

Apparatus, Method or Computer Program for Generating an Output Downmix Representation

5

Specification

The present invention is related to multichannel processing and, particularly, to multichannel processing providing the possibility for a mono output.

- 10 While a stereo encoded bitstream will usually be decoded to be played back on a stereo system, not all devices that are able to receive a stereo bitstream will always be able to output a stereo signal. A possible scenario would be playback of the stereo signal on a mobile phone with only a mono speaker. With the advent of multi-channel mobile communication scenarios as supported by the emerging 3GPP IVAS standard a stereo-to-
- 15 mono downmix is therefore required that is free of additional delay and complexity-wise as efficient as possible while also providing the best possible perceptual quality beyond what is achievable with a simple passive downmix.

- There are multiple ways of converting a stereo signal to a mono signal. The most direct
- 20 ways of doing it is by a passive downmix [1] in time-domain which generates a mid-signal by adding the left and right channels and scaling the result:

$$Mid = \frac{(Left + Right)}{2}$$

- Further more sophisticated (i.e. active) time-domain based downmixing methods include
- 25 energy-scaling in an effort to preserve the overall energy of the signal [2] [3], phase alignment to avoid cancellation effects [4] and prevention of comb-filter effects by coherence suppression [5].

- Another method is to do the energy-correction in a frequency-dependent manner by
- 30 calculation separate weighting factors for multiple spectral bands. For instance, this is done as part of the MPEG-H format converter [6], where the downmix is performed on a hybrid QMF subband representation of the signals with additional prior phase alignment of the channels. In [7], a similar band-wise downmix (including both phase and temporal

alignment) is already used for the parametric low-bitrate mode DFT Stereo where the weighting and mixing is applied in the DFT domain.

5 The solution of a passive stereo-to-mono downmix in time-domain after decoding the stereo signal is not ideal as it is well known that a purely passive downmix comes with certain shortcomings, e.g. phase cancellation effects or general loss of energy, which can - depending on the item - severely degrade the quality.

10 Other active downmixing methods that are purely time-domain based mitigate some of problems of the passive downmix but are still suboptimal due to the lack of frequency-dependent weighting.

15 With the implicit constraints for mobile communication codecs like IVAS (Immersive Voice and Audio Services) in terms of delay and complexity, having a dedicated post-processing stage like the MPEG-H format converter for applying a band-wise downmix is also not an option as the necessary transforms to frequency domain and back will inevitably cause an increase in both complexity and delay.

20 In a DFT-based stereo system as described in [8] that uses only parameter-based residual prediction to restore the stereo signal at the decoder and where the mid-signal is generated by an active downmix as described in [7], a sufficiently good mono signal is available at the decoder. However, if spectral parts of the signal rely on a coded residual signal for stereo restoration that was generated by an M/S transform, the mono signal available before the stereo upmix is not suitable anymore. In this case the mono signal will
25 spectrally consist in part of the mid-signal from the M/S transform (residual coding part) which is equal to a passive downmix and partially of an active downmix (residual prediction part). This mixture of two different downmixing methods leads to artifacts and energy imbalances in signal.

30 It is an object of the present invention to provide an improved concept for generating an output downmix representation for multichannel decoding.

This object is achieved by an apparatus for generating an output downmix representation of claim 1, a multichannel decoder of claim 19, a method of generating an output downmix
35 representation of claim 24, a method of multichannel decoding of claim 27 or a related computer program of claim 28.

An apparatus for generating an output downmix representation from an input downmix representation, where at least a portion of the input downmix representation is in accordance with a first downmixing scheme, comprises an upmixer for upmixing at least a
5 portion of the input downmix representation using an upmixing scheme corresponding to the first downmixing scheme to obtain at least one upmixed portion. Furthermore, the apparatus comprises a downmixer for downmixing the at least one upmixed portion in accordance with a second downmixing scheme different from the first downmixing scheme.

10

In another embodiment, the portion of the input downmix representation is in accordance with the downmixing scheme and, additionally, a second portion of the input downmix representation is in accordance with a second downmixing scheme being different from the first downmixing scheme. In this embodiment, the downmixer is configured for
15 downmixing the upmix portion in accordance with the second downmixing scheme or in accordance with a third downmixing scheme different from the downmixing scheme and the second downmixing scheme to obtain the first downmixed portion. Now, the situation with respect to the downmixed portion is such that the first downmixed portion and the second portion are related and, as one could say, in the same downmix scheme domain,
20 so that the first downmixed portion and the second downmixed portion or a downmixed portion derived from the second downmixed portion can be combined by a combiner to obtain the output downmix representation comprising an output representation for the first portion and an output representation for the second portion, where the output representation for the first portion and the output representation for the second portion are
25 based on the same downmixing scheme, i.e., are located in one and the same downmixing domain and are, therefore, "harmonized" with each other.

In a further embodiment, either the whole bandwidth or just a portion of the input downmix representation is based on a downmixing scheme relying on parameters and a residual
30 signal or only relying on a residual signal without parameters. In such a context, the input downmix representation comprises a core signal, a residual signal or a residual signal and parameters. This signal is upmixed using the side information, i.e., using the parameters and the residual signal or using just the residual signal. The upmix comprises all the available information including the residual signal and a downmix is performed into the
35 second downmixing scheme which is different from the first downmixing scheme, i.e., which is, preferably, an active downmix having measures for addressing energy

calculations or, in other words, a downmixing scheme that does not generate a residual signal and, preferably, does not generate a residual signal and any parameters. Such a downmix provides a good and pleasant and high quality audio mono rendering possibility, while the core signal of the input downmix representation when used without upmixing and subsequent downmixing does not provide any pleasant and high quality audio reproduction if rendered without advantageously taking into consideration the residual signal and the parameters.

In accordance with this embodiment, the apparatus for generating an output downmix representation performs a conversion of a residual-like downmixing scheme into a non-residual like downmixing scheme. This conversion can be performed either in the full band or can also be performed in a partial band. Typically, and in preferred embodiments, the lowband of a multichannel-encoded signal comprises a core signal, a residual signal and preferably parameters. However, in the highband, less precision is provided in favor of a lower bit rate and, therefore, in such a highband an active downmix is sufficient without any additional side information such as residual data or parameters. In such a context, the lowband which is in the residual-downmix domain is converted into the non-residual downmix domain and the result is combined with the highband that is already in the "correct" non-residual downmix domain.

In a further embodiment, it is not required that the first portion is converted from the first downmix domain into the same downmix domain, in which the second portion is located. Instead, in further embodiments, where the first portion is in the first downmix domain and the second portion of the input representation is in the second downmix domain, both these portions are converted into another third downmix domain by upmixing the first portion in accordance with the first upmixing scheme corresponding to the first downmixing scheme. Additionally, the second portion is upmixed in accordance with the second upmixing scheme corresponding to the second downmixing scheme, and both upmixes are downmixed, preferably by an active downmix without any residual or parametric data, into the third downmixing scheme, which is different from the first and the second downmixing schemes.

