

Apparatus and method for rendering audio objectsDescription5 Technical field

The invention relates to the technical field of audio reproduction. Specifically, reproduction of multichannel audio with reproduction of elevated or lowered height sounds is described herein.

10

Background

For sound reproduction, there are different kinds of systems which differ with regard to their complexity and reproduction quality. The reference for movie sound is the cinema. Cinemas provide multi-channel surround sound, with loudspeakers installed not only in the front of the listener (usually behind the screen), but additionally on the sides and rear, and recently also on the ceiling. The side and rear loudspeakers enable a horizontally enveloping sound reproduction, which can be further enhanced by vertically engulfing sound using height and ceiling loudspeakers.

20

With latest coding techniques, immersive, interactive, and object-based audio content can not only be used in professional environments, but can also conveniently be transmitted into the consumer's home, adding further features and dimensions, such as e.g. height reproduction.

Enhanced reproduction setups for realistic sound reproduction use loudspeakers not only mounted in the horizontal plane (usually at or close to ear-height of the listener), but additionally also loudspeakers spread in vertical direction. Those loudspeakers are e.g. elevated (mounted on the ceiling, or at some angle above head height) or are placed below the listener's ear height (e.g. on the floor, or on some intermediate or specific angle).

30

Often it is inconvenient or impossible to install loudspeakers at top or bottom directions.

In a home environment, likely only enthusiasts will install the number of loudspeakers needed to replicate the loudspeaker setups that are used in professional environments, research labs, or cinemas. Here, the term loudspeaker setup does also include devices and topologies like soundbars, TVs with built in loudspeakers, boomboxes, sound plates, loudspeaker arrays, smart speakers, and so forth.

35

Nonetheless, when rendering sound for an immersive sound experience or virtual reality, it is often desirable to render sound also in height (top and bottom) directions (denoted "top and bottom directions" in the following. Of course, not always both directions have to be processed, so this is equivalent to "(either) top or bottom directions" or "top/bottom directions").

5

Therefore, the need arises to render sound in top and bottom directions without having height loudspeakers, e.g. top loudspeakers and/or bottom loudspeakers.

10

A convenient alternative to those rather complex setups is compact reproduction systems that use signal processing means to generate a comparable or similar spatial auditory perception as the enhanced loudspeaker setups. Here, the term reproduction systems include all devices and topologies for audio reproduction like setups comprising a number of individual loudspeakers, soundbars, TVs with built in loudspeakers, boomboxes, sound plates, loudspeaker arrays, smart speakers, and so forth.

15

A practical method and an apparatus to achieve this is presented in the following.

Summary of the invention

20

It is an object of the present invention to provide a more efficient rendering of audio objects which allows 3D panning, wherein the increase in efficiency relates, for instance, to rendering stability, improved panning accuracy, computational efficiency and/or suitability with a larger number of loudspeaker setups, a changing number of loudspeakers, changing loudspeaker positions, changing listeners positions, changing object positions.

25

This object is achieved by the subject matter of the independent claims.

30

A more efficient rendering of audio objects, which allows 3D panning, is achieved by performing the panning in two stages, namely at least one horizontal in-layer panning leading to a first virtual (speaker) position and a second virtual or real (speaker) position, which is vertically offset, and another panning vertically between the two positions. Although acting in such a manner seems to increase the computational complexity, this staged processing increases, in fact, the stability of the rendering and the precision of localization of the intended virtual position. Moreover, the staged processing enables to perform, according to an embodiment, the panning by use of amplitude panning gains only, i.e. phase processing is not necessary, thereby rendering the computational complexity low. Even further, the rendering is flexible with respect to applicability to a variety of loudspeaker setups.

35

Embodiments of the present application refer to an apparatus for generating loudspeaker signals for a plurality of loudspeakers so that an application of the loudspeaker signals at the plurality of loudspeakers renders at least one audio object at an intended virtual position. The apparatus comprises an interface configured to receive an audio input signal which represents the at least one audio object. It may be one of a channel-based audio signal, object-based audio signal, and/or scene-based audio signal. A first panning gain determiner is configured to determine, depending on the intended virtual position, first panning gains for a first set of loudspeakers of the plurality of loudspeakers, which are arranged within a first layer set of one or more first horizontal layers, the first panning gains defining a derivation of first partial loudspeaker signals from the at least one audio input signal, which are associated with a rendering of the at least one audio object at a first virtual position upon application of the first partial loudspeaker signals onto the first set of loudspeakers. This is the afore-mentioned in-layer panning. A vertical panning gain determiner is configured to determine, depending on the intended virtual position, further panning gains for a panning (or fading) between the first partial loudspeaker signals and one or more second partial loudspeaker signals which is to be applied to a second set of one or more loudspeakers and is associated with a rendering of the at least one audio object at a second position, which is vertically offset relative to the first position, so as to pan between the first virtual position and the second position. This is the vertical panning. The one or more second partial loudspeaker signals may be the result of another in-layer panning in which case the second position is a second virtual position or the second position may be the real position of another one of the loudspeakers, which is positioned vertically offset to the first set of loudspeakers. The apparatus is configured to compose the loudspeaker signals from the first partial loudspeaker signals and the one or more second partial loudspeaker signals using the first panning gains and the further panning gains. That is, in the composition, the first and further panning gains are actually applied onto the audio input signal, thereby leading to the loudspeaker signals. There may possibly be one or more loudspeaker signals, for the generation of which just one of the panning gains is to be used, such as for the just-mentioned second loudspeaker positioned at the real loudspeaker position and fed with the second partial loudspeaker signal.

According to some embodiments, as said, the second set of one or more loudspeakers comprises more than one loudspeaker, and the one or more second partial loudspeaker signals comprise more than one second partial loudspeaker signals and the apparatus further comprises a second panning gain determiner, configured to determine, depending on the intended virtual position, second panning gains for the second set of loudspeakers, the second panning gains defining a derivation of second partial loudspeaker signals from the at least one

audio input signal, wherein the apparatus is configured to compose the loudspeaker signals from the first and second partial loudspeaker signals using the first and second panning gains and the further panning gains. Here, according to an embodiment, the second partial loudspeaker signals may be derived from the at least one audio signal by spectral shaping, so that the second position is a virtual position above or below the second layer set, such as not between or within any of the one or more first horizontal layers, and the one or more second horizontal layers, within which the second set of loudspeakers are arranged, but on one side, vertically, relative to these horizontal layers. In accordance with corresponding embodiments, an apparatus results which is for generating loudspeaker signals for a plurality of loudspeakers so that an application of the loudspeaker signals at the plurality of loudspeakers renders at least one audio object at an intended virtual position, wherein the plurality of loudspeakers are distributed onto one or more horizontal layers, the apparatus comprising an interface configured to receive an audio input signal which represents the at least one audio object, a first loudspeaker signal set determiner, configured to determine, depending on the intended virtual position, first panning gains, e.g., as said pure amplitude panning gains so that the first virtual position is in-between positions of the first set of loudspeakers, for a first set of loudspeakers of the plurality of loudspeakers, and use the first panning gains to derive first partial loudspeaker signals from the at least one audio input signal, which are associated with a rendering of the at least one audio object at a first virtual position upon application of the first partial loudspeaker signals onto the first set of loudspeakers, a second loudspeaker signal set determiner, configured to, by spectral shaping, derive second partial loudspeaker signals from the at least one audio signal, the second partial loudspeaker signals being associated with a rendering of the at least one audio object at a second virtual position upon application of the second partial loudspeaker signals onto the second set of loudspeakers, the second virtual position being above or below the one or more horizontal layers, e.g. not between or within any of the one or more horizontal layers, but on one side, vertically, relative to the one or more horizontal layers, and a vertical panning gain determiner configured to, depending on the intended virtual position, determine second panning gains for the first and second partial loudspeaker signals so as to pan between the first and second virtual positions, and a composer configured to compose the loudspeaker signals from the first and second partial loudspeaker signals using the second panning gains.

Embodiments set-out herein reveal, thus, a concept for rendering at least one audio object to a set of loudspeakers from at least one audio input signal. In brief, audio input signals may comprise information about audio objects that are to be output by the loudspeakers. For example, such an audio object can be a sound of a helicopter flying in a movie, sound of an instrument playing in an orchestra, or sound of a voice. The audio object is rendered using

loudspeakers. The audio input signal is processed to determine how the audio object is to be output at individual loudspeakers. For this each audio input signal is associated with position information of the at least one audio object. Such position information can be static, e.g. the violin is located on the left of the orchestra, the speaker is in front of the listener, or dynamic, e.g. the helicopter flies from right to left. The set of loudspeakers used to render the audio object may comprise one or more groups of loudspeakers, each group located in one horizontal layer. An additional loudspeaker may be a physical or virtual loudspeaker, located above or below the one or more groups.

That means that for the set of loudspeakers an association with layers and positions offset to the layers above or below the layers may be defined. For example, the setup can comprise four loudspeakers in one layer, e.g. all at the same height, and one physical or virtual loudspeaker higher, e.g. elevated, above the four other loudspeakers. This setup would then have one layer. Additional one or more layers are also possible.

Advantageous embodiments are the subject of dependent claims. In particular, preferred embodiments of the present application are described below with respect to the figures among which:

Fig. 1 shows a block diagram of an apparatus for audio rendering in accordance with an embodiment;

Fig. 2 shows another embodiment for an apparatus for audio rendering, here described to comprise the possibility of horizontal panning for both partial loudspeaker signal sets as well as the equalization for one of them;

Fig. 3 shows schematically an example loudspeaker setup and a listener positioned in between the loudspeakers, with additionally illustrating the consideration of a virtual top loudspeaker for audio rendering;

Fig. 4 shows a schematic diagram of the scenario of Fig. 3 with illustrating the first (horizontal) panning;

Fig. 5a shows the scenario of Fig. 3 with illustrating the usage of the equalization or spectral shaping in order to provide a monaural cue to achieve a virtual top loudspeaker;

- Fig. 5b shows the situation of Fig. 5a3 with illustrating the panning between loudspeakers recruited to participate in rendering the virtual top loudspeaker and the gains used to locate the virtual top loudspeaker;
- 5 Fig. 6 shows a block diagram of an apparatus for audio rendering varied compared to the embodiment of Fig. 2 by a different order between horizontal panning and equalization for the rendering of the top/bottom virtual loudspeaker;
- 10 Fig. 7 shows a block diagram of another embodiment for an apparatus for audio rendering or, shown differently, a block diagram of the elements of the apparatus of Fig. 1 participating in rendering the audio object for an intended virtual position in between two available loudspeaker layers;
- 15 Fig. 8 shows a block diagram illustrating, in addition to the elements of Fig. 7, the possibility of considering the listener's position;
- Fig. 9 shows a schematic top view of a possible loudspeaker setup, here a 5.0 loudspeaker setup;
- 20 Fig. 10 shows another schematic three-dimensional view of another example for a loudspeaker setup, here a 5.0+2H loudspeaker setup;
- Figs. 11, 12 show schematic diagrams so as to illustrate the two-stage process in performing the audio rendering of an object at an intended virtual position in between two
25 available layers, here for the example of using a 5.0+4H loudspeaker setup;
- Figs. 13, 14 illustrate the two-stage rendering of an object at an intended virtual position vertically offset to the available layers, here exemplary to the top of all layers, and
30
- Fig. 15 shows examples for shaping functions used in the equalization or spectral shaping so as to form a monaural cue for rendering the virtual top/bottom loudspeaker signal.
- 35 The following description starts with a description of an embodiment of an apparatus for generating loudspeaker signals for a plurality of loudspeakers. More specific embodiments are

outlined herein below along with a description of details which may, individually or in groups, apply to the apparatus of Fig. 1.

5 The apparatus of Fig. 1 is generally indicated using reference sign 10 and is for generating loudspeaker signals 12 for a plurality of loudspeakers 14 in a manner so that an application of the loudspeaker signals 12 at or to the plurality of loudspeakers 14 renders at least one audio object at an intended virtual position.

10 The apparatus 10 might be configured for a certain arrangement of loudspeakers 14, i.e., for certain positions in which the plurality of loudspeakers 14 are positioned or positioned and oriented. The apparatus may, however, alternatively be able to be configurable for different loudspeaker arrangements of loudspeakers 14. Likewise, the number of loudspeakers 14 may be two or more and the apparatus may be designed for a set number of loudspeakers 14 or may be configurable to deal with any number of loudspeakers 14.

