(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 January 2015 (29.01.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/010950 A1

(51) International Patent Classification:
G10L 21/0388 (2013.01) *G10L 19/02* (2013.01)
G10L 19/028 (2013.01)

(21) International Application Number:
PCT/EP2014/065112

(22) International Filing Date:
15 July 2014 (15.07.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

13177346.7	22 July 2013 (22.07.2013)	EP
13177350.9	22 July 2013 (22.07.2013)	EP
13177353.3	22 July 2013 (22.07.2013)	EP
13177348.3	22 July 2013 (22.07.2013)	EP
13189389.3	18 October 2013 (18.10.2013)	EP

(71) Applicant: **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Hansastraße 27c, 80686 München (DE).

(72) Inventors: **DISCH, Sascha**; Wilhelmstrasse 70, 90766 Fürth (DE). **GEIGER, Ralf**; Jakob-Herz-Weg 36, 91052 Erlangen (DE). **HELMRICH, Christian**; Hauptstraße 68, 91054 Erlangen (DE). **NAGEL, Frederik**; Wilhelmshavener Strasse 72, 90425 Nürnberg (DE). **NEUKAM, Christian**; Weißgasse 24, 90562 Kalchreuth (DE). **SCHMIDT, Konstantin**; Heerwagenstrasse 21, 90489 Nürnberg (DE). **FISCHER, Michael**; Haagstr. 24, 91054 Erlangen (DE).

(74) Agents: **ZINKLER, Franz** et al.; P. O. Box 246, 82043 Pullach (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

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

(54) Title: APPARATUS AND METHOD FOR DECODING AN ENCODED AUDIO SIGNAL USING A CROSS-OVER FILTER AROUND A TRANSITION FREQUENCY

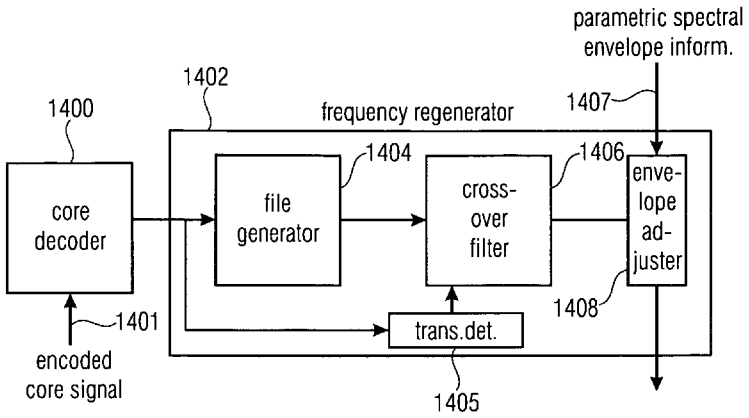


FIG 14A

(57) Abstract: Apparatus for decoding an encoded audio signal comprising an encoded core signal (1), comprising: a core decoder (1400) for decoding the encoded core signal (1401) to obtain a decoded core signal; a tile generator (1404) for generating one or more spectral tiles having frequencies not included in the decoded core signal using a spectral portion of the decoded core signal; and a cross-over filter (1406) for spectrally cross-over filtering the decoded core signal and a first frequency tile having frequencies extending from a gap filling frequency (309) to an upper border frequency or for spectrally cross-over filtering a first frequency tile and a second frequency tile.

WO 2015/010950 A1

Apparatus and method for decoding an encoded audio signal using a cross-over filter around a transition frequency

Specification

The present invention relates to audio coding/decoding and, particularly, to audio coding using Intelligent Gap Filling (IGF).

Audio coding is the domain of signal compression that deals with exploiting redundancy and irrelevancy in audio signals using psychoacoustic knowledge. Today audio codecs typically need around 60 kbps/channel for perceptually transparent coding of almost any type of audio signal. Newer codecs are aimed at reducing the coding bitrate by exploiting spectral similarities in the signal using techniques such as bandwidth extension (BWE). A BWE scheme uses a low bitrate parameter set to represent the high frequency (HF) components of an audio signal. The HF spectrum is filled up with spectral content from low frequency (LF) regions and the spectral shape, tilt and temporal continuity adjusted to maintain the timbre and color of the original signal. Such BWE methods enable audio codecs to retain good quality at even low bitrates of around 24 kbps/channel.

The inventive audio coding system efficiently codes arbitrary audio signals at a wide range of bitrates. Whereas, for high bitrates, the inventive system converges to transparency, for low bitrates perceptual annoyance is minimized. Therefore, the main share of available bitrate is used to waveform code just the perceptually most relevant structure of the signal in the encoder, and the resulting spectral gaps are filled in the decoder with signal content that roughly approximates the original spectrum. A very limited bit budget is consumed to control the parameter driven so-called spectral Intelligent Gap Filling (IGF) by dedicated side information transmitted from the encoder to the decoder.

Storage or transmission of audio signals is often subject to strict bitrate constraints. In the past, coders were forced to drastically reduce the transmitted audio bandwidth when only a very low bitrate was available.

Modern audio codecs are nowadays able to code wide-band signals by using bandwidth extension (BWE) methods [1]. These algorithms rely on a parametric representation of the high-frequency content (HF) - which is generated from the waveform coded low-frequency

part (LF) of the decoded signal by means of transposition into the HF spectral region ("patching") and application of a parameter driven post processing. In BWE schemes, the reconstruction of the HF spectral region above a given so-called cross-over frequency is often based on spectral patching. Typically, the HF region is composed of multiple adjacent patches and each of these patches is sourced from band-pass (BP) regions of the LF spectrum below the given cross-over frequency. State-of-the-art systems efficiently perform the patching within a filterbank representation, e.g. Quadrature Mirror Filterbank (QMF), by copying a set of adjacent subband coefficients from a source to the target region.

Another technique found in today's audio codecs that increases compression efficiency and thereby enables extended audio bandwidth at low bitrates is the parameter driven synthetic replacement of suitable parts of the audio spectra. For example, noise-like signal portions of the original audio signal can be replaced without substantial loss of subjective quality by artificial noise generated in the decoder and scaled by side information parameters. One example is the Perceptual Noise Substitution (PNS) tool contained in MPEG-4 Advanced Audio Coding (AAC) [5].

A further provision that also enables extended audio bandwidth at low bitrates is the noise filling technique contained in MPEG-D Unified Speech and Audio Coding (USAC) [7]. Spectral gaps (zeroes) that are inferred by the dead-zone of the quantizer due to a too coarse quantization, are subsequently filled with artificial noise in the decoder and scaled by a parameter-driven post-processing.

Another state-of-the-art system is termed Accurate Spectral Replacement (ASR) [2-4]. In addition to a waveform codec, ASR employs a dedicated signal synthesis stage which restores perceptually important sinusoidal portions of the signal at the decoder. Also, a system described in [5] relies on sinusoidal modeling in the HF region of a waveform coder to enable extended audio bandwidth having decent perceptual quality at low bitrates. All these methods involve transformation of the data into a second domain apart from the Modified Discrete Cosine Transform (MDCT) and also fairly complex analysis/synthesis stages for the preservation of HF sinusoidal components.

Fig. 13a illustrates a schematic diagram of an audio encoder for a bandwidth extension technology as, for example, used in High Efficiency Advanced Audio Coding (HE-AAC). An audio signal at line 1300 is input into a filter system comprising of a low pass 1302 and

a high pass 1304. The signal output by the high pass filter 1304 is input into a parameter extractor/coder 1306. The parameter extractor/coder 1306 is configured for calculating and coding parameters such as a spectral envelope parameter, a noise addition parameter, a missing harmonics parameter, or an inverse filtering parameter, for example. These extracted parameters are input into a bit stream multiplexer 1308. The low pass output signal is input into a processor typically comprising the functionality of a down sampler 1310 and a core coder 1312. The low pass 1302 restricts the bandwidth to be encoded to a significantly smaller bandwidth than occurring in the original input audio signal on line 1300. This provides a significant coding gain due to the fact that the whole functionalities occurring in the core coder only have to operate on a signal with a reduced bandwidth. When, for example, the bandwidth of the audio signal on line 1300 is 20 kHz and when the low pass filter 1302 exemplarily has a bandwidth of 4 kHz, in order to fulfill the sampling theorem, it is theoretically sufficient that the signal subsequent to the down sampler has a sampling frequency of 8 kHz, which is a substantial reduction to the sampling rate required for the audio signal 1300 which has to be at least 40 kHz.

Fig. 13b illustrates a schematic diagram of a corresponding bandwidth extension decoder. The decoder comprises a bitstream multiplexer 1320. The bitstream demultiplexer 1320 extracts an input signal for a core decoder 1322 and an input signal for a parameter decoder 1324. A core decoder output signal has, in the above example, a sampling rate of 8 kHz and, therefore, a bandwidth of 4 kHz while, for a complete bandwidth reconstruction, the output signal of a high frequency reconstructor 1330 must be at 20 kHz requiring a sampling rate of at least 40 kHz. In order to make this possible, a decoder processor having the functionality of an upsampler 1325 and a filterbank 1326 is required. The high frequency reconstructor 1330 then receives the frequency-analyzed low frequency signal output by the filterbank 1326 and reconstructs the frequency range defined by the high pass filter 1304 of Fig. 13a using the parametric representation of the high frequency band. The high frequency reconstructor 1330 has several functionalities such as the regeneration of the upper frequency range using the source range in the low frequency range, a spectral envelope adjustment, a noise addition functionality and a functionality to introduce missing harmonics in the upper frequency range and, if applied and calculated in the encoder of Fig. 13a, an inverse filtering operation in order to account for the fact that the higher frequency range is typically not as tonal as the lower frequency range. In HE-AAC, missing harmonics are re-synthesized on the decoder-side and are placed exactly in the middle of a reconstruction band. Hence, all missing harmonic lines that have been determined in a certain reconstruction band are not placed at the

frequency values where they were located in the original signal. Instead, those missing harmonic lines are placed at frequencies in the center of the certain band. Thus, when a missing harmonic line in the original signal was placed very close to the reconstruction band border in the original signal, the error in frequency introduced by placing this missing harmonics line in the reconstructed signal at the center of the band is close to 50% of the individual reconstruction band, for which parameters have been generated and transmitted.

Furthermore, even though the typical audio core coders operate in the spectral domain, the core decoder nevertheless generates a time domain signal which is then, again, converted into a spectral domain by the filter bank 1326 functionality. This introduces additional processing delays, may introduce artifacts due to tandem processing of firstly transforming from the spectral domain into the frequency domain and again transforming into typically a different frequency domain and, of course, this also requires a substantial amount of computation complexity and thereby electric power, which is specifically an issue when the bandwidth extension technology is applied in mobile devices such as mobile phones, tablet or laptop computers, etc.

Current audio codecs perform low bitrate audio coding using BWE as an integral part of the coding scheme. However, BWE techniques are restricted to replace high frequency (HF) content only. Furthermore, they do not allow perceptually important content above a given cross-over frequency to be waveform coded. Therefore, contemporary audio codecs either lose HF detail or timbre when the BWE is implemented, since the exact alignment of the tonal harmonics of the signal is not taken into consideration in most of the systems.

Another shortcoming of the current state of the art BWE systems is the need for transformation of the audio signal into a new domain for implementation of the BWE (e.g. transform from MDCT to QMF domain). This leads to complications of synchronization, additional computational complexity and increased memory requirements.

Storage or transmission of audio signals is often subject to strict bitrate constraints. In the past, coders were forced to drastically reduce the transmitted audio bandwidth when only a very low bitrate was available. Modern audio codecs are nowadays able to code wide-band signals by using bandwidth extension (BWE) methods [1-2]. These algorithms rely on a parametric representation of the high-frequency content (HF) - which is generated from the waveform coded low-frequency part (LF) of the decoded signal by means of

transposition into the HF spectral region ("patching") and application of a parameter driven post processing.

In BWE schemes, the reconstruction of the HF spectral region above a given so-called cross-over frequency is often based on spectral patching. Other schemes that are functional to fill spectral gaps, e.g. Intelligent Gap Filling (IGF), use neighboring so-called spectral tiles to regenerate parts of audio signal HF spectra. Typically, the HF region is composed of multiple adjacent patches or tiles and each of these patches or tiles is sourced from band-pass (BP) regions of the LF spectrum below the given cross-over frequency. State-of-the-art systems efficiently perform the patching or tiling within a filterbank representation by copying a set of adjacent subband coefficients from a source to the target region. Yet, for some signal content, the assemblage of the reconstructed signal from the LF band and adjacent patches within the HF band can lead to beating, dissonance and auditory roughness.

Therefore, in [19], the concept of dissonance guard-band filtering is presented in the context of a filterbank-based BWE system. It is suggested to effectively apply a notch filter of approx. 1 Bark bandwidth at the cross-over frequency between LF and BWE-regenerated HF to avoid the possibility of dissonance and replace the spectral content with zeros or noise.

However, the proposed solution in [19] has some drawbacks: First, the strict replacement of spectral content by either zeros or noise can also impair the perceptual quality of the signal. Moreover, the proposed processing is not signal adaptive and can therefore harm perceptual quality in some cases. For example, if the signal contains transients, this can lead to pre- and post-echoes.

Second, dissonances can also occur at transitions between consecutive HF patches. The proposed solution in [19] is only functional to remedy dissonances that occur at cross-over frequency between LF and BWE-regenerated HF.

Last, as opposed to filter bank based systems like proposed in [19], BWE systems can also be realized in transform based implementations, like e.g. the Modified Discrete Cosine Transform (MDCT). Transforms like MDCT are very prone to so-called warbling [20] or ringing artifacts that occur if bandpass regions of spectral coefficients are copied or spectral coefficients are set to zero like proposed in [19].

Particularly, US Patent 8,412,365 discloses to use, in filterbank based translation or folding, so-called guard-bands which are inserted and made of one or several subband channels set to zero. A number of filterbank channels is used as guard-bands, and a bandwidth of a guard-band should be 0,5 Bark. These dissonance guard-bands are partially reconstructed using random white noise signals, i.e., the subbands are fed with white noise instead of being zero. The guard bands are inserted irrespective of the current signal to processed.

Bandwidth extension systems are particularly problematic when they are realized in transform-based implementations like, for example, the Modified Discrete Cosine Transform (MDCT). Transforms like MDCT and other transforms as well are very prone to so-called warbling as discussed in [3] and ringing artifacts that occur if bandpass regions of spectral coefficients are copied or spectral coefficients are set to zero like proposed in [2].

It is the object of the present invention to provide an improved apparatus and method for decoding an encoded audio signal.

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

In accordance with the present invention, an apparatus for decoding an encoded audio signal comprises a core decoder, a tile generator for generating one or more spectral tiles having frequencies not included in the decoded core signal using a spectral portion of the decoded core signal and a cross-over filter for spectrally cross-over filtering the decoded core signal and a first frequency tile having frequencies extending from a gap filling frequency to a first tile stop frequency or for spectrally cross-over filtering a tile and a further frequency tile, the further frequency tile having a lower border frequency being frequency-adjacent to an upper border frequency of the frequency tile.

Preferably, this procedure is intended to be applied within a bandwidth extension based on a transform like the MDCT. However, the present invention is generally applicable and, particularly in a bandwidth extension scenario relying on a quadrature mirror filterbank (QMF), particularly if the system is critically sampled, for example when there is a real-

valued QMF representation as a time-frequency conversion or as a frequency-time conversion.

The present invention is particularly useful for transient-like signals, since for such transient-like signals, ringing is an audible and annoying artifact. Filter ringing artifacts are caused by the so-called brick-wall characteristic of a filter in the transition band, i.e., a steep transition from a pass band to a stop band at a cut-off frequency. Such filters can be efficiently implemented by setting one coefficient or groups of coefficients to zero in a frequency domain of a time- frequency transform. Therefore, the present invention relies on a cross-over filter at each transition frequency between patches/tiles or between a core band and a first patch/tile to reduce this ringing artifact. The cross-over filter is preferably implemented by spectral weighting in the transform domain employing suitable gain functions.

Preferably, the cross-over filter is signal-adaptive and consists of two filters, a fade-out filter, which is applied to the lower spectral region and a fade-in filter, which is applied to the higher spectral region. The filters can be symmetric or asymmetric depending on the specific implementation.

In a further embodiment, a frequency tile or frequency patch is not only subjected to cross-over filtering, but the tile generator preferably performs, before performing the cross-over filtering, a patch adaption comprising a setting of frequency borders at local spectral minima and a removal or attenuation of tonal portions remaining in transition ranges around the transition frequencies.

In this embodiment, a decoder-side signal analysis using an analyzer is performed for analyzing the decoded core signal before or after performing a frequency regeneration operation to provide an analysis result. Then, this analysis result is used by a frequency regenerator for regenerating spectral portions not included in the decoded core signal.

Thus, in contrast to a fixed decoder-setting, where the patching or frequency tiling is performed in a fixed way, i.e., where a certain source range is taken from the core signal and certain fixed frequency borders are applied to either set the frequency between the source range and the reconstruction range or the frequency border between two adjacent frequency patches or tiles within the reconstruction range, a signal-dependent patching or tiling is performed, in which, for example, the core signal can be analyzed to find local

minima in the core signal and, then, the core range is selected so that the frequency borders of the core range coincide with local minima in the core signal spectrum.

Alternatively or additionally, a signal analysis can be performed on a preliminary regenerated signal or preliminary frequency-patched or tiled signal, wherein, after the preliminary frequency regeneration procedure, the border between the core range and the reconstruction range is analyzed in order to detect any artifact-creating signal portions such as tonal portions being problematic in that they are quite close to each other to generate a beating artifact when being reconstructed. Alternatively or additionally, the borders can also be examined in such a way that a halfway-clipping of a tonal portion is detected and this clipping of a tonal portion would also create an artifact when being reconstructed as it is. In order to avoid these procedures, the frequency border of the reconstruction range and/or the source range and/or between two individual frequency tiles or patches in the reconstruction range can be modified by a signal manipulator in order to again perform a reconstruction with the newly set borders.

Additionally, or alternatively, the frequency regeneration is a regeneration based on the analysis result in that the frequency borders are left as they are and an elimination or at least attenuation of problematic tonal portions near the frequency borders between the source range and the reconstruction range or between two individual frequency tiles or patches within the reconstruction range is done. Such tonal portions can be close tones that would result in a beating artifact or could be clipped tonal portions.

Specifically, when a non-energy conserving transform is used such as an MDCT, a single tone does not directly map to a single spectral line. Instead, a single tone will map to a group of spectral lines with certain amplitudes depending on the phase of the tone. When a patching operation clips this tonal portion, then this will result in an artifact after reconstruction even though a perfect reconstruction is applied as in an MDCT reconstructor. This is due to the fact that the MDCT reconstructor would require the complete tonal pattern for a tone in order to finally correctly reconstruct this tone. Due to the fact that a clipping has taken place before, this is not possible anymore and, therefore, a time varying warbling artifact will be created. Based on the analysis in accordance with the present invention, the frequency regenerator will avoid this situation by attenuating the complete tonal portion creating an artifact or as discussed before, by changing corresponding border frequencies or by applying both measures or by even reconstructing the clipped portion based on a certain pre-knowledge on such tonal patterns.

The inventive approach is mainly intended to be applied within a BWE based on a transform like the MDCT. Nevertheless, the teachings of the invention are generally applicable, e.g. analogously within a Quadrature Mirror Filter bank (QMF) based system, especially if the system is critically sampled, e.g. a real-valued QMF representation.

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

Fig. 1a illustrates an apparatus for encoding an audio signal;

Fig. 1b illustrates a decoder for decoding an encoded audio signal matching with the encoder of Fig. 1a;

Fig. 2a illustrates a preferred implementation of the decoder;

Fig. 2b illustrates a preferred implementation of the encoder;

Fig. 3a illustrates a schematic representation of a spectrum as generated by the spectral domain decoder of Fig. 1b;

Fig. 3b illustrates a table indicating the relation between scale factors for scale factor bands and energies for reconstruction bands and noise filling information for a noise filling band;

Fig. 4a illustrates the functionality of the spectral domain encoder for applying the selection of spectral portions into the first and second sets of spectral portions;

Fig. 4b illustrates an implementation of the functionality of Fig. 4a;

Fig. 5a illustrates a functionality of an MDCT encoder;

Fig. 5b illustrates a functionality of the decoder with an MDCT technology;

Fig. 5c illustrates an implementation of the frequency regenerator;

Fig. 6a is an apparatus for decoding an encoded audio signal in accordance with one implementation;

Fig. 6b a further embodiment of an apparatus for decoding an encoded audio signal;

Fig. 7a illustrates a preferred implementation of the frequency regenerator of Fig. 6a or 6b;

Fig. 7b illustrates a further implementation of a cooperation between the analyzer and the frequency regenerator;

Fig. 8 illustrates a further implementation of the frequency regenerator;

Fig. 8b illustrates a further embodiment of the invention;

Fig. 9a illustrates a decoder with frequency regeneration technology using energy values for the regeneration frequency range;

Fig. 9b illustrates a more detailed implementation of the frequency regenerator of Fig. 9a;

Fig. 9c illustrates a schematic illustrating the functionality of Fig. 9b;

Fig. 9d illustrates a further implementation of the decoder of Fig. 9a;

Fig. 10a illustrates a block diagram of an encoder matching with the decoder of Fig. 9a;

Fig. 10b illustrates a block diagram for illustrating a further functionality of the parameter calculator of Fig. 10a;

Fig. 10c illustrates a block diagram illustrating a further functionality of the parametric calculator of Fig. 10a;

Fig. 10d illustrates a block diagram illustrating a further functionality of the parametric calculator of Fig. 10a;

Fig. 11a illustrates a spectrum of a filter ringing surrounding a transient;

Fig. 11b illustrates a spectrogram of a transient after applying bandwidth extension;

Fig. 11c illustrates a spectrogram of a transient after applying bandwidth extension with filter ringing reduction;

Fig. 12a illustrates a block diagram of an apparatus for decoding an encoded audio signal;

Fig. 12b illustrates magnitude spectra (stylized) of a tonal signal, a copy-up without patch/tile adaption, a copy-up with changed frequency borders and an additional elimination of artifact-creating tonal portions;

Fig. 12c illustrates an example cross-fade function;

Fig. 13a illustrates a prior art encoder with bandwidth extension; and

Fig. 13b illustrates a prior art decoder with bandwidth extension.

Fig. 14a illustrates a further apparatus for decoding an encoded audio signal using a cross-over filter;

Fig. 14b illustrates a more detailed illustration of an exemplary cross-over filter;

Fig. 6a illustrates an apparatus for decoding an encoded audio signal comprising an encoded core signal and parametric data. The apparatus comprises a core decoder 600 for decoding the encoded core signal to obtain a decoded core signal, an analyzer 602 for analyzing the decoded core signal before or after performing a frequency regeneration operation. The analyzer 602 is configured for providing an analysis result 603. The frequency regenerator 604 is configured for regenerating spectral portions not included in the decoded core signal using a spectral portion of the decoded core signal, envelope data 605 for the missing spectral portions and the analysis result 603. Thus, in contrast to

earlier implementations, the frequency regeneration is not performed on the decoder-side signal-independent, but is performed signal-dependent. This has the advantage that, when no problems exist, the frequency regeneration is performed as it is, but when problematic signal portions exist, then this is detected by the analysis result 603 and the frequency regenerator 604 then performs an adapted way of frequency regeneration which can, for example, be the change of an initial frequency border between the core region and the reconstruction band or the change of a frequency border between two individual tiles/patches within the reconstruction band. Contrary to the implementation of the guard-bands, this has the advantage that specific procedures are only performed when required and not, as in the guard-band implementation, all the time without any signal-dependency.

Preferably, the core decoder 600 is implemented as an entropy (e.g. Huffman or arithmetic decoder) decoding and dequantizing stage 612 as illustrated in Fig. 6b. The core decoder 600 then outputs a core signal spectrum and the spectrum is analyzed by the spectral analyzer 614 which is, quite similar to the analyzer 602 in Fig. 6a. implemented as a spectral analyzer rather than any arbitrary analyzer which could, as illustrated in Fig. 6a, also analyze a time domain signal. In the embodiment of Fig. 6b, the spectral analyzer is configured for analyzing the spectral signal so that local minima in the source band and/or in a target band, i.e., in the frequency patches or frequency tiles are determined. Then, the frequency regenerator 604 performs, as illustrated at 616, a frequency regeneration where the patch borders are placed to minima in the source band and/or the target band.

Subsequently, Fig. 7a is discussed in order to describe a preferred implementation of the frequency regenerator 604 of Fig. 6a. A preliminary signal regenerator 702 receives, as an input, source data from the source band and, additionally, preliminary patch information such as preliminary border frequencies. Then, a preliminary regenerated signal 703 is generated, which is detected by the detector 704 for detecting the tonal components within the preliminary reconstructed signal 703. Alternatively or additionally, the source data 705 can also be analyzed by the detector corresponding to the analyzer 602 of Fig. 6a. Then, the preliminary signal regeneration step would not be necessary. When there is a well-defined mapping from the source data to the reconstruction data, then the minima or tonal portions can be detected even by considering only the source data, whether there are tonal portions close to the upper border of the core range or at a frequency border

between two individually generated frequency tiles as will be discussed later with respect to Fig. 12b.

