

## Apparatus and Method for Audio Rendering Employing a Geometric Distance Definition

### Description

5

The present invention relates to audio signal processing, in particular, to an apparatus and a method for audio rendering, and, more particularly, to an apparatus and a method for audio rendering employing a geometric distance definition.

10

With increasing multimedia content consumption in daily life, the demand for sophisticated multimedia solutions steadily increases. In this context, positioning of audio objects plays an important role. An optimal positioning of audio objects for an existing loudspeaker setup would be desirable.

15

In the state of the art, audio objects are known. Audio objects may, e.g., be considered as sound tracks with associated metadata. The metadata may, e.g., describe the characteristics of the raw audio data, e.g., the desired playback position or the volume level. An advantage of object-based audio is that a predefined movement can be reproduced by a special rendering process on the playback side in the best way possible for all reproduction loudspeaker layouts.

20

25

Geometric metadata can be used to define where an audio object should be rendered, e.g., angles in azimuth or elevation or absolute positions relative to a reference point, e.g., the listener. The metadata is stored or transmitted along with the object audio signals.

30

In the context of MPEG-H, at the 105th MPEG meeting the audio group reviewed the requirements and timelines of different application standards (MPEG = Moving Picture Experts Group). According to that review, it would be essential to meet certain points in time and specific requirements for a next generation broadcast system. According to that, a system should be able to accept audio objects at the encoder input. Moreover, the system should support signaling, delivery and rendering of audio objects and should enable user control of objects, e.g., for dialog enhancement, alternative language tracks and audio description language.

35

In the state of the art, different concepts are known. A first concept is reflected sound rendering for object-based audio (see [2]). Snap to speaker location information is included in a metadata definition as useful rendering information. However, in [2], no

information is provided how the information is used in the playback process. Moreover, no information is provided how a distance between two positions is determined.

Another concept of the state of the art, system and tools for enhanced 3D audio authoring and rendering is described in [5]. Fig. 6B of document [5] is a diagram illustrating how a "snapping" to a speaker might be algorithmically realized. In detail, according to the document [5] if it is determined to snap the audio object position to a speaker location (see block 665 of Fig. 6B of document [5]), the audio object position will be mapped to a speaker location (see block 670 of Fig. 6B of document [5]), generally the one closest to the intended (x,y,z) position received for the audio object. According to [5], the snapping might be applied to a small group of reproduction speakers and/or to an individual reproduction speaker. However, [5] employs Cartesian (x,y,z) coordinates instead of spherical coordinates. Moreover, the renderer behavior is just described as map audio object position to a speaker location; if the snap flag is one, no detailed description is provided. Furthermore, no details are provided how the closest speaker is determined.

According to another prior art, System and Method for Adaptive Audio Signal Generation, Coding and Rendering, described in document [1], metadata information (metadata elements) specify that "one or more sound components are rendered to a speaker feed for playback through a speaker nearest an intended playback location of the sound component, as indicated by the position metadata". However, no information is provided, how the nearest speaker is determined.

In a further prior art, audio definition model, described in document [4], a metadata flag is defined called "channelLock". If set to 1, a renderer can lock the object to the nearest channel or speaker, rather than normal rendering. However, no determination of the nearest channel is described.

In another prior art, upmixing of object based audio is described (see [3]). Document [3] describes a method for the usage of a distance measure of speakers in a different field of application: Here it is used for upmixing object-based audio material. The rendering system is configured to determine, from an object based audio program (and knowledge of the positions of the speakers to be employed to play the program), the distance between each position of an audio source indicated by the program and the position of each of the speakers. Furthermore, the rendering system of [3] is configured to determine, for each actual source position (e.g., each source position along a source trajectory) indicated by the program, a subset of the full set of speakers (a "primary" subset) consisting of those speakers of the full set which are (or the speaker of the full set which

is) closest to the actual source position, where "closest" in this context is defined in some reasonably defined sense. However, no information is provided how the distance should be calculated.

5 The object of the present invention is to provide improved concepts for audio rendering. The object of the present invention is solved by an apparatus according to claim 1, by a decoder device according to claim 13, by a method according to claim 14 and by a computer program according to claim 15.

10 An apparatus for playing back an audio object associated with a position is provided. The apparatus comprises a distance calculator for calculating distances of the position to speakers or for reading the distances of the position to the speakers. The distance calculator is configured to take a solution with a smallest distance. The apparatus is configured to play back the audio object using the speaker corresponding to the solution.

15 According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers or to read the distances of the position to the speakers only if a closest speaker playout flag (mdae\_closestSpeakerPlayout), being received by the apparatus, is enabled. Moreover, the distance calculator may, e.g., be  
20 configured to take a solution with a smallest distance only if the closest speaker playout flag (mdae\_closestSpeakerPlayout) is enabled. Furthermore, the apparatus may, e.g., be configured to play back the audio object using the speaker corresponding to the solution only if the closest speaker playout flag (mdae\_closestSpeakerPlayout) is enabled.

25 In an embodiment, the apparatus may, e.g., be configured to not conduct any rendering on the audio object, if the closest speaker playout flag (mdae\_closestSpeakerPlayout) is enabled.

30 According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances depending on a distance function which returns a weighted Euclidian distance or a great-arc distance.

In an embodiment, the distance calculator may, e.g., be configured to calculate the distances depending on a distance function which returns weighted absolute differences  
35 in azimuth and elevation angles.

According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances depending on a distance function which returns weighted absolute

differences to the power  $p$ , wherein  $p$  is a number. In an embodiment,  $p$  may, e.g., be set to  $p = 2$ .

5 According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances depending on a distance function which returns a weighted angular difference.

In an embodiment, the distance function may, e.g., be defined according to

$$10 \quad \text{diffAngle} = \text{acos}(\cos(\text{azDiff}) * \cos(\text{elDiff})),$$

wherein  $\text{azDiff}$  indicates a difference of two azimuth angles, wherein  $\text{elDiff}$  indicates a difference of two elevation angles, and wherein  $\text{diffAngle}$  indicates the weighted angular difference.

