

Harmonicity-Dependent Controlling of a Harmonic Filter Tool

Description

The present application is concerned with the decision on controlling of a harmonic filter tool such as of the *pre/post filter* or *post-filter only* approach. Such tool is, for example, applicable to MPEG-D unified speech and audio coding (USAC) and the upcoming 3GPP EVS codec.

Transform-based audio codecs like AAC, MP3, or TCX generally introduce inter-harmonic quantization noise when processing harmonic audio signals, particularly at low bitrates.

This effect is further worsened when the transform-based audio codec operates at low delay, due to the worse frequency resolution and/or selectivity introduced by a shorter transform size and/or a worse window frequency response.

This inter-harmonic noise is generally perceived as a very annoying “warbling” artifact, which significantly reduces the performance of the transform-based audio codec when subjectively evaluated on highly tonal audio material like some music or voiced speech.

A common solution to this problem is to employ prediction-based techniques, preferably prediction using autoregressive (AR) modeling based on the addition or subtraction of past input or decoded samples, either in the transform-domain or in the time-domain.

However, using such techniques in signals with changing temporal structure again leads to unwanted effects such as temporal smearing of percussive musical events or speech plosives or even the creation of impulse trails due to the repetition of a single impulse-like transient. Thus, special care has to be taken for signals that contain both transient and harmonic components or for signals where there is ambiguity between transients and trains of pulses (the latter belonging to a harmonic signal composed of individual pulses of very short duration; such signals are also known as pulse-trains).

Several solutions exist to improve the subjective quality of transform-based audio codecs on harmonics audio signals. All of them exploit the long-term periodicity (pitch) of very harmonic, stationary waveforms, and are based on prediction-based techniques, either in the transform-domain or in the time-domain. Most of the solutions are known as either long-term prediction (LTP) or pitch prediction, characterized by a pair of filters being applied to the signal: a pre-filter in the encoder (usually as a first step in the time or frequency domain) and a post-filter

in the decoder (usually as a last step in the time or frequency domain). A few other solutions, however, apply only a single post-filtering process on the decoder side generally known as harmonic post-filter or bass-post-filter. All of these approaches, regardless of being pre- and post-filter pairs or only post-filters, will be denoted as a harmonic filter tool in the following.

Examples of transform-domain approaches are:

- [1] H. Fuchs, "Improving MPEG Audio Coding by Backward Adaptive Linear Stereo Prediction", 99th AES Convention, New York, 1995, Preprint 4086.
- [2] L. Yin, M. Suonio, M. Väänänen, "A New Backward Predictor for MPEG Audio Coding", 103rd AES Convention, New York, 1997, Preprint 4521.
- [3] Juha Ojanperä, Mauri Väänänen, Lin Yin, "Long Term Predictor for Transform Domain Perceptual Audio Coding", 107th AES Convention, New York, 1999, Preprint 5036.

Examples of time-domain approaches applying both pre- and post-filtering are:

- [4] Philip J. Wilson, Harprit Chhatwal, "Adaptive transform coder having long term predictor", U.S. Patent 5,012,517, April 30, 1991.
- [5] Jeongook Song, Chang-Heon Lee, Hyen-O Oh, Hong-Goo Kang, "Harmonic Enhancement in Low Bitrate Audio Coding Using an Efficient Long-Term Predictor", EURASIP Journal on Advances in Signal Processing, August 2010.
- [6] Juin-Hwey Chen, "Pitch-based pre-filtering and post-filtering for compression of audio signals", U.S. Patent 8,738,385, May 27, 2014.
- [7] Jean-Marc Valin, Koen Vos, Timothy B. Terriberry, "Definition of the Opus Audio Codec", ISSN: 2070-1721, IETF RFC 6716, September 2012.
- [8] Rakesh Taori, Robert J. Sluijter, Eric Kathmann "Transmission System with Speech Encoder with Improved Pitch Detection", U.S. Patent 5,963,895, October 5, 1999.

Examples of time-domain approaches where only post-filtering is applied are:

- [9] Juin-Hwey Chen, Allen Gersho, "Adaptive Postfiltering for Quality Enhancement of Coded Speech", IEEE Trans. on Speech and Audio Proc., vol. 3, January 1995.
- [10] Int. Telecommunication Union, "Frame error robust variable bit-rate coding of speech and audio from 8–32 kbit/s", Recommendation ITU-T G.718, June 2008. www.itu.int/rec/T-REC-G.718/e, section 7.4.1.
- [11] Int. Telecommunication Union, "Coding of speech at 8 kbit/s using conjugate structure algebraic CELP (CS-ACELP)", Recommendation ITU-T G.729, June 2012. www.itu.int/rec/T-REC-G.729/e, section 4.2.1.

- [12] Bruno Bessette et al., "Method and device for frequency-selective pitch enhancement of synthesized speech", U.S. Patent 7,529,660, May 30, 2003.

An example of a transient detector is:

- [13] Johannes Hilpert et al., "Method and Device for Detecting a Transient in a Discrete-Time Audio Signal", U.S. Patent 6,826,525, November 30, 2004.

Relevant literature on psychoacoustics:

- [14] Hugo Fastl, Eberhard Zwicker, "Psychoacoustics: Facts and Models", 3rd Edition, Springer, December 14, 2006.
- [15] Christoph Markus, "Background Noise Estimation", European Patent EP 2,226,794, March 6, 2009.

All the techniques described in the prior have decisions when to enable the prediction filter based on a single threshold decision (e.g. prediction gain [5] or pitch gain [4] or harmonicity which is basically proportional to the normalized correlation [6]). Furthermore, OPUS [7] employs hysteresis that increases the threshold if the pitch is changing and decreases the threshold if the gain in the previous frame was above a predefined fixed threshold. OPUS [7] also disables the long-term (pitch) predictor if a transient is detected in some specific frame configurations. The reason for this design seems to stem from the general belief that, in a mix of harmonic and transient signal components, the transient dominates the mix, and activating LTP or pitch prediction upon it would, as discussed earlier, subjectively cause more harm than improvement. However, for some mixtures of waveforms which will be discussed hereafter, activating the long-term or pitch predictor on transient audio frames significantly increases the coding quality or efficiency and thus is beneficial. Furthermore, it may be beneficial to, when activating the predictor, vary its strength based on instantaneous signal characteristics other than a prediction gain, the only approach in the state of the art.

Accordingly, it is an object of the present invention to provide a concept for a harmonicity-dependent controlling of a harmonic filter tool of an audio codec which results in an improved coding efficiency, e.g. improved objective coding gain or better perceptual quality or the like.

This object is achieved at the subject matter of the independent claims of the present application.

It is a basic finding of the present application that the coding efficiency of an audio codec using a controllable – switchable or even adjustable – harmonic filter tool may be improved

by performing the harmonicity-dependent controlling of this tool using a temporal structure measure in addition to a measure of harmonicity in order to control the harmonic filter tool. In particular, the temporal structure of the audio signal is evaluated in a manner which depends on the pitch. This enables to achieve a situation-adapted control of the harmonic filter tool such that in situations where a control made solely based on the measure of harmonicity would decide against or reduce the usage of this tool although using the harmonic filter tool would, in that situation, increase the coding efficiency, the harmonic filter tool is applied, while in other situations where the harmonic filter tool may be inefficient or even destructive, the control reduces the appliance of the harmonic filter tool appropriately.

Advantageous implementations of the present invention on the subject of the dependent claims and preferred embodiments of the present application are set out below with respect to the figures among which

- Fig. 1 shows a block diagram of an apparatus for controlling a harmonic filter tool in terms of filter gain in accordance with an embodiment;
- Fig. 2 shows an example for a possible predetermined condition to be met for applying the harmonic filter tool;
- Fig. 3 shows a flow diagram illustrating a possible implementation of a decision logic which, inter alias, could be parameterized so as to realize the condition example of Fig. 2;
- Fig. 4 shows a block diagram of an apparatus for performing a harmonicity (and temporal-measure) dependent controlling of a harmonic filter tool;
- Fig. 5 shows a schematic diagram illustrating the temporal position of a temporal region for determining the temporal structure measure in accordance with an embodiment;
- Fig. 6 shows schematically a graph of energy samples temporally sampling the energy of the audio signal within the temporal region in accordance with an embodiment;
- Fig. 7 shows a block diagram illustrating the usage of the apparatus of Fig. 4 in an audio codec by illustrating the encoder and the decoder of the audio codec,

respectively, when the encoder uses the apparatus of Fig. 4, in accordance with an embodiment wherein a harmonic pre-/post-filter tool is used;

