

CLAIMS

- 1) A method (600) for encoding a multi-channel input signal (101) which comprises N different channels; wherein the input signal (101) comprises a sequence of frames;
5 wherein the method (600) comprises, for a current frame of the sequence of frames,
– determining (601) whether the current frame is an active frame or an inactive frame using a signal and/or a voice activity detector;
– determining (602) a downmix signal (103) based on the multi-channel input signal (101); wherein the downmix signal (103) comprises less than or equal to N
10 channels;
– determining (603) upmixing metadata (105) comprising a set of parameters for generating, based on the downmix signal (103), a reconstructed multi-channel signal (111) comprising N channels; wherein the upmixing metadata (105) is determined in dependance of whether the current frame is an active frame or an
15 inactive frame; and
– encoding (604) the upmixing metadata (105) into a bitstream.
- 2) The method (600) of claim 1, wherein the method (600) comprises
– determining a covariance, in particular a covariance matrix, for the current frame
20 of the multi-channel input signal (101), by performing temporal smoothing over multiple frames from the sequence of frames, in dependance of whether the current frame is an active frame or an inactive frame; and
– determining one or more parameters of the upmixing metadata (105) based on the covariance for the current frame.
- 25 3) The method (600) of claim 2, wherein
– a forgetting factor for the temporal smoothing is dependent on whether the current frame is an active frame or an inactive frame; and/or
– if the current frame is an active frame, the forgetting factor is shorter than if the
30 current frame is an inactive frame; and/or

- if the current frame is an active frame, an amount of temporal smoothing is lower than if the current frame is an inactive frame.
- 4) The method (600) of claim 3, wherein the forgetting factor for temporal smoothing of the covariance is frequency dependent.
- 5) The method (600) of any of claims 2 to 4, wherein the method (600) comprises
- identifying one or more frames from the sequence of frames, which correspond to a transient noise or talk burst; and
 - ignoring the one or more identified frames when performing temporal smoothing, and/or removing the one or more identified frames from the temporal smoothing calculation for determining the covariance for the current frame.
- 6) The method (600) of any of claims 2 to 5, wherein the method (600) comprises
- determining that the current frame is an active frame following a set of one or more preceding inactive frames from the sequence of frames by less than a pre-determined burst duration threshold; and
 - determining the covariance for the current frame without taking into account the current frame.
- 7) The method (600) of claim 6, wherein the covariance for the current frame is determined based on a reference covariance which has been determined based on the set of one or more preceding inactive frames.
- 8) The method (600) of any of claims 2 to 5, wherein the method (600) comprises
- determining that the current frame is an active frame following a set of one or more preceding inactive frames from the sequence of frames by more than a pre-determined burst duration threshold; and
 - determining the covariance for the current frame based on samples, in particular only based on samples, from the current frame.

- 9) The method (600) of any of claims 2 to 5, wherein the method (600) comprises
- determining that the current frame is an inactive frame; and
 - storing the covariance for the current frame as a reference covariance for performing temporal smoothing for determining the covariance for a subsequent
- 5 frame from the sequence of frames, which follows the current frame.
- 10) The method (600) of any claims 2 to 9, wherein the method (600) comprises, for a subsequent frame from the sequence of frames, which follows the current frame, and if the current frame is an inactive frame,
- 10
- determining whether the subsequent frame is an active frame or an inactive frame using the signal and/or the voice activity detector;
 - if the subsequent frame is an inactive frame, determining the covariance for the subsequent frame based on samples of the current frame, in particular based on the covariance for the current frame, and based on samples of the subsequent frame.
- 15
- 11) The method (600) of any previous claim, wherein the method (600) comprises
- determining a covariance, in particular a covariance matrix, for the current frame of the multi-channel input signal (101) within a number of different frequency bands; wherein the number of frequency bands is dependent on whether the
- 20 current frame is an active frame or an inactive frame; and
- determining one or more parameters of the upmixing metadata (105) based on the covariance for the current frame.
- 12) The method (600) of claim 11, wherein if the current frame is an active frame, the number
- 25 of frequency bands is higher than if the current frame is an inactive frame.
- 13) The method (600) of any previous claim, wherein if the current frame is an active frame, the set of parameters of the upmixing parameters (105) is larger and/or comprises a higher number of different parameters than if the current frame is an inactive frame.
- 30
- 14) The method (600) of any previous claim, wherein the method (600) comprises,

