

[DESCRIPTION]

[Title of Invention] AUDIO SIGNAL PROCESSING METHOD, COMPUTER PROGRAM, AND AUDIO SIGNAL PROCESSING DEVICE

[Technical Field]

5 [0001]

The present disclosure relates to an audio signal processing method and the like.

[Background Art]

[0002]

10 In recent years, the spread of products and services that utilize extended reality (ER) (may be also expressed as "XR") including virtual reality (VR), augmented reality (AR), and mixed reality (MR) has advanced. Accompanying this, there has been growing demand for audio signal processing technologies that provide listeners with
15 immersive audio that, in a virtual space or a real-world space, assigns acoustic effects that are generated in accordance with the environment of the space to sounds emitted from a virtual sound source.

[0003]

20 Note that "listener" can also be expressed as "user". Furthermore, Patent Literature (PTL) 1, PTL 2, PTL 3, and Non Patent Literature (NPL) 1 disclose techniques that relate to the audio signal processing method and the like of the present disclosure.

[Citation List]

25 [Patent Literature]

[0004]

[PTL 1] Japanese Patent No. 6288100

[PTL 2] Japanese Unexamined Patent Application Publication No. 2019-22049

30 [PTL 3] WO Publication No. 2021/180938

[Non Patent Literature]

[0005]

[NPL 1] B.C.J. Moore, "An Introduction to the Psychology of Hearing", Seishin Shobo, 1994/4/20, Chapter 6: Space Perception, p. 225.

35 [Summary of Invention]

[Technical Problem]

[0006]

Incidentally, with the technique disclosed in PTL 1, it may be difficult to appropriately reduce the amount of computation and the computational load.

[0007]

Accordingly, an object of the present disclosure is to provide an audio signal processing method and the like that can appropriately reduce the amount of computation and the computational load.

[Solution to Problem]

[0008]

An audio signal processing method according to one aspect of the present disclosure is an audio signal processing method executed by an audio signal processing device, the audio signal processing method including: obtaining an audio signal including attribute information identifying an attribute of the audio signal; performing first determination processing and second determination processing when the attribute identified by the attribute information included in the audio signal obtained is information indicating an indirect sound, the first determination processing being processing to determine whether the audio signal obtained satisfies a first condition, the second determination processing being processing to determine whether the audio signal obtained satisfies a second condition different from the first condition; and outputting an output signal based on the audio signal obtained, when the audio signal obtained satisfies the first condition and the second condition.

[0009]

An audio signal processing method according to one aspect of the present disclosure is an audio signal processing method executed by an audio signal processing device, the audio signal processing method including: obtaining an audio signal indicating an indirect sound; performing, on the audio signal obtained, one or more first processes, first determination processing, second determination processing, and one or more second processes different from the one or more first processes; outputting an output signal based on the

audio signal obtained; calculating a first adjustment amount by
cumulatively adding adjustment amounts each determined by a
respective one of the one or more first processes, the adjustment
amounts each being an adjustment amount by which an amplitude of
5 the audio signal obtained is to be amplified; and calculating a second
adjustment amount by cumulatively adding adjustment amounts each
determined by a respective one of the one or more second processes,
the adjustment amounts each being an adjustment amount by which
the amplitude of the audio signal obtained is to be amplified, wherein
10 in the first determination processing, the audio signal obtained is
determined to satisfy a first condition when the first adjustment
amount calculated is greater than or equal to a first threshold value,
the second determination processing is not performed when the audio
signal obtained is determined not to satisfy the first condition, the
15 second determination processing is performed when the audio signal
obtained is determined to satisfy the first condition, the audio signal
obtained is determined to satisfy a second condition when, in the
second determination processing, a value based on the second
adjustment amount calculated is greater than or equal to a second
20 threshold value different from the first threshold value, and the
outputting is performed when the audio signal obtained is determined
to satisfy the second condition.

[0010]

Furthermore, a computer program according to one aspect of
25 the present disclosure causes a computer to execute the audio signal
processing method described above.

[0011]

An audio signal processing device according to one aspect of
the present disclosure is an audio signal processing device including:
30 an obtainer that obtains an audio signal including attribute
information identifying an attribute of the audio signal; a determiner
that performs first determination processing and second
determination processing when the attribute identified by the
attribute information included in the audio signal obtained is
35 information indicating an indirect sound, the first determination

processing being processing to determine whether the audio signal obtained satisfies a first condition, the second determination processing being processing to determine whether the audio signal obtained satisfies a second condition different from the first condition; and a reproducer that outputs an output signal based on the audio signal obtained, when the audio signal obtained satisfies the first condition and the second condition.

[0012]

Note that these comprehensive or specific aspects may be implemented as a system, a device, a method, an integrated circuit, a computer program, or a non-transitory computer-readable recording medium such as a CD-ROM, or may be implemented as any combination of a system, a device, a method, an integrated circuit, a computer program, and a recording medium.

[Advantageous Effects of Invention]

[0013]

The audio signal processing method and the like according to the one aspect of the present disclosure make it possible to appropriately reduce the amount of computation and the computational load.

[Brief Description of Drawings]

[0014]

[FIG. 1]

FIG. 1 is a diagram illustrating one example of a direct sound and reflected sounds generated in a sound space.

[FIG. 2]

FIG. 2 is a diagram illustrating an example of a three-dimensional sound reproduction system according to Embodiment 1.

[FIG. 3A]

FIG. 3A is a block diagram illustrating a configuration example of an encoding device according to Embodiment 1.

[FIG. 3B]

FIG. 3B is a block diagram illustrating a configuration example of a decoding device according to Embodiment 1.

[FIG. 3C]

FIG. 3C is a block diagram illustrating another configuration example of an encoding device according to Embodiment 1.

[FIG. 3D]

FIG. 3D is a block diagram illustrating another configuration example of a decoding device according to Embodiment 1.

[FIG. 4A]

FIG. 4A is a block diagram illustrating a configuration example of a decoder according to Embodiment 1.

[FIG. 4B]

FIG. 4B is a block diagram illustrating another configuration example of a decoder according to Embodiment 1.

[FIG. 5]

FIG. 5 is a diagram illustrating an example of a physical configuration of an audio signal processing device according to Embodiment 1.

[FIG. 6]

FIG. 6 is a diagram illustrating an example of a physical configuration of an encoding device according to Embodiment 1.

[FIG. 7]

FIG. 7 is a block diagram illustrating a configuration example of a renderer according to Embodiment 1.

[FIG. 8]

FIG. 8 is a flowchart illustrating an operation example of an audio signal processing device according to Embodiment 1.

[FIG. 9]

FIG. 9 is a diagram illustrating a comparatively distant positional relationship between a listener and an obstacle object.

[FIG. 10]

FIG. 10 is a diagram illustrating a comparatively close positional relationship between a listener and an obstacle object.

[FIG. 11]

FIG. 11 is a diagram illustrating relationships between time differences between direct sounds and reflected sounds, and threshold values.

[FIG. 12A]

FIG. 12A is a diagram illustrating a part of an example of a method for setting threshold value data.

[FIG. 12B]

FIG. 12B is a diagram illustrating a part of an example of a method for setting threshold value data.

[FIG. 12C]

FIG. 12C is a diagram illustrating a part of an example of a method for setting threshold value data.

[FIG. 13]

FIG. 13 is a diagram illustrating an example of a threshold value setting method.

[FIG. 14]

FIG. 14 is a flowchart illustrating an example of selection processing.

[FIG. 15]

FIG. 15 is a diagram illustrating relationships between directions of direct sounds, directions of reflected sounds, time differences, and threshold values.

[FIG. 16]

FIG. 16 is a diagram illustrating relationships between angular differences, time differences, and threshold values.

[FIG. 17]

FIG. 17 is a block diagram illustrating another configuration example of a renderer.

[FIG. 18]

FIG. 18 is a flowchart illustrating another example of selection processing.

[FIG. 19]

FIG. 19 is a flowchart illustrating yet another example of selection processing.

[FIG. 20]

FIG. 20 is a flowchart illustrating a first variation of operations of an audio signal processing device according to Embodiment 1.

[FIG. 21]

FIG. 21 is a flowchart illustrating a second variation of

operations of an audio signal processing device according to Embodiment 1.

[FIG. 22]

FIG. 22 is a diagram illustrating an arrangement example of an avatar, a sound source object, and an obstacle object.

[FIG. 23]

FIG. 23 is a flowchart illustrating yet another example of selection processing.

[FIG. 24]

FIG. 24 is a block diagram illustrating a configuration example for a renderer to perform pipeline processing.

[FIG. 25]

FIG. 25 is a diagram illustrating transmission and diffraction of a sound.

[FIG. 26]

FIG. 26 is a block diagram illustrating a configuration example of a renderer according to Embodiment 2.

[FIG. 27]

FIG. 27 is a flowchart illustrating an operation example of an audio signal processing device according to Embodiment 2.

[FIG. 28]

FIG. 28 is a graph showing threshold value data indicating a first threshold value according to Embodiment 2.

[FIG. 29]

FIG. 29 is a block diagram illustrating a configuration example of a renderer according to Embodiment 3.

[FIG. 30]

FIG. 30 is a block diagram illustrating a configuration example of a renderer according to Embodiment 4.

[FIG. 31]

FIG. 31 is a diagram illustrating the energy of the head-related transfer function according to Embodiment 4.

[FIG. 32]

FIG. 32 is a diagram illustrating the energy of the head-related transfer function according to Embodiment 4.

[FIG. 33]

FIG. 33 is a diagram illustrating the energy of the head-related transfer function according to Embodiment 4.

[FIG. 34]

5 FIG. 34 is a diagram illustrating the energy of the head-related transfer function according to Embodiment 4.

[FIG. 35]

FIG. 35 is a flowchart illustrating an operation example of an audio signal processing device according to Embodiment 4.

10 [FIG. 36]

FIG. 36 is a block diagram illustrating a configuration example of a renderer according to Embodiment 5.

[FIG. 37]

15 FIG. 37 is a diagram showing a table for illustrating the effects of changing the first threshold value and performing disabling processing, according to Embodiment 5.

[Description of Embodiments]

[0015]

(Underlying Knowledge Forming Basis of the Present Disclosure)

20 To date, investigations have been made into audio signal processing technologies that provide listeners with immersive audio by, in a virtual space or a real-world space, assigning acoustic effects that are generated in accordance with the environment of the space to sounds emitted from a virtual sound source.

25 [0016]

PTL 1 discloses such an audio signal processing technique. More specifically, PTL 1 discloses a technique for detecting the importance of audio signals (voice signals) and not outputting audio signals for which the detected importance is low. By thus not
30 outputting audio signals for which the importance is low, there are expectations for the audio signal processing technique to appropriately reduce the amount of computation and the computational load.

[0017]

35 Incidentally, in a sound space (a virtual space or a real-world

space), reflected sound is sometimes important.

[0018]

FIG. 1 is a diagram illustrating one example of a direct sound and reflected sounds generated in a sound space. In acoustic processing in which characteristics of a virtual space are expressed by a sound, it is effective to reproduce not only direct sounds, but also reflected sounds in order to express the size of the space, the material of the walls, and the like, as well as to allow for accurately grasping the location of the sound source (the positioning of the sound image).

[0019]

For example, when a sound is heard in a rectangular parallelepiped room such as that in FIG. 1, six primary reflected sounds, corresponding to the six walls, are generated for one sound source. Reproducing these reflected sounds provides a clue for appropriate understanding of the space and the sound image. Furthermore, for each reflected sound, a secondary reflected sound is generated by a surface other than the reflection surface that generated that reflected sound. These reflected sounds are also effective sensory clues.

[0020]

However, even when consideration is given no further than to secondary reflection, one direct sound and 36 ($6 + 6 \times 5$) reflected sounds are generated for one sound source. Thus, 37 sound rays are generated, and processing these sound rays requires a significant amount of computation.

[0021]

Furthermore, in applied products in recent years for which metaverses are imagined, such as virtual meetings, virtual shopping, virtual concerts, and the like, a plurality of sound sources are present out of necessity, whereby an even greater amount of computation is required.

[0022]

Moreover, the listener hearing the sounds in a virtual space uses headphones or VR goggles. In order to provide three-

dimensional sound to such a listener, binaural processing that assigns a sound pressure ratio and a phase difference between the two ears and reproduces the direction of arrival and distance sensation of the sounds is performed on each sound ray. Thus, if an attempt were
5 made to reproduce every reflected sound that is generated, the amount of computation would become immense.

[0023]

On the other hand, in light of convenience, a small storage battery is sometimes used as the battery for the VR goggles worn by
10 the listener who experiences the virtual space. Lessening the computational load resulting from the above-described processing makes it possible to further extend the life of the storage battery. To this end, the number of sound rays, which are emitted on a scale of hundreds, is desirably reduced, within a scope at which grasping the
15 space and the positioning of the sounds is not harmed.

[0024]

Furthermore, in a system that reproduces acoustics, a degree of freedom such as 6DoF (6 degrees of freedom) or the like may be allowed with respect to the position (in other words, the listening
20 position, which is the position where the listener is present) and orientation of the listener. In this case, the positional relationship between the listener, the sound sources, and the objects that reflect sounds cannot be fixed until the time of reproduction (the time of rendering). For this reason, the reflected sounds as well cannot be
25 fixed until the time of reproduction. Thus, it is difficult to determine the reflected sounds to be processed beforehand.

[0025]

Therefore, during reproduction, appropriately selecting and outputting (reproducing) one or more reflected sounds, from a
30 plurality of reflected sounds that are generated in a sound space, that are to be processed or are not to be processed is useful in appropriately reducing the amount of computation and the computational load.

[0026]

35 It should be noted that controlling whether to select a sound

corresponds to determining whether to select the sound, and more specifically corresponds to determining whether to select and output (reproduce) the sound. Furthermore, selecting a sound may be selecting the sound as a sound to be processed, or may be selecting the sound as a sound that is not to be processed.

[0027]

Incidentally, in PTL 1, the importance of the audio signal, and more specifically the importance of the direct sound that the audio signal represents, is detected, but the importance of reflected sound is not considered. Therefore, when indirect sounds such as reflected sounds are generated, as illustrated in FIG. 1, the amount of computation and the computational load increase; in other words, it may be difficult to appropriately reduce the amount of computation and the computational load.

[0028]

Therefore, there is a need for an audio signal processing method and the like that can appropriately reduce the amount of computation and the computational load in a sound space.

[0029]

Accordingly, an audio signal processing method according to a first aspect of the present disclosure is an audio signal processing method executed by an audio signal processing device, the audio signal processing method including: obtaining an audio signal including attribute information identifying an attribute of the audio signal; performing first determination processing and second determination processing when the attribute identified by the attribute information included in the audio signal obtained is information indicating an indirect sound, the first determination processing being processing to determine whether the audio signal obtained satisfies a first condition, the second determination processing being processing to determine whether the audio signal obtained satisfies a second condition different from the first condition; and outputting an output signal based on the audio signal obtained, when the audio signal obtained satisfies the first condition and the second condition.

[0030]

Thus, the first determination processing and the second determination processing are performed on an audio signal whose attribute is an indirect sound, and an output signal based on the audio signal obtained is output when the audio signal satisfies the first condition and the second condition. In other words, whether to output the output signal based on the audio signal for which the attribute is the indirect sound is appropriately determined. When the output signal is not output, the amount of computation and the computational load are reduced. In other words, it is possible to realize an audio signal processing method that can appropriately reduce the amount of computation and the computational load.

[0031]

An audio signal processing method according to a second aspect of the present disclosure is the audio signal processing method according to the first aspect, wherein the indirect sound is a reflected sound.

[0032]

This realizes an audio signal processing method that makes it possible to use reflected sound as the indirect sound.

[0033]

An audio signal processing method according to a third aspect of the present disclosure is the audio signal processing method according to the first or second aspect, wherein in the performing, the second determination processing is skipped when the attribute identified by the attribute information included in the audio signal obtained is information indicating a predetermined sound different from the indirect sound.

[0034]

Since the second determination processing is thus not performed on an audio signal whose attribute is the predetermined sound, the amount of computation and the computational load are further reduced. In other words, it is possible to realize an audio signal processing method that can appropriately reduce the amount of computation and the computational load.

[0035]

5 An audio signal processing method according to a fourth aspect of the present disclosure is the audio signal processing method according to the third aspect, wherein the predetermined sound is a direct sound.

[0036]

This realizes an audio signal processing method that makes it possible to use direct sound as the predetermined sound.

[0037]

10 An audio signal processing method according to a fifth aspect of the present disclosure is the audio signal processing method according to any one of the first to fourth aspects, wherein in the first determination processing, the audio signal obtained is determined to satisfy the first condition when an amplitude value of the audio signal
15 obtained is greater than or equal to a first threshold value, and in the second determination processing, the audio signal obtained is determined to satisfy the second condition when a sound volume ratio is greater than or equal to a second threshold value, the sound volume ratio being a sound volume ratio between a direct sound associated
20 with the indirect sound and the indirect sound when each of the direct sound and the indirect sound arrives at a listening position that is a position at which a listener is present, the second threshold value being a threshold value determined according to a time difference between when the direct sound arrives and when the indirect sound
25 arrives.

[0038]

Thus, in the first determination processing, the audio signal is determined to satisfy the first condition when the above-described amplitude value is greater than or equal to the first threshold value,
30 and in the second determination processing, the audio signal is determined to satisfy the second condition when the above-described sound volume ratio is greater than or equal to the second threshold value. An output signal based on the audio signal obtained is output when this second condition is satisfied. That is, whether to output
35 the output signal based on the audio signal is more appropriately

determined. In other words, it is possible to realize an audio signal processing method that can more appropriately reduce the amount of computation and the computational load.

[0039]

5 An audio signal processing method according to a sixth aspect of the present disclosure is the audio signal processing method according to any one of the first to fifth aspects, wherein the outputting includes: outputting a first output signal based on the audio signal when the attribute is a predetermined sound different
10 from the indirect sound; and outputting a second output signal based on the audio signal when the attribute is the indirect sound, in the outputting of the second output signal, the second output signal is output by performing diffusion filter processing on the audio signal obtained, the diffusion filter processing being processing that
15 improves realism of the indirect sound by diffusing the indirect sound, and in the outputting of the first output signal, the first output signal is output without applying the diffusion filter processing to the audio signal.

[0040]

20 Since the audio signal whose attribute is the indirect sound is thus subjected to diffusion filter processing and then output, the listener can hear indirect sounds with higher sound quality. Furthermore, since diffusion filter processing is not applied to the audio signal whose attribute is the predetermined sound, the amount
25 of computation and the computational load are further reduced. In other words, it is possible to realize an audio signal processing method that can appropriately reduce the amount of computation and the computational load.

[0041]

30 An audio signal processing method according to a seventh aspect of the present disclosure is the audio signal processing method according to any one of the first to sixth aspects, wherein the second determination processing is performed after the first determination processing.

35 [0042]

This realizes an audio signal processing method that makes it possible for the second determination processing to be performed after the first determination processing has been performed.

[0043]

5 Performing the determination processing in this order produces the following distinct effects. The main processing of the above-described first determination processing is processing to determine whether the amplitude value of the target audio signal is greater than or less than a predetermined threshold value; therefore, the
10 computational load of the first determination processing is extremely small. On the other hand, the above-described second determination processing requires calculating the sound volume ratio between the target indirect sound and the direct sound associated with the indirect sound, the time difference in their arrival, and the like, whereby the
15 computational load is significantly greater than that of the first determination processing. Therefore, in the first determination processing, for which the computational load is extremely small, a plurality of input signals are first filtered out, and only the remaining signals undergo determination in the second determination processing.
20 By doing so, the amount of computation for the entire determination step can be significantly reduced. Since the fundamental object of this disclosure is to reduce the amount of computation of audio signal processing, the order of this processing is extremely important.

[0044]

25 An audio signal processing method according to an eighth aspect of the present disclosure is an audio signal processing method executed by an audio signal processing device, the audio signal processing method including: obtaining an audio signal indicating an indirect sound; performing, on the audio signal obtained, one or more
30 first processes, first determination processing, second determination processing, and one or more second processes different from the one or more first processes; outputting an output signal based on the audio signal obtained; calculating a first adjustment amount by cumulatively adding adjustment amounts each determined by a
35 respective one of the one or more first processes, the adjustment

amounts each being an adjustment amount by which an amplitude of the audio signal obtained is to be amplified; and calculating a second adjustment amount by cumulatively adding adjustment amounts each determined by a respective one of the one or more second processes, the adjustment amounts each being an adjustment amount by which the amplitude of the audio signal obtained is to be amplified, wherein in the first determination processing, the audio signal obtained is determined to satisfy a first condition when the first adjustment amount calculated is greater than or equal to a first threshold value, the second determination processing is not performed when the audio signal obtained is determined not to satisfy the first condition, the second determination processing is performed when the audio signal obtained is determined to satisfy the first condition, the audio signal obtained is determined to satisfy a second condition when, in the second determination processing, a value based on the second adjustment amount calculated is greater than or equal to a second threshold value different from the first threshold value, and the outputting is performed when the audio signal obtained is determined to satisfy the second condition.

[0045]

As a result, the first determination processing and the second determination processing are performed on the audio signal indicating the indirect sound, and an output signal based on the audio signal obtained is output when the audio signal satisfies the first condition and the second condition. In other words, whether to output the output signal based on the audio signal indicating the indirect sound is appropriately determined. When the output signal is not output, the amount of computation and the computational load are reduced. In other words, it is possible to realize an audio signal processing method that can appropriately reduce the amount of computation and the computational load.

[0046]

An audio signal processing method according to a ninth aspect of the present disclosure is the audio signal processing method according to the eighth aspect, wherein the second adjustment

amount calculated in the calculating of the second adjustment amount includes an adjustment amount applied when diffusion filter processing has been performed, the diffusion filter processing being processing in which the indirect sound is diffused to improve realism of the indirect sound, the adjustment amount being an amount by which the amplitude of the audio signal obtained is to be amplified, the value is a ratio between a sound volume of a direct sound associated with the indirect sound and the second adjustment amount calculated in the calculating of the second adjustment amount, and the second threshold value is determined based on a time difference between when the direct sound arrives and when the indirect sound arrives.

[0047]

Accordingly, in the second determination processing, when the above-described ratio is greater than or equal to the second threshold value, the audio signal indicating the indirect sound is determined to satisfy the second condition, and an output signal based on the audio signal obtained is output when this second condition is satisfied. That is, whether to output the output signal based on the audio signal indicating the indirect sound is more appropriately determined. In other words, it is possible to realize an audio signal processing method that can more appropriately reduce the amount of computation and the computational load.

[0048]

An audio signal processing method according to a tenth aspect of the present disclosure is the audio signal processing method according to the eighth aspect, wherein the second adjustment amount calculated in the calculating of the second adjustment amount includes an amount of amplitude adjustment by a sound quality adjustment function set by selection by a listener, the value is a ratio between a sound volume of a direct sound associated with the indirect sound and the second adjustment amount calculated in the calculating of the second adjustment amount, and the second threshold value is determined based on a time difference between when the direct sound arrives and when the indirect sound arrives.

[0049]

The listener can thus listen to sound having his/her preferred sound quality, by performing selection to achieve his/her preferred sound quality, for example.

5 [0050]

An audio signal processing method according to an eleventh aspect of the present disclosure is the audio signal processing method according to any one of the eighth to tenth aspects, wherein the first threshold value is a value related to sound volume.

10 [0051]

This makes it possible to realize an audio signal processing method in which a value related to sound volume can be used as the first threshold value.

[0052]

15 An audio signal processing method according to a twelfth aspect of the present disclosure is the audio signal processing method according to any one of the eighth to eleventh aspects, wherein in the calculating of the second adjustment amount: a parameter indicating the adjustment amount determined by each of the one or more second
20 processes is obtained, the adjustment amount being an amount by which the amplitude of the audio signal obtained is to be amplified, and the second adjustment amount is calculated based on a plurality of parameters obtained, the plurality of parameters each being the parameter.

25 [0053]

The second adjustment amount is thus calculated according to the above-described parameters, whereby whether to output an output signal based on the audio signal whose attribute is the indirect sound is more appropriately determined. In other words, it is
30 possible to realize an audio signal processing method that can more appropriately reduce the amount of computation and the computational load.

[0054]

35 An audio signal processing method according to a thirteenth aspect of the present disclosure is the audio signal processing method

according to any one of the eighth to twelfth aspects, including:
setting the first threshold value; and performing disabling processing
to disable the second determination processing, wherein in the first
determination processing, the audio signal obtained is determined to
5 satisfy the first condition when the first adjustment amount calculated
is greater than or equal to the first threshold value set, and when the
disabling processing is performed, the outputting is performed upon
determining that the audio signal obtained satisfies the first condition.
[0055]

10 The second determination processing thus ceases to be
performed due to the disabling processing, thereby reducing the
amount of computation and the computational load required for
performing the second determination processing. In other words, it
is possible to realize an audio signal processing method that can more
15 appropriately reduce the amount of computation and the
computational load.
[0056]

An audio signal processing method according to a fourteenth
aspect of the present disclosure is the audio signal processing method
20 according to the thirteenth aspect, wherein a storage region storing
a value indicating the first threshold value set in the setting and a
storage region storing a signal instructing performance of the
disabling processing are adjacent regions.
[0057]

25 This not only allows the administrator of or the listener in the
virtual space to visually grasp the set values more easily, but also
yields the distinct advantage that, because the set values are adjacent
in terms of memory region, both values can be set simultaneously (in
a single memory access operation). In other words, by linking and
30 arranging the value indicating the first threshold value and the signal
instructing performance of the disabling processing in the upper bit
field and lower bit field of a region accessible by a single address, and
by writing the linked data sequence thus arranged in a single memory
access, the two data items can be set simultaneously.

35 [0058]

An audio signal processing method according to a fifteenth aspect of the present disclosure is the audio signal processing method according to the thirteenth aspect, wherein the first threshold value set when the disabling processing has been performed is greater than the first threshold value set when the disabling processing has not been performed.

[0059]

This makes it possible to realize an audio signal processing method that enables setting the magnitude of the first threshold value based on whether the disabling processing is performed.

[0060]

A computer program according to a sixteenth aspect of the present disclosure is a computer program for causing a computer to execute the audio signal processing method according to any one of the first to fifteenth aspects.

[0061]

This makes it possible for a computer to execute the above-described audio signal processing method, according to the computer program.

[0062]

An audio signal processing device according to a seventeenth aspect of the present disclosure is an audio signal processing device including: an obtainer that obtains an audio signal including attribute information identifying an attribute of the audio signal; a determiner that performs first determination processing and second determination processing when the attribute identified by the attribute information included in the audio signal obtained is information indicating an indirect sound, the first determination processing being processing to determine whether the audio signal obtained satisfies a first condition, the second determination processing being processing to determine whether the audio signal obtained satisfies a second condition different from the first condition; and a reproducer that outputs an output signal based on the audio signal obtained, when the audio signal obtained satisfies the first condition and the second condition.

[0063]

Thus, the first determination processing and the second determination processing are performed on an audio signal whose attribute is an indirect sound, and an output signal based on the audio signal obtained is output when the audio signal satisfies the first condition and the second condition. In other words, whether to output the output signal based on the audio signal for which the attribute is the indirect sound is appropriately determined. When the output signal is not output, the amount of computation and the computational load are reduced. In other words, it is possible to realize an audio signal processing device that can appropriately reduce the amount of computation and the computational load.

[0064]

[Embodiment 1]

(Example of Three-Dimensional Sound Reproduction System)

FIG. 2 is a diagram illustrating an example of three-dimensional sound reproduction system 1000. Specifically, FIG. 2 illustrates three-dimensional sound reproduction system 1000, which is an example of a system to which acoustic processing or decoding processing of the present disclosure can be applied. Three-dimensional sound is also expressed as immersive audio. Three-dimensional sound reproduction system 1000 includes audio signal processing device 1001 and audio presentation device 1002.

[0065]

Audio signal processing device 1001, which is also expressed as an acoustic processing device, applies acoustic processing to an audio signal emitted from a virtual sound source and generates an acoustic-processed audio signal to be presented to the listener. The audio signal is not limited to voices, and is acceptable as long as it is an audible sound. Acoustic processing is, for example, signal processing applied to an audio signal in order to reproduce one or more effects that a sound receives between when the sound is emitted from a sound source and when the sound arrives at the listener.

[0066]

Audio signal processing device 1001 performs acoustic

processing based on spatial information that describes the main factors for bringing about the above-described effects. Spatial information encompasses, for example: information that indicates the location of a sound source, a listener, and objects in the vicinity;
5 information that indicates the shape of a space; parameters regarding sound propagation; and the like. Audio signal processing device 1001 is, for example, a PC (personal computer), a smartphone, a tablet, a game console, or the like.

