(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 November 2009 (26.11.2009)

PCT

(10) International Publication Number
WO 2009/141775 A1

(51) International Patent Classification:
G10L 19/00 (2006.01) *H04S 3/02* (2006.01)

(21) International Application Number:
PCT/IB2009/052009

(22) International Filing Date:
14 May 2009 (14.05.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
08156801.6 23 May 2008 (23.05.2008) EP

(71) Applicant *(for all designated States except US)*: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **SCHUIJERS, Erik, G., P.** [NL/NL]; c/o High Tech Campus Building 44, NL-5656 AE Eindhoven (NL).

(74) Agents: **UITTENBOGAARD, Frank** et al.; High Tech Campus Building 44, NL-5656 AE Eindhoven (NL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:
— *with international search report (Art. 21(3))*

(54) Title: A PARAMETRIC STEREO UPMIX APPARATUS, A PARAMETRIC STEREO DECODER, A PARAMETRIC STEREO DOWNMIX APPARATUS, A PARAMETRIC STEREO ENCODER

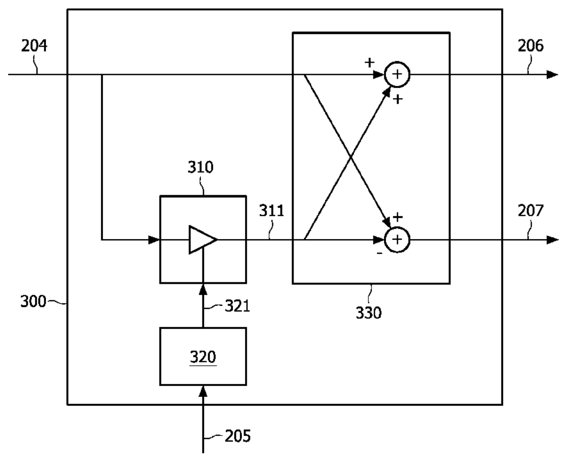

FIG. 3

(57) Abstract: A parametric stereo upmix apparatus (300, 400) generating a left signal (206) and a right signal (207) from a mono downmix signal (204) based on spatial parameters (205). Said parametric stereo upmix being characterized in that it comprises a means (310) for predicting a difference signal (311) comprising a difference between the left signal (206) and the right signal(207) based on the mono downmix signal (204) scaled with a prediction coefficient (321). Said prediction coefficient is derived from the spatial parameters (205). Said parametric stereo upmix apparatus (300, 400) further comprises an arithmetic means (330) for deriving the left signal (206) and the right signal (207) based on a sum and a difference of the mono downmix signal (204) and said difference signal (311).

WO 2009/141775 A1

A PARAMETRIC STEREO UPMIX APPARATUS, A PARAMETRIC STEREO DECODER, A PARAMETRIC STEREO DOWNMIX APPARATUS, A PARAMETRIC STEREO ENCODER

TECHNICAL FIELD

The invention relates to a parametric stereo upmix apparatus for generating a left signal and a right signal from a mono downmix signal based on spatial parameters. The invention further relates to a parametric stereo decoder comprising parametric stereo upmix apparatus, a method for generating a left signal and a right signal from a mono downmix signal based on spatial parameters, an audio playing device, a parametric stereo downmix apparatus, a parametric stereo encoder, a method for generating a prediction residual signal for a difference signal, and a computer program product.

TECHNICAL BACKGROUND

Parametric Stereo (PS) is one of the major advances in audio coding of the last couple of years. The basics of Parametric Stereo are explained in J. Breebaart, S. van de Par, A. Kohlrausch and E. Schuijers, "Parametric Coding of Stereo Audio", in *EURASIP J. Appl. Signal Process., vol 9*, pp. 1305-1322 (2004). Compared to traditional, a so-called discrete coding of audio signals, the PS encoder as depicted in Fig. 1 transforms a stereo signal pair *(l, r)* 101, 102 into a single mono downmix signal 104 plus a small amount of parameters 103 describing the spatial image. These parameters comprise Interchannel Intensity Differences (*iid*s), Interchannel Phase (or Time) Differences (*ipd*s/*itd*s) and Interchannel Coherence/Correlation (*icc*s). In the PS encoder 100 the spatial image of the stereo input signal *(l, r)* is analyzed resulting in *iid*, *ipd* and *icc* parameters. Preferably, the parameters are time and frequency dependent. For each time/frequency tile the *iid*, *ipd* and *icc* parameters are determined. These parameters are quantized and encoded 140 resulting in the PS bit-stream. Furthermore, the parameters are typically also used to control how the downmix of the stereo input signal is generated. The resulting mono sum signal (*s*) 104 is subsequently encoded using a legacy mono audio encoder 120. Finally the resulting mono and PS bit-stream are merged to construct the overall stereo bit-stream 107.

In the PS decoder 200 the stereo bit-stream is split into a mono bit-stream 202 and PS bit-stream 203. The mono audio signal is decoded resulting in a reconstruction of the mono downmix signal 204. The mono downmix signal is fed to the PS upmix 230 together with the decoded spatial image parameters 205. The PS upmix then generates the output stereo signal pair *(l, r)* 206, 207. In order to synthesize the *icc* cues, the PS upmix employs a so-called decorrelated signal (s_d), i.e., a signal is generated from the mono audio signal that has roughly the same spectral and temporal envelope, that however has a correlation of substantially zero with regard to the mono input signal. Then, based on the spatial image parameters, within the PS upmix for each time/frequency tile a 2x2 matrix is determined and applied:

$$\begin{bmatrix} l \\ r \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} s \\ s_d \end{bmatrix},$$

where H_{ij} represents an *(i, j)* upmix matrix *H* entry. The *H* matrix entries are functions of the PS parameters *iid*, *icc* and optionally *ipd*/*opd*. In the state-of-the-art PS system in case *ipd*/*opd* parameters are employed, the upmix matrix *H* can be decomposed as:

$$\begin{bmatrix} l \\ r \end{bmatrix} = \begin{bmatrix} e^{j\varphi_1} & 0 \\ 0 & e^{j\varphi_2} \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s \\ s_d \end{bmatrix},$$

where the left 2x2 matrix represents the phase rotations, a function of the *ipd* and *opd* parameters, and the right 2x2 matrix represents the part that reinstates the *iid* and *icc* parameters.

In WO2003090206 A1 it is proposed to equally distribute the *ipd* over the left and right channels in the decoder. Furthermore, it is proposed to generate a downmix signal by rotating the left and right signals both towards each other by half the measured *ipd* to obtain alignment. In practice, in case of nearly out of phase signals, this results for, both, the downmix generated in the encoder as well as the upmix generated in the decoder that the *ipd* over time varies slightly around 180 degrees, which due to wrapping may consist of a sequence of angles such as 179, 178, -179, 177, -179, ... As result of these jumps subsequent time/frequency tiles in the downmix exhibits phase discontinuities or in other words phase instability. Due to the inherent overlap-add synthesis structure this results in audible artefacts.

As an example, consider the downmix where in the one time/frequency tile the downmix is generated as:

$$s = le^{j(\pi/2-\varepsilon)} + re^{j(-\pi/2+\varepsilon)},$$

where ε is some arbitrary small angle, meaning that the *ipd* measured was close to 180 degrees, whereas for the next time-frequency tile the downmix is generated as:

$$s = le^{j(-\pi/2+\varepsilon)} + re^{j(\pi/2-\varepsilon)},$$

meaning that the measured *ipd* was close to -180 degrees. Using typical overlap-add synthesis a phase cancellation will occur in between the midpoints of the subsequent time/frequency tiles yielding artefacts.

A major disadvantage of the parametric stereo coding as discussed above is instability of a synthesis of the Interaural Phase Difference (*ipd*) cues in the PS decoder which are used in generating the output stereo pair. This instability has its source in phase modifications performed in the PS encoder in order to generate the downmix, and in the PS decoder in order to generate the output signal. As a result of this instability a lower audio quality of the output stereo pair is experienced.

In order to deal with this phase instability problem in practice the *ipd* synthesis is often discarded. However, this results in a reduced (spatial) audio quality of the reconstructed stereo signal.

Another alternative of dealing with this instability problem when *ipd* parameters are used is to incorporate so-called Overall Phase Differences (*opd*s) in the bitstream in order to provide the decoder with a phase reference. In this way the continuity over time/frequency tiles can be increased by allowing for a common phase rotation. This however happens at the expense of an increase of bitrate, and thus results in deterioration of the overall system performance.

SUMMARY OF THE INVENTION

It is an object of the invention to provide an enhanced parametric stereo upmix apparatus for generating a left signal and a right signal from a mono downmix signal that has improved audio quality of the generated left and right signals without additional bitrate increase, and does not suffer from the instabilities inferred by the interaural phase differences (*ipd*s) synthesis.

