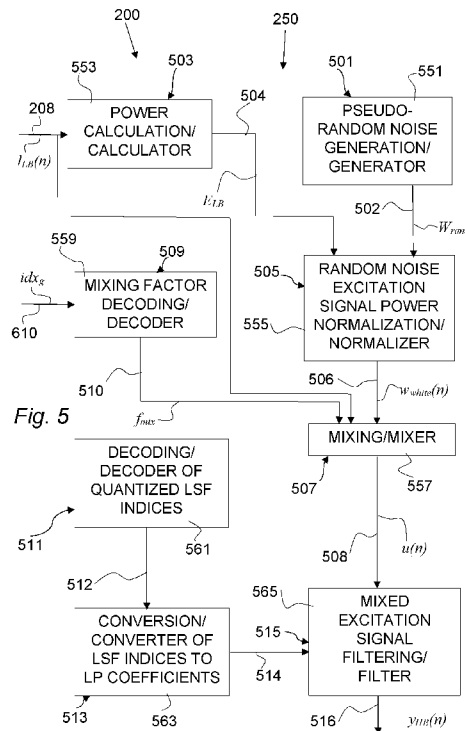




- (51) International Patent Classification: *G10L 19/00* (2013.01)
- (21) International Application Number: PCT/CA2023/050117
- (22) International Filing Date: 27 January 2023 (27.01.2023)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 63/306,291 03 February 2022 (03.02.2022) US
- (71) Applicant: **VOICEAGE CORPORATION** [CA/CA];
750 Lucerne Road, Suite 250, Town of Mount Royal,
Québec H3R 2H6 (CA).
- (72) Inventors: **MALENOVSKY, Vladimir**; 19, Kerova st.,
64100 BRNO (CZ). **JELINEK, Milan**; 1355 Emile-Nelligan,
Sherbrooke, Québec J1L 2W8 (CA).
- (74) Agent: **PRINCE, Gaetan**; 1100 Rene-Levesque boul.
West, 25th floor, Montreal, Québec H3B 5C9 (CA).
- (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, CV, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE,
KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU,
LY, MA, MD, 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, ST, SV, SY, TH,

(54) Title: TIME-DOMAIN SUPERWIDEBAND BANDWIDTH EXPANSION FOR CROSS-TALK SCENARIOS



(57) Abstract: A method for time-domain bandwidth expansion of an excitation signal during decoding of a cross-talk sound signal, comprises decoding a high-band mixing factor received in a bitstream, and mixing a low-band excitation signal and a random noise excitation signal using the high-band mixing factor to produce the time-domain expanded excitation signal. A method for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprises calculating a high-band residual signal using the sound signal and a temporal envelope of the high-band residual signal, calculating a high-band voicing factor based on the temporal envelope of the high-band residual signal, calculating a high-band mixing factor usable for mixing a low-band excitation signal and a random noise excitation signal to produce the time-domain expanded excitation signal, and estimating gain/shape parameters using the high-band voicing factor.



TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,
ZA, ZM, ZW.

- (84) Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, 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, ME, 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))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

TIME-DOMAIN SUPERWIDEBAND BANDWIDTH EXPANSION FOR CROSS-TALK SCENARIOS

TECHNICAL FIELD

[0001] The present disclosure relates to a method and device for time-domain bandwidth expansion of an excitation signal during encoding/decoding of a cross-talk sound signal.

[0002] In the present disclosure and the appended claims:

- The term “cross-talk” is generally intended to designate sound segments in which a first sound element is superposed to a second sound element, for example but not exclusively speech segments when a first person talks over a second person.
- The term “low-band” is intended to designate a lower frequency range. Although the 0 kHz - 6.4 kHz and 0 kHz - 8 kHz frequency ranges are given in the present disclosure as examples of “low-band”, the frequency boundaries of the low-band frequency range may obviously be modified/adapted to the bitrate of a codec and/or to achieve specific goals such as compliance with application-, system-, network- and design/business-related constraints.
- The term “high-band” is intended to designate a higher frequency range. Although the 6.4 kHz – 14 kHz and 8 kHz - 16 kHz frequency ranges are given in the present disclosure as examples of “high-band”, the frequency boundaries of the high-band frequency range may obviously be modified/adapted to the bitrate of a codec and/or to achieve specific goals such as compliance with application-, system-, network- and design/business-related constraints.

BACKGROUND

[0003] In many conversational applications there are often situations when one person talks over another person. As mentioned herein above, such situations are often referred to as “cross-talk”. Cross-talk speech segments may be problematic in modern speech encoding/decoding systems. Since the traditional speech encoding technologies have been designed and optimized mainly for single-talk content (only one person talking), the quality of cross-talk speech may be severely impacted by the encoding/decoding operations. As an example, one of the most serious issues in cross-talk speech encoding/decoding in the 3GPP EVS codec (Reference [1] or which the full content is incorporated herein by reference) is the occasional presence of “rattling noise”. “Rattling noise” is a strong annoying sound produced at frequencies from 8 kHz to 14 kHz, that is within the high-band frequency range examples as defined herein above.

[0004] At low bitrates of the 3GPP EVS codec the high-band frequency content is encoded/decoded using the superwideband bandwidth extension (SWB TBE) tool as described in Reference [1]. Due to the limited number of bits available for the SWB TBE tool the high-band excitation signal within the high-band frequency range is not encoded directly. Instead, the low-band excitation signal within the low-band frequency range is calculated using an ACELP (Algebraic Code-Excited Lineal Prediction) encoder (Reference [2] of which the full content is incorporated herein by reference), then upsampled and extended up to 14 kHz or 16 kHz depending on the high-band frequency range and used as a replacement for the high-band excitation signal. If there is a mismatch between the low-band excitation signal and the high-band excitation signal the synthesized sound may sound differently compared to the original sound. When the low-band excitation signal is voiced but the high-band excitation signal is unvoiced the synthesized sound will be perceived as the above defined rattling noise. The problem of rattling noise in the cross-talk content is illustrated in the spectral plot of Figure 1.

[0005] The plot in Figure 1 shows the power spectrum P versus frequency f of an exemplary cross-talk sound in which two speakers pronounce sounds of different types. While the sound from the first speaker (speaker 1) comprises dominantly voiced

content the sound from the second speaker (speaker 2) contains an unvoiced segment. Assuming a mono capturing device such as a smartphone or an omnidirectional microphone the sounds from the two speakers will be mixed together inside the capturing device. As a result the spectral content of the input sound signal, as seen by the encoder, will resemble the superset of the two spectra. A similar situation arises in a multi-channel capturing device such as a stereo microphone or an ambisonic microphone. If the encoder contains a downmixing module the resulting mono input signal might contain different types of sounds clearly distinguishable in the spectral domain.

SUMMARY

[0006] The present disclosure relates to the following aspects:

[0007] - A method for time-domain bandwidth expansion of an excitation signal during decoding of a cross-talk sound signal, comprising: decoding a high-band mixing factor received in a bitstream; and mixing a low-band excitation signal and a random noise excitation signal using the high-band mixing factor to produce the time-domain bandwidth expanded excitation signal.

[0008] - A method for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising: calculating (a) a high-band residual signal using the sound signal and (b) a temporal envelope of the high-band residual signal; and calculating a high-band voicing factor based on the temporal envelope of the high-band residual signal.

[0009] - A method for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising: calculating a high-band mixing factor usable for mixing a low-band excitation signal and a random noise excitation signal to produce the time-domain bandwidth expanded excitation signal.

[0010] - A method for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising: calculating (a) a high-band

residual signal using the sound signal and (b) a temporal envelope of the high-band residual signal; calculating a high-band voicing factor based on the temporal envelope of the high-band residual signal; calculating a high-band mixing factor usable for mixing a low-band excitation signal and a random noise excitation signal to produce the time-domain bandwidth expanded excitation signal; and estimating gain/shape parameters using the high-band voicing factor.

[0011] - A device for time-domain bandwidth expansion of an excitation signal during decoding of a cross-talk sound signal, comprising: a decoder of a high-band mixing factor received in a bitstream; and a mixer of a low-band excitation signal and a random noise excitation signal using the high-band mixing factor to produce the time-domain bandwidth expanded excitation signal.

[0012] - A device for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising: a calculator of (a) a high-band residual signal using the sound signal and (b) a temporal envelope of the high-band residual signal; and a calculator of a high-band voicing factor based on the temporal envelope of the high-band residual signal.

[0013] - A device for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising: a calculator of a high-band mixing factor usable for mixing a low-band excitation signal and a random noise excitation signal to produce the time-domain bandwidth expanded excitation signal.

[0014] - A device for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising: a calculator of (a) a high-band residual signal using the sound signal and (b) a temporal envelope of the high-band residual signal; a calculator of a high-band voicing factor based on the temporal envelope of the high-band residual signal; a calculator of a high-band mixing factor usable for mixing a low-band excitation signal and a random noise excitation signal to produce the time-domain bandwidth expanded excitation signal; and an estimator of gain/shape parameters using the high-band voicing factor.

[0015] The foregoing and other objects, advantages and features of the method and device for time-domain bandwidth expansion of an excitation signal during encoding/decoding of a cross-talk sound signal will become more apparent upon reading of the following non-restrictive description of illustrative embodiments thereof, given by way of example only with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] In the appended drawings:

[0017] Figure 1 is a plot showing the power spectrum P (dB) versus frequency f (kHz) of an exemplary cross-talk sound in which two speakers (speaker 1 and speaker 2) pronounce sounds of different types (VOICED and UNVOICED);

[0018] Figure 2 is a schematic block diagram illustrating concurrently a calculation/calculator of a high-band voicing factor in the method and the device for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal;

[0019] Figure 3 is a graph illustrating how a temporal envelope of a high-band residual signal is determined;

[0020] Figure 4 is a graph showing interpolation of segmental normalization factors calculated using mean values of consecutive segments of the down-sampled temporal envelope of the high-band residual signal;

[0021] Figure 5 is a schematic block diagram illustrating concurrently, at the decoder, a calculation/calculator of a time-domain bandwidth expanded excitation signal within the method and the device for time-domain bandwidth expansion of an excitation signal;

[0022] Figure 6 is a schematic block diagram illustrating concurrently, at the encoder, a calculation/calculator of a high-band mixing factor formed/represented by a

quantized normalized gain within the method and the device for time-domain bandwidth expansion of an excitation signal;

[0023] Figure 7 is a schematic block diagram of gain-shape estimation/estimator within the method and device for time-domain bandwidth expansion of an excitation signal;

[0024] Figure 8 is a graph illustrating interpolation of subframe gains; and

[0025] Figure 9 is a simplified block diagram of an example configuration of hardware components forming the method and device for time-domain bandwidth expansion of an excitation signal during encoding/decoding of a cross-talk sound signal.

DESCRIPTION

[0026] The following description relates to a technique for encoding/decoding cross-talk sound signals. In the present disclosure, the basis for the encoding/decoding technique is the SWB TBE tool of the 3GPP EVS codec as described in Reference [1]. However, it should be kept in mind that this technique may be used in conjunction with other encoding/decoding technologies.

[0027] More specifically, the present disclosure proposes a series of modifications to the SWB TBE tool. An objective of this series of modifications is to improve the quality of synthesized cross-talk sound signals, such as cross-talk speech signals, in particular but not exclusively to eliminate the above defined rattling noise. The series of modifications is concerned with time-domain bandwidth expansion of an excitation signal and is distributed in one or more of the following three areas:

- In the encoder, calculation of a high-band voicing factor using a temporal envelope of a high-band residual signal. In the SWB TBE tool, high-band corresponds to SHB (Super Higher-Band).
- In the encoder and decoder, calculation of a high-band mixing factor for a high-

band excitation signal.

- In the encoder and decoder, improvements in the estimation of gain/shape parameters and frame gain.

[0028] Calculation of the high-band voicing factor in accordance with the present disclosure uses a high-band autocorrelation function itself calculated from the temporal envelope of the high-band residual signal for example in the down-sampled domain. The high-band voicing factor is used in the encoder to replace the so-called voice factors derived from the low-band voicing parameter in the SWB TBE tool.

[0029] Calculation of the high-band mixing factor in accordance with the present disclosure replaces the corresponding method in the SWB TBE tool. The high-band mixing factor determines a proportion of a low-band excitation signal (for example from an ACELP core) and a random noise (which may also be defined as “white noise”) excitation signal for producing the time-domain bandwidth expanded excitation signal. In the disclosed implementation, the high-band mixing factor is calculated by means of MSE (Mean Squared Error) minimization between the temporal envelope of the random noise excitation signal and the temporal envelope of the low-band excitation signal, for example in the down-sampled domain. Quantization of the high-band mixing factor may be performed by the existing quantizer of the SWB TBE tool. The addition of the quantized high-band mixing factor to the SWB TBE bitstream results in a small increase of the bitrate. The mixing operation is performed both at the encoder and the decoder. Other properties of the mixing operation may comprise a re-scaling of the random noise excitation signal at the beginning of each frame and an interpolation of the high-band mixing factor to ensure smooth transitions between the current frame and the previous frame.