In case problematic tonal components have been discovered near frequency borders, a transition frequency adjuster 706 performs an adjustment of a transition frequency such as a transition frequency or cross-over frequency or gap filling start frequency between the core band and the reconstruction band or between individual frequency portions generated by one and the same source data in the reconstruction band. The output signal of block 706 is forwarded to a remover 708 of tonal components at borders. The remover is configured for removing remaining tonal components which are still there subsequent to the transition frequency adjustment by block 706. The result of the remover 708 is then forwarded to a cross-over filter 710 in order to address the filter ringing problem and the result of the cross-over filter 710 is then input into a spectral envelope shaping block 712 which performs a spectral envelope shaping in the reconstruction band.

As discussed in the context of Fig. 7a, the detection of tonal components in block 704 can be both performed on a source data 705 or a preliminary reconstructed signal 703. This embodiment is illustrated in Fig. 7b, where a preliminary regenerated signal is created as shown in block 718. The signal corresponding to signal 703 of Fig. 7a is then forwarded to a detector 720 which detects artifact-creating components. Although the detector 720 can be configured for being a detector for detecting tonal components at frequency borders as illustrated at 704 in Fig. 7a, the detector can also be implemented to detect other artifact-creating components. Such spectral components can be even other components than tonal components and a detection whether an artifact has been created can be performed by trying different regenerations and comparing the different regeneration results in order to find out which one has provided artifact-creating components.

The detector 720 now controls a manipulator 722 for manipulating the signal, i.e., the preliminary regenerated signal. This manipulation can be done by actually processing the preliminary regenerated signal by line 723 or by newly performing a regeneration, but now with, for example, the amended transition frequencies as illustrated by line 724.

One implementation of the manipulation procedure is that the transition frequency is adjusted as illustrated at 706 in Fig. 7a. A further implementation is illustrated in Fig. 8a, which can be performed instead of block 706 or together with block 706 of Fig. 7a. A detector 802 is provided for detecting start and end frequencies of a problematic tonal

portion. Then, an interpolator 804 is configured for interpolating and, preferably complex interpolating between the start and the end of the tonal portion within the spectral range. Then, as illustrated in Fig. 8a by block 806, the tonal portion is replaced by the interpolation result.

An alternative implementation is illustrated in Fig. 8a by blocks 808, 810. Instead of performing an interpolation, a random generation of spectral lines 808 is performed between the start and the end of the tonal portion. Then, an energy adjustment of the randomly generated spectral lines is performed as illustrated at 810, and the energy of the randomly generated spectral lines is set so that the energy is similar to the adjacent non-tonal spectral parts. Then, the tonal portion is replaced by envelope-adjusted randomly generated spectral lines. The spectral lines can be randomly generated or pseudo randomly generated in order to provide a replacement signal which is, as far as possible, artifact-free.

A further implementation is illustrated in Fig. 8b. A frequency tile generator located within the frequency regenerator 604 of Fig. 6a is illustrated at block 820. The frequency tile generator uses predetermined frequency borders. Then, the analyzer analyzes the signal generated by the frequency tile generator, and the frequency tile generator 820 is preferably configured for performing multiple tiling operations to generate multiple frequency tiles. Then, the manipulator 824 in Fig. 8b manipulates the result of the frequency tile generator in accordance with the analysis result output by the analyzer 822. The manipulation can be the change of frequency borders or the attenuation of individual portions. Then, a spectral envelope adjuster 826 performs a spectral envelope adjustment using the parametric information 605 as already discussed in the context of Fig. 6a.

Then, the spectrally adjusted signal output by block 826 is input into a frequency-time converter which, additionally, receives the first spectral portions, i.e., a spectral representation of the output signal of the core decoder 600. The output of the frequency-time converter 828 can then be used for storage or for transmitting to a loudspeaker for audio rendering.

The present invention can be applied either to known frequency regeneration procedures such as illustrated in Figs. 13a, 13b or can preferably be applied within the intelligent gap filling context, which is subsequently described with respect to Figs. 1a to 5b and 9a to 10d.

Fig. 1a illustrates an apparatus for encoding an audio signal 99. The audio signal 99 is input into a time spectrum converter 100 for converting an audio signal having a sampling rate into a spectral representation 101 output by the time spectrum converter. The spectrum 101 is input into a spectral analyzer 102 for analyzing the spectral representation 101. The spectral analyzer 101 is configured for determining a first set of first spectral portions 103 to be encoded with a first spectral resolution and a different second set of second spectral portions 105 to be encoded with a second spectral resolution. The second spectral resolution is smaller than the first spectral resolution. The second set of second spectral portions 105 is input into a parameter calculator or parametric coder 104 for calculating spectral envelope information having the second spectral resolution. Furthermore, a spectral domain audio coder 106 is provided for generating a first encoded representation 107 of the first set of first spectral portions having the first spectral resolution. Furthermore, the parameter calculator/parametric coder 104 is configured for generating a second encoded representation 109 of the second set of second spectral portions. The first encoded representation 107 and the second encoded representation 109 are input into a bit stream multiplexer or bit stream former 108 and block 108 finally outputs the encoded audio signal for transmission or storage on a storage device.

Typically, a first spectral portion such as 306 of Fig. 3a will be surrounded by two second spectral portions such as 307a, 307b. This is not the case in HE AAC, where the core coder frequency range is band limited

Fig. 1b illustrates a decoder matching with the encoder of Fig. 1a. The first encoded representation 107 is input into a spectral domain audio decoder 112 for generating a first decoded representation of a first set of first spectral portions, the decoded representation having a first spectral resolution. Furthermore, the second encoded representation 109 is input into a parametric decoder 114 for generating a second decoded representation of a second set of second spectral portions having a second spectral resolution being lower than the first spectral resolution.

The decoder further comprises a frequency regenerator 116 for regenerating a reconstructed second spectral portion having the first spectral resolution using a first spectral portion. The frequency regenerator 116 performs a tile filling operation, i.e., uses a tile or portion of the first set of first spectral portions and copies this first set of first

spectral portions into the reconstruction range or reconstruction band having the second spectral portion and typically performs spectral envelope shaping or another operation as indicated by the decoded second representation output by the parametric decoder 114, i.e., by using the information on the second set of second spectral portions. The decoded first set of first spectral portions and the reconstructed second set of spectral portions as indicated at the output of the frequency regenerator 116 on line 117 is input into a spectrum-time converter 118 configured for converting the first decoded representation and the reconstructed second spectral portion into a time representation 119, the time representation having a certain high sampling rate.

Fig. 2b illustrates an implementation of the Fig. 1a encoder. An audio input signal 99 is input into an analysis filterbank 220 corresponding to the time spectrum converter 100 of Fig. 1a. Then, a temporal noise shaping operation is performed in TNS block 222. Therefore, the input into the spectral analyzer 102 of Fig. 1a corresponding to a block tonal mask 226 of Fig. 2b can either be full spectral values, when the temporal noise shaping/ temporal tile shaping operation is not applied or can be spectral residual values, when the TNS operation as illustrated in Fig. 2b, block 222 is applied. For two-channel signals or multi-channel signals, a joint channel coding 228 can additionally be performed, so that the spectral domain encoder 106 of Fig. 1a may comprise the joint channel coding block 228. Furthermore, an entropy coder 232 for performing a lossless data compression is provided which is also a portion of the spectral domain encoder 106 of Fig. 1a.

The spectral analyzer/tonal mask 226 separates the output of TNS block 222 into the core band and the tonal components corresponding to the first set of first spectral portions 103 and the residual components corresponding to the second set of second spectral portions 105 of Fig. 1a. The block 224 indicated as IGF parameter extraction encoding corresponds to the parametric coder 104 of Fig. 1a and the bitstream multiplexer 230 corresponds to the bitstream multiplexer 108 of Fig. 1a.

Preferably, the analysis filterbank 222 is implemented as an MDCT (modified discrete cosine transform filterbank) and the MDCT is used to transform the signal 99 into a time-frequency domain with the modified discrete cosine transform acting as the frequency analysis tool.

The spectral analyzer 226 preferably applies a tonality mask. This tonality mask estimation stage is used to separate tonal components from the noise-like components in

the signal. This allows the core coder 228 to code all tonal components with a psycho-acoustic module. The tonality mask estimation stage can be implemented in numerous different ways and is preferably implemented similar in its functionality to the sinusoidal track estimation stage used in sine and noise-modeling for speech/audio coding [8, 9] or an HILN model based audio coder described in [10]. Preferably, an implementation is used which is easy to implement without the need to maintain birth-death trajectories, but any other tonality or noise detector can be used as well.

The IGF module calculates the similarity that exists between a source region and a target region. The target region will be represented by the spectrum from the source region. The measure of similarity between the source and target regions is done using a cross-correlation approach. The target region is split into $nTar$ non-overlapping frequency tiles. For every tile in the target region, $nSrc$ source tiles are created from a fixed start frequency. These source tiles overlap by a factor between 0 and 1, where 0 means 0% overlap and 1 means 100% overlap. Each of these source tiles is correlated with the target tile at various lags to find the source tile that best matches the target tile. The best matching tile number is stored in $tileNum[idx\_tar]$, the lag at which it best correlates with the target is stored in $xcorr\_lag[idx\_tar][idx\_src]$ and the sign of the correlation is stored in $xcorr\_sign[idx\_tar][idx\_src]$. In case the correlation is highly negative, the source tile needs to be multiplied by -1 before the tile filling process at the decoder. The IGF module also takes care of not overwriting the tonal components in the spectrum since the tonal components are preserved using the tonality mask. A band-wise energy parameter is used to store the energy of the target region enabling us to reconstruct the spectrum accurately.

This method has certain advantages over the classical SBR [1] in that the harmonic grid of a multi-tone signal is preserved by the core coder while only the gaps between the sinusoids is filled with the best matching "shaped noise" from the source region. Another advantage of this system compared to ASR (Accurate Spectral Replacement) [2-4] is the absence of a signal synthesis stage which creates the important portions of the signal at the decoder. Instead, this task is taken over by the core coder, enabling the preservation of important components of the spectrum. Another advantage of the proposed system is the continuous scalability that the features offer. Just using $tileNum[idx\_tar]$ and $xcorr\_lag = 0$, for every tile is called gross granularity matching and can be used for low bitrates while using variable $xcorr\_lag$ for every tile enables us to match the target and source spectra better.

In addition, a tile choice stabilization technique is proposed which removes frequency domain artifacts such as trilling and musical noise.

In case of stereo channel pairs an additional joint stereo processing is applied. This is necessary, because for a certain destination range the signal can a highly correlated panned sound source. In case the source regions chosen for this particular region are not well correlated, although the energies are matched for the destination regions, the spatial image can suffer due to the uncorrelated source regions. The encoder analyses each destination region energy band, typically performing a cross-correlation of the spectral values and if a certain threshold is exceeded, sets a joint flag for this energy band. In the decoder the left and right channel energy bands are treated individually if this joint stereo flag is not set. In case the joint stereo flag is set, both the energies and the patching are performed in the joint stereo domain. The joint stereo information for the IGF regions is signaled similar the joint stereo information for the core coding, including a flag indicating in case of prediction if the direction of the prediction is from downmix to residual or vice versa.

The energies can be calculated from the transmitted energies in the L/R-domain.

$$midNrg[k] = leftNrg[k] + rightNrg[k];$$
$$sideNrg[k] = leftNrg[k] - rightNrg[k];$$

with k being the frequency index in the transform domain.

Another solution is to calculate and transmit the energies directly in the joint stereo domain for bands where joint stereo is active, so no additional energy transformation is needed at the decoder side.

The source tiles are always created according to the Mid/Side-Matrix:

$$midTile[k] = 0.5 \cdot (leftTile[k] + rightTile[k])$$

$$sideTile[k] = 0.5 \cdot (leftTile[k] - rightTile[k])$$

Energy adjustment:

$$midTile[k] = midTile[k] * midNrg[k];$$

$$sideTile[k] = sideTile[k] * sideNrg[k];$$

Joint stereo -> LR transformation:

If no additional prediction parameter is coded:

$$leftTile[k] = midTile[k] + sideTile[k]$$

$$rightTile[k] = midTile[k] - sideTile[k]$$

If an additional prediction parameter is coded and if the signalled direction is from mid to side:

$$sideTile[k] = sideTile[k] - predictionCoeff \cdot midTile[k]$$
$$leftTile[k] = midTile[k] + sideTile[k]$$
$$rightTile[k] = midTile[k] - sideTile[k]$$

If the signalled direction is from side to mid:

$$midTile1[k] = midTile[k] - predictionCoeff \cdot sideTile[k]$$
$$leftTile[k] = midTile1[k] - sideTile[k]$$
$$rightTile[k] = midTile1[k] + sideTile[k]$$

This processing ensures that from the tiles used for regenerating highly correlated destination regions and panned destination regions, the resulting left and right channels still represent a correlated and panned sound source even if the source regions are not correlated, preserving the stereo image for such regions.

In other words, in the bitstream, joint stereo flags are transmitted that indicate whether L/R or M/S as an example for the general joint stereo coding shall be used. In the decoder, first, the core signal is decoded as indicated by the joint stereo flags for the core bands. Second, the core signal is stored in both L/R and M/S representation. For the IGF tile filling, the source tile representation is chosen to fit the target tile representation as indicated by the joint stereo information for the IGF bands.

Temporal Noise Shaping (TNS) is a standard technique and part of AAC [11 – 13]. TNS can be considered as an extension of the basic scheme of a perceptual coder, inserting an optional processing step between the filterbank and the quantization stage. The main task of the TNS module is to hide the produced quantization noise in the temporal masking region of transient like signals and thus it leads to a more efficient coding scheme. First, TNS calculates a set of prediction coefficients using "forward prediction" in the transform domain, e.g. MDCT. These coefficients are then used for flattening the temporal envelope of the signal. As the quantization affects the TNS filtered spectrum, also the quantization noise is temporarily flat. By applying the invers TNS filtering on decoder side, the quantization noise is shaped according to the temporal envelope of the TNS filter and therefore the quantization noise gets masked by the transient.

IGF is based on an MDCT representation. For efficient coding, preferably long blocks of approx. 20 ms have to be used. If the signal within such a long block contains transients, audible pre- and post-echoes occur in the IGF spectral bands due to the tile filling. Fig. 7c shows a typical pre-echo effect before the transient onset due to IGF. On the left side, the spectrogram of the original signal is shown and on the right side the spectrogram of the bandwidth extended signal without TNS filtering is shown.

This pre-echo effect is reduced by using TNS in the IGF context. Here, TNS is used as a temporal tile shaping (TTS) tool as the spectral regeneration in the decoder is performed on the TNS residual signal. The required TTS prediction coefficients are calculated and applied using the full spectrum on encoder side as usual. The TNS/TTS start and stop frequencies are not affected by the IGF start frequency $f_{IGFstart}$ of the IGF tool. In comparison to the legacy TNS, the TTS stop frequency is increased to the stop frequency of the IGF tool, which is higher than $f_{IGFstart}$. On decoder side the TNS/TTS coefficients are applied on the full spectrum again, i.e. the core spectrum plus the regenerated spectrum plus the tonal components from the tonality map (see Fig. 7e). The application of TTS is necessary to form the temporal envelope of the regenerated spectrum to match the envelope of the original signal again. So the shown pre-echoes are reduced. In addition, it still shapes the quantization noise in the signal below $f_{IGFstart}$ as usual with TNS.

In legacy decoders, spectral patching on an audio signal corrupts spectral correlation at the patch borders and thereby impairs the temporal envelope of the audio signal by introducing dispersion. Hence, another benefit of performing the IGF tile filling on the residual signal is that, after application of the shaping filter, tile borders are seamlessly correlated, resulting in a more faithful temporal reproduction of the signal.

In an inventive encoder, the spectrum having undergone TNS/TTS filtering, tonality mask processing and IGF parameter estimation is devoid of any signal above the IGF start frequency except for tonal components. This sparse spectrum is now coded by the core coder using principles of arithmetic coding and predictive coding. These coded components along with the signaling bits form the bitstream of the audio.

Fig. 2a illustrates the corresponding decoder implementation. The bitstream in Fig. 2a corresponding to the encoded audio signal is input into the demultiplexer/decoder which would be connected, with respect to Fig. 1b, to the blocks 112 and 114. The bitstream demultiplexer separates the input audio signal into the first encoded representation 107 of Fig. 1b and the second encoded representation 109 of Fig. 1b. The first encoded representation having the first set of first spectral portions is input into the joint channel decoding block 204 corresponding to the spectral domain decoder 112 of Fig. 1b. The second encoded representation is input into the parametric decoder 114 not illustrated in Fig. 2a and then input into the IGF block 202 corresponding to the frequency regenerator 116 of Fig. 1b. The first set of first spectral portions required for frequency regeneration are input into IGF block 202 via line 203. Furthermore, subsequent to joint channel decoding 204 the specific core decoding is applied in the tonal mask block 206 so that the output of tonal mask 206 corresponds to the output of the spectral domain decoder 112. Then, a combination by combiner 208 is performed, i.e., a frame building where the output of combiner 208 now has the full range spectrum, but still in the TNS/TTS filtered domain. Then, in block 210, an inverse TNS/TTS operation is performed using TNS/TTS filter information provided via line 109, i.e., the TTS side information is preferably included in the first encoded representation generated by the spectral domain encoder 106 which can, for example, be a straightforward AAC or USAC core encoder, or can also be included in the second encoded representation. At the output of block 210, a complete spectrum until the maximum frequency is provided which is the full range frequency defined by the sampling rate of the original input signal. Then, a spectrum/time conversion is performed in the synthesis filterbank 212 to finally obtain the audio output signal.

Fig. 3a illustrates a schematic representation of the spectrum. The spectrum is subdivided in scale factor bands SCB where there are seven scale factor bands SCB1 to SCB7 in the illustrated example of Fig. 3a. The scale factor bands can be AAC scale factor bands which are defined in the AAC standard and have an increasing bandwidth to upper frequencies as illustrated in Fig. 3a schematically. It is preferred to perform intelligent gap

filling not from the very beginning of the spectrum, i.e., at low frequencies, but to start the IGF operation at an IGF start frequency illustrated at 309. Therefore, the core frequency band extends from the lowest frequency to the IGF start frequency. Above the IGF start frequency, the spectrum analysis is applied to separate high resolution spectral components 304, 305, 306, 307 (the first set of first spectral portions) from low resolution components represented by the second set of second spectral portions. Fig. 3a illustrates a spectrum which is exemplarily input into the spectral domain encoder 106 or the joint channel coder 228, i.e., the core encoder operates in the full range, but encodes a significant amount of zero spectral values, i.e., these zero spectral values are quantized to zero or are set to zero before quantizing or subsequent to quantizing. Anyway, the core encoder operates in full range, i.e., as if the spectrum would be as illustrated, i.e., the core decoder does not necessarily have to be aware of any intelligent gap filling or encoding of the second set of second spectral portions with a lower spectral resolution.

Preferably, the high resolution is defined by a line-wise coding of spectral lines such as MDCT lines, while the second resolution or low resolution is defined by, for example, calculating only a single spectral value per scale factor band, where a scale factor band covers several frequency lines. Thus, the second low resolution is, with respect to its spectral resolution, much lower than the first or high resolution defined by the line-wise coding typically applied by the core encoder such as an AAC or USAC core encoder.

Regarding scale factor or energy calculation, the situation is illustrated in Fig. 3b. Due to the fact that the encoder is a core encoder and due to the fact that there can, but does not necessarily have to be, components of the first set of spectral portions in each band, the core encoder calculates a scale factor for each band not only in the core range below the IGF start frequency 309, but also above the IGF start frequency until the maximum frequency $f_{IGFstop}$ which is smaller or equal to the half of the sampling frequency, i.e., $f_{s/2}$. Thus, the encoded tonal portions 302, 304, 305, 306, 307 of Fig. 3a and, in this embodiment together with the scale factors SCB1 to SCB7 correspond to the high resolution spectral data. The low resolution spectral data are calculated starting from the IGF start frequency and correspond to the energy information values E_1, E_2, E_3, E_4, which are transmitted together with the scale factors SF4 to SF7.

Particularly, when the core encoder is under a low bitrate condition, an additional noise-filling operation in the core band, i.e., lower in frequency than the IGF start frequency, i.e., in scale factor bands SCB1 to SCB3 can be applied in addition. In noise-filling, there exist

several adjacent spectral lines which have been quantized to zero. On the decoder-side, these quantized to zero spectral values are re-synthesized and the re-synthesized spectral values are adjusted in their magnitude using a noise-filling energy such as NF_2 illustrated at 308 in Fig. 3b. The noise-filling energy, which can be given in absolute terms or in relative terms particularly with respect to the scale factor as in USAC corresponds to the energy of the set of spectral values quantized to zero. These noise-filling spectral lines can also be considered to be a third set of third spectral portions which are regenerated by straightforward noise-filling synthesis without any IGF operation relying on frequency regeneration using frequency tiles from other frequencies for reconstructing frequency tiles using spectral values from a source range and the energy information E_1, E_2, E_3, E_4.

Preferably, the bands, for which energy information is calculated coincide with the scale factor bands. In other embodiments, an energy information value grouping is applied so that, for example, for scale factor bands 4 and 5, only a single energy information value is transmitted, but even in this embodiment, the borders of the grouped reconstruction bands coincide with borders of the scale factor bands. If different band separations are applied, then certain re-calculations or synchronization calculations may be applied, and this can make sense depending on the certain implementation.

Preferably, the spectral domain encoder 106 of Fig. 1a is a psycho-acoustically driven encoder as illustrated in Fig. 4a. Typically, as for example illustrated in the MPEG2/4 AAC standard or MPEG1/2, Layer 3 standard, the to be encoded audio signal after having been transformed into the spectral range (401 in Fig. 4a) is forwarded to a scale factor calculator 400. The scale factor calculator is controlled by a psycho-acoustic model additionally receiving the to be quantized audio signal or receiving, as in the MPEG1/2 Layer 3 or MPEG AAC standard, a complex spectral representation of the audio signal. The psycho-acoustic model calculates, for each scale factor band, a scale factor representing the psycho-acoustic threshold. Additionally, the scale factors are then, by cooperation of the well-known inner and outer iteration loops or by any other suitable encoding procedure adjusted so that certain bitrate conditions are fulfilled. Then, the to be quantized spectral values on the one hand and the calculated scale factors on the other hand are input into a quantizer processor 404. In the straightforward audio encoder operation, the to be quantized spectral values are weighted by the scale factors and, the weighted spectral values are then input into a fixed quantizer typically having a compression functionality to upper amplitude ranges. Then, at the output of the quantizer processor there do exist quantization indices which are then forwarded into an entropy

encoder typically having specific and very efficient coding for a set of zero-quantization indices for adjacent frequency values or, as also called in the art, a "run" of zero values.

In the audio encoder of Fig. 1a, however, the quantizer processor typically receives information on the second spectral portions from the spectral analyzer. Thus, the quantizer processor 404 makes sure that, in the output of the quantizer processor 404, the second spectral portions as identified by the spectral analyzer 102 are zero or have a representation acknowledged by an encoder or a decoder as a zero representation which can be very efficiently coded, specifically when there exist "runs" of zero values in the spectrum.

Fig. 4b illustrates an implementation of the quantizer processor. The MDCT spectral values can be input into a set to zero block 410. Then, the second spectral portions are already set to zero before a weighting by the scale factors in block 412 is performed. In an additional implementation, block 410 is not provided, but the set to zero cooperation is performed in block 418 subsequent to the weighting block 412. In an even further implementation, the set to zero operation can also be performed in a set to zero block 422 subsequent to a quantization in the quantizer block 420. In this implementation, blocks 410 and 418 would not be present. Generally, at least one of the blocks 410, 418, 422 are provided depending on the specific implementation.

Then, at the output of block 422, a quantized spectrum is obtained corresponding to what is illustrated in Fig. 3a. This quantized spectrum is then input into an entropy coder such as 232 in Fig. 2b which can be a Huffman coder or an arithmetic coder as, for example, defined in the USAC standard.

The set to zero blocks 410, 418, 422, which are provided alternatively to each other or in parallel are controlled by the spectral analyzer 424. The spectral analyzer preferably comprises any implementation of a well-known tonality detector or comprises any different kind of detector operative for separating a spectrum into components to be encoded with a high resolution and components to be encoded with a low resolution. Other such algorithms implemented in the spectral analyzer can be a voice activity detector, a noise detector, a speech detector or any other detector deciding, depending on spectral information or associated metadata on the resolution requirements for different spectral portions.

