MULTI-CHANNEL AUDIO DECODER, MULTI-CHANNEL AUDIO ENCODER, METHODS AND COMPUTER PROGRAM USING A RESIDUAL-SIGNAL-BASED ADJUSTMENT OF A CONTRIBUTION OF A DECORRELATED SIGNAL

TECHNICAL FIELD

An embodiment according to the invention is related to a multi-channel audio decoder for providing at least two output audio signals on the basis of an encoded representation.

Another embodiment according to the invention is related to a multi-channel audio encoder for providing an encoded representation of a multi-channel audio signal.

Another embodiment according to the invention is related to a method for providing at least two output audio signals on the basis of an encoded representation.

Another embodiment according to the invention is related to a method for providing an encoded representation of a multi-channel audio signal.

Another embodiment according to the present invention is related to a computer program for performing one of the methods.

Generally, some embodiments according to the invention are related to a combined residual and parametric coding.

BACKGROUND OF THE INVENTION

In recent years, demand for storage and transmission of audio content has been steadily increasing. Moreover, the quality requirements for the storage and transmission of audio contents have also been increasing steadily. Accordingly, the concepts for the encoding and decoding of audio content have been enhanced. For example, the so-called “advanced audio coding” (AAC) has been developed, which is described, for example, in

the international standard ISO/IEC 13818-7: 2003.

Moreover, some spatial extensions have been created, like, for example, the so-called "MPEG surround" concept, which is described, for example, in the international standard ISO/IEC 23003-1:2007. Moreover additional improvements for the encoding and decoding of a spatial information of audio signals are described in the international standard ISO/IEC 23003-2:2010, which relates to the so-called spatial audio object coding. Moreover, a flexible (switchable) audio encoding/decoding concept, which provides the possibility to encode both general audio signals and speech signals with good coding efficiency and to handle multi-channel audio signals is defined in the international standard ISO/IEC 23003-3:2012, which describes the so-called "unified speech and audio coding" concept.

However, there is a desire to provide an even more advanced concept for an efficient encoding and decoding of multi-channel audio signals.

SUMMARY OF THE INVENTION

An embodiment according to the invention creates a multi-channel audio decoder for providing at least two output audio signals on the basis of an encoded representation. The multi-channel audio decoder is configured to obtain (at least) one of the output audio signals on the basis of an encoded representation of a downmix signal, a plurality of encoded spatial parameters and an encoded representation of a residual signal. The multi-channel audio decoder is configured to blend between a parametric coding and the residual coding in dependence on the residual signal. Accordingly, an intensity of the residual signal determines whether the decoding is mostly based on the spatial parameters in addition to the downmix signal, or the decoding is mostly based on the residual signal in addition to the downmix signal, or whether an intermediate state is taken in which both the spatial parameters and the residual signal affect a refinement of the output signal, to derive the output audio signals from the downmix signal. Thus, a very flexible audio decoding concept is achieved, wherein the best decoding mode (parametric coding and decoding versus residual coding and decoding) can be selected without additional signaling overhead. Moreover, the above explained consideration is also applied.

Another embodiment according to the invention creates a method for providing at least two output audio signals on the basis of an encoded representation. The method comprises obtaining (at least) one of the output audio signals on the basis of an encoded representation of a downmix signal, a plurality of encoded spatial parameters and an encoded representation of a residual signal. A blending (or fading) is performed between a parametric coding and a residual coding in dependence on the residual signal such that an intensity of the residual signal determines whether the decoding is mostly based on the spatial parameters in addition to the downmix signal, or whether the decoding is mostly based on the residual signal in addition to the downmix signal, or whether an intermediate state is taken in which both the spatial parameters and the residual signal affect a refinement of the output signal, to derive the output audio signals from the downmix signal. This method is also based on the same considerations as the above described audio decoder.

Further embodiments, according to the invention create computer programs for performing the methods described herein.

BRIEF DESCRIPTION OF THE FIGURES

Embodiments according the invention will subsequently be described taking reference to the enclosed figures, in which

Figure 1 shows a block schematic diagram of a multi-channel audio encoder, according to an embodiment of the invention;

Figure 2 shows a block schematic diagram of a multi-channel audio decoder, according to an embodiment of the invention;

Figure 3 shows a block schematic diagram of a multi-channel audio decoder, according to a another embodiment of the present invention;

Figure 4 shows a flow chart of a method for providing an encoded representation of a multi-channel audio signal, according to an embodiment of the invention;

Figure 5 shows a flow chart of a method for providing at least two output audio signals on the basis of an encoded representation, according to an embodiment of the invention;

Figure 6 shows a flow chart of a method for providing at least two output audio signals on the basis of an encoded representation, according to another embodiment of the invention; and

Figure 7 shows a flow diagram of a decoder, according to an embodiment of the present invention; and

Figure 8 shows a schematic representation of a Hybrid Residual Decoder.

DETAILED DESCRIPTION OF THE EMBODIMENTS

1. Multi-channel audio encoder according to figure 1

Figure 1 shows a block schematic diagram of a multi-channel audio encoder 100 for providing an encoded representation of a multi-channel signal.

The multi-channel audio encoder 100 is configured to receive a multi-channel audio signal 110 and to provide, on the basis theirs, an encoded representation 112 of the multi-channel audio signal 110. The multi-channel audio encoder 100 comprises a processor (or processing device) 120, which is configured to receive the multi-channel audio signal and to obtain a downmix signal 122 on the basis of the multi-channel audio signal 110. The processor 120 is further configured to provide parameters 124 describing dependencies between the channels of the multi-channel audio signal 110. Moreover, the processor 120 is configured to provide a residual signal 126. Furthermore, the multi-channel audio encoder comprises a residual signal processing 130, which is configured to vary an amount of residual signal included into the encoded representation 112 in dependence on the multi-channel audio signal 110.

However, it should be noted, that it is not necessary that the multi-channel audio decoder comprises a separate processor 120 and a separate residual signal processing 130. Rather, it is sufficient if the multi-channel audio encoder is somehow configured to perform the functionality of the processor 120 and of the residual signal processing 130.

Regarding the functionality of the multi-channel audio encoder 100, it can be noted that the channel signals of the multi-channel audio signal 110 are typically encoded using a multi-channel encoding, wherein the encoded representation 112 typically comprises (in an encoded form) the downmix signal 122, the parameters 124 describing dependencies between channels (or channel signals) of the multi-channel audio signal 110 and the residual signal 126. The downmix signal 122 may, for example, be based on a combination (for example, linear combination) of the channel signals of the multi-channel audio signal. However a signal downmix signal 122 may provided on the basis of a plurality of channel signals of the multi-channel audio signal. However, alternatively, two or more downmix signal may be associated with a larger number (typically larger than the number of downmix signals) of channel signals of the multi-channel audio signal 110. The parameters 124 may describe dependencies (for example, a correlation, a covariance, a level relationship or the like) between channels (or channel signals) of the multi-channel audio signal 110. Accordingly, the parameters 124 serve the purpose to derive a reconstructed version of the channel signals of the multi-channel audio signal 110 on the basis of the downmix signal 122 at the side of an audio decoder. For this purpose, the parameters 124 describe desired characteristics (for example, individual characteristics or relative characteristics) of the channel signals of the multi-channel audio signal, such that an audio encoder, which uses a parametric decoding, can reconstruct channel signals on the basis of the one or more downmix signals 122.

