Claims

1. Apparatus for processing an encoded audio scene representing a sound field related to a virtual listener position, the encoded audio scene comprising information on a transport signal and a first set of parameters related to the virtual listener position, the apparatus comprising:

a parameter converter for converting the first set of parameters into a second set of parameters related to a channel representation comprising two or more channels for a reproduction at predefined spatial positions for the two or more channels; and

an output interface for generating a processed audio scene using the second set of parameters and the information on the transport signal.

2. Apparatus of claim1,

wherein the output interface is configured for upmixing the transport signal using the second set of parameters into an upmix signal comprising the two or more channels.

3. Apparatus of claim 1, wherein the output interface is configured to generate the processed audio scene by combining the transport signal or the information on the transport signal and the second set of parameters to obtain a transcoded audio scene as the processed audio scene.

4. Apparatus of one of the preceding claims, wherein the first set of parameters comprises, for each input time frame of a plurality of input time frames and for each input frequency band of a plurality of input frequency bands, at least one DirAC parameter,

wherein the parameter converter is configured to calculate the second set of parameters as parametric stereo or multichannel parameters.

5. Apparatus of claim 4, wherein the at least one parameter comprises at least one of a direction of arrival parameter, a diffuseness parameter, a direction information parameter related to a sphere with the virtual listening position as an origin of the sphere, and a distance parameter, and

wherein the parametric stereo or multichannel parameters comprise at least one of a side gain parameter, a residual prediction gain parameter, an inter-channel level difference parameter, an inter-channel time difference parameter, an inter-channel phase difference parameter and an inter-channel coherence parameter.

6. Apparatus of one of the preceding claims, wherein an input time frame, to which the first set of parameters is related, comprises two or more input time subframes, and wherein an output time frame, to which the second set of parameters is related, is smaller than the input time frame and longer than an input time subframe of the two or more input time subframes, and

wherein the parameter converter is configured to calculate a raw parameter of the second set of parameters for each one of the two or more input time subframes being time-subsequent and to combine at least two raw parameters to derive a parameter of the second set of parameters related to the output subframe.

7. Apparatus of claim 6, wherein the parameter converter is configured to perform, in the combining of the at least two raw parameters, a weighted combination of the at least two raw parameters, wherein weighting factors for the weighted combination are derived based on an amplitude-related measure of the transport signal in the corresponding input time subframe.

8. Apparatus of claim 7, wherein the parameter converter is configured to use, as the amplitude-related measure, an energy or a power, and wherein a weighting factor for an input subframe is greater in case of a higher energy or power of the transport signal in the corresponding input time subframe compared to a weighting factor for an input subframe having a lower energy or power of the transport signal in the corresponding input time subframe.

9. Apparatus of one of the preceding claims, wherein the parameter converter is configured to calculate at least one raw parameter for each output time frame using at least one parameter of the first set of parameters for the input time frame,

wherein the parameter converter is configured to calculate a smoothing factor for each raw parameter in accordance with a smoothing rule, and

wherein the parameter converter is configured to apply a corresponding smoothing factor to the corresponding raw parameter to derive the parameter of the second set of parameters for the output time frame.

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

to calculate a long-term average over an amplitude-related measure of a first time portion of the transport signal, and

to calculate a short-term average over an amplitude-related measure of a second time portion of the transport signal, wherein the second time portion is shorter than the first time portion, and

to calculate a smoothing factor based on a ratio between the long-term average and the short-term average.

11. Apparatus of claim 9 or 10, wherein the parameter converter is configured to calculate a smoothing factor for a band using a compression function, the compression function being different for different frequency bands, and wherein a compression strength of the compression function is stronger for a lower frequency band than for a higher frequency band.

12. Apparatus of one of claims 9 to 11, wherein the parameter converter is configured to calculate the smoothing factor using different maximum bounds for different bands, wherein a maximum bound for a lower band is higher than a maximum bound for a higher band.

13. Apparatus of one of claims 9 to 12, wherein the parameter converter is configured to apply, as the smoothing rule, a recursive smoothing rule over time-subsequent output time frames, so that a smoothed parameter for a current output time frame is calculated by combining the parameter for a preceding output time frame weighted by a first weighting value and a raw parameter for the current output time frame weighted by a second weighting value, wherein the first weighting value and the second weighting value are derived from the smoothing factor for the current time frame.

