Claims

1. Apparatus for generating a description of a combined audio scene, comprising:

an input interface for receiving a first description of a first scene in a first format and a second description of a second scene in a second format, wherein the second format is different from the first format;

a format converter for converting the first description into a common format and for converting the second description into the common format, when the second format is different from the common format; and

a format combiner for combining the first description in the common format and the second description in the common format to obtain the combined audio scene.

2. Apparatus of claim 1,

wherein the first format and the second format are selected from a group of formats comprising a first order Ambisonics format, a high order Ambisonics format, the common format, a DirAC format, an audio object format and a multi-channel format.

3. Apparatus of claim 1 or 2,

wherein the format converter is configured to convert the first description into a first B-format signal representation and to convert the second description into a second B-format signal representation, and

wherein the format combiner is configured to combine the first and the second B-format signal representation by individually combining the individual components of the first and the second B-format signal representation.

4. Apparatus of one of the preceding claims,

wherein the format converter is configured to convert the first description into a first pressure/velocity signal representation and to convert the second description into a second pressure/velocity signal representation, and

wherein the format combiner is configured to combine the first and the second pressure/velocity signal representation by individually combining the individual components of the pressure/velocity signal representations to obtain a combined pressure/velocity signal representation.

5. Apparatus of one of the preceding claims,

wherein the format converter is configured to convert the first description into a first DirAC parameter representation and to convert the second description into a second DirAC parameter representation, when the second description is different from the DirAC parameter representation, and

wherein the format combiner is configured to combine the first and the second DirAC parameter representations by individually combining the individual components of the first and second DirAC parameter representations to obtain a combined DirAC parameter representation for the combined audio scene.

6. Apparatus of claim 5,

wherein the format combiner is configured to generate direction of arrival values for time-frequency tiles or direction of arrival values and diffuseness values for the time-frequency tiles representing the combined audio scene.

7. Apparatus of one of the preceding claims,

further comprising a DirAC analyzer for analyzing the combined audio scene to derive DirAC parameters for the combined audio scene,

wherein the DirAC parameters comprise direction of arrival values for time-frequency tiles or direction of arrival values and diffuseness values for the time-frequency tiles representing the combined audio scene.

8. Apparatus of one of the preceding claims,

further comprising a transport channel generator for generating a transport channel signal from the combined audio scene or from the first scene and the second scene, and

a transport channel encoder for core encoding the transport channel signal, or

wherein the transport channel generator is configured to generate a stereo signal from the first scene or the second scene being in a first order Ambisonics or a higher order Ambisonics format using a beam former being directed to a left position or the right position, respectively, or

wherein the transport channel generator is configured to generate a stereo signal from the first scene or the second scene being in a multichannel representation by downmixing three or more channels of the multichannel representation, or

wherein the transport channel generator is configured to generate a stereo signal from the first scene or the second scene being in an audio object representation by panning each object using a position of the object or by downmixing objects into a stereo downmix using information indicating, which object is located in which stereo channel, or

wherein the transport channel generator is configured to add only the left channel of the stereo signal to the left downmix transport channel and to add only the right channel of the stereo signal to obtain a right transport channel, or

wherein the common format is the B-format, and wherein the transport channel generator is configured to process a combined B-format representation to derive the transport channel signal, wherein the processing comprises performing a beamforming operation or extracting a subset of components of the B-format signal such as the omnidirectional component as the mono transport channel, or

wherein the processing comprises beamforming using the omnidirectional signal and the Y component with opposite signs of the B-format to calculate left and right channels, or

wherein the processing comprises a beamforming operation using the components of the B-format and the given azimuth angle and the given elevation angle, or

wherein the transport channel generator is configured to prove the B-format signals of the combined audio scene to the transport channel encoder, wherein any

spatial metadata are not included in the combined audio scene output by the format combiner.

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

a metadata encoder

for encoding DirAC metadata described in the combined audio scene to obtain encoded DirAC metadata, or

for encoding DirAC metadata derived from the first scene to obtain first encoded DirAC metadata and for encoding DirAC metadata derived from the second scene to obtain second encoded DirAC metadata.

