

DESCRIPTION
INFORMATION PROCESSING APPARATUS AND INFORMATION PROCESSING
METHOD

5 TECHNICAL FIELD
[0001]

The present disclosure relates to an information processing apparatus and an information processing method, and more particularly, to an information processing apparatus and an information processing method which are capable of
10 improving the efficiency of acquiring a predetermined type of audio data among a plurality of types of audio data.

BACKGROUND ART
15 [0002]

One of the recent most popular streaming services is the over-the-top video (OTT-V) via the Internet. The moving picture experts group phase-dynamic adaptive streaming over HTTP (MPEG-DASH) is widely used as its underlying technology
20 (see, for example, Non-Patent Document 1).

[0003]

In MPEG-DASH, a delivery server prepares a group of video data having different screen sizes and coding rates for one video content item, and a playback terminal requests a
25 group of video data having an optimal screen size and coding rate depending on transmission line conditions, thus adaptive streaming delivery is achieved.

CITATION LIST
30 NON-PATENT DOCUMENT
[0004]

Non-Patent Document 1: MPEG-DASH (Dynamic Adaptive Streaming over

HTTP) (URL:<http://mpeg.chiariglione.org/standards/mpeg-dash/media-presentation-description-and-segment-formats/text>

5 -isoiec-23009-12012-dam-1)

SUMMARY OF THE INVENTION

PROBLEMS TO BE SOLVED BY THE INVENTION

[0005]

10 However, no consideration is given to an improvement in efficiency of acquiring a predetermined type of audio data among a plurality of types of audio data of a video content.

[0006]

15 The present disclosure has been made in view of the above-mentioned circumstances and is capable of improving the efficiency of acquiring a predetermined type of audio data among a plurality of types of audio data.

SOLUTIONS TO PROBLEMS

20 [0007]

 An information processing apparatus according to a first aspect of the present disclosure is an information processing apparatus including an acquisition unit that acquires audio data in a predetermined track of a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types and the tracks are arranged.

[0008]

 An information processing method according to the first aspect of the present disclosure corresponds to the information processing apparatus according to the first aspect of the present disclosure.

30

[0009]

In the first aspect of the present disclosure, audio data of the predetermined track in the file in which the plurality of types of audio data are divided into the plurality of tracks depending on the types and the tracks
5 arranged is acquired.

[0010]

An information processing apparatus according to a second aspect of the present disclosure is an information
10 processing apparatus including a generation unit that generates a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types and the tracks are arranged.

[0011]

15 An information processing method according to the second aspect of the present disclosure corresponds to the information processing apparatus according to the second aspect of the present disclosure.

[0012]

20 In the second aspect of the present disclosure, the file in which the plurality of types of audio data are divided into the plurality of tracks depending on the types and the tracks are arranged is generated.

[0013]

25 Note that the information processing apparatuses according to the first and second aspects can be implemented by causing a computer to execute a program.

[0014]

30 Further, in order to achieve the information processing apparatuses according to the first and second aspects, a program executed by a computer can be provided by transmitting

the program via a transmission medium, or by recording the program in a recording medium.

EFFECTS OF THE INVENTION

5 [0015]

According to a first aspect of the present disclosure, audio data can be acquired. Further, according to the first aspect of the present disclosure, a predetermined type of audio data among a plurality of types of audio data can be
10 acquired efficiently.

[0016]

According to the second aspect of the present disclosure, a file can be generated. Further, according to the second aspect of the present disclosure, a file that
15 improves the efficiency of acquiring a predetermined type of audio data among a plurality of types of audio data can be generated.

BRIEF DESCRIPTION OF DRAWINGS

20 [0017]

Fig. 1 is a diagram illustrating an outline of a first example of an information processing system to which the present disclosure is applied.

Fig. 2 is a diagram showing an example of a file.

25 Fig. 3 is a diagram illustrating an object.

Fig. 4 is a diagram illustrating object position information.

Fig. 5 is a diagram illustrating image frame size information.

30 Fig. 6 is a diagram showing a structure of an MPD file.

Fig. 7 is a diagram showing a relationship among

"Period", "Representation", and "Segment".

Fig. 8 is a diagram showing a hierarchical structure of an MPD file.

Fig. 9 is a diagram showing a relationship between a
5 structure of an MPD file and a time axis.

Fig. 10 is a diagram illustrating an exemplary description of the MPD file.

Fig. 11 is a block diagram showing a configuration example of a file generation device.

10 Fig. 12 is a flowchart illustrating file generation process of the file generation device.

Fig. 13 is a block diagram showing a configuration example of a streaming playback unit.

15 Fig. 14 is a flowchart illustrating a streaming playback process of the streaming playback unit.

Fig. 15 is a diagram illustrating an exemplary description of the MPD file.

Fig. 16 is a diagram illustrating another exemplary description of the MPD file.

20 Fig. 17 is a diagram showing an arrangement example of an audio stream.

Fig. 18 is a diagram showing an exemplary description of gsix.

25 Fig. 19 is a diagram showing an example of information indicating a correspondence relation between a sample group entry and object ID.

Fig. 20 is a diagram showing an exemplary description of AudioObjectSampleGroupEntry.

30 Fig. 21 is a diagram showing an exemplary description of a type assignment box.

Fig. 22 is a diagram illustrating an outline of a second

example of the information processing system to which the present disclosure is applied.

Fig. 23 is a block diagram showing a configuration example of the streaming playback unit of the information processing system to which the present disclosure is applied.

Fig. 24 is a diagram illustrating a method of determining a position of an object.

Fig. 25 is a diagram illustrating a method of determining a position of an object.

Fig. 26 is a diagram illustrating a method of determining a position of an object.

Fig. 27 is a diagram showing a relationship between a horizontal angle θ_{Ai} and a horizontal angle θ_{Ai}' .

Fig. 28 is a flowchart illustrating the streaming playback process of the streaming playback unit shown in Fig. 23.

Fig. 29 is a flowchart illustrating details of a position determination process shown in Fig. 28.

Fig. 30 is a flowchart illustrating details of a horizontal angle θ_{Ai}' estimation process shown in Fig. 29.

Fig. 31 is a diagram illustrating an outline of tracks of a 3D audio file format of MP4.

Fig. 32 is a diagram showing a structure of a moov box.

Fig. 33 is a diagram illustrating an outline of tracks according to a first embodiment to which the present disclosure is applied.

Fig. 34 is a diagram showing an exemplary syntax of a sample entry of a base track shown in Fig. 33.

Fig. 35 is a diagram showing an exemplary syntax of a sample entry of a channel audio track shown in Fig. 33.

Fig. 36 is a diagram showing an exemplary syntax of a

sample entry of an object audio track shown in Fig. 33.

Fig. 37 is a diagram showing an exemplary syntax of a sample entry of an HOA audio track shown in Fig. 33.

Fig. 38 is a diagram showing an exemplary syntax of a
5 sample entry of an object metadata track shown in Fig. 33.

Fig. 39 is a diagram showing a first example of a segment structure.

Fig. 40 is a diagram showing a second example of the segment structure.

10 Fig. 41 is a diagram showing an exemplary description of a level assignment box.

Fig. 42 is a diagram showing an exemplary description of the MPD file in the first embodiment to which the present disclosure is applied.

15 Fig. 43 is a diagram showing a definition of Essential Property.

Fig. 44 is a diagram illustrating an outline of an information processing system in the first embodiment to which the present disclosure is applied.

20 Fig. 45 is a block diagram showing a configuration example of a file generation device shown in Fig. 44.

Fig. 46 is a flowchart illustrating a file generation process of the file generation device shown in Fig. 45.

Fig. 47 is a block diagram showing a configuration
25 example of a streaming playback unit implemented by a video playback terminal shown in Fig. 44.

Fig. 48 is a flowchart illustrating a channel audio playback process of the streaming playback unit shown in Fig. 47.

30 Fig. 49 is a flowchart illustrating an object specifying process of the streaming playback unit shown in

Fig. 47.

Fig. 50 is a flowchart illustrating a specific object audio playback process of the streaming playback unit shown in Fig. 47.

5 Fig. 51 is a diagram illustrating an outline of tracks in a second embodiment to which the present disclosure is applied.

Fig. 52 is a diagram showing an exemplary syntax of a sample entry of a base track shown in Fig. 51.

10 Fig. 53 is a diagram showing a structure of a base sample.

Fig. 54 is a diagram showing an exemplary syntax of a base sample.

15 Fig. 55 is a diagram showing an example of data of an extractor.

Fig. 56 is a diagram illustrating an outline of tracks in a third embodiment to which the present disclosure is applied.

20 Fig. 57 is a diagram illustrating an outline of tracks in a fourth embodiment to which the present disclosure is applied.

Fig. 58 is a diagram showing an exemplary description of an MPD file in the fourth embodiment to which the present disclosure is applied.

25 Fig. 59 is a diagram illustrating an outline of an information processing system in the fourth embodiment to which the present disclosure is applied.

Fig. 60 is a block diagram showing a configuration example of the file generation device shown in Fig. 59.

30 Fig. 61 is a flowchart illustrating a file generation process of the file generation device shown in Fig. 60.

Fig. 62 is a block diagram showing a configuration example of a streaming playback unit implemented by a video playback terminal shown in Fig. 59.

Fig. 63 is a flowchart illustrating an example of a
5 channel audio playback process of the streaming playback unit shown in Fig. 62.

Fig. 64 is a flowchart illustrating a first example of an object audio playback process of the streaming playback unit shown in Fig. 62.

10 Fig. 65 is a flowchart illustrating a second example of the object audio playback process of the streaming playback unit shown in Fig. 62.

Fig. 66 is a flowchart illustrating a third example of the object audio playback process of the streaming playback
15 unit shown in Fig. 62.

Fig. 67 is a diagram showing an example of an object selected on the basis of a priority.

Fig. 68 is a diagram illustrating an outline of tracks in a fifth embodiment to which the present disclosure is
20 applied.

Fig. 69 is a diagram illustrating an outline of tracks in a sixth embodiment to which the present disclosure is applied.

Fig. 70 is a diagram showing a hierarchical structure
25 of 3D audio.

Fig. 71 is a diagram illustrating a first example of a Web server process.

Fig. 72 is a flowchart illustrating a track division process of a Web server.

30 Fig. 73 is a diagram illustrating a first example of a process of an audio decoding processing unit.

Fig. 74 is a flowchart illustrating details of a first example of a decoding process of the audio decoding processing unit.

Fig. 75 is a diagram illustrating a second example of a process of the audio decoding processing unit.

Fig. 76 is a flowchart illustrating details of the second example of the decoding process of the audio decoding processing unit.

Fig. 77 is a diagram illustrating a second example of the Web server process.

Fig. 78 is a diagram illustrating a third example of the process of the audio decoding processing unit.

Fig. 79 is a flowchart illustrating details of the third example of the decoding process of the audio decoding processing unit.

Fig. 80 is a diagram showing a second example of syntax of Config information disposed in a base sample.

Fig. 81 is an exemplary syntax of Config information for Ext element shown in Fig. 80.

Fig. 82 is a diagram showing an exemplary syntax of Config information for Extractor shown in Fig. 81.

Fig. 83 is a diagram showing a second example of syntax of data of a frame unit disposed in a base sample.

Fig. 84 is a diagram showing an exemplary syntax of data of Extractor shown in Fig. 83.

Fig. 85 is a diagram showing a third example of syntax of Config information disposed in a base sample.

Fig. 86 is a diagram showing a third example of syntax of data of a frame unit disposed in a base sample.

Fig. 87 is a diagram showing a configuration example of an audio stream in a seventh embodiment of the information

processing system to which the present disclosure is applied.

Fig. 88 is a diagram illustrating an outline of tracks in the seventh embodiment.

Fig. 89 is a flowchart illustrating a file generation
5 process in the seventh embodiment.

Fig. 90 is a flowchart illustrating an audio playback process in the seventh embodiment.

Fig. 91 is a diagram illustrating an outline of tracks in an eighth embodiment of the information processing system
10 to which the present disclosure is applied.

Fig. 92 is a diagram showing a configuration example of an audio file.

Fig. 93 is a diagram showing another configuration example of the audio file.

15 Fig. 94 is a diagram showing still another configuration example of the audio file.

Fig. 95 is a block diagram showing a configuration example of hardware of a computer.

20 MODE FOR CARRYING OUT THE INVENTION

[0018]

Modes for carrying out the present disclosure (hereinafter referred to as embodiments) will be described below in the following order.

25 0. Premise of the present disclosure (Figs. 1 to 30)

1. First embodiment (Figs. 31 to 50)

2. Second embodiment (Figs. 51 to 55)

3. Third embodiment (Fig. 56)

4. Fourth embodiment (Figs. 57 to 67)

30 5. Fifth embodiment (Fig. 68)

6. Sixth embodiment (Fig. 69)

7. Explanation of Hierarchical Structure of 3D Audio (Fig. 70)

8. Explanation of First Example of Web Server Process (Figs. 71 and 72)

5 9. Explanation of First Example of Process of Audio Decoding Processing Unit (Figs. 73 and 74)

10. Explanation of Second Example of Process of Audio Decoding Processing Unit (Figs. 75 and 76)

10 11. Explanation of Second Example of Web Server Process (Fig. 77)

12. Explanation of Third Example of Process of Audio Decoding Processing Unit (Figs. 78 and 79)

13. Second Example of Syntax of Base Sample (Figs. 80 to 84)

14. Third Example of Syntax of Base Sample (Figs. 85 and 86)

15 15. Seventh embodiment (Figs. 87 to 90)

16. Eighth embodiment (Figs. 91 to 94)

17. Ninth embodiment (Fig. 95)

[0019]

<Premise of Present Disclosure>

20 (Outline of First Example of Information Processing System)

Fig. 1 is a diagram illustrating an outline of a first example of an information processing system to which the present disclosure is applied.

[0020]

25 An information processing system 10 shown in Fig. 1 has a configuration in which a Web server 12, which is connected to a file generation device 11, and a video playback terminal 14 are connected via the Internet 13.

[0021]

30 In the information processing system 10, the Web server 12 delivers (tiled streaming) image data of video content to

the video playback terminal 14 in units of tiles by a method compliant with MPEG-DASH.

[0022]

Specifically, the file generation device 11 acquires
5 the image data of video content and encodes the image data
in units of tiles to generate a video stream. The file
generation device 11 processes the video stream of each tile
into a file format at time intervals ranging from several
seconds to approximately ten seconds, which is called a
10 segment. The file generation device 11 uploads the resulting
image file of each tile to the Web server 12.

[0023]

Further, the file generation device 11 acquires audio
data of video content for each object (to be described in
15 detail later) and encodes the image data in units of objects
to generate an audio stream. The file generation device 11
processes the audio stream of each object into a file format
in units of segments, and uploads the resulting audio file
of each object to the Web server 12.

20 [0024]

Note that the object is a sound source. The audio data
of each object is acquired through a microphone or the like
attached to the object. The object may be an object such as
a fixed microphone stand, or may be a moving body such as a
25 person.

[0025]

The file generation device 11 encodes audio metadata
including object position information (audio position
information) indicating the position of each object (the
30 position at which audio data is acquired) and an object ID
that is an ID unique to the object. The file generation device

11 processes the encoded data obtained by encoding the audio metadata into a file format in unit of segments, and uploads the resulting audio metafile to the Web server 12.

[0026]

5 Further, the file generation device 11 generates a media presentation description (MPD) file (control information) managing an image file and audio file and including image frame size information that indicates the frame size of images of video content and tile position
10 information that indicates the position of each tile on an image. The file generation device 11 uploads the MPD file to the Web server 12.

[0027]

 The Web server 12 stores the image file, audio file,
15 audio metafile, and MPD file which are uploaded from the file generation device 11.

[0028]

 In the example shown in Fig. 1, the Web server 12 stores a segment group of a plurality of segments composed of image
20 files of a tile with a tile ID "1" and a segment group of a plurality of segments composed of image files of a tile with a tile ID "2". The Web server 12 also stores a segment group of a plurality of segments composed of audio files of an object with an object ID "1" and a segment group of a plurality of
25 segments composed of audio files of an object with an object ID "2". Although not shown, a segment group composed of audio metafiles is similarly stored.

[0029]

 Note that a file with a tile ID of *i* is hereinafter
30 referred to as "tile #*i*", and an object with an object ID of *i* is hereinafter referred to as "object #*i*".

[0030]

The Web server 12 functions as a transmitter and transmits the stored image file, audio file, audio metafile, MPD file, and the like to the video playback terminal 14 in response to a request from the video playback terminal 14.

[0031]

The video playback terminal 14 executes, for example, software for control of streaming data (hereinafter referred to as control software) 21, video playback software 22, and client software for hypertext transfer protocol (HTTP) access (hereinafter referred to as access software) 23.

[0032]

The control software 21 is software to control data delivered via streaming from the Web server 12. Specifically, the control software 21 allows the video playback terminal 14 to acquire the MPD file from the Web server 12.

[0033]

Further, the control software 21 specifies a tile in a display area on the basis of the display area that is an area in an image used to display a video content indicated by the video playback software 22 and the tile position information included in the MPD file. The control software 21 instructs the access software 23 to issue a request to transmit an image file of the specified tile.

[0034]

Further, the control software 21 instructs the access software 23 to issue a request to transmit the audio metafile. The control software 21 specifies an object corresponding to an image in the display area, on the basis of the display area, the image frame size information included in the MPD file, and the object position information included in the audio

metafile. The control software 21 instructs the access software 23 to issue a request to transmit an audio file of the specified object.

[0035]

5 The video playback software 22 is software to play back the image file and audio file acquired from the Web server 12. Specifically, when a user specifies a display area, the video playback software 22 indicates the specified display area to the control software 21. The video playback software
10 22 decodes the image file and audio file acquired from the Web server 12 in response to the indication, and the video playback software 22 synthesizes and outputs the decoded files.

[0036]

15 The access software 23 is software to control communication with the Web server 12 via the Internet 13 using HTTP. Specifically, the access software 23 allows the video playback terminal 14 to transmit the request to transmit the image file, audio file, and audio metafile in response to the
20 instruction from the control software 21. Further, the access software 23 allows the video playback terminal 14 to receive the image file, audio file, and audio metafile transmitted from the Web server 12 in response to the transmission request.

25 [0037]

(Example of Tile)

Fig. 2 is a diagram showing an example of a tile.

[0038]

As shown in Fig. 2, an image of video content is divided
30 into a plurality of tiles. A tile ID that is a sequential number starting from 1 is assigned to each tile. In the

example shown in Fig. 2, an image of video content is divided into four tiles #1 to #4.

[0039]

(Explanation of Object)

5 Fig. 3 is a diagram illustrating an object.

[0040]

 The example of Fig. 3 illustrates eight audio objects in an image acquired as an audio of video content. An object ID that is a sequential number starting from 1 is assigned to each object. Objects #1 to #5 are moving bodies, and
10 objects #6 to #8 are fixed material bodies. Further, in the example of Fig. 3, the image of video content is divided into 7 (width) × 5 (height) tiles.

[0041]

15 In this case, as shown in Fig. 3, when the user specifies a display area 31 composed of 3 (width) × 2 (height) tiles, the display area 31 includes only objects #1, #2, and #6. Thus, the video playback terminal 14 acquires and plays back, for example, only the audio files of the objects #1, #2, and #6
20 from the Web server 12.

[0042]

 The objects in the display area 31 can be specified on the basis of the image frame size information and the object position information as described below.

25 [0043]

(Explanation of Object Position Information)

 Fig. 4 is a diagram illustrating the object position information.

[0044]

30 As shown in Fig. 4, the object position information includes a horizontal angle θ_A ($-180^\circ \leq \theta_A \leq 180^\circ$), a vertical

angle γ_A ($-90^\circ \leq \gamma_A \leq 90^\circ$), and a distance r_A ($0 < r_A$) of an object 40. The horizontal angle θ_A is the angle in the horizontal direction formed by the straight line connecting the object 40 and an origin O and a YZ plane, for example,
 5 when a shooting position in the center of an image may be set to the origin (base point) O; the horizontal direction of the image is set to an X direction; the vertical direction of the image is set to a Y direction; and the depth direction perpendicular to the XY plane is set to a Z direction. The
 10 vertical angle γ_A is the angle in the vertical direction formed by the straight line connecting the object 40 and the origin O and the XZ plane. The distance r_A is the distance between the object 40 and the origin O.

[0045]

15 Furthermore, assume herein that the angle of the left and up rotation is set to a positive angle, and the angle of the right and down rotation is set to a negative angle.

[0046]

(Explanation of Image Frame Size Information)

20 Fig. 5 is a diagram illustrating the image frame size information.

[0047]

As shown in Fig. 5, the image frame size information includes a horizontal angle θ_{v1} of the left end, a horizontal
 25 angle θ_{v2} of the right end, a vertical angle γ_{v1} of the upper end, a vertical angle γ_{v2} of the lower end, and a distance r_v in the image frame.

[0048]

The horizontal angle θ_{v1} is the angle in the horizontal
 30 direction formed by the straight line connecting the left end of an image frame and the origin O and the YZ plane, for example,

when a shooting position in the center of an image is set to the origin O; the horizontal direction of the image is set to the X direction; the vertical direction of the image is set to the Y direction; and the depth direction perpendicular to the XY plane is set to the Z direction. The horizontal angle θ_{v2} is the angle in the horizontal direction formed by the straight line connecting the right end of an image frame and the origin O and the YZ plane. Thus, an angle obtained by combining the horizontal angle θ_{v1} and the horizontal angle θ_{v2} is a horizontal angle of view.

[0049]

The vertical angle γ_{v1} is the angle formed by the XZ plane and the straight line connecting the upper end of the image frame and the origin O, and the vertical angle γ_{v2} is the angle formed by the XZ plane and the straight line connecting the lower end of the image frame and the origin O. An angle obtained by combining the vertical angles γ_{v1} and γ_{v2} becomes a vertical angle of view. The distance r_v is the distance between the origin O and the image plane.

[0050]

As described above, the object position information represents the positional relationship between the object and the origin O, and the image frame size information represents the positional relationship between the image frame and the origin O. Thus, it is possible to detect (recognize) the position of each object on the image on the basis of the object position information and the image frame size information. As a result, it is possible to specify an object in the display area.

[0051]

(Explanation of Structure of MPD File)

Fig. 6 is a diagram illustrating the structure of an MPD file.

[0052]

In the analysis (parsing) of an MPD file, the video
5 playback terminal 14 selects an optimum one among attributes
of "Representation" included in "Period" of the MPD file
(Media Presentation in Fig. 6).

[0053]

The video playback terminal 14 acquires a file by
10 referring to a uniform resource locator (URL) or the like of
"Initialization Segment" at the head of the selected
"Representation", and processes the acquired file. Then,
the video playback terminal 14 acquires a file by referring
to the URL or the like of the subsequent "Media Segment", and
15 plays back the acquired file.

[0054]

Note that in the MPD file, the relationship among
"Period", "Representation", and "Segment" becomes as shown
in Fig. 7. In other words, a single video content item can
20 be managed in a longer time unit than the segment by "Period",
and can be managed in units of segments by "Segment" in each
"Period". Further, in each "Period", it is possible to manage
the video content in units of stream attributes by
"Representation".

25 [0055]

Thus, the MPD file has a hierarchical structure shown
in Fig. 8, starting from the "Period". Further, the structure
of the MPD file arranged on the time axis becomes the
configuration as shown in Fig. 9. As is clear from Fig. 9,
30 there are a plurality of "Representation" elements in the same
segment. The video playback terminal 14 selects any one from

among these elements adaptively, and thus it is possible to acquire an image file and audio file in the display area selected by the user and to play back the acquired file.