15 According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to

$$20 \quad \Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2|$$

$\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position, and  $\beta_2$  indicates an elevation angle of said one of the speakers. Or  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers, and  $\beta_2$  indicates an elevation angle of the position.

25 In an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to

$$30 \quad \Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|$$

35  $\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  $\beta_2$  indicates an elevation angle of said one of the speakers,  $r_1$  indicates a radius of the position and  $r_2$  indicates a radius of said one of the speakers. Or  $\alpha_1$  indicates an azimuth angle of said one of the

speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers,  $\beta_2$  indicates an elevation angle of the position,  $r_1$  indicates a radius of said one of the speakers and  $r_2$  indicates a radius of the position.

- 5 According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2|$$

10

- $\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  $\beta_2$  indicates an elevation angle of said one of the speakers,  $a$  is a first number, and  $b$  is a second number. Or  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers,  $\beta_2$  indicates an elevation angle of the position,  $a$  is a first number, and  $b$  is a second number.
- 15

- In an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to
- 20

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|$$

- $\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  $\beta_2$  indicates an elevation angle of said one of the speakers,  $r_1$  indicates a radius of the position,  $r_2$  indicates a radius of said one of the speakers,  $a$  is a first number, and  $b$  is a second number. Or,  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers, and  $\beta_2$  indicates an elevation angle of the position,  $r_1$  indicates a radius of said one of the speakers, and  $r_2$  indicates a radius of the position,  $a$  is a first number,  $b$  is a second number, and  $c$  is a third number.
- 25
- 30

- According to an embodiment, a decoder device is provided. The decoder device comprises a USAC decoder for decoding a bitstream to obtain one or more audio input channels, to obtain one or more input audio objects, to obtain compressed object metadata and to obtain one or more SAOC transport channels. Moreover, the decoder
- 35

device comprises an SAOC decoder for decoding the one or more SAOC transport channels to obtain a group of one or more rendered audio objects. Furthermore, the decoder device comprises an object metadata decoder for decoding the compressed object metadata to obtain uncompressed metadata. Moreover, the decoder device  
5 comprises a format converter for converting the one or more audio input channels to obtain one or more converted channels. Furthermore, the decoder device comprises a mixer for mixing the one or more rendered audio objects of the group of one or more rendered audio objects, the one or more input audio objects and the one or more converted channels to obtain one or more decoded audio channels. The object metadata  
10 decoder and the mixer together form an apparatus according to one of the above-described embodiments. The object metadata decoder comprises the distance calculator of the apparatus according to one of the above-described embodiments, wherein the distance calculator is configured, for each input audio object of the one or more input audio objects, to calculate distances of the position associated with said input audio  
15 object to speakers or for reading the distances of the position associated with said input audio object to the speakers, and to take a solution with a smallest distance. The mixer is configured to output each input audio object of the one or more input audio objects within one of the one or more decoded audio channels to the speaker corresponding to the solution determined by the distance calculator of the apparatus according to one of the  
20 above-described embodiments for said input audio object.

A method for playing back an audio object associated with a position, comprising:

- 25 - Calculating distances of the position to speakers or reading the distances of the position to the speakers.
- Taking a solution with a smallest distance. And:
- 30 - Playing back the audio object using the speaker corresponding to the solution.

Moreover, a computer program for implementing the above-described method when being executed on a computer or signal processor is provided.

In the following, embodiments of the present invention are described in more detail with  
35 reference to the figures, in which:

Fig. 1 is an apparatus according to an embodiment,

Fig. 2 illustrates an object renderer according to an embodiment,

Fig. 3 illustrates an object metadata processor according to an embodiment,

5 Fig. 4 illustrates an overview of a 3D-audio encoder,

Fig. 5 illustrates an overview of a 3D-Audio decoder according to an embodiment, and

10 Fig. 6 illustrates a structure of a format converter.

Fig. 1 illustrates an apparatus 100 for playing back an audio object associated with a position is provided.

15

The apparatus 100 comprises a distance calculator 110 for calculating distances of the position to speakers or for reading the distances of the position to the speakers. The distance calculator 110 is configured to take a solution with a smallest distance.

20 The apparatus 100 is configured to play back the audio object using the speaker corresponding to the solution.

For example, for each loudspeaker, a distance between the position (the audio object position) and said loudspeaker (the location of said loudspeaker) is determined.

25

According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers or to read the distances of the position to the speakers only if a closest speaker playout flag (mdae\_closestSpeakerPlayout), being received by the apparatus 100, is enabled. Moreover, the distance calculator may, e.g., be configured to take a solution with a smallest distance only if the closest speaker playout flag (mdae\_closestSpeakerPlayout) is enabled. Furthermore, the apparatus 100 may, e.g., be configured to play back the audio object using the speaker corresponding to the solution only if the closest speaker playout flag (mdae\_closestSpeakerPlayout) is enabled.

30

In an embodiment, the apparatus 100 may, e.g., be configured to not conduct any rendering on the audio object, if the closest speaker playout flag (mdae\_closestSpeakerPlayout) is enabled.

According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances depending on a distance function which returns a weighted Euclidian distance or a great-arc distance.

5

In an embodiment, the distance calculator may, e.g., be configured to calculate the distances depending on a distance function which returns weighted absolute differences in azimuth and elevation angles.

10 According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances depending on a distance function which returns weighted absolute differences to the power  $p$ , wherein  $p$  is a number. In an embodiment,  $p$  may, e.g., be set to  $p = 2$ .

15 According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances depending on a distance function which returns a weighted angular difference.

In an embodiment, the distance function may, e.g., be defined according to

20

$$\text{diffAngle} = \text{acos}(\cos(\text{azDiff}) * \cos(\text{elDiff})),$$

wherein  $\text{azDiff}$  indicates a difference of two azimuth angles, wherein  $\text{elDiff}$  indicates a difference of two elevation angles, and wherein  $\text{diffAngle}$  indicates the weighted angular

25

According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to

30

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2|$$

$\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position, and  $\beta_2$  indicates an elevation angle of said one of the speakers. Or,  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers, and  $\beta_2$  indicates an elevation angle of the position.