In further embodiments, more than two portions and, in particular, spectral portions or spectral bands, can be available that are in different downmix representations. By means of the present invention, where, preferably, the upmixing and subsequent downmixing is performed in the spectral domain, individual processings for individual bands can be

performed without interference from one spectral band to the other spectral band. At the output of the downmixer, all bands are in the same "downmix" domain and, therefore, a spectrum for the mono output downmix representation exists, which can be converted into a time domain representation by a spectrum-time-converter such as a synthesis bank, an
5 inverse discrete Fourier transform, an inverse MDCT domain or any other such transform. The combination of the individual bands and the conversion into the time domain can be implemented by means of such a synthesis filter bank. In particular, it is irrelevant whether the combination is performed before the actual conversion, i.e., in the spectral domain. In such a situation, the combination takes place before the spectrum-time transform, i.e., at
10 the input into the synthesis filter bank and only a single transform is performed to obtain a single time domain signal. However, the equivalent implementation consists in the implementation where the combiner performs a spectrum-time transform for each band individually, so that the time domain output of each such individual transform represents a time domain representation but in a certain bandwidth, and the individual time domain
15 outputs are combined in a sample-by-sample manner preferably subsequent to some kind of upsampling when critically sampled transforms have been implemented.

In a further implementation, the present invention is applied within a multichannel decoder that is operable in two different modes, i.e., in the multichannel output mode as the
20 "normal" mode and that is also operable in a second mode such as an "exceptional mode" which is the mono output mode. This mono output mode is particularly useful when the multichannel decoder is implemented within a device which only has a mono speaker output facility such as a mobile phone having a single speaker or which is implemented in a device that is in some kind of power saving mode where, in order to save battery power
25 or to save processing resources, only a mono output mode is provided even though the device would, basically, also have the possibility for a multichannel or a stereo output mode.

In such an implementation, the multichannel decoder comprises a first time-spectrum
30 transform for the decoded core signal and a second time-spectrum transform facility for the decoder residual signal. Two different upmixing facilities in the spectral domain for two different spectral portions being in two different downmix domains are provided and the corresponding left channel spectral lines are combined by a combiner such as a synthesis filterbank or an IDFT block and the other channel spectral lines are combined by an
35 additional or second synthesis filterbank or IDFT (inverse discrete Fourier transform) block.

- In order to enhance such a multichannel decoder, the downmixer for downmixing the at least one upmixed portion in accordance with a second downmixing scheme different from the first downmixing scheme that is preferably implemented as an active downmixer is provided. Additionally, in an embodiment, two switches and a controller are provided as well. The controller controls a first switch to bypass an upmixer for the highband portion and the second switch is implemented to feed the downmixer with the output of the upmixer. In such a mono output mode, the second combiner or synthesis filterbank is inactive and the upmixer for the highband is inactive as well in order to save processing power. However, in the stereo output mode, the first switch feeds the upmix for the highband and the second switch bypasses the (active) downmixer and both output synthesis filterbanks are active in order to obtain the left stereo output signal and the right output signal.
- 15 Since the mono output is calculated in the spectral domain such as the DFT domain, the generation of the mono output does not incur any additional delay compared to the generation of the stereo output, because any additional time-frequency transforms compared to the stereo processing mode are not necessary. Instead, one of the two stereo mode synthesis filterbanks are used for the mono mode as well. Furthermore, compared to the stereo output that, typically, provides an enhanced audio experience 20 compared to the mono output, the mono processing mode saves complexity and, in particular, processing resources and, therefore, battery power in a low power mode particularly useful for a battery-powered mobile device. This is true, since the highband upmixer that is normally required in the stereo mode can be deactivated and, additionally, 25 a second output filterbank that is also required for the stereo output mode is deactivated as well. Instead, only a low complexity and low delay active downmix block fully operating in the spectral domain is required as an additional processing block compared to the stereo mode. The additional processing resources required by this active downmix block, however, are significantly smaller than the processing resources that are saved by 30 deactivating the highband upmixer and the second synthesis filterbank or IDFT block.

- Embodiments aim at generating a harmonized mono output signal from a mono input signal that was created by a downmix of a stereo signal where the downmix was done 35 with different methods (e.g. active and passive) for at least two different spectral regions of the stereo signal. The harmonization is achieved by picking one downmix method as

the preferred method for the harmonized signal and transforming all spectral parts that were downmixed via different methods to the preferred method. This is achieved by first upmixing these spectral parts using all the side parameters necessary for the upmix to regain an LR representation in the respective spectral regions. Again using all the
5 necessary parameters required for the preferred downmix method, the spectral parts are converted to a mono representation by applying the preferred method to the stereo representation. A harmonized mono output signal is generated that avoids the problems a non-uniform downmix without additional delay and complexity.

10 Subsequently, preferred embodiments are discussed with respect to the accompanying drawings, in which:

Fig. 1 illustrates an apparatus for generating an output downmix representation in an embodiment;

15

Fig. 2 illustrates an apparatus for generating an output downmix representation in a further embodiment, in which the downmixing scheme is based on a residual signal or a residual signal and parameters;

20 Fig. 3 illustrates a further embodiment, where different downmixing schemes are performed for different portions such as spectral portions of the input downmix representation;

Fig. 4 illustrates a further embodiment illustrating the usage of different
25 downmixing schemes in different spectral portions for the input downmix representation and the procedure where the first downmixing scheme is based on residual data and the second downmixing scheme is an active downmixing scheme or a downmixing scheme without residual or parametric data;

30

Fig. 5 illustrates a preferred implementation of the upmixing scheme corresponding to the first downmixing scheme in an embodiment;

Fig. 6 illustrates a multichannel decoder operating in a stereo output mode;

