(19) DANMARK

Patent- og
Varemærkestyrelsen

(10) **DK/EP 3103116 T3**

(12) Oversættelse af
europæisk patentskrift

---

(51) Int.Cl.: ***G 10 L 21/038 (2013.01)***

(45) Oversættelsen bekendtgjort den: **2021-07-26**

(80) Dato for Den Europæiske Patentmyndigheds
bekendtgørelse om meddelelse af patentet: **2021-05-05**

(86) Europæisk ansøgning nr.: **15705687.0**

(86) Europæisk indleveringsdag: **2015-02-04**

(87) Den europæiske ansøgnings publiceringsdag: **2016-12-14**

(86) International ansøgning nr.: **FR2015050257**

(87) Internationalt publikationsnr.: **WO2015118260**

(30) Prioritet: **2014-02-07 FR 1450969**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **Koninklijke Philips N.V., High Tech Campus 52, 5656 AG Eindhoven, Holland**

(72) Opfinder: **RAGOT, Stéphane, Allegoat (Servel), F-Lannion 22300, Frankrig**
KANIEWSKA, Magdalena, Pakenstraat 3, 3001 LEUVEN, Belgien

(74) Fuldmægtig i Danmark: **Plougmann Vingtoft A/S, Strandvejen 70, 2900 Hellerup, Danmark**

(54) Benævnelse: **FORBEDRET UDVIDELSE AF FREKVENSBÅND I EN DEKODER FOR AUDIOFREKVENSSIGNALER**

(56) Fremdragne publikationer:
US-A1- 2001 044 722
ANNADANA RAGHURAM ET AL: "New Enhancements to the Audio Bandwidth Extension Toolkit (ABET)", AES CONVENTION 124; MAY 2008, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, 1 mai 2008 (2008-05-01), XP040508704,

IMPROVED FREQUENCY BAND EXTENSION IN AN AUDIO SIGNAL DECODER

The present invention relates to the field of coding/decoding and to the processing of audio-frequency signals (such as speech, music or other signals) for transmission or storage thereof.

More particularly, the invention concerns a method and a device for extending the frequency band in a decoder or processor to obtain an improved audio-frequency signal.

Numerous techniques are available for compressing (with loss) an audio-frequency signal such as speech or music.

Conventional coding methods for conversational applications are generally classified into *waveform coding*: Pulse Code Modulation (PCM), Adaptive Differential Pulse Code Modulation (ADPCM), transform coding…), *parametric coding* (Linear Predictive Coding (LPC), sinusoidal coding…) and hybrid parametric coding with quantification of parameters using “analysis-by-synthesis” for which Code Excited Linear Prediction (CELP) is the best-known example.

For non-conversational applications, (mono) audio signal coding in the prior art is composed of perceptual transform or sub-band coding, with parametric coding of high frequencies by Spectral Band Replication (SBR).

A review of conventional speech and audio coding methods can be found in the works by W.B. Kleijn and K.K. Paliwal (Eds.), Speech Coding and Synthesis, Elsevier, 1995; M. Bosi, R. E. Goldberg, Introduction to Digital Audio Coding and Standards, Springer 2002; J. Benesty, M.M. Sondhi, Y. Huand (Eds.), Handbook of Speech Processing, Springer 2008.

Particular attention is focused herein on the 3GPP standardized codec (coder and decoder) AMR-WB (Adaptive Multi-Rate Wideband) which operates at an input/output frequency of 16 kHz and in which the signal is divided into two sub-bands, the low band (0-6.4 kHz) sampled at 12.8 kHz and coded by the CELP model, and the high band (6.4-7 kHz) which is reconstructed parametrically by Bandwidth Extension (BWE) with or without additional information depending on the current frame mode. It can be noted here that limitation of the coded band of the

AMR-WB codec to 7 kHz is essentially related to the fact that the transmission frequency response for wideband terminals was approximated at the time of standardization (ETSI/3GPP then UIT-T) according to the frequency mask defined in standard UIT-T P.341, and more specifically using a so-called “P341” filter defined in standard UIT-T G.191 which cuts off frequencies above 7 kHz (this filter complies with the mask defined in UIT-T P.341). However, in theory, it is well known that a signal sampled at 16 kHz can have a defined audio band of 0 to 8000 Hz; the AMR-WB codec therefore introduces limitation of the high band in comparison with the theoretical bandwidth of 8 kHz.

The 3GPP AMR-WB speech codec was standardized in 2001 chiefly for telephone applications in circuit switched mode (CS) on GSM (2G) and UMTS (3G). This same codec was also standardized in 2003 by UIT-T as G.722.2 recommendation “Wideband coding speech at around 16 kbits/s using Adaptive Multi-Rate Wideband (AMR-WB)”.

It comprises nine bit rates called modes of 6.6 to 23.85 kbits/s, and comprises Discontinuous Transmission mechanisms (DTX) with Voice Activity Detection (VAD) and Comfort Noise Generation (CNG) from Silence Insertion Descriptor (SID) frames, and mechanisms for Frame Erasure Concealment (FEC) sometimes called Packet Loss Concealment (PLC).

The details will not be reproduced here of the AMR-WB coding and decoding algorithm, a detailed description of this codec can be found in 3GPP specifications (TS 26.190, 26.191, 26.192, 26.193, 26.194, 26.2024) and UIT-T specification G.722.2 (and the corresponding Annexes and Appendixes), and in the article by B. Bessette *et al* “The adaptive multirate wideband speech codec (AMR-WB)”, IEEE Transactions on Speech and Audio Processing, vol. 10, No. 8, 2002, pp. 620-636, and the source codes of associated standards 3GPP and UIT-T.

The principle of bandwidth extension in the AMR-WB codec is fairly rudimentary. The high band (6.4-7 kHz) is generated by shaping a white noise via a time envelope (applied in the form of gains per sub-frame) and frequency envelope (by application of a Linear Predictive Coding (LPC) synthesis filter). This bandwidth extension technique is illustrated in **Figure 1**.

White noise $u_{HB1}(n)$, $n = 0,\ldots,79$ is generated at 16 kHz per 5 ms sub-frame by a linear congruential generator (block 100). This noise $u_{HB1}(n)$, is shaped in the time by application of sub-frame gains; this operation is decomposed into two processing steps (blocks 102, 106 or 109):

- A first factor is calculated (block 101) to place white noise $u_{HB1}(n)$ (block 102) at a similar level to that of excitation $u(n)$, $n = 0,\cdots, 63$, decoded at 12.8 kHz in the low band:

$$u_{HB2}(n) = u_{HB1}(n)\sqrt{\frac{\sum_{l=0}^{63} u(l)^2}{\sum_{l=0}^{79} u_{HB1}(l)^2}}$$

It can be noted here that energy normalization is obtained by comparing blocks of different size (64 for $u(n)$ and 80 for $u_{HB1}(n)$) without compensation for differences in sampling frequencies (12.8 or 16 kHz).

- Excitation in the high band is then obtained (block 106 or 109) in the form:

$$u_{HB}(n) = \hat{g}_{HB} u_{HB2}(n)$$

where the gain $\hat{g}_{HB}$ is obtained differently depending on bit rate. If the bit rate of the current frame is <23.85 kbits/s, the gain $\hat{g}_{HB}$ is given "blind" estimation (i.e. without additional information); in this case, block 103 filters the low band decoded signal with a high-pass filter having a cut-off frequency at 400 Hz to obtain a signal $\hat{s}_{hp}(n), n=0,\ldots,63$ - this high-pass filter removes the influence of the very low frequencies which can distort the estimation calculated at block 104 – after which the "tilt" is calculated (spectral slope indicator), denoted e_{tilt}, of the signal $\hat{s}_{hp}(n)$,via normalized autocorrelation (block 104):

$$e_{tilt} = \frac{\sum_{n=1}^{63} \hat{s}_{hp}(n)\hat{s}_{hp}(n-1)}{\sum_{n=0}^{63} \hat{s}_{hp}(n)^2}$$

and finally $\hat{g}_{HB}$ is calculated in the form:

$$\hat{g}_{HB} = w_{SP} g_{SP} + (1 - w_{SP}) g_{BG}$$

where $g_{SP} = 1\text{-}e_{tilt}$ is the gain applied in the active speech frames (SP), $g_{BG} = 1.25 g_{SP}$ is the gain applied in the inactive speech frames associated with background noise (BG) and W_{SP} is a weighting function dependent on vocal activity detection (VAD). It will be understood that estimation of tilt (e_{tilt}) allows adapting of the level of the high band as a function of the spectral nature of the signal; this estimation is of particular importance when the spectral slope of the CELP decoded signal is such that average energy decreases when frequency increases (the case of a voiced signal when e_{tilt} is close to 1, hence $g_{SP}=1\text{-}e_{tilt}$ is reduced). It is also to be noted that the factor $\hat{g}_{HB}$ in AMR-WB decoding is limited to take values in the range [0.1, 1.0]. For signals with a spectrum having greater energy in high frequencies (e_{tilt} close to -1, g_{SP} close to 2), the gain $\hat{g}_{HB}$ is usally underestimated.

At 23.85 kbits/s, corrective information is transmitted by the AMR-WB coder and decoded (blocks 107. 108) to finetune the estimated gain per sub-frame (4 bits every 5 ms, i.e. 0.8 kbits/s.

Artificial excitation $u_{HB}^{(n)}$ is then filtered (block 111) with an LPC synthesis filter having a transfer function $1/A_{HB}(z)$ and operating at the sampling frequency of 16 kHz. The obtaining of this filter is dependent on the bit rate of the current frame:

- At 6.6 kbits/s, the filter $1/A_{HB}(z)$ is obtained by weighting, by a factor of γ=0.9, a 20th order LPC filter $1/\hat{A}^{ext}(z)$ which “extrapolates” the 16th order LPC filter $1/\hat{A}(z)$ decoded in the low band (at 12.8 kHz) – the details of extrapolation in the domain of ISF parameters (Immittance Spectral Frequency) are described in standard G.722.2 under section 6.3.2.1; in this case:

$$1/A_{HB}(z) = 1/\hat{A}^{ext}(z/\gamma)$$

- At bit rates > 6.6 kbits/s, the filter $1/A_{HB}(z)$ is of 16th order and simply corresponds to:

$$1/A_{HB}(z) = 1/\hat{A}(z/\gamma)$$

where $\gamma = 0.6$. It is to be noted that in this case the filter $1/\hat{A}(z/\gamma)$ is used at 16 kHz, which results in spreading (via homothety) of the frequency response of this filter from [0, 6.4 kHz] to [0, 8 kHz].

The result $S_{HB}(n)$ is finally processed with a bandpass filter (block 112) of FIR type (Finite Impulse Response) so as only to maintain the 6-7 kHz band; at 23.85 kbits/s a low-pass filter also of FIR type (block 113) is added to processing for further attenuation of the frequencies higher than 7 kHz. High frequency (HF) synthesis is finally added (block 130) to the low frequency (BF) synthesis obtained at blocks 120 to 123 and re-sampled at 16 kHz (block 123). Therefore, even if the high band theoretically extends from 6.4 to 7 kHz in the AMR-WB codec, HF synthesis is rather more included in the 6-7 kHz band before addition to low frequency synthesis BF.

