



(11)

EP 4 383 754 A1

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
12.06.2024 Bulletin 2024/24

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

(21) Application number: **22211764.0**

(52) Cooperative Patent Classification (CPC):
H04S 7/303; H04S 2400/11; H04S 2400/13

(22) Date of filing: **06.12.2022**

(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 ME MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA
Designated Validation States:
KH MA MD TN

(72) Inventors:
• **JELFS, Sam Martin**
Eindhoven (NL)
• **KOPPENS, Jeroen Gerardus Henricus**
Eindhoven (NL)
• **OUWELTJES, Okke**
Eindhoven (NL)

(71) Applicant: **Koninklijke Philips N.V.**
5656 AG Eindhoven (NL)

(74) Representative: **Philips Intellectual Property & Standards**
High Tech Campus 52
5656 AG Eindhoven (NL)

(54) **AN AUDIO APPARATUS AND METHOD OF RENDERING THEREFOR**

(57) An audio apparatus comprises a first receiver (501) receiving audio data for sources of a scene comprising acoustic environments divided by acoustically attenuating boundaries. A second receiver (503) receives metadata comprising a position indication for a transfer region of an acoustically attenuating boundary between a first acoustic environment and a second acoustic environment with the transfer region having less acoustic attenuation than the boundary. The metadata further includes an energy transfer indication for the transfer region indicating the proportion of energy at the transfer region from an omnidirectional point audio source at a reference position. A renderer (507) renders an audio signal for a position in the first acoustic environment by rendering audio data for a source of the second acoustic environment and adapting a level of the audio depending on the energy transfer indication. The level may be adapted based on a difference between a source position and the reference position.

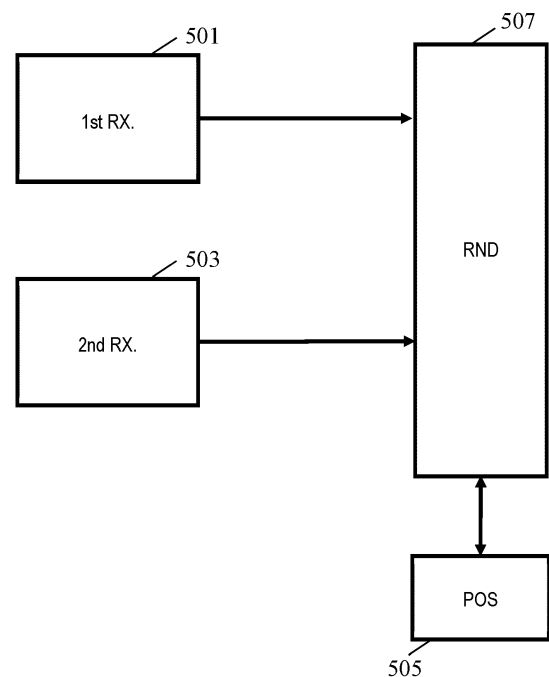


FIG. 5

Description

FIELD OF THE INVENTION

5 **[0001]** The invention relates to an apparatus and method for rendering an audio signal, and in particular, but not exclusively, for rendering audio for a multi-room scene as part of e.g. an eXtended Reality experience.

BACKGROUND OF THE INVENTION

10 **[0002]** The variety and range of experiences based on audiovisual content have increased substantially in recent years with new services and ways of utilizing and consuming such content continuously being developed and introduced. In particular, many spatial and interactive services, applications and experiences are being developed to give users a more involved and immersive experience.

15 **[0003]** Examples of such applications are eXtended Reality (XR) which is a common term referring to Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) applications, which are rapidly becoming mainstream, with a number of solutions being aimed at the consumer market. A number of standards are also under development by a number of standardization bodies. Such standardization activities are actively developing standards for the various aspects of VR/AR/MR systems including e.g. streaming, broadcasting, rendering, etc.

20 **[0004]** VR applications tend to provide user experiences corresponding to the user being in a different world/ environment/ scene whereas AR (including Mixed Reality MR) applications tend to provide user experiences corresponding to the user being in the current environment but with additional information or virtual objects or information being added. Thus, VR applications tend to provide a fully immersive synthetically generated world/ scene whereas AR applications tend to provide a partially synthetic world/ scene which is overlaid the real scene in which the user is physically present. However, the terms are often used interchangeably and have a high degree of overlap. In the following, the term eXtended Reality/ XR will be used to denote both Virtual Reality and Augmented/ Mixed Reality.

25 **[0005]** As an example, a service being increasingly popular is the provision of images and audio in such a way that a user is able to actively and dynamically interact with the system to change parameters of the rendering such that this will adapt to movement and changes in the user's position and orientation. A very appealing feature in many applications is the ability to change the effective viewing position and viewing direction of the viewer, such as for example allowing the viewer to move and "look around" in the scene being presented.

30 **[0006]** Such a feature can specifically allow a virtual reality experience to be provided to a user. This may allow the user to (relatively) freely move about in a virtual scene and dynamically change his position and where he is looking. Typically, such virtual reality applications are based on a three-dimensional model of the scene with the model being dynamically evaluated to provide the specific requested view. This approach is well known from e.g. game applications, such as in the category of first person shooters, for computers and consoles.

35 **[0007]** It is also desirable, in particular for virtual reality applications, that the image being presented is a three-dimensional image, typically presented using a stereoscopic display. Indeed, in order to optimize immersion of the viewer, it is typically preferred for the user to experience the presented scene as a three-dimensional scene. Indeed, a virtual reality experience should preferably allow a user to select his/her own position, viewpoint, and moment in time relative to a virtual world.

40 **[0008]** In addition to the visual rendering, most XR applications further provide a corresponding audio experience. In many applications, the audio preferably provides a spatial audio experience where audio sources are perceived to arrive from positions that correspond to the positions of the corresponding objects in the visual scene. Thus, the audio and video scenes are preferably perceived to be consistent and with both providing a full spatial experience.

45 **[0009]** For example, many immersive experiences are provided by a virtual audio scene being generated by headphone reproduction using binaural audio rendering technology. In many scenarios, such headphone reproduction may be based on headtracking such that the rendering can be made responsive to the user's head movements. This highly increases the sense of immersion.

50 **[0010]** An important feature for many applications is that of how to generate and/or distribute audio that can provide a natural and realistic perception of the audio scene. For example, when generating audio for a virtual reality application, it is important that not only are the desired audio sources generated but also that these are generated to provide a realistic perception of the audio environment including damping, reflection, coloration etc.

55 **[0011]** For room/ environment acoustics, reflections of sound waves off walls, floor, ceiling, objects etc. cause delayed and attenuated (typically frequency dependent) versions of the sound source signal to reach the listener (i.e. the user for a XR system) via different paths. The combined effect can be modelled by an impulse response which may be referred to as a Room Impulse Response (RIR).

[0012] As illustrated in FIG. 1, a RIR typically consists of a direct sound that depends on distance of the sound source to the listener, followed by a reflection portion that characterizes the acoustic properties of the room. The size and shape

of the room, the position of the sound source and listener in the room and the reflective properties of the room's surfaces all play a role in the characteristics of this reverberant portion.

[0013] The reflective portion can be broken down into two temporal regions, usually overlapping. The first region contains so-called early reflections, which represent isolated reflections of the sound source on walls or obstacles inside the room prior to reaching the listener. As the time lag/ (propagation) delay increases, the number of reflections present in a fixed time interval increases and the paths may include secondary or higher order reflections (e.g. reflections may be off several walls or both walls and ceiling etc).

[0014] The second region referred to as the reverberant portion is the part where the density of these reflections increases to a point where they cannot anymore be isolated by the human brain. This region is typically called the diffuse reverberation, late reverberation, or reverberation tail, or simply reverberation.

[0015] The RIR contains cues that give the auditory system information about the distance of the source, and of the size and acoustical properties of the room. The energy of the reverberant portion in relation to that of the anechoic portion largely determines the perceived distance of the sound source. The level and delay of the earliest reflections may provide cues about how close the sound source is to a wall, and the filtering by anthropometrics may strengthen the assessment of the specific wall, floor or ceiling.

[0016] The density of the (early-) reflections contributes to the perceived size of the room. The time that it takes for the reflections to drop 60 dB in energy level, indicated by the reverberation time T_{60} , is a frequently used measure for how fast reflections dissipate in the room. The reverberation time provides information on the acoustical properties of the room, such as specifically whether the walls are very reflective (e.g. bathroom) or there is much absorption of sound (e.g. bedroom with furniture, carpet and curtains).

[0017] Furthermore, RIRs may be dependent on a user's anthropometric properties when it is a part of a binaural room impulse response (BRIR), due to the RIR being filtered by the head, ears and shoulders; i.e. the head related impulse responses (HRIRs).

[0018] As the reflections in the late reverberation cannot be differentiated and isolated by a listener, they are often simulated and represented parametrically with, e.g., a parametric reverberator using a feedback delay network, as in the well-known Jot reverberator.

[0019] For early reflections, the direction of incidence and distance dependent delays are important cues to humans to extract information about the room and the relative position of the sound source. Therefore, the simulation of early reflections must be more explicit than the late reverberation. In efficient acoustic rendering algorithms, the early reflections are therefore simulated differently and separately from the later reverberation. A well-known method for early reflections is to mirror the sound sources in each of the room's boundaries to generate a virtual sound source that represents the reflection.

[0020] For early reflections, the position of the user and/or sound source with respect to the boundaries (walls, ceiling, floor) of a room is relevant, while for the late reverberation, the acoustic response of the room is diffuse and therefore tends to be homogeneous throughout the room. This allows simulation of late reverberation to often be more computationally efficient than early reflections.

[0021] Two main properties of the late reverberation are the slope and amplitude of the impulse response for times above a given threshold. These properties tend to be strongly frequency dependent in natural rooms. Often the reverberation is described using parameters that characterize these properties.

[0022] An example of parameters characterizing a reverberation is illustrated in FIG. 2. Examples of parameters that are traditionally used to indicate the slope and amplitude of the impulse response corresponding to diffuse reverberation include the known T_{60} value and the reverb level/ energy. More recently other indications of the amplitude level have been suggested, such as specifically parameters indicating the ratio between diffuse reverberation energy and the total emitted source energy.

[0023] Specifically, a Diffuse to Source Ratio, DSR, may be used to express the amount of diffuse reverberation energy or level of a source received by a user as a ratio of total emitted energy of that source. The DSR may represent the ratio between emitted source energy and a diffuse reverberation property, such as specifically the energy or the (initial) level of the diffuse reverberation signal:

$$DSR = \frac{\text{diffuse reverb energy}}{\text{total emitted energy}}$$

[0024] Henceforth this will be referred to as DSR (Diffuse-to-Source Ratio).

[0025] Such known approaches tend to provide efficient descriptions of audio propagation in a room and tend to lead to rendering of audio that is perceived as natural for the room in which the listener is (virtually) present.

[0026] However, whereas conventional approaches for representing and rendering sound in a room or individual acoustic environment may provide a suitable perception in many embodiments, it tends to not be suitable for all possible

scenarios. In particular, for audio scenes that may include different acoustic environments/ regions/ rooms, the generated audio signal using e.g. the described reverberation approach may not lead to an optimal experience or perception. It may typically lead to situations where the audio from other rooms is not sufficiently or accurately represented by the rendered audio resulting in a perception that may not fully reflect the acoustic scenario and scene.

[0027] Indeed, typically, the reverberation is modelled for a listener inside the room taking into account the properties of the room. When the listener is outside the room, or in a different room, the reverberator may be turned off or reconfigured for the other room's properties. Even when multiple reverberators can be run in parallel, the output of the reverberators typically is a diffuse binaural (or multi-loudspeaker) signal intended to be presented to the listener as being inside the room. However, such approaches tend to result in audio being generated which is often not perceived to be an accurate representation of the actual environment. This may for example lead to a perceived disconnect or even conflict between the visual perception of a scene and the associated audio being rendered.

Thus, whereas typical approaches for rendering audio may in many embodiments be suitable for rendering the audio of an environment, they tend to be suboptimal in some scenarios, including in particular when rendering audio for scenes that include different acoustic rooms or environments.

In particular, approaches for representing and rendering audio in one acoustic environment that originates from other acoustic environments tend to be suboptimal and/or be relatively impractical, including potentially requiring excessive computational resource or being relatively complex. Further, in applications where audio representing a multi acoustic environment (specifically a multi room scene) tends to be suboptimal in terms of not providing easy to use and low data rate information allowing multi acoustic environments to be represented and rendered.

[0028] Hence, an improved approach for rendering audio for a scene would be advantageous. In particular, an approach that allows improved operation, increased flexibility, reduced complexity, facilitated implementation, an improved audio experience, improved audio quality, reduced computational burden, improved representation of multi-acoustic environments, facilitated rendering, improved of audio from multiple acoustic environments, improved performance for virtual/mixed/ augmented reality applications, increased processing flexibility, improved representation and rendering of audio and audio properties of multiple rooms or other acoustic environments, a more natural sounding audio rendering, improved audio rendering for multi-room scenes, and/or improved performance and/or operation would be advantageous.

SUMMARY OF THE INVENTION

[0029] Accordingly, the Invention seeks to preferably mitigate, alleviate or eliminate one or more of the above-mentioned disadvantages singly or in any combination.

[0030] According to an aspect of the invention there is provided an audio apparatus: a first receiver arranged to receive audio data for audio sources of a scene comprising multiple acoustic environments, the acoustic environments being divided by acoustically attenuating boundaries; a second receiver arranged to receive metadata for the audio data, the metadata comprising: a position indication for at least a first transfer region of a first acoustically attenuating boundary between a first acoustic environment and a second acoustic environment, the first transfer region being a region of the first acoustically attenuating boundary having lower attenuation than an average attenuation of the first acoustically attenuating boundary outside of transfer regions; and an energy transfer indication for the first transfer region, the energy transfer indication being indicative of a proportion of energy at the first transfer region from an omnidirectional point audio source at a reference position, the reference position being a relative position with respect to the first transfer region; a renderer arranged to render an audio signal for a listening position in the first acoustic environment, the rendering including generating a first audio component by rendering audio data for a first audio source of the second acoustic environment and adapting a level of the first audio component in dependence on the energy transfer indication.

The approach may allow an audio signal being generated that provides an improved user experience for audio scenes with multiple acoustic environments, and often a more realistic and naturally sounding audio experience. The approach may allow an improved audio rendering for e.g. multi-room scenes. A more natural and/or accurate audio perception of a scene may be achieved in many scenarios.

[0031] The approach may provide improved and/or facilitated rendering of audio representing audio sources in other acoustic environments or rooms. The rendering of the audio signal may often be achieved with reduced complexity and reduced computational resource requirements.

[0032] The approach may provide improved, increased, and/or facilitated flexibility and/or adaptation of the processing and/or the rendered audio.

[0033] The approach may further allow improved and/or facilitated representation of multi-acoustic environment sound propagation data or properties. It may provide an improved and/or facilitated representation of sound propagation characteristics of transfer regions (such as portals) in acoustically attenuating boundaries.

[0034] The approach may in many embodiments and scenarios provide an efficient and low complexity approach for accurately representing acoustic properties for transfer region in acoustically attenuating boundaries, and for determining and rendering appropriate audio propagating into a given room from another room via such a transfer region.

[0035] An energy transfer indication indicating an energy transfer from a source to a transfer region energy is equivalent to an energy attenuation indication indicating an attenuation of the energy of the source signal/ audio at a transfer region. An increasing attenuation is indicative of a reduced proportion of audio energy from the source reaching the transfer region, corresponding to a reduced energy transfer. A decreasing attenuation is indicative of an increased proportion of audio energy from the source reaching the transfer region, corresponding to an increased energy transfer. The terms energy attenuation and energy transfer may thus be used interchangeable with the understanding that one is a monotonically decreasing function of the other.

[0036] The audio energy (or just energy) may specifically be represented by a level, amplitude, power, or time averaged energy measure.

[0037] An acoustically attenuating boundary may attenuate sound propagation through the acoustically attenuating boundary from one acoustic environment to the other acoustic environment. In many embodiments and scenarios the attenuation of the acoustically attenuating boundary outside of transfer regions may be no less than 3dB, 6 dB, 10dB, or even 20dB. The attenuation for a transfer region in an acoustically attenuating boundary may in many embodiments be no less than 3dB, 6 dB, 10dB, or even 20dB lower than the (average) attenuation of the acoustically attenuating boundary outside the transfer region(s).

[0038] The first acoustic environment and the second acoustic environment are different acoustic environments. The first audio source may be an audio source of the second acoustic environment, and may for example be an audio source corresponding to a diffuse reverberation sound, or a point source. The energy transfer indication may be frequency dependent. The energy transfer indication may be a nominal energy transfer indication. The reference position may be a normalized/ standardized/ predetermined position for the first transfer region.

[0039] In accordance with an optional feature of the invention, the renderer is arranged to adapt the level of the first audio component in response to a position of the first audio source relative to the reference position.

[0040] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene. It may in many embodiments allow a low complexity yet accurate determination of the energy level at the first transfer region from the first audio source.

[0041] In some embodiments, the energy transfer attenuation for the first pair of transfer regions is indicative of the proportion of audio energy incident on the second transfer region that propagates to exit the first transfer region (into the first acoustic environment).

[0042] In accordance with an optional feature of the invention, the renderer is arranged to adapt the level of the first audio component in dependence on a difference between a reference distance from the reference position to the first transfer region and a distance from a position of the first audio source to the first transfer region.

[0043] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene. It may in many embodiments allow a low complexity yet accurate determination of the energy level at the first transfer region from the first audio source.

[0044] In accordance with an optional feature of the invention, the renderer is arranged to adapt the level of the first audio component in dependence on a difference between a direction from the reference position to the first transfer region and a direction from the first audio source to the first transfer region.

[0045] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene. It may in many embodiments allow a low complexity yet accurate determination of the energy level at the first transfer region from the first audio source.

[0046] In accordance with an optional feature of the invention, the metadata further comprises data describing a directivity of sound radiation from the first audio source and the renderer is arranged to adapt the level of the first audio component in dependence on the directivity.

[0047] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene. It may in many embodiments allow a low complexity yet accurate determination of the energy level at the first transfer region from the first audio source.

[0048] In some embodiments, the renderer is arranged to generate a combined energy transfer attenuation by combining the energy transfer attenuation for the first pair of transfer regions and an energy transfer attenuation for a second pair of transfer regions comprising a third transfer region of a boundary of a third acoustic environment and the second transfer region; and to generate the first audio component by rendering a second audio source of the third acoustic environment in dependence on the combined energy transfer attenuation.