Fig. 5a illustrates a preferred implementation of the time spectrum converter 100 of Fig. 1a as, for example, implemented in AAC or USAC. The time spectrum converter 100 comprises a windower 502 controlled by a transient detector 504. When the transient detector 504 detects a transient, then a switchover from long windows to short windows is signaled to the windower. The windower 502 then calculates, for overlapping blocks, windowed frames, where each windowed frame typically has two N values such as 2048 values. Then, a transformation within a block transformer 506 is performed, and this block transformer typically additionally provides a decimation, so that a combined decimation/transform is performed to obtain a spectral frame with N values such as MDCT spectral values. Thus, for a long window operation, the frame at the input of block 506 comprises two N values such as 2048 values and a spectral frame then has 1024 values. Then, however, a switch is performed to short blocks, when eight short blocks are performed where each short block has 1/8 windowed time domain values compared to a long window and each spectral block has 1/8 spectral values compared to a long block. Thus, when this decimation is combined with a 50% overlap operation of the windower, the spectrum is a critically sampled version of the time domain audio signal 99.

Subsequently, reference is made to Fig. 5b illustrating a specific implementation of frequency regenerator 116 and the spectrum-time converter 118 of Fig. 1b, or of the combined operation of blocks 208, 212 of Fig. 2a. In Fig. 5b, a specific reconstruction band is considered such as scale factor band 6 of Fig. 3a. The first spectral portion in this reconstruction band, i.e., the first spectral portion 306 of Fig. 3a is input into the frame builder/adjustor block 510. Furthermore, a reconstructed second spectral portion for the scale factor band 6 is input into the frame builder/adjuster 510 as well. Furthermore, energy information such as E_3 of Fig. 3b for a scale factor band 6 is also input into block 510. The reconstructed second spectral portion in the reconstruction band has already been generated by frequency tile filling using a source range and the reconstruction band then corresponds to the target range. Now, an energy adjustment of the frame is performed to then finally obtain the complete reconstructed frame having the N values as, for example, obtained at the output of combiner 208 of Fig. 2a. Then, in block 512, an inverse block transform/interpolation is performed to obtain 248 time domain values for the for example 124 spectral values at the input of block 512. Then, a synthesis windowing operation is performed in block 514 which is again controlled by a long window/short window indication transmitted as side information in the encoded audio signal. Then, in block 516, an overlap/add operation with a previous time frame is performed. Preferably, MDCT applies a 50% overlap so that, for each new time frame of

2N values, N time domain values are finally output. A 50% overlap is heavily preferred due to the fact that it provides critical sampling and a continuous crossover from one frame to the next frame due to the overlap/add operation in block 516.

As illustrated at 301 in Fig. 3a, a noise-filling operation can additionally be applied not only below the IGF start frequency, but also above the IGF start frequency such as for the contemplated reconstruction band coinciding with scale factor band 6 of Fig. 3a. Then, noise-filling spectral values can also be input into the frame builder/adjuster 510 and the adjustment of the noise-filling spectral values can also be applied within this block or the noise-filling spectral values can already be adjusted using the noise-filling energy before being input into the frame builder/adjuster 510.

Preferably, an IGF operation, i.e., a frequency tile filling operation using spectral values from other portions can be applied in the complete spectrum. Thus, a spectral tile filling operation can not only be applied in the high band above an IGF start frequency but can also be applied in the low band. Furthermore, the noise-filling without frequency tile filling can also be applied not only below the IGF start frequency but also above the IGF start frequency. It has, however, been found that high quality and high efficient audio encoding can be obtained when the noise-filling operation is limited to the frequency range below the IGF start frequency and when the frequency tile filling operation is restricted to the frequency range above the IGF start frequency as illustrated in Fig. 3a.

Preferably, the target tiles (TT) (having frequencies greater than the IGF start frequency) are bound to scale factor band borders of the full rate coder. Source tiles (ST), from which information is taken, i.e., for frequencies lower than the IGF start frequency are not bound by scale factor band borders. The size of the ST should correspond to the size of the associated TT. This is illustrated using the following example. TT[0] has a length of 10 MDCT Bins. This exactly corresponds to the length of two subsequent SCBs (such as 4 + 6). Then, all possible ST that are to be correlated with TT[0], have a length of 10 bins, too. A second target tile TT[1] being adjacent to TT[0] has a length of 15 bins I (SCB having a length of 7 + 8). Then, the ST for that have a length of 15 bins rather than 10 bins as for TT[0].

Should the case arise that one cannot find a TT for an ST with the length of the target tile (when e.g. the length of TT is greater than the available source range), then a correlation is not calculated and the source range is copied a number of times into this TT (the

copying is done one after the other so that a frequency line for the lowest frequency of the second copy immediately follows - in frequency - the frequency line for the highest frequency of the first copy), until the target tile TT is completely filled up.

Subsequently, reference is made to Fig. 5c illustrating a further preferred embodiment of the frequency regenerator 116 of Fig. 1b or the IGF block 202 of Fig. 2a. Block 522 is a frequency tile generator receiving, not only a target band ID, but additionally receiving a source band ID. Exemplarily, it has been determined on the encoder-side that the scale factor band 3 of Fig. 3a is very well suited for reconstructing scale factor band 7. Thus, the source band ID would be 2 and the target band ID would be 7. Based on this information, the frequency tile generator 522 applies a copy up or harmonic tile filling operation or any other tile filling operation to generate the raw second portion of spectral components 523. The raw second portion of spectral components has a frequency resolution identical to the frequency resolution included in the first set of first spectral portions.

Then, the first spectral portion of the reconstruction band such as 307 of Fig. 3a is input into a frame builder 524 and the raw second portion 523 is also input into the frame builder 524. Then, the reconstructed frame is adjusted by the adjuster 526 using a gain factor for the reconstruction band calculated by the gain factor calculator 528. Importantly, however, the first spectral portion in the frame is not influenced by the adjuster 526, but only the raw second portion for the reconstruction frame is influenced by the adjuster 526. To this end, the gain factor calculator 528 analyzes the source band or the raw second portion 523 and additionally analyzes the first spectral portion in the reconstruction band to finally find the correct gain factor 527 so that the energy of the adjusted frame output by the adjuster 526 has the energy E_4 when a scale factor band 7 is contemplated.

In this context, it is very important to evaluate the high frequency reconstruction accuracy of the present invention compared to HE-AAC. This is explained with respect to scale factor band 7 in Fig. 3a. It is assumed that a prior art encoder such as illustrated in Fig. 13a would detect the spectral portion 307 to be encoded with a high resolution as a "missing harmonics". Then, the energy of this spectral component would be transmitted together with a spectral envelope information for the reconstruction band such as scale factor band 7 to the decoder. Then, the decoder would recreate the missing harmonic. However, the spectral value, at which the missing harmonic 307 would be reconstructed by the prior art decoder of Fig. 13b would be in the middle of band 7 at a frequency

indicated by reconstruction frequency 390. Thus, the present invention avoids a frequency error 391 which would be introduced by the prior art decoder of Fig. 13d.

In an implementation, the spectral analyzer is also implemented to calculating similarities between first spectral portions and second spectral portions and to determine, based on the calculated similarities, for a second spectral portion in a reconstruction range a first spectral portion matching with the second spectral portion as far as possible. Then, in this variable source range/destination range implementation, the parametric coder will additionally introduce into the second encoded representation a matching information indicating for each destination range a matching source range. On the decoder-side, this information would then be used by a frequency tile generator 522 of Fig. 5c illustrating a generation of a raw second portion 523 based on a source band ID and a target band ID.

Furthermore, as illustrated in Fig. 3a, the spectral analyzer is configured to analyze the spectral representation up to a maximum analysis frequency being only a small amount below half of the sampling frequency and preferably being at least one quarter of the sampling frequency or typically higher.

As illustrated, the encoder operates without downsampling and the decoder operates without upsampling. In other words, the spectral domain audio coder is configured to generate a spectral representation having a Nyquist frequency defined by the sampling rate of the originally input audio signal.

Furthermore, as illustrated in Fig. 3a, the spectral analyzer is configured to analyze the spectral representation starting with a gap filling start frequency and ending with a maximum frequency represented by a maximum frequency included in the spectral representation, wherein a spectral portion extending from a minimum frequency up to the gap filling start frequency belongs to the first set of spectral portions and wherein a further spectral portion such as 304, 305, 306, 307 having frequency values above the gap filling frequency additionally is included in the first set of first spectral portions.

As outlined, the spectral domain audio decoder 112 is configured so that a maximum frequency represented by a spectral value in the first decoded representation is equal to a maximum frequency included in the time representation having the sampling rate wherein the spectral value for the maximum frequency in the first set of first spectral portions is zero or different from zero. Anyway, for this maximum frequency in the first set of spectral

components a scale factor for the scale factor band exists, which is generated and transmitted irrespective of whether all spectral values in this scale factor band are set to zero or not as discussed in the context of Figs. 3a and 3b.

The invention is, therefore, advantageous that with respect to other parametric techniques to increase compression efficiency, e.g. noise substitution and noise filling (these techniques are exclusively for efficient representation of noise like local signal content) the invention allows an accurate frequency reproduction of tonal components. To date, no state-of-the-art technique addresses the efficient parametric representation of arbitrary signal content by spectral gap filling without the restriction of a fixed a-priory division in low band (LF) and high band (HF).

Embodiments of the inventive system improve the state-of-the-art approaches and thereby provides high compression efficiency, no or only a small perceptual annoyance and full audio bandwidth even for low bitrates.

The general system consists of

- full band core coding
- intelligent gap filling (tile filling or noise filling)
- sparse tonal parts in core selected by tonal mask
- joint stereo pair coding for full band, including tile filling
- TNS on tile
- spectral whitening in IGF range

A first step towards a more efficient system is to remove the need for transforming spectral data into a second transform domain different from the one of the core coder. As the majority of audio codecs, such as AAC for instance, use the MDCT as basic transform, it is useful to perform the BWE in the MDCT domain also. A second requirement for the BWE system would be the need to preserve the tonal grid whereby even HF tonal components are preserved and the quality of the coded audio is thus superior to the existing systems. To take care of both the above mentioned requirements for a BWE scheme, a new system is proposed called Intelligent Gap Filling (IGF). Fig. 2b shows the block diagram of the proposed system on the encoder-side and Fig. 2a shows the system on the decoder-side.

Fig. 9a illustrates an apparatus for decoding an encoded audio signal comprising an encoded representation of a first set of first spectral portions and an encoded

representation of parametric data indicating spectral energies for a second set of second spectral portions. The first set of first spectral portions is indicated at 901a in Fig. 9a, and the encoded representation of the parametric data is indicated at 901b in Fig. 9a. An audio decoder 900 is provided for decoding the encoded representation 901a of the first set of first spectral portions to obtain a decoded first set of first spectral portions 904 and for decoding the encoded representation of the parametric data to obtain a decoded parametric data 902 for the second set of second spectral portions indicating individual energies for individual reconstruction bands, where the second spectral portions are located in the reconstruction bands. Furthermore, a frequency regenerator 906 is provided for reconstructing spectral values of a reconstruction band comprising a second spectral portion. The frequency regenerator 906 uses a first spectral portion of the first set of first spectral portions and an individual energy information for the reconstruction band, where the reconstruction band comprises a first spectral portion and the second spectral portion. The frequency regenerator 906 comprises a calculator 912 for determining a survive energy information comprising an accumulated energy of the first spectral portion having frequencies in the reconstruction band. Furthermore, the frequency regenerator 906 comprises a calculator 918 for determining a tile energy information of further spectral portions of the reconstruction band and for frequency values being different from the first spectral portion, where these frequency values have frequencies in the reconstruction band, wherein the further spectral portions are to be generated by frequency regeneration using a first spectral portion different from the first spectral portion in the reconstruction band.

The frequency regenerator 906 further comprises a calculator 914 for a missing energy in the reconstruction band, and the calculator 914 operates using the individual energy for the reconstruction band and the survive energy generated by block 912. Furthermore, the frequency regenerator 906 comprises a spectral envelope adjuster 916 for adjusting the further spectral portions in the reconstruction band based on the missing energy information and the tile energy information generated by block 918.

Reference is made to Fig. 9c illustrating a certain reconstruction band 920. The reconstruction band comprises a first spectral portion in the reconstruction band such as the first spectral portion 306 in Fig. 3a schematically illustrated at 921. Furthermore, the rest of the spectral values in the reconstruction band 920 are to be generated using a source region, for example, from the scale factor band 1, 2, 3 below the intelligent gap filling start frequency 309 of Fig. 3a. The frequency regenerator 906 is configured for generating raw spectral values for the second spectral portions 922 and 923. Then, a gain factor g is calculated as illustrated in Fig. 9c in order to finally adjust the raw spectral

values in frequency bands 922, 923 in order to obtain the reconstructed and adjusted second spectral portions in the reconstruction band 920 which now have the same spectral resolution, i.e., the same line distance as the first spectral portion 921. It is important to understand that the first spectral portion in the reconstruction band illustrated at 921 in Fig. 9c is decoded by the audio decoder 900 and is not influenced by the envelope adjustment performed block 916 of Fig. 9b. Instead, the first spectral portion in the reconstruction band indicated at 921 is left as it is, since this first spectral portion is output by the full bandwidth or full rate audio decoder 900 via line 904.

Subsequently, a certain example with real numbers is discussed. The remaining survive energy as calculated by block 912 is, for example, five energy units and this energy is the energy of the exemplarily indicated four spectral lines in the first spectral portion 921.

Furthermore, the energy value E3 for the reconstruction band corresponding to scale factor band 6 of Fig. 3b or Fig. 3a is equal to 10 units. Importantly, the energy value not only comprises the energy of the spectral portions 922, 923, but the full energy of the reconstruction band 920 as calculated on the encoder-side, i.e., before performing the spectral analysis using, for example, the tonality mask. Therefore, the ten energy units cover the first and the second spectral portions in the reconstruction band. Then, it is assumed that the energy of the source range data for blocks 922, 923 or for the raw target range data for block 922, 923 is equal to eight energy units. Thus, a missing energy of five units is calculated.

Based on the missing energy divided by the tile energy tEk, a gain factor of 0.79 is calculated. Then, the raw spectral lines for the second spectral portions 922, 923 are multiplied by the calculated gain factor. Thus, only the spectral values for the second spectral portions 922, 923 are adjusted and the spectral lines for the first spectral portion 921 are not influenced by this envelope adjustment. Subsequent to multiplying the raw spectral values for the second spectral portions 922, 923, a complete reconstruction band has been calculated consisting of the first spectral portions in the reconstruction band, and consisting of spectral lines in the second spectral portions 922, 923 in the reconstruction band 920.

Preferably, the source range for generating the raw spectral data in bands 922, 923 is, with respect to frequency, below the IGF start frequency 309 and the reconstruction band 920 is above the IGF start frequency 309.

Furthermore, it is preferred that reconstruction band borders coincide with scale factor band borders. Thus, a reconstruction band has, in one embodiment, the size of corresponding scale factor bands of the core audio decoder or are sized so that, when energy pairing is applied, an energy value for a reconstruction band provides the energy of two or a higher integer number of scale factor bands. Thus, when is assumed that energy accumulation is performed for scale factor band 4, scale factor band 5 and scale factor band 6, then the lower frequency border of the reconstruction band 920 is equal to the lower border of scale factor band 4 and the higher frequency border of the reconstruction band 920 coincides with the higher border of scale factor band 6.

Subsequently, Fig. 9d is discussed in order to show further functionalities of the decoder of Fig. 9a. The audio decoder 900 receives the dequantized spectral values corresponding to first spectral portions of the first set of spectral portions and, additionally, scale factors for scale factor bands such as illustrated in Fig. 3b are provided to an inverse scaling block 940. The inverse scaling block 940 provides all first sets of first spectral portions below the IGF start frequency 309 of Fig. 3a and, additionally, the first spectral portions above the IGF start frequency, i.e., the first spectral portions 304, 305, 306, 307 of Fig. 3a which are all located in a reconstruction band as illustrated at 941 in Fig. 9d. Furthermore, the first spectral portions in the source band used for frequency tile filling in the reconstruction band are provided to the envelope adjuster/calculator 942 and this block additionally receives the energy information for the reconstruction band provided as parametric side information to the encoded audio signal as illustrated at 943 in Fig. 9d. Then, the envelope adjuster/calculator 942 provides the functionalities of Fig. 9b and 9c and finally outputs adjusted spectral values for the second spectral portions in the reconstruction band. These adjusted spectral values 922, 923 for the second spectral portions in the reconstruction band and the first spectral portions 921 in the reconstruction band indicated that line 941 in Fig. 9d jointly represent the complete spectral representation of the reconstruction band.

Subsequently, reference is made to Figs. 10a to 10b for explaining preferred embodiments of an audio encoder for encoding an audio signal to provide or generate an encoded audio signal. The encoder comprises a time/spectrum converter 1002 feeding a spectral analyzer 1004, and the spectral analyzer 1004 is connected to a parameter calculator 1006 on the one hand and an audio encoder 1008 on the other hand. The audio encoder 1008 provides the encoded representation of a first set of first spectral portions and does not cover the second set of second spectral portions. On the other hand, the parameter calculator 1006 provides energy information for a reconstruction band covering the first and second spectral portions. Furthermore, the audio encoder 1008 is configured

for generating a first encoded representation of the first set of first spectral portions having the first spectral resolution, where the audio encoder 1008 provides scale factors for all bands of the spectral representation generated by block 1002. Additionally, as illustrated in Fig. 3b, the encoder provides energy information at least for reconstruction bands located, with respect to frequency, above the IGF start frequency 309 as illustrated in Fig. 3a. Thus, for reconstruction bands preferably coinciding with scale factor bands or with groups of scale factor bands, two values are given, i.e., the corresponding scale factor from the audio encoder 1008 and, additionally, the energy information output by the parameter calculator 1006.

The audio encoder preferably has scale factor bands with different frequency bandwidths, i.e., with a different number of spectral values. Therefore, the parametric calculator comprise a normalizer 1012 for normalizing the energies for the different bandwidth with respect to the bandwidth of the specific reconstruction band. To this end, the normalizer 1012 receives, as inputs, an energy in the band and a number of spectral values in the band and the normalizer 1012 then outputs a normalized energy per reconstruction/scale factor band.

Furthermore, the parametric calculator 1006a of Fig. 10a comprises an energy value calculator receiving control information from the core or audio encoder 1008 as illustrated by line 1007 in Fig. 10a. This control information may comprise information on long/short blocks used by the audio encoder and/or grouping information. Hence, while the information on long/short blocks and grouping information on short windows relate to a "time" grouping, the grouping information may additionally refer to a spectral grouping, i.e., the grouping of two scale factor bands into a single reconstruction band. Hence, the energy value calculator 1014 outputs a single energy value for each grouped band covering a first and a second spectral portion when only the spectral portions have been grouped.

Fig. 10d illustrates a further embodiment for implementing the spectral grouping. To this end, block 1016 is configured for calculating energy values for two adjacent bands. Then, in block 1018, the energy values for the adjacent bands are compared and, when the energy values are not so much different or less different than defined by, for example, a threshold, then a single (normalized) value for both bands is generated as indicated in block 1020. As illustrated by line 1019, the block 1018 can be bypassed. Furthermore, the generation of a single value for two or more bands performed by block 1020 can be controlled by an encoder bitrate control 1024. Thus, when the bitrate is to be reduced, the encoded bitrate control 1024 controls block 1020 to generate a single normalized value for

two or more bands even though the comparison in block 1018 would not have been allowed to group the energy information values.

In case the audio encoder is performing the grouping of two or more short windows, this grouping is applied for the energy information as well. When the core encoder performs a grouping of two or more short blocks, then, for these two or more blocks, only a single set of scale factors is calculated and transmitted. On the decoder-side, the audio decoder then applies the same set of scale factors for both grouped windows.

Regarding the energy information calculation, the spectral values in the reconstruction band are accumulated over two or more short windows. In other words, this means that the spectral values in a certain reconstruction band for a short block and for the subsequent short block are accumulated together and only single energy information value is transmitted for this reconstruction band covering two short blocks. Then, on the decoder-side, the envelope adjustment discussed with respect to Fig. 9a to 9d is not performed individually for each short block but is performed together for the set of grouped short windows.

The corresponding normalization is then again applied so that even though any grouping in frequency or grouping in time has been performed, the normalization easily allows that, for the energy value information calculation on the decoder-side, only the energy information value on the one hand and the amount of spectral lines in the reconstruction band or in the set of grouped reconstruction bands has to be known.

Furthermore, it is emphasized that an information on spectral energies, an information on individual energies or an individual energy information, an information on a survive energy or a survive energy information, an information a tile energy or a tile energy information, or an information on a missing energy or a missing energy information may comprise not only an energy value, but also an (e.g. absolute) amplitude value, a level value or any other value, from which a final energy value can be derived. Hence, the information on an energy may e.g. comprise the energy value itself, and/or a value of a level and/or of an amplitude and/or of an absolute amplitude.

Fig. 12a illustrates a further implementation of the apparatus for decoding. A bitstream is received by a core decoder 1200 which can, for example, be an AAC decoder. The result is configured into a stage for performing a bandwidth extension patching or tiling 1202 corresponding to the frequency regenerator 604 for example. Then, a procedure of

patch/tile adaption and post-processing is performed, and, when a patch adaption has been performed, the frequency regenerator 1202 is controlled to perform a further frequency regeneration, but now with, for example adjusted frequency borders. Furthermore, when a patch processing is performed such as by the elimination or attenuation of tonal lines, the result is then forwarded to block 1206 performing the parameter-driven bandwidth envelope shaping as, for example, also discussed in the context of block 712 or 826. The result is then forwarded to a synthesis transform block 1208 for performing a transform into the final output domain which is, for example, a PCM output domain as illustrated in Fig. 12a.

Main features of embodiments of the invention are as follows:

The preferred embodiment is based on the MDCT that exhibits the above referenced warbling artifacts if tonal spectral areas are pruned by the unfortunate choice of cross-over frequency and/or patch margins, or tonal components get to be placed in too close vicinity at patch borders.

Fig. 12b shows how the newly proposed technique reduces artifacts found in state-of-the-art BWE methods. In Fig. 12 panel (2), the stylized magnitude spectrum of the output of a contemporary BWE method is shown. In this example, the signal is perceptually impaired by the beating caused by to two nearby tones, and also by the splitting of a tone. Both problematic spectral areas are marked with a circle each.

To overcome these problems, the new technique first detects the spectral location of the tonal components contained in the signal. Then, according to one aspect of the invention, it is attempted to adjust the transition frequencies between LF and all patches by individual shifts (within given limits) such that splitting or beating of tonal components is minimized. For that purpose, the transition frequency preferably has to match a local spectral minimum. This step is shown in Fig. 12b panel (2) and panel (3), where the transition frequency f_{x2} is shifted towards higher frequencies, resulting in f'_{x2}.

According to another aspect of the invention, if problematic spectral content in transition regions remains, at least one of the misplaced tonal components is removed to reduce either the beating artifact at the transition frequencies or the warbling. This is done via spectral extrapolation or interpolation / filtering, as shown in Figure 2 panel (3). A tonal component is thereby removed from foot-point to foot-point, i.e. from its left local minimum

to its right local minimum. The resulting spectrum after the application of the inventive technology is shown in Fig. 12b panel (4).

In other words, Fig. 12b illustrates, in the upper left corner, i.e., in panel (1), the original signal. In the upper right corner, i.e., in panel (2), a comparison bandwidth extended signal with problematic areas marked by ellipses 1220 and 1221 is shown. In the lower left corner, i.e., in panel (3), two preferred patch or frequency tile processing features are illustrated. The splitting of tonal portions has been addressed by increasing the frequency border f'_{x2} so that a clipping of the corresponding tonal portion is not there anymore. Furthermore, gain functions 1030 for eliminating the tonal portion 1031 and 1032 are applied or, alternatively, an interpolation illustrated by 1033 is indicated. Finally, the lower right corner of Fig. 12b, i.e., panel (4) depicts the improved signal resulting from a combination of tile/patch frequency adjusting on the one hand and elimination or at least attenuation of problematic tonal portions.

Panel (1) of Fig. 12b illustrates, as discussed before, the original spectrum, and the original spectrum has a core frequency range up to the cross-over or gap filing start frequency *fx1*.

Thus, a frequency f_{x1} illustrates a border frequency 1250 between the source range 1252 and a reconstruction range 1254 extending between the border frequency 1250 and a maximum frequency which is smaller than or equal to the Nyquist frequency $f_{Nyquist}$. On the encoder-side, it is assumed that a signal is bandwidth-limited at f_{x1} or, when the technology regarding intelligent gap filling is applied, it is assumed that f_{x1} corresponds to the gap filling start frequency 309 of Fig. 3a. Depending on the technology, the reconstruction range above f_{x1} will be empty (in case of the Fig. 13a, 13b implementation) or will comprise certain first spectral portions to be encoded with a high resolution as discussed in the context of Fig. 3a.