[0030] Estimation of the gain/shape parameters in accordance with the present disclosure comprises post-processing of the gain/shape parameters using adaptive smoothing of the unquantized gain/shape parameters (in the encoder) by means of weighting between original gain/shape parameters and interpolated gain/shape

parameters. Quantization of the gain/shape parameters may be performed by the existing quantizer of the SWB TBE tool. The adaptive smoothing is applied twice; it is first applied to the unquantized gain/shape parameters (in the encoder), and then to the quantized gain/shape parameters (both in the encoder and decoder). An adaptive attenuation is applied to the unquantized frame gain at the encoder. The adaptive attenuation is based on an MSE excess error which is a by-product of the SHB voicing parameter calculation in the SWB TBE tool.

[0031] Figure 2 is a schematic block diagram illustrating concurrently a calculation/calculator of a high-band voicing factor within the method 200 and the device 250 for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal.

1. Low-band Excitation Signal

[0032] Referring to Figure 2, the input sound signal $s_{inp}(n)$ to the 3GPP EVS codec is denoted, for example using the following relation (1):

$$s_{inp}(n), \quad n = 0, \dots, N_{32k} - 1 \quad (1)$$

[0033] where N_{32k} is the number of samples in the frame (frame length). In this particular non-limitative example, the input sound signal $s_{inp}(n)$ is sampled at the rate of $F_s = 32\text{kHz}$ and the length of a single frame is $N_{32k} = 640$ samples. This corresponds to a time interval of 20 ms. Frames of given duration, each including a given number of sub-frames and including a given number of successive sound signal samples, are used for processing sound signals in the field of sound signal encoding; further information about such frames can be found, for example, in Reference [1].

[0034] The method 200 comprises a downsampling operation 201 and the device 250 comprises a downsampler 251 for conducting operation 201. The downsampler 251 downsamples the input sound signal $s_{inp}(n)$ from 32 kHz to 12.8 kHz or 16kHz depending on the bitrate of the encoder. For example, the input sound signal

in the 3GPP EVS codec is downsampled to 12.8 kHz for all bitrates up to 24.4 kbps and to 16 kHz otherwise. The resulting signal is a low-band signal 202. The low-band signal 202 is encoded in an ACELP encoding operation 203 using an ACELP encoder 253.

[0035] The method 200 comprises the ACELP encoding operation 203 while the device 250 comprises the ACELP encoder 253 of the 3GPP EVS codec to perform the ACELP encoding. The ACELP encoder 253 generates two types of excitation signals, an adaptive codebook excitation signal 204 and a fixed codebook excitation signal 205 as described in Reference [1].

[0036] In the method 200 and device 250, the SWB TBE tool within the 3GPP EVS codec performs a low-band excitation signal generating operation 207 and comprises a corresponding generator 257 for generating the low-band excitation signal 208. The generator 257 uses the two excitation signals 204 and 205 as an input, mixes them together and applies a non-linear transformation to produce a mixed signal with flipped spectrum which is further processed in the SWB TBE tool to result into the low-band excitation signal 208 of Figure 2. Details about low-band excitation signal generation can be found in Reference [1]; specifically Section 5.2.6.1 describes SWB TBE encoding and Section 6.1.3.1 describes SWB TBE decoding.

[0037] As a non-limitative example, the low-band excitation signal 208 with flipped spectrum is sampled at 16 kHz and denoted using the following relation (2):

$$l_{LB}(n), \quad n = 0, \dots, N-1 \quad (2)$$

[0038] where $N = 320$ is the frame length.

2. High-Band Target Signal

[0039] Referring to Figure 2, a high-band target signal 210 is essentially an extract of the input sound signal $s_{inp}(n)$ containing spectral components in the frequency range of 6.4 kHz to 14 kHz or 8 kHz to 16 kHz depending on the bitrate of

the codec. The high-band target signal 210 is always sampled at 16 kHz regardless of the bitrate of the codec and its spectral content is flipped. Therefore, the first frequency bin of the high-band target spectrum corresponds to the last frequency bin of the spectrum and vice-versa. In the method 200 and device 250, the high-band target signal 210 may be generated for example using a QMF (Quadrature Mirror Filter) analysis operation 209 performed by the QMF analysis filter bank 259 of the 3GPP EVS codec as described in Reference [1]. Alternatively, the high-band target signal 210 may be generated by filtering the input sound signal $s_{inp}(n)$ with a pass-band filter, shifting it in frequency domain, flipping its spectral content as described above and finally downsampling it from 32 kHz to 16 kHz. In the present disclosure, the use of QMF processing will be assumed and the high-band target signal 210 is denoted, for example using the following relation (3):

$$s_{HB}(n), \quad n = 0, \dots, N - 1 \quad (3)$$

[0040] Following processing in the QMF filter bank 259, the method 200 comprises an operation 211 of estimating high-band filter coefficients 212 and the device 250 comprises an estimator 261 to perform operation 211. The estimator 261 estimates the high-band LP (Linear Prediction) filter coefficients 212 from the high-band target signal 210 in four consecutive subframes by frame where each subframe has the length of 80 samples. The estimator 261 calculates the high-band LP filter coefficients 212 using the Levinson-Durbin algorithm as described in Reference [1]. The high-band LP filter coefficients 212 may be denoted using the following relation (4):

$$\alpha_j^{HB}(k), \quad k = 0, \dots, P, j = 0, \dots, 3 \quad (4)$$

[0041] where $P = 10$ is the order of the high-band LP filter and $j = 0, \dots, 3$ is the subframe index. The first LP filter coefficient in each subframe is unitary, i.e. $\alpha_j^{HB}(0) = 1$.

[0042] The method 200 comprises an operation 213 of generating a high-band

residual signal 214 and the device 250 comprises a generator 263 of the high-band residual signal to conduct operation 213. The generator 263 produces the high-band residual signal 214 by filtering the high-band target signal 210 from the QMF analysis filter bank 259 with the high-band LP filter (LP filter coefficients 212) from estimator 261. The high-band residual signal 214 may be expressed, for example, using the following relation (5):

$$r_{HB}(n + j \cdot 80) = s_{HB}(n + j \cdot 80) - \sum_{k=1}^P \alpha_j^{HB}(k) s_{HB}(n + j \cdot 80 - k) \quad j = 0, \dots, 3, n = 0, \dots, 80 - 1 \quad (5)$$

[0043] The first P samples of the high-band residual signal 214 are calculated using the high-band target signal 210 from the previous frame. This is indicated by the negative index in $s_{HB}(-k), k = 1, \dots, P$ in the summation term. The negative indices refer to the samples of the high-band target signal 214 at the end of the previous frame.

3. High-Band Autocorrelation Function and Voicing Factor

[0044] Section 3 (High-Band Autocorrelation Function) relates to features of the encoder.

[0045] The high-band residual signal 214 calculated by the generator 263 using relation 5 is used to calculate a high-band autocorrelation function and a high-band voicing factor. The high-band autocorrelation function is not calculated directly on the high-band residual signal 214. Direct calculation of the high-band autocorrelation function requires significant computational resources. Furthermore, the dynamics of the high-band residual signal 214 are generally low and the spectral flipping process often leads to smearing the differences between voiced and unvoiced sound signals. To avoid these problems the high-band autocorrelation function is estimated on the temporal envelope of the high-band residual signal 214 for example in the downsampled domain.

[0046] The method 200 comprises an operation 215 of calculating the temporal envelope of the high band residual signal 214 and the device 250 comprises a calculator

265 to perform operation 215. To calculate the temporal envelope $R_{TD}(n)$ 216 of the high-band residual signal 214, the calculator 265 processes the high-band residual signal 214 through a sliding moving-average (MA) filter comprising in the example implementation $M = 20$ taps. The temporal envelope calculation can be expressed, for example by the following relation (6):

$$R_{TD}(n) = \frac{1}{M} \sum_{k=-M/2}^{M/2-1} |r_{HB}(n+k)|, \quad n = 0, \dots, N-M-1 \quad (6)$$

[0047] where the negative samples $r_{HB}(k)$, $k = -M/2, \dots, -1$ refer to the values of the high-band residual signal 214 in the previous frame. In mode switching scenarios it may happen that the high-band residual signal 214 in the previous frame is not calculated and the values are unknown. In that case the first $M/2$ values $r_{HB}(k)$, $k = 0, \dots, M/2-1$ are replicated and used as a replacement for the values $r_{HB}(k)$, $k = -M/2, \dots, -1$ of the previous frame. The calculator 265 approximates the last M values of the temporal envelope $R_{TD}(n)$ 216 in the current frame by means of IIR (Infinite Impulse Response) filtering. This can be done using the following relation (7):

$$R_{TD}(n) = 0.05 \cdot r_{HB}(n) + 0.95 \cdot R_{TD}(n-1), \quad n = N-M, \dots, N-1 \quad (7)$$

[0048] The operation 215 of calculating the temporal envelope $R_{TD}(n)$ 216 of the high-band residual signal 214 is illustrated in Figure 3.

[0049] The method 200 comprises a temporal envelope downsampling operation 217 and the device 250 comprises a downsampler 267 for conducting operation 217. The downsampler 267 downsamples the temporal envelope $R_{TD}(n)$ 216 by a factor of 4 using, for example, the following relation (8):

$$R_{4kHz}(n) = R_{TD}(4n), \quad n = 0, \dots, N/4-1 \quad (8)$$

[0050] The method 200 comprises a mean value calculating operation 219 and the device 250 comprises a calculator 269 for conducting operation 219. The calculator 269 divides the down-sampled temporal envelope $R_{4kHz}(n)$ 218 into four consecutive segments and calculates the mean value 220 of the down-sampled temporal envelope $R_{4kHz}(n)$ 218 in each segment using, for example, the following relation (9):

$$\bar{R}_k = \frac{1}{N/16} \sum_{n=0}^{N/16-1} R_{4kHz}(k \cdot \frac{N}{16} + n), \quad k = 0, \dots, 3 \quad (9)$$

[0051] where k is the index of the segment.

[0052] The calculator 269 limits all the mean values to a maximum value of 1.0.

[0053] The method 200 comprises a normalization factor calculating operation 221 and the device 250 comprises a calculator 271 for conducting operation 221. The calculator 271 uses the down-sampled temporal envelope mean values 220 to calculate, for the respective segments k , segmental normalization factors using, for example, the following relation (10):

$$\eta_k = \frac{1}{\bar{R}_k} \quad k = 0, \dots, 3 \quad (10)$$

[0054] The calculators 271 then linearly interpolates the segmental normalization factors from relation (10) within the entire interval of the current frame to produce interpolated normalization factors 222 using, for example, the following relation (11):

$$\begin{aligned} \gamma(n) &= \eta_{-1} + n \cdot \frac{\eta_0 - \eta_{-1}}{N/32}, & n = 0, \dots, N/32 - 1 \\ \gamma\left(\frac{N}{32} + k \cdot \frac{N}{16} + n\right) &= \eta_{k-1} + n \cdot \frac{\eta_k - \eta_{k-1}}{N/16}, & k = 0, \dots, 2, n = 0, \dots, N/16 - 1 \\ \gamma(n) &= \eta_3, & n = 7N/32, \dots, N/4 - 1 \end{aligned} \quad (11)$$

[0055] This interpolation process performed by operation 221 and calculator 271 is illustrated in Figure 4.

[0056] In relation (11), the term η_{-1} refers to the last segmental normalization factor in the previous frame. Therefore, η_{-1} is updated with η_3 after the interpolation process in each frame.

