(54) **AUDIO APPARATUS AND METHOD OF OPERATION THEREFOR**

(57)   An audio apparatus comprises a receiver (201) receiving an audio signal and a geometric model for an audio source object. The geometric model may be a polygon mesh. A position circuit (203) determines a set of point audio source positions for the audio source object based on the geometric model such that the point audio source positions are spatially distributed within the audio source object. A point audio signal generator (205) generates a point audio source signal for the point audio source positions from the audio signal. In some cases, the point audio signal may be a copy of the received audio signal. A data generator (207) generates a data representation of the audio source object comprising the set of point audio source positions and the at least one point audio source signal. A renderer (301) may render the audio source object by rendering point audio source signals from the point audio source positions.
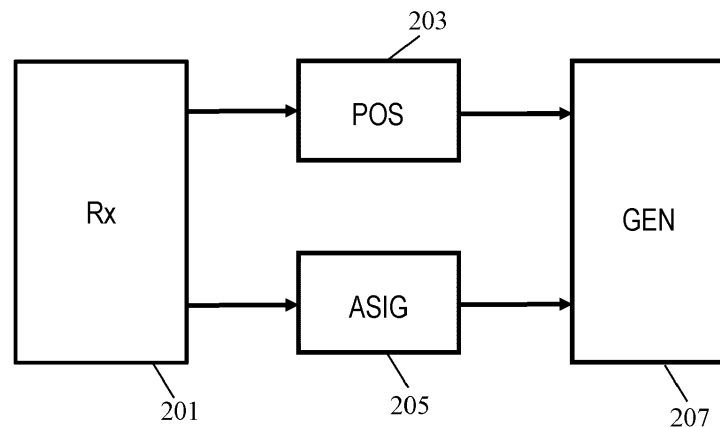
FIG. 2

Description

FIELD OF THE INVENTION

[0001] The invention relates to an apparatus and method of operation therefor, and in particular, but not exclusively, to an approach for generating a data representation of an audio source object in e.g. a Virtual Reality experience 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 often referred to as eXtended Reality (XR)), which are rapidly becoming mainstream, with a number of solutions being aimed at the consumer market. A number of standards are also under development by a number of standardization bodies. Such standardization activities are actively developing standards for the various aspects of VR/AR/MR systems including e.g. streaming, broadcasting, rendering, etc.

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

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

[0006] Such a feature can specifically allow a virtual reality experience to be provided to a user. This may allow the user to (relatively) freely move about in a virtual 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.

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

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

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

[0010] An important feature for many applications is that of how to generate and/or distribute audio that can provide a natural and realistic perception of the audio environment.

[0011] A particular challenge is to represent audio sources that are not limited to a single point source, i.e. which has a spatial acoustic extension/dimension.

[0012] In audio rendering the situation often occurs where one or more audio signals are meant to represent an object with large physical properties or a diffuse sound source. In traditional listening environments such as cinemas or home theatres, this is typically achieved by first converting the input signal into multiple output signals with varying levels of decorrelation, and then feeding those signals to individual loudspeakers or headphone channels so that they produce the perception of acoustic width, as e.g. described in US9654895B2. Varying the level of correlation between the differing signals can affect the size of the object as perceived by the listener. This method works well in the controlled listening environments where the listener position, the sound transducers, and the simulated object position are all known and controlled for.

[0013] However, in XR applications, users may have free movement in all dimensions, typically referred to as 6 Degrees of Freedom (6DoF). With 6DoF the user is

able to move freely during playback of the content, or in gaming during runtime of the application, and as such the content creator and any encoding algorithms do not know what the listening position may be at any moment in time, and as such where they are relative to the sound producing objects.

[0014] In a typical 6DoF, environment an audio object with a given location is typically rendered to the listener by first calculating the objects relative distance from the user and the direction from the listener to the object (e. g. as azimuth and elevation). For headphone applications, the audio signal associated with the object is then convolved with the matching Head Related Impulse Response (HRIR) or Head Related Transfer Function (HRTF). The resulting (stereo) signal is presented to the listener via headphones with the corresponding distance related time delay and level attenuation. Using this method will accurately represent an object as a point source, with all sound emanating from a single point in space.

[0015] Often sound should not emanate from a zero-dimensional point in space, rather it may emanate from a 1-dimensional (line), 2-dimensional (plane), or 3-dimensional (solid) audio object that should radiate energy uniformly from all points. When an object is used in this way, it is in the field referred to as the extent of the audio source. An extent audio source has a spatial extension. An extent audio source is a non- single point audio source. When describing a virtual environment an object may be described by using a simple geometric object (Line, Plane, Box, Sphere, Cone, Cylinder etc) or often by definition of a mesh consisting of vertices and faces. In particular, for more complex objects and environments, the spatial data is often represented by complex mesh structures which may comprise a large number of polygons and vertices, edges, and faces.

[0016] The relative size of the extent of the audio source with respect to the listener is constantly changing as the listener moves within the 6DoF environment, and as such the way in which it is rendered needs to constantly be adapted. As the sound source changes with respect to the user position, either through user motion or animation of the sound source, the perceived width of the source also needs to be adapted.

[0017] In order to represent such effects, the existing rendering methods require the calculation of the perceived source width, adapting the correlation and potentially other parameters of the audio signals associated with the source and any metadata describing the source, and applying these new correlation levels and parameters to the audio signals. This requires the audio rendering technology to have detailed information relating to the object representing the acoustic extent, and to calculate relative perceived widths in real time.

[0018] However, such processing and calculations tend to very complex and resource demanding and typically the rendering approaches will tend to be compromised on complexity, resource demands and/or the resulting audio quality (and specifically on the spatial per-

ception). For example, for a complex object described by a mesh representation, this may involve many thousands of vertices and faces in the description of the object which results in very resource demanding calculations being required to adapt the rendering to provide an acceptable audio perception.

[0019] Hence, an improved approach for rendering audio would be advantageous. In particular, an approach that allows improved operation, increased flexibility, reduced complexity, facilitated implementation, an improved audio experience, improved audio quality, reduced computational burden, improved suitability for varying positions, improved performance for virtual/mixed/ augmented/ extended reality applications, improved perceptual cues for spatial audio, increased and/or facilitated adaptability, increased processing flexibility, improved and/or facilitated rendering of the spatial extent of audio source and/or improved performance and/or operation would be advantageous.

SUMMARY OF THE INVENTION

[0020] Accordingly, the Invention seeks to preferably mitigate, alleviate or eliminate one or more of the above mentioned disadvantages singly or in any combination.

[0021] According to an aspect of the invention there is provided an audio apparatus comprising: a receiver arranged to receive an audio signal and a geometric model for an audio source object; a position circuit arranged to determine a set of point audio source positions for the audio source object in response to the geometric model, the point audio source positions being spatially distributed within the audio source object; a point audio signal generator arranged to generate at least one point audio source signal for the point audio source positions from the audio signal; and a data generator arranged to generate a data representation of the audio source object comprising the set of point audio source positions and the at least one point audio source signal.

[0022] The invention may allow improved and/or facilitated rendering of audio for audio source object having a spatial extent. The invention may in many embodiments and scenarios generate a more naturally perceived acoustic spatial extent of an audio source object. The approach may in many scenarios reduce computational resource requirements and usage substantially. In many embodiments, the approach may obviate the need for evaluating a geometric model of the audio source object for changing listening positions.

[0023] The approach may typically provide improved and/or reduced complexity processing in applications where listening positions may change dynamically, such as e.g. in many XR applications.

[0024] The generated data representation may be independent of the listening position. Rendering of audio based on the data representation may typically not require any processing /evaluation/ adaptation of a complex geometric model, such as e.g. a mesh model.

[0025] The rendering of audio output signals for the audio source object based on the generated data representation may often require reduced complexity and reduced computational resource requirements in comparison to a rendering of the same quality based directly on the audio signal and the geometric model.

[0026] The at least one point audio signal may for a given point audio source position represent audio of a point audio source located at the point audio source position.

[0027] The audio source object may be an audio source of a scene or environment. The scene or environment may be a virtual or real scene. The audio source object may represent audio for a virtual or real (scene) object.

[0028] The point audio source positions may be spread out in the audio source object, or in at least one region of the audio source object. The position circuit 203 may generate spatially distributed point audio source positions by determining the point audio source positions to have a minimum distance to a closest neighbor point audio source position. The minimum distance may be predetermined or may be determined in response to a spatial property of the audio source object.

[0029] In accordance with an optional feature of the invention, the position circuit is arranged to determine the point audio source positions such that a distance from any position of the set of point audio source positions to a surface of the audio source object does not exceed a distance threshold.

[0030] This may provide improved performance and/or reduced complexity/ resource demand. In many scenarios, it may allow fewer point audio source positions to be required for a given perceived spatial sound quality.

