



(11)

EP 4 498 704 A1

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
29.01.2025 Bulletin 2025/05

(51) International Patent Classification (IPC):
H04S 3/00 (2006.01) H04S 7/00 (2006.01)

(21) Application number: **23187995.8**

(52) Cooperative Patent Classification (CPC):
**H04S 7/302; H04S 3/008; H04S 7/304; H04S 7/306;
H04S 7/307**

(22) Date of filing: **27.07.2023**

(84) Designated Contracting States:
**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB
GR HR HU IE IS IT LI LT LU LV MC ME MK MT NL
NO PL PT RO RS SE SI SK SM TR**
Designated Extension States:
BA
Designated Validation States:
KH MA MD TN

(72) Inventors:
• **OUWELTJES, Okke**
Eindhoven (NL)
• **JELFS, Sam Martin**
Eindhoven (NL)
• **KOPPENS, Jeroen Gerardus Henricus**
5656AG Eindhoven (NL)
• **OOMEN, Arnoldus Werner Johannes**
Eindhoven (NL)

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

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

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

(57) An apparatus comprises a receiver (201) receiving a data signal comprising an audio signal and meta-data including a pose indication for an audio source of the first audio signal, and frequency equalization data indicative of a reference audio reproduction frequency response for the audio signal. A listener pose processor (205) determines a listening pose and a binaural transfer function processor (207) determines a binaural transfer function dependent on the listening pose and a pose of the audio source. A reproduction processor (211) determines an audio rendering frequency response indicative of an audio reproduction frequency response of an audio reproduction path for the output audio signal, such as headphone frequency response. An adapter (101) generates a binaural filter having a frequency response dependent on a combination of the audio rendering frequency response, the reference audio reproduction frequency response, and the binaural transfer function. A renderer (203) then generates an output audio signal using the binaural filter.

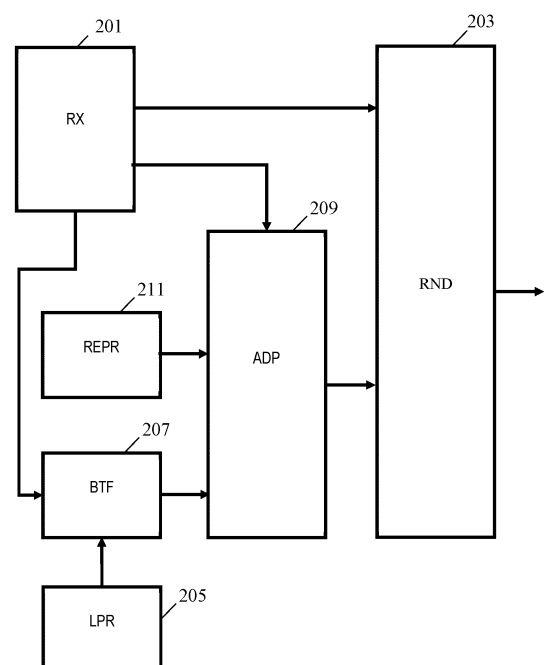


FIG. 2

EP 4 498 704 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 activities 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. In the following, the term Virtual Reality/ VR will be used to denote both Virtual Reality and Augmented Reality.

[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 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] However, whereas such applications may provide suitable user experiences in many embodiments they tend to not provide optimal user experiences in all situations. In particular, in many situations, a suboptimum audio quality may be provided, and typically a distortion may result for the rendered audio compared to the original or desired content side audio. In particular, in many applications, the approach may not accurately reflect or compensate for variations in signal processing or audio reproduction at the source or render side.

[0009] Hence, an improved approach for distribution and/or rendering and/or processing of audio signals, 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 audio reproduction functions, facilitated and/or improved adaptation to changes in listener position/orientation (e.g., a virtual listener position/orientation); an improved Virtual Reality experience, and/or improved performance and/or operation would be advantageous.

SUMMARY OF THE INVENTION

[0010] Accordingly, the Invention seeks to preferably mitigate, alleviate or eliminate one or more of the above mentioned disadvantages singly or in any combination.

[0011] 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 at least a first audio signal and metadata including: a pose indication for an audio source of the first audio signal, and frequency equalization data indicative of a reference audio reproduction frequency response for the first audio signal; a listener pose processor arranged to determine a listening pose; a binaural transfer function processor arranged to determine a binaural transfer function dependent on the listening pose and the pose indication for the audio source; a processor arranged to determine an audio rendering frequency response indicative of an audio

reproduction frequency response of an audio reproduction path for the output audio signal; an adapter arranged to generate a binaural filter having a frequency response dependent on a combination of the audio rendering frequency response, the reference audio reproduction frequency response, and the binaural transfer function; and a renderer arranged to generate the output audio signal in dependence on the binaural filter.

[0012] The approach may provide an improved output audio signal to be generated. The approach may in many embodiments and scenarios provide an improved audio quality and may e.g. reduce frequency distortion or degradations caused by the processing path. The approach may for example allow a rendering side compensation for distortion caused by the audio reproduction equipment (e.g. headphones) used by the content side to adapt/generate the audio signal at the encoding/creation side. For example, a content creator or sound engineer may adapt or generate the audio signal with a desired audio quality and content, and as this is based on the sound heard by the content creator/ sound engineer the frequency response of the used headphones (or other audio reproduction means) may impact the generated audio. The approach may allow rendering side compensation of such effects while at the same time compensating for the local rendering side audio reproduction means.

[0013] Further, the approach may allow for a very efficient implementation and operation and may in particular allow low complexity and/or resource usage. The approach may in particular allow a single filtering of the received audio signal to provide compensation for both content creator side audio reproduction effects and rendering side audio reproduction effects, as well as provide effective binaural filtering. Indeed, the approach may provide such improved effects with substantially no computational resource increase over a conventional generation of binaural signals. For example, the binaural filter may only be updated at low rate to reflect user movement (or audio source movement), and thus the updating may require very little computational resource in comparison to the resource required for continuous filtering of the audio signal.

[0014] A pose, which also may be referred to as a placement, may be a position and/or orientation. The listening pose may be the pose for which the (spatial) output audio signal is generated. The output audio signal may be a stereo audio signal.

[0015] The binaural transfer function processor may be arranged to determine a frequency response of a binaural transfer function dependent on the listening pose and the pose indication for the audio source. The binaural transfer function processor may be arranged to determine the binaural transfer function dependent on a difference between the listening pose and the pose indication for the audio source.

[0016] The audio reproduction path for the output audio signal may comprise or consist in a function converting the output audio signal from an electrical signal to

an audio/sound/acoustic signal. The audio reproduction path for the output audio signal may comprise or consist in an audio transducer, such as headphones and/or loudspeakers.

[0017] The audio reproduction frequency response for the output audio signal may comprise or consist in a frequency response of an audio transducer, such as a pair of headphones or loudspeaker(s).

