# Enhancement of Electrolaryngeal Speech by Reducing Leakage Noise Using Spectral Subtraction with Quantile Based Dynamic Estimation of Noise

*Prem C. Pandey, Santosh S. Pratapwar, and Parveen K. Lehana*

Department of Electrical Engineering
Indian Institute of Technology, Bombay, India
{pcpandey, santosh, lehana}@ee.iitb.ac.in

## Abstract

Transcervical electrolarynx is a vibrator held against the neck tissue in order to provide excitation to the vocal tract, as a substitute to that provided by a natural larynx. It is of great help in verbal communication to a large number of laryngectomy patients. Its intelligibility suffers from the presence of a background noise, caused by leakage of the acoustic energy from the vibrator. Pitch synchronous application of spectral subtraction method, normally used for enhancement of speech corrupted by uncorrelated random noise, can be used for reduction of the self leakage noise for enhancement of electrolaryngeal speech. Average magnitude spectrum of leakage noise, obtained with lips closed, is subtracted from the magnitude spectrum of the noisy speech and the signal is reconstructed using the original phase spectrum. However, the spectrum of the leakage noise varies because of variation in the application pressure and movement of the throat tissue. A quantile based dynamic estimation of the magnitude spectrum without the need for silence/voice detection was found to be effective in noise reduction.

## 1. Introduction

In normal speech production, the lungs provide the air stream, the vocal chords in the larynx provide the vibration source for the sound, and the vocal tract provides the spectral shaping of the resulting speech [1]. Laryngeal cancer often necessitates surgical removal of larynx. The patient (often known as a laryngectomee), needs external aids to communicate. An artificial larynx [2],[3] is a device used to provide excitation to the vocal tract, as a substitute to that provided by a natural larynx. The external electronic larynx or the transcervical electrolarynx is the widely used type of device, which is hand held and pressed against the neck. It consists of an electronic vibration generator. The vibrations are transmitted through the neck tissue to the vocal tract. Spectral shaping of the waveform by the vocal tract results in speech. The device is easy to use and portable, however the speaker needs to control the pitch and volume switches to prevent monotonic speech, and this needs practice. The speech produced is generally deficient in low frequency energy due to lower coupling efficiency through the throat tissue [4]. The unvoiced segments generally get substituted by the voiced segments. In addition to these, the major problem is that the speech output has a background noise, caused by leakage of the acoustic energy from the vibrator, which degrades the quality and intelligibility of the output speech considerably [5],[6],[7],[8],[9].

## 2. Electrolaryngeal speech

A model of the leakage sound generation during the use of transcervical larynx [8] is shown in Fig.1. The vibrations generated by the vibrator diaphragm have two paths. The first path is through the neck tissue and the vocal tract. Its impulse response $h_v(t)$ depends on the length and configuration of the vocal tract, the place of coupling of the vibrator, the amount of coupling, etc. Excitation $e(t)$ passing through this path results in speech signal $s(t)$. The second path of the vibrations is through the surroundings, and this leakage component $l(t)$ gets added to the useful speech $s(t)$, and deteriorates its intelligibility. Signal processing techniques can be used for reduction of noise by estimating noise present in the signal and subtracting it from the noisy signal. The main problem in noise subtraction is that the speech and noise, resulting from the same excitation, as shown in Fig.1, are highly correlated.

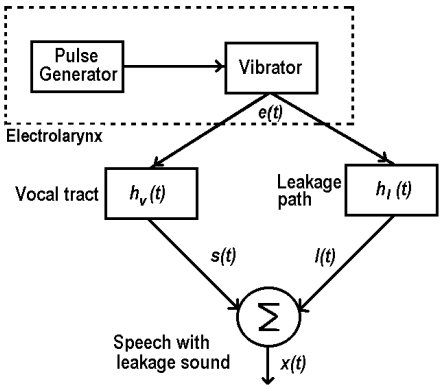


*Figure 1*: Model of self leakage or background noise generation in transcervical electrolarynx [8].

