

(19)



(11)

EP 4 492 828 A1

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
15.01.2025 Bulletin 2025/03

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

(21) Application number: **23185362.3**

(52) Cooperative Patent Classification (CPC):
H04S 7/00; H04S 2400/13

(22) Date of filing: **13.07.2023**

(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:
• **Hoffman, Michael**
San Francisco, CA 94103 (US)
• **Ganpule, Sachin**
San Francisco, CA 94103 (US)
• **Kannen, Daniel Edward**
San Francisco, CA 94103 (US)
• **Vaughin, Kyle Moise**
San Francisco, CA 94103 (US)

(71) Applicant: **Dolby Laboratories Licensing Corporation**
San Francisco, CA 94103 (US)

(74) Representative: **MERH-IP Matias Erny Reichl Hoffmann**
Patentanwälte PartG mbB
Paul-Heyse-Straße 29
80336 München (DE)

(54) **METHODS, APPARATUS, AND SYSTEMS FOR PERFORMING AUDIO LEVEL CONTROL OF VIRTUAL REALITY SCENES**

(57) Systems, methods, and computer program products for providing audio metadata for a virtual scene are provided. A representation of the virtual scene is obtained, wherein the representation of the virtual scene comprises at least one audio source. An anchor sound source is determined from the at least one audio source.

A target perceived loudness of the anchor sound source is determined. The anchor sound source and the target perceived loudness of the anchor sound source are provided as the audio metadata for the virtual scene for encoding by an encoder.

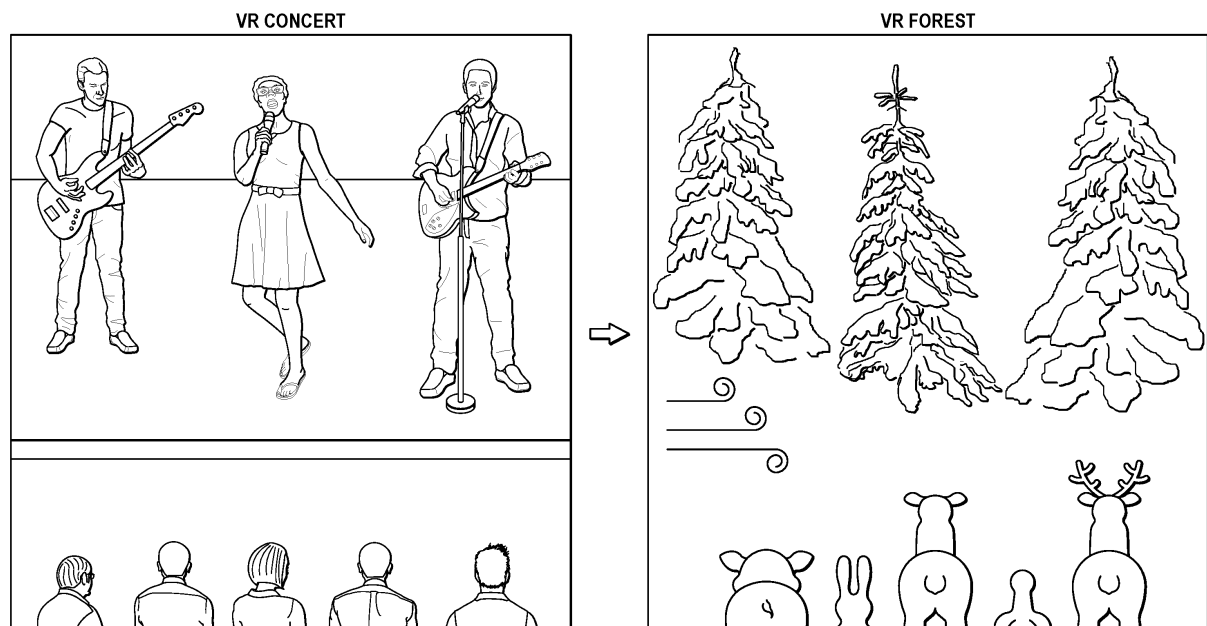


FIG. 1

Description

TECHNICAL FIELD

[0001] This disclosure pertains to systems, methods, and media for performing leveling control of audio elements in a virtual reality scene.

BACKGROUND

[0002] On a virtual reality rendering device, a consumer may switch between different virtual worlds/scenes, e.g., a forest and a concert. However, the providers of the different virtual worlds may be different, for example in a situation in which several worlds/scenes are accessible from a gateway scene. Thus, there exists a high likelihood that the sound levels of the different virtual worlds are not aligned. If the user experiences a change in the sound level between the different virtual worlds and possibly has to manually adjust the volume of his or her virtual reality rendering device, this may break the feeling of immersion and distract from the virtual reality experience. Thus, there is a need for improving the sound experience of a user of a virtual reality rendering device in particular in situations where the user switches between different virtual worlds.

[0003] Further, the different virtual worlds created by the content providers may be rendered on a wide array of consumer devices with different processing capabilities. Therefore, a virtual world created by a content provider may be too complex for a device type with low hardware capabilities. In addition, the audio level of a virtual world may cause listener hearing damage for a certain setting of a device type. Thus, there is a need for enabling rendering of a virtual world, irrespective of the hardware capabilities of the consumer devices, and at the same time preventing listener hearing damage when the consumer switches from one virtual world to another.

NOTATION AND NOMENCLATURE

[0004] Throughout this disclosure, including in the claims, the terms "audio source" and "audio object" are used synonymously to denote any sound-emitting object or person.

[0005] Throughout this disclosure, including in the claims, the expression "virtual scene" is understood as scene that can be rendered on a virtual reality or an augmented reality device. Further, the virtual scene may be a 6 degrees of freedom (6 DoF) virtual scene.

[0006] Throughout this disclosure including in the claims, a "virtual scene rendering device" is understood as any device that can render and playback a virtual scene.

[0007] Throughout this disclosure including in the claims, the expression "system" is used in a broad sense to denote a device, system, or subsystem. For example, a subsystem that implements a decoder may be referred to

as a decoder system, and a system including such a subsystem (e.g., a system that generates X output signals in response to multiple inputs, in which the subsystem generates M of the inputs and the other X - M inputs are received from an external source) may also be referred to as a decoder system.

[0008] Throughout this disclosure including in the claims, the term "processor" is used in a broad sense to denote a system or device programmable or otherwise configurable, such as with software or firmware, to perform operations on data, which may include audio, or video or other image data. Examples of processors include a field-programmable gate array (or other configurable integrated circuit or chip set), a digital signal processor programmed and/or otherwise configured to perform pipelined processing on audio or other sound data, a programmable general purpose processor or computer, and a programmable microprocessor chip or chip set.

SUMMARY

[0009] In view of the above, the present disclosure provides methods, apparatus, and programs, as well as computer-readable storage media for improving audio level control for virtual reality scenes, having the features of the respective independent claims.

[0010] One aspect of the disclosure relates to a method for providing audio metadata for a virtual scene. The method may include obtaining, i.e., receiving or creating, a representation of the virtual scene. The virtual scene may be a virtual reality scene or an augmented reality scene. The representation of the virtual scene may include at least one audio source. The at least one audio source may represent an object in the virtual scene, emitting an audio signal. Optionally, the representation of the virtual scene may include a location and a type of the at least one audio source. The type of the at least one audio source may be any one of an ambient sound, a voice or an instrument. An anchor audio source may be determined from the at least one audio source. The anchor audio source may be an audio source suitable for representing the acoustic environment of the virtual scene. A target perceived loudness of the anchor audio source may be determined. The target perceived loudness may be a loudness at which the anchor audio source should be perceived when the virtual scene is rendered at a virtual scene rendering device. The anchor audio source and the target perceived loudness of the anchor audio source may be provided as the audio metadata for the virtual scene for encoding by an encoder.

[0011] By providing metadata with an anchor audio source and a corresponding target perceived loudness, a virtual scene rendering device may be able to automatically level the loudness between different virtual scenes. Thereby a user experience while using a virtual scene rendering device may be improved, as a manual setting of the audio level after switching from one virtual

scene to another virtual scene may no longer be needed.

[0012] In some embodiments, determining an anchor audio source from the at least one audio source may include selecting the anchor audio source from the at least one audio source based on a relevancy of the at least one audio source for the virtual scene.

[0013] In some embodiments, the loudness at which the anchor audio source should be perceived may depend on a listener position in the virtual scene.

[0014] In some embodiments, determining a target perceived loudness of the anchor audio source may include measuring a loudness of the anchor audio source at multiple positions in the virtual scene. Further, the target perceived loudness of the anchor audio source may be determined based on the loudness of the anchor audio source at the multiple positions. Optionally, the multiple positions in the virtual scene may be positions in the virtual scene where a listener in the virtual scene is situated with a high likelihood.

[0015] In some embodiments, the method may further include encoding the representation of the virtual scene together with the audio metadata. Alternatively, the virtual scene itself may be encoded together with the audio metadata. The output of the encoding process may be an encoded bitstream.

[0016] In some embodiments, the method may further include assigning a priority level to each audio source of the at least one audio source. A highest priority level may be assigned to the anchor audio source. The priority level of each audio source may determine a rendering priority for a virtual scene rendering device.

[0017] Another aspect of the disclosure relates to a method for providing audio metadata for a virtual scene. The method may include obtaining, i.e., creating or receiving, a representation of the virtual scene. The representation of the virtual scene may comprise at least one audio source. A priority level may be assigned to each audio source of the at least one audio source. The priority level of each audio source may be provided as the audio metadata for the virtual scene for encoding by an encoder.

[0018] By providing different priority levels for different audio sources in a virtual scene, a virtual scene rendering device may choose the audio sources to be rendered based on the priority level and a criterion, such as a maximum loudness or a hardware constraint.

[0019] In some embodiments, the priority level may be assigned based on a relevancy of each audio source for the virtual scene. the priority level of each audio source may determine a rendering priority for a virtual scene rendering device.

[0020] In some embodiments, the method may further include encoding the representation of the virtual scene together with the audio metadata. Alternatively, the virtual scene itself may be encoded together with the audio metadata. The output of the encoding process may be an encoded bitstream.

[0021] Another aspect of the disclosure relates to a

method for controlling a perceived loudness between multiple virtual scenes. A virtual scene of the multiple virtual scenes may be a virtual reality scene or an augmented reality scene. The method may include obtaining audio metadata of each of the multiple virtual scenes. The audio metadata may include an anchor audio source and a target perceived loudness of the anchor audio source. The anchor audio source may be an audio source suitable for representing the acoustic environment of a virtual scene of the multiple virtual scenes. A loudness of a virtual scene rendering device may controlled based on the target perceived loudness of the anchor audio source in each of the multiple virtual scenes.

[0022] By controlling the loudness of the virtual scene rendering device in this way, the loudness between different virtual scenes may be automatically leveled. Thereby a user experience while using the virtual scene rendering device may be improved, as a manual setting of the audio level after switching from one virtual scene to another virtual scene may no longer be needed.

[0023] In some embodiments, obtaining audio metadata of each of the multiple virtual scenes may include decoding an encoded bitstream comprising the multiple virtual scenes and the corresponding audio metadata.

