(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 4 398 607 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
10.07.2024 Bulletin 2024/28

(21) Application number: **23150704.7**

(22) Date of filing: **09.01.2023**

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

(52) Cooperative Patent Classification (CPC):
H04S 7/302; H04S 7/30; H04S 7/305;
H04S 2400/15; H04S 2420/03

(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

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

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

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

(54) AN AUDIO APPARATUS AND METHOD OF OPERATION 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 (501) receives metadata comprising sets of reverberation parameters indicative of a level of reverberation resulting from audio sources in an acoustic environment and energy transfer parameter indicative of energy attenuation between acoustic environments. For a listening acoustic environment, the selector (513) selects a first set of set of reverberation parameters in response to parameter values of the first set of set of reverberation parameters and an energy transfer parameter indicative of the attenuation from the acoustic environment of the first set of set of reverberation parameters and the listening acoustic environment. A reverberator of a plurality of reverberators (509, 511) generates a reverberation audio signal based on the first set of reverberation parameters and a renderer (507) renders an audio signal for the listening acoustic environment including a component derived from the reverberation audio signal.

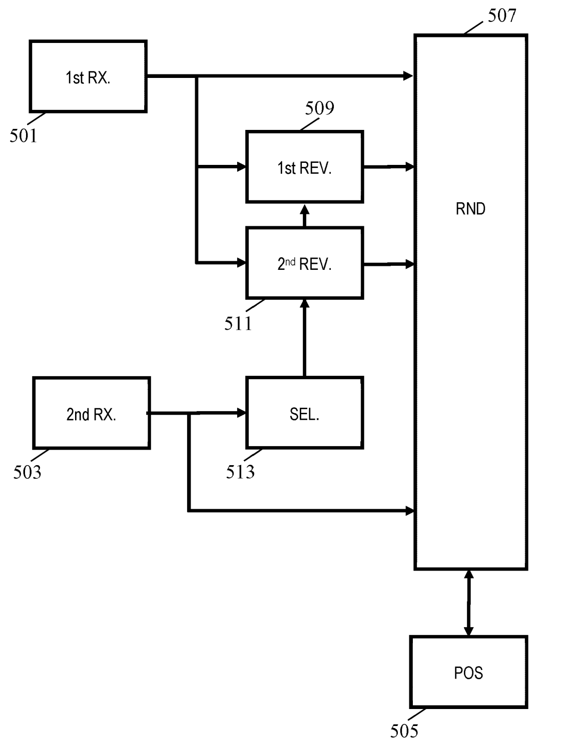

FIG. 5

Processed by Luminess, 75001 PARIS (FR)

Description

FIELD OF THE INVENTION

[0001] The invention relates to an audio apparatus and a method of operation therefor, 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

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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 anthropometries 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:

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

[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 fully 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 is typically 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.

[0028] In particular, in many scenarios, a scene may include multiple environments with the sound of some environments also being audible in other environments. For example, for a multi-room scene, the audio from other rooms can often be heard in the room in which the listener is present (referred to as the listener or listening room/ acoustic environment). It is therefore desirable to render audio to the listener that includes representations of the audio from other acoustic environments. In particular, it is often desirable to render audio representing the diffuse/reverberant audio from other acoustic environments/ rooms as this tend to result in a more realistic representation and perception. However, in order to provide a more realistic user perception, such a rendering should accordingly reflect the acoustic properties of the source room rather than (or possibly in addition to) the acoustic properties of the listening room.

[0029] Typical approaches for rendering audio tend to be suboptimal in some scenarios, including in particular when rendering audio for scenes that include different acoustic rooms or environments.

[0030] A particular issue in many scenarios is that audio rendering tends to be complex and time and/or resource demanding. In particular, when rendering audio from multiple acoustic environments, and needing to represent acoustic properties of these acoustic environments, multiple rendering paths are often required resulting in a high complexity and resource usage. For example, to include reverberation audio from other rooms than the listening room typically requires additional reverberators in order to generate the relevant reverberation sound. The trade-off between audio quality/ listener perception and complexity/ resource demand is a critical trade-off in many systems.

[0031] Hence, an improved approach for rendering audio for a scene with multiple acoustic environments 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, improved resources usage/ allocation, facilitated rendering, improved rendering of audio from multiple acoustic environments, improved performance for virtual/mixed/ augmented reality applications, increased processing flexibility, a more natural sounding audio rendering, improved audio rendering for multi-room scenes, improved trade-off between perceived audio quality/ realism and computational burden/ complexity, and/or improved performance and/or operation would be advantageous.

SUMMARY OF THE INVENTION

[0032] Accordingly, the Invention seeks to preferably mitigate, alleviate or eliminate one or more of the above-mentioned disadvantages singly or in any combination.

[0033] According to an aspect of the invention there is provided an audio apparatus comprising: 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 plurality of sets of reverberation parameters, each set of reverberation parameters being for one associated acoustic environment and comprising at least one reverberation parameter indicative of a relationship between a level of reverberation in the one associated acoustic environment and a level of an audio source in the one associated acoustic environment; and at least one energy transfer parameter, each energy transfer parameter being indicative of an energy attenuation for audio propagation between a source acoustic environment and a destination acoustic environment; a selector arranged to select for a first acoustic environment at least a first set of reverberation parameters of the multiple acoustic environments, the first set of reverberation parameters being for a different acoustic environment than the first acoustic environment and the selection being in response to parameter values of the first set of reverberation parameters and an energy transfer parameter indicative of energy attenuation for audio propagation between the different acoustic environment and the first acoustic environment; a plurality of reverberators arranged to generate reverberation audio signals, at least one reverberator being arranged to generate a reverberation audio signal based on the first set of reverberation parameters and an audio source of the different acoustic environment of the first set of reverberation parameters; and a renderer for generating an audio signal for the first acoustic environment based on the audio data, the audio signal including audio components derived from the reverberation audio signals.

[0034] The approach may allow an audio signal to be generated that provides an improved user experience for audio scenes with multiple acoustic environments, and often a more realistic and naturally sounding audio experience can be achieved. 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.

[0035] 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.

[0036] The approach may provide improved, increased, and/or facilitated flexibility and/or adaptation of the processing and/or the rendered audio.

[0037] An energy transfer parameter indicating an energy attenuation between a pair of transfer regions may be equivalent to an energy transfer parameter indicating an energy transfer between a pair of transfer regions. An increasing attenuation is indicative of a reduced proportion of audio energy from the source acoustic environment reaching the destination acoustic environment, corresponding to a reduced energy transfer. A decreasing attenuation is indicative of an increased proportion of audio energy from the source acoustic environment reaching the destination acoustic environment, 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.

[0038] The audio energy/level (or just energy/ level) may specifically be represented by a level, amplitude, power, or time averaged energy measure.

[0039] 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.

[0040] In some embodiments, an energy transfer parameter may be indicative of an energy attenuation for audio propagation between a transfer region of an acoustically attenuating boundary of a source acoustic environment and a transfer region of an acoustically attenuating boundary of a destination acoustic environment. 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).

[0041] The selector is arranged to select a plurality of sets of parameters, and in many scenarios/ embodiments for a plurality of acoustic environments other than the first acoustic environment. In some embodiments, the selector may be arranged to generate a rank value for a plurality of sets of reverberation parameters and to select the at least first set of reverberation parameters in response to the rank value, the rank value for a given set of reverberation parameters being determined in response to parameter values of the given set of reverberation parameters and an energy transfer parameter indicative of energy attenuation between an acoustic environment for which the set of reverberation parameters is provided and the first acoustic environment.

[0042] A set of reverberation parameters for a given acoustic environment comprise one or more parameters indicative of properties of reverberation in the given acoustic environment resulting from one (or more) audio sources in the given acoustic environment. The set of reverberation parameters may include parameters indicative of the level/ energy attenuation/ conversion from an audio source to the level/ energy of diffuse/ reverberation audio, a delay of diffuse/ reverberation audio, and/or a decay time for diffuse/ reverberation audio.

[0043] In accordance with an optional feature of the invention, the selector is arranged to select between sets of reverberation parameters based on a combined attenuation for each set of reverberation parameters, the combined attenuation for a given set of reverberation parameters being indicative of a level of audio in the first acoustic environment resulting from reverberation in a given acoustic environment of the given set of reverberation parameters resulting from reverberation in the given acoustic environment caused by an audio source in the given acoustic environment, the combined attenuation for the given set of reverberation parameters including a contribution from an attenuation indicated by the at least one reverberation parameter of the given set of reverberation parameters and an energy transfer parameter indicative of energy attenuation for audio propagation between the given acoustic environment and the first acoustic environment.

[0044] 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 provide a practical and/or improved selection of which reverberation to render. In many scenarios, it may provide an improved trade-off between complexity and resource usage on one hand and audio quality and user experience on the other.

[0045] In accordance with an optional feature of the invention, the selector is arranged to select a first set of reverberation parameters over a second set of reverberation parameters if the combined attenuation for the second set of reverberation parameters exceeds the combined attenuation for the first set of reverberation parameters.

[0046] 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. An improved utilization of limited computational resource can often be achieved.

[0047] In accordance with an optional feature of the invention, the selector is arranged to discard a set of reverberation parameters from selection if the combined attenuation exceeds a threshold.

[0048] 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. An improved utilization of limited computational resource can often be achieved.

[0049] In accordance with an optional feature of the invention, the metadata further comprise audio level data indicative of signal levels for audio sources, and the selector is arranged to determine resulting signal levels in the first acoustic environment for audio sources in acoustic environments of the sets of reverberation parameters based on signal levels of the audio sources and the combined attenuation for the sets of reverberation parameters ; and to select the at least one set of reverberation parameters in response to the resulting signal levels in the first acoustic environment.

[0050] This may provide improved selection in many embodiments and scenarios and may result in particularly advantageous performance and/or operation.

[0051] The selector may be arranged to determine a resulting signal level in the first acoustic environment for a given set of reverberation parameters in response to an audio signal level for at least one audio source in a given acoustic environment of the given set of reverberation parameters and the combined attenuation for the given set of reverberation parameters. The selector may be arranged to determine resulting signal levels for a plurality of sets of reverberation parameters, and possibly for all sets of reverberation parameters.

[0052] In accordance with an optional feature of the invention, the selector is arranged to combine levels of multiple sources in at least one acoustic environment into a combined audio source level and to determine the resulting signal level for a set of reverberation parameters of the at least one acoustic environment by applying the combined attenuation to the combined audio source level.

[0053] 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.

[0054] In accordance with an optional feature of the invention, the selector is arranged to determine a number of the plurality of reverberators available for generating reverberation signals for audio sources in acoustic environments other than the first acoustic environment; and to adapt a number of sets of reverberation parameters being selected to match the number of the plurality of reverberators available.

[0055] 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 allow a flexible and dynamic resource adaptation.

[0056] In accordance with an optional feature of the invention, the selector is arranged to determine a first number of reverberation signal sources in the first acoustic environment and to subtract the first number from a number of reverberators of the audio apparatus to determine the number of the plurality of reverberators available.

[0057] 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 allow a flexible and dynamic resource adaptation.

[0058] In accordance with an optional feature of the invention, at least one of the at least one reverberation parameter and the energy transfer parameter is frequency dependent.

[0059] This may provide improved performance and/or facilitated implementation in many scenarios.

[0060] In accordance with an optional feature of the invention, the selector is arranged to repeatedly perform selection of the at least one set of reverberation parameters.

[0061] 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 reselection may be performed with an interval between selections (e.g. on average) not exceeding 1,2,5,10, 20 seconds in different embodiments.