15 The apparatus 10 comprises an interface 16 at which apparatus 10 receives an audio signal 18 which represents the at least one audio object. For the time being, let's assume that the audio input signal 18 is a mono audio signal which represents the audio object such as the sound of a helicopter or the like. Additional examples and further details are provided below.
20 In any case, the audio signal 18 may represent the audio object in time domain, in frequency domain or in any other domain and it may represent the audio object in a compressed manner or without compression.

As depicted in Fig. 1, the apparatus 10 further comprises a position input for receiving the
25 intended virtual position. That is, at position input 20, the apparatus 10 is notified about the intended virtual position to which the audio object shall virtually be rendered by the application of the loudspeaker signals 12 at loudspeakers 14. That is, the apparatus 10 receives at input 20 the information of the intended virtual position, and this information may be provided relative to the arrangement/position of loudspeakers 14, relative to the position and/or head orientation
30 of the listener and/or relative to real-world coordinates. This information could e.g. be based on Cartesian coordinate systems, or polar coordinate systems. It could e.g. be based on a room centric coordinate system or a listener centric coordinate system, either as a cartesian, or polar coordinate system.

35 As depicted in Fig. 1, apparatus 10 comprises a first panning gain determiner 22 configured to determine, depending on the intended virtual position 21 received at input 20, first panning gains 24 for a first set 26 of loudspeakers out of the plurality of loudspeakers 14. This set 26

of loudspeakers is arranged within a first layer set of one or more first horizontal layers. That is, this set 26 of loudspeakers, quasi, are arranged at similar heights. The first panning gains 24 define a derivation of, or participate in a generation of, first partial loudspeaker signals 28 from the at least one audio input signal 18, which first partial loudspeaker signals 28 are associated with a rendering of the at least one audio object at a first virtual position upon an application of the first partial loudspeaker signals onto the first set 26 of loudspeakers. As outlined in more detail below, the first panning gain determiner 22 may, according to an embodiment, compute amplitude gains, one for each partial loudspeaker signal of the first partial loudspeaker signals 28, so that the first virtual position is panned between the loudspeakers of set 26 - including the possible case that, occasionally, the first virtual position coincides with one of the loudspeaker positions in which case merely the loudspeaker at that position might receive a non-zero panning gain. In even other words, the first panning gain determiner 22 is for computing amplitude gains for a horizontal panning within set 26, so that this horizontal panning results into a virtual rendering position within the first layer set of the set 26 of loudspeakers.

Apparatus 10 of Fig. 1 further comprises a vertical panning gain determiner 30 which is configured to determine, depending on the intended virtual position 21, further panning gains for a panning between the first partial loudspeaker signals 28 on the one hand and one or more second partial loudspeaker signals 34 on the other hand. The one or more second partial loudspeaker signals 34 are to be applied to a second set 36 of one or more loudspeakers out of loudspeakers 14, which comprises merely one loudspeaker or more than one.

Fig. 1 illustrates the case where the number of second partial loudspeaker signals 34 and loudspeakers within set 36 is more than one, but it may also be true that there is merely one loudspeaker within set 36 and, accordingly, merely one second partial loudspeaker signal 34. In the latter case, the single loudspeaker of set 36 would be external to set 26 of loudspeakers for which the first partial loudspeaker signals 28 are dedicated. In case of set 36 comprising more than one loudspeaker, sets 26 and 36 may be mutually disjoint, partially overlap, coincide or completely overlap, i.e., one may be a proper subset of the other. Examples are set out in more detail below. In any case, the second position is vertically offset relative to the first position. Different examples of how to achieve the vertical offset between first and second positions even in case of the first and second sets 26 and 36 coinciding, are set out herein below. Note that in the embodiments outlined with respect to the figures, each set 26 and 36 is made out of loudspeakers of one layer or even corresponds to one layer, so that in case of coincidence of sets 26 and 36, the layers sets, i.e. the layers of sets 26 and 32, coincide as

well. However, this correspondence between sets and layers may be varied so that any of sets 26 and 32 may be composed of loudspeakers of more than one layer.

5 The further panning gains 32 determined by vertical panning gain determiner 30 finally result into a panning between the first virtual position and the second position.

As shown in Fig. 1, apparatus 10 further comprises a composer 40 which is further configured to compose the loudspeaker signals 12 from the input audio signal 18 using the first panning gains 24 and the further panning gains 32. As said, the first panning gains may be simple
10 amplitude gains and accordingly, composer 40 may comprise a multiplier 42 for each partial loudspeaker signal 28 for a multiplication of the input audio signal 18 with the corresponding panning gain 24. The panning gains 24 are, accordingly, individual for partial loudspeaker signals 28. That is, there is one panning gain 24 per partial input signal 28. Similarly, and as further outlined below, the panning gains 32 output by vertical panning gain determiner 30 may
15 be simple amplitude gains, too. Here, there is one panning gain 32 per set 28 and 34, respectively. Accordingly, composer 40 may comprise one multiplier 44a, 44b for each of sets 28 and 34, respectively, with multiplier 44a multiplying each loudspeaker signal of set 28 with the panning gain 32 associated with that set 28, and multiplier 44b multiplying each partial loudspeaker signal out of set 34 with the panning gain 32 associated with that set 34.

20 A further task of composer 40 is the following: as mentioned above, loudspeaker sets 26 and 36 may or may not overlap. As a task of composer 40, composer 40 correctly distributes the partial loudspeaker signals 28 and 34, obtained by panning using panning gains 24 and 32, onto loudspeakers 14. For those partial loudspeaker signals of sets 28 and 34, which merely
25 belong to one of sets 28 and 34, the corresponding partial loudspeaker signal becomes one of the loudspeaker signals 12. For those one or more partial loudspeaker signals, however, which are associated with the same loudspeaker out of loudspeakers 14, however, composer 40 adds them up using an adder 46 so that the sum of mutually corresponding partial loudspeaker signals out of set 28 and 34, respectively, become one of the loudspeaker signals 12.

30 It should be noted that, owing to the associative and commutative properties of the multiplication, composer 40 is not restricted to perform the multiplications for each partial loudspeaker signal in the order depicted in Fig. 1. That is, although composer 40 of Fig. 1 is depicted to perform the partial loudspeaker signal individual multiplication with the first panning
35 gains 24 prior to the multiplication with the set-global panning gain 32, the multiplications may be performed in a different order.

Fig. 1 also illustrates details which are used according to embodiments further described hereinbelow. In particular, these details relate to the derivation or generation of partial loudspeaker signals 34 from input audio signal 18. Two further processing steps may be associated with a derivation/generation of partial loudspeaker signals 34 from audio input
5 signal 18. These two processing steps and the corresponding elements in Fig. 1, are optional and, accordingly, the input audio signal may represent one partial loudspeaker signal 34 directly, which is subject to the vertical panning by means of the corresponding panning gain 32. If present, merely one or both processing steps may apply and be embodied within apparatus 10.

10

The first processing step corresponds to a horizontal panning with respect to the partial loudspeaker signals 34 in a manner substantially corresponding to the horizontal panning realized by elements 22, 24 and 42 with respect to partial loudspeaker signals 28. That is, as shown in Fig. 1, apparatus 10 may comprise a second panning gain determiner 52 configured
15 to determine, depending on the intended virtual position 21, second panning gains 54 for the second set 36 of loudspeakers, the second panning gains 54 defining the derivation of the second partial loudspeaker signals 34 from the at least one audio input signal 18. Composer 40 would comprise corresponding multipliers 56, namely one per partial loudspeaker signal 34, which multiplies the corresponding panning gain 54 with the audio input signal. In other
20 words, composer 40 would subject the partial loudspeaker signal 34 for each loudspeaker within set 36 to a multiplication with the panning gain 54 associated with the corresponding loudspeaker within set 36. This would result into a horizontal panning and to a virtual loudspeaker position associated with the partial loudspeaker signals 34.

25 Additionally or alternatively relative to elements 52-56, apparatus 10 may comprise a spectral shaper 58 which performs spectral shaping to the input audio signal or intermediary or final products as a result of the horizontal panning at multipliers 56 and vertical panning at multiplier 44b, so that the second partial loudspeaker signals 34 are derived from the at least one audio input signal by this spectral shaping. The spectral shaping is, for instance, for each of the
30 partial loudspeaker signals 34 equal, i.e., the same spectral shaping function may be used. As outlined in more detail below, the spectral shaping function 60 used by spectral shaper 58, is selected so as to form a psycho-acoustical cue for the listener that the second virtual position associated with the second partial loudspeaker signals 34 is positioned above or below the second set 36 of loudspeakers.

35

The spectral shaping performed by spectral shaper 58 may be performed in spectral domain by means of a multiplication of the partial loudspeaker signals' spectrum with the shaping

function 60, or may be done in time domain such as by means of a time domain filter such as an IIR or FIR filter, which time domain filter then would have the frequency response corresponding to spectral shaping function 60. Further notes will be made with respect to the sets 26 and 36. The apparatus may select same depending on a current speaker setup. In
5 other words, the apparatus may be adaptive to different setups. The apparatus may select the first set 26 of loudspeakers out of the plurality of loudspeakers depending on a horizontal component of the intended virtual position such as out of one layer those speakers nearest to the intended virtual position (as far as its vertical projection into the one layer is concerned) or depending on the horizontal component of the intended virtual position and a vertical
10 component of the intended virtual position such as by selecting an outmost layer nearest to the intended virtual position and then selecting the speakers within that one layer. Additionally or alternatively, the second set 36 of loudspeakers may be selected out of the plurality of loudspeakers depending on a vertical component of the intended virtual position such as by selecting an outmost layer nearest to the intended virtual position and using all the speakers
15 belonging to that layer for set 36, or depending on the horizontal component of the intended virtual position and the vertical component of the intended virtual position such as by selecting an outmost layer nearest to the intended virtual position and selecting the set 36 out of the speakers of the layer so that same are nearest to the intended virtual position (as far as its vertical projection into the one layer is concerned).

20

As mentioned before with respect to the first partial loudspeaker signals 28, composer 40 may be configured to perform the multiplication 56 and 44b as well as the spectral shaping 58 in any order, i.e., may apply the three tasks in any order onto the audio input signal 18 in order to result into the corresponding partial loudspeaker signals 34.

25

Lastly, it should be noted that according to an example, it may be that the number of loudspeakers within set 36 and, thus, a number of partial loudspeaker signals 34, respectively, may be one, even in case of using the spectral shaper 58.

30

Before proceeding with the description of certain details and embodiments of the present application, which are described in the following by reusing the reference signs and the description brought forward above, the following note shall be made with respect to the composer 40: in case of Fig. 1, panning gain determiners 22, 30 and 52 form kind of intermediary modules for computing the panning gains on the basis of the intended virtual
35 position 21 while the actual application of the panning gains had been performed by composer 40. Additionally, spectral shaper 58 was shown to be included within composer 40 as a submodule thereof. However, as said above, modifications compared to the illustration of Fig.

1 are feasible. For instance, the spectral shaper 58 could be placed upstream elements 52, 54 and 56 so as to become, finally, a module external to, and especially upstream to, composer 40. Composer 40 would then, as far as the first loudspeaker set 36 is concerned, perform the composition of the loudspeaker signals 12 on the basis of a pre-shaped version of the audio
5 input signal 18. Additionally or alternatively, most of the subsequently explained embodiments make use of a composition, where the vertical panning is applied after the horizontal panning which, in turn, is realized by means of multipliers 42 and/or 56 and, if applicable, the spectral shaping 58, and in that case, composer 40 and its composition may involve elements 44a, 44b and, if applicable, adder 46, only, whereas elements 22, 24 and 42 form a first loudspeaker
10 signal set determiner 70 and elements 52, 54, 56, 58 and 60 (or parts thereof if the horizontal panning or the spectral shaping is missing) form a second loudspeaker signal determiner 72.

Before resuming the description with the announced further details and further detailed embodiments, a brief note shall be made with respect the achieved advantages resulting from
15 the concept of audio rendering as depicted in Fig. 1. In particular, as outlined above, the audio rendering of the concept of Fig. 1 allows the audio reproduction to get along without the usage and the associated computationally complex tasks of applying different HRTFs that are precisely adapted or selected based on or according to an exact angular variation of the intended virtual position 21. All horizontal and vertical panning is done by amplitude panning
20 only, and the spectral shaping 58 may use one spectral shaping or an equal spectral shaping function 60 for all partial loudspeaker signals 34 for all loudspeakers within set 36. In the embodiments described further below, apparatus 10 may either use continuously the same spectral shaping function 60 irrespective of the intended virtual position 21 (such as in case of the intended virtual position 21 being restricted to positions which are, in height, within,
25 between , or above, the listener position or the layers of the loudspeakers 14, or vice versa, in case of being restricted to positions which are, in height, within, between , or below, the listener position or the layers of the loudspeakers 14) or to discriminate between two spectral shaping functions 60, one being used in case of the intended virtual position 21 being higher than the listener's position or the highest loudspeaker layer, respectively, and the other in case of being
30 lower than the listener's position or the lowest loudspeaker layer, respectively. Thus, the computational complexity of the rendering of Fig. 1 is low. This is also true when making use of the optional spectral shaping 58.