[0031] The surface may be a boundary or edge of the audio source object. The distance threshold may be a predetermined distance, or may be dependent on a spatial property of the audio source object, such as a size or maximum dimension of the audio source object.

[0032] In some embodiments, the position circuit may be arranged to determine the point audio source positions such that a distance from any position of the set of point audio source positions to a nearest point on a surface of the audio source object does not exceed a distance threshold.

[0033] In accordance with an optional feature of the invention, the position circuit is arranged to determine a set of intersect positions of a grid and to select the set of point audio source positions from the intersect positions.

[0034] This may typically provide a low complexity yet high quality determination of point audio source positions. It may in many scenarios allow an improved spatial perception of the audio source object.

[0035] In accordance with an optional feature of the invention, the position circuit is arranged to determine the set of point audio source positions to satisfy at least one requirement of: a requirement that a maximum distance between each position of the set of point audio source positions and a nearest position of the set of point audio source positions is less than a first distance threshold; a requirement that a minimum distance between each position of the set of point audio source positions and a nearest position of the set of point audio source positions is more than a second distance threshold; a requirement that a number of points of the set of point audio source positions does not exceed a first number; a requirement that a number of points of the set of point audio source positions is not below a second number; and a requirement that a maximum distance from any point of a surface of the audio source object to a nearest point of the set of point audio source positions is less than a third distance threshold.

[0036] This may provide improved performance and/or reduced complexity/ resource demand. It may allow point audio source positions to be determined which may ensure that a sufficient perception of the spatial extent of the audio source object is provided.

[0037] In accordance with an optional feature of the invention, the point audio signal generator is arranged to determine at least one audio level for the set of point audio source positions in response to an audio level for audio source object and a number of positions in the set of point audio source positions.

[0038] This may allow an improved and typically more realistic spatial perception of the audio source object. It may reduce perceived distortion to the audio source object. The audio level for audio source object may be a desired/ target audio level for the audio source object.

[0039] In accordance with an optional feature of the invention, the point audio signal generator is arranged to determine the at least one audio level in response to positions in the set of point audio source positions.

[0040] This may allow an improved and typically more realistic spatial perception of the audio source object. It may reduce perceived distortion to the audio source object. The audio level for audio source object may be a desired/ target audio level for the audio source object.

[0041] In accordance with an optional feature of the invention, the data generator is arranged to generate the data representation to include a relative render priority for the set of point audio source positions.

[0042] An improved and typically more flexible operation can be achieved. For example, it may allow improved adaptation to resource availability of different renderers. It may in many scenarios assist in reducing the perceived impact of limited rendering resource. The relative render priority for one point audio source position may indicate a rendering priority relative to other point audio source positions.

[0043] In accordance with an optional feature of the invention, the data generator is arranged to generate the data representation to include directional sound propagation data for at least one position of the set of point audio source positions.

[0044] This may allow improved and/or more flexible rendering of the audio source object.

[0045] In accordance with an optional feature of the invention, the position circuit is arranged to determine the set of point audio source positions in response to a two dimensional extent of the audio source object when viewed from a given region relative to a position of the audio source object.

[0046] This may allow improved and/or facilitated operation in many embodiments.

[0047] In accordance with an optional feature of the invention, the position circuit is arranged to determine a number of dimensions for which the audio source object has an extent exceeding a threshold, and to generate the set of point audio source positions as a structure having the number of dimensions.

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

[0049] In accordance with an optional feature of the invention, the data generator is arranged to generate the data representation to include an indication of a propagation time parameter for the set of point audio source positions.

[0050] This may allow increased perceived audio quality in many scenarios. It may in particular in many scenarios allow a more consistent perception of the audio source object.

[0051] In accordance with an optional feature of the invention, the audio apparatus comprises a renderer arranged to render the audio source object by rendering the at least one point audio source signal from the set of point audio source positions.

[0052] The approach may in many scenarios allow an improved and/or facilitated rendering of an audio source object having a spatial extent.

[0053] In accordance with an optional feature of the invention, the audio apparatus further comprises an encoder for generating an encoded bitstream comprising the data representation of the audio source object.

[0054] The approach may in many scenarios allow an improved encoded bitstream representing an audio source object having a spatial extent to be generated.

[0055] According to an aspect of the invention there is provided an audio apparatus comprising: a receiver for receiving an audio bitstream comprising a data representation of an audio source object having a spatial extent, the data representation comprising: a set of point audio source positions being distributed within the audio source object, and at least one point audio source signal for the point audio source positions; and a renderer arranged to render the audio source object by rendering the at least one point audio source signal from the set of point audio source positions.

[0056] According to an aspect of the invention there is provided a method of operation for an audio apparatus, the method comprising: receiving an audio signal and a geometric model for an audio source object; determining a set of point audio source positions for the audio source object in response to the geometric model, the point audio source positions being spatially distributed within the au-

dio source object; generating at least one point audio source signal for the point audio source positions from the audio signal; and generating a data representation of the audio source object comprising the set of point audio source positions and the at least one point audio source signal.

[0057] According to an aspect of the invention there is provided a method of operation for an audio apparatus, the method comprising: receiving an audio bitstream comprising a data representation of an audio source object having a spatial extent, the data representation comprising: a set of point audio source positions being distributed within the audio source object, and at least one point audio source signal for the point audio source positions; and rendering the audio source object by rendering the at least one point audio source signal from the set of point audio source positions.

[0058] According to an aspect of the invention there is provided an audio bitstream comprising a data representation of an audio source object having a spatial extent; the data representation comprising: a set of point audio source positions being distributed within the audio source object; and at least one point audio source signal for the point audio source positions from the audio signal.

[0059] 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

[0060] 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 elements of an eXtended Reality system;
FIG. 2 illustrates an example of an audio apparatus in accordance with some embodiments of the invention;
FIG. 3 illustrates an example of an encoder audio apparatus accordance with some embodiments of the invention;
FIG. 4 illustrates an example of a renderer audio apparatus accordance with some embodiments of the invention;
FIG. 5 illustrates an example an audio object and point audio source signals;
FIG. 6 illustrates an example an audio object and point audio source signals; and
FIG. 7 illustrates an example an audio object and point audio source signals.

DETAILED DESCRIPTION OF SOME EMBODIMENTS OF THE INVENTION

[0061] The following description will focus on audio processing and rendering for an eXtended Reality (XR)

application, such as for a Virtual Reality (VR), Augmented Reality (AR), or Mixed Reality (MR) application. The described approach will focus on such applications where audio rendering is adapted to reflect acoustic variations and changes in the audio perception as a (possibly virtual) user / listener position changes. However, it will be appreciated that the described principles and concepts may be used in many other applications and embodiments.

[0062]   Semi or fully virtual experiences allowing a user to move around in a (possibly partially) virtual world are becoming increasingly popular and services are being developed to satisfy such a demand.

[0063]   In some systems, the XR application may be provided locally to a viewer by e.g. a stand-alone device that does not use, or even have any access to, any remote XR data or processing. For example, a device such as a games console may comprise a store for storing the scene data, input for receiving/ generating the viewer pose, and a processor for generating the corresponding images from the scene data.

[0064]   In other systems, the XR application may be implemented and performed remote from the viewer. For example, a device local to the user may detect/ receive movement/ pose data which is transmitted to a remote device that processes the data to generate the viewer pose. The remote device may then generate suitable view images and corresponding audio signals for the user pose based on scene data describing the scene. The view images and corresponding audio signals are then transmitted to the device local to the viewer where they are presented. For example, the remote device may directly generate a video stream (typically a stereo/ 3D video stream) and corresponding audio stream which is directly presented by the local device. Thus, in such an example, the local device may not perform any XR processing except for transmitting movement data and presenting received video data.

[0065]   In many systems, the functionality may be distributed across a local device and remote device. For example, the local device may process received input and sensor data to generate user poses that are continuously transmitted to the remote XR device. The remote XR device may then generate the corresponding view images and corresponding audio signals and transmit these to the local device for presentation. In other systems, the remote XR device may not directly generate the view images and corresponding audio signals but may select relevant scene data and transmit this to the local device, which may then generate the view images and corresponding audio signals that are presented. For example, the remote XR device may identify the closest capture point and extract the corresponding scene data (e.g. a set of object sources and their position metadata) and transmit this to the local device. The local device may then process the received scene data to generate the images and audio signals for the specific, current user pose. The user pose will typically correspond to the head pose, and references to the user pose may typically equivalently be considered to correspond to the references to the head pose.

