Claims

1. Apparatus for encoding a plurality of audio objects and related metadata indicating direction information on the plurality of audio objects, comprising:

a downmixer for downmixing the plurality of audio objects to obtain one or more transport channels;

a transport channel encoder for encoding one or more transport channels to obtain one or more encoded transport channels; and

an output interface for outputting an encoded audio signal comprising the one or more encoded transport channels,

wherein the downmixer is configured to downmix the plurality of audio objects in response to the direction information on the plurality of audio objects.

2. Apparatus of claim 1, wherein the downmixer is configured

to generate two transport channels as two virtual microphone signals arranged at the same position and having different orientations or at two different positions with respect to a reference position or orientation such as a virtual listener position or orientation, or

to generate three transport channels as three virtual microphone signals arranged at the same position and having different orientations or at three different positions with respect to a reference position or orientation such as a virtual listener position or orientation, or

to generate four transport channels as four virtual microphone signals arranged at the same position and having different orientations or at four different positions with respect to a reference position or orientation such as a virtual listener position or orientation, or

wherein the virtual microphone signals are virtual first order microphone signals, or virtual cardioid microphone signals, or virtual figure of 8 or dipole or bidirectional microphone signals, or virtual directional microphone signals, or virtual subcardioid microphone

signals, or virtual unidirectional microphone signals, or virtual hypercardioid microphone signals, or virtual omnidirectional microphone signals.

3. Apparatus of claim 1 or 2, wherein the downmixer is configured

to derive, for each audio object of the plurality of audio objects, a weighting information for each transport channel using the direction information for the corresponding audio object;

to weight the corresponding audio object using the weighting information for the audio object for a specific transport channel to obtain an object contribution for the specific transport channel, and

to combine the object contributions for the specific transport channel from the plurality of audio objects to obtain the specific transport channel.

4. Apparatus of one of the preceding claims,

wherein the downmixer is configured to calculate the one or more transport channels as one or more virtual microphone signals arranged at the same position and having different orientations or at different positions with respect to a reference position or orientation such as a virtual listener position or orientation, to which the direction information is related,

wherein the different positions or orientations are on or to a left side of a center line and on or to a right side of the center line, or wherein the different positions or orientations are equally or non-equally distributed to horizontal positions or orientations such as +90 degrees or -90 degrees with respect to the center line or -120 degrees, 0 degrees and +120 degrees with respect to the center line, or wherein the different positions or orientations comprise at least one position or orientation being directed upwards or downwards with respect to a horizontal plane in which a virtual listener is placed, wherein the direction information on the plurality of audio objects is related to the virtual listener position or reference position or orientation.

5. Apparatus in accordance with one of the preceding claims, further comprising:

a parameter processor for quantizing the metadata indicating the direction information on the plurality of audio objects to obtain quantized direction items for the plurality of audio objects,

wherein the downmixer is configured to operate in response to the quantized direction items as the direction information, and

wherein the output interface is configured to introduce information on the quantized direction items into the encoded audio signal.

6. Apparatus of one of the preceding claims,

wherein the downmixer is configured to perform an analysis of the direction information on the plurality of audio objects and to place one or more virtual microphones for the generation of the transport channels depending on a result of the analysis.

7. Apparatus of one of the preceding claims,

wherein the downmixer is configured to downmix using a downmixing rule being static over the plurality of time frames, or

wherein the direction information is variable over a plurality of time frames, and wherein the downmixer is configured to downmix using a downmixing rule being variable over the plurality of time frames.

8. Apparatus of one of the preceding claims,

wherein the downmixer is configured to downmix in a time domain using a sample-by-sample weighting and combining of samples of the plurality of audio objects.

9. Apparatus of one of the preceding claims, further comprising:

an object parameter calculator configured for calculating, for one or more frequency bins of a plurality of frequency bins related to a time frame, parameter data for at least two

relevant audio objects, wherein a number of the at least two relevant audio objects is lower than a total number of the plurality of audio objects, and

wherein the output interface is configured to introduce information on the parameter data for the at least two relevant audio objects for the one or more frequency bins into the encoded audio signal.

10. Apparatus of claim 9, wherein the object parameter calculator is configured

to convert each audio object of the plurality of audio objects into a spectral representation having the plurality of frequency bins,

to calculate a selection information from each audio object for the one or more frequency bins, and

to derive object identifications as the parameter data indicating the at least two relevant audio objects, based on the selection information, and

wherein the output interface is configured to introduce information on the object identifications into the encoded audio signal.