[0062] In accordance with an optional feature of the invention, at least a first reverberator of the plurality of reverberators is arranged to continue to generate a reverberation audio signal for a first set of reverberation parameters for a time period following deselection of the first set of reverberation parameters.

[0063] 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. It may in particular, reduce or mitigate the perceptual impact of switching/ reselection effects.

[0064] The level of the reverberation audio signal may be gradually reduced during the time period. The reverberation audio signal may be faded out during the time period.

[0065] In accordance with an optional feature of the invention, the first reverberator has lower computational resource usage than a reverberator of the plurality of reverberators used to generate reverberation audio signals for the first set of reverberation parameters prior to deselection of the first set of reverberation parameters.

[0066] This may provide particularly advantageous operation and/or performance in many embodiments.

[0067] In accordance with an optional feature of the invention, the at least one reverberation parameter comprises a

Diffuse to Source Ratio parameter indicative of a ratio between emitted audio source energy and a diffuse reverberation energy.

[0068] This may provide particularly advantageous operation and/or performance in many embodiments.

[0069] According to an aspect of the invention there is provided a method of operation for an audio apparatus 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 plurality of sets of reverberation parameters, each set of reverberation parameters being for one associated acoustic environment and comprising at least one reverberation parameter indicative of a relationship between a level of reverberation in the one associated acoustic environment and a level of an audio source in the one associated acoustic environment; and at least one energy transfer parameter, each energy transfer parameter being indicative of an energy attenuation for audio propagation between a source acoustic environment and a destination acoustic environment; selecting, for a first acoustic environment, at least a first set of reverberation parameters of the multiple acoustic environments, the first sets of reverberation parameters being for a different acoustic environment than the first acoustic environment and the selection being in response to parameter values of the first set of reverberation parameters and an energy transfer parameter indicative of energy attenuation for audio propagation between the different acoustic environment and the first acoustic environment; a plurality of reverberators (509, 511) generating reverberation audio signals, at least one reverberator being arranged to generate a reverberation audio signal based on the first set of reverberation parameters and an audio source of the different acoustic environment of the first set of reverberation parameters; and generating an audio signal for the first acoustic environment, the audio signal including audio components derived from the reverberation audio signals.

[0070] 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; and metadata for the audio data, the metadata comprising: a plurality of sets of reverberation parameters, each set of reverberation parameters being for one associated acoustic environment and comprising at least one reverberation parameter indicative of a relationship between a level of reverberation in the one associated acoustic environment and a level of an audio source in the one associated acoustic environment; and at least one energy transfer parameter, each energy transfer parameter being indicative of an energy attenuation for audio propagation between a source acoustic environment and a destination acoustic environment.

[0071] 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

[0072] 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 Jot reverberator;
FIG. 7 illustrates an example of a Feedback Delay Network reverberator;
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 attenuation for sound propagation between different acoustic environments;
FIG. 10 illustrates an example of a sound attenuation for sound propagation between different acoustic environments; and
FIG. 11 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

[0073] 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.

[0074] 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.

[0075] 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.

[0076] 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.

[0077] 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.

[0078] 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.

[0079] 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.

[0080] 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.

[0081] 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.

[0082] 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.

[0083] 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.

[0084] 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.

[0085] 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 (T60) or a reverberation coloration.

[0086] 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.

[0087] 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.

[0088] 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.

[0089] 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.

[0090] 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.

[0091] 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.

[0092] In the following, advantageous approaches will be described for rendering an audio scene that includes multiple rooms/ acoustic environments.

[0093] FIG. 5 illustrates an example of an audio apparatus that is arranged to render audio for 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.

[0094] 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.

[0095] 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.

[0096] 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.

[0097] 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.

[0098] 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.

[0099] 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.

[0100] 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.

[0101] In the present example, much of the description related to reverberation and diffuse audio is dependent mainly on which acoustic environment/ room is the listening acoustic environment/ room rather than on the specific position of a listener within the listening acoustic environment/ room. Accordingly, in some embodiments, the listening position may simply represent the listening acoustic environment/ room, i.e. the acoustic environment/ room in which the listener is present.

[0102] 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 as diffuse signals with no specific position, etc.

[0103] The audio apparatus of FIG. 5 comprises a plurality of reverberators. FIG. 5 illustrates a first reverberator 509 and a second reverberator 511 but it will be appreciated that this is merely for illustration and that an audio apparatus in other embodiments may include more than two reverberators. For brevity and clarity, much of the following description focusses on embodiments where the audio apparatus includes only two reverberators.

[0104] The audio apparatus thus includes a bank/ plurality of reverberators. Each of the reverberators is arranged to generate a diffuse/ reverberation signal. Each of the reverberators may receive an audio signal corresponding to an audio source or a plurality of audio sources and generate a reverberation/ diffuse audio signal. The reverberator may further receive a set of reverberation parameters that are used to adapt the reverberator to generate a reverberation/ diffuse signal with properties that reflect the acoustic environment/ room for which the signal is generated. For example, a parameter indicating e.g. a proportion of the input audio signal energy or emitted source energy that converts into reverberation energy may be provided (e.g. a DSR value), a parameter indicating an onset/ delay time for reverberation/ diffuse energy (e.g. a delay for the reverberation tail of the room impulse response), and/or a parameter indicating a decay speed for the amplitude (such as T60 value) may be provided.

[0105] In many other embodiments, the audio signal being processed by the reverberators may be a downmix of the audio signals/ sources in the relevant room/ acoustic environment. For example, in the example of FIG. 5, each reverberator may include a downmixer and/or a delay that generates the audio signal to be processed by the reverberator from the signals of the audio sources in the audio data. The audio signals from the first receiver (still separated into a signal per source) may be fed to the reverberator(s). In that case each reverberator may generate a suitable weighted downmix from these signals as needed for that acoustic environment, and additionally have a (reverberator-specific) delay to align with the direct path rendering.

[0106] The reverberators 509, 511 are coupled to the renderer 507 to which the reverberation signals are fed. The renderer 507 may then proceed to combine the reverberation signals with path signals representing the individual paths for audio sources in the listening room to generate a combined audio signal that represents the combined sound in the environment as perceived by the listener.

[0107] Each of the reverberators of the bank of reverberators may comprise (or be) a parametric reverberator, such as a Feedback Delay Network (FDN) reverberator, and specifically a Jot Reverberator.

[0108] An example of a suitable reverberator is the Jot reverberator illustrated in FIG. 6. This reverberator includes a loop input vector b and a loop extraction matrix C to control how input samples are distributed over the feedback loops of the reverberator and how the output signals are generated from the loops.

[0109] An example of a reverberator based on a feedback delay network of the input signal is illustrated in FIG. 7 where three feedback loops are illustrated. A Jot reverberator can be considered a specific example of a feedback delay network reverberator.

[0110] A feedback delay network may comprise a plurality of feedback loops where each (or at least one) feedback loop has an input receiving an input audio signal and where each feedback loop implements a loop transfer function (which specifically may be a delay), a feedback network feeding output signals of the feedback loops back to inputs of the loops to be combined with the input audio signal, and an output circuit arranged to generate an output signal of the feedback delay network as a combination of the output signals of the feedback loops. The feedback network may for each feedback loop implement a feedback path for the output signal of the feedback loop to an input of the feedback loop, and typically may also implement a feedback path to one or more inputs of other feedback loops. In many embodiments, the feedback network may implement a feedback path from the output of each feedback loop to each input of all feedback loops. Each feedback path typically implements an attenuation factor (or equivalently a gain factor) but may in some embodiments provide a more complex feedback path, such as e.g. implementing a frequency dependent gain (e.g. it may implement a filter function). In some embodiments, the loop transfer function may be a filter implementing both the desired frequency response and gain factors and the feedback bath may simply be a flat unity gain feedback (e.g. corresponding to a feedback matrix representing the feedbacks having coefficients of one on the diagonal). In many embodiments, the feedback network may be represented by a feedback matrix having a coefficient for each feedback loop pair combination.

[0111] Feedback delay networks are typically based on feedback loops with different delays in them. Input signals are inserted in the loops and with appropriate feedback gains, the signals are fed back into the loops. Output signals are extracted by combining signals in the loops. Signals fed in are therefore continuously repeated with different delays. Using delays that are mutually prime and having a feedback matrix that mixes signals between loops can create a pattern that is similar to reverberation in real spaces, and is particularly suitable for generating diffuse reverberation as in the example of a Jot or other parametric reverberator.

[0112] The absolute value of the elements in the feedback matrix are designed to be below one in order to achieve a stable, decaying impulse response. The coefficients can be set in combination with the delays to achieve a desired reverberation time (T60). In many implementations, additional gains or filters are included in the loops. These filters can control the attenuation instead of the matrix. Using filters has the benefit that the decaying response can be different for different frequencies. Thus, the gains, delays, path transfer functions can be set based on reverberation parameters characterizing the desired properties of the reverberation (and typically the room properties).

[0113] 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.

[0114] Specifically, the renderer 507 may generate audio for the current listening acoustic environment/ room which represents reverberant diffuse audio of other acoustic environments/ rooms. Such reverberant diffuse audio of a different acoustic environment/ room may be rendered as a localized source (e.g. positioned at a door or window in a wall of the listening room), or may be rendered as reverberant/ diffuse audio in the listening acoustic environment/room.

[0115] 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.

[0116] 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 various such approaches.

[0117] In particular, the audio apparatus may be arranged to render an audio signal for a listener in a given listener room with the audio signal comprising different audio components resulting from different audio sources and/or acoustic environments.

[0118] In particular, the renderer 507 may receive audio data and metadata for audio sources that are present in the listening room and render these audio sources using conventional approaches. For example, some audio sources in the room are rendered as point audio sources with a position as indicated in the metadata. It will be appreciated that many such different rendering algorithms are known and that any suitable approach may be used without detracting from the invention.

[0119] In addition, a reverberation signal component may in some scenarios and embodiments also be generated for audio sources that are present in the room. For example, a set of reverberation parameters may be provided for the listening room and these may be used to adapt a first reverberator 509 of the bank of reverberators 509, 511 which may also be fed a combined audio signal representing all the audio sources in (or contributing to) the listening room. The first reverberator 509 may then generate a listening room reverberation signal component which is fed to the renderer 507 and which by the renderer 507 is combined with the point source audio rendering.

[0120] In addition, the audio apparatus may include signal components representing audio sources in other rooms, and specifically it may include signal components that represents reverberation audio from other rooms. The audio apparatus may specifically use one or more of the reverberators 509, 511 to generate reverberation signals that represent the reverberation in another room. For example, the second reverberator 511 may be used to generate a reverberation signal for another room by providing an audio signal representing one or more audio sources in a second room and adapting the second reverberator 511 based on a set of reverberation parameters representing reverberation properties for the second room. The resulting reverberation signal may then be adapted, and specifically attenuated, to reflect that the reverberation audio of the second room propagates to the listening room. The resulting reverberation audio signal may then be included in the output audio signal by the renderer 507.

[0121] In many embodiments, the reverberation signal from another acoustic environment may be included in the output signal by rendering the signal as being a localizable audio source, such as a point source, or an audio source with a (limited) extent, e.g. positioned at a transfer region of an acoustically attenuating boundary, such as an opening in a wall of the room. In some embodiments and scenarios, the reverberation signal may be included as a reverberation signal, and indeed it may in some cases be included directly (e.g. after a suitable scaling/ amplification). In some embodiments, the generated reverberation signal for the second room may even be used as an input signal for a reverberator that is adapted based on reverberation parameters for the listening room. This may represent a scenario where the reverberation sound from another room propagates to the listening room where it is subjected to further reverberation. However, in most embodiments, a realistic and high quality audio rendering and perception is achieved without necessitating a second reverberation processing for the listening room thereby allowing reduced complexity and resource usage.

