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

(11) **EP 4 539 509 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
16.04.2025 Bulletin 2025/16

(21) Application number: **23203277.1**

(22) Date of filing: **12.10.2023**

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

(52) Cooperative Patent Classification (CPC):
H04S 7/307; H04S 7/305; G10L 19/008;
H04S 7/304

(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:
- **SZCZERBA, Marek Zbigniew**
  Eindhoven (NL)
- **OOMEN, Arnoldus Werner Johannes**
  Eindhoven (NL)
- **SCHUIJERS, Erik Gosuinus Petrus**
  Eindhoven (NL)
- **DILLEN, Paulus Henricus Antonius**
  Eindhoven (NL)

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

(54) **GENERATING AN AUDIO DATA SIGNAL**

(57) An apparatus comprises a receiver (201) receiving a data signal comprising audio data for at least a first audio signal and first and second acoustic environment data for an acoustic environment, where a data size of sets of acoustic environment parameters is larger for the first acoustic environment data and an update rate is higher for the second acoustic environment data. An acoustic data generator (203) selects between the first and second acoustic environment data and to generate rendering acoustic environment data. For example, the second acoustic environment data may be selected only if corresponding first acoustic environment data is not received. A renderer (205) generates an audio output signal by rendering the audio signal based on the rendering acoustic environment data. A reduced delay in rendering the acoustic environment may be achieved without sacrificing long term accuracy.

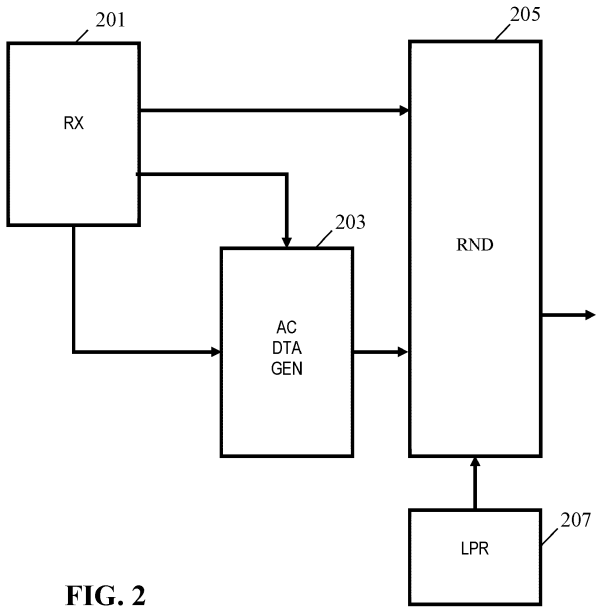


FIG. 2

EP 4 539 509 A1

Description

FIELD OF THE INVENTION

[0001] The invention relates to generating an audio signal and/or an audio data signal, and in particular, but not exclusively, to generating such signals to support e.g., an eXtended Reality application.

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 Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) applications (commonly referred to as eXtended Reality XR 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 bodies are actively developing standards for the various aspects of VR/AR/MR/XR 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 and are commonly referred to as XR applications. In the following, the terms will be used interchangeably.

[0005] VR applications typically provide a virtual reality experience to a user allowing the user to (relatively) freely move about in a virtual environment 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.

[0006] In addition to the visual rendering, most XR (and in particular VR) 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 (including both objects that are currently visible and objects that are not currently visible (e.g., behind the user)). Thus, the audio and video scenes are preferably perceived to be consistent and with both providing a full spatial experience.

[0007] For audio, headphone reproduction using binaural audio rendering technology is widely used. In many scenarios, headphone reproduction enables a highly immersive, personalized experience to the user. Using headtracking, the rendering can be made responsive to the user's head movements, which highly increases the sense of immersion.

[0008] Often, audio data is provided together with metadata describing an acoustic environment, such as the acoustic properties of a room etc. This allows the rendering of the audio to be adapted to provide a perception of a more realistic environment.

[0009] However, whereas such approaches may provide suitable user experiences in many practical applications they tend to not provide optimal user experiences in all scenarios. In particular, in many situations, a suboptimum audio quality/perception/user experience may result. For example, the representation of the acoustic environment may not be perceived to be accurate or realistic, or there may be an undesirable delay before such perception can be achieved.

[0010] Hence, an improved approach for distribution and/or rendering and/or processing of audio signals/environment data, in particular for a Virtual/ Augmented/ Mixed/ eXtended Reality experience/ application, would be advantageous. In particular, an approach that allows improved operation, increased flexibility, reduced complexity, facilitated implementation, an improved user experience, improved audio quality, improved adaptation to different acoustic environments, improved trade-off between data size/rate and quality of description of an acoustic environment, facilitated and/or improved and/or faster adaptation to acoustics environment properties; an improved eXtended Reality (XR) experience, and/or improved performance and/or operation would be advantageous.

SUMMARY OF THE INVENTION

[0011] Accordingly, the invention seeks to preferably mitigate, alleviate or eliminate one or more of the above-mentioned disadvantages singly or in any combination.

[0012] According to an aspect of the invention there is provided an apparatus for generating an output audio signal, the

apparatus comprising: a receiver arranged to receive a data signal comprising audio data for at least a first audio signal and metadata including: first acoustic environment data for an acoustic environment, the first acoustic environment data comprising repeated first sets of acoustic environment parameters, each first set of acoustic environment parameters providing a description of the acoustic environment; second acoustic environment data for the acoustic environment, the second acoustic environment data comprising repeated second sets of acoustic environment parameters, each second set of acoustic environment parameters providing a description of the acoustic environment, a data size of the first sets of acoustic environment parameters exceeding a data size for the second sets of acoustic environment parameters and an update rate for the second sets of acoustic environment parameters higher than an update rate for the first sets of acoustic environment parameters; an acoustic data generator being arranged to select between the first acoustic environment data and the second acoustic environment data to generate rendering acoustic environment data; and a renderer arranged to generate the audio output signal by rendering the audio signal based on the rendering acoustic environment data.

[0013] The approach may allow an improved output audio signal to be generated. The approach may in many embodiments and scenarios provide an improved audio quality and may in particular provide an audio output signal that provides an improved representation of the acoustic environment. The approach may in many scenarios provide an improved and more immersive user experience, and in particular may in many applications provide an improved XR, and specifically VR experience. For example, it may allow a user experience which may provide a more consistent user perception of an environment based on generation of both audio and video.

[0014] The approach may facilitate and/or improve distribution of audio and may e.g., allow an improved ratio between data size/rate and access delay to and/or accuracy and detail of a characterization of the acoustic environment. It may in many embodiments and scenarios allow faster initialization of rendering to reflect a particular acoustic environment without sacrificing longer term precision and accuracy of such a representation.

[0015] The data signal may be received as a plurality of data packets. Each data packet may be received individually/ separate from other data packets. A set of the first and/or second sets of acoustic environment parameters may be comprised in a single data packet. In many embodiments, each set of the first sets of acoustic environment parameters may be distributed over a plurality of data packets.

[0016] In some embodiments, the selector may be arranged to generate the rendering acoustic environment data to include data of both the first acoustic environment data and the second acoustic environment data.

[0017] The acoustic data generator may be arranged to generate the rendering acoustic environment data to include, or consist, of a rendering set of acoustic environment parameters, the rendering set of acoustic environment parameters comprising parameters selected from one set of the first sets of acoustic environment parameters and/or from one set of second sets of acoustic environment parameters.

[0018] The second acoustic environment data may provide a coarser/ less detailed and/or less accurate representation of the acoustic environment than the first acoustic environment data.

[0019] The data signal may comprise repeated sets of the first sets of acoustic environment parameters. In some embodiments, some or all of the first sets of acoustic environment parameters may be identical. The update rate may be a rate of the sets of the first sets of acoustic environment parameters in the data signal (e.g., represented by duration between the sets). The update rate may be a repetition rate.

[0020] The data signal may comprise repeated sets of the second sets of acoustic environment parameters. In some embodiments, some or all of the second sets of acoustic environment parameters may be identical. The update rate may be a rate of the (e.g., represented as duration between) sets of the second sets of acoustic environment parameters in the data signal. The update rate may be a repetition rate.

[0021] The update rate may be a rate/duration of data providing a complete representation of a set of the first/second sets of acoustic environment parameters.

[0022] The acoustic environment data may specifically be (or include) reverberation data indicative of reverberation properties of the acoustic environment. The rendering may include reverberation rendering based on the rendering acoustic environment data.

[0023] In many embodiments, the update rate for the first sets of acoustic environment parameters is not less than 30 seconds and the update rate for the second acoustic environment data is no higher than 10 secs. The audio data signal may be a data bitstream (including e.g., a plurality of data packets) at least comprising audio data.

[0024] The first sets of acoustic environment parameters and the second sets of acoustic environment parameters may comprise at least one common acoustic environment parameter/property. At least one acoustic environment parameter/property may be included in both the first sets of acoustic environment parameters and the second sets of acoustic environment parameters. The first sets of acoustic environment parameters and the second sets of acoustic environment parameters may comprise values for at least one common acoustic environment parameter/value.

[0025] The first sets of acoustic environment parameters and the second sets of acoustic environment parameters may both comprise at least one reverberation decay parameter value for the acoustic environment. The first sets of acoustic environment parameters and the second sets of acoustic environment parameters may both comprise at least one reverberation delay parameter value for the acoustic environment. The first sets of acoustic environment parameters and

the second sets of acoustic environment parameters may both comprise at least one reverberation energy rate parameter value for the acoustic environment.

[0026] In accordance with an optional feature of the invention, a quantization of at least one parameter of the second set of acoustic environment parameters is coarser than a corresponding parameter of the first set of acoustic environment parameters.

[0027] This may provide improved and/or facilitated operation or performance in many embodiments.

[0028] The coarser quantization may e.g., be a coarser frequency quantization/resolution, and/or a coarser quantization of one or more parameter values. The first and second sets of acoustic environment parameters may comprise parameters representing the same property with at least one parameter of the second set (a set of the second sets of acoustic environment parameters) being coarser than for the first set (a set of the first sets of acoustic environment parameters). Corresponding (acoustic) parameters may be (acoustic) parameters representing/indicating the same (acoustic) property.

[0029] In accordance with an optional feature of the invention, the audio apparatus further comprises: a listener pose processor arranged to determine a listener pose; and wherein the renderer is arranged to render the audio signal in dependence on the listener pose.

[0030] This may provide improved and/or facilitated operation or performance in many embodiments.

[0031] A pose, which also may be referred to as a placement, may be a position and/or orientation. The listener pose may be the pose for which the (spatial) output audio signal is generated. The output audio signal may be a stereo audio signal.

[0032] In accordance with an optional feature of the invention, data for each set of the first sets of acoustic environment parameters is distributed over a plurality of non-contiguous data segments.

[0033] This may provide improved and/or facilitated operation or performance in many embodiments. It may typically reduce the peak data rate and/or provide a more consistent data flow. It may typically facilitate communication/distribution of audio data for e.g., XR applications.

[0034] In many embodiments, the different data segments may correspond to different data packets.

[0035] The data for each set of the second sets of acoustic environment parameters may be included in a single data segment, which specifically may be a single data packet.

[0036] In accordance with an optional feature of the invention, at least one data segment for a first set of the first sets of acoustic environment parameters comprises an indication of a start position for data for the first set, and the acoustic data generator is arranged to generate the rendering acoustic environment data to represent the first set in dependence on the indication of the start position.

[0037] This may provide improved and/or facilitated operation or performance in many embodiments. It may in many embodiments reduce a delay before the first sets of acoustic environment parameters can be used by the rendering.