Several disadvantages can be identified in the bandwidth extension technique of the AMR-WB codec:

- The signal in the high band is shaped white noise (by sub-frame time gains, by $1/A_{HB}(z)$ filtering and bandpass filtering), which is not a good general model of the signal in the 6.4-7 kHz band. For example, there are highly harmonic music signals for which the 6.4-7 kHz band contains sinusoidal components (or tones) and no noise (or little noise); for these signals, the band extension of the AMR-WB codec strongly degrades quality.
- The low-pass filter at 7 kHz (block 113) introduces a shift of nearly 1 ms between the low and high bands, which can potentially degrade the quality of some signals by slightly de-synchronizing the two bands at 23.85 kbits/s – this de-synchronization can also cause a problem when switching from the bit rate of 23.85 kbits/s to other modes.
- The estimation of gains per sub-frame (blocks 101, 103 to 105) is not optimal. It is based in part on equalization of “absolute” energy per sub-frame (block 101) between signals at different frequencies: artificial equalization at 16 kHz (white noise) and a signal at 12.8 kHz (decoded ACELP excitation). It can be noted in particular that this approach implicitly induces attenuation of high band excitation (by a ratio of 12.8/16 = 0.8); it will also be noted that no

deaccentuation (or de-emphasis) is performed on the high band in the AMR-WB codec, which implicitly induces relative amplification close to 0.6 (which corresponds to the value of the frequency response of $1/(1\text{-}0.68z^{-1})$ at 6400 Hz).

In fact the factors of 1/0.8 and 0.6 approximately offset each other:

- For speech, characterization tests of the 3GPP AMR-WB codec, documented in the 3GPP TR 26.976 report, showed that the mode at 23.85 kbits/s has lesser quality than at 23.05 kbits/s, the quality thereof in fact being similar to that of the mode at 15.85 kbits/s. This particularly shows that the level of the artificial HF signal must be very carefully controlled since quality is degraded at 23.85 kbits/s, whereas the 4 bits per frame are supposed to allow closer approaching of the energy of the original high frequencies.
- Limitation of the coded band to 7 kHz results from application of a strict model of transmission response by acoustic terminals (P.341 filter in the UIT-T G.191 standard). Yet, for a sampling frequency of 16 kHz, the frequency content in the 7-8 kHz band remains high in particular for music signals to ensure a good quality level.

The AMR-WB decoding algorithm was partly improved with the development of the scalable UIT-T G.718 codec which was standardized in 2008.

The UIT-T G.718 standard comprises a so-called interoperable mode, for which core coding is compatible with G.722.2 (AMR-WB) coding at 12.65 kbits/s; in addition, the G.718 decoder has the particular aspect of being able to decode an AMR-WB/G.722.2 bitstream at all the possible bit rates of the AMR-WB codec (from 6.6 to 23.85 kbits/s).

The interoperable G.718 decoder in low delay mode (G.718-LD) is illustrated in **Figure 2**. Listed below are the improvements to the operating functions of AMR-WB bitstream decoding brought by the G.718 decoder, with reference to Figure 1 whenever necessary:

Bandwidth extension (described for example in clause 7.13.1 of recommendation G.718, block 206) is identical to that of the AMR-WB decoder except that the 6-7 kHz low-pass filter and 1/AHB(z) synthesis filter (blocks 111 and 112) are in reverse order. In addition, at 23.85 kbits/s the 4 bits transmitted per sub-frame by the AMR-

WB are not used in the interoperable G.718 decoder; synthesis of the high frequencies (HF) at 23.85 kbits/s is therefore the same as at 23.05 kbits/s which avoids the known problem of AMR-WB decoding quality at 23.85 kbits/s. *A fortiori*, the low-pass filter at 7 kHz (block 113) is not used, and specific decoding of the mode at 23.85 kbits/s is omitted (blocks 107 to 109).

Post-processing of synthesis at 16 kHz (see clause 7.14 of G.718) is implemented in G.718 by a noise gate in block 208 (to “improve” silence quality via level reduction), by high-pass filtering (block 209), bass post-filtering in block 210 attenuating low-frequency inter-harmonic noise and conversion to 16-bit integers with saturation control (automatic gain control – AGC) in block 211.

However, bandwidth extension in the AMR-WB codec and/or G.718 codec (interoperable mode) still remains limited in several aspects.

In particular, synthesis of high frequencies by shaped white noise (via a time approach of LPC source-filter type) is a very limited model of the signal in the band of frequencies higher than 6.4 kHz.

Only the 6.4-7 kHz band is artificially re-synthesized whereas in practice a wider band (up to 8 kHz) is theoretically possible at the sampling frequency of 16 kHz, which can potentially improve signal quality if signals are not pre-processed with a filter of P.341 type (50-7000 Hz) such as defined in the UIT-T Software Tool Library (standard G.191).

New Enhancements to the Audio Bandwidth Extension Toolkit (ABET) is an article presenting a set of tools for acoustic signal processing and some possible improvements. ABET comprises a method of Accurate Spectral Replacement (ASR), a Fractal Self Similarity Model (FSSM) and Multi-Band Temporal Amplitude Coding (MBTAC).

There is therefore a need to improve band extension in a codec of AMR-WB type or an interoperable version of this codec, or more generally to improve band extension of an audio signal in particular to improve the frequency content of band extension.

The present invention sets out to improve the situation.

For this purpose, the invention proposes a method for extending the frequency band of an audio-frequency signal in a decoding or improvement process, comprising a step of obtaining the decoded signal in a first frequency band referred to as low band. The method is such that it comprises the steps of claim 1.

It will be noted that in the remainder hereof “band extension” is taken in its broad meaning and includes not only the case of extending a sub-band into high frequencies, but also the case of replacing sub-bands set to zero (of noise filling type in transform coding).

Therefore, the taking into account of tonal components and of an ambient signal extracted from a signal originating from the decoded low band signal, allows band extension to be performed with a signal model adapted to the true nature of the signal contrary to the use of an artificial noise. The quality of band extension is thereby improved, and in particular for certain types of signals such as music signals.

The decoded signal in the low band comprises a portion corresponding to ambient sounds which can be transposed into high frequency so that mixing of the harmonic components and existing ambience ensures a coherent reconstructed high band.

It will be observed that even if the invention is prompted by improving the quality of band extension in the context of interoperable AMR-WB coding, the different embodiments apply to the more general case of bandwidth extension of an audio signal, in particular in an improvement device performing analysis of the audio signal to extract the parameters needed for bandwidth extension.

The different particular embodiments mentioned above can be added independently or in combination with each other, at the steps of the extension method defined above.

In one embodiment, bandwidth extension is obtained in the excitation domain and the decoded low band signal is a decoded low band excitation signal.

The advantage of this embodiment is that transforming without windowing (or in equivalent manner with an implicit rectangular window having the length of the frame) is possible in the excitation domain. In this case, no artefact (block effects) is audible.

The extraction of tonal components and of the ambient signal is carried out with the following steps:

- obtaining the ambient signal by calculating an average value of the spectrum of the decoded, or decoded and extended low band signal;
- obtaining the tonal components by subtracting the calculated ambient signal from the decoded, or decoded and extended low band signal.

In one embodiment of the combination step, an energy-level control factor used for adaptive mixing is calculated as a function of the total energy of the decoded, or decoded and extended low band signal and of the tonal components.

The application of this control factor, at the combination step, allows adapting to the characteristics of the signal to optimise the relative proportion of the ambient signal in the mix. The energy level is therefore controlled to prevent audible artefacts.

The decoded low band signal undergoes a decomposition step into sub-bands by transform or filter bank, the extraction and combination steps then taking place in the frequency domain or in sub-bands.

The application of bandwidth extension in the frequency domain allows finetuning of frequency analysis, which is not possible with a time approach, and also provides sufficient frequency resolution to detect the tonal components.

In one detailed embodiment, the decoded and extended low band signal is obtained with the following equation:

$$U_{HB1}(k) = \begin{cases} 0 & k = 0, \cdots, 199 \\ U(k) & k = 200, \cdots, 239 \\ U(k + start\_band - 240) & k = 240, \cdots, 319 \end{cases}$$

with *k* the sample index, *U(k)* the spectrum of the signal obtained after a transform step, $U_{HB}1^{(k)}$ the spectrum of the extended signal, and *start_band* a predefined variable. Therefore, this function comprises re-sampling of the signal by adding samples to the spectrum of this signal. Other methods of extending the signal are possible however, for example by translation in sub-band processing.

The present invention also concerns a device for extending the frequency band of an audio-frequency signal, the signal having been decoded in a first frequency band referred to as low band. The device is such that it has the characteristics of claim 6.

This device has the same advantages as the previously described method which it implements.

The invention concerns a decoder comprising a device such as described.

It concerns a computer programme comprising code instructions for the implementation of the steps of the frequency band extension method such as described, when these instructions are executed by a processor.

Finally, the invention relates to a storage medium readable by a processor whether or not integrated in the band extension device, optionally removable, memorising a computer programme implementing a band extension method such as previously described.

Other characteristics and advantages of the invention will become more clearly apparent on reading the following description given solely as a nonlimiting example with reference to the appended drawings in which:

- Figure 1 illustrates part of a decoder of AMR-WB type implementing steps of prior art frequency band extension such as previously described;
- Figure 2 illustrates a decoder of interoperable type G.718-LD at 16 kHz in the prior art and such as previously described;
- Figure 3 illustrates an interoperable decoder with AMR-WB coding and integrating a band extension device according to one embodiment of the invention;
- Figure 4 in the form of a flow chart illustrates the main steps of a band extension method according to one embodiment of the invention;
- Figure 5 illustrates one embodiment in the frequency domain of a band extension device according to the invention integrated in a decoder; and
- Figure 6 illustrates a physical embodiment of a band extension device of the invention.

Figure 3 illustrates an example of a decoder compatible with the AMR-WB/G.722.2 standard, providing post-processing similar to that introduced by G.718 described with reference to Figure 2, and improved band extension according to the extension method of the invention implemented by the band extension device illustrated in block 309.

Contrary to AMR-WB decoding which operates with an output sampling frequency of 16 kHz, and to G.718 decoding which operates at 8 or 16 kHz, consideration is given here to a decoder which can operate with an output (synthesis) signal at the frequency *fs* = 8, 16, 32 or 48 kHz. It is to be noted that it is assumed here that coding was performed according to the AMR-WB algorithm with an internal frequency of 12.8 kHz for low band CELP coding, and at 23.85 kbits/s for gain coding per sub-frame at the frequency of 16 kHz, but interoperable variants of the AMR-WB coder are also possible; even if the invention is described herein with respect to decoding, it is assumed that coding can also operate with an input signal at frequency *fs* = 8, 16, 32 or 48 kHz, and adequate re-sampling operations beyond the scope of the invention are applied to coding as a function of the value of *fs*. It can be noted that when *fs*= 8 kHz at the decoder, in the event of decoding compatible with AMR-WB, it is not necessary to extend the low band 0-6.4 kHz, since the audio band reconstructed at frequency *fs* is limited to 0-4000 Hz.

