INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Subband Kalman Filtering with DNN Estimated Parameters for Speech Enhancement

Hongjiang Yu[1], *Wei-Ping Zhu*[1] *and Benoit Champagne*[2]

[1] Department of Electrical and Computer Engineering, Concordia University, Canada
[2] Department of Electrical and Computer Engineering, McGill University, Canada

ho_yu@encs.concordia.ca, weiping@ece.concordia.ca, benoit.champagne@mcgill.ca

Abstract

In this paper, we present a novel deep neural network (DNN) assisted subband Kalman filtering system for speech enhancement. In the off-line phase, a DNN is trained to explore the relationships between the features of the noisy subband speech and the linear prediction coefficients of the clean ones, which are the key parameters in Kalman filtering. In the on-line phase, the input noisy speech is firstly decomposed into subbands, and then Kalman filtering is applied to each subband speech for denoising. The final enhanced speech is obtained by synthesizing the enhanced subband speeches. Experimental results show that our proposed system outperforms three Kalman filtering based methods in terms of both speech quality and intelligibility.

Index Terms: speech enhancement, subband Kalman filter, deep neural network, wavelet transform

1. Introduction

Speech enhancement aims at removing the background noise in noise-corrupted speech to improve its quality and intelligibility. It has been widely adopted in many applications including speech/speaker recognition, hearing aids and speech communication. During the past decades, researchers have proposed a variety of speech enhancement techniques.

Kalman filtering based speech enhancement was first proposed in [1] and has attracted researchers' great interest because of its capability to enhance the time-domain and non-stationary speech signals. In this kind of method, the clean speech is often characterized as an autoregressive (AR) process and the Kalman filter is viewed as a linear MMSE estimator of the original clean speech. The performance of Kalman filtering is largely dependent on the estimation accuracy of the AR parameters, i.e., the linear prediction coefficients (LPCs) and the driving noise variance. Experiments demonstrated that the AR parameters estimated from the clean speech could achieve excellent performance [1], However, the clean speech is not accessible in practice. As such, various estimation algorithms have been proposed to obtain the estimated parameters from noisy observation [2–5].

In recent years, the deep neural network (DNN) based signal processing methodology has largely advanced the research in speech enhancement. Compared with the unsupervised techniques, the DNN based approaches can achieve a better enhancement performance under the complex noise environment and/or low signal-to-noise ratio (SNR) conditions due to the powerful learning capability of the DNN. The earliest work employing DNN to learn the relationship between the magnitude spectra of the noisy speech and that of the clean speech was found in [6], where the enhanced speech is reconstructed with the DNN-estimated magnitude and the noisy speech's phase. Subsequet works have utilized DNN to estimate the key parameters in traditional speech enhancment methods in order to improve the performances [7–9]. For example, in [9], the DNN is employed to estimate a complex ideal ratio mask (cIRM), which is then applied to the noisy spectrogram to suppress the additive noise. Recently in [10], we have proposed a DNN assisted Kalman filter for speech enhancement, where the DNN is trained to estimate the AR parameters for Kalman filtering. Although the performance is significantly improved compared with the iterative Kalman filtering [2], the high-frequency components of the enhanced speech are still degraded. One possible reason is that the noise does not affect the speech signal uniformly over the whole spectrum [11], while the Kalman filtering is not performed with respect to the different frequencies.

Subband analysis is widely adopted in speech processing such as speech coding. It has also been applied to speech enhancement to separately reduce the background noise in different subbands [12–15]. In [12], a subband Kalman filtering method was proposed, which applies low-order Kalman filters to the noisy subband speeches, and reconstructs the fullband clean speech by synthesizing the enhanced subband speeches. In [13], the authors proposed a multiband spectral subtraction, where the noisy speech spectrum is divided into several non-overlapping bands, and then spectral subtraction is performed independently in each band. The authors of [12] and [13] have shown that the their subband methods yeild better performances compared to their respective full-band counterparts [2, 16].

