Speech Enhancement Using Noise Estimation Based on Dynamic Quantile Tracking for Hearing Impaired Listeners

Nitya Tiwari and Prem C. Pandey
Indian Institute of Technology Bombay
Mumbai 400076, India
Email: { nitya, pcpandey } @ ee.iitb.ac.in

***Abstract*— Persons with sensorineural hearing loss experience difficulty in speech perception in noisy environments. They can benefit from signal processing for suppressing the background noise in their hearing aids. For such an application, a speech enhancement technique using spectral subtraction and noise spectrum estimation based on dynamic quantile tracking is presented. It does not involve voice activity detection or storage and sorting of past spectral samples. Enhancement of speech corrupted with different types of additive stationary and non-stationary noise showed improvement in speech quality to be equivalent to an SNR advantage of 3 – 6 dB. The technique is implemented and tested for satisfactory real-time operation, with sampling frequency of 10 kHz, on a 16-bit fixed-point DSP processor with on-chip FFT hardware.**

***Keywords*— hearing aid; quantile estimation; spectral subtraction; speech enhancement**

I. Introduction

Hearing impaired listeners most likely to benefit from use of hearing aids generally have moderate-to-severe loss with varying combination of conductive and sensorineural losses. Conductive loss is caused by disorders of the middle ear and is characterized by frequency-dependent elevation of hearing thresholds without any change in the dynamic range of hearing. It can be compensated by frequency-selective amplification. Sensorineural loss is caused by degeneration of the sensory hair cells of the inner ear or the auditory nerve. It occurs due to aging, excessive exposure to noise, infection, or congenital defects. Persons with this kind of loss, in addition to having frequency-dependent elevation of hearing thresholds, have significantly reduced dynamic range of hearing and abnormal loudness growth leading to distorted loudness relationship among speech components, increased temporal masking leading to poor detection of acoustic landmarks, and increased spectral masking leading to reduced ability to sense spectral shapes [1]–[3]. Persons with such loss experience difficulty in speech perception, particularly in noisy environments.

Digital hearing aids generally have processing for frequency-selective amplification, dynamic range compression, and tone suppression, but not for decreasing the adverse effects of increased temporal and spectral masking and for suppressing wideband non-stationary noise. Several signal processing techniques, such as binaural dichotic presentation [4], [5], spectral contrast enhancement [6], multiband frequency compression [7], [8], and enhancement of consonant-vowel ratio [9], have been reported for improving speech perception by such listeners. These techniques assume noise-free speech signal to be available as the input. Thus processing for suppression of wide-band non-stationary background noise as part of the signal processing in hearing aids can serve as a practical solution for improving speech quality and intelligibility for persons with sensorineural or mixed loss and it can also be used as a pre-processing stage for many other techniques.

For implementing the noise suppression technique on a low-power processor in a hearing aid, it should have low algorithmic delay and low computational complexity. Spectral subtraction [10], [11] is a single-input speech enhancement technique, and it can be considered as a good candidate for this application. A large number of variations of the basic technique have been developed for use in audio codecs and speech recognition [12]–[14]. The processing steps are estimating the noise spectrum, subtracting it from the noisy speech spectrum, and re-synthesizing the speech signal. Due to non-stationary nature of the interfering noise, its spectrum needs to be dynamically estimated. Under-estimation of the noise results in residual noise and over-estimation results in distortion leading to degraded quality and reduced intelligibility. Noise can be estimated during the silence intervals identified by a voice activity detector, but the detection may not be satisfactory under low-SNR conditions and the method may not correctly track the noise spectrum during long speech segments.

Several techniques based on minimum statistics for estimating the noise spectrum, without voice activity detection, have been reported [15]–[17]. These techniques involve tracking the noise as minima of the magnitude spectra of the past frames and are suitable for real-time operation. However, they often underestimate the noise and need estimation of an SNR-dependent subtraction factor. In the absence of significant silence segments, processing may remove some parts of the speech signal during the weaker segments. It has been reported [18] that a quantile-based estimation of the noise spectrum from the spectrum of the noisy speech can be used for spectral subtraction based noise

The research is supported by "National Program on Perception Engineering," sponsored by the Department of Electronics & Information Technology, MCIT, Government of India.

