(19) DANMARK

(10) **DK/EP 4307124 T3**

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: ***G 10 L 19/008 (2013.01)*** ***G 06 F 12/0815 (2016.01)*** *H 04 S 3/02 (2006.01)*

(45) Oversættelsen bekendtgjort den: **2024-11-04**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2024-08-07**

(86) Europæisk ansøgning nr.: **23214132.5**

(86) Europæisk indleveringsdag: **2006-06-21**

(87) Den europæiske ansøgnings publiceringsdag: **2024-01-17**

(30) Prioritet: **2005-07-19 US 701001 P** **2005-12-29 US 323965**

(62) Stamansøgningsnr: **23180543.3**

(84) Designerede stater: **AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR**

(73) Patenthaver: **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., Hansastraße 27c, 80686 München, Tyskland**
Dolby Laboratories Licensing Corporation, 1275 Market Street, San Francisco, CA 94103, USA

(72) Opfinder: **Hilpert, Johannes, , 91058 Erlangen, Tyskland**
Faller, Christof, , 91058 Erlangen, Tyskland
Linzmeier, Karsten, , 90491 Nürnberg, Tyskland
SPERSCHNEIDER, Ralph, , 91058 Erlangen, Tyskland

(74) Fuldmægtig i Danmark: **Larsen & Birkeholm A/S Skandinavisk Patentbureau, Banegårdspladsen 1, 1570 København V, Danmark**

(54) Benævnelse: **Koncept til at bygge bro mellem parametrisk multikanalaudiokodning og matrix-surround-multikanalkodning**

(56) Fremdragne publikationer:
EP-A1- 1 376 538
HERRE J ET AL: "THE REFERENCE MODEL ARCHITECTURE FOR MPEG SPATIAL AUDIO CODING", AUDIO ENGINEERING SOCIETY CONVENTION PAPER, NEW YORK, NY, US, 28 May 2005 (2005-05-28), pages 1 - 13, XP009059973
MEARES D J ET AL: "MATRIXED SURROUND SOUND IN AN MPEG DIGITAL WORLD", JOURNAL OF THE AUDIO ENGINEERING SOCIETY, AUDIO ENGINEERING SOCIETY, NEW YORK, NY, US, vol. 46, no. 4, April 1998 (1998-04-01), pages 331 - 335, XP000783409, ISSN: 1549-4950
BREEBART J ET AL: "MPEG Spatial Audio Coding / MPEG surround: Overview and Current Status", AUDIO ENGINEERING SOCIETY CONVENTION PAPER, NEW YORK, NY, US, 7 October 2005 (2005-10-07), pages 1 - 17, XP002379094

DESCRIPTION

Description

Field of the invention

[0001] The present invention relates to multi-channel audio coding and transmission, and in particular to techniques to encode and decode multi-channel audio in a manner that is fully backwards compatible with stereo devices and formats, allowing for an efficient coding of multi-channel audio.

Background of the invention and prior art

[0002] Parametric coding of multi-channel audio signals is an ongoing topic of research. Generally two approaches to encode multi-channel audio signals can be distinguished. The Moving Pictures Experts Group (MPEG), a subgroup of the International Organization for Standardization (ISO), is currently working on the standardization of technology for the reconstruction of multi-channel audio content from stereo or even mono down-mix signals by adding only a small amount of helper information to the down-mix signals.

[0003] In parallel stereo to multi-channel up-mix methods are being developed which do not need any additional side-information that is not already (implicitly) contained in the down-mix signal in order to reconstruct the spatial image of the original multi-channel audio signal.

[0004] Existing methods for stereo-compatible multi-channel transmission without additional side-information that gained practical relevance can mostly be characterized as matrixed-surround methods, such as Dolby Pro Logic (Dolby Pro Logic II) and Logic-7, as described in more detail in "Dolby Surround Pro Logic II Decoder - Principles of Operation", http://www.dolby.com/assets/pdf/tech_library/209_Dolby_Surround_Pro_Logic_II_Decoder_Principles_of_Operation.pdf and in "Multichannel Matrix Surround Decoders for Two-Eared Listeners",

[0005] Griesinger, D., 101st AES Convention, Los Angeles, USA, 1996, Preprint 4402. The common principle of these methods is that they make use of dedicated ways of multi-channel or stereo down-mixing where the encoder applies phase shifts to the surround channels prior to mixing them together with front and center channels to form a stereo down-mix signal. The generation of the down-mix signal (Lt, Rt) is depicted in the following equation:

$$\begin{bmatrix} Lt \\ Rt \end{bmatrix} = \begin{bmatrix} 1 & 0 & q & a \cdot j & b \cdot j \\ 0 & 1 & q & b \cdot j & a \cdot j \end{bmatrix} \begin{bmatrix} Lf \\ Rf \\ C \\ Ls \\ Rs \end{bmatrix} \quad (1)$$

[0006] The left down-mix signal (Lt) consists of the left-front signal (Lf), the center signal (C) multiplied by a factor q, the left-surround signal (Ls) phase rotated by 90 degrees (‚j') and scaled by a factor a, and the right-surround signal (Rs) which is also phase rotated by 90 degrees and scaled by a factor b. The right down-mix signal (Rt) is generated similarly. Typical down-mix factors are 0.707 for q and a, and 0.408 for b. The rationale for the different signs of the surround channels for the right down-mix signal (Rt) and the left down-mix signal (Lt) is, that it is advantageous to mix the surround channels in anti-phase in the down-mix pair (Lt, Rt).

This property helps the decoder to discriminate between front and rear channels from the down-mix signal pair. Hence the down-mix matrix allows for a partial reconstruction of a multi-channel output signal out from the stereo down-mix within the decoder by applying a de-matrixing operation. How close the re-created multi-channel signal resembles the original encoder input signal, however, depends on the specific properties of the multi-channel audio content.

[0007] An example for a coding method adding helper information, also called side information, is MPEG Surround audio coding. This efficient way for parametric multi-channel audio coding is for example described in "The Reference Model Architecture for MPEG Spatial Audio Coding", Herre, J., Purnhagen, H., Breebaart, J., Faller, C., Disch, S., Kjoerling, K., Schuijers, E., Hilpert, J., Myburg, F., Proc. 118th AES Convention, Barcelona, Spain, 2005 and in "Text of Working Draft for Spatial Audio Coding (SAC)", ISO/IEC JTC1/SC29/WG11 (MPEG), Document N7136, Busan, Korea, 2005.

[0008] A schematic overview of an encoder used in spatial audio coding is shown in Fig. 6. The encoder splits incoming signals 10 (input 1, ... input N) in separate time-frequency tiles by means of Quadrature Mirror Filters 12 (QMF). Groups of the resulting frequency tiles (bands) are referred to as "parameter bands". For every parameter band, a number of spatial parameters 14 are determined by a parameter estimator 16 that describes the properties of the spatial image, e.g. level differences between pairs of channels (CLD), cross correlation between pairs of channels (ICC) or information on signal envelopes (CPC). These parameters are subsequently quantized, encoded and compiled jointly into a bit-stream of spatial data. Depending on the operation mode, this bit-stream can cover a wide range of bit-rates, starting from a few kBit/s for good quality multi-channel audio up to tenths of kBit/s for near-transparent quality.

