Speech Enhancement Using Noise Estimation with Adaptive Dynamic Quantile Tracking for Use in Hearing Aids

Nitya Tiwari

Dept. of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India
nitya@ee.iitb.ac.in

Abstract

Background noise significantly degrades speech perception for persons with sensorineural hearing impairment. A single-input speech enhancement using adaptive dynamic quantile tracking for estimating the noise spectrum is presented for improving performance of hearing aids. A histogram for each spectral sample is estimated by dynamically tracking ten quantiles and the histogram peak is used as the adaptive quantile for estimating the noise at each spectral sample. It does not involve storage and sorting of past spectral samples for finding the quantiles and thus is suitable for real-time processing. Speech enhancement is carried out using the geometric approach-based spectral subtraction. 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 algorithmic and computational delays introduced by the processing are acceptable for face-to-face communication.

Index Terms: noise suppression, speech enhancement, hearing aids

1. Introduction

Sensorineural hearing loss significantly reduces the dynamic range of hearing and results in abnormal loudness growth [1]-[3]. The hearing aids for compensating sensorineural loss provide frequency-selective amplification along with dynamic range compression with the objective of presenting all the sounds comfortably within the limited dynamic range of the listener [4]-[5]. Sensorineural loss is usually associated with increased temporal and spectral masking, leading to difficulty in speech perception, particularly in noisy environments. Therefore, background noise needs to be suppressed for improving speech quality and intelligibility for hearing impaired listeners.

The noise suppression technique for use in a hearing aid should have low algorithmic delay and low computational complexity. It involves estimating the noise spectrum, removal of estimated noise using a suppression rule, and re-synthesizing the speech signal. Dynamic estimation of the noise spectrum is important for effective noise suppression. Under-estimation results in excessive residual noise and over-estimation results in perceptible distortion, leading to degraded quality and poor intelligibility. Use of voice activity detection [6] for tracking noise during silence intervals may not work satisfactory under low-SNR conditions and during long speech segments. Several techniques based on statistical properties of speech and noise have been reported for estimating noise spectrum without voice activity detection [7]–[13].

Minimum statistics based noise estimation techniques [7]–[9] have low computational complexity, but they often under-estimate the noise and need estimation of an SNR-dependent over-subtraction factor. It has been earlier reported in the literature that a quantile-based dynamic estimation of the noise spectrum from the spectrum of the noisy speech without using a voice activity detector can be used for noise suppression. These techniques use a quantile of the noisy speech spectral sample as the noise estimate [10]–[13]. They are based on the observation that the speech energy in a particular frequency bin is low in most of the frames and high only in 10−20% frames. Use of frequency-dependent quantiles is needed for estimation of non-stationary noises [13]−[14]. Several histogram-based techniques that estimate noise as the maximum of the distribution of energy values in each frequency bin have been reported [14]−[16]. The quantile-based techniques are not suitable for use in hearing aids due to large memory requirement and high computational complexity involved in storing and sorting the spectral samples. The histogram-based techniques pose even higher implementation challenges as they require estimation of multiple quantiles.

The research objective is to develop quantile-based noise estimation technique for use in hearing aids. As quantile values are to be estimated in real-time for each frequency bin, the quantile tracking technique should have low computational complexity and low storage requirement. Towards, this objective a technique for noise spectrum estimation based on dynamic quantile tracking as an approximation to the sample quantile, without involving storage and sorting of past samples was developed earlier [17], [18]. An improved noise estimation technique that selects the quantile adaptively is presented. Histogram is tracked dynamically and its peak is used as the adaptive quantile for estimating the noise at each spectral sample for speech enhancement. The proposed technique, test results, and future work are presented in following subsections.

2. Signal processing technique

2.1. Estimation of noise spectrum

The most frequent energy value, obtained as the maximum of histogram, in individual frequency bins is reported to be related to the noise level in the specified frequency bins [14]. The proposed noise estimation technique dynamically estimates histogram and the histogram peak is used as the adaptive quantile for estimating the noise at each spectral sample. The histogram is estimated by dynamically tracking multiple quantile values (q_1, q_2, ... q_M) for a set of evenly spaced probabilities (p_1, p_2, ... p_M). The desired quantile corresponding to the peak of the histogram is obtained by finding p_i for which the difference between neighboring quantile values is minimum. The estimate of noise spectrum, $D(n, k)$, at nth frame and kth frequency bin is obtained as

$$D(n,k) = \arg\min_{q_i(n,k)}[q_i(n,k) - q_{i-1}(n,k)];\ i = 2,3,...,M \quad (1)$$

For estimating each of the quantiles, we use a previously reported computationally efficient technique, named as dynamic quantile tracking using range estimation [17], [18]. In this technique, an estimate of the quantile of a data stream is obtained without storage and sorting of past samples. The quantile $q_i(n, k)$ is estimated by applying an increment or a decrement on its previous estimate, selected to be a fraction 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.

