SELECTIVE BASS POST FILTER

Technical field

The present invention generally relates to digital audio coding and more precisely to coding techniques for audio signals containing components of different characters.

Background

A widespread class of coding method for audio signals containing speech or singing includes *code excited linear prediction* (CELP) applied in time alternation with different coding methods, including frequency-domain coding methods especially adapted for music or methods of a general nature, to account for variations in character between successive time periods of the audio signal. For example, a simplified Moving Pictures Experts Group (MPEG) Unified Speech and Audio Coding (USAC; see standard ISO/IEC 23003-3) decoder is operable in at least three decoding modes, Advanced Audio Coding (AAC; see standard ISO/IEC 13818-7), algebraic CELP (ACELP) and transform-coded excitation (TCX), as shown in the upper portion of accompanying figure 2.

The various embodiments of CELP are adapted to the properties of the human organs of speech and, possibly, to the human auditory sense. As used in this application, CELP will refer to all possible embodiments and variants, including but not limited to ACELP, wide- and narrow-band CELP, SB-CELP (sub-band CELP), low- and high-rate CELP, RCELP (relaxed CELP), LD-CELP (low-delay CELP), CS-CELP (conjugate-structure CELP), CS-ACELP (conjugate-structure ACELP), PSI-CELP (pitch-synchronous innovation CELP) and VSELP (vector sum excited linear prediction). The principles of CELP are discussed by R. Schroeder and S. Atal in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 10, pp. 937–940, 1985, and some of its applications are described in references 25–29 cited in Chen and Gersho, *IEEE Transactions on*

Speech and Audio Processing, vol. 3, no. 1, 1995. As further detailed in the former paper, a CELP decoder (or, analogously, a CELP speech synthesizer) may include a pitch predictor, which restores the periodic component of an encoded speech signal, and an pulse codebook, from which an innovation sequence is added. The pitch predictor may in turn include a long-delay predictor for restoring the pitch and a short-delay predictor for restoring formants by spectral envelope shaping. In this context, the pitch is generally understood as the fundamental frequency of the tonal sound component produced by the vocal chords and further coloured by resonating portions of the vocal tract. This frequency together with its harmonics will dominate speech or singing. Generally speaking, CELP methods are best suited for processing solo or one-part singing, for which the pitch frequency is well-defined and relatively easy to determine.

To improve the perceived quality of CELP-coded speech, it is common practice to combine it with *post filtering* (or *pitch enhancement* by another term). U.S. Patent No. 4 969 192 and section II of the paper by Chen and Gersho disclose desirable properties of such post filters, namely their ability to suppress noise components located between the harmonics of the detected voice pitch (long-term portion; see section IV). It is believed that an important portion of this noise stems from the spectral envelope shaping. The long-term portion of a simple post filter may be designed to have the following transfer function:

$$H_E(z) = 1 + \alpha\left(\frac{z^T + z^{-T}}{2} - 1\right),$$

where T is an estimated pitch period in terms of number of samples and α is a gain of the post filter, as shown in figures 1 and 2. In a manner similar to a comb filter, such a filter attenuates frequencies 1/(2T), 3/(2T), 5/(2T), …, which are located midway between harmonics of the pitch frequency, and adjacent frequencies. The attenuation depends on the value of the gain α. Slightly more sophisticated post filters apply this attenuation only to low frequencies – hence the commonly used term *bass* post filter – where the noise is most perceptible. This can be expressed by cascading the transfer function

H_E described above and a low-pass filter H_{LP}. Thus, the post-processed decoded S_E provided by the post filter will be given, in the transform domain, by

$$S_E(z) = S(z) - \alpha S(z) P_{LT}(z) H_{LP}(z),$$

where

$$P_{LT}(z) = 1 - \frac{z^T + z^{-T}}{2}$$

and S is the decoded signal which is supplied as input to the post filter. Figure 3 shows an embodiment of a post filter with these characteristics, which is further discussed in section 6.1.3 of the Technical Specification ETSI TS 126 290, version 6.3.0, release 6. As this figure suggests, the pitch information is encoded as a parameter in the bit stream signal and is retrieved by a pitch tracking module communicatively connected to the long-term prediction filter carrying out the operations expressed by P_{LT}.

The long-term portion described in the previous paragraph may be used alone. Alternatively, it is arranged in series with a noise-shaping filter that preserves components in frequency intervals corresponding to the formants and attenuates noise in other spectral regions (short-term portion; see section III), that is, in the 'spectral valleys' of the formant envelope. As another possible variation, this filter aggregate is further supplemented by a gradual high-pass-type filter to reduce a perceived deterioration due to spectral tilt of the short-term portion.

Audio signals containing a mixture of components of different origins – e.g., tonal, non-tonal, vocal, instrumental, non-musical – are not always reproduced by available digital coding technologies in a satisfactory manner. It has more precisely been noted that available technologies are deficient in handling such non-homogeneous audio material, generally favouring one of the components to the detriment of the other. In particular, music containing singing accompanied by one or more instruments or choir parts which has been encoded by methods of the nature described above, will often be decoded with perceptible artefacts spoiling part of the listening experience.

Summary of the invention

In order to mitigate at least some of the drawbacks outlined in the previous section, it is an object of the present invention to provide methods and

devices adapted for audio encoding and decoding of signals containing a mixture of components of different origins. As particular objects, the invention seeks to provide such methods and devices that are suitable from the point of view of coding efficiency or (perceived) reproduction fidelity or both.

The invention achieves at least one of these objects by providing an encoder system, a decoder system, an encoding method, a decoding method and computer program products for carrying out each of the methods, as defined in the independent claims. The dependent claims define embodiments of the invention.

