



A new double backward distributive weighted adaptive filtering approach for speech quality improvement

V. Srinivasarao¹ · Umesh Ghanekar¹

Received: 11 January 2021 / Accepted: 21 August 2021 / Published online: 20 October 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

In modern telecommunication systems the presence of noise in background degrades the overall intelligibility and quality of the speech signal. The problem of enhancing speech signals and reducing acoustic noise from the noisy environment using adaptive filtering algorithms with incorporation of blind source separation approach has drawn a particular attention in the recent past. In this paper a dual channel double backward distributive weighted adaptive filtering algorithm is proposed for speech quality enhancement. The proposed method has been evaluated using the objective measures such as Perceptual Evaluation of Speech Quality (PESQ) and Short Time Objective Intelligibility (STOI) in different noise setup and the results achieved indicate that this is a better method for speech quality improvement.

Keywords Adaptive filtering · Speech quality · Noise reduction · Speech enhancement

1 Introduction

For the purpose of speech quality improvement and acoustic noise reduction, a number of approaches have been proposed in the literature. The least-mean-square family (LMS and normalized LMS) are the popular ones (Ozeki & Umeda, 1984; Widrow & Hoff, 1960), because of their simplicity and robust performance. The blind source separation scheme (BSS) is one of the efficient methods, which has shown good performances. These methods usually estimate the original source signals using the information of mixture signals from available input channels. Recently, this BSS scheme has been combined with adaptive filtering mechanism for enhancing the speech signal and to reduce the acoustic noise. Several two-channel BSS techniques were proposed to enhance noisy speech, basing on adaptive filters (Bendoumia & Djendi, 2015; Djendi & Bendoumia, 2013, 2014; Mitra, 2020).

The adaptive identification of unknown impulse response in the cross coupling filters is quite equivalent to blind

source separation problem (Gerven & Compennolle, 1995; Mitra, 2020). This same principle of full band algorithms is valid and used with sub band techniques as well using analysis and synthesis filter banks employing decimation and interpolation processes that yields better performance. Speech signal noise removal (Shrawankar & Thakare, 2010) utilized an optimal filtering technique with consideration of time or transform domain technique. This developed filtering approach estimate the noise signal and helps reducing the signal noise with improved speech signal characteristics.

Two channel forward blind source separation is one such speech quality enhancement method, which uses two mixed signals measured at two distinct points, to implement speech quality improvement. In a two channel forward blind source separation, two adaptive finite impulse response filters (FIR) are used to obtain the speech signal estimate from two mixtures of speech and noise. Adaptive filters are usually updated using a forward normalized least mean square algorithm (Priyanka Kajla & George, 2020). Speech enhancement by the consideration of different statistical characteristic properties of noise on the basis of noise classification is presented in Adam and Babikir, (2020).

Enhancement algorithms generally endeavor noise suppression and late reverberation, as early reverberation is perceived, not as a separate sound source and usually improves speech quality and intelligibility. An adaptive denoising and dereverberation kalman filtering framework that tracks the speech and

✉ V. Srinivasarao
sraoyaar@yahoo.co.in
Umesh Ghanekar
ugnitk@nitkkr.ac.in

¹ ECE Department, National Institute of Technology,
Kurukshetra, Haryana, India

reverberation spectral log-magnitudes is presented in Nicolas and Mike B", (2019). Computational complexity of fast transversal filtering algorithm can be further reduced where the adaptation gain is obtained by discarding completely the forward and backward predictors (Benallal & Arezki, 2014).

Several adaptive algorithms were presented earlier in order to reduce noise and enhance the speech signal (Djendi et al., 2007) (Rakesh & Kumar, 2015). The popular algorithms are the recursive least square RLS (Cioffi & Kailath, 1984) 16] and the least mean square LMS and its normalized version (Al-Kindi & Dunlop, 1989; Bashar, 2019; Manoharan, 2019). The NLMS algorithm is famous for its ease of implementation and less computational complexity in comparison with the RLS. However the RLS has better convergence speed. Another approach that is widely used in literature to resolve the problem of corrupted speech signals is the blind source separation (BSS) techniques. Two widely employed BSS structures are forward BSS and backward BSS. These structures are often combined with different adaptive algorithms and used for applications such as acoustic noise reduction and speech enhancement (Mirchandani et al., 1992) (Ghribi et al., 2016).

