Claims

1. Apparatus for generating an encoded audio scene from an audio signal having a first frame and a second frame, comprising:

a soundfield parameter generator for determining a first soundfield parameter representation for the first frame from the audio signal in the first frame and a second soundfield parameter representation for the second frame from the audio signal in the second frame; and

an activity detector for analyzing the audio signal to determine, depending on the audio signal, that the first frame is an active frame and the second frame is an inactive frame,

wherein the soundfield parameter generator is configured to determine, from the second frame of the audio signal, individual sound source(s) and to determine, for each sound source, a parametric description for the second frame,

wherein the soundfield parameter generator is configured to decompose the second frame into frequency bin(s), each frequency bin representing an individual sound source of the individual sound source(s), and to determine, for each frequency bin, at least one inactive spatial parameter as the second soundfield parameter representation for the second frame, the at least one inactive spatial parameter comprising a direction parameter, a direction of arrival parameter, a diffuseness parameter, or an energy ratio parameter,

the apparatus further comprising:

an audio signal encoder for generating an encoded audio signal, the encoded audio signal providing an encoded audio signal for the first frame being the active frame and the parametric description for the second frame being the inactive frame; and

an encoded signal former for composing the encoded audio scene by bringing together the first soundfield parameter representation for the first frame, the second soundfield parameter representation for the second frame, the encoded audio signal for the first frame, and the parametric description for the second frame.

2. Apparatus of claim 1, wherein the soundfield parameter generator is configured to determine, from the second frame of the audio signal, a plurality of individual sound sources and to determine, for each sound source, the parametric description for the second frame, each frequency bin representing an individual sound source of the plurality of individual sound sources.

3. Apparatus of claim 1 or 2, wherein the soundfield parameter generator is configured to generate the second soundfield parameter representation so that the second soundfield parameter representation comprises a parameter indicating a characteristic of the audio signal with respect to a listener position.

4. Apparatus of claim 1 or 2 or 3, wherein the first soundfield parameter representation comprises one or more direction parameters indicating a direction of sound with respect to a listener position in the first frame, or one or more diffuseness parameters indicating a portion of a diffuse sound with respect to a direct sound in the first frame, or one or more energy ratio parameters indicating an energy ratio of a direct sound and a diffuse sound in the first frame, or an inter-channel/surround coherence parameter in the first frame.

5. Apparatus of one of the preceding claims, wherein the audio signal for the first frame and the second frame comprises an input format having a plurality of components representing a soundfield with respect to a listener,

wherein the soundfield parameter generator is configured to calculate one or more transport channels for the first frame and the second frame, using a downmix of the plurality of components, and to analyze the input format to determine the first parameter representation related to the one or more transport channels, or

wherein the soundfield parameter generator is configured to calculate one or more transport channels, using a downmix of the plurality of components, and

wherein the activity detector is configured to analyze the one or more transport channels derived from the audio signal in the second frame.

6. Apparatus of one of the preceding claims,

wherein the audio signal for the first frame or the second frame comprises an input format having, for each frame of the first and second frames, one or more transport channels and metadata associated with each frame,

wherein the soundfield parameter generator is configured to read the metadata from the first frame and the second frame and to use or process the metadata for the first frame as the first soundfield parameter representation and to process the metadata of the second frame to obtain the second soundfield parameter representation, wherein the processing to obtain the second soundfield parameter representation is such that an amount of information units required for the transmission of the metadata for the second frame is reduced with respect to an amount required before the processing.

7. Apparatus of claim 6,

wherein the soundfield parameter generator is configured to process the metadata for the second frame to reduce a number of information items in the metadata or to resample the information items in the metadata to a lower resolution, such as a time resolution or a frequency resolution, or to requantize the information units of the metadata for the second frame to a coarser representation with respect to a situation before requantization.