11. Apparatus of claim 9 or 10, wherein the object parameter calculator is configured to quantize and encode one or more amplitude related measures or one or more combined values derived from the amplitude related measures of the relevant audio objects in the one or more frequency bins as the parameter data, and

wherein the output interface is configured to introduce the quantized one or more amplitude related measure or the quantized one or more combined values into the encoded audio signal.

12. Apparatus of claim 10 or 11,

wherein the selection information is an amplitude-related measure such as an amplitude value, a power value or a loudness value or an amplitude raised to a power being different from one for the audio object, and

wherein the object parameter calculator is configured to calculate a combined value such as a ratio from an amplitude related measure of a relevant audio object and a sum of two or more amplitude related measures of the relevant audio objects, and

wherein the output interface is configured to introduce an information on the combined value into the encoded audio signal, wherein a number of information items on the combined values in the encoded audio signal is equal to at least one and is lower than the number of relevant audio objects for the one or more frequency bins.

13. Apparatus of one of claims 10 to 12,

wherein the object parameter calculator is configured to select the object identifications based on an order of the selection information of the plurality of audio objects in the one or more frequency bins.

14. Apparatus of one of claims 10 to 13, wherein the object parameter calculator is configured

to calculate a signal power as the selection information,

to derive the object identifications for the two or more audio objects having the greatest signal power values in the corresponding one or more frequency bins for each frequency bin separately,

to calculate a power ratio between the sum of the signal powers of the two or more audio objects having the greatest signal power values and the signal power of at least one of the audio objects having the derived object identifications as the parameter data, and

to quantize and encode the power ratio, and

wherein the output interface is configured to introduce the quantized and encoded power ratio into the encoded audio signal.

15. Apparatus of one of claims 10 to 14, wherein the output interface is configured to introduce, into the encoded audio signal,

one or more encoded transport channels,

as the parameter data, two or more encoded object identifications for the relevant audio objects for each one of the one or more frequency bins of the plurality of frequency bins in the time frame, and one or more encoded combined values or encoded amplitude related measures, and

quantized and encoded direction data for each audio object in the time frame, the direction data being constant for all frequency bins of the one or more frequency bins.

16. Apparatus of one of claims 9 to 15, wherein the object parameter calculator is configured to calculate the parameter data for at least the most dominant object and the second most dominant object in the one or more frequency bins, or

wherein a number of audio objects of the plurality of audio objects is three or more, the plurality of audio objects comprising a first audio object, a second audio object and a third audio object, and

wherein the object parameter calculator is configured to calculate for a first one of the one or more frequency bins, as the relevant audio objects, only a first group of audio objects such as the first audio object and the second audio object, and to calculate, as the relevant audio objects for a second frequency bin of the one or more frequency bins, only a second group of audio objects, such as the second audio object and the third audio object or the first audio object and the third audio object, wherein the first group of audio objects is different from the second group of audio objects at least with respect to one group member.

17. Apparatus of one of claims 9 to 16, wherein the object parameter calculator is configured

to calculate raw parametric data with a first time or frequency resolution and to combine the raw parametric data into combined parametric data having a second time or frequency resolution being lower than the first time of frequency resolution, and, and to calculate the parameter data for the at least two relevant audio objects with respect to the combined parametric data having the second time or frequency resolution, or

to determine parameter bands having a second time or frequency resolution being different from a first time or frequency resolution used in a time or frequency decomposition of the plurality of audio objects, and to calculate the parameter data for the at least two relevant audio objects for the parameter bands having the second time or frequency resolution.

18. Decoder for decoding an encoded audio signal comprising one or more transport channels and direction information for a plurality of audio objects, and, for one or more frequency bins of a time frame, parameter data for an audio object, the decoder comprising:

an input interface for providing the one or more transport channels in a spectral representation having, in the time frame, the plurality of frequency bins; and

an audio renderer for rendering the one or more transport channels into a number of audio channels using the direction information,

wherein the audio renderer is configured to calculate a direct response information from the one or more audio objects per each frequency bin of the plurality of frequency bins and the direction information associated with the relevant one or more audio objects in the frequency bins.

19. Decoder of claim 18,

wherein the audio renderer is configured to calculate a covariance synthesis information using the direct response information and an information on the number of audio channels, and to apply the covariance synthesis information to the one or more transport channels to obtain the number of audio channels, or

wherein the direct response information is a direct response vector for each one or more audio object, and wherein the covariance synthesis information is a covariance synthesis matrix, and wherein the audio renderer is configured to perform a matrix operation per frequency bin in applying the covariance synthesis information.

20. Decoder of claim 18 or 19, wherein the audio renderer is configured

to derive, in the calculation of the direct response information, a direct response vector for the one or more audio objects and to calculate, for the one or more audio objects, a covariance matrix from each direct response vector,

to derive, in the calculation of the covariance synthesis information, a target covariance information from

the covariance matrix of the one audio object or the covariant matrices from more audio objects,

a power information on the respective one or more audio objects, and

a power information derived from the one or more transport channels.