14. Apparatus of one of the preceding claims, wherein the output interface is configured

to perform a conversion of a time portion of the transport signal corresponding to an output time frame into a spectral representation, wherein the portion is shorter than an input time frame, in which the parameters of the first set of parameters are organized,

to perform an upmixing operation of the spectral representation using the second set of parameters to obtain the two or more channels in the spectral representation; and

to convert each channel of the two or more channels in the spectral representation into a time representation.

15. Apparatus of claim 14, wherein the output interface is configured

to convert into a complex discrete Fourier transform domain,

to perform the upmixing operation in the complex discrete Fourier transform domain, and

to perform the conversion from the complex discrete Fourier transform domain into a real-valued time domain representation.

16. Apparatus of claim 14 or 15, wherein the output interface is configured to perform the upmixing operation based on the following equation:

$$\tilde{L}_{t,k} = \frac{(\tilde{M}_{t,k}(1+\tilde{g}_{t,b})+\tilde{r}_{t,b}g_{norm}\tilde{\rho}_{t,k})}{\sqrt{2}}$$

and

$$\tilde{R}_{t,k} = \frac{(\tilde{M}_{t,k}(1-\tilde{g}_{t,b})-\tilde{r}_{t,b}g_{norm}\tilde{\rho}_{t,k})}{\sqrt{2}},$$

wherein $\widetilde{M}_{t,k}$ is the transport signal for a frame t and a frequency bin k, wherein $\tilde{L}_{t,k}$ is a first channel of the two or more channels in the spectral representation for the frame t and the frequency bin k, wherein $\tilde{R}_{t,k}$ is a second channel of the two or more channels in the spectral representation for the frame t and the frequency bin k, wherein $\tilde{g}_{t,b}$ is a side gain parameter for the frame t and a subband b, wherein $\tilde{r}_{t,b}$ is a residual prediction gain parameter for the frame t and the suband b, wherein g_{norm} is an energy adjusting factor that can be there or not, and wherein $\tilde{\rho}_{t,k}$ is a raw residual signal for the frame t and the frequency bin k.

17. Apparatus of one of the preceding claims,

wherein the first set of parameters is a direction of arrival parameter for an input frequency band, and wherein the second set of parameters comprises a side gain parameter per input frequency band, and

wherein the parameter converter is configured to calculate the side gain parameter for the output frequency band using the following equation:

$$sidegain[b] = \sin\left(azimuth[b] * \frac{\pi}{180}\right)\cos\left(elevation[b] * \frac{\pi}{180}\right),$$

wherein b is the output frequency band, wherein sidegain is the side gain parameter, wherein azimuth is an azimuth component of the direction of arrival parameter, and wherein elevation is an elevation component of the direction of arrival parameter.

18. Apparatus of claim 17,

wherein the first set of parameters additionally comprises a diffuseness parameter for the input frequency band, and wherein the parameter converter is configured to calculate the side gain parameter for the output frequency band using the following equation

$$sidegain[b] = \sqrt{1 - diff(b)}\sin\left(azimuth[b] * \frac{\pi}{180}\right)\cos\left(elevation[b] * \frac{\pi}{180}\right)$$

wherein *diff(b)* is the diffuseness parameter for the input frequency band b.

19. Apparatus of one of the preceding claims,

wherein the first set of parameters comprises a diffuseness parameter per input frequency band, and

wherein the second set of parameters comprises a residual prediction gain parameter for an output frequency band, and

wherein the parameter converter is to use, as the residual prediction gain parameter for the output parameter band, the diffuseness parameter from the input parameter band, when the input parameter band and the output parameter band are equal to each other, or to derive, from the diffuseness parameter for the input parameter band, a diffuseness parameter for the output parameter band, and to then use the diffuseness parameter for the output parameter band as the residual prediction gain parameter for the output parameter band.

20. Apparatus of one of claims 14 to 19, wherein the information on the transport signal comprises a core encoded audio signal, and wherein the apparatus further comprises:

a core decoder for core decoding the core encoded audio signal to obtain the transport signal.