[0066]   In many applications, especially for broadcast services, a source may transmit or stream scene data in the form of an image (including video) and audio representation of the scene which is independent of the user pose. For example, signals and metadata corresponding to audio sources within the confines of a certain virtual room may be transmitted or streamed to a plurality of clients. The individual clients may then locally synthesize audio signals corresponding to the current user pose. Similarly, the source may transmit a general description of the audio environment including describing audio sources in the environment and acoustic characteristics of the environment. An audio representation may then be generated locally and presented to the user, for example using binaural rendering and processing.

[0067]   FIG. 1 illustrates such an example of a XR system in which a remote XR client device 101 liaises with a XR server 103 e.g. via a network 105, such as the Internet. The server 103 may be arranged to simultaneously support a potentially large number of client devices 101.

[0068]   The XR server 103 may for example support a broadcast experience by transmitting an image signal comprising an image representation in the form of image data that can be used by the client devices to locally synthesize view images corresponding to the appropriate user poses (a pose refers to a position and/or orientation). Similarly, the XR server 103 may transmit an audio representation of the scene allowing the audio to be locally synthesized for the user poses. Specifically, as the user moves around in the virtual environment, the image and audio synthesized and presented to the user is updated to reflect the current (virtual) position and orientation of the user in the (virtual) environment.

[0069]   In many applications, such as that of FIG. 1, it may thus be desirable to model a scene and generate an efficient image and audio representation that can be efficiently included in a data signal that can then be transmitted or streamed to various devices which can locally synthesize views and audio for different poses than the capture poses.

[0070]   The audio representation may include an audio representation for a plurality of different sound sources in the environment. This may include some sound sources corresponding to general diffuse and non-localized sound, such as ambient and backgrounds sounds for the environment as a whole. It will typically also comprise a number of audio representations for point size audio sources to be rendered from specific positions in the scene. However, in addition, the sound representation may include a number of audio sources that have a spatial extent (which in many cases may equivalently be referred to as having spatial extension), and which should preferably be rendered such that a user is provided with a perception of the spatial extent and perceived dimen-

sion (often width) of the sound source.

[0071] Typically, such extent audio sources are represented by a geometric model that describes the spatial properties of the audio source object. Thus, audio data together with data describing a geometric model may provide a representation of an extent audio source object.

[0072] The geometric model may typically be a model that may also be used to describe visual properties of the object corresponding to the audio source object and the same geometric model and data may be used to represent spatial properties of the object for both audio and visual properties and rendering.

[0073] An often used approach is to represent objects, and the scene in general, is a mesh. Specifically, a polygon mesh may be a collection of vertices, edges and faces that defines the shape of a polyhedral object. For visual properties, texture data, color, brightness, and possibly other visual properties may be provided for the polygons. The visual rendering of such an object typically involves processing the mesh and applying the visual properties for the given viewing pose as is well known in the art.

[0074] When rendering audio for an audio source object having a spatial extent, a renderer may process the geometric model to determine spatial characteristics that are then used to adapt properties of the rendered audio such as diffusion, correlation etc. However, such a determination of spatial properties and the associated signal properties tend to be very complex and resource demanding. For example, evaluating complex mesh models to determine how the perceived spatial width of an object varies as the user moves may require millions of operations and calculations, which may require a high computational resource. Indeed, in many cases it may even require dedicated hardware in order to allow real time processing and adaptation.

[0075] In the following, an approach will be described which may generate and use a representation of audio source objects with spatial extent that in many situations, scenarios, and applications may provide improved and/or facilitated operation. It may for example, reduce complexity and may provide an audio representation that does not require the complex processing of polygon meshes or other geometric models.

[0076] FIG. 2 illustrates an audio apparatus that can generate a data representation of an audio source object which has a spatial extent. The audio apparatus may for example be part of the server 103 or the client 101 of FIG. 1.

[0077] The audio apparatus comprises a receiver 201 which is arranged to receive at least one audio signal for an audio source object. In addition, the receiver receives a geometric model which provides a description of the spatial properties of the audio source object.

[0078] The geometric model may specifically be a polygon mesh description of the audio source object. Such models are frequently used in the field to represent spatial properties for objects in a real or virtual environment. It

may typically allow accurate representations and is frequently used in e.g. computer vision. In other embodiments, other geometric models may be used. For example, the geometric model may be defined as a simple 3D object such as a cuboid, sphere, cylinder, etc.

[0079] The audio signal for the audio source object may be provided as an encoded audio data signal describing the audio to be produced by the audio source object. In some cases, only a single audio signal may be provided to represent the audio to be produced by the audio source object However, in some scenarios, more than one audio signal may be provided for one audio source object. For example, an audio signal may be provided to represent audio originating from one part (e.g. the left part) of the audio source object and a second audio signal may be provided to represent audio originating from a different part (e.g. the right part) of the audio source object. Such audio signals may possibly be closely correlated but differ in some aspects, such as e.g. by having different levels, frequency spectra (filtering), include different signal components etc. In some cases, the audio signals may be provided as completely different audio signals and in other cases they may be provided as a first signal and data describing the difference of the second audio signal such as a filter or level adjustment to be applied to the first signal.

[0080] The audio apparatus further comprises a position circuit 203 which is arranged to determine a set of point audio source positions for the audio source object. The position circuit determines the point audio sources as spatially distributed positions within the audio source object.

[0081] The audio apparatus further comprises a point audio signal generator 205 which is arranged to generate at least one point audio source signal for the single point audio positions. The point audio signal generator 205 may in some embodiments, simply generate the point audio source signal as the first audio signal, i.e. in some embodiments and scenarios, the point audio signal may directly be the same as the audio signal received and representing the audio of the audio source object. In other embodiments, the point audio signal generator 205 may be arranged to generate the point audio source signal by processing the received first audio signal, such as for example by applying a filtering or level adaptation. Each point audio source position of the set of point audio source positions indicates a position for a single point audio source of a set of single point audio sources which together form an audio source producing the audio source object.

[0082] The audio apparatus also comprises a data generator 207 which is coupled to the position circuit 203 and the point audio signal generator 205. The data generator 207 is arranged to generate a data representation of the audio source object comprising the set of point audio source positions and the (at least one) point audio source signal. Thus, the audio apparatus may generate a new representation of the audio source object with this

being represented by a plurality of spatially distributed point audio source positions and associated point audio source signal(s). Thus, the audio apparatus of FIG. 2 may receive a representation of a spatial extent audio source object given by an audio signal representing the audio of the audio source object and a geometric model for the audio source object spatial extent. It may from this generate a different representation of the spatially extended audio source object as a plurality of spatially distributed point audio source positions and associate point audio source signal(s). The audio apparatus may generate a new representation of the spatially extent audio source object where the spatial extent is represented by a plurality of point audio sources. Rather than representing the audio source object as a single distributed audio source, the approach may represent the audio source object as a plurality of point audio sources (each point audio source position corresponding to a position of a point audio source position of this plurality of audio sources).

[0083] A rendering of the audio source object may be based on this modified data representation and specifically the rendering may be performed by rendering a point source audio signal from each point audio source position based on the point audio source signal. Thus, the rendering may for each point audio source position generate one rendered audio signal corresponding to a point audio source being positioned at the point audio source position. The individual rendering for each audio source position may thus be performed without considering any spatial properties of the audio source object (except for the point audio source position). The individual rendering audio signals for the different point audio source positions may then be combined such that the combined signal of the different positions represent the audio source object.

[0084] Although such an approach does not generate individual signals that reflect the spatial extent of the audio source object, but rather generates multiple audio signals that each represent a point audio source, the Inventor has realized that this in practice tends to provide spatial cues that provide a perception of the spatial extent of the audio source object. Indeed, it has been found that the perceived spatial properties may often very closely reflect those of the audio source object.

[0085] The approach may also in many cases provide a computationally efficient process of representing spatial extents of audio objects. In particular, it may require only point source audio rendering and may obviate the necessity for evaluating a complex geometric model with respect to the current listening pose. Rather, the geometric model may in many embodiments only be evaluated when generating the point audio source positions for the new representation of the audio source object. Such an evaluation is typically performed only once as it is independent of the listening position whereas a traditional approach requires the model to be evaluated for each new listening position. Thus, whereas some additional complexity may sometimes be required for the ren-

dering of multiple point source audio signals rather than when rendering fewer audio signals representing distributed audio sources, the overall computational reduction is typically very significant. This is especially the case for complex geometric models and shapes, such as specifically mesh representations of complex shapes.

[0086] Further, in many embodiments, the generation of the modified data representation may be performed once for multiple rendering devices and operations, such as for example by a central server serving multiple clients.

[0087] In many embodiments, the approach may provide improved audio quality. For example, a reduced complexity may allow more resource to be allocated to more accurate rendering (including potentially of other audio sources in the environment). Further, in itself, the perception provided by multiple point sources may often be considered more accurate than that which is achieved by traditional approaches.