The inventors have realized that some artefacts perceived in decoded audio signals of non-homogeneous origin derive from an inappropriate switching between several coding modes of which at least one includes post filtering at the decoder and at least one does not. More precisely, available post filters remove not only interharmonic noise (and, where applicable, noise in spectral valleys) but also signal components representing instrumental or vocal accompaniment and other material of a 'desirable' nature. The fact that the *just noticeable difference* in spectral valleys may be as large as 10 dB (as noted by Ghitza and Goldstein, *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-4, pp. 697–708, 1986) may have been taken as a justification by many designers to filter these frequency bands severely. The quality degradation by the interharmonic (and spectral-valley) attenuation itself may however be less important than that of the switching occasions. When the post filter is switched on, the background of a singing voice sounds suddenly muffled, and when the filter is deactivated, the background instantly becomes more sonorous. If the switching takes place frequently, due to the nature of the audio signal or to the configuration of the coding device, there will be a switching artefact. As one example, a USAC decoder may be operable either in an ACELP mode combined with post filtering or in a TCX mode without post filtering. The ACELP mode is used in episodes where a dominant vocal component is present. Thus, the switching into the ACELP mode may be triggered by the onset of singing, such as at the beginning of a new musical phrase, at the beginning of a new verse, or simply after an episode where the accompaniment is deemed to drown the singing voice in the sense that the vocal com-

ponent is no longer prominent. Experiments have confirmed that an alternative solution, or rather circumvention of the problem, by which TCX coding is used throughout (and the ACELP mode is disabled) does not remedy the problem, as reverb-like artefacts appear.

Accordingly, in a first and a second aspect, the invention provides an audio encoding method (and an audio encoding system with the corresponding features) characterized by a decision being made as to whether the device which will decode the bit stream, which is output by the encoding method, should apply post filtering including attenuation of interharmonic noise. The outcome of the decision is encoded in the bit stream and is accessible to the decoding device.

By the invention, the decision whether to use the post filter is taken separately from the decision as to the most suitable coding mode. This makes it possible to maintain one post filtering status throughout a period of such length that the switching will not annoy the listener. Thus, the encoding method may prescribe that the post filter will be kept inactive even though it switches into a coding mode where the filter is conventionally active.

It is noted that the decision whether to apply post filtering is normally taken frame-wise. Thus, firstly, post filtering is not applied for less than one frame at a time. Secondly, the decision whether to disable post filtering is only valid for the duration of a current frame and may be either maintained or reassessed for the subsequent frame. In a coding format enabling a main frame format and a reduced format, which is a fraction of the normal format, e.g., 1/8 of its length, it may not be necessary to take post-filtering decisions for individual reduced frames. Instead, a number of reduced frames summing up to a normal frame may be considered, and the parameters relevant for the filtering decision may be obtained by computing the mean or median of the reduced frames comprised therein.

In a third and a fourth aspect of the invention, there is provided an audio decoding method (and an audio decoding system with corresponding features) with a decoding step followed by a post-filtering step, which includes interharmonic noise attenuation, and being characterized in a step of dis-

abling the post filter in accordance with post filtering information encoded in the bit stream signal.

A decoding method with these characteristics is well suited for coding of mixed-origin audio signals by virtue of its capability to deactivate the post filter in dependence of the post filtering information only, hence independently of factors such as the current coding mode. When applied to coding techniques wherein post filter activity is conventionally associated with particular coding modes, the post-filtering disabling capability enables a new operative mode, namely the unfiltered application of a conventionally filtered decoding mode.

In a further aspect, the invention also provides a computer program product for performing one of the above methods. Further still, the invention provides a post filter for attenuating interharmonic noise which is operable in either an active mode or a pass-through mode, as indicated by a post-filtering signal supplied to the post filter. The post filter may include a decision section for autonomously controlling the post filtering activity.

As the skilled person will appreciate, an encoder adapted to cooperate with a decoder is equipped with functionally equivalent modules, so as to enable faithful reproduction of the encoded signal. Such equivalent modules may be identical or similar modules or modules having identical or similar transfer characteristics. In particular, the modules in the encoder and decoder, respectively, may be similar or dissimilar processing units executing respective computer programs that perform equivalent sets of mathematical operations.

In one embodiment, encoding the present method includes decision making as to whether a post filter which further includes attenuation of spectral valleys (with respect to the formant envelope, see above). This corresponds to the short-term portion of the post filter. It is then advantageous to adapt the criterion on which the decision is based to the nature of the post filter.

One embodiment is directed to a encoder particularly adapted for speech coding. As some of the problems motivating the invention have been observed when a mixture of vocal and other components is coded, the com-

bination of speech coding and the independent decision-making regarding post filtering afforded by the invention is particularly advantageous. In particular, such a decoder may include a code-excited linear prediction encoding module.

In one embodiment, the encoder bases its decision on a detected simultaneous presence of a signal component with dominant fundamental frequency (pitch) and another signal component located below the fundamental frequency. The detection may also be aimed at finding the co-occurrence of a component with dominant fundamental frequency and another component with energy between the harmonics of this fundamental frequency. This is a situation wherein artefacts of the type under consideration are frequently encountered. Thus, if such simultaneous presence is established, the encoder will decide that post filtering is not suitable, which will be indicated accordingly by post filtering information contained in the bit stream.

One embodiment uses as its detection criterion the total signal power content in the audio time signal below a pitch frequency, possibly a pitch frequency estimated by a long-term prediction in the encoder. If this is greater than a predetermined threshold, it is considered that there are other relevant components than the pitch component (including harmonics), which will cause the post filter to be disabled.

In an encoder comprising a CELP module, use can be made of the fact that such a module estimates the pitch frequency of the audio time signal. Then, a further detection criterion is to check for energy content between or below the harmonics of this frequency, as described in more detail above.

As a further development of the preceding embodiment including a CELP module, the decision may include a comparison between an estimated power of the audio signal when CELP-coded (i.e., encoded and decoded) and an estimated power of the audio signal when CELP-coded and post-filtered. If the power difference is larger than a threshold, which may indicate that a relevant, non-noise component of the signal will be lost, and the encoder will decide to disable the post filter.

