Device and method for generating an encoded stereo signal of an audio piece or audio datastream

Description

The present invention relates to multi-channel audio technology and, in particular, to multi-channel audio applications in connection with headphone technologies.

The international patent applications WO 99/49574 and WO 99/14983 disclose audio signal processing technologies for driving a pair of oppositely arranged headphone loudspeakers in order for a user to get a spatial perception of the audio scene via the two headphones, which is not only a stereo representation but a multi-channel representation. Thus, the listener will get, via his or her headphones, a spatial perception of an audio piece which in the best case equals his or her spatial perception, should the user be sitting in a reproduction room which is exemplarily equipped with a 5.1 audio system. For this purpose, for each headphone loudspeaker, each channel of the multi-channel audio piece or the multi-channel audio datastream, as is illustrated in Fig. 2, is supplied to a separate filter, whereupon the respective filtered channels belonging together are added, as will be illustrated subsequently.

On a left side in Fig. 2, there are the multi-channel inputs 20 which together represent a multi-channel representation of the audio piece or the audio datastream. Such a scenario is exemplarily schematically shown in Fig. 10. Fig. 10 shows a reproduction space 200 in which a so-called 5.1 audio system is arranged. The 5.1 audio system includes a center loudspeaker 201, a front-left loudspeaker 202, a front-right loudspeaker 203, a back-left loudspeaker 204 and a back-right loudspeaker 205. A 5.1 audio system comprises an additional subwoofer 206 which is also referred to as low-frequency enhancement channel. In the

so-called "sweet spot" of the reproduction space 200, there is a listener 207 wearing a headphone 208 comprising a left headphone loudspeaker 209 and a right headphone loudspeaker 210.

The processing means shown in Fig. 2 is formed to filter each channel 1, 2, 3 of the multi-channel inputs 20 by a filter H_{iL} describing the sound channel from the loudspeaker to the left loudspeaker 209 in Fig. 10 and to additionally filter the same channel by a filter H_{iR} representing the sound from one of the five loudspeakers to the right ear or the right loudspeaker 210 of the headphone 208.

If, for example, channel 1 in Fig. 2 were the front-left channel emitted by the loudspeaker 202 in Fig. 10, the filter H_{iL} would represent the channel indicated by a broken line 212, whereas the filter H_{1R} would represent the channel indicated by a broken line 213. As is exemplarily indicated in Fig. 10 by a broken line 214, the left headphone loudspeaker 209 does not only receive the direct sound, but also early reflections at an edge of the reproduction space and, of course, also late reflections expressed in a diffuse reverberation.

Such a filter representation is illustrated in Fig. 11. In particular, Fig. 11 shows a schematic example of an impulse response of a filter, such as, for example, of the filter H_{1L} of Fig. 2. The direct or primary sound illustrated in Fig. 11 by the line 212 is represented by a peak at the beginning of the filter, whereas early reflections, as are illustrated exemplarily in Fig. 10 by 214, are reproduced by a center region having several (discrete) small peaks in Fig. 11. The diffuse reverberation is typically no longer resolved for individual peaks, since the sound of the loudspeaker 202 in principle is reflected arbitrarily frequently, wherein the energy of course decreases with each reflection and additional propagation distance, as is

illustrated by the decreasing energy in the back portion which in Fig. 11 is referred to as "diffuse reverberation".

Each filter shown in Fig. 2 thus includes a filter impulse response roughly having a profile as is shown by the schematic impulse response illustration of Fig. 11. It is obvious that the individual filter impulse response will depend on the reproduction space, the positioning of the loudspeakers, possible attenuation features in the reproduction space, for example due to several persons present or due to furniture in the reproduction space, and ideally also on the characteristics of the individual loudspeakers 201 to 206.

The fact that the signals of all loudspeakers are superposed at the ear of the listener 207 is illustrated by the adders 22 and 23 in Fig. 2. Thus, each channel is filtered by a corresponding filter for the left ear to then simply add up the signals output by the filters which are destined for the left ear to obtain the headphone output signal for the left ear L. In analogy, an addition by the adder 23 for the right ear or the right headphone loudspeaker 210 in Fig. 10 is performed to obtain the headphone output signal for the right ear by superposing all the loudspeaker signals filtered by a corresponding filter for the right ear.

