

(51) International Patent Classification:
G10L 19/00 (2006.01)(21) International Application Number:
PCT/EP2008/068371(22) International Filing Date:
30 December 2008 (30.12.2008)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **NOKIA CORPORATION** [—/FI]; Keilalahdentie 4, FI-02150 Espoo (FI).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **OJANPERÄ, Juha Petteri** [FI/FI]; Luuppikatu 3, FI-37120 Nokia (FI).(74) Agent: **COHAUSZ & FLORACK**; Bleichstrasse 14, 40211 Düsseldorf (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

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

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

Published:

— with international search report (Art. 21(3))

(54) Title: PARAMETRIC AUDIO CODING

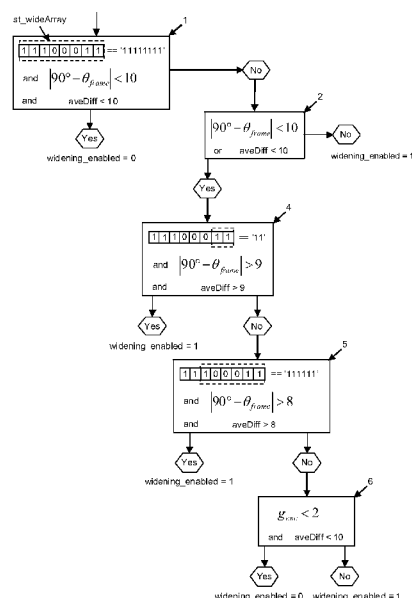


Fig. 5

(57) Abstract: It is disclosed to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair.

5 **Parametric Audio Coding**

FIELD

This invention relates to parametric audio coding.

10 BACKGROUND

Channel capacity for transmitting an audio signal or storage capacity for storing an audio signal is often limited. In consequence, audio coding is applied to reduce the required bitrate. It has been the industries
15 constant goal to develop audio coding techniques enabling high quality audio signal reconstruction while reducing the required bitrate to a minimum. This is especially true for multichannel audio coding where, compared to single channel audio signals, an even larger amount of
20 information has to be dealt with.

SUMMARY OF SOME EXEMPLARY EMBODIMENTS OF THE INVENTION

According to a first aspect of the present invention, a first apparatus is disclosed, comprising a processor
25 configured to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second
30 channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of

the audio signal frame pair and at least one preceding audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio
5 signal frame pair.

According to the first aspect of the present invention, further a second apparatus is disclosed, comprising means for determining whether a gain should be applied to at
10 least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a
15 parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding
20 audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair.

25 The means for determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed
30 first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising

a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on
5 a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair, may for instance comprise a processor configured to determine whether a gain should be applied to at least one frequency subband
10 of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio
15 signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on
20 a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair, but said means is not limited thereto.

25 The first apparatus as well as the second apparatus according to the first aspect of the present invention can be a module that forms part of or is to form part of another apparatus (such as for instance an audio encoder or decoder), for instance a processor, or it can be a
30 separate apparatus.

According to a second aspect of the present invention, further a method is disclosed, comprising determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair.

According to a third aspect of the present invention, further a program is disclosed, comprising program code for performing the method according to the second aspect of the present invention, when the program is executed on a processor. The program may for instance be a computer program that is readable by a computer or processor. The program code may then for instance be computer program code. The program may for instance be distributed via a network, such as for instance the Internet. The program may for instance be stored on a tangible readable medium, for instance a computer-readable or processor-readable medium. The readable medium may for instance be embodied as an electric, magnetic, electro-magnetic, optic or other storage medium, and may either be a removable

medium or a medium that is fixedly installed in an apparatus or device.

According to a fourth aspect of the present invention,
5 further a readable storage medium encoded with instructions that, when executed by a processor, perform the method according to the second aspect of the present invention is disclosed.

10 The readable storage medium may for instance be a computer-readable or processor-readable storage medium. It may be embodied as an electric, magnetic, electro-magnetic, optic or other storage medium, and may either be a removable storage medium or a storage medium that is
15 fixedly installed in an apparatus or device.

According to a fifth aspect of the present invention, further a program is disclosed which causes a processor to determine whether a gain should be applied to at least
20 one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically
25 encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio
30 signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair.

The program may for instance be a computer program that is readable and/or executable by a computer or processor. The program may for instance be a computer program with
5 computer program code. The program may be stored on a tangible readable medium, for instance a computer-readable or processor-readable medium. The readable medium may for instance be embodied as an electric, magnetic, electro-magnetic, optic or other storage
10 medium, and may either be a removable medium or a medium that is fixedly installed in an apparatus or device.

The term auditory spatial image may be used to refer to the listener's spatial perception of components of an
15 audio signal.

Human listeners are capable of perceiving an auditory spatial image by interpreting differences of the signals received by the left and by the right ear. Among these
20 differences are the interaural level difference (ILD) and the interaural time difference (ITD). Diffraction and reflection may be caused by at least some of the listener's body parts, thus affecting the audio signal on its way from a sound source to the listener's eardrum. A
25 head related transfer function (HRTF) describes the transform an audio signal may be subject to due to these diffraction and reflection phenomena. It is direction-dependent. Hence, being used to the listener's individual HRTF, the listener's brain can derive directional cues
30 from the interaural signal difference caused by the HRTF.

Many audio signal reproduction technologies aim at obtaining an auditory spatial image of the reproduced audio signal that resembles the auditory spatial image of the original audio signal as closely as possible.

5 Multichannel audio signals exist. An exemplary multichannel audio signal is a stereo audio signal. A stereo audio signal comprises a first channel audio signal and a second channel audio signal. The first channel and the second channel are often referred to as a
10 left channel and a right channel, respectively. A stereo audio signal is usually intended to be reproduced via a set of speakers, wherein for each channel there is at least one speaker provided. In accordance with the naming of the channels, these two speakers can be called a left
15 speaker and a right speaker, respectively. The left and the right speaker can be arranged with a certain offset to one another. A user of the audio signal reproduction system comprising said left and said right speaker, who may also be referred to as a listener or an auditor, may
20 then localize whether a certain component of a replayed audio signal is provided through the left or through the right speaker. Of course, the audio signal may also be provided through both speakers, possibly with a different sound level for each speaker, so that the listener does
25 not only perceive said signal component exclusively on one side. In consequence, the listener can perceive an auditory spatial image. In the case of two channels being provided, the auditory spatial image the listener perceives may also be called a stereo image.

30

For audio reproduction by means of speakers or headphones, the audio signal provided to the speakers or

headphones may, among other parameters, exhibit a certain inter-channel level difference (ICLD) and a certain inter-channel time difference (ICTD) and a certain inter-channel correlation (ICC). If reproduced by means of headphones, the ICLD and the ICTD may correspond to the ILD and to the ITD, respectively. In loudspeaker reproduction, the ICLD and the ICTD may only partially determine the ILD and the ITD, respectively. It is to be understood that the above explanations represent only some of the factors contributing to the complex system of human sound localization. Furthermore, it is readily clear to a skilled person that recording or encoding audio signals may involve using much more elaborate spatial information capturing or encoding techniques.

15

Using more than two channels may enable an even better spatial reproduction of an audio signal. For instance, four channel audio systems exist. A set of four loudspeakers, for example a front right loudspeaker, a front left loudspeaker, a rear right loudspeaker and a rear left loudspeaker, may be provided for audio signal reproduction. Five channel audio signal reproduction systems often comprise a fifth loudspeaker arranged in between the front right loudspeaker and the front left loudspeaker. It is noted that an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame may be formed by selecting two channels of a plurality of provided channels. For instance, if five channels are provided, the left and right front channels may form one channel pair and the left and right rear channels form another channel pair. It is also possible that the front left and

20
25
30

rear left channels form a channel pair and the front right and rear right channels form another channel pair. In the following, exemplary embodiments of the present invention are explained with respect to a scenario in
5 which only two channels are provided. It is to be understood that, as stated above, exemplary embodiments according to all aspects of the present invention may be applied in scenarios in which more than two channels are provided by forming at least one channel pair, each
10 channel pair comprising two selected channels.

In order to reduce the bitrate required by an audio signal, various audio signal coding approaches exist. In the field of audio coding, the auditory spatial image a
15 listener will perceive if provided with the decoded audio signal that has been obtained by decoding the previously encoded audio signal may be referred to as a reconstructed auditory spatial image. Again, it is often desired to obtain a reconstructed auditory spatial image
20 that closely resembles the auditory spatial image of the original (unmodified and uncoded) audio signal. On the other hand, in some cases it may be desirable to modify the perceived auditory spatial image to be different from the original auditory spatial image, for example by
25 introducing a wider or narrower auditory spatial image than the one present in the original audio signal.

Parametric audio coding may be thought of as generating a downmix signal from a multichannel audio signal and
30 providing spatial extension information that allows reconstruction of the multichannel audio signal from the downmix signal. A downmix signal is generated from a

multichannel audio signal in such a way that M input channels are used to generate N downmix channels, with $N < M$. As an example, a single-channel downmix signal - a mono signal - is generated from a stereo audio signal, implying $N=1$ and $M=2$. In the course of audio signal encoding, the audio signal may be divided into a sequence of frames, each frame comprising a limited number of samples of the audio signal. In parametric audio coding, an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame may be encoded as a single mono audio signal frame and spatial extension information. One may say that the parametrically encoded audio signal represents the audio signal frame pair.

15

The spatial extension information may contain parameters such as ICLDs, ICTDs and ICCs, as well as other parameters. Any parametric audio coding technology may be used in the scope of the present invention. Exemplary parametric audio coding technologies comprise Binaural Cue Coding (BCC) and Spatial Audio Coding (SAC). An advantage of parametric audio coding is that bitrate requirements can often be reduced significantly. Instead of separately encoding a full set of input channel signals, a smaller number of signals and the spatial extension information have to be provided for transmission or storage. For example instead of two audio channel signals only a mono signal, i.e. a single channel signal, and the spatial extension information has to be provided for transmission or storage. The bitrate required for the spatial extension information is usually smaller than the bitrate required for transmitting

30

another mono audio signal. One may therefore say that parametric audio coding exploits channel interrelations to reduce the bitrate required by the parametrically encoded audio signal compared to the bitrate required by the original multichannel audio signal.

Any pertinent mono codec may be used for mono signal coding. Among them are Advanced Audio Coding (AAC), Advanced Audio Coding with spectral band replication (AAC+) and the International Telecommunication Union (ITU) G.718 mono codec, to give only some examples of mono codecs.

Decoding a parametrically encoded audio signal may involve decoding the mono signal. Exploiting the spatial extension information, the mono signal can serve as a basis for reconstructing a first channel audio signal and a second channel audio signal. The reconstructed first channel audio signal frame may be a representation of the first channel audio signal frame, while the reconstructed second channel audio signal frame may be a representation of the second channel audio signal frame. Hence, one may say that the reconstructed first channel audio signal frame and the reconstructed second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame.

An audio signal frame may be divided into a plurality of frequency subbands. In the frequency domain, a signal is represented by a set of frequency components. A frequency

subband of the signal may then be thought of as comprising each frequency component falling within an upper frequency bound and a lower frequency bound limiting the respective frequency subband. The frequency subbands, i.e. their bounds, may be identical for the reconstructed first channel audio signal frame and the reconstructed second channel audio signal frame. Moreover, they may also be identical to the frequency subbands of the audio signal frame pair.

10

Exemplary embodiments according to all aspects of the present invention involve determination whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame.

15

One possibility to apply a gain to at least one frequency subband of a reconstructed audio signal frame is to multiply each frequency component of the respective frequency subband by a gain factor. Thus, the listener's perception of an audio signal can be modified by applying a gain to a certain frequency band. In particular, the auditory spatial image perceived by the listener may be affected by applying a gain to a frequency subband of either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame.

20

25

30

A listener's perception may exhibit different characteristics in different frequency subbands. In consequence, it can be relevant to which specific
5 frequency subband or frequency subbands a gain is applied. In addition, it can also be relevant whether a gain is applied to a frequency subband of a reconstructed first channel audio signal frame or to a frequency
10 subband of a reconstructed second channel audio signal frame. Audio signal components falling within the subband of the reconstructed audio signal frame to which a gain is applied may be perceived more prominently in the respective channel than audio signal components falling within the subband of the reconstructed audio signal
15 frame of the other channel.

According to all aspects of the present invention, the decision whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio
20 signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame is based on a first parameter and a set of second parameters.

The first parameter may be indicative of a long-term
25 estimate of a perceived direction of the audio signal frame pair. For instance, the first parameter may be indicative of a perceived direction of the (current) audio signal frame pair and at least one preceding audio signal frame pair. No limitations pertain to the first
30 parameter as long as it is indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair. It is readily clear

that different listeners are likely to have a different audio signal perception. Hence, there is probably not a single parameter that can exactly describe the perceived direction of an audio signal for all possible listeners.

5 Moreover, with a variety of parameters contributing to the auditory spatial image a listener perceives and complex interrelations among these parameters, a model describing human audio signal localization is likely to be afflicted with at least some amount of inaccuracy. In

10 consequence, various parameters can be determined that are indicative of a perceived direction of an audio signal, while it may be difficult to determine a parameter exactly representing said direction. In this respect, other examples for directional sensing of sounds

15 may include techniques based on subspace decomposition of a covariance matrix of received signals; ESPRIT and MUSIC. The former is introduced in „A. Paulraj and R. R. Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational Invariance Techniques," IEEE Trans. Acoust.,

20 Speech, Signal Processing, vol.37, no.7, pp.984-995, 1989." And the latter in „R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," IEEE Trans. Antennas and Propagation, vol.34, no.3, 1986."

25 According to all aspects of the present invention, the first parameter is indicative of a perceived direction of the audio signal frame pair and at least one preceding frame pair. In exemplary embodiments according to all aspects of the present invention, this may be understood

30 as taking into account not only the perceived direction of a current audio signal frame pair but also considering the audio signal history. In consequence, the first

parameter may change more slowly than, for instance, a parameter indicative of a perceived direction of the current audio signal frame pair only. It is noted that according to all aspects of the present invention, any

5 number of preceding audio signal frames may be taken into account. Of course, if the current audio signal frame pair does not have a predecessor, a preceding audio signal frame pair cannot be taken into account.