[0024] In some embodiments, controlling a loudness of a virtual scene rendering device based on the target perceived loudness of the anchor audio source in each of the multiple virtual scenes may include controlling a loudness of the virtual scene rendering device such that a loudness difference between the anchor audio source in each of the multiple virtual scenes corresponds to the difference between the target perceived loudness of the anchor audio source in each of the multiple virtual scenes.

[0025] In some embodiments, controlling a loudness of the virtual scene rendering device may include controlling the loudness of the anchor audio source in a current virtual scene of the multiple virtual scenes, wherein the current virtual scene is a virtual scene currently rendered by the virtual scene rendering device.

[0026] In some embodiments, controlling a loudness of the virtual scene rendering device may include controlling a loudness of an audio source other than the anchor audio source in a current virtual scene of the multiple virtual scenes based on the loudness of the anchor audio source.

[0027] In some embodiments, the method may further include rendering a virtual scene of the multiple virtual scenes.

[0028] In some embodiments, the method may further include switching from a current virtual scene rendered by the virtual scene rendering device, to another virtual scene of the multiple virtual scenes. The loudness of the anchor audio source in the other virtual scene may be controlled such that the difference in target perceived loudness between the anchor audio source of the current virtual scene and the anchor audio source of the other virtual scene is preserved.

[0029] In some embodiments, the audio metadata of each of the multiple virtual scenes may include a priority level for each audio source in a virtual scene of the multiple virtual scenes. The anchor audio source of each of the multiple virtual scenes may have a highest priority level. Optionally, the priority level may include a mandatory level or an optional level.

[0030] In some embodiments, rendering of an audio source in a virtual scene of the multiple virtual scenes based on the priority level may include rendering the audio source if the priority level assigned to the audio source is above a threshold. The threshold may depend on a hardware capability of the virtual scene rendering device. Alternatively or additionally, the threshold may depend on a target loudness of the virtual scene. Alternatively or additionally, the threshold may depend on a maximum safe listening level.

[0031] In some embodiments, rendering of an audio source in a virtual scene of the multiple virtual scenes based on the priority level may include rendering an audio source if the mandatory level is assigned to it. Further, an audio range of an audio source may be rendered and compressed if the optional level is assigned to it, if the audio source with the mandatory level assigned to it exceeds a loudness threshold.

[0032] Another aspect of the disclosure relates to a method for controlling a number of audio sources rendered by a virtual scene rendering device for a virtual scene. The method may include obtaining audio metadata of the virtual scene. The audio metadata may include a priority level assigned to each audio source of at least two audio sources in the virtual scene. The priority level may include a mandatory level or an optional level. Further, for each audio source of the at least two audio sources, the audio source may be rendered based on the priority level of the audio source.

[0033] By rendering the audio sources based on their priority level, different objectives can be achieved by the virtual scene rendering device, such as limiting the loudness to a maximum loudness or enable rendering of the virtual scene on a hardware constrained virtual scene rendering device.

[0034] In some embodiments, rendering the audio source based on the priority level of the audio source may include rendering the audio source if the priority level assigned to the audio source is above a threshold. The threshold may depend on a hardware capability of the virtual scene rendering device. Alternatively or additionally, the threshold may depend on a target loudness of the virtual scene. Alternatively or additionally, the threshold may depend on a maximum safe listening level.

[0035] In some embodiments, rendering of an audio source in a virtual scene of the multiple virtual scenes based on the priority level may include rendering an audio source if the mandatory level is assigned to it. Further, an audio range of an audio source may be rendered and compressed if the optional level is assigned to it, if the audio source with the mandatory level assigned to it

exceeds a loudness threshold.

[0036] It should be noted that the methods and systems including its preferred embodiments as outlined in the present disclosure may be used stand-alone or in combination with the other methods and systems disclosed in this document. Furthermore, all aspects of the methods and systems outlined in the present disclosure may be arbitrarily combined. In particular, the features of the claims may be combined with one another in an arbitrary manner.

[0037] It will be appreciated that apparatus features and method steps may be interchanged in many ways. In particular, the details of the disclosed method(s) can be realized by the corresponding apparatus, and vice versa, as the skilled person will appreciate. Moreover, any of the above statements made with respect to the method(s) (and, e.g., their steps) are understood to likewise apply to the corresponding apparatus (and, e.g., their blocks, stages, units), and vice versa.

[0038] Some or all of the operations, functions and/or methods described herein may be performed by one or more devices according to instructions (e.g., software) stored on one or more non-transitory media. Such non-transitory media may include memory devices such as those described herein, including but not limited to random access memory (RAM) devices, read-only memory (ROM) devices, etc. Accordingly, some innovative aspects of the subject matter described in this disclosure can be implemented via one or more non-transitory media having software stored thereon.

[0039] At least some aspects of the present disclosure may be implemented via an apparatus. For example, one or more devices may be capable of performing, at least in part, the methods disclosed herein. In some implementations, an apparatus is, or includes, an audio processing system having an interface system and a control system. The control system may include one or more general purpose single- or multi-chip processors, digital signal processors (DSPs), application specific integrated circuits (ASICs), field programmable gate arrays (FPGAs) or other programmable logic devices, discrete gates or transistor logic, discrete hardware components, or combinations thereof.

[0040] Details of one or more implementations of the subject matter described in this specification are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages will become apparent from the description, the drawings, and the claims. Note that the relative dimensions of the following figures may not be drawn to scale.

BRIEF DESCRIPTION OF THE DRAWINGS

[0041] The invention is explained below in an exemplary manner with reference to the accompanying drawings, wherein

Figure 1 is an illustrative example for two different

virtual worlds rendered in a virtual world rendering device.

Figure 2 is an illustrative schematic display of an example system for providing audio metadata for a virtual scene in accordance with some embodiments.

Figure 3 is an illustrative schematic display of another example system for providing audio metadata for a virtual scene in accordance with some embodiments.

Figure 4 is a block diagram of an example system for providing and utilizing audio metadata for virtual scenes in accordance with some embodiments.

Figure 5 is an illustrative example for assigned anchor sound sources in different virtual worlds in accordance with some embodiments.

Figure 6 is a block diagram of an example for using priority levels of audio source at a virtual scene rendering device in accordance with some embodiments.

Figure 7 is a block diagram of another example for using priority levels of audio source at a virtual scene rendering device in accordance with some embodiments.

Figure 8 is a flowchart of an example process for providing audio metadata for a virtual scene in accordance with some embodiments.

Figure 9 is a flowchart of another example process for providing audio metadata for a virtual scene in accordance with some embodiments.

Figure 10 is a flowchart of an example process for utilizing audio metadata for a virtual scene in accordance with some embodiments.

Figure 11 is a flowchart of another example process for utilizing audio metadata for a virtual scene in accordance with some embodiments.

[0042] Like reference numbers and designations in the various drawings indicate like elements.

DETAILED DESCRIPTION OF EMBODIMENTS

[0043] The rendering of virtual scenes through virtual reality or augmented reality poses many challenges. To provide a user with a wide variety of virtual content, it is expected that the virtual content will be created by various different artists or providers. Further, these artists and providers may use different tools to create the virtual

scenes and may have different methods to determine the loudness of different audio objects in the virtual scene. This may lead to an inconsistent audio level when a user switches between the different virtual scenes on a virtual scene rendering device.

[0044] Fig. 1 shows an example for two different virtual scenes to illustrate the problem of different audio levels in different virtual scenes. The first scene is a concert which may comprise many different sources of audio, such as a singer, different music instruments, and noise produced by the audience of the concert. An audio source may be seen as a sound emitted by a specific object or person, a by a group of objects or persons with similar properties. An audio source may be seen as a point source if the origin of the sound can be attributed to a small area or as an area/volume source if the sound is emitted over a relatively large area. The type of audio source may be an ambient sound, a voice, or an instrument, for example. The audience of the concert may be an example for the area/volume source, while the singer may be an example of a point audio source. The second scene in the example may be a forest scene, which comprises different audio sources, such as wind, rustling leaves and animal noises.

[0045] The expectation of the user, when switching from the first to the second scene, may be that the sound level change functions similar to a sound level change that would happen in real life, i.e., from a loud concert to comparatively quiet forest. Further, the user may have set the sound level of the virtual scene rendering device such that the sound level of the first scene is to the user's satisfaction. Under the assumption the concert and the forest have been created by different content providers, a similar tuning of the sound level of the virtual scene can however not be expected. Therefore, when switching from the first scene to the second scene, the sound level of the second virtual scene may either be too low or too high for the user's expectation. In both cases, the immersion experienced by the user will be broken and the user may have to manually change the sound level by operating the virtual scene rendering device. To solve this problem, dynamic range compression may be used in some scenarios. The dynamic range compression may however change the overall sound experience of the virtual scene. It is therefore an object of the present disclosure to provide a method for aligning the sound levels of different virtual scenes.

[0046] Further, when rendering a virtual scene on a virtual scene rendering device, the hardware capability of the device may be considered. Some virtual scenes may be more hardware demanding than other virtual scenes. For example, a virtual scene may have many audio sources so that rendering the virtual scene may not be possible by all virtual scene rendering devices. It is therefore also an object of this disclosure to provide a method for rendering complex virtual scenes on virtual scene rendering devices with relatively low hardware capability.

[0047] Fig. 2 depicts an illustrative schematic display of an example system for providing audio leveling informa-

tion at the creation side of the virtual scene. On the right side of the interface, a representation of the virtual scene is displayed. The representation of the virtual scene may be a representation of the virtual scene suitable for being displayed on a flat display, i.e., not a virtual scene rendering device. In other words, the interface may be a visual interface displayed on a generic, i.e., flat, display of a workstation, laptop, or a similar device. Alternatively, the interface may be a visual interface overlaid on the representation of virtual scene displayed by a virtual scene rendering device. Thereby, a creator or editor of the virtual scene may edit the metadata of the virtual scene while observing the virtual scene in the same way as a user of a virtual scene rendering device. Further, the representation may not represent all details of the final virtual scene. For example, the representation of the virtual scene may be focused on the audio sources in the virtual scene, and other objects effecting the audio experience. The representation of the virtual scene may comprise a location and a type of each audio source.

[0048] Optionally, the representation of the virtual scene may include likely positions of a user in the virtual scene. These positions may be predefined by for example limiting the movement of the user in the virtual scene. Additionally, or alternatively, the likely positions of the user in the virtual scene may be determined based on the content of the virtual scene. For example, a user will most likely be in front of the stage of a concert and not behind the stage of a concert. The determination of the likely positions of the user may further be based on known statistics of similar virtual scenes.

[0049] On the right side of the interface may be a list of all the audio sources in the virtual scene. The list of the audio sources may be an unordered list. In order to align the audio experience of different audio scenes, an anchor audio source may be determined. The anchor audio source may be determined from any audio source in the virtual scene. The anchor audio source should be an audio source that has a high relevancy for audio experience in the virtual scene. For example, at a virtual concert the most relevant audio source for the user experience may be a singer. The most relevant audio source or audio sources may predominantly determine whether a user of a virtual scene rendering device is satisfied with the overall loudness of the virtual scene. The anchor audio source may be determined based on the content of the virtual scene. Further, the anchor audio source may be determined based on statistics from similar virtual scenes. Alternatively, the creator of the virtual scene may select the anchor audio source from the list of audio sources.

