



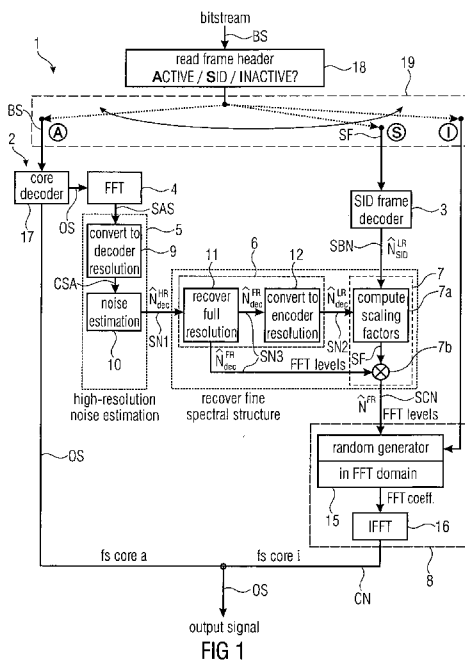
- (51) International Patent Classification:  
G10L 19/012 (2013.01)
- (21) International Application Number:  
PCT/EP2013/077525
- (22) International Filing Date:  
19 December 2013 (19.12.2013)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
61/740,857 21 December 2012 (21.12.2012) US
- (71) Applicant: FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V. [DE/DE]; HansasträÙe 27c, 80686 München (DE).
- (72) Inventors: LOMBARD, Anthony; Taurusstr. 59, 91056 Erlangen (DE). DIETZ, Martin; Deutschhermstraße 37, 90429 Nürnberg (DE). WILDE, Stephan; Sandrartstraße 4, 90419 Nürnberg (DE). RAVELLI, Emmanuel; Brandnerweg 7, 91058 Erlangen (DE). SETIAWAN, Panji; Südliche Stadtmauerstrasse 15c, 91054 Erlangen (DE). MULTRUS, Markus; Etzlaubweg 7, 90469 Nürnberg (DE).
- (74) Agents: ZINKLER, Franz et al.; Postfach 246, 82043 Pullach (DE).

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

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

Published:  
— with international search report (Art. 21(3))

(54) Title: GENERATION OF A COMFORT NOISE WITH HIGH SPECTRO-TEMPORAL RESOLUTION IN DISCONTINUOUS TRANSMISSION OF AUDIO SIGNALS



(57) Abstract: the invention provides an audio decoder being configured for decoding a bit-stream so as to produce therefrom an audio output signal, the bitstream comprising at least an active phase followed by at least an inactive phase, wherein the bitstream has encoded therein at least a silence insertion descriptor frame which describes a spectrum of a background noise, the audio decoder comprising: a silence insertion descriptor decoder configured to decode the silence insertion descriptor frame so as to reconstruct a spectrum of the background noise; a decoding device configured to reconstruct the audio output signal from the bitstream during the active phase; a spectral converter configured to determine a spectrum of the audio output signal a noise estimator or device configured to determine a first spectrum of the noise of the audio output signal based on the spectrum of the audio output signal provided by the spectral converter, wherein the first spectrum of the noise of the audio output signal has a higher spectral resolution than the spectrum of the background noise; a resolution converter configured to establish a second spectrum of the noise of the audio output signal based on the first spectrum of the noise of the audio output signal, wherein the second spectrum of the noise of the audio output signal has a same spectral resolution as the spectrum of the background noise; a comfort noise spectrum estimation device having a scaling factor computing device configured to compute scaling factors for a spectrum for a comfort noise based on the spectrum of the background noise as provided by the silence insertion descriptor decoder and based on the second spectrum of the noise of the audio output signal as provided by the resolution converter and having a comfort noise spectrum generator configured to compute the spectrum for a comfort noise based on the scaling factors; and a comfort noise generator configured to produce the comfort noise during the inactive phase based on the spectrum for the comfort noise.

WO 2014/096279 A1

## Generation of a comfort noise with high spectro-temporal resolution in discontinuous transmission of audio signals

### Description

5

The present invention relates to audio signal processing, and, in particular, to comfort noise addition to audio signals.

Comfort noise generators are usually used in discontinuous transmission  
10 (DTX) of audio signals, in particular of audio signals containing speech. In such a mode the audio signal is first classified in active and inactive frames by a voice activity detector (VAD). Based on the VAD result, only the active speech frames are coded and transmitted at the nominal bit-rate. During long  
15 pauses, where only the background noise is present, the bit-rate is lowered or zeroed and the background noise is coded episodically and parametrically using silence insertion descriptor frames (SID frames). The average bit-rate is then significantly reduced.

The noise is generated during the inactive frames at the decoder side by a  
20 comfort noise generator (CNG). The size of an SID frame is very limited in practice. Therefore, the number of parameters describing the background noise has to be kept as small as possible. To this aim, the noise estimation is not applied directly in the output of the spectral transforms. Instead, it is applied at a lower spectral resolution by averaging the input power spectrum  
25 among groups of bands, e.g., following the Bark scale. The averaging can be achieved either by arithmetic or geometric means. Unfortunately, the limited number of parameters transmitted in the SID frames does not allow to capture the fine spectral structure of the background noise. Hence only the smooth spectral envelope of the noise can be reproduced by the CNG. When  
30 the VAD triggers a CNG frame, the discrepancy between the smooth spectrum of the reconstructed comfort noise and the spectrum of the actual background noise can become very audible at the transitions between active

frames (involving regular coding and decoding of a noisy speech portion of the signal) and CNG frames.

5 An object of the present invention is to provide improved concepts for audio signal processing. More particular, an object of the present invention is to provide improved concepts for comfort noise addition to audio signals. The object of the present invention is achieved by an audio decoder according to claim 1, by a system according to claim 17, by a method according to claim 18 and by a computer program according to claim 19.

10

In one aspect the invention provides an audio decoder being configured for decoding a bitstream so as to produce therefrom an audio output signal, the bitstream comprising at least an active phase followed by at least an inactive phase, wherein the bitstream has encoded therein at least a silence insertion descriptor frame which describes a spectrum of a background noise, the audio decoder comprising:

15

a silence insertion descriptor decoder configured to decode the silence insertion descriptor frame so as to reconstruct a spectrum of the background noise;

20

a decoding device configured to reconstruct the audio output signal from the bitstream during the active phase;

25

a spectral converter configured to determine a spectrum of the audio output signal;

30

a noise estimator device configured to determine a first spectrum of the noise of the audio output signal based on the spectrum of the audio output signal provided by the spectral converter, wherein the first spectrum of the noise of the audio output signal has a higher spectral resolution than the spectrum of the background noise as provided by the silence insertion descriptor decod-

er;

a resolution converter configured to establish a second spectrum of the noise of the audio output signal based on the first spectrum of the noise of the audio output signal, wherein the second spectrum of the noise of the audio output signal has a same spectral resolution as the spectrum of the background noise as provided by the silence insertion descriptor decoder;

a comfort noise spectrum estimation device having a scaling factor computing device configured to compute scaling factors for a spectrum for a comfort noise based on the spectrum of the background noise as provided by the silence insertion descriptor decoder and based on the second spectrum of the noise of the audio output signal as provided by the resolution converter and having a comfort noise spectrum generator configured to compute the spectrum for a comfort noise based on the scaling factors; and

a comfort noise generator configured to produce the comfort noise during the inactive phase based on the spectrum for the comfort noise.

The bitstream contains active phases and inactive phases, wherein an active phase is a phase, which contains wanted components of the audio information, such as speech or music, whereas an inactive phase is a phase, which does not contain any wanted components of the audio information. Inactive phases usually occur during pauses, where no wanted components, such as music or speech, are present. Therefore, inactive phases usually contain solely background noise. The information in the bitstream containing an encoded audio signal is embedded in so called frames, wherein each of these frames contain audio information referring to a certain time. During active phases active frames comprising audio information including audio information regarding the wanted signal may be transmitted within the bitstream. In contrast of that, during inactive phases silence insertion descriptor frames comprising noise information may be transmitted within the bitstream

at a lower average bit-rate compared to the average bit-rate of the active phases.

5 The silence insertion descriptor decoder is configured to decode the silence insertion descriptor frames so as to reconstruct a spectrum of the background noise. However, this spectrum of the background noise does not allow to capture the fine spectral structure of the background noise due to a limited number of parameters transmitted in the silence insertion descriptor frames.

10

The decoding device may be a device or a computer program capable of decoding the audio bitstream, which is a digital data stream containing audio information, during active phases. The decoding process may result in a digital decoded audio output signal, which may be fed to a D/A converter to produce an analogous audio signal, which then may be fed to a loudspeaker, in order to produce an audible signal.

15

The spectral converter may obtain a spectrum of the audio output signal which has a significantly higher spectral resolution than the spectrum of the background noise as provided by the silence insertion descriptor decoder.

20

Therefore, the noise estimator may determine a first spectrum of the noise of the audio output signal based on the spectrum of the audio output signal provided by the spectral converter, wherein the first spectrum of the noise of the audio output signal has a higher spectral resolution than the spectrum of the background noise as provided by the silence insertion descriptor decoder.

25

Further, the resolution converter may establish a second spectrum of the noise of the audio output signal based on the first spectrum of the noise of the audio output signal, wherein the second spectrum of the noise of the audio output signal has a same spectral resolution as the spectrum of the background noise as provided by the silence insertion descriptor decoder.

30

The scaling factor computing device may easily compute scaling factors for a spectrum for a comfort noise based on the spectrum of the background noise as provided by the silence insertion descriptor decoder and based on the  
5 second spectrum of the noise of the audio output signal as provided by the resolution converter as the spectrum of the background noise as provided by the silence insertion descriptor decoder and the second spectrum of the noise of the audio output signal have the same spectral resolution.

10 The comfort noise spectrum generator may establish the spectrum for the comfort noise based on the scaling factors and based on the first spectrum of the noise of the audio output signal as provided by the noise estimation device.

15 Furthermore, the comfort noise generator may produce the comfort noise during the inactive phase based on the spectrum for the comfort noise.

