Interspeech 2018
2-6 September 2018, Hyderabad

A Non-convolutive NMF Model for Speech Dereverberation

Nikhil Mohanan, Rajbabu Velmurugan, Preeti Rao

Indian Institute of Technology Bombay
{nikhilm, rajbabu, prao}@ee.iitb.ac.in

Abstract

Reverberation corrupts speech recorded using distant microphones, resulting in poor speech intelligibility. We propose a single-channel, supervised non-negative matrix factorization (NMF) based dereverberation method, in contrast to the convolutive NMF (CNMF) based methods in literature. Recent supervised approaches use a CNMF model for reverberation and a NMF model for clean speech spectrogram to obtain enhanced speech by directly estimating the clean speech activations. In the proposed method, with a separability assumption on the room impulse response (RIR) spectrogram, the reverb speech can be decomposed into bases and activations using conventional NMF. Using these reverb activations, the clean speech activations are estimated to obtain enhanced speech. The proposed model (i) helps in imposing meaningful constraints on the RIR in both frequency- and time-domains to achieve improved enhancement (ii) leads to a framework that can include a NMF model for noise. (iii) gives a better interpretation of the effects of reverberation in the NMF context. We evaluate and compare the enhancement performance of the algorithm on reverb and noisy conditions, simulated using TIMIT utterances and REVERB challenge RIRs. The proposed method performs better than existing C-NMF based methods in objective measures, such as cepstral distance (CD) and speech-to-reverberation modulation energy ratio (SRMR).

Index Terms: NMF, distant speech recording, reverberation, noise

1. Introduction

In many real-world applications such as smart homes, robots, conference meetings, and voice-controlled personal assistants speech recordings are done using microphones placed few meters away from the source. Such distant speech recordings (DSRs) are severely affected by reverberation and background noise [1]. This degrades speech intelligibility and performance of automatic speech recognition performance (ASR) systems. Speech enhancement helps in improving speech intelligibility and can be used as a pre-processing step for improving ASR [2]. The effects of reverberation depend on the properties of speech and room impulse response (RIR). Speech dereverberation can be done using single- or multi-channel data depending on the application of interest. In this work, we address single-channel dereverberation in the DSR scenario.

Dereverberation methods proposed in the literature include reverberation cancellation methods, blind deconvolution based methods, and reverberation suppression methods such as spectral subtraction, linear prediction (LP) and non-negative matrix factorization (NMF) based methods [1]. The earliest work on NMF based dereverberation [3] uses a convolutive NMF (referred as C-NMF) model for the reverb spectrogram. Since then many modifications to this have been proposed both in single-channel [4, 5, 6, 7, 8] and multi-channel scenario [9]. In [8] the C-NMF model is shown as a special case of NMF decomposition. The C-NMF model for speech dereverberation was improved by additionally incorporating a NMF model for clean speech [5, 6]. Various supervised approaches to handle reverberation in noisy environments have also been proposed [7, 10, 11]. Different regularization on RIRs in single-channel [11, 12] and multi-channel [13] scenario have been proposed leading to better speech enhancement. In contrast to these methods that use C-NMF for dereverberation, we propose a NMF model for reverberation. The model uses magnitude spectrogram of the reverb speech and learned clean speech bases to estimate the enhanced speech. Such an approach will allow us to incorporate meaningful constraints in the frequency- and time-domain. This leads to a better speech enhancement, as it has direct control over the estimates of clean speech activations and better RIR estimates. Another advantage of such a model is that it can be easily extended to handle additive noise making it suitable for a noisy reverberant scenario.