Moreover, although the decomposition of the 3D panning into horizontal panning on the one
35 hand and vertical panning on the other hand might appear to result in a more complex rendering procedure, the resulting computational complexity is still low, while the rendering

accuracy in terms of positioning the intended virtual position is still high even at this computational moderate complexity.

That is, embodiments described herein provide an alternative to the rather complex setups set-out in the introductory portion of the specification and form a compact reproduction that uses signal processing means to generate a comparable or similar spatial auditory perception as more complex loudspeaker setups. The concepts presented above and in the following are capable of

10 (1) perceptually replacing missing loudspeakers / loudspeaker arrays by consideration of one or more virtual loudspeakers. The generation of those virtual loudspeakers is described herein.

(2) efficiently rendering sound in 3D loudspeaker setups, wherein the rendering can be used if the virtual loudspeaker (1) is used, as well as in scenarios where the necessary loudspeakers are available physically. The benefit of (2) is the flexibility and efficiency, which makes it also applicable in scenarios where the listener position is tracked in real time, and the rendering is adapted in real time to the listener's current position.

20 Note that the embodiments described herein are independent of the reproduction environment and could, e.g., also be used e.g. in an automotive environment. Furthermore, the embodiments are independent of the specific type of transducer or topology used for reproduction. That is, the embodiments could be applied e.g. in headphone reproduction, as well as in reproduction using specific loudspeakers such as loudspeaker arrays, soundbars, smart speakers, etc.

That is, the just-made notes render clear that the loudspeakers 14 may be headphone loudspeakers or stereo loudspeakers, but may, as well, form a loudspeaker array, a soundbar, or a set of loudspeakers, smart speakers, or a set of smart speakers, from a surround sound setup or may be individual loudspeakers, wherein combinations may be feasible as well. Moreover, the description made clear that apparatus 10 operates adaptive in order to adapt, in real-time, the composition of the loudspeaker signals 12 to the intended virtual position 21 which may vary in time.

35 In this regard, it shall briefly be noted that, while embodiments of the rendering apparatuses may be pre-configured for certain loudspeaker setups, i.e. that they expect a predefined set of loudspeakers 14 to be positioned at predefined positions, it might also be that the apparatuses

described herein are adaptive to different loudspeaker setups, differing in number of loudspeakers and/or speaker positions, in terms of an initialization of the apparatus and/or in terms of an adaptation to moving loudspeaker positions. In the former case, the apparatus may, after initialization, assume the loudspeaker setup to be constant. The latter case, the apparatus may even adapt to speaker setup variations during runtime. Even the number of speakers could vary in runtime. Accordingly, the apparatus may receive information on the loudspeaker positions with this optional circumstance, however, not being explicitly shown in the figures. Thus, similar to the optional reception of the listener position information, apparatus of Fig. 1 (and subsequently shown embodiments) may comprise a further position input for receiving the loudspeaker setup information revealing number of speakers and positions thereof. This information may be provided relative to the position and/or head orientation of the listener and/or relative to real-world coordinates. This information could e.g. be based on Cartesian coordinate systems, or polar coordinate systems. It could e.g. be based on a room centric coordinate system or a listener centric coordinate system, either as a Cartesian, or polar coordinate system.

Commonly used methods for rendering are amplitude panning techniques. To generate the perception of an auditory object at positions that are not covered by loudspeakers (e.g. not between two or more loudspeakers), rendering techniques such as crosstalk cancelation can be utilized. Crosstalk cancellation (XTC) [1-7] has the goal to control the left and right ear signals of a listener by means of loudspeakers. This is achieved by "cancelling the crosstalk between the ears" which occurs when a loudspeaker's signal reaches a listener. Once the ear signals can directly be controlled, binaural techniques [8, 9] can be applied to render sound at top and bottom directions. There are two major limitations of the before mentioned technique. Firstly, XTC has limitations related to sound coloration, extremely small sweet spot, and high dependence on loudspeaker positions relative to the listener. Secondly, without head tracking / listener tracking and/or individualized head related transfer functions (HRTFs) or binaural room impulse responses (BRIRs), binaural techniques are limited in the achievable quality/performance. Both of these would add high complexity, cost, and user inconvenience to the system.

Enhancements to conventional amplitude panning have been proposed, using virtual loudspeakers in dimensions not covered by the loudspeaker setup, see e.g. [14, 15]. Height panning using such techniques is not entirely realistic as timbre deviates from sources truly rendered at height.

Vertical Hemispherical Amplitude Panning (VHAP) [10, 11] uses two lateral loudspeakers to render objects with height and on top of a listener. As the loudspeakers have to be at ± 90 degrees lateral directions, VHAP is inflexible in terms of listener position.

- 5 In this specification, the term virtual loudspeaker is used for a non-existent loudspeaker which is considered during the process of panning an object.

The concept of Fig. 1 makes use of concepts for top and/or bottom rendering with the following advantages over the state-of-the-art techniques just mentioned:

- 10
- Equalization (spectral shaping 58) is applied to the top/bottom virtual loudspeaker signals for a more faithful top/bottom/height perception
 - Any loudspeaker setup can be used for speakers 14, and nevertheless an enhancement for (virtual) top and bottom rendering is achievable. For example, a stereo setup or a 5.1 setup may be used as a basis for speakers 14. Even loudspeaker

15 setups with height loudspeakers, e.g. 5.1+4H, can be enhanced using the concept of Fig. 1, such as with respect to Top rendering (e.g. "voice of god" loudspeaker), or lower layer rendering. In contrast to this, VHAP requires, for instance, a precise and specific loudspeaker setup with loudspeakers at each side of a listener (± 90 degrees). - Moreover, the top and bottom rendering of Fig. 1 does not rely on specific loudspeaker

20 positions relative to the listener. In other words, the scheme of Fig. 1 can be applied also in a scenario where a listener moves, e.g. tracked rendering.

The embodiments described herein allow for very straight forward implementations of virtual height rendering.

25

That is, object panning according to Fig. 1, may be implemented in a manner leading to a rendering apparatus or object panning processor according to Fig. 2, which generates the loudspeaker signals 12 at the output of composer 40 with two paths which provide partial loudspeaker signals 34 on the one hand and partial loudspeaker signals 28 on the other hand

30 to composer 40, namely one path comprising partial loudspeaker set determiner 70 which receives audio input signal 18 and intended virtual position 21 and outputs the partial loudspeaker signals 28, and another path comprising module 72 which generates partial loudspeaker signals 34 on the basis of the two inputs 18 and 21, and which apparatus and so forth renders an object in 3D space over ANY loudspeaker setup, by

- 35
- considering at least one virtual loudspeaker (Top or Bottom) at a vertical (top or bottom) direction. This is done or achieved by the spectral shaping 58 which, as outlined in more detail below, leads to a psycho-acoustical cue for the

listener that the sound reproduced by the first partial loudspeaker signals 34 arrives from top or bottom, respectively.

- amplitude panning the object, considering the loudspeaker setup plus one or more virtual loudspeakers. The amplitude panning is performed by the vertical panning within composer 40, and the horizontal panning within module 70 and within module 72.
- applying equalization to virtual and/or real loudspeaker signals. The equalization is done by this spectral shaping within spectral shaper 58.
- reproducing each virtual loudspeaker signal over a subset or all loudspeakers of the setup as explained with respect to Fig. 1, the second loudspeaker set 36 may coincide with set 26 and, thus, involve all loudspeakers 14, or may relate only to a subset of loudspeakers 14.

In the following, the concept of embodiments of the present application is visualized three-dimensionally. See Fig. 3. In Fig. 3, the listener is indicated by reference sign 100. The individual loudspeakers 14 are distinguished from one another by small letters. In Fig. 3, the loudspeaker setup comprises, exemplary, four loudspeakers. Fig. 3 shows one virtual loudspeaker 102 on top of, or above, listener 100. Fig. 3 is, naturally, just an example. A virtual loudspeaker 102 in the bottom or below listener 100 may be considered, alternatively. Moreover, the virtual loudspeaker 102 may be positioned right above listener 100 even with allowing the listener 100 to move horizontally, namely by means of tracking the listener position, or listener's 100 position may be fixed by default irrespective of the listener 100 being, actually, right below/above the virtual loudspeaker 102.

Stated differently, Fig. 3 shows an example for a positioning of loudspeakers 14, here exemplary four loudspeakers 14a to 14d, and explain that the embodiments shown in Figs. 1 and 2, may involve a virtual loudspeaker positioned at a virtual position which is the aforementioned virtual position of rendering associated with the first partial loudspeaker signals 34. That is, Fig. 3 illustrates that the embodiment of Fig. 2 as well as the embodiment of Fig. 1, as far as making use of spectral shaper 58, additionally considers a virtual loudspeaker 102 in addition to the available loudspeakers 14.

Figs. 4, 5a and 5b show, decomposed into individual sub-concepts or steps, as to how the rendering at an intended virtual position 104 using the available loudspeakers 14a to 14d and the virtual loudspeaker 102 is done.

Fig. 4 illustrated the intended virtual position 104. This position 104 is indicated to be vertically above the layer or plane within which the loudspeakers 14a to 14d are. Fig. 4 also shows the projection of the intended virtual position 104 into the layer or plane of the loudspeakers 14a to 14d, i.e., the projection 104 along vertical direction into the layer or plane of loudspeakers 14a to 14d. The resulting projected position 106, i.e., the projection of the intended virtual position 104, into the layer of loudspeakers 14a to 14d, is indicated using reference sign 106. Module 70 may use amplitude panning so as to result in partial loudspeaker signals which are associated with a rendering of the audio object at this projected virtual position 106. Thus, Fig. 4 illustrates another circumstance not yet having been described with respect to Figs. 1 and 2 so far. In particular, the apparatus of Figs. 1 and 2, respectively, may be configured to select 26 out of all available loudspeakers 14 or out of a group of loudspeakers such as the group of loudspeakers belonging to a certain layer such as loudspeakers 14a to 14d here in Fig. 4. In particular, as illustrated by use of hatching, only two loudspeakers 14c and 14d may be selected, namely those of the group of loudspeakers belonging to the horizontal plane of listener 100 are selected to receive corresponding partial loudspeaker signals 28, which are nearest to the projected virtual position 106. According to a different view, the horizontal panning, while resulting in non-zero weights only with respect to a subset of the corresponding loudspeaker layer set, continuously relates to all loudspeakers of the corresponding layer set. Here, only loudspeakers 14c and 14d would be associated with non-zero weights for horizontal panning, while the other two speakers 14a and 14b would be associated with zero weights, thereby not participating in the horizontal panning. The two loudspeakers 14c and 14d of the loudspeaker setup are, thus, used, in addition to the virtual loudspeaker 102. Fig. 4 concentrated on the horizontal panning achieved by module 70 or by determiner 22, respectively, whereas the following figures concentrate on module 72 and its contribution to the final rendering. That is, the following figures will reveal as to how the two loudspeakers 14c and 14d of the loudspeaker setup along with a virtual top loudspeaker 102 are used for amplitude panning the object at the intended virtual position 104.

Note that the distance of the intended virtual position 104 does not play a major role in the context of this application and that, accordingly, position 104 is depicted as being far away from the listener for sake of an easier perspective representation only. The rendition may, optionally, operate dependent on the direction towards position 104 only.

Fig. 5a shows the sub-concept or step according to which equalization or spectral shaping 58 is used for, or applied to, the loudspeaker signal(s) for the virtual loudspeaker 102. Again, Fig. 3 to 5b concentrate on an example where this virtual loudspeaker 102 is a virtual top

loudspeaker, but this is only an example. The equalization or spectral shaping 58 may likewise be used in order to form a virtual bottom loudspeaker.