[0088] In the approach an audio source object to be rendered as emanating from a geometric extent object may be rendered with the spatial extension of the audio source object being represented using a plurality of individual audio point sources.

[0089] In some embodiments, the audio apparatus of FIG. 2 may for example be included in a rendering device such as for example the client 101 of FIG. 1. For example, the server 203 may generate an XR audio visual data stream which includes a mesh model for various objects in the environment/ scene. At least one of these objects may correspond to an audio object with spatial extent for which the audio visual data stream further comprises audio data.

[0090] The XR audio visual data stream may be transmitted to the client 101 which in the example includes the audio apparatus 200 of FIG. 2 as illustrated in the example of FIG. 3.

[0091] The audio apparatus 200 may perform the described operation to generate a modified representation of the spatial extent audio source object by a plurality of distributed point audio sources and associated point audio source data.

[0092] The client 101 may include a renderer 301 which is arranged to render audio for the scene. Specifically, the renderer 301 is arranged to include functionality for spatially rendering audio from point sources. It will be appreciated that many algorithms and operations for such rendering is known to the skilled person, including for example HRTF and BRIR based rendering for headphones, and that these for brevity will not be described further herein.

[0093] The renderer 301 may specifically receive a listener pose and for that pose render a point source audio signal from each point audio source generated for the audio source object. Thus, rather than the audio for the spatial extent signal being rendered as a single diffuse signal, it is rendered as a plurality of point source audio signals from different positions within the audio source object. The audio signal rendered for each point audio

source signal may be rendered independently of any other audio signal rendering for any other point audio source position. The rendering of one point source audio signal may thus be based only on the point source audio signal associated with the point audio source position and on the point audio source position. Each point audio signal may be rendered independently of other point audio source positions and point audio signals. The generated point audio signals may then be combined for rendering to the user e.g. via loudspeakers or headphones.

[0094] Thus, each rendering may be performed separately and taking into account only the point audio source position and the listening position/ pose. No consideration of the geometric extent of the audio source object is needed for the rendering. Thus, although the approach requires multiple audio signals to be rendered from different positions for each audio source object, each rendering may use well-known and efficient rendering algorithms and approaches. Further, no evaluation of a geometric model is required for each listening position.

[0095] Not only may such an approach provide a very efficient audio processing and rendering but it has also been found that it provides highly advantageous and realistic spatial perception of the audio in many scenarios and embodiments. It may typically provide very accurate perception of the spatial extent of an audio source as perceived from the listening position.

[0096] The approach may represent an extent object as a number of discrete points, and then the standard object rendering method may be used to give the perception of source width without the need for decorrelation of the audio signals or for calculations of perceived width and decorrelation metrics, or for the storage of object surface descriptions.

[0097] In some embodiments, the audio apparatus may be part of an audio visual bitstream generator, such as for example an encoder or server. For example, the audio apparatus 200 may in some embodiments be part of the server 103 of FIG. 1.

[0098] In such a case the server 103 may receive or generate an audio visual data that includes a mesh model for various objects in an environment/ scene. At least one of these objects may correspond to a spatial extent audio object for which audio data is also provided/ generated.

[0099] Such data may be received by the audio apparatus 200 of the server 103 which may include elements as illustrated in FIG. 4. The audio apparatus of the server 103 may proceed to perform the described operations to generate a representation of the audio object using a plurality of point audio source positions and associated audio signal(s). This representation may then be provided to an encoding unit 401 arranged to generate an encoded bitstream comprising the data representation of the audio source object.

[0100] The encoder may encode the point audio source positions and point audio signals in any suitable form, and it will be appreciated that many approaches for encoding audio visual signals are known and appro-priate. For example, the audio signals may be encoded using a known audio encoding algorithm and format, and the point audio source positions may be encoded as metadata associated with the individual point audio source signal.

[0101] Thus, the encoder, and in the specific example the server 103, may generate a audio bitstream that comprises a data representation of an audio source object having a spatial extent with the data representation comprising: a set of point audio source positions being distributed within the audio source object, and at least one point audio source signal for the single point audio positions from the first audio signal.

[0102] An advantage of the approach is in many embodiments that the point source based representation can be generated without any consideration of the specific listening position, i.e. the representation can be listening position independent.

[0103] Different approaches, principles, rules, and requirements may be used to generate the spatially distributed point audio source positions in different embodiments and scenarios.

[0104] In many embodiments, the position circuit 203 may be arranged to determine a set of intersect positions of a grid and to select the set of point audio source positions from the intersect positions. The grid may in many embodiments be a regular grid but may in some embodiments be an irregular or unstructured grid (such as e.g. a Ruppert's grid).

[0105] A regular grid may be a tessellation of n-dimensional space by parallelotopes and an irregular grid may be based on other shapes than parallelotopes. The intersect points may be corners/ vertices of the parallelotopes/ shapes. The intersect positions may be positions where lines or curves describing or defining the tessellation intersect. The grid may specifically be a Cartesian grid. In many embodiments, the grid may be a one, two, or three dimensional grid of Euclidian space.

[0106] In many embodiments, the point audio source positions may be determined as equidistant positions. The distance from a point audio source position to a nearest neighbor point audio source position may in some embodiments be constant for different point audio source positions.

[0107] As an example, in some embodiments, the position circuit 203 may align a predetermined regular grid with equidistant positions/ intersection points to the audio source object, and then evaluate the geometric model to identify the positions/ intersections that fall within the audio source object. These positions may then be selected as the point audio source positions for the audio source object. The alignment between the audio source object and the grid may in some embodiments be in accordance with a suitable algorithm or criteria (e.g. a reference position of the grid is aligned with e.g. a lowest/ highest etc. point of the audio source object; a reference position of the grid is positioned within the object, such as e.g. at a center). In most embodiments, an arbitrary alignment

may be applied between the object and the grid (this may in particular be useful for embodiments where the distance between grid positions is much smaller than the spatial extent of the audio source object).

[0108]   An example of such an approach is illustrated in FIG. 5. In the example, the audio source object is an elongated object in which only one row/ column of positions falls within the object. FIG. 6 illustrates an example for an object that has a significant extent in two directions but not in a third direction.

[0109]   In some embodiments, the position circuit 203 may be arranged to determine a number of dimensions for which the audio source object has a perceived extent exceeding a threshold. For example, for the audio source object of FIG. 5, the position circuit 203 may determine that the audio source object only has a significant extent in only one direction, and thus that it is essentially a one dimensional object. For the object of FIG. 6, the position circuit 203 may determine that it is an object with significant spatial extent in two directions and thus essentially is a two dimensional object.

[0110]   The position circuit 203 may then generate the set of point audio source positions as a structure having the same number of dimensions as determined by the position circuit 203. For example, a grid having the same number of dimensions as determined by the position circuit 203 may be used to determine the point audio source positions.

[0111]   As a specific example, the audio apparatus may proceed through the following steps

    1. Given any n-dimensional geometric object the audio apparatus may first check whether the dimensions are over a given threshold for rendering, reducing the object to the minimum number of dimensions required for rendering.
    2. If the resulting object is 0-dimensional (i.e. a point source) then no further processing is required.
    3. For a 1-dimensional (line) object, the line is subdivided into a number of discrete points, and each point is stored as a set of coordinates (e.g. corresponding to the example FIG. 5). For 2 or 3 dimensional objects each dimension is subdivided into a number of discrete points and a grid of points is created to cover the object uniformly (e.g. corresponding to the example FIG. 6).
    4. All points are checked to ensure that they exist within or on the object surface and any points outside of the object are discarded.
    5. The points are stored as a set of metadata defining the position and the corresponding audio signal to be associated with them, as well as possibly an audio presentation level reduction based on the number of additional sources that have been added.
    6. The metadata is presented to the audio rendering technology which renders each object as a discrete point source, providing a perception of acoustic width.

[0112]   In some embodiments, the position circuit 203 is arranged to determine the point audio source positions such that a distance from any of these positions to a boundary or surface of the audio source object is less than a distance. Thus, it may be required that a position is only included if it is sufficiently close to the surface of the object.

[0113]   In such an example, the audio apparatus may check that the point audio source positions are located within a prescribed distance of the surface of the audio source object, and it may remove any positions that are in the center of the object.

[0114]   This may for some objects substantially reduce the required processing for the rendering as the number of point audio sources that are rendered is reduced. However, in many scenarios, the spatial perception, and in particular the perception of the extent of the object, will not be substantially affected by the internal audio sources being excluded. Thus, an improved complexity and resource demand versus perceived audio quality can be achieved.

[0115]   In many embodiments, the position circuit 203 may be arranged to determine the point audio source positions to satisfy one or more requirements.