This object is achieved by a parametric stereo (PS) upmix apparatus comprising a means for predicting a difference signal comprising a difference between the left signal and the right signal based on the mono downmix signal scaled with a prediction coefficient. Said prediction coefficient is derived from the spatial parameters. Said PS upmix apparatus further comprises an arithmetic means for deriving the left signal and the right signal based on a sum and a difference of the mono downmix signal and said difference signal.

The proposed PS upmix apparatus offers a different way of derivation of the left signal and the right signal to this of the known PS decoder. Instead of applying the spatial parameters to reinstate the correct spatial image in a statistical sense as done in the known PS decoder, the proposed PS upmix apparatus constructs the difference signal from the mono downmix signal and the spatial parameters. Both the known and the proposed PS aim at reinstating the correct power ratios (*iids*), cross correlations (*iccs*) and phase relations (*ipds*). However, the known PS decoder does not strive to obtain the most accurate waveform match. Instead it ensures that the measured encoder parameters statistically match to the reinstated decoder parameters. In the proposed PS upmix by simple arithmetic operations, such as a sum and a difference, applied to the mono downmix signal and the estimated difference signal the left signal and the right signal are obtained. Such construction gives much better results for the quality and stability of the reconstructed left and right signals since it provides a close waveform match reinstating the original phase behavior of the signal.

In an embodiment, said prediction coefficient is based on waveform matching the downmix signal onto the difference signal. Waveform matching as such does not suffer from instabilities as the statistical approach used in known PS decoder for *ipd* and *opd* synthesis does since it inherently provides phase preservation. Thus by using the difference signal derived as a (complex-valued) scaled mono downmix signal and deriving the prediction coefficient based on waveform matching the source of instabilities of the known PS decoder is removed. Said waveform matching comprises e.g. a least-squares match of the mono downmix signal onto the difference signal, calculating the difference signal as:

$$d = \alpha \cdot s,$$

where s is the downmix signal and α is the prediction coefficient. It is well known that the least-squares prediction solution is given by:

$$\alpha = \frac{\langle s, d \rangle^*}{\langle s, s \rangle},$$

where $\langle s, d \rangle^*$ represents the complex conjugate of the cross correlation of the downmix and the difference signal and $\langle s, s \rangle$ represents the power of the downmix signal.

In a further embodiment, the prediction coefficient is given as a function of the spatial parameters:

$$\alpha = \frac{iid - 1 - j \cdot 2 \cdot \sin(ipd) \cdot icc \cdot \sqrt{iid}}{iid + 1 + 2 \cdot \cos(ipd) \cdot icc \cdot \sqrt{iid}}$$

whereby *iid*, *ipd*, and *icc* are the spatial parameters, and *iid* is an interchannel intensity difference, *ipd* is an interchannel phase difference, and *icc* is an interchannel coherence. It is generally difficult to quantize the complex-valued prediction coefficient α in a perceptually meaningful sense since the required accuracy depends on the properties of the left and right audio signals to be reconstructed. Hence, the advantage of this embodiment is that in contrast to the complex prediction coefficient α , the required quantization accuracies for the spatial parameters are well known from psycho-acoustics. As such, optimal use of the psycho-acoustic knowledge can be employed to efficiently, i.e. with the least steps possible, quantize the prediction coefficient to lower the bit rate. Furthermore, this embodiment allows for upmixing using backward compatible PS content.

In a further embodiment, the means for predicting the difference signal are arranged to enhance the difference signal by adding a scaled decorrelated mono downmix signal. Since in general it is not possible to completely predict the original encoder difference signal from the mono downmix signal, it gives a rise to a residual signal. This residual signal has no correlation with the downmix signal as otherwise it would have been taken into account by means of the prediction coefficient. In many cases the residual signal comprises a reverberant sound field of a recording. The residual signal can be effectively synthesized using a decorrelated mono downmix signal, derived from the mono downmix signal.

In a further embodiment, said decorrelated mono downmix is obtained by means of filtering the mono downmix signal. The goal of this filtering is to effectively generate a signal with a similar spectral and temporal envelope as the mono downmix signal, but with a correlation substantially close to zero such that it corresponds to a synthetic variant of the residual component derived in the encoder. This can e.g. be achieved by means of allpass filtering, delays, lattice reverberation filters, feedback delay networks or a combination thereof. Additionally, power normalization can be applied to the decorrelated signal in order to ensure that the power for each time/frequency tile of the decorrelated signal closely corresponds to that of the mono downmix signal. In this way it is ensured that the decoder output signal will contain the correct amount of decorrelated signal power.

In a further embodiment, a scaling factor applied to the decorrelated mono downmix is set to compensate for a prediction energy loss. The scaling factor applied to the decorrelated mono downmix ensures that the overall signal power of the left signal and right signal at the decoder side matches the signal power of the left and right signal power at the encoder side, respectively. As such the scaling factor β can also be interpreted as a prediction energy loss compensation factor.

In a further embodiment, the scaling factor applied to the decorrelated mono downmix is given as a function of the spatial parameters:

$$\beta = \sqrt{\frac{iid + 1 - 2 \cdot \cos(ipd) \cdot icc \cdot \sqrt{iid}}{iid + 1 + 2 \cdot \cos(ipd) \cdot icc \cdot \sqrt{iid}} - |\alpha|^2}$$

whereby *iid*, *ipd*, and *icc* are the spatial parameters, and *iid* is an interchannel intensity difference, *ipd* is an interchannel phase difference, *icc* is an interchannel coherence, and α is the prediction coefficient. Similarly as in case of the prediction coefficient, expressing the decorrelated scaling factor β as a function of the spatial parameters enables the use of the knowledge about the required quantization accuracies of these spatial parameters. As such, optimal use of the psycho-acoustic knowledge can be employed to lower the bit rate.

In a further embodiment, said parametric stereo upmix has a prediction residual signal for the difference signal as an additional input, whereby the arithmetic means are arranged for deriving the left signal and the right signal also based on said prediction residual signal for the difference signal. To avoid long names of signals a prediction residual signal is used for the prediction residual signal for the difference signal throughout the remainder of the patent application. The prediction residual signal operates as a replacement for the synthetic decorrelation signal by its original encoder counterpart. It allows reinstating the original stereo signal in the decoder. This however is at the cost of additional bitrate since the prediction signal needs to be encoded and transmitted to the decoder. Therefore, typically the bandwidth of the prediction residual signal is limited. The prediction residual signal can either completely replace the decorrelated mono downmix signal for a given time/frequency tile or it can work in a complementary fashion. The latter can be beneficial in case the prediction residual signal is only sparsely coded, e.g. only a few of the most significant frequency bins are encoded. In that case, compared to the encoder situation, still energy will be missing. This lack of energy will be filled by the decorrelated signal. A new decorrelated scaling factor β' is then calculated as:

$$\beta' = \sqrt{\beta^2 - \frac{\langle d_{res,cod}, d_{res,cod} \rangle}{\langle s, s \rangle}},$$

where $\langle d_{res,cod}, d_{res,cod} \rangle$ is the signal power of the coded prediction residual signal and $\langle s, s \rangle$ is the power of the mono downmix signal. These signal powers can be measured at the decoder side and thus need not need to be transmitted as signal parameters.

The invention further provides a parametric stereo decoder comprising said parametric stereo upmix apparatus and an audio playing device comprising said parametric stereo decoder.

The invention also provides a parametric stereo downmix apparatus and a parametric stereo encoder comprising said parametric stereo downmix apparatus.

The invention further provides method claims as well as a computer program product enabling a programmable device to perform the method according to the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments shown in the drawings, in which:

Fig. 1 schematically shows an architecture of a parametric stereo encoder (prior art);

Fig. 2 schematically shows an architecture of a parametric stereo decoder (prior art);

Fig. 3 shows a parametric stereo upmix apparatus according to the invention, said parametric stereo upmix apparatus generating a left signal and a right signal from a mono downmix signal based on spatial parameters;

Fig. 4 shows the parametric stereo upmix apparatus comprising a prediction means being arranged to enhance the difference signal by adding a scaled decorrelated mono downmix signal;

Fig. 5 shows the parametric stereo upmix apparatus having a prediction residual signal for the difference signal as an additional input;

Fig. 6 shows the parametric stereo decoder comprising the parametric stereo upmix apparatus according to the invention;

Fig. 7 shows a flow chart for a method for generating the left signal and the right signal from the mono downmix signal based on spatial parameters according to the invention;

Fig. 8 shows a parametric stereo downmix apparatus according to the invention, said parametric stereo downmix apparatus generating a mono downmix signal from the left signal and the right signal based on spatial parameters;

Fig. 9 shows the parametric stereo encoder comprising the parametric stereo downmix apparatus according to the invention.

Throughout the figures, same reference numerals indicate similar or corresponding features. Some of the features indicated in the drawings are typically implemented in software, and as such represent software entities, such as software modules or objects.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Fig. 3 shows a parametric stereo upmix apparatus 300 according to the invention. Said parametric stereo upmix apparatus 300 generates a left signal 206 and right signal 207 from a mono downmix signal 204 based on spatial parameters 205.