[0038] The acoustic data generator may be arranged to parse the first set of acoustic environment data depending on the indication of the start position.

[0039] In accordance with an optional feature of the invention, at least one data segment for a first set of the first sets of acoustic environment parameters comprises an indication of a data size for data for the first set, and the acoustic data generator is arranged to generate the rendering acoustic environment data to represent the first set in dependence on the indication of the data size.

[0040] This may provide improved and/or facilitated operation or performance in many embodiments.

[0041] The acoustic data generator may be arranged to parse the first set of acoustic environment data depending on the data size.

[0042] In accordance with an optional feature of the invention, the receiver is arranged to store data from data segments as these are received, and the acoustic data generator is arranged to generate the rendering acoustic environment data to represent the first set from stored data from the data segments.

[0043] This may provide improved and/or facilitated operation or performance in many embodiments. It may in many embodiments reduce a delay before the first sets of acoustic environment parameters can be used by the rendering.

[0044] In accordance with an optional feature of the invention, first acoustic environment data for a given set of the first sets of acoustic environment parameters comprises a data integrity verification value, and the acoustic data generator is arranged to generate the acoustic environment data from a previously received set of the first sets of acoustic environment parameters rather than from the given set if the data integrity verification value matches a data integrity verification value generated from data of the previously received set.

[0045] This may provide improved and/or facilitated operation or performance in many embodiments. The data integrity verification value may specifically be a checksum value.

[0046] In some embodiments, the first acoustic data comprises a first acoustic environment identifier and the second acoustic data comprises a second acoustics environment identifier, the first acoustic environment identifier and the second acoustic environment identifier both being indicative of the acoustic environment, and the acoustic data generator is arranged to generate the rendering acoustic environment data in dependence on the first acoustic environment identifier and the second acoustic environment identifier.

[0047] In accordance with an optional feature of the invention, the second sets of acoustic environment parameters comprise fewer parameters than the first sets of acoustic environment parameters.

[0048] This may provide improved and/or facilitated operation or performance in many embodiments.

[0049] In accordance with an optional feature of the invention, at least one set of the first sets of acoustic environment parameters comprise at least one parameter differentially encoded relative to a parameter of a set of the second sets of acoustic environment parameters.

[0050] This may provide improved and/or facilitated operation or performance in many embodiments.

[0051] According to an aspect of the invention, there is provided a data signal comprising at least a first audio signal and metadata including: first acoustic environment data for an acoustic environment, the first acoustic environment data comprising repeated first sets of acoustic environment parameters, each first set of parameters providing a description of the acoustic environment; second acoustic environment data for the acoustic environment, the second acoustic environment data comprising repeated second sets of acoustic environment parameters, each second set of parameters providing a description of the acoustic environment, a data size of the first sets of parameters exceeding a data size for the second sets of parameters and an update rate for the second sets of parameters is higher than an update rate for the first sets of parameters.

[0052] There may be provided an apparatus for generating such a data signal.

[0053] In accordance with an optional feature of the invention, the apparatus further comprises: a transmitter arranged to transmit the data signal over a communication channel; a determiner arranged to determine a communication channel property for the communication channel; and a controller arranged to adapt a property of the first acoustic environment data in dependence on the communication channel property.

[0054] This may provide improved and/or facilitated operation or performance in many embodiments. It may in many embodiments reduce the impact of communication conditions. The communication channel may be a connection of a network and the communication channel property may be a (network/connection) capacity, data rate, error rate, etc.

[0055] According to an aspect of the invention there is provided a method of generating an output audio signal, the method comprising: receiving a data signal comprising audio data for at least a first audio signal and metadata including: first acoustic environment data for an acoustic environment, the first acoustic environment data comprising repeated first sets of acoustic environment parameters, each first set of acoustic environment parameters providing a description of the acoustic environment; second acoustic environment data for the acoustic environment, the second acoustic environment data comprising repeated second sets of acoustic environment parameters, each second set of acoustic environment parameters providing a description of the acoustic environment, a data size of the first sets of acoustic environment parameters exceeding a data size for the second sets of acoustic environment parameters and an update rate for the second sets of acoustic environment parameters is higher than an update rate for the first sets of acoustic environment parameters; selecting between the first acoustic environment data and the second acoustic environment data to generate rendering acoustic environment data; and generating the audio output signal by rendering the audio signal based on the rendering acoustic environment data.

[0056] 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

[0057] 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 client server based Virtual Reality system;

FIG. 2 illustrates an example of elements of an audio rendering apparatus in accordance with some embodiments of the invention;

FIG. 3 illustrates an example of elements of an audio data signal generating apparatus in accordance with some embodiments of the invention; and

FIG. 4 illustrates an example of a data structure for acoustic environment data;

FIG. 5 illustrates an example of a data structure for acoustic environment data in accordance with some embodiments of the invention;

FIG. 6 illustrates an example of a data structure for acoustic environment data in accordance with some embodiments of the invention;

FIG. 7 illustrates an example of a data structure for acoustic environment data in accordance with some embodiments of the invention;

FIG. 8 illustrates an example of a data structure for acoustic environment data in accordance with some embodiments of the invention;

FIG. 9 illustrates an example of some elements of an audio data signal generating apparatus in accordance with some embodiments of the invention; and

FIG. 10 illustrates some elements of a possible processor arrangement for implementing elements of an apparatus in accordance with some embodiments of the invention.

DETAILED DESCRIPTION OF SOME EMBODIMENTS OF THE INVENTION

[0058] The following description will focus on eXtended Reality applications where audio is rendered, following a user position in an audio scene to provide an immersive user experience. Typically, the audio rendering may be accompanied by a rendering of images such that a complete audiovisual experience is provided to the user. However, it will be appreciated that the described approaches may be used in many other applications including applications where a user position is not explicitly considered.

[0059] EXtended Reality (including Virtual Augmented and Mixed Reality) experiences allowing a user to move around in a virtual or augmented world are becoming increasingly popular and services are being developed to improve such applications. In many such approaches, visual and audio data may dynamically be generated to reflect a user's (or viewer's) current pose.

[0060] In the field, the terms placement and pose are used as a common term for position and/or orientation / direction. The combination of the position and direction/ orientation of e.g., an object, a camera, a head, or a view may be referred to as a pose or placement. Thus, a placement or pose indication may comprise up to six values/ components/ degrees of freedom with each value/ component typically describing an individual property of the position/ location or the orientation/ direction of the corresponding object. Of course, in many situations, a placement or pose may be represented by fewer components, for example if one or more components is considered fixed or irrelevant (e.g., if all objects are considered to be at the same height and have a horizontal orientation, four components may provide a full representation of the pose of an object). In the following, the term pose is used to refer to a position and/or orientation which may be represented by one to six values (corresponding to the maximum possible degrees of freedom).

[0061] Many XR applications are based on a pose having the maximum degrees of freedom, i.e., three degrees of freedom of each of the position and the orientation resulting in a total of six degrees of freedom. A pose may thus be represented by a set or vector of six values representing the six degrees of freedom and thus a pose vector may provide a three-dimensional position and/or a three-dimensional direction indication. However, it will be appreciated that in other embodiments, the pose may be represented by fewer values.

[0062] A system or entity based on providing the maximum degree of freedom for the viewer is typically referred to as having 6 Degrees of Freedom (6DoF). Many systems and entities provide only an orientation or position and these are typically known as having 3 Degrees of Freedom (3DoF).

[0063] Typically, the Virtual Reality application generates a three-dimensional output in the form of separate view images for the left and the right eyes. These may then be presented to the user by suitable means, such as typically individual left and right eye displays of a VR headset. In other embodiments, one or more view images may e.g., be presented on an autostereoscopic display, or indeed in some embodiments only a single two-dimensional image may be generated (e.g., using a conventional two-dimensional display).

[0064] Similarly, for a given viewer/ user/ listener pose, an audio representation of the scene may be provided. The audio scene is typically rendered to provide a spatial experience where audio sources are perceived to originate from desired positions. As audio sources may be static in the scene, changes in the user pose will result in a change in the relative position of the audio source with respect to the user's pose. Accordingly, the spatial perception of the audio source may change to reflect the new position relative to the user. The audio rendering may accordingly be adapted depending on the user pose.

[0065] The listener pose input may be determined in different ways in different applications. In many embodiments, the physical movement of a user may be tracked directly. For example, a camera surveying a user area may detect and track the user's head (or even eyes (eye-tracking)). In many embodiments, the user may wear a VR headset which can be tracked by external and/or internal means. For example, the headset may comprise accelerometers and gyroscopes providing information on the movement and rotation of the headset and thus the head. In some examples, the VR headset may transmit signals or comprise (e.g., visual) identifiers that enable an external sensor to determine the position and orientation of the VR headset.

[0066] In many systems, the VR/ scene data may be provided from a remote device or server. For example, a remote server may generate audio data representing an audio scene and may transmit audio signals corresponding to audio components/ objects/ channels, or other audio elements corresponding to different audio sources in the audio scene together with position information indicative of the position of these (which may e.g., dynamically change for moving objects). The audio signals/elements may include elements associated with specific positions but may also include elements for more distributed or diffuse audio sources. For example, audio elements may be provided representing generic (non-localized) background sound, ambient sound, diffuse reverberation etc.

[0067] The local VR device may then render the audio elements appropriately, and specifically by applying appropriate binaural processing reflecting the relative position of the audio sources for the audio components.

[0068] Similarly, a remote device may generate visual/video data representing a visual audio scene and may transmit visual scene components/ objects/ signals or other visual elements corresponding to different objects in the visual scene together with position information indicative of the position of these (which may e.g., dynamically change for moving objects). The visual items may include elements associated with specific positions but may also include video items for more distributed sources.

[0069] In some embodiments, the visual items may be provided as individual and separate items, such as e.g., as descriptions of individual scene objects (e.g., dimensions, texture, opaqueness, reflectivity etc.). Alternatively or additionally, visual items may be represented as part of an overall model of the scene e.g., including descriptions of different objects and their relationship to each other.

[0070] For a VR service, a central server may accordingly in some embodiments generate audiovisual data representing a three dimensional scene, and may specifically represent the audio by a number of audio signals representing audio sources in the scene which can then be rendered by the local client/ device.

[0071] FIG. 1 illustrates an example of a VR/XR system in which a central server 101 liaises with a number of remote clients 103 e.g., via a network 105, such as e.g., the Internet. The central server 101 may be arranged to simultaneously support a potentially large number of remote clients 103.

[0072] Such an approach may in many scenarios provide an improved trade-off e.g., between complexity and resource demands for different devices, communication requirements etc. For example, the scene data may be transmitted only once or relatively infrequently with the local rendering device (the remote client 103) receiving a viewer pose and locally processing the scene data to render audio and/or video to reflect changes in the viewer pose. This approach may provide for an efficient system and attractive user experience. It may for example substantially reduce the required communication bandwidth while providing a low latency real time experience while allowing the scene data to be centrally stored, generated, and maintained. It may for example be suitable for applications where a VR experience is provided to a plurality of remote devices.

[0073] FIG. 2 illustrates elements of an apparatus for generating an output audio signal, henceforth also referred to as an audio rendering apparatus, which may generate an improved output audio signal in many applications and scenarios. In particular, the audio rendering apparatus may provide improved rendering for many VR applications, and the audio rendering apparatus may specifically be arranged to perform the audio processing and rendering for a VR client 103 of FIG. 1.

[0074] The audio apparatus of FIG. 2 is arranged to render audio of a three dimensional scene to provide a three dimensional perception of the scene. The specific description will focus on audio rendering, but it will be appreciated that in many embodiments this may be supplemented by a visual rendering of the scene. Specifically, images may in many embodiments be generated and presented to the user.

