Single-Channel Non-Causal Speech Enhancement to Suppress Reverberation and Background Noise

Myung-Suk Song[†] and Hong-Goo Kang

School of Electrical and Electronic Engineering, Yonsei University
(Received August 14, 2012; revised October 5, 2012; accepted October 22, 2012)

ABSTRACT: This paper proposes a speech enhancement algorithm to improve the speech intelligibility by suppressing both reverberation and background noise. The algorithm adopts a non-causal single-channel minimum variance distortionless response (MVDR) filter to exploit an additional information that is included in the noisy-reverberant signals in subsequent frames. The noisy-reverberant signals are decomposed into the parts of the desired signal and the interference that is not correlated to the desired signal. Then, the filter equation is derived based on the MVDR criterion to minimize the residual interference without bringing speech distortion. The estimation of the correlation parameter, which plays an important role to determine the overall performance of the system, is mathematically derived based on the general statistical reverberation model. Furthermore, the practical implementation methods to estimate sub-parameters required to estimate the correlation parameter are developed. The efficiency of the proposed enhancement algorithm is verified by performance evaluation. From the results, the proposed algorithm achieves significant performance improvement in all studied conditions and shows the superiority especially for the severely noisy and strongly reverberant environment.

Key words: Single channel speech enhancement, Dereverberation, Non-causal

ASK subject classification: Speech Signal Processing (2)

I. Introduction

In speech signal processing systems, such as voice-controlled systems, hands-free mobile telephones, and hearing aids, the received microphone signals are generally contaminated by environmental artifacts.[1] The detrimental effects, such as background noise, interfering signal, and channel distortion, degrade the overall performance of the system.[2] Many researchers are still struggling to remove the undesired components from the acquired input signal.

One critical obstacle is reverberation caused by a multi-path propagation of an acoustic sound from its source to microphone.[3-5] An acoustic channel between the source and the microphone can be described by the acoustic impulse response (AIR) and it can be divided into three segments: a direct path, early reflections, and late reverberation.[1] While the early reflections, which is the combination of the direct and the early reflections, impacts only to the coloration of the speech, late reverberation causes the lengthening of speech phonemes. Consequently, the previous phonemes overlap following phonemes, so that it results in speech intelligibility (and also recognition rate) degradation.[2]

A number of techniques have been proposed to reduce the detrimental effects of the reverberation. If the AIR is known as a priori knowledge, the dereverberation can be ideally achieved by taking an inverse filtering such as multiple input/output inverse theorem (MINT).[6-8] The problem of speech dereverberation in unknown acoustic environments has also received a lot of attention. Cepstrum based decon-

†**Corresponding author:** Myung-Suk Song (earth112@dsp.yonsei.ac.kr)
School of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, Republic of Korea
(Tel: 82-2-2123-4534; Fax: 82-2-364-4870)

volution techniques utilize the idea that deconvolution in the time domain is identical to subtraction in the cepstrum domain.[9-12] Methods to enhance a residual of linear prediction (LP) filtering have been introduced.[13,14] An algorithm to employ the harmonic structure of speech, which is called harmonicity based dereverberation (HERB), has been also proposed.[15-17]

The spectral enhancement technique is known as the most famous approach for single channel dereverberation techniques.[4,18-20] The spectral enhancement based dereverberation approaches have been developed to reduce the late reverberation or in other words to estimate the early speech component from the acquired input signal. They are derived under the assumption that the early speech component and the late reverberation are uncorrelated, and the processings are commonly performed in the frequency domain by estimating the late reverberation spectral variance (LRSV).

Several techniques have been proposed to suppress both the reverberation and the noise.[21] In Habets' approach, the noise power spectrum is first suppressed, and then the LRSV is obtained from the denoised reverberant speech.[4,22] The output signal is obtained by applying a spectral enhancement. Erkelens et al. proposed a late reverberation suppression rule in noisy and reverberant environments by exploiting the longterm correlation coefficients between the current reverberant spectrum and enhanced ones in the previous frames.[23] They extended their works to design the suppression rule in noisy and non-stationary acoustical environments by assuming that the AIR has time varying characteristic.[24]

In,[25] an efficient dereverberation algorithm was introduced and verified its superiority in reverberant environment. The basic idea of the algorithm was that the reverberant signals in the following frames contain the desired speech at the current frame since the desired speech at the current frame is convolved

with a relatively long time interval of the AIR in reverberant environment. The algorithm decomposed the observed reverberant signal into a component correlated to the desired speech signal and interference that did not have correlation with the desired signal. The non-causal filter minimizes interference, while maintaining speech quantity by gathering the correlated component.

In this paper, the single-channel non-causal enhancement algorithm to suppress both reverbeartion and background noise is proposed. The dereverberation algorithm in[25] is extended to enhance the desired speech signal in noisy reverberant speech. The proposed algorithm utilizes a non-causal MVDR filter to exploit the correlation information that lies in subsequent frames. The noisy-reverberant signals are decomposed into the parts of the desired signal and the interference that does not correlated to the desired signal. The interference consists of two different components. One is a reverberant interference that is the reverberant signal and uncorrelated to the desired speech signal. The other is additive noise interference that is assumed to be uncorrelated with the desired speech. Then, the filter equation is derived based on the MVDR criterion to minimize the residual interference without bringing speech distortion.

The late reverberation and additive noise are suppressed by estimating correlation coefficient, which is the main parameter to determine the overall performance of the proposed algorithm. The correlation parameter is derived by employing a statistical reverberant model, composed of the late reverberation spectral variance and another sub-parameters. The efficient method to estimate the correlation parameter including the sub-parameters are described and practically implemented.