- Fig. 8 shows a block diagram illustrating the usage of the apparatus of Fig. 4 in an audio codec by illustrating the encoder and the decoder of the audio codec, respectively, when the encoder uses the apparatus of Fig. 4, in accordance with an embodiment wherein a harmonic post-filter tool is used;
- Fig. 9 shows a block diagram of the controller of Fig. 4 in accordance with an embodiment;
- Fig. 10 shows a block diagram of a system illustrating the possibility that the apparatus of Fig. 4 shares the use of the energy samples of Fig. 6 with a transient detector;
- Fig. 11 shows a graph of a time-domain portion (portion of the waveform) out of an audio signal as an example of a low pitched signal with additionally illustrating the pitch dependent positioning of the temporal region for determining the at least one temporal structure measure;
- Fig. 12 shows a graph of a time-domain portion out of an audio signal as an example of a high pitched signal with additionally illustrating the pitch dependent positioning of the temporal region for determining the at least one temporal structure measure;
- Fig. 13 shows an exemplary spectrogram of an impulse and step transient within a harmonic signal;
- Fig. 14 shows an exemplary spectrogram to illustrate an LTP influence on impulse and step transient;
- Fig. 15 shows, one upon the other, time-domain portions of the audio signal shown in Fig. 14, and its low pass filtered and high-pass filtered version thereof, respectively, in order to illustrate the control according to Fig. 2, 3, 16 and 17 for impulse and for step transient;

- Fig. 16 shows a bar chart of an example for temporal sequence of energies of segments - sequence of energy samples - for an impulse like transient and the placement of the temporal region for determining the at least one temporal structure measure in accordance with Fig. 2 and 3;
- Fig. 17 shows a bar chart of an example for temporal sequence of energies of segments - sequence of energy samples - for a step like transient and the placement of the temporal region for determining the at least one temporal structure measure in accordance with Fig. 2 and 3;
- Fig. 18 shows an exemplary spectrogram of a train of pulses (excerpt using short FFT spectrogram);
- Fig. 19 shows an exemplary waveform of the train of pulses;
- Fig. 20 shows an original Short FFT spectrogram of the train of pulses; and
- Fig. 21 shows an original Long FFT spectrogram of the train of pulses.

The following description starts with a first detailed embodiment of a harmonic filter tool control. A brief survey of thoughts, which led to this first embodiment, are presented. These thoughts, however, also apply to the subsequently explained embodiments. Thereinafter, generalizing embodiments are presented, followed by specific concrete examples for audio signal portions in order to more concretely outline the effects resulting from embodiments of the present application.

The decision mechanism for enabling or controlling a harmonic filter tool of, for example, a prediction based technique, is, based on a combination of a harmonicity measure such as a normalized correlation or prediction gain and a temporal structure measure, e.g. temporal flatness measure or energy change.

The decision may, as outlined below, not be dependent just on the harmonicity measure from the current frame, but also on a harmonicity measure from the previous frame and on a temporal structure measure from the current and, optionally, from the previous frame.

The decision scheme may be designed such that the prediction based technique is enabled also for transients, whenever using it would be psychoacoustically beneficial as concluded by a respective model.

Thresholds used for enabling the prediction based technique may be, in one embodiment, dependent on the current pitch instead on the pitch change.

The decision scheme allows, for example, to avoid repetition of a specific transient, but allow prediction based technique for some transients and for signals with specific temporal structures where a transient detector would normally signal short transform blocks (i.e. the existence of one or more transients).

The decision technique presented below may be applied to any of the prediction-based methods described above, either in the transform-domain or in the time-domain, either pre-filter plus post-filter or post-filter only approaches. Moreover, it can be applied to predictors operating band-limited (with lowpass) or in subbands (with bandpass characteristics).

The overall objective regarding the activating of LTP, pitch prediction, or harmonic post-filtering is that both of the following conditions are achieved:

- An objective or subjective benefit is obtained by activating the filter,
- No significant artifacts are introduced by the activation of said filter.

Determining whether there is an objective benefit to using the filter usually performed by means of autocorrelation and/or prediction gain measures on the target signal and is well known [1–7].

The measurement of a subjective benefit is also straightforward at least for stationary signals, since perceptual improvement data obtained through listening tests are typically proportional to the corresponding objective measures, i.e. the abovementioned correlation and/or prediction gain.

Identifying or predicting the existence of artifacts caused by the filtering, though, requires more sophisticated techniques than simple comparisons of objective measures like frame type (long transforms for stationary vs. short transforms for transient frames) or prediction gain to certain thresholds, as is done in the state of the art. Essentially, in order to prevent artifacts one has to ensure that the changes the filtering causes in the target waveform do not significantly exceed a time-varying spectro-temporal masking threshold anywhere in time

or frequency. The decision scheme in accordance with some of the embodiments presented below, thus, uses the following filter decision and control scheme consisting of three algorithmic blocks to be executed in series for each frame of the audio signal to be coded and/or subjected to the filtering:

A harmonicity measurement block which calculates commonly used harmonic filter data such as normalized correlation or gain values (referred to as "prediction gain" hereafter). As noted again later, the word "gain" is meant as a generalization for any parameter commonly associated with a filter's strength, e.g. an explicit gain factor or the absolute or relative magnitude of a set of one or more filter coefficients.

A T/F envelope measurement block which computes time-frequency (T/F) amplitude or energy or flatness data with a predefined spectral and temporal resolution (this may also include measures of frame transientness used for frame type decisions, as noted above). The pitch obtained in the harmonicity measurement block is input to the T/F envelope measurement block since the region of the audio signal used for filtering of the current frame, typically using past signal samples, depends on the pitch (and, correspondingly, so does the computed T/F envelope).

A filter gain computation block performing the final decision about which filter gain to use (and thus to transmit in the bit-stream) for the filtering. Ideally, this block should compute, for each transmittable filter gain less than or equal to the prediction gain, a spectro-temporal excitation-pattern-like envelope of the target signal after filtering with said filter gain, and should compare this "actual" envelope with an excitation-pattern envelope of the original signal. Then, one may use for coding/transmission the largest filter gain whose corresponding spectro-temporal "actual" envelope does not differ from the "original" envelope by more than a certain amount. This filter gain we shall call psychoacoustically optimal.

In other embodiments described later, the three-block structure is a little bit modified.

In other words, harmonicity and T/F envelope measures are obtained in corresponding blocks, which are subsequently used to derive psychoacoustic excitation patterns of both the input and filtered output frames, and finally the filter gain is adapted such that a masking threshold, given by a ratio between the "actual" and the "original" envelope, is not significantly exceeded. To appreciate this, it should be noted that an excitation pattern in this context is very similar to a spectrogram-like representation of the signal being examined, but

exhibits temporal smoothing modeled after certain characteristics of human hearing and manifesting itself as “post-masking”.

Fig. 1 illustrates the connection between the three blocks introduced above. Unfortunately, a frame-wise derivation of two excitation patterns and a brute-force search for the best filter gain often is computationally complex. Therefore simplifications are presented in the following description.

In order to avoid expensive computations of excitation patterns in the proposed filter-activation decision scheme, low-complexity envelope measures are used as estimates of the characteristics of the excitation patterns. It was found that in the T/F envelope measurement block, data such as segmental energies (SE), temporal flatness measure (TFM), maximum energy change (MEC) or traditional frame configuration info such as the frame type (long/stationary or short/transient) suffice to derive estimates of psychoacoustic criteria. These estimates then can be utilized in the filter gain computation block to determine, with high accuracy, an optimal filter gain to be employed for coding or transmission. In order to prevent a computationally intensive search for the globally optimal gain, a rate-distortion loop over all possible filter gains (or a sub-set thereof) can be substituted by one-time conditional operators. Such “cheap” operators serve to decide whether some filter gain, computed using data from the harmonicity and T/F envelope measurement blocks, shall be set to zero (decision not to use harmonic filtering) or not (decision to use harmonic filtering). Note that the harmonicity measurement block can remain unchanged. A step-by-step realization of this low-complexity embodiment is described hereafter.

As noted, the “initial” filter gain subjected to the one-time conditional operators is derived using data from the harmonicity and T/F envelope measurement blocks. More specifically, the “initial” filter gain may be equal to the product of the time-varying prediction gain (from the harmonicity measurement block) and a time-varying scale factor (from the psychoacoustic envelope data of the T/F envelope measurement block). In order to further reduce the computational load a fixed, constant scale factor such as 0.625 may be used instead of the signal-adaptive time-variant one. This typically retains sufficient quality and is also taken into account in the following realization.

A step-by-step description of a concrete embodiment for controlling of the filter tool is laid out now.