[0075] FIG. 3 illustrates an example of an apparatus, also henceforth referred to as an audio encoding apparatus, for generating an audio data signal that includes (data describing) one or more audio signals as well as metadata that describes an acoustic environment. In particular, the audio encoding apparatus may provide improved representation of an audio source or scene for many VR applications, and the audio encoding apparatus may specifically be arranged to perform the audio processing and function for a VR server 101 of FIG. 1 The audio encoding apparatus may provide the audio data signal which is transmitted to the audio rendering apparatus which proceeds to render the audio signal based on the audio signal data and the acoustic environment data.

[0076] The approach of transmitting audio together with acoustic environment data has been proposed in different contexts. It has for example been proposed for the Immersive Voice and Audio Services (IVAS) standard under development by 3GPP that room reverberation parameters may be transmitted and used to locally render room reverberation audio.

[0077] However, an issue related to the transmission of such acoustic environment data is that it tends to require a relatively large amount of data to be transmitted in order to provide accurate and detailed information of the acoustic environment. In order to allow easy access, and potentially dynamically updated information, it is further desired that the acoustic environment data is repeatedly included in the data signal which results in a high data rate being required for the acoustic environment data resulting in inefficient bandwidth utilization.

[0078] Indeed, for several use-cases it is required that e.g., room acoustics should be in-line with the audio content. This might become especially relevant in case of VR/XR gaming/social communications/streaming, etc. Not all room acoustics information may be available upfront and should therefore be communicated together with the audio in a dynamic and continuous way. Further, users may in many applications and scenarios join or rejoin an existing session instantly, leaving little time to share metadata, including environment acoustics information. As the room acoustics information may not be instantly available, the synthesized room/environment acoustics may not match the audiovisual content until the actual acoustics data has been received.

[0079] The apparatuses of FIGs. 2 and 3 use an approach that may provide improved communication and usage of acoustic environment data, and in particular may generate and process an (audio) data signal that may provide both audio and acoustic environment data which may allow improved trade-off between the conflicting requirements and desires,

such as an improved trade-off between data rate and delay associated with rendering of audio using acoustic environment data.

[0080] The audio rendering apparatus of FIG. 2 comprises a receiver 201 which is arranged to receive and (audio) data signal. The data signal may for example be received from the audio encoding apparatus of FIG. 3. The audio apparatus of FIG. 2 is specifically arranged to render audio of a three dimensional scene to provide a three dimensional perception of the scene. The specific description will focus on audio rendering, but it will be appreciated that in many embodiments this may be supplemented by a visual rendering of the scene. Specifically view images may in many embodiments be generated and presented to the user.

[0081] The receiver 201 receives the data signal which comprises at least a first audio signal. The audio signal may specifically comprise audio data describing/representing audio for one or more audio sources of an audio scene. The audio data may specifically provide audio data for audio objects, audio channel signals, non-spatial audio (e.g. diffuse/ambient sources), etc.

[0082] In addition, the data signal further comprises acoustic environment data and indeed comprises two different versions/formats of acoustic environment data.

[0083] The acoustic environment data specifically comprises first acoustic environment data for an acoustic environment. The first acoustic environment data comprises repeated first sets of acoustic environment parameters where each first set of acoustic environment parameters provide a description of the acoustic environment. A set of the first sets of acoustic environment parameters will also for conciseness be referred to as a first set of parameters, or simply as a first set.

[0084] Further, in addition to the first acoustic environment data, the data signal also comprises second acoustic environment data for the acoustic environment. The second acoustic environment data comprises repeated second sets of acoustic environment parameters where each second set provides a description of the acoustic environment. In many embodiments, the second set may comprise data for the same parameters as the first sets. For example, the second sets may also comprise parameters reflecting a reverberation decay rate, a reverberation energy rate, and/or optionally a reverberation delay indication. Indeed, in many embodiments, one, more, or all of the parameters in the first and second sets may be alternative parameters that describe the same property. A set of the second sets of acoustic environment parameters will also for conciseness be referred to as a second set of parameters, or simply as a second set.

[0085] Each set of parameters may provide a full/complete representation/description of the acoustic environment. A rendering of audio for the acoustic environment may be based on a single set of acoustic environment parameters without requiring information from any other set of acoustic environment parameters. Thus, each set of acoustic environment parameters may provide an independent/separate representation/description of the acoustic environment. Thus, one complete first set of acoustic environment parameters may provide a full description with no requirement for, or additional information being provided by other first sets. Similarly, one complete second set of acoustic environment parameters may provide a full description with no requirement for, or additional information being provided by other second sets.

[0086] The acoustic environment parameters may thus provide parameters that reflect acoustic properties of the acoustic environment, and specifically the parameters may provide a complete set of parameters. Specifically, rendering of the audio signal may be performed based on the parameters of a single set of acoustic environment parameters.

[0087] In many embodiments, the sets of acoustic environment parameters may comprise, or consist in, parameters describing reverberation properties of the acoustic environment. The acoustic environment data may be or include acoustic reverberation data. In some cases, such parameters may for example simply be indicative of properties of a room, such as the dimensions and reflectivity of walls of a room etc. In many embodiments, the sets of acoustic environment parameters may include parameters that directly describe impulse response properties for a room/environment, and specifically properties of the reverberating part of the impulse response. For example, the parameters may in particular include parameters indicating relative energy of the diffuse reverberation and/or a decay rate and/or a delay of the diffuse reverberation tail etc.

[0088] In many embodiments, the sets of acoustic environment parameters may specifically comprise a T_{60} parameter and a DSR parameter. A T_{60} parameter is indicative of a reverberation time/ decay and specifically indicates the time for the level to reduce by 60dB. DSR parameter may indicate a Diffuse to Source Ratio which indicates a relationship, and specifically a ratio, between the energy of the diffuse reverberation and the total energy of the audio source. Such a measure may provide a particularly advantageous operation as it may be analogous to how the physics works, where all sound source energy emitted from a source position in all directions contributes to reflections that combine to create the diffuse field. With this generic characterization of an environment's reverberation properties, there is no dependency on the direct path between certain source- and receiver positions within the environment. Also, the directivity pattern is included in the source energy, as it considers the total energy emitted into the environment from all directions of the source.

[0089] These parameters may be provided per frequency band. These frequency bands can be specified in the data set. Alternatively, the data set can specify an identifier of one of the default data grids to be used. Next to T_{60} parameter and a DSR parameter, a (pre)delay parameter may be provided, which specifies the delay at which the DSR parameters were computed. Additionally, the set of acoustic environment parameters may comprise shoebox model parameters for early reflections synthesis. These parameters may include physical room properties, such as room dimensions, individual wall

absorption coefficients, etc. Additionally, the set of acoustic environment parameters may also include default listener pose (position and orientation).

[0090] The data size of (each of) the first sets exceeds the data size of (each of) the second sets. Specifically, the number of bits used to represent a first set exceeds the number of bits used to represent a second set. Each set of the first and second sets may provide a complete representation of the acoustic environment but with the first sets providing a more accurate and/or detailed description of the acoustic environment than the second sets. Thus, the second sets provide a coarser and less data intense description than the first sets. In many embodiments, the data size for each of the first sets may exceed the data size for each of the second sets by a factor of no less than 2, 5, 10, or even 25 times.

[0091] Further, the update rate for the first sets is lower than an update rate for the second sets, and the second sets are received/ included with a higher update rate in the data signal than the first sets. The update rate may be the rate of transmissions of complete sets of the first and second sets respectively. The duration between the inclusion of consecutive first sets in the data signal is higher than the duration between the inclusion of consecutive second sets. The data signal includes repeated first sets and repeated second sets with the duration between the first sets being higher, and typically substantially higher, than the duration between second sets. The update rate for the first sets may be a rate/frequency of the repetitions of (the complete data for) the first sets and similarly the update rate for the second sets may be a rate/frequency of the repetitions of (the complete data for) the second sets.

[0092] In many embodiments, the update rate for the first sets of acoustic environment parameters is no less than 30 seconds, or sometimes even 1 or 5 minutes, whereas the update rate for the second set of acoustic environment parameters is no higher than 10 secs.

[0093] Thus, in many embodiments, a data signal is received which comprises two forms/versions of acoustic environment data. It includes first acoustic environment data that includes first sets of acoustic environment parameters which provide detailed and high quality description of the acoustic environment with the first sets having a relatively large data size but being infrequently updated. This first acoustic environment data is supplemented by second acoustic environment data which provides a less detailed and lower quality description of the acoustic environment with the second sets having a significantly smaller data size but being more frequently updated. The first and second acoustic environment data typically provide alternative representations.

[0094] The receiver 201 is coupled to an acoustic data generator 203 which is arranged to generate rendering acoustic environment data which includes data selected from the first and/or second acoustic environment data, and specifically the acoustic data generator 203 may be arranged to select between generating the rendering acoustic environment data from the received first acoustic environment data or from the received second acoustic environment data.

[0095] The acoustic data generator 203 may specifically be arranged to generate a rendering set of acoustic environment parameters, also sometimes just referred to as a rendering set, by selecting parameter values from the first sets of acoustic environment parameters and the second sets of acoustic environment parameters, and may specifically generate the rendering set of acoustic environment parameters by selecting either a first set or a second set.

[0096] In many embodiments, the selection may be based on the data that has been received and specifically on which sets of acoustic environment parameters that have been received. The acoustic data generator 203 may for example be arranged to generate a rendering set of acoustic environment parameters by selecting a received set of acoustic environment parameters except if/until a first set of acoustic environment parameters has been received. For example, when initializing a rendering operation, the receiver 201 may initially receive a second set of acoustic environment parameters as this has a high update rate. Accordingly, the acoustic data generator 203 may proceed to generate the rendering set of acoustic environment parameters to be identical to (include the parameters of) the received second set. However, when subsequently a first set of acoustic environment parameters is received, the acoustic data generator 203 may proceed to instead generate the rendering set of acoustic environment parameters to be identical to (including the parameters of) the received first set. In some embodiments, the acoustic data generator 203 may be arranged to select acoustic environment data/ a set of parameters/ parameters from the first acoustic environment data if received and to select environment data/ a set of parameters/ parameters from the second acoustic environment data if (corresponding) first acoustic environment data has not been received.

[0097] Such an approach may reduce the delay (in particular when initiating rendering) before the acoustic environment can be reflected in the rendering while still allowing high quality rendering of the acoustic environment during normal operation when the first acoustic environment data is subsequently received.

[0098] The receiver 201 and the acoustic data generator 203 are coupled to a renderer 205 which is arranged to generate an audio output signal by rendering the received audio signal in dependence on the rendering acoustic environment data, and specifically based on the rendering set of acoustic environment parameters.

[0099] The renderer 205 may proceed to render the audio scene based on the received audio signal which may include various audio items/elements including audio objects linked with a position, ambient audio elements, channel-based audio associated with nominal positions etc. In case of encoded data, the renderer 205 may also be arranged to decode the audio data (or in some embodiments decoding may be performed by the receiver 201).

[0100] The renderer 205 is arranged to render the audio scene by generating audio signals based on the received audio

data representing audio from various audio sources including diffuse audio sources representing e.g., ambient noise and sounds. The audio is generated to reflect the acoustic environment, such as specifically to reflect the reverberation properties of the acoustic environment.

[0101] In the example, the renderer 205 is specifically a binaural audio renderer which generates binaural audio signals for a left and right ear of a user. The binaural audio signals are generated to provide a desired spatial experience and are typically reproduced by headphones or earphones that specifically may be part of a headset worn by a user (the headset typically also comprises left and right eye displays).