The rest of this paper is organized as follows. The problem is formulated in section 2. Here, the observed noisy reverberant signal is decomposed into three uncorrelated components: the part correlated to the

desired signal, reverberant interference, and noise interference. The non-causal single-channel MVDR filter to suppress the noise-plus-reverberant inter-ference is derived in section 3. In section 4, the complete algorithm to estimate the early speech component is described. The correlation parameter is derived using the statistical reverberation model and implemented by using estimates of sub-parameters. Section 5 demonstrates the performance evaluation and the summary follows in section 6.

II. Problem formulation

The observed noisy reverberant signal is assumed to be first convolved with the acoustic transfer function (ATF) $H(k, m)$, and then corrupted by the uncor-related noise $V(k, m)$, as follows

$$Y(k,m) = Z(k,m) + V(k,m), \qquad (1)$$

with

$$Z(k,m) = \sum_{l=0}^{\infty} H(k,l)S(k,m-l). \qquad (2)$$

Where $H(k, 0) = 1$ and k and m mean frequency-bin and time-frame, respectively. The speech signal $S(k, m)$ is assumed to be uncorrelated to the speech signal at different time-frame.

In reverberant environment, the desired signal $S(k, m)$ is first delayed and attenuated by the AIR, and then accumulated into the subsequent reverberant signal $Z(k, m+l)$, $l > 0$. The reverberation components in future frames, which are highly correlated with the desired signal of current frame, should be taken into account in the derivation process of dereverberation algorithms. For that purpose, a non-causal filter is employed as

$$\hat{S}(k,m) = \sum_{l=0}^{L-1} W_l^*(k,m)Y(k,m+l) = \mathbf{w}^H(k,m)\mathbf{y}(k,m), \quad (3)$$

where

$$\mathbf{y}(k,m) = \left[Y(k,m)\cdots Y(k,m+L-1)\right]^T, \qquad (4)$$

$$\mathbf{w}(k,m) = \left[W(k,m)\cdots W(k,m+L-1)\right]^T, \qquad (5)$$

are the vector of the observed signal and the vector of the filter coefficient, relatively. The filter order L is required to be determined based on reverberant level, that is the reverberation time RT_{60}. It is obvious that the large value of L promises ideal performance. However, the limitation such as complexity in real application forces to choose appropriate value. In this paper, L is chosen to be 12 (i.e., 48 ms).

The observed signal $Y(k, m+l)$ contains two parts; a part that has a correlation with the desired signal $S(k, m)$ and a component that is uncorrelated to $S(k, m)$. Precisely, the observed signal vector $\mathbf{y}(k, m)$ contains :

1-1) The desired signal $S(k, m)$ itself.

1-2) The reverberant components that are correlated with $S(k, m)$. $H(k, 1)S(k, m)$ contained in $Y(k, m+1)$ is a good example.

2-1) The reverberant components that do not have correlation with the desired signal $S(k, m)$. This category contains undesired speech signals in subsequent frames $S(k, m+l)$, $l > 0$ and the reverberant signal caused by all undesired speech signals at the earlier time-frames. For instance, the rest components of $Y(k, m+1)$ except for $H(k, 1)S(k, m)$, such as the speech signal $S(k, m+1)$ and the reverberant signal from earlier time-frame $H(k, 2)S(k, m-1)$, are included in this category.

2-2) The uncorrelated additive noise $V(k, m)$.

The fact above inspires us to decompose the observed noisy reverberant signal into two orthogonal components corresponding to the part of the desired

signal and interference.

$$Y(k, m+l) = \gamma_S^*(k, m, l)S(k, m) + S_{Y}{}'(k, m+l), \quad (6)$$

where

$$\gamma_S(k, m, l) = \frac{E\left[S(k, m)Y^*(k, m+l)\right]}{E\left[\left|S(k, m)\right|^2\right]} \quad (7)$$

is correlation coefficient between the desired signal $S(k, m)$ and the subsequent observed signal $Y(k, m+l)$ and $S_{Y}{}'(k, m+l)$ is the interference signal.[26] Note that the desired signal and interference signal is uncorrelated:

$$E\left[S(k, m)S_{Y}{}'^*(k, m+l)\right] = 0. \quad (8)$$

The interference signal $S_{Y}{}'(k, m+l)$ is a superposition of both the reverberant interference $S_{Z}{}'(k, m+l)$, which is refered as (2-1) in previous page, and the noise interference $S_{V}{}'(k, m+l)$, which is (2-2), such that

$$S_{Y}{}'(k, m+l) = S_{Z}{}'(k, m+l) + S_{V}{}'(k, m+l), \quad (9)$$

where

$$S_{Z}{}'(k, m+l) = Z(k, m+l) - \gamma_S^*(k, m, l)S(k, m) \quad (10)$$

and

$$S_{V}{}'(k, m+l) = V(k, m+l). \quad (11)$$

In,[25] it was described that the observed noise-free reverberant signal was decomposed into the one correlated to the desired signal and the other referred as the interference. A dissimilarity between the decomposition in Eq. (5) and one in,[25] is that the

interference $S_{Y}{}'(k, m+l)$ in Eq. (5) contains not only the uncorrelated reverberant component $S_{Z}{}'(k, m+l)$ but also the additive background noise $S_{V}{}'(k, m+l)$ since the background noise is assumed to be uncorrelated to the desired speech signal.