35

In an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to

5

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|$$

$\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  $\beta_2$  indicates an elevation angle of said one of the speakers,  $r_1$  indicates a radius of the position and  $r_2$  indicates a radius of said one of the speakers. Or  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers,  $\beta_2$  indicates an elevation angle of the position,  $r_1$  indicates a radius of said one of the speakers and  $r_2$  indicates a radius of the position.

15

According to an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to

20

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2|$$

$\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  $\beta_2$  indicates an elevation angle of said one of the speakers,  $a$  is a first number, and  $b$  is a second number. Or  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers,  $\beta_2$  indicates an elevation angle of the position,  $a$  is a first number, and  $b$  is a second number.

25

In an embodiment, the distance calculator may, e.g., be configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to

30

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|$$

35

$\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  $\beta_2$  indicates an elevation angle of said one of the speakers,  $r_1$  indicates a radius of the position,  $r_2$  indicates a

radius of said one of the speakers,  $a$  is a first number,  $b$  is a second number, and  $c$  is a third number. Or,  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers, and  $\beta_2$  indicates an elevation angle of the position,  $r_1$  indicates a radius of said  
5 one of the speakers, and  $r_2$  indicates a radius of the position,  $a$  is a first number,  $b$  is a second number, and  $c$  is a third number.

In the following, embodiments of the present invention are described. The embodiments provide concepts for using a geometric distance definition for audio rendering.  
10

Object metadata can be used to define either:

- 1) where in space an object should be rendered, or
- 15 2) which loudspeaker should be used to play back the object.

If the position of the object indicated in the metadata does not fall on a single speaker, the object renderer would create the output signal based by using multiple loudspeakers and defined panning rules. Panning is suboptimal in terms of localizing sounds or the sound  
20 color.

Therefore, it may be desirable by the producer of object based content, to define that a certain sound should come from a single loudspeaker from a certain direction.

25 It may happen that this loudspeaker does not exist in the users loudspeaker setup. Then a flag is set in the metadata that forces the sound to be played back by the nearest available loudspeaker without rendering.

The invention describes how the closest loudspeaker can be found allowing for some weighting to account for a tolerable deviation from the desired object position.  
30

Fig. 2 illustrates an object renderer according to an embodiment.

In object-based audio formats metadata are stored or transmitted along with object  
35 signals. The audio objects are rendered on the playback side using the metadata and information about the playback environment. Such information is e.g. the number of loudspeakers or the size of the screen.

Table 1 – Example metadata:

|                 |                                                                                                                                                                                                                                                                                                                   |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                 | ObjectID                                                                                                                                                                                                                                                                                                          |
| Dynamic<br>OAM  | Azimuth<br>Elevation<br>Gain<br>Distance                                                                                                                                                                                                                                                                          |
| Interactivity   | AllowOnOff<br>AllowPositionInteractivity<br>AllowGainInteractivity<br>DefaultOnOff<br>DefaultGain<br>InteractivityMinGain<br>InteractivityMaxGain<br>InteractivityMinAzOffset<br>InteractivityMaxAzOffset<br>InteractivityMinElOffset<br>InteractivityMaxElOffset<br>InteractivityMinDist<br>InteractivityMaxDist |
| Playout         | IsSpeakerRelatedGroup<br>SpeakerConfig3D<br><br>AzimuthScreenRelated<br><br>ElevationScreenRelated<br><br>ClosestSpeakerPlayout                                                                                                                                                                                   |
| Content         | ContentKind<br>ContentLanguage                                                                                                                                                                                                                                                                                    |
| Group           | GroupID<br>GroupDescription<br>GroupNumMembers<br>GroupMembers<br>Priority                                                                                                                                                                                                                                        |
| Switch<br>Group | SwitchGroupID<br>SwitchGroupDescription<br>SwitchGroupDefault<br>SwitchGroupNumMembers<br>SwitchGroupMembers                                                                                                                                                                                                      |
| Audio<br>Scene  | NumGroupsTotal<br>IsMainScene<br>NumGroupsPresent<br>NumSwitchGroups                                                                                                                                                                                                                                              |

For objects geometric metadata can be used to define how they should be rendered, e.g. angles in azimuth or elevation or absolute positions relative to a reference point, e.g. the listener. The renderer calculates loudspeaker signals on the basis of the geometric data and the available speakers and their position.

If an audio-object (audio signal associated with a position in the 3D space, e.g. azimuth, elevation and distance given) should not be rendered to its associated position, but instead played back by a loudspeaker that exists in the local loudspeaker setup, one way would be to define the loudspeaker where the object should be played back by means of metadata.

Nevertheless, there are cases where the producer does not want the object content to be played-back by a specific speaker, but rather by the next available speaker, i.e. the "geometrically nearest" speaker. This allows for a discrete playback without the necessity to define which speaker corresponds to which audio signal or to do rendering between multiple loudspeakers.

Embodiments according to the present invention emerge from the above in the following manner.

Metadata fields:

|                              |                                                                                                                                     |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| <b>ClosestSpeakerPlayout</b> | object should be played back by geometrically nearest speaker, no rendering (only for dynamic objects (IsSpeakerRelatedGroup == 0)) |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|

25

**Table 2 — Syntax of GroupDefinition():**

| Syntax                                    | No. of bits | Mnemonic |
|-------------------------------------------|-------------|----------|
| mdae_GroupDefinition( numGroups )         |             |          |
| {                                         |             |          |
| for ( grp = 0; grp < numGroups; grp++ ) { |             |          |
| mdae_groupID[grp];                        | 7           | uimsbf   |
| ...                                       |             |          |

|                                        |          |               |
|----------------------------------------|----------|---------------|
| <b>mdae_groupPriority[grp];</b>        | <b>3</b> | <b>uimsbf</b> |
| <b>mdae_closestSpeakerPayout[grp];</b> | <b>1</b> | <b>bslbf</b>  |
| ...                                    |          |               |
| }                                      |          |               |
| }                                      |          |               |

**mdae\_closestSpeakerPayout** This flag defines that the members of the metadata element group should not be rendered but directly be played back by the speakers which are nearest to the geometric position of the members.

