Highly Training Algorithm for Enhancement of Speech Signal Data (HTA-ESSD)

Kumbhar Trupti Sambhaji
Research Scholar, Dept. of ECE
SaIT, Bengaluru

Veena C.S., PhD
Professor, Dept. of ECE
SaIT, Bengaluru

ABSTRACT
The enhancement of speech aims to maximize the quality of speech by utilizing HTA (Highly Training Algorithm). The main aim of enhancement is to maximize the intelligibility or perceptual quality of the speech signal data. We represent HTA, aimed at fast removal and very effective of background noise from the signal-channel of speech signal data based on analytically determined output-weights and randomly selected-hidden units. The feature learning with HTA may not be effective for the natural signals, even with the larger number of the hidden nodes, C-HTA (Classified Highly Training Algorithm) are employed by leveraging the sparse auto-encoders. This work is mainly to apply C-HTA and HTA to enhance the speech-signal data. The proposed HTA is evaluated on Aurora database at three SNRs. We also compare our introduced algorithm with many state-art-methods.

Keywords
Enhancement of speech, HTA, C-HTA

1. INTRODUCTION
Presently, the transmission of speech signal from one device to another is been designed very well and largely utilized as an applicable tool for interacting with others because of its portable behavior. Anyways, the speech transmission of devices is undoubtedly affected by changing the surrounding noises. Many SEA (Speech-Enhancement-Algorithm) have been researched in recent decades for resolving the removal of surrounding noises because of their understanding purpose [1]. SE, the operation of enhancing perceptual features of sound spoiled by additional noise is a concept that has gained attention of processing of speech signal from many years. SE is utilized in every latest communication device. It is general that when speech signal is sent, the quality of it may be spoiled because of surrounding noises from where it is going through. Few of the surrounding noise cause the degradation of the speech quality of sent along with acoustic addition noise, white Gaussian noise or acoustic reverberation. It is a necessary element in the transmission process. It enhances speech signal and minimizes noise; it is utilized in various domains such as help in listening and other utilities like cell phones, hearing aids, tele-conferencing devices, and voice interaction devices. SE is nearly associated to restoration of the speech as it rebuilds and restores the speech after getting spoiled [2]. Anyways, there is a small vary between SE and restoration of speech. Speech restoration is used to transform the noisy signal into the original signal i.e. before additive noise. Whereas SE is used to improve the original signal. Furthermore, an original degraded signal can be enhanced but cannot be restored. The motive of SEA is to enhance perceptual features of speech data, which is spoiled by additional noise like intelligibility or complete quality with motive of minimizing problem of the listener [3]. SE can be utilized in various configurations like in places with interrupting background noise on noisy roads, buildings or streets at the places where automobiles are passing. These interrupting noises spoil the quality of original signal in a way where no clean signal is remained. There are different kind of noise that cause degradation of original speech data. Few of the following techniques are wiener filter, MMSE, log MMSE, spectral subtraction and decision direct methods used for minimizing the noise [5]. These old methods were mainly designed with the help of less amount of a priori details regarding noise and speech. Thus, they use to produce limited performance of improvement, particularly when the speech is degraded by additional noise like under minimum input SNR or non-linear noise conditions.

To resolve these constraints, method of machine learning is used to SE operation and represented very good performance in previous years. NMF technique is one of the SE technique, example [4] that decomposes a provided matrix into activation and basis matrix with positive elements. In a supervised infrastructure, the source vector of sound speech and sources of noise are gained a priori from learning data, and then utilized while improvement steps. In past few years, DNN (Deep-Neural-Network) techniques have obtained more attention, and got various usages like classification of image, speech recognition automatically and improvement of speech quality [5]. DNN has gained achievement in image and speech recognition after Hinton proposed in year 2006. Later, DNN based SE got too much attention due to its capability of complete research on the nonlinear relation among clean speech and noisy speech by training offline that advantages from huge datasets. Or else, it is proven that the technique ignores the local optimization and de-noise efficiently. A DNN based SE structure [6] came up, where spectrum of log-power was trained as features of speech and the task of DNN was to map function to obtain clean speech signal from the noisy signal. Supervised DNN learning targets at assuming the non-linear function for mapping, given with biases and weights of hidden layers of operating network that compares the input and output features. Different DNN infrastructure like feed-forward DNN [5], deep auto-encoder [6], and CNN (Convolutional-Neural-Network) [7] was been used for SE. In [10], author introduced a DNN based on regression for improving the quality of the speech. A function for mapping is implemented in middle of the clean and noisy features. Various hidden layers are used for evaluating the SNR. In [11], an ILMSAF (Improved-LMS-Adaptive-Filter) integrated with DNN for improving the speech quality. DBN (Deep-Belief-Network) is used to predict the coefficients of adaptive filter and the improved speech is obtained via ILMSAF. Reinforcement training can be utilized for enhancement of huge amount of training datasets of DNN. The cochlear implant is implemented depending on the DNN application utilized for improving the quality of the speech. In [12], author done the improvement of the speech using DNN

having three hidden layer. Improvement of audio and video can be performed using DNN.