The observed signal vector $y(k, m)$ is given as

$$\begin{aligned} \mathbf{y}(k, m) &= S(k, m)\gamma_s^*(k, m) + \mathbf{s}_y{}'(k, m) \\ &= \mathbf{s}_d(k, m) + \mathbf{s}_y{}'(k, m) \end{aligned} \quad (12)$$

where the normalized correlation vector $\gamma_s(k, m)$ is

$$\boldsymbol{\gamma}_s(k, m) = \left[\gamma_S(k, m, 0)\cdots\gamma_S(k, m, L-1)\right]^T, \quad (13)$$

$S_d(k, m)$ is the desired signal vector, and

$$\mathbf{s}_y{}'(k, m) = \left[S_{Y}{}'(k, m)\cdots S_{Y}{}'(k, m+L-1)\right]^T \quad (14)$$

denotes the reverberant-plus-noise interference signal vector.

$\mathbf{s}_y{}'(k, m)$ consists of reverberant-interference signal vector $\mathbf{s}_z{}'(k, m)$ and noise-interference signal vector $\mathbf{s}_v{}'(k, m)$ such as

$$\mathbf{s}_y{}'(k, m) = \mathbf{s}_z{}'(k, m) + \mathbf{s}_v{}'(k, m), \quad (15)$$

where

$$\begin{aligned} \mathbf{s}_z{}'(k, m) &= \left[S_{Z}{}'(k, m)\cdots S_{Z}{}'(k, m+L-1)\right]^T, \\ \mathbf{s}_v{}'(k, m) &= \left[S_{V}{}'(k, m)\cdots S_{V}{}'(k, m+L-1)\right]^T. \end{aligned} \quad (16)$$

From Eq. (3) and (12), one can write the estimate $\hat{S}(k, m)$ into the following form:

$$\begin{aligned} \hat{S}(k, m) &= \mathbf{w}^H(k, m)\mathbf{y}(k, m) \\ &= \mathbf{w}^H(k, m)\left[\mathbf{s}_d(k, m) + \mathbf{s}_y{}'(k, m)\right] \\ &= S(k, m)\mathbf{w}^H(k, m)\gamma_s^*(k, m) + \mathbf{w}^H(k, m)\mathbf{s}_y{}'(k, m) \\ &= S_{fd}(k, m) + S_{ri}{}'(k, m), \end{aligned} \quad (17)$$

where $S_{fd}(k, m)$ and $S_{ri}`(k, m)$ are the filtered desired signal and the residual (reverberation-plus-noise) interference, respectively.

By Eq. (15), the residual interference $S'_{ri}(k, m)$ can be rewritten as

$$
\begin{aligned}
S_{ri}`(k,m) &= \mathbf{w}^H(k,m)\mathbf{s}_y{}'(k,m) \\
&= \mathbf{w}^H(k,m)\mathbf{s}_z{}'(k,m) + \mathbf{w}^H(k,m)\mathbf{s}_v{}'(k,m) \quad (18) \\
&= S'_{rr}(k,m) + S'_{rn}(k,m)
\end{aligned}
$$

where $S'_{rr}(k, m)$ and $S'_{rn}(k, m)$ are the the residual reverberation and the residual noise, respectively. From Eq. (17) and Eq. (18), it is observed that the estimate of the desired signal is the sum of three mutually uncorrelated terms, which are the filtered desired signal, the residual reverberant signal, and the residual noise signals.

Therefore, the variance of $\hat{S}(k,m)$ is

$$
\lambda_{\hat{S}}(k,m) = \lambda_{S_{fd}}(k,m) + \lambda_{S_{rr}}(k,m) + \lambda_{S_{rn}}(k,m), \quad (19)
$$

where

$$
\begin{aligned}
\lambda_{S_{fd}}(k,m) &= \lambda_S(k,m)\left|\mathbf{w}^H(k,m)\boldsymbol{\gamma}_s^*(k,m)\right|^2, \\
\lambda_{S_{rr}}(k,m) &= \mathbf{w}^H(k,m)\boldsymbol{\Phi}_{s_z{}'}(k,m)\mathbf{w}(k,m) \\
&= \mathbf{w}^H(k,m)\boldsymbol{\Phi}_z(k,m)\mathbf{w}(k,m) - \lambda_S(k,m)\left|\mathbf{w}^H(k,m)\boldsymbol{\gamma}_s^*(k,m)\right|^2, \quad (20) \\
\lambda_{S_{rn}}(k,m) &= \mathbf{w}^H(k,m)\boldsymbol{\Phi}_v(k,m)\mathbf{w}(k,m),
\end{aligned}
$$

and $\boldsymbol{\Phi}_a(k, m) = E[\mathbf{a}(k, m)\mathbf{a}^H(k, m)]$ is the correlation matrix of $\mathbf{a}(k, m) \in \{\mathbf{s}_z{}'(k, m), \mathbf{z}(k, m), \mathbf{v}(k, m)\}$.

III. Non-causal single-channel MVDR filter

In order to derive the filter coefficients, the error signal between the estimated and desired signals is defined as

$$
\varepsilon(k,m) = \hat{S}(k,m) - S(k,m) = \varepsilon_d(k,m) + \varepsilon_r(k,m), \quad (21)
$$

where

$$
\varepsilon_d(k,m) = S_{fd}(k,m) - S(k,m) \quad (22)
$$

is the signal distortion due to the complex non-causal filter, which is difference between the filtered desired signal $S_{fd}(k, m)$ in Eq. (17) and the desired signal, and

$$
\varepsilon_r(k,m) = S'_{ri}(k,m) = S'_{rr}(k,m) + S'_{rn}(k,m) \quad (23)
$$

represents the residual (reverberation-plus-noise) interferences (See Eq. (17) and Eq. (18)).