[0018] The reference audio reproduction frequency response for the first audio signal may comprise or consist in a frequency response of an audio transducer, such as a pair of headphones or loudspeaker(s).

[0019] The adapter may be arranged to generate the binaural filter to have a frequency response which is the combination (in particular a series coupling) of a filter having a frequency response matching the audio rendering frequency response, a filter having a frequency response matching the reference audio reproduction frequency response, and a filter having a frequency response matching a frequency response of the binaural transfer function.

[0020] In accordance with an optional feature of the invention, the reference audio reproduction frequency response is a target frequency response for the first audio signal.

[0021] This may provide improved and/or facilitated operation or performance in many embodiments. The target frequency response may be indicative of a (desired/target) frequency response for the entire audio path or e.g., only for the rendering audio path.

[0022] In accordance with an optional feature of the invention, the reference audio reproduction frequency response is a source frequency response applied in generation of the first audio signal.

[0023] This may provide improved and/or facilitated operation or performance in many embodiments.

[0024] In accordance with an optional feature of the invention, the frequency equalization data comprises an indication of an audio reproduction device, and the adapter is arranged to determine the reference reproduction frequency response as a predetermined frequency response for the audio reproduction device.

[0025] This may provide improved and/or facilitated operation or performance in many embodiments. The audio reproduction device may be (or include) a headphone and/or loudspeaker(s). The predetermined frequency response may for example be a stored frequency response for the audio reproduction device. The predetermined frequency response may for example be stored internally or remote from the audio apparatus. For example, the adapter may be arranged to retrieve the frequency response from a remote server.

[0026] In accordance with an optional feature of the invention, the frequency equalization data comprises data describing a Finite Impulse Response, FIR, filter having a frequency response matching the reference audio reproduction frequency response.

[0027] This may provide improved and/or facilitated

operation or performance in many embodiments. It may allow a particularly advantageous representation that may be indicated effectively and with a relatively low data rate. It may further provide improved audio quality by allowing an accurate representation.

[0028] In accordance with an optional feature of the invention, the frequency equalization data comprises data describing a combination of a set of sections, each section being a First Order Section or a Second Order Section and the combination having a frequency response matching the reference audio reproduction frequency response.

[0029] This may provide improved and/or facilitated operation or performance in many embodiments. It may allow a particularly advantageous representation that may be indicated effectively and with a relatively low data rate. It may further provide improved audio quality by allowing an accurate representation.

[0030] In accordance with an optional feature of the invention, the frequency equalization data comprises data describing a set of parallel filters, the set of parallel filters having a frequency response matching the reference audio reproduction frequency response.

[0031] This may provide improved and/or facilitated operation or performance in many embodiments. It may allow a particularly advantageous representation that may be indicated effectively and with a relatively low data rate. It may further provide improved audio quality by allowing an accurate representation. For example, the approach may facilitate processing reflecting perceptual significance, such as e.g. with the reference audio reproduction frequency response reflecting perceptual frequency bands etc.

[0032] In accordance with an optional feature of the invention, the frequency equalization data comprises frequency response identification data, and the adapter is arranged to access a remote server to retrieve the reference reproduction frequency response based on the frequency response identification data.

[0033] This may provide improved and/or facilitated operation or performance in many embodiments.

[0034] In accordance with an optional feature of the invention, the metadata further comprises a reference playback level; and wherein audio apparatus is arranged to adapt the frequency response of the binaural filter in dependence on a first play back level for a reproduction of the output audio signal relative to the reference playback level.

[0035] This may provide improved and/or facilitated operation or performance in many embodiments.

[0036] In accordance with an optional feature of the invention, the metadata includes data indicating a dependency of the reference audio reproduction frequency response on a playback level and the adapter is arranged to adapt the reference audio reproduction frequency response in dependence on the playback level.

[0037] This may provide improved and/or facilitated operation or performance in many embodiments.

[0038] According to an aspect of the invention there is provided an apparatus for generating an audio data signal, the apparatus comprising: a receiver arranged to receive at least a first audio signal from a first audio source having a pose in a scene; a reproduction processor (arranged to determine a reference audio reproduction frequency response for the first audio signal; a data signal generator arranged to generate the audio data signal to comprise: audio data for the first audio signal; a pose indication indicative of the pose of the audio source; and frequency equalization data indicative of the reference reproduction frequency response.

[0039] 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 at least a first audio signal and metadata including: a pose indication for an audio source of the first audio signal, and frequency equalization data indicative of a reference audio reproduction frequency response for the first audio signal; determining a listening pose; determining a binaural transfer function dependent on the listening pose and a pose indication for the audio source; determining an audio rendering frequency response indicative of an audio reproduction frequency response of an audio reproduction path for the output audio signal; generating a binaural filter having a frequency response dependent on a combination of the audio rendering frequency response, the reference audio reproduction frequency response, and the binaural transfer function; and generating the output audio signal in dependence on the binaural filter.

[0040] According to an aspect of the invention there is provided an audio data signal, the method comprising: receiving at least a first audio signal from a first audio source having a pose in a scene; determining a reference reproduction frequency response for the first audio signal; a data signal generator arranged to generate the audio data signal to comprise: audio data for the first audio signal; a pose indication indicative of the pose of the audio source; and frequency equalization data indicative of the reference reproduction frequency response.

[0041] According to an aspect of the invention there is provided an audio data signal comprising at least a first audio signal and metadata including: a pose indication for an audio source of the first audio signal; and frequency equalization data indicative of a reference reproduction frequency response for the first audio signal.

[0042] 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

[0043] 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 some elements of a possible arrangement of a processor for implementing elements of an apparatus in accordance with some embodiments of the invention.

DETAILED DESCRIPTION OF SOME EMBODIMENTS OF THE INVENTION

[0044] The following description will focus on eXtended Reality applications where audio is rendered to 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.

[0045] 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.

[0046] 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).

[0047] 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. How-

ever, it will be appreciated that in other embodiments, the pose may be represented by fewer values.

[0048] 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).

[0049] 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).

[0050] 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.

[0051] 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.

[0052] 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.

[0053] 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.

[0054] 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.

[0055] 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.

[0056] 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.

[0057] 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.

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.

[0058] 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 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. FIG. 3 illustrates an example of an apparatus, also henceforth referred to as an audio encoding apparatus, for generating an audio data signal that represents audio for one or more audio sources of a scene. 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

[0059] 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 view images may in many embodiments be generated and presented to the user.

[0060] The audio rendering apparatus comprises a first receiver 201 which is arranged to receive audio items from a local or remote source. In the specific example, the first receiver 201 receives data describing the audio items from the server 101. The first receiver 201 may be arranged to receive data describing the virtual scene, and specifically the audio scene. The data may include data providing a visual description of the scene and may include data providing an audio description of the scene. Thus, an audio scene description and a visual scene description may be provided by the received data.

