Low-Frequency Emphasis for LPC-based Coding in Frequency Domain

Description

It is well-known that non-speech signals, e.g. musical sound, can be more complicated in processing than human vocal sound, occupying a wider band of frequency. Recent state-of-the-art audio coding systems such as AMR-WB+ [3] and xHE-AAC [4] offer a transform coding tool for music and other generic, non-speech signals. This tool is commonly known as transform coded excitation (TCX) and is based on the principle of transmission of a linear predictive coding (LPC) residual, termed excitation, quantized and entropy coded in the frequency domain. Due to the limited order of the predictor used in the LPC stage, however, artifacts can occur in the decoded signal especially at low frequencies, where human hearing is very sensitive. To this end, a low-frequency emphasis and de-emphasis scheme was introduced in [1–3].

Said prior-art adaptive low-frequency emphasis (ALFE) scheme amplifies low-frequency spectral lines prior to quantization in the encoder. In particular, low-frequency lines are grouped into bands, the energy of each band is computed, and the band with the local energy maximum is found. Based on the value and location of the energy maximum, bands below the maximum-energy band are boosted so that they are quantized more accurately in the subsequent quantization.

The low-frequency de-emphasis performed to invert the ALFE in a corresponding decoder is conceptually very similar. As done in the encoder, low-frequency bands are established and a band with maximum energy is determined. Unlike in the encoder, the bands below the energy peak are now attenuated. This procedure roughly restores the line energies of the original spectrum.

It is worth noting that in the prior art, the band-energy calculation in the encoder is performed before quantization, i.e. on the input spectrum, whereas in the decoder it is conducted on the inversely quantized lines, i.e. the decoded spectrum. Although the quantization operation can be designed such that spectral energy is preserved on average, exact energy preservation cannot be assured for individual spectral lines. Hence, the ALFE cannot be perfectly inverted. Moreover, a square-root operation is required in a preferred implementation of the prior-art ALFE in both encoder and decoder. Avoiding such relatively complex operations is desirable.

An object of the present invention is to provide improved concepts for audio signal processing. More particularly, an object of the present invention is to provide improved concepts for adaptive low-frequency emphasis and de-emphasis. The object of the present invention is achieved by an audio encoder according to claim 1, an audio decoder according to claim 11, by a system according to claim 21, by methods according to claims 22 and 23 and by a computer program according to claim 24.

In one aspect the invention provides an audio encoder for encoding a non-speech audio signal so as to produce therefrom a bitstream, the audio encoder comprising:

a combination of a linear predictive coding filter having a plurality of linear predictive coding coefficients and a time-frequency converter, wherein the combination is configured to filter and to convert a frame of the audio signal into a frequency domain in order to output a spectrum based on the frame and on the linear predictive coding coefficients;

a low-frequency emphasizer configured to calculate a processed spectrum based on the spectrum, wherein spectral lines of the processed spectrum representing a lower frequency than a reference spectral line are emphasized; and

a control device configured to control the calculation of the processed spectrum by the low-frequency emphasizer depending on the linear predictive coding coefficients of the linear predictive coding filter.

A linear predictive coding filter (LPC filter) is a tool used in audio signal processing and speech processing for representing the spectral envelope of a framed digital signal of sound in compressed form, using the information of a linear predictive model.

A time-frequency converter is a tool for converting in particular a framed digital signal from the time domain into a frequency domain so as to estimate a spectrum of the signal. The time-frequency converter may use a modified discrete cosine transform (MDCT), which is a lapped transform based on the type-IV discrete cosine transform (DCT-IV), with the additional property of being lapped: it is designed to be performed on consecutive frames of a larger dataset, where subsequent frames are overlapped so that the last half of one frame coincides with the first half of the next frame. This overlapping, in addition to the energy-compaction qualities of the DCT, makes the MDCT especially attractive for signal compression applications, since it helps to avoid artifacts stemming from the frame boundaries.

The low-frequency emphasizer is configured to calculate a processed spectrum based on the spectrum, wherein spectral lines of the processed spectrum representing a lower frequency than a reference spectral line are emphasized so that only low frequencies contained in the processed spectrum are emphasized. The reference spectral line may be predefined based on empirical experience.

The control device is configured to control the calculation of the processed spectrum by the low-frequency emphasizer depending on the linear predictive coding coefficients of the linear predictive coding filter. Therefore, the

encoder according to the invention does not need to analyze the spectrum of the audio signal for the purpose of low-frequency emphasis. Further, since identical linear predictive coding coefficients may be used in the encoder and in a subsequent decoder, the adaptive low-frequency emphasis is fully invertible regardless of spectrum quantization as long as the linear predictive coding coefficients are transmitted to the decoder in the bitstream which is produced by the encoder or by any other means. In general the linear predictive coding coefficients have to be transmitted in the bitstream anyway for the purpose of reconstructing an audio output signal from the bitstream by a respective decoder. Therefore, the bit rate of the bitstream will not be increased by the low-frequency emphasis as described herein.

The adaptive low-frequency emphasis system described herein may be implemented in the TCX core-coder of LD-USAC (EVS), a low-delay variant of xHE-AAC [4] which can switch between time-domain and MDCT-domain coding on a per-frame basis.

According to a preferred embodiment of the invention the frame of the audio signal is input to the linear predictive coding filter, wherein a filtered frame is output by the linear predictive coding filter and wherein the time-frequency converter is configured to estimate the spectrum based on the filtered frame. Accordingly, the linear predictive coding filter may operate in the time domain, having the audio signal as its input.

According to a preferred embodiment of the invention the frame of the audio signal is input to the time-frequency converter, wherein a converted frame is output by the time-frequency converter and wherein the linear predictive coding filter is configured to estimate the spectrum based on the converted frame. Alternatively but equivalently, to the first embodiment of the inventive encoder having a low-frequency emphasizer, the encoder may calculate a processed spectrum based on the spectrum of a frame produced by means of frequency-domain noise shaping (FDNS), as disclosed for example in [5].

More specifically, the tool ordering here is modified: the time-frequency converter such as the above-mentioned one may be configured to estimate a converted frame based on the frame of the audio signal and the linear predictive coding filter is configured to estimate the audio spectrum based on the converted frame, which is output by the time-frequency converter. Accordingly, the linear predictive coding filter may operate in the frequency domain (instead of the time domain), having the converted frame as its input, with the linear predictive coding filter applied via multiplication by a spectral representation of the linear predictive coding coefficients.

It should be evident to those skilled in the art that these two approaches – a linear filtering in the time domain followed by time-frequency conversion vs. time-frequency conversion followed by linear filtering via spectral weighting in the frequency domain – can be implemented such that they are equivalent.

According to a preferred embodiment of the invention the audio encoder comprises a quantization device configured to produce a quantized spectrum based on the processed spectrum and a bitstream producer configured to embed the quantized spectrum and the linear predictive coding coefficients into the bitstream. Quantization, in digital signal processing, is the process of mapping a large set of input values to a (countable) smaller set – such as rounding values to some unit of precision. A device or algorithmic function that performs quantization is called a quantization device. The bitstream producer may be any device which is capable of embedding digital data from different sources into a unitary bitstream. By these features a bitstream produced with an adaptive low-frequency emphasis may be produced easily, wherein the adaptive low-frequency emphasis is fully invertible by a subsequent decoder solely using information already contained in the bitstream.

In a preferred embodiment of the invention the control device comprises a spectral analyzer configured to estimate a spectral representation of the linear predictive coding coefficients, a minimum-maximum analyzer configured to

estimate a minimum of the spectral representation and a maximum of the spectral representation below a further reference spectral line, and an emphasis factor calculator configured to calculate spectral line emphasis factors for calculating the spectral lines of the processed spectrum representing a lower frequency than the reference spectral line based on the minimum and on the maximum, wherein the spectral lines of the processed spectrum are emphasized by applying the spectral line emphasis factors to spectral lines of the spectrum of the filtered frame. The spectral analyzer may be a time-frequency converter as described above. The spectral representation is the transfer function of the linear predictive coding filter and may be, but does not have to be, the same spectral representation as the one utilized for FDNS, as described above. The spectral representation may be computed from an odd discrete Fourier transform (ODFT) of the linear predictive coding coefficients. In xHE-AAC and LD-USAC, the transfer function may be approximated by 32 or 64 MDCT-domain gains that cover the entire spectral representation.

In a preferred embodiment of the invention the emphasis factor calculator is configured in such a way that the spectral line emphasis factors increase in a direction from the reference spectral line to the spectral line representing the lowest frequency of the spectrum. This means that the spectral line representing the lowest frequency is amplified the most whereas the spectral line adjacent to the reference spectral line is amplified the least. The reference spectral line and spectral lines representing higher frequencies than the reference spectral line are not emphasized at all. This reduces the computational complexity without any audible disadvantages.