suppression. It is based on the observation that the signal energy in a particular frequency bin is low in most of the frames and high only in 10-20% frames corresponding to voiced speech segments. Stahl et al. [18] reported a 26% decrease in word error in speech recognition of utterances degraded by car noise, when 0.55-quantile was used to estimate the noise spectrum. Using a time-frequency quantile based noise estimation, Evans and Mason [19] reported 35% improvement in word accuracy in a speech recognition task. A two-pass quantile based noise estimation technique by Bai and Wan [20] involved estimating SNR for each time frequency point using a fixed quantile, and using it to determine a new quantile level for each frequency sub-band. These quantile-based noise estimation techniques use quantiles obtained by ordering the spectral samples or from dynamically generated histograms. Due to large memory space required for storing the spectral samples and high computational complexity, they are not suited for use in hearing aids. Use of median, i.e. 0.5-quantile, considerably reduces the computation requirement, but still does not permit real-time implementation. Waddi et al. [21] used a cascaded-median as an approximation to median for real-time implementation of speech enhancement. Processing resulted in significant improvements in objective PESQ score for speech quality for different types of noises: 0.48 – 0.90 for white noise and 0.13 – 0.33 for babble noise. The results indicate the need for using frequency-bin dependent quantiles for suppression of non-white and non-stationary noises.

We present a technique for noise spectrum estimation based on dynamic quantile tracking as an approximation to the quantile value obtained by sorting, without involving storage and sorting of past samples. It is used for speech enhancement using spectral subtraction and implemented using a fixed-point DSP chip for possible use in hearing aids. The following sections present the technique, its implementation for real-time speech enhancement, test results, and conclusions.

II. Signal Processing Technique

A. Estimation of noise spectrum by dynamic quantile tracking

In the proposed technique, the signal is processed using overlapping analysis windows or frames. The quantile is estimated at each frame by applying an increment or a decrement on the previous estimate. The increment and decrement are selected to be appropriate fractions of the range such that the estimate after a sufficiently large number of input frames matches the sample quantile. As the underlying distribution of the spectral samples is unknown, the range also needs to be dynamically estimated.

Let the kth spectral sample of the noise magnitude spectrum $D_n(k)$ be estimated as the $p(k)$-quantile of the magnitude spectrum $|X_n(k)|$. It is tracked dynamically as

$$D_n(k) = D_{n-1}(k) + d_n(k) \tag{1}$$

where the change $d_n(k)$ is given as

$$d_n(k) = \begin{cases} \Delta_+(k), & |X_n(k)| \geq D_{n-1}(k) \\ -\Delta_-(k), & \text{otherwise} \end{cases} \tag{2}$$

The values of $\Delta_+(k)$ and $\Delta_-(k)$ should be such that the quantile estimate approaches the sample quantile and sum of the changes in the estimate approaches zero, i.e. $\sum d_n(k) \approx 0$. For stationary input and sufficiently large number of frames M, $d_n(k)$ is expected to be $-\Delta_-(k)$ for $p(k)M$ frames and $\Delta_+(k)$ for $(1-p(k))M$ frames. Therefore,

$$(1-p(k))M\ \Delta_+(k) - p(k)M\ \Delta_-(k) \approx 0 \tag{3}$$

Thus the ratio of the increment to the decrement should satisfy the following condition:

$$\Delta_+(k)/\Delta_-(k) = p(k)/(1-p(k)) \tag{4}$$

and therefore $\Delta_+(k)$ and $\Delta_-(k)$ may be selected as

$$\Delta_+(k) = \lambda p(k)R \tag{5}$$

$$\Delta_-(k) = \lambda(1 - p(k))R \tag{6}$$

where R is the range (difference between the maximum and minimum values of the sequence of spectral values in a particular frequency bin) and λ is a factor which controls the step size during tracking.

As the sample quantile may be overestimated by $\Delta_+(k)$ or underestimated by $\Delta_-(k)$, the ripple in the estimated value is given as

$$\begin{aligned} \delta &= \Delta_+(k) + \Delta_-(k) \\ &= \lambda R \end{aligned} \tag{7}$$

During tracking, the number of steps needed for the estimated value to change from initial value $D_i(k)$ to final value $D_f(k)$ is given as

$$S = \max\left[\frac{D_f(k) - D_i(k)}{\Delta_+(k)},\ \frac{D_i(k) - D_f(k)}{\Delta_-(k)}\right] \tag{8}$$

Since $(|D_f(k) - D_i(k)|)_{\max} = R$, the maximum value of S is given as

$$S_{\max} = \max\left[\frac{1}{\lambda p(k)},\ \frac{1}{\lambda(1-p(k))}\right] \tag{9}$$

The factor λ can be considered as the convergence factor and its value should be selected for an appropriate tradeoff between δ and $S_{\max}$. It may be noted that the convergence becomes slow for very low or high values of $p(k)$.