[0061] The audio signals/items may be encoded audio data, such as encoded audio signals. The audio signals may be different types of audio elements including different types of elements and components, and indeed in many embodiments the first receiver 201 may receive audio data which defines different types/ formats of audio. For example, the audio data may include audio represented by audio channel signals, individual audio objects, scene based audio such as Higher Order Ambisonics (HOA) etc. The audio may for example be represented as encoded audio for a given audio component that is to be rendered.

[0062] The received data signal further comprises data indicative of the poses (position and/or orientation) of the audio sources for which the audio signals are provided. In many embodiments, at least some of the audio signals are linked to data describing the position of the audio source in the scene. In some cases, data may alternatively, or typically additionally, comprise data defining the orientation of the audio sources (e.g., allowing non-omni directional audio sources).

[0063] The received signal may specifically comprise metadata which includes pose data that indicates a position and/or orientation of the audio items and specifically the position and/or orientation of an audio source and/or a visual scene object or element. The pose data may for example include absolute position and/or orientation data defining a position of each, or at least some of the audio sources.

[0064] The first receiver 201 is coupled to a renderer 203 which proceeds to render the audio scene based on the received data describing the audio items. In case of encoded data, the renderer 203 may also be arranged to decode the audio data (or in some embodiments decoding may be performed by the first receiver 201).

[0065] The renderer 203 is arranged to render the audio scene by generating audio signals based on the received audio data for the audio sources. In the example, the renderer 203 is 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).

[0066] Thus, in many embodiments, the audio rendering by the renderer 203 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 203 may be arranged to generate an audio component to be perceived to arrive from a specific position using binaural processing.

[0067] 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 cancellation 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.

[0068] 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.

[0069] 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 present during the recording.

[0070] 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.

[0071] 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), or Binaural Room Transfer Functions (BRTFs). 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 room 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.

[0072] The renderer 203 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.

[0073] The audio rendering apparatus further comprises a listener pose processor 205 which is arranged to determine a listening 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.

[0074] 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.

[0075] The renderer 203 of the apparatus of FIG.2 is arranged to generate the output audio signal by a binaural rendering process such that spatial cues are provided from e.g., rendering using headphones. The renderer may filter the audio signal by suitable binaural filters corresponding to respectively the left and right ear in order to generate a binaural output signal that when

presented to the ears of a user will provide a spatial audio experience.

[0076] The listener pose processor 205 is coupled to a binaural transfer function processor 207 which receives the listener pose from the listener pose processor 205 and which is arranged to determine a binaural transfer function for this listener pose and for a given audio source pose. The binaural transfer function processor 207 is arranged to determine the appropriate binaural transfer function for a given audio signal and audio source dependent on the relative position/pose of the audio source relative to the listener pose.

[0077] The binaural transfer function processor 207 may comprise a store/ service/ database with binaural transfer functions for a, typically high, number of different positions, and sometimes orientations, with each binaural transfer function providing information on how an audio signal should be processed/ filtered in order to be perceived to originate from that position.

[0078] The binaural transfer function processor 207 may for a given audio signal that is to be perceived to originate from a given position relative to the user's head, select and retrieve the stored binaural transfer function that most closely matches the desired relative position, i.e., that most closely matches the audio source position relative to the listener position. In some embodiments, the binaural transfer function processor 207 may be arranged to generate the binaural transfer function by interpolating between a plurality of close stored binaural transfer functions. It may then apply the selected binaural transfer function to the audio signal of the audio element thereby generating an audio signal for the left ear and an audio signal for the right ear.

[0079] However, in the apparatus of FIG. 2, the extracted binaural transfer function is not used directly by the renderer 203 but rather it is adapted prior to the rendering. The binaural transfer function processor 207 is coupled to an adapter 209 which is arranged to adapt the binaural transfer function and specifically it is arranged to generate a binaural filter that has a frequency response which corresponds to the frequency response of the binaural transfer function but with this being adapted/modified. The binaural filter is then fed to the renderer 203 where it is used to perform the binaural rendering. Typically, the binaural filter is a stereo filter comprising both a transfer function/ impulse response/ frequency response for generating a left ear signal and for generating a right ear signal (equivalently the adapter 207 may generate two filters that are used by the renderer 203 to generate respectively the left ear signal and the right ear signal. The described process may in such a case be applied for both binaural filters).

[0080] The renderer may specifically generate an output stereo signal by filtering the audio signal by the binaural filter(s). The generated output stereo signal in the form of the left and right ear signal is then suitable for headphone rendering and may be amplified to generate drive signals that are fed to the headset of a user. The

user will then perceive the audio signal/source to originate from the desired position.

[0081] It will be appreciated that the audio signals may in some embodiments also be processed to e.g., add acoustic environment effects. For example, the audio signal may be processed to add reverberation or e.g., decorrelation/ diffuseness. In many embodiments, this processing may be performed on the generated binaural signal rather than directly on the audio element signal.

[0082] Thus, the renderer 203 may be arranged to generate the output audio (stereo) signal such that a given audio source is rendered with a user wearing headphones perceiving the audio to be received from a desired position. The approach is typically applied to a plurality of audio sources and signals which are then combined by the renderer 203 to generate the output signal.

[0083] It will be appreciated that many algorithms and approaches for rendering of spatial audio, e.g., using headphones, and specifically for binaural rendering, will be known to the skilled person and that any suitable approach may be used without detracting from the invention.

[0084] The adapter 209 may specifically be arranged to compensate/adapt the binaural transfer functions for properties of audio reproduction at both the rendering and source side.

[0085] The apparatus of FIG. 2 specifically comprises a reproduction processor 211 which is arranged to determine an audio rendering frequency response which reflects the audio reproduction frequency response of the audio reproduction path for the output audio signal. The audio rendering frequency response may specifically reflect a frequency response of a pair of headphones that are used to render the generated output audio signal to the user. More generally, the audio rendering frequency response typically is or includes a frequency response of an audio transducer arranged to convert an electrical signal to an acoustic signal. The audio transducer may e.g., be a headphone and/or loudspeaker.

[0086] In the approach, the received data signal further comprises frequency equalization data that is indicative of a reference audio reproduction frequency response for the generated audio. The received data may comprise frequency equalization data that for example indicate a desired frequency spectrum compensation, a target frequency response, and/or a frequency response of an audio reproduction device that was used at the content side, e.g., in the generation or post-processing of the audio of the audio source.

[0087] The reproduction processor 211 is coupled to the adapter 209 which is fed the audio rendering frequency response, and which further is fed the reference audio reproduction frequency response from the receiver 201. The adapter 209 is arranged to adapt the binaural transfer function in dependence on both the audio rendering frequency response and the reference audio re-

production frequency response to generate the binaural filter which is fed to the renderer 203 and applied to the audio signal to generate the binaural audio signal.

[0088] The exact adaptation will depend on the specific preferences and requirements of the individual embodiment, and will depend on e.g., the format and specific properties of the provided audio rendering frequency response and reference audio reproduction frequency response.