In a preferred embodiment of the invention the emphasis factor calculator comprises a first stage configured to calculate a basis emphasis factor according to a first formula $\gamma = (\alpha \cdot \min / \max)^{\beta}$, wherein α is a first preset value, with $\alpha > 1$, β is a second preset value, with $0 < \beta \leq 1$, min is the minimum of the spectral representation, max is the maximum of the spectral representation, and γ is the basis emphasis factor, and wherein the emphasis factor

calculator comprises a second stage configured to calculate spectral line emphasis factors according to a second formula $\varepsilon_i = \gamma^{i'-i}$, wherein i' is a number of the spectral lines to be emphasized, i is an index of the respective spectral line, the index increases with the frequencies of the spectral lines, with i = 0 to i'-1, γ is the basis emphasis factor and ε_i is the spectral line emphasis factor with index i. The basis emphasis factor is calculated from a ratio of the minimum and the maximum by the first formula in an easy way. The basis emphasis factor serves as a basis for the calculation of all spectral line emphasis factors, wherein the second formula ensures that the spectral line emphasis factors increase in a direction from the reference spectral line to the spectral line representing the lowest frequency of the spectrum. In contrast to prior-art solutions the proposed solution does not require a per-spectral-band square-root or similar complex operation. Only 2 division and 2 power operators are needed, one of each on encoder and decoder side.

In a preferred embodiment of the invention the first preset value is smaller than 42 and larger than 22, in particular smaller than 38 and larger than 26, more particular smaller 34 and larger than 30. The aforementioned intervals are based on empirical experiments. Best results may be achieved when the first preset value is set to 32.

In a preferred embodiment of the invention the second preset value is determined according to the formula $\beta = 1 / (\theta \cdot i')$, wherein i' is a number of the spectral lines being emphasized, θ is a factor between 3 and 5, in particular between 3,4 and 4,6, more particular between 3,8 and 4,2. Also these intervals are based on empirical experiments. It has been found the best results may be achieved when the second preset value is set to 4.

In a preferred embodiment of the invention the reference spectral line represents a frequency between 600 Hz and 1000 Hz, in particular between 700 Hz and 900 Hz, more particular between 750 Hz and 850 Hz. These empirically found intervals ensure sufficient low-frequency emphasis as well as a

low computational complexity of the system. These intervals ensure in particular that in densely populated spectra, the lower-frequency lines are coded with sufficient accuracy. In a preferred embodiment the reference spectral line represents 800 Hz, wherein 32 spectral lines are emphasized.

In a preferred embodiment of the invention the further reference spectral line represents the same or a higher frequency than the reference spectral line. These features ensure that the estimation of the minimum and the maximum is done in the relevant frequency range.

In the preferred embodiment of the invention the control device is configured in such a way that the spectral lines of the processed spectrum representing a lower frequency than the reference spectral are emphasized only if the maximum is less than the minimum multiplied with α, the first preset value. These features ensure that low-frequency emphasis is only executed when needed so that the work load of the encoder may be minimized and no bits are wasted on perceptually unimportant regions during spectral quantization.

In one aspect the invention provides an audio decoder for decoding a bitstream based on a non-speech audio signal so as to produce from the bitstream a decoded non-speech audio output signal, in particular for decoding a bitstream produced by an audio encoder according to the invention, the bitstream containing quantized spectrums and a plurality of linear predictive coding coefficients, the audio decoder comprising:

a bitstream receiver configured to extract the quantized spectrum and the linear predictive coding coefficients from the bitstream;

a de-quantization device configured to produce a de-quantized spectrum based on the quantized spectrum;

a low-frequency de-emphasizer configured to calculate a reverse processed

spectrum based on the de-quantized spectrum, wherein spectral lines of the reverse processed spectrum representing a lower frequency than a reference spectral line are de-emphasized; and

a control device configured to control the calculation of the reverse processed spectrum by the low-frequency de-emphasizer depending on the linear predictive coding coefficients contained in the bitstream.

The bitstream receiver may be any device which is capable of classifying digital data from a unitary bitstream so as to send the classified data to the appropriate subsequent processing stage. In particular, the bitstream receiver is configured to extract the quantized spectrum, which then is forwarded to the de-quantization device, and the linear predictive coding coefficients, which then are forwarded to the control device, from the bitstream.

The de-quantization device is configured to produce a de-quantized spectrum based on the quantized spectrum, wherein de-quantization is an inverse process with respect to quantization as explained above.

The low-frequency de-emphasizer is configured to calculate a reverse processed spectrum based on the de-quantized spectrum, wherein spectral lines of the reverse processed spectrum representing a lower frequency than a reference spectral line are de-emphasized so that only low frequencies contained in the reverse processed spectrum are de-emphasized. The reference spectral line may be predefined based on empirical experience. It has to be noted that the reference spectral line of the decoder should represent the same frequency as the reference spectral line of the encoder as explained above. However, the frequency to which the reference spectral line refers may be stored on the decoder side so that it is not necessary to transmit this frequency in the bitstream.

The control device is configured to control the calculation of the reverse processed spectrum by the low-frequency de-emphasizer depending on the linear predictive coding coefficients of the linear predictive coding filter. Since identical linear predictive coding coefficients may be used in the encoder producing the bitstream and in the decoder, the adaptive low-frequency emphasis is fully invertible regardless of spectrum quantization as long as the linear predictive coding coefficients are transmitted to the decoder in the bitstream. In general the linear predictive coding coefficients have to be transmitted in the bitstream anyway for the purpose of reconstructing the audio output signal from the bitstream by the decoder. Therefore, the bit rate of the bitstream will not be increased by the low-frequency emphasis and the low-frequency de-emphasis as described herein.

The adaptive low-frequency de-emphasis system described herein may be implemented in the TCX core-coder of LD-USAC, a low-delay variant of xHE-AAC [4] which can switch between time-domain and MDCT-domain coding.

By these features a bitstream produced with an adaptive low-frequency emphasis may be decoded easily, wherein the adaptive low-frequency de-emphasis may be done by the decoder solely using information already contained in the bitstream.

According to a preferred embodiment of the invention the audio decoder comprises combination of a frequency-time converter and an inverse linear predictive coding filter receiving the plurality of linear predictive coding coefficients contained in the bitstream, wherein the combination is configured to inverse-filter and to convert the reverse processed spectrum into a time domain in order to output the output signal based on the reverse processed spectrum and on the linear predictive coding coefficients.

A frequency-time converter is a tool for executing an inverse operation of the operation of a time-frequency converter as explained above. It is a tool for

converting in particular a spectrum of a signal in a frequency domain into a framed digital signal in the time domain so as to estimate the original signal. The frequency-time converter may use an inverse modified discrete cosine transform (inverse MDCT), wherein the modified discrete cosine transform is a lapped transform based on the type-IV discrete cosine transform (DCT-IV), with the additional property of being lapped: it is designed to be performed on consecutive frames of a larger dataset, where subsequent frames are overlapped so that the last half of one frame coincides with the first half of the next frame. This overlapping, in addition to the energy-compaction qualities of the DCT, makes the MDCT especially attractive for signal compression applications, since it helps to avoid artifacts stemming from the frame boundaries. Those skilled in the art will understand that other transforms are possible. However, the transform in the decoder should be an inverse transform of the transform in the encoder.

An inverse linear predictive coding filter is a tool for executing an inverse operation to the operation done by the linear predictive coding filter (LPC filter) as explained above. It is a tool used in audio signal processing and speech processing for decoding of the spectral envelope of a framed digital signal in order to reconstruct the digital signal, using the information of a linear predictive model. Linear predictive coding and decoding is fully invertible as long as the same linear predictive coding coefficients are used, which may be ensured by transmitting the linear predictive coding coefficients from the encoder to the decoder embedded in the bitstream as described herein.

By these features the output signal may be processed in an easy way.

According to a preferred embodiment of the invention the frequency-time converter is configured to estimate a time signal based on the reverse processed spectrum, wherein the inverse linear predictive coding filter is configured to output the output signal based on the time signal. Accordingly, the

inverse linear predictive coding filter may operate in the time domain, having the reverse processed spectrum as its input.

According to a preferred embodiment of the invention the inverse linear predictive coding filter is configured to estimate an inverse filtered signal based on the reverse processed spectrum, wherein the frequency-time converter is configured to output the output signal based on the inverse filtered signal.

Alternatively and equivalently, and analogous to the above-described FDNS procedure performed on the encoder side, the order of the frequency-time converter and the inverse linear predictive coding filter may be reversed such that the latter is operated first and in the frequency domain (instead of the time domain). More specifically, the inverse linear predictive coding filter may output an inverse filtered signal based on the reverse processed spectrum, with the inverse linear predictive coding filter applied via multiplication (or division) by a spectral representation of the linear predictive coding coefficients, as in [5]. Accordingly, a frequency-time converter such as the above-mentioned one may be configured to estimate a frame of the output signal based on the inverse filtered signal, which is input to the time-frequency converter.

It should be evident to those skilled in the art that these two approaches – a linear inverse filtering in the frequency domain followed by frequency-time conversion vs. frequency-time conversion followed by linear filtering via spectral weighting in the time domain – can be implemented such that they are equivalent.