In an advantageous embodiment, the encoder comprises a CELP module and a TCX module. As is known in the art, TCX coding is advanta-

geous in respect of certain kinds of signals, notably non-vocal signals. It is not common practice to apply post-filtering to a TCX-coded signal. Thus, the encoder may select either TCX coding, CELP coding with post filtering or CELP coding without post filtering, thereby covering a considerable range of signal types.

As one further development of the preceding embodiment, the decision between the three coding modes is taken on the basis of a rate–distortion criterion, that is, applying an optimization procedure known *per se* in the art.

In another further development of the preceding embodiment, the encoder further comprises an Advanced Audio Coding (AAC) coder, which is also known to be particularly suitable for certain types of signals. Preferably, the decision whether to apply AAC (frequency-domain) coding is made separately from the decision as to which of the other (linear-prediction) modes to use. Thus, the encoder can be apprehended as being operable in two supermodes, AAC or TCX/CELP, in the latter of which the encoder will select between TCX, post-filtered CELP or non-filtered CELP. This embodiment enables processing of an even wider range of audio signal types.

In one embodiment, the encoder can decide that a post filtering at decoding is to be applied gradually, that is, with gradually increasing gain. Likewise, it may decide that post filtering is to be removed gradually. Such gradual application and removal makes switching between regimes with and without post filtering less perceptible. As one example, a singing episode, for which post-filtered CELP coding is found to be suitable, may be preceded by an instrumental episode, wherein TCX coding is optimal; a decoder according to the invention may then apply post filtering gradually at or near the beginning of the singing episode, so that the benefits of post filtering are preserved even though annoying switching artefacts are avoided.

In one embodiment, the decision as to whether post filtering is to be applied is based on an approximate difference signal, which approximates that signal component which is to be removed from a future decoded signal by the post filter. As one option, the approximate difference signal is computed as the difference between the audio time signal and the audio time signal when subjected to (simulated) post filtering. As another option, an encod-

ing section extracts an intermediate decoded signal, whereby the approximate difference signal can be computed as the difference between the audio time signal and the intermediate decoded signal when subjected to post filtering. The intermediate decoded signal may be stored in a long-term prediction buffer of the encoder. It may further represent the excitation of the signal, implying that further synthesis filtering (vocal tract, resonances) would need to be applied to obtain the final decoded signal. The point in using an intermediate decoded signal is that it captures some of the particularities, notably weaknesses, of the coding method, thereby allowing a more realistic estimation of the effect of the post filter. As a third option, a decoding section extracts an intermediate decoded signal, whereby the approximate difference signal can be computed as the difference between the intermediate decoded signal and the intermediate decoded signal when subjected to post filtering. This procedure probably gives a less reliable estimation than the two first options, but can on the other hand be carried out by the decoder in a standalone fashion.

The approximate difference signal thus obtained is then assessed with respect to one of the following criteria, which when settled in the affirmative will lead to a decision to disable the post filter:

a) whether the power of the approximate difference signal exceeds a predetermined threshold, indicating that a significant part of the signal would be removed by the post filter;

b) whether the character of the approximate difference signal is rather tonal than noise-like;

c) whether a difference between magnitude frequency spectra of the approximate difference signal and of the audio time signal is unevenly distributed with respect to frequency, suggesting that it is not noise but rather a signal that would make sense to a human listener;

d) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a predetermined relevance envelope, based on what can usually be expected from a signal of the type to be processed; and

e) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a relevance envelope obtained by thresholding a magnitude frequency spectrum of the audio time signal by a magnitude of the largest signal component therein downscaled by a predetermined scale factor.

When evaluating criterion e), it is advantageous to apply peak tracking in the magnitude spectrum, that is, to distinguish portions having peak-like shapes normally associated with tonal components rather than noise. Components identified by peak tracking, which may take place by some algorithm known *per se* in the art, may be further sorted by applying a threshold to the peak height, whereby the remaining components are tonal material of a certain magnitude. Such components usually represent relevant signal content rather than noise, which motivates a decision to disable the post filter.

In one embodiment of the invention as a decoder, the decision to disable the post filter is executed by a switch controllable by the control section and capable of bypassing the post filter in the circuit. In another embodiment, the post filter has variable gain controllable by the control section, or a gain controller therein, wherein the decision to disable is carried out by setting the post filter gain (see previous section) to zero or by setting its absolute value below a predetermined threshold.

In one embodiment, decoding according to the present invention includes extracting post filtering information from the bit stream signal which is being decoded. More precisely, the post filtering information may be encoded in a data field comprising at least one bit in a format suitable for transmission. Advantageously, the data field is an existing field defined by an applicable standard but not in use, so that the post filtering information does not increase the payload to be transmitted.

It is noted that the methods and apparatus disclosed in this section may be applied, after appropriate modifications within the skilled person's abilities including routine experimentation, to coding of signals having several components, possibly corresponding to different channels, such as stereo channels. Throughout the present application, pitch enhancement and post filtering are used as synonyms. It is further noted that AAC is discussed as a

representative example of frequency-domain coding methods. Indeed, applying the invention to a decoder or encoder operable in a frequency-domain coding mode other than AAC will only require small modifications, if any, within the skilled person's abilities. Similarly, TCX is mentioned as an example of weighted linear prediction transform coding and of transform coding in general.

Features from two or more embodiments described hereinabove can be combined, unless they are clearly complementary, in further embodiments. The fact that two features are recited in different claims does not preclude that they can be combined to advantage. Likewise, further embodiments can also be provided by the omission of certain features that are not necessary or not essential for the desired purpose.

Brief description of the drawings

Embodiments of the present invention will now be described with reference to the accompanying drawings, on which:

figures 1 is a block diagram showing a conventional decoder with post filter;

figure 2 is a schematic block diagram of a conventional decoder operable in AAC, ACELP and TCX mode and including a post filter permanently connected downstream of the ACELP module;

figure 3 is a block diagram illustrating the structure of a post filter;

figures 4 and 5 are block diagrams of two decoders according to the invention;

figure 6 and 7 are block diagrams illustrating differences between a conventional decoder (figure 6) and a decoder (figure 7) according to the invention;

figure 8 is a block diagram of an encoder according to the invention;

figures 9 and 10 are a block diagrams illustrating differences between a conventional decoder (figure 9) and a decoder (figure 10) according to the invention; and

figure 11 is a block diagram of an autonomous post filter which can be selectively activated and deactivated.