[0089] In many embodiments, the reference audio reproduction frequency response may be indicative of a frequency response of an audio reproduction device, such as a headphone, that was used by a content creator to generate (e.g., equalize) the audio signal. Similarly, the reference audio reproduction frequency response may characterize the frequency response of a set of headphones or a loudspeaker that are used to render the output stereo signal to the user. In such a case, the adapter 209 may be arranged to compensate the frequency response of the extracted binaural transfer function to compensate for the frequency responses of the headphones/ loudspeaker. For example, the gain of the binaural transfer function for a first frequency may be multiplied by a (normalized) gain of the reference audio reproduction frequency response and divided by a (normalized) gain of the audio rendering frequency response. This may be repeated for all frequency values resulting in a frequency response of the binaural filter. This frequency response thus represents an audio reproduction compensated binaural transfer function/ filter that, when applied to the audio signal may provide improved audio quality.

[0090] In the approach, the rendering side may accordingly adapt/ compensate for frequency distortion at both the encoder and decoder side. The compensation and audio quality may be aided or controlled by the content side thereby allowing a content creator to affect the rendering properties. Further, the adaptation of both the content creator side and the rendering side properties are synergistically combined with the binaural processing. Indeed, the binaural filter pre-processing (rather than e.g., applying a filter as postprocessing of the output stereo signal), provides a highly efficient processing and operation which may provide improved audio quality for a low implementation complexity and resource usage. In particular, in the approach only a single filtering of the audio signal is required to achieve the beneficial effects.

[0091] The encoder of FIG. 3 comprises an encoder receiver 301 arranged to receive at least a first audio signal from a first audio source having a given pose in a(n audio) scene. Typically, the receiver will receive a plurality of audio sources such as audio sources at different positions and/or with different orientations, and/or audio sources with no specific position and/or orientation, such as e.g., a source representing the diffuse reverberation audio in the audio scene.

[0092] The encoder further comprises an encoder reproduction processor 303 which is arranged to determine

a reference audio reproduction frequency response for the first audio signal. The reference audio reproduction frequency response may be determined in dependence on a frequency response of an audio reproduction path or device that has been used at the encoder/content side. For example, the reference audio reproduction frequency response may be determined on the basis of the frequency response of a pair of headphones or a loudspeaker that was used by a content creator to generate the audio signal. In many embodiments, the reference audio reproduction frequency response may be or include a frequency response of an audio transducer arranged to convert an electrical signal to an acoustic signal. The audio transducer may e.g., be a headphone and/or loudspeaker.

[0093] The encoder reproduction processor 303 may for example include a user interface allowing a user or content creator to manually input data describing a frequency response, e.g., of a headphone or speaker. As another example, in some embodiments, the encoder reproduction processor 303 may be arranged to control a test system to measure a frequency response of a loudspeaker. For example, the encoder reproduction processor 303 may comprise a variable frequency tone generator that may be coupled to an external audio reproduction device, e.g., headphone or speaker. A frequency sweep may be performed, and a microphone positioned appropriately for a given audio reproduction device may generate a microphone signal capturing the generated audio. The level of the captured audio for different frequencies may then be used to determine the frequency response of the reproduction device/headphone. As another example, the reproduction device may have functionality for identifying e.g., a brand, model, type etc. and the encoder reproduction processor 303 may be arranged to retrieve a stored nominal frequency response for that device, e.g., from a remote database. For example, a Bluetooth™ headphone may provide identification data that allows the encoder reproduction processor 303 to access e.g., a remote database to retrieve the frequency response for the headphone.

[0094] The encoder reproduction processor 303 and the encoder receiver 301 are coupled to a data signal generator 305 which is arranged to generate an output audio data signal which may be transmitted or distributed to the decoder of FIG. 1 (as well as possibly to other decoder devices). The data signal generator 305 may encode the audio signal to generate encoded audio data that is included in the audio data signal. It may typically encode other audio sources to provide a combined audio representation of the audio scene. The data signal generator 305 may further receive pose data for the audio source of the first audio signal, and typically for other audio sources, and may be arranged to include pose data for the audio sources in the output audio signal.

[0095] The data signal generator 305 may further receive the frequency equalization data indicative of the reference reproduction frequency response, such as

specifically the frequency response of a reproduction device used by the content creation side. The data signal generator 305 may then include data in the output audio data signal that indicates and describes the reference audio reproduction frequency response.

[0096] In some embodiments, the reference audio reproduction frequency response may be a target frequency response for the reproduction of the first audio signal at the decoder side. For example, the reference audio reproduction frequency response may indicate a desired frequency response for the overall audio path, including the frequency response of the reproduction path at the rendering device.

[0097] For example, a content creator may manually process and adapt an audio signal to sound as desired and such a process typically include frequency equalization and filtering, i.e., the content creator may adapt the frequency spectrum of the audio signal to result in a desired sound. It may further be desired that the overall frequency response of the entire signal processing is a flat frequency response such that the rendered sound corresponds more directly to the audio manually generated by the content creator. However, the content creator will inherently use an audio reproduction device, such as headphones or speakers, and this will inherently have a frequency response that will tend to not be flat. Accordingly, the audio as heard by the content creator may be colored by the reproduction. In the described approach, the reference audio reproduction frequency response may be used to compensate for such distortion by indicating a target reproduction frequency response that will compensate for the frequency response of the content side reproduction path.

[0098] Specifically, the target reference audio reproduction frequency response may be indicated as frequency response that compensates, and e.g., has the inverse frequency response of the determined content side reproduction device frequency response. For example, when the content side reproduction frequency response has increased attenuation, the target reference audio reproduction frequency response may be determined to have the corresponding gain. In particular, the reference audio reproduction frequency response may be determined as the inverse/ reciprocal of the determined reproduction frequency response for the content side reproduction device. When receiving such a target reference audio reproduction frequency response, the rendering apparatus of FIG. 2 may proceed to perform a frequency equalization that includes applying the corresponding reference audio reproduction frequency response. The rendering apparatus may further include a compensation for the audio rendering frequency response. Thus, specifically, the binaural transfer function/HRTF may be modified to provide a desired target response while compensating for the local audio reproduction. As a specific example, the frequency response of a binaural filter used by the renderer 203 may be determined as:

$$H_{RND}(f) = \frac{H_T(f)}{H_R(f)} H_B(f)$$

5 where $H_T(f)$ indicates the received target reference audio reproduction frequency response, $H_R(f)$ indicates the local audio rendering frequency response, and $H_B(f)$ indicates the determined binaural transfer function (specifically HRTF) for the position of the audio source relative to the listening position.

10 **[0099]** In some embodiments, the reference audio reproduction frequency response is a source frequency response applied in generation of the first audio signal. Specifically, the reference audio reproduction frequency response may provide an indication of the frequency response of e.g., a headphone or a loudspeaker that has been used by e.g., a sound engineer or content creator on the content side. In such a case, the adapter 209 may be arranged to adapt the frequency response of the binaural transfer function to compensate for the filter response. For example, the frequency response of a binaural filter used by the renderer 203 may be determined using the same approach as described above (bearing in mind that the impact of the frequency response of the headphone/loudspeaker on the transmitted signal is the inverse of the frequency response, and thus that the frequency response of the headphone is also the desired target frequency response to compensate for the headphone).