In our work, we are developing a novel technique Enhancement of Speech Signal Data based on effective and unique of HTA, which is highly good generalization, fast training and capability of the classification/approximation. The HTA are very suitable for wide range of feature-mapping applications. We deploy the Enhancement of Speech Signal Data with C-HTAs. This work is mainly applying C-HTA and HTA to enhance the speech task. To estimate the capability of noise reduction of C-HTA and HTA, we conducted the experiments on Aurora databases [15]. The rest of the paper is organized in such a way that section-2 represents related work, section-3 represents HTA/CHTA based enhancement of speech signal data. Section-4 shows our experimental outcomes and finally conclusions are discussed in section 5.

2. RELATED WORK

SEAs (Speech-Enhancement-Algorithms) can be divided into two important sections, known as processing of the speech techniques and training of data techniques. In this section, we will study on both techniques. Initially, the SE issues will be represented more generally via the speech signal restoration technique. After that, we will study about training the data techniques.

2.1 Conservative Speech Signal Rebuilding Techniques

SEAs consist of modification of speech signal added with noise into speech signal domain to obtain the required clean speech. A noisy signal $ns[t]$ is combination of additional noise $n[t]$ and clean signal $s[t]$.

$$ns[t] = s[t] + n[t], \quad (1)$$

Here t is the index of time. STFT (Short-Time-Fourier-Transform) is performed on noisy speech to obtain the phase and frequency factors. By performing STFT, the speech is separated into small frames with the help of windowing function $win(t)$. The alternative STFT speech can be represented as

$$NS[i,j] = S[i,j] + N[i,j], \quad (2)$$

Here $S[i,j]$, $NS[i,j]$, and $N[i,j]$ are the ith bins of frequency for clean, noisy, and additive noise signal of jth frame, respectively, with respect to the frequency $freq_i$. Where $freq_i = \frac{(2\pi * i)}{I}, i = 0\ to\ (I-1)$. The motive of noise minimizing is to obtain $s[t]$ signal from $ns[t]$ signal. For rebuilding the speech signal, a gain function $gain[i,j]$ is predicted based on the evaluated *posteriori* and *priori* SNR statistic. The improved speech signal $\hat{S}[i,j]$ is gained by refining $NS[i,j]$ by $gain[i,j]$. The phase factor of noisy signal is copied and utilized to make the phase of improved speech signal. A reverse STFT known as ISTFT (Inverse-STFT) is used to transform $\hat{S}[i,j], i = 0\ to\ I; j = 1\ to\ J$ and phase factor for getting the improved signal $\hat{s}$.

2.2 Training of the data techniques:

- **Factorization of Non-negative Matrix (FNM):**

The improvement of speech signal based on FNM (Factorization of Non-negative Matrix), a speech signal's data in matrix form $NS \in \mathbb{R}^{I \times J}$ having I number of bins of frequency and J number of frames of the signal is represented on the space, which is linear integration of vector sets $NS \approx weight \times coeff$. Here,

$$weight = [weight_S weight_N] \in \mathbb{R}^{I \times (vector_s + vector_n)}$$

Where, $weight_N$ and $weight_S$ represents the weight matrices of noise and clean signal, respectively, and,

$$coeff = \left[coeff_{\hat{S}}^{transpose} coeff_{\hat{N}}^{transpose}\right]^{transpose} \in \mathbb{R}^{(vector_s + vector_n) \times J}$$

Here, $vector_s, vector_n \leq \min(I,J)$ are the amount of training vectors w.r.t clean signal and noise. And, $coeff_{\hat{S}}$ and $coeff_{\hat{N}}$ represents the matrices of trained coefficient with respect to clean signal and noise. FNM approximation is obtained with the help of two different minimizing conditions:

(i) The minimum square condition to reduce $\|N - (weight \times coeff)\|^2$ based on $weight$ and $coeff$;

(ii) The universal KL (Kullback-Leibler) divergence to reduce $diverg\big(N \parallel (weight \times coeff)\big)$.

