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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 August 2014 (07.08.2014)

WIPO | PCT

(10) International Publication Number
**WO 2014/118160 A1**

(51) **International Patent Classification**:
*G10L 21/038* (2013.01)

(21) **International Application Number**:
PCT/EP2014/051601

(22) **International Filing Date**:
28 January 2014 (28.01.2014)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/758,090 29 January 2013 (29.01.2013) US

(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). **MULTRUS, Markus**; Etzlaubweg 7, 90469 Nürnberg (DE). **SCHMIDT, Konstantin**; Heerwagenstrasse 21, 90489 Nürnberg (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 GENERATING A FREQUENCY ENHANCED SIGNAL USING TEMPORAL SMOOTHING OF SUBBANDS

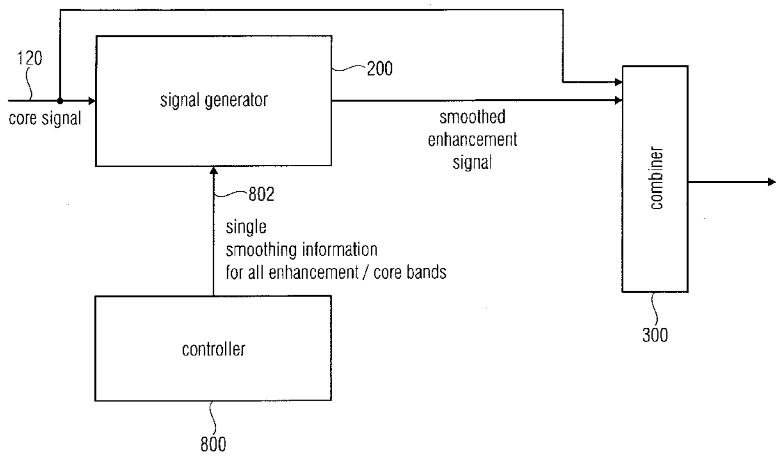


FIG 8

(57) **Abstract**: An apparatus for generating a frequency enhancement signal (130) comprises: a signal generator (200) for generating an enhancement signal from a core signal (120, 110), the enhancement signal comprising an enhancement frequency range not included in the core signal, wherein a current time portion (320, 340) of the enhancement signal or the core signal comprises subband signals for a plurality of subbands; a controller (800) for calculating the same smoothing information (802) for the plurality of subband signals of the enhancement frequency range or the core signal, and wherein the signal generator (200) is configured for smoothing the plurality of subband signals of the enhancement frequency range or the core signal using the same smoothing information (802).

WO 2014/118160 A1

# Apparatus and Method for Generating a Frequency Enhanced Signal using Temporal Smoothing of Subbands

## Specification

The present invention is based on audio coding and in particular on frequency enhancement procedures such as bandwidth extension, spectral band replication or intelligent gap filling.

The present invention is particularly related to non-guided frequency enhancement procedures, i.e. where the decoder-side operates without side information or only with a minimum amount of side information.

Perceptual audio codecs often quantize and code only a lowpass part of the whole perceivable frequency range of an audio signal, especially when operated at (relatively) low bitrates. Although this approach guarantees an acceptable quality for the coded low-frequency signal, most listeners perceive the missing of the highpass part as a quality degradation. To overcome this issue, the missing high-frequency part can by synthesized by bandwidth extension schemes.

State of the art codecs often use either a waveform-preserving coder, such as AAC, or a parametric coder, such as a speech coder, to code the low-frequency signal. These coders operate up to a certain stop frequency. This frequency is called crossover frequency. The frequency portion below the crossover frequency is called low band. The signal above the crossover frequency, which is synthesized by means of a bandwidth extension scheme, is called high band.

A bandwidth extension typically synthesizes the missing bandwidth (high band) by means of the transmitted signal (low band) and extra side information. If applied in the field of low-bitrate audio coding, the extra information should consume as little as possible extra bitrate. Thus, usually a parametric representation is chosen for the extra information. This parametric representation is either transmitted from the encoder at comparably low bitrate (guided bandwidth extension) or estimated at the decoder based on specific signal characteristics (non-guided bandwidth extension). In the latter case, the parameters consume no bitrate at all.

The synthesis of the high band typically consists of two parts:

1. Generation of the high-frequency content. This can be done by either copying or flipping (parts of) the low frequency content to the high band, or inserting white or shaped noise or other artificial signal portions into the high band.

2. Adjustment of the generated high frequency content according to the parametric information. This includes manipulation of shape, tonality/noisiness and energy according to the parametric representation.

The goal of the synthesis process is usually to achieve a signal that is perceptually close to the original signal. If this goal can't be matched, the synthesized portion should be least disturbing for the listener.

Other than a guided BWE scheme, a non-guided bandwidth extension can‘t rely on extra information for the synthesis of the high band. Instead, it typically uses empirical rules to exploit correlation between low band and high band. Whereas most music pieces and voiced speech segments exhibit a high correlation between high and low frequency band, this is usually not the case for unvoiced or fricative speech segments. Fricative sounds have very few energy in the lower frequency range while having high energy above a certain frequency. If this frequency is close to the crossover frequency, then it can be problematic to generate the artificial signal above the crossover frequency since in that case the lowband does contain little relevant signal parts. To cope with this problem, a good detection of such sounds is helpful.

HE-AAC is a well-known codec that consists of a waveform preserving codec for the low band (AAC) and a parametric codec for the high band (SBR). At decoder side, the high band signal is generated by transforming the decoded AAC signal into the frequency domain using a QMF filterbank. Subsequently, subbands of the low band signal are copied to the high band (generation of high frequency content). This high band signal is then adjusted in spectral envelope, tonality and noise floor based on the transmitted parametric side-information (adjustment of the generated high frequency content). Since this method uses a guided BWE approach, a weak correlation between high and low band is in general not problematic and can be overcome be transmitting the appropriate parameter sets. However, this requires additional bitrate, which might not be acceptable for a given application scenario.

The ITU Standard G.722.2 is a speech codec that operates in time domain only, i.e. without performing any calculations in frequency domain. Such a decoder outputs a time domain signal with a sampling rate of 12.8 kHz, which is subsequently upsampled to 16 kHz. The generation of the high frequency content (6.4 – 7.0 kHz) is based on inserting bandpass noise. In most operation modes the spectral shaping of the noise is done without using any side-information, only in the operation mode with highest bitrate information about the noise energy is transmitted in the bitstream. For reasons of simplicity, and since not all application scenarios can afford the transmission of extra parameters sets, in the following only the generation of the high band signal without using any side-information is described.

For generating the high band signal, a noise signal is scaled to have the same energy as the core excitation signal. In order to give more energy to unvoiced parts of the signal, a spectral tilt e is calculated:

$$e = \frac{\sum_{n=1}^{63} s(n)\, s(n-1)}{\sum_{n=0}^{63} s^2\,(n)}$$

where $s$ is the high-pass filtered decoded core signal with cut-off frequency of 400 Hz. $n$ is the sample index. In case of voiced segments where less energy is present at high frequencies, $e$ approaches 1, while for unvoiced segments e is close to zero. In order to have more energy in the high band signal, for unvoiced speech the energy of the noise is multiplied by $(1 - e)$. Finally, the scaled noise signal is filtered by a filter which is derived from the core Linear Predictive Coding (LPC) filter by extrapolation in the Line Spectral Frequency (LSF) domain.

The non-guided bandwidth extension from G.722.2, which entirely operates in time domain, has the following drawbacks:

1. The generated HF content is based on noise. This creates audible artifacts if the HF signal is combined with a tonal, harmonic low-frequency signal (e.g. music). To avoid such artifacts, G.722.2 strongly limits the energy of the generated HF signal, which also limits potential benefits of the bandwidth extension. Thus, unfortunately

also the maximum possible improvement of the brightness of a sound or the maximum obtainable increase in intelligibility of a speech signal is limited.

2. Since this non-guided bandwidth extension operates in the time domain, the filter operations cause additional algorithmic delay. This additional delay lowers the quality of the user experience in bi-directional communication scenarios or might not be allowed by the terms of requirement of a given communication technology standard.

3. Also, since the signal processing is performed in time domain, the filter operations are prone to instabilities. Moreover, the time domain filters have a high computational complexity.

4. Since only the overall sum of the energy of the high band signal is adapted to the energy of the core signal (and further weighted by the spectral tilt), there might be a significant local mismatch of energy at the crossover frequency between upper frequency range of the core signal (the signal just below the crossover frequency) and the high band signal. For example, this will be the case especially for tonal signals that exhibit an energy concentration in the very low frequency range but contain little energy in the upper frequency range.

5. Furthermore, it is computationally complex to estimate a spectral slope in a time domain representation. In frequency domain, an extrapolation of a spectral slope can be done very efficiently. Since most of the energy of e.g. fricatives is concentrated in the high frequency range, these may sound dull if a conservative energy and spectral slope estimation strategy like in G.722.2 is applied (see 1.).

To summarize, the prior art non-guided or blind bandwidth extension schemes may require a significant computational complexity on the decoder side and nevertheless result in a limited audio quality specifically for problematic speech sounds such as fricatives. Furthermore, guided bandwidth extension schemes, although providing a better audio quality and sometimes requiring less computational complexity on the decoder side cannot provide the substantial bitrate reductions due to the fact that the additional parametric information on the high band can require a significant amount of additional bitrate with respect to the encoded core audio signal.

It is therefore an object of the present invention to provide an improved concept for audio processing in the context of non-guided frequency enhancement technologies.

This object is achieved by an apparatus for generating a frequency enhanced signal of claim 1, a method of generating a frequency enhanced signal of claim 11, a system comprising an encoder and an apparatus for generating a frequency enhanced signal of claim 12, a related method of claim 13, or a computer program of claim 14.

The present invention provides a frequency enhancement scheme such as a bandwidth extension scheme for audio codecs. This scheme aims at extending the frequency bandwidth of an audio codec without the need of extra side-information or with only a minimum amount significantly reduced compared to a full parametric description of missing bands as in guided bandwidth extension schemes.

An apparatus for generating a frequency enhanced signal comprises a calculator for calculating a value describing an energy distribution with respect to frequency in a core signal. A signal generator for generating an enhancement signal comprising an enhancement frequency range not included in the core signal operates using the core signal and then performs a shaping of the enhancement signal or the core signal so that the spectral envelope of the enhancement signal depends on the value describing the energy distribution.

Thus, the envelope of the enhancement signal, or the enhancement signal is shaped based on this value describing the energy distribution. This value can be easily calculated and this value then defines the full envelope shape or the full shape of the enhancement signal. Thus, the decoder can operate with a low complexity and at the same time a good audio quality is obtained. Specifically, the energy distribution in the core signal when used for the spectral shaping of the frequency enhancement signal results in a good audio quality even though the processing of calculating the value on the energy distribution such as a spectral centroid in the core signal and the adjustment of the enhancement signal based on this spectral centroid is a procedure which is straightforward and can be performed with low computational resources.

Furthermore, this procedure allows that the absolute energy and the slope (roll-off) of the high band signal are derived from the absolute energy and the slope (roll-off) of the core signal, respectively. It is preferred to perform these operations in the frequency domain so that they can be done in the computationally efficient way, since the shaping of a spectral envelope is equivalent to simply multiplying the frequency representation with a gain curve, and this gain curve is derived from the value describing the energy distribution with respect to frequency in the core signal.

Furthermore it is computationally complex to precisely estimate and extrapolate a given spectral shape in the time domain. Thus, such operations are preferably performed in the frequency domain. Fricative sounds for example have typically only a low amount of energy at low frequencies and a high amount of energy at high frequencies. The rise in energy is dependent on the actual fricative sound and might start only little below the crossover frequency. In the time domain, it is difficult to detect this situation and computationally complex to obtain a valid extrapolation from it. For non-fricative sounds it is assured that the energy of the artificial generated spectrum always drops with rising frequency.

In a further aspect, a temporal smoothing procedure is applied. A signal generator for generating an enhancement signal from a core signal is provided. A time portion of the enhancement signal or the core signal comprises subband signals for a plurality of subbands. A controller for calculating the same smoothing information for the plurality of subband signals of the enhancement frequency range is provided and this smoothing information is then used by the signal generator for smoothing the plurality of subband signals of the enhancement frequency range, particularly using the same smoothing information or, alternatively, when the smoothing is performed before the high frequency generation, then the plurality of subband signals of the core signal are smoothed all using the same smoothing information. This temporal smoothing avoids the continuation of smaller fast energy fluctuations, which are inherited from the low-band, to the high-band, and thus leads to a more pleasant perceptual impression. The low-band energy fluctuations are usually caused by quantization errors of the underlying core-coder that lead to instabilities. The smoothing is signal adaptive since it is dependent on the (long-term) stationary of the signal. Furthermore, the usage of one and the same smoothing information for all individual subbands makes sure that the coherency between the subbands is not changed by the temporal smoothing. Instead, all subbands are smoothed in the same way, and the smoothing information is derived from all subbands or from only the subbands in the enhancement frequency range. Thus, a significantly better audio quality compared to an individual smoothing of each subband signal individually is obtained.