35

- Fig. 7 illustrates a multichannel encoder in accordance with an embodiment that is switchable between the multichannel output mode or the mono output mode;
- 5 Fig. 8a illustrates a preferred implementation for the second downmixing scheme;
- Fig. 8b illustrates a further embodiment of the second downmixing scheme; and
- 10 Fig. 9 illustrates the separation of an input downmix representation into the portion of the input downmix representation in the first downmixing scheme indicated as the first portion and into the second portion of the input downmixing representation that relies on a downmixing scheme with weights.
- 15 Fig. 1 illustrates an apparatus for generating an output downmix representation from an input downmix representation, where at least a portion of the input downmix representation is in accordance with a first downmixing scheme. The apparatus comprises an upmixer 200 for upmixing at least the portion of the input downmix representation using an upmixing scheme corresponding to the first downmixing scheme to obtain at least one
- 20 upmixed portion at the output of block 200. The apparatus furthermore comprises a downmixer 300 for downmixing the at least one upmixed portion in accordance with a second downmixing scheme being different from the first downmixing scheme. Preferably, the output of the downmixer 300 is forwarded to an output stage 500 for generating a mono output. The output stage is, for example, an output interface for outputting the
- 25 output downmix representation to a rendering device or the output stage 500 actually comprises a rendering device for rendering the output downmix representation as a mono replay signal.
- The apparatus illustrated in Fig. 1 provides a conversion from a downmix representation in
- 30 a first "downmix domain" into another second downmix domain. As will be illustrated in other figures, the conversion can be valid only for a limited part of the spectrum such as the first portion illustrated, for example, in Fig. 9 for the exemplarily given lowest three bands b_1 , b_2 and b_3 . Alternatively, the apparatus can also perform a conversion from one downmix domain to another downmix domain for the full band, i.e., for all bands b_1 to b_6
- 35 exemplarily illustrated in Fig. 9. The portion can be any portion of the signal such as a

spectral portion, a time portion such as a time block or frame, or any other portion of the signal.

Fig. 2 illustrates an embodiment where the first downmixing scheme relies on a residual signal only or on a residual signal and parametric information. Fig. 2 comprises an input interface 10 where the input interface receives an encoded multichannel signal that comprises an encoded core signal and an encoded side information part. The core signal is decoded by a core decoder 20 to provide the input downmix representation without side information. Additionally, the side information part from the encoded multichannel signal is provided and processed by the side information decoder 30 within the input interface, and the side information decoder 30 provides the residual signal or the residual signal and parameters as indicated at 210 in Fig. 2. The data, i.e., the input downmix that corresponds to the decoded core signal and the residual data are both input the upmixer 200 and the upmixer 200 generates an upmix signal that has a first channel and a second channel and the first channel and the second channel data are high quality audio data, since the high quality audio data are generated not only by the core signal and some kind of passive upmix, but are generated additionally using the residual data or the residual data and the parameters, i.e., all data available from the encoded multichannel signal. The output of the upmixer 200 is downmixed by the downmixer 300 using, for example, an active downmix or, generally, a downmixing scheme that does not generate a residual signal or that does not generate any parameters but that generates a downmix or mono signal that is energy-compensated, i.e., that does not suffer from energy fluctuations that are normally a significant problem when only a passive downmix is performed as is, for example, the case with the core signal generated by the core decoder 20 of Fig. 2. The output of the downmixer 300 is forwarded, for example, to a renderer for rendering the mono signal or, for example, to the output stage 500 illustrated in Fig. 1.

Fig. 3 illustrates a further embodiment where, again referring to Fig. 9, the first portion is available in the first downmixing scheme such as a downmixing scheme with residual data and where there is a second spectral portion that is available, for example in a second downmixing scheme without any residuals, i.e., that has been generated by an active downmix using, for example, downmix weights derived based on energy considerations to combat any fluctuations that otherwise would occur if a passive downmix would be applied.

35

The first portion of the downmix representation is input into the upmixer 200 that upmixes corresponding to the first downmixing scheme and the first portion is forwarded, as discussed with respect to Fig. 1 or Fig. 2, into the downmixer 300 that now performs a downmix in the second downmixing scheme. The second portion illustrated in Fig. 3 can be, for example, in the second downmixing scheme but can also be in a third, i.e., any other downmixing scheme, from the downmixing scheme of the portion input into the upmixer 200 or the second downmixing scheme output by the downmixer 300. In case of the downmixing domain being the same for the second portion and the output of the downmixer 300, any second portion processor 600 is not required. Instead, the second portion can be forwarded into a combiner 400 for combining the first and the second portion that are now harmonized with respect to their downmixing schemes. However, when the second portion is in a downmixing domain, i.e., has an underlying downmixing scheme being different from the downmixing scheme in which the output of the downmixer 300 is available, the second portion processor 600 is provided. Generally, the second portion processor 600 also comprises an upmixer for upmixing the second portion being in a third downmixing scheme and the second portion processor 600 additionally comprises a downmixer for downmixing the upmixer representation into the same downmixing domain, i.e., using the same downmixing scheme, as is available from the downmixer 300. The second portion processor 600 can be implemented using the upmixer 200 and the subsequently connected downmixer 300 so that a full harmonization of the data input into the combiner 400 is obtained. The combiner 400 outputs, preferably, a spectral representation of the mono output downmix representation which is converted into the time domain by means of a spectrum-time-converter such as a filterbank, an IDFT, an IMDCT, etc. Alternatively, the combiner 400 is configured for combining the individual inputs into individual time domain signals, and the time domain signals are combined in the time domain to obtain a time domain mono output downmix representation.

Fig. 4 comprises an input interface that may include a first time-to-spectrum converter 100 such as DFT block as illustrated in Fig. 4 and a second time-to-spectrum converter 120 such as the second DFT block in Fig. 4. The first block 100 is configured for converting the decoded core signal as, for example, output by the core decoder 20 of Fig. 2 into a spectral representation. Furthermore, the second time-to-spectral converter 120 is configured to convert the decoded residual signal as, for example, output by the side information decoder 30 of Fig. 2 into a spectral representation illustrated at 210a. Furthermore, line 210b illustrates optionally provided additional parametric data such as side gains that are also output by the side information decoder 30 of Fig. 2 for example.

- The upmixer 200 of Fig. 4 generates an upmixed left channel and an upmixed right channel for a lowband, i.e., exemplary for the first three band b_1, b_2, b_3 of Fig. 9. Furthermore, the lowband upmix at the output of block 200 is input into the downmixer 300 preferably performing an active downmix so that a lowband representation for the
- 5 exemplarily illustrated three bands b_1, b_2, b_3 of Fig. 9 is provided. This lowband downmix is now in the same domain as the highband downmix generated already by the DFT block 100. The output of block 100 for the highband would, in the example of Fig. 9, correspond to the downmix representation for bands b_4, b_5, b_6 . Now, at the input into the combiner 400, illustrated in Fig. 4 as an IDFT 400, the lowband representation and the highband
- 10 representation of the downmix are in the same "downmix domain", and have been generated with the same downmixing scheme. Now, the lowband and the highband of the harmonized downmix representation can be combined and preferably converted into the time domain to provide the mono output signal at the output of block 400.
- 15 A mostly parametric stereo scheme as described in [8] is built around the idea of only transmitting a single downmixed channel and recreating the stereo image via side parameters. This downmix at the encoder side is done in an active manner by dynamically calculating weights for both channels in the DFT domain [7]. These weights are computed band-wise using the respective energies of the two channels and their cross-correlation.
- 20 The target energy that has to be preserved by the downmix is equal to the energy of the phase-rotated mid-channel:

$$E_{target} = \left| \frac{L + Re^{-j\varphi}}{2} \right|^2 = \frac{\langle L, L \rangle + \langle R, R \rangle + 2|\langle L, R \rangle|}{4} = \frac{|L|^2 + |R|^2 + 2|\langle L, R \rangle|}{4},$$