In (Sayoud et al., 2018), fast NLMS (FNLMS) algorithm combined with the FBSS structure has been presented that showed best performance in comparison with the classical double forward NLMS (DNLMS) one. Classical LMS adaptive algorithms suffer from weak performance for nonstationary signals and RLS algorithms suffer from large computational complexity. In (Rahima et al., 2018) a dual backward adaptive algorithm has been presented that employs simplified fast transversal filter structure and forward prediction to calculate the adaptive gain to yield better values for computational metrics.

In this paper, double backward distributive weighted adaptive filtering approach is proposed for speech quality improvement by way of a two channel convolutive mixture model. In this scheme, two mixed speech signals are used as inputs to estimate the original signals which created these mixtures.

The rest of the paper is organized as follows: The development of the proposed method is presented in Sect. 2. Section 3 provides discussion on the simulated results and the performance evaluation of this proposed method with considered objective measures and the concluding remarks are drawn in Sect. 4.

2 Proposed method of speech enhancement

As discussed earlier, the speech quality in the hearing devices can be enhanced by employing double microphone sound acquisition. With this motivation the proposed method has been presented.

In this section, the proposed method for speech quality improvement is presented with mathematical formulations. In a two channel two microphone model scenario (Priyanka Kajla & George, 2020) may be considered equivalent to the comprehensive convolutive mixing model shown in Fig. 1.

Here $s(n)$ is the speech signal and $i(n)$ is the noise. These signals are passed through the mixture model with two forward impulse responses and two cross coupled impulse responses. In order to simplify the depicted model further, unit impulse response considered for the forward paths and additional background noise is assumed zero. Under this generalization, the outputs $y1(n)$ and $y2(n)$ of the mixture models are given by

$$y1(n) = s(n) + h21(n) * i(n) \quad (1)$$

$$y2(n) = i(n) + h12(n) * s(n) \quad (2)$$

Now these mixture signals are passed through an adaptive filtering block for speech enhancement. This block will deconvolve the mixture signals (Sayoud et al., 2018) using two adaptive filters with variable step sizes and a variable updation algorithm as shown in Fig. 2.

The deconvolved signals are denoted as $z1(n)$ and $z2(n)$ respectively with the corresponding weights of the adaptive filters as

$$p21(n) = [p21;0(n), p21;1(n), p21;2(n), \dots, \dots, p21;M-1(n)]^T \quad (3)$$

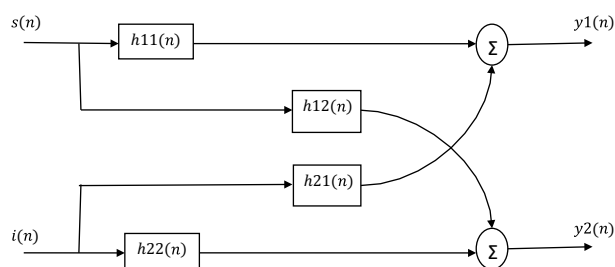
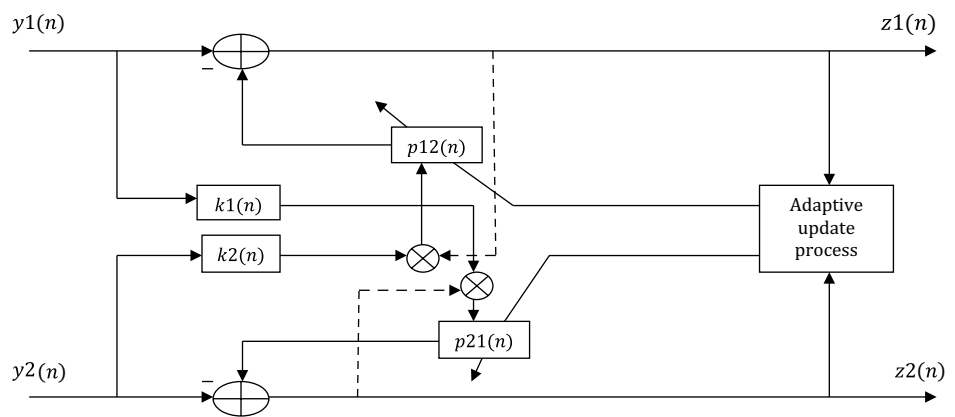


Fig. 1 Complete convolutive mixing model

Fig. 2 Deconvolution (filtering) mechanism with updation process



$$p12(n) = [p12;0(n), p12;1(n), p12;2(n), \dots, p12;M-1(n)]^T \tag{4}$$

These weights are updated in order to remove the signal content from the noise correlated component. Now the estimated output speech is given as

$$z1(n) = y1(n) - p12^T z2(n) \tag{5}$$

where $z2(n) = [z2(n), z2(n-1), z2(n-2), \dots, z2(n-M+1)]^T$ is the tapped delay mixed signal. In a similar way, noise estimation is done. In the blind deconvolution procedure the weights of two adaptive filters $p12(n)$ and $p21(n)$ are updated (Rahima et al., 2018) using normalized least mean square algorithm and here the weights are updated as follows.