A further aspect is related to performing an energy limitation, preferably at the end of the whole procedure for generating the enhancement signal. A signal generator for generating an enhancement signal from a core signal is provided, where the enhancement signal comprises an enhancement frequency range not included in the core signal, where a time portion of the enhancement signal comprises subband signals for one or a plurality of subbands. A synthesis filterbank for generating the frequency enhancement signal using the enhancement signal is provided, where the signal generator is configured for performing an energy limitation in order to make sure that the frequency enhancement signal obtained by the synthesis filterbank is so that an energy of a higher band is, at the most, equal to an energy in a lower band or greater than, at the most, by a predefined threshold. This may apply for a single extension band. Then, the comparison or energy limitation is done using the energy of the highest core band. This may also apply for a plurality of extension bands. Then a lowest extension band is energy limited using the highest core band, and a highest extension band is energy limited with respect to the second to highest extension band.

This procedure is particularly useful for non-guided bandwidth extension schemes, but can also help in guided bandwidth extension schemes, since the non-guided bandwidth extension schemes are prone to artifacts caused by spectral components which stick out unnaturally, especially at segments which have a negative spectral tilt. These components might lead to high-frequency noise-bursts. To avoid such a situation, the energy limitation is preferably applied at the end of the processing, which limits the energy increment over frequency. In an implementation, the energy at a QMF (Quadrature Mirror Filtering) subband k must not exceed the energy at a QMF subband k-1. This energy limiting might be performed on a time-slot base or to save on complexity, only once per frame. Thus, it is made sure that any unnatural situations in bandwidth extension schemes are avoided, since it is very unnatural that a higher frequency band has more energy than the lower frequency band or that the energy of a higher frequency band is higher by more than the predefined threshold, such as a threshold of 3dB, than the energy in the lower band. Typically, all speech/music signals have a low-pass characteristic, i.e. have a more or less monotonically decreasing energy content over frequency. This may apply for a single extension band. Then, the comparison or energy limitation is done using the energy of the highest core band. This may also apply for a plurality of extension bands. Then a lowest extension band is energy limited using the highest core band, and a highest extension band is energy limited with respect to the second to highest extension band.

Although the technologies of shaping of the frequency enhancement signal, temporal smoothing of the frequency enhancement subband signals and energy limitation can be performed individually and separately from each other, these procedures can also be performed all together within preferably a non-guided frequency enhancement scheme.

Furthermore, reference is made to the dependent claims which refer to specific embodiments.

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

Fig. 1 illustrates an embodiment comprising the technologies of shaping a frequency enhancement signal, the smoothing of the subband signal and the energy limitation;

Fig. 2a-2c illustrate different implementations of the signal generator of Fig. 1;

Fig. 3 illustrates individual time portions, where a frame has a long time portion and a slot has a short time portion and each frame comprises a plurality of slots;

Fig. 4 illustrates a spectral chart indicating the spectral position of a core signal and an enhancement signal in an implementation of a bandwidth extension application;

Fig. 5 illustrates an apparatus for generating the frequency enhanced signal using a spectral shaping based on the value describing an energy distribution of the core signal;

Fig. 6 illustrates an implementation of the shaping technology;

Fig. 7 illustrates different roll-offs determined by a certain spectral centroid;

Fig. 8 illustrates an apparatus for generating the frequency enhanced signal comprising the same smoothing information for smoothing the subband signals of the core signal or the frequency enhancement signal;

Fig. 9 illustrates a preferred procedure applied by the controller and the signal generator of Fig. 8;

Fig. 10 illustrates a further procedure applied by the controller and the signal generator of Fig. 8;

Fig. 11 illustrates an apparatus for generating a frequency enhanced signal, which performs an energy limitation procedure in the enhancement signal so that a higher band of the enhancement signal may, at the most, have the same energy of the adjacent lower band or is, at the most, higher in energy by a predefined threshold;

Fig. 12a illustrates the spectrum of the enhancement signal before limitation;

Fig. 12b illustrates the spectrum of Fig. 12a subsequent to the limitation;

Fig. 13 illustrates a process performed by the signal generator in an implementation;

Fig. 14 illustrates the concurrent application of the technologies of shaping, smoothing and energy limitation within a filterbank domain; and

Fig. 15 illustrates a system comprising an encoder and a non-guided frequency enhancement decoder.

Fig. 1 illustrates an apparatus for generating a frequency enhanced signal 140 in a preferred implementation, in which the technologies of shaping, temporal smoothing and energy limitation are performed all together. However, these technologies can also be individually applied as discussed in the context of Figs. 5 to 7 for the shaping technology, Figs. 8 to 10 for the smoothing technology and Figs. 11 to 13 for the energy limitation technology.

Preferably, the apparatus for generating the frequency enhanced signal 140 of Fig. 1 comprises an analysis filterbank or a core decoder 100 or any other device for providing the core signal in the filterbank domain such as in a QMF domain, when the core decoder outputs QMF subband signals. Alternatively, the analysis filterbank 100 can be a QMF filterbank or another analysis filterbank, when the core signal is a time domain signal or is provided in any other domain than a spectral or subband domain.

The individual subband signals of the core signal 110 which are available at 120 are then input into a signal generator 200 and the output of the signal generator 200 is an enhancement signal 130. This enhancement signal 130 comprises an enhancement frequency range which is not included in the core signal 110 and the signal generator generates this enhancement signal not e.g. by (only) shaping noise or so, but using the core signal 110 or preferably the core signal subbands 120. The synthesis filterbank then combines the core signal subbands 120 and the frequency enhancement signal 130, and the synthesis filterbank 300 then outputs the frequency enhanced signal.

Basically, the signal generator 200 comprises a signal generation block 202 which is indicated as “HF generation” where HF stands for high frequency. However, the frequency enhancement in Fig. 1 is not limited to the technology that a high frequency is generated. Instead, also a low frequency or an intermediate frequency can be generated and there can even be a regeneration of a spectral hole in the core signal, i.e. when the core signal has a higher band and a lower band and when there is a missing intermediate band, as is for example known from intelligent gap filling (IGF). The signal generation 202 may comprise copy-up procedures as known from HE-AAC or mirroring procedures, i.e. where, in order to generate the high frequency range or frequency enhancement range, the core signal is mirrored rather than copied up.

Furthermore, the signal generator comprises a shaping functionality 204, which is controlled by the calculation for calculating a value indicating the energy distribution with respect to frequency in the core signal 120. This shaping may be a shaping of the signal generated by block 202 or alternatively the shaping of the low frequency, when the order between functionality 202 and 204 is reversed as discussed in the context of Fig. 2a to Fig. 2c.

A further functionality is the temporal smoothing functionality 206, which is controlled by a smoothing controller 800. An energy limitation 208 is preferably performed at the end of the procedure, but the energy limitation can also be placed at any other position in the chain of processing functionalities 202 to 208 as long as it is made sure that the combined signal output by the synthesis filterbank 300 fulfills the energy limitation criterion such as that a higher frequency band must not have more energy than the adjacent lower frequency band or that the higher frequency band must not have more energy compared to the adjacent lower frequency band, where the increment is limited, at the most, to a predefined threshold such as 3dB

Fig. 2a illustrates a different order, in which the shaping 204 is performed together with the temporal smoothing 206 and the energy limitation 208 before performing the HF generation 202. Thus, the core signal is shaped/smoothed/limited and then the already completed shaped/smoothed/limited signal is copied-up or mirrored into the enhancement frequency range. Furthermore, it is important to understand that the order of blocks 204, 206, 208 can be performed in any way as can also be seen when Fig. 2a is compared to the order of the corresponding blocks in Fig. 1.

Fig. 2b illustrates a situation, in which the temporal smoothing and the shaping is performed on the low frequency or core signal, and the HF generation 202 is then performed before the energy limitation 208. Furthermore, Fig. 2c illustrates a situation where the shaping of the signal is performed to the low frequency signal and a subsequent HF generation such as by copy-up or mirroring is performed in order to obtain the signal for the enhancement frequency range, and this signal is then smoothed 206 and energy-limited 208.

Furthermore, it is to be emphasized that the functionalities of shaping, temporal smoothing and energy limiting may all be performed by applying certain factors to a subband signal as, for example, illustrated in Fig. 14. The shaping is implemented by multipliers 1402a, 1401a and 1400a for individual bands i, i + 1, i + 2.

Furthermore, the temporal smoothing is performed by multipliers 1402b, 1401b and 1400b. Additionally, the energy limitation is performed by limitation factors 1402c, 1401c and 1400c for the individual bands i + 2, i + 1 and i. Due to the fact that all of these functionalities are implemented in this embodiment by multiplication factors, it is to be noted that all these functionalities can also be applied to the individual subband signals by a single multiplication factor 1402, 1401, 1400 for each individual band, and this single "master" multiplication factor would then be a product of the individual factors 1402a, 1402b and 1402c for a band i + 2, and the situation would be analogous to the other bands i + 1 and i. Thus, the real/imaginary subband samples values for the subbands are then multiplied by this single "master" multiplication factor and the output is obtained as multiplied real/imaginary subband sample values at the output of block 1402, 1401 or 1400, which are then introduced into the synthesis filterbank 300 of Fig. 1. Thus, the output of blocks 1400, 1401, 1402 corresponds to the enhancement signal 1300 typically covering the enhancement frequency range not included in the core signal.

Fig. 3 illustrates a chart indicating different time resolutions used in the process of signal generation. Basically, the signal is processed frame-wise. This means that the analysis filterbank 100 is preferably implemented to generate time-subsequent frames 320 of subband signals, where each frame 320 of subband signals comprises a one or a plurality of slots or filterbank slots 340. Although Fig. 3 illustrates four slots per frame, there can also be 2, 3 or even more than four slots per frame. As illustrated in Fig. 14, the shaping of the enhancement signal or the core signal based on the energy distribution of the core signal is performed once per frame. On the other hand, the temporal smoothing is performed with a high time resolution, i.e. preferably once per slot 340 and the energy limitation can once again be performed once per frame when a low complexity is required, or once per slot when a higher complexity is non-problematic for the specific implementation.

Fig. 4 illustrates a representation of a spectrum having five subbands 1, 2, 3, 4, 5 in the core signal frequency range. Furthermore, the example in Fig. 4 has four subband signals or subbands 6, 7, 8, 9 in the enhancement signal range and the core signal range and the enhancement signal range are separated by a crossover frequency 420. Furthermore, a start frequency band 410 is illustrated, which is used for calculating the value describing an energy distribution with respect to frequency for the purpose of shaping 204, as will be discussed later on. This procedure makes sure that the lowest or a plurality of lowest subbands are not used for the calculation of the value describing the energy distribution with respect to frequency in order to obtain a better enhancement signal adjustment.

Subsequently, an implementation of the generation 202 of the enhancement frequency range not included in the core signal using the core signal is illustrated.

In order to generate the artificial signal above the crossover frequency, typically QMF values from the frequency range below the crossover frequency are copied ("patched") up into the high band. This copy-operation can be done by just shifting QMF samples from the lower frequency range up to the area above the crossover frequency or by additionally mirroring these samples. The advantage of the mirroring is that the signal just below the crossover frequency and the artificial generated signal will have a very similar energy and harmonic structure at the crossover frequency. The mirroring or copy up can be applied to a single subband of the core signal or to a plurality of subbands of the core signal.

In the case of said QMF filterbank, the mirrored patch preferably consists of the negative complex conjugate of the base band in order to minimize subband aliasing in the transition region:

$$Qr(t, xover + f - 1) = -Qr(t, xover - f); f = 1..nBands$$
$$Qi(t, xover + f - 1) = Qi(t, xover - f); f = 1..nBands$$

Here, $Qr(t,f)$ is the real value of the QMF at time-index $t$ and subband-index $f$ and $Qi(t,f)$ is the imaginary value; $xover$ is the QMF subband referring to the crossover frequency; $nBands$ is the integer number of bands to be extrapolated. The minus sign in the real part denotes the negative conjugate complex operation.

Preferably, the HF generation 202 or generally the generation of the enhancement frequency range relies on a subband representation provided by block 100. Preferably, the inventive apparatus for generating a frequency enhanced signal should be a multi-bandwidth decoder which is able to resample the decoded signal 110 to vary sampling frequencies, to support, for example narrow band, wideband and super-wideband output. Therefore, the QMF filterbank 100 takes the decoded time domain signal as input. By padding zeroes in the frequency domain, the QMF filterbank can be used to resample the decoded signal, and the same QMF filterbank is preferably also used to create the high band signal.