The remapping is done in an object metadata processor that takes the local loudspeaker setup into account and performs a routing of the signals to the corresponding renderers with specific information by which loudspeaker or from which direction a sound should be rendered.

Fig. 3 illustrates an object metadata processor according to an embodiment.

A strategy for distance calculation is described as follows:

- if closest loudspeaker metadata flag is set, sound is played back over the closest speaker
- to this end, the distance to next speakers is calculated (or read from a pre-stored table)
- solution with smallest distance is taken
- distance function can be, for instance (but not limited to):
  - weighted euclidian or great-arc distance
  - weighted absolute differences in azimuth and elevation angle
  - weighted absolute differences to the power  $p$  ( $p=2 \Rightarrow$  Least Squares Solution)

- weighted angular difference, e.g.  $\text{diffAngle} = \text{acos}(\cos(\text{azDiff}) * \cos(\text{elDiff}))$

Examples for closest speaker calculation are set out below.

5

If the *mdae\_closestSpeakerLayout* flag of an audio element group is enabled, the members of the audio element group shall each be played back by the speaker that is nearest to the given position of the audio element. No rendering is applied.

- 10 The distance of two positions  $P_1$  and  $P_2$  in a spherical coordinate system is defined as the absolute difference of their azimuth angles  $\alpha$  and elevation angles  $\beta$ .

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|$$

- 15 This distance has to be calculated for all known positions  $P_1$  to  $P_N$  of the N output speakers with respect to the wanted position of the audio element  $P_{\text{wanted}}$ .

The nearest known loudspeaker position is the one, where the distance to the wanted position of the audio element gets minimal

20

$$P_{\text{next}} = \min(\Delta(P_{\text{wanted}}, P_1), \Delta(P_{\text{wanted}}, P_2), \dots, \Delta(P_{\text{wanted}}, P_N))$$

With this formula, it is possible to add weights to elevation, azimuth and/or radius. In that way it is possible to state that an azimuth deviation should be less tolerable than an elevation deviation by weighting the azimuth deviation by a high number:

25

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|$$

An example concerns a closest loudspeaker calculation for binaural rendering.

30

If audio content should be played back as a binaural stereo signal over headphones or a stereo speaker setup, each channel of the audio content is traditionally mathematically combined with a binaural room impulse response or a head-related impulse response.

- 35 The measuring position of this impulse response has to correspond to the direction from which the audio content of the associated channel should be perceived. In multi-channel audio systems or object-based audio there is the case that the number of definable

positions (either by a speaker or by an object position) is larger than the number of available impulse responses. In that case, an appropriate impulse response has to be chosen if there is no dedicated one available for the channel position or the object position. To inflict only minimum positional changes in the perception, the chosen impulse response should be the “geometrically nearest” impulse response.

It is in both cases needed to determine, which of the list of known positions (i.e. playback speakers or BRIRs) is the next to the wanted position (BRIR = Binaural Room Impulse Response). Therefore a “distance” between different positions has to be defined.

The distance between different positions is here defined as the absolute difference of their azimuth and elevation angles.

The following formula is used to calculate a distance of two positions  $P_1, P_2$  in a coordinate system that is defined by elevation  $\alpha$  and azimuth  $\beta$ :

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2|$$

It is possible to add the radius  $r$  as a third variable:

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|$$

The nearest known position is the one, where the distance to the wanted position gets minimal

$$P_{next} = \min(\Delta(P_{wanted}, P_1), \Delta(P_{wanted}, P_2), \dots, \Delta(P_{wanted}, P_N)).$$

In an embodiment, weights may, e.g., be added to elevation, azimuth and/or radius:

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|.$$

According to some embodiments, the closest speaker may, e.g., be determined as follows:

The distance of two positions  $P_1$  and  $P_2$  in a spherical coordinate system may, e.g., be defined as the absolute difference of their azimuth angles  $\varphi$  and elevation angles  $\theta$ .

$$\Delta(P_1, P_2) = |\theta_1 - \theta_2| + |\varphi_1 - \varphi_2|$$

This distance has to be calculated for all known position  $P_1$  to  $P_N$  of the N output speakers with respect to the wanted position of the audio element  $P_{wanted}$ .

5

The nearest known loudspeaker position is the one, where the distance to the wanted position of the audio element gets minimal:

$$P_{next} = \min(\Delta(P_{wanted}, P_1), \Delta(P_{wanted}, P_2), \dots, \Delta(P_{wanted}, P_N))$$

10

For example, according to some embodiments, the closest speaker ployout processing according to some embodiments may be conducted by determining the position of the closest existing loudspeaker for each member of the group of audio objects, if the ClosestSpeakerPloyout flag is equal to one.

15

The closest speaker ployout processing may, e.g., be particularly meaningful for groups of elements with dynamic position data. The nearest known loudspeaker position may, e.g., be the one, where the distance to the desired/wanted position of the audio element gets minimal.

20

In the following, a system overview of a 3D audio codec system is provided. Embodiments of the present invention may be employed in such a 3D audio codec system. The 3D audio codec system may, e.g., be based on an MPEG-D USAC Codec for coding of channel and object signals.

25

According to embodiments, to increase the efficiency for coding a large amount of objects, MPEG SAOC technology has been adapted (SAOC = Spatial Audio Object Coding). For example, according to some embodiments, three types of renderers may, e.g., perform the tasks of rendering objects to channels, rendering channels to headphones or rendering channels to a different loudspeaker setup.

30

When object signals are explicitly transmitted or parametrically encoded using SAOC, the corresponding object metadata information is compressed and multiplexed into the 3D-audio bitstream.

