ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 2, March 2017

A Novel Approach to Speech Enhancement using Modified Spectral Subtraction

Rajalakshmi.P[1], Kopperundevi.P[2]

PG Student [APE], Dept. of ECE, Sri Venkateswara College of Engineering, Chennai, Tamilnadu, India [1]

PG Student [APE], Dept. of ECE, Sri Venkateswara College of Engineering, Chennai, Tamilnadu, India [2]

ABSTRACT: Unseen noise estimation is a key, yet challenging step to make the speech enhancement algorithm work in adverse environments. The only prior knowledge known about the encountered noise is that it is different from the involved speech. Therefore, by subtracting the components which cannot be adequately represented by a well-defined speech model, the noises can be estimated and removed. The proposed work consists of two parts, offline training and online enhancement. In offline training, the clean speech data and input noisy speech are collected. The training of these signals are done using Support Vector Machine (SVM). In online enhancement, these signals are compared and their spectrum is estimated. The Modified Spectral Subtraction (MSS) method is employed for the estimation and noise removal. Finally the enhanced speech signal is obtained by transforming the estimated spectrum into time domain.

KEYWORDS: Speech enhancement, SVM, MSS, PSNR.

I.INTRODUCTION

Speech Enhancement is an important stage to improve speech quality and intelligibility of a noisy speech signal degraded in adverse conditions. The main aim of this paper is speech enhancement and the removal of noise with the estimation of noise power spectrum. Speech Enhancement improves overall perceptual quality of degraded speech signal using audio signal processing techniques and used for many applications such as mobile phones, VoIP, teleconferencing systems, speech recognition, and hearing aids, etc., The main problem in speech enhancement is the separation of speech and noise, for which a commonly deployed technique is estimation and removal of the noise spectrum from the input noisy speech spectrum. Another difficulty in speech enhancement is that many types of noises are non-stationary. The spectral properties of non-stationary ones are difficult to predict and estimate, which makes noise removal challenging.

In this paper, the only assumption is that the noise taken is different from the involved speech. A supervised learning algorithm called Support Vector Machine (SVM) is used for the accurate modeling of the clean speech spectrum. In the enhancement stage, Modified Spectral Subtraction (MSS) is used for the removal of noise and estimation of the noise power spectrum and finally reconstructing the original clean speech signal.

II. METHODOLOGY

The proposed system consists of two parts, offline training and online enhancement. In offline training, the clean speech data and the noisy speech magnitude are collected and then trained using offline algorithm called SVM. In online enhancement, these signals are compared and estimation and removal of noise is done using MSS. The spectrum is reconstructed in time domain.

The central methods for enhancing speech are the removal of background noise, echo suppression and the process of artificially bringing certain frequencies into the speech signal. It will be focused on the removal of background noise after briefly discussing what the other methods are all about. When the background noise is suppressed, it is crucial not to harm or garble the speech signal or at least not very badly. Another thing to remember is that quiet natural background noise sounds more comfortable than more quiet unnatural twisted noise. If the speech signal is not intended to be listened by humans, but driven for instance to a speech recognizer, then the comfort is not the issue. It is crucial then to keep the background noise low. Background noise suppression has many applications. Using telephone in a noisy environment like in streets of in a car is an obvious application. Traditionally, the background noise has been suppressed when sending speech from the cockpit of an airplane to the ground or to the cabin. It is easy to come up with similar examples.

It is also a good idea to enhance speech for coding and recognition purposes. Speech codes have been optimized for speech and they usually make the background noise sound weird. Moreover, enhanced speech can be compressed in fewer bits than non-enhanced. Speech recognition systems whose operation relies on the features extracted from speech will be disturbed by extra noise sounds.

All the speech enhancement methods are aimed at suppressing the background noise are naturally based in one way or the other on the estimation of the background noise. If the background noise is evolving more slowly than the speech, i.e., if the noise is more stationary than the speech, it is easy to estimate the noise during the pauses in speech. Several speech enhancement methods were developed over the past several years. Spectral subtraction is one such method which subtracts an estimate of the short-term noise spectrum to produce an estimated spectrum of the clean speech.