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

an output interface for generating an encoded output signal representing the combined audio scene, the output signal comprising encoded DirAC metadata and one or more encoded transport channels.

11. Apparatus of one of the preceding claims,

wherein the format converter is configured to convert a high order Ambisonics or a first order Ambisonics format into the B-format, wherein the high order Ambisonics format is truncated before being converted into the B-format, or

wherein the format converter is configured to project an object or a channel on spherical harmonics at a reference position to obtain projected signals, and wherein the format combiner is configured to combine the projection signals to obtain B-format coefficients, wherein the object or the channel is located in space at a specified position and has an optional individual distance from a reference position, or

wherein the format converter is configured to perform a DirAC analysis comprising a time-frequency analysis of B-format components and a determination of pressure and velocity vectors, and wherein the format combiner is configured to combine different pressure/velocity vectors and wherein the format combiner further com-

prises a DirAC analyzer for deriving DirAC metadata from the combined pressure/velocity data, or

wherein the format converter is configured to extract DirAC parameters from object metadata of an audio object format as the first or second format, wherein the pressure vector is the object waveform signal and the direction is derived from the object position in space or the diffuseness is directly given in the object metadata or is set to a default value such as 0 value, or

wherein the format converter is configured to convert DirAC parameters derived from the object data format into pressure/velocity data and the format combiner is configured to combine the pressure/velocity data with pressure/velocity data derived from a different description of one or more different audio objects, or

wherein the format converter is configured to directly derive DirAC parameters, and wherein the format combiner is configured to combine the DirAC parameters to obtain the combined audio scene.

12. Apparatus of one of the preceding claims, wherein the format converter comprises:

a DirAC analyzer for a first order Ambisonics or a high order Ambisonics input format or a multi-channel signal format;

a metadata converter for converting object metadata into DirAC metadata or for converting a multi-channel signal having a time-invariant position into the DirAC metadata; and

a metadata combiner for combining individual DirAC metadata streams or combining direction of arrival metadata from several streams by a weighted addition, the weighting of the weighted addition being done in accordance to energies of associated pressure signal energies, or for combining diffuseness metadata from several streams by a weighted addition, the weighting of the weighted addition being done in accordance with energies of associated pressure signal energies, or

wherein the metadata combiner is configured to calculate, for a time/frequency bin of the first description of the first scene, an energy value, and direction of arrival value, and to calculate, for the time/frequency bin of the second description of the second scene, an energy value and a direction of arrival value, and wherein the

format combiner is configured to multiply the first energy to the first direction of arrival value and to add a multiplication result of the second energy value and the second direction of arrival value to obtain the combined direction of arrival value or, alternatively, to select the direction of arrival value among the first direction of arrival value and the second direction of arrival value that is associated with the higher energy as the combined direction of arrival value.

13. Apparatus of one of the preceding claims,

further comprising an output interface for adding to the combined format, a separate object description for an audio object, the object description comprising at least one of a direction, a distance, a diffuseness or any other object attribute, wherein the object has a single direction throughout all frequency bands and is either static or moving slower than a velocity threshold.

14. Method for generating a description of a combined audio scene, comprising:

receiving a first description of a first scene in a first format and receiving a second description of a second scene in a second format, wherein the second format is different from the first format;

converting the first description into a common format and converting the second description into the common format, when the second format is different from the common format; and

combining the first description in the common format and the second description in the common format to obtain the combined audio scene.

15. Computer program for performing, when running on a computer or a processor, the method of claim 14.

16. Apparatus for performing a synthesis of a plurality of audio scenes, comprising:

an input interface for receiving a first DirAC description of a first scene and for receiving a second DirAC description of a second scene and one or more transport channels; and

a DirAC synthesizer for synthesizing the plurality of audio scenes in a spectral domain to obtain a spectral domain audio signal representing the plurality of audio scenes; and

a spectrum-time converter for converting the spectral domain audio signal into a time-domain.