[0102] Thus, in many embodiments, the audio rendering by the renderer 205 is a binaural render process using suitable binaural transfer functions to provide the desired spatial effect for a user wearing a headphone. For example, the renderer 205 may be arranged to generate an audio component to be perceived to arrive from a specific position using binaural processing.

[0103] Binaural processing is known to be used to provide a spatial experience by virtual positioning of sound sources using individual signals for the listener's ears. With an appropriate binaural rendering processing, the signals required at the eardrums in order for the listener to perceive sound from any desired direction can be calculated, and the signals can be rendered such that they provide the desired effect. These signals are then recreated at the eardrum using either headphones or a crosstalk cancelation method (suitable for rendering over closely spaced speakers). Binaural rendering can be considered to be an approach for generating signals for the ears of a listener resulting in tricking the human auditory system into perceiving that a sound is coming from the desired positions.

[0104] The binaural rendering is based on binaural transfer functions which vary from person to person due to the acoustic properties of the head, ears and reflective surfaces, such as the shoulders. Binaural transfer functions may therefore be personalized for an optimal binaural experience. For example, binaural filters can be used to create a binaural recording simulating multiple sources at various locations. This can be realized by convolving each sound source with the pair of e.g., Head Related Impulse Responses (HRIRs) that correspond to the position of the sound source.

[0105] A well-known method to determine binaural transfer functions is binaural recording. It is a method of recording sound that uses a dedicated microphone arrangement and is intended for replay using headphones. The recording is made by either placing microphones in the ear canal of a subject or using a dummy head with built-in microphones, a bust that includes pinnae (outer ears). The use of such dummy head including pinnae provides a very similar spatial impression as if the person listening to the recordings was physically present during the recording.

[0106] By measuring e.g., the responses from a sound source at a specific location in 2D or 3D space to microphones placed in or near the human ears, the appropriate binaural filters can be determined. Based on such measurements, binaural filters reflecting the acoustic transfer functions to the user's ears can be generated. The binaural filters can be used to create a binaural recording simulating multiple sources at various locations. This can be realized e.g., by convolving each sound source with the pair of measured impulse responses for a desired position of the sound source. In order to create the illusion that a sound source is moving around the listener, a large number of binaural filters is typically required with a certain spatial resolution, e.g., 10 degrees.

[0107] The head related binaural transfer functions may be represented e.g., as Head Related Impulse Responses (HRIR), or equivalently as Head Related Transfer Functions (HRTFs) or, Binaural Room Impulse Responses (BRIRs). The (e.g., estimated or assumed) transfer function from a given position to the listener's ears (or eardrums) may for example be represented in the frequency domain in which case it is typically referred to as an HRTF or BRTF, or in the time domain in which case it is typically referred to as a HRIR or BRIR. In some scenarios, the head related binaural transfer functions are determined to include aspects or properties of the acoustic environment and specifically of the environment in which the measurements are made, whereas in other examples only the user characteristics are considered. Examples of the first type of functions are the BRIRs and BRTFs.

[0108] The renderer 205 may be arranged to individually apply binaural processing to a plurality of audio signals/sources and may then combine the results into a single binaural output audio signal representing the audio scene with a number of audio sources positioned at appropriate positions in the sound stage.

[0109] The audio rendering apparatus further comprises a listener pose processor 207 which is arranged to determine a listener pose for which the output audio signal is generated. The listener pose may accordingly correspond to the position in the scene of the user/listener.

[0110] The listener pose may specifically be determined in response to sensor input, e.g., from suitable sensors being part of a headset. It will be appreciated that many suitable algorithms will be known to the skilled person and for brevity this will not be described in more detail.

[0111] It will be appreciated that many different spatial rendering algorithms are known and that any suitable approach may be used. It will also be appreciated that rendering approaches and techniques for generating audio to represent the acoustic environment are well known.

[0112] For example, for a rendering set of acoustic environment parameters representing reverberation properties, such as a diffuse reverberation energy rate and a decay rate (T_{60}), it is well known to implement parametric reverberators, such as a Jot reverberator, which can be initialized with the determined parameter values. More description of such a

reverberator may e.g. be found in Jot, J.-M. and Chaigne, A. 1991. "Digital delay networks for designing artificial reverberators". Proc. 90th Conv. Audio Eng. Soc.

[0113] As another example, for binaural rendering, BRIRs may be stored for different rendering parameter values and rendering may include selecting the BRIR(s) that most closely represents/matches the parameters of the rendering set, and then proceeding to render the audio signal using the extracted BRIR(s).

[0114] It will be appreciated that many other rendering algorithms and approaches may be used and that many such techniques will be known to the skilled person and therefore will for conciseness not be described further herein.

[0115] FIG. 3 illustrates an example of an audio encoding apparatus that may generate the data signal that is provided to the audio rendering apparatus of FIG. 2.

[0116] The audio encoding apparatus comprises an audio data receiver 301 which is arranged to receive audio data representing a number of different audio sources. The audio data receiver 301 is arranged to generate the audio signal data for inclusion in the data signal. The audio data receiver 301 may in some embodiments be arranged to merely combine received audio data or may in some embodiments be arranged to process the received audio data to generate the audio data for the data signal. Such processing may for example include combining or separating audio from different audio sources, performing audio encoding, etc.

[0117] The audio encoding apparatus further comprises an acoustic environment data receiver 303 which is arranged to receive data providing information on the acoustic environment. The acoustic environment data receiver 303 is arranged to generate the acoustic environment data and specifically is arranged to generate both the first acoustic environment data and the second acoustic environment data. In some embodiments, the acoustic environment data receiver 303 may directly receive data corresponding to the first and second acoustic environment data, but in other embodiments it may be arranged to process the received data to generate the acoustic environment data. For example, the acoustic environment data receiver 303 may receive data describing physical properties of a room, such as dimensions and wall reflectivity, and may therefrom determine the acoustic environment data (e.g., it may use simulation or formulas for determining a reverberation energy, delay and/or decay rates).

[0118] In some embodiments, the acoustic environment data receiver 303 may receive acoustic environment data that directly may be used as the first acoustic environment data. It may then proceed to generate the second acoustic environment data from the received first acoustic environment data. For example, the acoustic environment data receiver 303 may proceed to generate coarse values by reducing a level quantization, frequency resolution, combining parameters, etc.

[0119] The audio data receiver 301 and the acoustic environment data receiver 303 are coupled to a data signal generator 305 which is arranged to generate the output data signal which may be transmitted or distributed to the audio rendering apparatus of FIG. 2. The data signal generator 305 may combine the audio data as appropriate, typically including generation of repeated inclusions of the first and second sets of set of acoustic environment parameters with different update/repetition rates.

[0120] In many applications, the data signal may be a real time data signal providing audio data that is rendered in real time. Thus, the audio rendering apparatus may receive the data signal and proceed to render it as it is being received. The audio encoding apparatus, the audio rendering apparatus, and the data signal may all operate in accordance with a real time protocol allowing rendering in real time as the data signal is being transmitted.

[0121] In many embodiments, the data signal may be transmitted/communicated in a plurality of data packets with each data packet being transmitted individually and separate to other data packets. For example, for a distribution via a network, e.g., including the Internet, the data packets may reach the audio rendering apparatus via different paths, and indeed in some embodiments each data packet may be routed individually and independently of the routing of other data packets.

[0122] In many embodiments, a quantization of one, more, and possibly all parameters of the second sets of parameters are coarser than corresponding parameters of the first sets of parameters.

[0123] Specifically, one or more parameters representing the same property in respectively the first sets and the second sets may have a coarser quantization in the second sets. In many embodiments, the first and second sets may provide data for the same parameter, such as for example a parameter representing a property, such as a reverberation delay, energy rate or a reverberation decay rate, of the impulse response of the acoustic environment/ room.

[0124] The quantization of such a parameter/property for which a value is provided in both the first and in the second set, the quantization of the parameter may be coarser in the second set than in the first set. For example, the frequency quantization may be finer for the parameter in the first sets rather than in the second sets. For example, in the first set, values may be provided for a reverberation decay rate and/or delay for a plurality of different frequency bands, such as for example for, say, 8-32 different frequency bands in the audio range. However, in contrast, in the second set, a single value may be included for the parameter, and specifically only a single reverberation energy rate and/or delay parameter value may be included.

[0125] This may allow an advantageous trade-off where an accurate reflection of the frequency dependency of the parameters is represented for most of the rendering while also allowing quick adaptation when starting the rendering based on the received acoustic environment data.

[0126] In many embodiments, the values of one or more parameters may be represented by a coarser level quantization for the second set relative to the first set. In particular, the parameter values may be represented by data words having fewer bits in the second set than in the first set.

[0127] As a specific example for T_{60} and DSR parameters, values may be provided for a relatively high number of frequency bands in the first set, e.g., for 8, 16, 32, 64 or even more frequency bands whereas it for the second sets may be limited to three frequency bands: low, medium, and high frequency. For instance, a grid of three center frequencies of 25, 250, and 2500Hz may be used. Similarly, the reverb parameters may in the second set be represented by low-resolution variants of higher resolution variants in the first set. For example, RT_{60} and DSR values can be represented using code words of 8 bits on average for the second set. Such an approach would require less than 10 bits per second of bandwidth to send rough/coarse room acoustics data every 5 seconds.

[0128] In some embodiments, the second sets may comprise fewer acoustic environment parameters than the first sets. For example, the first sets may include a parameter that describes a particular acoustic property which may not be included in the second set. The audio rendering apparatus may be arranged to perform the rendering based on a nominal or predetermined value of that parameter if only second sets have been received. For example, a pre-delay value corresponding to an average sized room may be used until a first set of acoustic environment parameters is received which provides an actual specific pre-delay value for the current acoustic environment. The nominal values may then be replaced by the values of the first sets to provide a more accurate rendering of the acoustic environment.

[0129] In some embodiments, such a nominal or predetermined value may be an absolute value that is used directly, e.g., after being retrieved from memory. In other cases, the nominal or predetermined value may be a relative value which may e.g., be determined from one of the received values.

[0130] In some embodiments, at least some of the first acoustic environment data may be differentially encoded with respect to the second acoustic environment data. For example, at least one of the parameters of a first set may be represented by a relative value with respect to the corresponding parameter of a second set.

[0131] As an example, the second sets may include a single, say 5 bit word, value of a T_{60} decay value. The second set may then include data which provides values for the T_{60} decay for a plurality of different frequency bands by indicating the difference for the different frequency bands relative to the single T_{60} decay value of the second set. Further, the first sets may include further data bits providing a finer level quantization of the values. For example, the first sets may include, say 6 bit, values corresponding to the three LSBs of the value of the second set and to three additional LSBs to provide a more accurate and nuanced values for a plurality of frequency bands.

[0132] Thus, in some embodiments, the approach could be using differential coding of room acoustics data, where high-resolution representation could be coded as deltas with respect to low-resolution representation.