Due to the fact that, apart from the direct sound, there are also early reflections and, in particular, a diffuse reverberation, which is of particularly high importance for the space perception, in order for the tone not to sound synthetic or "awkward" but to give the listener the impression that he or she is actually sitting in a concert room with its acoustic characteristics, impulse responses of the individual filters 21 will all be of considerable lengths. The convolution of each individual multi-channel of the multi-channel representation having two filters already results in a considerable computing task. Since two

filters are required for each individual multi-channel, namely one for the left ear and another one for the right ear, when the subwoofer channel is also treated separately, a total amount of 12 completely different filters is required for a headphone reproduction of a 5.1 multi-channel representation. All filters have, as becomes obvious from Fig. 11, a very long impulse response to be able to not only consider the direct sound but also early reflections and the diffuse reverberation, which really only gives an audio piece the proper sound reproduction and a good spatial impression.

In order to put the well-known concept into practice, apart from a multi-channel player 220, as is shown in Fig. 10, very complicated virtual sound processing 222 is required, which provides the signals for the two loudspeakers 209 and 210 represented by lines 224 and 226 in Fig. 10.

Headphone systems for generating a multi-channel headphone sound are complicated, bulky and expensive, which is due to the high computing power, the high current requirement for the high computing power required and the high working memory requirements for the evaluations to be performed of the impulse response and the high volume or expensive elements for the player connected thereto. Applications of this kind are thus tied to home PC sound cards or laptop sound cards or home stereo systems.

In particular, the multi-channel headphone sound remains inaccessible for the continually increasing market of mobile players, such as, for example, mobile CD players, or, in particular, hardware players, since the calculating requirements for filtering the multi-channels with exemplarily 12 different filters cannot be realized in this price segment neither with regard to the processor resources nor with regard to the current requirements of typically battery-driven apparatuses. This refers to a price segment at the bottom (lower) end of the scale.

However, this very price segment is economically very interesting due to the high numbers of pieces.

The object of the present invention is to provide an efficient signal-processing concept allowing a multi-channel quality headphone reproduction on simple reproduction apparatuses.

This object is achieved by a device for generating an encoded stereo signal according to claim 1 or by a method for generating an encoded stereo signal according to claim 11 or by a computer program according to claim 12.

The present invention is based on the finding that the high-quality and attractive multi-channel headphone sound can be made available to all players available, such as, for example, CD players or hardware players, by subjecting a multi-channel representation of an audio piece or audio datastream, i.e. exemplarily a 5.1 representation of an audio piece, to headphone signal processing outside a hardware player, i.e. exemplarily in a computer of a provider having a high calculating power. According to the invention, the result of a headphone signal processing is, however, not simply played but supplied to a typical audio stereo encoder which then generates an encoded stereo signal from the left headphone channel and the right headphone channel.

This encoded stereo signal may then, like any other encoded stereo signal not comprising a multi-channel representation, be supplied to the hardware player or, for example, a mobile CD player in the form of a CD. The reproduction or replay apparatus will then provide the user with a headphone multi-channel sound without any additional resources or means having to be added to devices already existing. Inventively, the result of the headphone signal processing, i.e. the left and the right headphone signal, is not reproduced in a headphone, as has been the case in

the prior art, but encoded and output as encoded stereo data.

Such an output may be storage, transmission or the like. Such a file having encoded stereo data may then easily be supplied to any reproduction device designed for stereo reproduction, without the user having to perform any changes on his device.

The inventive concept of generating an encoded stereo signal from the result of the headphone signal processing thus allows multi-channel representation providing a considerably improved and more real quality for the user, to be also employed on all simple and widespread and, in future, even more widespread hardware players.

In a preferred embodiment of the present invention, the starting point is an encoded multi-channel representation, i.e. a parametric representation comprising one or typically two basic channels and additionally comprising parametric data to generate the multi-channels of the multi-channel representation on the basis of the basic channels and the parametric data. Since a frequency domain-based method for multi-channel decoding is preferred, the headphone signal processing is, according to the invention, not performed in the time domain by convoluting the time signal by an impulse response, but in the frequency domain by multiplication by the filter transmission function.