At kth spectral sample, $q_i(n, k)$ is tracked as the $p_i(k)$-quantile of the magnitude spectrum $|X(n, k)|$ as

$$q_i(n,k) = q_i(n-1,k) + d_i(n,k) \tag{2}$$

The change $d_i(n, k)$ is given as

$$d_i(n,k) = \begin{cases} \Delta_i^+(k), & |X(n,k)| \geq q_i(n-1,k) \\ -\Delta_i^-(k), & \text{otherwise} \end{cases} \tag{3}$$

The values of $\Delta_i^+(k)$ and $\Delta_i^-(k)$ should be such that the ratio $\Delta_i^+(k)/\Delta_i^-(k) = p_i(k)/(1-p_i(k))$ to ensure that the quantile estimate approaches the sample quantile and sum of the changes in the estimate approaches zero, i.e. $\sum d_i(n, k) \approx 0$. Therefore $\Delta_i^+(k)$ and $\Delta_i^-(k)$ may be selected as

$$\Delta_i^+(k) = \lambda p_i(k) R(n, k) \tag{4}$$

$$\Delta_i^-(k) = \lambda(1 - p_i(k)) R(n, k) \tag{5}$$

where R is the range (difference between the maximum and minimum values of the sequence of spectral values in a particular frequency bin). λ is a convergence factor which controls the step size during tracking and it is selected for tradeoff between ripple in the estimated quantile value and the number of steps needed for convergence as described in [17].

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:

$$P(n,k) = \begin{cases} \tau P(n-1,k) + (1-\tau)|X(n,k)|, & |X(n,k)| \geq P(n-1,k) \\ \sigma P(n-1,k) + (1-\sigma)V(n-1,k), & \text{otherwise} \end{cases} \tag{6}$$

$$V(n,k) = \begin{cases} \tau V(n-1,k) + (1-\tau)|X(n,k)|, & |X(n,k)| \leq V(n-1,k) \\ \sigma V(n-1,k) + (1-\sigma)P(n-1,k), & \text{otherwise} \end{cases} \tag{7}$$

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

$$R(n, k) = P(n, k) - V(n, k) \tag{8}$$

The dynamic quantile tracking to estimate quantile $q_i(n, k)$ as given by (2), (3), and (8) can be written as the following:

$$q_i(n,k) = \begin{cases} q_i(n-1,k) + \lambda p_i(k) R(n,k), & |X(n,k)| \geq q_i(n-1,k) \\ q_i(n-1,k) - \lambda(1-p_i(k)) R(n,k), & \text{otherwise} \end{cases} \tag{9}$$

To estimate the histogram, quantiles coressponding to the set of probabilities (p_1, p_2, ... p_M), are obtained using (9) with a common range tracked using (6), (7), and (8).

2.2. Speech enhancement by spectral subtraction

Geometric approach based spectral subtraction [19] is used for suppression of background noise, as this technique results in negligible residual noise. The processing consists of noise spectrum estimation, enhanced magnitude spectrum calculation, and estimating the enhanced complex spectrum without explicit phase estimation.

Table 1: *Improvement in PESQ scores*

Noise	SNR	Unprocessed		Improvement	
		Mean	S. D.	Mean	S. D.
Airport	6	2.21	0.15	0.29	0.16
	3	2.01	0.17	0.33	0.16
	0	1.81	0.18	0.35	0.18
Babble	6	1.96	0.13	0.26	0.14
	3	1.78	0.15	0.23	0.20
	0	1.61	0.19	0.17	0.24
Street	6	2.28	0.15	0.27	0.15
	3	2.08	0.17	0.30	0.16
	0	1.86	0.19	0.35	0.17