Fig. 12b, panel (2) illustrates a preliminary regenerated signal, for example generated by block 702 of Fig. 7a which has two problematic portions. One problematic portion is illustrated at 1220. the frequency distance between the tonal portion within the core region illustrated at 1220a and the tonal portion at the start of the frequency tile illustrated at 1220b is too small so that a beating artifact would be created. The further problem is that at the upper border of the first frequency tile generated by the first patching operation or frequency tiling operation illustrated at 1225 is a halfway-clipped or split tonal portion

1226. When this tonal portion 1226 is compared to the other tonal portions in Fig. 12b, it becomes clear that the width is smaller than the width of a typical tonal portion and this means that this tonal portion has been split by setting the frequency border between the first frequency tile 1225 and the second frequency tile 1227 at the wrong place in the source range 1252. In order to address this issue, the border frequency f_{x2} has been modified to become a little bit greater as illustrated in panel (3) in Fig. 12b, so that a clipping of this tonal portion does not occur.

On the other hand, this procedure, in which f'_{x2} has been changed does not effectively address the beating problem which, therefore, is addressed by a removal of the tonal components by filtering or interpolation or any other procedures as discussed in the context of block 708 of Fig. 7a. Thus, Fig. 12b illustrates a sequential application of the transition frequency adjustment 706 and the removal of tonal components at borders illustrated at 708.

Another option would have been to set the transition border f_{x1} so that it is a little bit lower so that the tonal portion 1220a is not in the core range anymore. Then, the tonal portion 1220a has also been removed or eliminated by setting the transition frequency f_{x1} at a lower value.

This procedure would also have worked for addressing the issue with the problematic tonal component 1032. By setting f'_{x2} even higher, the spectral portion where the tonal portion 1032 is located could have been regenerated within the first patching operation 1225 and, therefore, two adjacent or neighboring tonal portions would not have occurred.

Basically, the beating problem depends on the amplitudes and the distance in frequency of adjacent tonal portions. The detector 704, 720 or stated more general, the analyzer 602 is preferably configured in such a way that an analysis of the lower spectral portion located in the frequency below the transition frequency such as f_{x1}, f_{x2}, f'_{x2} is analyzed in order to locate any tonal component. Furthermore, the spectral range above the transition frequency is also analyzed in order to detect a tonal component. When the detection results in two tonal components, one to the left of the transition frequency with respect to frequency and one to the right (with respect to ascending frequency), then the remover of tonal components at borders illustrated at 708 in Fig. 7a is activated. The detection of tonal components is performed in a certain detection range which extends, from the transition frequency, in both directions at least 20% with respect to the bandwidth of the

corresponding band and preferably only extends up to 10% downwards to the left of the transition frequency and upwards to the right of the transition frequency related to the corresponding bandwidth, i.e., the bandwidth of the source range on the one hand and the reconstruction range on the other hand or, when the transition frequency is the transition frequency between two frequency tiles 1225, 1227, a corresponding 10% amount of the corresponding frequency tile. In a further embodiment, the predetermined detection bandwidth is one Bark. It should be possible to remove tonal portions within a range of 1 Bark around a patch border, so that the complete detection range is 2 Bark, i.e., one Bark in the lower band and one Bark in the higher band, where the one Bark in the lower band is immediately adjacent to the one Bark in the higher band.

According to another aspect of the invention, to reduce the filter ringing artifact, a cross-over filter in the frequency domain is applied to two consecutive spectral regions, i.e. between the core band and the first patch or between two patches. Preferably, the cross-over filter is signal adaptive.

The cross over filter consists of two filters, a fade-out filter $\boldsymbol{h}_{out}$, which is applied to the lower spectral region, and a fade-in filter $\boldsymbol{h}_{in}$, which is applied to the higher spectral region.

Each of the filters has length N.

In addition, the slope of both filters is characterized by a signal adaptive value called $Xbias$ determining the notch characteristic of the cross-over filter, with $0 \leq Xbias \leq N$:

If $Xbias = 0$, then the sum of both filters is equal to 1, i.e. there is no notch filter characteristic in the resulting filter.

If $Xbias = N$, then both filters are completely zero.

The basic design of the cross-over filters is constraint to the following equations:

$$\boldsymbol{h}_{out}(k) = \boldsymbol{h}_{in}(N-1-k), \forall\, Xbias$$
$$\boldsymbol{h}_{out}(k) + \boldsymbol{h}_{in}(k) = 1, Xbias = 0$$

with $k = 0, 1, \ldots, N-1$ being the frequency index. Fig. 12c shows an example of such a cross-over filter.

In this example, the following equation is used to create the filter h_{out}:

$$\boldsymbol{h}_{out}(k) = 0.5 + 0.5 \cdot cos\left(\frac{k}{N-1-Xbias} \cdot \pi\right), k = 0, 1, \ldots, N-1-Xbias$$

The following equation describes how the filters h_{in} and h_{out} are then applied,

$$\begin{aligned} \boldsymbol{Y}(k_t - (N-1) + k) = \quad & \boldsymbol{LF}(k_t - (N-1) + k) \cdot \boldsymbol{h}_{out}(k) + \\ & \boldsymbol{HF}(k_t - (N-1) + k) \cdot \boldsymbol{h}_{in}(k), \qquad k = 0, 1, \ldots, N-1 \end{aligned}$$

with $\boldsymbol{Y}$ denoting the assembled spectrum, k_t being the transition frequency, $\boldsymbol{LF}$ being the low frequency content and $\boldsymbol{HF}$ being the high frequency content.

Next, evidence of the benefit of this technique will be presented. The original signal in the following examples is a transient-like signal, in particular a low pass filtered version thereof, with a cut-off frequency of 22 kHz. First, this transient is band limited to 6 kHz in the transform domain. Subsequently, the bandwidth of the low pass filtered original signal is extended to 24 kHz. The bandwidth extension is accomplished through copying the LF band three times to entirely fill the frequency range that is available above 6 kHz within the transform.

Fig. 11a shows the spectrum of this signal, which can be considered as a typical spectrum of a filter ringing artifact that spectrally surrounds the transient due to said brick-wall characteristic of the transform (speech peaks 1100). By applying the inventive approach, the filter ringing is reduced by approx. 20 dB at each transition frequency (reduced speech peaks).

The same effect, yet in a different illustration, is shown in Fig. 11b, 11c. Fig. 11b shows the spectrogram of the mentioned transient like signal with the filter ringing artifact that temporally precedes and succeeds the transient after applying the above described BWE technique without any filter ringing reduction. Each of the horizontal lines represents the filter ringing at the transition frequency between consecutive patches. Figure 6 shows the same signal after applying the inventive approach within the BWE. Through the application of ringing reduction, the filter ringing is reduced by approx. 20 dB compared to the signal displayed in the previous Figure.

Subsequently, Figs. 14a, 14b are discussed in order to further illustrate the cross-over filter invention aspect already discussed in the context with the analyzer feature. However,

the cross-over filter 710 can also be implemented independent of the invention discussed in the context of Figs. 6a-7b.

Fig. 14a illustrates an apparatus for decoding an encoded audio signal comprising an encoded core signal and information on parametric data. The apparatus comprises a core decoder 1400 for decoding the encoded core signal to obtain a decoded core signal. The decoded core signal can be bandwidth limited in the context of the Fig. 13a, Fig. 13b implementation or the core decoder can be a full frequency range or full rate coder in the context of Figs. 1 to 5c or 9a-10d.

Furthermore, a tile generator 1404 for regenerating one or more spectral tiles having frequencies not included in the decoded core signal are generated using a spectral portion of the decoded core signal. The tiles can be reconstructed second spectral portions within a reconstruction band as, for example, illustrated in the context of Fig. 3a or which can include first spectral portions to be reconstructed with a high resolution but, alternatively, the spectral tiles can also comprise completely empty frequency bands when the encoder has performed a hard band limitation as illustrated in Fig. 13a.

Furthermore, a cross-over filter 1406 is provided for spectrally cross-over filtering the decoded core signal and a first frequency tile having frequencies extending from a gap filling frequency 309 to a first tile stop frequency or for spectrally cross-over filtering a first frequency tile 1225 and a second frequency tile 1221, the second frequency tile having a lower border frequency being frequency-adjacent to an upper border frequency of the first frequency tile 1225.

In a further implementation, the cross-over filter 1406 output signal is fed into an envelope adjuster 1408 which applies parametric spectral envelope information included in an encoded audio signal as parametric side information to finally obtain an envelope-adjusted regenerated signal. Elements 1404, 1406, 1408 can be implemented as a frequency regenerator as, for example, illustrated in Fig. 13b, Fig. 1b or Fig. 6a, for example.

Fig. 14b illustrates a further implementation of the cross-over filter 1406. The cross-over filter 1406 comprises a fade-out subfilter receiving a first input signal IN1, and a second fade-in subfilter 1422 receiving a second input IN2 and the results or outputs of both filters 1420 and 1422 are provided to a combiner 1424 which is, for example, an adder. The adder or combiner 1424 outputs the spectral values for the frequency bins. Fig. 12c

illustrates an example cross-fade function comprising the fade-out subfilter characteristic 1420a and the fade-in subfilter characteristic 1422a. Both filters have a certain frequency overlap in the example in Fig. 12c equal to 21, i.e., N=21. Thus, other frequency values of, for example, the source region 1252 are not influenced. Only the highest 21 frequency bins of the source range 1252 are influenced by the fade-out function 1420a.

On the other hand, only the lowest 21 frequency lines of the first frequency tile 1225 are influenced by the fade-in function 1422a.

Additionally, it becomes clear from the cross-fade functions that the frequency lines between 9 and 13 are influenced, but the fade-in function actually does not influence the frequency lines between 1 and 9 and face-out function 1420a does not influence the frequency lines between 13 and 21. This means that only an overlap would be necessary between frequency lines 9 and 13, and the cross-over frequency such as f_{x1} would be placed at frequency sample or frequency bin 11. Thus, only an overlap of two frequency bins or frequency values between the source range and the first frequency tile would be required in order to implement the cross-over or cross-fade function.

Depending on the specific implementation, a higher or lower overlap can be applied and, additionally, other fading functions apart from a cosine function can be used. Furthermore, as illustrated in Fig. 12c, it is preferred to apply a certain notch in the cross-over range. Stated differently, the energy in the border ranges will be reduced due to the fact that both filter functions do not add up to unity as it would be the case in a notch-free cross-fade function. This loss of energy for the borders of the frequency tile, i.e., the first frequency tile will be attenuated at the lower border and at the upper border, the energies concentrated more to the middle of the bands. Due to the fact, however, that the spectral envelope adjustment takes place subsequent to the processing by the cross-over filter, the overall frequency is not touched, but is defined by the spectral envelope data such as the corresponding scale factors as discussed in the context of Fig. 3a. In other words, the calculator 918 of Fig. 9b would then calculate the "already generated raw target range", which is the output of the cross-over filter. Furthermore, the energy loss due to the removal of a tonal portion by interpolation would also be compensated for due to the fact that this removal then results in a lower tile energy and the gain factor for the complete reconstruction band will become higher. On the other hand, however, the cross-over frequency results in a concentration of energy more to the middle of a frequency tile and

this, in the end, effectively reduces the artifacts, particularly caused by transients as discussed in the context of Figs. 11a-11c.

Fig. 14b illustrates different input combinations. For a filtering at the border between the source frequency range and the frequency tile, input 1 is the upper spectral portion of the core range and input 2 is the lower spectral portion of the first frequency tile or of the single frequency tile, when only a single frequency tile exists. Furthermore, the input can be the first frequency tile and the transition frequency can be the upper frequency border of the first tile and the input into the subfilter 1422 will be the lower portion of the second frequency tile. When an additional third frequency tile exists, then a further transition frequency will be the frequency border between the second frequency tile and the third frequency tile and the input into the fade-out subfilter 1421 will be the upper spectral range of the second frequency tile as determined by filter parameter, when the Fig. 12c characteristic is used, and the input into the fade-in subfilter 1422 will be the lower portion of the third frequency tile and, in the example of Fig. 12c, the lowest 21 spectral lines.

As illustrated in Fig. 12c, it is preferred to have the parameter N equal for the fade-out subfilter and the fade-in subfilter. This, however, is not necessary. The values for N can vary and the result will then be that the filter "notch" will be asymmetric between the lower and the upper range. Additionally, the fade-in/fade-out functions do not necessarily have to be in the same characteristic as in Fig. 12c. Instead, asymmetric characteristics can also be used.

Furthermore, it is preferred to make the cross-over filter characteristic signal-adaptive. Therefore, based on a signal analysis, the filter characteristic is adapted. Due to the fact that the cross-over filter is particularly useful for transient signals, it is detected whether transient signals occur. When transient signals occur, then a filter characteristic such as illustrated in Fig. 12c could be used. When, however, a non-transient signal is detected, it is preferred to change the filter characteristic to reduce the influence of the cross-over filter. This could, for example, be obtained by setting N to zero or by setting X_{bias} to zero so that the sum of both filters is equal to 1, i.e., there is no notch filter characteristic in the resulting filter. Alternatively, the cross-over filter 1406 could simply be bypassed in case of non-transient signals. Preferably, however, a relatively slow changing filter characteristic by changing parameters N, X_{bias} is preferred in order to avoid artifacts obtained by the quickly changing filter characteristics. Furthermore, a low-pass filter is preferred for only allowing such relatively small filter characteristic changes even though the signal is

changing more rapidly as detected by a certain transient/tonality detector. The detector is illustrated at 1405 in Fig. 14a. It may receive an input signal into a tile generator or an output signal of the tile generator 1404 or it can even be connected to the core decoder 1400 in order to obtain a transient/non-transient information such as a short block indication from AAC decoding, for example. Naturally, any other crossover filter different from the one shown in Fig. 12c can be used as well.

Then, based on the transient detection, or based on a tonality detection or based on any other signal characteristic detection, the cross-over filter 1406 characteristic is changed as discussed.

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

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

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

Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing

one of the methods when the computer program product runs on a computer. The program code may, for example, be stored on a machine readable carrier.

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

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

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

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

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

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

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

In some embodiments, a programmable logic device (for example, a field programmable gate array) may be used to perform some or all of the functionalities of the methods

described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

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

List of citations

[1] Dietz, L. Liljeryd, K. Kjörling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in 112th AES Convention, Munich, May 2002.

[2] Ferreira, D. Sinha, "Accurate Spectral Replacement", Audio Engineering Society Convention, Barcelona, Spain 2005.

[3] D. Sinha, A. Ferreira1 and E. Harinarayanan, "A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)", Audio Engineering Society Convention, Paris, France 2006.

[4] R. Annadana, E. Harinarayanan, A. Ferreira and D. Sinha, "New Results in Low Bit Rate Speech Coding and Bandwidth Extension", Audio Engineering Society Convention, San Francisco, USA 2006.

[5] T. Żernicki, M. Bartkowiak, "Audio bandwidth extension by frequency scaling of sinusoidal partials", Audio Engineering Society Convention, San Francisco, USA 2008.

[6] J. Herre, D. Schulz, Extending the MPEG-4 AAC Codec by Perceptual Noise Substitution, 104th AES Convention, Amsterdam, 1998, Preprint 4720.

[7] M. Neuendorf, M. Multrus, N. Rettelbach, et al., MPEG Unified Speech and Audio Coding-The ISO/MPEG Standard for High-Efficiency Audio Coding of all Content Types, 132nd AES Convention, Budapest, Hungary, April, 2012.

[8] McAulay, Robert J., Quatieri, Thomas F. "Speech Analysis/Synthesis Based on a Sinusoidal Representation". IEEE Transactions on Acoustics, Speech, And Signal Processing, Vol 34(4), August 1986.

[9] Smith, J.O., Serra, X. "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation", Proceedings of the International Computer Music Conference, 1987.

[10] Purnhagen, H.; Meine, Nikolaus, "HILN-the MPEG-4 parametric audio coding tools," *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on* , vol.3, no., pp.201,204 vol.3, 2000

[11] International Standard ISO/IEC 13818-3, Generic Coding of Moving Pictures and Associated Audio: Audio", Geneva, 1998.

[12] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Oikawa: "MPEG-2 Advanced Audio Coding", 101st AES Convention, Los Angeles 1996

[13] J. Herre, "Temporal Noise Shaping, Quantization and Coding methods in Perceptual Audio Coding: A Tutorial introduction", 17th AES International Conference on High Quality Audio Coding, August 1999

[14] J. Herre, "Temporal Noise Shaping, Quantization and Coding methods in Perceptual Audio Coding: A Tutorial introduction", 17th AES International Conference on High Quality Audio Coding, August 1999

[15] International Standard ISO/IEC 23001-3:2010, Unified speech and audio coding Audio, Geneva, 2010.

[16] International Standard ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects - Part 3: Audio, Geneva, 2005.

[17] P. Ekstrand, "Bandwidth Extension of Audio Signals by Spectral Band Replication", in Proceedings of 1st IEEE Benelux Workshop on MPCA, Leuven, November 2002

[18] F. Nagel, S. Disch, S. Wilde, A continuous modulated single sideband bandwidth extension, ICASSP International Conference on Acoustics, Speech and Signal Processing, Dallas, Texas (USA), April 2010

[19] Liljeryd, Lars; Ekstrand, Per; Henn, Fredrik; Kjorling, Kristofer: Spectral translation/folding in the subband domain, United States Patent 8,412,365, April 2, 2013.

[20] Daudet, L.; Sandler, M.; "MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction," Speech and Audio Processing, IEEE Transactions on , vol.12, no.3, pp. 302- 312, May 2004.

Claims

1. Apparatus for decoding an encoded audio signal comprising an encoded core signal (1), comprising:

a core decoder (1400) for decoding the encoded core signal (1401) to obtain a decoded core signal;

a tile generator (1404) for generating one or more spectral tiles having frequencies not included in the decoded core signal using a spectral portion of the decoded core signal; and

a cross-over filter (1406) for spectrally cross-over filtering the decoded core signal and a first frequency tile having frequencies extending from a gap filling frequency (309) to an upper border frequency or for spectrally cross-over filtering a first frequency tile and a second frequency tile.

2. Apparatus of claim 1,

wherein the cross-over filter (1406) is configured to perform a frequency-wise weighted addition (1424) of the decoded core signal filtered by a fade-out subfilter (1420) and at least a portion of the first frequency tile filtered by a fade-in filter (1422) within a cross-over range extending over at least three frequency values or to perform a frequency-wise weighted addition (1424) of at least a part of a first frequency tile filtered by the fade-out subfilter (1420) and at least a part of a second frequency tile filtered by the fade-in subfilter (1422) within a cross-over range extending over at least three frequency values.

3. Apparatus of claim 1 or 2,

wherein a spectral portion of the decoded core signal, a spectral portion of the first frequency tile or a spectral portion of the second frequency tile influenced by the cross-over filter (1406) is smaller than 30% of the spectral portion covered by a total spectral band of the decoded core frequency band or a total spectral band of

the first or second frequency tiles and is greater than or equal to a band defined by at least 5 adjacent frequency values.

4. Apparatus of claim 1, 2, or 3,

wherein the cross-over filter (1406) is configured for applying a cosine-like filter characteristic for fading-in and fading-out.

5. Apparatus in accordance with one of the preceding claims comprising an envelope adjuster (1408) for envelope adjusting a cross-over filtered spectral signal in a spectral range defined by spectral ranges of the one or more spectral tiles using parametric spectral envelope information (1407) included in the encoded audio signal.

6. Apparatus of one of the preceding claims,

further comprising a frequency-time converter (828) for converting an envelope-adjusted signal together with the decoded core signal into a time representation.

7. Apparatus in accordance with claim 6, wherein the frequency-time converter is configured for applying an inverse modified discrete cosine transform (512, 514, 516) comprising an overlap/add processing (516) of a current frame with a preceding time frame.

8. Apparatus in accordance with one of the preceding claims, wherein the cross-over filter is a controllable filter,

wherein the apparatus further comprises a signal characteristics detector (1405), and

wherein the signal characteristics detector (1405) is configured for controlling a filter characteristic of the cross-over filter (1406) in accordance with a detection result derived from the decoded core signal.

9. Apparatus of claim 8,

wherein the signal characteristics detector (1405) is a transient detector, and wherein the transient detector (1405) is configured to control the cross-over filter in such a way that, for a more transient signal portion, the cross-over filter has a higher impact on a cross-over filter input signal and that the cross-over filter (1406) has a lower impact on the cross-over filter input signal for a less-transient signal portion.

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

wherein a characteristic of the cross-over filter (1406) is defined by a fade-out subfilter characteristic (1420a) and a fade-in subfilter characteristic (1422a),

wherein the fade-in subfilter characteristic $\mathbf{h}_{in}(k)$, and the fade-out subfilter characteristic $\mathbf{h}_{out}(k)$ are defined based on the following equations:

$$\boldsymbol{h}_{out}(k) = \boldsymbol{h}_{in}(N-1-k), \forall\, Xbias$$
$$\boldsymbol{h}_{out}(k) + \boldsymbol{h}_{in}(k) = 1, Xbias = 0$$
$$\boldsymbol{h}_{out}(k) = 0.5 + 0.5 \cdot cos\left(\frac{k}{N-1-Xbias} \cdot \pi\right), k = 0, 1, \dots, N-1-Xbias,$$

wherein *Xbias* is an integer defining a slope of both filters extending between zero and an integer *N*, wherein *k* is a frequency index extending between zero and *N*-1, and wherein *N* is an additional integer, and wherein different values for *N* and *Xbias* result in different cross-over filter characteristics.

11. Apparatus of claim 10,

wherein *Xbias* is set between 2 and 20 and wherein *N* is set between 10 and 50.

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

wherein the tile generator (1404) is configured to generate a preliminary frequency tile (703), wherein an analyzer (702) is configured for analyzing the preliminary frequency tile, wherein the tile generator is additionally configured for generating a regenerated signal having attenuated or eliminated artifact creating tonal portions in relation to the preliminary frequency tile, wherein the file generator is configured

to eliminate or attenuate tonal components near frequency tile borders (708) to obtain an input signal into the cross-over filter (1406).

13. Apparatus of claim 12, wherein the tile generator is configured to detect and remove or attenuate tonal spectral portions within a detection range being less than 20% of a bandwidth of a frequency tile or a source range for the regeneration.

14. Apparatus of one of the preceding claims, wherein the cross-over filter (1406) is configured to cross-over filter within an overlapping range, the overlapping range comprising an upper freqeuncy portion of the decoded core signal and a lower frequency portion of the first frequency tile, or

wherein the cross-over filter (1406) is configured to cross-over filter within an overlapping range, the overlapping range comprising an upper frequency portion of a first frequency tile and a lower frequency portion of a second frequency tile

15. Method of decoding an encoded audio signal comprising an encoded core signal (1), comprising:

decoding (1400) the encoded core signal (1401) to obtain a decoded core signal;

generating (1404) one or more spectral tiles having frequencies not included in the decoded core signal using a spectral portion of the decoded core signal; and

spectrally cross-over filtering (1406) the decoded core signal and a first frequency tile having frequencies extending from a gap filling frequency (309) to an upper border frequency or for spectrally cross-over filtering a first frequency tile and a second frequency tile.