[0067]

10 An acoustic-processed signal is presented from audio presentation device 1002 to the listener. Audio presentation device 1002 is connected to audio signal processing device 1001 via wireless or wired communication. The acoustic-processed audio signal generated by audio signal processing device 1001 is transmitted to
15 audio presentation device 1002 via wireless or wired communication.

[0068]

When audio presentation device 1002 includes a plurality of devices such as, for example, a device for the right ear and a device for the left ear, or the like, the plurality of devices present sound in
20 synchronization by means of communication between the plurality of devices or communication between each of the plurality of devices and audio signal processing device 1001. Audio presentation device 1002 is, for example, headphones, earphones, or a head-mounted display worn on the head of the listener, surround speakers including
25 a plurality of fixed speakers, or the like.

[0069]

Note that three-dimensional sound reproduction system 1000 may be used in combination with an image presentation device or a stereoscopic image presentation device that visually provides an ER
30 experience that includes AR/VR. For example, a space handled by spatial information is a virtual space in which the positions of sound sources, the listener, and objects in the space are virtual positions of virtual sound sources, a virtual listener, and virtual objects in a virtual space. The space can also be expressed as a sound space.
35 Furthermore, the spatial information can also be expressed as sound

space information.

[0070]

Furthermore, FIG. 2 illustrates a system configuration example in which audio signal processing device 1001 and audio presentation device 1002 are separate devices, but three-dimensional sound reproduction system 1000 to which the acoustic processing method (the audio signal processing method) or the decoding method of the present disclosure can be applied is not limited to the configuration in FIG. 2. For example, audio signal processing device 1001 may be included in audio presentation device 1002, and audio presentation device 1002 may perform both acoustic processing and sound presentation.

[0071]

Furthermore, audio signal processing device 1001 and audio presentation device 1002 may, in a shared manner, perform the acoustic processing described in the present disclosure. Furthermore, a server connected to audio signal processing device 1001 or audio presentation device 1002 over a network may perform a part or all of the acoustic processing described in the present disclosure.

[0072]

Furthermore, audio signal processing device 1001 may perform the acoustic processing by decoding a bitstream that has been generated by encoding at least a part of data of the audio signal and the spatial information used in the acoustic processing. Thus, audio signal processing device 1001 may be expressed as a decoding device.

[0073]

(Example of Encoding Device)

FIG. 3A is a block diagram illustrating a configuration example of encoding device 1100. Specifically, FIG. 3A illustrates the configuration of encoding device 1100, which is an example of the encoding device of the present disclosure.

[0074]

Input data 1101 is data to be encoded, and includes spatial information and/or an audio signal to be inputted into encoder 1102. Details regarding the spatial information will be described later.

[0075]

Encoder 1102 encodes input data 1101 to generate encoded data 1103. Encoded data 1103 is, for example, a bitstream generated by means of encoding processing.

5 [0076]

Memory 1104 stores encoded data 1103. Memory 1104 may be, for example, a hard disk or an SSD (solid-state drive), or may be another type of memory.

[0077]

10 Note that in the above description, a bitstream generated by means of encoding processing was given as an example of encoded data 1103 stored in memory 1104, but encoded data 1103 may be data other than a bitstream. For example, encoding device 1100 may store, in memory 1104, converted data generated by converting the
15 bitstream into a predetermined data format. The converted data may be, for example, a file or multiplexed stream that corresponds to one or more bitstreams.

[0078]

Here, the file is a file having a file format of, for example, ISO
20 base media file format (ISOBMFF) or the like. Furthermore, encoded data 1103 may be in the form of a plurality of packets generated by splitting the above-described bitstream or file.

[0079]

For example, the bitstream generated by encoder 1102 may be
25 converted to data that is different from the bitstream. In this case, encoding device 1100 may include a converter, not illustrated, and the converter may perform conversion processing, or conversion processing may be performed by a central processing unit (CPU) that is an example of a processor, described later.

30 [0080]

(Example of Decoding Device)

FIG. 3B is a block diagram illustrating a configuration example of decoding device 1110. Specifically, FIG. 3B illustrates the configuration of decoding device 1110, which is an example of the
35 decoding device of the present disclosure.

[0081]

Memory 1114 stores, for example, the same data as encoded data 1103 generated by encoding device 1100. The stored data is read from memory 1114 and inputted into decoder 1112 as input data 1113. Input data 1113 is, for example, a bitstream that is to be decoded. Memory 1114 may be, for example, a hard disk or an SSD, or may be another type of memory.

[0082]

Note that decoding device 1110 may not directly input, to decoder 1112, data read from memory 1114 as input data 1113, and may instead convert the data read and then input the converted data to decoder 1112 as input data 1113. The data before conversion may be, for example, multiplexed data that includes one or more bitstreams. Here, the multiplexed data may be, for example, a file having a file format such as ISOBMFF or the like.

[0083]

Furthermore, the data before conversion may be a plurality of packets generated by splitting the above-described bitstream or file. Data that is different from the bitstream may be read from memory 1114 and then converted into a bitstream. In this case, decoding device 1110 may include a converter, not illustrated, and the converter may perform conversion processing, or conversion processing may be performed by a CPU that is an example of a processor, described later.

[0084]

Decoder 1112 decodes input data 1113 to generate audio signal 1111 that indicates audio to be presented to the listener.

[0085]

(Other Example of Encoding Device)

FIG. 3C is a block diagram illustrating a configuration example of another encoding device. Specifically, FIG. 3C illustrates the configuration of encoding device 1120, which is another example of the encoding device of the present disclosure. In FIG. 3C, constituent elements that are the same as the constituent elements in FIG. 3A have been given the same reference signs as in FIG. 3A,

and description of these constituent elements is omitted.

[0086]

Encoding device 1100 stores encoded data 1103 in memory 1104. On the other hand, encoding device 1120 is different from
5 encoding device 1100 in the respect that encoding device 1120 includes transmitter 1121 that transmits encoded data 1103 externally.

[0087]

Transmitter 1121 transmits, to a different device or server,
10 transmission signal 1122 that is generated based on data converted from encoded data 1103 or encoded data 1103 to a different file format. The data used in generating transmission signal 1122 is, for example, the bitstream, multiplexed data, file, or packet described in relation to encoding device 1100.

[0088]

(Other Example of Decoding Device)

FIG. 3D is a block diagram illustrating another configuration example of a decoding device. Specifically, FIG. 3D illustrates the configuration of decoding device 1130, which is another example of
20 the decoding device of the present disclosure. In FIG. 3D, constituent elements that are the same as the constituent elements in FIG. 3B have been given the same reference signs as in FIG. 3B, and description of these constituent elements is omitted.

[0089]

Decoding device 1110 reads input data 1113 from memory 1114. On the other hand, decoding device 1130 is different from decoding device 1110 in the respect that decoding device 1130 includes receiver 1131, which receives input data 1113 from an external source.

[0090]

Receiver 1131 receives reception signal 1132 to obtain reception data, and outputs input data 1113 to be inputted into decoder 1112. The reception data may be the same as input data 1113 inputted into decoder 1112, or may be data in a data format
35 that is different from that of input data 1113.

[0091]

When the data format of the reception data is different from the data format of input data 1113, receiver 1131 may convert the reception data into input data 1113. Alternatively, a converter or a CPU, each not illustrated, of decoding device 1130 may convert the reception data into input data 1113. The reception data is, for example, the bitstream, multiplexed data, file, or packet described in relation to encoding device 1120.

[0092]

10 (Example of Decoder)

FIG. 4A is a block diagram illustrating a configuration example of decoder 1200. Specifically, FIG. 4A illustrates the configuration of decoder 1200, which is an example of decoder 1112 in FIG. 3B and FIG. 3D.

15 [0093]

Input data 1113 is an encoded bitstream, and includes encoded audio data that is an audio signal that has been encoded, and metadata used in acoustic processing.

[0094]

20 Spatial information manager 1201 obtains the metadata included in input data 1113 and analyzes the metadata. The metadata includes information that describes the main factors that act on the sounds arranged in the sound space. Spatial information manager 1201 manages the spatial information that is obtained by analyzing the metadata and is used in the acoustic processing, and provides the spatial information to renderer 1203.

[0095]

Note that in the present disclosure, the information used in the acoustic processing is expressed as spatial information, but another expression may be used. For example, the information used in the acoustic processing may be expressed as sound space information, or may be expressed as scene information. Furthermore, when the information used in the acoustic processing changes over time, the spatial information inputted into renderer 1203 may be information expressed as a spatial state, a sound space state, a scene state, or

the like.

[0096]

Furthermore, the spatial information may be managed for each sound space or for each scene. For example, when each of a plurality of mutually differing rooms is expressed as a virtual space, the plurality of rooms may be managed as a plurality of scenes that mutually differ. Furthermore, spatial information may be managed such that even the same room is managed as a different scene in accordance with the expressed state.

[0097]

Thus, a plurality of items of spatial information may be managed with respect to a plurality of sound spaces or a plurality of scenes. In management of a plurality of items of spatial information, an identifier that identifies each item of the plurality of items of spatial information may be assigned to the spatial information.

[0098]

The spatial information data may be included in a bitstream that is an example of input data 1113. Alternatively, the bitstream may include an identifier of the spatial information, and the spatial information data may be obtained from an information source other than the bitstream. Specifically, when the bitstream includes only the identifier of the spatial information, in the rendering, the spatial information data stored in the memory inside the device or in an external server may be obtained as input data 1113, using the identifier of the spatial information.

[0099]

Note that the information managed by spatial information manager 1201 is not limited to information included in the bitstream. For example, input data 1113 may include, as data not included in the bitstream, data that indicates the characteristics and structure of a space obtained from a VR or AR software application or server.

[0100]

Furthermore, input data 1113 may include data that indicates the characteristics, position, and/or the like of the listener or an object. Moreover, input data 1113 may include information on the

position of the listener, obtained using a sensor included in a terminal including a decoding device (1110, 1130), or may include information that indicates the position of the terminal, estimated based on information obtained using the sensor.

5 [0101]

In other words, spatial information manager 1201 may communicate with an external system or server to obtain spatial information and the position of the listener (in other words, the listening position). Furthermore, spatial information manager 1201
10 may obtain clock synchronization information from an external system and perform processing to perform synchronization with a clock in renderer 1203.

[0102]

Note that the space in the above description may be a virtually
15 formed space, i.e., a VR space, or may be a real-world space or a virtual space that corresponds to a real-world space, i.e., an AR space or an MR space. Furthermore, the virtual space may be expressed as a sound field or a sound space. Moreover, the information indicating position in the above description may be information on
20 coordinates or the like that indicate a position in a space, may be information that indicates a relative position with respect to a predetermined reference position, or may be information that indicates movement or acceleration of a position in a space.

[0103]

25 Audio data decoder 1202 decodes encoded audio data included in input data 1113 to obtain an audio signal.

[0104]

The encoded audio data obtained by three-dimensional sound reproduction system 1000 is, for example, a bitstream encoded in a
30 predetermined format such as MPEG-H 3D Audio (ISO/IEC 23008-3). Note that MPEG-H 3D Audio is merely an example of an encoding method that can be used when generating the encoded audio data included in the bitstream. The encoded audio data may be a bitstream encoded by another encoding method.

35 [0105]

For example, the encoding method may be a lossy codec such as MPEG-1 Audio Layer III (MP3), Advanced Audio Coding (AAC), Windows Media Audio (WMA), Audio Codec 3 (AC3), Vorbis, or the like. Alternatively, the encoding method may be a lossless codec such as Apple Lossless Audio Codec (ALAC), Free Lossless Audio Codec (FLAC), or the like.

[0106]

Alternatively, any encoding method other than the above-described may be used. For example, PCM data may be a type of the encoded audio data. In this case, when, for example, the quantization bit rate of the PCM data is N, the decoding processing may be processing in which the N-bit binary number is converted into a numerical format (for example, floating-point format) that can be processed by renderer 1203.

[0107]

Renderer 1203 obtains the audio signal and the spatial information, applies acoustic processing to the audio signal using the spatial information, and outputs an acoustic-processed audio signal (audio signal 1111).

[0108]

Before starting the rendering, spatial information manager 1201 reads the metadata of the input signal, detects rendering items such as objects and sounds specified by the spatial information, and transmits the rendering items to renderer 1203. After the start of rendering, spatial information manager 1201 grasps the change over time of the spatial information and the position of the listener, and updates and manages the spatial information. Then, the updated spatial information is transmitted to renderer 1203.

[0109]

Renderer 1203 generates and outputs an acoustic processing-added audio signal based on the audio signal included in input data 1113 and the spatial information received from spatial information manager 1201.

[0110]

Update processing of the spatial information and output

processing of the acoustic processing-added audio signal may be performed in the same thread. Furthermore, processing may be allocated to spatial information manager 1201 and renderer 1203 in mutually independent threads. When spatial information manager
5 1201 and renderer 1203 perform the update processing of the spatial information and the output processing of the acoustic processing-added audio signal in different threads, the thread activation frequency may be set separately, or processing may be performed in parallel.

10 [0111]

When spatial information manager 1201 and renderer 1203 perform processing in independent threads that are different, renderer 1203 can be preferentially allotted computation resources. This makes it possible to safely execute sound output processing for
15 which even a slight delay is unallowable, i.e., for which a popping noise would be generated with a delay of even one sample (0.02 msec).

[0112]

At this time, allotment of computation resources to spatial
20 information manager 1201 is limited. However, since compared to the output processing of the audio signal, updating of the spatial information is processing that is infrequent (for example, processing such as updating the orientation of a listener's face), it is not necessary for the output processing of the audio signal to be
25 performed instantaneously. Thus, even when the allotment of computation resources is limited, the acoustic quality is not greatly impacted.

[0113]

The updating of the spatial information may be periodically
30 performed with each elapse of a preset time period or term, or may be performed when a preset condition is satisfied. Furthermore, the updating of the spatial information may be performed manually by the listener or a sound space manager, or may be performed by being triggered by a change to an external system.

35 [0114]

For example, the spatial information may be updated when a controller is being operated by the listener and the position of the listener's avatar is instantaneously warped or time is instantaneously progressed or reversed. Alternatively, the spatial information may be updated when an operation to suddenly change the field environment is performed by the manager of the virtual space. In these cases, the thread for updating the spatial information managed by spatial information manager 1201 may be activated as one-time interrupt processing in addition to the periodic activation.

[0115]

The role of the information update thread that performs update processing of the spatial information is, for example: processing to update, based on the position or orientation of the VR goggles worn by the listener, the position or orientation of the listener's avatar positioned in the virtual space; updating of the positions of objects that have moved within the virtual space; and the like. This is handled within a processing thread that operates at a relatively low frequency of around several tens of Hz. Processing that reflects the characteristics of direct sound may also be performed in such a low-frequency processing thread. This is because the characteristics of a direct sound vary less frequently than audio processing frames for audio output occur. Rather, by doing so, the computational load of such processing can be made relatively low, and the risk of pulsive noise can be avoided, since updating information at an unduly high frequency would generate the risk of pulsive noise occurring.

[0116]

FIG. 4B is a block diagram illustrating another configuration example of a decoder. Specifically, FIG. 4B illustrates the configuration of decoder 1210, which is another example of decoder 1112 in FIG. 3B and FIG. 3D.

[0117]

FIG. 4B is different from FIG. 4A in the respect that input data 1113 includes not encoded audio data, but an unencoded audio signal. Input data 1113 includes an audio signal and a bitstream including metadata.

[0118]

Spatial information manager 1211 is the same as spatial information manager 1201 in FIG. 4A; therefore, description thereof has been omitted.

5 [0119]

Renderer 1213 is the same as renderer 1203 in FIG. 4A; therefore, description thereof has been omitted.

[0120]

10 Note that decoders 1112, 1200, and 1210 may be expressed as the acoustic processor that performs the acoustic processing. Furthermore, decoding devices 1110 and 1130 may be audio signal processing device 1001, or may be expressed as the acoustic processing device.

[0121]

15 (Physical Configuration of Audio Signal Processing Device)

FIG. 5 is a diagram illustrating an example of a physical configuration of audio signal processing device 1001. Note that audio signal processing device 1001 in FIG. 5 may be decoding device 1110 in FIG. 3B or decoding device 1130 in FIG. 3D. A plurality of the constituent elements illustrated in FIG. 3B or FIG. 3D may be implemented by a plurality of the constituent elements illustrated in FIG. 5. Furthermore, a part of the configuration described here may be included in audio presentation device 1002.

[0122]

25 Audio signal processing device 1001 in FIG. 5 includes processor 1402, memory 1404, communication interface (I/F) 1403, sensor 1405, and loudspeaker 1401.

[0123]

30 Processor 1402 is, for example, a CPU, a digital signal processor (DSP), or a graphics processing unit (GPU). The acoustic processing or the decoding processing of the present disclosure may be performed by the CPU, the DSP, or the GPU executing a program stored in memory 1404. Furthermore, processor 1402 is, for example, a circuit that performs information processing. Processor
35 1402 may be a dedicated circuit that performs signal processing on

audio signals, including the acoustic processing of the present disclosure.

[0124]

Memory 1404 includes, for example, random access memory (RAM) or read-only memory (ROM). Memory 1404 may include, for example, magnetic storage media, exemplified by a hard disk, or semiconductor memory, exemplified by an SSD. Furthermore, memory 1404 may be an internal memory incorporated into the CPU or GPU. Moreover, spatial information managed by spatial information manager 1201 and 1211 and/or the like may be stored in memory 1404. Furthermore, threshold value data, described later, may be stored.

[0125]

Communication I/F 1403 is, for example, a communication module that supports a communication method such as Bluetooth (registered trademark) or WiGig (registered trademark). Audio signal processing device 1001 communicates with other communication devices via communication I/F 1403, and obtains a bitstream to be decoded. The obtained bitstream is, for example, stored in memory 1404.

[0126]

Communication I/F 1403 includes, for example, a signal processing circuit that supports the communication method, and an antenna. The communication method is not limited to Bluetooth (registered trademark) or WiGig (registered trademark), and may be Long Term Evolution (LTE), New Radio (NR), Wi-Fi (registered trademark), or the like.

[0127]

The communication method is not limited to the wireless communication methods described above, and may be a wired communication method such as Ethernet (registered trademark), Universal Serial Bus (USB), High-Definition Multimedia Interface (HDMI) (registered trademark), or the like.

[0128]

Sensor 1405 performs sensing to estimate the position or

orientation of the listener. Specifically, sensor 1405 estimates the position and/or orientation of the listener based on one or more detection results of one or more of the position, orientation, movement, velocity, angular velocity, acceleration, or the like of a part or all of the listener's body, and generates position/orientation information indicating the position and/or orientation of the listener. [0129]

Note that a device outside of audio signal processing device 1001 may include sensor 1405. The part of the body may be the listener's head or the like. The position/orientation information may be information indicating the position and/or orientation of the listener in real-world space, or may be information indicating the displacement of the position and/or orientation of the listener with respect to the position and/or orientation of the listener at a predetermined time point. Furthermore, the position/orientation information may be information indicating a position and/or orientation relative to three-dimensional sound reproduction system 1000 or an external device including sensor 1405. [0130]

Sensor 1405 may be, for example, an imaging device such as a camera or a distance measuring device such as a laser imaging detection and ranging (LiDAR) distance measuring device. Sensor 1405 may capture an image of the movement of the listener's head and detect the movement of the listener's head by processing the captured image. Furthermore, a device that performs position estimation using radio waves in any given frequency band such as millimeter waves may be used as sensor 1405. [0131]

Furthermore, audio signal processing device 1001 may obtain position information via communication I/F 1403 from an external device including sensor 1405. In this case, audio signal processing device 1001 need not include sensor 1405. Here, the external device refers to, for example, audio presentation device 1002 described in FIG. 2, or a stereoscopic image reproduction device worn on the listener's head. In this case, sensor 1405 is configured as a

combination of various sensors, such as a gyro sensor and an acceleration sensor, for example.

[0132]

5 As the speed of the movement of the listener's head, sensor 1405 may detect, for example, the angular speed of rotation about at least one of three mutually orthogonal axes in the sound space as the axis of rotation or the acceleration of displacement in at least one of the three axes as the direction of displacement.

[0133]

10 As the amount of the movement of the listener's head, sensor 1405 may detect, for example, the amount of rotation about at least one of three mutually orthogonal axes in the sound space as the axis of rotation or the amount of displacement in at least one of the three axes as the direction of displacement. Specifically, sensor 1405
15 detects 6DoF positions (x, y, z) and angles (yaw, pitch, roll) as the position of the listener. Sensor 1405 is configured as a combination of various sensors used for detecting movement, such as a gyro sensor and an acceleration sensor.

[0134]

20 Note that sensor 1405 may implemented by, e.g., a camera or a Global Positioning System (GPS) receiver for detecting the position of the listener. Position information obtained by performing self-position estimation, by using LiDAR or the like as sensor 1405, may be used. For example, when three-dimensional sound reproduction
25 system 1000 is implemented by a smartphone, sensor 1405 is included in the smartphone.

[0135]

Furthermore, sensor 1405 may include a temperature sensor such as a thermocouple that detects the temperature of audio signal
30 processing device 1001. Moreover, sensor 1405 may include, for example, a sensor that detects the remaining level of a battery included in audio signal processing device 1001 or a battery connected to audio signal processing device 1001.

[0136]

35 Loudspeaker 1401 includes, for example, a diaphragm, a

driving mechanism such as a magnet or a voice coil, and an amplifier, and presents the acoustic-processed audio signal as sound to the listener. Loudspeaker 1401 operates the driving mechanism according to the audio signal (more specifically, a waveform signal indicating the waveform of the sound) amplified via the amplifier, and vibrates the diaphragm by means of the driving mechanism. In this way, the diaphragm vibrating according to the audio signal generates sound waves, which propagate through the air and are transmitted to the listener's ears, allowing the listener to perceive the sound.

[0137]

Note that although here, an example in which audio signal processing device 1001 includes loudspeaker 1401 and presents the acoustic-processed audio signal via loudspeaker 1401 was given, the means for providing the audio signal is not limited to this configuration.

[0138]

For example, the acoustic-processed audio signal may be outputted to external audio presentation device 1002 connected via a communication module. The communication performed by the communication module may be wired or wireless. As another example, audio signal processing device 1001 may include a terminal that outputs an analog audio signal, and may present the audio signal from earphones or the like by connecting the earphone cable to the terminal.

[0139]

In this case, audio presentation device 1002 may be headphones, earphones, a head-mounted display, neck speakers, wearable speakers, or the like, each worn on the listener's head or a part of the listener's body. Alternatively, audio presentation device 1002 may be surround speakers configured with a plurality of fixed speakers, or the like. Audio presentation device 1002 may reproduce the audio signal.

[0140]

(Physical Configuration of Encoding Device)

FIG. 6 is a diagram illustrating an example of a physical

configuration of encoding device 1500. Encoding device 1500 in FIG. 6 may be encoding device 1100 in FIG. 3A or encoding device 1120 in FIG. 3C, or a plurality of the constituent elements illustrated in FIG. 3A or FIG. 3C may be implemented by a plurality of the constituent elements illustrated in FIG. 6.

[0141]

Encoding device 1500 in FIG. 6 includes processor 1501, memory 1503, and communication I/F 1502.

[0142]

Processor 1501 is, for example, a CPU, a DSP, or a GPU. The encoding processing of the present disclosure may be performed by the CPU, the DSP, or the GPU executing a program stored in memory 1503. Furthermore, processor 1501 is, for example, a circuit that performs information processing. Processor 1501 may be a dedicated circuit that performs signal processing on audio signals, including the encoding processing of the present disclosure.

[0143]

Memory 1503 includes, for example, RAM or ROM. Memory 1503 may include, for example, magnetic storage media, exemplified by a hard disk, or semiconductor memory, exemplified by an SSD. Furthermore, memory 1503 may be an internal memory incorporated into the CPU or GPU.

[0144]

Communication I/F 1502 is, for example, a communication module that supports a communication method such as Bluetooth (registered trademark) or WiGig (registered trademark). For example, encoding device 1500 communicates with other communication devices via communication I/F 1502, and transmits an encoded bitstream.

[0145]

Communication I/F 1502 includes, for example, a signal processing circuit that supports the communication method, and an antenna. The communication method is not limited to Bluetooth (registered trademark) or WiGig (registered trademark), and may be LTE, NR, Wi-Fi (registered trademark), or the like. The

communication method is not limited to wireless communication methods. The communication method may be a wired communication method such as Ethernet (registered trademark), USB, HDMI (registered trademark), or the like.

5 [0146]

The communication module includes, for example, a signal processing circuit that supports the communication method, and an antenna. In the above example, Bluetooth (registered trademark) and WiGig (registered trademark) were given as examples of the communication method, but a communication method such as Long Term Evolution (LTE), New Radio (NR), Wi-Fi (registered trademark), or the like may be supported. Furthermore, the communication I/F may be not the wireless communication methods described above, but a wired communication method such as Ethernet (registered trademark), Universal Serial Bus (USB), High-Definition Multimedia Interface (HDMI) (registered trademark), or the like.

10
15

[0147]

[Configuration of Renderer]

FIG. 7 is a block diagram illustrating a configuration example of renderer 1300. Specifically, FIG. 7 illustrates an example of the detailed configuration of renderer 1300, which corresponds to renderers 1203 and 1213 in FIG. 4A and FIG. 4B.

20

[0148]

Renderer 1300 includes analyzer 1301, determiner 1302, and reproducer 1303, and adds acoustic processing to sound data included in the input signal and outputs the sound data.

25

[0149]

The input signal includes, for example, spatial information, sensor information, and sound data. The input signal may include a bitstream that includes sound data and metadata (control information), and in this case, the spatial information may be included in the metadata.

30

[0150]

The spatial information is information related to the sound space (three-dimensional sound field) created by three-dimensional

35

sound reproduction system 1000, and includes information about objects included in the sound space and information about the listener. The objects include sound source objects that emit sound and serve as sound sources, and non-sound-emitting objects that do not emit sound. The sound source objects may be expressed as simply sound sources.

[0151]

The non-sound-emitting object serves as an obstacle object that reflects sound emitted by the sound source object, but a sound source object may also serve as an obstacle object that reflects sound emitted by another sound source object. The obstacle object may also be expressed as a reflection object.

[0152]

Information assigned in common to both sound source objects and non-sound-emitting objects includes position information, geometry information, and the attenuation rate of sound volume when the object reflects sound.

[0153]

The position information is represented by coordinate values of three axes, for example, the X-axis, the Y-axis, and the Z-axis of Euclidean space, but it does not necessarily have to be three-dimensional information. For example, the position information may be two-dimensional information represented by coordinate values of the two axes of the X-axis and the Y-axis. The position information of the object is defined by a representative position of the shape expressed by a mesh or voxel.

[0154]

The geometry information may include information about the material of the surface.

[0155]

The attenuation rate may be expressed as a real number greater than or equal to 0 and less than or equal to 1, or may be expressed as a negative decibel value. Since sound volume does not increase from reflection in real-world space, the attenuation rate is set to a negative decibel value. However, for example, to create an

eerie atmosphere in a non-realistic space, an attenuation rate greater than or equal to 1, that is, a positive decibel value, may be intentionally set.

[0156]

5 Furthermore, the attenuation rate may be set such that each frequency band included in a plurality of frequency bands has a different value, or values may be independently set for each frequency band. Furthermore, when the attenuation rate is set for each type of material of an object surface, the value of the
10 corresponding attenuation rate may be used based on information about the surface material.

[0157]

 Furthermore, the spatial information may include, for example, information indicating whether the object belongs to an animate thing
15 or information indicating whether the object is a mobile body. When the object is a mobile body, the position indicated by the position information may move over time. In this case, information on the changed position or the amount of change is transmitted to renderer
1300.

20 [0158]

 Information related to the sound source object includes, in addition to information assigned in common to both sound source objects and non-sound-emitting objects, sound data and information necessary for radiating the sound data into the sound space. The
25 sound data is data representing sound perceived by the listener, and indicates information such as the frequency and intensity of the sound.

[0159]

 The sound data is typically a PCM signal, but may also be data compressed using an encoding method such as MP3. In this case,
30 since the signal needs to be decoded at least before arriving at reproducer 1303, renderer 1300 may include a decoder (not illustrated). Alternatively, the signal may be decoded by audio data decoder 1202.

[0160]

35 One item of sound data may be set for one sound source object,

or a plurality of items of sound data may be set. Furthermore, identification information for identifying each item of sound data may be assigned to the sound data, or information related to the sound source object may include identification information for the sound data.

[0161]

As information necessary for radiating sound data into the sound space, for example, information on a reference sound volume that is used as a standard in reproducing the sound data, information indicating a characteristic of sound data, information related to the position of the sound source object, information related to the orientation of the sound source object (in other words, information related to the directivity of the sound emitted by the sound source object), and the like may be included.

[0162]

The information on the reference sound volume may be, for example, the root mean square value of the amplitude value of the sound data at the sound source position at the time of radiating the sound data into the sound space, and may be expressed as a floating-point decibel (dB) value.