[0116]   For example, in some embodiments, point audio source positions may be determined as random positions, and these may then be evaluated to see if they meet a set of one or more requirements. For example, an iterative operation may be performed where a new point audio source position is randomly generated at each iteration. The random point audio source position is then evaluated in accordance with the set of requirements, and if all are met, the point audio source position is stored as part of the data representation for the audio source object, and otherwise it is discarded. The process may continue until e.g. a given number of point audio source positions have been determined (with the number potentially being determined based on a spatial property of the audio source object, such as based on a size or volume of the audio source object).

[0117]   The position circuit 203 may be arranged to determine the set of point audio source positions to satisfy one, more or all of the following:

- A requirement that a maximum distance between each position of the set of point audio source positions and a nearest position of the set of point audio source positions is less than a first distance. The position circuit 203 may determine the audio source positions such that the maximum distance between at least two points (and possibly of two points that meet the other requirements) is less than a given threshold. Thus, a given density of points may be ensured and a more homogenous perception of the entire audio source object can often be achieved.

- A requirement that a minimum distance between each position of the set of point audio source positions and a nearest position of the set of point audio

source positions is more than a second distance. The position circuit 203 may determine the audio source positions such that the minimum distance between two neighboring points (and possibly of two neighboring points that meet the other requirements) is more than a given threshold. Thus, the position circuit 203 may determine the point audio source positions such that the points are not too close together, and thus such that they may represent a larger region with fewer point audio source positions.

• A requirement that a number of points of the set of point audio source positions does not exceed a given number. This may limit the total number of point audio source positions and thus may ensure that the resource requirement for the rendering is maintained sufficiently low (with the number potentially being determined based on a spatial property of the audio source object, such as based on a size or volume of the audio source object).

• A requirement that a number of points of the set of point audio source positions does not fall below a given number (with the number potentially being determined based on a spatial property of the audio source object, such as based on a size or volume of the audio source object).

• A requirement that a maximum distance from any point of a surface of the audio source object to a nearest point of the set of point audio source positions is less than a distance. As previously described, the point audio source positions may be determined to be close to the surface of the audio source object.

[0118]   In some embodiments, the placement of point audio source positions is not controlled using a regular pattern, but instead the point audio source positions may be randomly generated within a certain set of rules, for example but not limited to, a minimum distance between individual sources, a maximum number of sources, or a minimum distance to the surface of the extent.

[0119]   In some embodiments only a single audio signal is received for the audio source object. A point audio source signal for the point audio source positions may then be generated from this received audio signal and indeed in some embodiments the received audio signal may be used directly as the point audio source signal.

[0120]   In other embodiments, some processing may be included such as for example a filtering or level adjustment.

[0121]   In some embodiments, a plurality of point audio source signal may be generated for the point audio source positions for the audio source object. For example, slightly different audio signals may be generated for different parts of the audio source object (e.g. depending on how close they are to the surface or on which surface is the closest). The multiple point audio source signals may be generated from a single received audio signal or in some cases multiple point audio source signals may be generated from received multiple audio signals. When multiple point audio source signals are generated from an audio source object, each point audio source position is typically assigned one of the generated point audio source signals.

[0122]   The point source audio signals may specifically be a duplication of the input audio signal, possibly with a level adjustment. If there is only one input audio signal, then all point audio source positions will typically be assigned the same point audio signal generator 205. If there are multiple input audio signals with associated metadata to indicate from which part of the audio source object they should be rendered, then the point audio source signal for the individual point audio source position may be determined based on audio signal that most closely matches the point audio source position.

[0123]   In many embodiments, the level of the point audio source signals may typically be adapted. This adaptation may be included to compensate for the audio of the audio source object being represented by multiple audio sources. The audio levels of the point audio source signals may thus specifically depend on the number of point audio source positions and on their position. For example, in order to render the audio source object with a given audio level using a plurality of point audio sources, the level for each point audio source may be adjusted such that the combined effect of all point audio sources combine to provide the desired audio level.

[0124]   In some embodiments, the point audio signal generator 205 may be arranged to determine at least one audio level for the set of point audio source positions in response to an audio level of audio source object and a number of positions in the set of point audio source positions. The level may further be determined in response to positions in the set of point audio source positions.

[0125]   As a specific example, the original sound source level for the audio source object may be defined by a reference distance, with this being the distance at which the level of the audio signal is known either as an absolute reproduction level or a gain level that should be applied at that distance from the object.

[0126]   In such an example, a number of measurement positions may be determined at the reference distance from the surface of the audio source object (see FIG. 7), and the audio level for the point audio source positions/point audio source signals may be adjusted such that when all point audio source signals are rendered, the level received across all of the measurement positions is equal to that which would be received were a single object rendered at the nearest point on the audio source object surface to the measurement position. Such a determination may be performed using an automated optimization function or other estimation paradigm.

[0127]   In some embodiments, the data generator 207 may be arranged to generate the data representation to include a relative render priority for the set of point audio source positions.

[0128]   For example, for each point audio source position, metadata may be generated which indicates a rel-

ative priority of the point audio source position relative to other point audio source positions. The priority could for example be provided as a ranking of all of the point audio source positions in order of importance for the rendering in accordance with any suitable criterion. As another example, the point audio source positions may be allocated a rendering priority from a given set of possible rendering priorities. For example, each point audio source position may be indicated as being mandatory, preferred, or optional.

[0129] The renderer may in such embodiments select a set of point audio source positions from which to render audio signals in response to the relative render priority. The rendering for the audio source object may then be by rendering of the selected set of point audio source positions. The renderer may in some scenarios select a subset of point audio source positions to render based on the rendering priorities.

[0130] For example, if the renderer has limited resources and is only able to render a given number of audio signals, it may, if more point audio source positions are received than can be rendered, proceed to select a subset based on the rendering priorities. In this way, the rendering can be controlled to typically provide an improved audio experience.

[0131] For example, the rendering priority may be determined based on spatial relationships between the point audio source positions and possibly relative to spatial properties of the audio source object.

[0132] For example, all point audio source positions that are closest to a part of the surface of the audio source object may be indicated to be mandatory or have the highest rendering priority. Subsequently, a preferred rendering priority may be given to point audio source positions that have a given minimum distance to the point audio source positions that were given the highest rendering priority, and a given minimum distance to other point audio source positions that are given a preferred rendering priority. Finally, all other point audio source positions may be assigned an optional rendering priority.

[0133] A constrained renderer may in this case select which point audio source positions to render by first selecting all point audio source positions that are indicated to be mandatory, then point audio source positions that are indicated to be preferred, and finally if sufficient resource is still available may select point audio source positions that are indicated to be optional. If only some point audio source positions of a given category can be selected (e.g. due to resource constraints), a suitable selection criterion may be used (including e.g. merely randomly selecting point audio source positions within a given category, or e.g. selecting point audio source positions to have the maximum distance between them).

[0134] Such an approach may provide improved spatial audio perception for constrained rendering by e.g. ensuring that point audio source positions contributing most significantly to the perception of the spatial extent are rendered.

[0135] In some embodiments the priority indication may be dependent on a distance between the audio source object/ point audio source position and a listening position.

[0136] Such a priority may be used to reduce the number of point audio source positions that are rendered based on the user-source distance, i.e. the distance between the listening position and the audio source object/point audio source position. As the listening position moves further from the audio source object, the perceived relative width of the audio source object decreases and as such the number of points needed to give an adequate perceived width also decreases. Such an approach may allow a reduction in rendering complexity with increasing source distance. The renderer may for example determine an audio source object or point audio source position to listening position distance and then select the point audio source positions for rendering in response to this distance and the relative distance dependent rendering priorities.

[0137] In some embodiments, the data generator 207 may be arranged to generate the data representation to include directional sound propagation data for at least one position of the set of point audio source positions. The rendering may then render the audio signal from that point audio source position in response to the directional sound propagation data for the position.

[0138] In some embodiments, directional response data may be added to the metadata for each of the additional point audio source positions such that e.g. the reproduction when the listener is external to the audio source object may be controlled separately to the internal reproduction. Such a directional response may be described as a polar pattern, or some other representation known to the rendering technology.

[0139] For example, a gain as a function of direction may be provided for a point audio source position and when rendering a point audio signal from that point audio source position the gain of the rendered signal may be determined in response to the gain provided for the direction from the point audio source position to the listening position.

[0140] In some embodiments, the audio apparatus may be arranged to generate the point audio source positions based on a spatial relationship between the audio source object and a reference viewing/ listening region. The reference viewing/ listening region may specifically be a region to which the listener is constrained, or e.g. within which it is likely that the listener is positioned.

[0141] In same embodiments, the position circuit 203 may be arranged to determine the set of point audio source positions in response to a two dimensional extent of the audio source object when viewed from a given region relative to a position of the audio source object. In same embodiments, the position circuit 203 may be arranged to determine the set of point audio source positions in response to a representation of the audio source object when viewed from a given listening region relative

to a position of the audio source object.

