

Apparatus and Method for Processing a Decoded Audio Signal in a Spectral Domain

Specification

5

The present invention relates to audio processing and, in particular, to the processing of a decoded audio signal for the purpose of quality enhancement.

10 Recently, further developments regarding switched audio codecs have been achieved. A high quality and low bit rate switched audio codec is the unified speech and audio coding concept (USAC concept). There is a common pre/post-processing consisting of an MPEG surround (MPEGs) functional unit to handle a stereo or multichannel processing and an enhanced SBR (eSBR) unit which handles the parametric representation of the higher audio frequencies in the input signal. Subsequently there are two branches, one consisting
15 of an advanced audio coding (AAC) tool path and the other consisting of a linear prediction coding (LP or LPC domain) based path which, in turn, features either a frequency domain representation or a time domain representation of the LPC residual. All transmitted spectra for both AAC and LPC are represented in the MDCT domain following quantization and arithmetic coding. The time domain representation uses an ACELP
20 excitation coding scheme. Block diagrams of the encoder and the decoder are given in Fig. 1.1 and Fig. 1.2 of ISO/IEC CD 23003-3.

An additional example for a switched audio codec is the extended adaptive multi-rate-wide band (AMR-WB+) codec as described in 3GPP TS 26.290 V10.0.0 (2011-3). The AMR-
25 WB+ audio codec processes input frames equal to 2048 samples at an internal sampling frequency F_s . The internal sampling frequencies are limited to the range 12800 to 38400 Hz. The 2048-sample frames are split into two critically sampled equal frequency bands. This results in two super frames of 1024 samples corresponding to the low frequency (LF) and high frequency (HF) band. Each super frame is divided into four 256-sample frames.
30 Sampling at the internal sampling rate is obtained by using a variable sampling conversion scheme which re-samples the input signal. The LF and HF signals are then encoded using two different approaches: the LF is encoded and decoded using a “core” encoder/decoder, based on switched ACELP and transform coded excitation (TCX). In the ACELP mode, the standard AMR-WB codec is used. The HF signal is encoded with relatively few bits
35 (16 bits per frame) using a bandwidth extension (BWE) method. The AMR-WB coder includes a pre-processing functionality, an LPC analysis, an open loop search functionality, an adaptive codebook search functionality, an innovative codebook search functionality and memories update. The ACELP decoder comprises several functionalities such as

decoding the adaptive codebook, decoding gains, decoding the innovative codebook, decode ISP, a long term prediction filter (LTP filter), the construct excitation functionality, an interpolation of ISP for four sub-frames, a post-processing, a synthesis filter, a de-emphasis and an up-sampling block in order to finally obtain the lower band portion of the speech output. The higher band portion of the speech output is generated by gains scaling using an HB gain index, a VAD flag, and a 16 kHz random excitation. Furthermore, an HB synthesis filter is used followed by a band pass filter. More details are in Fig. 3 of G.722.2.

This scheme has been enhanced in the AMR-WB+ by performing a post-processing of the mono low-band signal. Reference is made to Figs. 7, 8 and 9 illustrating the functionality in AMR-WB+. Fig. 7 illustrates pitch enhancer 700, a low pass filter 702, a high pass filter 704, a pitch tracking stage 706 and an adder 708. The blocks are connected as illustrated in Fig. 7 and are fed by the decoded signal.