$$p12(n) = p12(n) + \sigma1 y2(n) k2(n) z1(n) \tag{6}$$

$$p21(n) = p21(n) + \sigma2 y1(n) k1(n) z2(n) \tag{7}$$

here $k1(n)$ and $k2(n)$ are the adaptation gain vectors with variable step sizes ($\sigma1, \sigma2$), which are calculated from the likelihood variables and kalman gain. The above update rules are termed as double backward normalized least mean square in this paper. The cross coupling model used for the mixing is a familiar acoustic path model where in impulse response will be generally smaller in nature. So as to take the advantage over these impulse responses, a new updating rule may be derived as

$$p21(n+1) = p21(n) - y2(n) k2(n) \cdot z1(n) - \lambda \{ \text{rect}(p21(n)) / \| p21(n) \|^2 \} \tag{8}$$

$$p12(n+1) = p12(n) - y1(n) k1(n) \cdot z2(n) - \epsilon \{ \text{rect}(p12(n)) / \| p12(n) \|^2 \} \tag{9}$$

where λ and ϵ are the loss factors for the update filters while minimizing the cost functions. The adaptation gains $k1(n)$

and $k2(n)$ are obtained by using the calculation of dual kalman variables while discarding the forward and backward predictors.

3 Simulation results

For the purpose of the simulation of the proposed method, speech signal is taken from GRID corpus database and noise is taken from NOISEX92 database. Simulation parameters have been computed for different input SNR scenarios are PESQ (Rix et al., 2001) and STOI (Taal et al., 2010), both have shown significant improvement. PESQ is used for evaluating the quality of the processed speech. The higher the PESQ score, the better will be the quality. The short time objective intelligibility is used to evaluate intelligibility of speech. The STOI is shown to have a high correlation with the speech intelligibility. The larger the STOI score, the more intelligible will be the speech. The simulation results for babble noise are presented in the following Tables 1, 2 and are compared with the available methods (Raj, 2019). The computed PESQ and STOI performance measures are depicted graphically in the following Figs. 3 and 4.

The simulation results for white noise are presented in the following Tables 3, 4 and are compared with the existing methods (Raj, 2019).

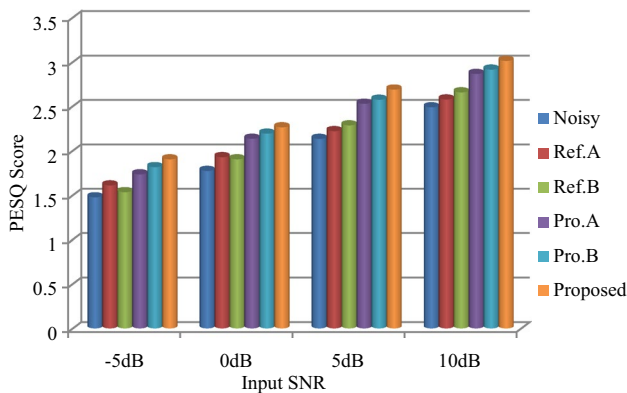
From the Figs. 5 and 6 presented above, it is clearly evident that PESQ and STOI scores have improved

Table 1 Comparison of the results of PESQ for different methods

Input SNR (dB)	Noisy	Ref.A	Ref.B	Pro.A	Pro.B	Proposed
- 5	1.4755	1.6104	1.5334	1.7323	1.8131	1.9015
0	1.7705	1.9263	1.9018	2.1343	2.1923	2.2643
5	2.1332	2.2207	2.2861	2.5266	2.5725	2.6874
10	2.4865	2.5756	2.6562	2.8642	2.9134	3.0121

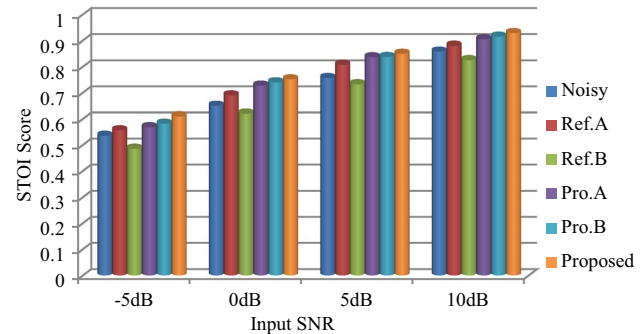
Table 2 Comparison of the results of STOI for different methods