[0122] 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.

[0123] The metadata may include various data that allows the rendering of audio for the scene.

[0124] In particular, the metadata comprises a plurality of sets of reverberation parameters where each set is provided for one acoustic environment. For example, in some embodiments, one set of reverberation parameters may be provided for each acoustic environment of the scene, such as specifically for each room of a building. However, it will be appreciated that in some embodiments, no set of reverberation parameters may be provided for some acoustic environments, and/or more than one sets of reverberation parameters may be provided for some acoustic environments (for example different sets of reverberation parameters may be provided depending on the position of the listener in the acoustic environment).

[0125] The sets of reverberation parameters include parameters that characterize how audio sources/ signals in the acoustic environment for which the set of reverberation parameters is provided result in/ convert to reverberation. The parameters of a set of reverberation parameters may thus provide information of the reverberation that results in a given acoustic environment from an audio source in that acoustic environment. The sets of reverberation parameters may for example include parameters indicative of room dimensions and acoustic properties of the rooms.

[0126] Each set of reverberation parameters specifically comprise a parameter which is indicative of a relationship between a level/ energy of reverberation and a level/ energy of audio sources in the acoustic environment for which the set of reverberation parameters is provided. This parameter will for brevity also be referred to as a level parameter.

[0127] The level parameter for a given acoustic environment thus indicates the level of reverberation in an acoustic environment/ room that results from a given level of an audio source in the acoustic environment. The level parameter may thus be useful for determining the level of reverberation that is present in a given acoustic environment considering the audio sources therein.

[0128] The sets of reverberation parameters may specifically include a DSR value that directly indicates the ratio between the diffuse audio energy/ level and the (emitted) audio source energy/ level. Using a DSR value may be particularly advantageous in many embodiments and the following description will focus on examples where the level parameter is a DSR parameter.

[0129] In many embodiments, each set of reverberation parameters may comprise a parameter indicating e.g. a proportion of the input audio signal energy that converts into reverberation energy (e.g. a DSR value) as well as often a parameter indicating an onset/ delay time for reverberation/ diffuse energy (e.g. a delay/ lag for the assumed starting point of the reverberation tail of the room impulse response), and/or a parameter indicating a decay speed for the amplitude (such as T60 value). The set of reverberation parameters may in many embodiments provide all the information required for adapting a reverberator to generate a reverberation signal representing reverberation in a given acoustic environment resulting from audio sources in that acoustic environment. Thus, by setting up a reverberator based on a set of reverberation parameters and applying an input signal representing the audio sources in the acoustic environment of the set of reverberation parameters, the reverberator can generate a reverberation signal for the acoustic environment. The reverberation signal may be generated to have a level (energy/ amplitude), delay, and decay time that reflects the audio source and the acoustic environment characteristics.

[0130] The metadata further comprises one or typically more energy transfer parameters where each energy transfer parameter is indicative of an energy attenuation for audio propagation between a source acoustic environment and a destination acoustic environment. The energy transfer parameter may indicate the level attenuation from audio in the source acoustic environment to audio in the destination acoustic environment. The energy transfer parameter for a pair of acoustic environments may thus reflect the sound level in the destination acoustic environment from sound in the source acoustic environment, and thus indicate how loud the sounds of the source acoustic environment are in the destination acoustic environment.

[0131] The energy transfer parameters may directly reflect all parameters of sound propagation from the source acoustic environment to the destination acoustic environment, and thus include how sound may propagate in free space, through acoustically attenuating boundaries, through openings in the acoustically attenuating boundaries, etc.

[0132] In many embodiments, audio from one acoustic environment to another acoustic environment may predominantly propagate through transfer regions of acoustically attenuating boundaries between the acoustic environments. For example, for a building, sound may predominantly propagate between rooms through openings in the walls, such as through door and window openings. In such cases, the reverberation from a different room may in the listening room be rendered as an audio signal with the audio properties of the reverberation but with a position corresponding to an opening in the wall. E.g. the reverberation of a second room may be heard as reaching the user from the opening in the wall. The energy transfer parameter(s) may in some embodiments directly indicate the energy attenuation for audio propagation between a transfer region of the source acoustic environment and a transfer region of the destination acoustic environment.

[0133] The energy transfer parameter may thus indicate an energy attenuation between a pair of acoustic environments. The energy attenuation for a pair of acoustic environments may be indicative of a proportion of audio energy in one acoustic environment of the pair of acoustic environments that propagates to the other acoustic environment of the pair of acoustic environments.

[0134] The audio apparatus of FIG. 5 may as mentioned be arranged to render audio for a position in a first/ listening acoustic environment. The rendered audio may include rendering of audio sources with specific positions in the listening acoustic environment (including rendering direct path and reflected path audio components). In addition, the audio apparatus may optionally generate a reverberation signal component for reverberation caused by audio sources in the listening acoustic environment. Further, the audio apparatus may also be arranged to generate signal components that reflect reverberation audio in other acoustic environments than the listening acoustic environment. The audio apparatus may specifically generate a reverberation signal for one or more other acoustic environments and include them as reverberation signal components in the rendered output audio signal and/or render them as audio sources with positions (possible as extended audio sources).

[0135] The audio apparatus includes a bank of reverberators 509, 511 that are used to generate the reverberation signal components based on the set of reverberation parameters for the corresponding acoustic environment. Reverberators may be used both for generating the reverberation signal of the listening acoustic environment as well as for other acoustic environments. For other acoustic environment than the listening acoustic environment, the level of the generated reverberation signal component is determined based on the energy transfer parameter for the given acoustic environment and the listening acoustic environment. Specifically, the reverberation signal of the acoustic environment is attenuated by a value determined from the energy transfer parameter.

[0136] In the ideal case, a reverberation signal is generated for each set of reverberation parameters, i.e. for all acoustic environments for which a set of reverberation parameters is provided (including generating multiple reverberation signals for acoustic environments having multiple associated parameter sets). However, such an approach would typically require a large number of reverberators which are typically quite complex and has a high resource usage. Thus, such an approach is impractical for all but the simplest scenes and most powerful processing platforms.

[0137] The audio apparatus of FIG. 5 comprises a selector 513 which is arranged to select a number of the sets of reverberation parameters for which to generate a reverberation signal component. The selector 513 may select the sets of reverberation parameters from the sets of reverberation parameters for which the listening acoustic environment is the destination acoustic environment, i.e. for which the set of reverberation parameters includes an energy transfer parameter that indicates an attenuation from another acoustic environment to the listening acoustic environment.

[0138] The selector 513 is arranged to select the sets of reverberation parameters based on the parameters of the sets themselves. For example, the selector 513 may include a criterion or equation for generating a rank value (or equivalently a cost value or strength value) based on one or more parameter values of the sets of reverberation parameters. The rank value may specifically be one generated to provide an indication of the perceptual significance of the reverberation sound from the corresponding other acoustic environment to the listening acoustic environment. This may for example include a weighting of cost values reflecting the energy of the diffuse signal (e.g. represented by a DSR) and the decay time (e.g. represented by a T60), as well as e.g. the propagation proportion (e.g. represented by the energy transfer parameter) such that low level and quickly decaying reverberations are rejected in favor of higher level and slower decaying reverberations. It will be appreciated that the specific equation or algorithm used to determine the rank value (or cost value) will depend on the specific preferences and requirements of the individual embodiment.

[0139] The selector 513 may then proceed to select a given number of sets of reverberation parameters as the ones that have the highest rank values. It may then allocate a reverberator to each of the selected sets of reverberation parameters and proceed to generate the corresponding reverberation signal components for inclusion in the rendered output audio signal.

[0140] The selector 513 may thus be used to select a subset of sets of reverberation parameters that are rendered. The selection may be based on a ranking of the sets of reverberation parameters that may reflect the estimated perceptual significance of the corresponding reverberation in the listening acoustic environment.

[0141] The number of sets of reverberation parameters that are selected may be different in different embodiments.

[0142] In many embodiments, the selector 513 may determine how many reverberators are available for generating reverberation signals for other acoustic environments. The number of sets of reverberation parameters that are selected may then be adapted accordingly and specifically in many embodiments the same number of sets of reverberation parameters are selected as there are reverberators available.

[0143] In some embodiments, the audio apparatus may comprise a fixed number of reverberators allocated to generate reverberation signals for other acoustic environments and the selector 513 may be arranged to fixedly select a corresponding number of sets of reverberation parameters.

[0144] In other embodiments, the number of available reverberators may vary and the selector 513 may be arranged to dynamically determine the number of available reverberators and to select a corresponding number of sets of reverberation parameters.

[0145] For example, in some embodiments, the reverberators may be implemented as software and the number of reverberators that can be implemented at a given time may depend on how much computational resource is available at a given time. For example, if a large number of other complex operations are being performed, there may only be sufficient resources for a relatively small number of reverberators to be implemented and thus only a low number of reverberators may be available to the selector 513. If instead, no other complex operations are currently being performed, the number of reverberators that can be implemented may be higher and thus more reverberators may be available to the selector 513. The selector 513 may accordingly determine the number of reverberators available in dependence on a (current) computational loading of the audio processing apparatus. The selector 513 may then select a corresponding number of sets of reverberation parameters and allocate one selected set of reverberation parameters to each available reverberator.

[0146] In some embodiments, the number of available reverberators may be dependent on a number of reverberators that are used to render reverberation in the listening acoustic environment. For example, for the current listening acoustic environment there may be none, one, or more sets of reverberation parameters depending on how reverberant the acoustic environment is (and on whether this may e.g. be different for different areas of the acoustic environment).

[0147] The audio apparatus may be arranged to prioritize local reverberation higher than reverberation from other acoustic environments, and accordingly it may allocate a reverberator to each set of reverberation parameters. Accordingly, the audio apparatus may allocate none, one, or more reverberators to generating reverberation signals for audio sources in the listening acoustic environment. The selector 513 may then proceed to determine how many reverberators remain and are available for rendering of reverberation from other acoustic environments. The selector 513 may then proceed to select that number of sets of reverberation parameters from other acoustic environment and allocate a reverberator to each of the selected set of reverberation parameters. Thus, in such an embodiment, the generation of local reverberation for local sources in the listening acoustic environment is prioritized higher than reverberation in other acoustic environments. However, reverberators that are not used for generating local reverberation signals are allocated to generate reverberation in other acoustic environments.

[0148] Thus, in some embodiments, the selector 513 may be arranged to determine a number of sets of reverberation parameters that are provided for the listening acoustic environment and to subtract this number from the total number of reverberators of the audio apparatus to determine the number of reverberators that are available.

[0149] In some embodiments, the selector 513 may be arranged to rank all sets of reverberation parameters that include the listening acoustic environment as a destination acoustic environment for an energy transfer parameter that has the acoustic environment of the set of reverberation parameters as a source acoustic environment, i.e. for all sets of reverberation parameters that may result in a contribution of audio to the listening acoustic environment. The selector 513 may then select the highest ranked sets of reverberation parameters and allocate them to a reverberator until all reverberators have been allocated. Such an approach may be advantageous in many embodiments and may provide an improved perceived realism with reverberation being generated for the most likely signals to have a perceptual impact.

[0150] In many embodiments, the selector 513 may be arranged to determine a (combined) attenuation indicative of at least part of a reduction in the level of an audio source in a source acoustic environment to a level of reverberant audio in the listening acoustic environment. The combined attenuation may specifically be indicative of a reduction in the level of reverberation in the source acoustic environment relative to the level of the audio source and a reduction in the level of the reverberation sound resulting from the propagation from the source acoustic environment to the listening acoustic environment.