20 **[0100]** The frequency equalization data may use different approaches for indicating the reference audio reproduction frequency response.

25 **[0101]** In some embodiments, the frequency equalization data may comprise data that describes a Finite Impulse Response, FIR, filter having a frequency response matching the reference audio reproduction frequency response. The reference audio reproduction frequency response may specifically be described as a FIR filter that has a given frequency response. The FIR filter may be described in the frequency domain and/or in the time domain. For example, the frequency equalization data may include a set of coefficients for a FIR filter, and specifically the coefficients may correspond to suitable samples of an impulse response for a FIR filter having a frequency response that represents/matches the reference audio reproduction frequency response.

30 **[0102]** Such an approach may provide a highly efficient representation that is easy to process at the renderer side. It may also be generated with relatively low complexity at the content side. It may further be advantageous in that a FIR filter may effectively provide equalization of both amplitude and phase, and indeed even with these being done independently from each other.

35 **[0103]** In some embodiments, the frequency equalization data may include data that describes the reference audio reproduction frequency response as an Infinite Impulse Response, IIR, filter. This may in some scenarios provide for a more compressed representation and/or

lower complexity implementation.

[0104] In some embodiments, the frequency equalization data may comprise data that describes a combination of a set of sections with the combination having a frequency response matching the reference audio reproduction frequency response. Each section may be a First Order Section, FOS, or a Second Order Section, SOS. The combination may specifically be a series of the sections of FOS and SOSs. Each section may thus correspond to a first order or second order filter, and the combination may correspond to a filter formed by the series of these first and second order filters. The combined frequency response of the series of filters may be considered by the renderer 203 as the reference audio reproduction frequency response.

[0105] Such an approach may be particularly advantageous in many embodiments and scenarios and may in particular allow computationally efficient implementations and functions.

[0106] In some embodiments, the frequency equalization data may include data describing a set of parallel filters and the reference audio reproduction frequency response may be given as the combined frequency response of these parallel filters. For example, a set of parallel band pass filters may be provided with each bandpass filter providing the frequency response for a given frequency band and with other the pass bands adding up to the overall frequency response. As a specific example, parallel filters may for example be provided an octave apart. In some embodiments, parallel filters may be provided with different bandwidths and an advantage of such an approach is that the filter set may e.g., be designed on a perceptually motivated logarithmic frequency scale, which is an advantage over FIR filters that are designed on a linear frequency scale. For example, the filters and filter bandwidths may be determined to correspond to e.g., a logarithmic, octave, 1/3rd octave, equivalent rectangular bandwidth (ERB), Bark scale, etc.

[0107] As a specific example, the audio data signal may comprise frequency equalization data in the form of an EQ ((frequency EQualization) field specifier that describes e.g., a type and method of frequency equalization. Such an EQ field may indicate a method of frequency equalization and the equalization filters that can or may be used. The frequency equalization data may for example include data indications for one or more of the following examples of frequency equalization indications:

The reference audio reproduction frequency response represented as the coefficients of an FIR filter specified at certain sample rate.

[0108] The reference audio reproduction frequency response represented as a series of one or more of First and Second Order Sections (FOS & SOS) specified either by their coefficients and sample rate or by their descriptive parameters. For example, there may be provided a second order peak-EQ filter with a center frequency, a gain and a quality factor, followed by first order

shelving filter with a given gain and center frequency, etc.

[0109] The reference audio reproduction frequency response represented as a sum of parallel filters.

[0110] It will be appreciated that other approaches may be used depending on the requirements and preferences of the individual embodiment.

[0111] In some embodiments, the audio rendering apparatus may receive frequency equalization data that comprises frequency response identification data. This identification data may not fully describe a frequency response but may provide an identification of a frequency response, such as specifically a specific name and/or number.

[0112] In some embodiments, the audio rendering apparatus may be arranged to extract a corresponding reference audio reproduction frequency response from a set of locally stored reference audio reproduction frequency responses. For example, during manufacturing, a plurality of reference audio reproduction frequency responses may be stored in local memory together with an associated identity for each reference audio reproduction frequency response. During operation, the rendering apparatus may receive an audio data signal with an identification of a reference audio reproduction frequency response, and it may proceed to access the memory to attract the reference audio reproduction frequency response that is stored for the received identification. It may then proceed to use this reference audio reproduction frequency response to modify the binaural transfer function.

[0113] In some embodiments, the adapter 209 may be arranged to access a remote server to retrieve a reference reproduction frequency response that matches the frequency response identification received in the data audio signal. For example, the adapter 209 may comprise a network interface that e.g., may allow it to connect to the Internet to access a remote server with a request that includes the frequency response identity. The server may extract the corresponding reference audio reproduction frequency response and transmit it back to the rendering apparatus.

[0114] Such an approach may be highly advantageous in many embodiments as it may allow a single (or a few) central servers to serve a large number of rendering devices. This may substantially facilitate the distribution and updating of the reference audio reproduction frequency responses.

[0115] In many embodiments, the identification may specifically be an identification of an audio reproduction device, and specifically of an audio reproduction device that may be used at the content side operation. For example, the identification may identify headphones, speakers, amplifiers etc. that may be used as part of the content side setup. The identification may in many examples simply be data describing the make and model of the device. The adapter 209 may then access a remote database/ server to retrieve a frequency response for that device and the binaural transfer function may be modified

to compensate for this response.

[0116] As a specific example, if a headphone brand and type is indicated, the adapter 209 may access a (third-party) database containing the preferred equalization filter frequency response or the headphone frequency response and this may then be used to provide a desired target frequency response, such as specifically a flat frequency response.

[0117] In some embodiments, the encoder may further be arranged to generate the audio data signal to include metadata that provides an indication of a reference playback level. The reference playback level may be a nominal value or may e.g., be dynamically determined to directly indicate the level of the playback that has been applied, e.g., by a content creator or sound engineer manually adjusting the frequency spectrum of the first audio signal. For example, a sound engineer may perform a manual frequency equalization while listening to the stereo audio signal. The level of the reproduction of the stereo signal may be measured or determined (e.g., from a volume setting by the sound engineer) and data may be included in the audio data signal indicating this playback level.

[0118] The playback level may for example be provided as a relative or absolute value. For example, in some embodiments, a general audio level in a predetermined range (e.g., from 0 to 11) may be provided. In other embodiments, the reference playback level may be provided as an indication of a Sound Pressure Level, SPL.

[0119] The renderer may receive such an audio data signal and extract the reference playback level from the metadata of the audio data signal and the adapter 209 may be arranged to adapt the binaural filter in dependence on the reference playback level and a playback level for the generated output audio signal.

