A Constrained MMSE LP Residual Estimator for Speech Dereverberation in Noisy Environments

Chengshi Zheng, *Member, IEEE*, Renhua Peng, Jian Li, and Xiaodong Li

***Abstract*—After revealing that both late reverberation and noise are additive interference components in the residual domain, this paper proposes to suppress these additive interference components by using a constrained minimum mean square error linear prediction (LP) residual estimator, where the optimal filter can be obtained by the generalized singular value decomposition. We propose to estimate the LP residuals for both late reverberation and noise continuously, which is based on the non-VAD related noise power spectral density estimator and the incessant late reverberant spectral variance estimator. The non-intrusive objective measure and the PESQ show that the proposed algorithm is better than traditional LP residual-based algorithms and spectral subtraction-based algorithms.**

***Index Terms*—Additive, linear prediction residual, noise reduction, speech dereverberation.**

I. Introduction

SPEECH dereverberation and noise reduction are extremely important for hands-free speech communication systems, especially when the desired talker is far away from the microphone in a closed room [1]–[5]. This is because that both the reverberation and the noise may seriously reduce speech intelligibility and quality [6], [7]. In the last half-century, lots of researchers have proposed numerous effective algorithms to suppress the late reverberant and the noisy components.

In fact, speech dereverberation and noise reduction can be considered separately or together. Some researchers only deal with the noise, where the reverberant components may be still preserved even when the noise components have already been removed thoroughly [8]–[22]. Some only handle the reverberation in noise-free environments, where the noise components may seriously degrade the performance of some dereverberation algorithms [23]–[30]. Others propose joint denoising and dereverberation techniques to suppress both the noise and the reverberant components [31], [32].

This paper focuses on linear prediction residual estimator (LPRE) for speech dereverberation in noisy environments. To the best of our knowledge, the traditional LPRE-based algorithms could not suppress the reverberation and the noise simultaneously. In [21] and [22], only the additive noise components are removed. Yegnanarayana *et al.* propose to selectively enhance the linear prediction (LP) residuals related to the high signal-to-noise-ratio (SNR) regions in the noisy speech [21], while Jin and Scordilis propose to estimate the LP residuals by using a constrained optimization criterion [22]. In [28] and [29], Yegnanarayana *et al.* suppress the LP residuals related to low signal-to-reverberant component ratio (SRR) regions via entropy weighting and Hilbert envelope weighting, respectively. The performance of these existing algorithms may degrade when both the noise and the reverberation are presented in practical situation.

Generalized singular value decomposition (GSVD) has already been widely used in speech enhancement since Doclo and Moonen proposed this method in 2002 [17]. Most of GSVD-based algorithms are proposed to reduce the noise in the time/frequency domain directly and others are applied to dereverberate the speech signal by estimating the transfer functions from the desired talker to the multiple microphones. This paper introduces a novel speech dereverberation and noise reduction algorithm in the LP residual domain by using a constrained minimum mean square error (MMSE) LPRE, which is based on the fact that both late reverberation and noise are additive components in the residual domain. To implement the constrained MMSE GSVD-based LPRE (CMMSE-GSVD-LPRE) algorithm in practice, we need to estimate the LP residuals continuously for both the reverberation and the noise, where a non-VAD noise power spectral density (NPSD) estimator and an incessant late reverberant spectral variance (LRSV) estimator are proposed to achieve this goal.

The remainder of this paper is organized as follows. Section II formulates the problem. Section III presents the proposed algorithm, and the detailed procedure that estimates the LP residuals for both late reverberation and noise is also presented in this section. Experimental results and conclusions are presented in Section IV and Section V, respectively.

II. Problem Formulation

In noisy and reverberant environments, the microphone signal can be given by:

$$\begin{aligned} x(n) &= h(n) \otimes s(n) + d(n) \\ &= \sum\nolimits_i h(i)s\left(n-i\right) + d(n), \end{aligned} \tag{1}$$

where $s(n)$ is the clean speech and $h(n)$ is the transfer function from the clean speech to the microphone. $\otimes$ is the convolution operator and $d(n)$ is the noise in the microphone. For a causal system, $h(i) = 0$ for $i < 0$ should hold true.