[0142] As an example, in an application the listener may be limited to a region of a virtual environment, and the audio apparatus may be aware of this region. The position circuit 203 may determine the maximum visible angle of the audio source object with respect to one or more positions within the listeners region and it may only consider the dimensions where the viewable angle exceeds a given threshold. This viewable angle can also inform the audio apparatus of the number of point audio source positions that should be used to provide the desired perceived source width.

[0143] In some embodiments, the data generator 207 may further be arranged to include an indication of a propagation time parameter for the set of point audio source positions.

[0144] The renderer may be arranged to render audio signals from the point audio source positions in response to the propagation time parameter. Specifically, in many embodiments, the renderer may be arranged to render the audio signals from all point audio source positions with the same propagation time, and with this propagation time being determined from the propagation time parameter.

[0145] Thus, although different point audio source positions will have different path lengths to the listening position, the audio sources are generated to have the same propagation time. The audio signals may be generated with other spatial cues reflecting the different positions of the point audio source positions. For example, differential delays between a right ear signal and a left ear signal (for HRTF processing) may reflect the actual position within the audio source object etc. Indeed, such an approach has been found to provide improved spatial perception in many scenarios and to in particular provide a perception of a cohesive audio source object with spatial extent.

[0146] In some embodiments, the distance dependent delay that is used by the renderer to simulate the time taken for the audio to propagate from the audio source object to the listener may be controlled by additional metadata linked to point audio source positions. The metadata may indicate that the time of flight for all sources should be considered to be the same as e.g. that of the nearest source to the listener or some other metric.

[0147] An extent of an audio source may be a spatial extension of an audio source. An extent audio source may be an audio source having a spatial extent. An extent audio source may be a spatially extended audio source. An audio source having an extent may be referred to as an audio source having a spatial extension.

[0148] 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.

[0149] 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.

[0150] 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.

[0151] Furthermore, although individually listed, a plurality of means, elements, circuits or method steps may be implemented by e.g. a single circuit, unit or processor. Additionally, although individual features may be included in different claims, these may possibly be advantageously combined, and the inclusion in different claims does not imply that a combination of features is not feasible and/or advantageous. Also, the inclusion of a feature in one category of claims does not imply a limitation to this category but rather indicates that the feature is equally applicable to other claim categories as appropriate. Furthermore, the order of features in the claims do not imply any specific order in which the features must be worked and in particular the order of individual steps in a method claim does not imply that the steps must be performed in this order. Rather, the steps may be performed in any suitable order. In addition, singular references do not exclude a plurality. Thus references to "a", "an", "first", "second" etc. do not preclude a plurality. Reference signs in the claims are provided merely as a clarifying example shall not be construed as limiting the scope of the claims in any way.

Claims

1. An audio apparatus comprising:

a receiver (201) arranged to receive an audio signal and a geometric model for an audio source object;

a position circuit (203) arranged to determine a set of point audio source positions for the audio source object in response to the geometric model, the point audio source positions being spatially distributed within the audio source object;

a point audio signal generator (205) arranged to generate at least one point audio source signal for the point audio source positions from the audio signal; and

a data generator (207) arranged to generate a data representation of the audio source object comprising the set of point audio source positions and the at least one point audio source signal.

2. The audio apparatus of claim 1 wherein the position circuit (203) is arranged to determine the point audio source positions such that a distance from any position of the set of point audio source positions to a surface of the audio source object does not exceed a distance threshold.

3. The audio apparatus of any previous claim wherein the position circuit (203) is arranged to determine a set of intersect positions of a grid and to select the set of point audio source positions from the intersect positions.

4. The audio apparatus of any previous claim wherein the position circuit (203) is arranged to determine the set of point audio source positions to satisfy at least one requirement of:

a requirement that a maximum distance between each position of the set of point audio source positions and a nearest position of the set of point audio source positions is less than a first distance threshold;

a requirement that a minimum distance between each position of the set of point audio source positions and a nearest position of the set of point audio source positions is more than a second distance threshold;

a requirement that a number of points of the set of point audio source positions does not exceed a first number;

a requirement that a number of points of the set of point audio source positions is not below a second number; and

a requirement that a maximum distance from any point of a surface of the audio source object to a nearest point of the set of point audio source positions is less than a third distance threshold.

5. The audio apparatus of any previous claim wherein

the point audio signal generator (205) is arranged to determine at least one audio level for the set of point audio source positions in response to an audio level for audio source object and a number of positions in the set of point audio source positions.

6. The audio apparatus of claim 5 wherein the point audio signal generator (205) is arranged to determine the at least one audio level in response to positions in the set of point audio source positions.

7. The audio apparatus of any previous claim wherein the data generator (209) is arranged to generate the data representation to include a relative render priority for the set of point audio source positions.

8. The audio apparatus of any previous claim wherein the data generator (209) is arranged to generate the data representation to include directional sound propagation data for at least one position of the set of point audio source positions.

9. The audio apparatus of any previous claim wherein the position circuit (203) is arranged to determine the set of point audio source positions in response to a two dimensional extent of the audio source object when viewed from a given region relative to a position of the audio source object.

10. The audio apparatus of any previous claim wherein the position circuit (203) is arranged to determine a number of dimensions for which the audio source object has an extent exceeding a threshold, and to generate the set of point audio source positions as a structure having the number of dimensions.

11. The audio apparatus of any previous claim wherein the data generator (207) is arranged to generate the data representation to include an indication of a propagation time parameter for the set of point audio source positions.

12. The audio apparatus of any previous claim further comprising a renderer (301) arranged to render the audio source object by rendering the at least one point audio source signal from the set of point audio source positions.

13. The audio apparatus of any previous claim further comprising an encoder (401) for generating an encoded bitstream comprising the data representation of the audio source object.

14. An audio apparatus comprising:

a receiver for receiving an audio bitstream comprising a data representation of an audio source object having a spatial extent, the data repre-

sentation comprising:

a set of point audio source positions being distributed within the audio source object, and

at least one point audio source signal for the point audio source positions; and

a renderer arranged to render the audio source object by rendering the at least one point audio source signal from the set of point audio source positions.

15. A method of operation for an audio apparatus, the method comprising

receiving an audio signal and a geometric model for an audio source object;
determining a set of point audio source positions for the audio source object in response to the geometric model, the point audio source positions being spatially distributed within the audio source object;
generating at least one point audio source signal for the point audio source positions from the audio signal; and
generating a data representation of the audio source object comprising the set of point audio source positions and the at least one point audio source signal.

16. A method of operation for an audio apparatus, the method comprising:

receiving an audio bitstream comprising a data representation of an audio source object having a spatial extent, the data representation comprising:

a set of point audio source positions being distributed within the audio source object, and
at least one point audio source signal for the point audio source positions; and

rendering the audio source object by rendering the at least one point audio source signal from the set of point audio source positions.

17. A computer program product comprising computer program code means adapted to perform all the steps of claim 15 or 16 when said program is run on a computer.

18. An audio bitstream comprising a data representation of an audio source object having a spatial extent; the data representation comprising
a set of point audio source positions being distributed