[0056]

5 (Explanation of Description of MPD File)

Fig. 10 is a diagram illustrating the description of an MPD file.

[0057]

As described above, in the information processing
10 system 10, the image frame size information is included in the MPD file to allow an object in the display area to be specified by the video playback terminal 14. As shown in Fig. 10, Scheme (urn:mpeg:DASH:viewingAngle:2013) used to define new image frame size information (viewing angle) is extended
15 by utilizing a DescriptorType element of Viewpoint, and thus the image frame size information is arranged in an "Adaptation Set" for audio and an "Adaptation Set" for image. The image frame size information may be arranged only in the "Adaptation Set" for image.

20 [0058]

Further, the "Representation" for audio metafile is described in the "Adaptation Set" for audio of the MPD file. A URL or the like as information for specifying the audio metafile (audiometadata.mp4) is described in "Segment" of the
25 "Representation". In this case, it is described that the file to be specified in "Segment" is the audio metafile (objectaudiometadata) by utilizing Role element.

[0059]

The "Representation" for audio file of each object is
30 also described in "Adaptation Set" for audio of the MPD file. A URL or the like as information for specifying the audio file

(audioObj1.mp4, audioObj5.mp4) of each object is described in "Segment" of the "Representation". In this case, object IDs (1 and 5) of the objects corresponding to the audio file are also described by extending Viewpoint.

5 [0060]

Note that, although not shown, the tile position information is arranged in the "Adaptation Set" for image.

[0061]

(Configuration Example of File Generation Device)

10 Fig. 11 is a block diagram showing a configuration example of the file generation device 11 shown in Fig. 1.

[0062]

The file generation device 11 shown in Fig. 11 includes a screen split processing unit 51, an image coding processing unit 52, an image file generation unit 53, an image information generation unit 54, an audio coding processing unit 55, an audio file generation unit 56, an MPD generation unit 57, and a server upload processing unit 58.

[0063]

20 The screen split processing unit 51 of the file generation device 11 splits image data of video content input from the outside into tile units. The screen split processing unit 51 supplies the image information generation unit 54 with the tile position information. Further, the screen split processing unit 51 supplies the image coding processing unit 52 with the image data configured in units of tiles.

[0064]

30 The image coding processing unit 52 encodes the image data, which is configured in units of tiles and is supplied from the screen split processing unit 51, for each tile to generate a video stream. The image coding processing unit

52 supplies the image file generation unit 53 with the video stream of each tile.

[0065]

5 The image file generation unit 53 processes the video stream of each tile supplied from the image coding processing unit 52 into a file format in units of segments and supplies the MPD generation unit 57 with the resulting image file of each tile.

[0066]

10 The image information generation unit 54 supplies the MPD generation unit 57 with the tile position information supplied from the screen split processing unit 51 and with the image frame size information input from the outside as image information.

15 [0067]

The audio coding processing unit 55 encodes audio data, which is configured in units of objects of video content input from the outside, for each object, and generates an audio stream. Further, the audio coding processing unit 55 encodes
20 the object position information of each object input from the outside and the audio metadata including the object ID and the like to generate encoded data. The audio coding processing unit 55 supplies the audio file generation unit 56 with the audio stream of each object and the encoded data
25 of the audio metadata.

[0068]

The audio file generation unit 56 functions as an audio file generation unit, processes the audio stream of each object supplied from the audio coding processing unit 55 into
30 a file format in units of segments, and supplies the MPD generation unit 57 with the resulting audio file of each

object.

[0069]

Further, the audio file generation unit 56 functions as a metafile generation unit, processes the encoded data of audio metadata supplied from the audio coding processing unit 55 into a file format in units of segments, and supplies the MPD generation unit 57 with the resulting audio metafile.

[0070]

The MPD generation unit 57 determines the URL or the like of the Web server 12 for storing the image file of each tile supplied from the image file generation unit 53. Further, the MPD generation unit 57 determines the URL or the like of the Web server 12 for storing the audio file of each object and the audio metafile which are supplied from the audio file generation unit 56.

[0071]

The MPD generation unit 57 arranges the image information supplied from the image information generation unit 54 in "AdaptationSet" for an image of the MPD file. Further, the MPD generation unit 57 arranges the image frame size information among the pieces of image information in "AdaptationSet" for audio of the MPD file. The MPD generation unit 57 arranges the URL or the like of the image file of each tile in "Segment" of "Representation" for the image file of the tile.

[0072]

The MPD generation unit 57 arranges the URL or the like of the audio file of each object in "Segment" of "Representation" for audio file of the object. Further, the MPD generation unit 57 functions as an information generation unit, and arranges a URL or the like as information for

specifying an audio metafile in "Segment" of "Representation" for audio metafile. The MPD generation unit 57 supplies the server upload processing unit 58 with the MPD file in which various types of information are arranged as described above,
5 the image file, the audio file, and the audio metafile.

[0073]

The server upload processing unit 58 uploads the image file of each tile, the audio file of each object, the audio metafile, and the MPD file, which are supplied from the MPD
10 generation unit 57, to the Web server 12.

[0074]

(Explanation of Process of File Generation Device)

Fig. 12 is a flowchart illustrating a file generation process of the file generation device 11 shown in Fig. 11.

15 [0075]

In step S11 of Fig. 12, the screen split processing unit 51 of the file generation device 11 splits image data of video content input from the outside into tile units. The screen split processing unit 51 supplies the image information
20 generation unit 54 with the tile position information. Further, the screen split processing unit 51 supplies the image coding processing unit 52 with the image data configured in units of tiles.

[0076]

25 In step S12, the image coding processing unit 52 encodes the image data, which is configured in units of tiles and is supplied from the screen split processing unit 51, for each tile to generate a video stream of each tile. The image coding processing unit 52 supplies the image file generation unit
30 53 with the video stream of each tile.

[0077]

In step S13, the image file generation unit 53 processes the video stream of each tile supplied from the image coding processing unit 52 into a file format in units of segments to generate an image file of each tile. The image file
5 generation unit 53 supplies the MPD generation unit 57 with the image file of each tile.

[0078]

In step S14, the image information generation unit 54 acquires the image frame size information from the outside.

10 In step S15, the image information generation unit 54 generates image information including the tile position information supplied from the screen split processing unit 51 and the image frame size information, and supplies the MPD generation unit 57 with the image information.

15 [0079]

In step S16, the audio coding processing unit 55 encodes audio data, which is configured in units of objects of video content input from the outside, for each object, and generates an audio stream of each object. Further, the audio coding
20 processing unit 55 encodes the object position information of each object input from the outside and the audio metadata including the object ID to generate encoded data. The audio coding processing unit 55 supplies the audio file generation unit 56 with the audio stream of each object and the encoded
25 data of the audio metadata.

[0080]

In step S17, the audio file generation unit 56 processes the audio stream of each object supplied from the audio coding processing unit 55 into a file format in units of segments to generate an audio file of each object. Further, the audio
30 file generation unit 56 processes the encoded data of the

audio metadata supplied from the audio coding processing unit 55 into a file format in units of segments to generate an audio metafile. The audio file generation unit 56 supplies the MPD generation unit 57 with the audio file of each object and the

5 audio metafile.
[0081]

In step S18, the MPD generation unit 57 generates an MPD file including the image information supplied from the image information generation unit 54, the URL or the like of each file. The MPD generation unit 57 supplies the server upload processing unit 58 with the MPD file, the image file of each tile, the audio file of each object, and the audio metafile.

[0082]

15 In step S19, the server upload processing unit 58 uploads the image file of each tile, the audio file of each object, the audio metafile, and the MPD file, which are supplied from the MPD generation unit 57, to the Web server 12. Then, the process is terminated.

20 [0083]

(Functional Configuration Example of Video Playback Terminal)

Fig. 13 is a block diagram showing a configuration example of the streaming playback unit which is implemented in such a manner that the video playback terminal 14 shown in Fig. 1 executes the control software 21, the video playback software 22, and the access software 23.

[0084]

A streaming playback unit 90 shown in Fig. 13 includes an MPD acquisition unit 91, an MPD processing unit 92, a metafile acquisition unit 93, an audio selection unit 94, an

30

audio file acquisition unit 95, an audio decoding processing unit 96, an audio synthesis processing unit 97, an image selection unit 98, an image file acquisition unit 99, an image decoding processing unit 100, and an image synthesis processing unit 101.

[0085]

The MPD acquisition unit 91 of the streaming playback unit 90 functions as a receiver, acquires an MPD file from the Web server 12, and supplies the MPD processing unit 92 with the MPD file.

[0086]

The MPD processing unit 92 extracts information such as a URL, which is described in "Segment" for audio metafile, from the MPD file supplied from the MPD acquisition unit 91, and supplies the metafile acquisition unit 93 with the extracted information. Further, the MPD processing unit 92 extracts image frame size information, which is described in "AdaptationSet" for image, from the MPD file, and supplies the audio selection unit 94 with the extracted information. The MPD processing unit 92 extracts information such as a URL, which is described in "Segment" for audio file of the object requested from the audio selection unit 94, from the MPD file, and supplies the audio selection unit 94 with the extracted information.

[0087]

The MPD processing unit 92 extracts the tile position information described in "AdaptationSet" for image from the MPD file and supplies the image selection unit 98 with the extracted information. The MPD processing unit 92 extracts information such as a URL, which is described in "Segment" for the image file of the tile requested from the image

selection unit 98, from the MPD file, and supplies the image selection unit 98 with the extracted information.

[0088]

On the basis of the information such as a URL supplied
5 from the MPD processing unit 92, the metafile acquisition unit 93 requests the Web server 12 to send an audio metafile specified by the URL, and acquires the audio metafile. The metafile acquisition unit 93 supplies the audio selection unit 94 with object position information included in the audio
10 metafile.

[0089]

The audio selection unit 94 functions as a position determination unit, and calculates a position of each object on the image on the basis of the image frame size information
15 supplied from the MPD processing unit 92 and the object position information supplied from the metafile acquisition unit 93. The audio selection unit 94 selects an object in the display area designated by the user on the basis of the position of each object on the image. The audio selection
20 unit 94 requests the MPD processing unit 92 to send information such as the URL of the audio file of the selected object. The audio selection unit 94 supplies the audio file acquisition unit 95 with the information such as the URL supplied from the MPD processing unit 92 in response to the
25 request.

[0090]

The audio file acquisition unit 95 functions as a receiver. On the basis of the information such as a URL supplied from the audio selection unit 94, the audio file
30 acquisition unit 95 requests the Web server 12 to send an audio file, which is specified by the URL and configured in units

of objects, and acquires the audio file. The audio file acquisition unit 95 supplies the audio decoding processing unit 96 with the acquired audio file in units of objects.
[0091]

5 The audio decoding processing unit 96 decodes an audio stream included in the audio file, which is supplied from the audio file acquisition unit 95 and configured in units of objects, to generate audio data in units of objects. The audio decoding processing unit 96 supplies the audio
10 synthesis processing unit 97 with the audio data in units of objects.
[0092]

 The audio synthesis processing unit 97 synthesizes the audio data, which is supplied from the audio decoding
15 processing unit 96 and configured in units of objects, and outputs the synthesized data.
[0093]

 The image selection unit 98 selects a tile in the display area designated by the user on the basis of the tile
20 position information supplied from the MPD processing unit 92. The image selection unit 98 requests the MPD processing unit 92 to send information such as a URL for the image file of the selected tile. The image selection unit 98 supplies the image file acquisition unit 99 with the information such
25 as a URL supplied from the MPD processing unit 92 in response to the request.
[0094]

 On the basis of the information such as a URL supplied from the image selection unit 98, the image file acquisition
30 unit 99 requests the Web server 12 to send an image file, which is specified by the URL and configured in units of tiles, and

acquires the image file. The image file acquisition unit 99 supplies the image decoding processing unit 100 with the acquired image file in units of tiles.

[0095]

5 The image decoding processing unit 100 decodes a video stream included in the image file, which is supplied from the image file acquisition unit 99 and configured in units of tiles, to generate image data in units of tiles. The image decoding processing unit 100 supplies the image synthesis processing unit 101 with the image data in units of tiles.

[0096]

 The image synthesis processing unit 101 synthesizes the image data, which is supplied from the image decoding processing unit 100 and configured in units of tiles, and
15 outputs the synthesized data.

[0097]

(Explanation of Process of Moving Image Playback Terminal)

 Fig. 14 is a flowchart illustrating a streaming playback process of the streaming playback unit 90 (Fig. 13)
20 of the video playback terminal 14.

[0098]

 In step S31 of Fig. 14, the MPD acquisition unit 91 of the streaming playback unit 90 acquires the MPD file from the Web server 12 and supplies the MPD processing unit 92 with
25 the MPD file.

[0099]

 In step S32, the MPD processing unit 92 acquires the image frame size information and the tile position information, which are described in "AdaptationSet" for image,
30 from the MPD file supplied from the MPD acquisition unit 91. The MPD processing unit 92 supplies the audio selection unit

94 with the image frame size information and supplies the image selection unit 98 with the tile position information. Further, the MPD processing unit 92 extracts information such as a URL described in "Segment" for audio metafile and
5 supplies the metafile acquisition unit 93 with the extracted information.

[0100]

In step S33, on the basis of the information such as a URL supplied from the MPD processing unit 92, the metafile
10 acquisition unit 93 requests the Web server 12 to send an audio metafile specified by the URL, and acquires the audio metafile. The metafile acquisition unit 93 supplies the audio selection unit 94 with object position information included in the audio metafile.

15 [0101]

In step S34, the audio selection unit 94 selects an object in the display area designated by the user on the basis of the image frame size information supplied from the MPD processing unit 92 and the object position information
20 supplied from the metafile acquisition unit 93. The audio selection unit 94 requests the MPD processing unit 92 to send the information such as a URL for the audio file of the selected object.

[0102]

25 The MPD processing unit 92 extracts information such as a URL, which is described in "Segment" for audio file of the object requested from the audio selection unit 94, from the MPD file, and supplies the audio selection unit 94 with the extracted information. The audio selection unit 94
30 supplies the audio file acquisition unit 95 with the information such as a URL supplied from the MPD processing

unit 92.

[0103]

In step S35, on the basis of the information such as a URL supplied from the audio selection unit 94, the audio file acquisition unit 95 requests the Web server 12 to send an audio file of the selected object which is specified by the URL, and acquires the audio file. The audio file acquisition unit 95 supplies the audio decoding processing unit 96 with the acquired audio file in units of objects.

10 [0104]

In step S36, the image selection unit 98 selects a tile in the display area designated by the user on the basis of the tile position information supplied from the MPD processing unit 92. The image selection unit 98 requests the MPD processing unit 92 to send information such as a URL for the image file of the selected tile.

[0105]

The MPD processing unit 92 extracts information such as a URL, which is described in "Segment" for image file of the object requested from the image selection unit 98, from the MPD file, and supplies the image selection unit 98 with the extracted information. The image selection unit 98 supplies the image file acquisition unit 99 with the information such as a URL supplied from the MPD processing unit 92.

25

[0106]

In step S37, on the basis of the information such as a URL supplied from the image selection unit 98, the image file acquisition unit 99 requests the Web server 12 to send an image file of the selected tile which is specified by the URL, and acquires the image file. The image file acquisition

30

unit 99 supplies the image decoding processing unit 100 with the acquired image file in units of tiles.

[0107]

In step S38, the audio decoding processing unit 96
5 decodes an audio stream included in the audio file, which is supplied from the audio file acquisition unit 95 and configured in units of objects, to generate audio data in units of objects. The audio decoding processing unit 96
10 supplies the audio synthesis processing unit 97 with the audio data in units of objects.

[0108]

In step S39, the image decoding processing unit 100
decodes a video stream included in the image file, which is supplied from the image file acquisition unit 99 and
15 configured in units of tiles, to generate image data in units of tiles. The image decoding processing unit 100 supplies the image synthesis processing unit 101 with the image data in units of tiles.

[0109]

20 In step S40, the audio synthesis processing unit 97 synthesizes the audio data, which is supplied from the audio decoding processing unit 96 and configured in units of objects, and outputs the synthesized data. In step S41, the image synthesis processing unit 101 synthesizes the image data,
25 which is supplied from the image decoding processing unit 100 and configured in units of tiles, and outputs the synthesized data. Then, the process is terminated.

[0110]

30 As described above, the Web server 12 transmits the image frame size information and the object position information. Thus, the video playback terminal 14 can

specify, for example, an object in the display area to selectively acquire an audio file of the specified object so that the audio file corresponds to the image in the display area. This allows the video playback terminal 14 to acquire
5 only a necessary audio file, which leads to an improvement in transmission efficiency.

[0111]

Note that as shown in Fig. 15, an object ID (object specifying information) may be described in "AdaptationSet"
10 for an image of the MPD file as information for specifying an object corresponding to audio to play back at the same time with the image. The object ID may be described by extending Scheme (urn:mpeg:DASH:audioObj:2013) for defining new object ID information (audioObj) by utilizing a
15 DescriptorType element of Viewpoint. In this case, the video playback terminal 14 selects an audio file of the object corresponding to the object ID described in "AdaptationSet" for image, and acquires the audio file for playback.

[0112]

20 Instead of generating an audio file in units of objects, the encoded data of all objects may be multiplexed into a single audio stream to generate a single audio file.

[0113]

In this case, as shown in Fig. 16, one "Representation"
25 for audio file is provided in "AdaptationSet" for audio of the MPD file, and a URL or the like for the audio file (audioObj.mp4) including the encoded data of all objects is described in "Segment". At this time, object IDs (1, 2, 3, 4, and 5) of all objects corresponding to the audio file are
30 described by extending Viewpoint.

[0114]

In addition, in this case, as shown in Fig. 17, the encoded data (Audio object) of each object is arranged, as a sub-sample, in an mdat box of the audio file (hereinafter also referred to as an audio media file, as appropriate) acquired by referring to "Media Segment" of the MPD file.
[0115]

Specifically, data is arranged in the audio media file in units of subsegments that are any time shorter than a segment. The position of data in units of subsegments is specified by an sidx box. Further, the data in units of subsegments is composed of a moof box and an mdat box. The mdat box is composed of a plurality of samples, and the encoded data of each object is arranged as each sub-sample of the sample.

[0116]

Further, a gsix box in which information on a sample is described is arranged next to the sidx box of the audio media file. In this manner, the gsix box in which the information on the sample is described is provided separately from the moof box, and thus the video playback terminal 14 can acquire the information on the sample rapidly.

[0117]

As shown in Fig. 18, grouping_type representing the types of Sample group entries each composed of one or more samples or sub-samples managed by the gsix box is described in the gsix box. For example, when the Sample group entry is a sub-sample of the encoded data in units of objects, the type of the Sample group entry is "obja" as shown in Fig. 17. A plurality of gsix boxes of grouping_type are arranged in the audio media file.

[0118]

Further, as shown in Fig. 18, an index (`entry_index`) of each Sample group entry and a byte range (`range_size`) as data position information indicating the position in the audio media file are described in the `gsix` box. Note that

5 when the index (`entry_index`) is 0, the corresponding byte range indicates a byte range of the `moof` box (a1 in the example of Fig. 17).

[0119]

Information indicating which object is used for

10 allowing each Sample group entry to correspond to a sub-sample of encoded data is described in the audio file acquired by referring to "Initialization Segment" of the MPD file (hereinafter also referred to as an audio initialization file, as appropriate).

15 [0120]

Specifically, as shown in Fig. 19, this information is indicated by using a type assignment box (`typ`) of an `mvex` box that is associated with `AudioObjectSampleGroupEntry` of a sample group description box (`sgpd`) in an `sbt1` box of the

20 audio initialization file.

[0121]

In other words, as shown in A of Fig. 20, an object ID (`audio_object_id`) corresponding to the encoded data included in the sample is described in each

25 `AudioObjectSampleGroupEntry` box. For example, as shown in B of Fig. 20, object IDs 1, 2, 3, and 4 are described in each of four `AudioObjectSampleGroupEntry` boxes.

[0122]

On the other hand, as shown in Fig. 21, in the type

30 assignment box, an index as a parameter (`grouping_type_parameter`) of the Sample group entry

corresponding to the AudioObjectSampleGroupEntry is described for each AudioObjectSampleGroupEntry.

[0123]

5 The audio media file and the audio initialization file are configured as described above. Thus, when the video playback terminal 14 acquires the encoded data of the object selected as an object in the display area, the AudioObjectSampleGroupEntry in which the object ID of the selected object is described is retrieved from the stbl box of the audio initialization file. Then, the index of the Sample group entry corresponding to the retrieved AudioObjectSampleGroupEntry is read from the mvex box. Then, the position of data in units of subsegments is read from the sidx of the audio file, and the byte range of the Sample group entry of the read index is read from the gsix. Then, the encoded data arranged in the mdat is acquired on the basis of the position of data in units of subsegments and the byte range. Thus, the encoded data of the selected object is acquired.

20 [0124]

Although, in the above description, the index of Sample group entry and the object ID of AudioObjectSampleGroupEntry are associated with each other through the mvex box, they may be associated with each other directly. In this case, the index of Sample group entry is described in the AudioObjectSampleGroupEntry.

[0125]

Further, when the audio file is composed of a plurality of tracks, the sgpd can be stored in the mvex, which allows the sgpd to be shared among the tracks.

[0126]

(Outline of Second Example of Information Processing System)

Fig. 22 is a diagram illustrating an outline of a second example of the information processing system to which the present disclosure is applied.

5 [0127]

Note that the elements shown in Fig. 22 that are the same as those in Fig. 3 are denoted by the same reference numerals.

[0128]

10 In the example shown in Fig. 22, as is the case with Fig. 3, the image of video content is divided into 7 (width) × 5 (height) tiles, and audios of objects #1 to #8 are acquired as the audio of video content.

[0129]

15 In this case, when the user specifies the display area 31 composed of 3 (width) × 2 (height) tiles, the display area 31 is converted (extended) to an area having the same size as the size of the image of video content, thereby obtaining a display image 111 in a second example as shown in Fig. 22.

20 The audios of the objects #1 to #8 are synthesized on the basis of the positions of the objects #1 to #8 in a display image 111 and are output together with the display image 111. In other words, the audios of the objects #3 to #5, #7, and #8, which are outside the display area 31, are output, in addition to the audios of the objects #1, #2, and #6, which are inside the display area 31.

[0130]

(Configuration Example of Streaming Playback Unit)

30 The configuration of the second example of the information processing system to which the present disclosure is applied is the same as the configuration of the information

processing system 10 shown in Fig. 1 except for the configuration of the streaming playback unit, and thus only of the streaming playback unit will be described below.

[0131]

5 Fig. 23 is a block diagram showing a configuration example of the streaming playback unit of the information processing system to which the present disclosure is applied.

[0132]

10 The components shown in Fig. 23 that are the same as those in Fig. 13 are denoted by the same reference numerals, and repeated explanation is omitted as appropriate.

[0133]

15 The configuration of the streaming playback unit 120 shown in Fig. 23 differs from the configuration of the streaming playback unit 90 shown in Fig. 13 in that an MPD processing unit 121, an audio synthesis processing unit 123, and an image synthesis processing unit 124 are newly provided instead of the MPD processing unit 92, the audio synthesis processing unit 97, and the image synthesis processing unit 101, respectively, and a position determination unit 122 is additionally provided.

[0134]

25 The MPD processing unit 121 of the streaming playback unit 120 extracts information such as a URL, which is described in "Segment" for audio metafile, from the MPD file supplied from the MPD acquisition unit 91, and supplies the metafile acquisition unit 93 with the extracted information. Further, the MPD processing unit 121 extracts image frame size information of an image of the video content (hereinafter referred to as content image frame size information) that is
30 described in "AdaptationSet" for image from the MPD file and

supplies the position determination unit 122 with the extracted information. The MPD processing unit 121 extracts information such as a URL, which is described in "Segment" for audio file of all objects, from the MPD file, and supplies
5 the audio file acquisition unit 95 with the extracted information.

