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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 January 2015 (22.01.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/007889 A2

(51) International Patent Classification:
H04S 3/02 (2006.01)

(21) International Application Number:
PCT/EP2014/065517

(22) International Filing Date:
18 July 2014 (18.07.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13306042.6 19 July 2013 (19.07.2013) EP

(71) Applicant: **THOMSON LICENSING** [FR/FR]; 1-5 rue Jeanne d'Arc, F-92130 Issy-les-Moulineaux (FR).

(72) Inventor: **BOEHM, Johannes**; Sieberweg 35, 37081 Göttingen (DE).

(74) Agent: **KÖNIG, Uwe**; Deutsche Thomson OHG, European Patent Operations, Karl-Wiechert-Allee 74, 30625 Hannover (DE).

(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, BN, BR, BW, BY, BZ, CA, CH, CL, 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, IR, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, 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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title**: METHOD FOR RENDERING MULTI-CHANNEL AUDIO SIGNALS FOR L1 CHANNELS TO A DIFFERENT NUMBER L2 OF LOUDSPEAKER CHANNELS AND APPARATUS FOR RENDERING MULTI-CHANNEL AUDIO SIGNALS FOR L1 CHANNELS TO A DIFFERENT NUMBER L2 OF LOUDSPEAKER CHANNELS

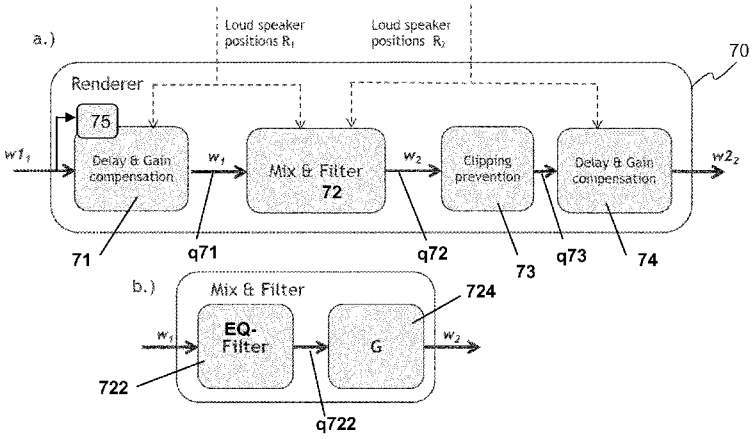


Fig.7

(57) **Abstract**: Multi-channel audio content is mixed for a particular loudspeaker setup. However, a consumer's audio setup is very likely to use a different placement of speakers. The present invention provides a method of rendering multi-channel audio that assures replay of the spatial signal components with equal loudness of the signal. A method for obtaining an energy preserving mixing matrix (G) for mixing L1 input audio channels to L2 output channels comprises steps of obtaining (s711) a first mixing matrix G, performing (s712) a singular value decomposition on the first mixing matrix $\hat{G}$ to obtain a singularity matrix S, processing (s713) the singularity matrix S to obtain a processed singularity matrix $\hat{S}$, determining (s715) a scaling factor a, and calculating (s716) an improved mixing matrix G according to $G = a\, U\, \hat{S}\, V^T$. The perceived sound, loudness, timbre and spatial impression of multi-channel audio replayed on an arbitrary loudspeaker setup practically equals that of the original speaker setup.

WO 2015/007889 A2

METHOD FOR RENDERING MULTI-CHANNEL AUDIO SIGNALS FOR L1 CHANNELS TO A DIFFERENT NUMBER L2 OF LOUDSPEAKER CHANNELS AND APPARATUS FOR RENDERING MULTI-CHANNEL AUDIO SIGNALS FOR L1 CHANNELS TO A DIFFERENT NUMBER L2 OF LOUDSPEAKER CHANNELS

Field of the invention

This invention relates to a method for rendering multi-channel audio signals, and an apparatus for rendering multi-channel audio signals. In particular, the invention relates to a method and apparatus for rendering multi-channel audio signals for L1 channels to a different number L2 of loudspeaker channels.

Background

New 3D channel based Audio formats provide audio mixes for loudspeaker channels that not only surround the listening position, but also include channels positioned above (height) and below in respect to the listening position (sweet spot). The mixes are suited for a special positioning of these speakers. Common formats are 22.2 (i.e. 22 channels) or 11.1 (i.e. 11 channels).

Fig.1 shows two examples of ideal speaker positions in different speaker setups: a 22-channel speaker setup (left) and a 12-channel speaker setup (right). Every node shows the virtual position of a loudspeaker. Real speaker positions that differ in distance to the sweet spot are mapped to the virtual positions by gain and delay compensation.

A renderer for channel based audio receives L_1 digital audio signals $\boldsymbol{w_1}$ and processes the output to L_2 output signals $\boldsymbol{w_2}$. Fig.2 shows, in an embodiment, the integration of a renderer 21 into a reproduction chain. The renderer output signal $\boldsymbol{w_2}$ is converted to an analog signal in a D/A converter 22, amplified in an amplifier 23 and reproduced by loudspeakers 24.

The renderer 21 uses the position information of the input speaker setup and the position information of the output loudspeaker 24 setup as input to initialize the chain of processing. This is shown in Fig.3. Two main processing blocks are a Mixing & Filtering block 31 and a Delay & Gain Compensation block 32.

The speaker position information can be given e.g. in Cartesian or spherical coordinates. The position for the output configuration $\boldsymbol{R}_2$ may be entered manually, or derived via microphone measurements with special test signals, or by any other method. The positions of the input configuration $\boldsymbol{R}_1$ can come with the content by table entry, like an indicator e.g. for 5-channel surround. Ideal standardized loudspeaker positions [9] are

assumed. The positions might also be signaled directly using spherical angle positions. A constant radius is assumed for the input configuration.

Let $\boldsymbol{R}_2 = [\mathbf{r2}_1, \mathbf{r2}_2, \ldots, \mathbf{r2}_{L_2}]$ with $\mathbf{r2}_l = [r2_l, \theta 2_l, \phi 2_l]^T = \left[r2_l, \widehat{\boldsymbol{\Omega}}_l^T\right]^T$ be the positions of the output configuration in spherical coordinates. Origin of the coordinate system is the sweet spot (i.e. listening position). $r2_l$ is the distance between the listening position and a speaker l, and θ_l, ϕ_l are the related spherical angles that indicate the spatial direction of the speaker l relative to the listening position.

Delay and gain compensation

The distances are used to derive delays and gains g_l that are applied to the loudspeaker feeds by amplification/attenuation elements and a delay line with d_l unit sample delay steps. First, the maximal distance between a speaker and the sweet spot is determined: $r2_{max} = \max([r2_1, \ldots r2_{L_2}])$.

For each speaker feed the delay is calculated by:

$$d_l = \lfloor (r2_{max} - r2_l) f_s / c + 0.5 \rfloor \quad (1)$$

with sampling rate f_s, speed of sound c ($c \cong 343 m/s$ at 20°celsius temperature) and $\lfloor x + 0.5 \rfloor$ indicates rounding to next integer. The loudspeaker gains g_l are determined by

$$g_l = \frac{r2_l}{r2_{max}} \quad (2)$$

The task of the Delay and Gain Compensation building block 32 is to attenuate and delay speakers that are closer to the listener than other speakers, so that these closer speakers do not dominate the sound direction perceived. The speakers are thus arranged on a virtual sphere, as shown in Fig.1. The Mix & Filter block 31 now can use virtual speaker positions $\widehat{\boldsymbol{R}}_2 = [\widehat{\mathbf{r2}}_1, \widehat{\mathbf{r2}}_2, \ldots, \widehat{\mathbf{r2}}_{L_2}]$ with $\widehat{\mathbf{r2}}_l = \left[r2_{max}, \widehat{\boldsymbol{\Omega}}_l^T\right]^T$ with a constant speaker distance.

Mix & Filter

In an initialization phase, the speaker positions of the input and idealized output configurations $\boldsymbol{R}_1$, $\widehat{\boldsymbol{R}}_2$ are used to derive a $L_2 \; x \; L_1$ mixing matrix $\boldsymbol{G}$. During the process of rendering, this mixing matrix is applied to the input signals to derive the speaker output signals. As shown in Fig.4, two general approaches exist. In the first approach shown in Fig.4 a), the mixing matrix is independent from the audio frequency and the output is derived by:

$$\boldsymbol{W}_2 = \boldsymbol{G}\, \boldsymbol{W}_1, \quad (3)$$

where $\boldsymbol{W_1} \in \mathcal{R}^{L_1 x\, \tau}$, $\boldsymbol{W_2} \in \mathcal{R}^{L_2 x\, \tau}$ denote the input and output signals of L_1, L_2 audio channels and τ time samples in matrix notation. The most prominent method is Vector Base Amplitude Panning (VBAP) [1].

In the second approach, the mixing matrix becomes frequency dependent ($\boldsymbol{G}(f)$), as shown in Fig.4 b). Then, a filter bank of sufficient resolution is needed, and a mixing matrix is applied to every frequency band sample according to eq.(3).

Examples for the latter approach are known [2],[3],[4]. For deriving the mixing matrix, the following approach is used: A virtual microphone array 51 as depicted in Fig.5, is placed around the sweet spot. The microphone signals M_1 of sound received from the input configuration (the original directions, left-hand side) is compared to the microphone signals M_2 of sound received from the desired speaker configuration (right-hand side). Let $\boldsymbol{\mathcal{M}_1} \in \mathcal{R}^{Mx\, \tau}$ denote M microphone signals receiving the sound radiated from the input configuration, and $\boldsymbol{\mathcal{M}_2} \in \mathcal{R}^{Mx\, \tau}$ be M microphone signals of the sound from the output configuration. They can be derived by

$$\boldsymbol{\mathcal{M}_1} = \boldsymbol{H_{M,L_1}} \boldsymbol{W_1} \tag{4}$$

and

$$\boldsymbol{\mathcal{M}_2} = \boldsymbol{H_{M,L_2}} \boldsymbol{W_2} \tag{5}$$

with $\boldsymbol{H_{M,L_1}} \in \mathbb{C}^{Mx\, L_1}$, $\boldsymbol{H_{M,L_2}} \in \mathbb{C}^{Mx\, L_2}$ being the complex transfer function of the ideal sound radiation in the free field, assuming spherical wave or plane wave radiation. The transfer functions are frequency dependent. Selecting a mid-frequency f_m related to a filter bank, eq.(4) and eq. (5) can be equated using eq.(3). For every f_m the following equation needs to be solved to derive $\boldsymbol{G}(f_m)$:

$$\boldsymbol{H_{M,L_1}} \boldsymbol{W_1} = \boldsymbol{H_{M,L_2}} \boldsymbol{G}\, \boldsymbol{W_1} \tag{6}$$

A solution that is independent of the input signals and that uses the pseudo inverse matrix of $\boldsymbol{H_{M,L_2}}$ can be derived as:

$$\boldsymbol{G} = \boldsymbol{H^{+}_{M,L_2}} \boldsymbol{H_{M,L_1}}\ . \tag{7}$$

Usually this produces non-satisfying results, and [2] and [5] present more sophisticated approached to solve eq.(6) for $\boldsymbol{G}$.