Said parametric stereo upmix apparatus 300 comprises a means 310 for predicting a difference signal 311 comprising a difference between the left signal 206 and the right signal 207 based on the mono downmix signal 204 scaled with a prediction coefficient 321, whereby said prediction coefficient 321 is derived from the spatial parameters 205 in a unit 320 and an arithmetic means 330 for deriving the left signal 206 and the right signal 207 based on a sum and a difference of the mono downmix signal 204 and said difference signal 311.

The left signal 206 and right signal 207 are preferably reconstructed as follows:

$$l = s + d,$$

$$r = s - d,$$

where s is the mono downmix signal, and d is the difference signal. This is under the assumption that the encoder sum signal is calculated as:

$$s = \frac{l + r}{2}.$$

In practice gain normalization is often applied when constructing the left signal 206 and the right signal 207:

$$l = \frac{1}{2c} \cdot (s + d),$$

$$r = \frac{1}{2c} \cdot (s - d),$$

where c is a gain normalization constant and is a function of the spatial parameters. Gain normalization ensures that a power of the mono downmix signal 204 is equal to a sum of powers of the left signal 206 and the right signal 207. In this case the encoder sum signal was calculated as:

$$s = c \cdot (l + r).$$

The spatial parameters are determined in an encoder beforehand and transmitted to the decoder comprising a parametric stereo upmix 300. Said spatial parameters are determined on a frame-by-frame basis for each time/frequency tile as:

$$iid = \frac{\langle l,l \rangle}{\langle r,r \rangle},$$

$$icc = \frac{|\langle l,r \rangle|}{\sqrt{\langle l,l \rangle \cdot \langle r,r \rangle}},$$

$$ipd = \angle\langle l,r \rangle,$$

where *iid* is an interchannel intensity difference, *icc* is an interchannel coherence, *ipd* is an interchannel phase difference, and $\langle l,l \rangle$ and $\langle r,r \rangle$ are the left and right signal powers respectively and $\langle l,r \rangle$ represents the non-normalized complex-valued covariance coefficient between the left and right signals.

For a typical complex-valued frequency domain such as the DFT (FFT), these powers are measured as:

$$\langle l,l \rangle = \sum_{k \in k_{tile}} l[k] \cdot l^*[k],$$

$$\langle r,r \rangle = \sum_{k \in k_{tile}} r[k] \cdot r^*[k],$$

$$\langle l,r \rangle = \sum_{k \in k_{tile}} l[k] \cdot r^*[k],$$

where k_{tile} represents the DFT bins corresponding to a parameter band. It is to be noted that also other complex domain representation could be used, such as e.g. a complex exponentially modulated QMF bank as described in P. Ekstrand, "Bandwidth extension of audio signals by spectral band replication", in *Proc. 1st IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002)*, Leuven, Belgium, Nov. 2002, pp. 73 – 79.

For low frequencies up to 1.5-2 kHz the above equations hold. However, for higher frequencies the *ipd* parameters are not relevant for perception and therefore they are set to a zero value resulting in:

$$iid = \frac{\langle l,l \rangle}{\langle r,r \rangle},$$

$$icc = \frac{\Re\{\langle l,r \rangle\}}{\sqrt{\langle l,l \rangle \cdot \langle r,r \rangle}},$$

$$ipd = 0.$$

Alternatively, since at higher frequencies, rather the broadband envelope than the phase differences are important for perception, the *icc* is calculated as:

$$icc = \frac{\left|\langle l,r\rangle\right|}{\sqrt{\langle l,l\rangle \cdot \langle r,r\rangle}}.$$

The gain normalization constant c is expressed as:

$$c = \sqrt{\frac{iid+1}{iid+1+2\cdot icc\cdot \cos(ipd)\cdot \sqrt{iid}}}.$$

Since c may approach infinity due to left and right signals being out of phase, the value of the gain normalization constant c is typically limited as:

$$c = \min\left(\sqrt{\frac{iid+1}{iid+1+2\cdot icc\cdot \cos(ipd)\cdot \sqrt{iid}}}, c_{\max}\right),$$

with $c_{\max}$ being the maximum amplification factor, e.g. $c_{\max} = 2$.

In an embodiment, said prediction coefficient is based on estimating the difference signal 311 from the mono downmix signal 204 using waveform matching. Said waveform matching comprises e.g. a least-squares match of the mono downmix signal 204 onto the difference signal 311, resulting in the difference signal provided as:

$$d = \alpha \cdot s,$$

where s is the mono downmix signal 204 and α is the prediction coefficient 321.

Beside the least-squares matching a waveform matching using a different norm from L_2-norm can be used. Alternatively, the p-norm error $\left\| d - \alpha \cdot s \right\|^p$ could be e.g. perceptually weighted. However, the least-squares matching is advantageous as it results in relatively simple calculations for deriving the prediction coefficient from the transmitted spatial image parameters.

It is well known that the least-squares prediction solution for the prediction coefficient α is given by:

$$\alpha = \frac{\langle s,d\rangle^*}{\langle s,s\rangle},$$

where $\langle s,d\rangle^*$ represents the complex conjugate of the cross correlation of the mono downmix signal 204 and the difference signal 311 and $\langle s,s\rangle$ represents the power of the mono downmix signal.

In a further embodiment, the prediction coefficient 321 is given as a function of the spatial parameters:

$$\alpha = \frac{iid - 1 - j \cdot 2 \cdot \sin(ipd) \cdot icc \cdot \sqrt{iid}}{iid + 1 + 2 \cdot \cos(ipd) \cdot icc \cdot \sqrt{iid}}.$$

Said prediction coefficient is calculated in unit 320 according to the above formula.

Fig. 4 shows the parametric stereo upmix apparatus 300 comprising a prediction means 310 being arranged to enhance the difference signal by adding a scaled decorrelated mono downmix signal. The mono downmix signal 204 is provided to the unit 340 for decorrelating. As a result the decorrelated mono downmix signal 341 is provided at the output of the unit 340. In the prediction means 310 a first part of the difference signal is calculated by scaling the mono downmix signal 204 with the prediction coefficient 321. Additionally the decorrelated mono downmix signal 341 is also scaled in the prediction means 310 with the scale factor 322. A resulting second part of the difference signal is consequently added to the first part of the difference signal resulting in the enhanced difference signal 311. The mono downmix signal 204 and the enhanced difference signal 311 are provided to the arithmetic means 330, which calculate the left signal 206 and the right signal 207.

In general it is not possible to accurately predict the difference signal from the mono downmix signal by just scaling with the prediction coefficient. This gives rise to a residual signal $d_{res} = d - \alpha \cdot s$. This residual signal has no correlation with the downmix signal as otherwise it would have been taken into account by means of the prediction coefficient. In many cases the residual signal comprises a reverberant sound field of a recording. The residual signal is effectively synthesized using a decorrelated mono downmix signal, derived from the mono downmix signal. Said decorrelated signal is the second part of the difference signal that is calculated in the prediction means 310.

In a further embodiment, said decorrelated mono downmix 341 is obtained by means of filtering the mono downmix signal 204. Said filtering is performed in the unit 340. This filtering generates a signal with a similar spectral and temporal envelope as the mono downmix signal 204, but with a correlation substantially close to zero such that it corresponds to a synthetic variant of the residual component derived in the encoder. This effect is achieved by means of e.g. allpass filtering, delays, lattice reverberation filters, feedback delay networks or a combination thereof.

In a further embodiment, a scaling factor 322 applied to the decorrelated mono downmix 341 is set to compensate for a prediction energy loss. The scaling factor 322 applied to the decorrelated mono downmix 341 ensures that the overall signal power of the left signal 206 and right signal 207 at the output of the parametric stereo upmix apparatus 300 matches the signal power of the left and right signal power at the encoder side, respectively. As such the scaling factor 322 indicated further as β is interpreted as a prediction energy loss compensation factor. The difference signal d is then expressed as:

$$d = \alpha \cdot s + \beta \cdot s_d ,$$

where s_d is the decorrelated mono downmix signal.

It can be shown that said scaling factor 322 can be expressed as:

$$\beta = \sqrt{\frac{\langle d,d \rangle}{\langle s,s \rangle} - |\alpha|^2}$$

in terms of signal powers corresponding to the difference signal d and the mono downmix signal s.

In a further embodiment, the scaling factor 322 applied to the decorrelated mono downmix 341 is given as a function of the spatial parameters 205:

$$\beta = \sqrt{\frac{iid + 1 - 2 \cdot \cos(ipd) \cdot icc \cdot \sqrt{iid}}{iid + 1 + 2 \cdot \cos(ipd) \cdot icc \cdot \sqrt{iid}} - |\alpha|^2} .$$

Said scaling factor 322 is derived in unit 320.