The noise estimates obtained at the decoder contain information about the spectral structure of the background noise, which is more accurate than the  
20 information about the smooth spectral envelope of the background noise contained in the SID frames. However, these estimates cannot be updated during inactive phases since the noise estimation is carried out on the decoded audio output signal during active phases. In contrast, the SID frames deliver new information about the spectral envelope during inactive phases. The decoder according to the invention combines these two sources of information.  
25 The scaling factors may be updated during active phases depending on the noise estimates at the decoder side and during inactive phases depending on the noise estimates contained in the SID frames. The continuous update of the scaling factors ensures that there are no sudden changes of the characteristics of the produced comfort noise.  
30

As the spectrum of the background noise as contained in the SID frames and the second spectrum of the noise of the audio output signal have the same spectral resolution the update of the scaling factors and, hence, of the comfort noise can be done in an easy way, as for each frequency band group of the spectrum of the background noise as contained in the SID frames exactly one frequency band group exists in the second spectrum of the noise of the audio output signal. It has to be noted that in a preferred embodiment the frequency band groups of the spectrum of the background noise as contained in the SID frames and the frequency band groups of the second spectrum of the noise of the audio output signal correspond to each other.

Further, as the spectrum of the background noise as contained in the SID frames and the second spectrum of the noise of the audio output signal have the same spectral resolution the update of the scaling factors produces no or only barely audible artifacts.

According to a preferred embodiment of the invention the spectral analyzer comprises a fast Fourier transformation device. A fast Fourier transform (FFT) is an algorithm to compute a discrete Fourier transform (DFT) and its inverse, which requires only low computational effort. Therefore, the fast Fourier transformation device may calculate the spectrum of the audio output signal in an easy way.

According to a preferred embodiment of the invention the noise estimator device at the decoder comprises a converter device configured to convert the spectrum of the audio output signal into a converted spectrum of the audio output signal which has in general a much lower spectral resolution.. By providing the converted spectrum of the audio output signal the complexity of subsequent computational steps may be reduced.

30

According to a preferred embodiment of the invention the noise estimator device comprises a noise estimator configured to determine the first spec-

trum of the noise of the audio output signal based on the converted spectrum of the audio output signal provided by the converter device. When the converted spectrum of the audio output signal is used as a basis for the noise estimation at the decoder computational efforts may be reduced without lowering the quality of the noise estimation.

According to a preferred embodiment of the invention the scaling factor computing device is configured to compute the scaling factors according to the formula

$$\hat{S}^{LR}(i) = \frac{\hat{N}_{SID}^{LR}(i)}{\hat{N}_{dec}^{LR}(i)},$$

wherein  $\hat{S}^{FR}(i)$  denotes a scaling factor for a frequency band group  $i$  of the comfort noise, wherein  $\hat{N}_{SID}^{LR}(i)$  denotes a level of a frequency band group  $i$  of the spectrum of the background noise as contained in the SID frames, wherein  $\hat{N}_{dec}^{LR}(i)$  denotes a level of a frequency band group  $i$  of the second spectrum of the noise of the audio output signal, wherein  $i = 0, \dots, L^{LR} - 1$ , wherein  $L^{LR}$  is the number of frequency band groups of the spectrum of the background noise as contained in the SID frames and of the second spectrum of the noise of the audio output signal. By these features the scaling factors may be computed in an easy manner.

According to a preferred embodiment of the invention the comfort noise spectrum generator is configured to compute the spectrum of the comfort noise based on the scaling factors and based on the first spectrum of the noise of the audio output signal as provided by the noise estimation device. By these features the comfort noise spectrum may be computed in such way that it has the spectral resolution of the first spectrum of the noise of the audio output signal, which is in general much higher than the spectral resolution obtained from SID frames.

According to a preferred embodiment of the invention the comfort noise spectrum generator is configured to compute the spectrum of the comfort noise according to the formula  $\hat{N}^{FR}(k) = \hat{S}^{LR}(i) \cdot \hat{N}_{dec}^{HR}(k)$ , wherein  $\hat{N}^{FR}(k)$  denotes



a level of a frequency band  $k$  of the spectrum of the comfort noise, wherein  $\hat{S}^{LR}(i)$  denotes a scaling factor of a frequency band group  $i$  of the spectrum of the background noise as contained in the SID frames and of the second spectrum of the noise of the audio output signal, wherein  $\hat{N}_{dec}^{HR}(k)$  denotes a  
5 level of a frequency band  $k$  of the first spectrum of the noise of the audio output signal, wherein  $k = b^{LR}(i), \dots, b^{LR}(i + 1) - 1$ , wherein  $b^{LR}(i)$  is a first frequency band of one of the frequency band groups, wherein  $i = 0, \dots, L^{LR} - 1$ , wherein  $L^{LR}$  is the number of frequency band groups of the spectrum of the background noise as contained in the SID frames and of the second spec-  
10 trum of the noise of the audio output signal. By these features the spectrum of the comfort noise may be computed at the high-resolution in an easy way.

According to a preferred embodiment of the invention the resolution converter comprises a first converter stage configured to establish a third spectrum  
15 of the noise of the audio output signal based on the first spectrum of the noise of the audio output signal, wherein the spectral resolution of the third spectrum of the noise of the audio output signal is higher or the same as the spectral resolution of the first spectrum of the noise of the audio output signal, and wherein the resolution converter comprises a second converter  
20 stage configured to establish the second spectrum of the noise of the audio output signal.

According to a preferred embodiment of the invention the comfort noise spectrum generator is configured to compute the spectrum of the comfort noise  
25 based on the scaling factors and based on the third spectrum of the noise of the audio output signal as provided by the first converter stage of the resolution converter. By these features a comfort noise spectrum may be obtained during inactive phases which has a higher spectral resolution than spectral resolution of the first spectrum of the noise of the audio output signal during  
30 active phases.

According to a preferred embodiment of the invention the comfort noise spectrum generator is configured to compute the spectrum of the comfort noise according to the formula  $\hat{N}^{FR}(k) = \hat{S}^{LR}(i) \cdot \hat{N}_{dec}^{FR}(k)$ , wherein  $\hat{N}^{FR}(k)$  denotes a level of a frequency band  $k$  of the spectrum of the comfort noise, wherein  
5  $\hat{S}^{LR}(i)$  denotes a scaling factor of a frequency band group  $i$  of the spectrum of the background noise as contained in the SID frames and of the second spectrum of the noise of the audio output signal, wherein  $\hat{N}_{dec}^{FR}(k)$  denotes a level of a frequency band  $k$  of the third spectrum of the noise of the audio output signal, wherein  $k = b^{LR}(i), \dots, b^{LR}(i + 1) - 1$ , wherein  $b^{LR}(i)$  is a first  
10 frequency band of a frequency band group, wherein  $i = 0, \dots, L^{LR} - 1$ , wherein  $L^{LR}$  is the number of frequency band groups of the spectrum of the background noise as contained in the SID frames and of the second spectrum of the noise of the audio output signal. By these features the spectrum of the comfort noise may be computed at the high-resolution in an easy way.

15

According to a preferred embodiment of the invention the comfort noise generator comprises a first fast Fourier converter configured to adjust levels of frequency bands of the comfort noise in a fast Fourier transformation domain and a second fast Fourier converter to produce at least a part of the comfort  
20 noise based on an output of the first fast Fourier converter. By these features the background noise can be produced in an easy way.

According to a preferred embodiment of the invention the decoding device comprises a core decoder configured to produce the audio output signal during the active phase. By these features a simple structure of the decoder may  
25 be achieved which is suitable for narrowband (NB) and wideband (WB) applications.

According to a preferred embodiment of the invention the decoding device  
30 comprises a core decoder configured to produce an audio signal and a bandwidth extension module configured to produce the audio output signal

based on the audio signal as provided by the core decoder. By these features a simple structure of the decoder may be achieved which is suitable for super wideband (SWB) applications.

- 5 According to a preferred embodiment of the invention the bandwidth extension module comprises a spectral band replication decoder, a quadrature mirror filter analyzer, and/or a quadrature mirror filter synthesizer.

According to a preferred embodiment of the invention the comfort noise as  
10 provided by the fast Fourier converter is fed to the bandwidth extension module. By this feature the comfort noise as provided by the fast Fourier converter may be transformed into a comfort noise with a higher bandwidth.

According to a preferred embodiment of the invention the comfort noise gen-  
15 erator comprises a quadrature mirror filter adjuster device configured to adjust levels of frequency bands of the comfort noise in a quadrature mirror filter domain, wherein an output of the quadrature mirror filter synthesizer is fed to the bandwidth extension module. By these features noise information  
20 transmitted by the silence insertion descriptor frames related to noise frequencies above the bandwidth of the core decoder may be used to further improve the comfort noise.

In a further aspect the invention relates to a system comprising a decoder  
and an encoder, wherein the decoder is designed according to the invention.  
25

In another aspect the invention relates to a method of decoding an audio bit-  
stream so as to produce therefrom an audio output signal, the bitstream  
comprising at least an active phase followed by at least an inactive phase,  
wherein the bitstream has encoded therein at least a silence insertion de-  
30 scriptor frame which describes a spectrum of a background noise, the method comprising the steps:

decoding the silence insertion descriptor frame so as to reconstruct a spectrum of the background noise;

5 reconstructing the audio output signal from the bitstream during the active phase;

determining a spectrum of the audio output signal;

10 determining a first spectrum of the noise of the audio output signal based on the spectrum of the audio output signal, wherein the first spectrum of the noise of the audio output signal has a higher spectral resolution than the spectrum of the background noise as provided by the silence insertion descriptor decoder;

15 establishing a second spectrum of the noise of the audio output signal based on the first spectrum of the noise of the audio output signal, wherein the second spectrum of the noise of the audio output signal has the same spectral resolution as the spectrum of the background noise as provided by the silence insertion descriptor decoder;

20 computing scaling factors for a spectrum for a comfort noise based on the spectrum of the background noise as provided by the silence insertion descriptor decoder and based on the second spectrum of the noise of the audio output signal; and

25 producing the comfort noise during the inactive phase based on the spectrum for the comfort noise.

30 In a further aspect the invention relates to a computer program for performing, when running on a computer or a processor, the inventive method.

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

- 5 Fig. 1 illustrates a first embodiment of a decoder according to the invention;
- Fig. 2 illustrates a second embodiment of a decoder according to the invention;
- 10 Fig. 3 illustrates a third embodiment of a decoder according to the invention;
- Fig. 4 illustrates a first embodiment of an encoder suitable for an inventive system; and
- 15 Fig. 5 illustrates a second embodiment of an encoder suitable for an inventive system.