In addition, the multi-channel audio decoder 100 provides the residual signal 126, which typically represents signal components that, according to the expectation or estimation of the multi-channel audio encoder, cannot be reconstructed by an audio decoder (for example, by an audio decoder following a certain processing rule) on the basis of the downmix signal 122 and the parameters 124. Accordingly, the residual signal 126 can typically be considered as a refinement signal, which allows for a wave from reconstruction, or at least for a partial wave from reconstruction, at the side of an audio decoder.

However, the multi-channel audio encoder 100 is configured to vary an amount of residual signal included into the encoded representation 112 in dependence on the multi-channel audio signal 110. In other words, the multi-channel audio encoder may, for example, decide about the intensity (or the energy) of the residual signal 126 which is included into the encoded representation 112. Additionally or alternatively, the multi-channel audio encoder 100 may decide, for which frequency bands and/or for how many frequency bands the residual signal is included into the encoded representation 112. By varying the "amount" of residual signal 126 included into the encoded representation 112 in dependence on the multi-channel audio signal (and/or in dependence on an available bitrate), the multi-channel audio encoder 100 can flexibly determine with which accuracy the channel signals of the multi-channel audio signal 110 can be reconstructed at the side of an audio decoder on the basis of the encoded representation 112. Thus, the accuracy with which the channel signals of the multi-channel audio signal 110 can be reconstructed, can be adapted to a psychoacoustic relevance of different signal portions of the channel signals of the multi-channel audio signal 110 (like, for example, temporal portions, frequency portions and/or time/frequency portions). Thus, signal portions of high psychoacoustic relevance (like, for example, tonal signal portions or signal portions comprising transient events can be encoded with particularly high resolution by including a "large amount" of the residual signal 126 into the encoded representation. For example, it can be achieved that a residual signal with a comparatively high energy is included in the encoded representation 112 for signal portions of high psychoacoustic relevance. Moreover, it can be achieved that a residual signal of high energy is included in the encoded representation 112 if the downmix signal 122 comprises a "poor quality", for example, if there is a substantial cancellation of signal components when combining the channel signals of the multi-channel audio signal 112 into the downmix signal 122. In other words, the multi-channel audio decoder 100 can selectively embed a "larger amount" of residual signal (for example, a residual signal having a comparatively high energy) into the encoded representation 112 for signal portions of the multi-channel audio signal 110 for which the provision of a comparatively large amount of the residual signal brings along a significant improvement of the reconstructed channel signals (reconstructed at the side of an audio decoder).

Accordingly, the variation of the amount of residual signal included in the encoded

representation in dependence on the multi-channel audio signal 110 allows to adapt the encoded representation 112 (for example, the residual signal 126, which is included into the encoded representation in an encoded form) of the multi-channel audio signal 110, such that a good trade off between bitrate efficiency and audio quality of the reconstructed multi-channel audio signal (reconstructed at the side of an audio decoder) can be achieved.

It should be noted, that the multi-channel audio encoder 100 can be optionally improved in many different ways. For example the multi-channel audio encoder may be configured to vary a bandwidth of the residual signal 126 (which is included into the encoded representation) in dependence on the multi-channel audio signal 110. Accordingly, the amount of residual signal included into the encoded representation 112 may be adapted to perceptually most important frequency bands.

Optionally, the multi-channel audio decoder may be configured to select frequency bands for which the residual signal 126 is included into the encoded representation 112 in dependence on the multi-channel audio signal 110. Accordingly, the encoded representation 120 (more precisely, the amount of residual signal included into the encoded representation 112) may be adapted to the multi-channel audio signal, for example, to the perceptually most important frequency bands of the multi-channel audio signal 110.

Optionally, the multi-channel audio encoder may be configured to including the residual signal 126 into the encoded representation for frequency bands for which the multi-channel audio signal is tonal. In addition, the multi-channel audio encoder may be configured to not include the residual signal 126 into the encoded representation 112 for frequency bands in which the multi-channel audio signal is non-tonal (unless any other specific condition is fulfilled which causes an inclusion of the residual signal into the encoded representation for a specific frequency band). Thus, the residual signal may be selectively included into the encoded representation for perceptually important tonal frequency bands.

Optionally, the multi-channel audio encoder 100 may be configured to selectively include the residual signal into the encoded representation for time portions and/or for frequency

bands in which the formation of the downmix signal results in a cancellation of signal components of the multi-channel audio signal. For example, the multi-channel audio encoder may be configured to detect a cancellation of signal components of the multi-channel audio signal 110 in the downmix signal 122, and to activate the provision of the residual signal 126 (for example, the inclusion of the residual signal 126 into the encoded representation 112) in response to the result of the detection. Accordingly, if the downmixing (or any other typically linear combination) of channel signals of the multi-channel audio signal 110 into the downmix signal 122 results in a cancellation of signal components of the multi-channel audio signal 112 (which may be caused, for example, by signal components of different channel signals which are phase-shifted by 180 degrees), the residual signal 126, which helps to overcome the detrimental effect of this cancellation when reconstructing the multi-channel audio signal 110 in an audio decoder, will be included into the encoded representation 112. For example, the residual signal 126 may be selectively included in the encoded representation 112 for frequency bands for which there is such a cancellation.

Optionally, the multi-channel audio encoder may be configured to compute the residual signal using a linear combination of at least two channel signals of the multi-channel audio signal and in dependence on upmix coefficients to be used at the side of a multi-channel audio decoder. Such a computation of a residual signal is efficient and allows for a simple reconstruction of the channel signals at the side of an audio decoder.

Optionally, the multi-channel audio encoder may be configured to encode the upmix coefficients using the parameter 124 describing dependencies between the channels of the multi-channel audio signal, or to derive the upmix coefficients from the parameters describing dependencies between the channels of the multi-channel audio signal. Accordingly, the parameters 124 (which may, for example, be intra-channel level difference parameters, intra-channel correlation parameters, or the like) may be used both for the parametric coding (encoding or decoding) and for the residual signal-assisted coding (encoding or decoding). Thus, the usage of the residual signal 126 does not bring along an additional signaling overhead. Rather, the parameters 124, which are used for the parametric coding (encoding/decoding) anyway, are re-used also for the residual coding (encoding/decoding). Thus high coding efficiency can be achieved.

Optionally, the multi-channel audio decoder may be configured to time-variantly determine the amount of residual signal included into the encoded representation using a psychoacoustic model. Accordingly, the encoding precision can be adapted to psychoacoustic characteristics of the signal, which typically results in a good bitrate efficiency.

However, it should be noted, that the multi-channel audio encoder can optionally be supplemented by any of the features or functionalities described herein (both in the description and in the claims). Moreover, the multi-channel audio encoder can also be adapted in parallel with the audio decoder described herein, to cooperate with the audio decoder.

2. Multi-channel audio decoder according to figure 2

Figure 2 shows a block schematic diagram of a multi-channel audio decoder 200 according to an embodiment of the present invention.