In a preferred embodiment of the invention the control device comprises a spectral analyzer configured to estimate a spectral representation of the linear predictive coding coefficients, a minimum-maximum analyzer configured to estimate a minimum of the spectral representation and a maximum of the spectral representation below a further reference spectral line and a de-

emphasis factor calculator configured to calculate spectral line de-emphasis factors for calculating the spectral lines of the reverse processed spectrum representing a lower frequency than the reference spectral line based on the minimum and on the maximum, wherein the spectral lines of the reverse processed spectrum are de-emphasized by applying the spectral line de-emphasis factors to spectral lines of the de-quantized spectrum. The spectral analyzer may be a time-frequency converter as described above. The spectral representation is the transfer function of the linear predictive coding filter and may be, but does not have to be, the same spectral representation as the one utilized for FDNS, as described above. The spectral representation may be computed from an odd discrete Fourier transform (ODFT) of the linear predictive coding coefficients. In xHE-AAC and LD-USAC, the transfer function may be approximated by 32 or 64 MDCT-domain gains that cover the entire spectral representation.

In a preferred embodiment of the invention the de-emphasis factor calculator is configured in such a way that the spectral line de-emphasis factors decrease in a direction from the reference spectral line to the spectral line representing the lowest frequency of the reverse processed spectrum. This means that the spectral line representing the lowest frequency is attenuated the most whereas the spectral line adjacent to the reference spectral line is attenuated the least. The reference spectral line and spectral lines representing higher frequencies than the reference spectral line are not de-emphasized at all. This reduces the computational complexity without any audible disadvantages.

In a preferred embodiment of the invention the de-emphasis factor calculator comprises a first stage configured to calculate a basis de-emphasis factor according to a first formula $\delta = (\alpha \cdot \min / \max)^{-\beta}$, wherein α is a first preset value, with $\alpha > 1$, β is a second preset value, with $0 < \beta \leq 1$, min is the minimum of the spectral representation, max is the maximum of the spectral representation and δ is the basis de-emphasis factor, and wherein the de-

emphasis factor calculator comprises a second stage configured to calculate spectral line de-emphasis factors according to a second formula $\zeta_i = \delta^{i'-i}$, wherein i' is a number of the spectral lines to be de-emphasized, i is an index of the respective spectral line, the index increases with the frequencies of the spectral lines, with i = 0 to i'-1, δ is the basis de-emphasis factor and ζ_i is the spectral line de-emphasis factor with index i. The operation of the de-emphasis factor calculator is inverse to the operation of the emphasis factor calculator as described above. The basis de-emphasis factor is calculated from a ratio of the minimum and the maximum by the first formula in an easy way. The basis de-emphasis factor serves as a basis for the calculation of all spectral line de-emphasis factors, wherein the second formula ensures that the spectral line de-emphasis factors decrease in a direction from the reference spectral line to the spectral line representing the lowest frequency of the reverse processed spectrum. In contrast to prior-art solutions the proposed solution does not require a per-spectral-band square-root or similar complex operation. Only 2 division and 2 power operators are needed, one of each on encoder and decoder side.

In a preferred embodiment of the invention the first preset value is smaller than 42 and larger than 22, in particular smaller than 38 and larger than 26, more particular smaller 34 and larger than 30. The aforementioned intervals are based on empirical experiments. Best results may be achieved when the first preset value is set to 32. Note, that the first preset value of the decoder should be the same as the first preset value of the encoder.

In a preferred embodiment of the invention the second preset value is determined according to the formula $\beta = 1 / (\theta \cdot i')$, wherein i' is the number of the spectral lines being de-emphasized, θ is a factor between 3 and 5, in particular between 3,4 and 4,6, more particular between 3,8 and 4,2. Best results may be achieved when the second preset value is set to 4. Note, that the second preset value of the decoder should be the same as the second preset value of the encoder.

In a preferred embodiment of the invention the reference spectral line represents a frequency between 600 Hz and 1000 Hz, in particular between 700 Hz and 900 Hz, more particular between 750 Hz and 850 Hz. These empirically found intervals ensure sufficient low-frequency emphasis as well as a low computational complexity of the system. These intervals ensure in particular that in densely populated spectra, the lower-frequency lines are coded with sufficient accuracy. In a preferred embodiment the reference spectral line represents 800 Hz, wherein 32 spectral lines are de-emphasized. It is obvious that the reference spectral line of the decoder should represent the same frequency as the reference spectral line of the encoder.

In a preferred embodiment of the invention the further reference spectral line represents the same or a higher frequency than the reference spectral line. These features ensure that the estimation of the minimum and the maximum is done in the relevant frequency range, as is the case in the encoder.

In a preferred embodiment of the invention the control device is configured in such a way that the spectral lines of the reverse processed spectrum representing a lower frequency than the reference spectral line are de-emphasized only if the maximum is less than the minimum multiplied with the first preset value α. These features ensure that low-frequency de-emphasis is only executed when needed so that the work load of the decoder may be minimized and no bits are wasted on perceptually irrelevant regions during quantization.

In one aspect the invention provides a system comprising a decoder and an encoder, wherein the encoder is designed according to the invention and/or the decoder is designed according to the invention.

In one aspect the invention provides a method for encoding a non-speech audio signal so as to produce therefrom a bitstream, the method comprising the steps:

filtering with a linear predictive coding filter having a plurality of linear predictive coding coefficients and converting a frame of the audio signal into a frequency domain in order to output a spectrum based on the frame and on the linear predictive coding coefficients;

calculating a processed spectrum based on the spectrum of the filtered frame, wherein spectral lines of the processed spectrum representing a lower frequency than a reference spectral line are emphasized; and

controlling the calculation of the processed spectrum depending on the linear predictive coding coefficients of the linear predictive coding filter.

In one aspect the invention provides a method for decoding a bitstream based on a non-speech audio signal so as to produce from the bitstream a non-speech audio output signal, in particular for decoding a bitstream produced by the method according to the preceding claim, the bitstream containing quantized spectrums and a plurality of linear predictive coding coefficients, the method comprising the steps:

extracting the quantized spectrum and the linear predictive coding coefficients from the bitstream;

producing a de-quantized spectrum based on the quantized spectrum;

calculating a reverse processed spectrum based on the de-quantized spectrum, wherein spectral lines of the reverse processed spectrum representing a lower frequency than a reference spectral line are de-emphasized; and

controlling the calculation of the reverse processed spectrum depending on the linear predictive coding coefficients contained in the bitstream.

In one aspect the invention provides a computer program for performing, when running on a computer or a processor, the inventive method.

Preferred embodiments of the invention are subsequently discussed with respect to the accompanying drawings, in which:

Fig. 1a illustrates a first embodiment of an audio encoder according to the invention;

Fig. 1b illustrates a second embodiment of an audio encoder according to the invention;

Fig. 2 illustrates a first example for low-frequency emphasis executed by an audio encoder according to the invention;

Fig. 3 illustrates a second example for low-frequency emphasis executed by an audio encoder according to the invention;

Fig. 4 illustrates a third example for low-frequency emphasis executed by an audio encoder according to the invention;

Fig. 5a illustrates a first embodiment of an audio decoder according to the invention;

Fig. 5b illustrates a second embodiment of an audio decoder according to the invention;

Fig. 6 illustrates a first example for low-frequency de-emphasis executed by an audio decoder according to the invention;

Fig. 7 illustrates a second example for low-frequency de-emphasis executed by an audio decoder according to the invention; and

Fig. 8 illustrates a third example for low-frequency de-emphasis executed by an audio decoder according to the invention.

5 Fig. 1a illustrates a first embodiment of an audio encoder 1 according to the invention. The audio encoder 1 for encoding a non-speech audio signal AS so as to produce therefrom a bitstream BS comprises

a combination 2, 3 of a linear predictive coding filter 2 having a plurality of
10 linear predictive coding coefficients LC and a time-frequency converter 3, wherein the combination 2, 3 is configured to filter and to convert a frame FI of the audio signal AS into a frequency domain in order to output a spectrum SP based on the frame FI and on the linear predictive coding coefficients LC;

15 a low frequency emphasizer 4 configured to calculate a processed spectrum PS based on the spectrum SP, wherein spectral lines SL (see Fig. 2) of the processed spectrum PS representing a lower frequency than a reference spectral line RSL (see Fig.2) are emphasized; and

20 a control device 5 configured to control the calculation of the processed spectrum PS by the low frequency emphasizer 4 depending on the linear predictive coding coefficients LC of the linear predictive coding filter 2.

A linear predictive coding filter (LPC filter) 2 is a tool used in audio signal
25 processing and speech processing for representing the spectral envelope of a framed digital signal of sound in compressed form, using the information of a linear predictive model.

A time-frequency converter 3 is a tool for converting in particular a framed
30 digital signal from time domain into a frequency domain so as to estimate a spectrum of the signal. The time-frequency converter 3 may use a modified discrete cosine transform (MDCT), which is a lapped transform based on the

type-IV discrete cosine transform (DCT-IV), with the additional property of being lapped: it is designed to be performed on consecutive frames of a larger dataset, where subsequent frames are overlapped so that the last half of one frame coincides with the first half of the next frame. This overlapping, in addition to the energy-compaction qualities of the DCT, makes the MDCT especially attractive for signal compression applications, since it helps to avoid artifacts stemming from the frame boundaries.