[0163]

For example, when the reference sound volume is 0 dB, the information on the reference sound volume may indicate that the sound is to be radiated into the sound space from the position indicated by the information related to the position of the sound source object, at the same sound volume as the signal level indicated by the sound data, without increase or decrease. Furthermore, when the reference sound volume is -6 dB, this may indicate that the sound is to be radiated into the sound space from the position indicated by the information related to the position of the sound source object at approximately half the sound volume of the signal level indicated by the sound data.

[0164]

The information on the reference sound volume may be assigned to each item of sound data, or may be assigned collectively

to a plurality of items of sound data.

[0165]

The information indicating a characteristic of sound data is, for example, information on the sound volume of a sound source, and
5 may be information indicating the chronological variation in the sound volume of a sound source.

[0166]

For example, when the sound space is a virtual conference room and the sound source is a speaker, the sound volume transitions
10 intermittently over short periods of time. In other words sound portions and silent portions occur alternately. Furthermore, when the sound space is a concert hall and the sound source is a performer, the sound volume is maintained over a certain duration of time. Moreover, when the sound space is a battlefield and the sound source
15 is an explosive, the sound volume of the explosion sound becomes large for only an instant and then continues to be silent or in a quiet state thereafter.

[0167]

In this way, the sound volume information on the sound source
20 may include not only information on the magnitude of sound but also information on the transition of the sound magnitude. Such information may be used as the information indicating a characteristic of the sound data.

[0168]

The information on the transition may be expressed by data
25 indicating frequency characteristics in chronological order. The information on the transition may be expressed by data indicating the duration of a sound interval. The information on the transition may be expressed by data indicating the chronological order of durations
30 of sound intervals and durations of silent intervals. The information on the transition may be expressed by, for example, data that enumerates, in chronological order, a plurality of sets of a duration for which the amplitude of the sound signal can be considered stationary (can be considered approximately constant) and the
35 amplitude value of said signal during that duration.

[0169]

The information on the transition may be expressed by data of a duration during which the frequency characteristics of the sound signal can be considered stationary. The information on the transition may be expressed by, for example, data that enumerates, in chronological order, a plurality of sets of a duration during which the frequency characteristics of the sound signal can be considered stationary and the frequency characteristics during that duration. The information on the transition may be expressed in the format of, for example, data indicating the general shape of a spectrogram.

[0170]

Furthermore, the sound volume that is used as the standard for the above-mentioned frequency characteristics may be used as the reference sound volume. The information on the reference sound volume and the information indicating a characteristic of the sound data may be used for calculation processing of the sound volume of direct sound or reflected sound to be perceived by the listener, and/or may be used for selection processing (also referred to as determination processing) for selecting whether to make the listener perceive the sound. Other examples of and usage methods for the information indicating a characteristic of sound data will be described later.

[0171]

It should be noted that the reflected sound according to the present embodiment is an example of indirect sound. Indirect sound may be reflected sound, diffracted sound, or the like. Furthermore, the direct sound according to the present embodiment is an example of a predetermined sound different from the above-mentioned indirect sound. The predetermined sound may be a direct sound, or may be a higher-order ambisonic (HOA) sound. Alternatively, by combining a plurality of audio signals (e.g., through mixing), a representative sound that represents a plurality of sounds indicated by the plurality of audio signals may be created, and this representative sound may be used as the predetermined sound. In this case, the representative sound may be referred to as the representative sound.

[0172]

In the present embodiment, reflected sound, which is an example of indirect sound, and direct sound, which is an example of predetermined sound, are used for description, but the same processing is performed even if indirect sound is used instead of reflected sound and predetermined sound is used instead of direct sound.

[0173]

Information regarding the orientation of a sound source object (orientation information) is typically expressed in terms of yaw, pitch, and roll. Alternatively, the rotation of roll may be omitted, and the orientation information of a sound source object may be expressed in terms of azimuth (yaw) and elevation (pitch). The orientation information of a sound source object may change over time, and when changed, the orientation information is transmitted to renderer 1300.

[0174]

Information related to the listener is information regarding the position and orientation of the listener in the sound space. The information regarding the position (position information) is represented by the position on the X-, Y-, and Z-axes of Euclidean space, but need not necessarily be three-dimensional information and may be two-dimensional information. Information regarding the orientation of the listener (orientation information) is typically expressed in terms of yaw, pitch, and roll. Alternatively, the rotation of roll may be omitted, and the listener orientation information may be expressed in terms of azimuth (yaw) and elevation (pitch).

[0175]

The position information and orientation information regarding a listener may change over time, and when changed, the position information and orientation information are transmitted to renderer 1300.

[0176]

The sensor information is information that includes, e.g., the rotation amount or displacement amount detected by sensor 1405 worn by the listener, and the position and orientation of the listener.

The sensor information is transmitted to renderer 1300, and renderer 1300 updates the information on the position and orientation of the listener based on the sensor information. The sensor information may include position information obtained by performing self-localization estimation by a mobile terminal using GPS, a camera, or LiDAR, for example.

[0177]

Furthermore, information obtained not from sensor 1405, but from an external source through a communication module, may also be detected as sensor information. Information indicating the temperature of audio signal processing device 1001, and information indicating the remaining level of the battery may be obtained from sensor 1405. Moreover, computational resources (CPU capability, memory resources, PC performance, and the like) of audio signal processing device 1001 or audio presentation device 1002 may be obtained in real time.

[0178]

Analyzer 1301 analyzes an audio signal included in the input signal and spatial information received from spatial information managers 1201 and 1211 to calculate the information required for generating direct sounds and reflected sounds with reproducer 1303, and the information required for determining (selecting) whether to generate reflected sounds.

[0179]

The information required for generating direct sounds and reflected sounds is, for example, for each of direct sounds and reflected sounds, values related to the path until arriving at the listening position, the time period taken until arrival, the sound volume at the arrival time, and the like. The values related to, e.g., the path until arriving at the listening position, the time period taken until arrival, and the sound volume at the arrival time are, for example, values representing the path until arriving at the listening position, the time period taken until arrival, and the sound volume at the arrival time, respectively.

[0180]

The information required for selecting a reflected sound to be output is information indicating the relationship between the direct sound and the reflected sound, and is, for example, a value regarding a time difference between the direct sound and the reflected sound, a value regarding a sound volume ratio of the reflected sound to the direct sound at the listening position, and/or the like. The value related to the time difference between a direct sound and a reflected sound and the value related to the sound volume ratio between a direct sound and a reflected sound at the listening position are, for example, a value representing the time difference between the direct sound and the reflected sound and a value representing the sound volume ratio of the reflected sound to the direct sound at the listening position, respectively.

[0181]

Note that it goes without saying that when the sound volume is expressed in units of decibels on a logarithmic scale (when the sound volume is expressed in the decibel domain), the sound volume ratio between the two signals is expressed as a decibel value difference. Specifically, the sound volume ratio between the two signals may be the difference when the amplitude value of each signal is expressed in the decibel domain. That value may be calculated based on, e.g., an energy value, a power value, or the like. Furthermore, this difference can be referred to as a difference in gain or simply a gain difference, in the decibel domain.

[0182]

In other words, the sound volume ratio in the present disclosure is essentially the ratio between the amplitudes of signals; thus, the sound volume ratio may be expressed as a loudness ratio, a volume ratio, an amplitude ratio, a sound level ratio, a sound intensity ratio, a gain ratio, or the like. Furthermore, when the unit of sound volume is decibels, it goes without saying that the sound volume ratio in the present disclosure may be rephrased as the sound volume difference.

[0183]

In the present disclosure, the "sound volume ratio" typically

means the gain difference when the sound volume of each of two sounds is expressed in the unit of decibels, and in the examples of the embodiment, the threshold value data is also typically specified by the gain difference expressed in the decibel domain. However, the sound volume ratio is not limited to the gain difference in the decibel domain. When a sound volume ratio that is not expressed by the decibel domain is used, threshold value data specified in the decibel domain may be used by converting the threshold value data into the unit of the sound volume ratio calculated. Alternatively, threshold value data specified beforehand in each unit may be stored in the memory.

[0184]

In other words, for example, even if a ratio between energy values, power value, or the like is used instead of the sound volume ratio, it is obvious that the algorithm in the present disclosure can be applied to solve the problem of the present disclosure.

[0185]

The time difference between the arrival of a direct sound and the arrival of a reflected sound is, for example, the time difference between a direct sound arrival time period (arrival time) and a reflected sound arrival time period (arrival time). It should be noted that for simplicity, the time difference between the arrival of a direct sound and the arrival of a reflected sound may be referred to as the time difference between the direct sound and the reflected sound. The time difference between a direct sound and a reflected sound may be the time difference between the times at which each of the direct sound and the reflected sound arrive at the listening position, the difference in the time periods taken until each of the direct sound and the reflected sound arrive at the listening position, or the time difference between the time when emission of the direct sound ends and the time when the reflected sound arrives at the listening position. The methods for calculating these values will be described later.

[0186]

Determiner 1302 performs at least one of first determination processing using a first threshold value and second determination

processing using a second threshold value. In the present embodiment, determiner 1302 performs the second determination processing on reflected sound (that is, the audio signal indicating the reflected sound). Below, the second determination processing is
5 described in greater detail.

[0187]

Determiner 1302 determines whether reproducer 1303 is to generate a reflected sound by using information calculated by analyzer 1301 and the threshold value data indicating the second
10 threshold value. To put it differently, determiner 1302 determines whether to select a reflected sound as a reflected sound to be generated. To put it still differently, determiner 1302 selects which reflected sounds reproducer 1303 is to generate, from a plurality of reflected sounds. It should be noted that below, when determiner
15 1302 determines that a reflected sound is to be generated by reproducer 1303, this may be described as determiner 1302 selecting the reflected sound, or determiner 1302 selecting to generate the reflected sound.

[0188]

The threshold value data is, for example, a graph having a horizontal axis that indicates the time difference between a direct sound and reflected sounds and a vertical axis that indicates the sound volume ratios of reflected sounds to a direct sound, and is expressed as a boundary (threshold value) that demarcates whether
20 each reflected sound is perceived. For example, the threshold value data may be expressed as an approximation formula that includes the time difference between a direct sound and a reflected sound as a variable, or may be expressed as an arrangement that includes values of time differences between direct sounds and reflected sounds as an
25 index, and corresponding threshold values. It should be noted that in Embodiment 1, the second threshold value may be referred to simply as the threshold value.

[0189]

Determiner 1302 selects the generation of a reflected sound
35 when, for example, at the time difference between the arrival time of

a direct sound and the arrival time of a reflected sound, the sound volume ratio of the arrival time sound volume of the reflected sound to the arrival time sound volume of the direct sound is a value that is larger than a threshold value set with reference to threshold value data. It should be noted that the arrival time sound volume means the sound volume when the sound arrives at the listening position.

[0190]

To put it differently, the time difference between the arrival time of a direct sound and the arrival time of a reflected sound is the difference in the amount of time taken for the direct sound and the reflected sound to arrive at the listening position. Furthermore, the time difference between the time point at which emission of the direct sound stops and the time point at which the reflected sound arrives at the listening position may be used as the time difference between the direct sound and the reflected sound. In this case, threshold value data that is different from the threshold value data determined by using, as a reference, the time difference between the direct sound arrival time and the reflected sound arrive time may be used, or common threshold value data may be used.

[0191]

The threshold value data may be obtained from memory 1404 of audio signal processing device 1001, or may be obtained from an external storage device via a communication module. The threshold value data storage method and the threshold value setting method will be described later.

[0192]

Reproducer 1303 synthesizes the audio signals of direct sounds and the audio signals of reflected sounds selected for generation by determiner 1302.

[0193]

Specifically, reproducer 1303 processes the inputted audio signals to generate direct sounds, based on information on the direct sound arrival time and the direct sound arrival time sound volume calculated by analyzer 1301. Furthermore, reproducer 1303 processes the inputted audio signals to generate reflected sounds,

based on information on the reflected sound arrival time and the reflected sound arrival time sound volume related to the reflected sounds selected by determiner 1302. Then, reproducer 1303 synthesizes and outputs the direct sounds and reflected sounds that were generated.

[0194]

[Operation Example of Renderer]

FIG. 8 is a flowchart illustrating an operation example of audio signal processing device 1001. FIG. 8 illustrates the processing performed mainly by renderer 1300 of audio signal processing device 1001.

[0195]

In the analysis processing of the input signal (S101 in FIG. 8), analyzer 1301 analyzes the input signal inputted into audio signal processing device 1001 to detect direct sounds and reflected sounds that may be generated in the sound space. The reflected sounds detected here are candidates for the reflected sounds to be selected by determiner 1302 as the reflected sounds to be ultimately generated by reproducer 1303. Furthermore, analyzer 1301 analyzes the input signal to calculate information necessary for generating direct sound and reflected sound, and information necessary for selecting the reflected sounds to be generated.

[0196]

First, the characteristics of each of the direct sound and the reflected sound are calculated. Specifically, the arrival time period and the arrival time sound volume when each of the direct sound and the reflected sound arrives at the listener are calculated. When a plurality of objects are present in the sound space as reflection objects, reflected sound characteristics with respect to each of the plurality of objects are calculated.

[0197]

The direct sound arrival time period (t_d) is calculated based on the direct sound arrival path (p_d). The direct sound arrival path (p_d) is a path that connects position information $S(x_s, y_s, z_s)$ of a sound source object with position information $A1(x_a, y_a, z_a)$ of the listener.

The direct sound arrival time period (t_d) is a value obtained by dividing the length of the path that connects position information S (x_s, y_s, z_s) with position information A1 (x_a, y_a, z_a), by the speed of sound (approximately 340 m/sec).

5 [0198]

For example, the path length (X) is determined by the expression $((x_s - x_a)^2 + (y_s - y_a)^2 + (z_s - z_a)^2)^{0.5}$. The sound volume attenuates in inverse proportion to the distance. Thus, when the sound volume at position information S (x_s, y_s, z_s) of a sound source object is denoted by N and the unit distance is denoted by U , the direct sound arrival time sound volume (I_d) is determined by the expression $I_d = N * U/X$.

10 [0199]

Sound volume N at the sound source position may be the reference sound volume described above.

15 [0200]

The reflected sound arrival time period (t_r) is calculated based on the reflected sound arrival path (p_r). The reflected sound arrival path (p_r) is a path that connects the position of the sound image of a reflected sound with position information A1 (x_a, y_a, z_a).

20 [0201]

Note that the position of the sound image of the reflected sound may be derived by using, for example, a "mirror image method" or a "ray tracing method", or by using any other method for deriving sound image positions. The mirror image method is a method that simulates a sound image by assuming that a reflected wave on the wall in a room has a mirror image in a position symmetrical to the sound source relative to the wall, and that sound waves are emitted from the position of that mirror image. The ray tracing method is a method that simulates, for example, an image (sound image) observed at a certain point by tracing waves that are transmitted in a linear manner, such as light rays or sound rays.

25 [0202]

FIG. 9 is a diagram illustrating a comparatively distant positional relationship between a listener and an obstacle object.

35 - 52 -

FIG. 10 is a diagram illustrating a comparatively close positional relationship between a listener and an obstacle object. In other words, each of FIG. 9 and FIG. 10 illustrate an example in which the sound image of a reflected sound is formed in a position symmetrical to the sound source position, with a wall interposed therebetween. Based on such a relationship, by determining the position of the sound image of the reflected sound on the x-, y-, and z-axes, the reflected sound arrival time period can be determined in the same manner as the method for calculating the direct sound arrival time period.

[0203]

The reflected sound arrival time period (t_r) is a value obtained by dividing the length of the path that connects the position of the sound image of a reflected sound with position information A1 (x_a , y_a , z_a), by the speed of sound (approximately 340 m/sec). The sound volume attenuates in inverse proportion to the distance. Thus, when the sound volume at the sound source position is denoted by N , the unit distance is denoted by U , and the attenuation rate of the sound volume at the reflection is denoted by G , the reflected sound arrival time sound volume (I_r) is determined by the expression $I_r = N * G * U/Y$.

[0204]

As described above, attenuation rate G may be expressed as a real number greater than or equal to 0 and less than or equal to 1, or may be expressed as a negative decibel value. In this case, the sound volume of the signal as a whole attenuates by the amount of G . Furthermore, the attenuation rate may be set for each frequency band included in a plurality of frequency bands. In this case, analyzer 1301 multiplies each frequency component of the signal by the specified attenuation rate. Furthermore, in order to reduce the amount of computation, analyzer 1301 may, by using, as an overall attenuation rate, a representative value, an average value, or the like of a plurality of attenuation rates of a plurality of frequency bands, cause the sound volume of the signal as a whole to attenuate by that amount.

[0205]

Next, analyzer 1301 calculates the sound volume ratio (L), which is the ratio of the reflected sound arrival time sound volume (Ir) to the direct sound arrival time sound volume (Id), and the time difference (T) between the direct sound and the reflected sound, each
5 of the sound volume ratio (L) and the time difference (T) being required for selection of the reflected sound to be generated.

[0206]

The sound volume ratio (L), which is the ratio of the above-described Ir to the direct sound arrival time sound volume (Id), is, for
10 example, a value obtained by dividing the reflected sound arrival time sound volume (Ir) by the direct sound arrival time sound volume (Id), and is determined by: $L = (N * G * U/Y)/(N * U/X) = G * X/Y$. Since the value to be determined is a sound volume ratio, the values of N and U may be any predetermined values.

15 [0207]

The time difference (T) between a direct sound and a reflected sound may be, for example, the time difference between the time periods each of the direct sound and the reflected sound take to arrive at the listening position. For example, the time difference (T)
20 between the time periods taken for each of a direct sound and a reflected sound to arrive at the listening position is determined by $T = t_r - t_d$.

[0208]

Furthermore, the time difference (T) may be the difference
25 between the times at which each of a direct sound and a reflected sound arrive at the listening position. Moreover, the time difference (T) may be the time difference between the time at which the emission of the direct sound ends and the time at which the reflected sound arrives at the listening position. In other words, the time
30 difference (T) may be the time difference, at the listening position, between the time at which the direct sound ends and the time at which the reflected sound begins.

[0209]

Next, in reflected sound selection processing (S102 in FIG. 8),
35 determiner 1302 selects whether reproducer 1303 is to generate a

reflected sound calculated by analyzer 1301. To put it differently, determiner 1302 determines whether to select a reflected sound as a reflected sound to be generated. When there are a plurality of reflected sounds, determiner 1302 selects whether to generate each reflected sound. As the result of selecting whether to generate each reflected sound, determiner 1302 may select one or more reflected sounds to be generated from the plurality of reflected sounds, or may select one reflected sound to be generated.

[0210]

Note that determiner 1302 may select reflected sounds to which other processing is to be applied, not limited to generation processing. For example, determiner 1302 may select reflected sounds to which binaural processing is to be applied. Furthermore, determiner 1302 fundamentally selects only the one or more reflected sounds that are to be processed. However, determiner 1302 may select only one or more reflected sounds that are not to be processed. Processing may then be applied to the one or more reflected sounds that were not selected.

[0211]

For example, the selection of reflected sounds may be performed based on the sound volume ratio (L) and the time difference (T) calculated by analyzer 1301. Due to the selection processing being performed based on the time difference (T) between direct sounds and reflected sounds, it is possible to more appropriately select reflected sounds that have a large degree of influence on the listener's perception, in comparison to when performing the selection processing based only on the sound volume difference between direct sounds and reflected sounds.

[0212]

Specifically, the selection of whether to generate a reflected sound is performed by comparing, to a preset threshold value, the sound volume ratio of a reflected sound to a direct sound, the sound volume ratio corresponding to the time difference between the direct sound and the reflected sound. The threshold value is set with reference to the threshold value data. The threshold value data is

an indicator indicating the boundary that demarcates whether a reflected sound corresponding to a direct sound is perceived by the listener, and is defined as the ratio between the direct sound arrival time sound volume (I_d) and the reflected sound arrival time sound volume (I_r).

[0213]

Note that the threshold value corresponds to a value expressed by, e.g., a numerical value determined based at the time difference (T). The threshold value data corresponds to the relationship between the time difference (T) and a threshold value, and corresponds to table data or a relational expression used for identifying or calculating the threshold value at the time difference (T). The format and type of the threshold value data is not limited to table data or a relational expression.

[0214]

FIG. 11 is a diagram illustrating relationships between time differences between direct sounds and reflected sounds, and threshold values. For example, threshold value data of predetermined sound volume ratios may be referenced for each value of the time difference between a direct sound and a reflected sound, as illustrated in FIG. 11. Alternatively, threshold value data obtained by, e.g., interpolating or extrapolating from the threshold value data illustrated in FIG. 11 may be referenced.

[0215]

Furthermore, the threshold value of the sound volume ratio at the time difference (T) calculated by analyzer 1301 is identified from the threshold value data. Moreover, determiner 1302 determines whether to select a reflected sound as a reflected sound to be generated based on whether the sound volume ratio (L) of the reflected sound to the direct sound calculated by analyzer 1301 exceeds the threshold value.

[0216]

Due to performing the selection processing by using the threshold value data of the sound volume ratio that is predetermined for each value of the time difference between a direct sound and a

reflected sound, selection processing that considers post-masking or the precedence effect can be achieved. The type, format, storage method, setting method, and the like of the threshold value data will be described in detail later.

5 [0217]

Next, in the generation processing of direct sounds and reflected sounds (S103 in FIG. 8), reproducer 1303 generates and synthesizes the audio signals for direct sounds and the audio signals for reflected sounds that have been selected by determiner 1302 as reflected sounds to be generated.

[0218]

The audio signals for direct sounds are generated by applying the direct sound arrival time period (t_d) and the direct sound arrival time sound volume (I_d) calculated by analyzer 1301 to the sound data for the sound source objects included in the input signal. Specifically, processing in which the sound data is delayed by the amount of the direct sound arrival time period (t_d) and multiplied by the direct sound arrival time sound volume (I_d) is performed. The processing to delay the sound data is processing in which the position of the sound data is moved forward or backward on the time axis. Processing in which the sound data is delayed may be applied without causing the sound quality to deteriorate, such as was disclosed in PTL 2.

[0219]

The audio signals for reflected sounds are, similarly to the direct sounds, generated by applying the reflected sound arrival time period (t_r) and the reflected sound arrival time sound volume (I_r) calculated by analyzer 1301 to the sound data for the sound source objects.

[0220]

However, the reflected sound arrival time sound volume (I_r) in the generation of reflected sounds differs from the arrival time sound volume of the direct sounds in that the arrival time sound volume of the reflected sounds is a value to which attenuation rate G of the sound volume in the reflection has been applied. G may be an attenuation rate that is applied globally to all frequency bands.

Alternatively, in order to reflect the biases of frequency components generated by reflection, the reflectance may be defined for each predetermined frequency band. In this case, the processing to apply the reflected sound arrival time sound volume (I_r) may be performed as frequency equalizer processing, which is processing that involves multiplying each band by the attenuation rate.

[0221]

In the above example, for each of the direct sounds and the reflected sound candidates, the path length when arriving at the listener is calculated. Furthermore, the arrival time period and the arrival time sound volume are calculated based on each path length. The selection processing of the reflected sound candidates is then performed based on the time differences and the sound volume ratios of these.

[0222]

Note that as a different example, the selection processing may be performed based on the path lengths when each of the direct sound and the reflected sound arrives at the listener, and the calculation of the arrival time period and the arrival time sound volume of each of the direct sound and the reflected sound and the calculation of the time difference and the sound volume ratio may be omitted. In this case, threshold values according to path length differences may be determined beforehand with respect to path length ratios. Then, selection processing may be performed based on whether the path length ratio calculated is greater than or equal to the threshold value according to the path length difference calculated. This makes it possible to perform selection processing based on path length differences that correspond to time differences, while reducing the amount of computation.

[0223]

Furthermore, a parameter that indicates sound propagation speed or a parameter that has an impact on the sound propagation speed parameter may be used in addition to the path length difference.

[0224]

(Details of Selection Processing)

The selection processing that determines whether reflected sounds are generated will be explained in detail.

[0225]

The selection of a reflected sound is performed by comparing, with the sound volume ratio (L) calculated by analyzer 1301, the threshold value determined for the sound volume ratio, which is the ratio of the reflected sound arrival time sound volume to the direct sound arrival time sound volume, at the time difference (T) between the direct sound and the reflected sound. For example, of threshold values of sound volume ratios that were determined beforehand for each value of a time difference between a direct sound and a reflected sound, the threshold value of the sound volume ratio at the time difference (T) between the direct sound and the reflected sound calculated by analyzer 1301 is referenced. Then, determination of whether the reflected sound is selected as a reflected sound to be generated is made based on whether the sound volume ratio (L) calculated by analyzer 1301 exceeds the threshold value.

[0226]

The time difference (T) may be any of, for example, the difference in the times at which each of a direct sound and a reflected sound arrive at the listening position, the time difference between the time periods taken when each of a direct sound and a reflected sound arrives at the listening position, or the time difference between the time point when emission of a direct sound stops and the time point when a reflected sound arrives at the listening position. Here, the direct sound end time may be determined by adding the duration of a direct sound to the arrival time of the direct sound.

[0227]

The threshold value data may be determined based on the minimum time difference at which the perception of the listener is able to detect the divergence of two sounds due to an action of the auditory nerve or a cognitive effect in the brain, and more specifically due to the precedence effect, described later, the temporal masking phenomenon, described later, or a combination of both. Specific numerical values may be derived from research results into the

temporal masking effect, the precedence effect, the echo detection limit, etc. that are already known, or may be determined by an auditory test performed with the premise of application in the virtual space.

5 [0228]

FIG. 12A, FIG. 12B, and FIG. 12C are diagrams illustrating examples of threshold value data setting methods. As illustrated in FIG. 12A, FIG. 12B, and FIG. 12C, the threshold value data represents the boundaries (threshold values) determining whether reflected
10 sound is perceived or not perceived, in a graph having a horizontal axis that indicates the time difference between direct sound and reflected sound and a vertical axis that indicates the sound volume ratio of the reflected sound to the direct sound.

[0229]

15 The threshold value data may be expressed by an approximation formula that includes the time difference between direct sound and reflected sound as a variable. Furthermore, as illustrated in FIG. 11, the threshold value data may be stored in the domain of memory 1404 as an arrangement of an index of time
20 differences between direct sounds and reflected sounds, and threshold values corresponding to the index.

[0230]

Note that when the height of a line parallel to the horizontal axis in Example 4 in FIG. 12C (the minimum audible limit) is used as
25 the threshold value, what is compared to the threshold value is not the sound volume ratio (L) between a direct sound and a reflected sound, but the sound volume of the reflected sound itself. The reason for this is that the threshold value indicates the sound volume of the boundary demarcating whether a sound is perceivable to the
30 listener, and is a threshold value for determining a sound having a lower sound volume than the threshold value to be a sound that is not to be reproduced. In other words, the threshold value corresponding to the minimum audible limit is not a threshold value with respect to the ratio of the sound volume of a reflected sound to
35 the sound volume of a direct sound.

[0231]

When the minimum audible limit is used as the threshold value, the threshold value is constant regardless of the time difference (T); thus, the time difference (T) need not be calculated.

5 [0232]

Note that when a plurality of reflected sounds are generated in the analysis processing (S101 in FIG. 8), the selection processing may be performed on all of the reflected sounds, or the selection processing may be performed on only the reflected sounds having
10 high evaluation values based on the evaluation values derived for each reflected sound by means of a preset evaluation method. Here, the evaluation value of a reflected sound corresponds to the sensory level of importance of the reflected sound. Note that the evaluation value being high corresponds to the evaluation value being large, and
15 these expressions may be used interchangeably.

[0233]

Determiner 1302 may calculate an evaluation value for each reflected sound by an evaluation method set beforehand based on, for example, the sound volume of the sound source, the visual
20 properties of the sound source, the positionality of the sound source, the visual properties of the reflection object (the obstacle object), the geometrical relationship between the direct sound and the reflected sound, and/or the like.

[0234]

Specifically, the evaluation value may become higher as the sound volume of the sound source is greater. Furthermore, in order to cause visual positioning and acoustic positioning to match each other, the evaluation value may be high when a sound source object or a reflection object (obstacle object) is visible from the listener, or
30 when the positionality of a sound source object is high.

[0235]

Moreover, the size of the arrival angle formed by a direct sound and a reflected sound and the difference between the arrival time periods of a direct sound and a reflected sound greatly affect the
35 grasping of the space. Thus, the evaluation value may be high when

the size of the angle formed by the arrival of a direct sound and the arrival of a reflected sound is large, or when the difference between the arrival time periods of a direct sound and a reflected sound is large.

5 [0236]

The sound volume information on the sound source may indicate the reference sound volume determined for each content, a temporal transition of sound volume, or both.