[0050] Further, in order to be able to align the sound experience of different virtual scenes, a target perceived loudness may be assigned to the anchor audio source. The target perceived loudness may specify a specific loudness at which the anchor audio source should be perceived by the user of the virtual scene rendering device. The target perceived loudness may depend on

a user position in the virtual scene. The target perceived loudness may be determined by measuring the loudness of the anchor audio source. Optionally, measuring the loudness of the anchor audio source is performed at different locations of the virtual scene. The different locations of the virtual scene may be the likely positions of a user. By assigning a target perceived loudness to an anchor audio source in every virtual scene, the loudness of the different virtual scenes can be aligned at the virtual scene rendering device, i.e., the difference in perceived loudness of the anchor sound sources in two different virtual scenes, when switching between the virtual scenes, corresponds to the difference between the target perceived loudness of the two anchor audio sources. Therefore, the determination of an anchor audio source together with a target perceived loudness may be able to improve the sound experience of a user operating a virtual scene rendering device, when switching between different virtual scenes.

[0051] Optionally, the loudness of all other audio sources in a virtual scene, i.e., audio sources other than the anchor audio source, may be aligned with the target perceived loudness of the anchor sound source. That is, the loudness of an audio source may be expressed as a loudness difference or ratio to the target perceived loudness of the anchor audio source.

[0052] Fig. 3 depicts an illustrative schematic display of an example system for providing audio source priority information at the creation side of the virtual scene. The interface on the right side and on the left side may be identical to the interface depicted in Fig. 2. Therefore, identical elements will not be repeated again. In the interface of Fig. 3, different priority levels can be assigned to the different audio sources. A priority level may be binary, e.g., a mandatory and an optional priority level. Alternatively, the priority level may order the audio sources from least important to most important. The importance of an audio source for a virtual scene may depend on the content of the virtual scene. For example, in a concert, the singer and different instruments may be most important, while the crowd noise of the audience may be less important. The priority levels may be assigned based on the content of the virtual scene. Additionally, the priority levels may be assigned based on statistics of similar virtual scenes. Alternatively, the creator of the virtual scene may assign the priority of the audio sources.

[0053] By providing priority levels for the different audio sources, a virtual scene rendering device may determine which audio sources are to be rendered based on the priority level and the hardware capabilities of the virtual scene rendering device. Additionally, audio sources with a lower priority level may not be rendered or/and may be compressed in their dynamic range to prevent hearing damage at a virtual scene rendering device. Therefore, hearing damage may be prevented at a virtual scene rendering device without impacting the audio sources with a high priority level and thereby keeping the impact

on the audio experience at a minimum.

[0054] Further, the capabilities of the interfaces of **Fig. 2** and **Fig. 3** may be combined. In this case the anchor audio source may be an audio source with the highest priority level.

[0055] **Fig. 4** depicts an example system 100 comprising a creation side of the virtual scene as well as the rendering side of the virtual scene according to some implementations. The creation side of the virtual scene may include a metadata generator 101. A representation of a virtual scene may be the input of the metadata generator 101. As already noted with respect to **Fig. 2**, the representation of the virtual scene may be a representation focused on the audio aspects of the virtual scene. Then, the metadata generator 101 may generate metadata suitable for audio level control of the virtual scene at a virtual scene rendering device. The content of the metadata may be the data determined according to **Fig. 2** and/or **Fig. 3**. Accordingly, the metadata may include an anchor audio source and a corresponding target perceived loudness. Additionally, or alternatively, the metadata may include a priority level for each audio source in the virtual scene.

[0056] In a next step, the metadata output by the metadata generator 101 may be input to an encoder 102, together with the virtual scene. In contrast to the representation of the virtual scene, the virtual scene may have all necessary features in order to be rendered by a virtual scene rendering device. Encoder 102 may be any encoder suitable for encoding virtual scenes together with metadata. Encoder 102 encodes the virtual scene together with the metadata and outputs a bitstream.

[0057] The bitstream may be saved at a server (not shown) or may be directly transmitted to a receiving device. The receiving device may be a local device or a remote device.

[0058] The metadata generator 101 and the encoder 102 may be part of a virtual scene creation device. Alternatively, only metadata generator 101 may be part of a virtual scene creation device and the encoder 102 may be an external device. The virtual scene creation device may be a generic computer or a specialized computing device.

[0059] In a next step, the bitstream and other bitstreams also including an encoded virtual scene with metadata may be multiplexed by a multiplexer 103. The multiplexer 103 may be part of a network structure, e.g., a server.

[0060] In the following the receiving/rendering side for the virtual scene will be described. The multiplexed bitstream may be received by a demultiplexer 104. Demultiplexer 104 may reverse the operation of the multiplexer 103, i.e., demultiplex the bitstream. Demultiplexer 104 may put out multiple encoded bitstreams, each including an encoded virtual scene together with metadata.

[0061] A decoder 105 may receive the encoded bitstreams. Decoder 105 may decode all bitstreams or may only decode a subset of the bitstreams. For example,

decoder 105 may only decode a bitstream corresponding to a virtual scene that should be currently rendered by a virtual scene rendering device. In this case, decoder 105 may put out a virtual scene together with the corresponding metadata. Decoder 105 may be any decoder that is suitable for decoding virtual scenes together with metadata.

[0062] Finally, renderer 106 may receive the virtual scene together with the metadata. Renderer 106 may then render the virtual scene and control a loudness of a virtual scene rendering device based on the metadata. Details on the loudness control will be provided in the following with respect to **Fig. 5** to **Fig. 7**. The renderer 106, the decoder 105 and the demultiplexer 106 may be part of a virtual scene rendering device. Alternatively, only renderer 106 may be part of the virtual scene rendering device and decoder 105 and demultiplexer 106 may be part of an external device. The virtual scene rendering device may be a device that is capable of playing back audio data in the virtual scene, that is, the audio data corresponding to the audio sources in the virtual scene. To playback the audio data, the virtual scene rendering device may comprise headphones or any other audio playback device that may be used for experiencing audio data of a virtual scene.

[0063] **Fig. 5** depicts an illustrative example for switching from one virtual scene to another virtual scene by a virtual scene rendering device. The virtual scenes depicted in **Fig. 5** are the same virtual scenes depicted in the example of **Fig. 1**. The left scene may be a virtual scene currently viewed by a user on a virtual scene rendering device. For the left scene, the audio metadata may contain the information that the audio source corresponding to the singer of the band is the anchor audio source. The metadata further may contain a target perceived loudness for the singer. It is assumed that the user of the virtual scene rendering device has set the loudness to a satisfying level. In a next step, the user wants to switch to the right scene, i.e., the forest scene. In the forest scene, noises of an animal may be the anchor audio source. For the animal noise, also a target perceived loudness may be included in the audio metadata. When the virtual scene rendering device switches from the first scene to the right scene, a difference in the target perceived loudness of the singer and the animal noise may be determined. This audio level difference may then be applied to the current loudness setting of the virtual scene rendering device. By controlling the audio level in this way, each virtual scene to be rendered can be experienced in the intended audio level in relation to the first satisfactory audio setting by the user. The user therefore keeps immersed in the virtual world even with switching between different virtual scenes with completely different audio elements. In addition, no dynamic range compression is needed to align the audio level of the different virtual scenes.

[0064] **Fig. 6** shows an example for using priorities of audio sources at a virtual scene rendering device in

dependence of the hardware capability of the virtual scene rendering device in accordance with some embodiments. In this case, the virtual scene rendering device receives a priority level for each audio source in the audio metadata for the virtual scene. In the example of **Fig. 6** the audio sources in the virtual scene are either assigned the mandatory priority level or the optional priority level. Audio sources with the mandatory priority level may be assigned to audio sources that are essential for the content of the virtual scene, e.g., a band at a concert. The optional priority level may be assigned to audio sources that may not be essential for the audio experience, e.g., crowd noises at a concert.

[0065] When rendering the virtual scene, the virtual scene rendering device may render all audio sources with the mandatory priority level and may render *M* audio sources of the total number of audio sources with the optional priority level, wherein *M* denotes the number of audio sources that the virtual scene rendering device is able to render in addition to the audio sources with the mandatory priority level. In other words, the virtual scene rendering device may render additional audio sources with the optional priority level until 100% of allocated processing power for rendering is used. In still other words, the hardware capability of the virtual scene rendering device may define a threshold for the number of audio sources to be rendered.

[0066] Alternatively, the audio metadata may include a different priority system for the audio sources of the virtual scene. For example, the audio sources may be ordered from most important to least important. The virtual scene rendering device may then render the audio sources from most important to least important until 100% of allocated processing power for rendering is used.

[0067] By using the priority system in this way, a larger variety of virtual scene rendering devices may be able to render a particular virtual scene.

[0068] **Fig. 7** shows an example for using priorities of audio sources at a virtual scene rendering device to achieve a target loudness in accordance with some embodiments. As in the example of **Fig. 6**, the audio metadata may have two different priority levels assigned to the audio sources of the virtual scene, i.e., the mandatory level and the optional level. When the virtual scene rendering device detects that a virtual scene would exceed a target loudness at the current loudness setting of the virtual scene rendering device, only *M* of the total number of audio sources with the optional priority level would be rendered until the target loudness is reached. In addition, the *M* audio sources may additionally undergo dynamic range compression. In other words, the target loudness may define a threshold for the number of audio sources to be rendered.

[0069] Optionally, the target loudness may be a maximum allowed loudness to prevent hearing damage.

[0070] Optionally, when the audio sources with the mandatory audio level would already exceed the target

loudness, a dynamic range compression may be applied to the audio sources with the mandatory priority level. In addition, the dynamic range compression of the *M* audio sources with the optional audio level may be determined based on the dynamic range compression of the audio sources with the mandatory audio level such that the overall loudness does not exceed the target loudness.

[0071] Alternatively, the audio metadata may include a different priority system for the audio sources of the virtual scene. For example, the audio sources may be ordered from most important to least important. The virtual scene rendering device may then render the audio sources from most important to least important until 100% of the maximum allowed loudness is reached.

[0072] **Fig. 8** is a flowchart of an example process 200 for providing audio metadata for a virtual scene in accordance with some embodiments. In some implementations, blocks of process 200 may be performed by an encoder device. Alternatively, blocks of process 200 may be performed by a device without an encoding functionality.

[0073] In 202, process 200 may obtain a representation of a virtual scene. The representation of the virtual scene may comprise at least one audio source. A typical virtual scene may comprise multiple audio sources. As process 200 may provide means to align the audio level between different virtual scenes by providing a target perceived loudness for a single audio source, one audio source per virtual scene may be sufficient for process 200. The representation of the virtual scene may be the virtual scene itself or a representation focused on audio aspects of the virtual scene. Obtaining the representation of the virtual scene may be understood as creating the representation of the virtual scene. Alternatively, the representation of the virtual scene may be received.