[0151] The level reduction resulting from the conversion from the audio source energy to reverberation energy in the source acoustic environment may be indicated by a reverberation parameter and the combined attenuation may be determined in response to the reverberation parameter. As a specific example, the sets of reverberation parameters may comprise a DSR value that indicates the ratio of reverberation/ diffuse energy resulting from a given source energy.

[0152] The combined attenuation may thus be determined in dependence on a reverberation parameter that is indicative of a relationship between a level of reverberation in the source acoustic environment and a level of audio sources in the source acoustic environment, i.e. based on the level parameter.

[0153] The attenuation from the source acoustic environment to the listening acoustic environment may be represented by the energy transfer parameter for the source acoustic environment and the listening acoustic environment. The combined attenuation may be determined in response to this energy transfer parameter.

[0154] The combined attenuation may thus be indicative of the combined effect of attenuation in the source acoustic environment when converting from the audio source energy to reverberation energy and the effect of attenuation of propagation from the source acoustic environment to the destination/ listening acoustic environment.

[0155] In some embodiments, the combined attenuation may for example be determined directly as the combined attenuation of the attenuation represented by the level parameter and the attenuation represented by the energy transfer parameter between the source acoustic environment and the listening acoustic environment. The combined attenuation may in some embodiments simply be determined as these two attenuations being combined by addition (when in a logarithmic/ dB domain) or multiplication (when in a linear domain).

[0156] In some cases, the selector 513 may be arranged to consider other aspects in determining a combined attenuation. For example, if there is more than one sets of reverberation parameters for the source acoustic environment, the selector 513 may be arranged to perform interpolation between these to determine suitable attenuations to include.

[0157] The selector 513 may be arranged to select between the sets of reverberation parameters based on the combined attenuation, and indeed in some cases based only on the combined attenuation. In particular, in some embodiments, the selector 513 may be arranged to perform the selection without considering the signal level of the audio sources.

[0158] For example, in some embodiments, the selector 513 may for a given listening acoustic environment be arranged to determine a combined attenuation for each set of reverberation parameters for an acoustic environment that is a source acoustic environment for an energy transfer parameter that further has the listening acoustic environment as a destination acoustic environment. It may then rank all the determined combined attenuations and select between these in the order of increasing attenuation. Thus, the selector 513 may be arranged to select a first set of reverberation parameters over a second set of reverberation parameters if the first set has a combined attenuation that is lower than the combined attenuation of the second set of reverberation parameters.

[0159] For example, the number of available reverberators may be determined and the corresponding number of sets of reverberation parameters that have the lowest combined attenuation may be selected.

[0160] In some embodiments, the selector 513 may be arranged to discard sets of reverberation parameters from selection if the combined attenuation for the set of reverberation parameters is above a threshold. For example, it may be determined if the combined attenuation is simply too high, the corresponding audio in the listening acoustic environment will not be sufficiently loud to have a significant perceptual effect no matter how high the signal source level is. For example, if the combined attenuation is above, say, 60dB, the resulting sound in the listening acoustic environment will tend to be insignificant even if the audio source is loud, and accordingly the set of reverberation parameters can be ignored.

[0161] In some embodiments, the threshold may be a dynamic threshold that can be varied. For example, in many embodiments, the selector 513 may be arranged to determine the threshold based on a (e.g. combined) level/ energy of audio sources in the listening acoustic environment (thus whether reverberation audio from other acoustic environments are rendered may depend on how loud the local sources are). Such an approach may thus take into account the masking effect of local audio sources.

[0162] The approach may be highly advantageous in many embodiments. It may for example allow the resource allocated to the reverberation processing to be adapted to reflect the perceptual significance. For example, in some embodiments, the number of reverberators that are used may be limited to only sets of reverberation parameters that are likely to have a significant perceptual impact. In some embodiments, computational resource may dynamically be re-allocated to other functions if not required for generation of reverberation components.

[0163] In some embodiments, the selector 513 may be arranged to determine resulting signal levels in the listening acoustic environment and the selection may be based on these signal levels.

[0164] Specifically, in some embodiments, the metadata may include audio level data which indicates the signal levels of audio sources. The selector 513 may, based on these signal levels and the combined attenuation, proceed to determine the resulting signal levels in the listening acoustic environment. For example, in some embodiments, the resulting signal level may simply be determined as the signal level of the audio source attenuated by a value corresponding to the combined attenuation.

[0165] In many embodiments, the selector 513 may be arranged to generate the signal levels that result from all the reverberation in the source acoustic environments rather than from individual signal sources. The selector 513 may be arranged to combine the levels of multiple sources in the source acoustic environment into a combined audio source level for the source acoustic environment. For example, in some cases, the levels may simply be added together to generate a combined audio source level. The combined attenuation may then be applied to the combined audio source level to generate the resulting (combined) signal level at the listening acoustic environment.

[0166] The selector 513 may in some embodiments be arranged to select the sets of reverberation parameters based on the resulting signal levels in the listening acoustic environment. For example, a given number of sets of reverberation parameters may be selected as the number of sets of reverberation parameters that have the highest signal level.

[0167] The approach may in many embodiments and scenarios provide an improved rendering of a scene that includes multiple environments. Reverberation for environments(s) that the listener is not in, may be rendered e.g. as a localizable source. Typically, this may reflect the reverb being audible through portals (openings through which audio propagates between environments). However, the approach takes into consideration the realization that in many cases it may not be necessary to render audio from all other environments because they do not contribute significantly to the listener's perception of the scene. This may be used to reduce computational complexity for rendering such scenes, or to scale computational complexity to available or assigned capability.

[0168] In the approach, the reverberation of other acoustic environments are represented by sets of reverberation parameters. Some individual acoustic environments may be represented by more than one set of parameters. This often means a reverberator is instantiated for each of the set of parameters, and that their output is interpolated depending on the listener's position. Also here, it can happen that some contribute very little to the listener's perception of the (virtual) environment.

[0169] The described approach may provide an efficient and high performance way of determining which set of reverberation parameters/ acoustic environments to render. The selection may be influenced by several factors:

- Presence of audio sources in the environment.
- The level of the reverberation audio.
- The transfer function between the environment's portal(s) to the listener position.
- The level of the reverberation audio combined with the transfer function, relative to the level of other audio presented to the listener.

[0170] As an example, a floorplan of rooms is shown in FIG. 8. In the example, some environments / rooms (F, G) may not have any sources by themselves, and may therefore not emit any reverberation. However, they may receive reverberation from external sources (e.g. room G's reverberation may be significant from the source in room H).

[0171] With the listener in room E, the transfer of reverberation energy over a long distance and/or with obstacles may be too weak to be perceivable (e.g. from room H).

[0172] Sources in some nearby rooms may give rise to too weak reverberation energy and thus have little audible reverberation energy at the listening position (e.g. room A), while a much stronger reverberation in another environment (e.g. room B) may yield significant energy at the listening position for similar source levels in room A and B.

[0173] The audio apparatus of FIG. 5 may dynamically adapt the rendering to reflect such properties and characteristics.

[0174] As sources may be animated, doors may open or close and the listener moves around, the selection of rooms may constantly be updated. The disabling and (re-) enabling of reverberators may also be controlled so as to avoid switching artefacts.

[0175] As a specific example, the selector 513 may be arranged to generate a prioritized list of sets of reverberation parameters to allocate to reverberators. Such a prioritized list may include:

- Reverb of the current acoustic environment (in-room reverb).
- Reverb for all other acoustic environment.
    - ◦ Prioritized in accordance with a rank value that may reflect combined attenuation and/or resulting signal levels.

[0176] Prioritization of the reverbs/ sets of reverberation parameters may be based on a single strength metric that is based on the reverberation energy at the listener position/ in the listening acoustic environment. Some modifications may be made, for example applying a bonus weight factor (e.g. 1.5) for all in-room reverbs.

[0177] A reverberator's rank value or 'strength weight' can be chosen to be the resulting reverberation energy/ level perceived in the listening acoustic environment. This may specifically be the amount of reverberator energy received at a listening position (or acoustic environment with the listening position). This perceived/received reverberator energy can be derived for each set of reverberation parameters from metadata as indicated in FIG. 9.

[0178] In a first step, a (normalized) emitted source energy can be calculated for every source contributing to a source acoustic environment reverberation.

[0179] All source energies contributing to a given source acoustic environment may be combined at this point (sum, average, max) for a total source acoustic environment energy. The DSR may then translate this to a corresponding reverb energy/ level in the source acoustic environment.

[0180] In case of multiple parameter sets per room, an interpolation coefficient may isolate the contribution of a specific set of reverberation parameters and this may be applied at this stage. For example, each parameter set of the room may be considered as an individual reverberator similar to other parameter sets of acoustic environments represented by a single set or reverberation parameters. If the source acoustic environment has only one set of reverberation parameters, this step may be skipped, or the interpolation coefficient may be assumed to be 1.

[0181] The energy transfer parameter representing the acoustic transfer/ propagation from a source acoustic environment to the listening acoustic environment may then be applied to determine the perceived reverberator energy in the listening acoustic environment. This step may include combining portal transfer parameters from all portals of the source acoustic environment to all portals of the listening acoustic environment. For example, in the example of FIG. 8, for listening room E and considering perceived reverberator energy from room D, the energy transfer parameter for portal 7 and the energy transfer parameter from portal 3 to 4 may be combined into a combined energy transfer parameter (aka room to room transfer parameter).

[0182] In the previous example, the source energies were considered to be nominal and identical. Thus, effectively the selection is based directly on the combined attenuation. In some embodiments the signal level of the sources may however be available. This may provide a much better estimate of the perceived reverberation energy. The signal level may be a signal energy or signal loudness that is calculated directly from the signals or received as metadata.

[0183] In such a case the approach of FIG. 9 may be modified as shown in FIG. 10. In this case, the actual signal source energies/ levels may be used rather than assuming nominal signal energies/ levels.

[0184] The signal level could be indicated as a loudness (e.g. according to ITU-R BS. 1770) within a medium-term window of the signal, but it could also be an RMS level or signal energy. Level indications may be provided per signal or per source, where a source may be represented by multiple signals. For example, for an HOA (Higher Order Ambisonics) source a single value may be calculated for the omnidirectional representation. Based on loudness standards, e.g., a range of -62 - 0 dBFS may be appropriate with an additional value for indicating $-\infty$ dBFS. As a result, a single value of about 6 bits per signal/source would be needed per update. The update may e.g. be made once per second. With these assumptions, the required bitrate would be 6 bits per second per source. Or, for an example scene with 30 signals/sources this would be 180 bps. Even for scenes with more signals/sources, the bitrate would not exceed 1 kbps until reaching 171 loudness values. Further optimizations such as dedicated Huffman coding, non-homogeneous quantization and differential coding can make it even more efficient to transmit.

[0185] The values may represent perceptual loudness according to a model, which is mostly a single value representing the full bandwidth. However, in such a case, the individual loudness indications may possibly not be very suitable for simply being summed to get an accurate combined loudness, and more conservative audibility thresholds may be needed. In other, more accurate, embodiments it may be desired to transfer signal energy for different frequencies instead. This may be done at a different frequency resolution. A trade-off can be made between bitrate and performance. For example, choosing octave bands would require at most 11 values per signal/source. This results in a maximum bitrate of 2 kbps for the example of 30 values, without any entropy coding.