[0009] Besides the extraction of parameters, the encoder also generates a mono or stereo down-mix from the multi-channel input signal. Moreover, in case of a stereo down-mix, the user has the choice of a conventional (ITU-style) stereo down-mix or of a down-mix that is compatible with matrixed-surround systems. Finally, the stereo down-mix is transferred to the time-domain by means of QMF synthesis banks 18. The resulting down-mix can be transmitted to a decoder, accompanied by the spatial parameters or the spatial parameter bit-stream 14. Preferably, the down-mix is also encoded before transmission (using a conventional mono or stereo core coder), while the bit-streams of the core coder and the spatial parameters might additionally be combined (multiplexed) to form a single output bit-stream.

[0010] A decoder, as sketched in Fig. 7, in principle performs the reverse process of the encoder. An input-stream is split into a core coder bit-stream and a parameter bit-stream. This is not shown in Fig. 7. Subsequently, the decoded down-mix 20 is processed by a QMF analysis bank 22 to derive parameter bands that are the same as those applied in the encoder. A spatial synthesis stage 24 reconstructs the multi-channel signal by means of control data 26 (i.e., the transmitted spatial parameters). Finally, the QMF-domain signals are transferred to the time domain by means of a QMF synthesis bank 27 that derives the final multi-channel output signals 28.

[0011] Fig. 8 shows a simple example of a QMF analysis, as it is performed within the prior art encoder in Fig. 6 and the prior art decoder in Fig. 7. An audio sample 30, sampled in the time domain and having four sample values is input into a filter bank 32. The filter bank 32 derives three output samples 34a, 34b and 34c having four sample values each. In an ideal case, the filter bank 32 derives the output samples 34a to 34c such that the samples within the output signals do only comprise information on discrete frequency ranges of the underlying audio signal 30. In the case shown in Fig. 8, the sample 34a has information on the frequency interval ranging from f0 to f1, the sample 34b has information of the frequency interval [f1, f2] and the sample 34c has information on the frequency interval [f2, f3]. Although the frequency

intervals in Fig. 8 do not overlap, in a more general case the frequency intervals of the output samples coming out of a filter bank may very well have a frequency overlap.

[0012] A prior art encoder can, as already described above, deliver either an ITU-style down-mix or a matrixed-surround compatible down-mix, when a two-channel down-mix is desired. In the case of a matrixed-surround compatible down-mix (using for example the matrixing approach given in Equation 1), one possibility would be that the encoder generates a matrixed-surround compatible down-mix directly.

[0013] Fig. 9 shows an alternative approach to generate a matrixed-surround compatible down-mix using a down-mix post processing unit 30 working on a regular stereo down-mix 32. The matrixed-surround processor 30 (MTX encoder) modifies the regular stereo down-mix 32 to make it matrixed-surround compatible guided by the spatial parameters 14 extracted by the parameter extraction stage 16. For transmission, a matrixed-surround compatible down-mix 34 is transferred to the time domain by a QMF synthesis using the QMF synthesis bank 18.

[0014] Deriving the matrixed-surround compatible signal by post-processing a regular stereo down-mix has the advantage that the matrixed-surround compatibility processing can be fully reversed at a decoder side if the spatial parameters are available.

[0015] Although both of the approaches are suited to transmit a multi-channel signal, there are specific drawbacks of state of the art systems. Matrixed-surround methods are very efficient (since no additional parameters are required) at the price of a very limited multi-channel reconstruction quality.

[0016] Parametric multi-channel approaches on the other hand require a higher bit-rate due to the side information, which becomes a problem when a limit is set as a maximum acceptable bit-rate for the parametric representation. When the encoded parameters require a comparatively high amount of bit-rate, the only possible way to stay within such a bit-rate limit is to decrease the quality of an encoded down-mix channel by increasing the compression of the channel. Hence, the result is a general loss in audio quality, which may be unacceptably high. In other words, for parametric multi-channel approaches, there is often a hard limit of the minimum bit-rate that is required for the spatial parameter layer, which may in some cases be unacceptably high.

[0017] Although principle backwards compatibility between matrixed-surround methods and spatial audio methods can be achieved by a prior art encoder as illustrated in Fig. 9, no additional bit-rate can be saved with this approach when only matrix-based decoding is required. Even then the full set of spatial parameters has to be transmitted, wasting transmission bandwidth.

[0018] Whereas the bit-rate that has to be spent when applying the parametric method may be too high in case of certain application scenarios, the audio quality delivered by the methods without transmission of side-information might not be sufficient.

[0019] The US Patent Application 2005157883 is showing an apparatus for constructing a multi-channel audio signal using an input signal and parametric side information, the input signal including the first input channel and the second input channel derived from an original multi-channel signal, and the parametric side information describing interrelations between channels of the multi-channel original signal.

[0020] US Patent Application 2005/157883 A1 discloses an apparatus for constructing a multi-channel audio signal using an input signal and parametric side information, the input signal including the first input channel and the second input channel derived from an original multi-channel signal, and the parametric side information describing interrelations between channels

of the original multi-channel signal.

[0021] EP 1376538 A1 discloses a procedure, in which a part of the spectrum of two or more input signals is encoded using conventional coding techniques, while the rest of the spectrum is encoded using binaural cue coding (BCC). In BCC coding, spectral components of the input signals are downmixed and BCC parameters (e.g., inter-channel level and/or time differences) are generated. In a stereo implementation, after converting the left and right channels to the frequency domain, pairs of left- and right-channel spectral components are downmixed to mono. The mono components are then converted back to the time domain, along with those left- and right-channel spectral components that were not downmixed, to form hybrid stereo signals, which can then be encoded using conventional coding techniques. For playback, the encoded bitstream is decoded using conventional decoding techniques. BCC synthesis techniques may then apply the BCC parameters to synthesize an auditory scene based on the mono components as well as the unmixed stereo components.

Summary of the Invention

[0022] It is the object of the present invention to provide a concept for more efficient coding of multi-channel audio signals while being backwards compatible to matrix-based coding solutions.

[0023] This object is achieved by a multi-channel audio decoder of claim 1.

[0024] The present invention is based on the finding that a multi-channel audio signal can be efficiently represented by a parametric representation, when a first deriving rule is used for deriving first parametric data of the parametric representation describing a first portion of the multi-channel signal, and when for a second portion of the multi-channel signal second parametric data or no parametric data is included in the parametric representation, whereas the second parametric data is requiring less information units than the first parametric data when describing an identical portion of the multi-channel signal.

[0025] Thus, a first portion of the multi channel signal is represented by first parameters allowing for a reconstruction of the multi channel signal with higher quality and a second portion can be represented by second parameters allowing for a reconstruction with slightly lower quality. The bit-rate consumed by the first parametric data is consequently higher than the bit rate consumed by the second parametric data when both parametric data is to describe the same portion of a multi-channel signal. In other words, the first parameters require more bit rate per signal portion than the second parameters.