10 Each second parameter of the set of second parameters may be indicative of a short-term estimate of a perceived direction of a frequency subband (or several frequency subbands) of the audio signal frame pair. For instance, each second parameter may be indicative of a perceived

15 direction of a frequency subband of the audio signal frame pair. In exemplary embodiments according to all aspects of the present invention, each second parameter may be understood as a parameter that does not describe the perceived direction of the entire audio signal frame

20 pair, but of one or more specific frequency subbands of the audio signal frame pair. In case of one specific subband, one may also call each second parameter a frequency subband specific parameter. Again, as already explained with regard to the first parameter indicative

25 of a perceived direction of the audio signal frame pair, different parameters indicative of a perceived direction of a frequency subband of the audio signal frame pair can be employed. The set of second parameters may comprise at least two second parameters respectively indicative of a

30 perceived direction of a frequency subband of the audio signal frame pair but may also comprise more than two second parameters. If the set of second parameters is

only derived from the current audio signal frame pair and not also from a preceding audio signal frame pair, it may not only be thought of as comprising frequency subband specific information but it may also change more quickly
5 than the first parameter for a sequence of audio signal frame pairs.

It is noted that according to all aspects of the present invention, any suitable processing of the first parameter
10 and the set of second parameters may be employed in the course of determining whether a gain should be applied to at least one frequency subband of the reconstructed first channel audio signal frame or to at least one frequency subband of the reconstructed second channel audio signal
15 frame. Any condition that has to be met in order to cause determination to yield a positive result can be imposed on the first parameter and the set of second parameters.

According to all aspects of the present invention, no
20 restrictions regarding the location at which it is determined whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame
25 apply. For example, determination may be performed at a parametric audio encoder or at a parametric audio decoder. It is also possible that it takes place at a separate module neither forming part of an encoder nor of a decoder, such as for instance a network element located
30 in a transmission path between an encoder and a decoder, a separate module co-located with the encoder, or a separate module co-located with the decoder.

The auditory spatial image a listener perceives when provided with the reconstructed first channel audio signal frame and the reconstructed second channel audio signal frame, which will in the following be referred to as the reconstructed auditory spatial image, may differ from the auditory spatial image the listener would perceive if provided directly with the first channel audio signal frame and the second channel audio signal frame. The latter auditory spatial image will be in the following referred to as the original auditory spatial image.

One reason for the occurrence of a difference between the reconstructed auditory spatial image and the original auditory spatial image may be that parametric coding may possibly not allow exact reconstruction of the auditory spatial image. This may be, for instance, due to the spatial extension information only representing some but not every aspect of the interrelations of the first channel audio signal frame and the second channel audio signal frame. Coarse quantization of the spatial extension information may also adversely affect the reconstructed first channel audio signal frame and the reconstructed second channel audio signal frame with regard to the auditory spatial image.

The reconstructed auditory spatial image may be perceived by the listeners as a being different from the desired auditory spatial image. The desired reconstructed auditory spatial image may be the original auditory spatial image or for example a wider or a narrower

auditory spatial image than the original one. Exemplary embodiments according to all aspects of the present invention can be employed to determine whether auditory spatial image width modification should be applied to the reconstructed auditory spatial image. As an example, the reconstructed auditory spatial image may be perceived as being narrow compared to the original auditory spatial image, and it may be desirable to apply auditory spatial image widening to the reconstructed auditory spatial image. Applying a gain to at least one frequency subband of the reconstructed first channel audio signal frame or to at least one frequency subband of the reconstructed second channel audio signal frame may yield a widened auditory spatial image, for example by means of a thereby obtained subband and channel specific audio signal amplification. Due to auditory spatial image widening, even a widened reconstructed auditory spatial image obtained from a parametrically encoded low bitrate audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame may closely resemble the original auditory spatial image.

However, so as to attain a desired auditory spatial image, for example the one closely resembling the original auditory spatial image, well targeted auditory spatial image width modification can be necessary. In order to minimize the risk of adversely affecting the auditory spatial image, exemplary embodiments according to all aspects of the present invention take into account parameters indicative of a perceived direction of the audio signal frame pair, of a preceding audio signal

frame pair and also of a frequency subband of the audio signal frame pair.

5 Thereby, a hypothetical listener's directional audio signal perception can serve as a criterion whether a gain should be applied to at least one frequency subband of the reconstructed first channel audio signal frame or to at least one frequency subband of the reconstructed second channel audio signal frame. An advantage of this approach
10 can be that an auditory spatial image width modification result can be obtained that causes the listener's perception of the modified reconstructed auditory spatial image to closely resemble that of the original auditory spatial image. Also, the determination criteria used may
15 help applying width modification only if it is beneficial. Only applying a gain in these cases can also have the advantage of reducing the signal processing load by potentially omitting gain application if the criteria are not met.

20

 According to exemplary embodiments of the first apparatus according to the first aspect of the present invention, the first apparatus comprises a processor configured to calculate the first parameter indicative of a perceived
25 direction of the audio signal frame pair and the set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair. According to exemplary embodiments of the second apparatus according to the first aspect of the
30 present invention, the second apparatus further comprises means for calculating the first parameter indicative of a perceived direction of the audio signal frame pair and

the set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair. These means may for instance be embodied as a processor configured to

5 calculate the first parameter indicative of a perceived direction of the audio signal frame pair and the set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair, but are not limited thereto.

10

An advantage of these embodiments can be that it may then not be necessary to transmit the first parameter and the set of second parameters from another apparatus or device that is configured to determine the first parameter and

15 the set of second parameters to the first apparatus or the second apparatus. The processor configured to calculate the first parameter indicative of a perceived direction of the audio signal frame pair and the set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio

20 signal frame pair, or the corresponding means do not necessarily have to be separate entities. For example, the processor configured to calculate the first parameter and the set of second parameters may form part of the

25 processor configured to determine whether a gain should be applied to a frequency subband of a reconstructed first channel audio signal frame or to a frequency subband of a reconstructed second channel audio signal frame.

30

According to exemplary embodiments of the method according to the second aspect of the present invention,

the method comprises calculating the first parameter indicative of a perceived direction of the audio signal frame pair and the set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair. This feature also pertains to exemplary embodiments of the program according to the third aspect of the present invention and to the readable storage medium according to the fourth aspect of the present invention.

10

According to exemplary embodiments of the program according to the fifth aspect of the present invention, the program causes a processor to calculate the first parameter indicative of a perceived direction of the audio signal frame pair and the set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair.

15

According to exemplary embodiments of the first apparatus according to the first aspect of the present invention, the apparatus forms part of a parametric audio encoder. According to exemplary embodiments of the second apparatus according to the first aspect of the present invention, the apparatus forms part of means for parametric audio encoding. The means for parametric audio coding may for instance be embodied as a parametric audio encoder, but are not limited thereto.

20

25

According to exemplary embodiments of the method according to the second aspect of the present invention, determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel

30

audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame is performed on a parametric audio encoder side. This feature also pertains to exemplary embodiments of the program according to the third aspect of the present invention and to the readable storage medium according to the fourth aspect of the present invention.

According to exemplary embodiments of the program according to the fifth aspect of the present invention, the program causes a processor forming part of a parametric audio encoder to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame.

For parametric audio encoding, a parametric audio encoder can be provided with an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame. Thus, information that may be necessary for calculating a first parameter indicative of a perceived direction of the audio signal frame pair and a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair, may be available at the parametric audio encoder. The first parameter and the set of second parameters may then be computed on the parametric audio encoder side, too. This approach can render it unnecessary to forward these parameters to another module serving for determining whether a gain should be applied to at least one

frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame. It may suffice to generate a flag indicating the determination
5 result and to provide it to an entity configured to carry out the actual application of the gain if permitted by the flag value.

According to exemplary embodiments according to all
10 aspects of the present invention, the first parameter and the second parameters are at least partially obtainable from an assumed or known position of a listener relative to a first loudspeaker to which the first channel audio signal frame is assigned and relative to a second
15 loudspeaker to which the second channel audio signal frame is assigned.

For instance and as already elucidated above, for stereo audio reproduction, i.e. two channel audio reproduction,
20 two speakers referred to as a left speaker and as a right speaker are used. First channel audio signal frames, for example left channel audio signal frames, are replayed by means of the left loudspeaker, while second channel audio signal frames, i.e. right channel audio signal frames,
25 are replayed by means of the right loudspeaker. Put differently, the first channel audio signal frames are assigned to a first loudspeaker and the second channel audio signal frames are assigned to a second loudspeaker.

30 A listener's position can be expressed relative to the first loudspeaker and the second loudspeaker. It is possible that the listener's perceived auditory spatial

image depends on his position relative to the first
loudspeaker and to the second loudspeaker. As an example,
being located closer to the first loudspeaker than to the
second loudspeaker, audio signals emitted from the first
5 loudspeaker may reach the listener with a smaller time
difference and a higher sound intensity than audio
signals emitted from the second speaker.

According to exemplary embodiments according to all
10 aspects of the present invention, the listener's position
relative to the first loudspeaker and the second
loudspeaker can be assumed, for instance based on one or
more pre-defined assumptions. To this end, it can be for
instance assumed that the listener is positioned with the
15 same distance to both the first and the second
loudspeaker. Moreover, it may be assumed that the
distance between the first loudspeaker and the second
loudspeaker corresponds to the distance from the
listener's position to each of the loudspeakers. Thus,
20 the listener's position, the position of the first
loudspeaker and the position of the second loudspeaker
form the vertices of an equilateral triangle.

It is also possible, that the position of a listener
25 relative to a first loudspeaker and a second loudspeaker
is known. For instance, reconstructed and possibly
widened first and second audio signal frames can be
reproduced by means of the two loudspeakers. Since the
reconstructed first channel audio signal frame can be
30 seen as a representation of the first audio signal frame
and since the reconstructed second channel audio signal
frame can be seen as a representation of the second audio

signal frame, also the first channel audio signal frame is assigned to the first loudspeaker and the second channel audio signal frame is assigned to the second loudspeaker.

5

The listener's position relative to these two loudspeakers may then be measured and provided to an entity configured to calculate the first parameter and the set of second parameters. In turn, perceived

10 direction determination of the audio signal frame pair may be adapted to the reconstructed first and second audio signal frame reproduction scenario.

By taking into account a listener's assumed or known
15 position relative to a first loudspeaker to which the first channel audio signal frame is assigned and relative to a second loudspeaker to which the second channel audio signal frame is assigned, determining whether a gain should be applied to at least one frequency subband of a
20 reconstructed first or second channel audio signal frame can aim at reproduction of an auditory spatial image that could be perceived with the underlying assumed or known listener and loudspeaker position configuration.

25 Exemplary embodiments according to all aspects of the present invention comprise that the first parameter is obtainable from the direction from the assumed or known position of the listener to the first loudspeaker weighted with the sum of the energy of the first channel
30 audio signal frame and the energy of a first channel audio signal frame of the at least one preceding audio signal frame pair, and the direction from the assumed or

known position of the listener to the second loudspeaker weighted with the sum of the energy of the second channel audio signal frame and the energy of a second channel audio signal frame of the at least one preceding audio
5 signal frame pair. Therein, in the calculation of the energies, all frequency subbands, or only a subset thereof (for instance a subset that comprises the frequency subbands that are considered important, such as for instance low frequencies, wherein this consideration
10 may depend on the sample rate applied) may be considered.

These embodiments may provide a good representation of the perceived direction of the audio signal frame pair and the at least one preceding audio signal frame pair.
15 Knowing the direction from the assumed or known position of the listener to the first loudspeaker and the direction from the assumed or known position of the listener to the second loudspeaker, the first parameter may be determined with low computational effort.

20

According to exemplary embodiments according to all aspects of the present invention, the second parameters of the set of second parameters are obtainable from the direction from the assumed or known position of the
25 listener to the first loudspeaker weighted with the first channel audio signal frame energy within the respective frequency subband, and the direction from the assumed or known position of the listener to the second loudspeaker weighted with the second channel audio signal frame
30 energy within the respective frequency subband.

Exemplary embodiments according to all aspects of the present invention comprise that the first parameter and the second parameters are indicative of a perceived direction relative to a reference direction, for example
5 relative to a first imaginary line passing through the assumed or known position of the listener, the first imaginary line being perpendicular to a second imaginary line connecting the first loudspeaker and the second loudspeaker. In other words, the first parameter can be
10 indicative of a perceived direction of an audio signal frame pair and at least one preceding audio signal frame pair relative to said first imaginary line and each second parameter of the set of second parameters can be indicative of a perceived direction of a frequency
15 subband of the audio signal frame pair relative to said first imaginary line.

By describing the above perceived directions relative to the first imaginary line, arranging the perceived
20 directions in two groups, wherein the first group comprises perceived directions having a positive angular offset relative to the first imaginary line and the second group comprises perceived directions having a negative angular offset relative to the first imaginary
25 line, is enabled. In exemplary embodiments according to all aspects of the present invention, the angular offset relative to the first imaginary line may also be thought of as an indicator for the amount of spatial effects contained in the entire audio signal frame pair and in a
30 specific subband of the audio signal frame pair, respectively. This indicator may also be considered in deciding whether a gain should be applied to a specific

subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame.

- 5 According to exemplary embodiments of the first apparatus according to the first aspect of the present invention, the processor configured to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at
- 10 least one frequency subband of a reconstructed second channel audio signal frame is configured to base the determination on averaged second parameters. According to exemplary embodiments of the second apparatus according to the first aspect of the present invention, the means
- 15 for determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame are configured to base the determination on
- 20 averaged second parameters.

- According to exemplary embodiments of the method according to the second aspect of the present invention, determining whether a gain should be applied to at least
- 25 one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame is based on averaged second parameters. This feature also pertains to exemplary embodiments of the program
- 30 according to the third aspect of the present invention and to the readable storage medium according to the fourth aspect of the present invention.

According to exemplary embodiments of the program according to the fifth aspect of the present invention, the program causes a processor to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame based on averaged second parameters.

10

Averaging may be based on any existing averaging approach. Besides of computing the mean value or the median value, averaging may also comprise assigning weights to the second parameters before summing them up.