35

Fig. 4 and Fig. 5 show the different algorithmic blocks of the 3D-Audio system. In particular, Fig. 4 illustrates an overview of a 3D-audio encoder. Fig. 5 illustrates an overview of a 3D-Audio decoder according to an embodiment.

5 Possible embodiments of the modules of Fig. 4 and Fig. 5 are now described.

In Fig. 4, a prerenderer 810 (also referred to as mixer) is illustrated. In the configuration of Fig. 4, the prerenderer 810 (mixer) is optional. The prerenderer 810 can be optionally used to convert a Channel+Object input scene into a channel scene before encoding.  
10 Functionally the prerenderer 810 on the encoder side may, e.g., be related to the functionality of object renderer/mixer 920 on the decoder side, which is described below. Prerendering of objects ensures a deterministic signal entropy at the encoder input that is basically independent of the number of simultaneously active object signals. With prerendering of objects, no object metadata transmission is required. Discrete Object  
15 Signals are rendered to the Channel Layout that the encoder is configured to use. The weights of the objects for each channel are obtained from the associated object metadata (OAM).

The core codec for loudspeaker-channel signals, discrete object signals, object downmix signals and pre-rendered signals is based on MPEG-D USAC technology (USAC Core  
20 Codec). The USAC encoder 820 (e.g., illustrated in Fig. 4) handles the coding of the multitude of signals by creating channel- and object mapping information based on the geometric and semantic information of the input's channel and object assignment. This mapping information describes, how input channels and objects are mapped to USAC-  
25 Channel Elements (CPEs, SCEs, LFEs) and the corresponding information is transmitted to the decoder.

All additional payloads like SAOC data or object metadata have been passed through extension elements and may, e.g., be considered in the USAC encoder's rate control.  
30

The coding of objects is possible in different ways, depending on the rate/distortion requirements and the interactivity requirements for the renderer. The following object coding variants are possible:

- 35 - Prerendered objects: Object signals are prerendered and mixed to the 22.2 channel signals before encoding. The subsequent coding chain sees 22.2 channel signals.

- Discrete object waveforms: Objects are supplied as monophonic waveforms to the USAC encoder 820. The USAC encoder 820 uses single channel elements SCEs to transmit the objects in addition to the channel signals. The decoded objects are rendered and mixed at the receiver side. Compressed object metadata information is transmitted to the receiver/renderer alongside.
- Parametric object waveforms: Object properties and their relation to each other are described by means of SAOC parameters. The down-mix of the object signals is coded with USAC by the USAC encoder 820. The parametric information is transmitted alongside. The number of downmix channels is chosen depending on the number of objects and the overall data rate. Compressed object metadata information is transmitted to the SAOC renderer.

On the decoder side, a USAC decoder 910 conducts USAC decoding.

Moreover, according to embodiments, a decoder is provided, see Fig. 5. The decoder comprises a USAC decoder 910 for decoding a bitstream to obtain one or more audio input channels, to obtain one or more audio objects, to obtain compressed object metadata and to obtain one or more SAOC transport channels.

Furthermore, the decoder comprises an SAOC decoder 915 for decoding the one or more SAOC transport channels to obtain a first group of one or more rendered audio objects.

Furthermore, the decoder comprises a format converter 922 for converting the one or more audio input channels to obtain one or more converted channels.

Moreover, the decoder comprises a mixer 930 for mixing the audio objects of the first group of one or more rendered audio objects, the audio object of the second group of one or more rendered audio objects and the one or more converted channels to obtain one or more decoded audio channels.

In Fig. 5 a particular embodiment of a decoder is illustrated. The SAOC encoder 815 (the SAOC encoder 815 is optional, see Fig. 4) and the SAOC decoder 915 (see Fig. 5) for object signals are based on MPEG SAOC technology. The system is capable of recreating, modifying and rendering a number of audio objects based on a smaller number of transmitted channels and additional parametric data (OLDs, IOCs, DMGs) (OLD = object level difference, IOC = inter object correlation, DMG = downmix gain). The

additional parametric data exhibits a significantly lower data rate than required for transmitting all objects individually, making the coding very efficient.

5 The SAOC encoder 815 takes as input the object/channel signals as monophonic waveforms and outputs the parametric information (which is packed into the 3D-Audio bitstream) and the SAOC transport channels (which are encoded using single channel elements and transmitted).

10 The SAOC decoder 915 reconstructs the object/channel signals from the decoded SAOC transport channels and parametric information, and generates the output audio scene based on the reproduction layout, the decompressed object metadata information and optionally on the user interaction information.

15 Regarding object metadata codec, for each object, the associated metadata that specifies the geometrical position and spread of the object in 3D space is efficiently coded by quantization of the object properties in time and space, e.g., by the metadata encoder 818 of Fig. 4. The compressed object metadata cOAM (cOAM = compressed audio object metadata) is transmitted to the receiver as side information. At the receiver the cOAM is decoded by the metadata decoder 918.

20 For example, in Fig. 5, the metadata decoder 918 may, e.g., implement the distance calculator 110 of Fig. 1 according to one of the above-described embodiments.

25 An object renderer, e.g., object renderer 920 of Fig. 5, utilizes the compressed object metadata to generate object waveforms according to the given reproduction format. Each object is rendered to certain output channels according to its metadata. The output of this block results from the sum of the partial results. In some embodiments, if determination of the closest loudspeaker is conducted, the object renderer 920, may, for example, pass the audio objects, received from the USAC-3D decoder 910, without rendering them to the mixer 930. The mixer 930 may, for example, pass the audio objects to the loudspeaker that was determined by the distance calculator (e.g., implemented within the meta-data decoder 918) to the loudspeakers. By this according to an embodiment, the meta-data decoder 918 which may, e.g., comprise a distance calculator, the mixer 930 and, optionally, the object renderer 920 may together implement the apparatus 100 of Fig. 1.