The mean-square error (MSE) is then

$$
E\left[\left|\varepsilon(k,m)\right|^2\right] = E\left[\left|\varepsilon_d(k,m)\right|^2\right] + E\left[\left|\varepsilon_r(k,m)\right|^2\right], \quad (24)
$$

where

$$
\begin{aligned}
E\left[\left|\varepsilon_d(k,m)\right|^2\right] &= E\left[\left|S_{fd}(k,m) - S(k,m)\right|^2\right] \\
&= E\left[\left|S(k,m)\mathbf{w}^H(k,m)\boldsymbol{\gamma}_s^*(k,m) - S(k,m)\right|^2\right] \quad (25) \\
&= \lambda_S(k,m)\left|\mathbf{w}^H(k,m)\boldsymbol{\gamma}_s^*(k,m) - 1\right|^2,
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[\left|\varepsilon_r(k,m)\right|^2\right] &= E\left[\left|S'_{rr}(k,m) + S'_{rn}(k,m)\right|^2\right] \\
&= \lambda_{S_{rr}}(k,m) + \lambda_{S_{rn}}(k,m) \quad (26) \\
&= \mathbf{w}^H(k,m)\boldsymbol{\Phi}_{in}(k,m)\mathbf{w}(k,m),
\end{aligned}
$$

with

$$
\boldsymbol{\Phi}_{in}(k,m) = \boldsymbol{\Phi}_{s_z{}'}(k,m) + \boldsymbol{\Phi}_v(k,m) \quad (27)
$$

being the reverberation-plus-noise interference covariance matrix.

The MVDR filter can be derived by minimizing the MSE of the residual interference, $E[\,|\varepsilon_r(k, m)|^2]$, with the constraint that the desired signal is not

distorted.

$$\min_{w(k,m)} \mathbf{w}^H(k,m)\mathbf{\Phi}_{\text{in}}(k,m)\mathbf{w}(k,m)$$
$$\text{subject to } \mathbf{w}^H(k,m)\boldsymbol{\gamma}_s^*(k,m)=1, \qquad (28)$$

for which the solution is[26,38,39]

$$\mathbf{w}_{\text{MVDR}}(k,m) = \frac{\mathbf{\Phi}_{\text{in}}^{-1}(k,m)\boldsymbol{\gamma}_s^*(k,m)}{\boldsymbol{\gamma}_s^T(k,m)\mathbf{\Phi}_{\text{in}}^{-1}(k,m)\boldsymbol{\gamma}_s^*(k,m)}$$
$$= \frac{\mathbf{\Phi}_y^{-1}(k,m)\boldsymbol{\gamma}_s^*(k,m)}{\boldsymbol{\gamma}_s^T(k,m)\mathbf{\Phi}_y^{-1}(k,m)\boldsymbol{\gamma}_s^*(k,m)}, \qquad (29)$$

where $\mathbf{\Phi}_y(k,m) = E[\mathbf{y}(k,m)\mathbf{y}^H(k,m)]$ is the correlation matrix of $\mathbf{y}(k,m)$.

Note that the filter equation is actually identical to the one for the noise-free environment in,[25] except for replacement of $\mathbf{\Phi}_z^{-1}(k,m)$ by $\mathbf{\Phi}_y^{-1}(k,m)$. This is expected since the additive noise is assumed to be uncorrelated to the speech signal and thus regarded as the interference. It apparently shows that the proposed MVDR filter in Eq. (29) is primarily designed to minimize every components uncorrelated with desired speech, so that it is an algorithm robust to the noise.

IV. Suppression of late reverberation and additive noise

In this section, the correlation vector $\boldsymbol{\gamma}_s(k,m)$ in Eq. (29) is derived using a statistical reverberation model to suppress the late reverberation. And the practical methods to obtain the subparameters which are required to construct $\boldsymbol{\gamma}_s(k,m)$ are introduced. The subparameters include the variances and the correlation coefficients of the observed signal, the late reverberation, and the noise signal.

4.1 Derivation of the correlation parameter

The summation in Eq. (2) can be split into a

contribution of the early speech component $X(k,m)$ and the late reverberation $R(k,m)$ as follows[4,18-20]

$$X(k,m) = \sum_{l'=0}^{N_e-1} H(k,l')S(k,m-l'), \qquad (30)$$

and

$$R(k,m) = \sum_{l'=N_e}^{\infty} H(k,l')S(k,m-l'), \qquad (31)$$

where N_e determines the start time of the AIR that may be considered as reverberation. If N_e is decided big enough, it can be assumed that the correlation between $R(k,m)$ and $S(k,m)$ is negligible. The time instance N_e usually ranges from 32 to 64 ms.[4] In this paper, we empirically choose $N_e = 12$ (i.e., 48 ms), which is identical to one in Habets' work,[20] so that $R(k,m)$ in the Eq. (31) consists of only late reverberation.

A new desired signal, that is, the early speech component $X(k,m)$ is given by

$$X(k,m) = Y(k,m) - R(k,m) - V(k,m). \qquad (32)$$

Suppressing the late reverberation can be achieved by recovering $X(k,m)$. From Eq. (7) and (32), the estimated correlation coefficient to estimate the early speech component is given by

$$\tilde{\gamma}_s(k,m,l) = \frac{E\left[X(k,m)Y^*(k,m+l)\right]}{E\left[|S(k,m)|^2\right]}$$
$$= \frac{E\left[\{Y(k,m)-R(k,m)-V(k,m)\}Y^*(k,m+l)\right]}{E\left[|Y(k,m)-R(k,m)-V(k,m)|^2\right]}$$
$$= \frac{E\left[Y(k,m)Y^*(k,m+l)-R(k,m)Y^*(k,m+l)-V(k,m)Y^*(k,m+l)\right]}{\lambda_Y(k,m)-\lambda_R(k,m)-\lambda_V(k,m)}$$
$$= \frac{E\left[Y(k,m)Y^*(k,m+l)\right]-E\left[R(k,m)R^*(k,m+l)\right]-E\left[V(k,m)V^*(k,m+l)\right]}{\lambda_Y(k,m)-\lambda_R(k,m)-\lambda_V(k,m)} \qquad (33)$$

due to

$$E\left[R(k,m)Y^*(k,m+l)\right] = E\left[R(k,m)R^*(k,m+l)\right],$$
$$E\left[V(k,m)Y^*(k,m+l)\right] = E\left[V(k,m)V^*(k,m+l)\right]. \qquad (34)$$

