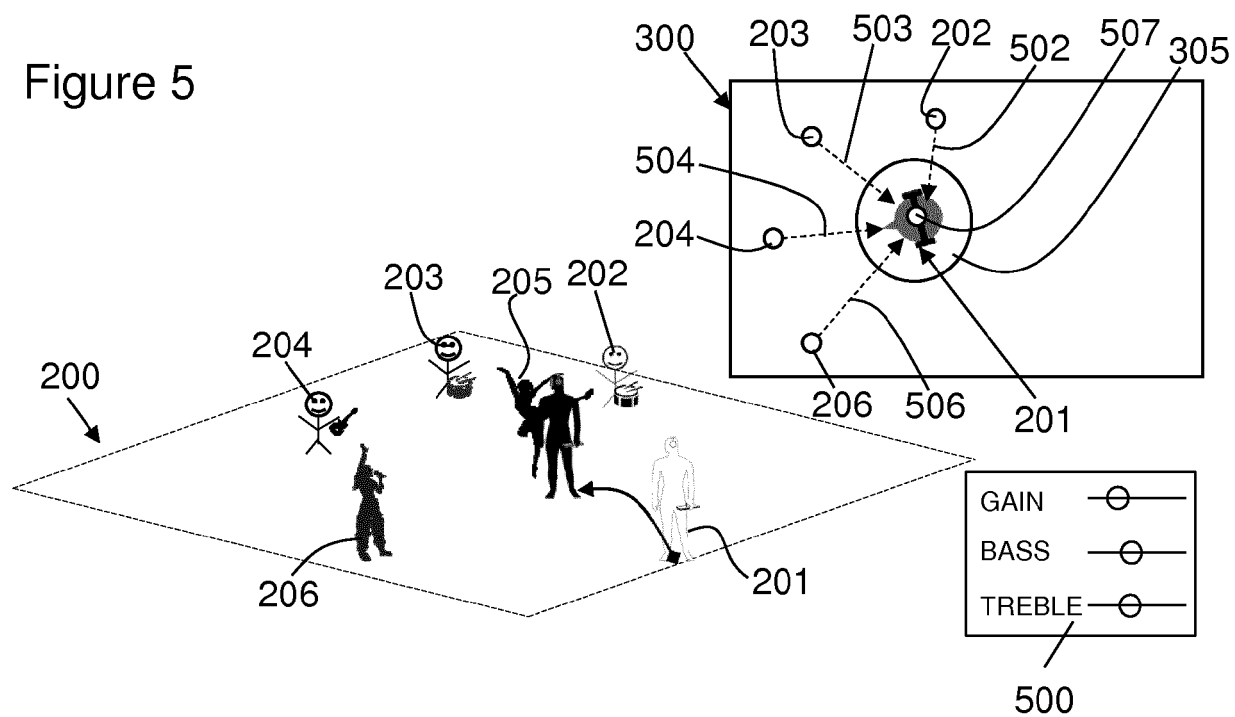


Figure 6

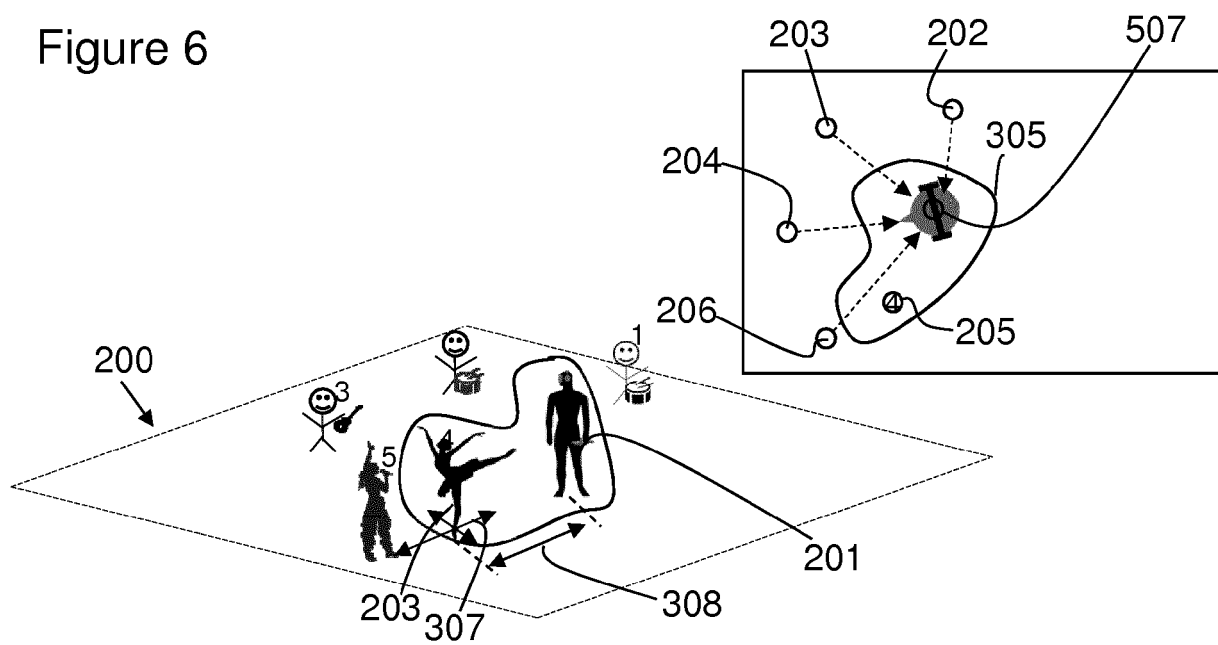