This allows at least one retransformation before the headphone signal processing to be saved and is of particular advantage when the subsequent stereo encoder also operates in the frequency domain, such that the stereo encoding of the headphone stereo signal, without ever having to go to the time domain, may also take place without going to the time domain. The processing from the multi-channel representation to the encoded stereo signal, without the time domain taking part or by an at least

reduced number of transformations, is interesting not only with regard to the calculating time efficiency, but puts a limit to quality losses since fewer processing stages will introduce fewer artefacts into the audio signal.

In particular in block-based methods performing quantization considering a psycho-acoustic masking threshold, as is preferred for the stereo encoder, it is important to prevent as may tandem encoding artefacts as possible.

In a particularly preferred embodiment of the present invention, a BCC representation having one or preferably two basic channels is used as a multi-channel representation. Since the BCC method operates in the frequency domain, the multi-channels are not transformed to the time domain after synthesis, as is usually done in a BCC decoder. Instead, the spectral representation of the multi-channels in the form of blocks is used and subjected to the headphone signal processing. For this, the transformation functions of the filters, i.e. the Fourier transforms of the impulse responses, are used to perform a multiplication of the spectral representation of the multi-channels by the filter transformation functions. When the impulse responses of the filters are, in time, longer than a block of spectral components at the output of the BCC decoder, a block-wise filter processing is preferred where the impulse responses of the filters are separated in the time domain and are transformed block by block in order to then perform corresponding spectrum weightings required for measures of this kind, as is, for example, disclosed in WO 94/01933.

Preferred embodiments of the present invention will be detailed subsequently referring to the appended drawings, in which:

Fig. 1 shows a block circuit diagram of the inventive device for generating an encoded stereo signal;

Fig. 2 is a detailed illustration of an implementation of the headphone signal processing of Fig. 1;

Fig. 3 shows a well-known joint stereo encoder for generating channel data and parametric multi-channel information;

Fig. 4 is an illustration of a scheme for determining ICLD, ICTD and ICC parameters for BCC encoding/decoding;

Fig. 5 is a block diagram illustration of a BCC encoder/decoder chain;

Fig. 6 shows a block diagram of an implementation of the BCC synthesis block of Fig. 5;

Fig. 7 shows cascading between a multi-channel decoder and the headphone signal processing without any transformation to the time domain;

Fig. 8 shows cascading between the headphone signal processing and a stereo encoder without any transformation to the time domain;

Fig. 9 shows a principle block diagram of a preferred stereo encoder;

Fig. 10 is a principle illustration of a reproduction scenario for determining the filter functions of Fig. 2; and

Fig. 11 is a principle illustration of an expected impulse response of a filter determined according to Fig. 10.

Fig. 1 shows a principle block circuit diagram of an inventive device for generating an encoded stereo signal of an audio piece or an audio datastream. The stereo signal includes, in an uncoded form, an uncoded first stereo channel 10a and an uncoded second stereo channel 10b and is generated from a multi-channel representation of the audio piece or the audio data stream, wherein the multi-channel representation comprises information on more than two multi-channels. As will be explained later, the multi-channel representation may be in an uncoded or an encoded form. If the multi-channel representation is in an uncoded form, it will include three or more multi-channels. With a preferred application scenario, the multi-channel representation includes five channels and one subwoofer channel.

If the multi-channel representation is, however, in an encoded form, this encoded form will typically include one or several basic channels as well as parameters for synthesizing the three or more multi-channels from the one or two basic channels. A multi-channel decoder 11 thus is an example of means for providing the more than two multi-channels from the multi-channel representation. If the multi-channel representation is, however, already in an uncoded form, i.e., for example, in the form of 5+1 PCM channels, the means for providing corresponds to an input terminal for means 12 for performing headphone signal processing to generate the uncoded stereo signal with the uncoded first stereo channel 10a and the uncoded second stereo channel 10b.

Preferably, the means 12 for performing headphone signal processing is formed to evaluate the multi-channels of the multi-channel representation each by a first filter function for the first stereo channel and by a second filter function for the second stereo channel and to add the respective evaluated multi-channels to obtain the

uncoded first stereo channel and the uncoded second stereo channel, as is illustrated referring to Fig. 2. Downstream of the means 12 for performing the headphone signal processing is a stereo encoder 13 which is formed to encode the first uncoded stereo channel 10a and the second uncoded stereo channel 10b to obtain the encoded stereo signal at an output 14 of the stereo encoder 13. The stereo encoder performs a data rate reduction such that a data rate required for transmitting the encoded stereo signal is smaller than a data rate required for transmitting the uncoded stereo signal.