Manuscript received April 19, 2014; revised July 12, 2014; accepted July 13, 2014. Date of publication July 17, 2014; date of current version July 30, 2014. This work was supported in part by the National Science Fund of China (NSFC) under Grants 61201403 and 61302126 and in part by the Tri-Networks Integration Grant KGZD-EW-103-5(3). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rita Singh.

The authors are with the Communication Acoustics Laboratory, Institute of Acoustics, Chinese Academy of Science, Beijing 100190, Beijing, and also with the Acoustics and Information Technology Laboratory, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, Shanghai (e-mail: cszheng@mail.ioa.ac.cn; pengrenhua@mail.ioa.ac.cn; lijian@mail.ioa.ac.cn; lxd@mail.ioa.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2340396

Eq. (1) can be rewritten as:

$$x(n) = s_{\rm early}(n) + s_{\rm late}(n) + d(n), \tag{2}$$

where $s_{\rm early}(n) = \sum_{i=0}^{D-1} h(i)s(n-i)$ and $s_{\rm late}(n) = \sum_{i=D}^{\infty} h(i)s(n-i)$ are the direct-and-early reflection of the speech and the late reflection of the speech, respectively. D defines the filter length of the early reflection. In this paper, only the late reverberation and the noise are considered to be removed from the microphone signal, while the early reflection of the speech will still be preserved.

Similar to the noise, the late reverberation is also additive component in the time domain. Therefore, the late reverberation can be removed by using the same algorithms that are proposed to suppress the noise, where the only problem is that the LRSV is highly non-stationary [23].

By using the LP model, $x(n)$ can be given by:

$$x(n) = \sum\nolimits_{p=1}^{P} a_p^x x\left(n-p\right) + r_x(n), \tag{3}$$

where P is the order of the LP model and a_p^x, with $p = 1, \cdots, P$ are the AR coefficients of $x(n)$. $r_x(n)$ is the LP residual of $x(n)$. Assuming that a_p^s and $a_p^{s_{\rm early}}$ are the AR coefficients of $s(n)$ and $s_{\rm early}(n)$, respectively. Using statistical room acoustics (SRA) theory, Gaubitch *et al.* prove that $E\{a_p^x\} = E\{a_p^s\}$ holds true [30]. Note that $E\{a_p^x\} = E\{a_p^s\}$ holds true if and only if $d(n) = 0$. In noisy environments, we use an approximate assumption:

$$E\left\{a_p^x\right\} \approx E\left\{a_p^s\right\} \approx E\left\{a_p^{s_{\rm early}}\right\}, \tag{4}$$

where (4) is a common assumption that for the traditional LP-based algorithms [5], [22], [21], [28], [29], [30], although Huang *et al.* have already pointed out that the assumption in (4) is not accurate enough [4, Ch. 46]. Note that, although a_p^x may not represent the all-pole filter of the clean speech accurately, we still can use it to reconstruct the enhanced speech based on the previous study results [21], [33]. Note that speech quality may be improved if a_p^s can be optimally estimated from $x(n)$ [34].

Using the same AR coefficients $a_p^x|_{p=1,\cdots,P}$ on each term of the right side in (2), we have:

$$r_x(n) = r_{s_{\rm early}}(n) + r_{s_{\rm late}}(n) + r_d(n), \tag{5}$$

where $r_{s_{\rm early}}(n)$, $r_{s_{\rm late}}(n)$ and $r_d(n)$ are, respectively, the LP residuals of $s_{\rm early}(n)$, $s_{\rm late}(n)$ and $d(n)$. Traditional LP-based algorithms could not remove $r_{s_{\rm late}}(n)$ and $r_d(n)$ simultaneously via the same criterion, since the late reverberation and the noise have significantly different influences on the LP residuals, which can be found from the empirical studies presented in [21] and [28].

To extend the traditional LP algorithms, this paper proposes a CMMSE-GSVD-LPRE algorithm to suppress both the late reverberation and the noise. A detailed description of the proposed algorithm is presented in the next section.