Figure 7

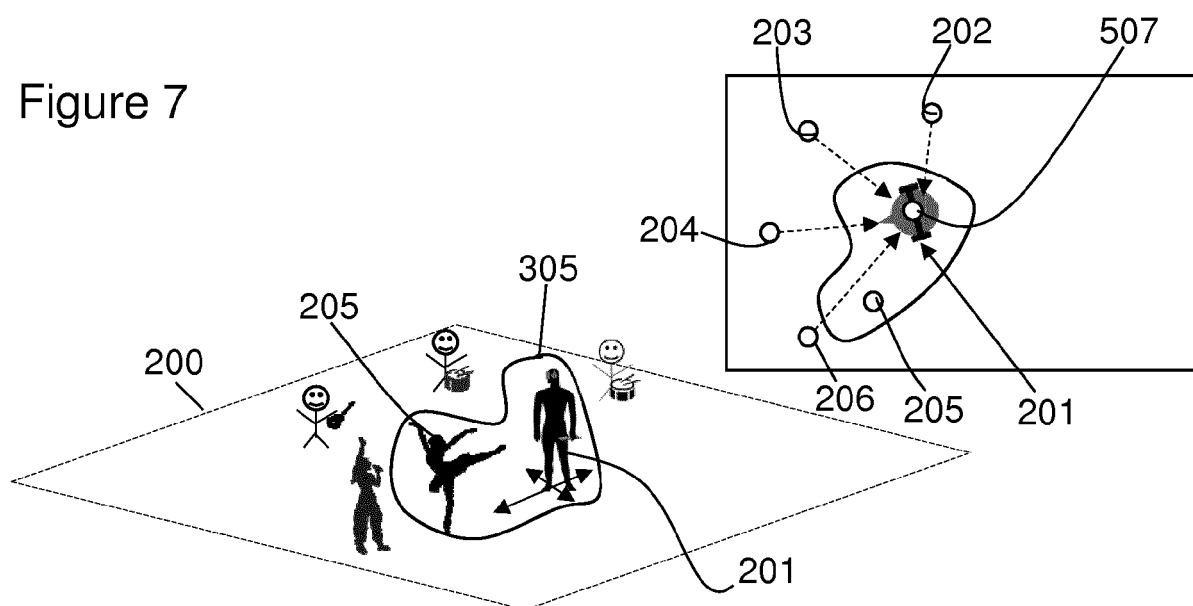


Figure 8

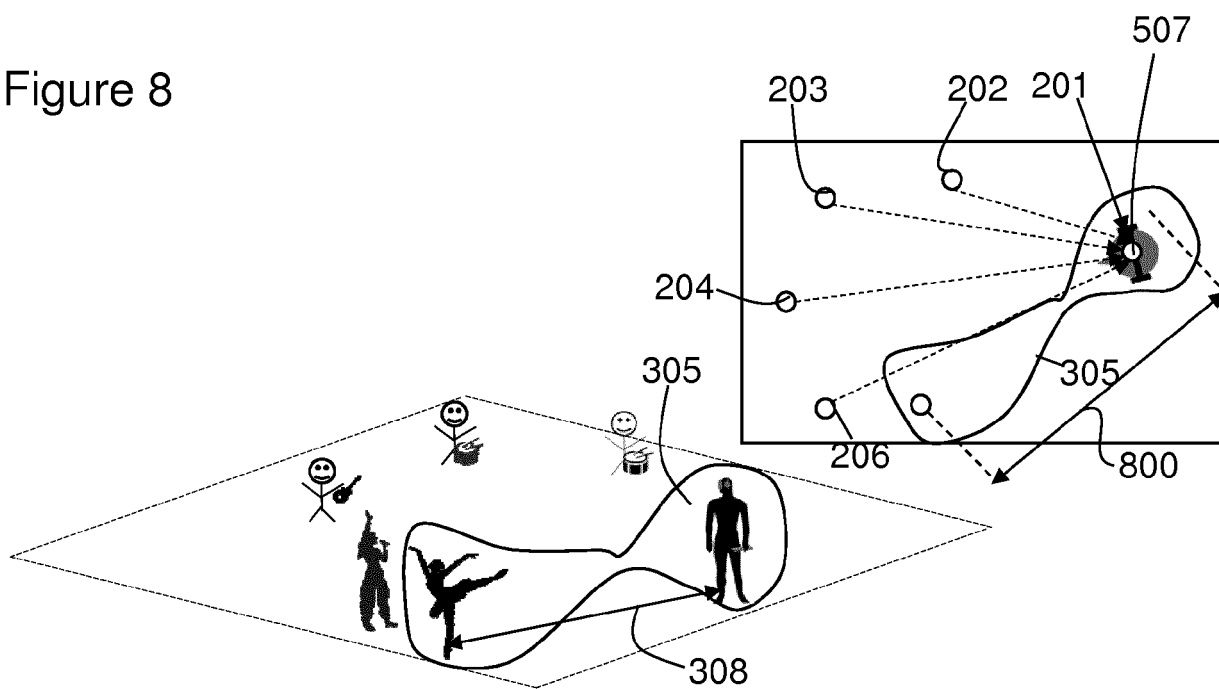