Further, there is a completely different way of signal adaptive rendering, where the directional signals of the incoming audio content is extracted and rendered like audio objects. The residual signal is panned and de-correlated to the output speakers. This kind of audio rendering is much more expensive in terms of computational complexity, and often not free from artifacts. Signal adaptive rendering is not used and only mentioned here for completeness.

One problem is that a consumer's home setup is very likely to use a different placement of speakers due to real world constraints of a living room. Also the number of speakers may be different. The task of a renderer is thus to adapt the channel based audio signals to a new setup such that the perceived sound, loudness, timbre and spatial impression comes as close as possible to the original channel based audio as replayed on its original speaker setup, like e.g. in the mixing room.

Summary of the Invention

The present invention provides a preferably computer-implemented method of rendering multi-channel audio signals that assures replay (i.e. reproduction) of the spatial signal components with correct loudness of the signal (ie. equal to the original setup). Thus, a directional signal that is perceived in the original mix coming from a direction is also perceived equally loud when rendered to the new loudspeaker setup. In addition, filters are provided that equalize the input signals to reproduce a timbre as close as possible as it would be perceived when listening to the original setup.

In one aspect, the invention relates to a method for rendering L1 channel-based input audio signals to L2 loudspeaker channels, where L1 is different from L2, as disclosed in claim 1. In one embodiment, a step of mixing the delay and gain compensated input audio signal for L2 audio channels uses a mixing matrix that is generated as disclosed in claim 5. A corresponding apparatus according to the invention is disclosed in claim 8 and claim 12, respectively.

In one aspect, the invention relates to a method for generating an energy preserving mixing matrix ***G*** for mixing input channel-based audio signals for L1 audio channels to L2 loudspeaker channels, as disclosed in claim 7. A corresponding apparatus for generating an energy preserving mixing matrix ***G*** according to the invention is disclosed in claim 14. In one aspect, the invention relates to a computer readable medium having stored thereon executable instructions to cause a computer to perform a method according to claim 1, or a method according to claim 7.

In one embodiment of the invention, a computer-implemented method for generating an energy preserving mixing matrix ***G*** for mixing input channel-based audio signals for L1 audio channels to L2 loudspeaker channels comprises computer-executed steps of obtaining a first mixing matrix $\hat{G}$ from virtual source directions $\widehat{R_1}$ and target speaker directions $\widehat{R_2}$, performing a singular value decomposition on the first mixing matrix $\hat{G}$ to

obtain a singularity matrix $\boldsymbol{S}$, processing the singularity matrix $\boldsymbol{S}$ to obtain a processed singularity matrix $\widehat{\boldsymbol{S}}$ with s_m non-zero diagonal elements, determining from the number of non-zero diagonal elements a scaling factor a according to $a = \sqrt{\frac{L_1}{s_m}}$ (for $L2 \leq L1$) or $a = \sqrt{\frac{L_2}{s_m}}$ (for $L2 > L1$), and calculating a mixing matrix $\boldsymbol{G}$ by using the scaling factor according to $\boldsymbol{G} = a\,\boldsymbol{U}\,\widehat{\boldsymbol{S}}\,\boldsymbol{V}^{\boldsymbol{T}}$. As a result, the perceived sound, loudness, timbre and spatial impression of multi-channel audio replayed on an arbitrary loudspeaker setup is improved, and in particular comes as close as possible to the original channel based audio as if replayed on its original speaker setup.

Further objects, features and advantages of the invention will become apparent from a consideration of the following description and the appended claims when taken in connection with the accompanying drawings.

Brief description of the drawings

Exemplary embodiments of the invention are described with reference to the accompanying drawings, which show in

Fig.1 two examples of loudspeaker setups;

Fig.2 a known general structure for rendering content for a new loudspeaker setup;

Fig.3 a general known structure for channel based audio rendering;

Fig.4 two approaches to mix L_1 channels to L_2 output channels, using a) a frequency-independent mixing matrix $\boldsymbol{G}$, and b) a frequency dependent mixing matrix $\boldsymbol{G}(\boldsymbol{f})$;

Fig.5 a virtual microphone array used to compare the sound radiated from the original setup (input configuration) to a desired output configuration;

Fig.6 a) a flow-chart of a method for rendering L1 channel-based input audio signals to L2 loudspeaker channels according to the invention;

Fig.6 b) a flow-chart of a method for generating an energy preserving mixing matrix $\boldsymbol{G}$ according to the invention;

Fig.7 a rendering architecture according to one embodiment of the invention;

Fig.8 the structure of one embodiment of a filter in the Mix&Filter block;

Fig.9 exemplary frequency responses for a remix of five channels; and

Fig.10 exemplary frequency responses for a remix of twenty-two channels.

Detailed description of the invention

Fig.6 a) shows a flow-chart of a method for rendering a first number L1 of channel-based input audio signals to a different second number L2 of loudspeaker channels according to

one embodiment of the invention. The method for rendering L1 channel-based input audio signals $w1_1$ to L2 loudspeaker channels, where the number L1 of channel-based input audio signals is different from the number L2 of loudspeaker channels, comprises steps of determining s60 a mix type of the L1 input audio signals, performing a first delay and gain compensation s61 on the L1 input audio signals according to the determined mix type, wherein a delay and gain compensated input audio signal with the first number L1 of channels and with a defined mix type is obtained, mixing s624 the delay and gain compensated input audio signal for the second number L2 of audio channels, wherein a remixed audio signal for the second number L2 of audio channels is obtained, clipping s63 the remixed audio signal, wherein a clipped remixed audio signal for the second number L2 of audio channels is obtained, and performing a second delay and gain compensation s64 on the clipped remixed audio signal for the second number L2 of audio channels, wherein the second number L2 of loudspeaker channels $w2_2$ are obtained. Possible mix types include at least one of spherical, cylindrical and rectangular (or, more general, cubic). In one embodiment, the method comprises a further step of filtering s622 the delay and gain compensated input audio signal q71 having the first number L1 of channels in an equalization filter (or equalizer filter), wherein a filtered delay and gain compensated input audio signal is obtained. While the equalization filtering is in principle independent from the usage of, and can be used without, an energy preserving mixing matrix, it is particularly advantageous to use both in combination.

Fig.6 b) shows a flow-chart of a method for generating an energy preserving mixing matrix $\boldsymbol{G}$ according to one embodiment of the invention. The method s710 for obtaining an energy preserving mixing matrix $\boldsymbol{G}$ for mixing input channel-based audio signals for a first number L1 of audio channels to a second number L2 of loudspeaker channels comprises steps of obtaining s711 a first mixing matrix $\hat{G}$ from virtual source positions/ directions $\widehat{R_1}$ and target speaker positions/directions $\widehat{R_2}$ wherein a panning method is used, performing s712 a singular value decomposition on the first mixing matrix $\hat{G}$ according to $\widehat{\boldsymbol{G}} = \boldsymbol{U}\,\boldsymbol{S}\,\boldsymbol{V}^{\boldsymbol{T}}$, wherein $\boldsymbol{U} \in \mathcal{R}^{L_2 x\, L_2}$ and $\boldsymbol{V} \in \mathcal{R}^{L_1 x\, L_1}$ are orthogonal matrices and $\boldsymbol{S} \in \mathcal{R}^{L_1 x\, L_2}$ is a singularity matrix and has s first diagonal elements being the singular values of $\boldsymbol{G}$ in descending order and all other elements of $\boldsymbol{S}$ are zero, processing s713 the singularity matrix $\boldsymbol{S}$, wherein a quantized singularity matrix $\hat{S}$ is obtained with diagonal elements that are above a threshold set to one and diagonal elements that are below a threshold set to zero, determining s714 a number s_m of diagonal elements that are set to one in the quantized singularity matrix $\hat{S}$, determining

s715 a scaling factor a according to $a = \sqrt{\frac{L_1}{s_m}}\ for\ (L2 \leq L1)$ or $a = \sqrt{\frac{L_2}{s_m}}$ for $(L2 > L1)$, and calculating s716 a mixing matrix $\boldsymbol{G}$ according to $\boldsymbol{G} = a\ \boldsymbol{U}\ \widehat{\boldsymbol{S}}\ \boldsymbol{V^T}$. The steps of any of the above-mentioned methods can be performed by one or more processing elements, such as microprocessors, threads of a GPU etc.

Fig.7 shows a rendering architecture 70 according to one embodiment of the invention. In the rendering architecture according to the embodiment shown in Fig.7a), an additional "Gain and Delay Compensation" block 71 is used for preprocessing different input setups, such as spherical, cylindrical or rectangular input setups. Further, a modified "Mix & Filter" block 72 that is capable of preserving the original loudness is used. In one embodiment, the "Mix & Filter" block 72 comprises an equalization filter 722. The "Mix & Filter" block 72 is described in more detail with respect to Fig.7b) and Fig.8. A clipping prevention block 73 prevents signal overflow, which may occur due to the modified mixing matrix. A determining unit 75 determines a mix type of the input audio signals.

Fig.7b) shows the Mix&Filter block 72 incorporating an equalization filter 722 and a mixer unit 724. Fig.8 shows the structure of the equalization filter 722 in the Mix&Filter block. The equalization filter is in principle a filter bank with L_1 filters $EF_1,\ldots,EF_{L1}$, one for each input channel. The design and characteristics of the filters are described below. All blocks mentioned may be implemented by one or more processors or processing elements that may be controlled by software instructions.

The renderer according to the invention solves at least one of the following problems: First, new 3D audio channel based content can be mixed for at least one of spherical, rectangular or cylindrical speaker setups. The setup information needs to be transmitted alongside e.g. with an index for a table entry signaling the input configuration (which assumes a constant speaker radius) to be able to calculate the real input speaker positions. In an alternative embodiment, full input speaker position coordinates can be transmitted along with the content as metadata. To use mixing matrices independent of the mixing type, a gain and delay compensation is provided for the input configuration. Second, the invention provides an energy preserving mixing matrix $\boldsymbol{G}$. Conventionally, the mixing matrix is not energy preserving. Energy preservation assures that the content has the same loudness after rendering, compared to the content loudness in the mixing room when using the same calibration of a replay system [6],[7],[8]. This also assures that e.g. 22-channel input or 10-channel input with equal 'Loudness, K-weighted, relative to Full Scale' (LKFS) content loudness appears equally loud after rendering.

One advantage of the invention is that it allows generating energy (and loudness) preserving, frequency independent mixing matrices. It is noted that the same principle

can also be used for frequency dependent mixing matrices, which however are not so desirable. A frequency independent mixing matrix is beneficial in terms of computational complexity, but often a drawback can be a in change in timbre after remix. In one embodiment, simple filters are applied to each input loudspeaker channel before mixing, in order to avoid this timbre mismatching after mixing. This is the equalization filter 722. A method for designing such filters is disclosed below.

Energy preserving rendering has a drawback that signal overload is possible for peak audio signal components. In one embodiment of the present invention, an additional clipping prevention block 73 prevents such overload. In a simple realization, this can be a saturation, while in more sophisticated realizations this block is a dynamics processor for peak audio.