The low frequency emphasizer 4 is configured to calculate a processed spectrum PS based on the spectrum SP of the filtered frame FF, wherein spectral lines SL of the processed spectrum PS representing a lower frequency than a reference spectral line RSL are emphasized so that only low frequencies contained in the processed spectrum PS are emphasized. The reference spectral line RSL may be predefined based on empirical experience.

The control device 5 is configured to control the calculation of the processed spectrum SP by the low frequency emphasizer 4 depending on the linear predictive coding coefficients LC of the linear predictive coding filter 2. Therefore, the encoder 1 according to the invention does not need to analyze the spectrum SP of the audio signal AS for the purpose of low-frequency emphasis. Further, since identical linear predictive coding coefficients LC may be used in the encoder 1 and in a subsequent decoder 12 (see Fig. 5), the adaptive low-frequency emphasis is fully invertible regardless of spectrum quantization as long as the linear predictive coding coefficients LC are transmitted to the decoder 12 in the bitstream BS which is produced by the encoder 1 or by any other means. In general the linear predictive coding coefficients LC have to be transmitted in the bitstream BS anyway for the purpose of reconstructing an audio output signal OS (see Fig. 5) from the bitstream BS by a respective decoder 12. Therefore, the bit rate of the bitstream BS will not be increased by the low-frequency emphasis as described herein.

The adaptive low-frequency emphasis system described herein may be implemented in the TCX core-coder of LD-USAC, a low-delay variant of xHE-AAC [4] which can switch between time-domain and MDCT-domain coding on a per-frame basis.

According to a preferred embodiment of the invention the frame FI of the audio signal AS is input to the linear predictive coding filter 2, wherein a filtered frame FF is output by the linear predictive coding filter 2 and wherein the time-frequency converter 3 is configured to estimate the spectrum SP based on the filtered frame FF. Accordingly, the linear predictive coding filter 2 may operate in the time domain, having the audio signal AS as its input.

According to a preferred embodiment of the invention the audio encoder 1 comprises a quantization device 6 configured to produce a quantized spectrum QS based on the processed spectrum BS and a bitstream producer 7 and configured to embed the quantized spectrum QS and the linear predictive coding coefficients LC into the bitstream BS. Quantization, in digital signal processing, is the process of mapping a large set of input values to a (countable) smaller set – such as rounding values to some unit of precision. A device or algorithmic function that performs quantization is called a quantization device 6. The bitstream producer 7 may be any device which is capable of embedding digital data from different sources 2, 6 into a unitary bitstream BS. By these features a bitstream BS produced with an adaptive low-frequency emphasis may be produced easily, wherein the adaptive low-frequency emphasis is fully invertible by a subsequent decoder 12 solely using information contained in the bitstream BS.

In a preferred embodiment of the invention the control device 5 comprises a spectral analyzer 8 configured to estimate a spectral representation SR of the linear predictive coding coefficients LC, a minimum-maximum analyzer 9 configured to estimate a minimum MI of the spectral representation SR and a maximum MA of the spectral representation SR below a further reference

spectral line and an emphasis factor calculator 10, 11 configured to calculate spectral line emphasis factors SEF for calculating the spectral lines SL of the processed spectrum PS representing a lower frequency than the reference spectral line RSL based on the minimum MI and on the maximum MA, wherein the spectral lines SL of the processed spectrum PS are emphasized by applying the spectral line emphasis factors SL to spectral lines of the spectrum SP of the filtered frame FF. The spectral analyzer may be a time-frequency converter as described above The spectral representation SR is the transfer function of the linear predictive coding filter 2. The spectral representation SR may be computed from an odd discrete Fourier transform (ODFT) of the linear predictive coding coefficients. In xHE-AAC and LD-USAC, the transfer function may be approximated by 32 or 64 MDCT-domain gains that cover the entire spectral representation SR.

In a preferred embodiment of the invention the emphasis factor calculator 10, 11 is configured in such way that the spectral line emphasis factors SEF increase in a direction from the reference spectral line RSL to the spectral line SL_0 representing the lowest frequency of the processed spectrum PS. That means that the spectral line SL_0 representing the lowest frequency is amplified the most whereas the spectral line $SL_{i'-1}$ adjacent to the reference spectral line is amplified the least. The reference spectral line RSL and spectral lines $SL_{i'+1}$ representing higher frequencies than the reference spectral line RSL are not emphasized at all. This reduces the computational complexity without any audible disadvantages.

In a preferred embodiment of the invention the emphasis factor calculator 10, 11 comprises a first stage 10 configured to calculate a basis emphasis factor BEF according to a first formula $\gamma = (\alpha \cdot \min / \max)^{\beta}$, wherein α is a first preset value, with $\alpha > 1$, β is a second preset value, with $0 < \beta \leq 1$, min is the minimum MI of the of the spectral representation SR, max is the maximum MA of the spectral representation SR and γ is the basis emphasis factor BEF, and wherein the emphasis factor calculator 10, 11 comprises a second stage

11 configured to calculate spectral line emphasis factors SEF according to a second formula $\varepsilon_i = \gamma^{i'-i}$, wherein i' is a number of the spectral lines SL to be emphasized, i is an index of the respective spectral line SL, the index increases with the frequencies of the spectral lines SL, with i = 0 to i'-1, γ is the basis emphasis factor BEF and ε_i is the spectral line emphasis factor SEF with index i. The basis emphasis factor is calculated from a ratio in the minimum and the maximum by the first formula in an easy way. The basis emphasis factor BEF serves as a basis for the calculation of all spectral line emphasis factors SEF, wherein the second formula ensures that the spectral line emphasis factors SEF increase in a direction from the reference spectral line RSL to the spectral line SL_0 representing the lowest frequency of the spectrum PS. In contrast to prior art solutions the proposed solution does not require a per-spectral-band square-root or similar complex operation. Only 2 division and 2 power operators are needed, one of each on encoder and decoder side.

In a preferred embodiment of the invention the first preset value is smaller than 42 and larger than 22, in particular smaller than 38 and larger than 26, more particular smaller 34 and larger than 30. The aforementioned intervals are based on empirical experiments. Best results may be achieved when the first preset value is set to 32.

In a preferred embodiment of the invention the second preset value is determined according to the formula $\beta = 1 / (\theta \cdot i')$, wherein i' is a number of the spectral lines SL being emphasized, θ is a factor between 3 and 5, in particular between 3,4 and 4,6, more particular between 3,8 and 4,2. Also these intervals are based on empirical experiments. It has been found the best results may be achieved than the second preset value is set to 4.

In a preferred embodiment of the invention the reference spectral line RSL represents a frequency between 600 Hz and 1000Hz, in particular between 700 Hz and 900 Hz, more particular between 750 Hz and 850 Hz. These em-

pirically found intervals ensure sufficient low-frequency emphasis as well as a low computational complexity of the system. These intervals ensure in particular that in densely populated spectra, the lower-frequency lines are coded with sufficient accuracy. In a preferred embodiment the reference spectral line represents 800 Hz, wherein 32 spectral lines are emphasized.

The calculation of the spectral line emphasis factors SEF may be done by the following income of the program code:

```
max = tmp = lpcGains[0];

/* find minimum (tmp) and maximum (max) of LPC gains in low frequencies */
for (i = 1; i < 9; i++) {

  if (tmp > lpcGains[i]) {
    tmp = lpcGains[i];
  }
  if (max < lpcGains[i]) {
    max = lpcGains[i];
  }
}
tmp *= 32.0f;
if ((max < tmp) && (max > FLT_MIN)) {
  fac = tmp = (float)pow(tmp / max, 0.0078125f);

  /* gradual boosting of lowest 32 bins; DC is boosted by (tmp/max)^1/4 */
  for (i = 31; i >= 0; i--) {
    x[i] *= fac;
    fac  *= tmp;
  }
}
```

In a preferred embodiment of the invention the further reference spectral line represents a higher frequency than the reference spectral line RSL. These features ensure that the estimation of the minimum MI and the maximum MA is done in the relevant frequency range.

Fig. 1b illustrates a second embodiment of an audio encoder 1 according to the invention. The second embodiment is based on the first embodiment. In the following only the differences between the two embodiments will be explained.

According to a preferred embodiment of the invention the frame FI of the audio signal AS is input to the time-frequency converter 3, wherein a converted frame CF is output by the time-frequency converter 3 and wherein the linear predictive coding filter 2 is configured to estimate the spectrum SP based on the converted frame CF. Alternatively but equivalently to the first embodiment of the inventive encoder 1having a low-frequency emphasizer, the encoder 1 may calculate a processed spectrum PS based on the spectrum SP of a frame FI produced by means of frequency-domain noise shaping (FDNS), as disclosed for example in [5]. More specifically, the tool ordering here is modified: the time-frequency converter 3 such as the above-mentioned one may be configured to estimate a converted frame FC based on the frame FI of the audio signal AS and the linear predictive coding filter 2 is configured to estimate the audio spectrum SP based on the converted frame FC, which is output by the time-frequency converter 3. Accordingly, the linear predictive coding filter 2 may operate in the frequency domain (instead of the time domain), having the converted frame FC as its input, with the linear predictive coding filter 2 applied via multiplication by a spectral representation of the linear predictive coding coefficients LC.