8. Apparatus of one of the preceding claims,

wherein the audio signal encoder is configured to determine a silence information description for the inactive frame as the parametric description,

wherein the silence information description comprises an amplitude-related information, such as an energy, a power or a loudness for the second frame, and a shaping information, such as a spectral shaping information, or an amplitude-related information for the second frame, such as an energy, a power, or a loudness, and linear prediction coding, LPC, parameters for the second frame, or scale parameters for the second frame with a varying associated frequency resolution so that different scale parameters refer to frequency bands with different widths.

9. Apparatus of one of the preceding claims,

wherein the audio signal encoder is configured to encode, for the first frame, the audio signal using a time domain or frequency domain encoding mode, the encoded audio signal comprising encoded time domain samples, encoded spectral domain samples, encoded LPC domain samples and side information obtained from components of the audio signal or obtained from one or more transport channels derived from the components of the audio signal by a downmixing operation.

10. Apparatus of one of the preceding claims,

wherein the audio signal comprises an input format being a first order Ambisonics format, a higher order Ambisonics format, a multi-channel format associated with a given loudspeaker setup, such as 5.1 or 7.1 or 7.1 + 4, or one or more audio channels representing one or several different audio objects localized in a space as indicated by information included in associated metadata, or an input format being a metadata associated spatial audio representation,

wherein the soundfield parameter generator is configured for determining the first soundfield parameter representation and the second soundfield representation so that the parameters represent a soundfield with respect to a defined listener position.

11. Apparatus of one of the preceding claims ,

wherein the audio signal comprises an input format being a first order Ambisonics format, a higher order Ambisonics format, a multi-channel format associated with a given loudspeaker setup, such as 5.1 or 7.1 or 7.1 + 4, or one or more audio channels representing one or several different audio objects localized in a space as indicated by information included in associated metadata, or an input format being a metadata associated spatial audio representation,

wherein the audio signal comprises a microphone signal as picked up by real microphone or a virtual microphone or a synthetically created microphone signal e.g. being in a first order Ambisonics format, or a higher order Ambisonics format.

12. Apparatus of one of the preceding claims,

wherein the activity detector is configured for detecting an inactivity phase over the second frame and one or more frames following the second frame, and

wherein the activity detector is configured for determining an inactive phase comprising the second frame and eight frames following the second frame, and wherein the audio signal encoder is configured for generating a parametric description for an inactive frame only at every eighth frame, and wherein the soundfield parameter generator is configured for generating a soundfield parameter representation for each eighth inactive frame.

13. Apparatus of one of claims 1-11,

wherein the activity detector is configured for detecting an inactivity phase over the second frame and one or more frames following the second frame, and

wherein the soundfield parameter generator is configured for generating a soundfield parameter representation for each inactive frame even when the audio signal encoder does not generate a parametric description for an inactive frame.

14. Apparatus of one of claims 1-11,

wherein the soundfield parameter generator is configured for determining a parameter representation with a higher frame rate than the audio signal encoder generates the parametric description for one or more inactive frames.

15. Apparatus of one of the preceding claims,

wherein the soundfield parameter generator is configured for determining the second soundfield parameter representation for the second frame

using spatial parameters for one or more directions in frequency bands and associated energy ratios in frequency bands corresponding to a ratio of one directional component over a total energy.

16. Apparatus of one of claims 1-14,

wherein the soundfield parameter generator is configured for determining the second soundfield parameter representation for the second frame

to determine a diffuseness parameter indicating a ratio of diffuse sound or direct sound.

17. Apparatus of one of claims 1-14,

wherein the soundfield parameter generator is configured for determining the second soundfield parameter representation for the second frame

to determine a direction information using a coarser quantization scheme compared to a quantization in the first frame).

18. Apparatus of one of claims 1-14,

wherein the soundfield parameter generator is configured for determining the second soundfield parameter representation for the second frame

using an averaging of a direction over time or frequency for obtaining a coarser time or frequency resolution.

19. Apparatus of one of claims 1-14,

wherein the soundfield parameter generator is configured for determining the second soundfield parameter representation for the second frame

to determine a soundfield parameter representation for one or more inactive frames with the same frequency resolution as in the first soundfield parameter representation for an active frame, and with a time occurrence that is lower than the time occurrence for active frames with respect to a direction information in the soundfield parameter representation for the inactive frame.