Fig. 1 illustrates a first embodiment of a decoder 1 according to the invention.  
20 The audio decoder 1 depicted in Fig. 1 is configured for decoding a bitstream BS so as to produce therefrom an audio output signal OS, the bitstream BS comprising at least an active phase followed by at least an inactive phase, wherein the bitstream BS has encoded therein at least a silence insertion descriptor frame SI which describes a spectrum SBN of a background noise,  
25 the audio decoder 1 comprising:

- a decoding device 2 configured to reconstruct the audio output signal OS from the bitstream BS during the active phase;
- 30 a silence insertion descriptor decoder 3 configured to decode the silence insertion descriptor frame SI so as to reconstruct the spectrum SBN of the background noise;

a spectral converter 4 configured to determine a spectrum SAS of the audio output signal OS;

5 a noise estimator device 5 configured to determine a first spectrum SN1 of the noise of the audio output signal OS based on the spectrum SAS of the audio output signal AS provided by the spectral converter 4, wherein the first spectrum SN1 of the noise of the audio output signal OS has a higher spectral resolution than the spectrum SBN of the background noise;

10

a resolution converter 6 configured to establish a second spectrum SN2 of the noise of the audio output signal OS based on the first spectrum SN1 of the noise of the audio output signal OS, wherein the second spectrum SN2 of the noise of the audio output signal OS has a same spectral resolution as the spectrum SBN of the background noise;

15

a comfort noise spectrum estimation device 7 having a scaling factor computing device 7a configured to compute scaling factors SF for a spectrum SCN for a comfort noise CN based on the spectrum SBN of the background noise as provided by the silence insertion descriptor decoder 3 and based on the second spectrum SN2 of the noise of the audio output signal OS as provided by the resolution converter 6 and having a comfort noise spectrum generator 7b configured to compute the spectrum SCN for a comfort noise CN based on the scaling factors SF ; and

25

a comfort noise generator 8 configured to produce the comfort noise CN during the inactive phase based on the spectrum SCN for the comfort noise CN.

The bitstream BS contains active phases and inactive phases, wherein an active phase is a phase, which contains wanted components of the audio information, such as speech or music, whereas an inactive phase is a phase, which does not contain any wanted components of the audio information.

30

Inactive phases usually occur during pauses, where no wanted components, such as music or speech, are present. Therefore, inactive phases usually contain solely background noise. The information in the bitstream BS containing an encoded audio signal is embedded in so called frames, wherein  
5 each of these frames contain audio information referring to a certain time. During active phases active frames comprising audio information including audio information regarding the wanted signal may be transmitted within the bitstream BS. In contrast of that, during inactive phases silence insertion descriptor frames SI comprising noise information may be transmitted within the  
10 bitstream at a lower average bit-rate compared to the average bit-rate of the active phases.

The decoding device 2 may be a device or a computer program capable of decoding the audio bitstream BS, which is a digital data stream containing  
15 audio information, during active phases. The decoding process may result in a digital decoded audio output signal OS, which may be fed to a D/A converter to produce an analogous audio signal, which then may be fed to a loudspeaker, in order to produce an audible signal.

20 The silence insertion descriptor decoder 3 is configured to decode the silence insertion descriptor frames SI so as to reconstruct a spectrum SBN of the background noise. However, this spectrum SBN of the background noise does not allow to capture the fine spectral structure of the background noise due to a limited number of parameters transmitted in the silence insertion  
25 descriptor frames SI.

The spectral converter 4 may obtain a spectrum SAS of the audio output signal OS which has a significantly higher spectral resolution than the spectrum SBN of the background noise as provided by the silence insertion descriptor  
30 decoder 3.

Therefore, the noise estimator 10 may determine a first spectrum SN1 of the noise of the audio output signal OS based on the spectrum SAS of the audio output signal OS provided by the spectral converter 4, wherein the first spectrum SN1 of the noise of the audio output signal OS has a higher spectral  
5 resolution than the spectrum of the background noise SBN.

Further, the resolution converter 6 may establish a second spectrum SN2 of the noise of the audio output signal OS based on the first spectrum SN1 of the noise of the audio output signal OS, wherein the second spectrum SN2 of  
10 the noise of the audio output signal OS has a same spectral resolution as the spectrum of the background noise SBN.

The scaling factor computing device 7a may easily compute scaling factors SF for a spectrum SCN for a comfort noise CN based on the spectrum SBN  
15 of the background noise as provided by the silence insertion descriptor decoder 3 and based on the second spectrum SN2 of the noise of the audio output signal OS as provided by the resolution converter 6 as the spectrum SBN of the background noise and the second spectrum SN2 of the noise of the audio output signal OS have the same spectral resolution.

20 The comfort noise spectrum generator 7b may establish the spectrum SCN for the comfort noise CN based on the scaling factors SF.

Furthermore, the comfort noise generator 8 may produce the comfort noise  
25 CN during the inactive phase based on the spectrum SCN for the comfort noise.

The noise estimates obtained at the decoder 1 contain information about the spectral structure of the background noise, which is more accurate than the  
30 information about the spectral structure of the background noise contained in the SID frames SI. However, these estimates cannot be adapted during inactive phases since the noise estimation is carried out on the decoded audio



output signal OS. In contrast, the SID frames deliver new information about the spectral envelope at regular intervals during inactive phases. The decoder 1 according to the invention combines these two sources of information. The scaling factors SF may be updated during active phases depending on the noise estimates at the decoder side and during inactive phases depending on the noise estimates contained in the SID frames SI. The continuous update of the scaling factors SF ensures that there are no sudden changes of the characteristics of the produced comfort noise CN.

10 As the spectrum SBN of the background noise as contained in the SID frames SI and the second spectrum SN2 of the noise of the audio output signal OS have the same spectral resolution the update of the scaling factors SF and, hence, of the comfort noise CN can be done in an easy way, as for each frequency band group of the spectrum SBN of the background noise as contained in the SID frames SI exactly one frequency band group exists in the second spectrum SN2 of the noise of the audio output signal OS. It has to be noted that in a preferred embodiment the frequency band groups of the spectrum of the background noise as contained in the SID frames SI and the frequency band groups of the second spectrum SN2 of the noise of the audio output signal OS correspond to each other.

Further, as the spectrum SBN of the background noise as contained in the SID frames SI and the second spectrum SN2 of the noise of the audio output signal OS have the same spectral resolution the update of the scaling factors SF produces no or only barely audible artifacts.

According to a preferred embodiment of the invention the spectral analyzer 4 comprises a fast Fourier transformation device. A fast Fourier transform (FFT) is an algorithm to compute a discrete Fourier transform (DFT) and its inverse, which requires only low computational effort. Therefore, the fast Fourier transformation device may calculate the spectrum SAS of the audio output signal OS in an easy way.

According to a preferred embodiment of the invention the noise estimator device 5 comprises a converter device 9 configured to convert the spectrum SAS of the audio output signal OS into a converted spectrum CSA of the audio output signal OS which has the same spectral resolution as the core decoder 17. In general the spectral resolution of the spectrum SAS of the audio output signal OS obtained by a spectral converter 4 is much higher than the spectral resolution of the core decoder 17. By providing the converted spectrum CSA of the audio output signal OS the complexity of subsequent computational steps may be reduced.

According to a preferred embodiment of the invention the noise estimator device 5 comprises a noise estimator 10 configured to determine the first spectrum SN1 of the noise of the audio output signal OS based on the converted spectrum CAS of the audio output signal OS provided by the converter device 9. When the converted spectrum CSA of the audio output signal OS is used as a basis for the noise estimation at the decoder computational efforts may be reduced without lowering the quality of the noise estimation.

According to a preferred embodiment of the invention the scaling factor computing device 7a is configured to compute the scaling factors SF according to the formula

$$\hat{S}^{LR}(i) = \frac{\hat{N}_{SID}^{LR}(i)}{\hat{N}_{dec}^{LR}(i)},$$

wherein  $\hat{S}^{FR}(i)$  denotes a scaling factor SF for a frequency band group  $i$  of the comfort noise CN, wherein  $\hat{N}_{SID}^{LR}(i)$  denotes a level of a frequency band group  $i$  of the spectrum SBN of the background noise, wherein  $\hat{N}_{dec}^{LR}(i)$  denotes a level of a frequency band group  $i$  of the second spectrum SN2 of the noise of the audio output signal, wherein  $i = 0, \dots, L^{LR} - 1$ , wherein  $L^{LR}$  is the number of frequency band groups of the spectrum SBN of the background noise and of the second spectrum SN2 of the noise of the audio output signal OS. By these features the scaling factors SF may be computed in an easy manner.

According to a preferred embodiment of the invention the comfort noise spectrum generator 7b is configured to compute the spectrum SCN of the comfort noise CN based on the scaling factors SF and based on the first spectrum SN1 of the noise of the audio output signal OS as provided by the noise estimation device 5. By these features the comfort noise spectrum SCN may be computed in such way that it has the spectral resolution of the first spectrum SN1 of the noise of the audio output signal OS.

According to a preferred embodiment of the invention the comfort noise spectrum generator 7b is configured to compute the spectrum SCN of the comfort noise CN according to the formula  $\widehat{N}^{FR}(k) = \widehat{S}^{LR}(i) \cdot \widehat{N}_{dec}^{HR}(k)$ , wherein  $\widehat{N}^{FR}(k)$  denotes a level of a frequency band  $k$  of the spectrum SCN of the comfort noise CN, wherein  $\widehat{S}^{LR}(i)$  denotes a scaling factor SF of a frequency band group  $i$  of the spectrum SBN of the background noise and of the second spectrum SN2 of the noise of the audio output signal OS, wherein  $\widehat{N}_{dec}^{HR}(k)$  denotes a level of a frequency band  $k$  of the first spectrum SN1 of the noise of the audio output signal OS, wherein  $k = b^{LR}(i), \dots, b^{LR}(i + 1) - 1$ , wherein  $b^{LR}(i)$  is a first frequency band of one of the frequency band groups, in  $i = 0, \dots, L^{LR} - 1$ , wherein  $L^{LR}$  is the number of frequency band groups of the spectrum SBN of the background noise and of the second spectrum SN2 of the noise of the audio output signal. By these features the spectrum SCN of the comfort noise CN may be computed at a high-resolution in an easy way.

According to a preferred embodiment of the invention the resolution converter 6 comprises a first converter stage 11 configured to establish a third spectrum SN3 of the noise of the audio output signal OS based on the first spectrum SN1 of the noise of the audio output signal OS, wherein the spectral resolution of the third spectrum SN3 of the noise of the audio output signal OS is same or higher as the spectral resolution of the first spectrum SN1 of

the noise of the audio output signal OS, and wherein the resolution converter 6 comprises a second converter stage 12 configured to establish the second spectrum SN2 of the noise of the audio output signal OS.