15 The weights may control the influence of each second parameter on the averaged value. An advantage of basing determination whether a gain should be applied to at least one frequency subband of a reconstructed first or second channel audio signal frame on average second parameters can be that fast determination may be enabled. Checking if each second parameter meets a certain condition may require more time. This may be even more significant if several conditions have to be met.

25 According to exemplary embodiments of the first apparatus according to the first aspect of the present invention, the processor configured to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame is further configured to consider whether a gain has been applied to at least one

30

frequency subband of a preceding reconstructed first channel audio signal frame or a preceding reconstructed second channel audio signal frame in determining whether a gain should be applied to at least one frequency

- 5 subband of the reconstructed first channel audio signal frame or to at least one frequency subband of the reconstructed second channel audio signal frame.

According to exemplary embodiments of the second

- 10 apparatus according to the first aspect of the present invention, the means for determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second
- 15 channel audio signal frame are configured to consider whether a gain has been applied to at least one frequency subband of a preceding reconstructed first channel audio signal frame or a preceding reconstructed second channel audio signal frame in determining whether a gain should
- 20 be applied to at least one frequency subband of the reconstructed first channel audio signal frame or to at least one frequency subband of the reconstructed second channel audio signal frame.

- 25 According to exemplary embodiments of the method according to the second aspect of the present invention, determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband
- 30 of a reconstructed second channel audio signal frame involves considering whether a gain has been applied to at least one frequency subband of a preceding

reconstructed first channel audio signal frame or a preceding reconstructed second channel audio signal frame. This feature also pertains to exemplary embodiments of the program according to the third aspect
5 of the present invention and to the readable storage medium according to the fourth aspect of the present invention.

According to exemplary embodiments of the program
10 according to the fifth aspect of the present invention, the program causes a processor to consider whether a gain has been applied to at least one frequency subband of a preceding reconstructed first channel audio signal frame or a preceding reconstructed second channel audio signal
15 frame in determining whether a gain should be applied to at least one frequency subband of the reconstructed first channel audio signal frame or to at least one frequency subband of the reconstructed second channel audio signal frame.

20

A listener's perception of the auditory spatial images of an audio signal he is provided with at a certain time may depend on the history of the audio signal. For example, due to auditory spatial image widening having been
25 applied to a comparatively large number of preceding reconstructed audio signal frames, the listener may continue to perceive a widened auditory spatial image even no gain is applied to any of the subbands of the currently perceived reconstructed audio signal frames.
30 Thus, it may not be necessary to apply a gain to any of the subbands of the currently perceived reconstructed audio signal frames. On the other hand, scenarios are

imaginable in which neither the first parameter nor the set of second parameters would cause that it is determined that a gain should be applied to at least one frequency subband of a reconstructed first or second channel audio signal frame. However, if a gain has been applied to a subband of a preceding reconstructed audio signal frame it may be advisable to also apply a gain to at least one subband of a current reconstructed first or second channel audio signal frame so as not to cause sudden changes in the perceived auditory spatial image. An advantage of the embodiments of the present invention currently discussed can thus be that the history of the audio signal is taken into account, thereby allowing adequate use of auditory spatial image width modification.

In further exemplary embodiments according to all aspects of the present invention, the frequency subbands of the audio signal frame pair are non-uniform frequency subbands corresponding to the human auditory filters. The human auditory system can be modeled as a system comprising a plurality of auditory filters. The bandwidths of these auditory filters increase with increasing audio signal frequencies. An example of non-uniform frequency subbands corresponding to the human auditory filters are equivalent rectangular bandwidth (ERB) frequency subbands. By splitting the audio signal frame pair into non-uniform frequency subbands corresponding to the human auditory filters, determination whether a gain should be applied to a frequency subband of a reconstructed first or second channel audio signal frame can take into account

psychoacoustic parameters of the human auditory system, thus possibly attaining enhanced modified auditory spatial images.

5 According to exemplary embodiments of the first apparatus according to the first aspect of the present invention, the apparatus comprises a processor configured to determine for each frequency subband whether a gain should be applied to either the reconstructed first
10 channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal
15 frame is based on the second parameter associated with the respective frequency subband. This processor may either be a separate module or it may form part of the processor configured to determine whether a gain should be applied to a frequency subband of a reconstructed
20 first or second channel audio signal frame. According to exemplary embodiments of the second apparatus according to the first aspect of the present invention, the apparatus comprises means for determining for each frequency subband whether a gain should be applied to
25 either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second
30 channel audio signal frame is based on the second parameter associated with the respective frequency subband. These means may for example be embodied as a

processor configured to determine for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein
5 discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the second parameter associated with the respective frequency subband, but the
10 means are not limited thereto.

According to exemplary embodiments of the method according to the second aspect of the present invention, the method comprises determining for each frequency
15 subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel
20 audio signal frame or of the reconstructed second channel audio signal frame is based on the second parameter associated with the respective frequency subband. This feature also pertains to exemplary embodiments of the program according to the third aspect of the present
25 invention and to the readable storage medium according to the fourth aspect of the present invention.

According to exemplary embodiments of the program according to the fifth aspect of the present invention,
30 the program causes a processor to determine for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame

or the reconstructed second channel audio signal frame,
wherein discrimination whether the gain should be applied
to the respective subband of the reconstructed first
channel audio signal frame or of the reconstructed second
5 channel audio signal frame is based on the second
parameter associated with the respective frequency
subband.

These embodiments may for instance enable applying a gain
10 to the reconstructed audio signal frame associated with
the audio signal frame of the audio signal frame pair
that is assigned to the loudspeaker for which the
direction from an assumed or known position of a listener
to said loudspeaker is closer to the perceived direction
15 in the respective frequency subband as indicated by the
second parameter of said frequency subband.

As an example, the perceived direction as indicated by
the second parameter of a certain frequency subband of an
20 audio signal frame pair can be 110° . The direction is
measured from an assumed or known position of a listener
relative to a pair of speakers. For instance and as
already elucidated above, for stereo audio reproduction,
i.e. two channel audio reproduction, two speakers
25 referred to as a left speaker and as a right speaker are
used. First channel audio signal frames, for example left
channel audio signal frames, are replayed by means of the
left loudspeaker, while second channel audio signal
frames, i.e. right channel audio signal frames, are
30 replayed by means of the right loudspeaker. Put
differently, the first channel audio signal frames are

assigned to a first loudspeaker and the second channel audio signal frames are assigned to a second loudspeaker.

5 An angle of 90° corresponds to the direction of a first imaginary line perpendicular to a second imaginary line connecting the left and the right loudspeaker, the first imaginary line passing through the assumed or known position of the listener. Therefore, directions describable by an angle greater than 90° are closer to
10 the direction from the assumed or known position of the listener to the left speaker, while directions describable by an angle of less than 90° are closer to the direction from the assumed or known position of the listener to the right speaker.

15

The perceived direction of 110° is thus closer to the direction from the assumed or known position of the listener to the left loudspeaker. According to the exemplary embodiments of the present invention currently
20 discussed, for the specific subband for which the perceived direction as indicated by the second parameter of the said subband is 110° it is determined to apply a gain to the subband of the reconstructed first channel audio signal frame. Said frame is associated with the
25 left, i.e. first, channel audio signal frame of the audio signal frame pair that is assigned to the left loudspeaker because it may be thought of as a representation of the left channel audio signal frame.

30 Hence, the embodiments of the present invention currently discussed may enable emphasizing the content of the reconstructed audio signal frame contributing more

significantly to the spatial auditory effects in the respective frequency subband. The modified auditory spatial image may then be close to the desired auditory spatial image, it may for example closely resemble the original auditory spatial image. Of course it does not have to be determined for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame if it has previously been determined that no gain should be applied at all.

According to exemplary embodiments of the first apparatus according to the first aspect of the present invention, the apparatus comprises a processor configured to determine for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the inter-channel level difference in said frequency subband of the reconstructed first channel audio signal frame and on the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame. This processor may either be a separate module or it may form part of the processor configured to determine whether a gain should be applied to a frequency subband of a reconstructed first channel audio signal frame or to a frequency subband of a reconstructed second channel audio signal frame.

According to exemplary embodiments of the second apparatus according to the first aspect of the present invention, the apparatus comprises means for determining
5 for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the
10 reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the inter-channel level difference in said frequency subband of the reconstructed first channel audio signal frame and on the inter-channel level difference in said
15 frequency subband of the reconstructed second channel audio signal frame. These means may for example be embodied as a processor configured to determine for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame
20 or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the inter-channel
25 level difference in said frequency subband of the reconstructed first channel audio signal frame and on the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame, but the means are not limited thereto.

30

According to exemplary embodiments of the method according to the second aspect of the present invention,

the method comprises determining for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein
5 discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the inter-channel level difference in said frequency subband of the reconstructed
10 first channel audio signal frame and on the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame. This feature also pertains to exemplary embodiments of the program according to the third aspect of the present
15 invention and to the readable storage medium according to the fourth aspect of the present invention.

According to exemplary embodiments of the program according to the fifth aspect of the present invention,
20 the program causes a processor to determine for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied
25 to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the inter-channel level difference in said frequency subband of the reconstructed first channel audio signal frame and on the
30 inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame.

These embodiments may be advantageous because the inter-channel level differences may form part of the spatial extension information that may be needed in order to establish these differences between the reconstructed
5 first channel audio signal frame and the reconstructed second channel audio signal frame. It may be possible to determine for each frequency subband whether a gain should be applied to either the reconstructed first or second channel audio signal frame at the entity
10 configured to perform reconstruction of the reconstructed first and second channel audio signal frame. To this end, it may then not be necessary to provide said entity with additional parameters. Also, the gain can be applied to at least one frequency subband of the reconstructed first
15 or second channel audio signal frame at said entity in exemplary embodiments according to all aspects of the present invention.

According to exemplary embodiments according to all
20 aspects of the present invention, the gain is obtainable from a reciprocal average value of frequency subband specific indicator ratios, wherein each indicator ratio is the ratio of an indicator of the maximum of the energy in the frequency subband of the first channel audio
25 signal frame and the energy in said frequency subband of the second channel audio signal frame, and an indicator of the minimum of the energy in the frequency subband of the first channel audio signal frame and the energy in said frequency subband of the second channel audio signal
30 frame.

Thus, due to the computation of the reciprocal average value of frequency subband specific indicator ratios, it may be achieved that the gain is comparatively small if the subband specific indicator ratios are large. Large
5 subband specific indicator ratios may occur if the indicator of the maximum of the energy in the frequency subband of the first channel audio signal frame and the energy in said frequency subband of the second channel audio signal frame is much greater than the indicator of
10 the minimum of the energy in the frequency subband of the first channel audio signal frame and the energy in said frequency subband of the second channel audio signal frame. In other words, large subband specific indicator ratios may occur if the energies of the first channel
15 audio signal frame and the second channel audio signal frame in the respective subband differ significantly. In consequence, applying the gain may yield a widened auditory spatial image, in which subtle inter-channel differences are emphasized more than prominent inter-
20 channel differences. Thus the widened auditory spatial image may closely resemble the original auditory spatial image.

Other exemplary embodiments according to all aspects of
25 the present invention comprise that the gain is obtainable from a reciprocal average value of frequency subband specific indicator ratios, wherein each indicator ratio is the ratio of an indicator of the maximum of the inter-channel level difference in the frequency subband
30 of the reconstructed first channel audio signal frame and the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal

frame, and an indicator of the minimum of the inter-channel level difference in the frequency subband of the reconstructed first channel audio signal frame and the inter-channel level difference in said frequency subband
5 of the reconstructed second channel audio signal frame.

These embodiments may be advantageous because the inter-channel level differences may form part of the spatial extension information that may be needed in order to
10 establish these differences between the reconstructed first channel audio signal frame and the reconstructed second channel audio signal frame. The gain may then be computed at the entity configured to perform reconstruction of the reconstructed first and second
15 channel audio signal frame. To this end, it may then not be necessary to provide said entity with additional parameters for gain computation. Also, the gain can be applied to at least one frequency subband of the reconstructed first or second channel audio signal frame
20 at said entity in exemplary embodiments according to all aspects of the present invention.

Exemplary embodiments according to all aspects of the present invention comprise that the gain is a quantized
25 value. Gain quantization may enable representation of the gain by a reduced number of bits. Thus, transmitting the gain to an entity configured to apply it to at least one frequency subband of a reconstructed first or second channel audio signal frame may require a reduced
30 bandwidth.

According to exemplary embodiments according to all aspects of the present invention, the gain is obtainable from a lookup table. An advantage of these embodiments can be that instead of transmitting the gain to an entity
5 configured to apply it to at least one frequency subband of a reconstructed first or second channel audio signal frame, it may suffice to merely transmit a lookup table index, which can possibly be represented by fewer bits than the quantized gain itself.

10

The features of the present invention and of its exemplary embodiments as presented above shall also be understood to be disclosed in all possible combinations with each other.

15

It is to be noted that the above description of embodiments of the present invention is to be understood to be merely exemplary and non-limiting.

20 Further aspects of the invention will be apparent from and elucidated with reference to the detailed description presented hereinafter.

BRIEF DESCRIPTION OF THE FIGURES

25 In the figures show:

Fig. 1: An exemplary illustration of a method for parametric audio coding;

30 Fig. 2: A more detailed representation of method step 101 of Fig. 1;

- Fig. 3: A more detailed representation of method step 102 of Fig. 1;
- 5 Fig. 4: A schematic exemplary illustration of the assumed or known position of a listener relative to a first loudspeaker to which a first channel audio signal frame is assigned and relative to a second loudspeaker to which a second channel audio signal frame is assigned;
- 10 Fig. 5: A flowchart illustrating an exemplary embodiment of a first method according to the second aspect of the present invention;
- 15 Fig. 6: A flowchart illustrating an exemplary embodiment of a second method according to the second aspect of the present invention;
- 20 Fig. 7: A more detailed illustration of method step 112 of Fig. 6;
- 25 Fig. 8: A flowchart exemplarily illustrating decoding of a bitstream generated according to the method of Fig. 6 including gain application;
- 30 Fig. 9: A schematic illustration of a first exemplary embodiment of an apparatus according to the first aspect of the present invention;
- Fig. 10: A schematic illustration of a second exemplary embodiment of an apparatus according to the first aspect of the present invention;

Fig. 11: A schematic illustration of an exemplary embodiment of a readable medium according to the fourth aspect of the present invention.