16. Computer program for performing, when running on a computer or a processor, the method of claim 15.

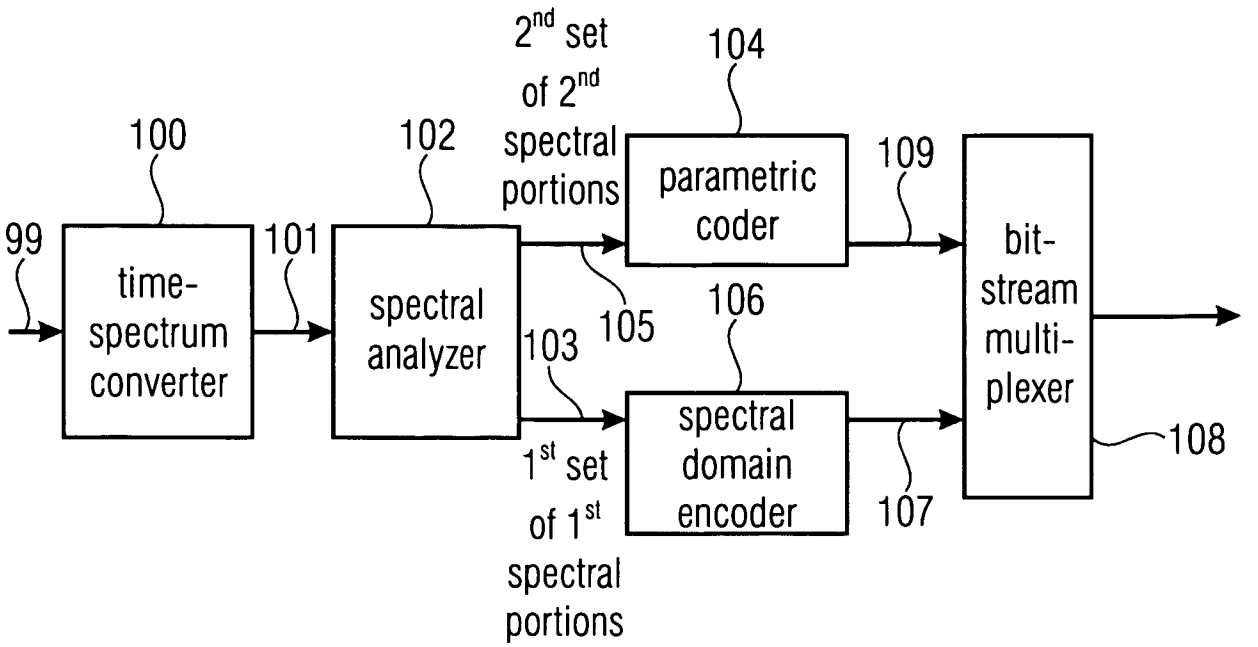

99
100
time-spectrum converter
101
102
spectral analyzer
2nd set of 2nd spectral portions
104
parametric coder
109
bit-stream multi-plexer
108
105
103
106
spectral domain encoder
107
1st set of 1st spectral portions

FIG 1A

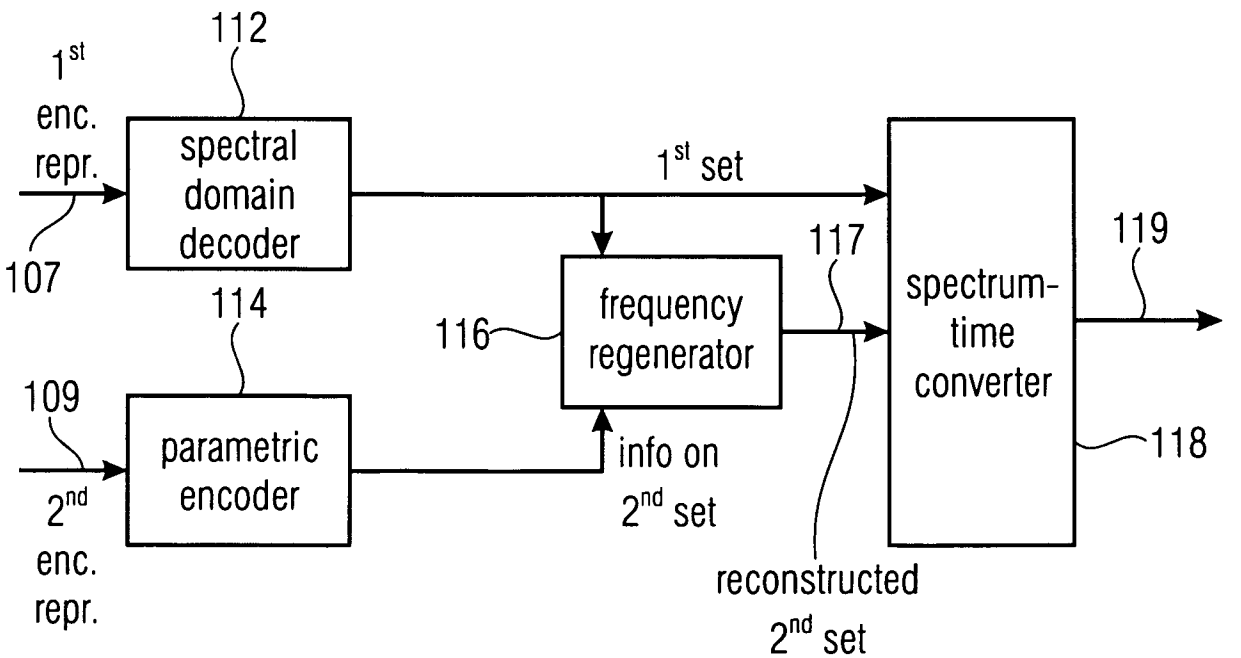

1st enc. repr.
107
112
spectral domain decoder
1st set
117
119
spectrum-time converter
116
frequency regenerator
114
109
parametric encoder
info on 2nd set
118
2nd enc. repr.
reconstructed 2nd set

FIG 1B

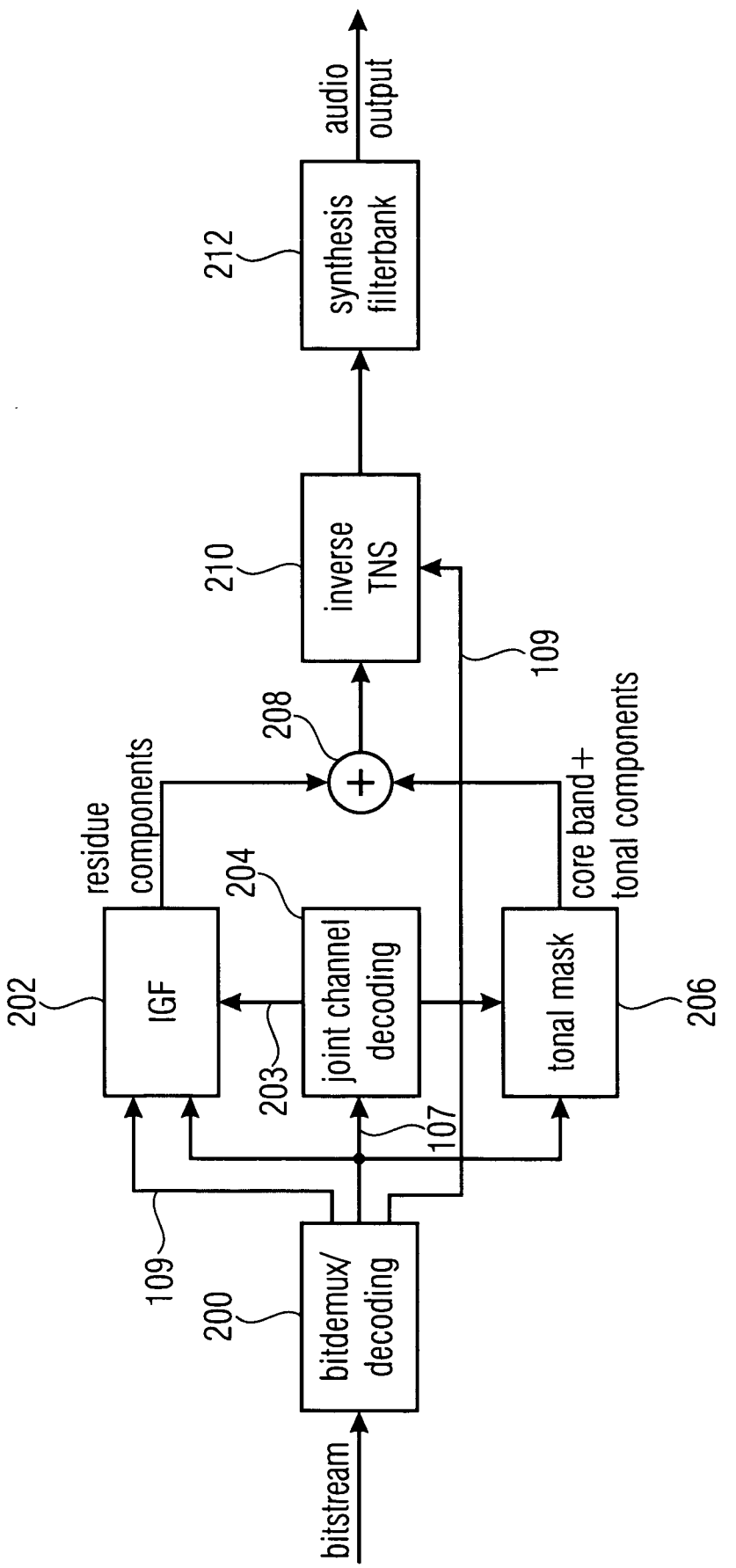

bitstream
bitdemux/ decoding
200
109
IGF
202
residue components
joint channel decoding
204
203
107
tonal mask
206
core band + tonal components
208
inverse TNS
210
109
synthesis filterbank
212
audio output

FIG 2A

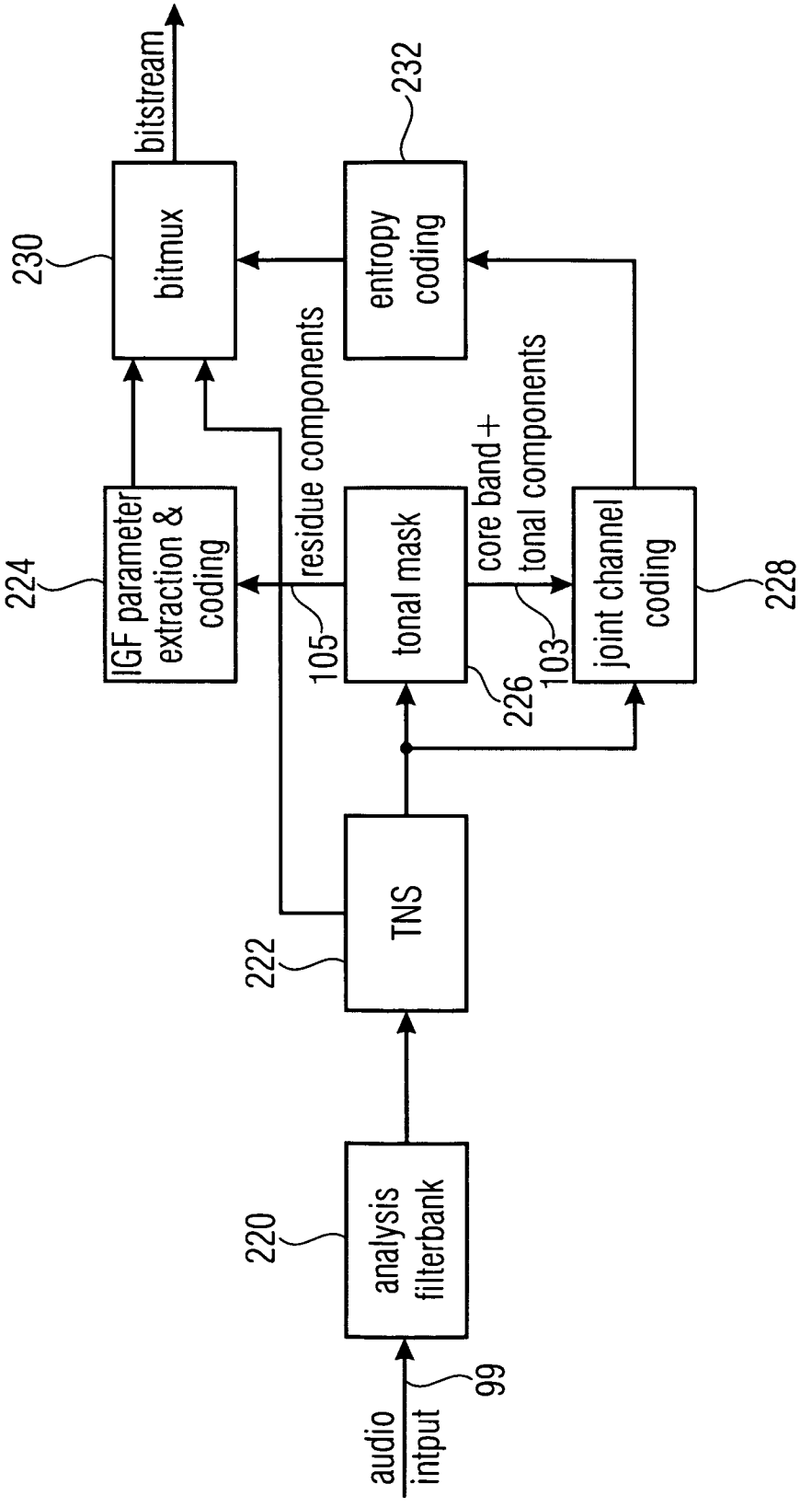

audio intput
99
analysis filterbank
220
TNS
222
IGF parameter extraction & coding
224
bitmux
230
bitstream
entropy coding
232
residue components
105
tonal mask
226
core band + tonal components
103
joint channel coding
228

FIG 2B

- 1st resolution (high resolution) for „envelope“ of the 1st set (line-wise coding);
- 2nd resolution (low resolution) for „envelope“ of the 2nd set (scale factor per SCB);

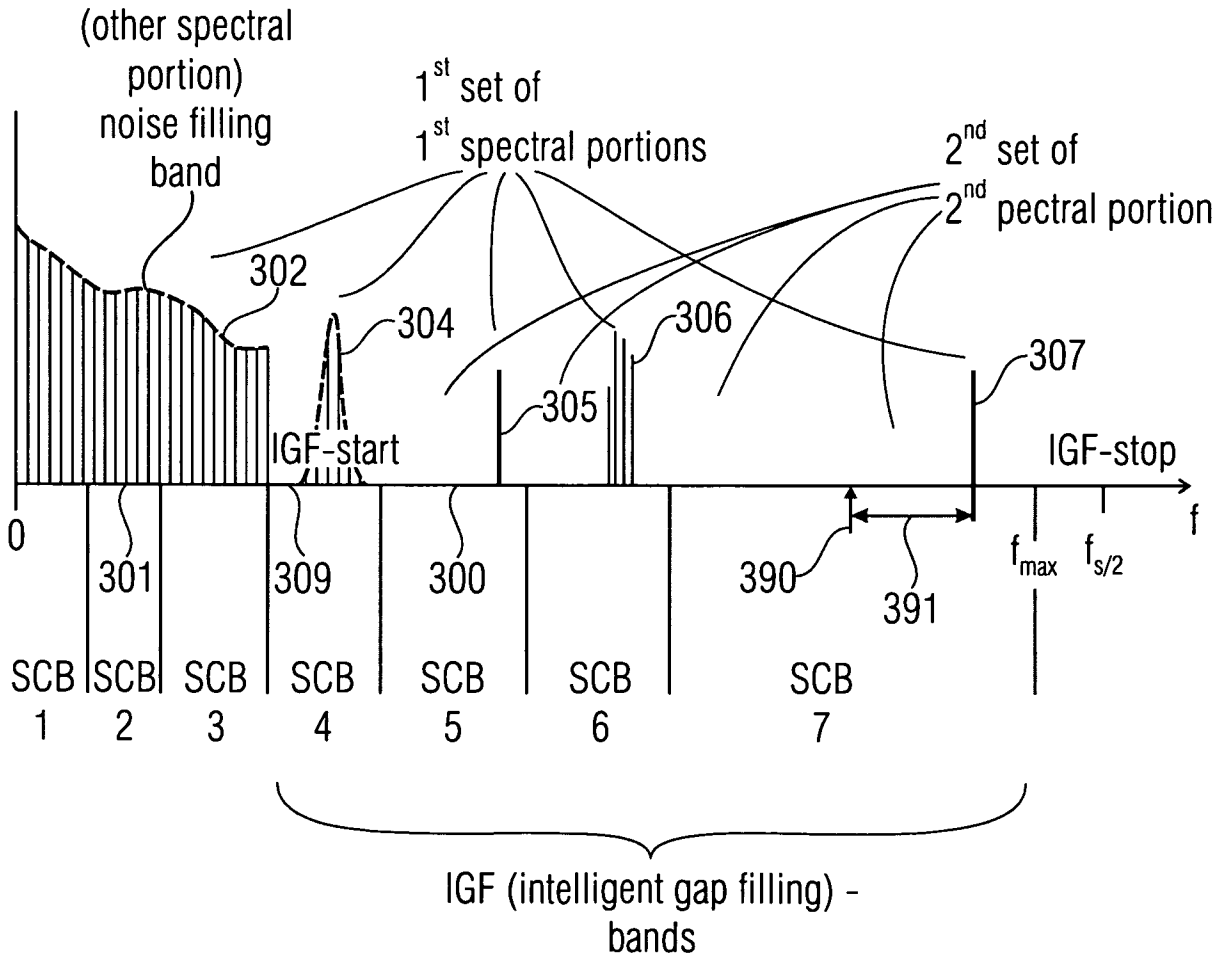


FIG 3A

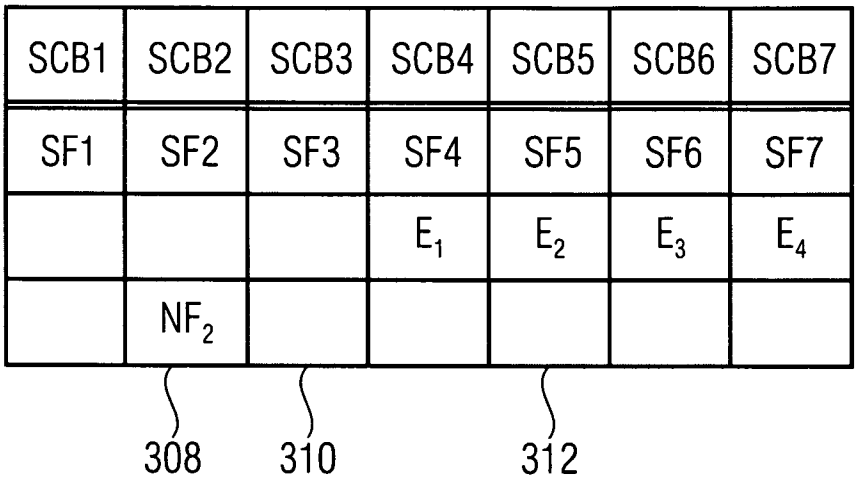

SCB1
SCB2
SCB3
SCB4
SCB5
SCB6
SCB7
SF1
SF2
SF3
SF4
SF5
SF6
SF7
E1
E2
E3
E4
NF2
308
310
312

FIG 3B

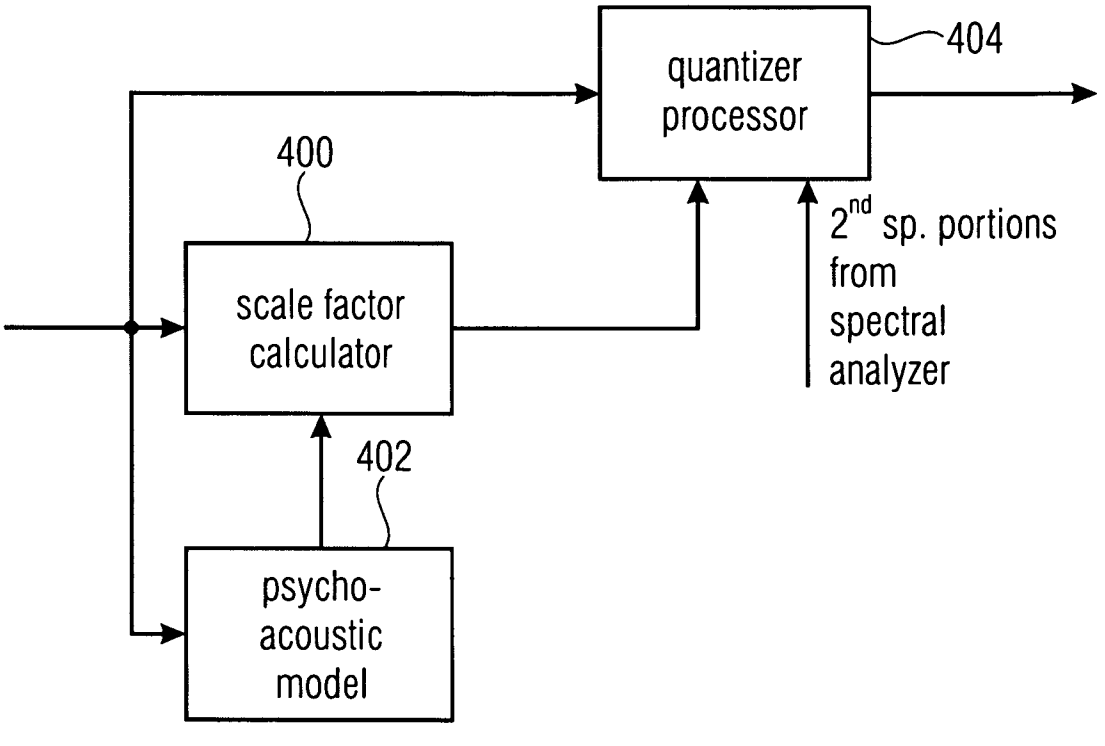

quantizer processor
404
400
scale factor calculator
2nd sp. portions from spectral analyzer
402
psycho-acoustic model

FIG 4A

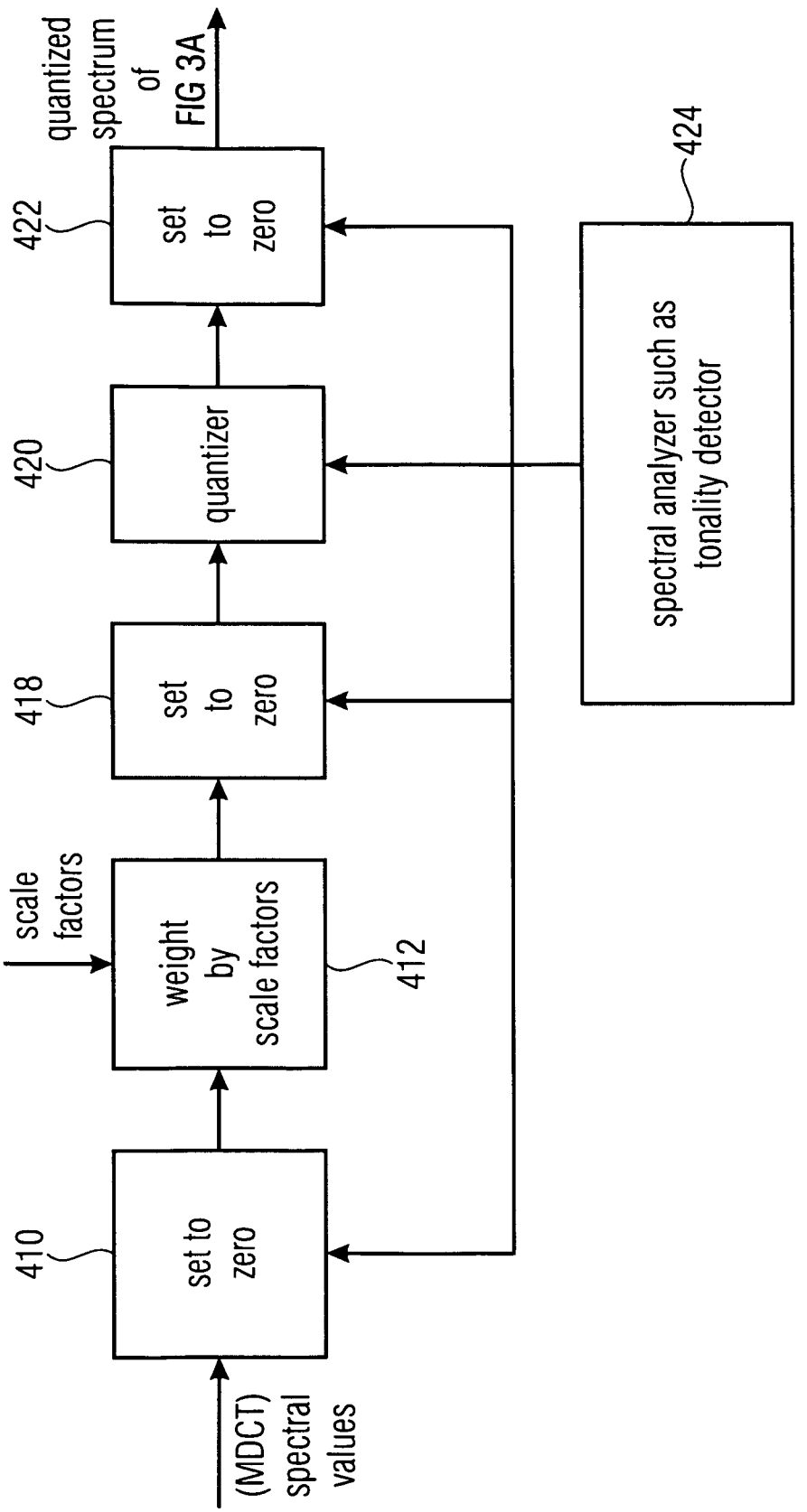


FIG 4B
(QUANTIZER PROCESSOR)