In case, no downmix normalization was applied in the encoder, i.e., the downmix signal was calculated as $s = \frac{1}{2}(l + r)$, the left signal 206 and the right signal 207 are then expressed as:

$$\begin{bmatrix} l \\ r \end{bmatrix} = \begin{bmatrix} 1+\alpha & \beta \\ 1-\alpha & -\beta \end{bmatrix} \begin{bmatrix} s \\ s_d \end{bmatrix} .$$

In case downmix normalization was applied, i.e., the downmix signal was calculated as $s = c(l + r)$, the left signal 206 and the right signal 207 are expressed as:

$$\begin{bmatrix} l \\ r \end{bmatrix} = \begin{bmatrix} 1/2c & 0 \\ 0 & 1/2c \end{bmatrix} \begin{bmatrix} 1+\alpha & \beta \\ 1-\alpha & -\beta \end{bmatrix} \begin{bmatrix} s \\ s_d \end{bmatrix} .$$

Fig. 5 shows the parametric stereo upmix apparatus 500 having a prediction residual signal for the difference signal 331 as an additional input. The arithmetic means 330 are arranged for deriving the left signal 206 and the right signal 207 based on the mono downmix signal 204, the difference signal 311, and said prediction residual signal 331. The means 310 predict a difference signal 311 based on the mono downmix signal 204 scaled with a prediction coefficient 321. Said prediction coefficient 321 is derived in the unit 320 based on the spatial parameters 205.

The left signal 206 and the right signal 207, respectively, are given as:

$$l = s + d + d_{res},$$

$$r = s - d - d_{res},$$

where d_{res} is the prediction residual signal.

Alternatively, in case power normalization was applied to the downmix, but not to the residual signal the left signal and the right signal can be derived as:

$$l = \frac{1}{2c} \cdot (s + d) + d_{res},$$

$$r = \frac{1}{2c} \cdot (s - d) - d_{res}.$$

The prediction residual signal 331 operates as a replacement for the synthetic decorrelation signal 341 by its original encoder counterpart. It allows reinstating the original stereo signal by the parametric stereo upmix apparatus 300. The prediction residual signal 331 can either completely replace the decorrelated mono downmix signal 341 for a given time/frequency tile or it can work in a complementary fashion. The latter is beneficial in case the prediction residual signal is only sparsely coded, e.g. only a few of most significant frequency bins are encoded. In this case energy still is missing as compared with the encoder prediction residual signal. This lack of energy is filled by the decorrelated signal 341. A new decorrelated scaling factor β' is then calculated as:

$$\beta' = \sqrt{\beta^2 - \frac{\langle d_{res,cod}, d_{res,cod} \rangle}{\langle s, s \rangle}},$$

where $\langle d_{res,cod}, d_{res,cod} \rangle$ is the signal power of the coded prediction residual signal and $\langle s, s \rangle$ is the power of the mono downmix signal 204.

The parametric stereo upmix apparatus 300 can be used in the state of the art architecture of the parametric stereo decoder without any additional adaptations. The parametric stereo upmix apparatus 300 replaces then the upmix unit 230 as depicted in Fig. 2. When the prediction residual signal 331 is used by the parametric stereo upmix 400 a couple of adaptations are required, which are depicted in Fig. 6.

Fig. 6 shows the parametric stereo decoder comprising the parametric stereo upmix apparatus 400 according to the invention. A parametric stereo decoder comprises a de-multiplexing means 210 for splitting the input bitstream into a mono bitstream 202, a prediction residual bitstream 332, and parameter bitstream 203. A mono decoding means 220 decode said mono bitstream 202 into a mono downmix signal 204. The mono decoding means is further configured to decode the prediction residual bitstream 332 into the prediction residual signal 331. A parameter decoding means 240 decode the parameter bitstream 203 into spatial parameters 205. The parametric stereo upmix apparatus 400 generates a left signal 206 and a right signal 207 from the mono downmix signal 204 and the prediction residual signal 331 based on spatial parameters 205. Although the decoding of the mono downmix signal 204 and the prediction residual signal is performed by the decoding means 220, it is possible that said decoding is performed by a separate decoding software and/or hardware for each of the signals to be decoded.

Fig. 7 shows a flow chart for a method for generating the left signal 206 and the right signal 207 from the mono downmix signal 204 based on spatial parameters according to the invention. In a first step 710 a difference signal 311 comprising a difference between the left signal 206 and the right signal 207 is predicted based on the mono downmix signal 204 scaled with a prediction coefficient 321, whereby said prediction coefficient is derived from the spatial parameters 205. In a second step 720 the left signal 206 and the right signal 207 are derived based on a sum and a difference of the mono downmix signal 204 and said difference signal 311.

When the prediction residual signal is available in the second step 720 the prediction residual signal next to the mono downmix signal 204 and the difference signal 311 is used to derive the left signal 206 and the right signal 207.

When the parametric stereo upmix 300 is used in the parametric stereo decoder no modifications to the parametric stereo encoder are required. The parametric stereo encoder as known in the prior art can be used.

However, when the parametric stereo upmix 400 is used the parametric stereo encoder must be adapted to provide the prediction residual signal in the bitstream.

Fig. 8 shows a parametric stereo downmix apparatus 800 according to the invention, said parametric stereo downmix apparatus generating a mono downmix signal from the left signal and the right signal based on spatial parameters. Said parametric stereo downmix apparatus 800 outputs next to the mono downmix signal 104 an additional signal 801, which is the prediction residual signal. Said parametric stereo downmix apparatus 800 comprises a further arithmetic means 810 for deriving the mono downmix signal 104 and a difference signal 811 comprising a difference between the left signal 101 and the right signal

102. Said parametric stereo downmix apparatus 800 comprises further a further prediction means 820 for deriving a prediction residual signal (for the difference signal) 801 as a difference between the difference signal 811 and the mono downmix signal 104 scaled with a predetermined prediction coefficient 831 derived from the spatial parameters 103. Said predetermined prediction coefficient is determined in a unit 830. The predetermined prediction coefficient is chosen to provide the prediction residual signal 801 that is orthogonal to the mono downmix signal 104. In addition power normalization of the downmix signal can be employed (not shown in Fig. 8).

Although the numbering of the signals corresponding to the mono downmix and the prediction residual have different reference numbers in the parametric stereo upmix apparatus and the parametric stereo downmix apparatus, it should be clear that the mono downmix signals 204 and 104 correspond to each other and the prediction residual signal 331 and 801 as well correspond to each other.

Fig. 9 shows the parametric stereo encoder comprising the parametric stereo downmix apparatus 800 according to the invention. Said parametric stereo encoder comprises:

- an estimation means 130 for deriving spatial parameters 103 from the left signal 101 and the right signal 102,
- a parametric stereo downmix apparatus 110 according to the invention for generating a mono downmix signal 104 from the left signal 101 and the right signal 102 based on spatial parameters 103,
- a mono encoding means 120 for encoding said mono downmix signal 104 into a mono bitstream 105, said mono encoding means 120 being further arranged to encode the prediction residual signal 801 into a prediction residual bitstream 802,
- a parameter encoding means 140 for encoding spatial parameters 103 into a parameter bitstream 106, and
- a multiplexing means 150 for merging the mono bitstream 105, the parameter bitstream 106 and the prediction residual bitstream 802 into an output bitstream 107.

Although the encoding of the mono downmix signal 104 and the prediction residual signal 801 is performed by the encoding means 120, it is possible that said encoding is performed by a separate decoding software and/or hardware for each of the signals to be encoded.

Furthermore, although individually listed, a plurality of means, elements or method steps may be implemented by e.g. a single unit or processor. Additionally, although individual features may be included in different claims, these may possibly be advantageously combined, and the inclusion in different claims does not imply that a combination of features is not feasible and/or advantageous. Also the inclusion of a feature in one category of claims does not imply a limitation to this category but rather indicates that the feature is equally applicable to other claim categories as appropriate. Furthermore, the order of features in the claims do not imply any specific order in which the features must be worked and in particular the order of individual steps in a method claim does not imply that the steps must be performed in this order. Rather, the steps may be performed in any suitable order. In addition, singular references do not exclude a plurality. Thus references to "a", "an", "first", "second" etc do not preclude a plurality. Reference signs in the claims are provided merely as a clarifying example shall not be construed as limiting the scope of the claims in any way.

CLAIMS:

1. A parametric stereo upmix apparatus (300, 400) for generating a left signal (206) and a right signal (207) from a mono downmix signal (204) based on spatial parameters (205), characterized in that said parametric stereo upmix apparatus (300, 400) comprises a means (310) for predicting a difference signal (311) comprising a difference between the left signal (206) and the right signal (207) based on the mono downmix signal (204) scaled with a prediction coefficient (321), whereby said prediction coefficient is derived from the spatial parameters (205), and an arithmetic means (330) for deriving the left signal (206) and the right signal (207) based on a sum and a difference of the mono downmix signal (204) and said difference signal (311).