The multi-channel audio decoder 200 is configured to receive an encoded representation 210 and to provide, on the basis thereof, at least two output audio signals 212, 214. The multi-channel audio decoder 200 may, for example, comprise a weighting combiner 220, which is configured to perform a weighted combination of a downmix signal 222, a decorrelated signal 224 and a residual signal 226, to obtain (at least) one of the output signals, for example, the first output audio signal 212. It should be noted here, that the downmix signal 212, the decorrelated signal 224 and the residual signal 226 may, for example, be derived from the encoded representation 210, wherein the encoded representation 210 may carry an encoded representation of the downmix signal 220 and an encoded representation of the residual signal 226. Moreover, the decorrelated signal 224 may, for example, be derived from the downmix signal 222 or may be derived using additional information included in the encoded representation 210. However, the decorrelated signal may also be provided without any dedicated information from the encoded representation 210.

The multi-channel audio decoder 200 is also configured to determine a weight describing a contribution of the decorrelated signal 224 in the weighted combination in dependence

on the residual signal 226. For example, the multi-channel audio decoder 200 may comprise a weight determinator 230, which is configured to determine a weight 232 describing the contribution of the decorrelated signal 224 in the weighted combination (for example, the contribution of the decorrelated signal 224 to the first output audio signal 212) on the basis of the residual signal 226.

Regarding the functionality of the multi-channel audio decoder 200, it should be noted, that the contribution of the decorrelated signal 224 to the weighted combination, and consequently to the first output audio signal 212, is adjusted in a flexible (for example, temporally variable and frequency-dependent) manner in dependence on the residual signal 226, without additional signaling overhead. Accordingly, the amount of decorrelated signal 224, which is included into the first output audio signal 212, is adapted in dependence on the amount of residual signal 226 which is included into the first output audio signal 212, such that a good quality of the first output audio signal 212 is achieved. Accordingly, it is possible to obtain an appropriate weighting of the decorrelated signal 224 under any circumstances and without an additional signaling overhead. Thus, using the multi-channel audio decoder 200, a good quality of the decoded output audio signal 212 can be achieved with moderate bitrate. A precision of the reconstruction can be flexibly adjusted by an audio encoder, wherein the audio encoder can determine an amount of residual signal 226 which is included in the encoded representation 212 (for example, how big the energy of the residual signal 226 included in the encoded representation 210 is, or to how many frequency bands the residual signal 226 included in the encoded representation 210 relates), and the multi-channel audio decoder 200 can react accordingly and adjust the weighting of the decorrelated signal 224 to fit the amount of residual signal 226 included in the encoded representation 210. Consequently, if there is a large amount of residual signal 226 included in the encoded representation 210 (for example, for a specific frequency band, or for specific temporal portion), the weighted combination 220 may predominantly (or exclusively) consider the residual signal 226 while giving little weight (or no weight) to the decorrelated signal 224. In contrast, if there is only a smaller amount of a residual signal 226 included in the encoded representation 210, the weighted combination 220 may predominantly (or exclusively) consider the decorrelated signal 224 but only to a comparatively small degree (or not at all) the residual signal 226 in addition to the downmix signal 222. Thus, the multi-channel audio decoder 200 can flexible cooperate with an appropriate multi-channel audio encoder and adjust the

weighted combination 220 to achieve the best possible audio quality under any circumstances (irrespective of whether a smaller amount or a larger amount of residual signal 226 is included in the encoded representation 210).

It should be noted, that the second output audio signal 214 may be generated in a similar manner. However, it is not necessary to apply the same mechanisms to the second output audio signal 214, for example, if there are different quality requirements with respect to the second output audio signal.

In an optional improvement, the multi-channel audio decoder may be configured to determine the weight 232 describing the contribution of the decorrelated signal 224 in the weighted combination in dependence on the decorrelated signal 224. In other words, the weight 232 may be dependent both on the residual signal 226 and the decorrelated signal 224. Accordingly, the weight 232 may be even better adapted to a currently decoded audio signal without additional signaling overhead.

As another optional improvement, the multi-channel audio decoder may be configured to obtain upmix parameters on the basis of the encoded representation 212 and to determine the weight 232 describing the contribution of the decorrelated signal in the weighted combination in dependence on the upmix parameters. Accordingly, the weight 232 may be additionally dependent on the upmix parameters, such that an even better adaptation of the weight 232 can be achieved.

As another optional improvement, the multi-channel audio decoder may be configured to determine the weight describing the contribution of the decorrelated signal in the weighted combination such that the weight of the decorrelated signal decreases with increasing energy of the residual signal. Accordingly, a blending or fading can be performed between a decoding which is predominantly based on the decorrelated signal 224 (in addition to a downmix signal 222) and a decoding which is predominantly based on the residual signal 226 (in addition to a downmix signal 222).

As another optional improvement, the multi-channel audio decoder 200 may be configured to determine the weight 232 such that a maximum weight, which is determined by a decorrelated signal upmix parameter (which may be included in, or derived from, the

encoded representation 210) is associated to the decorrelated signal 224 if an energy of the residual signal 226 is zero, and that such that a zero weight is associated to the decorrelated signal 224 if an energy of the residual signal 226, weighted with the residual signal weighting coefficient (or a residual signal upmix parameter), is larger than or equal to an energy of the decorrelated signal 224, weighted with the decorrelated signal upmix parameter. Accordingly, it is possible to completely blend (or fade) between a decoding based on the decorrelated signal 224 and a decoding based on the residual signal 226. If the residual signal 226 is judged to be strong enough (for example, when the energy of the weighted residual signal is equal to or larger than the energy of the weighted decorrelated signal 224), the weighted combination may fully rely on the residual signal 226 to refine the downmix signal 222 while leaving the decorrelated signal 224 out of consideration. In this case, a particularly good (at least partial) wave form reconstruction at the side of the multi-channel audio decoder 200 can be performed, since the consideration of the decorrelated signal 224 typically prevents a particularly good wave form reconstruction while the usage of the residual signal 226 typically allows for a good wave form reconstruction.

In another optional improvement, the multi-channel audio decoder 200 may be configured to compute a weighted energy value of a decorrelated signal, weighted in dependence on one or more decorrelated signal upmix parameters, and to compute a weighted energy value of the residual signal, weighted using one or more residual signal upmix parameters. In this case, the multi-channel audio decoder may be configured to determine a factor in dependence on the weighted energy value of the decorrelated signal and the weighted energy value of the residual signal and to obtain a weight describing the contribution of the decorrelated signal 224 to one of the output audio signals (for example, the first output audio signal 212) on the basis of the factor. Thus, the weight determination 230 may provide particularly well-adapted weighting values 232.

In an optional improvement, the multi-channel audio decoder 200 (or the weight determinator 230 thereof) may be configured to multiply the factor with the decorrelated signal upmix parameter (which may be included in the encoded representation 210, or derived from the encoded representation 210), to obtain the weight (or weighting value) 232 describing the contribution of the decorrelated signal 224 to one of the output audio signals (for example the first output audio signal 212).

In an optional improvement, the multi-channel audio decoder (or the weight determinator 230 thereof) may be configured to compute the energy of the decorrelated signal 224, weighted using decorrelated signal upmix parameters (which may be included in the encoded representation 210, or which may be derived from the encoded representation 210), over a plurality of upmix channels and time slots, to obtain the weighted energy value of the decorrelated signal.

As a further optional improvement, the multi-channel audio decoder 200 may be configured to compute the energy of the residual signal 224, weighted using residual signal upmix parameters (which may be included in the encoded representation 210 or which may be derived from the encoded representation 210) over a plurality of upmix channels and time slots, to obtain the weighted energy value of the residual signal.