[0120] For example, the current level setting (e.g., a volume setting) for the rendering/ reproduction of the output stereo signal may be provided to the adapter 209 and depending on this and the reference playback level, the adapter 209 may for example modify the frequency response adaptation applied to the binaural transfer function. For example, if the reference playback level is indicative of a low level and the current volume setting is set to a low reproduction sound level, the adapter 209 not provide any additional sound level frequency adaptation. However, if the volume setting is currently set to correspond to a high reproduction sound level, the adapter 209 may proceed to perform an additional frequency response adjustment to attenuate low and high frequencies (e.g., to prevent hearing damage). If instead, the reference playback level indicates a high sound level, the adapter 209 may be arranged to not include any additional frequency response adaptation for a high volume setting, but may increase the gain for high and low frequencies for a low volume setting (e.g., to provide a loudness effect).

[0121] Such an approach may provide an improved listening experience in many embodiments.

[0122] In some embodiments, the metadata may include data indicative of a dependency of the reference audio reproduction frequency response on a playback level. Specifically, the metadata may not only provide data describing a reference audio reproduction frequency response but also include data describing how this may vary for different reproduction sound levels.

[0123] As a low complexity example, the metadata may include different reference audio reproduction frequency responses for different sound reproduction levels. For example, the metadata may include a first reference audio reproduction frequency response for a first (e.g., low) reference playback level and a second reference audio reproduction frequency response for a second (e.g., high) reference playback level. As another example, the metadata may include a reference audio reproduction frequency response with a gain for one or more frequencies being provided as a function of a sound/playback level.

[0124] The adapter 209 may then adapt the reference audio reproduction frequency response that is used for modifying the binaural transfer function in dependence on a current playback level for the output audio signal. For example, dependent on the current volume setting (e.g., whether it is a "low range" or a "high range" - e.g., whether it is below or above a setting of 5), the adapter 209 may select between different reference audio reproduction frequency responses provided in the metadata. As another example, in some embodiments, the actual current sound reproduction level may be measured (e.g., by a microphone) and the adapter 209 may proceed to determine various gains of the reference audio reproduction frequency response by evaluating the functions provided by the metadata for the given measured sound reproduction level.

[0125] Such an approach may provide a highly advantageous user experience in many embodiments and may in particular allow content side control or assistance in adapting render side systems to provide a desired audio reproduction.

[0126] FIG. 4 is a block diagram illustrating an example processor 400 according to embodiments of the disclosure. Processor 400 may be used to implement one or more processors implementing an apparatus as previously described or elements thereof. Processor 400 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.

[0127] The processor 400 may include one or more cores 402. The core 402 may include one or more Arithmetic Logic Units (ALU) 404. In some embodiments, the core 402 may include a Floating Point Logic Unit (FPLU) 406 and/or a Digital Signal Processing Unit

(DSPU) 408 in addition to or instead of the ALU 404.

[0128] The processor 400 may include one or more registers 312 communicatively coupled to the core 402. The registers 412 may be implemented using dedicated logic gate circuits (e.g., flip-flops) and/or any memory technology. In some embodiments the registers 412 may be implemented using static memory. The register may provide data, instructions and addresses to the core 402.

[0129] In some embodiments, processor 400 may include one or more levels of cache memory 410 communicatively coupled to the core 402. The cache memory 410 may provide computer-readable instructions to the core 402 for execution. The cache memory 410 may provide data for processing by the core 402. In some embodiments, the computer-readable instructions may have been provided to the cache memory 410 by a local memory, for example, local memory attached to the external bus 416. The cache memory 410 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.

[0130] The processor 400 may include a controller 414, which may control input to the processor 400 from other processors and/or components included in a system and/or outputs from the processor 400 to other processors and/or components included in the system. Controller 414 may control the data paths in the ALU 404, FPLU 406 and/or DSPU 408. Controller 414 may be implemented as one or more state machines, data paths and/or dedicated control logic. The gates of controller 414 may be implemented as standalone gates, FPGA, ASIC or any other suitable technology.

[0131] The registers 412 and the cache 410 may communicate with controller 414 and core 402 via internal connections 420A, 420B, 420C and 420D. Internal connections may be implemented as a bus, multiplexer, crossbar switch, and/or any other suitable connection technology.

[0132] Inputs and outputs for the processor 400 may be provided via a bus 416, which may include one or more conductive lines. The bus 416 may be communicatively coupled to one or more components of processor 400, for example the controller 414, cache 410, and/or register 412. The bus 416 may be coupled to one or more components of the system.

[0133] The bus 416 may be coupled to one or more external memories. The external memories may include Read Only Memory (ROM) 432. ROM 432 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) 433. RAM 433 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) 435. The external memory may

include Flash memory 434. The External memory may include a magnetic storage device such as disc 436. In some embodiments, the external memories may be included in a system.

[0134] 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.

[0135] 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.

[0136] 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.

[0137] 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 per-

formed 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 at least a first audio signal and metadata including:

a pose indication for an audio source of the first audio signal, and
frequency equalization data indicative of a reference audio reproduction frequency response for the first audio signal;

a listener pose processor (205) arranged to determine a listening pose;

a binaural transfer function processor (207) arranged to determine a binaural transfer function dependent on the listening pose and the pose indication for the audio source;

a processor (211) arranged to determine an audio rendering frequency response indicative of an audio reproduction frequency response of an audio reproduction path for the output audio signal;

an adapter (209) arranged to generate a binaural filter having a frequency response dependent on a combination of the audio rendering frequency response, the reference audio reproduction frequency response, and the binaural transfer function; and

a renderer (203) arranged to generate the output audio signal in dependence on the binaural filter.

2. The audio apparatus of claim 1 wherein the reference audio reproduction frequency response is a target frequency response for the first audio signal.

3. The audio apparatus of claim 1 wherein the reference audio reproduction frequency response is a source frequency response applied in generation of the first audio signal.

4. The audio apparatus of any previous claim wherein the frequency equalization data comprises an indication of an audio reproduction device, and the adapter (209) is arranged to determine the reference reproduction frequency response as a predetermined frequency response for the audio reproduction device.

tion device.

5. The audio apparatus of any previous claim wherein the frequency equalization data comprises data describing a Finite Impulse Response, FIR, filter having a frequency response matching the reference audio reproduction frequency response.

6. The audio apparatus of any previous claim wherein the frequency equalization data comprises data describing a combination of a set of sections, each section being a First Order Section or a Second Order Section and the combination having a frequency response matching the reference audio reproduction frequency response.

7. The audio apparatus of any previous claim wherein the frequency equalization data comprises data describing a set of parallel filters, the set of parallel filters having a frequency response matching the reference audio reproduction frequency response.

8. The audio apparatus of any previous claim wherein the frequency equalization data comprises frequency response identification data, and the adapter (209) is arranged to access a remote server to retrieve the reference reproduction frequency response based on the frequency response identification data.