[0057] The method 200 comprises a downsampled temporal envelope normalizing operation 223 and the device 250 comprises a normalizer 273 for conducting operation 223. The normalizer 273 processes the down-sampled temporal envelope $R_{4kHz}(n)$ 218 from the downsampler 267 with the interpolated normalization factors $\gamma(n)$ 222 using, for example, the following relation (12):

$$R_\gamma(n) = R_{4kHz}(n) \cdot \gamma(n), \quad n = 0, \dots, N/4 - 1 \quad (12)$$

[0058] The normalizer 273 then subtracts the global mean value \bar{R}_γ (relation (13)) of the normalized envelope from the value $R_\gamma(n)$ of relation (12) to complete the downsampled temporal envelope normalization process ($R_{norm}(n)$ 224 of Figure 2) in operation 223. This can be expressed by relation (13):

$$\bar{R}_\gamma = \frac{1}{N/4} \sum_{n=0}^{N/4-1} R_\gamma(n) \quad (13)$$

$$R_{norm}(n) = R_\gamma(n) - \bar{R}_\gamma, \quad n = 0, \dots, N/4 - 1$$

[0059] It is useful to estimate the tilt of the temporal envelope of the high-band residual signal. For that purpose, the method 200 comprises a temporal envelope tilt estimation operation 225 and the device 250 comprises an estimator 275 for conducting operation 225. The temporal envelope tilt estimation can be done by fitting a linear curve to the segmental mean values \bar{R}_k calculated in relation (9) with the linear least squares (LLS) method. The tilt 226 of the temporal envelope is then the slope of the linear curve. The linear curve calculated with the LLS method is defined as:

$$y_k^{[LLS]} = a \cdot k + b, \quad k = 0, \dots, 3 \quad (14)$$

[0060] According to the LLS method, the objective is to minimize the sum of squared differences between \bar{R}_k and $y_k^{[LLS]}$ for all $k = 0, \dots, 3$. This can be expressed using the following relation (15):

$$(a_{LLS}, b_{LLS}) = \min_{a,b} \left[\sum_{k=0}^3 (\bar{R}_k - y_k^{[LLS]})^2 \right] \quad (15)$$

[0061] The optimal slope a_{LLS} (tilt 226) can be calculated by the estimator 275 using relation (16):

$$a_{LLS} = \frac{4 \sum_{k=0}^3 k \cdot \bar{R}_k - 6 \sum_{k=0}^3 \bar{R}_k}{4 \cdot 14 - 6 \cdot 6} = \frac{1}{5} \sum_{k=0}^3 k \cdot \bar{R}_k - \frac{3}{10} \sum_{k=0}^3 \bar{R}_k \quad (16)$$

[0062] The method 200 comprises a high-band autocorrelation function calculating operation 227 and the device 250 comprises a calculator 277 for conducting operation 227. The calculator 277 calculates the high-band autocorrelation function X_{corr} 228 based on the normalized temporal envelope using, for example, relation (17):

$$X_{corr}(k-40) = \frac{1}{\max(E_f, E_f^{[-1]})} \sum_{n=0}^{N/4-1} R_{norm}(n) R_{norm}(n-k), \quad k = 40, \dots, 80-1 \quad (17)$$

[0063] where E_f is the energy of the normalized temporal envelope $R_{norm}(n)$ 224 in the current frame and $E_f^{[-1]}$ is the energy of the normalized temporal envelope $R_{norm}(n)$ 224 in the previous frame. The calculator 277 may use the following relation (18) to calculate the energy:

$$E_f = \sum_{n=0}^{N/4-1} (R_{norm}(n))^2 + 1.0 \quad (18)$$

[0064] In case of mode switching the factor in front of the summation term in relation (17) is set to $1/E_f$ because the energy of the normalized temporal envelope $R_{norm}(n)$ 224 in the previous frame is unknown.

[0065] The method 200 comprises a high-band voicing factor calculating operation 229 and the device 250 comprises a calculator 279 for conducting operation 229.

[0066] The voicing of the high-band residual signal is closely related to the variance σ_{corr} of the high-band autocorrelation function X_{corr} 228. The calculator 279 calculates the variance σ_{corr} using, for example, the following relation (19):

$$\sigma_{corr} = \frac{1}{40} \sum_{k=0}^{40-1} (X_{corr}(k))^2 \quad (19)$$

[0067] To improve the discriminative potential (VOICED/UNVOICED decision) of the voicing parameter v_{mult} , the calculator 279 multiplies the variance σ_{corr} with the maximum value of the high-band autocorrelation function X_{corr} 228 as expressed in the following relation (20):

$$v_{mult} = \sigma_{corr} \cdot \max_k (X_{corr}(k)) \quad (20)$$

[0068] The calculator 279 then transforms the voicing parameter v_{mult} from relation (20) with the sigmoid function to limit its dynamic range and obtain a high-band voicing factor v_{HB} 230 using, for example, the following relation (21):

$$v_{HB} = \frac{2.0}{1.0 + \exp(-\beta \cdot v_{mult})} - 1.0 \quad (21)$$

[0069] where the factor β is estimated experimentally and set, for example, to a constant value of 25.0. The high-band voicing factor v_{HB} 230 as calculated from

relation (21) above is then limited to the range of $\langle 0.0;1.0 \rangle$ and transmitted to the decoder

4. Excitation Mixing Factor

[0070] Figure 5 is a schematic block diagram illustrating concurrently, at the decoder, a calculation/calculator of a time-domain bandwidth expanded excitation signal within the method 200 and the device 250.

[0071] Section 4 (Excitation Mixing Factor) relates to features of both the encoder and decoder.

[0072] The SWB TBE tool in the 3GPP EVS codec uses the low-band excitation signal 208 (Figure 2) described in Section 1 (Low-Band Excitation Signal) to predict the high-band residual signal 214 (Figure 2) described in Section 2 (High-Band Target Signal). At lower bitrates of the EVS codec, below 24.4 kbps, the SWB TBE tool uses 19 bits to encode the spectral envelope and the energy of the predicted high-band residual signal. With a frame length of 20 ms this results in a bitrate of 0.95 kbps. At bitrates higher than 24.4 kbps the SWB TBE tool uses 32 bits to encode the spectral envelope and the energy of the predicted high-band residual signal. With a frame length of 20 ms this results in a bitrate of 1.6 kbps. At both bitrates (0.95 and 1.6 kbps) of the SWB TBE tool no bits are used to encode the high-band residual signal 214 or the high-band target signal 210.

[0073] Referring to Figure 5, the method 200 comprises a pseudo-random noise generating operation 501 and the device 250 comprises a pseudo-random noise generator 551 to perform operation 501.

[0074] The pseudo-random noise generator 551 produces a random noise excitation signal 502 with uniform distribution. For example, the generator of pseudo-random numbers of the 3GPP EVS codec as described in Reference [1] can be used as pseudo-random noise generator 551. The random noise excitation signal w_{rand} 502

can be expressed using the following relation (22):

$$w_{rand}(n) \in U[-32767; 32768], \quad n = 0, \dots, N-1 \quad (22)$$

[0075] The random noise excitation signal w_{rand} 502 has zero mean and a non-zero variance $\sigma_{rand} = 1.14e+11$. It should be noted that the variance is only approximate and represents an average value over 100 frames.

[0076] The method 200 comprises an operation 503 of calculating the power of the low-band excitation signal $l_{LB}(n)$ 208 and a power calculator 553 to perform operation 503.

[0077] The power calculator 503 calculates that power 504 of the low-band excitation signal $l_{LB}(n)$ 208 transmitted from the encoder using, for example, the following relation (23):

$$E_{LB} = \sum_{n=0}^{N-1} (l_{LB}(n))^2 \quad (23)$$

[0078] The method 200 comprises an operation 505 of normalizing the power of the random noise excitation signal 502 and a power normalizer 555 to perform operation 505.

[0079] The power normalizer 555 normalizes the power of the random noise excitation signal 502 to the power 504 of the low-band excitation signal 208 using, for example, the following relation (24):

$$w_{white}(n) = \sqrt{\frac{E_{LB}}{\sigma_{rand}}} \cdot w_{rand}(n), \quad n = 0, \dots, N-1 \quad (24)$$

[0080] Although the true variance of the random noise excitation signal 502 varies from frame to frame, the exact value is not needed for power normalization. Instead, the above defined approximate value of the variance is used in the above

relation (24) to save computational resources.

[0081] The method 200 comprises an operation 507 of mixing the low-band excitation signal $I_{LB}(n)$ 208 with the power normalized random noise excitation signal $w_{white}(n)$ 506 and a mixer 557 to perform operation 507.

[0082] The mixer 557 produces the time-domain bandwidth expanded excitation signal 508 by mixing the low-band excitation signal $I_{LB}(n)$ 208 with the power normalized random noise excitation signal $w_{white}(n)$ 506 using a high-band mixing factor to be described later in the present disclosure.

[0083] Figure 6 is a schematic block diagram illustrating concurrently, at the encoder, a calculation/calculator of a high-band mixing factor formed/represented by a quantized normalized gain within the method and the device for time-domain bandwidth expansion of an excitation signal.

[0084] Referring to Figure 6, at the encoder,

- the method 200 comprises an operation 602 of calculating the temporal envelope of the power-normalized random noise excitation signal $w_{white}(n)$ 506, an operation 604 of calculating the temporal envelope of the low-band excitation signal $I_{LB}(n)$ 208, and a mean squared error (MSE) minimizing operation 601, and a gain quantizing operation 607; and
- the device 250 comprises a temporal envelope calculator 652 to perform operation 602, a temporal envelope calculator 654 to perform operation 604, an MSE minimizer 651 to perform operation 601, and a gain quantizer 657 to perform operation 607.

[0085] As illustrated in Figure 6, to save computational resources, optimal gains g_l^*, g_w^* are calculated based on the temporal envelopes of the signals in the downsampled domain using a mean squared error (MSE) minimization process.

Another advantage of this approach is higher robustness against background noise.

[0086] The calculator 652 calculates the downsampled temporal envelope $W_{4kHz}(n)$ 606 of the power-normalized random noise excitation signal $w_{white}(n)$ 506 (which is also calculated at the encoder as shown in Figure 5 and corresponding description) using the same algorithm as described in Section 3 (High-Band Autocorrelation Function and Voicing Factor) upon calculating (operation 215 and calculator 265 of Figure 2) the temporal envelope of the high-band residual signal 214 and downsampling (operation 217 and downsampler 267 of Figure 2) the temporal envelope. The downsampling factor being used is, for example, 4. The downsampled temporal envelope of the power-normalized random noise excitation signal can be denoted using the following relation (25):

$$W_{4kHz}(n), \quad n = 0, \dots, N/4 - 1 \quad (25)$$

[0087] Similarly, the calculator 654 calculates temporal envelope $L_{4kHz}(n)$ 605 of the low-band excitation signal $I_{LB}(n)$ 208 downsampled at 4 kHz again using the same algorithm as described in Section 3 (High-Band Autocorrelation Function and Voicing Factor). The downsampled temporal envelope 606 of the low-band excitation signal $I_{LB}(n)$ 208 can be denoted as follows:

$$L_{4kHz}(n), \quad n = 0, \dots, N/4 - 1 \quad (26)$$

[0088] The objective of the MSE minimization operation 601 is to find an optimal pair of gains g_l^*, g_w^* minimizing the energy of the error between (a) the combined temporal envelope ($L_{4kHz}(n), W_{4kHz}(n)$) and (b) the temporal envelope $R_{4kHz}(n)$ of the high-band residual signal $r_{HB}(n)$ 214. This can be mathematically expressed using relation (27):

$$(g_l^*, g_w^*) = \min_{g_l, g_w} \left[\sum_{n=0}^{N/4-1} (R_{4kHz} - g_l \cdot L_{4kHz}(n) - g_w \cdot W_{4kHz}(n))^2 \right] \quad (27)$$

[0089] For that purpose, the MSE minimizer 651 solves a system of linear equations. The solution is found in the scientific literature. For example, the optimal pair of gains g_l^* , g_w^* can be calculated using relation (28):

$$\begin{aligned}
 D &= 4.0 \cdot c_0 \cdot c_2 - c_4 \cdot c_4 \\
 g_l^* &= \frac{c_3 \cdot c_4 - 2.0 \cdot c_1 \cdot c_2}{D} \\
 g_w^* &= \frac{c_1 \cdot c_4 - 2.0 \cdot c_0 \cdot c_3}{D}
 \end{aligned} \tag{28}$$

[0090] where the values c_0, \dots, c_4 , and c_5 are given by

$$\begin{aligned}
 c_0 &= \sum_{n=0}^{N/4-1} (L_{4kHz}(n))^2, \quad c_1 = -2.0 \sum_{n=0}^{N/4-1} L_{4kHz}(n) \cdot R_{4kHz}(n), \quad c_2 = \sum_{n=0}^{N/4-1} (W_{4kHz}(n))^2 \\
 c_3 &= -2.0 \sum_{n=0}^{N/4-1} R_{4kHz}(n) \cdot W_{4kHz}(n), \quad c_4 = 2.0 \sum_{n=0}^{N/4-1} L_{4kHz}(n) \cdot W_{4kHz}(n), \quad c_5 = \sum_{n=0}^{N/4-1} (R_{4kHz}(n))^2
 \end{aligned} \tag{29}$$

[0091] The MSE minimizer 651 then calculates the minimum MSE error energy (excess error) using, for example, the following relation (30):

$$E_{err} = (g_l^*)^2 c_0 + g_l^* c_1 + (g_w^*)^2 c_2 + g_w^* c_3 + g_l^* g_w^* c_4 + c_5 \tag{30}$$

[0092] For further processing, the gain quantizer 657 scales the optimal gains g_l^* , g_w^* in such a way that a gain g_{ln} associated with the temporal envelope $L_{4kHz}(n)$ 605 of the low-band excitation signal $I_{LB}(n)$ becomes unitary, with a gain g_{wn} associated with the temporal envelope $W_{4kHz}(n)$ 606 of the power-normalized random noise excitation signal $w_{white}(n)$ 506 given using, for example the following relation (31):

$$\begin{aligned}
 g_{ln} &= 1.0 \\
 g_{wn} &= \frac{g_w^*}{g_l^* + g_w^*}
 \end{aligned} \tag{31}$$

[0093] The result/advantage of the re-scaling of relation (31) is that only one parameter, the normalized gain g_{wn} , needs to be coded and transmitted in the bitstream from the encoder to the decoder instead of two parameters. Therefore, scaling of the gains using relation (31) reduces bit consumption and simplifies the quantization process. On the other hand, the energy of the combined temporal envelopes ($L_{4kHz}(n)$ and $W_{4kHz}(n)$) will not match the energy of the temporal envelope $R_{4kHz}(n)$ of the high-band residual signal 214. This is not a problem since the SWB TBE tool uses subframe gains and a global gain containing the information about energy of the high-band residual signal. The calculation of subframe gains and the global gain is described in Section 6 (Gain/Shape Estimation) of the present disclosure.