5

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS OF THE INVENTION

In the following detailed description, exemplary embodiments of the present invention will be described.

10

An input audio signal comprising two channels, namely L_i and R_i , is provided to method step 101. Method step 101 comprises parametrically encoding the input audio signals L_i and R_i . As an output, method step 101 delivers a

15 parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame. In step 102, the parametrically encoded audio signal is decoded. In decoding the parametrically encoded audio signal frame pair, a reconstructed first channel audio signal frame and a reconstructed second channel audio signal frame can be obtained.

20

Fig. 2 is a more detailed representation of method step 25 101 of Fig. 1.

The input signals L_i and R_i are divided into frames in step 103. Successive frames may overlap in time, for example half of their duration. As a result, the first 30 channel audio signal frame L_i and the second channel audio

signal frame R_i are obtained. They comprise samples $\bar{t}_L$ and $\bar{t}_R$, respectively.

In step 104 windowing is applied to the audio signal frames L_i and R_i . Windowing may serve for suppressing framing induced artefacts in time-to-frequency transformed representations of L_i and R_i . Any pertinent windowing function $\bar{w}^T$ can be employed. An exemplary windowing function is the sinusoidal window given by:

$$w(i) = \sin\left((i+0.5) \cdot \frac{\pi}{N}\right), \quad 0 \leq i < N, \quad (1)$$

wherein N is the number of samples in each frame L_i and R_i . As a result, the windowed first channel audio signal frame L_w and the windowed second channel audio signal frame R_w are obtained.

In step 105 the windowed audio signal frames L_w and R_w are transformed to the frequency domain. To this end, any transform TF that provides complex valued output may be used. For example, Discrete Fourier Transform (DFT), Modified Discrete Cosine Transform (MDCT), Modified Discrete Sine Transform (MDST) or Quadrature Mirror Filtering (QMF) may be used. The transformed windowed audio signal frames are referred to as L_f and R_f . The transformed windowed audio signal frames comprise the samples $\bar{f}_L$ and $\bar{f}_R$, respectively. How they are obtained from the samples $\bar{t}_L$ and $\bar{t}_R$ can be described by the following equations:

$$\bar{f}_L = TF(\bar{w}^T \cdot \bar{t}_L), \quad \bar{f}_R = TF(\bar{w}^T \cdot \bar{t}_R). \quad (2)$$

In step 106 the audio signal frames L_f and R_f are transformed to a downmix signal, which in this example embodiment is a mono signal M_f .

$$M_f = 0.5 \cdot (L_f + R_f) \quad (3)$$

5

Step 107 comprises encoding the mono signal M_f . Any pertinent mono codec, such as Advanced Audio Coding (AAC), Advanced Audio Coding with spectral band replication (AAC+) and the International
10 Telecommunication Union (ITU) G.718 mono codec, may be employed.

In step 108, spatial extension information is derived from the first channel audio signal frame L_f and the

15 second channel audio signal frame R_f . Exemplary spatial extension parameters are the inter-channel level difference (ICLD), the inter-channel time difference (ICTD) and a the inter-channel correlation (ICC).

20 By multiplexing the encoded mono signal and the spatial extension information, a bitstream can be formed. The bitstream representative of an input frame may comprise one or more encoded frames or packets. It can be thought of as a parametrically encoded audio signal representing
25 an audio signal frame pair comprising a first channel audio signal frame L_f or L_t and a second channel audio signal frame R_f or R_t .

Fig. 3 is a more detailed representation of method step 102 of Fig. 1.

According to Fig. 3, the bitstream provided from the
5 encoding step 101 is demultiplexed in step 109. Thereby,
an encoded mono signal and spatial extension information
are obtained. In step 110 the mono signal is decoded,
yielding the decoded mono signal $\tilde{M}_f$. Step 111 comprises
applying the spatial extension information, thus
10 obtaining the reconstructed first channel audio signal
frame $\tilde{L}_f$ and the reconstructed second channel audio
signal frame $\tilde{R}_f$.

Fig. 4 schematically illustrates the assumed or known
15 position A of a listener relative to a first loudspeaker
LS to which a first channel audio signal frame L_i is
assigned and relative to a second loudspeaker RS to which
a second channel audio signal frame R_i is assigned;

20 For stereo audio reproduction, i.e. two channel audio
reproduction, two loudspeakers referred to as a left
loudspeaker LS and as a right loudspeaker RS are used.
First channel audio signal frames, for example left
channel audio signal frames L_i , are replayed by means of
25 the left loudspeaker LS, while second channel audio
signal frames, i.e. right channel audio signal frames R_i ,
are replayed by means of the right loudspeaker RS. Put
differently, the first channel audio signal frames L_i are
assigned to the first loudspeaker LS, while the second

channel audio signal frames R_i are assigned to the second loudspeaker RS.

Also reconstructed and possibly widened first and second
 5 audio signal frames $\tilde{L}_f$ and $\tilde{R}_f$ can be reproduced by means
 of the two loudspeakers LS and RS. Since the
 reconstructed first channel audio signal frame $\tilde{L}_f$ can be
 seen as a representation of the first audio signal frame
 L_i and since the reconstructed second channel audio signal
 10 frame $\tilde{R}_f$ can be seen as a representation of the second
 audio signal frame R_i , the first channel audio signal
 frame L_i is assigned to the first loudspeaker LS and the
 second channel audio signal frame R_i is assigned to the
 second loudspeaker RS in this case, too.

15 The position A can be expressed relative to the
 loudspeakers LS and RS. For instance, the distance from
 position A to the first loudspeaker LS is s_1 , while the
 distance from position A to the second loudspeaker RS is
 20 s_2 . The distances s_1 and s_2 are equal. Furthermore, the
 direction from position A to the first loudspeaker LS can
 be described by the angle θ_L and the direction from
 position A to the first loudspeaker LS can be described
 by the angle θ_R . In the present example, θ_L is 120° and
 25 θ_R is 60° . Thus, the triangle A, RS, LS is an equilateral
 triangle. The first imaginary line s_3 is perpendicular to
 the second imaginary line s_4 that connects the first
 loudspeaker LS and the second loudspeaker RS. The first
 imaginary line s_3 passes through the position A.

As an example, the listener may localize a sound component of the replayed first channel audio signal frame L_i or of the second channel audio signal frame R_i at
 5 a position P which can be described by its distance to the position A of the listener and the direction θ as seen from that position A.

In the following it is explained how a first parameter
 10 indicative of a perceived direction of an audio signal frame pair and at least one preceding audio signal frame pair may be obtained.

In a first step, the energy $e_{L_{frame_1}}$ of the first channel
 15 audio signal frame L_i and the energy $e_{R_{frame_1}}$ of the second channel audio signal frame R_i are computed. This can be done by calculating the sum of the squared absolute values of the frequency transformed samples $\bar{f}_L$ or $\bar{f}_R$ of the respective frame and then extracting the square root
 20 of this sum according to equation (4).

$$e_{L_{frame_1}} = \sqrt{\sum_{j=0}^{N-1} |\bar{f}_L(j)|^2}, \quad e_{R_{frame_1}} = \sqrt{\sum_{j=0}^{N-1} |\bar{f}_R(j)|^2} \quad (4)$$

It should be noted that, instead of considering all frequency subbands in the calculation of the energy $e_{L_{frame_1}}$ of the first channel audio signal frame L_i and the energy
 25 $e_{R_{frame_1}}$ of the second channel audio signal frame R_i , only the most important frequency bands may be considered, for instance the frequency subbands corresponding to the low

frequencies. Which frequency subbands are important may depend on the sample rate of the signal. For instance, half of the frequency subbands could be left out in case of a sample rate of 48kHz, whereas, at a sample rate of 8kHz, the frequency band coverage may only be up to 4kHz so in this case more than half of the bands may probably have to be included in the calculations.

The energies $e_{L_{frame}}$ and $e_{R_{frame}}$ of preceding frames are also taken into account according to the pseudocode given in equation (5).

$$e_{L_{frame}} = e_{L_{frame}} \cdot \beta + e_{L_{frame_1}} \cdot (1 - \beta), \quad e_{R_{frame}} = e_{R_{frame}} \cdot \beta + e_{R_{frame_1}} \cdot (1 - \beta) \quad (5)$$

According to equation (5), with each update the sum of the former value of the variable $e_{L_{frame}}$ weighted with β and the value of $e_{L_{frame_1}}$ weighted with $(1 - \beta)$ is assigned to $e_{L_{frame}}$. $e_{R_{frame}}$ is computed accordingly. As an example, β may have the value 0.95. It may thus happen that $e_{L_{frame}}$ and $e_{R_{frame}}$ change only slightly with each update, and the $e_{L_{frame}}$ and $e_{R_{frame}}$ can be considered to represent the long-term energies of the first and second channels, respectively. $e_{L_{frame}}$ and $e_{R_{frame}}$ are initialized to zero so that they can be determined even if there has not been a preceding audio signal frame pair.

In the next step, the direction θ_L from the assumed or known position A of the listener to the first loudspeaker LS is weighted with the sum of the energy $e_{L_{frame_1}}$ of the

first channel audio signal frame L_t and the energy $e_{L_{frame}}$ of a first channel audio signal frame of at least one preceding audio signal frame pair. Accordingly, the direction θ_R from the assumed or known position A of the listener to the second loudspeaker RS is weighted with the sum of the energy $e_{R_{frame_1}}$ of the first channel audio signal frame R_t and the energy $e_{R_{frame}}$ of a second channel audio signal frame of at least one preceding audio signal frame pair. Normalization with the sum of the energies $e_{L_{frame}}$ and $e_{R_{frame_1}}$ is performed.

$$\alpha_{-r_{frame}} = \frac{e_{L_{frame}} \cdot \cos(\theta_L) + e_{R_{frame}} \cdot \cos(\theta_R)}{e_{L_{frame}} + e_{R_{frame}}}, \quad \alpha_{-i_{frame}} = \frac{e_{L_{frame}} \cdot \sin(\theta_L) + e_{R_{frame}} \cdot \sin(\theta_R)}{e_{L_{frame}} + e_{R_{frame}}} \quad (6)$$

$\alpha_{-r_{frame}}$ and $\alpha_{-i_{frame}}$ can be seen as a coordinate pair representing the overall localized position of sound components within the pair of audio signal frames L_t and R_t and preceding audio signal frames. For instance, $\alpha_{-r_{frame}}$ and $\alpha_{-i_{frame}}$ can describe the position of the point P in Fig. 4. The angle or direction θ_{frame} is an indicator of the perceived direction of the audio signal frame pair L_t and R_t .

θ_{frame} can be a good representation of the perceived direction of the audio signal frame pair L_t and R_t and the at least one preceding audio signal frame pair:

$$\theta_{frame} = \angle(\alpha_{-r_{frame}}, \alpha_{-i_{frame}}). \quad (7)$$

In other words, θ_{frame} may be seen as a representation of a long-term estimate of a perceived direction of the audio signal.

- 5 Knowing the direction from the assumed or known position A of the listener to the first loudspeaker LS and the direction from the assumed or known position A of the listener to the second loudspeaker RS, the first parameter θ_{frame} may be determined with low computational
10 effort.

The difference $|\theta_{frame} - 90^\circ|$ can also be seen as a first parameter indicative of a perceived direction of an audio signal frame pair L_f and R_f and at least one preceding
15 audio signal frame pair or, in the time domain, of the pair L_t and R_t and at least one preceding audio signal frame pair.

It is indicative of a perceived direction of the audio
20 signal frame pair L_f and R_f and at least one preceding audio signal frame pair relative to a first imaginary line s_3 passing through the assumed or known position A of the listener, the first imaginary line s_3 being perpendicular to a second imaginary line s_4 connecting the
25 first loudspeaker LS and the second loudspeaker RS.

The angular offset $|\theta_{frame} - 90^\circ|$ relative to the first imaginary line s_3 may also be thought of as an indicator for the amount of spatial effects contained in the entire

audio signal frame pair L_f and R_f and at least one preceding audio signal frame pair. This indicator may also be considered in deciding whether a gain should be applied to a specific subband of the reconstructed first
5 channel audio signal frame $\tilde{L}_f$ or of the reconstructed second channel audio signal frame $\tilde{R}_f$.

In the following it is explained how a set of second parameters, each second parameter indicative of a
10 perceived direction of a frequency subband of a audio signal frame pair, can be obtained.

First, the frequency domain audio signal frames L_f and R_f are divided into m frequency subbands, wherein $0 \leq m < M$.
15 $sbOffset[m]$ is the lower bound of frequency subband m .
The frequency subbands m , i.e. their bounds, may be identical for the first channel audio signal frame L_f and the second channel audio signal frame R_f . Moreover, they may also be identical to the frequency subbands of the
20 reconstructed audio signal frame pair $\tilde{L}_f$ and $\tilde{R}_f$.

The frequency subbands m of the audio signal frame pair L_f and R_f are non-uniform frequency subbands corresponding to the human auditory filters. Namely, they
25 are equivalent rectangular bandwidth (ERB) frequency subbands. The human auditory system can be modeled as a system comprising a plurality of auditory filters. The bandwidths of these auditory filters increase with increasing audio signal frequencies. By splitting the

audio signal frame pair L_f and R_f into non-uniform frequency subbands m corresponding to the human auditory filters, determination whether a gain should be applied to a frequency subband m of the reconstructed first

5 channel audio signal frame $\tilde{L}_f$ or of the reconstructed second channel audio signal frame $\tilde{R}_f$ can take into account psychoacoustic parameters of the human auditory system, thus possibly attaining enhanced widened auditory spatial images.

10

In the next step the energies e_{L_m} in each frequency subband m of the first channel audio signal frame L_f and the energies e_{R_m} in each frequency subband m of the second channel audio signal frame R_f are determined:

$$e_{L_m} = \sqrt{\sum_{j=sbOffset[m]}^{sbOffset[m+1]-1} |\tilde{f}_L(j)|^2}, \quad e_{R_m} = \sqrt{\sum_{j=sbOffset[m]}^{sbOffset[m+1]-1} |\tilde{f}_R(j)|^2}. \quad (8)$$

15

The energies e_{L_m} and e_{R_m} can be considered to represent the short-term energies of the first and second channels in each frequency subband m , respectively.