In Figure 3, CELP decoding (BF for low frequencies) always operates at the internal frequency of 12.8 kHz,as in AMR-WB and G.718, and band extension (HF for high frequencies) subject of the invention operates at the frequency of 16 kHz, the syntheses BF and HF are combined (block 312) at frequency *fs* after adequate re-sampling (blocks 307 and 311). In some variants of the invention, the combining of the low and high bands can be made at 16 kHz after re-sampling the low band from 12.8 to 16 kHz, before re-sampling the combined signal at frequency *fs*.

The decoding in Figure 3 is dependent on the AMR-WB mode (or bit rate) associated with the received current frame. By way of indication and without impacting block 309, decoding of the CELP portion in the low band comprises the following steps:

- Demultiplexing the coded parameters (block 300) if frame is correctly received (*bfi*=0 where *bfi* is the “bad frame indicator” of value 0 for a received frame and 1 for a lost frame).
- Decoding ISF parameters with interpolation and conversion to LPC coefficients (block 301) as described in clause 6.1 of standard G.722.2.
- Decoding CELP excitation (block 302) with a fixed, adaptive part to reconstruct excitation (exc or *u’(n))* in each sub-frame of length 64 at 12.8 kHz

$$u'(n) = \hat{g}_p v(n) + \hat{g}_c c(n) \;,\; n = 0, \cdots, 63$$

following the notations of clause 7.1.2.1. in G.718 concerning CELP decoding, where v(n) and *c(n)* are respectively the codes of the adaptive and fixed codebook, and $\hat{g}_p$ and $\hat{g}_c$ are the associated decoded gains. This excitation *u'(n)* is used in the adaptive codebook of the following sub-frame; it is then post-processed and, as in G.718, a distinction is made between excitation *u'(n)* (also denoted exc) and its modified post-processed version *u(n)* (also denoted exc2) which is used as input to the synthesis filter 1/*Â(z)* in block 303. In variants, which can be implemented for the invention, the post-processing operations applied to excitation can be modified (for example phase dispersion can be improved) or these post-processing operations can be extended (for example inter-harmonic noise can be reduced) without impacting the inherent characteristics of the band extension method of the invention.

- Synthesis filtering by 1/*Â(z)* (block 303) where the decoded LPC filter Â(z) is of 16th order.
- Narrow band post-processing (block 304) in accordance with clause 7.3 of G.718 if *fs*=8 kHz.
- Deemphasis (block 305) by the filter $1/(1\text{-}0.68^{z^{-1}})$
- Post -processing of the low frequencies (block 306) such as described in clause 7.14.1.1. of G.718. This processing introduces a delay which is taken into account in decoding of the high band (>6.4 kHz).
- Re-sampling of the internal frequency from 12.8 kHz to the output frequency *fs* (block 307). Several methods are possible. Without loss of generality, it is considered here as an example that if *fs*=8 or 16 kHz, the re-sampling described in clause 7.6 of G.718 is reproduced here, and if *fs*=32 or 48 kHz additional Finite Impulse Response filters (FIRs) are used.
- Calculation of the noise gate parameters (block 308) preferably performed as described in clause 7.14.3 of G.718.

In variants which can be implemented for the invention, the post-processing operations applied to excitation can be modified (for example phase dispersion can be improved) or such post-processing can be extended (for example inter-harmonic noise can be reduced) without impacting the inherent characteristics of band extension. The description is not given here of low band decoding when the current frame is lost (*bfi*=1) which is informative in standard 3GPP AMR-WB; in general, whether for the AMR-WB decoder or a general decoder based on the source-filter model, the purpose is typically to best estimate LPC excitation and the coefficients of the LPC synthesis filter in order to reconstruct the lost signal whilst maintaining the source-filter model. When *bfi*=1, it is considered here that band extension (block 309) can function as in the case *bfi*=0 and bit rate < 23.85 kbits/s; therefore, it is assumed hereafter in the description of the invention without loss of generality that *bfi*=0.

It can be noted that the use of blocks 306, 308, 314 is optional.

It will also be noted that the low band decoding described above assumes a so-called “active” current frame with a bit rate between 6.6 and 23.85 kbits/s. When the DTX mode (discontinuous transmission) is activated, some frames can be coded as “inactive” and in this case either a silence descriptor can be transmitted (on 35 bits) or nothing is transmitted. In particular, it is recalled that the SID frame of the AMR-WB coder describes several parameters: ISF parameters averaged over 8 frames, average energy over 8 frames, dither flag for reconstruction of non-stationary noise. In all cases, at the decoder, the same decoding model is found as for an active frame, with reconstruction of excitation and of an LPC filter for the current frame, which allows the invention to be applied even to inactive frames. The same applies to the decoding of “lost frames” (or FEC, PLC) for which the LPC model is applied.

This decoder example functions in the excitation domain and therefore comprises a step to decode the low band excitation signal. The band extension device and band extension method in the meaning of the invention also function in a domain differing from the excitation domain, and in particular with a low band decoded direct signal or a signal weighted with a perceptual filter.

Contrary to AMR-WB or G.718 decoding, the describe decoder allows extension of the decoded low band (50-6400 Hz taking into account high-pass filtering at 50 Hz at the decoder, 0-6400 Hz in the general case) into an extended band of varying width, approximately ranging from 50-6900 Hz to 50-7700 Hz as a function of the mode implemented in the current frame. It can then be spoken in terms of a first frequency band of 0 to 6400 Hz and a second frequency band of 6400 to 8000 Hz. In reality, in the preferred embodiment, excitation for the high frequencies is generated in the frequency domain in a band of 5000 to 8000 Hz to allow bandpass filtering of width 6000 to 6900 or 7700 Hz having a slope that is not too steep in the rejected upper band.

The high band synthesis part is performed in block 309 illustrating the band extension device of the invention and which is detailed in Figure 5 in one embodiment.

To align the decoded low and high bands, a delay (block 310) is inserted to synchronize the outputs of blocks 306 and 309, and the high band synthesized at 16 Hz is re-sampled from 16 Hz to the frequency *fs* (output of block 311). The value of the delay *T* must be adapted for the other cases (*fs*=32, 48 kHz) as a function of the processing applied. It is recalled that when *fs*=8 kHz it is not necessary to apply blocks 309 to 311 since the band of the signal output by the decoder is limited to 0-4000 Hz.

It is to be noted that the extension method of the invention implemented in block 309 according to the first embodiment preferably does not introduce any additional delay in relation to the low band reconstructed at 12.8 kHz; however, in variants of the invention (for example using time/frequency transforming with overlap) a delay could be introduced. Therefore, in general, the value of *T* in block 310 is to be adjusted as a function of specific implementation. For example, if post-processing of the low frequencies (block 306) is not used, the delay to be introduced for *fs*=16 kHz can be set at *T*=15.

The low and high bands are then combined (added) in block 312, and the synthesis obtained is post-processed by high-pass filtering at 50 Hz (of IIR type) of 2nd order, the coefficients of which are dependent on frequency *fs* (block 313), and output post-processing with optional application of the noise gate in similar fashion to G.718 (block 314).

The band extension device of the invention, illustrated by block 309 in the decoder embodiment in Figure 5, performs a band extension method (in the broad meaning) that is now described with reference to **Figure 4**.

This extension device can also be independent of the decoder and can implement the method described in Figure 4 to perform band extension of an existing audio signal stored or transmitted to the device, with analysis of the audio signal to extract therefrom an excitation and an LPC filter for example.

At its input, this device receives a decoded signal in a first frequency band called low band *u(n)* which can be in the excitation domain or in the signal domain. In the embodiment described here, a decomposition step into sub-bands (E401b) by time frequency transform or filter bank is applied to the decoded low band signal to obtain the spectrum of the decoded low band signal *U(k)* for use in the frequency domain.

A step E401a to extend the decoded low band signal into a second frequency band higher than the first frequency band, to obtain an extended decoded low band signal $U_{HB1}{}^{(k)}$ can be performed on this decoded low band signal before or after the analysis step (decomposition into sub-bands). This extension step can comprise both a re-sampling step and an extension step, or simply a frequency translation or transposition step as a function of the input signal. It is to be noted that in some variants, step E401a can be performed at the end of the processing described in Figure 4 i.e. on the combined signal, this processing then being chiefly performed on the low band signal before extension, the result being equivalent.

This step is detailed below in the embodiment described with reference to Figure 5.

A step E402 to extract an ambient signal $(U_{HBA}(k))$ and tonal components (y(k)) is performed on the decoded $(U(k))$ or decoded and extended $(U_{HB1}(k))$ low band signal. Ambience is defined here as the residual signal obtained by deleting the main (or dominant) harmonics (or tonal components) in the existing signal.

In most wideband signals (sampled at 16 kHz) the high band (>6 kHz) contains ambience information which is generally similar to that contained in the low band.

This step is obtained by:

- obtaining the ambient signal by calculating an average value of the decoded (or decoded and extended) low band signal; and
- obtaining tonal components by subtracting the calculated ambient signal from the decoded (or decoded and extended) low band signal.

The tonal components and ambient signal are then adaptively combined using energy-level control factors at step E403 to obtain a so-called combined signal $(U_{HB2}(k))$. The extension step E401a can then be carried out if not already performed on the decoded low band signal.

Therefore, the combining of these two types of signals allows a combined signal to be obtained having characteristics better adapted to certain types of signals such as music signals, and with richer frequency content in the extended frequency band corresponding to the entire frequency band including the first and second frequency bands.

The band extension of the method improves quality for these types of signals compared with the extension described in standard AMR-WB.

By using a combination of ambient signal and tonal components, it is possible to enrich this extension signal to bring it closer to the characteristics of the true signal and not an artificial signal.

This combination step will be detailed below with reference to Figure 5.

A synthesis step, which corresponds to the analysis at 401b, is performed at E404b to bring the signal into the time domain.

Optionally, a step to adjust the energy level of the high band signal can be performed at E404a, before and/or after the synthesis step, by applying a gain and/or by adequate filtering. This step is explained in more detail in the embodiment described in Figure 5 for blocks 501 to 507.

In one example of embodiment, the band extension device 500 is now described with reference to **Figure 5** illustrating both this device and processing modules adapted for application in a decoder of interoperable type with AMR-WB coding. This device 500 implements the band extension method previously described with reference to Figure 4.

The processing block 510 therefore receives a decoded low band signal *(u(n))*. In one particular embodiment, band extension uses excitation decoded at 12.8 kHz (exc2 or *u(n))* at the output of block 302 in Figure 3.