[0026] The purpose of the invention is to bridge the gap between both prior art worlds by gradually improving the sound of the up-mix signal while raising the bit-rate consumed by the side-information starting from 0 up to the bit-rates of the parametric methods. That is, the present invention aims at bridging the gap in bit-rates and perceptual quality between fully parametric methods and matrixed-surround methods. More specifically, it provides a method of flexibly choosing an "operating point" somewhere between matrixed-surround (no side-information, limited audio quality) and fully parametric reconstruction (full side-information rate required, good quality). This operating point can be chosen dynamically (i.e. varying in time) and in response to the permissible side-information rate, as it is dictated by the individual application.

[0027] By dynamically choosing the size of the first portion of the multi-channel audio signal which is the part of the multi-channel audio signal that is represented by the spatial audio parameters, the demanded bit-rate can be varied within a broad range. Representing major parts of a multi-channel signal by the spatial audio parameters will consume a comparatively

high bit-rate at the benefit of a good perceptual quality. Since for the second portion of the multi-channel audio signal a parameter deriving rule is chosen that results in parameters consuming less bit-rate, the resulting total bit-rate can be decreased by increasing the size of the second portion of the multi-channel signal. In a preferred embodiment of the present invention, no parametric data at all is transmitted for the second portion of the multi-channel signal, which is of course most bit-saving. Therefore, by dynamically shifting the size of the first portion with respect to the size of the second portion, the bit-rate (or the perceptual quality) can be dynamically adjusted to the needs.

[0028] A down-mix signal is derived in a matrix compatible way. Therefore, the first portion of the multi-channel audio signal can be reproduced with high perceptual quality using the spatial audio parameters and the second portion of the multi-channel signal can be reproduced using matrix-based solutions. This allows for a high-quality reproduction of parts of the signals requiring higher quality. At the same time, the overall bit-rate is decreased by relying on a matrix-based reproduction for signal parts less vital for the quality of a reproduced signal. In the present invention, the inventive concept is applied on the decoder side within a QMF representation of a received down-mix signal. The up-mixing process can principally be sub-divided into three steps:

Pre-processing of the input signals (down-mix signals received in the QMF domain) by application of a pre-de-correlator matrix;

de-correlation of part of the pre-processed signals; and mixing the thus derived signals (pre-processed signals and de-correlated signals) within a mix matrix, the output of the mixing being the channels of the up-mix signal.

[0029] Both, the pre-de-correlator matrix as well as the mixed-matrix are two-dimensional matrices with the dimensions "number of time slots" on the one hand and "number of parameter bands" on the other hand. Within a decoding process, the elements of these matrices are filled up with values that are derived from the parameters read from the spatial bit-stream, i.e. by the first parametric data. When the first parametric data is only received for a first portion of the multi-channel signal, only that portion of a reconstruction of a multi-channel signal can be derived using the first parametric data submitted. The matrix elements for deriving the second part of the reconstruction of the multi-channel signal are, according to the present invention, derived using matrix compatible coding schemes. These matrix elements can therefore either be derived based only on knowledge achieved from the down-mix signal or be replaced by predefined values.

[0030] In a preferred embodiment, a multi-channel audio decoder according to the present invention recognizes by the amount of the transmitted first parametric data, which part of the matrix or which part of the multi-channel audio signal is to be processed by the rule depending on the spatial parameters and which part is to be processed by the matrix based solution.

[0031] In another example, an audio encoder creates window information, indicating which parts of a multi-channel signal are being processed by the matrix based solution or by the spatial audio compatible approach. The window information is included in the parametric representation of a multi-channel signal.

[0032] An inventive decoder, therefore, is able to receive and to process the window information created to apply the appropriate up-mixing rules on the portions of the multi-channel audio signal indicated by the window information.

[0033] The inventive concept is applied in the QMF domain during the signal processing, i.e. in a domain where the signals are represented by multiple representations each representation

holding information on a certain frequency band.

[0034] In a further preferred embodiment of the present invention, the side-information free method (matrix based approach) is applied only to the higher frequency parts while applying (explicit) parametric information (i.e. the first encoding and decoding rule) for a proper reproduction of the low-frequency parts. This is advantageous due to the property of the human hearing to notice small deviations of two similar signals (e.g. phase deviations) a lot easier for low frequencies than for high frequencies.

[0035] A great benefit of the present invention is that a backwards compatibility of a spatial audio encoding and decoding scheme with matrix based solutions is achieved without having to introduce additional hard- or software when the encoding and decoding rules of the spatial audio coders are chosen appropriately.

[0036] Furthermore, the compatibility is achieved without having to transmit additional data, as it is the case in other prior art attempts. The coding scheme according to the present invention is furthermore extremely flexible, as it allows a seamless adjustment of the bit-rate or the quality, i.e. a smooth transition between full matrix based coding to full spatial audio coding of a given signal. That is, the coding scheme applied can be adjusted to the actual needs, either with respect to the required bit-rate or with respect to the desired quality.

Brief description of the drawings

[0037] Preferred embodiments of the present invention are subsequently described by referring to the enclosed drawings, wherein:

Fig.1
: shows an encoder;

Fig. 2
: shows an example of a parameter bit-stream created by the exemplary concept;

Fig. 2a
: shows a transcoder;

Fig. 3
: shows an inventive decoder;

Fig. 4
: shows an example of a spatial audio decoder implementing the inventive concept;

Fig. 5
: illustrates the use of the different coding schemes on a decoder side;

Fig. 6
: shows a prior art encoder;

Fig. 7
: shows a prior art decoder;

Fig.8
: shows a block diagram of a filterbank; and

Fig. 9
: shows a further example of a prior art encoder.

Detailed description of the preferred embodiments

[0038] Fig. 1 shows a multi-channel encoder. The multi-channel encoder 100 is having a parameter generator 102 and an output interface 104.

[0039] A multi-channel audio signal 106 is input into the encoder 100, where a first portion 108 and a second portion 110 of the multi-channel signal 106 are processed. The parameter generator 102 receives the first portion 108 and the second portion 110 and derives spatial parameters describing spatial properties of the multi-channel signal 106.

[0040] The spatial parameters are transferred to the output interface 104 that derives a parametric representation 112 of the multi-channel signal 106 such that the parametric representation 112 includes first parametric data for a first portion 108 of the multi-channel signal and wherein for a second portion 110 of the multi-channel signal 106 second parametric data requiring less information than the first parametric data or no parametric data is included in the parametric representation 112.

[0041] Several variations of the multi-channel encoder 100 are possible to achieve the same goal. For example, the parameter generator 102 can apply two different parameter deriving rules on the first portion 108 and on the second portion 110 that result in different parameter sets that are then transferred to the output interface 104 that combines the different parameter sets into the parametric representation 112. A special and preferred case is that for the second portion 110 no parameters are included in the parametric representation (and therefore not derived by the parameter generator 102) since on a decoder side the decoder derives the required decoding parameters by some heuristic rules.

[0042] Another possibility is that the parameter generator 102 derives a full set of spatial audio parameters as well for the first portion 108 as for the second portion 110. Hence, the output interface 104 would have to process the spatial parameters such that the second parametric data require less bits than the first parametric data.