Figure 9

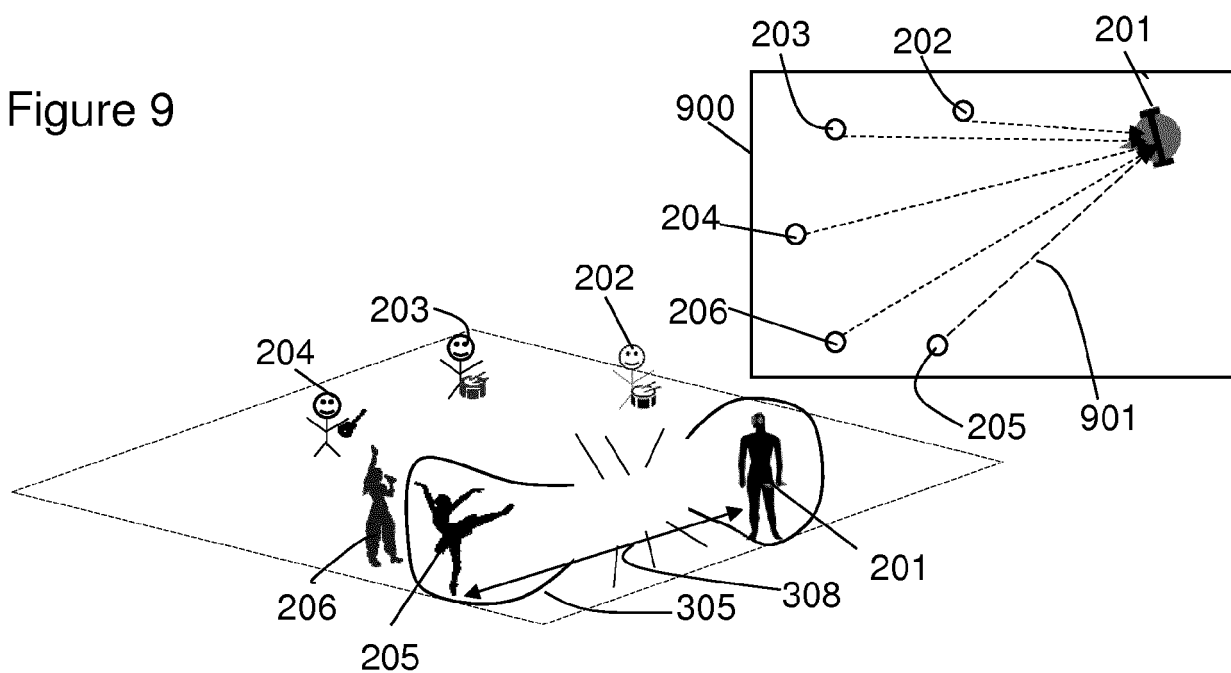


Figure 10

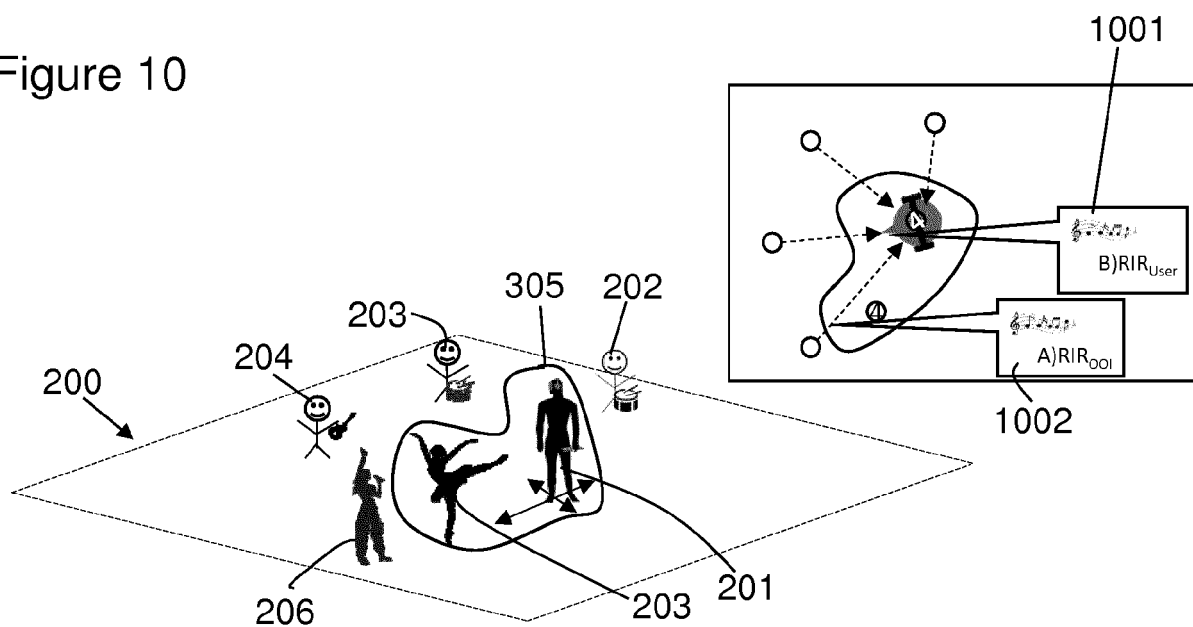


