A LOW-COMPLEXITY NOISE SUPPRESSOR WITH NONUNIFORM SUBBANDS AND A FREQUENCY-DOMAIN HIGHPASS FILTER

Masanori Kato and Akihiko Sugiyama

Media and Information Research Labs., NEC Corporation
Kawasaki 211-8666, JAPAN

ABSTRACT

This paper proposes a low-complexity noise suppressor with nonuniform subbands and a frequency domain highpass filter. A highpass filter to suppress undesirable low frequency components in the input signal is replaced with a simple DC offset canceler in the time-domain combined with a frequency-domain weighting function. Frequency bins after Fourier transform are nonuniformly grouped to reduce the computations for calculating the spectral gain. A new decomposition pattern is developed for higher quality of the enhanced speech. The number of operations on typical DSP chips is 3.2 MIPS that is approximately 50% off the conventional load. Subjective evaluation results demonstrates that the low-complexity noise suppressor achieves comparable quality of the enhanced speech to that of the conventional noise suppressor.

1. INTRODUCTION

The third generation (3G) mobile communication service has been becoming more and more popular all over the world. 3GPP (The 3rd Generation Partnership Project) has standardized, for its standard AMR (Adaptive Multi-Rate) codec[1], an AMR noise suppressor. However, a single noise suppressor is not standardized. Instead, the minimum performance requirements for noise suppressor and the evaluation procedure had been standardized[2]. The mobile network operator or the terminal manufacturer can choose whichever noise suppressor algorithm they like.

Several noise suppressors[3]-[6] that satisfy all the minimum performance requirements have been developed. Their complexities are all higher than 5 wMOPS (weighted million operations per second) on the 3GPP virtual chip. There is only one report on DSP (digital signal processor) implementation, where [4] and [5] consume more than 6 MIPS[7]. In view of a fact that the AMR codec can be implemented with 12 MIPS, there is considerable motivation to develop a low-complexity noise suppressor that provides good enhanced-speech quality.

This paper proposes a low-complexity noise suppressor whose computational requirement is approximately 3 MIPS while preserving an equivalent subjective quality to a 3GPP endorsed noise suppressor. In the new noise suppressor, nonuniform subbands and a frequency domain highpass filter are newly introduced. In the following section, one of the 3GPP endorsed noise suppressors is reviewed as the basis for new development. Section 3 presents the low-complexity noise suppressor. Considerable reduction in computations is demonstrated in Section 4, followed by subjective listening test results in Section 5.

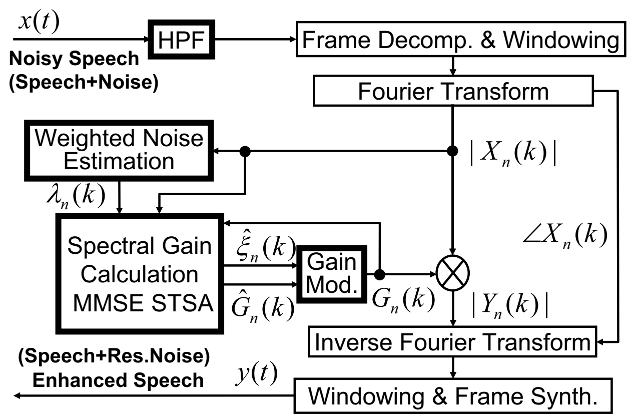

Fig. 1. A Conventional 3GPP Noise Suppressor.

2. CONVENTIONAL 3GPP NOISE SUPPRESSOR

2.1. Noise Suppression Algorithm

One of the 3GPP endorsed noise suppressors[5] was used as the basis in the development of a low-complexity version because of its good balance between the enhanced-speech quality and the complexity. It is based on MMSE STSA (Minimum Mean Square Error Short Time Spectral Amplitude) originally proposed by Ephraim and Malah[8]. A block diagram of the noise suppressor is depicted in Fig. 1.

It features weighted noise estimation and synthesis windowing. Noise estimation is carried out using an estimated signal-to-noise ratio (SNR). It enables continuous noise estimation even during speech sections resulting in better tracking capability for nonstationary noise. A synthesis windowing function is applied between inverse transform and overlap-add processing for smooth transition from a frame to the next by flattening out the gaps at frame boundaries.