[0135]

The MPD processing unit 121 extracts the tile position information described in "AdaptationSet" for image from the
10 MPD file and supplies the image selection unit 98 with the extracted information. The MPD processing unit 121 extracts information such as a URL, which is described in "Segment" for the image file of the tile requested from the image selection unit 98, from the MPD file, and supplies the image
15 selection unit 98 with the extracted information.

[0136]

The position determination unit 122 acquires the object position information that is included in the audio metafile obtained by the metafile acquisition unit 93 and the content
20 image frame size information that is supplied from the MPD processing unit 121. Further, the position determination unit 122 acquires display area image frame size information that is the image frame size information of the display area designated by the user. The position determination unit 122
25 determines (recognizes) the position of each object in the display area on the basis of the object position information, the content image frame size information, and the display area image frame size information. The position determination unit 122 supplies the audio synthesis processing unit 123 with
30 the determined position of each object.

[0137]

The audio synthesis processing unit 123 synthesizes audio data in units of objects supplied from the audio decoding processing unit 96 on the basis of the object position supplied from the position determination unit 122.

5 Specifically, the audio synthesis processing unit 123 determines audio data to be allocated to each speaker for each object on the basis of the object position and the position of each speaker that outputs sound. The audio synthesis processing unit 123 synthesizes audio data of each object for
10 each speaker and outputs the synthesized audio data as audio data for each speaker. A detailed description of the method of synthesizing audio data of each object on the basis of the object position is disclosed in, for example, Ville Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude
15 Panning", Journal of AES, vol. 45, no. 6, pp. 456-466, 1997. [0138]

The image synthesis processing unit 124 synthesizes image data in units of tiles supplied from the image decoding processing unit 100. The image synthesis processing unit 124
20 functions as a converter, and converts the size of the image corresponding to the synthesized image data to the size of the video content to generate a display image. The image synthesis processing unit 124 outputs the display image. [0139]

25 (Explanation of Object Position Determination Method)

Figs. 24 to 26 are diagrams each illustrating the object position determination method by the position determination unit 122 shown in Fig. 23.

[0140]

30 The display area 31 is extracted from the video content and the size of the display area 31 is converted to the size

of the video content, so that the display image 111 is generated. Thus, the display image 111 has a size equivalent to the size obtained by shifting the center C of the display area 31 to the center C' of the display image 111 as shown in Fig. 24 and by converting the size of the display area 31 to the size of the video content as shown in Fig. 25.

[0141]

Thus, the position determination unit 122 calculates, by the following Formula (1), a shift amount θ_{shift} in horizontal direction when the center O of the display area 31 is shifted to the center O' of the display image 111.

[0142]

[Mathematical Formula 1]

$$\theta_{\text{shift}} = \frac{\theta_{v1}' + \theta_{v2}' - \theta_{v1} - \theta_{v2}}{2} \quad \dots (1)$$

[0143]

In Formula (1), θ_{v1}' represents a horizontal angle at a left end of the display area 31 included in the display area image frame size information, and θ_{v2}' represents a horizontal angle at a right end of the display area 31 included in the display area image frame size information. Further, θ_{v1} represents a horizontal angle at a left end in the content image frame size information, and θ_{v2} represents a horizontal angle at a right end in the content image frame size information.

[0144]

Next, the position determination unit 122 calculates, by the following Formula (2), a horizontal angle θ_{v1_shift}' at the left end of the display area 31 and a horizontal angle θ_{v2_shift}' at the right end thereof after the center O of the display area 31 is shifted to the center O' of the display

image 111 by using the shift mount θ_{shift} .

[0145]

[Mathematical Formula 2]

$$\theta_{v1_shift}' = \text{mod}(\theta_{v1}' + \theta_{\text{shift}} + 180^\circ, 360^\circ) - 180^\circ$$

$$5 \quad \theta_{v2_shift}' = \text{mod}(\theta_{v2}' + \theta_{\text{shift}} + 180^\circ, 360^\circ) - 180^\circ \dots (2)$$

[0146]

According to Formula (2), the horizontal angle θ_{v1_shift}' and the horizontal angle θ_{v2_shift}' are calculated so as not to exceed the range of -180° to 180° .

10 [0147]

Note that, as described above, the display image 111 has a size equivalent to the size obtained by shifting the center O of the display area 31 to the center O' of the display image 111 and by converting the size of the display area 31 to the size of the video content. Thus, the following Formula (3) is satisfied for the horizontal angles θ_{v1} and θ_{v2} .

[0148]

[Mathematical Formula 3]

$$\theta_{v1} = \frac{\theta_{v1} - \theta_{v2}}{\theta_{v1_shift}' - \theta_{v2_shift}'}, * \theta_{v1_shift}'$$

$$\theta_{v2} = \frac{\theta_{v1} - \theta_{v2}}{\theta_{v1_shift}' - \theta_{v2_shift}'}, * \theta_{v2_shift}' \dots (3)$$

20 [0149]

The position determination unit 122 calculates the shift amount θ_{shift} , the horizontal angle θ_{v1_shift}' , and the horizontal angles θ_{v2_shift}' , in the manner as described above, and then calculates a horizontal angle of each object in the display image 111. Specifically, the position determination unit 122 calculates, by the following Formula (4), a horizontal angle θ_{Ai_shift} of the object #i after the center C of the display area 31 is shifted to the center C' of the

display image 111 by using the shift mount θ_{shift} .

[0150]

[Mathematical Formula 4]

$$\theta_{\text{Ai_shift}} = \text{mod}(\theta_{\text{Ai}} + \theta_{\text{shift}} + 180^\circ, 360^\circ) - 180^\circ \quad \dots (4)$$

5 [0151]

In Formula (4), θ_{Ai} represents the horizontal angle of the object #i included in the object position information. Further, according to Formula (4), the horizontal angle $\theta_{\text{Ai_shift}}$ is calculated so as not to exceed the range of -180° to 180° .

10

[0152]

Next, when the object #i is present in the display area 31, that is, when the condition of $\theta_{\text{v2_shift}}' < \theta_{\text{Ai_shift}} < \theta_{\text{v1_shift}}'$ is satisfied, the position determination unit 122 calculates a horizontal angle θ_{Ai}' of the object #i in the display image 111 by the following Formula (5).

15

[0153]

[Mathematical Formula 5]

$$\theta_{\text{Ai}}' = \frac{\theta_{\text{v1}} - \theta_{\text{v2}}}{\theta_{\text{v1_shift}}' - \theta_{\text{v2_shift}}'} \cdot \left(\theta_{\text{Ai_shift}} - \frac{\theta_{\text{v1}} + \theta_{\text{v2}}}{2} \right) \quad \dots (5)$$

20 [0154]

According to Formula (5), the horizontal angle θ_{Ai}' is calculated by extending the distance between the position of the object #i in the display image 111 and the center C' of the display image 111 according to the ratio between the size of the display area 31 and the size of the display image 111.

25

[0155]

On the other hand, when no object #i is present in the display area 31, that is, when the condition of $-180^\circ \leq \theta_{\text{Ai_shift}} \leq \theta_{\text{v2_shift}}'$ or $\theta_{\text{v1_shift}}' \leq \theta_{\text{Ai_shift}} \leq 180^\circ$ is satisfied, the position determination unit 122 calculates the horizontal

30

angle θ_{Ai}' of the object #i in the display image 111 by the following Formula (6).

[0156]

[Mathematical Formula 6]

$$\theta_{Ai}' = \frac{\theta_{v2} + 180^\circ}{\theta_{v2_shift}' + 180^\circ} \cdot (\theta_{Ai_shift} + 180^\circ) - 180^\circ$$

(when $-180^\circ \leq \theta_{Ai_shift} \leq \theta_{v2_shift}'$)

$$\theta_{Ai}' = \frac{180^\circ - \theta_{v1}}{180^\circ - \theta_{v1_shift}} \cdot (\theta_{Ai_shift} - 180^\circ) + 180^\circ$$

(when $\theta_{v1_shift}' \leq \theta_{Ai_shift} \leq 180^\circ$)

5

... (6)

[0157]

According to Formula (6), when the object #i is present at a position 151 on the right side of the display area 31 ($-180^\circ \leq \theta_{Ai_shift} \leq \theta_{v2_shift}'$) as shown in Fig. 26, the horizontal angle θ_{Ai}' is calculated by extending the horizontal angle θ_{Ai_shift} according to the ratio between an angle R1 and an angle R2. Note that the angle R1 is the angle measured from the right end of the display image 111 to a position 154 just behind a viewer 153, and the angle R2 is the angle measured from the right end of the display area 31 whose center is shifted to the position 154.

[0158]

Further, according to Formula (6), when the object #i is present at a position 155 on the left side of the display area 31 ($\theta_{v1_shift}' \leq \theta_{Ai_shift} \leq 180^\circ$), the horizontal angle θ_{Ai}' is calculated by extending the horizontal angle θ_{Ai_shift} according to the ratio between an angle R3 and an angle R4. Note that the angle R3 is the angle measured from the left end of the display image 111 to the position 154, and the angle

R4 is the angle measured from the left end of the display area 31 whose center is shifted to the position 154.

[0159]

Further, the position determination unit 122 calculates a vertical angle γ_{Ai}' in a similar manner to the horizontal angle θ_{Ai}' . Specifically, the position determination unit 122 calculates, by the following Formula (7), a movement amount γ_{shift} in the vertical direction when the center C of the display area 31 is shifted to the center C' of the display image 111.

[0160]

[Mathematical Formula 7]

$$\gamma_{shift} = \frac{\gamma_{v1}' + \gamma_{v2}' - \gamma_{v1} - \gamma_{v2}}{2} \quad \dots (7)$$

[0161]

In Formula (7), γ_{v1}' represents the vertical angle at an upper end of the display area 31 included in the display area image frame size information, and γ_{v2}' represents the vertical angle at a lower end thereof. Further, γ_{v1} represents the vertical angle at an upper end in the content image frame size information, and γ_{v2} represents the vertical angle at a lower end in the content image frame size information.

[0162]

Next, the position determination unit 122 calculates, by the following Formula (8), a vertical angle γ_{v1_shift}' at an upper end of the display area 31 and a vertical angle γ_{v2_shift}' at a lower end thereof after the center C of the display area 31 is shifted to the center C' of the display image 111 by using the movement amount γ_{shift} .

[0163]

[Mathematical Formula 8]

$$\begin{aligned} \gamma_{v1_shift}' &= \text{mod}(\gamma_{v1}' + \gamma_{shift} + 90^\circ, 180^\circ) - 90^\circ \\ \gamma_{v2_shift}' &= \text{mod}(\gamma_{v2}' + \gamma_{shift} + 90^\circ, 180^\circ) - 90^\circ \quad \dots (8) \end{aligned}$$

[0164]

5 According to Formula (8), the vertical angle γ_{v1_shift}' and the vertical angle γ_{v2_shift}' are calculated so as not to exceed the range of -90° to 90° .

[0165]

10 The position determination unit 122 calculates the movement amount γ_{shift} , the vertical angle γ_{v1_shift}' , and the vertical angle γ_{v2_shift}' in the manner as described above, and then calculates the position of each object in the display image 111. Specifically, the position determination unit 122 calculates, by the following Formula (9), a vertical angle γ_{Ai_shift} of the object #i after the center C of the display area 31 is shifted to the center C' of the display image 111 by using the movement amount γ_{shift} .

[0166]

[Mathematical Formula 9]

$$20 \quad \gamma_{Ai_shift} = \text{mod}(\gamma_{Ai} + \gamma_{shift} + 90^\circ, 180^\circ) - 90^\circ \quad \dots (9)$$

[0167]

25 In Formula (9), γ_{Ai} represents the vertical angle of the object #i included in the object position information. Further, according to Formula (9), the vertical angle γ_{Ai_shift} is calculated so as not to exceed the range of -90° to 90° .

[0168]

Next, the position determination unit 122 calculates a vertical angle γ_{A1}' of the object #i in the display image 111 by the following Formula (10).

30 [0169]

[Mathematical Formula 10]

$$\gamma_{Ai}' = \begin{cases} \frac{\gamma_{v2}+90^\circ}{\gamma_{v2_shift}'+90^\circ} \cdot (\gamma_{Ai_shift}+90^\circ) - 90^\circ & \text{(when } -90^\circ \leq \gamma_{Ai_shift} \leq \gamma_{v2_shift}') \\ \frac{\gamma_{v1}-\gamma_{v2}}{\gamma_{v1_shift}'-\gamma_{v2_shift}'} \cdot \left(\gamma_{Ai_shift} - \frac{\gamma_{v1}+\gamma_{v2}}{2} \right) & \text{(when } \gamma_{v2_shift}' < \gamma_{Ai_shift} < \gamma_{v1_shift}') \\ \frac{90^\circ-\gamma_{v1}}{90^\circ-\gamma_{v1_shift}'} \cdot (\gamma_{Ai_shift}-90^\circ) + 90^\circ & \text{(when } \gamma_{v1_shift}' \leq \gamma_{Ai_shift} \leq 90^\circ) \end{cases} \dots (10)$$

[0170]

Further, the position determination unit 122 determines a distance r_{A1}' of the object #i in the display image 111 to be a distance r_{A1} of the object #i included in the object position information. The position determination unit 122 supplies the audio synthesis processing unit 123 with the horizontal angle θ_{Ai}' , the vertical angle γ_{A1}' , and the distance r_{A1} of the object #i, which are obtained as described above, as the position of the object #i.

[0171]

Fig. 27 is a diagram showing the relationship between the horizontal angle θ_{Ai} and the horizontal angle θ_{Ai}' .

[0172]

In the graph of Fig. 27, the horizontal axis represents the horizontal angle θ_{Ai} , and the vertical axis represents the horizontal angle θ_{Ai}' .

[0173]

As shown in Fig. 27, when the condition of $\theta_{v2}' < \theta_{Ai} < \theta_{v1}'$ is satisfied, the horizontal angle θ_{Ai} is shifted by

the movement amount θ_{shift} and is extend, and then the horizontal angle θ_{Ai} becomes equal to the horizontal angle θ_{Ai}' . Further, when the condition of $-180^\circ \leq \theta_{Ai} \leq \theta_{v2}'$ or $\theta_{v1}' \leq \theta_{Ai} \leq 180^\circ$ is satisfied, the horizontal angle θ_{Ai} is shifted
 5 by the movement amount θ_{shift} and is reduced, and then the horizontal angle θ_{Ai} becomes equal to the horizontal angle θ_{Ai}' .

[0174]

(Explanation of Process of Streaming Playback Unit)

10 Fig. 28 is a flowchart illustrating a streaming playback process of the streaming playback unit 120 shown in Fig. 23.

[0175]

In step S131 of Fig. 28, the MPD acquisition unit 91
 15 of the streaming playback unit 120 acquires the MPD file from the Web server 12 and supplies the MPD processing unit 121 with the MPD file.

[0176]

In step S132, the MPD processing unit 121 acquires the
 20 content image frame size information and the tile position information, which are described in "AdaptationSet" for image, from the MPD file supplied from the MPD acquisition unit 91. The MPD processing unit 121 supplies the position
 25 determination unit 122 with the image frame size information, and supplies the image selection unit 98 with the tile position information. Further, the MPD processing unit 121 extracts information such as a URL described in "Segment" for audio metafile, and supplies the extracted information to the metafile acquisition unit 93.

30 [0177]

In step S133, the metafile acquisition unit 93 requests

the Web server 12 to send the audio metafile specified by the URL on the basis of the information such as the URL supplied from the MPD processing unit 121, and acquires the audio metafile. The metafile acquisition unit 93 supplies the
5 position determination unit 122 with the object position information included in the audio metafile.

[0178]

In step S134, the position determination unit 122 performs a position determination process for determining the
10 position of each object in the display image on the basis of the object position information, the content image frame size information, and the display area image frame size information. The position determination process will be described in detail with reference to Fig. 29 which is
15 described later.

[0179]

In step S135, the MPD processing unit 121 extracts information such as a URL described in "Segment" for audio file of all objects from the MPD file, and supplies the audio
20 file acquisition unit 95 with the extracted information.

[0180]

In step S136, the audio file acquisition unit 95 requests the Web server 12 to send an audio file of all objects specified by the URL on the basis of the information such as
25 the URL supplied from the MPD processing unit 121, and acquires the audio file. The audio file acquisition unit 95 supplies the audio decoding processing unit 96 with the acquired audio file in units of objects.

[0181]

30 The process of steps S137 to S140 is similar to the process of steps S36 to S39 shown in Fig. 14, and thus the

descriptions thereof will be omitted.

[0182]

In step S141, the audio synthesis processing unit 123 synthesizes and outputs the audio data in units of objects
5 supplied from the audio decoding processing unit 96 on the basis of the position of each object supplied from the position determination unit 122.

[0183]

In step S142, the image synthesis processing unit 124
10 synthesizes the image data in units of tiles supplied from the image decoding processing unit 100.

[0184]

In step S143, the image synthesis processing unit 124 converts the size of the image corresponding to the
15 synthesized image data into the size of the video content, and generates the display image. Then, the image synthesis processing unit 124 outputs the display image, and the process is terminated.

[0185]

20 Fig. 29 is a flowchart illustrating details of the position determination process in step S134 of Fig. 28. This position determination process is carried out, for example, for each object.

[0186]

25 In step S151 of Fig. 29, the position determination unit 122 performs a horizontal angle θ_{Ai}' estimation process for estimating the horizontal angle θ_{Ai}' in the display image. Details of the horizontal angle θ_{Ai}' estimation process will be described with reference to Fig. 30 which is described
30 later.

[0187]

In step S152, the position determination unit 122 performs a vertical angle γ_{Ai}' estimation process for estimating the vertical angle γ_{Ai}' in the display image. Details of the vertical angle γ_{Ai}' estimation process are similar to those of the horizontal angle θ_{Ai}' estimation process in step S151, except that the vertical direction is used in place of the horizontal direction, and thus a detailed description thereof will be omitted.

[0188]

10 In step S153, the position determination unit 122 determines a distance r_{Ai}' in the display image to be a distance r_{Ai} included in the object position information supplied from the metafile acquisition unit 93.

[0189]

15 In step S154, the position determination unit 122 outputs, to the audio synthesis processing unit 123, the horizontal angle θ_{Ai}' , the vertical angle γ_{Ai}' , and the distance r_{Ai} , as the position of the object #i. Then, the process returns to step S134 of Fig. 28 and proceeds to step
20 S135.

[0190]

Fig. 30 is a flowchart illustrating details of the horizontal angle θ_{Ai}' estimation process in step S151 of Fig. 29.

25 [0191]

In step S171 shown in Fig. 30, the position determination unit 122 acquires the horizontal angle θ_{Ai} included in the object position information supplied from the metafile acquisition unit 93.

30 [0192]

In step S172, the position determination unit 122

acquires the content image frame size information supplied from the MPD processing unit 121 and the display area image frame size information specified by the user.

[0193]

5 In step S173, the position determination unit 122 calculates the movement amount θ_{shift} by the above-mentioned Formula (1) on the basis of the content image frame size information and the display area image frame size information.

10 [0194]

 In step S174, the position determination unit 122 calculates horizontal angles θ_{v1_shift}' and θ_{v2_shift}' by the above-mentioned Formula (2) using the movement amount θ_{shift} and the display area image frame size.

15 [0195]

 In step S175, the position determination unit 122 calculates the horizontal angle θ_{Ai_shift} by the above-mentioned Formula (4) using the horizontal angle θ_{Ai} and the movement amount θ_{shift} .

20 [0196]

 In step S176, the position determination unit 122 determines whether the object #i is present in the display area 31 (the horizontal angle of the object #i is between the horizontal angles at both ends of the display area 31), i.e.,
25 whether the condition of $\theta_{v2_shift}' < \theta_{Ai_shift} < \theta_{v1_shift}'$ is satisfied or not.

[0197]

 When it is determined in step S176 that the object #i is present in the display area 31, that is, when the condition
30 of $\theta_{v2_shift}' < \theta_{Ai_shift} < \theta_{v1_shift}'$ is satisfied, the process proceeds to step S177. In step S177, the position

determination unit 122 calculates the horizontal angle θ_{A1}' by the above-mentioned Formula (5) on the basis of the content image frame size information, the horizontal angles θ_{v1_shift}' and θ_{v2_shift}' , and the horizontal angle θ_{Ai_shift} .

5 [0198]

On the other hand, when it is determined in step S176 that the object #i is not present in the display area 31, that is, when the condition of $-180^\circ \leq \theta_{Ai_shift} \leq \theta_{v2_shift}'$ or $\theta_{v1_shift}' \leq \theta_{Ai_shift} \leq 180^\circ$ is satisfied, the process proceeds to step S178. In step S178, the position determination unit 122 calculates the horizontal angle θ_{Ai}' by the above-mentioned Formula (6) on the basis of the content image frame size information, the horizontal angle θ_{v1_shift}' or θ_{v2_shift}' , and the horizontal angle θ_{Ai_shift} .

15 [0199]

After the process of step S177 or step S178, the process returns to step S151 of Fig. 29 and proceeds to step S152.
[0200]

Note that in the second example, the size of the display image is the same as the size of the video content, but instead the size of the display image may be different from the size of the video content.

[0201]

Further, in the second example, the audio data of all objects is not synthesized and output, but instead only the audio data of some objects (for example, an object in the display area, an object in a predetermined range from the display area, etc.). The method for selecting an object of audio data to be output may be determined in advance, or may be specified by the user.

30

[0202]

Further, in the above description, only the audio data in units objects is used, but the audio data may include audio data of channel audio, audio data of higher-order ambisonics (HOA) audio, audio data of spatial audio object coding (SAOC), and metadata (scene information, dynamic or static metadata) of audio data. In this case, for example, not only the coded data of each object, but also the coded data of these pieces of data are arranged as sub-samples.

[0203]

10 <First Embodiment>

(Outline of 3D Audio File Format)

Prior to the description of the first embodiment to which the present disclosure is applied, the outline of tracks of the 3D audio file format of MP4 will be described with reference to Fig. 31.

15 [0204]

In the MP4 file, the codec information of the video content and the position information indicating the position in the file can be managed for each track. In the 3D audio file format of MP4, all audio streams (elementary stream (ES)) of 3D audio (Channel audio/Object audio/HOA audio/metadata) are recorded as one track in units of samples (frames). Further, the codec information (Profile/level/audio configuration) of 3D audio is stored as a sample entry.

25 [0205]

Channel audio constituting the 3D audio is audio data in units of channels; Object audio is audio data in units of objects; HOA audio is spherical audio data; and metadata is metadata of Channel audio/Object audio/HOA audio. In this case, audio data in units of objects is used as Object audio, but instead audio data of SAOC may be used.

[0206]

(Structure of moov box)

Fig. 32 shows a structure of a moov box of an MP4 file.

[0207]

5 As shown in Fig. 32, in the MP4 file, the image data and the audio data are recorded in different tracks. Fig. 32 does not illustrate the details of the track of the audio data, but the track of the audio data is similar to the track of the image data. The sample entry is included in the sample description arranged in an stsd box within the moov box.

[0208]

Incidentally, in broadcasting or local storage playback, the Web server delivers all audio streams, and the video playback terminal (client) decodes audio streams of necessary 3D audio, while parsing all audio streams, and outputs (rendering) the audio streams. When the bit rate (Bitrate) is high, or when there is a limitation on the rate of reading of a local storage, there is a demand for reducing the load on the decode process by acquiring only the audio stream of necessary 3D audio.