According to the invention, a concept is achieved which allows supplying a multi-channel tone, which is also referred to as "surround", to stereo headphones via simple players, such as, for example, hardware players.

The sum of certain channels may exemplarily be formed as simple headphone signal processing to obtain the output channels for the stereo data. Improved methods operate with more complex algorithms which in turn obtain an improved reproduction quality.

It is to be mentioned that the inventive concept allows the calculating-intense steps for multi-channel decoding and for performing the headphone signal processing not to be performed in the player itself but to be performed externally. The result of the inventive concept is an encoded stereo file which is, for example, an MP3 file, an AAC file, an HE-AAC file or some other stereo file.

In other embodiments, the multi-channel decoding, headphone signal processing and stereo encoding may be performed on different devices since the output data and input data, respectively, of the individual blocks may be ported easily and be generated and stored in a standardized way.

Subsequently, reference will be made to Fig. 7 showing a preferred embodiment of the present invention where the multi-channel decoder 11 comprises a filter bank or FFT function such that the multi-channel representation is provided in the frequency domain. In particular, the individual multi-channels are generated as blocks of spectral values for each channel. Inventively, the headphone signal processing is not performed in the time domain by convoluting the temporal channels with the filter impulse responses, but a multiplication of the frequency domain representation of the multi-channels by a spectral representation of the filter impulse response is performed. An uncoded stereo signal is achieved at the output of the headphone signal processing, which is, however, not in the time domain but includes a left and a right stereo channel, wherein such a stereo channel is given as a sequence of blocks of spectral values, each block of spectral values representing a short-term spectrum of the stereo channel.

In the embodiment shown in Fig. 8, the headphone signal-processing block 12 is, on the input side, supplied with either time-domain or frequency-domain data. On the output side, the uncoded stereo channels are generated in the frequency domain, i.e. again as a sequence of blocks of spectral values. A stereo encoder which is based on a transformation, i.e. which processes spectral values without a frequency/time conversion and a subsequent time/frequency conversion being necessary between the headphone signal processing 12 and the stereo encoder 13, is preferred as the stereo encoder 13 in this case. On the output side, the stereo encoder 13 then outputs a file with the encoded stereo signal which, apart from side information, includes an encoded form of spectral values.

In a particularly preferred embodiment of the present invention, a continuous frequency domain processing is performed on the way from the multi-channel representation at the input of block 11 of Fig. 1 to the encoded stereo

file at the output 14 of the means of Fig. 1, without a transformation to the time domain and, possibly, a re-transformation to the frequency domain having to take place. When an MP3 encoder or an AAC encoder is used as the stereo encoder, it will be preferred to transform the Fourier spectrum at the output of the headphone signal-processing block to an MDCT spectrum. Thus, it is ensured according to the invention that the phase information required in a precise form for the convolution/evaluation of the channels in the headphone signal-processing block is converted to the MDCT representation not operating in such a phase-correct way, such that means for transforming from the time domain to the frequency domain, i.e. to the MDCT spectrum, is not required for the stereo encoder, in contrast to a normal MP3 encoder or a normal AAC encoder.

Fig. 9 shows a general block circuit diagram for a preferred stereo encoder. The stereo encoder includes, on the input side, a joint stereo module 15 which is preferably determining in an adaptive way whether a common stereo encoding, for example in the form of a center/side encoding, provides a higher encoding gain than a separate processing of the left and right channels. The joint stereo module 15 may further be formed to perform an intensity stereo encoding, wherein an intensity stereo encoding, in particular with higher frequencies, provides a considerable encoding gain without audible artefacts arising. The output of the joint stereo module 15 is then processed further using different other redundancy-reducing measures, such as, for example, TNS filtering, noise substitution, etc., to then supply the results to a quantizer 16 which achieves a quantization of the spectral values using a psycho-acoustic masking threshold. The quantizer step size here is selected such that the noise introduced by quantizing remains below the psycho-acoustic masking threshold, such that a data rate reduction is achieved without the distortions introduced by the lossy quantization to be audible. Downstream of the quantizer 16, there is an

entropy encoder 17 performing lossless entropy encoding of the quantized spectral values. At the output of the entropy encoder, there is the encoded stereo signal which, apart from the entropy-coded spectral values, includes side information required for decoding.