Preferably, the apparatus for generating a frequency enhanced signal is operative to perform all operations in the frequency domain. Thus, an existing system already having an internal frequency domain representation at a decoder side is extended as illustrated in Fig. 1 by indicating block 100 as a “core decoder” which provides, for example, already a QMF filterbank domain output signal.

This representation is simply re-used for additional tasks like sampling rate conversion and other signal manipulations which are preferably done in the frequency domain (e.g. insertion of shaped comfort noise, high-pass/low-pass filtering). Thus, no additional time-frequency transformation needs to be calculated.

Instead of using noise for the HF content, the high-band signal is generated based on the low-band signal only in this embodiment. This can be done by means of a copy-up or folding-up (mirroring) operation in the frequency domain. Thus, a high band signal with the same harmonic and temporal fine-structure as the low band signal is assured. This avoids a computationally costly folding of the time-domain signal and additional delay.

Subsequently, the functionality of the shaping 204 technology of Fig. 1 is discussed in the context of Figs. 5, 6, and 7, where the shaping can be performed in the context of Fig. 1, 2a-2c or separately and individually together with other functionalities known from other guided or non-guided frequency enhancement technologies.

Fig. 5 illustrates an apparatus for generating a frequency enhanced signal 140 comprising a calculator 500 for calculating a value describing an energy distribution with respect to frequency in a core signal 120. Furthermore, the signal generator 200 is configured for generating an enhancement signal comprising an enhancement frequency range not included in the core signal from the core signal as illustrated by line 502. Furthermore, the signal generator 200 is configured for shaping the enhancement signal such as output by block 202 in Fig. 1 or the core signal 120 in the context of Fig. 2a so that a spectral envelope of the enhancement signal depends on the value describing the energy distribution.

Preferably, the apparatus additionally comprises a combiner 300 for combining the enhancement signal 130 output by block 200 and the core signal 120 to obtain the frequency enhanced signal 140. Additional operations such as temporal smoothing 206 or energy limitation 208 are preferred to further process the shaped signal, but are not necessarily required in certain implementations.

The signal generator 200 is configured to shape the enhancement signal so that a first spectral envelope decrease from a first frequency in the enhancement frequency range to a second higher frequency in the enhancement frequency range is obtained for a first value describing the energy distribution. Furthermore, a second spectral envelope decrease from the first frequency in the enhancement range to the second frequency in the enhancement range is obtained for a second value describing a second energy distribution. If the second frequency is greater than the first frequency, and the second spectral envelope decrease is greater than the first spectral envelope decrease, then the first value indicates that the core signal has an energy concentration at a higher frequency range of the core signal compared to the second value describing an energy concentration at a lower frequency range of the core signal.

Preferably, the calculator 500 is configured to calculate a measure for a spectral centroid of a current frame as the information value on the energy distribution. Then, the signal generator 200 shapes in accordance with this measure for the spectral centroid so that a spectral centroid at a higher frequency results in a more shallow slope of the spectral envelope compared to a spectral centroid at a lower frequency.

The information on the energy distribution calculated by the energy distribution calculator 500 is calculated on a frequency portion of the core signal starting at the first frequency and ending at the second frequency being higher than the first frequency. The first frequency is lower than a lowest frequency in the core signal, as for example illustrated at 410 in Fig. 4. Preferably, the second frequency is the crossover frequency 420 but can also be a frequency lower than the crossover frequency 420 as the case may be. However, extending the second frequency used for calculating the measure for the spectral distribution as much as possible to the crossover frequency 420 is preferred and results in the best audio quality.

In an embodiment, the procedure of Fig. 6 is applied by the energy distribution calculator 500 and the signal generator 200. In step 602, an energy value for each band of the core signal indicated at $E(i)$ is calculated. Then, a single energy distribution value such as *sp* used for the adjustment of all bands of the enhancement frequency range is calculated in block 604. Then, in step 606, weighting factors are calculated for all bands of the enhancement frequency range using for this a single value, where the weighting factors are preferably $att^f$.

Then, in step 608 performed by the signal generator 208, the weighting factors are applied to real and imaginary parts of the subband samples.

Fricative sounds are detected by calculating the spectral centroid of the current frame in the QMF domain. The spectral centroid is a measure that has a range of 0.0 to 1.0. A high spectral centroid (a value close to one) means that the spectral envelope of the sound has a rising slope. For speech signals this means that the current frame most likely contains a fricative. The closer the value of the spectral centroid approaches one, the steeper is the slope of the spectral envelope or the more energy is concentrated in the higher frequency range.

The spectral centroid is calculated according to:

$$sp = \frac{\sum_{i=start}^{xover} i * E(i)}{(xover - start + 1) * \sum_{i=start}^{xover} E(i)}$$

where $E(i)$ is the energy of QMF subband $i$ and $start$ is the QMF subband-index referring to 1 kHz. The copied QMF subbands are weighted with the factor $att^f$:

$$\widehat{Qr}(t, xover + f) = Qr(t, xover + f) * att^f; f = 1..nBands$$

where $att = 0.5 * sp + 0.5$. Generally, att can be calculated using the following equation:

$$att = p\ (sp),$$

wherein p is a polynomial. Preferably, the polynomial has degree 1:

$$att = a * sp + b,$$

wherein a, b or generally the polynomial coefficients are all between 0 and 1.

Apart from the above equation, other equations having a comparable performance can be applied. Such other equations are as follows:

$$sp = \frac{\sum_{i=start}^{xover} ai * E(i)}{bi * \sum_{i=start}^{xover} E(i)}$$

In particular, the value $a_i$ should be so that the value is higher for higher i and, importantly, the values $b_i$ are lower than the values $a_i$ at least for the index $i$ > 1. Thus, a similar result, but with a different equation compared to the above equation, is obtained. Generally, ai, bi are monotonically increasing or decreasing values with i.

Furthermore, reference is made to Fig. 7. Fig. 7 illustrates individual weighting factors $att^f$ for different energy distribution values *sp*. When *sp* is equal to 1, then the whole energy of the core signal is concentrated at the highest band the core signal. Then, *att* is equal to 1 and the weighting factors $att^f$ are constant over frequency as illustrated at 700. When, on the other hand, the complete energy in the core signal is concentrated at the lowest band of the core signal, then *sp* is equal to 0 and *att* is equal to 0.5 and the corresponding course of the adjustment factors over frequency illustrated at 706.

Courses of shaping factors over frequency indicated at 702 and 704 are for correspondingly increasing spectral distribution values. Thus, for item 704, the energy distribution value is greater than 0 but smaller than the energy distribution value for item 702 as indicated by parametric arrow 708.

Fig. 8 illustrates an apparatus for generating a frequency enhanced signal using the temporal smoothing technology. The apparatus comprises a signal generator 200 for generating an enhancement signal from a core signal 120, 110, where the enhancement signal comprises an enhancement frequency range not included in the core signal. A current time portion such as a frame 320 and preferably a slot 340 of the enhancement signal or the core signal comprises subband signals for a plurality of subbands.

A controller 800 is for calculating the same smoothing information 802 for the plurality of subband signals of the enhancement frequency range or the core signal. Furthermore, the signal generator 200 is configured for smoothing the plurality of subband signals of the enhancement frequency range using the same smoothing information 802 or for smoothing the plurality of subband signals of the core signal using the same smoothing information 802. The output of the signal generator 200 is, in Fig. 8, a smooth enhancement signal which can then be input into a combiner 300. As discussed in the context of Figs. 2a-2c, the smoothing 206 can be performed at any place in the processing chain of Fig. 1 or can even be performed individually in the context of any other frequency enhancement scheme.

The controller 800 is preferably configured to calculate the smoothing information using a combined energy of the plurality of subband signals the core signal and the frequency enhancement signal or using only the frequency enhancement signal of the time portion. Furthermore, an average energy of the plurality of subband signals of the core signal and the frequency enhancement signal or of the core signal only of one or more earlier time portions preceding the current time portion is used. The smoothing information is a single correction factor for the plurality of subband signals of the enhancement frequency range in all bands and therefore the signal generator 200 is configured to apply the correction factor to the plurality of subband signals of the enhancement frequency range.

As discussed in the context of Fig. 1, the apparatus furthermore comprises a filterbank 100 or a provider for providing the plurality of subband signals of the core signal for a plurality of time-subsequent filterbank slots. Furthermore, the signal generator is configured to derive the plurality of subband signals of the enhancement frequency range for the plurality of time-subsequent filterbank slots using the plurality of subband signals of the core signal and the controller 800 is configured to calculate an individual smoothing information 802 for each filterbank slot and the smoothing is then performed, for each filterbank slot, with a new individual smoothing information.

The controller 800 is configured to calculate a smoothing intensity control value based on the core signal or the frequency enhanced signal of the current time portion and based on one or more preceding time portions and the controller 800 is then configured to calculate the smoothing information using the smoothing control value such that the smoothing intensity varies depending on a difference between an energy of the core signal or the frequency enhancement signal of the current time portion and the average energy of the core signal or the frequency enhancement signal of the one or more preceding time portions.

Reference is made to Fig. 9 illustrating a procedure performed by the controller 800 and the signal generator 200. Step 900, which is performed by the controller 800, comprises finding a decision about smoothing intensity which may, for example, be found based on a difference between the energy in the current time portion and an average energy in one or more preceding time portions, but any other procedures for deciding about the smoothing intensity can be used as well. One alternative is to used, instead or in addition future time slots. A further alternative is that one only has a single transform per frame and one would then smooth over timely subsequent frames. Both these alternatives, however, can introduce a delay. This can be non-problematic in applications, where delay is not a problem, such as streaming application. For applications, where a delay is problematic such as for a two way communication e.g. using mobile phones, the past or preceding frames are preferred over future frames, since the usage of the past frames does not introduce a delay.

Then, in step 902, a smoothing information is calculated based on the decision of the smoothing intensity of the step 900. This step 902 is also performed by the controller 800. Then, the signal generator 200 performs 904 comprising the application of the smoothing information to several bands, where one and the same smoothing information 802 is applied to these several bands either in the core signal or in the enhancement frequency range.

Fig. 10 illustrates a preferred procedure of the implementation of the Fig. 9 sequence of steps. In step 1000, an energy of a current slot is calculated. Then, in step 1020, an average energy of one or more previous slots is calculated. Then, in step 1040, a smoothing coefficient for the current slot is determined based on the difference between the values obtained by block 1000 and 1020. Then, step 1060 comprises the calculation of a correction factor for the current slot and the steps 1000 to 1060 are all performed by the controller 800. Then, in step 1080, which is performed by the signal generator 200, the actual smoothing operation is performed, i.e. the corresponding correction factor is applied to all subband signals within one slot.

In an embodiment, the temporal smoothing is performed in two steps:

*Decision about smoothing intensity.* For the decision about the smoothing intensity, the stationary of the signal over time is evaluated. A possible way to perform this evaluation is to compare the energy of the current short-term window or QMF time-slot with averaged energy values of previous short-term windows or QMF time-slots. To save on complexity, this might be evaluated for the high-band portion only. The closer the compared energy values are, the lower should be the intensity of smoothing. This is reflected in a smoothing coefficient $a$, where $0 < a \leq 1$. The greater $a$, the higher is the intensity of smoothing.

*Application of smoothing to the high-band.* The smoothing is applied for the high-band portion on a QMF time-slot base. Therefore, the high-band energy of the current time-slot $Ecurr_t$ is adapted to an averaged high-band energy $Eavg_t$ of one or multiple previous QMF time-slots:

$$\widehat{Ecurr_t} = a\,Ecurr_t + (1-a)\,Eavg_t$$

$Ecurr$ is calculated as the sum of high-band QMF energies in one timeslot:

$$Ecurr_t = \sum_{f=xover}^{xover+nBands} {Qr_{t,f}}^2 + {Qi_{t,f}}^2.$$

$Eavg$ is the moving average over time of the energies:

$$Eavg = \frac{1}{stop-start}\sum_{t=start}^{stop} Ecurr_t$$

where $start$ and $stop$ are the borders of the interval used for calculating the moving average.

The real and imaginary QMF values used for synthesis are multiplied with a correction factor $currFac$:

$$\widehat{Qr}_{t,f} = currFac\;Qr_{t,f}$$

$$\widehat{Qi}_{t,f} = currFac\;Qi_{t,f}$$

which is derived from $Ecurr$ and $Eavg$:

$$currFac = \sqrt{\frac{a\,Ecurr_t + (1-a)Eavg_t}{Ecurr_t}}$$

The factor $a$ may be fixed or dependent on the difference of the energy of $Ecurr$ and $Eavg$.

