FREQUENCY-DOMAIN COMFORT NOISE GENERATION FOR DISCONTINUOUS TRANSMISSION IN EVS

Anthony Lombard, Stephan Wilde, Emmanuel Ravelli, Stefan Döhla, Guillaume Fuchs, Martin Dietz

Fraunhofer IIS
Am Wolfsmantel 33
91058 Erlangen, Germany

{anthony.lombard,stephan.wilde,emmanuel.ravelli,stefan.doehla,guillaume.fuchs}@iis.fhg.de
martin.dietz@iis-extern.fraunhofer.de

ABSTRACT

Discontinuous Transmission (DTX) is an efficient way to drastically reduce the transmission rate of a communication codec in the absence of voice input. In this mode, most frames that are determined to consist of background noise only are dropped from transmission and replaced by some Comfort Noise Generation (CNG) in the decoder. In this paper, we propose a novel CNG approach combining information gained about the actual background noise at both encoder and decoder side. It is able to better reproduce background noise types showing a pronounced spectral tilt, which is difficult for traditional schemes based on a linear prediction model. The proposed technique operates in the frequency domain. It is part of the Enhanced Voice Services (EVS) codec, where it is known as FD-CNG. Listening tests show the superior quality of FD-CNG over existing approaches for certain background noise such as car noise.

Index Terms— speech coding, audio coding, CNG, DTX, EVS

1. INTRODUCTION

With the increasing availability of powerful mobile devices and the introduction of LTE (Long Term Evolution) for high-speed cellular access to the telecommunication networks, a new speech and audio communication codec has been developed in 3GPP for the Enhanced Voice Services (EVS) of LTE [1]. The EVS codec enhances the transmission efficiency and provides substantial quality improvements over existing codecs for low-delay conversational services. It supports a variety of bitrates ranging from 128 down to 5.9 kbps, and accepts fullband (FB), superwideband (SWB), wideband (WB) or narrowband (NB) signals sampled at 48, 32, 16 or 8 kHz, respectively. The EVS codec provides therefore a high-quality complement to narrowband codecs like the Adaptive Multi-Rate (AMR) codec [2, 3] or its WB extension known as AMR-WB [4, 5].

To further reduce the transmission rate, the EVS codec is equipped with a Discontinuous Transmission (DTX) mode applying Comfort Noise Generation (CNG) for inactive frames, i.e. frames that are determined to consist of background noise only. For these frames, a low-rate parametric representation of the signal is conveyed by Silence Insertion Descriptor (SID) frames at most every 8 frames (160 ms). This allows the CNG in the decoder to produce an artificial noise resembling the actual background noise.

In EVS, CNG can be achieved using either a linear predictive scheme (LP-CNG) or a frequency-domain scheme (FD-CNG), depending on the spectral characteristics of the background noise. This paper describes the FD-CNG approach, which is chosen by a CNG mode selector for noise types exhibiting a pronounced spectral tilt, such as a car noise. Contrary to exising DTX systems relying on linear prediction [6, 7], FD-CNG makes use of a frequency-domain noise estimation algorithm followed by a vector quantization of the background noise's smoothed spectral envelope. The decoded envelope is refined in the decoder by running a second frequency-domain noise estimator. Since a purely parametric representation is used during inactive frames, the noise signal is not available at the decoder in this case. Hence the noise estimation is performed during active phases only, i.e. on noisy speech content.

Noise estimation has been widely used over the last few decades in speech enhancement, where knowledge of the noise power spectral density allows to derive filters aiming at attenuating the noise while minimizing speech distortion [8, 9, 10]. A variety of noise estimation techniques have been proposed that do not rely on a speech pause detector [11, 12, 13]. Such schemes can be applied continuously on every frame, i.e. regardless of the speech activity, and therefore deliver meaningful information about the noise spectrum at any time. This is especially advantageous in the DTX context at the encoder side, since this minimizes the risk of transmitting wrong information at the beginning of a CNG phase, which may cause disturbing noise bursts or audible transitions between active frames and CNG frames. In FD-CNG, noise estimation is performed at encoder and decoder sides based on the minimum statistics algorithm [11].