Input SNR (dB)	Noisy	Ref.A	Ref.B	Pro.A	Pro.B	Proposed
- 5	0.5353	0.5563	0.4855	0.5683	0.5816	0.6092
0	0.6494	0.6894	0.6196	0.7271	0.7396	0.7512
5	0.7563	0.8072	0.7325	0.8362	0.8373	0.8496
10	0.8578	0.8796	0.8252	0.9043	0.9142	0.9279

**Fig. 3** PESQ Score comparison for babble noise in different SNR scenario

significantly, in comparison to the existing methodologies. The performance metrics used here hold the highest scores for all the conditions.

In the conventional speech enhancement methods, it is easy to take out the harmonic structure at high frequencies owing to its weak energy so that quality and intelligibility is reduced. Considering the temporal relevance of the speech spectra between the adjacent frames and the application of double backward least mean square algorithm to update the weights of the adaptive filter which result in better noise removal, assist in attaining the better results by the proposed method.

**Fig. 4** STOI Score comparison for babble noise in different SNR scenario

4 Concluding remarks

In this paper, a double backward distributive weighted adaptive filtering scheme has been proposed for quality improvement of the speech signal. The proposed approach provides significant lead over the distributive nature of the impulse responses used in the mixing scenario. This scheme has been shown to provide improved speech quality with babble and white noise simulation in terms of PESQ and STOI when compared to the traditional methods. Hence, this is a better method for speech enhancement in terms of speech quality

Table 3 Comparison of the results of PESQ for different methods

Input SNR (dB)	Noisy	Ref.A	Ref.B	Pro.A	Pro.B	Proposed
-5	1.2852	1.5205	1.9592	2.1265	2.1832	2.2815
0	1.4935	1.8285	2.2072	2.4913	2.5634	2.6543
5	1.7863	2.1672	2.4373	2.7538	2.8025	2.8532
10	2.1064	2.5948	2.7436	2.9218	2.9440	2.9712

Table 4 Comparison of the results of STOI for different methods

Input SNR (dB)	Noisy	Ref.A	Ref.B	Pro.A	Pro.B	Proposed
- 5	0.5186	0.6747	0.5397	0.6861	0.7123	0.7497
0	0.6324	0.7095	0.6535	0.7714	0.7943	0.8276
5	0.7578	0.8118	0.7537	0.8482	0.8593	0.8736
10	0.8593	0.8883	0.8446	0.8942	0.9013	0.9184

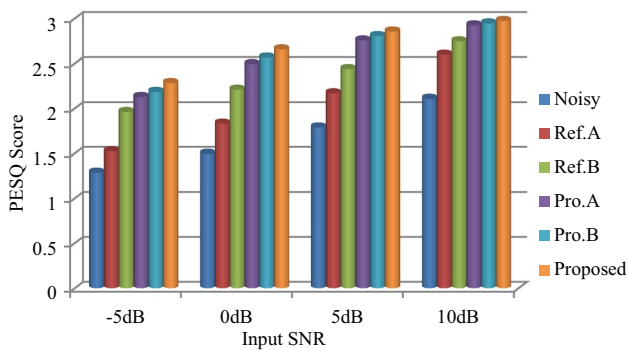


Fig. 5 PESQ Score comparison for white noise in different SNR scenario

and intelligibility. This method can be implemented for the

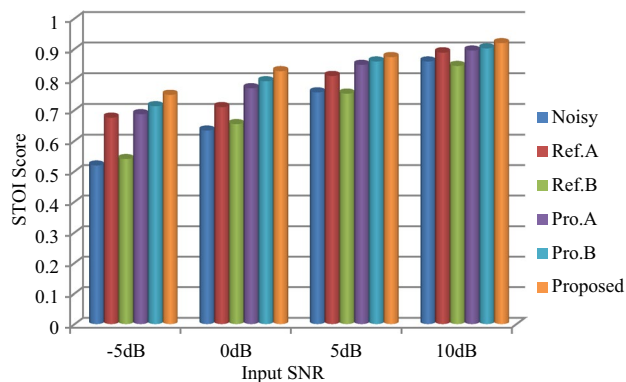


Fig. 6 STOI Score comparison for white noise in different SNR scenario

effectiveness by incorporating a modified filter structure as well as with the other database signals.

References

Adam, E. E. B. (2020). Deep learning based NLP techniques in text to speech synthesis for communication recognition. *Journal of Soft Computing Paradigm (JSCP)*, 2(04), 209–215.