- where L and R represent the left and right channel. Based on this target energy the
- 25 weights for the channels can be computed per band b as follows:

$$w_{R,b} = \frac{1}{2\sqrt{2}} \frac{\sqrt{|L_b|^2 + |R_b|^2 + 2|\langle L_b, R_b \rangle|}}{|L_b| + |R_b|}$$

And

$$w_{L,b} = w_{R,b} + 1 - \frac{|L_b + R_b|}{|L_b| + |R_b|}.$$

30

$|L|$ and $|R|$ are computed for each band b as

$$|L_b| = \sqrt{\sum_{i \text{ in } b} (L_{real,i,b}^2 + L_{imag,i,b}^2)}, |R_b| = \sqrt{\sum_{i \text{ in } b} (R_{real,i,b}^2 + R_{imag,i,b}^2)}$$

$|L + R|$ is computed as

$$|L_b + R_b| = \sqrt{|L_b|^2 + |R_b|^2 + 2 \text{dotprod}_{real}^2}$$

5

and $|\langle L, R \rangle|$ is computed as the absolute of the complex dot product

$$|\langle L_b, R_b \rangle| = \sqrt{\text{dotprod}_{real,b}^2 + \text{dotprod}_{imag,b}^2}$$

with

$$\text{dotprod}_{real,b} = \sum_{i \text{ in } b} (L_{real,i,b} R_{real,i,b} + L_{imag,i,b} R_{imag,i,b})$$

10

and

$$\text{dotprod}_{imag,b} = \sum_{i \text{ in } b} (L_{imag,i,b} R_{real,i,b} - L_{real,i,b} R_{imag,i,b})$$

where i specifies the bin number inside spectral band b .

- 15 The downmixed spectrum is obtained for each band by adding the weighted spectral bins of left and right channel:

$$DMX_{real,i,b} = w_{L,b} L_{real,i,b} + w_{R,b} R_{real,i,b}$$

and

$$DMX_{imag,i,b} = w_{L,b} L_{imag,i,b} + w_{R,b} R_{imag,i,b}.$$

- 20 If all the stereo processing in such a system is entirely reliant on parameters and the described active downmix is done on the whole spectrum, a mono signal that satisfies the given quality requirements by avoiding the problems of a passive downmix is already available after the core decoding. This means that in most cases it suffices to skip all decoder stereo processing and output the signal without going into DFT domain.

However, for higher bitrates this kind of system also supports the coding of a residual signal for the lower spectral bands. The residual signal can be seen as the side-signal of an MS-transform of these lowest bands while the core signal is the complementary mid-signal, basically a passive downmix of left and right. To keep the side signal as small as possible, a compensation of the interaural level differences (ILDs) between the channels is applied to it using side gains that are computed per band.

The downmixed mid-channel is computed at the encoder side for every spectral bin i inside the residual coding spectrum as

10

$$mid_i = \frac{L_i + R_i}{2}$$

while the complementary side channel is computed as

$$side_i = \frac{L_i - R_i}{2}.$$

The residual signal is obtained by subtracting the predicted part due to an ILD between left and right:

$$res_i = side_i - g_b * mid_i$$

with side gain g_b of the current spectral band b given as

20

$$g_b = \frac{|L_b|^2 - |R_b|^2}{|L_b + R_b|^2}.$$

The full-band signal going into the core coder is a mixture of passive downmix in lower bands and active downmix in all higher bands. Listening tests have shown that there are perceptual issues when playing back such a mixed signal. A way of harmonizing the different signal parts is therefore required.

Fig. 5 illustrates a representation of the upmixing scheme relying on residual data res_i and parametric data illustrated by bandwise side gain indices g_b . i stands for spectral values and b stands for a certain band. Fig. 5 illustrates a situation, which is also illustrated in Fig. 9, where each band b_i has several spectral lines. In particular, in order to calculate the

30

spectral value L_i , the mid-signal spectral value, i.e., the corresponding spectral value with index i of the output of the core decoder 20 or the output of DFT block 100 of Fig. 4 is used. Furthermore, the corresponding parameter g_b for the corresponding band, in which the spectral value i is located, is required as illustrated in Fig. 4 by line 210b and the residual spectral value as generated by block 120 and as illustrated at line 210a for the certain spectral value with index i and for the respective band b is required as well.

The L-R representations of the lowband signal with residual coding are thereby regained as follows:

10

$$L_i = mid_i + \widehat{g}_b * mid_i + res_i$$

and

$$R_i = mid_i - \widehat{g}_b * mid_i - res_i.$$

Subsequently, the active downmix is applied as described above, only the weights are calculated from the upmixed decoded spectra L and R . The lowband is combined with the already actively downmixed highband to create a harmonized signal which is brought back to time domain via IDFT.

Fig. 6 illustrates an implementation of a multichannel decoder for a stereo output. The multichannel decoder comprises elements of Fig. 4 that are indicated with the same reference numbers. Additionally, the stereo multichannel decoder comprises a second upmixer 220 for upmixing the highband downmix, i.e., the second portion into a second upmix representation comprising, for example, a left channel and a right channel for a stereo output as one implementation of the multichannel decoder. For another implementation of the multichannel decoder, where there are more than two output channels, such as three or more output channels, the upmixer 220 as well as the upmixer 200 would generate a corresponding higher number of output channels rather than only the left channel and the right channel.

Furthermore, a second combiner 420 is illustrated in Fig. 6 for the multichannel decoder, i.e., for the illustrated stereo decoder. In case of more than two outputs, a further combiner would be there for the third output channel and another one for the fourth output channel and so on. In contrast to Fig. 6, however, the downmixer 300 of Fig. 4 is not necessary for the multichannel output.

Fig. 7 illustrates a preferred implementation of a switchable multichannel decoder which is switchable by means of the actuation of a controller 700, between a mono mode or a stereo/multichannel output mode. Furthermore, in contrast to Fig. 6, the multichannel decoder additionally comprises the downmixer 300 already described with respect to Fig. 4 or the other figures. Furthermore, in the switchable implementation, one option is to provide two individual switches S1, S2. However, the switching functionalities illustrated at the bottom of Fig. 7 can also be implemented by other switching means such as combined switches or even more than two switches. Generally, switch 1 is configured to operate in the mono output mode, so that the second upmixer 220 also indicated as "upmix high" is bypassed. Furthermore, the second switch S2 is configured by the second control signal CTRL₂ to feed the active downmix 300 with the output of the upmixer 200 indicated as "upmix low" in Fig. 7. Furthermore, in the mono output mode, the upmix high block 220 described with respect to Fig. 6 is inactive and, additionally, the second combiner 420 indicated as "IDFT_R" is inactive as well, since only a single combiner 400 for the generation of the single mono output signal is required.

