Claims

1. Audio decoder for decoding a multi-object audio signal consisting of a downmix signal ~~(X)~~ and side information ~~(PSI)~~, the side information comprising object-specific side information ~~(PSI_i)~~ for at least one audio object ~~(s_i)~~ in at least one time/frequency region ~~($R(t_R,f_R)$)~~, and object-specific time/frequency resolution information ~~($TFRI_i$)~~ indicative of an object-specific time/frequency resolution ~~(TFR_h)~~ of the object-specific side information for the at least one audio object ~~(s_i)~~ in the at least one time/frequency region ~~($R(t_R,f_R)$)~~, the audio decoder comprising:

an object-specific time/frequency resolution determiner ~~(110)~~ configured to determine the object-specific time/frequency resolution information ~~($TFRI_i$)~~ from the side information ~~(PSI)~~ for the at least one audio object ~~(s_i)~~; and

an object separator ~~(120)~~ configured to separate the at least one audio object ~~(s_i)~~ from the downmix signal ~~(X)~~ using the object-specific side information in accordance with the object-specific time/frequency resolution ~~($TFRI_i$)~~.

2. Audio decoder according to claim 1, wherein the object-specific side information is a fine structure object-specific side information ~~($fsI_i^{n,k}$, $fsI_j^{n,k}$)~~ for the at least one audio object ~~(s_i)~~ in the at least one time/frequency region ~~($R(t_R,f_R)$)~~, and wherein the side information ~~(PSI)~~ further comprises coarse object-specific side information for the at least one audio object ~~(s_i)~~ in the at least one time/frequency region ~~($R(t_R,f_R)$)~~, the coarse object-specific side information being constant within the at least one time/frequency region ~~($R(t_R,f_R)$)~~.

3. Audio decoder according to claim 1, wherein the fine structure object-specific side information ~~($fsI_i^{n,k}$)~~ describes a difference between the coarse object-specific side information and the at least one audio object ~~(s_i)~~.

4. Audio decoder according to any one of the preceding claims, wherein the downmix signal ~~(X)~~ is sampled in the time/frequency domain into a plurality of time-slots and a plurality of (hybrid) sub-bands, wherein the time/frequency region ~~($R(t_R,f_R)$)~~ extends over at least two samples of the downmix signal ~~(X)~~, and wherein the object-specific time/frequency resolution ~~(TFR_h)~~ is finer in at least one of both dimensions than the time/frequency region ~~($R(t_R,f_R)$)~~.

5. Audio decoder according to any one of the preceding claims, wherein the object separator ~~(120)~~ is configured to determine an estimated covariance matrix ~~($\mathbb{E}^{\eta,\kappa}$)~~ with elements $e_{i,j}^{\eta,\kappa}$ of the at least one audio object ~~(s_i)~~ and at least one further audio object ~~(s_j)~~ according to

$$e_{i,j}^{\eta,\kappa} = \sqrt{fsl_i^{\eta,\kappa} fsl_j^{\eta,\kappa}}\, fsc_{i,j}^{\eta,\kappa},$$

wherein

$e_{i,j}^{\eta,\kappa}$ is the estimated covariance of audio objects i and j for fine-structure time-slot η and fine-structure (hybrid) sub-band κ;

$fsl_i^{\eta,\kappa}$ and $fsl_j^{\eta,\kappa}$ are the object-specific side information of the audio objects i and j for fine-structure time-slot η and fine-structure (hybrid) sub-band κ;

$fsc_{i,j}^{\eta,\kappa}$ is an inter object correlation information of the audio objects i and j, respectively, fine-structure time-slot η and fine-structure (hybrid) sub-band κ;

wherein at least one of $fsl_i^{\eta,\kappa}$, $fsl_j^{\eta,\kappa}$, and $fsc_{i,j}^{\eta,\kappa}$ varies within the time/frequency region ~~($R(t_R, f_R)$)~~ according to the object-specific time/frequency resolution ~~(TFR_h)~~ for the audio objects i and j indicated by the object-specific time/frequency resolution information ~~($TFRI_i$, $TFRI_j$)~~, and

wherein the object separator ~~(120)~~ is further configured to separate the at least one audio object ~~(s_i)~~ from the downmix signal ~~(X)~~ using the estimated covariance matrix ~~($\mathbb{E}^{\eta,\kappa}$)~~.