[0237]

10 For example, when the virtual space is a virtual conference room and the direct sound is a speaking voice, the sound volume transitions intermittently over short periods of time. In other words, sound portions and silent portions occur alternately. Furthermore, when the virtual space is a concert hall and the direct sound is the
15 performance of a musical piece, the sound volume is maintained over a certain duration of time. Moreover, when the virtual space is a battlefield and the direct sound is an explosion sound, the sound volume becomes large for only an instant and then continues to be silent or in a quiet state thereafter.

20 [0238]

In this way, the sound volume information on the sound source may include not only information on the reference sound volume corresponding to the volume setting when the sound is radiated into the virtual space, but also information on the transition of the sound
25 magnitude.

[0239]

The information on the transition may be expressed by data indicating frequency characteristics in chronological order. The information on the transition may be expressed by data indicating the
30 duration of a sound interval. The information on the transition may be expressed by data indicating the chronological order of durations of sound intervals and durations of silent intervals. The information on the transition may be expressed by, for example, data that enumerates, in chronological order, a plurality of sets of a duration
35 for which the amplitude of the sound signal can be considered

stationary (can be considered approximately constant) and the amplitude value of said signal during that duration.

[0240]

5 The information on the transition may be expressed by data of a duration during which the frequency characteristics of the sound signal can be considered stationary. The information on the transition may be expressed by, for example, data that enumerates, in chronological order, a plurality of sets of a duration for which the frequency characteristics of the sound signal can be considered
10 stationary and the frequency characteristics during that duration.

[0241]

Furthermore, efforts to use temporal transitions in the frequency characteristics of signals for the acoustic processing of virtual spaces have been conventionally widely performed (PTL 1,
15 etc.). When considering such conventional techniques, it goes without saying that the above-described sets may be sets that include durations of time in which the frequency characteristics are constant and those frequency characteristics.

[0242]

20 The geometrical relationships may be positional relationships between a sound source, the listener, and a reflection object in a virtual space. The path lengths of each of the direct sound and the reflected sound arriving can be geometrically calculated based on these relationships. Therefore, using the relationship that sound
25 volume is inversely proportional to distance, it is possible to calculate the reference sound volume of a reflected sound with respect to the reference sound volume of a direct sound.

[0243]

A reflection coefficient of the reflection object may be used for
30 calculation of the reference sound volume of a reflected sound. Furthermore, a generally used typical value may be used as the reflection coefficient. On the other hand, when special conditions are present, such as the reflection object being covered by a sound-absorbing material or the like, a specially assigned reflection
35 coefficient may be used as the reflection coefficient of the reflection

object.

[0244]

Each reflected sound may be evaluated based on the sound volume of the reflected sound. The sound volume of the reflected sound may, as described above, be determined from the geometrical relationship between the direct sound and the reflected sound, and an indicator assigned to the reflection object. The reflected sound may be evaluated by comparing that sound volume with a predetermined threshold value.

[0245]

Furthermore, information that indicates a temporal transition in the sound volume of the sound source may be reflected in the evaluation. For example, in a case in which the information that indicates a temporal transition in the sound volume of the sound source indicates the duration of a sound interval, the reflected sound evaluation value may be maintained as-is when the time is within the sound interval. On the other hand, when the time is outside of the sound interval, processing to reduce or make zero the evaluation value of the reflected sound may be performed, even if the reference sound volume of the reflected sound exceeds the threshold value.

[0246]

Alternatively, the information that indicates a temporal transition in the sound volume of the sound source may be data that lists, in chronological order, a plurality of sets of a duration during which it is considered that the amplitude of the sound signal is largely constant and amplitude values of the signal during that period. In that case, processing may be performed such that reflected sounds are evaluated by changing the reference sound volume of the reflected sounds in coordination with changes in the amplitude values in the data.

[0247]

Furthermore, both the information on the reference sound volume and the information on the temporally transitioning sound volume may be used as the information indicating the sound volume of a direct sound. For example, an evaluation value can be calculated

based on the information on the reference sound volume, and then the evaluation value can be corrected by using the information on the transitioning sound volume.

[0248]

5 In the evaluation of reflected sounds, all of the methods described above may be performed, or only some of the methods may be performed. For example, reflected sounds may be evaluated using a plurality of evaluation methods, or reflected sounds may be evaluated using one evaluation method.

10 [0249]

 When reflected sounds are evaluated using a plurality of evaluation methods, whether to select each reflected sound may be determined based on an evaluation value comprehensively determined by the plurality of evaluation methods, or may be
15 determined based on the evaluation value of each of the plurality of evaluation methods.

[0250]

 When whether to select each reflected sound is determined based on each of the plurality of evaluation methods, audio signal
20 processing device 1001 may select a sound when all of the plurality of evaluation results based on the plurality of evaluation methods indicate selecting the sound. Alternatively, audio signal processing device 1001 may select a sound when any one of the plurality of the evaluation results based on the plurality of evaluation methods
25 indicates selecting the sound.

[0251]

 Furthermore, for example, an order of priority may be assigned to the first to third evaluation methods. When it is determined by the first evaluation method that a sound is not to be selected, audio
30 signal processing device 1001 may then make a final determination that the sound is not to be selected, without depending on the determination results in the second and third evaluation methods. Moreover, when, although it has been determined by one of the second and third evaluation methods that a sound is not to be selected,
35 it has been determined by the other of the second and third evaluation

methods that the sound is to be selected, audio signal processing device 1001 may make a final determination that the sound is to be selected.

[0252]

5 Furthermore, the selection processing and the evaluation processing may be performed independently of each other, or only one of these may be performed. Moreover, the evaluation processing may be performed only on reflected sounds that have been determined to be selected in the selection processing, and whether to
10 select each reflected sound may be redetermined in the evaluation processing. Alternatively, the evaluation processing may be performed only on reflected sounds that have been determined to not be selected in the selection processing, and whether to select each reflected sound may be redetermined in the evaluation processing.

15 [0253]

 The selection processing described above can be interpreted as processing in which a reflected sound is selected in accordance with a characteristic of a direct sound. For example, in processing in which a reflected sound is selected in accordance with a characteristic
20 of a direct sound, the threshold value used in selection of the reflected sound is set or adjusted in accordance with a characteristic of the direct sound. Alternatively, the evaluation value used in the selection of a reflected sound may be calculated based on one or more of, for example, the sound volume of the sound source, the visual
25 properties of the sound source, the positionality of the sound source, the visual properties of the reflection object (the obstacle object), the geometrical relationship between the direct sound and the reflected sound, and/or the like.

[0254]

30 Furthermore, the processing in which a reflected sound is selected based on a characteristic of a direct sound is not limited to processing in which the threshold value is set or adjusted in accordance with a characteristic of the direct sound and processing in which the evaluation value used for selection of the reflected sound
35 to be processed is calculated, and other processes may be performed.

Furthermore, even when performing the processing in which the threshold value is set or adjusted in accordance with a characteristic of the direct sound or the processing in which the evaluation value used in selection of the reflected sounds to be processed is calculated, the processing may be partially changed, or new processing may be added.

[0255]

Note that setting the threshold value may include adjusting the threshold value, changing the threshold value, and the like.

[0256]

[Threshold Value Setting Method]

The threshold value data used in the selection processing may be set with reference to the value of an echo detection limit based on a known precedence effect or a masking threshold value based on the post-masking effect.

[0257]

The precedence effect is a phenomenon in which, when sounds are heard from two locations, it is perceived that the sound source is present at the location from which the first sound was heard. If two short sounds fuse together to be heard as one sound, the position (localization position) from which the overall sound is heard is, for the most part, determined by the position of the first sound. The echo detection limit is a phenomenon that occurs due to the precedence effect, and is the minimum time difference at which the listener's perception detects the divergence of two sounds.

[0258]

In Example 2 of FIG. 12C, the horizontal axis corresponds to the arrival time period of reflected sound (echo), and specifically corresponds to the delay time period from the arrival time of direct sound to the arrival time of reflected sound. The vertical axis corresponds to the sound volume ratio of detectable reflected sound to direct sound, and specifically corresponds to the threshold value that determines whether reflected sound that has arrived with a delay time period is detectable.

[0259]

FIG. 13 is a diagram illustrating an example of a threshold value setting method. The horizontal axis in FIG. 13 corresponds to the arrival time period of reflected sound, and specifically corresponds to the time differences (T) between direct sound and reflected sound.

5 The vertical axis in FIG. 13 corresponds to the sound volume of reflected sound. Specifically, the vertical axis in FIG. 13 may correspond to the sound volume (sound volume ratios) of reflected sound determined in relation to direct sound, or may correspond to the sound volume of reflected sound determined absolutely without
10 depending on the sound volume of the direct sound.

[0260]

For example, when, as illustrated in FIG. 9, the listener and an obstacle object are comparatively far from each other, the arrival time period of the reflected sound becomes longer, and, as illustrated in C
15 in FIG. 13, the threshold value is set to be low. As a result, in the case of FIG. 9, the reflected sound is generated. On the other hand, when, as illustrated in FIG. 10, the listener and the obstacle object are comparatively close to each other, the arrival time period of the reflected sound is shorter than that in the case of FIG. 9, and as
20 illustrated in B in FIG. 13, the threshold value is set to be high. As a result, in the case of FIG. 10, the reflected sound is not generated.

[0261]

Furthermore, the threshold value data may be stored in memory 1404, obtained from memory 1404 at the time of the
25 selection processing, and used in the selection processing.

[0262]

FIG. 14 is a flowchart illustrating an example of selection processing. First, determiner 1302 specifies a reflected sound detected by analyzer 1301 (S201). Determiner 1302 then detects
30 the sound volume ratio (L) of the reflected sound to the direct sound, and the time difference (T) between the direct sound and the reflected sound (S202 and S203).

[0263]

The time difference (T) may be any of, for example, the time
35 difference between the time periods each of the direct sound and the

reflected sound take to arrive at the listening position, the time difference between the direct sound arrival time and the reflected sound arrival time, and the time difference between the time when emission of the direct sound ends and the time when the reflected sound arrives at the listening position. Here, an example will be described based on the time difference between the direct sound arrival time and the reflected sound arrival time.

[0264]

Specifically, based on: the position information on the sound source object and the listener; and the position information and geometry information on the obstacle object, determiner 1302 calculates the difference between the length of the path of the direct sound and the length of the path of the reflected sound. By dividing the difference between the lengths by the speed of sound, determiner 1302 then detects the time difference (T) between the time when the direct sound arrives at the listener's position and the time when the reflected sound arrives at the listener's position.

[0265]

The sound volume when arriving at the listener attenuates, with respect to the sound volume of the sound source, in proportion to the distance to the listener (in inverse proportion to the distance). Therefore, the sound volume of the direct sound is obtained by dividing the sound volume of the sound source by the length of the path of the direct sound. The sound volume of the reflected sound is obtained by dividing the sound volume of the sound source by the length of the path of the reflected sound, and then further multiplying by the attenuation rate assigned to the virtual obstacle object. Determiner 1302 detects the sound volume ratio by calculating the ratio between these sound volumes.

[0266]

Furthermore, using the threshold value data, determiner 1302 identifies the threshold value corresponding to the time difference (T) (S204). Determiner 1302 then determines whether the sound volume ratio (L) detected is greater than or equal to the threshold value (S205).

[0267]

When the sound volume ratio (L) is greater than or equal to the threshold value ("Yes" in S205), determiner 1302 selects the reflected sound as a reflected sound to be generated (S206). When the sound volume ratio (L) is less than the threshold value ("No" in S205), determiner 1302 skips selecting the reflected sound as a reflected sound to be generated (S207). That is, in this case, determiner 1302 determines the reflected sound to be a reflected sound that is not to be generated, i.e., a reflected sound to be culled.

[0268]

Subsequently, determiner 1302 determines whether there are any unspecified reflected sounds (S208). If there are unspecified reflected sounds ("Yes" in S208), determiner 1302 repeats the above-described processing (S201 to S207). If there are no unspecified reflected sounds ("No" in S208), determiner 1302 ends the processing.

[0269]

This selection processing may be performed on all of the reflected sounds generated in the analysis processing, or may be performed on only the reflected sounds for which the above-described evaluation value is high.

[0270]

[Details of Threshold Value Storing Method]

The threshold value data according to the present embodiment is stored in memory 1404 of audio signal processing device 1001.

The format and type of the threshold value data to be stored may be any format and any type. When threshold values having a plurality of formats and a plurality of types are stored, in the selection processing, the format and the type of the threshold values to be used in the selection processing of the reflected sounds may be decided.

The method for determining which items of threshold value data to use in the selection processing will be described later.

[0271]

Furthermore, a plurality of formats and a plurality of types of threshold value data may be stored in combination. The combined threshold value data may be read from spatial information managers

1201 and 1211 to set the threshold values to be used in the selection processing. Note that the threshold value data to be stored in memory 1404 may be stored in spatial information managers 1201 and 1211.

5 [0272]

For example, the threshold value data may be stored as threshold values at each time difference, so as to plot a line between the threshold values as illustrated in [Example 1] and [Example 2] of FIG. 12C.

10 [0273]

Furthermore, the threshold value data may be stored as table data in which, as illustrated in FIG. 11, the threshold values and the time differences (T) are associated with each other. In other words, the threshold value data may be stored as table data that includes the time differences (T) as an index. Naturally, the threshold values illustrated in FIG. 11 are examples, and the threshold values are not limited to the examples in FIG. 11. Furthermore, the threshold values may be approximated by functions that include the time differences (T) as variables, and coefficients of the functions may be stored, without storing the threshold values themselves. Moreover, a plurality of approximation expressions may be combined and stored.

[0274]

For example, the threshold value data may be expressed by a formula such as that shown below, where timeDiff denotes the time difference (T), and gainThresh denotes the threshold value.

[0275]

[Math. 1]

$$\text{gainThresh} = \begin{cases} -0.10 * \text{timeDiff} - 17.5 & , \text{if } 1.0 < \text{timeDiff} < 10.0 \\ -0.55 * \text{timeDiff} - 13.0 & , \text{if } 10.0 \leq \text{timeDiff} < 20.0 \\ -0.60 * \text{timeDiff} - 12.0 & \text{if } 20.0 \leq \text{timeDiff} < 30.0 \\ -0.80 * \text{timeDiff} - 6.0 & , \text{if } 30.0 \leq \text{timeDiff} \leq 40.0 \end{cases}$$

[0276]

30 The threshold value is defined only over the time range in which the precedence effect is considered to occur. When the time difference is outside of that time range (in the above formula, a value

of 1 ms or less or 40 ms or greater), the determination may be performed not by gainThresh, but solely by a threshold value representing the minimum sound volume for reproduction in the virtual space, described later.

5 [0277]

Experiments by the present inventors have clarified that in the time range in which the precedence effect is considered to occur, the threshold value may be approximated by a function that is convex upward. The above-described formula is an example of an approximation formula generated based on these experiments.

10 [0278]

Information on a relational expression that indicates the relationship between time differences (T) and threshold values may be stored in memory 1404. In other words, an expression that includes the time difference (T) as a variable may be stored. The threshold values of the time differences (T) may be approximated by a straight line or a curved line, and a parameter that indicates the geometrical shape of the straight line or the curved line may be stored. For example, when the geometrical shape is a straight line, the start point and the slope for expressing the straight line may be stored.

20 [0279]

Furthermore, the threshold value data may be stored having the type and format thereof defined for each characteristic of direct sound. Moreover, parameters for adjusting threshold values based on a characteristic of the direct sound and using the threshold values in the selection processing may be stored. Processing to adjust threshold values in accordance with a characteristic of the direct sound and use the threshold values in the selection processing is described later, as a variation of the threshold value setting method.

30 [0280]

As an example in which a plurality of types of threshold value data are stored in combination, as illustrated in [Example 3] in FIG. 12C, for each time difference (T), the larger value of the masking threshold value and the echo detection limit threshold value may be stored. As illustrated in [Example 4] in FIG. 12C, for each time

difference (T), the larger value of the minimum sound volume for reproduction in a virtual space and the echo detection limit threshold value may be stored.

[0281]

5 The combination of the plurality of types of the threshold value data is not limited to these. For example, in a plurality of items of threshold value data, information on the maximum value may be stored for each time difference (T).

[0282]

10 Furthermore, in the above description, the information on threshold values has time period items as a one-dimensional index. The information on threshold values may have a two-dimensional or three-dimensional index that further includes variables related to the direction of arrival.

15 [0283]

FIG. 15 is a diagram illustrating relationships between directions of direct sounds, directions of reflected sounds, time differences, and threshold values. For example, as illustrated in FIG. 15, threshold values pre-calculated in accordance with the relationship between the direct sound direction (θ), the reflected sound direction (γ), the time difference (T), and the sound volume ratio (L) may be stored.

[0284]

25 The direct sound direction (θ) corresponds to the angle, with respect to the listener, of the direction of arrival of a direct sound. The reflected sound direction (γ) corresponds to the angle, with respect to the listener, of the direction of arrival of a reflected sound. Here, the direction in which the listener is facing is defined as 0 degrees. The time difference (T) corresponds to the difference, until reaching the listening position, of the direct sound arrival time period and the reflected sound arrival time period. The sound volume ratio (L) corresponds to the sound volume ratio of the reflected sound arrival time sound volume to the direct sound arrival time sound volume.

35 [0285]

Naturally, the threshold values illustrated in FIG. 15 are examples, and the threshold values are not limited to the examples in FIG. 15. Furthermore, in FIG. 15, mainly threshold values when the angle (θ) of the direct sound arrival direction is 0 degrees are exemplified. However, threshold values when the direct sound arrival direction (θ) is not 0 degrees are also stored in memory 1404.

[0286]

Further, in the above description, the threshold values are stored in an arrangement that has the angle (θ) of the direct sound (more specifically, the angle (θ) of the direct sound arrival direction) and the angle (γ) of the reflected sound (more specifically, the angle (γ) of the reflected sound arrival direction) as independent variables or indexes. However, the angle (θ) of the direct sound arrival direction and the angle (γ) of the reflected sound arrival direction need not be used as independent variables.

[0287]

For example, the angular difference between the angle (θ) of the direct sound arrival direction and the angle (γ) of the reflected sound arrival direction may be used. This angular difference corresponds to the angle formed between the direct sound arrival direction and the reflected sound arrival direction, and may be expressed as the arrival angle between a direct sound and a reflected sound.

[0288]

FIG. 16 is a diagram illustrating relationships between angular differences, time differences, and threshold values. For example, threshold values pre-calculated by using, as a variable, the angular difference (Φ) between the angle (θ) of the direct sound arrival direction and the angle (γ) of the reflected sound arrival direction may be stored as in the example illustrated in FIG. 16. Naturally, the threshold values illustrated in FIG. 16 are examples, and the threshold values are not limited to the examples in FIG. 16.

[0289]

In the example in FIG. 16, the number of variables used for deriving threshold values may be reduced. Thus, it is possible to

reduce the number of threshold values stored in memory 1404. Therefore, it is possible to decrease the amount of data stored in memory 1404.

[0290]

5 Furthermore, when the angular difference (Φ) between the angle (θ) of the direct sound arrival direction and the angle (γ) of the reflected sound arrival direction is used, the threshold value data may be stored in a two-dimensional arrangement. Moreover, in the selection processing, the difference between the angle (θ) of the
10 direct sound arrival direction and the angle (γ) of the reflected sound arrival direction may be calculated by using a three-dimensional arrangement.

[0291]

 The method for selecting reflected sounds using threshold
15 values based on the directions of arrival will be described later.

[0292]

[First Variation of Threshold Value Setting Method]

 In the examples in FIG. 12A, FIG. 12B, and FIG. 12C, threshold values in a plurality of formats and of a plurality of types may be
20 stored in spatial information managers 1201 and 1211. Then, of the threshold values having a plurality of formats and a plurality of types, the format and the type of the threshold values to be used in the selection processing of the reflected sounds may be decided. Specifically, as illustrated in Example 3 of FIG. 12C, in the time
25 differences (T) corresponding to the reflected sound arrival times, the largest threshold value may be adopted.

[0293]

 Moreover, as illustrated in Example 4, the masking threshold value, the echo detection limit threshold value, and a threshold value
30 indicating the minimum sound volume for reproduction in the virtual space may be stored. Then, the largest threshold value at the time difference (T) corresponding to the reflected sound arrival time may be adopted.

[0294]

35 [Second Variation of Threshold Value Setting Method]

As another example of the threshold value setting method, a method for setting threshold values in accordance with a characteristic of direct sounds will be described.

[0295]

5 FIG. 17 is a block diagram illustrating another configuration example of renderer 1300 illustrated in FIG. 7. Renderer 1300 in FIG. 17 is different from renderer 1300 in FIG. 7 in the respect that renderer 1300 in FIG. 17 includes threshold value adjuster 1304. The description other than threshold value adjuster 1304 is the same as the matters described regarding FIG. 7, and has thus been omitted.

[0296]

Threshold value adjuster 1304 selects, from the threshold value data, threshold values that are to be used by determiner 1302, based on information indicating a characteristic of an audio signal. Alternatively, threshold value adjuster 1304 may adjust the threshold values included in the threshold value data, based on the information indicating a characteristic of the audio signal.

[0297]

20 The information indicating a characteristic of the audio signal may be included in the input signal. Then, threshold value adjuster 1304 may obtain the information indicating a characteristic of the audio signal from the input signal. Alternatively, analyzer 1301 may derive a characteristic of the audio signal by analyzing the audio signal included in the input signal accepted by analyzer 1301, and output the information indicating the characteristic of the audio signal to threshold value adjuster 1304.

[0298]

30 The information indicating a characteristic of the audio signal may be obtained before starting the rendering processing, or may be obtained each time during rendering.

[0299]

Furthermore, threshold value adjuster 1304 need not be included in audio signal processing device 1001; another transmission device may have the role of threshold value adjuster 1304. In this case, analyzer 1301 or determiner 1302 may obtain, from the other

transmission device via communication I/F 1403, the information indicating a characteristic of the audio signal, the threshold value data corresponding to the characteristic, or information for adjusting the threshold value data in accordance with the characteristic.

5 [0300]

FIG. 18 is a flowchart illustrating another example of selection processing. FIG. 19 is a flowchart illustrating yet another example of selection processing. In FIG. 18 and FIG. 19, the threshold value is set in accordance with a characteristic of the direct sound. Specifically, in FIG. 18, threshold value adjuster 1304 identifies a threshold value from the threshold value data, based on the time difference (T) and a characteristic of the audio signal. In FIG. 19, threshold value adjuster 1304 adjusts, based on a characteristic of the audio signal, the threshold value identified from the threshold value data based on the time difference (T).

15 [0301]

Hereinafter, the operations of each example will be described. Note that description has been omitted for processes that are shared with the example in FIG. 14.

20 [0302]

First, an example of the processing illustrated in FIG. 18 will be described. Here, the threshold value data is stored beforehand in memory 1404 for each nature of direct sound. Accordingly, a plurality of items of threshold value data corresponding to a plurality of natures are stored beforehand in memory 1404. Then, threshold value adjuster 1304 identifies, from the plurality of items of threshold value data, the threshold value data to be used in the selection processing of reflected sounds.

[0303]

30 For example, threshold value adjuster 1304 obtains a characteristic of a direct sound based on the input signal (S211). Threshold value adjuster 1304 may obtain a characteristic of the direct sound that is associated with the input signal. Threshold value adjuster 1304 may then identify the threshold value corresponding to the time difference (T) and the characteristic of the direct sound

35

(S212).

[0304]

Furthermore, as illustrated in FIG. 19, threshold value adjuster 1304 may adjust the threshold value identified by determiner 1302, based on the characteristic of the direct sound (S221).

[0305]

In any of these cases, the information indicating a characteristic of the audio signal, the information for adjusting the threshold value in accordance with the characteristic of the audio signal, or both of these may be included in the input signal. Threshold value adjuster 1304 may adjust the threshold value using one or both of these.

[0306]

Furthermore, the information indicating a characteristic of the audio signal, the information for adjusting the threshold value, or both of these may be transmitted by another input signal aside from the input signal that includes the audio signal. In this case, information for associating the other input signal aside from the input signal may be included in the input signal that includes the audio signal, or information for associating the other input signal with the input signal may be stored in memory 1404 together with the information on threshold values.

[0307]

In the examples in FIG. 18 and FIG. 19, the threshold value used in selecting each reflected sound is set in accordance with a characteristic of the direct sound, that is, a characteristic of the audio signal. Threshold value data preset for each characteristic may be used, as in FIG. 18, or the threshold value may be adjusted in accordance with the characteristics of the audio signal, as in FIG. 19. Furthermore, threshold value data parameters may be adjusted in accordance with the characteristic of the audio signal.

[0308]

Moreover, the operations performed by threshold value adjuster 1304 may be performed by analyzer 1301 or determiner 1302. For example, analyzer 1301 may obtain a characteristic of the audio

signal. Furthermore, determiner 1302 may set threshold values in accordance with the characteristic of the audio signal.

[0309]

5 Next, the relationship between the characteristic of the audio signal and the threshold value will be described.

[0310]

10 Two short sounds that arrive at the listener's ears in succession are heard as one sound if the time period interval between the two short sounds is sufficiently short. This phenomenon is referred to as the precedence effect. The precedence effect is known to only occur with respect to unconnected sounds, that is, transient sounds (NPL 1). Thus, when an audio signal indicates a stationary sound, the echo detection limit may be set lower than when the audio signal indicates a non-stationary sound.

15 [0311]

 In other words, the threshold value is set low in accordance with the characteristics of this precedence effect when, for example, a direct sound is a stationary sound. Furthermore, the threshold value may be set lower as the stationarity is greater.

20 [0312]

 An example of processing when the characteristic of the audio signal is stationary will be explained. First, threshold value adjuster 1304 or analyzer 1301 makes a determination on the stationarity based on the amount of variation in a frequency component of an audio signal accompanying the passage of time. For example, when the amount of variation is small, it is determined that the stationarity is high. Conversely, when the amount of variation is great, it is determined that the stationarity is low. As a result of the determination, a graph indicating the level of stationarity may be set, or a parameter indicating the stationarity in accordance with the amount of variation may be set.

30 [0313]

 Next, threshold value adjuster 1304 adjusts the threshold value data or the threshold values based on information indicating the stationarity, such as the graph or the parameter indicating the

35 stationarity, such as the graph or the parameter indicating the

stationarity of the audio signal, and sets the adjusted threshold value data or threshold values as threshold value data or threshold values to be used by determiner 1302.

[0314]

5 Alternatively, a parameter for setting the threshold value data in accordance with the information indicating direct sound stationarity may be stored beforehand in memory 1404. In this case, threshold value adjuster 1304 may make a determination on the stationarity of the audio signal and set the threshold value data to be used in the
10 selection of reflected sounds, based on the information indicating stationarity and the parameter.

[0315]

 Alternatively, a plurality of parameters for threshold value data may be stored beforehand in memory 1404, corresponding to a
15 plurality of patterns of direct sound stationarity. In this case, threshold value adjuster 1304 may make a determination on the stationarity of the audio signal, select the threshold value data parameter based on the pattern of direct sound stationarity, and set the threshold value data to be used in the selection of reflected
20 sounds, based on the threshold value data parameter.

[0316]

 Note that a determination on the stationarity of an audio signal may be made based on the amount of variation of the frequency component of the audio signal, each time an audio signal is inputted.

25 [0317]

 Alternatively, a determination on the stationarity of an audio signal may be made based on information indicating stationarity that is pre-associated with the audio signal. In other words, the information indicating audio signal stationarity may be associated
30 with the audio signal and pre-stored in memory 1404. Analyzer 1301 may, each time an audio signal is inputted, obtain information indicating stationarity that is associated with the audio signal. Threshold value adjuster 1304 may then adjust the threshold values based on the information indicating stationarity that is associated
35 with the audio signal.

[0318]

As another example of threshold values being set in accordance with a characteristic of the audio signal, when an audio signal indicates short sounds (clicking sounds, etc.), the application scope of the echo detection limit may be set shorter than when an audio signal indicates long sounds. This processing is based on the characteristics of the precedence effect.

[0319]

It is known that due to the precedence effect, two short sounds that arrive at the listener's ears in succession are heard as one sound if the time period interval between the two short sounds is sufficiently short. The upper limit of this time period interval is dependent on the length of the sounds. For example, the upper limit of this time period interval is about 5 ms for clicking sounds, but for complex sounds such as a human voice or music, the upper limit may be 40 ms (NPL 1).

[0320]

In accordance with the characteristics of this precedence effect, for example, in the case of a sound for which the duration of a direct sound is short, threshold values for short time period lengths are set. Furthermore, threshold values for shorter time period lengths are set as the duration of the direct sound is shorter.

[0321]

Threshold values for short time period lengths being set means that within a range in which the time difference (T) between a direct sound and a reflected sound is small, threshold values corresponding to an echo detection limit based on the characteristics of the precedence effect are set. Threshold values corresponding to the echo detection limit based on the characteristics of the precedence effect are not set outside of this range. In other words, outside of this range, threshold values are low. Thus, threshold values for short time period lengths being set for short sounds can correspond to low threshold values being set for short sounds.