Epsy-Wilson *et al.* [5] reported a technique for enhancement of electrolaryngeal speech using two-input LMS algorithm. Authors have reported that by carrying out noise adaptation during non-sonorant or low energy segments, the noise cancellation was effective, and most

of leakage noise was cancelled. During the sonorant sounds, there was an improvement in the output quality, though the leakage noise was not removed fully. Processing resulted in improvement in speech intelligibility.

We have earlier reported [8] a single-input noise cancellation technique based on spectral subtraction applied in a pitch synchronous manner. In this technique, the noise spectrum is estimated by averaging the noise spectra over several segments of the self-leakage noise acquired with speaker keeping the lips closed. Because of variations in the noise characteristics, effective cancellation requires frequent acquisitions of noise. We have applied quantile based noise spectrum estimation (QBNE) for continuous updating of noise spectrum [12],[13]. In this paper, we are presenting results of investigations with various types of QBNE based noise spectra for spectral subtraction.

## 3. Spectral subtraction for reducing leakage noise

Spectral subtraction technique is one of the important techniques for enhancing noisy speech [10],[11]. The basic assumption in this technique is that the clean speech and the noise are uncorrelated, and therefore the power spectrum of the noisy speech equals the sum of power spectra of noise and clean speech. In case of electrolaryngeal speech, speech signal and leakage interference are not uncorrelated. It has been shown earlier [8] that, considering the impulse response of the vocal tract and leakage path to be uncorrelated, if the short-time spectra are evaluated using pitch synchronous window, power spectrum of noisy speech will be sum of power spectrum of noise and speech, and spectral subtraction can be employed.

Mean squared spectrum averaged over non-speech segments can be used for spectral subtraction during the noisy speech segments. For implementation of the technique, squared magnitudes of the FFT of a number of adjacent windowed segments in non-speech duration are averaged to get the mean squared noise spectrum. During speech, the noisy speech is windowed by the same window as in earlier mode, and its magnitude and phase spectra are obtained. The phase spectrum is retained for resynthesis. From the squared magnitude spectrum, the mean squared spectrum of noise, determined during the noise estimation mode, is subtracted. The resulting magnitude spectrum is combined with the earlier phase spectrum, and its inverse FFT is taken as the clean speech signal during the window duration.

Assumptions regarding impulse response of vocal tract and impulse response of leakage path being uncorrelated may be valid over long period, but not necessarily over short segments. This may result in some of the frequency components becoming negative, causing narrow random spikes during non-speech segment, known as residual noise. In order to reduce it, modified spectral subtraction method [11] as shown in schematically in Fig.2. is used. The processing parameters involved are subtraction factor $\alpha$, spectral floor factor $\beta$, and exponent factor $\gamma$, and these need to be empirically selected for an output free from broadband as well as residual noise. Optimal values for reduction of leakage noise as reported earlier [8] are: window length = twice the pitch period, spectral subtraction factor $\alpha \approx 2$, spectral floor factor $\beta \approx 0.001$, $\gamma = 1$.

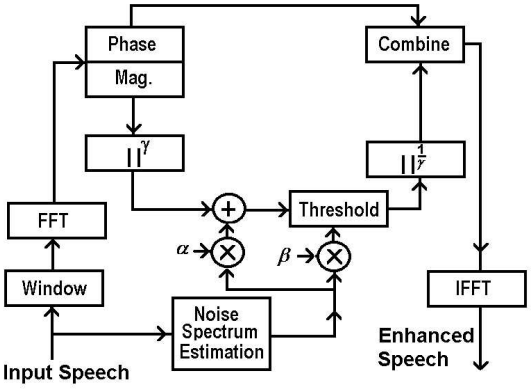


*Figure 2*: Spectral subtraction scheme [8].

