Claims

1. Apparatus for processing an audio scene representing a sound field, the audio scene comprising information on a transport signal and a set of parameters, the apparatus comprising:

an output interface for generating a raw representation of two or more channels using the set of parameters and the transport signal;

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

a transport signal enhancer for generating an enhancement transport signal wherein the transport signal 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;

an upmixer for upmixing the enhancement transport signal to obtain an upmixed enhancement transport signal; and

a multichannel filler for generating a multichannel filled transport signal using the enhancement transport signal and at least one parameter of the set of parameters,

wherein the upmixed enhancement transport signal and the multichannel filled transport signal represent the enhancement representation of the two or more channels; and

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 a processed audio scene.

2. Apparatus of claim 1, 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 set of parameters and the decoded raw transport signal.

3. Apparatus of claim 1, wherein the transport signal enhancer and the upmixer are configured to operate in parallel with the output interface when generating the raw representation of the two or more channels.

4. Apparatus of one of claims 1 to 3, wherein the output interface is configured to generate the 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, and 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.

5. Apparatus of claim 5, wherein the first domain is a time domain and the second domain is a spectral domain.

6. Apparatus of one of claims 2 to 5, wherein the transport signal 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 and the upmixer, 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.

7. Apparatus of one of the claims 1 to 6, 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 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, the time domain representation representing the raw representation of the two or more channels, and

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

8. Apparatus of one of the preceding claims,

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

9. 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 a received set of parameters are organized,

to perform an upmixing operation of the spectral representation using the 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 representing the raw representation of the two or more channels.

10. Apparatus of claim 9, 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.

11. Apparatus of claim 9 or 10, 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 $\tilde{M}_{t,k}$ 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.

12. Apparatus of claim 1,

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, as the raw representation of the two or more channels, a low band representation of the two or more channels,

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

wherein the multichannel filler is configured to apply a stereo filling operation to the high band signal from the transport signal in the time domain to obtain the multichannel filled transport signal,

wherein the upmixer is configured to apply a broad band panning in the time domain to the high band signal from the transport signal using the at least one parameter of the set of parameters to obtain the upmixed enhancement transport signal, and

wherein the signal combiner is configured for combining, in the time domain, the upmixed enhancement transport signal, the multichannel filled transport signal, and, as the raw representation of two or more channels, the low band representation of the two or more channels to obtain a full band multichannel signal in the time domain as the processed audio scene.

13. Apparatus of claim 12, wherein the multichannel filler is configured to generate a left channel of the multichannel filled transport signal using the equation:

$$PRED_{hb,left}[i] = g_{pred,hb} * g_{norm} * HB_{dmx}[i-d], \text{ and}$$

wherein the multichannel filler is configured to generate a right channel of the multichannel filled transport signal using the equation:

$$PRED_{hb,right}[i] = -g_{pred,hb} * g_{norm} * HB_{dmx}[i-d],$$

wherein $PRED_{hb,left}[i]$ is a left channel of the multichannel filled transport signal for a sample number i in a current frame, wherein $PRED_{hb,right}[i]$ is a right channel of the multichannel filled transport signal for the sample number i in the current frame, wherein $g_{pred,hb}$ is a high band prediction gain as the at least one parameter, wherein g_{norm} is an

energy normalization factor, wherein i is the sample number in the current frame, and wherein d is a number of samples by which the high band signal HB_{dmx} from the enhancement transport signal is delayed.

14. Method of processing an audio scene representing a sound field related to a virtual listener position, the audio scene comprising information on a transport signal and a set of parameters, the method comprising:

generating a raw representation of two or more channels using the set of parameters and the transport signal,

multichannel generating an enhancement representation of the two or more channels using the transport signal, wherein the multichannel generating comprises:

generating an enhancement transport signal using performing 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,

upmixing the enhancement transport signal to obtain an upmixed enhancement transport signal; and

multichannel filling for generating a multichannel filled transport signal using the enhancement transport signal and at least one parameter of the set of parameters,

wherein the upmixed enhancement transport signal and the multichannel filled transport signal represent the enhancement representation of the two or more channels, and

combining the raw representation of the two or more channels and the enhancement representation of the two or more channels to obtain a processed audio scene.

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