[0209]

Further, in stream playback, there is a demand for the video playback terminal (client) to acquire only the coded data of necessary 3D audio to thereby acquire an audio stream of a coding rate optimum for the playback environment.

[0210]

Accordingly, in the present disclosure, the coded data of 3D audio is divided into tracks for each type of the data and the tracks are arranged in the audio file, which makes it possible to efficiently acquire only a predetermined type of coded data. Thus, the load on the system can be reduced

in broadcasting and local storage playback. Further, in stream playback, the highest-quality coded data of necessary 3D audio can be played back according to the frequency band. Further, since it is only necessary to record the position information of the audio stream of 3D audio within the audio file in units of tracks of subsegments, the amount of position information can be reduced as compared with the case where the coded data in units of objects are arranged in the sub-sample.

10 [0211]

(Outline of Tracks)

Fig. 33 is a diagram illustrating the outline of tracks in the first embodiment to which the present disclosure is applied.

15 [0212]

As shown in Fig. 33, in the first embodiment, the Channel audio/Object audio/HOA audio/metadata constituting the 3D audio are respectively set as audio streams of different tracks (Channel audio track/Object audio track(s)/HOA audio track/Object metadata track). The audio stream of audio metadata is arranged in the object metadata track.

[0213]

Further, as a track for arranging information about the entire 3D audio, a base track (Base Track) is provided. In the base track shown in Fig. 33, the information about the entire 3D audio is arranged in the sample entry, while no sample is arranged in the sample entry. Further, the Base track, Channel audio track, Object audio track(s), HOA audio track, and Object metadata track are recorded as the same audio file (3daudio.mp4).

[0214]

Track Reference is arranged in, for example, a track box, and represents a reference relationship between a corresponding track and another track. Specifically, Track
 5 Reference represents an ID unique to a track in other referred tracks (hereinafter referred to as a track ID). In the example shown in Fig. 33, the track IDs of Base track, Channel audio track, HOA audio track, Object metadata track, and Object audio track(s) are 1, 2, 3, 4, 10 ..., respectively.
 10 Track References of Base track are 2, 3, 4, 10 ..., and Track References of Channel audio track/HOA audio track/Object metadata track/Object audio track(s) are 1 which corresponds to the track ID of Base track.

[0215]

15 Accordingly, Base track and Channel audio track/HOA audio track/Object metadata track/Object audio track(s) have a reference relationship. Specifically, Base track is referred to during playback of Channel audio track/HOA audio track/Object metadata track/Object audio track(s).

20 [0216]

(Exemplary Syntax of Sample Entry of Base Track)

Fig. 34 is a diagram showing an exemplary syntax of a sample entry of the base track shown in Fig. 33.

[0217]

25 As information about the entire 3D audio, configurationVersion, MPEGHAudioProfile, and MPEGHAudioLevel shown in Fig. 34 represent config information, profile information, and level information, respectively, of the entire audio stream of 3D audio (audio stream of normal
 30 3D audio). Further, as information about the entire 3D audio, the width and the height shown in Fig. 34 represent the number

of pixels in the horizontal direction of the video content and the number of pixels in the vertical direction of the video content, respectively. As information about the entire 3D audio, θ_{v1} , θ_{v2} , γ_{v1} , and γ_{v2} represent the
 5 horizontal angle θ_{v1} at the left end of the image frame, the horizontal angle θ_{v2} at the right end of the image frame, the vertical angle γ_{v1} at the upper end of the image frame, and the vertical angle γ_{v2} at the lower end of the image frame, respectively, in the image frame size information of the video
 10 content.

[0218]

(Exemplary Syntax of Sample Entry of Channel Audio Track)

Fig. 35 is a diagram showing an exemplary syntax of a sample entry of the channel audio track (Channel audio track)
 15 shown in Fig. 33.

[0219]

Fig. 35 shows `configurationVersion`, `MPEGHAudioProfile`, and `MPEGHAudioLevel` represent config information, profile information, and level information, respectively, of Channel
 20 Audio.

[0220]

(Exemplary Syntax of Sample Entry of Object Audio Track)

Fig. 36 is a diagram showing an exemplary syntax of a sample entry of the object audio track (Object audio track)
 25 shown in Fig. 33.

[0221]

`ConfigurationVersion`, `MPEGHAudioProfile`, and `MPEGHAudioLevel` shown in Fig. 36 represent config information, profile information, and level information, respectively, in
 30 one or more Object audios included in the object audio track. `object_is_fixed` indicates whether one or more Object audio

objects included in the object audio track are fixed or not. When `object_is_fixed` indicates 1, it indicates that the object is fixed, and when `object_is_fixed` indicates 0, it indicates that the object is shifted. `mpegh3daConfig`
 5 represents config of identification information of one or more Object audio objects included in the object audio track.
 [0222]

Further,
`objectTheta1/objectTheta2/objectGamma1/objectGamma2/objectRength`
 10 `tRength` represents object information of one or more Object audios included in the object audio track. This object information is information which is valid when `Object_is_fixed = 1` holds.
 [0223]

15 `maxobjectTheta1, maxobjectTheta2, maxobjectGamma1, maxobjectGamma2/,` and `maxobjectRength` represent maximum values of object information when one or more Object audio objects included in the object audio track are shifted.
 [0224]

20 (Exemplary Syntax of Sample Entry of HOA Audio Track)

Fig. 37 is a diagram showing an exemplary syntax of a sample entry of the HOA audio track shown in Fig. 33.
 [0225]

`ConfigurationVersion, MPEGHAudioProfile,` and
 25 `MPEGHAudioLevel` shown in Fig. 37 represent config information, profile information, and level information, respectively, of HOA audio.
 [0226]

(Exemplary Syntax of Sample Entry of Object Metadata Track)

30 Fig. 38 is a diagram showing an exemplary syntax of a sample entry of the object metadata track (Object metadata

track) shown in Fig. 33.

[0227]

ConfigurationVersion shown in Fig. 38 represents config information of metadata.

5 [0228]

(First Example of Segment Structure of Audio File of 3D Audio)

Fig. 39 is a diagram showing a first example of a segment structure of an audio file of 3D audio in the first embodiment to which the present disclosure is applied.

10 [0229]

In the segment structure shown in Fig. 39, Initial segment is composed of an ftyp box and a moov box. trak boxes for each track included in the audio file are arranged in the moov box. An mvex box including information indicating the correspondence relation between the track ID of each track and the level used in an ssix box within the media segment is arranged in the moov box.

[0230]

Further, the media segment is composed of the sidx box, the ssix box, and one or more subsegments. Position information indicating a position in the audio file of each subsegment is arranged in the sidx box. The ssix box includes position information of an audio stream at each level arranged in the mdat box. Note that each level corresponds to each track. Further, the position information of a first track is the position information of data composed of audio streams of the moof box and the first track.

[0231]

The subsegment is provided for any length of time. A pair of a moof box and an mdat box which are common to all tracks is provided in the subsegment. In the mdat box, audio

streams of all tracks are collectively arranged for any length of time. In the moof box, management information of the audio streams is arranged. The audio streams of each track arranged in the mdat box are continuous for each track.

5 [0232]

In the example of Fig. 39, Track1 having the track ID of 1 is a base Track, and Track2 to trackN having track IDs of 2 to N are Channel Audio Track, Object audio track(s), HOA audio track, and object metadata track, respectively. The same holds true for Fig. 40 to be described later.

[0233]

(Second Example of Segment Structure of Audio File of 3D Audio)

Fig. 40 is a diagram showing a second example of the segment structure of the audio file of 3D audio in the first embodiment to which the present disclosure is applied.

[0234]

The segment structure shown in Fig. 40 is different from the segment structure shown in Fig. 39 in that a moof box and an mdat box are provided for each track.

[0235]

Specifically, Initial segment shown in Fig. 40 is similar to Initial segment shown in Fig. 39. Like the media segment shown in Fig. 39, the media segment shown in Fig. 40 is composed of the sidx box, the ssix box, and one or more subsegments. Further, like the sidx box shown in Fig. 39, the position information of each subsegment is arranged in the sidx box. The ssix box includes the position information of data of each level that is composed of the moof box and the mdat box.

[0236]

The subsegment is provided for any length of time. A pair of a moof box and an mdat box is provided for each track in the subsegment. Specifically, audio streams of each track are collectively arranged (interleaved and stored) for any length of time in the mdat box of each track, and management information of the audio streams is arranged in the moof box.

[0237]

As shown in Figs. 39 and 40, the audio streams for each track are collectively arranged for any length of time, so that the efficiency of acquiring audio streams via HTTP or the like is improved as compared with the case where audio streams are collectively arranged in units of samples.

[0238]

(Exemplary Description of mvex Box)

Fig. 41 is a diagram showing an exemplary description of a level assignment box arranged in the mvex box shown in Figs. 39 and 40.

[0239]

The level assignment box is a box for associating the track ID of each track with the level used in the ssix box. In the example of Fig. 41, the base track having the track ID of 1 is associated with the level 0, and the channel audio track having the track ID of 2 is associated with the level 1. Further, the HOA audio track having the track ID of 3 is associated with the level 2, and the object metadata track having the track ID of 4 is associated with the level 3. Furthermore, the object audio track having the track ID of 10 is associated with the level 4.

[0240]

(Exemplary Description of MPD File)

Fig. 42 is a diagram showing an exemplary description

of an MPD file in the first embodiment to which the present disclosure is applied.

[0241]

As shown in Fig. 42, "Representation" for managing the
5 segment of the audio file (3daudio.mp4) of 3D audio,
"SubRepresentation" for managing the tracks included in the
segment, and the like are described in the MPD file.

[0242]

In "Representation" and "SubRepresentation", "codecs"
10 representing a type of codec of a corresponding segment or
track in a code defined in a 3D audio file format is included.
Further, "id", "associationId", and "associationType" are
included in "Representation".

[0243]

15 "id" represents the ID of "Representation" in which
"id" is included. "associationId" represents information
indicating a reference relationship between a corresponding
track and another track, and represents "id" of the reference
track. "associationType" represents a code indicating the
20 meaning of a reference relationship (dependent relationship)
with respect to the reference track. For example, the same
value as the value of the track reference of MP4 is used.

[0244]

Further, in "SubRepresentation", "level" which is a
25 value set in the level assignment box as the value
representing the corresponding track and the corresponding
level is included. In "SubRepresentation",
"dependencyLevel" which is a value representing a level
corresponding to another track having a reference
30 relationship (dependency) (hereinafter referred to as a
reference track) is included.

[0245]

Further, "SubRepresentation" includes
 <EssentialProperty schemeIdUri=
 "urn:mpeg:DASH:3daudio:2014" value= "audioType,
 5 contentkind, priority"> as information necessary for
 selection of 3D audio.

[0246]

Further, "SubRepresentation" in the Object audio track
 includes <EssentialProperty schemeIdUri=
 10 "urn:mpeg:DASH:viewingAngle:2014" value= " θ , γ , r ">. When
 the object corresponding to "SubRepresentation" is fixed, θ ,
 γ , and r represent a horizontal angle, a vertical angle, and
 a distance, respectively, in the object position information.
 On the other hand, when the object is shifted, the values θ ,
 15 γ , and r represent the maximum value of the horizontal angle,
 the maximum value of the vertical angle, and the maximum value
 of the distance, respectively, among the maximum values of
 the object position information.

[0247]

20 Fig. 43 is a diagram showing a definition of Essential
 Property shown in Fig. 42.

[0248]

On the upper left side of Fig. 43, AudioType of
 <EssentialProperty schemeIdUri=
 25 "urn:mpeg:DASH:3daudio:2014" value= " audioType,
 contentkind, priority"> is defined. AudioType represents
 the type of 3D audio of the corresponding track.

[0249]

In the example of Fig. 43, when AudioType indicates 1,
 30 it indicates that the audio data of the corresponding track
 is Channel audio of 3D audio, and when AudioType indicates

2, it indicates that the audio data of the corresponding track is HOA audio. Further, when AudioType indicates 3, it indicates that the audio data of the corresponding track is Object audio, and when AudioType is 4, it indicates that the audio data of the corresponding track is metadata.

[0250]

Further, on the right side of Fig. 43, contentkind of <EssentialProperty schemeIdUri=

"urn:mpeg:DASH:3daudio:2014" value= " audioType,

contentkind, priority"> is defined. The contentkind represents the content of the corresponding audio. In the example of Fig. 43, for example, when the contentkind indicates 3, the corresponding audio is music.

[0251]

As shown in the lower left of Fig. 43, Priority is defined by 23008-3 and represents the processing priority of the corresponding Object. A value representing the processing priority of Object is described as Priority only when the value is not changed during the audio stream, while when the value is changed during the audio stream, a value "0" is described.

[0252]

(Outline of Information Processing System)

Fig. 44 is a diagram illustrating an outline of an information processing system according to the first embodiment to which the present disclosure is applied.

[0253]

The components shown in Fig. 44 that are the same as the components shown in Fig. 1 are denoted by the same reference numerals. Repeated explanation is omitted as appropriate.

[0254]

An information processing system 140 shown in Fig. 44 has a configuration in which a Web server 142, which is connected to a file generation device 141 is connected to a video playback terminal 144 via the Internet 13.

[0255]

In the information processing system 140, the Web server 142 delivers (tiled streaming) a video stream of video content to the video playback terminal 144 in units of tiles by a method in conformity with MPEG-DASH. Further, in the information processing system 140, the Web server 142 delivers an audio stream of Object audio, Channel audio, or HOA audio corresponding to the tile to be played back to the video playback terminal 144.

15 [0256]

The file generation device 141 of the information processing system 140 is similar to the file generation device 11 shown in Fig. 11, except that, for example, the audio file generation unit 56 generates an audio file in the first embodiment and the MPD generation unit 57 generates the MPD file in the first embodiment.

[0257]

Specifically, the file generation device 141 acquires the image data of video content and encodes the image data in units of tiles to generate a video stream. The file generation device 141 processes the video stream of each tile into a file format. The file generation device 141 uploads the image file of each tile obtained as a result of the process to the Web server 142.

30 [0258]

Further, the file generation device 141 acquires 3D

audio of video content and encodes the 3D audio for each type (Channel audio/Object audio/HOA audio/metadata) of 3D audio to generate an audio stream. The file generation device 141 allocates tracks to the audio stream for each type of 3D audio.

5 The file generation device 141 generates the audio file of the segment structure shown in Fig. 39 or 40 in which the audio stream of each track is arranged in units of subsegments, and uploads the audio file to the Web server 142.

[0259]

10 The file generation device 141 generates an MPD file including image frame size information, tile position information, and object position information. The file generation device 141 uploads the MPD file to the Web server 142.

15 [0260]

The Web server 142 stores the image file, the audio file, and the MPD file which are uploaded from the file generation device 141.

[0261]

20 In the example of Fig. 44, the Web server 142 stores a segment group formed of image files of a plurality of segments of the tile #1 and a segment group formed of image files of a plurality of segments of the tile #2. The Web server 142 also stores a segment group formed of audio files
25 of 3D audio.

[0262]

The Web server 142 transmits, to the video playback terminal 144, the image file, the audio file, the MPD file, and the like stored in the Web server, in response to a request
30 from the video playback terminal 144.

[0263]

The video playback terminal 144 executes control software 161, video playback software 162, access software 163, and the like.

[0264]

5 The control software 161 is software for controlling data to be streamed from the Web server 142. Specifically, the control software 161 causes the video playback terminal 144 to acquire the MPD file from the Web server 142.

[0265]

10 Further, the control software 161 specifies a tile in the display area on the basis of the display area instructed from the video playback software 162 and the tile position information included in the MPD file. Then, the control software 161 instructs the access software 163 to transmit
15 a request for the image file of the tile.

[0266]

When Object audio is to be played back, the control software 161 instructs the access software 163 to transmit a request for the image frame size information in the audio
20 file. Further, the control software 161 instructs the access software 163 to transmit a request for the audio stream of metadata. The control software 161 specifies the object corresponding to the image in the display area on the basis of the image frame size information and the object position
25 information included in the audio stream of metadata, which are transmitted from the Web server 142 according to the instruction, and the display area. Then, the control software 161 instructs the access software 163 to transmit a request for the audio stream of the object.

30 [0267]

Further, when Channel audio or HOA audio is to be played

back, the control software 161 instructs the access software 163 to transmit a request for the audio stream of Channel audio or HOA audio.

[0268]

5 The video playback software 162 is software for playing back the image file and the audio file which are acquired from the Web server 142. Specifically, when the display area is specified by the user, the video playback software 162 instructs the control software 161 to transmit the display
10 area. Further, the video playback software 162 decodes the image file and the audio file which are acquired from the Web server 142 according to the instruction. The video playback software 162 synthesizes and outputs the image data in units of tiles obtained as a result of decoding. Further, the video
15 playback software 162 synthesizes and outputs, as needed, the Object audio, Channel audio, or HOA audio, which are obtained as a result of decoding.

[0269]

 The access software 163 is software for controlling the
20 communication with the Web server 142 via the Internet 13 using HTTP. Specifically, the access software 163 causes the video playback terminal 144 to transmit a request for the image frame size information or predetermined audio stream in the image file and audio file in response to the instruction
25 from the control software 161. Further, the access software 163 causes the video playback terminal 144 to receive the image frame size information or predetermined audio stream in the image file and audio file, which are transmitted from the Web server 142, in response to the transmission request.

30 [0270]

(Configuration Example of File Generation Device)

Fig. 45 is a block diagram showing a configuration example of the file generation device 141 shown in Fig. 44.
[0271]

The components shown in Fig. 45 that are the same as
5 the components shown in Fig. 11 are denoted by the same reference numerals. Repeated explanation is omitted as appropriate.
[0272]

The configuration of the file generation device 141
10 shown in Fig. 45 is different from the configuration of the file generation device 11 shown in Fig. 11 in that an audio coding processing unit 171, an audio file generation unit 172, an MPD generation unit 173, and a server upload processing unit 174 are provided instead of the audio coding processing
15 unit 55, the audio file generation unit 56, the MPD generation unit 57, and the server upload processing unit 58.
[0273]

Specifically, the audio coding processing unit 171 of the file generation device 141 encodes the 3D audio of video
20 content input from the outside for each type (Channel audio/Object audio/HOA audio/metadata) to generate an audio stream. The audio coding processing unit 171 supplies the audio file generation unit 172 with the audio stream for each type of the 3D audio.
25 [0274]

The audio file generation unit 172 allocates tracks to the audio stream, which is supplied from the audio coding processing unit 171, for each type of the 3D audio. The audio file generation unit 172 generates the audio file of the
30 segment structure shown in Fig. 39 or 40 in which the audio stream of each track is arranged in units of subsegments. At

this time, the audio file generation unit 172 stores the image frame size information input from the outside in the sample entry. The audio file generation unit 172 supplies the MPD generation unit 173 with the generated audio file.

5 [0275]

The MPD generation unit 173 determines the URL or the like of the Web server 142 that stores the image file of each tile supplied from the image file generation unit 53. Further, the MPD generation unit 173 determines the URL or the like
10 of the Web server 142 that stores the audio file supplied from the audio file generation unit 172.

[0276]

The MPD generation unit 173 arranges the image information supplied from the image information generation
15 unit 54 in "AdaptationSet" for image of the MPD file. Further, the MPD generation unit 173 arranges the URL or the like of the image file of each tile in "Segment" of "Representation" for the image file of the tile.

[0277]

20 The MPD generation unit 173 arranges the URL or the like of the audio file in "Segment" of "Representation" for the audio file. Further, the MPD generation unit 173 arranges the object position information or the like of each object input from the outside in "Sub Representation" for the Object
25 metadata track of the object. The MPD generation unit 173 supplies the server upload processing unit 174 with the MPD file, in which various pieces of information arranged as described above, and the image file and audio file.

[0278]

30 The server upload processing unit 174 uploads the image file, the audio file, and the MPD file of each tile supplied

from the MPD generation unit 173 to the Web server 142.

[0279]

(Explanation of Process of File Generation Device)

Fig. 46 is a flowchart illustrating a file generation
5 process of the file generation device 141 shown in Fig. 45.

[0280]

The process of steps S191 to S195 shown in Fig. 46 is
similar to the process of steps S11 to S15 shown in Fig. 12,
and thus the description thereof is omitted.

10 [0281]

In step S196, the audio coding processing unit 171
encodes the 3D audio of video content input from the outside
for each type (Channel audio/Object audio/HOA
audio/metadata) to generate an audio stream. The audio
15 coding processing unit 171 supplies the audio file generation
unit 172 with the audio stream for each type of the 3D audio.

[0282]

In step S197, the audio file generation unit 172
allocates tracks to the audio stream, which is supplied from
20 the audio coding processing unit 171, for each type of the
3D audio.

[0283]

In step S198, the audio file generation unit 172
generates the audio file of the segment structure shown in
25 Fig. 39 or 40 in which the audio stream of each track is
arranged in units of subsegments. At this time, the audio
file generation unit 172 stores the image frame size
information input from the outside in the sample entry. The
audio file generation unit 172 supplies the MPD generation
30 unit 173 with the generated audio file.

[0284]

In step S199, the MPD generation unit 173 generates the MPD file including the image information supplied from the image information generation unit 54, the URL of each file, and the object position information. The MPD generation unit
5 173 supplies the server upload processing unit 174 with the image file, the audio file, and the MPD file.

[0285]

In step S200, the server upload processing unit 174 uploads the image file, the audio file, and the MPD file, which
10 are supplied from the MPD generation unit 173, to the Web server 142. Then, the process is terminated.

[0286]

(Functional Configuration Example of Video Playback Terminal)

15 Fig. 47 is a block diagram showing a configuration example of the streaming playback unit which is implemented in such a manner that the video playback terminal 144 shown in Fig. 44 executes the control software 161, the video playback software 162, and the access software 163.

20 [0287]

The components shown in Fig. 47 that are the same as the components shown in Fig. 13 are denoted by the same reference numerals. Repeated explanation is omitted as appropriate.

25 [0288]

The configuration of the streaming playback unit 190 shown in Fig. 47 is different from the configuration of the streaming playback unit 90 shown in Fig. 13 in that an MPD processing unit 191, an audio selection unit 193, an audio
30 file acquisition unit 192, an audio decoding processing unit 194, and an audio synthesis processing unit 195 are provided

instead of the MPD processing unit 92, the audio selection unit 94, the audio file acquisition unit 95, the audio decoding processing unit 96, and the audio synthesis processing unit 97 and the metafile acquisition unit 93 is not provided.

[0289]

The streaming playback unit 190 is similar to the streaming playback unit 90 shown in Fig. 13, except for, for example, the method of acquiring the audio data to be played back of the selected object.

[0290]

Specifically, the MPD processing unit 191 of the streaming playback unit 190 extracts information, such as the URL of the audio file of the segment to be played back that is described in "Segment" for audio file, from the MPD file supplied from the MPD acquisition unit 91, and supplies the audio file acquisition unit 192 with the extracted information.

[0291]

The MPD processing unit 191 extracts the tile position information described in "AdaptationSet" for image from the MPD file, and supplies the image selection unit 98 with the extracted information. The MPD processing unit 191 extracts information, such as the URL described in "Segment" for the image file of the tile requested from the image selection unit 98, from the MPD file, and supplies the image selection unit 98 with the extracted information.

[0292]

When Object audio is to be played back, the audio file acquisition unit 192 requests the Web server 142 to transmit Initial Segment of Base track in the audio file specified by

the URL on the basis of the information such as the URL supplied from the MPD processing unit 191, and acquires the Initial Segment of Base track.