At the training stage of speech signal improvement, FNM is used individually on noisy data and clean data, where evaluation of magnitude for noise ($|N[i,j]|$) and clean signal ($|S[i,j]|$) are performed. Then, the distance of Euclidean between factored matrices and magnitude spectrum is reduced by the following modification rule:

$$coeff \leftarrow coeff \otimes \frac{weight^{transpose} \times NS}{weight^{transpose} \times weight \times coeff}$$

$$weight \leftarrow weight \otimes \frac{NS \times coeff^{transpose}}{weight \times coeff \times coeff^{transpose}} \quad (3)$$

In the improvement stage, a speech signal gain is considered and the improved speech signal is gained as

$$\hat{S}[i,j] = gain[i,j] \times NS[i,j] \quad (4)$$

Here, the function of gain $gain[i,j]$ is equated with the help of particular optimality conditional and statistical model.

3. PROPOSED METHODOLOGY

The Highly-Training-Algorithm (HTA) is used for one-layer feed-forward networks, which gives an effective and rapid training process that does not need huge refinement of parameters.

3.1 Single Highly Training Algorithm (HTA)

In One-Layer Feed-forward Networks (OLFNs), the biases and weights of the input in hidden layer can be selected randomly to train *X* different cases. Where *X* different cases is of (ns_p, s_p). Here, $ns_p = [ns_{p1}, ns_{p2}, \dots, ns_{pQ}]^{transpose} \in \mathbb{R}^Q$ and $s_p = [s_{p1}, s_{p2}, \dots, s_{pP}]^{transpose} \in \mathbb{R}^P$, the result of on-layer feed-forward networks can be given as:

$$func(ns_p) = \sum_{l}^{L\_total} out_l \times active\left((in_l \times ns_p) + bais_l\right) \quad (5)$$

Here, $active(*)$ is the function of activation used in training hidden layers, $in_l = [in_{1l}, in_{2l}, \dots, in_{Ql}]^{transpose} \in \mathbb{R}^Q$ is the weight vector between input and l-th hidden node, $bias_l$ is the bias of l-th hidden node, and $out_l = [out_{l1}, out_{l2}, \dots, out_{lP}]^{transpose} \in \mathbb{R}^P$ is the weight vector between the l-th and output nodes. L_total is the total amount of hidden nodes. For the p-th vector of input, a standard OLFN targets to produce zero error, represented as

$$\sum_{p=1}^{X} \left\| func(ns_p) - s_p \right\| = 0 \quad (6)$$

The equation (6) can be simplified as given below,

$$Hidden \times Bias = S \quad (7)$$

Here,

$$Hidden = \begin{bmatrix} active(in_1 * ns_1 + bias_1) & \cdots & active(in_{L\_total} * ns_1 \\ \vdots & & \vdots \\ active(in_1 * ns_X + bias_1) & \cdots & active(in_{L\_total} * ns_X \end{bmatrix} \quad (7.1)$$

$$Bias = \begin{bmatrix} out_1^{transpose} \\ \vdots \\ out_{L\_total}^{transpose} \end{bmatrix}_{L\_total \times P}, S = \begin{bmatrix} s_1^{transpose} \\ \vdots \\ s_X^{transpose} \end{bmatrix}_{X \times P}$$

The weight matrix of output $Bias$ is evaluated by simplifying as given below,

$$Bias = S \times Hidden^{inverse} \quad (8)$$

Here, $Hidden^{inverse}$ is the pseudo-inverse of the matrix $Hidden$ that can be evaluated using orthogonal projection techniques. For example,

$Hidden^{inverse} = \frac{Hidden^{transpose}}{Hidden \times Hidden^{transpose}}$, here $(Hidden \times Hidden^{transpose})$ must be non-single.

3.2 Classified Highly Training Algorithm (C-HTA)

Using DNNs concept, where the extraction of the features with the help of multi-layer infrastructure along with invalid initialization, HTA was further extended and introduced C-HTA for more than one layer perceptron's (learning binary classifications). The architecture of C-HTA is shown in figure-1. The C-HTA techniques has two levels, one is unsupervised extraction of features and another one is supervised regression of the features. In extraction of unsupervised features, features that are high level are extracted with the help of auto-encoder based on HTA by assuming every single layer as separate layer. The input data is given to HTA feature space for extraction of features, for the purpose of making utilization of information from learning data. The result obtained from unsupervised extraction of features level is given to supervised regression of features level as input for obtaining the result depending on the training from both the levels.