2. A parametric stereo upmix apparatus as claimed in claim 1, whereby said prediction coefficient (321) is based on waveform matching the downmix signal (204) onto the difference signal (311).

3. A parametric stereo upmix apparatus as claimed in claim 2, whereby the prediction coefficient (321) is given as a function of the spatial parameters (205):

$$\alpha = \frac{iid - 1 - j \cdot 2 \cdot \sin(ipd) \cdot icc \cdot \sqrt{iid}}{iid + 1 + 2 \cdot \cos(ipd) \cdot icc \cdot \sqrt{iid}}$$

whereby *iid*, *ipd*, and *icc* are the spatial parameters, and *iid* is an interchannel intensity difference, *ipd* is an interchannel phase difference, and *icc* is an interchannel coherence.

4. A parametric stereo upmix apparatus as claimed in claim 1 to 3, whereby the means (310) for predicting the difference signal (311) are arranged to enhance the difference signal by adding a scaled decorrelated mono downmix signal.

5. A parametric stereo upmix apparatus as claimed in claim 4, whereby said decorrelated mono downmix (341) is obtained by means of filtering the mono downmix signal (204).

6. A parametric stereo upmix as claimed in claim 4, whereby the scaling factor (322) applied to the decorrelated mono downmix (341) is set to compensate for a prediction energy loss.

7. A parametric stereo upmix apparatus as claimed in claim 6, whereby a scaling factor (322) applied to the decorrelated mono downmix (341) is given as a function of the spatial parameters:

$$\beta = \sqrt{\frac{iid + 1 - 2 \cdot \cos(ipd) \cdot icc \cdot \sqrt{iid}}{iid + 1 + 2 \cdot \cos(ipd) \cdot icc \cdot \sqrt{iid}} - |\alpha|^2}$$

whereby *iid*, *ipd*, and *icc* are the spatial parameters, and *iid* is an interchannel intensity difference, *ipd* is an interchannel phase difference, *icc* is an interchannel coherence, and α is the prediction coefficient (321).

8. A parametric stereo upmix apparatus according to claim 1 to 7, whereby said parametric stereo upmix (300, 400) has a prediction residual signal for the difference signal (331) as an additional input, whereby the arithmetic means (330) are arranged for deriving the left signal (206) and the right signal (207) based on the mono downmix signal (204), said difference signal (311), and said prediction residual signal for the difference signal (331).

9. A parametric stereo decoder comprising a de-multiplexing means (210) for splitting the input bitstream (201) into a mono bitstream (202) and parameter bitstream (203), a mono decoding means (220) for decoding said mono bitstream into a mono downmix signal (204), a parameter decoding means (240) for decoding said parameter bitstream into spatial parameters (205), and a parametric stereo upmix means (230) for generating a left signal (206) and a right signal (207) from a mono downmix signal (204) based on spatial parameters (205), said parametric stereo decoder further comprising the parametric stereo upmix apparatus (300) according to claims 1-7.

10. A parametric stereo decoder comprising a de-multiplexing means (210) for splitting the input bitstream (201) into a mono bitstream (202) and parameter bitstream (203), a mono decoding means (220) for decoding said mono bitstream into a mono downmix signal (204), a parameter decoding means (240) for decoding parameter bitstream into spatial parameters (205), and a parametric stereo upmix means (230) for generating a left signal (206) and a right signal (207) from a mono downmix signal (204) based on spatial parameters

(205), characterized in that the de-multiplexing means (210) are further arranged for extracting a prediction residual bitstream (332) from the input bitstream, the mono decoding means (220) are further arranged to decode a prediction residual signal for the difference signal (331) from the prediction residual bitstream, and the parametric stereo upmix means (230) are being the parametric stereo upmix apparatus according to claim 8.

11. A method for generating a left signal and a right signal from a mono downmix signal based on spatial parameters, characterized by:

- predicting a difference signal comprising a difference between the left signal and the right signal based on the mono downmix signal scaled with a prediction coefficient, whereby said prediction coefficient is derived from the spatial parameters;
- deriving the left signal and the right signal based on a sum and a difference of the mono downmix signal and said difference signal.

12. A method for generating a left signal and a right signal from a mono downmix signal based on spatial parameters as claimed in claim 11, whereby the step of deriving the left signal and the right signal is also based on the prediction residual signal for the difference signal.

13. An audio playing device comprising a parametric stereo decoder according to claim 9 or 10.

14. A parametric stereo downmix apparatus (800) for generating a mono downmix signal (104) from a left signal (101) and a right signal (102) based on spatial parameters (103), characterized in that said parametric stereo downmix apparatus (800) has a prediction residual signal for a difference signal (801) as an additional output, whereby said parametric stereo downmix apparatus comprises a further arithmetic means (810) for deriving the mono downmix signal (104) and a difference signal (811) comprising a difference between the left signal and the right signal, and a further prediction means (820) for deriving a prediction residual signal for the difference signal (801) as a difference between the difference signal (811) and the mono downmix signal (104) scaled with a predetermined prediction coefficient (831) derived from the spatial parameters (103).

15. A parametric stereo encoder comprising an estimation means (130) for deriving spatial parameters (103) from a left signal (101) and a right signal (102), a parametric stereo downmix means (110) for generating a mono downmix signal (104) from the left signal and the right signal based on spatial parameters, a mono encoding means (120) for encoding said mono downmix signal into a mono bitstream (105), a parameter encoding means (140) for encoding spatial parameters into a parameter bitstream (106), and a multiplexing means (150) for merging the mono bitstream and the parameter bitstream into an output bitstream, characterized in that the parametric stereo downmix means (110) are being the parametric stereo downmix apparatus according to claim 14, and the mono encoding means (220) are further arranged to encode the prediction residual signal for the difference signal (801) into a prediction residual bitstream (802), and the multiplexing means (150) are further arranged to merge the prediction bitstream into the output stream.

16. A method for generating a prediction residual signal for a difference signal from a left signal and a right signal based on spatial parameters, characterized by:
- deriving the difference signal between the left signal and the right signal;
- deriving a prediction residual signal for the difference signal as a difference between the difference signal and the mono downmix signal scaled with a prediction coefficient derived from the spatial parameters.