The rest of the paper is organized as follows. The EVS encoder and decoder for FD-CNG is described in Section 2 and Section 3, respectively. Listening test results are presented in Section 4, computational complexity is discussed in Section 5, and conclusions are given in Section 6.

2. EVS ENCODING FOR FD-CNG

The EVS encoder is depicted in Fig. 1 for DTX/FD-CNG operation, where each frame is classified either as Ⓐ=active, Ⓢ=SID, or Ⓩ=zero (no data) frame. To obtain an artificial noise resembling the actual input background noise in terms of spectro-temporal characteristics, the FD-CNG makes use of a noise estimation algorithm to track the energy of the background noise present at the encoder input. The noise estimates are transmitted in the form of SID frames which are used at the decoder to update the amplitude of the random sequences generated in each frequency band during inactive phases, as shown in Section 3.

However, the size of an SID frame is very limited in practice (48 bits in the EVS codec). Hence, to reduce the number of parameters describing the noise, the input power spectrum is accumulated

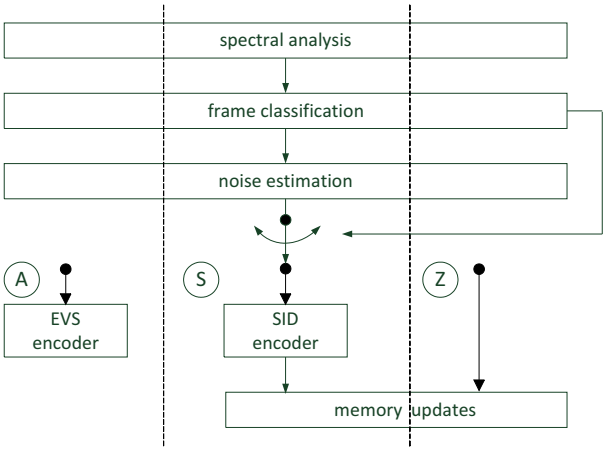


Fig. 1. EVS encoder in DTX/FD-CNG operation.

among up to 24 critical bands [14], as explained in the sequel.

2.1. Spectral analysis at the encoder

The FD-CNG relies on a hybrid spectral analysis approach. Low frequencies up to 6.4 kHz are covered by a 256-point FFT analysis, whereas higher frequencies are captured by a Complex Low-Delay Filter Bank (CLDFB) which exhibits a significantly lower spectral resolution of 400 Hz.

For each frame, the resulting power spectrum is divided into critical bands [14]. To enable rate switching at any time, the number L_{enc} of critical bands does not depend on the bitrate at this stage, in contrast to the SID encoder described in Section 2.3. This leads to a set of $L_{\mathrm{enc}} = 17$ energy values for an input sampling rate $f_{\mathrm{s,in}} = 8$ kHz, $L_{\mathrm{enc}} = 21$ for $f_{\mathrm{s,in}} = 16$ kHz, and $L_{\mathrm{enc}} = 24$ for $f_{\mathrm{s,in}} \geq 32$ kHz, regardless of the actual operation mode of the EVS codec. Note that FFT and CLDFB transforms are already computed for the preprocessing in the EVS codec and can be therefore re-used, as well as the building of the critical bands for the FFT spectrum.

2.2. Noise estimation at the encoder

A noise estimation algorithm is used to analyze the set of energy values over time and provides estimates of the background noise level for each band k at any given frame n. This is achieved on the basis of the minimum statistics algorithm [11] in a slightly modified form.