[0094] The gain quantizer 657 limits the normalized gain g_{wn} between a maximum threshold of 1.0 and a minimum threshold of 0.0. The gain quantizer 657 quantizes the normalized gain g_{wn} using, for example, a 3-bit uniform scalar quantizer described by the following relation (32):

$$idx_g = \left\lfloor \frac{g_{wn}}{1/7} + 0.5 \right\rfloor \quad (32)$$

[0095] and the resulting index idx_g 610 is limited to the interval $\langle 0; 7 \rangle$ to form/represent the high-band mixing factor and is transmitted in the SWB TBE bitstream together with the existing indices of the SWB TBE encoder at 0.95 kbps or 1.6 kbps.

[0096] Referring back to Figure 5, the method 200 comprises, at the decoder, a mixing factor decoding operation 509, and the device 250 comprises a mixing factor decoder 559 to perform operation 509.

[0097] The mixing factor decoder 559 produces from the received index idx_g 610 a decoded gain using, for example, the following relation (33):

$$f_{mix} = \frac{idx_g}{7} \quad (33)$$

[0098] The decoded gain from relation (33) forms the high-band mixing factor f_{mix} 510.

[0099] The low-band excitation signal $l_{LB}(n)$ 208, sampled for example at 16 kHz, and the normalized random noise excitation signal $w_{white}(n)$ 506, sampled for example at 16 kHz, are mixed together in the mixer 557. However, both the energy of the low-band excitation signal $l_{LB}(n)$ 208 and the energy of the random noise excitation signal w_{rand} 502 vary from frame to frame. The fluctuation of energy could eventually generate audible artifacts at frame borders if the low-band excitation signal $l_{LB}(n)$ 208 and the random noise excitation signal w_{rand} 502 were mixed directly using the high-band mixing factor f_{mix} 510 obtained from relation (33). To ensure smooth transitions the energy of the random noise excitation signal w_{rand} 502 is linearly interpolated in generator 551 between the previous frame and the current frame. This can be done by scaling the random noise excitation signal w_{rand} 502 in the first half of the current frame with the following interpolation factor:

$$\zeta_w(n) = \sqrt{\frac{E_{LB}^{[-1]}}{E_{LB}}} + n \frac{1.0 - \sqrt{\frac{E_{LB}^{[-1]}}{E_{LB}}}}{N/2}, \quad n = 0, \dots, N/2 - 1 \quad (34)$$

[00100] where E_{LB} is the energy of the low-band excitation signal $l_{LB}(n)$ 208 in the current frame and $E_{LB}^{[-1]}$ is the energy of the low-band excitation signal $l_{LB}(n)$ 208 in the previous frame.

[00101] To further smooth the transitions between the previous and the current frame the decoder 559 also linearly interpolates the high-band mixing factor f_{mix} 510.

This can be done by introducing the scaling factor $\beta_{mix}(n)$ calculated, for example, using the following relation:

$$\beta_{mix}(n) = f_{mix}^{[-1]} + n \cdot \frac{f_{mix} - f_{mix}^{[-1]}}{N/2}, \quad n = 0, \dots, N/2 - 1 \quad (35)$$

[00102] where $f_{mix}^{[-1]}$ is the value of the high-band mixing factor in the previous frame. Note that the interpolation factor $\zeta_w(n)$ calculated in relation (34) and the scaling factor $\beta_{mix}(n)$ calculated in relation (35) are defined for $n = 0, \dots, N/2 - 1$.

[00103] The mixing of the low-band excitation signal $l_{LB}(n)$ 208 and the random noise excitation signal $w_{white}(n)$ 506 is finally done by the mixer 557 using, for example, relation (36) to obtain a time-domain bandwidth expanded excitation signal $u(n)$ 508.

$$\begin{aligned} u(n) &= \beta_{mix}(n) \cdot \zeta_w(n) \cdot l_{LB}(n) + (1 - \beta_{mix}(n)) \cdot w_{white}(n), & n = 0, \dots, N/2 - 1 \\ u(n) &= f_{mix} \cdot l_{LB}(n) + (1 - f_{mix}) \cdot w_{white}(n), & n = N/2, \dots, N - 1 \end{aligned} \quad (36)$$

5. High-band synthesis (LP synthesis)

[00104] The high-band LP filter coefficients $a_f^{HB}(n)$ 212 calculated by means of the LP analysis on the high-band input signal $s_{HB}(n)$ in relation (4) are converted in the encoder of the SWB TBE tool into LSF parameters and quantized. At the bitrate of 0.95 kbps the SWB TBE encoder uses 8 bits to quantize the LSF indices. At the bitrate of 1.6 kbps the SWB TBE encoder uses 21 bits to quantize the LSF indices.

[00105] Referring back to Figure 5, at the decoder:

- The method 200 comprises a decoding operation 511 and the device 250 comprises a corresponding decoder 561 to decode the quantized LSF indices; and

- The method 200 comprises a conversion operation 513 and the device 250 comprises a corresponding converter 563 to convert the decoded LSF indices 512 into high-band LP filter coefficients 514.

[00106] The decoded high-band LP filter coefficients 512 can be denoted as:

$$\alpha_{qj}^{HB}(k), \quad k=0,\dots,P, j=0,\dots,3 \quad (37)$$

[00107] where $P=10$ is the order of the LP filter. The first decoded LP filter coefficient in each subframe is unitary, i.e. $\alpha_{qj}^{HB}(0)=1.0, j=0,\dots,3$.

[00108] The method 200 comprises a filtering operation 515 and the device 250 comprises a corresponding synthesis filter 565 using the decoded high-band LP filter coefficients 514 to filter the mixed time-domain bandwidth expanded excitation signal 508 of relation (36) using for example the following relation (38) to obtain a LP-filtered high-band signal y_{HB} 516:

$$y_{HB}(n+j\cdot 80) = u(n+j\cdot 80) - \sum_{k=1}^P \alpha_{qj}^{HB}(k) y_{HB}(n+j\cdot 80-k) \quad j=0,\dots,3, n=0,\dots,80-1 \quad (38)$$

6. Gain/Shape Estimation (Figure 7)

[00109] A gain/shape parameter smoothing is applied both at the encoder and at the decoder. The adaptive attenuation of the frame gain is applied at the encoder only.

[00110] The spectral shape of the high-band target signal $s_{HB}(n)$ 210 is encoded with the quantized LSF coefficients. Referring to Figure 7, the SWB TBE tool also comprises an estimation operation 701/estimator 751 for estimating temporal subframe gains 702 of the high-band target signal $s_{HB}(n)$ 210 as described in Reference [1]. The estimator 751 normalizes the estimated temporal subframe gains to unit energy.

[00111] The normalized estimated temporal subframe gains 702 from estimator

751 can be denoted using relation (39):

$$g_k, \quad k = 0, \dots, 3 \quad (39)$$

[00112] The method 200 comprises a calculating operation 703 and the device 250 comprise a corresponding calculator 753 for determining a temporal tilt 704 of the normalized estimated temporal subframe gains g_k 702 by means of linear least squares (LLS) interpolation. As illustrated in Figure 8, this interpolation process can be done by fitting a linear curve 801 to the true subframe gains 702 in four consecutive subframes (subframes 0-3 in Figure 8) and calculating its slope.

[00113] The linear curve 801 built with the LLS interpolation method can be defined using the following relation (40):

$$g_k^{[LLS]} = c_{LLS} \cdot k + d_{LLS}, \quad k = 0, \dots, 3 \quad (40)$$

[00114] where the parameters c_{LLS} and d_{LLS} are found by minimizing the sum of squared differences between the true subframe gains g_k 702 and the corresponding points on the linear curve for all $k = 0, \dots, 3$ subframes. This can be expressed using the following relation (41):

$$(c_{LLS}, d_{LLS}) = \min_{c,d} \left[\sum_{k=0}^3 (g_k - g_k^{[LLS]})^2 \right] \quad (41)$$

[00115] By expanding relation (41) it is possible to express the temporal tilt g_{tilt} of the estimated temporal subframe gains g_k 702. The temporal tilt g_{tilt} 702 is, in fact, equal to the optimal slope c_{LLS} of the linear curve. The temporal tilt g_{tilt} can be calculated in the calculator 753 using the following relation (42):

$$g_{\text{tilt}} = c_{LLS} = \frac{4 \sum_{k=0}^3 k \cdot g_k - 6 \sum_{k=0}^3 g_k}{4 \cdot 14 - 6 \cdot 6} = \frac{1}{5} \sum_{k=0}^3 k \cdot g_k - \frac{3}{10} \sum_{k=0}^3 g_k \quad (42)$$

[00116] The method 200 comprises a smoothing operation 705 and the device 250 comprises a corresponding smoother 755 for smoothing the temporal subframe gains g_k 702 with the interpolated (LLS) gains $g_k^{[LLS]}$ from relation (40) when, for example, the following condition is true:

$$v_{HB} < 0.4 \text{ AND } idx_g \geq 5 \text{ AND } |g_{\text{tilt}}| < 0.2 \quad (43)$$

[00117] The smoothing of the temporal subframe gains g_k 702 is then done by the smoother 755 using, for example, the following relation (44):

$$\bar{g}_k = (1 - \kappa) g_k + \kappa \cdot g_k^{[LLS]}, \quad k = 0, \dots, 3 \quad (44)$$

[00118] where the weight κ is proportional to the voicing parameter v_{HB} 230 (Figure 2) given by relation (21). For example, the weight κ may be calculated using the following relation (45):

$$\kappa = \frac{-0.95}{0.4} (v_{HB} - 0.4) \quad (45)$$

[00119] and limited to a maximum value of 1.0 and a minimum value of 0.0.

[00120] The method 200 comprises a gain-shape quantizing operation 707 and the device 250 comprises a corresponding gain-shape quantizer 757 for quantizing the smoothed temporal subframe gains \bar{g}_k 706. For that purpose, the gain-shape quantizer of the encoder of the SWB TBE tool as described in Reference [1] using, for example 5 bits, can be used as the quantizer 757. The quantized temporal subframe gains \hat{g}_k 708 from the quantizer 757 can be denoted using the following relation (46):

$$\hat{g}_k, \quad k = 0, \dots, 3 \quad (46)$$

[00121] The method 200 comprises an interpolation operation 709 and the device 250 comprises a corresponding interpolator 759 for interpolating, after the quantization operation 707, the quantized temporal subframe gains \hat{g}_k 708 again using the same LLS interpolation procedure as described in relations (40) and (41). The interpolated quantized subframe gains 710 in the four consecutive subframes in a frame can be denoted using the following relation (47):

$$\hat{g}_k^{[LLS]}, \quad k = 0, \dots, 3 \quad (47)$$

[00122] The method 200 comprises a tilt calculation operation 711 and the device 250 comprises a corresponding tilt calculator 761 for calculating the tilt of the interpolated quantized temporal subframe gains $\hat{g}_k^{[LLS]}$ 710 using, for example, relation (42). The tilt of the interpolated quantized temporal subframe gains $\hat{g}_k^{[LLS]}$ 710 can be denoted as $\hat{g}_{tilt}^{[LLS]}$.