Contrary thereto, in the stereo output mode or, generally, in the multichannel output mode, the controller 700 is configured to activate, via control signal CTRL₁, the first switch so that the output of the first time-to-frequency converter 100 is fed into the second upmixer 220 indicated as "upmix high" in Fig. 7. By means of the actuation of switch S1, the second combiner 220 is activated. Furthermore, the controller 700 is configured to control the second switch S2 720 so that the output of block 200 is not input into the active downmixer 300, but the downmixer 300 is bypassed. The left channel (lowband) portion of the output of block 200 is forwarded as the lowband portion for the combiner 400 and the right channel lowband portion at the output of block 200 is forwarded to the lowband input of the second combiner 420 as illustrated in Fig. 7. Furthermore, in the stereo/multichannel output mode, the downmix 300 is inactive.

Fig. 8a illustrates a flow chart for an embodiment used in the downmix 300 for performing an active downmix. In a step 800, weights w_R and w_L are calculated based on a target energy. This is done per band such that a weight w_R for the right channel and a weight w_L for the left channel are obtained for each band.

In block 820, the weights are applied to the upmixed signal over the whole bandwidth of the signal under consideration or only in the corresponding portion per spectral bin. To

this end, block 820 receives the spectral domain (complex) signals or bins or spectral values. Subsequent to the application of the weights and, particularly, an addition of the weighted values to obtain the downmix, a conversion 840 to the time domain is performed. Depending on whether only a portion or the full band is processed in block 5 820, the conversion to the time domain takes place without any other portion or takes place with the other portion particularly in the context of a harmonized downmix as, for example, illustrated and discussed with respect to Fig. 3 or Fig. 4.

Fig. 8b illustrates a preferred implementation of the functionalities performed in block 800 of Fig. 8a. In particular, for the calculation of the weights w_R and w_L for each band, an amplitude-related measure for L is calculated for a band. To this end, the individual spectral lines for the left channel, i.e., for the left channel as output by block 200 of any of the Figs. 1 to 7 are input. In block 804, the same procedure is performed for the second channel or right channel in the same band b. Furthermore, in block 806, another amplitude-related measure is calculated for a linear combination of L and R in the band b. In block 806, once again, the spectral values of the first channel L, the spectral values for the second channel R are required for the band under consideration. In block 808, a cross-correlation measure is calculated between the left channel and the right channel or, generally, between the first channel and the second channel in the corresponding band b. 10 20 To this end, once again, the spectral values at indices e for the first and the second channels are required for the corresponding band.

As illustrated, the amplitude-related measure can be the square root over the squared magnitudes of the spectral values in a band. This is illustrated as $|L_b|$. Another amplitude-related measure would, for example, be the sum over the magnitudes of the spectral lines in the band without any square root or with an exponent being different from 1/2 such as an exponent being between 0 and 1 but excluding 0 and 1. Furthermore, the amplitude-related measure could also refer to a sum over exponentiated magnitudes of spectral lines where the exponent is different from 2. For example, using an exponent of 3 would correspond to the loudness in psychoacoustic terms. However, other exponents being greater than 1 would be useful as well. 25 30

The same is true for the amplitude-related measure calculated in block 804 or the amplitude-related measure calculated in block 806. 35

Furthermore, with respect to the cross-correlation measure calculated in block 808, the corresponding mathematical equation illustrated before also relies on a squaring of the dot products and the calculation of a square root. However, other exponents for the dot products different from 2 such as exponents equal to 3 corresponding to a loudness domain or exponents greater than 1 can be used as well. At the same time, instead of the square root, other exponents different from 1/2 can be used such as 1/3 or, generally, any exponent being between 0 and 1.

Furthermore, block 810 indicates the calculation of w_R and w_L based on the three amplitude-related measures and the cross-correlation measure. Although it has been indicated that the target energy is preserved by the downmix and is equal to the energy of the phase-rotated mid-channel, it is not necessary, neither for the calculation of w_R and w_L nor for the calculation of the actual downmix signal that such a rotation with a rotation angle is actually performed. Instead, the only thing being required when the actual rotation with the rotation angle Φ is not performed is the calculation of the cross-correlation measure between L and R in the corresponding bands b . In the previously described embodiment, although it has been indicated that an energy of a phase-rotated mid-channel is used as the target energy, any other target energies can be used or any phase rotation has not to be performed at all. With respect to other target energies, these target energies are energies that make sure that an energy of the downmix signal generated by the downmix 300 is fluctuating for the same signal less than the energy of a passive downmix as, for example, underlying the decoded core signal input into block 100 of Fig. 4.

Fig. 9 illustrates a general representation of a spectrum indicating a lowband first portion that is provided, with respect to the input downmix representation, as a downmix with residual data and indicating a second portion that is provided, with respect to the input downmix representation, by a downmix generated with weights as discussed before with respect to Fig. 8a, 8b. Although Fig. 9 illustrates only six bands, where three bands are for the first portion and three bands are for the second portion, and although Fig. 9 illustrates certain bandwidths that increase from lower bands to higher bands, the specific numbers, the specific bandwidths and the separation of the spectrum into the first portion and into the second portion are only exemplary. In a real scenario, a significantly higher number of bands will be there and, additionally, the first portion that, additionally, has the residual signal will be less than 50 % of the number of bands b .

Preferably, the time-to-spectral converters 100, 120 of Figs. 4, 6 and 7 and the combiner 400, 420 are implemented as DFT or IDFT blocks that preferably implement an FFT or IFFT algorithm. For the processing of a continuous decoded signal input into blocks 100, 120, a block wise processing is performed where overlapping blocks are formed, analysis
5 filtered, transformed into the spectral domain, processed and, in the combiners 400, 420 synthesis filtered, and combined, once again with a 50% overlap. The combination of a 50% overlap on the synthesis side will typically be performed by an overlap add operation with a cross fading from one block to the other where, preferably, the cross fading weights are already included in the analysis/synthesis windows. However, when this is not the
10 case, an actual cross fading is performed at the output of block 400, for example, or 420, for example, of Fig. 7 or Fig. 6, so that each time domain output sample of either the mono output signal or the left output signal or the right output signal is generated by an addition of two values of two different blocks. For an overlap of more than 50 %, an overlap between three or corresponding even more blocks can be performed as well.