As another optional improvement, the multi-channel audio decoder 200 (or the weight determinator 232 thereof) may be configured to compute the factor mentioned above in dependence on a difference between the weighted energy value of the decorrelated signal and the weighted energy value of the residual signal. It has been found, that such computation is an efficient solution to determine the weighting values 232.

As an optional improvement, the multi-channel audio decoder may be configured to compute the factor in dependence on a ratio between a difference between the weighted energy value of the decorrelated signal 224 and the weighted energy value of the residual signal 226, and the weighted energy value of the decorrelated signal 224. It has been found, that such a computation for the factor brings along good results for blending between a predominantly decorrelation signal based refinement of the downmix signal 222 and a predominantly residual signal based refinement of the downmix signal 222.

As an optional improvement, the multi-channel audio decoder 200 may be configured to determine weights describing contributions of the decorrelated signals to two or more output audio signals, like, for example, the first output audio signal 212 and the second output audio signal 214. In this case, the multi-channel audio decoder may be configured to determine a contribution of the decorrelated signal 224 to the first output audio signal 212 on the basis of the weighted energy value of the decorrelated signal 224 and a first-

channel decorrelated signal upmix parameter. Moreover, the multi-channel audio decoder may be configured to determine a contribution of the decorrelated signal 224 to the second output audio signal 214 on the basis of the weighted energy value of the decorrelated signal 224 and a second-channel decorrelated signal upmix parameter. In other words, different decorrelated signal upmix parameters may be used for providing the first output audio signal 212 and the second output audio signal 214. However, the same weighted energy value of the decorrelated signal may be used for determining the contribution of the decorrelated signal to the first output audio signal 212 and the contribution of the decorrelated signal to the second output audio signal 214. Thus, an efficient adjustment is possible, wherein nevertheless different characteristics of the two output audio signals 212, 214 can be considered by different decorrelated signal upmix parameters.

As an optional improvement, the multi-channel audio decoder 200 may be configured to disable a contribution of the decorrelated signal 224 to the weighted combination if a residual energy (for example, an energy of the residual signal 226 or of a weighted version of the residual signal 226) exceeds a decorrelated energy (for example, an energy of the decorrelated signal 224 or of a weighted version of the decorrelated signal 224).

As a further optional improvement, the audio decoder may be configured to band-wisely determine the weight 232 describing a contribution of the decorrelated signal 224 in the weighted combination in dependence on a band-wise determination of a weighted energy value of the residual signal. Accordingly a fine-tuned adjustment of the multi-channel audio decoder 200 to the signals to be decoded can be performed.

In another optional improvement, the audio decoder may be configured to determine the weight describing a contribution of the decorrelated signal in the weighted combination for each frame of the output audio signal 212, 214. Accordingly, a good temporal resolution can be achieved.

In a further optional improvement, the determination of the weighting value 232 may be performed in accordance with some of the equations provided below.

Moreover, it should be noted, that the multi-channel audio decoder 200 can be

supplemented by any of the features or functionalities described herein, also with respect to other embodiments.

3. Multi-channel audio decoder according to figure 3

Figure 3 shows a block schematic diagram of a multi-channel audio decoder 300 according to an embodiment of the invention. The multi-channel audio decoder 300 is configured to receive an encoded representation 310 and to provide, on the basis thereof, two or more output audio signals 312, 314. The encoded representation 310 may, for example, comprise an encoded representation of a downmix signal, an encoded representation of one or more spatial parameters and an encoded representation of a residual signal. The multi-channel audio decoder 300 is configured to obtain (at least) one of the output audio signals, for example, a first output audio signal 312 and/or a second output audio signal 314, on the basis of the encoded representation of the downmix signal, a plurality of encoded spatial parameters and an encoded representation of the residual signal.

In particular, the multi-channel audio decoder 300 is configured to blend between a parametric coding and a residual coding in dependence on the residual signal (which is included, in an encoded form, in the encoded representation 310). In other words, the multi-channel audio decoder 300 may blend between a decoding mode in which the provision of the output audio signals 312, 314 is performed on the basis of the downmix signal and using spatial parameters which describe a desired relationship between the output audio signals 312, 314 (for example, a desired inter-channel level difference or a desired inter-channel correlation of the output audio signals 312, 314), and a decoding mode in which the output audio signals 312, 314 are reconstructed on the basis of the downmix signal using the residual signal. Thus, the intensity (for example, energy) of the residual signal, which is included in the encoded representation 310, may determine whether the decoding is mostly (or exclusively) based on the spatial parameters (in addition to the downmix signal) or whether the decoding is mostly (or exclusively) based on the residual signal (in addition to the downmix signal), or whether an intermediate state is taken in which both the spatial parameters and the residual signal affect the refinement of the downmix signal, to derive the output audio signals 312, 314 from the downmix

signal.

Moreover, the multi-channel audio decoder 300 allows for a decoding which is well-adapted to the current audio content without high signaling overhead by blending between the parametric coding, (in which, typically, a comparatively high weight is given to a decorrelated signal when providing the output audio signals 312, 314) and a residual coding (in which, typically, a comparatively small weight is given to a decorrelated signal) in dependence on the residual signal.

Moreover, it should be noted, that the multi-channel audio decoder 300 is based on similar considerations as the multi-channel audio decoder 200 and that optional improvements described above with respect to the multi-channel audio decoder 200 can also be applied to the multi-channel audio decoder 300.

4. Method for providing an encoded representation of a multi-channel audio signal according to figure 4

Figure 4 shows a flow chart of a method 400 for providing an encoded representation of a multi-channel audio signal.

The method 400 comprises a step 410 of obtaining a downmix signal on the basis of a multi-channel audio signal. The method 400 also comprises a step 420 of providing parameters describing dependencies between the channels of the multi-channel audio signal. For example, inter-channel-level-difference parameters and/or inter-channel correlation parameters (or covariance parameters) may be provided, which describe dependencies between channels of the multi-channel audio signal. The method 400 also comprises a step 430 of providing a residual signal. Moreover, the method comprises a step 440 of a varying an amount of residual signal included into the encoded representation in dependence on the multi-channel audio signal.

It should be noted, that the method 400 is based on the same considerations as the audio encoder 100 according to figure 1. Moreover, the method 400 can be supplemented by any of the features and functionalities described herein with respect to the inventive

apparatuses.

5. Method for providing at least two output audio signals on the basis of an encoded representation according to figure 5.

Figure 5 shows a flow chart of a method 500 for providing at least two output audio signals on the basis of an encoded representation. The method 500 comprises determining 510 a weight describing a contribution of a decorrelated signal in a weighted combination in dependence on a residual signal. The method 500 also comprises performing 520 a weighted combination of a downmix signal, a decorrelated signal and a residual signal, to obtain one of the output audio signals.

It should be noted, that the method 500 can be supplemented by any of the features and functionalities described herein with respect to the inventive apparatuses.

6. Method for providing at least two output audio signals on the basis of an encoded representation according to figure 6.

Figure 6 shows a flow chart of a method 600 for providing at least two output audio signals on the basis of an encoded representation. The method 600 comprises obtaining 610 one of the output audio signals on the basis of an encoded representation of a downmix signal, a plurality of encoded spatial parameters and an encoded representation of a residual signal. Obtaining 610 one of the output audio signals comprises performing 620 a blending between a parametric coding and a residual coding in dependence on the residual signal.

