Claims

1. A decoder for generating an audio output signal comprising one or more audio output channels from a downmix signal comprising one or more downmix channels, wherein the downmix signal encodes two or more audio object signals, wherein the decoder comprises:

a threshold determiner for determining a threshold value depending on a signal energy or a noise energy of at least one of the two or more audio object signals or depending on a signal energy or a noise energy of at least one of the one or more downmix channels, and

a processing unit for generating the one or more audio output channels from the one or more downmix channels depending on the threshold value,

wherein the processing unit is configured to generate the one or more audio output channels from the one or more downmix channels depending on an object covariance matrix ($\mathbf{E}$) of the one or more audio object signals, depending on a downmix matrix ($\mathbf{D}$) for downmixing the two or more audio object signals to obtain the one or more downmix channels, and depending on the threshold value,

wherein the processing unit is configured to generate the one or more audio output channels from the one or more downmix channels by applying the threshold value in a function to inverse a downmix channel cross correlation matrix $\mathbf{Q}$,

wherein $\mathbf{Q}$ is defined as $\mathbf{Q}=\mathbf{DED}^{*}$,

wherein $\mathbf{D}$ is the downmix matrix for downmixing the two or more audio object signals to obtain the two or more downmix channels,

wherein $\mathbf{E}$ is the object covariance matrix of the one or more audio object signals, and

wherein the processing unit is configured to generate the one or more audio output channels from the one or more downmix channels by computing the eigenvalues of the downmix channel cross correlation matrix $\mathbf{Q}$.

2. A decoder according to claim 1,

wherein the downmix signal comprises two or more downmix channels, and

wherein the threshold determiner is configured to determine the threshold value depending on a noise energy of each of the two or more downmix channels.

3. A decoder according to claim 2, wherein the threshold determiner is configured to determine the threshold value depending on the sum of all noise energy in the two or more downmix channels.

4. A decoder according to any one of the preceding claims,

wherein threshold determiner is configured to determine the threshold value depending on a signal energy of the audio object signal of the two or more audio object signals which has the greatest signal energy of the two or more audio object signals.

5. A decoder according to any one of the preceding claims,

wherein the downmix signal encodes the two or more audio object signals for each time-frequency tile of a plurality of time-frequency tiles,

wherein the threshold determiner is configured to determine a threshold value for each time-frequency tile of the plurality of time-frequency tiles depending on the signal energy or the noise energy of at least one of the two or more audio object signals or depending on the signal energy or the noise energy of at least one of the one or more downmix channels, wherein a first threshold value of a first time-frequency tile of the plurality of time-frequency tiles differs from a second time-frequency tile of the plurality of time-frequency tiles, and

wherein the processing unit is configured to generate for each time-frequency tile of the plurality of time-frequency tiles a channel value of each of the one or more audio output channels from the one or more downmix channels depending on the threshold value of said time-frequency tile.

6. A decoder according to any one of the preceding claims,

wherein the downmix signal comprises two or more downmix channels,

wherein the decoder is configured to determine the threshold value $\mathbf{T}$ in decibel according to the formula

$\mathbf{T}[dB] = \mathbf{E}_{noise}[dB] - \mathbf{E}_{ref}[dB] - \mathbf{Z}$ or according to the formula

$\mathbf{T}[dB] = \mathbf{E}_{noise}[dB] - \mathbf{E}_{ref}[dB]$,

wherein $\mathbf{T}[dB]$ indicates the threshold value in decibel,

wherein $\mathbf{E}_{noise}[dB]$ indicates the sum of all noise energy in the two or more downmix channels in decibel, or $\mathbf{E}_{noise}[dB]$ indicates the sum of all noise energy in the two or more downmix channels in decibel divided by the number of the two or more downmix channels,

wherein $\mathbf{E}_{ref}[dB]$ indicates the signal energy of one of the audio object signals in decibel, and

wherein $\mathbf{Z}$ indicates an additional parameter being a number.

7. A decoder according to any one of claims 1 to 5,

wherein the downmix signal comprises two or more downmix channels,

wherein the decoder is configured to determine the threshold value $\mathbf{T}$ according to the formula

$\mathbf{T} = \frac{\mathbf{E}_{noise}}{\mathbf{E}_{ref} \cdot \mathbf{Z}}$ or according to the formula

$\mathbf{T} = \frac{\mathbf{E}_{noise}}{\mathbf{E}_{ref}}$,

wherein $\mathbf{T}$ indicates the threshold value,

wherein $\mathbf{E}_{noise}$ indicates the sum of all noise energy in the two or more downmix channels, or $\mathbf{E}_{noise}$ in decibel indicates the sum of all noise energy in the two or

more downmix channels in decibel divided by the number of the two or more downmix channels,

wherein $\mathbf{E}_{ref}$ indicates the signal energy of one of the audio object signals, and

wherein $\mathbf{Z}$ indicates an additional parameter being a number.

8. A decoder according to any one of the preceding claims, wherein the processing unit is configured to generate the one or more audio output channels from the one or more downmix channels by multiplying the largest eigenvalue of the eigenvalues of the downmix channel cross correlation matrix $\mathbf{Q}$ with the threshold value to obtain a relative threshold.

9. A decoder according to claim 8,

wherein the processing unit is configured to generate the one or more audio output channels from the one or more downmix channels by generating a modified matrix,

wherein the processing unit is configured to generate the modified matrix depending on only those eigenvectors of the downmix channel cross correlation matrix $\mathbf{Q}$, which have an eigenvalue of the eigenvalues of the downmix channel cross correlation matrix $\mathbf{Q}$, which is greater than or equal to the relative threshold,

wherein the processing unit is configured to conduct a matrix inversion of the modified matrix to obtain an inverted matrix, and

wherein the processing unit is configured to apply the inverted matrix on one or more of the downmix channels to generate the one or more audio output channels.

10. A method for generating an audio output signal comprising one or more audio output channels from a downmix signal comprising one or more downmix channels, wherein the downmix signal encodes two or more audio object signals, wherein the method comprises:

determining a threshold value depending on a signal energy or a noise energy of at least one of the two or more audio object signals or depending on a signal energy or a noise energy of at least one of the one or more downmix channels, and

generating the one or more audio output channels from the one or more downmix channels depending on the threshold value,

wherein generating the one or more audio output channels from the one or more downmix channels depending on an object covariance matrix ($\mathbf{E}$) of the one or more audio object signals is conducted depending on a downmix matrix ($\mathbf{D}$) for downmixing the two or more audio object signals to obtain the one or more downmix channels, and depending on the threshold value,

wherein generating the one or more audio output channels from the one or more downmix channels is conducted by applying the threshold value in a function to inverse a downmix channel cross correlation matrix $\mathbf{Q}$,

wherein $\mathbf{Q}$ is defined as $\mathbf{Q}=\mathbf{D}\mathbf{E}\mathbf{D}^{*}$,

wherein $\mathbf{D}$ is the downmix matrix for downmixing the two or more audio object signals to obtain the two or more downmix channels,

wherein $\mathbf{E}$ is the object covariance matrix of the one or more audio object signals, and

wherein generating the one or more audio output channels from the one or more downmix channels is conducted by computing the eigenvalues of the downmix channel cross correlation matrix $\mathbf{Q}$.

11. A computer program for implementing the method of claim 10 when being executed on a computer or signal processor.