As already discussed in Fig. 14, the time resolution for the temporal smoothing is set to be higher than the time resolution of the shaping or the time resolution of the energy limitation technology. This makes sure that a temporally smooth course of the subband signals is obtained while, at the same time, the computationally more intensive shaping is to be performed only once per frame. However, any smoothing from one subband to the other subband, i.e. in the frequency direction, is not performed, since, as has been found, this substantially reduces the subjective listening quality.

It is preferred to use the same smoothing information such as the correction factor for all subbands in the enhancement range. However, it can also be an implementation, in which the same smoothing information is applied not for all bands but for a group of bands wherein such a group has at least two subbands.

Fig. 11 illustrates a further aspect directed to the energy limitation technology 208 illustrated in Fig. 1. Specifically, Fig. 11 illustrates an apparatus for generating a frequency enhanced signal comprising the signal generator 200 for generating an enhancement signal, the enhancement signal comprising an enhancement frequency range not included in the core signal. Furthermore, a time portion of the enhancement signal comprises subband signals for a plurality of subbands. Additionally, the apparatus comprises a synthesis filterbank 300 for generating the frequency enhanced signal 140 using the enhancement signal 130.

In order to implement the energy limitation procedure, the signal generator 200 is configured for performing an energy limitation in order to make sure that the frequency enhanced signal 140 obtained by the synthesis filterbank 300 is so that an energy of a higher band is, at the most, equal to an energy in a lower band or greater than the energy in a lower band, at the most, by a predefined threshold.

The signal generator is preferable implemented to make sure that a higher QMF subband k must not exceed the energy at a QMF subband k – 1. Nevertheless, the signal generator 200 can also be implemented to allow a certain incremental increase which can preferably be a threshold of 3dB and a threshold can preferably be 2dB and even more preferably 1dB or even smaller. The predetermined threshold may be a constant for each band or dependent on the spectral centroid calculated previously. A preferred dependence is that the threshold becomes lower, when the centroid approaches lower frequencies, i.e. becomes smaller, while the threshold can become greater the closer the centroid approaches higher frequencies or sp approaches 1.

In a further implementation, the signal generator 200 is configured to examine a first subband signal in a first subband and to examine a subband signal in a second subband being adjacent in frequency to the first subband and having a center frequency being higher than a center frequency of the first subband and the signal generator will not limit the second subband signal, when an energy of the second subband signal is equal to an energy of the first subband signal or when the energy of the second subband signal is greater than the energy of the first subband signal by less than the predefined threshold.

Furthermore, the signal generator is configured to form a plurality of processing operations in a sequence as illustrated, for example, in Fig. 1 or Figs. 2a-2c. Then, the signal generator preferably performs the energy limitation at an end of the sequence to obtain the enhancement signal 130 input into the synthesis filterbank 300. Thus, the synthesis filterbank 300 is configured to receive, as an input, the enhancement signal 130 generated at the end of the sequence by the final process of the energy limitation.

Furthermore, the signal generator is configured to perform spectral shaping 204 or temporal smoothing 206 before the energy limitation.

In a preferred embodiment, the signal generator 200 is configured to generate the plurality of subband signals of the enhancement signal by mirroring a plurality of subbands of the core signal.

For the mirroring, preferably the procedure of negating either the real part or the imaginary part is performed as discussed earlier.

In a further embodiment, the signal generator is configured for calculating a correction factor *limFac* and this limitation factor *limFac* is then applied to the subband signals of the core or the enhancement frequency range as follows:

Let $E_f$ be the energy of one band averaged over a time span $stop - start$:

$$E_f = \sum_{t=start}^{stop} {Qr_{t,f}}^2 + {Qi_{t,f}}^2$$

If this energy exceeds the average energy of the previous band by some level, the energy of this band is multiplied by a correction/limitation factor $limFac$:

$$if\ E_f > fac * E_{f-1}$$

$$limFac = \sqrt{\frac{fac * E_{f-1}}{E_f}}$$

and the real and imaginary QMF values are corrected by:

$$\widehat{Qr}_{t,f} = limFac\ Qr_{t,f}$$
$$\widehat{Qi}_{t,f} = limFac\ Qi_{t,f}$$

The factor or predetermined threshold $fac$ may be a constant for each band or dependent on the spectral centroid calculated previously.

$\hat{Q}r_{t,f}$ is the energy limited real part of subband signal at the subband indicated by *f*. $\hat{Q}i_{t,f}$ is the corresponding imaginary part of a subband signal subsequent to energy limitation in a subband *f*. $Qr_{t,f}$ and $Qi_{t,f}$ are corresponding real and imaginary parts of the subband signals before energy limitation such as the subband signals directly when any shaping or temporal smoothing is not performed or the shaped and temporally smoothed subband signals.

In another implementation, the limitation factor *limFac* is calculated using the following equation:

$$limFac = \sqrt{\frac{E_{lim}}{E_f(i)}}.$$

In this equation, $E_{lim}$ is the limitation energy, which is typically the energy of the lower band or the energy of the lower band incremented by the certain threshold fac. $E_f(i)$ is the energy of the current band *f* or *i*.

Reference is made to Figs. 12a and 12b illustrating a certain example where there are seven bands in the enhancement frequency range. Band 1202 is greater than band 1201 with respect to energy. Thus, as becomes clear from Fig. 12b, band 1202 is energy-limited as indicated at 1250 in Fig. 12b for this band. Furthermore, bands 1205, 1204 and 1206 are all greater than band 1203. Thus, all three bands are energy-limited as illustrated as 1250 in Fig. 12b. The only non-limited bands that remain are bands 1201 (this is the first band in the reconstruction range) and bands 1203 and 1207.

As outlined, Fig. 12a/12b illustrates the situation where the limitation is so that a higher band must not have more energy than a lower band. However, the situation would look a bit different if a certain increment would have been allowed.

The energy limitation may apply for a single extension band. Then, the comparison or energy limitation is done using the energy of the highest core band. This may also apply for a plurality of extension bands. Then a lowest extension band is energy limited using the highest core band, and a highest extension band is energy limited with respect to the second to highest extension band.

Fig. 15 illustrates a transmission system or, generally, a system comprising an encoder 1500 and a decoder 1510. The encoder is preferably an encoder for generating the encoded core signal which performs a bandwidth reduction, or generally which deletes several frequency ranges in the original audio signal 1501, which do not necessarily have to be a complete upper frequency range or upper band, but which can also be any frequency band in between core frequency bands. Then, the encoded core signal is transmitted from the encoder 1500 to the decoder 1510 without any side information and the decoder 1510 then performs a non-guided frequency enhancement to obtain the frequency enhancement signal 140. Thus, the decoder can be implemented as discussed in any of the Figs. 1 to 14.

Although the present invention has been described in the context of block diagrams where the blocks represent actual or logical hardware components, the present invention can also be implemented by a computer-implemented method. In the latter case, the blocks represent corresponding method steps where these steps stand for the functionalities performed by corresponding logical or physical hardware blocks.

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

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

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

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

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

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

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

A further embodiment of the inventive method is, therefore, a data carrier (or a non-transitory storage medium such as 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.

## Claims

1. Apparatus for generating a frequency enhancement signal (130), comprising:

a signal generator (200) for generating an enhancement signal from a core signal (120, 110), the enhancement signal comprising an enhancement frequency range not included in the core signal, wherein a current time portion (320, 340) of the enhancement signal or the core signal comprises subband signals for a plurality of subbands;

a controller (800) for calculating the same smoothing information (802) for the plurality of subband signals of the enhancement frequency range or the core signal, and

wherein the signal generator (200) is configured for smoothing the plurality of subband signals of the enhancement frequency range or the core signal using the same smoothing information (802).

2. Apparatus of claim 1,

wherein the controller (800) is configured to calculate the smoothing information (802) using a combined energy of the plurality of subband signals of the core signal and the frequency enhancement signal or using only the frequency enhancement signal of the current time portion, and

using an average energy of the plurality of subband signals of the core signal and the frequency enhancement signal or of the core signal only of one or more earlier time portions preceding the current time portion or one or more later time portions following the current time portion.

3. Apparatus of claim 1 or 2,

wherein the smoothing information 802 is a single correction factor (1402b, 1401b, 1400b) for the plurality of subband signals of the enhancement frequency range, and wherein the signal generator (200) is configured to apply the correction factor to the plurality of subband signals of the enhancement frequency range.

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

further comprising a filterbank or a provider (100) for providing the plurality of subband signals of the core signal for a plurality of time-subsequent filterbank slots (340),

wherein the signal generator (200) is configured to derive the plurality of subband signals of the enhancement frequency range for the plurality of time-subsequent filterbank slots (340) using the plurality of subband signals of the core signal (120), and

wherein the controller (800) is configured to calculate an individual smoothing information for each filterbank slot (340).

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

wherein the controller (800) is configured to calculate a smoothing intensity control value (1040) based on the core signal or the frequency enhancement signal of the current time portion and one or more preceding time portions, and

wherein the controller (800) is configured to calculate the smoothing information (802) using the smoothing control value (1060) in such a way that the smoothing intensity varies dependent on a difference between an energy of the core signal or the frequency enhancement signal in a current time portion and an average energy in the core signal or the frequency enhancement signal of one or more preceding time portions.

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

wherein the controller (800) is configured to calculate the smoothing information (802) based on the following equation:

$$currFac = \sqrt{\frac{a\ Ecurr_t + (1-a)Eavg_t}{Ecurr_t}},$$

wherein $Ecurr_t$ is an energy in the current time portion, wherein $Eavg_t$ is an average of one or more preceding or later time portions, and wherein $a$ is a parameter controlling the smoothing intensity, and

wherein the signal generator is configured to apply the smoothing information on each subband sample of the plurality of subbands of the frequency enhanced signal.

7. Apparatus in accordance with one of the preceding claims, wherein the signal generator (200) is configured for shaping (204) the core signal or the enhancement signal in addition to smoothing.

8. Apparatus of claim 7,

wherein the current time portion and at least one further time portion form a frame (340),

wherein the signal generator (200) is configured for applying the same shaping information for a whole frame (340), and wherein the signal generator (200) is configured for smoothing using an individual smoothing information (802) for each time portion (340) within the frame (320).

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

wherein the signal generator (200) is configured for performing an energy limitation on the frequency enhancement signal or the core signal in order to make sure that a signal obtained by a synthesis filterbank (300) is so that an energy of a higher band is, at the most, equal to an energy in a lower band or greater than, at the most, by a predefined threshold of 3dB or less.

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

wherein the signal generator (200) is configured for mirroring (202) a single subband signal of the core signal or the plurality of subband signals of the core signal when calculating the plurality of subband signals of the frequency enhancement signal.

11. Method of generating a frequency enhancement signal (130), comprising:

generating (200) an enhancement signal from a core signal (120, 110), the enhancement signal comprising an enhancement frequency range not included in the core signal, wherein a current time portion (320, 340) of the enhancement signal or the core signal comprises subband signals for a plurality of subbands;

calculating (800) the same smoothing information (802) for the plurality of subband signals of the enhancement frequency range or the core signal, and

wherein the generating (200) comprises smoothing the plurality of subband signals of the enhancement frequency range or the core signal using the same smoothing information (802).

12. System for processing audio signals, comprising:

an encoder (1500) for generating an encoded core signal (110); and

an apparatus for generating a frequency enhancement signal of any one of claims 1 to 10.

13. Method of processing audio signals, comprising:

generating (1500) an encoded core signal (110); and

generating a frequency enhancement signal using a method of claim 11.