- 5 According to a preferred embodiment of the invention the comfort noise spectrum generator 7b is configured to compute the spectrum SCN of the comfort noise CN based on the scaling factors SF and based on the third spectrum SN3 of the noise of the audio output signal OS as provided by the first converter stage 11 of the resolution converter 6. By these features a comfort  
10 noise spectrum SCN may be obtained which has a higher spectral resolution than the background noise spectrum SBN provided by the silence insertion descriptor decoder 3.

According to a preferred embodiment of the invention the comfort noise spectrum generator 7b is configured to compute the spectrum SCN of the comfort  
15 noise according to the formula  $\hat{N}^{FR}(k) = \hat{S}^{LR}(i) \cdot \hat{N}_{dec}^{FR}(k)$ , wherein  $\hat{N}^{FR}(k)$  denotes a level of a frequency band  $k$  of the spectrum SCN of the comfort noise CN, wherein  $\hat{S}^{LR}(i)$  denotes a scaling factor SF of a frequency band group  $i$  of the spectrum SCN of the background noise and of the second  
20 spectrum SN2 of the noise of the audio output signal OS, wherein  $\hat{N}_{dec}^{FR}(k)$  denotes a level of a frequency band  $k$  of the third spectrum SN3 of the noise of the audio output signal OS, wherein  $k = b^{LR}(i), \dots, b^{LR}(i + 1) - 1$ , wherein  $b^{LR}(i)$  is a first frequency band of a frequency band group, wherein  $i = 0, \dots, L^{LR} - 1$ , wherein  $L^{LR}$  is the number of frequency band groups of the  
25 spectrum SBN of the background noise and of the second spectrum SN2 of the noise of the audio output signal OS. By these features the spectrum SCN of the comfort noise may be computed at the high-resolution in an easy way.

- 30 According to a preferred embodiment of the invention the comfort noise generator 8 comprises a first fast Fourier converter 15 configured to adjust levels

of frequency bands of the comfort noise CN in a fast Fourier transformation domain and a second fast Fourier converter 16 to produce at least a part of the comfort noise CN based on an output of the first fast Fourier converter 15. By these features the comfort noise can be produced in an easy way.

5

According to a preferred embodiment of the invention the decoding device 2 comprises a core decoder 17 configured to produce the audio output signal OS during the active phase. By these features a simple structure of the decoder may be achieved which is suitable for narrowband (NB) and wideband (WB) applications.

10

According to the preferred embodiment of the invention the audio decoder 1 comprises a header reading device 18, which is configured to discriminate between active phases and inactive phase. The header reading device 18 is further configured to switch a switch device 19 in such way that the bitstream BS during active phases is fed to the core decoder 17 and that the silence insertion descriptor frames during the inactive phases are fed to the silence insertion descriptor decoder 3. Additionally, an inactive phase flag is transmitted to the background noise generator 8 so that the generation of the comfort noise CN may be triggered.

15  
20

Fig. 2 illustrates a second embodiment of an audio decoder 1 according to the invention. The decoder 1 depicted in Fig. 2 is based on the decoder 1 of Fig. 1. In the following only the differences will be explained. The audio decoder 1 of a second embodiment of the invention comprises a bandwidth extension module 20 to which the output signal of the core decoder 17 is fed. The bandwidth extension module 20 is configured to produce a bandwidth extended output signal EOS based on the audio output signal OS. By these features a simple structure of the decoder 1 may be achieved which is suitable for super wideband (SWB) applications.

25

30

According to a preferred embodiment of the invention the comfort noise CN as provided by the fast Fourier converter 16 is fed to the bandwidth extension module 20. By this feature the comfort noise CN as provided by the fast Fourier converter 16 may be transformed into a comfort noise CN with a higher  
5 bandwidth.

According to a preferred embodiment of the invention the comfort noise generator 8 comprises a quadrature mirror filter adjuster device 24 configured to adjust levels of frequency bands of the comfort noise CN in a quadrature mirror filter domain, wherein an output of the quadrature mirror filter synthesizer  
10 24 is fed to the bandwidth extension module 20 as an additional comfort noise CN'. QMF levels contained in the silence insertion descriptor frames SI may be fed to the quadrature mirror filter synthesizer device 24. By these features noise information transmitted by the silence insertion descriptor  
15 frames SI related to noise frequencies above the bandwidth of the core decoder 17 may be used to further improve the comfort noise CN.

According to a preferred embodiment of the invention the bandwidth extension module 20 comprises a spectral band replication decoder 21, a quadrature mirror filter analyzer 22, and/or a quadrature mirror filter synthesizer 23.  
20

Fig. 3 illustrates a third embodiment of a decoder 1 according to the invention. The decoder 1 of Fig. 3 is based on the decoder 1 of Fig. 2. The following only the differences to be discussed.  
25

According to a preferred embodiment of the invention the decoding device 2 comprises a core decoder 17 configured to produce an audio signal AS and a bandwidth extension module 20 configured to produce the audio output signal OS based on the audio signal AS as provided by the core decoder 17. By  
30 these features a simple structure of the decoder may be achieved which is suitable for super wideband (SWB) applications.

In principle the bandwidth extension module 20 of Fig. 3 is the same as the bandwidth extension module 20 of Fig. 2. However, in the third embodiment of the audio decoder 1 according to the invention the bandwidth extension module 20 is used to produce the audio output signal OS, which is fed to the spectral converter 4. By these features the entire bandwidth can be used for producing comfort noise.

Regarding the three embodiments of the audio decoder according to the invention it may be added: At the decoder side, a random generator 8 may be applied to excite each individual spectral band in the FFT domain, as well as in the QMF domain for SWB modes. The amplitude of the random sequences should be individually computed in each band such that the spectrum of the generated comfort noise CN resembles the spectrum of the actual background noise present in the bitstream.

The high-resolution noise estimates obtained at the decoder 1 capture information about the fine spectral structure of the background noise. However, these estimates cannot be adapted during inactive phases since the noise estimation is carried out on the decoded signal OS. In contrast, the SID frames SI deliver new information about the spectral envelope at regular intervals during inactive phases. The present decoder 1 combines these two sources of information in an effort to reproduce the fine spectral structure captured from the background noise present during active phases, while updating only the spectral envelope of the comfort noise CN during inactive parts with the help of the SID information.

To achieve this goal, an additional noise estimator 5 is used in the decoder 1, as shown in Figs. 1 to 3. Hence, noise estimation is carried out at both sides of the transmission system, but applying a higher spectral resolution at the decoder 1 than at the encoder 100. One way to obtain a high spectral resolution at the decoder 1 is to simply consider each spectral band individually (full resolution) instead of grouping them via averaging like in the encoder 100.

Alternatively, a trade-off between spectral resolution and computational complexity can be obtained by carrying out the spectral grouping also in the decoder 1 but using an increased number of spectral groups compared to the encoder 100, yielding thereby a finer quantization of the frequency axis in the decoder.

Note that the decoder-side noise estimation operates on the decoded signal OS. In a DTX-based system, it should be therefore capable of operating during active phases only, i.e., necessarily on clean speech or noisy speech contents (in contrast to noise only).

The high-resolution (HR) noise power spectrum  $\widehat{N}_{\text{dec}}^{\text{HR}}$  computed at the decoder may be first interpolated (e.g., using linear interpolation) to provide a full-resolution (FR) power spectrum  $\widehat{N}_{\text{dec}}^{\text{FR}}$ . It may then be converted to a low-resolution (LR) power spectrum  $\widehat{N}_{\text{dec}}^{\text{LR}}$  by spectral grouping (i.e., averaging) just as done in the encoder. The power spectrum  $\widehat{N}_{\text{dec}}^{\text{LR}}$  exhibits therefore the same spectral resolution as the noise levels  $\widehat{N}_{\text{SID}}^{\text{LR}}$  gained from the SID frames SI. Comparing the low-resolution noise spectra  $\widehat{N}_{\text{dec}}^{\text{LR}}$  and  $\widehat{N}_{\text{SID}}^{\text{LR}}$ , the full-resolution noise spectrum  $\widehat{N}_{\text{dec}}^{\text{FR}}$  can be finally scaled to yield a full-resolution power spectrum as follows:

$$\widehat{N}^{\text{FR}}(k) = \frac{\widehat{N}_{\text{SID}}^{\text{LR}}(i)}{\widehat{N}_{\text{dec}}^{\text{LR}}(i)} \cdot \widehat{N}_{\text{dec}}^{\text{FR}}(k) \quad k = b^{\text{LR}}(i), \dots, b^{\text{LR}}(i+1) - 1,$$

$$i = 0, \dots, L^{\text{LR}} - 1,$$

where  $L^{\text{LR}}$  is the number of spectral groups used by the low-resolution noise estimation in the encoder, and  $b^{\text{LR}}(i)$  denotes the first spectral band of the  $i$ th spectral group,  $i = 0, \dots, L^{\text{LR}} - 1$ . The full-resolution noise power spectrum  $\widehat{N}^{\text{FR}}(k)$  can finally be used to accurately adjust the level of comfort noise generated in each individual FFT or QMF band (the latter for SWB modes only).



In Figs. 1 and 2, the above mechanism is applied to the FFT coefficients only. Hence, for SWB systems, it is not applied in the QMF bands capturing the high-frequency content left over by the core. Since these frequencies are perceptually less relevant, reproducing the smooth spectral envelope of the noise for these frequencies is sufficient in general.

To adjust the level of comfort noise applied in the QMF domain for frequencies which are above the core bandwidth in SWB modes, the system relies solely on the information transmitted by the SID frames. The SBR module is thus bypassed when the VAD triggers a CNG frame. In WB modes, the CNG module does not take the QMF bands into account since blind bandwidth extension is applied to recover the desired bandwidth.

Nevertheless, the scheme can be easily extended to cover the entire bandwidth by applying the decoder-side noise estimator at the output of the bandwidth extension module instead of applying it at the output of the core decoder. This extension as shown in Fig. 3 causes an increase in computational complexity since the high frequencies captured by the QMF filterbank have to be considered as well.

Fig. 4 illustrates a first embodiment of an encoder 100 suitable for an inventive system. The input audio signal IS is fed to a first spectral converter 25 configured to transfer that time domain signal IS into a frequency domain. The first spectral converter 25 may be a quadrature mirror filter analyzer. The output of the first spectral converter 25 is fed to a second spectral converter 26 which is configured to transfer the output of the first spectral converter 25 to a domain. The second spectral converter 26 may be a quadrature mirror filter synthesizer. The output of the second spectral converter 26 is fed to a third spectral converter 27 which may be a fast Fourier transforming device. The output of the third spectral converter 27 is fed to a noise estimator device 28 which consists of a convert device 29 and a noise estimator 30.