[0322]

As another example of threshold values being set in accordance

with a characteristic of a direct sound, when a direct sound is an intermittent sound (such as speech), threshold values may be set lower than when a direct sound is a continuous sound (such as music).
[0323]

5 For example, when a direct sound corresponds to speech, sound portions and silent portions repeat, and in the silent portions, only the post-masking effect occurs as the masking effect. On the other hand, when the direct sound is a continuous sound such as musical content, the masking effects that occur include both the post-masking
10 effect and a simultaneous masking effect that results from sound occurring at that time. Consequently, the overall masking effect is greater in the case of music, etc. than in the case of speech, etc.
[0324]

 In accordance with masking effect characteristics such as those
15 described above, threshold values may be set higher in the case of music, etc. than in the case of speech, etc. Conversely, threshold values may be set lower in the case of speech, etc. than in the case of music, etc. That is, threshold values may be set to be low when a direct sound has numerous intermittent portions.

20 [0325]

 As described above, the information indicating a characteristic of a direct sound may be information indicating the stationarity, intermittency, duration, etc. of the direct sound. Furthermore, the information indicating a characteristic of a direct sound may be any
25 combination of these. Furthermore, the information indicating a characteristic of a direct sound may be information indicating the time variation of one of these, or may be information indicating the time variation of any combination of these. That is, the information indicating a characteristic of a direct sound may be information
30 indicating the time variation of the direct sound.

[0326]

 For example, as indicated in the description of the stationarity determination, the information indicating a characteristic of a direct sound may be chronological data on the frequency characteristics.
35 Here, the frequency characteristics may be expressed in a commonly

used form such as a gain value per frequency band, a Fourier series with respect to a time axis signal, a linear predictive coding (LPC) coefficient or cepstral coefficient for determining a frequency envelope, or the like.

5 [0327]

Furthermore, the information indicating a characteristic of a direct sound may be, as information indicating the intermittency of direct sound, information (an amplitude envelope outline) enumerating, in chronological order, a plurality of sets of: the duration
10 for which the amplitude of a signal is stationary; and the amplitude value of the signal for that duration. Here, the amplitude value may be expressed as a ratio with respect to the reference sound volume.

[0328]

Furthermore, the information indicating a characteristic of a
15 direct sound may be information on the frequency characteristics of the direct sound. For example, the information indicating a characteristic of a direct sound may be information indicating the stationarity of the frequency characteristics of the direct sound. Specifically the information indicating a characteristic of a direct
20 sound may be information (a spectrogram outline) enumerating, in chronological order, a plurality of sets of: a duration for which the variation in frequency characteristics is small; and the frequency characteristics of the signal for the duration. Here, the sound volume used as the reference for the frequency characteristics may be the
25 reference sound volume.

[0329]

For example, the information indicating the time variation of a direct sound may be information indicating a direct sound envelope. The information indicating the time variation of a direct sound may
30 be used when the "minimum audible limit" described in [Example 4] of FIG. 12C is the threshold value. The signal compared to the minimum audible limit is the sound volume of a reflected sound.

[0330]

The sound volume of a reflected sound is obtained by
35 geometrical calculation from information on the positions of the sound

source, the listener, and the reflection object. Specifically, the reference sound volume of the reflected sound is obtained with respect to the reference sound volume of the sound source. By adjusting the reference sound volume of the reflected sound using, as the information on a characteristic of a direct sound, information on the sound magnitude transition of the sound source, the sound volume of the reflected sound at each moment can be accurately determined. This is because the variation in the sound volume of the sound source is reflected in the variation in the sound volume of the reflected sound.

[0331]

By comparing the sound volume of a reflected sound with the threshold value after adjusting the sound volume of the reflected sound, the reflected sounds that are auditorily necessary can be appropriately selected with more accuracy.

[0332]

It goes without saying that naturally, the same result can be obtained by, without adjusting the reference sound volume of a reflected sound, adjusting the threshold value based on the inverse of the information on the sound magnitude transition of the sound source, and comparing the adjusted threshold value to the reference sound volume of the reflected sound. That is, the reference sound volume of a reflected sound may be adjusted using the information on the sound magnitude transition of the sound source, or the threshold value may be adjusted using the information on the sound magnitude transition of the sound source. The adjustment of the reference sound volume of a reflected sound and the adjustment of a threshold value correspond to each other.

[0333]

Depending on the composition of the surface of an object that reflects sound, the reflectance of sound (the attenuation rate accompanying the reflection) is different for each frequency band. Accordingly, as described later, the reflectance (attenuation rate) of sound may be associated with each frequency band, for an object that reflects sound. Whether to select the reflected sound can be more

accurately determined based on such reflectance information and spectrogram information. For example, processing such as the following is performed.

[0334]

5 Specifically, for example, it is indicated by spectrogram information that a high-frequency component is more dominant than a low-frequency component at a certain time interval. Furthermore, for example, it is indicated by sound reflectance information that the reflectance is very low at the high-frequency component, compared
10 to the low-frequency component.

[0335]

 In this case, there is a possibility that even if the amplitude of the sound source signal on the time axis is high, the sound volume of the reflected sound will be low, the sound volume of the reflected
15 sound being obtained by multiplying, by the attenuation rate for each frequency band indicated by the reflectance information, the frequency component indicated by the spectrogram information. Thus, the reflected sound will not be selected.

[0336]

20 As described above, the information indicating a characteristic of a direct sound may be information indicating the time variation of the direct sound. For example, the information indicating a characteristic of a direct sound may indicate a value obtained by analyzing the direct sound at a predetermined time length.

25 [0337]

 Specifically, the information indicating a characteristic of a direct sound may be information obtained by calculating the average energy or the average amplitude of the direct sound for each predetermined time length. Furthermore, the information indicating
30 a characteristic of a direct sound may be information obtained by calculating the energy or average amplitude of the direct sound for each short time analysis length, and calculating the weighted average of the energy or the average amplitude for each long time analysis length that is longer than the short time analysis length.

35 [0338]

More specifically, for example, the information indicating the time variation of a direct sound may be information obtained by calculating the energy or the average amplitude of the direct sound for each predetermined short time length (for example, 5 ms; hereinafter, time length frames are expressed as analysis frames). Furthermore, the information indicating the time variation of a direct sound may be information represented by a weighted average of the energy or average amplitude calculated using the past N-1 analysis frames.

10 [0339]

When the energy of the n-th analysis frame is expressed by E(n), information I(n) indicating a characteristic of the direct sound is determined in accordance with the following formula.

[0340]

15 [Math. 2]

$$I(n) = a(0) \cdot E(n) + \sum_{i=1}^{N-1} a(i) \cdot E(n-i)$$

• • • (Formula 1)

[0341]

Here, parameter a(i) denotes a weighting factor. Typically, a(i) is set such that a(i) ≥ 0 and the sum of a(i) is 1. However, the method for setting a(i) is not limited thereto.

[0342]

Note that information I(n) indicating a characteristic of direct sound is calculated each time 5 ms of direct sound is imported. That is to say, the time variation of information I(n) indicating a characteristic of a direct sound can be calculated with low delay. Thus, this method is suitably applied to an application requiring real-time performance.

[0343]

Furthermore, information I(n) indicating a characteristic of a direct sound may be determined in accordance with the following formula.

[0344]

[Math. 3]

$$I(n) = b(0) \cdot E(n) + \sum_{i=1}^{N-1} b(i) \cdot I(n-i)$$

• • • (Formula 2)

[0345]

Here, parameter $b(i)$ denotes a weighting factor. Typically,
5 $b(i)$ is set such that $b(i) \geq 0$ and the sum of $b(i)$ is 1. However, the method for setting $b(i)$ is not limited thereto.

[0346]

In this formula, information $I(n)$ indicating a characteristic of a direct sound is recursively determined. Thus, the average energy for
10 a long time length can be calculated with a small amount of computation.

[0347]

Formula 1 and Formula 2 above can be considered to be filters in which $E(n)$ is the input signal and $I(n)$ is the output signal. In this
15 case, Formula 1 is a moving average (MA) model filter and Formula 2 is an autoregressive (AR) model filter, and both of these have the characteristics of a low-pass filter. Furthermore, an ARMA model filter, in which both of these are combined, may be used.

[0348]

Note that the method for deriving the information indicating the time variation of a direct sound is not limited to the above-described formulae or filters, and other well-known methods may be used. As described above, the information indicating the time
20 variation of a direct sound indicates a value obtained by analyzing a direct sound at a predetermined time length. A direct sound may be analyzed from perspectives other than average energy.

[0349]

Furthermore, as described above, the information indicating a characteristic of a direct sound may be information on the frequency
30 characteristics of the direct sound. The information on the frequency characteristics of a direct sound may be information calculated by using the frequency characteristics of the direct sound. For example,

the information on the frequency characteristics of a direct sound may be information obtained as the average energy of a low-frequency component, by averaging the low-frequency component of the direct sound at a predetermined analysis length.

5 [0350]

Specifically, the low-frequency component of a direct sound can be determined by applying, to a direct sound contained in an analysis frame length, a filter having low-pass characteristics. From the energy or average amplitude of this low-frequency component, the information indicating a characteristic of a direct sound can be derived, similarly to Formula 1 described above.

[0351]

When the energy of the low-frequency component of the n-th analysis frame is expressed by $E_L(n)$, information $I(n)$ indicating a characteristic of the direct sound can be determined in accordance with the following formula.

[0352]

[Math. 4]

$$I(n) = c(0) \cdot E_L(n) + \sum_{i=1}^{N-1} c(i) \cdot E_L(n-i)$$

• • • (Formula 3)

20 [0353]

Here, parameter $c(i)$ denotes a weighting factor. Typically, $c(i)$ is set such that $c(i) \geq 0$ and the sum of $c(i)$ is 1. However, the method for setting $c(i)$ is not limited thereto.

[0354]

25 Note that information $I(n)$ indicating a characteristic of direct sound is calculated each time 5 ms of direct sound is imported. That is to say, the time variation of information $I(n)$ indicating a characteristic of a direct sound can be calculated with low delay. Thus, this method is suitably applied to an application requiring real-time performance.

30

[0355]

Furthermore, similarly to Formula 2, information $I(n)$ indicating

a characteristic of a direct sound can be determined in accordance with the following formula.

[0356]

[Math. 5]

$$I(n) = d(0) \cdot E_L(n) + \sum_{i=1}^{N-1} d(i) \cdot I(n-i)$$

5 · · · (Formula 4)

[0357]

Here, parameter $d(i)$ denotes a weighting factor. Typically, $d(i)$ is set such that $d(i) \geq 0$ and the sum of $d(i)$ is 1. However, the method for setting $d(i)$ is not limited thereto.

10 [0358]

In this formula, information $I(n)$ indicating a characteristic of a direct sound is recursively determined. Thus, the average energy for a long time length can be calculated with a small amount of computation.

15 [0359]

Formula 3 and Formula 4 above can be considered to be filters in which $E(n)$ is the input signal and $I(n)$ is the output signal. In this case, Formula 3 is a moving average (MA) model filter and Formula 4 is an autoregressive (AR) model filter, and both of these have the characteristics of a low-pass filter. Furthermore, an ARMA model filter, in which both of these are combined, may be used.

[0360]

In the above description, a filter having low-pass characteristics was used in the method for determining the low-frequency component of a direct sound, but the method for determining the low-frequency component of a direct sound is not limited thereto. Furthermore, the method for deriving the information indicating the time variation of direct sound is not limited to the above-described formulae or filters, and other well-known methods may be used. Furthermore, the spectrum of a direct sound can be calculated by applying frequency conversion to the direct sound. The energy or average amplitude of the low-frequency

component of the spectrum can then be calculated.

[0361]

Furthermore, in the above description, the MA model or the AR model is used for deriving the information indicating the time variation of a direct sound. The coefficients of these models may be
5 predetermined fixed values, or may be variable values, which are values that temporally vary.

[0362]

Furthermore, the relationship between the analysis frame
10 length and the interval at which the information update threads are created may be as described below.

[0363]

For example, when the time length of an analysis frame is TA (msec) and the interval at which information update threads are
15 created is TU (msec), the value of N in (Formula 1) and (Formula 3), described above, in the MA filter may be approximately a value given by TU/TA. Furthermore, $b(i)$ and $d(i)$ ($1 \leq i < N$) in (Formula 2) and (Formula 4), described above, in the AR filter may be a value that results in the time constant of the filter being about TU (msec).

20 [0364]

The reason for this setting is that within the information update interval period, convergence of the filter is expected.

[0365]

On the other hand, when, in the information indicating the time
25 variation of a direct sound with the above-described setting, the value varies too sharply, $I(n)$ may be precalculated. Furthermore, precalculated $I(n)$ may be applied to the selection processing of reflected sounds. For example, in the processing of the t-th time frame, $I(t + \tau)$ may be used. Here, τ is a value defined in
30 accordance with the characteristics of filter convergence. When the convergence is slow, the value of τ is large in comparison to when the convergence is fast.

[0366]

Furthermore, as the information indicating the characteristics
35 of a direct sound, information on auditory masking (frequency

masking) calculated from the direct sound may be used. The auditory masking information indicates a threshold value of an amplitude value at a frequency region masked by a direct sound. Processing in which the amplitude values of reflected sounds in the same frequency range are compared to the threshold value, and a reflected sound having an amplitude value lower than the threshold value is not selected may be performed. The amplitude value of a reflected sound in a frequency range may be obtained by analyzer 1301 as information indicating the characteristics of a reflected sound.

10 [0367]

When threshold values to be used in selecting reflected sounds are thus set in accordance with the characteristics of direct sound, it is possible to appropriately select reflected sounds that are auditorily necessary, and auditory characteristics can be effectively reflected in three-dimensional sound reproduction system 1000. Processing to detect characteristics of direct sound, processing to determine threshold values in accordance with the characteristics, and processing to adjust the threshold values in accordance with the characteristics may be performed during the rendering processing, or may be performed before starting the rendering processing.

[0368]

For example, these processes may be performed, for example, during virtual space creation (during software creation), when starting processing of the virtual space (when launching the software or starting rendering), or when there is an occurrence of an information update thread that periodically occurs in processing of the virtual space. Furthermore, the time of virtual space creation may be when the virtual space is built before starting acoustic processing, may be when information (spatial information) on the virtual space is obtained, or may be when software is obtained.

30 [0369]

Here, in the information update thread, processing to update the spatial information managed by spatial information managers 1201 and 1211 is performed.

35 [0370]

The role of the information update thread is, for example, processing to update, based on the position and orientation of the VR goggles worn by the listener, the position and orientation of the listener's avatar positioned in the virtual space, updating of the positions of objects that have moved within the virtual space, and the like. Such processing is covered within a processing thread that launches relatively infrequently, at approximately tens of Hz.

[0371]

Processing to update the information indicating a characteristic of a direct sound may be performed by such a processing thread that is created infrequently. The reason for this is that characteristics of direct sound vary less frequently than audio processing frames for audio output occur. This makes it possible to relatively reduce the computational load of the processing. Furthermore, when the information is updated at an unduly high frequency, there is a risk of pulsive noise being generated. Updating the information at a low frequency also makes it possible to avoid such a risk.

[0372]

[Third Variation of Threshold Value Setting Method]

As another example of a threshold value setting method, threshold values may be set in accordance with computation resources (CPU capability, memory resources, PC performance, remaining level of battery, etc.) for processing reproduction of the virtual space. More specifically, sensor 1405 of audio signal processing device 1001 detects the amount of computation resources, and when the amount of computation resources is low, the threshold values are set to be high. Since consequently, the sound volume of a greater number of reflected sounds falls below the threshold values, the number of reflected sounds on which binaural processing is to be performed can be reduced, whereby the amount of computation can be reduced.

[0373]

Alternatively, when the signal processing is performed by equipment that is driven by a storage battery, such as a smartphone or VR goggles, it is expected that priority is given to allowing

processing to be performed for a longer duration, and computation resources are used economically. In such a case, it is not necessary to detect the amount or remaining level of computation resources, and the threshold values may be set to be high.

5 [0374]

[Fourth Variation of Threshold Value Setting Method]

As another example of a threshold value setting method, by including a threshold value setter, not illustrated, in audio signal processing device 1001 or audio presentation device 1002, threshold
10 values can be set by the manager of the virtual space or the listener.
[0375]

For example, an “energy-saving mode”, in which there are few reflected sounds to be heard and the amount of computation is low, or a “high-performance mode”, in which there are many reflected
15 sounds to be heard and the amount of computation is high, may be selectable by the listener to whom audio presentation device 1002 is equipped. Alternatively, the mode may be selectable by the manager who manages three-dimensional sound reproduction system 1000 or by the creator of the three-dimensional sound content. Furthermore,
20 not the mode, but the threshold values or the threshold value data may be directly selectable.

[0376]

[First Variation of Operations of Renderer]

FIG. 20 is a flowchart illustrating a first variation of operations
25 of audio signal processing device 1001. FIG. 20 illustrates the processing performed mainly by renderer 1300 of audio signal processing device 1001. In this variation, sound volume compensation processing is added to the operations of renderer 1300.
[0377]

30 For example, analyzer 1301 obtains data (the input signal) (S301). Next, analyzer 1301 analyzes the data (S302). Next, determiner 1302 determines whether to select reflected sounds based on the analysis results (S303). Next, reproducer 1303 performs sound volume compensation processing based on the reflected sounds
35 that were not selected (S304). Next, reproducer 1303 performs

acoustic processing on the direct sounds and the reflected sounds (S305). Reproducer 1303 then outputs the direct sounds and the reflected sounds as audio (S306).

[0378]

5 The above-described processes (S301 to S306) other than the sound volume compensation processing (S304) are processes that are shared with the other examples described above; thus, explanation thereof has been omitted.

[0379]

10 The sound volume compensation processing is performed in accordance with the reflected sounds that were not selected in the selection processing. For example, due to not selecting a reflected sound in the selection processing, an absence emerges in the sound volume sensation. The sound volume compensation processing
15 reduces the incongruity that accompanies this absence in the sound volume sensation. As an example of compensating the sound volume sensation, the following two methods are disclosed. Either of these two methods may be used.

[0380]

20 First, a method in which the sound volume sensation is compensated for by raising the sound volume of a direct sound will be described. Reproducer 1303 raises the sound volume of the direct sound by the amount of the sound volume of a reflected sound that was not selected, and generates the direct sound. Accordingly, the
25 sound volume sensation lost due to the reflected sound not being generated is compensated for.

[0381]

 At the time of raising the sound volume, reproducer 1303 may raise the sound volume of each frequency component in accordance
30 with the frequency characteristics of the reflected sound. In order to make such processing possible, an attenuation rate of the sound volume attenuated by the reflection object may be assigned to each of predetermined frequency bands. This makes it possible to derive the frequency characteristics of the reflected sound.

35 [0382]

Next, a method in which the sound volume sensation is compensated for by causing a reflected sound to be synthesized in a direct sound will be described. In this method, reproducer 1303 adds, to a direct sound, a reflected sound that was not selected and generates the direct sound to compensate for the sound volume sensation that results from the reflected sound not being generated. The sound volume (amplitude), frequency, delay, and the like of the reflected sound that was not selected are reflected in the generated direct sound.

10 [0383]

In the case of the method for raising the sound volume of the direct sound, while the amount of computation for the compensation processing is extremely slight, only the sound volume is compensated for. In the case of the method of causing a reflected sound to be synthesized in a direct sound, the amount of computation for the compensation processing is large compared to the method of raising the sound volume of the direct sound, but the characteristics of the reflected sound are more accurately compensated for.

[0384]

20 Since in both cases, only the direct sound is generated and the reflected sound is not generated, the total amount of computation is reduced. In particular, since the amount of computation required for binaural processing, which includes processing to implement a head-related transfer function (HRTF), is reduced, the total amount of computation is greatly reduced. The reason for this is that the amount of computation required for binaural processing is much greater than the amount of processing required for the above-described compensation processing.

[0385]

30 Note that when the reason for a reflected sound not being selected is that the sound volume of the reflected sound is less than the masking threshold value, the sound volume sensation is not lost; thus, the reflected sound may be simply removed without performing compensation processing.

35 [0386]

[Second Variation of Operations of Renderer]

FIG. 21 is a flowchart illustrating a second variation of operations of audio signal processing device 1001. FIG. 21 illustrates the processing performed mainly by renderer 1300 of audio signal processing device 1001. In this variation, left-right sound volume difference adjustment processing is added to the operations of renderer 1300.

[0387]

For example, analyzer 1301 analyzes the input signal (S401).
10 Next, analyzer 1301 detects the direction of arrival of sounds (S402).
Next, determiner 1302 adjusts the difference in sound volume between the sounds perceived by the left and right ears (S403).
Furthermore, determiner 1302 adjusts the difference in the arrival time periods (delay) between the sounds perceived by the left and
15 right ears (S404). Determiner 1302 determines whether to select reflected sounds based on information on the adjusted sounds (S405).

[0388]

The above-described processes (S401 to S405) other than the left-right sound volume difference adjustment processing (S403) and
20 the delay adjustment processing (S404) are processes that are shared with the other examples described above; thus, explanation thereof has been omitted.

[0389]

FIG. 22 is a diagram illustrating an arrangement example of an
25 avatar, a sound source object, and an obstacle object. For example, in a case in which the front direction of the listener is 0 degrees, when, as in FIG. 22, the polarities (for example, positive-negative) of the direction of arrival (θ) of the direct sound and the direction of arrival (γ) of the reflected sound (the direction (γ) of the reflected sound)
30 are different, the sound volume difference that occurs between the ears is corrected.

[0390]

Specifically, when the polarities of θ and γ are different, the ear which mainly (first) perceives the sound is different for each of
35 the direct sound and the reflected sound. In this case, as the left-

right sound volume difference adjustment processing (S403), determiner 1302 adjusts the sound volume of the direct sound in accordance with the position of the ear that mainly perceives the reflected sound. For example, by multiplying the sound volume when the direct sound arrives at the listener by $(1.0 - 0.3 \sin(\theta))$ ($0 \leq \theta \leq 180$), determiner 1302 causes attenuation of the sound volume when the direct sound arrives at the listener.

[0391]

By calculating the sound volume ratio of the sound volume of the reflected sound to the sound volume of the direct sound, corrected as described above, and comparing the calculated sound volume ratio with threshold values, determiner 1302 determines whether to select reflected sounds. Accordingly, the sound volume difference that occurs between the ears is corrected, the sound volume of direct sounds that affect reflected sounds is more accurately derived, and the determination of whether to select reflected sounds is more accurately performed.

[0392]

Furthermore, in addition to the left-right sound volume difference adjustment processing (S403), determiner 1302 may, as delay adjustment processing (S404), delay the direct sound arrival time period in accordance with the positions of the ears at which a reflected sound is perceived. Specifically, determiner 1302 may delay the direct sound arrival time period by adding, to the direct sound arrival time period, $(a (\sin\theta + \theta)/c)$ ms (where a is the radius of the head and c is the speed of sound).

[0393]

[Third Variation of Operations of Renderer]

A method for setting threshold values in accordance with directions of arrival will be described.

[0394]

FIG. 23 is a flowchart illustrating yet another example of the selection processing. Description has been omitted for processes that are shared with the example in FIG. 14. In the example in FIG. 23, determiner 1302 selects reflected sounds by using threshold

values in accordance with directions of arrival.

[0395]

Specifically, from the direct sound arrival path (pd), the reflected sound arrival path (pr), and avatar orientation information D1, each calculated by analyzer 1301, determiner 1302 calculates the direct sound arrival direction (θ) and the reflected sound arrival direction (γ) (the direction (γ) of the reflected sound), each defined using the orientation of an avatar as reference. In other words, determiner 1302 detects the direct sound arrival direction (θ) and the reflected sound arrival direction (γ) (S231). The orientation of the avatar corresponds to the orientation of the listener. Avatar orientation information D1 may be included in the input signal.

[0396]

By using three indexes including the time difference (T), in addition to the direct sound arrival direction (θ) and the reflected sound arrival direction (γ), determiner 1302 identifies, from a three-dimensional arrangement such as that illustrated in FIG. 15, the threshold values to be used in the selection processing (S232).

[0397]

As an example, a method for setting threshold values to be used in selection processing when, as in FIG. 22, an avatar, a sound source object, and an obstacle object are arranged will be described.

[0398]

Position information on the avatar, the sound source object, and the obstacle object, and avatar orientation information D1 are obtained from the input information. The direction (θ) of the direct sound and the direction (γ) of the sound image of the reflected sound when the orientation of the avatar is determined to be 0 degrees are calculated by using these items of position information and orientation information D1. In the case of FIG. 22, the direction (θ) of the direct sound is about 20 degrees, and the direction (γ) of the sound image of the reflected sound is about 265 degrees (-95 degrees).

[0399]

Next, referencing the threshold value data stored in the three-

dimensional arrangement illustrated in FIG. 15, a threshold value is identified from an arrangement domain that corresponds to the values of the two directions (θ) and (γ), and the value of the time difference (T) calculated by analyzer 1301. When there is no index that corresponds to the values of (θ), (γ), and (T) that were calculated, the threshold value corresponding to the index that is closest may be identified.

[0400]

As another method, threshold values may be identified by performing processing such as interpolation or extrapolation, based on one or more threshold values that correspond to one or more indexes that are closest to the values of (θ), (γ), and (T) that were calculated. For example, a threshold value corresponding to (20 degrees, 265 degrees, T) may be identified based on the four threshold values corresponding to the four indexes of (0 degrees, 225 degrees, T), (0 degrees, 270 degrees, T), (45 degrees, 225 degrees, T), and (45 degrees, 270 degrees, T).

[0401]

Selection processing based on the difference between the direct sound arrival direction angle (θ) and the reflected sound arrival direction angle (γ) will be described.

[0402]

For example, as illustrated in FIG. 16, threshold value data having, as a two-dimensional index arrangement: the angular difference (Φ) between the direct sound arrival direction (θ) and the reflected sound arrival direction (γ); and the time difference (T) may be pre-created and set. In this case, the angular difference (Φ) and the time difference (T) are referenced in the selection processing. Alternatively, the angular difference (Φ) between the angle (θ) of the direct sound arrival direction and the angle (γ) of the reflected sound arrival direction may be calculated in the selection processing, and the angular difference (Φ) calculated may be used to identify the threshold value.

[0403]

Alternatively, threshold value data having, as an index

arrangement, a combination of the angular difference (Φ), the direct sound arrival direction (θ), and the time difference (T), or a combination of the angular difference (Φ), the reflected sound arrival direction (γ), and the time difference (T) may be set.

5 [0404]

Alternatively, as illustrated in FIG. 15, threshold value data having, as a three-dimensional index arrangement, values of (θ), (γ), and (T) may be set.

[0405]

10 [Fourth Variation of Operations of Renderer]

The processing performed by the above-described analyzer 1301, determiner 1302, and reproducer 1303 may, for example, be performed as pipeline processing as described in PTL 3.

[0406]

15 FIG. 24 is a block diagram illustrating a configuration example for renderer 1300 to perform pipeline processing.

[0407]

Renderer 1300 in FIG. 24 includes reverberation processor 1311, early reflection processor 1312, distance attenuation processor 1313, determiner 1314, generator 1315, and binaural processor 1316. These constituent elements may be configured as a plurality of the constituent elements of renderer 1300 illustrated in FIG. 7, or may be configured as at least a part of the plurality of constituent elements of audio signal processing device 1001 illustrated in FIG. 5.

25 [0408]

Pipeline processing refers to dividing the processing for applying acoustic effects into a plurality of processes and executing each of the plurality of processes one by one in order. The plurality of processes include, for example, signal processing on the audio signal, generation of parameters used for signal processing, and the like.

[0409]

35 Renderer 1300 may perform reverberation processing, early reflection processing, distance attenuation processing, binaural processing, and the like as pipeline processing. However, these

types of processing are examples, and the pipeline processing may include processes other than these, or may not include some of these processes. For example, the pipeline processing may include diffraction processing and occlusion processing. Furthermore, for
5 example, the reverberation processing may be omitted when unneeded.

[0410]

Furthermore, each process may be expressed as a stage. Moreover, the audio signals of the reflected sounds and the like
10 generated as the result of the processes may be expressed as rendering items. The plurality of stages and the order of these stages in the pipeline processing are not limited to the example illustrated in FIG. 24.

[0411]

Here, the parameters (the arrival paths, the arrival time periods, and the sound volume ratios related to direct sounds and reflected sounds) used in the selection processing are calculated in one of the plurality of stages for generating the rendering items. In other words, the parameters used for selecting the reflected sounds
15 are calculated as a part of the pipeline processing for generating the rendering items. Note that it is not necessary for all of the stages to be performed by renderer 1300. For example, a part of the stages may be omitted, or may be performed by an element other than renderer 1300.
20

[0412]

The reverberation processing, the early reflection processing, the distance attenuation processing, the selection processing, the generation processing, and the binaural processing that may be included as stages in the pipeline processing will be described. In
30 each stage, the metadata included in the input signal may be analyzed, and the parameters used for generating the reflected sounds may be calculated.