within the audio source object; and at least one point audio source signal for the point audio source positions from the audio signal.
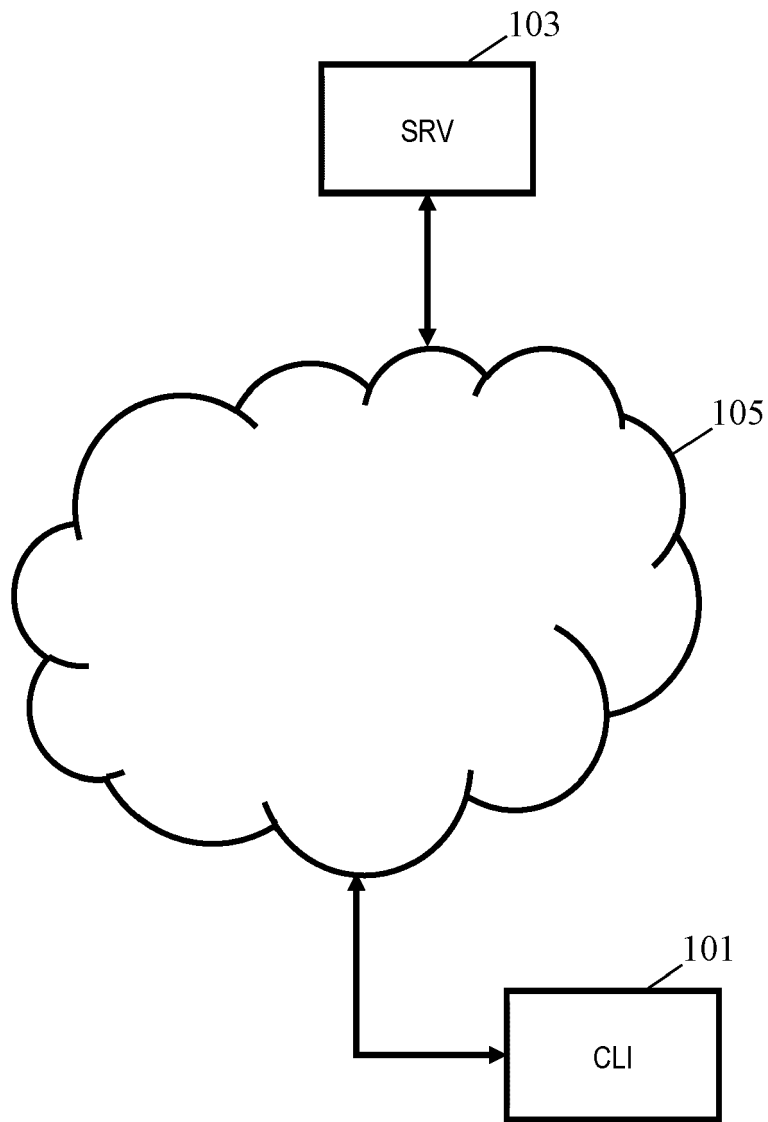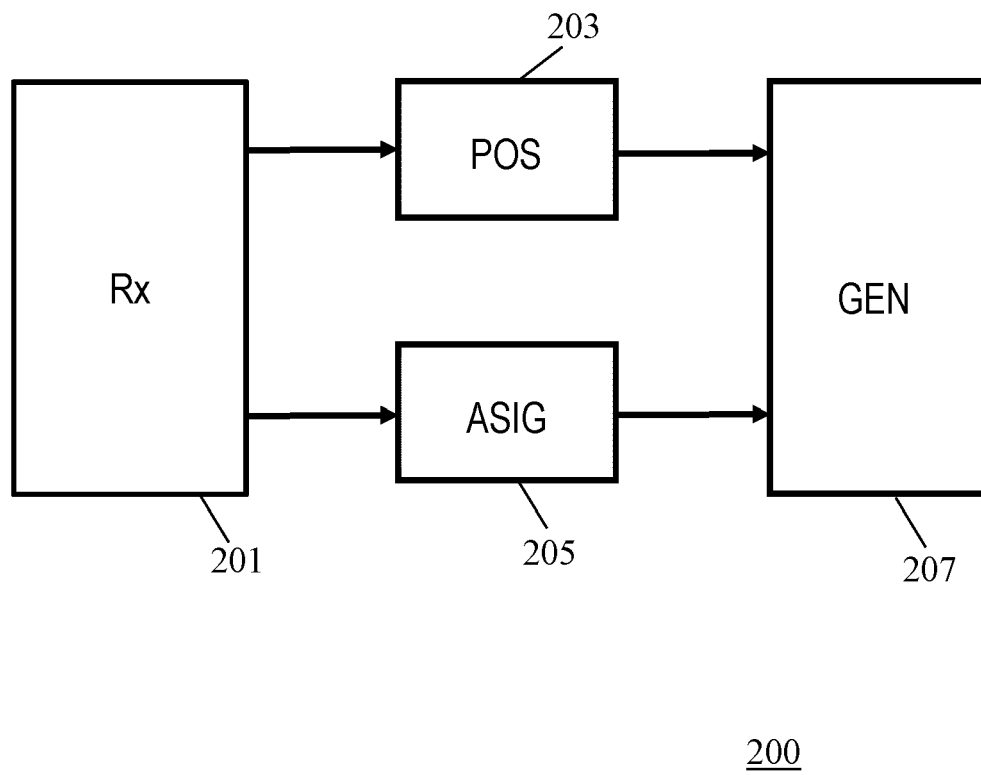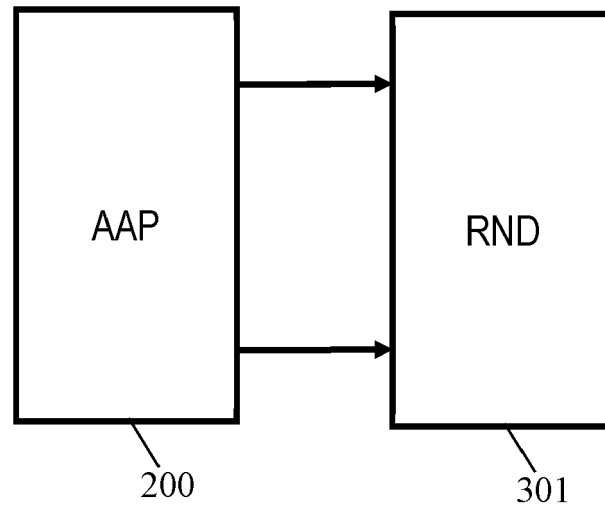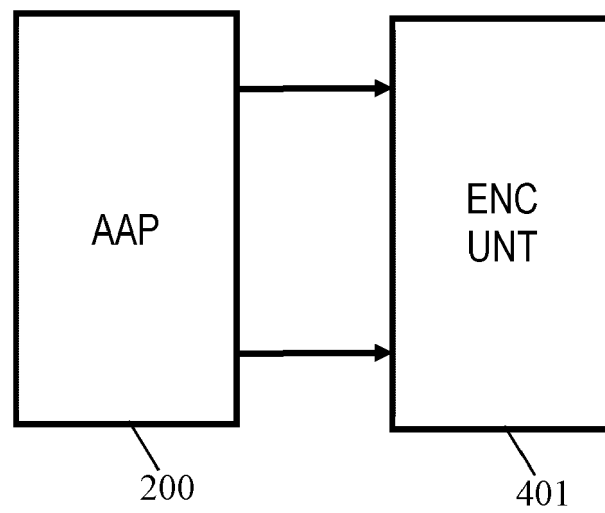
103