Detailed description of embodiments

Figure 4 is a schematic drawing of a decoder system 400 according to an embodiment of the invention, having as its input a bit stream signal and as its output an audio signal. As in the conventional decoders shown in figure 1, a post filter 440 is arranged downstream of a decoding module 410 but can be switched into or out of the decoding path by operating a switch 442. The post filter is enabled in the switch position shown in the figure. It would be disabled if the switch was set in the opposite position, whereby the signal from the decoding module 410 would instead be conducted over the bypass line 444. As an inventive contribution, the switch 442 is controllable by post filtering information contained in the bit stream signal, so that post filtering may be applied and removed irrespectively of the current status of the decoding module 410. Because a post filter 440 operates at some delay – for example, the post filter shown in figure 3 will introduce a delay amounting to at least the pitch period T – a compensation delay module 443 is arranged on the bypass line 444 to maintain the modules in a synchronized condition at switching. The delay module 443 delays the signal by the same period as the post filter 440 would, but does not otherwise process the signal. To minimize the change-over time, the compensation delay module 443 receives the same signal as the post filter 440 at all times. In an alternative embodiment where the post filter 440 is replaced by a zero-delay post filter (e.g., a causal filter, such as a filter with two taps, independent of future signal values), the compensation delay module 443 can be omitted.

Figure 5 illustrates a further development according to the teachings of the invention of the triple-mode decoder system 500 of figure 2. An ACELP decoding module 511 is arranged in parallel with a TCX decoding module 512 and an AAC decoding module 513. In series with the ACELP decoding module 511 is arranged a post filter 540 for attenuating noise, particularly noise located between harmonics of a pitch frequency directly or indirectly derivable from the bit stream signal for which the decoder system 500 is adapted. The bit stream signal also encodes post filtering information governing the positions of an upper switch 541 operable to switch the post filter 540 out of the

processing path and replace it with a compensation delay 543 like in figure 4. A lower switch 542 is used for switching between different decoding modes. With this structure, the position of the upper switch 541 is immaterial when one of the TCX or AAC modules 512, 513 is used; hence, the post filtering information does not necessary indicate this position except in the ACELP mode. Whatever decoding mode is currently used, the signal is supplied from the downstream connection point of the lower switch 542 to a spectral band replication (SBR) module 550, which outputs an audio signal. The skilled person will realize that the drawing is of a conceptual nature, as is clear notably from the switches which are shown schematically as separate physical entities with movable contacting means. In a possible realistic implementation of the decoder system, the switches as well as the other modules will be embodied by computer-readable instructions.

Figures 6 and 7 are also block diagrams of two triple-mode decoder systems operable in an ACELP, TCX or frequency-domain decoding mode. With reference to the latter figure, which shows an embodiment of the invention, a bit stream signal is supplied to an input point 701, which is in turn permanently connected via respective branches to the three decoding modules 711, 712, 713. The input point 701 also has a connecting branch 702 (not present in the conventional decoding system of figure 6) to a pitch enhancement module 740, which acts as a post filter of the general type described above. As is common practice in the art, a first transition windowing module 703 is arranged downstream of the ACELP and TCX modules 711, 712, to carry out transitions between the decoding modules. A second transition module 704 is arranged downstream of the frequency-domain decoding module 713 and the first transition windowing module 703, to carry out transition between the two super-modes. Further a SBR module 750 is provided immediately upstream of the output point 705. Clearly, the bit stream signal is supplied directly (or after demultiplexing, as appropriate) to all three decoding modules 711, 712, 713 and to the pitch enhancement module 740. Information contained in the bit stream controls what decoding module is to be active. By the invention however, the pitch enhancement module 740 performs an analogous self actuation , which responsive to post filtering information in the

bit stream may act as a post filter or simply as a pass-through. This may for instance be realized through the provision of a control section (not shown) in the pitch enhancement module 740, by means of which the post filtering action can be turned on or off. The pitch enhancement module 740 is always in its pass-through mode when the decoder system operates in the frequency-domain or TCX decoding mode, wherein strictly speaking no post filtering information is necessary. It is understood that modules not forming part of the inventive contribution and whose presence is obvious to the skilled person, e.g., a demultiplexer, have been omitted from figure 7 and other similar drawings to increase clarity.

As a variation, the decoder system of figure 7 may be equipped with a control module (not shown) for deciding whether post filtering is to be applied using an analysis-by-synthesis approach. Such control module is communicatively connected to the pitch enhancement module 740 and to the ACELP module 711, from which it extracts an intermediate decoded signal $s_{i_DEC}(n)$ representing an intermediate stage in the decoding process, preferably one corresponding to the excitation of the signal. The detection module has the necessary information to simulate the action of the pitch enhancement module 740, as defined by the transfer functions $P_{LT}(z)$ and $H_{LP}(z)$ (cf. *Background* section and figure 3), or equivalently their filter impulse responses $p_{LT}(z)$ and $h_{LP}(n)$. As follows by the discussion in the *Background* section, the component to be subtracted at post filtering can be estimated by an approximate difference signal $s_{AD}(n)$ which is proportional to $[(s_{i_DEC} * p_{LT}) * h_{LP}](n)$, where * denotes discrete convolution. This is an approximation of the true difference between the original audio signal and the post-filtered decoded signal, namely

$$s_{ORIG}(n) - s_E(n) = s_{ORIG}(n) - (s_{DEC}(n) - \alpha[s_{DEC} * p_{LT} * h_{LP}](n)),$$

where α is the post filter gain. By studying the total energy, low-band energy, tonality, actual magnitude spectrum or past magnitude spectra of this signal, as disclosed in the *Summary* section and the claims, the control section may find a basis for the decision whether to activate or deactivate the pitch enhancement module 740.