In the following, details about the mix type determining unit 75 and the Input Gain and Delay compensation 71 are described. If the input configuration is signaled by a table entry plus mix room information, like e.g. rectangular, cylindrical or spherical, the configuration coordinates are read from special prepared tables (e.g. RAM) as spherical coordinates. If the coordinates are transmitted directly, they are converted to spherical coordinates. A determining unit 75 determines a mix type of the input audio signals. Let $\boldsymbol{R}_1 = [\mathbf{r1}_1, \mathbf{r1}_2, \ldots, \mathbf{r1}_{L_1}]$ with $\mathbf{r1}_l = [r1_l, \theta 1_l, \phi 1_l]^T = \left[r1_l, \boldsymbol{\Omega}_l^T\right]^T$ being the positions of this input configuration.

In a first step the maximum radius is detected: $\mathrm{r1}_{\max} = \max(\left[\mathrm{r1}_1, \ldots \mathrm{r1}_{\mathrm{L}_2}\right]$. Because only relative differences are of interest for this building block, the radii are $\mathrm{r1}_\mathrm{l}$ scaled by $r2_{max}$ that is available from the gain and delay compensation initialization of the output configuration:

$$\widehat{r1}_l = r1_l \frac{r2_{max}}{r1_{max}} \tag{8}$$

The number of delay tabs $\check{d}_l$ and the gain values $\check{g}_l$ for every speaker are calculated as follows with $\widehat{r1}_{max} = r2_{max}$:

$$\check{d}_l = \left\lfloor \left(r2_{max} - \widehat{r1}_l\right) f_s / c + 0.5 \right\rfloor \tag{9}$$

with sampling rate f_s, speed of sound c ($c \cong 343 m/s$ at 20°celsius temperature) and $\lfloor x + 0.5 \rfloor$ indicates rounding to next integer.

The loudspeaker gains $\check{g}_l$ are determined by

$$\check{g}_l = \frac{\widehat{r1}_l}{\widehat{r1}_{max}} \tag{10}$$

The Mix & Filter block now can use virtual speaker positions $\widehat{\boldsymbol{R}}_1 = [\widehat{\mathbf{r1}}_1, \widehat{\mathbf{r1}}_2, \ldots, \widehat{\mathbf{r1}}_{L_1}]$ with $\widehat{\mathbf{r1}}_l = \left[\widehat{r1}_{max}, \boldsymbol{\Omega}_l^T\right]^T$ with a constant speaker distance.

In the following, the Mixing Matrix design is explained.

First, the energy of the speaker signals and perceived loudness are discussed.

Fig.7a) shows a block diagram defining the descriptive variables. L_1 loudspeakers signals have to be processed to L_2 signals (usually, $L_2 \leq L_1$). Replay of the loudspeaker feed signals W_2 (shown as W_{2_2} in Fig.7) should ideally be perceived with the same loudness as if listening to a replay in the mixing room, with the optimal speaker setup. Let $\boldsymbol{W_1}$ be a matrix of L_1 loudspeaker channels (rows) and τ samples (columns).

The energy of the signal $\boldsymbol{W_1}$, of the τ-time sample block is defined as follows:

$$\mathrm{E}_{w_1} = \left|\left|\boldsymbol{W_1}\right|\right|_{fro}^2 = \sum_{i=1}^{\tau}\sum_{l=1}^{L_1} W_{1\,l,i}^{\,2} = \sum_{i=1}^{\tau} \boldsymbol{w}_{1_t}^{\,T}\, \boldsymbol{w}_{1_t} \qquad (11)$$

Here $W_{l,i}$ are the matrix elements of $\boldsymbol{W_1}$, l denotes the speaker index, i denotes the sample index, $\left|\left|\ \right|\right|_{fro}$ denotes the Frobenius matrix norm, $\boldsymbol{w}_{1_t}$ is the t^{th} column vector of $\boldsymbol{W_1}$ and $[\,]^T$ denotes vector or matrix transposition.

This energy E_w gives a fair estimate of the loudness measure of a channel based audio as defined in [6],[7],[8], where the K-filter suppresses frequencies lower than 200Hz.

Mixing of the signals $\boldsymbol{W_1}$ provides signals $\boldsymbol{W_2}$. The signal energy after mixing becomes:

$$\mathrm{E}_{w_2} = \left|\left|\boldsymbol{W_2}\right|\right|_{fro}^2 = \sum_{i=1}^{\tau}\sum_{l=1}^{L_2} W_{2\,l,i}^{\,2} \qquad (12)$$

where L_2 is the new number of loudspeakers, with $L_2 \leq L_1$.

The process of rendering is assumed to be performed by a mixing matrix $\boldsymbol{G}$, signals $\boldsymbol{W_2}$ are derived from $\boldsymbol{W_1}$ as follows:

$$\boldsymbol{W_2} = \boldsymbol{G}\,\boldsymbol{W_1} \qquad (13)$$

Evaluating E_{w_2} and using the columns vector decomposition of $\boldsymbol{W_1} = \left[\boldsymbol{w}_{1_1}, .., \boldsymbol{w}_{1_t}, .., \boldsymbol{w}_{1_\tau}\right]$ with $\boldsymbol{w}_{1_t} = [w_{1_{t,1}}, .., w_{1_{t,l}}, .., w_{1_{t,L}}]^T$ then leads to:

$$\mathrm{E}_{w_2} = \sum_{i=1}^{\tau}\sum_{l=1}^{L} W_{2\,l,i}^{\,2} = \sum_{i=1}^{\tau}[\boldsymbol{G}\boldsymbol{w}_{1_t}]^T \boldsymbol{M}\boldsymbol{w}_{1_t} = \sum_{i=1}^{\tau} \boldsymbol{w}_{1_t}^{\,T}\boldsymbol{G}^T\,\boldsymbol{G}\,\boldsymbol{w}_{1_t} \qquad (14)$$

In one embodiment, loudness preservation is then obtained as follows.

The loudness of the original signal mix is preserved in the new rendered signal if:

$$E_1 = E_2 \qquad (15)$$

From eq.(14) it becomes apparent that mixing matrix $\boldsymbol{M}$ needs to be orthogonal and

$$\boldsymbol{G}^T\boldsymbol{G} = \boldsymbol{I} \qquad (16)$$

with $\boldsymbol{I}$ being the $L_1\ x\ L_1$ unit matrix.

An optimal rendering matrix (also called mixing matrix or decode matrix) can be obtained as follows, according to one embodiment of the invention.

Step 1: A conventional mixing matrix $\widehat{\boldsymbol{G}}$ is derived by using panning methods. A single loudspeaker l_1 from the set of original loudspeakers is viewed as a sound source to be reproduced by L_2 speakers of the new speaker setup. Preferred panning methods are VBAP [1] or robust panning [2] for a constant frequency (i.e. a known technology can be used for this step). To determine the mixing matrix $\widehat{\boldsymbol{G}}$, the modified speaker positions $\widehat{\boldsymbol{R}}_2$, $\widehat{\boldsymbol{R}}_1$are used, $\widehat{\boldsymbol{R}}_2$for the output configuration and $\widehat{\boldsymbol{R}}_1$ for the virtual source directions.

Step 2: Using compact singular value decomposition, the mixing matrix is expressed as a product of three matrices:

$$\widehat{\boldsymbol{G}} = \boldsymbol{U}\,\boldsymbol{S}\,\boldsymbol{V}^{\boldsymbol{T}} \tag{17}$$

$\boldsymbol{U} \in \mathcal{R}^{L_2 x\, L_2}$ and $\boldsymbol{V} \in \mathcal{R}^{L_1 x\, L_1}$ are orthogonal matrices and $\boldsymbol{S} \in \mathcal{R}^{L_1 x\, L_2}$ has s first diagonal elements (the singular values in descending order), with $s \leq L_2$. The other matrix elements are zeros.

Note that this holds for the case of $L_2 \leq L_1$, (remix $L_2 = L_1$, downmix $L_2 < L_1$). For the case of upmix ($L_2 > L_1$), L_2 needs to be replaced by L_1 in this section.

Step3: A new matrix $\widehat{\boldsymbol{S}}$ is formed from $\boldsymbol{S}$ where the diagonal elements are replaced by a value of one, but very low valued singular values $s_s << s_{max}$ are replaced by zeros. A threshold in the range of -10dB ... -30dB or less is usually selected (e.g. -20dB is a typical value). The threshold becomes apparent from actual numbers in realistic examples, since there will occur two groups of diagonal elements: elements with larger value and elements with considerably smaller value. The threshold is for distinguishing among these two groups.

For most speaker settings, the number of non-zero diagonal elements s_m is $s_m = L_2$, but for some settings it becomes lower and then $s_m < L_2$. This means that $L_2 - s_m$ speakers will not be used to replay content; there is simply no audio information for them, and they remain silent.

Let s_m denote the last singular value to be replaced by one. Then the mixing matrix $\boldsymbol{G}$ is determined by:

$$\boldsymbol{G} = a\,\boldsymbol{U}\,\widehat{\boldsymbol{S}}\,\boldsymbol{V}^{\boldsymbol{T}} \tag{18}$$

with the scaling factor

$$a = \sqrt{\frac{L_1}{s_m}}\ for\ (L2 \leq L1) \tag{19}$$

or, respectively,

$$a = \sqrt{\frac{L_2}{s_m}} \text{ for } (L2 > L1) \qquad (19')$$

The scaling factor is derived from: $\boldsymbol{G}^T\boldsymbol{G} = a^2\boldsymbol{V}\hat{\boldsymbol{S}}^2\boldsymbol{V}^T = a^2\boldsymbol{V}\boldsymbol{V}^T$, where $\boldsymbol{V}\boldsymbol{V}^T$ has s_m Eigenvalues equal to one. That means that $|\boldsymbol{V}\boldsymbol{V}^T|_{fro} = \sqrt{s_m}$. Thus, simply down mixing the L_1 signals to s_m signals will reduce the energy, unless $s_m = L_1$ (in other words: when the number of output speakers matches the number of input speakers). With $|\boldsymbol{I}_{L_1}|_{fro} = \sqrt{L_1}$, a scaling factor $a = \sqrt{\frac{L_1}{s_m}}$ compensates the loss of energy during down-mixing.

As an example, processing of a singularity matrix is described in the following. E.g., an initial (conventional) mixing matrix for L loudspeakers is decomposed using compact singular value decomposition according to eq.(17): $\hat{\boldsymbol{G}} = \boldsymbol{U}\,\boldsymbol{S}\,\boldsymbol{V}^T$. The singularity matrix $\boldsymbol{S}$ is square (with LxL elements, L=min{L_1,L_2} for compact singular value decomposition) and is a diagonal matrix of the form

$\boldsymbol{S} = \begin{bmatrix} s_1 & \dots & 0 \\ 0 & s_2 & \vdots \\ \vdots & \ddots & 0 \\ 0 & \dots & s_L \end{bmatrix}$ with $s_1 \geq s_2 \geq \dots \geq s_L$ (i.e., $s_1=s_{max}$). Then the singularity matrix is processed by setting the coefficients $s_1,s_2,\dots,s_L$ to be either 1 or 0, depending whether each coefficient is above a threshold of e.g. $0.06*s_{max}$. This is similar to a relative quantization of the coefficients. The threshold factor is exemplary 0.06, but can be (when expressed in decibel) e.g. in the range of -10dB or lower.

For a case with e.g. L=5 and e.g. only s_1 and s_2 being above the threshold and s_3, s_4 and s_5 being below the threshold, the resulting processed (or "quantized") singularity matrix $\hat{S}$ is

$\hat{S} = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. Thus, the number of its non-zero diagonal coefficients s_m is two.