[0074] In 204, process 200 may determine an anchor audio source from the at least one audio source. The anchor audio source may be the audio source most important for the acoustic experience for the virtual scene. The importance of an audio source may be determined by the creator of the virtual scene or by analyzing statistics of similar virtual scenes.

[0075] In 206, process 200 may determine a target perceived loudness of the anchor audio source. The target perceived loudness may be determined by measuring the loudness of the audio source in the virtual scene. Optionally, the loudness of the audio source may be measured at different positions in the virtual scene, which may be positions where a user is likely to be.

[0076] In 208, process 200 may provide the anchor audio source and the target perceived loudness of the anchor audio source as audio metadata for the virtual scene for encoding by an encoder. The encoder may encode the audio metadata and output an encoded audio bitstream.

[0077] **Fig. 9** is a flowchart of another example process 300 for providing audio metadata for a virtual scene in accordance with some embodiments. Processes 200

and 300 may be combined or performed independently. In some implementations, blocks of process 300 may be performed by an encoder device. Alternatively, blocks of process 300 may be performed by a device without an encoding functionality.

[0078] In 302, process 300 may obtain a representation of a virtual scene. The representation of the virtual scene may comprise at least one audio source. Details of the representation of the virtual scene and the number of audio sources are identical to process 200.

[0079] In 304, process 300 may assign a priority level to each audio source of the at least one audio source. The priority level may be assigned based on an importance of an audio source for the acoustic experience of the virtual scene. The importance may be determined based on statistics of similar virtual scenes. The priority level may comprise at least two different levels.

[0080] In 306, process 300 may provide the priority level of each audio source as the audio metadata for the virtual scene for encoding by an encoder. The encoder may encode the audio metadata and output an encoded audio bitstream.

[0081] Fig. 10 is a flowchart of an example process 400 for utilizing audio metadata for a virtual scene in accordance with some embodiments. In some implementations, blocks of process 400 may be performed by a decoding device. Alternatively, blocks of process 300 may be performed by a device without a decoding functionality.

[0082] In 402, process 400 may obtain audio metadata of each of multiple virtual scenes. The audio metadata may comprise an anchor audio source and a target perceived loudness of the anchor audio source. Audio metadata may be obtained by decoding an encoded audio bitstream or may be received by an external decoder.

[0083] In 404, process 400 may control a loudness of a virtual scene rendering device based on the target perceived loudness of the anchor audio source in each of the multiple virtual scenes. Controlling the loudness of the virtual device may be understood as controlling the overall loudness of the device and/or the loudness of particular audio sources in a virtual scene. To align the loudness of different virtual scenes, the target perceived loudness of the anchor elements in each virtual scene are compared. The difference between the target perceived loudness may then be used to control the loudness setting of the virtual scene rendering device when the user of the virtual scene rendering device switches from one virtual scene to another virtual scene. In other words, the loudness setting is either increased or decreased, depending on the difference between the target perceived loudness of the anchor audio sources in the two virtual scenes.

[0084] Fig. 11 is a flowchart of another example process 500 for utilizing audio metadata for a virtual scene in accordance with some embodiments. Processes 400 and 500 may be combined or performed independently.

In some implementations, blocks of process 500 may be performed by a decoding device. Alternatively, blocks of process 500 may be performed by a device without a decoding functionality.

[0085] In 502, process 500 may obtain audio metadata of a virtual scene. The audio metadata may comprise a priority level assigned to each audio source of at least two audio sources in the virtual scene. The priority level may be in the form as described in process 300. Audio metadata may be obtained by decoding an encoded audio bitstream or may be received by an external decoder.

[0086] In 504, process 500 may, for each audio source of the at least two audio sources, render the audio source based on the priority level of the audio source. An audio source may be rendered based on a hardware capability of a virtual scene rendering device. In other words, if the hardware capability of the virtual scene rendering device is not sufficient to render all audio sources in the virtual scene, the audio sources are rendered from highest to lowest priority level, until all processing resources of the virtual scene rendering device are utilized.

[0087] Alternatively, an audio source may be rendered based on a maximum loudness threshold. In other words, if the combined loudness of the audio sources in the virtual scene at a current loudness setting of the device would exceed the maximum loudness threshold, the audio sources are rendered from highest to lowest priority level, until the maximum loudness threshold is reached.

[0088] It should be noted that the processing at the virtual scene rendering device with respect to priority levels and anchor sound sources with target perceived loudness can be combined. In this case, the anchor sound source may have the highest priority level.

[0089] Some aspects of present disclosure include a system or device configured, e.g., programmed, to perform one or more examples of the disclosed methods, and a tangible computer readable medium, e.g., a disc, which stores code for implementing one or more examples of the disclosed methods or steps thereof. For example, some disclosed systems can be or include a programmable general purpose processor, digital signal processor, or microprocessor, programmed with software or firmware and/or otherwise configured to perform any of a variety of operations on data, including an embodiment of disclosed methods or steps thereof. Such a general purpose processor may be or include a computer system including an input device, a memory, and a processing subsystem that is programmed (and/or otherwise configured) to perform one or more examples of the disclosed methods (or steps thereof) in response to data asserted thereto.

[0090] Some embodiments may be implemented as a configurable (e.g., programmable) digital signal processor (DSP) that is configured (e.g., programmed and otherwise configured) to perform required processing on audio signal(s), including performance of one or more examples of the disclosed methods. Alternatively, embo-

diments of the disclosed systems (or elements thereof) may be implemented as a general purpose processor, e.g., a personal computer (PC) or other computer system or microprocessor, which may include an input device and a memory, which is programmed with software or firmware and/or otherwise configured to perform any of a variety of operations including one or more examples of the disclosed methods. Alternatively, elements of some embodiments of the inventive system are implemented as a general purpose processor or DSP configured (e.g., programmed) to perform one or more examples of the disclosed methods, and the system also includes other elements. The other elements may include one or more loudspeakers and/or one or more microphones. A general purpose processor configured to perform one or more examples of the disclosed methods may be coupled to an input device. Examples of input devices include, e.g., a mouse and/or a keyboard. The general purpose processor may be coupled to a memory, a display device, etc.

[0091] Another aspect of present disclosure is a computer readable medium, such as a disc or other tangible storage medium, which stores code for performing, e.g., by a coder executable to perform, one or more examples of the disclosed methods or steps thereof.

[0092] While specific embodiments of the present disclosure and applications of the disclosure have been described herein, it will be apparent to those of ordinary skill in the art that many variations on the embodiments and applications described herein are possible without departing from the scope of the disclosure described and claimed herein. It should be understood that while certain forms of the disclosure have been shown and described, the disclosure is not to be limited to the specific embodiments described and shown or the specific methods described.

[0093] Various aspects and implementations of the present disclosure may also be appreciated from the following enumerated example embodiments (EEEs), which are not claims.

EEE1. A method for providing audio metadata for a virtual scene, the method comprising:

obtaining a representation of the virtual scene, wherein the representation of the virtual scene comprises at least one audio source;
determining an anchor audio source from the at least one audio source;
determining a target perceived loudness of the anchor audio source;
providing the anchor audio source and the target perceived loudness of the anchor audio source as the audio metadata for the virtual scene for encoding by an encoder.

EEE2. The method of any previous EEE, wherein the virtual scene is a virtual reality scene or an augmen-

ted reality scene.

EEE3. The method of any previous EEE, wherein the representation of the virtual scene comprises a location and a type of the at least one audio source.

EEE4. The method of EEE 3, wherein the type of the at least one audio source is any one of an ambient sound, voice, or instrument.

EEE5. The method of any previous EEE, wherein obtaining a representation of the virtual scene comprises:

creating the representation of the virtual scene;
or
receiving the representation of the virtual scene.

EEE6. The method of any previous EEE, wherein the anchor audio source is an audio source suitable for representing the acoustic environment of the virtual scene.

EEE7. The method of any previous EEE, wherein the at least one audio source represents an object in the virtual scene, emitting an audio signal.

EEE8. The method of any previous EEE, wherein determining an anchor audio source from the at least one audio source comprises:
selecting the anchor audio source from the at least one audio source based on a relevancy of the at least one audio source for the virtual scene.

EEE9. The method of any previous EEE, wherein the target perceived loudness is a loudness at which the anchor audio source should be perceived when the virtual scene is rendered at a virtual scene rendering device.

EEE10. The method of EEE 9, wherein the loudness at which the anchor audio source should be perceived depends on a listener position in the virtual scene.

EEE11. The method of any previous EEE, wherein determining a target perceived loudness of the anchor audio source comprises:

measuring a loudness of the anchor audio source at multiple positions in the virtual scene;
and
determining the target perceived loudness of the anchor audio source based on the loudness of the anchor audio source at the multiple positions.

EEE12. The method of EEE 11, wherein the multiple

positions in the virtual scene are positions in the virtual scene where a listener in the virtual scene is situated with a high likelihood.

EEE13. The method of any previous EEE, wherein the method further comprises:
encoding the representation of the virtual scene together with the audio metadata to output an encoded bitstream.

EEE14. The method of any previous EEE, wherein the method further comprises:

assigning a priority level to each audio source of the at least one audio source; and
providing the priority level of each audio source to the audio metadata.

EEE15. The method of EEE 14, wherein a highest priority level is assigned to the anchor audio source.

EEE16. The method of EEEs 14 or 15, wherein the priority level of each audio source determines a rendering priority for a virtual scene rendering device.

EEE17. A method for providing audio metadata for a virtual scene, the method comprising:

obtaining a representation of the virtual scene, wherein the representation of the virtual scene comprises at least one audio source;
assigning a priority level to each audio source of the at least one audio source; and
providing the priority level of each audio source as the audio metadata for the virtual scene for encoding by an encoder.

EEE18. The method of EEE 17, wherein the priority level is assigned based on a relevancy of each audio source for the virtual scene.

EEE19. The method of EEEs 17 or 18, wherein the priority level of each audio source determines a rendering priority for a virtual scene rendering device.

EEE20. The method of any one of EEEs 17 to 19, wherein the method further comprises:
encoding the representation of the virtual scene together with the audio metadata to output an encoded bitstream.

EEE21. A method for controlling a perceived loudness between multiple virtual scenes, the method comprising:

obtaining audio metadata of each of the multiple

virtual scenes, wherein the audio metadata comprises an anchor audio source and a target perceived loudness of the anchor audio source; controlling a loudness of a virtual scene rendering device based on the target perceived loudness of the anchor audio source in each of the multiple virtual scenes.

EEE22. The method of EEE 21, wherein a virtual scene of the multiple virtual scenes is a virtual reality scene or an augmented reality scene.

EEE23. The method of EEEs 21 or 22, wherein obtaining audio metadata of each of the multiple virtual scenes comprises decoding an encoded bitstream comprising the multiple virtual scenes and the corresponding audio metadata.

EEE24. The method of any one of EEEs 21 to 23, wherein the anchor audio source is an audio source suitable for representing the acoustic environment of a virtual scene of the multiple virtual scenes.

EEE25. The method of any one of EEEs 21 to 24, wherein controlling a loudness of a virtual scene rendering device based on the target perceived loudness of the anchor audio source in each of the multiple virtual scenes comprises:
controlling a loudness of the virtual scene rendering device such that a loudness difference between the anchor audio source in each of the multiple virtual scenes corresponds to the difference between the target perceived loudness of the anchor audio source in each of the multiple virtual scenes.