This signal is decomposed into frequency sub-bands by the sub-band decomposition module 510 (which performs step E401b in Figure 4) generally using a transform or applying a filter bank to obtain decomposition into sub-bands *U(k)* of the signal *u(n)*.

In one particular embodiment, a transform of DCT-IV type (Discrete Cosine Transform – Type IV) (block 510) is applied to the current frame of 20 ms (256 samples) without windowing, which amounts to directly transforming *u(n)* with n=0,…, 255 according to the following formula:

$$U(k)=\sum_{n=0}^{N-1}u(n)\cos\left(\frac{\pi}{N}\left(n+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\right)$$

where:

$N=256$ and $k=0,\cdots,255$.

Transforming without windowing (or in equivalent manner with an implicit rectangular window of frame length) is possible when processing is performed in the excitation domain and not in the signal domain. In this case, no artefact (block effects) is audible, which provides an important advantage of this embodiment of the invention.

In this embodiment, the DCT-IV transform is performed by FFT using the algorithm called *Evolved DCT(EDCT)* described in the article by D.N. Zhang, H.T. Li, A Low Complexity Transform – Evolved DCT, IEEE 14th International Conference on Computational Science and Engineering (CSE), Aug. 2011, pp. 144-149 and implemented in standards UIT-T G.718 Annex B and G.729.1 Annex E.

In some variants of the invention, and without loss of generality, the DCT-IV transform can be replaced by other short-term time-frequency transforms of same length and in the excitation domain or signal domain, such as the *Fast Fourier Transform* (FFT) or *Discrete Cosine Transform – Type II* (DCT-II). Alternatively, DCT-IV can be replaced on the frame by transforming with overlap-addition and windowing of longer length than the length of the current frame, for example using a *Modified Discrete Cosine Transform* (MDCT). In this case, the delay T in block 310 in Figure 3 must be adequately adjusted (reduced) as a function of the additional delay due to analysis/synthesis by this transform.

In another embodiment, decomposition into sub-bands is performed by application of a bank of real or complex filters e.g. of PQMF type (Pseudo-QMF). For some filter banks, for each sub-band in a given frame, the value obtained is not a spectral value but a series of time values associated with the sub-band; in this case, the preferred embodiment of the invention can be applied for example by performing a transform of each sub-band and calculating the ambient signal in the domain of absolute values, the tonal components always being obtained by the difference between the signal (in absolute value) and the ambient signal. For a complex filter bank, the complex module of the samples will replace the absolute value.

In other embodiments, the invention is applied in a system using two sub-bands, the low band being analysed by transform or by filter bank.

If a DCT is used, the DCT spectrum *U(k)* of 256 samples covering the band 0-6400 Hz (at 12.8 kHz) is then extended (block 511) into a spectrum of 320 samples covering the band 0-8000 Hz (at 16 Hz) in the following form:

$$U_{HB1}(k) = \begin{cases} 0 & k = 0, \cdots, 199 \\ U(k) & k = 200, \cdots, 239 \\ U(k + start\_band - 240) & k = 240, \cdots, 319 \end{cases}$$

where preferably *start_band* = 160.

Block 511 performs step E401a in Figure 4 i.e. extension of the decoded low band signal. This step can also comprise re-sampling from 12.8 to 16 kHz in the frequency domain, by adding ¼ of samples (*k*= 240,...,319) to the spectrum, the ratio between 16 and 12.8 being 5/4.

In the frequency band corresponding to the samples ranging from indices 200 to 239, the original spectrum is maintained for application thereto of a progressive attenuation response by the high-pass filter in this frequency band, and also to avoid introducing audible defects at the addition step of low frequency synthesis to high frequency synthesis.

It will be noted that in this embodiment, generation of the over-sampled extended spectrum takes place in a frequency band ranging from 5 to 8 kHz, therefore including a second frequency band (6.4-8 kHz) higher than the first frequency band (0-6.4 kHz).

As a result, extension of the decoded low band signal is performed at least on the second frequency band but also on part of the first frequency band.

Evidently, the values defining these frequency bands can differ depending on the decoder or processing device in which the invention is applied.

In addition, block 511 performs implicit high-pass filtering in the 0-5000 Hz band since the 200 first samples of $U_{HB1}(k)$ are set to zero; as explained below, this high-pass filtering can also be supplemented with progressive attenuation of the spectral values of indices $k = 200,\ldots,255$ in the 5000-6400 Hz band, this progressive attenuation being implemented in block 501 but could be performed separately outside block 501. In equivalent manner, and in some variants of the invention, separate implementation of high-pass filtering in blocks of coefficients of index $k = 0,\ldots,199$ set to zero, of attenuated coefficients $k = 200,\ldots,255$ in the transform domain can therefore be performed in a single step.

In this example of embodiment and according to the definition of $U_{HB1}(k)$, it is observed that the 5000-6000 Hz band of $U_{HB1}(k)$ (which corresponds to indices $k = 200,\ldots,239$) is copied from the 5000-6000 Hz band of $U(k)$. With this approach, it is possible to maintain the original spectrum in this band, and it avoids introducing distortions in the 5000-6000 Hz band when adding high-frequency synthesis to low-frequency synthesis – in particular the signal phase (implicitly represented in the DCT-IV domain) in this band is preserved.

The 6000-8000 Hz band of $U_{HB1}(k)$ is defined here by copying the 4000-6000 Hz band of $U(k)$ since the value of *start_band* is preferably set at 160.

In one variant of the embodiment, the value of *start_band* could be made adaptive around the value of 160, without modifying the inherent characteristics of the invention. The details for adapting the *start_band* value are not described herein since they lie outside but do not modify the scope of the invention.

In most wideband signals (sampled at 16 Hz), the high band (>6 kHz) contains ambience information which is naturally similar to that contained in the low band. Ambience is defined here as the residual signal obtained by deleting the main (or dominant) harmonics in the existing signal. The level of harmonicity in the 6000-8000 Hz band is generally correlated with that of the lower frequency bands.

This decoded and extended low band signal is input into the extension device 500 and in particular into the module 512. Therefore block 512, which extracts tonal components and an ambient signal, performs step E402 in Figure 4 in the frequency domain. The ambient signal $U_{HBA}(k)$ for k = 240,...,319 (80 samples) is therefore obtained for a second frequency band called high frequency, so that it can subsequently be adaptively combined with the extracted tonal components y(k) in the combination block 513.

In one particular embodiment, extraction of the tonal components and of the ambient signal (in the 6000-8000 Hz band) is performed with the following operations:

- Calculating the total energy of the extended, decoded low band signal $ener_{HB}$

  $$ener_{HB} = \sum_{k=240}^{319} U_{HB1}(k)^2 + \varepsilon$$

  where ε = 0.1 (this value can be different, it is fixed here as an example).

- Calculating ambience (in absolute value) which here corresponds to the average level of the spectrum *lev(i)* (line by line), and calculating the energy $ener_{tonal}$ of the dominant tonal parts (in the high frequency spectrum).

  For i = 0...L-1, this average level is obtained with the following equation:

  $$lev(i) = \frac{1}{fn(i) - fb(i) + 1} \sum_{j=fb(i)}^{fn(i)} \left| U_{HB1}(j + 240) \right|$$

This corresponds to the average level (in absolute value) and therefore represents a kind of spectral envelope. In this embodiment, L = 80 and represents the length of the spectrum, and the index i of 0 to L-1 corresponds to the indices j+240 from 240 to 319, i.e. the spectrum of 6 to 8 kHz.

In general $fb(i) = i - 7$ and $fn(i) = i + 7$, however the 7 first and last indices (i = 0,...,6 and $i = L-7, \dots L-1$) require special processing, and without loss of generality the following is therefore defined:

$$fb(i) = 0 \text{ and } fn(i) = i + 7 \text{ for } i = 0,\dots 6$$

$$fb(i) = i - 7 \text{ and } fn(i) = L-1 \text{ for } i = L-7,\dots L-1.$$

In some variants of the invention, the average of $|U_{HB1}(j + 240)|$, $j = fb(i), \ldots, fn(i)$ could be replaced by a median value on the same set of values, i.e. $lev(i) = median_{j=fb(i),\ldots,fn(i)}(|U_{HB1}(j + 240)|)$. This variant has the drawback of being more complex (in terms of number of computations) than a moving average. In other variants, non-uniform weighting could be applied to the averaged terms, or median filtering could be replaced for example by other nonlinear filters of *stack filter* type.

The residual signal is also calculated:

$$y(i) = |U_{HB1}(i+240)| - lev(i)\,,\ i = 0,\ldots,L-1$$

which (approximately) corresponds to the tonal components if the value $y(i)$ at a given line i is positive ($y(i)>0$). This calculation therefore involves implicit detection of the tonal components. The tonal parts are therefore implicitly detected by means of the intermediate term y(i) representing an adaptive threshold. The condition of detection being $y(i)>0$. In some variants of the invention, this condition could be changed, for example by defining an adaptive threshold that is a function of the local envelope of the signal or in the form $y(i) > lev(i) + x\ dB$ where x has a predefined value (for example $x = 10$ dB).

The energy of the dominant tonal parts is defined by the following equation:

$$ener_{tonal} = \sum_{i=0..7 | y(i)>0} y(i)^2$$

Other methods for extracting the ambient signal can evidently be envisaged. For example, this ambient signal can be extracted from a low frequency signal or possibly another frequency band (or several frequency bands).

The detection of peaks or tonal components could be performed differently.

The extraction of this ambient signal could also be performed on decoded but non-extended excitation i.e. before the spectral extension or translation step i.e. for example on a portion of the low frequency signal rather than directly on the high frequency signal.

In other variants of the invention, the absolute value of the spectral values can be replaced for example by the square of spectral values, without changing the principle of the invention; in this case, a square root is required to return to the signal domain, which is more complex to carry out.

The combination module 513 performs a combination step by adaptive mixing of the ambient signal and tonal components. For this purpose, an ambience-level control factor Γ is defined by the following equation:

$$\Gamma = \beta \frac{ener_{HB} - ener_{tonal}}{ener_{HB} - \beta ener_{tonal}}$$

β being a factor of which a computing example is given below.

To obtain the extended signal, first the combined signal is obtained in absolute values for $i = 0 \ldots L-1$:

$$y'(i) = \begin{cases} \Gamma y(i) + \frac{1}{\Gamma} lev(i) & y(i) > 0 \\ y(i) + \frac{1}{\Gamma} lev(i) & y(i) \leq 0 \end{cases}$$

to which the signs $U_{HB1}(k)$ are applied:

$$y''(i) = \mathrm{sgn}\left(U_{HB1}(i+240)\right).y'(i)$$

where the function sgn(.) gives the sign:

$$\mathrm{sgn}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

By definition the factor Γ is >1. The tonal components, detected line by line by the condition $y(i) > 0$ are reduced by factor Γ; the average level is amplified by the factor 1/ Γ.

In the adaptive mixing block 513, an energy-level control factor is computed as a function of the total energy of the decoded (or decoded and extended) low band signal and of the tonal components.