The correlation of the reverberant component $E[R(k, m)R^*(k, m+l)]$ can be represented by the multiplication of the variance of the late reverberation and a parameter that is exponentially decaying due to l [25]:

$$E\left[R(k,m)R^*(k,m+l)\right] = \lambda_R(k,m)e^{-\alpha(k)lR}. \quad (35)$$

Then, the estimated correlation coefficient $\tilde{\gamma}_S(k,m,l)$ in Eq. (33) is reformulated as

$$\tilde{\gamma}_S(k,m,l) = \frac{\lambda_Y(k,m)}{\lambda_X(k,m)}\gamma_Y(k,m,l) \\ -\frac{\lambda_R(k,m)}{\lambda_X(k,m)}\gamma_R(k,m,l) - \frac{\lambda_V(k,m)}{\lambda_X(k,m)}\gamma_V(k,m,l) \quad (36)$$

where

$$\lambda_X(k,m) = \lambda_Y(k,m) - \lambda_R(k,m) - \lambda_V(k,m) \quad (37)$$

and

$$\gamma_R(k,m,l) = e^{-\alpha(k)lN}. \quad (38)$$

The proposed algorithm to estimate the correlation parameter in Eq. (36) requires subparameters, such as the variance of the late reverberation $\hat{\lambda}_R(k, m)$, the variance of noise $\hat{\lambda}_V(k, m)$, the correlation of the late reverberation $\gamma_R(k, m, l)$, and the correlation of the noise signal $\gamma_V(k, m, l)$ at the subsequent frame. While the variance terms play a classical role in attenuating the spectral component, the correlation parameters give additional aggressiveness to the proposed algorithm, so that it dynamically suppresses the late reverberation and the noise by estimating the changes, for example speech on/off set or noise fluctuations, which may occur at the subsequent frames.

4.2 Estimation of subparameters

The power spectrum of the early speech component in Eq. (37) can be estimated by following power spectral subtraction method

$$\lambda_X(k,m) = \left(G_P(k,m)\right)^2 \left|Y(k,m)\right|^2. \quad (39)$$

As shown in,[27] the spectral gain function is given by

$$G_P(k,m) = \sqrt{\frac{\xi_P(k,m)}{1+\xi_P(k,m)}\left(\frac{1}{\gamma_P(k,m)} + \frac{\xi_P(k,m)}{1+\xi_P(k,m)}\right)}, \quad (40)$$

where

$$\xi_P(k,m) = \frac{\lambda_X(k,m)}{\lambda_R(k,m) + \lambda_V(k,m)} \quad (41)$$

and

$$\gamma_P(k,m) = \frac{\left|Y(k,m)\right|^2}{\lambda_R(k,m) + \lambda_V(k,m)} \quad (42)$$

denote the *a priori* and *a posteriori* SIR, respectively. The *a priori* SIR can be estimated by using the decision-directed method.[4,37]

From the general statisitical reverberant model in,[4,18-20] the late reverberant spectral variance $\lambda_R(k, m)$ is given by

$$\lambda_R(k,m) = e^{-2\alpha N(N_e-1)}\lambda_Z(k,m-N_e+1), \quad (43)$$

where

$$\lambda_Z(k,m) \\ = (1-\kappa)e^{-2\alpha N}\lambda_Z(k,m-1) + \kappa e^{-2\alpha N}\lambda_Y(k,m-1). \quad (44)$$

The estimate $\lambda_R(k, m)$ can then be used to estimate

the spectral variance of the early speech component $\lambda_X(k, m)$ in Eq. (39-42) and to estimate the correlation coefficient $\tilde{\gamma}_S(k, m, l)$ in Eq. (36).

For the estimation of the late reverberation spectral variance $\lambda_X(k, m)$ in Eq. (39), an estimate of the power spectrum of the late reverberation $\lambda_R(k, m)$ is required. The late reverberant spectral variance $\lambda_R(k, m)$ can be attained by Eq. (43). For the estimation of the late reverberation spectral variance $\lambda_R(k, m)$ in Eq. (43), an estimate of the power spectrum $\lambda_Z(k, m)$ in Eq. (44) is required. The power spectrum of the reverberant spectral component $Z(k, m)$ can be estimated by the power spectral subtraction method given by

$$\lambda_Z(k,m) = |Z(k,m)|^2 = \left(G_{SP}(k,m)\right)^2 |Y(k,m)|^2. \quad (45)$$

with

$$G_{SP}(k,m) = \sqrt{\frac{\xi_{SP}(k,m)}{1+\xi_{SP}(k,m)}\left(\frac{1}{\gamma_{SP}(k,m)} + \frac{\xi_{SP}(k,m)}{1+\xi_{SP}(k,m)}\right)}, \quad (46)$$

where

$$\xi_{SP}(k,m) = \frac{\lambda_Z(k,m)}{\lambda_V(k,m)} \quad (47)$$

and

$$\gamma_{SP}(k,m) = \frac{|Y(k,m)|^2}{\lambda_V(k,m)} \quad (48)$$

denote the *a priori* and *a posteriori* SNR, respectively. The noise spectral variance $\lambda_V(k, m)$ is estimated from the observed noisy reverberant signal $Y(k, m)$ by using noise power spectrum estimation methods.[30-36]

A diagram of the proposed single-channel non-causal dereverberation algorithm is depicted in Figure 1. The output signal $\hat{X}(k, m)$ is made by filtering the input vector of the reverberant signal $Z(k, m)$ with the correlation vector $\hat{\gamma}_S(k, m)$. The correlation coefficient is estimated based on the input signal and the sub-parameters such as $\hat{\gamma}_R(k, m)$, $\hat{\lambda}_R(k, m)$, and $\hat{\lambda}_X(k, m)$.