Fig. 5b concentrates on the reproduction of the audio object at the position of the virtual
5 loudspeaker 102. A loudspeaker signal which would be applied to the virtual loudspeaker 102
directly, namely the audio input signal, is subject to the equalizing or spectral shaping 58 and
to the horizontal panning here illustrated by the corresponding multipliers 56a to 56d. The latter
multipliers are optional. They are only necessary if the virtual loudspeaker position 102 is not
10 static, but positioned so as to be vertically adjusted to the listener position of listener 100, i.e.,
to be horizontally located such that its vertical projection into the plane of loudspeakers 14a to
14d coincides with the position of the listener 100 within this plane or layer of loudspeakers
14a to 14d. Fig. 5b exemplary illustrates that the set 36 may encompass all loudspeakers 14a
to 14d or at least all loudspeakers of the corresponding group within one horizontal layer. That
15 is, 5b illustrates the reproduction of each second partial loudspeaker signal 34 over a subset
or, as illustrated in Fig. 5b, all loudspeakers 14a to 14d of the setup. Since the virtual
loudspeaker(s) 102 is not physically available, corresponding equalized signals 34 are
reproduced over the mentioned subset of loudspeakers. The gains are applied in total or for
each loudspeaker individually to adjust level and resulting direction vector for virtual direction.
An alternative implementation that is beneficial due to its reduced computational costs has
20 already been mentioned above and is depicted in Fig. 6. That is, Fig. 6 shows another example
for an apparatus for rendering or an alternative embodiment for an object panning processor,
namely one where, compared to Fig. 2, the equalization or spectral shaping 58 is performed
upstream the horizontal panning by elements 52, 54 and 56 within a module 72. That is, the
equalization or spectral shaping so as to result in psycho acoustical cues for the listener, to
25 result in top or bottom loudspeakers 102, is applied to the audio input signal 18 directly rather
than onto each partial loudspeaker signal 34 individually. That is, the audio input signal 18 is
subject to the equalization or spectral shaping, where upon the panning may be applied such
as, optionally, the horizontal panning to control the position of virtual position 102 horizontally,
and the vertical panning achieved using the vertical panning factors or gains provided by the
30 vertical panning gain determiner. An even lower computational complexity is achieved if the
vertical panning gain for partial loudspeaker signals 34 is applied prior to the optional horizontal
panning in between loudspeaker set 36. In the latter case, the equalized or frequency shaped
and level-aligned signal may be copied and distributed onto the loudspeakers that have been
selected for reproduction of the virtual height loudspeaker 102.

35

According to the concepts set forth above, the efficient generation of a virtual height reproduction is part of a panning algorithm that allows for using the corresponding virtual height speaker in arbitrary loudspeaker setups. Further details are described in the following.

- 5 An (object) panning algorithm / panning processor or an apparatus according to any of Figs. 1, 2 and 6, can be used for positioning the perceived location of auditory objects within a 3D reproduction space both for static, as well as for moving sound sources.

Due to the efficiency of the underlying concept, it can also be used for static as well as moving
10 listener positions, i.e. also for applications, for instance, in which the position of the listener 100 is tracked, and the rendering by the apparatus is adapted to the listener position. Adaptation examples are set-out below. Furthermore, an apparatus as described herein could even be applied to scenarios with static as well as moving loudspeakers 14.

15 In typical reproduction scenarios, the loudspeaker positions are fixed, but the listener's 100 position may continuously change. In such a case, the angles under which the listener 100 sees the loudspeakers 14, as well as the respective angles between loudspeakers change as a function of the listener's 100 position.

20 Conventional panning algorithms, such as VBAP, typically need initialization for their considered invariant sweet spot and loudspeaker positions. During initialization phase, some complex operations are used, such as mapping loudspeakers to pair, triplet, or quadruplet panning groups.

25 Since in a tracking scenario, relative positioning of loudspeakers 14 and listener 100 frequently changes, it is undesirable to have a complex initialization phase and fixed mapping. The described panning according to Figs. 1, 2 and 6 addresses these issues and includes a few other novelties related to panning, especially at positions that do not lie inside an area that is covered / surrounded by loudspeakers.

30

In particular, the following steps assist in achieving an efficient rendering and to deal with speaker setups with more than one layer of speakers 14a-d as exemplarily shown in Figs. 3-5b and may be added as functionalities two the apparatuses described herein:

- 35
- Amplitude panning gains are computed for a horizontal loudspeaker layer, such as in any of the horizontal panning stages in 70 and 72. It might be, that the apparatus is responsive to whether the number of layers of speakers is one or not. If only one layer

exists, elements 52,54,56 are not used or are only for positioning the top/bottom virtual speaker position 102 right above/below listener 100. If more than one layer exists, the following is true.

- If more than one layer of speakers 14 is present, then
 - 5 ○ amplitude panning gains for more than one loudspeaker layer may be computed such as for a height layer and a bottom layer using module 70 and 72, respectively. This may be done, for instance, if the intended virtual position points to a position vertically inbetween both layers. Note that even more than two layers may be treated that way.
 - 10 ○ In the panning, any rendered horizontal / azimuthal virtual position of the object, such as 106 in Fig. 4, namely in each layer for which horizontal panning is performed, is considered in the rendering, namely in the vertical panning. Two layers, i.e. two groups of speakers 14, each of which is associated with another horizontal layer at different heights, may, for instance, be selected, one forming
 - 15 set 26, or being used for selecting set 26 thereout, the other forming set 36, or being used for selecting set 36 thereout. The selection out of several (more than two) available layers may be done as described below, namely by taking the layers nearest to the intended virtual positions. The “rendered object position” such as 106 in Fig. 4 for the one exemplary layer shown therein, on each one
 - 20 of the layers may then be used as a virtual loudspeaker for vertically panning the object between the layers. Details are illustrated below.
 - If the object position is above the highest layer or below the lowest layer, then the object is horizontally panned only on one layer (i.e. on the highest, or on the
 - 25 lowest layer, respectively). In that case, module 72 operates for the virtual top/bottom speaker 102 and the horizontal panning is for adjusting the horizontal position of the top/bottom speaker 102 to the listener position 100 only, if this option is used at all (alternatives are described below according to which this listener position adaptivity is not used), and module 70 operates for the horizontal panning in the used vertically outermost speaker layer or outmost
 - 30 group of speakers 14 forming a horizontal layer. Both modules 70 and 72 would have their sets 26 and 36 of speakers 14 be selected to correspond to, or be part of the mentioned vertically outermost speaker layer or outmost group of speakers 14.
 - Thus, if the object position 104, 21 lies above (below) the highest (lowest) loudspeaker
 - 35 layer (or in the case that only one loudspeaker layer (e.g. at roughly ear height) is available), then a virtual vertical top (vertical bottom) loudspeaker 102 is considered to perceptually render the auditory object above (below) the loudspeaker layer(s)

- A top or bottom equalizer, i.e. a spectral shaping 58 using a corresponding function 60, is applied to the object audio signal and distributed to the loudspeakers that have been selected for top or bottom direction reproduction, i.e. set 36.

5 The steps/functions/blocks participating in the rendering between two layers, or speakers of two layers, is depicted in Fig. 7. To be more precise, Fig. 7 either illustrates an apparatus according to an additional embodiment capable of three-dimensionally panning an audio object to be rendered between two layers of speakers, or Fig. 7 illustrates the cooperation of those portions of the apparatus of Fig. 1, which participate in the rendering in case of the intended virtual position 21 being between two such speaker layers, while the other element shown in Fig. 1 such as the spectral shaper/equalizer 58 do not participate in the rendering in this case (but rather in case of the intended virtual position lying above all speaker layers of speakers 14 or below those available speaker layers). As shown, the input is the audio input signal 18. Horizontal panning is performed by module 70 with respect to one layer and elements 52, 54 and 56 is part of module 72 for the other layer. The corresponding partial loudspeaker signals 28 and 34, respectively, are composed to result into loudspeaker signals 12 by composer 40, with additionally performing the vertical panning using the panning gains provided by determiner 30. The speaker sets 36 and 26, for which the partial loudspeaker signals 34 and 28, respectively, are, may be mutually disjoint as illustrated in Fig. 7 as they belong to different layers. However, it should be noted that the association of speakers 14 to "layers" may be such that one speaker 14 may be associated with different layers. In other words, the grouping of speakers 14 into layer groups of speakers may be such that they overlap. Insofar, the illustration of Fig. 7 is merely an example and may be modified.

25 The cooperation of the individual elements of Fig. 7 is described in more detail below. As shown and as explained above, the panning, both horizontal and vertical pannings, are controlled by way of the positional information 21. It can either be delivered as additional information such as in form of additional information in a separate data stream, namely separate relative to the audio input signal 18, e.g., as an audio object including at least one channel of audio information and associated metadata defining the intended position. If the audio input signal 18 is a multichannel file without metadata, the intended position 21 of different elements included in the audio signal can be estimated and extracted based on a signal analysis given the known target loudspeaker layout the signal has been produced for. For instance, the audio input signal 18 may comprise a channel associated with a loudspeaker position at the top and/or at the bottom, but the speakers 14 available do not have such speakers. In that case, the intended virtual position 21 is the position of that channel's speaker's position. Other examples are, naturally, available as well. This may be done for all channels

conveyed. The mutual speaker positions to which the channels relate may be maintained by the rendering apparatus.

In accordance with an embodiment, both horizontal pannings, namely the one or more module
5 70 with respect to partial loudspeaker signals 28 and the one regarding the other partial
loudspeaker signals 34 by way of elements 52 to 56 use the same azimuth angle for panning.
That is, the same azimuth angle is used for both layers. In other words, the horizontal panning
is done in a manner so that the projected virtual positions 106 depicted in Fig. 4 coincide in a
vertical projection onto one another. Naturally, this may be implemented differently. The
10 restriction is not necessary and different azimuth angles may be used for different layers.

A beneficial feature of the embodiments discussed herein is the fact that they do not require
extensive initialization. Instead, panning parameters are computed directly from given or
changing listener and loudspeaker coordinates or positions. The initialization of the rendering
15 is not dependent on predefined pairs, triplets, or quadruplets of loudspeakers.

Fig. 8 illustrates the fact that both, horizontal and vertical panning, may be controlled by
information on the listener position, namely information 110. To be more precise, imagine the
intended virtual position 21 is represented by solid angles indicating a certain direction from
20 which the listener 100 shall perceive the audio object to be rendered. Depending on the listener
position 110, aside from any adaptation of the virtual top/bottom speaker's position to the listener
position, if any, a horizontal panning, which is dependent on the listener position, might be
applied in order to attain this perception direction for the listener. Same is true in case of the
listener position information 110 being indicative of the position of listener 100 not only in terms
25 of horizontal position but also in terms of height such as the height of the position of the
listener's ears.

As is clear from the above description, apparatuses according to embodiments of the present
application are not restricted to deal with loudspeaker setups where the available loudspeakers
30 14 are arranged in one layer only. The latter example had been depicted in Figs. 3 to 5b.
Rather, loudspeakers 14 being available for the apparatus, may be associated with different
layers. The partial loudspeaker signals 34 on the one hand and partial loudspeaker signals 28
on the other hand which have been discussed above, or, differently speaking, the two paths
into which module 70 and 72, respectively, are serially connected, may be associated with one
35 or more of such speaker layers. For the following description, we assume that each of same is
associated with one speaker layer. That is, each is associated with one group of loudspeakers
forming one layer. Some loudspeakers may be associated with more than one layer as will

become clear from the following description and has already been stated above. The attribution or association of layers to the individual paths, namely path of module 70 and path of module 72, may be fixed or may be subject to adaptation to the intended virtual position 21 and/or the listener position 110. This has already been discussed above: If there are more than two layers
5 available, two layers may be selected in case of the intended virtual position being in between a pair of these layers and these layers are associated with the two paths. In case of the intended virtual position 21 exceeding all layers available, and there is no real top or bottom speaker available, then the outermost layer nearest to the intended virtual position is selected as the loudspeaker layer for which both paths are used.

10

Given an arbitrary loudspeaker setup, initialization may involve only that each loudspeaker 14 is classified as belonging to one or more of the following categories:

Layer 1:

15 Typically this loudspeaker layer is used for panning objects horizontally (approx. on ear height of a seated listener).

Layer 2 to N:

20 Optionally, loudspeakers in a second layer can be defined, such as loudspeakers in a height (top or bottom) layer. These are layers vertically above or below Layer 1. The loudspeaker layers can, thus, be more than two. The distinction between Layer 1, being on ear height, and any other layer or the other layers is optional.

Top:

25 Loudspeaker(s) over which vertical top direction is reproduced. This can be a dedicated loudspeaker, or a subset of loudspeakers of other layers.

Bottom:

30 Loudspeaker(s) over which vertical bottom direction is reproduced. This can be a dedicated loudspeaker, or a subset of other layers.