20 Subsequently, the direction θ_L from the assumed or known position A of the listener to the first loudspeaker LS weighted with the energy e_{L_m} of the first channel audio signal frame L_f , which corresponds to the energy of said frame in the time domain L_t , is calculated for each

25 subband m . Accordingly, the direction θ_R from the assumed or known position A of the listener to the second

loudspeaker RS weighted with the energy e_{R_m} of the second channel audio signal frame R_f , which corresponds to the energy of said frame in the time domain R_t , is calculated for each subband m . Normalization with the sum of the
 5 energies e_{L_m} and e_{R_m} is performed:

$$\alpha_{-r_m} = \frac{e_{L_m} \cdot \cos(\theta_L) + e_{R_m} \cdot \cos(\theta_R)}{e_{L_m} + e_{R_m}}, \quad \alpha_{-i_m} = \frac{e_{L_m} \cdot \sin(\theta_L) + e_{R_m} \cdot \sin(\theta_R)}{e_{L_m} + e_{R_m}}. \quad (9)$$

Similar to the explanations given above with respect to the first parameter, a directional angle θ_m can be derived from α_{-r_m} and α_{-i_m} .

$$\theta_m = \angle(\alpha_{-r_m}, \alpha_{-i_m}) \quad (10)$$

10

The angles or direction angles θ_m are second parameters indicative of the perceived direction of the respective frequency subband m of the audio signal frame pair L_f and R_f , and thus of the audio signal frame pair L_t and R_t in
 15 the time domain. In other words, the angles or direction angles θ_m may be seen as indicative of a short-term estimate of a perceived direction of the respective frequency subband m of the audio signal frame pair.

20 The second parameters θ_m of the set of second parameters are obtainable from the direction from the assumed or known position of the listener A to the first loudspeaker LS weighted with the first channel audio signal frame energy e_{L_m} within the respective frequency subband m , and
 25 the direction from the assumed or known position A of the listener to the second loudspeaker RS weighted with the

second channel audio signal frame energy e_{R_m} within the respective frequency subband m .

The differences $|\theta_m - 90^\circ|$ can also be seen as a set of
5 second parameters indicative of a perceived direction of the respective frequency subband m of the audio signal frame pair L_f and R_f .

They are indicative of a perceived direction of a
10 respective frequency subband m of the audio signal frame pair L_f and R_f relative to a first imaginary line s_3 passing through the assumed or known position A of the listener, the first imaginary line s_3 being perpendicular to a second imaginary line s_4 connecting the first
15 loudspeaker LS and the second loudspeaker RS.

By describing the above perceived directions θ_m relative to the first imaginary line s_3 , arranging the perceived directions θ_m in two groups, wherein the first group
20 comprises perceived directions having a positive angular offset relative to the first imaginary s_3 line and the second group comprises perceived directions having a negative angular offset relative to the first imaginary line s_3 , is enabled. The angular offset relative to the
25 first imaginary line s_3 may also be thought of as an indicator for the amount of spatial effects contained in the respective subband m of the audio signal frame pair L_f and R_f . This indicator may also be considered in
30 subband m of the reconstructed first channel audio signal

frame $\tilde{L}_f$ or of the reconstructed second channel audio signal frame $\tilde{R}_f$.

The parameter *aveDiff* can be seen as an averaged
 5 representation of the set of second parameters θ_m :

$$aveDiff = \frac{1}{M} \cdot \sum_{m=0}^{M-1} (\theta_m - 90^\circ). \quad (11)$$

Alternatively, the computation of parameter *aveDiff* may consider only a subset of the second parameters θ_m , for example θ_m corresponding to a selected number of lowest
 10 frequency bands, instead of considering the full set of second parameters θ_m for the M frequency bands as in equation (11).

Considering *aveDiff* in determining whether a gain should be
 15 applied to at least one frequency subband m of the reconstructed first channel audio signal frame $\tilde{L}_f$ or to at least one frequency subband m of a reconstructed second channel audio signal frame $\tilde{R}_f$, may thus correspond to basing the determination on averaged second
 20 parameters.

If *aveDiff* assumes the value 0, this can be understood as indicating that no spatial effects are contained within the pair of audio signal frames L_f and R_f , i.e. that the
 25 first audio signal frame L_f and the second audio signal frame R_f are identical. On the other hand, an assumed

value of 90° indicates maximum spatial effects within the pair of audio signal frames L_f and R_f , i.e. they differ completely from each other and are not correlated.

- 5 An advantage of basing determination whether a gain should be applied to at least one frequency subband m can be that can be that fast determination may be enabled. Checking if each second parameter θ_m meets a certain condition may require more time or higher processing
- 10 power. This may be even more significant if several conditions have to be met.

Fig. 5 is a flowchart illustrating an exemplary embodiment of a first method according to the second

15 aspect of the present invention.

According to this embodiment, whether a gain has been applied to at least one frequency subband m of a preceding reconstructed first or second channel audio

20 signal frame is taken into account in determining whether a gain should be applied to at least one frequency subband m of the reconstructed first or second channel audio signal frame $\tilde{L}_f$ and $\tilde{R}_f$, respectively.

- 25 To this end, it is stored in the array `st_wideArray`, whether a gain has been applied to at least one frequency subband m of a preceding reconstructed first or second channel audio signal frame. `st_wideArray` is initialized to zero start up. It is updated after each audio signal
- 30 frame pair by performing a left shift, thereby discarding the flag indicating whether a gain has been applied to

the oldest reconstructed audio signal frame pair represented by `st_wideArray`. The flag for the current reconstructed audio signal frame pair $\tilde{L}_f$ and $\tilde{R}_f$ is stored at the rightmost position `st_wideArray[0]` in the array.

5 In the present example, eight flags are stored in `st_wideArray`.

In step 1, it is checked whether a gain has been applied to the last eight frames by comparing `st_wideArray` to the
10 binary number 11111111. It is also checked whether the conditions $aveDiff < 10^\circ$ and $|\theta_{frame} - 90^\circ| < 10^\circ$ are also met. If true, applying a gain to at least one frequency subband m of the reconstructed first channel audio signal frame $\tilde{L}_f$ or of the reconstructed second channel audio signal frame
15 $\tilde{R}_f$ is not enabled. In other words, auditory spatial image widening is not enabled. The flag `widening_enabled` is used to indicate the whether or not a gain should be applied. The value 1 corresponds to enabled auditory spatial image widening and the value 0 corresponds to
20 disabled auditory spatial image widening.

A listener's perception of the auditory spatial image of an audio signal he is provided with at a certain time may depend on the history of the audio signal. For example,
25 due to auditory spatial image widening having been applied to preceding reconstructed audio signal frames, the listener may continue to perceive a widened auditory spatial image even no gain is applied to any of the subbands m of the current reconstructed audio signal
30 frames $\tilde{L}_f$ and $\tilde{R}_f$. Thus, it may not be necessary to apply

a gain to any of the subbands m of the currently perceived reconstructed audio signal frames $\tilde{L}_f$ and $\tilde{R}_f$.

In step 2, it is checked whether the conditions $aveDiff < 10^\circ$
 5 or $|\theta_{frame} - 90^\circ| < 10^\circ$ are met.

If false, i.e. if at least one of them is greater than 10° , it is determined that auditory spatial image widening should be applied. Being greater than 10° , the
 10 amount of spatial effects in the audio signal frame pair L_t and R_t is rather large. Thus, a gain should be applied to at least one frequency subband m of the reconstructed first channel audio signal frame $\tilde{L}_f$ or to at least one frequency subband m of the reconstructed second channel
 15 audio signal frame $\tilde{R}_f$ in order to obtain a reconstructed auditory spatial image closely resembling the original auditory spatial image of the audio signal frame pair L_t and R_t .

20 If the check of step 2 yields a positive result, a third check is performed in step 4. This time, it is determined if the conditions $aveDiff > 9^\circ$ and $|\theta_{frame} - 90^\circ| > 9^\circ$ are met and if widening has been enabled for the past two frames. Thus, a somewhat lesser amount of spatial effects in the
 25 audio signal frame pair L_t and R_t is sufficient to cause determination that widening should be applied. Among other reasons, this stems from the fact that if a gain has been applied to a subband of a preceding reconstructed audio signal frame it may be advisable to

also apply a gain to at least one subband m of current reconstructed audio signal frames $\tilde{L}_f$ and $\tilde{R}_f$ so as not to cause sudden changes in the perceived auditory spatial image. Moreover, as widening has not been applied to the
5 past eight frames, but only to a smaller number of frames, it will make a perceptual difference whether or not widening is applied to at least one subband m of the current reconstructed audio signal frames $\tilde{L}_f$ and $\tilde{R}_f$.

10 Similarly, step 5 comprises checking if the conditions $aveDiff > 8^\circ$ and $|\theta_{frame} - 90^\circ| > 8^\circ$ are met and if widening has been enabled for the past six frames.

If the checks in step 5 are negative, step 6 is entered.

15 It is checked if $g_{enc} < 2$ and $aveDiff < 10^\circ$ hold, with

$$g_{enc} = \frac{1}{M} \cdot \sum_{m=0}^{M-1} \left(\frac{MAX(a_m, b_m)}{MIN(a_m, b_m)} \right). \text{ The variables } a_m \text{ and } b_m \text{ and the}$$

functions MAX() and MIN() will be explained in more detail below (see the discussion of equation (13)). If this is not the case, it is determined that auditory
20 spatial image widening should be applied.

Taking into account the amount of spatial effects in the audio signal frame pair L_i and R_i as well as frequency subband specific spatial effect information and the audio
25 signal history according to the exemplary embodiment of the present invention currently discussed, determination whether auditory spatial image widening should be applied is based on properties of human auditory perception. In consequence, reconstructed auditory spatial images may be

obtained that closely resemble the original auditory spatial image.

In the following it is explained how it can be determined
5 for each frequency subband m whether a gain should be applied to either the reconstructed first channel audio signal frame $\tilde{L}_f$ or the reconstructed second channel audio signal frame $\tilde{R}_f$, wherein discrimination whether the gain should be applied to the respective subband m of the
10 reconstructed first channel audio signal frame $\tilde{L}_f$ or of the reconstructed second channel audio signal frame $\tilde{R}_f$ is based on the second parameter θ_m associated with the respective frequency subband m .

15 To this end, if it has been determined that a gain should be applied to at least one frequency subband m of the reconstructed first channel audio signal frame $\tilde{L}_f$, i.e. the left channel audio signal frame, or to at least one frequency subband m of the reconstructed second channel
20 audio signal frame $\tilde{R}_f$, i.e. the right channel audio signal frame, a binary value $flags_{widening}(m)$ is determined for each frequency subband m indicating whether a gain should be applied to the said frequency subband m of the reconstructed left channel audio signal frame $\tilde{L}_f$ or to
25 said frequency subband m of the reconstructed right channel audio signal frame $\tilde{R}_f$. If θ_m is greater than 90° , a gain should be applied to the respective frequency subband m of the reconstructed left channel audio signal

frame $\tilde{L}_f$. Otherwise, it should be applied to the respective frequency subband m of the reconstructed right channel audio signal frame $\tilde{R}_f$.

$$flags_{widening}(m) = \begin{cases} Left, & \theta_m \geq 90^\circ \\ Right, & otherwise \end{cases}, \quad 0 \leq m < M \quad (12)$$

- 5 Thus, applying a gain to the reconstructed audio signal frame associated with the audio signal frame of the audio signal frame pair L_t and R_t that is assigned to the loudspeaker for which the direction from an assumed or known position A of a listener to said loudspeaker is
 10 closer to the perceived direction in the respective frequency subband m as indicated by the second parameter θ_m of said frequency subband m is enabled.

- In the following it is explained how the gain can be
 15 obtained from a reciprocal average value of frequency subband specific indicator ratios, wherein each indicator ratio is the ratio of an indicator of the maximum of the energy e_{L_m} in the frequency subband m of the first channel audio signal frame L_f and the energy e_{R_m} in said
 20 frequency subband m of the second channel audio signal frame R_f , and an indicator of the minimum of the energy e_{L_m} in the frequency subband m of the first channel audio signal frame L_f and the energy e_{R_m} in said frequency subband m of the second channel audio signal frame R_f .

$$a_m = \sqrt{\frac{e_{L_m}}{e_{L_m} + e_{R_m}}}, \quad b_m = \sqrt{\frac{e_{R_m}}{e_{L_m} + e_{R_m}}} \quad (13)$$

Defining a_m and b_m according to equation (13), $MAX(a_m, b_m)$ is an indicator of the maximum of the energy e_{L_m} in the frequency subband m of the first channel audio signal frame L_f and the energy e_{R_m} in said frequency subband m of the second channel audio signal frame R_f . Similarly, $MIN(a_m, b_m)$ is an indicator of the minimum of the energy e_{L_m} in the frequency subband m of the first channel audio signal frame L_f and the energy e_{R_m} in said frequency subband m of the second channel audio signal frame R_f .

The gain $g_{widening}$ is then given by equation (14):

$$g_{widening} = \frac{1}{\sqrt{\frac{1}{M} \cdot \sum_{m=0}^{M-1} \left(\frac{MAX(a_m, b_m)}{MIN(a_m, b_m)} \right)}} \quad (14)$$

$g_{widening}$ is a reciprocal average value of the frequency subband specific indicator ratios of $MAX(a_m, b_m)$ and $MIN(a_m, b_m)$.

15

With $g_{widening}$ being the reciprocal average value of the frequency subband specific indicator ratios, the gain $g_{widening}$ is comparatively small if the subband specific indicator ratios are large. Large subband specific indicator ratios may occur if the indicator $MAX(a_m, b_m)$ of is greater than the indicator $MIN(a_m, b_m)$. In other words, large subband specific indicator ratios may occur if the

20

energies of the first channel audio signal frame L_f and the second channel audio signal frame R_f in the respective subband m differ significantly. In consequence, applying the gain g_{widening} may yield a widened reconstructed auditory spatial image in which subtle inter-channel differences are emphasized more than prominent inter-channel differences. Thus the widened reconstructed auditory spatial image may closely resemble the original auditory spatial image.

10

Gain quantization may enable representation of the gain by a reduced number of bits. If the gain is to be applied to a frequency subband, transmitting it to an entity configured to apply it may only require a reduced bandwidth.