[0133] The approach may exploit a consideration that having a rough approximation of e.g., room acoustics can provide an improved perception than using default room acoustic parameters or disabling room acoustics synthesis before the complete room acoustics parameters are available. Therefore, a complete room acoustics parameter block in the form of a first set of acoustic environment parameters may be transmitted with very low frequency (e.g., once per minute) to minimize bandwidth utilization. An example of such a full block/ first set that particularly describes reverberation properties are shown in FIG. 4. A possible pseudo-syntax may be as follows:

```
frequency_grid() {
    …
}

early_reflections() {
    …
}

room_acoustics_parameters_full(){
    id;                                    ??   bslbf
    frequency_grid();
    for (n=0; n < N; n++) {
        rt_sixty[n];                       ??   uimsbf
        dsr[n];                            ??   uimsbf
    }
    pre_delay;                             ??   uimsbf
    if (early_reflections_flag) {          1    blsbf
        early_reflections();
    }
}
```

[0134] In the above syntax, a full set of room acoustic parameters consist of an id, a frequency grid description frequency_grid(), the T_{60} and DSR parameters for a number of N bands, as specified by the frequency grid description, a pre-delay parameter, and a set of early reflections early _reflections() that can be toggled on or off.

[0135] Additionally, compact room acoustics parameter blocks in the form of second sets of acoustic environment parameters may be sent with much higher frequency (e.g., once per 5 seconds). Such a block may contain only a limited number of e.g., roughly quantized RT60 and DSR data, with optionally a room acoustics environment identifier, as e.g., illustrated in FIG. 5 for an IVAS (Immersive Voice and Audio Services) session. A possible syntax is:

```
room_acoustics_parameters_compact(){
    id;                                    ??   bslbf
    for (m=0; m < 3; m++) {
        rt_sixty[m];                       ??   blsbf
        dsr[m];                            ??   bslbf
    }
}
```

[0136] In the example, when the IVAS session starts or gets reestablished, such rough room acoustics data will become available within a few seconds. This will allow the audio to be rendered using room acoustics parameters that are possibly much closer to the target room acoustics than in the case of using default parameters. This can be also useful if the room acoustic environment gets changed abruptly, for example in case of a sudden change in a gaming/social VR/XR application.

[0137] Each set of the second set of acoustic environment parameters may typically be transmitted as a single block of data, and specifically in a single data packet. Similarly, in some embodiments, each first set of the first set of acoustic environment parameters may be transmitted as a single block of data, and specifically in a single data packet. In such cases, the data blocks/packets for the second sets may be transmitted more frequently than the data blocks/packets for the first sets.

[0138] However, in many embodiments, each set of the first sets of acoustic environment parameters are distributed over a plurality of non-contiguous data segments, and specifically over a plurality of non-contiguous data blocks/fragments. In particular, in many embodiments, a single first set is distributed over a plurality of data packets.

[0139] In particular, for the specific example of room acoustics, the complete room acoustics parameters may be transmitted in a single data packet with very low frequency. However, this might create a local peak in bandwidth usage. To

mitigate this, the complete room acoustics parameters can be fragmented and transported as multiple extensions in multiple compact room acoustics parameters packets. Examples of data structures for such an approach are shown in FIG. 6. An exemplary data structure syntax could be:

```
room_acoustics_parameters (){
     id;                                        ??    bslbf
     for (m=0; m < 3; m++) {
          rt_sixty[m];                          ??    blsbf
          dsr[m];                               ??    bslbf
     }
     if (extension_fragment_flag) {             1     blsbf
          extension_fragment();
     }
}
```

[0140] An extension fragment should be very compact to minimize bandwidth usage. It should also enable reconstruction of fragments once all are collected. The extension fragment format could be based on a complete room acoustics data bitstream, as provided with IVAS deliverables. This can be accompanied with a data integrity verification value, such as specifically a checksum, for data completeness and integrity check. Consolidated extension fragments sequence could look as indicated in FIG. 6.

[0141] In the example, the data segments, and specifically data segments/extension fragments, may comprise an indication of a start (and accordingly also the end) position for the data of the first set. Thus, the received data may also indicate in which packet/extension fragment a given set ends and a new set begins.

[0142] Alternatively or additionally, the data signal may include an indication of the data size for data for the first set. For example, one data packet or extension fragment may indicate a length/size of the data that is included for a single set. For example, it may be indicated how many data packets are transmitted for each set.

[0143] The audio rendering apparatus may then, based on such information, be arranged to extract and combine the data to provide a full first set. The approach may allow the audio rendering apparatus to continuously receive and store the data of the first set, and to determine when complete data for a given set has been received. For example, the receiver 201 may continuously receive and store the first set data in a local buffer/memory. It may then, based on the start point and/or data size, determine when a full set of first set data has been received and stored. The acoustic data generator 203 may then access the memory to generate a full first set to use for the generation of the rendering set of acoustic environment parameters.

[0144] For example, in some embodiments, when initializing rendering, the audio rendering apparatus may start rendering audio based on a rendering set of acoustic environment parameters that comprise nominal acoustic environment parameters. With a short duration (e.g., within 5 seconds), a second set of acoustic environment parameters may be received, and the rendering set of acoustic environment parameters may be generated to include the parameters thereof. This may provide an improved rendering more closely reflecting the acoustic environment. The audio rendering apparatus may further proceed to begin to receive and store data for a first set of set of acoustic environment parameters. When a full first set has been received (e.g., within a minute), the acoustic data generator 203 may proceed to generate the rendering set of acoustic environment parameters to comprise the parameters of the received first set thereby providing a more accurate representation of the acoustic environment.

[0145] In some embodiments, the acoustic data generator 203 may proceed to gradually replace/add parameters to the rendering set of acoustic environment parameters as more parameters of the first set are received.

[0146] In the example of FIG. 7, an EF_{begin} fragment may signal the beginning of a sequence of datalfragments that together describe a first set of acoustic environment parameters. Such an indication can for instance contain a unique bit sequence followed by the length of the extension sequence and checksum data. Subsequent extension fragments may then contain room acoustics data. A possible data signal syntax for the first fragment for a new set could be:

```
extension_fragment_begin(){
     extension_uid;                          ??    bslbf
     nr_extension_fragments;                 ??    uimsbf
     extension_checksum;                     ??    bslbf
     frequency_grid();
}
```

[0147] In this syntax, extension_fragment_begin() may be a special case of an extension_fragment(), extension _uid may be fixed bit string, nr_extension _fragments may indicate how many fragments there are, extension_checksum may be a checksum over concatenation of all extension fragments, and frequency_grid() may be sent at beginning.

[0148] A possible data signal syntax for another fragment of a set could be:

```
extension_fragment(){
     extension_index;                        ??    uimsbf
     nr_bands;                               ??    uimsbf
     for (b = 0; b < nr_bands; b++) {
          rt_sixty[b];                       ??    bslbf
          dsr[b];                            ??    bslbf
     }
     if (pre_delay_flag)                     1     bslbf
          pre_delay;                         ??    uimsbf
     }
     if (early_reflections_flag) {           1     blsbf
          early_reflections_fragment();
     }
}
```

[0149] In this syntax, extension _index may be the index of the extension, the pre-delay may be flagged as the above assumes no knowledge of "how far" one is in decoding the T_{60}/DSR parameters.

[0150] A decoding process for such extension fragments could be:

1) As extension fragments are encountered, the bit-stream chunks are kept in a sufficiently large array of memory using the extension _index for bookkeeping.
2) As the first extension_uid is found, signalling the extension_fragment_ begin, it is known how many fragments are to be required.
3) After this extension fragment (the extension_fragment_ begin) or any subsequent extension _fragment is found; it is put in the above list until all extension fragments form a complete list.
4) If a complete list is formed, the checksum is checked.
5) If the checksum is correct, all data is decoded (if not, wait for next iteration).

[0151] Another syntax could be:

```
extension_fragment_begin(){
     extension_uid;                          ??    bslbf
     nr_extension_fragments;                 ??    uimsbf
     extension_checksum;                     ??    bslbf
}
```

and

```
extension_fragment(){
     extension_index;                    ??          uimsbf
     nr_bits;                            ??          uimsbf
     extension_bitstream;                nr_bits     bslbf
}
```

[0152] The audio rendering apparatus may for example operate a process including the following steps/considerations:

- If no fragment is yet received, all the received extension fragments are ignored,
- Once the EF_{begin} fragment is received, a memory buffer is allocated reflecting expected extension data length,
- If a received checksum is equal to one of a previously received sequence, subsequent extension fragments are ignored,
- Subsequent fragments are read and stored in the allocated buffer,
- Once the number of fragments is equal to the length provided at the beginning of the sequence, the checksum of data read is computed and compared with the received checksum,
- Room acoustics data can be decoded eventually,
- Once EF_{begin} data is received unexpectedly, the extension fragment receive process gets reset.

[0153] In the worst-case scenario, the receiver requires almost two cycles of extension fragments to read the complete room acoustics data. A possible improvement is to begin reading extension fragments as soon as they start to appear and store them in a buffer that is big enough to accommodate the presumed maximum amount of extension fragments. Once EF_{begin} data is read, the reader will be provided with the sequence length and the checksum information. The difference between the sequence length provided in the EF_{begin} data and the number of EF fragments already received will indicate how many fragments are still missing. The receiver 201 can then continue reading until the whole sequence is read followed by sequence validation using the checksum data. This approach is illustrated in FIG. 7.

[0154] In some embodiments, the first sets may comprise a checksum value and the acoustic data generator 203 may generate the acoustic environment data from a previously received first set rather than from the current set if the checksum value matches a checksum generated from data of the previously received set.

[0155] For example, for the previously received first set, a checksum value may be generated from the received data. When a new first set is then received, it may include a checksum value for the new first set. This checksum value may be included in the first data fragment/packet for the new first set. If the checksum value for the previous first set matches that of the new first set, the acoustic data generator 203 may conclude that the two sets are identical, i.e., that the acoustic environment properties have not changed. It may accordingly proceed to use the previously received first set instead.

[0156] It will be appreciated that rather than a checksum value, other forms of a data integrity verification value may be used.

[0157] In some embodiments, acoustic environment data may be provided for a plurality of acoustic environments, such as from a plurality of rooms. In such cases, the set of acoustic environment parameters may be linked to acoustic environment identifiers that allow the data for different acoustic environments be combined to generate the full set of acoustic environment parameters for a given acoustic environment.

[0158] The provision of acoustics environment identifiers may for example allow handling of multiple room acoustics environments. In case an acoustics environment identifier is transmitted with compact room acoustics data, it may also indicate the current environment selection. This can be overridden by an application-level selection. The extension fragments may be used to transport multiple acoustics environments. To minimize the time required to collect complete room acoustic data for the currently selected environment, individual room acoustics environments can be transmitted separately while the data for the selected acoustics environment can be transmitted much more frequently. Such an approach is illustrated in FIG. 8 (where RAx refers to the currently selected room acoustics environment).

[0159] In some embodiments, the audio encoding apparatus may be arranged to adapt/determine a property of the first acoustic environment data, and specifically the first sets of acoustic environment parameters, depending on a property of the communication channel over which the data signal is transmitted. Similarly, in some embodiments, the audio encoding apparatus may be arranged to adapt/determine a property of the second acoustic environment data, and specifically the second sets of acoustic environment parameters, depending on a property of the communication channel over which the data signal is transmitted.

[0160] For example, the audio encoding apparatus may be arranged to adapt a data size of the first (and/or second) sets in dependence on a communication bandwidth of a communication channel over which the data signal is transmitted.

[0161] The data signal generator 305 may for example as illustrated in FIG. 9 comprise a transmitter 901 which is arranged to transmit the data signal over a communication channel. The communication channel may for example be a communication channel formed by a communication network, including e.g., the Internet, and the transmitter 901 may

include a network interface.

[0162] The data signal generator 305 further comprises a determiner 903 which is arranged to determine a communication channel property for a communication channel, such as specifically a communication bandwidth, capacity, data rate etc. of the connection from the audio encoding apparatus to the audio rendering apparatus. It will be appreciated that many different approaches for determining such parameters will be known to the skilled person.