[0043] Furthermore, the output interface 104 could add an additional window signal to the parametric representation 112 that shall signal to a decoder, how the multi-channel signal 106 was split into the first portion 108 and into the second portion 110 during the encoding. In a modification of the preferred embodiment of a multi-channel encoder 100, the multi-channel encoder 100 may additionally have a portion decider for deciding, which part of the multi-channel signal 106 is used as the first portion 108 and which part is used as the second portion 110, the decision being based on a quality criterion.

[0044] The quality criterion can be derived with respect to a resulting total bit-rate of the parametric representation 112 or with respect to quality aspects, taking into account the perceptual quality of a reproduction of the multi-channel signal 106 based on the parametric representation 112.

[0045] A major advantage is that the bit-rate consumed by the parametric representation can thus be varied in time, assuring that the quality criterion is met at any time during the encoding while allowing for an overall reduction of the required bit-rate compared to prior art methods.

[0046] Fig. 2 shows an example of a parametric representation 112 created by an encoder.

[0047] As mentioned above, the processing of the audio signals is done block-wise, i.e. a number of subsequent samples of the multi-channel signal in the time domain, building a so-called frame, is processed in one step. Fig. 2 shows a parameter bit-stream, i.e. a parametric representation for two consecutive frames. The parameter bit-stream is having a representation of a high-quality frame 120 and a representation of a lower quality frame 122. During the encoding of the high-quality frame 120, the decision was taken that the first portion 108 , which is being represented by parametric data has to be big compared to the second portion, which may for example be the case if the audio scene to encode is rather complex. The parameter bit-stream of Fig. 2 is furthermore created under the assumption that an exemplary encoder is used that does not derive any parametric data for the second portion

110 of the multi-channel signal 106. As can be seen in Fig. 2, 28 spatial parameters ICC and ICLD are included in the parametric representation to describe the high-quality frame 120. For example, the 28 spatial parameters describe the lower frequency bands of a QMF representation of the multi-channel signal.

[0048] The lower quality frame 122 comprises only 21 spatial parameter sets having ICC and ICLD parameters as this was found to be sufficient for the desired perceptual quality.

[0049] Fig. 2a shows a transcoder 150. The transcoder receives as an input an input bit stream 152 having a full set of spatial parameters describing a first frame 154 and a second frame 156 of a multi-channel audio signal.

[0050] The transcoder 150 generates a bit stream 158 holding a parametric representation representing the spatial properties of the multi-channel audio signal. In the example shown in Fig. 2a, the transcoder 150 derives the parametric representation such that for the first frame the number of parameters 160 is only slightly decreased. The number of parameters 162 describing the second frame corresponding to the input parameters 156 are strongly decreased, which reduces the amount of bit rate needed by the resulting parametric representation significantly. Such a transcoder 150 can therefore be used to post-process an already existing bit stream of spatial parameters to derive an inventive parametric representation requiring less bit rate during transmission or less storage space when stored on a computer-readable medium. It should be noted here that it is of course also possible to implement a transcoder for trans-coding in the other direction, i.e. using the parametric representation to generate spatial parameters.

[0051] The transcoder 150 can be implemented in various different ways, as for example by reducing the amount of parameters with a given rule or by additionally receiving the multi-channel audio signal to analyze the reduction of bit rate possible without disturbing the perceptual quality beyond an acceptable limit.

[0052] Fig. 3 shows an inventive multi-channel audio decoder 200 having a processor 202.

[0053] The processor is receiving as an input a down-mix signal 204 derived from a multi-channel audio signal, first parametric data 206 describing a first portion of the multi-channel signal and, for a second portion of the multi-channel signal, optional second parametric data 208 requiring less bits than the first parametric data 206. The processor 202 is deriving an intermediate signal 210 from the down-mix signal 204 using a first deriving rule for deriving a high-quality portion 212 of the intermediate signal, wherein the high-quality portion 212 of the intermediate signal 212 is corresponding to the first portion of the multi-channel audio signal. The processor 202 is using a second deriving rule for a second portion 214 of the intermediate signal 210, wherein the second deriving rule is using the second parametric data or no parametric data and wherein the first deriving rule is depending on the first parametric data 206.

[0054] The intermediate signal 210 derived by the processor 202 is built from a combination of the high-quality portion 212 and of the second portion 214.

[0055] The multi-channel audio decoder 200 may derive by itself, which portions of the down-mix signal 204 are to be processed with the first parametric data 206 by applying some appropriate rules, for example counting the number of spatial parameters included in the first parametric data 206. Alternatively, the processor 202 may be signaled the fractions of the high-quality portion 212 and of the second portion 214 within the down-mix signal 204 by some additional window information which is derived on an encoder side and that is additionally transmitted to the multi-channel audio decoder 200.

[0056] In a preferred embodiment, the second parametric data 208 is omitted and the processor 202 derives the second deriving rule from information already contained in the down-mix signal 204.

[0057] Fig. 4 shows a further embodiment of the present invention that combines the inventive feature of matrix compatibility in a spatial audio decoder. The multi-channel audio decoder 600 comprises a pre-de-correlator 601, a de-correlator 602 and a mix-matrix 603.

[0058] The multi-channel audio decoder 600 is a flexible device allowing to operate in different modes depending on the configuration of input signals 605 input into the pre-de-correlator 601. Generally, the pre-de-correlator 601 derives intermediate signals 607 that serve as input for the de-correlator 602 and that are partially transmitted unaltered to form, together with decorrelated signals calculated by the de-correlator 602, input signals 608. The input signals 608 are the signals input into the mix-matrix 603 that derives output channel configurations 610a or 610b, depending on the input channel configuration 605.

[0059] In a 1-to-5 configuration, a down-mix signal and an optional residual signal is supplied to the pre-de-correlator 601, that derives four intermediate signals (e1 to e4) that are used as an input of the de-correlator, which derives four de-correlated signals (d1 to d4) that form the input parameters 608 together with a directly transmitted signal m derived from the input signal.

[0060] It may be noted, that in the case where an additional residual signal is supplied as input, the de-correlator 602 that is generally working in a sub-band domain, may be operative to simply forward the residual signal instead of deriving a de-correlated signal. This may also be done in a frequency selective manner for certain frequency bands only.

[0061] In the 2-to-5 configuration the input signals 605 comprise a left channel, a right channel and optionally a residual signal. In that configuration, the pre-de-correlator matrix 601 derives a left, a right and a center channel and in addition two intermediate channels (e1, e2). Hence, the input signals to the mix-matrix 603 are formed by the left channel, the right channel, the center channel, and two de-correlated signals (d1 and d2).

[0062] In a further modification, the pre-de-correlator matrix may derive an additional intermediate signal (e5) that is used as an input for a de-correlator (D5) whose output is a combination of the de-correlated signal (d5) derived from the signal (e5) and the de-correlated signals (d1 and d2). In this case, an additional de-correlation can be guaranteed between the center channel and the left and the right channel.

[0063] The inventive audio decoder 600 implements the inventive concept in the 2-to-5 configuration. The transmitted parametric representation is used in the pre-de-correlation matrix 601 and in the mix-matrix 603. There, the inventive concept can be implemented in different ways as shown in more detail in Fig. 5.