14. Computer program for performing, when running on a computer or a processor, the method of claim 11 or claim 13.

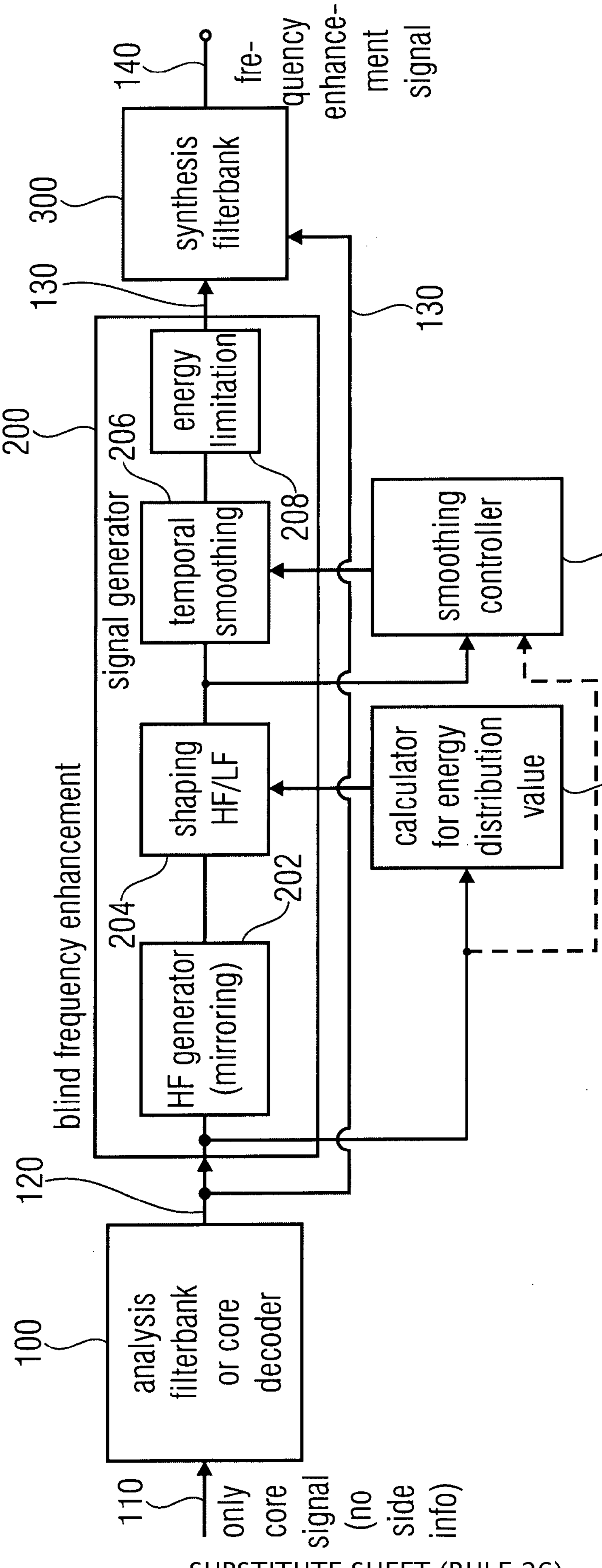
110
only core signal (no side info)
100
analysis filterbank or core decoder
120
200
blind frequency enhancement
signal generator
HF generator (mirroring)
202
204
shaping HF/LF
temporal smoothing
206
energy limitation
208
130
300
synthesis filterbank
140
fre-quency enhance-ment signal
130
calculator for energy distribution value
500
smoothing controller
800

FIG 1

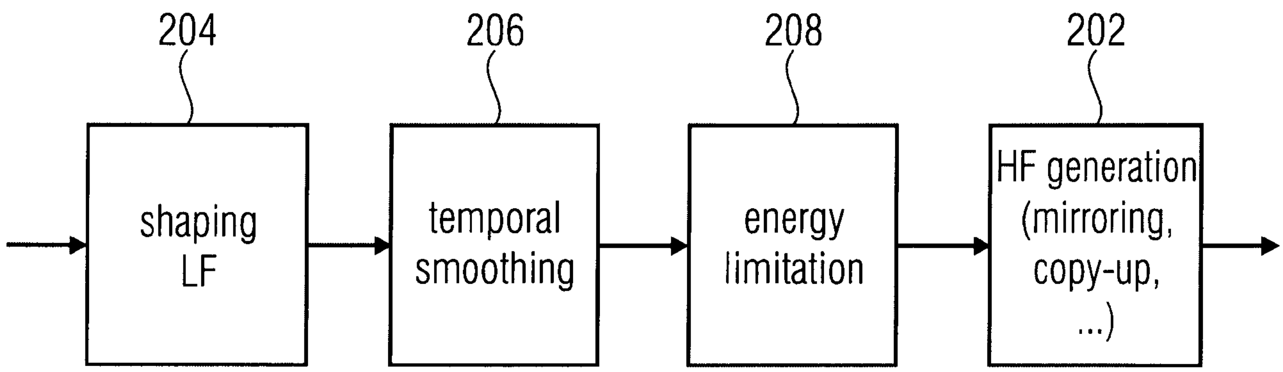


FIG 2A

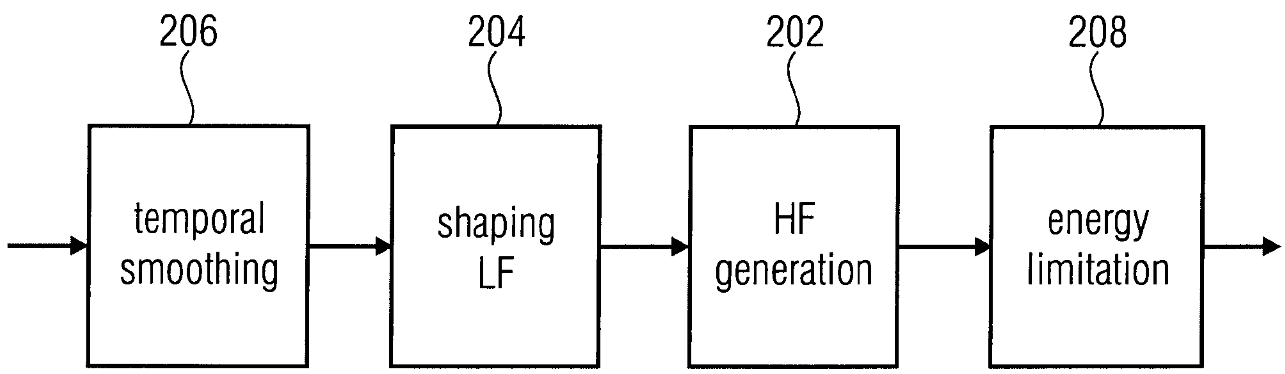


FIG 2B

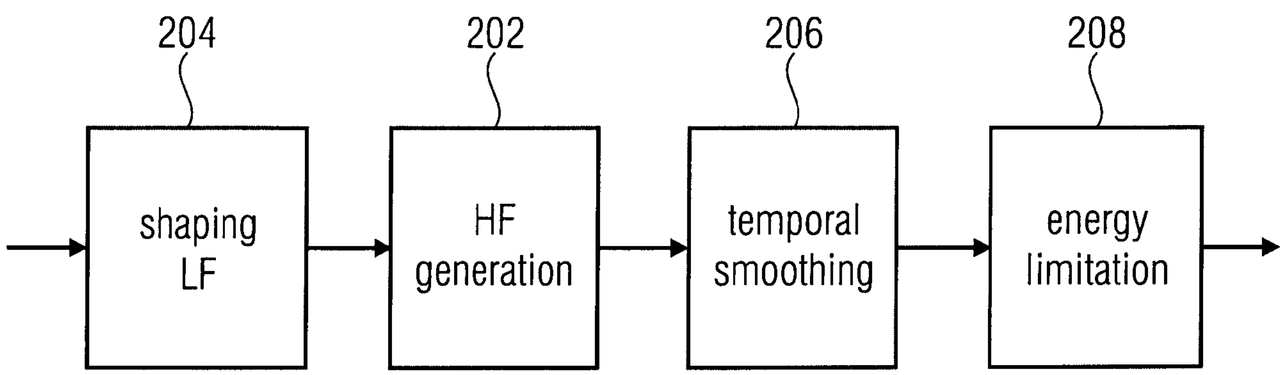


FIG 2C

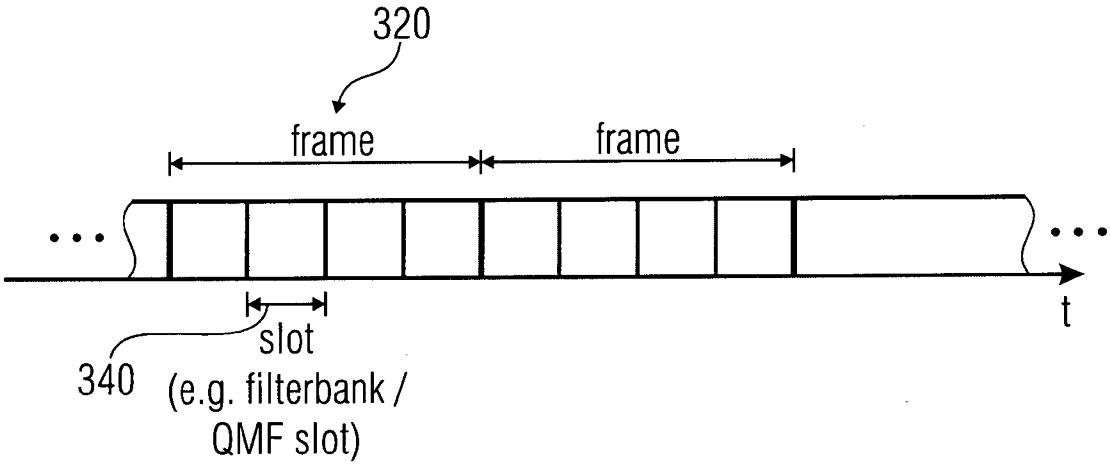
320
frame
frame
t
340
slot
(e.g. filterbank /
QMF slot)

FIG 3

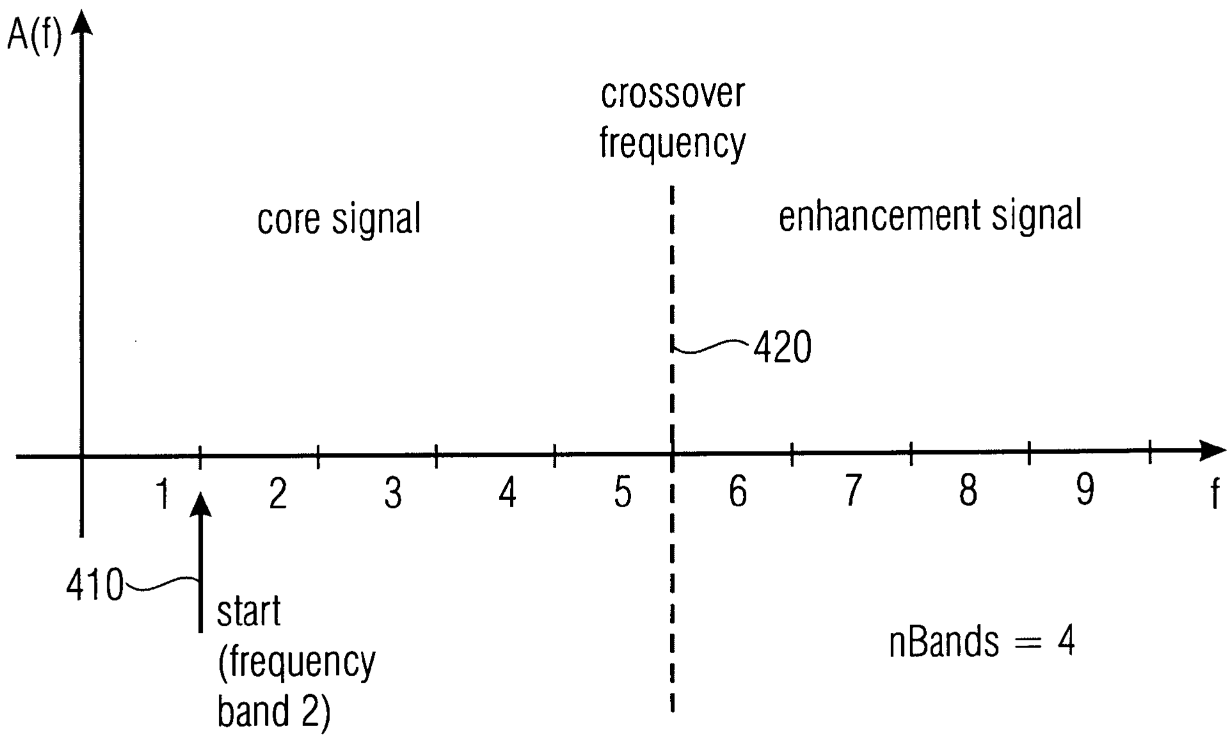
A(f)
crossover
frequency
core signal
enhancement signal
420
1
2
3
4
5
6
7
8
9
f
410
start
(frequency
band 2)
nBands = 4