1. Transient detection and temporal measures

The input signal $s_{HP}(n)$ is input to the time-domain transient detector. The input signal $s_{HP}(n)$ is high-pass filtered. The transfer function of the transient detection's HP filter is given by

$$H_{TD}(z) = 0.375 - 0.5z^{-1} + 0.125z^{-2} \quad (1)$$

The signal, filtered by the transient detection's HP filter, is denoted as $s_{TD}(n)$. The HP-filtered signal $s_{TD}(n)$ is segmented into 8 consecutive segments of the same length. The energy of the HP-filtered signal $s_{TD}(n)$ for each segment is calculated as:

$$E_{TD}(i) = \sum_{n=0}^{L_{segment}-1} (s_{TD}(iL_{segment} + n))^2, \quad i = 0, \dots, 7 \quad (2)$$

where $L_{segment} = \frac{L}{8}$ is the number of samples in 2.5 milliseconds segment at the input sampling frequency.

An accumulated energy is calculated using:

$$E_{Acc} = \max(E_{TD}(i-1), 0.8125E_{Acc}) \quad (3)$$

An attack is detected if the energy of a segment $E_{TD}(i)$ exceeds the accumulated energy by a constant factor $attackRatio = 8.5$ and the attackIndex is set to i :

$$E_{TD}(i) > attackRatio \cdot E_{Acc} \quad (4)$$

If no attack is detected based on the criteria above, but a strong energy increase is detected in segment i , the attackIndex is set to i without indicating the presence of an attack. The attackIndex is basically set to the position of the last attack in a frame with some additional restrictions.

The energy change for each segment is calculated as:

$$E_{chg}(i) = \begin{cases} \frac{E_{TD}(i)}{E_{TD}(i-1)}, & E_{TD}(i) > E_{TD}(i-1) \\ \frac{E_{TD}(i-1)}{E_{TD}(i)}, & E_{TD}(i-1) > E_{TD}(i) \end{cases} \quad (5)$$

The temporal flatness measure is calculated as:

$$TFM(N_{past}) = \frac{1}{8 + N_{past}} \sum_{i=-N_{past}}^7 E_{chg}(i) \quad (6)$$

The maximum energy change is calculated as:

$$MEC(N_{past}, N_{new}) = \max(E_{chg}(-N_{past}), E_{chg}(-N_{past} + 1), \dots, E_{chg}(N_{new} - 1)) \quad (7)$$

If index of $E_{chg}(i)$ or $E_{TD}(i)$ is negative then it indicates a value from the previous segment, with segment indexing relative to the current frame.

N_{past} is the number of the segments from the past frames. It is equal to 0 if the temporal flatness measure is calculated for the usage in ACELP/TCX decision. If the temporal flatness measure is calculate for the TCX LTP decision then it is equal to:

$$N_{past} = 1 + \min\left(8, \left\lceil 8 \frac{pitch}{L} + 0.5 \right\rceil\right) \quad (8)$$

N_{new} is the number of segments from the current frame. It is equal to 8 for non-transient frames. For transient frames first the locations of the segments with the maximum and the minimum energy are found:

$$i_{\max} = \arg \max_{i \in \{-N_{past}, \dots, 7\}} E_{TD}(i) \quad (9)$$

$$i_{\min} = \arg \min_{i \in \{-N_{past}, \dots, 7\}} E_{TD}(i) \quad (10)$$

If $E_{TD}(i_{\min}) > 0.375 E_{TD}(i_{\max})$ then N_{new} is set to $i_{\max} - 3$, otherwise N_{new} is set to 8.

2. Transform block length switching

The overlap length and the transform block length of the TCX are dependent on the existence of a transient and its location.

Table 1: Coding of the overlap and the transform length based on the transient position

attack Index	Overlap with the first window of the following frame	Short/Long Transform decision (binary coded) 0 – Long, 1 - Short	Binary code for the overlap width	Overlap code
none	ALDO	0	0	00
-2	FULL	1	0	10
-1	FULL	1	0	10
0	FULL	1	0	10
1	FULL	1	0	10
2	MINIMAL	1	10	110
3	HALF	1	11	111
4	HALF	1	11	111
5	MINIMAL	1	10	110
6	MINIMAL	0	10	010
7	HALF	0	11	011

The transient detector described above basically returns the index of the last attack with the restriction that if there are multiple transients then MINIMAL overlap is preferred over HALF overlap which is preferred over FULL overlap. If an attack at position 2 or 6 is not strong enough then HALF overlap is chosen instead of the MINIMAL overlap.

3. Pitch estimation

One pitch lag (integer part + fractional part) per frame is estimated (frame size e.g. 20ms). This is done in 3 steps to reduce complexity and improves estimation accuracy.

a. First Estimation of the integer part of the pitch lag

A pitch analysis algorithm that produces a smooth pitch evolution contour is used (e.g. Open-loop pitch analysis described in Rec. ITU-T G.718, sec. 6.6). This analysis is generally done on a subframe basis (subframe size e.g. 10ms), and produces one pitch lag estimate per subframe. Note that these pitch lag estimates do not have any fractional part and are generally estimated on a downsampled signal (sampling rate e.g. 6400Hz). The signal used can be any audio signal, e.g. a LPC weighted audio signal as described in Rec. ITU-T G.718, sec. 6.5.

b. Refinement of the integer part of the pitch lag

The final integer part of the pitch lag is estimated on an audio signal $x[n]$ running at the core encoder sampling rate, which is generally higher than the sampling rate of the downsampled signal used in a. (e.g. 12.8kHz, 16kHz, 32kHz...). The signal $x[n]$ can be any audio signal e.g. a LPC weighted audio signal.

The integer part of the pitch lag is then the lag T_{int} that maximizes the autocorrelation function

$$C(d) = \sum_{n=0}^L x[n]x[n-d]$$

with d around a pitch lag T estimated in step 1.a.

$$T - \delta_1 \leq d \leq T + \delta_2$$

c. Estimation of the fractional part of the pitch lag

The fractional part is found by interpolating the autocorrelation function $C(d)$ computed in step 2.b. and selecting the fractional pitch lag T_{fr} which maximizes the interpolated autocorrelation function. The interpolation can be performed using a low-pass FIR filter as described in e.g. Rec. ITU-T G.718, sec. 6.6.7.

4. Decision bit

If the input audio signal does not contain any harmonic content or if a prediction based technique would introduce distortions in time structure (e.g. repetition of a short transient), then no parameters are encoded in the bitstream. Only 1 bit is sent such that the decoder knows whether he has to decode the filter parameters or not. The decision is made based on several parameters:

Normalized correlation at the integer pitch-lag estimated in step 3.b.

$$norm_corr = \frac{\sum_{n=0}^L x[n]x[n-T_{int}]}{\sqrt{\sum_{n=0}^L x[n]x[n]} \sqrt{\sum_{n=0}^L x[n-T_{int}]x[n-T_{int}]}}$$

The normalized correlation is 1 if the input signal is perfectly predictable by the integer pitch-lag, and 0 if it is not predictable at all. A high value (close to 1) would then indicate a harmonic signal. For a more robust decision, beside the normalized correlation for the current frame ($norm_corr(curr)$) the normalized correlation of the past frame ($norm_corr(prev)$) can also be used in the decision., e.g.:

$$\text{If } (norm_corr(curr) * norm_corr(prev)) > 0.25$$

or

If $\max(\text{norm_corr}(\text{curr}), \text{norm_corr}(\text{prev})) > 0.5$,

then the current frame contains some harmonic content (bit=1)

- a. Features computed by a transient detector (e.g. Temporal flatness measure (6), Maximal energy change (7)), to avoid activating the postfilter on a signal containing a strong transient or big temporal changes. The temporal features are calculated on the signal containing the current frame (N_{new} segments) and the past frame up to the pitch lag (N_{past} segments). For step like transients that are slowly decaying, all or some of the features are calculated only up to the location of the transient ($i_{\text{max}} - 3$) because the distortions in the non-harmonic part of the spectrum introduced by the LTP filtering would be suppressed by the masking of the strong long lasting transient (e.g. crash cymbal).
- b. Pulse trains for low pitched signals can be detected as a transient by a transient detector. For the signals with low pitch the features from the transient detector are thus ignored and there is instead additional threshold for the normalized correlation that depends on the pitch lag, e.g.:

If $\text{norm_corr} \leq 1.2 \cdot T_{\text{int}}/L$, then set the bit=0 and do not send any parameters.

One example decision is shown in Fig. 2 where b1 is some bitrate, for example 48 kbps, where TCX_20 indicates that the frame is coded using single long block, where TCX_10 indicates that the frame is coded using 2,3,4 or more short blocks, where TCX_20/TCX_10 decision is based on the output of the transient detector described above. tempFlatness is the Temporal Flatness Measure as defined in (6), maxEnergyChange is the Maximum Energy Change as defined in (7). The condition $\text{norm_corr}(\text{curr}) > 1.2 \cdot T_{\text{int}}/L$ could also be written as $(1.2 - \text{norm_corr}(\text{curr})) \cdot L < T_{\text{int}}$.