III. Proposed Algorithm

Before we describe the proposed CMMSE-GSVD-LPRE algorithm, (5) is rewritten as:

$$r_x(n) = r_{\rm des}(n) + r_{\rm int}(n), \tag{6}$$

where $r_{\rm des}(n) = r_{s_{\rm early}}(n)$ is the desired LP residual that needs to be preserved, and $r_{\rm int}(n) = r_{s_{\rm late}}(n) + r_d(n)$ is the interference LP residual that needs to be suppressed. There are three parts in this section. In the first part, the CMMSE-GSVD-LPRE algorithm is presented under the assumption that $r_{\rm int}(n)$ is a *priori*. In the second part, $r_{\rm int}(n)$ is estimated frame by frame. We summarize the implementation steps in the last part.

A. CMMSE-GSVD-LPRE Algorithm

For the λth frame, the LP residuals are:

$$r_{x|{\rm des}|{\rm int}}\left(\lambda, \mu\right) = r_{x|{\rm des}|{\rm int}}\left(\lambda M + \mu\right), \tag{7}$$

where $r_{x|{\rm des}|{\rm int}}(\lambda, \mu)$ means $r_x(\lambda, \mu)$, $r_{\rm des}(\lambda, \mu)$ and $r_{\rm int}(\lambda, \mu)$ for compact notation. M is the frame shift and $\mu = 0, 1, \cdots, N-1$ with N the frame length. We further define $\Xi = \{x, {\rm des}, {\rm int}\}$ in the paper for simplicity.

The Hankel-form sample matrices of the LP residuals in the λth frame can be given by:

$$\mathbf{H}_{r_\Xi}(\lambda) = \begin{bmatrix} r_\Xi(\lambda M) & r_\Xi\left(\lambda M+1\right) & \cdots & r_\Xi\left(\lambda M+K-1\right) \\ r_\Xi\left(\lambda M+1\right) & r_\Xi\left(\lambda M+2\right) & \cdots & r_\Xi\left(\lambda M+K\right) \\ \cdots & \vdots & \ddots & \vdots \\ r_\Xi\left(\lambda M+L-1\right) & & \cdots & r_\Xi\left(\lambda M+N-1\right) \end{bmatrix}, \tag{8}$$

where $L = N - K + 1$.

To estimate $\mathbf{H}_{r_{\rm des}}(\lambda)$ in the constrained MMSE sense, the optimal filter is designed by:

$$\mathbf{W}_{\rm opt}(\lambda) = \min_{\mathbf{W}(\lambda)} \left\{ \left\| \mathbf{H}_{r_{\rm des}}(\lambda) - \mathbf{H}_{r_x}(\lambda)\mathbf{W}(\lambda) \right\|_F^2 \right\}$$
$$\text{subject to :} E\left\{ \left| \mathbf{w}_{{\rm opt,k}}^{\#}(\lambda)\mathbf{r}_{\rm int}(\lambda) \right|^2 \right\} \le \alpha_k \sigma_{r_{{\rm int},k}}^2(\lambda), \tag{9}$$

where $\mathbf{W}_{\rm opt}(\lambda) = [\,\mathbf{w}_1(\lambda) \;\; \mathbf{w}_2(\lambda) \;\; \cdots \;\; \mathbf{w}_K(\lambda)\,]$, $\mathbf{W}_{\rm opt}(\lambda) \in \mathfrak{R}^{\mathfrak{K}\times\mathfrak{K}}$, and ${\rm Rank}(\mathbf{H}_{\rm des}(\lambda)) = K$. $\sigma_{r_{{\rm int},k}}^2(\lambda)$ is the square of the kth singular value of $\mathbf{H}_{r_{\rm int}}(\lambda)$.