V. Performance evaluation

In this section, the performance of the proposed algorithm for the noisy reverberant environments is verified. We compare the proposed method with the Habet's method.[4] The evaluation is performed based on three major objective measurements - the signal to interference ratio (SIR) in time domain, the signal to
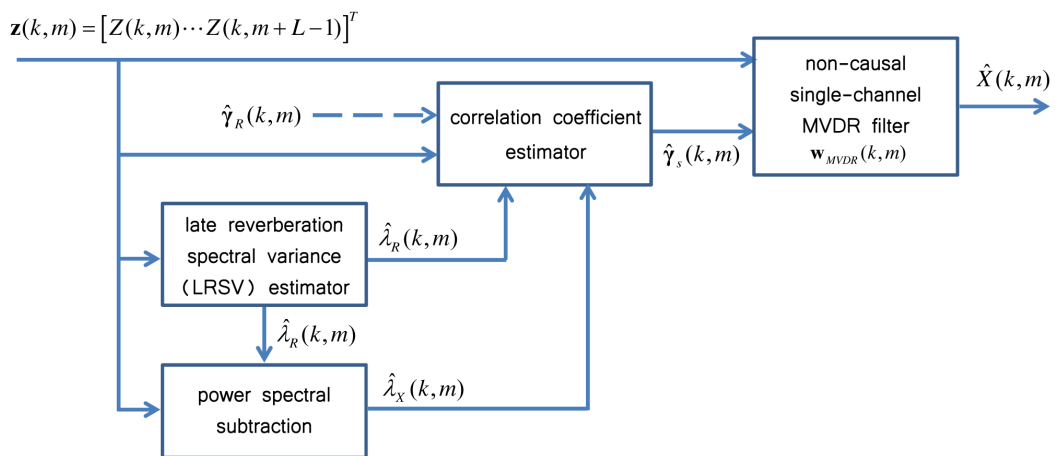

Fig. 1. Block diagram of the proposed system.

interference ratio (SIR) in frequency domain, and speech distortion (SD) index. As the interference consists of both the reverberant interference and the noise interference, the SIR in frequency domain can be divided into the signal to reverberant ratio (SRR) and the signal to noise ratio (SNR). By using the SRR and the SNR measurements, the performance can be analyzed separately for each interference.

The rest of this section is organized as follows. The simulation set-ups described in section 5.1. Section 5.2 represents the evaluation of the proposed algorithm in the noisy reverberant environment.

5.1 Experimental set−ups

The clean speech signal is created by concatenating 5 different utterances, which are spoken by 5 different speakers, from AURORA2 database. The signal is sampled at 8 kHz, 15 s-long, and it is transformed into the short time Fourier transform (STFT) domain using 75% overlapping (i.e., N=32). The Kaiser window of 128 samples is used.

The speech signal is convolved with different AIRs in order to generate the reverberant signals. The AIRs are synthesized under different environments using the image method.[28] The source-microphone distance D = 4.5 m, RT_{60} = {600, 800, 1000, 1200} ms, and the room size is set to 6×8×5 m (length × width × height).

The noisy reverberant signals are generated by first convolving the speech signal with the AIRs and then corrupted by the additive noise. Gaussian random noise and destroyer-engine noise (from NOISEX-92 database[40]) are added to the reverberant signal at a specified input SNR. Ten independent trials are conducted to examine the consistency of the evaluation.

The reverberation time RT_{60} is assumed to be known in the simulation, which can be estimated by using blind estimation procedures in practice.[18,29] Preliminary experiments confirm that the proposed algorithm is robust to the estimation error of RT_{60},

although further analysis remains as future work. The forgetting factor for the variance of the late reverberation is set to $\kappa = 0.2$.

The estimates of $\Phi_y(k, m)$ are recursively updated as follows:

$$
\begin{aligned}
&\Phi_y(k,m) \\
&= \mu\Phi_y(k,m-1) + (1-\mu)\mathbf{y}(k,m)\mathbf{y}^H(k,m),
\end{aligned} \tag{49}
$$

where μ represents the forgetting factor. The forgetting factor μ has an important role to control the trade-off between the singular or ill-conditioned correlation matrix $\Phi_y(k, m)$ (with a small μ) and smoothing of the short-term variation of speech signals (for μ close to 1). Unless we specifically mention, μ is empirically fixed as 0.6 to guarantee relatively high output SIR and good listening quality.

To compute the inverse of $\Phi_y(k, m)$, the regularization technique is used, so that $\Phi_y^{-1}(k, m)$ is replaced by

$$
\left\{ \Phi_y^{-1}(k,m) + \frac{\rho \cdot tr\left[\left|\Phi_y(k,m)\right|\right]}{L} I_{L\times L} \right\}^{-1}, \tag{50}
$$

where $\rho > 0$, $tr[\cdot]$, and $I_{L\times L}$ denote the regularization parameter, trace operation, and L by L identity matrix, respectively. We use the first 10 frames (i.e., 40 ms) to compute the initial estimates of $\Phi_y(k, m)$. The rest of signal frames is then used for performance evaluation.