[0049] In accordance with an optional feature of the invention, the renderer is arranged to scale the level of the first audio component in dependence on a relative directivity gain for the first audio source in a direction from the first audio source to the first transfer region, the relative directivity gain being indicative of a gain relative to an omnidirectional source.

[0050] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene. It may in many embodiments allow a low complexity yet accurate determination of the energy level at the first transfer region from the first audio source.

[0051] In accordance with an optional feature of the invention, the first audio source represents audio reaching the second acoustic environment from a third acoustic environment via a second transfer region of a second boundary separating the third acoustic environment from the second acoustic environment.

[0052] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene. The approach may in particular allow efficient representation of audio propagation through multiple intermediate acoustic environments.

[0053] In accordance with an optional feature of the invention, the metadata further comprises: energy transfer parameters, each energy transfer parameter indicating an energy attenuation between a pair of transfer regions, the energy attenuation for a pair of transfer regions being indicative of a proportion of audio energy at one transfer region of the pair of transfer regions propagating to the other transfer region of the pair of transfer regions; and the renderer is arranged to render a second audio component of the audio signal by rendering a second audio source of a third acoustic environment in dependence on an energy attenuation for a pair of transfer regions comprising a transfer region of an acoustically attenuating boundary of the first acoustic environment and a transfer region of a second acoustically attenuating boundary being a boundary of the second acoustic environment.

[0054] This may provide particularly advantageous performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene. The approach may in particular allow improved rendering of audio reaching the listener via paths including a plurality of acoustic environments and transfer regions.

[0055] In accordance with an optional feature of the invention, the metadata includes a coupling coefficient for the first transfer region, and the renderer is arranged to render the first audio component as originating from an audio source at a position proximal to the first transfer region and in dependence on the coupling factor.

[0056] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene. The approach may in particular allow efficient rendering of audio from other acoustic environment reaching the listening acoustic environment via a coupled area, such as e.g. a window or similar.

[0057] In accordance with an optional feature of the invention, the renderer is arranged to render the first audio component as a reverberation audio component of the first acoustic environment.

[0058] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene. The approach is particularly advantageous for generating a reverberant/ diffuse/ background sound reflecting audio sources in other acoustic environments.

[0059] In accordance with an optional feature of the invention, the renderer is arranged to generate a reverberation audio signal for the second environment to include a first component from the first audio source, the renderer further being arranged to determine an energy loss estimate as a proportion of energy of the first audio source that reaches the first transfer region, the proportion of energy being determined in dependence on the energy transfer indication and a position of the first audio source relative to the reference position; and the renderer being arranged to reduce a level of the reverberation audio signal by an amount depending on the energy loss estimate.

[0060] This may provide particularly advantageous performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene.

[0061] In accordance with an optional feature of the invention, the energy transfer indication reflects a proportion of a sphere that is covered by the first transfer region, the sphere being centered on the reference position and having a radius corresponding to a distance from the reference position to the first transfer region.

[0062] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene.

[0063] In accordance with an optional feature of the invention, the renderer is arranged to render the first audio component as a non-direct audio component.

[0064] This may provide improved performance and/or facilitated implementation in many scenarios. It may assist in providing an improved user experience when rendering audio for a multi-acoustic environment scene.

[0065] According to an aspect of the invention there is provided a method of rendering an audio signal, the method comprising: receiving audio data for audio sources of a scene comprising multiple acoustic environments, the acoustic environments being divided by acoustically attenuating boundaries; receiving metadata for the audio data, the metadata comprising: a position indication for at least a first transfer region of a first acoustically attenuating boundary between a first acoustic environment and a second acoustic environment, the first transfer region being a region of the first acoustically attenuating boundary having lower attenuation than an average attenuation of the first acoustically attenuating

boundary outside of transfer regions; and an energy transfer indication for the first transfer region, the energy transfer indication being indicative of a proportion of energy at the first transfer region from an omnidirectional point audio source at a reference position, the reference position being a relative position with respect to the first transfer region; rendering an audio signal for a listening position in the first acoustic environment, the rendering including generating a first audio component by rendering audio data for a first audio source of the second acoustic environment and adapting a level of the first audio component in dependence on the energy transfer indication.

[0066] According to an aspect of the invention there is provided an audio data signal comprising: audio data for audio sources of a scene comprising multiple acoustic environments, the acoustic environments being divided by acoustically attenuating boundaries; metadata for the audio data, the metadata comprising: a position indication for at least a first transfer region of a first acoustically attenuating boundary between a first acoustic environment and a second acoustic environment, the first transfer region being a region of the first acoustically attenuating boundary having lower attenuation than an average attenuation of the first acoustically attenuating boundary outside of transfer regions; and an energy transfer indication for the first transfer region, the energy transfer indication being indicative of a proportion of energy at the first transfer region from an omnidirectional point audio source at a reference position, the reference position being a relative position with respect to the first transfer region.

[0067] These and other aspects, features and advantages of the invention will be apparent from and elucidated with reference to the embodiment(s) described hereinafter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0068] Embodiments of the invention will be described, by way of example only, with reference to the drawings, in which

FIG. 1 illustrates an example of a room impulse response;

FIG. 2 illustrates an example of a room impulse response;

FIG. 3 illustrates an example of elements of virtual reality system;

FIG. 4 illustrates an example of a scene with three rooms;

FIG. 5 illustrates an example of an audio apparatus for generating an audio signal in accordance with some embodiments of the invention;

FIG. 6 illustrates an example of a scene with multiple rooms separated by walls with sound portals;

FIG. 7 illustrates an example of a sound propagation from an audio source towards a wall with a sound portal;

FIG. 8 illustrates an example of a scene with multiple rooms separated by walls with sound portals;

FIG. 9 illustrates an example of a sound propagation from an audio source towards a wall with a sound portal;

FIG. 10 illustrates an example of a sound propagation from an audio source towards a wall with a sound portal;

FIG. 11 illustrates an example of a scene with multiple rooms separated by walls with sound portals;

FIG. 12 illustrates an example of a scene with multiple rooms separated by walls with sound portals;

FIG. 13 illustrates an example of a scene with multiple rooms separated by walls with sound portals;

FIG. 14 illustrates an example of a scene with multiple rooms separated by walls with sound portals;

FIG. 15 illustrates an example of a scene with multiple rooms separated by walls with sound portals;

FIG. 16 illustrates an example of a scene with multiple rooms separated by walls with sound portals; and

FIG. 17 illustrates an example of room connection graphs for the examples of FIGs.13-16;

FIG. 18 illustrates some elements of a possible arrangement of a processor for implementing elements of an apparatus in accordance with some embodiments of the invention.

DETAILED DESCRIPTION OF SOME EMBODIMENTS OF THE INVENTION

[0069] The following description will focus on audio processing and rendering for an eXtended Reality application, but it will be appreciated that the described principles and concepts may be used in many other applications and embodiments.

[0070] Virtual experiences allowing a user to move around in a virtual world are becoming increasingly popular and services are being developed to satisfy such a demand.

[0071] In some systems, the VR application may be provided locally to a viewer by e.g. a standalone device that does not use, or even have any access to, any remote VR data or processing. For example, a device such as a games console may comprise a store for storing the scene data, input for receiving/ generating the viewer pose, and a processor for generating the corresponding images from the scene data.

[0072] In other systems, the VR application may be implemented and performed remotely from the viewer. For example, a device local to the user may detect/ receive movement/ pose data which is transmitted to a remote device that processes the data to generate the viewer pose. The remote device may then generate suitable view images and corresponding audio signals for the user pose based on scene data describing the scene. The view images and corresponding audio signals are then transmitted to the device local to the viewer where they are presented. For example, the remote device

may directly generate a video stream (typically a stereoscopic / 3D video stream) and corresponding audio stream which is directly presented by the local device. Thus, in such an example, the local device may not perform any VR processing except for transmitting movement data and presenting received video data.

[0073] In many systems, the functionality may be distributed across a local device and remote device. For example, the local device may process received input and sensor data to generate user poses that are continuously transmitted to the remote VR device. The remote VR device may then generate the corresponding view images and corresponding audio signals and transmit these to the local device for presentation. In other systems, the remote VR device may not directly generate the view images and corresponding audio signals but may select relevant scene data and transmit this to the local device, which may then generate the view images and corresponding audio signals that are presented. For example, the remote VR device may identify the closest capture point and extract the corresponding scene data (e.g. a set of object sources and their position metadata) and transmit this to the local device. The local device may then process the received scene data to generate the images and audio signals for the specific, current user pose. The user pose will typically correspond to the head pose, and references to the user pose may typically equivalently be considered to correspond to the references to the head pose.

[0074] In many applications, especially for broadcast services, a source may transmit or stream scene data in the form of an image (including video) and audio representation of the scene which is independent of the user pose. For example, signals and metadata corresponding to audio sources within the confines of a certain virtual room may be transmitted or streamed to a plurality of clients. The individual clients may then locally synthesize audio signals corresponding to the current user pose. Similarly, the source may transmit a general description of the audio environment including describing audio sources in the environment and acoustic characteristics of the environment. An audio representation may then be generated locally and presented to the user, for example using binaural rendering and processing.

[0075] FIG. 3 illustrates such an example of a VR system in which a remote VR client device 301 liaises with a VR server 303 e.g. via a network 305, such as the Internet. The server 303 may be arranged to simultaneously support a potentially large number of client devices 301.

[0076] The VR server 303 may for example support a broadcast experience by transmitting an image signal comprising an image representation in the form of image data that can be used by the client devices to locally synthesize view images corresponding to the appropriate user poses (a pose refers to a position and/or orientation). Similarly, the VR server 303 may transmit an audio representation of the scene allowing the audio to be locally synthesized for the user poses. Specifically, as the user moves around in the virtual environment, the image and audio synthesized and presented to the user is updated to reflect the current (virtual) position and orientation of the user in the (virtual) environment.

[0077] In many applications, such as that of FIG.3, it may thus be desirable to model a scene and generate an efficient image and audio representation that can be efficiently included in a data signal that can then be transmitted or streamed to various devices which can locally synthesize views and audio for different poses than the capture poses.

[0078] In some embodiments, a model representing a scene may for example be stored locally and may be used locally to synthesize appropriate images and audio. For example, an audio model of a room may include an indication of properties of audio sources that can be heard in the room as well as acoustic properties of the room. The model data may then be used to synthesize the appropriate audio for a specific position.

[0079] In many scenarios, the scene may include a plurality of different acoustic environments or regions that have different acoustic properties and specifically have different reverberation properties. Specifically, the scene may include or be divided into different acoustic environments/ regions that each have homogenous reverberation but between which the reverberation is different. For all positions within an acoustic environment/ region, a reverberation component of audio received at the positions may be homogeneous, and specifically may be substantially the same (except potentially for a gain difference). An acoustic environment/ region may be a set of positions for which a reverberation component of audio is homogeneous. An acoustic environment/ region may be a set of positions for which a reverberation component of the audio propagation impulse response for audio sources in the acoustic environment is homogeneous. Specifically, an acoustic environment/ region may be a set of positions for which a reverberation component of the audio propagation impulse response for audio sources in the acoustic environment has the same frequency dependent slope- and/or amplitude properties except for possibly a gain difference. Specifically, an acoustic environment/ region may be a set of positions for which a reverberation component of the audio propagation impulse response for audio sources in the acoustic environment is the same except for possibly a gain difference.

[0080] An acoustic environment/ region may typically be a set of positions (typically a 2D or 3D region) having the same rendering reverberation parameters. The reverberation parameters used for rendering a reverberation component may be the same for all positions in an acoustic environment/region. In particular, the same reverberation decay parameter (e.g. T_{60}) or Diffuse-to-Source Ratio, DSR, may apply to all positions within an acoustic environment/ region.

[0081] Impulse responses may be different between different positions in a room/ acoustic environment/ region due to the 'noisy' characteristic resulting from many various reflections of different orders causing the reverberation. However, even in such a case, the frequency dependent slope- and/or amplitude properties may be the same (except for possibly a gain difference), especially when represented by e.g. the reverberation time (T_{60}) or a reverberation coloration.

[0082] In many scenarios, acoustic environments may be separated by an acoustically attenuating boundary. Indeed, in many scenarios different acoustic environments may be determined by the presence of acoustically attenuating boundaries. An acoustically attenuating boundary may divide a region into different acoustic environments, and different acoustic environments may be formed by the presence of one or more acoustically attenuating boundaries. Two acoustic environments may be created by an acoustically attenuating boundary with the two acoustic environments being on opposite sides of the acoustically attenuating boundary. Such acoustically attenuating boundaries may for example be formed by walls or by any other structure that provides an acoustic attenuation that divides a space into multiple acoustic environments.

[0083] Acoustic environments/ regions may also be referred to as acoustic rooms or simply as rooms. A room may be considered an environment/ region as described above.

[0084] In many embodiments, a scene may be provided where acoustic rooms correspond to different virtual or real rooms between which a user may (e.g. virtually) move. An example of a scene with three rooms A, B, C is illustrated in FIG. 4. In the example, a user may move between the three rooms, or outside any room, through doorways and openings.

[0085] For a room to have substantial reverberation properties, it tends to represent a spatial region which is sufficiently bounded by geometric surfaces with wholly or partially reflecting properties such that a substantial part of the reflection in this room keep reflecting back into the region to generate a diffuse field of reflections in the region, having no significant directional properties. The geometric surfaces need not be aligned to any visual elements.

[0086] Audio rendering aimed at providing natural and realistic effects to a listener typically includes rendering of an acoustic scene. For many environments, this includes the representation and rendering of diffuse reverberation present in the environment, such as in a room where the listener is. The rendering and representation of such diffuse reverberation has been found to have a significant effect on the perception of the environment, such as on whether the audio is perceived to represent a natural and realistic environment.

[0087] In situations where the scene includes multiple rooms, the approach is typically to render the audio and reverberation only for the room in which the listener is present and to ignore any audio from other rooms. However, this tends to lead to audio experiences that are not perceived to be optimal and tends to not provide an optimal natural experience, particularly when the user transitions between rooms. Although some applications have been implemented to include rendering of audio from adjacent rooms, they have been found to be suboptimal. The audio from other rooms may in some embodiments have a substantial effect on the perceived audio scene. In particular, audio from other rooms may in many scenarios provide a significant contribution to the reverberation or diffuse (background) sound in a room and a suboptimal rendering of such audio may result in a degraded user experience.

[0088] In the following, advantageous approaches will be described for rendering an audio scene that includes multiple rooms.

[0089] FIG. 5 illustrates an example of an audio apparatus that is arranged to render an audio scene. The audio apparatus may receive audio data describing audio and audio sources in a scene (such as e.g. the one of FIG. 4). Based on the received audio data, the audio apparatus may render audio signals representing the scene for a given listening position. The rendered audio may include contributions both from audio generated in the room in which the listener is present as well as contributions from other neighbor, and typically adjacent, rooms.

[0090] The audio apparatus is arranged to generate an audio output signal that represents audio in the scene. Specifically, the audio apparatus may generate audio representing the audio perceived by a user moving around in the scene with a number of audio sources and with given acoustic properties. Each audio source is represented by an audio signal representing the sound from the audio source as well as metadata that may describe characteristics of the audio source (such as providing a level indication for the audio signal). In addition, metadata is provided to characterize the scene.

[0091] The renderer is in the example part of an audio apparatus which is arranged to receive audio data and metadata for a scene and to render audio representing at least part of the environment based on the received data.

[0092] The audio apparatus of FIG. 5 comprises a first receiver 501 which is arranged to receive audio data for audio sources in the scene, and thus it may receive audio data for multiple acoustic environments/ rooms that are divided by acoustically attenuating boundaries. The audio data may include audio data describing a plurality of audio signals from different audio sources in the scene. Typically, a number of e.g. point sources may be provided with audio data that reflects the sound to be rendered from those audio (point) sources. In some embodiments, audio data may also be provided for more diffuse audio sources, such as e.g. a background or ambient sound source, or sound sources with a spatial extent.

[0093] The audio apparatus comprises a second receiver 503 which is arranged to receive metadata for the audio data, and which specifically may receive metadata for the audio sources represented by the audio data. As will be described in more detail later, the metadata may include various information of the scene, including specifically related to different acoustic environments and boundaries between such.

[0094] The apparatus further comprises a position circuit 505 arranged to determine a listening position in the scene. The listening position typically reflects the (virtual) position of the user in the scene. For example, the position circuit

505 may be coupled to a user tracking device, such as a VR headset, an eye tracking device, a motion capture camera etc., and may from this receive user movement (including or possibly limited to head movement and/or eye movement) data. The position circuit 505 may from this data continuously determine a current listening position.

[0095] This listening position may alternatively be represented by or augmented with controller input with which a user can move or teleport the listening position in the scene.

[0096] It will be appreciated that many approaches and techniques are known and used for determining listening positions in a scene for various applications, and that any suitable approach may be used without detracting from the invention.

[0097] The audio apparatus comprises a renderer 507 which is arranged to generate an audio output signal representing the audio of the scene at the listening position. Typically, the audio signal may be generated to include audio components for a range of different audio sources in the scene. For example, point audio sources in the same room may be rendered as point audio sources having direct acoustic paths, reverberation components may be rendered, or generated etc.

[0098] In the following an approach will be described in which the rendered audio signal includes audio signals/ components that represent audio from other rooms than the one comprising the listening position. The description will focus on the generation of this audio component, but it will be appreciated that the rendered audio signal presented to the user may include many other components and audio sources. These may be generated and processed in accordance with any suitable algorithm or approach, and it will be appreciated that the skilled person will be aware of a large number of such approaches.

[0099] The renderer (507) is arranged to render the audio signal for a listening position being in an acoustic environment, in the following referred to as the first acoustic environment, based on the received audio data and metadata. The rendering is further such that it includes at least one audio component generated by rendering an audio source of another acoustic environment, i.e. the generated audio signal for the listening position in the first acoustic environment is generated to include a component from an audio source in a second acoustic environment (different from the first acoustic environment). Specifically, in situations/ embodiments where the different acoustic environments are different rooms, the rendering of the audio signal for a listening position includes rendering contributions from audio sources in other rooms.

[0100] In many cases the rendering of the audio and audio sources of other acoustic environments/ rooms than the first acoustic environment may be at least partly as diffuse or reverberation audio. In some cases, the rendering may be as reverberant diffuse audio which is the same for all positions in the first acoustic environment, i.e. the audio may be substantially independent of the exact listening position in the first acoustic environment. In such cases, rendering the audio for the listening position may be achieved simply by rendering the diffuse audio without this being specifically dependent on the listening position.