To reduce the dynamic range of the input energies and hence facilitate the fixed-point implementation of the noise estimation algorithm, the input energies $\sigma^2_{x,\mathrm{enc}}(n,k), k = 0, ..., L_{\mathrm{enc}} - 1$ are first processed by a non-linear transform and quantized with 9-bit resolution. Omitting the time index n for simplicity, like in the rest of this paper, it reads:

$$\sigma^2_{\tilde{x},\mathrm{enc}}(k) = \frac{\lfloor \log_2 \left(1 + \sigma^2_{x,\mathrm{enc}}(k)\right) 2^9 \rfloor}{2^9}, \qquad (1)$$

where $\lfloor \cdot \rfloor$ denotes the floor operator. The noise estimator is hence fed with logarithmic input data instead of linear data, which necessitates only minor parameter adjustments of the algorithm described in [11]. Note that the constant 1 inside the $\log_2(\cdot)$ ensures that $\sigma^2_{\tilde{x},\mathrm{enc}}(k)$ remains positive. This is especially important as the noise estimator relies on a statistical model of the noise energy. Performing noise estimation on negative values would hence strongly violate the model and would result in an unexpected behaviour.

Table 1. Encoder configurations for FD-CNG.

bandwidth	bitrate [kbps]	L_{SID}	$=$	$L_{\mathrm{SID}}^{[\mathrm{FFT}]}$	$+$	$L_{\mathrm{SID}}^{[\mathrm{CLDFB}]}$
NB	all	17	$=$	17	$+$	0
WB	≤ 8	20	$=$	20	$+$	0
	$8 < \cdot \leq 13.2$	21	$=$	20	$+$	1
	> 13.2	21	$=$	21	$+$	0
SWB/FB	≤ 13.2	24	$=$	20	$+$	4
	> 13.2	24	$=$	21	$+$	3

The resulting noise estimates $\hat{\sigma}^2_{\tilde{n},\mathrm{enc}}(k), k = 0, ..., L_{\mathrm{enc}} - 1$ are then converted back to a linear representation as follows:

$$\hat{\sigma}^2_{n,\mathrm{enc}}(k) = 2^{\hat{\sigma}^2_{\tilde{n},\mathrm{enc}}(k) - 1}, \qquad (2)$$

which reverses the non-linear transform applied in (1), except for the 9-bit quantization. Note that the base-2 logarithm in (1) and its inverse in (2) are both very well suited for fixed-point implementation.

2.3. Encoding SID frames

Among the L_{enc} noise estimates computed in (2), only the first L_{SID} values are quantized and transmitted via SID frames. Among those L_{SID} values, $L_{\mathrm{SID}}^{[\mathrm{FFT}]}$ values cover the FFT spectrum, and $L_{\mathrm{SID}}^{[\mathrm{CLDFB}]}$ cover the CLDFB spectrum, depending on the bitrate and the bandwidth used by the EVS codec, as shown in Table 1.

The L_{SID} noise estimates are first converted into dB and normalized by a global gain to capture the shape information of the spectrum. The global gain is quantized on 7 bits, whereas the normalized vector is encoded by a Multi-Stage Vector Quantizer (MSVQ) with 6 stages, see e.g. [15]. An M-best search algorithm is used to search for $M = 24$ survivors in each stage: 7 bits are allocated in the first stage, and 6 bits are allocated for each of the remaining 5 stages. The SID frames contain also a header of 4 bits to signal the selected CNG type, the codec bandwidth, and the bandwidth of the core encoder. Note that a single set of MSVQ codebooks serves for all configurations. The codebook vectors of length 24 are hence truncated to the appropriate length L_{SID} according to the EVS codec mode.

2.4. Memory updates

To ensure seamless transitions between active and inactive frames, the memories of the EVS encoder are updated during inactive frames. To this end, the quantized noise estimates obtained by decoding the SID frames are used to perform a local FD-CNG synthesis in the encoder.

3. EVS DECODING FOR FD-CNG

The DTX/FD-CNG decoder is depicted in Fig. 2. Based on the information transmitted in the SID frames, a comfort noise is generated in each FFT and CLDFB band. The amplitude of the random sequence is individually adjusted for each band to match the spectrum of the actual background noise, as explained in Section 3.3.