6. Audio decoder according to any one of the preceding claims, further comprising:

a downmix signal time/frequency transformer configured to transform the downmix signal ~~(X)~~ within the time/frequency region ~~($R(t_R, f_R)$)~~ from a downmix signal time/frequency resolution to at least the object-specific time/frequency resolution ~~(TFR_h)~~ of the at least one audio object ~~(s_i)~~ to obtain a re-transformed downmix signal ~~($X^{\eta,\kappa}$)~~;

an inverse time/frequency transformer configured to time/frequency transform the at least one audio object ~~(s_i)~~ within the time/frequency region ~~($R(t_R, f_R)$)~~ from the object-specific time/frequency resolution ~~(TFR_h)~~ back to a common t/f-resolution or the downmix signal time/frequency resolution;

wherein the object separator (120) is configured to separate the at least one audio object (s_i) from the downmix signal (X) at the object-specific time/frequency resolution (TFR_h).

7. Audio encoder for encoding a plurality of audio objects (s_i) into a downmix signal (X) and side information (PSI), the audio encoder comprising:

a time-to-frequency transformer configured to transform the plurality of audio objects (s_i) at least to a first plurality of corresponding transformations ($s_{1,1}(t,f), \ldots s_{N,1}(t,f)$) using a first time/frequency resolution (TFR_1) and to a second plurality of corresponding transformations ($s_{1,2}(t,f), \ldots s_{N,2}(t,f)$) using a second time/frequency resolution (TFR_2);

a side information determiner (t/f-SIE) configured to determine at least a first side information for the first plurality of corresponding transformations ($s_{1,1}(t,f) \ldots s_{N,1}(t,f)$) and a second side information for the second plurality of corresponding transformations ($s_{1,2}(t,f) \ldots s_{N,2}(t,f)$), the first and second side information indicating a relation of the plurality of audio objects (s_i) to each other in the first and second time/frequency resolutions (TFR_1, TFR_2), respectively, in a time/frequency region ($R(t_R, f_R)$); and

a side information selector (SI-AS) configured to select, for at least one audio object (s_i) of the plurality of audio objects, one object-specific side information from at least the first and second side information on the basis of a suitability criterion indicative of a suitability of at least the first or second time/frequency resolution for representing the audio object (s_i) in the time/frequency domain, the object-specific side information being inserted into the side information (PSI) output by the audio encoder.

8. Audio encoder according to claim 7, wherein the suitability criterion is based on a source estimation and wherein the side information selector (SI-AS) comprises:

a source estimator configured to estimate at least a selected audio object of the plurality of audio objects (s_i) using the downmix signal (X) and at least the first information and the second information corresponding to the first and second time/frequency resolutions (TFR_1, TFR_2), respectively, the source estimator thus providing at least a first estimated audio object ($s_{i,estim1}$) and a second estimated audio object ($s_{i,estim2}$);

a quality assessor configured to assess a quality of at least the first estimated audio object ~~($s_{i,estim1}$)~~ and the second estimated audio object ~~($s_{i,estim2}$)~~.

9. Audio encoder according to claim 8, wherein the quality assessor is configured to assess the quality of at least the first estimated audio object ~~($s_{i,estim1}$)~~ and the second estimated audio object ~~($s_{i,estim2}$)~~ on the basis of a signal-to-distortion ratio ~~(SDR)~~ as a source estimation performance measure, the signal-to-distortion ratio ~~(SDR)~~ being determined solely on the basis of the side information ~~(PSI)~~.