20. Apparatus of one of claims 1-14,

wherein the soundfield parameter generator is configured for determining the second soundfield parameter representation for the second frame

to determine the second soundfield parameter representation having a diffuseness parameter, where the diffuseness parameter is transmitted with the same time or frequency resolution as for active frames, but with a coarser quantization.

21. Apparatus of one of claims 1-14,

wherein the soundfield parameter generator is configured for determining the second soundfield parameter representation for the second frame

to quantize a diffuseness parameter for the second soundfield representation with a first number of bits, and wherein only a second number of bits of each quantization index is transmitted, the second number of bits being smaller than the first number of bits.

22. Apparatus of one of claims 1-14,

wherein the soundfield parameter generator is configured for determining the second soundfield parameter representation for the second frame

to determine, for the second soundfield parameter representation, an inter-channel coherence if the audio signal has input channels corresponding to channels positioned in a spatial domain or inter-channel level differences if the audio signal has input channels corresponding to channels positioned in the spatial domain.

23. Apparatus of one of claims 1-14,

wherein the soundfield parameter generator is configured for determining the second soundfield parameter representation for the second frame

to determine a surround coherence being defined as a ratio of diffuse energy being coherent in a soundfield represented by the audio signal.

24. Method of generating an encoded audio scene from an audio signal having a first frame and a second frame, comprising:

determining a first soundfield parameter representation for the first frame from the audio signal in the first frame and a second soundfield parameter representation for the second frame from the audio signal in the second frame; and

analyzing the audio signal to determine, depending on the audio signal, that the first frame is an active frame and the second frame is an inactive frame,

wherein determining the first soundfield parameter representation includes determining, from the second frame of the audio signal, for each sound source, a parametric description for the second frame,

wherein determining the first soundfield parameter representation includes decomposing the second frame into frequency bin(s), each frequency bin representing an individual sound source, and determining, for each frequency bin, at least one inactive spatial parameter as the second soundfield parameter representation for the second frame, the at least one inactive spatial parameter comprising a direction parameter, a direction of arrival parameter, a diffuseness parameter, or an energy ratio parameter,

the method further comprising:

generating an encoded audio signal, the encoded audio signal providing an encoded audio signal for the first frame being the active frame and the parametric description for the second frame being the inactive frame; and

composing the encoded audio scene by bringing together the first soundfield parameter representation for the first frame, the second soundfield parameter representation for the second frame, the encoded audio signal for the first frame, and the parametric description for the second frame.

25. Method of claim 24, from the second frame of the audio signal, a plurality of individual sound sources and determining, for each sound source, a parametric description for the second frame, wherein determining the first soundfield parameter representation includes decomposing the second frame into a plurality of frequency bins, each frequency bin representing an individual sound source.

26. Apparatus for processing an encoded audio scene comprising, in a first frame, a first soundfield parameter representation and an encoded audio signal, wherein a second frame is an inactive frame, the apparatus comprising:

an activity detector for detecting that the second frame is the inactive frame;

a synthetic signal synthesizer for synthesizing a synthetic audio signal for the second frame using a parametric description for the second frame;

an audio decoder for decoding the encoded audio signal for the first frame; and

a transcoder for generating a meta data assisted output format comprising the audio signal for the first frame, the first soundfield parameter representation for the first frame, the synthetic audio signal for the second frame, and a second soundfield parameter representation for the second frame.

27. Apparatus for processing an encoded audio scene comprising, in a first frame, a first soundfield parameter representation and an encoded audio signal, and in a second frame an inactive frame, the second frame being decomposed frequency bin(s), and, for each frequency bin, at least one inactive spatial parameter being determined as second soundfield parameter representation for the second frame, the at least one inactive spatial parameter comprising a direction parameter, a direction of arrival parameter, a diffuseness parameter, or an energy ratio parameter,

the apparatus comprising:

an activity detector for detecting that the second frame is the inactive frame;

a synthetic signal synthesizer for synthesizing a synthetic audio signal for the second frame using a parametric description for the second frame;

an audio decoder for decoding the encoded audio signal for the first frame; and

a spatial renderer for spatially rendering the audio signal for the first frame using the first soundfield parameter representation and using the synthetic audio signal and the second soundfield parameter representation for the second frame,

wherein the synthetic signal generator is configured to generate one or more transport channels for the second frame as the synthetic audio signal, and

wherein the spatial renderer is configured to spatially render the one or more transport channels for the second frame.