The above description is not limited to regular setups, where regular would e.g. imply that an equal number of loudspeakers is present in every layer, having equal angles/distances between them, or that all layers completely surround the listener, or that all layers have
35 loudspeakers arranged at exactly the same vertical angle as seen from the listener.

Actually, as mentioned before, any arbitrary setup can be used. The different loudspeakers could be positioned at different / arbitrary azimuth angles, and at different / arbitrary elevation

angles (i.e. different heights). Loudspeakers considered to be part of one layer do not necessarily need to lie within a plane. Variations in their vertical positioning is allowed.

Fig. 9 and 10 show example realizations / example classifications. These figures shall exemplify the procedure of allocating the different available loudspeakers to the different layers. Those are only examples, different mappings in the same situation(s) would be possible and are subject to the user's preferences.

Fig. 9 shows a classification using a 5.0 loudspeaker setup. Here as well as in following figures, the following identifiers are used for simplicity to indicate available speakers 14: The horizontally arranged loudspeakers, that would usually form the setup that is installed at roughly ear height of a listener is labeled in the form "M_X", where M is an indicator for MIDDLE, hinting that this layer is usually between the upper and lower loudspeaker layers. This would, thus, be a Layer 1 in the above nomenclature. The X identifies the specific loudspeaker in this layer, e.g. M_L would be the "front left loudspeaker in the middle layer". Similarly, we identify an upper layer loudspeaker as "U_X", so "U_Rs" would be the "right surround loudspeaker in the upper layer". Loudspeakers in a lower layer would be identified by "L_X". U and L speakers are, thus, speakers of Layers 2...N in the above nomenclature. A loudspeaker mounted at the ceiling (i.e. either directly above the listener, or directly above the center of the loudspeaker array) is denoted Top. Respectively, the term Bottom is used for loudspeakers directly below the listener, or directly below the center of the loudspeaker array. In Fig. 9, the classification of speakers would be:

25

<u>Loudspeakers</u>	<u>Categories</u>
M_L , M_R	Layer 1, Top, Bottom
C	Layer 1
M_Ls , M_Rs	Layer 1, Top, Bottom

30

Horizontal panning by module 70 would be done using all available loudspeakers (Layer 1). Top and Bottom directions are rendered using module 72 over all loudspeakers except the center (C). That is, set 36 would comprise all loudspeakers except the center, while set 28 would encompass all speakers.

35

Please note that this is an explicit decision for this example. Of course, the center loudspeaker could also be used for height rendering.

A further classification using a 5.0+2H loudspeaker setup is depicted in Fig. 10. Here, two layers exist in the available set-up and the classification or association would be:

5

<u>Loudspeakers</u>	<u>Categories</u>
M_L , M_R	Layer 1, Bottom
C	Layer 1
M_Ls , M_Rs	Layer 1, Layer 2, Top, Bottom
U_L , U_R	Layer 2, Top

10

In this example, the middle layer surround loudspeakers (M_Ls and M_Rs) are used for both layers (Layer 1 and Layer2), since otherwise Layer 2 would not surround the listener. That is, Layer 1 and Layer 2 speakers would be used for inter-layer panning as illustrated in Fig. 7 and 8, e.g. those of Layer 1 for set 26 and those of Layer 2 for set 36 or vice versa, and as soon as the intended virtual position is outside both layers, to the top or bottom thereof, then speakers belonging to the class Top are used for set 36 with active equalization 58 and with using Layer 2 speakers for set 26, or the class Bottom speakers are used for set 36 with active equalization 58 and with using Layer 1 speakers for set 26.

20

Alternative classifications in this setup could be to decide for rendering without a Layer 2. The Top could be rendered using only the elevated loudspeakers U_L and U_R, or alternatively, the top could also be rendered by a combination of the U_L, U_R, M_Ls, and M_Rs as described before.

25

Further examples are readily derivable. E.g. with bottom layer loudspeakers, or with more or less elevated loudspeakers, or with more or less loudspeakers in the middle layer, or with more arbitrary or irregular loudspeaker setups.

30

In the following, the case of rendering an object in 3D is explained for an example case where the object is panned in a direction (as seen from the listener) that lies between two physically present loudspeakers layers (which are at different height). This had already been discussed above with respect to Fig. 7 and 8, but it is illustrated more clearly in Fig. 11 and 12. A 5.0+4H loudspeaker setup is exemplarily illustrated here. Examples for a position of the listener 100 and the position of the audio object 104 are indicated. The speakers are classified into two separate layers discriminated using different line types, dashed for second layer and continuous for first layer.

35

The object is amplitude panned in the first layer by giving the object signal to loudspeakers in this layer with different gains 24, e.g. by giving the object signal to M_L and M_Ls such that it is amplitude panned to bottom layer gray dot position 106₁ in Fig. 11. Similarly, the object is amplitude panned in the second layer to the height layer gray dot position 106₂ in Fig. 11. As
5 can be seen, positions 106₁ and 106₂ may be selected so that they vertically overlay each other and/or so that the vertical projection of intended position 104 and the positions 106₁ and 106₂ coincide as well.

Fig. 12 illustrates rendering the final object direction by applying amplitude panning between
10 the layers, i.e. illustrates the vertical panning. Considering the virtual objects at positions 106₁ and 106₂ as virtual loudspeakers, amplitude panning by elements 30 and 40 is applied to render the virtual object at intended position 104, between the two layers appearing in the direction of the object. The result of this amplitude panning between the layers are two gain
15 factors 32 with which the two layers' signals 34 and 28 are weighted.

This weighting for the horizontal panning between (real) loudspeaker layers can additionally be frequency dependent to compensate for the effect that in vertical panning different frequency ranges may be perceived at different elevation [13].

20 Rendering Objects above or below a layer or outmost layer is further inspected now, as an additional information relative to the description set forth above.

An object may have a direction or position 104 which is not within the range of directions between two layers as discussed wrt Fig. 11 and 12. This case is discussed wrt Fig. 13 and
25 14. An object's intended position 104 is above or below a (physically present) layer, here above any available layer and, in particular, above the upper one indicated in dashed lines. As an example, the object has a direction/position 104 above the top loudspeaker layer of the 5.0+4H setup which has been used as an example set-up in Fig. 11 and 12 as well.

30 In this case, horizontal amplitude panning is applied by module 70 to the height layer to render the object in that layer. The resulting position 106₁ of the rendered object is indicated as height layer gray dot position 106₁ in Fig. 13.

Then, panning is applied between position 106₁ in the height layer and the vertical
35 direction/position 106₂, indicated as gray dot position 106₂ in Fig. 14. The resulting 3D panned virtual object is indicated as gray dot position 104'.

Since there is no real loudspeaker at the vertical top or bottom direction, the vertical signal at 106₂ is equalized by module 58 to mimic coloration of top or bottom sound respectively (see

subsequent explanation for more details on the equalization). The vertical signal is then given to the loudspeakers designated for top/bottom direction, i.e. set 36.

As to the rendering of the virtual Top or Bottom loudspeakers 102 the following may be said.

5

In general, different approaches can be chosen to render the virtual vertical Top or Bottom loudspeakers.

In general, two different approaches can be chosen:

- 10 (1) Virtual top/bottom always rendered above the actual listening position as indicated by 110.
- (2) Virtual top/bottom speaker is always rendered above a "sweet spot" or a center of the (main) loudspeaker array

15 As application examples, (1) could be beneficially chosen, if the listener position can be tracked, while (2) could be chosen if the possibility for listener tracking is not available.

20 A simple implementation uses the same gain for each loudspeaker selected for Top or Bottom rendering, i.e. the gains 54 would be chosen to be equal. This scheme works well. (It can e.g. be used as the simplest implementation and is especially useful, when the listener position is not tracked and such not known.)

Especially when the listener is not centrally located within the loudspeaker setup, then the following considerations can improve top and bottom rendering:

- 25 • If there is a height layer and one wants to pan above that height layer, gain factors 54 applied to the (height-layer) loudspeakers 36 may be used for the top direction, such that the resulting panning direction vector points vertically upwards (or alternatively towards a virtual top loudspeaker position 102), i.e. so that 102 is right above the listener 100.
- Same for bottom direction, when there is a bottom loudspeaker layer.
- 30 • If there is no height layer and one wants to pan above the horizontal layer, gains are applied to the loudspeakers such that the amplitude panning vector vanishes (no horizontal direction bias). Simpler, one can apply gains 54 to the loudspeakers such that signal amplitude or power at the listener is the same for each top/bottom rendering loudspeaker.
- Same for bottom direction, when there is no bottom loudspeaker layer.

35

In the following, the equalizer (or spectral shaper) 58 is further exemplified using further details. The main cues enabling the listener 100 to localize a sound source in the horizontal

plane are differences between the left and right ear input signals (interaural time differences (ITDs) and interaural level differences (ILDs)). The primary cues for estimating the vertical position of a sound source are spectral variations due to reflections produced by the listener's head, torso, and pinnae. Such cues are often called monaural cues (MCs), called psycho-acoustical cue in the above description.

The specific ILDs, ITDs, and MCs, which occur due to the unique body features of each individual and the considered direction of incidence, are commonly sub-summed under the term Head Related Transfer Functions (HRTFs). Especially the MCs are highly individual. Still, there are some common features that influence the height perception in general.

By shaping the frequency content of a specific source signal that is received from one direction, the illusion that this sound actually comes from a different elevation and/or front-back-orientation on the same cone of confusion can be supported. This corresponds to changing MCs and is the purpose of the equalizer (EQ).

A simple but well working implementation of the concept of using virtual top/bottom loudspeakers, and equalization of these signals, uses a specific static EQ for the top and bottom direction respectively.

Fig. 15 shows two such heuristically determined equalizers as examples or, differently speaking, shows a shaping function 60a for virtual top speaker rendering and a shaping function 60b for virtual bottom speaker rendering. These have been determined by analysis of measured HRTF data, corresponding to cues implying a source above or below a listener. HRTFs of many subjects were considered and the EQs were determined by ignoring spectral changes which vary too much between subjects.

The equalizer 60a for top direction typically has one or more notches and/or peaks. Typically there is a notch below 1kHz and one or more peaks at higher frequencies. An equalizer 60b for bottom direction includes the effect of "body shadowing", that is, overall high frequencies are attenuated. In other words, by function 60a, the second partial loudspeaker signals 34 are, relative to the audio input signal 18, dampened in a notch spectral range between 200 and 1000 Hz and amplified within one or more in peak spectral ranges 122_1 and 122_2 – here there are exemplarily two – lying between 1000 and 10 kHz. By function 60b, the second partial loudspeaker signals 34 are, relative to the at least one audio signal, dampened in a spectral range above 1000 Hz with a reduction of the dampening within a spectral subrange within the spectral range 124, which subrange is located between 5 and 10 kHz. Further, function 60b may, as depicted in Fig. 15, lead to an amplification of the signals 34 within a

spectral range 128 between 500 Hz and 1 kHz. Naturally, the ranges and examples may be varied.

5 The effective overall spectrum of the acoustic signal arriving at the listener is determined partially by non-EQ'ed signal (amplitude panning within a layer) 28 and partially by EQ'ed signal (signal from virtual top/bottom) 34. Thus the effective overall EQ is a linear combination of unity and the top/bottom EQs 60a/60b. In that way, the EQing at the listener is fading in as a source 104 moves towards top position (or correspondingly towards bottom position).

10 Such a continuous fade/change in the amount of EQing is specifically beneficial, since the human auditory system can use those changes in the spectrum of the received signal to judge its location. Especially in tracked scenarios, this changes can be used to distinguish whether a specific spectral feature is a property of the actual signal, or changes while the listener is moving, and it can such be interpreted as a feature related to the source location.

15

Summarizing, a reproduction of object based audio or multichannel audio with reproduction of elevated or lowered height sounds (top and bottom) is enabled. A playback of input audio signals (featuring sound intended for reproduction over elevated or lower loudspeaker layers) over arbitrary loudspeaker setups is possible. Here, „loudspeaker setups“ does also include
20 devices and topologies like soundbars, TVs with built in loudspeakers, boomboxes, soundplates, loudspeaker arrays, smart speakers, and so forth. There is no need to have elevated or lower loudspeaker layers. Thus, a perceptual effect of top or bottom sounds in almost any arbitrary loudspeaker setup (even without elevated or lower loudspeakers) is made possible.

25

The embodiments are computationally efficient, such that it can also be beneficially used in scenarios where the (changing) listener position is known and/or (constantly) tracked by the playback system.