9. The audio apparatus of any previous claim wherein the metadata further comprises a reference playback level; and wherein audio apparatus is arranged to adapt the frequency response of the binaural filter in dependence on a first play back level for a reproduction of the output audio signal relative to the reference playback level.

10. The audio apparatus of claim 9 wherein the metadata includes data indicating a dependency of the reference audio reproduction frequency response on a playback level and the adapter (209) is arranged to adapt the reference audio reproduction frequency response in dependence on the playback level.

11. An apparatus for generating an audio data signal, the apparatus comprising:

a receiver (301) arranged to receive at least a first audio signal from a first audio source having a pose in a scene;

a reproduction processor (303) arranged to determine a reference audio reproduction frequency response for the first audio signal;

a data signal generator (305) arranged to generate the audio data signal to comprise:

audio data for the first audio signal;

a pose indication indicative of the pose of the audio source; and
frequency equalization data indicative of the reference reproduction frequency response.

5

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

10

a pose indication for an audio source of the first audio signal, and
frequency equalization data indicative of a reference audio reproduction frequency response for the first audio signal;
determining a listening pose;
determining a binaural transfer function dependent on the listening pose and a pose indication for the audio source;
determining an audio rendering frequency response indicative of an audio reproduction frequency response of an audio reproduction path for the output audio signal;
generating a binaural filter having a frequency response dependent on a combination of the audio rendering frequency response, the reference audio reproduction frequency response, and the binaural transfer function; and
generating the output audio signal in dependence on the binaural filter.

15

20

25

30

- 13.** A method of generating an audio data signal, the method comprising:

35

receiving at least a first audio signal from a first audio source having a pose in a scene;
determining a reference reproduction frequency response for the first audio signal;
a data signal generator arranged to generate the audio data signal to comprise:

40

audio data for the first audio signal;
a pose indication indicative of the pose of the audio source; and
frequency equalization data indicative of the reference reproduction frequency response.

45

- 14.** An audio data signal comprising at least a first audio signal and metadata including

50

a pose indication for an audio source of the first audio signal; and
frequency equalization data indicative of a reference reproduction frequency response for the first audio signal.

55

- 15.** A computer program product comprising computer program code means adapted to perform all the steps of claim 13 or 14 when said program is run on a computer.

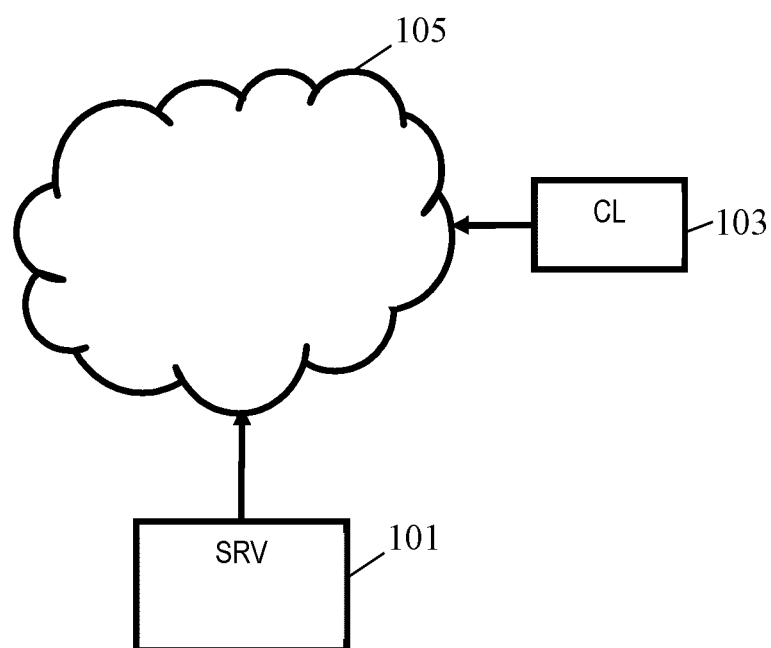
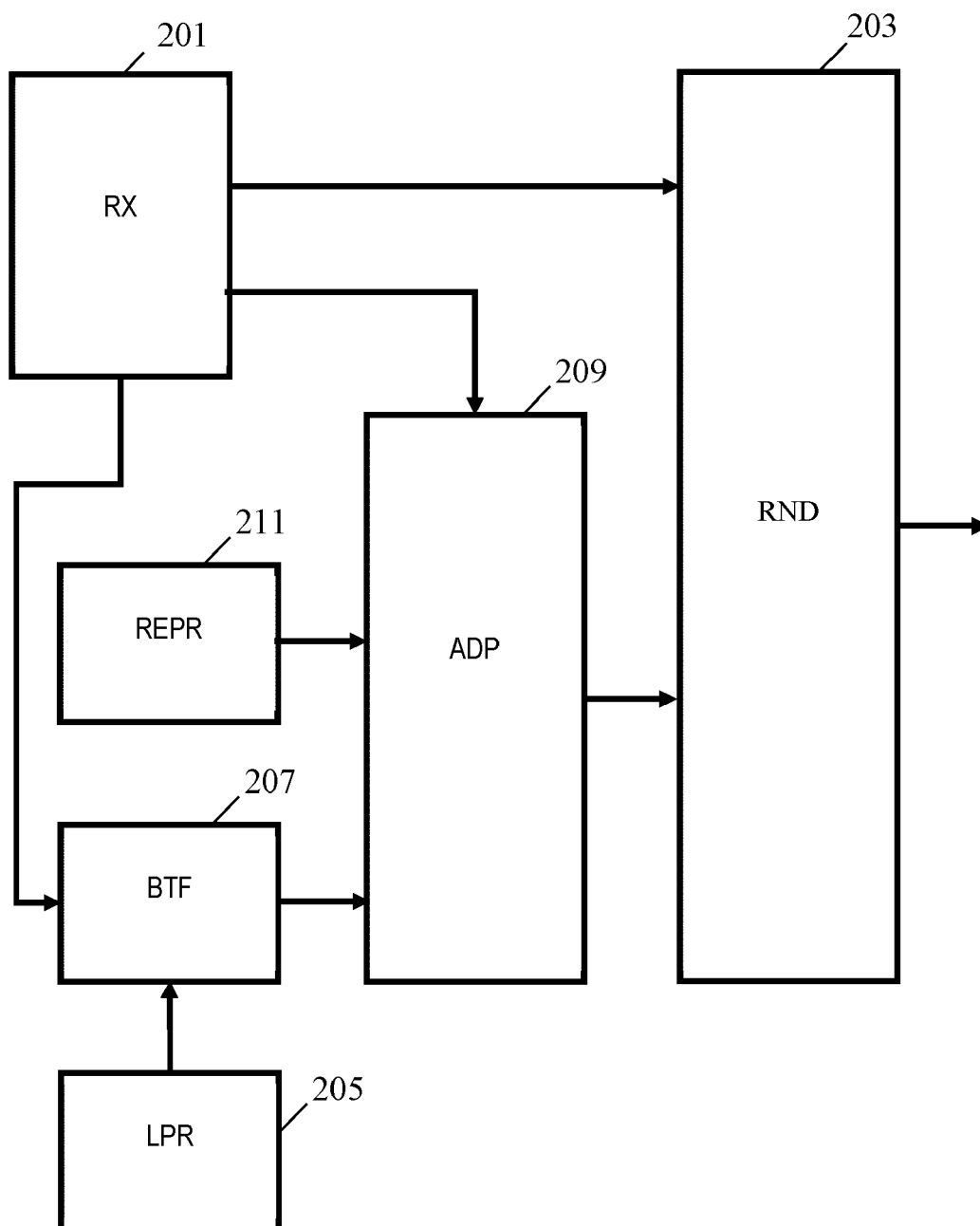