2. NMF based dereverberation

Reverberated speech $y(t)$ recorded at a microphone is expressed as the convolution of clean speech $s(t)$ and the RIR $h(t)$ [1]. In the absence of noise, speech degradation due to reverberation can be modeled in the magnitude spectrogram domain by utilizing the modulation transfer function (MTF) model for reverberation [3]. According to the MTF model, the magnitude envelope for each subband of reverberant speech magnitude spectrogram ($\mathbf{Y}$) can be approximated as the convolution of the corresponding subband magnitude envelopes of the RIR ($\mathbf{H}$) and the clean speech ($\mathbf{S}$) spectrograms [4]. Accordingly,

$$Y(k,n) \approx H(k,n) *_n S(k,n) = \sum_{l=0}^{L_h-1} H(k,l)S(k,n-l), \tag{1}$$

where, $Y(k,n)$, $H(k,n)$ and $S(k,n)$ represent the (n,k)-th element of $\mathbf{Y}$, $\mathbf{H}$ and $\mathbf{S}$, respectively, L_h represents the number of frames used to represent the RIR spectrogram $\mathbf{H}$ and $*_n$ represents convolution across frame index. The model in (1) can be viewed as a convolutive NMF (C-NMF) decomposition where $\mathbf{H}$ and $\mathbf{S}$ can be obtained using multiplicative updates [3].

The speech enhancement results using the model in (1) were improved by incorporating a NMF model for the magnitude spectrogram of clean speech [5, 6]. They exploit the low-rank nature of clean speech spectrogram by having a NMF decomposition on clean speech spectrogram $S(k,n)$ as,

$$S(k,n) \approx \sum_{r=1}^{R} W_s(k,r)X_s(r,n), \tag{2}$$

where $W_s(k,r)$, $X_s(r,n)$ are elements of the bases, activations

10.21437/Interspeech.2018-1834

and R is the rank of the decomposition. Using (2) in (1),

$$Y(k,n) \approx \sum_{l=0}^{L_h-1} H(k,l)\Big(\sum_{r=1}^{R} W_s(k,r)X_s(r,n-l)\Big). \quad (3)$$

From (3), enhanced speech is obtained by solving for $H(k,l)$, $W_s(k,r)$ and $X_s(r,n)$ iteratively. This method will be referred to as C-NMF+NMF. In [5, 6], several approaches to obtain bases are experimented with. In online (unsupervised) methods, the bases are learned from the reverberant speech. In offline (supervised) methods, the bases are learned from clean speech utterances. The proposed approach uses supervised bases, so we use the offline method as a baseline for comparison. The joint speech dereverberation and denoising methods using C-NMF model [7, 10, 11] are not compared as they use exemplar bases, which require a large number of bases when compared to learned bases to represent clean speech.

2.1. Proposed non-convolutive NMF model

We propose a method to perform dereverberation by representing the reverb spectrogram using a non-convolutive NMF as, $\tilde{\mathbf{Y}} = \mathbf{W}_R\mathbf{X}_R$, where $\mathbf{W}_R$ and $\mathbf{X}_R$ represent the bases and activation of this decomposition. Such a model is made possible by having a separability assumption on the RIR spectrogram $H(n,k) = H_1(k)H_2(n)$. This approximation is based on the following observations. Firstly, the RIR magnitude spectrum across frequencies for different frames is similar and has a decaying structure across time as is observed in literature [16]. Secondly, the subband magnitudes of the RIR for different frequencies decay with time. The rate of decay with time for different subband is assumed to be same. Combining these observations, a simplifying model will be to write $H(n,k)$ as having a frequency envelope $H_1(k)$ with a gain $H_2(n)$ for different frames. With this, we have

$$\begin{aligned}\tilde{Y}(k,n) &= \sum_{l=0}^{L_h-1} H_1(k)H_2(l)\sum_{r=1}^{R} W_s(k,r)X_s(r,n-l) \\ &= \sum_{r=1}^{R} \underbrace{W_s(k,r)H_1(k)}_{W_R(k,r)} \underbrace{\sum_{l=0}^{L_h-1} H_2(n)X_s(r,n-l)}_{X_R(r,n)} \end{aligned} \quad (4)$$