In light of the successes of the previous subband techniques, in this paper, we propose a novel DNN assisted subband Kalman filtering system for speech enhancement, where the noisy speech is divided into subband speeches using discrete wavelet transform (DWT). For each noisy subband speech, the DNN is employed for the estimation of AR parameters and the Kalman filter is then applied to obtain the enhanced subband speech. The inverse DWT (IDWT) is finally used to obtain the enhanced full-band speech. Compared with our previous work in [10], the proposed system performs denoising at each subband, and is thus able to not only suppress the background noise but also reduce the speech distortion in the enhanced speech, especially at higher frequencies. Computer simulations under various conditions show that the new system can yield better speech quality and intelligibility than previous Kalman filter based algorithms.

2. Proposed speech enhancement system

The overall block diagram of our DNN assisted subband Kalman filtering system is depicted in Fig.1. It contains four

The authors acknowledge the support from China Scholarships Council (CSC No.201606270200) and the NSERC of Canada under a CRD project sponsored by Microchip in Ottawa, Canada.

http://dx.doi.org/10.21437/Interspeech.2020-1913

parts: subband analysis, DNN based line spectrum frequencies (LSFs) estimation, Kalman filtering and subband synthesis. The details of each part are introduced in the following subsections.

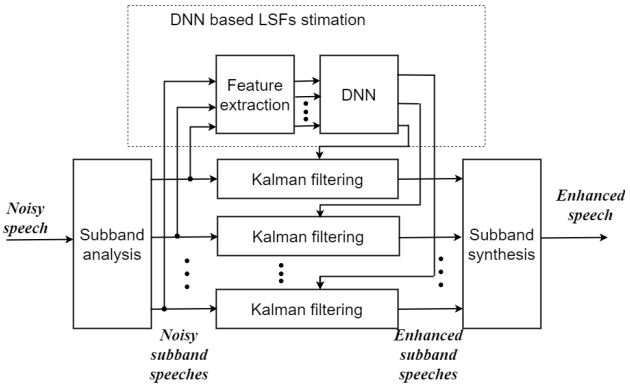

Figure 1: *Block diagram of proposed speech enhancement system.*

2.1. Subband analysis and synthesis

Since the Kalman filter is viewed as a time-domain estimator, we adopt the DWT to directly decompose the time-domain noisy speech. This way can avoid the short-time Fourier transform (STFT) operation, which brings moderate distortion to the time-domain signal due to the necessary segmentation and windowing processes [15].

The DWT and IDWT are performed by a set of well-defined low-pass/high-pass filters together with a down/up-sampling process, which are regarded as distortionless analysis/synthesis for a time-domain signal. Taking a 2-level DWT for an example, in the first level, DWT decomposes a full-band signal x into two subband signals with respect to the low and high frequency information components. In the next level, the decomposition operation is further applied to the low frequency subband signal, while the high frequency subband remains untouched. That is, with a J-level of DWT, we will obtain $J+1$ subband signals, as denoted by

$$x_b = DWT^J\{x\}, \quad b = 1, 2, \cdots, J+1, \tag{1}$$

where x_b is the b-th subband signal produced by DWT with b denoting the subband index.

Similarly, for subband synthesis, the IDWT is adopted to reconstruct a full-band signal $\hat{x}$ from the subband signals, which is given by

$$\bar{x} = IDWT^J\{\mathbf{x}\}, \tag{2}$$

where $\mathbf{x}$ denotes the set of all subband signals $\{x_b\}_{b=1}^{J+1}$. The reconstructed signal $\hat{x}$ is identical to the original signal x in the perfect reconstruction case.

2.2. Kalman filtering

While the noisy speech $y(n)$ is decomposed into subband speeches $\{y_b(n)\}_{b=1}^{J+1}$ in the subband analysis, Kalman filter is then applied to each noisy subband speech for denoising. To illustrate the Kalman filtering algorithm, we take an arbitrary noisy subband speech $y_b\,(n)$ for example, which is viewed as a mixture of the clean subband speech $s_b\,(n)$ and the additive noise $w_b\,(n)$,

$$y_b\,(n) = s_b\,(n) + w_b\,(n)\,, \tag{3}$$

where n is the discrete time index. The clean subband speech $s_b\,(n)$ is usually modelled as a dynamic process with the AR system,

$$s_b\,(n) = \sum_{i=1}^{p} a_{b,i} s_b\,(n-i) + v_b\,(n)\,, \tag{4}$$

where $a_{b,i}$ are LPCs of the clean subband speech, p the order of the model, and $v_b\,(n)$ the driving noise with variance σ_v^2.