[0413]

In the reverberation processing, reverberation processor 1311
35 generates an audio signal indicating reverberation sound or the

parameters used in generating the audio signal. Reverberation sound is a sound that arrives at the listener as reverberation after the direct sound. As one example, the reverberation sound is a sound that arrives at the listener at a relatively late stage (for example, approximately 100 to 200 ms after the arrival of the direct sound) after the early reflected sound (to be described later) arrives at the listener, and after undergoing more reflections (for example, several tens of times) than the early reflected sound.

[0414]

Reverberation processor 1311 refers to the audio signal and spatial information included in the input signal, and calculates reverberation sound by using, as a function for generating reverberation sound, a predetermined function prepared beforehand.

[0415]

Reverberation processor 1311 may generate reverberation sound by applying a known reverberation generation method to the audio signal included in the input signal. One example of a known reverberation generation method is the Schroeder method, but the known reverberation generation method is not limited to the Schroeder method. Furthermore, reverberation processor 1311 uses the shape and acoustic characteristics of a sound reproduction space indicated by the spatial information when applying the known reverberation generation method. In this way, reverberation processor 1311 can calculate parameters for generating reverberation sound.

[0416]

In the early reflection processing, early reflection processor 1312 calculates parameters for generating early reflection sounds based on the spatial information. The early reflected sound is reflected sound that arrives at the listener at a relatively early stage (for example, approximately several tens of ms after the arrival of the direct sound) after the direct sound from the sound source object arrives at the listener, and after undergoing one or more reflections.

[0417]

Early reflection processor 1312 references, for example, the

audio signal and metadata, and calculates the path, from reflection objects, of reflected sound that arrives at the listener after being reflected by the reflection objects. For example, in calculation of the path, the shape of the three-dimensional sound field (space), the size of the three-dimensional sound field, the positions of reflection objects such as structures, the reflectance of reflection objects, and the like may be used.

[0418]

Early reflection processor 1312 may calculate the path of the direct sound. The information of said path may be used as a parameter for early reflection processor 1312 to generate the early reflected sound, and may be used as a parameter for determiner 1314 to select reflected sounds.

[0419]

In the distance attenuation processing, distance attenuation processor 1313 calculates the sound volume of the direct sound and the reflected sound that arrive at the listener, based on the lengths of the paths of the direct sound and the reflected sound. The sound volume of the direct sound and the reflected sound that arrive at the listener attenuate, with respect to the sound volume of the sound source, in proportion to the distance of the path to the listener (in inverse proportion to the distance). Thus, distance attenuation processor 1313 is able to calculate the sound volume of the direct sound by dividing the sound volume of the sound source by the length of the direct sound path, and is able to calculate the sound volume of the reflected sound by dividing the sound volume of the sound source by the length of the path of the reflected sound.

[0420]

In the selection processing, determiner 1314 selects the reflected sounds to be generated, based on the parameters calculated before the selection processing. One of the selection methods of the present disclosure may be used for selection of the reflected sounds to be generated.

[0421]

The selection processing may be performed on all of the

reflected sounds, or may be performed only on the reflected sounds having high evaluation values based on the evaluation processing, as described above. In other words, the reflected sounds having low evaluation values may be determined as not selected, without performing the selection processing. For example, reflected sounds for which the sound volume is extremely low may be considered to be reflected sounds having low evaluation values, and may be determined as not selected.

[0422]

Furthermore, for example, the selection processing may be performed on all of the reflected sounds. Then, the evaluation values of the reflected sounds selected in the selection processing may be determined, and the reflected sounds having low determined evaluation values may be redetermined as not selected.

[0423]

The selection processing and the evaluation processing may each be performed independently, or may be performed in combination with each other. When the selection processing and the evaluation processing are performed in combination with each other, either of the two processes may be performed first.

[0424]

In the generation processing, generator 1315 generates direct sounds and reflected sounds. For example, generator 1315 generates direct sounds based on the direct sound arrival times and arrival time sound volume, from the audio signal included in the input signal. Furthermore, for each reflected sound selected in the selection processing, generator 1315 generates the reflected sound based on the reflected sound arrival time and the arrival time sound volume, from the audio signal included in the input signal.

[0425]

In the binaural processing, binaural processor 1316 performs signal processing so that the audio signal of the direct sound is perceived as sound arriving at the listener from the direction of the sound source object. Furthermore, binaural processor 1316 performs signal processing so that the reflected sounds selected by

determiner 1314 are perceived as sounds arriving at the listener from the reflection object.

[0426]

For example, based on the position and orientation of the listener in the sound space, binaural processor 1316 performs processing to apply an HRIR DB so that sound arrives at the listener from the position of the sound source object or the position of the obstacle object.

[0427]

Note that Head-Related Impulse Response (HRIR) is the response characteristic when one impulse is generated. Specifically, HRIR is the response characteristic obtained by converting from an expression in the frequency domain to an expression in the time domain by Fourier transforming the head-related transfer function, in which the change in sound caused by surrounding objects including the auricle, the head, and the shoulders is expressed as a transfer function. The HRIR DB is a database including such information.

[0428]

Furthermore, the position and orientation of the listener in the sound space are, for example, the position and orientation of a virtual listener in a virtual sound space. The position and orientation of the virtual listener in the virtual sound space may change in accordance with movement of the head of the listener. Furthermore, the position and orientation of the virtual listener in the virtual sound space may be determined based on information obtained from sensor 1405.

[0429]

The program(s), spatial information, HRIR DB, threshold value data, other parameters, and/or the like used in the above-described processing are obtained from memory 1404 included in audio signal processing device 1001, or from outside of audio signal processing device 1001.

[0430]

Furthermore, the pipeline processing may contain other processes. Moreover, renderer 1300 may contain a processor that is not illustrated, for performing another process included in the pipeline

processing. For example, renderer 1300 may include a diffraction processor and an occlusion processor.

[0431]

The diffraction processor executes processing to generate an audio signal indicating sound including diffracted sound caused by an obstacle object between the listener and the sound source object in a three-dimensional sound field (space). Diffracted sound is sound that, when an obstacle object is present between the sound source object and the listener, arrives at the listener from the sound source object by going around the obstacle object.

[0432]

The diffraction processor references, for example, the audio signal and metadata, and calculates the path by which diffracted sound arrives at the listener from the sound source object by detouring around the obstacle object, and generates diffracted sound based on the calculated path. In the calculation of the path, the sound source object in the three-dimensional sound field (space), the positions of the listener and the obstacle object, the shape and size of the obstacle object, and the like may be used.

[0433]

When a sound source object is present on the other side of an obstacle object, the occlusion processor generates an audio signal for a sound that passes from the sound source object through the obstacle object and is audible therethrough, based on the spatial information and information on the material, etc. of the obstacle object.

[0434]

[Example of Sound Source Object]

As described above, in the position information assigned to the sound source object, a "point" in the virtual space indicates the position of a sound source object. In other words, as described above, the sound source is defined as a "point sound source".

[0435]

On the other hand, a sound source in a virtual space may be defined as an object that has a length, size, shape, and the like, i.e.,

as a sound source that is not a point sound source, but a spatially extended sound source. In this case, the distance between the listener and the sound source, and the direction of arrival of the sound are not determined. Consequently, reflected sounds originating from such a sound source may be limited to being selected by determiner 1302 without performing analysis by analyzer 1301, or regardless of the analysis result. By doing so, it is possible to avoid the sound quality degradation that might occur by not selecting the reflected sound.

10 [0436]

Alternatively, a representative point such as the center of gravity of the object may be determined, and the processing of the present disclosure may be applied on the assumption that sound is generated from that representative point. In this case, the threshold value may be adjusted in accordance with information on the spatial extension of the sound source.

[0437]

[Examples of Direct Sound and Reflected Sound]

For example, direct sound is sound that has not been reflected by a reflection object, and reflected sound is sound that has been reflected by a reflection object. Direct sound may be sound that has arrived at the listener from a sound source without being reflected by a reflection objection, and reflected sound may be sound that has arrived at the listener from a sound source due to being reflected by a reflection object.

[0438]

Furthermore, each of direct sound and reflected sound are not limited to being sound that has arrived at the listener, and may each be sound that will arrive at the listener. For example, direct sound may be sound that has been outputted from a sound source, or to put it differently, a sound source sound.

[0439]

FIG. 25 is a diagram illustrating transmission and diffraction of a sound. As illustrated in FIG. 25, a direct sound may not arrive at the listener due to the presence of an obstacle object between the

sound source object and the listener. In this case, a sound that arrives at the listener after being emitted from the sound source object and passing through the obstacle object may be considered to be a direct sound. Furthermore, a sound that arrives at the listener after being emitted from the sound source object and diffracted by the obstacle object may be considered to be a reflected sound.

[0440]

Furthermore, the two sounds compared in the selection processing are not limited to a direct sound and a reflected sound based on sound emitted from one sound source. For example, the selection of a sound may be performed by performing a comparison between two reflected sounds based on a sound emitted from one sound source. In this case, the direct sound in the present disclosure may be understood to be the sound that reaches the listener first, and the reflected sound in the present disclosure may be understood to be the sound that reaches the listener afterward.

[0441]

[Example of Bitstream Structure]

The bitstream includes, for example, an audio signal and metadata. The audio signal is sound data in which sound is expressed, and indicates, e.g., information on the frequency and intensity of sound. Furthermore, metadata includes spatial information on the sound space, which is the space of the sound field.

[0442]

For example, the spatial information is information on the space in which the listener who hears sound based on the audio signal is positioned. Specifically, the spatial information is information about a predetermined position (localization position) in the sound space (for example, a three-dimensional sound field) for localizing the sound image of the sound at that predetermined position, that is, for causing the listener to perceive the sound as arriving from a direction that corresponds to the predetermined position. The spatial information includes, for example, sound source object information and position information indicating the position of the listener.

[0443]

The sound source object information is information on a sound source object that generates sound based on the audio signal. In other words, the sound source object information is information on an object (a sound source object) that reproduces the audio signal, and is information on a virtual sound source object located in a virtual sound space. Here, the virtual sound space may correspond to real-world space in which an object that generates sound is located, and the sound source object in the virtual sound space may correspond to an object that generates sound in a real-world space.

10 [0444]

The sound source object information may indicate, for example, the position of the sound source object located in the sound space, the orientation of the sound source object, the directivity of the sound emitted by the sound source object, whether the sound source object belongs to an animate thing, whether the sound source object is a mobile body, and the like. For example, the audio signal is associated with one or more sound source objects indicated by the sound source object information.

[0445]

20 The bitstream includes, for example, metadata (control information) and an audio signal.

[0446]

The audio signal and metadata may be contained in a single bitstream or may be separately contained in a plurality of bitstreams. Furthermore, the audio signal and metadata may be contained in a single file or may be separately contained in a plurality of files.

[0447]

The bitstream may exist for each sound source or may exist for each playback time. Even in a case in which bitstreams exist for each playback time, a plurality of bitstreams may be processed in parallel simultaneously.

[0448]

Metadata may be assigned to each bitstream, or may be collectively assigned to a plurality of bitstreams as information for controlling the plurality of bitstreams. In this case, the plurality of

35

bitstreams may share the metadata. Furthermore, the metadata may be assigned for each playback time.

[0449]

When a plurality of bitstreams or a plurality of files exist,
5 information indicating a relevant bitstream or a relevant file may be contained in one or more bitstreams or one or more files. Alternatively, information indicating a relevant bitstream or a relevant file may be contained in each of all of the bitstreams or each of all of the files.

10 [0450]

Here, the relevant bitstream or the relevant file is, for example, a bitstream or file that may be used simultaneously during acoustic processing. Furthermore, a bitstream or file that collectively describes the information indicating the relevant bitstream or the
15 relevant file may be included.

[0451]

Here, the information indicating the relevant bitstream or the relevant file may be, for example, an identifier indicating a relevant bitstream or a relevant file. Furthermore, the information indicating
20 the relevant bitstream or the relevant file may be, for example, a file name indicating a relevant bitstream or a relevant file, a uniform resource locator (URL), a uniform resource identifier (URI), or the like.

[0452]

In this case, an obtainer identifies and obtains a relevant
25 bitstream or a relevant file based on the information indicating the relevant bitstream or the relevant file. Furthermore, the information indicating the relevant bitstream or the relevant file may be included in a bitstream or a file, and the information indicating the relevant bitstream or the relevant file may be included in a different bitstream
30 or a different file.

[0453]

Here, the file including the information indicating the relevant bitstream or the relevant file may be, for example, a control file such as a manifest file used in content distribution.

35 [0454]

Note that the entire metadata or part of the metadata may be obtained from somewhere other than a bitstream of the audio signal. For example, either one of metadata for controlling an acoustic sound or metadata for controlling a video may be obtained from somewhere
5 other than from a bitstream, or both may be obtained from somewhere other than from a bitstream.

[0455]

Furthermore, the metadata for controlling a video may be included in the bitstream obtained by three-dimensional sound
10 reproduction system 1000. In this case, three-dimensional sound reproduction system 1000 may output the metadata for controlling a video to a display device that displays images or a stereoscopic video reproduction device that reproduces stereoscopic videos.

[0456]

15 [Examples of Information Included in Metadata]

The metadata may be information used for describing a scene expressed in the sound space. As used herein, the term "scene" refers to a collection of all elements that represent three-dimensional video and acoustic events in the sound space, which are modeled in
20 three-dimensional sound reproduction system 1000 using metadata.

[0457]

Thus, the metadata may include not only information for controlling acoustic processing, but also information for controlling video processing. The metadata may include only one among the
25 information for controlling acoustic processing or the information for controlling video processing, or may include both.

[0458]

Three-dimensional sound reproduction system 1000 generates virtual acoustic effects by performing acoustic processing on the audio
30 signal using the metadata included in the bitstream and additionally obtained interactive listener position information. Early reflection processing, obstacle processing, diffraction processing, occlusion processing, and reverberation processing may be performed as acoustic effects, and other acoustic processing may be performed
35 using the metadata. For example, an acoustic effect such as a

distance decay effect, localization, or a Doppler effect may be added.
[0459]

In addition, information for switching between on and off of all or one or more of the acoustic effects, and priority information
5 regarding a plurality of processes for the acoustic effects may be added to the metadata.
[0460]

As an example, the metadata includes information about a sound space including a sound source object and an obstacle object
10 and information about a localization position for localizing the sound image at a predetermined position in the sound space (that is, causing the listener to perceive the sound as arriving from a predetermined direction).
[0461]

15 Here, an obstacle object is an object that can influence a sound emitted by a sound source object and perceived by the listener, by, for example, blocking or reflecting the sound between the sound source object and the listener. The obstacle object can include an animal or a movable body such as a machine, in addition to a
20 stationary object. The animal may be a person or the like.
[0462]

Furthermore, when a plurality of sound source objects are present in a sound space, another sound source object may be an obstacle object for a certain sound source object. In other words,
25 non-sound-emitting objects such as building materials or inanimate objects, and sound source objects that emit sound can both be obstacle objects.
[0463]

The metadata includes information indicating all or part of the
30 shape of the sound space, the shapes and positions of obstacle objects in the sound space, the shapes and positions of sound source objects in the sound space, and the position and orientation of the listener in the sound space.
[0464]

35 The sound space may be either a closed space or an open space.

Furthermore, the metadata may include information indicating the reflectance of each obstacle object that can reflect sound in the sound space. For example, the floor, walls, ceiling, and the like constituting the boundaries of the sound space can be included in the obstacle objects.

[0465]

The reflectance is an energy ratio between a reflected sound and an incident sound, and may be set for each sound frequency band. Of course, the reflectance may be uniformly set, irrespective of the sound frequency band. Note that when the sound space is an open space, for example, parameters such as a uniformly set attenuation rate, diffracted sound, and early reflected sound may be used.

[0466]

The metadata may include information other than reflectance as a parameter with regard to an obstacle object or a sound source object. For example, the metadata may include information on the material of an object as a parameter related to both of a sound source object and a non-sound-emitting object. Specifically, the metadata may include information such as the diffusivity, transmittance, and sound absorption rate.

[0467]

For example, information on a sound source object may include information indicating, for example, sound volume, radiation characteristics (directivity), a reproduction condition, the number and types of sound sources of one object, and a sound source region of an object. The reproduction condition may determine whether a sound is, for example, a sound that is continuously being emitted or is emitted at an event. The sound source region of an object may be determined by the relative relationship between the position of the listener and the position of the object, or may be determined using the object as a reference.

[0468]

For example, when the sound source region is determined by the relative relationship between the position of the listener and the position of the object, it is possible to cause the listener to perceive

sound E from the right side of the object and sound F from the left side of the object, the right side and the left side being as seen from the listener.

[0469]

5 Furthermore, when the sound source region is determined using the object as a reference, it is possible to fix what sound is emitted from what region of the object, using the object as a reference. For example, it is possible, when the listener sees the object from the front, to cause the listener to perceive a high sound
10 from the right side of the object and a low sound from the left side of the object. Furthermore, it is possible, when the listener sees the object from the rear, to cause the listener to perceive a low sound from the right side of the object and a high sound from the left side of the object.

15 [0470]

 Metadata related to the space may include the time period until early reflected sound, the reverberation time period, the ratio of direct sound to diffuse sound, and the like. When the ratio between a direct sound and a diffuse sound is zero, the listener can be caused
20 to perceive only the direct sound.

[0471]

[Embodiment 2]

 Embodiment 2 is described below. The description below is centered on the points of difference from Embodiment 1, and
25 descriptions of points in common are omitted or simplified.

[0472]

[Configuration of Renderer]

 First, the configuration of renderer 2300 according to the present embodiment is described. FIG. 26 is a block diagram
30 illustrating a configuration example of renderer 2300 according to the present embodiment.

[0473]

 Renderer 2300 includes analyzer 2301, determiner 2302, and reproducer 2303. It should be noted that as described above, the
35 audio signal processing device according to the present embodiment

is an example of a decoding device. The decoding device includes a decoder, and the decoder includes renderer 2300. In other words, it can be stated that the audio signal processing device according to the present embodiment includes analyzer 2301, determiner 2302, and
5 reproducer 2303. Renderer 2300 applies acoustic processing to sound data included in the input signal, and outputs the result.

[0474]

Similarly to Embodiment 1, the input signal includes, for example, spatial information, sensor information, and sound data.
10 The spatial information also includes physical information such as reflection coefficients, transmission coefficients, and diffraction coefficients of non-emitting objects (obstacle objects).

[0475]

Analyzer 2301 may be able to perform all of the processing or
15 some of the processing performed by analyzer 1301 according to Embodiment 1. Furthermore, analyzer 2301 creates an audio signal indicating reflected sound and an audio signal indicating direct sound, and stores the created audio signal indicating reflected sound and the created audio signal indicating direct sound.

20 [0476]

It should be noted that in the present embodiment, mainly reflected sound, which is an example of indirect sound, and direct sound, which is an example of predetermined sound, are used. In other words, in the present embodiment, reflected sound can be used
25 as the indirect sound and direct sound can be used as the predetermined sound. It should be noted that here, the description mainly uses reflected sound and direct sound, but the same processing is performed even if indirect sound is used instead of reflected sound and predetermined sound is used instead of direct
30 sound. Furthermore, indirect sound includes, for example, reflected sound, diffracted sound, and the like, and predetermined sound is sound different from indirect sound, including, for example, direct sound, HOA, and representative sound.

[0477]

35 Analyzer 2301 has propagation path detector 2301a and

memory 2301b.

[0478]

Propagation path detector 2301a creates an audio signal indicating reflected sound and an audio signal indicating direct sound,
5 based on the spatial information and the sound data.

[0479]

More specifically, propagation path detector 2301a creates an audio signal indicating reflected sound and an audio signal indicating direct sound based on: the position information of the sound source
10 object, the position information of the non-emitting object (obstacle object), and the position information and physical information of the listener that are included in the spatial information; and the sound data.

[0480]

15 In other words, propagation path detector 2301a creates an audio signal generated within the virtual space, based on the spatial information and the sound data, and then assigns attribute information indicating an attribute identifying the audio signal to the created audio signal, thereby creating an audio signal containing the
20 attribute information. The attribute indicates whether the sound indicated by the audio signal is a direct sound (predetermined sound) or a reflected sound (indirect sound).

[0481]

Sound that directly reaches the listener's head from one sound
25 source is direct sound, and sound that reaches the listener's head after being output from that one sound source and then reflected by a non-emitting object or diffracted by a non-emitting object is indirect sound (reflected sound or diffracted sound).

[0482]

30 Furthermore, a direct sound associated with an indirect sound means a direct sound that originates from the same sound source as that indirect sound. An indirect sound associated with a direct sound means an indirect sound that originates from the same sound source as that direct sound.

35 [0483]

An audio signal whose attribute is a reflected sound (indirect sound) includes information indicating an audio signal for the direct sound associated with that reflected sound (indirect sound).

[0484]

5 Propagation path detector 2301a stores the created audio signals in memory 2301b. FIG. 26 shows audio signals A, B, C, and D stored in memory 2301b, along with the respective attributes of audio signals A, B, C, and D. It should be noted that the audio signals may also be stored in a storage device other than memory 2301b.

10 [0485]

 Furthermore, as in Embodiment 1, propagation path detector 2301a may calculate, for each of the direct sound and reflected sound, values related to: the path until arriving at the listening position; the time period taken until arrival; the sound volume at arrival; and the
15 like. Similarly, propagation path detector 2301a may calculate information indicating the relationship between the direct sound and the reflected sound, such as a value related to the time difference between when the direct sound arrives and when the reflected sound arrives (the time difference between the direct sound and the
20 reflected sound), a value related to the sound volume ratio between the direct sound and the reflected sound at the listening position, and the like.

[0486]

 Each of an audio signal whose attribute is a reflected sound and
25 an audio signal whose attribute is a direct sound may contain information indicating the sound volume of the sound indicated by the audio signal, at the listening position. An audio signal whose attribute is a reflected sound may include information indicating the reflected sound arrival time sound volume (I_r) as the sound volume
30 of the reflected sound indicated by the audio signal at the listening position. An audio signal whose attribute is a direct sound may include information indicating the direct sound arrival time sound volume (I_d) as the sound volume of the direct sound indicated by the audio signal at the listening position.

35 [0487]

Determiner 2302 may be able to perform all of the processing or some of the processing performed by determiner 1302 according to Embodiment 1. Furthermore, determiner 2302 determines whether the output signal based on the audio signal created by analyzer 2301 (more specifically, propagation path detector 2301a) is to be output (reproduced) by reproducer 2303.

[0488]

Determiner 2302 includes classifier 2302a, first determiner 2302b, and second determiner 2302c.

[0489]

Classifier 2302a obtains an audio signal including attribute information created by propagation path detector 2301a and stored in memory 2301b. Classifier 2302a outputs the audio signal obtained to either first determiner 2302b or second determiner 2302c according to the attribute identified by the attribute information included in the audio signal.

[0490]

When the attribute identified by the attribute information included in the audio signal obtained is information indicating a direct sound (predetermined sound), classifier 2302a outputs the audio signal to first determiner 2302b. When the attribute identified by the attribute information included in the audio signal obtained is information indicating a reflected sound (indirect sound), classifier 2302a outputs the audio signal to second determiner 2302c.

[0491]

First determiner 2302b performs first determination processing to determine whether the audio signal output from classifier 2302a satisfies a first condition. When the audio signal satisfies the first condition, first determiner 2302b outputs the audio signal to reproducer 2303.

[0492]

Furthermore, first determiner 2302b does not perform the second determination processing. In other words, when the attribute identified by the attribute information included in the audio signal obtained is information indicating a predetermined sound, the second

determination processing is not performed. Thus, when the second determination processing is not performed, the amount of computation and the computational load can be reduced.

[0493]

5 Second determiner 2302c performs the first determination processing to determine whether the audio signal output from classifier 2302a satisfies the first condition, and the second determination processing to determine whether the audio signal satisfies a second condition different from the first condition. When
10 the audio signal satisfies both the first condition and the second condition, second determiner 2302c outputs the audio signal to reproducer 2303.

[0494]

15 Reproducer 2303 may be able to perform all of the processing or some of the processing performed by reproducer 1303 according to Embodiment 1. Furthermore, reproducer 2303 obtains the audio signal output from determiner 2302 and outputs an output signal based on the audio signal obtained.

[0495]

20 Reproducer 2303 has first reproducer 2303a and second reproducer 2303b. First reproducer 2303a outputs an output signal (first output signal) based on an audio signal whose attribute is a predetermined sound different from an indirect sound. Second reproducer 2303b outputs an output signal (second output signal)
25 based on an audio signal whose attribute is an indirect sound.

[0496]

 In other words, first reproducer 2303a obtains the audio signal output from first determiner 2302b and outputs the first output signal. Second reproducer 2303b obtains the audio signal output from second
30 determiner 2302c and outputs the second output signal.

[0497]

 First reproducer 2303a generates the first output signal by performing binaural filter processing on the audio signal obtained, and outputs the first output signal. Binaural filtering processing is
35 realized, for example, by subjecting the audio signal obtained to

processing using a head-related transfer function.

[0498]

Second reproducer 2303b generates the second output signal by performing binaural filter processing and diffusion filter processing on the audio signal obtained, and outputs the second output signal. The diffusion filter processing is, for example, processing that improves the realism of indirect sound by diffusing the indirect sound indicated by the audio signal in the audio signal obtained. Furthermore, the diffusion filter processing is processing in which a filter is used to simulate the auditory intensity of the sound diffusion indicated by the audio signal obtained (i.e., to simulate the auditory intensity of the sound diffusion as perceived by the listener). Finite impulse filters and/or infinite impulse filters are used in the diffusion filter processing.

[0499]

First reproducer 2303a outputs the first output signal without applying diffusion filter processing to the audio signal obtained.

[0500]

Since reproducer 2303 has the above-described configuration, audio signals whose attribute is indirect sound are subjected to diffusion filter processing and then output, thereby allowing listeners to hear indirect sounds with higher sound quality. Furthermore, since diffusion filter processing is not applied to audio signals whose attribute is the predetermined sound, the amount of computation and the computational load are further reduced.

[0501]

Hereinafter, an example of the operation of the audio signal processing method performed by the audio signal processing device according to the present embodiment (more specifically, renderer 2300) is described.

[0502]

[Operation Example of Renderer]

FIG. 27 is a flowchart illustrating an operation example of the audio signal processing device according to the present embodiment.

FIG. 27 illustrates the processing performed mainly by renderer 2300

included in the audio signal processing device according to the present embodiment.

[0503]

First, analyzer 2301 performs analysis processing to analyze the input signal (S501). More specifically, propagation path detector 2301a of analyzer 2301 creates an audio signal indicating reflected sound and an audio signal indicating direct sound, based on the spatial information and the sound data. Propagation path detector 2301a stores the audio signal created in memory 2301b.

[0504]

Analyzer 2301 analyzes the input signal and calculates, for each of the direct sound and reflected sound, values related to, e.g., the path until arrival at the listening position, the time taken to arrive, and the sound volume at arrival; a value related to the time difference between the direct sound and the reflected sound; and a value related to the sound volume ratio between the direct sound and the reflected sound at the listening position.

[0505]

First, propagation path detector 2301a calculates the characteristics of each of the direct sound indicated by the audio signal created and the reflected sound indicated by the audio signal created. Specifically, the arrival time period and the arrival time sound volume when each of the direct sound and the reflected sound arrives at the listener (listening position) are calculated. It should be noted that the method shown in Embodiment 1 may be used as the method for calculating these arrival time periods and arrival time sound volumes.

[0506]

Next, propagation path detector 2301a calculates the sound volume ratio (L), which is the ratio between the direct sound arrival time sound volume (I_d) and the reflected sound arrival time sound volume (I_r), and the time difference (T) between the direct sound and the reflected sound (the time difference (T) between when the direct sound arrives and when the reflected sound arrives). The direct sound arrival time sound volume (I_d) refers to the sound volume of a

direct sound, which is an example of a predetermined sound, at the time of arrival at the listening position, which is the location at which the listener is present in a virtual space. In other words, the direct sound arrival time sound volume (I_d) is the sound volume of the predetermined sound (direct sound) at the listening position. The reflected sound arrival time sound volume (I_r) refers to the sound volume of a reflected sound, which is an example of an indirect sound, at the time of arrival at the listening position. In other words, the reflected sound arrival time sound volume (I_r) is the sound volume of the indirect sound (reflected sound) at the listening position. That is, the sound volume ratio (L) is the sound volume ratio between the sound volume of the direct sound and the sound volume of the reflected sound (indirect sound) at the listening position. It should be noted that the methods shown in Embodiment 1 may be used as the methods for calculating the sound volume ratio (L) and the time difference (T).