35 For example, the meta-data decoder 918 comprises a distance calculator (not shown) and said distance calculator or the meta-data decoder 918 may signal, e.g., by a connection (not shown) to the mixer 930, the closest loudspeaker for each audio object of the one or

more audio objects received from the USAC-3D decoder. The mixer 930 may then output the audio object within a loudspeaker channel only to the closest loudspeaker (determined by the distance calculator) of the plurality of loudspeakers.

- 5 In some other embodiments, the closest loudspeaker is only signaled for one or more of the audio objects by the distance calculator or the meta-data decoder 918 to the mixer 930.

10 If both channel based content as well as discrete/parametric objects are decoded, the channel based waveforms and the rendered object waveforms are mixed before outputting the resulting waveforms, e.g., by mixer 930 of Fig. 5 (or before feeding them to a postprocessor module like the binaural renderer or the loudspeaker renderer module).

15 A binaural renderer module 940, may, e.g., produce a binaural downmix of the multichannel audio material, such that each input channel is represented by a virtual sound source. The processing is conducted frame-wise in QMF domain. The binauralization may, e.g., be based on measured binaural room impulse responses.

20 A loudspeaker renderer 922 may, e.g., convert between the transmitted channel configuration and the desired reproduction format. It is thus called format converter 922 in the following. The format converter 922 performs conversions to lower numbers of output channels, e.g., it creates downmixes. The system automatically generates optimized downmix matrices for the given combination of input and output formats and applies these matrices in a downmix process. The format converter 922 allows for standard loudspeaker configurations as well as for random configurations with non-standard loudspeaker positions.

30 According to embodiments, a decoder device is provided. The decoder device comprises a USAC decoder 910 for decoding a bitstream to obtain one or more audio input channels, to obtain one or more input audio objects, to obtain compressed object metadata and to obtain one or more SAOC transport channels.

Moreover, the decoder device comprises an SAOC decoder 915 for decoding the one or more SAOC transport channels to obtain a group of one or more rendered audio objects.

35

Furthermore, the decoder device comprises an object metadata decoder 918 for decoding the compressed object metadata to obtain uncompressed metadata.

Moreover, the decoder device comprises a format converter 922 for converting the one or more audio input channels to obtain one or more converted channels.

5 Furthermore, the decoder device comprises a mixer 930 for mixing the one or more rendered audio objects of the group of one or more rendered audio objects, the one or more input audio objects and the one or more converted channels to obtain one or more decoded audio channels.

10 The object metadata decoder 918 and the mixer 930 together form an apparatus 100 according to one of the above-described embodiments, e.g., according to the embodiment of Fig. 1.

15 The object metadata decoder 918 comprises the distance calculator 110 of the apparatus 100 according to one of the above-described embodiments, wherein the distance calculator 110 is configured, for each input audio object of the one or more input audio objects, to calculate distances of the position associated with said input audio object to speakers or for reading the distances of the position associated with said input audio object to the speakers, and to take a solution with a smallest distance.

20 The mixer 930 is configured to output each input audio object of the one or more input audio objects within one of the one or more decoded audio channels to the speaker corresponding to the solution determined by the distance calculator 110 of the apparatus 100 according to one of the above-described embodiments for said input audio object.

25 In such embodiments, the object renderer 920 may, e.g., be optional. In some embodiments, the object renderer 920 may be present, but may only render input audio objects if metadata information indicates that a closest speaker ployout is deactivated. If metadata information indicates that closest speaker ployout is activated, then the object renderer 920 may, e.g., pass the input audio objects directly to the mixer without  
30 rendering the input audio objects.

Fig. 6 illustrates a structure of a format converter. Fig. 6 illustrates a downmix configurator 1010 and a downmix processor for processing the downmix in the QMF domain (QMF domain = quadrature mirror filter domain).

35

In the following, further embodiments and concepts of embodiments of the present invention are described.

In embodiments, the audio objects may, e.g., be rendered, e.g., by an object renderer, on the playback side using the metadata and information about the playback environment. Such information may, e.g., be the number of loudspeakers or the size of the screen. The object renderer may, e.g., calculate loudspeaker signals on the basis of the geometric data and the available speakers and their positions.

User control of objects may, e.g., be realized by descriptive metadata, e.g., by information about the existence of an object inside the bitstream and high-level properties of objects, or, may, e.g., be realized by restrictive metadata, e.g., information on how interaction is possible or enabled by the content creator.

According to embodiments, signaling, delivery and rendering of audio objects may, e.g., be realized by positional metadata, e.g., by structural metadata, for example, grouping and hierarchy of objects, e.g., by the ability to render to specific speaker and to signal channel content as objects, and, e.g., by means to adapt object scene to screen size.

Therefore, new metadata fields were developed in addition to the already defined geometrical position and level of the object in 3D space.

In general, the position of an object is defined by a position in 3D space that is indicated in the metadata.

This playback loudspeaker can be a specific speaker that exists in the local loudspeaker setup. In this case the wanted loudspeaker can be directly defined by the means of metadata.

Nevertheless, there are cases where the producer does not want the object content to be played-back by a specific speaker, but rather by the next available speaker, e.g., the "geometrically nearest" speaker. This allows for a discrete playback without the necessity to define which speaker corresponds to which audio signal. This is useful as the reproduction loudspeaker layout may be unknown to the producer, such that he might not know which speakers he can choose of.

Embodiments provides a simple definition of a distance function that does not need any square root operations or cos/sin functions. In embodiments, the distance function works in angular domain (azimuth, elevation, distance), so no transform to any other coordinate system (Cartesian, longitude/latitude) is needed. According to embodiments, there are weights in the function that provide a possibility to shift the focus between azimuth

deviation, elevation deviation and radius deviation. The weights in the function might, e.g., be adjusted to the abilities of human hearing (e.g. adjust weights according to the just noticeable difference in azimuth and elevation direction). The function could not only be applied for the determination of the closest speaker, but also for choosing a binaural room impulse response or head-related impulse response for binaural rendering. No interpolation of impulse responses is needed in this case, instead the "closest" impulse response can be used.