[0101] It will be appreciated that in many cases the audio data and metadata may be received as part of the same bitstream and the first and second receivers 501, 503 may be implemented by the same functionality and effectively the same receiver functionality may implement both the first and second receiver. The audio apparatus of FIG. 5 may specifically correspond to, or be part of, the client device 301 of FIG. 3 and may receive the audio data and metadata in a single bitstream transmitted from the server 303.

[0102] The metadata may describe acoustic elements and properties of the scene, and specifically for the different acoustic environments. For example, it may include data describing room dimensions, acoustic properties of the rooms (e.g. T60, DSR, material properties), the relationships between rooms etc. The metadata may further describe positions and orientations of some or all of the audio sources.

[0103] The metadata includes data that reflects how sound can propagate or spread between different acoustic environments, such as between different rooms. It may specifically include metadata related to transfer regions of the acoustically attenuating boundaries.

[0104] In particular, it may include data describing one or more transfer regions for at least one, and typically for more or even all, of the acoustically attenuating boundaries of the scene. A transfer region may specifically be a region for which an acoustic transmission level of sound from one acoustic environment to a neighbor acoustic environment (specifically from one room to a neighbor room) exceeds a threshold. Specifically, a transfer region may be a region (typically an area) of an acoustically attenuating boundary between two acoustic environments for which the attenuation by/ across the boundary is less than a given threshold whereas it may be higher outside the region. A transfer region is a region of an acoustically attenuating boundary having lower attenuation than an average attenuation of the acoustically attenuating boundary outside of transfer regions.

[0105] Thus, the transfer regions may define regions of the boundary between two acoustic environments/ rooms for which an acoustic propagation/ transmission/ transparency/ coupling exceeds a threshold. Parts of the boundary that are not included in a transfer region may have an acoustic propagation/ transmission/ transparency/ coupling below the threshold. Correspondingly, the transfer regions may define regions of the boundary between two acoustic environments/ rooms for which an acoustic attenuation is below a threshold. Parts of the boundary that are not included in a transfer region may have an acoustic attenuation above the threshold. The transfer regions may also be referred to as portals (in the acoustically attenuating boundaries).

[0106] A portal is associated with at least two acoustic environments, such as specifically two rooms. It may provide an acoustic link between the two acoustic environments/ rooms. Apart from indicating a link between acoustic environments, it may also include or reference acoustic properties of this link.

[0107] The following description will focus on an example where the acoustic environments are rooms, and the acoustically attenuating boundaries are walls of the rooms. However, it will be appreciated that this is merely exemplary and that acoustic environments may be other acoustic environments that are at least partially separated by acoustically attenuating boundaries.

[0108] The transfer region may thus indicate regions of a boundary for which the acoustic transparency is relatively high whereas it may be low outside the regions. A transfer region may for example correspond to an opening in the boundary. For example, for conventional rooms formed by acoustically attenuating boundaries in the form of walls, a transfer region may e.g. correspond to a doorway, an open window, or a hole etc. in a wall separating the two rooms.

[0109] A transfer region may be a three-dimensional or two-dimensional region. In many embodiments, boundaries between rooms are represented as two dimensional objects (e.g. walls considered to have no thickness) and a transfer region may in such a case be a two-dimensional shape or area of the boundary which has a low acoustic attenuation.

[0110] The acoustic transparency can be expressed on a scale. Full transparency means there is no acoustic suppression present (e.g. an open doorway). Partial transparency could introduce an attenuation to the energy what transiting from one room to the other (e.g. a thick curtain in a doorway, or a single pane window). On the other end of the scale are room separating materials that do not allow any (significant) acoustic leakage between rooms (e.g. a thick concrete wall).

[0111] The approach may thus (in the form of transfer regions) in some embodiments provide acoustic linking metadata that describes how two rooms are acoustically linked. This data may be derived locally, or may e.g. be obtained from a received bitstream. The data may be manually provided by a content author, or derived indirectly from a geometric description of the room (e.g. boxes, meshes, voxelized representation, etc.) including acoustic properties such as material properties indicating how much audio energy is transmitted through the material, or coupled into vibrations of the material causing an acoustic link from one room to another. The transfer region may in many cases be considered to indicate room leaks, where acoustic energy may be exchanged between two rooms.

[0112] FIG. 6 shows an example of a scene to which the described approach may be applied. FIG. 6 shows an example of a scene comprising a building with a number of rooms A-H. In the building, some audio sources are present in different rooms (indicated by circles 601). The audio apparatus of FIG. 5 may in this case determine a listener position 603 in room E and render audio for this listening position. The rendered audio signal includes different audio components in other rooms. The sound from such sources may specifically reach room E through a number of transfer regions 605, e.g. corresponding to (open) doors or windows in the walls forming the rooms.

[0113] Rendering of audio sources within the same room as the listening position is well established and many algorithms are known and may be used by the renderer without detracting from the invention. Rendering of audio from audio sources positioned in other rooms may for example be performed by representing the audio from the other rooms as an audio source that e.g. has no position (specifically for diffuse reverberation) or which e.g. has been assigned a position proximal to a portal. For example, a sound component from an audio source 605 may be considered to reach a given first room E comprising the listening position 603 via a first portal 4 of the first room E. The signal level reduction that results from the propagation to the first portal 4 may be determined and used to determine a level of the corresponding sound component at the first portal. The audio source may then be rendered as an audio signal component having a level corresponding to the determined level at the first portal E. As mentioned, in some embodiments, the sound source may be rendered as a spatially defined audio source, e.g. even as a point source positioned at the position of the first portal, or as a source with a spatial extent similar to, and proximal to, the portal. In other embodiments, the sound component may be considered a diffuse sound and may be rendered as diffuse reverberation in the first room E.

[0114] Such an approach may for example be used to render an audio signal component audio/ sound from room C as heard from the listening position in room E. It may for example also be used to render audio sources that are distanced by more than one room, such as e.g. an audio source from room A, if the resulting signal level after propagation through multiple portals is determined.

[0115] Rendering of audio sources as point sources, spatially extended sources, distributed sources, or diffuse sources with a given signal level are known in the art and will for brevity not be described in detail.

[0116] In some embodiments, the metadata may specifically include data that describes a position of at least one of transfer region of an acoustically attenuating boundary. The position may for example be described relative to, e.g., the room or as a relative position on the acoustically attenuating boundary in which the transfer region is formed (which, e.g., may be defined by a position within the room).

[0117] In many embodiments, the metadata may for example describe the scene topologically and/or geometrically including describing rooms, acoustically attenuating boundaries, and transfer regions in these. In some embodiments, a geometric description may be included which, e.g., describes sizes of all rooms (forming acoustic environments), extensions and positions of walls (forming the acoustically attenuating boundaries), and sizes, shapes, and positions of

portals (forming transfer regions).

[0118] However, in other embodiments, the metadata may additionally or alternatively include a topologic description of the scene. Such data may for example list a number of rooms and for each room provide some acoustic properties (Such as a BRIR or parameters describing reverberation). It may in addition define a number of portals/ transfer regions and for each transfer region may describe which two rooms the transfer region is connecting.

[0119] In many embodiments, the metadata comprises an energy transfer indication (which also may be referred to as a nominal energy transfer indication based on a nominal reference audio source) for at least a first transfer region formed in an acoustically attenuating boundary that separates two acoustic environments/ rooms. The nominal energy transfer indication is indicative of a proportion of energy of an omnidirectional point audio source at a reference position that would propagate to the first transfer region where the reference position is a relative position with respect to the first transfer region. The nominal energy transfer indication may thus indicate the amount of energy that would be radiated from an omnidirectional source at the reference position and which would be at the portal/ transfer region for which the nominal energy transfer indication is provided. The nominal energy transfer indication thus provides a description of an acoustic property of the transfer region based on a reference omnidirectional source. The acoustic property may specifically be an indication of the transfer of audio between the two acoustic environments.

[0120] In some embodiments, the nominal energy transfer indication for a first transfer region may be an indication of the proportion of energy that reaches the first transfer region, i.e. it may indicate the proportion of energy that is incident on the first transfer region from the reference audio source. In some embodiments, the nominal energy transfer indication may be an indication of the proportion of energy that exits the first transfer region, i.e. it may indicate the proportion of energy that is leaving the first transfer region into the first room from the reference audio source. It will be appreciated that such measures may be identical in case where the first transfer region does not include any attenuation or has any other acoustic effect, such as for example if the first transfer region is an empty opening in the first acoustically attenuating boundary. In other embodiments, the measures may for example differ due to an acoustic effect or attenuation of the first transfer region, such as, e.g., if the first transfer region is formed by a material that may have some acoustic effect yet allow some sound to propagate through. It will also be appreciated that such indications may be equivalent, i.e., an indication of energy incident on a transfer region may equivalently be considered an indication of an energy leaving the transfer region, and vice versa. Typically, one value/property can be determined directly from the other by considering the acoustic effect of the transfer region (such as, e.g., by compensating for an attenuation of sound by the transfer region).

[0121] The representation of acoustic information of the transfer regions using a nominal reference audio source as described may provide a particularly advantageous operation in many embodiments. It may typically allow for a low complexity and efficient (e.g. low data rate) description of acoustic properties resulting from the presence of transfer regions in acoustic environment dividing acoustic environments. It may further be provided in a way that allows easy processing to provide data suitable for rendering the specific audio sources that are present in the scene.

[0122] The approach is highly advantageous for rendering to include contributions from sources in one acoustic environment when rendering audio for another acoustic environment where the environments are divided by an acoustically attenuating boundary which includes a transfer region.

[0123] In the approach, the specific transfer region/ portal geometry may not be described by the metadata or used in the rendering but rather the transfer region/ portal may be described by the acoustic transfer properties as expressed by the reference to the reference audio source.

[0124] In many embodiments and for many transfer regions, the reference position may be within the second acoustic environment, but it will be appreciated that this is not necessary and indeed that the reference position could be outside of the second acoustic environment. For example, the reference audio source position for a portal between two rooms may be within the room, but could in some cases also be outside the room.

[0125] As a specific example, metadata for each portal/ transfer region may include the following data/ indications:

- portalFactor - an indication of normalized energy transfer from a reference source position to the portal. Thus the portalFactor may be an example of the nominal energy transfer indication.

[0126] Optionally it may also include one or more of the following:

- the portal position (for distance and angle impact)
- the portal orientation, (e.g. normal vector, for angle impact)
- an indication of portal dimensions (e.g. width and height, for determining angle impact)
- indications with which acoustic environments the portal is associated
- an indication of the acoustic environments separated by the acoustically attenuating boundary in which the portal is formed.

[0127] The nominal energy transfer indication may accordingly in some embodiments be represented by a data field/

value that may be referred to as a portalFactor. This portalFactor may indicate a proportion of an omnidirectional source that reaches a transfer region/ portal where the omnidirectional source is positioned at a reference position relative to the transfer region/ portal. The reference source is at a reference distance and reference angle with respect to the transfer region's position and orientation. Typically, the reference angle is advantageously chosen to be substantially perpendicular to the portal's orientation, but may also be at a different angle (e.g. in the range $\pm 10^\circ$, 20° , 30° , or 45° from a direction that is perpendicular to the portal (or to an acoustically attenuating boundary in which the portal is formed)).

[0128] As exemplified by FIG. 7 in two dimensions (whereas most embodiments will be in three dimensions), which shows a first transfer region 701 and a corresponding reference source 703, audio energy radiating omnidirectionally from a reference source position results in the energy spreading on a sphere and with only a portion of the energy being radiated transferring through the portal to the other acoustic environment. Thus, the proportion of energy from a reference source at a given position (in particular being perpendicular to a plane of extension for the transfer region) that reaches the transfer region can (for the distance variation over transfer region being negligible) be determined. The nominal energy transfer indication may specifically reflect a proportion of a sphere that is covered by the first transfer region where the sphere is centered on the reference position and has a radius corresponding to a distance from the reference position to the transfer region. In many embodiments, the distance from the reference position to the transfer region varies only negligibly over the transfer region. In cases where the distance variation is significant over the transfer region, a maximum distance may often advantageously be used, although in other embodiments e.g. a minimum or average distance may be used.

[0129] In the following an example of determining a nominal energy transfer indication/ portalFactor representing such a value will be described.

[0130] The opening from the portal covers a certain angular proportion. The portals may be assumed to be rectangular (or a rectangular equivalent can be derived based on surface area and aspect ratio), and may cover different angles in width and height. The angles can be derived from the reference distance and portal dimensions (width and height). From that, the proportion of that patch, compared to the sphere's surface can be derived, which is the portalFactor.

[0131] The width of the portal w gives the azimuth angle from the relation:

$$\sin(0.5 \cdot \theta_a) = \frac{0.5 \cdot w}{r}$$

yielding:

$$\theta_a = 2 \cdot \arcsin \left(\frac{0.5 \cdot w}{r} \right)$$

where r is the radius of the sphere. The radius is typically between the smallest distance from the reference source to the transfer region, and the largest distance from the reference source to the transfer region.

[0132] In many embodiments, the reference position is chosen to be perpendicular to the portal/ transfer region and in the middle of its (rectangular equivalent) width and height. With those conditions, the best radius is equal to the largest distance, corresponding with the distance to any of the four corner points on the (rectangular equivalent) of the transfer region.

[0133] This radius can thus be calculated from the (rectangular equivalent) width and height and the reference source's reference distance as:

$$r = \sqrt{d_{ref}^2 + 0.25 \cdot w^2 + 0.25 \cdot h^2}$$

[0134] Similarly, the height of the portal h gives the elevation angle

$$\theta_e = 2 \cdot \arcsin \left(\frac{0.5 \cdot h}{r} \right)$$

[0135] With those angles, the nominal energy transfer indication may be estimated with:

$$portalFactor = \frac{\theta_a}{2\pi} \cdot \frac{\theta_e}{\pi}$$

or, more accurately, be based on a surface integral:

$$portalFactor = \int_{e=-0.5\theta_e}^{0.5\theta_e} \int_{a=-0.5\theta_a}^{0.5\theta_a} r^2 \cdot \cos(e) \cdot da \cdot de$$

$$portalFactor = \int_{e=-0.5\theta_e}^{0.5\theta_e} [r^2 \cdot \cos(e) \cdot a]_{-0.5\theta_a}^{0.5\theta_a} \cdot de$$

$$portalFactor = \int_{e=-0.5\theta_e}^{0.5\theta_e} r^2 \cdot \cos(e) \cdot \theta_a \cdot de$$

$$portalFactor = r^2 \cdot \theta_a \cdot [\sin(e)]_{-0.5\theta_e}^{0.5\theta_e}$$

$$portalFactor = 2r^2 \cdot \theta_a \cdot \sin(0.5\theta_e)$$

[0136] Metadata providing such descriptions for transfer regions may be highly suitable for representing transfer regions and may form a highly efficient basis for determining sound propagation through transfer regions for audio sources of the scene, and specifically for audio sources in the same room as the transfer region and propagating through to the neighbor room. For example, in the scenario of FIG. 8, the nominal energy transfer indication may be provided for the first transfer region 1 based on reference position 801. The sound energy from a scene audio source 803 that reaches the first transfer region 1 and which propagates through this into room A may be determined based on this nominal energy transfer indication. The rendering of an audio signal for a listening position in room A is then determined to include a contribution from the audio source 803 in room B based on the propagated energy measure. In particular, the renderer may generate an audio component based on the audio for the scene audio source 803 and adapt the level of this in dependence on the determined propagated energy measure.

[0137] As will be described in more detail later, the renderer 203 may determine an energy reduction factor for a given transfer region/ portal being formed in an acoustic environment/ portal formed in an acoustically attenuating boundary/ wall separating a first acoustic environment/ room comprising a listening position for which an audio signal is generated and a second acoustic environment/ room comprising an audio source generating the audio. For clarity and brevity, the following description will focus on a scenario of a building where audio from different rooms is rendered in other rooms and the corresponding terminology will be used. However, it will be appreciated that the terms can be substituted for the alternative terms as indicated above.

[0138] The renderer 203 may, when rendering an audio signal component in a target room from an audio source in a source room via a (target) portal, proceed to determine an energy reduction factor for the target portal/ room and the rendering may be performed using the energy reduction factor. The renderer 507 specifically adapts the level of the rendered audio component to reflect the energy attenuation, and specifically the higher the energy attenuation given, the lower the level of the corresponding rendered audio signal component.

[0139] The energy reduction factor F_{tgt} may for example be applied to a source signal. For example as:

$$S_{in} = S_{src} \cdot \sqrt{F_{tgt}}$$

where S_{in} may be an input contribution of the source represented by signal S_{src} to a rendering algorithm (e.g. reverberation, coupled source rendering).

[0140] In many embodiments, the renderer may be rendering an immersive reverberation signal for the acoustic environment in which the listener is, denoted in-room reverberation. Typically, all energy emitted by sources inside the room is contributing to that reverberation. The nominal energy transfer indication may be used to determine how much energy reaches transfer regions of this room. These proportions of source energy may additionally, or alternatively, be used to reduce the source energy contributed to the in-room reverberation of that room.

[0141] In many cases the reduction is obtained by subtracting the proportions of source energy from the source energy contributing to the in-room reverberation. This may further be dependent on material properties associated with the transfer region. I.e., when the reflective properties of the transfer region are non-zero, the reduction of source energy may be limited. E.g. where F_{tgt} indicates the total energy of a source that is reaching the (only) transfer region of the room, the reduction of the in-room reverberation for that source may be determined as:

$$S_{rev} = S_{src} \cdot \sqrt{c_{sig2nrg} - F_{tgt} \cdot (1 - c_{refl})}$$

where S_{rev} is the input signal to the in-room reverb, S_{src} the source signal, $c_{sig2nrg}$ is a conversion coefficient indicating the ratio between emitted source energy and signal energy, and c_{refl} is a reflection coefficient associated with the transfer region. The coefficients and F_{tgt} may be frequency dependent.

[0142] Thus, in some embodiments, often when the listening position is in the second acoustic environment, the renderer may be arranged to render a diffuse audio signal component for the second acoustic environment (in which the audio source is present). The renderer 507 may in this case be arranged to adapt the level of the diffuse audio signal component dependent on the nominal energy transfer indication. The renderer may determine a energy estimate (which may be relatively) for an amount of energy reaching the transfer region from the audio source and reduce the level of the diffuse audio signal component by a corresponding amount.

[0143] The renderer 507 may be arranged to adapt a level of the audio component that is generated for a given portal based on an audio source in another room based on/ in dependence on the position of the audio source relative to the reference position for the nominal audio source. It may in many embodiments, be arranged to adjust the signal level of this audio component based on a difference in distance between the reference and actual audio source positions and the portal, based on the angular difference between directions between these sources and the portal, and/or based on a directivity (gain) for the audio source (in a direction towards the portal).