[00123] The quantized temporal subframe gains \hat{g}_k 708 are then smoothed when the condition of the following condition (48) is true, where idx_g is the index from relation (32):

$$idx_g \geq 6 \text{ AND } |\hat{g}_{tilt}^{[LLS]}| < 0.12 \quad (48)$$

[00124] For that purpose, the method 200 comprises a quantized gains smoothing operation 713 and the device 250 comprises a corresponding smoother 714 for smoothing the quantized temporal subframe gains \hat{g}_k 708 by means of averaging using, for example, the interpolated temporal subframe gains $\hat{g}_k^{[LLS]}$ 710 from relation (47). For that purpose, the following relation (49) can be used:

$$\tilde{g}_k = 0.7 \cdot \hat{g}_k + 0.3 \cdot \hat{g}_k^{[LLS]}, \quad k = 0, \dots, 3 \quad (49)$$

[00125] The method 200 comprises a frame gain estimating operation 715 and the device 250 comprises a corresponding frame gain estimator 765. The SWB TBE tool uses the frame gain to control the global energy of the synthesized high-band sound signal. The frame gain is estimated by means of energy-matching between (a) the LP-filtered high-band signal y_{HB} 516 of relation (38) multiplied by the smoothed quantized temporal subframe gains \tilde{g}_k 714 from relation (49) and (b) the high-band target signal $s_{HB}(n)$ 210 of relation (3). The LP-filtered high-band signal y_{HB} 516 of relation (38) is multiplied by the smoothed quantized temporal subframe gains \tilde{g}_k 714 using, for example, the following relation (50):

$$\tilde{y}_{HB}(n+k \cdot 80) = \tilde{g}_k \cdot y_{HB}(n+k \cdot 80), \quad k = 0, \dots, 3, n = 0, \dots, 80-1 \quad (50)$$

[00126] The details of the frame gain estimation operation 715 are described in Reference [1]. The estimated frame gain parameter is denoted as g_f (see 716).

[00127] The method 200 comprises an operation 717 of calculating a synthesis high-band signal 718 and the device 250 comprises a calculator 767 for performing the operation 717. The calculator 767 may modify the estimated frame gain g_f 717 under some specific conditions. For example, the frame gain g_f can be attenuated according to relation (51) under given values of high-band voicing factor v_{HB} 230 (Figure 2) and MSE excess error energy E_{err} as shown in relation (51):

$$g_f \leftarrow f_{att} \cdot g_f, \quad \text{if } v_{HB} > 0.1 \text{ AND } E_{err} > 5.0 \quad (51)$$

[00128] where E_{err} is the MSE excess error energy calculated in relation (30) and f_{att} is an attenuation factor for example calculated as:

$$f_{att} = 1.0 - 0.04(E_{err} - 5.0) \quad (52)$$

[00129] Further modifications to the frame gain g_f under some specific conditions are described in Reference [1].

[00130] The calculator 767 then quantizes the modified frame gain using the frame gain quantizer of the encoder of the SWB TBE tool of Reference [1].

[00131] Finally, the calculator 767 determines the synthesized high-band sound signal 718 using, for example, the following relation (53):

$$\tilde{s}_{HB}(n) = g_f \cdot \tilde{y}_{HB}(n), \quad n = 0, \dots, N-1 \quad (53)$$

7. Example Configuration of Hardware Components

[00132] Figure 9 is a simplified block diagram of an example configuration of hardware components forming the above-described method 200 and device 250 for time-domain bandwidth extension of an excitation signal during encoding/decoding of a cross-talk signal (herein after “method 200 and device 250”).

[00133] The method 200 and device 250 may be implemented as a part of a mobile terminal, as a part of a portable media player, or in any similar device. The device 250 (identified as 900 in Figure 9) comprises an input 902, an output 904, a processor 906 and a memory 908.

[00134] The input 902 is configured to receive the input signal. The output 904 is configured to supply the time-domain bandwidth expanded excitation signal. The input 902 and the output 904 may be implemented in a common module, for example a serial input/output device.

[00135] The processor 906 is operatively connected to the input 902, to the output 904, and to the memory 908. The processor 906 is realized as one or more processors for executing code instructions in support of the functions of the various operations and elements of the above described method 200 and device 250 as shown in the accompanying figures and/or as described in the present disclosure.

[00136] The memory 908 may comprise a non-transient memory for storing code instructions executable by the processor 906, specifically, a processor-readable memory comprising/storing non-transitory instructions that, when executed, cause a processor to implement the operations and elements of the method 200 and device 250. The memory 908 may also comprise a random access memory or buffer(s) to store intermediate processing data from the various functions performed by the processor 908.

[00137] Those of ordinary skill in the art will realize that the description of the method 200 and device 250 are illustrative only and are not intended to be in any way limiting. Other embodiments will readily suggest themselves to such persons with ordinary skill in the art having the benefit of the present disclosure. Furthermore, the disclosed method 200 and device 250 may be customized to offer valuable solutions to existing needs and problems of encoding and decoding sound.

[00138] In the interest of clarity, not all of the routine features of the implementations of the method 200 and device 250 are shown and described. It will, of course, be appreciated that in the development of any such actual implementation of the method 200 and device 250, numerous implementation-specific decisions may need to be made in order to achieve the developer's specific goals, such as compliance with application-, system-, network- and business-related constraints, and that these specific goals will vary from one implementation to another and from one developer to another. Moreover, it will be appreciated that a development effort might be complex and time-consuming, but would nevertheless be a routine undertaking of engineering for those of ordinary skill in the field of sound processing having the benefit of the present disclosure.

[00139] In accordance with the present disclosure, the elements, processing operations, and/or data structures described herein may be implemented using various types of operating systems, computing platforms, network devices, computer programs, and/or general purpose machines. In addition, those of ordinary skill in the art will recognize that devices of a less general purpose nature, such as hardwired devices,

field programmable gate arrays (FPGAs), application specific integrated circuits (ASICs), or the like, may also be used. Where a method comprising a series of operations and sub-operations is implemented by a processor, computer or a machine and those operations and sub-operations may be stored as a series of non-transitory code instructions readable by the processor, computer or machine, they may be stored on a tangible and/or non-transient medium.

[00140] Processing operations and elements of the method 200 and device 250 as described herein may comprise software, firmware, hardware, or any combination(s) of software, firmware, or hardware suitable for the purposes described herein.

[00141] In the method 200 and device 250, the various processing operations and sub-operations may be performed in various orders and some of the processing operations and sub-operations may be optional.

[00142] Although the present disclosure has been described hereinabove by way of non-restrictive, illustrative embodiments thereof, these embodiments may be modified at will within the scope of the appended claims without departing from the spirit and nature of the present disclosure.

8. References

[00143] The present disclosure mentions the following references, of which the full content is incorporated herein by reference:

- [1] 3GPP TS 26.445, "EVS Codec Detailed Algorithmic Description," 3GPP Technical Specification (Release 12) (2014) - Sections 5.2.6.1 and 6.1.3.1.
- [2] Bessette, B., Lefebvre, R., Salami, R. et al. "Techniques for high-quality ACELP coding of wideband speech". Int. Conference EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event, Aalborg, Denmark, September 3-7, 2001.

What is claimed is:

1. A method for time-domain bandwidth expansion of an excitation signal during decoding of a cross-talk sound signal, comprising:
 - decoding a high-band mixing factor received in a bitstream; and
 - mixing a low-band excitation signal and a random noise excitation signal using the high-band mixing factor to produce the time-domain bandwidth expanded excitation signal.
2. The method according to claim 1, wherein decoding the high-band mixing factor comprises decoding a quantized normalized gain received in the bitstream and calculating the high-band mixing factor using the decoded quantized normalized gain.
3. The method according to claim 1 or 2, comprising interpolating an energy of the random noise excitation signal between a previous frame and a current frame of the sound signal to smoothen transition between the previous and current frames.
4. The method according to claim 3, comprising, for interpolating the energy of the random noise excitation signal, scaling the random noise signal in a portion of the current frame.
5. The method according to any one of claims 1 to 4, comprising interpolating the high-band mixing factor between a previous and a current frame of the sound signal to ensure smooth transition between the previous and current frames.
6. The method according to any one of claims 1 to 4, comprising estimating quantized gain/shape parameters.
7. A method for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising:

calculating (a) a high-band residual signal using the sound signal and (b) a temporal envelope of the high-band residual signal;

calculating a high-band voicing factor based on the temporal envelope of the high-band residual signal;

calculating a high-band mixing factor usable for mixing a low-band excitation signal and a random noise excitation signal to produce the time-domain bandwidth expanded excitation signal; and

estimating gain/shape parameters using the high-band voicing factor.

8. A method for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising:

calculating (a) a high-band residual signal using the sound signal and (b) a temporal envelope of the high-band residual signal; and

calculating a high-band voicing factor based on the temporal envelope of the high-band residual signal.

9. A method for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising:

calculating a high-band mixing factor usable for mixing a low-band excitation signal and a random noise excitation signal to produce the time-domain bandwidth expanded excitation signal.

10. The method according to claim 7 or 8, wherein calculating the high-band voicing factor comprises (a) calculating a high-band autocorrelation function based on the temporal envelope, and (b) using the high-band autocorrelation function to calculate the high-band voicing factor.

11. The method according to any one of claims 7, 8 and 10, wherein calculating the high-band voicing factor comprises downsampling the temporal envelope of the high-band residual signal by a given factor.

12. The method according to claim 11, wherein calculating the high-band voicing factor comprises dividing the downsampled temporal envelope into a number of segments and calculating a mean value of each segment of the downsampled temporal envelope.

13. The method according to claim 12, wherein calculating the high-band voicing factor comprises per-segment normalization of the downsampled temporal envelope of the high-band residual signal.

14. The method according to claim 13, wherein per-segment normalization of the downsampled temporal envelope comprises (a) calculating segmental normalization factors from the calculated means values, (b) interpolating the segmental normalization factors in a current frame, and (c) normalizing the downsampled temporal envelope using the interpolated segmental normalization factors.

15. The method according to any one of claims 7, 8 and 10 to 14, comprising calculating a tilt of the temporal envelope of the high-band residual signal based on a linear least squares method.

16. The method according to claim 14, wherein calculating the high-band voicing factor comprises (a) calculating a high-band autocorrelation function based on the normalized temporal envelope, and (b) using the high-band autocorrelation function to calculate the high-band voicing factor.

17. The method according to any one of claims 7 and 9 to 16, wherein calculating the high-band mixing factor comprises calculating and quantizing a gain from which the high-band mixing factor is obtained.

18. The method according to claim 17, wherein calculating the high-band mixing factor comprises generating a random noise excitation signal.
19. The method according to claim 18, wherein generating the random noise excitation signal comprises power-normalizing the random noise excitation signal.
20. The method according to claim 18 or 19, wherein calculating the high-band mixing factor comprises (a) mixing the low-band excitation signal with the random noise excitation signal, and (b) minimizing a mean squared error between the mixed excitation signal and a high-band residual signal calculated from the sound signal.
21. The method according to any one of claims 18 to 20, wherein calculating the high-band mixing factor comprises (a) calculating a temporal envelope of the random noise excitation signal, (b) calculating a temporal envelope of the low-band excitation signal, and (c) finding respective gains for the temporal envelopes of the random noise excitation signal and the low-band excitation signal by means of mean squared error minimization process.
22. The method according to claim 21, wherein calculating the high-band mixing factor comprises scaling the gains for the temporal envelopes of the random noise excitation signal and the low-band excitation signal.
23. The method according to claim 22, wherein scaling the gains comprises obtaining a single gain parameter and wherein calculating the high-band mixing factor comprises quantizing the single gain parameter to obtain the said quantized gain from which the high-band mixing factor is obtained.
24. The method according to claim 7 or 8, comprising estimating gain/shape parameters using the high-band voicing factor.

25. The method according to any one of claims 7 to 24, wherein the gain/shape parameters are selected from the group comprising:
- a spectral shape of a high-band target signal;
 - subframe gains of the high-band target signal;
 - a frame gain parameter.
26. The method according to any one of claims 7 to 25, wherein estimating the gain/shape parameters comprises calculating a temporal tilt of the gain/shape parameters.
27. The method according to claim 26, wherein calculating the temporal tilt comprises interpolating the gain/shape parameters.
28. The method according to claim 27, wherein interpolating the gain/shape parameters comprises using a linear least squares method.
29. The method according to any one of claims 7 to 28, wherein estimating the gain/shape parameters comprises smoothing the gain/shape parameters using an adaptive weight parameter.
30. The method according to claim 29, comprising calculating the adaptive weight parameter using the high-band voicing factor.
31. The method according to claim 29 or 30, comprising smoothing of the gain/shape parameters using the adaptive weight parameter in response to a given condition involving the high-band voicing factor.
32. The method according to any one of claims 29 to 31, wherein estimating the gain/shape parameters comprises quantizing the smoothed gain/shape parameters.