[0186] If no audio source signal level is received or calculated, the selection may be based on normalized source energies or equivalently on the combined attenuation. This actual reverberation energy in the listening acoustic environment will depend on the unknown signal level, so in some embodiments this selection may only be performed if there is a very substantial combined attenuation.

[0187] In many embodiments, one or more of the reverberation parameters and/or the energy transfer parameter may be frequency dependent. For example, a plurality of different values may be provided for a given parameter with each value being associated with a given frequency interval. In such cases, the previously described approach may be performed for each frequency interval to determine frequency dependent signal levels and/or combined attenuations. A single value may then be determined by combining the values of the individual frequency ranges. For example, a weighted average/ summation of the values of the different frequency intervals may be performed. The averaging/ summation may be weighted by coefficients that reflect the size of the different frequency intervals.

[0188] In some embodiments, the combination may be a perceptually weighted combination. For example, the coefficients for a weighted summation may be set to reflect how perceptually significant the individual frequency interval is. For example, frequency intervals in the 500Hz - 1000Hz range may be weighted higher than similar size frequency intervals in the 2kHz-3kHz range.

[0189] Thus, in some cases, some or all of the parameters used to derive a perceived reverberator energy in the listening acoustic environment (as e.g. indicated in FIG. 9 or 10) may be frequency dependent. In such cases it may be beneficial to calculate the perceived energy as frequency dependent. Further, the different frequency values may be summed to obtain a single metric. While doing so, a perceptual weighting may be applied to the different frequencies, e.g. A-weighting, to get a more perceptually relevant metric.

[0190] The selector 513 may in many embodiments be arranged to repeatedly perform a selection of the sets of reverberation parameters for which to render reverberation signal components. For example, a (re)selection may be performed with an interval not exceeding 1,2,5,10, 20 seconds in different embodiments.

[0191] Such an approach may provide advantageous operation in many scenarios where the rendered audio is dynamically adapted and modified to provide a perceptually more accurate and realistic sounding audio scene.

[0192] In order to mitigate and reduce the perceptual impact of the continued reselection and the change of sets of reverberation parameters that are rendered, the audio apparatus may include functionality to perform a gradual transition.

[0193] For example, after the a given set of reverberation parameters has been deselected, the audio apparatus may not immediately stop generating the corresponding reverberation signal but may proceed to generate the reverberation audio signal for that set of reverberation parameters for a given time period after deselection. Further, during the time period, the audio apparatus may proceed to generate a reverberation signal for a newly selected set of reverberation parameters and a gradual transition may be achieved by the former signal being increasingly attenuated and the latter signal being decreasingly attenuated during the time period. Thus, both the old and the new reverberation signal may be generated during a transitional time period in which a gradual crossfading is further performed.

[0194] In some embodiments, two of the reverberators may have different complexity and resource usage (or equivalently the complexity and resource usage of the individual reverberator may be changed, e.g. by each reverberator being arranged to operate in, and switch between, different modes of operation).

[0195] In such embodiments, the generation of the set of reverberation parameters that are deselected may be continued to be rendered for a period of time but using a reverberator having a lower complexity and/or resource usage. For example, during a transitional time period in which a given set of reverberation parameters is faded out after deselection, the rendering of the reverberation signal may not only be of reducing level but may also be a lower resource usage rendering. The quality of the generated reverberation signal may be reduced due to the reduced resource usage. This may in some cases allow other functionality to be performed.

[0196] For example, in some embodiments, the generation of the reverberation signal for the newly selected set of reverberation parameters may also be performed using a lower resource usage reverberator. For example, during the transitional period, a single reverberator may be replaced by two lower complexity reverberators that during the transitional period generate both the previously selected reverberation signal and the newly selected reverberation signal. After a crossfade, the reverberators may be reconfigured such that a more complex and resource demanding generation of the reverberation signal for the newly selected set of reverberation parameters is performed while no resource is allocated to the deselected set of reverberation parameters. Thus, during normal operation, high quality generation of the reverberation signals is performed but during a transitional period, two lower quality reverberation signals may be generated and a crossfading between these may be performed.

[0197] In some embodiments, the selector 513 may be arranged to include a hysteresis effect in the selection. This may for example be introduced to a decision threshold for the strength metric/energy.

[0198] In the previous examples, the energy transfer parameter has predominantly been considered to directly reflect the propagation attenuation between acoustic environments, and specifically to be indicative of how much audio energy of a source acoustic environment propagates to a destination acoustic environment. In the example, the energy transfer parameter indicates the loss involved in propagation but there is no consideration of how the propagation occurs (or through which media).

[0199] However, in many embodiments, sound propagation between environments is predominantly via transfer regions in the acoustically attenuating boundaries between the acoustic environments. For example, for rooms in a building, sound propagation is predominantly via openings (e.g. for doors or windows) in the walls.

[0200] Accordingly, in some embodiments, the metadata may include data related to transfer regions of the acoustically attenuating boundaries and some or all energy transfer parameters may be indicative of a propagation loss/ attenuation of propagation between a transfer region of the source acoustic environment and a transfer region of the destination acoustic environment.

[0201] The metadata may in such embodiments 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.

[0202] 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.

[0203] 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).

[0204] 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.

[0205] 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.

[0206] 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.

[0207] 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.

[0208] 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 transitioning 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).

[0209] The energy transfer parameters may thus (in the form of transfer regions) in some embodiments provide acoustic linking metadata that describes how two rooms are acoustically linked via transfer regions. 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.

[0210] FIG. 8 shows an example of a scene to which the described approach may be applied. FIG. 8 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). The audio apparatus of FIG. 5 may in this case determine a listening position in room E and render audio for this listening position. The rendered audio signal includes different audio components in other rooms, and specifically reverberant audio from other rooms. The sound from such sources may specifically reach room E through a number of transfer regions, e.g. corresponding to (open) doors or windows in the walls forming the rooms.

[0211] 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 reverberation sound component from another acoustic environment may be considered to reach a given first room E comprising the listener 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. In the specific example, the sound component from other rooms is a reverberant sound with the corresponding signal components being generated by one of the bank of reverberators 509, 511.

[0212] Such an approach may for example be used to render the reverberation 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.

[0213] 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).

[0214] 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).

[0215] 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 sets of reverberation 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.

[0216] In some embodiments, the energy transfer parameters may be indicative of at least one energy attenuation between a pair of transfer regions with one being a transfer region of a source acoustic environment and the other being a transfer region of a destination acoustic environment, 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).

[0217] The energy transfer parameters may provide energy attenuation indications, and these may be specifically 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 (intermediate) acoustic environment (e.g. a energy transfer parameter may be provided for portal 1 and 4 of FIG. 8. The energy transfer parameter indicates propagation through the portals from room A t room E via portals to a shared intermediate room C). 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.

[0218] 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/ acoustically attenuating boundaries that do not share any acoustic environment.

[0219] Indeed, energy transfer parameters providing energy attenuation indications may be provided for any pair of transfer regions or acoustic environments 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 and/or between any two rooms/ acoustic environments in the scene.

[0220] 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.

[0221] 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 (or more generally acoustic environment to acoustic environment 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.

[0222] 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*log10(energy attenuation [i, j]) < -60 dB.

[0223] Each energy transfer parameter may provide an energy attenuation indication for two transfer regions/ portals/ acoustic environments, and the metadata may specifically include the energy attenuation indication and the identification of the associated transfer regions/ acoustic environments. 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/ acoustic environments 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.

[0224] 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.

[0225] 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.

[0226] 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.

[0227] 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.

[0228] 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.

[0229] 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.

[0230] 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 (e.g. for all pairs of portals of room C). These transfers may then be combined to represent higher order room to room transfers (i.e. including more than one shared/ intermediate room).

[0231] 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.

[0232] 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, it may result from the described reverberation rendering for the source room.

[0233] 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). Signal S_{src} may typically correspond to the reverberation component generated by one of the reverberators.

[0234] 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.

[0235] 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.

[0236] 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.

[0237] 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 may simply include extracting and using the appropriate energy attenuation indication for that transfer region.

[0238] 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.

[0239] 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, , 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.

[0240] 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.

[0241] FIG. 11 is a block diagram illustrating an example processor 1100 according to embodiments of the disclosure. Processor 1100 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 1100 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.

[0242] The processor 1100 may include one or more cores 1102. The core 1102 may include one or more Arithmetic Logic Units (ALU) 1104. In some embodiments, the core 1102 may include a Floating Point Logic Unit (FPLU) 1106 and/or a Digital Signal Processing Unit (DSPU) 1108 in addition to or instead of the ALU 1104.

[0243] The processor 1100 may include one or more registers 1112 communicatively coupled to the core 1102. The registers 1112 may be implemented using dedicated logic gate circuits (e.g., flip-flops) and/or any memory technology. In some embodiments the registers 1112 may be implemented using static memory. The register may provide data, instructions and addresses to the core 1102.

[0244] In some embodiments, processor 1100 may include one or more levels of cache memory 1110 communicatively coupled to the core 1102. The cache memory 1110 may provide computer-readable instructions to the core 1102 for execution. The cache memory 1110 may provide data for processing by the core 1102. In some embodiments, the computer-readable instructions may have been provided to the cache memory 1110 by a local memory, for example, local memory attached to the external bus 1116. The cache memory 1110 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.

[0245] The processor 1100 may include a controller 1114, which may control input to the processor 1100 from other processors and/or components included in a system and/or outputs from the processor 1100 to other processors and/or components included in the system. Controller 1114 may control the data paths in the ALU 1104, FPLU 1106 and/or DSPU 1108. Controller 1114 may be implemented as one or more state machines, data paths and/or dedicated control logic. The gates of controller 1114 may be implemented as standalone gates, FPGA, ASIC or any other suitable technology.

[0246] The registers 1112 and the cache 1110 may communicate with controller 1114 and core 1102 via internal connections 1120A, 1120B, 1120C and 1120D. Internal connections may be implemented as a bus, multiplexer, crossbar switch, and/or any other suitable connection technology.

[0247] Inputs and outputs for the processor 1100 may be provided via a bus 1116, which may include one or more conductive lines. The bus 1116 may be communicatively coupled to one or more components of processor 1100, for example the controller 1114, cache 1110, and/or register 1112. The bus 1116 may be coupled to one or more components of the system.

[0248] The bus 1116 may be coupled to one or more external memories. The external memories may include Read Only Memory (ROM) 1132. ROM 1132 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) 1133. RAM 1133 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) 1135. The external memory may include Flash memory 1134. The External memory may include a magnetic storage device such as disc 1136. In some embodiments, the external memories may be included in a system.

[0249] 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.

[0250] 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.

[0251] 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.

[0252] 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.

[0253] 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 comprising:

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 (503) arranged to receive metadata for the audio data, the metadata comprising:

a plurality of sets of reverberation parameters, each set of reverberation parameters being for one associated acoustic environment and comprising at least one reverberation parameter indicative of a relationship between a level of reverberation in the one associated acoustic environment and a level of an audio source in the one associated acoustic environment; and
at least one energy transfer parameter, each energy transfer parameter being indicative of an energy attenuation for audio propagation between a source acoustic environment and a destination acoustic environment;
a selector (513) arranged to select for a first acoustic environment at least a first set of reverberation parameters of the multiple acoustic environments, the first set of reverberation parameters being for a different acoustic environment than the first acoustic environment and the selection being in response to parameter values of the first set of reverberation parameters and an energy transfer parameter indicative of energy attenuation for audio propagation between the different acoustic environment and the first acoustic environment;
a plurality of reverberators (509, 511) arranged to generate reverberation audio signals, at least one reverberator being arranged to generate a reverberation audio signal based on the first set of reverberation parameters and audio data for an audio source of the different acoustic environment of the first set of reverberation parameters; and
a renderer (507) for generating an audio signal for the first acoustic environment based on the audio data, the audio signal including audio components derived from the reverberation audio signals.