[0144] For example, if the nominal energy transfer indication represents a portalFactor that indicates the portion of source energy that is lost through the portal for an omnidirectional source at the reference source position, then the portion of source energy lost for a target source at a different position in the room, may be calculated as:

$$F_{tgt} = portalFactor \cdot G_{dist} \cdot G_{angle} \cdot G_{dir}$$

where G_{dist} compensates for distance difference, G_{angle} compensates for angle difference and G_{dir} compensates for directivity pattern. Some embodiments may use a subset of these compensations or may employ additional ones.

[0145] The renderer is in many embodiments arranged to adapt the level of the audio component for the audio source in the second room as a function of the difference between a reference distance which is from the reference position to a first transfer region and a source distance which is from the scene audio source to the first transfer region.

[0146] An approach may be described with reference to the scenario in FIG. 9 where a reference position 901 is positioned perpendicularly to a portal 903 whereas a scene audio source is at a source position 905 which differs from the reference position.

[0147] The effect of distance is related to the physical phenomenon that sources further away are sounding quieter. For this, typically, an $1/r$ law is used, meaning that the Root Mean Square (RMS) amplitude level (i.e. not energy) is inversely proportional to distance (r). The reasoning is that the source energy is represented by the surface of a sphere, and at a greater distance, with a larger sphere radius, the energy drops by approximately 6 dB because the surface of the sphere is four times as large.

[0148] Using this effect as a basis for adjusting for the distance effect gives:

$$G_{dist} = \left(\frac{d_{ref}}{d_{tgt}} \right)^2$$

[0149] Variations of the $1/r$ law may be used, or a decay curve may be used instead, where the decay curve may be

represented as an equation, function or look-up table indicating a distance attenuation gain for a given distance from the source. In such a case, the adjustment factor may be:

5

$$G_{dist} = \left(\frac{f(d_{tgt})}{f(d_{ref})} \right)^2$$

where $f()$ denotes the decay curve as an equation, function or look-up table.

10 **[0150]** When the portal size is relatively small and uniformly sized, or there is a need for a lower complexity approach, the distance from the scene audio source to the portal may be calculated once as the distance between the centre point of the portal and the source, in 3D space. The distance d between two points $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ may be calculated by:

15

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

20 **[0151]** In other embodiments the approach may include determining an average distance based on first calculating the distance across a number of uniformly distributed positions on the portal, such as the corners of the bounding box, or the nodes of the mesh describing it, and taking the mean of those calculated distances.

[0152] Other approaches, specifically when the size of the portal is relatively large relative to the source to portal distance (e.g. maximum portal dimension $\geq 0.9 * d_{tgt}$), may use the shortest distance between the target source and a point in the portal.

25 **[0153]** The renderer is in many embodiments arranged to adapt the level of the audio component for the audio source in the second room as a function of the difference between a direction from the reference position to the transfer region and a direction from the scene audio source position to the transfer region.

[0154] Similarly to distance, the angle between the scene audio source position and the portal impacts how much energy is lost through the portal. At a narrower angle, the effective surface of the portal, as seen from the source is smaller, and thus less energy will be lost.

30 **[0155]** Some embodiments may apply a simple linear relation:

35

$$G_{angle} = \frac{\min(|\varphi_{tgt} - \varphi_{ref}|, 0.5\pi)}{0.5\pi}$$

or, more typically, in 3D:

40

$$G_{angle} = \frac{\min(|\varphi_{tgt,a} - \varphi_{ref,a}|, 0.5\pi)}{0.5\pi} \cdot \frac{\min(|\varphi_{tgt,e} - \varphi_{ref,e}|, 0.5\pi)}{0.5\pi}$$

where the a and e subscripts indicate azimuth and elevation angles between the source position and the portal plane.

45 **[0156]** Other embodiments may consider that the energy reduction is stronger when the angle is further away from the nominal angle. For example, by:

$$G_{angle} = \cos(\min(|\varphi_{tgt} - \varphi_{ref}|, 0.5\pi))$$

50 or:

$$G_{angle} = \cos(\min(|\varphi_{tgt,a} - \varphi_{ref,a}|, 0.5\pi)) \cdot \cos(\min(|\varphi_{tgt,e} - \varphi_{ref,e}|, 0.5\pi))$$

55 or:

$$G_{angle} = \frac{\operatorname{atan}\left(\frac{d_{ref}L}{d_{ref}^2 - \frac{1}{4}L^2} \cdot \cos(\varphi_{tgt} - \varphi_{ref})\right)}{2 * \operatorname{atan}\left(\frac{L}{d_{ref}}\right)}$$

[0157] Calculating the angle of the scene audio source can be done in various ways, depending on which point on the transfer region is used. A simple approach may be to use the middle of the transfer region as a reference for the angle calculation. In many cases, the angle with the closest point on the transfer region may be particularly beneficial for sources close to the portal. Other embodiments may interpolate between these two angles dependent on the source's distance to the transfer region.

[0158] In more elaborate methods, the angle may be an averaged angle based on multiple points in the transfer region. For example, the four corners of (a rectangular equivalent of) the transfer region, or the nodes of a mesh describing the transfer region. This may be particularly beneficial for estimating realistic energy proportions for a wide range of source positions.

[0159] Determining an angle based on two coordinates means that the two coordinates form a line, and the angle of that line with respect to a reference orientation is calculated. The reference orientation may be defined as part of the coordinate system used for defining the scene. For example, the negative z-axis. Alternatively, the angle can be calculated with respect to the normal vector of the portal. Calculating the angle between two vectors is well known in the art and will not be described further.

[0160] When a nominal source is not at a 90° (0.5π radians) angle with the transfer region, a simple approach is to do the inverse compensation of the reference angle to 90° followed by a compensation from 90° to the target angle. For example:

$$G_{angle} = \frac{\cos(\min(|\varphi_{tgt} - 0.5\pi|, 0.5\pi))}{\cos(\min(|0.5\pi - \varphi_{ref}|, 0.5\pi))}$$

[0161] In many embodiments, the metadata may comprise data describing a directivity of the scene audio source and the renderer 507 may be arranged to adapt the level of the audio component generated for the scene audio source as a function of/ depending on the directivity. The directivity may typically indicate a variation in the gain/ signal level in different directions from the scene audio source.

[0162] The renderer 507 may specifically be arranged to scale the level of the audio component representing the scene audio source as a function of a relative directivity gain for the first audio source in a direction from the scene audio source to the transfer region where the relative directivity gain is indicative of a gain relative to an omnidirectional source.

[0163] A directivity pattern also influences the amount of energy that is leaking through the portal, and which may be frequency dependent.

[0164] The directivity may be given as a directivity pattern representing the amount of energy radiated in a range of azimuth and elevation directions relative to an omnidirectional pattern and nominal frontal direction. In a low complexity approach the effect of the directivity pattern can be taken as the mean energy level in the azimuth and elevation range that is covered by the portal.

$$G_{dir} = \frac{\sum_{a=a_{min}}^{a_{max}} \sum_{e=e_{min}}^{e_{max}} L_{a,e}}{q * n}$$

where a and e represent the azimuth and elevation angles covered by the ranges a_{min} to a_{max} and e_{min} to e_{max} respectively, which are defined by the relative position of the portal to the nominal frontal direction of the source (and directivity pattern), and q and n represent the number of azimuth and elevation angles considered. $L_{a,e}$ is the directivity gain associated with azimuth a and elevation e as specified in the directivity pattern. An exemplary scenario is shown in FIG. 10 which illustrates a scene audio source 1001 relative to a portal 1003.

[0165] In many embodiments, G_{dir} may be frequency dependent, and may be calculated per frequency band in which the directivity pattern is specified.

[0166] In some embodiments, the audio source that is rendered may specifically be an audio source that represents audio from a third acoustic environment, and specifically it may represent the audio that reaches the second acoustic environment via a portal between the third acoustic environment and the second acoustic environment. For example, for a scene audio source in the third acoustic environment, the described approach may be used to determine a level at a portal between the third acoustic environment and the second acoustic environment. The resulting audio signal, i.e. the audio signal from the scene audio source after level compensation, may thus represent the audio from the scene audio source that will propagate into the second acoustic environment via the second portal. This sound may further propagate into the first acoustic environment by the first portal. This effect may be emulated by positioning an audio source at the second portal with the signal corresponding to that entering the second acoustic environment from the first acoustic environment. This audio source may then be processed as described previously thereby allowing the audio entering the first acoustic environment to be determined and rendered.

[0167] The approach may in this way be used to represent sound/ audio propagation through multiple rooms.

[0168] Alternatively, or additionally, to the metadata described previously, the metadata received by the second receiver 503 may in some embodiments include transfer region data that describes transfer regions in the acoustically attenuating boundaries of the scene, and which may further include energy transfer parameters. Each energy transfer parameter is indicative of at least one energy attenuation between a pair of transfer regions, and specifically of an energy attenuation between two transfer regions of different acoustically attenuating boundaries. The energy attenuation for a pair of transfer regions is indicative of a proportion of audio energy at one transfer region of the pair of transfer regions that propagates to the other transfer region of the pair of transfer regions. Thus, each energy transfer parameter may comprise one energy attenuation indication for the pair of transfer regions (or, as will be described later, two energy attenuation indications).

[0169] Thus, whereas the nominal energy transfer indication may indicate the proportion of energy that reaches a transfer region from a given nominal omnidirectional audio source, the energy transfer parameter, and specifically the energy attenuation indication, reflects the proportion of audio energy at a second transfer region that transfers to a first transfer region. Similar to the nominal energy transfer indication, the energy attenuation indication may reflect the proportion of energy incident on the first transfer region and/or the proportion of energy radiating/ exiting the first transfer region (into the first acoustic environment). In general, the comments provided for the nominal energy transfer indications apply equally to the energy attenuation indications, mutatis mutandis.

[0170] The renderer 507 may render an audio source from the second acoustic environment in the first acoustic environment based on the energy attenuation indication for a pair of transfer regions that are part of acoustically attenuating boundaries of the first acoustic environment and of the second acoustic environment. Specifically, the renderer 507 may determine a level in the first acoustic environment of a signal component for an audio source in the second acoustic environment based on the energy attenuation indication. This signal component accordingly may represent audio that propagates from the second acoustic environment to the first acoustic environment through the first and the second transfer regions.

[0171] The renderer 507 may for example determine the signal energy for a given audio source that is incident on the second transfer region. For example, in some embodiments the level/ energy of reverberant audio in the second acoustic environment may be determined and converted into an energy/ signal level for reverberant audio that is considered to reach the second transfer region. As another example, the energy/ signal level at the second transfer region may be determined for a given specific, and e.g. point, audio source. In particular, the energy/ signal level at the second transfer region from an audio source in the second acoustic environment may be determined based on a nominal energy transfer indication for the audio source. Indeed, the energy/ signal level of the audio source that reaches the second transfer region can be determined using an approach based on a nominal energy transfer indication as previously described. The resulting energy/ signal level for the signal at the first transfer region can be determined by directly applying the energy attenuation indication for the transfer region pair, and the renderer 507 can adapt the signal level of the rendered signal component to reflect this attenuation. The previously described rendering approaches may for example be used as described but with an attenuation being introduced as determined by the energy attenuation indication.

[0172] As a specific example, in the example of FIG. 6, an audio source 607 in a second acoustic environment which in the specific example is room A may reach the listening room E via first the transfer region/ portal 1 between room A and C and then the transfer region/ portal 4 between room C and E. In this example, a nominal energy transfer indication may for example be provided for portal 1 and based on this, the energy at portal 1 from the audio source 607 may be determined as previously described. This may provide a first attenuation factor for the energy/ signal level from the audio source. The attenuation may then be increased based on an energy attenuation indication provided in the metadata for portals 1 and 4, or equivalently the energy/ signal level at transfer region 1 may be reduced by an amount given by the energy attenuation indication. The audio from the audio source in room A may then be rendered for the listening position in room E but with a reduced level that reflects the attenuation associated with the propagation through the two portals.

[0173] The acoustic environments of the two transfer regions/ portals of a given pair of transfer regions for which an energy transfer parameter is provided (and thus the acoustically attenuating boundaries of the acoustically attenuating

boundaries in which the transfer regions/ portals are formed) may have a shared acoustic environment, i.e. the two acoustic environments may be divided by a single acoustic environment, and thus the two acoustically attenuating boundaries in which the portals are formed may both be boundaries of a single shared acoustic environment. Specifically, as in the example of FIG. 6, the two portals 1 and 4 may be for acoustically attenuating boundaries that are of different rooms (namely room A and E), but which are also both boundaries of the same room, namely room C.

[0174] The energy transfer parameters and energy attenuation indications may be useful to describe sound propagation between different rooms via portals to an interconnected room. In some embodiments, the metadata comprises energy attenuation parameters only for pairs of transfer regions of boundaries sharing an acoustic environment, i.e. for which the portals/ transfer regions are formed in acoustically attenuating boundaries that are boundaries of the same acoustic environment. This may provide a reduced data rate for the metadata and may limit data representations to the most likely audio propagations between acoustic environments. Further, in some embodiments, if sound propagation is desired to be determined for acoustic environments that are further apart, such energy transfer parameters/ energy attenuation indications may be combined as described in more detail later.

[0175] A particular advantage of the approach is that it may be suitable for, and applied to, many different topologies and connections between different acoustic environments, including providing information on sound propagations between acoustic environments that do not have a shared acoustic environment. Indeed, in many embodiments, one or more of the energy transfer parameters/ energy attenuation indications are provided for transfer regions of acoustically attenuating boundaries that do not share any acoustic environment. For example, as illustrated in FIG. 11, an energy attenuation indication may be provided for two portals 1 and 3 that are separated by two acoustic environments/ rooms B and C, and thus for which there is no shared adjacent acoustic environment. This may allow facilitated rendering of audio in room A resulting from an audio source 1101 in room D as the properties of the full path of sound propagation through different acoustic environments may be combined and represented by a single energy attenuation indication.

[0176] Indeed, energy transfer parameters providing energy attenuation indications may be provided for any pair of transfer regions to indicate the sound propagation that may occur between these, and indeed in some embodiments an energy attenuation indication may be provided for each possible pair of transfer regions between any two rooms/ acoustic environments in the scene.

[0177] In many typical applications, the sound propagation may be symmetric and thus the energy attenuation indication for propagation from transfer region x to transfer region y is the same as the propagation from transfer region y to transfer region x. In such a case, the same energy attenuation indication may be used for rendering an audio signal in a first acoustic environment from an audio source in a second acoustic environment and for rendering an audio signal in the second acoustic environment from an audio source in the first acoustic environment.

[0178] Such symmetry is typically present in many physical or virtual scenes, and in particular for diffuse or reverberant audio that tends to not be associated with specific positions. The symmetry may be used to reduce the amount of data that is included in the metadata to describe transfer region to transfer region sound propagation. For example, the energy attenuation indications for all transfer region pairs may in such a case be represented by a symmetric matrix, such as

$$T = \begin{pmatrix} 1 & t_{12} & t_{13} & t_{14} & t_{15} & t_{16} & t_{17} & t_{18} \\ t_{12} & 1 & t_{23} & t_{24} & t_{25} & t_{26} & t_{27} & t_{28} \\ t_{13} & t_{23} & 1 & t_{34} & t_{35} & t_{36} & t_{37} & t_{38} \\ t_{14} & t_{24} & t_{34} & 1 & t_{45} & t_{46} & t_{47} & t_{48} \\ t_{15} & t_{25} & t_{35} & t_{45} & 1 & t_{56} & t_{57} & t_{58} \\ t_{16} & t_{26} & t_{36} & t_{46} & t_{56} & 1 & t_{67} & t_{68} \\ t_{17} & t_{27} & t_{37} & t_{47} & t_{57} & t_{67} & 1 & t_{78} \\ t_{18} & t_{28} & t_{38} & t_{48} & t_{58} & t_{68} & t_{78} & 1 \end{pmatrix}$$

where $t_{xy} = t_{yx}$ indicates the energy attenuation indication from transfer region x to transfer region y and from transfer region y to transfer region x.

[0179] The energy attenuation data may be efficiently represented as a matrix as above, but may for example also be represented by a direct indication as a set of portal pairs and the corresponding transfer region to transfer region energy attenuation indication, or in other suitable ways. A matrix such as the above may be sparsely populated or the set of portal pairs may not be a complete set of possible pairs. This is often beneficial for scenes with many acoustic environments. Entries with high energy attenuation values may for example be excluded. E.g. when $10 \cdot \log_{10}(\text{energy attenuation } [i, j]) < -60 \text{ dB}$.

[0180] Each energy attenuation indication is provided for two transfer regions/ portals and the metadata provides the energy attenuation indication and the identification of the transfer regions. The energy attenuation indication may also

be considered as an inverse energy transfer indication, i.e. the higher the energy attenuation, the lower the energy transfer. The energy attenuation indication between two transfer regions may typically indicate an increasing attenuation for an increasing distance between the transfer regions and depending on how many intermediate acoustic environments and transfer regions the sound must cross to reach the destination transfer region. Further if the two transfer regions

are not aligned (around corners or occluded by obstacles), the corresponding energy attenuation indication may indicate a higher attenuation to reflect the higher loss of the sound attenuation.

[0181] Further, the energy attenuation indication may in some embodiments indicate time varying values or e.g. values that are dependent on dynamically changing properties of the scene. For example, if portals like doors are opened, closed, or moved, the energy attenuation indication may change.

[0182] The approach may include the consideration that a portal may be assumed to radiate sound uniformly across its surface into a receiving room. When the receiving room has other portals, a portion of the sound from the first portal will reach such a second portal and may leak into the next receiving room. The amount of sound that is transferred may be linked to the relative positions and sizes of the other portals with respect to the first portal and the total room surface area.

[0183] This information may be used to efficiently determine how much energy of sources in one room contributes to other rooms, and this information may be captured by the energy attenuation indications. For example, each row in the matrix above may indicate for a portal of an associated room how much it contributes to all the other rooms.

[0184] In many embodiments, the transfer region positions (as well as the acoustically attenuating boundaries) may be assumed to be fixed and to not move, and accordingly the energy attenuation indications can be precalculated for their specific positions. A simple method is to calculate the visible area of the receiving portal relative to the center point of the source portal, and compare that area to the area of a hemisphere with radius equal to the distance between portals. It is assumed that the portal is a subsection of a larger plane, therefore it may often be assumed to radiate hemispherically rather than omnidirectionally as for the nominal energy transfer indication.