where, $\tilde{Y}(i,j)$, $W_R(i,j)$, and $X_R(i,j)$ are the (i,j)-th element of $\tilde{\mathbf{Y}}$, $\mathbf{W}_R$ and $\mathbf{X}_R$, respectively. Equation (4) is a NMF decomposition with rank R of the reverb spectrogram. The set of bases and activations obtained from this decomposition is related to clean speech bases and activations. The reverb bases ($W_R(k,r) = W_s(k,r)H_1(k)$) are the clean speech bases $W_s(k,r)$ modified by the frequency envelope of the RIR spectrogram $H_1(k)$. The reverb activations are obtained as the convolution of clean speech activations with the time-dependent envelope of the RIR ($X_R(r,n) = X_s(r,n) *_n H_2(n)$). Estimation of these parameters is done in two steps. In the first step, with the knowledge of learned speech bases, $H_1(k)$ and reverb activations are learned from the reverb spectrogram. Generalized Kullback-Leibler (KL) divergence is used as the distance measure in the first stage as this is related to speech [5]. The NMF cost function (C) is given as,

$$C = \sum_{n,k}\left[Y(k,n)\ln\left(\frac{Y(k,n)}{\tilde{Y}(k,n)}\right) - Y(k,n) + \tilde{Y}(k,n)\right] \quad (5)$$

Multiplicative update rules are obtained for $H_1(k)$ and $X_R(r,n)$ using the cost function in (5). The updates obtained are,

$$\begin{aligned} H_1(k) &\leftarrow H_1(k)\frac{\sum_{n,r}\frac{Y(k,n)}{\tilde{Y}(k,n)}W_s(k,r)X_R(r,n)}{\sum_{n,r}W_s(k,r)X_R(r,n)}, \text{ and} \\ X_R(r,n) &\leftarrow X_R(r,n)\frac{\sum_{k}\frac{Y(k,n)}{\tilde{Y}(k,n)}H_1(k)W_s(k,r)}{\sum_{k}H_1(k)W_s(k,r)} \end{aligned} \quad (6)$$

In the second stage, clean activations are learned from reverb activations. The second stage uses Euclidean distance as the distance measure to estimate clean activations from reverb activations, since the estimated parameters are no longer related to speech as in (5) and can be viewed as a general signal. The NMF cost function in this case is defined as,

$$C_1 = \sum_{r,n}(X_R(r,n) - H_2(n) *_n X_s(r,n))^2 \quad (7)$$

Simultaneous estimation of $X_s(r,n)$ and $H_2(n)$ from $X_R(r,n)$ leads to the trivial solution of $X_s(r,n) = X_R(r,n)$ and $H_2(n)$ being an impulse. To get a meaningful solution, $H_2(n)$ is initialized using prior knowledge of room and source-microphone distance in the RIR structure (Sec 4.1.3 in [2]). The update for $X_s(r,n)$ is given by,

$$X_s(r,p) \leftarrow X_s(r,p)\frac{\sum_n X_R(r,n)H_2(n-p)}{\sum_n \tilde{X}_R(r,n)H_2(n-p)} \quad (8)$$

where, $\tilde{X}_R(r,n)$ is the estimated reverb activation ($\tilde{X}_R(r,n) = X_s(r,n) *_n H_2(n)$). The clean speech spectrogram can be estimated as $\hat{S}(k,n) = G(k,n)Y(k,n)$, where the gain function $G(k,n)$ is written as,

$$G(k,n) = \frac{\sum_r W_s(k,r)X_s(r,n)}{\tilde{Y}(n,k)} \quad (9)$$

The proposed reverberation model using NMF will be referred to as R-NMF.

We now justify the proposed model using an illustrative example, considering a reverberated TIMIT utterance obtained using a RIR with $T_{60} \approx 700$ ms and source-microphone distance (d) of 0.5 m. Using the first step, one can see that the reverb bases $W_R(k,r)$ are indeed the clean speech bases $W_s(k,r)$ acted upon by the frequency envelope of the RIR $H_1(k)$. The estimated frequency envelope of the RIR obtained using the proposed model is compared with the true frequency envelope $H_1^{\text{True}}(k)$ in Figure 1. $H_1^{\text{True}}(k)$ is obtained as the average value of normalized frequency spectrum for the three frames of RIR spectrogram with maximum energy. From the figure, it is clear that for most frequencies, the estimated $H_1(k)$ is very close to $H_1^{\text{True}}(k)$. A similar behavior was observed for other RIRs.

