Claims

1. Apparatus for encoding a spatial audio representation representing an audio scene to obtain an encoded audio signal, the apparatus comprising:

a transport representation generator (600) for generating a transport representation from the spatial audio representation, and for generating transport metadata related to the generation of the transport representation; and

an output interface (640) for generating the encoded audio signal, the encoded audio signal comprising information on the transport representation, and information on the transport metadata,

wherein the spatial audio representation is a first order Ambisonics or a higher order Ambisonics representation comprising a multitude of coefficient signals, or a multi-channel representation comprising a plurality of audio channels,

wherein the transport representation generator (600) is configured to combine two more coefficient signals from the higher order Ambisonics representation or the first order Ambisonics representation, or to combine two or more audio channels from the multichannel representation, and

wherein the transport representation generator (600) is configured to generate, as the transport metadata, information on how the two or more coefficient signals from the higher order Ambisonics representation or the first order Ambisonics representation or the two or more audio channels from the multichannel representation have been combined, or which ones of the two or more coefficient signals from the first order Ambisonics representation or the higher order Ambisonics representation or which ones of the two or more audio channels from the multichannel representation have been combined.

2. Apparatus of claim 1, further comprising a parameter processor (620) for deriving spatial parameters from the spatial audio representation,

wherein the encoded audio signal additionally comprises information on the spatial parameters.

3. Apparatus of claim 1 or 2,

wherein the transport representation generator (600) is configured to determine whether a dominant sound energy originates from a specific sector or hemisphere such as a left or right hemisphere or a forward or backward hemisphere,

wherein the transport representation generator (600) is configured to generate a first transport signal from the specific sector or hemisphere, where the dominant sound energy originates, and a second transport signal from a different sector or hemisphere such as the sector or hemisphere having an opposite direction with respect to a reference location and with respect to the specific sector or hemisphere, and

wherein the transport representation generator (600) is configured to determine the transport metadata so that the transport metadata comprises information identifying the specific sector or hemisphere, or identifying the different sector or hemisphere.

4. Apparatus of one of the preceding claims,

wherein the transport representation generator (600) is configured to combine the coefficient signals of the spatial audio representation so that a first resulting signal being a first transport signal corresponds to a directional microphone signal directed to a specific sector or hemisphere, and a second resulting signal being a second transport signal corresponds to a directional microphone signal directed to a different sector or hemisphere.

5. Apparatus of one of the preceding claims,

wherein the transport representation generator (600) is configured to generate the transport representation and the transport metadata in a time-variant or frequency-dependent way, so that the transport representation and the transport metadata for a first frame is different from the transport representation and the transport metadata for a second frame, or so that the transport representation and the transport metadata for a first frequency band is different from a transport representation and the transport metadata for a second different frequency band.

6. Apparatus of one of the preceding claims,

wherein the transport representation generator (600) is configured to generate one or two transport signals by a weighted combination (602) of two or more than two coefficient signals of the spatial audio representation, and

wherein the transport representation generator (600) is configured to calculate the transport metadata so that the transport metadata comprises information on weights used in the weighted combination, or information on an azimuth and/or elevation angle as a look direction of a generated directional microphone signal, or information on a shape parameter indicating a directional characteristic of a directional microphone signal.

7. Apparatus of one of the preceding claims,

wherein the transport representation generator (600) is configured to generate quantitative transport metadata, to quantize (605a) the quantitative transport metadata to obtain quantized transport metadata, and to entropy encode (605b) the quantized transport metadata, and wherein the output interface (640) is configured to include the encoded transport metadata into the encoded audio signal.

8. Apparatus of one of claims 1 to 6,

wherein the transport representation generator (600) is configured to transform the transport metadata into a table index or a preset parameter, and

wherein the output interface (640) is configured to include the table index or the preset parameter into the encoded audio signal.

9. Apparatus of one of the preceding claims,

wherein the spatial audio representation comprises at least two audio signals and spatial parameters,

wherein a parameter processor (620) is configured to derive the spatial parameters from the spatial audio representation by extracting the spatial parameters from the spatial audio representation,

wherein the output interface (640) is configured to include information on the spatial parameters into the encoded audio signal or to include information on processed

spatial parameters derived from the spatial parameters into the encoded audio signal, or

wherein the transport representation generator (600) is configured to combine the at least two audio signals or a subset of the at least two audio signals and to calculate the transport metadata such that the transport metadata includes information on the combination of the audio signals performed for calculating the transport representation of the spatial audio representation.

10. Apparatus of one of the preceding claims,

wherein the transport representation generator (600) is configured to combine (602) audio signals included in the spatial audio representation using spatial filtering or beamforming, and

wherein the transport representation generator (600) is configured to include information on a look direction of the transport representation or information on beamforming weights used in calculating the transport representation into the transport metadata.