The range is estimated using dynamic peak and valley detectors. The peak $P_n(k)$ and the valley $V_n(k)$ are updated, using the following first-order recursive relations:

$$P_n(k) = \begin{cases} \tau P_{n-1}(k) + (1-\tau)\,|X_n(k)|, & |X_n(k)| \geq P_{n-1}(k) \\ \sigma P_{n-1}(k) + (1-\sigma) V_{n-1}(k), & \text{otherwise} \end{cases} \tag{10}$$

$$V_n(k) = \begin{cases} \tau V_{n-1}(k) + (1-\tau)\,|X_n(k)|, & |X_n(k)| \leq V_{n-1}(k) \\ \sigma V_{n-1}(k) + (1-\sigma) P_{n-1}(k), & \text{otherwise} \end{cases} \tag{11}$$

The constants τ and σ are selected in the range [0, 1] to control the rise and fall times of the detection. As the peak and valley samples may occur after long intervals, τ should be small to provide fast detector responses to an increase in the range and σ should be relatively large to avoid ripples.

The range is tracked as

$$R_n(k) = P_n(k) - V_n(k) \tag{12}$$

The dynamic quantile tracking for estimating the noise spectrum as given by (1), (2), (5), and (6) can be written as the following:

$$D_n(k) = \begin{cases} D_{n-1}(k) + \lambda p(k) R_n(k), & |X_n(k)| \geq D_{n-1}(k) \\ D_{n-1}(k) - \lambda(1-p(k)) R_n(k), & \text{otherwise} \end{cases} \tag{13}$$

The computation steps of the technique as given by (10) – (13) are shown as a block diagram in Fig. 1.

B. Speech enhancement by spectral subtraction

Several types of spectral subtraction techniques for suppressing additive noise have been reported [10]–[14]. Fig. 2 shows a block diagram of the technique as used by us. The processing blocks are: windowing, FFT calculation, noise spectrum estimation, enhanced magnitude spectrum calculation, estimating enhanced complex spectrum without explicit phase estimation, and re-synthesis using IFFT with overlap-add.

Windowed segments of the input $x(n)$ are used as the analysis frames and FFT is used to obtain the spectra. The noise magnitude spectrum $D_n(k)$ is estimated using dynamic quantile tracking. The enhanced magnitude spectrum $|Y_n(k)|$ is computed, using generalized spectral subtraction, as

$$|Y_n(k)| = \begin{cases} \beta^{1/\gamma} D_n(k), & |X_n(k)| < (\alpha+\beta)^{1/\gamma} D_n(k) \\ [|X_n(k)|^\gamma - \alpha (D_n(k))^\gamma]^{1/\gamma}, & \text{otherwise} \end{cases} \tag{14}$$

The exponent factor γ may be selected appropriately for power subtraction ($\gamma = 2$) or magnitude subtraction ($\gamma = 1$). Choosing subtraction factor $\alpha > 1$ helps in reducing the broadband peaks in the residual noise, but it may result in deep valleys, causing warbling or musical noise which is masked by a floor noise controlled by the spectral floor factor β.

The complex spectrum is obtained by combining the enhanced magnitude spectrum with the original noisy phase. In order to avoid phase calculation, the complex spectrum is calculated using

$$Y_n(k) = |Y_n(k)|\, X_n(k) / |X_n(k)| \tag{15}$$

The speech signal is re-synthesized using IFFT. The signal segments corresponding to the modified complex spectra of the consecutive frames may have discontinuities due to modification of short-time Fourier transform involved in spectral subtraction, and overlap-add is used to mask them.