Al-Kindi, M. J., & Dunlop, J. (1989). Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel. *Signal Processing*, 17(3), 241–250.

Bashar, A. (2019). Survey on evolving deep learning neural network architectures. *Journal of Artificial Intelligence*, 1(02), 73–82.

Benallal, A., & Arezki, M. (2014). A fast convergence normalized least-mean-square type algorithm for adaptive filtering. *International Journal of Adaptive Control and Signal Processing*, 28(10), 1073–1080.

Bendoumia, R., & Djendi, M. (2015). Two-channel variable-step-size forward and backward adaptive algorithms for acoustic noise reduction and speech enhancement. *Signal Processing*, 108, 226–244.

Cioffi, J., & Kailath, T. (1984). Fast recursive least squares transversal filters for adaptive filtering. *IEEE Transactions on Acoustic Speech Signal Processing ASSP*, 32, 304–337.

Djendi, M., Henni, R., & Sayoud, A. (2016). A new dual forward BSS based RLS algorithm for speech enhancement. In *International Conference on Engineering and MIS, ICEMIS 2016*, Agadir, Morocco.

Djendi, M., & Bendoumia, R. (2013). A new adaptive filtering sub-band algorithm for two-channel acoustic noise reduction and speech enhancement. *Computers & Electrical Engineering*, 39(8), 2531–2550.

Djendi, M., & Bendoumia, R. (2014). A new efficient two-channel backward algorithm for speech intelligibility enhancement: A sub-band approach. *Applied Acoustics*, 76, 209–222.

Djendi, M., Gilloire, A., & Scalart, P. (2007). New frequency domain post-filters for noise cancellation using two closely spaced microphones. *Proc EUSIPCO, Poznan*, 1, 218–221.

Gerven, S. V., & Compernelle, D. V. (1995). Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness. *IEEE Transactions on Signal Processing*, 43(7), 1602–1612.

Ghribi, K., Djendi, M., & Berkani, D. (2016). A New wavelet-based forward BSS algorithm for acoustic noise reduction and speech quality enhancement. *Applied Acoustics*, 105, 55–66.

- Kajla, P., & George, N. V. (2020). Speech quality enhancement using a two channel sparse adaptive filtering approach. *Applied Acoustics*, 158, 107035.
- Manoharan, S. (2019). A smart image processing algorithm for text recognition information extraction and vocalization for the visually challenged. *Journal of Innovative Image Processing (JIIP)*, 1(01), 31–38.
- Mirchandani, G., Zinser, R. L., & Evans, J. B. (1992). A new adaptive noise cancellation scheme in the presence of crosstalk. *IEEE Transactions on Circuits and Systems*, 39(10), 681–694.
- Mitra, A. (2020). Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset). *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, 2(03), 145–152.
- Nicolas, D., & Mike, B. (2019). Modulation domain kalman filtering for monaural blind speech denoising and dereverberation. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 27(4), 799–814.
- Ozeki, K., & Umeda, T. (1984). An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties. *Electronics and Communications in Japan*, 67(5), 19–27.
- Rahima, H., Mohamed, D., & Djebbari, M. (2018). A dual backward adaptive algorithm for speech enhancement and acoustic noise reduction. In *Proceedings of the Fourth International Conference on Engineering & MIS 2018* (pp. 1–4).
- Raj, J. S. (2019). A comprehensive survey on the computational intelligence techniques and its applications. *Journal of ISMAC*, 1(03), 147–159.
- Rakesh, P., & Kumar, T. K. (2015). A novel RLS adaptive filtering method for speech enhancement. *Electrical, Computer, Energetic*, 9(2), 225–229.
- Rix, A. W., Beerends, J. G., Hollier, M. P., & Hekstra, A. P. (2001). Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In *2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221)* (Vol. 2, pp. 749–752). IEEE
- Sayoud, A., Djendi, M., Medahi, S., & Guessoum, A. (2018). A dual fast nlms adaptive filtering algorithm for blind speech quality enhancement. *Applied Acoustics*, 135, 101–110.
- Shrawankar, U., & Thakare, V. (2010). Noise estimation and noise removal techniques for speech recognition in adverse environment. In *International Conference on Intelligent Information Processing* (pp. 336–342). Springer, Berlin,
- Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2010). A short-time objective intelligibility measure for time-frequency weighted noisy speech. In *2010 IEEE international conference on acoustics, speech and signal processing* (pp. 4214–4217).
- Widrow, B., & Hoff, M. (1960). Adaptive switching circuits. In *Proceedings of IRE Western Electronic Show and Convention* (Part 4, pp. 96–104).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.