[0064] Fig. 5 shows the pre-de-correlator, implemented as pre-decorrelator-matrix 601 and the mix-matrix 603 in a principle sketch, wherein the other components of the multi-channel audio decoder 600 are omitted.

[0065] The matrix used to perform the pre-de-correlation and the mixing has columns that represent time slots, i.e. the individual time samples of a signal and rows that represent the different parameter bands, i.e. each row is associated with one parameter band of an audio signal.

[0066] According to the concept of the present invention, the matrix elements of the matrices 601 and 603 are only partly derived from transmitted parametric data, wherein the remaining

matrix elements are derived by the decoder, based for example on knowledge of the down-mix signal. Fig. 5 shows one example where below a given frequency border line 622 the elements of the pre-de-correlator matrix 601 and the mix-matrix 603 are derived from parameters 620 that are read from the bit-stream, i.e. based on information transmitted from the encoder. Above the frequency borderline 622 the matrix elements are derived in the decoder based on knowledge of the down-mix signal only.

[0067] The border frequency (or in general: the amount of matrix elements derived from transmitted data) can be freely adapted according to the quality and/or bit-rate constraints that have to be met for the particular application scenario.

[0068] It is preferred for the novel coding method outlined here that a side-information free up-mix process may be performed with the same structure that has been outlined in the MPEG Spatial Audio Coding Reference Model 0. This invention may consist in describing a method for side-information free up-mix, but preferably provides a method for seamless and advantageous combination of such concepts with methods for side-information assisted up-mix.

[0069] In contrary to the MPEG Spatial Audio Coding Reference Model 0, in the side-information free up-mix process the elements of the matrices M1 (601) and M2(603) are preferably not derived from data transmitted in a bit-stream but by different means without the help of side-information, e.g. by applying heuristic rules based only on knowledge achieved from the down-mix signal.

[0070] In this way it is possible to achieve a gradual scaling between both techniques - in terms of bit-rate as well as in terms of sound quality - by acquiring only parts of the matrices based on the transmitted parameters and applying the rules of the method without side-information to fill up the remaining parts. Conceptually speaking, this corresponds to transmitting for certain parts of the matrices the spatial parameters and for other parts generating them at the decoder.

[0071] The determination of the parts of matrices that are to be derived from either the one or the other method can be done in a lot of different ways, such as

- deriving the parts of the matrices below a given horizontal border line by one method and above this border line by the other method;
- deriving the parts of the matrices left of a given vertical border line by one method and right from this border line by the other method;
- determining arbitrary time-frequency tiles within both matrices the elements of which are derived by one method and deriving the elements of the remaining time-frequency tiles by means of the other method.

[0072] It has been detailed in the above paragraphs that it is advantageous to describe all frequency parts of a multi-channel signal up to a certain border frequency by spatial parameters whereas the remaining frequency parts of the multi-channel signal are not represented by spatial parameters. This takes into account the characteristics of the human ear that has a better perception of lower frequencies than of higher frequencies. Of course, the present invention is by no means limited to this splitting of the multi-channel signal into a first portion and a second portion as it may also be advantageous or appropriate to describe higher frequency parts of the signal with better accuracy. This may especially be the case when in the lower frequency region only little energy is contained in the signal since most of the energy is contained in a high-frequency domain of the audio signal. Due to masking effects the low-frequency part will be mostly dominated by the high frequency parts then and it may be advantageous to provide the possibility for a high-quality reproduction of the high-frequency

part of the signal.

REFERENCES CITED IN THE DESCRIPTION

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US2005157883A [0019]
- US2006157883A1 [0020]
- EP1376538A1 [0021]

Non-patent literature cited in the description

- **HERRE, J.PURNHAGEN, H.BREEBAART, J.FALLER, C.DISCH, S.KJOERLING, K.SCHUIJERS, EHILPERT, J.MYBURG, F.**The Reference Model Architecture for MPEG Spatial Audio CodingProc. 118th AES Convention, Barcelona, Spain, 2005, [0007]
- Text of Working Draft for Spatial Audio Coding (SAC)ISO/IEC JTC1/SC29/WG11 (MPEG), Document N7136, Busan, Korea, 2005, [0007]

PATENTKRAV

1. Multikanalaudiodekoder (200) til behandling af et nedmiksningsaudiosignal (204) afledt på en matrixkompatibel måde og til behandling af første parametriske data (206), som beskriver en første andel af et multikanalsignal, hvor ingen parametriske eller andre parametriske data (208) behandles for en anden andel af multikanalsignalet, hvor de andre parametriske data (208) kræver færre informationsenheder end de første parametriske data (206), når en identisk andel af multikanalsignalet beskrives, omfattende:

en processor (202) til afledning af et mellemsignal (210) fra nedmiksningsaudiosignalet (204), hvor nedmiksningsaudiosignalet (204) er i en QMF, kvadraturspejlfilterbank, -visning,

ved anvendelse af en første afledningsregel til afledning af en første andel af mellemsignalet (210), hvor den første andel af mellemsignalet (210) svarer til den første andel af multikanalaudiosignalet, hvor den første afledningsregel er afhængig af de første parametriske data (206); og

ved anvendelse af en anden afledningsregel til afledning af en anden andel af mellemsignalet (210), hvor den anden afledningsregel ikke anvender nogen parametriske data eller anvender de andre parametriske data (208),

hvor den første andel er en tidsandel eller en frekvensandel; og

hvor de første parametriske data omfatter et eller flere rumlige audioparametre valgt blandt følgende liste over rumlige audioparametre: ICC (interkanalkorrelation), CLD (kanalniveauforskel) og CPC (kanalprædiktionskoefficient).

2. Multikanalaudiodekoder ifølge krav 1, hvor processoren (202) er konfigureret til som en indlæsning at modtage nedmiksningssignalet (204) og de første

parametriske data (206) og til at opbygge mellemsignalet ved en kombination af den første andel af mellemsignalet (210), som er en andel med høj kvalitet, og den anden andel af mellemsignalet (210).

3. Multikanalaudiodekoder (200) ifølge krav 1, som kan anvendes til at behandle første parametriske data (206), som omfatter en beskrivelse af en tidsandel af multikanalaudiosignalet, hvor kun information om et forudbestemt frekvensbånd af multikanalaudiosignalet er indeholdt i beskrivelsen.

4. Multikanalaudiodekoder (200) ifølge krav 1, hvor processoren (202) kan anvendes til at anvende den første afledningsregel til afledning af en frekvensandel af mellemsignalet (210), som rækker fra en lavfrekvensgrænse til en højfrekvensgrænse.

5. Multikanalaudiodekoder (200) ifølge krav 4, som kan anvendes til supplerende at behandle en vinduesinformation, som signalerer mindst højfrekvensgrænsen.

6. Multikanalaudiodekoder (200) ifølge krav 4, som kan anvendes til at aflede mindst højfrekvensgrænsen fra de første parametriske data (206) ved anvendelse af en vinduesestimeringsregel.

7. Multikanalaudiodekoder (200) ifølge krav 6, hvor vinduesestimeringsreglen indbefatter tælling af mængden af parametriske data fremlagt for en signalandel af nedmiksningsaudiosignalet (204) og sammenligning af den talte mængde af parametriske data med en nominel mængde af parametriske data.