10. Audio encoder according to any one of claims 7 to 9, wherein the suitability criterion for the at least one audio object ~~(s_i)~~ among the plurality of audio objects is based on degrees of sparseness of more than one t/f-resolution representations of the at least one audio object according to at least the first time/frequency resolution ~~(TFR_1)~~ and the second time/frequency resolution ~~(TFR_2)~~, and wherein the side information selector ~~(SI-AS)~~ is configured to select the side information among at least the first and second side information that is associated with the most sparse t/f-representation of the at least one audio object ~~(s_i)~~.

11. Audio encoder according to any one of claims 7 to 10, wherein the side information determiner ~~(t/f-SIE)~~ is further configured to provide fine structure object-specific side information ~~($fsl_i^{n,k}$)~~ and coarse object-specific side information as a part of at least one of the first side information and the second side information, the coarse object-specific side information being constant within the at least one time/frequency region ~~($R(t_R,f_R)$)~~.

12. Audio encoder according to claim 11, wherein the fine structure object-specific side information ~~($fsl_i^{n,k}$)~~ describes a difference between the coarse object-specific side information and the at least one audio object ~~(s_i)~~.

13. Audio encoder according to any one of the claims 7 to 12, further comprising a downmix signal processor configured to transform the downmix signal ~~(X)~~ to a representation that is sampled in the time/frequency domain into a plurality of time-slots and a plurality of (hybrid) sub-bands, wherein the time/frequency region ~~($R(t_R,f_R)$)~~ extends over at least two samples of the downmix signal ~~(X)~~, and wherein an object-specific time/frequency resolution ~~(TFR_h)~~ specified for at least one audio object is finer in at least one of both dimensions than the time/frequency region ~~($R(t_R,f_R)$)~~.

14. Method for decoding a multi-object audio signal consisting of a downmix signal ~~(X)~~ and side information ~~(PSI)~~, the side information comprising object-specific side infor-

mation (PSI_i) for at least one audio object (s_i) in at least one time/frequency region ($R(t_R,f_R)$), and object-specific time/frequency resolution information ($TFRI_i$) indicative of an object-specific time/frequency resolution (TFR_h) of the object-specific side information for the at least one audio object (s_i) in the at least one time/frequency region ($R(t_R,f_R)$), the method comprising:

determining the object-specific time/frequency resolution information ($TFRI_i$) from the side information (PSI) for the at least one audio object (s_i); and

separating the at least one audio object (s_i) from the downmix signal (X) using the object-specific side information in accordance with the object-specific time/frequency resolution ($TFRI_i$).

15. Method for encoding a plurality of audio object (s_i) to a downmix signal (X) and side information (PSI), the method comprising:

transforming the plurality of audio object (s_i) at least to a first plurality of corresponding transformations ($s_{1,1}(t,f)\ldots s_{N,1}(t,f)$) using a first time/frequency resolution (TFR_1) and to a second plurality of corresponding transformations ($s_{1,2}(t,f)\ldots s_{N,2}(t,f)$) using a second time/frequency resolution (TFR_2);

determining at least a first side information for the first plurality of corresponding transformations ($s_{1,1}(t,f)\ldots s_{N,1}(t,f)$) and a second side information for the second plurality of corresponding transformations ($s_{1,2}(t,f)\ldots s_{N,2}(t,f)$), the first and second side information indicating a relation of the plurality of audio object (s_i) to each other in the first and second time/frequency resolutions (TFR_1, TFR_2), respectively, in a time/frequency region ($R(t_R,f_R)$); and

selecting, for at least one audio object (s_i) of the plurality of audio objects, one object-specific side information from at least the first and second side information on the basis of a suitability criterion indicative of a suitability of at least the first or second time/frequency resolution for representing the audio object (s_i) in the time/frequency domain, the object-specific side information being inserted into the side information (PSI) output by the audio encoder.