2.2. Bottlenecks in Computations

The highpass filter has a sharp cut-off frequency response near 100Hz and therefore, is generally implemented as an IIR (infinite impulse response) filter for smaller number of

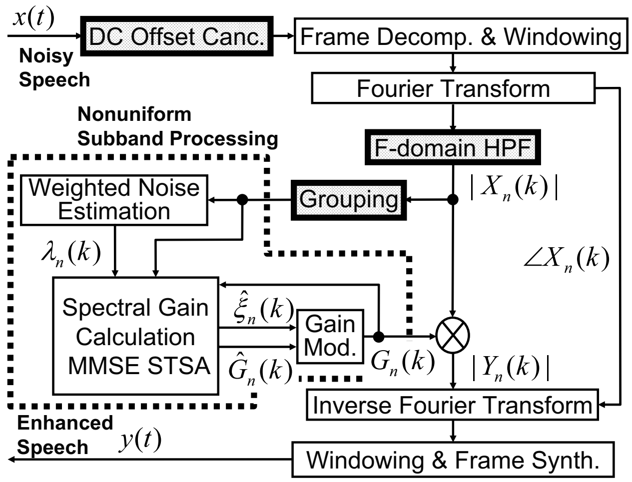

Fig. 2. Low-Complexity Noise Suppressor.

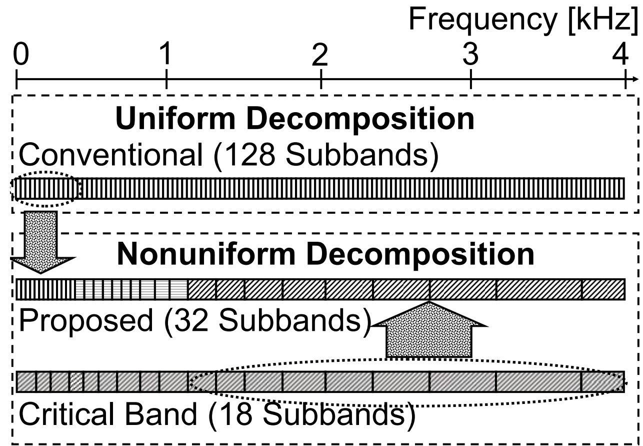

Fig. 3. Nonuniform Subband Decomposition.

computations than an FIR (finite impulse response) filter. Its transfer function is a rational function whose denominator is sensitive to an error in the coefficients. To guarantee sufficient precision for the sharp cut-off characteristic in the fixed point implementation, double-precision has been used for the denominator. One double-precision operation is usually implemented as multiple single-precision operations, leading to heavier computational load. On the other hand, if the highpass filter is removed for reduced computations, the input-signal linearity is hard to be preserved, which may result in degraded enhanced-speech quality.

Another possible bottle neck lies in noise estimation, spectral gain calculation, and gain modification, which consume approximately 50% of the total computations[1]. These operations are repeated for all frequency bins. If this repetition can be reduced, the resulting reduction in total computations will be significant. It is known as a psychoacoustical fact that human ears are insensitive in high frequencies[10]. Therefore, it should be worth trying to share the same spectral gain among multiple frequency bins in high frequencies.

3. LOW-COMPLEXTY NOISE SUPPRESSOR

The low-complexity noise suppressor carries out highpass filtering in the frequency domain and calculates spectral gains for a smaller number of frequency bins. A blockdiagram of the low-complexity noise suppressor is depicted in Fig. 2. New functions are highlighted as shaded boxes surrounded by a bold line. A highpass filter, which is usually placed before the noise suppressor, is decomposed into a time-domain DC offset canceller (DC Offset Canc.) and a frequency-domain highpass filter (F-domain HPF). A grouping unit (Grouping) nonuniformly combines frequency bins after Fourier transform for weighted noise estimation, spectral gain calculation, and gain modification. These three functions, which are surrounded by a dashed line, all operate in a reduced number of subbands, sharing the same spectral gain among multiple frequency bins.

[1]Details are presented in Fig. 4.