Figure 8 shows an encoder system 800 according to an embodiment of the invention. The encoder system 800 is adapted to process digital audio

signals, which are generally obtained by capturing a sound wave by a microphone and transducing the wave into an analog electric signal. The electric signal is then sampled into a digital signal susceptible to be provided, in a suitable format, to the encoder system 800. The system generally consists of an encoding module 810, a decision module 820 and a multiplexer 830. By virtue of switches 814, 815 (symbolically represented), the encoding module 810 is operable in either a CELP, a TCX or an AAC mode, by selectively activating modules 811, 812, 813. The decision module 820 applies one or more predefined criteria to decide whether a bit stream signal produced by the encoder system 800 to encode an audio signal. For this purpose, the decision module 820 may examine the audio signal directly or may receive data from the encoding module 810 via a connection line 816. A signal indicative of the decision taken by the decision module 820 is provided, together with the encoded audio signal from the encoding module 810, to a multiplexer 830, which concatenates the signals into a bit stream constituting the output of the encoder system 800.

Preferably, the decision module 820 bases its decision on an approximate difference signal computed from an intermediate decoded signal s_{i_DEC}, which can be subtracted from the encoding module 810. The intermediate decoded signal represents an intermediate stage in the *decoding* process, as discussed in preceding paragraphs, but may be extracted from a corresponding stage of the *encoding* process. However, in the encoder system 800 the original audio signal s_{ORIG} is available so that, advantageously, the approximate difference signal is formed as:

$$s_{ORIG}(n) - (s_{i_DEC}(n) - \alpha\,[(s_{i_DEC} * p_{LT}) * h_{LP}](n)).$$

The approximation resides in the fact that the intermediate decoded signal is used in lieu of the final decoded signal. This enables an appraisal of the nature of the component that a post filter would remove at decoding, and by applying one of the criteria discussed in the *Summary* section, the decision module 820 will be able to take a decision whether to disable post filtering.

As a variation to this, the decision module 820 may use the original signal in place of an intermediate decoded signal, so that the approximate difference signal will be $[(s_{i_DEC} * p_{LT}) * h_{LP}](n)$. This is likely to be a less faith-

ful approximation but on the other hand makes the presence of a connection line 816 between the decision module 820 and the encoding module 810 optional.

In such other variations of this embodiment where the decision module 820 studies the audio signal directly, one or more of the following criteria may be applied:

- Does the audio signal contain both a component with dominant fundamental frequency and a component located below the fundamental frequency? (The fundamental frequency may be supplied as a by-product of the encoding module 810.)
- Does the audio signal contain both a component with dominant fundamental frequency and a component located between the harmonics of the fundamental frequency?
- Does the audio signal contain significant signal energy below the fundamental frequency?
- Is post-filtered decoding (likely to be) preferable to unfiltered decoding with respect to rate–distortion optimality?

In all the described variations of the encoder structure shown in figure 8 – that is, irrespectively of the basis of the detection criterion – the decision section 820 may be enabled to decide on a gradual onset or gradual removal of post filtering, so as to achieve smooth transitions. The gradual onset and removal may be controlled by adjusting the post filter gain.

Figure 9 shows a conventional decoder operable in a frequency-decoding mode and a CELP decoding mode depending on the bit stream signal supplied to the decoder. Post filtering is applied whenever the CELP decoding mode is selected. An improvement of this decoder is illustrated in figure 10, which shows an decoder 1000 according to an embodiment of the invention. This decoder is operable not only in a frequency-domain-based decoding mode, wherein the frequency-domain decoding module 1013 is active, and a filtered CELP decoding mode, wherein the CELP decoding module 1011 and the post filter 1040 are active, but also in an unfiltered CELP mode, in which the CELP module 1011 supplies its signal to a compensation delay module 1043 via a bypass line 1044. A switch 1042 controls what decoding

mode is currently used responsive to post filtering information contained in the bit stream signal provided to the decoder 1000. In this decoder and that of figure 9, the last processing step is effected by an SBR module 1050, from which the final audio signal is output.

Figure 11 shows a post filter 1100 suitable to be arranged downstream of a decoder 1199. The filter 1100 includes a post filtering module 1140, which is enabled or disabled by a control module (not shown), notably a binary or non-binary gain controller, in response to a post filtering signal received from a decision module 1120 within the post filter 1100. The decision module performs one or more tests on the signal obtained from the decoder to arrive at a decision whether the post filtering module 1140 is to be active or inactive. The decision may be taken along the lines of the functionality of the decision module 820 in figure 8, which uses the original signal and/or an intermediate decoded signal to predict the action of the post filter. The decision of the decision module 1120 may also be based on similar information as the decision modules uses in those embodiments where an intermediate decoded signal is formed. As one example, the decision module 1120 may estimate a pitch frequency (unless this is readily extractable from the bit stream signal) and compute the energy content in the signal below the pitch frequency and between its harmonics. If this energy content is significant, it probably represents a relevant signal component rather than noise, which motivates a decision to disable the post filtering module 1140.

A 6-person listening test has been carried out, during which music samples encoded and decoded according to the invention were compared with reference samples containing the same music coded while applying post filtering in the conventional fashion but maintaining all other parameters unchanged. The results confirm a perceived quality improvement.

Further embodiments of the present invention will become apparent to a person skilled in the art after reading the description above. Even though the present description and drawings disclose embodiments and examples, the invention is not restricted to these specific examples. Numerous modifications and variations can be made without departing from the scope of the present invention, which is defined by the accompanying claims.

The systems and methods disclosed hereinabove may be implemented as software, firmware, hardware or a combination thereof. Certain components or all components may be implemented as software executed by a digital signal processor or microprocessor, or be implemented as hardware or as an application-specific integrated circuit. Such software may be distributed on computer readable media, which may comprise computer storage media (or non-transitory media) and communication media (or transitory media). As is well known to a person skilled in the art, computer storage media includes both volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by a computer. Further, it is well known to the skilled person that communication media typically embodies computer readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media.