EEE26. The method of any one of EEEs 21 to 25, wherein controlling a loudness of the virtual scene rendering device comprises controlling the loudness of the anchor audio source in a current virtual scene of the multiple virtual scenes, wherein the current virtual scene is a virtual scene currently rendered by the virtual scene rendering device.

EEE27. The method of any one of EEEs 21 to 26, wherein controlling a loudness of the virtual scene rendering device comprises controlling a loudness of an audio source other than the anchor audio source in a current virtual scene of the multiple virtual scenes based on the loudness of the anchor audio source.

EEE28. The method of any one of EEEs 21 to 27, wherein the method further comprises rendering a virtual scene of the multiple virtual scenes.

EEE29. The method of any one of EEEs 21 to 28, wherein the method further comprises switching from a current virtual scene rendered by the virtual

scene rendering device, to another virtual scene of the multiple virtual scenes.

EEE30. The method of EEE 29, wherein the loudness of the anchor audio source in the other virtual scene is controlled such that the difference in target perceived loudness between the anchor audio source of the current virtual scene and the anchor audio source of the other virtual scene is preserved.

EEE31. The method of any one of EEEs 21 to 30, wherein the audio metadata of each of the multiple virtual scenes comprises a priority level for each audio source in a virtual scene of the multiple virtual scenes.

EEE32. The method of EEE 31, wherein the anchor audio source of each of the multiple virtual scenes has a highest priority level.

EEE33. The method of any one of EEEs 31 to 32, wherein rendering a virtual scene of the multiple virtual scenes comprises rendering an audio source in a virtual scene of the multiple virtual scenes based on the priority level.

EEE34. The method of any one of EEEs 31 to 33, wherein the priority level comprises a mandatory level or an optional level.

EEE35. The method of any one of EEEs 33 to 34, wherein the rendering of an audio source in a virtual scene of the multiple virtual scenes based on the priority level comprises:
rendering the audio source if the priority level assigned to the audio source is above a threshold.

EEE36. The method of EEE 35, wherein the threshold depends on a hardware capability of the virtual scene rendering device.

EEE37. The method of EEE 35, wherein the threshold depends on a target loudness of the virtual scene.

EEE38. The method of EEE 35, wherein the threshold depends on a maximum safe listening level.

EEE39. The method of EEE 34, wherein the rendering of an audio source in a virtual scene of the multiple virtual scenes based on the priority level comprises:

rendering an audio source if the mandatory level is assigned to it; and
rendering and compressing an audio range of an audio source if the optional level is assigned to it, if the audio source with the mandatory level assigned to it exceeds a loudness threshold.

EEE40. A method for controlling a number of audio sources rendered by a virtual scene rendering device for a virtual scene, the method comprising:

obtaining audio metadata of the virtual scene, wherein the audio metadata comprises a priority level assigned to each audio source of at least two audio sources in the virtual scene; and for each audio source of the at least two audio sources:
rendering the audio source based on the priority level of the audio source.

EEE41. The method of EEE 40, wherein the priority level comprises a mandatory level or an optional level.

EEE42. The method of EEEs 40 or 41, wherein rendering the audio source based on the priority level of the audio source comprises:
rendering the audio source if the priority level assigned to the audio source is above a threshold.

EEE43. The method of EEE 42, wherein the threshold depends on a hardware capability of the virtual scene rendering device.

EEE44. The method of EEE 42, wherein the threshold depends on a target loudness of the virtual scene.

EEE45. The method of EEE 42, wherein the threshold depends on a maximum safe listening level.

EEE46. The method of EEE 41, wherein rendering the audio source based on the priority level of the audio source comprises:

rendering the audio source if the mandatory level is assigned to it; or
rendering and compressing an audio range of the audio source if the optional level is assigned to it, if an audio source with the mandatory level assigned to it exceeds a loudness threshold.

EEE47. An apparatus configured for implementing the method of any one of EEEs 1 to 46.

EEE48. A program comprising instructions that when executed by a processing device cause the processing device to carry out the method according to any one of EEEs 1 to 46.

EEE49. A computer-readable storage medium storing the program of EEE 48.

Claims

1. A method for providing audio metadata for a virtual scene, the method comprising:

obtaining a representation of the virtual scene, wherein the representation of the virtual scene comprises at least one audio source;
determining an anchor audio source from the at least one audio source;
determining a target perceived loudness of the anchor audio source;
providing the anchor audio source and the target perceived loudness of the anchor audio source as the audio metadata for the virtual scene for encoding by an encoder.

2. The method of any previous claim, wherein the anchor audio source is an audio source suitable for representing the acoustic environment of the virtual scene.

3. The method of any previous claim, wherein the target perceived loudness is a loudness at which the anchor audio source should be perceived when the virtual scene is rendered at a virtual scene rendering device.

4. The method of any previous claim, wherein determining a target perceived loudness of the anchor audio source comprises:

measuring a loudness of the anchor audio source at multiple positions in the virtual scene; and
determining the target perceived loudness of the anchor audio source based on the loudness of the anchor audio source at the multiple positions.

5. The method of any previous claim, wherein the method further comprises:
encoding the representation of the virtual scene together with the audio metadata to output an encoded bitstream.

6. A method for providing audio metadata for a virtual scene, the method comprising:

obtaining a representation of the virtual scene, wherein the representation of the virtual scene comprises at least one audio source;
assigning a priority level to each audio source of the at least one audio source; and
providing the priority level of each audio source as the audio metadata for the virtual scene for encoding by an encoder.

7. The method of claim 6, wherein the priority level of each audio source determines a rendering priority for a virtual scene rendering device.

8. The method of any one of claims 6 to 7, wherein the method further comprises:
encoding the representation of the virtual scene together with the audio metadata to output an encoded bitstream.

9. A method for controlling a perceived loudness between multiple virtual scenes, the method comprising:

obtaining audio metadata of each of the multiple virtual scenes, wherein the audio metadata comprises an anchor audio source and a target perceived loudness of the anchor audio source;
controlling a loudness of a virtual scene rendering device based on the target perceived loudness of the anchor audio source in each of the multiple virtual scenes.

10. The method of claim 9, wherein obtaining audio metadata of each of the multiple virtual scenes comprises decoding an encoded bitstream comprising the multiple virtual scenes and the corresponding audio metadata.

11. The method of any one of claims 9 to 10, wherein controlling a loudness of a virtual scene rendering device based on the target perceived loudness of the anchor audio source in each of the multiple virtual scenes comprises:
controlling a loudness of the virtual scene rendering device such that a loudness difference between the anchor audio source in each of the multiple virtual scenes corresponds to the difference between the target perceived loudness of the anchor audio source in each of the multiple virtual scenes.

12. The method of any one of claims 9 to 11, wherein controlling a loudness of the virtual scene rendering device comprises controlling the loudness of the anchor audio source in a current virtual scene of the multiple virtual scenes, wherein the current virtual scene is a virtual scene currently rendered by the virtual scene rendering device.

13. A method for controlling a number of audio sources rendered by a virtual scene rendering device for a virtual scene, the method comprising:

obtaining audio metadata of the virtual scene, wherein the audio metadata comprises a priority level assigned to each audio source of at least two audio sources in the virtual scene; and for each audio source of the at least two audio

sources:

rendering the audio source based on the priority level of the audio source.

- 14.** The method of claim 13, wherein rendering the audio source based on the priority level of the audio source comprises:
rendering the audio source if the priority level assigned to the audio source is above a threshold.

10

- 15.** The method of claim 14, wherein the threshold depends on a hardware capability of the virtual scene rendering device; or

the threshold depends on a target loudness of the virtual scene; or
the threshold depends on a maximum safe listening level.