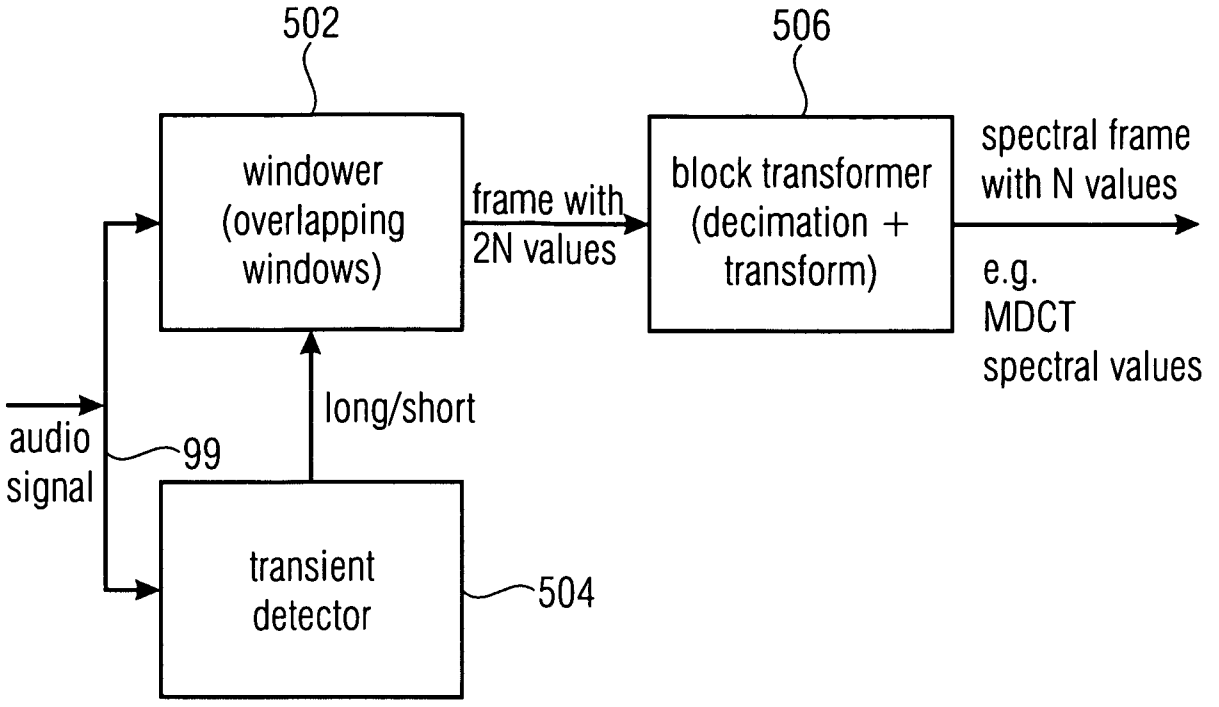


FIG 5A
(OTHER SPECTRAL PORTIONS)

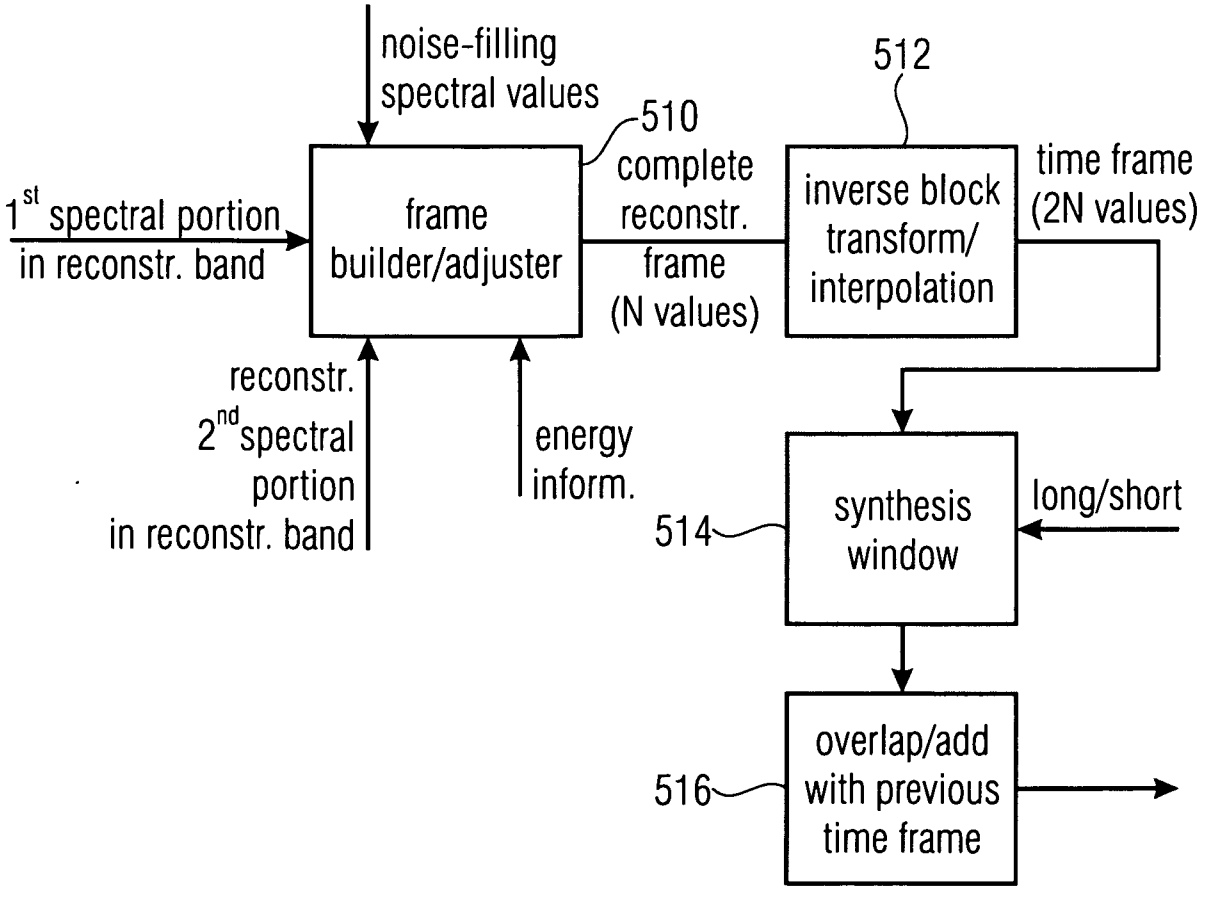


FIG 5B

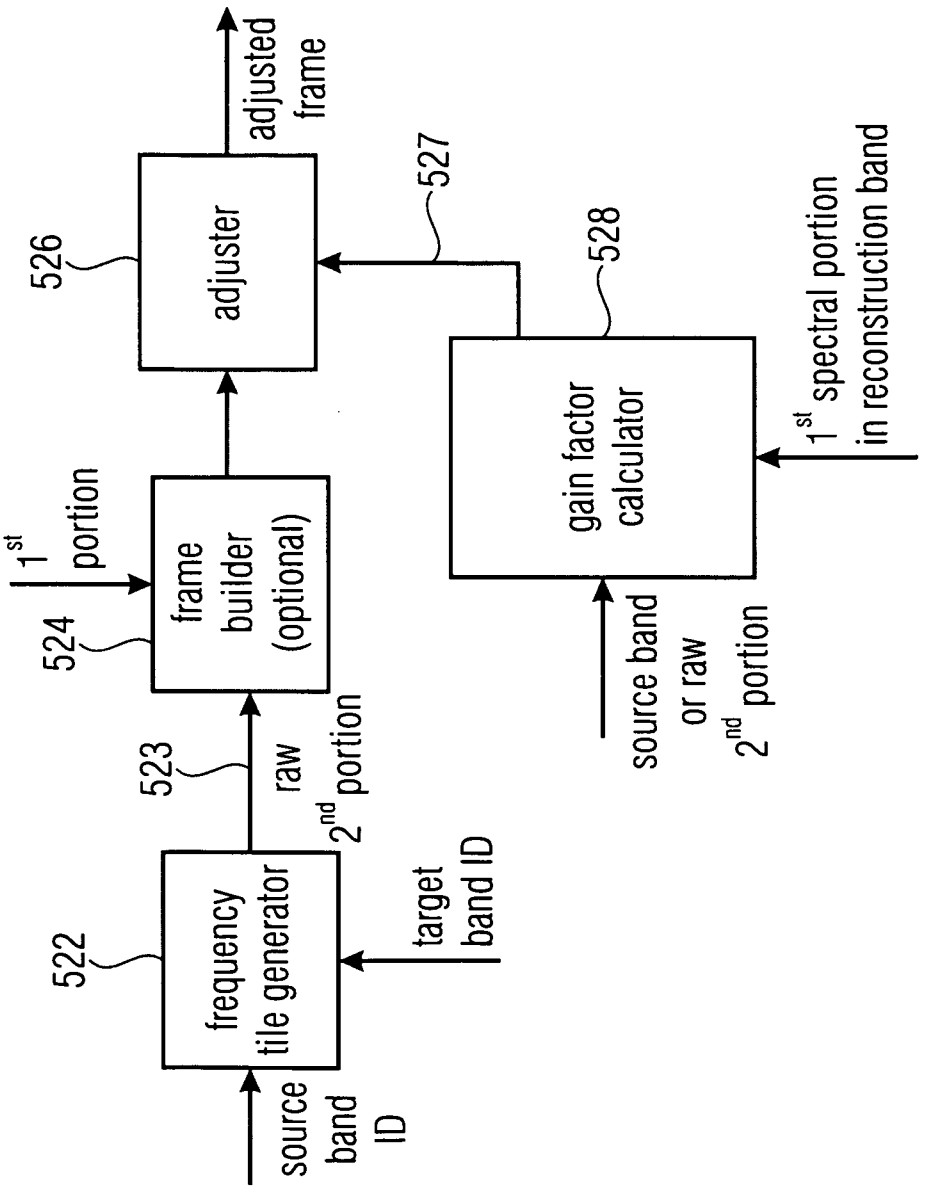

source band ID
frequency tile generator
522
target band ID
raw 2nd portion
523
frame builder (optional)
524
1st portion
adjuster
526
adjusted frame
527
gain factor calculator
528
source band or raw 2nd portion
1st spectral portion in reconstruction band

FIG 5C

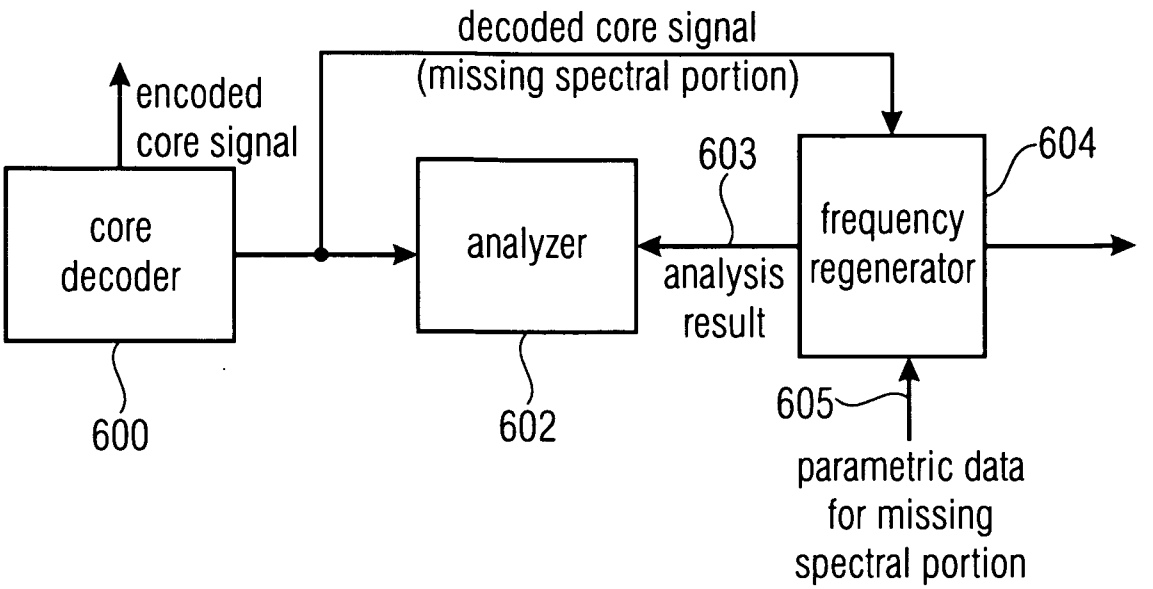

decoded core signal
(missing spectral portion)
encoded
core signal
core
decoder
analyzer
603
analysis
result
frequency
regenerator
604
600
602
605
parametric data
for missing
spectral portion

FIG 6A

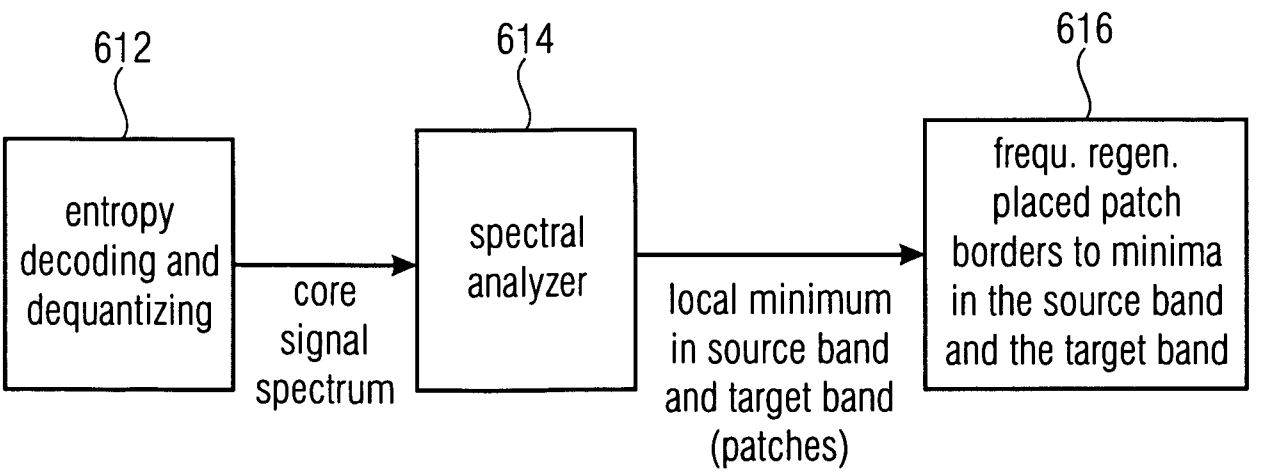

612
614
616
entropy
decoding and
dequantizing
core
signal
spectrum
spectral
analyzer
local minimum
in source band
and target band
(patches)
frequ. regen.
placed patch
borders to minima
in the source band
and the target band

FIG 6B

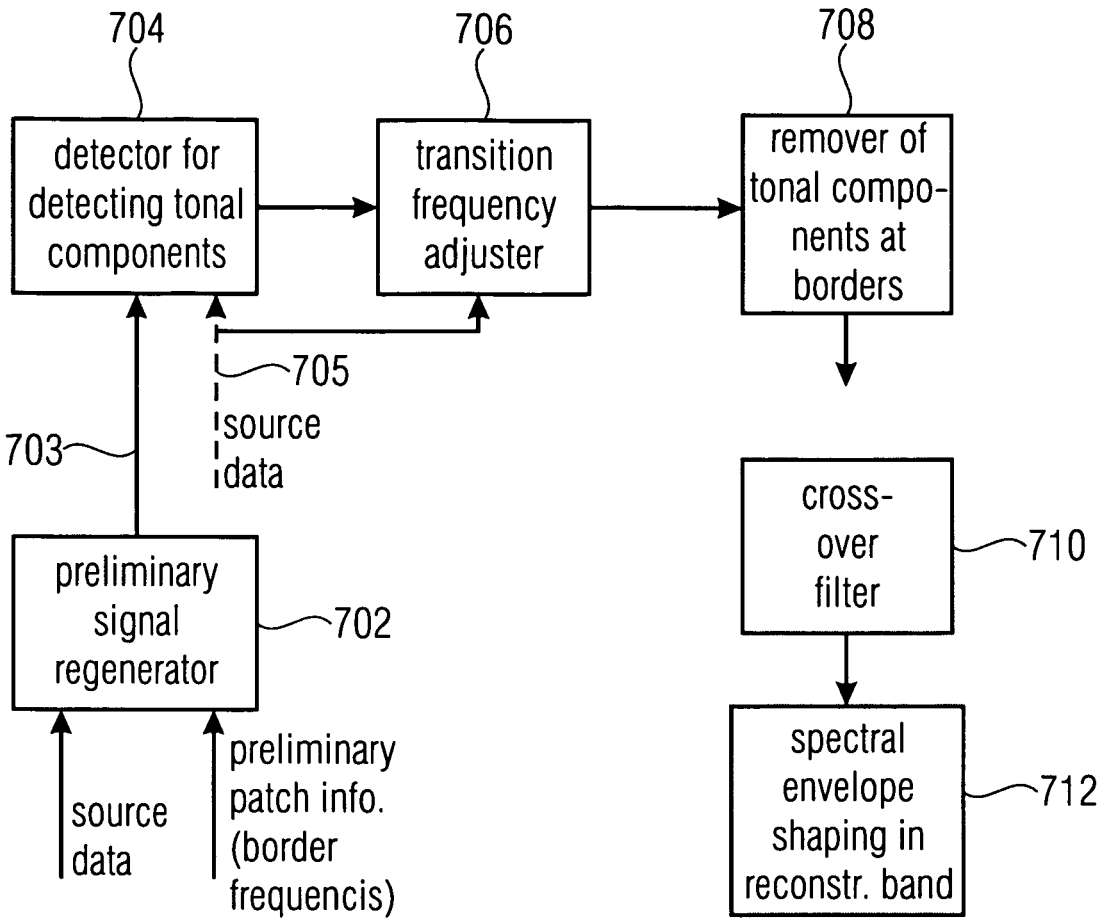

704
706
708
detector for detecting tonal components
transition frequency adjuster
remover of tonal compo-nents at borders
705
source data
703
cross-over filter
710
preliminary signal regenerator
702
source data
preliminary patch info. (border frequencis)
spectral envelope shaping in reconstr. band
712

FIG 7A

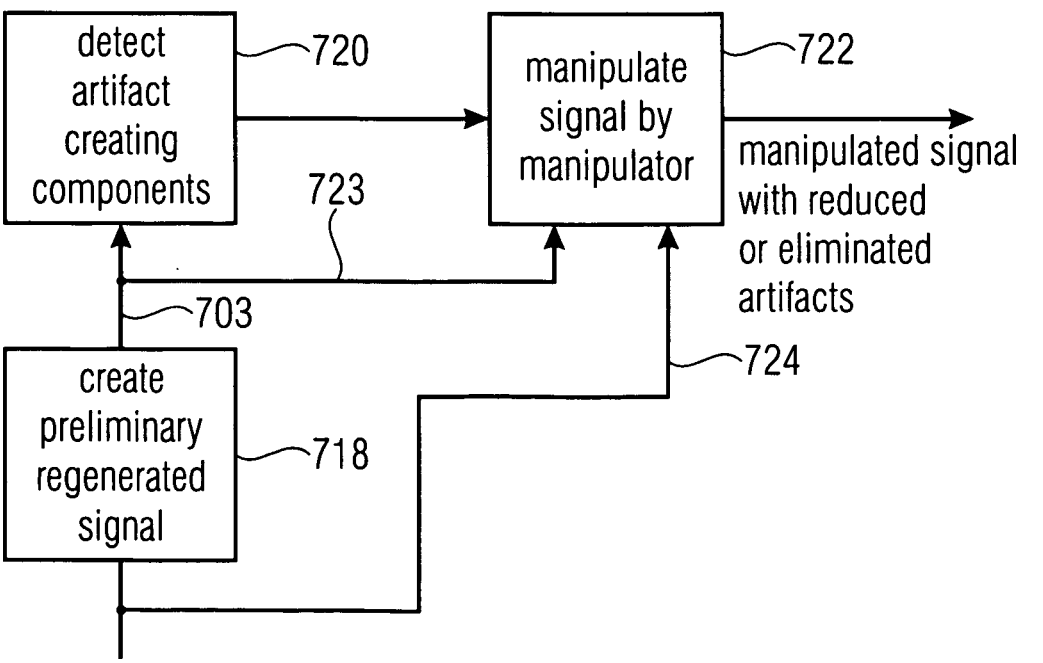

detect artifact creating components
720
manipulate signal by manipulator
722
manipulated signal with reduced or eliminated artifacts
723
703
724
create preliminary regenerated signal
718

FIG 7B

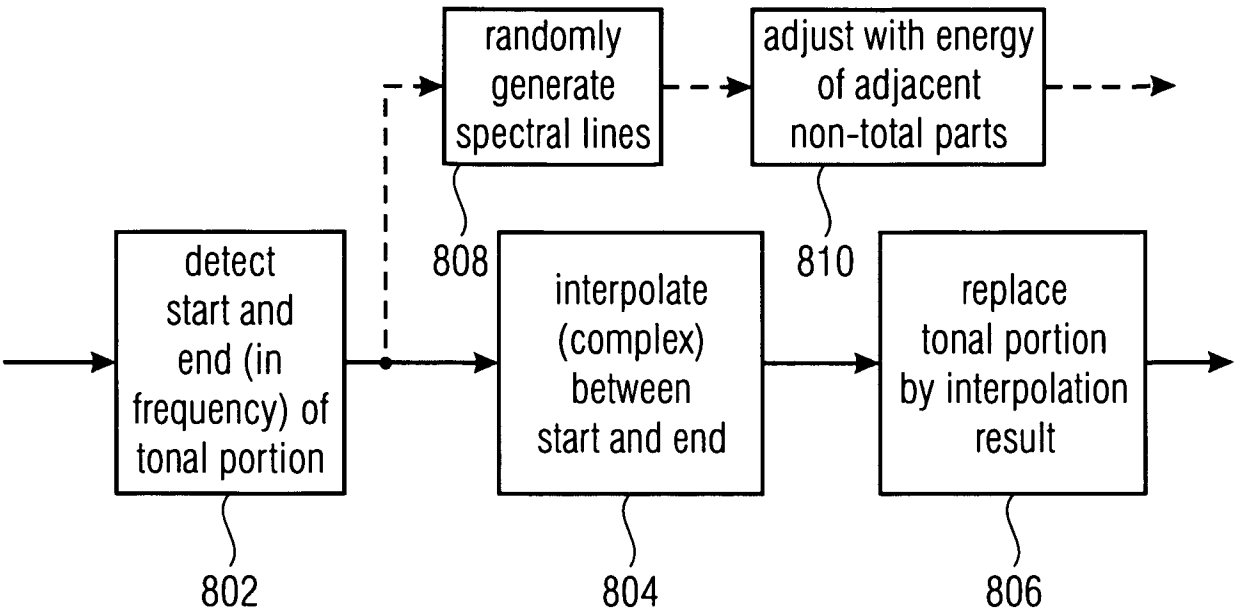

randomly generate spectral lines
adjust with energy of adjacent non-total parts
detect start and end (in frequency) of tonal portion
interpolate (complex) between start and end
replace tonal portion by interpolation result
808
810
802
804
806

FIG 8A

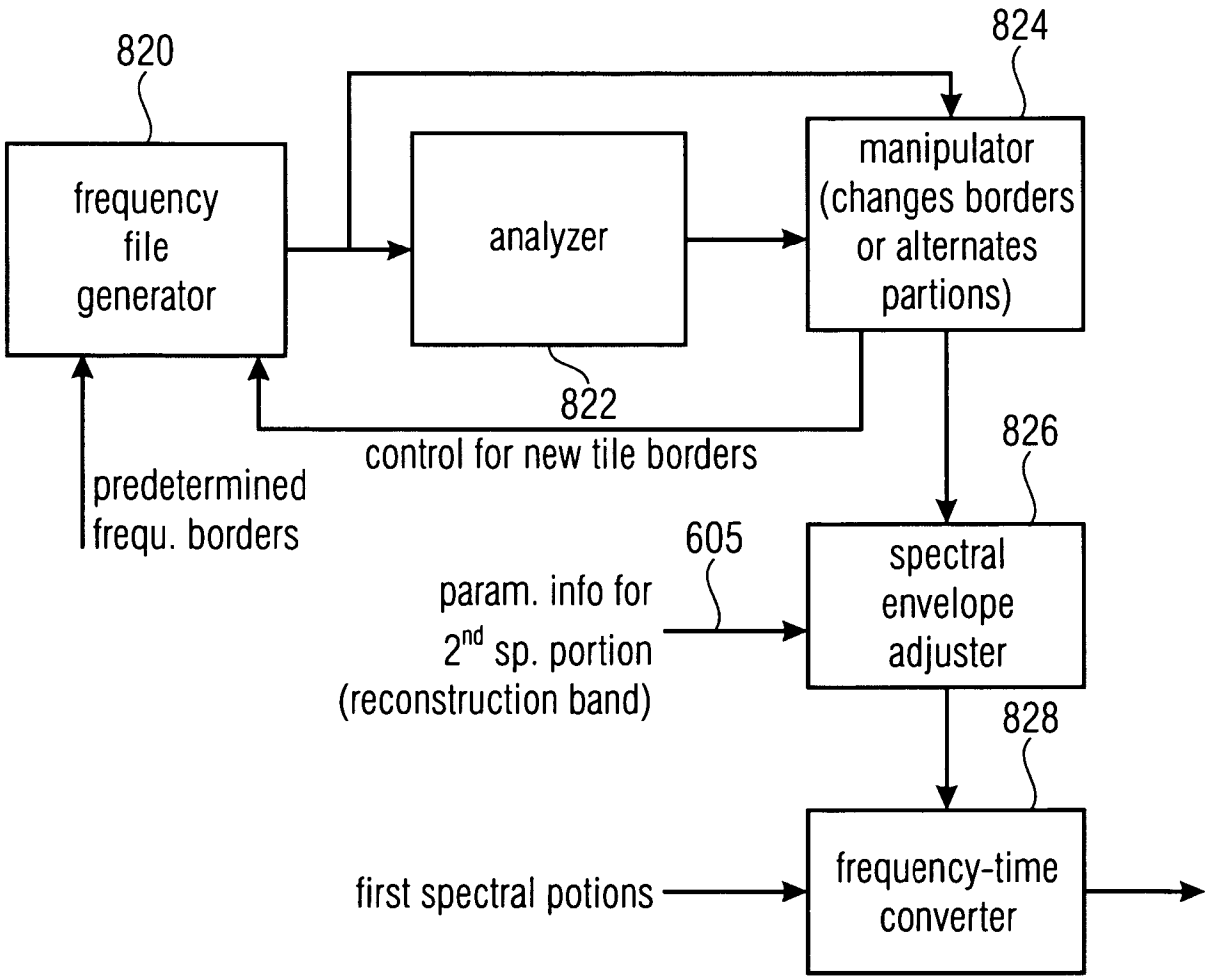

820
824
frequency file generator
analyzer
manipulator (changes borders or alternates partions)
822
control for new tile borders
predetermined frequ. borders
826
605
param. info for 2nd sp. portion (reconstruction band)
spectral envelope adjuster
828
first spectral potions
frequency-time converter