It should be evident to those skilled in the art that the first and the second embodiment– a linear filtering in the time domain followed by time-frequency conversion vs. time-frequency conversion followed by linear filtering via spectral weighting in the frequency domain – can be implemented such that they are equivalent.

Fig. 2 illustrates a first example for low-frequency emphasis executed by an encoder according to the invention. Fig. 2 shows an exemplary spectrum SP, exemplary spectral line emphasis factors SEF and an exemplary processed spectrum SP in a common coordinate system, wherein the frequency is plotted against the x-axis and amplitude depending on the frequency is plotted against the y-axis. The spectral lines SL_0 to $SL_{i'-1}$, which represents frequencies lower than the reference spectrum line RSL, are amplified, whereas the

reference spectral line RSL and the spectral line $SL_{i'+1}$, which represents a frequency higher than the reference spectrum RSL, are not amplified. Fig. 2 depicts a situation in which the ratio of the minimum MI and the maximum MA of the spectral representation SR of the linear predictive coding coefficients LC is close to 1. Therefore, a maximum spectral line emphasis factor SEF for the spectral line SL_0 is about 2.5.

Fig. 3illustrates a second example for low-frequency emphasis executed by an encoder according to the invention. The difference to the low-frequency emphasis as is stated in Fig. 2 is that the ratio of the minimum MI and the maximum MA of the spectral representation SR of the linear predictive coding coefficients LC is smaller. Therefore, a maximum spectral line emphasis factor SEF for the spectral line SL_0 is smaller, e.g. below 2.0.

Fig. 4 illustrates a third example for low-frequency emphasis executed by an encoder according to the invention. In the preferred embodiment of the invention the control device 5 is configured in such way that the spectral lines SL of the processed spectrum SP representing a lower frequency than the reference spectral RSL are emphasized only if the maximum is less than the minimum multiplied with the first preset value. These features ensure that low-frequency emphasis is only executed when needed so that the work load of the encoder may be minimized. In Fig. 4 these conditions are met so that no low-frequency emphasis executed.

Fig. 5 illustrates an embodiment of a decoder according to the invention. The audio decoder 12 is configured for decoding a bitstream BS based on a non-speech audio signal so as to produce from the bitstream BS a non-speech audio output signal OS, in particular for decoding a bitstream BS produced by an audio encoder 1 according to the invention, wherein the bitstream BS contains quantized spectrums QS and a plurality of linear predictive coding coefficient LC. The audio decoder 12 comprises:

a bitstream receiver 13 configured to extract the quantized spectrum QS and the linear predictive coding coefficients LC from the bitstream BS;

a de-quantization device 14 configured to produce a de-quantized spectrum DQ based on the quantized spectrum QS;

a low frequency de-emphasizer 15 configured to calculate a reverse processed spectrum RS based on the de-quantized spectrum DQ, wherein spectral lines SLD of the reverse processed spectrum RS representing a lower frequency than a reference spectral line RSLD are deemphasized; and

a control device 16 configured to control the calculation of the reverse processed spectrum RS by the low frequency de-emphasizer 15 depending on the linear predictive coding coefficients LC contained in the bitstream BS.

The bitstream receiver 13 may be any device which is capable of classifying digital data from a unitary bitstream BS so as to send the classified data to the appropriate subsequent processing stage. In particular the bitstream receiver 13 is configured to extract the quantized spectrum QS, which then is forwarded to the de-quantization device 14, and the linear predictive coding coefficients LC, which then are forwarded to the control device 16, from the bitstream BS.

The de-quantization device 16 is configured to produce a de-quantized spectrum DQ based on the quantized spectrum QS, wherein de-quantization is an inverse process with respect to quantization as explained above.

The low frequency de-emphasizer 15 is configured to calculate a reverse processed spectrum RS based on the de-quantized spectrum QS, wherein spectral lines SLD of the reverse processed spectrum RS representing a lower frequency than a reference spectral line RSLD are deemphasized so that only low frequencies contained in the reverse processed spectrum RS

are de-emphasized. The reference spectral line RSLD may be predefined based on empirical experience. It has to be noted that the reference spectral line RSLD of the decoder 12 should represent the same frequency as the reference spectral line RSL of the encoder 1 as explained above. However, the frequency to which the reference spectral line RSLD refers may be stored on the decoder side so that it is not necessary to transmit this frequency in the bitstream BS.

The control device 16 is configured to control the calculation of the reverse processed spectrum RS by the low frequency de-emphasizer 15 depending on the linear predictive coding coefficients LS of the linear predictive coding filter 2. Since identical linear predictive coding coefficients LC may be used in the encoder 1 producing the bitstream BS and in the decoder 12, the adaptive low-frequency emphasis is fully invertible regardless of spectrum quantization as long as the linear predictive coding coefficients are transmitted to the decoder 12 in the bitstream BS. In general the linear predictive coding coefficients LC have to be transmitted in the bitstream BS anyway for the purpose of reconstructing the audio output signal OS from the bitstream BS by the decoder 12. Therefore, the bit rate of the bitstream BS will not be increased by the low-frequency emphasis and the low-frequency de-emphasis as described herein.

The adaptive low-frequency de-emphasis system described herein may be implemented in the TCX core-coder of LD-USAC, a low-delay variant of xHE-AAC [4] which can switch between time-domain and MDCT-domain coding on a per-frame basis.

By these features a bitstream BS produced with an adaptive low-frequency emphasis may be decoded easily, wherein the adaptive low-frequency de-emphasis may be done by the decoder 12 solely using information contained in the bitstream BS.

According to a preferred embodiment of the invention the audio decoder 12 comprises combination 17, 18 of a frequency-time converter 17 and an inverse linear predictive coding filter 18 receiving the plurality of linear predictive coding coefficients LC contained in the bitstream BS, wherein the combination 17, 18 is configured to inverse-filter and to convert the reverse processed spectrum RS into a time domain in order to output the output signal OS based on the reverse processed spectrum RS and on the linear predictive coding coefficients LC.

A frequency-time converter 17 is a tool for executing an inverse operation of the operation of a time-frequency converter 3 as explained above. It is a tool for converting in particular a spectrum of a signal in a frequency domain into a framed digital signal in her time domain so as to estimate the original signal. The frequency-time converter may use an inverse modified discrete cosine transform (inverse MDCT), wherein the modified discrete cosine transform is a lapped transform based on the type-IV discrete cosine transform (DCT-IV), with the additional property of being lapped: it is designed to be performed on consecutive frames of a larger dataset, where subsequent frames are overlapped so that the last half of one frame coincides with the first half of the next frame. This overlapping, in addition to the energy-compaction qualities of the DCT, makes the MDCT especially attractive for signal compression applications, since it helps to avoid artifacts stemming from the frame boundaries. Those skilled in the art will understand that other transforms are possible. However, the transform in the decoder 12 should be an inverse transform of the transform in the encoder 1.

An inverse linear predictive coding filter 18 is a tool for executing an inverse operation to the operation done by the linear predictive coding filter (LPC filter) 2 as explained above. It is a tool used in audio signal and speech signal processing for decoding of the spectral envelope of a framed digital signal in order to reconstruct the digital signal, using the information of a linear predictive model. Linear predictive coding and decoding is fully invertible as known

as the same linear predictive coding coefficients used, which may be ensured by transmitting the linear predictive coding coefficients LC from the encoder 1 to the decoder 12 embedded in the bitstream BS as described herein.

By these features the output signal OS may be processed in an easy way.

According to a preferred embodiment of the invention the frequency-time converter 17 is configured to estimate a time signal TS based on the reverse processed spectrum RS, wherein the inverse linear predictive coding filter 18 is configured to output the output signal OS based on the time signal TS. Accordingly, the inverse linear predictive coding filter 18 may operate in the time domain, having the time signal TS as its input.

In a preferred embodiment of the invention the control device 16 comprises a spectral analyzer 19 configured to estimate a spectral representation SR of the linear predictive coding coefficients LC, a minimum-maximum analyzer 20 configured to estimate a minimum MI of the spectral representation SR and a maximum MA of the spectral representation SR below a further reference spectral line and a de-emphasis factor calculator 21, 22 configured to calculate spectral line de-emphasis factors SDF for calculating the spectral lines SLD of the reverse processed spectrum RS representing a lower frequency than the reference spectral line RSLD based on the minimum MI and on the maximum MA, wherein the spectral lines SLD of the reverse processed spectrum RS are de-emphasized by applying the spectral line de-emphasis factors SDF to spectral lines of the de-quantized spectrum DQ. The spectral analyzer may be a time-frequency converter as described above The spectral representation is the transfer function of the linear predictive coding filter. The spectral representation may be computed from an odd discrete Fourier transform (ODFT) of the linear predictive coding coefficients. In xHE-AAC and LD-USAC, the transfer function may be approximated by 32 or 64 MDCT-domain gains that cover the entire spectral representation.

