US 20130003998A1

(19) United States

(12) Patent Application Publication

Kirkeby et al.

(10) Pub. No.: US 2013/0003998 A1

(43) Pub. Date: Jan. 3, 2013

(54) **MODIFYING SPATIAL IMAGE OF A PLURALITY OF AUDIO SIGNALS**

(75) Inventors: **Ole Kirkeby**, Espoo (FI); **Jussi Virolainen**, Espoo (FI)

(73) Assignee: **NOKIA CORPORATION**, Espoo (FI)

(21) Appl. No.: **13/581,303**

(22) PCT Filed: **Feb. 26, 2010**

(86) PCT No.: **PCT/FI2010/050146**

§ 371 (c)(1), (2), (4) Date: **Sep. 14, 2012**

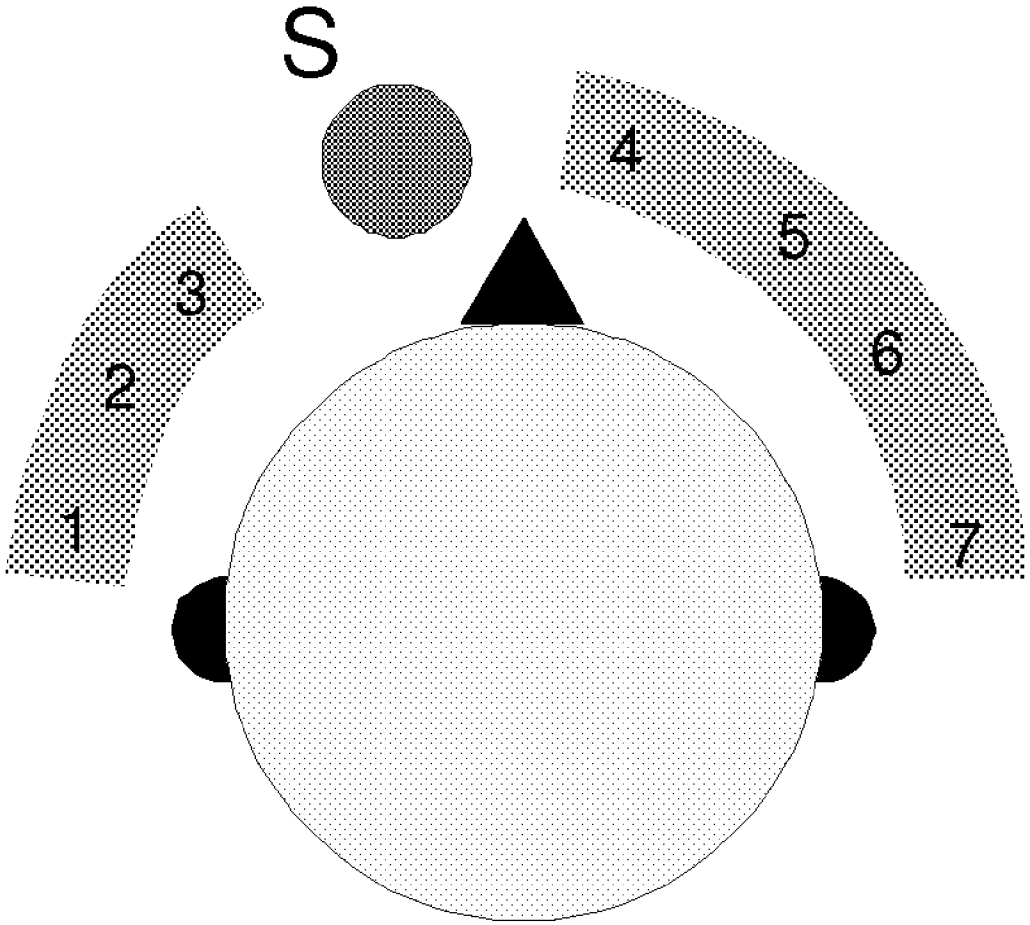
Publication Classification

(51) **Int. Cl.**
H04R 5/02 (2006.01)

(52) **U.S. Cl.** **381/300**

(57) **ABSTRACT**

A method comprising: modifying a sound stage produced by an input audio signal comprising two or more audio channels such that spatial room is relieved for one or more additional sound sources; and inserting said one or more additional sound sources in the relieved spatial room of the modified sound stage of the input audio signal without introducing spatial interference with the modified sound stage of the input audio signal.