[0163] The data signal generator 305 further comprises a controller 905 which is arranged to adapt a property of the first and or second acoustic environment data that is included in the data signal with the property being adapted in dependence on the communication channel property. For example, the controller 905 may be arranged to remove or include (or e.g., combine) some parameters of the acoustic environment data depending on the available communication bandwidth.

[0164] In some embodiments, the frequency/update rate of the first sets of acoustic environment parameters may be adapted in dependence on the communication property, such as in dependence on the bandwidth usage and the average time required to access room acoustics data.

[0165] The approach may for example be used to provide graceful degradation scenarios. For example, if the session capacity drops, a sender can stop sending complete environment acoustics data or can provide it with reduced resolution.

[0166] In the above example, the rendering of the audio signal(s) is based on a listener pose and it may be dynamically adapted to reflect the position of the listener e.g. moving around in the acoustic environment/audio scene. However, it will be appreciated that this is merely an option and that the described approach, and the principles behind the application and provision of the acoustic environment data, is not dependent thereon, but may equally apply to applications and embodiments where a listener pose is not actively determined and considered in the rendering. For example, in some embodiments, the renderer may render the audio based on the acoustic environment data but without considering the position of the listener within the acoustic environment. Such approaches may in particular be appropriate for scenarios in which the rendering based on the acoustic environment data is a reverberation rendering, and specifically a rendering of a diffuse and ambient reverberation sound in an acoustic environment. Such sound may typically be relatively independent of the listener's particular position and orientation, and the rendering may be independent of such information.

[0167] FIG. 10 is a block diagram illustrating an example processor 1000 according to embodiments of the disclosure. Processor 1000 may be used to implement one or more processors implementing an apparatus as previously described or elements thereof. Processor 1000 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.

[0168] The processor 1000 may include one or more cores 1002. The core 1002 may include one or more Arithmetic Logic Units (ALU) 1004. In some embodiments, the core 1002 may include a Floating Point Logic Unit (FPLU) 1006 and/or a Digital Signal Processing Unit (DSPU) 1008 in addition to or instead of the ALU 1004.

[0169] The processor 1000 may include one or more registers 1012 communicatively coupled to the core 1002. The registers 1012 may be implemented using dedicated logic gate circuits (e.g., flip-flops) and/or any memory technology. In some embodiments the registers 1012 may be implemented using static memory. The register may provide data, instructions and addresses to the core 1002.

[0170] In some embodiments, processor 1000 may include one or more levels of cache memory 1010 communicatively coupled to the core 1002. The cache memory 1010 may provide computer-readable instructions to the core 1002 for execution. The cache memory 1010 may provide data for processing by the core 1002. In some embodiments, the computer-readable instructions may have been provided to the cache memory 1010 by a local memory, for example, local memory attached to the external bus 1016. The cache memory 1010 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.

[0171] The processor 1000 may include a controller 1014, which may control input to the processor 1000 from other processors and/or components included in a system and/or outputs from the processor 1000 to other processors and/or components included in the system. Controller 1014 may control the data paths in the ALU 1004, FPLU 1006 and/or DSPU 1008. Controller 1014 may be implemented as one or more state machines, data paths and/or dedicated control logic. The gates of controller 1014 may be implemented as standalone gates, FPGA, ASIC or any other suitable technology.

[0172] The registers 1012 and the cache 1010 may communicate with controller 1014 and core 1002 via internal connections 1020A, 1020B, 1020C and 1020D. Internal connections may be implemented as a bus, multiplexer, crossbar switch, and/or any other suitable connection technology.

[0173] Inputs and outputs for the processor 1000 may be provided via a bus 1016, which may include one or more conductive lines. The bus 1016 may be communicatively coupled to one or more components of processor 1000, for example the controller 1014, cache 1010, and/or register 1012. The bus 1016 may be coupled to one or more components of the system.

[0174] The bus 1016 may be coupled to one or more external memories. The external memories may include Read Only Memory (ROM) 1032. ROM 1032 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) 1033. RAM 1033 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) 1035. The external memory may include Flash memory 1034. The External memory may include a magnetic storage device such as disc 1036. In some embodiments, the external memories may be included in a system.

[0175] 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.

[0176] 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.

[0177] 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.

[0178] 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 apparatus for generating an output audio signal, the apparatus comprising:

a receiver (201) arranged to receive a data signal comprising audio data for at least a first audio signal and metadata including:

first acoustic environment data for an acoustic environment, the first acoustic environment data comprising repeated first sets of acoustic environment parameters, each first set of acoustic environment parameters providing a description of the acoustic environment;
second acoustic environment data for the acoustic environment, the second acoustic environment data comprising repeated second sets of acoustic environment parameters, each second set of acoustic environment parameters providing a description of the acoustic environment, a data size of the first sets of acoustic environment parameters exceeding a data size for the second sets of acoustic environment parameters and an update rate for the second sets of acoustic environment parameters being higher than an update rate for the first sets of acoustic environment parameters;
an acoustic data generator (203) being arranged to select between the first acoustic environment data and the second acoustic environment data to generate rendering acoustic environment data; and

a renderer (205) arranged to generate the audio output signal by rendering the audio signal based on the rendering acoustic environment data.

2. The audio apparatus of claim 1 wherein a quantization of at least one parameter of the second set of acoustic

environment parameters is coarser than a corresponding parameter of the first set of acoustic environment parameters.

3. The audio apparatus of claim 1 or 2 further comprising:

a listener pose processor (207) arranged to determine a listener pose; and wherein
the renderer (205) is arranged to render the audio signal in dependence on the listener pose.

4. The audio apparatus of any previous claim wherein data for each set of the first sets of acoustic environment parameters is distributed over a plurality of non-contiguous data segments.

5. The audio apparatus of claim 4 wherein at least one data segment for a first set of the first sets of acoustic environment parameters comprises an indication of a start position for data for the first set, and the acoustic data generator is arranged to generate the rendering acoustic environment data to represent the first set in dependence on the indication of the start position.

6. The audio apparatus of claim 4 or 5 wherein at least one data segment for a first set of the first sets of acoustic environment parameters comprises an indication of a data size for data for the first set, and the acoustic data generator is arranged to generate the rendering acoustic environment data to represent the first set in dependence on the indication of the data size.

7. The audio apparatus of claim 5 or 6 wherein the receiver (201) is arranged to store data from data segments as these are received, and the acoustic data generator (203) is arranged to generate the rendering acoustic environment data to represent the first set from stored data from the data segments.

8. The apparatus of any previous claim wherein first acoustic environment data for a given set of the first sets of acoustic environment parameters comprises a data integrity verification value, and the acoustic data generator (203) is arranged to generate the acoustic environment data from a previously received set of the first sets of acoustic environment parameters rather than from the given set if the data integrity verification value matches a data integrity verification value generated from data of the previously received set.

9. The apparatus of any previous claim wherein the second sets of acoustic environment parameters comprise fewer parameters than the first sets of acoustic environment parameters.

10. The apparatus of any previous claim wherein at least one set of the first sets of acoustic environment parameters comprise at least one parameter differentially encoded relative to a parameter of a set of the second sets of acoustic environment parameters.

11. A data signal comprising at least a first audio signal and metadata including:

first acoustic environment data for an acoustic environment, the first acoustic environment data comprising repeated first sets of acoustic environment parameters, each first set of parameters providing a description of the acoustic environment;
second acoustic environment data for the acoustic environment, the second acoustic environment data comprising repeated second sets of acoustic environment parameters, each second set of parameters providing a description of the acoustic environment, a data size of the first sets of parameters exceeding a data size for the second sets of parameters and an update rate for the second sets of parameters is higher than an update rate for the first sets of parameters.

12. An apparatus for generating the data signal of claim 11.

13. The apparatus of claim 12 further comprising:

a transmitter (901) arranged to transmit the data signal over a communication channel;
a determiner (903) arranged to determine a communication channel property for the communication channel; and
a controller (905) arranged to adapt a property of the first acoustic environment data in dependence on the communication channel property.

14. A method of generating an output audio signal, the method comprising:
receiving a data signal comprising audio data for at least a first audio signal and metadata including:

first acoustic environment data for an acoustic environment, the first acoustic environment data comprising repeated first sets of acoustic environment parameters, each first set of acoustic environment parameters providing a description of the acoustic environment;
second acoustic environment data for the acoustic environment, the second acoustic environment data comprising repeated second sets of acoustic environment parameters, each second set of acoustic environment parameters providing a description of the acoustic environment, a data size of the first sets of acoustic environment parameters exceeding a data size for the second sets of acoustic environment parameters and an update rate for the second sets of acoustic environment parameters is higher than an update rate for the first sets of acoustic environment parameters;
selecting between the first acoustic environment data and the second acoustic environment data to generate rendering acoustic environment data; and
generating the audio output signal by rendering the audio signal based on the rendering acoustic environment data.

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.