Further, the encoder 100 comprises a signal activity detector 31 which is configured to switch the switch device 32 in such way that during active phases input signal is fed to a core encoder 33 and that in SID frames during inactive phases a noise estimation created by the noise estimating device 28 is fed to a silence insertion descriptor encoder 35. Further, in inactive phases an inactivity flag is fed to a core updater 34.

The encoder 100 further comprises a bitstream producer 36 which receives silence insertion descriptor frames SI from the silence insertion descriptor encoder 35 and an encoded input signal ISE from the core encoder 33 in order to produce the bitstream BS therefrom.

Fig. 5 illustrates a second embodiment of an encoder 100 suitable for an inventive system which is based on the encoder 100 of first embodiment. The additional features of a second embodiment will briefly be explained in the following. The output of the first converter 25 is also fed to the noise estimator device 28. Further, during active phases, a spectral band replication encoder 37 produces an enhancement signal ES which contains information about higher frequencies in the input audio signal IS. That enhancement signal 37 is also transferred to the bitstream producer 36 so as to embed that enhancement signal ES into the bitstream BS.

Regarding the encoders shown in Figs. 4 and 5 following information may be added: In case the VAD triggers a CNG phase, SID frames containing information about the input background noise are transmitted. This should allow the decoder to generate an artificial noise resembling the actual background noise in terms of spectro-temporal characteristics. To this aim, a noise estimator 28 is applied at the encoder side to track the spectral shape of the background noise present in the input signal IS, as shown in Figs. 4 and 5

30

In principle, noise estimation can be applied with any spectro-temporal analysis tool decomposing a time-domain signal into multiple spectral bands, as

long as it offers sufficient spectral resolution. In the present system, a QMF filterbank is used as a resampling tool to downsample the input signal to the core sampling rate. It exhibits a significantly lower spectral resolution than the FFT which is applied to the downsampled core signal.

5

Since the core encoder 33 already covers the entire NB bandwidth and since WB modes rely on blind bandwidth extension, the frequencies above the core bandwidth are irrelevant and can be simply discarded for NB and WB systems. In SWB modes, in contrast, those frequencies are captured by the up-  
per QMF bands and need to be taken into account explicitly.

10

The size of an SID frame SI is very limited in practice. Therefore, the number of parameters describing the background noise has to be kept as small as possible. To this aim, the noise estimation is not applied directly in the output of the spectral transforms. Instead, it is applied at a lower spectral resolution by averaging the input power spectrum among groups of bands, e.g., following the Bark scale. The averaging can be achieved either by arithmetic or geometric means. In the SWB case, the spectral grouping is carried out for the FFT and QMF domains separately, whereas the NB and WB modes rely  
on the FFT domain only.

15

20

Note that reducing the spectral resolution is also advantageous in terms of computational complexity since the noise estimation needs to be applied to only a small number of spectral groups instead of considering each spectral  
band individually.

25

The estimated noise levels (one for each spectral group) can be jointly encoded in SID frames using vector quantization techniques. In NB and WB modes, only the FFT domain is exploited. In contrast, for SWB modes, the  
encoding of SID frames can be performed for both FFT and QMF domains jointly using vector quantization, i.e., resorting to a single codebook covering  
both domains.

30

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

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

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

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

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

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

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

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

20

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

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

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

memory device or the like. The apparatus or system may, for example, comprise a file server for transferring the computer program to the receiver .

5 In some embodiments, a programmable logic device (for example, a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

10

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

Reference signs:

- 20 1 audio decoder  
2 decoding device  
3 silence insertion descriptor decoder  
4 spectral converter  
5 noise estimator device  
25 6 resolution converter  
7 comfort noise spectrum estimation device  
7a scaling factor computing device  
7b comfort noise spectrum generator  
8 comfort noise generator  
30 9 converter device  
10 noise estimator  
11 first converter stage

- 12 second converter stage  
15 first fast Fourier converter  
16 second fast Fourier analyzer  
17 core decoder  
5 18 header reading device  
19 switch device  
20 bandwidth extension module  
21 spectral band replication decoder  
22 quadrature mirror filter analyzer  
10 23 quadrature mirror filter synthesizer  
24 quadrature mirror filter adjuster device  
25 first spectral converter  
26 second spectral converter  
27 third spectral converter  
15 28 noise estimator device  
29 converter device  
30 noise estimator  
31 signal activity detector  
32 switch device  
20 33 core encoder  
34 core updater  
35 silence insertion descriptor encoder  
36 bitstream producer  
37 spectral band replication encoder  
25 100 encoder
- BS bitstream  
OS audio output signal  
SI silence insertion descriptor frame  
30 SBN spectrum of the background noise  
SAS spectrum of the audio signal  
SN1 first spectrum of the noise of the audio signal

- SN2 second spectrum of the noise of the audio signal
- SF scaling factors
- SCN spectrum of the comfort noise
- CN comfort noise
- 5 AS output signal
- CSA converted spectrum of the audio signal
- SN3 third spectrum of the noise of the audio signal
- EOS bandwidth extended output signal
- IS input audio signal
- 10 ISE encoded input signal
- ES enhancement signal



Claims

1. Audio decoder for decoding a bitstream (BS) so as to produce there-  
from an audio output signal (OS), the bitstream (BS) comprising at least  
5 an active phase followed by at least an inactive phase, wherein the bit-  
stream (BS) has encoded therein at least a silence insertion descriptor  
frame (SI) which describes a spectrum of a background noise (SBN),  
the audio decoder (1) comprising:
- 10 a silence insertion descriptor decoder (3) configured to decode the si-  
lence insertion descriptor frame (SI) so as to reconstruct the spectrum  
(SBN) of the background noise;
- a decoding device (2) configured to reconstruct the audio output signal  
15 (OS) from the bitstream during the active phase;
- a spectral converter (4) configured to determine a spectrum (SAS) of  
the audio output signal (OS);
- 20 a noise estimator device (5) configured to determine a first spectrum  
(SN1) of the noise of the audio output signal (OS) based on the spec-  
trum (SAS) of the audio output signal (OS) provided by the spectral  
converter (4), wherein the first spectrum (SN1) of the noise of the audio  
output signal (OS) has a higher spectral resolution than the spectrum  
25 (SBN) of the background noise;
- a resolution converter (6) configured to establish a second spectrum  
(SN2) of the noise of the audio output signal (OS) based on the first  
spectrum (SN1) of the noise of the audio output signal (OS), wherein  
30 the second spectrum (SN2) of the noise of the audio output signal (OS)  
has a same spectral resolution as the spectrum (SBN) of the back-  
ground noise;

5 a comfort noise spectrum estimation device (7) having a scaling factor  
computing device (7a) configured to compute scaling factors (SF) for a  
spectrum (SCN) for a comfort noise (CN) based on the spectrum (SBN)  
of the background noise as provided by the silence insertion descriptor  
decoder (3) and based on the second spectrum (SN2) of the noise of  
the audio output signal (OS) as provided by the resolution converter (6)  
and having a comfort noise spectrum generator (7b) configured to com-  
pute the spectrum (SCN) for a comfort noise (CN) based on the scaling  
10 factors (SF); and

a comfort noise generator (8) configured to produce the comfort noise  
(CN) during the inactive phase based on the spectrum (SCN) for the  
comfort noise (CN).

15

2. Audio decoder according to the preceding claim, wherein the spectral  
analyzer (4) comprises a fast Fourier transformation device (4).

3. Audio decoder according to one of the preceding claims, wherein the  
20 noise estimator device (5) comprises a converter device (9) configured  
to convert the spectrum (SAS) of the audio output signal (OS) into a  
converted spectrum (CSA) of the audio output signal (OS) which has  
same or lower spectral resolution than the spectrum (SAS) of the output  
audio signal and a higher spectral resolution than the spectrum (SBN)  
25 of the background noise.

4. Audio decoder according to the preceding claim, wherein the noise es-  
timator device (5) comprises a noise estimator (10) configured to de-  
termine the first spectrum (SN1) of the noise of the audio output signal  
30 (OS) based on the converted spectrum (CSA) of the audio output signal  
(OS) provided by the converter device (9).

5. Audio decoder according to one of the preceding claims, wherein the scaling factor computing device (7a) is configured to compute the scaling factors (SF) according to the formula
- $$\hat{S}^{LR}(i) = \frac{\hat{N}_{SID}^{LR}(i)}{\hat{N}_{dec}^{LR}(i)},$$
- wherein  $\hat{S}^{FR}(i)$  denotes a scaling factor (SF) for a frequency band group  $i$  of the comfort noise (CN), wherein  $\hat{N}_{SID}^{LR}(i)$  denotes a level of a frequency band group  $i$  of the spectrum (SBN) of the background noise, wherein  $\hat{N}_{dec}^{LR}(i)$  denotes a level of a frequency band group  $i$  of the second spectrum (SN2) of the noise of the audio output signal (OS), wherein  $i = 0, \dots, L^{LR} - 1$ , wherein  $L^{LR}$  is the number of frequency band groups of the spectrum (SBN) of the background noise and of the second spectrum (SN2) of the noise of the audio output signal (OS).
6. Audio decoder according to one of the preceding claims, wherein the comfort noise spectrum generator (7b) is configured to compute the spectrum of the comfort noise (SCN) based on the scaling factors (SF) and based on the first spectrum (SN1) of the noise of the audio output signal (OS) as provided by the noise estimation device (5).
7. Audio decoder according to one of the preceding claims, wherein the comfort noise spectrum generator (7b) is configured to compute the spectrum (SCN) of the comfort noise according to the formula  $\hat{N}^{FR}(k) = \hat{S}^{LR}(i) \cdot \hat{N}_{dec}^{HR}(k)$ , wherein  $\hat{N}^{FR}(k)$  denotes a level of a frequency band  $k$  of the spectrum of the comfort noise (SCN), wherein  $\hat{S}^{LR}(i)$  denotes a scaling factor (SF) of a frequency band group  $i$  of the spectrum (SBN) of the background noise and of the second spectrum (SN2) of the noise of the audio output signal, wherein  $\hat{N}_{dec}^{HR}(k)$  denotes a level of a frequency band  $k$  of the first spectrum (SN1) of the noise of the audio output signal (OS), wherein  $k = b^{LR}(i), \dots, b^{LR}(i + 1) - 1$ , wherein  $b^{LR}(i)$  is a first frequency band of one of the frequency band groups, wherein  $i =$

in  $i = 0, \dots, L^{LR} - 1$ , wherein  $L^{LR}$  is the number of frequency band groups of the spectrum (SBN) of the background noise and of the second spectrum (SN2) of the noise of the audio output signal (OS).