Subsequently, reference will be made to preferred implementations of the multi-channel decoder and to preferred multi-channel illustrations using Figs. 3 to 6.

There are several techniques for reducing the amount of data required for transmitting a multi-channel audio signal. Such techniques are also called joint stereo techniques. For this purpose, reference is made to Fig. 3 showing a joint stereo device 60. This device may be a device implementing, for example, the intensity stereo (IS) technique or the binaural cue encoding technique (BCC). Such a device generally receives at least two channels CH1, CH2, ..., CHn as input signal and outputs a single carrier channel and parametric multi-channel information. The parametric data are defined so that an approximation of an original channel (CH1, CH2, ..., CHn) may be calculated in a decoder.

Normally, the carrier channel will include subband samples, spectral coefficients, time domain samples, etc., which provide a relatively fine representation of the underlying signal, whereas the parametric data do not include such samples or spectral coefficients, but control parameters for controlling a certain reconstruction algorithm, such as, for example, weighting by multiplication, time shifting, frequency shifting, etc. The parametric multi-channel information thus includes a relatively rough representation of the signal or the associated channel. Expressed in numbers, the amount of data required by a carrier channel is in the range of 60 to 70 kbits/s, whereas the amount of data required by parametric side information for a channel is in the range from 1.5 to 2.5

kbits/sec. It is to be mentioned that the above numbers apply to compressed data. A non-compressed CD channel of course requires approximately tenfold data rates. An example of parametric data are the known scale factors, intensity stereo information or BCC parameters, as will be described below.

The intensity stereo encoding technique is described in the AES Preprint 3799 entitled "Intensity Stereo Coding" by J. Herre, K.H. Brandenburg, D. Lederer, February 1994, Amsterdam. In general, the concept of intensity stereo is based on a main axis transform which is to be applied to data of the two stereophonic audio channels. If most data points are concentrated around the first main axis, an encoding gain may be achieved by rotating both signals by a certain angle before encoding takes place. However, this does not always apply to real stereophonic reproduction techniques. Thus, this technique is modified in that the second orthogonal component is excluded from being transmitted in the bitstream. Thus, the reconstructed signals for the left and right channels consist of differently weighted or scaled versions of the same transmitted signal. Nevertheless, the reconstructed signals differ in amplitude, but they are identical with respect to their phase information. The energy time envelopes of both original audio channels, however, are maintained by means of the selective scaling operation typically operating in a frequency-selective manner. This corresponds to human sound perception at high frequencies where the dominant spatial information is determined by the energy envelopes.

In addition, in practical implementations, the transmitted signal, i.e. the carrier channel, is produced from the sum signal of the left channel and the right channel instead of rotating both components. Additionally, this processing, i.e. generating intensity stereo parameters for performing the scaling operations, is performed in a frequency-selective manner, i.e. independently for each scale factor

band, i.e. for each encoder frequency partition. Preferably, both channels are combined to form a combined or "carrier" channel and, in addition to the combined channel, the intensity stereo information. The intensity stereo information depends on the energy of the first channel, the energy of the second channel or the energy of the combined channel.

The BCC technique is described in the AES Convention Paper 5574 entitled "Binaural Cue Coding applied to stereo and multichannel audio compression" by T. Faller, F. Baumgarte, May 2002, Munich. In BCC encoding, a number of audio input channels are converted to a spectral representation using a DFT-based transform with overlapping windows. The resulting spectrum is divided into non-overlapping portions, of which each has an index. Each partition has a bandwidth which is proportional to the equivalent right-angled bandwidth (ERB). The inter-channel level differences (ICLD) and the inter-channel time differences (ICTD) are determined for each partition and for each frame k. The ICLD and ICTD are quantized and encoded to finally reach a BCC bitstream as side information. The inter-channel level differences and the inter-channel time differences are given for each channel with regard to a reference channel. Then, the parameters are calculated according to predetermined formulae depending on the particular partitions of the signal to be processed.

On the decoder side, the decoder typically receives a mono-signal and the BCC bitstream. The mono-signal is transformed to the frequency domain and input into a spatial synthesis block which also receives decoded ICLD and ICTD values. In the spatial synthesis block, the BCC parameters (ICLD and ICTD) are used to perform a weighting operation of the mono-signal, to synthesize the multi-channel signals which, after a frequency/time conversion, represent a reconstruction of the original multi-channel audio signal.