Samsung v. Zophonos
IPR2026-00084
Exhibit 1019

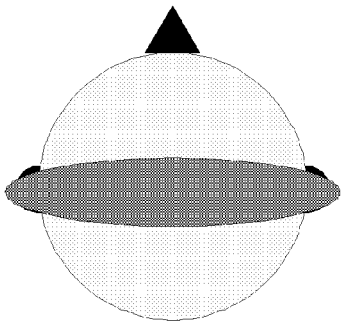


Fig.1a

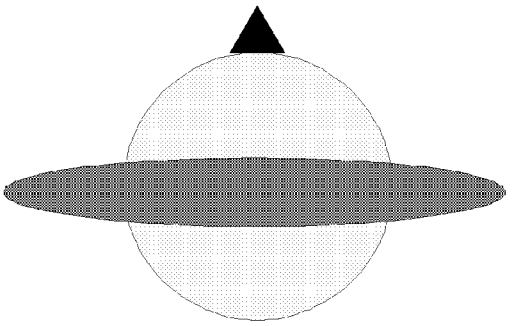


Fig.1b

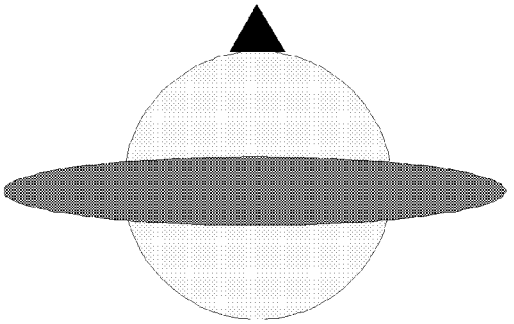


Fig. 2a

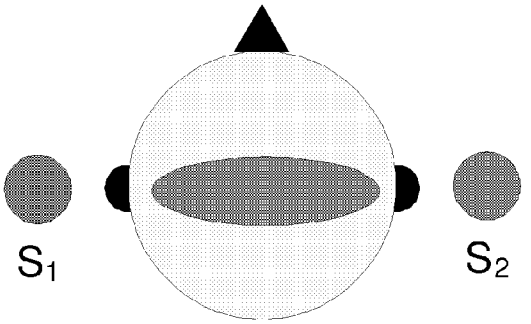


Fig.2b

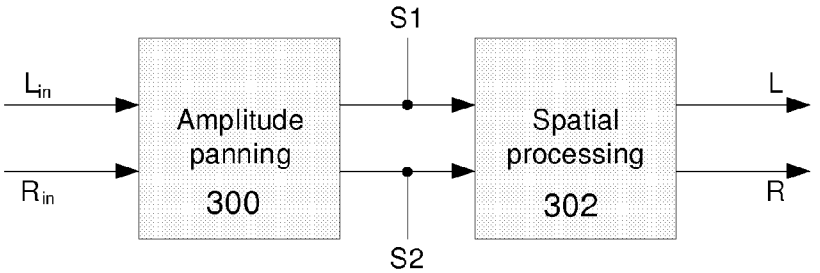


Fig. 3

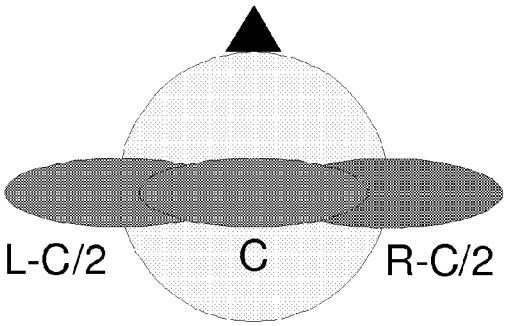


Fig. 4a

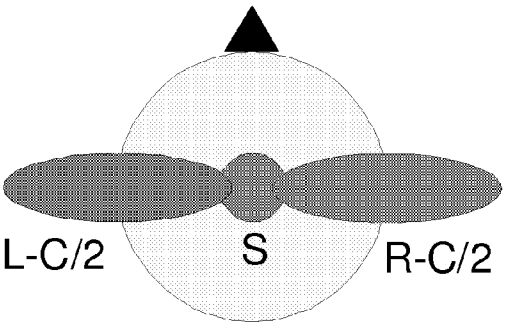


Fig. 4b

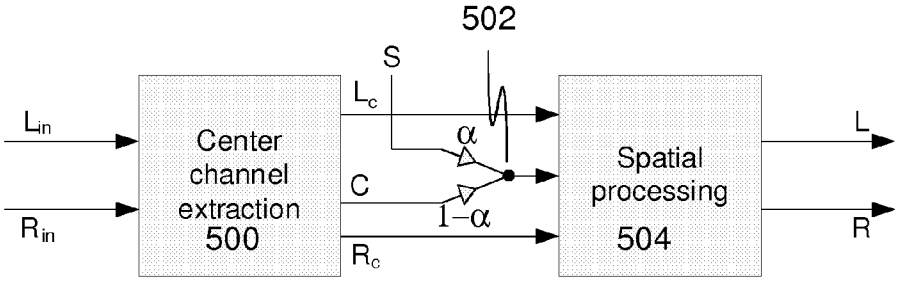


Fig. 5

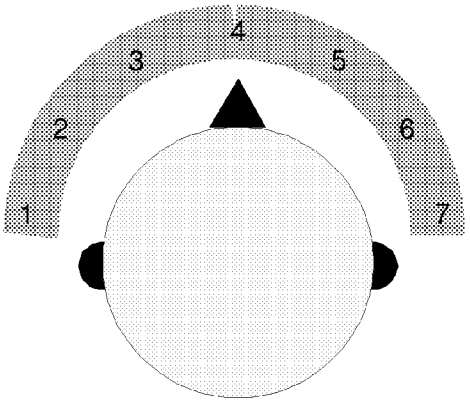


Fig. 6a

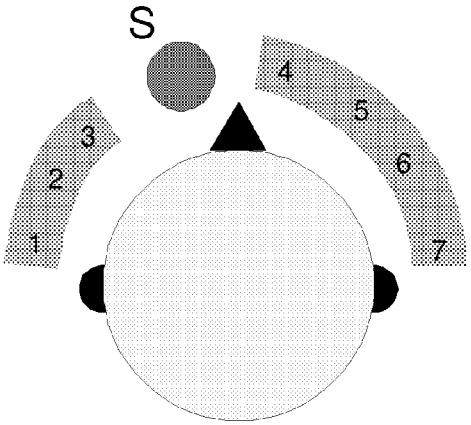


Fig. 6b

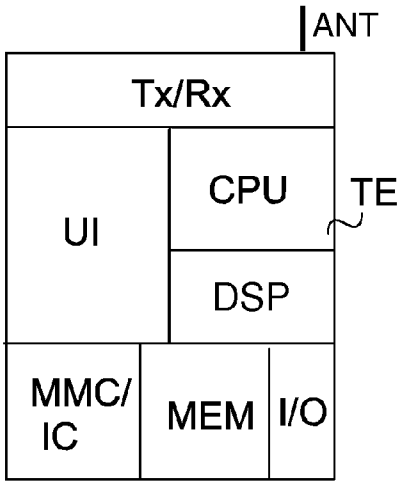


Fig. 7

MODIFYING SPATIAL IMAGE OF A PLURALITY OF AUDIO SIGNALS

FIELD OF THE INVENTION

[0001] The present invention relates to audio processing, and more particularly to modifying spatial image of a plurality of audio signals.

BACKGROUND OF THE INVENTION

[0002] The human auditory system is very good at focusing attention on a sound source according to its position. This is sometimes referred to as the 'cocktail-party effect': in a noisy crowded room it is possible to have a conversation, since the listener can shut out most of the distracting sound coming from directions other than that of the person they are talking to.

[0003] It is much harder for a listener to separate sounds that come from the same direction. For example, when listening to stereo music over headphones the sound does not appear to come from a single position but is rather smeared out over a wide sound stage. In that case it is difficult to understand speech, if the voice is superimposed on the music without any attempt to separate the two spatially.

[0004] This may imply problems when using, for example, mobile phones. Contemporary mobile terminals include features, which enable to listen to high quality music reproduction via headphones. However, if a phone call is received during music reproduction, either the music is muted or the phone call is superimposed on the music. Consequently, a phone call or a voice message cannot be mixed in with a stereo music track without reducing intelligibility. It is therefore desirable to be able to modify the audio streams spatially so that the speech is easy to understand while the music track is still playing.

SUMMARY OF THE INVENTION

[0005] Now there has been invented an improved method and technical equipment implementing the method, by which the intelligibility of speech or any other audio signal is increased when mixed with another audio signal. Various aspects of the invention include a method, an apparatus and a computer program, which are characterized by what is stated in the independent claims. Various embodiments of the invention are disclosed in the dependent claims.

[0006] According to a first aspect, a method according to the invention is based on the idea of modifying a sound stage produced by an input audio signal comprising two or more audio channels such that spatial room is relieved for one or more additional sound sources; and inserting said one or more additional sound sources in the relieved spatial room of the modified sound stage of the input audio signal without introducing spatial interference with the modified sound stage of the input audio signal.