By using GSVD algorithm, $\mathbf{W}_{\rm opt}(\lambda)$ can be given by:

$$\mathbf{W}_{\rm opt}(\lambda) = \mathbf{X}^{-1}(\lambda)\mathbf{Q}(\lambda)\mathbf{X}(\lambda), \tag{10}$$

where

$$\mathbf{U}^T(\lambda)\mathbf{H}_{r_x}(\lambda)\mathbf{X}(\lambda) = {\rm diag}\left\{ \sigma_{r_{x,1}}(\lambda) \;\; \cdots \;\; \sigma_{r_{x,K}}(\lambda) \right\}, \tag{11}$$

$$\mathbf{V}^T(\lambda)\mathbf{H}_{r_{\rm int}}(\lambda)\mathbf{X}(\lambda) = {\rm diag}\left\{ \sigma_{r_{{\rm int},1}}(\lambda) \;\; \cdots \;\; \sigma_{r_{{\rm int},K}}(\lambda) \right\} \tag{12}$$

and $\mathbf{Q}(\lambda) = {\rm diag}\{\, q_{11}(\lambda) \;\; q_{22}(\lambda) \;\; \cdots \;\; q_{KK}(\lambda)\,\}$, where $q_{kk}(\lambda)$, with $k = 1, \cdots, K$, are determined by the constraint in (9). $\mathbf{U}(\lambda)$ and $\mathbf{V}(\lambda)$ are two orthogonal matrices, and $\mathbf{X}(\lambda)$ is an invertible matrix. By using the same derivation method in ([19, (34)-(45)]), we have:

$$q_{kk}(\lambda) = \alpha_k^{1/2}. \tag{13}$$

We propose to use a more aggressive suppression gain function that was first proposed in [19, (44)], which is:

$$\alpha_k = \exp\left\{ -\frac{\nu\sigma_{r_{{\rm int},k}}^2(\lambda)}{\max\left\{ \sigma_{r_{x,k}}^2(\lambda) - \sigma_{r_{{\rm int},k}}^2(\lambda), \sigma_{\min}^2 \right\}} \right\}, \tag{14}$$

where $\sigma^2_{\min}$ is a small positive value avoiding division by zero and the typical value of ν ranges from 1 to 5, where the larger value ν, the more amounts of reduction and speech distortion.

Note that although (14) has nearly the same expression as [19, (44)]. There also exist two obvious differences. First, we don't assume that the interference LP residuals are white, while the noise is assumed to be white in [19]. In other words, (14) can deal with colored interference, which is more practical. Second, the LP residuals are applied to obtain the optimal filter, while the time-domain signals are directly used to obtain the optimal filter in [19]. Many researchers have already pointed out that noise reduction in the residual domain is better than that in the time/frequency domain [21], [22]. This paper also show that the late reverberation and the noise can be better suppressed by using the residual domain than by using the time/frequency domain.

As shown in (9)–(14), we needs to estimate the Hankel matrix of the interference LP residuals. Since the late reverberation is highly non-stationary, it is impossible to estimate the LRSV in speech-absent segments even assuming that the noise is stationary. In the following part, we propose to estimate $\mathbf{H}_{r_{\rm int}}(\lambda)$ in an efficient way.

B. Estimation of the Interference LP Residual

As mentioned above, we need to estimate the interference LP residual frame by frame since the late reverberation is generally non-stationary and the noise may also be non-stationary. Note that most of traditional algorithms use the speech-absent segments to estimate the noise covariance matrix for noise reduction [17], [19], [22].

In this part, the NPSD and the LRSV are estimated separately frame by frame. After that, we use the overlap-add method to obtain the time-domain interference signal and its corresponding LP residual.

1) Estimation of the NPSD: As a matter of fact, both the VAD-based and the non-VAD-based NPSD estimators can be applied in estimating the NPSD. This paper proposes to use the unbiased MMSE NPSD estimator, which is proposed by Gerkmann and Hendriks in [11], for its simplicity and accuracy. If ignoring the computational load, the improved version of the unbiased MMSE NPSD estimator can be used in practice, which is proposed in [13]. We assume that the estimated NPSD in the λth frame is $\hat{\sigma}^2_D(\omega,\lambda)$, where ω is the frequency bin index.

2) Estimation of the LRSV: Lots of LRSV estimators have already been proposed in recent years [23]–[27], [32]. Some of these LRSV estimators need to estimate the reverberation time in the time/frequency domain [23], [25], [26]. Others do not estimate any explicit parameters in estimating the LRSV [24], [27]. This paper proposes to use the LRSV estimator proposed by Wu and Wang in [24] for its simplicity and efficiency. We assume that the estimated LRSV in the λth frame is $\hat{\sigma}^2_{S_{\rm late}}(\omega,\lambda)$. Note that the LRSV is estimated after removing the noise, which could reduce the influence of the noise on estimating the LRSV.