21. Decoder of claim 20, wherein the audio renderer is configured

to derive, in the calculation of the direct response information, a direct response vector for the one or more audio object and to calculate, for each one or more audio objects, a covariance matrix from each direct response vector,

to derive an input covariance information from the transport channels, and

to derive a mixing information from the target covariance information, the input covariance information and the information on the number of channels, and

to apply the mixing information to the transport channels for each frequency bin in the time frame.

22. Decoder of claim 21, wherein a result of the application of the mixing information for each frequency bin in the time frame is converted into a time domain to obtain the number of audio channels in the time domain.

23. Decoder of one of claims 18 to 22, wherein the audio renderer is configured

to only use main diagonal elements of an input covariance matrix derived from the transport channels in a decomposition of the input covariance matrix, or

to perform a decomposition of a target covariance matrix using a direct response matrix and a matrix of powers of the objects or transport channels, or

to perform a decomposition of the input covariance matrix by taking the root of each main diagonal element of the input covariance matrix, or

to calculate a regularized inverse of decomposed input covariance matrix, or

to perform a singular value decomposition in calculating an optimum matrix to be used in an energy compensation without an extended identity matrix.

24. Decoder of one of claims 18 to 23, wherein the parameter data for the one or more audio objects comprise parameter data for at least two relevant audio objects, wherein a number of the at least two relevant audio objects is lower than a total number of the plurality of audio objects, and

wherein the audio renderer is configured to calculate, for each one of the one or more frequency bins, a contribution from the one or more transport channels in accordance with a first direction information associated with a first one of the at least two relevant audio objects and in accordance with a second direction information associated with a second one of the at least two relevant audio objects.

25. Decoder of claim 24,

wherein the audio renderer is configured to ignore, for the one or more frequency bins, a direction information of an audio object different from the at least two relevant audio objects.

26. Decoder of claim 24 or 25,

wherein the encoded audio signal comprises an amplitude related measure for each relevant audio object or a combined value related to at least two relevant audio objects in the parameter data, and

wherein the audio renderer is configured to operate so that a contribution from the one or more transport channels in accordance with a first direction information associated with a first one of the at least two relevant audio objects and in accordance with a second direction information associated with a second one of the at least two relevant audio objects is accounted for, or to determine a quantitative contribution of the one or more transport channels in accordance with the amplitude-related measure or the combined value.

27. Decoder of claim 26, wherein the encoded signal comprises the combined value in the parameter data, and

wherein the audio renderer is configured to determine the contribution of the one or more transport channels using the combined value for one of the relevant audio objects and the direction information for the one relevant audio object, and

wherein the audio renderer is configured to determine the contribution for the one or more transport channels using a value derived from the combined value for another of the relevant audio objects in the one or more frequency bins and the direction information of the other relevant audio object.

28. Decoder of one of claims 24 to 27, wherein the audio renderer is configured

to calculate the direct response information from the relevant audio objects per each frequency bin of the plurality of frequency bins and the direction information associated with the relevant audio objects in the frequency bins,

29. Decoder of claim 28,

wherein the audio renderer is configured to determine a diffuse signal per each frequency bin of the plurality of frequency bins using a diffuseness information such as a diffuseness parameter included in the metadata or a decorrelation rule and to combine a direct response as determined by the direct response information and the diffuse signal to obtain a spectral domain rendered signal for a channel of the number of channels.

30. Method of encoding a plurality of audio objects and related metadata indicating direction information on the plurality of audio objects, comprising:

downmixing the plurality of audio objects to obtain one or more transport channels;

encoding the one or more transport channels to obtain one or more encoded transport channels; and

outputting an encoded audio signal comprising the one or more encoded transport channels,

wherein the downmixing comprises downmixing the plurality of audio objects in response to the direction information on the plurality of audio objects.

31. Method of decoding an encoded audio signal comprising one or more transport channels and direction information for a plurality of audio objects, and, for one or more frequency bins of a time frame, parameter data for an audio object, the method comprising:

providing the one or more transport channels in a spectral representation having, in the time frame, the plurality of frequency bins; and

audio rendering the one or more transport channels into a number of audio channels using the direction information,

wherein the audio rendering comprises calculating a direct response information from the one or more audio objects per each frequency bin of the plurality of frequency bins and the direction information associated with the relevant one or more audio objects in the frequency bins.

32. Computer program for performing, when running on a computer or a processor, the method of claim 30 or the method of claim 31.