[0007] According to an embodiment, the input audio signal comprises a two- channel stereo signal, the method further comprising: narrowing the sound stage produced by the two-channel stereo signal by applying an amplitude panning process to input audio signal; and inserting one additional sound source at least on either side of the narrowed sound stage.

[0008] According to an embodiment, the amplitude panning process is applied to input signal components of said two-channel stereo signal according to

$$\begin{pmatrix} L_{out} \\ R_{out} \end{pmatrix} = \begin{pmatrix} 1-\square & \square \\ \square & 1-\square \end{pmatrix} \begin{pmatrix} L_{in} \\ R_{in} \end{pmatrix},$$

wherein L_{in}, L_{out}, R_{in} and R_{out} are input and output signal components of left and right stereo channels, respectively, and $0 \leqq \square 0.5$.

[0009] According to an embodiment, if the one or more additional sound sources are based on speech signals, the value of □ is adjusted to be approximately 0.3 or higher.

[0010] According to an embodiment, wherein the input audio signal comprises a two-channel stereo signal, the method further comprises: determining a center channel audio component based on audio components common to the stereo signals; narrowing the sound stage produced by the two-channel stereo signal by removing the center channel audio component; and inserting an additional sound source in a non-interfering spatial space between the extremes of the sound stage.

[0011] According to an embodiment, said removing the center channel audio component and said inserting the additional sound source is performed proportionally to each other according to factors 1-$\alpha$and α, respectively.

[0012] According to an embodiment, the value of α is adjusted in a time-varyingly.

[0013] According to an embodiment, upon determining that an additional sound source should be included in the sound stage produced by the two-channel stereo signal, the method further comprises: increasing the value of α gradually to a predetermined value, such as its maximum value, within a first predetermined period, for example one second.

[0014] According to an embodiment, the method further comprises: delaying feeding of the additional sound source for said first predetermined period.

[0015] According to an embodiment, upon determining that no active additional signal producing said additional sound source has been detected for a second predetermined period, the method further comprises: decreasing the value of α gradually to zero.

[0016] According to an embodiment, the input audio signal comprises Binaural cue coded downmixed signals, the method further comprising: suppressing audio signals arriving from at least one virtual audio source by selecting sub-bands having inter-channel time difference parameters within a predetermined range to be suppressed; and inserting said one or more additional sound sources in the Binaural cue coded downmixed signals instead of said suppressed audio signals.

[0017] According to an embodiment, the input audio signal comprises Directional audio coded signals, the method further comprising: suppressing audio signals arriving from at least one virtual audio source by selecting sub-bands having azimuth and/or elevation parameters within a predetermined range to be suppressed; and inserting said one or more additional sound sources in the Directional audio coded signals instead of said suppressed audio signals.