[0507]

Determiner 2302 makes a determination on the audio signal ($S502$). More specifically, determiner 2302 determines whether the audio signal created by propagation path detector 2301a is to be output by reproducer 2303. That is, determiner 2302 performs the determination processing (selection processing).

[0508]

First, classifier 2302a obtains an audio signal including attribute information created by propagation path detector 2301a and stored in memory 2301b. Classifier 2302a outputs an audio signal whose attribute is a direct sound (predetermined sound) to first determiner 2302b, and outputs an audio signal whose attribute is a reflected sound (indirect sound) to second determiner 2302c.

[0509]

Furthermore, classifier 2302a also obtains the direct sound arrival time sound volume (I_d), the reflected sound arrival time sound volume (I_r), the sound volume ratio (L), and the time difference (T) calculated by propagation path detector 2301a.

[0510]

Next, first determiner 2302b performs the first determination processing on the audio signal output from classifier 2302a. Furthermore, second determiner 2302c performs the first determination processing and the second determination processing on the audio signal output from classifier 2302a. In other words, determiner 2302 (second determiner 2302c) performs the first determination processing and the second determination processing when the attribute identified by the attribute information included in the audio signal obtained is information indicating a reflected sound (indirect sound).

[0511]

Here, the first determination processing and the second determination processing are described. First, the second determination processing is described. It should be noted that in the present embodiment, the second determination processing is processing performed only on audio signals whose attribute is a reflected sound (indirect sound).

[0512]

The second determination processing is processing to determine whether the audio signal obtained satisfies the second condition. In the second determination processing, the audio signal obtained is determined to satisfy the second condition when the sound volume ratio between the direct sound (predetermined sound) and the reflected sound (indirect sound) arriving at the listening position is greater than or equal to the second threshold value, which is determined based on the time difference between the direct sound associated with the reflected sound, and the reflected sound. That is, the second determination processing corresponds to the processing performed by determiner 1302 in Embodiment 1. It should be noted that as described above, the time difference is calculated by propagation path detector 2301a, and second determiner 2302c obtains the time difference calculated and performs the second determination processing.

[0513]

As described above, the reflected sound and the direct sound

associated with that reflected sound originate from the same sound source.

[0514]

As described in Embodiment 1, the time difference between a direct sound and a reflected sound is, for example, the time difference between a direct sound arrival time period (arrival time) and a reflected sound arrival time period (arrival time), but is not limited thereto. The sound volume ratio between the direct sound and the reflected sound at the time at which the direct sound and the reflected sound arrive at the listening position corresponds to the sound volume ratio (L) in Embodiment 1, which is the ratio between the direct sound arrival time sound volume (I_d) and the reflected sound arrival time sound volume (I_r).

[0515]

Furthermore, the sound volume ratio (L) is calculated using the same method as in Embodiment 1.

[0516]

The second threshold value is a value determined based on the time difference between the direct sound associated with the reflected sound (indirect sound) and the reflected sound (indirect sound); in other words, the second threshold value is a value dependent on the time difference and is the value indicated by the threshold value data of Embodiment 1. The threshold value data is, for example, a graph having a horizontal axis that indicates the time difference between a direct sound and reflected sounds and a vertical axis that indicates the sound volume ratios of reflected sounds to a direct sound, and is expressed as a threshold value (second threshold value) that demarcates whether each reflected sound is perceived.

[0517]

More specifically, the threshold value data indicating the second threshold value is the data shown in FIG. 11 to FIG. 13 and the like.

[0518]

In the second determination processing, the audio signal obtained is determined to satisfy the second condition when the sound volume ratio between the direct sound and the reflected sound

(indirect sound), indicated by the audio signal obtained, is greater than or equal to the second threshold value.

[0519]

Next, the first determination processing is described. In the present embodiment, the first determination processing is processing performed on audio signals whose attribute is an indirect sound (reflected sound) and on audio signals whose attribute is information indicating a predetermined sound (direct sound).

[0520]

The first determination processing determines whether the audio signal obtained satisfies the first condition. In the first determination processing, the audio signal obtained is determined to satisfy the first condition when the amplitude value of the audio signal obtained is greater than or equal to the first threshold value. That is, since the amplitude value of an audio signal corresponds to the sound volume of the sound (reflected sound (indirect sound) or direct sound (predetermined sound)) indicated by that audio signal, when, in the first determination processing, the sound volume of the sound indicated by the audio signal obtained is greater than or equal to a certain level, the audio signal obtained is determined to satisfy the first condition.

[0521]

The first threshold value, unlike the second threshold value, is a value independent of the time difference between the direct sound associated with the reflected sound, and the reflected sound (indirect sound), and is a fixed value. The first threshold value is a value related to the sound volume of the audio signal obtained; in other words, the first threshold value is a value related to the amplitude value. Furthermore, the first threshold value indicates the sound volume of the boundary demarcating whether a sound is perceivable to the listener, and is a threshold value for determining a sound having a lower sound volume than the threshold value to be a sound that is not to be reproduced. FIG. 28 is a graph showing threshold value data indicating the first threshold value according to the present embodiment. For example, the first threshold value is -70 dB.

Since the amplitude value of audio signal B shown in FIG. 28 is greater than or equal to the first threshold value, audio signal B is determined to satisfy the first condition. Furthermore, since the amplitude value of audio signal A shown in FIG. 28 is less than the first threshold value, audio signal A is determined not to satisfy the first condition.
[0522]

It should be noted that for example, the first threshold value may be set (determined) by the administrator of or the listener in the virtual space using the above-mentioned threshold value setter.

[0523]

The first threshold value used in first determiner 2302b and the first threshold value used in second determiner 2302c may be the same value or may be different values.

[0524]

Determiner 2302 performs the first determination processing and the second determination processing as described above.

[0525]

Then, first determiner 2302b performs the first determination processing on the audio signal output from classifier 2302a. When the audio signal satisfies the first condition, first determiner 2302b outputs the audio signal to reproducer 2303.

[0526]

Second determiner 2302c performs the first determination processing to determine whether the audio signal output from classifier 2302a satisfies the first condition, and the second determination processing to determine whether the audio signal satisfies a second condition different from the first condition. It should be noted that in the present embodiment, the second determination processing is performed after the first determination processing is performed.

[0527]

Therefore, first, second determiner 2302c performs the first determination processing to determine whether the audio signal output from classifier 2302a satisfies the first condition, and performs the second determination processing on that audio signal, when the

audio signal satisfies the first condition. Then, second determiner 2302c outputs the audio signal to reproducer 2303, when the audio signal satisfies the second condition.

[0528]

5 Furthermore, reproducer 2303 obtains the audio signal output from determiner 2302 and outputs an output signal based on the audio signal (S503).

[0529]

10 More specifically, first reproducer 2303a generates the first output signal by obtaining the audio signal output from first determiner 2302b and performing binaural filter processing on the audio signal obtained, and outputs the first output signal. Second reproducer 2303b obtains the audio signal output from second determiner 2302c, applies binaural filter processing and diffusion
15 filter processing to the audio signal to generate a second output signal, and outputs the second output signal. In other words, reproducer 2303 (more specifically, second reproducer 2303b) outputs an output signal (the second output signal) based on the audio signal obtained, when the audio signal satisfies the first condition and the second
20 condition.

[0530]

It should be noted that first determiner 2302b does not output the audio signal to reproducer 2303 (first reproducer 2303a) when the audio signal does not satisfy the first condition. Since, in such
25 cases, first reproducer 2303a does not output an output signal based on the audio signal, the amount of computation and the computational load are reduced.

[0531]

Furthermore, second determiner 2302c does not output the
30 audio signal to reproducer 2303 (second reproducer 2303b) when the audio signal does not satisfy the first condition, and does not output the audio signal to reproducer 2303 (second reproducer 2303b) when the audio signal does not satisfy the second condition. Since, in such cases, second reproducer 2303b does not output an output signal
35 based on the audio signal, the amount of computation and the

computational load are reduced.

[0532]

The audio signal processing method according to the present embodiment is thus an audio signal processing method performed by
5 an audio signal processing device (renderer 2300), and includes an obtaining step, a determining step, and a reproducing step.

[0533]

In the obtaining step, an audio signal that includes attribute information identifying an attribute of the audio signal is obtained.
10 In the determining step, first determination processing and second determination processing are performed when the attribute identified by the attribute information included in the audio signal obtained is information indicating an indirect sound. The first determination processing is processing to determine whether the audio signal
15 obtained satisfies a first condition, and the second determination processing is processing to determine whether the audio signal obtained satisfies a second condition different from the first condition. In the reproducing step, an output signal based on the audio signal obtained is output when the audio signal obtained satisfies the first
20 condition and the second condition.

[0534]

In other words, when the first determination processing and the second determination processing are performed on an audio signal whose attribute is an indirect sound (reflected sound) and the audio
25 signal satisfies the first condition and the second condition, an output signal based on the audio signal obtained is output. That is, whether to output the output signal based on the audio signal for which the attribute is an indirect sound (reflected sound) is appropriately determined. When the output signal is not output, the amount of
30 computation and the computational load are reduced. In other words, it is possible to realize an audio signal processing method that can appropriately reduce the amount of computation and the computational load.

[0535]

35 Furthermore, in the present embodiment, in the first

determination processing, the audio signal obtained is determined to satisfy the first condition when the amplitude value of the audio signal obtained is greater than or equal to the first threshold value. In the second determination processing, the audio signal obtained is determined to satisfy the second condition when the sound volume ratio between a direct sound associated with an indirect sound and the indirect sound, when each arrives at the listening position that is the position at which the listener is present, is greater than or equal to the second threshold value, which is determined based on the time difference between when the direct sound arrives and when the indirect sound arrives.

[0536]

Thus, in the first determination processing, the audio signal is determined to satisfy the first condition when the above-described amplitude value is greater than or equal to the first threshold value, and in the second determination processing, the audio signal is determined to satisfy the second condition when the above-described sound volume ratio is greater than or equal to the second threshold value. An output signal based on the audio signal obtained is output when this second condition is satisfied. That is, whether to output the output signal based on the audio signal is more appropriately determined. In other words, it is possible to realize an audio signal processing method that can more appropriately reduce the amount of computation and the computational load.

[0537]

It should be noted that in the present embodiment, first determiner 2302b and second determiner 2302c are provided, but this is not intended to be limiting. For example, first determiner 2302b may be omitted, and second determiner 2302c may be provided. In this case, second determiner 2302c obtains both audio signals whose attribute is a predetermined sound (direct sound) and audio signals whose attribute is an indirect sound (reflected sound), and performs the first determination processing on both. When, in the first determination processing, an audio signal whose attribute is a predetermined sound (direct sound) is determined to satisfy the first

condition, the audio signal is output to first reproducer 2303a, and first reproducer 2303a outputs a first output signal based on the audio signal output. When, in the first determination processing, it is determined that an audio signal whose attribute is an indirect sound (reflected sound) satisfies the first condition, the second determination processing is further performed on that audio signal. When, in the second determination processing, the audio signal is determined to satisfy the second condition, the audio signal is output to second reproducer 2303b, and second reproducer 2303b outputs a second output signal based on the audio signal output. It can be stated that such a renderer in which first determiner 2302b is omitted and second determiner 2302c is provided performs essentially the same processing as renderer 2300 according to the present embodiment.

[0538]

[Embodiment 3]

Embodiment 3 is described below. The description below is centered on the points of difference from Embodiment 2, and descriptions of points in common are omitted or simplified.

[0539]

[Configuration of Renderer]

First, the configuration of renderer 3300 according to the present embodiment is described. FIG. 29 is a block diagram illustrating a configuration example of renderer 3300 according to the present embodiment.

[0540]

Renderer 3300 includes analyzer 2301, determiner 3302, and reproducer 3303.

[0541]

Determiner 3302 has the same configuration as determiner 2302 according to Embodiment 2, except that determiner 3302 has second determiner 3302c instead of second determiner 2302c.

[0542]

Reproducer 3303 has the same configuration as reproducer 2303 according to Embodiment 2, except that reproducer 3303 further

has gain setter 3303c.

[0543]

Gain setter 3303c sets (determines) the gain in the diffusion filter processing by second reproducer 2303b. That is, gain setter
5 3303c sets (determines) the amplification rate of the amplitude of the audio signal in the diffusion filter processing. In the present embodiment, second reproducer 2303b performs the diffusion filter processing using the gain determined by gain setter 3303c.

[0544]

10 For example, gain setter 3303c obtains the gain set (determined) by the administrator or the listener in the virtual space, and sets (determines) the gain obtained as the gain in the diffusion filter processing.

[0545]

15 Gain setter 3303c outputs the gain determined to determiner 3302 (more specifically, second determiner 3302c).

[0546]

In the present embodiment as well, when the attribute identified by the attribute information included in the audio signal
20 obtained is an indirect sound, classifier 2302a outputs the audio signal to second determiner 3302c.

[0547]

Second determiner 3302c performs the first determination processing to determine whether the audio signal output from
25 classifier 2302a satisfies the first condition. In the present embodiment, second determiner 3302c does not perform the second determination processing.

[0548]

Second determiner 3302c obtains the gain output by gain setter
30 3303c before performing the first determination processing. Second determiner 3302c adds the gain obtained to the audio signal output from classifier 2302a and performs the first determination processing on the audio signal to which the gain has been added. When the audio signal satisfies the first condition, second determiner 3302c
35 outputs the audio signal to reproducer 3303 (more specifically, second

reproducer 2303b).

[0549]

[Embodiment 4]

Embodiment 4 is described below. The description below is
5 centered on the points of difference from Embodiment 2, and
descriptions of points in common are omitted or simplified.

[0550]

[Configuration of Renderer]

First, the configuration of renderer 4300 according to the
10 present embodiment is described. FIG. 30 is a block diagram
illustrating a configuration example of renderer 4300 according to the
present embodiment.

[0551]

Renderer 4300 includes analyzer 2301, renderer pipeline 4304,
15 reproducer 4303, first gain accumulator 4305, and second gain
accumulator 4306.

[0552]

It should be noted that in the present embodiment as well,
reflected sound, which is an example of indirect sound, and direct
20 sound, which is an example of predetermined sound, are mainly used
for description, but the same processing is performed even if indirect
sound is used instead of reflected sound and predetermined sound is
used instead of direct sound.

[0553]

25 In the present embodiment, analyzer 2301 outputs the audio
signal stored in memory 2301b to renderer pipeline 4304.

[0554]

Renderer pipeline 4304 obtains the audio signal output from
analyzer 2301. For example, renderer pipeline 4304 obtains an audio
30 signal whose attribute is an indirect sound (reflected sound) (i.e., an
audio signal indicating indirect sound (reflected sound)).

[0555]

Furthermore, analyzer 2301 outputs the reference sound
volume included in the spatial information included in the input signal.
35 More specifically, analyzer 2301 outputs, to first gain accumulator

4305, the reference sound volume of the indirect sound (reflected sound) indicated by the audio signal, and first gain accumulator 4305 obtains the reference sound volume. Renderer pipeline 4304 performs, on the audio signal obtained, one or more first processes (here, a plurality of first processes), the first determination processing, the second determination processing, and one or more second processes (here, a plurality of second processes).

[0556]

Renderer pipeline 4304 has one or more first processors 4304a (here, a plurality of first processors 4304a), determiner 4302, and one or more second processors 4304b (here, a plurality of second processors 4304b).

[0557]

The plurality of first processors 4304a include first processor 4304a1, first processor 4304a2, first processor 4304a3, and first processor 4304a4. Each of the plurality of first processors 4304a performs the first process on the audio signal obtained. In the present embodiment, the first process is a process that determines the adjustment amount by which the amplitude of the audio signal is to be amplified, when processing based on physical characteristics has been performed on the audio signal. It should be noted that the first process is not limited thereto, and need not be a process that determines the adjustment amount by which the amplitude of the audio signal is to be amplified.

[0558]

The first processes respectively performed by the plurality of first processors 4304a may be different from one another. For example, first processors 4304a1 to 4304a4 respectively perform the following first processes.

[0559]

First processor 4304a1 performs, as the first process, processing to determine the adjustment amount by which the amplitude of the audio signal obtained is to be amplified, when the above-mentioned early reflection processor 1312 has performed processing on the audio signal.

[0560]

First processor 4304a2 performs, as the first process, processing to determine the adjustment amount by which the amplitude of the audio signal obtained is to be amplified, when the
5 above-mentioned diffraction processor has performed processing on the audio signal.

[0561]

First processor 4304a3 performs, as the first process, processing to determine the adjustment amount by which the
10 amplitude of the audio signal obtained is to be amplified, when the above-mentioned distance attenuation processor 1313 has performed processing on the audio signal.

[0562]

First processor 4304a4 performs, as the first process, processing to determine the adjustment amount by which the
15 amplitude of the audio signal obtained is to be amplified, when the above-described reverberation processor 1311 has performed processing on the audio signal.

[0563]

It should be noted that, as another example, first processor 4304a may perform, as the first process, processing to determine the adjustment amount by which the amplitude of the audio signal obtained is to be amplified, when processing such as pass-through, directivity, sound image localization, or diffusion has been performed
20 on the audio signal.

[0564]

Each of first processors 4304a1 to 4304a4 performs the above-described first process and outputs the adjustment amount determined to first gain accumulator 4305.

30 [0565]

In the present embodiment, after the first process is performed by each of the plurality of first processors 4304a, the audio signal obtained is output to determiner 4302, and processing is performed by determiner 4302.

35 [0566]

Determiner 4302 performs the first determination processing and the second determination processing. It should be noted that the first determination processing and the second determination processing according to the present embodiment are once again described, after describing first gain accumulator 4305 and second gain accumulator 4306.

[0567]

The plurality of second processors 4304b include second processor 4304b1 and second processor 4304b2. Each of the plurality of second processors 4304b performs the second process on the audio signal obtained. The second process determines the adjustment amount by which the amplitude of the audio signal is to be amplified, when processing based on the characteristics of the auditory perception of the listener has been performed on the audio signal. Furthermore, the second process may be a sound quality adjustment function performed based on selection by the listener, such as the preference or convenience of the listener, and may be processing accompanied by adjustment of the amplitude of the signal. It should be noted that the second process is not limited thereto, and need not be a process that determines the adjustment amount by which the amplitude of the audio signal is to be amplified.

[0568]

The second processes respectively performed by the plurality of second processors 4304b may be different from one another. For example, second processors 4304b1 and 4304b2 respectively perform the following second processes.

[0569]

Second processor 4304b1 performs, as the second process, processing to determine the adjustment amount by which the amplitude of the audio signal obtained is to be amplified, when diffusion filter processing has been performed on the audio signal. The diffusion filter processing is, for example, processing that improves the realism of indirect sound by diffusing the indirect sound indicated by the audio signal in the audio signal obtained, and uses a finite impulse response filter or an infinite impulse response filter.

[0570]

Second processor 4304b2 performs, as the second process, processing to determine the adjustment amount by which the amplitude of the audio signal obtained is to be amplified, when the
5 above-described binaural filter processing has been performed on the audio signal.

[0571]

Here, the second process performed by second processor 4304b2 is described with reference to FIG. 31 to FIG. 34. This is
10 described here by using an example of a head-related transfer function employed in binaural filter processing.

[0572]

FIG. 31 to FIG. 34 are each a diagram illustrating the energy of the head-related transfer function according to the present
15 embodiment. FIG. 31 to FIG. 34 each show listening position P of the listener. Furthermore, FIG. 31 to FIG. 34 each indicate the front, back, left, right, up, and down directions relative to the listener.

[0573]

FIG. 31 schematically illustrates a sphere centered on listening
20 position P, and on the sphere, circles are depicted along the listener's horizontal plane, midline plane, and frontal plane, respectively. In FIG. 32, the circle along the horizontal plane of the listener is shown with a thick line. In FIG. 33, the circle along the midline plane of the listener is shown with a thick line. In FIG. 34, the circle along the
25 frontal plane of the listener is shown with a thick line.

[0574]

In each of FIG. 31 through FIG. 34, the direction where the energy of the head-related transfer function is maximized is indicated by a circle filled with dots, while the direction where the energy of
30 the head-related transfer function is minimized is indicated by an open circle. Furthermore, FIG. 31 through FIG. 34 each show the value at which the energy of the head-related transfer function is maximized and the value at which the energy of the head-related transfer function is minimized.

35 [0575]

As shown in FIG. 31 to FIG. 34, the energy of the head-related transfer function differs for each localization direction. In FIG. 31 to FIG. 34, the energy of the head-related transfer function is displayed in dB relative to the energy when a transfer function with a transfer characteristic of 1 is set to 0 dB.

[0576]

FIG. 31 to FIG. 34 show the energy of one example of a head-related transfer function, but this energy varies significantly depending on the head-related transfer function used. In the second process performed by second processor 4304b2, the processing to determine the adjustment amount is performed based on the head-related transfer function used.

[0577]

Each of second processors 4304b1 and 4304b2 performs the above-described second process and outputs the adjustment amount determined to second gain accumulator 4306.

[0578]

First gain accumulator 4305 and second gain accumulator 4306 are further described.

[0579]

First gain accumulator 4305 calculates a first adjustment amount. The first adjustment amount is the cumulative value of the adjustment amounts determined by each of the plurality of first processes (more specifically, each of the plurality of first processors 4304a), each adjustment amount being the amount by which the amplitude of the audio signal obtained is to be amplified. More specifically, first gain accumulator 4305 calculates the first adjustment amount by cumulatively adding the respective adjustment amounts to the reference sound volume. That is, the first adjustment amount is the cumulative value of: the adjustment amounts determined by each of the plurality of first processors 4304a; and the reference sound volume.

[0580]

As described above, the plurality of first processors 4304a determine a plurality of (in this case, four) adjustment amounts for

the audio signal obtained. Each of the plurality of first processors 4304a outputs the adjustment amount determined to first gain accumulator 4305. First gain accumulator 4305 obtains the plurality of adjustment amounts and calculates the first adjustment amount by
5 cumulatively adding the plurality of (four) adjustment amounts and the reference sound volume. First gain accumulator 4305 outputs the first gain adjustment amount calculated to second gain accumulator 4306, and second gain accumulator 4306 obtains the first adjustment amount output.

10 [0581]

Second gain accumulator 4306 calculates the second adjustment amount. The second adjustment amount is the cumulative value of the adjustment amounts determined by each of the plurality of second processes (more specifically, each of the
15 plurality of second processors 4304b), each adjustment amount being the amount by which the amplitude of the audio signal obtained is to be amplified. More specifically, second gain accumulator 4306 cumulatively adds the respective adjustment amounts and calculates the second adjustment amount by further cumulatively adding, to the
20 accumulated adjustment amount, the first adjustment amount output from first gain accumulator 4305. That is, the second adjustment amount is the cumulative value of: the adjustment amounts determined by each of the plurality of second processors 4304b; and the first adjustment amount.

25 [0582]

As described above, the plurality of second processors 4304b determine a plurality of (in this case, two) adjustment amounts for the audio signal obtained. Each of the plurality of second processors 4304b outputs a parameter indicating the adjustment amount
30 determined to second gain accumulator 4306. Second gain accumulator 4306 obtains the plurality of parameters output and calculates the second adjustment amount by cumulatively adding: the adjustment amounts indicated by each of the plurality of parameters; and the first adjustment amount.

35 [0583]

It should be noted that each of the plurality of second processors 4304b may output the adjustment amount determined to second gain accumulator 4306, without employing the above-described parameters. In this case, second gain accumulator 4306
5 obtains the plurality of adjustment amounts output and calculates the second adjustment amount by cumulatively adding the plurality of (two) adjustment amounts and the first adjustment amount.

[0584]

Furthermore, as described above, the second process may be
10 a sound quality adjustment function performed based on selection by the listener, and may be processing accompanied by adjustment of the amplitude of the signal. For example, the second adjustment amount may include an amount of amplitude adjustment by the sound quality adjustment function, set based on selection by the listener in
15 the virtual space. In this case, the listener may perform selection, for example, by operating an operation reception component, and second gain accumulator 4306 may calculate the second adjustment amount based on the selection received by the operation reception component. The listener performing the selection to achieve his/her
20 preferred sound quality, for example, allows the listener to listen to sound having his/her preferred sound quality. Furthermore, the second adjustment amount may further include an amount of amplitude adjustment by the sound quality adjustment function, set based on selection by the administrator of the virtual space instead
25 of by the listener in the virtual space.

[0585]

Once again, the first determination processing and the second determination processing according to the present embodiment are described.

30 [0586]

First, the first determination processing is described.

[0587]

Determiner 4302 performs the first determination processing on the audio signal obtained by renderer pipeline 4304. In the first
35 determination processing, the audio signal obtained is determined to

satisfy the first condition when the first adjustment amount (G1) calculated by first gain accumulator 4305 is greater than or equal to the first threshold value (T1).

[0588]

5 The first threshold value according to the present embodiment is a fixed value related to the sound volume of the audio signal; in other words, the first threshold value is a value related to the amplitude value. In other words, the first threshold value according to the present embodiment is the same as the first threshold value
10 described in Embodiment 2.

[0589]

 It should be noted that in the first determination processing according to Embodiment 2, the audio signal obtained was determined to satisfy the first condition when the amplitude value of the audio
15 signal obtained was greater than or equal to the first threshold value. That is, the first determination processing according to the present embodiment is the same as the first determination processing according to Embodiment 2, except that the amplitude value of the audio signal obtained is changed to the first adjustment amount (G1).

20 [0590]

 Furthermore, the second determination processing is described.

[0591]

 Determiner 4302 performs the second determination processing on the audio signal obtained by renderer pipeline 4304.
25 In the second determination processing, the audio signal obtained is determined to satisfy the second condition when a value based on the second adjustment amount, which was calculated by second gain accumulator 4306, is greater than or equal to the second threshold value.

30 [0592]

 The second threshold value according to the present embodiment is a value different from the first threshold value, and corresponds to the second threshold value described in Embodiment 2. That is, the second threshold value is a value determined based
35 on the time difference between: the arrival of the direct sound

associated with the reflected sound (indirect sound) indicated by the audio signal obtained; and the arrival of the reflected sound (indirect sound).

[0593]

5 The second adjustment amount calculated by second gain accumulator 4306 corresponds to the reflected sound arrival time sound volume (I_r) indicated in Embodiments 1 and 2. That is, the second adjustment amount indicates the sound volume when the reflected sound indicated by the audio signal obtained arrives at the
10 listening position.

[0594]

 The value based on the second adjustment amount is the ratio (sound volume ratio) between the sound volume of the direct sound associated with the indirect sound (reflected sound) indicated by the
15 audio signal obtained and the second adjustment amount calculated by second gain accumulator 4306. That is, the value based on the second adjustment amount is equivalent to the sound volume ratio (L) indicated in Embodiment 2, which is the ratio between the direct sound arrival time sound volume (I_d) and the reflected sound arrival
20 time sound volume (I_r).

[0595]

 It should be noted that in the second determination processing according to Embodiment 2, the audio signal obtained is determined to satisfy the second condition, when the sound volume ratio (L)
25 between the direct sound and the indirect sound when each of the direct sound and the indirect sound arrives at the listening position is greater than or equal to the second threshold value. That is, the second determination processing according to the present embodiment is the same as the second determination processing
30 according to Embodiment 2, except that the sound volume ratio between the direct sound and the indirect sound is changed to the value based on the second adjustment amount.

[0596]

 Furthermore, the second adjustment amount and the value
35 based on the second adjustment amount according to the present

embodiment can be calculated using the same method as the reflected sound arrival time sound volume (I_r) and sound volume ratio (L), each described in Embodiment 2.

[0597]

5 In the present embodiment, the second determination processing is skipped when the audio signal obtained is determined not to satisfy the first condition in the first determination processing. Furthermore, the second determination processing is performed when the audio signal obtained is determined to satisfy the first condition
10 in the first determination processing.

[0598]

 It should be noted that as described above, the second adjustment amount is calculated before the second determination processing is performed. Due to utilizing processing based on the
15 characteristics of the auditory perception of the listener, the second processing performed by second processor 4304b determines the adjustment amount independently of the audio signal (in other words, statically). Therefore, the second adjustment amount can be calculated before the second determination processing is performed.

20 [0599]

 Then, when the audio signal obtained is determined to satisfy the second condition in the second determination processing, determiner 4302 outputs the audio signal obtained to a plurality of second processors 4304b, and further, renderer pipeline 4304 outputs
25 the audio signal obtained to reproducer 4303. Furthermore, in this case, renderer pipeline 4304 may output the second adjustment amount calculated by second gain accumulator 4306 to reproducer 4303. Then, processing by reproducer 4303 is performed.