In one preferred embodiment of adaptive mixing, energy adjustment is performed in the following manner:

$$U_{HB2}(k) = fac.y''(k-240)\,,\ k = 240,\cdots,319$$

$U_{HB2}(k)$ being the combined band extension signal.

The adjustment factor is defined by the following equation:

$$fac = \gamma\sqrt{\frac{ener_{HB}}{\sum_{i=0}^{L-1} y''(i)}}$$

where γ allows avoiding of energy over-estimation. In one example of embodiment, β is computed so as to keep the same level of ambient signal in relation to the energy of the tonal components in the consecutive bands of the signal. The energy of the tonal components is computed in three bands: 2000-4000 Hz, 4000-6000 Hz and 6000-8000 Hz, with

$$E_{N2\sim4} = \sum_{k\in N(80,159)} U'^{2}(k)$$

$$E_{N4\sim6} = \sum_{k\in N(160,239)} U'^{2}(k)$$

$$E_{N4\sim6} = \sum_{k\in N(240,319)} U'^{2}(k)$$

where:

$$U'(k)=\begin{cases}\sqrt{\dfrac{\sum_{k=160}^{239}U^2(k)}{\sum_{k=80}^{159}U^2(k)}}\,U(k) & k=80,\ldots,159\\ U(k) & k=160,\ldots,239\\ \sqrt{\dfrac{\sum_{k=160}^{239}U^2(k)}{\sum_{k=240}^{319}U_{HB1}{}^2(k)}}\,U_{HB1}(k) & k=240,\ldots,319\end{cases}$$

And where N(k_1,k_2) is the set of indices k for which the coefficient of index k is classified as being associated with the tonal components. This set can be obtained for example by detecting the local peaks in $U'(k)$ verifying $|U'(k)|>lev(k)$, or $lev(k)$ is computed as the average level of the spectrum line by line.

It can be noted that other methods for computing the energy of the tonal components are possible, for example by taking the median value of the spectrum on the band under consideration.

β is fixed such that the ratio between the energy of the tonal components in the bands 4-6 kHz and 6-8 kHz is the same as between the bands 2-4 kHz and 4-6 kHz:

$$\beta=\frac{\rho-E_{N6-8}}{\sum_{k=160}^{239}U^2(k)-E_{N6-8}}$$

$$E_{N4-6}=\max(E_{N4-6},E_{N2-4}),\quad \rho=\frac{E_{N4-6}^2}{E_{N2-4}},\quad \rho=\max(\rho,E_{N6-8})$$

and max (..,..) is the function which gives the maximum of both arguments.

In some variants of the invention, the computing of β could be replaced by other methods. For example, in one variant, different features could be extracted (computed) characterizing the low band signal including a “tilt” feature similar to the one computed in the AMR-WB codec, and factor β is estimated as a function of a linear regression from these different features limiting its value to between 0 and 1. For example, linear regression could be estimated in supervised manner by estimating factor β using the original high band in a training dataset. It is noted that the computing mode of β does not limit the inherent characteristics of the invention. Next, parameter β can be used to compute γ taking into account the fact that a signal with an ambient signal added in a given band is generally perceived as being stronger than a harmonic signal having the same energy in the same band. If α is defined as the quantity of ambient signal added to the harmonic signal:

$$\alpha = \sqrt{1-\beta}$$

γ can be computed as a decreasing function of α, for example $\gamma = b - a\sqrt{\alpha}$, b=1.1, a = 1.2, and γ limited to 0.3 to 1. Here again, other definitions of α and γ are possible in the invention.

At the output of the band extension device 500, in one particular embodiment block 501 optionally performs a dual operation of applying bandpass filter frequency response, and deaccentuation (or deemphasis) filtering in the frequency domain.

In one variant of the invention, deemphasis filtering can be performed in the time domain, after block 502 even before block 510; however, in this case, bandpass filtering performed in block 501 can leave some low frequency components of very low level which become amplified by deemphasis, which may modify the decoded low band in slightly perceptible manner. On this account, it is preferred here to perform deemphasis in the frequency domain. In the preferred embodiment, the coefficients of index $k = 0,\ldots,199$ are set to zero, therefore deemphasis is limited to the higher coefficients.

Excitation is first deemphasised according to the following equation:

$$U_{HB2}'(k) = \begin{cases} 0 & k = 0, \cdots, 199 \\ G_{deemph}(k-200)U_{HB2}(k) & k = 200, \cdots, 255 \\ G_{deemph}(55)U_{HB2}(k) & k = 256, \cdots, 319 \end{cases}$$

where $G_{deemph}(k)$ is the frequency response of the $1/(1\text{-}0.68^{z-1})$ filter on a restricted discrete frequency band. Taking into account the (odd) discrete frequencies of DC-IV, $G_{deemph}(k)$ is defined here as:

$$G_{deemph}(k) = \frac{1}{\left|e^{j\theta_k} - 0.68\right|}, \ k = 0, \cdots, 255$$

where:

$$\theta_k = \frac{256 - 80 + k + \frac{1}{2}}{256}.$$

In the event that a transform other than DCT-IV is used, the definition of θ_k can be adjusted (for example for even frequencies).

It is noted that deemphasis is applied in two phases for k = 200,…,255 corresponding to the frequency band 5000-6400 Hz where the response $1/(0.68^{z-1})$ is applied as at 12.8 kHz, and for k = 256 …,319 corresponding to the frequency band 6400-8000 Hz where the response is extended here from 16 kHz to a constant value in the 6.4-8 kHz band.

It can be noted that in the AMR-WB codec, HF synthesis is not deemphasised. In the embodiment presented here, the high frequency signal on the contrary is deemphasised to bring it into a coherent domain with the low frequency signal (0-6.4 kHz) output from block 305 in Figure 3. This is of importance for estimation and subsequent adjustment of the energy of HF synthesis.

In one variant of the embodiment, to reduce complexity, $G_{deemph}(k)$ could be set at a constant value independent of k, taking for example $G_{deemph}(k)$ = 0.6, which approximately corresponds to the average value of $G_{deemph}(k)$ for k = 200,…,319 under the conditions of the above-described embodiment.

In another variant of the embodiment of the decoder, deemphasis could be performed in equivalent manner in the time domain after inverse DCT.

In addition to deemphasis, bandpass filtering is applied in two separate parts: one fixed high-pass, the other adaptive low-pass (as a function of bit rate).

This filtering is performed in the frequency domain.

In the preferred embodiment, the partial low-pass filter response is computed in the frequency domain as follows:

$$G_{lp}(k) = 1 - 0.999\frac{k}{N_{lp} - 1}$$

where $N_{lp} = 60$ at 6.6 kbits/s, 40 at 8.85 kbits/s, 20 at bit rates > 8.85 bits/s. Next, a bandpass filter is applied in the form:

$$U_{HB3}(k) = \begin{cases} 0 & k = 0,\cdots,199 \\ G_{lp}(k-200)U_{HB2}'(k) & k = 200,\cdots,255 \\ U_{HB2}'(k) & k = 256,\cdots,319 - N_{lp} \\ G_{lp}(k - 320 - N_{lp})U_{HB2}'(k) & k = 320 - N_{lp},\cdots,319 \end{cases}$$

The definition of $G_{hp}(k)$, $k = 0,\ldots,55$ is given for example in Table 1 below:

Table 1

K	$g_{hp}(k)$	K	$g_{hp}(k)$	K	$g_{hp}(k)$	K	$g_{hp}(k)$
0	0.001622428	14	0.114057967	28	0.403990611	42	0.776551214
1	0.004717458	15	0.128865425	29	0.430149896	43	0.800503267
2	0.008410494	16	0.144662643	30	0.456722014	44	0.823611104
3	0.012747280	17	0.161445005	31	0.483628433	45	0.845788355
4	0.017772424	18	0.179202219	32	0.510787115	46	0.866951597
5	0.023528982	19	0.197918220	33	0.538112915	47	0.887020781
6	0.030058032	20	0.217571104	34	0.565518011	48	0.905919644
7	0.037398264	21	0.238133114	35	0.592912340	49	0.923576092
8	0.045585564	22	0.259570657	36	0.620204057	50	0.939922577
9	0.054652620	23	0.281844373	37	0.647300005	51	0.954896429
10	0.064628539	24	0.304909235	38	0.674106188	52	0.968440179
11	0.075538482	25	0.328714699	39	0.700528260	53	0.980501849
12	0.087403328	26	0.353204886	40	0.726472003	54	0.991035206
13	0.100239356	27	0.378318805	41	0.751843820	55	1.000000000

It will be noted that in some variants of the invention, the values of $Ghp(k)$ could be modified whilst maintaining progressive attenuation. Similarly, low-pass filtering with variable bandwidth $Glp(k)$ could be adjusted with different values or frequency support without changing the principle of this filtering step.

It is to be noted also that low-pass filtering can be adapted by defining a single filtering step which combines high-pass and low-pass filtering.

In another embodiment, low-pass filtering can be performed in equivalent manner in the time domain (as in block 112 in Figure 1) with different filter coefficients according to bit rate, after an inverse DCT step. However, it will be noted that it is advantageous to perform this step directly in the frequency domain since filtering is performed in the LPC excitation domain, the problems of circular convolution and edge effects being very limited in this domain.

The inverse transform block 502 performs inverse DCT on 320 samples to find the high frequency signal sampled at 16 kHz. Implementation thereof is the same as in block 510 since the DCT-IV is orthonormal except that the length of the transform is 320 instead of 256, giving:

$$u_{HB}(n) = \sum_{k=0}^{N_{16k}-1} U_{HB3}(\mathrm{k}) \cos\left(\frac{\pi}{N_{16k}}\left(k+\frac{1}{2}\right)\left(n+\frac{1}{2}\right)\right)$$

where:

$$N_{16k} = 320 \text{ and } k = 0,\cdots,319 .$$

In the event that block 510 is not a DCT but another transform or decomposition into sub-bands, block 502 performs the synthesis corresponding to the analysis performed in block 510.

The signal sampled at 16 kHz is then optionally scaled by gains defined per sub-frame of 80 samples (block 504).

In one preferred embodiment, first (in block 503) a gain $g_{HB1}(m)$ is computed per sub-frame using energy ratios of the sub-frames, such that in each sub-frame of index $m = 0, 1, 2$ or 3 of the current frame:

$$g_{HB1}(m) = \sqrt{\frac{e_3(m)}{e_2(m)}}$$

where:

$$e_1(m) = \sum_{n=0}^{63} u(n+64m)^2 + \varepsilon$$

$$e_2(m) = \sum_{n=0}^{79} u_{HB}(n+80m)^2 + \varepsilon$$

$$e_3(m) = e_1(m)\frac{\sum_{n=0}^{319} u_{HB}(n)^2 + \varepsilon}{\sum_{n=0}^{255} u(n)^2 + \varepsilon}$$

with ε = 0.01. The gain per sub-frame $g_{HB1}(m)$ can be written in the form:

$$g_{HB1}(m) = \sqrt{\frac{\dfrac{\sum_{n=0}^{63} u(n+64m)^2 + \varepsilon}{\sum_{n=0}^{255} u(n)^2 + \varepsilon}}{\dfrac{\sum_{n=0}^{79} u_{HB}(n+80m)^2 + \varepsilon}{\sum_{n=0}^{319} u_{HB}(n)^2 + \varepsilon}}}$$

which shows that in the signal u_{HB} the same ratio is ensured between energy per sub-frame and energy per frame as in the signal $u(n)$.