The second step of the proposed approach can be justified by showing that the clean speech activations can be obtained from a deconvolution of the reverb activations and $H_2(k)$. Figure 2 compares the activations obtained using the different NMF models for a specific basis, which is obtained from the NMF decomposition of the reverberated utterance. The clean speech activations (shown in Figure 2(a)) spread due to the effect of reverberation as is shown in Figure 2(b). Dereverberation using the proposed approach helps in reducing this effect. This is evident from the activations estimated using the R-NMF method

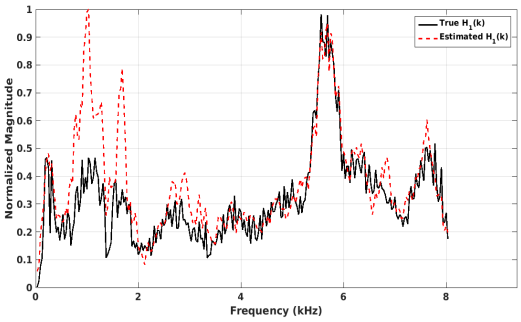

Figure 1: *Comparison of the estimated frequency envelope $H_1(k)$ to that of the true frequency envelope $H_1^{True}(k)$ for a measured RIR with $T_{60} \approx 700$ ms and $d = 0.5$ m.*

as shown in Figure 2(d). The C-NMF+NMF method (shown in Figure 2(c)) was unable to completely recover the clean activations in this case. In our experiments, it was observed that the R-NMF consistently obtained the clean activations, whereas the C-NMF+NMF was not consistent. The overall enhancement

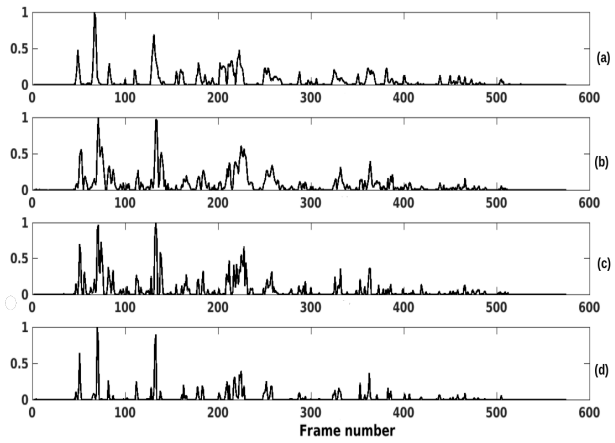

Figure 2: *Normalized activations obtained for a test utterance. (a) Clean utterance, (b) Reverb utterance, (c) Dereverberation using C-NMF+NMF, (d) Dereverberation using R-NMF. The estimated activations in (d) are similar to the true activations in (a).*

obtained using the R-NMF and C-NMF+NMF is in Figure 3 by comparing the enhanced spectrograms with the clean and reverb spectrograms. It can be seen that the R-NMF was more effective in removing the artifacts caused due to reverberation. This is clearly visible in the silence regions as indicated by the red boxes.

2.2. Proposed model for reverberation and noise

One of the advantages of having the NMF model for reverberation as proposed in (4) is that it can be easily extended to include a NMF model for additive noise. The time-domain degraded speech can be represented as $y_D(t) = y(t) + z(t)$. The corresponding magnitude spectrogram $\mathbf{Y}_D$ can be approximated as the sum of reverberation spectrogram $\tilde{\mathbf{Y}}$ and noise spectrogram $\mathbf{Z}$ [7, 10, 18]. Further, the noise spectrogram can also be