In the low-frequency pitch enhancement, two-band decomposition is used and adaptive filtering is applied only to the lower band. This results in a total post-processing that is mostly targeted at frequencies near the first harmonics of the synthesized speech signal. Fig. 7 shows the block diagram of the two-band pitch enhancer. In the higher branch the decoded signal is filtered by the high pass filter 704 to produce the higher band signals s_H . In the lower branch, the decoded signal is first processed through the adaptive pitch enhancer 700 and then filtered through the low pass filter 702 to obtain the lower band post-process signal (s_{LEE}). The post-process decoded signal is obtained by adding the lower band post-process signal and the higher band signal. The object of the pitch enhancer is to reduce the inter-harmonic noise in the decoded signal which is achieved by a time-varying linear filter with a transfer function H_E indicated in the first line of Fig. 9 and described by the equation in the second line of Fig. 9. α is a coefficient that controls the inter-harmonic attenuation. T is the pitch period of the input signal $\hat{S}(n)$ and $s_{LE}(n)$ is the output signal of the pitch enhancer. Parameters T and α vary with time and are given by the pitch tracking module 706 with a value of $\alpha = 1$, the gain of the filter described by the equation in the second line of Fig. 9 is exactly zero at frequencies $1/(2T)$, $3/(2T)$, $5/(2T)$, etc, i.e., at the mid-point between the DC (0 Hz) and the harmonic frequencies $1/T$, $3/T$, $5/T$, etc. When α approaches zero, the attenuation between the harmonics produced by the filter as defined in the second line of Fig. 9 decreases. When α is zero, the filter has no effect and is an all-pass. To confine the post-processing to the low frequency region, the enhanced signal s_{LE} is low pass filtered to produce the signal s_{LEF} which is added to the high pass filter signal s_H to obtain the post-process synthesis signal s_E .

Another configuration equivalent to the illustration in Fig. 7 is illustrated in Fig. 8 and the configuration in Fig. 8 eliminates the need to high pass filtering. This is explained with respect to the third equation for s_E in Fig. 9. The $h_{LP}(n)$ is the impulse response of the low pass filter and $h_{HP}(n)$ is the impulse response of the complementary high pass filter. Then, the post-process signal $s_{E(n)}$ is given by the third equation in Fig. 9. Thus, the post processing is equivalent to subtracting the scaled low pass filtered long-term error signal $\alpha \cdot e_{LT}(n)$ from the synthesis signal $\hat{s}(n)$. The transfer function of the long-term prediction filter is given as indicated in the last line of Fig. 9. This alternative post-processing configuration is illustrated in Fig. 8. The value T is given by the received closed-loop pitch lag in each subframe (the fractional pitch lag rounded to the nearest integer). A simple tracking for checking pitch doubling is performed. If the normalized pitch correlation at delay $T/2$ is larger than 0.95 then the value $T/2$ is used as the new pitch lag for post-processing. The factor α is given by $\alpha = 0.5g_p$, constrained to α greater than or equal to zero and lower than or equal to 0.5. g_p is the decoded pitch gain bounded between 0 and 1. In TCX mode, the value of α is set to zero. A linear phase FIR low pass filter with 25 coefficients is used with the cut-off frequency of about 500 Hz. The filter delay is 12 samples). The upper branch needs to introduce a delay corresponding to the delay of the processing in the lower branch in order to keep the signals in the two branches time aligned before performing the subtraction. In AMR-WB+ $F_s=2x$ sampling rate of the core. The core sampling rate is equal to 12800 Hz. So the cut-off frequency is equal to 500Hz.

It has been found that, particularly for low delay applications, the filter delay of 12 samples introduced by the linear phase FIR low pass filter contributes to the overall delay of the encoding/decoding scheme. There are other sources of systematic delays at other places in the encoding/decoding chain, and the FIR filter delay accumulates with the other sources.

It is an object of the present invention to provide an improved audio signal processing concept which is better suited for real time applications or two-way communication scenarios such as mobile phone scenarios.

This object is achieved by an apparatus for processing a decoded audio signal in accordance with claim 1 or a method of processing a decoded audio signal in accordance with claim 15 or a computer program in accordance with claim 16.

The present invention is based on the finding that the contribution of the low pass filter in the bass post filtering of the decoded signal to the overall delay is problematic and has to be reduced. To this end, the filtered audio signal is not low pass filtered in the time domain but is low pass filtered in the spectral domain such as a QMF domain or any other spectral

domain, for example, an MDCT domain, an FFT domain, etc. It has been found that the transform from the spectral domain into the frequency domain and, for example, into a low resolution frequency domain such as a QMF domain can be performed with low delay and the frequency-selectivity of the filter to be implemented in the spectral domain can be implemented by just weighting individual subband signals from the frequency domain representation of the filtered audio signal. This “impression” of the frequency-selected characteristic is, therefore, performed without any systematic delay since a multiplying or weighting operation with a subband signal does not incur any delay. The subtraction of the filtered audio signal and the original audio signal is performed in the spectral domain as well. Furthermore, it is preferred to perform additional operations which are, for example, necessary anyway, such as a spectral band replication decoding or a stereo or a multichannel decoding are additionally performed in one and the same QMF domain. A frequency-time conversion is performed only at the end of the decoding chain in order to bring the finally produced audio signal back into the time domain. Hence, depending on the application, the result audio signal generated by the subtractor can be converted back into the time domain as it is when no additional processing operations in the QMF domain are required anymore. However, when the decoding algorithm has additional processing operations in the QMF domain, then the frequency-time converter is not connected to the subtractor output but is connected to the output of the last frequency domain processing device.