[0018] According to an embodiment, the input audio signal comprises Directional audio coded (DirAC) signals or Binaural cue coded (BCC) downmixed signals, the method further comprising: applying a repanning process to said input audio signal in order to re-allocate energy of one or more predefined DirAC or BCC signals to new spatial positions;

and inserting said one or more additional sound sources in the spatial positions relieved by said one or more predefined DirAC or BCC signals.

[0019] The arrangement according to the invention provides many advantages. It enables to include one or more additional sound sources based on audio signals, e.g. speech signals, in a sound stage produced by an original input audio signal(s) such that the additional sound sources are intelligible even if the original audio signal(s), e.g.

[0020] stereo music, belonging to the sound stage are still reproduced. Especially in a case of a stereo sound stage, there is provided straightforward methods for relieving non-interfering spatial room for one or two speech signals to be intelligibly mixed with the underlying sound stage. This provides an entertaining feature, for example, for social music services, wherein a push-to-talk feature could be available on a "Now listening to" page so that user's friends could instantaneously comment on the listened music.

[0021] According to a second aspect, there is provided an apparatus comprising at least one processor and at least one memory storing computer program code, wherein the at least one memory and stored computer program code are configured to, with the at least one processor, cause the apparatus to at least: modify a sound stage produced by an input audio signal comprising two or more audio channels such that spatial room is relieved for one or more additional sound sources; and insert said one or more additional sound sources in the relieved spatial room of the modified sound stage of the input audio signal without introducing spatial interference with the modified sound stage of the input audio signal.

[0022] According to a third aspect, there is provided a computer program product, stored on a computer readable medium and executable in a data processing device, for processing audio signals, the computer program product comprising: a computer program code section for modifying a sound stage produced by an input audio signal comprising two or more audio channels such that spatial room is relieved for one or more additional sound sources; and a computer program code section for inserting said one or more additional sound sources in the relieved spatial room of the modified sound stage of the input audio signal without introducing spatial interference with the modified sound stage of the input audio signal.

[0023] These and other aspects of the invention and the embodiments related thereto will become apparent in view of the detailed disclosure of the embodiments further below.

LIST OF DRAWINGS

[0024] In the following, various embodiments of the invention will be described in more detail with reference to the appended drawings, in which

[0025] FIGS. **1***a*, **1***b* show how the listener may perceive the spatial properties of stereo music when played back over headphones, without spatial processing and with spatial processing, respectively;

[0026] FIG. **2***a* shows a stereo widened sound stage;

[0027] FIG. **2***b* shows how the stereo widened sound stage of FIG. **2***a* is narrowed in order to make room for an additional signal;

[0028] FIG. **3** shows a reduced block diagram for the processing components required to produce the spatial effect of FIG. **2***b* according to an embodiment;

[0029] FIG. **4***a* shows the principle of a center channel common audio component for a stereo signal;

[0030] FIG. **4***b* shows how the sound stage of FIG. **4***a* is narrowed by removing the center channel common audio component in order to make room for an additional signal;

[0031] FIG. **5** shows a reduced block diagram for the processing components required to produce the spatial effect of FIG. **4***b* according to an embodiment;

[0032] FIGS. **6***a*, **6***b* illustrate a repanning-based embodiment for relieving spatial room between a plurality of virtual audio sources; and

[0033] FIG. **7** shows a reduced block chart of an apparatus according to an embodiment.

DESCRIPTION OF EMBODIMENTS

[0034] In the following, the invention will be illustrated by referring to (stereo) music as the source material, wherein spatial room is created for the insertion of an additional sound source based on a speech signal. It is, however, noted that the invention is not limited to music as the source material solely, but it can be implemented in any type of multi-channel audio with spatial content, including movie sound tracks, TV broadcasts, and games. Furthermore, the speech signals can be replaced by other types of material that take priority over the spatial sound track, for example UI sounds and alerts.

[0035] The first implementation examples are described on the basis of two-channel (stereo) input audio signal, but the basic aspects are applicable to multi-channel input audio signal as well, as illustrated in the implementation examples further below. It is also generally known that the sound stage created by a stereo signal can be modified in such a way that the listener perceives the sound stage as extending beyond the positions of the speakers at both sides. This process is generally referred to as stereo widening, wherein the widening effect is typically created by introducing cross-talk from the left input to the right loudspeaker, and from the right input to the left loudspeaker. There are known stereo widening schemes for both loudspeaker playback and headphone playback.

[0036] In the following, headphone playback is used as an example but the principle is the same with two closely spaced loudspeakers. In both cases, the positions of the sound sources can be assumed to be distributed along a line, or arc, extending from the left to the right relative to the listener, symmetrically around the median plane, in a way similar to what is experienced when sitting in front of a conventional stereo setup where the loudspeakers span an angle of 60 degrees as seen by the listener.

[0037] In the enclosed figures, the head of the listener is depicted from above, the triangle denoting the listener's nose and the two hemispheres denoting listener's ears, and the sound stage perceived by the listener is depicted by the area of the ellipsis.