The principle of the decision logic is depicted in the block diagram in Fig. 3. It should be noted that Fig. 3 is more general than Fig. 2 in sense that the thresholds are not restricted. They may be set according to Fig. 2 or differently. Moreover, Fig. 3 illustrates that the exemplary bitrate dependency of Fig. 2 may be left-off. Naturally, the decision logic of Fig. 3 could be varied to include the bitrate dependency of Fig. 2. Further, Fig. 3 has been held unspecific with regard to the usage of only the current or also the past pitch. Insofar, Fig. 3 shows that the embodiment of Fig. 2 may be varied in this regard.

The "threshold" in Fig. 3 corresponds to different thresholds used for tempFlatness and maxEnergyChange in Fig. 2. The "threshold_1" in Fig. 3 corresponds to $1.2 \cdot T_{\text{int}}/L$ in Fig. 2. The

"threshold_2" in Fig. 3 corresponds to 0.44 or $\max(\text{norm_corr}(\text{curr}), \text{norm_corr}(\text{prev})) > 0.5$ or $(\text{norm_corr}(\text{curr})$

* $\text{norm_corr_prev}) > 0.25$ in Fig. 2

It is obvious from the examples above that the detection of a transient affects which decision mechanism for the long term prediction will be used and what part of the signal will be used for the measurements used in the decision, and not that it directly triggers disabling of the long term prediction.

The temporal measures used for the transform length decision may be completely different from the temporal measures used for the LTP decision or they may overlap or be exactly the same but calculated in different regions.

For low pitched signals the detection of transients is completely ignored if the threshold for the normalized correlation that depends on the pitch lag is reached.

5. Gain estimation and quantization

The gain is generally estimated on the input audio signal at the core encoder sampling rate, but it can also be any audio signal like the LPC weighted audio signal. This signal is noted $y[n]$ and can be the same or different than $x[n]$.

The prediction $y_P[n]$ of $y[n]$ is first found by filtering $y[n]$ with the following filter

$$P(z) = B(z, T_{fr})z^{-T_{int}}$$

with T_{int} the integer part of the pitch lag (estimated in 0) and $B(z, T_{fr})$ a low-pass FIR filter whose coefficients depend on the fractional part of the pitch lag T_{fr} (estimated in 0).

One example of $B(z)$ when the pitch lag resolution is $\frac{1}{4}$:

$$\begin{aligned} T_{fr} = \frac{0}{4} & \quad B(z) = 0.0000z^{-2} + 0.2325z^{-1} + 0.5349z^0 + 0.2325z^1 \\ T_{fr} = \frac{1}{4} & \quad B(z) = 0.0152z^{-2} + 0.3400z^{-1} + 0.5094z^0 + 0.1353z^1 \\ T_{fr} = \frac{2}{4} & \quad B(z) = 0.0609z^{-2} + 0.4391z^{-1} + 0.4391z^0 + 0.0609z^1 \\ T_{fr} = \frac{3}{4} & \quad B(z) = 0.1353z^{-2} + 0.5094z^{-1} + 0.3400z^0 + 0.0152z^1 \end{aligned}$$

The gain g is then computed as follows:

$$g = \frac{\sum_{n=0}^{L-1} y[n]y_P[n]}{\sum_{n=0}^{L-1} y_P[n]y_P[n]}$$

and limited between 0 and 1.

Finally, the gain is quantized e.g. on 2 bits, using e.g. uniform quantization.

If the gain is quantized to 0, then no parameters are encoded in the bitstream, only the 1 decision bit (bit=0).

The description brought forward so far motivated and outlined the advantages of embodiments of the present application for a harmonicity-dependent control of a harmonic filter tool, also for the ones outlined below which represent generalized embodiments to the step-by-step embodiment above. Sometimes the description brought forward so far was very specific although the harmonicity-dependent control concept may also advantageously be used in the framework of other audio codecs and may be varied relative to the specific details outlined in the foregoing. For this reason, embodiments of the present application are described again in the following in a more generic manner. Nevertheless, from time to time the following description refers back to the detailed description brought forward above in order to use the above details in order to reveal as to how the generically described elements occurring below may be implemented in accordance with further embodiments. In doing so, it should be noted that all of these specific implementation details may be individually transferred from the above description towards the elements described below. Accordingly, whenever in the description outlined below reference is made to the description brought forward above, this reference is meant to be independent from further references to the above description.

Thus, a more generic embodiment which emerges from the above detailed description is depicted in Fig. 4. In particular, Fig. 4 shows an apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool, such as a harmonic pre/post filter or harmonic post-filter tool, of an audio codec. The apparatus is generally indicated using reference sign 10. Apparatus 10 receives the audio signal 12 to be processed by the audio codec and outputs a control signal 14 to fulfill the controlling task of apparatus 10. Apparatus 10 comprises a pitch estimator 16 configured to determine a current pitch lag 18 of the audio signal 12, and a harmonicity measurer 20 configured to determine a measure 22 of harmonicity of the audio signal 12 using a current pitch lag 18. In particular, the harmonicity measure may be a prediction gain or may be embodied by one (single-) or more (multi-tap) filter coefficients or a maximum normalized correlation. The harmonicity measure calculation block of Fig. 1 comprised the tasks of both pitch estimator 16 and harmonicity measurer 20.

The apparatus 10 further comprises a temporal structure analyzer 24 configured to determine at least one temporal structure measure 26 in a manner dependent on the pitch lag 18, measure 26 measuring a characteristic of a temporal structure of the audio signal 12. For example, the dependency may rely in the positioning of the temporal region within which measure 26 measures the characteristic of a temporal structure of the audio signal 12, as described above and later in more detail. For sake of completeness, however, it is briefly noted that the dependency of the determination of measure 26 on the pitch-lag 18 may also be embodied differently to the description above and below. For example, instead of positioning the temporal portion, i.e. the determination window, in a manner dependent on the pitch-lag, the dependency could merely temporally vary weights at which a respective time-interval of the audio signal within a window positioned independently from the pitch-lag relative to the current frame, contribute to the measure 26. Relating to the description below, this may mean that the determination window 36 could be steadily located to correspond to the concatenation of the current and previous frames, and that the pitch-dependently located portion merely functions as a window of increased weight at which the temporal structure of the audio signal influences the measure 26. However, for the time being, it is assumed that the temporal window is located positioned according to the pitch-lag. Temporal structure analyzer 24 corresponds to the T/F envelope measure calculation block of Fig. 1.

Finally, the apparatus of Fig. 4 comprises a controller 28 configured to output control signal 14 depending on the temporal structure measure 26 and the measure 22 of harmonicity so as to thereby control the harmonic pre/post filter or harmonic post-filter. When comparing Fig. 4 with Fig. 1, the optimal filter gain computation block corresponds to, or represents a possible implementation of, controller 28.

The mode of operation of apparatus 10 is as follows. In particular, the task of apparatus 10 is to control the harmonic filter tool of an audio codec, and although the above-outlined more detailed description with respect to Figs. 1 to 3 reveals a gradual control or adaptation of this tool in terms of its filter strength or filter gain, for example, controller 28 is not restricted to that type of gradual control. Generally speaking, the control by controller 28 may gradually adapt the filter strength or gain of the harmonicity filter tool between 0 and a maximum value, both inclusively, as it was the case in the above specific examples with respect to Figs. 1 to 3, but different possibilities are feasible as well, such as a gradual control between two non-zero filter gain values, a step-wise control or a binary control such as a switching between enablement (non-zero) or disablement (zero gain) to switch on or off the harmonic filter tool.

As became clear from the above discussion, the harmonic filter tool which is illustrated in Fig. 4 by dashed lines 30 aims at improving the subjective quality of an audio codec such as a transform-based audio codec, especially with respect to harmonic phases of the audio signal. In particular, such a tool 30 is especially useful in low bitrate scenarios where a quantization noise introduced would, without tool 30, lead in such harmonic phases to audible artifacts. It is important, however, that filter tool 30 does not negatively affect other temporal phases of the audio signal which are not predominately harmonic. Further, as outlined above, filter tool 30 may be of the post-filter approach or pre-filter plus post-filter approach. Pre and/or post-filters may operate in transform domain or time domain. For example, a post-filter of tool 30 may, for example, have a transfer function having local maxima arranged at spectral distances corresponding to, or being set dependent on, pitch lag 18. The implementation of pre-filter and/or post-filter in the form of an LTP filter, in the form of, for example, an FIR and IIR filter, respectively, is also feasible. The pre-filter may have a transfer function being substantially the inverse of the transfer function of the post-filter. In effect, the pre-filter seeks to hide the quantization noise within the harmonic component of the audio signal by increasing the quantization noise within the harmonic of the current pitch of the audio signal and the post-filter reshapes the transmitted spectrum accordingly. In case of the post-filter only approach, the post-filter really modifies the transmitted audio signal so as to filter quantization noise occurring the between the harmonics of the audio signal's pitch.