Unfortunately, the limited number of parameters transmitted in SID frames does not allow to capture the fine spectral structure of the noise. At the output of a DTX system, the discrepancy between the smoothed spectrum of the artificial noise and the spectrum of the actual background noise can become very audible at the transitions between active and CNG frames. It is therefore highly desirable to

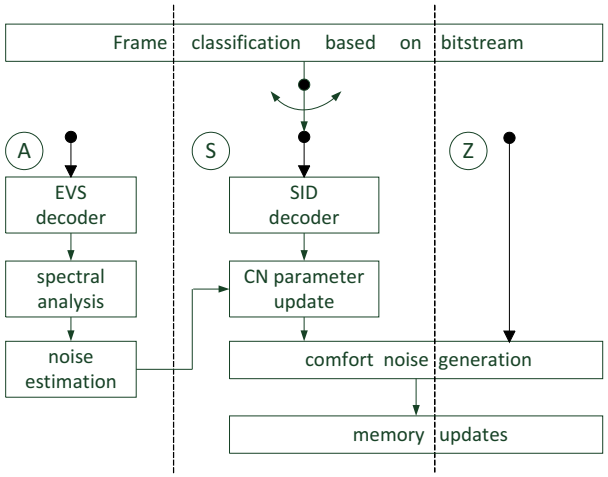


Fig. 2. EVS decoder in DTX/FD-CNG operation.

recover the information about the fine spectral structure of the noise in the decoder. This is achieved in FD-CNG by running a noise estimator at the output of the core EVS decoder during active frames.

3.1. Noise estimation at the decoder during active frames

The same noise estimation algorithm is used in encoder (see Section 2.2) and decoder, but with a significantly higher spectral resolution at the decoder to capture information about the fine spectral structure of the background noise.

The noise estimation is performed at the output of the EVS core decoder in the FFT domain, up to a certain boundary frequency f_{b} which depends on bandwidth and bitrate. The FFT length is adjusted to provide a constant frequency resolution of 25 Hz, regardless of the sampling frequency of the EVS core output signal which varies for the different bitrates and bandwidths.

Frequencies below 1 kHz are covered by the plain FFT resolution, i.e. with no grouping of FFT bin energies. For frequencies above 1 kHz, the FFT bins are grouped into spectral partitions for complexity reasons. While the 24 critical bands formed in the encoder correspond to the 24 integer indices of the Bark scale, critical bands are computed in the decoder using fractional Bark indices with roughly double spectral resolution, for frequencies between 1 kHz and the boundary frequency f_{b}. This offers a good trade-off between spectral resolution and computational complexity, leading to a number of spectral partitions L_{dec} ranging between 56 and 62, as shown in Table 2.

3.2. Combining encoder and decoder information

Thanks to its better spectral resolution, the decoder-side noise estimator captures information about the fine spectral structure of the background noise present during active frames, up to the boundary frequency f_{b}. However, it cannot adapt to changes in the background noise during inactive phases. In contrast, the SID frames provide some information about the evolution of the background noise's spectral envelope during inactive frames and over the entire bandwidth. The FD-CNG combines therefore these two sources of information in an effort to reproduce the fine spectral structure of the noise present during active phases (up to the boundary frequency f_{b}), while updating only the spectral envelope of the comfort noise during inactive parts.

Table 2. Decoder configurations for FD-CNG.

bandwidth	bitrate [kbps]	L_{dec}	f_{b}
NB	all	56	4.0
WB/SWB/FB	≤ 13.2	62	6.4
	> 13.2	61	8.0

First, a full-resolution FFT energy spectrum $\hat{\sigma}^2_{n,\mathrm{FR}}(\nu), \nu = 0, ..., L_{\mathrm{FFT}}/2$ is derived from the decoder-side noise estimates $\hat{\sigma}^2_n(k), k = 0, ..., L_{\mathrm{dec}} - 1$ by using a log-domain linear interpolation approach. The full-resolution spectrum is subsequently converted to the (lower) resolution of the encoder. The resulting noise energy spectrum $\hat{\sigma}^2_{n,\mathrm{LR}}(k), k = 0, \ldots, L^{[\mathrm{FFT}]}_{\mathrm{SID}} - 1$ exhibits therefore the same spectral resolution as the SID parameters $\hat{\sigma}^2_{n,\mathrm{SID}}(k), k = 0, \ldots, L^{[\mathrm{FFT}]}_{\mathrm{SID}} - 1$. Hence, both sets are comparable and can be easily combined. If $\nu_{\min}(k)$ and $\nu_{\max}(k)$ denote the indices of the first and last FFT bins in the critical band $k \in [0, \ldots, L^{[\mathrm{FFT}]}_{\mathrm{SID}} - 1]$, we have for the corresponding FFT bins $\nu = \nu_{\min}(k), \ldots, \nu_{\max}(k)$:

$$\hat{\sigma}^2_{n,\mathrm{CNG}}(\nu) = \frac{\hat{\sigma}^2_{n,\mathrm{SID}}(k)}{\hat{\sigma}^2_{n,\mathrm{LR}}(k)} \, \hat{\sigma}^2_{n,\mathrm{FR}}(\nu). \tag{3}$$

The $L^{[\mathrm{CLDFB}]}_{\mathrm{SID}}$ remaining SID parameters $\hat{\sigma}^2_{n,\mathrm{SID}}(k), k = L^{[\mathrm{FFT}]}_{\mathrm{SID}} - 1, \ldots, L_{\mathrm{SID}} - 1$ are linearly interpolated in the log-domain to yield a CLDFB energy spectrum at the desired spectral resolution of 400 Hz. Hence, the CNG levels for frequencies above the boundary frequency f_{b} are solely derived from the SID frames.

3.3. Comfort noise generation during inactive frames

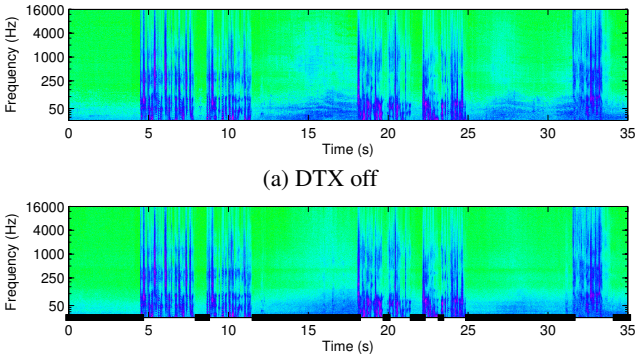
In FD-CNG, the noise generation is performed in the FFT domain for frequencies below the boundary frequency f_{b}, and in the CLDFB domain for higher frequencies. The CNG levels are recomputed after each SID update, as described above. They are used to generate a random Gaussian noise of zero mean and variance $\hat{\sigma}^2_{n,\mathrm{CNG}}(\nu)/2$ separately for the real and imaginary parts of each FFT and CLDFB coefficient. Finally, a hybrid spectral synthesis is used to convert the FFT/CLDFB spectrum into a time-domain comfort noise signal.

To avoid block artefacts at the transitions between active and inactive frames, a windowing and overlap-add mechanism is applied and the memories of the EVS decoder are updated during inactive frames based on the time-domain comfort noise signal.

For illustration, the spectrogram of an EVS decoder output is shown in Fig. 3 with DTX enabled or disabled. In DTX on mode, the CNG mode selector choosing between FD-CNG and LP-CNG was forced to choose FD-CNG for all inactive frames. We see that the two spectrograms are visually very similar, which indicates that the comfort noise is able to reproduce accurately the actual background noise during inactive frames. This visual impression is confirmed by the listening impression and no audible artefacts can be noticed at the transitions between active and inactive frames.

4. LISTENING TEST RESULTS

In order to evaluate the performance of the proposed CNG scheme, subjective evaluations of the EVS codec were conducted. The P.800 test methodology [16] was followed using the Degradation Category Rating (DCR). Three bandwidths (NB, WB and SWB) were tested in three separate tests using three different kinds of car noise with a



(a) DTX off

(b) DTX on

Fig. 3. SWB output of the EVS decoder in a noisy speech scenario (here a non-stationary street noise) at 24.4 kbps, with (a) DTX off and (b) DTX on. FD-CNG frames are signaled by a solid line along the x-axis.