To facilitate the Kalman filter presentation for speech enhancement, $s_b(n)$ and $y_b(n)$ are expressed in a state-space form as,

$$\begin{cases} \mathbf{s}_b\,(n) = \mathbf{F}\mathbf{s}_b\,(n-1) + \mathbf{G}v_b\,(n) \\ y_b\,(n) = \mathbf{H}^T\mathbf{s}_b\,(n) + w_b\,(n) \end{cases}, \tag{5}$$

where $\mathbf{s}_b\,(n) = [s_b\,(n-p+1)\,, \ldots, s_b\,(n)]^T$ denotes the speech state vector, $\mathbf{F}$ is the transition matrix given by

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ a_{b,p} & a_{b,p-1} & \cdots & a_{b,2} & a_{b,1} \end{bmatrix}, \tag{6}$$

and $\mathbf{H} = \mathbf{G} = [0, \cdots, 0, 1]^T \in \mathbb{R}^p$.

Given the corrupted subband speech $y_b(n)$, the Kalman filter recursively calculates an unbiased and linear MMSE estimate of the state vector $\mathbf{s}_b(n)$. The denoising process can be summarized by the following recursive equations:

$$\begin{cases} e\,(n) = y_b\,(n) - \mathbf{H}^T\hat{\mathbf{s}}_b\,(n|n-1) \\ \mathbf{K}\,(n) = \mathbf{P}\,(n|n-1)\,\mathbf{H}\left(\sigma_w^2 + \mathbf{H}^T\mathbf{P}\,(n|n-1)\,\mathbf{H}\right)^{-1} \\ \hat{\mathbf{s}}_b\,(n|n) = \hat{\mathbf{s}}_b\,(n|n-1) + \mathbf{K}\,(n)\,e\,(n) \\ \mathbf{P}\,(n|n) = \left(\mathbf{I} - \mathbf{K}\,(n)\,\mathbf{H}^T\right)\mathbf{P}\,(n|n-1) \\ \hat{\mathbf{s}}_b\,(n+1|n) = \mathbf{F}\hat{\mathbf{s}}_b\,(n|n) \\ \mathbf{P}\,(n+1|n) = \mathbf{F}\mathbf{P}\,(n|n)\,\mathbf{F}^T + \sigma_v^2\mathbf{G}\mathbf{G}^T \end{cases}, \tag{7}$$

where $\hat{\mathbf{s}}_b\,(n|n-1)$ is a priori estimate of the current state vector $\mathbf{s}_b\,(n)$; $\mathbf{P}\,(n|n-1)$ the predicted state error correlation matrix of $\hat{\mathbf{s}}\,(n|n-1)$, $e\,(n)$ the innovation, $\mathbf{K}\,(n)$ the Kalman gain matrix, $\hat{\mathbf{s}}_b\,(n|n)$ the filtered estimate of state vector $\mathbf{s}_b\,(n)$, and $\mathbf{P}\,(n|n)$ the filtered state error covariance matrix of $\hat{\mathbf{s}}_b\,(n|n)$. The enhanced subband speech $\hat{s}_b\,(n)$ is finally given by

$$\hat{s}_b\,(n) = \mathbf{G}^T\hat{\mathbf{s}}_b\,(n|n)\,. \tag{8}$$

To perform Kalman filtering, three parameters in Eq. (7) should be determined beforehand, that is, the additive noise variance σ_w^2, the driving noise variance σ_v^2, and the transition matrix $\mathbf{F}$ with the LPCs of the clean subband speech. In our method, the additive noise variance is estimated and updated during the unvoiced frames and the driving noise variance is given by

$$\sigma_v^2 = \sigma_y^2 - \sigma_w^2, \tag{9}$$

where σ_y^2 can be computed from the noisy observation with Levinson-Durbin algorithm [17]. The estimation of the LPCs of $s_b(n)$ is introduced in the following subsection.