[0038] FIGS. **1***a* and **1***b* show how the listener may perceive the spatial properties of stereo music when played back over headphones. Without spatial processing (FIG. **1***a*), all sound sources of the sound stage extend from the left ear to the right ear across the center of the head. With a spatial effect created by the stereo widening (FIG. **1***b*), the extremes of the sound stage are externalised so that some sound sources appear to be heard outside the head. Regardless of whether spatial processing is used or not, the sound stage (i.e. the spatial image) of a typical stereo music track is dense, with no gaps in which to squeeze in an additional sound source. This is depicted by the solid ellipsis area.

[0039] Now according to an embodiment applicable particularly to stereo signals, the spatial image of the original stereo input signal is modified such that spatial room is relieved for one or more additional audio sound sources, based on e.g. one or more additional signals, in such a way that the one or more additional sound sources may be inserted in the relieved spatial room without introducing spatial interference with the modified spatial image of the original stereo signal. Thus, by relieving spatial room from the original sound stage comprising e.g. music, it is possible to include contents of one or more additional audio signals, e.g. speech signals, in the sound stage of the original two-channel stereo signal as additional sound sources such that the additional sound sources are intelligible even if the stereo signal, e.g. music, is still reproduced.

[0040] According to an embodiment, the sound stage is narrowed so that there is room in the spatial image for additional (e.g. speech) signals on both sides. Stereo widening has little or no effect on stereo signals in a case when the audio in the left channel, L, is identical to the right, R. Consequently, the sound stage can be narrowed artificially by mixing the left and right channels together so that the two channels of the stereo signal that are input to the stereo widening network are more similar than in the original recording. This is a standard operation usually referred to as amplitude panning. Control of the width of the sound stage is achieved when amplitude panning is applied to both channels according to

$$\begin{pmatrix} L_{out} \\ R_{out} \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix} \begin{pmatrix} L_{in} \\ R_{in} \end{pmatrix}, \quad (1)$$

where α is a parameter that varies between 0-0.5. As seen in the equation (1), when $\alpha=0$, there is no effect on the stereo input; i.e. $L_{out}=L_{in}$ and $R_{out=Rin}$. Likewise, when $\alpha=0.5$, the two output signals are made identical; i.e. $L_{out}=R_{out}=0.5*L_{in}+0.5*R_{in}$. The experiments have shown that when a value of a becomes greater than approximately 0.3, the sound stage of an average stereo signal is narrowed enough in order to add a speech signal on both the left and right side of the listener. This enables e.g. two callers, or voice messages, to be heard simultaneously and yet intelligibly with the underlying audio signal of the sound stage.

[0041] This is illustrated in FIGS. **2***a* and **2***b*, wherein the (stereo widened) sound stage of FIG. **2***a* is narrowed in order to make room for speech signals S_1 and S_2 on both sides of the listener.

[0042] It is to be noted that depending on the nature of the additional audio signal (e.g. a non-speech signal) to be added as a sound source to the sound stage, it may be possible to add one or more additional sound sources on one or both sides of the listener with significantly smaller value of a than 0.3. For some type of additional audio signals, for example various alerts or user interface sounds, even a value of α less than 0.1 may be sufficient.

[0043] FIG. **3** shows an embodiment of an exemplified block diagram for the processing components required to produce the spatial effect of FIG. **2***b*. First the two stereo input channels L_{in} and R_{in} are fed in an amplitude panning unit 300, which controls the amplitude panning process by the value of α as described above. With the suitable value of α, the sound stage output from the amplitude panning unit **300** is narrowed enough so as to insert an additional sound source based on audio signals S**1**, S**2** on one or both sides of the narrowed sound stage. The narrowed sound stage produced from the two stereo input channels L_{in} and R_{in} and the one or two additional sound sources based on audio signals S**1**, S**2** are then fed into the spatial processing unit **302**. The spatial processing unit **302** then creates a 3-D spatial audio image, manifested by the left L and right R audio signals, to be reproduced via headphone playback.

[0044] According to another embodiment, the sound stage can be narrowed by making room in the middle of the sound stage. A sound source based on e.g. a speech signal can be added in the middle of the sound stage, instead of at one of the sides, by subtracting out the component common to the two channels in the stereo input. FIG. **4***a* illustrates an example, wherein the common component C of a sound stage has been determined according to a center channel extraction algorithm.

[0045] There are many known algorithms available for center channel extraction, and they are typically dependent on the used surround sound process. In the sound stage, the left ear component L-C/2 and the right ear component R-C/2 are at least partly overlapping with the center channel (common component) C. Typically, the center channel extraction cannot be made perfectly, and in order to avoid processing artifacts, it is preferable to allow the common component to be relatively wide (as shown in FIG. **4***a*) by adjusting the parameters of the center channel extraction algorithm appropriately.

[0046] As seen in FIG. **4***a*, the result of the application of the center channel extraction algorithm is that the left ear component L-C/2 and the right ear component R-C/2 are not spatially interfering each other, but there is spatial room between them, if the center channel (common) component C is removed. This is illustrated in FIG. **4***b*, wherein the sound stage is narrowed by dividing it into two parts L-C/2 and R-C/2 having spatial room there between, whereby an additional audio signal S can be inserted as an additional sound source to the sound stage without spatial interference with the modified spatial image of the original stereo signal while still allowing the additional audio signal to be intelligibly heard.

[0047] According to an embodiment, it is preferable to limit the number of simultaneously appearing sound sources to one, since typically there is room for only a single additional sound source in the center of the sound stage. For instance in case of the additional sound sources are based on speech signals, if several people are speaking at the same time, it is difficult to identify the active talker, i.e. the phenomenon familiar with the conventional teleconferencing equipment with mono playback.