[0185] In a more complex method, rather than calculating the visible area relative to only the center point of the source, the area of the source portal may be taken into account. This may be by calculating the visible area across a number of locations bounded by the source portal and taken the average visible area, or by other means.

[0186] In many embodiments, the energy attenuation indication may be calculated at an encoder side or with an offline process where computational complexity is more amply available (e.g. it may be calculated at the VR server 303). In such cases, acoustic models of various complexity levels may be used to determine how much energy from the first transfer region reaches the second transfer region. This may include occlusion and/or diffraction modelling.

[0187] Some embodiments may focus on calculating the energy transfers/ attenuations from all transfer regions of a room to all other transfer regions of the same room. These transfers may then be combined to represent higher order room to room transfers (i.e. including more than one shared/ intermediate room). For example, when room A is associated with transfer regions 4 and 5, and room B is associated with transfer regions 5 and 2, the transfer from transfer region 4 to 2 can be obtained by combining the transfer from transfer region 4 to 5 calculated for room A with the transfer from transfer region 5 to 2 calculated for room B. Some embodiments may further include a transfer- or material property of transfer region 5.

[0188] The energy attenuation indications may be directly used to determine an energy reduction factor for sound in one acoustic environment reaching another acoustic environment, and the rendering may be performed using the energy reduction factor.

[0189] Specifically, for a given audio source in a source room, the energy incident on a transfer region may be determined. This may e.g. be done using the previously described approach or by other means. For example, the data may come from an audio source defined in a bitstream as a low complexity replacement for several sources in a source room, may be calculated using another method, or may be resulting from reverberation rendering in a source room.

[0190] The resulting energy reduction factor F_{tgt} may by the renderer 507 be applied to a signal, e.g. as:

$$S_{in} = S_{src} \cdot \sqrt{F_{tgt}}$$

where S_{in} may be an input contribution of the source represented by signal S_{src} to a rendering algorithm (e.g. reverberation, coupled source rendering).

[0191] A particular advantage of the approach is that it does not require detailed geometric information of the scene, and in particular of rooms, acoustically attenuating boundaries, transfer regions etc., or indeed of specific acoustic properties of the scene. Indeed, information on the exact connections between the rooms or the acoustic properties of these are not necessary. Rather, the energy transfer parameters can be considered topological properties that simply connect two transfer regions and provide information of sound propagation between these. This may allow a much facilitated operation and rendering with much reduced complexity and resource usage being possible.

[0192] In many embodiments, the energy attenuation indication for a pair of portals may indicate the proportion of

audio energy incident on the one transfer region that will propagate to be incident on the other transfer region. This may be advantageous in allowing the energy attenuation indication to be symmetric thereby allowing one indication to be used in both directions, and thus the amount of metadata may be reduced. It may also allow for the rendering to be adapted based on specific acoustic properties of the transfer region. For example, if the transfer region is dynamically covered by a fabric (e.g. a curtain) this can be reflected by introducing an additional attenuation factor that can be left out when the transfer region is not covered.

[0193] In other embodiments, the energy attenuation indication may indicate the energy attenuation for the output of the receiving transfer region, i.e. it may represent the energy exiting/ radiating from a given transfer region for a given energy being incident on another transfer region. This may allow reduced complexity rendering in many situations.

[0194] In many embodiments, the renderer 507 may be arranged to generate an audio source by combining two, more, or all audio sources in an acoustic environment into a single audio source. Such an audio source may for example be generated by determining relative sound levels at a given audio source position and generating the audio as a weighted summation of the audio signals from the individual audio sources with the weights reflecting the relative sound levels at the source position. The source position may specifically be generated to correspond to the position of a transfer region.

[0195] For example, the previously described approach of determining a sound level at the transfer region based on a nominal energy transfer indication and the actual position of the individual audio source may be performed for all audio sources in the acoustic environment. The audio signals may then be weighted accordingly and summed to result in all audio of the acoustic environment being represented by a single audio source positioned at the transfer region.

[0196] The sound propagation to the listening acoustic environment may then be determined based on the energy transfer parameters as previously described and the renderer 507 may render the resulting signal, e.g. as reverberation and diffuse sound.

[0197] Thus, in some embodiments, the sources of each acoustic environment can be combined into a single source at the related transfer region (e.g. one for each transfer region associated with the environment). The renderer 507 may then for each transfer region of the acoustic environment determine the sound in the listening room based on applying the energy attenuation indications of the energy transfer parameters and subsequently proceed to render all of these audio signal components. This may provide a lower complexity of rendering audio in one acoustic environment originating in another acoustic environment and considering sound propagation through multiple, and possibly all acoustic paths between the acoustic environments.

[0198] In some embodiments, energy transfer parameters may be provided for all transfer region pairs for which some sound transfer/ propagation is possible and rendering of a signal component representing inter-room propagation through transfer regions, the renderer 507 may simply extract and use the appropriate energy attenuation indication for that transfer region.

[0199] However, in some embodiments, energy transfer parameters may only be provided for a subset of transfer regions, such as e.g. only for transfer regions that share a common acoustic environment. This may allow a reduced data rate and/or may substantially alleviate the requirement for determining accurate energy attenuation indications. For example, if these are based on measurements in a real building, the number of measurement operations that are required can be reduced substantially.

[0200] In such embodiments, energy attenuation indications for other transfer region pairs may e.g. in some cases be determined by combining energy attenuation indications for other transfer region pairs. Thus, in some embodiments, the renderer 507 is arranged to generate a combined energy transfer attenuation by combining the energy transfer attenuation for a first pair of transfer regions and for a second pair of transfer regions where the two pairs include a transfer region that is common. For example, as illustrated in the example of FIG. 12, the first pair of transfer regions may be for a transfer region between a first and second transfer region thereby providing an indication of the energy transfer/ attenuation between a first and second environment. The second pair of transfer regions may be between a third transfer region and the second transfer region thereby providing an indication of the energy transfer/ attenuation between the third transfer region and the second transfer region, and thus an indication of the energy transfer/ attenuation between a second transfer region and the third acoustic environment. The energy attenuation indications of the two transfer region pairs may be combined, e.g. simply by combining the attenuations (e.g. by multiplying the two energy attenuations in the linear domain or adding them in the logarithmic domain for attenuation values). The resulting combined value thus indicate the energy attenuation from the third transfer region to the first transfer region and thus indicates the sound propagation from the third acoustic environment to the first acoustic environment. The combined energy attenuation may accordingly be used for rendering audio for a listening position in the first acoustic environment from an audio source in the third acoustic environment in the same way as if a direct energy attenuation indication was provided for the pair of the first transfer region and the third transfer region. Some embodiments may further include a transfer- or material property of the second transfer region.

[0201] In some embodiments, the energy transfer parameter for a given pair of transfer regions may comprise a plurality of energy attenuation indications with different energy attenuation indications being provided for the different

acoustic environments that are separated by an acoustically attenuating boundary in which one of the transfer regions of the pair of transfer regions is provided.

[0202] For example, the energy transfer parameter for a first transfer region of a pair of transfer regions may comprise an energy attenuation indication for both of the acoustic environments that are separated by a given transfer region/ acoustically attenuating boundary. Thus, rather than merely providing an energy attenuation indication for the transfer region pairs, different energy attenuation indications may be provided for sound that reaches the source transfer region from one acoustic environment and for sound that reaches the source transfer region from the other acoustic environment. For example, for a portal in a wall dividing two rooms, a separate energy attenuation indication may be provided for each of the rooms. The renderer 507 may then render sound from the two acoustic environments differently.

[0203] This may provide improved performance in many scenarios and may in particular reflect that sound to different acoustic environments/ rooms from other acoustic environment/ rooms may depend on the direction of the incident sound. Indeed, in many embodiments, sound from a given room to another given room may only be possible/ suitable for sound passing through a given portal in one direction but not in the other. In many embodiments, one of the directional energy attenuation indications for a given transfer region pair may be zero.

[0204] Such an approach of directional, separate energy attenuation indications may be particularly suitable for scenarios in which energy attenuations for multiple transfer region pairs are combined to provide a path from a source acoustic environment to a destination listening acoustic environment.

[0205] Indeed, portal to portal transfer (transfer region to transfer region) is often dependent on the direction of sound incidence onto the portal (transfer region). For example, for a building comprising a number of rooms, a rendering algorithm may be arranged to proceed to determine the room that each audio source is in, and then for each source determine all the portals in that room. It may then determine the audio source energy at (specifically incident on) each portal, and then continue to apply the energy attenuations for each of the portals in the source room to each of the portals in the listening room.

[0206] However, for some scenarios, some such approaches may result in undesired behavior and this may be addressed by having the energy transfer parameters indicate directional energy attenuations which are dependent on the direction of incidence of sound on the (source) transfer region.

[0207] For example, the room layout of FIG. 13 may be considered. With the listener in room A and source s_1 in room C, it can be seen that the transfer p_{21} (from portal 2 to portal 1) is relevant, but the transfer p_{31} is not relevant for the listening position in room A. For source s_2 in room D, the transfer p_{31} is relevant. The topology of the rooms contributes to determining which transfers are important.

[0208] The relevance can be pre-determined and represented in the received metadata by the metadata reflecting different energy attenuation for different acoustic environments of at least one transfer region of at least one pair of transfer regions. The different acoustic environments of one transfer region are the acoustic environments separated by the acoustically attenuating boundary in which the transfer region is present.

[0209] As a portal/ transfer region is a region in an acoustically attenuating boundary connecting/ separating two rooms, a source will only be on one of two sides of the portal. Therefore, the metadata can provide two different energy attenuation values, where the first corresponds to the first room connected to the portal and the second value corresponds to the second room connected to the portal.

[0210] This could be represented by two values per portal pair, or two matrices (or a 3D matrix where one dimension has size 2). The relation with the rooms could be pre-determined, for example when portals are defined with IDs to two environments, the first energy attenuation value could correspond with the first environment and the second value with the second environment. It will be appreciated that any way of the metadata indicating different/ directional energy attenuations may be used.

[0211] In the example of FIG. 13, the value for p_{31} could specifically indicate infinite attenuation (zero energy transfer) for room C and typically a non-infinite value (indicating some energy transfer) for room D.

[0212] In the example of FIG. 14, there may always be an acoustic path from any acoustic environment to any other acoustic environment, but this can still mean that some pairs of portals represent irrelevant/ invalid paths. For example, p_{65} is relevant when the source s_3 is in room M, but not when the source would be in room L. This is the case because portals 1 and 2 are both related to source room L, but also because there is a path between the portals outside room L that passes through the listening room.

[0213] Therefore, in some embodiments there may be four values provided, depending on which side of the first transfer region the source is and on which side of the second transfer region the listener is.

[0214] A second layout example, shown in FIG. 15, shows an example where for source s_4 only $p_{10,7}$ is relevant, and for source s_5 it is not, but p_{87} and $p_{11,7}$ are. The energy attenuation values for the non-relevant transfer can be indicated to be infinite.

[0215] In the fourth example shown in FIG. 16 (which is a variation of the third example), it can be seen that $p_{10,7}$ is relevant for both s_4 as well as s_5 , but that the values will likely be different for these.

[0216] Instead of relying on additional and explicit metadata providing the different directional energy attenuation

values, some embodiments may determine relevant transfers based on metadata that is already available and used for other purposes. Specifically, based on the information of which portals connect which environments, a connectivity graph can be made. This graph indicates how the different environments are connected through portals and can be used to determine relevance.

[0217] The graphs for the four examples from FIG. 13 through FIG. 16 are shown in FIG. 17. Each node represents a room and each vertex a portal. Known graph techniques can be used to determine whether there is a connection from a particular room through a particular first portal, where each vertex may be crossed only once.

[0218] With that it can be seen e.g. that from T to P there is no path through portal 10 for the third example, but such a path does exist in the fourth example.

[0219] Such graphs may also be used when energy transfer parameters are provided given for first order transfers (i.e. only through 1 room), by using a path finding algorithm that collects the relevant transfer factors on the one or more paths they find.

[0220] In many embodiments, the audio signal component determined as described above may be rendered as a non-direct audio component, i.e. it may be rendered to not merely be rendered as an audio source that propagates by other means than (just) a direct line of sight propagation.

[0221] Specifically, the rendering may be as a reverberation audio component in the first acoustic environment. Specifically, the audio signal may be level compensated and the resulting signal rendered using a suitable rendering approach for generating reverberation audio. It will be appreciated that a large number of algorithms for rendering audio signals as reverberant audio/ sound are known and may be used.

[0222] Thus, in many embodiments the approach may be used to generate reverberant audio in a room/ acoustic environment that results from audio sources in other rooms. This may provide a particularly advantageous approach in many scenarios and may reflect a more natural experience in many situations.

[0223] In many embodiments, the renderer 507 may be arranged to render the signal component to reflect all the sound energy that reaches the corresponding transfer region. In particular, the rendering may be such that it is considered that the transfer region has no other impact on the rendered audio, and indeed that apart from the extension of the transfer region in the acoustically attenuating boundary, the transfer region has no other acoustic properties or characteristics that need to be considered. Indeed, the transfer region/ portal may simply be considered to correspond to an opening in the acoustically attenuating boundary/ wall, and may be considered to have no acoustic impact.

[0224] In such cases, the energy reaching a transfer region may be considered equal to the energy that exits the transfer region. The determined energy attenuation (from another transfer region or from a specific sound source) may be considered to be the same for the incident energy and for the radiated energy entering the first acoustic environment. For example, the nominal energy transfer indication or the energy attenuation may inherently indicate both of these (as they may be the same).

[0225] However, in other embodiments, the transfer region itself may be considered to have an acoustic property that affects the amount of sound energy that passes through the transfer region. For example, in some embodiments, the transfer region may not be a complete opening but may have some attenuation which however is less than the surrounding acoustically attenuating boundary. For example, a wall may include a door which is covered by a drape that provides some acoustic attenuation. The renderer 507 may be arranged to take such attenuation into account, and specifically may reduce the signal level accordingly. In some cases, the acoustic effects of a transfer region may vary, and the rendering may dynamically be adapted to reflect this.

[0226] Portals may represent features such as windows or doors, and as such whether they are open or closed may be changed during runtime. When a user or other element of the rendering system interacts with the portal, such as partially closing a door, an additional weighing function may be applied to the calculated total gain, such that 1 is with the portal fully open, and 0 or a factor related to a material property with the portal fully closed. For example, a transmission coefficient of the scene element covering the portal may be used.

[0227] A similar approach may be used when a portal or other surface has a non-zero coupling coefficient. Energy reaching a closed portal may not be fully blocked by the portal, but a proportion of the energy may couple with the surface and be re-radiated. As an extreme example a single layer glass window will vibrate when a loud noise is made on the opposite side, reproducing some portion of that noise, even though there is no direct path for the sound to travel. Thus, in some embodiments, sound propagation through a transfer region/ portal may fully or partially be via an acoustic coupling effect. In some such embodiments, the renderer 507 may be arranged to render the corresponding sound from a source in the neighbor room by rendering a sound source at the position of the portal and having an energy level that is dependent on, and e.g. proportional to, the signal energy reaching the transfer region compensated by the attenuation occurring as a result of the coupling propagation effect.

[0228] Often acoustic properties of the transfer region may be in the form of material-related properties such as reflectiveness, absorptiveness, transmissiveness or related effects. Reflectiveness can indicate a proportion of incident sound that is reflected in a specular and/or diffuse way. Absorption can relate to dissipation in the material or translation into material vibrations (which may be re-emitted as a coupled source). Transmission typically indicates how much

energy is passed through.

[0229] Thus, in some embodiments, the metadata (e.g. the nominal energy transfer indication and the energy transfer parameter) for a transfer region may be indicative of the energy that reaches the transfer region, and thus the incident energy on the transfer region. This energy may then be reduced/ modified based on the acoustic properties of the transfer region when determining a suitable signal level for the resulting audio signal component. This may for example provide improved flexibility and e.g. allow dynamic variations in the transfer region to easily be accommodated.

[0230] However, in other embodiments, the metadata (e.g. the nominal energy transfer indication and/or the energy transfer parameter) for a transfer region may be indicative of the energy that exits the transfer region. Thus, in some embodiments the metadata (e.g. the nominal energy transfer indication and/or the energy transfer parameter) may reflect/ include a contribution from an acoustic property of the transfer region itself. Thus, in some embodiments, different acoustic properties for the transfer region need not be explicitly be considered or taken into account when rendering, but rather may implicitly be specified by the received metadata and no specific adaptation of the rendering itself may be necessary.

[0231] It will be appreciated that the use of (nominal) energy transfer indications as described previously may be in combination with, or separate to, the use of energy transfer parameters as described above. Similarly, it will be appreciated that the use of energy transfer parameters as described previously may be in combination with, or separate to, the use of (nominal) energy transfer indications as described above. The principles, approaches, functions, uses etc. described above for respectively nominal energy transfer indications and energy transfer parameters thus (as appropriate) apply to the individual uses and do not imply or require that such functions related to nominal energy transfer indications must be combined with such functions related to energy transfer parameters. The different metadata and applications are independent and separate. However, it will also be appreciated that particularly advantageous and synergistic operation may be achieved for embodiments using both functions related to nominal energy transfer indications and energy transfer parameters.

[0232] The apparatus(s) may specifically be implemented in one or more suitably programmed processors. In particular, the artificial neural networks may be implemented in one more such suitably programmed processors. The different functional blocks may be implemented in separate processors and/or may e.g. be implemented in the same processor. An example of a suitable processor is provided in the following.

[0233] FIG. 18 is a block diagram illustrating an example processor 1800 according to embodiments of the disclosure. Processor 1800 may be used to implement one or more processors implementing an apparatus as previously described or elements thereof (including in particular one more artificial neural network). Processor 1800 may be any suitable processor type including, but not limited to, a microprocessor, a microcontroller, a Digital Signal Processor (DSP), a Field Programmable Array (FPGA) where the FPGA has been programmed to form a processor, a Graphical Processing Unit (GPU), an Application Specific Integrated Circuit (ASIC) where the ASIC has been designed to form a processor, or a combination thereof.

[0234] The processor 1800 may include one or more cores 1802. The core 1802 may include one or more Arithmetic Logic Units (ALU) 1804. In some embodiments, the core 1802 may include a Floating Point Logic Unit (FPLU) 1806 and/or a Digital Signal Processing Unit (DSPU) 1808 in addition to or instead of the ALU 1804.