The SIRs in both the time domain and the frequency domain are utilized for performance evaluation. The SIR in the time domain between the clean speech $s(n)$ and the processed signal $\hat{s}(n)$ is defined as

$$
SIR_{time}(\hat{s}) = \frac{1}{M}\sum_{m=0}^{M-1} 10\log_{10}\left(\frac{\sum_{n=mR}^{mR+N-1} s^2(n)}{\sum_{n=mR}^{mR+N-1}\left(\hat{s}(n)-s(n)\right)^2} \right), \tag{51}
$$

and the SIR in the time domain between the clean speech $s(n)$ and the observed noisy reverberant signal $y(n)$ is calculated by

$$SIR_{time}(y) = \frac{1}{M}\sum_{m=0}^{M-1}10\log_{10}\left(\frac{\sum_{n=mR}^{mR+N-1}s^2(n)}{\sum_{n=mR}^{mR+N-1}\left(y(n)-s(n)\right)^2}\right). \quad (52)$$

Accordingly, the improvement of the time domain SIR is defined by

$$\Delta\text{SIR}_{time} = \text{SIR}_{time}(\hat{s}) - \text{SIR}_{time}(y). \quad (53)$$

The large ΔSIR value represents that the output signal $\hat{s}(n)$ is much more similar to the desired signal $s(n)$ compared to the observed signal $y(n)$.

The input SIR in the frequency domain is defined by

$$iSIR_{freq} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_S(k,m)}{\sum_{K=0}^{K-1}|Y(k,m)-S(k,m)|^2}\right) \quad (54)$$

and

$$oSIR_{freq} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_{S_{fd}}(k,m)}{\sum_{K=0}^{K-1}\lambda_{S_{ri}'}(k,m)}\right)$$
$$= 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\left|S(k,m)\mathbf{w}^H(k,m)\boldsymbol{\gamma}_s^*(k,m)\right|^2}{\sum_{K=0}^{K-1}\left|\mathbf{w}^H(k,m)\mathbf{s}'(k,m)\right|^2}\right) \quad (55)$$

is the output SIR in the frequency domain, which is ratio between variance of the filtered desired signal $\lambda_{S_{fd}}(k,m)$ and variance of the reverberation-plus-noise residual signal $\lambda_{S_{ri}'}$. The improvement of the frequency domain SIR is defined by

$$\Delta SIR_{freq} = oSIR_{freq} - iSIR_{freq}. \quad (56)$$

The proposed algorithm suppresses both of the reverberation and the background noise at the same time. However, the performance measures described above are not able to distinguish the reverberation reduction and the noise suppression. Thus, additional measures to separately analyze the effect of the proposed method on the reverberation reduction and the noise reduction are required.

The reverberation-plus-noise residual $S'_{ri}(k,m)$ can be decomposed into the residual reverberation $S'_{rr}(k,m)$ and the residual noise $S'_{rn}(k,m)$ as in Eq. (18). We define the input SRR by the ratio between variance of the desired signal and variance of reverberant signal as followed

$$iSRR_{freq} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_S(k,m)}{\sum_{K=0}^{K-1}|Z(k,m)-S(k,m)|^2}\right) \quad (57)$$

and

$$oSRR_{freq} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_{S_{fd}}(k,m)}{\sum_{K=0}^{K-1}\lambda_{S_{ri}'}(k,m)}\right)$$
$$= 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\left|S(k,m)\mathbf{w}^H(k,m)\boldsymbol{\gamma}_s^*(k,m)\right|^2}{\sum_{K=0}^{K-1}\left|\mathbf{w}^H(k,m)\mathbf{s}_z'(k,m)\right|^2}\right) \quad (58)$$

is the output SRR, which denotes the ratio between variance of the filtered desired signal $\lambda_{S_{fd}}(k,m)$ and variance of the reverberation residual signal $\lambda_{S_{rr}'}$. The improvement of the frequency domain SRR is defined by

$$\Delta SRR_{freq} = oSRR_{freq} - iSRR_{freq}. \quad (59)$$

Similarly, the improvement of the frequency domain SNR is defined by

$$\Delta SNR_{freq} = oSNR_{freq} - iSNR_{freq}, \quad (60)$$

where

$$iSNR_{freq} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_S(k,m)}{\sum_{K=0}^{K-1}\lambda_V(k,m)}\right) \qquad (61)$$

is the input SNR and

$$oSNR_{freq} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_{S_{fd}}(k,m)}{\sum_{K=0}^{K-1}\lambda_{S_{rn}}(k,m)}\right)$$
$$= 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\left|S(k,m)\mathbf{w}^H(k,m)\boldsymbol{\gamma}_s^*(k,m)\right|^2}{\sum_{K=0}^{K-1}\left|\mathbf{w}^H(k,m)\mathbf{s}_v'(k,m)\right|^2}\right) \qquad (62)$$

is the output SNR representing the ratio between the filtered desired signal variance and the variance of the noise residual.

Another useful performance measure is the speech distortion index defined as

$$\upsilon_{sd} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\upsilon_{sd}(m)\right), \qquad (63)$$

where

$$\upsilon_{sd}(m) = \frac{\sum_{k=0}^{K-1}\left|S_{fd}(k,m)-S(k,m)\right|^2}{\sum_{k=0}^{K-1}\lambda_S(k,m)} \qquad (64)$$
$$= \left|\mathbf{w}^H(k,m)\boldsymbol{\gamma}_s^*(k,m)-1\right|^2$$

is the speech distortion at the time-frame m. The speech distortion $\upsilon_{sd}(m)$ is always greater than or equal to 0 and should be upper bounded by 1 for optimal filters. So the higher is its value, the more the desired signal is distorted. For the proposed filter, it is clear that $\upsilon_{sd}(m)\approx0$, so that $\upsilon_{sd}\approx-\infty$.

The objective measures explained above are summarized in Table 1.