It should be noted that Fig. 4 is, in some sense, drawn in a simplifying manner. For example, although Fig. 4 suggests that pitch estimator 16, harmonicity measurer 20 and temporal structure analyzer 24 operate, i.e. perform their tasks, on the audio signal 12 directly, or at least at the same version thereof, this does not need to be the case. Actually, pitch-estimator 16, temporal structure analyzer 24 and harmonicity measurer 20 may operate on different versions of the audio signal 12 such as different ones of the original audio signal and some pre-modified version thereof, wherein these versions may vary among elements 16, 20 and 24 internally and also with respect to the audio codec as well, which may also operate on some modified version of the original audio signal. For example, the temporal structure analyzer 24 may operate on the audio signal 12 at the input sampling rate thereof, i.e. the original sampling rate of audio signal 12, or it may operate on an internally coded/decoded version thereof. The audio codec, in turn, may operate at some internal core sampling rate which is usually lower than the input sampling rate. The pitch-estimator 16, in turn, may perform its pitch estimation task on a pre-modified version of the audio signal, such as, for example, on a psychoacoustically weighted version of the audio signal 12 so as to improve the pitch estimation with respect to spectral components which are, in terms of perceptibility, more significant than other spectral components. For example, as described above, the

pitch-estimator 16 may be configured to determine the pitch lag 18 in stages comprising a first stage and a second stage, the first stage resulting in a preliminary estimation of the pitch lag which is then refined in the second stage. For example, as it has been described above, pitch estimator 16 may determine a preliminary estimation of the pitch lag at a down-sampled domain corresponding to a first sample rate, and then refining the preliminary estimation of the pitch lag at a second sample rate which is higher than the first sample rate.

As far as the harmonicity measurer 20 is concerned, it has become clear from the discussion above with respect to Figs. 1 to 3 that it may determine the measure 22 of harmonicity by computing a normalized correlation of the audio signal or a pre-modified version thereof at the pitch lag 18. It should be noted that harmonicity measurer 20 may even be configured to compute the normalized correlation even at several correlation time distances besides the pitch lag 18 such as in a temporal delay interval including and surrounding the pitch lag 18. This may be favorable, for example, in case of filter tool 30 using a multi-tap LTP or possible LTP with fractional pitch. In that case, harmonicity measurer 20 may analyze or evaluate the correlation even at lag indices neighboring the actual pitch lag 18, such as the integer pitch lag in the concrete example outlined above with respect to Figs. 1 to 3.

For further details and possible implementations of the pitch estimator 16, reference is made to the section "pitch estimation" brought forward above. Possible implementations of the harmonicity measurer 20 were discussed above with respect to the equation of norm.corr. However, as also described above, the term "harmonicity measure" shall include not only a normalized correlation but also hints at measuring the harmonicity such as a prediction gain of the harmonic filter, wherein that harmonic filter may be equal to or may be different to the pre-filter of filter 230 in case of using the pre/post-filter approach and irrespective of the audio codec using this harmonic filter or as to whether this harmonic filter is merely used by harmonicity measurer 20 so as to determine measure 22.

As was described above with respect to Figs. 1 to 3, the temporal structure analyzer 24 may be configured to determine the at least one temporal structure measure 26 within a temporal region temporally placed depending on the pitch lag 18. In order to illustrate this further, see Fig. 5. Fig. 5 illustrates a spectrogram 32 of the audio signal, i.e. its spectral decomposition up to some highest frequency f_H depending on, for example, the sample rate of the version of the audio signal internally used by the temporal structure analyzer 24, temporally sampled at some transform block rate which may or may not coincide with an audio codec's transform block rate, if any. For illustration purposes, Fig. 5 illustrates the spectrogram 32 as being temporally subdivided into frames in units of which the controller may, for example, perform

its controlling of filter tool 30, which frame subdivision may, for example, also coincide with the frame subdivision used by the audio codec comprising or using filter tool 30.

For the time being, it is illustratively assumed that the current frame for which the controlling task of controller 28 is performed, is frame 34a. As was described above and as is illustrated in Fig. 5, the temporal region 36, within which temporal structure analyzer determiner determines the at least one temporal structure measure 26, does not necessarily coincide with current frames 34a. Rather, both the temporally past-heading end 38 as well as the temporally future-heading end 40 of the temporal region 36 may deviate from the temporally past-heading and future heading ends 42 and 44 of the current frame 34a. As has been described above, the temporal structure analyzer 24 may position the temporally past-heading end 38 of the temporal region 36 depending on the pitch lag 18 determined by pitch estimator 16 which determines the pitch lag 18 for each frame 34, for current frame 34a. As became clear from the discussion above, the temporal structure analyzer 24 may position the temporal past-heading end 38 of the temporal region such that the temporally past-heading end 38 is displaced into a past direction relative to the current frame's 34a past-heading end 42, for example, by a temporal amount 46 which monotonically increases with an increase of the pitch lag 18. In other words, the greater the pitch lag 18 is, the greater amount 46 is. As became clear from the discussion above with respect to Figs. 1 to 3, the amount may be set according to equation 8, where N_{past} is a measure for the temporal displacement 46.

The temporally future-heading end 40 of temporal region 36, in turn, may be set by temporal structure analyzer 24 depending on the temporal structure of the audio signal within a temporal candidate region 48 extending from the temporally past-heading end 38 of the temporal region 36 to the temporally future-heading end of the current frame, 44. In particular, as has been discussed above, the temporal structure analyzer 24 may evaluate a disparity measure of energy samples of the audio signal within the temporal candidate region 48 so as to decide on the position of the temporally future-heading end 40 of temporal region 36. In the above specific details presented with respect to Figs. 1 to 3, a measure for a difference between maximum and minimum energy samples within the temporal candidate region 48 were used as the disparity measure, such an amplitude ratio therebetween. In particular, in the above concrete example, variable N_{new} measured the position of the temporally future-heading end 40 of temporal future 36 with respect to the temporally past-heading end 42 of the current frame 34a as indicated at 50 in Fig. 5.

As became clear from the above discussion, the placement of the temporal region 36 dependent on pitch lag 18 is advantageous in that the apparatus's 10 ability to correctly

identify situations where the harmonic filter tool 30 may advantageously be used is increased. In particular, the correct detection of such situations is made more reliable, i.e. such situations are detected at higher probability without substantially increasing falsely positive detection.

As was described above with respect to Figs. 1 to 3, the temporal structure analyzer 24 may determine the at least one temporal structure measure within the temporal region 36 on the basis of a temporal sampling of the audio signal's energy within that temporal region 36. This is illustrated in Fig. 6, where the energy samples are indicated by dots plotted in a time/energy plane spanned by arbitrary time and energy axes. As explained above, the energy samples 52 may have been obtained by sampling the energy of the audio signal at a sample rate higher than the frame rate of frames 34. In determining the at least one temporal structure measure 26, analyzer 24 may, as described above, compute for example a set of energy change values during a change between pairs of immediately consecutive energy samples 52 within temporal region 36. In the above description, equation 5 was used to this end. By way of this measure, an energy change value may be obtained from each pair of immediately consecutive energy samples 52. Analyzer 24 may then subject the set of energy change values obtained from the energy samples 52 within temporal region 36 to a scalar function to obtain the at least one structural energy measure 26. In the above concrete example, the temporal flatness measure, for example, has been determined on the basis of a sum over addends, each of which depends on exactly one of the set of energy change values. The maximum energy change, in turn, was determined according to equation 7 using a maximum operator applied onto the energy change values.

As already noted above, the energy samples 52 do not necessarily measure the energy of the audio signal 12 in its original, unmodified version. Rather, the energy sample 52 may measure the energy of the audio signal in some modified domain. In the concrete example above, for example, the energy samples measured the energy of the audio signal as obtained after high pass filtering the same. Accordingly, the audio signal's energy at a spectrally lower region influences the energy samples 52 less than spectrally higher components of the audio signal. Other possibilities exist, however, as well. In particular, it should be noted that the example where the temporal structure analyzer 24 merely uses one value of the at least one temporal structure measure 26 per sample time instant in accordance with the examples presented so far, is merely one embodiment and alternatives exist according to which the temporal structure analyzer determine the temporal structure measure in a spectrally discriminating manner so as to obtain one value of the at least one temporal structure measure per spectral band of a plurality of spectral bands. Accordingly,

the temporal structure analyzer 24 would then provide to the controller 28 more than one value of the at least one temporal structure measure 26 for the current frame 34a as determined within the temporal region 36, namely one per such spectral band, wherein the spectral bands partition, for example, the overall spectral interval of spectrogram 32.