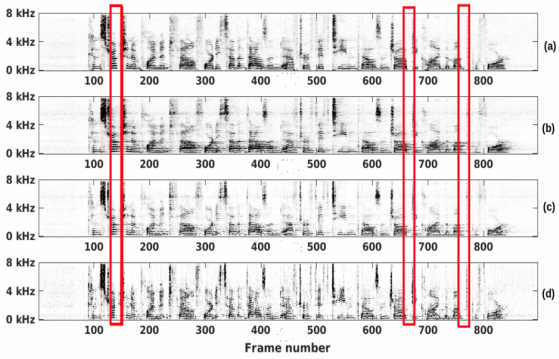

Figure 3: *Spectrogram of (a) Clean speech, (b) Reverb speech, (c) Enhanced speech using C-NMF+NMF, and (d) Enhanced speech using the proposed R-NMF. The regions where R-NMF performs better is shown using red boxes.*

decomposed using a NMF model ($\mathbf{Z} = \mathbf{W}_n\mathbf{X}_n$) [19]. Hence,

$$\tilde{Y}_D(k,n) = \tilde{Y}(k,n) + Z(k,n), \tag{10}$$

$$\text{where } Z(k,r) = \sum_{r=1}^{R_n} W_n(k,r)X_n(r,n). \tag{11}$$

$\tilde{Y}_D(k,n)$, $Z(k,n)$ represent the (k,n)-th element of $\tilde{\mathbf{Y}}_D$, $\mathbf{Z}$, respectively, with $W_n(k,r)$ and $X_n(r,n)$ representing the bases and activations for the noise spectrogram and R_n the rank of NMF decomposition for $\mathbf{Z}$. The model in (10) can be written as,

$$\tilde{\mathbf{Y}}_D = \mathbf{W}_R\mathbf{X}_R + \mathbf{W}_n\mathbf{X}_n = [\mathbf{W}_R|\mathbf{W}_n][\mathbf{X}_R^T|\mathbf{X}_n^T]^T \tag{12}$$

The bases for the decomposition are the combined bases of reverb and noise spectrograms. The reverb basis $W_R(k,r)$ depends on clean speech basis $W_s(k,r)$ as $W_R(k,r) = W_s(k,r)H_1(k)$. Activations for the decomposition in (12) are the combined activations of reverberation and noise activations. The parameters $H_1(k)$, $\mathbf{X}_R$ and $\mathbf{X}_n$ are estimated using multiplicative update rules for the cost function, similar to (5), except that $Y(n,k)$ and $\tilde{Y}(n,k)$ is replaced by $Y_D(n,k)$ and $\tilde{Y}_D(n,k)$. The update rules for $H_1(k)$ and $\mathbf{X}_R$ are similar to the updates in equation (6), except that $Y(k,n)$ and $\tilde{Y}(k,n)$ are replaced by $Y_D(k,n)$ and $\tilde{Y}_D(k,n)$, respectively. Update of $\mathbf{X}_n$ is obtained as,

$$X_n(n,r) \leftarrow X_n(n,r)\frac{\sum_k \frac{Y_D(k,n)}{\tilde{Y}_D(k,n)}W_n(k,r)}{\sum_k W_n(k,r)} \tag{13}$$

We will refer to this model as R-NMF+NMF.

3. Results

The performance of the algorithms in Sec. 2 was compared using speech enhancement measures. As mentioned earlier, we do not consider [7, 10, 11] as these are exemplar based methods.

3.1. Dataset and experiments

The speech enhancement performance was assessed using a subset of 16 speakers, with ten utterances per speaker from the TIMIT database [20]. One utterance spoken by each speaker

Table 1: *Comparison of objective measures for the reverberant and noisy speech, with stationary noise at* 10 *dB SNR*

	CD				SRMR			
	RIR1	RIR2	RIR3	RIR4	RIR1	RIR2	RIR3	RIR4
Degraded speech	4.98	5.43	5.07	5.46	3.24	2.08	3.03	1.92
C-NMF+NMF [5, 6]	5.28	5.35	5.38	5.45	4.69	3.71	4.70	3.85
R-NMF	4.85	5.16	4.97	5.26	4.16	3.43	3.99	3.73
R-NMF+NMF	**4.40**	**4.49**	**4.49**	**4.48**	**5.37**	**4.08**	**5.20**	**4.12**