FIG 4

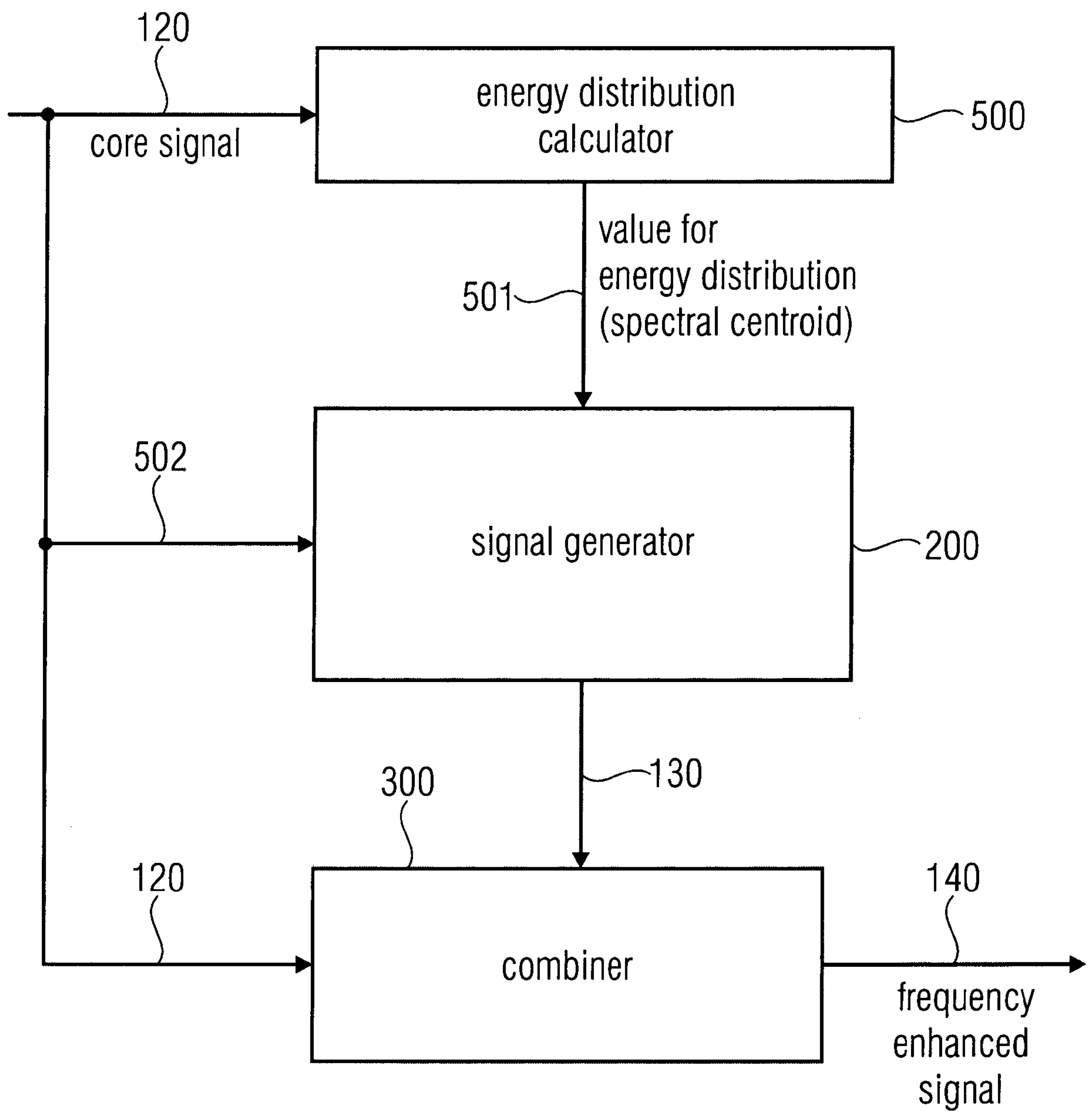
120
core signal
energy distribution calculator
500
value for
energy distribution
(spectral centroid)
501
502
signal generator
200
130
300
120
combiner
140
frequency
enhanced
signal

FIG 5

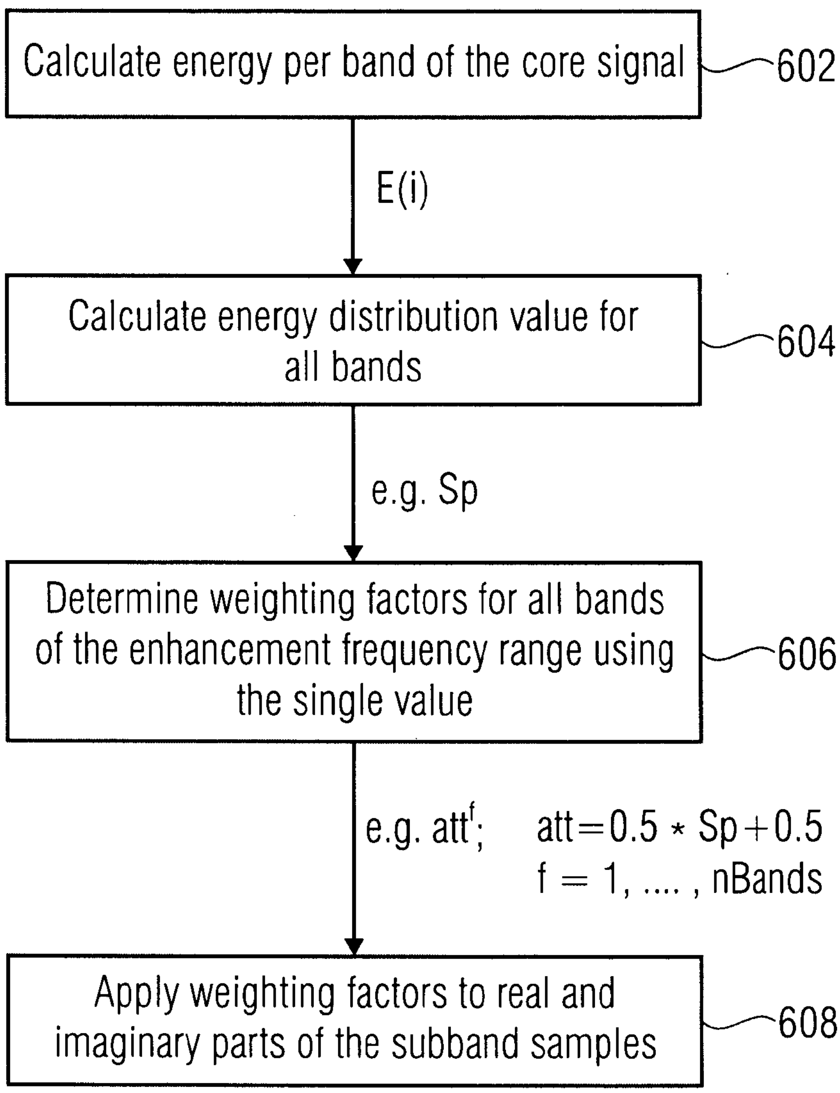
Calculate energy per band of the core signal
602
E(i)
Calculate energy distribution value for all bands
604
e.g. Sp
Determine weighting factors for all bands of the enhancement frequency range using the single value
606
e.g. att^f; att=0.5 * Sp+0.5
f = 1, .... , nBands
Apply weighting factors to real and imaginary parts of the subband samples
608

FIG 6

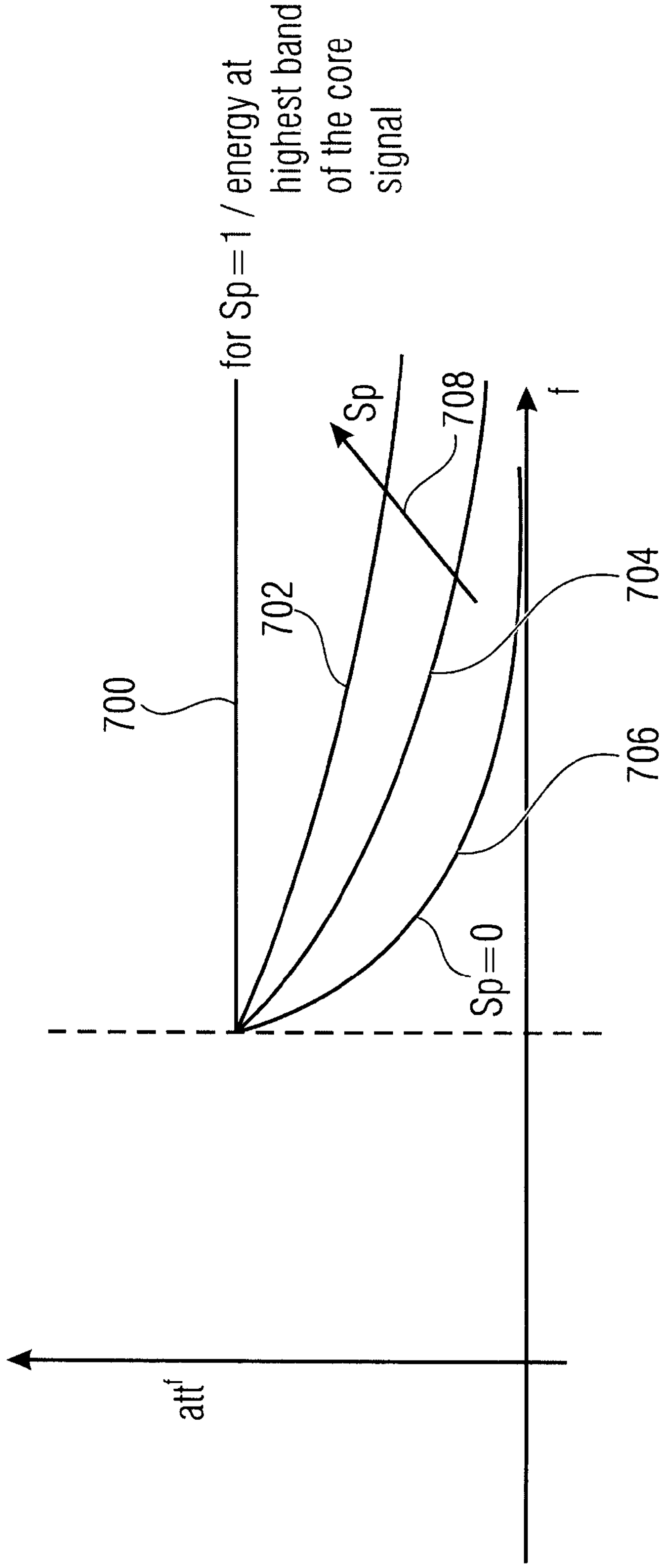
for Sp=1 / energy at highest band of the core signal
700
702
704
706
708
Sp
Sp=0
f
att^f

FIG 7

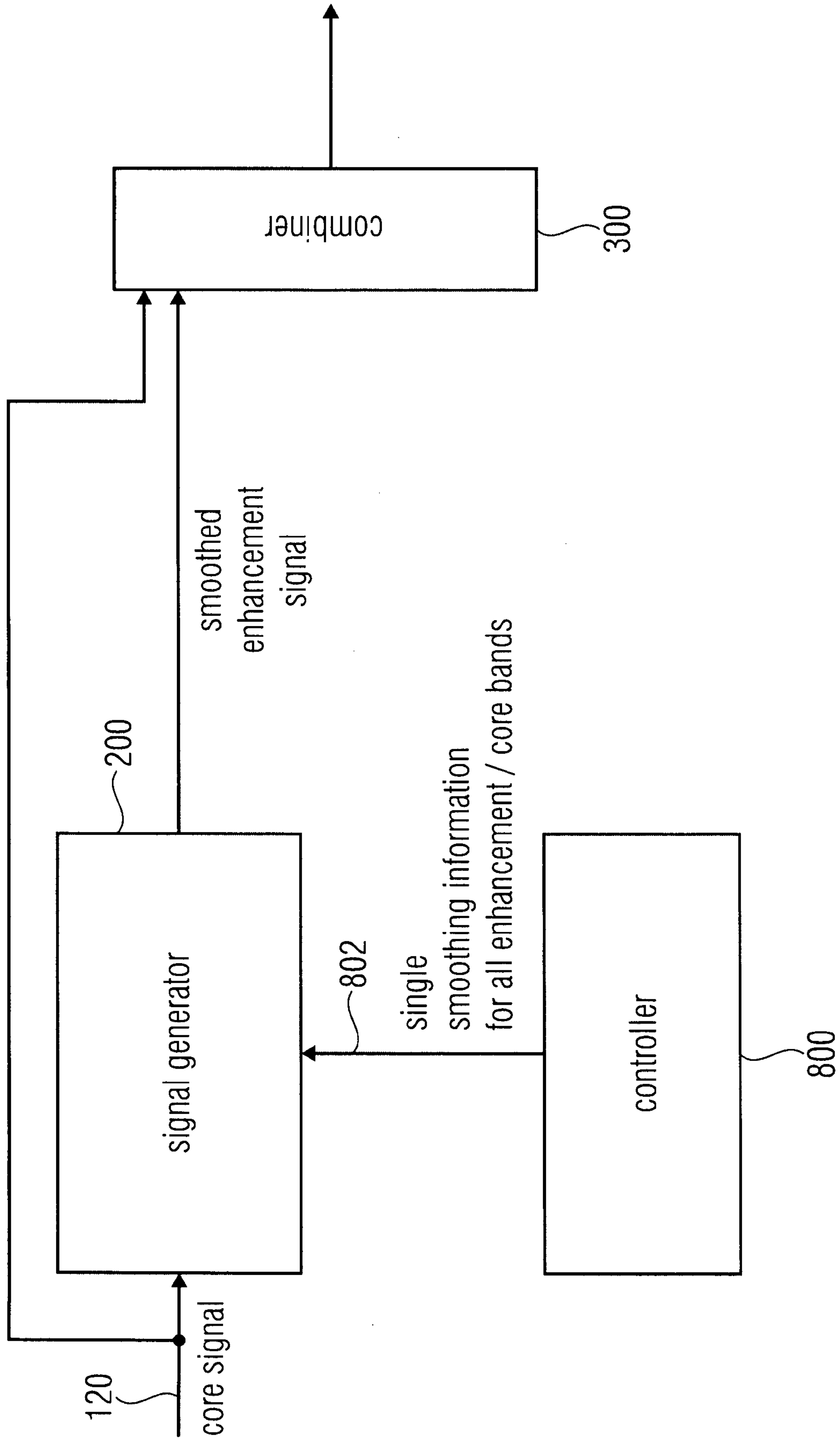
combiner
300
smoothed
enhancement
signal
200
signal generator
802
single
smoothing information
for all enhancement / core bands
controller
800
120
core signal