3) Estimation of the Interference LP Residual: After estimating the NPSD and the LRSV, the interference power spectral density (IPSD) can be given by:

$$\hat{\sigma}^2_{\rm INT}(\omega,\lambda) = \min\left\{\alpha\hat{\sigma}^2_D(\omega,\lambda) + \beta\hat{\sigma}^2_{S_{\rm late}}(\omega,\lambda), \sigma^2_X(\omega,\lambda)\right\}, \quad (15)$$

TABLE I
REAL TIME IMPLEMENTATION OF THE PROPOSED ALGORITHM

1) Calculate the FFT of $x(n)$ in the λth frame to obtain $X(\omega,\lambda)$;
2) Estimate the NPSD by using the unbiased MMSE NPSD estimator;
3) Estimate the LRSV by using the method presented in [24];
4) Estimate the IPSD by using (15);
5) Estimate the time-domain interference by using (16);
6) Estimate the LP coefficients of $x(n)$ in the λth frame, $a^x_p(\lambda)$, with $p = 1, \cdots, P$, by using Levinson-Durbin recursion;
7) Estimate the interference LP residual by using (17);
8) Use (7)-(14) to obtain the optimal filter $\mathrm{W}_{\rm opt}(\lambda)$;
9) Reconstruct the desired LP residual from the Hankel-form matrix, $\hat{\mathbf{H}}_{r_{\rm des}}(\lambda) = \mathbf{H}_{r_x}(\lambda)\,\mathrm{W}_{\rm opt}(\lambda)$;
10) Synthesize the enhanced speech by using $\hat{r}_{\rm des}(\lambda M+\mu)$ and $a^x_p(\lambda)$.

where $\alpha = \beta = 1$. $\sigma^2_X(\omega,\lambda)$ is the smoothed version of the raw periodogram of $x(n)$ in the ωth bin of the λth frame, which is applied to avoid overestimating the IPSD in practice.

The time-domain interference can be estimated by:

$$\hat{x}_{\rm int}(n) = \mathrm{IFFT}\left\{\frac{\sqrt{\hat{\sigma}^2_{\rm INT}(\omega,\lambda)}}{|X(\omega,\lambda)|}X(\omega,\lambda)\right\}, \quad (16)$$

where $X(\omega,\lambda)$ is the FFT of $x(n)$ in the ωth bin of the λth frame. To obtain $\hat{x}_{\rm int}(n)$ for all values of n, the overlap-add method should be used frame by frame.

Assuming that the LP coefficients of $x(n)$ in the λth frame are $a^x_p(\lambda)$, with $p = 1, \cdots, P$. The interference LP residual can be given by:

$$\hat{x}_{\rm int}(\lambda M+\mu) = \hat{r}_{\rm int}(\lambda M+\mu) + \sum_{p=1}^{P} a^x_p(\lambda)\hat{x}_{\rm int}(\lambda M+\mu-p), \quad (17)$$

where $\hat{r}_{\rm int}(\lambda M+\mu)$ is an estimate of $r_{\rm int}(n)$ in the λth frame.

C. Real Time Implementation

We summarize the proposed algorithm in Table I. This table clearly indicates that the proposed algorithm is casual, which can be implemented frame by frame.

We want to emphasize that the proposed CMMSE-GSVD-LPRE algorithm can also be applied to reduce only the noise or only the reverberation in a convenient way. If $\alpha = 1$ and $\beta = 0$ in (15), the proposed algorithm could only remove the noise components. While $\alpha = 0$ and $\beta = 1$ in (15), the proposed algorithm could only remove the reverberant components. Compared with the traditional LP residual-based algorithms, the proposed CMMSE-GSVD-LPRE algorithm could suppress both the noise and the reverberation in a unified criterion.