In a preferred embodiment of the invention the de-emphasis factor calculator is configured in such way that the spectral line de-emphasis factors decrease in a direction from the reference spectral line to the spectral line representing the lowest frequency of the reverse process spectrum. This means that the spectral line representing the lowest frequency is attenuated the most whereas the spectral line adjacent to the reference spectral line is attenuated the least. The reference spectral line and spectral lines representing higher frequencies than the reference spectral line are not de-emphasized at all. This reduces the computational complexity without any audible disadvantages.

In a preferred embodiment of the invention the de-emphasis factor calculator 21, 22 comprises a first stage 21 configured to calculate a basis de-emphasis factor BDF according to a first formula $\delta = (\alpha \cdot \min / \max)^{-\beta}$, wherein α is a first preset value, with $\alpha > 1$, β is a second preset value, with $0 < \beta \leq 1$, min is the minimum MI of the of the spectral representation SR, max is the maximum MA of the spectral representation SR and δ is the basis de-emphasis factor BDF, and wherein the de-emphasis factor calculator 21, 22 comprises a second stage 22 configured to calculate spectral line de-emphasis factors SDF according to a second formula $\zeta_i = \delta^{i'-i}$, wherein i' is a number of the spectral lines SLD to be de-emphasized, i is an index of the respective spectral line SLD, the index increases with the frequencies of the spectral lines SLD, with i = 0 to i'-1, δ is the basis de-emphasis factor and ζ_i is the spectral line de-emphasis factor SDF with index i. The operation of the de-emphasis factor calculator 21, 22 is inverse to the operation of the emphasis factor calculator 10, 11 as described above. The basis de-emphasis factor BDF is calculated from a ratio in the minimum MI and the maximum MA by the first formula in an easy way. The basis de-emphasis factor BDF serves as a basis for the calculation of all spectral line de-emphasis factors SDF, wherein the second formula ensures that the spectral line de-emphasis factors SDF decrease in a direction from the reference spectral line RSLD to the spectral line SL_0 representing the lowest frequency of the reverse processed spectrum RS. In contrast to prior art solutions the proposed solution does not re-

quire a per-spectral-band square-root or similar complex operation. Only 2 division and 2 power operators are needed, one of each on encoder and decoder side.

In a preferred embodiment of the invention the first preset value is smaller than 42 and larger than 22, in particular smaller than 38 and larger than 26, more particular smaller 34 and larger than 30. The aforementioned intervals are based on empirical experiments. Best results may be achieved when the first preset value is set to 32. Note, that the first preset value of the decoder 12 should be the same as the first preset value of the encoder 1.

In a preferred embodiment of the invention the second preset value is determined according to the formula $\beta = 1 / (\theta \cdot i')$, wherein i' is the number of the spectral lines being de-emphasized, θ is a factor between 3 and 5, in particular between 3,4 and 4,6, more particular between 3,8 and 4,2. Best results may be achieved when the second preset value is set to 4. Note, that the second preset value of the decoder 12 should be the same as the second preset value of the encoder 1.

In a preferred embodiment of the invention the reference spectral line represents RSLD a frequency between 600 Hz and 1000Hz, in particular between 700 Hz and 900 Hz, more particular between 750 Hz and 850 Hz. These empirically found intervals ensure sufficient low-frequency emphasis as well as a low computational complexity of the system. These intervals ensure in particular that in densely populated spectra, the lower-frequency lines are coded with sufficient accuracy. In a preferred embodiment the reference spectral line RSLD represents 800 Hz, wherein 32 spectral lines SL are de-emphasized. It is obvious that the reference spectral line RSLD of decoder 12 should represent the same frequency than the reference spectral line RSL of the encoder.

The calculation of the spectral line emphasis factors SEF may be done by the following income of the program code:

```
max = tmp = lpcGains[0];

/* find minimum (tmp) and maximum (max) of LPC gains in low frequencies */
for (i = 1; i < 9; i++) {

  if (tmp > lpcGains[i]) {
    tmp = lpcGains[i];
  }
  if (max < lpcGains[i]) {
    max = lpcGains[i];
  }
}
tmp *= 32.0f;
if ((max < tmp) && (tmp > FLT_MIN)) {
  fac = tmp = (float)pow(max / tmp, 0.0078125f);

  /* gradual lowering of lowest 32 bins; DC is lowered by (max/tmp)^1/4 */
  for (i = 31; i >= 0; i--) {
    x[i] *= fac;

    fac  *= tmp;
  }
}
```

In a preferred embodiment of the invention the further reference spectral line represents the same or a higher frequency than the reference spectral line RSLD. These features ensure that the estimation of the minimum MI and the maximum MA is done in the relevant frequency range.

Fig. 5b illustrates a second embodiment of an audio decoder 12 according to the invention. The second embodiment is based on the first embodiment. In the following only the differences between the two embodiments will be explained.

According to a preferred embodiment of the invention the inverse linear predictive coding filter 18 is configured to estimate an inverse filtered signal IFS based on the reverse processed spectrum RS, wherein the frequency-time

converter 17 is configured to output the output signal OS based on the inverse filtered signal IFS.

Alternatively and equivalently, and analogous to the above-described FDNS procedure performed on the encoder side, the order of the frequency-time 17 converter and the inverse linear predictive coding filter 18 may be reversed such that the latter is operated first and in the frequency domain (instead of the time domain). More specifically, the inverse linear predictive coding filter 18 may output an inverse filtered signal IFS based on the reverse processed spectrum RS, with the inverse linear predictive coding filter 2 applied via multiplication (or division) by a spectral representation of the linear predictive coding coefficients LC, as in [5]. Accordingly, a frequency-time converter 17 such as the above-mentioned one may be configured to estimate a frame of the output signal OS based on the inverse filtered signal IFS, which is input to the time-frequency converter 17.

It should be evident to those skilled in the art that these two approaches – a linear inverse filtering in the frequency domain followed by frequency-time conversion vs. frequency-time conversion followed by linear filtering via spectral weighting in the time domain – can be implemented such that they are equivalent.

Fig. 6 illustrates a first example for low-frequency de-emphasis executed by a decoder according to the invention. Fig. 2 shows a de-quantized spectrum DQ, exemplary spectral line de-emphasis factors SDF and an exemplary of reverse processed spectrum RS in a common coordinate system, wherein the frequency is plotted against the x-axis and amplitude depending on the frequency is plotted against the y-axis. The spectral lines SLD_0 to $SLD_{i'-1}$, which represents frequencies lower than the reference spectrum line RSLD, are deemphasized, whereas the reference spectral line RSLD and the spectral line $SLD_{i'+1}$, which represents a frequency higher than the reference spectrum RSLD, are not deemphasize. Fig. 6 depicts a situation in which the

ratio of the minimum MI and the maximum MA of the spectral representation SR of the linear predictive coding coefficients LC is close to 1. Therefore, a maximum spectral line emphasis factor SEF for the spectral line SL_0 is about 0.4. Additionally Fig. 6 shows the quantization error QE, depending on the frequency. Due to the strong low-frequency de-emphasis the quantization error QE is very low at lower frequencies.

Fig. 7 illustrates a second example for low-frequency de-emphasis executed by a decoder according to the invention. The difference to the low-frequency emphasis as is stated in Fig. 6 is that the ratio of the minimum MI and the maximum MA of the spectral representation SR of the linear predictive coding coefficients LC is smaller. Therefore, a maximum spectral line de-emphasis factor SDF for the spectral line SL_0 is launcher, e.g. above 0.5. The quantization error QE is higher in this case but that is not critical as it is well below the amplitude of the reverse processed spectrum RS.

Fig. 8 illustrates a third example for low-frequency de-emphasis executed by a decoder according to the invention. In a preferred embodiment of the invention the control device 16 is configured in such way that the spectral lines SLD of the reverse processed spectrum RS representing a lower frequency than the reference spectral line RSLD are de-emphasized only if the maximum MA is less than the minimum MI multiplied with the first preset value. These features ensure that low-frequency de-emphasis is only executed when needed so that the work load of the decoder 12 may be minimized. These features ensure that low-frequency de-emphasis is only executed when needed so that the work load of the encoder may be minimized. In Fig. 8 these conditions are met so that no low-frequency emphasis executed.

As a solution to the above mentioned problem of relatively high complexity (possibly causing implementation issues on low-power mobile devices) and lack of perfect invertibility (risking sufficient fidelity) of the prior-art ALFE ap-

proach, a modified adaptive low-frequency emphasis (ALFE) design is proposed which

- does not require a per-spectral-band square-root or similar complex operation. Only 2 division and 2 power operators are needed, one of each on encoder and decoder side.
- utilizes a spectral representation of the LPC filter coefficients as control information for the (de)emphasis, not the spectrum itself. Since identical LPC coefficients are used in encoder and decoder, the ALFE is fully invertible regardless of spectrum quantization.