2.3. DNN based LSFs estimation

To begin with, the LPCs are converted to the LSFs in DNN based estimation since the well-behaved dynamic range of LSFs is suitable for a stable DNN training process [7, 10]. The DNN based LSFs estimation is divided into off-line and on-line

phases. In the off-line phase, a DNN is trained to learn the mapping between the acoustic features of noisy subband speeches and the LSFs of the clean counterparts. In the on-line phase, given the features of a noisy subband speech, the well-trained DNN predicts the LSFs of the clean subband speech. It should be mentioned that instead of training several DNNs for different subbands separately, we employ a single DNN for all the subband speeches to better exploit the relationships within different subbands as well as to reduce the computational and structural complexity.

For the input of DNN, we extract LSFs along with four acoustic features [18] of the noisy speech to represent the speech characteristics. The input features are extracted for each frame of the noisy speech. To make a full use of the temporal information of speech, we incorporate the features of the adjacent two frames into a single extended feature vector, which is then normalized for effective training.

The structure of the DNN used in our proposed system is shown in Fig.2. It is fully-connected and consists of one input layer, three hidden layers with 1024 units in each layer, and one output layer. The activation function used in the hidden layer is the rectified linear unit (ReLU), while a linear function is used in the output layer.

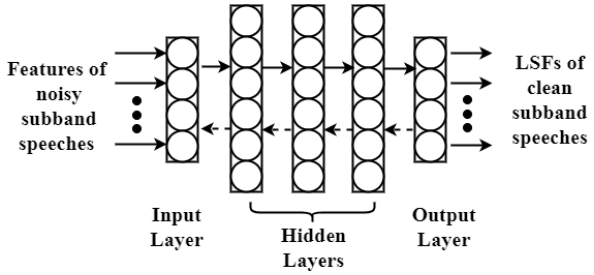

Figure 2: *Proposed DNN for LSFs estimation.*

Back propagation is used to find the optimal weights and biases of the DNN to minimize the cost function, which is defined as the mean square error (MSE) between the reference LSFs and the estimated ones for all subbands,

$$E_r = \frac{1}{J+1}\sum_{b=1}^{J+1}\left\{\frac{1}{M_b}\sum_{m=1}^{M_b}\left\{\frac{1}{p}\sum_{i=1}^{p}\left[\hat{L}_{b,i}(m) - L_{b,i}(m)\right]^2\right\}\right\} \tag{10}$$

where M_b denotes the total number of frames for the b-th noisy subband speech, $L_{b,i}(m)$ and $\hat{L}_{b,i}(m)$ are the reference and the estimated LSFs for frame m, respectively, where $i \in \{1, ..., p\}$ is the order index of the clean speech AR model.

In summary, the proposed DNN assisted subband Kalman filtering system includes an off-line training phase and an on-line enhancement phase. The former trains a DNN with subband noisy and clean speech pairs, while the latter is described in detail below.

- Decompose the full-band noisy speech $y(n)$ into the subband versions $\{y_b(n)\}$ with DWT.
- Extract the features of each noisy subband speech and employ the trained DNN to obtain the estimated LSFs, which are converted to the LPCs to form the transition matrix $\mathbf{F}$.
- Estimate the additive noise variance σ_w^2 during unvoiced frames and compute the driving noise variance σ_v^2 using Eq. (9).
- Perform Kalman filtering with Eq. (7) for each noisy subband speech $y_b(n)$ to obtain the enhanced counterpart $\hat{s}_b(n)$.
- Synthesize the enhanced subband speeches $\{\hat{s}_b(n)\}$ to reconstruct the final enhanced speech $\hat{s}(n)$ with IDWT.