[0048] FIG. **5** shows an embodiment of an exemplified block diagram for the processing components required to produce the spatial effect of FIG. **4***b*. First the two stereo input channels L_{in} and R_{in} are fed in a center channel extraction unit **500**, which produces output signal components L_c, C and R_c representing substantially the sound stage illustrated in

[0049] FIG. **4***a*. The mutually non-interfering left-ear component L_c and right-ear component R_c are fed into a spatial processing unit **504** as such, but the center channel (common) component C is multiplied by 1-α and the additional audio signal S is, in turn, multiplied by a before feeding the both signals into a summing unit **502**. Thus, by adjusting the value of α, it can be determined whether the center channel component C, the additional sound source based on the audio signal S or a mix of said signals C and S in fed into the spatial processing unit **504**. The spatial processing unit **504** then

creates a 3-D spatial audio image, manifested by the left L and right R audio signals, to be reproduced via headphone playback.

[0050] A skilled person immediately appreciates that the spatial processing method applied by the spatial processing units **302** in FIGS. **3** and **504** in FIG. **5** may vary depending on the application used. Moreover, since the basic aspects are applicable in loudspeaker playback as well, the spatial processing method applied in loudspeaker playback is preferably different than in headphones playback. Thus, the applied spatial processing method as such is not relevant for embodiments described herein.

[0051] In the above embodiments of narrowing the sound stage, if there is no additional sound source(s) based on audio signal(s) S to be included, the spatial content of the original audio signal, e.g. music, is perceived by the listener in a reduced and thus in an unsatisfactory manner. Therefore, it is advantageous to modify the sound stage and make room for an additional sound source only when additional signal(s) with audible content is/are present, e.g. in case an additional signal based on which an additional sound source is to be introduced is a speech signal, the sound stage may be modified to make room for the additional sound source only when there is voice activity in the respective signal.

[0052] According to an embodiment, this is implemented by making the parameter α time-varying. In the embodiments described in FIGS. **3** and **5**, when $\alpha=0$, there is no room for an additional sound source in the sound stage and the speech channel(s) S based on which additional sound source(s) is/are to be introduced are muted. According to an embodiment, upon determining that an additional sound source should be included in the sound stage, the value of α is gradually increased to a predetermined value providing desired width of sound stage for the original audio signal within a first predetermined period, for example one second. Thereby, a pleasant and entertaining spatial effect is achieved. It should be noted that the maximum value of α is 0.5 for the narrowing of the sound stage and 1 for removing the center channel

[0053] According to a further embodiment, feeding of the additional sound source(s) based on signal(s) S is/are delayed by the same (first) predetermined period as it takes to increase α to the predetermined value. This enables to modify the sound stage before the additional sound source, e.g. speech, is heard.

[0054] According to an embodiment, when there has been no active additional signal for a second predetermined period, for example five seconds, then the value of α is reduced to zero again using the same gradual update scheme as when it is increased, but naturally in a reversed manner.

[0055] The above embodiments have been described in view of two-channel (stereo) input audio signal, but as mentioned above, the basic aspects are applicable to multi-channel input audio signal as well. A skilled person is aware that there are different ways to implement the spatial processing, and for example stereo widening may be considered merely a special case that works on a two-channel input.

[0056] Thus, the basic aspect of the embodiments can be generalized as modifying the spatial image of an input audio signal comprising two or more audio channels such that spatial room is relieved for one or more additional sound sources, based on e.g. one or more additional audio signals, in such a way that the one or more additional sound sources may be inserted in the relieved spatial room without introducing spatial interference with the modified spatial image of the original input signal, and inserting said one or more additional sound sources in the relieved spatial room of the modified spatial image of the input audio signal. Thus, also in the case of multi-channel input audio having more than two channels it is possible to insert one or more additional sound sources into the sound stage such that the additional sound source(s) are intelligible even if the multi-channel audio signal(s) are still reproduced.

[0057] A number of audio processing algorithms, referred to as 'virtual surround', utilize the properties of the human auditory system to create the perception that the sound stage is created by more audio sources than are actually present. These algorithms may be based on the utilization of head-related transfer functions (HRTFs), parametric audio coding techniques like Binaural Cue Coding (BCC), reflections or diffuse sound sources or a combination of those. Many of these algorithms may include, at least in some stage of the processing, more than two channel signals.

[0058] In Binaural cue coding (BCC) the encoder transforms input signals into the frequency domain using for example the Fourier transform or QMF filterbank techniques, and then performs spatial analysis. Inter-channel level difference (ILD) and time difference (ITD) parameters as well as additional parameters are estimated for each frequency sub-band in each input frame. These parameters are transmitted as side information along with a downmixed audio signal that is created by combining the input signals.

[0059] In Directional Audio coding (DirAC) the signals from spatial microphone system, such as the B-format Sound Field microphone, are analysed by dividing the input signals into frequency bands. Direction-of-arrival and the diffuseness are estimated individually for each time instance and frequency band. The spatial side information which consists of azimuth, elevation, and diffuseness values for each frequency band are transmitted with omnidirectional microphone signal.

[0060] According to an embodiment, if the audio signal is already BCC or DirAC coded, it is possible to suppress sounds that are coming from certain (virtual) spatial direction (s). For example, from N spatial directions, one or two spatial directions could be suppressed to make room for one or more additional sound source (s) to be mixed therein, and the additional sound sources based on e.g. additional audio signal(s) may then be inserted instead of the suppressed virtual audio sources. In practise, this can be implemented by manipulating the side information in the parametric domain. For example, in BCC coded signals sub-bands that have ITD at certain range can be suppressed. In DirAC coded signals, sub-bands having certain azimuth and/or elevation values can be suppressed.