A speech processor based on this technique has a noise estimation mode during which speaker keeps the lips closed and the acquired signal consists of only the self leakage noise. After this, the device automatically switches to speech enhancement mode: the earlier estimated noise spectrum is used in noise subtraction. The noise spectrum is taken to be constant over the entire duration of enhancement mode. But actually the leakage noise varies because of variations in the place of coupling of vibrator to the neck tissue and the amount of coupling. This results in variations in the effectiveness of noise enhancement over extended period. Hence a continuous updating of the estimated noise spectrum is required. Recursive averaging of spectra during silence segments may be used for noise spectrum estimation [10],[11]. However, speech/silence detection in electro-laryngeal speech is rather difficult. Quantile-based noise estimation (QBNE) technique [12],[13] does not need speech/non-speech classification and we have investigated its use for noise estimation in electro-laryngeal speech.

## 4. Investigation with QBNE

Quantile-based noise estimation (QBNE) makes use of the fact that even during speech periods, frequency bins tend not to be permanently occupied by speech i.e. exhibit high energy levels [12],[13]. Speech/non-speech boundaries are detected implicitly on a per-frequency bin basis, and noise spectrum estimates are updated throughout speech/non-speech periods. Further investi-

gations have reported using time-frequency quantiles [14] and signal-to-noise ratio based quantiles [15]

The degraded signal may be analyzed on a frame-by-frame basis, to obtain an array of the magnitude spectral values for each frequency sample, for a certain number of the past frames. Sorting of magnitude values in this array may be used for obtaining a particular quantile value. Accuracy of noise estimation and the rate at which QBNE reacts to changes in the noise depend on the number of past frames used. The buffer for all the frequency samples have to be reconstructed and resorted at each frame and this is computationally expensive. For a faster processing, an efficient indexing algorithm [16] was implemented for obtaining quantile based noise estimation continuously [17].

The recordings were done with the microphone positioned at the center between the mouth and the artificial larynx position. During first 2 s, speaker kept the lips closed, and the recorded speech contained only the leakage noise. This segment was used for training the QBNE method. Following methods were investigated for selection of quantile values for each frequency sample.

*Single quantile value:* Quantile value was selected for the best visual match between the quantile derived spectrum and the averaged noise spectrum.

*Two-band quantile values:* It was found that a better match could be obtained by partitioning the spectrum into two at about 900 Hz and selecting one quantile value for each band. The spectral subtraction was found to be better [17].

*Frequency dependent quantile values:* The averaged noise spectrum can be matched over the entire band to the quantile derived spectrum by having frequency dependent quantile values. Sorted array of power spectrum of noisy speech for each frequency was formed. Index of array of each frequency sample with value nearest to the averaged noise spectrum was selected as optimum quantile value. The quantile values as a function of frequency were smoothened by a 9-point symmetrical averaging.

*SNR based dynamic selection of quantile values*: The leakage noise characteristics change slowly with the method of application of the vibrator and the vocal tract configuration. It was observed that the spectral subtraction based on fixed quantile values was less effective during weak and non-speech segments. A dynamic selection of quantile values based on signal strength was investigated. It can be relatively more effective in case of long pauses in speech. A plot of signal-plus-noise to noise ratio (SNR), for recordings with different applications of the vibrator, and corresponding frequency dependent quantile values showed a large SNR dependence. SNR used here is the ratio of averaged spectrum of noisy speech to the averaged spectrum of noise, and hence it is 0 dB for the noise part. Frequency dependent quantile values were obtained for the noise segment and for the noisy speech segment. A dynamic estimation of frequency dependent quantile values was carried out by linear interpolation based on the short-time SNR (dB) in the analysis frame.

## 5. Test results

QBNE technique was used for speech enhancement of electrolaryngeal speech, recorded with 16-bit quantization and 11.025 kSa/s sampling rate. Electrolarynx NP-1 (manufactured by NP Voice, India) was used for this purpose. The vibrator of the electrolarynx had a fixed pitch of 90.3 Hz, i.e. pitch period of 122 samples. The degraded signal was analyzed on frame-by-frame basis, with frame size of twice pitch period i.e. 22.1 ms, with 50 % overlap.