[0293]

5 Further, on the basis of the information such as the URL of the audio file, the audio file acquisition unit 192 requests the Web server 142 to transmit the audio stream of the object metadata track in the audio file specified by the URL, and acquires the audio stream of the object metadata
10 track. The audio file acquisition unit 192 supplies the audio selection unit 193 with the object position information included in the audio stream of the object metadata track, the image frame size information included in Initial Segment of Base track, and the information such as the URL of the audio
15 file.

[0294]

 Further, when Channel audio is to be played back, the audio file acquisition unit 192 requests the Web server 142 to transmit the audio stream of Channel audio track in the
20 audio file specified by the URL on the basis of the information such as the URL of the audio file, and acquires the audio stream of Channel audio track. The audio file acquisition unit 192 supplies the audio decoding processing unit 194 with the acquired audio stream of Channel audio track.

25 [0295]

 When HOA audio is to be played back, the audio file acquisition unit 192 performs a process similar to that performed when Channel audio is to be played back. As a result, the audio stream of the HOA audio track is supplied to the
30 audio decoding processing unit 194.

[0296]

Note that it is determined which one of Object audio, Channel audio, and HOA audio is to be played back, for example, according to an instruction from a user.

[0297]

5 The audio selection unit 193 calculates the position of each object on the image on the basis of the image frame size information and object position information supplied from the audio file acquisition unit 192. The audio selection unit 193 selects an object in the display area designated by
10 the user on the basis of the position of each object on the image. On the basis of the information such as the URL of the audio file supplied from the audio file acquisition unit 192, the audio selection unit 193 requests the Web server 142 to transmit the audio stream of the Object audio track of the
15 selected object in the audio file specified by the URL, and acquires the audio stream of the Object audio track. The audio selection unit 193 supplies the audio decoding processing unit 194 with the acquired audio stream of the Object audio track.

20 [0298]

 The audio decoding processing unit 194 decodes the audio stream of the Channel audio track or HOA audio track supplied from the audio file acquisition unit 192, or decodes the audio stream of the Object audio track supplied from the
25 audio selection unit 193. The audio decoding processing unit 194 supplies the audio synthesis processing unit 195 with one of the Channel audio, the HOA audio, and the Object audio which are obtained as a result of decoding.

[0299]

30 The audio synthesis processing unit 195 synthesizes and outputs the Object audio, the Channel audio, or the HOA audio

supplied from the audio decoding processing unit 194, as needed.

[0300]

(Explanation of Process of Video Playback Terminal)

5 Fig. 48 is a flowchart illustrating the channel audio playback process of the streaming playback unit 190 shown in Fig. 47. This channel audio playback process is performed, for example, when the user selects the Channel audio as an object to be played back.

10 [0301]

 In step S221 of Fig. 48, the MPD processing unit 191 analyzes the MPD file supplied from the MPD acquisition unit 91, and specifies "SubRepresentation" of Channel audio of the segment to be played back on the basis of the essential
15 property and codec described in "SubRepresentation". Further, the MPD processing unit 191 extracts, from the MPD file, information such as the URL described in "Segment" for the audio file of the segment to be played back, and supplies the audio file acquisition unit 192 with the extracted
20 information.

[0302]

 In step S222, the MPD processing unit 191 specifies the level of the Base track, which is a reference track, on the basis of the dependencyLevel of "SubRepresentation"
25 specified in step S221, and supplies the audio file acquisition unit 192 with the specified level of the Base track.

[0303]

 In step S223, the audio file acquisition unit 192
30 requests the Web server 142 to transmit Initial Segment of the segment to be played back on the basis of the information

such as the URL supplied from the MPD processing unit 191,
and acquires the Initial Segment.

[0304]

In step S224, the audio file acquisition unit 192
5 acquires, from the Level assignment box in the Initial Segment,
the track IDs corresponding to the levels of the channel audio
track and the Base track which is a reference track.

[0305]

In step S225, the audio file acquisition unit 192
10 acquires the sample entry of the Initial Segment in the trak
box corresponding to the track ID of Initial Segment, on the
basis of the track IDs of the channel audio track and the Base
track which is a reference track. The audio file acquisition
unit 192 supplies the audio decoding processing unit 194 with
15 the codec information included in the acquired sample entry.

[0306]

In step S226, on the basis of the information such as
the URL supplied from the MPD processing unit 191, the audio
file acquisition unit 192 sends a request to the Web server
20 142 and acquires the sidx box and the ssix box from the head
of the audio file of the segment to be played back.

[0307]

In step S227, the audio file acquisition unit 192
acquires the position information of the reference track and
25 the channel audio track of the segment to be played back, from
the sidx box and the ssix box which are acquired in step S223.
In this case, since the Base track which is a reference track
does not include any audio stream, there is no position
information of the reference track.

30 [0308]

In step S228, the audio file acquisition unit 192

requests the Web server 142 to transmit the audio stream of the channel audio track arranged in the mdat box, on the basis of the position information of the channel audio track and the information such as the URL of the audio file of the segment to be played back, and acquires the audio stream of the channel audio track. The audio file acquisition unit 192 supplies the audio decoding processing unit 194 with the acquired audio stream of the channel audio track.

[0309]

10 In step S229, the audio decoding processing unit 194 decodes the audio stream of the channel audio track on the basis of the codec information supplied from the audio file acquisition unit 192. The audio file acquisition unit 192 supplies the audio synthesis processing unit 195 with the channel audio obtained as a result of decoding.

[0310]

 In step S230, the audio synthesis processing unit 195 outputs the channel audio. Then, the process is terminated.

[0311]

20 Note that, although not shown, an HOA audio playback process for playing back the HOA audio by the streaming playback unit 190 is performed in a manner similar to the channel audio playback process shown in Fig. 48.

[0312]

25 Fig. 49 is a flowchart illustrating the object specifying process of the streaming playback unit 190 shown in Fig. 47. This object specifying process is performed, for example, when the user selects the Object audio as an object to be played back and the playback area is changed.

30 [0313]

 In step S251 of Fig. 49, the audio selection unit 193

acquires the display area designated by the user through the user's operation or the like.

[0314]

In step S252, the MPD processing unit 191 analyzes the
5 MPD file supplied from the MPD acquisition unit 91, and
specifies "SubRepresentation" of metadata of the segment to
be played back, on the basis of the essential property and
codec described in "SubRepresentation". Further, the MPD
processing unit 191 extracts, from the MPD file, information
10 such as the URL of the audio file of the segment to be played
back that is described in "Segment" for audio file, and
supplies the audio file acquisition unit 192 with the
extracted information.

[0315]

15 In step S253, the MPD processing unit 191 specify the
level of the Base track, which is a reference track, on the
basis of the dependencyLevel of "SubRepresentation"
specified in step S252, and supplies the audio file
acquisition unit 192 with the specified level of the Base
20 track.

[0316]

In step S254, the audio file acquisition unit 192
requests the Web server 142 to transmit Initial Segment of
the segment to be played back, on the basis of the information
25 such as the URL supplied from the MPD processing unit 191,
and acquires the Initial Segment.

[0317]

In step S255, the audio file acquisition unit 192
acquires, from the Level assignment box in the Initial Segment,
30 the track IDs corresponding to the levels of the object
metadata track and the Base track which is a reference track.

[0318]

In step S256, the audio file acquisition unit 192 acquires the sample entry of Initial Segment in the trak box corresponding to the track ID of the Initial Segment on the basis of the track IDs of the object metadata track and the Base track which is a reference track. The audio file acquisition unit 192 supplies the audio selection unit 193 with the image frame size information included in the sample entry of the Base track which is a reference track. Further, the audio file acquisition unit 192 supplies the audio selection unit 193 with the Initial Segment.

[0319]

In step S257, on the basis of the information such as the URL supplied from the MPD processing unit 191, the audio file acquisition unit 192 sends a request to the Web server 142 and acquires the sidx box and the ssix box from the head of the audio file of the segment to be played back.

[0320]

In step S258, the audio file acquisition unit 192 acquires, from the sidx box and ssix box acquired in step S257, the position information of the reference track and the object metadata track of the subsegment to be played back. In this case, since the Base track which is a reference track does not include any audio stream, there is no position information of the reference track. The audio file acquisition unit 192 supplies the audio selection unit 193 with the sidx box and the ssix box.

[0321]

In step S259, the audio file acquisition unit 192 requests the Web server 142 to transmit the audio stream of the object metadata track arranged in the mdat box, on the

basis of the position information of the object metadata track and the information such as the URL of the audio file of the segment to be played back, and acquires the audio stream of the object metadata track.

5 [0322]

In step S260, the audio file acquisition unit 192 decodes the audio stream of the object metadata track acquired in step S259, on the basis of the codec information included in the sample entry acquired in step S256. The audio file acquisition unit 192 supplies the audio selection unit 193 with the object position information included in the metadata obtained as a result of decoding. Further, the audio file acquisition unit 192 supplies the audio selection unit 193 with the information such as the URL of the audio file supplied from the MPD processing unit 191.

[0323]

In step S261, the audio selection unit 193 selects an object in the display area on the basis of the image frame size information and object position information supplied from the audio file acquisition unit 192 and on the basis of the display area designated by the user. Then, the process is terminated.

[0324]

Fig. 50 is a flowchart illustrating a specific object audio playback process performed by the streaming playback unit 190 after the object specifying process shown in Fig. 49.

[0325]

In step S281 of Fig. 50, the MPD processing unit 191 analyzes the MPD file supplied from the MPD acquisition unit 91, and specifies "SubRepresentation" of the object audio of

the selected object on the basis of the essential property and codec described in "SubRepresentation".

[0326]

5 In step S282, the MPD processing unit 191 specifies the level of the Base track, which is a reference track, on the basis of the dependencyLevel of "SubRepresentation" specified in step S281, and supplies the audio file acquisition unit 192 with the specified level of the Base track.

10 [0327]

In step S283, the audio file acquisition unit 192 acquires, from the Level assignment box in the Initial Segment, the track IDs corresponding to the levels of the object audio track and the Base track which is a reference track, and supplies the audio selection unit 193 with the track IDs.

[0328]

20 In step S284, the audio selection unit 193 acquires the sample entry of Initial Segment in the trak box corresponding to the track ID of the Initial Segment, on the basis of the track IDs of the object audio track and the Base track which is a reference track. This Initial Segment is supplied from the audio file acquisition unit 192 in step S256 shown in Fig. 49. The audio selection unit 193 supplies the audio decoding processing unit 194 with the codec information included in the acquired sample entry.

[0329]

30 In step S285, the audio selection unit 193 acquires, from the sidx box and ssix box supplied from the audio file acquisition unit 192 in step S258, the position information of the reference track and the object audio track of the selected object of the subsegment to be played back. In this

case, since the Base track which is a reference track does not include any audio stream, there is no position information of the reference track.

[0330]

5 In step S286, the audio selection unit 193 requests the Web server 142 to transmit the audio stream of the object audio track of the selected object, which is arranged in the mdat box, on the basis of the position information of the object audio track and the information such as the URL of the audio
10 file of the segment to be played back, and acquires the audio stream of the object audio track. The audio selection unit 193 supplies the audio decoding processing unit 194 with the acquired audio stream of the object audio track.

[0331]

15 In step S287, the audio decoding processing unit 194 decodes the audio stream of the object audio track on the basis of the codec information supplied from the audio selection unit 193. The audio selection unit 193 supplies the audio synthesis processing unit 195 with the object audio obtained
20 as a result of decoding.

[0332]

 In step S288, the audio synthesis processing unit 195 synthesizes and outputs the object audio supplied from the audio decoding processing unit 194. Then, the process is
25 terminated.

[0333]

 As described above, in the information processing system 140, the file generation device 141 generates an audio file in which 3D audio is divided into a plurality of tracks
30 depending on the types of the 3D audio and the tracks are arranged. The video playback terminal 144 acquires the audio

stream of a predetermined type of 3D audio in the audio file. Accordingly, the video playback terminal 144 can efficiently acquire the audio stream of the predetermined types of 3D audio. Therefore, it can be said that the file generation
5 device 141 generates the audio file capable of improving the efficiency of acquiring the audio stream of the predetermined type of 3D audio.

[0334]

<Second Embodiment>

10 (Outline of Tracks)

Fig. 51 is a diagram illustrating the outline of tracks in a second embodiment to which the present disclosure is applied.

[0335]

15 As shown in Fig. 51, the second embodiment differs from the first embodiment in that the base sample is recorded as a sample of Base track. The base sample is formed of information to be referred to for the sample of Channel
20 audio/Object audio/HOA audio/metadata. The sample of Channel audio/Object audio/HOA audio/metadata to be referred to by the reference information included in the base sample is arranged in the order of arrangement of the reference information, thereby making it possible to generate the audio stream of 3D audio before the 3D audio is divided into tracks.

25 [0336]

(Exemplary Syntax of Sample Entry of Base Track)

Fig. 52 is a diagram showing an exemplary syntax of the sample entry of the base track shown in Fig. 51.

[0337]

30 The syntax shown in Fig. 52 is the same as the syntax shown in Fig. 34, except that "mha2" representing that the

sample entry is the sample entry of the Base track shown in Fig. 51 is described instead of "mha1" representing that the sample entry is the sample entry of the Base track shown in Fig. 33.

5 [0338]

(Exemplary Structure of Base Sample)

Fig. 53 is a diagram showing an exemplary structure of the base sample.

[0339]

10 As shown in Fig. 53, the base sample is configured using the extractor of Channel audio/Object audio/HOA audio/metadata in units of samples as a sub-sample. The extractor of Channel audio/Object audio/HOA audio/metadata is composed of the type of the extractor, and the offset and
15 size of the sub-sample of the corresponding Channel audio track/Object audio track(s)/HOA audio track/Object metadata track. This offset is a difference between the position of the base sample in the file of sub-sample of the base sample and the position of Channel audio track/Object audio
20 track(s)/HOA audio track/Object metadata track in the file of the sample. In other words, the offset is information indicating the position within the file of a sample of another track corresponding to the sub-sample of the base sample including the offset.

25 [0340]

Fig. 54 is a diagram showing an exemplary syntax of the base sample.

[0341]

As shown in Fig. 54, in the base sample, SCE element
30 for storing the object audio in the sample of the Object audio track is replaced by EXT element for storing the extractor.

[0342]

Fig. 55 is a diagram showing an example of extractor data.

[0343]

5 As shown in Fig. 55, the type of the extractor and the offset and size of the sub-sample of the corresponding Channel audio track/Object audio track(s)/HOA audio track/Object metadata track are described in the extractor.

[0344]

10 Note that the extractor may be extended by utilizing Network Abstraction Layer (NAL) structure, which is defined in Advanced Video Coding (AVC)/High Efficiency Video Coding (HEVC), so that audio elementary and config information can be stored.

15 [0345]

The information processing system and the process performed by the information processing system in the second embodiment are similar to those of the first embodiment, and thus the descriptions thereof are omitted.

20 [0346]

<Third Embodiment>

(Outline of Tracks)

25 Fig. 56 is a diagram illustrating the outline of tracks in a third embodiment to which the present disclosure is applied.

[0347]

30 As shown in Fig. 56, the third embodiment differs from the first embodiment in that the base sample and the sample of metadata are recorded as the sample of the Base track and the Object metadata track is not provided.

[0348]

The information processing system and the process performed by the information processing system in the third embodiment are similar to those of the first embodiment, except that the audio stream of the Base track is acquired instead of the Object metadata track so as to acquire the object position information. Accordingly, the descriptions thereof are omitted.

[0349]

<Fourth Embodiment>

10 (Outline of Tracks)

Fig. 57 is a diagram illustrating the outline of tracks in a fourth embodiment to which the present disclosure is applied.

[0350]

15 As shown in Fig. 57, the fourth embodiment differs from the first embodiment in that the tracks are recoded as different files

(3da_base.mp4/3da_channel.mp4/3da_object_1.mp4/3da_hoa.mp4/3da_meta.mp4). In this case, only the audio data of a desired track can be acquired by acquiring a file of a desired track via HTTP. Accordingly, the audio data of a desired track can be efficiently acquired via HTTP.

[0351]

(Exemplary Description of MPD File)

25 Fig. 58 is a diagram showing an exemplary description of the MPD file according to the fourth embodiment to which the present disclosure is applied.

[0352]

As shown in Fig. 58, "Representation" or the like that manages the segment of each audio file

(3da_base.mp4/3da_channel.mp4/3da_object_1.mp4/3da_hoa.mp

4/3da_meta.mp4) of 3D audio is described in the MPD file.
[0353]

The "Representation" includes "codecs", "id",
"associationId", and "associationType". Further, the
5 "Representation" of Channel audio track/Object audio
track(s)/HOA audio track/Object metadata track also includes
"<EssentialProperty schemeIdUri=
"urn:mpeg:DASH:3daudio:2014" value=" audioType,
contentkind, priority">". Further, the "Representation" of
10 Object audio track(s) includes <EssentialProperty
schemeIdUri="urn:mpeg:DASH:viewingAngle:2014" value="θ, γ,
r">.
[0354]

(Outline of Information Processing System)

15 Fig. 59 is a diagram illustrating the outline of the
information processing system in the fourth embodiment to
which the present disclosure is applied.
[0355]

The components shown in Fig. 59 that are the same as
20 the components shown in Fig. 1 are denoted by the same
reference numerals. Repeated explanation is omitted as
appropriate.
[0356]

The information processing system 210 shown in Fig. 59
25 has a configuration in which a Web server 212, which is
connected to a file generation device 211, and a video
playback terminal 214 are connected via the Internet 13.
[0357]

In the information processing system 210, the Web
30 server 212 delivers (tilted streaming) a video stream of video
content to the video playback terminal 214 in units of tiles

by a method in conformity with MPEG-DASH. Further, in the information processing system 210, the Web server 212 delivers the audio file of Object audio, Channel audio, or HOA audio corresponding to the file to be played back to the video playback terminal 214.

[0358]

Specifically, the file generation device 211 acquires the image data of video content and encodes the image data in units of tiles to generate a video stream. The file generation device 211 processes the video stream of each tile into a file format for each segment. The file generation device 211 uploads the image file of each file obtained as a result of the above process to the Web server 212.

[0359]

Further, the file generation device 211 acquires the 3D audio of video content, and encodes the 3D audio for each type (Channel audio/Object audio/HOA audio/metadata) of the 3D audio to generate an audio stream. The file generation device 211 allocates the tracks to the audio stream for each type of the 3D audio. The file generation device 211 generates an audio file in which the audio stream is arranged for each track, and uploads the generated audio file to the Web server 212.

[0360]

The file generation device 211 generates the MPD file including the image frame size information, the tile position information, and the object position information. The file generation device 211 uploads the MPD file to the Web server 212.

[0361]

The Web server 212 stores the image file uploaded from

the file generation device 211, the audio file for each type of 3D audio, and the MPD file.

[0362]

In the example of Fig. 59, the Web server 212 stores
5 a segment group formed of image files of a plurality of segments of the tile #1, and a segment group formed of image files of a plurality of segments of the tile #2. The Web server 212 also stores a segment group formed of the audio file of Channel audio and a segment group of the audio file
10 of the object #1.

[0363]

The Web server 212 transmits, to the video playback terminal 214, the image file, the predetermined type of audio file of 3D audio, the MPD file, and the like, which are stored
15 in the Web server, in response to a request from the video playback terminal 214.

[0364]

The video playback terminal 214 executes control software 221, video playback software 222, access software
20 223, and the like.

[0365]

The control software 221 is software for controlling data to be streamed from the Web server 212. Specifically, the control software 221 causes the video playback terminal
25 214 to acquire the MPD file from the Web server 212.

[0366]

Further, the control software 221 specifies a tile in the MPD file on the basis of the display area instructed from the video playback software 222 and the tile position
30 information included in the MPD file. Then, the control software 221 instructs the access software 223 to send a

request for transmitting the image file of the tile.

[0367]

When Object audio is to be played back, the control software 221 instructs the access software 223 to send a request for transmitting the audio file of the Base track. Further, the control software 221 instructs the access software 223 to send a request for transmitting the audio file of the Object metadata track. The control software 221 acquires the image frame size information in the audio file of the Base track, which is transmitted from the Web server 142 according to the instruction, and the object position information included in the audio file of metadata. The control software 221 specifies the object corresponding to the image in the display area on the basis of the image frame size information, the object position information, and the display area. Further, the control software 221 instructs the access software 223 to send a request for transmitting the audio file of the object.

[0368]

Further, when Channel audio or HOA audio is to be played back, the control software 221 instructs the access software 223 to send a request for transmitting the audio file of the Channel audio or HOA audio.

[0369]

The video playback software 222 is software for playing back the image file and audio file acquired from the Web server 212. Specifically, when the display area is specified by the user, the video playback software 222 gives an instruction on the display area to the control software 221. Further, the video playback software 222 decodes the image file and audio file acquired from the Web server 212 according to the

instruction. The video playback software 222 synthesizes and outputs the image data in units of tiles obtained as a result of decoding. Further, the video playback software 222 synthesizes and outputs, as needed, the Object audio, Channel
5 audio, or HOA audio obtained as a result of decoding.

[0370]

The access software 223 is software for controlling the communication with the Web server 212 via the Internet 13 using HTTP. Specifically, the access software 223 causes the
10 video playback terminal 214 to transmit a request for transmitting the image file and the predetermined audio file in response to an instruction from the control software 221. Further, the access software 223 causes the video playback terminal 214 to receive the image file and the predetermined
15 audio file, which are transmitted from the Web server 212, according to the transmission request.

[0371]

(Configuration Example of File Generation Device)

Fig. 60 is a block diagram of the file generation device
20 211 shown in Fig. 59.

[0372]

The components shown in Fig. 60 that are the same as the components shown in Fig. 45 are denoted by the same reference numerals. Repeated explanation is omitted as
25 appropriate.

[0373]

The configuration of the file generation device 211 shown in Fig. 60 is different from the configuration of the file generation device 141 shown in Fig. 45 in that an audio
30 file generation unit 241, an MPD generation unit 242, and a server upload processing unit 243 are provided instead of the

audio file generation unit 172, the MPD generation unit 173, and the server upload processing unit 174, respectively.

[0374]

Specifically, the audio file generation unit 241 of the
5 file generation device 211 allocates the tracks to the audio
stream, which is supplied from the audio coding processing
unit 171, for each type of the 3D audio. The audio file
generation unit 241 generates an audio file in which the audio
stream is arranged for each track. At this time, the audio
10 file generation unit 241 stores the image frame size
information input from the outside in the sample entry of the
Base track. The audio file generation unit 241 supplies the
MPD generation unit 242 with the audio file for each type of
the 3D audio.

15 [0375]

The MPD generation unit 242 determines the URL or the
like of the Web server 212 that stores the image file of each
tile supplied from the image file generation unit 53. Further,
the MPD generation unit 242 determines, for each type of the
20 3D audio, the URL or the like of the Web server 212 that stores
the audio file supplied from the audio file generation unit
241.

[0376]

The MPD generation unit 242 arranges, in
25 "AdaptationSet" for the image of the MPD file, the image
information supplied from the image information generation
unit 54. Further, the MPD generation unit 242 arranges the
URL or the like of the image file of each tile in "Segment"
of "Representation" for the image file of the tile.,

30 [0377]

The MPD generation unit 242 arranges, for each type of

the 3D audio, the URL or the like of the audio file in "Segment" of "Representation" for the audio file. Further, the MPD generation unit 242 arranges the object position information or the like of each object input from the outside in

5 "Representation" for the Object metadata track of the object. The MPD generation unit 242 supplies the server upload processing unit 243 with the MPD file, in which various pieces of information are arranged as described above, the image file, and the audio file for each type of the 3D audio.

10 [0378]

The server upload processing unit 243 uploads the image file of each tile supplied from the MPD generation unit 242, the audio file for each type of the 3D audio, and the MPD file to the Web server 212.