IV. PERFORMANCE EVALUATION

This section evaluates the performance of the proposed algorithm and compares it with the spectral subtraction (SS)-based algorithm in [24] and the traditional LP residual-based algorithm in [21]. The SS-based algorithm can easily be extended to suppress both the noise and the reverberation, although Wu and Wang only consider the reverberation in [24]. However, the

TABLE II
COMPARISON RESULTS OF THE THREE ALGORITHMS INCLUDING THE SS, THE LP-SD, AND THE CMMSE-GSVD-LPRE IN THE NOISE-FREE CASE

	SRMR				PESQ			
T_{60} [ms]	400	600	800	1000	400	600	800	1000
Noisy	2.19	1.67	1.42	1.30	2.53	2.35	2.25	2.18
SS	3.24	2.59	2.26	2.09	2.56	2.42	2.33	2.27
LP-SD	3.16	2.63	2.26	2.04	**2.57**	2.35	2.27	2.18
Proposed	**3.80**	**3.20**	**2.87**	**2.68**	2.56	**2.45**	**2.38**	**2.31**

TABLE III
COMPARISON RESULTS OF THE THREE ALGORITHMS IN THE REVERBERANT AND NOISY ENVIRONMENT

	SRMR				PESQ			
SNR [dB]	-5	5	15	25	-5	5	15	25
Noisy	0.51	1.40	2.05	2.17	1.56	1.83	2.28	2.43
SS	1.43	2.82	3.18	3.23	1.58	2.09	2.40	**2.50**
LP-SD	0.81	2.00	2.98	3.15	1.53	1.82	2.27	2.43
Proposed	**2.59**	**3.72**	**3.82**	**3.80**	**1.61**	**2.19**	**2.41**	2.49

traditional LP-based noise reduction algorithm in [21] fails to suppress the reverberation in most cases due to that the additive noise and the reverberation have significantly different impacts on spectral flatness characteristics [28]. In this paper, the algorithm presented in [24] will be referred as SS, while the algorithm presented in [28] will be referred as LP-SD, where SD is short for speech dereverberation.

We evaluate the three algorithms in two different environments including the noise-free, the reverberant and noisy environments. The clean speech samples are taken from the TIMIT database [35], while the noise samples are taken from the NOISEX92 database [36]. To obtain the reverberant speech signals, we need to generate the simulated room impulse responses (RIRs) using the image method [37]. In all the results, the simulated rectangular room with dimensions $[5 \times 4 \times 3]$(m) is used and all six wall surfaces of this room have the same reflection coefficient. Based on Sabine's reverberation equation, different reverberation time T_{60} can be achieved by properly choosing the value of the reflection coefficient. To give quantitative comparisons, the speech to reverberation modulation energy ratio (SRMR) is selected as a non-intrusive measure since it can measure the perceived amount of reverberation efficiently [38], [39], [40]. We also use the PESQ score to compare the three algorithms for its high correlation with the perceived amount of reverberation and its wide application [40], [41]. The comparison results are presented in the following two parts.

A. Comparison in the Noise-Free Environment

By properly setting the value of the reflection coefficient, the reverberation time ranging from 400 ms to 1000 ms is considered to evaluate the performances of the three algorithms. The results are presented in Table II. As shown in this table, all of the three algorithms can improve the values of the SRMR, while the proposed algorithm has the largest values of the SRMR among the three algorithms. The PESQ scores in Table II also show that the proposed algorithm has the largest values in most cases.

B. Comparison in the Reverberant and Noisy Environment

In this part, both the noise and the reverberation are considered to show the validity of the proposed algorithm in the reverberant and noisy environment. The reverberation time $T_{60} = 400$ ms is a fixed value, and only the white Gaussian noise is added to the reverberant speech in this paper for the space limitation, where the input SNR ranges from -5 dB to 25 dB. The comparison results are presented in Table III. The same as Table II, the proposed CMMSE-GSVD-LPRE has the largest values of the SRMR among the three algorithms. Notice should be given that the LP-SD algorithm degrades its performance