Preferably, the filter for filtering the decoded audio signal is a long term prediction filter. Furthermore, it is preferred that the spectral representation is a QMF representation and it is additionally preferred that the frequency-selectivity is a low pass characteristic.

However, any other filters different from a long term prediction filter, any other spectral representations different from a QMF representation or any other frequency-selectivity different from a low pass characteristic can be used in order to obtain a low-delay post-processing of a decoded audio signal.

Preferred embodiments of the present invention are subsequently described with respect to the accompanying drawings in which:

Fig. 1a is a block diagram of an apparatus for processing a decoded audio signal in accordance with an embodiment;

Fig. 1b is a block diagram of a preferred embodiment for the apparatus for processing a decoded audio signal;

- Fig. 2a illustrates a frequency-selective characteristic exemplarily as a low pass characteristic;
- 5 Fig. 2b illustrates weighting coefficients and associated subbands;
- Fig. 2c illustrates a cascade of the time/spectral converter and a subsequently connected weighter for applying weighting coefficients to each individual subband signal;
- 10 Fig. 3 illustrates an impulse response in the frequency response of the low pass filter in AMR-WB+ illustrated in Fig. 8;
- Fig. 4 illustrates an impulse response and the frequency response transformed into the QMF domain;
- 15 Fig. 5 illustrates weighting factors for the weighters for the example of 32 QMF subbands;
- 20 Fig. 6 illustrates the frequency response for 16 QMF bands and the associated 16 weighting factors;
- Fig. 7 illustrates a block diagram of the low frequency pitch enhancer of AMR-WB+;
- 25 Fig. 8 illustrates an implemented post-processing configuration of AMR-WB+;
- Fig. 9 illustrates a derivation of the implementation of Fig. 8; and
- 30 Fig. 10 illustrates a low delay implementation of the long term prediction filter in accordance with an embodiment.

Fig. 1a illustrates an apparatus for processing a decoded audio signal on line 100. The decoded audio signal on line 100 is input into the filter 102 for filtering the decoded audio signal to obtain a filtered audio signal on line 104. The filter 102 is connected to a time-spectral converter stage 106 illustrated as two individual time-spectral converters 106a for the filtered audio signal and 106b for the decoded audio signal on line 100. The time-spectral converter stage is configured for converting the audio signal and the filtered audio

35

signal into a corresponding spectral representation each having a plurality of subband signals. This is indicated by double lines in Fig. 1a, which indicates that the output of blocks 106a, 106b comprises a plurality of individual subband signals rather than a single signal as illustrated for the input into blocks 106a, 106b.

5

The apparatus for processing additionally comprises a weighter 108 for performing a frequency-selective weighting of the filtered audio signal output by block 106a by multiplying individual subband signals by respective weighting coefficients to obtain a weighted filtered audio signal on line 110.

10

Furthermore, a subtractor 112 is provided. The subtractor is configured for performing a subband-wise subtraction between the weighted filtered audio signal and the spectral representation of the audio signal generated by block 106b.

15 Furthermore, a spectral-time converter 114 is provided. The spectral-time conversion performed by block 114 is so that the result audio signal generated by the subtractor 112 or a signal derived from the result audio signal is converted into a time domain representation to obtain the processed decoded audio signal on line 116.

20 Although Fig. 1a indicates that the delay by time-spectral conversion and weighting is significantly lower than delay by FIR filtering, this is not necessary in all circumstances, since in situations, in which the QMF is absolutely necessary cumulating the delays of FIR filtering and of QMF is avoided. Hence, the present invention is also useful, when the delay by time-spectral conversion weighting is even higher than the delay of an FIR filter
25 for bass post filtering.