8. Multikanalaudiodekoder (200) ifølge krav 1, hvor processoren (202) kan anvendes til at beregne den anden afledningsregel fra informationen om nedmiksningsaudiosignalet (204) eller fra de andre parametriske data (208).

9. Multikanalaudiodekoder (200) ifølge krav 1, hvor processoren (202) kan anvendes til at anvende en foruddefineret regel som den anden afledningsregel.

10. Multikanalaudiodekoder (200) ifølge krav 1, som endvidere omfatter en dekorrelator til at aflede et dekorreleret signal fra mellemsignalet (210) ved anvendelse af et dekorrelationsfilter.

11. Multikanalaudiodekoder (200) ifølge krav 1, som endvidere omfatter en opmikser til beregning af en rekonstruktion af multikanalsignalet baseret på mellemsignalet (210)

ved anvendelse af en første opmiksningsregel til beregning af en første andel af rekonstruktionen af multikanalsignalet svarende til den første andel af multikanalsignalet, hvor den første opmiksningsregel er afhængig af de første parametriske data (206); og

ved anvendelse af en anden opmiksningsregel til beregning af en anden andel af rekonstruktionen af multikanalsignalet, hvor den anden opmiksningsregel ikke anvender nogen parametriske data eller anvender de andre parametriske data (208).

DRAWINGS

Drawing

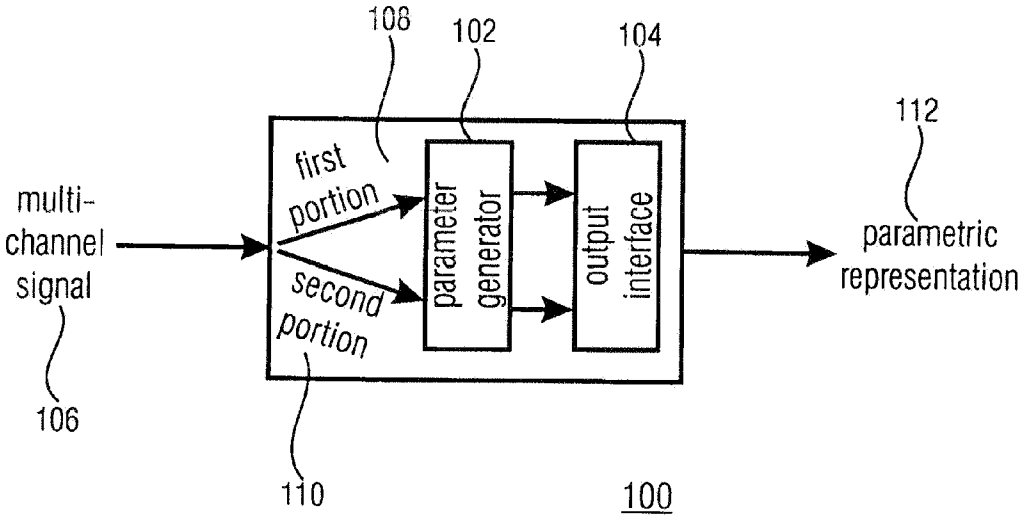


FIGURE 1

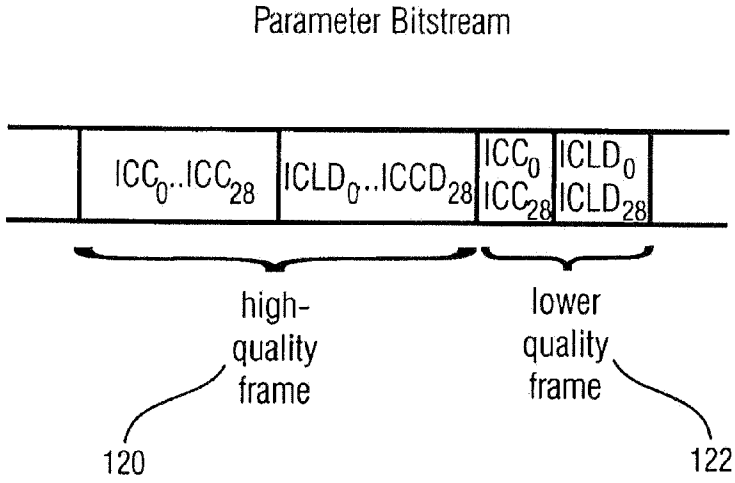

Parameter Bitstream
ICC0..ICC28
ICLD0..ICCD28
ICC0 ICC28
ICLD0 ICLD28
high-quality frame
lower quality frame
120
122