[0061] Repanning is an audio processing method, basically applied to stereo music tracks, which maps energy in a specific spatial position to new spatial position. According to an embodiment, repanning is applied for BCC or DirAC coded signals. Thus, by re-allocating energy of certain BCC or DirAC coded signals to new spatial positions, spatial room may be relieved from the sound stage allowing one or more additional sound sources to be included in the sound stage, while still preserving substantially all content in the original signal.

[0062] The principle of this embodiment is illustrated in FIGS. **6***a* and **6***b*. In FIG. **6***a*, the virtual audio sources of the sound stage, denoted by numbers **1** to **7**, are equally distributed in the sound stage. In FIG. **6***b*, as a result of the repanning process, the virtual audio sources **1** to **3** and **4** to **7**, respectively, are squeezed together and pulled apart into two groups in order to make room for an additional audio signal S slightly to the left of the listener.

[0063] The process for making spatial room through repanning is described more in detail in the patent application

publication US2008/0298610, "Parameter Space Re-Panning for Spatial Audio", which is incorporated in its entirety herein by reference.

[0064] According to an embodiment, the sound stage is not limited to be located in front/sides of the listener, but it can also extend behind the listener, if an advanced rendering technology, for example with head-tracking, is used.

[0065] A skilled man appreciates that any of the embodiments described above may be implemented as a combination with one or more of the other embodiments, unless there is explicitly or implicitly stated that certain embodiments are only alternatives to each other.

[0066] FIG. **7** illustrates a simplified structure of an apparatus, i.e. a data processing device (TE), wherein the sound stage modifying method according to the embodiments can be implemented. The data processing device (TE) can be, for example, a mobile terminal, a PDA device or a personal computer (PC). The data processing unit (TE) comprises I/O means (I/O), a central processing unit (CPU) and memory (MEM). The memory (MEM) comprises a read-only memory ROM portion and a rewriteable portion, such as a random access memory RAM and FLASH memory. The information used to communicate with different external parties, e.g. a CD-ROM, other devices and the user, is transmitted through the I/O means (I/O) to/from the central processing unit (CPU). If the data processing device is implemented as a mobile station, it typically includes a transceiver Tx/Rx, which communicates with the wireless network, typically with a base transceiver station (BTS) through an antenna (ANT). User Interface (UI) equipment typically includes a display, a keypad, a microphone and connecting means for headphones. The data processing device may further comprise connecting means MMC, such as a standard form slot, for various hardware modules or as integrated circuits IC, which may provide various applications to be run in the data processing device.

[0067] Accordingly, the sound stage modifying method according to the embodiments may be executed in a central processing unit CPU or in a dedicated digital signal processor DSP (a parametric code processor) of the data processing device, and in at least one memory MEM storing computer program code, wherein the at least one memory and stored computer program code are configured to, with the at least one processor, cause the apparatus to at least modify a spatial image of two or more audio signals such that spatial room is relieved for one or more additional audio signals, which spatial room has no spatial interference between said two or more audio signals and insert said one or more additional audio signals in the relieved spatial room of the spatial image of the two or more audio signals.

[0068] Thus, the functionalities of the embodiments may be implemented in an apparatus, such as a mobile station, as a computer program which, when executed in a central processing unit CPU or in a dedicated digital signal processor DSP, affects the terminal device to implement procedures of the invention. Functions of the computer program SW may be distributed to several separate program components communicating with one another. The computer software may be stored into any memory means, such as the hard disk of a PC or a CD-ROM disc, from where it can be loaded into the memory of mobile terminal. The computer software can also be loaded through a network, for instance using a TCP/IP protocol stack.

[0069] It is also possible to use hardware solutions or a combination of hardware and software solutions to implement the inventive means. Accordingly, the above computer program product can be at least partly implemented as a hardware solution, for example as ASIC or FPGA circuits, in a hardware module comprising connecting means for connecting the module to an electronic device, or as one or more integrated circuits IC, the hardware module or the ICs further including various means for performing said program code tasks, said means being implemented as hardware and/or software.

[0070] It is obvious that the present invention is not limited solely to the above-presented embodiments, but it can be modified within the scope of the appended claims.

1-27. (canceled)

28. A method comprising:

modifying a sound stage produced by an input audio signal comprising two or more audio channels such that spatial room is relieved for one or more additional sound sources; and

inserting said one or more additional sound sources in the relieved spatial room of the modified sound stage of the input audio signal without introducing spatial interference with the modified sound stage of the input audio signal.

29. The method according to claim **28**, wherein the input audio signal comprises a two-channel stereo signal, the method further comprising:

narrowing the sound stage produced by the two-channel stereo signal by applying an amplitude panning process to input audio signal; and

inserting one additional sound source at least on either side of the narrowed sound stage.

30. The method according to claim **29**, wherein the amplitude panning process is applied to input signal components of said two-channel stereo signal according to

$$\begin{pmatrix} L_{out} \\ R_{out} \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix} \begin{pmatrix} L_{in} \\ R_{in} \end{pmatrix},$$

wherein L_{in}, L_{out}, R_{in} and R_{out} are input and output signal components of left and right stereo channels, respectively, and $0 \leqq \alpha \leqq 0.5$.