Fig. 1b illustrates a preferred embodiment of the present invention in the context of the USAC decoder or the AMR-WB+ decoder. The apparatus illustrated in Fig. 1b comprises an ACELP decoder stage 120, a TCX decoder stage 122 and a connection point 124 where
30 the outputs of the decoders 120, 122 are connected. Connection point 124 starts two individual branches. The first branch comprises the filter 102 which is, preferably, configured as a long term prediction filter which is set by the pitch lag T followed by an amplifier 129 of an adaptive gain α . Furthermore, the first branch comprises the time-spectral converter 106a which is preferably implemented as a QMF analysis filterbank.
35 Furthermore, the first branch comprises the weighter 108 which is configured for weighting the subband signals generated by the QMF analysis filterbank 106a.

In the second branch, the decoded audio signal is converted into the spectral domain by the QMF analysis filterbank 106b.

Although the individual QMF blocks 106a, 106b are illustrated as two separate elements, it is noted that, for analyzing the filtered audio signal and the audio signal, it is not necessarily required to have two individual QMF analysis filterbanks. Instead, a single QMF analysis filterbank and a memory may be sufficient, when the signals are transformed one after the other. However, for very low delay implementations, it is preferred to use individual QMF analysis filterbanks for each signal so that the single QMF block does not form the bottleneck of the algorithm.

Preferably, the conversion into the spectral domain and back into the time domain is performed by an algorithm, having a delay for the forward and backward transform being smaller than the delay of the filtering in the time domain with the frequency selective characteristic. Hence, the transforms should have an overall delay being smaller than the delay of the filter in question. Particularly useful are low resolution transforms such as QMF-based transforms, since the low frequency resolution results in the need for a small transform window, i.e., in a reduced systematic delay. Preferred applications only require a low resolution transform decomposing the signal in less than 40 subbands, such as 32 or only 16 subbands. However, even in applications where the time-spectral conversion and weighting introduce a higher delay than the low pass filter, an advantage is obtained due to the fact that a cumulating of delays for the low pass filter and the time-spectral conversion necessary anyway for other procedures is avoided.

For applications, however, which anyway require a time frequency conversion due to other processing operations, such as resampling, SBR or MPS, a delay reduction is obtained irrespective of the delay incurred by the time-frequency or frequency-time conversion, since the "inclusion" of the filter implementation into the spectral domain, the time domain filter delay is completely saved due to the fact that the subband-wise weighting is performed without any systematic delay.

The adaptive amplifier 129 is controlled by a controller 130. The controller 130 is configured for setting the gain α of amplifier 129 to zero, when the input signal is a TCX-decoded signal. Typically, in switched audio codecs such as USAC or AMR-WB+, the decoded signal at connection point 124 is typically either from the TCX-decoder 122 or from the ACELP-decoder 120. Hence, there is a time-multiplex of decoded output signals of the two decoders 120, 122. The controller 130 is configured for determining for a current time instant, whether the output signal is from a TCX-decoded signal or an

ACELP-decoded signal. When it is determined that there is a TCX signal, then the adaptive gain α is set to zero so that the first branch consisting of elements 102, 129, 106a, 108 does not have any significance. This is due to the fact that the specific kind of post filtering used in AMR-WB+ or USAC is only required for the ACELP-coded signal.

5 However, when other post filtering implementations apart from harmonic filtering or pitch enhancing is performed, then a variable gain α can be set differently depending on the needs.

10 When, however, the controller 130 determines that the currently available signal is an ACELP-decoded signal, then the value of amplifier 129 is set to the right value for α which typically is between 0 and 0.5. In this case, the first branch is significant and the output signal of the subtractor 112 is substantially different from the originally decoded audio signal at connection point 124.

15 The pitch information (pitch lag and gain alpha) used in filter 120 and amplifier 128 can come from the decoder and/or a dedicated pitch tracker. Preferably, the information are coming from the decoder and then re-processed (refined) through a dedicated pitch tracker/long term prediction analysis of the decoded signal.