28. Apparatus of claim 27, wherein there are determined, for the second frame of the audio signal, individual sound source(s) and there are determined, for each sound source, the parametric description for the second frame, each frequency bin representing an individual sound source.

29. Apparatus of claim 27 or 28, wherein the encoded audio scene comprises, for the second frame, a second soundfield parameter description, and wherein the parameter processor for deriving one or more soundfield parameters from the second soundfield parameter representation, and wherein the spatial renderer is configured to use, for the rendering of the synthetic audio signal for the second frame, the one or more soundfield parameters for the second frame.

30. Apparatus of claim 27 or 28 or 29,

wherein the parameter processor is configured to store one or more soundfield parameter representationsfor several frames occurring in time before the second frame or occurring in time subsequent to the second frame to extrapolate or interpolate using the

at least two soundfield parameter representations of the one or more soundfield parameter representations for several frames to determine the one or more soundfield parameters for the second frame, and

wherein the spatial renderer is configured to use, for the rendering of the synthetic audio signal for the second frame, the one or more soundfield parameters for the second frame.

31. Apparatus of claim 30,

wherein the parameter processor is configured to perform a dithering with directions included in the at least two soundfield parameter representations occurring in time before or after the second frame, when extrapolating or interpolating to determine the one or more soundfield parameters for the second frame.

32. Apparatus of any of claims 27-31,

wherein the synthetic signal generator is configured to generate, for the second frame, a plurality of synthetic component audio signals for individual components related to an audio output format of the spatial renderer as the synthetic audio signal.

33. Apparatus of claim 32, wherein the synthetic signal generator is configured to generate, at least for each one of a subset of at least two individual components related to the audio output format, an individual synthetic component audio signal,

wherein a first individual synthetic component audio signal is decorrelated from a second individual synthetic component audio signal, and

wherein the spatial renderer is configured to render a component of the audio output format using a combination of the first individual synthetic component audio signal and the second individual synthetic component audio signal.

34. Apparatus of claim 33,

wherein the spatial renderer is configured to apply a covariance method.

35. Apparatus of claim 34,

wherein the spatial renderer is configured to not use any decorrelator processing or to control a decorrelator processing so that only an amount of decorrelated signals generated by the decorrelator processing as indicated by the covariance method is used in generating a component of the audio output format.

36. Apparatus of one of claims 27-35, wherein the synthetic signal generator is a comfort noise generator.

37. Apparatus of one of claims 33-36, wherein the synthetic signal generator comprises a noise generator and the first individual synthetic component audio signal is generated by a first sampling of the noise generator and the second individual synthetic component audio signal is generated by a second sampling of the noise generator, wherein the second sampling is different from the first sampling.

38. Apparatus of claim 37, wherein the noise generator comprises a noise table, and wherein the first individual synthetic component audio signal is generated by taking a first portion of the noise table, and wherein the second individual synthetic component audio signal is generated by taking a second portion of the noise table, wherein the second portion of the noise table is different from the first portion of the noise table.

39. Apparatus of claim 37, wherein the noise generator comprises a pseudo noise generator, and wherein the first individual synthetic component audio signal is generated by using a first seed for the pseudo noise generator, and wherein the second individual synthetic component audio signal is generated using a second seed for the pseudo noise generator.

40. Apparatus of one of claims 27-39

wherein the encoded audio scene comprises, for the first frame, two or more transport channels, and

wherein the synthetic signal generator comprises a noise generator and is configured to generate, using the parametric description for the second frame, a first transport channel by sampling the noise generator and a second transport channel by sampling the noise generator, wherein the first and the second transport channels as determined by sampling the noise generator are weighted using the same parametric description for the second frame.