33. The method according to claim 32, wherein estimating the gain/shape parameters comprises interpolating the quantized gain/shape parameters.
34. The method according to claim 32 or 33, wherein estimating the gain/shape parameters comprises smoothing the quantized gain/shape parameters.
35. The method according to claim 34, wherein smoothing the quantized gain/shape parameters is performed by means of averaging of the quantized interpolated gain/shape parameters.
36. The method according to any one of claims 7 to 35, wherein estimating the gain/shape parameters comprises adaptive attenuation of a frame gain parameter using a MSE excess error.
37. A device for time-domain bandwidth expansion of an excitation signal during decoding of a cross-talk sound signal, comprising:
a decoder of a high-band mixing factor received in a bitstream; and
a mixer of a low-band excitation signal and a random noise excitation signal using the high-band mixing factor to produce the time-domain bandwidth expanded excitation signal.
38. The device according to claim 37, wherein the decoder of the high-band mixing factor decodes a quantized normalized gain received in the bitstream and calculates the high-band mixing factor using the decoded quantized normalized gain.
39. The device according to claim 37 or 38, comprising a generator of the random noise excitation signal which interpolates an energy of the random noise excitation signal between a previous frame and a current frame of the sound signal to smoothen transition between the previous and current frames.

40. The method according to claim 39, comprising, for interpolating the energy of the random noise excitation signal, the generator of the random noise excitation signal scales the random noise signal in a portion of the current frame.

41. The device according to any one of claims 37 to 40, wherein the decoder of the high-band mixing factor interpolates the high-band mixing factor between a previous and a current frame of the sound signal to ensure smooth transition between the previous and current frames.

42. The device according to any one of claims 37 to 40, comprising an estimator of quantized gain/shape parameters.

43. A device for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising:

- a calculator of (a) a high-band residual signal using the sound signal and (b) a temporal envelope of the high-band residual signal;

- a calculator of a high-band voicing factor based on the temporal envelope of the high-band residual signal;

- a calculator of a high-band mixing factor usable for mixing a low-band excitation signal and a random noise excitation signal to produce the time-domain bandwidth expanded excitation signal; and

- an estimator of gain/shape parameters using the high-band voicing factor.

44. A device for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising:

- a calculator of (a) a high-band residual signal using the sound signal and (b) a temporal envelope of the high-band residual signal; and

- a calculator of a high-band voicing factor based on the temporal envelope of the high-band residual signal.

45. A device for time-domain bandwidth expansion of an excitation signal during encoding of a cross-talk sound signal, comprising:

a calculator of a high-band mixing factor usable for mixing a low-band excitation signal and a random noise excitation signal to produce the time-domain bandwidth expanded excitation signal.

46. The device according to claim 43 or 44, wherein the calculator of the high-band voicing factor calculates a high-band autocorrelation function based on the temporal envelope, and uses the high-band autocorrelation function to calculate the high-band voicing factor.

47. The device according to any one of claims 43, 44 and 46, wherein the calculator of the high-band voicing factor comprises a downsampler of the temporal envelope of the high-band residual signal by a given factor.

48. The device according to claim 47, wherein the calculator of the high-band voicing factor comprises a divider of the downsampled temporal envelope into a number of segments and a calculator of a mean value of each segment of the downsampled temporal envelope.

49. The device according to claim 48, wherein the calculator of the high-band voicing factor comprises a per-segment normalizer of the downsampled temporal envelope of the high-band residual signal.

50. The device according to claim 49, wherein the per-segment normalizer (a) calculates segmental normalization factors from the calculated means values, (b) interpolates the segmental normalization factors in a current frame, and (c) normalizes the downsampled temporal envelope using the interpolated segmental normalization factors.

51. The device according to any one of claims 43, 44 and 46 to 50, comprising a calculator of a tilt of the temporal envelope of the high-band residual signal based on a linear least squares method.

52. The device according to claim 50, wherein the calculator of the high-band voicing factor comprises a calculator of a high-band autocorrelation function based on the normalized temporal envelope, and uses the high-band autocorrelation function to calculate the high-band voicing factor.

53. The device according to any one of claims 43 and 45 to 52, wherein the calculator or the high-band mixing factor calculates and quantizes a gain forming the high-band mixing factor.

54. The device according to claim 53, wherein the calculator of the high-band mixing factor comprises a generator of a random noise excitation signal.

55. The device according to claim 54, comprising a power-normalizer of the random noise excitation signal to a power of the low-band excitation signal.

56. The device according to claim 54 or 55, wherein the calculator of the high-band mixing factor (a) combines the low-band excitation signal with the random noise excitation signal, and (b) minimizes a mean squared error between the mixed excitation signal and a high-band residual signal calculated from the sound signal.

57. The device according to any one of claims 54 to 56, wherein the calculator of the high-band mixing factor (a) comprises a calculator of a temporal envelope of the random noise excitation signal and a calculator of a temporal envelope of the low-band excitation signal, and (b) finds respective gains for the temporal envelopes of the random noise excitation signal and the low-band excitation signal by means of a mean squared error minimization process.

58. The device according to claim 57, wherein the calculator of the high-band mixing factor scales the gains for the temporal envelopes of the random noise excitation signal and the low-band excitation signal.

59. The device according to claim 58, wherein, to scale the gains for the temporal envelopes of the random noise excitation signal and the low-band excitation signal, the calculator of the high-band mixing factor calculates a single gain parameter and quantizes the single gain parameter to obtain the said quantized gain forming the high-band mixing factor.

60. The device according to claim 43 or 44, comprising an estimator of gain/shape parameters using the high-band voicing factor.

61. The device according to any one of claims 43 to 60, wherein the gain/shape parameters are selected from the group comprising:

- a spectral shape of a high-band target signal;
- subframe gains of the high-band target signal;
- a frame gain parameter.

62. The device according to any one of claims 43 to 61, wherein the gain/shape parameters comprise subframe gains of the high-band target signal, and wherein the estimator of the gain/shape parameters comprises a calculator of a temporal tilt of the subframe gains.

63. The device according to claim 62, wherein the calculator of the temporal tilt comprises an interpolator of the subframe gains.

64. The device according to claim 63, wherein the interpolator of the subframe gains uses a linear least squares method.

65. The device according to any one of claims 43 to 64, wherein the gain/shape parameters comprise subframe gains of the high-band target signal, and the estimator of the gain/shape parameters comprises a smoother of the subframe gains using an adaptive weight parameter.

66. The device according to claim 65, wherein the smoother of the subframe gains calculates the adaptive weight parameter using the high-band voicing factor.

67. The device according to claim 29 or 30, wherein the smoother of the gain/shape parameters performs smoothing of the gain/shape parameters using the adaptive weight parameter in response to a given condition involving the high-band voicing factor.

68. The device according to any one of claims 65 to 67, wherein the estimator of the gain/shape parameters comprises a quantizer of the subframe gains.

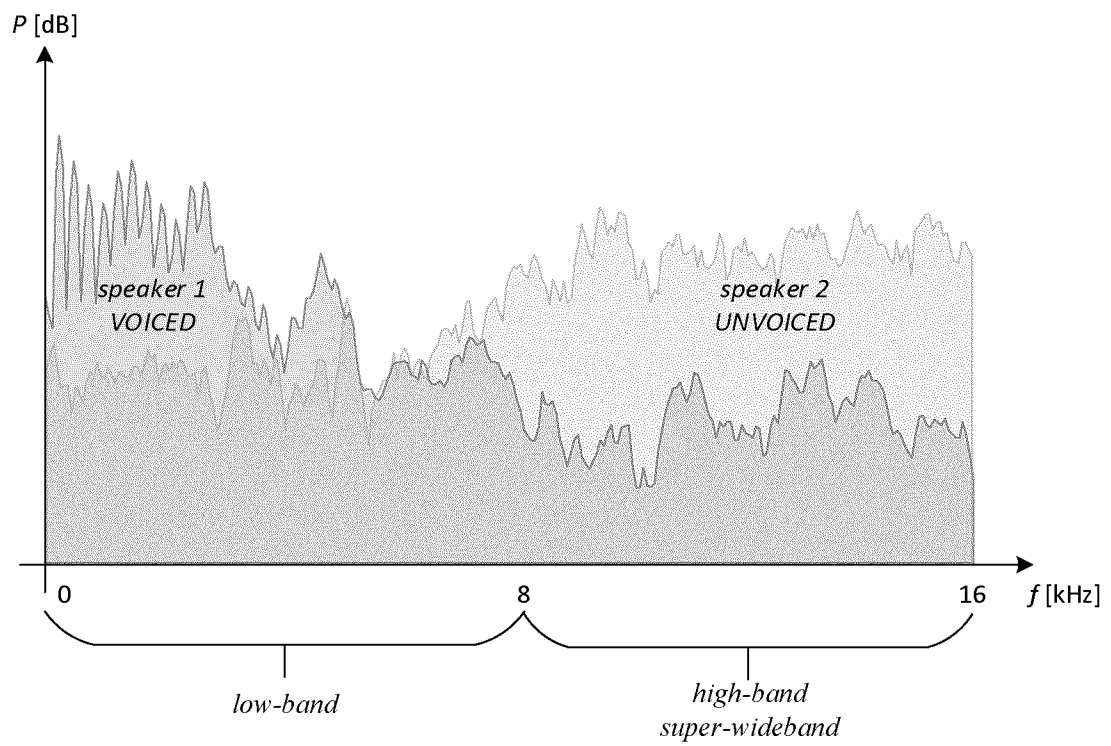
69. The device according to claim 68, wherein the estimator of the gain/shape parameters comprises an interpolator of the quantized subframe gains.

70. The device according to claim 68 or 69, wherein the estimator of the gain/shape parameters comprises a smoother of the subframe gains.

71. The device according to claim 70, wherein the smoother of the subframe gains smoothes the quantized gain/shape parameters by means of averaging of the quantized interpolated gain/shape parameters.

72. The device according to any one of claims 43 to 71, wherein the gain/shape parameters comprise subframe gains of the high-band target signal, and wherein the

estimator of the gain/shape parameters performs adaptive attenuation of a frame gain parameter using a MSE excess error.