In the following, the Equalization Filter 722 is described.

When mixing between different 3D setups, and especially when mixing from 3D setups to 2D setups, timbre may change. E.g. for 3D to 2D, a sound originally coming from above is now reproduced using only speakers on the horizontal plane. The task of the equalization filter is to minimize this timbre mismatch and maximize energy preservation. Individual filters F_l are applied to each channel of the L_1 channels of the input configuration before

applying the mixing matrix, as shown in Fig.7 b). The following shows the theoretical deviation and describes how the frequency response of the filters is derived.

A model according to Fig.7 and eqs. (4) and (5) is used. Both equations are reprinted here for convenience:

$$\boldsymbol{\mathcal{M}_1} = \boldsymbol{H_{M,L_1}}\boldsymbol{W_1} \tag{20}$$

and

$$\boldsymbol{\mathcal{M}_2} = \boldsymbol{H_{M,L_2}}\boldsymbol{W_2} \tag{21}$$

with $\boldsymbol{H_{M,L_1}} \in \mathbb{C}^{Mx\,L_1}$, $\boldsymbol{H_{M,L_2}} \in \mathbb{C}^{Mx\,L_2}$ being the complex transfer function of the ideal sound radiation in the free field assuming spherical wave or plane wave radiation. These matrices are functions of frequency, and they can be calculated using the position information $\widehat{\boldsymbol{R}}_2$, $\widehat{\boldsymbol{R}}_1$. We define $\boldsymbol{W_2} = \widetilde{\boldsymbol{G}}\boldsymbol{W_1}$, where $\widetilde{\boldsymbol{G}}$ is a function of frequency.

Instead of equating eqs.(4) and (5), as mentioned in the background section, we will equate the energies. And since we want to equalize for the sound of the speaker directions of the input configuration, we can solve the considerations for each input speaker at a time (loop over L_1).

The energy measured at the virtual microphones for the input setup, if only one speaker l is active, is given by

$$\left|\boldsymbol{\mathcal{M}_{1,l}}\right|_{fro}^2 = \left|\boldsymbol{h_{M,l}}\,\boldsymbol{w_{1\,l}}\right|_{fro}^2 \tag{22}$$

with $\boldsymbol{h_{M,l}}$ representing the lth column of $\boldsymbol{H_{M,L_1}}$ and $\boldsymbol{w_{1\,l}}$ one row of $\boldsymbol{W_1}$, i.e. the time signal of speaker l with τ samples. Rewriting the Frobenius norm analog to eq.(11), we can further evaluate eq.(22) to:

$$\left|\boldsymbol{\mathcal{M}_{1,l}}\right|_{fro}^2 = \sum_{i=1}^{\tau} \boldsymbol{w_{1\,l}^T}\,\boldsymbol{w_{1\,l}}\;\boldsymbol{h_{M,l}^H}\,\boldsymbol{h_{M,l}} = E_{wl}\boldsymbol{h_{M,l}^H}\,\boldsymbol{h_{M,l}} \tag{23}$$

where $(\,)^H$ is conjugate complex transposed (Hermitian transposed) and E_{wl} is the energy of speaker signal l. The vector $\boldsymbol{h_{M,l}}$ is composed out of complex exponentials (see eqs.(31), (32)) and the multiplication of an element with its conjugate complex equals one, thus $\boldsymbol{h_{M,l}^H}\,\boldsymbol{h_{M,l}} = L_1$:

$$\left|\boldsymbol{\mathcal{M}_{1,l}}\right|_{fro}^2 = E_{wl}L_1 \tag{24}$$

The measures at the virtual microphones after mixing are given by $\boldsymbol{\mathcal{M}_2} = \boldsymbol{H_{M,L_2}}\widetilde{\boldsymbol{G}}\boldsymbol{W_1}$. If only one speaker is active, we can rewrite to:

$$\boldsymbol{\mathcal{M}_{2,l}} = \boldsymbol{H_{M,L_2}}\,\widetilde{\boldsymbol{g}}_{\boldsymbol{l}}\,\boldsymbol{w_{1\,l}} \tag{25}$$

with $\widetilde{\boldsymbol{g}}_{\boldsymbol{l}}$ being the lth column of $\widetilde{\boldsymbol{G}}$. We define $\widetilde{\boldsymbol{G}}$ to be decomposable into a frequency dependent part related to speaker l and mixing matrix $\boldsymbol{G}$ derived from eq.(24):

$$\widetilde{\boldsymbol{G}}(f) = diag\big(\boldsymbol{b}(f)\big)\,\boldsymbol{G} \tag{26}$$

with $\boldsymbol{b}$ as a frequency dependent vector of L_1 complex elements and (f) denoting frequency dependency, which is neglected in the following for simplicity. With this, eq.(25) becomes:

$$\boldsymbol{\mathcal{M}}_{2,l} = \boldsymbol{H}_{M,L_2}\, b_l\, \boldsymbol{g}\, \boldsymbol{w}_{1\,l} \tag{27}$$

where $\boldsymbol{g}$ is the l^{th} column of $\boldsymbol{G}$ and b_l the l^{th} element of $\boldsymbol{b}$. Using the same considerations of the Frobenius norm as above, the energy at the virtual microphones becomes:

$$\left|\boldsymbol{\mathcal{M}}_{2,l}\right|_{fro}^2 = E_{wl}\left(\boldsymbol{H}_{M,L_2} b_l\, \boldsymbol{g}\right)^H \left(\boldsymbol{H}_{M,L_2} b_l\, \boldsymbol{g}\right) \tag{28}$$

which can be evaluated to:

$$\left|\boldsymbol{\mathcal{M}}_{2,l}\right|_{fro}^2 = E_{wl}\, b_l^2\, \boldsymbol{g}^T\, \boldsymbol{H}_{M,L_2}^H\, \boldsymbol{H}_{M,L_2}\boldsymbol{g} \tag{29}$$

We can now equate the energies according to eq.(24) and eq.(29) respectively, and solve for b_l for each frequency f:

$$b_l = \sqrt{\frac{L_1}{\boldsymbol{g}^T\, \boldsymbol{H}_{M,L_2}^H\, \boldsymbol{H}_{M,L_2}\boldsymbol{g}}} \tag{30}$$

The b_l of eq.(30) are frequency-dependent gain factors or scaling factors, and can be used as coefficients of the equalization filter 722 for each frequency band, since b_l and $\boldsymbol{H}_{M,L_2}^H\, \boldsymbol{H}_{M,L_2}$ are frequency-dependent.

In the following, practical filter design for the equalization filter 722 is described.

Virtual microphone array radius and transfer function are taken into account as follows. To match the perceptual timbre effects of humans best, a microphone radius r_M of 0.09m is selected (the mean diameter of a human head is commonly assumed to be about 0.18m). $M \gg L1$ virtual microphones are placed on a sphere or radius r_M around the origin (sweet spot, listening position). Suitable positions are known [11]. One additional virtual microphone is added at the origin of the coordinate system.

The transfer matrices $\boldsymbol{H}_{M,L_2} \in \mathbb{C}^{Mx\,L_2}$ are designed using a plane wave or spherical wave model. For the latter, the amplitude attenuation effects can be neglected due to the gain and delay compensation stages. Let $h_{m,l}$ be an abstract matrix element of the transfer matrices $\boldsymbol{H}_{M,L}$, for the free field transfer function from speaker l to microphone m (which also indicate column and row indices of the matrices). The plane wave transfer function is given by

$$h_{m,l} = e^{ikr_m \cos(\gamma_{l,m})} \tag{31}$$

with i the imaginary unit, r_m the radius of the microphone position (ether r_M or zero for the origin position) and $\cos(\gamma_{l,m})$=$\cos\theta_l \cos\theta_m + \sin\theta_l \sin\theta_m \cos(\phi_l - \phi_m)$ the cosine of the spherical angles of the positions of speaker l and microphone m. The frequency dependency is given by $k = \frac{2\pi f}{c}$, with f the frequency and c the speed of sound. The spherical wave transfer function is given by:

$$h_{m,l} = e^{-ikr_{l,m}} \quad (32)$$

with $r_{l,m}$ the distance speaker l to microphone m.

The frequency response $B_{resp} \in \mathbb{C}^{L_1 x F_N}$ of the filter is calculated using a loop over F_N discrete frequencies and a loop over all input configuration speakers L_1:

Calculate $\mathbf{G}$ according to the above description (3-step procedure for design of optimal rendering matrices):

for (f=0; f=f+fstep; f<$F_N$fstep) /* loop over frequencies */
    k=2*pi*f/342;
    (… calculate $\boldsymbol{H}_{M,L_2}$(f) according to eq.(31) or eq.(32))
    $\breve{\boldsymbol{H}} = \boldsymbol{H}_{M,L_2}^{H} \boldsymbol{H}_{M,L_2}$
    for (l=1; l++; l<=L_1) /* loop over input channels */
        g= $\boldsymbol{G}$(:,l)
        $\boldsymbol{B}_{resp}$(l,f)= $\sqrt{\frac{L_1}{\boldsymbol{g}^T \breve{\boldsymbol{H}} \boldsymbol{g}}}$
    end
end

The filter responses can be derived from the frequency responses $\boldsymbol{B}_{resp}(l,f)$ using standard technologies. Typically, it is possible to derive a FIR filter design of order equal or less than 64, or IIR filter designs using cascaded bi-quads with even less computational complexity. Fig.9 and 10 show design examples.

In Fig.9, example frequency responses of filters for a remix of 5-channels ITU setup [9] (L,R,C,Ls,Rs) to +/- 30° 2-channel stereo, and an exemplary resulting 2x5 mixing matrix $\boldsymbol{G}$ are shown. The mixing matrix was derived as described above, using [2] for 500Hz. A plane wave model was used for the transfer functions. As shown, two of the filters (upper row, for two of the channels) have in principle low-pass (LP) characteristics, and three of the filters (lower rows, for the remaining three channels) have in principle high-pass (HP)

characteristics. It is intended that the filters do not have ideal HP or LP characteristics, because together they form an equalization filter (or equalization filter bank). Generally, not all the filters have substantially same characteristics, so that at least one LP and at least one HP filter is employed for the different channels.

In Fig.10, example responses of filters for a remix of 22 channels of the 22.2 NHK setup [10] to ITU 5-channel surround [9] are shown. In Fig.10b), the three filters of the first row of Fig.10a) are exemplarily shown. Also a resulting 5x22 mixing matrix G is shown, as obtained by the present invention.

The present invention can be used to adjust audio channel based content with arbitrary defined L_1 loudspeaker positions to enable replay to L_2 real-world loudspeaker positions. In one aspect, the invention relates to a method of rendering channel based audio of L_1 channels to L_2 channels, wherein a loudness & energy preserving mixing matrix is used. The matrix is derived by singular value decomposition, as described above in the section about design of optimal rendering matrices. In one embodiment, the singular value decomposition is applied to a conventionally derived mixing matrix.

In one embodiment, the matrix is scaled according to eq.(19) or (19') by a factor of $\sqrt{\frac{L_1}{s_m}}$ (for $L_1 \geq L_2$) , or by a factor of $\sqrt{\frac{L_2}{s_m}}$ (for $L_1 < L_2$).