15

A quantized representation of g_{widening} can be obtained using the quantization table $qTbl(k)$:

$$qTbl(k) = 2^{-0.25 \cdot k}, \quad 0 \leq k < K. \quad (15)$$

20 In the present example, K is 16. The quantized gain can be chosen as the quantized value $qTbl(k)$ minimizing the error distance to the unquantized gain value g_{widening} .

25

To further reduce the required bitrate for representing the gain for example for transmission or storage, a lookup table can be used. The entity configured to determine the gain can then derive the index $\text{widening}_{\text{gain_idx}}$ of the respective quantized value. An entity configured

to apply the gain can retrieve the quantized gain $\tilde{g}_{\text{widening}}$ from the lookup table using the index $\text{widening}_{\text{gain_idx}}$.

$$\tilde{g}_{\text{widening}} = qTbl[\text{widening}_{\text{gain_idx}}] \quad (16)$$

The index $\text{widening}_{\text{gain_idx}}$ can possibly be represented by fewer
5 bits than the quantized gain $\tilde{g}_{\text{widening}}$ itself.

Fig. 6 is a flowchart illustrating an exemplary embodiment of a second method according to the second aspect of the present invention.

10

It is similar to the flowchart of Fig. 2. Hence, corresponding method steps are not repeatedly discussed. In contrast to the flowchart of Fig. 2, it is determined in step 112 whether a gain should be applied to at least
15 one frequency subband m of the reconstructed first channel audio signal frame $\tilde{L}_f$ or of the reconstructed second channel audio signal frame $\tilde{R}_f$. The results obtained in method step 112 are also incorporated into a bitstream to be provided for a parametric audio decoder.

20

Fig. 7 is a more detailed illustration of method step 112 of Fig. 6.

In step 113, the first parameter θ_{frame} as defined by
25 equation (7) is calculated. Step 114 comprises calculating the set of second parameters θ_m as defined by equation (10). In step 115 the value $|\theta_{\text{frame}} - 90^\circ|$ is computed, while in step 116 the values $|\theta_m - 90^\circ|$ are

determined. Subsequently, in step 117, $aveDiff$ is determined according to equation (11). In step 118, the actual decision whether a gain should be applied is performed as elucidated above with respect to Fig. 5. If gain application is not be carried out, merely the flag $widening_enabled = 0$ comprising this information is to form part of the bitstream to be provided for a parametric audio decoder. If a gain should be applied, it is computed in step 120 in accordance with equations (13) to (16). Also, the flags $flags_{widening}(m)$ according to equation (12) are determined in step 119. The index $widening_{gain_idx}$ and the flags $flags_{widening}(m)$ are multiplexed into the bitstream along with the flag $widening_enabled = 1$ indicating that the gain should be applied, the mono signal and the spatial extension information.

Fig. 8 is a flowchart exemplarily illustrating decoding of a bitstream generated according to the method of Fig. 6 including gain application. It is similar to the flowchart of Fig. 3, but in step 121 the index $widening_{gain_idx}$, the flags $flags_{widening}(m)$ and the flag $widening_enabled = 1$ are used to apply a gain to frequency subbands m of the reconstructed first channel audio signal frame $\tilde{L}_f$ or of the reconstructed second channel audio signal frame $\tilde{R}_f$. Audio signal frames $\tilde{L}_t$ and $\tilde{R}_t$ are obtained in step 122 by frequency-to-time transformation.

The actual gain application can be performed by multiplying the samples $\tilde{f}_L$ of the reconstructed first channel audio signal frame $\tilde{L}_f$ or the samples $\tilde{f}_R$ of the reconstructed second channel audio signal frame $\tilde{R}_f$ by

- 5 $\tilde{g}_{\text{widening}}$ if the flags $\text{flags}_{\text{widening}}(m)$ indicate to do so.

This following pseudocode also describes this process.

```

for(i = 0; i < M i++)
{
  if widening_enabled == '1' bit
  {
    for(j = sbOffset[i]; j < sbOffset[i + 1]; j++)
    {
      if(flags_widening(i) == Right)
         $\tilde{f}_R(j) = \tilde{f}_R(j) \cdot \tilde{g}_{\text{widening}}$ 
      Else
         $\tilde{f}_L(j) = \tilde{f}_L(j) \cdot \tilde{g}_{\text{widening}}$ 
    }
  }
}

```

- In another embodiment, it is also possible to determine for each frequency subband m whether a gain should be
- 10 applied to either the reconstructed first channel audio signal frame $\tilde{L}_f$ or the reconstructed second channel audio signal frame $\tilde{R}_f$, wherein discrimination whether the gain should be applied to the respective subband m of the reconstructed first channel audio signal frame $\tilde{L}_f$ or of
- 15 the reconstructed second channel audio signal frame $\tilde{R}_f$ is based on the inter-channel level difference $\text{icld}_L(m)$ in said frequency subband m of the reconstructed first

channel audio signal frame $\tilde{L}_f$ and on the inter-channel level difference $icld_R(m)$ in said frequency subband m of the reconstructed second channel audio signal frame $\tilde{R}_f$.

$$flags_{widening}(m) = \begin{cases} Left, & icld_L(m) \geq icld_R(m) \\ Right, & otherwise \end{cases}, \quad 0 \leq m < M \quad (17)$$

- 5 The gain $\tilde{g}_{widening}$ can also be obtained from a reciprocal average value of frequency subband specific indicator ratios, wherein each indicator ratio is the ratio of an indicator of the maximum $MAX(icld_L(i), icld_R(i))$ of the inter-channel level difference $icld_L(m)$ in the frequency subband
- 10 m of the reconstructed first channel audio signal frame $\tilde{L}_f$ and the inter-channel level difference $icld_R(m)$ in said frequency subband m of the reconstructed second channel audio signal frame $\tilde{R}_f$, and an indicator of the minimum $MIN(icld_L(i), icld_R(i))$ of the inter-channel level difference
- 15 $icld_L(m)$ in the frequency subband m of the reconstructed first channel audio signal frame $\tilde{L}_f$ and the inter-channel level difference $icld_R(m)$ in said frequency subband m of the reconstructed second channel audio signal frame $\tilde{R}_f$:

$$\tilde{g}_{widening} = \frac{1}{\sqrt{\frac{1}{M} \cdot \sum_{i=0}^{M-1} \left(\frac{MAX(icld_L(i), icld_R(i))}{MIN(icld_L(i), icld_R(i))} \right)}}. \quad (18)$$

- 20 The parameters determining the inter-channel level differences $icld_L(m)$ and $icld_R(m)$ may anyway form part of the spatial extension information needed in order to establish these differences between the reconstructed

first channel audio signal frame $\tilde{L}_f$ and the reconstructed second channel audio signal frame $\tilde{R}_f$. It may be possible to determine for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame $\tilde{L}_f$ or the reconstructed second channel audio signal frame $\tilde{R}_f$ at the entity configured to perform reconstruction of the reconstructed first and second channel audio signal frames $\tilde{L}_f$ and $\tilde{R}_f$, i.e. at a decoder. To this end, it may not be necessary to provide said entity with additional parameters for determining for each frequency subband m whether a gain should be applied to either the reconstructed first channel audio signal frame $\tilde{L}_f$ or the reconstructed second channel audio signal frame $\tilde{R}_f$. Also, the gain can be applied to at least one frequency subband m of the reconstructed first channel audio signal frame $\tilde{L}_f$ and the reconstructed second channel audio signal frame $\tilde{R}_f$ at the decoder.

Fig. 9 is a schematic illustration of a first exemplary embodiment of an apparatus according to the first aspect of the present invention.

The apparatus comprises a processor 201 configured to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically

encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair. The processor 201 can also be seen as means for determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair. The processor 201 may be part of an audio encoder or an audio decoder, to name but a few examples. For instance, processor 201 may be configured to implement the steps of the flowcharts of Figs. 5-7 or 8.

Fig. 10 is a schematic illustration of a second exemplary embodiment of an apparatus 203 according to the first aspect of the present invention.

The apparatus 203 forms part of a parametric audio encoder 202. It comprises a processor 204. In turn, the processor 204 comprises a frequency subband selection circuit 205 and a parameter determination circuit 206.

5

The processor 204 is configured to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair. Processor 204 thus may be configured to implement the steps of the flowcharts of Fig. 5-7.

The first parameter and the second parameter can be determined by means of the parameter determination circuit 206. Processor 204 is thus also configured to calculate the first parameter and the set of second parameters (see steps 113 and 114 of Fig. 7). The parameter determination circuit 206 may also be thought of as means for calculating the first parameter and the set of second parameters and so may the processor 204.

30

For parametric audio encoding, the parametric audio encoder 202 can be provided with an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame. Thus, information that
5 may be necessary for the parameter determination circuit 206 for calculating a first parameter indicative of a perceived direction of the audio signal frame pair and a set of second parameters, each second parameter indicative of a perceived direction of a frequency
10 subband of the audio signal frame pair, may be available at the parametric audio encoder 202. As the processor 204 and thus the parameter determination circuit 206 form part of the parametric audio encoder 202, this information is available to the parameter determination
15 circuit 206.

The frequency subband selection circuit 205 is configured to determine for each frequency subband whether a gain should be applied to either the reconstructed first
20 channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal
25 frame is based on the second parameter associated with the respective frequency subband. Hence, the processor 204 can be considered to be configured accordingly. The frequency subband selection circuit 205 or the processor 204 may thus also be seen as means for determining for
30 each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal

frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the second
5 parameter associated with the respective frequency subband.

Fig. 11 is a schematic illustration of an embodiment of a readable medium 300 according to the fourth aspect of the
10 present invention.

In this example the readable medium 300 is a computer-readable medium. A program 301 according to the fifth aspect of the present invention is stored thereon. The
15 program 301 comprises program code 302. When executed by a processor, the instructions of the program code 302 cause a processor (for instance processor 201 of Fig. 9 or processor 204 of Fig. 10) to determine whether a gain should be applied to at least one frequency subband of a
20 reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio
25 signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on
30 a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair.

The program 301 can also be considered as a program according to the third aspect of the present invention.

5 It is to be understood that with respect to all of the above embodiments that relate to a processor the processor may for instance be implemented in hardware alone, may have certain aspects in software alone, or may be a combination of hardware and software. The processor
10 may either be a separate module or it may be a subcomponent of a module such as, for example, a processor or an application specific integrated circuit (ASIC) that has other functional components or structures, too.

15 Furthermore, it is readily clear for a person skilled in the art that the logical blocks in the schematic block diagrams as well as the flowchart and algorithm steps presented in the above description may at least partially
20 be implemented in electronic hardware and/or computer software, wherein it may depend on the functionality of the logical block, flowchart step and algorithm step and on design constraints imposed on the respective devices to which degree a logical block, a flowchart step or
25 algorithm step is implemented in hardware or software. The presented logical blocks, flowchart steps and algorithm steps may for instance be implemented in one or more digital signal processors (DSPs), application specific integrated circuits (ASICs), field programmable
30 gate arrays (FPGAs) or other programmable devices. The computer software may be stored in a variety of computer-readable storage media of electric, magnetic, electro-

magnetic or optic type and may be read and executed by a processor, such as for instance a microprocessor. To this end, the processor and the storage medium may be coupled to interchange information, or the storage medium may be
5 included in the processor.

The invention has been described above by means of embodiments, which shall be understood to be exemplary and non-limiting. In particular, it should be noted that
10 there are alternative ways and variations which are obvious to a skilled person in the art and can be implemented without deviating from the scope and spirit of the appended claims. It should also be understood that the sequence of all method steps presented above is not
15 mandatory, also alternative sequences may be possible.

CLAIMS

1. An apparatus, comprising:
 - a processor configured to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair.
2. The apparatus of claim 1, wherein the apparatus comprises a processor configured to calculate the first parameter and the set of second parameters.
3. The apparatus of any of the claims 1-2, wherein the apparatus forms part of a parametric audio encoder.
4. The apparatus of any of any of the claims 1-3, wherein the first parameter and the second parameters are at least partially obtainable from an

assumed or known position of a listener with respect to a first loudspeaker to which the first channel audio signal frame is assigned and with respect to a second loudspeaker to which the second channel audio signal frame is assigned.

5. The apparatus of claim 4, wherein the first parameter is obtainable from:
 - the direction from the assumed or known position of the listener to the first loudspeaker weighted with the sum of the energy of the first channel audio signal frame and the energy of a first channel audio signal frame of the at least one preceding audio signal frame pair, and
 - the direction from the assumed or known position of the listener to the second loudspeaker weighted with the sum of the energy of the second channel audio signal frame and the energy of a second channel audio signal frame of the at least one preceding audio signal frame pair.
6. The apparatus of claim 4 or 5, wherein the second parameters of the set of second parameters are obtainable from the direction from:
 - the assumed or known position of the listener to the first loudspeaker weighted with the first channel audio signal frame energy within the respective frequency subband, and
 - the direction from the assumed or known position of the listener to the second loudspeaker weighted with the second channel audio signal frame energy within the respective frequency subband.

7. The apparatus of any of the claims 4-6, wherein the first parameter and the second parameters are indicative of a perceived direction relative to a first imaginary line passing through the assumed or known position of the listener, the first imaginary line being perpendicular to a second imaginary line connecting the first loudspeaker and the second loudspeaker.
8. The apparatus of any of the claims 1-7, wherein the processor configured to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame is configured to base the determination on averaged second parameters.
9. The apparatus of any of the claims 1-8, wherein the processor configured to determine whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame is further configured to consider whether a gain has been applied to at least one frequency subband of a preceding reconstructed first channel audio signal frame or a preceding reconstructed second channel audio signal frame in determining whether a gain should be applied to at least one frequency subband

of the reconstructed first channel audio signal frame or to at least one frequency subband of the reconstructed second channel audio signal frame.

10. The apparatus of any of the claims 1-9, wherein the frequency subbands of the audio signal frame pair are non-uniform frequency subbands corresponding to the human auditory filters.
11. The apparatus of any of the claims 1-10, wherein the apparatus comprises a processor configured to determine for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the second parameter associated with the respective frequency subband.
12. The apparatus of any of the claims 1-10, wherein the apparatus comprises a processor configured to determine for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the

inter-channel level difference in said frequency subband of the reconstructed first channel audio signal frame and on the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame.