To examine the effect of the processing parameters, the technique was implemented using Matlab for offline processing. Implementation was carried out using magnitude subtraction (exponent factor $\gamma = 1$) as it showed higher tolerances to variation in the values of α and β [21]. Processing was carried out with sampling frequency of 10 kHz

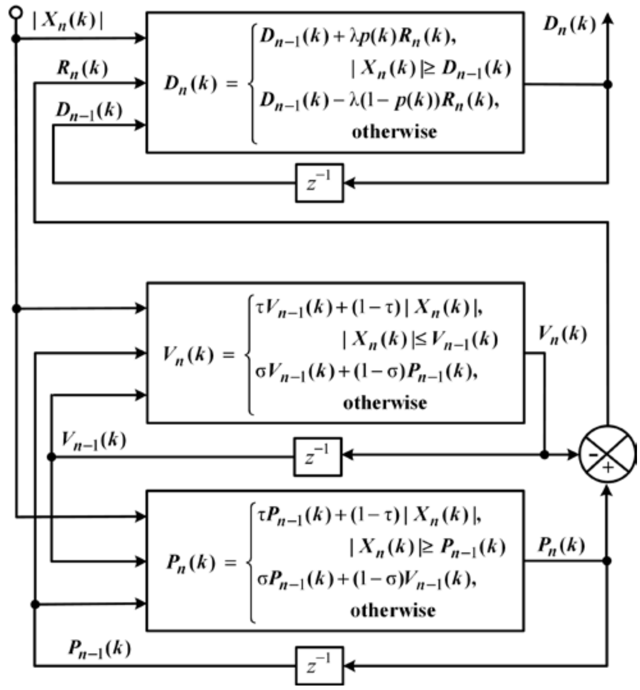

Fig. 1. Block diagram of the dynamic quantile tracking technique.

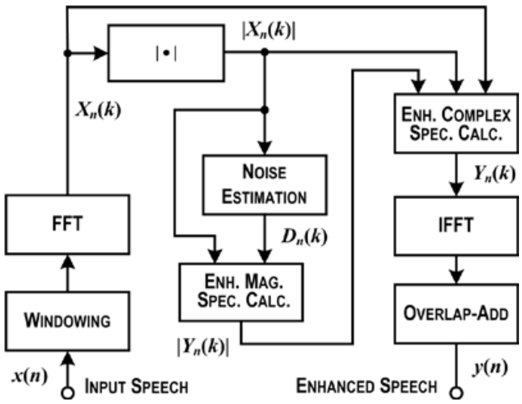

Fig. 2. Speech enhancement by spectral subtraction.

and window length of 25.6 ms (i.e. 256 samples) with 75% overlap. As the processed outputs with FFT length $N = 512$ and higher were indistinguishable, $N = 512$ was used. The offline processing was used to get an optimal combination of α and β for spectral subtraction, and that of τ and σ for noise estimation. These empirically obtained optimal values were used in the implementation for real-time processing.

III. Implementation for Real-time Processing

The spectral subtraction technique presented in the previous section needs to be implemented for real-time processing on a low-power DSP chip in order to use it in aids for the hearing impaired. The 16-bit fixed point processor TI/TMS320C5515 [22] is selected for this purpose. It has several features, including DMA-based I/O and on-chip hardware for 8 to 1024-point FFT, making it particularly suited for implementing our denoising technique for real-time

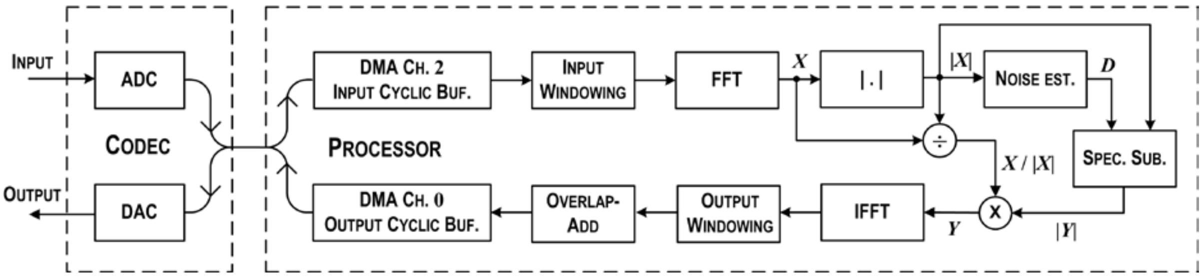

Fig. 3. Implementation of spectral subtraction on the DSP board [21]

processing. It has a maximum clock rate of 120 MHz. The implementation was carried out using DSP board "eZdsp" [23] with codec TLV320AIC3204 [24] supporting 16/20/24/32-bit stereo ADC and DAC with sampling frequency of 8 – 192 kHz. TI's 'CCStudio, ver. 4.0' was used as the development environment for programming in C. The implementation uses one channel of the codec, with 16-bit quantization and 10 kHz sampling.