20 The result audio signal generated by subtractor 112 performing the per band or per subband subsection is not immediately performed back into the time domain. Instead, the signal is forwarded to an SBR decoder module 128. Module 128 is connected to a mono-stereo or mono-multichannel decoder such as an MPS decoder 131, where MPS stands for MPEG surround.

25 Typically, the number of bands is enhanced by the spectral bandwidth replication decoder which is indicated by the three additional lines 132 at the output of block 128.

30 Furthermore, the number of outputs is additionally enhanced by block 131. Block 131 generates, from the mono-signal at the output of block 129 a, for example, 5-channel signal or any other signal having two or more channels. Exemplarily, a 5-channel scenario have a left channel L, a right channel R, a center channel C, a left surround channel L_s and a right surround channel R_s is illustrated. The spectral-time converter 114 exists, therefore, for each of the individual channels, i.e., exists five times in Fig. 1b in order to convert each individual channel signal from the spectral domain which is, in the Fig. 1b example, the QMF domain, back into the time domain at the output of block 114. Again, there is not necessarily a plurality of individual spectral-time converters. There can be a single one as well which processes the conversions one after the other. However, when a very low delay

35

implementation is required, it is preferred to use an individual spectral time converter for each channel.

5 The present invention is advantageous in that the delay introduced by the bass post filter and, specifically, by the implementation of the low pass filter FIR filter is reduced. Hence, any kind of frequency-selective filtering does not introduce an additional delay with respect to the delay required for the QMF or, stated generally, the time/frequency transform.

10 The present invention is particularly advantageous, when a QMF or, generally, a time-frequency transform is required anyway as, for example, in the case of Fig. 1b, where the SBR functionality and the MPS functionality are performed in the spectral domain anyway. An alternative implementation, where a QMF is required is, when a resampling is performed with the decoded signal, and when, for the purpose of resampling, a QMF
15 analysis filterbank and a QMF synthesis filterbank with a different number of filterbank channels is required.

Furthermore, a constant framing between ACELP and TCX is maintained due to the fact that both signals, i.e., TCX and ACELP now have the same delay.

20

The functionality of a bandwidth extension decoder 129 is described in detail in section 6.5 of ISO/IEC CD 23003-3. The functionality of the multichannel decoder 131 is described in detail, for example, in section 6.11 of ISO/IEC CD 23003-3. The functionalities behind the TCX decoder and ACELP decoder are described in detail in blocks 6.12 to 6.17 of
25 ISO/IEC CD 23003-3.

Subsequently, Figs. 2a to 2c are discussed in order to illustrate a schematic example. Fig. 2a illustrates a frequency-selected frequency response of a schematic low pass filter.

30 Fig. 2b illustrates the weighting indices for the subband numbers or subbands indicated in Fig. 2a. In the schematic case of Fig. 2a, subbands 1 to 6 have weighting coefficients equal to 1, i.e., no weighting and bands 7 to 10 have decreasing weighting coefficients and bands 11 to 14 have zeros.

35 A corresponding implementation of a cascade of a time-spectral converter such as 106a and the subsequently connector weighter 108 is illustrated in Fig. 2c. Each subband 1, 2 ..., 14 is input into an individual weighting block indicated by $W_1, W_2, \dots, W_{14}$. The weighter 108 applies the weighting factor of the table of Fig. 2b to each individual subband

signal by multiplying each sampling of the subband signal by the weighting coefficient. Then, at the output of the weighter, there exist weighted subband signals which are then input into the subtractor 112 of Fig. 1a which additionally performs a subtraction in the spectral domain.

5

Fig. 3 illustrates the impulse response and the frequency response of the low pass filter in Fig. 8 of the AMR-WB+ encoder. The low pass filter $h_{LP}(n)$ in the time domain is defined in AMR-WB+ by the following coefficients.

10 $a[13] = [0.088250, 0.086410, 0.081074, 0.072768, 0.062294, 0.050623, 0.038774, 0.027692, 0.018130, 0.010578, 0.005221, 0.001946, 0.000385];$

$h_{LP}(n)=a(13-n)$ for n from 1 to 12

$h_{LP}(n)=a(n-12)$ for n from 13 to 25

15

The impulse response and the frequency response illustrated in Fig. 3 are for a situation, when the filter is applied to a time-domain signal sample that 12.8 kHz. The generated delay is then a delay of 12 samples, i.e., 0.9375 ms.