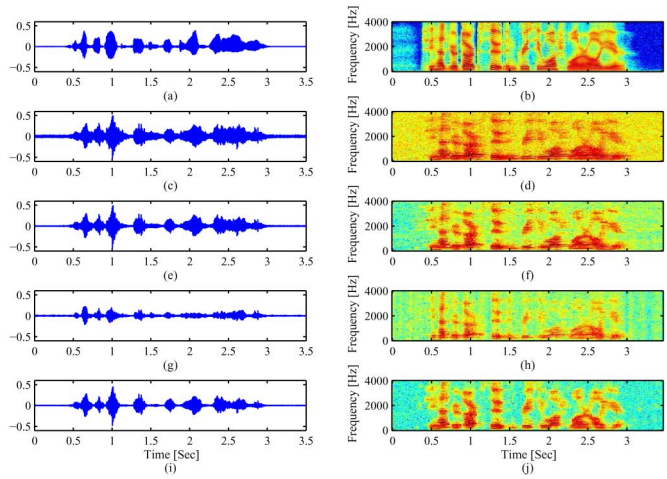

Fig. 1. Waveforms and spectrograms of (a)(b) the clean speech, (c)(d) the reverberant and noisy speech with $T_{60} = 400$ ms and the input $\mathrm{SegSNR} = 15$ dB, (e)(f) enhanced by SS, (g)(h) enhanced by LP-SD, (i)(j) enhanced by the proposed approach.

significantly especially when the input SNR is extremely low. Table III also shows that the LP-SD could not improve the PESQ score in noisy environments in all cases, while both the SS and the proposed algorithms can improve the PESQ score.

By studying the periodograms of the enhanced speech signals by the three algorithms as shown in Fig. 1, we find that the SS contains lots of audible *musical noise* components, while the LP-SD could not remove the noise in most cases and only the late reverberation is partially removed. Compared with the LP-SD, the proposed CMMSE-GSVD-LPRE can remove both the noise and the reverberation. Compared with the SS, the proposed algorithm has less speech distortion and does not suffer from serious *musical noise* problem.

V. CONCLUSION

This paper proposes to suppress both the noise and the reverberation in the residual domain, where a constrained MMSE GSVD-LPRE algorithm is proposed to enhance the LP residual. Experimental results verify the validity of the proposed algorithm. Further work should concentrate on studying the influence of the NPSD estimator and the LRSV estimator on the proposed algorithm to further improve its performance.

ACKNOWLEDGMENT

The authors appreciate the Associate Editor and the three reviewers for their valuable comments in improving this letter. The authors also would like to thank Dr. Jinqiu Sang for proofreading this letter.

REFERENCES

[1] M. Brandstein and D. Ward, *Microphone arrays: Signal processing techniques and applications*. Berlin, Germany: Springer-Verlag, 2001.

[2] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*. Berlin, Germany: Springer-Verlag, 2005.

[3] P. C. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL, USA: CRC, 2013.

[4] J. Benesty, M. M. Sondhi, and Y. Huang, *Springer Handbook of Speech Processing*. Berlin, Germany: Springer-Verlag, 2007.

[5] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*. London, U.K.: Spinger-Verlag, 2010.

[6] D. N. Kalikow, K. N. Stevens, and L. L. Elliott, "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Amer.*, vol. 61, pp. 1337–1351, 1977.

[7] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech in auditoria," *J. Acoust. Soc. Amer.*, vol. 77, pp. 1069–1077, 1985.

[8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.

[9] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP–32, no. 6, pp. 1109–1121, Dec. 1984.

[10] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.

[11] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[12] J. Taghia, J. Taghia, N. Mohammadiha, J. Sang, V. Bouse, and R. Martin, "An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 4640–4643.

[13] X. Hu, S. Wang, C. Zheng, and x. Li, "A cepstrum-based preprocessing and postprocessing for speech enhancement in adverse environments," *Appl. Acoust.*, vol. 74, no. 12, pp. 1458–1462, 2013.

[14] I. Cohen, "Relaxed statistical model for speech enhancement and *a priori* SNR estimation," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 870–881, Sep. 2005.

[15] Y. Hu and P. C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Trans. Speech Audio Process*, vol. 12, no. 1, pp. 59–67, Jan. 2004.