3. Experimental results

3.1. Experimental setup

Clean speeches are selected from the IEEE sentence database [19], among them 670 utterances are used for the off-line training and 50 different utterances for the on-line enhancement. Eight types of noise from NOISEX-92 database [20] are picked to generate the noisy speeches, in which four types (babble, white, street, factory) are used as seen noise, and another four types (pink, buccaneer2, destroyerengine, hfchannel) as unseen noise. The mixing SNR levels are set to -3dB, 0dB, 3dB and 6dB. In the off-line phase, only seen noise is mixed with the training clean speech, which results in 10720 noisy and clean speech pairs. In the on-line phase, both seen and unseen noise are mixed with testing clean speech, giving 800 noisy speeches for both seen and unseen noise. The sampling frequency is set to 16 kHz for both clean speech and noise. It should be noted that since our proposed system aims to enhance the noisy subband speeches in the on-line phase, the noisy and clean speeches in off-line phase are also decomposed into their respective subband signals for the DNN training. A rectangular window is used to divide the audio signals into 20 ms frames with no overlap. For subband Kalman filtering, we set $\mathbf{s}_b(0|0) = \mathbf{0}$, $\mathbf{P}(0|0) = \mathbf{I}$, and the AR model order as $p = 12$.

To evaluate the enhancement performance, two objective metrics are selected: the perceptual evaluation of speech quality (PESQ) measure [21] and the short-time objective intelligibility (STOI) measure [22]. PESQ and STOI evaluate the processed speech from speech quality and intelligibility perspectives, respectively. For both metrics, a higher score means a better speech quality or intelligibility.

3.2. Level of subband analysis

First, we decompose the noisy speech at different levels to find the optimal subband analysis level J in our system. Three levels ($J = 1, 2, 3$) are tested under seen noise. Table 1 shows the objective results under the assumption that the clean speech is accessible to obtain ideal AR parameters for Kalman filtering. In this case, the three subband Kalman filters outperform the full-band processing, which indicates that denoising in each subband indeed removes the additive noise better and introduces less speech distortion. In addition, decomposing the speech with a deeper level contributes to a better performance when the ideal parameters are available.

Table 2 shows the objective results where the DNN-estimated AR parameters are employed for Kalman filtering. We find that adopting 1-level DWT for subband analysis, namely, decomposing the noisy speech into two subband speeches, leads to the best result. As the subband analysis level gets deeper, the number of the input subband signals is increased, which requires a more complex structure to perfectly learn the relationships between more input features and the targets. As such, the 1-level led to better enhancement result. Another possible reason is that if we decompose at a deeper level,

more Kalman filters are required. Since the parameters cannot be ideally estimated for each Kalman filter, the estimation error leads to a degradation of the enhanced subband speeches. Thus, the synthesized full-band speech suffers more performance decrease for high-level subband analysis cases. As a result, we choose 1-level decomposition in our system.

Table 1: *Objective results using **ideal** AR parameters for Kalman filtering under seen noise*

		-3dB	0dB	3dB	6dB
PESQ	Noisy	1.41	1.52	1.68	1.86
	Full-band processing	2.37	2.54	2.70	2.86
	1-level analysis	2.38	2.55	2.72	2.90
	2-level analysis	2.39	2.56	2.74	2.91
	3-level analysis	**2.40**	**2.57**	**2.75**	**2.93**
STOI	Noisy	0.66	0.72	0.78	0.83
	Full-band processing	0.84	0.87	0.89	0.90
	1-level analysis	0.86	0.88	0.90	0.92
	2-level analysis	0.87	0.89	0.91	0.92
	3-level analysis	**0.88**	**0.90**	**0.92**	**0.93**

Table 2: *Objective results using **estimated** AR parameters for Kalman filtering under seen noise*

		-3dB	0dB	3dB	6dB
PESQ	Noisy	1.41	1.52	1.68	1.86
	Full-band processing	1.70	1.93	2.13	2.30
	1-level analysis	**1.92**	**2.16**	**2.36**	**2.57**
	2-level analysis	1.81	2.05	2.27	2.50
	3-level analysis	1.68	1.88	2.12	2.36
STOI	Noisy	0.66	0.72	0.78	0.83
	Full-band processing	0.71	0.77	0.81	0.85
	1-level analysis	**0.72**	**0.78**	**0.84**	**0.87**
	2-level analysis	0.71	0.77	0.83	**0.87**
	3-level analysis	0.69	0.75	0.82	0.86