- if it is determined that the current frame is an inactive frame, determining whether a preceding frame from the sequence of frames, which directly precedes the current frame, has been an inactive frame; and
- if the preceding frame has been an inactive frame and if a number of consecutive inactive frames since a last transmission of upmixing metadata (105) is less than a pre-computed number, refraining from encoding (604) the upmixing metadata (105) for the current frame into the bitstream; and/or
- if the preceding frame has been an active frame or if the number of consecutive inactive frames since the last transmission of upmixing metadata (105) is equal to the pre-computed number, encoding (604) the upmixing metadata (105) for the current frame into the bitstream.

- 15) The method (600) of any previous claim, wherein the method (600) comprises, if a preceding frame from the sequence of frames, which directly precedes the current frame, has been an active frame,
- determining the number of channels of the downmix signal (103) for the preceding claim; and
 - maintaining the same number of channels of the downmix signal (103) for the inactive current frame, as for the downmix signal (103) for the preceding claim.

- 16) The method (600) of any previous claims, wherein the set of parameters of the upmixing metadata (105) describes and/or models a spatial characteristic of audio content, in particular of noise, comprised within the current frame of the multi-channel input signal (101).

- 17) The method (600) of any previous claims, wherein
- the multi-channel input signal (101) comprises a soundfield representation, referred to as SR, signal; and/or
 - the upmixing metadata (105) comprises Spatial Reconstructor, referred to as SPAR, metadata.

- 18) The method (600) of any previous claims, wherein

- the multi-channel input signal (101) comprises an ambisonics signal with a W channel, a Y channel, a Z channel and an X channel; and
- the set of parameters of the upmixing metadata (105) comprises prediction coefficients for predicting the Y channel, the Z channel and the X channel based on a representation of the W channel, thereby providing residual channels, referred to as Y' channel, Z' channel and X' channel, respectively.

19) The method (600) of claim 18, wherein

- the method (600) comprises, if the downmix signal (103) comprises a first residual channel in addition to the representation of the W channel, determining a cross-prediction parameter as part of the upmixing metadata (105), which is dependent of a covariance between the first residual channel and the one or more remaining residual channels;
- the cross-prediction parameter is not part of the upmixing metadata (105) that is encoded into the bitstream for the current frame, if the current frame is an inactive frame; and
- the cross-prediction parameter is part of the upmixing metadata (105) that is encoded into the bitstream for the current frame, if the current frame is an active frame.

20) The method (600) of any of claims 18 to 19, wherein the method (600) comprises determining a decorrelation parameter for generating a decorrelated channel when reconstructing a residual channel that has not been included into the downmix signal (103), based on a covariance of the residual channel.

21) The method (600) of claim 20, wherein if more than one residual channel has not been included into the downmix signal (103),

- the set of parameters of the upmixing metadata (105) for the current frame comprises a decorrelation parameter for each possible combination of a non-included residual channel either with itself or with another one of the non-included residual channels, if the current frame is an active frame; and

- the set of parameters of the upmixing metadata (105) for the current frame comprises a decorrelation parameter only for the combinations of a non-included residual channel with itself, if the current frame is an inactive frame.

5 22) The method (600) of any previous claims, wherein

- the set of parameters comprises corresponding parameters for a number of different frequency bands; and
- if the current frame is an active frame, the number of different frequency bands is higher than if the current frame is an inactive frame.