Fig. 7 illustrates the apparatus 10 and its usage in an audio codec supporting the harmonic filter tool 30 according to the harmonic pre/post filter approach. Fig. 7 shows a transform-based encoder 70 as well as a transform-based decoder 72 with the encoder 70 encoding audio signal 12 into a data stream 74 and decoder 72 receiving the data stream 74 so as to reconstruct the audio signal either in spectral domain as illustrated at 76 or, optionally, in time-domain illustrated at 78. It should be clear that encoder and decoder 70 and 72 are discrete/separate entities and shown in Fig. 7 concurrently merely for illustration purposes.

The transform-based encoder 70 comprises a transformer 80 which subjects the audio signal 12 to a transform. Transformer 80 may use a lapped transform such a critically sampled lapped transform, an example of which is MDCT. In the example of Fig. 7, the transform-based audio encoder 70 also comprises a spectral shaper 82 which spectrally shapes the audio signal's spectrum as output by transformer 80. Spectral shaper 82 may spectrally shape the spectrum of the audio signal in accordance with a transfer function being substantially an inverse of a spectral perceptual function. The spectral perceptual function may be derived by way of linear prediction and thus, the information concerning the spectral perceptual function may be conveyed to the decoder 72 within data stream 74 in the form of, for example, linear prediction coefficients in the form of, for example, quantized line spectral pair of line spectral frequency values. Alternatively, a perceptual model may be used to determine the spectral perceptual function in the form of scale factors, one scale factor per scale factor band, which scale factor bands may, for example, coincide with bark bands. The encoder 70 also comprises a quantizer 84 which quantizes the spectrally shaped spectrum with, for example, a quantization function which is equal for all spectral lines. The thus spectrally shaped and quantized spectrum is conveyed within data stream 74 to decoder 72.

For the sake of completeness only, it should be noted that the order among transformer 80 and spectral shaper 82 has been chosen in Fig. 7 for illustration purposes only. Theoretically, spectral shaper 82 could cause the spectral shaping in fact within the time-domain, i.e. upstream transformer 80. Further, in order to determine the spectral perceptual function, spectral shaper 82 could have access to the audio signal 12 in time-domain although not specifically indicated in Fig. 7. At the decoder side, decoder 72 is illustrated in Fig. 7 as comprising a spectral shaper 86 configured to shape the inbound spectrally shaped and

quantized spectrum as obtained from data stream 74 with the inverse of the transfer function of spectral shaper 82, i.e. substantially with the spectral perceptual function, followed by an optional inverse transformer 88. The inverse transformer 88 performs the inverse transformation relative to transformer 80 and may, for example, to this end perform a transform block-based inverse transformation followed by an overlap-add-process in order to perform time-domain aliasing cancellation, thereby reconstructing the audio signal in time-domain.

As illustrated in Fig. 7, a harmonic pre-filter may be comprised by encoder 70 at a position upstream or downstream transformer 80. For example, a harmonic pre-filter 90 upstream transformer 80 may subject the audio signal 12 within the time-domain to a filtering so as to effectively attenuate the audio signal's spectrum at the harmonics in addition to the transfer function or spectral shaper 82. Alternatively, the harmonic pre-filter may be positioned downstream transformer 80 with such pre-filter 92 performing or causing the same attenuation in the spectral domain. As shown in Fig. 7, corresponding post-filters 94 and 96 are positioned within the decoder 72: in case of pre-filter 92, within spectral domain post-filter 94 positioned upstream inverse transformer 88 inversely shapes the audio signal's spectrum, inverse to the transfer function of pre-filter 92, and in case of pre-filter 90 being used, post filter 96 performs a filtering of the reconstructed audio signal in the time-domain, downstream inverse transformer 88, with a transfer function inverse to the transfer function of pre-filter 90.

In the case of Fig. 7, apparatus 10 controls the audio codec's harmonic filter tool implemented by pair 90 and 96 or 92 and 94 by explicitly signaling control signals 98 via the audio codec's data stream 74 to the decoding side for controlling the respective post-filter and, in line with the control of the post-filter at the decoding side, controlling the pre-filter at the encoder side.

For the sake of completeness, Fig. 8 illustrates the usage of apparatus 10 using a transform-based audio codec also involving elements 80, 82, 84, 86 and 88, however, here illustrating the case where the audio codec supports the harmonic post-filter-only approach. Here, the harmonic filter tool 30 may be embodied by a post-filter 100 positioned upstream the inverse transformer 88 within decoder 72, so as to perform harmonic post filtering in the spectral domain, or by use of a post-filter 102 positioned downstream inverse transformer 88 so as to perform the harmonic post-filtering within decoder 72 within the time-domain. The mode of operation of post-filters 100 and 102 is substantially the same as the one of post-filters 94 and 96: the aim of these post-filters is to attenuate the quantization noise between the harmonics. Apparatus 10 controls these post-filters via explicit signaling within data stream 74, the explicit signaling indicated in Fig. 8 using reference sign 104.

As already described above, the control signal 98 or 104 is sent, for example, on a regular basis, such as per frame 34. As to the frames, it is noted that same are not necessarily of equal length. The length of the frames 34 may also vary.

The above description, especially the one with regard to Fig. 2 and 3, revealed possibilities as to how controller 28 controls the harmonic filter tool. As became clear from that discussion, it may be that the at least one temporal structure measure measures an average or maximum energy variation of the audio signal within the temporal region 36. Further, the controller 28 may include, within its control options, the disablement of the harmonic filter tool 30. This is illustrated in Fig. 9. Fig. 9 shows the controller 28 as comprising a logic 120 configured to check whether a predetermined condition is met by the at least one temporal structure measure and the harmonicity measure, so as to obtain a check result 122, which is of binary nature and indicates whether or not the predetermined condition is fulfilled. Controller 28 is shown as comprising a switch 124 configured to switch between enabling and disabling the harmonic filter tool depending on the check result 122. If the check result 122 indicates that the predetermined condition has been approved to be met by logic 120, switch 124 either directly indicates the situation by way of control signal 14, or switch 124 indicates the situation along with a degree of filter gain for the harmonic filter tool 30. That is, in the latter case, switch 124 would not switch between switching off the harmonic filter tool 30 completely and switching on the harmonic filter tool 30 completely, only, but would set the harmonic filter tool 30 to some intermediate state varying in the filter strength or filter gain, respectively. In that case, i.e. if switch 124 also adapts/controls the harmonic filter tool 30 somewhere between completely switching off and completely switching on tool 30, switch 124 may rely on the at least temporal structure measure 26 and the harmonicity measure 22 so as to determine the intermediate states of control signal 14, i.e. so as to adapt tool 30. In other words, switch 124 could determine the gain factor or adaptation factor for controlling the harmonic filter tool 30 also on the basis of measures 26 and 22. Alternatively, switch 124 uses for all states of control signal 14 not indicating the off state of harmonic filter tool 30, the audio signal 12 directly. If the check result 122 indicates that a predetermined condition is not met, then the control signal 14 indicates the disablement of the harmonic filter tool 30.

As became clear from the above description of Figs. 2 and 3, the predetermined condition may be met if both the at least one temporal structure measure is smaller than a predetermined first threshold and the measure of harmonicity is, for a current frame and/or a previous frame, above a second threshold. An alternative may also exist: the predetermined condition may additionally be met if the measure of harmonicity is, for a current frame, above

a third threshold and the measure of harmonicity is, for a current frame and/or a previous frame, above a fourth threshold which decreases with an increase of the pitch lag.

In particular, in the example of Figs. 2 and 3, there were actually three alternatives for which the predetermined condition is met, the alternatives being dependent on the at least one temporal structure measure:

1. One temporal structure measure < threshold and combined harmonicity for current and previous frame > second threshold;
2. One temporal structure measure < third threshold and (harmonicity for current or previous frame) > fourth threshold;
3. (One temporal structure measure < fifth threshold or all temp. measures < thresholds) and harmonicity for current frame > sixth threshold.

Thus, Fig. 2 and Fig. 3, reveal possible implementation examples for logic 124.