20 The filter illustrated in Fig. 3 has a frequency response in the QMF domain, where each QMF has a resolution of 400 Hz. 32 QMF bands cover the bandwidth of the signal sample at 12.8 kHz. The frequency response and the QMF domain are illustrated in Fig. 4.

25 The amplitude frequency response with a resolution of 400 Hz forms the weights used when applying the low pass filter in the QMF domain. The weights for the weighter 108 are, for the above exemplary parameters as outlined in Fig. 5.

These weights can be calculated as follows:

30 $W=\text{abs}(\text{DFT}(h_{LP}(n), 64))$, where $\text{DFT}(x,N)$ stands for the Discrete Fourier Transform of length N of the signal x . If x is shorter than N , the signal is padded with N -size of x zeros. The length N of the DFT corresponds to two times the number of QMF sub-bands. Since $h_{LP}(n)$ is a signal of real coefficients, W shows a Hermitian symmetry and $N/2$ frequency coefficients between the frequency 0 and the Nyquist frequency.

35

By analysing the frequency response of the filter coefficients, it corresponds about to a cut-off frequency of $2\pi \cdot 10/256$. This is used for designing the filter. The coefficients were

then quantized for writing them on 14 bits for saving some ROM consumption and in view of a fixed point implementation.

5 The filtering in QMF domain is then performed as follows:

Y=post-processed signal in QMF domain

X= decoded signal in QMF signal from core-coder

E=inter-harmonic noise generated in TD to remove from X

10

$Y(k) = X(k) - W(k) \cdot E(k)$ for k from 1 to 32

Fig. 6 illustrates a further example, where the QMF has a resolution of 800 Hz, so that 16 bands cover the full bandwidth of the signal sampled at 12.8 kHz. The coefficients W are then as indicated in Fig. 6 below the plot. The filtering is done in the same way as discussed with respect to Fig. 6, but k only goes from 1 to 16.

The frequency response of the filter in the 16 bands QMF is plotted as illustrated in Fig. 6.

20 Fig. 10 illustrates a further enhancement of the long term prediction filter illustrated at 102 in Fig. 1b.

Particularly, for a low delay implementation, the term $\hat{s}(n+T)$ in the third to last line of Fig. 9 is problematic. This is due to the fact that the T samples are in the future with respect to the actual time n. Therefore, in order to address situations, where, due to the low delay implementation, the future values are not available yet, $\hat{s}(n+T)$ is replaced by $\hat{s}$ as indicated in Fig. 10. Then, the long term prediction filter approximates the long term prediction of the prior art, but with less or zero delay. It has been found that the approximation is good enough and that the gain with respect to the reduced delay is more advantageous than the slight loss in pitch enhancing.

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.

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 CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control
5 signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed.

Some embodiments according to the invention comprise a non-transitory data carrier having electronically readable control signals, which are capable of cooperating with a
10 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
15 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.
20

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.

25 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.

A further embodiment of the inventive method is, therefore, a data stream or a sequence of
30 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
35 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.

5 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.

10 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
15 herein.

Claims

1. Apparatus for processing a decoded audio signal (100), comprising:

5 a filter (102) for filtering the decoded audio signal to obtain a filtered audio signal (104);

10 a time-spectral converter stage (106) for converting the decoded audio signal and the filtered audio signal into corresponding spectral representations, each spectral representation having a plurality of subband signals;

15 a weighter (108) for performing a frequency selective weighting of the spectral representation of the filtered audio signal by multiplying subband signals by respective weighting coefficients to obtain a weighted filtered audio signal;

a subtractor (112) for performing a subband-wise subtraction between the weighted filtered audio signal and the spectral representation of the audio signal to obtain a result audio signal; and

20 a spectral-time converter (114) for converting the result audio signal or a signal derived from the result audio signal into a time domain representation to obtain a processed decoded audio signal (116).