10

23) The method (600) of any previous claims, wherein the method (600) comprises,

- determining that the current frame is an inactive frame following a subsequence of one or more previous inactive frames;
- determining whether a spatial and/or spectral characteristic of background noise comprised within the current frame and/or whether a signal-to-noise ratio of the current frame has changed with regards to the subsequence of one or more previous inactive frames; and
- encoding (604) the upmixing metadata (105) for the current frame into the bitstream, if, in particular only if, it is determined that the spatial and/or spectral characteristic of background noise comprised within the current frame and/or the signal-to-noise ratio of the current frame has changed with regards to the subsequence of one or more previous inactive frames.

15

20

24) The method (600) of any previous claims, wherein the method (600) comprises,

- 25 – determining that the current frame is an inactive frame following one or more previous inactive frames;
- determining a value of a distance measure between the covariance and/or the upmixing metadata (105) for the current frame and a previous covariance and/or previous upmixing metadata (105) for the one or more previous inactive frames;
- 30 – determining whether the value of the distance measure is greater than a pre-determined distance threshold; and

- encoding (604) the upmixing metadata (105) for the current frame into the bitstream, if, in particular only if, the value of the distance measure is greater than the pre-determined distance threshold; and/or
- refraining from encoding (604) the upmixing metadata (105) for the current frame into the bitstream, if, in particular only if, the value of the distance measure is smaller than the pre-determined distance threshold.

25) The method (600) of any previous claims, wherein

- the method (600) comprises quantizing the parameters from the set of parameters for encoding (604) the upmixing metadata (105) for the current frame into the bitstream, using a quantizer; and
- the quantizer, in particular a quantization step size and/or a number of quantization steps of the quantizer, is dependent on
 - whether the current frame is an active frame or an inactive frame; and/or
 - the number of channels of the downmix signal (103); and/or
 - a type of channel for which the parameters are to be quantized.

26) The method (600) of any previous claims, wherein the method (600) comprises,

- encoding (604) the one or more channels of the downmix signal (103) individually using one or more instances of a single channel audio encoder to provide audio data (106) to be inserted into the bitstream; and/or
- entropy encoding the set of parameters of the upmixing metadata (105) to provide coded metadata (107) to be inserted into the bitstream.

27) The method (600) of any previous claims, wherein the method (600) comprises encoding the downmix signal (103) for the current frame into the bitstream, if, in particular only if, the current frame is an active frame.

28) The method (600) of any previous claims, wherein the method (600) comprises, if the current frame is an inactive frame,

- determining spectral data for each one of the one or more channels of the downmix signal (103) individually using one or more instances of a single channel audio encoder; and
- inserting the spectral data into the bitstream.

5

29) The method (600) of any previous claims, wherein the method (600) is repeated for each frame from the sequence of frames.

30) The method (600) of any previous claims, wherein $N > 2$.

10

31) The method (600) of any previous claims, wherein

- the multi-channel input signal (101) comprises an ambisonics signal with a W channel, a Y channel, a Z channel and an X channel; and
- the method (600) comprises mixing the W channel with the Y channel, the Z channel and/or the X channel using a mixing factor, to generate a channel of the downmix signal (103); wherein the mixing factor is dependent on whether the current frame is an active frame or an inactive frame.

15

32) The method (600) of claim 31, wherein if the current frame is an active frame, the mixing factor is higher than if the current frame is an inactive frame.

20

33) A method (610) for decoding a bitstream which is indicative of a reconstructed multi-channel signal (111) comprising N channels; wherein the reconstructed signal (111) comprises a sequence of frames; wherein the method (610) comprises, for a current frame of the sequence of frames,

25

- determining (611), based on the bitstream, whether the current frame is an active frame or an inactive frame;
- determining (612) a reconstructed downmix signal (114); wherein the reconstructed downmix signal (114) comprises less than or equal to N channels; and
- generating (613) the reconstructed multi-channel signal (111) based on the reconstructed downmix signal (114) and based on upmixing metadata (105)

30

comprised within the bitstream; wherein the reconstructed multi-channel signal (111) is generated in dependance of whether the current frame is an active frame or an inactive frame.