According to an embodiment, a "ClosestSpeakerPayout" flag called *mae\_closestSpeakerPayout* may, e.g., be defined in the object-based metadata that forces the sound to be played back by the nearest available loudspeaker without rendering. An object may, e.g., be marked for playback by the closest speaker if its "ClosestSpeakerPayout" flag is set to one. The "ClosestSpeakerPayout" flag may, e.g., be defined on a level of a "group" of objects. A group of objects is a concept of a gathering of related objects that should be rendered or modified as a union. If this flag is set to one, it is applicable for all members of the group.

According to embodiments, for determining the closest speaker, if the *mae\_closestSpeakerPayout* flag of a group, e.g., a group of audio objects, is enabled, the members of the group shall each be played back by the speaker that is nearest to the given position of the object. No rendering is applied. If the "ClosestSpeakerPayout" is enabled for a group, then the following processing is conducted:

For each of the group members, the geometric position of the member is determined (from the dynamic object metadata (OAM)), and the closest speaker is determined, either by lookup in a pre-stored table or by calculation with help of a distance measure. The distance of the member's position to every (or only a subset) of the existing speakers is calculated. The speaker that yields the minimum distance is defined to be the closest speaker, and the member is routed to its closest speaker. The group members are played back each by its closest speaker.

As already described, the distance measures for the determination of the closest speaker may, for example, be implemented as:

- The weighted absolute differences in azimuth and elevation angle
- The weighted absolute differences in azimuth, elevation and radius/distance and for instance (but not limited to):
- The weighted absolute differences to the power  $p$  ( $p=2 \Rightarrow$  Least Squares Solution)

- (Weighted) Pythagorean Theorem / Euclidean Distance

The distance  $d$  for Cartesian coordinates may, e.g., be realized by employing the formula

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

with  $x_1, y_1, z_1$  being the x-, y- and z-coordinate values of a first position, with  $x_2, y_2, z_2$  being the x-, y- and z-coordinate values of a second position, and with  $d$  being the distance between the first and the second position.

A distance measure  $d$  for polar coordinates may, e.g., be realized by employing the formula:

$$d = \sqrt{a \cdot (\alpha_1 - \alpha_2)^2 + b \cdot (\beta_1 - \beta_2)^2 + c \cdot (r_1 - r_2)^2}.$$

with  $\alpha_1, \beta_1$  and  $r_1$  being the polar coordinates of a first position, with  $\alpha_2, \beta_2$  and  $r_2$  being the polar coordinates of a second position, and with  $d$  being the distance between the first and the second position.

The weighted angular difference may, e.g., be defined according to

$$\text{diffAngle} = \text{acos}(\cos(\alpha_1 - \alpha_2) \cdot \cos(\beta_1 - \beta_2))$$

Regarding the orthodromic distance, the Great-Arc Distance, or the Great-Circle Distance, the distance measured along the surface of a sphere (as opposed to a straight line through the sphere's interior). Square root operations and trigonometric functions may, e.g., be employed. Coordinates may, e.g., be transformed to latitude and longitude.

Returning to the formula presented above:

30

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|,$$

the formula can be seen as a modified *Taxicab geometry* using polar coordinates instead of Cartesian coordinates as in the original taxicab geometry definition

35

$$\Delta(P_1, P_2) = |x_1 - x_2| + |y_1 - y_2|.$$

With this formula, it is possible to add weights to elevation, azimuth and/or radius. In that way it is possible to state that an azimuth deviation should be less tolerable than an elevation deviation by weighting the azimuth deviation by a high number:

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|.$$

As a further side remark, it should be noted, that in embodiments, the "rendered object audio" of Fig. 2 may, e.g., be considered as "rendered object-based audio". In Fig. 2, the usacConfigExtention regarding static object metadata and the usacExtension are only used as examples of particular embodiments.

Regarding Fig. 3. It should be noted that in some embodiments, the dynamic object metadata of Fig. 3 may, e.g., positional OAM (audio object metadata, positional data + gain). In some embodiments, the "route signals" may, e.g., be conducted by routing signals to a format converter or to an object renderer.

Although some aspects have been described in the context of an apparatus, it is clear that these aspects also represent a description of the corresponding method, where a block or device corresponds to a method step or a feature of a method step. Analogously, aspects described in the context of a method step also represent a description of a corresponding block or item or feature of a corresponding apparatus.

The inventive decomposed signal can be stored on a digital storage medium or can be transmitted on a transmission medium such as a wireless transmission medium or a wired transmission medium such as the Internet.

Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a digital storage medium, for example a floppy disk, a DVD, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed.

Some embodiments according to the invention comprise a non-transitory data carrier having electronically readable control signals, which are capable of cooperating with a programmable computer system, such that one of the methods described herein is performed.

Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing one of the methods when the computer program product runs on a computer. The program code may for example be stored on a machine readable carrier.

5

Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine readable carrier.

10

In other words, an embodiment of the inventive method is, therefore, a computer program having a program code for performing one of the methods described herein, when the computer program runs on a computer.

15

A further embodiment of the inventive methods is, therefore, a data carrier (or a digital storage medium, or a computer-readable medium) comprising, recorded thereon, the computer program for performing one of the methods described herein.

20

A further embodiment of the inventive method is, therefore, a data stream or a sequence of signals representing the computer program for performing one of the methods described herein. The data stream or the sequence of signals may for example be configured to be transferred via a data communication connection, for example via the Internet.

25

A further embodiment comprises a processing means, for example a computer, or a programmable logic device, configured to or adapted to perform one of the methods described herein.