In the case of BCC, the joint stereo module 60 is operative to output the channel-side information such that the parametric channel data are quantized and encoded ICLD or ICTD parameters, wherein one of the original channels is used as a reference channel for encoding the channel-side information.

Normally, the carrier signal is formed of the sum of the participating original channels.

The above techniques of course only provide a mono-representation for a decoder which can only process the carrier channel, but which is not able to process parametric data for generating one or several approximations of more than one input channel.

The BCC technique is also described in the US patent publication US 2003/0219130 A1, US 2003/0026441 A1 and US 2003/0035553 A1. Additionally, reference is made to the expert publication "Binaural Cue Coding. Part II: Schemes and Applications" by T. Faller and F. Baumgarte, IEEE Trans. On Audio and Speech Proc., Vol. 11, No. 6, November 2003.

Subsequently, a typical BCC scheme for multi-channel audio encoding will be illustrated in greater detail referring to Figs. 4 to 6.

Fig. 5 shows such a BCC scheme for encoding/transmitting multi-channel audio signals. The multi-channel audio input signal at an input 110 of a BCC encoder 112 is mixed down in a so-called downmix block 114. With this example, the original multi-channel signal at the input 110 is a 5-channel surround signal having a front-left channel, a front-right channel, a left surround channel, a right surround channel and a center channel. In the preferred embodiment of the present invention, the downmix block 114

generates a sum signal by means of a simple addition of these five channels into one mono-signal.

Other downmix schemes are known in the art, so that using a multi-channel input signal, a downmix channel having a single channel is obtained.

This single channel is output on a sum signal line 115. Side information obtained from the BCC analysis block 116 is output on a side-information line 117.

Inter-channel level differences (ICLD) and inter-channel time differences (ICTD) are calculated in the BCC analysis block, as has been illustrated above. Now, the BCC analysis block 116 is also able to calculate inter-channel correlation values (ICC values). The sum signal and the side information are transmitted to a BCC decoder 120 in a quantized and encoded format. The BCC decoder splits the transmitted sum signal into a number of subbands and performs scalings, delays and further processing steps to provide the subbands of the multi-channel audio channels to be output. This processing is performed such that the ICLD, ICTD and ICC parameters (cues) of a reconstructed multi-channel signal at the output 121 match the corresponding cues for the original multi-channel signal at the input 110 in the BCC encoder 112. For this purpose, the BCC decoder 120 includes a BCC synthesis block 122 and a side information-processing block 123.

Subsequently, the internal setup of the BCC synthesis block 122 will be illustrated referring to Fig. 6. The sum signal on the line 115 is supplied to a time/frequency conversion unit or filter bank FB 125. At the output of block 125, there is a number N of subband signals or, in an extreme case, a block of spectral coefficients when the audio filter bank 125 performs a 1:1 transformation, i.e. a transformation generating N spectral coefficients from N time domain samples.

The BCC synthesis block 122 further includes a delay stage 126, a level modification stage 127, a correlation processing stage 128 and an inverse filter bank stage IFB 129. At the output of stage 129, the reconstructed multi-channel audio signal having, for example, five channels in the case of a 5-channel surround system, may be output to a set of loudspeakers 124, as are illustrated in Fig. 5 or Fig. 4.

The input signal sn is converted to the frequency domain or the filter bank domain by means of the element 125. The signal output by the element 125 is copied such that several versions of the same signal are obtained, as is illustrated by the copy node 130. The number of versions of the original signal equals the number of output channels in the output signal. Then, each version of the original signal at the node 130 is subjected to a certain delay d_1, d_2, ..., d_i, ..., d_N. The delay parameters are calculated by the side information-processing block 123 in Fig. 5 and derived from the inter-channel time differences as they were calculated by the BCC analysis block 116 of Fig. 5.

The same applies to the multiplication parameters a_1, a_2, ..., a_i, ..., a_N, which are also calculated by the side information-processing block 123 based on the inter-channel level differences as they were calculated by the BCC analysis block 116.

The ICC parameters calculated by the BCC analysis block 116 are used for controlling the functionality of block 128 so that certain correlations between the delayed and level-manipulated signals are obtained at the outputs of block 128. It is to be noted here that the order of the stages 126, 127, 128 may differ from the order shown in Fig. 6.