11. Apparatus of one of the preceding claims,

wherein the spatial audio representation is a description of a sound field related to a reference position, and

wherein a parameter processor (620) is configured to derive spatial parameters from the spatial audio representation, wherein the spatial parameters define time-variant or frequency-dependent parameters on a direction of arrival of sound at the reference position or time-variant or frequency-dependent parameters on a diffuseness of the sound field at the reference position, or

wherein the transport representation generator (600) comprises a downmixer (601) for generating, as the transport representation, a downmix representation having a second number of individual signals being smaller than a first number of individual signals included in the spatial audio representation, wherein the downmixer (601) is

configured to combine the individual signals included in the spatial audio representation in order to decrease the first number of signals to the second number of signals.

12. Apparatus of one of the preceding claims,

wherein a parameter processor (620) comprises a spatial audio analyzer (621) for deriving the spatial parameters from the spatial audio representation by performing an audio signal analysis, and

wherein the transport representation generator (600) is configured to generate the transport representation based on the result of the spatial audio analyzer (621), or

wherein the transport representation comprises a core encoder (603) for core encoding one or more audio signals of the transport signals of the transport representation, or

wherein the parameter processor (620) is configured to quantize and entropy encode (622) the spatial parameters, and

wherein the output interface (640) is configured to include a core-encoded transport representation (611) as the information on the transport representation into the encoded audio signal or to include the entropy encoded spatial parameters (612) as the information on spatial parameters into the encoded audio signal.

13. Apparatus for decoding an encoded audio signal, comprising:

an input interface (700) for receiving the encoded audio signal comprising information on a transport representation and information on transport metadata; and

a spatial audio synthesizer (750) for synthesizing a spatial audio representation using the information on the transport representation comprising transport signals and the information on the transport metadata,

wherein the transport metadata indicates a first transport signal as referring to a first sector or hemisphere related to a reference position of the spatial audio representa-

tion and a second transport signal as referring to a second different sector or hemisphere related to the reference position of the spatial audio representation, wherein the spatial audio synthesizer (750) is configured

to generate (915) a component signal of the spatial audio representation associated with the first sector or hemisphere using the first transport signal and without using the second transport signal, or

to generate (915) another component signal of the spatial audio representation associated with the second sector or hemisphere using the second transport signal and not using the first transport signal, or

to generate (915) another component signal of the spatial audio representation associated with the first sector or hemisphere using a first combination of the first and the second transport signal, or

to generate (915) another component signal of the spatial audio representation associated with a different second sector or hemisphere using a second combination of the first and the second transport signals, wherein the first combination is influenced by the first transport signal stronger than the second combination, or wherein the second combination is influenced by the second transport signal stronger than the first combination,

or

wherein the transport metadata indicate information on weighting coefficients or look directions of the transport signals of the transport representation, and wherein the spatial audio synthesizer (750) is configured to calculate (932) different first order Ambisonics components of the spatial audio representation using the information on the look directions or the weighting coefficients, and using the transport signals.

14. Apparatus of claim 13, wherein the input interface (700) is configured for receiving the encoded audio signal additionally comprising information on spatial parameters, and

wherein the spatial audio synthesizer (750) is configured for synthesizing the spatial audio representation additionally using the information on the spatial parameters.

15. Apparatus of claim 13 or 14, wherein the spatial audio synthesizer (750) comprises:

a core decoder (751) for core decoding two or more encoded transport signals representing the information on the transport representation to obtain two or more decoded transport signals, or

wherein the spatial audio synthesizer (750) is configured to calculate a first order Ambisonics or a higher order Ambisonics representation (754) or a multi-channel signal (755) or an object representation (756) or a binaural representation of the spatial audio representation, or

wherein the spatial audio synthesizer (750) comprises a metadata decoder (752) for decoding the information on the transport metadata to derive the decoded transport metadata (720) or for decoding information on spatial parameters (722) to obtain decoded spatial parameters.

16. Apparatus of claim 13, 14, or 15,

wherein the spatial audio representation comprises a plurality of component signals,

wherein the spatial audio synthesizer (750) is configured to determine (760), for a component signal of the spatial audio representation, a reference signal using the information on the transport representation (711) and the information on the transport metadata (710), and

to calculate (770) the component signal of the spatial audio representation using the reference signal and information on spatial parameters, or to calculate (770) the component signal of the spatial audio representation using the reference signal.