- 5 8. Audio decoder according to one of the preceding claims, wherein the resolution converter (6) comprises a first converter stage (11) configured to establish a third spectrum (SN3) of the noise of the audio output signal (OS) based on the first spectrum (SN1) of the noise of the audio output signal (OS), wherein the spectral resolution of the third spectrum
- 10 (SN3) of the noise of the audio output signal (OS) is same or higher as the spectral resolution of the first spectrum (SN1) of the noise of the audio output signal (OS), and wherein the resolution converter (6) comprises a second converter stage (12) configured to establish the second spectrum (SN2) of the noise of the audio output signal (OS).
- 15
9. Audio decoder according to the preceding claim, wherein the comfort noise spectrum generator (7b) is configured to compute the spectrum of the comfort noise (SCN) based on the scaling factors (SF) and based on the third spectrum (SN3) of the noise of the audio output signal (OS)
- 20 as provided by the first converter stage (11) of the resolution converter (6).
10. Audio decoder according to claim eight or nine, wherein the comfort noise spectrum generator (7b) is configured to compute the spectrum
- 25 (SCN) of the comfort noise according to the formula  $\hat{N}^{FR}(k) = \hat{S}^{LR}(i) \cdot \hat{N}_{dec}^{FR}(k)$ , wherein  $\hat{N}^{FR}(k)$  denotes a level of a frequency band  $k$  of the spectrum of the comfort noise (SCN), wherein  $\hat{S}^{LR}(i)$  denotes a scaling factor (SF) of a frequency band group  $i$  of the spectrum (SBN) of the background noise and of the second spectrum (SN2) of the noise of the audio output signal, wherein  $\hat{N}_{dec}^{FR}(k)$  denotes a level of a frequency band  $k$  of the third spectrum (SN3) of the noise of the audio
- 30

output signal (OS), wherein  $k = b^{LR}(i), \dots, b^{LR}(i + 1) - 1$ , wherein  $b^{LR}(i)$  is a first frequency band of a frequency band group, wherein  $i =$   
in  $i = 0, \dots, L^{LR} - 1$ , wherein  $L^{LR}$  is the number of frequency band groups  
of the spectrum (SBN) of the background noise and of the second spec-  
trum (SN2) of the noise of the audio output signal (OS).

- 5 11. Audio decoder according to one of the preceding claims, wherein the comfort noise generator (8) comprises a first fast Fourier converter (15) configured to adjust levels of frequency bands of the comfort noise (CN) in a fast Fourier transformation domain and a second fast Fourier converter (16) to produce at least a part of the comfort noise based on an output of the first fast Fourier converter (15).
- 10 12. Audio decoder according to one of the preceding claims, wherein the decoding device (2) comprises a core decoder (17) configured to produce the audio output signal (OS) during the active phase.
- 15 13. Audio decoder according to one of the claims 1 to 11, wherein the decoding device (2) comprises a core decoder (17) configured to produce an audio signal (AS) and a bandwidth extension module (20) configured to produce the audio output signal (OS) based on the audio signal (AS) as provided by the core decoder (17).
- 20 14. Audio decoder according to the preceding claim, wherein the bandwidth extension module (20) comprises a spectral band replication decoder (21), a quadrature mirror filter analyzer (22), and/or a quadrature mirror filter synthesizer (23).
- 25 15. Audio decoder according to claim 13 or 14, wherein the comfort noise (CN) as provided by the fast Fourier synthesizer (15) is fed to the bandwidth extension module (17).
- 30

16. Audio decoder according to one of the claims 13 to 15, wherein the comfort noise generator (8) comprises a quadrature mirror filter adjuster device (24) configured to adjust levels of frequency bands of the comfort noise (CN) in a quadrature mirror filter domain, wherein an output of the quadrature mirror filter synthesizer (24) is fed to the bandwidth extension module (20).

17. A system comprising a decoder (1) and an encoder (100), wherein the decoder (1) is designed according to one of the claims 1 to 16.

18. A method of decoding an audio bitstream (BS) so as to produce therefrom an audio output signal (OS), the bitstream (BS) comprising at least an active phase followed by at least an inactive phase, wherein the bitstream (BS) has encoded therein at least a silence insertion descriptor frame (SI) which describes a spectrum of a background noise (SBN), the method comprising the steps:

decoding the silence insertion descriptor frame (SI) so as to reconstruct the spectrum (SBN) of the background noise;

reconstructing the audio output signal (OS) from the bitstream during the active phase;

determining a spectrum (SAS) of the audio output signal (OS);

determining a first spectrum (SN1) of the noise of the audio output signal (OS) based on the spectrum (SAS) of the audio output signal (OS), wherein the first spectrum (SN1) of the noise of the audio output signal (OS) has a higher spectral resolution than the spectrum (SBN) of the background noise;

establishing a second spectrum (SN2) of the noise of the audio output

signal (OS) based on the first spectrum (SN1) of the noise of the audio output signal (OS), wherein the second spectrum (SN2) of the noise of the audio output signal (OS) has a same spectral resolution as the spectrum (SBN) of the background noise;

5

computing scaling factors for a spectrum (SCN) for a comfort noise (CN) based on the spectrum (SBN) of the background noise and based on the second spectrum (SN2) of the noise of the audio output signal (OS); and

10

producing the comfort noise (CN) during the inactive phase based on the spectrum (SCN) for the comfort noise (CN).