21. Apparatus of claim 20, wherein the core decoder is in an ACELP decoder, or

wherein the output interface is configured to convert the transport signal being a low band signal into a spectral representation, to upmix the spectral representation and to convert an upmixed spectral representation in a time domain to obtain a low band representation of the two or more channels,

wherein the apparatus comprises a bandwidth extension processor for generating a high band signal from the transport signal in the time domain,

wherein the apparatus comprises a multichannel filler to apply a multichannel filling operation to the transport signal in the time domain,

wherein the apparatus comprises an upmixer to apply a broad band panning in the time domain to the high band signal using at least one parameter from the second set of parameters, and

wherein the apparatus comprises a signal combiner for combining, in the time domain, a result of the broad band panning, a result of the stereo filling, and the low band representation of the two or more channels to obtain a full band multichannel signal in the time domain as the channel representation.

22. Apparatus of one of the preceding claims,

wherein the output interface is configured to generate a raw representation of the two or more channels using the second set of parameters and the transport signal,

wherein the apparatus further comprises a multichannel enhancer for generating an enhancement representation of the two or more channels using the transport signal, and

wherein the apparatus further comprises a signal combiner for combining the raw representation of the two or more channels and the enhancement representation of the two or more channels to obtain the processed audio scene.

23. Apparatus of claim 22, wherein the multichannel enhancer is configured for generating an enhancement representation of the two or more channels using the enhancement transport signal and the second set of parameters, or

wherein the multichannel enhancer comprises a transport signal enhancer for generating an enhancement transport signal, and an upmixer for upmixing the enhancement transport signal.

24. Apparatus of claim 23, wherein the transport signal is an encoded transport signal, and wherein the apparatus further comprises:

a transport signal core decoder for generating a decoded raw transport signal,

wherein the transport signal enhancer is configured for generating the enhancement transport signal using the decoded raw transport signal, and

wherein the output interface is configured to generate the raw representation of the two or more channels using the second set of parameters and the decoded raw transport signal.

25. Apparatus of claim 22 or claim 23 or claim 24, wherein the multichannel enhancer comprises either the upmixer or a multichannel filler or both the upmixer or the multichannel filler for generating the enhancement representation of the two or more channels using the transport signal or the enhancement transport signal and at least one parameter of the second set of parameters.

26. Apparatus of one of claims 22, 23, 24, or 25, wherein the output interface is configured to generate a raw representation of the two or more channels using an upmix in a second domain,

wherein the transport signal enhancer is configured for generating the enhancement transport signal in a first domain being different from the second domain, or wherein the multichannel enhancer is configured for generating the enhancement representation of the two or more channels using the enhancement transport signal in the first domain, and

wherein the signal combiner is configured to combine the raw representation of the two or more channels and the enhancement representation of the two or more channels in the first domain.

27. Apparatus of claim 26, wherein the first domain is a time domain and the second domain is a spectral domain.

28. Apparatus of one of claims 22 to 27, wherein the transport signal enhancer or the multichannel enhancer is configured to perform at least one operation of a group of operations comprising a bandwidth extension operation, a gap filling operation, a quality enhancement operation, or an interpolation operation.

29. Apparatus of one of claims 22 to 28,

wherein the transport signal enhancer or the multichannel enhancer is configured to operate in parallel with the output interface when generating the raw representation, or

wherein the parameter converter is configured to operate in parallel with the transport signal enhancer.

30. Apparatus of one of claims 24 to 29, wherein the core decoder is configured to feed the decoded raw transport signal in two parallel branches, a first branch of the two parallel branches comprising the output interface, and a second branch of the two parallel branches comprising the transport signal enhancer or the multichannel enhancer or both, and wherein the signal combiner is configured to receive a first input to be combined from the first branch and a second input to be combined from the second branch.

31. Apparatus of one of the claims 1 to 30, wherein the output interface is configured

to perform a conversion of a time portion of the transport signal corresponding to an output time frame into a spectral representation,

to perform an upmixing operation of the spectral representation using the second set of parameters to obtain the two or more channels in the spectral representation; and

to convert each channel of the two or more channels in the spectral representation into a time representation to obtain a raw time representation of the two or more channels, and

wherein the signal combiner is configured for combining the raw time representation of the two or more channels and the enhancement time representation of the two or more channels.

32. Method of processing an encoded audio scene representing a sound field related to a virtual listener position, the encoded audio scene comprising information on a

transport signal and a first set of parameters related to the virtual listener position, the method comprising:

converting the first set of parameters into a second set of parameters related to a channel representation comprising two or more channels for a reproduction at predefined spatial positions for the two or more channels; and

generating a processed audio scene using the second set of parameters and the information on the transport signal.

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