Table 2: *Comparison of objective measures for the reverberant and noisy speech, with non-stationary noise (Factory) at* 10 *dB SNR*

	CD				SRMR			
	RIR1	RIR2	RIR3	RIR4	RIR1	RIR2	RIR3	RIR4
Degraded speech	5.28	5.63	5.36	5.68	3.53	2.22	3.25	2.05
C-NMF+NMF [5, 6]	5.60	5.65	5.75	5.82	4.81	3.83	4.83	3.99
R-NMF	5.16	5.43	5.28	5.55	4.50	3.53	4.30	3.84
R-NMF+NMF	**4.77**	**5.12**	**4.80**	**5.17**	**5.50**	**4.32**	**5.15**	**4.31**

was used for testing. For the NMF representation, 100 speaker specific clean speech bases were learned from 9 utterances different from the one used in testing of that speaker. Four measured RIRs from the REVERB challenge [2] were used for the evaluation. The RIRs correspond to two different rooms and two different source-microphone distances : near (0.5 m) and far (2 m) in each room. RIR1 and RIR2 correspond to near and far RIR recordings in the room with $T_{60} \approx 600$ ms. RIR3 and RIR4 correspond to near and far RIR recordings for the room with $T_{60} \approx 700$ ms. Each test sentence was convolved with these RIRs to obtain a total of 64 reverberated recordings. Stationary noise available from REVERB challenge [2] or non-stationary noise (factory noise) from [21] was added to the reverberant data at different signal-to-noise ratios (SNRs). 100 noise bases were learned from these noise recordings.

The magnitude spectrogram of the 64 reverberant signals was obtained using a 64 ms window with a hop-size of 16 ms. The square root of Hanning window was used in analysis and synthesis. L_h in (1) was experimentally fixed as 40. $H_1(k)$ was initialized to 1 for all k, $\mathbf{X}_R$ and $\mathbf{X}_n$ were initialized to random values. With the knowledge of RIR, $H_2(n)$ was obtained as the average of $H(n,k)$ for different subbands. Each subband is normalized to have a maximum value 1. The enhanced speech is reconstructed using original noisy phase. The improvement in speech enhancement task was compared using the objective measures of CD and SRMR [14, 15]. Note, we do not include PESQ as it does not provide consistent estimates, as observed in [2]. Dereverberated speech has larger SRMR scores and lower CD when compared to the reverberated speech. The performance of the algorithms is reflected in these measures.

3.2. Speech enhancement using the proposed methods

The performance of the proposed algorithms was compared to the C-NMF+NMF method, for the various reverberation conditions with stationary or non-stationary noise added at different SNRs. For want of space, we do not include the results for 20 dB noise where all the methods behave similarly, with R-NMF+NMF performing slightly better. Table 1 provides objective measures for the degraded speech with a stationary noise at 10 dB SNR and the enhanced speech obtained using the various methods. It can be seen that R-NMF enhanced speech shows improvement in all the objective measures. Considering SRMR, the performance of C-NMF+NMF and R-NMF methods are comparable. However, R-NMF leads to improvements in CD, whereas C-NMF+NMF results in poor CD. The R-NMF+NMF provides significantly better improvements in both the measures when compared to C-NMF+NMF.

Table 2 compares the enhancement results obtained using the proposed methods when non-stationary (factory) noise is added at 10 dB SNR. It can be seen that the R-NMF method performs relatively better compared to C-NMF+NMF. Here too the proposed R-NMF+NMF based enhancement significantly improved the performance, as is evident from the CD and SRMR improvements.

In the presence of noise, the enhancement performance of C-NMF+NMF and R-NMF are comparable, which is expected as there is no explicit model for noise. But, this also shows that the proposed non-convolutive NMF model for dereverberation is equivalent to existing C-NMF model.