FIG 8B

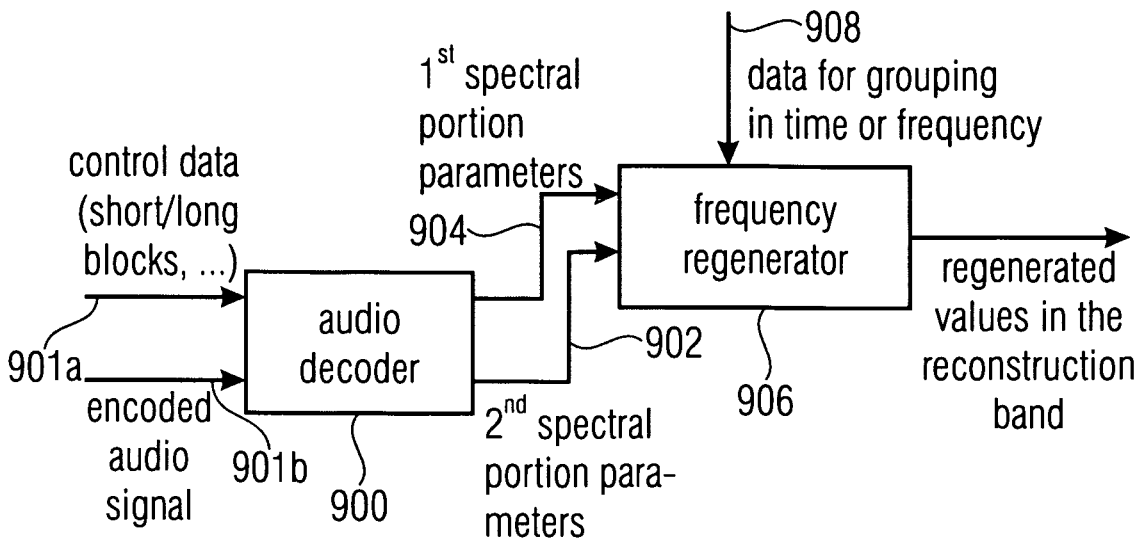

908
data for grouping in time or frequency
1st spectral portion parameters
904
control data (short/long blocks, ...)
audio decoder
frequency regenerator
regenerated values in the reconstruction band
902
906
901a
encoded audio signal
901b
900
2nd spectral portion parameters

FIG 9A

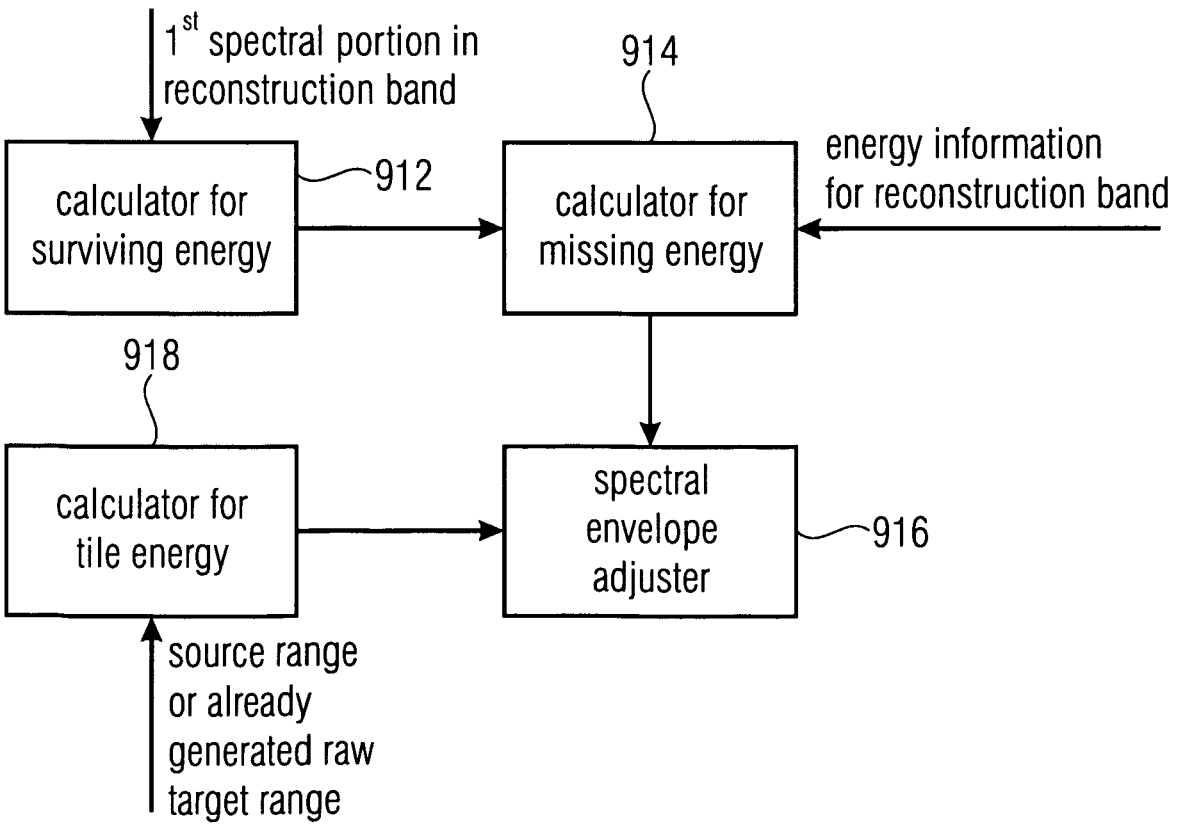
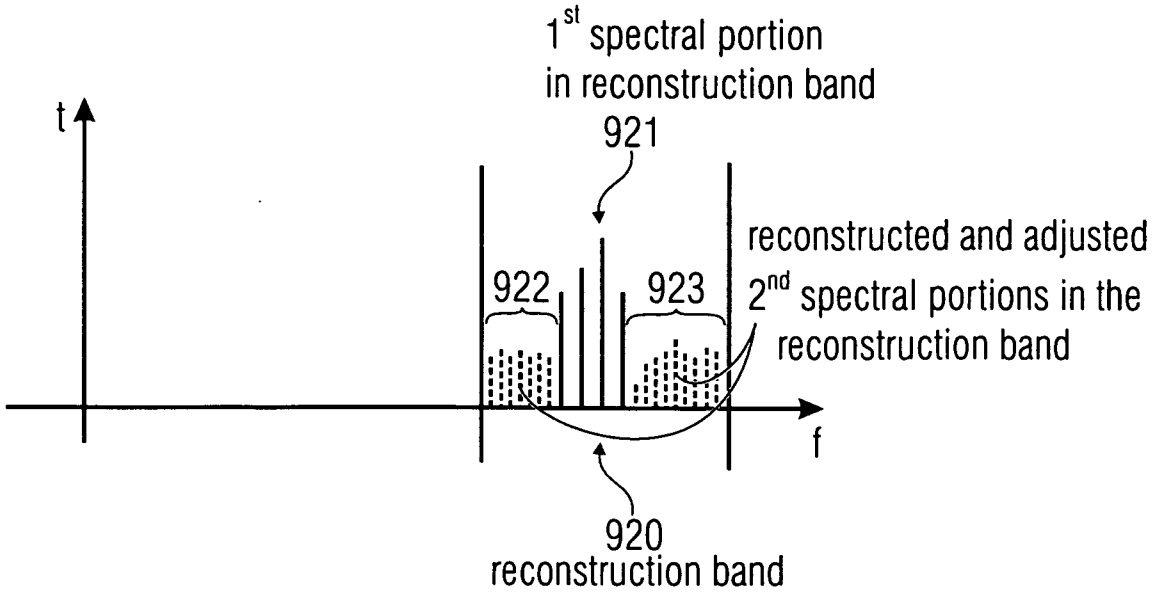

1st spectral portion in reconstruction band
914
calculator for surviving energy
912
calculator for missing energy
energy information for reconstruction band
918
calculator for tile energy
spectral envelope adjuster
916
source range or already generated raw target range

FIG 9B



e.g.

- surviving energy: 5 units
- energy value for reconstr. band: 10 units (covers 1st and 2nd spectral portions in the reconstruction band)
- energy of source range data or raw target range data: 8 units
- missing energy: 5 units
- gain factor: $g := \sqrt{\frac{mE_k}{pE_k}} = 0.79$

→ only spectral values for the 2nd spectral portions are adjusted

→ 1st spectral portion is not influenced by the envelope adjustment

FIG 9C

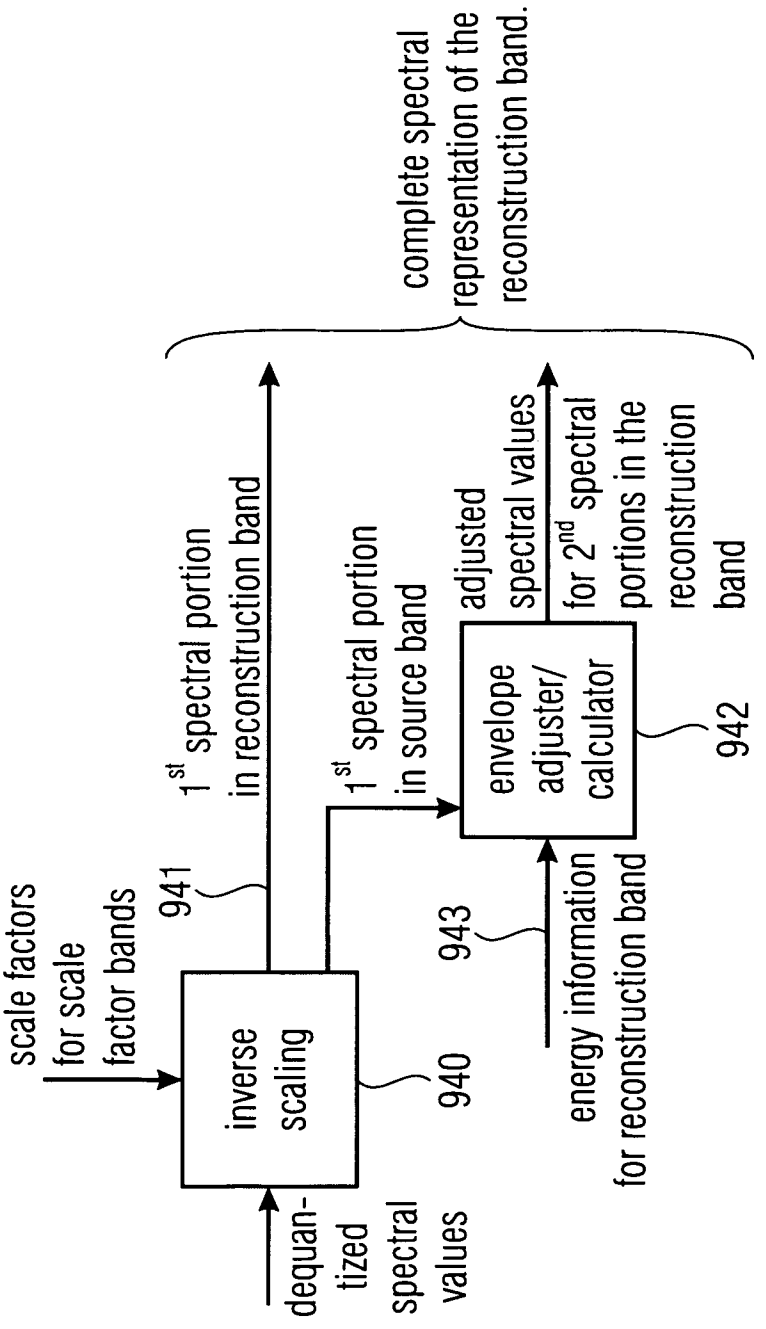

scale factors for scale factor bands
dequan-tized spectral values
inverse scaling
940
941
1st spectral portion in reconstruction band
1st spectral portion in source band
943
energy information for reconstruction band
envelope adjuster/ calculator
942
adjusted spectral values for 2nd spectral portions in the reconstruction band
complete spectral representation of the reconstruction band.

FIG 9D

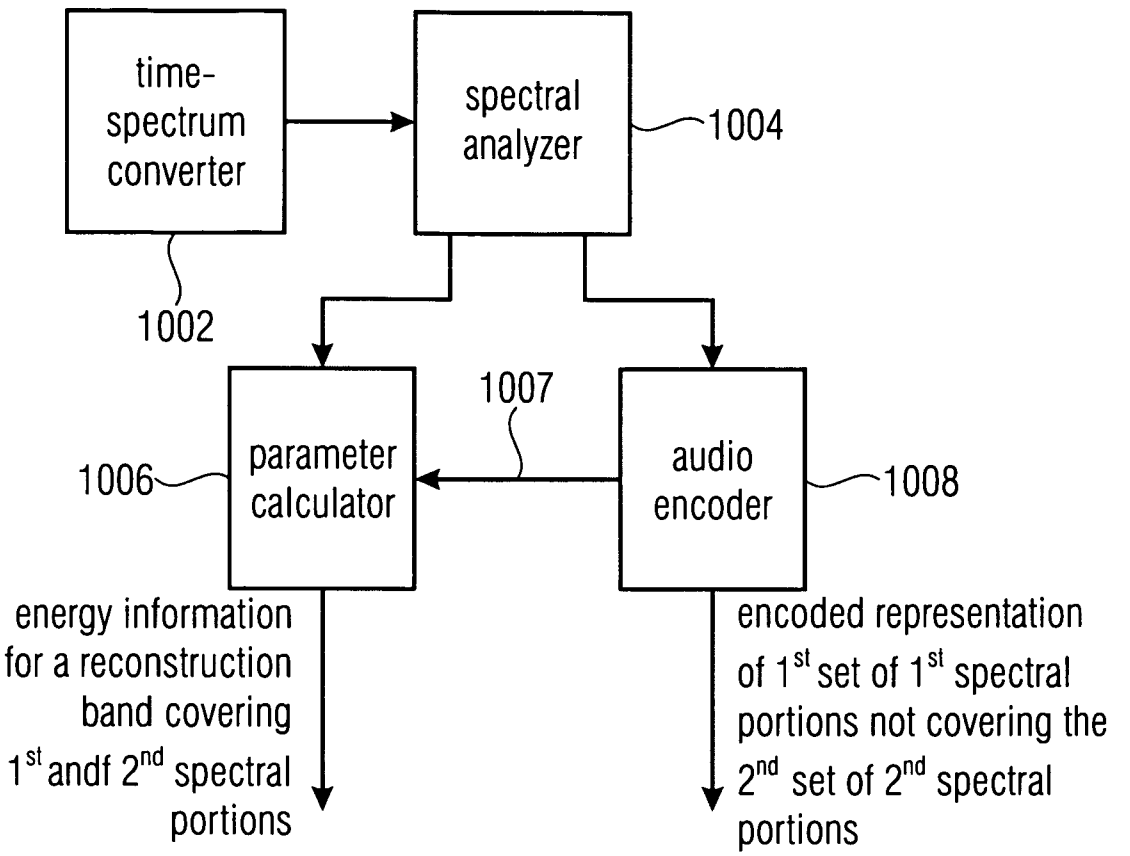


FIG 10A

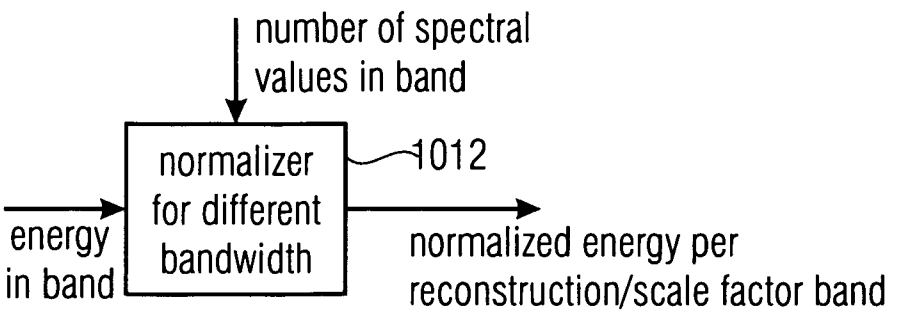


FIG 10B

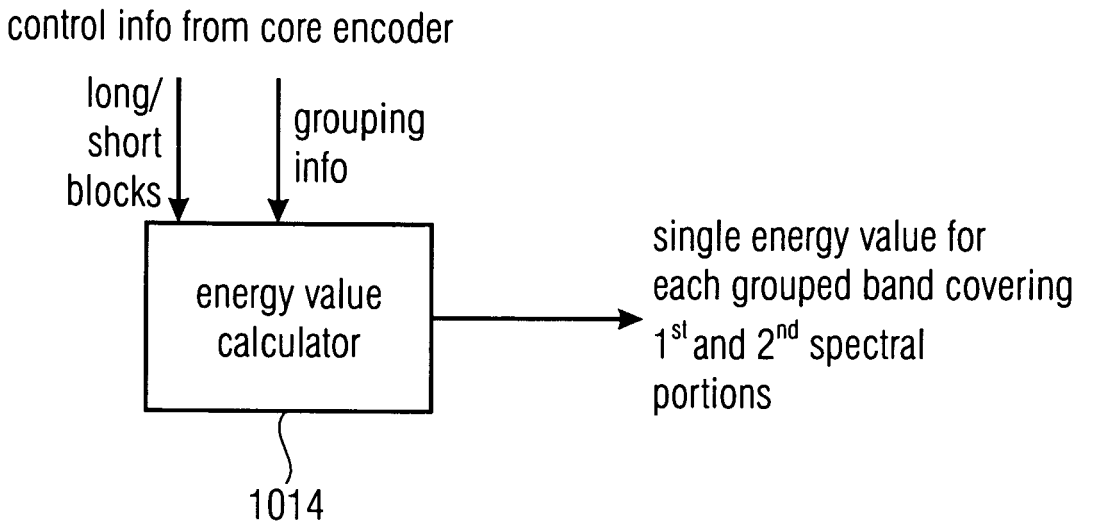


FIG 10C

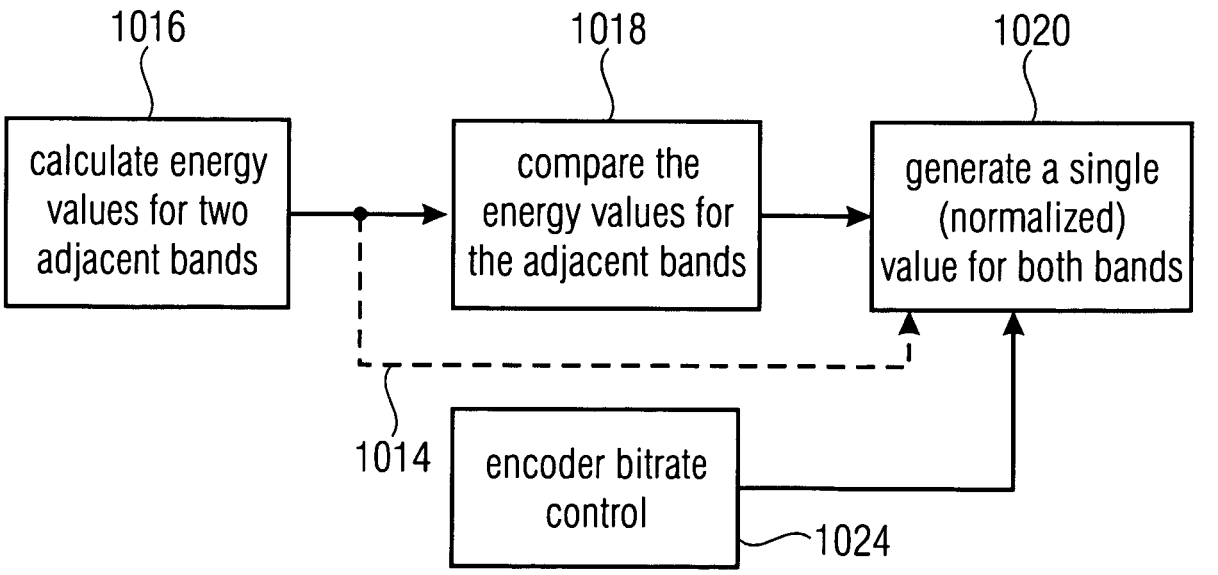


FIG 10D

1100

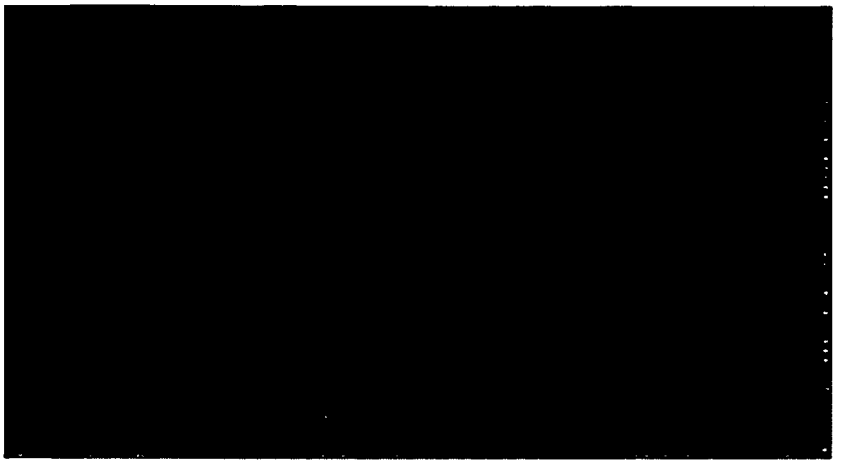
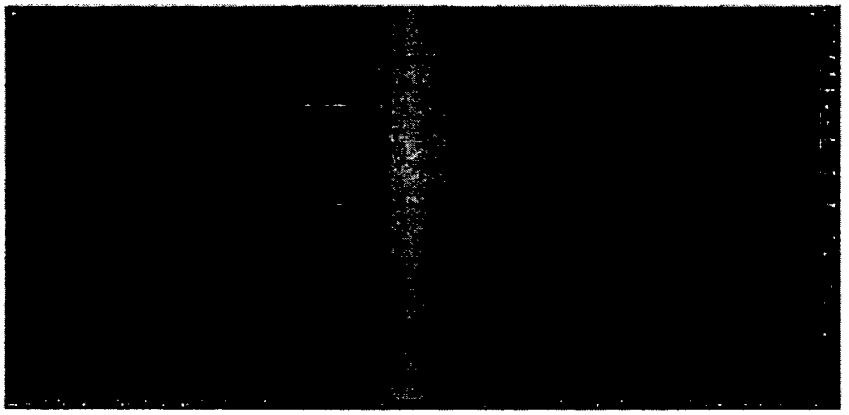
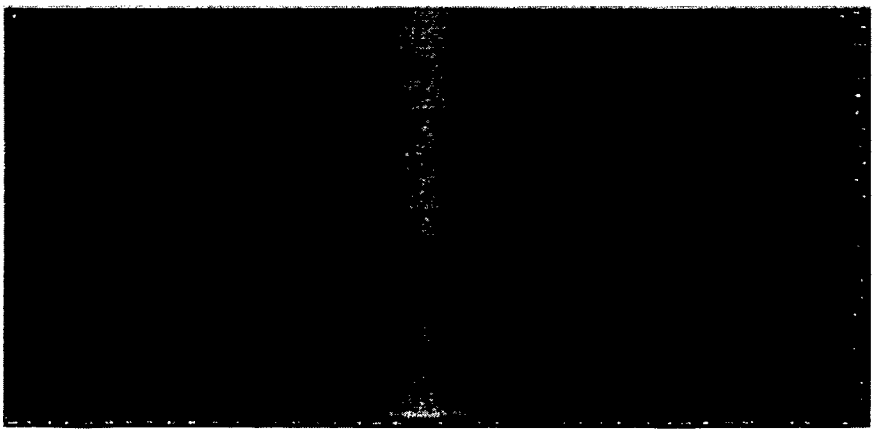


FIG 11A



Spectrogram of a transient after applying BWE.
The x-axis represents time, the y-axis frequency.