Fig. 3 shows a block diagram of the implementation, with L-sample window and N-point FFT (L = 256, N = 512). It uses input-output operations similar to that described in [21]. ADC values are read into the input cyclic buffer through DMA channel-2 and output cyclic buffer values are written to DAC through channel-0 at the set sampling frequency. In order to reduce the conversion overheads, the input samples, spectral values, and the processed samples are all stored as 4-byte words with 16-bit real and 16-bit imaginary parts. Fig. 4 shows the input, output, data transfer, and buffering operations devised for an efficient realization of the processing with 75% overlap and zero padding. The input samples are acquired using a 5-block input cyclic buffer and the processed samples are output using a 2-block cyclic buffer, with S-word blocks and $S = L/4$. The current input and just-filled input blocks are tracked using pointers with cyclic values (.., 1, 2, 3, 4, 5, 1, ..) which are initialized to 1 and 5, respectively. Pointers with toggling values of 1 and 2 are used to track current output and write-to output blocks. They are initialized as 1 and 2, respectively. A DMA interrupt is generated when the current input block gets filled. All pointers are incremented cyclically. The DMA-mediated reading from ADC and writing to DAC are continued. The samples of the just-filled and the previous blocks are copied to the input data buffer, and are padded with $N-L$ zero-valued samples to serve as input to N-point FFT. The processing for noise estimation, spectral subtraction, and re-synthesis of output signal is implemented with due care to avoid overflows.

IV. Test Results

Informal listening and objective evaluation using perceptual evaluation of speech quality (PESQ) measure [25] were used for evaluation of the proposed technique. The PESQ score (scale: 0 – 4.5) is calculated from the difference between the loudness spectra of level-equalized and time aligned noise-free reference and test signals. The speech material consisted of a recording with three isolated vowels, a Hindi sentence, and an English sentence (-/a/-/i/-/u/– "aayiye

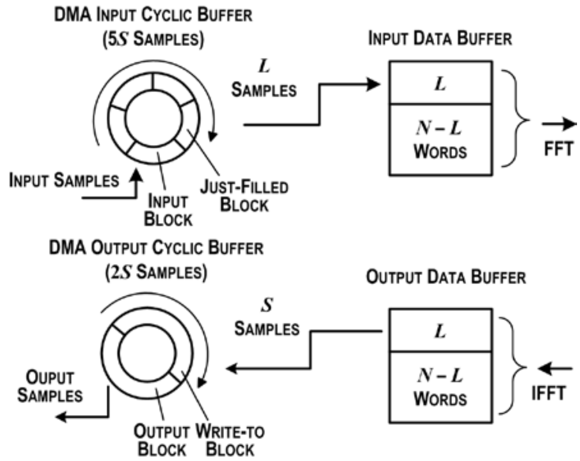

Fig. 4. Data transfer and buffering (S = L/4) [21].

aap kaa naam kyaa hai?" – "Where were you a year ago?") from a male speaker. A longer test sequence was generated by speech-speech-silence-speech concatenation of the recording for informal listening test. Testing involved processing of speech with additive white, street, babble, car, and train noises at SNR of 15, 12, 9, 6, 3, 0, –3, –6, –9, and –12 dB. The processing was carried out using τ = 0.1 and $\sigma = (0.9)^{1/1024}$ which corresponds to rise time of one frame shift and a fall time of 1024 frame shift.

To find the most suitable quantile for noise estimation and number of frames over which this quantile should be estimated, the offline processing was carried out using sample quantile. Informal listening test showed that the processing significantly enhanced the speech for all noises and there was no audible roughness. For objective evaluation of the processed outputs, PESQ scores were obtained for the processed output with β = 0, α in the range of 0.4 to 6, and with quantile p = 0.1, 0.25, 0.5, 0.75, and 0.9. The quantile values were obtained using previous M frames, where M = 32, 64, 128, 256, and 512. For fixed values of SNR, α, and p, the highest PESQ scores were obtained for M = 128. Lower values of M resulted in attenuation of speech signal and larger values were unable to track non-stationary noise. The investigations were repeated using dynamic quantile tracking. The PESQ scores of the processed output with convergence factor λ = 1/256 were found to be nearly equal to the PESQ scores obtained using sample quantile with M = 128. It was further observed that noise estimation with p = 0.25 resulted in nearly

Table 1. PESQ scores of the unprocessed (Unpr.) noisy speech with babble (a non-stationary noise) and processed (Pr.) signals with noise estimation by sample quantile (SQ) with M = 128 and dynamic quantile tracking (DQT) with λ = 1/256.