17. Apparatus of one of claims 13 to 16,

wherein the transport metadata comprises information on a directional characteristic associated with the transport signals of the transport representation,

wherein the spatial audio synthesizer (750) is configured to calculate (911) virtual microphone signals using the first order Ambisonics or the higher order Ambisonics signals, loudspeaker positions and the transport metadata, or

wherein the spatial audio synthesizer (750) is configured to determine (911) the directional characteristics of the transport signals using the transport metadata and to determine the first order Ambisonics or a higher order Ambisonics component (754) from the transport signals in line with the determined directional characteristics of the transport signals, and

to determine (911) another first order Ambisonics or another higher order Ambisonics component (754) not associated with the directional characteristics of the transport signals in accordance with a fallback process.

18. Apparatus of one of claims 13 to 17,

wherein the transport metadata comprises an information on a first look direction associated with the first transport signal, and an information on a second look direction associated with the second transport signal,

wherein the spatial audio synthesizer (750) is configured to select (771) a reference signal for the calculation of the component signal or the other component signal of the spatial audio representation based on the transport metadata and the position of a loudspeaker associated with the component signal or the other component signal of the spatial audio representation.

19. Apparatus of claim 18,

wherein the first look direction indicates a left or a front hemisphere, wherein the second look direction indicates a right or a back hemisphere,

wherein, for the calculation of the component signal or the other component signal for a loudspeaker in the left hemisphere, the first transport signal and not the second transport signal is used (771), or wherein for the calculation of the component signal or the other component signal for a loudspeaker signal in the right hemisphere, the second transport signal and not the first transport signal is used (771), or

wherein for the calculation of the component signal or the other component signal for a loudspeaker in a front hemisphere, the first transport signal and not the second

transport signal is used (771), or wherein for the calculation of a loudspeaker in a back hemisphere, the second transport signal and not the first transport signal is used (771), or

wherein for the calculation of the component signal or the other component signal for a loudspeaker in a center region, a combination of the left transport signal and the second transport signal is used (771), or wherein for the calculation of the component signal or the other component signal for a loudspeaker signal associated with a loudspeaker in a region between the front hemisphere and the back hemisphere, a combination of the first transport signal and the second transport signal is used (771).

20. Apparatus of one of claims 13 to 19,

wherein the information on the transport metadata indicates, as a first look direction, a left direction for the first transport signal and indicates, as a second look direction, a right look direction for the second transport signal,

wherein the spatial audio synthesizer (750) is configured to calculate a first Ambisonics component by adding (920) the first transport signal and the second transport signal, or to calculate a second Ambisonics component by subtracting (921) the first transport signal and the second transport signal, or wherein another Ambisonics component is calculated (922) using a sum of the first transport signal and the second transport signal.

21. Apparatus of one of claims 13 to 17,

wherein the transport metadata indicates, for the first transport signal, a front look direction and indicates, for the second transport signal, a back look direction,

wherein the spatial audio synthesizer (750) is configured to calculate a first order Ambisonics component for an x direction by performing the calculation of a difference (921) between the first transport signal and the second transport signal, and to calculate an omnidirectional first order Ambisonics component using an addition (920) of the first transport signal and the second transport signal, and

to calculate (922) another first order Ambisonics component using a sum of the first transport signal and the second transport signal.

22. Apparatus of claim 13,

wherein the spatial audio synthesizer (750) is configured to calculate (932) the different first order Ambisonics components of the spatial audio representation additionally using spatial parameters.

23. Apparatus of one of claims 13 to 22,

wherein the transport metadata include information on the transport signals being derived from microphone signals at two different positions or with different look directions,

wherein the spatial audio synthesizer (750) is configured to select (931) a reference signal that has a position that is closest to a loudspeaker position, or to select (932) a reference signal that has a closest look direction with respect to the direction from a reference position of the spatial audio representation and a loudspeaker position, or to perform a linear combination (771) with the transport signals to determine a reference signal for a loudspeaker being placed between two look directions indicated by the transport metadata.

24. Apparatus of one of claims 13 to 23,

wherein the transport metadata includes information on a distance between microphone positions associated with the transport signals,

wherein the spatial audio synthesizer (750) comprises a diffuse signal generator (830, 823, 824), and wherein the diffuse signal generator (830, 823, 824) is configured to control an amount of a decorrelated signal in a diffuse signal generated by the diffuse signal generator using the information on the distance, so that, for a first distance, a higher amount of decorrelated signal is included in the diffuse signal compared to an amount of decorrelated signal for a second distance, wherein the first distance is lower than the second distance, or

wherein the spatial audio synthesizer (750) is configured to calculate, for a first distance between the microphone positions, a component signal for the spatial audio representation using an output signal of a decorrelation filter (823) configured for

decorrelating a reference signal or a scaled reference signal and the reference signal weighted (822) using a gain derived from a sound direction of arrival information and to calculate, for a second distance between the microphone positions, a component signal for the spatial audio representation using the reference signal weighted (822) using a gain derived from a sound direction of arrival information without any decorrelation processing, the second distance being greater than the first distance or being greater than a distance threshold.