17. A data bitstream comprising merged a mono downmix stream, a parameter stream, and a prediction residual stream.

18. A computer program product for executing the method of any of the claims 11, 12, or 16.

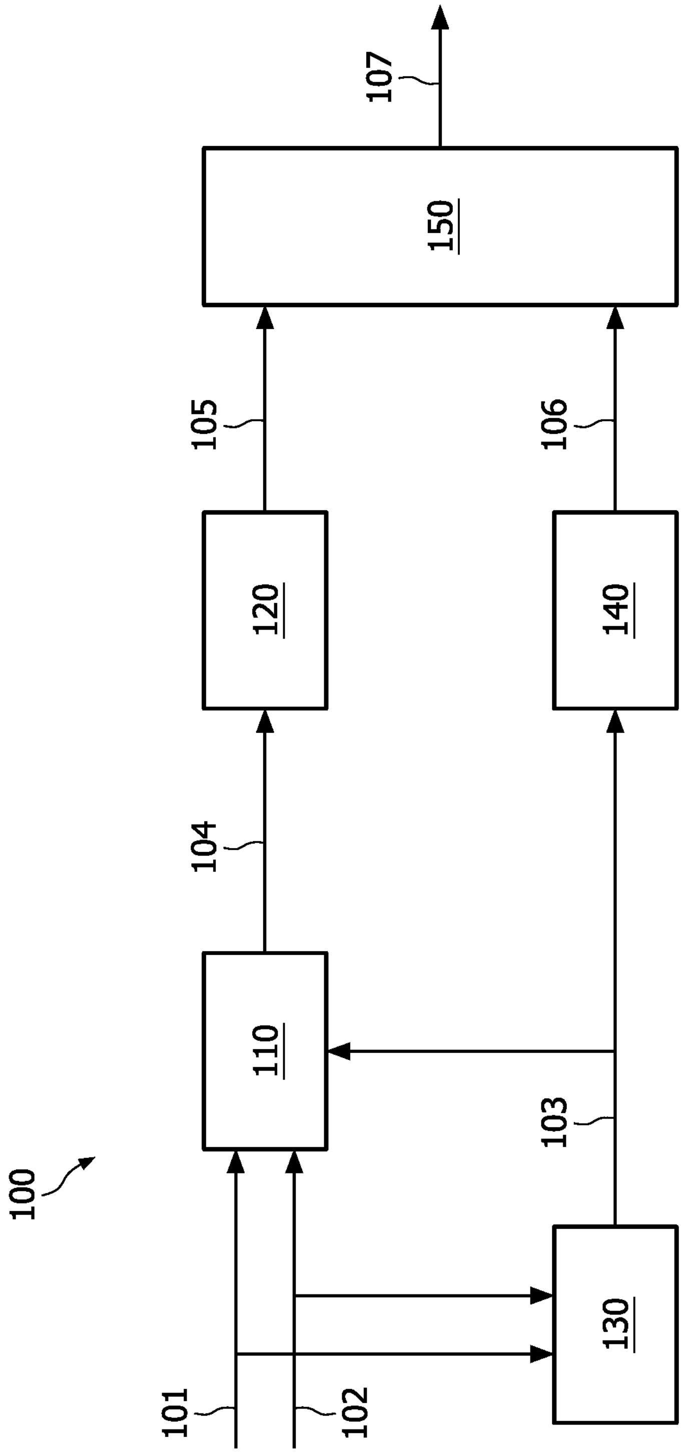

FIG. 1

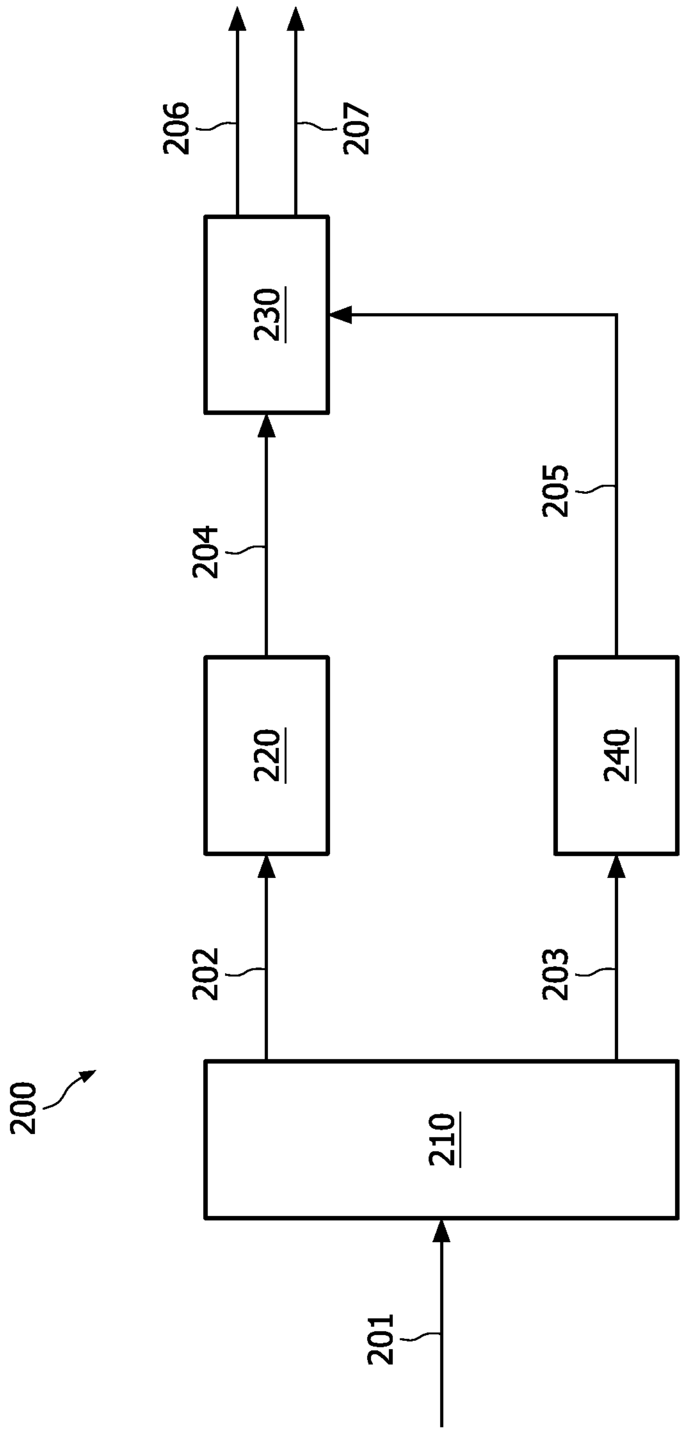

FIG. 2

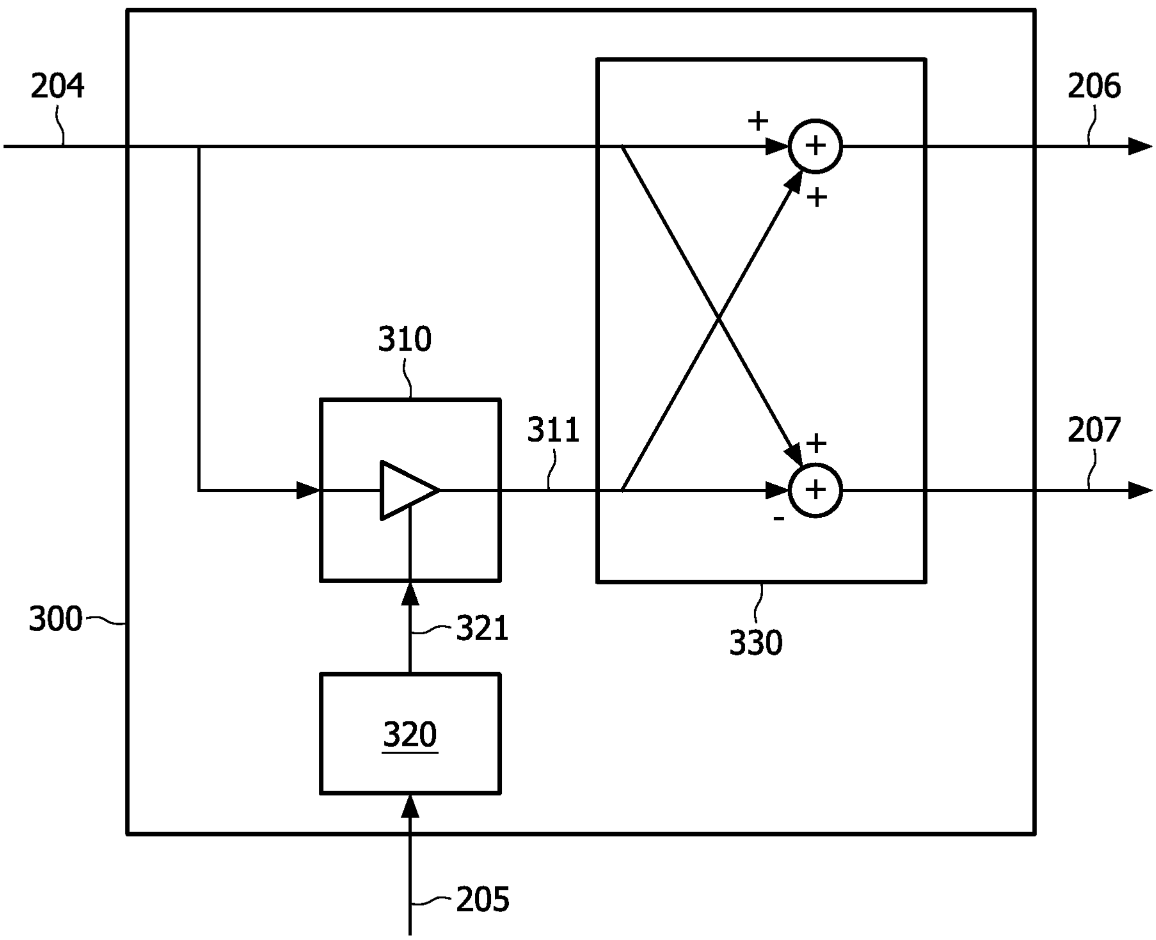

FIG. 3

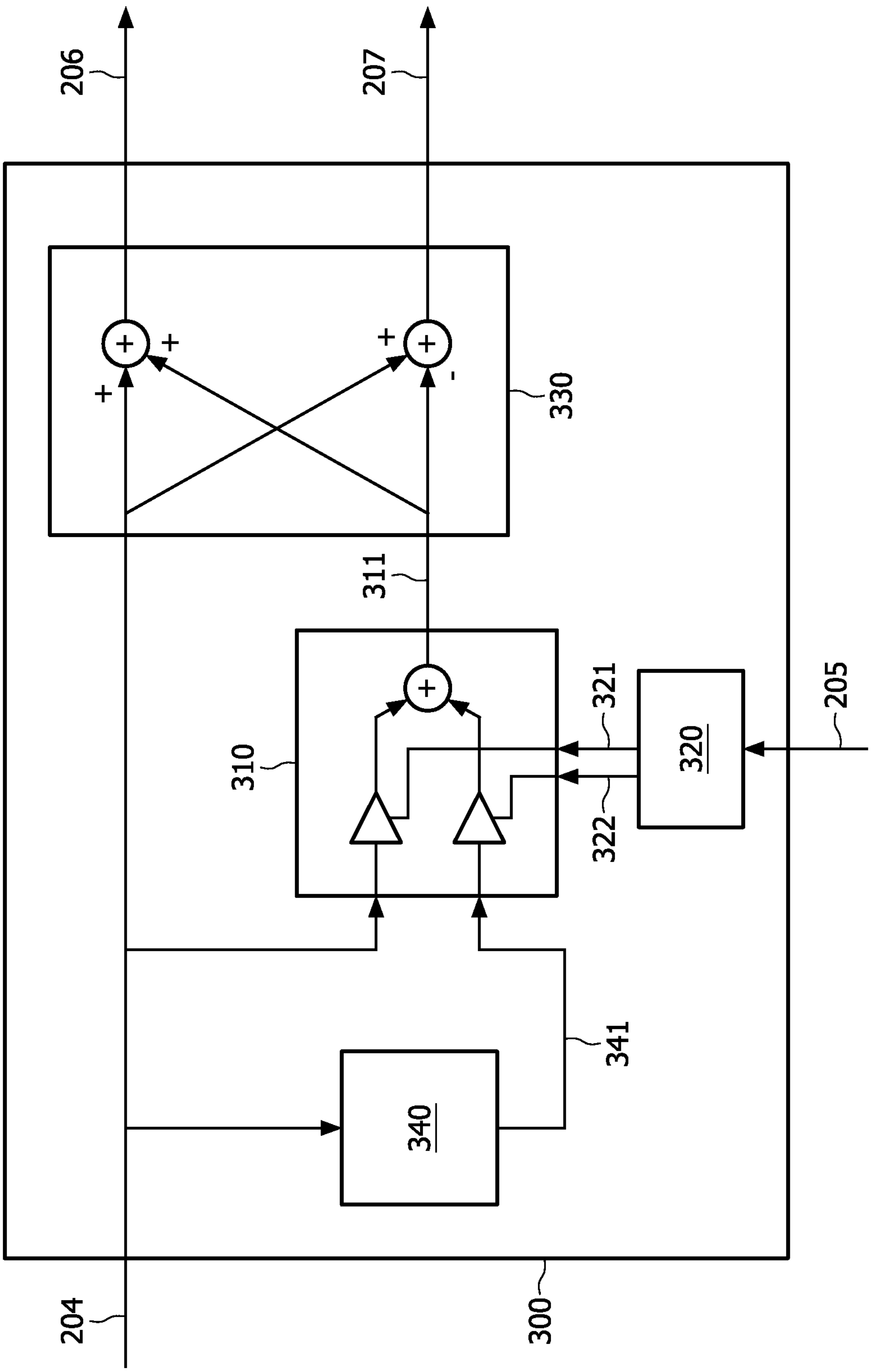

FIG. 4

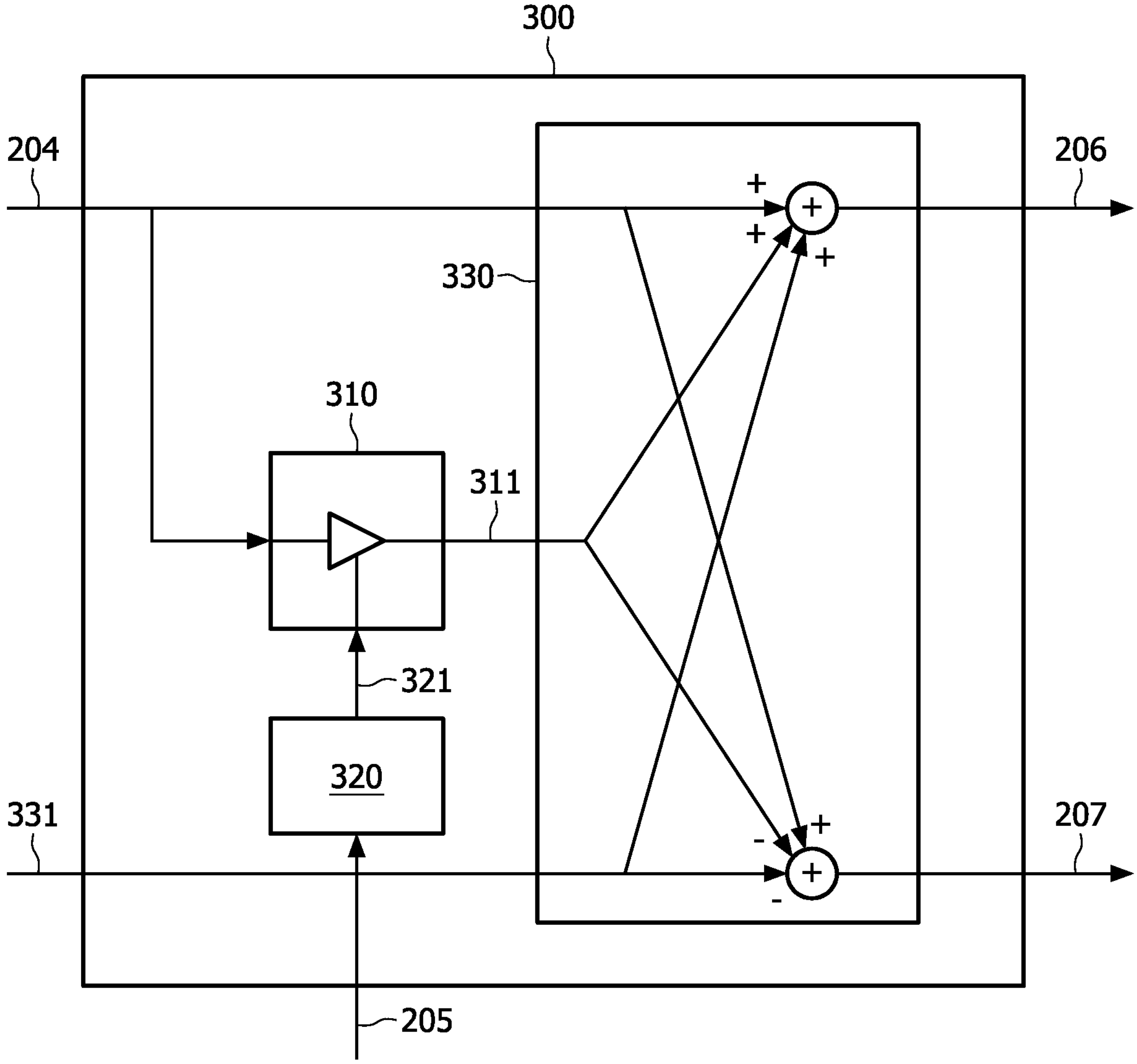

FIG. 5

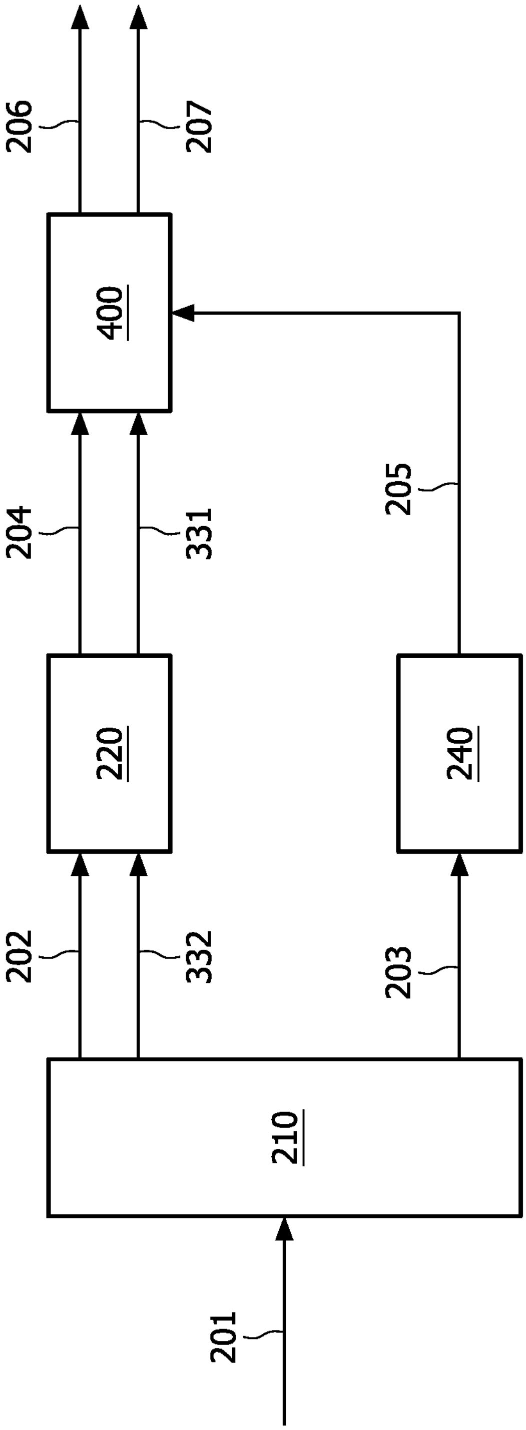

FIG. 6

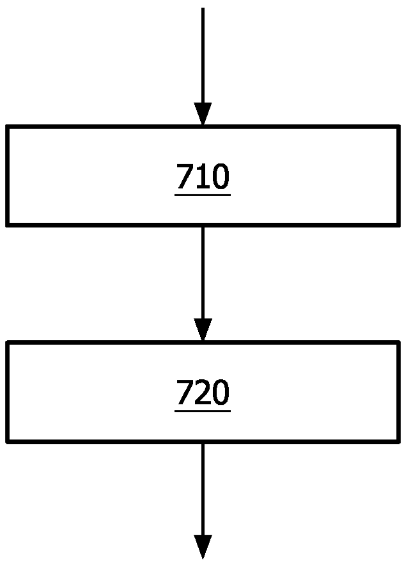

FIG. 7

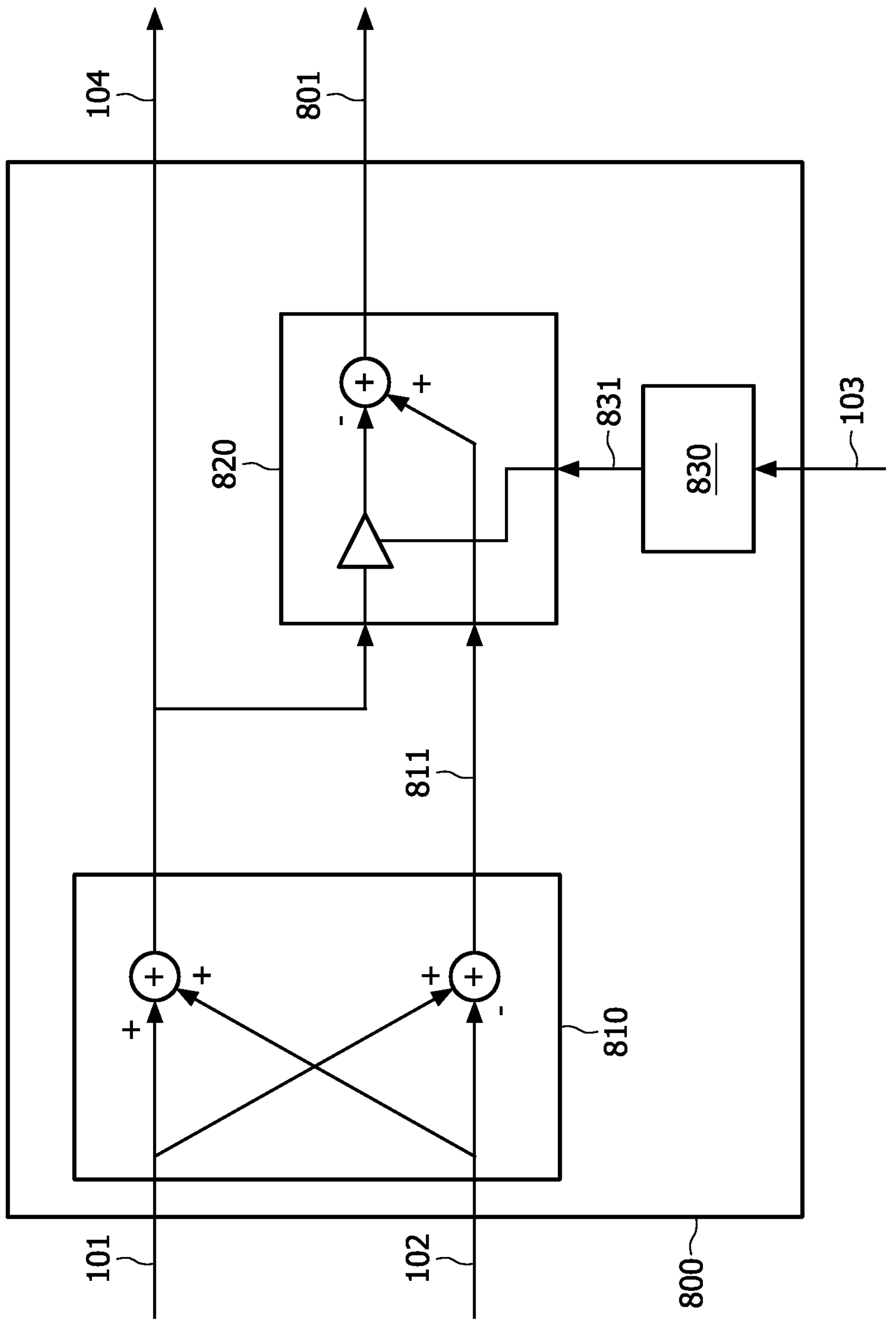

FIG. 8

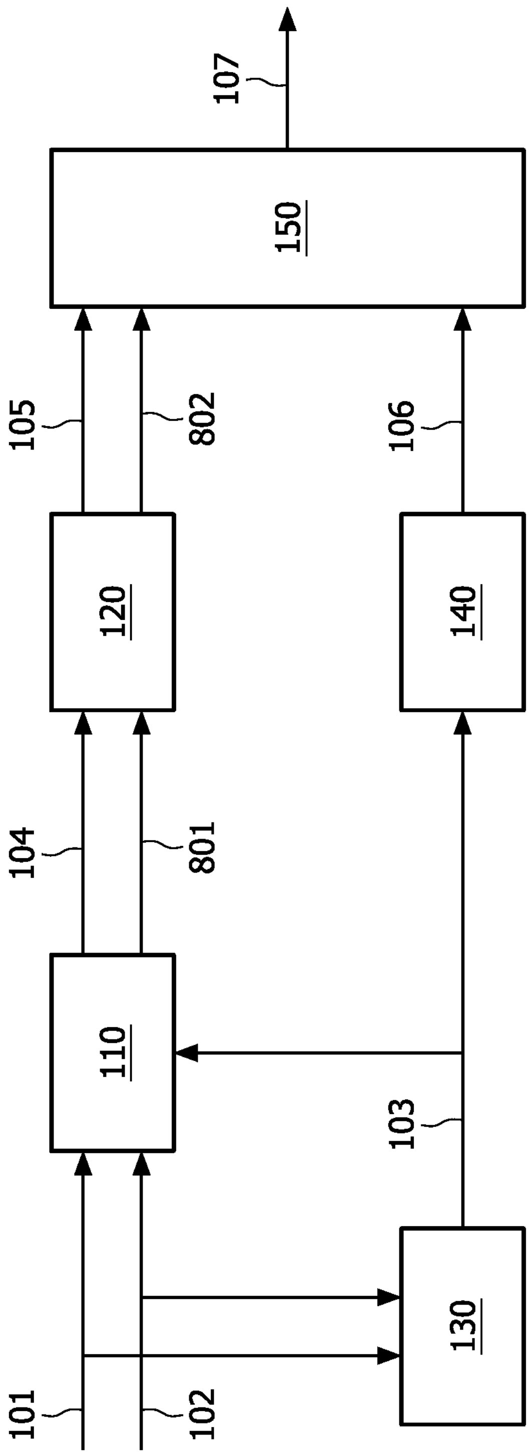

FIG. 9

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2009/052009