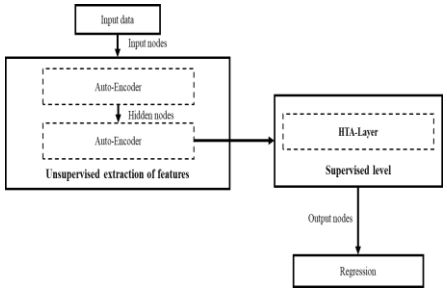

Figure 1: Architecture of C-HTA

3.3 Speech Improvement using HTA and C-HTA

In this, we explain the usage of HTA and C-HTA for the framework of regression to improve the speech. Figure-2 shows the architecture of the implemented speech improving method based on HTA/C-HTA.

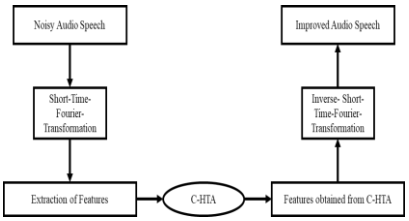

Figure 2: Overall Architecture of implemented speech improving method

The main operation is to utilize HTA/C-HTA framework to convert noisy audio speech into clean audio speech. The complete framework has offline and online levels.

In offline level, a collection of noisy-clean audio speech sets is made. The clean and noisy audio speech signals are transformed into frequency domain using Short-Time-Fourier-Transformation to differentiate the phase and frequency parts of the speech signal. The LPS (Logarithm-Power-Spectra) of clean and noise speech signal are then assigned at the output and input sides of HTA framework, respectively. Mainly, the aim of HTA/C-HTA framework is to rebuild the clean spectrum from the spectrum of noise by reducing the error of reconstruction. For example,

$$Error = \left\| S - \hat{S} \right\|_F^2 \quad (9)$$

Here $\hat{S}$ is the assumed speech spectrum and S is the reference clean speech spectrum. Based on HTA methodology, any linear target function can be assumed as $\sum_{j=1}^{X} \left\| func(NS[j]) - \hat{S}[j] \right\| = 0$, here $\hat{S}[j]$ and $NS[j]$ are the j-th Logarithm Amplitude Vectors of the assumed clean spectrum and input noisy spectrum, respectively. The correlation in equation (5) can be represented as below:

$$func(NS[j]) = \sum_{l-1}^{L\_total} out_l \times active(in_l * NS[j] + bias_l) \quad (10)$$

Here, in_l is the weight vector, bias is represented as $bias_l$ and output weight vector is given as out_l for l-th hidden neuron (node). The correlation in equation (7) can be simply written as,

$$Hidden \times Bias = \hat{S} \quad (11)$$

Here, $Hidden$ is the final layer of hidden layers, output weight is given as $Bias$ and the assumed speech spectrum is represented as $\hat{S}$, represented as:

$$Hidden = \begin{bmatrix} active(in_1 * NS[1] + bias_1) & \cdots & active(in_{L\_tota} \\ \vdots & & \\ active(in_1 * NS[X] + bias_1) & \cdots & active(in_{L\_tota} \end{bmatrix}$$

$$Bias = \begin{bmatrix} out_1^{transpose} \\ \vdots \\ out_{L\_total}^{transpose} \end{bmatrix}_{L\_total \times Y} \quad S = \begin{bmatrix} \hat{S}^{transpose}[1] \\ \vdots \\ \hat{S}^{transpose}[X] \end{bmatrix}_{X \times Y}$$

The alternative matrix of output weight for the assumed speech spectrum can be evaluated as,

$$\widehat{Bias} = Hidden^{inverse} \times \hat{S} \quad (12)$$

Here, $Hidden^{inverse}$ is the pseudo-inverse of the matrix $Hidden$ that can be evaluated using orthogonal projection techniques, output weight matrix is represented as $\widehat{bias}$, and assumed speech spectrum is represented as $\hat{S}$.

In the online level, the spectrum of noise is initially transformed into LPS and phase sets. The features of noisy LPS are converted to get the improved features by performing the equation (10) and (11) for the HTA/C-HTA frameworks ($Hidden$ and $\widehat{Bias}$) assumed in the offline level. The phase set of the noise spectrum is utilized to make the phase of the improved speech spectrum. An Inverse-Short-Time-Fourier-Transformation is used to get the improved speech spectrums.