Figure 11

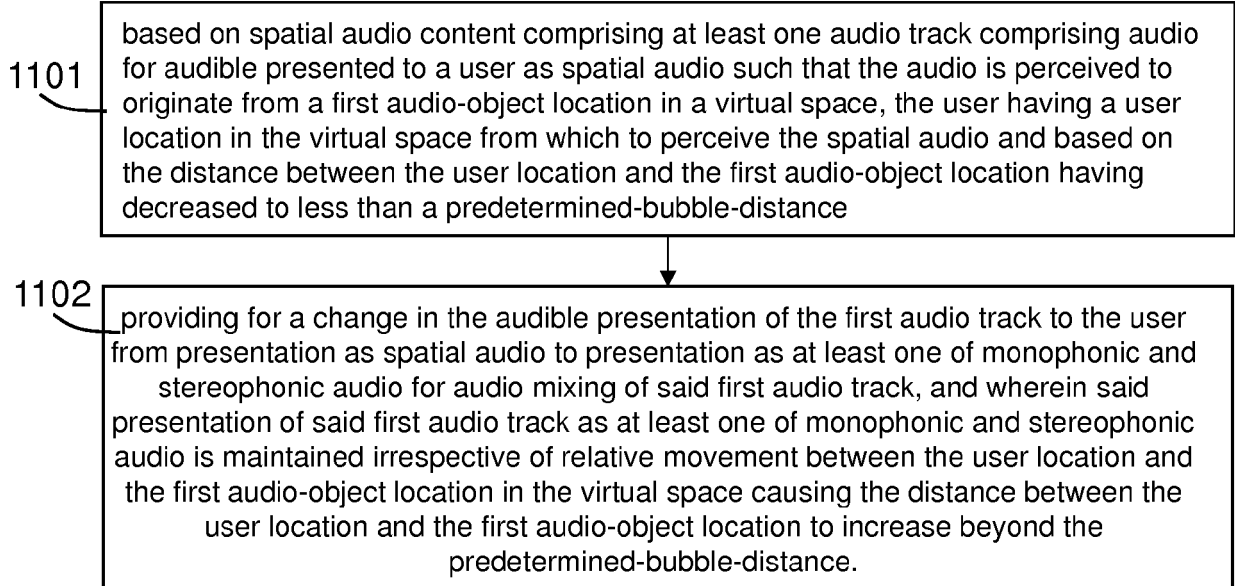
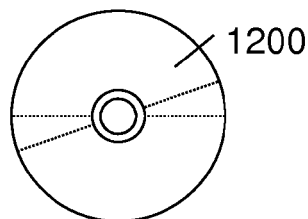