As has been illustrated above with respect to Figs. 1 to 3, it is feasible that apparatus 10 is not only used for controlling a harmonic filter tool of an audio codec. Rather, the apparatus 10 may form, along with a transient detection, a system able to perform both control of the harmonic filter tool as well as detecting transients. Fig. 10 illustrates this possibility. Fig. 10 shows a system 150 composed of apparatus 10 and a transient detector 152, and while apparatus 10 outputs control signal 14 as discussed above, transient detector 152 is configured to detect transients in the audio signal 12. To do this, however, the transient detector 152 exploits an intermediate result occurring within apparatus 10: the transient detector 152 uses for its detection the energy samples 52 temporally or, alternatively, spectro-temporally sampling the energy of the audio signal, with, however, optionally evaluating the energy samples within a temporal region other than temporal region 36 such as within current frame 34a, for example. On the basis of these energy samples, transient detector 152 performs the transient detection and signals the transients detected by way of a detection signal 154. In case of the above example, the transient detection signal substantially indicated positions where the condition of equation 4 is fulfilled, i.e. where an energy change of temporally consecutive energy samples exceeds some threshold.

As also became clear from the above discussion, a transform-based encoder such as the one depicted in Fig. 8 or a transform-coded excitation encoder, may comprise or use the system of Fig. 10 so as to switch a transform block and/or overlap length depending on the transient detection signal 154. Further, additionally or alternatively, an audio encoder

comprising or using the system of Fig. 10 may be of a switching mode type. For example, USAC and EVS use switching between modes. Thus, such an encoder could be configured to support switching between a transform coded excitation mode and a code excited linear prediction mode and the encoder could be configured to perform the switching dependent on the transient detection signal 154 of the system of Fig. 10. As far as the transform coded excitation mode is concerned, the switching of the transform block and/or overlap length could, again, be dependent on the transient detection signal 154.

Examples for the advantages of the above embodiments

Example 1:

The size of the region in which temporal measures for the LTP decision are calculated is dependent on the pitch (see equation (8)) and this region is different from the region where temporal measures for the transform length are calculated (usually current frame plus look-ahead).

In the example in Fig. 11 the transient is inside the region where the temporal measures are calculated and thus influences the LTP decision. The motivation, as stated above, is that a LTP for the current frame, utilizing past samples from the segment denoted by "pitch lag", would reach into a portion of the transient.

In the example in Fig. 12 the transient is outside the region where the temporal measures are calculated and thus doesn't influence the LTP decision. This is reasonable since, unlike in the previous figure, a LTP for the current frame would not reach into the transient.

In both examples (Fig. 11 and Fig. 12) the transform length configuration is decided on temporal measures only within the current frame, i.e. the region marked with "frame length". This means that in both examples, no transient would be detected in the current frame and preferably, a single long transform (instead of many successive short transforms) would be employed.

Example 2:

Here we discuss the behavior of the LTP for impulse and step transients within harmonic signal, of which one example is given by signal's spectrogram in Fig. 13.

When coding the signal includes the LTP for the complete signal (because the LTP decision is based only on the pitch gain), the spectrogram of the output looks as presented in Fig. 14.

The waveform of the signal, which spectrogram is in Fig. 14, is presented in Fig. 15. The Fig. 15 also includes the same signal Low-pass (LP) filtered and High-pass (HP) filtered. In the LP filtered signal the harmonic structure becomes clearer and in the HP filtered signal the location of the impulse like transient and its trail is more evident. The level of the complete signal, LP signal and HP signal is modified in the figure for the sake of the presentation.

For short impulse like transients (as the first transient in Fig. 13), the long term prediction produces repetitions of the transient as can be seen in Fig. 14 and Fig. 15. Using the long term prediction during the step like long transients (as the second transient in Fig. 13) doesn't introduce any additional distortions as the transient is strong enough for longer period and thus masks (simultaneous and post-masking) the portions of the signal constructed using the long term prediction. The decision mechanism enables the LTP for step like transients (to exploit the benefit of prediction) and disables the LTP for short impulse like transient (to prevent artifacts).

In Fig. 16 and Fig. 17, the energies of segments computed in transient detector are shown. Fig. 16 shows impulse like transient Fig. 17 shows step like transient. For impulse like transient in Fig. 16 the temporal features are calculated on the signal containing the current frame (N_{new} segments) and the past frame up to the pitch lag (N_{past} segments), since the

ratio $\frac{E_{TD}(i_{max})}{E_{TD}(i_{min})}$ is above the threshold ($\frac{1}{0.375}$). For the step like transient in Fig. 17, the ratio

$\frac{E_{TD}(i_{max})}{E_{TD}(i_{min})}$ is below the threshold ($\frac{1}{0.375}$) and thus only the energies from segments -8, -7 and

-6 are used in the calculation of the temporal measures. These different choices of the segments where the temporal measures are calculated, leads to determination of much higher energy fluctuations for impulse like transients and thus to disabling the LTP for impulse like transients and enabling the LTP for step like transients.

Example 3:

However in some cases the usage of the temporal measures may be disadvantageous. The spectrogram in Fig. 18 and the waveform in Fig. 19 display an excerpt of about 35 milliseconds from the beginning of "Kalifornia" by Fatboy Slim.

The LTP decision that is dependent on the Temporal Flatness Measure and on the Maximum Energy Change disables the LTP for this type of signal as it detects huge temporal fluctuations of energy.

This sample is an example of ambiguity between transients and train of pulses that form low pitched signal.

As can be seen in Fig. 20, where the 600 milliseconds excerpt from the same signal the signal is presented, the signal contains repeated very short impulse like transient (the spectrogram is produced using short length FFT).

As can be seen in the same 600 milliseconds excerpt in Fig. 21 the signal looks as if it contains very harmonic signal with low and changing pitch (the spectrogram is produced using long length FFT).

This kind of signals benefit from the LTP as there is clear repetitive structure (equivalent to clear harmonic structure). Since there is clear energy fluctuation (that can be seen in Fig. 18, Fig. 19 and Fig. 20), the LTP would be disabled due to exceeding threshold for the Temporal Flatness Measure or for the Maximum Energy Change. However, in our proposal, the LTP is enabled due to the normalized correlation exceeding the threshold dependent on the pitch lag ($\text{norm_corr}(\text{curr}) \leq 1.2 - T_{\text{int}}/L$).

Thus, above embodiments, inter alias, revealed, for example, a concept for a better harmonic filter decision for audio coding. It must be restated in passing that slight deviations from said concept are feasible. In particular, as noted above, the audio signal 12 may be a speech or music signal and may be replaced by a pre-processed version of signal 12 for the purpose of pitch estimation, harmonicity measurement, or temporal structure analysis or measurement. Also, the pitch estimation may not be limited to measurements of pitch lags but, as should be known to those skilled in the art, may also be performed via measurements of a fundamental frequency, in the time or a spectral domain, which can easily be converted into an equivalent pitch lag by way of an equation such as "pitch lag = sampling frequency / pitch frequency". Thus, generally speaking, the pitch estimator 16 estimates the audio signal's pitch which, in turn, is manifests itself in pitch-lag and pitch frequency.

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.

The inventive encoded audio signal can be stored on a digital storage medium or can be transmitted on a transmission medium such as a wireless transmission medium or a wired transmission medium such as the Internet.

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

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

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

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

Claims

1. Apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising

a pitch estimator configured to determine a pitch of an audio signal to be processed by the audio codec;

a harmonicity measurer configured to determine a measure of harmonicity of the audio signal using the pitch;

a temporal structure analyzer configured to determine, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

a controller configured to control the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the temporal structure analyzer is configured to determine the at least one temporal structure measure within a temporal region temporally placed depending on the pitch, and

wherein the temporal structure analyzer is configured to position a temporally past-heading end of the temporal region, or of a region of higher influence onto the determination of the temporal structure measure, depending on the pitch.
2. Apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising

a pitch estimator configured to determine a pitch of an audio signal to be processed by the audio codec;

a harmonicity measurer configured to determine a measure of harmonicity of the audio signal using the pitch;

a temporal structure analyzer configured to determine, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

a controller configured to control the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the harmonicity measurer is configured to determine the measure of harmonicity by computing a normalized correlation of the audio signal or a pre-modified version thereof at or around a pitch-lag of the pitch.

3. Apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising

a pitch estimator configured to determine a pitch of an audio signal to be processed by the audio codec;

a harmonicity measurer configured to determine a measure of harmonicity of the audio signal using the pitch;

a temporal structure analyzer configured to determine, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

a controller configured to control the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the pitch estimator is configured to determine the pitch in stages comprising a first stage and a second stage.