List of embodiments

1. A decoder system (400; 500; 700; 1000) for decoding a bit stream signal as an audio time signal, including:

a decoding section (410; 511, 512, 513; 711, 712, 713; 1011, 1013) for decoding a bit stream signal as a preliminary audio time signal; and

an interharmonic noise attenuation post filter (440; 540; 740; 1040) for filtering the preliminary audio time signal to obtain an audio time signal,

characterized by a control section adapted to disable the post filter responsive to post-filtering information encoded in the bit stream signal, wherein the preliminary audio time signal is output as the audio time signal.

2. The decoder system of embodiment 1, wherein the post filter is further adapted to attenuate noise located in spectral valleys.

3. The decoder system of embodiment 1, wherein the control section includes a switch (442; 541; 1042) for selectively excluding the post filter from the signal processing path of the decoder system, whereby the post filter is disabled.

4. The decoder system of embodiment 1, wherein the post filter has variable gain determining the interharmonic attenuation and the control section includes a gain controller operable to set the absolute value of the gain below a predetermined threshold, whereby the post filter is disabled.

5. The decoder system of embodiment 1, said decoding section including a speech decoding module.

6. The decoder system of embodiment 1, said decoding section including a code-excited linear prediction, CELP, decoding module (511; 711; 1011).

7. The decoder system of embodiment 5, wherein a pitch frequency estimated by a long-term prediction section in the encoder is encoded in the bit stream signal.

8. The decoder system of embodiment 7, wherein the post filter is adapted to attenuate spectral components located between harmonics of the pitch frequency.

9. The decoder system of embodiment 1, wherein the bit stream signal contains a representation of a pitch frequency and the post filter is adapted to attenuate spectral components located between harmonics of the pitch frequency.

10. The decoder system of embodiment 8 or 9, wherein the post filter is adapted to attenuate only such spectral components which are located below a predetermined cut-off frequency.

11. The decoder system of embodiment 6,

the decoding section further comprising a transform-coded excitation, TCX, decoding module (512; 712) for decoding a bit stream signal as an audio time signal,

the control section being adapted operate the decoder system in at least the following modes:

a) the TCX module is enabled and the post filter is disabled;

b) the CELP module and the post filter are enabled; and

c) the CELP module is enabled and the post filter is disabled, wherein the preliminary audio time signal and the audio time signal coincide.

12. The decoder system of embodiment 10,

the decoding section further comprising an Advanced Audio Coding, AAC, decoding module (513; 713) for decoding a bit stream signal as an audio time signal,

the control section being adapted to operate the decoder also in the following mode:

d) the AAC module is enabled and the post filter is disabled.

13. The decoder system of embodiment 1, wherein the bit stream signal is segmented into time frames and the control section is adapted to disable an entire time frame or a sequence of entire time frames.

14. The decoder system of embodiment 13, wherein the control section is further adapted to receive, for each time frame in a Moving Pictures Experts Group, MPEG, bit stream, a data field associated with this time frame and is operable, responsive to the value of the data field, to disable the post filter.

15. The decoder system of embodiment 4, wherein the control section is adapted to decrease and/or increase the gain of the post filter gradually.

16. A decoder system (400; 500; 700; 1000) comprising:

a decoding section (410; 511, 512, 513; 711, 712, 713; 1011, 1013) for decoding a bit stream signal as a preliminary audio time signal; and

an interharmonic noise attenuation post filter (440; 540; 740; 1040) for filtering the preliminary audio time signal to obtain an audio time signal,

characterized in that

the decoding section is adapted to generate an intermediate decoded signal representing excitation and to provide this to the control section; and

the control section is adapted to compute an approximate difference signal, which approximates the signal component which is to be removed from the decoded signal by the post filter, as a difference between the intermediate decoded signal and the intermediate decoded signal when subjected to post filtering and to assess at least one of the following criteria:

a) whether the power of the approximate difference signal exceeds a predetermined threshold;

b) whether the character of the approximate difference signal is tonal;

c) whether a difference between magnitude frequency spectra of the approximate difference signal and of the audio time signal is unevenly distributed with respect to frequency;

d) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a predetermined relevance envelope; and

e) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a relevance envelope obtained by thresholding a magnitude frequency spectrum of the audio time signal by a magnitude of the largest signal component therein downscaled by a predetermined scale factor;

and, responsive to a positive determination, to disable the post filter, whereby the preliminary audio time signal is output as the audio time signal.

17. An interharmonic noise attenuation post filter (440; 550; 740; 1040; 1140) adapted to receive an input signal, which comprises a preliminary audio signal, and to supply an output audio signal,

characterized by a control section for selectively, in accordance with the value of a post-filtering signal, operating the post filter in one of the following modes:

i) a filtering mode, wherein it filters the preliminary audio signal to obtain a filtered signal and supplies this as output audio signal; and

ii) a pass-through mode, wherein it supplies the preliminary audio signal as output audio signal.

18. The post filter of embodiment 17, wherein the post-filtering signal is included in the input signal.

19. The post filter of embodiment 17, further comprising a decision module (1120) adapted to estimate a pitch frequency of the preliminary audio signal and to assess at least one of the following criteria:

a) whether the power of spectral components below the pitch frequency exceed a predetermined threshold;

b) whether spectral components below the pitch frequency are tonal;

c) whether the power of spectral components between harmonics of the pitch frequency exceed a predetermined threshold; and

d) whether spectral components between harmonics of the pitch frequency are tonal;

and, responsive to a positive determination, to take a decision to generate a negative post-filtering signal disabling the post filter.

20. A method of decoding a bit stream signal as an audio time signal, including the steps of:

decoding a bit stream signal as a preliminary audio time signal; and

post-filtering the preliminary audio time signal by attenuating interharmonic noise, thereby obtaining an audio time signal,

characterized in that the post-filtering step is selectively omitted responsive to post-filtering information encoded in the bit stream signal.