The ALFE system described herein was implemented in the TCX core-coder of LD-USAC, a low-delay variant of xHE-AAC [4] which can switch between time-domain and MDCT-domain coding on a per-frame basis. The process in encoder and decoder is summarized as follows:

1. In the encoder, the minimum and maximum of the spectral representation of the LPC coefficients is found below a certain frequency. The spectral representation of a filter generally adopted in signal processing is the filter's transfer function. In xHE-AAC and LD-USAC, the transfer function is approximated by 32 or 64 MDCT-domain gains that cover the entire spectrum, computed from an odd DFT (ODFT) of the filter coefficients.

2. If the maximum is greater than a certain global minimum (e.g. 0) and less than α times larger than the minimum, with $\alpha > 1$ (e.g. 32), the following 2 ALFE steps are executed.

3. A low-frequency emphasis factor γ is computed from the ratio between minimum and maximum as $\gamma = (\alpha \cdot \text{minimum} / \text{maximum})\beta$, where $0 < \beta \leq 1$ and β is dependent on α.

4. The MDCT lines with indices i lower than an index i' representing a certain frequency (i.e. all lines below that frequency, preferably the same frequency used in step 1) are now multiplied by $\gamma^{i'-i}$. This implies that the line closest to i' is amplified the least, while the first line, the one closest to direct current, is amplified the most. Preferably, i' = 32.

5. In the decoder, steps 1 and 2 are carried out like in the encoder (same frequency limit).

6. Analogous to step 3, a low-frequency de-emphasis factor, the inverse of the emphasis factor γ, is computed as $\delta = (\alpha \cdot \text{minimum} / \text{maximum})^{-\beta} = (\text{maximum} / (\alpha \cdot \text{minimum}))^{\beta}$.

7. The MDCT lines with indices i lower than index i', with i' chosen as in the encoder, are finally multiplied by $\delta^{i'-i}$. The result is that the line closest to i' is attenuated the least, the first line is attenuated the most, and overall the encoder-side ALFE is fully inverted.

Essentially, the proposed ALFE system ensures that in densely populated spectra, the lower-frequency lines are coded with sufficient accuracy. Three cases can serve to illustrate this, as depicted in Fig. 8. When the maximum is more than α times larger than the minimum, no ALFE is performed. This occurs when the low-frequency LPC shape contains a strong peak, probably originating from a strong isolated low-pitch tone in the input signal. LPC coders are typically able to reproduce such a signal relatively well, so an ALFE is not necessary.

In case the LPC shape is flat, i.e. the maximum approaches the minimum, the ALFE is the strongest as depicted in Fig. 6 and can avoid coding artifacts like musical noise.

When the LPC shape is neither fully flat nor peaky, e.g. on harmonic signals with closely spaced tones, only gentle ALFE is performed as depicted in Fig. 7. It must be noted that the application of the exponential factors γ in step 4 and δ in step 7 does not require power instructions but can be incrementally performed using only multiplications. Hence, the per-spectral-line complexity called for by the inventive ALFE scheme is very low.

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

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

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

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

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

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

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

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

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

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

A further embodiment according to the invention comprises an apparatus or a system configured to transfer (for example, electronically or optically) a computer program for performing one of the methods described herein to a receiver. The receiver may, for example, be a computer, a mobile device, a memory device or the like. The apparatus or system may, for example, comprise a file server for transferring the computer program to the receiver .

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

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

Reference signs:

1 audio encoder
2 linear predictive coding filter
3 time-frequency converter
4 low frequency emphasizer
5 control device
6 quantization device
7 bitstream producer
8 spectral analyzer

9	minimum-maximum analyzer
10	first stage of the emphasis factor calculator
11	second stage of the emphasis factor calculator
12	audio decoder
13	bitstream receiver
14	de-quantization device
15	low frequency de-emphasizer
16	control device is
17	frequency-time converter
18	inverse linear predictive coding filter
19	spectral analyzer
20	minimum-maximum analyzer
21	first stage of the de-emphasis factor calculator
22	second stage of the de-emphasis factor calculator

AS	audio signal
LC	linear predictive coding coefficients
FF	filtered frame
FI	frame
SP	spectrum
PS	processed spectrum
QS	quantized spectrum
SR	spectral representation
MI	minimum of the spectral representation
MA	maximum of the spectral representation
SEF	spectral line emphasis factors
BEF	phases emphasis factor
FC	frame converted to time domain
RSL	reference spectral line
SL	spectral line

DQ de-quantized spectrum
RS reverse processed spectrum
TS time signal
SDF spectral line de-emphasis factors
BDF basis de-emphasis factor
IFS inverse filtered signal

SLD spectral line
RSLD reference spectral line
QE quantization error

References:

[1] 3GPP TS 26.290, "Extended AMR Wideband Codec - Transcoding Functions," Dec. 2004.

[2] B. Bessette, U.S. Patent 7,933,769 B2, "Methods and devices for low-frequency emphasis during audio compression based on ACELP/TCX", Apr. 2011.

[3] J. Mäkinen et al., "AMR-WB+: A New Audio Coding Standard for 3rd Generation Mobile Audio Services," in Proc. ICASSP 2005, Philadelphia, USA, Mar. 2005.

[4] M. Neuendorf et al., "MPEG Unified Speech and Audio Coding – The ISO/MPEG Standard for High-Efficiency Audio Coding of All Content Types," in Proc. 132nd Convention of the AES, Budapest, Hungary, Apr. 2012. Also to appear in the Journal of the AES, 2013.

[5] T. Baeckstroem et al., European Patent EP 2 471 061 B1, "Multi-mode audio signal decoder, multi-mode audio signal encoder, methods and

computer program using linear prediction coding based noise shaping".

Claims

1. Audio encoder for encoding a non-speech audio signal (AS) so as to produce therefrom a bitstream (BS), the audio encoder (1) comprising:

   a combination (2, 3) of a linear predictive coding filter (2) having a plurality of linear predictive coding coefficients (LC) and a time-frequency converter (3), wherein the combination (2, 3) is configured to filter and to convert a frame (FI) of the audio signal (AS) into a frequency domain in order to output a spectrum (SP) based on the frame (FI) and on the linear predictive coding coefficients (LC);

   a low frequency emphasizer (4) configured to calculate a processed spectrum (PS) based on the spectrum (SP), wherein spectral lines (SL) of the processed spectrum (PS) representing a lower frequency than a reference spectral line (RSL) are emphasized; and

   a control device (5) configured to control the calculation of the processed spectrum (PS) by the low frequency emphasizer (4) depending on the linear predictive coding coefficients (LC) of the linear predictive coding filter (2).

2. Audio encoder according to the preceding claim, wherein the frame (FI) of the audio signal (AS) is input to the linear predictive coding filter (2), wherein a filtered frame (FF) is output by the linear predictive coding filter (2) and wherein the time-frequency converter (3) is configured to estimate the spectrum (SP) based on the filtered frame (FF).

3. Audio encoder according to claim 1, wherein the frame (FI) of the audio signal (AS) is input to the time-frequency converter (3), wherein a converted frame (FC) is output by the time-frequency converter (3) and

wherein the linear predictive coding filter (2) is configured to estimate the spectrum (SP) based on the converted frame (FC).

4. Audio encoder according to one of the preceding claims, wherein the audio encoder (1) comprises a quantization device (6) configured to produce a quantized spectrum (QS) based on the processed spectrum (PS) and a bitstream producer (7) configured to embed the quantized spectrum (QS) and the linear predictive coding coefficients (LC) into the bitstream (BS).

5. Audio encoder according to one of the preceding claims, wherein the control device (5) comprises a spectral analyzer (8) configured to estimate a spectral representation (SR) of the linear predictive coding coefficients (LC), a minimum-maximum analyzer (9) configured to estimate a minimum (MI) of the spectral representation (SR) and a maximum (MA) of the spectral representation (SR) below a further reference spectral line and an emphasis factor calculator (10, 11) configured to calculate spectral line emphasis factors (SEF) for calculating the spectral lines (SL) of the processed spectrum (PS) representing a lower frequency than the reference spectral line (RSL) based on the minimum (MI) and on the maximum (MA), wherein the spectral lines (SL) of the processed spectrum (PS) are emphasized by applying the spectral line emphasis factors (SEF) to spectral lines of the spectrum of the filtered frame.

6. Audio encoder according to the preceding claim, wherein the emphasis factor calculator (10, 11) is configured in such way that the spectral line emphasis factors (SEF) increase in a direction from the reference spectral line (RSL) to the spectral line (SL) representing the lowest frequency of the spectrum (SP).

7. Audio encoder according to claim 5 or 6, wherein the emphasis factor calculator (10, 11) comprises a first stage (10) configured to calculate a basis emphasis factor (BEF) according to a first formula $\gamma = (\alpha \cdot \min / \max)^{\beta}$,

wherein α is a first preset value, with $\alpha > 1$, β is a second preset value, with $0 < \beta \leq 1$, min is the minimum (MI) of the of the spectral representation (SR), max is the maximum (MA) of the spectral representation (SR) and γ is the basis emphasis factor (BEF), and wherein the emphasis factor calculator (10, 11) comprises a second stage (11) configured to calculate spectral line emphasis factors (SEF) according to a second formula $\varepsilon_i = \gamma^{i'-i}$, wherein i' is a number of the spectral lines (SL) to be emphasized, i is an index of the respective spectral line (SL), the index increases with the frequencies of the spectral lines, with i = 0 to i'-1, γ is the basis emphasis factor (BEF) and ε_i is the spectral line emphasis factor (SEF) with index i.