15 Alternatively, when the time-to-spectral conversion on the one hand and the spectral-time-conversion on the other hand are performed with, for example, a modified discrete cosine transform, an overlap processing is used as well. On the spectral-to-time conversion side, an overlap-add processing is performed so that, once again, each output time domain
20 sample is obtained by summing corresponding time domain samples from two (or more) different IMDCT blocks.

Preferably, the harmonization of the downmixing schemes is performed fully in the spectral domain as illustrated in Figs. 4, 6 and 7. Any additional time-spectrum-transform
25 or spectrum-time-transform is not required when switching from mono to stereo or from stereo to mono as illustrated in Fig. 7. Only manipulations of data in the spectral domain either by the downmixer 300 for the mono output mode or by the second upmixer 220 (upmix high) for the stereo output mode have to be done. The whole delay of the processing is the same either for mono or stereo output and this is also a significant
30 advantage since any subsequent processing operations or preceding processing operations do not have to be aware of whether there is a mono or a stereo output signal.

Preferred embodiments remove artifacts and spectral loudness imbalances that stem from having different downmix methods in different spectral bands in the decoded core signal
35 of a system as described in [8] without the additional delay and significantly higher complexity that a dedicated post-processing stage would bring about.

Embodiments provide, in an aspect, an upmix and a subsequent downmix at the decoder of one (or more) spectral or time parts of a mono signal, that was downmixed using one or more than one downmix method, in order to harmonize all spectral or time parts of the
5 signal.

The present invention provides, in an aspect, a harmonization of a stereo-to-mono downmix at the decoder side.

10 In an embodiment, the output downmix is for a replay device that receives the downmix included in the output representation and feeds this downmix of the output representation into a digital to analog converter and the analog downmix signal is rendered by one or more loudspeakers included in the replay device. The replay device may be a mono device such as a mobile phone, a tablet, a digital clock, a Bluetooth speaker etc.

15

It is to be mentioned here that all alternatives or aspects as discussed before and all aspects as defined by independent claims in the following claims can be used individually, i.e., without any other alternative or object than the contemplated alternative, object or independent claim. However, in other embodiments, two or more of the alternatives or the
20 aspects or the independent claims can be combined with each other and, in other embodiments, all aspects, or alternatives and all independent claims can be combined to each other.

Although some aspects have been described in the context of an apparatus, it is clear that
25 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.

30 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 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
35 computer system such that the respective method is performed.

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.

5

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.

10

Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine readable carrier or a non-transitory storage medium.

15 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.

20 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.

25 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.

30 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.

35

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.

5 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.

10

References

- [1] ITU-R BS.775-2, *Multichannel Stereophonic Sound System With And Without Accompanying Picture*, 07/2006.
- [2] F. Baumgarte, C. Faller und P. Kroon, „Audio Coder Enhancement using Scalable Binaural Cue Coding with Equalized Mixing," in *116th Convention of the AES*, Berlin, 2004.
- [3] G. Stoll, J. Groh, M. Link, J. Deigmöller, B. Runow, M. Keil, R. Stoll, M. Stoll und C. Stoll, „Method for Generating a Downward-Compatible Sound Format". USA Patent US 2012/0 014 526, 2012.
- [4] M. Kim, E. Oh und H. Shim, „Stereo audio coding improved by phase parameters," in *129th Convention of the AES*, San Francisco, 2010.
- [5] A. Adami, E. Habets und J. Herre, „Down-mixing using coherence suppression," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, 2014.
- [6] ISO/IEC 23008-3:, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio*, 2019.
- [7] S. Bayer, C. Borß, J. Büthe, S. Disch, B. Edler, G. Fuchs, F. Ghido und M. Multrus, „DOWNMIXER AND METHOD FOR DOWNMIXING AT LEAST TWO CHANNELS AND MULTICHANNEL ENCODER AND MULTICHANNEL DECODER". Patent WO18086946, 17 05 2018.
- [8] S. Bayer, M. Dietz, S. Döhla, E. Fotopoulou, G. Fuchs, W. Jaegers, G. Markovic, M. Multrus, E. Ravelli und M. Schnell, „ APPARATUS AND METHOD FOR ESTIMATING AN INTER-CHANNEL TIME DIFFERENCE". Patent WO17125563, 27 07 2017.

Claims

1. Apparatus for generating an output downmix representation from an input downmix representation, wherein at least a portion of the input downmix representation is in accordance with a first downmixing scheme, the apparatus comprising:
- an upmixer (200) for upmixing at least the portion of the input downmix representation using an upmixing scheme corresponding to the first downmixing scheme to obtain at least one upmixed portion; and
- a downmixer (300) for downmixing the at least one upmixed portion in accordance with a second downmixing scheme different from the first downmixing scheme to obtain a first downmixed portion representing the output downmix representation for at least the portion of the input downmix representation.
2. Apparatus of claim 1, wherein only the portion of the input downmix representation is in accordance with the first downmixing scheme and a second portion of the input downmix representation is in accordance with the second downmixing scheme,
- wherein the downmixer (300) is configured for downmixing the at least one upmixed portion in accordance with the second downmixing scheme to obtain the first downmixed portion; and
- further comprising a combiner (400) for combining the first downmixed portion and the second portion of the input downmix representation or a downmixed portion derived from the second portion of the input downmix representation to obtain the output downmix representation comprising a first output representation for only the portion of the input downmix representation and a second output representation for the second portion of the input downmix representation, wherein the first output representation for only the portion of the input downmix representation and the second output representation for the second portion of the input downmix representation are based on the same downmixing scheme.
3. Apparatus of claim 1 or 2,

wherein the at least the portion of the input downmix representation or only the portion of the input downmix representation is a first frequency band, wherein the first downmixing scheme is a downmixing scheme relying on a residual signal, and

5 wherein the upmixer (200) is configured to perform an upmix using the residual signal.

4. Apparatus of claim 1, 2, or 3, wherein the second downmixing scheme is a fully parametric scheme, and wherein the downmixer (300) is configured to apply the
10 second downmixing scheme.

5. Apparatus of claim 2, 3, or 4, wherein the second portion of the input downmix representation is a second frequency band, and wherein the combiner (400) is configured to combine the first downmixed portion and the second portion of the
15 input downmix representation to obtain the output downmix representation.