FIGURE 2

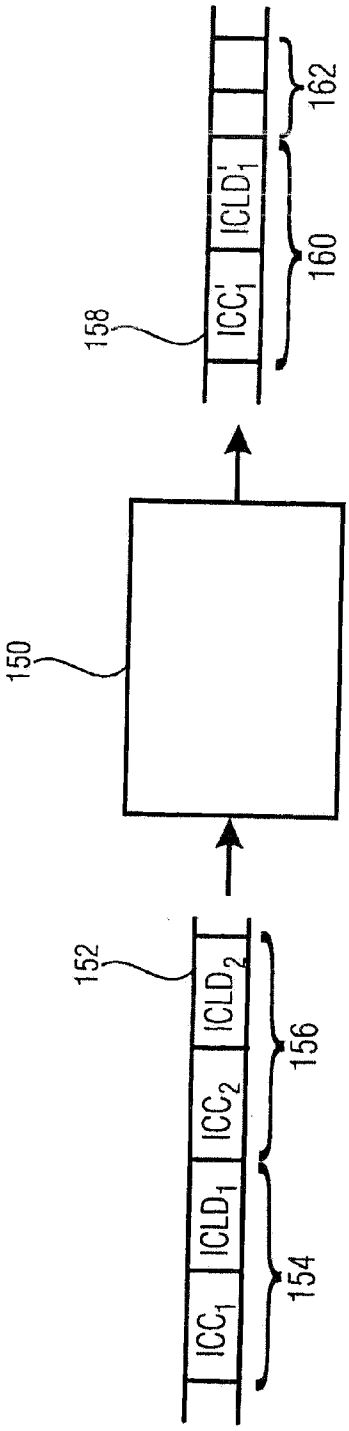


FIGURE 2A

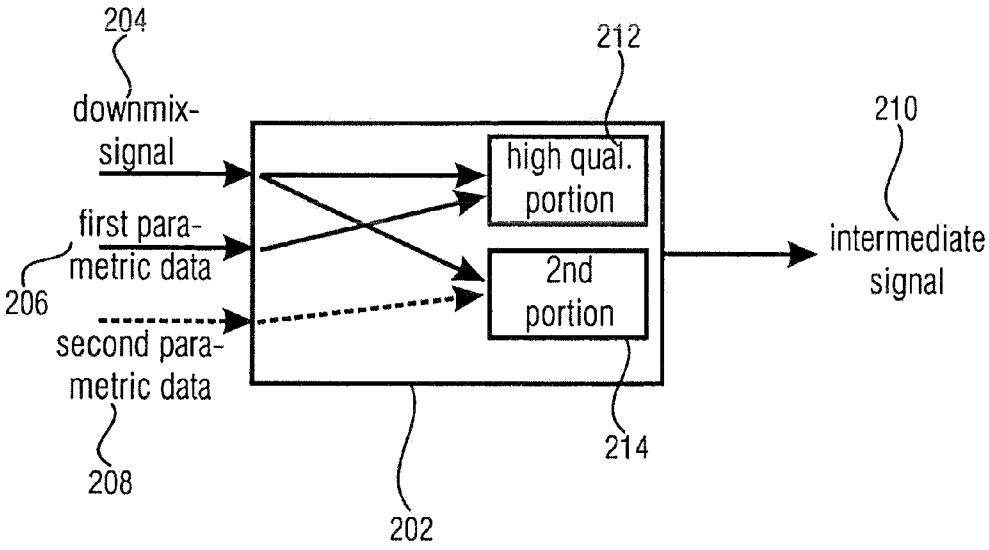
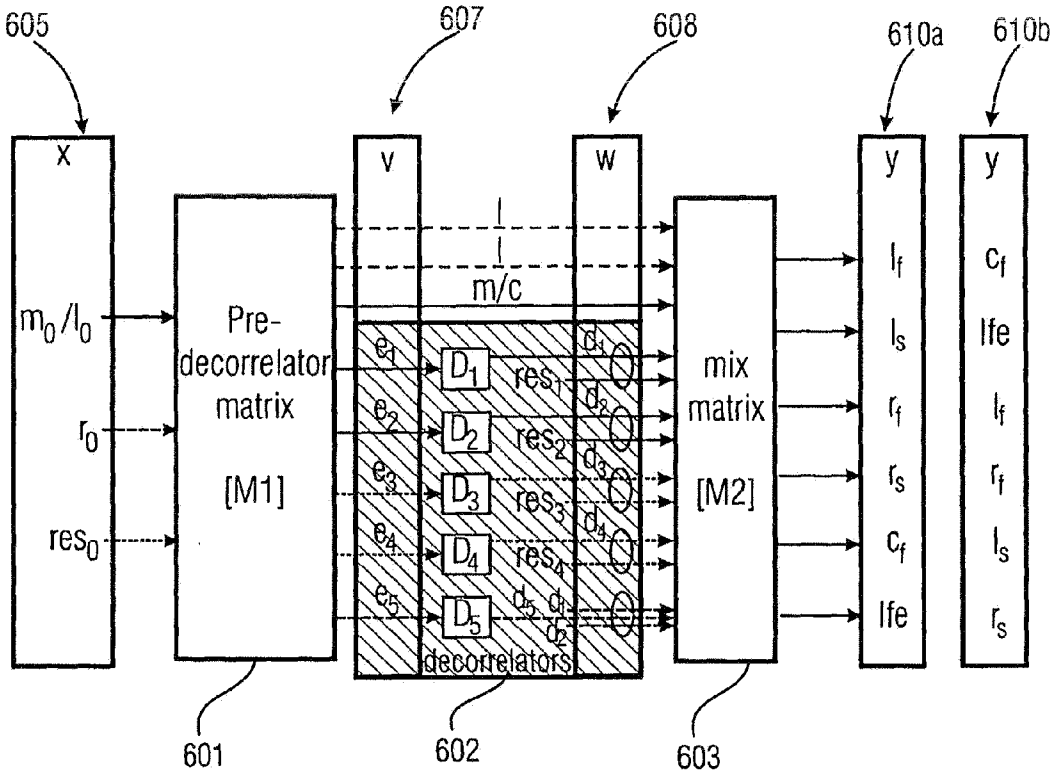

204
downmix-signal
212
high qual. portion
210
intermediate signal
first para-metric data
206
2nd portion
second para-metric data
208
214
202
200

FIGURE 3


605
607
608
610a
610b
x
v
w
y
y
m_0/l_0
r_0
res_0
Pre-decorrelator matrix
[M1]
l
l
m/c
e_1
e_2
e_3
e_4
e_5
D_1
D_2
D_3
D_4
D_5
res_1
res_2
res_3
res_4
d_5 d_1 d_2
decorrelators
mix matrix
[M2]
l_f
l_s
r_f
r_s
c_f
lfe
c_f
lfe
l_f
r_f
l_s
r_s
601
602
603

600

FIGURE 4

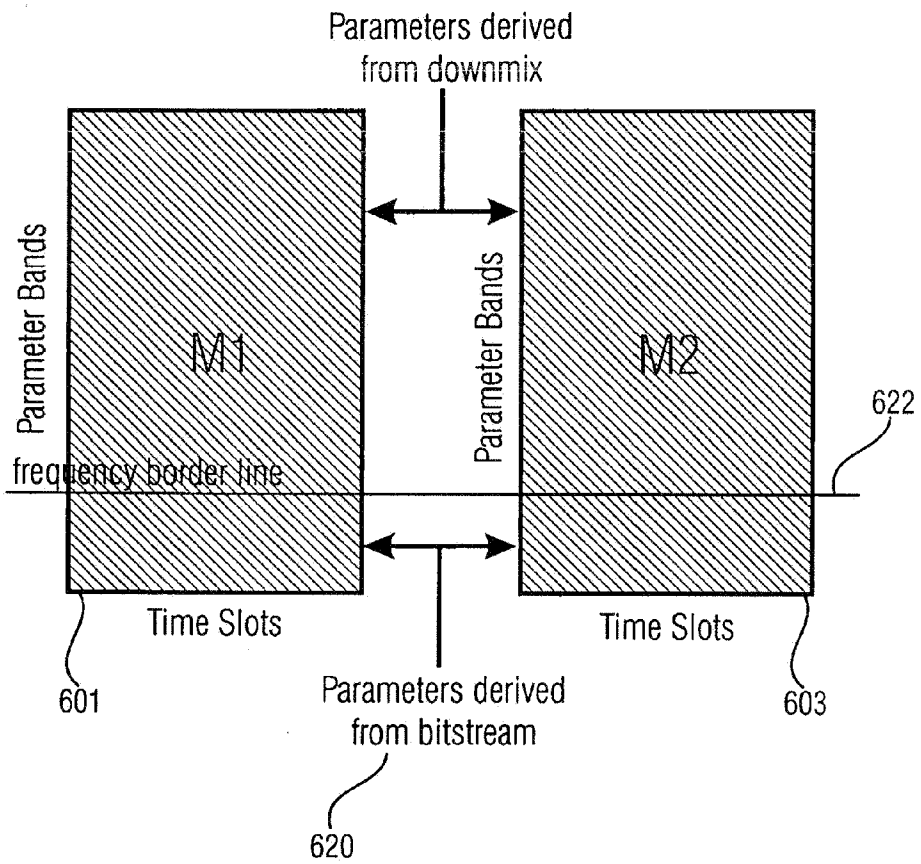

Parameters derived
from downmix
Parameter Bands
M1
Parameter Bands
M2
622
frequency border line
Time Slots
Time Slots
601
603
Parameters derived
from bitstream
620