[0600]

30 Reproducer 4303 obtains the audio signal output from renderer pipeline 4304 and the second adjustment amount. Reproducer 4303 outputs an output signal based on the audio signal obtained. More specifically, reproducer 4303 may generate an output signal by multiplying the audio signal obtained by the second adjustment
35 amount obtained, and output the output signal. In the present

embodiment, reproducer 4303 thus outputs the output signal when the audio signal output satisfies both the first condition and the second condition.

[0601]

5 Hereinafter, an example of the operation of the audio signal processing method performed by the audio signal processing device according to the present embodiment (more specifically, renderer 4300) is described.

[0602]

10 [Operation Example of Renderer]

FIG. 35 is a flowchart illustrating an operation example of the audio signal processing device according to the present embodiment. FIG. 35 illustrates the processing performed mainly by renderer 4300 included in the audio signal processing device according to the
15 present embodiment.

[0603]

First, analyzer 2301 performs analysis processing to analyze the input signal (S601).

[0604]

20 Furthermore, renderer pipeline 4304 obtains the audio signal output from analyzer 2301 (S602). It should be noted that renderer pipeline 4304 obtains, for example, an audio signal indicating indirect sound.

[0605]

25 Then, each of the plurality of first processors 4304a determines the adjustment amount by which the amplitude of the audio signal obtained is to be amplified (S603). Each of the plurality of first processors 4304a outputs the adjustment amount determined to first gain accumulator 4305. Furthermore, analyzer 2301 outputs the
30 reference sound volume of the indirect sound (reflected sound) to first gain accumulator 4305.

[0606]

First gain accumulator 4305 obtains the plurality of (four) adjustment amounts output and the reference sound volume, and
35 then cumulatively adds the plurality of adjustment amounts obtained

and the reference sound volume obtained to calculate the first adjustment amount (S604). Then, first gain accumulator 4305 outputs the first adjustment amount calculated to determiner 4302. Furthermore, first gain accumulator 4305 outputs the first adjustment amount calculated to second gain accumulator 4306.

[0607]

Next, each of the plurality of second processors 4304b determines the adjustment amount by which the amplitude of the audio signal obtained is to be amplified (S605). Each of the plurality of second processors 4304b outputs the adjustment amount determined to second gain accumulator 4306.

[0608]

Second gain accumulator 4306 obtains the plurality of (two) adjustment amounts output and the first adjustment amount calculated, and calculates the second adjustment amount by cumulatively adding the plurality of adjustment amounts obtained and the first adjustment amount obtained (S606). Then, second gain accumulator 4306 outputs the second adjustment amount calculated to determiner 4302.

[0609]

Determiner 4302 performs the first determination processing (S607). In other words, determiner 4302 determines whether the audio signal obtained satisfies the first condition, by determining whether the first adjustment amount calculated is greater than or equal to the first threshold value.

[0610]

Determiner 4302 performs the second determination processing (S608). In other words, determiner 4302 determines whether the audio signal obtained satisfies the second condition, by determining whether the value based on the second adjustment amount calculated is greater than or equal to the second threshold value.

[0611]

It should be noted that in the present embodiment, the second determination processing is performed when the audio signal obtained

satisfies the first condition.

[0612]

Furthermore, reproducer 4303 outputs an output signal based on the audio signal obtained, when the audio signal obtained satisfies
5 the second condition (S609).

[0613]

Furthermore, when the audio signal does not satisfy the first condition, the audio signal is not output to reproducer 4303, and when the audio signal does not satisfy the second condition, the audio
10 signal is not output to reproducer 4303. In such cases, reproducer 4303 does not output an output signal based on the audio signal, thereby reducing the amount of computation and computational load.

[0614]

It should be noted that in the present embodiment, renderer
15 pipeline 4304 may include classifier 2302a described in Embodiment 2. Classifier 2302a may, for example, be provided between the plurality of first processors 4304a and determiner 4302, but is not limited thereto.

[0615]

As described above, the audio signal processing method according to the present embodiment is an audio signal processing method executed by an audio signal processing device, and includes: an obtaining step; a renderer pipeline step; a reproducing step; a first gain accumulation step; and a second gain accumulation step.
20

25 [0616]

In the obtaining step, an audio signal indicating indirect sound is obtained. In the renderer pipeline step, one or more first processes, first determination processing, second determination processing, and one or more second processes different from the one
30 or more first processes are performed on the audio signal obtained.

[0617]

In the first gain accumulation step, a first adjustment amount is calculated by cumulatively adding the adjustment amounts each determined by a respective one of the one or more first processes.
35 The adjustment amounts are each an adjustment amount by which

the amplitude of the audio signal obtained is to be amplified. In the second gain accumulation step, a second adjustment amount is calculated by cumulatively adding the adjustment amounts each determined by a respective one of the one or more second processes.

- 5 The adjustment amounts are each an adjustment amount by which the amplitude of the audio signal obtained is to be amplified. In the first determination processing, the audio signal obtained is determined to satisfy the first condition when the first adjustment amount calculated is greater than or equal to the first threshold value.
- 10 The second determination processing is skipped when the audio signal obtained is determined not to satisfy the first condition. The second determination processing is performed when the audio signal obtained is determined to satisfy the first condition. In the second determination processing, the audio signal obtained is determined to
- 15 satisfy the second condition when the value based on the second adjustment amount is greater than or equal to a second threshold value different from the first threshold value, and the reproducing step is performed when the audio signal obtained is determined to satisfy the second condition.

20 [0618]

- As a result, the first determination processing and the second determination processing are performed on the audio signal indicating the indirect sound, and an output signal based on the audio signal obtained is output when the audio signal satisfies the first condition
- 25 and the second condition. In other words, whether to output the output signal based on the audio signal indicating the indirect sound is appropriately determined. When the output signal is not output, the amount of computation and the computational load are reduced. In other words, it is possible to realize an audio signal processing
- 30 method that can appropriately reduce the amount of computation and the computational load.

[0619]

- In the present embodiment, the second adjustment amount calculated in the second gain accumulation step is the adjustment
- 35 amount applied when diffusion filter processing that enhances the

realism of an indirect sound by diffusing the indirect sound is performed. The second adjustment amount includes the adjustment amounts by which the amplitude of the audio signal obtained is to be amplified. The above-described value is the ratio between the sound volume of the direct sound associated with the indirect sound and the second adjustment amount calculated in the second gain accumulation step, and the second threshold value is determined according to the time difference between when the direct sound arrives and when the indirect sound arrives.

10 [0620]

Accordingly, in the second determination processing, when the above-described ratio is greater than or equal to the second threshold value, the audio signal indicating the indirect sound is determined to satisfy the second condition, and an output signal based on the audio signal obtained is output when this second condition is satisfied. That is, whether to output the output signal based on the audio signal indicating the indirect sound is more appropriately determined. In other words, it is possible to realize an audio signal processing method that can more appropriately reduce the amount of computation and the computational load.

20 [0621]

In the present embodiment, the second adjustment amount calculated in the second gain accumulation step includes the amount of amplitude adjustment by the sound quality adjustment function, set based on selection by the listener. The above-described value is the ratio between the sound volume of the direct sound associated with the indirect sound and the second adjustment amount calculated in the second gain accumulation step, and the second threshold value is determined according to the time difference between when the direct sound arrives and when the indirect sound arrives.

30 [0622]

The listener can thus listen to sound having his/her preferred sound quality, by performing selection to achieve his/her preferred sound quality, for example.

35 [0623]

In the present embodiment, the first threshold value is a value related to sound volume.

[0624]

5 This makes it possible to realize an audio signal processing method in which a value related to sound volume can be used as the first threshold value.

[0625]

10 In the present embodiment, in the second gain accumulation step, parameters that indicate the adjustment amounts by which the amplitude of the audio signal obtained is to be amplified and that have been determined by each of one or more second processes are obtained, and the second adjustment amount is calculated according to the plurality of parameters obtained.

[0626]

15 The second adjustment amount is thus calculated according to the above-described parameters, whereby whether to output an output signal based on the audio signal whose attribute is an indirect sound is more appropriately determined. In other words, it is possible to realize an audio signal processing method that can more
20 appropriately reduce the amount of computation and the computational load.

[0627]

[Embodiment 5]

25 Embodiment 5 is described below. The description below is centered on the points of difference from Embodiment 4, and descriptions of points in common are omitted or simplified.

[0628]

[Configuration of Renderer]

30 First, the configuration of renderer 5300 according to the present embodiment is described. FIG. 36 is a block diagram illustrating a configuration example of renderer 5300 according to the present embodiment.

[0629]

35 Renderer 5300 has the same configuration as renderer 4300 according to Embodiment 4, except that renderer 5300 further

includes changer 5307 and disabler 5308.

[0630]

Changer 5307 performs the processing to set (determine) the first threshold value used in the first determination processing. For example, changer 5307 outputs an instruction to set the first threshold value to determiner 4302. Determiner 4302 obtains the instruction output and sets the first threshold value according to that instruction. Thus, in the first determination processing, the audio signal obtained is determined to satisfy the first condition when the first adjustment amount calculated is greater than or equal to the first threshold value set.

[0631]

Disabler 5308 performs disabling processing to disable the second determination processing. When the disabling processing is performed, the second determination processing is disabled, or in other words, no longer performed. For example, disabler 5308 outputs an instruction indicating the disabling processing to determiner 4302. When determiner 4302 obtains the instruction output, determiner 4302 ceases to perform the second determination processing.

[0632]

Here, a case in which, in the present embodiment, the audio signal obtained in the first determination processing is determined to satisfy the first condition and the disabling processing has been performed is described. In this case, determiner 4302 outputs the audio signal obtained to the plurality of second processors 4304b, and further, renderer pipeline 4304 outputs the audio signal obtained to reproducer 4303. In other words, in Embodiment 4, reproducer 4303 output the output signal when the audio signal output satisfies both the first condition and the second condition. However, in this case, reproducer 4303 outputs the output signal when the audio signal output satisfies the first condition.

[0633]

Furthermore, the first threshold value set by changer 5307 may have different values between a case in which the disabling processing

has been performed and a case in which the disabling processing has not been performed. More specifically, the first threshold value set when the disabling processing is performed may be greater than the first threshold value set when the disabling processing has not been performed. Below, the effects due to the setting of the first threshold value and the disabling processing are described with reference to FIG. 37.

[0634]

FIG. 37 is a diagram showing a table for illustrating the effects of setting the first threshold value and performing disabling processing, according to the present embodiment.

[0635]

FIG. 37 illustrates a case in which the first threshold value has become larger ("First threshold value: large") and a case in which the first threshold value has become smaller ("First threshold value: small"), due to the setting of the first threshold value. Furthermore, FIG. 37 illustrates a case in which the disabling processing has been performed ("Second determination processing disabled") and a case in which the disabling processing has not been performed ("Second determination processing enabled"). Furthermore, FIG. 37 shows the processing complexity, sound quality maintenance, and decimation factor.

[0636]

The processing complexity indicates the complexity of the processing by renderer 5300. When the processing complexity is low, meaning processing is simple, "○" is indicated, and when the processing complexity is high, meaning processing is difficult, "x" is indicated.

[0637]

Sound quality maintenance indicates whether the sound quality of the signal output from reproducer 4303 is maintained. "○" indicates that the sound quality is maintained, while "x" indicates that the sound quality decreases.

[0638]

The decimation factor indicates the number of audio signals,

among the audio signals (more specifically, the plurality of audio signals) obtained by renderer pipeline 4304, that are not output as output signals from reproducer 4303. When the decimation factor is high, that is, when a large number of audio signals are not output, thereby reducing the amount of computation and the computational load, "○" is indicated. When the decimation factor is low, that is, when the number of audio signals not output is low, thereby making it difficult to reduce the amount of computation and the computational load, "×" is indicated.

10 [0639]

As shown in FIG. 37, in the case of "First threshold value: large" and "Second determination processing disabled", "○" is indicated for the processing complexity, "×" is indicated for the sound quality maintenance, and "○" is indicated for the decimation factor. In other words, in this case, the processing by renderer 5300 is simple, and the amount of computation and the computational load can be sufficiently reduced.

[0640]

Similarly, in the case of "First threshold value: large" and "Second determination processing enabled", "×" is indicated for the processing complexity, "×" is indicated for the sound quality maintenance, and "○" is indicated for the decimation factor. In other words, in this case, the amount of computation and computational load can be sufficiently reduced.

25 [0641]

Similarly, in the case of "First threshold value: small" and "Second determination processing disabled", "○" is indicated for the processing complexity, "○" is indicated for the sound quality maintenance, and "×" is indicated for the decimation factor. In other words, in this case, high sound quality can be maintained.

[0642]

Similarly, in the case of "First threshold value: small" and "Second determination processing enabled", "×" is indicated for the processing complexity, "○" is indicated for the sound quality maintenance, and "○" is indicated for the decimation factor. In other

words, in this case, high sound quality can be maintained, and the amount of computation and the computational load can be sufficiently reduced.

[0643]

5 It should be noted that whether changer 5307 performs the processing to set the first threshold value, and whether disabler 5308 performs the disabling processing, may be set (determined) by the administrator of or the listener in the virtual space. Furthermore, the spatial information may include information indicating whether
10 changer 5307 performs the processing to set the first threshold value and whether disabler 5308 performs the disabling processing, and the determination may be made in accordance with this information. By doing so, the administrator of or the listener in the virtual space can select the most appropriate effects from among various effects such
15 as those shown in FIG. 37, through trial and error. In order to enable the repetition of trial and error to be performed effectively, the respective setting regions may be adjacent to each other. Specifically, the storage region for storing the value indicating the first threshold value set by changer 5307 and the storage region for
20 storing the signal instructing that the disabling processing be performed may be adjacent regions. This not only makes it easier to visually grasp the set values due to their adjacency, but also yields the distinct advantage that, because the set values are adjacent in terms of memory region, both values can be set simultaneously (in a
25 single memory access operation). In other words, by linking and arranging the value indicating the first threshold value and the signal instructing performance of the disabling processing in the upper bit field and lower bit field of a region accessible by a single address, and by writing the linked data sequence thus arranged in a single memory
30 access, the two data items can be set simultaneously.

[0644]

 Furthermore, the processing by which changer 5307 sets the first threshold value and the disabling processing performed by disabler 5308 may be performed during the update of the above-
35 described thread.

[0645]

As described above, the audio signal processing method according to the present embodiment includes a changing step for setting the first threshold value and a disabling step for performing the disabling processing that disables the second determination processing. In the first determination processing, the audio signal obtained is determined to satisfy the first condition when the first adjustment amount calculated is greater than or equal to the first threshold value set. When the disabling processing is performed, the reproducing step is performed upon determining that the audio signal obtained satisfies the first condition.

[0646]

The second determination processing thus ceases to be performed due to the disabling processing, thereby reducing the amount of computation and the computational load required for performing the second determination processing. In other words, it is possible to realize an audio signal processing method that can more appropriately reduce the amount of computation and the computational load.

[0647]

In the audio signal processing method according to the present embodiment, the first threshold value set when the disabling processing has been performed is greater than the first threshold value set when the disabling processing has not been performed.

[0648]

This makes it possible to realize an audio signal processing method that enables setting the magnitude of the first threshold value based on whether the disabling processing is performed.

[0649]

(Supplement)

Note that the aspects understood based on the present disclosure are not limited to the embodiment, and various changes may be performed.

[0650]

For example, a process performed by a certain constituent

element in the embodiment may be performed by another constituent element instead of the specific constituent element. Furthermore, the order of a plurality of processes may be changed, or a plurality of processes may be performed in parallel.

5 [0651]

Moreover, ordinals such as first and second used for description may be interchanged, removed, or newly assigned as appropriate. These ordinals do not necessarily correspond to meaningful orders, and may be used to distinguish between elements.

10 [0652]

Furthermore, for example, in comparisons between threshold values, "greater than or equal to" a threshold value and "greater than" a threshold value may be read interchangeably. Similarly, "less than or equal to" a threshold value and "less than" a threshold value may be read interchangeably. Moreover, for example, there may be cases in which the terms "time period" and "time" are read interchangeably.

[0653]

Furthermore, in a process for selecting one or more sounds to be processed from a plurality of sounds, no sounds need be selected as a sound to be processed if no sounds that satisfy the conditions exist. In other words, a case in which no sounds to be processed are selected may be included in the process for selecting one or more sounds to be processed from a plurality of sounds.

[0654]

25 Furthermore, at least one of a first element, a second element, or a third element can correspond to the first element, the second element, the third element, or any combination of these.

[0655]

30 In addition, for example, in the embodiment, the case in which the aspects that are understood based on the present disclosure are implemented as an audio signal processing device, an encoding device, or a decoding device has been described. However, the aspects that are understood based on the present disclosure are not limited thereto, and may be implemented as software for executing the audio signal processing method, the encoding method, or the decoding

35

method.

[0656]

For example, a program for executing the above-described audio signal processing method, encoding method, or decoding method may be stored beforehand in ROM. Then, a CPU may operate according to this program.

[0657]

Furthermore, a program for executing the above-described audio signal processing method, encoding method, or decoding method may be stored on a computer-readable recording medium. Then, a computer may record, in computer RAM, the program stored on the recording medium, and operate according to this program.

[0658]

Moreover, each of the above-described constituent elements may be expressed typically as a large-scale integration (LSI), which is an integrated circuit (IC) having an input terminal and an output terminal. These may take the form of individual chips, or all or one or more constituent elements of the embodiment may be encapsulated in a single chip. Depending upon the level of integration, the LSI may be expressed as an IC, a system LSI, a super LSI, or an ultra LSI.

[0659]

Furthermore, such IC is not limited to an LSI, and a dedicated circuit or a general-purpose processor may be used. Alternatively, a field programmable gate array (FPGA) that allows for programming after the manufacture of an LSI, or a reconfigurable processor that allows for reconfiguration of the connection and the setting of circuit cells inside an LSI may be employed. Furthermore, when a circuit integration technology that replaces LSIs comes along owing to advances in semiconductor technology or to a separate derivative technology, the constituent elements should naturally be integrated using that technology. The adaptation of biotechnology, and the like are also conceivable as possibilities.

[0660]

Moreover, an FPGA, a CPU, or the like may, by means of wireless

communication or wired communication, download all or a part of the software for executing the audio signal processing method, the encoding method, or the decoding method described in the present disclosure. Furthermore, all or a part of software for updating may
5 be downloaded by means of wireless communication or wired communication. Moreover, an FPGA, a CPU, or the like may execute the digital signal processing described in the present disclosure by storing the downloaded software in memory and operating based on the stored software.

10 [0661]

At this time, the machine that includes the FPGA, the CPU, or the like may be connected wirelessly or in a wired manner to a signal processing device, or may be connected to a signal processing server over a network. Accordingly, this machine and the signal processing
15 device or the signal processing server may perform the audio signal processing method, the encoding method, or the decoding method described in the present disclosure.

[0662]

For example, the audio signal processing device, the encoding
20 device, or the decoding device in the present disclosure may include an FPGA, a CPU, or the like. Furthermore, the audio signal processing device, the encoding device, or the decoding device may include: an interface for acquiring, from an external source, the software for causing the FPGA, the CPU, or the like to operate; and
25 memory for storing the acquired software. The FPGA, the CPU, or the like may perform the signal processing described in the present disclosure by operating based on the stored software.

[0663]

A server may provide the software related to the acoustic
30 processing, the encoding processing, or the decoding processing of the present disclosure. Furthermore, a terminal or a machine may operate as the audio signal processing device, the encoding device, or the decoding device described in the present disclosure by installing the software. Note that the terminal or the machine may
35 install the software by connecting to a server over a network.

[0664]

Furthermore, the software may be installed on the terminal or the machine by means of another device that is different from the terminal or the machine obtaining data for installing the software by connecting to a server over a network and providing the data for installing the software to the terminal or the machine. Note that VR software or AR software for causing a terminal or a machine to execute the audio signal processing method described by way of the embodiment may be an example of the software.

[0665]

Note that in the foregoing embodiment, each constituent element may be configured from dedicated hardware, or may be implemented by executing a software program suitable for each constituent element. Each constituent element may be implemented by means of a program executor such as a CPU or a processor loading and executing a software program recorded on a recording medium such as a hard disk or semiconductor memory.

[0666]

Thus, the device and the like according to one or more aspects have been described by way of the embodiment, but the aspects understood based on the present disclosure are not limited to the embodiment. The one or more aspects may thus include forms obtained by making various modifications to the above embodiments that can be conceived by those skilled in the art, as well as forms obtained by combining constituent elements in different variations, without materially departing from the spirit of the present disclosure.

[Industrial Applicability]

[0667]

The present disclosure includes aspects that can be applied to, for example, an audio signal processing device, an encoding device, a decoding device, or a terminal or equipment that includes any of these.

[Reference Signs List]

[0668]

1000 three-dimensional sound reproduction system

1001 audio signal processing device (acoustic processing device)
 1002 audio presentation device
 1100, 1120, 1500 encoding device
 1101, 1113 input data
 5 1102 encoder
 1103 encoded data
 1104, 1114, 1404, 1503, 2301b memory
 1110, 1130 decoding device
 1111 audio signal
 10 1112, 1200, 1210 decoder
 1121 transmitter
 1122 transmission signal
 1131 receiver
 1132 reception signal
 15 1201, 1211 spatial information manager
 1202 audio data decoder
 1203, 1213, 1300, 2300, 3300, 4300, 5300 renderer
 1301, 2301 analyzer
 1302, 1314, 2302, 3302, 4302 determiner
 20 1303, 2303, 3303, 4303 reproducer
 1304 threshold value adjuster
 1311 reverberation processor
 1312 early reflection processor
 1313 distance attenuation processor
 25 1315 generator
 1316 binaural processor
 1401 loudspeaker
 1402, 1501 processor
 1403, 1502 communication I/F
 30 1405 sensor
 2301a propagation path detector
 2302a classifier
 2302b first determiner
 2302c, 3302c second determiner
 35 2303a first reproducer

2303b second reproducer
3303c gain setter
4304 renderer pipeline
4304a, 4304a1, 4304a2, 4304a3, 4304a4 first processor
5 4304b, 4304b1, 4304b2 second processor
4305 first gain accumulator
4306 second gain accumulator
5307 changer
5308 disabler

[CLAIMS]

[Claim 1]

An audio signal processing method executed by an audio signal processing device, the audio signal processing method comprising:

- 5 obtaining an audio signal including attribute information
identifying an attribute of the audio signal;
 performing first determination processing and second
determination processing when the attribute identified by the
attribute information included in the audio signal obtained is
10 information indicating an indirect sound, the first determination
processing being processing to determine whether the audio signal
obtained satisfies a first condition, the second determination
processing being processing to determine whether the audio signal
obtained satisfies a second condition different from the first condition;
15 and
 outputting an output signal based on the audio signal obtained,
when the audio signal obtained satisfies the first condition and the
second condition.

20 [Claim 2]

 The audio signal processing method according to claim 1,
wherein
 the indirect sound is a reflected sound.

25 [Claim 3]

 The audio signal processing method according to claim 1 or 2,
wherein
 in the performing, the second determination processing is
skipped when the attribute identified by the attribute information
30 included in the audio signal obtained is information indicating a
predetermined sound different from the indirect sound.

[Claim 4]

 The audio signal processing method according to claim 3,
35 wherein

the predetermined sound is a direct sound.

[Claim 5]

5 The audio signal processing method according to any one of claims 1 to 4, wherein

in the first determination processing, the audio signal obtained is determined to satisfy the first condition when an amplitude value of the audio signal obtained is greater than or equal to a first threshold value, and

10 in the second determination processing, the audio signal obtained is determined to satisfy the second condition when a sound volume ratio is greater than or equal to a second threshold value, the sound volume ratio being a sound volume ratio between a direct sound associated with the indirect sound and the indirect sound when each
15 of the direct sound and the indirect sound arrives at a listening position that is a position at which a listener is present, the second threshold value being a threshold value determined according to a time difference between when the direct sound arrives and when the indirect sound arrives.

20

[Claim 6]

The audio signal processing method according to any one of claims 1 to 5, wherein

the outputting includes:

25 outputting a first output signal based on the audio signal when the attribute is a predetermined sound different from the indirect sound; and

outputting a second output signal based on the audio signal when the attribute is the indirect sound,

30 in the outputting of the second output signal, the second output signal is output by performing diffusion filter processing on the audio signal obtained, the diffusion filter processing being processing that improves realism of the indirect sound by diffusing the indirect sound, and

35 in the outputting of the first output signal, the first output

signal is output without applying the diffusion filter processing to the audio signal.

[Claim 7]

- 5 The audio signal processing method according to any one of claims 1 to 6, wherein
 the second determination processing is performed after the first determination processing.

10 [Claim 8]

- An audio signal processing method executed by an audio signal processing device, the audio signal processing method comprising:
 obtaining an audio signal indicating an indirect sound;
 performing, on the audio signal obtained, one or more first
15 processes, first determination processing, second determination processing, and one or more second processes different from the one or more first processes;
 outputting an output signal based on the audio signal obtained;
 calculating a first adjustment amount by cumulatively adding
20 adjustment amounts each determined by a respective one of the one or more first processes, the adjustment amounts each being an adjustment amount by which an amplitude of the audio signal obtained is to be amplified; and
 calculating a second adjustment amount by cumulatively
25 adding adjustment amounts each determined by a respective one of the one or more second processes, the adjustment amounts each being an adjustment amount by which the amplitude of the audio signal obtained is to be amplified, wherein
 in the first determination processing, the audio signal obtained
30 is determined to satisfy a first condition when the first adjustment amount calculated is greater than or equal to a first threshold value,
 the second determination processing is not performed when the audio signal obtained is determined not to satisfy the first condition,
 the second determination processing is performed when the
35 audio signal obtained is determined to satisfy the first condition,

the audio signal obtained is determined to satisfy a second condition when, in the second determination processing, a value based on the second adjustment amount calculated is greater than or equal to a second threshold value different from the first threshold value, and

the outputting is performed when the audio signal obtained is determined to satisfy the second condition.

[Claim 9]

The audio signal processing method according to claim 8, wherein

the second adjustment amount calculated in the calculating of the second adjustment amount includes an adjustment amount applied when diffusion filter processing has been performed, the diffusion filter processing being processing in which the indirect sound is diffused to improve realism of the indirect sound, the adjustment amount being an amount by which the amplitude of the audio signal obtained is to be amplified,

the value is a ratio between a sound volume of a direct sound associated with the indirect sound and the second adjustment amount calculated in the calculating of the second adjustment amount, and

the second threshold value is determined based on a time difference between when the direct sound arrives and when the indirect sound arrives.

[Claim 10]

The audio signal processing method according to claim 8, wherein

the second adjustment amount calculated in the calculating of the second adjustment amount includes an amount of amplitude adjustment by a sound quality adjustment function set by selection by a listener,

the value is a ratio between a sound volume of a direct sound associated with the indirect sound and the second adjustment amount calculated in the calculating of the second adjustment amount, and

the second threshold value is determined based on a time difference between when the direct sound arrives and when the indirect sound arrives.

5 [Claim 11]

The audio signal processing method according to any one of claims 8 to 10, wherein

the first threshold value is a value related to sound volume.

10 [Claim 12]

The audio signal processing method according to any one of claims 8 to 11, wherein

in the calculating of the second adjustment amount:

15 a parameter indicating the adjustment amount determined by each of the one or more second processes is obtained, the adjustment amount being an amount by which the amplitude of the audio signal obtained is to be amplified, and

the second adjustment amount is calculated based on a plurality of parameters obtained, the plurality of parameters each
20 being the parameter.

[Claim 13]

The audio signal processing method according to any one of claims 8 to 12, comprising:

25 setting the first threshold value; and

performing disabling processing to disable the second determination processing, wherein

in the first determination processing, the audio signal obtained is determined to satisfy the first condition when the first adjustment
30 amount calculated is greater than or equal to the first threshold value set, and

when the disabling processing is performed, the outputting is performed upon determining that the audio signal obtained satisfies the first condition.

35

[Claim 14]

The audio signal processing method according to claim 13, wherein

5 a storage region storing a value indicating the first threshold value set in the setting and a storage region storing a signal instructing performance of the disabling processing are adjacent regions.

[Claim 15]

10 The audio signal processing method according to claim 13, wherein

the first threshold value set when the disabling processing has been performed is greater than the first threshold value set when the disabling processing has not been performed.

15

[Claim 16]

A computer program for causing a computer to execute the audio signal processing method according to any one of claims 1 to 15.

20

[Claim 17]

An audio signal processing device comprising:

an obtainer that obtains an audio signal including attribute information identifying an attribute of the audio signal;

25 a determiner that performs first determination processing and second determination processing when the attribute identified by the attribute information included in the audio signal obtained is information indicating an indirect sound, the first determination processing being processing to determine whether the audio signal obtained satisfies a first condition, the second determination processing being processing to determine whether the audio signal
30 obtained satisfies a second condition different from the first condition; and

35 a reproducer that outputs an output signal based on the audio signal obtained, when the audio signal obtained satisfies the first

condition and the second condition.