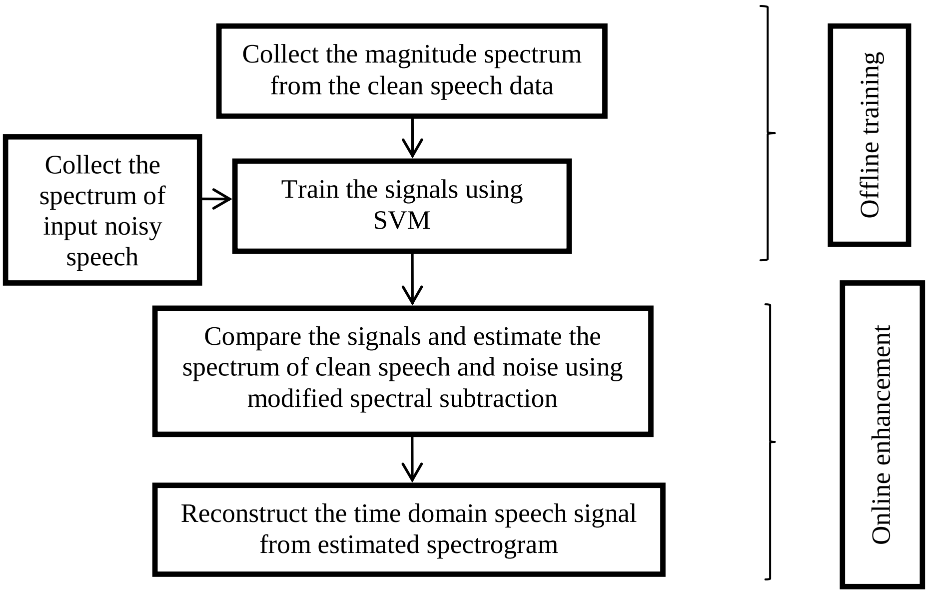

Figure 1 Block Diagram for the Proposed Method

A.SUPPORT VECTOR *MACHINE* (SVM)

SVM are supervised machine learning algorithms that analyse data for classification and regression analysis. The goal is to decide which class a new data point will be in. SVM training algorithm builds a model that assigns new examples into one category or other. SVM segregates two classes.
SVM can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. SVMs maximize the margin around the separating hyper plane. The decision function is fully specified by a subset of training samples, the support vectors. Support vectors are the data points that lie closest to the decision surface. They are the most difficult to classify. They have direct bearing on the optimum location of the decision surface.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 2, March 2017

The goal of the SVM is to train a model that assigns new unseen objects into a particular category. It achieves this by creating a linear partition of the feature space into two categories. Based on the features in the new unseen objects (e.g. documents/emails), it places an object "above" or "below" the separation plane, leading to a categorization. However, much of the benefit of SVMs comes from the fact that they are not restricted to being linear classifiers.

B.MODIFIED SPECTRAL SUBTRACTION (MSS)

Spectral subtraction (SS) is based on the principle that one can obtain an estimate of the clean signal spectrum by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The spectral subtraction method is a simple and effective method of noise reduction. In method, an average signal spectrum and average noise spectrum are estimated in parts of the recording and subtracted from each other, so that average signal-to-noise ratio (SNR) is improved. It is assumed that the signal is distorted by a wide-band, stationary, additive noise, the noise estimate is the same during the analysis and the restoration and the phase is the same in the original and restored signal.

$$Y(n) = X(n) + D(n) \qquad (1)$$

where Y(n) – noisy speech , X(n) – speech signal and D(n) – noise

The noisy speech is segmented into overlapping frames. Then Hamming window is applied on each segment and a set of Fourier coefficients using short-time fast Fourier transform is generated. Noise spectrum is estimated during periods when no speech is present in the input signal. This condition is recognized by Voice Activity Detector (VAD) to produce a control signal which permits the updating of store with spectrum when speech is absent from the current segment. This spectrum is smoothed by making each frequency samples of the average of adjacent frequency samples. This smoothed spectrum then will be used to update a spectral estimate of noise, which consists of a proportion of the previous noise and a portion of the smoothed short-term spectrum of current segment. Thus the noise spectrum gradually adapts to changes in the actual spectrum noise. After noise estimation and subtraction , the a root of the output terms is taken to provide corresponding Fourier Amplitude components and the time-domain signal segments reconstructed by an inverse Fourier transform unit from these along with phase components φ directly from the FFT unit. The windowed speech segments are overlapped to provide the reconstructed output signal at an output.

III. METRICS FOR EVALUATION

Two metrics were computed to evaluate the performance of the speech enhancement algorithms.

A.PESQ SCORE

PESQ stands for Perceptual Evaluation of Speech Quality. PESQ is the new ITU-T standard for measuring the voice quality of communications networks. It measures the subjective speech quality. It is calculated by comparing the enhanced speech with the clean reference speech. The value ranges from -0.5 to 4.5. It analyses the speech signal sample by sample. It provides numerical measure of the quality of human speech. PESQ can be used in a wide range of measurement applications since it is fast and repeatable.