FIG 8

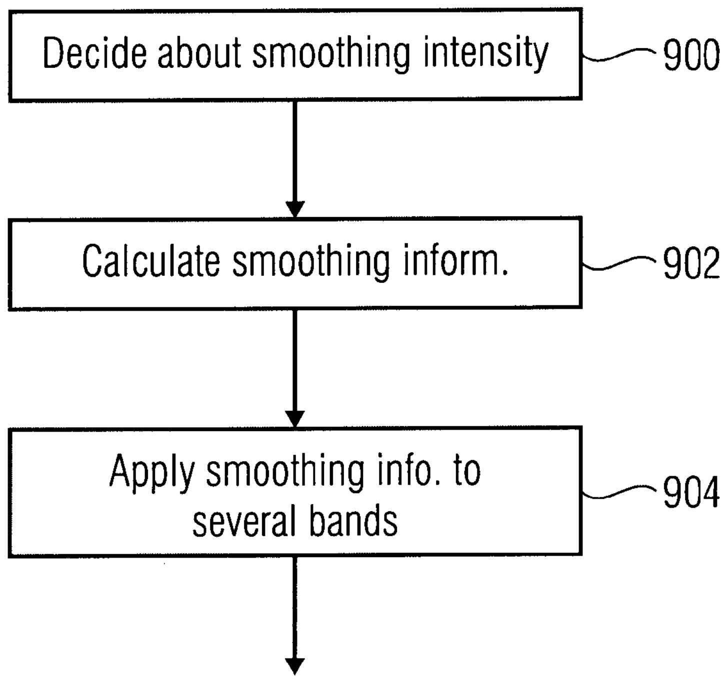
Decide about smoothing intensity
900
Calculate smoothing inform.
902
Apply smoothing info. to
several bands
904

FIG 9

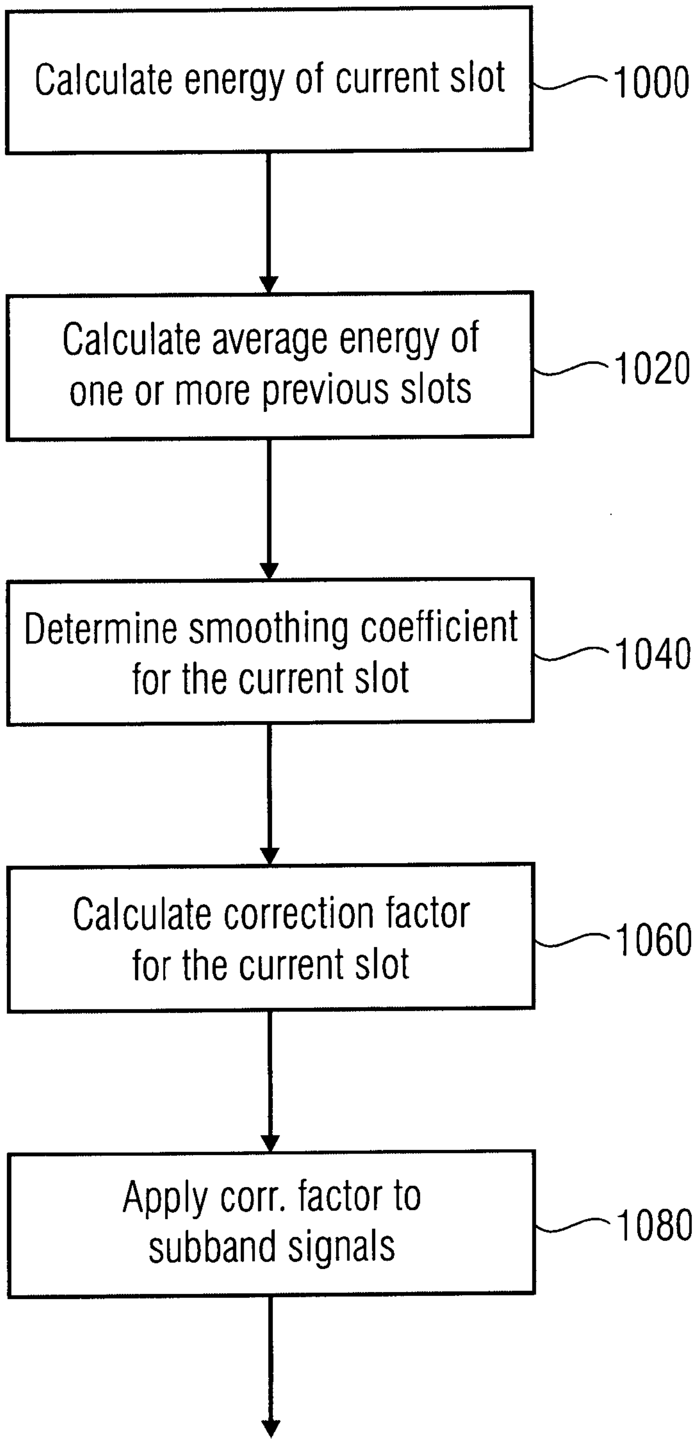
Calculate energy of current slot
1000
Calculate average energy of one or more previous slots
1020
Determine smoothing coefficient for the current slot
1040
Calculate correction factor for the current slot
1060
Apply corr. factor to subband signals
1080

FIG 10

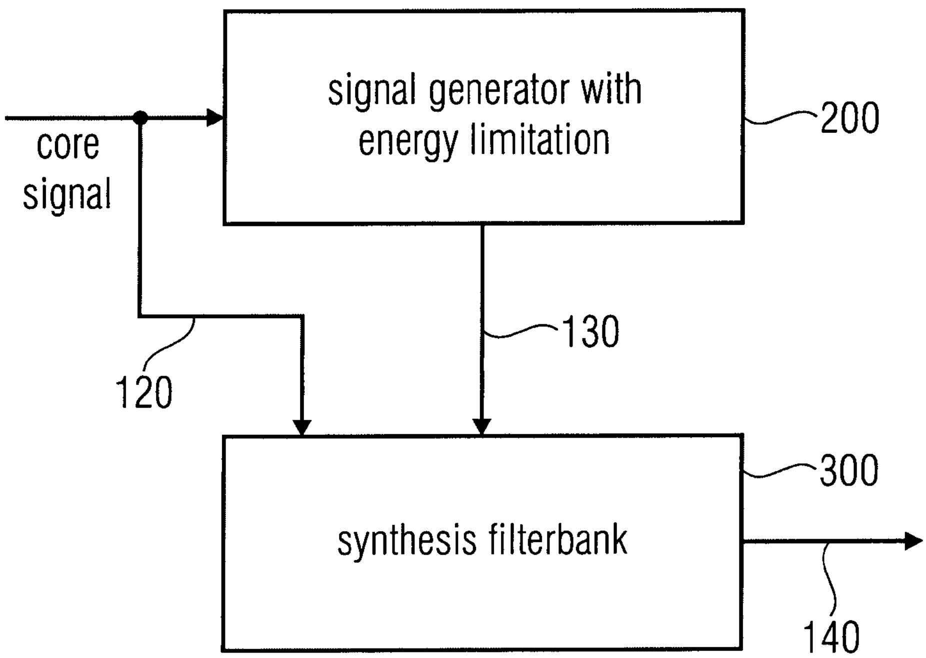
signal generator with
energy limitation
200
core
signal
130
120
300
synthesis filterbank
140