Conventional matrices can be derived by using various panning methods, e.g. VBAP or robust panning. Further, conventional matrices use idealized input and output speaker positions (spherical projection, see above). Therefore, in one aspect, the invention relates to a method of filtering the L_1 input channels before applying the mixing matrix. In one embodiment, input signals that use different speaker positions are mapped to a spherical projection in a Delay & Gain Compensation block 71.

In one embodiment, equalization filters are derived from the frequency responses as described above.

In one embodiment, a device for rendering a first number L_1 of channels of channel-based audio signals (or content) to a second number L_2 of channels of channel-based audio signals (or content) is assembled out of at least the following building blocks/ processing blocks:

- input (and output) gain and delay compensation blocks 71,74, having the purpose to map the input and output speaker positions to a virtual sphere. Such spherical structure is required for the above-described mixing matrix to be applicable;
- equalization filters 722 derived by the method described above for filtering the first number L_1 of channels after input gain and delay compensation;
- a mixer unit 72 for mixing the first number L_1 of input channels to the second number L_2 of output channels by applying the energy preserving mixing matrix 724 as derived by the method described above. The equalization filters 722 may be part of the mixer unit 72, or may be a separate module;
- a signal overflow detection and clipping prevention block (or clipping unit) 73 to prevent signal overload to the signals of L_2 channels; and
- an output gain and delay correction block 74 (already mentioned above).

In one embodiment, a method for obtaining or generating an energy preserving mixing matrix $\boldsymbol{G}$ for mixing L1 input audio channels to L2 output channels comprises steps of obtaining s711 a first mixing matrix $\widehat{\boldsymbol{G}}$, performing s712 a singular value decomposition on the first mixing matrix $\widehat{\boldsymbol{G}}$ to obtain a singularity matrix $\boldsymbol{S}$, processing s713 the singularity matrix $\boldsymbol{S}$ to obtain a processed singularity matrix $\widehat{\boldsymbol{S}}$, determining s715 a scaling factor a, and calculating s716 an improved mixing matrix $\boldsymbol{G}$ according to $\boldsymbol{G} = a\,\boldsymbol{U}\,\widehat{\boldsymbol{S}}\,\boldsymbol{V}^{\boldsymbol{T}}$. One advantage of the improved mixing mode matrix $\boldsymbol{G}$ is that the perceived sound, loudness, timbre and spatial impression of multi-channel audio replayed on an arbitrary loudspeaker setup practically equals that of the original speaker setup. Thus, it is not required any more to locate loudspeakers strictly according to a predefined setup for enjoying a maximum sound quality and optimal perception of directional sound signals.

In one embodiment, an apparatus for rendering L1 channel-based input audio signals to L2 loudspeaker channels, where L1 is different from L2, comprises at least one of each of a determining unit for determining a mix type of the L1 input audio signals, wherein possible mix types include at least one of spherical, cylindrical and rectangular;
a first delay and gain compensation unit for performing a first delay and gain compensation on the L1 input audio signals according to the determined mix type, wherein a delay and gain compensated input audio signal with L1 channels and with a defined mix type is obtained;
a mixer unit for mixing the delay and gain compensated input audio signal for L2 audio channels, wherein a remixed audio signal for L2 audio channels is obtained;

a clipping unit for clipping the remixed audio signal, wherein a clipped remixed audio signal for L2 audio channels is obtained; and
a second delay and gain compensation unit for performing a second delay and gain compensation on the clipped remixed audio signal for L2 audio channels, wherein L2 loudspeaker channels are obtained.

Further, in one embodiment of the invention, an apparatus for obtaining an energy preserving mixing matrix $\boldsymbol{G}$ for mixing input channel-based audio signals for L1 audio channels to L2 loudspeaker channels comprises at least one processing element and memory for storing software instructions for implementing
a first calculation module for obtaining a first mixing matrix $\hat{G}$ from virtual source directions $\widehat{R_1}$ and target speaker directions $\widehat{R_2}$ wherein a panning method is used;
a singular value decomposition module for performing a singular value decomposition on the first mixing matrix $\hat{G}$ according to $\widehat{\boldsymbol{G}} = \boldsymbol{U}\,\boldsymbol{S}\,\boldsymbol{V^T}$, wherein $\boldsymbol{U} \in \mathcal{R}^{L_2 x\, L_2}$ and $\boldsymbol{V} \in \mathcal{R}^{L_1 x\, L_1}$ are orthogonal matrices and $\boldsymbol{S} \in \mathcal{R}^{L_1 x\, L_2}$ is a singularity matrix and has s first diagonal elements being the singular values of G in descending order and all other elements of $\boldsymbol{S}$ are zero;
a processing module processing the singularity matrix $\boldsymbol{S}$, wherein a quantized singularity matrix $\hat{S}$ is obtained with diagonal elements that are above a threshold set to one and diagonal elements that are below a threshold set to zero;
a counting module for determining a number s_m of diagonal elements that are set to one in the quantized singularity matrix $\hat{S}$;
a second calculation module for determining a scaling factor a according to
$a = \sqrt{\frac{L_1}{s_m}}\ for\ (L2 \leq L1)$ or $a = \sqrt{\frac{L_2}{s_m}}$ for $(L2 > L1)$; and
a third calculation module for calculating a mixing matrix $\boldsymbol{G}$ according to
$\boldsymbol{G} = a\,\boldsymbol{U}\,\widehat{\boldsymbol{S}}\,\boldsymbol{V^T}$.

Advantageously, the invention is usable for content loudness level calibration. If the replay levels of a mixing facility and of presentation venues are setup in the manner as described, switching between items or programs is possible without further level adjustments. For channel based content, this is simply achieved if the content is tuned to a pleasant loudness level at the mixing site. The reference for such pleasant listening level can either be the loudness of the whole item itself or an anchor signal.
If the reference is the whole item itself, this is useful for 'short form content', if the content is stored as a file. Besides adjustment by listening, a measurement of the loudness in

Loudness Units Full Scale (LUFS) according to EBU R128 [6] can be used to loudness adjust the content. Another name for LUFS is 'Loudness, K-weighted, relative to Full Scale' from ITU-R BS.1770 [7] (1 LUFS = 1 LKFS). Unfortunately [6] only supports content for setups up to 5-channel surround. It has not been investigated yet if loudness measures of 22-channel files correlate with perceived loudness if all 22 channels are factored by equal channel weights of one.

If the above-mentioned reference is an anchor signal, such as in a dialog, the level is selected in relation to this signal. This is useful for 'long form content' such as film sound, live recordings and broadcasts. An additional requirement, extending the pleasant listening level, is intelligibility of the spoken word here. Again, besides an adjustment by listening, the content may be normalized related a loudness measure, such as defined in ATSC A/85 [8]. First parts of the content are identified as anchor parts. Then a measure as defined in [7] is computed or these signals and a gain factor to reach the target loudness is determined. The gain factor is used to scale the complete item. Unfortunately, again the maximum number of channels supported is restricted to five.

Out of artistic considerations, content should be adjusted by listening at the mixing studio. Loudness measures can be used as a support and to show that a specified loudness is not exceeded. The energy E_w according to eq.(11) gives a fair estimate of the perceived loudness of such an anchor signal for frequencies over 200Hz. Because the K-filter suppresses frequencies lower than 200Hz [5], E_w is approximately proportional to the loudness measure.

It is noted that when a “speaker” is mentioned herein, a loudspeaker is meant. Generally, a speaker or loudspeaker is a synonym for any sound emitting device. It is noted that usually where speaker directions are mentioned in the specification or the claims, also speaker positions can be equivalently used (and vice versa).

While there has been shown, described, and pointed out fundamental novel features of the present invention as applied to preferred embodiments thereof, it will be understood that various omissions and substitutions and changes in the apparatus and method described, in the form and details of the devices disclosed, and in their operation, may be made by those skilled in the art without departing from the spirit of the present invention. E.g., although in the above embodiments, the number L1 of channels of the channel-based input audio signals is usually different from the number L2 of loudspeaker channels, it is clear that the invention can also be applied in cases where both numbers

are equal (so-called remix). This may be useful in several cases, e.g. if directional sound should be optimized for any irregular loudspeaker setup. Further, it is generally advantageous to use an energy preserving rendering matrix for rendering. It is expressly intended that all combinations of those elements that perform substantially the same function in substantially the same way to achieve the same results are within the scope of the invention.

Substitutions of elements from one described embodiment to another are also fully intended and contemplated. It will be understood that the present invention has been described purely by way of example, and modifications of detail can be made without departing from the scope of the invention.

Each feature disclosed in the description and (where appropriate) the claims and drawings may be provided independently or in any appropriate combination. Features may, where appropriate be implemented in hardware, software, or a combination of the two. Connections may, where applicable, be implemented as wireless connections or wired, not necessarily direct or dedicated, connections.

Reference numerals appearing in the claims are by way of illustration only and shall have no limiting effect on the scope of the claims.

Cited References

[1] Pulkki, V., "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", J. Audio Eng. Soc., vol. 45, pp. 456-466 (1997 June).

[2] Poletti, M., "Robust two-dimensional surround sound reproduction for non-uniform loudspeaker layouts". J. Audio Eng. Soc., 55(7/8):598–610, July/August 2007.

[3] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," J. Acoust. Soc. Am. 94 (5), 2992-3000 (1993).

[4] Fazi, F.; Yamada, T; Kamdar, S.; Nelson P.A.; Otto, P., "Surround Sound Panning Technique Based on a Virtual Microphone Array", AES Convention:128 (May 2010)Paper Number:8119

[5] Shin, M.; Fazi, F.; Seo, J.; Nelson, P. A. "Efficient 3-D Sound Field Reproduction", AES Convention:130 (May 2011)Paper Number:8404

[6] EBU Technical Recommendation R128, "Loudness Normalization and Permitted Maximum Level of Audio Signals", Geneva, 2010 [http://tech.ebu.ch/docs/r/r128.pdf]

[7] ITU-R Recommendation BS.1770-2, "Algorithms to measure audio programme loudness and true-peak audio level", Geneva, 2011.

[8] ATSC A/85, "Techniques for Establishing and Maintaining Audio Loudness for Digital Television", Advanced Television Systems Committee, Washington, D.C., July 25, 2011.

[9] ITU-R BS 775-1 (1994)

[10] Hamasaki, K.; Nishiguchi,T.; Okumura, R.; Nakayama, Y. ; Ando, A. "A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV)," SMPTE Motion Imaging J., pp.40-49, Apr. 2008.

[11] Jörg Fliege and Ulrike Maier. A two-stage approach for computing cubature formulae for the sphere. Technical report, Fachbereich Mathematik, Universität Dortmund, 1999. Node numbers & report can be found at http://www.personal.soton.ac.uk/jf1w07/nodes/nodes.html

Claims

1. A method for rendering L1 channel-based input audio signals ($w1_1$) to L2 loudspeaker channels, where L1 is different from L2, the method comprising steps of
   - determining (s60) a mix type of the L1 input audio signals, wherein possible mix types include at least one of spherical, cylindrical and rectangular;
   - performing a first delay and gain compensation (s61) on the L1 input audio signals according to the determined mix type, wherein a delay and gain compensated input audio signal (q71) with L1 channels and with a defined mix type is obtained;
   - mixing (s624) the delay and gain compensated input audio signal for L2 audio channels, wherein a remixed audio signal for L2 audio channels is obtained;
   - clipping (s63) the remixed audio signal, wherein a clipped remixed audio signal for L2 audio channels is obtained; and
   - performing a second delay and gain compensation (s64) on the clipped remixed audio signal for L2 audio channels, wherein L2 loudspeaker channels ($w2_2$) are obtained.