30 The embodiments can be used for channel-based audio, object-based audio, and scene-based audio (e.g. Ambisonics) input format signals.

Compared to rendering methods which are HRTF based, it is to be emphasized that the embodiments do not aim at simulating detailed specific binaural cues for specific object
35 positions in all possible directions (which might be difficult to achieve over a wide range). Instead, a good simulation of cues is produced that evoke the perception of a sound source above or below the listener (i.e., produce a virtual source above or below) at one specific position/direction. Thus, it is tried to mimic the perception for those two directions (top/bottom 102) in a very good / convincing way. A benefit of these two specific directions chosen is that,

besides the spectral cues, the two other dominant spatial audio cues (i.e. ITDs and ILDs) are minimal; theoretically, no ITD and no ILD occurs for sound sources perfectly above or below a listener, i.e., the particle velocity in horizontal direction is close to zero for the direct sound from the sound source. Thus, the two stage approach with panning horizontally and vertically,
5 potentially with virtually rendering the top/bottom speaker 102, is stable and leads to high accuracy.

In the following, we describe some further example selection criteria how loudspeakers of the plurality of loudspeakers could automatically be assigned to a set or a layer of loudspeakers
10 for reproduction of a virtual loudspeaker

- Criteria for selecting the loudspeakers for the sets/layers:
 - Chose every layer such, that preferably a 360 degree panning around a listener is possible.
- 15 ○ Choice of loudspeakers for the reproduction of the virtual height channel:
 - Use multiple loudspeakers, such that
 - 1) preferably choose loudspeakers that are already at elevated positions
 - 2) considering 1), select (further) loudspeakers to achieve an array surrounding
20 the listener
 - The selected loudspeakers should as good as possible enable that they can reproduce the signal for the virtual height channel such that: the generated soundfield at the listener position has zero or small particle velocity in horizontal direction.
 - 25 ▪ If multiple suitable loudspeakers are available, either *all* of them can be used, or the selection procedure could be as follows:
 - If possible, select loudspeakers symmetrically around the listener (ideally as (rotationally) symmetrical as possible)
 - If loudspeaker are available that are already arranged at elevated
30 positions (up or down) towards the desired elevation position of the intended virtual height source
 - the elevation angle of the loudspeakers should be as large as possible, i.e., always select the loudspeakers with the biggest elevation angles (as vertical as possible)
- 35 ○ Ideally, select as few loudspeakers as possible to fulfill the above criteria
- Of course, the loudspeakers can also be selected/assigned by the user “by hand”.

Possible input parameters for (possibly adaptive) rendering are:

- The angles (azimuth and elevation) from the listener position to the loudspeakers
 - This is under the assumption that all loudspeakers are equally far away and produce similar level at the listening position
 - If they are not equally far away, the level and/or delay can be balanced to achieve equal level / time of arrival at the listener position
 - In a scenario where the listener is tracked, also the distance to each loudspeaker is needed in addition to the angles, so that level and/or delay can be adapted.
 - Such a level and delay adaption in a tracked scenario can also be beneficial to achieve the above mentioned “small particle velocity in horizontal direction” criterium for the reproduction of the virtual height signals.
- To conclude, the embodiments described herein can optionally be supplemented by any of the important points or aspects described here. However, it is noted that the important points and aspects described here can either be used individually or in combination and can be introduced into any of the embodiments described herein, both individually and in combination. As an outcome of the latter, the above description inter alia, includes an apparatus for generating loudspeaker signals 12 for a plurality of loudspeakers 14 so that an application of the loudspeaker signals 12 at the plurality of loudspeakers 14 renders at least one audio object at an intended virtual position 104, the apparatus comprising an interface 16 configured to receive an audio input signal 18 which represents the at least one audio object, a first panning gain determiner 22, configured to determine, depending on the intended virtual position, first panning gains 24 for a first set 26 of loudspeakers of the plurality of loudspeakers, which are arranged within, or form, a first horizontal layer, the first panning gains 24 defining a derivation of first partial loudspeaker signals 28 from the at least one audio input signal 18, which are associated with a rendering of the at least one audio object at a first virtual position 106 upon application of the first partial loudspeaker signals 28 onto the first set 26 of loudspeakers, a vertical panning gain determiner 30, configured to determine, depending on the intended virtual position, further panning gains 32 for a panning between the first partial loudspeaker signals 28 and second partial loudspeaker signals 34 which are to be applied to a second set 36 of loudspeakers, which is vertically offset relative to the first layer set, so as to be arranged in, or form, a second horizontal layer, and is associated with a rendering of the at least one audio object at a second position 102 so as to pan between the first virtual position 106 and the second position 102, wherein the apparatus is configured to compose the loudspeaker signals 12 from the audio input signal 18 using the first panning gains 24 and the further panning gains

32. A second panning gain determiner 52 is also comprised, which is configured to determine, depending on the intended virtual position, second panning gains 54 for the second set of loudspeakers, the second panning gains 54 defining a derivation of the second partial loudspeaker signals 34 from the at least one audio input signal, and the apparatus is configured to compose the loudspeaker signals 12 from the audio input signal 18 using the first and second panning gains and the further panning gains. The first and second panning gain determiners 22, 52 are configured to select the first and second sets 26, 36 of loudspeakers of the plurality of loudspeakers so that the first and second layer sets have, among horizontal layers which the plurality of loudspeakers are distributed onto, the intended virtual position 104 vertically therebetween. Note that the first set 26 of loudspeakers and the second set 36 of loudspeakers may partially overlap, i.e. one loudspeaker may be contained by both sets 26 and 36. To be more precise, the plurality of loudspeakers may be distributed onto the horizontal layers in a manner that, for each horizontal layers, the loudspeakers belonging to that horizontal layer surround, horizontally (i.e. in horizontal projection) a listener position, or, differently speaking, allow for, horizontally, a 360 degree panning around the listener position, and for sake of achieving this circumstance, for instance, at least one pair of horizontal layers may share one or more of their loudspeakers. That is, horizontality and vertical offsetness of the horizontal layers may be abstracted to an extent that sometimes, such as for at least one pair of horizontal layers, one or more loudspeakers belong to more than one of the horizontal layers, respectively. In even other words, the above description, inter alia, includes an apparatus for generating loudspeaker signals 12 for a plurality of loudspeakers 14 so that an application of the loudspeaker signals 12 at the plurality of loudspeakers 14 renders at least one audio object at an intended virtual position 104, wherein the plurality of loudspeakers are distributed onto one or more horizontal layers, the apparatus comprising an interface 16 configured to receive an audio input signal 18 which represents the at least one audio object, a first loudspeaker signal set determiner 70, configured to determine, depending on the intended virtual position, first panning gains 24 for a first set of loudspeakers 26 of the plurality of loudspeakers, and use the first panning gains 24 to derive first partial loudspeaker signals 28 from the at least one audio input signal 18, which are associated with a rendering of the at least one audio object at a first virtual position 106 upon application of the first partial loudspeaker signals onto the first set 26 of loudspeakers, a second loudspeaker signal set determiner 72, configured to, by spectral shaping, derive second partial loudspeaker signals 34 from the at least one audio input signal 18, the second partial loudspeaker signals 34 being associated with a rendering of the at least one audio object at a second virtual position 102 upon application of the second partial loudspeaker signals 34 onto a second set of loudspeakers 36, the second virtual position being above or below the one or more horizontal layers, and a vertical panning gain determiner 30 configured to, depending on the intended

virtual position, determine further panning gains 32 for the first and second partial loudspeaker signals so as to pan between the first and second virtual positions, and a composer 40 configured to compose the loudspeaker signals from the first and second partial loudspeaker signals using the further panning gains 32. Again, note that the first set 26 of loudspeakers and the second set 36 of loudspeakers may partially overlap, i.e. one loudspeaker may be contained by both sets 26 and 36. To be more precise, the plurality of loudspeakers may be distributed onto the horizontal layers in a manner that, for each horizontal layer, the loudspeakers belonging to that horizontal layer surround, horizontally (i.e. in horizontal projection) a listener position, or, differently speaking, allow for, horizontally, a 360 degree panning around the listener position, and for sake of achieving this circumstance, for instance, at least one pair of horizontal layers may share one or more of their loudspeakers. That is, horizontality and vertical offsetness of the horizontal layers may be abstracted to an extent that sometimes, such as for at least one pair of horizontal layers, one or more loudspeakers belong to more than of the horizontal layers, respectively. All the other modifications described above and mentioned in the subsequent claims are feasible as well, such as the usage of spectral shaping 58 so as to derive the second partial loudspeaker signals 34 from the at least one audio signal 18 in order to result into the second position being a virtual position 102 above the highest one or below the lowest one of the horizontal layers.

Although some aspects have been described in the context of an apparatus, it is clear that these aspects also represent a description of the corresponding method, where a device or a part thereof corresponds to a method step or a feature of a method step. Analogously, aspects described in the context of a method step also represent a description of a corresponding apparatus or part of an apparatus or item or feature of a corresponding apparatus. Some or all of the method steps may be executed by (or using) a hardware apparatus, like for example, a microprocessor, a programmable computer or an electronic circuit. In some embodiments, one or more of the most important method steps may be executed by such an apparatus.

Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a digital storage medium, for example a floppy disk, a DVD, a Blu-Ray, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed. Therefore, the digital storage medium may be computer readable.

Some embodiments according to the invention comprise a data carrier having electronically readable control signals, which are capable of cooperating with a programmable computer system, such that one of the methods described herein is performed.

5 Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing one of the methods when the computer program product runs on a computer. The program code may for example be stored on a machine-readable carrier.

10 Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine-readable carrier.

In other words, an embodiment of the inventive method is, therefore, a computer program having a program code for performing one of the methods described herein, when the
15 computer program runs on a computer.

A further embodiment of the inventive methods is, therefore, a data carrier (or a digital storage medium, or a computer-readable medium) comprising, recorded thereon, the computer program for performing one of the methods described herein. The data carrier, the digital
20 storage medium or the recorded medium are typically tangible and/or non-transitionary.

A further embodiment of the inventive method is, therefore, a data stream or a sequence of signals representing the computer program for performing one of the methods described herein. The data stream or the sequence of signals may for example be configured to be
25 transferred via a data communication connection, for example via the Internet.

A further embodiment comprises a processing means, for example a computer, or a programmable logic device, configured to or adapted to perform one of the methods described herein.
30

A further embodiment comprises a computer having installed thereon the computer program for performing one of the methods described herein.

A further embodiment according to the invention comprises an apparatus or a system
35 configured to transfer (for example, electronically or optically) a computer program for performing one of the methods described herein to a receiver. The receiver may, for example,

be a computer, a mobile device, a memory device or the like. The apparatus or system may, for example, comprise a file server for transferring the computer program to the receiver.

5 In some embodiments, a programmable logic device (for example a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

10 The apparatus described herein may be implemented using a hardware apparatus, or using a computer, or using a combination of a hardware apparatus and a computer.

The apparatus described herein, or any components of the apparatus described herein, may be implemented at least partially in hardware and/or in software.

15 The methods described herein may be performed using a hardware apparatus, or using a computer, or using a combination of a hardware apparatus and a computer.

20 The methods described herein, or any parts of the methods described herein, may be performed at least partially by hardware and/or by software.

The above described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

References

30 [1] A.B. S and S.M. R. Apparent sound source translator. February 1966. US Patent 3,236,949.

[2] Philip A Nelson, Hareo Hamada, and Stephen J Elliott. Adaptive inverse filters for stereophonic sound reproduction. IEEE Transactions on Signal Processing, 40(7):1621-
35 1632, 1992.

- [3] P. A. Nelson and J. F. W. Rose. Errors in two-point sound reproduction. The Journal of the Acoustical Society of America, 118(1):193, 2005.
- [4] Takashi Takeuchi and Philip A. Nelson. Optimal source distribution for binaural synthesis over loudspeakers. The Journal of the Acoustical Society of America, 112(6):2786, 2002.
- [5] Hironori Tokuno, Ole Kirkeby, Philip A Nelson, and Hareo Hamada. Inverse filter of sound reproduction systems using regularization. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 80(5):809-820, 1997.
- [6] Ole Kirkeby, Philip A. Nelson, Hareo Hamada, and Felipe Orduna-Bustamante. Fast deconvolution of multichannel systems using regularization. IEEE Transactions on Speech and Audio Processing, 6(2):189-194, 1998.
- [7] Edgar Y Choueiri. Optimal crosstalk cancellation for binaural audio with two loudspeakers. Princeton University, page 28, 2008.
- [8] B. B. Bauer. Stereophonic earphones and binaural loudspeakers. J. Audio Eng. Soc., 9:148-151, 1961.
- [9] J. Huopaniemi. Virtual Acoustics and 3D Sound in Multimedia Signal Processing. PhD thesis, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland, 1999. Rep. 53.
- [10] Hyunkook Lee. Sound source and loudspeaker base angle dependency of phantom image elevation effect. J. Audio Eng. Soc, 65(9):733-748, 2017.
- [11] Hyunkook Lee, Dale Johnson, and Maksims Mironovs. Virtual hemispherical amplitude panning (vhap): A method for 3d panning without elevated loudspeakers. In Audio Engineering Society Convention 144, May 2018.
- [12] Young Woo Lee et al., "Virtual Height Speaker Rendering for Samsung 10.2-channel Vertical Surround System". In Audio Engineering Society Convention 131, October 2011.