It should be noted, that the method 600 can be supplemented by any of the features and functionalities described herein with respect to the inventive apparatuses**.**

7. Further embodiments

In the following, some general considerations and some further embodiments will be described.

7.1 General considerations

Embodiments according to the invention are based on the idea that, instead of using a fixed residual bandwidth, a decoder (for example, a multi-channel audio decoder) detects the amount of transmitted residual signal by measuring its energy band-wise for each frame (or, generally, at least for a plurality of frequency ranges and/or for a plurality of temporal portions). Depending on the transmitted spatial parameters, a decorrelated output is added where residual energy “is missing”, to achieve a required (or desired) amount of output energy and decorrelation. This allows a variable residual bandwidth as well as band pass-style residual signals. For example, it is possible to only use residual coding for tonal bands. To be able to use the simplified downmix for parametric coding as well as for wave form-preserving coding (which is also designated as residual coding), a residual signal for the simplified downmix is defined herein.

7.2 Calculation of the residual signal for the simplified downmix

In the following, some considerations regarding the calculation of the residual signal and regarding the construction of channel signals of a multi-channel audio signal will be described.

In unified-speech- and audio-coding (USAC), there is no residual signal defined when a so-called “simplified downmix” is used. Thus, no partially waveform preserving coding is possible. However, in the following, a method for a calculating a residual signal for the so-called “simplified downmix” will be described.

“Simplified downmix” weights d_1, d_2 are calculated per scale factor band, whereas parametric upmix coefficients u_{d1}, u_{d2} are calculated per parameter band. Thus, coefficients $w_{r1,}$ w_{r2}, for calculating the residual signal cannot be directly computed from the spatial parameters (as it is the case for a classic MPEG surround), but may need to be determined scale factor band-wise from the down- and upmix coefficients.

With L, R being the input channels and D being the downmix channel, a residual signal res should fulfill the following properties:

$$D = d_1 L + d_2 R \tag{1}$$
$$L = u_{d,1} D + u_{r,1} \mathrm{res} \tag{2}$$
$$R = u_{d,2} D + u_{r,2} \mathrm{res} \tag{3}$$

This is achieved by calculating the residual as

$$\mathrm{res} = w_{r,1} L + w_{r,2} R \tag{4}$$

using the downmix weights

$$w_{r,1} = \frac{1}{2}\left(\frac{1 - u_{d,1} d_1}{u_{r,1}} - \frac{u_{d,2} d_1}{u_{r,2}}\right) \tag{5}$$
$$w_{r,2} = \frac{1}{2}\left(\frac{1 - u_{d,2} d_2}{u_{r,2}} - \frac{u_{d,1} d_2}{u_{r,1}}\right). \tag{6}$$

The residual upmix coefficients $u_{r,1}$, $u_{r,2}$ used by the decoder are preferably chosen in a way to ensure robust decoding. Since the simplified downmix has asymmetric properties (as opposed to MPEG Surround with fixed weights) an upmix depending on the spatial parameters is applied, e.g. using the following upmix coefficients:

$$u_{r,1} = \max\{u_{d,1}, 0.5\} \tag{7}$$
$$u_{r,2} = -\max\{u_{d,2}, 0.5\} \tag{8}$$

Another option is to define the residual upmix coefficients to be orthogonal to the downmix signal's upmix coefficients, so that:

$$\left\langle \begin{pmatrix} u_{d,1} \\ u_{d,2} \end{pmatrix}, \begin{pmatrix} u_{r,1} \\ u_{r,2} \end{pmatrix} \right\rangle \stackrel{!}{=} 0 \tag{9}$$

In other words, an audio decoder may obtain the downmix signal D using a linear combination of a left channel signal L (first channel signal) and a right channel signal R (second channel signal). Similarly, the residual signal res is obtained using a linear combination of the left channel L and the right channel signal R (or, generally, of a first

channel signal and a second channel signal of the multi-channel audio signal).

It can be seen, for example, in Equations (5) and (6), the downmix weights $w_{r,1}$ and $w_{r,2}$ for obtaining the residual signal res can be obtained when the simplified downmix weights d_1, d_2, the parametric upmix coefficients $u_{d,1}$ and $u_{d,2}$ and the residual upmix coefficients $u_{r,1}$ and $u_{r,2}$ are determined. Moreover it can be seen, that $u_{r,1}$ and $u_{r,2}$ can be derived from $u_{d,1}$ and $u_{d,2}$ using equations (7) and (8) or equation (9). The simplified downmix weights d_1 and d_2, as well as the parametric upmix coefficients $u_{d,1}$ and $u_{d,2}$ can be obtained in the usual manner.

7.3 Encoding process

In the following, some details regarding the encoding process will be described. The encoding may, for example, be performed by the multi-channel audio encoder 100 or by any other appropriate means or computer programs.

Preferably, the amount of a residual that is transmitted is determined by a psychoacoustic model of the encoder (for example, multi-channel audio encoder), depending on the audio signal (for example, depending on the channel signals of the multi-channel audio signal 110) and an available bitrate. The transmitted residual signal can, for example, be used for partial wave form preservation or to avoid signal cancellation caused by the used downmixing method (for example, the downmixing method described by equation (1) above).

7.3.1 Partial wave form preservation

In the following, it is described how a partial wave form preservation can be achieved. For example, the calculated residual (for example, the residual res according to equation (4)) is transmitted full-band or band-limited to provide partial wave form preservation within the residual bandwidth. Residual parts, which are detected as perceptually irrelevant by the psychoacoustic model may, for example, be quantized to zero (for example, when providing the encoded representation 112 on the basis of the residual signal 126). This includes, but is not limited to, reducing the transmitted residual bandwidth at runtime (which may be considered as varying an amount of residual signal which is included into

the encoded representation). This system may also allow band-pass-style deletion of residual signal parts, as missing signal energy will be reconstructed by the decoder (for example, by the multi-channel audio decoder 200 or the multi-channel audio decoder 300). Thus, for example, residual coding may be only applied to tonal components of the signal, preserving their phase-relations, whereas background noise can be parametrically coded to reduce the residual bitrate. In other words, the residual signal 126 may only be included into the encoded representation 112 (for example, by the residual signal processing 130) for frequency bands and/or temporal portions for which the multi-channel audio signal 110 (or at least one of the channel signals of the multi-channel audio signal 110) are found to be tonal. In contrast, the residual signal 126 may not be included into the encoded representation 112 for frequency bands and/or temporal portions for which the multi-channel audio signal 110 (or at least one or more channel signals of the multi-channel audio signal 110) are identified as being noise-like. Thus, an amount of residual signal included into the encoded representation is varied in dependence on the multi-channel audio signal.

7.3.2 Prevention of signal cancellation in downmix

In the following, it will be described how a signal cancellation in the downmix can be prevented (or compensated).