4. RESULTS

The clean-NB (Narrow-Band) speech samples for evaluation of introduced algorithm were taken from Aurora dataset [13]. The Aurora is developing standards for DSR (Distributed Speech Recognition) in recognition and telecommunication terminal at central area in telecom-network. The samples of noisy speech were generated by adding 3-real-world noises such as Airport, babble and car at different SNRs (0, 5, 10 dB) to clean the speech. Our experimental results were performed on MATLAB-platform with windows-10 OS and hardware compatible with 16GB RAM. Our proposed HTA is compared with various algorithm such as PK-VRE [14], PKV [15], KF [16], PK-OBC [17]. In KF [16], the bidirectional-KF is improved to utilize the model of system dynamics, which uses future and past measurements to estimate the current system-time state. It gives efficient recursive to estimate the process state, which reduces the MSE.

4.1Weighted-spectral-slope-distance (d_{WSS})

The measure of $\mathbf{d_{WSS}}$ (weighted-spectral-slope-distance) has higher correlation along with subjective-quality ratings and calculated as:

$$d_{WSS} = 0.5\mathcal{M} \sum_{m=0}^{\mathcal{M}-1} \frac{\sum_{j=1}^{25} \mathcal{W}(j,m)\left(S_a(j,m) - \bar{S}_{\hat{a}}(j,m)\right)^2}{\sum_{j=1}^{25} \mathcal{W}(j,m)}$$

Where, $\bar{S}_{\hat{a}}(j,m)$ and $S_a(j,m)$ is defined as spectral-slopes of enhanced and clean signals of mth and jth frame. Here, weight is the $\mathcal{W}(j,m)$.The values of $\mathbf{d_{WSS}}$ are shown in Table-1 for NB speech. The values of $\mathbf{d_{WSS}}$ clearly represent that HTA output provides much better clarity compared to other given algorithms utilized in analysis. The below table-1 provides lowest values of $\mathbf{d_{WSS}}$ for our introduced algorithm in every-case. The values of $\mathbf{d_{WSS}}$ represent that our HTA performs much better for all of the types of noises. Also, the values represent that vowels can be differentiate much better in output of HTA speech compared to other algorithms utilized in analysis that showing its higher-intelligibility.

4.2 Log-likelihood-ratio-distance ($\mathbf{d_{LLR}}$)

The $\mathbf{d_{LLR}}$ is defined as the LPC based objective. It is defined as:

$$d_{LRR}(x_c, x_p) = log\left(\frac{x_p \mathcal{R}_c x_p^T}{x_c \mathcal{R}_c x_c^T}\right)$$

Where, x_c, x_p are LPC-vectors of enhanced and original speech of signal frames and $\mathcal{R}_c$ is defined as auto-correlation-matrix of actual speech signal. The lower $\mathbf{d_{LLR}}$ specifies good quality of speech. Below Table-2, represents that our HTA has less values of $\mathbf{d_{LLR}}$ in all of the cases. All of the given values represents that the samples of clean speech are same to their corresponding that improved the speech samples of our proposed HTA than other algorithms.

Table 1 Values of $\mathbf{d_{WSS}}$ for NB Speech

dB	Types of Noise	d_{WSS}				
		PK-VRE [14]	PKV [15]	KF [16]	PK-OBC [17]	Proposed
0	Airport	117	98	92	86	61.44
0	Babble	113	99	91	80	60.82
0	Car	112	95	84	80	56.24
5	Airport	91	77	72	63	47.59
5	Babble	94	78	82	64	47.67
5	Car	86	73	78	62	46.17
10	Airport	69	59	60	50	40.68
10	Babble	69	60	55	52	42.56
10	Car	68	59	61	54	41.52

Table 2 Values of $\mathbf{d_{LLR}}$ for NB Speech

dB	Types of Noise	d_{LLR}				
		PK-VRE [14]	PKV [15]	KF [16]	PK-OBC [17]	Proposed
0	Airport	1.27	1.19	1.02	0.91	0.75
0	Babble	1.21	1.19	1.00	0.88	0.81
0	Car	1.44	1.33	1.13	0.98	0.78
5	Airport	1.01	0.98	0.90	0.68	0.56
5	Babble	1.00	0.99	0.90	0.70	0.60
5	Car	1.06	1.08	0.91	0.80	0.62
10	Airport	0.78	0.78	0.68	0.50	0.44
10	Babble	0.79	0.80	0.78	0.64	0.48
10	Car	0.86	0.90	0.79	0.60	0.51