*Fig. 1*

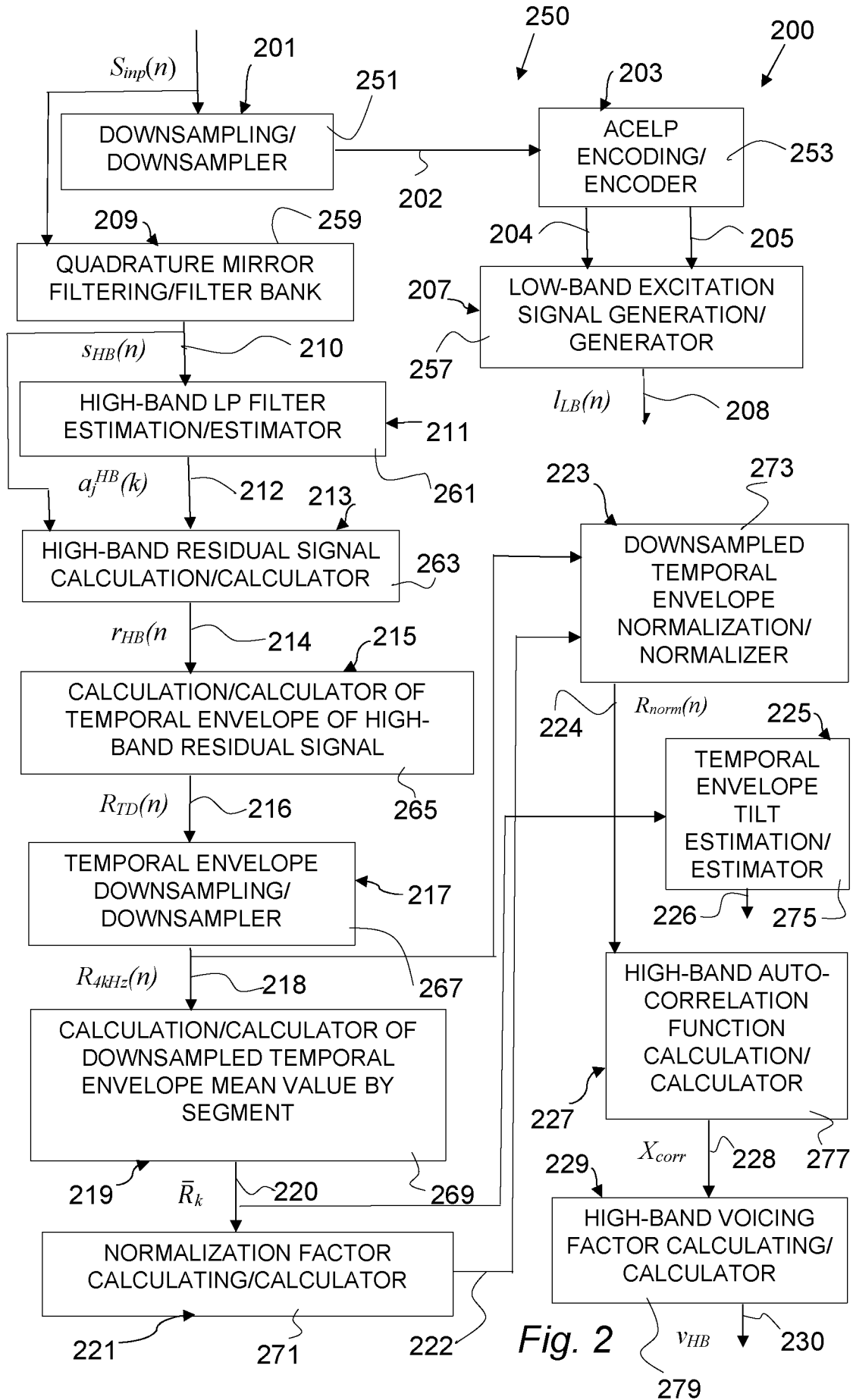


Fig. 2

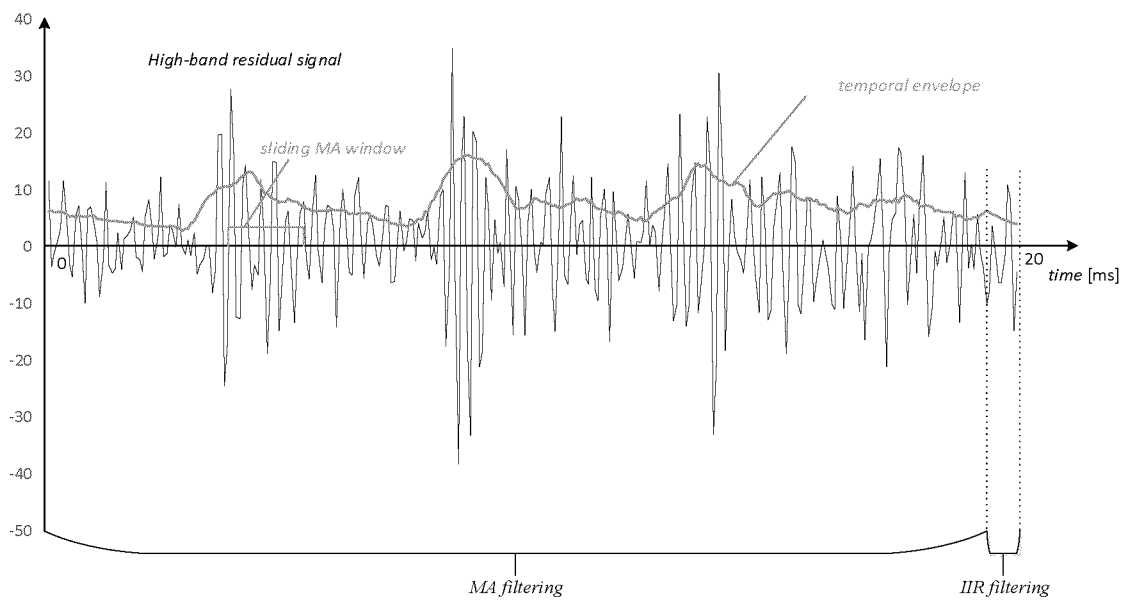


Fig. 3

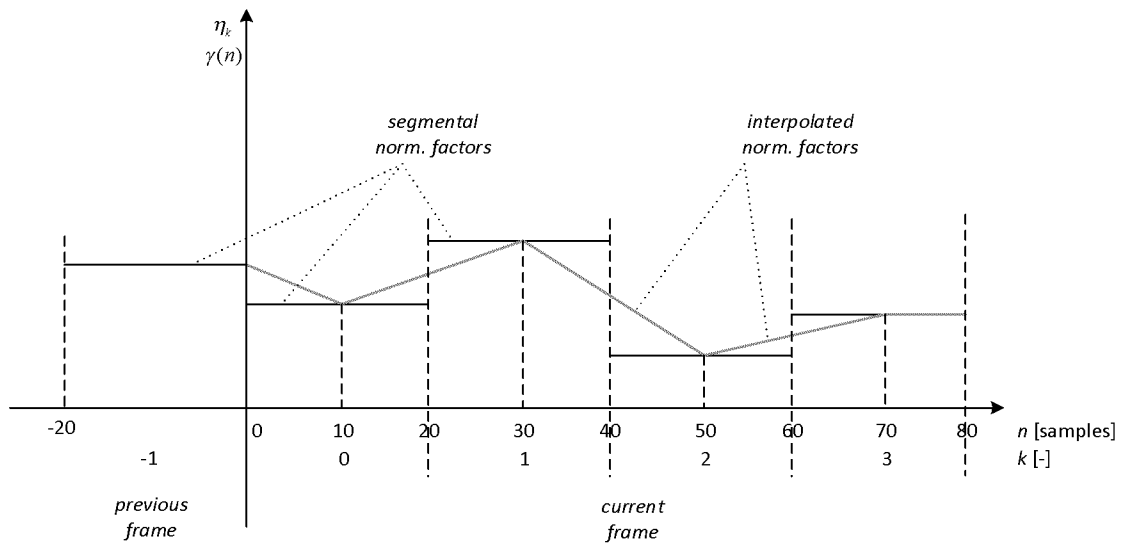


Fig. 4

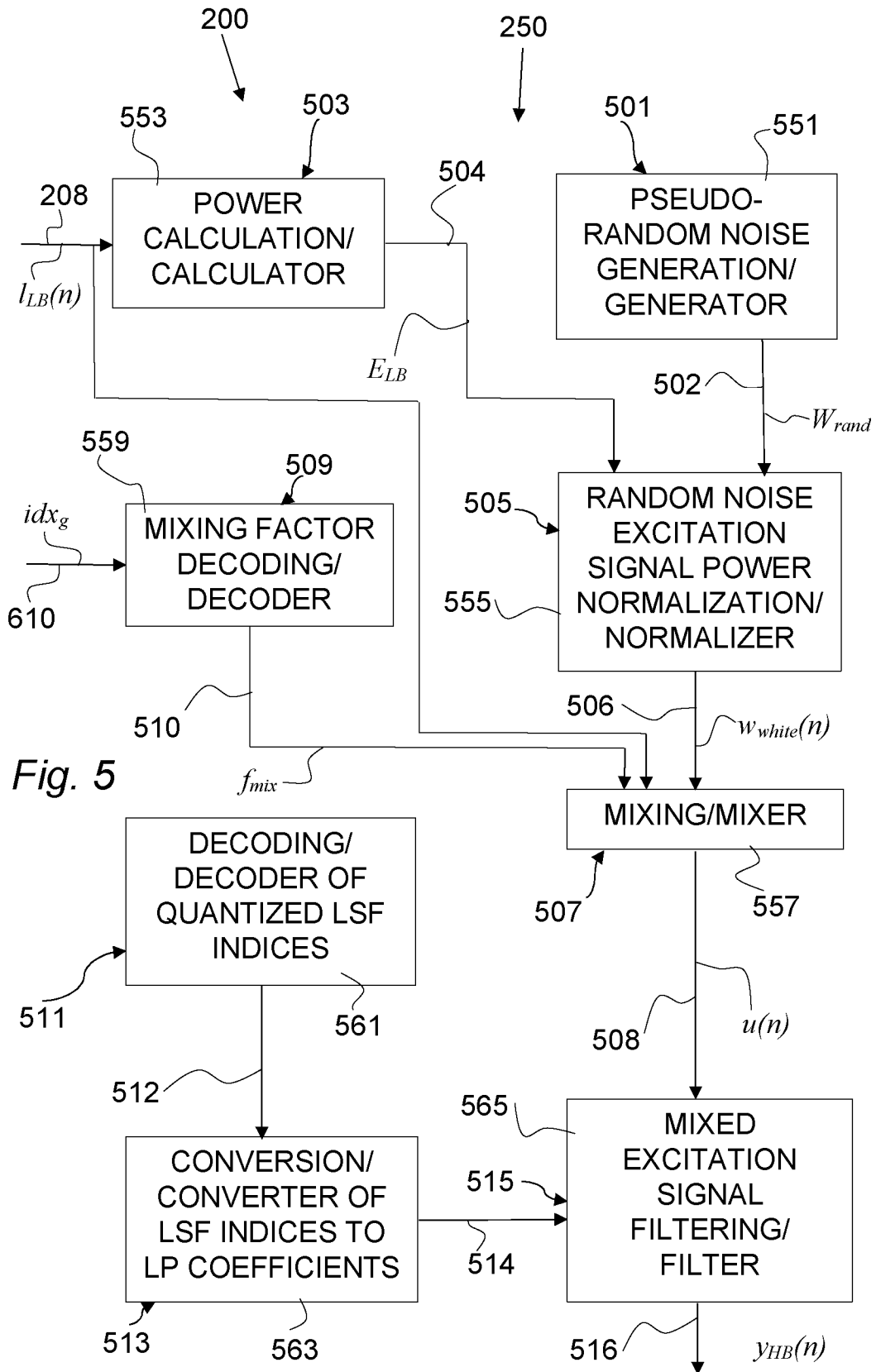


Fig. 5

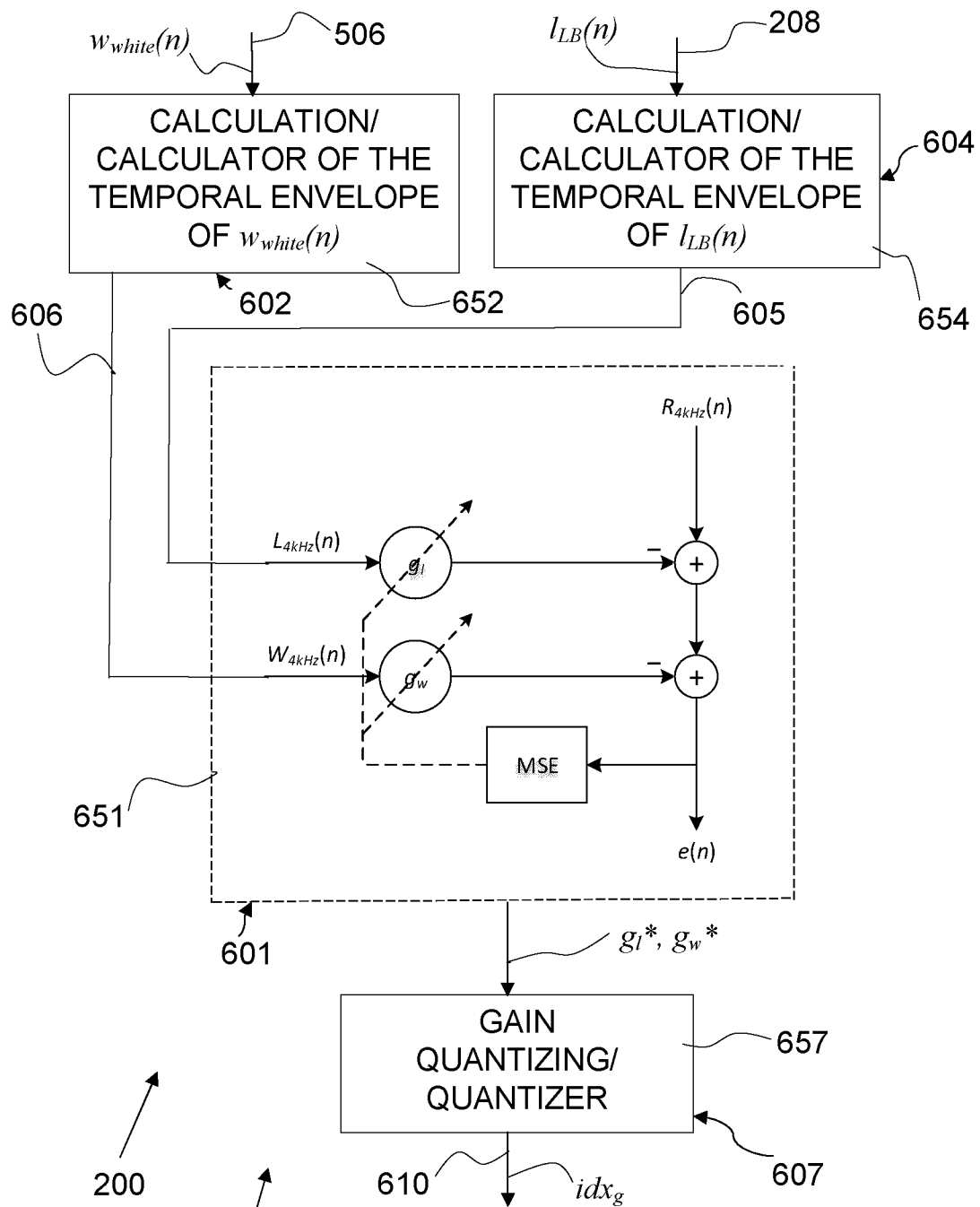


Fig. 6

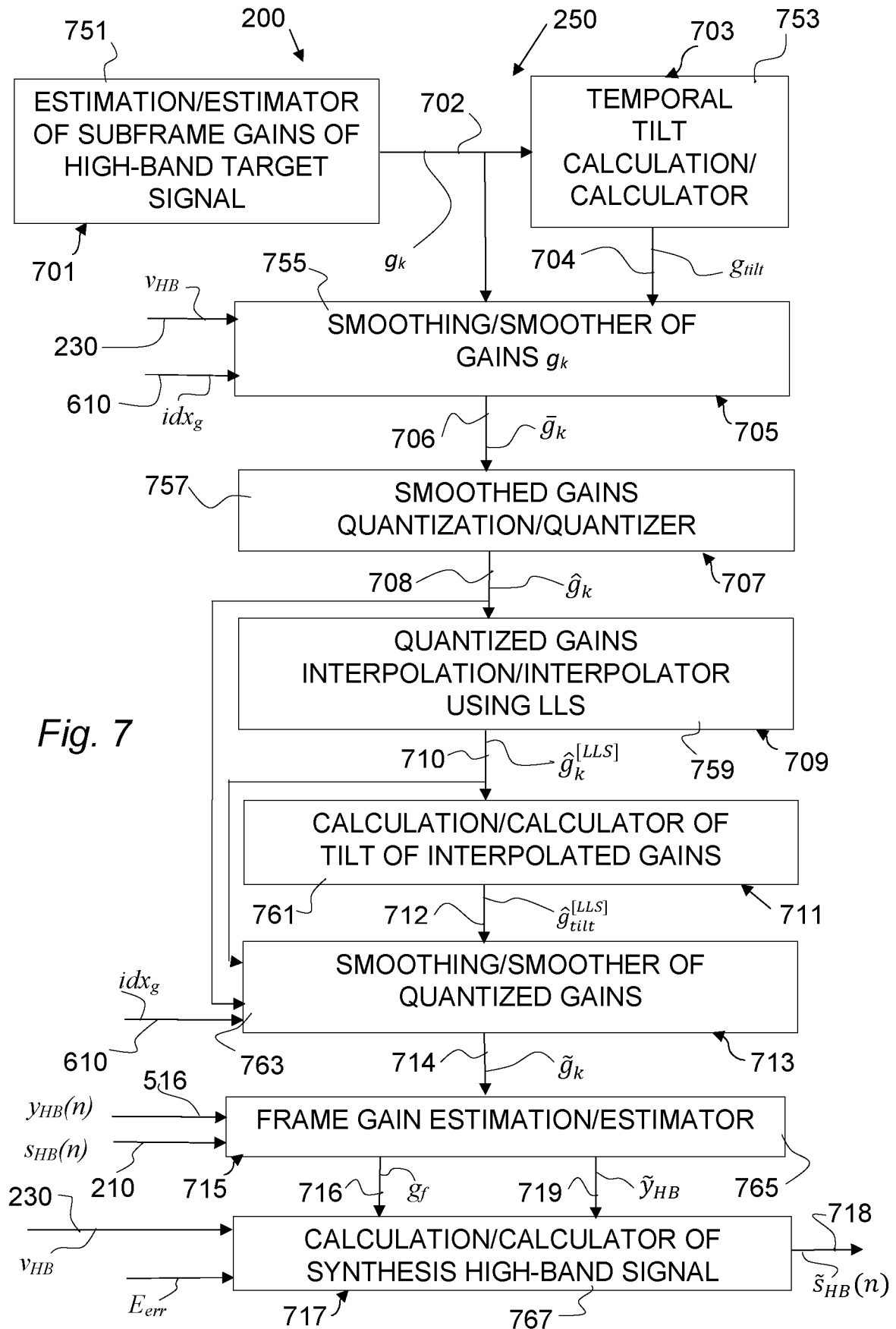


Fig. 7

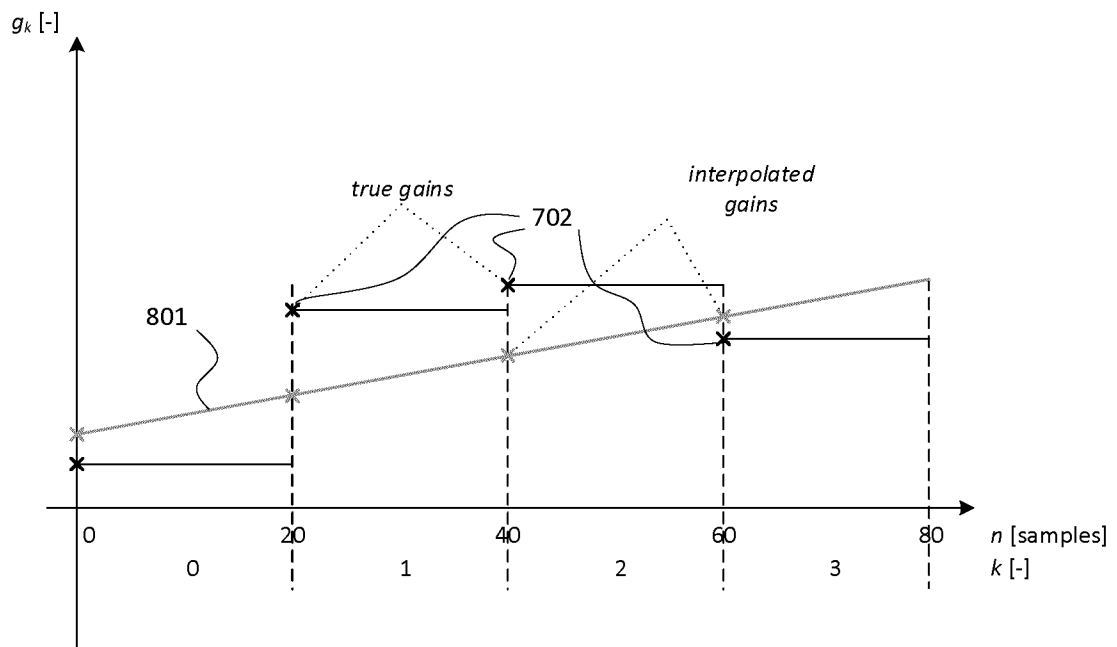


Fig. 8

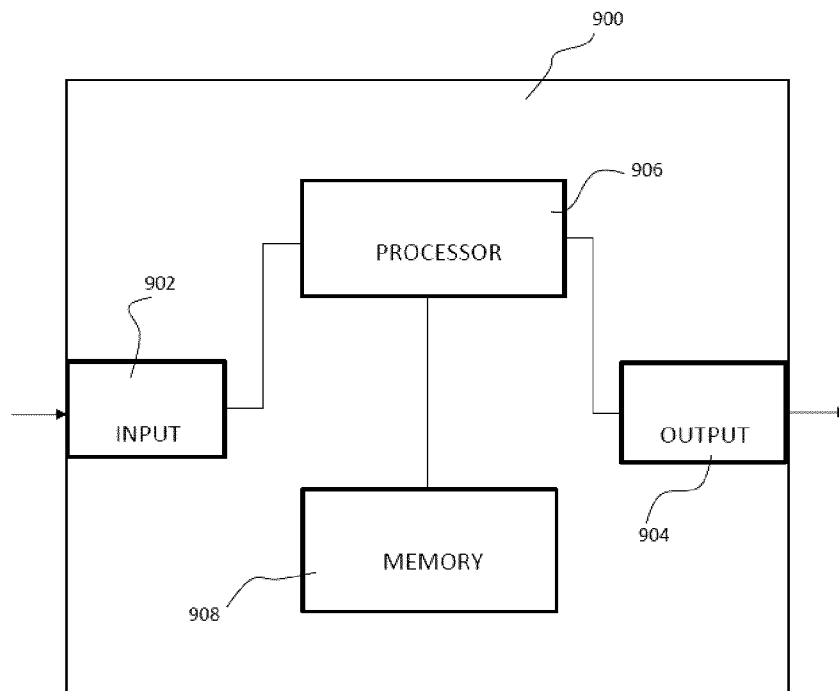


FIGURE 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2023/050117

A. CLASSIFICATION OF SUBJECT MATTER

IPC: **G10L 19/00** (2013.01)

CPC: **G10L 19/00** (2020.01)

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)

IPC: **G10L 19/00** (2013.01)

CPC: **G10L 19/00** (2020.01)

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

Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used)

Databases: Questel Orbit – FAMPAT, SCOPUS, IEEE Xplore

Keywords : crosstalk, bandwidth expansion, SWB TBE, 3GPP EVS, rattling noise, mixing factor, gain, shape, envelope, time-domain

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2015/0162010 A1 (ISHIKAWA, T. et al.) 11 June 2015 (11-06-2015) * Entire Document *	1-72
A	KANIEWSKA, M. et al., " <i>Enhanced AMR-WB bandwidth extension in 3GPP EVS codec</i> ". 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 25 February 2016 (25-02-2016), pp. 652-656, [online] [retrieved on 20 March 2023 (20-03-2023)]. Retrieved from the Internet: < https://ieeexplore.ieee.org/document/7418277 > * Entire Document *	1-72

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"D" document cited by the applicant in the international application	"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
"E" earlier application or patent but published on or after the international filing date	"&" document member of the same patent family
"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	

Date of the actual completion of the international search
15 April 2023 (15-04-2023)

Date of mailing of the international search report
15 April 2023 (15-04-2023)

Name and mailing address of the ISA/CA
Canadian Intellectual Property Office
Place du Portage I, C114 - 1st Floor, Box PCT
50 Victoria Street
Gatineau, Quebec K1A 0C9
Facsimile No.: 819-953-2476

Authorized officer