For low bitrate applications, parametric coding (which predominantly or exclusively relies on the parameters 124, describing dependencies between channels of the multi-channel audio signal) instead of wave form preserving coding (which, for example, predominantly relies on the residual signal 126, in addition to the downmix signal 122) is applied. Here, the residual signal 126 is only used to compensate for signal cancellations in the downmix 122, to minimize the bit usage of the residual. As long as no signal cancellations in the downmix 122 are detected, the system runs in parametric mode using decorrelators (at the side of the audio decoder). When signal cancellations occur, for example, for phasing tonal signals, a residual signal 126 is transmitted for the impaired signal parts (for example, frequency bands and/or temporal portions). Thus, the signal energy can be restored by the decoder.

7.4 Decoding process

7.4.1 Overview

In the decoder (for example, in the multi-channel audio decoder 200 or in the multi-channel audio decoder 300), the transmitted downmix and residual signals (for example, downmix signal 222 or residual signal 226) are decoded by a core decoder and fed into an MPEG surround decoder together with the decoded MPEG surround payload. Residual upmix coefficients for the classic MPS downmix are unchanged, and residual upmix coefficient for the simplified downmix are defined in equations (7) and (8) and/or (9). Additionally, decorrelator outputs and its weighting coefficients are calculated, as for parametric decoding. The residual signal and the decorrelator outputs are weighted and both mixed to the output signal. Therefore, weighting factors are determined by measuring the energies of the residual and decorrelator signals.

In other words, residual upmix factors (or coefficients) may be determined by measuring the energies of the residual and decorrelated signals.

For example, the downmix signal 222 is provided on the basis of the encoded representation 210, and the decorrelated signal 224 is derived from the downmix signal 222 or generated on the basis of parameters included in the encoded representation 210 (or otherwise). The residual upmix coefficients may, for example be derived from the parametric upmix coefficients $u_{d,1}$ and $u_{d,2}$ in accordance with equations (7) and (8) by the decoder, wherein the parametric upmix coefficients $u_{d,1}$ $u_{d,2}$ may be obtained on the basis of the encoded representation 210, for example, directly or by deriving them from spatial data included in the encoded representation 210 (for example, from inter-channel correlation coefficients and inter-channel level difference coefficients, or from inter-object correlation coefficients and inter-object level differences).

Upmixing coefficients for the decorrelator output (or outputs) may be obtained as for conventional MPEG surround decoding. However, weighting factors for weighting the decorrelator output (or decorrelator outputs) may be determined on the basis of the energies of the residual signal (and possibly also on the basis of the energies of the decorrelator signal or signals) such that a weight describing a contribution of the

decorrelated signal in the weighted combination is determined in dependence on the residual signal.

7.4.2 Example Implementation

In the following, an example implementation will be described taking reference to figure 7. However, it should be noted, that the concept described herein can also be applied in the multi-channel audio decoders 200 or 300 according to figures 2 and 3.

Figure 7 shows a block schematic diagram (or flow diagram) of a decoder (for example, of a multi-channel audio decoder). The decoder according to figure 7 is designated with 700 in its entirety**.** The decoder 700 is configured to receive a bit stream 710 and to provide, on the basis thereof, a first output channel signal 712 and a second output channel signal 714. The decoder 700 comprises a core decoder 720, which is configured to receive the bit stream 710 and to provide, on the basis thereof, a downmix signal 722, a residual signal 724 and spatial data 726. For example, the core decoder 720 may provide, as the downmix signal, a time domain representation or transform domain representation (for example, frequency domain representation, MDCT domain representation, QMF domain representation) of the downmix signal represented by the bit stream 710. Similarly, the core decoder 720 may provide a time domain representation or transform domain representation of the residual signal 724, which is represented by the bit stream 710. Moreover, the core decoder 720 may provide one or more spatial parameters 726, like, for example, one or more inter-channel-correlation parameter, inter-channel-level difference parameters, or the like.

The decoder 700 also comprises a decorrelator 730, which is configured to provide a decorrelated signal 732 on the basis of the downmix signal 722. Any of the known decorrelation concepts may be used by the decorrelator 730. Moreover, the decoder 700 also comprises an upmix coefficient calculator 740, which is configured to receive spatial data 726 and to provide upmix parameters (for example, upmix parameters $u_{dmx,1}$, $u_{dmx,2}$, $u_{dec,1}$ and $u_{dec,2}$). Moreover, the decoder 700 comprises an upmixer 750, which is configured to apply the upmix parameters 742 (also designated as upmix coefficients) which are provided by the upmix coefficient calculator 740 on the basis of the spatial data 726. For example, the upmixer 750 may scale the downmix signal 722 using two

downmix-signal upmix coefficients (for example the $u_{dmx,1}$, $u_{dmx,2}$), to obtain two upmixed versions 752, 754 of the downmix signal 722. Moreover, the upmixer 750 is also configured to apply one or more upmix parameters (for example two upmix parameters) to the decorrelated signal 732 provided by the decorrelator 730, to obtain a first upmixed (scaled) version 756 and a second upmixed (scaled) version 758 of the decorrelated signal 732. Moreover, the upmixer 750 is configured to apply one or more upmix coefficients (for example, two upmix coefficients) to the residual signal 724, to obtain a first upmixed (scaled) version 760 and a second upmixed (scaled) version 762 of the residual signal 724.

The decoder 700 also comprises a weight calculator 770, which is configured to measure energies of the upmixed (scaled) versions 756, 758 of the decorrelated signal 752 and of the upmixed (scaled) version 760, 762 of the residual signal 724. Moreover, the weight calculator 770 is configured to provide one or more weighting values 772 to a weighter 780. The weighter 780 is configured to obtain a first upmixed (scaled) and weighted version 782 of the decorrelated signal 732, a second upmixed (scaled) and a weighted version 784 of the decorrelated signal 732, a first upmixed (scaled) and weighted version 786 of the residual signal 724 and a second upmixed (scaled) and weighted version 788 of the residual signal 724 using one or more weighting values 772 provided by the weight calculator 770. The decoder also comprises a first adder 790, which is configured to add up the first upmixed (scaled) version 752 of the downmix signal 720, the first upmixed (scaled) and weighted version 782 of the decorrelated signal 732 and the first upmixed (scaled) and weighted version 786 of the residual signal 724, to obtain the first output channel signal 712. Moreover, the decoder comprises a second adder 792, which is configured to add up the second upmixed version 754 of the downmix signal 720, the second upmixed (scaled) and weighted version 784 of the decorrelated signal 732 and the second upmixed (scaled) and weighted version 788 of the residual signal 724, to obtain the second output channel signal 714.

However, it should be noted, that it is not necessary that the weighter 780 weights all of the signals 756, 758, 760, 762. For example, in some embodiments it may be sufficient to weight only the signals 756, 758, while leaving the signals 760, 762 unaffected (such that, effectively, the signals 760, 762 are directly applied to the adders 790, 792. Alternatively, however, the weighting of the residual signals 760, 762 may be varied over time. For

example, the residual signals may be faded in or faded out. For example, the weighting (or the weighting factors) of the decorrelated signals may be smoothened over time, and the residual signals may be faded in or faded out correspondingly.

Moreover, it should be noted, that the weighting, which is performed by the weighter 780 and the upmixing, which is applied by the upmixer 750, may also be performed as a combined operation, wherein the weight calculation may be performed directly using the decorrelated signal 732 and the residual signal 724.

In the following, some further details regarding the functionality of the decoder 700 will be described.