15 [0379]

(Explanation of Process of File Generation Device)

Fig. 61 is a flowchart illustrating a file generation process of the file generation device 211 shown in Fig. 60.

[0380]

20 The process of steps S301 to S307 shown in Fig. 61 is similar to the process of steps S191 to S197 shown in Fig. 46, and thus the description thereof is omitted.

[0381]

In step S308, the audio file generation unit 241
25 generates an audio file in which an audio stream is arranged for each track. At this time, the audio file generation unit 241 stores the image frame size information input from the outside in the sample entry in the audio file of the Base track. The audio file generation unit 241 supplies the MPD generation
30 unit 242 with the generated audio file for each type of the 3D audio.

[0382]

In step S309, the MPD generation unit 242 generates an MPD file including the image information supplied from the image information generation unit 54, the URL of each file, and the object position information. The MPD generation unit 242 supplies the server upload processing unit 243 with the image file, the audio file for each type of the 3D audio, and the MPD file.

[0383]

In step S310, the server upload processing unit 243 uploads the image file supplied from the MPD generation unit 242, the audio file for each type of the 3D audio, and the MPD file to the Web server 212. Then, the process is terminated.

[0384]

(Functional Configuration Example of Video Playback Terminal)

Fig. 62 is a block diagram showing a configuration example of a streaming playback unit which is implemented in such a manner that the video playback terminal 214 shown in Fig. 59 executes the control software 221, the video playback software 222, and the access software 223.

[0385]

The components shown in Fig. 62 that are the same as the components shown in Figs. 13 and 47 are denoted by the same reference numerals. Repeated explanation is omitted as appropriate.

[0386]

The configuration of the streaming playback unit 260 shown in Fig. 62 is different from the configuration of the streaming playback unit 90 shown in Fig. 13 in that an MPD

processing unit 261, a metafile acquisition unit 262, an audio selection unit 263, an audio file acquisition unit 264, an audio decoding processing unit 194, and an audio synthesis processing unit 195 are provided instead of the MPD processing unit 92, the metafile acquisition unit 93, the audio selection unit 94, the audio file acquisition unit 95, the audio decoding processing unit 96, and the audio synthesis processing unit 97, respectively.

[0387]

Specifically, when Object audio is to be played back, the MPD processing unit 261 of the streaming playback unit 260 extracts, from the MPD file supplied from the MPD acquisition unit 91, information such as the URL described in "Segment" of the audio file of the object metadata track of the segment to be played back, and supplies the metafile acquisition unit 262 with the extracted information. Further, the MPD processing unit 261 extracts, from the MPD file, information such as the URL described in "Segment" of the audio file of the object audio track of the object requested from the audio selection unit 263, and supplies the audio selection unit 263 with the extracted information. Furthermore, the MPD processing unit 261 extracts, from the MPD file, information such as the URL described in "Segment" of the audio file of the Base track of the segment to be played back, and supplies the metafile acquisition unit 262 with the extracted information.

[0388]

Further, when Channel audio or HOA audio is to be played back, the MPD processing unit 261 extracts, from the MPD file, information such as the URL described in "Segment" of the audio file of the Channel audio track or HOA audio track of

the segment to be played back. The MPD processing unit 261 supplies the audio file acquisition unit 264 with the information such as the URL via the audio selection unit 263. [0389]

5 Note that it is determined which one of Object audio, Channel audio, and HOA audio is to be played back, for example, according to an instruction from a user. [0390]

10 The MPD processing unit 261 extracts, from the MPD file, the tile position information described in "AdaptationSet" for image, and supplies the image selection unit 98 with the extracted tile position information. The MPD processing unit 261 extracts, from the MPD file, information such as the URL described in "Segment" for the image file of the tile
15 requested from the image selection unit 98, and supplies the image selection unit 98 with the extracted information. [0391]

 On the basis of the information such as the URL supplied from the MPD processing unit 261, the metafile acquisition
20 unit 262 requests the Web server 212 to transmit the audio file of the object metadata track specified by the URL, and acquires the audio file of the object metadata track. The metafile acquisition unit 93 supplies the audio selection unit 263 with the object position information included in the
25 audio file of the object metadata track. [0392]

 Further, on the basis of the information such as the URL of the audio file, the metafile acquisition unit 262 requests the Web server 142 to transmit the Initial Segment
30 of the audio file of the Base track specified by the URL, and acquires the Initial Segment. The metafile acquisition unit

262 supplies the audio selection unit 263 with the image frame size information included in the sample entry of the Initial Segment.

[0393]

5 The audio selection unit 263 calculates the position of each object on the image on the basis of the image frame size information and the object position information supplied from the metafile acquisition unit 262. The audio selection unit 263 selects an object in the display area designated by
10 the user, on the basis of the position of each object on the image. The audio selection unit 263 requests the MPD processing unit 261 to transmit the information such as the URL of the audio file of the object audio track of the selected object. The audio selection unit 263 supplies the audio file
15 acquisition unit 264 with the information such as the URL supplied from the MPD processing unit 261 according to the request.

[0394]

 On the basis of the information, such as the URL of the
20 audio file of the object audio track, Channel audio track, or HOA audio track supplied from the audio selection unit 263, the audio file acquisition unit 264 requests the Web server 12 to transmit the audio stream of the audio file specified by the URL, and acquires the audio stream of the audio file.
25 The audio file acquisition unit 95 supplies the audio decoding processing unit 194 with the acquired audio file in units of objects.

[0395]

(Explanation of Process of Video Playback Terminal)

30 Fig. 63 is a flowchart illustrating a channel audio playback process of the streaming playback unit 260 shown in

Fig. 62. This channel audio playback process is performed, for example, when Channel audio is selected by the user as an object to be played back.

[0396]

5 In step S331 of Fig. 63, the MPD processing unit 261 analyzes the MPD file supplied from the MPD acquisition unit 91, and specifies "Representation" of the Channel audio of the segment to be played back on the basis of the essential property and codec described in "Representation". Further,
10 the MPD processing unit 261 extracts information such as the URL of the audio file of the Channel audio track of the segment to be played back that is described in "Segment" included in the "Representation", and supplies the audio file acquisition unit 264 with the extracted information via the audio
15 selection unit 263.

[0397]

 In step S332, the MPD processing unit 261 specifies "Representation" of the Base track, which is a reference track, on the basis of the associationId of "Representation"
20 specified in step S331. The MPD processing unit 261 extracts information such as the URL of the audio file of the reference track described in "Segment" included in the "Representation", and supplies the audio file acquisition unit 264 with the extracted file via the audio selection unit 263.

25 [0398]

 In step S333, the audio file acquisition unit 264 requests the Web server 212 to transmit the Initial Segment of the audio files of the Channel audio track of the segment to be played back and the reference track on the basis of the
30 information such as the URL supplied from the audio selection unit 263, and acquires the Initial Segment.

[0399]

In step S334, the audio file acquisition unit 264 acquires the sample entry in the trak box of the acquired Initial Segment. The audio file acquisition unit 264
5 supplies the audio decoding processing unit 194 with the codec information included in the acquired sample entry.

[0400]

In step S335, the audio file acquisition unit 264 sends a request to the Web server 142 on the basis of the information
10 such as the URL supplied from the audio selection unit 263, and acquires the sidx box and the ssix box from the head of the audio file of the Channel audio track of the segment to be played back.

[0401]

15 In step S336, the audio file acquisition unit 264 acquires the position information of the subsegment to be played back from the sidx box and ssix box acquired in step S333.

[0402]

20 In step S337, the audio selection unit 263 requests the Web server 142 to transmit the audio stream of the channel audio track arranged in the mdat box in the audio file, on the basis of the position information acquired in step S337 and the information such as the URL of the audio file of the
25 channel audio track of the segment to be played back, and acquires the audio stream of the channel audio track. The audio selection unit 263 supplies the audio decoding processing unit 194 with the acquired audio stream of the channel audio track.

30 [0403]

In step S338, the audio decoding processing unit 194

decodes the audio stream of the channel audio track supplied from the audio selection unit 263 on the basis of the codec information supplied from the audio file acquisition unit 264. The audio selection unit 263 supplies the audio synthesis processing unit 195 with the channel audio obtained as a result of decoding.

[0404]

In step S339, the audio synthesis processing unit 195 outputs the channel audio. Then, the process is terminated.

10 [0405]

Although not shown, the HOA audio playback process for playing back HOA audio by the streaming playback unit 260 is performed in a manner similar to the channel audio playback process shown in Fig. 63.

15 [0406]

Fig. 64 is a flowchart illustrating an object audio playback process of the streaming playback unit 260 shown in Fig. 62. This object audio playback process is performed, for example, when the user selects Object audio as an object to be played back and the playback area is changed.

20

[0407]

In step S351 of Fig. 64, the audio selection unit 263 acquires the display area designated by the user through the user's operation or the like.

25 [0408]

In step S352, the MPD processing unit 261 analyzes the MPD file supplied from the MPD acquisition unit 91, and specifies "Representation" of the metadata of the segment to be played back, on the basis of the essential property and codec described in "Representation". Further, the MPD processing unit 261 extracts information such as the URL of

30

the audio file of the object metadata track of the segment to be played back that is described in "Segment" included in the "Representation", and supplies the metafile acquisition unit 262 with the extracted information.

5 [0409]

In step S353, the MPD processing unit 261 specifies "Representation" of the Base track, which is a reference track, on the basis of the associationId of "Representation" specified in step S352. The MPD processing unit 261 extracts
10 information such as the URL of the audio file of the reference track described in "Segment" included in the "Representation", and supplies the metafile acquisition unit 262 with the extracted information.

[0410]

15 In step S354, the metafile acquisition unit 262 requests the Web server 212 to transmit the Initial Segment of the audio files of the object metadata track of the segment to be played back and the reference track, on the basis of the information such as the URL supplied from the MPD
20 processing unit 261, and acquires the Initial Segment.

[0411]

In step S355, the metafile acquisition unit 262 acquires the sample entry in the trak box of the acquired Initial Segment. The metafile acquisition unit 262 supplies
25 the audio file acquisition unit 264 with the image frame size information included in the sample entry of the Base track which is a reference track.

[0412]

In step S356, the metafile acquisition unit 262 sends
30 a request to the Web server 142 on the basis of the information such as the URL supplied from the MPD processing unit 261,

and acquires the sidx box and the ssix box from the head of the audio file of the object metadata track of the segment to be played back.

[0413]

5 In step S357, the metafile acquisition unit 262 acquires the position information of the subsegment to be played back from the sidx box and ssix box acquired in step S356.

[0414]

10 In step S358, the metafile acquisition unit 262 requests the Web server 142 to transmit the audio stream of the object metadata track arranged in the mdat box in the audio file, on the basis of the position information acquired in step S357 and the information such as the URL of the audio
15 file of the object metadata track of the segment to be played back, acquires the audio stream of the object metadata track.

[0415]

 In step S359, the metafile acquisition unit 262 decodes the audio stream of the object metadata track acquired in step
20 S358, on the basis of the codec information included in the sample entry acquired in step S355. The metafile acquisition unit 262 supplies the audio selection unit 263 with the object position information included in the metadata obtained as a result of decoding.

25 [0416]

 In step S360, the audio selection unit 263 selects an object in the display area on the basis of the image frame size information and object position information supplied from the metafile acquisition unit 262 and on the basis of
30 the display area designated by the user. The audio selection unit 263 requests the MPD processing unit 261 to transmit the

information such as the URL of the audio file of the object audio track of the selected object.

[0417]

In step S361, the MPD processing unit 261 analyzes the MPD file supplied from the MPD acquisition unit 91, and specifies "Representation" of the object audio of the selected object on the basis of the essential property and codec described in "Representation". Further, the MPD processing unit 261 extracts information such as the URL of the audio file of the object audio track of the selected object of the segment to be played back that is described in "Segment" included in the "Representation", and supplies the audio file acquisition unit 264 with the extracted information via the audio selection unit 263.

15 [0418]

In step S362, the MPD processing unit 261 specifies "Representation" of the Base track, which is a reference track, on the basis of the associationId of "Representation" specified in step S361. The MPD processing unit 261 extracts information such as the URL of the audio file of the reference track described in "Segment" included in the "Representation", and supplies the audio file acquisition unit 264 with the extracted information via the audio selection unit 263.

[0419]

25 In step S363, the audio file acquisition unit 264 requests the Web server 212 to transmit the Initial Segment of the audio files of the object audio track of the segment to be played back and the reference track, on the basis of the information such as the URL supplied from the audio selection unit 263, and acquires the Initial Segment.

30 [0420]

In step S364, the audio file acquisition unit 264 acquires the sample entry in the trak box of the acquired Initial Segment. The audio file acquisition unit 264 supplies the audio decoding processing unit 194 with the codec information included in the sample entry.
[0421]

In step S365, the audio file acquisition unit 264 sends a request to the Web server 142 on the basis of the information such as the URL supplied from the audio selection unit 263, and acquires the sidx box and the ssix box from the head of the audio file of the object audio track of the segment to be played back.
[0422]

In step S366, the audio file acquisition unit 264 acquires the position information of the subsegment to be played back from the sidx box and ssix box acquired in step S365.
[0423]

In step S367, the audio file acquisition unit 264 requests the Web server 142 to transmit the audio stream of the object audio track arranged in the mdat box within the audio file, on the basis of the position information acquired in step S366 and the information such as the URL of the audio file of the object audio track of the segment to be played back, and acquires the audio stream of the object audio track. The audio file acquisition unit 264 supplies the audio decoding processing unit 194 with the acquired audio stream of the object audio track.
[0424]

The process of steps S368 and S369 is similar to the process of steps S287 and S288 shown in Fig. 50, and thus the

description thereof is omitted.

[0425]

Note that in the above description, the audio selection unit 263 selects all objects in the display area. However, 5 the audio selection unit 263 may select only objects with a high processing priority in the display area, or may select only an audio object of a predetermined content.

[0426]

Fig. 65 is a flowchart illustrating an object audio 10 playback process when the audio selection unit 263 selects only objects with a high processing priority among the objects in the display area.

[0427]

The object audio playback process shown in Fig. 65 is 15 similar to the object audio playback process shown in Fig. 64, except that the process of step S390 shown in Fig. 65 is performed instead of step S360 shown in Fig. 64. Specifically, the process of steps S381 to S389 and steps S391 to S399 shown in Fig. 65 is similar to the process of steps S351 to S359 20 and steps S361 to S369 shown in Fig. 64. Accordingly, only the process of step S390 will be described below.

[0428]

In step S390 shown in Fig. 65, the audio file 25 acquisition unit 264 selects an object with a high processing priority in the display area on the basis of the image frame size information, the object position information, the display area, and the priority of each object. Specifically, the audio file acquisition unit 264 specifies each object with the display area on the basis of the image frame size 30 information, the object position information, and the display area. The audio file acquisition unit 264 selects, from among

the specified objects, an object having a priority equal to or higher than a predetermined value. Note that, for example, the MPD processing unit 261 analyzes the MPD file, thereby acquiring the priority from "Representation" of the object audio of the specified object. The audio selection unit 263 requests the MPD processing unit 261 to transmit information such as the URL of the audio file of the object audio track of the selected object.

[0429]

10 Fig. 66 is a flowchart illustrating the object audio playback process when the audio selection unit 263 selects only the audio object of the predetermined content with a high processing priority among the objects in the display area.

[0430]

15 The object audio playback process shown in Fig. 66 is similar to the object audio playback process shown in Fig. 64, except that the process of step S420 shown in Fig. 66 is performed instead of step S360 shown in Fig. 64. Specifically, the process of steps S381 to S389 and steps S391 to S399 shown in Fig. 66 is similar to the process of steps S411 to S419 and steps S421 to S429 shown in Fig. 64. Accordingly, only the process of step S420 will be described below.

[0431]

25 In step S420 shown in Fig. 66, the audio file acquisition unit 264 selects the audio object of the predetermined content with a high processing priority in the display area on the basis of the image frame size information, the object position information, the display area, the priority of each object, and the contentkind of each object. Specifically, the audio file acquisition unit 264 specifies each object in the display area on the basis of the image frame

30

size information, the object position information, and the display area. The audio file acquisition unit 264 selects, from among the specified objects, an object that has a priority equal to or higher than a predetermined value and
5 has a contentkind indicated by a predetermined value.
[0432]

Note that, for example, the MPD processing unit 261 analyzes the MPD file, thereby acquiring the priority and contentkind from "Representation" of the object audio of the
10 specified object. The audio selection unit 263 requests the MPD processing unit 261 to transmit information such as the URL of the audio file of the object audio track of the selected object.
[0433]

15 Fig. 67 is a diagram showing an example of the object selected on the basis of the priority.
[0434]

In the example of Fig. 67, the objects #1 (object1) to #4 (object4) are objects in the display area, and objects
20 having a priority equal to or lower than 2 are selected from among the objects in the display area. Assume that the smaller the number, the higher the processing priority. Further, in Fig. 67, the circled number represents the value of the priority of the corresponding object.
25 [0435]

In the example shown in Fig. 67, when the priorities of the objects #1 to #4 are 1, 2, 3, and 4, respectively, the object #1 and the object #2 are selected. Further, the priorities of the objects #1 to #4 are changed to 3, 2, 1,
30 and 4, respectively, the object #2 and the object #3 are selected. Further, when the priorities of the objects #1 to

#4 are changed to 3, 4, 1, and 2, the object #3 and the object #4 are selected.

[0436]

As described above, only the audio stream of the object
5 audio of objects having a high processing priority are
selectively acquired from among the objects in the display
area, the frequency band between the Web server 142 (212) and
the video playback terminal 144 (214) is efficiently utilized.
The same holds true when an object is selected on the basis
10 of the contentkind of the object.

[0437]

<Fifth Embodiment>

(Outline of Tracks)

Fig. 68 is a diagram illustrating the outline of tracks
15 in a fifth embodiment to which the present disclosure is
applied.

[0438]

As shown in Fig. 68, the fifth embodiment differs from
the second embodiment in that the tracks are recorded as
20 different files

(3da_base.mp4/3da_channel.mp4/3da_object_1.mp4/3da_hoa.mp
4/3da_meta.mp4).

[0439]

The information processing system and the process
25 performed by the information processing system according to
the fifth embodiment are similar to those of the fourth
embodiment, and thus the descriptions thereof are omitted.

[0440]

<Sixth Embodiment>

30 (Outline of Tracks)

Fig. 69 is a diagram illustrating the outline of tracks

in a sixth embodiment to which the present disclosure is applied.

[0441]

As shown in Fig. 69, the sixth embodiment differs from the third embodiment in that the tracks are recorded as different files

(3da_basemeta.mp4/3da_channel.mp4/3da_object_1.mp4/3da_ho
a.mp4).

[0442]

10 The information processing system and the process performed by the information processing system according to the sixth embodiment are similar to those of the fourth embodiment, except that the audio stream of the Base track is acquired instead of the Object metadata track so as to
15 acquire the object position information. Accordingly, the descriptions thereof are omitted.

[0443]

Note that also in the first to third embodiments, the fifth embodiment, and the sixth embodiment, an object in the
20 display area can be selected on the basis of the priority or contentkind of the object.

[0444]

Further, in the first to sixth embodiments, the streaming playback unit may acquire the audio stream of
25 objects outside the display area and synthesize and output the object audio of the objects, like the streaming playback unit 120 shown in Fig. 23.

[0445]

Further, in the first to sixth embodiments, the object
30 position information is acquired from the metadata, but instead the object position information may be acquired from

the MPD file.

[0446]

<Explanation of Hierarchical Structure of 3D Audio>

Fig. 70 is a diagram showing a hierarchical structure
5 of 3D audio.

[0447]

As shown in Fig. 70, audio elements (Elements) which
are different for each audio data are used as the audio data
of 3D audio. As the types of the audio elements, there are
10 Single Channel Element (SCE) and Channel Pair Element (CPE).
The type of the audio element of audio data for one channel
is SCE, and the type of the audio element corresponding to
the audio data for two channels is CPE.

[0448]

15 The audio elements of the same audio type
(Channel/Object/SAOC Objects/HOA) form a group. Examples of
the group type (GroupType) include Channels, Objects, SAOC
Objects, and HOA. Two or more groups can form a switch Group
or a group Preset, as needed.

20 [0449]

The switch Group defines a group of audio elements to
be exclusively played back. Specifically, as shown in Fig.
70, when an Object audio group for English (EN) and an Object
audio group for French (FR) are present, one of the groups
25 is to be played back. Accordingly, a switch Group is formed
of the Object audio group for English having a group ID of
2 and the Object audio group for French having a group ID of
3. Thus, the Object audio for English and the Object audio
for French are exclusively played back.

30 [0450]

On the other hand, the group Preset defines a

combination of groups intended by a content producer.

[0451]

Ext elements (Ext Elements) which are different for each metadata are used as the metadata of 3D audio. Examples of the type of the Ext elements include Object Metadata, SAOC 3D Metadata, HOA Metadata, DRC Metadata, SpatialFrame, and SaocFrame. Ext elements of Object Metadata are all metadata of Object audio, and Ext elements of SAOC 3D Metadata are all metadata of SAOC audio. Further, Ext elements of HOA Metadata are all metadata of HOA audio, and Ext elements of Dynamic Range Control (DRC) Metadata are all metadata of Object audio, SAOC audio, and HOA audio.

[0452]

As described above, the audio data of 3D audio is divided in units of audio elements, group types, groups, switch Groups, and group Presets. Accordingly, the audio data may be divided into audio elements, groups, switch Groups, or group Presets, instead of dividing the audio data into tracks for each group type (in this case, however, the object audio is divided for each object) like in the first to sixth embodiments.

[0453]

Further, the metadata of 3D audio is divided in units of Ext element type (ExtElementType) or audio element corresponding to the metadata. Accordingly, the metadata may be divided for each audio element corresponding to the metadata, instead of dividing the metadata for each type of Ext element like in the first to sixth embodiments.

[0454]

Assume that, in the following description, audio data is divided for each audio element; metadata is divided for

each type of Ext element; and data of different tracks are arranged. The same holds true when other division units are used.

[0455]

5 <Explanation of First Example of Web Server Process>

Fig. 71 is a diagram illustrating a first example of the process of the Web server 142 (212).

[0456]

In the example of Fig. 71, the 3D audio corresponding
10 to the audio file uploaded from the file generation device 141 (211) is composed of the channel audio of five channels, the object audio of three objects, and metadata of the object audio (Object Metadata).

[0457]

15 The channel audio of five channels is divided into a channel audio of a front center (FC) channel, a channel audio of front left/right (FL, FR) channels, and a channel audio of rear left/right (RL, RR) channels, which are arranged as data of different tracks. Further, the object audio of each
20 object is arranged as data of different tracks. Furthermore, Object Metadata is arranged as data of one track.

[0458]

Further, as shown in Fig. 71, each audio stream of 3D audio is composed of config information and data in units of
25 frames (samples). In the example of Fig. 71, in the audio stream of the audio file, a channel audio of five channels, an object audio of three objects, and config information of Object Metadata are collectively arranged, and data items of each frame are collectively arranged.

30 [0459]

In this case, as shown in Fig. 71, the Web server 142

(212) divides, for each track, the audio stream of the audio file uploaded from the file generation device 141 (211), and generates the audio stream of seven tracks. Specifically, the Web server 142 (212) extracts, from the audio stream of the audio file, the config information of each track and audio data according to the information such as the ssix box, and the audio stream of each track is generated. The audio stream of each track is composed of the config information of the track and the audio data of each frame.

10 [0460]

Fig. 72 is a flowchart illustrating a track division process of the Web server 142 (212). This track division process is started, for example, when the audio file is uploaded from the file generation device 141 (211).