4. Discussion and Summary

The proposed R-NMF and R-NMF+NMF methods model reverberation using a non-convolutive NMF model. Assuming a NMF model for noise, R-NMF+NMF method jointly handles noise and reverberation. Such a method provides improved speech enhancement compared to a C-NMF model that does not handle noise. Further, the enhancement results for R-NMF demonstrates the effectiveness of using a NMF model to perform dereverberation. This leads to simple updates and less computational complexity. In addition, using the proposed model we also showed a convincing interpretation of the effects of reverberation on the clean speech activations. The NMF based methods proposed here used learned bases for each speaker. As part of future work, we will look at incorporating speaker independent and exemplar bases. The improvements in CD and SRMR indicate that the proposed method will lead to improved single-channel ASR systems for reverberant and noisy conditions.

5. Acknowledgement

Part of the work was supported by Bharti Centre for Communication in IIT Bombay, Council of Scientific and Industrial Research (CSIR), India and Tata Consultancy Services (TCS), India.

6. References

[1] N. Patrick and G. Nikolay, *Speech Dereverberation*. New York: Springer, 2010.

[2] K. Kinoshita *et al.*, "A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, p. 7, 2016.

[3] H. Kameoka, T. Nakatani, and T. Yoshioka, "Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 45–48.

[4] K. Kumar, R. Singh, B. Raj, and R. Stern, "Gammatone sub-band magnitude-domain dereverberation for ASR," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.

[5] N. Mohammadiha and S. Doclo, "Speech dereverberation using non-negative convolutive transfer function and spectro-temporal modeling," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 24, no. 2, pp. 276–289, 2016.

[6] N. Mohammadiha, P. Smaragdis, and S. Doclo, "Joint acoustic and spectral modeling for speech dereverberation using non-negative representations," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[7] D. Baby, T. Virtanen, and J. F. Gemmeke, "Coupled dictionaries for exemplar-based speech enhancement and automatic speech recognition," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 23, no. 11, pp. 1788–1799, 2015.

[8] H. Kallasjoki, J. F. Gemmeke, K. J. Palomaki, A. V. Beeston, and G. J. Brown, "Recognition of reverberant speech by missing data imputation and NMF feature enhancement," in *Proc. REVERB Workshop*, May 2014.

[9] S. Mirsamadi and J. H. L. Hansen, "Multichannel speech dereverberation based on convolutive nonnegative tensor factorization for ASR applications," in *Proc. Fifteenth Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2014.

[10] D. Baby, "Non-negative sparse representations for speech enhancement and recognition," Ph.D. dissertation, University of Leuven, 2016.

[11] D. Baby and V. H. Hugo, "Joint denoising and dereverberation using exemplar-based sparse representations and decaying norm constraint," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 10, pp. 2024–2035, 2017.

[12] N. Mohanan, R. Velmurugan, and P. Rao, "Speech dereverberation using NMF with regularized room impulse response," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 4955–4959.

[13] M. Yu and F. K. Soong, "Speech dereverberation by constrained and regularized multi-channel spectral decomposition: evaluated on REVERB challenge," in *Proc. REVERB Workshop*, May 2014.

[14] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on audio, speech, and language processing*, vol. 16, no. 1, pp. 229–238, 2008.

[15] T. H. Falk, C. Zheng, and W.-Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1766–1774, 2010.

[16] J. Y. Wen, E. A. Habets, and P. A. Naylor, "Blind estimation of reverberation time based on the distribution of signal decay rates," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 329–332.

[17] M. Jeub, M. Schäfer, H. Krüger, C. Nelke, C. Beaugeant, and P. Vary, "Do we need dereverberation for hand-held telephony?" in *Proc. Int. Congress on Acoustics (ICA), Sydney, Australia*, 2010.

[18] D. Baby and H. V. Hamme, "Supervised speech dereverberation in noisy environments using exemplar-based sparse representations," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 156–160.

[19] K. W. Wilson, B. Raj, P. Smaragdis, and A. Divakaran, "Speech denoising using nonnegative matrix factorization with priors," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 4029–4032.

[20] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," *Linguistic data consortium, Philadelphia*, vol. 33, 1993.

[21] 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.