30

In some embodiments, a programmable logic device (for example a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

35

The above described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent,

therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

Literature

- 5 [1] "System and Method for Adaptive Audio Signal Generation, Coding and Rendering", Patent application number: US20140133683 A1 (Claim 48)
- [2] "Reflected sound rendering for object-based audio", Patent application number: WO2014036085 A1 (Chapter Playback Applications)
- 10 [3] "Upmixing object based audio", Patent application number: US20140133682 A1 (BRIEF DESCRIPTION OF EXEMPLARY EMBODIMENTS + Claim 71 b))
- [4] "Audio Definition Model", EBU-TECH 3364,  
<https://tech.ebu.ch/docs/tech/tech3364.pdf>
- 15 [5] "System and Tools for Enhanced 3D Audio Authoring and Rendering", Patent application number: US20140119581 A1

Claims

1. An apparatus for playing back an audio object associated with a position, comprising:

5

a distance calculator for calculating distances of the position to speakers,

wherein the distance calculator is configured to take a solution with a smallest distance, and

10

wherein the apparatus is configured to play back the audio object using the speaker corresponding to the solution,

wherein the distance calculator is configured to calculate the distances depending on a distance function which returns a great-arc distance, or which returns weighted absolute differences in azimuth and elevation angles, or which returns a weighted angular difference.

15

2. An apparatus according to claim 1,

20

wherein the distance calculator is configured to calculate the distances of the position to the speakers only if a closest speaker playout flag, being received by the apparatus, is enabled,

wherein the distance calculator is configured to take a solution with a smallest distance only if the closest speaker playout flag is enabled, and

25

wherein the apparatus is configured to play back the audio object using the speaker corresponding to the solution only if the closest speaker playout flag is enabled.

30

3. An apparatus according to claim 2, wherein the apparatus is configured to not conduct any rendering on the audio object, if the closest speaker playout flag is enabled.

35

4. An apparatus according to one of the preceding claims, wherein the distance function is defined according to

$$\text{diffAngle} = \text{acos}(\cos(\text{azDiff}) * \cos(\text{elDiff})),$$

wherein azDiff indicates a difference of two azimuth angles,

5 wherein elDiff indicates a difference of two elevation angles, and

wherein diffAngle indicates the weighted angular difference.

5. An apparatus according to one of the preceding claims, wherein the distance  
10 calculator is configured to calculate the distances of the position to the speakers,  
so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated  
according to

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2|$$

15

wherein  $\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth  
angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  
and  $\beta_2$  indicates an elevation angle of said one of the speakers, or

20

wherein  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an  
azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the  
speakers, and  $\beta_2$  indicates an elevation angle of the position.

25

6. An apparatus according to one of claims 1 to 4,

25

wherein the distance calculator is configured to calculate the distances of the  
position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of  
the speakers is calculated according to

30

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|$$

35

wherein  $\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth  
angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  
 $\beta_2$  indicates an elevation angle of said one of the speakers,  $r_1$  indicates a radius of  
the position and  $r_2$  indicates a radius of said one of the speakers, or

wherein  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers,  $\beta_2$  indicates an elevation angle of the position,  $r_1$  indicates a radius of said one of the speakers and  $r_2$  indicates a radius of the position.

7. An apparatus according to one of claims 1 to 4,

wherein the distance calculator is configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2|$$

wherein  $\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  $\beta_2$  indicates an elevation angle of said one of the speakers,  $a$  is a first number, and  $b$  is a second number, or

wherein  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers,  $\beta_2$  indicates an elevation angle of the position,  $a$  is a first number, and  $b$  is a second number.

8. An apparatus according to one of claims 1 to 4,

wherein the distance calculator is configured to calculate the distances of the position to the speakers, so that each distance  $\Delta(P_1, P_2)$  of the position to one of the speakers is calculated according to

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|$$

wherein  $\alpha_1$  indicates an azimuth angle of the position,  $\alpha_2$  indicates an azimuth angle of said one of the speakers,  $\beta_1$  indicates an elevation angle of the position,  $\beta_2$  indicates an elevation angle of said one of the speakers,  $r_1$  indicates a radius of the position,  $r_2$  indicates a radius of said one of the speakers,  $a$  is a first number,  $b$  is a second number, and  $c$  is a third number, or

wherein  $\alpha_1$  indicates an azimuth angle of said one of the speakers,  $\alpha_2$  indicates an azimuth angle of the position,  $\beta_1$  indicates an elevation angle of said one of the speakers, and  $\beta_2$  indicates an elevation angle of the position,  $r_1$  indicates a radius of said one of the speakers, and  $r_2$  indicates a radius of the position,  $a$  is a first number,  $b$  is a second number, and  $c$  is a third number.

9. A decoder device comprising:

a USAC decoder for decoding a bitstream to obtain one or more audio input channels, to obtain one or more input audio objects, to obtain compressed object metadata and to obtain one or more SAOC transport channels,

an SAOC decoder for decoding the one or more SAOC transport channels to obtain a group of one or more rendered audio objects,

an object metadata decoder, for decoding the compressed object metadata to obtain uncompressed metadata,

a format converter for converting the one or more audio input channels to obtain one or more converted channels, and

a mixer for mixing the one or more rendered audio objects of the group of one or more rendered audio objects, the one or more input audio objects and the one or more converted channels to obtain one or more decoded audio channels,

wherein the object metadata decoder and the mixer together form an apparatus according to one of the preceding claims,

wherein the object metadata decoder comprises the distance calculator of the apparatus according to one of the preceding claims, wherein the distance calculator is configured, for each input audio object of the one or more input audio objects, to calculate distances of the position associated with said input audio object to speakers, and to take a solution with a smallest distance, and

wherein the mixer is configured to output each input audio object of the one or more input audio objects within one of the one or more decoded audio channels to the speaker corresponding to the solution determined by the distance calculator

of the apparatus according to one of the preceding claims for said input audio object.

10. A method for playing back an audio object associated with a position, comprising:

5

calculating distances of the position to speakers,

taking a solution with a smallest distance, and

10

playing back the audio object using the speaker corresponding to the solution,

wherein calculating the distances is conducted depending on a distance function which returns a great-arc distance, or which returns weighted absolute differences in azimuth and elevation angles, or which returns a weighted angular difference.

15

11. A computer program for implementing the method of claim 10 when being executed on a computer or signal processor.