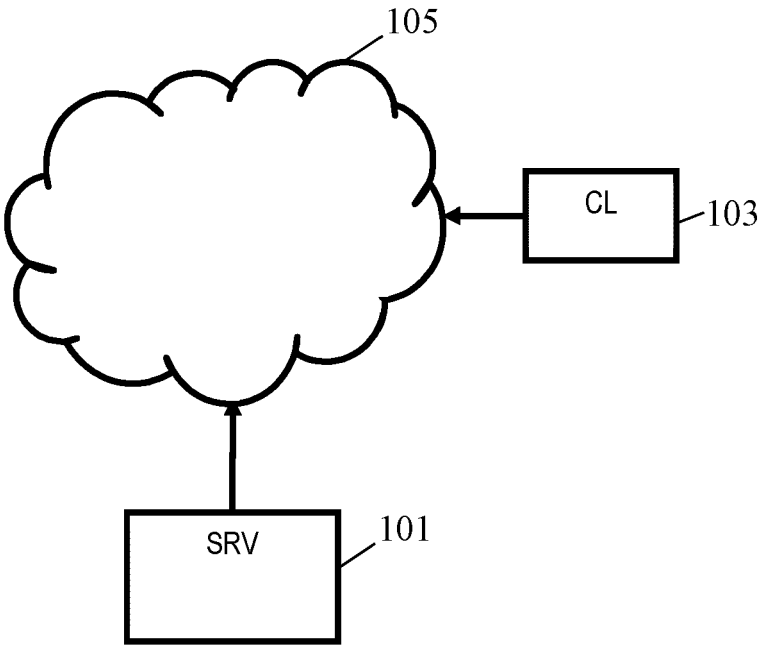


FIG. 1

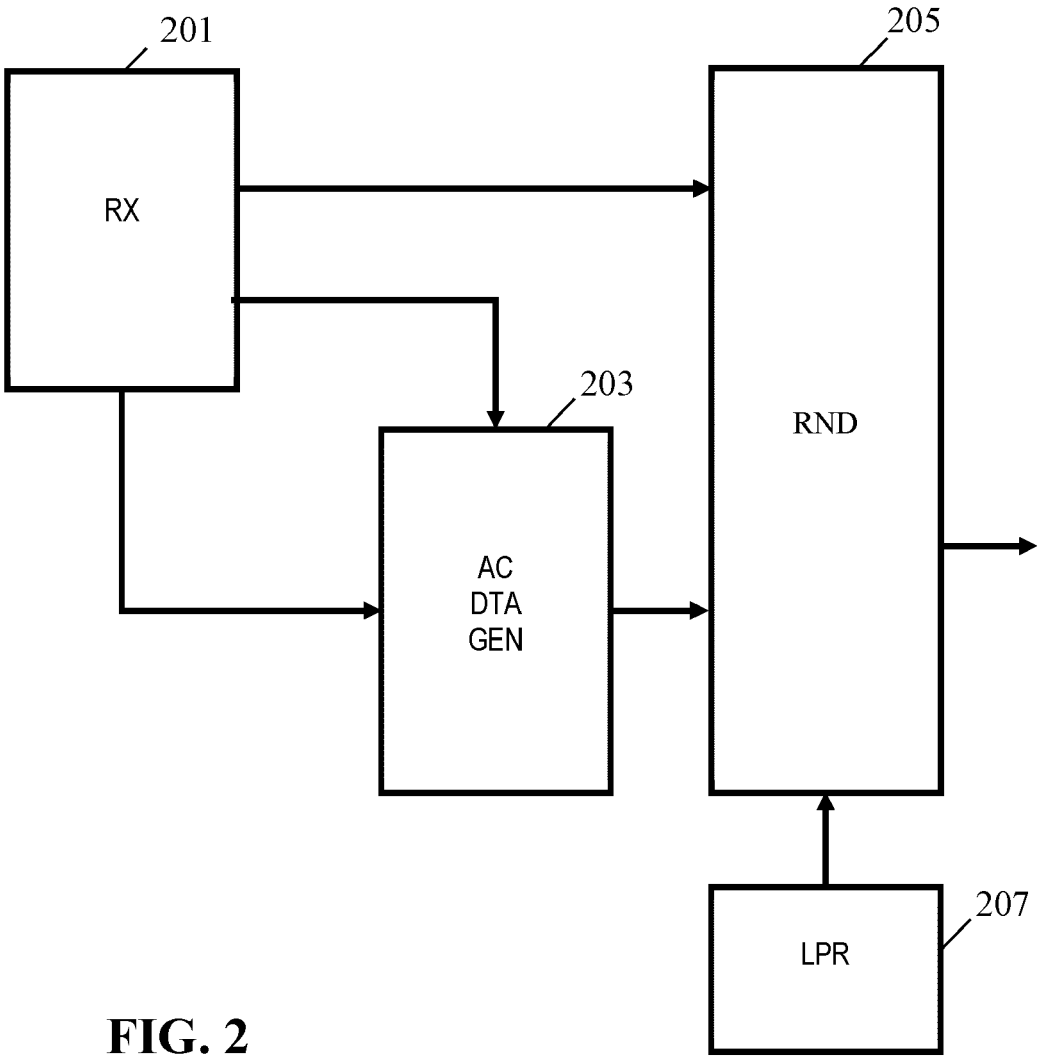


FIG. 2

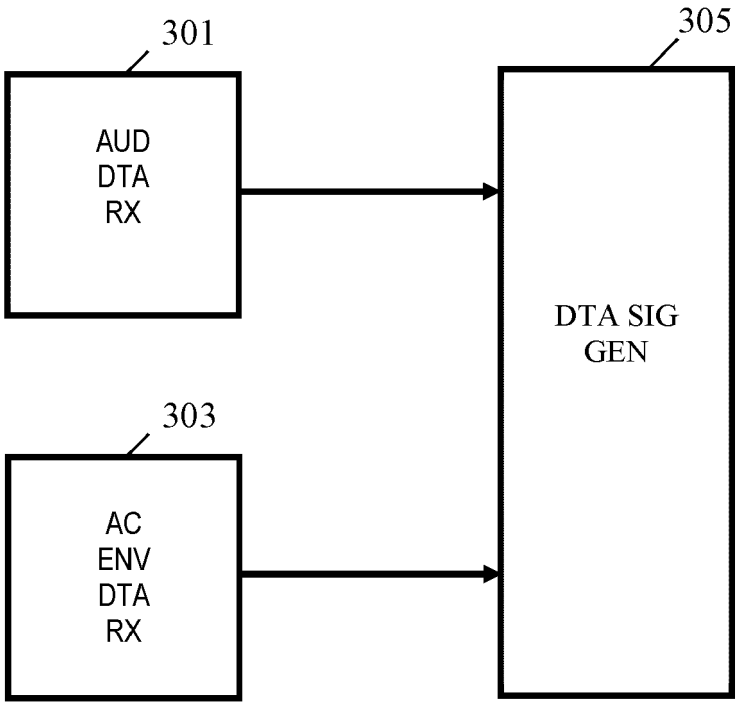


FIG. 3

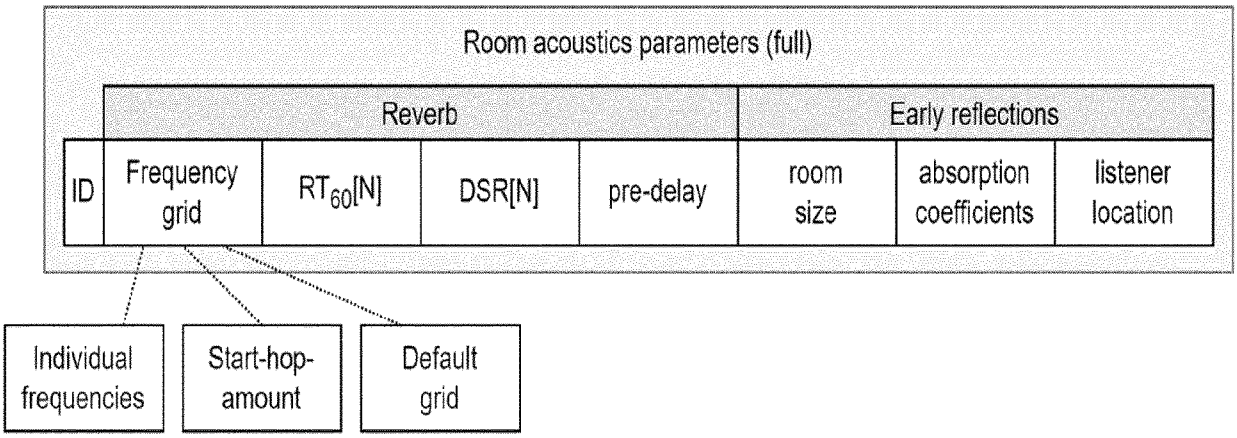
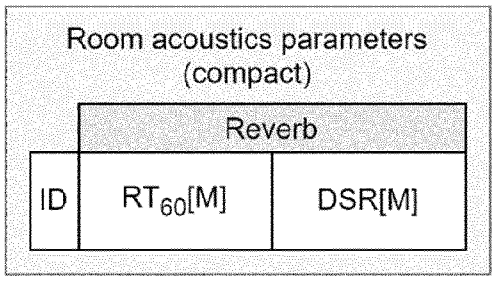

Room acoustics parameters (full)
Reverb
Early reflections
ID
Frequency grid
RT60[N]
DSR[N]
pre-delay
room size
absorption coefficients
listener location
Individual frequencies
Start-hop-amount
Default grid

FIG. 4


Room acoustics parameters
(compact)
Reverb
ID
RT60[M]
DSR[M]

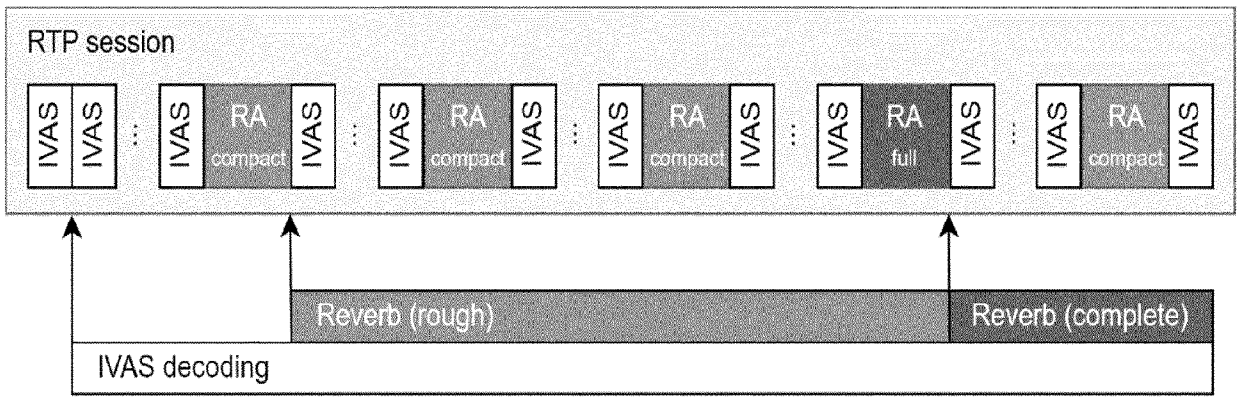

RTP session
IVAS
IVAS
IVAS
RA
compact
IVAS
IVAS
RA
compact
IVAS
IVAS
RA
compact
IVAS
IVAS
RA
full
IVAS
IVAS
RA
compact
IVAS
Reverb (rough)
Reverb (complete)
IVAS decoding

FIG. 5

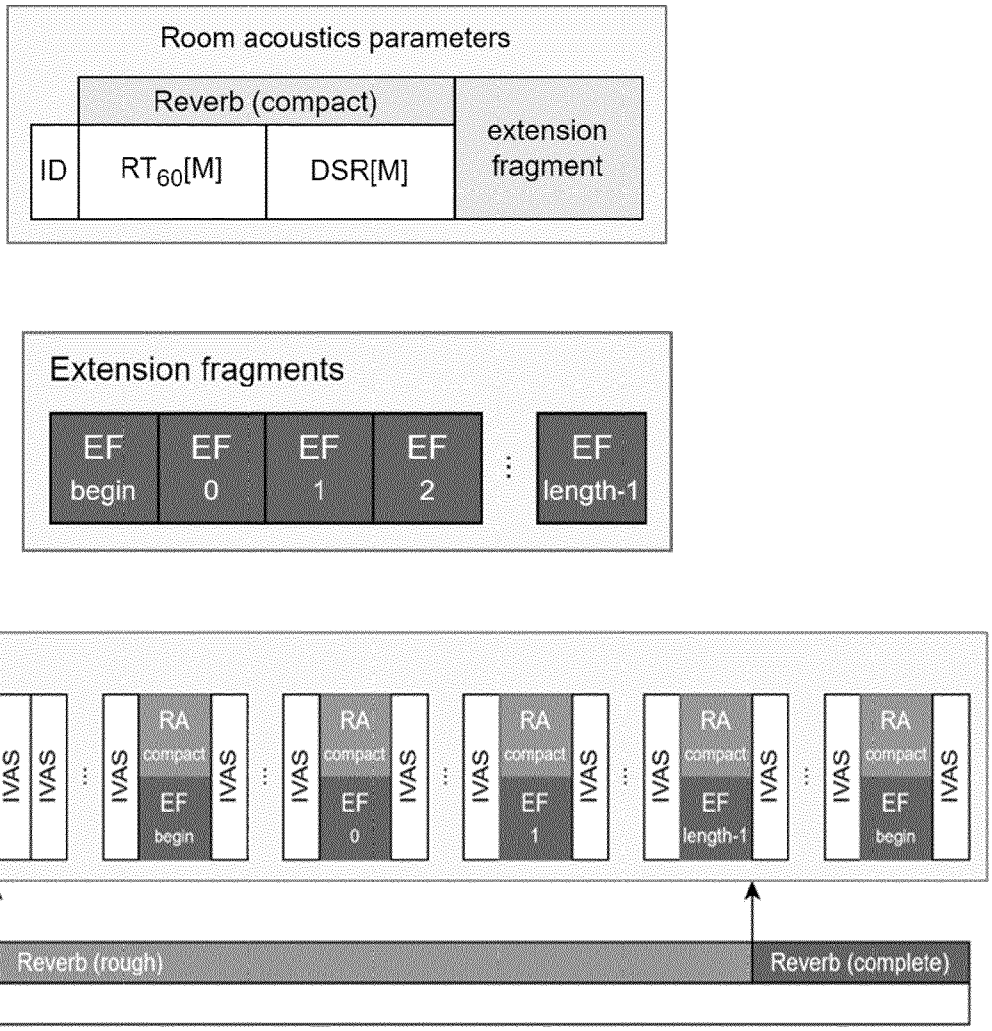

Room acoustics parameters
Reverb (compact)
ID
RT60[M]
DSR[M]
extension fragment
Extension fragments
EF begin
EF 0
EF 1
EF 2
EF length-1
RTP session
IVAS
RA compact
EF length-1
EF begin
EF 0
EF 1
Reverb (rough)
Reverb (complete)
IVAS decoding