20

25

30

35

40

45

50

55

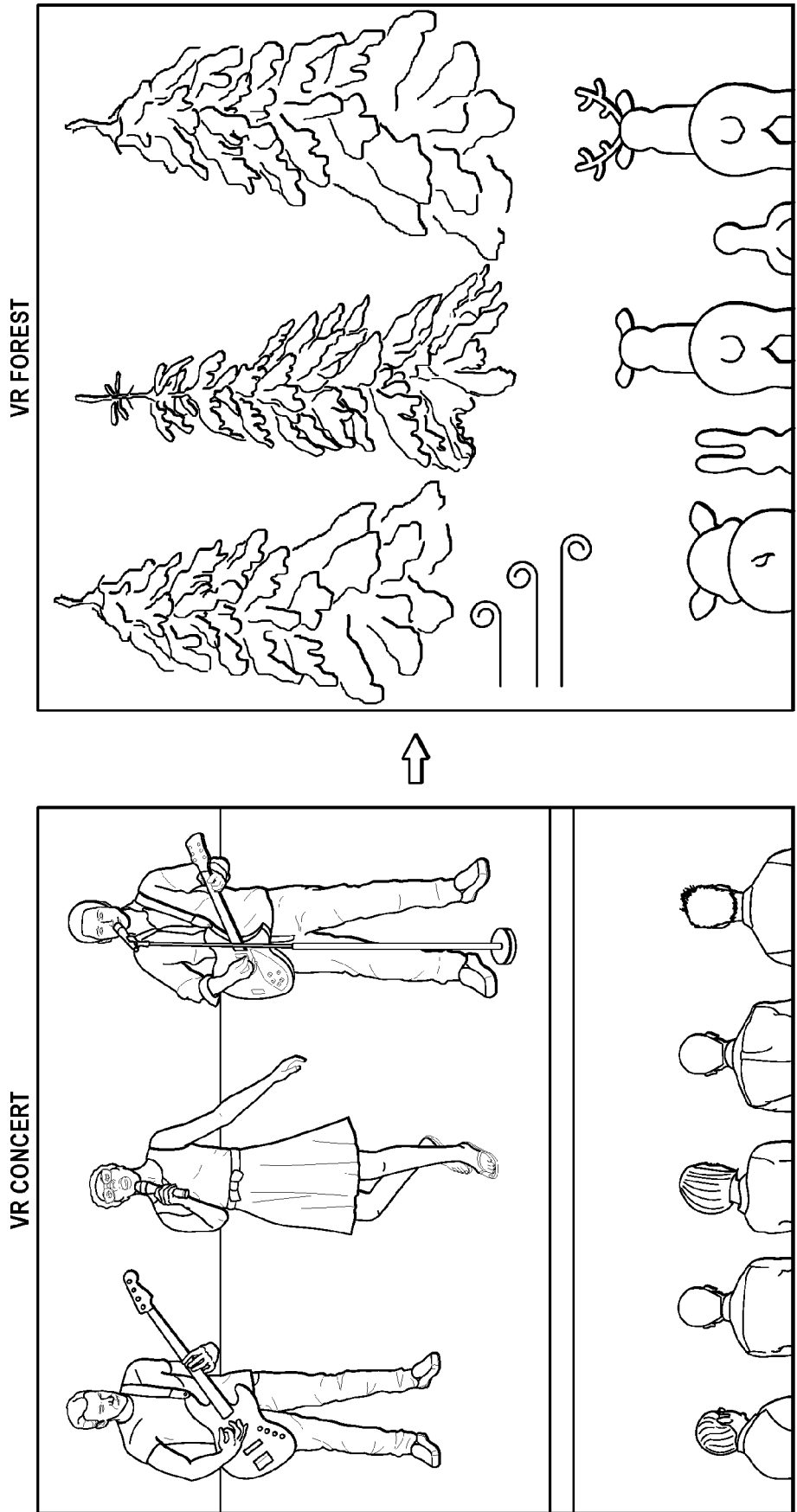


FIG. 1

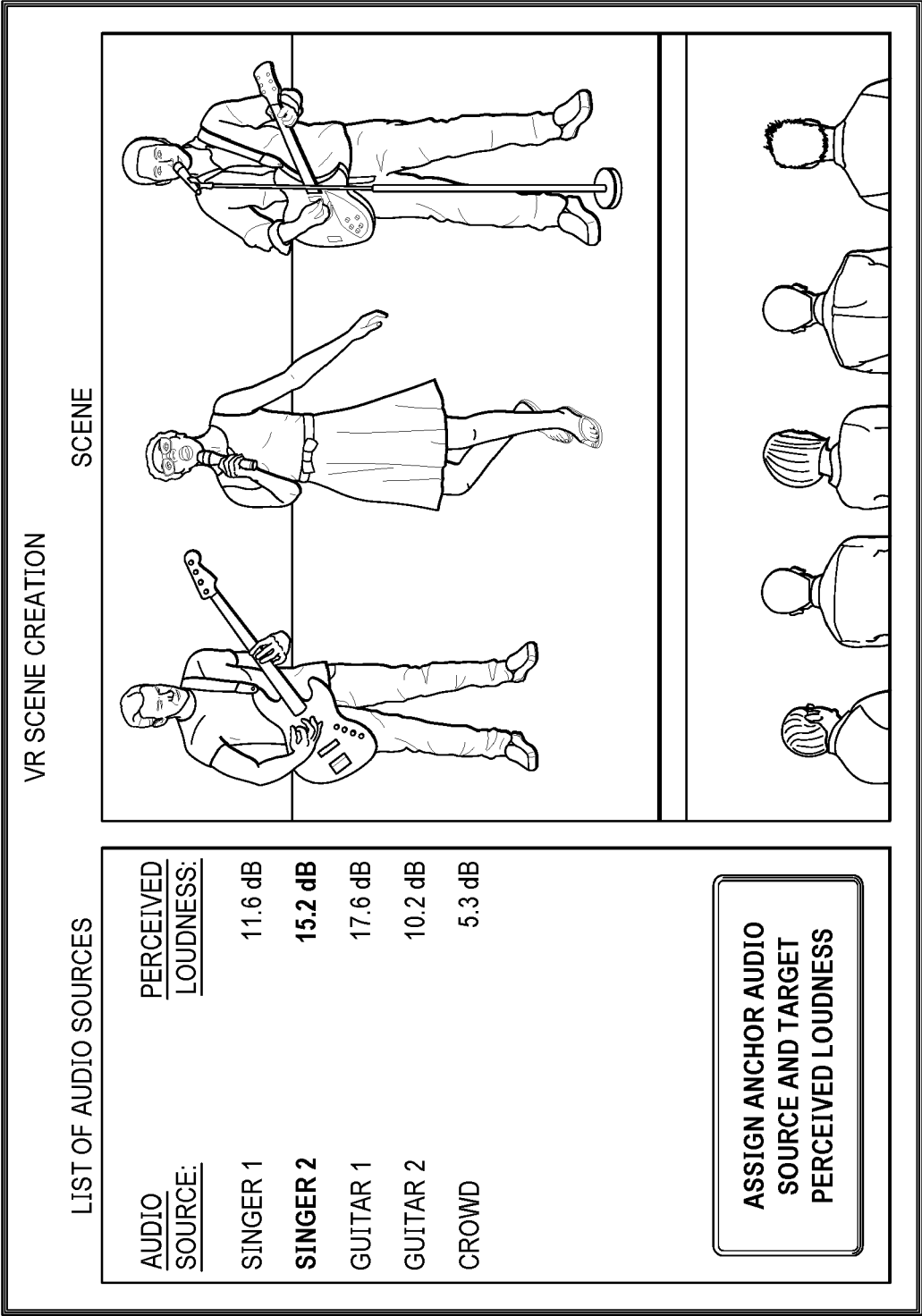


FIG. 2

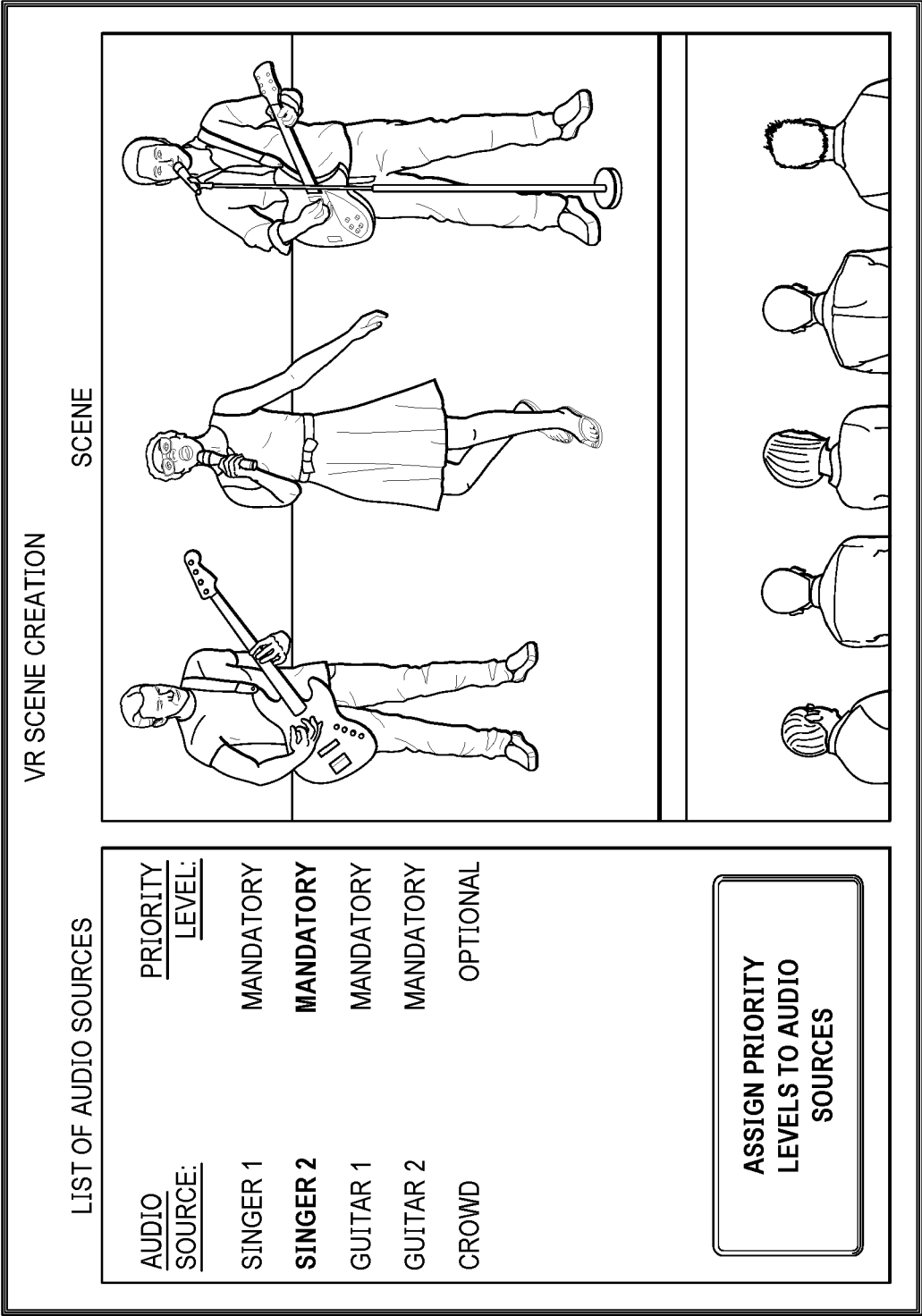


FIG. 3

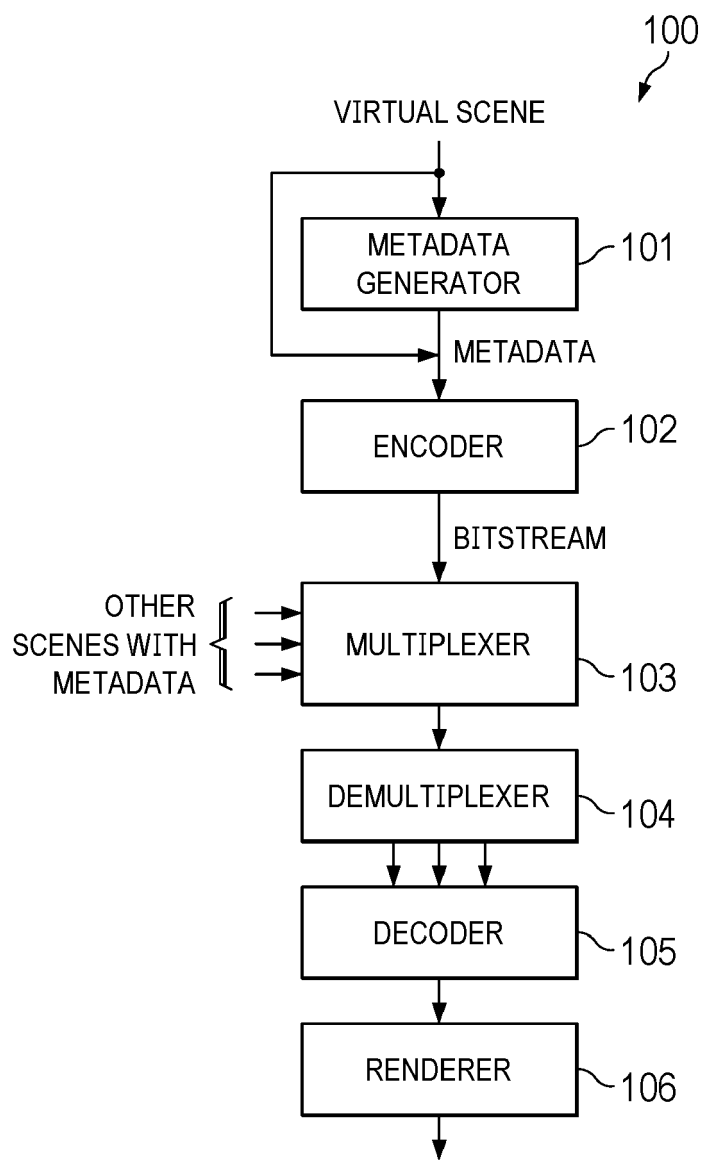


FIG. 4

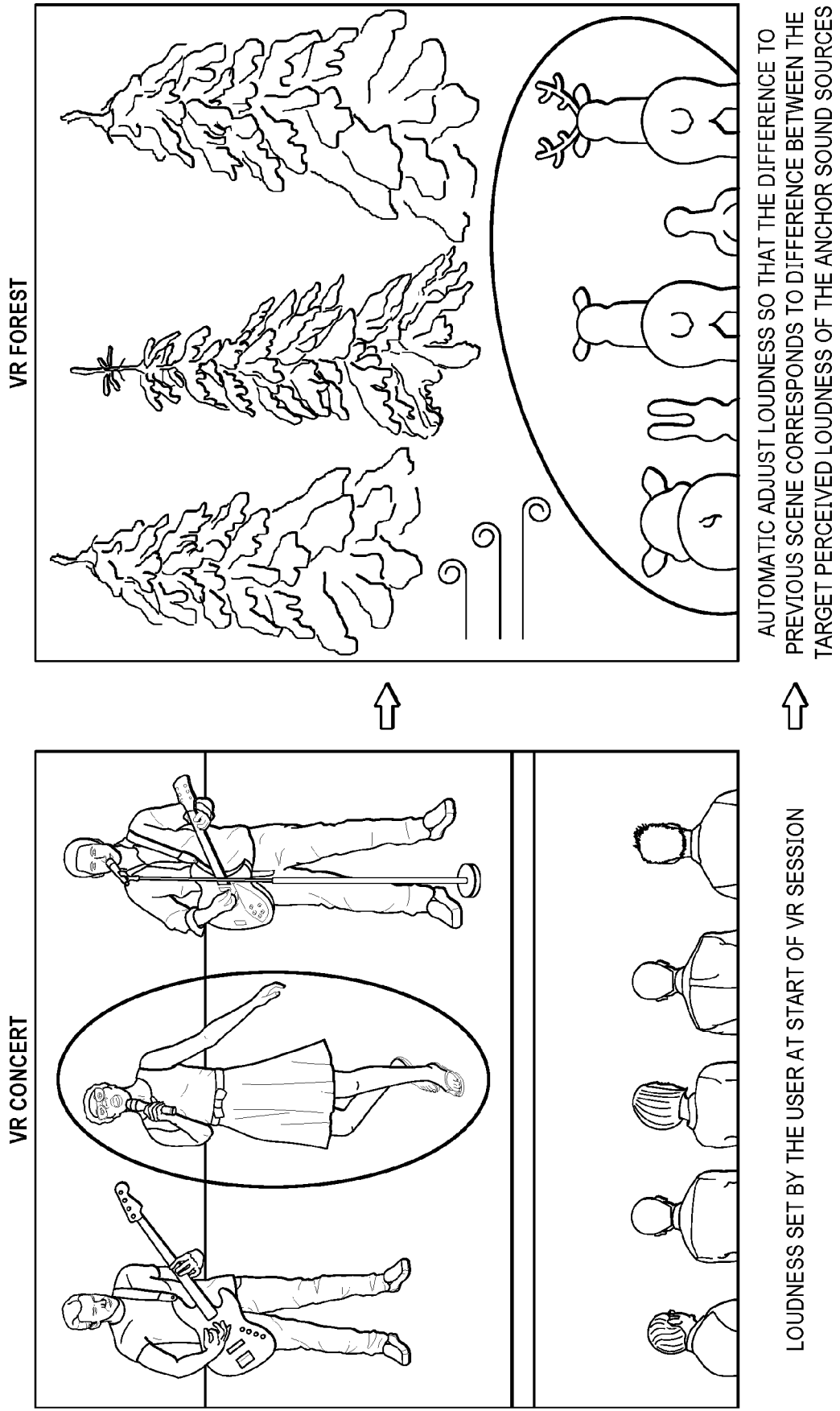


FIG. 5

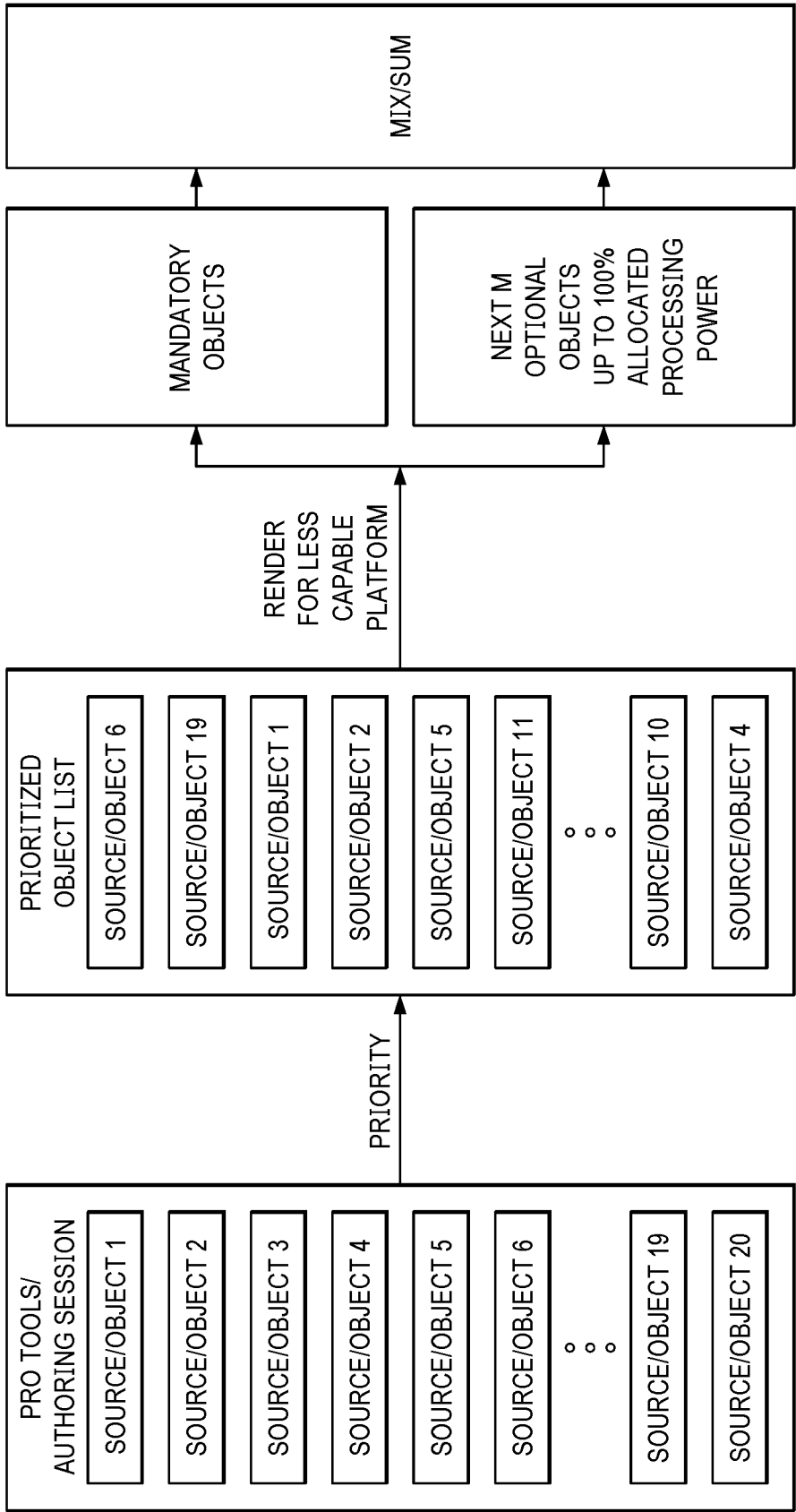


FIG. 6

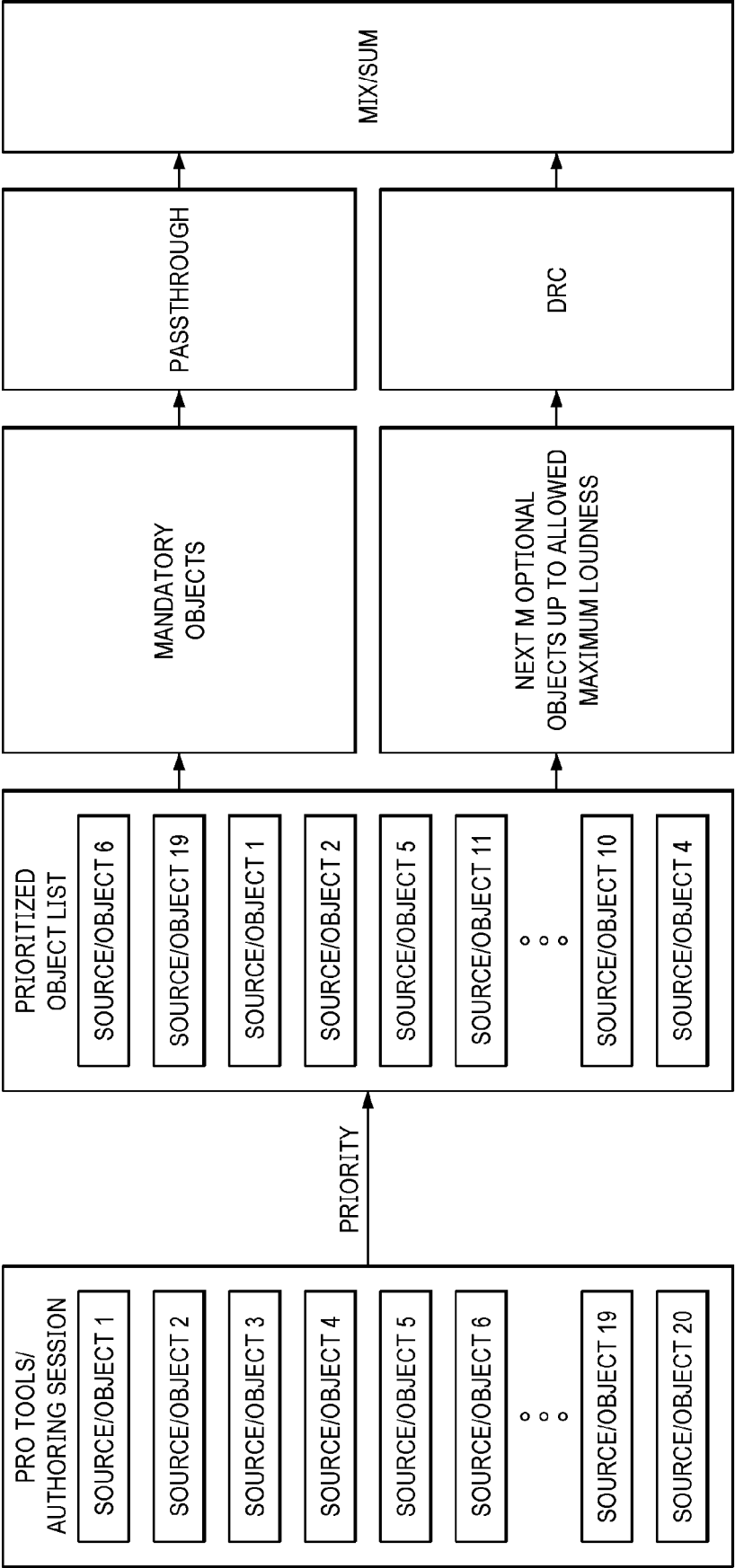


FIG. 7

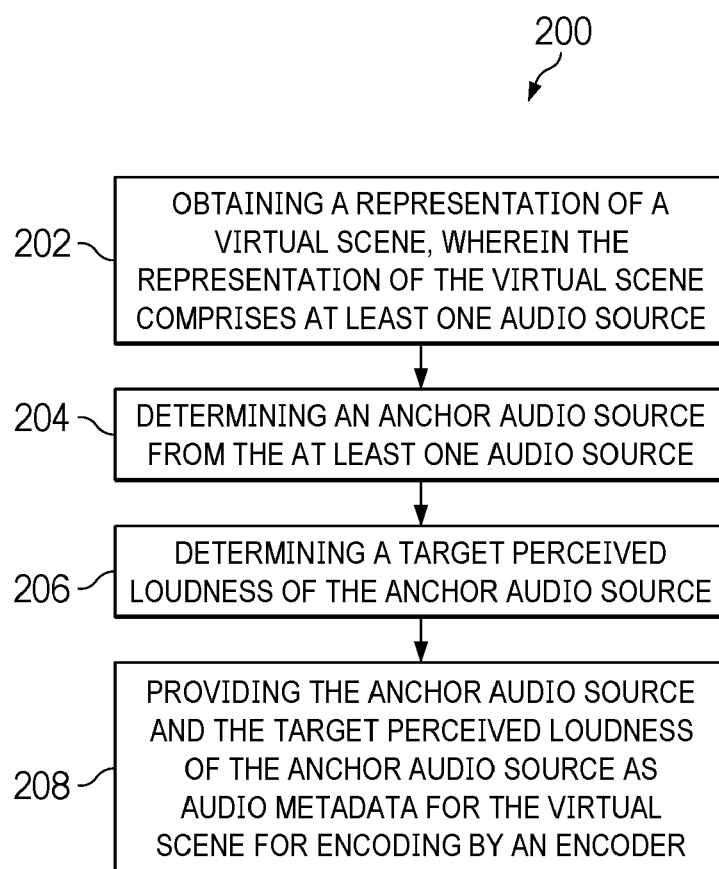


FIG. 8

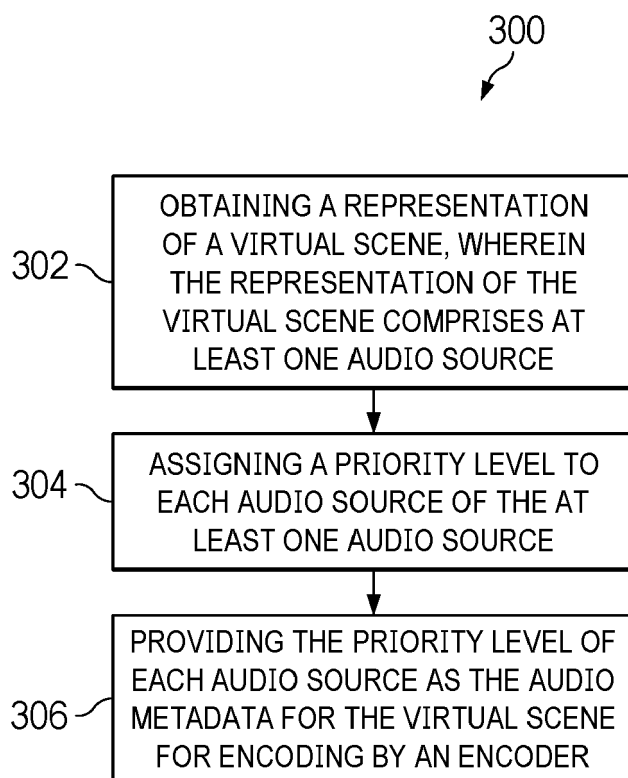


FIG. 9

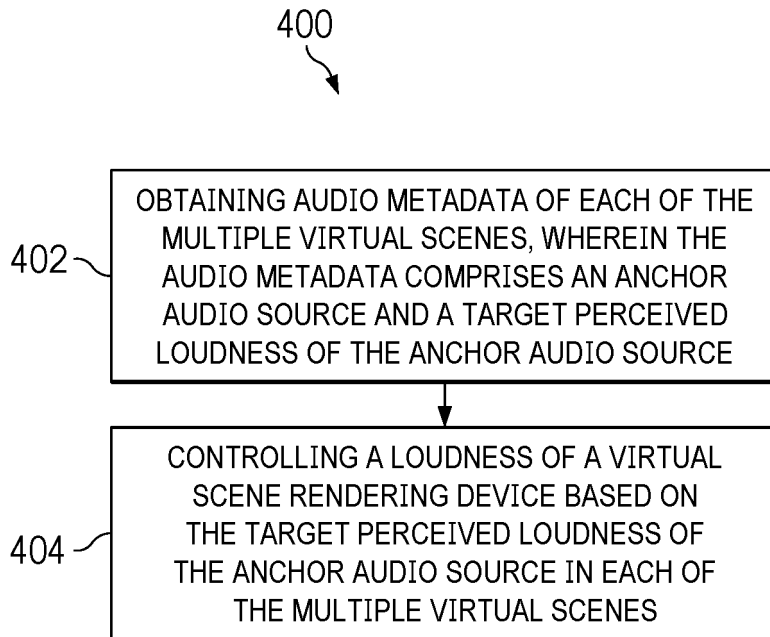


FIG. 10

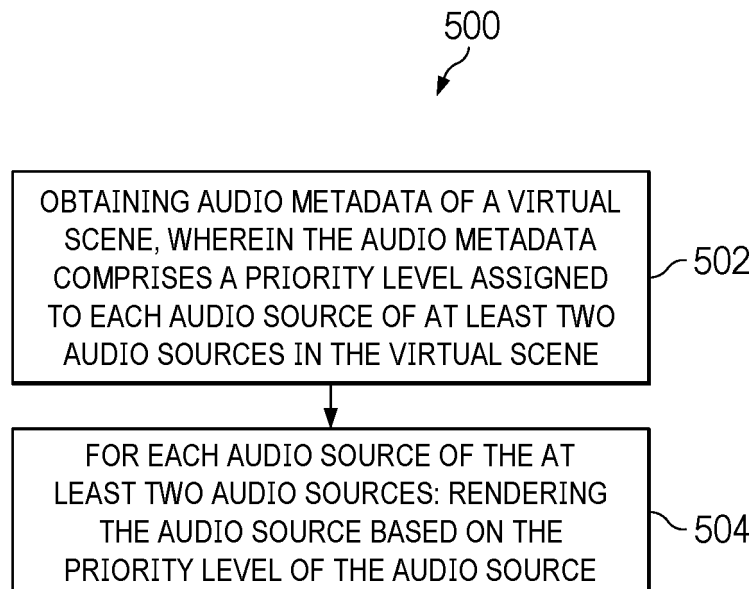


FIG. 11



EUROPEAN SEARCH REPORT

Application Number

EP 23 18 5362

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	"Loudness Management in MPEG-I Immersive Audio CFP", 134. MPEG MEETING; 20210426 - 20210430; ONLINE; (MOTION PICTURE EXPERT GROUP OR ISO/IEC JTC1/SC29/WG11), , no. n20436 4 May 2021 (2021-05-04), XP030294717, Retrieved from the Internet: URL:https://dms.mpeg.expert/doc_end_user/documents/134_OnLine/wg11/MDS20436_WG06_N00053.zip N0053_Loudness management.v5/N0053_Loudness management.v5.docx [retrieved on 2021-05-04] * the whole document *	1-5, 9-12	INV. H04S7/00
X	US 2022/286751 A1 (TIAN JUN [US] ET AL) 8 September 2022 (2022-09-08) * paragraphs [0007], [0013] * * paragraphs [0022] - [0024], [0031], [0048], [0049] * * paragraph [0078] - paragraph [0083]; figure 2 *	1-5, 9-12	TECHNICAL FIELDS SEARCHED (IPC) H04S G06F
X	US 2021/160644 A1 (OLIVIERI FERDINANDO [US] ET AL) 27 May 2021 (2021-05-27) * paragraphs [0004], [0006] * * paragraph [0092] - paragraph [0112]; figure 4A * * paragraph [0113] - paragraph [0128]; figure 4B *	6-8, 13-15	