B.PEAK SNR (PSNR)

It refers to Peak Signal-to-noise ratio. It is the ratio between the maximum possible power of a signal and the power of corrupting noise. Usually it is expressed in terms of the logarithmic decibel scale. It is most easily defined via the mean squared error (MSE). Lower the error, higher will be the PSNR. The PSNR value 40 dB or more refers to good quality of the signal.

$$PSNR = 20 \log 10 (MAX_i) – 10 \log 10 (MSE) \qquad (2)$$

Where
MSE – mean squared error
MAX_i – max possible value of the signal

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 2, March 2017

$$MSE = \frac{1}{N}\sum_{i=0}^{N}(xi - yi)2 \qquad (3)$$

Where
x_i and y_i are the original and noisy signals
N – no of signal samples

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments were carried out on the MATLAB. Two databases were used – noizeus and super seded. The samples were experimented under both supervised and unsupervised conditions.

A.OFFLINE TRAINING RESULTS

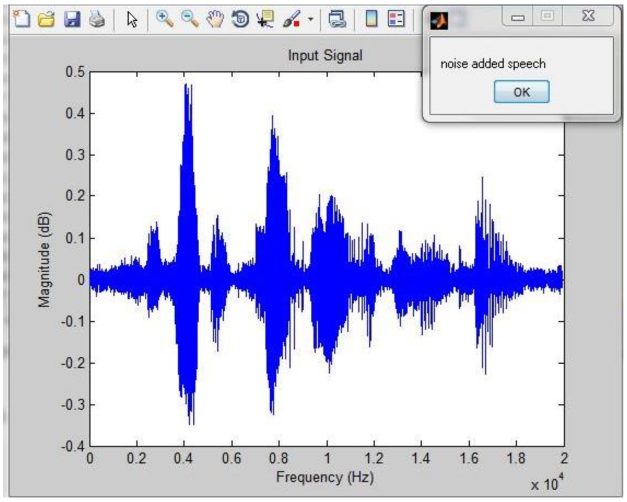

Figure 2 Output of offline training process

The above figure 2 shows the spectrum of an input speech of the offline training process. It shows the training of the input speech which is given as input to the system of speech enhancement.

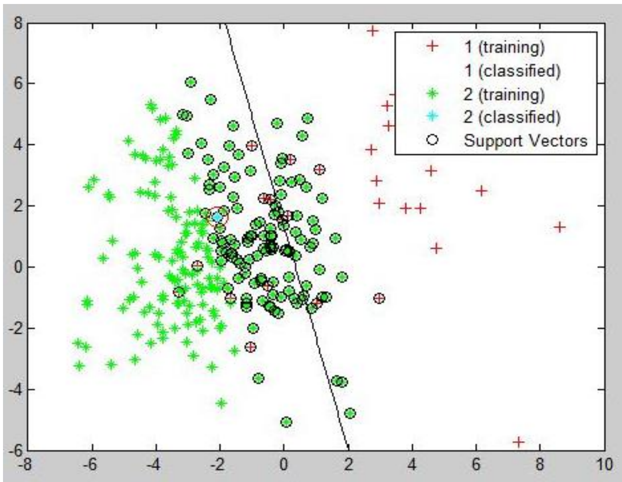

Figure 3 SVM plot

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 2, March 2017

The above figure 3 shows the SVM plot of a noisy speech. From this figure, it is observed that the samples given are classified into two classes as clean and noisy speech using the SVM training and classification algorithms.

B.ONLINE ENHANCEMENT RESULTS

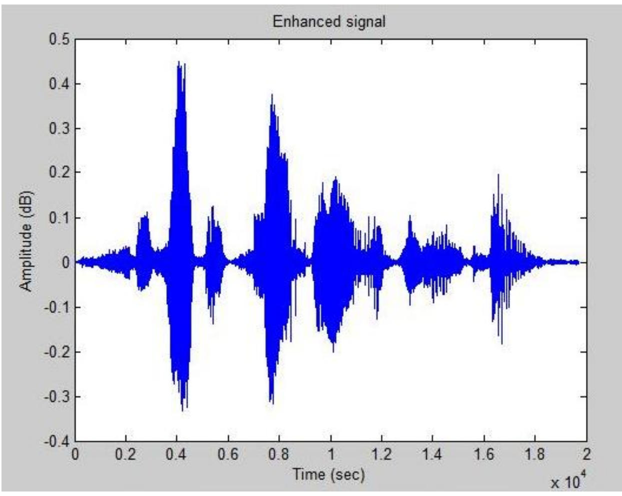

Figure 4 Enhanced output

The above figure 4 shows the enhanced output of the process. It is the reconstructed time domain signal after noise estimation and noise removal.