5. CONCLUSION

This paper introduces the HTA/C-HTA based enhancement of Speech Signal Data that eliminates noise and retains clean speech signal from noisy speech. We utilized the hierarchical framework to improve the HTA ability by replacing one-layer to multi-layer where the levels of distortion were properly controlled to give better performance alongside the desirable

speech intelligibility and speech quality. Our introduced algorithms clearly outperform much better than other state-art of the speech enhancement-signals based methods for NB (Narrow Band) which is based on intelligibility and quality. The low values of $\boldsymbol{d_{WSS}}$ and $\boldsymbol{d_{LLR}}$ for enhanced the outputs of the introduced HTA represents its superior quality over the other algorithms.

6. REFERENCES

[1] Mosayyebpour, S.; Esmaeili, M.; Gulliver, T.A. Single-microphone early and late reverberation suppression in noisy speech. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *21*, 322–335.

[2] Shishir Banchhor, Jimish Dodia, and Darshana Gowda. Gui based performance analysis of speech enhancement techniques. International Journal of Scientific and Research Publications, 3(9):1, 2013.

[3] Hardik Panchmatia, Karan Gaikar, and Dharmesh Patel. Comparison of different speech enhancement techniques. Imperial Journal of Interdisciplinary Research, 2(5), 2016.

[4] P. C. Loizou, Speech Enhancement: Theory and Practice. NewYork: CRC, 2007.

[5] H. Chung. R. Badeau, E. Plourde and B. Champagne, "Training and compensation of class-conditioned NMF bases for speech enhancement," Neurocomputing, vol. 284, pp. 107-118, Apr. 2018.

[6] Xu, Yong, et al. "An Experimental Study on Speech Enhancement Based on Deep Neural Networks." Signal Processing Letters IEEE 21.1(2014):65-68.

[7] X. Lu, Y. Tsao, S. Matsuda and C. Hori, "Speech enhancement based on deep denoising autoencoder," in Proc. Interspeech, pp. 436-440, Aug. 2013.

[8] S. -W. Fu, Y. Tsao and X. Lu, "SNR-aware convolutional neural network modeling for speech enhancement," in Proc. Interspeech, pp. 3768-3772, Sep. 2016.

[9] M. Kolbaek, D. Yu, Z. -H. Tan and J. Jensen, "Joint separation and denoising of noisy multi-talker speech using recurrent neural networks and permutation invariant training," in Proc. MLSP, six pages, Sep. 2017.

[10] S. Nie, S. Liang, H. Li, X. Zhang, Z. Zhang,W. J. Liu and L. K. Dong, "Exploiting spectro-temporal structures using NMF for DNN-based supervised speech separation," in Proc. ICASSP, pp. 469-473, Mar. 2016.

[11] W. Han, X. Zhang, M. Sun, W. Shi, X. Chen and Y. Hu, "Perceptual improvement of deep neural networks for monaural speech enhancement," in Proc. Int. Workshop on Acoustic Signal Enhancement, five pages, Sep. 2016.

[12] R. Ram, M. N.Mohanty, Deep Neural Network based Speech Enhancement. Int. Conf. On Cognitive Informatics & Soft Computing, 2017. (Accepted).

[13] H. Hirsch, and D. Pearce (2000). "The Aurora Experimental Framework for the Performance Evaluation of Speech Recognition Systems under Noisy Conditions." ISCA ITRW ASR2000, Paris, France, September 18-20.

[14] Adda Saadoune, Abderrahmane Amrouche, and Sid-Ahmed Selouani. Perceptual subspace speech enhancement using variance of the reconstruction error. Digital Signal Processing, 24:187 – 196, 2014.

[15] Sudeep Surendran and T. Kishore Kumar. Variance normalized perceptual subspace speech enhancement. AEU - International Journal of Electronics and Communications, 74(Supplement C):44 – 54, 2017.

[16] Y. H. Goh, P. Raveendran, and Y. L. Goh. Robust speech recognition system using bidirectional kalman filter. IET Signal Processing, 9(6):491–497, 2015.

[17] S. Surendran and T. K. Kumar, "Oblique Projection and Cepstral Subtraction in Signal Subspace Speech Enhancement for Colored Noise Reduction," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 12, pp. 2328-2340, Dec. 2018.