[0235] The processor 1800 may include one or more registers 1812 communicatively coupled to the core 1802. The registers 1812 may be implemented using dedicated logic gate circuits (e.g., flip-flops) and/or any memory technology. In some embodiments the registers 1812 may be implemented using static memory. The register may provide data, instructions and addresses to the core 1802.

[0236] In some embodiments, processor 1800 may include one or more levels of cache memory 1810 communicatively coupled to the core 1802. The cache memory 1810 may provide computer-readable instructions to the core 1802 for execution. The cache memory 1810 may provide data for processing by the core 1802. In some embodiments, the computer-readable instructions may have been provided to the cache memory 1810 by a local memory, for example, local memory attached to the external bus 1816. The cache memory 1810 may be implemented with any suitable cache memory type, for example, Metal-Oxide Semiconductor (MOS) memory such as Static Random Access Memory (SRAM), Dynamic Random Access Memory (DRAM), and/or any other suitable memory technology.

[0237] The processor 1800 may include a controller 1814, which may control input to the processor 1800 from other processors and/or components included in a system and/or outputs from the processor 1800 to other processors and/or components included in the system. Controller 1814 may control the data paths in the ALU 1804, FPLU 1806 and/or DSPU 1808. Controller 1814 may be implemented as one or more state machines, data paths and/or dedicated control logic. The gates of controller 1814 may be implemented as standalone gates, FPGA, ASIC or any other suitable technology.

[0238] The registers 1812 and the cache 1810 may communicate with controller 1814 and core 1802 via internal connections 1820A, 1820B, 1820C and 1820D. Internal connections may be implemented as a bus, multiplexer, crossbar switch, and/or any other suitable connection technology.

[0239] Inputs and outputs for the processor 1800 may be provided via a bus 1816, which may include one or more

conductive lines. The bus 1816 may be communicatively coupled to one or more components of processor 1800, for example the controller 1814, cache 1810, and/or register 1812. The bus 1816 may be coupled to one or more components of the system.

[0240] The bus 1816 may be coupled to one or more external memories. The external memories may include Read Only Memory (ROM) 1832. ROM 1832 may be a masked ROM, Electronically Programmable Read Only Memory (EPROM) or any other suitable technology. The external memory may include Random Access Memory (RAM) 1833. RAM 1833 may be a static RAM, battery backed up static RAM, Dynamic RAM (DRAM) or any other suitable technology. The external memory may include Electrically Erasable Programmable Read Only Memory (EEPROM) 1835. The external memory may include Flash memory 1834. The External memory may include a magnetic storage device such as disc 1836. In some embodiments, the external memories may be included in a system.

[0241] The terms audio and sound may be considered equivalent and interchangeable and may both refer to respectively physical sound pressure and/or electrical signal representations of such as appropriate in the context.

[0242] It will be appreciated that the above description for clarity has described embodiments of the invention with reference to different functional circuits, units and processors. However, it will be apparent that any suitable distribution of functionality between different functional circuits, units or processors may be used without detracting from the invention. For example, functionality illustrated to be performed by separate processors or controllers may be performed by the same processor or controllers. Hence, references to specific functional units or circuits are only to be seen as references to suitable means for providing the described functionality rather than indicative of a strict logical or physical structure or organization.

[0243] The invention can be implemented in any suitable form including hardware, software, firmware or any combination of these. The invention may optionally be implemented at least partly as computer software running on one or more data processors and/or digital signal processors. The elements and components of an embodiment of the invention may be physically, functionally and logically implemented in any suitable way. Indeed, the functionality may be implemented in a single unit, in a plurality of units or as part of other functional units. As such, the invention may be implemented in a single unit or may be physically and functionally distributed between different units, circuits and processors.

[0244] Although the present invention has been described in connection with some embodiments, it is not intended to be limited to the specific form set forth herein. Rather, the scope of the present invention is limited only by the accompanying claims. Additionally, although a feature may appear to be described in connection with particular embodiments, one skilled in the art would recognize that various features of the described embodiments may be combined in accordance with the invention. In the claims, the term comprising does not exclude the presence of other elements or steps.

[0245] Furthermore, although individually listed, a plurality of means, elements, circuits or method steps may be implemented by e.g. a single circuit, unit or processor. Additionally, although individual features may be included in different claims, these may possibly be advantageously combined, and the inclusion in different claims does not imply that a combination of features is not feasible and/or advantageous. Also, the inclusion of a feature in one category of claims does not imply a limitation to this category but rather indicates that the feature is equally applicable to other claim categories as appropriate. Furthermore, the order of features in the claims do not imply any specific order in which the features must be worked and in particular the order of individual steps in a method claim does not imply that the steps must be performed in this order. Rather, the steps may be performed in any suitable order. In addition, singular references do not exclude a plurality. Thus, references to "a", "an", "first", "second" etc. do not preclude a plurality. Reference signs in the claims are provided merely as a clarifying example shall not be construed as limiting the scope of the claims in any way.

Claims

1. An audio apparatus:

a first receiver (501) arranged to receive audio data for audio sources of a scene comprising multiple acoustic environments, the acoustic environments being divided by acoustically attenuating boundaries;
a second receiver (501) arranged to receive metadata for the audio data, the metadata comprising:

a position indication for at least a first transfer region of a first acoustically attenuating boundary between a first acoustic environment and a second acoustic environment, the first transfer region being a region of the first acoustically attenuating boundary having lower attenuation than an average attenuation of the first acoustically attenuating boundary outside of transfer regions; and

an energy transfer indication for the first transfer region, the energy transfer indication being indicative of a proportion of energy at the first transfer region from an omnidirectional point audio source at a reference

position, the reference position being a relative position with respect to the first transfer region;

a renderer (507) arranged to render an audio signal for a listening position in the first acoustic environment, the rendering including generating a first audio component by rendering audio data for a first audio source of the second acoustic environment and adapting a level of the first audio component in dependence on the energy transfer indication.

2. The audio apparatus of claim 1 wherein the renderer (507) is arranged to adapt the level of the first audio component in response to a position of the first audio source relative to the reference position.

3. The audio apparatus of claim 1 or 2 wherein the renderer (507) is arranged to adapt the level of the first audio component in dependence on a difference between a reference distance from the reference position to the first transfer region and a distance from a position of the first audio source to the first transfer region.

4. The audio apparatus of any previous claim wherein the renderer (507) is arranged to adapt the level of the first audio component in dependence on a difference between a direction from the reference position to the first transfer region and a direction from the first audio source to the first transfer region.

5. The audio apparatus any previous claim wherein the metadata further comprises data describing a directivity of sound radiation from the first audio source and the renderer (507) is arranged to adapt the level of the first audio component in dependence on the directivity.

6. The audio apparatus of any previous claim wherein the renderer (507) is arranged to scale the level of the first audio component in dependence on a relative directivity gain for the first audio source in a direction from the first audio source to the first transfer region, the relative directivity gain being indicative of a gain relative to an omnidirectional source.

7. The audio apparatus of any previous claim wherein the first audio source represents audio reaching the second acoustic environment from a third acoustic environment via a second transfer region of a second boundary separating the third acoustic environment from the second acoustic environment.

8. The audio apparatus of any previous claim wherein the metadata further comprises:

energy transfer parameters, each energy transfer parameter indicating an energy attenuation between a pair of transfer regions, the energy attenuation for a pair of transfer regions being indicative of a proportion of audio energy at one transfer region of the pair of transfer regions propagating to the other transfer region of the pair of transfer regions;

and the renderer (507) is arranged to render a second audio component of the audio signal by rendering a second audio source of a third acoustic environment in dependence on an energy attenuation for a pair of transfer regions comprising a transfer region of an acoustically attenuating boundary of the first acoustic environment and a transfer region of a second acoustically attenuating boundary being a boundary of the second acoustic environment.

9. The audio apparatus of any previous claim wherein the metadata includes a coupling coefficient for the first transfer region, and the renderer (507) is arranged to render the first audio component as originating from an audio source at a position proximal to the first transfer region and in dependence on the coupling factor.

10. The audio apparatus of any previous claim wherein the renderer (507) is arranged to render the first audio component as a reverberation audio component of the first acoustic environment.

11. The audio apparatus of any previous claim wherein the renderer (507) is arranged to generate a reverberation audio signal for the second environment to include a first component from the first audio source, the renderer (507) further being arranged to determine an energy loss estimate as a proportion of energy of the first audio source that reaches the first transfer region, the proportion of energy being determined in dependence on the energy transfer indication and a position of the first audio source relative to the reference position; and the renderer (507) being arranged to reduce a level of the reverberation audio signal by an amount depending on the energy loss estimate.

12. The audio apparatus of any previous claim wherein the energy transfer indication reflects a proportion of a sphere

that is covered by the first transfer region, the sphere being centered on the reference position and having a radius corresponding to a distance from the reference position to the first transfer region.

5 13. The audio apparatus of any previous claim wherein the renderer (507) is arranged to render the first audio component as a non-direct audio component.

14. A method of rendering an audio signal, the method comprising:

10 receiving audio data for audio sources of a scene comprising multiple acoustic environments, the acoustic environments being divided by acoustically attenuating boundaries;
receiving metadata for the audio data, the metadata comprising:

15 a position indication for at least a first transfer region of a first acoustically attenuating boundary between a first acoustic environment and a second acoustic environment, the first transfer region being a region of the first acoustically attenuating boundary having lower attenuation than an average attenuation of the first acoustically attenuating boundary outside of transfer regions; and
an energy transfer indication for the first transfer region, the energy transfer indication being indicative of a proportion of energy at the first transfer region from an omnidirectional point audio source at a reference position, the reference position being a relative position with respect to the first transfer region;

20 rendering an audio signal for a listening position in the first acoustic environment, the rendering including generating a first audio component by rendering audio data for a first audio source of the second acoustic environment and adapting a level of the first audio component in dependence on the energy transfer indication.

25 15. An audio data signal comprising:

audio data for audio sources of a scene comprising multiple acoustic environments, the acoustic environments being divided by acoustically attenuating boundaries;
30 metadata for the audio data, the metadata comprising:

a position indication for at least a first transfer region of a first acoustically attenuating boundary between a first acoustic environment and a second acoustic environment, the first transfer region being a region of the first acoustically attenuating boundary having lower attenuation than an average attenuation of the first acoustically attenuating boundary outside of transfer regions; and
35 an energy transfer indication for the first transfer region, the energy transfer indication being indicative of a proportion of energy at the first transfer region from an omnidirectional point audio source at a reference position, the reference position being a relative position with respect to the first transfer region.

40 16. An audio apparatus arranged to generate an audio data signal in accordance with claim 15.

45 17. A computer program product comprising computer program code means adapted to perform all the steps of claim 15 when said program is run on a computer.