2. Method according to claim 1, further comprising a step of filtering (s622) the delay and gain compensated input audio signal (q71) with L1 channels, wherein a filtered delay and gain compensated input audio signal is obtained, and wherein the step of mixing (s624) uses the filtered delay and gain compensated input audio signal.

3. Method according to claim 2, wherein the filtering (s622) of the delay and gain compensated input audio signal with L1 channels uses an equalizer filter with different types of filters for the channels, wherein at least one channel uses a high-pass filter and at least one channel uses a low-pass filter.

4. Method according to one of the claims 1-3, wherein the defined mix type is spherical.

5. Method according to one of the claims 1-4, wherein the step of mixing (s624) the delay and gain compensated input audio signal for L2 audio channels uses an energy preserving mixing matrix $\boldsymbol{G}$ that is obtained by steps of
   - obtaining a first mixing matrix $\hat{G}$ from virtual source directions $\widehat{R_1}$ and target speaker directions $\widehat{R_2}$ using a panning method;
   - performing a singular value decomposition on the first mixing matrix $\hat{G}$ according to $\widehat{\boldsymbol{G}} = \boldsymbol{U}\,\boldsymbol{S}\,\boldsymbol{V}^{\boldsymbol{T}}$, wherein $\boldsymbol{U} \in \mathcal{R}^{L_2 x\, L_2}$ and $\boldsymbol{V} \in \mathcal{R}^{L_1 x\, L_1}$ are orthogonal matrices and

$\boldsymbol{S} \in \mathcal{R}^{L_1 x\, L_2}$ is a singularity matrix and has s first diagonal elements being the singular values of $\boldsymbol{G}$ in descending order and all other elements of $\boldsymbol{S}$ are zero;

- processing the singularity matrix $\boldsymbol{S}$, wherein a quantized singularity matrix $\widehat{\boldsymbol{S}}$ is obtained with diagonal elements that are above a threshold set to one and diagonal elements that are below a threshold set to zero;
- determining a number $\mathcal{s}_m$ of diagonal elements that are set to one in the quantized singularity matrix $\hat{S}$;
- determining a scaling factor a according to $a = \sqrt{\frac{L_1}{\mathcal{s}_m}}\ for\ (L2 \leq L1)$ or $a = \sqrt{\frac{L_2}{\mathcal{s}_m}}$ for $(L2 > L1)$; and
- calculating a mixing matrix $\boldsymbol{G}$ according to $\boldsymbol{G} = a\, \boldsymbol{U}\, \widehat{\boldsymbol{S}}\, \boldsymbol{V^T}$.

6. Method according to one of the claims 1-5, wherein the input signal is optimized for L1 regular loudspeaker positions and the rendering is optimized for L2 arbitrary loudspeaker positions, wherein at least one of the arbitrary loudspeaker positions is different from the regular loudspeaker positions.

7. A computer-implemented method (s710) for generating an energy preserving mixing matrix $\boldsymbol{G}$ for mixing input channel-based audio signals for L1 audio channels to L2 loudspeaker channels, the method comprising steps executed by the computer of

- obtaining (s711) a first mixing matrix $\hat{G}$ from virtual source directions $\widehat{R_1}$ and target speaker directions $\widehat{R_2}$ wherein a panning method is used;
- performing (s712) a singular value decomposition on the first mixing matrix $\hat{G}$ according to $\widehat{\boldsymbol{G}} = \boldsymbol{U}\, \boldsymbol{S}\, \boldsymbol{V^T}$, wherein $\boldsymbol{U} \in \mathcal{R}^{L_2 x\, L_2}$ and $\boldsymbol{V} \in \mathcal{R}^{L_1 x\, L_1}$ are orthogonal matrices and $\boldsymbol{S} \in \mathcal{R}^{L_1 x\, L_2}$ is a singularity matrix and has s first diagonal elements being the singular values of G in descending order and all other elements of $\boldsymbol{S}$ are zero;
- processing (s713) the singularity matrix $\boldsymbol{S}$, wherein a quantized singularity matrix $\hat{S}$ is obtained with diagonal elements that are above a threshold set to one and diagonal elements that are below a threshold set to zero;
- determining (s714) a number $\mathcal{s}_m$ of diagonal elements that are set to one in the quantized singularity matrix $\hat{S}$;
- determining (s715) a scaling factor a according to $a = \sqrt{\frac{L_1}{\mathcal{s}_m}}\ for\ (L2 \leq L1)$ or $a = \sqrt{\frac{L_2}{\mathcal{s}_m}}$ for $(L2 > L1)$; and
- calculating (s716) a mixing matrix $\boldsymbol{G}$ according to $\boldsymbol{G} = a\, \boldsymbol{U}\, \widehat{\boldsymbol{S}}\, \boldsymbol{V^T}$.

8. An apparatus (70) for rendering L1 channel-based input audio signals ($w1_1$) to L2 loudspeaker channels, where L1 is different from L2, the apparatus comprising at least one of each of
   - a determining unit (75) for determining a mix type of the L1 input audio signals, wherein possible mix types include at least one of spherical, cylindrical and rectangular;
   - a first delay and gain compensation unit (71) for performing a first delay and gain compensation on the L1 input audio signals according to the determined mix type, wherein a delay and gain compensated input audio signal (q71) with L1 channels and with a defined mix type (q72) is obtained;
   - a mixer unit (72) for mixing the delay and gain compensated input audio signal (q71) for L2 audio channels, wherein a remixed audio signal (q72) for L2 audio channels is obtained;
   - a clipping unit (73) for clipping the remixed audio signal (q72), wherein a clipped remixed audio signal (q73) for L2 audio channels is obtained; and
   - a second delay and gain compensation unit (74) for performing a second delay and gain compensation on the clipped remixed audio signal (q73) for L2 audio channels, wherein L2 loudspeaker channels ($w2_2$) are obtained.

9. Apparatus according to claim 8, further comprising an equalization filter (722) for filtering the delay and gain compensated input audio signal (q71) with L1 channels, wherein a filtered delay and gain compensated input audio signal (q722) is obtained.

10. Apparatus according to claim 9, wherein the equalization filter (722) comprises different types of filters that are used for the channels, wherein at least one channel uses a high-pass filter and at least one channel uses a low-pass filter.

11. Apparatus according to one of the claims 8-10, wherein the defined mix type is spherical.

12. Apparatus according to one of the claims 8-11, wherein the mixer unit (724) mixes the delay and gain compensated input audio signal (q71) for L2 audio channels uses an energy preserving mixing matrix $\boldsymbol{G}$ that is obtained by a mixing matrix generation unit that comprises one or more processors for implementing
   - a first calculating module for obtaining a first mixing matrix $\hat{G}$ from virtual source directions $\widehat{R_1}$ and target speaker directions $\widehat{R_2}$ using a panning method;

- a singular value decomposition module for performing a singular value decomposition on the first mixing matrix $\hat{G}$ according to $\widehat{\boldsymbol{G}} = \boldsymbol{U}\,\boldsymbol{S}\,\boldsymbol{V}^{\boldsymbol{T}}$, wherein $\boldsymbol{U} \in \mathcal{R}^{L_2 x\, L_2}$ and $\boldsymbol{V} \in \mathcal{R}^{L_1 x\, L_1}$ are orthogonal matrices and $\boldsymbol{S} \in \mathcal{R}^{L_1 x\, L_2}$ is a singularity matrix and has s first diagonal elements being the singular values of G in descending order and all other elements of $\boldsymbol{S}$ are zero;
- a processing module for processing the singularity matrix $\boldsymbol{S}$, wherein a quantized singularity matrix $\hat{S}$ is obtained with diagonal elements that are above a threshold set to one and diagonal elements that are below a threshold set to zero;
- a counting module for determining a number s_m of diagonal elements that are set to one in the quantized singularity matrix $\hat{S}$;
- a second calculating module for determining a scaling factor a according to $a = \sqrt{\frac{L_1}{s_m}}\ for\ (L2 \leq L1)$ or $a = \sqrt{\frac{L_2}{s_m}}$ for $(L2 > L1)$; and
- a third calculating module for calculating a mixing matrix $\boldsymbol{G}$ according to $\boldsymbol{G} = a\,\boldsymbol{U}\,\widehat{\boldsymbol{S}}\,\boldsymbol{V}^{\boldsymbol{T}}$.

13. Apparatus according to one of the claims 8-12, wherein the input signal is optimized for L1 regular loudspeaker positions and the rendering is optimized for L2 arbitrary loudspeaker positions, wherein at least one of the arbitrary loudspeaker positions is different from the regular loudspeaker positions.

14. Apparatus for obtaining an energy preserving mixing matrix $\boldsymbol{G}$ for mixing input channel-based audio signals for L1 audio channels to L2 loudspeaker channels, comprising at least one processing element for implementing
    - a first calculation module for obtaining a first mixing matrix $\hat{G}$ from virtual source directions $\widehat{R_1}$ and target speaker directions $\widehat{R_2}$ wherein a panning method is used;
    - a singular value decomposition module for performing a singular value decomposition on the first mixing matrix $\hat{G}$ according to $\widehat{\boldsymbol{G}} = \boldsymbol{U}\,\boldsymbol{S}\,\boldsymbol{V}^{\boldsymbol{T}}$, wherein $\boldsymbol{U} \in \mathcal{R}^{L_2 x\, L_2}$ and $\boldsymbol{V} \in \mathcal{R}^{L_1 x\, L_1}$ are orthogonal matrices and $\boldsymbol{S} \in \mathcal{R}^{L_1 x\, L_2}$ is a singularity matrix and has s first diagonal elements being the singular values of G in descending order and all other elements of $\boldsymbol{S}$ are zero;
    - a processing module processing the singularity matrix $\boldsymbol{S}$, wherein a quantized singularity matrix $\hat{S}$ is obtained with diagonal elements that are above a threshold set to one and diagonal elements that are below a threshold set to zero;
    - a counting module for determining a number s_m of diagonal elements that are set to one in the quantized singularity matrix $\hat{S}$;

- a second calculation module for determining a scaling factor a according to $a = \sqrt{\frac{L_1}{s_m}}\ for\ (L2 \leq L1)$ or $a = \sqrt{\frac{L_2}{s_m}}$ for $(L2 > L1)$; and
- a third calculation module for calculating a mixing matrix $\boldsymbol{G}$ according to $\boldsymbol{G} = a\ \boldsymbol{U}\ \widehat{\boldsymbol{S}}\ \boldsymbol{V^T}$.

15. A computer readable storage medium having stored thereon instructions that when executed on a computer cause the computer to perform a method according to one of the claims 1-6.