21. The method of embodiment 20, wherein the step of post-filtering further includes attenuating noise located in spectral valleys.

22. The method of embodiment 20, wherein the decoding step includes applying a coding method adapted for speech coding.

23. The method of embodiment 20, wherein the decoding step includes applying code-excited linear prediction, CELP, decoding.

24. The method of embodiment 22 or 23, wherein the post-filtering step includes attenuating spectral components located between harmonics of the pitch frequency, the pitch frequency being extracted from the bit stream signal or estimated in the decoding step.

25. The method of embodiment 20, wherein the post-filtering step includes attenuating only such spectral components which are located below a predetermined cut-off frequency.

26. The method of embodiment 23, wherein the steps of decoding and post-filtering selectively perform one of the following:

a) TCX decoding;

b) CELP decoding with post filtering; and

c) CELP decoding without post filtering.

27. The method of embodiment 26, wherein the steps of decoding and post-filtering selectively perform one of modes a), b), c) and

d) Advanced Audio Coding, AAC, decoding.

28. The method of embodiment 20, wherein the bit stream signal is segmented into time frames and the post-filtering step is omitted for an entire time frame or a sequence of entire time frames.

29. The method of embodiment 28, wherein:

the bit stream signal is a Moving Pictures Experts Group, MPEG, bit stream and includes, for each time frame, an associated data field; and

the post-filtering step is omitted in a time frame responsive to the value of the associated data field.

30. The method of embodiment 20, wherein said omission of the post-filtering includes one of the following: full omission of attenuation,

partial omission of attenuation,

gradually increasing attenuation, and

gradually decreasing attenuation.

31. A method of decoding a bit stream signal as an audio time signal, including the steps of:

decoding a bit stream signal as a preliminary audio time signal; and

post-filtering the preliminary audio time signal by attenuating interharmonic noise, thereby obtaining an audio time signal,

characterized in that the step of decoding includes:

extracting an intermediate decoded signal representing excitation;

computing an approximate difference signal, which approximates the signal component which is to be removed from the decoded signal by the post filter, as a difference between the intermediate decoded signal and the intermediate decoded signal when subjected to post filtering;

assessing at least one of the following criteria:

a) whether the power of the approximate difference signal exceeds a predetermined threshold;

b) whether the character of the approximate difference signal is tonal;

c) whether a difference between magnitude frequency spectra of the approximate difference signal and of the audio time signal is unevenly distributed with respect to frequency;

d) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a predetermined relevance envelope;

e) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a relevance envelope obtained by thresholding a magnitude frequency spectrum of the audio time signal by a magnitude of the largest signal component therein downscaled by a predetermined scale factor;

and, responsive to a positive determination, to disable the post filter, whereby the preliminary audio signal is output as the audio time signal.

32. An encoder system (800) for encoding an audio time signal as a bit stream signal, including an encoding section (810) for encoding an audio time signal as a bit stream signal,

characterized by a decision section (820) adapted to decide whether post filtering, which includes attenuation of interharmonic noise, is to be disabled at decoding of the bit stream signal and to encode this decision in the bit stream signal as post filtering information.

33. The encoder system of embodiment 32, the decision section being adapted to decide whether to disable post filtering which further includes attenuation of noise located in spectral valleys.

34. The encoder system of embodiment 32, the encoding section including a speech coding module.

35. The encoder system of embodiment 32, the encoding section including a code-excited linear prediction, CELP, encoding module.

36. The encoder system of embodiment 32, the decision section being adapted to:

detect a co-presence of a signal component with dominant fundamental frequency and a signal component located below the fundamental frequency and, optionally, between its harmonics; and

responsive thereto, to take a decision to disable.

37. The encoder system of embodiment 35,

the CELP encoding module being adapted to estimate a pitch frequency in the audio time signal; and

the decision section being adapted to detect spectral components located below the estimated pitch frequency and, responsive thereto, to take a decision to disable.

38. The encoder system of embodiment 35, the decision section being adapted

to compute a difference between a predicted power of the audio time signal when CELP-coded and a predicted power of the audio time signal when CELP-coded and post-filtered, and,

responsive to this difference exceeding a predetermined threshold, to take a decision to disable.

39. The encoder system of embodiment 35,

said encoding section further including a transform-coded excitation, TCX, encoding module,

wherein the decision section is adapted to select one of the following coding modes:

a) TCX coding;

b) CELP coding with post filtering; and

c) CELP coding without post filtering.

40. The encoder system of embodiment 39, further comprising a coding selector (814) adapted to select one of the following super-modes:

i) Advanced Audio Coding, AAC coding, wherein the decision section is disabled; and

ii) TCX/CELP coding, wherein the decision section is enabled to select one of coding modes a), b) and c).

41. The encoder system of embodiment 39, the decision section being adapted to decide which mode to use on the basis of a rate–distortion optimization.

42. The encoder system of embodiment 32,

further adapted to segment the bit stream signal into time frames,

the decision section being adapted to decide to disable the post filter in time segments consisting of entire frames.

43. The encoder system of embodiment 32, the decision section being adapted to decide to gradually decrease and/or increase the attenuation of the post filter.

44. The encoder system of embodiment 32, the decision section being adapted to:

compute the power of the audio time signal below an estimated pitch frequency; and

responsive to this power exceeding a predetermined threshold, to take a decision to disable.

45. The encoder system of embodiment 32, where the decision section is adapted to:

derive, from the audio time signal, an approximate difference signal approximating the signal component which is to be removed from a future decoded signal by the post filter;

assess at least one of the following criteria:

a) whether the power of the approximate difference signal exceeds a predetermined threshold;

b) whether the character of the approximate difference signal is tonal;

c) whether a difference between magnitude frequency spectra of the approximate difference signal and of the audio time signal is unevenly distributed with respect to frequency;

d) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a predetermined relevance envelope; and

e) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a relevance envelope obtained by thresholding a magnitude frequency spectrum of the audio time signal by a magnitude of the largest signal component therein downscaled by a predetermined scale factor;

and, responsive to a positive determination, to take a decision to disable the post filter.