13. The apparatus of any of the claims 1-12, wherein the gain is obtainable from a reciprocal average value of frequency subband specific indicator ratios, wherein each indicator ratio is the ratio of:
 - an indicator of the maximum of the energy in the frequency subband of the first channel audio signal frame and the energy in said frequency subband of the second channel audio signal frame, and
 - an indicator of the minimum of the energy in the frequency subband of the first channel audio signal frame and the energy in said frequency subband of the second channel audio signal frame.
14. The apparatus of any of the claims 1-12, wherein the gain is obtainable from a reciprocal average value of frequency subband specific indicator ratios, wherein each indicator ratio is the ratio of:
 - an indicator of the maximum of the inter-channel level difference in the frequency subband of the reconstructed first channel audio signal frame and the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame, and
 - an indicator of the minimum of the inter-channel level difference in the frequency subband of the reconstructed first channel audio signal frame and

the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame.

15. The apparatus of any of the claims 1-14, wherein the gain is a quantized value.
16. The apparatus of claim 15, wherein the gain is obtainable from a lookup table.
17. An apparatus, comprising:
 - means for determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair.
18. A method, comprising:
 - determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one

frequency subband of a reconstructed second channel audio signal frame, wherein said reconstructed first and second channel audio signal frame are obtainable by decoding a parametrically encoded audio signal representing an audio signal frame pair comprising a first channel audio signal frame and a second channel audio signal frame, based on a first parameter indicative of a perceived direction of the audio signal frame pair and at least one preceding audio signal frame pair and on a set of second parameters, each second parameter indicative of a perceived direction of a frequency subband of the audio signal frame pair.

19. The method of claim 18, wherein it further comprises calculating the first parameter and the set of second parameters.
20. The method of any of the claims 18-19, wherein determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame is performed on a parametric audio encoder side.
21. The method of any of the claims 18-20, wherein the first parameter and the second parameters are at least partially obtainable from an assumed or known position of a listener relative to a first loudspeaker to which the first channel audio signal frame is assigned and relative to a second

loudspeaker to which the second channel audio signal frame is assigned.

22. The method of claim 21, wherein the first parameter is obtainable from:
- the direction from the assumed or known position of the listener to the first loudspeaker weighted with the sum of the energy of the first channel audio signal frame and the energy of a first channel audio signal frame of the at least one preceding audio signal frame pair, and
 - the direction from the assumed or known position of the listener to the second loudspeaker weighted with the sum of the energy of the second channel audio signal frame and the energy of a second channel audio signal frame of the at least one preceding audio signal frame pair.
23. The method of any of the claims 21-22, wherein the second parameters of the set of second parameters are obtainable from the direction from:
- the assumed or known position of the listener to the first loudspeaker weighted with the first channel audio signal frame energy within the respective frequency subband, and
 - the direction from the assumed or known position of the listener to the second loudspeaker weighted with the second channel audio signal frame energy within the respective frequency subband.
24. The method of any of the claims 21-23, wherein the first parameter and the second parameters are

indicative of a perceived direction relative to a first imaginary line passing through the assumed or known position of the listener, the first imaginary line being perpendicular to a second imaginary line connecting the first loudspeaker and the second loudspeaker.

25. The method of any of the claims 18-24, wherein determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame is based on averaged second parameters.
26. The method of any of the claims 18-25, wherein determining whether a gain should be applied to at least one frequency subband of a reconstructed first channel audio signal frame or to at least one frequency subband of a reconstructed second channel audio signal frame involves considering whether a gain has been applied to at least one frequency subband of a preceding reconstructed first channel audio signal frame or a preceding reconstructed second channel audio signal frame.
27. The method of any of the claims 18-26, wherein the frequency subbands of the audio signal frame pair are non-uniform frequency subbands corresponding to the human auditory filters.

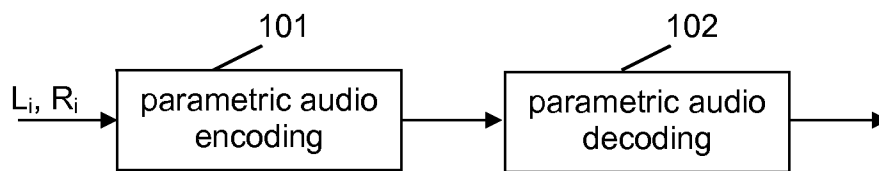
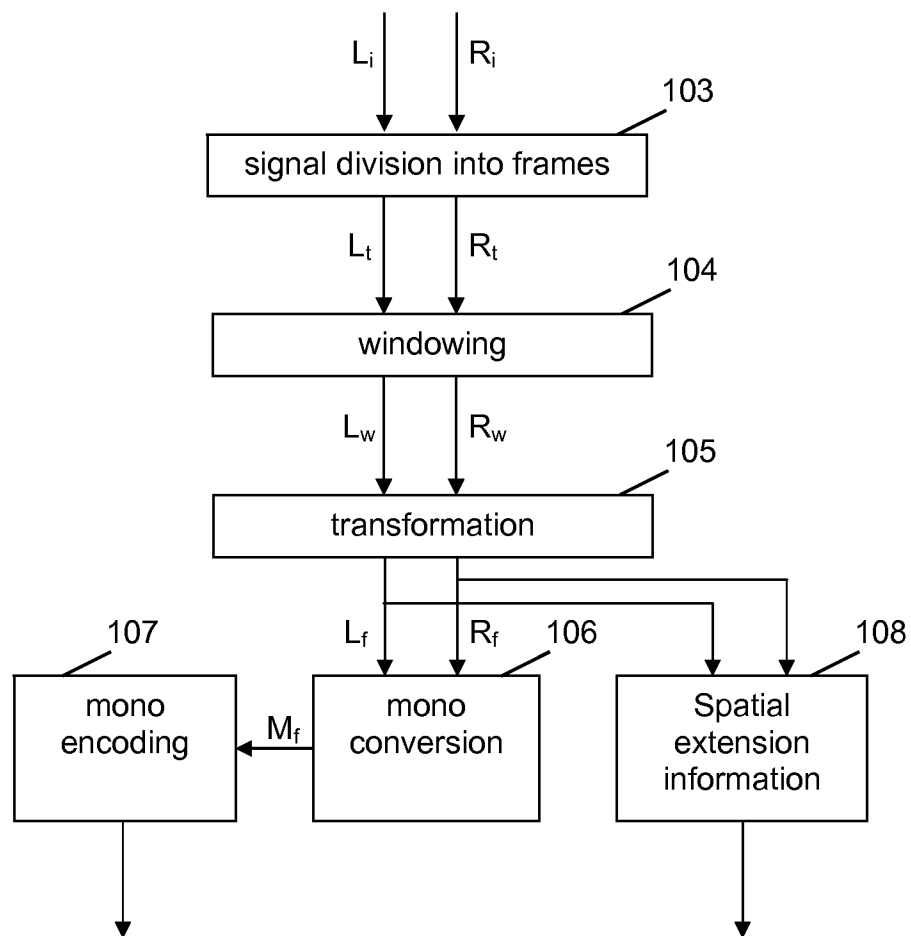
28. The method of any of the claims 18-27, wherein it comprises determining for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the second parameter associated with the respective frequency subband.
29. The method of any of the claims 18-27, wherein it comprises determining for each frequency subband whether a gain should be applied to either the reconstructed first channel audio signal frame or the reconstructed second channel audio signal frame, wherein discrimination whether the gain should be applied to the respective subband of the reconstructed first channel audio signal frame or of the reconstructed second channel audio signal frame is based on the inter-channel level difference in said frequency subband of the reconstructed first channel audio signal frame and on the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame
30. The method of any of the claims 18-29, wherein the gain is obtainable from a reciprocal average value of frequency subband specific indicator ratios, wherein each indicator ratio is the ratio of:

- an indicator of the maximum of the energy in the frequency subband of the first channel audio signal frame and the energy in said frequency subband of the second channel audio signal frame, and
 - an indicator of the minimum of the energy in the frequency subband of the first channel audio signal frame and the energy in said frequency subband of the second channel audio signal frame.
31. The method of any of the claims 18-29, wherein the gain is obtainable from a reciprocal average value of frequency subband specific indicator ratios, wherein each indicator ratio is the ratio of:
- an indicator of the maximum of the inter-channel level difference in the frequency subband of the reconstructed first channel audio signal frame and the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame, and
 - an indicator of the minimum of the inter-channel level difference in the frequency subband of the reconstructed first channel audio signal frame and the inter-channel level difference in said frequency subband of the reconstructed second channel audio signal frame.
32. The method of any of the claims 18-31, wherein the gain is a quantized value.
33. The method of claim 32, wherein the gain is obtainable from a lookup table.

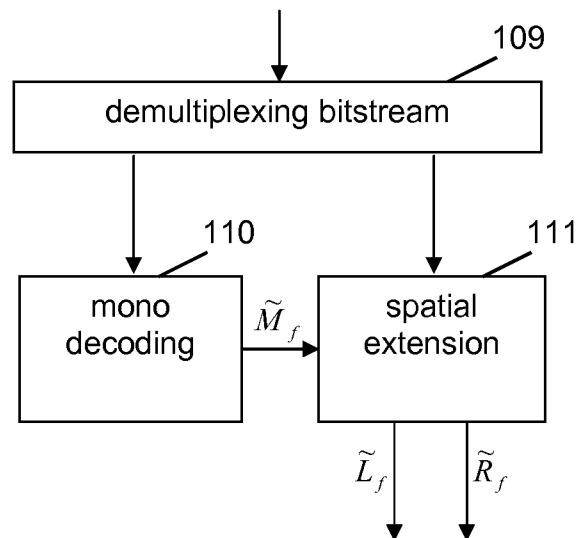
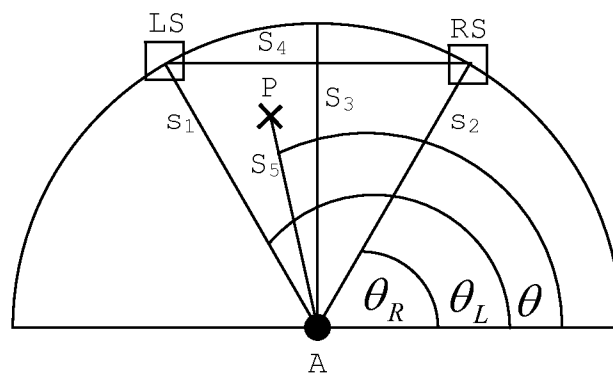
34. A program comprising:

- program code for performing the method according to any of the claims 18-33, when the program is executed on a processor.