3. Test results

The proposed technique was implemented for offline processing using MATLAB using sampling frequency of 10 kHz, window length of 25.6 ms with 75% overlap, and FFT length of 512. The histogram is dynamically tracked for each frequency bin by estimating ten quantiles for p = 0.25, 0.30, 0.35, ... , 0.75 using $\lambda = 1/256$, $\tau = 0.1$, and $\sigma = (0.9)^{1/1024}$. Informal listening and objective evaluation using perceptual evaluation of speech quality (PESQ) measure (scale: 0 – 4.5) [20] were used for evaluation, with sentences from NOIZEUS database [21] as speech material. Testing involved processing of speech with additive noises from AURORA database [22]. PESQ scores were obtained for processed outputs for various noises. The SNR improvements at PESQ score of 2 (considered as lowest score for acceptable speech quality) are as 6, 3, 3, 5, and 5 dB for airport, babble, car, street, and whites. Mean improvement in PESQ scores and corresponding standard deviations are given in Table II. The improvements in PESQ score are significant with $p<0.001$.

A smartphone app [23] was developed earlier with processing for dynamic range compression on Nexus 5X with Android 7.1 OS. The noise suppression technique has been incorporated as part of this app. The outputs from the app and the offline processing showed no perceptible differences. The audio latency of the app, measured using DSO and using a 1 kHz tone burst of 200 ms as the input, was found to be 45 ms, making it suitable for face-to-face communication [24].

4. Future work

Investigations are to be carried out for improving the performance of noise suppression by reducing the residual noise, musical noise, and speech distortions. The performance of the noise estimation technique needs to be compared with existing techniques using objective measures and by conducting listening tests on normal-hearing subjects. The 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.

5. Acknowledgements

I would like to express my sincere gratitude towards my supervisor Prof. P. C. Pandey for his invaluable guidance and support. I would also like to thank Saketh Sharma for smartphone app based implementation of the work and for sharing interesting discussions with me and providing support whenever needed.

6. 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] R. E. Sandlin, Textbook of Hearing Aid Amplification, San Diego, Cal.: Singular 2000, pp. 210–220.

[5] L. D.Braida, N. I. Durlach, R. P. Lippmann, B. L. Hicks, W. M. Rabinowitz, and C. M. Reed, "Hearing aids–a review of past research on linear amplification, amplitude compression, and frequency lowering," *Journal of the American Speech and Hearing Association Monographs*, vol. 19, pp. 1–114, 1979.

[6] D. Malah, R. V. Cox, and A. J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in nonstationary noise environments," in *Proc. IEEE ICASSP 1999*, Phoenix, Arizona, 1999, pp. 789-792.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] H. Hirsch, C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in *Proc. IEEE ICASSP 1995*, Detroit, Michigan, 1995, pp. 153-156.

[15] R. J. MacAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust. Speech Sig. Process*., vol. 28, no. 2, pp. 137-145, 1980.

[16] C. Ris and S. Dupont, "Assessing local noise level estimation methods: application to noise robust ASR," Speech Commun., vol. 34, no. 1-2, pp. 141–158, 2001.

[17] N. Tiwari and P. C. Pandey, "Speech enhancement using noise estimation based on dynamic quantile tracking for hearing impaired listeners," in *Proc. 21st Nat. Conf. Commun. (NCC 2015)*, Mumbai, India, 2015, paper no. 1570056299.

[18] P. C. Pandey and N. Tiwari, "Method and system for suppressing noise in speech signals in hearing aids and speech communication devices, US Patent application publication no. US20170032803 A1, 2017.

[19] Y. Lu and P. C. Loizou, "A geometric approach to spectral subtraction," *Speech Commun*., vol. 50, pp. 453–466, 2008.

[20] 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.

[21] Y. Hu and P. Loizou, "Subjective comparision of speech enhancement algorithms," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.,* Toulouse, France, 2006, pp. 153-156.

[22] H. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *Proc. ISCA Tutorial and Research Workshop on Automatic Speech Recognition: Challenges for the New Millenium 2000*, (ASR 2000), Paris, France, pp. 181–188.

[23] S. Sharma, N. Tiwari, and P. C. Pandey, "Implementation of a digital hearing aid with user-settable frequency response and sliding-band dynamic range compression as a smartphone app," in *Proc. Int. Conf. Intelligent Human Computer Interaction 2016 (IHCI 2016)*, Pilani, India, Dec. 12 - 13, 2016, paper no. 81.

[24] *International Telecommunication Union: Relative timing of sound and vision for broadcasting*, ITU Rec. ITU-R BT.1359 1998.