15 [0461]

In step S441 shown in Fig. 72, the Web server 142 (212) stores the audio file uploaded from the file generation device 141.

[0462]

20 In step S442, the Web server 142 (212) divides the audio stream constituting the audio file for each track according to the information such as the ssix box of the audio file.

[0463]

In step S443, the Web server 142 (212) holds the audio stream of each track. Then, the process is terminated. This audio stream is transmitted to the video playback terminal 144 (214) from the Web server 142 (212) when the audio stream is requested from the audio file acquisition unit 192 (264) of the video playback terminal 144 (214).

30 [0464]

<Explanation of First Example of Process of Audio Decoding

Processing Unit>

Fig. 73 is a diagram illustrating a first example of the process of the audio decoding processing unit 194 when the Web server 142 (212) performs the process described above with reference to Figs. 71 and 72.

[0465]

In the example of Fig. 73, the Web server 142 (212) holds the audio stream of each track shown in Fig. 71. The tracks to be played back are the tracks of the channel audio of the front left/right channels, the channel audio of the rear left/right channels, the object audio of a first object, and Object Metadata. The same holds true for Fig. 75 to be described later.

[0466]

In this case, the audio file acquisition unit 192 (264) acquires the tracks of the channel audio of the front left/right channels, the channel audio of the rear left/right channels, the object audio of the first object, and Object Metadata.

[0467]

The audio decoding processing unit 194 first extracts the audio stream of the metadata of the object audio of the first object from the audio stream of the track of Object Metadata acquired by the audio file acquisition unit 192 (264).

[0468]

Next, as shown in Fig. 73, the audio decoding processing unit 194 synthesizes the audio stream of the track of the audio to be played back and the extracted audio stream of the metadata. Specifically, the audio decoding processing unit 194 generates the audio stream in which Config information

items included in all audio streams are collectively arranged and the data items of each frame are collectively arranged. Further, the audio decoding processing unit 194 decodes the generated audio stream.

5 [0469]

As described above, when the audio streams to be played back include an audio stream other than the audio stream of one channel audio track, audio streams of two or more tracks are to be played back. Accordingly, the audio streams are synthesized before decoding.

10

[0470]

On the other hand, when only the audio stream of the track of one channel audio is to be played back, there is no need to synthesize the audio stream. Accordingly, the audio decoding processing unit 194 directly decodes the audio stream acquired by the audio file acquisition unit 192 (264).

15

[0471]

Fig. 74 is a flowchart illustrating details of the first example of a decoding process of the audio decoding processing unit 194 when the Web server 142 (212) performs the process described above with reference to Figs. 71 and 72. This decoding process is at least one of the processes of step S229 shown in Fig. 48 and step S287 shown in Fig. 50 which are carried out when the tracks to be played back include a track other than one channel audio track.

20

25

[0472]

In step S461 of Fig. 74, the audio decoding processing unit 194 sets "0" to all element numbers representing the number of elements included in the generated audio stream. In step S462, the audio decoding processing unit 194 resets (clears) all element type information indicating the type of

30

elements included in the generated audio stream.

[0473]

In step S463, the audio decoding processing unit 194 sets, as a track to be processed, the track which has not been
5 determined to be the track to be processed among the tracks to be played back. In step S464, the audio decoding processing unit 194 acquires the number and type of elements included in the track to be processed from, for example, the audio stream of the track to be processed.

10 [0474]

In step S465, the audio decoding processing unit 194 adds the number of acquired elements to the total number of elements. In step S466, the audio decoding processing unit 194 adds the type of acquired elements to the all element type
15 information.

[0475]

In step S467, the audio decoding processing unit 194 determines whether all tracks to be played back are set as tracks to be processed. When it is determined in step S467
20 that not all the tracks to be played back are set as the track to be processed, the process returns to step S463 and the process of steps S463 to S467 is repeated until all tracks to be played back are set as the track to be processed.

[0476]

25 On the other hand, when it is determined in step S467 that all tracks to be played back are set as tracks to be processed, the process proceeds to step S468. In step S468, the audio decoding processing unit 194 arranges the total number of elements and all element type information at a
30 predetermined position on the generated audio stream.

[0477]

In step S469, the audio decoding processing unit 194 sets, as a track to be processed, the track which has not been determined to be the track to be processed among the tracks to be played back. In step S470, the audio decoding
5 processing unit 194 sets, as an element to be processed, the element which has not been determined to be the element to be processed among the elements included in the track to be processed.

[0478]

10 In step S471, the audio decoding processing unit 194 acquires, from the audio stream of tracks to be processed, Config information of the elements to be processed, and arranges the Config information on the generated audio stream. At this time, the Config information items of all elements
15 of all tracks to be played back are successively arranged.

[0479]

In step S472, the audio decoding processing unit 194 determines whether all elements included in the track to be processed are set as elements to be processed. When it is
20 determined in step S472 that not all the elements are set as the element to be processed, the process returns to step S470 and the process of steps S470 to S472 is repeated until all elements are set as the element to be processed.

[0480]

25 On the other hand, when it is determined in step S472 that all elements are set as elements to be processed, the process proceeds to step S473. In step S473, the audio decoding processing unit 194 determines whether all tracks to be played back are set as tracks to be processed. When
30 it is determined in step S473 that not all the tracks to be played back are set as the track to be processed, the process

returns to step S469 and the process of steps S469 to S473 is repeated until all tracks to be played back are set as the track to be processed.

[0481]

5 On the other hand, when it is determined in step S473 that all tracks to be played back are set as tracks to be processed, the process proceeds to step S474. In step S474, the audio decoding processing unit 194 determines a frame to be processed. In the process of step S474 of the first time,
10 the head frame is determined to be the frame to be processed. In the process of step S474 of the second and subsequent times, the frame next to the current frame to be processed is determined to be a new frame to be processed.

[0482]

15 In step S475, the audio decoding processing unit 194 sets, as a track to be processed, the track which has not been determined to be the track to be processed among the tracks to be played back. In step S476, the audio decoding processing unit 194 sets, as an element to be processed, the
20 element which has not been determined to be the element to be processed among the elements included in the track to be processed.

[0483]

 In step S477, the audio decoding processing unit 194
25 determines whether the element to be processed is an EXT element. When it is determined in step S477 that the element to be processed is not the EXT element, the process proceeds to step S478.

[0484]

30 In step S478, the audio decoding processing unit 194 acquires, from the audio stream of tracks to be processed,

the audio data of the frame to be processed of the element to be processed, and arranges the audio data on the generated audio stream. At this time, the data in the same frame of all elements of all tracks to be played back are successively
5 arranged. After the process of step S478, the process proceeds to step S481.

[0485]

On the other hand, when it is determined in step S477 that the element to be processed is the EXT element, the
10 process proceeds to step S479. In step S479, the audio decoding processing unit 194 acquires, from the audio stream of tracks to be processed, the metadata of all objects in the frame to be processed of the element to be processed.

[0486]

15 In step S480, the audio decoding processing unit 194 arranges the metadata of objects to be played back among the acquired metadata of all objects on the generated audio stream. At this time, the data items in the same frame of all elements of all tracks to be played back are successively arranged.
20 After the process of step S480, the process proceeds to step S481.

[0487]

In step S481, the audio decoding processing unit 194 determines whether all elements included in the track to be
25 processed are set as elements to be processed. When it is determined in step S481 that not all the elements are set as the element to be processed, the process returns to step S476 and the process of steps S476 to S481 is repeated until all elements are set as the element to be processed.

30 [0488]

On the other hand, when it is determined in step S481

that all elements are set as elements to be processed, the process proceeds to step S482. In step S482, the audio decoding processing unit 194 determines whether all tracks to be played back are set as tracks to be processed. When
5 it is determined in step S482 that not all the tracks to be played back are set as the track to be processed, the process returns to step S475 and the process of steps S475 to S482 is repeated until all tracks to be played back are set as the track to be processed.

10 [0489]

On the other hand, when it is determined in step S482 that all tracks to be played back are set as tracks to be processed, the process proceeds to step S483.

[0490]

15 In step S483, the audio decoding processing unit 194 determines whether all frames are set as frames to be processed. When it is determined in step S483 that not all the frames are set as the frame to be processed, the process returns to step S474 and the process of steps S474 to S483
20 is repeated until all frames are set as the frame to be processed.

[0491]

On the other hand, when it is determined in step S483 that all frames are set as frames to be processed, the process
25 proceeds to step S484. In step S484, the audio decoding processing unit 194 decodes the generated audio stream. Specifically, the audio decoding processing unit 194 decodes the audio stream in which the total number of elements, all element type information, Config information, audio data, and
30 metadata of objects to be played back are arranged. The audio decoding processing unit 194 supplies the audio synthesis

processing unit 195 with the audio data (Object audio, Channel audio, HOA audio) obtained as a result of decoding. Then, the process is terminated.

[0492]

5 <Explanation of Second Example of Process of Audio Decoding Processing Unit>

Fig. 75 is a diagram illustrating a second example of the process of the audio decoding processing unit 194 when the Web server 142 (212) performs the process described above
10 with reference to Figs. 71 and 72.

[0493]

As shown in Fig. 75, the second example of the process of the audio decoding processing unit 194 differs from the first example thereof in that audio streams of all tracks are
15 arranged on the generated audio stream and a stream or flag indicating a decoding result of zero (hereinafter referred to as a zero stream) is arranged as an audio stream of tracks which are not to be played back.

[0494]

20 Specifically, the audio file acquisition unit 192 (264) acquires Config information included in the audio streams of all tracks held in the Web server 142 (212), and data of each frame included in the audio streams of tracks to be played back.

25 [0495]

As shown in Fig. 75, the audio decoding processing unit 194 arranges the Config information items of all tracks collectively on the generated audio stream. Further, the audio decoding processing unit 194 arranges, on the generated
30 audio stream, the data of each frame of tracks to be played back and the zero stream as data of each frame of tracks which

are not to be played back.

[0496]

As described above, since the audio decoding processing unit 194 arranges, on the generated audio stream, the zero stream as the audio stream of tracks which are not to be played back, the audio stream of objects which are not to be played back is also present. Accordingly, it is possible to include the metadata of objects which are not to be played back in the generated audio stream. This eliminates the need for the audio decoding processing unit 194 to extract the audio stream of the metadata of objects to be played back from the audio stream of the track of Object Metadata.

[0497]

Note that the zero stream may be arranged as Config information of tracks which are not to be played back.

[0498]

Fig. 76 is a flowchart illustrating details of the second example of the decoding process of the audio decoding processing unit 194 when the Web server 142 (212) performs the process described above with reference to Figs. 71 and 72. This decoding process is at least one of the process of step S229 shown in Fig. 48 and the process of step S287 shown in Fig. 50 which are carried out when the tracks to be played back include a track other than one channel audio track.

[0499]

The process of steps S501 and S502 shown in Fig. 76 is similar to the process of steps S461 and S462 shown in Fig. 74, and thus the description thereof is omitted.

[0500]

In step S503, the audio decoding processing unit 194 sets, as a track to be processed, the track which has not been

determined to be the track to be processed among the tracks corresponding to the audio streams held in the Web server 142 (212).

[0501]

5 The process of steps S504 to S506 is similar to the process of steps S464 to S466, and thus the description thereof is omitted.

[0502]

10 In step S507, the audio decoding processing unit 194 determines whether all tracks corresponding to the audio streams held in the Web server 142 (212) are set as tracks to be processed. When it is determined in step S507 that not all the tracks are set as the track to be processed, the process returns to step S503 and the process of steps S503 to S507
15 is repeated until all tracks are set as the track to be processed.

[0503]

 On the other hand, when it is determined in step S507 that all tracks are set as tracks to be processed, the process
20 proceeds to step S508. In step S508, the audio decoding processing unit 194 arranges the total number of elements and all element type information at a predetermined position of the generated audio stream.

[0504]

25 In step S509, the audio decoding processing unit 194 sets, as a track to be processed, the track which has not been determined to be the track to be processed among the tracks corresponding to the audio streams held in the Web server 142 (212). In step S510, the audio decoding processing unit 194
30 sets, as an element to be processed, the element which has not been determined to be the element to be processed among

the elements included in the track to be processed.

[0505]

In step S511, the audio decoding processing unit 194 acquires Config information of an element to be processed from the audio stream of the track to be processed, and generates the Config information on the generated audio stream. At this time, the Config information items of all elements of all tracks corresponding to the audio streams held in the Web server 142 (212) are successively arranged.

10 [0506]

In step S512, the audio decoding processing unit 194 determines whether all elements included in the tracks to be processes are set as elements to be processed. When it is determined in step S512 that not all the elements are set as the element to be processed, the process returns to step S510 and the process of steps S510 to S512 is repeated until all elements are set as the element to be processed.

[0507]

On the other hand, when it is determined in step S512 that all elements are set as elements to be processed, the process proceeds to step S513. In step S513, the audio decoding processing unit 194 determines whether all tracks corresponding to the audio streams held in the Web server 142 (212) are set as tracks to be processed. When it is determined in step S513 that not all the tracks are set as the track to be processed, the process returns to step S509 and the process of steps S509 to S513 is repeated until all tracks are set as the track to be processed.

[0508]

30 On the other hand, when it is determined in step S513 that all tracks are set as tracks to be processed, the process

proceeds to step S514. In step S514, the audio decoding processing unit 194 determines a frame to be processed. In the process of step S514 of the first time, the head frame is determined to be the frame to be processed. In the process
5 of step S514 of the second and subsequent times, the frame next to the current frame to be processed is determined to be a new frame to be processed.

[0509]

10 In step S515, the audio decoding processing unit 194 sets, as a track to be processed, the track which has not been determined to be the track to be processed among the tracks corresponding to the audio streams held in the Web server 142 (212).

[0510]

15 In step S516, the audio decoding processing unit 194 determines whether the track to be processed is the track to be played back. When it is determined in step S516 that the track to be processed is the track to be played back, the process proceeds to step S517.

20 [0511]

In step S517, the audio decoding processing unit 194 sets, as an element to be processed, the element which has not been determined to be the element to be processed among the elements included in the track to be processed.

25 [0512]

In step S518, the audio decoding processing unit 194 acquires, from the audio stream of the track to be processed, the audio data of the frame to be processed of the element to be processed, and arranges the audio stream on the
30 generated audio stream. At this time, the data items in the same frame of all elements of all tracks corresponding to the

audio streams held in the Web server 142 (212) are successively arranged.

[0513]

In step S519, the audio decoding processing unit 194
5 determines whether all elements included in the track to be processed are set as the element to be processed. When it is determined in step S519 that not all the elements are set as the element to be processed, the process returns to step S517 and the process of steps S517 to S519 is repeated until
10 all elements are set as the element to be processed.

[0514]

On the other hand, when it is determined in step S519 that all elements are set as the element to be processed, the process proceeds to step S523.

15 [0515]

Further, when it is determined in step S516 that the track to be processed is not the track to be played back, the process proceeds to step S520. In step S520, the audio decoding processing unit 194 sets, as an element to be
20 processed, the element which has not been determined to be the element to be processed among the elements included in the track to be processed.

[0516]

In step S521, the audio decoding processing unit 194
25 arranges the zero stream as the data of the frame to be processed of the element to be processed on the generated audio stream. At this time, the data items in the same frame of all elements of all tracks corresponding to the audio streams held in the Web server 142 (212) are successively
30 arranged.

[0517]

In step S522, the audio decoding processing unit 194 determines whether all elements included in the track to be processed are set as the element to be processed. When it is determined in step S522 that not all the elements are set
5 as the element to be processed, the process returns to step S520 and the process of steps S520 to S522 is repeated until all elements are set as the element to be processed.

[0518]

On the other hand, when it is determined in step S522
10 that all elements are set as the element to be processed, the process proceeds to step S523.

[0519]

In step S523, the audio decoding processing unit 194 determines whether all tracks corresponding to the audio
15 streams held in the Web server 142 (212) are set as the track to be processed. When it is determined in step S522 that not all the tracks are set as the track to be processed, the process returns to step S515 and the process of steps S515 to S523 is repeated until all tracks to be played back are set as the
20 track to be processed.

[0520]

On the other hand, when it is determined in step S523 that all tracks are set as the track to be processed, the process proceeds to step S524.

25 [0521]

In step S524, the audio decoding processing unit 194 determines whether all frames are set as the frame to be processed. When it is determined in step S524 that not all the frames are set as the frame to be processed, the process
30 returns to step S514 and the process of steps S514 to S524 is repeated until all frames are set as the frame to be

processed.

[0522]

On the other hand, when it is determined in step S524 that all frames are set as the frame to be processed, the process proceeds to step S525. In step S525, the audio decoding processing unit 194 decodes the generated audio stream. Specifically, the audio decoding processing unit 194 decodes the audio stream in which the total number of elements, all element type information, and Config information and data of all tracks corresponding to the audio streams held in the Web server 142 (212) are arranged. The audio decoding processing unit 194 supplies the audio synthesis processing unit 195 with the audio data (Object audio, Channel audio, HOA audio) obtained as a result of decoding. Then, the process is terminated.

[0523]

<Explanation of Second Example of Web Server Process>

Fig. 77 is a diagram illustrating a second example of the process of the Web server 142 (212).

20 [0524]

The second example of the process of the Web server 142 (212) shown in Fig. 77 is the same as the first example shown in Fig. 71, except that Object Metadata of each object is arranged in the audio file as data of different tracks.

25 [0525]

Accordingly, as shown in Fig. 77, the Web server 142 (212) divides, for each track, the audio stream of the audio file uploaded from the file generation device 141 (211), and generates the audio stream of nine tracks.

30 [0526]

In this case, the track division process of the Web

server 142 (212) is similar to the track division process shown in Fig. 72, and thus the description thereof is omitted. [0527]

<Explanation of Third Example of Audio Decoding Processing Unit>

Fig. 78 is a diagram illustrating the process of the audio decoding processing unit 194 when the Web server 142 (212) performs the process described above with reference to Fig. 77.

10 [0528]

In the example of Fig. 78, the Web server 142 (212) holds the audio stream of each track shown in Fig. 77. The tracks to be played back are the tracks of the channel audio of the front left/right channels, the channel audio of the rear

15 left/right channels, the object audio of the first object, and the Object Metadata of the first object.

[0529]

In this case, the audio file acquisition unit 192 (264) acquires the audio streams of the tracks of the channel audio of the front left/right channels, the channel audio of the

20 rear left/right channels, the object audio of the first object, and the Object Metadata of the first object. The audio decoding processing unit 194 synthesizes the acquired audio streams of the tracks to be played back, and decodes the

25 generated audio stream.

[0530]

As described above, when the Object Metadata is arranged as data of different tracks for each object, there is no need for the audio decoding processing unit 194 to

30 extract the audio stream of the Object Metadata of objects to be played back. Accordingly, audio decoding processing

unit 194 can easily generate the audio stream to be decoded.
[0531]

Fig. 79 is a flowchart illustrating details of the decoding process of the audio decoding processing unit 194 when the Web server 142 (212) performs the process described above with reference to Fig. 77. This decoding process is one of the process of step S229 shown in Fig. 48 and the process of step S287 shown in Fig. 50 which are carried out when the tracks to be played back include a track other than one channel audio track.

[0532]

The decoding process shown in Fig. 79 is similar to the decoding process shown in Fig. 74, except that processes in steps S477, S479, and S480 are not carried out and not only audio data but also metadata are arranged in the process of step S478. Specifically, the process of steps S541 to S556 shown in Fig. 79 is similar to steps S461 to S476 shown in Fig. 74. In the process of step S557 shown in Fig. 79, the data of the frame to be processed of the element to be processed is arranged, like in the process of step S478. Further, the process of steps S558 to S561 is similar to the process of steps S481 to S484 shown in Fig. 74.

[0533]

Note that in the above description, the video playback terminal 144 (214) generates the audio stream to be decoded, but instead the Web server 142 (212) may generate a combination of audio streams which are assumed as a combination of tracks to be played back. In this case, the video playback terminal 144 (214) can play back the audio of the tracks to be played back only by acquiring the audio streams with a combination of tracks to be played back from

the Web server 142 (212) and decoding the audio streams.
[0534]

Further, the audio decoding processing unit 194 may
decode, for each track, the audio streams of the tracks to
5 be played back that are acquired from the Web server 142 (212).
In this case, the audio decoding processing unit 194 needs
to synthesize the audio data and metadata obtained as a result
of decoding.

[0535]

10 <Second Example of Syntax of Base Sample>

(Second Example of Syntax of Config information Arranged in
Base Sample)

Fig. 80 is a diagram showing a second example of syntax
of Config information arranged in a base sample.

15 [0536]

In the example of Fig. 80, the number of elements
(numElements) arranged in the base sample is described as
Config information. Further, as the type of each element
(usacElementType) arranged in the base sample, "ID_USAC_EXT"
20 representing the Ext element is described and Config
information for Ext element of each element
(mpeg3daExtElementConfig) is also described.

[0537]

Fig. 81 is a diagram showing an exemplary syntax of
25 Config information (mpeg3daExtElementConfig) for Ext
element shown in Fig. 80.

[0538]

As shown in Fig. 81, "ID_EXT_ELE_EXTRACTOR"
representing Extractor as the type of the Ext element is
30 described as Config information for Ext element
(mpeg3daExtElementConfig) shown in Fig. 80. Further,

Config information for Extractor (ExtractorConfig) is described.

[0539]

Fig. 82 is a diagram showing an exemplary syntax of
5 Config information for Extractor (ExtractorConfig) shown in Fig. 81.

[0540]

As shown in Fig. 82, as Config information for Extractor (ExtractorConfig) shown in Fig. 81, the type of the element
10 (usac Element Type Extractor) to be referred to by the Extractor is described. Further, when the type of the element (usac Element Type Extractor) is "ID_USAC_EXT" which represents the Ext element, the type of the Ext element (usacExtElementTypeExtractor) is described. Furthermore,
15 the size (configLength) and position (configOffset) of the Config information of the element (sub-sample) to be referred to are described.

[0541]

(Second Example of Syntax of Data of Frame Unit Arranged in
20 Base Sample)

Fig. 83 is a diagram showing a second example of syntax of data in units of frames arranged in the base sample.

[0542]

As shown in Fig. 83, as the data in units of frames arranged in the base sample, "ID_EXT_ELE_EXTRACTOR" which represents Extractor as the type of the Ext element which is the data element is described. Extractor data (Extractor Metadata) is also described.

[0543]

Fig. 84 is a diagram showing an exemplary syntax of
30 Extractor data (Extractor Metadata) shown in Fig. 83.

[0544]

As shown in Fig. 84, the size (elementLength) and position (elementOffset) of the data of the element to be referred to by the Extractor are described as Extractor data (Extractor Metadata) shown in Fig. 83.

[0545]

<Third Example of Syntax of Base Sample>

(Third Example of Syntax of Config Information Arranged in Base Sample)

Fig. 85 is a diagram showing a third example of syntax of Config information arranged in the base sample.

[0546]

In the example of Fig. 85, the number of elements (numElements) arranged in the base sample is described as Config information. Further, "1" indicating Extractor is described as an Extractor flag (flag Extractor) indicating whether the sample in which Config information is arranged is Extractor. Furthermore, "1" is described as elementLengthPresent.

[0547]

Further, the type of the element to be referred to by the element is described as the type of each element (usacElementType) arranged in the base sample. When the type of the element (usacElementType) is "ID_USAC_EXT" which represents the Ext element, the type of the Ext element (usacExtElementType) is described. Furthermore, the size (configLength) and position (configOffset) of Config information of the element to be referred to are described.

[0548]

(Third Example of Syntax of Data in Units of Frames Arranged in Base Sample)

Fig. 86 is a diagram showing a third example of syntax of data in units of frames arranged in the base sample.