3.1. Frequency-Domain Highpass Filter

The proposed noise suppressor performs highpass filtering in the frequency domain combined with time-domain offset canceling. The offset canceller removes a DC offset that may degrade the accuracy in Fourier transform implemented with a limited wordlength. The offset $\bar{x}_o(n)$ is calculated as an averaged time-domain samples in each frame as

$$x_o(n) = \frac{1}{L} \sum_{t=L(n-1)+1}^{nL} x(t), \tag{1}$$

$$\bar{x}_o(n) = \begin{cases} x_o(n), & n = 0 \\ \alpha x_o(n-1) + (1-\alpha) x_o(n), & n \neq 0 \end{cases} \tag{2}$$

where n is the index to the current frame with $x_o(0) = 0$ and L, the frame size. α is a constant satisfying $0 \leq \alpha \leq 1$. This offset is subtracted from all samples of the input noisy speech $x(t)$ in the n-th frame such that $x(t) - \bar{x}_o(n)$ will be framed and Fourier-transformed.

The Fourier transformed samples are weighted by the F-domain HPF in the frequency-domain. The weight for each bin should correspond to the frequency response of the magnitude of the original highpass filter. However, only a small number of low-frequency bins are actually weighted because all others are equal to unity. In the case of 256-point Fourier transform, 15 bins out of 128 are to be weighted.

All together, the frequency-domain highpass filter requires two additions and two multiplications for DC-offset calculation, one subtraction for DC-offset removal, both per sample, as well as 15 multiplications for weighting per frame. Thanks to simple scaling operations in the frequency domain, double-precision operations are no longer needed.

3.2. Nonuniform Subbands

The principle of psychoacoustics[10] suggests that a spectral gain may be shared among adjacent high-frequency components. This fact naturally leads to nonuniform grouping of frequency components after Fourier transform. However, it is not straightforward. Informal listening tests in

Table 1. Nonuniform Subband Decomposition by Frequency-Bin Combination (8 kHz Sampling and 256-point Fourier Transform)

Subband Index	Frequency Bin Index (Number of Bins)	Frequency [Hz]
0	0 (1)	0 – 31
1	1 (1)	31 – 62
···	···	···
12	12 (1)	375 – 406
13	13 – 14 (2)	406 – 469
14	15 – 16 (2)	469 – 531
15	17 – 18 (2)	531 – 594
16	19 – 20 (2)	594 – 656
17	21 – 22 (2)	656 – 719
18	23 – 24 (2)	719 – 781
19	25 – 26 (2)	781 – 844
20	27 – 29 (3)	844 – 938
21	30 – 32 (3)	938 – 1031
22	33 – 36 (4)	1031 – 1156
23	37 – 42 (6)	1156 – 1344
24	43 – 48 (6)	1344 – 1531
25	49 – 56 (8)	1531 – 1781
26	57 – 65 (9)	1781 – 2063
27	66 – 75 (10)	2063 – 2375
28	76 – 87 (12)	2375 – 2750
29	88 – 101 (14)	2750 – 3188
30	102 – 119 (18)	3188 – 3750
31	120 – 128 (9)	3750 – 4000

the development revealed that grouping of frequency bins based on the critical band[10] does not provide as good enhanced-speech quality as the conventional noise suppressor [5]. Such degradation comes from insufficient frequency resolution, which leads to oversuppression of spectral peaks.

This problem is alleviated in two ways; higher resolution in low frequency range and adjustment to the estimated noise in the high frequency range. The strategy in subband decomposition is summarized in Fig. 3. The subband decomposition in the high frequency is copied from the critical band, while each low-frequency bin is kept independent. Mid-frequency bins are grouped with a smaller number of bins than the critical band. The estimated noise is scaled down in the frequency range where the critical band is preserved. A smaller noise leads to a higher SNR, resulting in weak suppression and low distortion.

Table 1 shows how each frequency bin is combined in the case of 8 kHz sampling with 256-point Fourier transform. The frequency bins higher than approximately 1030 Hz are grouped with more than four bins according to the critical band. The proposed nonuniform subbands lower than 1031 Hz consist of 13 subbands with a single bin, 7 subbands with two bins, 2 subbands with three bins. The total number of subbands is reduced by 75% from 128 to 32.