Block 504 performs scaling of the combined signal (included in step E404a in Figure 4) as per the following equation:

$$u_{HB}'(n) = g_{HB1}(m)u_{HB}(n),\ n = 80m, \cdots, 80(m+1)-1$$

It will be noted that implementation by block 503 differs from that of block 101 in Figure 1, since energy at the current frame is taken into account in addition to that of the sub-frame. This makes it possible to have the energy ratio of each sub-frame in relation to frame energy. A comparison is therefore made between energy ratios (or relative energies) rather than absolute energies, between low band and high band.

This scaling step therefore allows maintaining in the high band of the energy ratio between the sub-frame and the frame in the same manner as in the low band.

Optionally, block 506 then scales the signal (included in step E404a in Figure 4) as per the following equation:

$$u_{HB}''(n) = g_{HB2}(m)u_{HB}'(n)\,,\ n = 80m,\cdots,80(m+1)-1$$

where the gain $g_{HB2}(m)$ is obtained from block 505 by executing blocks 103, 104 and 105 of the AMR-WB codec (the input to block 103 being low band decoded excitation $u(n)$). Blocks 505 and 506 are useful for adjusting the level of the LPC synthesis filter (block 507) here as a function of signal tilt. Other methods for computing gain $g_{HB2}(m)$ are possible without changing the inherent characteristics of the invention.

Finally, the signal $u_{HB}'(n)$ or $u_{HB}''(n)$ is filtered by the filtering module 507 which can be performed here taking $1/\hat{A}(z/\gamma)$ as transfer function, where $\gamma = 0.9$ at 6.6 kbits/s and $\gamma = 0.6$ at the other bit rates, which limits the order of the filter to the 16th order. In one variant, this filtering can be carried out in the same manner as described for block 111 in Figure 1 for the AMR-WB decoder, however the order of the filter increases to 20 at the bit rate of 6.6 which does not significantly change the quality of the synthesized signal. In another variant, LPC synthesis filtering can be performed in the frequency domain after calculating the frequency response of the filter used in block 507.

In some variants of embodiment of the invention, coding of the low band (0-6.4 kHz) could be replaced by a CELP coder other than the one used in AMR-WB e.g. the CELP coder in G.718 at 8 kbits/s. Without loss of generality, other wideband coders or operating at frequencies higher than 16 kHz could be used, in which coding of the low band takes place at an internal frequency of 12.8 kHz. Additionally, the invention can evidently be adapted to sampling frequencies other than 12.8 kHz when a low frequency coder operates at a sampling frequency lower than that of the original or reconstructed signal. If low band decoding does not use linear prediction, there is no excitation signal to be heard, in which case LPC analysis of the reconstructed signal can be performed in the current frame and LPC excitation is computed so that the invention can be applied.

Finally, in another variant of the invention, excitation or the low band signal $(u(n))$ is re-sampled, for example by linear interpolation or cubic spline from 12.8 to 16 kHz before transforming (e.g. DCT-IV) of length 320. This variant has the drawback of being more complex, since the (DCT-IV) transform of excitation or of the signal is then calculated over a longer length and re-sampling is not performed in the domain of the transform.

Additionally, in some variants of the invention, all the calculations needed for estimation of gains (G_{HBN}, $g_{HB1}(m)$, $g_{HB2}(m)$, $\mathrm{g}_{\mathrm{HBN}}$ ...) can be carried out in a logarithmic domain.

Figure 6 illustrates an example of physical implementation of a band extension device 600 according to the invention. It can be an integral part of an audio-frequency signal decoder or of equipment receiving audio-frequency signals whether or not decoded.

This type of device comprises a processor PROC cooperating with a memory block BM comprising a storage and/or working memory MEM.

Said device comprises an input module E able to receive a decoded or extracted audio signal in a first frequency band called low band brought into the frequency domain $(U(k))$. It comprises an output module S able to transmit the extension signal in a second frequency band $(U_{HB2}(k))$, for example to a filtering module 501 of Figure 5.

The memory block can advantageously comprise a computer programme comprising code instructions for implementing the steps of the band extension method in the meaning of the invention, when these instructions are executed by the processor PROC, and in particular the steps of extracting (E402) tonal components and an ambient signal from a signal derived from the decoded low band signal $(U)k))$, combining (E 403) the tonal components (y(k)) and the ambient signal

($U_{HBA}(k)$) by adaptive mixing using energy-level control factors to obtain an audio signal called combined signal ($U_{HB2}(k)$), extending (E401a) - on at least one second frequency band higher than the first frequency band - the decoded low band signal before the extraction step, or the combined signal after the combining step.

Typically, the description in Figure 4 reproduces the steps of an algorithm of said computer programme. The computer programme can also be stored on a memory medium readable by a reader of the device, or downloadable into the memory space thereof.

The memory MEM generally records all the data required for implementing the method.

In one possible embodiment, the device thus described may also comprise the functions of low band decoding and other processing functions described for example in Figures 5 and 3, in addition to the band extension functions of the invention.

Patentkrav

1. Fremgangsmåde til udvidelse af frekvensbåndet af et audiofrekvenssignal under en dekodnings- eller forbedringsproces med et trin til at opnå det dekodede signal i et første frekvensbånd betegnet det lave bånd, idet fremgangsmåden omfatter følgende trin, der udføres i frekvensområdet:

- udtrækning (E402) af tonekomponenter og af et omgivelsessignal ud fra et signal, der stammer fra det dekodede signal fra det lave bånd;
- kombination (E403) af tonekomponenterne og omgivelsessignalet ved adaptiv blanding under anvendelse af styrefaktorer på energiniveau for at opnå et kombineret signal;
- udvidelse (E401a) over mindst et andet frekvensbånd, der er højere end det første frekvensbånd af det dekodede signal fra det lave bånd inden udtrækningstrinnet eller for det kombinerede signal efter kombinationstrinnet;
- syntese (E404b) af et audiosignal for til tidsområdet at tilbageføre:
  - det kombinerede signal, hvis udvidelsestrinnet (E401a) udføres inden udtrækningstrinnet (E402), eller
  - det udvidede kombinerede signal, hvis udvidelsestrinnet (E401a) udføres efter kombinationstrinnet (E403);

idet fremgangsmåden er **kendetegnet ved, at** trinnet til udtrækning af tonekomponenterne og af omgivelsessignalet udføres ifølge følgende trin:

- opnåelse af omgivelsessignalet ved beregning af en gennemsnitsværdi af spektret for det dekodede eller det dekodede og udvidede signal fra det lave bånd;
- opnåelse af af tonekomponenterne ved subtraktion af det beregnede omgivelsessignal fra det dekodede eller det dekodede og udvidede signal fra det lave bånd.

2. Fremgangsmåde ifølge krav 1, **kendetegnet ved, at** det dekodede signal fra det lave bånd er et dekodet excitationssignal fra det lave bånd.

3. Fremgangsmåde ifølge krav 1, **kendetegnet ved, at** en til den adaptive blanding anvendt energiniveaureguleringsfaktor beregnes som funktion af den totale energi af det dekodede eller det dekodede og udvidede signal fra det lave bånd og af tonekomponenterne.

4. Fremgangsmåde ifølge et af de foregående krav, **kendetegnet ved, at** det dekodede signal fra det lave bånd undergår et trin til spaltning i underbånd ved transformation eller ved filterbank, idet udtræknings- og kombinationstrinnene så udføres i frekvensområdet eller i underbånd.

5. Fremgangsmåde ifølge et af de foregående krav, **kendetegnet ved, at** udvidelsestrinnet for det dekodede signal fra det lave bånd udføres efter følgende ligning:

$$U_{HB1}(k)=\begin{cases}0 & k=0,\cdots,199\\ U(k) & k=200,\cdots,239\\ U(k+start\_band-240) & k=240,\cdots,319\end{cases}$$

hvor k er prøvens index, U(k) er spektret for det efter et transformationstrin opnåede dekodede signal fra det lave bånd, $U_{HB1}(k)$ er spektret for det udvidede signal, og *start_band* er en foruddefineret variabel.

6. Apparat til udvidelse af frekvensbåndet af et audiofrekvenssignal, idet signalet er blevet dekodet i et første frekvensbånd betegnet det lave bånd, idet apparatet omfatter:

- et modul (512) til udtrækning af tonekomponenter og af et omgivelsessignal i frekvensområdet ud fra et signal, der stammer fra det dekodede signal fra det lave bånd;
- et modul (513) til kombination af tonekomponenterne og omgivelsessignalet i frekvensområdet ved adaptiv blanding under anvendelse af styrefaktorer på energiniveau for at opnå et kombineret signal;
- et udvidelsesmodul (511) i frekvensområdet på mindst et andet frekvensbånd, der er højere end det første frekvensbånd, anvendt på det dekodede signal fra det lave bånd før udtrækningsmodulet eller på det kombinerede signal efter dekombinationsmodulet,
- et modul (502) til syntese af et audiosignal for i tidsområdet at tilbageføre:
    - det kombinerede signal, hvis udvidelsesmodulet (511) anvendes inden udtrækningsmodulet (512), eller

- det udvidede kombinerede signal, hvis udvidelsesmodulet (511) anvendes efter kombinationsmodulet (513);

idet apparatet er **kendetegnet ved, at** modulet til udtrækning af tonekomponenterne og omgivelsessignalet er konfigureret til at

- opnå omgivelsessignalet ved beregning af en gennemsnitsværdi af spektret for det dekodede eller det dekodede og udvidede signal fra det lave bånd;
- opnå tonekomponenter ved subtraktion af det beregnede omgivelsessignal fra det dekodede eller det dekodede og udvidede signal fra det lave bånd.

7. Audiofrekenssignaldekoder, **kendetegnet ved, at** den omfatter et apparat til udvidelse af frekvensbåndet ifølge krav 6.

8. Computerprogram med kodeinstruktioner til udførelse af trinnene i fremgangsmåden til udvidelse af frekvensbåndet ifølge et af kravene 1 til 5, når disse instruktioner udføres af en processor.

9. Lagermedium, der kan læses af et apparat til udvidelse af frekvensbåndet, og på hvilket et computerprogram er optegnet, som omfatter kodeinstruktioner til udførelse af trinnene i fremgangsmåden til udvidelse af frekvensbåndet ifølge et af kravene 1 til 5.