31. The method according to claim **30**, wherein if the one or more additional sound sources are based on speech signals, the value of a is adjusted to be approximately 0.3 or higher.

32. The method according to claim **28**, wherein the input audio signal comprises a two-channel stereo signal, the method further comprising:

determining a center channel audio component based on audio components common to the stereo signals;

narrowing the sound stage produced by the two-channel stereo signal by removing the center channel audio component; and

inserting an additional sound source in a non-interfering spatial space between the extremes of the sound stage.

33. The method according to claim **32**, wherein said removing the center channel audio component and said inserting the additional sound source is performed proportionally to each other according to factors $1-\alpha$ and α, respectively.

34. An apparatus comprising at least one processor and at least one memory storing computer program code, wherein the at least one memory and stored computer program code are configured to, with the at least one processor, cause the apparatus to at least:

modify a sound stage produced by an input audio signal comprising two or more audio channels such that spatial room is relieved for one or more additional sound sources; and

insert said one or more additional sound sources in the relieved spatial room of the modified sound stage of the input audio signal without introducing spatial interference with the modified sound stage of the input audio signal.

35. The apparatus according to claim **34**, wherein the input audio signal comprise a two-channel stereo signal, wherein the at least one memory and stored computer program code are further configured to, with the at least one processor, cause the apparatus to at least:

narrow the sound stage produced by the two-channel stereo signal by applying an amplitude panning process to input audio signal; and

insert one additional sound source at least on either side of the narrowed sound stage.

36. The apparatus according to claim **35**, wherein the amplitude panning process is arranged to be applied to input signal components of said two-channel stereo signal according to

$$\begin{pmatrix} L_{out} \\ R_{out} \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix} \begin{pmatrix} L_{in} \\ R_{in} \end{pmatrix},$$

wherein L_{in}, L_{out}, R_{in} and R_{out} are input and output signal components of left and right stereo channels, respectively, and $0 \leqq \alpha \leqq 0.5$.

37. The apparatus according to claim **36**, wherein if the one or more additional sound sources are based on speech signals, the value of α is arranged to be adjusted to be approximately 0.3 or higher.

38. The apparatus according to claim **34**, wherein the input audio signal comprises two-channel stereo signal, wherein the at least one memory and stored computer program code are further configured to, with the at least one processor, cause the apparatus to at least:

determine a center channel audio component based on audio components common to the stereo signals;

narrow the sound stage produced by the two-channel stereo signal by removing the center channel audio component; and

insert an additional sound source in a non-interfering spatial space between the extremes of the sound stage.

39. The apparatus according to claim **38**, wherein said removing the center channel audio component and said inserting the additional sound source are arranged to be performed proportionally to each other according to factors $1-\alpha$ and α, respectively.

40. The apparatus according to claim **36**, wherein the value of α is arranged to be adjusted in time-varyingly.

41. The apparatus according to claim **40**, wherein upon determining that an additional sound source should be included in the sound stage produced by the two-channel stereo signal, the at least one memory and stored computer program code are further configured to, with the at least one processor, cause the apparatus to at least:

increase the value of α gradually to a predetermined value, such as its maximum value, within a first predetermined period, for example one second.

42. The apparatus according to claim **41**, further comprising:

delaying feeding of the additional sound source for said first predetermined period.

43. The apparatus according to claim **41**, wherein upon determining that no active additional signal producing said additional sound source has been detected for a second predetermined period, the at least one memory and stored computer program code are further configured to, with the at least one processor, cause the apparatus to at least:

decrease the value of α gradually to zero.

44. The apparatus according to claim **34**, wherein the input audio signal comprises binaural cue coded downmixed signals, the at least one memory and stored computer program code are further configured to, with the at least one processor, cause the apparatus to at least:

suppress audio signals arriving from at least one virtual audio source by selecting sub-bands having inter-channel time difference parameters within a predetermined range to be suppressed; and

insert said one or more additional sound sources in the binaural cue coded downmixed signals instead of said suppressed audio signals.

45. The apparatus according to claim **34**, wherein the input audio signal comprises directional audio coded signals, the at least one memory and stored computer program code are further configured to, with the at least one processor, cause the apparatus to at least:

suppressing audio signals arriving from at least one virtual audio source by selecting sub-bands having azimuth and/or elevation parameters within a predetermined range to be suppressed; and

inserting said one or more additional sound sources in the directional audio coded signals instead of said suppressed audio signals.

46. The apparatus according to claim **34**, wherein the input audio signal comprise directional audio coded signals or binaural cue coded downmixed signals, the at least one memory and stored computer program code are further configured to, with the at least one processor, cause the apparatus to at least:

applying a repanning process to said input audio signal in order to re-allocate energy of one or more predefined directional audio coded or binaural cue coded signals to new spatial positions; and

inserting said one or more additional sound sources in the spatial positions relieved by said one or more predefined directional audio coded or binaural cue coded signals.

47. A computer program product, stored on a computer readable medium and executable in a data processing device, for processing audio signals, the computer program product comprising:

a computer program code section for modifying a sound stage produced by an input audio signal comprising two or more audio channels such that spatial room is relieved for one or more additional sound sources; and

a computer program code section for inserting said one or more additional sound sources in the relieved spatial room of the modified sound stage of the input audio signal without introducing spatial interference with the modified sound stage of the input audio signal.

* * * * *