[0549]

As shown in Fig. 86, as the data in units of frames arranged in the base sample, the size (elementLength) and position (elementOffset) of the data of the element to be referred to by the data are described.

[0550]

<Seventh Embodiment>

10 (Configuration Example of Audio Stream)

Fig. 87 is a diagram showing a configuration example of the audio stream stored in the audio file in a seventh embodiment of an information processing system to which the present disclosure is applied.

15 [0551]

As shown in Fig. 87, in the seventh embodiment, the audio file stores coded data in units of samples of 3D audio for each group type (in this case, however, the object audio is stored for each object) and an audio stream (3D audio stream) arranged as a sub-sample.

[0552]

Further, the audio file stores a clue stream (3D audio hint stream) in which the extractor including the size, position, and group type of the coded data in units of samples of 3D audio for each group type is set as a sub-sample. The configuration of the extractor is similar to the configuration described above, and the group type is described as the type of the extractor.

[0553]

30 (Outline of Tracks)

Fig. 88 is a diagram illustrating the outline of tracks

in the seventh embodiment.

[0554]

As shown in Fig. 88, in the seventh embodiment, different tracks are allocated to an audio stream and a clue stream, respectively. The track ID "2" of the track of the
5 corresponding clue stream is described as Track Reference of the track of the audio stream. Further, the track ID "1" of the track of the corresponding audio stream is described as Track Reference of the track of the clue stream.

10 [0555]

The syntax of the sample entry of the track of the audio stream is the syntax shown in Fig. 34, and the syntax of the sample entry of the track of the clue stream includes the syntax shown in Figs. 35 to 38.

15 [0556]

(Explanation of Process of File Generation Device)

Fig. 89 is a flowchart illustrating a file generation process of the file generation device in the seventh embodiment.

20 [0557]

Note that the file generation device according to the seventh embodiment is the same as the file generation device 141 shown in Fig. 45, except for the processes of the audio coding processing unit 171 and the audio file generation unit
25 172. Accordingly, the file generation device, the audio coding processing unit, and the audio file generation unit according to the seventh embodiment are hereinafter referred to as a file generation device 301, an audio coding processing unit 341, and an audio file generation unit 342, respectively.

30 [0558]

The process of steps S601 to S605 shown in Fig. 89 is

similar to the process of steps S191 to S195 shown in Fig. 46, and thus the description thereof is omitted.

[0559]

In step S606, the audio coding processing unit 341
5 encodes, for each group type, the 3D audio of the video content input from the outside, and generates the audio stream shown in Fig. 87. The audio coding processing unit 341 supplies the audio file generation unit 342 with the generated audio stream.

10 [0560]

In step S607, the audio file generation unit 342 acquires sub-sample information from the audio stream supplied from the audio coding processing unit 341. The sub-sample information indicates the size, position, and
15 group type of the coded data in units of samples of the 3D audio of each group type.

[0561]

In step S608, the audio file generation unit 342 generates the clue stream shown in Fig. 87 on the basis of
20 the sub-sample information. In step S609, the audio file generation unit 342 multiplexes the audio stream and the clue stream as different tracks, and generates an audio file. At this time, the audio file generation unit 342 stores the image frame size information input from the outside in the sample
25 entry. The audio file generation unit 342 supplies the MPD generation unit 173 with the generated audio file.

[0562]

The process of steps S610 and S611 is similar to the process of steps S199 and S200 shown in Fig. 46, and thus the
30 description thereof is omitted.

[0563]

(Explanation of Process of Video Playback Terminal)

Fig. 90 is a flowchart illustrating an audio playback process of the stream playback unit of the video playback terminal in the seventh embodiment.

5 [0564]

Note that the streaming playback unit according to the seventh embodiment is the same as the streaming playback unit 190 shown in Fig. 47, except that the processes of the MPD processing unit 191, the audio file acquisition unit 192, and the audio decoding processing unit 194 are different and the audio selection unit 193 is not provided. Accordingly, the streaming playback unit, the MPD processing unit, the audio file acquisition unit, and the audio decoding processing unit according to the seventh embodiment are hereinafter referred to as a streaming playback unit 360, an MPD processing unit 381, an audio file acquisition unit 382, and an audio decoding processing unit 383, respectively.

[0565]

In step S621 shown in Fig. 90, the MPD processing unit 381 of the streaming playback unit 360 analyzes the MPD file supplied from the MPD acquisition unit 91, acquires information such as the URL of the audio file of the segment to be played back, and supplies the audio file acquisition unit 382 with the acquired information.

25 [0566]

In step S622, the audio file acquisition unit 382 requests the Web server to transmit Initial Segment of the segment to be played back on the basis of the information such as the URL supplied from the MPD processing unit 381, and acquires the Initial Segment.

[0567]

In step S623, the audio file acquisition unit 382 acquires the track ID of the track of the audio stream as the reference track from the sample entry of the track of the clue stream (hereinafter referred to as a clue track) of the moov box in the Initial Segment.
[0568]

In step S624, the audio file acquisition unit 382 requests the Web server to transmit the sidx box and the ssix box from the head of the media segment of the segment to be played back on the basis of the information such as the URL supplied from the MPD processing unit 381, and acquires the sidx box and the ssix box.
[0569]

In step S625, the audio file acquisition unit 382 acquires the position information of the clue track from the sidx box and the ssix box which are acquired in step S624.
[0570]

In step S626, the audio file acquisition unit 382 requests the Web server to transmit the clue stream on the basis of the position information of the clue track acquired in step S625, and acquires the clue stream. Further, the audio file acquisition unit 382 acquires, from the clue stream, the extractor of the group type of the 3D audio to be played back. Note that when the 3D audio to be played back is the object audio, the object to be played back is selected on the basis of the image frame size information and object position information.
[0571]

In step S627, the audio file acquisition unit 382 acquires the position information of the reference track from the sidx box and the ssix box which are acquired in step S624.

In step S628, the audio file acquisition unit 382 determines the position information of the audio stream of the group type of the 3D audio to be played back on the basis of the position information of the reference track acquired in step S627 and
5 the sub-sample information included in the acquired extractor.

[0572]

In step S629, the audio file acquisition unit 382 requests the Web server to transmit the audio stream of the
10 group type of the 3D audio to be played back on the basis of the position information determined in step S627, and acquires the audio stream. The audio file acquisition unit 382 supplies the audio decoding processing unit 383 with the acquired audio stream.

15 [0573]

In step S630, the audio decoding processing unit 383 decodes the audio stream supplied from the audio file acquisition unit 382, and supplies the audio synthesis processing unit 195 with the audio data obtained as a result
20 of decoding.

[0574]

In step S631, the audio synthesis processing unit 195 outputs the audio data. Then, the process is terminated.

[0575]

25 Note that in the seventh embodiment, the track of the audio stream and the clue track are stored in the same audio file, but may be stored in different files.

[0576]

<Eighth Embodiment>

30 (Outline of Tracks)

Fig. 91 is a diagram illustrating the outline of tracks

in an eighth embodiment of an information processing system to which the present disclosure is applied.

[0577]

The audio file of the eighth embodiment is different
5 from the audio file of the seventh embodiment in that the stored clue stream is a stream for each group type. Specifically, the clue stream of the eighth embodiment is generated for each group type, and the extractor including the size, position, and group type of the coded data in units
10 of samples of the 3D audio of each group type is arranged as a sample in each clue stream. Note that when the 3D audio includes object audios of a plurality of objects, the extractor is arranged as a sub-sample for each object.

[0578]

15 Further, as shown in Fig. 91, in the eighth embodiment, different tracks are allocated to the audio stream and each clue stream. The track of the audio stream is the same as the track of the audio stream shown in Fig. 88, and thus the description thereof is omitted.

20 [0579]

As Track Reference of the clue track of the group types of "Channels", "Objects", "HOA", and "metadata", the track ID "1" of the track of the corresponding audio stream is described.

25 [0580]

The syntax of the sample entry of the clue track of each of the group types of "Channels", "Objects", "HOA", and "metadata" is the same as the syntax shown in Figs. 35 to 38, except for the information indicating the type of the sample
30 entry. The information indicating the type of the sample entry of the clue track of each of the group types of "Channels",

"Objects", "HOA", and "metadata" is similar to the information shown in Figs. 35 to 38, except that the number "1" of the information is replaced by "2". The number "2" represents the sample entry of the clue track.

5 [0581]

(Configuration Example of Audio File)

Fig. 92 is a diagram showing a configuration example of the audio file.

[0582]

10 As shown in Fig. 92, the audio file stores all tracks shown in Fig. 91. Specifically, the audio file stores the audio stream and the clue stream of each group type.

[0583]

15 The file generation process of the file generation device according to the eighth embodiment is similar to the file generation process shown in Fig. 89, except that the clue stream is generated for each group type, instead of the clue stream shown in Fig. 87.

[0584]

20 Further, the audio playback process of the streaming playback unit of the video playback terminal according to the eighth embodiment is similar to the audio playback process shown in Fig. 90, except that the track ID of the clue track of the group type to be played back, as well as the track ID
25 of the reference track is acquired in step S623; the position information of the clue track of the group type to be played back in step S625 is acquired; and the clue stream of the group type to be played back is acquired in step S626.

[0585]

30 Note that in the eighth embodiment, the track of the audio stream and the clue track are stored in the same audio

file, but may be stored in different files.

[0586]

For example, as shown in Fig. 93, the track of the audio stream may be stored in one audio file (3D audio stream MP4 File), and the clue track may be stored in one audio file (3D audio hint stream MP4 File). Further, as shown in Fig. 94, the clue track may be divided into a plurality of audio files to be stored. In the example of Fig. 94, the clue tracks are stored in different audio files.

10 [0587]

Further, in the eighth embodiment, the clue stream is generated for each group type even when the group type indicates objects. However, when the group type indicates objects, the clue stream may be generated for each object. In this case, different tracks are allocated to the clue streams of each object.

[0588]

As described above, in the audio file of the seventh and eighth embodiments, all the audio streams of 3D audio are stored in one track. Accordingly, the video playback terminal can play back all the audio streams of 3D audios by acquiring the track.

[0589]

Further, the clue stream is stored in the audio file of the seventh and eighth embodiments. Accordingly, the video playback terminal acquires only the audio stream of a desired group type among all the audio streams of 3D audio without referring to the moof box in which a table that associates a sub-sample with the size or position of the sub-sample is described, thereby making it possible to play back the audio stream.

30

[0590]

Further, in the audio file of the seventh and eighth embodiments, the video playback terminal can be caused to acquire the audio stream for each group type, only by storing
5 all the audio streams of 3D audio and the clue stream. Accordingly, there is no need to prepare the audio stream of 3D audio for each group type separately from all the generated audio streams of 3D audio for the purpose of broadcasting or local storage so as to enable acquisition of the audio stream
10 for each group type.

[0591]

Note that in the seventh and eighth embodiments, the extractor is generated for each group type, but may be generated in units of audio elements, groups, switch Groups,
15 or group Presets.

[0592]

When the extractor is generated in units of groups, the sample entry of each clue track of the eighth embodiment includes information about the corresponding group. The
20 information about the group is composed of, for example, information indicating the ID of the group and the content of data of the element classified as the group. When the group forms the switch Group, the sample entry of the clue track of the group also includes information about the switch Group.
25 The information about the switch Group is composed of, for example, the ID of the switch Group and the ID of the group that forms the switch Group. The sample entry of the clue track of the seventh embodiment includes the information included in the sample entries of all clue tracks of the eighth
30 embodiment.

[0593]

Further, the segment structures in the seventh and eighth embodiments are the same as the segment structures shown in Figs. 39 and 40.

[0594]

5 <Ninth Embodiment>

(Explanation of Computer to Which The Present Disclosure Is Applied)

A series of processes of the Web server described above can also be executed by hardware or software. When the series
10 of processes is executed by software, a program constituting the software is installed in a computer. Examples of the computer include a computer incorporated in dedicated hardware and a general-purpose personal computer capable of executing various functions by installing various programs
15 therein.

[0595]

Fig. 95 is a block diagram showing a configuration example of hardware of a computer that executes a series of processes for the Web server by using a program.

20 [0596]

In the computer, a central processing unit (CPU) 601, a read only memory (ROM) 602, and a random access memory (RAM) 603 are interconnected via a bus 604.

[0597]

25 The bus 604 is also connected to an input/output interface 605. The input/output interface 605 is connected to each of an input unit 606, an output unit 607, a storage unit 608, a communication unit 609, and a drive 610.

[0598]

30 The input unit 606 is formed with a keyboard, a mouse, a microphone, and the like. The output unit 607 is formed

with a display, a speaker, and the like. The storage unit 608 is formed with hardware, a non-volatile memory, and the like. The communication unit 609 is formed with a network interface and the like. The drive 610 drives a removable
5 medium 611 such as a magnetic disk, an optical disk, a magneto-optical disk, or a semiconductor memory.

[0599]

In the computer configured as described above, a CPU 601 loads, for example, the program stored in the storage unit
10 608 in the RAM 603 via the input/output interface 605 and the bus 604, and executes the program, thereby performing the series of processes described above.

[0600]

The program executed by the computer (CPU 601) can be
15 provided by being recorded in the removable medium 611 serving as, for example, a package medium or the like. In addition, the program can be provided via a wired or wireless transmission medium such as a local area network, the Internet, or digital satellite broadcasting.

20 [0601]

The program can be installed in the storage unit 608 via the input/output interface 605 by loading the removable medium 611 in the drive 610. Further, the program can be received by the communication unit 609 and installed in the
25 storage unit 608 via the wired or wireless transmission medium. In addition, the program can be installed in advance in the ROM 602 or the storage unit 608.

[0602]

Note that the program executed by the computer may be
30 a program which performs the processes in a time series manner in the order described in the present description, or may be

a program which performs the processes in parallel or at necessary timings when they are invoked, for example.

[0603]

The video playback terminal described above may have
5 a hardware configuration that is similar to that of the computer shown in Fig. 95. In this case, for example, the CPU 601 can execute the control software 161 (221), the video playback software 162 (222), and the access software 163 (223). The process of the video playback terminal 144 (214) may be
10 executed by hardware.

[0604]

In the present description, a system has the meaning of a set of a plurality of components (such as an apparatus or a module (part)), and does not take into account whether
15 or not all the components are in the same casing. Therefore, the system may be either a plurality of apparatuses, which are stored in separate casings and connected through a network, or a plurality of modules within a single casing.

[0605]

20 Note that embodiments of the present disclosure are not limited to the above-described embodiments, and can be modified in various ways without departing from the gist of the present disclosure.

[0606]

25 For example, the file generation device 141 (211) may generate the video stream by multiplexing the coded data of all tiles to generate one image file, instead of generating the image file in units of tiles.

[0607]

30 The present disclosure can be applied not only to MPEG-H 3D audio, but also to general audio codecs capable of creating

a stream for each object.

[0608]

Further, the present disclosure can also be applied to an information processing system that performs broadcasting
5 and local storage playback, as well as streaming playback.

[0609]

Furthermore, the present disclosure may have the following configurations.

[0610]

10 (1)

An information processing apparatus including an acquisition unit that acquires audio data of a predetermined track in a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types
15 and the tracks are arranged.

(2)

The information processing apparatus according to the above item (1), in which the types are configured to be an element of the audio data, a type of the element, or a group
20 into which the element is classified.

(3)

The information processing apparatus according to the above item (1) or (2), further including a decoding unit that decodes the audio data of the predetermined track acquired
25 by the acquisition unit.

(4)

The information processing apparatus according to the above item (3), in which when there are a plurality of predetermined tracks, the decoding unit synthesizes the audio
30 data of the predetermined tracks acquired by the acquisition unit, and decodes the synthesized audio data.

(5)

The information processing apparatus according to the above item (4), in which

the file is configured in such a manner that audio data
5 in units of a plurality of objects is divided into the tracks
different for each object and the tracks are arranged, and
metadata items of all the audio data in units of objects are
collectively arranged in a track different from the track,

the acquisition unit is configured to acquire the audio
10 data of the track of the object to be played back as the audio
data of the predetermined track, and acquires the metadata,
and

the decoding unit is configured to extract the metadata
of the object to be played back from the metadata acquired
15 by the acquisition unit, and synthesize the metadata with the
audio data acquired by the acquisition unit.

(6)

The information processing apparatus according to the above item (4), in which

20 the file is configured in such a manner that audio data
in units of a plurality of objects is divided into the tracks
different for each object and the tracks are arranged, and
metadata items of all the audio data in units of objects are
collectively arranged in a track different from the track,

25 and

the acquisition unit is configured to acquire the audio
data of the track of the object to be played back as the audio
data of the predetermined track, and acquires the metadata,
and

30 the decoding unit is configured to synthesize zero data
with the audio data and the metadata acquired by the

acquisition unit, the zero data indicating a decoding result of zero as the audio data of the track which is not to be played back.

(7)

5 The information processing apparatus according to the above item (4), in which

 the file is configured in such a manner that audio data in units of a plurality of objects is divided into the tracks different for each object and the tracks are arranged, and
10 metadata items of the audio data in units of objects are arranged in different tracks for each object,

 the acquisition unit is configured to acquire the audio data of the track of the object to be played back as the audio data of the predetermined track, and acquires the metadata of
15 the object to be played back, and

 the decoding unit is configured to synthesize the audio data and the metadata acquired by the acquisition unit.

(8)

 The information processing apparatus according to any
20 one of the above items (1) to (7), in which the audio data items of the plurality of tracks are configured to be arranged in one file.

(9)

 The information processing apparatus according to any
25 one of the above items (1) to (7), in which the audio data items of the plurality of tracks are configured to be arranged in the different files for each track.

(10)

 The information processing apparatus according to any
30 one of the above items (1) to (9), in which the file is configured in such a manner that information about the

plurality of types of the audio data is arranged as a track different from the plurality of tracks.

(11)

The information processing apparatus according to the
5 above item (10), in which information about the plurality of types of the audio data is configured to include image frame size information indicating an image frame size of image data corresponding to the audio data.

(12)

10 The information processing apparatus according to any one of the above items (1) to (9), in which the file is configured in such a manner that, as the audio data of a track different from the plurality of tracks, information indicating a position of the audio data of another track
15 corresponding to the audio data is arranged.

(13)

The information processing apparatus according to any one of the above items (1) to (9), in which the file is configured in such a manner that, as the data of a track
20 different from the plurality of tracks, information indicating a position of the audio data of another track corresponding to the data and metadata of the audio data of the other track are arranged.

(14)

25 The information processing apparatus according to the above item (13), in which the metadata of the audio data is configured to include information indicating a position at which the audio data is acquired.

(15)

30 The information processing apparatus according to any one of the above items (1) to (14), in which the file is

configured to include information indicating a reference relationship between the track and the other track.

(16)

The information processing apparatus according to any
5 one of the above items (1) to (15), in which the file is configured to include codec information of the audio data of each track.

(17)

The information processing apparatus according to any
10 one of the above items (1) to (16), in which the predetermined type of audio data is information indicating a position at which another type of audio data is acquired.

(18)

An information processing method including an
15 acquisition step of acquiring, by an information processing apparatus, audio data of a predetermined track in a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types and the tracks are arranged.

20 (19)

An information processing apparatus including a generation unit that generates a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types and the tracks are arranged.

25 (20)

An information processing method including a generation step of generating, by an information processing apparatus, a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types
30 and the tracks are arranged.

REFERENCE SIGNS LIST

[0611]

	141	File generation device
	144	Moving image playback terminal
5	172	Audio file generation unit
	192	Audio file acquisition unit
	193	Audio selection unit
	211	File generation device
	214	Moving image playback terminal
10	241	Audio file generation unit
	264	Audio file acquisition unit

CLAIMS

1. An information processing apparatus comprising an acquisition unit that acquires audio data of a predetermined track in a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types and the tracks are arranged.
5
2. The information processing apparatus according to claim 1, wherein the types are configured to be an element of the audio data, a type of the element, or a group into which the element is classified.
10
3. The information processing apparatus according to claim 1, further comprising a decoding unit that decodes the audio data of the predetermined track acquired by the acquisition unit.
15
4. The information processing apparatus according to claim 3, wherein when there are a plurality of the predetermined tracks, the decoding unit synthesizes the audio data of the predetermined tracks acquired by the acquisition unit, and decodes the synthesized audio data.
20
5. The information processing apparatus according to claim 4, wherein
25
the file is configured in such a manner that audio data in units of a plurality of objects is divided into the tracks different for each object and the tracks are arranged, and
30
metadata items of all the audio data in units of objects are collectively arranged in a track different from the track,

the acquisition unit is configured to acquire the audio data of the track of an object to be played back as the audio data of the predetermined track, and acquire the metadata, and

5 the decoding unit is configured to extract the metadata of the object to be played back from the metadata acquired by the acquisition unit, and synthesize the metadata with the audio data acquired by the acquisition unit.

10 6. The information processing apparatus according to claim 4, wherein

the file is configured in such a manner that audio data in units of a plurality of objects is divided into the tracks different for each object and the tracks are arranged, and
15 metadata of all the audio data in units of objects are collectively arranged in a track different from the track,

the acquisition unit is configured to acquire the audio data of the track of the object to be played back as the audio data of the predetermined track, and acquire the metadata,
20 and

the decoding unit is configured to synthesize the audio data and the metadata acquired by the acquisition unit with zero data indicating a decoding result of zero as the audio data of the track which is not to be played back.

25

7. The information processing apparatus according to claim 4, wherein

the file is configured in such a manner that audio data in units of a plurality of objects is divided into the tracks different for each object and the tracks are arranged,
30 metadata of the audio data in units of objects is arranged

in tracks different for each object,

the acquisition unit is configured to acquire the audio data in the track of an object to be played back as the audio data of the predetermined track, and acquire metadata of the object to be played back, and

the decoding unit synthesizes the audio data and the metadata acquired by the acquisition unit.

8. The information processing apparatus according to claim 1, wherein the audio data of the plurality of tracks is arranged in one file.

9. The information processing apparatus according to claim 1, wherein the audio data of the plurality of tracks is arranged in the files different for each track.

10. The information processing apparatus according to claim 1, wherein the file is configured in such a manner that information about the plurality of types of the audio data is arranged as tracks different from the plurality of tracks.

11. The information processing apparatus according to claim 10, wherein information about the plurality of types of the audio data includes image frame size information indicating an image frame size of image data corresponding to the audio data.

12. The information processing apparatus according to claim 1, wherein the file is configured in such a manner that information indicating a position of the audio data in another track corresponding to the audio data is arranged as the audio

data of a track different from the plurality of tracks.

13. The information processing apparatus according to claim 1, wherein the file is configured in such a manner that
5 information indicating a position of the audio data of another track corresponding to the data and metadata of the audio data of the other track are arranged as the data of a track different from the plurality of tracks.

10 14. The information processing apparatus according to claim 13, wherein the metadata of the audio data is configured to include information indicating a position at which the audio data is acquired.

15 15. The information processing apparatus according to claim 1, wherein the file is configured to include information indicating a reference relationship between the track and the other track.

20 16. The information processing apparatus according to claim 1, wherein the file is configured to include codec information of the audio data of each track.

25 17. The information processing apparatus according to claim 1, wherein the predetermined type of audio data is configured as information indicating a position at which another type of audio data is acquired.

30 18. An information processing method comprising an acquisition step of acquiring, by an information processing apparatus, acquiring audio data of a predetermined track in

a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types and the tracks are arranged.

5 19. An information processing apparatus comprising a generation unit that generates a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types and the tracks are arranged.

10 20. An information processing method comprising a generation step of generating, by an information processing apparatus, a file in which a plurality of types of audio data are divided into a plurality of tracks depending on the types and the tracks are arranged.

15