FIG 11B



Spectrogram of a transient after applying BWE. The x-axis represents time, the y-axis frequency. Through the application of filter ringing reduction, the filter ringing is reduced by approx. 20dB.

FIG 11C

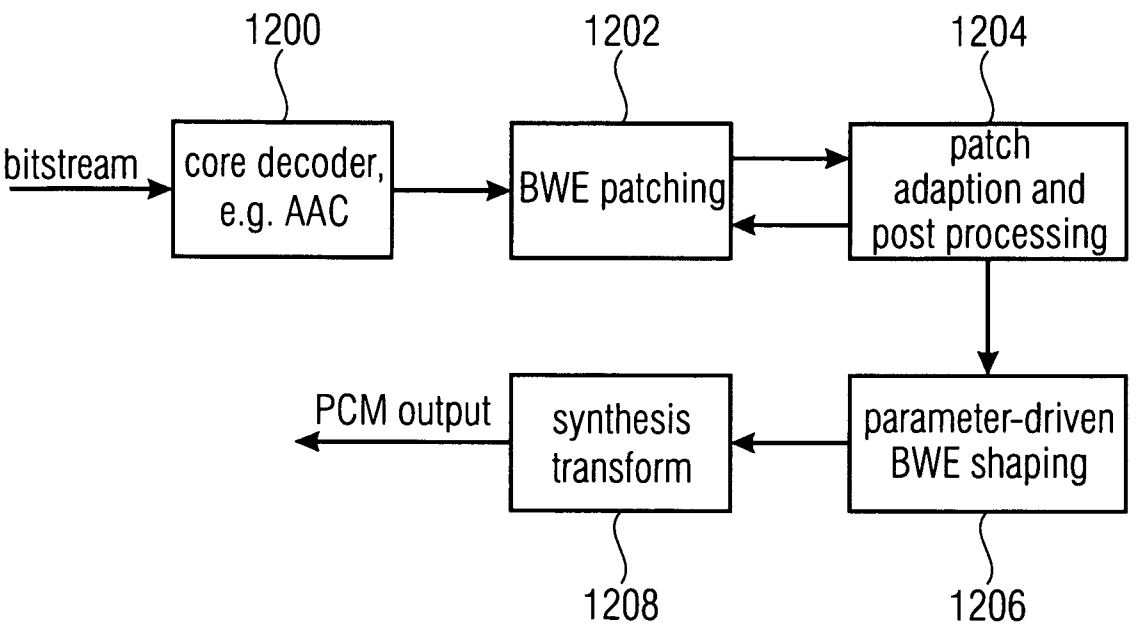

1200
1202
1204
bitstream
core decoder, e.g. AAC
BWE patching
patch adaption and post processing
PCM output
synthesis transform
parameter-driven BWE shaping
1208
1206

FIG 12A

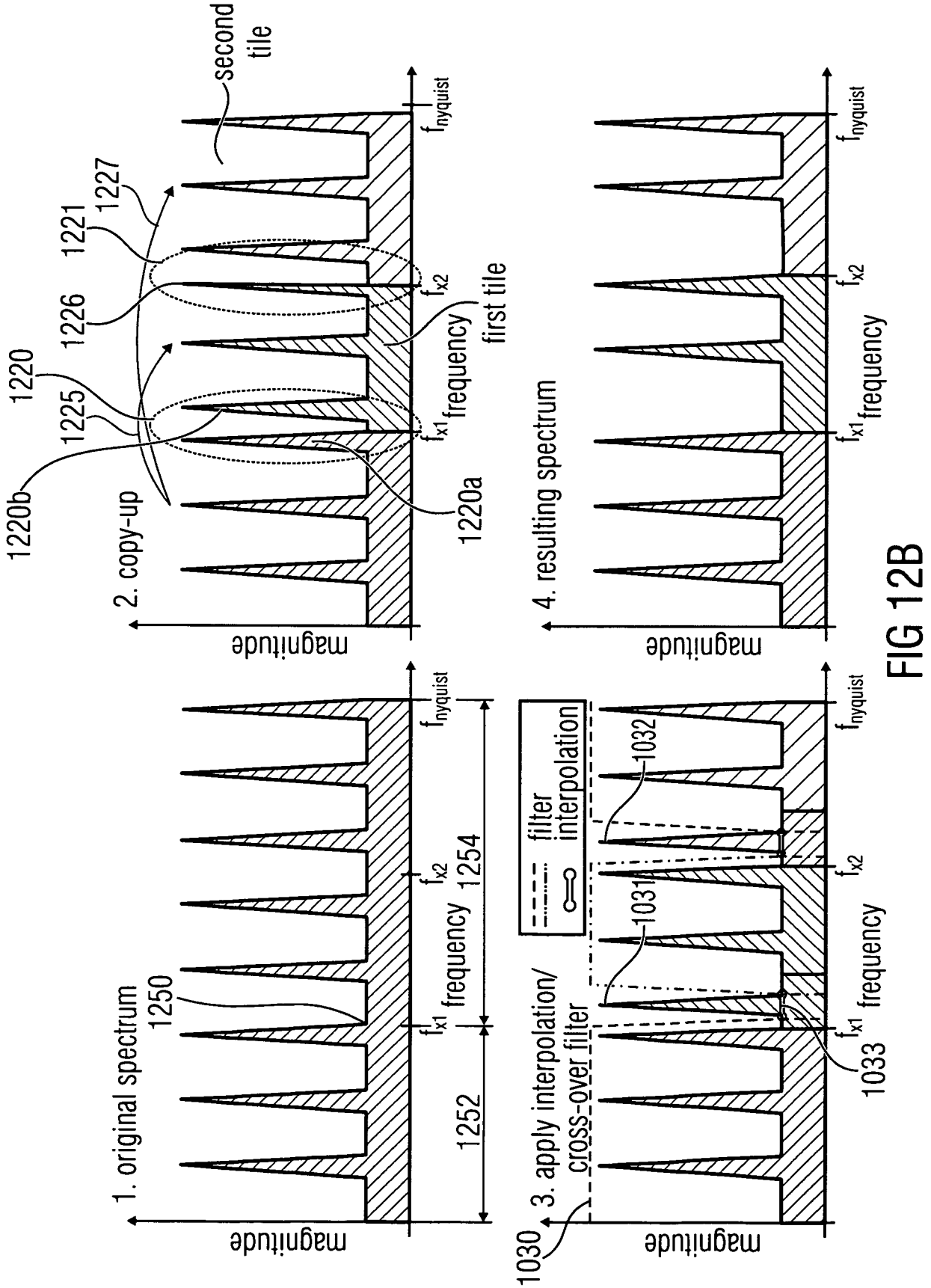

1. original spectrum
1250
magnitude
$f_{nyquist}$
f_{x1} frequency
f_{x2}
1254
1252
2. copy-up
1220b
1220
1225
1226
1221
1227
second tile
first tile
1220a
3. apply interpolation/ cross-over filter
1030
filter
interpolation
1031
1032
1033
4. resulting spectrum

FIG 12B

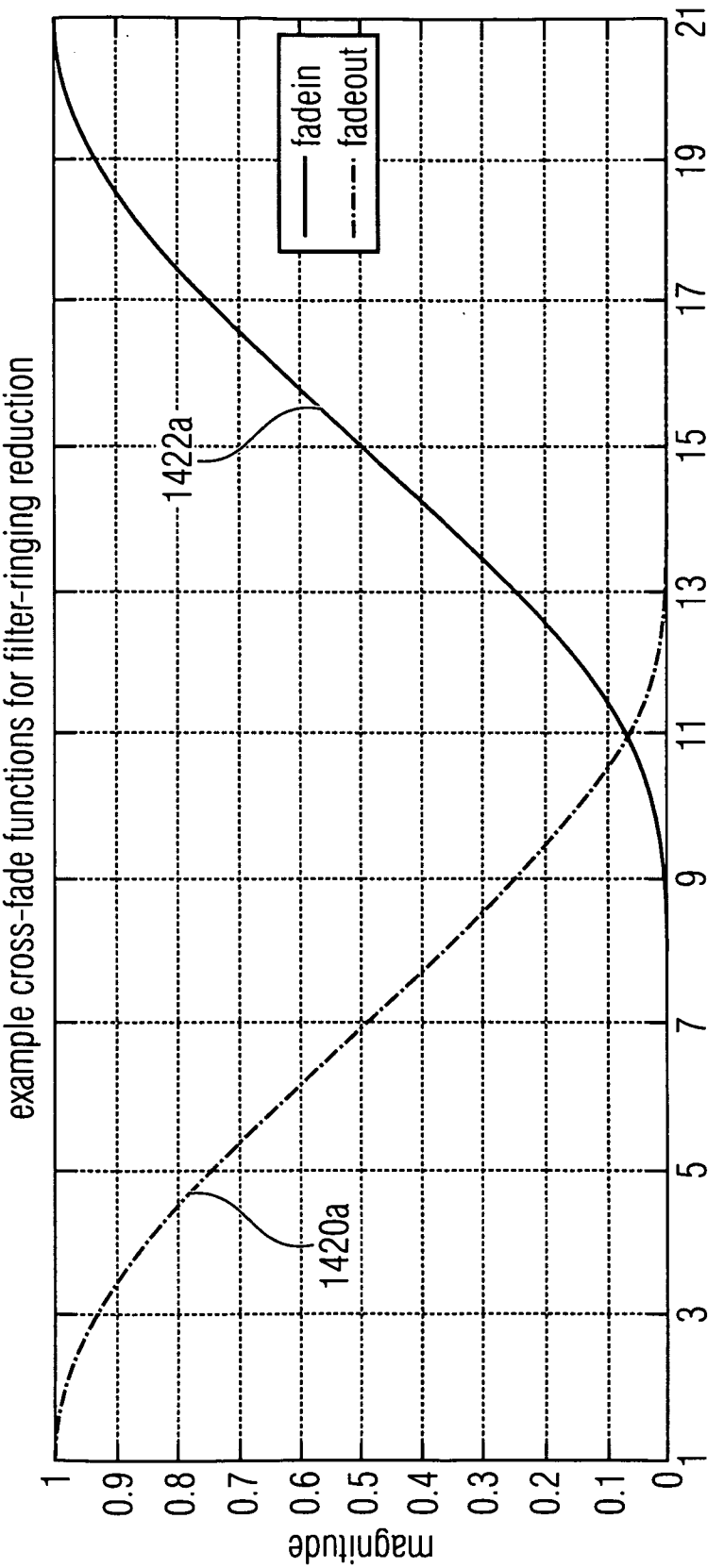

example cross-fade functions for filter-ringing reduction
fadein
fadeout
1422a
1420a
magnitude
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
3
5
7
9
11
13
15
17
19
21
samples [bin]
example cross-fade functions, N = 21, Xbias = 8.

FIG 12C

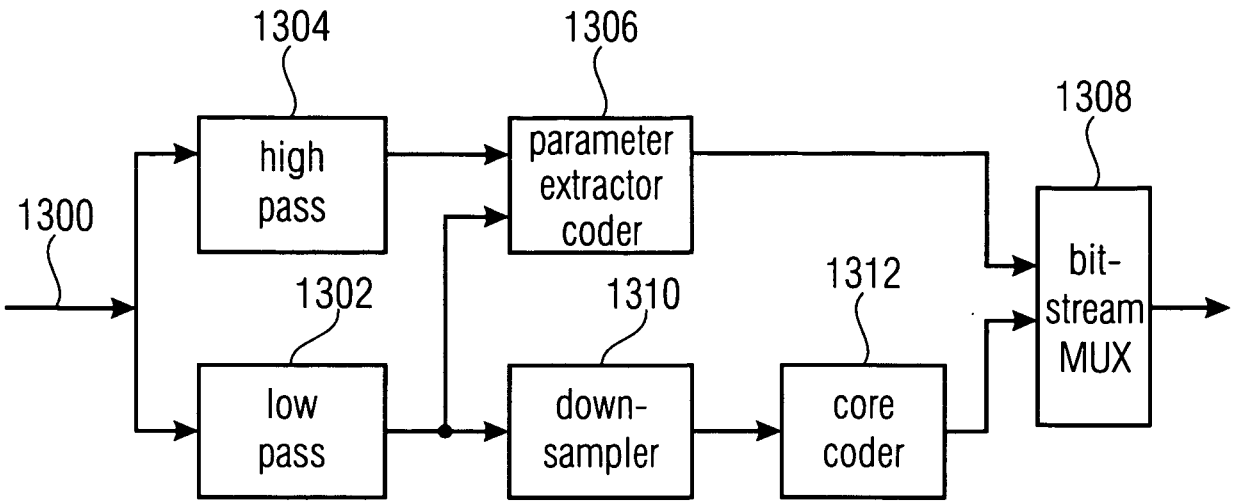


FIG 13A
(PRIOR ART)

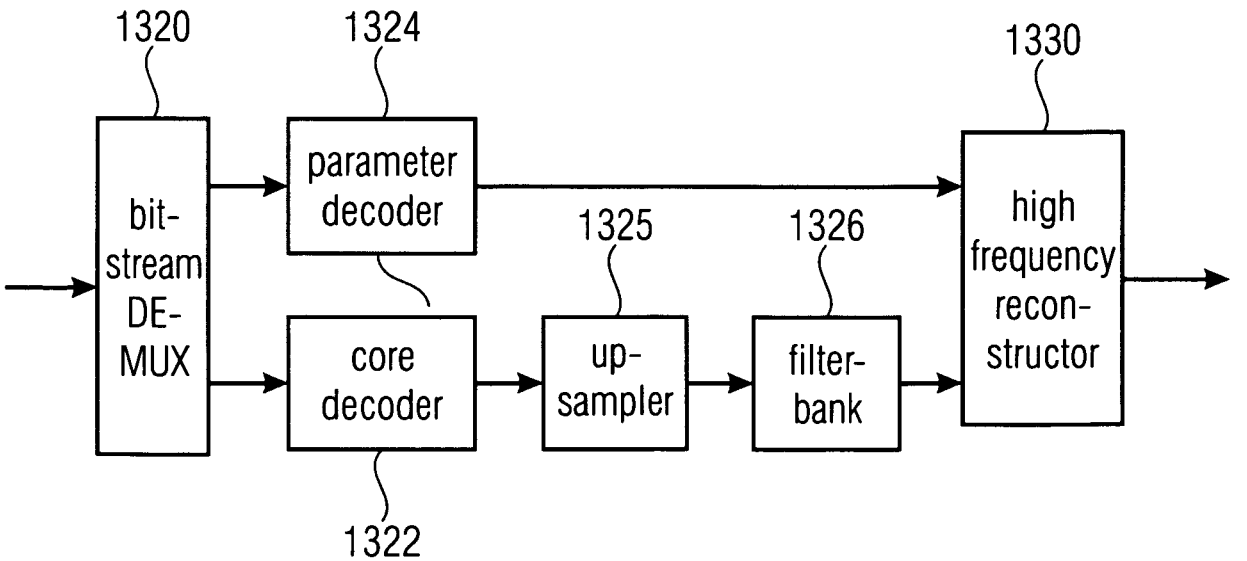


FIG 13B
(PRIOR ART)

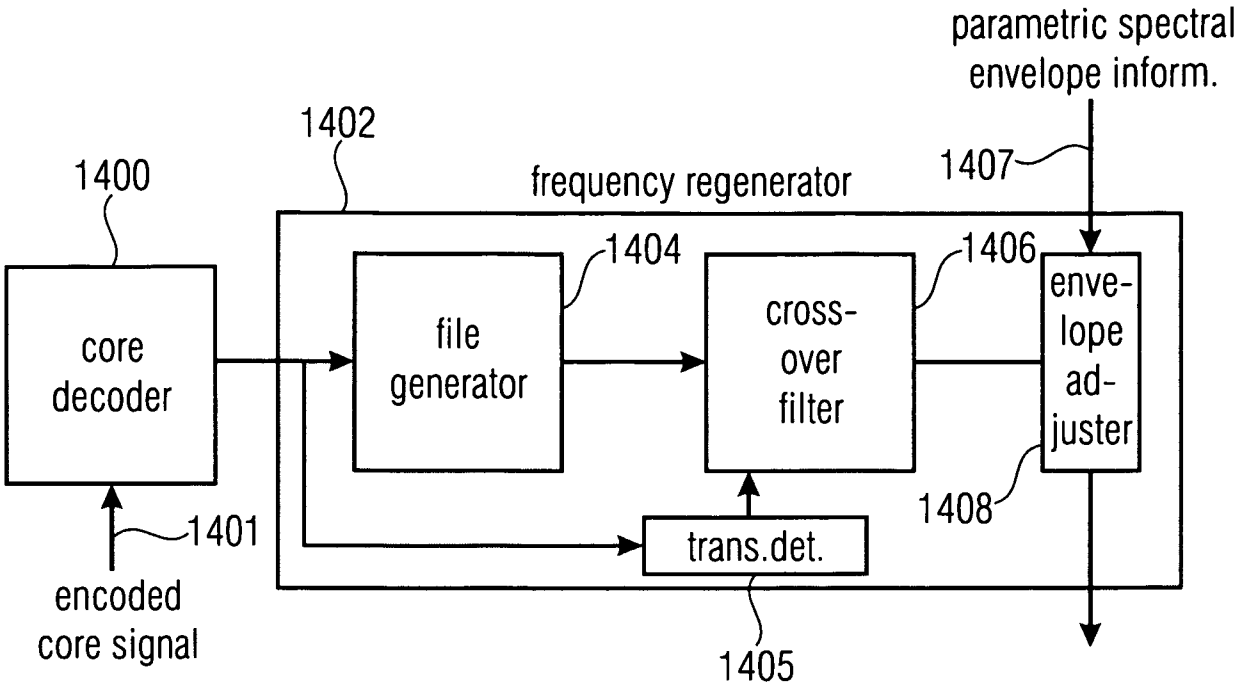
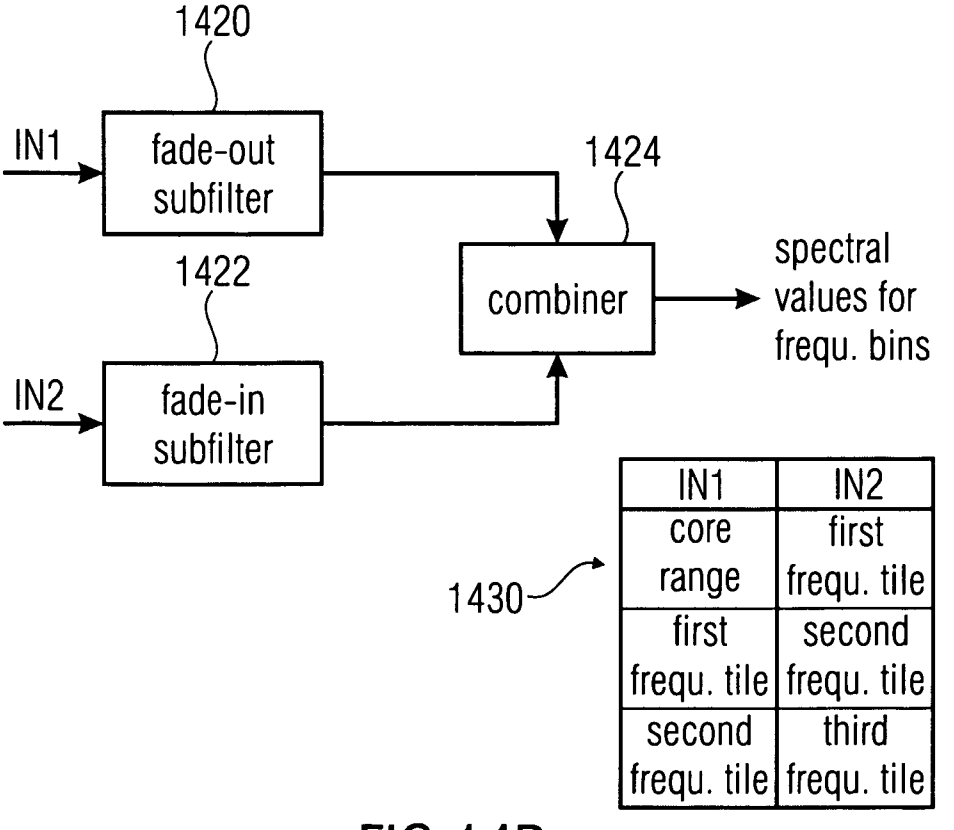


FIG 14A



IN1	IN2
core range	first frequ. tile
first frequ. tile	second frequ. tile
second frequ. tile	third frequ. tile

FIG 14B

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/065112

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L21/0388
ADD. G10L19/028 G10L19/02

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, COMPENDEX, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2010/136459 A1 (DOLBY INT AB [NL]; EKSTRAND PER [SE]; VILLEMOES LARS [SE]; HEDELIN PER) 2 December 2010 (2010-12-02)	1-7, 10-16
A	figures 4,5,12b page 19, line 20 - page 20, line 8 page 13, lines 26-31 page 23, lines 12-31 page 24, lines 14-26	8,9

X	US 8 412 365 B2 (LILJERYD LARS [SE] ET AL) 2 April 2013 (2013-04-02) cited in the application	1,15,16
A	figures 1,2 column 4, lines 15-26 column 4, lines 49-67 column 5, line 58 - column 6, line 37	2-14

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

"E" earlier application or patent but published on or after the international filing date

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 September 2014	06/10/2014

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Chétry, Nicolas

4

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2009/144062 A1 (RAMABADRAN TENKASI V [US] ET AL) 4 June 2009 (2009-06-04)	1,15,16
A	figure 1 paragraph [0027]	2-14

A	FREDERIK NAGEL ET AL: "A HARMONIC BANDWIDTH EXTENSION METHOD FOR AUDIO CODECS", INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING 2009, TAIPEI, 19 April 2009 (2009-04-19), pages 145-148, XP002527507, section 2.3, "Bandwidth extension artifacts" section 3.1, "Spectral stretching"	1-16

4

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2014/065112

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2010136459 A1	02-12-2010	AR 076799 A1	06-07-2011
		AU 2010252028 A1	01-12-2011
		CN 102449692 A	09-05-2012
		CN 103971699 A	06-08-2014
		EP 2436005 A1	04-04-2012
		JP 5363648 B2	11-12-2013
		JP 2012528344 A	12-11-2012
		JP 2014013408 A	23-01-2014
		KR 20120018341 A	02-03-2012
		RU 2011147676 A	27-05-2013
		SG 175975 A1	29-12-2011
		TW 201117196 A	16-05-2011
		US 2012065983 A1	15-03-2012
		WO 2010136459 A1	02-12-2010
US 8412365 B2	02-04-2013	AT 250272 T	15-10-2003
		AU 6283601 A	03-12-2001
		BR 0111362 A	20-05-2003
		CN 1430777 A	16-07-2003
		DE 60100813 D1	23-10-2003
		DE 60100813 T2	15-07-2004
		EP 1285436 A1	26-02-2003
		HK 1067954 A1	28-10-2005
		JP 4289815 B2	01-07-2009
		JP 5090390 B2	05-12-2012
		JP 2003534577 A	18-11-2003
		JP 2009122699 A	04-06-2009
		US 2004131203 A1	08-07-2004
		US 2009041111 A1	12-02-2009
		US 2010211399 A1	19-08-2010
		US 2012213378 A1	23-08-2012
		US 2013339037 A1	19-12-2013
		WO 0191111 A1	29-11-2001
US 2009144062 A1	04-06-2009	CN 101878416 A	03-11-2010
		CN 102646419 A	22-08-2012
		EP 2232223 A1	29-09-2010
		KR 20100086018 A	29-07-2010
		KR 20120055746 A	31-05-2012
		RU 2010126497 A	10-01-2012
		US 2009144062 A1	04-06-2009
		WO 2009070387 A1	04-06-2009