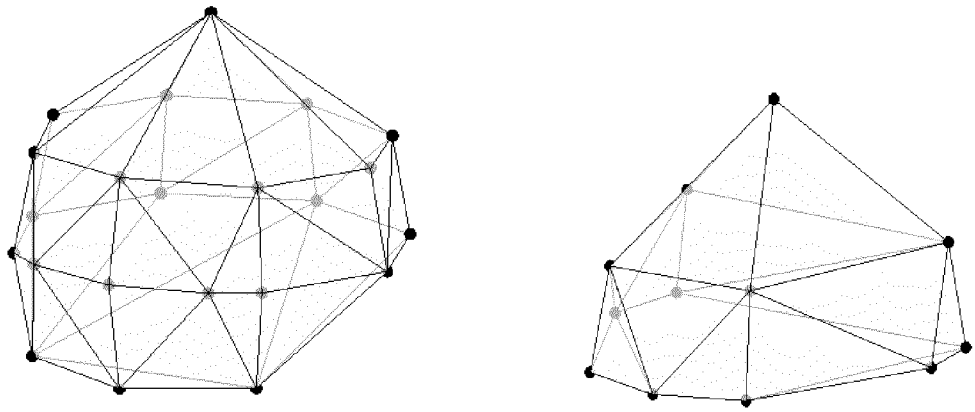


Fig.1

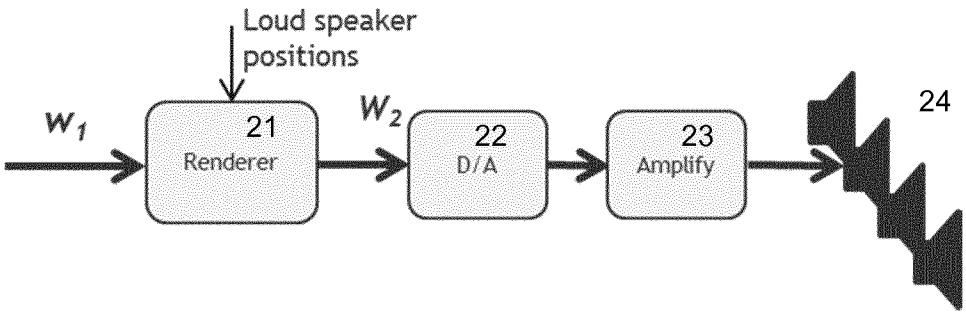


Fig.2

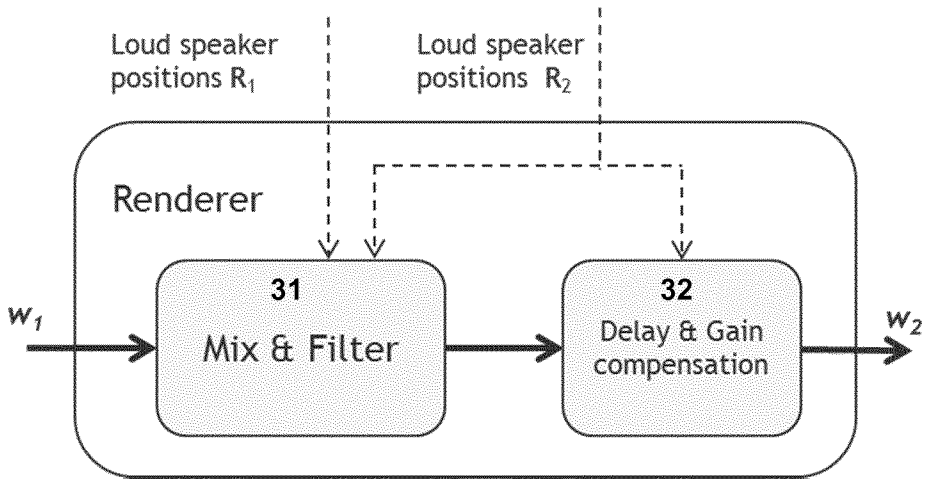


Fig.3

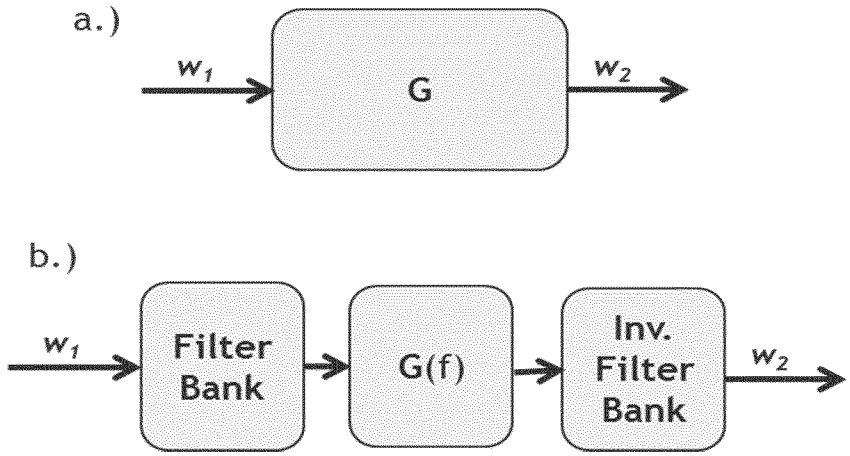


Fig.4

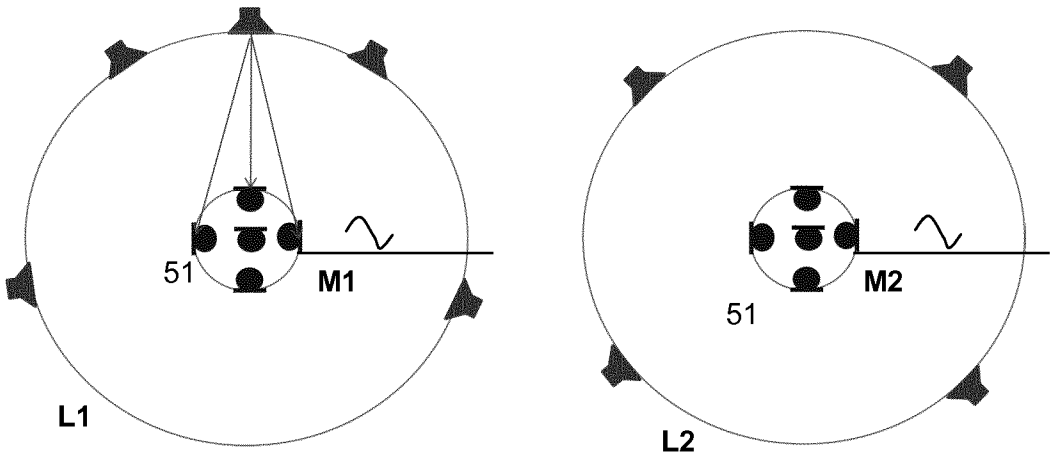


Fig.5

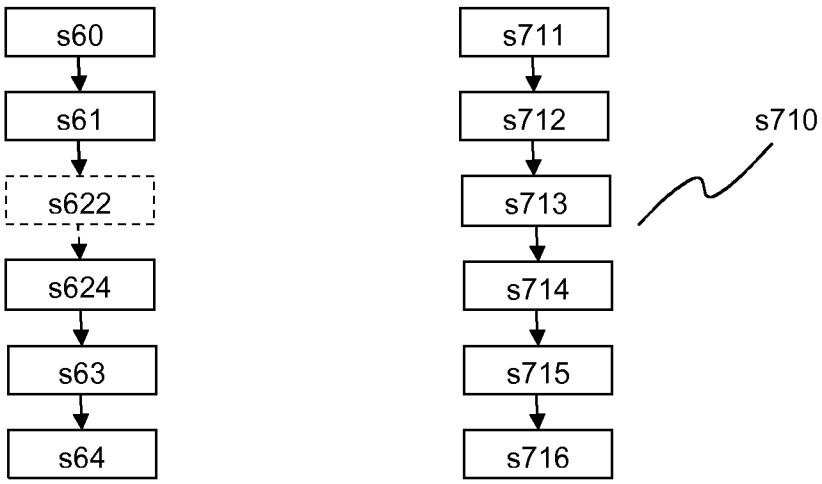


Fig.6

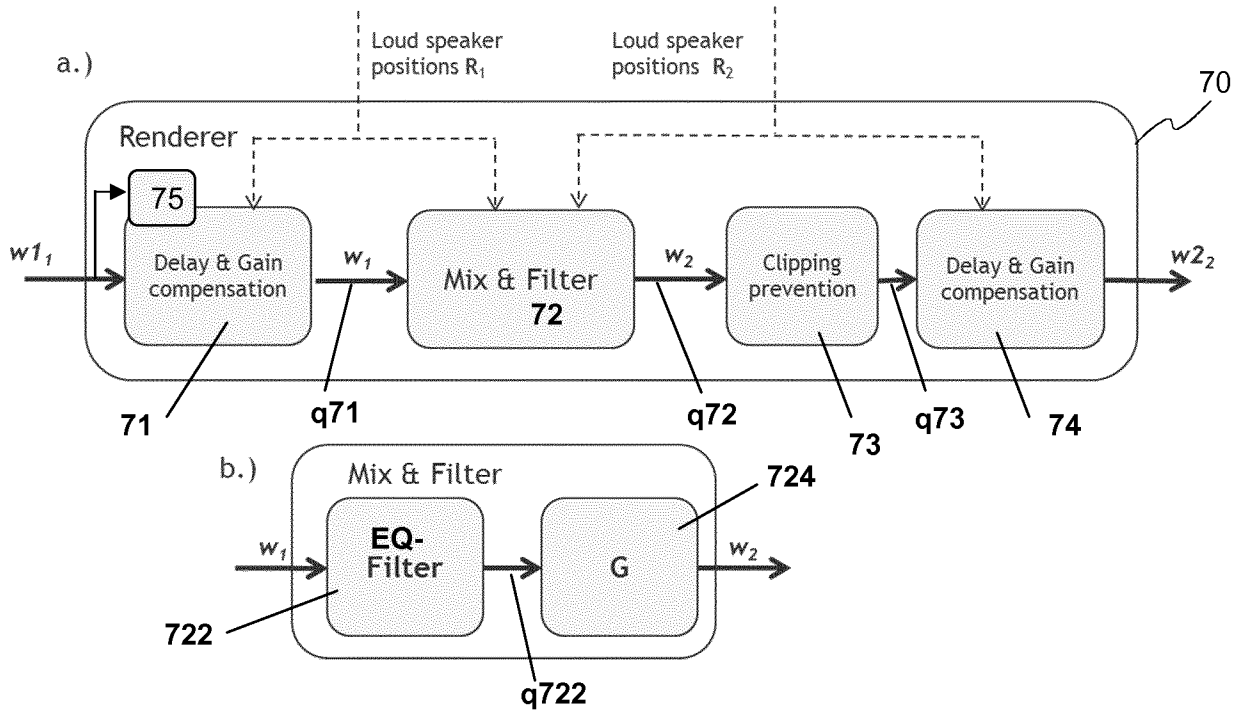

a.)
Loud speaker positions R_1
Loud speaker positions R_2
70
Renderer
75
$w1_1$
Delay & Gain compensation
w_1
Mix & Filter
72
w_2
Clipping prevention
Delay & Gain compensation
$w2_2$
71
q71
q72
73
q73
74
b.)
Mix & Filter
724
w_1
EQ-Filter
G
w_2
722
q722