A combined residual and parametric coding mode may, for example, be signaled in a semi-backwards compatible way, for example, by signaling a residual bandwidth of one parameter band in the bit stream. Thus, a legacy decoder will still pass and decode the bit stream by switching to parametric decoding above the first parameter band. Legacy bit streams using a residual bandwidth of one would not contain residual energy above the first parameter band, leading to a parametric decoding in the proposed new decoder.

However, within a 3D audio codec system, the combined residual and parametric coding may be used in combination with other core decoder tools like a quad channel element, enabling the decoder to explicitly detect legacy bit streams and decode them in regular band-limited residual coding mode. An actual residual bandwidth is preferably not explicitly signaled, as it is determined by the decoder at run time. The calculation of the upmix coefficients is set to parametric mode instead of a residual coding mode. The energies of the weighted decorrelator output E_{dec} and weighted residual signal E_{res} are calculated per hybrid band hb over all time slots ts and upmix channels ch for each frame:

$$E_{\mathrm{dec}}(\mathrm{hb}) = \sum_{\mathrm{ch}} \sum_{\mathrm{ts}} \| u_{\mathrm{dec}}(\mathrm{hb}, \mathrm{ts}, \mathrm{ch}) \cdot x_{\mathrm{dec}}(\mathrm{hb}, \mathrm{ts}, \mathrm{ch}) \| \tag{10}$$

$$E_{\mathrm{res}}(\mathrm{hb}) = \sum_{\mathrm{ch}} \sum_{\mathrm{ts}} \| u_{\mathrm{res}}(\mathrm{hb}, \mathrm{ts}, \mathrm{ch}) \cdot x_{\mathrm{res}}(\mathrm{hb}, \mathrm{ts}, \mathrm{ch}) \| \tag{11}$$

Here, u_{dec} designates a decorrelated signal upmix parameter for a frequency band hb, for a time slot ts and for an upmix channel ch, $\sum_{ch}$ designates a sum over upmix channels, and $\sum_{ts}$ designates a sum over time slots. x_{dec} designates a value (for example, a complex transform domain value) of the decorrelated signal for a frequency band hb, for a time slot ts and for an upmix channel ch.

The residual signal (for example, the upmixed residual signal 760 or the upmixed residual signal 762) is added to output channels (for example, to output channels 712, 714) with a weight of one. The decorrelator signal (for example the upmixed decorrelator signal 756 or the upmixed decorellator signal 758) may be weighted with a factor r (for example by the weighter 780) that is calculated as

$$r = \sqrt{\left| \frac{E_{\mathrm{dec}}(\mathrm{hb}) - E_{\mathrm{res}}(\mathrm{hb})}{E_{\mathrm{dec}}(\mathrm{hb})} \right|} \tag{12}$$

$$\tag{13}$$

wherein $E_{dec}(hb)$ represents a weighted energy value of the decorrelated signal x_{dec} for a frequency band hb, and wherein $E_{res}(hb)$ represents a weighted energy value of the residual signal x_{res} for a frequency band hb.

If no residual (for example, no residual signal 724) has been transmitted, for example, if $E_{res} = 0$, r (the factor which may be applied by the weighter 780, and which may be considered as a weighting value 772) becomes 1, which is equivalent to a purely parametric decoding. If the residual energy (for example, the energy of the upmixed residual signal 760 and/or of the upmixed residual signal 762) exceeds the decorrelator energy (for example, the energy of the upmixed decorrelated signal 756 or of the upmixed decorrelated signal 758), for example, if $E_{res} > E_{dec}$, the factor r may be set to zero, thus disabling the decorrelator and enabling partially wave form preserving decoding (which may be considered as residual coding). In the upmixing process, the weighted decorrelator output (for example, signals 782 and 784) and the residual signal (for example, signals 786, 788 or signals 760, 762) are both added to the output channels (for

example, signals 712, 714).

In conclusion, this leads to an upmix rule in matrix form

$$\begin{pmatrix} \mathrm{ch}_1 \\ \mathrm{ch}_2 \end{pmatrix} = \begin{bmatrix} u_{\mathrm{dmx},1} & r \cdot u_{\mathrm{dec},1} & \max\{u_{\mathrm{dmx},1}, 0.5\} \\ u_{\mathrm{dmx},2} & r \cdot u_{\mathrm{dec},2} & -\max\{u_{\mathrm{dmx},2}, 0.5\} \end{bmatrix} \cdot \begin{pmatrix} x_{\mathrm{dmx}} \\ x_{\mathrm{dec}} \\ x_{\mathrm{res}} \end{pmatrix} \qquad (14)$$

wherein ch1 represents one or more time domain samples or transform domain samples of a first output audio signal, wherein ch2 represents one or more time domain samples or transform domain samples of a second output audio signal, wherein x_{dmx} represents one or more time domain samples or transform domain samples of a downmix signal, wherein x_{dec} represents one or more time domain samples or transform domain samples of a decorrelated signal, wherein x_{res} represents one or more time domain samples or transform domain samples of a residual signal, wherein $u_{dmx,1}$ represents a downmix signal upmix parameter for the first output audio signal, wherein $u_{dmx,2}$ represents a downmix signal upmix parameter for the second output audio signal, wherein $u_{dec,1}$ represents a decorrelated signal upmix parameter for the first output audio signal, wherein $u_{dec,2}$ represents a decorrelated signal upmix parameter for the second output audio signal, wherein max represents a maximum operator, and wherein r represents a factor describing a weighting of the decorrelated signal in dependence on the residual signal.

The upmix coefficients $U_{dmx,1}$, $U_{dmx,2}$, $U_{dec,1}$,, $U_{dec,2}$ are calculated as for the MPS two-one-two (2-1-2) parametric mode. For details, reference is made to the above referenced standard of the MPEG surround concept.

To summarize, an embodiment according to the invention creates a concept to provide output channel signals on the basis of a downmix signal, a residual signal and spatial data, wherein a weighting of the decorrelated signal is flexibly adjusted without any significant signaling overhead.

7.5 Implementation alternatives

Although some aspects have been described in the context of an apparatus, it is clear that

these aspects also represent a description of the corresponding method, where a block or device corresponds to a method step or a feature of a method step. Analogously, aspects described in the context of a method step also represent a description of a corresponding block or item or feature of a corresponding apparatus. Some or all of the method steps may be executed by (or using) a hardware apparatus, like for example, a microprocessor, a programmable computer or an electronic circuit. In some embodiments, some one or more of the most important method steps may be executed by such an apparatus.

The inventive encoded audio signal can be stored on a digital storage medium or can be transmitted on a transmission medium such as a wireless transmission medium or a wired transmission medium such as the Internet.

Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a digital storage medium, for example a floppy disk, a DVD, a Blu-Ray, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed. Therefore, the digital storage medium may be computer readable.

Some embodiments according to the invention comprise a data carrier having electronically readable control signals, which are capable of cooperating with a programmable computer system, such that one of the methods described herein is performed.

Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing one of the methods when the computer program product runs on a computer. The program code may for example be stored on a machine readable carrier.

Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine readable carrier.

In other words, an embodiment of the inventive method is, therefore, a computer program

having a program code for performing one of the methods described herein, when the computer program runs on a computer.

A further embodiment of the inventive methods is, therefore, a data carrier (or a digital storage medium, or a computer-readable medium) comprising, recorded thereon, the computer program for performing one of the methods described herein. The data carrier, the digital storage medium or the recorded medium are typically tangible and/or non–transitory.