The present search report has been drawn up for all claims			
Place of search		Date of completion of the search	Examiner
The Hague		2 February 2024	Betgen, Benjamin
CATEGORY OF CITED DOCUMENTS		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	
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			



EUROPEAN SEARCH REPORT

Application Number

EP 23 18 5362

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 2018/115850 A1 (DE BURGH MARK DAVID [AU] ET AL) 26 April 2018 (2018-04-26) * paragraph [0005] * * paragraph [0144] - paragraph [0148]; figure 9C * * paragraph [0110] - paragraph [0120]; figure 8 * -----	6-8, 13-15	
			TECHNICAL FIELDS SEARCHED (IPC)
The present search report has been drawn up for all claims			
Place of search		Date of completion of the search	Examiner
The Hague		2 February 2024	Betgen, Benjamin
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			

 2
 EPO FORM 1503 03.82 (P04C01)



Application Number

EP 23 18 5362**CLAIMS INCURRING FEES**

The present European patent application comprised at the time of filing claims for which payment was due.

☐ Only part of the claims have been paid within the prescribed time limit. The present European search report has been drawn up for those claims for which no payment was due and for those claims for which claims fees have been paid, namely claim(s):

☐ No claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for those claims for which no payment was due.

LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

see sheet B

☒ All further search fees have been paid within the fixed time limit. The present European search report has been drawn up for all claims.

☐ As all searchable claims could be searched without effort justifying an additional fee, the Search Division did not invite payment of any additional fee.

☐ Only part of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the inventions in respect of which search fees have been paid, namely claims:

☐ None of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims, namely claims:

☐ The present supplementary European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims (Rule 164 (1) EPC).

**LACK OF UNITY OF INVENTION
SHEET B**

Application Number

EP 23 18 5362

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

1. claims: 1-5, 9-12**Volume control in immersive audio encoding/rendering**

2. claims: 6-8, 13-15**Sound object prioritisation in immersive audio
encoding/rendering**

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

EP 23 18 5362

5

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.

02-02-2024

10

15

20

25

30

35

40

45

50

55

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2022286751 A1	08-09-2022	CN 115486096 A	16-12-2022
		EP 4101181 A1	14-12-2022
		JP 2023521689 A	25-05-2023
		KR 20220160638 A	06-12-2022
		US 2022286751 A1	08-09-2022
		WO 2022191876 A1	15-09-2022

US 2021160644 A1	27-05-2021	CN 114730564 A	08-07-2022
		EP 4062404 A1	28-09-2022
		US 2021160644 A1	27-05-2021
		WO 2021102132 A1	27-05-2021

US 2018115850 A1	26-04-2018	EP 3286929 A1	28-02-2018
		US 2018115850 A1	26-04-2018
		WO 2016172111 A1	27-10-2016

EPO FORM P0459

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