4. COMPLEXITY EVALUATION

The time-domain highpass filtering requires 21 multiplications per sample assuming a 3-stage biquad IIRs. This is equivalent to approximately 3400 multiplications per frame that consists of 160 samples. On the contrary, the frequency-domain highpass filtering necessitates 15 multiplications per frame. More than 99% reduction is expected in multiplications alone in highpass filtering. In noise estimation, SNR calculation, and spectral gain calculation, the computational load should be reduced by 75% because the number of subbands has been reduced from 128 to 32.

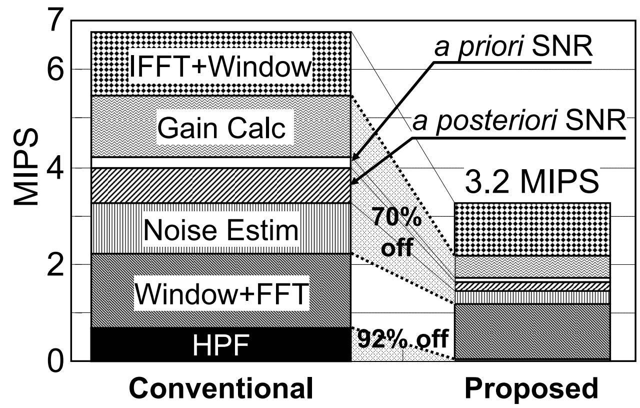

Fig. 4. Computational Reduction.

Table 2. Number of Operations on Typical DSPs

	μPD77210	TMS320VC5510		
Fs	8	8		16
FFT Size	256	256	512	512
Frame Size	160	160	256	256
Overlap	40	40	256	256
MIPS(mono)	3.2	3.2	1.8	3.6

The low-complexity noise suppressor was implemented, for different applications, on μPD77210 by NEC [11] and TMS320VC5510 by Texas Instruments [12]. The number of operations in total for μPD77210 has been reduced by approximately 50% to 3.2 MIPS in case of 8kHz sampling and a frame size of 160 with a 40-sample overlap. Detailed MIPS for each function is summarized in Fig. 4. It should be noted that table-look-up has been applied to both implementations wherever possible. The average saving for Gain Calc., *a priori* and *a posteriori* SNR calculations, and Noise Estim is approximately 70%, which almost agrees with the estimation described earlier in this section. For highpass filtering (HPF), the reduction is slightly higher than 90%. This is smaller than the estimated result. The difference mainly comes from data-handling operations such as data load and store. When the number of operations is evaluated in wMOPS, it is 4.33 wMOPS, resulting in 42% reduction from 7.43 wMOPS.

With a similar number of operations, the low-complexity noise suppressor can be implemented by TMS320VC5510. A different frame size, a different overlap size, and/or a different sampling frequency requires a different number of operations. The total computations on these typical DSPs are summarized in Tab. 2. When a larger frame size of 256 and 256-sample overlap are used, the total computations becomes 1.8 MIPS. For a sampling frequency of 16kHz, it is doubled.

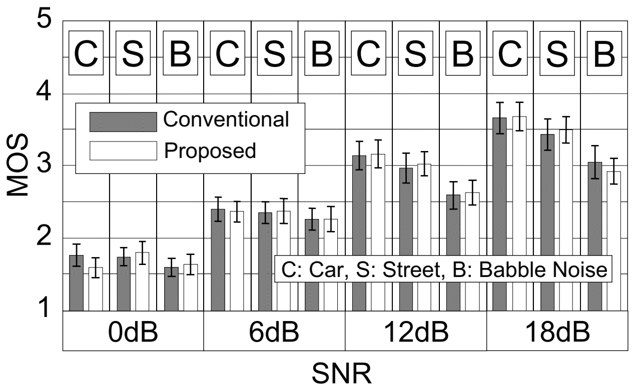

Fig. 5. Subjective Quality of the Enhanced Speech (ACR).