Nicholas Armstrong (819) 962-8763

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CA2023/050117

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ATTI, V. et al., " <i>Super-wideband bandwidth extension for speech in the 3GPP EVS codec</i> ". <i>2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)</i> , South Brisbane, QLD, Australia, 6 August 2015 (06-08-2015), pp. 5927-5931, [online] [retrieved on 20 March 2023 (20-03-2023)]. Retrieved from the Internet: < https://ieeexplore.ieee.org/document/7179109 > * Entire Document *	1-72
A	US 2015/0162008 A1 (VILLETTE, S. et al.) 11 June 2015 (11-06-2015) * Entire Document *	1-72

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CA2023/050117

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
US2015162010A1	11 June 2015 (11-06-2015)	US9424847B2	23 August 2016 (23-08-2016)
		CN104584124A	29 April 2015 (29-04-2015)
		CN104584124B	16 April 2019 (16-04-2019)
		EP2950308A1	02 December 2015 (02-12-2015)
		EP2950308A4	24 February 2016 (24-02-2016)
		EP2950308B1	19 February 2020 (19-02-2020)
		JPWO2014115225A1	19 January 2017 (19-01-2017)
		JP6262668B2	17 January 2018 (17-01-2018)
		WO2014115225A1	31 July 2014 (31-07-2014)
US2015162008A1	11 June 2015 (11-06-2015)	US9293143B2	22 March 2016 (22-03-2016)
		CN105814629A	27 July 2016 (27-07-2016)
		EP3080804A1	19 October 2016 (19-10-2016)
		JP2017503192A	26 January 2017 (26-01-2017)
		KR20160096119A	12 August 2016 (12-08-2016)
		WO2015088919A1	18 June 2015 (18-06-2015)



(12) 发明专利申请

(10) 申请公布号 CN 118742956 A

(43) 申请公布日 2024. 10. 01

(21) 申请号 202380020248.1

(74) 专利代理机构 北京市柳沈律师事务所

(22) 申请日 2023.01.27

11105

专利代理师 胡琪

(30) 优先权数据

63/306,291 2022.02.03 US

(51) Int.Cl.

G10L 21/038 (2013.01)

(85) PCT国际申请进入国家阶段日

G10L 19/08 (2013.01)

2024.08.02

(86) PCT国际申请的申请数据

PCT/CA2023/050117 2023.01.27

(87) PCT国际申请的公布数据

WO2023/147650 EN 2023.08.10

(71) 申请人 沃伊斯基公司

地址 加拿大魁北克

(72) 发明人 V·马列诺夫斯基 M·耶利内克

权利要求书5页 说明书15页 附图7页

(54) 发明名称

针对串扰场景的时域超宽带带宽扩展

(57) 摘要

一种用于在串扰声音信号的解码期间对激励信号进行时域带宽扩展的方法,包括:对在比特流中接收的高频带混合因子进行解码,以及使用该高频带混合因子混合低频带激励信号和随机噪声激励信号以产生时域扩展的激励信号。一种用于在串扰声音信号的编码期间对激励信号进行时域带宽扩展的方法,包括:使用声音信号计算高频带残差信号和计算高频带残差信号的时间包络;基于高频带残差信号的时间包络来计算高频带发声因子;计算可用于混合低频带激励信号和随机噪声激励信号以产生时域扩展的激励信号的高频带混合因子;以及使用高频带发声因子来估计增益/形状参数。