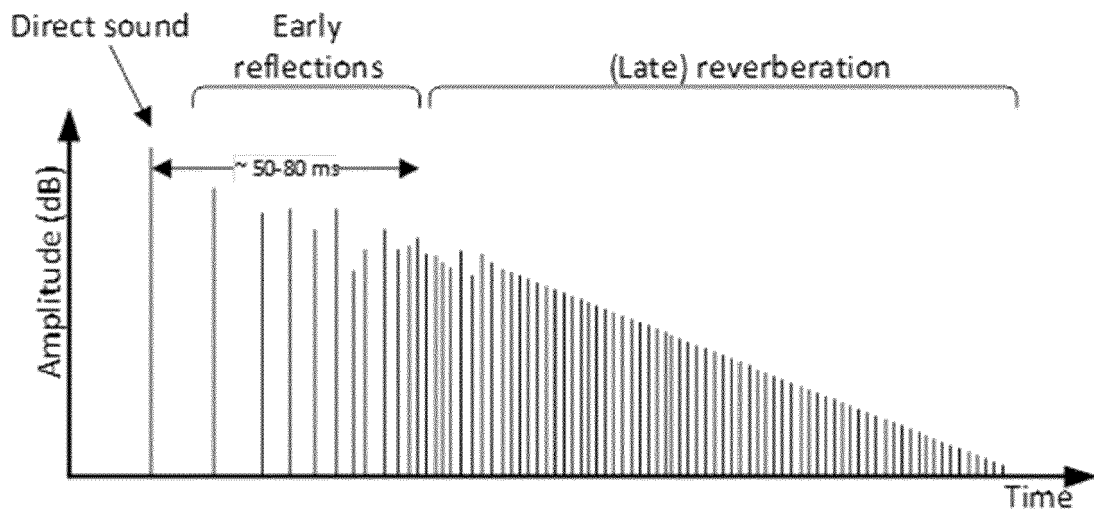


FIG. 1

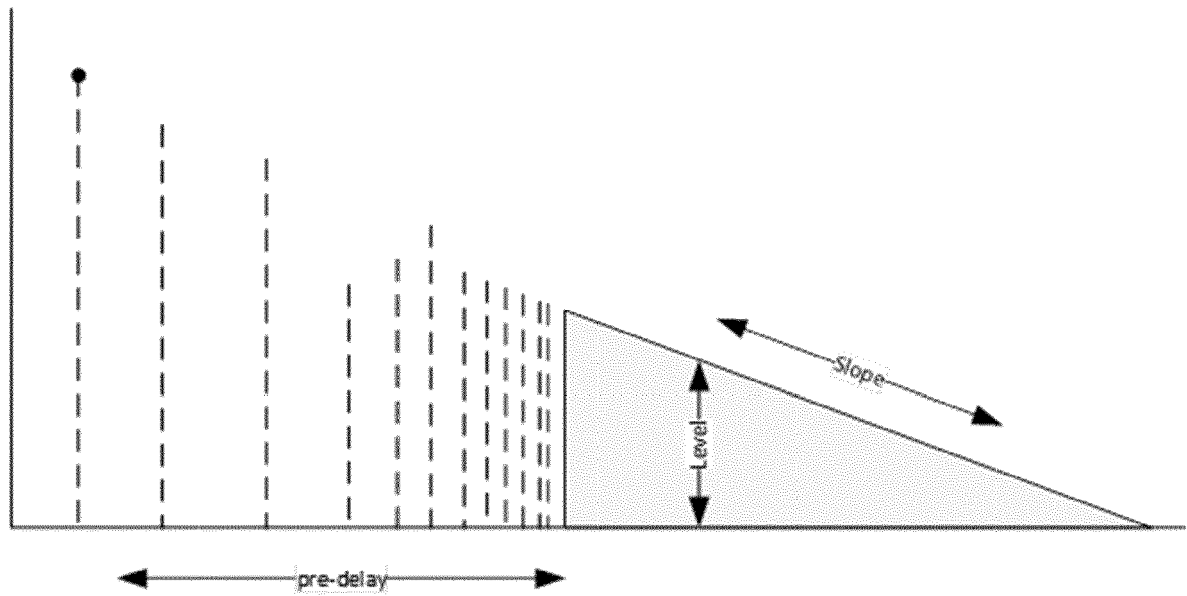


FIG. 2

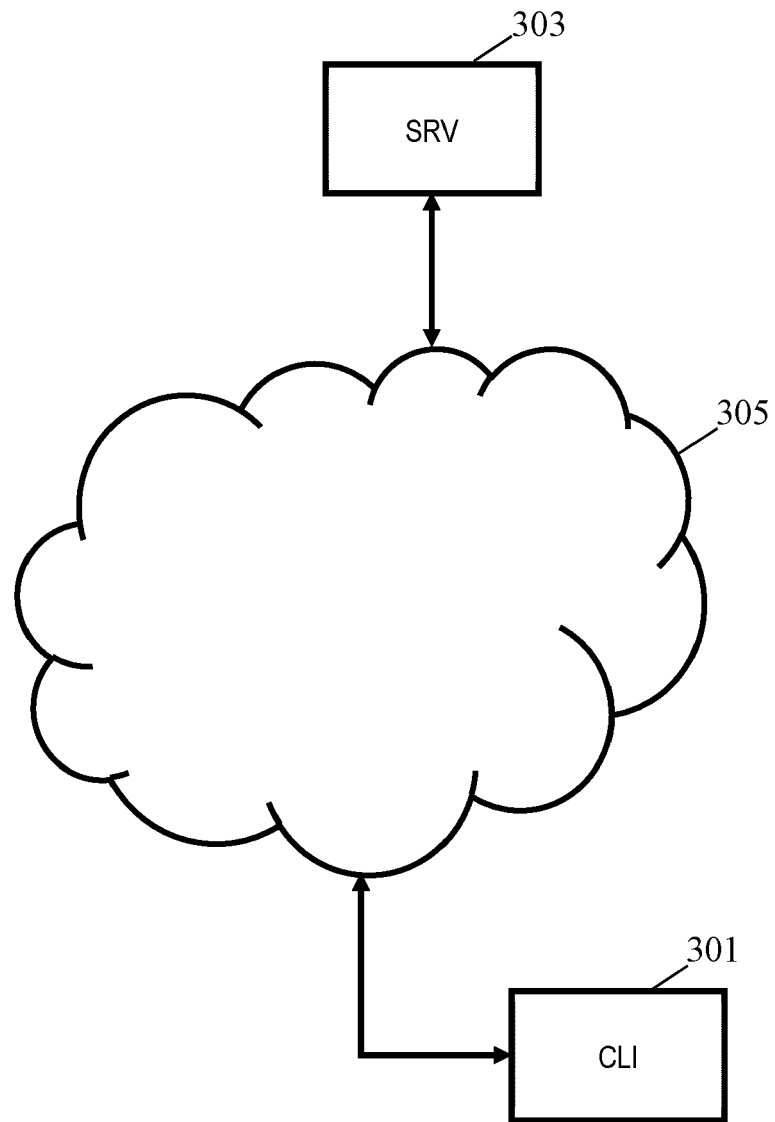


FIG. 3

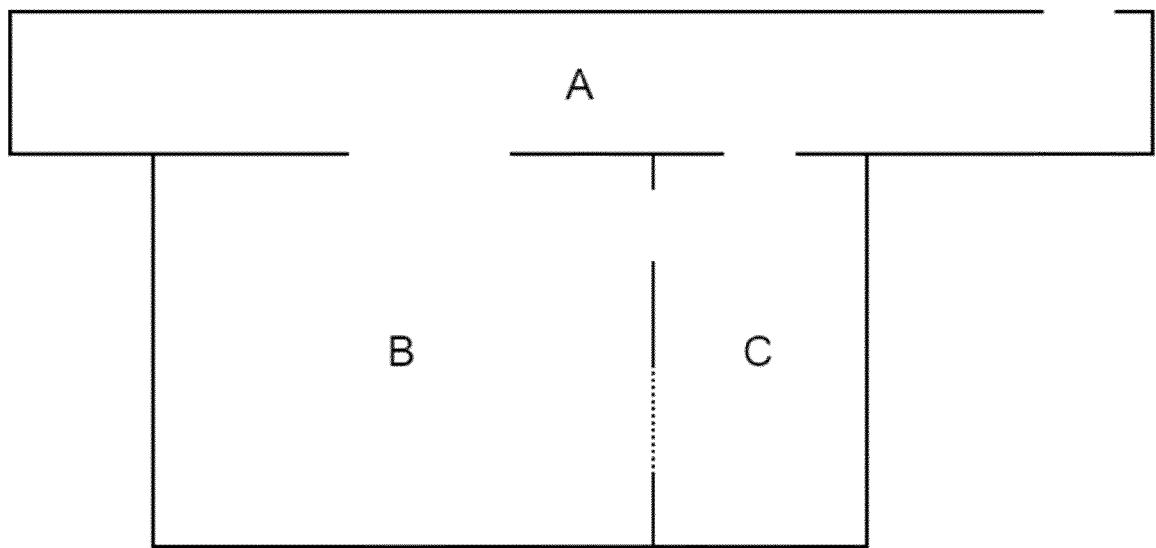


FIG. 4

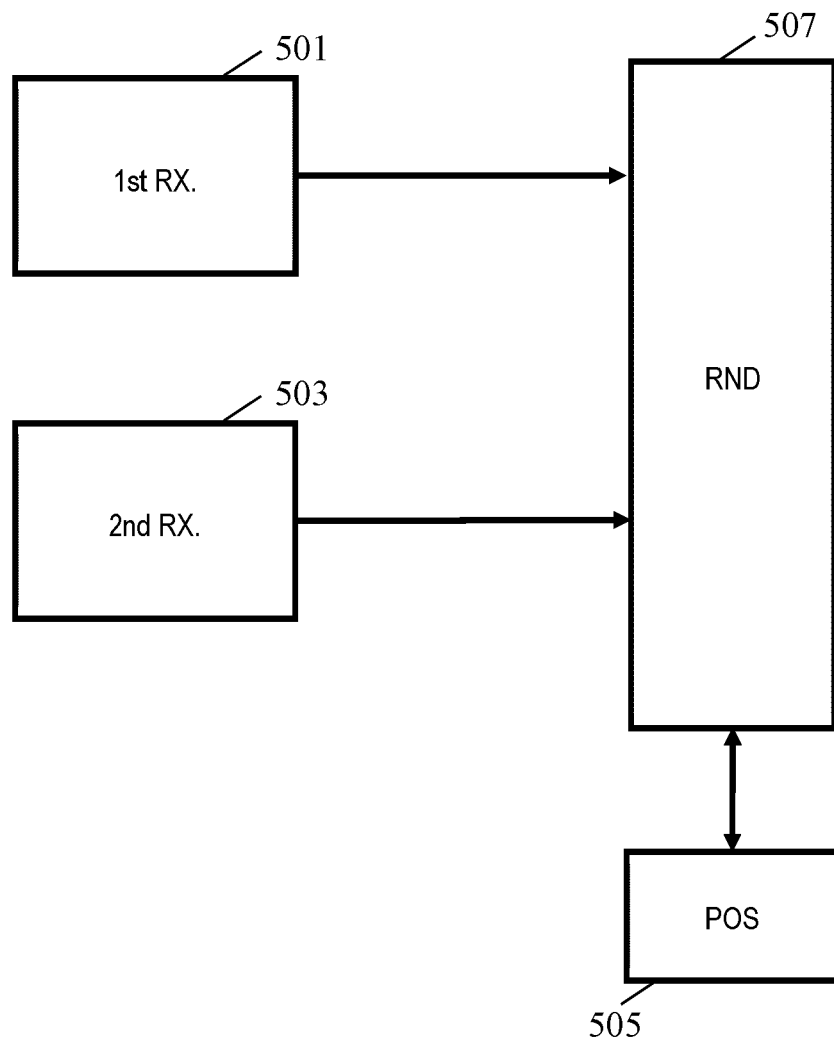


FIG. 5

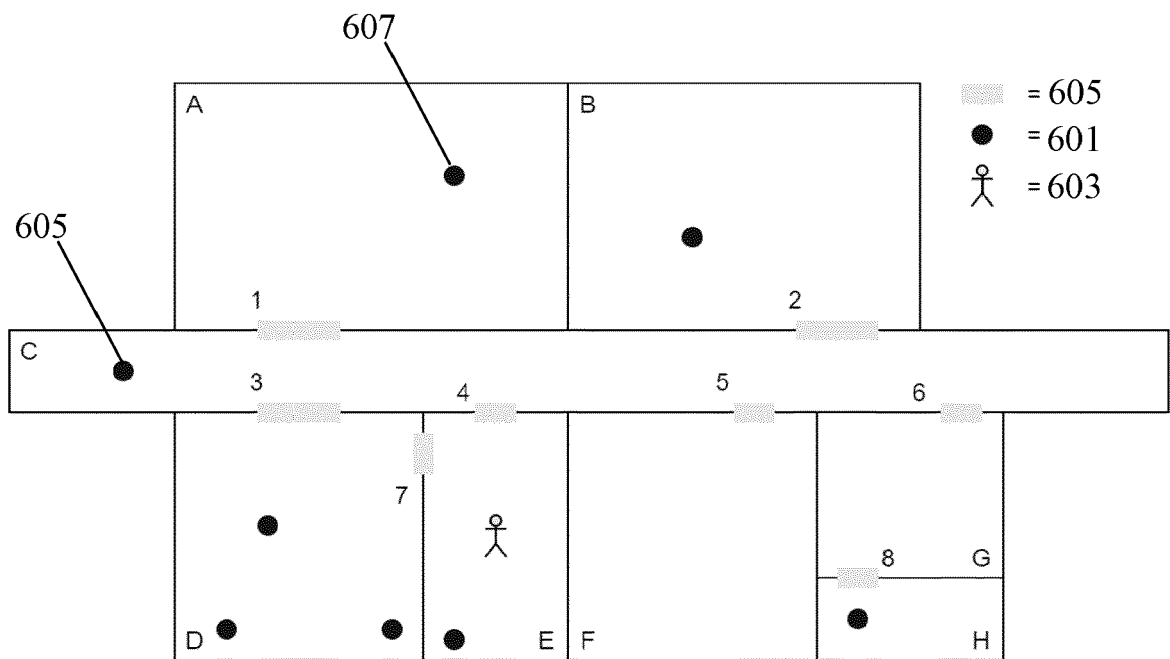


FIG. 6

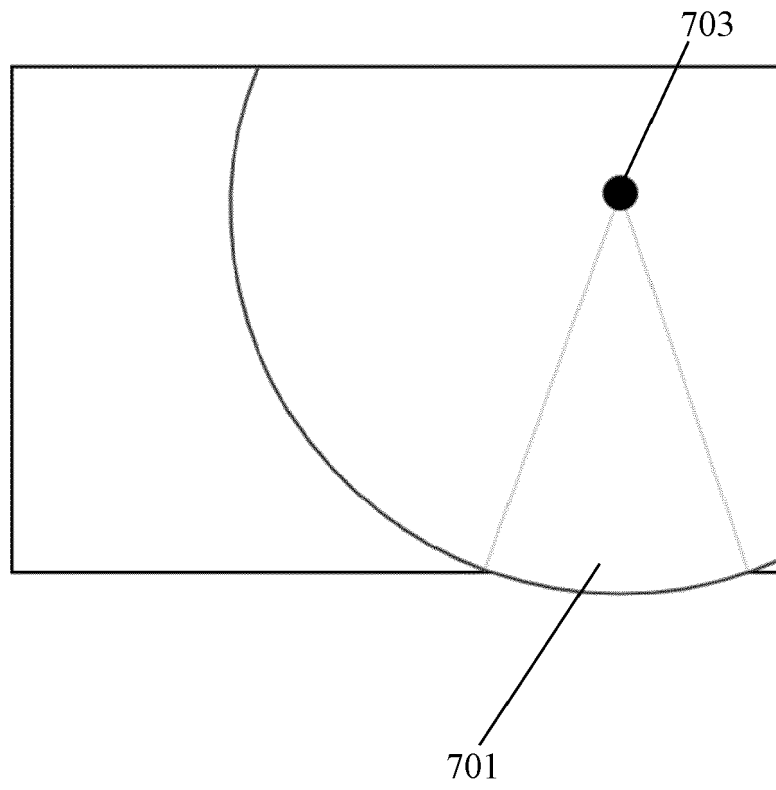


FIG. 7

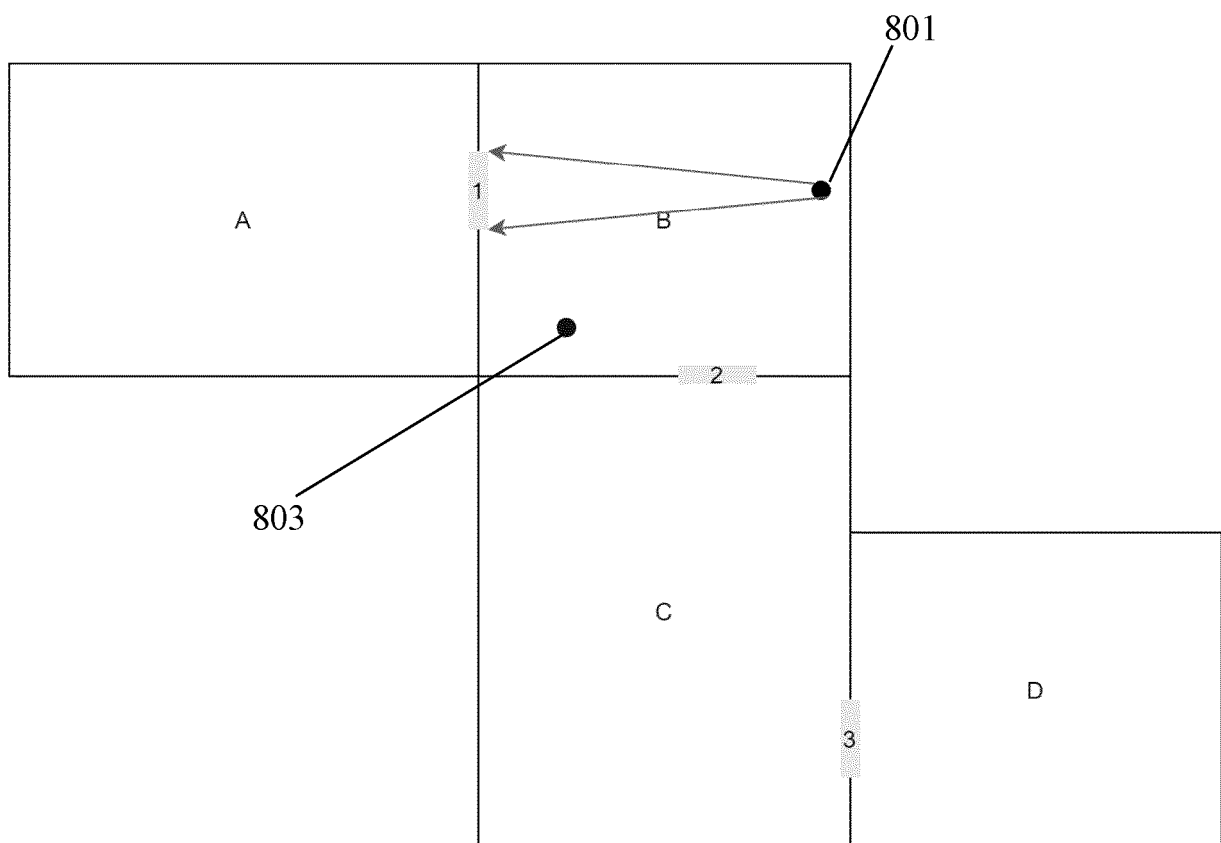


FIG. 8

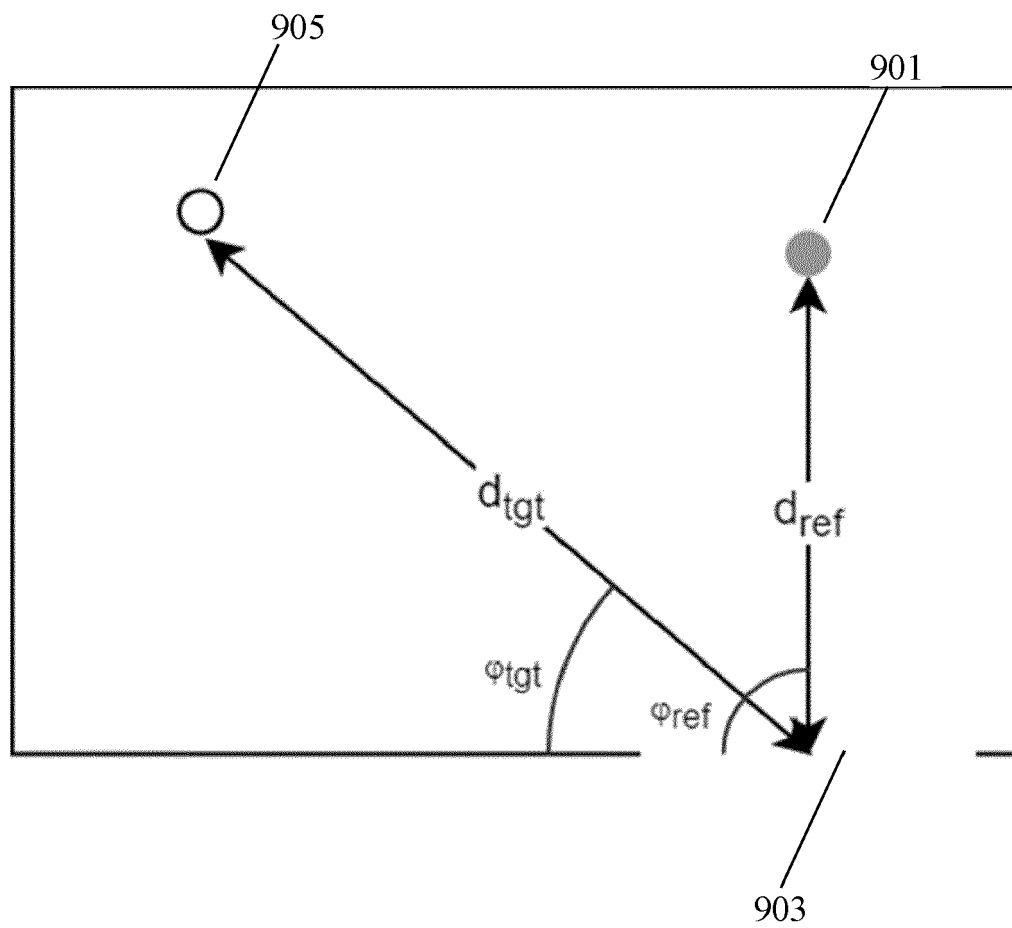


FIG. 9

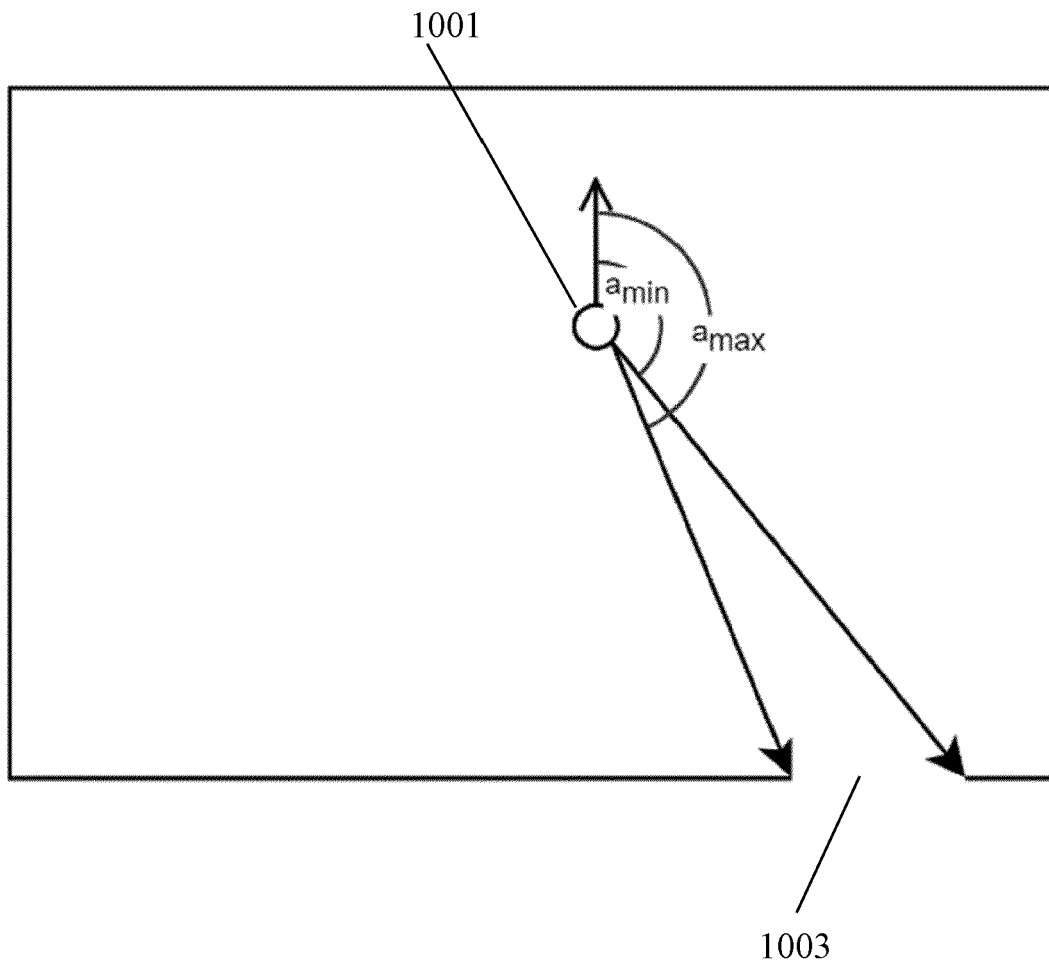


FIG. 10

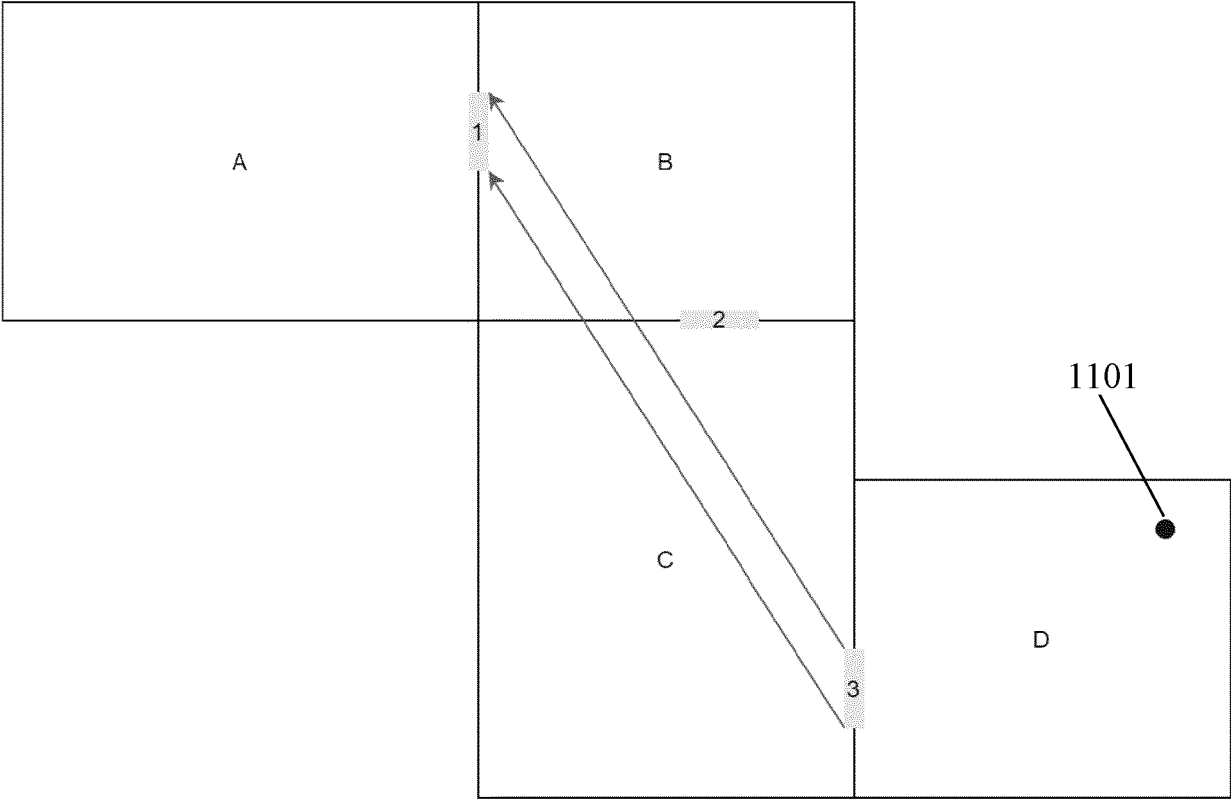


FIG. 11

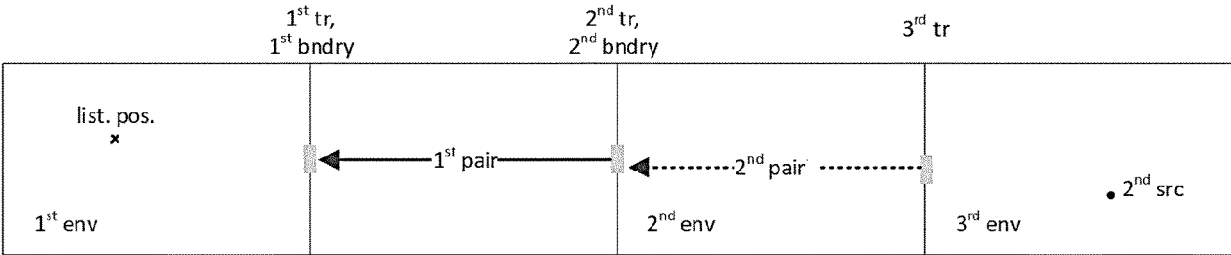


FIG. 12

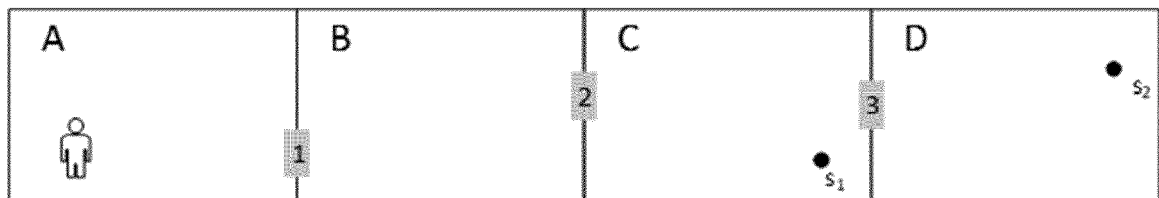


FIG. 13

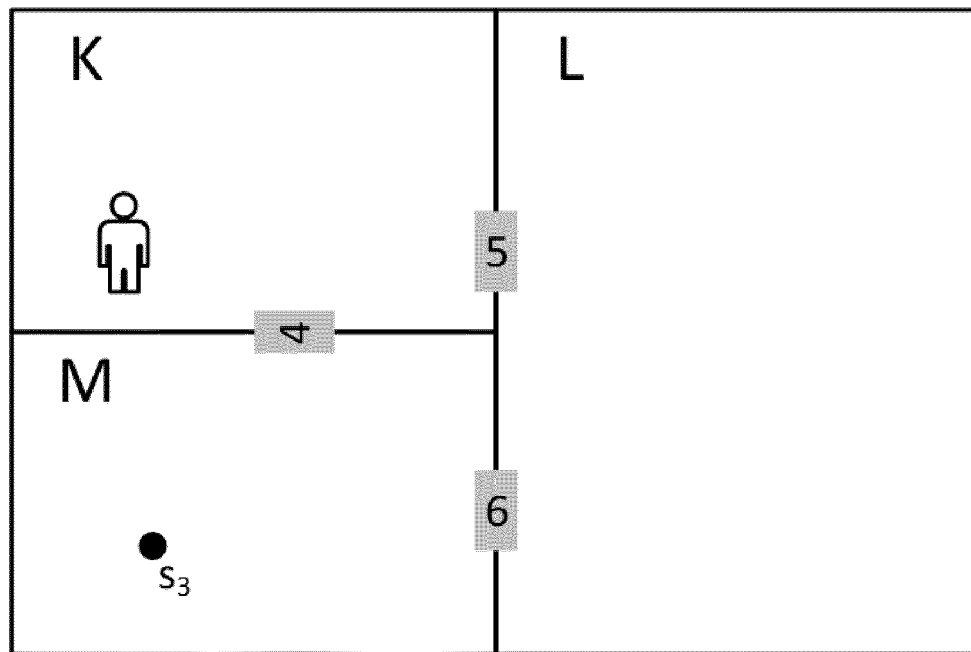


FIG. 14

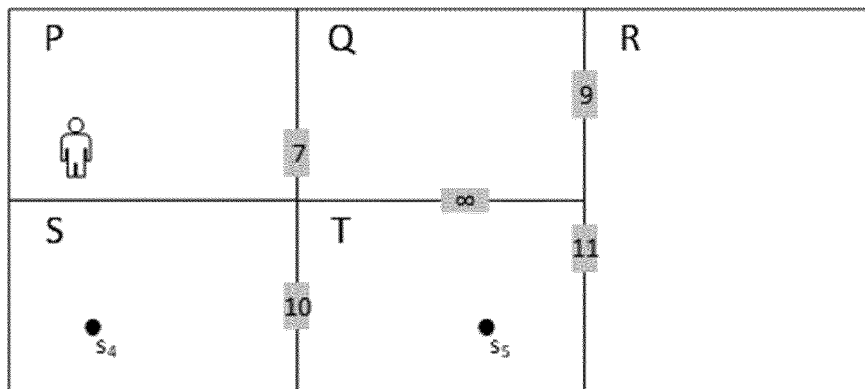


FIG. 15

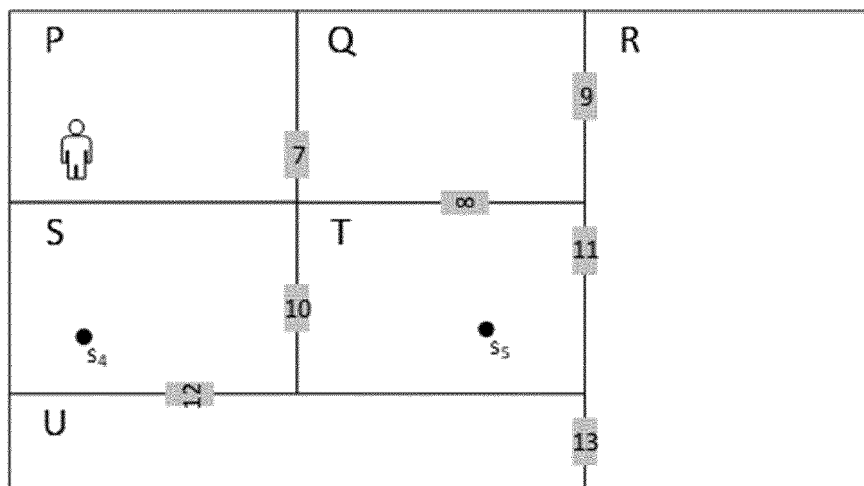
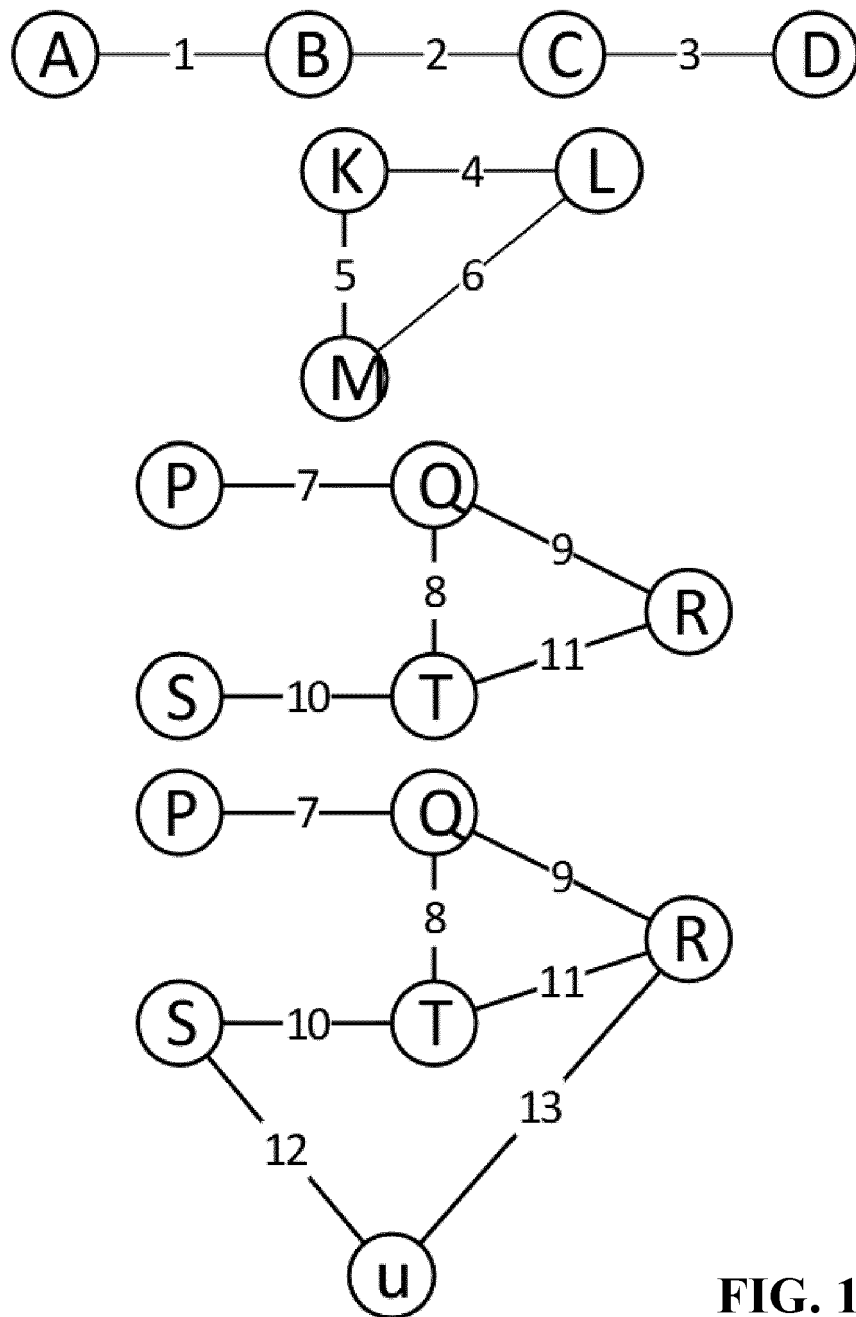


FIG. 16

**FIG. 17**

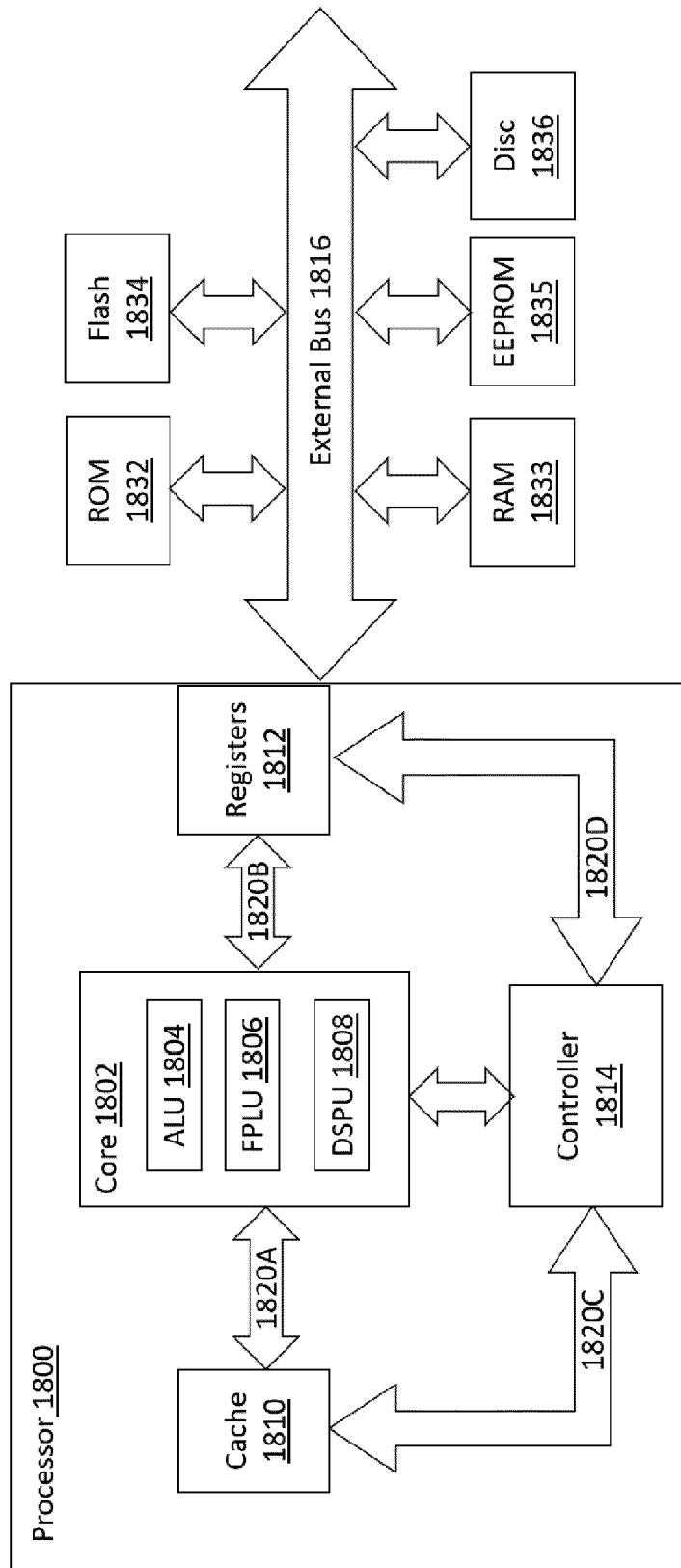


FIG. 18



EUROPEAN SEARCH REPORT

Application Number

EP 22 21 1764

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	SCHRÖDER DIRK ET AL: "Hybrid method for room acoustic simulation in real-time", , vol. 7, 2 September 2007 (2007-09-02), pages 4521-4526, XP093046083, ISBN: 978-1-61567-707-8 Retrieved from the Internet: URL:https://www.researchgate.net/publication/258098589_Hybrid_method_for_room_acoustic_simulation_in_real-time> * page 2 - page 5 *	1-3,7,9, 10,12-17	INV. H04S7/00
Y		8	
A		11	