SNR (dB)	PESQ Score						
	Unpr.	Pr., α=1,β=0		Pr., α=2,β=0		Pr., α=3, β=0	
		SQ	DQT	SQ	DQT	SQ	DQT
-6	1.68	1.72	1.66	1.71	1.75	1.62	1.57
0	1.97	2.00	2.13	2.20	2.19	2.17	2.28
6	2.39	2.54	2.53	2.70	2.65	2.69	2.67

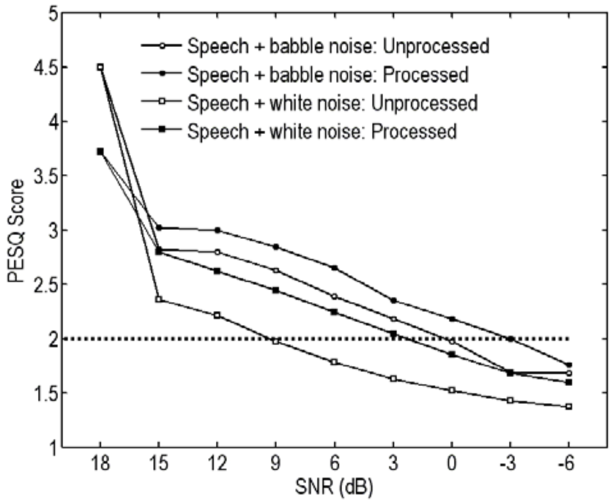

Fig. 5. PESQ score vs SNR for noisy and enhanced speech.

the best scores for different types of noises at all SNRs. Table 1 shows the PESQ scores for babble as an example of non-stationary noise. Further, the PESQ scores for the processed outputs obtained using 0.25-quantile were not very sensitive to changes in α. Thus the combination of λ = 1/256, p = 0.25, and α = 2 was used for a more detailed examination of the scores.

The PESQ score vs. SNR plots of unprocessed and processed signals for speech signal added with white and babble noises are shown in Fig. 5. For unprocessed speech, the score decreases with decrease in SNR. The scores for white noise were lower than the corresponding ones for babble noise. After processing, the scores of the noisy speech increased by 0.24 – 0.46 for white noise and by 0.08 – 0.32 for babble noise. For a score of 2 (generally considered as lowest score for acceptable speech), processing resulted in SNR advantage of approximately 6 dB for white noise and 3 dB for babble noise. SNR advantage for other types of noise was between these two values. Informal listening showed that spectral floor factor β = 0.001 reduced the musical noise without degrading the speech quality.

The real-time processing was tested using speech mixed with white, babble, car, street, and train noises at different SNRs. A PC sound card was used to acquire the processed output signal from DSP board. Fig. 6 shows an example of processing showing the noise-free speech, noisy speech with white noise at SNR of 3 dB, output from offline processing,