17. Apparatus of claim 16, wherein the DirAC synthesizer comprises;

a scene combiner for combining the first DirAC description and the second DirAC description into a combined DirAC description; and

a DirAC renderer for rendering the combined DirAC description using one or more transport channels to obtain the spectral domain audio signal, or

wherein the scene combiner is configured to calculate, for a time/frequency bin of the first description of the first scene, an energy value, and direction of arrival value, and to calculate, for the time/frequency bin of the second description of the second scene, an energy value and a direction of arrival value, and wherein the scene combiner is configured to multiply the first energy to the first direction of arrival value and to add a multiplication result of the second energy value and the second direction of arrival value to obtain the combined direction of arrival value or, alternatively, to select the direction of arrival value among the first direction of arrival value and the second direction of arrival value that is associated with the higher energy as the combined direction of arrival value.

18. Apparatus of claim 16,

wherein the input interface is configured to receive, for a DirAC description, a separate transport channel and separate DirAC metadata,

wherein the DirAC synthesizer is configured to render each description using the transport channel and the metadata for the corresponding DirAC description to obtain a spectral domain audio signal for each description, and to combine the spec-

tral domain audio signal for each description to obtain the spectral domain audio signal.

19. Apparatus of one claims 16 to 18, wherein the input interface is configured to receive extra audio object metadata for an audio object, and

wherein the DirAC synthesizer is configured to selectively manipulate the extra audio object metadata or object data related to the metadata to perform a directional filtering based on object data included in the object metadata or based on user-given direction information, or

wherein the DirAC synthesizer is configured for performing, in the spectral domain a zero-phase gain function, the zero-phase gain function depending upon a direction of an audio object, wherein the direction is contained in a bitstream if directions of objects are transmitted as side information, or wherein the direction is received from a user interface.

20. Method for performing a synthesis of a plurality of audio scenes, comprising:

receiving a first DirAC description of a first scene and receiving a second DirAC description of a second scene and one or more transport channels; and

synthesizing the plurality of audio scenes in a spectral domain to obtain a spectral domain audio signal representing the plurality of audio scenes; and

spectral-time converting the spectral domain audio signal into a time-domain.

21. Computer program for performing, when running on a computer or a processor, the method of claim 20.

22. Audio data converter, comprising:

an input interface for receiving an object description of an audio object having audio object metadata;

a metadata converter for converting the audio object metadata into DirAC metadata; and

an output interface for transmitting or storing the DirAC metadata.

23. Audio data converter of claim 22, in which the audio object metadata has an object position, and wherein the DirAC metadata has a direction of arrival with respect to a reference position.

24. Audio data converter of one of claims 22 or 23,

wherein the metadata converter is configured to convert DirAC parameters derived from the object data format into pressure/velocity data and wherein the metadata converter is configured to apply a DirAC analysis to the pressure/velocity data.

25. Audio data converter in accordance with one of claims 22 to 24,

wherein the input interface is configured to receive a plurality of audio object descriptions,

wherein the metadata converter is configured to convert each object metadata description into an individual DirAC data description, and

wherein the metadata converter is configured to combine the individual DirAC metadata descriptions to obtain a combined DirAC description as the DirAC metadata.

26. Audio data converter in accordance with claim 25, wherein the metadata converter is configured to combine the individual DirAC metadata descriptions, each metadata description comprising direction of arrival metadata or direction of arrival metadata and diffuseness metadata, by individually combining the direction of arrival metadata from different metadata descriptions by a weighted addition, wherein the weighting of the weighted addition is being done in accordance with energies of associated pressure signal energies, or by combining diffuseness metadata from the different DirAC metadata descriptions by a weighted addition, the weighting of the weighted addition being done in accordance with energies of associated pressure signal energies, or, alternatively, to select the direction of arrival value among the first direction of arrival value and the second direction of arrival value that is associated with the higher energy as the combined direction of arrival value.

27. Audio data converter is accordance with one of claims 22 to 26,

wherein the input interface is configured to receive, for each audio object, an audio object wave form signal in addition to this object metadata,

wherein the audio data converter further comprises a downmixer for downmixing the audio object wave form signals into one or more transport channels, and

wherein the output interface is configured to transmit or store the one or more transport channels in association with the DirAC metadata.