It is also to be noted that in a frame-wise processing of the audio signal, the BCC analysis is also performed frame-

wise, i.e. temporally variable, and that further a frequency-wise BCC analysis is obtained, as can be seen by the filter bank division of Fig. 6. This means that the BCC parameters are obtained for each spectral band. This also means that in the case that the audio filter bank 125 breaks down the input signal into, for example, 32 band-pass signals, the BCC analysis block obtains a set of BCC parameters for each of the 32 bands. Of course, the BCC synthesis block 122 of Fig. 5, which is illustrated in greater detail in Fig. 6, also performs a reconstruction which is also based on the exemplarily mentioned 32 bands.

Subsequently, a scenario used for determining individual BCC parameters will be illustrated referring to Fig. 4. Normally, the ICLD, ICTD and ICC parameters may be defined between channel pairs. It is, however, preferred that the ICLD and ICTD parameters are determined between a reference channel and each other channel. This is illustrated in Fig. 4A.

ICC parameters may be defined in different manners. In general, ICC parameters may be determined in the encoder between all possible channel pairs, as is illustrated in Fig. 4B. There has been the suggestion to calculate only ICC parameters between the two strongest channels at any time, as is illustrated in Fig. 4C, which shows an example in which, at any time, an ICC parameter between the channels 1 and 2 is calculated and, at another time, an ICC parameter between the channels 1 and 5 is calculated. The decoder then synthesizes the inter-channel correlation between the strongest channels in the decoder and uses certain heuristic rules for calculating and synthesizing the inter-channel coherence for the remaining channel pairs.

With respect to the calculation of, for example, the multiplication parameters a_1, a_N based on the transmitted ICLD parameters, reference is made to the AES Convention

Paper No. 5574. The ICLD parameters represent an energy distribution of an original multi-channel signal. Without loss of generality, it is preferred, as is shown in Fig. 4A, to take 4 ICLD parameters representing the energy difference between the respective channels and the front-left channel. In the side information-processing block 122, the multiplication parameters a_1, ..., a_N are derived from the ICLD parameters so that the total energy of all reconstructed output channels is the same (or proportional to the energy of the sum signal transmitted).

In the embodiment shown in Fig. 7, the frequency/time conversion obtained by the inverse filter banks IFB 129 of Fig. 6 is dispensed with. Instead, the spectral representations of the individual channels at the input of these inverse filter banks are used and supplied to the headphone signal-processing device of Fig. 7 to perform the evaluation of the individual multi-channels with the respective two filters per multi-channel without an additional frequency/time transformation.

With regard to a complete processing taking place in the frequency domain, it is to be noted that in this case the multi-channel decoder, i.e., for example, the filter bank 125 of Fig. 6, and the stereo encoder should have the same time/frequency resolution. Additionally, it is preferred to use one and the same filter bank, which is particularly of advantage in that only a single filter bank is required for the entire processing, as is illustrated in Fig. 1. In this case, the result is a particularly efficient processing since the transformations in the multi-channel decoder and the stereo encoder need not be calculated.

The input data and output data, respectively, in the inventive concept are thus preferably encoded in the frequency domain by means of transformation/filter bank and are encoded under psycho-acoustic guidelines using masking effects, wherein in particular in the decoder there should

be a spectral representation of the signals. Examples of this are MP3 files, AAC files or AC3 files. However, the input data and output data, respectively, may also be encoded by forming the sum and difference, as is the case in so-called matrixed processes. Examples of this are Dolby ProLogic, Logic7 or Circle Surround. The data of, in particular, the multi-channel representation may additionally be encoded by means of parametric methods, as is the case in MP3 surround, wherein this method is based on the BCC technique.

Depending on the circumstances, the inventive method for generating may be implemented in either hardware or software. The implementation may be on a digital storage medium, in particular on a disc or CD having control signals which can be read out electronically, which can cooperate with a programmable computer system such that the method will be executed. In general, the invention also is in a computer program product having a program encode stored on a machine-readable carrier for performing an inventive method when the computer program product runs on a computer. Put differently, the invention may also be realized as a computer program having a program encode for performing the method when the computer program runs on a computer.