Signal-to-Noise Ratio (SNR) of 10, 15 and 15 dB, respectively. The listening test participants were asked to grade the quality of noisy speech signals coded by the EVS codec in various conditions. The results of the listening tests are shown in Fig. 4. As references, signals coded with the AMR-WB [4] and G722.1 [17] codecs were included in each test, as well as Modulated Noise Reference Unit (MNRU) [18] and direct (i.e, uncompressed) conditions.

For the NB test, the EVS codec was assessed with DTX enabled and DTX disabled. In DTX on conditions, FD-CNG was compared to LP-CNG by forcing the CNG mode selection. In the WB test, EVS is also used both with DTX on and off. This time, the CNG mode selector was used to automatically choose between FD-CNG and LP-CNG, but in fact the selection logic systematically opts for FD-CNG for this kind of noise. In the SWB test, similar conditions were tested, but with 3 % of packet loss and concealment enabled.

The NB test demonstrates the superior quality of FD-CNG compared to LP-CNG on car noise. As this is not necessarily the case for other noise types, the CNG mode selector in the EVS codec has been designed to select FD-CNG rather than LP-CNG in the presence of car noise, where the higher frequency resolution offered by FD-CNG is particularly advantageous. In addition, all the tests show that EVS with DTX on and FD-CNG provides similar quality as EVS with DTX off at the corresponding bitrate. In other words, FD-CNG is able to encode the background noise with the same quality as the EVS codec in normal operation, i.e active transmission, but with a much lower bit consumption. Furthermore, the NB and WB tests show that EVS can achieve very high quality, i.e. almost as good as the direct condition, at 24.4 kbps with DTX on and FD-CNG selected. Finally, the robustness of FD-CNG to frame loss during inactive frames is demonstrated in the SWB test. FB results are not available but it is expected to behave similarly to the SWB case. Overall, we see also that the EVS codec improves over reference codecs in all conditions, both in terms of quality and bit consumption.

5. COMPUTATIONAL COMPLEXITY

Enabling DTX in EVS provides not only a gain in terms of bit consumption, but also a significant reduction of the workload. For instance, in SWB mode at 13.2 kbps with DTX on and FD-CNG selected, coding active frames requires about 45 WMOPS (Weighted Millions of Operations per Second) at the encoder, and 25 WMOPS

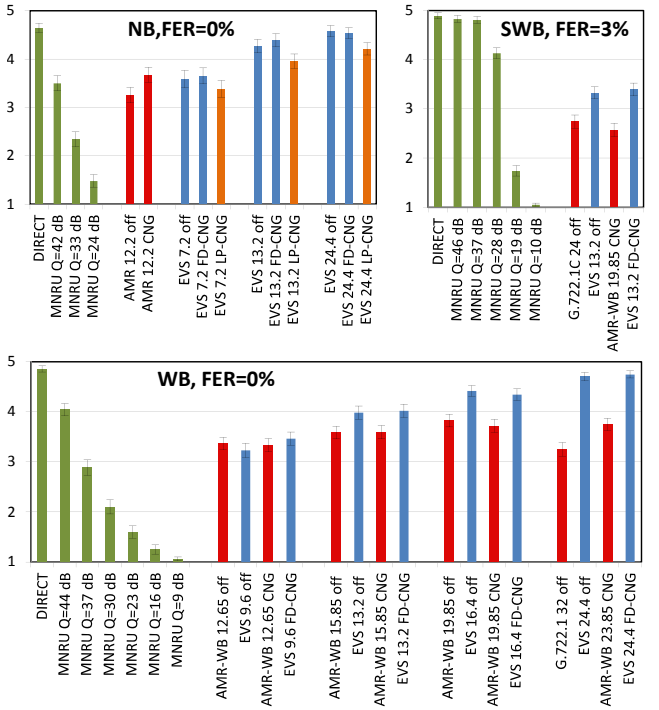


Fig. 4. P.800 listening test results comparing the EVS codec with reference codecs at various bitrates, with DTX off, DTX on with FD-CNG enabled, or DTX on with LP-CNG enabled.

at the decoder. For CNG frames, the workload decreases to 25 WMOPS at the encoder, and 16 WMOPS at the decoder. Hence, the average computational complexity of the codec can be reduced by up to about 40%, depending on the ratio of active to inactive frames. Note that LP-CNG enables a similar gain.