8. Audio encoder according to the preceding claim, wherein the first preset value is smaller than 42 and larger than 22, in particular smaller than 38 and larger than 26, more particular smaller 34 and larger than 30.

9. Audio encoder according to claim 7 or 8, wherein the second preset value is determined according to the formula $\beta = 1 / (\theta \cdot i')$, wherein i' is the number of the spectral lines being emphasized, θ is a factor between 3 and 5, in particular between 3,4 and 4,6, more particular between 3,8 and 4,2.

10. Audio encoder according to one of the preceding claims, wherein the reference spectral line (RSL) represents a frequency between 600 Hz and 1000Hz, in particular between 700 Hz and 900 Hz, more particular between 750 Hz and 850 Hz.

11. Audio encoder according to one of the claims 5 to 10, wherein the further reference spectral line represents the same or a higher frequency than the reference spectral line (RSL).

12. Audio encoder according to one of the preceding claims, wherein the control device (5) is configured in such way that the spectral lines (SL) of the processed spectrum (PS) representing a lower frequency than the reference spectral line (RSL) are emphasized only if the maximum (MA) is less than the minimum (MI) multiplied with the first preset value.

13. Audio decoder for decoding a bitstream (BS) based on a non-speech audio signal (AS) so as to produce from the bitstream (BS) a non-speech audio output signal (OS), in particular for decoding a bitstream (BS) produced by an audio encoder (1) according to claims 1 to 12, the bitstream (BS) containing quantized spectrums (QS) and a plurality of linear predictive coding coefficients (LC), the audio decoder (12) comprising:

a bitstream receiver (13) configured to extract the quantized spectrum (QS) and the linear predictive coding coefficients (LC) from the bitstream (BS);

a de-quantization device (14) configured to produce a de-quantized spectrum (DQ) based on the quantized spectrum (QS);

a low frequency de-emphasizer (15) configured to calculate a reverse processed spectrum (RS) based on the de-quantized spectrum (DQ), wherein spectral lines (SLD) of the reverse processed spectrum (RS) representing a lower frequency than a reference spectral line (RSLD) are deemphasized; and

a control device (16) configured to control the calculation of the reverse processed spectrum (RS) by the low frequency de-emphasizer (15) depending on the linear predictive coding coefficients (LC) contained in the bitstream (BS).

14. Audio decoder according to the preceding claim, wherein the audio decoder (12) comprises combination (17, 18) of a frequency-time converter (17) and an inverse linear predictive coding filter (18) receiving the plurality of linear predictive coding coefficients (LC) contained in the bitstream (BS), wherein the combination (17, 18) is configured to inverse-filter and to convert the reverse processed spectrum (RS) into a time domain in order to output the output signal (OS) based on the reverse processed spectrum (RS) and on the linear predictive coding coefficients (LC).

15. Audio decoder according to the preceding claim, wherein the frequency-time converter (17) is configured to estimate a time signal (TS) based on the reverse processed spectrum (RS) and wherein the inverse linear predictive coding filter (18) is configured to output the output signal (OS) based on the time signal (TS).

16. Audio decoder according to claim 14, wherein the inverse linear predictive coding filter (18) is configured to estimate an inverse filtered signal (IFS) based on the reverse processed spectrum (RS) and wherein the frequency-time converter (17) is configured to output the output signal (OS) based on the inverse filtered signal (IFS).

17. Audio decoder according to one of the claims 13 to 16, wherein the control device (16) comprises a spectral analyzer (19) configured to estimate a spectral representation (SR) of the linear predictive coding coefficients (LC), a minimum-maximum analyzer (20) configured to estimate a minimum (MI) of the spectral representation (SR) and a maximum (MA) of the spectral representation (SR) below a further reference spectral line and a de-emphasis factor calculator (21, 22) configured to calculate spectral line de-emphasis factors (SDF) for calculating the spectral lines (SLD) of the reverse processed spectrum (RS) representing a lower frequency than the reference spectral line (RSLD) based on the minimum (MI) and on the maximum (MA), wherein the spectral lines (SLD) of the reverse pro-

cessed spectrum (RS) are de-emphasized by applying the spectral line de-emphasis factors (SDF) to spectral lines of the spectrum of the de-quantized spectrum (DQ).

18. Audio decoder according to the preceding claim, wherein the de-emphasis factor calculator (21, 22) is configured in such way that the spectral line de-emphasis factors (SDF) decrease in a direction from the reference spectral line (RSLD) to the spectral line (SL) representing the lowest frequency of the reverse process spectrum (RS).

19. Audio decoder according to claim 17 or 18, wherein the de-emphasis factor calculator (21, 22) comprises a first stage (21) configured to calculate a basis de-emphasis factor (BDF) according to a first formula $\delta = (\alpha \cdot \min / \max)^{-\beta}$, wherein α is a first preset value, with $\alpha > 1$, β is a second preset value, with $0 < \beta \leq 1$, min is the minimum (MI) of the of the spectral representation (SR), max is the maximum (MA) of the spectral representation (SR) and δ is the basis de-emphasis factor (BDF), and wherein the de-emphasis factor calculator (21, 22) comprises a second stage (22) configured to calculate spectral line de-emphasis factors (SDF) according to a second formula $\zeta_i = \delta^{i'-i}$, wherein i' is a number of the spectral lines (SLD) to be de-emphasized, i is an index of the respective spectral line (SLD), the index increases with the frequencies of the spectral lines, with i = 0 to i'-1, δ is the basis de-emphasis factor (BDF) and ζ_i is the spectral line de-emphasis factor (SDF) with index i.

20. Audio decoder according to the preceding claim, wherein the first preset value is smaller than 42 and larger than 22, in particular smaller than 38 and larger than 26, more particular smaller 34 and larger than 30.

21. Audio decoder according to claim 19 or 20, wherein the second preset value is determined according to the formula $\beta = 1 / (\theta \cdot i')$, wherein i' is the number of the spectral lines (SLD) being de-emphasized, θ is a factor

between 3 and 5, in particular between 3,4 and 4,6, more particular between 3,8 and 4,2.

22. Audio decoder according to one of the claims 13 to 21, wherein the reference spectral line (RSLD) represents a frequency between 600 Hz and 1000Hz, in particular between 700 Hz and 900 Hz, more particular between 750 Hz and 850 Hz.

23. Audio decoder according to one of the claims 17 to 22, wherein the further reference spectral line represents the same or a higher frequency than the reference spectral line (RSLD).

24. Audio decoder according to one of the claims 13 to 23, wherein the control device (16) is configured in such way that the spectral lines (SLD) of the reverse processed spectrum (RS) representing a lower frequency than the reference spectral line (RSLD) are de-emphasized only if the maximum (MA) is less than the minimum (MI) multiplied with the first preset value.

25. A system comprising a decoder (1) and an encoder (12), wherein the encoder (1) is designed according to one of the claims 1 to 12 and/or the decoder is designed according to one of the claims 13 to 24.

26. Method for encoding a non-speech audio signal (AS) so as to produce therefrom a bitstream (BS), the method comprising the steps:

filtering with a linear predictive coding filter (2) having a plurality of linear predictive coding coefficients (LC) and converting a frame (FI) of the audio signal (AS) into a frequency domain in order to output a spectrum (SP) based on the frame (FI) and on the linear predictive coding coefficients (LC);

calculating a processed spectrum (PS) based on the spectrum (SP), wherein spectral lines (SL) of the processed spectrum (PS) representing a lower frequency than a reference spectral line (RSL) are emphasized; and

controlling the calculation of the processed spectrum (PS) depending on the linear predictive coding coefficients (LC) of the linear predictive coding filter (2).

27. Method for decoding a bitstream (BS) based on a non-speech audio signal (AS) so as to produce from the bitstream (BS) a non-speech audio output signal (OS), in particular for decoding a bitstream (BS) produced by the method according to the preceding claim, the bitstream (BS) containing quantized spectrums (QS) and a plurality of linear predictive coding coefficients (LC), the method comprising the steps:

extracting the quantized spectrum (QS) and the linear predictive coding coefficients (LC) from the bitstream (BS);

producing a de-quantized spectrum (DQ) based on the quantized spectrum (QS);

calculating a reverse processed spectrum (RS) based on the de-quantized spectrum (DQ), wherein spectral lines (SLD) of the reverse processed spectrum (RS) representing a lower frequency than a reference spectral line (RSLD) are deemphasized; and

controlling the calculation of the reverse processed spectrum (RS) depending on the linear predictive coding coefficients (LC) contained in the bitstream (BS).

28. Computer program for performing, when running on a computer or a processor, the method of claim 26 or 27.