3.3. Performance evaluation

Four existing algorithms are adopted as referebce methods to compare with our proposed system (DNN-SKF). They are the iterative Kalman filter (I-KF) [2], the perceptual Kalman filter (P-KF) [3], the DNN assisted Kalman filter (DNN-KF) [10] and the DNN based complex ideal ratio masking method (DNN-CIRM) [9]. Since the traditional Kalman filters do not involve a training stage for parameter estimation, experiments are conducted on unseen noise only for fair comparison.

Table 3 shows the average objective scores of different Kalman filter based algorithms. Obviously, DNN-KF and our DNN-SKF outperform the other two traditional unsupervised Kalman filter based algorithms, which indicates that the employment of DNN can provide more accurate LPCs estimates, and thus improve the enhancement performance of Kalman filtering. Moreover, our DNN-SKF has better objective scores than DNN-KF in most cases. It can be inferred that the subband Kalman filtering still works better even on unseen noise. Comparing the results of DNN-SKF in Table 3 with that in Table 2 (the 1-level analysis case under seen noise), the performance under unseen noise does not decrease notably, which demonstrates a good generalization capability of our DNN-SKF. Comparing DNN-SKF with DNN-CIRM, DNN-SKF achieves better perceptual quality, while DNN-CIRM gives a better score of objective speech intelligibility.

At last, the spectrograms of the enhanced speeches resulting from the DNN-KF and the proposed DNN-SKF are plotted. The selected noisy speech is corrupted by babble noise at 0 dB. As shown in Fig.3, without subband analysis, Kalman filtering removes the noise for the whole spectrum. The high-frequency component of the DNN-KF enhanced speech, which has relative low power, would be removed together with the noise and thus suffers severe speech distortion. Contrarily, DNN-SKF, which applies Kalman filtering for each subband signal, can better retain the harmonic structures in high frequencies and the spectrogram exhibits a high similarity to the original one.

Table 3: *Objective results of different enhancement methods under unseen noise*

		-3dB	0dB	3dB	6dB
PESQ	Noisy	1.37	1.51	1.65	1.82
	I-KF	1.64	1.84	2.04	2.26
	P-KF	1.67	1.88	2.09	2.32
	DNN-KF	1.73	1.95	2.21	2.38
	DNN-CIRM	1.77	2.01	2.23	2.43
	DNN-SKF	**1.87**	**2.10**	**2.33**	**2.55**
STOI	Noisy	0.65	0.72	0.78	0.83
	I-KF	0.68	0.75	0.81	0.85
	P-KF	0.69	0.76	0.81	0.85
	DNN-KF	**0.71**	0.77	0.82	0.86
	DNN-CIRM	**0.71**	**0.78**	**0.85**	**0.89**
	DNN-SKF	**0.71**	**0.78**	0.83	0.88

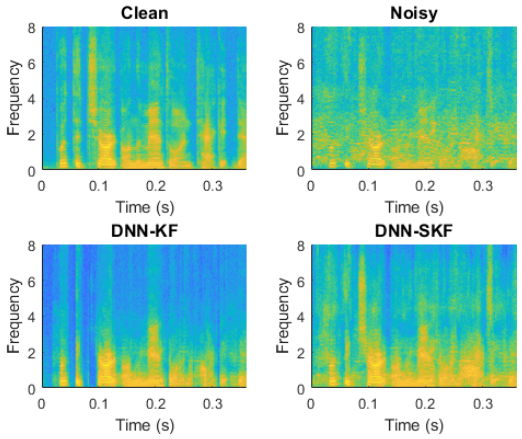

Figure 3: *Spectrograms of clean, noisy and enhanced speeches.*

4. Conclusions

In this paper, a DNN assisted subband Kalman filtering system has been proposed for speech enhancement, which first decomposes the noisy speech into subbands and then performs Kalman filtering in each subband. A DNN has been introduced for LSFs estimation in order to provide more accurate LPCs for Kalman filtering. Experiments have shown that our system outperforms several existing Kalman filter based algorithms. There are two possible reasons behind this improvement. Firstly, the powerful learning ability of DNN helps to estimate LPCs with high accuracy, Secondly, the subband Kalman filtering reduces the speech distortion and efficiently removes the background noise with respect to different frequencies.