FIG. 6

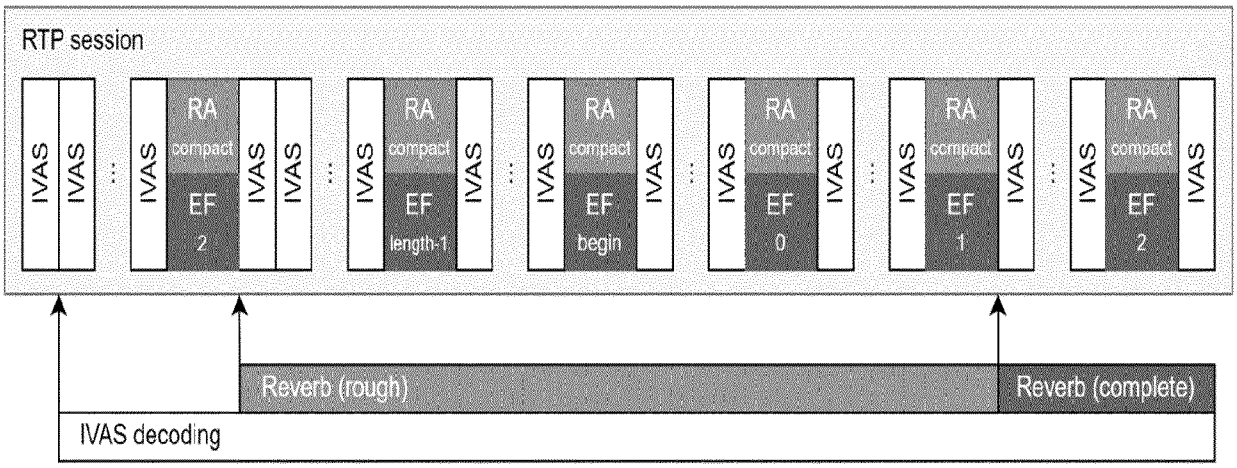

RTP session
IVAS
IVAS
...
IVAS
RA
compact
EF
2
IVAS
IVAS
...
IVAS
RA
compact
EF
length-1
IVAS
...
IVAS
RA
compact
EF
begin
IVAS
...
IVAS
RA
compact
EF
0
IVAS
...
IVAS
RA
compact
EF
1
IVAS
...
IVAS
RA
compact
EF
2
IVAS
Reverb (rough)
Reverb (complete)
IVAS decoding

FIG. 7

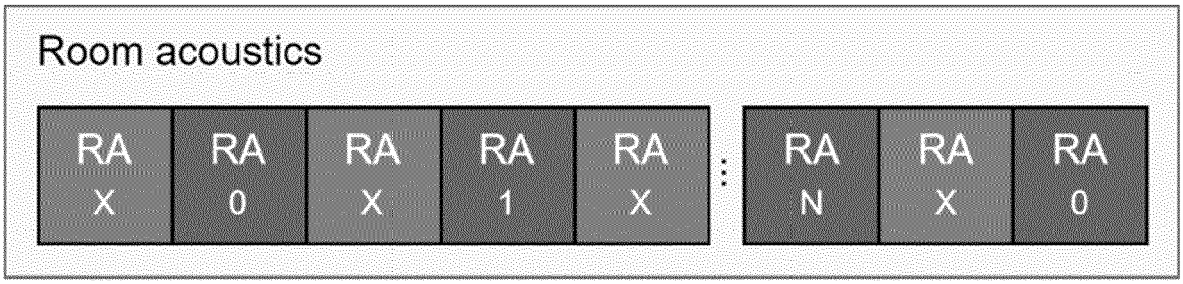


FIG. 8

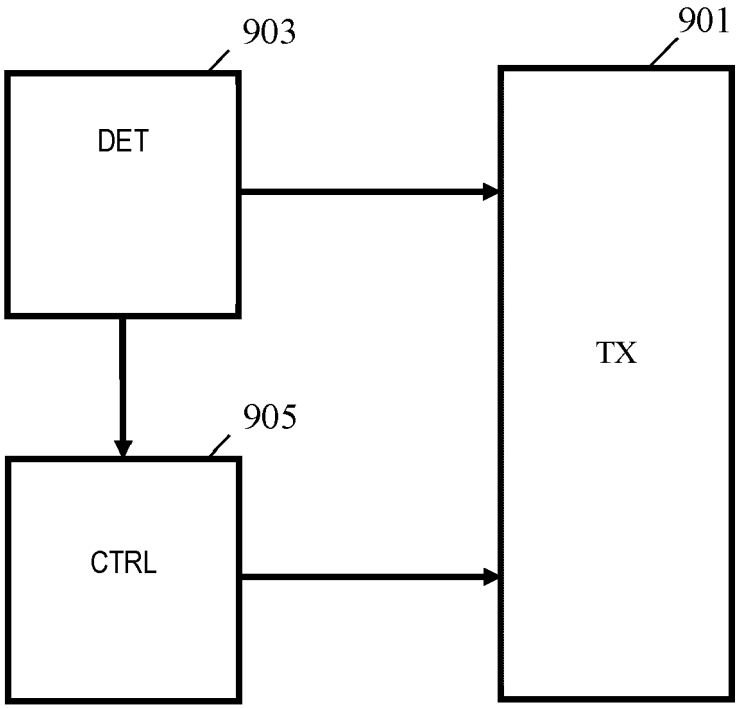


FIG. 9

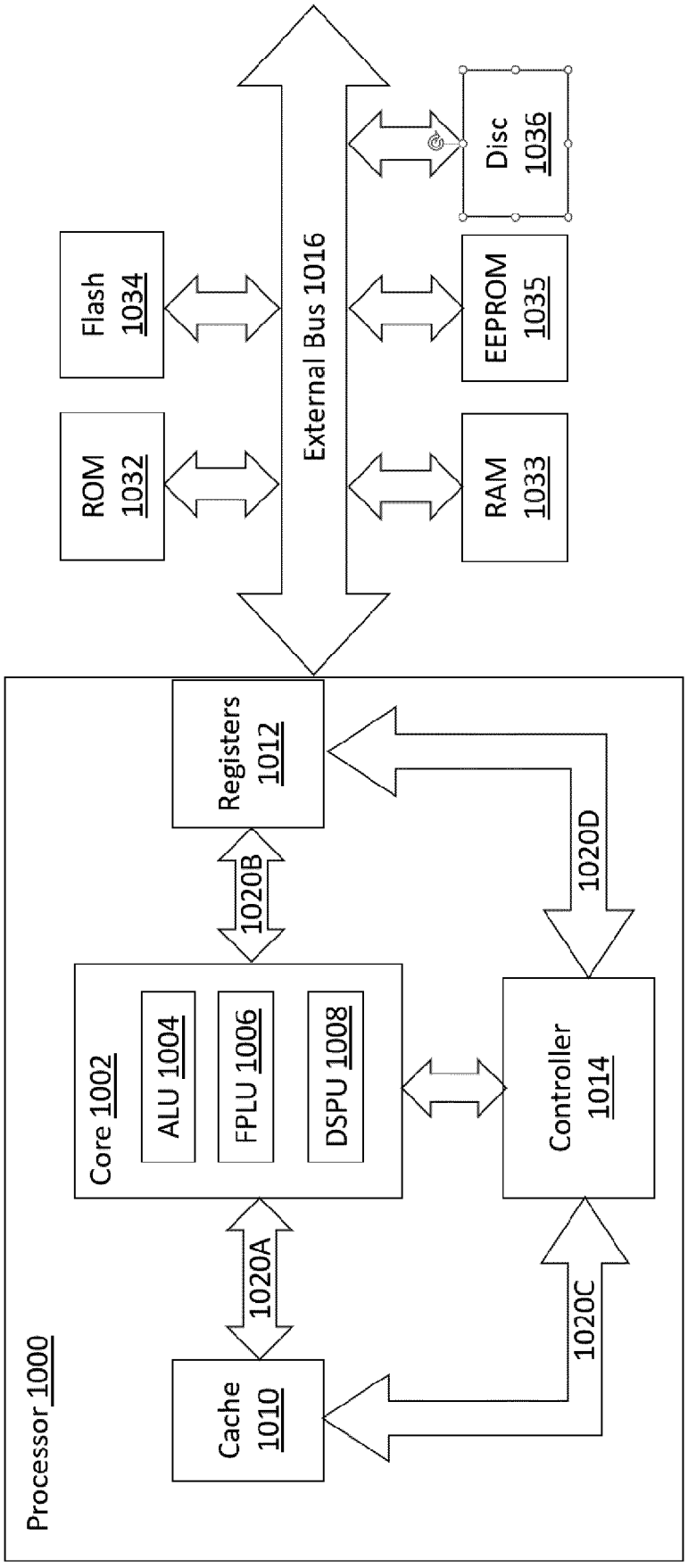

Processor 1000
Cache 1010
Core 1002
ALU 1004
FPLU 1006
DSPU 1008
Registers 1012
Controller 1014
1020A
1020B
1020C
1020D
External Bus 1016
ROM 1032
Flash 1034
RAM 1033
EEPROM 1035
Disc 1036

FIG. 10

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 23 20 3277

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X A	US 2021/287651 A1 (ERONEN ANTTI [FI] ET AL) 16 September 2021 (2021-09-16) * paragraphs [0083], [0084]; figures 2,3 * -----	12 1-11, 13-15	INV. H04S7/00
A	GB 2 616 424 A (NOKIA TECHNOLOGIES OY [FI]) 13 September 2023 (2023-09-13) * page 24, line 26 - page 27, line 8 * -----	1-15	
A	QUACKENBUSH SCHUYLER R ET AL: "MPEG Standards for Compressed Representation of Immersive Audio", PROCEEDINGS OF THE IEEE, IEEE. NEW YORK, US, vol. 109, no. 9, 28 May 2021 (2021-05-28), pages 1578-1589, XP011873330, ISSN: 0018-9219, DOI: 10.1109/JPROC.2021.3075390 [retrieved on 2021-08-18] * page 1584 - page 1585 * -----	1-15	
			TECHNICAL FIELDS SEARCHED (IPC) H04S G10L

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	11 March 2024	Righetti, Marco

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 20 3277

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.

11-03-2024

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2021287651 A1	16-09-2021	EP 4121957 A1	25-01-2023
		US 2021287651 A1	16-09-2021
		WO 2021186107 A1	23-09-2021
GB 2616424 A	13-09-2023	GB 2616424 A	13-09-2023
		WO 2023169819 A2	14-09-2023

EPO FORM P0459

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

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **JOT, J.-M.** ; **CHAIGNE, A.** Digital delay networks for designing artificial reverberators. *Proc. 90th Conv. Audio Eng. Soc.*, 1991 **[0112]**