- [13] Reinhard Gretzki and Andreas Silzle, "A new method for elevation panning reducing the size of the resulting auditory events", TecniAcustica, Bilbao, 2003.
- 5 [14] Christian Borß, "A Polygon-Based Panning Method for 3D Loudspeaker Setups," Audio Engineering Society Convention 137, Oct, 2014.
- [15] MPEG-H Standard, ISO/IEC 23008-3:2015(E).

Claims

1. An apparatus for generating loudspeaker signals (12) for a plurality of loudspeakers (14) so that an application of the loudspeaker signals (12) at the plurality of loudspeakers (14) renders at least one audio object at an intended virtual position (104), the apparatus comprising
- an interface (16) configured to receive an audio input signal (18) which represents the at least one audio object,
- a first panning gain determiner (22), configured to determine, depending on the intended virtual position, first panning gains (24) for a first set (26) of loudspeakers of the plurality of loudspeakers, which are arranged within a first layer set of one or more first horizontal layers, the first panning gains (24) defining a derivation of first partial loudspeaker signals (28) from the at least one audio input signal (18), which are associated with a rendering of the at least one audio object at a first virtual position (106) upon application of the first partial loudspeaker signals (28) onto the first set (26) of loudspeakers,
- a vertical panning gain determiner (30), configured to determine, depending on the intended virtual position, further panning gains (32) for a panning between the first partial loudspeaker signals (28) and one or more second partial loudspeaker signals (34) which is to be applied to a second set (36) of one or more loudspeakers, which is vertically offset relative to the first layer set, and is associated with a rendering of the at least one audio object at a second position (102) so as to pan between the first virtual position (106) and the second position (102),
- wherein the apparatus is configured to compose the loudspeaker signals (12) from the audio input signal (18) using the first panning gains (24) and the further panning gains (32).
2. Apparatus according to claim 1, wherein the second set (36) of one or more loudspeakers comprises more than one loudspeaker, the one or more second partial loudspeaker signals (34) comprise more than one second partial loudspeaker signals and the apparatus further comprises
- a second panning gain determiner (52), configured to determine, depending on the intended virtual position, second panning gains (54) for the second set of loudspeakers,

the second panning gains (54) defining a derivation of the second partial loudspeaker signals (34) from the at least one audio input signal, and

5 wherein the apparatus is configured to compose the loudspeaker signals (12) from the audio input signal (18) using the first and second panning gains and the further panning gains.

3. Apparatus according to claim 2,

10 wherein the second set (36) of loudspeakers are within a second layer set of one or more horizontal layers and the first and second layer sets are vertically offset to each other.

15 4. Apparatus according to any of claims 2 to 3,

wherein the second set (36) of loudspeakers are within a second layer set of one or more horizontal layers and the first and second layer sets are vertically offset to each other with the intended virtual position (104) being vertically therebetween.

20 5. Apparatus according to any of claims 2 to 4,

25 wherein the second set (36) of loudspeakers are within a second layer set of one or more horizontal layers and the first and second panning gain determiners (22, 52) are configured to select the first and second sets (26, 36) of loudspeakers of the plurality of loudspeakers so that the first and second layer sets are, among horizontal layers which the plurality of loudspeakers are distributed onto, vertically nearest to the intended virtual position (104) and vertically offset to each other with the intended virtual position (104) being vertically therebetween.

30 6. Apparatus according to any of claims 2 to 5, wherein the first and second panning gain determiners (22, 52) are configured to derive the first and second panning gains (24, 54) so that the first virtual position (106₁) and the second position (106₂) coincide in a vertical projection.

35 7. Apparatus according to claim 2 or 3, wherein the apparatus is configured to

derive the second partial loudspeaker signals (34) from the at least one audio signal (18) by spectral shaping (58), so that the second position is a virtual position (102) above or below the second layer set.

8. Apparatus according to claim 7, wherein
- 5 the spectral shaping (58) mimics properties of a Head Related Transfer Function, HRTF, along a perception direction from the second position (102).
9. Apparatus according to any one of claims 7 to 8, configured
- 10 so that the second position is vertically above the second layer set and to perform the spectral shaping (58) so that the second partial loudspeaker signals (34) are, relative to the at least one audio input signal, dampened in a notch spectral range (120) between 200 and 1000 Hz and amplified within one or more in peak spectral ranges (122₁, 122₂) between 1000 and 10 kHz, or
- 15 so that the second position is vertically below the second layer set and to perform the spectral shaping so that the second partial loudspeaker signals (34) are, relative to the at least one audio signal, dampened in a spectral range above 1000 Hz.
10. Apparatus according to any one of claims 7 to 9, configured
- 20 so that the second position is vertically above the second layer set and to perform the spectral shaping (58) so that the second partial loudspeaker signals (34) are, relative to the at least one audio input signal, dampened in a notch spectral range (120) between 200 and 1000 Hz and amplified within one or more in peak spectral ranges
- 25 (122₁, 122₂) between 1000 and 10 kHz, or
- so that the second position is vertically below the second layer set and to perform the spectral shaping so that the second partial loudspeaker signals (34) are, relative to the at least one audio signal, dampened in a spectral range (124) above 1000 Hz with an
- 30 intermediate reduction of the dampening within a spectral subrange (126) within the spectral range, located between 5 and 10 kHz, and amplified (128) between 500 Hz and 1 kHz.
11. Apparatus according to any one of claims 7 to 10, configured to
- 35 if the intended virtual position (104) is vertically above the second layer set, position the second position to be vertically above the second layer set, perform the spectral shaping so that the second partial loudspeaker signals are, relative to the at least one audio signal, dampened in a notch spectral range between

200 and 1000 Hz and amplified within one or more in peak spectral ranges
Between 1000 and 10 kHz, and

5 if the intended virtual position is vertically below the second layer set, position
the second position to be vertically below the second layer set, perform the
spectral shaping so that the second partial loudspeaker signals are, relative to
the at least one audio signal, dampened in a spectral range above 1000 Hz.

12. Apparatus according to any one of claims 7 to 11,
10 wherein the plurality of loudspeakers (14) forms a setup in which the loudspeakers are
associated with horizontal layers, and the apparatus is configured to be responsive to
a change of the intended virtual position so as to

15 if the intended virtual position is between two horizontal layers,

select the first layer set to be a first of the two horizontal layers and the second
layer set to be a second of the two horizontal layers, and the first set (26) out of
loudspeakers associated with the first horizontal layer and the second set (36)
20 out of loudspeakers associated with the second horizontal layer, wherein the
first and second panning gain determiner (22, 52) are configured to determine,
depending on the intended virtual position, the first and second panning gains,
and the spectral shaping (58) is switched off, so that the first virtual position is
within the first horizontal layer and the second virtual position is within the
25 second horizontal layer, and

if the intended virtual position is vertically offset to all horizontal layers towards above
or below the horizontal layers,

30 select the first layer set and the second layer set to be an outmost layer of the
horizontal layers nearest to the intended virtual position, and the first set (26)
and the second set (36) out of loudspeakers associated with the outmost layer,
wherein the first panning gain determiner (22) is configured to determine,
depending on the intended virtual position, the first panning gains and the
35 spectral shaping (58) being used, so that the second position is a virtual position
(102) vertically offset relative to the outmost layer towards a direction at which
the intended virtual position (104) lies.

13. Apparatus according to claim 12,

wherein the apparatus is configured to be responsive to a change of the intended virtual position so as to

5 if the intended virtual position is between two horizontal layers,

the first and second panning gain determiner (22, 52) are configured to determine, depending on the intended virtual position, the first and second panning gains so that the first virtual position (106₁) and the second position (106₂) coincide in a vertical projection, and the spectral shaping (58) is switched off, and/or

10

if the intended virtual position is vertically offset to all horizontal layers towards above or below the horizontal layers,

15 the first panning gain determiner (22) is configured to determine, depending on the intended virtual position, the first panning gains so that the first virtual position (106) coincides in a vertical projection, with the intended virtual position.

20 14. Apparatus according to any one of claims 7 to 13,

wherein the plurality of loudspeakers (14) forms a setup in which the loudspeakers are associated with one or more horizontal layers, and the apparatus is configured to be responsive to a number of the one or more horizontal layers and a change of the intended virtual position so as to

25

if the number of one or more horizontal layers is larger than one,

if the intended virtual position is between two horizontal layers,

30

select the first layer set to be a first of the two horizontal layers and the second layer set to be a second of the two horizontal layers, and the first set (26) out of loudspeakers associated with the first horizontal layer and the second set (36) out of loudspeakers associated with the second horizontal layer, wherein the first and second panning gain determiner (22, 52) are configured to determine, depending on the intended virtual position, the first and second panning gains, and the spectral shaping (58) is switched off, so that the first virtual position is within the first horizontal layer and the second virtual position is within the second horizontal layer, and

35

if the intended virtual position is vertically offset to all horizontal layers towards above or below the horizontal layers,

5 select the first layer set and the second layer set to be an outmost layer of the horizontal layers nearest to the intended virtual position, and the first set (26) and the second set (36) out of loudspeakers associated with the outmost layer, wherein the first panning gain determiner (22) is configured to determine, depending on the intended virtual position, the first panning gains and the spectral shaping (58) is used, so that the second position is a virtual position (102) vertically offset relative to the outmost layer towards a direction at which the intended virtual position (104) lies, and

10 if the number of one or more horizontal layers is one,

15

if the intended virtual position is within the one horizontal layer,

compose the loudspeaker signals (12) purely from the first partial loudspeaker signals, and

20

if the intended virtual position is vertically offset to the one horizontal layer,

select the first layer set and the second layer set to be the one horizontal layer, and the first set (26) and the second set (36) out of loudspeakers associated with the one horizontal layer, wherein the first panning gain determiner (22) is configured to determine, depending on the intended virtual position, the first panning gains and the spectral shaping (58) is used, so that the second position is a virtual position (102) vertically offset relative to the one horizontal layer towards a direction at which the intended virtual position (104) lies.

25

30

15. Apparatus according to claim 14,

wherein the apparatus is configured to be responsive to a number of the one or more horizontal layers and a change of the intended virtual position so as to

35

if the number of one or more horizontal layers is larger than one,

if the intended virtual position is between two horizontal layers,

the first and second panning gain determiner (22, 52) are configured to determine, depending on the intended virtual position, the first and second panning gains so that the first virtual position (106₁) and the second position (106₂) coincide in a vertical projection, and/or

5

if the intended virtual position is vertically offset to all horizontal layers towards above or below the horizontal layers,

10

the first panning gain determiner (22) is configured to determine, depending on the intended virtual position, the first panning gains so that the first virtual position (106) coincides in a vertical projection, with the intended virtual position, and/or

if the number of one or more horizontal layers is one,

15

if the intended virtual position is vertically offset to the one horizontal layer,

20

the first panning gain determiner (22) is configured to determine, depending on the intended virtual position, the first panning gains so that the first virtual position (106) coincides in a vertical projection, with the intended virtual position.

16. Apparatus according to any of claims 1 to 15,

25

wherein the first set (26) of loudspeakers is included in the second set (36) of one or more loudspeakers, and/or

wherein the second set (36) of one or more loudspeakers is included in the first set (26) of loudspeakers, and/or

30

wherein the first set (26) of loudspeakers and the second set (36) of one or more loudspeakers coincide, and/or

wherein the first set (26) of loudspeakers and the second set (36) of one or more loudspeakers partially overlap, and/or

35

wherein the first set (26) of loudspeakers and the second set (36) of one or more loudspeakers are disjoint sets.