41. Apparatus of one of claims 27-40, wherein the spatial renderer is configured to operate

in a first mode for the first frame using a mixing of a direct signal and a diffuse signal generated by a decorrelator from the direct signal under a control of the first soundfield parameter representation, and

in a second mode for the second frame using a mixing of a first synthetic component signal and the second synthetic component signal, wherein the first and the second

synthetic component signals are generated by the synthetic signal synthesizer by different realizations of a noise process or a pseudo noise process.

42. Apparatus of claim 41, wherein the spatial renderer is configured to control the mixing in the second mode by a diffuseness parameter, an energy distribution parameter, or a coherence parameter derived for the second frame by the parameter processor.

43. Apparatus of one of claims 27-42,

wherein the synthetic signal generator is configured to generate a synthetic audio signal for the first frame using the parametric description for the second frame, and

wherein the spatial renderer is configured to perform a weighted combination of the audio signal for the first frame and the synthetic audio signal for the first frame before or after the spatial rendering, wherein, in the weighted combination, an intensity of the synthetic audio signal for the first frame is reduced with respect to an intensity of the synthetic audio signal for the second frame.

44. Apparatus of one of claims 27-43,

wherein a parameter processor is configured to determine, for the second inactive frame, a surround coherence being defined as a ratio of diffuse energy being coherent in a soundfield represented by the second frame, wherein the spatial renderer is configured for re-distributing an energy between direct and diffuse signals in the second frame based on the sound coherence, wherein an energy of sound surround coherent components is removed from the diffuse energy to be re-distributed to directional components, and wherein the directional components are panned in a reproduction space.

45. Apparatus of one of claims 27-44, further comprising an output interface for converting an audio output format generated by the spatial renderer into a transcoded output format such as an output format comprising a number of output channels dedicated for loudspeakers to be placed at predefined positions, or a transcoded output format comprising FOA or HOA data.

46. Apparatus of one of claims 27-45, further comprising a parameter processor configured for deriving one or more second soundfield parameters for the second frame, wherein the parameter processor is configured to store the first soundfield parameter representation for the first frame and to synthesize one or more second soundfield parameters for the

second frame using the stored first soundfield parameter representation for the first frame, wherein the second frame) follows the first frame) in time.

47. Method of processing an encoded audio scene comprising, in a first frame, a first soundfield parameter representation and an encoded audio signal, and, in a second frame, an inactive frame, the encoded audio scene comprising one or more transport channels for the first frame, the second frame being decomposed into frequency bin(s), and, for each frequency bin, at least one inactive spatial parameter being determined as second soundfield parameter representation for the second frame, the method comprising:

detecting that the second frame is the inactive frame;

synthesizing a synthetic audio signal for the second frame using a parametric description for the second frame;

decoding the encoded audio signal for the first frame; and

spatially rendering the audio signal for the first frame using the first soundfield parameter representation and using the synthetic audio signal for the second frame) and the synthetic audio signal for the second frame,

the method further comprising generating one or more transport channels for the second frame as the synthetic audio signal, and spatially rendering the one or more transport channels for the second frame,

the method further comprising deriving one or more second soundfield parameters for the second frame, wherein the parameter processor is configured to store the first soundfield parameter representation for the first frame and to synthesize one or more second soundfield parameters for the second frame using the stored first soundfield parameter representation for the first frame, wherein the second frame follows the first frame in time.

48. The method of claim 47, further comprising providing a parametric description for the second frame.

49. Encoded audio scene comprising:

a first soundfield parameter representation for a first frame;

a second soundfield parameter representation for a second frame;

an encoded audio signal for the first frame; and

a parametric description for the second frame), decomposed into frequency bin(s),

wherein it is determined, for each frequency bin, at least one inactive spatial parameter as the second soundfield parameter representation for the second frame, the at least one inactive

spatial parameter comprising a direction parameter, a direction of arrival parameter, a diffuseness parameter, or an energy ratio parameter.

50. Computer program for performing, when running on a computer or processor, the method of claim 24 or claim 47.