Figure 12





EUROPEAN SEARCH REPORT

 Application Number
EP 17 19 2113

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 2009/282335 A1 (ALEXANDERSSON PETTER [SE]) 12 November 2009 (2009-11-12) * paragraph [0007] - paragraph [0009] * * paragraph [0012] - paragraph [0016] * * paragraph [0019] - paragraph [0020] * * paragraphs [0040], [0045], [0060], [0065] * * paragraph [0068] - paragraph [0072] * * paragraph [0076] - paragraph [0079] * * paragraph [0086] - paragraph [0090] * * paragraphs [0093], [0095], [0102] * * figures 2,3,5,7 *	1-15	INV. H04S7/00 G06F3/01
A	US 2012/230534 A1 (PAN DAVIS Y [US] ET AL) 13 September 2012 (2012-09-13) * paragraph [0003] * * paragraph [0012] - paragraph [0016] * * paragraph [0019] * * figures 2,3 *	4	
A	US 2014/226842 A1 (SHENOY RAVI [IN] ET AL) 14 August 2014 (2014-08-14) * paragraphs [0080], [0113], [0124] * * figures 3, 4, 6-9, 11 *	4,6-8	TECHNICAL FIELDS SEARCHED (IPC) H04S G06F H04R
A	US 2007/172086 A1 (DICKINS GLEN N [AU] ET AL) 26 July 2007 (2007-07-26) * paragraph [0179] - paragraph [0180] * * figures 27,28 *	9	
A	US 2016/316308 A1 (SANDLER MARK BRIAN [GB] ET AL) 27 October 2016 (2016-10-27) * paragraphs [0053], [0054], [0056] * * paragraph [0058] - paragraph [0061] * * figures 3,4 *	10	
The present search report has been drawn up for all claims			
Place of search The Hague		Date of completion of the search 26 February 2018	Examiner Valenzuela, Miriam
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	

EPO FORM 1503 03.82 (P04C01)

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

EP 17 19 2113

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.

26-02-2018

10

15

20

25

30

35

40

45

50

55

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009282335 A1	12-11-2009	US 2009282335 A1	12-11-2009
		WO 2009136227 A1	12-11-2009
US 2012230534 A1	13-09-2012	CN 101796855 A	04-08-2010
		CN 105188012 A	23-12-2015
		EP 2181552 A1	05-05-2010
		JP 5314021 B2	16-10-2013
		JP 2010536274 A	25-11-2010
		US 2009060208 A1	05-03-2009
		US 2012230534 A1	13-09-2012
		WO 2009029471 A1	05-03-2009
US 2014226842 A1	14-08-2014	EP 2716021 A1	09-04-2014
		US 2014226842 A1	14-08-2014
		WO 2012164153 A1	06-12-2012
US 2007172086 A1	26-07-2007	DK 1025743 T3	05-08-2013
		EP 1025743 A1	09-08-2000
		JP 4477081 B2	09-06-2010
		JP 4627880 B2	09-02-2011
		JP 2001517050 A	02-10-2001
		JP 2009010995 A	15-01-2009
		KR 20010030608 A	16-04-2001
		US 2007172086 A1	26-07-2007
		US 2007223751 A1	27-09-2007
		WO 9914983 A1	25-03-1999
US 2016316308 A1	27-10-2016	EP 2319253 A1	11-05-2011
		US 2011208331 A1	25-08-2011
		US 2015339099 A1	26-11-2015
		US 2016316308 A1	27-10-2016
		WO 2010020788 A1	25-02-2010