Fig.7

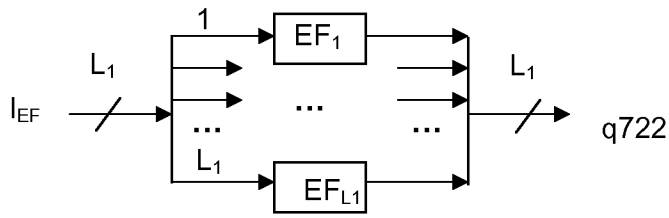
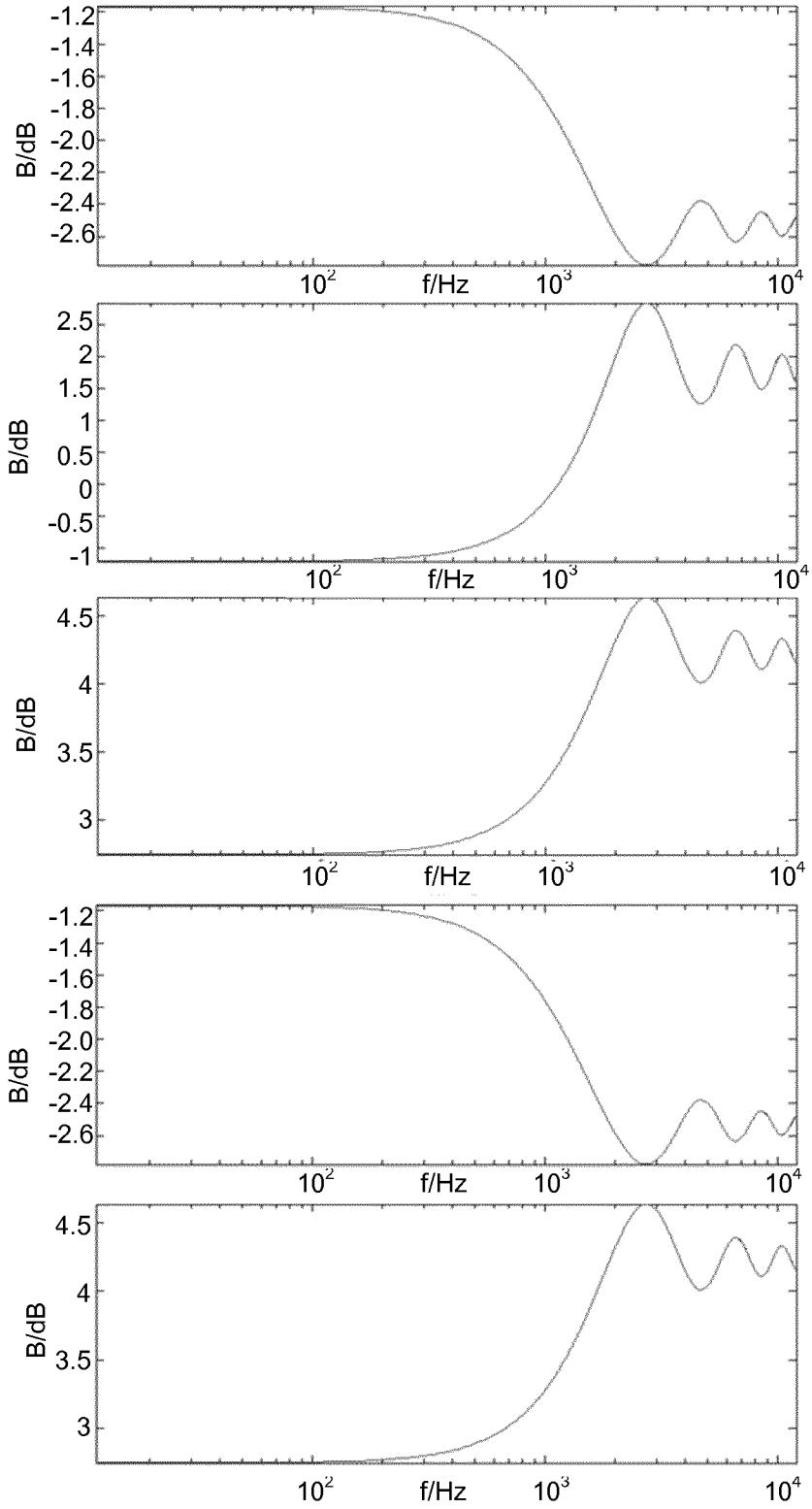

1
EF_1
L_1
I_{EF}
L_1
L_1
EF_{L1}
q722

Fig.8



a) Filter l=1

b) Filter l=2

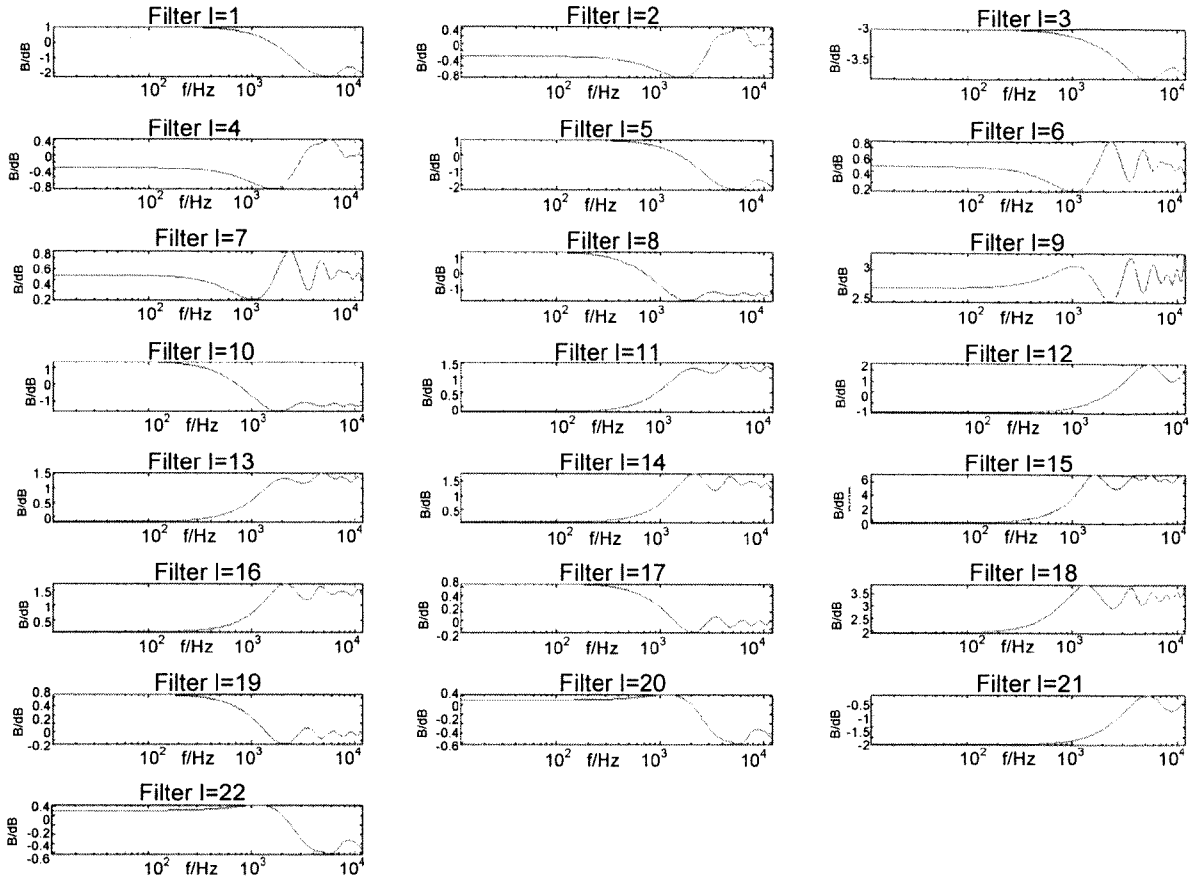
c) Filter l=3

d) Filter l=4

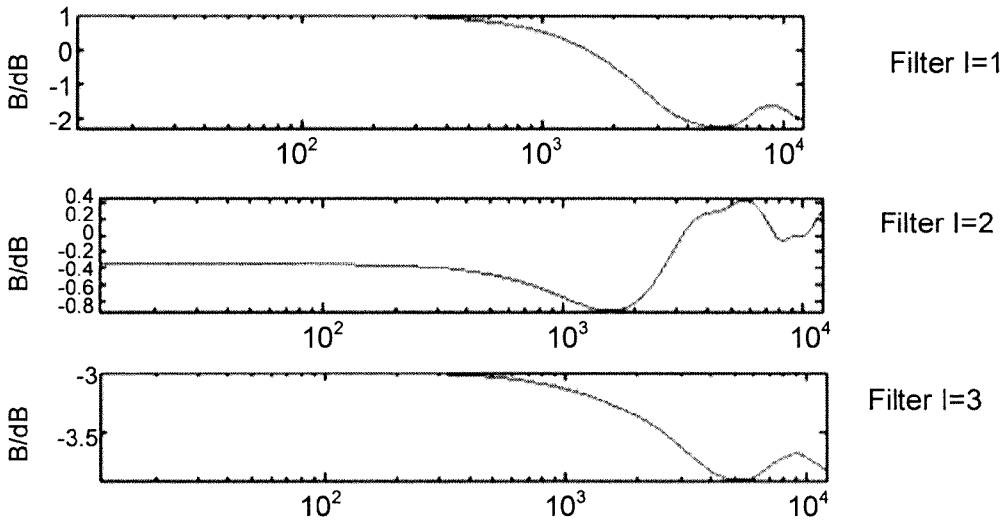
e) Filter l=5

$$\boldsymbol{G} = \begin{bmatrix} 1.33 & -0.18 & 0.57 & 0.60 & 0.12 \\ -0.18 & 1.38 & 0.57 & 0.12 & 0.60 \end{bmatrix}$$

Fig.9



a)



b)

$$G = \begin{bmatrix} 1.21 & 0.96 & -0.06 & 0.00 & -0.04 & 0.16 & 0.01 & -0.18 & -0.09 & 0.00 & 0.85 & 0.23 & 0.00 & 0.17 & 0.14 & 0.03 & -0.12 & -0.06 & 0.02 & 1.04 & 0.13 & -0.02 \\ -0.04 & 0.00 & -0.06 & 0.96 & 1.21 & 0.01 & 0.16 & 0.00 & -0.09 & -0.18 & 0.00 & 0.23 & 0.85 & 0.03 & 0.14 & 0.17 & 0.02 & -0.06 & -0.12 & -0.02 & 0.13 & 1.04 \\ -0.31 & 0.17 & 1.54 & 0.17 & -0.31 & 0.02 & 0.02 & -0.04 & -0.07 & -0.04 & 0.05 & 0.76 & 0.05 & 0.07 & 0.20 & 0.07 & -0.00 & -0.03 & -0.00 & -0.12 & 1.05 & -0.12 \\ 0.03 & -0.11 & -0.00 & 0.01 & -0.00 & 0.91 & -0.16 & 1.13 & 0.50 & -0.06 & 0.11 & 0.01 & 0.01 & 0.82 & 0.28 & -0.10 & 1.00 & 0.48 & 0.01 & 0.04 & -0.00 & 0.01 \\ 0.00 & 0.01 & -0.00 & -0.11 & 0.03 & -0.16 & 0.91 & -0.06 & 0.50 & 1.13 & 0.01 & 0.01 & 0.11 & -0.10 & 0.28 & 0.82 & 0.01 & 0.48 & 1.00 & 0.01 & -0.00 & 0.04 \end{bmatrix}$$

c)

Fig.10