6. CONCLUSIONS

In this paper, we introduced a novel approach to generate a high-quality comfort noise signal in a communication codec operating in DTX mode. This technique, denoted as FD-CNG, has been recently standardized as part of the EVS codec [1]. The proposed method can cope with a variety of background noise types, but it is of particular interest for noise signals exhibiting a pronounced spectral tilt at low frequencies, where FD-CNG benefits from a better spectral resolution compared to other existing methods relying on LP models. Listening tests demonstrate the superior quality of FD-CNG for noisy speech items in car noise environments, for NB, WB as well as for SWB modes. While minimizing bit consumption and workload, we showed that enabling DTX with FD-CNG does not degrade quality compared to DTX off conditions, even for high bitrates.

A particularity of the proposed FD-CNG scheme is that it relies on a noise estimator in the decoder. This source of information is exploited for seamless transitions between active and CNG frames, but not only. It also opens the way for new enhancement tools in the decoder, where some comfort noise can be injected even in the absence of encoder information, as for active phases or also when DTX is disabled. In fact, this idea is exploited in the EVS codec and is denoted as Comfort Noise Addition (CNA) [1]. It is applied for low bitrates, where some low-level comfort noise is generated in an FD-CNG fashion and added to the decoded signal to mask potential coding artifacts, regardless of whether DTX is enabled or not.

7. REFERENCES

[1] *EVS Codec Detailed Algorithmic Description*, 3GPP Technical Specification 26.445, Sep. 2014. [Online]. Available: http://www.3gpp.org/DynaReport/26445.htm

[2] *Mandatory speech CODEC speech processing functions; AMR speech codec; General description*, 3GPP Technical Specification 26.071. [Online]. Available: http://www.3gpp.org/DynaReport/26071.htm

[3] K. Järvinen, "Standardisation of the adaptive multi-rate codec," in *Proc. Eur. Signal Processing Conf. (EUSIPCO)*, Sep. 2000.

[4] *Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description*, 3GPP Technical Specification 26.171, Mar. 2001. [Online]. Available: http://www.3gpp.org/DynaReport/26171.htm

[5] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Trans. Speech Audio Processing*, vol. 10, no. 8, pp. 620–636, Nov. 2002.

[6] *Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Comfort noise aspects*, 3GPP Technical Specification 26.192, Mar. 2001. [Online]. Available: http://www.3gpp.org/DynaReport/26192.htm

[7] *Frame error robust narrow-band and wideband embedded variable bit rate coding of speech and audio from 8-32 kbit/s*, ITU-T Recommendation G.718, Jun. 2008. [Online]. Available: http://www.itu.int/rec/T-REC-G.718

[8] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. 27, no. 2, pp. 113–120, Apr. 1979.

[9] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[10] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, ser. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2005.

[11] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 5, pp. 504–512, Jul. 2001.

[12] T. Gerkmann and R. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Language Processing*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[13] L. Lin, W. H. Holmes, and E. Ambikairajah, "Adaptive noise estimation algorithm for speech enhancement," *Electronics Letters*, vol. 39, no. 9, pp. 754–755, May 2003.

[14] E. Zwicker, "Subdivision of the audible frequency range into critical bands (frequenzgruppen)," *The Journal of the Acoustical Society of America*, vol. 33, no. 2, 1961.

[15] W. LeBlanc, B. Bhattacharya, S. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 4, pp. 373–385, Oct. 1993.

[16] *Methods for subjective determination of transmission quality*, ITU-T Recommendation P.800, Aug. 1982. [Online]. Available: http://www.itu.int/rec/T-REC-P.800

[17] *Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss*, ITU-T Recommendation G.722.1, May 2005. [Online]. Available: http://www.itu.int/rec/T-REC-G.722.1

[18] *Modulated noise reference unit (MNRU)*, ITU-T Recommendation P.810, Feb. 1996. [Online]. Available: www.itu.int/rec/T-REC-P.810