19. Computer program for performing, when running on a computer or a processor, the method of claim 18.

15

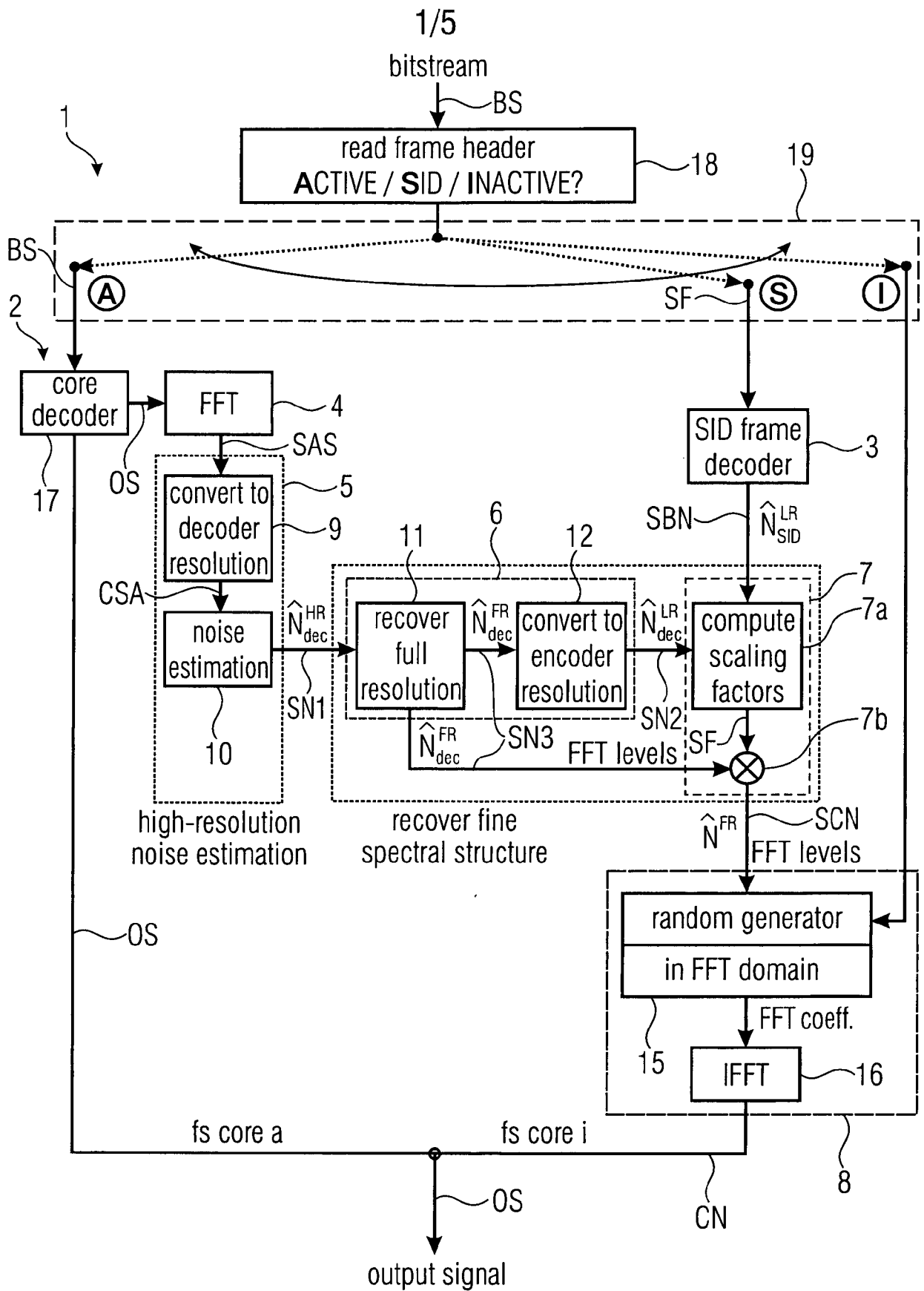


FIG 1



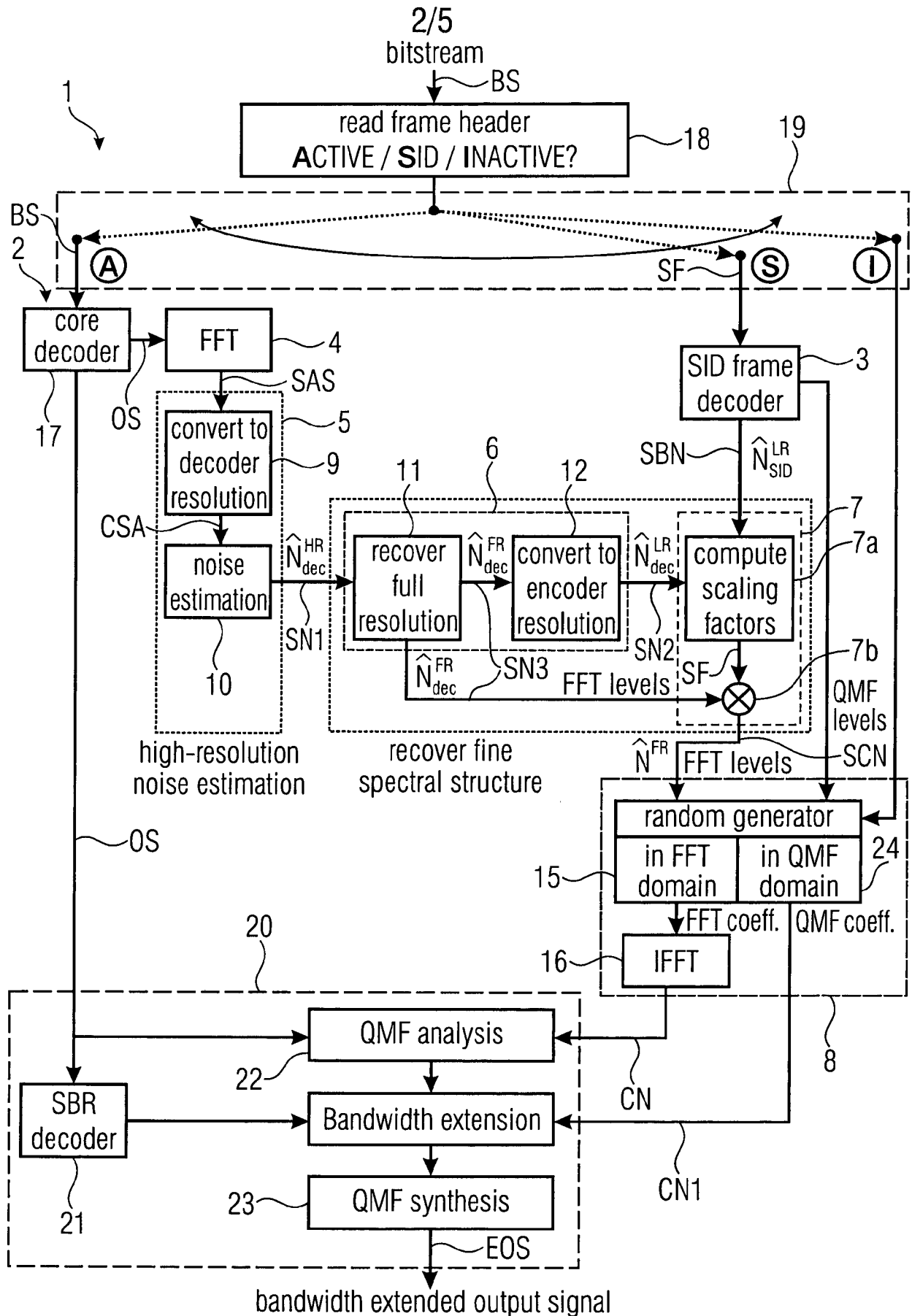


FIG 2

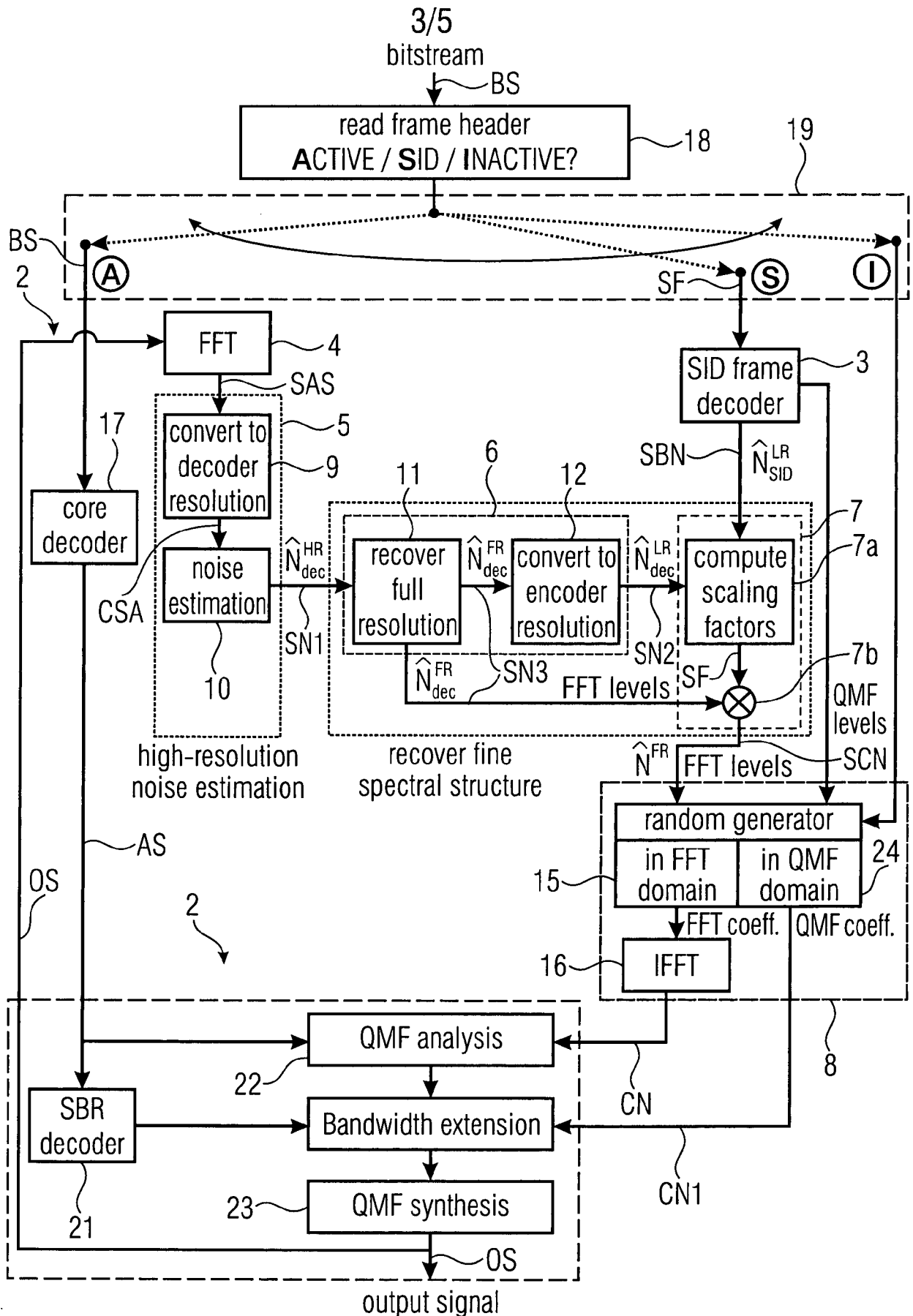


FIG 3

4/5

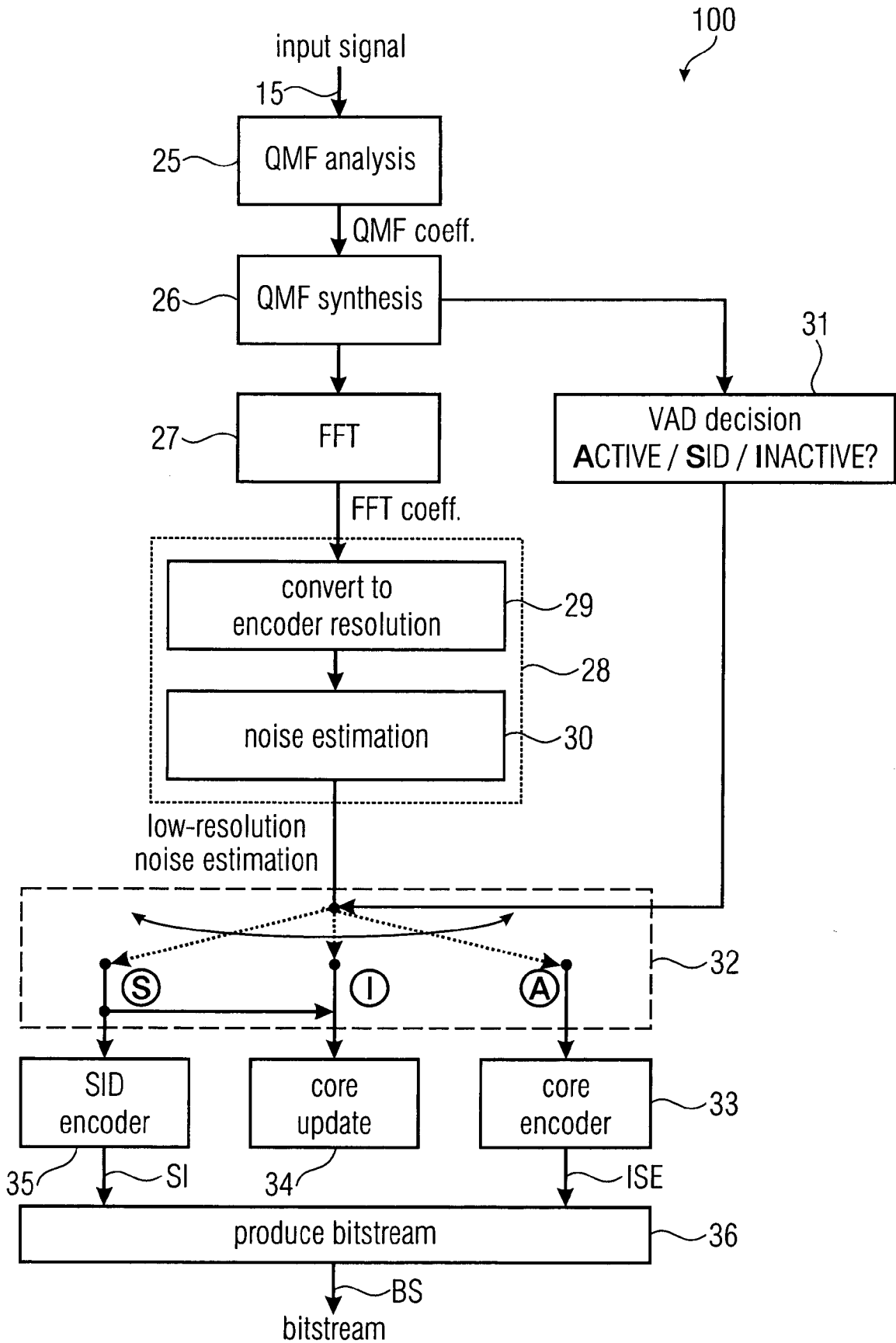


FIG 4

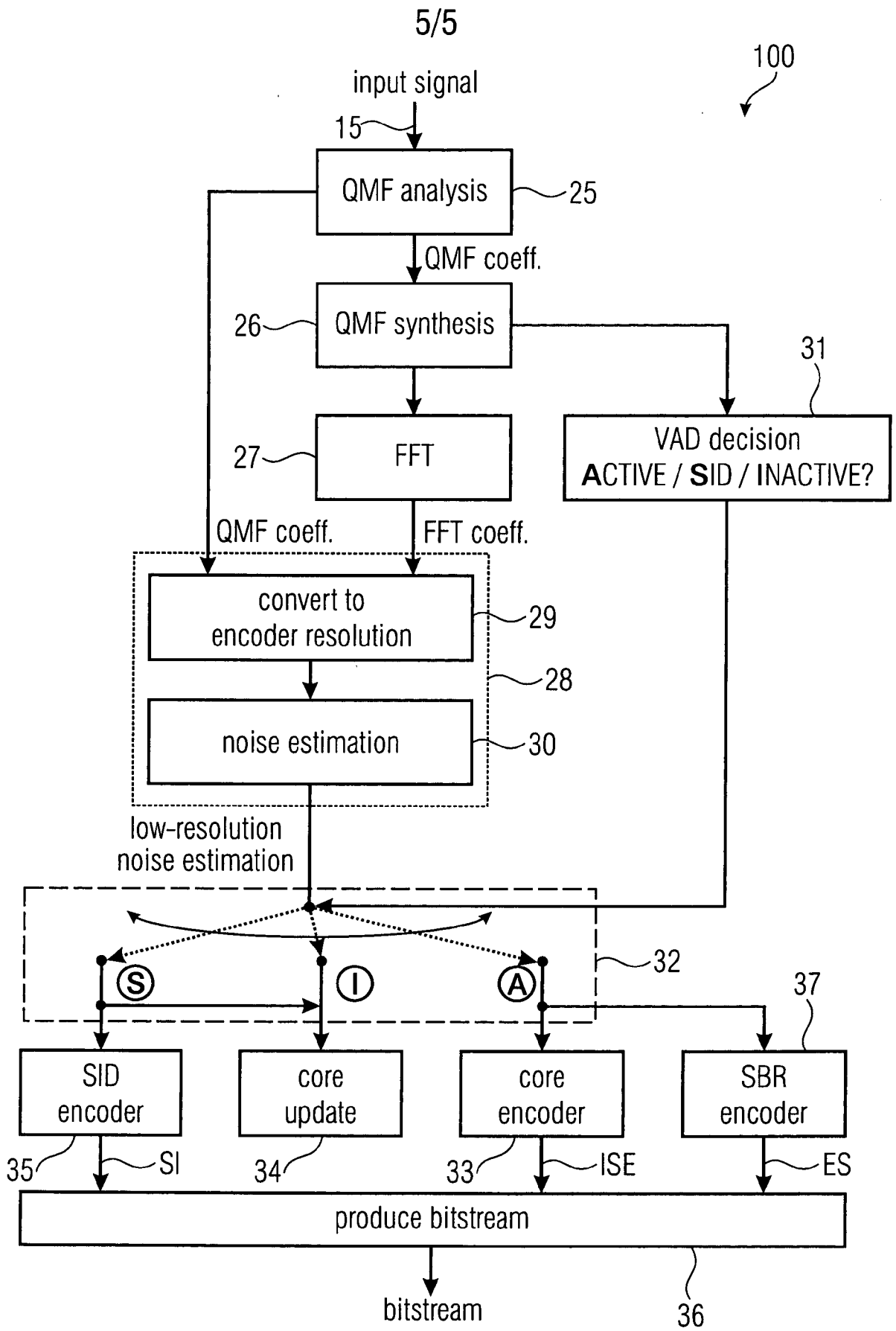


FIG 5

INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2013/077525

A. CLASSIFICATION OF SUBJECT MATTER  
INV. G10L19/012  
ADD.  
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED  
Minimum documentation searched (classification system followed by classification symbols)  
G10L  
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2012/110482 A2 (FRAUNHOFER GES FORSCHUNG [DE]; SETIAWAN PANJI [DE]; WILDE STEPHAN [DE]) 23 August 2012 (2012-08-23) abstract page 1, lines 6-20 page 28, lines 4-23 figure 13	1-19
A	US 6 873 604 B1 (SURAZSKI LUKE [US] ET AL) 29 March 2005 (2005-03-29) column 1, lines 8-13 column 3, lines 35-65 column 6, lines 21-49 figure 6	1,17-19
	----- -/--	

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents :

<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"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</p> <p>"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</p> <p>"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</p> <p>"&amp;" document member of the same patent family</p>
---	---

Date of the actual completion of the international search <b>17 March 2014</b>	Date of mailing of the international search report <b>27/03/2014</b>
---	---

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

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2013/077525

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 1 229 520 A2 (TELOGY NETWORKS INC [US]) 7 August 2002 (2002-08-07) abstract page 2, paragraph [0003] - page 3, paragraph [0007] figure 1	1,17-19
A	----- WO 2006/136901 A2 (NOKIA CORP [FI]; GREER STEVEN CRAIG [US]; GOURNAY PHILIPPE [CA]; JELIN) 28 December 2006 (2006-12-28) abstract page 2, line 1 - page 3, line 5 figure 1 -----	1,17-19

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2013/077525

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012110482	A2	23-08-2012	
		AR 085217 A1	18-09-2013
		AR 085218 A1	18-09-2013
		AR 085221 A1	18-09-2013
		AR 085222 A1	18-09-2013
		AR 085223 A1	18-09-2013
		AR 085224 A1	18-09-2013
		AR 085361 A1	25-09-2013
		AR 085362 A1	25-09-2013
		AR 085794 A1	30-10-2013
		AR 085895 A1	06-11-2013
		AU 2012217153 A1	10-10-2013
		AU 2012217156 A1	29-08-2013
		AU 2012217158 A1	13-12-2012
		AU 2012217160 A1	10-10-2013
		AU 2012217161 A1	26-09-2013
		AU 2012217162 A1	29-08-2013
		AU 2012217184 A1	19-09-2013
		AU 2012217215 A1	29-08-2013
		AU 2012217216 A1	26-09-2013
		AU 2012217269 A1	05-09-2013
		CA 2799343 A1	23-08-2012
		CA 2827000 A1	23-08-2012
		CA 2827156 A1	23-08-2012
		CA 2827249 A1	23-08-2012