25. Apparatus of one of claims 13 to 24,

wherein the transport metadata comprises information on a beamforming or a spatial filtering associated with the transport signals of the transport representation, and

wherein the spatial audio synthesizer (750) is configured to generate (932) a loudspeaker signal for a loudspeaker using the transport signal having a look direction being closest to a look direction from a reference position of the spatial audio representation to the loudspeaker.

26. Apparatus of one of claims13 to 25,

wherein the spatial audio synthesizer (750) is configured to determine component signals of the spatial audio representation as a combination (825) of a direct sound component and a diffuse sound component, wherein the direct sound component is obtained by scaling (822) a reference signal with a factor depending on a diffuseness parameter or a directional parameter, wherein the directional parameter depends on a direction of arrival of sound, wherein the determination of the reference signal is performed (821, 760) based on the information on the transport metadata, and wherein the diffuse sound component is determined (823, 824) using the same reference signal and the diffuseness parameter.

27. Apparatus of one of claims 13 to 26,

wherein the spatial audio synthesizer (750) is configured to determine component signals of the spatial audio representation as a combination (825) of a direct sound component and a diffuse sound component, wherein the direct sound component is obtained by scaling (822) a reference signal with a factor depending on a diffuseness parameter or a directional parameter, wherein the directional parameter depends on a direction of arrival of sound, wherein the determination of the reference signal is

performed (821, 760) based on the information on the transport metadata, and wherein the diffuse sound component is determined (823, 824) using a decorrelation filter (823), the same reference signal, and the diffuseness parameter.

28. Apparatus of one of claims 13 to 27, wherein the transport representation comprises at least two different microphone signals,

wherein the transport metadata comprises information indicating, whether the at least two different microphone signals are at least one of omnidirectional signals, dipole signals or cardioid signals, and

wherein the spatial audio synthesizer is configured for adapting (915) a reference signal determination to the transport metadata to determine, for components of the spatial audio representation, individual reference signals and for calculating (916) the respective component using the individual reference signal determined for the respective component.

29. Method for encoding a spatial audio representation representing an audio scene to obtain an encoded audio signal, the method comprising:

generating a transport representation from the spatial audio representation;

generating transport metadata related to the generation of the transport representation; and

generating the encoded audio signal, the encoded audio signal comprising information on the transport representation, and information on the transport metadata

wherein the spatial audio representation is a first order Ambisonics or a higher order Ambisonics representation comprising a multitude of coefficient signals, or a multichannel representation comprising a plurality of audio channels,

wherein the generating the transport comprises combining two more coefficient signals from the higher order Ambisonics representation or the first order Ambisonics representation, or combining two or more audio channels from the multichannel representation, and

wherein the generating the transport metadata comprises generating, as the transport metadata, information on how the two or more coefficient signals from the higher order Ambisonics representation or the first order Ambisonics representation or the two or more audio channels from the multichannel representation have been combined, or which ones of the two or more coefficient signals from the first order Ambisonics representation or the higher order Ambisonics representation or which ones of the two or more audio channels from the multichannel representation have been combined.

30. Method for decoding an encoded audio signal, the method comprising:

receiving the encoded audio signal comprising information on a transport representation comprising transport signals and information on transport metadata; and

synthesizing a spatial audio representation using the information on the transport representation and the information on the transport metadata,

wherein the transport metadata indicates a first transport signal as referring to a first sector or hemisphere related to a reference position of the spatial audio representation and a second transport signal as referring to a second different sector or hemisphere related to the reference position of the spatial audio representation, wherein the synthesizing comprises

generating (915) a component signal of the spatial audio representation associated with the first sector or hemisphere using the first transport signal and without using the second transport signal, or

generating (915) another component signal of the spatial audio representation associated with the second sector or hemisphere using the second transport signal and not using the first transport signal, or

generating (915) another component signal of the spatial audio representation associated with the first sector or hemisphere using a first combination of the first and the second transport signal, or

generating (915) another component signal of the spatial audio representation associated with a different second sector or hemisphere using a second

combination of the first and the second transport signals, wherein the first combination is influenced by the first transport signal stronger than the second combination, or wherein the second combination is influenced by the second transport signal stronger than the first combination,

or

wherein the transport metadata indicate information on weighting coefficients or look directions of the transport signals of the transport representation, and wherein the synthesizing comprises calculating (932) different first order Ambisonics components of the spatial audio representation using the information on the look directions or the weighting coefficients, and using the transport signals.

31. Computer program comprising instructions which, when the computer program is executed by a computer or a processor, cause the computer or the processor to carry out the method of claim 29 or claim 30.