FIG. 1

**FIG. 2**

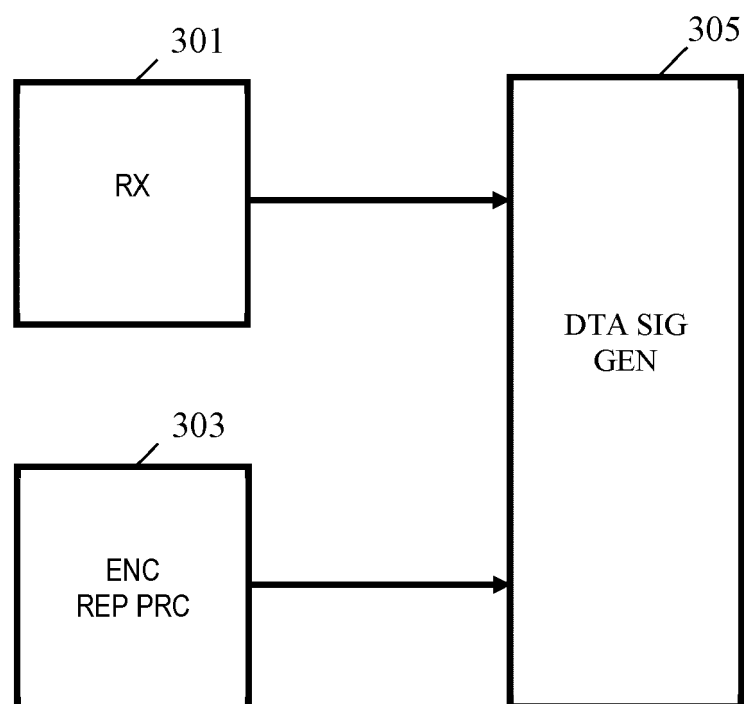
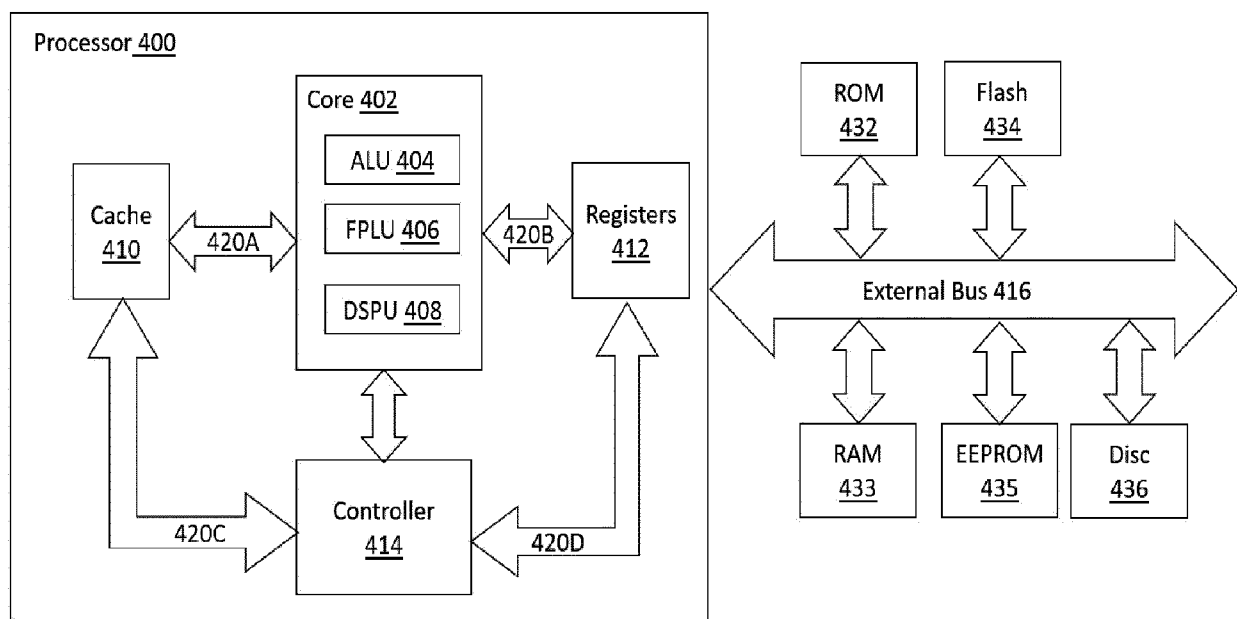


FIG. 3

**FIG. 4**



EUROPEAN SEARCH REPORT

Application Number

EP 23 18 7995

DOCUMENTS CONSIDERED TO BE RELEVANT

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
Y	US 2023/224661 A1 (LEE YONG JU [KR] ET AL) 13 July 2023 (2023-07-13)	11,13-15	INV.
A	* the whole document * -----	1-10,12	H04S3/00 H04S7/00
Y	US 2022/014869 A1 (LEE YONG JU [KR] ET AL) 13 January 2022 (2022-01-13)	11,13-15	
	* the whole document * -----		
Y	US 2017/032801 A1 (BAUMGARTE FRANK [US]) 2 February 2017 (2017-02-02)	11,13-15	
	* the whole document * -----		

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

EP 23 18 7995

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.

02-01-2024

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2023224661 A1	13-07-2023	KR 20230106986 A	14-07-2023
		US 2023224661 A1	13-07-2023

US 2022014869 A1	13-01-2022	KR 20220006744 A	18-01-2022
		US 2022014869 A1	13-01-2022

US 2017032801 A1	02-02-2017	CN 107851449 A	27-03-2018
		CN 111477253 A	31-07-2020
		EP 3329489 A1	06-06-2018
		EP 4290888 A2	13-12-2023
		JP 6499374 B2	10-04-2019
		JP 6750061 B2	02-09-2020
		JP 2018528459 A	27-09-2018
		JP 2019091095 A	13-06-2019
		KR 20180020295 A	27-02-2018
		KR 20200074243 A	24-06-2020
		US 2017032801 A1	02-02-2017
		US 2018218746 A1	02-08-2018
		US 2020342886 A1	29-10-2020
		WO 2017023423 A1	09-02-2017