- 5 34) The method (610) of claim 33, wherein the method (610) comprises,
- generating a decorrelated channel for a channel of the reconstructed multi-channel signal (111) that is not included in the reconstructed downmix signal (114); and
 - adding a spatial property to the decorrelated channel based on the upmixing metadata (105), to generate a reconstructed channel of the reconstructed multi-
- 10 channel signal (111), other than the one or more reconstructed channels comprised within the reconstructed downmix signal (114).

- 35) The method (610) of claim 34, wherein the decorrelated channel is generated based on the reconstructed downmix signal (114).

15

- 36) The method (610) of any of claims 33 to 35, wherein
- the reconstructed downmix signal (114) is generated based on audio data (106) comprised within the bitstream, if the current frame is an active frame; and/or
 - the reconstructed downmix signal (114) is generated using a random noise
- 20 generator and spectral data comprised within the bitstream, if the current frame is an inactive frame.

- 37) The method (610) of any of claims 33 to 36, wherein the method (610) comprises,
- determining that the current frame is an inactive frame following one or more
- 25 previous inactive frames; wherein previous upmixing metadata (105) was used for generating the reconstructed multi-channel signal (111) for the one or more previous inactive frames;- determining that the bitstream comprises updated upmixing metadata (105) for the current frame; and
- generating (613) the reconstructed multi-channel signal (111) for the current

30 frame based on the updated upmixing metadata (105).

38) The method (610) of claim 37, wherein the method (610) comprises,

- performing metadata smoothing based on the updated upmixing metadata (105) and the previous upmixing metadata (105), to determine smoothed upmixing metadata (105); and
- 5 – generating (613) the reconstructed multi-channel signal (111) for the current frame based on the smoothed upmixing metadata (105).

39) The method (610) of claim 38, wherein metadata smoothing comprises crossfading from the previous upmixing metadata (105) to the updated upmixing metadata (105) within one
10 or more inactive frames starting with the current frame.

40) An encoding unit (100) for encoding a multi-channel input signal (101) which comprises N different channels; wherein the input signal (101) comprises a sequence of frames; wherein for a current frame of the sequence of frames, the encoding unit (100) is
15 configured to

- determine whether the current frame is an active frame or an inactive frame using a signal and/or a voice activity detector;
- determine a downmix signal (103) based on the multi-channel input signal (101); wherein the downmix signal (103) comprises less than or equal to N channels;
- 20 – determine upmixing metadata (105) comprising a set of parameters for generating, based on the downmix signal (103), a reconstructed multi-channel signal (111) comprising N channels; wherein the upmixing metadata (105) is determined in dependance of whether the current frame is an active frame or an inactive frame; and
- 25 – encode the upmixing metadata (105) into a bitstream.

41) A decoding unit (150) for decoding a bitstream which is indicative of a reconstructed multi-channel signal (111) comprising N channels; wherein the reconstructed signal (111) comprises a sequence of frames; wherein for a current frame of the sequence of frames,
30 the decoding unit (150) is configured to

- determine, based on the bitstream, whether the current frame is an active frame or an inactive frame;

- determine a reconstructed downmix signal (114); wherein the reconstructed downmix signal (114) comprises less than or equal to N channels; and
- generate the reconstructed multi-channel signal (111) based on the reconstructed downmix signal (114) and based on upmixing metadata (105) comprised within the bitstream; wherein the reconstructed multi-channel signal (111) is generated in dependance of whether the current frame is an active frame or an inactive frame.

42) A non-transitory computer-readable medium storing instructions that, when executed by one or more processors, cause the one or more processors to perform operations of any one of claims 1 to 39.

15