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L19/00
ADD. H04S3/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L H04S

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 434 948 A (HOLT CHRISTOPHER E [GB]; MUNDAY EDWARD [GB]; CHEETHAM BARRY M G [GB]) 18 July 1995 (1995-07-18)	17
A	column 2, lines 35-68 column 5, lines 2-18	1-16,18
A	US 5 717 764 A (JOHNSTON JAMES DAVID [US]; SINHA DEEPEN [US]) 10 February 1998 (1998-02-10) column 5, lines 16-26 column 5, line 43 - column 6, line 7	1-18

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 August 2009	02/09/2009

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Fax: (+31–70) 340–3016	Bensa, Julien

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2009/052009

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BREEBAART J; ET AL: "Parametric Coding of Stereo Audio" INTERNET CITATION 1 June 2005 (2005-06-01), pages 1305-1322, XP002514252 ISSN: 1110-8657 Retrieved from the Internet: URL:http://www.jeroenbreebaart.com/papers/jasp/jasp2005.pdf> [retrieved on 2009-02-10] *Section 5.3, 5.4*	1-18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2009/052009

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
US 5434948	A	18-07-1995	NONE			
US 5717764	A	10-02-1998	CA	2117829	A1	24-05-1995
			DE	69432012	D1	20-02-2003
			DE	69432012	T2	06-11-2003
			EP	0655876	A1	31-05-1995
			JP	3970342	B2	05-09-2007
			JP	7199993	A	04-08-1995
			US	5488665	A	30-01-1996