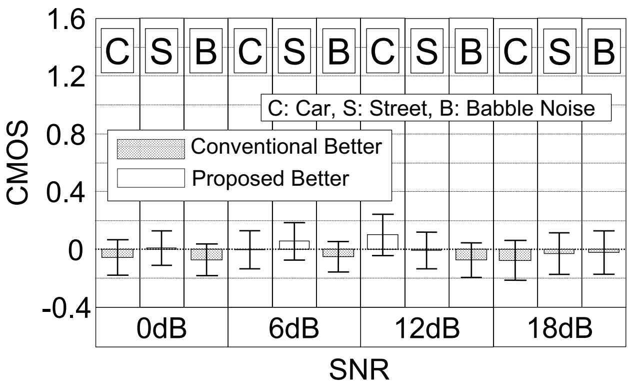

Fig. 6. Subjective Quality of the Enhanced Speech (CCR).

5. SUBJECTIVE EVALUATIONS

Subjective evaluations have been performed to assess the enhanced-speech quality. Car, street, and babble noise signals were used as in the 3GPP evaluation. The SNR was set to 0, 6, 12, and 18dB. 22 subjects at ages between 20 and 40 were asked to score the enhanced speech by the conventional [4] and the low-complexity noise suppressor. Evaluations were based on two different ratings, namely, Absolute Category Rating (ACR) based on a 5-grade mean opinion score (MOS) and Comparison Category Rating (CCR) based on a 7-grade CMOS.

Figure 5 exhibits the ACR results. The height of the bar represents the average score and the vertical line at the top of the bar exhibits the 95% confidence interval. It is clear from Fig. 5 that there is no statistically significant difference between the two, as is shown by overlapping 95% confidence intervals. The CCR results are illustrated in Fig. 6. The height of the bar from the 0 line represents the average CCR. The 95% confidence interval is expressed by a vertical line. Because all the vertical lines cross the 0 line, the subjective quality of the low-complexity noise suppressor is statistically comparable to that of the conventional one [4] for all the tested conditions. Moreover, the low-complexity noise suppressor satisfies all the 3GPP minimum performance requirements[13]. The evaluation results have already been endorsed by 3GPP[14].

6. CONCLUSION

A low-complexity noise suppressor with nonuniform subbands and a frequency domain highpass filter has been proposed. A new subband decomposition has been developed for higher quality of the enhanced speech. The number of operations on a typical DSP chip has been reduced by approximately 50% to 3.2 MIPS. Subjective evaluation results have demonstrated that the proposed noise suppressor achieves comparable quality of the enhanced speech to the conventional noise suppressor. 3GPP endorsement has already been obtained to show that the low-complexity noise suppressor satisfies the minimum performance requirements.

7. REFERENCES

[1] "Digital cellular telecommunications system (Phase 2+); Adaptive Multi-Rate (AMR); speech processing functions; General description," 3GPP TS 06.71 Release 98.

[2] "Minimum performance requirements for noise suppresser application to the AMR speech encoder," 3GPP TS 06.77 V8.1.1, Apr. 2001.

[3] "Test results of Mitsubishi AMR-NS solution based on TS 26.077," 3GPP Tdoc S4-020251, May 2002.

[4] "Test results of NEC AMR-NS solution based on TS 26.077," 3GPP Tdoc S4-020415, July 2002.

[5] "Test results of NEC low-complexity AMR-NS solution based on TS 26.077," 3GPP Tdoc S4-020417, July 2002.

[6] "Test results of Fujitsu AMR-NS solution based on TS 26.077 Specification," 3GPP Tdoc S4-050016, Feb. 2005.

[7] M. Kato, A. Sugiyama, and M. Serizawa, "A Family of 3GPP-Standard Noise Suppressors for the AMR Codec and Its Evaluation Results," ICASSP'03, pp. 916-919, Apr. 2003.

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

[9] "Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems," TR-45, IS-127, July 1996.

[10] E. Zwicker and H. Fastl, "Psychoacoustics," Springer-Verlag, Jun. 1990.

[11] "μPD77210,77213 Data Sheet," U15203EJ-3V0DS00, NEC, Nov. 2001.

[12] "TMS320VC5510 Fixed-Point Digital Signal Processor DATA Manual," SPRS076C, Texas Instruments, May 2002.

[13] "Test results of NEC's new low-complexity AMR-NS solution based on TS 26.077," 3GPP Tdoc S4-030601, Sep. 2003.

[14] "TSG SA WG4 status report at TSG SA#21," TSGS#21-030433, Sep. 2003.