1/7

**Fig. 1****Fig. 2**

2/7

**Fig. 3****Fig. 4**

3/7

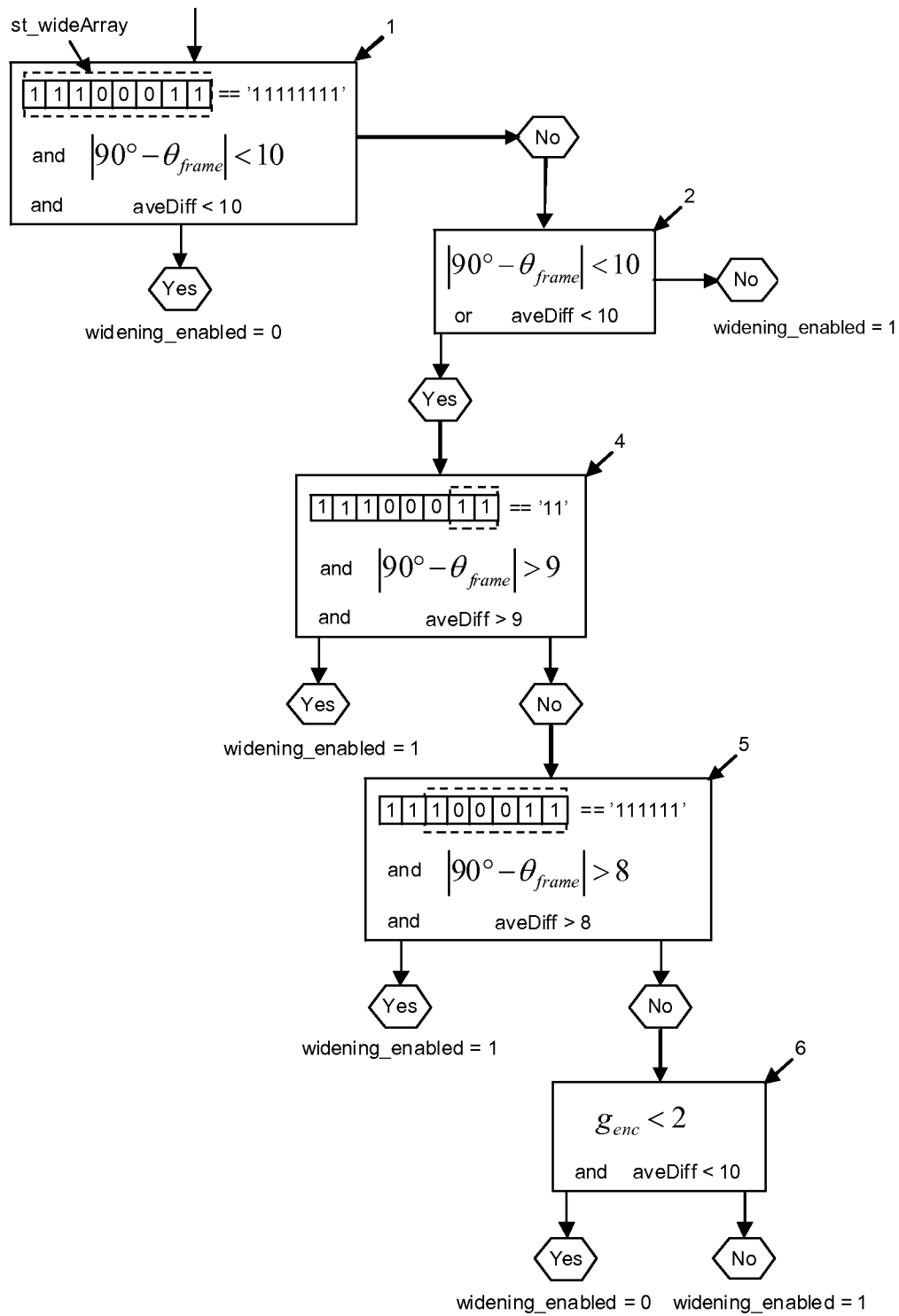
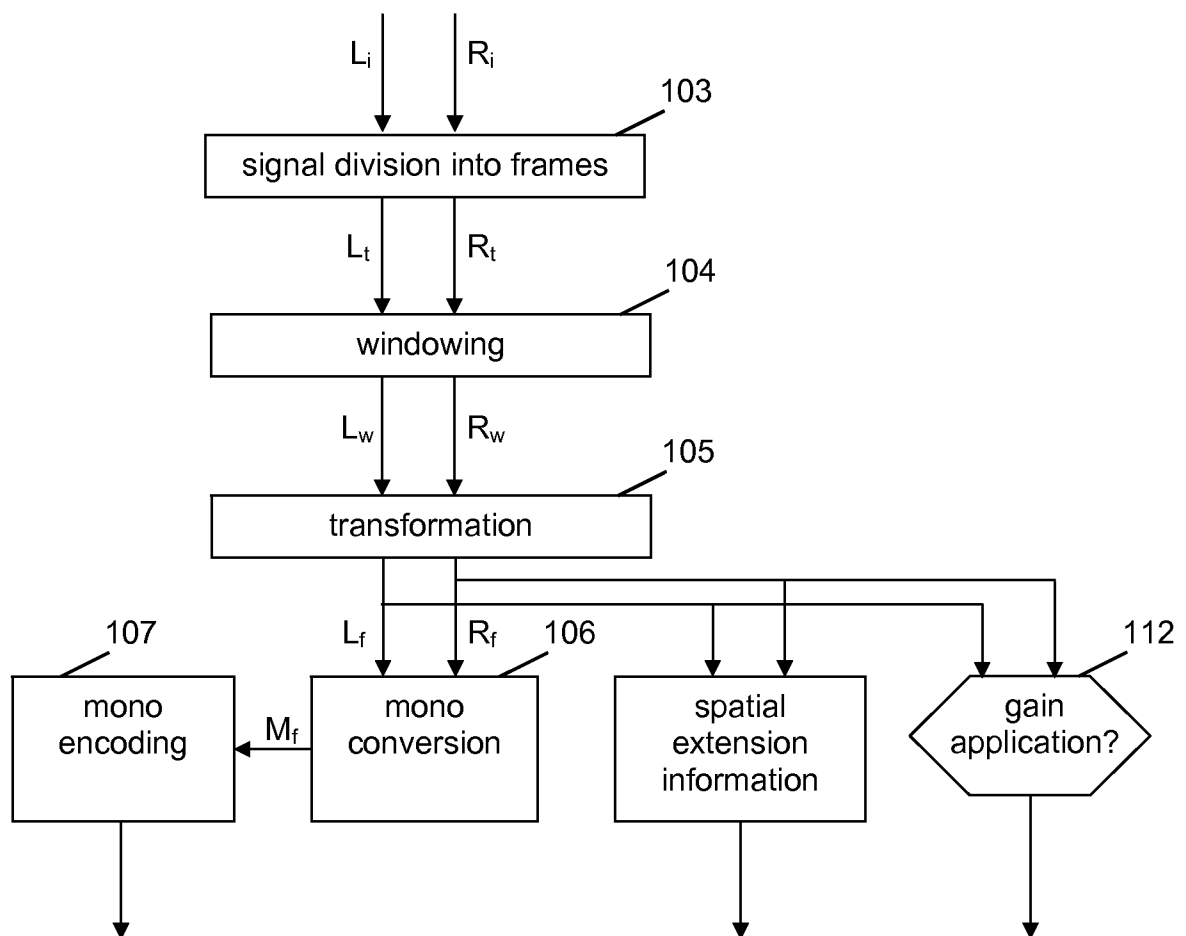
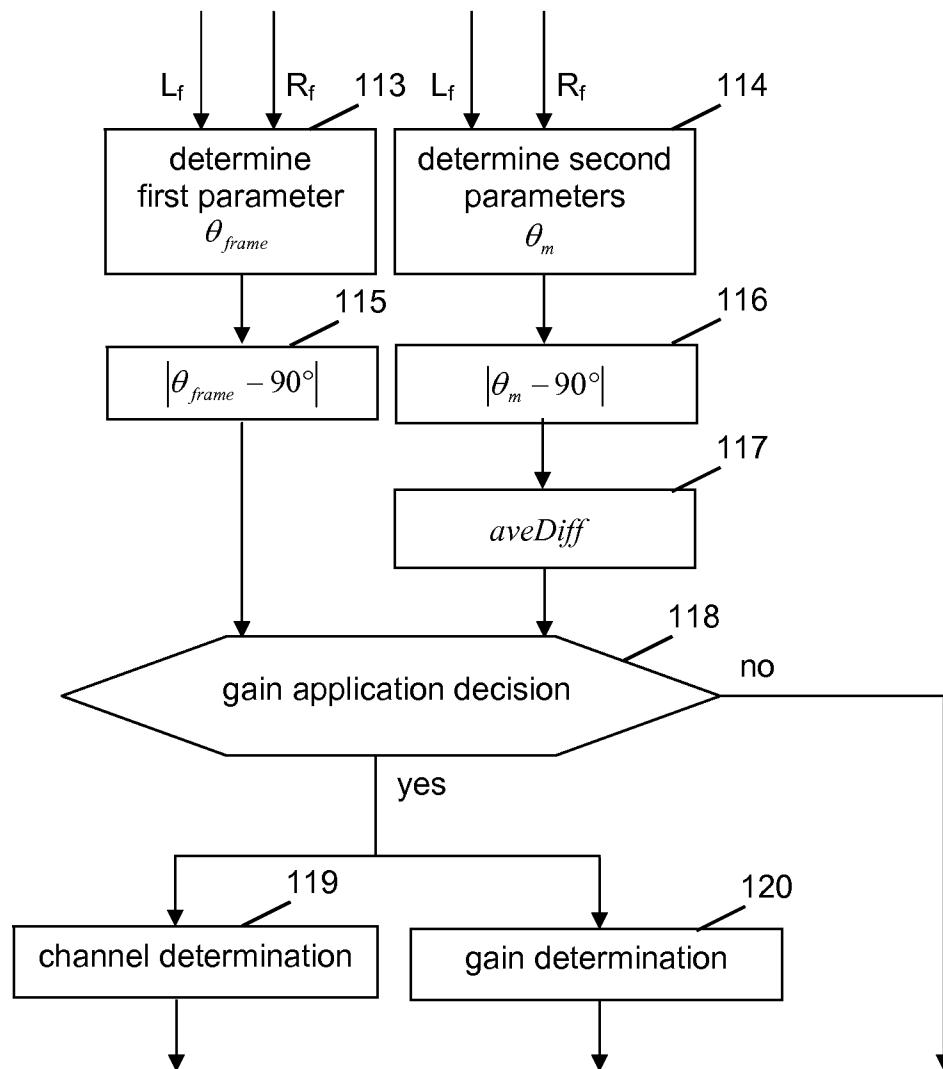


Fig. 5

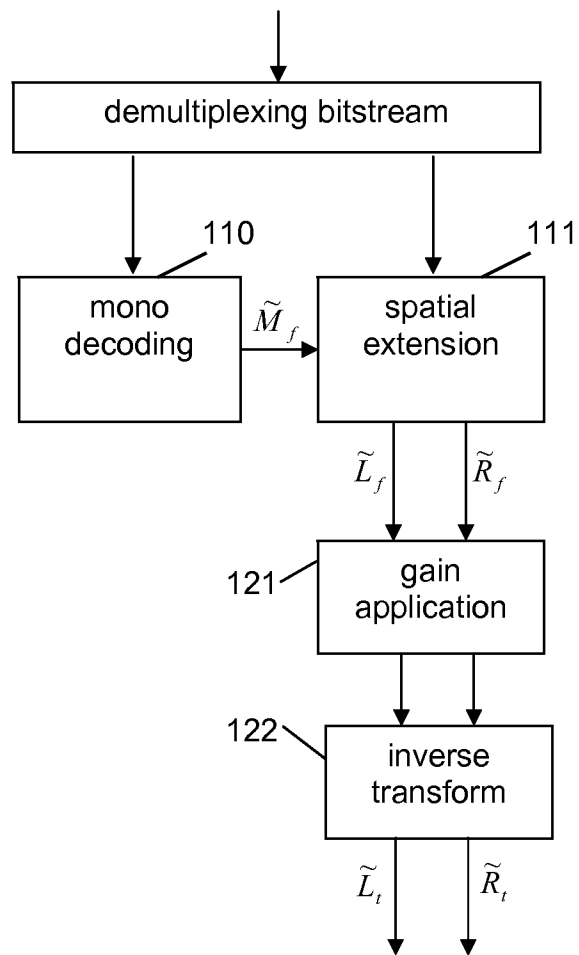
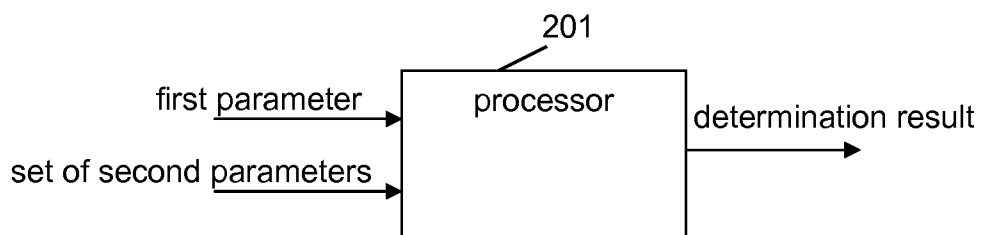
4/7

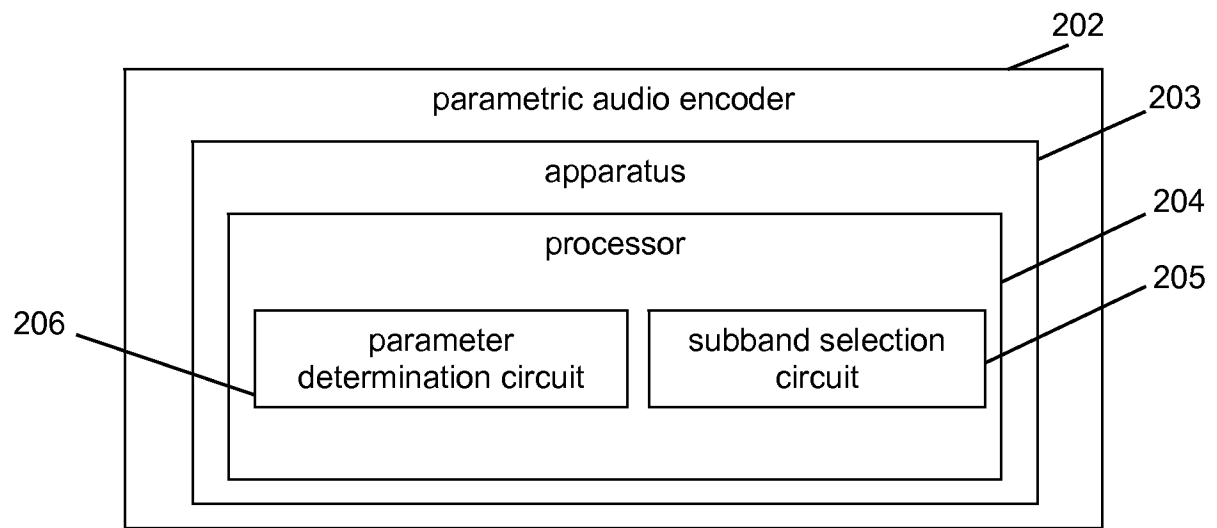
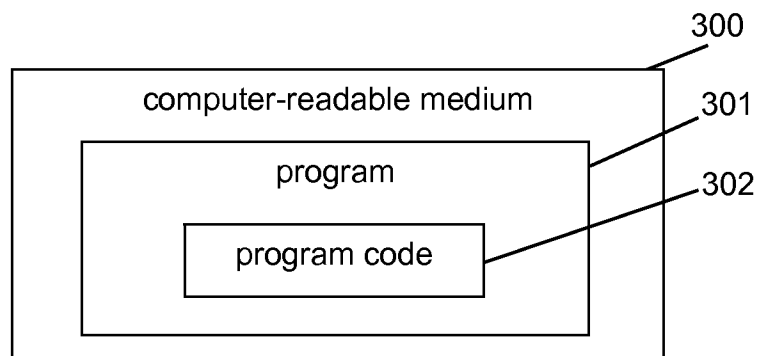
**Fig. 6**

5/7

112**Fig. 7**

6/7

**Fig. 8****Fig. 9**

**Fig. 10****Fig. 11**

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2008/068371

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L19/00

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

B. FIELDS SEARCHED

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

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

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

EPO-Internal, WPI Data, IBM-TDB, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2007/080225 A (NOKIA CORP [FI]; OJALA PASI [FI]; TURKU JULIA [FI]; VAEAENAENEN MAURI) 19 July 2007 (2007-07-19)	1-4, 10-12, 15-21, 27-29, 32-34
Y	page 3, line 10 - page 4, line 9	5-7, 22-24
A	page 5, line 23 - page 11, line 17; figure 3 page 13, line 6 - line 15 ----- -/--	8,9,13, 14,25, 26,30,31

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

G document member of the same patent family

Date of the actual completion of the International search

16 March 2009

Date of mailing of the international search report

06/04/2009

Name and mailing address of the ISA/
European Patent Office, P.B. 5618 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Dobler, Ervin

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2008/068371

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JAKKA J: "Binaural to Multichannel Audio Upmix" THESIS HELSINKI UNIVERSITY OF TECHNOLOGY, 6 June 2005 (2005-06-06), page 34, XP007907636 page 16, line 12 - page 17, line 1; figure 3.1 -----	5-7, 22-24
A	VILLE PULKKI: "Spatial Sound Generation and Perception by Amplitude Panning Techniques" PHD DISSERTATION, 3 August 2001 (2001-08-03), pages 1-42, XP007907630 Helsinki, FI page 11, line 1 - page 16, line 9 -----	1-34

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2008/068371

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007080225 A	19-07-2007	NONE	