4. Apparatus according to claim 3, wherein the pitch estimator is configured to, within the first stage, determine a preliminary estimation of the pitch at a down-sampled domain of a first sample rate and, within the second stage, refine the preliminary estimation of the pitch at a second sample rate, higher than the first sample rate.
5. Apparatus according to any of the previous claims, wherein the pitch estimator is configured to determine the pitch using autocorrelation.
6. Apparatus according to any of the previous claims, wherein the temporal structure analyzer is configured to determine the at least one temporal structure measure within a temporal region temporally placed depending on the pitch.
7. Apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising

a pitch estimator configured to determine a pitch of an audio signal to be processed by the audio codec;

a harmonicity measurer configured to determine a measure of harmonicity of the audio signal using the pitch;

a temporal structure analyzer configured to determine, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

a controller configured to control the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the temporal structure analyzer is configured to determine the at least one temporal structure measure within a temporal region temporally placed depending on the pitch, and

wherein the temporal structure analyzer is configured to position the temporal past-heading end of the temporal region or, of the region of higher influence onto the determination of the temporal structure measure, such that the temporally past-heading end of the temporal region or, of the region of higher influence onto the determination of the temporal structure measure, is displaced into past direction by a temporal amount monotonically increasing with a decrease of the pitch.

8. Apparatus according to claim 7, wherein the temporal structure analyzer is configured to position a temporally future-heading end of the temporal region or, of the region of higher influence onto the determination of the temporal structure measure, depending on the temporal structure of the audio signal within a temporal candidate region extending from the temporally past-heading end of the temporal region, or of the region of higher influence onto the determination of the temporal structure measure, to a temporally future-heading end of a current frame.

9. Apparatus according to claim 8, wherein the temporal structure analyzer is configured to use an amplitude or ratio between maximum and minimum energy samples within the temporal candidate region in order to position the temporally future-heading end of the temporal region or, of the region of higher influence onto the determination of the temporal structure measure.
10. Apparatus according to any of the previous claims, wherein the controller comprises
 - a logic configured to check whether a predetermined condition is met by the at least one temporal structure measure and the measure of harmonicity so as to obtain a check result; and
 - a switch configured to switch between enabling and disabling the harmonic filter tool depending on the check result.
11. Apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising
 - a pitch estimator configured to determine a pitch of an audio signal to be processed by the audio codec;
 - a harmonicity measurer configured to determine a measure of harmonicity of the audio signal using the pitch;
 - a temporal structure analyzer configured to determine, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

a controller configured to control the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the controller comprises

a logic configured to check whether a predetermined condition is met by the at least one temporal structure measure and the measure of harmonicity so as to obtain a check result; and

a switch configured to switch between enabling and disabling the harmonic filter tool depending on the check result,

wherein the at least one temporal structure measure measures an average or maximum energy variation of the audio signal within the temporal region and the logic is configured such that the predetermined condition is met if

both the at least one temporal structure measure is smaller than a predetermined first threshold and the measure of harmonicity is, for a current frame and/or a previous frame, above a second threshold.

12. Apparatus according to claim 11, wherein the logic is configured such that the predetermined condition is also met if

the measure of harmonicity is, for a current frame, above a third threshold, and the measure of harmonicity is, for a current frame and/or a previous frame, above a fourth threshold which decreases with an increase of a pitch lag of the pitch.

13. Apparatus according to any of the previous claims, wherein the controller is configured to control the harmonic filter tool by

explicitly signaling a control signal via an audio codec's data stream to a decoding side; or

explicitly signaling a control signal via an audio codec's data stream to a decoding side for controlling a post-filter at the decoding side and, in line with the control of the post-filter at the decoding side, controlling a pre-filter at an encoder side.

14. Apparatus according to any of the previous claims, wherein the temporal structure analyzer is configured to determine the at least one temporal structure measure in a spectrally discriminating manner so as to obtain one value of the at least one temporal structure measure per spectral band of a plurality of spectral bands.

15. Apparatus according to any of the previous claims, wherein the controller is configured to control the harmonic filter tool at units of frames, and the temporal structure analyzer is configured to sample an energy of the audio signal at a sample rate higher than a frame rate of the frames so as to obtain energy samples of the audio signal and to determine the at least one temporal structure measure on the basis of the energy samples.

16. Apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising

a pitch estimator configured to determine a pitch of an audio signal to be processed by the audio codec;

a harmonicity measurer configured to determine a measure of harmonicity of the audio signal using the pitch;

a temporal structure analyzer configured to determine, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

a controller configured to control the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the controller is configured to control the harmonic filter tool at units of frames, and the temporal structure analyzer is configured to sample an energy of the audio signal at a sample rate higher than a frame rate of the frames so as to obtain energy samples of the audio signal and to determine the at least one temporal structure measure on the basis of the energy samples,

wherein the temporal structure analyzer is configured to determine the at least one temporal structure measure within a temporal region temporally placed depending on the pitch and the temporal structure analyzer is configured to determine the at least one temporal structure measure on the basis of the energy samples by computing a set of energy change values measuring a change between pairs of immediately consecutive energy samples of the energy samples within the temporal region and subjecting the set of energy change values to a scalar function including a maximum operator or a sum over addends each of which depends on exactly one of the set of energy change values.

17. Apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising

a pitch estimator configured to determine a pitch of an audio signal to be processed by the audio codec;

a harmonicity measurer configured to determine a measure of harmonicity of the audio signal using the pitch;

a temporal structure analyzer configured to determine, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

a controller configured to control the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the controller is configured to control the harmonic filter tool at units of frames, and the temporal structure analyzer is configured to sample an energy of the audio signal at a sample rate higher than a frame rate of the frames so as to obtain energy samples of the audio signal and to determine the at least one temporal structure measure on the basis of the energy samples,

wherein the temporal spectrum analyzer is configured to perform the sampling of the energy of the audio signal within a high-pass filtered domain.

18. Apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising

a pitch estimator configured to determine a pitch of an audio signal to be processed by the audio codec;

a harmonicity measurer configured to determine a measure of harmonicity of the audio signal using the pitch;

a temporal structure analyzer configured to determine, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

a controller configured to control the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the pitch estimator, the harmonicity measurer and the temporal structure analyzer perform its determination based on different versions of the audio signal including the original audio signal and some pre-modified version thereof.

19. Apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising

a pitch estimator configured to determine a pitch of an audio signal to be processed by the audio codec;

a harmonicity measurer configured to determine a measure of harmonicity of the audio signal using the pitch;

a temporal structure analyzer configured to determine, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

a controller configured to control the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the controller is configured to, in controlling the harmonic filter tool, depending on the temporal structure measure and the measure of harmonicity

switch between enabling and disabling a pre-filter and/or a post-filter of the harmonic filter tool, or

gradually adapt a filter strength of the pre-filter and/or the post-filter of the harmonic filter tool,

wherein the harmonic filter tool is of a pre-filter plus post-filter approach and the pre-filter of the harmonic filter tool is configured to increase the quantization noise within a harmonic of the pitch of the audio signal and the post-filter of the harmonic filter tool is configured to reshape a transmitted spectrum accordingly, or the harmonic filter tool is of a post-filter only approach and the post-filter of the harmonic filter tool is configured to filter quantization noise occurring between the harmonics of the pitch of the audio signal.

20. Audio encoder or audio decoder, comprising a harmonic filter tool and the apparatus for performing a harmonicity-dependent controlling of the harmonic filter tool according to any of the previous claims.

21. System comprising

an apparatus for performing a harmonicity-dependent controlling of a harmonic filter tool according to any of claims 15 to 17, and

a transient detector configured to detect transients in an audio signal to be processed by the audio codec on the basis of the energy samples.

22. Transform-based encoder comprising the system of claim 21, configured to switch a transform block and/or overlap length depending on the detected transients.
23. Audio encoder comprising the system of claim 21, configured to support switching between a transform coded excitation mode and a code excited linear prediction mode depending on the detected transients.
24. Audio encoder according to claim 23, configured to switch a transform block and/or overlap length in the transform coded excitation mode depending on the detected transients.
25. Method for performing a harmonicity-dependent controlling of a harmonic filter tool of an audio codec, comprising

determining a pitch of an audio signal to be processed by the audio codec;

determining a measure of harmonicity of the audio signal using the pitch;

determining, depending on the pitch, at least one temporal structure measure measuring a characteristic of a temporal structure of the audio signal;

controlling the harmonic filter tool depending on the temporal structure measure and the measure of harmonicity,

wherein the at least one temporal structure measure is determined within a temporal region temporally placed depending on the pitch, and

wherein a temporally past-heading end of the temporal region, or of a region of higher influence onto the determination of the temporal structure measure, is positioned depending on the pitch.

26. Computer program having a program code for performing, when running on a computer, a method according to claim 25.