16. Audio decoder for decoding a multi-object audio signal consisting of a downmix signal (X) and side information (PSI), the side information comprising object-specific side information (PSI_i) for at least one audio object (s_i) in at least one time/frequency region

~~($R(t_R,f_R)$)~~, and object-specific time/frequency resolution information ~~($TFRI_i$)~~ indicative of an object-specific time/frequency resolution ~~(TFR_h)~~ of the object-specific side information for the at least one audio object ~~(s_i)~~ in the at least one time/frequency region ~~($R(t_R,f_R)$)~~, the audio decoder comprising:

an object-specific time/frequency resolution determiner ~~(110)~~ configured to determine the object-specific time/frequency resolution information ~~($TFRI_i$)~~ from the side information ~~(PSI)~~ for the at least one audio object ~~(s_i)~~; and

an object separator ~~(120)~~ configured to separate the at least one audio object ~~(s_i)~~ from the downmix signal ~~(X)~~ using the object-specific side information in accordance with the object-specific time/frequency resolution ~~($TFRI_i$)~~, wherein object-specific side information for at least one other audio object ~~(s_j)~~ within the downmix signal has a different object-specific time/frequency resolution ~~(TFR)~~.

17. Method for decoding a multi-object audio signal consisting of a downmix signal ~~(X)~~ and side information ~~(PSI)~~, the side information comprising object-specific side information ~~(PSI_i)~~ for at least one audio object ~~(s_i)~~ in at least one time/frequency region ~~($R(t_R,f_R)$)~~, and object-specific time/frequency resolution information ~~($TFRI_i$)~~ indicative of an object-specific time/frequency resolution ~~(TFR_h)~~ of the object-specific side information for the at least one audio object ~~(s_i)~~ in the at least one time/frequency region ~~($R(t_R,f_R)$)~~, the method comprising:

determining the object-specific time/frequency resolution information ~~($TFRI_i$)~~ from the side information ~~(PSI)~~ for the at least one audio object ~~(s_i)~~; and

separating the at least one audio object ~~(s_i)~~ from the downmix signal ~~(X)~~ using the object-specific side information in accordance with the object-specific time/frequency resolution ~~($TFRI_i$)~~, wherein object-specific side information for at least one other audio object ~~(s_j)~~ within the downmix signal has a different object-specific time/frequency resolution ~~(TFR)~~.

18. Computer program for performing the method according to claim 14, 15, or 17 when the computer program runs on a computer.

19. Audio decoder for decoding a multi-object audio signal consisting of a downmix signal and side information, the side information comprising object-specific side information for at least one audio object in at least one

time/frequency region, and object-specific time/frequency resolution information indicative of an object-specific time/frequency resolution of the object-specific side information for the at least one audio object in the at least one time/frequency region, the audio decoder comprising:

an object-specific time/frequency resolution determiner configured to deter-mine the object-specific time/frequency resolution information from the side information for the at least one audio object; and

an object separator configured to separate the at least one audio object from the downmix signal using the object-specific side information in accord-ance with the object-specific time/frequency resolution,

wherein the object-specific side information is a fine structure object-specific side information for the at least one audio object in the at least one time/frequency region, and wherein the side information further comprises coarse object-specific side information for the at least one audio object in the at least one time/frequency region, the coarse object-specific side information being constant within the at least one time/frequency region, or

wherein the fine structure object-specific side information describes a difference between the coarse object-specific side information and the at least one audio object.

20. Method for decoding a multi-object audio signal consisting of a downmix signal and side information, the side information comprising object-specific side information for at least one audio object in at least one time/frequency region, and object-specific time/frequency resolution information indicative of an object-specific time/frequency resolution of the object-specific side information for the at least one audio object in the at least one time/frequency region, the method comprising:

determining the object-specific time/frequency resolution information from the side information for the at least one audio object; and

separating the at least one audio object from the downmix signal using the object-specific side information in accordance with the object-specific time/frequency resolution,

wherein the object-specific side information is a fine structure object-specific side information for the at least one audio object in the at least one time/frequency region, and wherein the side information further comprises coarse object-specific side information for the at least one audio object in the at least one time/frequency region, the coarse object-specific side information being constant within the at least one time/frequency region, or

wherein the fine structure object-specific side information describes a difference between the coarse object-specific side information and the at least one audio object.