17. Apparatus according to any of claims 1 to 16,

configured to select the first set (26) of loudspeakers out of the plurality of loudspeakers depending on a horizontal component of the intended virtual position or depending on the horizontal component of the intended virtual position and a vertical component of the intended virtual position, and/or

5

configured to select the second set (36) of one or more loudspeakers out of the plurality of loudspeakers depending on a vertical component of the intended virtual position or depending on the horizontal component of the intended virtual position and the vertical component of the intended virtual position.

10

18. Apparatus according to any of claims 1 to 17 wherein the second set of one or more loudspeakers comprises one or more loudspeakers at, or horizontally surrounding, the second position, and horizontally arranged between the first set of loudspeakers.

15

19. Apparatus according to any of claims 1 to 18,

wherein the first and/or second panning gain determiners (22, 52) are configured to determine the first and/or second panning gains (24, 54) further depending on a listener position.

20

20. Apparatus according to any previous claim,

wherein the plurality of loudspeakers refer to any one of, or a combination of, one or more loudspeaker arrays, one or more soundbars, one or more smart speakers, one or more stereo speakers, one or more surround sound setups, or one or more sets of individual loudspeakers.

25

21. Apparatus according to any previous claim,

30

wherein the audio input signal is one of a channel-based audio signal, object-based audio signal, and/or scene-based audio signal.

22. Apparatus according to any previous claim,

35

configured to derive the intended virtual position from the audio input signal.

23. Apparatus according to any previous claim,

wherein the panning gains are amplitude panning gains.

24. Apparatus according to any one of claims 1 to 23,
wherein the audio input signal is a channel-based audio signal defining an audio signal
for each of signal-specific loudspeaker positions,
wherein the apparatus is configured to treat each of a selection of one or more (or all)
out of the audio signals for the signal-specific loudspeaker positions as one of the at
least one audio object.
25. Apparatus according to claim 24, configured to
derive the intended virtual position of the one audio object from the loudspeaker
position of the respective audio signal.
26. Apparatus according to claim 25, configured to
wherein the intended virtual position of the one audio object is derived from the
loudspeaker position of the respective audio signal in a manner so that a mutual
positional relationship between the signal-specific loudspeaker position is maintained.
27. Apparatus according to any one of claims 1 to 26,
wherein the audio input signal is an object-based audio signal defining one or more
renderable audio objects,
wherein the apparatus is configured to use a selection of one or more (or all) out of the
one or more renderable audio objects as one of the at least one audio object.
28. Apparatus according to any previous claim,
configured to receive information on a change of the plurality of loudspeakers in terms
of loudspeaker position and to take the change into account in subsequent generation
of the loudspeaker signals and / or
configured to receive information on a change of the plurality of loudspeakers in terms
of number of loudspeakers and to take the change into account in subsequent
generation of the loudspeaker signals.

29. An apparatus for generating loudspeaker signals (12) for a plurality of loudspeakers (14) so that an application of the loudspeaker signals (12) at the plurality of loudspeakers (14) renders at least one audio object at an intended virtual position (104), wherein the plurality of loudspeakers are distributed onto one or more horizontal layers, the apparatus comprising
- 5
- an interface (16) configured to receive an audio input signal (18) which represents the at least one audio object,
- 10
- a first loudspeaker signal set determiner (70), configured to determine, depending on the intended virtual position, first panning gains (24) for a first set of loudspeakers (26) of the plurality of loudspeakers, and use the first panning gains (24) to derive first partial loudspeaker signals (28) from the at least one audio input signal (18), which are associated with a rendering of the at least one audio object at a first virtual position (106) upon application of the first partial loudspeaker signals onto the first set (26) of loudspeakers,
- 15
- a second loudspeaker signal set determiner (72), configured to, by spectral shaping, derive second partial loudspeaker signals (34) from the at least one audio input signal (18), the second partial loudspeaker signals (34) being associated with a rendering of the at least one audio object at a second virtual position (102) upon application of the second partial loudspeaker signals (34) onto a second set of loudspeakers (36), the second virtual position being above or below the one or more horizontal layers, and
- 20
- a vertical panning gain determiner (30) configured to, depending on the intended virtual position, determine further panning gains (32) for the first and second partial loudspeaker signals so as to pan between the first and second virtual positions, and
- 25
- a composer (40) configured to compose the loudspeaker signals from the first and second partial loudspeaker signals using the further panning gains (32).
- 30
30. Apparatus according to claim 29,
- 35
- wherein the first set of loudspeakers is within one or more horizontal layers which is/are, among the one or more horizontal layers, vertically nearest to the intended virtual position.
31. Apparatus according to claim 29 or 30,

wherein the first loudspeaker signal set determiner (70) is configured to select the first set of loudspeakers (26) of the plurality of loudspeakers so that the first set of loudspeakers is within one or more horizontal layers which is/are, among the one or more horizontal layers, vertically nearest to the intended virtual position.

5

32. Apparatus according to claim 29 or 30,

wherein the first loudspeaker signal set determiner (70) is configured so that the first set of loudspeakers is within one horizontal layer and to determine the first panning gains further depending on positions of the first set of loudspeakers within the one horizontal layer.

10

33. Apparatus according to any of claims 29 to 32,

wherein the first loudspeaker signal set determiner (70) is configured so that the first panning gains implement a pure amplitude panning so that the first virtual position is between positions of the set of first loudspeakers.

15

34. Apparatus according to any of claims 29 to 33,

wherein the first loudspeaker signal set determiner (70) is configured to determine the first panning gains further depending on a listener position.

20

35. Apparatus according to any of claims 29 to 34,

wherein the second loudspeaker signal set determiner (72) is configured so that the spectral shaping mimics properties of a Head Related Transfer Function, HRTF, along a perception direction from the second virtual position.

25

30 36. Apparatus according to any of claims 29 to 35,

wherein the second loudspeaker signal set determiner (72) is configured to derive the second partial loudspeaker signals from the at least one audio signal

35

so that the second partial loudspeaker signals are generated from the at least one audio signal using an amplitude gain factor which is equal for all of the second partial loudspeaker signals, or

by panning using panning gains which correspond to a horizontal central position or sweet spot position in-between the second set of loudspeakers, or

5 by panning gains which correspond to a horizontal position which coincides with a listener position along a vertical projection.

37. Apparatus according to any of claims 29 to 36,

10 wherein the first set of loudspeakers is included in the second set of loudspeakers, and/or

wherein the second set (36) of loudspeakers is included in the first set (26) of loudspeakers, and/or

15 wherein the first set of loudspeakers and the second set of loudspeakers coincide, and/or

20 wherein the first set (26) of loudspeakers and the second set (36) of loudspeakers partially overlap, and/or

wherein the first set of loudspeakers and the second set of loudspeakers are mutually exclusive.

25 38. Apparatus according to any of claims 29 to 37,

30 configured to select the first set (26) of loudspeakers out of the plurality of loudspeakers depending on a horizontal component of the intended virtual position or depending on the horizontal component of the intended virtual position and a vertical component of the intended virtual position, and/or

35 configured to select the second set (36) of loudspeakers out of the plurality of loudspeakers depending on a vertical component of the intended virtual position or depending on the horizontal component of the intended virtual position and the vertical component of the intended virtual position.

39. Apparatus according to any one of claims 29 to 38,

wherein the second loudspeaker signal set determiner (72) is configured so that the second virtual position is vertically above the one or more horizontal layers and to

perform the spectral shaping so that the second partial loudspeaker signals are, relative to the at least one audio signal, dampened in a notch spectral range between 200 and 1000 Hz and amplified within one or more in peak spectral ranges between 1000 and 10 kHz, or

5

wherein the second loudspeaker signal set determiner (72) is configured so that the second virtual position is vertically below the one or more horizontal layers and to perform the spectral shaping so that the second partial loudspeaker signals are, relative to the at least one audio signal, dampened in a spectral range above 1000 Hz.

10

40. Apparatus according to any one of claims 29 to 39,

wherein the second loudspeaker signal set determiner (72) is configured so that the second virtual position is vertically above the one or more horizontal layers and to perform the spectral shaping so that the second partial loudspeaker signals are, relative to the at least one audio signal, dampened in a notch spectral range between 200 and 1000 Hz and amplified within one or more in peak spectral ranges between 1000 and 10 kHz, or

15

20

wherein the second loudspeaker signal set determiner (72) is configured so that the second virtual position is vertically below the one or more horizontal layers and to perform the spectral shaping so that the second partial loudspeaker signals are, relative to the at least one audio signal, dampened in a spectral range above 1000 Hz with an intermediate reduction of the dampening within a spectral subrange within the spectral range, located between 5 and 10 kHz, and amplified between 500 Hz and 1 kHz.

25

41. Apparatus according to any one of claims 29 to 40,

wherein the second loudspeaker signal set determiner (72) is configured to,

30

if the intended virtual position is vertically above the one or more horizontal layers, position the second virtual position to be vertically above the one or more horizontal layers, perform the spectral shaping so that the second partial loudspeaker signals are, relative to the at least one audio signal, dampened in a notch spectral range between 200 and 1000 Hz and amplified within one or more in peak spectral ranges Between 1000 and 10 kHz, and

35

if the intended virtual position is vertically below the one or more horizontal layers, position the second virtual position to be vertically below the one or more

horizontal layers, perform the spectral shaping so that the second partial loudspeaker signals are, relative to the at least one audio signal, dampened in a spectral range above 1000 Hz.

5 42. Apparatus according to any one of claims 29 to 41,

wherein the composer is configured to be responsive to a change of the intended virtual position from an in-layer position, which is vertically within or in-between the one or more layers, to a position vertically offset from the one or more horizontal layers, by

10

controlling the further panning gains so as to fade from composing the loudspeaker signals purely from the first partial loudspeaker signals to composing the loudspeaker signals from the first and second partial loudspeaker signals so that the further panning gains pan from the first virtual position towards the second virtual position.

15

43. A system comprising:

a plurality of loudspeakers, and

20

an apparatus according to one of the preceding claims.

44. Method for generating loudspeaker signals (12) for a plurality of loudspeakers (14) so that an application of the loudspeaker signals (12) at the plurality of loudspeakers (14) renders at least one audio object at an intended virtual position (104), the method comprising

25

receiving an audio input signal (18) which represents the at least one audio object,

30 determining, depending on the intended virtual position, first panning gains (24) for a first set (26) of loudspeakers of the plurality of loudspeakers, which are arranged within a first layer set of one or more first horizontal layers, the first panning gains (24) defining a derivation of first partial loudspeaker signals (28) from the at least one audio input signal (18), which are associated with a rendering of the at least one audio object at a first virtual position (106) upon application of the first partial loudspeaker signals (28) onto the first set (26) of loudspeakers,

35

determining, depending on the intended virtual position, further panning gains (32) for a panning between the first partial loudspeaker signals (28) and one or more second partial loudspeaker signals (34) which is to be applied to a second set (36) of one or more

loudspeakers, which is vertically offset relative to the first layer set, and is associated with a rendering of the at least one audio object at a second position (102) so as to pan between the first virtual position (106) and the second position (102),

- 5 composing the loudspeaker signals (12) from the audio input signal (18) using the first panning gains (24) and the further panning gains (32).

45. Method for generating loudspeaker signals (12) for a plurality of loudspeakers (14) so that an application of the loudspeaker signals (12) at the plurality of loudspeakers (14) renders
10 at least one audio object at an intended virtual position (104), wherein the plurality of loudspeakers are distributed onto one or more horizontal layers, the method comprising

receiving an audio input signal (18) which represents the at least one audio object,

- 15 determining, depending on the intended virtual position, first panning gains (24) for a first set of loudspeakers (26) of the plurality of loudspeakers, and use the first panning gains (24) to derive first partial loudspeaker signals (28) from the at least one audio input signal (18), which are associated with a rendering of the at least one audio object at a first virtual position (106) upon application of the first partial loudspeaker signals onto the first set (26) of loudspeakers,

- 20 by spectral shaping, deriving second partial loudspeaker signals (34) from the at least one audio input signal (18), the second partial loudspeaker signals (34) being associated with a rendering of the at least one audio object at a second virtual position (102) upon application of the second partial loudspeaker signals (34) onto a second set of loudspeakers (36), the second
25 virtual position being above or below the one or more horizontal layers, and

- depending on the intended virtual position, determining further panning gains (32) for the first and second partial loudspeaker signals so as to pan between the first and second virtual positions, and

- 30 composing the loudspeaker signals from the first and second partial loudspeaker signals using the further panning gains (32).

46. Computer readable digital storage medium having stored thereon a computer program
35 having a program code for performing, when running on a computer, a method according to claims 44 or 45.