6. Apparatus of any one of the preceding claims, further comprising an audio decoder (10) for generating a decoded core signal for at least the portion of the input downmix representation or only the portion of the input downmix representation, and a decoded residual signal for at least the portion of the input downmix representation or only the portion of the input downmix representation,
20

wherein the upmixer (200) is configured to use, in the upmixing scheme, the decoded core signal for at least the portion of the input downmix representation or only the portion of the input downmix representation and the decoded residual signal for at least the portion of the input downmix representation or only the portion of the input downmix representation,
25

wherein the downmixer (300) is configured for receiving the at least one upmixed portion comprising more channels than the input downmix representation.
30

7. Apparatus of claim 6, wherein the second portion of the input downmix representation is in accordance with the second downmixing scheme, wherein the audio decoder (10) is configured for generating a decoded core signal for the second portion of the input downmix representation and a decoded residual signal for at least the portion of the input downmix representation or only the portion of
35

the input downmix representation only, and wherein the combiner (400) is configured to combine the first downmixed portion and the decoded core signal for the second portion of the input downmix representation.

- 5 8. Apparatus of one of the preceding claims, further comprising:

10 a time-to-spectrum converter (100) for converting a time domain input downmix representation of at least the portion of the input downmix representation or only the portion of the input downmix representation into a spectral domain; and a spectrum-to-time converter (400) for converting an output signal into a time domain to obtain the output downmix representation, wherein the time-to-spectrum converter (100) or the spectrum-to-time converter (400) is configured to perform an overlap and add processing or to perform a crossover processing from an earlier time block to a later time block, or

15 further comprising an output interface (500) for outputting the output downmix representation to a rendering device or further comprising a rendering device for rendering the output downmix representation as a mono replay signal, or

20 wherein the downmixer (300) is configured to apply, as the second downmixing scheme, an active downmixing scheme, an energy conserving downmixing scheme, or a downmixing scheme, in which a target energy of the downmix signal is in a predetermined ratio to an energy of a mid-channel derived from a first channel and a second channel, wherein at least one of the first channel and the second channel is phase rotated before being added together to form the input downmix representation.

25 9. Apparatus of claim 8, wherein the second portion of the input downmix representation is in accordance with the second downmixing, wherein the time-to-spectrum converter (100) is configured for converting a time domain input downmix representation of the second portion of the input downmix representation into the spectral domain, or

30 wherein the predetermined ratio indicates an equality or a deviation range being 3 dB related to a higher energy of energies of a first original channel and a second original channel.

35

10. Apparatus of one of the preceding claims, wherein at least the portion of the input downmix representation is in accordance with the first downmixing scheme relying on a residual signal or on a residual signal and parametric information,

5

wherein the upmixer (200) is configured for upmixing the input downmix representation of at least the portion of the input downmix representation using the upmixing scheme corresponding to the first downmixing scheme and using the residual signal or the residual signal and the parametric information, respectively to obtain the at least one upmixed portion; and

10

wherein the downmixer (300) is configured for downmixing the at least one upmixed portion in accordance with the second downmixing scheme different from the first downmixing scheme, wherein the second downmixing scheme is an active downmixing scheme or a fully parametric downmixing scheme to obtain the output downmix representation comprising at least one downmixed portion.

15

11. Apparatus of claim 10, further comprising an output interface (500) for outputting the output downmix representation to a rendering device or further comprising a rendering device for rendering the output downmix representation as a mono replay signal.

20

12. Apparatus of claim 10 or 11, wherein the downmixer (300) is configured to apply, as the active downmixing scheme, an energy conserving downmixing scheme, or a downmixing scheme, in which a target energy of the downmix signal is in a predetermined ratio to an energy of a mid-channel derived from a first channel and a second channel, wherein at least one of the first channel and the second channel is phase rotated before being added together.

25

13. Apparatus of claim 10, 11, or 12, wherein at least the portion of the input downmix representation comprises the full bandwidth of the input downmix representation.

30

14. Apparatus of one of the preceding claims,

35

wherein the downmixer (300) is configured to perform the second downmixing scheme, the second downmixing scheme comprising:

calculating (800) a first weight for a first channel and a second weight for a second channel for a spectral band of the at least one upmixed portion, the spectral band comprising a plurality of spectral lines, and

5

applying (820) the first weight to spectral lines of the spectral band of the first channel and applying the second weight to spectral lines of the spectral band of the second channel, and adding first weighted lines and second weighted lines to obtain downmixed spectral lines in the spectral band, and

10

wherein the apparatus is configured to convert (840) the downmixed spectral lines to a time domain to obtain time domain samples of the output downmix representation.

15 15. Apparatus of claim 14, wherein the calculation of the first weight and the second weight is performed band wise using energies of the first channel and the second channel and a target energy.

16. Apparatus of claim 15, wherein the target energy is equal to an energy of a phase-rotated mid-channel or is derived from the energies of the first channel, the second channel and from a correlation value between the first channel and the second channel.

20

17. Apparatus of one of claims 14 to 16, wherein calculating the first weight and the second weight comprises, for a spectral band:

25

calculating (802) an amplitude-related measure for the first channel in the spectral band;

30 calculating (804) an amplitude-related measure for the second channel in the spectral band:

calculating (806) an amplitude-related measure for a linear combination of the first channel and the second channel in the spectral band;

35

calculating (808) a cross-correlation measure between the first channel and the second channel in the spectral band; and

5 calculating (810) the first weight and the second weight using the amplitude-related measure for the first channel, the amplitude-related measure for the second channel, the amplitude-related measure for the linear combination and the cross-correlation measure.

18. Apparatus of one of the preceding claims, wherein the upmixer (200) is configured
10 to perform the upmixing scheme, the upmixing scheme comprising:

calculating first channel spectral lines for a spectral band of at least the portion of
the input downmix representation or only the portion of the input downmix
representation from spectral lines of the spectral band of at least the portion of the
15 input downmix representation or only the portion of the input downmix
representation using a prediction parameter for the spectral band and residual
signal lines for the spectral band and a first calculation rule, and

calculating second channel spectral lines for the spectral band of at least the
20 portion of the input downmix representation or only the portion of the input
downmix representation from the spectral lines of the spectral band of at least the
portion of the input downmix representation or only the portion of the input
downmix representation using the prediction parameter for the spectral band and
the residual signal lines for the spectral band and a second calculation rule,

25 wherein the first calculation rule is different from the second calculation rule.

19. Apparatus of claim 18, wherein the first calculation rule comprises one of an
addition and a subtraction and the second calculation rule comprises the other one
30 of the addition and the subtraction.

20. Multichannel decoder, comprising:

an input interface (100, 120) for providing an input downmix representation and
35 parametric data at least for a second portion of the input downmix representation;
and

the apparatus of one of the preceding claims,

5 wherein the multichannel decoder is configured to upmix, with the upmixer (200),
the input downmix representation for at least the portion of the input downmix
representation or only the portion of the input downmix representation in
accordance with the upmixing scheme corresponding to the first downmixing
scheme to obtain the at least one upmixed portion, and/or to upmix (220) the input
10 downmix representation for the second portion and the parametric data using a
second upmixing scheme corresponding to the second downmixing scheme to
obtain an upmixed second portion, and

wherein a combiner (400, 420) is configured to combine the at least one upmixed
15 portion and the upmixed second portion to obtain a multichannel output signal.