2. The audio apparatus of claim 1 wherein the selector (513) is arranged to select between sets of reverberation parameters based on a combined attenuation for each set of reverberation parameters, the combined attenuation for a given set of reverberation parameters being indicative of a level attenuation to audio in the first acoustic environment resulting from reverberation in a given acoustic environment of the given set of reverberation parameters resulting from reverberation in the given acoustic environment caused by an audio source in the given acoustic environment, the combined attenuation for the given set of reverberation parameters including a contribution from an attenuation indicated by the at least one reverberation parameter of the given set of reverberation parameters and an energy transfer parameter indicative of energy attenuation for audio propagation between the given acoustic environment and the first acoustic environment.

3. The audio apparatus of claim 2 wherein the selector (513) is arranged to select a first set of reverberation parameters over a second set of reverberation parameters if the combined attenuation for the second set of reverberation parameters exceeds the combined attenuation for the first set of reverberation parameters.

4. The audio apparatus of claim 2 or 3 wherein the selector (513) is arranged to discard a set of reverberation parameters from selection if the combined attenuation exceeds a threshold.

5. The audio apparatus of any of claims 2 to 4 wherein the metadata further comprises audio level data indicative of signal levels for audio sources, and the selector (513) is arranged to determine resulting signal levels in the first acoustic environment for audio sources in acoustic environments of the sets of reverberation parameters based on signal levels of the audio sources and the combined attenuation for the sets of reverberation parameters ; and to select the at least one set of reverberation parameters in response to the resulting signal levels in the first acoustic environment.

6. The audio apparatus of claim 5 wherein the selector (513) is arranged to combine levels of multiple sources in at least one acoustic environment into a combined audio source level and to determine the resulting signal level for a set of reverberation parameters of the at least one acoustic environment by applying the combined attenuation for the at least one acoustic environment to the combined audio source level.

7. The audio apparatus of any of the previous claims wherein the selector (513) is arranged to determine a number of the plurality of reverberators that are available for generating reverberation signals for audio sources in acoustic environments other than the first acoustic environment; and to adapt a number of sets of reverberation parameters being selected to match the number of the plurality of reverberators available.

8. The audio apparatus of claim 7 wherein the selector (513) is arranged to determine a first number of reverberation signal sources in the first acoustic environment and to subtract the first number from a number of reverberators of the audio apparatus to determine the number of the plurality of reverberators available.

9. The audio apparatus of any of the previous claims wherein at least one of the at least one reverberation parameter and the energy transfer parameter is frequency dependent.

10. The audio apparatus of any of the previous claims wherein the selector (513) is arranged to repeatedly perform selection of the at least one set of reverberation parameters.

11. The audio apparatus of claim 10 wherein at least a first reverberator (509) of the plurality of reverberators (509, 511) is arranged to continue to generate a reverberation audio signal for a first set of reverberation parameters for a time period following deselection of the first set of reverberation parameters.

12. The audio apparatus of claim 11 wherein the first reverberator (509) has lower computational resource usage than a reverberator of the plurality of reverberators (509, 511) used to generate reverberation audio signals for the first set of reverberation parameters prior to deselection of the first set of reverberation parameters.

13. The audio apparatus of any previous claim wherein the at least one reverberation parameter comprises a Diffuse to Source Ratio parameter indicative of a ratio between emitted audio source energy and a diffuse reverberation energy.

14. A method of operation for an audio apparatus 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 plurality of sets of reverberation parameters, each set of reverberation parameters being for one associated acoustic environment and comprising at least one reverberation parameter indicative of a relationship between a level of reverberation in the one associated acoustic environment and a level of an audio source in the one associated acoustic environment; and
at least one energy transfer parameter, each energy transfer parameter being indicative of an energy attenuation for audio propagation between a source acoustic environment and a destination acoustic environment;
selecting, for a first acoustic environment, at least a first set of reverberation parameters of the multiple acoustic environments, the first sets of reverberation parameters being for a different acoustic environment than the first acoustic environment and the selection being in response to parameter values of the first set

of reverberation parameters and an energy transfer parameter indicative of energy attenuation for audio propagation between the different acoustic environment and the first acoustic environment;

a plurality of reverberators (509, 511) generating reverberation audio signals, at least one reverberator being arranged to generate a reverberation audio signal based on the first set of reverberation parameters and audio data for an audio source of the different acoustic environment of the first set of reverberation parameters; and

generating an audio signal for the first acoustic environment based on the audio data, the audio signal including audio components derived from the reverberation audio signals.

15. A computer program product comprising computer program code means adapted to perform all the steps of claim 14 when said program is run on a computer.

16. 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; and

metadata for the audio data, the metadata comprising:

a plurality of sets of reverberation parameters, each set of reverberation parameters being for one associated acoustic environment and comprising at least one reverberation parameter indicative of a relationship between a level of reverberation in the one associated acoustic environment and a level of an audio source in the one associated acoustic environment; and

at least one energy transfer parameter, each energy transfer parameter being indicative of an energy attenuation for audio propagation between a source acoustic environment and a destination acoustic environment.

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

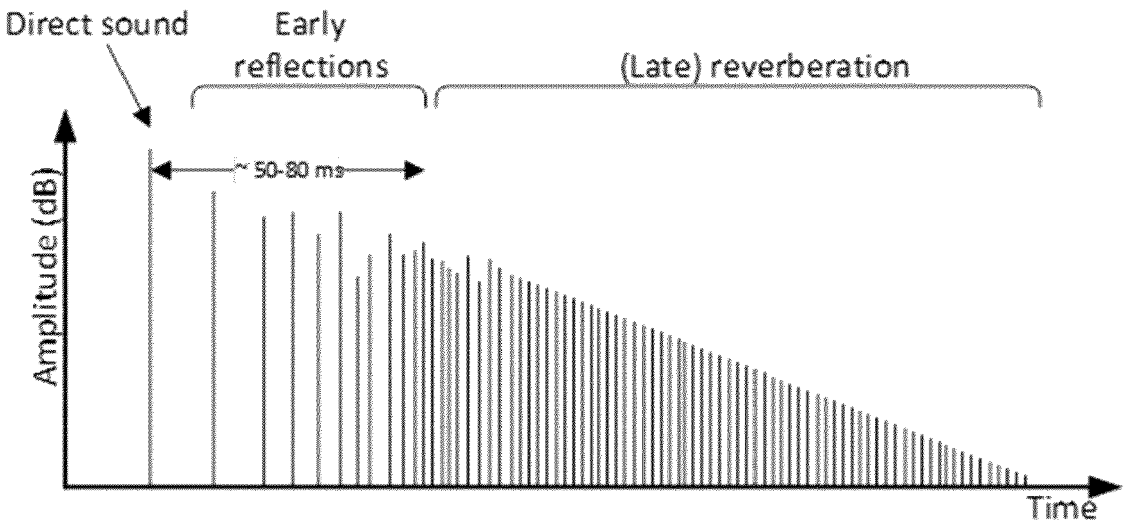
Direct sound
Early reflections
(Late) reverberation
~ 50-80 ms
Amplitude (dB)
Time