		CA 2827266 A1	23-08-2012
		CA 2827272 A1	23-08-2012
		CA 2827277 A1	23-08-2012
		CA 2827296 A1	23-08-2012
		CA 2827305 A1	23-08-2012
		CA 2827335 A1	23-08-2012
		CN 102959620 A	06-03-2013
		CN 103460284 A	18-12-2013
		CN 103477386 A	25-12-2013
		CN 103477387 A	25-12-2013
		CN 103493129 A	01-01-2014
		CN 103503061 A	08-01-2014
		CN 103503062 A	08-01-2014
		CN 103534754 A	22-01-2014
		CN 103548078 A	29-01-2014
		CN 103620672 A	05-03-2014
		EP 2550653 A1	30-01-2013
		EP 2661745 A1	13-11-2013
		EP 2676262 A2	25-12-2013
		EP 2676264 A1	25-12-2013
		EP 2676265 A1	25-12-2013
		EP 2676266 A1	25-12-2013
		EP 2676267 A1	25-12-2013
		EP 2676268 A1	25-12-2013
		EP 2676269 A1	25-12-2013
		EP 2676270 A1	25-12-2013
		JP 2013531820 A	08-08-2013
		KR 20130007651 A	18-01-2013
		KR 20130126708 A	20-11-2013
		KR 20130126711 A	20-11-2013
		KR 20130133843 A	09-12-2013
		KR 20130133846 A	09-12-2013
		KR 20130133847 A	09-12-2013
		KR 20130133848 A	09-12-2013

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No  
PCT/EP2013/077525

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		KR 20130138362 A	18-12-2013
		KR 20140000322 A	02-01-2014
		KR 20140005277 A	14-01-2014
		SG 185519 A1	28-12-2012
		SG 192714 A1	30-09-2013
		SG 192715 A1	30-09-2013
		SG 192718 A1	30-09-2013
		SG 192721 A1	30-09-2013
		SG 192734 A1	30-09-2013
		SG 192745 A1	30-09-2013
		SG 192746 A1	30-09-2013
		SG 192747 A1	30-09-2013
		SG 192748 A1	30-09-2013
		TW 201237848 A	16-09-2012
		TW 201241823 A	16-10-2012
		TW 201246189 A	16-11-2012
		TW 201248615 A	01-12-2012
		TW 201248616 A	01-12-2012
		TW 201248617 A	01-12-2012
		TW 201250671 A	16-12-2012
		TW 201301262 A	01-01-2013
		TW 201301265 A	01-01-2013
		US 2013064383 A1	14-03-2013
		US 2013332148 A1	12-12-2013
		US 2013332151 A1	12-12-2013
		US 2013332152 A1	12-12-2013
		US 2013332153 A1	12-12-2013
		US 2013332174 A1	12-12-2013
		US 2013332175 A1	12-12-2013
		US 2013332176 A1	12-12-2013
		US 2013332177 A1	12-12-2013
		US 2013339036 A1	19-12-2013
		WO 2012110415 A1	23-08-2012
		WO 2012110416 A1	23-08-2012
		WO 2012110447 A1	23-08-2012
		WO 2012110448 A1	23-08-2012
		WO 2012110473 A1	23-08-2012
		WO 2012110476 A1	23-08-2012
		WO 2012110478 A1	23-08-2012
		WO 2012110480 A1	23-08-2012
		WO 2012110481 A1	23-08-2012
		WO 2012110482 A2	23-08-2012
-----			
US 6873604	B1	29-03-2005	NONE
-----			
EP 1229520	A2	07-08-2002	EP 1229520 A2 07-08-2002
			JP 2002237785 A 23-08-2002
			US 6807525 B1 19-10-2004
-----			
WO 2006136901	A2	28-12-2006	CA 2609945 A1 28-12-2006
			CN 101213591 A 02-07-2008
			EP 1897085 A2 12-03-2008
			JP 5351206 B2 27-11-2013
			JP 2008546341 A 18-12-2008
			JP 2011199875 A 06-10-2011
			MY 144785 A 15-11-2011
			TW I327852 B 21-07-2010
			US 2006293885 A1 28-12-2006



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2013/077525

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		WO 2006136901 A2	28-12-2006
		ZA 200800461 A	31-12-2008