28. Method for performing an audio data conversion, comprising:

receiving an object description of an audio object having audio object metadata;

converting the audio object metadata into DirAC metadata; and

transmitting or storing the DirAC metadata.

29. Computer program for performing, when running on a computer or a processor, the method of claim 28.

30. Audio scene encoder, comprising:

an input interface for receiving a DirAC description of an audio scene having DirAC metadata and for receiving an object signal having object metadata;

a metadata generator for generating a combined metadata description comprising the DirAC metadata and the object metadata, wherein the DirAC metadata comprises a direction of arrival for individual time-frequency tiles and the object metadata comprises a direction or additionally a distance or a diffuseness of an individual object.

31. Audio scene encoder of claim 30, wherein the input interface is configured for receiving a transport signal associated with the DirAC description of the audio scene

and wherein the input interface is configured for receiving an object wave form signal associated with the object signal, and

wherein the audio scene encoder further comprises a transport signal encoder for encoding the transport signal and the object wave form signal.

32. Audio scene encoder of one of claims 30 and 31,

wherein the metadata generator comprises a metadata converter as described in any of the claims 12 to 27.

33. An audio scene encoder of one of claims 30 to 32,

wherein the metadata generator is configured to generate, for the object metadata, a single broadband direction per time and wherein the metadata generator is configured to refresh the single broadband direction per time less frequently than the DirAC metadata.

34. Method of encoding an audio scene, comprising:

receiving a DirAC description of an audio scene having DirAC metadata and receiving an object signal having audio object metadata; and

generating a combined metadata description comprising the DirAC metadata and the object metadata, wherein the DirAC metadata comprises a direction of arrival for individual time-frequency tiles and wherein the object metadata comprises a direction or, additionally, a distance or a diffuseness of an individual object.

35. Computer program for performing, when running on a computer or a processor, the method of claim 34.

36. Apparatus for performing a synthesis of audio data, comprising:

an input interface for receiving a DirAC description of one or more audio objects or a multi-channel signal or a first order Ambisonics signal or a high order Ambisonics signal, wherein the DirAC description comprises position information of the one or more objects or side information for the first order Ambisonics signal or the high

order Ambisonics signal or a position information for the multi-channel signal as side information or from a user interface;

a manipulator for manipulating the DirAC description of the one or more audio objects, the multi-channel signal, the first order Ambisonics signal or the high order Ambisonics signal to obtain a manipulated DirAC description; and

a DirAC synthesizer for synthesizing the manipulated DirAC description to obtain synthesized audio data.

37. Apparatus of claim 36,

wherein the DirAC synthesizer comprises a DirAC renderer for performing a DirAC rendering using the manipulated DirAC description to obtain a spectral domain audio signal; and

a spectral-time converter to convert the spectral domain audio signal into a time-domain.

38. Apparatus of claim 36 or 37,

wherein the manipulator is configured to perform a position-dependent weighting operation prior to DirAC rendering.

39. Apparatus of one of claims 36 to 38,

wherein the DirAC synthesizer is configured to output a plurality of objects or a first order Ambisonics signal or a high order Ambisonics signal or a multi-channel signal, and wherein the DirAC synthesizer is configured to use a separate spectral-time converter for each object or each component of the first order Ambisonics signal or the high order Ambisonics signal or for each channel of the multi-channel signal.

40. Method for performing a synthesis of audio data, comprising:

receiving a DirAC description of one or more audio objects or a multi-channel signal or a first order Ambisonics signal or a high order Ambisonics signal, wherein the DirAC description comprising position information of the one or more objects or

of the multi-channel signal or additional information for the first order Ambisonics signal or the high order Ambisonics signal as side information or for a user interface;

manipulating the DirAC description to obtain a manipulated DirAC description; and

synthesizing the manipulated DirAC description to obtain synthesized audio data.

41. Computer program for performing, when running on a computer or a processor, the method of claim 40.