FIGURE 5

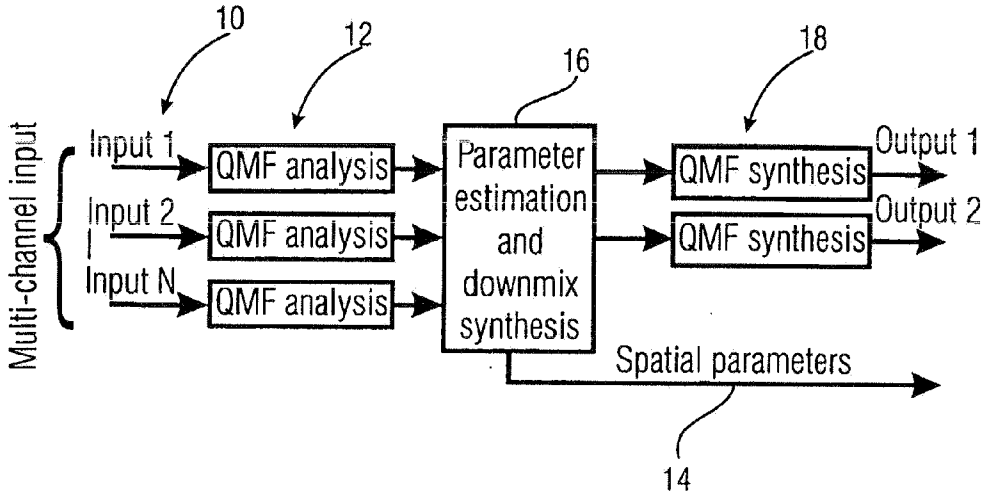


FIGURE 6

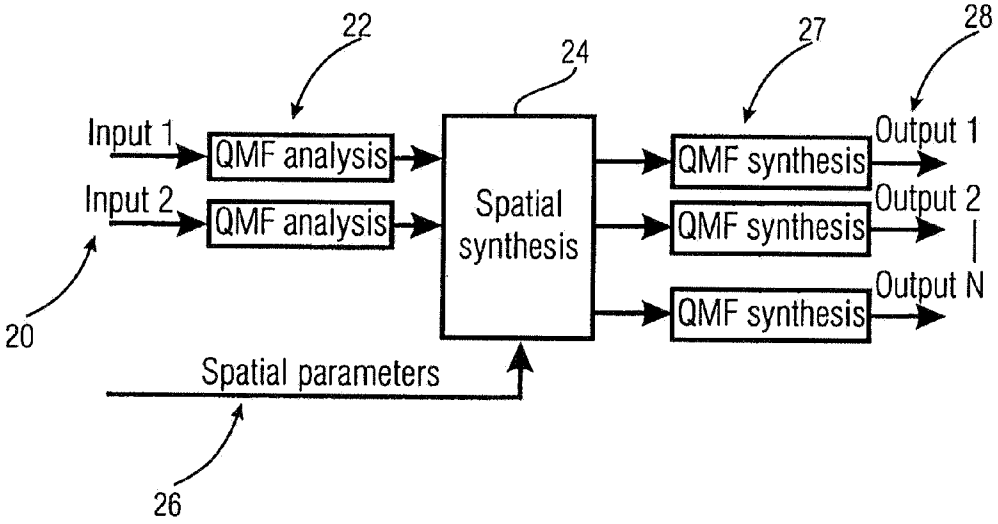


FIGURE 7

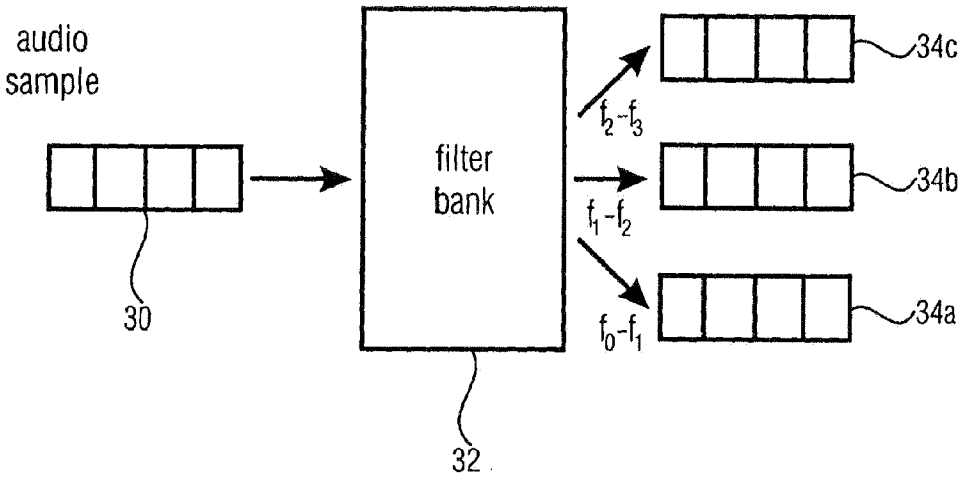


FIGURE 8

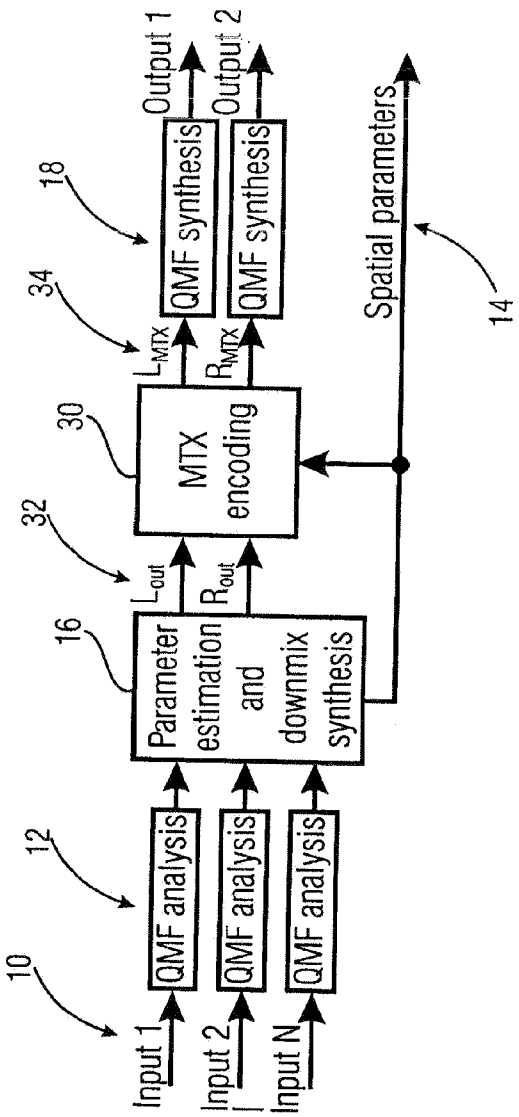

Input 1
Input 2
Input N
QMF analysis
QMF analysis
QMF analysis
Parameter estimation and downmix synthesis
L_{out}
R_{out}
MTX encoding
L_{MTX}
R_{MTX}
QMF synthesis
QMF synthesis
Output 1
Output 2
Spatial parameters
10
12
16
18
30
32
34
14

FIGURE 9