FIG 11

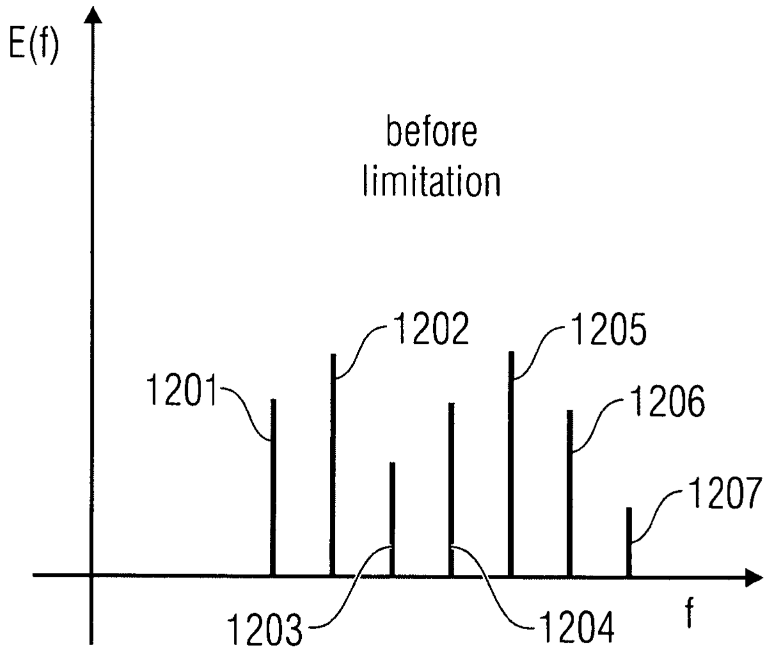


FIG 12A

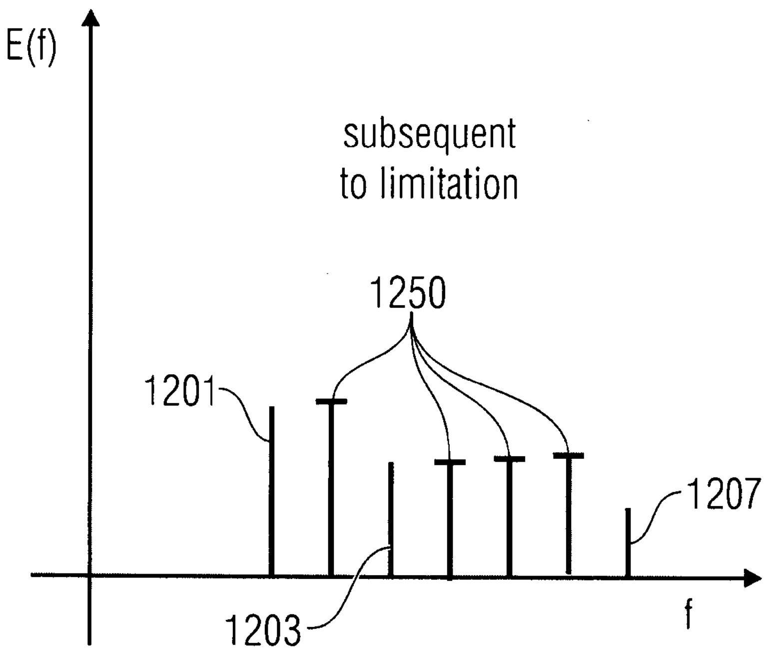


FIG 12B

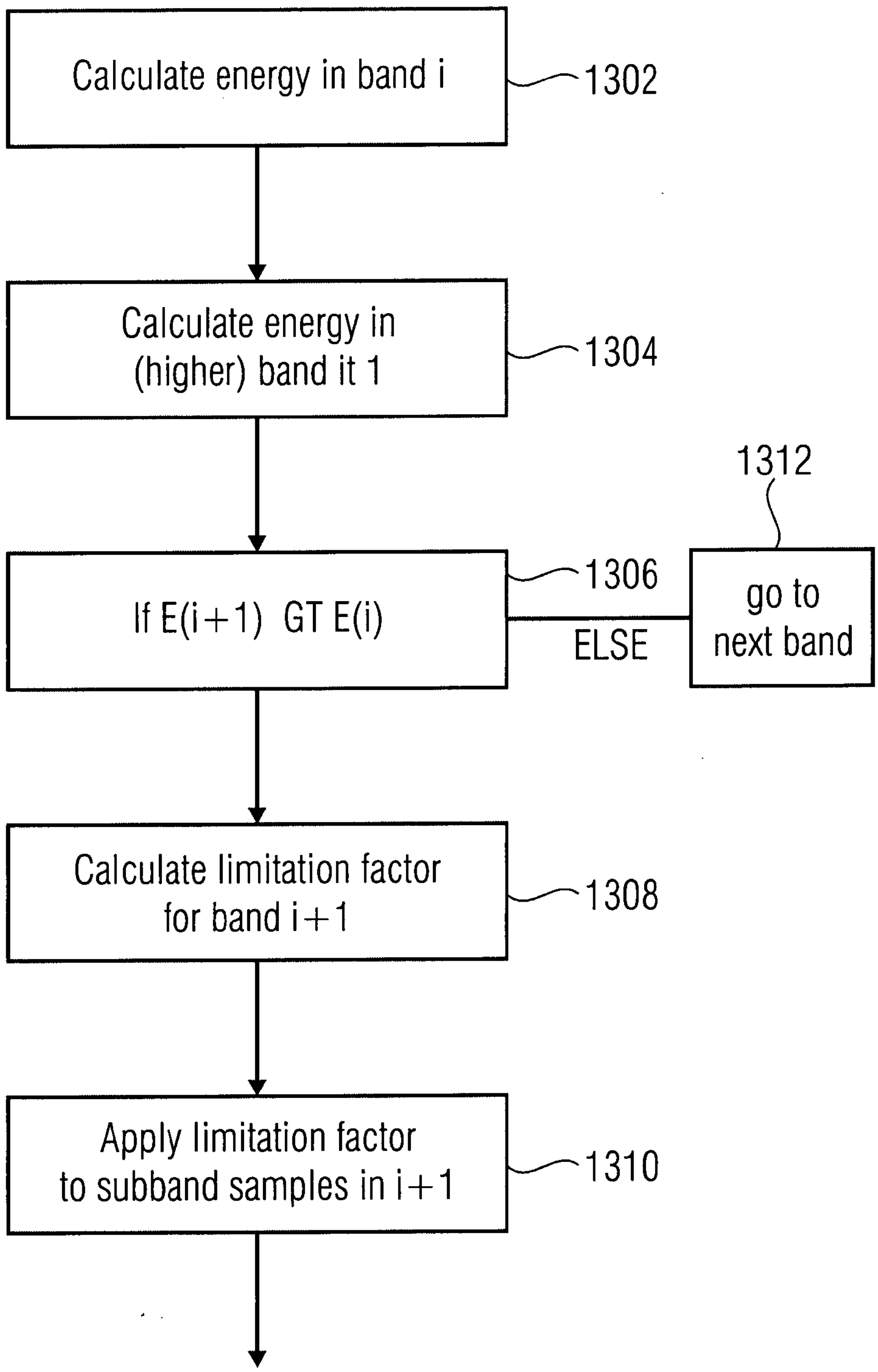
Calculate energy in band i
1302
Calculate energy in (higher) band it 1
1304
1312
1306
If E(i+1) GT E(i)
ELSE
go to next band
Calculate limitation factor for band i+1
1308
Apply limitation factor to subband samples in i+1
1310

FIG 13

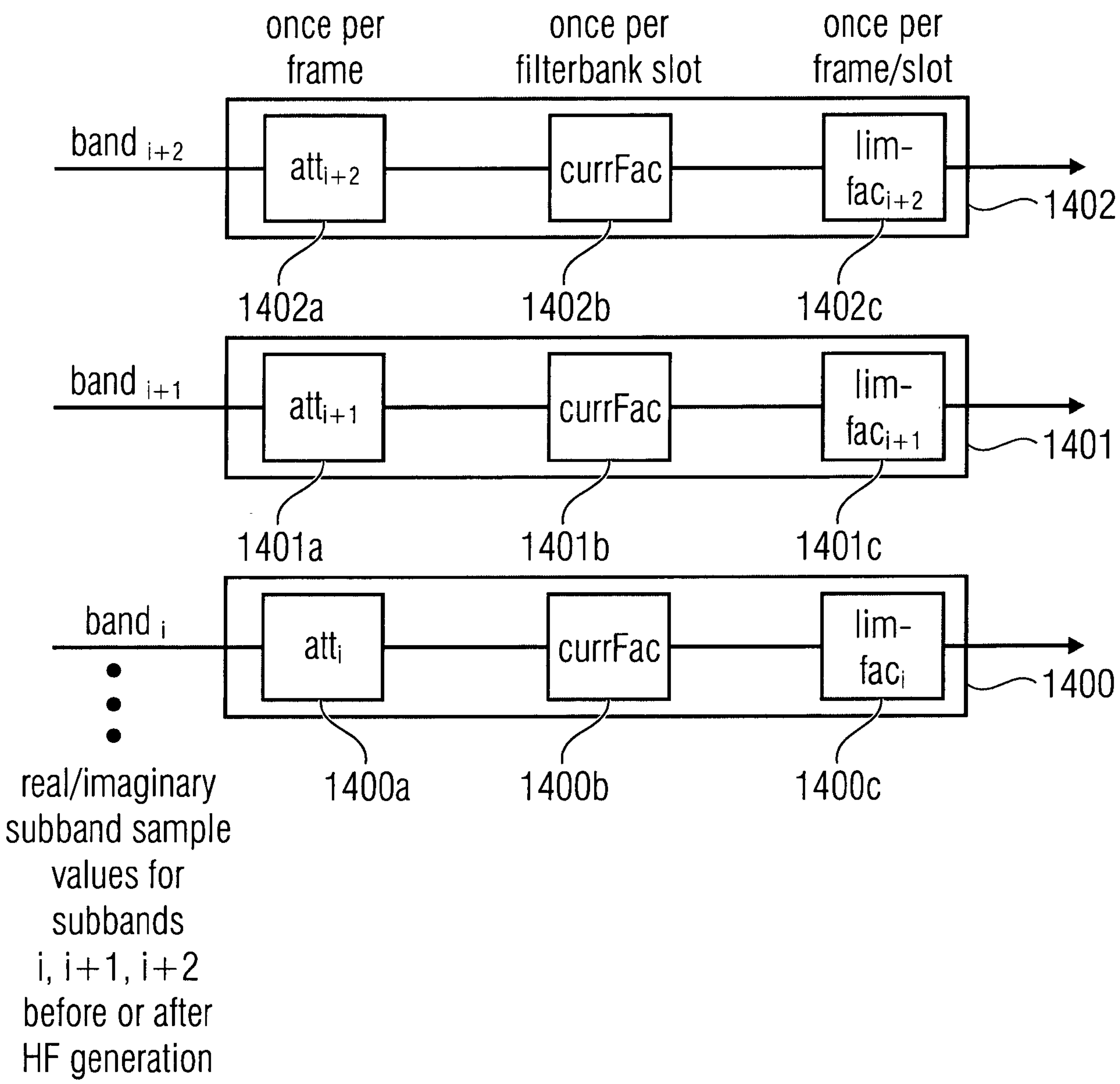
once per
frame
once per
filterbank slot
once per
frame/slot
band i+2
att i+2
currFac
lim-
fac i+2
1402
1402a
1402b
1402c
band i+1
att i+1
currFac
lim-
fac i+1
1401
1401a
1401b
1401c
band i
att i
currFac
lim-
fac i
1400
1400a
1400b
1400c
real/imaginary
subband sample
values for
subbands
i, i+1, i+2
before or after
HF generation

FIG 14

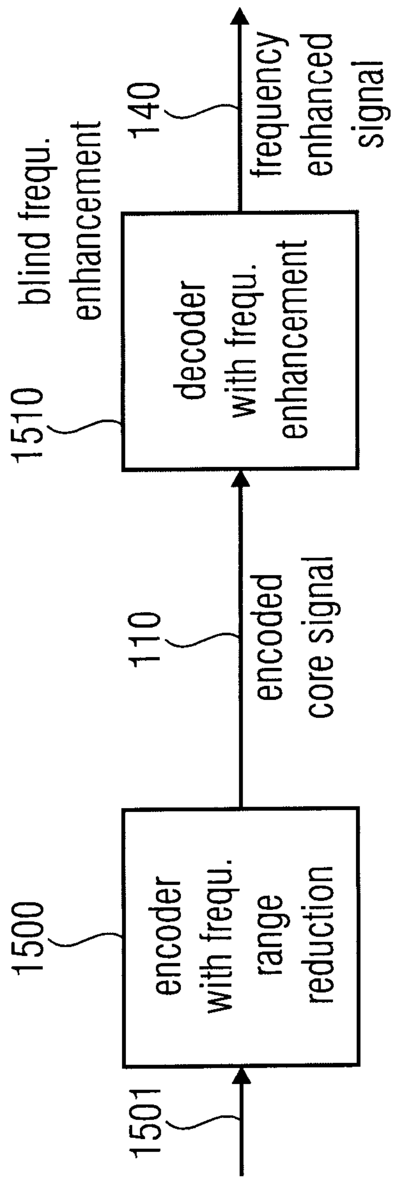
1501
1500
encoder with frequ. range reduction
110
encoded core signal
1510
decoder with frequ. enhancement
blind frequ. enhancement
140
frequency enhanced signal

FIG 15

# INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/051601

**A. CLASSIFICATION OF SUBJECT MATTER**
INV. G10L21/038
ADD.

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

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 2011/148230 A1 (NOKIA CORP [FI]; MYLLYLA VILLE MIKAEL [FI]; LAAKSONEN LAURA [FI]; PULA) 1 December 2011 (2011-12-01) | 1-8, 10-14 |
| Y | In particular pages 23, 31, 33, 34. Figures 2 and 3 | 9 |
| | ----- | |
| X | KORNAGEL ET AL: "Techniques for artificial bandwidth extension of telephone speech", SIGNAL PROCESSING, ELSEVIER SCIENCE PUBLISHERS B.V. AMSTERDAM, NL, vol. 86, no. 6, 1 June 2006 (2006-06-01), pages 1296-1306, XP024997679, ISSN: 0165-1684, DOI: 10.1016/J.SIGPRO.2005.07.039 [retrieved on 2006-06-01] | 1-8, 10-14 |
| Y | In particular Figure 1, equations 4 and 7. Sections 6.3 and 6.5 | 9 |
| | ----- | |
| | -/-- | |

[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 April 2014 | 02/05/2014 |

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

# INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/051601

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 2010/198588 A1 (SUDO TAKASHI [JP] ET AL) 5 August 2010 (2010-08-05) In particular paragraph 0138 | 1-3, 11-14 |
| X | US 2009/201983 A1 (JASIUK MARK A [US] ET AL) 13 August 2009 (2009-08-13) In particular paragraphs 0077 and 0078 | 1-3, 11-14 |
| Y | WO 2012/017621 A1 (SONY CORP [JP]; YAMAMOTO YUKI [JP]; CHINEN TORU [JP]; HATANAKA MITSUYU) 9 February 2012 (2012-02-09) In particular paragraphs 0021 to 0026 | 9 |

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2014/051601

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 2011148230 A1 | 01-12-2011 | CA 2800208 A1 | 01-12-2011 |
| | | CN 103026407 A | 03-04-2013 |
| | | EP 2577656 A1 | 10-04-2013 |
| | | KR 20130031849 A | 29-03-2013 |
| | | SG 185606 A1 | 28-12-2012 |
| | | US 2013144614 A1 | 06-06-2013 |
| | | WO 2011148230 A1 | 01-12-2011 |
| US 2010198588 A1 | 05-08-2010 | JP 4945586 B2 | 06-06-2012 |
| | | JP 2010176090 A | 12-08-2010 |
| | | US 2010198588 A1 | 05-08-2010 |
| US 2009201983 A1 | 13-08-2009 | CN 101939783 A | 05-01-2011 |
| | | EP 2238593 A1 | 13-10-2010 |
| | | KR 20100123712 A | 24-11-2010 |
| | | RU 2010137104 A | 20-03-2012 |
| | | US 2009201983 A1 | 13-08-2009 |
| | | US 2011112844 A1 | 12-05-2011 |
| | | US 2011112845 A1 | 12-05-2011 |
| | | WO 2009100182 A1 | 13-08-2009 |
| WO 2012017621 A1 | 09-02-2012 | AR 082447 A1 | 05-12-2012 |
| | | AU 2011287140 A1 | 19-04-2012 |
| | | CA 2775314 A1 | 09-02-2012 |
| | | CN 102549658 A | 04-07-2012 |
| | | CO 6531467 A2 | 28-09-2012 |
| | | EP 2471063 A1 | 04-07-2012 |
| | | JP 2012037582 A | 23-02-2012 |
| | | KR 20130107190 A | 01-10-2013 |
| | | RU 2012111784 A | 27-10-2013 |
| | | US 2013124214 A1 | 16-05-2013 |
| | | WO 2012017621 A1 | 09-02-2012 |