FIG. 1

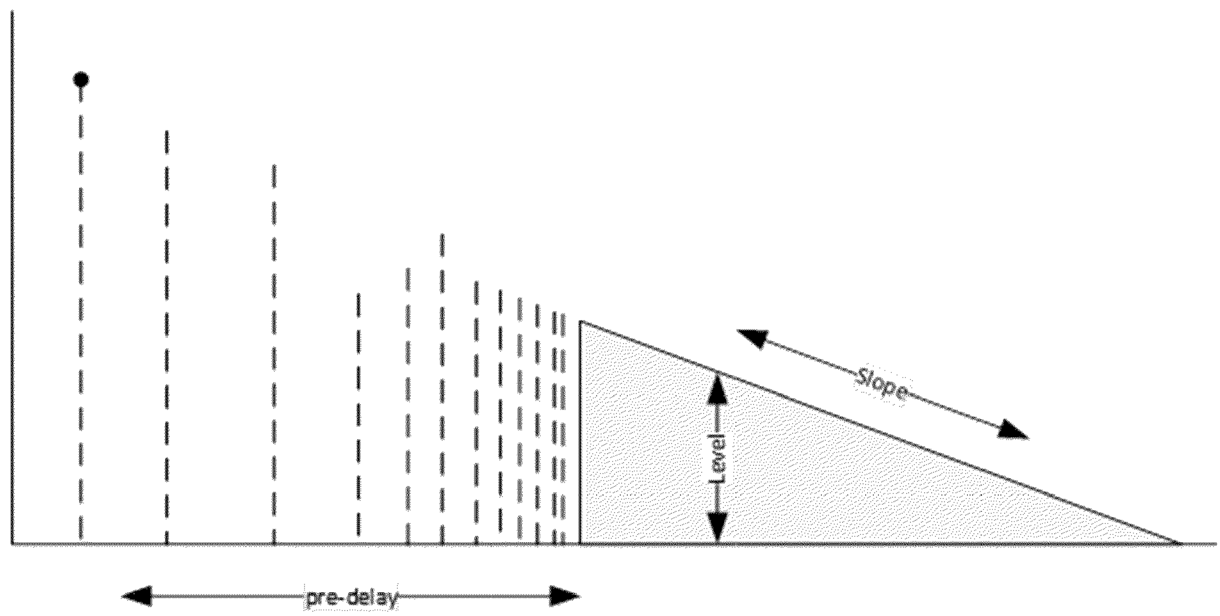

FIG. 2

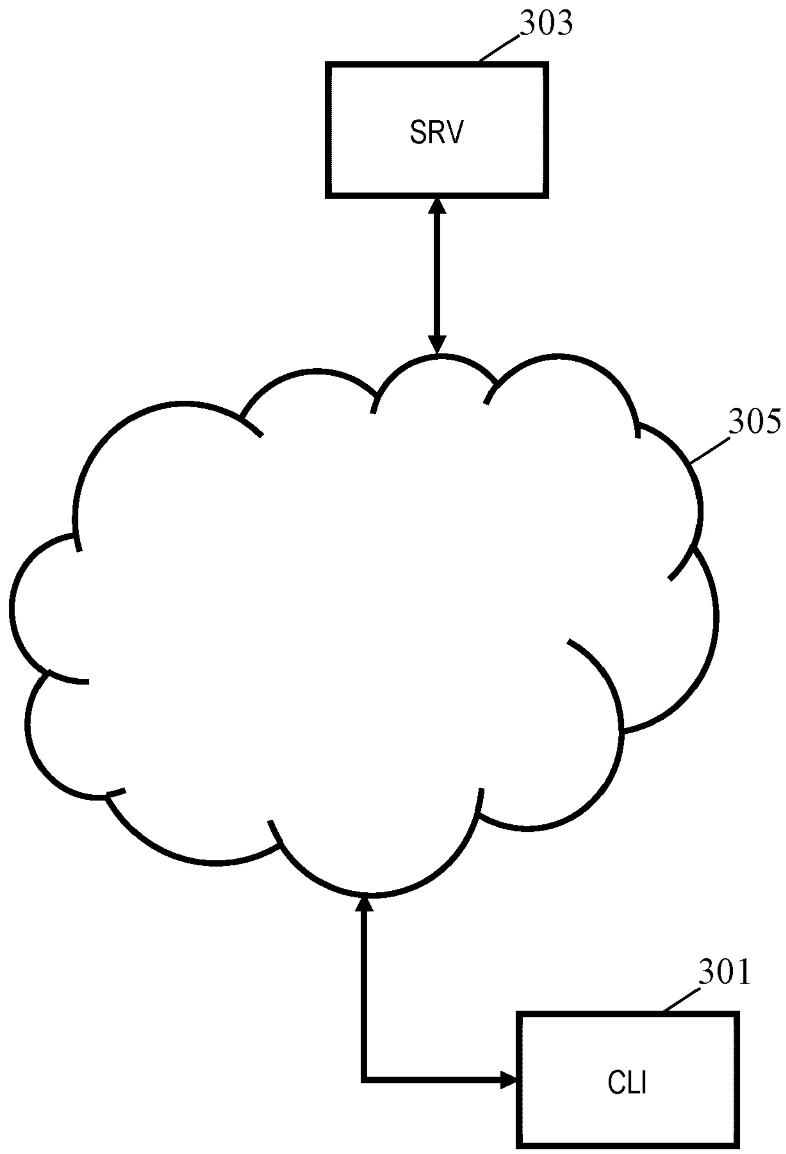

FIG. 3

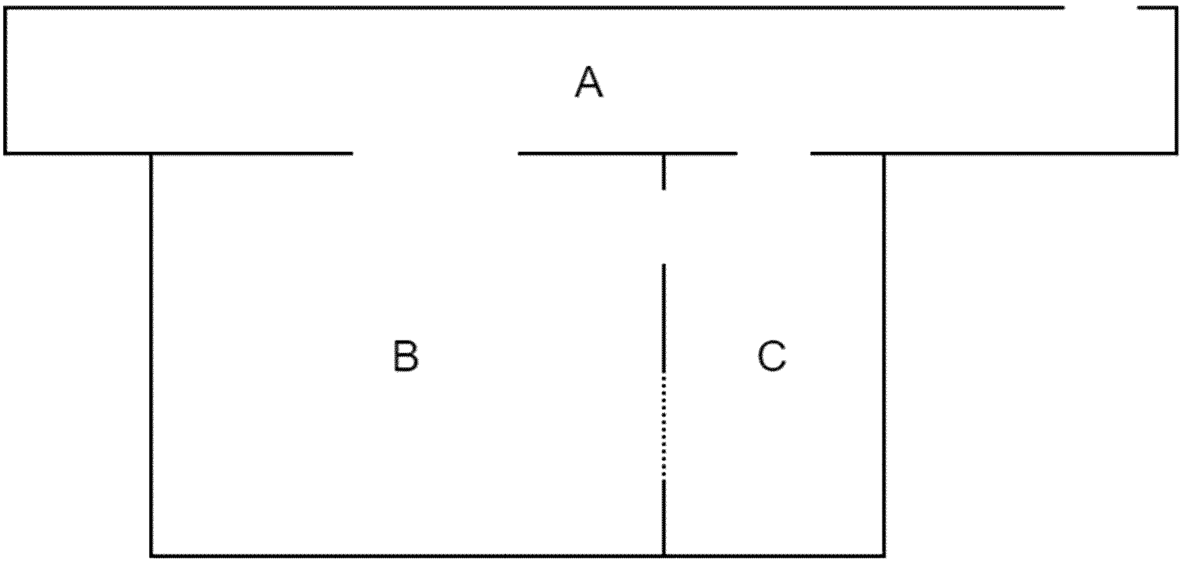

FIG. 4

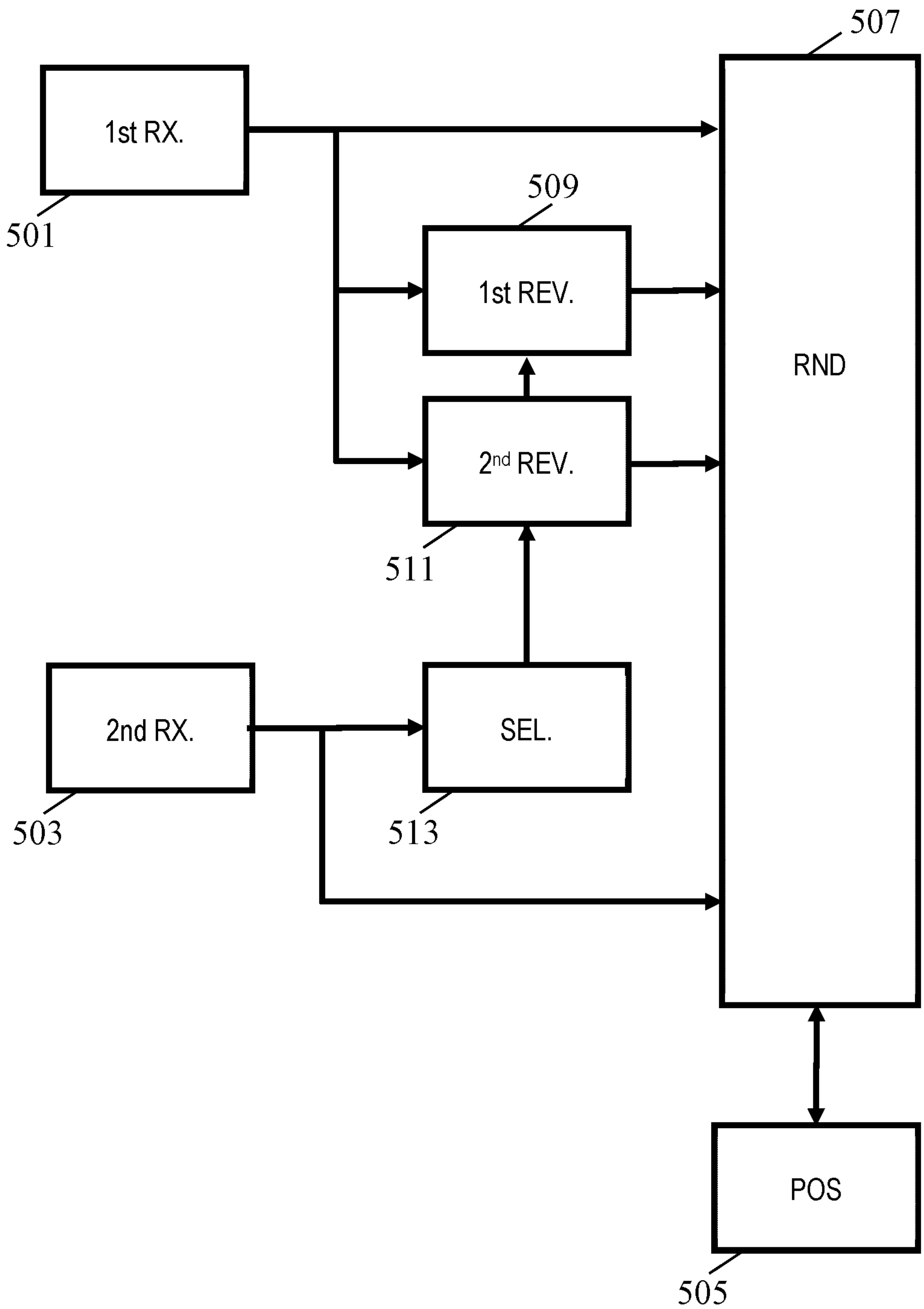
507
1st RX.
501
509
1st REV.
RND
2nd REV.
511
2nd RX.
SEL.
503
513
POS
505

FIG. 5

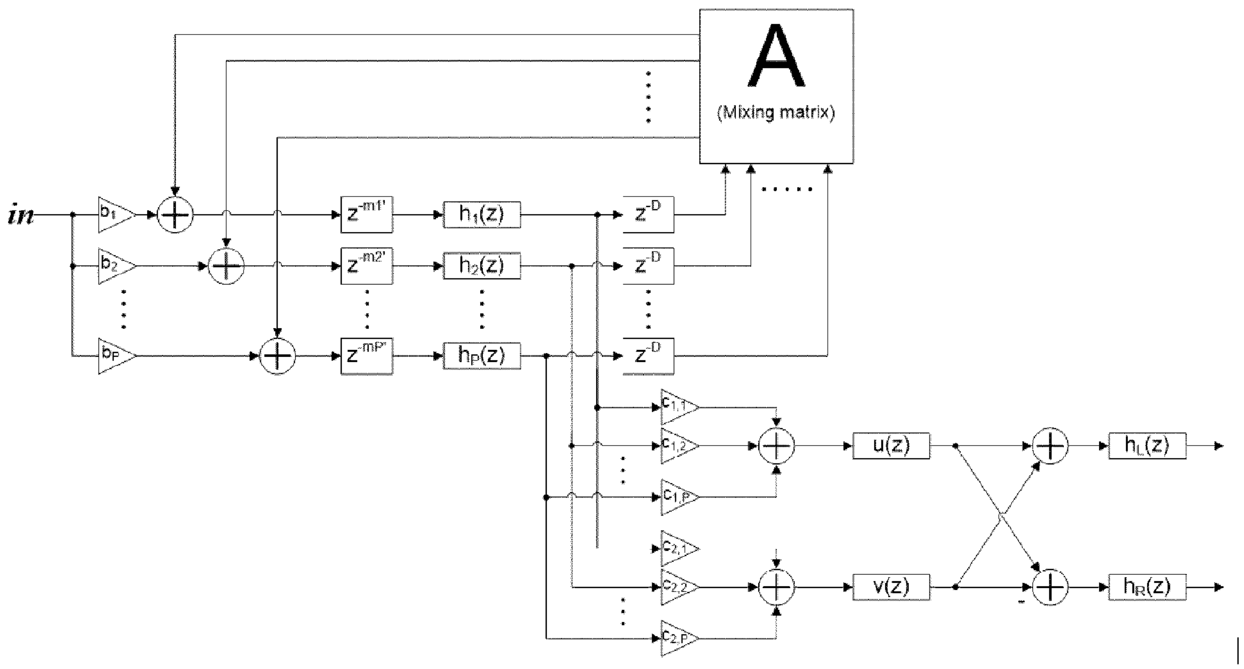
A
(Mixing matrix)
in
b1
b2
bP
z-m1'
z-m2'
z-mP'
h1(z)
h2(z)
hP(z)
z-D
z-D
z-D
c1,1
c1,2
c1,P
c2,1
c2,2
c2,P
u(z)
v(z)
hL(z)
hR(z)
-

FIG. 6

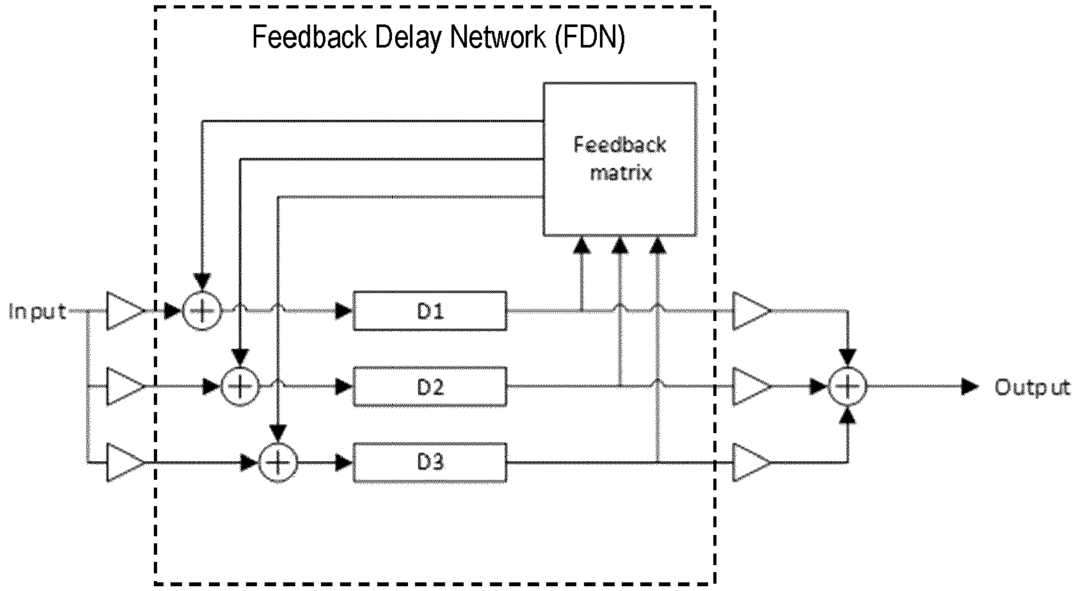
Feedback Delay Network (FDN)
Feedback matrix
Input
D1
D2
D3
Output