Fig.3. shows a recording of a question-answer pair and enhancements carried out with different estimates of noise. All the enhancement results shown here were obtained using $\alpha = 1.4$, $\beta = 0.001$, and $\gamma = 1$. In the unprocessed speech (Fig. 3a), the initial 2 s segment was the leakage noise (sound produced with the speaker's mouth closed). The subsequent 2.4 s segment was the noisy speech. Enhancement using spectral subtraction (Fig. 3b-f) showed a significant reduction in noise. Listening indicated an improvement in intelligibility and quality for all the methods. Compared to averaged estimate of noise, single quantile noise estimation showed better speech quality. However, the residual noise during non-speech segments was higher. Frequency dependent quantile values show better overall reduction. SNR based dynamic estimation of quantile noise resulted in similar intelligibility and quality, but it was more effective during long pauses.

## 6. Conclusion

We have earlier reported [8] application of spectral subtraction in a pitch synchronous manner for enhancement of electrolaryngeal speech, by using an averaged estimate of leakage. Quantile based noise estimation was later applied for obtaining a continuously updated estimate of noise spectrum without speech vs non-speech detection [15]. In this paper, we have presented an investigation involving different quantile based estimates of noise spectrum for providing a significant reduction in leakage noise without appreciable reduction or clipping of signal. It is found that noise spectrum estimated with matched quantile values smoothened along frequency samples meets this requirement. A dynamic estimation of quantile values based on signal strength results in perceptually similar improvement in quality and intelligibility, and is particularly effective during long non-speech segments (pauses). Further, it adapts to changes in the application of the vibrator. More listening tests are needed for determining the improvement in quality and intelligibility brought about by this enhancement technique.

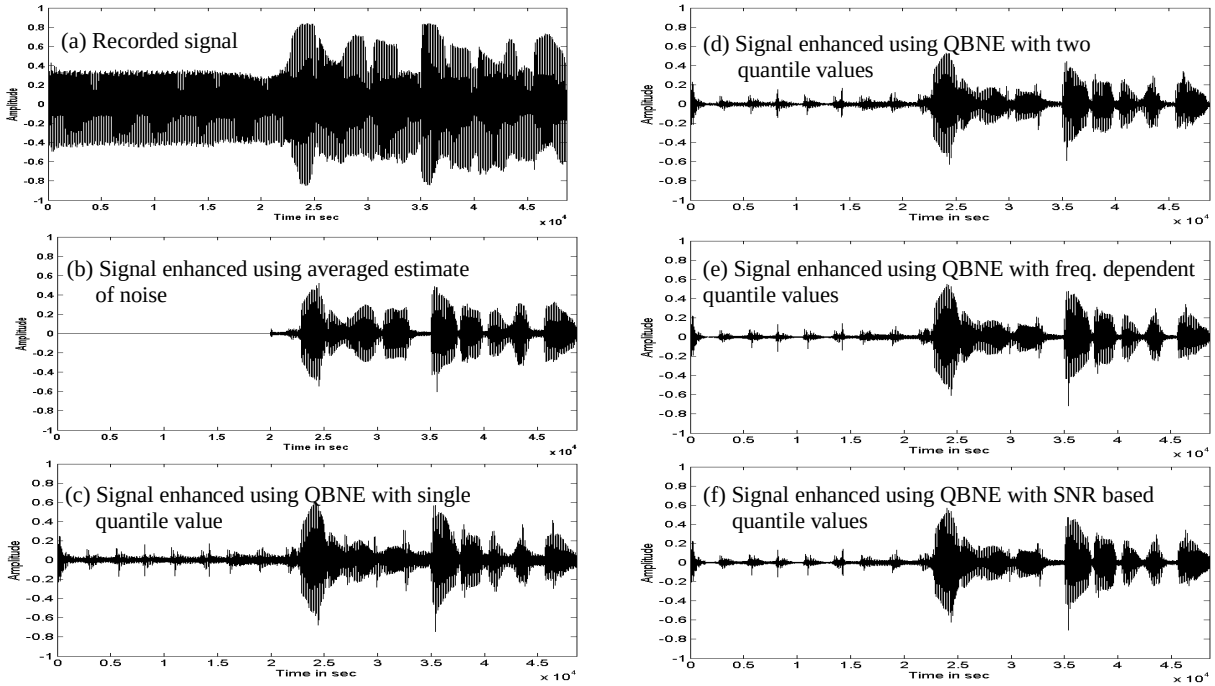


*Figure 3*: Recorded and enhanced speech with five different estimates of noise spectrum. Speaker SP, material: question-answer pair in English, "*What is your name? My name is Santosh*".