A further embodiment of the inventive method is, therefore, a data stream or a sequence of signals representing the computer program for performing one of the methods described herein. The data stream or the sequence of signals may for example be configured to be transferred via a data communication connection, for example via the Internet.

A further embodiment comprises a processing means, for example a computer, or a programmable logic device, configured to or adapted to perform one of the methods described herein.

A further embodiment comprises a computer having installed thereon the computer program for performing one of the methods described herein.

A further embodiment according to the invention comprises an apparatus or a system configured to transfer (for example, electronically or optically) a computer program for performing one of the methods described herein to a receiver. The receiver may, for example, be a computer, a mobile device, a memory device or the like. The apparatus or system may, for example, comprise a file server for transferring the computer program to the receiver.

In some embodiments, a programmable logic device (for example a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

The above described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

7.6 Further embodiment

In the following, another embodiment according to the invention will be described taking reference to Fig. 8, which shows a block schematic diagram of a so-called Hybrid Residual Decoder.

The Hybrid Residual Decoder 800 according to Fig. 8 is very similar to the Decoder 700 according to Fig. 7, such that reference is made to the above explanations. However, in the Hybrid Residual Decoder 800, an additional weighting (in addition to the application of the upmix parameters) is only applied to the upmixed decorrelated signals (which correspond to the signals 756,758 in the decoder 700), but not to the upmixed residual signals (which correspond to the signals 760, 762 in the decoder 700). Thus, the weighter in the Hybrid Residual Decoder 800 is somewhat simpler than the weighter in the decoder 700, but is well in agreement, for example, with the weighting according to equation (14).

In the following, the combined Parametric and Residual Decoding (Hybrid Residual Coding) according to Fig. 8 will be explained in some more detail.

However, firstly, an overview will be provided.

In addition to using either decorrelator-based mono-to-stereo upmixing or residual coding as described in ISO/IEC 23003-3, subclause 7.11.1, Hybrid Residual Coding allows a signal dependent combination of both modes. Residual signal and decorrelator output are blended together, using time and frequency dependent weighting factors depending on the signal energies and the spatial parameters, as illustrated in Fig. 8.

In the following, the decoding process will be described.

Hybrid Residual Coding mode is indicated by the syntax elements bsResidualCoding == 1 and bsResidualBands == 1 in Mps212Config(). In other words, the usage of the Hybrid Residual coding may be signaled using a bitstream element of the encoded representation. The calculation of mix-matrix M2 is performed as if bsResidualCoding == 0, following the calculation in ISO/IEC 23003-3, subclause 7.11.2.3. The matrix $\mathrm{R}_2^{\mathrm{l,m}}$ for the decorrelator based part is defined as

$$\mathrm{R}_2^{\mathrm{l,m}} = \begin{bmatrix} H11_{OTT}^{l,m} & H12_{OTT}^{l,m} \\ H21_{OTT}^{l,m} & H22_{OTT}^{l,m} \end{bmatrix}$$

The upmixing process is split up into Downmix, decorrelator output and residual. The upmixed Downmix u_{dmx} is calculated using:

$$\mathrm{R}_{2,\mathrm{dmx}}^{\mathrm{l,m}} = \begin{bmatrix} H11_{OTT}^{l,m} & 0 \\ H21_{OTT}^{l,m} & 0 \end{bmatrix}$$

The upmixed decorrelator output $\mathrm{u}_{\mathrm{dec}}$ is calculated using:

$$\mathrm{R}_{2,\mathrm{dec}}^{\mathrm{l,m}} = \begin{bmatrix} 0 & H12_{OTT}^{l,m} \\ 0 & H22_{OTT}^{l,m} \end{bmatrix}$$

The upmixed residual signal $\mathrm{u}_{\mathrm{res}}$ is calculated using:

$$\mathrm{R}_{2,\mathrm{res}}^{\mathrm{l,m}} = \begin{bmatrix} 0 & H12_{RES}^{l,m} \\ 0 & H22_{RES}^{l,m} \end{bmatrix} = \begin{bmatrix} 0 & \max\{0.5, H11_{OTT}^{l,m}\} \\ 0 & -\max\{0.5, H21_{OTT}^{l,m}\} \end{bmatrix}$$

The energies of the upmixed residual signal E_{res} and of the upmixed decorrelator output E_{dec} are calculated per hybrid band as sum over both output channels ch and all timeslots ts and of one frame as:

$$\mathrm{E}_{\mathrm{res}} = \sum_{ch}\sum_{ts}\left\|\mathrm{u}_{\mathrm{res}}(ch,ts)\right\|$$

$$\mathrm{E}_{\mathrm{dec}} = \sum_{ch}\sum_{ts}\left\|\mathrm{u}_{\mathrm{dec}}(ch,ts)\right\|$$

The upmixed decorrelator output is weighted using a weighting factor r_{dec} calculated for each hybrid band per frame as:

$$r_{dec} = \begin{cases} 0 & if \quad E_{res} > E_{dec} \\ 1 & if \quad E_{res} < \varepsilon \\ \sqrt{\left|\dfrac{E_{dec} - E_{res} + \varepsilon}{E_{dec} + \varepsilon}\right|} & else \end{cases}$$

with ε a small number to prevent division by zero (for example, ε = 1e-9, or 0<ε<=1e-5). However, in some embodiments, ε may be set to zero (replacing "$E_{res} < \varepsilon$" by "$E_{res} = 0$").

All three upmix signals are added to form the decoded output signal.

8. Conclusions

To conclude, embodiments according to the invention create a combined residual and parametric coding.

The present invention creates a method for a signal dependent combination of parametric and residual coding for joint stereo coding, which is based on the USAC unified stereo tool. Instead of using a fixed residual bandwidth, the amount of transmitted residual is determined signal dependently by an encoder, time and frequency variant. On decoder side, the required amount of decorrelation between the output channels is generated by mixing residual signal and decorrelator output. Thus, a corresponding audio coding/decoding system is able to blend between fully parametric coding and wave form preserving residual coding at run time, depending on the encoded signal.

Embodiments according to the invention outperform conventional solutions. For example, in USAC, an MPEG surround two-one-two (2-1-2) system is used for parametric stereo coding, or unified stereo, transmitting a band-limited or full-bandwidth residual signal for partial wave form preservation. If a band-limited residual is transmitted, parametric

upmixing with the use of decorrelators is applied above the residual bandwidth. The drawback of this method is, that the residual bandwidth is set to a fixed value at the encoder initialization.

In contrast, embodiments according to the invention allow for a signal dependent adaptation of the residual bandwidth or switching to parametric coding. Moreover, if the downmixing process in parametric coding mode produces signal cancellations for ill-conditioned phase relations, embodiments according to the invention allow to reconstruct missing signal parts (for example, by providing an appropriate residual signal). It should be noted, that the simplified downmix method produces less signal cancellations than the classic MPS downmix for parametric coding. However, while the conventional simplified downmix cannot be used for partial wave form preservation, since no residual signal is defined in USAC, embodiments according to the invention allow for a wave form reconstruction (for example, a selective partial wave form reconstruction for signal portions in which partial wave form reconstruction appears to be important).

To further conclude, embodiments according to the invention create an apparatus, a method or a computer program for audio encoding or decoding as described herein.