Claims

1. A device for generating an encoded stereo signal of an audio piece or an audio datastream having a first stereo channel and a second stereo channel from a multi-channel representation of the audio piece or the audio datastream comprising information on more than two multi-channels, comprising:

means (11) for providing the more than two multi-channels from the multi-channel representation;

means (12) for performing headphone signal processing to generate an uncoded stereo signal with an uncoded first stereo channel (10a) and an uncoded second stereo channel (10b); and

a stereo encoder (13) for encoding the first uncoded stereo channel (10a) and the uncoded second stereo channel (10b) to obtain the encoded stereo signal (14), the stereo encoder being formed such that a data rate required for transmitting the encoded stereo signal is smaller than a data rate required for transmitting the uncoded stereo signal.

2. The device according to claim 1, wherein the means (12) for performing is formed:

to evaluate each multi-channel by a first filter function (H_{iL}) for the first stereo channel and by a second filter function (H_{iR}) for the second stereo channel to generate a first evaluated channel and a second evaluated channel for each multi-channel,

to add up (22) all evaluated first channels to obtain the first uncoded stereo channel (10a), and

to add up (23) all evaluated second channels to obtain the second uncoded stereo channel (10b).

3. The device according to claim 2, wherein to each multi-channel a separate pair of a first and a second filter function is associated, wherein the first filter function is derived from a virtual position of a loudspeaker for reproducing the multi-channel and a virtual first ear position of a listener, and

wherein the second filter function is derived from a virtual position of the loudspeaker and a virtual second ear position of the listener, the two virtual ear positions of the listener being different.

4. The device according to one of the preceding claims,

wherein the multi-channel representation comprises one or several basic channels as well as parametric information for calculating the multi-channels from one or several basic channels, and

wherein the means (11) for providing is formed to calculate the at least three multi-channels from the one or the several basic channels and the parametric information.

5. The device according to claim 4,

wherein the means (11) for providing is formed to provide, on the output side, a block-wise frequency domain representation for each multi-channel, and

wherein the means (12) for performing is formed to evaluate the block-wise frequency domain representation by a frequency domain representation of the first and second filter functions.

6. The device according to one of the preceding claims,

wherein the means (12) for performing is formed to provide a block-wise frequency domain representation of the uncoded first stereo channel and the uncoded second stereo channel, and

wherein the stereo encoder (13) is a transformation-based encoder and is also formed to process the block-wise frequency domain representation of the uncoded first stereo channel and the uncoded second stereo channel without a conversion from the frequency domain representation to a temporal representation.

7. The device according to one of the preceding claims,

wherein the stereo encoder (13) is formed to perform a common stereo encoding (15) of the first and second stereo channels.

8. The device according to one of the preceding claims,

wherein the stereo encoder (13) is formed to quantize (16) a block of spectral values using a psycho-acoustic masking threshold and subject it to entropy encoding (17) to obtain the encoded stereo signal.

9. The device according to one of the preceding claims,

wherein the means (11) for providing is formed as a BCC decoder.

10. The device according to one of the preceding claims,

wherein the means (11) for providing is formed as a multi-channel decoder comprising a filter bank having several outputs,

wherein the means (12) for performing is formed to evaluate signals at the filter bank outputs by the first and second filter functions, and

wherein the stereo encoder (13) is formed to quantize (16) the uncoded first stereo channel in the frequency domain and the uncoded second stereo channel in the frequency domain and subject it to entropy encoding (17) to obtain the encoded stereo signal.

11. A method for generating an encoded stereo signal of an audio piece or an audio datastream having a first stereo channel and a second stereo channel from a multi-channel representation of the audio piece or the audio datastream comprising information on more than two multi-channels, comprising the steps of:

providing (11) the more than two multi-channels from the multi-channel representation;

performing (12) headphone signal processing to generate an uncoded stereo signal with an uncoded first stereo channel (10a) and an uncoded second stereo channel (10b); and

stereo-coding (13) the first uncoded stereo channel (10a) and the uncoded second stereo channel (10b) to obtain the encoded stereo signal (14), the step of stereo-coding being executed such that a data rate required for transmitting the encoded stereo signal is smaller than a data rate required for transmitting the uncoded stereo signal.

12. A computer program having a program code for performing the method for generating an encoded stereo signal according to claim 11, when the computer program runs on a computer.