5. References

[1] K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 12, April, 1987, pp. 177–180.

[2] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and sequential Kalman filter-based speech enhancement algorithms," *IEEE Trans. on speech and audio processing*, vol. 6, no. 4, pp. 373–385, 1998.

[3] N. Ma, M. Bouchard, and R. A. Goubran, "Speech enhancement using a masking threshold constrained Kalman filter and its heuristic implementations," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 19–32, 2005.

[4] Y. Xia and J. Wang, "Low-dimensional recurrent neural network-based Kalman filter for speech enhancement," *Neural Networks*, vol. 67, pp. 131–139, 2015.

[5] M. S. Kavalekalam, M. G. Christensen, F. Gran, and J. B. Boldt, "Kalman filter for speech enhancement in cocktail party scenarios using a codebook-based approach," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, March, 2016, pp. 191–195.

[6] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE/ACM Trans. on Audio, Speech and Language Processing (TASLP)*, vol. 23, no. 1, pp. 7–19, 2015.

[7] Y. Li and S. Kang, "Deep neural network-based linear predictive parameter estimations for speech enhancement," *IET Signal Processing*, vol. 11, no. 4, pp. 469–476, 2016.

[8] Z. Ouyang, H. Yu, W.-P. Zhu, and B. Champagne, "A deep neural network based harmonic noise model for speech enhancement," *Proceedings of Interspeech*, pp. 3224–3228, September, 2018.

[9] D. S. Williamson and D. Wang, "Time-frequency masking in the complex domain for speech dereverberation and denoising," *IEEE/ACM transactions on audio, speech, and language processing (TALSP)*, vol. 25, no. 7, pp. 1492–1501, 2017.

[10] H. Yu, Z. Ouyang, W.-P. Zhu, and B. Champagne, "A deep neural network based Kalman filter for time domain speech enhancement," *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, May, 2019.

[11] P. C. Loizou, *Speech enhancement: theory and practice*. CRC press, 2013.

[12] W.-R. Wu and P.-C. Chen, "Subband Kalman filtering for speech enhancement," *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 1072–1083, 1998.

[13] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May, 2002, pp. 4160–4164.

[14] S. K. Roy, W.-P. Zhu, and B. Champagne, "Single channel speech enhancement using subband iterative Kalman filter," *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 762–765, May, 2016.

[15] S.-S. Wang, A. Chern, Y. Tsao, J.-w. Hung, X. Lu, Y.-H. Lai, and B. Su, "Wavelet speech enhancement based on nonnegative matrix factorization," *IEEE Signal Processing Letters*, vol. 23, no. 8, pp. 1101–1105, 2016.

[16] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[17] T. Shimamura, N. Kunieda, and J. Suzuki, "A robust linear prediction method for noisy speech," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 4, May,1998, pp. 257–260.

[18] Y. Wang, K. Han, and D. Wang, "Exploring monaural features for classification-based speech segregation," *IEEE/ACM Trans. on Audio, Speech and Language Processing (TASLP)*, vol. 21, no. 2, pp. 270–279, 2013.

[19] IEEE Subcommittee, "IEEE recommended practice for speech quality measurements," *IEEE Trans. on Audio and Electroacoustics*, vol. 17, no. 3, pp. 225–246, 1969.

[20] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247–251, 1993.

[21] ITU-R, "Perceptual evaluation of speech quality (PESQ) an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *Recommendation P.862*, 2001.

[22] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE/ACM Trans. on Audio, Speech and Language Processing (TASLP)*, vol. 19, no. 7, pp. 2125–2136, 2011.