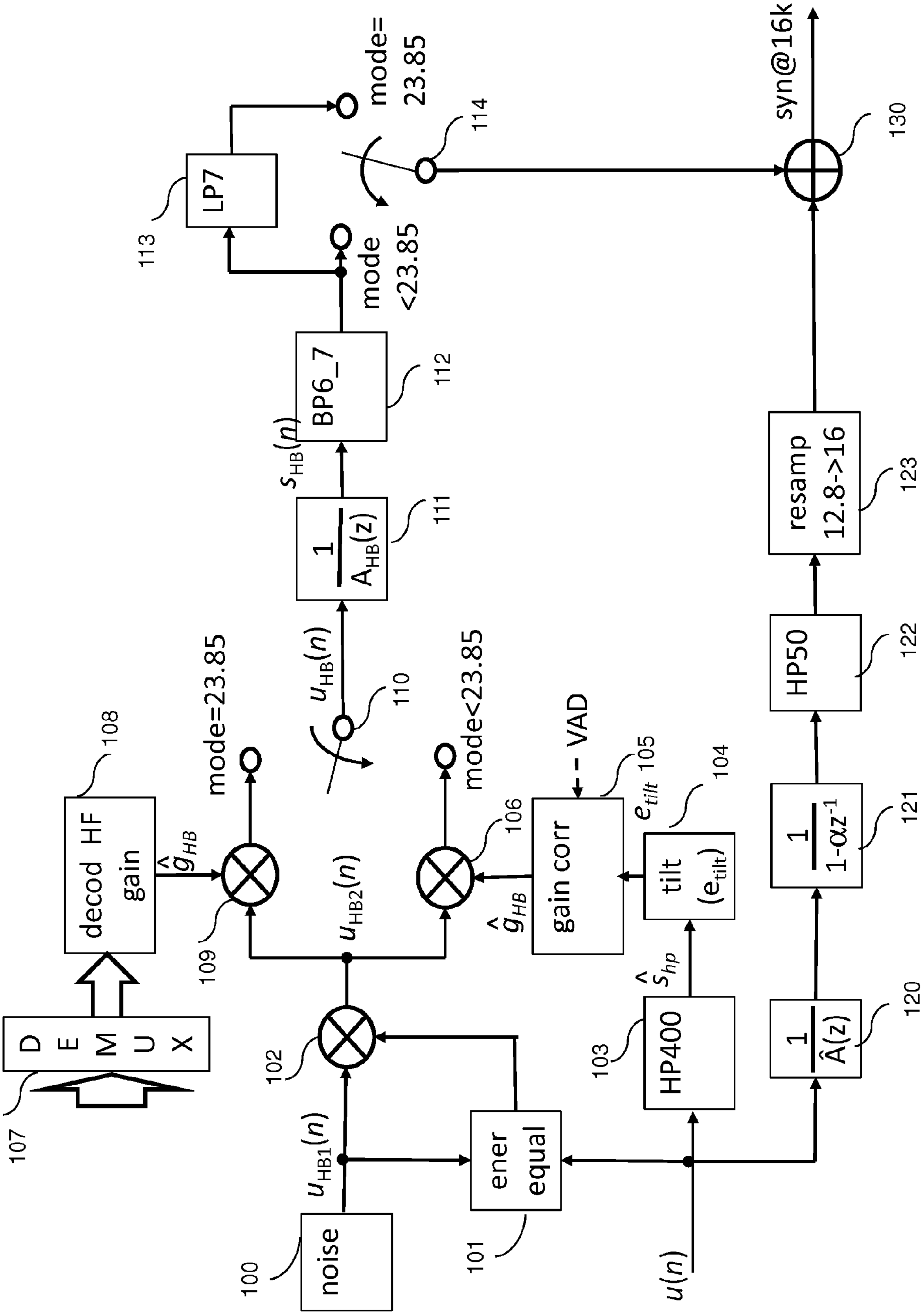
noise
100
$u_{HB1}(n)$
ener equal
101
$u(n)$
102
$u_{HB2}(n)$
D E M U X
107
decod HF gain
108
$\hat{g}_{HB}$
109
mode=23.85
106
mode<23.85
gain corr
105
VAD
$\hat{g}_{HB}$
tilt (e_{tilt})
104
e_{tilt}
HP400
103
$\hat{s}_{hp}$
110
$u_{HB}(n)$
$\frac{1}{A_{HB}(z)}$
111
$s_{HB}(n)$
BP6_7
112
mode <23.85
LP7
113
mode= 23.85
114
$\frac{1}{\hat{A}(z)}$
120
$\frac{1}{1-\alpha z^{-1}}$
121
HP50
122
resamp 12.8->16
123
130
syn@16k

Fig.1

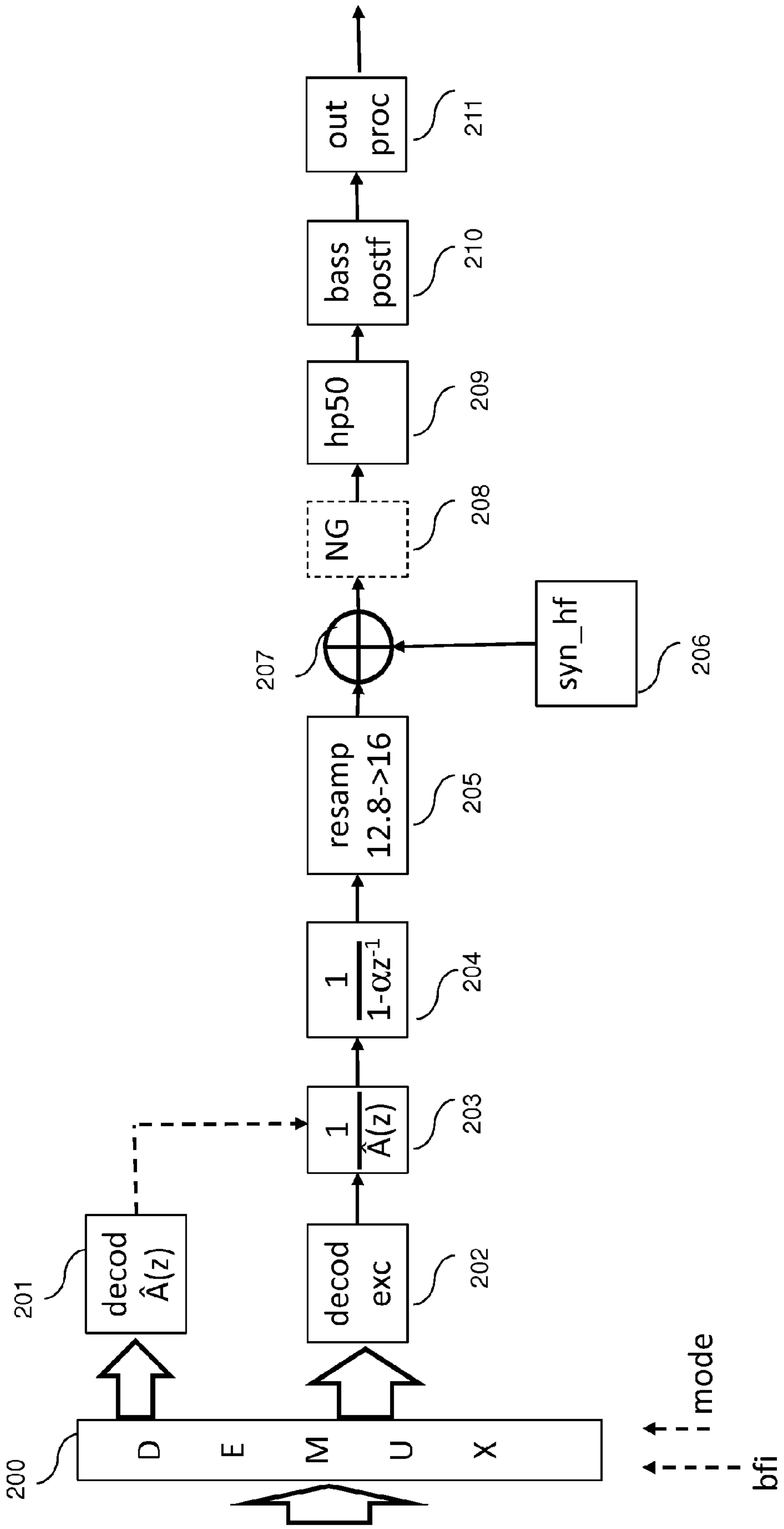
200
D E M U X
bfi
mode
201
decod
Â(z)
202
decod
exc
203
1/Â(z)
204
1/(1-αz⁻¹)
205
resamp
12.8->16
206
syn_hf
207
208
NG
209
hp50
210
bass
postf
211
out
proc

Fig.2

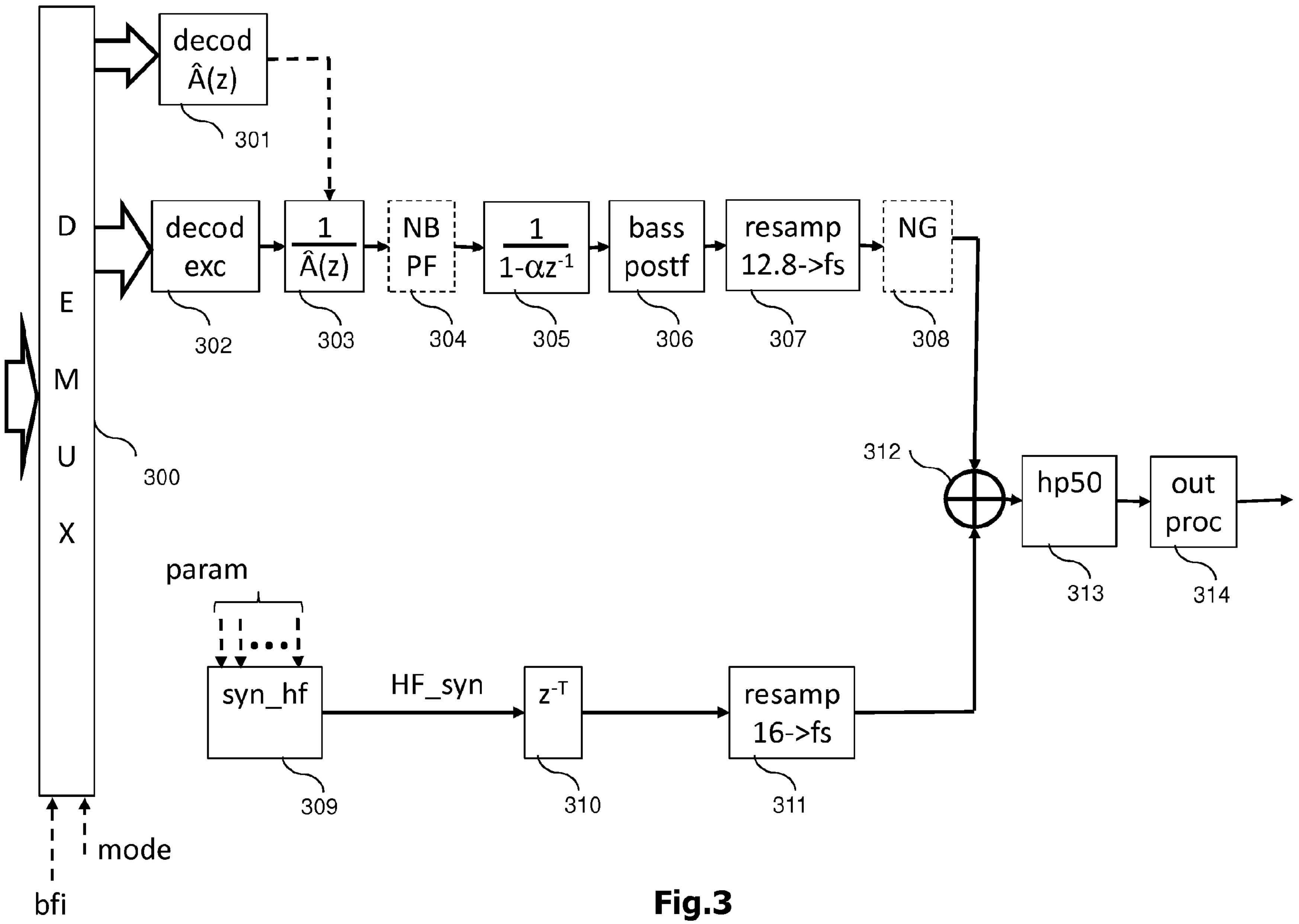
decod
Â(z)
301
D
E
M
U
X
300
decod
exc
302
1
Â(z)
303
NB
PF
304
1
1-αz-1
305
bass
postf
306
resamp
12.8->fs
307
NG
308
312
hp50
313
out
proc
314
param
syn_hf
309
HF_syn
z-T
310
resamp
16->fs
311
mode
bfi