46. The encoder system of embodiment 45, wherein the decision section is adapted to compute the approximate difference signal as a difference between the audio time signal and the audio time signal when subjected to post filtering.

47. The encoder system of embodiment 45, wherein:

the encoding section is adapted to extract an intermediate decoded signal representing excitation and to provide this to the decision section; and

the decision section is adapted to compute the approximate difference signal as a difference between the audio time signal and the intermediate decoded signal when subjected to post filtering.

48. A method of encoding an audio time signal as a bit stream signal, the method including the step of encoding an audio time signal as a bit stream signal,

characterized by the further step of deciding whether post filtering, which includes attenuation of interharmonic noise, is to be disabled at decod-

ing of the bit stream and encoding this decision in the bit stream signal as post filtering information.

49. The method of embodiment 48, wherein the step of deciding relates to post filtering which further includes attenuation of noise located in spectral valleys.

50. The method of embodiment 48, wherein the step of encoding includes applying a coding method adapted for speech coding.

51. The method of embodiment 48, wherein the step of encoding includes applying code-excited linear prediction, CELP, coding.

52. The method of embodiment 48,

further comprising the step of detecting a co-presence of a signal component with dominant fundamental frequency and a signal component located below the fundamental frequency and, optionally, between its harmonics,

wherein a decision to disable post filtering is made in the case of a positive detection outcome.

53. The method of embodiment 51, wherein:

said step of CELP coding includes estimating a pitch frequency in the audio time signal; and

the step of deciding includes detecting spectral components located below the estimated pitch frequency and a decision to disable post filtering is made in the case of a positive detection outcome.

54. The method of embodiment 51,

further including the step of computing a difference between a predicted power of the audio time signal when CELP-coded and a predicted power of the audio time signal when CELP-coded and post-filtered,

wherein a decision to disable post filtering is made if this difference exceeds a predetermined threshold.

55. The method of embodiment 51, wherein:

the step of encoding includes selectively applying either CELP coding or transform-coded excitation, TCX, coding; and

the step of deciding whether post filtering is to be disabled is performed only when CELP coding is applied.

56. The method of embodiment 55, wherein the step of deciding includes selecting, on the basis of a rate–distortion optimization, one of the following operating modes:

a) TCX coding;

b) CELP coding with post filtering; and

c) CELP coding without post filtering.

57. The method of embodiment 55, wherein the step of deciding includes selecting, on the basis of a rate–distortion optimization, one of the following operating modes:

a) TCX coding;

b) CELP coding with post filtering;

c) CELP coding without post filtering; and

d) Advanced Audio Coding, AAC coding.

58. The method of embodiment 48, wherein:

the step of encoding includes segmenting the audio time signal into time frames and to form a bit stream signal having corresponding time frames; and

the step of deciding that post filtering is to be disabled is carried out once in every time frame.

59. The method of embodiment 48, wherein the outcome of the step of deciding that post filtering is to be disabled is chosen from:

no attenuation,

full attenuation,

partial attenuation,
gradually increasing attenuation, and
gradually decreasing attenuation.

60. The method of embodiment 48, wherein the step of deciding includes computing the power of the audio time signal below and estimated pitch frequency and, responsive to this power exceeding a predetermined threshold, to disable the post filter.

61. The method of embodiment 48, wherein:

the step of encoding includes deriving, from the audio time signal, an approximate difference signal approximating the signal component which is to be removed from a future decoded signal by the post filter; and

the step of deciding includes assessing at least one of the following criteria:

a) whether the power of the approximate difference signal exceeds a predetermined threshold;

b) whether the character of the approximate difference signal is tonal;

c) whether a difference between magnitude frequency spectra of the approximate difference signal and of the audio time signal is unevenly distributed with respect to frequency;

d) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a predetermined relevance envelope; and

e) whether a magnitude frequency spectrum of the approximate difference signal is localized to frequency intervals within a relevance envelope obtained by thresholding a magnitude frequency spectrum of the audio time signal by a magnitude of the largest signal component therein downscaled by a predetermined scale factor;

and, responsive to at least a positive determination, to disable the post filter.

62. The method of embodiment 61, wherein the approximate difference signal is computed as a difference between the audio time signal and the audio time signal when subjected to post filtering.

63. The method of embodiment 61, wherein:

the step of encoding includes extracting an intermediate decoded signal representing excitation; and

the step of deciding includes computing the approximate difference signal as a difference between the audio time signal and the intermediate decoded signal when subjected to post filtering.

64. A computer-program product including a data carrier storing instructions for performing the method of any one of embodiment 20 to 31 and 48 to 63.

CLAIMS

1. A method of decoding a bit stream signal as an audio time signal, including the steps of:

decoding a bit stream signal as a preliminary audio time signal according to a coding mode selected from a plurality of coding modes, wherein the plurality of coding modes includes at least a first coding mode which includes a post-filtering step and at least a second coding mode which does not include the post-filtering step,

wherein the post-filtering step applies a pitch-enhancement filter to the preliminary audio time signal, thereby obtaining an audio time signal, wherein the post-filtering is applied only to low frequencies, and

wherein the post-filtering step is selectively omitted responsive to post-filtering information encoded in the bit stream signal, the post-filtering information being indicative of an encoder-side decision of whether or not to omit the post-filtering step, whereby the post-filtering step is selectively omitted in the first coding mode.

2. The method of claim 1, wherein the low frequencies are below a predetermined cutoff frequency.

3. The method of claim 1, wherein the decoding step includes applying code-excited linear prediction, CELP, decoding.

4. The method of claim 1, wherein the bit stream signal is segmented into time frames and the post-filtering step is omitted for an entire time frame or a sequence of entire time frames.

5. A computer-program product including a data carrier storing instructions which, when executed by a digital signal processor, cause the digital signal processor to perform the method of any one of claims 1-4.

6. A decoding system configured to perform the method of any one of claims 1-4.