## 7. References


[1] Rabiner, L. R., and Schafer R. W., *Digital Processing of Speech Signals*, Prentice Hall, Englewood Cliffs, New Jersey, 1978.

[2] Lebrun, Y., "History and development of laryngeal prosthetic devices", *The Artificial Larynx*, Swets and Zeitlinger, Amsterdam, 1973, pp. 19-76.

[3] Goldstein, L. P., "History and development of laryngeal prosthetic devices", *Electrostatic Analysis and Enhancement of Alaryngeal Speech*, Charles C. Thomas, Springfield, Ill., 1982, pp. 137-165.

[4] Yingyong , Qi, and Weinberg, B., "Low-frequency energy deficit in electro laryngeal speech",*J. Speech and Hearing Res.*, Vol. 34, 1991, pp. 1250- 1256.

[5] Espy-Wilson, C. Y., Chari, V. R., and Huang, C. B., "Enhancement of alaryngeal speech by adaptive filtering", *Proc. ICSLP*, 1996, pp.764-771.

[6] Barney, H. L., Haworth, F. E., and Dunn, H. K., "An experimental transistorized artificial larynx", *Bell Systems Tech. J.*, Vol. 38(6), 1959, pp. 1337-1356.

[7] Weiss, M., Komshian, G. Y., and Heinz, J., "Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx", *J. Acoust. Soc. Amer.*,Vol. 65(5), 1979, pp.1298-1308.

[8] Pandey, P. C., Bhandarkar, S. M., Bachher, G. K., and Lehana, P. K., "Enhancement of alaryngeal speech using spectral subtraction", *Proc. 14th Int. Conf.DSP'2002*, Santorini,Greece,2002,pp.591-594.

[9] Meltzner, G. S., and Hillman, R. E., "Impact of abnormal acoustic properties on the perceived quality of electrolaryngeal speech", *Proc. ICSA Tut. and Res. Workshop*, Geneva, 2003 pp .73-78

[10] Boll, S. F., "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. ASSP*, Vol. 27(2), 1979, pp. 113-120.

[11] Berouti, M., Schwartz, R., and Makhoul, J., "Enhancement of speech corrupted by acoustic noise", *Proc. ICASSP*, 1979, pp. 208-211.

[12] Stahl, V., Fischer, A. and Bippus, R., "Quantile based noise estimation for spectral subtraction and Wiener filtering", *Proc. ICASSP*, Vol. 3, 2000, pp. 1875-1878.

[13] Evans, N. W. D., Mason, J. S. and B. Fauve, "Efficient real-time noise estimation without speech, non-speech detection: An assessment on the Aurora corpus", *Proc. 14th Int. Conf. DSP 2002*, Santorini, Greece, 2002, pp. 985-988.

[14] Evans, N. W. D., and Mason, J. S. "Time-frequency quantile-based noise estimation", *Proc. EUSIPCO '02*, 2002.

[15] Houwu, B., and Wan, E. A., "Two-pass quantile-based noise spectrum estimation", *Oregon Health & Science Univ. (OHSU) Tech. Report*, 2002. http://citeseer.nj.nec.com/560702.html

[16] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., *Numerical Recipes in C*, Cambridge University Press, Cambridge, UK, 1992.

[17] Pratapwar, S. S., Pandey, P. C., and Lehana, P. K., "Reduction of background noise in alaryngeal speech using spectral subtraction with quantile based noise estimation", *Proc. 7th World Conf. SCI 2003*, Orlando, Florida, 2003.