SRV

105

101

CLI

FIG. 1

203

Rx

POS

ASIG

GEN

201

205

207

200

FIG. 2

AAP

RND

200

301

FIG. 3

FIG. 4

FIG. 5

FIG. 6

FIG. 7

Europäisches
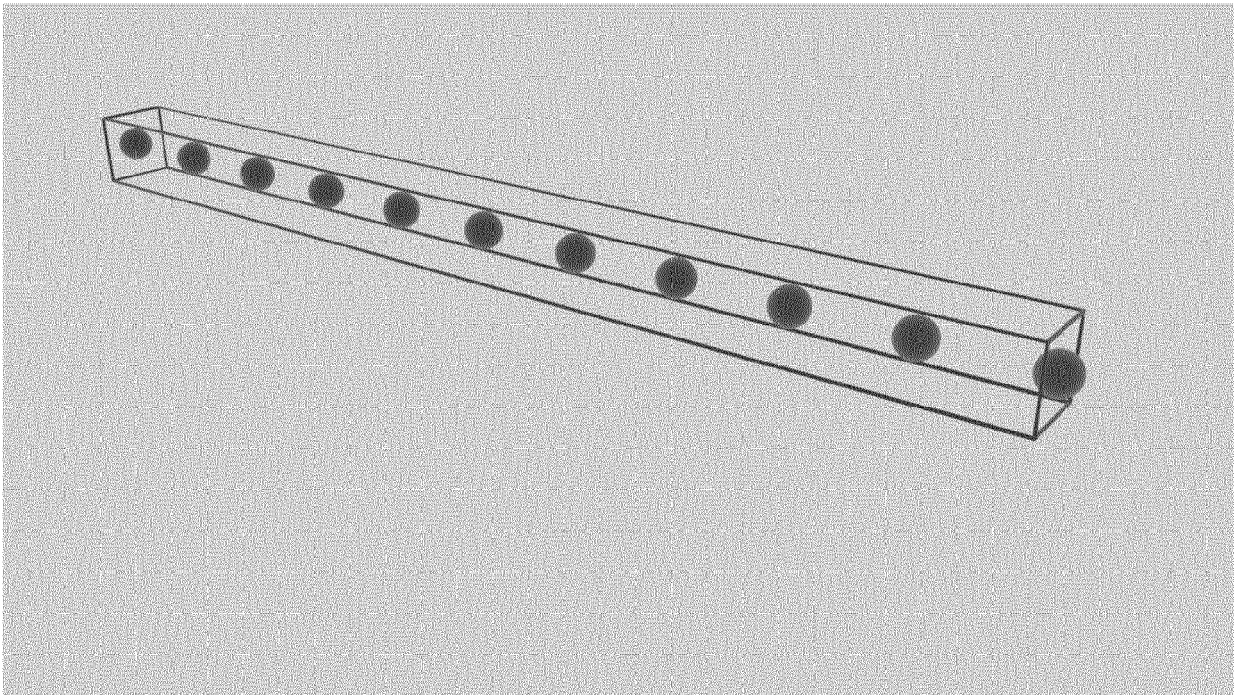Patentamt

European
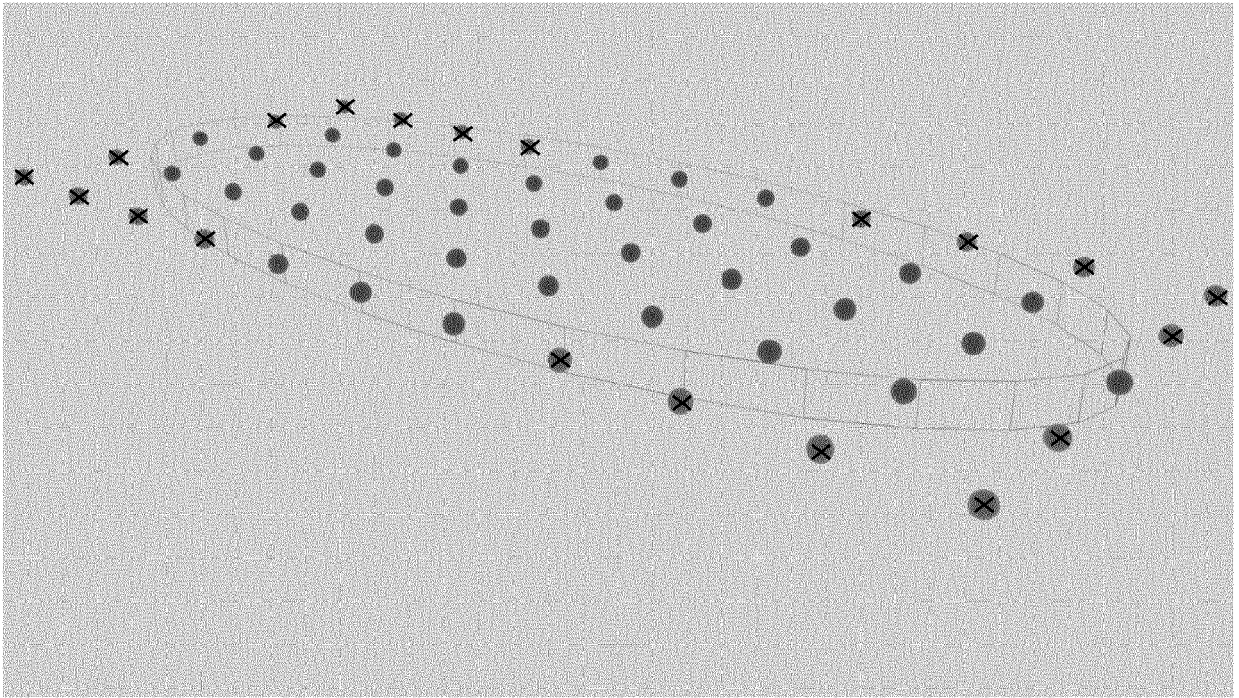Patent Office

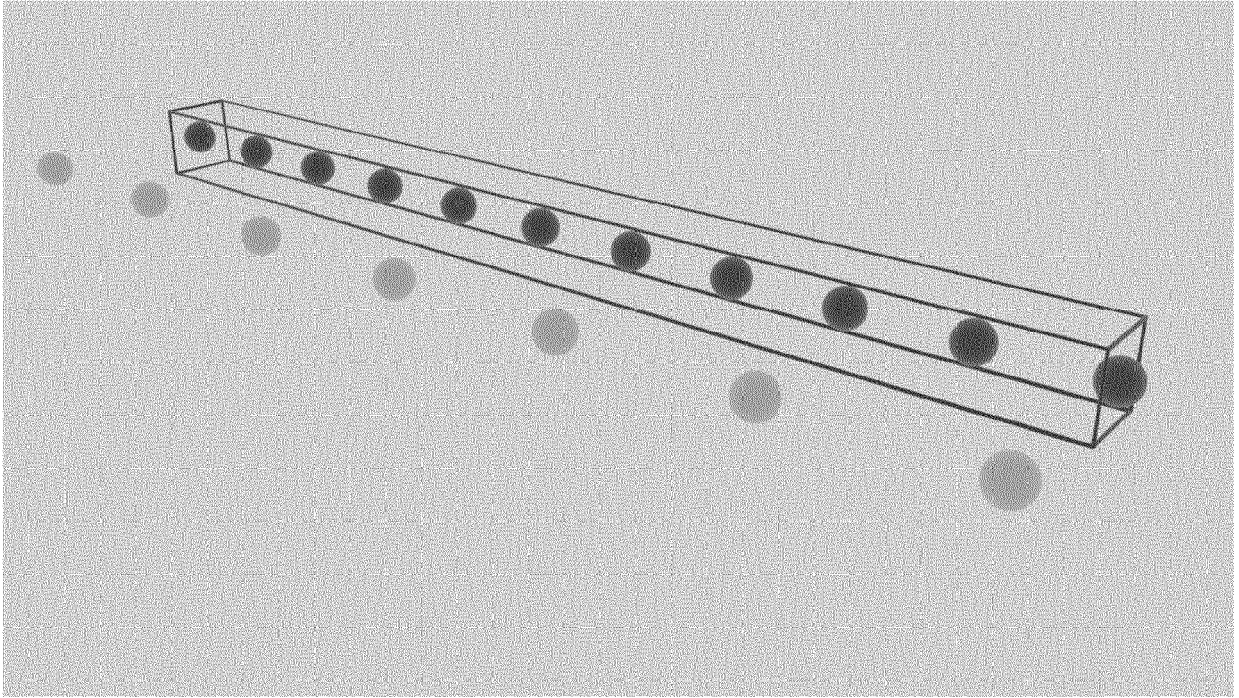Office européen
des brevets

EUROPEAN SEARCH REPORT

Application Number

EP 22 15 0861

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2017/223476 A1 (BREEBAART DIRK JEROEN [AU] ET AL) 3 August 2017 (2017-08-03) * paragraphs [0107] - [0131]; figures 9,10A,10B *	1-18	INV. H04S7/00
X A	US 2006/165238 A1 (SPILLE JENS [DE] ET AL) 27 July 2006 (2006-07-27) * paragraphs [0017] - [0026]; figures 1,2 *	1,15 2-14, 16-18	
X A	US 2020/145773 A1 (ARTEAGA DANIEL [ES] ET AL) 7 May 2020 (2020-05-07) * paragraphs [0018] - [0025] *	1,15 2-14, 16-18	

TECHNICAL FIELDS SEARCHED (IPC)

H04S

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	23 June 2022	Fruhmann, Markus

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)

2

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

EP 22 15 0861

23-06-2022

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2017223476 A1	03-08-2017	BR 112016001738 A2	01-08-2017
		CN 105431900 A	23-03-2016
		CN 110797037 A	14-02-2020
		CN 110808055 A	18-02-2020
		EP 3028273 A1	08-06-2016
		EP 3564951 A1	06-11-2019
		HK 1229945 A1	24-11-2017
		JP 6388939 B2	12-09-2018
		JP 6804495 B2	23-12-2020
		JP 2016530803 A	29-09-2016
		JP 2018174590 A	08-11-2018
		JP 2021036729 A	04-03-2021
		KR 20160021892 A	26-02-2016
		KR 20160140971 A	07-12-2016
		KR 20210141766 A	23-11-2021
		KR 20220061284 A	12-05-2022
		RU 2016106913 A	01-09-2017
		RU 2018104812 A	26-02-2019
		US 2016192105 A1	30-06-2016
		US 2017223476 A1	03-08-2017
		US 2018295464 A1	11-10-2018
		US 2020221249 A1	09-07-2020
		US 2022046378 A1	10-02-2022
		WO 2015017235 A1	05-02-2015
US 2006165238 A1	27-07-2006	AT 357043 T	15-04-2007
		AU 2003273981 A1	04-05-2004
		BR 0315326 A	16-08-2005
		CN 1973318 A	30-05-2007
		DE 60312553 T2	29-11-2007
		EP 1570462 A1	07-09-2005
		ES 2283815 T3	01-11-2007
		JP 4751722 B2	17-08-2011
		JP 2006516164 A	22-06-2006
		JP 2010198033 A	09-09-2010
		KR 20050055012 A	10-06-2005
		US 2006165238 A1	27-07-2006
		WO 2004036548 A1	29-04-2004
US 2020145773 A1	07-05-2020	CN 110603821 A	20-12-2019
		EP 3619922 A1	11-03-2020
		US 2020145773 A1	07-05-2020
		US 2022103961 A1	31-03-2022

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.

Patent documents cited in the description

- US 9654895 B2 **[0012]**