Fig.3

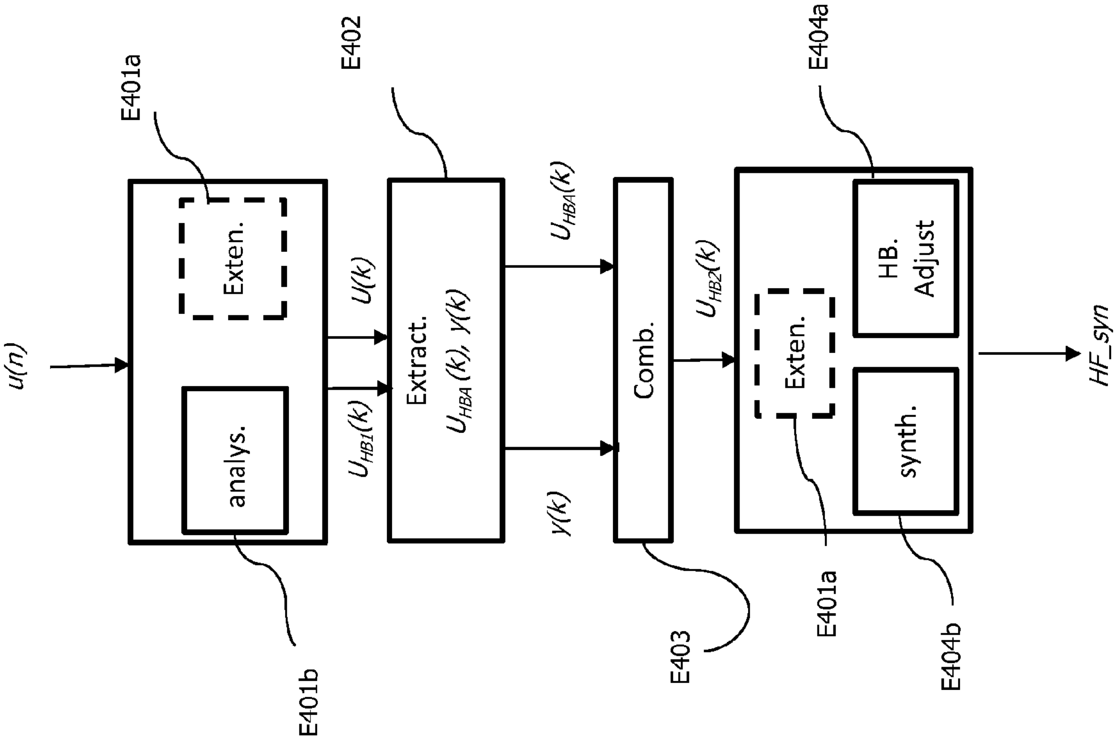
u(n)
E401a
Exten.
analys.
E401b
U(k)
$U_{HB1}(k)$
E402
Extract.
$U_{HBA}(k)$, y(k)
$U_{HBA}(k)$
y(k)
Comb.
E403
$U_{HB2}(k)$
E404a
Exten.
E401a
HB. Adjust
synth.
E404b
HF_syn

Fig.4

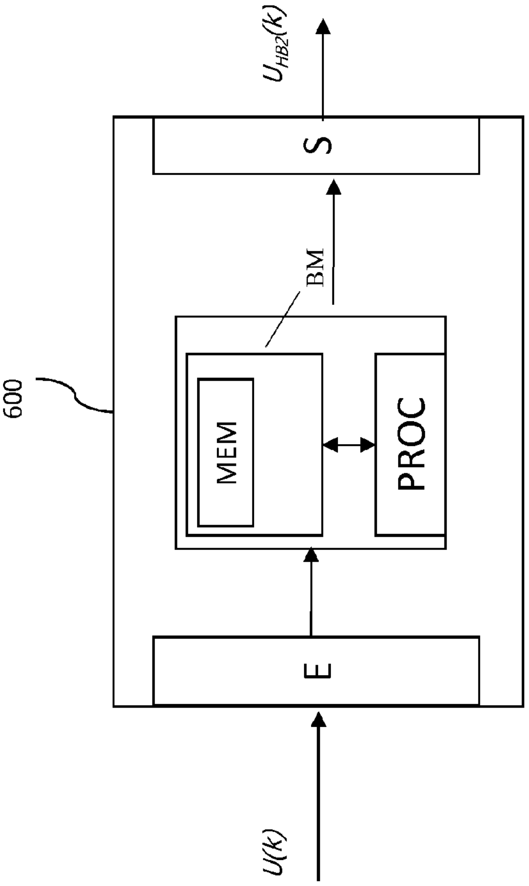
600
U(k)
E
MEM
PROC
BM
S
$U_{HB2}(k)$

Fig.6

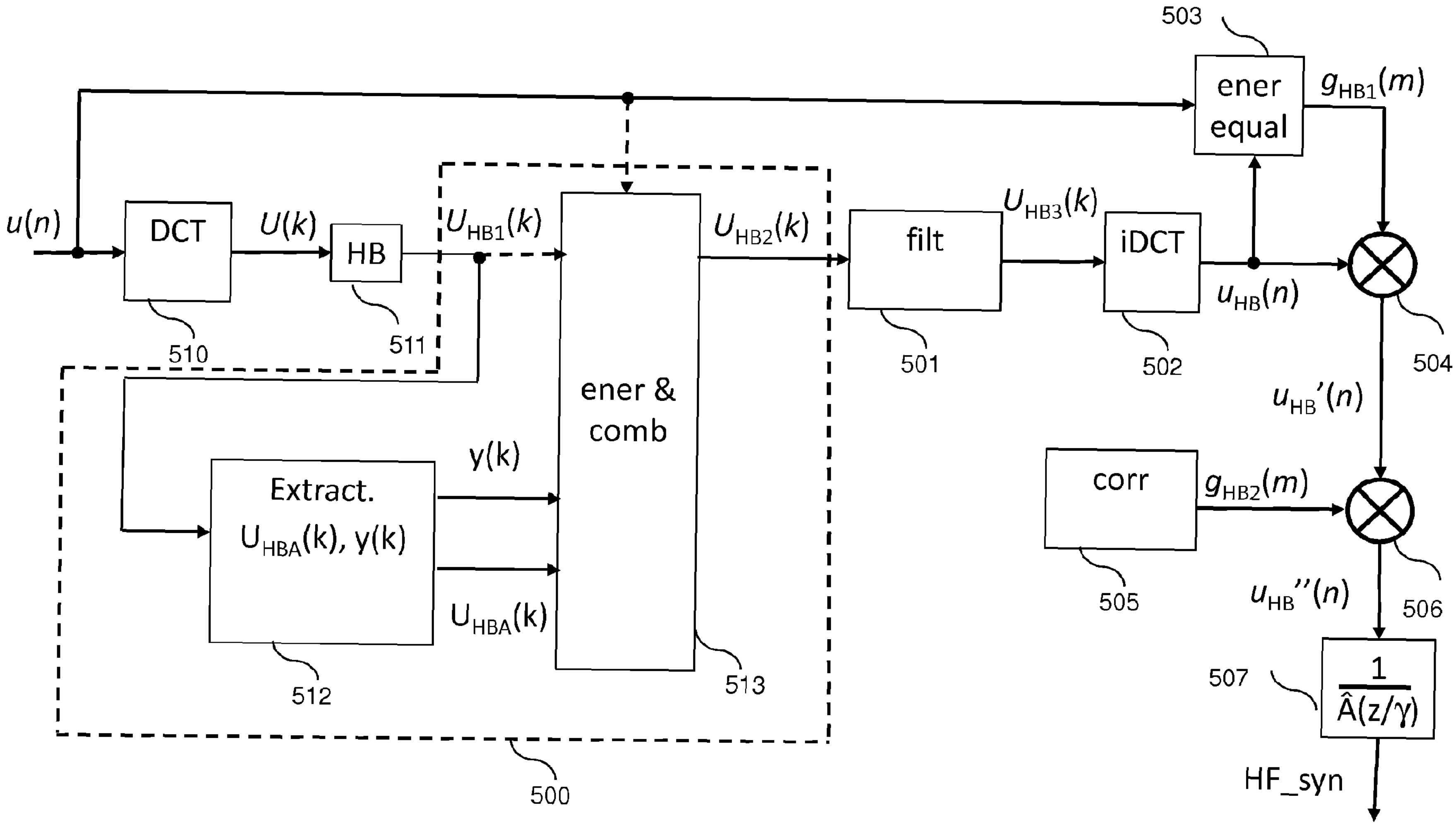
u(n)
DCT
510
U(k)
HB
511
$U_{HB1}(k)$
ener & comb
513
$U_{HB2}(k)$
Extract.
$U_{HBA}(k)$, y(k)
512
y(k)
$U_{HBA}(k)$
500
filt
501
$U_{HB3}(k)$
iDCT
502
$u_{HB}(n)$
ener equal
503
$g_{HB1}(m)$
504
$u_{HB}'(n)$
corr
505
$g_{HB2}(m)$
506
$u_{HB}''(n)$
$\frac{1}{\hat{A}(z/\gamma)}$
507
HF_syn

Fig.5