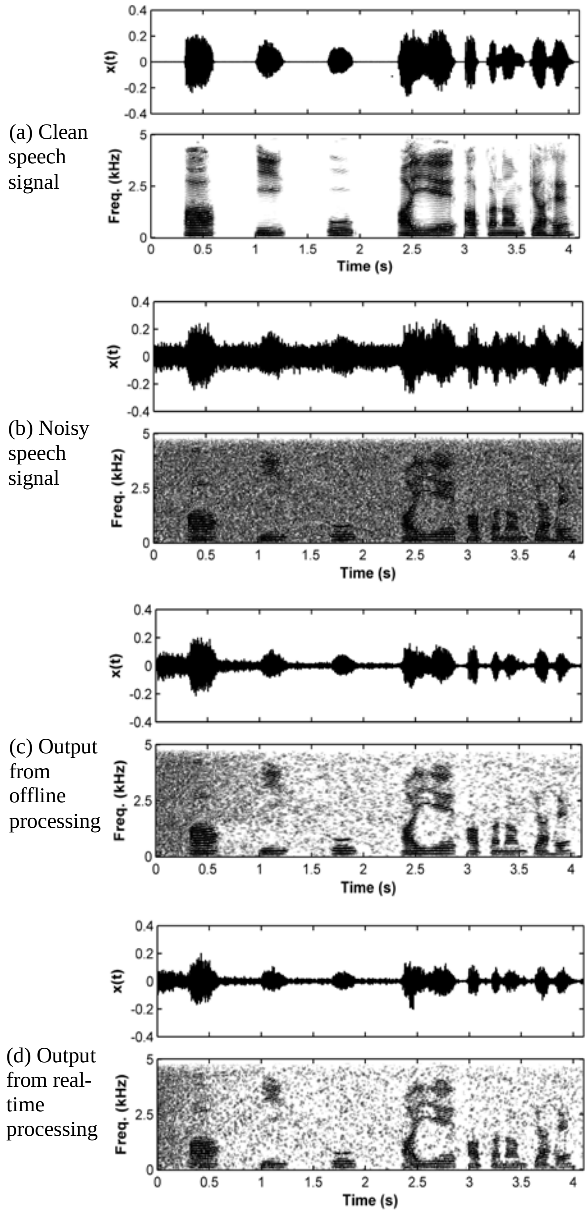

Fig. 6. Processing of the sequence ("-/a/-/i/-/u/– "aayiye aap kaa naam kyaa hai?" – "Where were you a year ago?"", from a male speaker) with white noise at SNR of 3 dB: signals and spectrograms [26].

and output from real-time processing [26]. The output of the real-time processing was found to have a close match with the corresponding output of offline processing. The match was also confirmed by high PESQ scores (greater than 3.5) for real-time processing with offline processing as the reference. Total signal delay (consisting of algorithmic delay, computation delay, and input-output delay) was found to be approximately 36 ms which may be considered as acceptable for its use in the hearing aids along with lipreading.

For an empirical estimation of the capacity of the processor needed for implementing the proposed denoising technique, the clock frequency of the processor was progressively decreased from its maximum value of 120 MHz and the output was examined for satisfactory execution of the code. For comparison, the processing was also implemented without noise estimation (zero-valued spectral samples for the estimated noise and the code for noise estimation bypassed) and cascaded-median based noise estimation as reported in [21]. The minimum clock frequencies needed for processing with bypassed noise estimation, cascaded-median based noise estimation, and dynamic-quantile-tracking based noise estimation were 38, 45, and 50 MHz, respectively, indicating a requirement of approximately 32%, 38% and 41% of the processor capacity. Thus the results show that the proposed technique for dynamic quantile tracking can be used for noise estimation with only a marginal increase in the processing capacity as required for cascaded-median based noise estimation. As the proposed processing needs only 41% of the available capacity, the rest can be used in implementing other processing as needed for a hearing aid.

V. CONCLUSION

A speech enhancement technique for suppressing stationary and non-stationary background noise using spectral subtraction and noise spectrum estimation without voice activity detection has been presented. The noise spectrum estimation is based on dynamic quantile tracking without involving storage and sorting of past samples. Test results using the implementation for offline processing of noisy speech with different types of additive stationary and non-stationary noises showed that use of 0.25-quantile worked well for all of them. The improvement in speech quality due to the enhancement was equivalent to SNR advantage of 3 – 6 dB. The technique has been implemented on the 16-bit fixed-point processor TI/TMS320C5515 for sampling frequency of 10 kHz. The real-time processing results in a signal delay of 36 ms and uses about 41% of the processor capacity, indicating the potential of its use in hearing aids.

The technique permits use of a different quantile at each frequency bin for noise estimation. Use of frequency dependent quantile values may further improve the performance of the technique without introducing any processing overheads. The proposed speech enhancement technique may be combined with other signal processing techniques used in the hearing aids and tested for improving perception of different speech materials by the hearing-impaired listeners. Its implementation using other processors may also be investigated.

REFERENCES

[1] H. Levitt, J. M. Pickett, and R. A. Houde (eds.), *Senosry Aids for the Hearing Impaired.* New York: IEEE Press, 1980.

[2] J. M. Pickett, *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Boston, Mass.: Allyn Bacon, 1999, pp. 289–323.

[3] H. Dillon, *Hearing Aids*. New York: Thieme Medical, 2001.

[4] T. Lunner, S. Arlinger, and J. Hellgren, “8-channel digital filter bank for hearing aid use: preliminary results in monaural, diotic, and dichotic modes,” *Scand. Audiol. Suppl*., vol. 38, pp. 75–81, 1993.