[16] P. C. Loizou, G. Kim, and G. , "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," *IEEE Trans. Audio, Speech, Lang. Process*, vol. 19, no. 1, pp. 47–56, Jan. 2011.

[17] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[18] G. H. Ju and L. S. Lee, "A perceptually constrained GSVD-based approach for enhancing speech corrupted by colored noise," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 119–134, Jan. 2007.

[19] Y. Ephraim and H. L. V. Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[20] Y. Hu and P. C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," *IEEE Signal Process. Lett.*, vol. 9, no. 7, pp. 204–207, Jul. 2002.

[21] B. Yegnanarayana, C. Avendano, H. Hermansky, and P. S. Murthy, "Speech enhancement using linear prediction residual," *Speech Commun.*, vol. 28, pp. 25–42, 1999.

[22] W. Jin and S. Scordilis, "Speech enhancement by residual domain constrained optimization," *Speech Commun.*, vol. 48, pp. 1349–1364, 2006.

[23] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acust. United with Acust.*, vol. 87, no. 3, pp. 359–366, 2001.

[24] M. Wu and D. Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 3, pp. 774–784, May 2006.

[25] E. A. P. Habets, "Single- and multi-microphone speech dereveration using spectral enhancement," Ph.D. dissertation, Technische Univ. Eindhoven, Eindhoven, The Netherlands, Jun. 25, 2007.

[26] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 770–773, Sep. 2009.

[27] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Suppression of late reverberation effect on speech using long-term multiple-step linear prediction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 4, pp. 1–12, May 2009.

[28] B. Yegnanarayana and P. S. Murthy, "Enhancement of reverberant speech using LP residual signal," *IEEE Trans. Speech Audio Process*, vol. 8, no. 3, pp. 267–281, May 2000.

[29] B. Yegnanarayana, S. R. Mahadeva Prasanna, and K. Sreenivasa Rao, "Speech enhancement using excitation source information," in *Proc. IEEE Int. Conf. Audio, Speech, Signal Process.*, 2002, vol. 1, pp. 541–544.

[30] N. D. Gaubitch, P. A. Naylor, and D. B. Ward, "On the use of linear prediction for dereverberation of speech," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2003, vol. 1, pp. 99–102.

[31] S. Doclo, "Multimicrophone noise reduction and dereverberation techniques for speech applications," Ph.D. dissertation, Dept. Elect. Eng., Katholieke Univ. Leuven, Leuven, Belgium, May 2003.

[32] S. Mosayyebpour, M. Esmaeili, and T. A. Gulliver, "Single-microphone early and late reverberation suppression in noisy speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 2, pp. 322–335, Feb. 2013.

[33] C. Hamon, E. Moulines, and F. J. Charpentier, "A diphone synthesis system based on time domain prosodic modifications of speech," in *Proc. IEEE Int. Conf. Audio, Speech, Signal Process.*, 1989, pp. 238–241.

[34] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 3, pp. 197–210, Jun. 1978.

[35] J. S. Garofolo, "Getting started with the DARPA TIMIT CD-ROM: An acoustic-phonetic continuous speech database," Nat. Inst. of Standards and Technology (NIST). Gaithersburg, MD, USA, 1993.

[36] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognnition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, pp. 247–251, 1993.

[37] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[38] T. H. Falk, C. Zheng, and W. Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Trans. Audio, Speech, Lang. Process,*, vol. 18, no. 7, pp. 1766–1774, Sep. 2010.

[39] [Online]. Available: http://reverb2014.dereverberation.com/download.html

[40] K. Kinoshita, M. Delcroix, T. Yoshioka, E. Habets, R. H-Umbach, V. Leutnant, A. Sehr, W. Kellermann, R. Maas, S. Gannot, and B. Raj, "Summary of the REVERB chanllenge," in *Proc. REVERB Workshop Int. Conf. Audio, Speech, Lang. Process.*, Florance, Italy, May 10, 2014 [Online]. Available: http://reverb2014.dereverberation.com/workshop/slides/reverb-summary.pdf

[41] *Perceptual evaluation of speech quality: An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, P.862, Int. Telecomm. Union, 2001.