FIG. 7

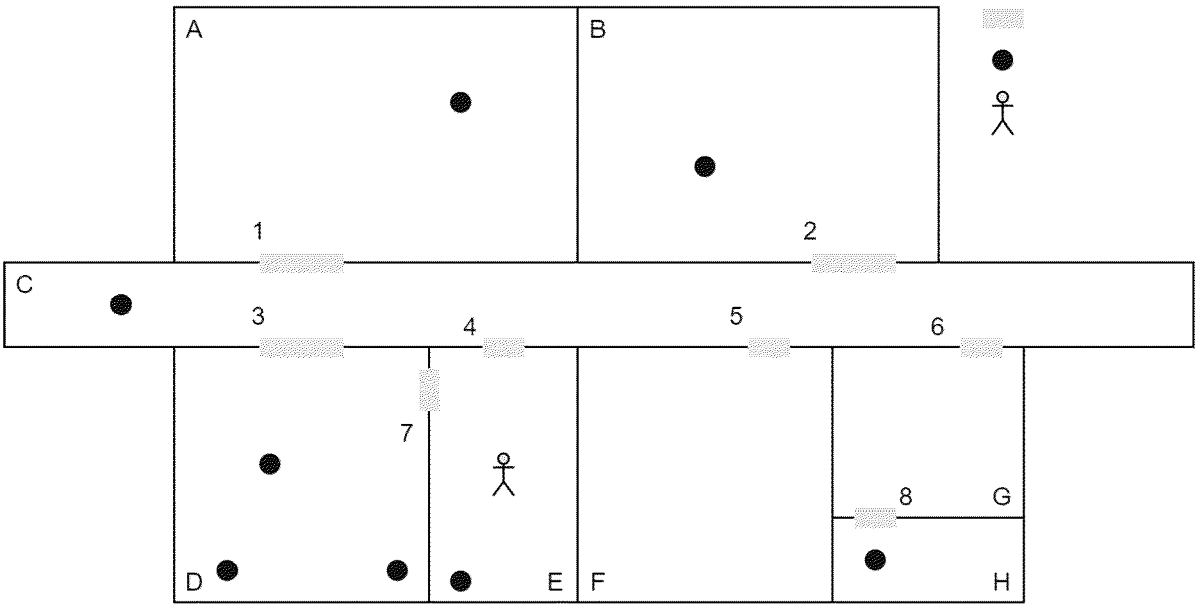

FIG. 8

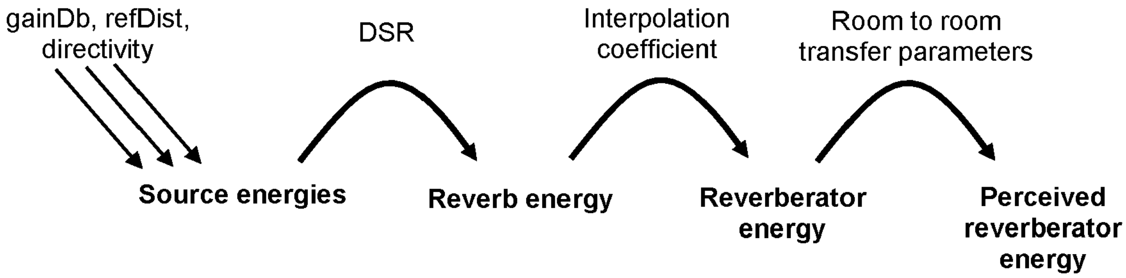
gainDb, refDist, directivity
DSR
Interpolation coefficient
Room to room transfer parameters
Source energies
Reverb energy
Reverberator energy
Perceived reverberator energy

FIG. 9

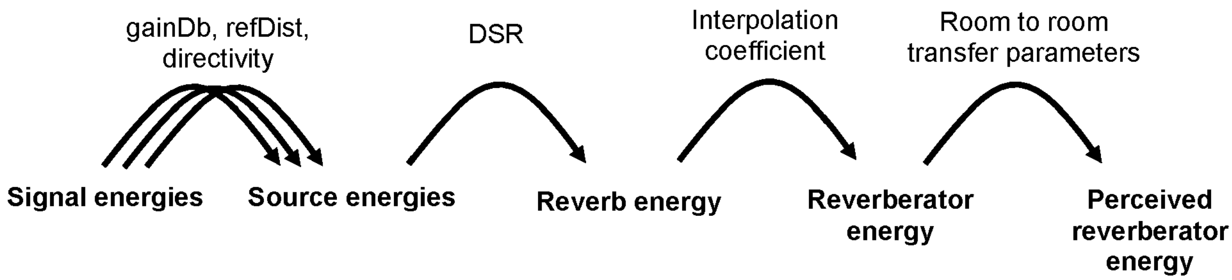
gainDb, refDist, directivity
DSR
Interpolation coefficient
Room to room transfer parameters
Signal energies
Source energies
Reverb energy
Reverberator energy
Perceived reverberator energy

FIG. 10

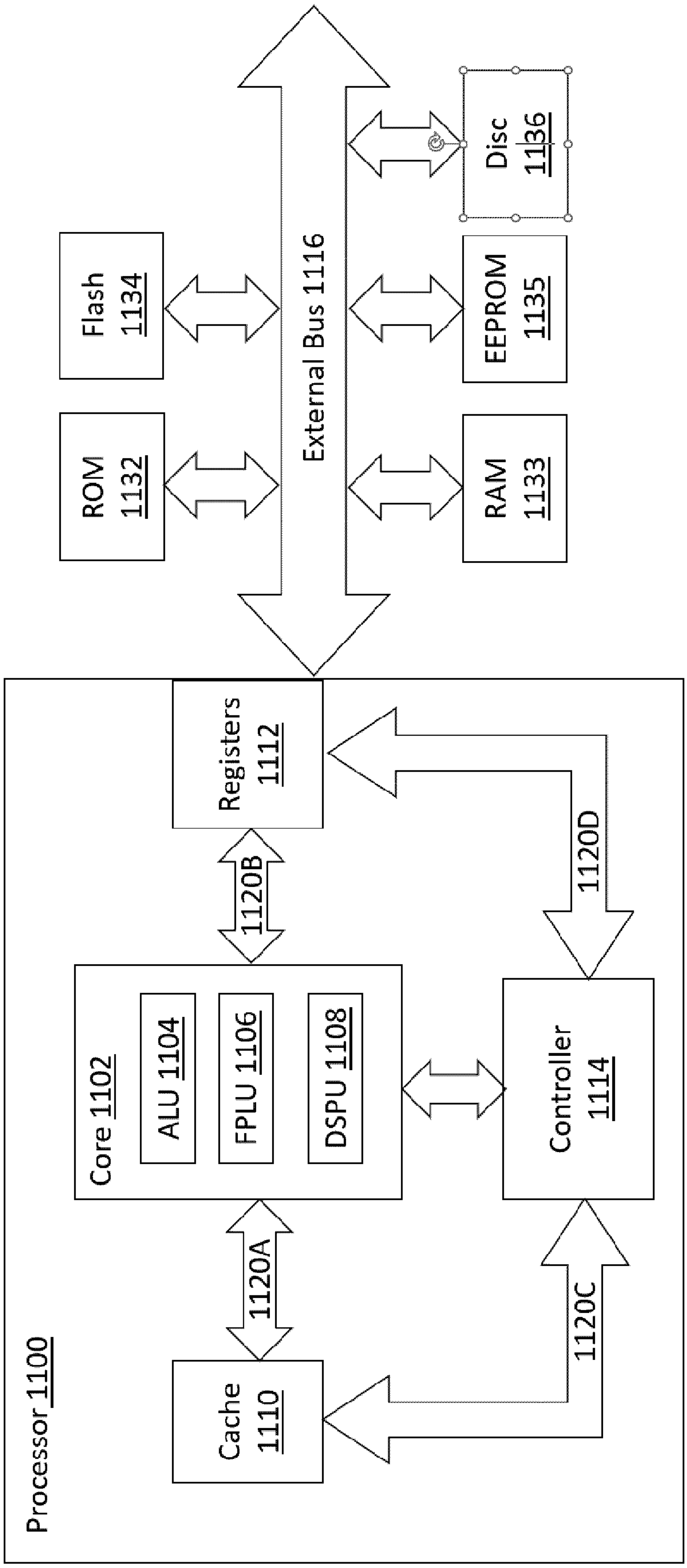
External Bus 1116
ROM 1132
Flash 1134
RAM 1133
EEPROM 1135
Disc 1136
Processor 1100
Registers 1112
1120B
1120D
Core 1102
ALU 1104
FPLU 1106
DSPU 1108
Controller 1114
1120A
1120C
Cache 1110

FIG. 11

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 23 15 0704

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	EP 3 617 871 A1 (KONINKLIJKE PHILIPS NV [NL]) 4 March 2020 (2020-03-04) * paragraph [0018] - paragraph [0050]; claims 1-15 * -----	1-17	INV. H04S7/00
A	EP 4 072 163 A1 (KONINKLIJKE PHILIPS NV [NL]) 12 October 2022 (2022-10-12) * paragraph [0027] - paragraph [0262]; claims 1-18 * -----	1-17	
A	EP 3 930 349 A1 (KONINKLIJKE PHILIPS NV [NL]) 29 December 2021 (2021-12-29) * paragraph [0030] - paragraph [0204]; claims 1-16 * -----	1-17	

TECHNICAL FIELDS SEARCHED (IPC)

G06F
H04R
H04S
G10K
G10L

The present search report has been drawn up for all claims

1

Place of search	Date of completion of the search	Examiner
The Hague	12 June 2023	Timms, Olegs

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

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

EP 23 15 0704

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

12-06-2023

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 3617871 A1	04-03-2020	BR 112021003449 A2	18-05-2021
		CN 112602053 A	02-04-2021
		EP 3617871 A1	04-03-2020
		EP 3844606 A1	07-07-2021
		JP 2022502886 A	11-01-2022
		US 2021258690 A1	19-08-2021
		US 2023179919 A1	08-06-2023
		WO 2020043539 A1	05-03-2020
EP 4072163 A1	12-10-2022	EP 4072163 A1	12-10-2022
		WO 2022214270 A1	13-10-2022
EP 3930349 A1	29-12-2021	BR 112022026158 A2	17-01-2023
		CA 3187637 A1	30-12-2021
		CN 115769603 A	07-03-2023
		EP 3930349 A1	29-12-2021
		EP 4169267 A1	26-04-2023
		KR 20230027273 A	27-02-2023
		WO 2021259829 A1	30-12-2021

EPO FORM P0459

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