25 2. Apparatus in accordance with claim 1, further comprising a bandwidth enhancement decoder (129) or a mono-stereo or a mono-multichannel decoder (131) to calculate the signal derived from the result audio signal,

30 wherein the spectral-time converter (114) is configured for not converting the result audio signal but the signal derived from the result audio signal into the time domain so that all processing by the bandwidth enhancement decoder (129) or the mono-stereo or mono-multichannel decoder (131) is performed in the same spectral domain as defined by the time-spectral converter stage (106).

35 3. Apparatus of claim 1 or 2,

wherein the decoded audio signal is an ACELP-decoded output signal, and

wherein the filter (102) is a long term prediction filter controlled by pitch information.

4. Apparatus in accordance with one of the preceding claims,

wherein the weighter (108) is configured for weighting the filtered audio signal so that lower frequency subbands are less attenuated or not attenuated than higher frequency subbands so that the frequency-selective weighting impresses a low pass characteristic to the filtered audio signal.

5. Apparatus in accordance with one of the preceding claims,

wherein the time-spectral converter stage (106) and the spectral-time converter (114) are configured to implement a QMF analysis filterbank and a QMF synthesis filterbank, respectively.

6. Apparatus in accordance with one the preceding claims,

wherein the subtractor (112) is configured for subtracting a subband signal of the weighted filtered audio signal from the corresponding subband signal of the audio signal to obtain a subband of the result audio signal, the subbands belonging to the same filterbank channel.

7. Apparatus in accordance with one of the preceding claims,

wherein the filter (102) is configured to perform a weighted combination of the audio signal and at least the audio signal shifted in time by a pitch period.

8. Apparatus of claim 7,

wherein the filter (102) is configured for performing the weighted combination by only combining the audio signal and the audio signal existing at earlier time instants.

9. Apparatus in accordance with one of the preceding claims,

wherein the spectral-time converter (114) has a different number of input channels with respect to the time-spectral converter stage (106) so that a sample-rate

conversion is obtained, wherein an upsampling is obtained, when the number of input channels into the spectral-time converter is higher than the number of output channels of the time-spectral converter stage and wherein a downsampling is performed, when the number of input channels into the spectral-time converter is smaller than the number of output channels from the time-spectral converter stage.

10. Apparatus in accordance with one of the preceding claims, further comprising:

a first decoder (120) for providing the decoded audio signal in a first time portion;

a second decoder (122) for providing a further decoded audio signal in a different second time portion;

a first processing branch connected to the first decoder (120) and the second decoder (122);

a second processing branch connected to the first decoder (120) and the second decoder (122),

wherein the second processing branch comprises the filter (102) and the weighter (108) and, additionally, comprises a controllable gain stage (129) and a controller (130), wherein the controller (130) is configured for setting a gain of the gain stage (129) to a first value for the first time portion and to a second value or to zero for the second time portion, which is lower than the first value.

11. Apparatus in accordance with one of the preceding claims, further comprising a pitch tracker for providing a pitch lag and for setting the filter (102) based on the pitch lag as the pitch information.

12. Apparatus in accordance with one of claims 10 or 11, wherein the first decoder (120) is configured for providing the pitch information or a part of the pitch information for setting the filter (102).

13. Apparatus in accordance with claim 10, 11 or 12, wherein an output of the first processing branch and an output of the second processing branch are connected to inputs of the subtractor (112).

14. Apparatus in accordance with one of the preceding claims, wherein the decoded audio signal is provided by an ACELP decoder (120) included in the apparatus, and
5 wherein the apparatus further comprises a further decoder (122) implemented as a TCX decoder.
15. Method of processing a decoded audio signal (100), comprising:
10 filtering (102) the decoded audio signal to obtain a filtered audio signal (104);
converting (106) the decoded audio signal and the filtered audio signal into
corresponding spectral representations, each spectral representation having a
plurality of subband signals;
15 performing (108) a frequency selective weighting of the filtered audio signal by
multiplying subband signals by respective weighting coefficients to obtain a
weighted filtered audio signal;
20 performing (112) a subband-wise subtraction between the weighted filtered audio
signal and the spectral representation of the audio signal to obtain a result audio
signal; and
converting (114) the result audio signal or a signal derived from the result audio
25 signal into a time domain representation to obtain a processed decoded audio signal
(116).
16. Computer program having a program code for performing, when running on a
computer, the method of processing a decoded audio signal in accordance with
claim 15.
30