X	Dirk Schröder: "PHYSICALLY BASED REAL-TIME AURALIZATION OF INTERACTIVE VIRTUAL ENVIRONMENTS", , 4 February 2011 (2011-02-04), XP055593422, Berlin Retrieved from the Internet: URL:http://publications.rwth-aachen.de/record/50580/files/3875.pdf [retrieved on 2019-06-03] * Section 5.5.2; page 47 *	1,4-7,9, 10,12-17	TECHNICAL FIELDS SEARCHED (IPC) H04S
Y		8	
A		11	
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 16 May 2023	Examiner Joder, Cyril
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	



EUROPEAN SEARCH REPORT

Application Number

EP 22 21 1764

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	STAVRAKIS EFSTATHIOS ET AL: "Topological Sound Propagation with Reverberation Graphs", ACUSTICA UNITED WITH ACTA ACUSTICA, vol. 94, no. 6, 1 November 2008 (2008-11-01), pages 921-932, XP093046042, DE ISSN: 1610-1928, DOI: 10.3813/AAA.918109 Retrieved from the Internet: URL:https://www-sop.inria.fr/revs/project s/revGraphs/final_revgraphs.pdf> * Sections 4 and 6 *	1,2, 7-10, 12-17	
Y	CAMERON FOALE ET AL: "Portal-based sound propagation for first-person computer games", INTERACTIVE ENTERTAINMENT, RMIT UNIVERSITY, GPO BOX 2476V MELBOURNE VIC 3001 AUSTRALIA, 3 December 2007 (2007-12-03), pages 1-8, XP058325482, ISBN: 978-1-921166-87-7 * Section 9 *	8	TECHNICAL FIELDS SEARCHED (IPC)
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 16 May 2023	Examiner Joder, Cyril
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	

1
EPO FORM 1503 03.82 (P04C01)