[5] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, “Binaural dichotic presentation to reduce the effects of spectral masking in moderate bilateral sensorineural hearing loss,” *Int. J. Audiol*., vol. 51, no. 4, pp. 334–344, 2012.

[6] J. Yang, F. Luo, and A. Nehorai, “Spectral contrast enhancement: Algorithms and comparisons,” *Speech Commun*., vol. 39, no. 1–2, pp. 33–46, 2003.

[7] T. Arai, K. Yasu, and N. Hodoshima, “Effective speech processing for various impaired listeners,” in *Proc. 18th Int. Cong. Acoust*., 2004, Kyoto, Japan, pp. 1389–1392.

[8] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, “Multi-band frequency compression for improving speech perception by listeners with moderate sensorineural hearing loss,” *Speech Commun*., vol. 54, no. 3 pp. 341–350, 2012.

[9] A. R. Jayan and P. C. Pandey, “Automated modification of consonant-vowel ratio of stops for improving speech intelligibility,” *Int. J. Speech Technol.*, 2014, [online] DOI 10.1007/s10772-014-9254-4.

[10] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” in *Proc. IEEE ICASSP 1979*, Washington, D.C., pp. 208-211.

[11] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113-120, 1979.

[12] P. C. Loizou, *Speech Enhancement: Theory and Practice*. New York: CRC, 2007.

[13] Y. Lu and P. C. Loizou, “A geometric approach to spectral subtraction,” *Speech Commun.*, vol. 50, no. 6, pp. 453-466, 2008.

[14] K. Paliwal, K. Wójcicki, and B. Schwerin, “Single-channel speech enhancement using spectral subtraction in the short-time modulation domain,” *Speech Commun.*, vol. 52, no. 5, pp. 450–475, 2010.

[15] R. Martin, “Spectral subtraction based on minimum statistics,” in *Proc. 6th Eur. Signal Process. Conf. (EUSIPCO 1994)*, Edinburgh, U.K., 1994, pp. 1182-1185.

[16] I. Cohen, “Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging,” *IEEE Trans. Speech Audio Process*., vol. 11, no. 5, pp. 466-475, 2003.

[17] G. Doblinger, “Computationally efficient speech enhancement by spectral minima tracking in subbands,” in *Proc. EUROSPEECH 1995*, Madrid, Spain, pp. 1513-1516.

[18] V. Stahl, A. Fisher, and R. Bipus, “Quantile based noise estimation for spectral subtraction and Wiener filtering,” in *Proc. IEEE ICASSP 2000*, Istanbul, Turkey, pp. 1875-1878.

[19] N. W. Evans and J. S. Mason, "Time-frequency quantile-based noise estimation," *in Proc. 11th Eur. Signal Process. Conf. (EUSIPCO 2002)*, Toulouse, France, 2002, pp. 539-542.

[20] H. Bai and E. A. Wan, "Two-pass quantile based noise spectrum estimation," Center of spoken language understanding, OGI School of Science and Engineering at OHSU (2003), [online] Available: http://speech.bme.ogi.edu/publications/ps/bai03.pdf.

[21] S. K. Waddi, P. C. Pandey, and N. Tiwari, “Speech enhancement using spectral subtraction and cascaded-median based noise estimation for hearing impaired listeners,” in *Proc. 19th Nat. Conf. Commun. (NCC 2013)*, Delhi, India, 2013, paper no. 1569696063.

[22] Texas Instruments, Inc., “TMS320C5515 Fixed-Point Digital Signal Processor,” 2011, [online] Available: focus.ti.com/lit/ds/symlink/ tms320c5515.pdf.

[23] Spectrum Digital, Inc., “TMS320C5515 eZdsp USB Stick Technical Reference,” 2010, [online] Available: support.spectrumdigital.com/ boards/usbstk5515/reva/files/usbstk5515_TechRef_RevA.pdf

[24] Texas Instruments, Inc., “TLV320AIC3204 Ultra Low Power Stereo Audio Codec,” 2008, [online] Available: focus.ti.com/lit/ds/ symlink/tlv320aic3204.pdf.

[25] ITU, “Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” *ITU-T Rec*., P.862, 2001.

[26] N. Tiwari, “Speech enhancement using noise estimation based on dynamic quantile tracking for hearing impaired listeners: Processing results”, 2015, [online] Available: www.ee.iitb.ac.in/~spilab/material /nitya/ncc2015.