TABLE I
Comparison of PSNR Values for Noizeus Database

Noises	PSNR(dB) of Supervised samples	PSNR(dB) of Unsupervised samples
Noise 1	87.586	86.120
Noise 2	89.196	87.859
Noise 3	86.929	85.165
Noise 4	87.544	85.943
Noise 5	90.308	87.877
Noise 6	90.991	88.785
Noise 7	85.615	84.832
Noise 8	88.337	87.262

The above Table I shows the comparison of the obtained PSNR values of the supervised and unsupervised samples of the Noizeus database. It is inferred that the performance of the latter is quite close to that of the first. Hence enhancement is efficiently carried out.

V. CONCLUSION

Modified spectral subtraction method was proposed for the estimation and removal of noise for speech enhancement. Initially the input signals are trained using SVM. Compared to previous methods, this method copes up with both stationary and non-stationary noises. Both supervised and unsupervised learning methods were investigated. Experimental evaluation on PESQ score and PSNR on the two databases i.e., Noizeus and Superseded are

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 2, March 2017

demonstrated. The results of unsupervised samples are similar to the supervised one. This method is highly efficient for learning real world datasets. The noises are reduced without affecting the signal power and the SNR is improved.

In the future work, classification the types of noises present in the signal will be carried out. Then the features are extracted in Automatic Speech Recognition systems. The goal of an ASR system is to accurately and efficiently convert a speech signal into a text message transcription of the spoken words independent of the speaker, environment or the device used to record the speech. Future applications of automatic speech recognition will contribute substantially to the quality of life among deaf children and adults, as well as public and private sectors of the business community who will benefit from this technology.

REFERENCES

[1] Meng Sun, Xiongwei Zhang, Hugo Van hamme, and Thomas Fang Zheng, "Unseen Noise Estimation Using Separable Deep Auto Encoder for Speech Enhancement," IEEE/ACM Transactions on Audio , Speech and Language Processing, Vol 24, no 1, pp 93-104, January 2016.
[2] Y.Xu, J.Du, L-R. Dai, and C-H. Lee, "A regression approach to speech enhancement based on deep neural networks," IEEE/ACM Transactions on Audio , Speech and Language Processing, Vol 23,no 1, pp 7-19, January 2015.
[3] N.Mohammadiha, P.smaragdis, and A.Leijon, "Supervised and unsupervised speech enhancement using non negative matrix factorisation," IEEE/ACM Transactions on Audio , Speech and Language Processing, Vol 21,no 10, pp 2140-2151, October 2014.
[4] Karam M., Khazaal H.F., Aglan H. and Cole C., "Noise Removal in Speech Processing Using Spectral Subtraction," Journal of Signal and Information Processing, Vol 5, pp. 32-41,2014.
[5] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising auto encoder," in Proc. INTERSPEECH, 2013, pp.436–440.
[6] Z. Chen and D. P. Ellis, "Speech enhancement by sparse, low-rank, and dictionary spectrogram decomposition," in Proc. IEEE Workshop Application Signal Process. Audio Acoust., 2013, pp. 1–4.
[7] Ekaterina Verteletskaya, and Boris Simak, "Noise Reduction Based on Modified Spectral Subtraction Method," IAENG International Journal of Computer Science, Vol 38, pp. 231-239,2011.
[8] J. Bai and M. Brookes, "Adaptive hidden Markov models for noise modelling," in Proc. 19th Eur. Signal Process. Conf. (EUSIPCO'11),Aug. 2011, pp. 494–499.
[9] K. Paliwal, K. Wjcicki, 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.
[10] D.Y.Zhao, W.B. Kleijn, A.Ypma, and B.de Vries, "Online noise estimation using stochastic-gain HMM for speech enhancement," IEEE/ACM Transactions on Audio , Speech and Language Processing, Vol 16,no 4, pp 835-846, May 2008.
[11] Yang Lu and Philipos C. Loizou, "A geometric approach to spectral subtraction," in speech communication,2008.
[12] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook driven short-term predictor parameter estimation for speech enhancement," IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 1, pp. 163–176,January 2006.
[13] P.Smaragdis, "Non-negative matrix factor deconvolution: Extraction of multiple sound sources from monophonic inputs,"5th Int.Conf. Ind. Compon. Anal., September 2004,pp 494-499.
[14] Anuja Chougule and V. V. Patil , " Survey of Noise Estimation Algorithms for Speech Enhancement Using Spectral Subtraction," in International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2, Issue: 12, pp. 157-168,2004.
[15] K. Lebart, and J. M. Boucher, "A New method based on spectral subtraction for speech enhancement," Acustica, Vol. 87, pp. 359-366,2001.
[16] Y. Ephraim, "A signal subspace approach for speech enhancement," IEEE Trans. on speech and audio processing, pp. 251-266,1995.