21. Multichannel decoder of claim 20, wherein the input interface (100, 120)
comprises:

20 a first time-spectrum converter (100) for converting a first spectral representation of
the at least the portion of the input downmix representation or only the portion of
the input downmix representation and a second spectral representation of a
second portion of the input downmix representation, the second portion of the input
downmix representation comprising spectral values for higher frequencies than at
least the portion of the input downmix representation or only the portion of the
25 input downmix representation of the first spectral representation;

a second time-spectrum-converter (120) for generating a spectral representation of
a residual signal for the at least the portion of the input downmix representation or
only the portion of the input downmix representation,

30 wherein the upmixer (200) is configured to upmix the first spectral representation
using the spectral representation of the residual signal to obtain the at least one
upmixed portion in the spectral domain,

35 wherein the downmixer (300) is configured to downmix the at least one upmixed
portion to obtain the first downmixed portion in the spectral domain, and

5 wherein the combiner (400) comprises a spectrum-time converter for combining the first downmixed portion and the spectral representation of the second portion of the input downmix representation and for converting into the time domain to obtain the output downmix representation.

22. Multichannel decoder of claim 20 or 21, further comprising:

10 a second upmixer (220) for upmixing the second portion of the input downmix representation to obtain the upmixed second portion,

15 wherein, in a multichannel output mode, the combiner (400) is configured to combine a first channel of the at least one upmixed portion and the first channel of the upmixed second portion and to convert into a time domain to obtain a first channel of a multichannel output,

20 wherein the multichannel decoder further comprises a second combiner (420) configured to combine, in the multichannel output mode, a second channel of the at least one upmixed portion and a second channel of the upmixed second portion and to convert into the time domain to obtain a second channel of the multichannel output.

23. Multichannel decoder of claim 21, further comprising:

25 a second upmixer (220) for upmixing the second portion of the input downmix representation to obtain the upmixed second portion,

30 wherein, in a multichannel output mode, the combiner (400) is configured to combine a first channel of the at least one upmixed portion and the first channel of the upmixed second portion and to convert into a time domain to obtain a first channel of a multichannel output,

35 wherein the multichannel decoder further comprises a second combiner (420) configured to combine, in the multichannel output mode, a second channel of the at least one upmixed portion and a second channel of the upmixed second portion

and to convert into the time domain to obtain a second channel of the multichannel output,

5 a switch (710) connected between the first time-spectrum-converter (100) and the second upmixer (220), and

10 a controller (700), wherein the controller (700) is configured to control, in a mono-output mode, the switch (710) to connect an output of the first time-spectrum-converter (100) to the combiner (400) or to bypass the second upmixer (220) and to connect an output of the upmixer (200) to an input of the downmixer (300), or to control, in the multichannel output mode, the switch (710) to connect an output of the first time-spectrum-converter (100) to an input of the second upmixer (220).

15 24. Multichannel decoder of one of claims 22, 23, further comprising a second switch (720) connected between the upmixer (200) and the downmixer (300); and

20 a controller (700), wherein the controller (700) is configured to control, in the mono-output mode, the second switch (720) to connect an output of the upmixer (200) to an input of the downmixer (300) and to control, in the multichannel output mode, the second switch (720) to connect an output of the upmixer (200) to an input of the second combiner (420) or to bypass the downmixer (300).

25 25. Method for generating an output downmix representation from an input downmix representation, wherein at least a portion of the input downmix representation is in accordance with a first downmixing scheme, the method comprising:

upmixing the input downmix representation of at least the portion of the input downmix representation using an upmixing scheme corresponding to the first downmixing scheme to obtain an at least one upmixed portion; and

30 downmixing the at least one upmixed portion in accordance with a second downmixing scheme different from the first downmixing scheme to obtain a first downmixed portion representing the output downmix representation for at least the portion of the input downmix representation.

35

26. Method of claim 25, wherein a second portion of the input downmix representation is in accordance with a second downmixing scheme,

5 wherein the downmixing comprises downmixing the at least one upmixed portion in accordance with the second downmixing scheme to obtain the first downmixed portion; and

10 wherein the method further comprises combining the first downmixed portion and the second portion or a downmixed portion derived from the second portion to obtain the output downmix representation, wherein the output downmix representation for at least the portion of the input downmix representation and the output representation for the second portion are based on the same downmixing scheme.

- 15 27. Method of claim 25 or 26, wherein at least the portion of the input downmix representation is in accordance with the first downmixing scheme relying on a residual signal or on a residual signal and parametric information,

20 wherein the upmixing comprises upmixing the input downmix representation of at least the portion of the input downmix representation using an upmixing scheme corresponding to the first downmixing scheme and using the residual signal or the residual signal and the parametric information, respectively to obtain the at least one upmixed portion; and

25 wherein the downmixing comprises downmixing the at least one upmixed portion in accordance with the second downmixing scheme different from the first downmixing scheme, wherein the second downmixing scheme is an active downmixing scheme or a fully parametric downmixing scheme to obtain the output downmix representation for at least the portion of the input downmix representation.
30

28. Method of multichannel decoding, comprising:

35 providing an input downmix representation and parametric data at least for a second portion of the input downmix representation;

the method of any one of claims 25 to 27,

wherein the method comprises the upmixing the input downmix representation for at least the portion of the input downmix representation or only the portion of the input downmix representation in accordance with the upmixing scheme corresponding to the first downmixing scheme to obtain the at least one upmixed portion, and/or upmixing the second portion of the input downmix representation and the parametric data using a second upmixing scheme corresponding to the second downmixing scheme to obtain an upmixed second portion, and

combining the at least one upmixed portion and the upmixed second portion to obtain a multichannel output signal.

29. Computer program for performing, when running on a computer or a processor, the method of any one of claims 25 to 28.

30. Apparatus for generating an output downmix representation from an input downmix representation, wherein a first portion of the input downmix representation is in accordance with a first downmixing scheme and a second portion of the input downmix representation is in accordance with the second downmixing scheme, the apparatus comprising:

an upmixer (200) for upmixing the first portion of the input downmix representation using a first upmixing scheme corresponding to the first downmixing scheme to obtain a first upmixed portion and for upmixing the second portion of the input downmix representation using a second upmixing scheme corresponding to the second downmixing scheme to obtain a second upmixed portion; and

a downmixer (300) for downmixing the first upmixed portion and the second upmixed portion in accordance with a third downmixing scheme different from the first downmixing scheme and the second downmixing scheme to obtain the output downmix representation, wherein the output representation for the first portion of the input downmix representation and the output representation for the second portion of the input downmix representation are based on the same downmixing scheme of the input downmix representation.