Table 1. Summary of objective measures

Number	Objective measurement
Eq. (53)	$\Delta\text{SIR}_{time} = \frac{1}{M}\sum_{m=0}^{M-1}10\log_{10}\left(\frac{\sum_{n=mR}^{mR+N-1}s^2(n)}{\sum_{n=mR}^{mR+N-1}(\hat{s}(n)-s(n))^2}\right) - \frac{1}{M}\sum_{m=0}^{M-1}10\log_{10}\left(\frac{\sum_{n=mR}^{mR+N-1}s^2(n)}{\sum_{n=mR}^{mR+N-1}(y(n)-s(n))^2}\right)$
Eq.(56)	$\Delta SIR_{freq} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_{S_{fd}}(k,m)}{\sum_{K=0}^{K-1}\lambda_{S_{ri}}(k,m)}\right) - 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_S(k,m)}{\sum_{K=0}^{K-1}\left|Y(k,m)-S(k,m)\right|^2}\right)$
Eq.(59)	$\Delta SRR_{freq} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_{S_{fd}}(k,m)}{\sum_{K=0}^{K-1}\lambda_{S_{ri}}(k,m)}\right) - 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_S(k,m)}{\sum_{K=0}^{K-1}\left|Z(k,m)-S(k,m)\right|^2}\right)$
Eq.(60)	$\Delta SNR_{freq} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_{S_{fd}}(k,m)}{\sum_{K=0}^{K-1}\lambda_{S_{rn}}(k,m)}\right) - 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{K=0}^{K-1}\lambda_S(k,m)}{\sum_{K=0}^{K-1}\lambda_V(k,m)}\right)$
Eq.(63)	$\upsilon_{sd} = 10\log_{10}\left(\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{k=0}^{K-1}\left|S_{fd}(k,m)-S(k,m)\right|^2}{\sum_{k=0}^{K-1}\lambda_S(k,m)}\right)$

5.2 Experimental results

Figure 2 and Figure 3 show the basic simulation results. Figure 2 depicts the waveforms of the output signal processed by the conventional and the proposed method and the improvements of the time domain SIR by those algorithms. The same results are represented as spectrograms in Figure 3. The simulations are conducted under environment that the reverberation time RT_{60} is 0.9 s and 20 dB additive white noise. The proposed algorithm works with L=12.

As shown in the figures, the improvements of SIR_{time} by both algorithms are observed mostly for non-speech region. Note that the ΔSIR_{time} by the proposed method tends to increasing for the region of the reverberant tail, compared to the one by the

Habets' algorithm. This is one of the strong points of the proposed algorithm to dynamically suppress the reverberation by detecting the speech onset or offset, since it quantifies the variation in the subsequent frames by adopting $\hat{\gamma}_Y(k, m)$ and $\hat{\gamma}_R(k, m)$ in Eq. (36).

Figure 4 shows the effect of the forgetting factor μ on the performance of the proposed algorithm. The performance of the proposed algorithms tended to monotonically increasing as the order of the FIR filter increased. Using a large μ, the temporal variation of non-stationary speech signal can not be captured, so that both of SIR_{time} and SRR_{freq} gradually decrease. In contrast, the noise in the experiments is stationary such that a large μ is more advantageous for the performance of the noise reduction, so that
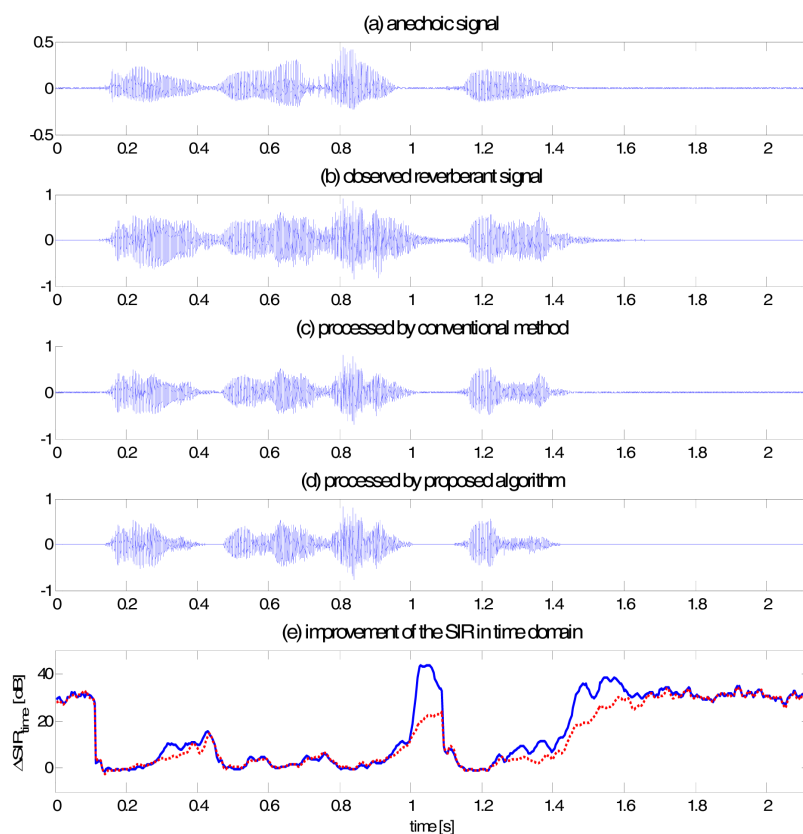

Fig. 2. The waveforms and the improvement of the time domain SIR (white noise case). (a) anechoic speech signal, (b) observed noisy reverberant signal, (c) processed by the Lebart's method, (d) processed by the proposed algorithm, and (e) improvements of the time domain SIRs by the Habets' and the proposed enhancement approaches.
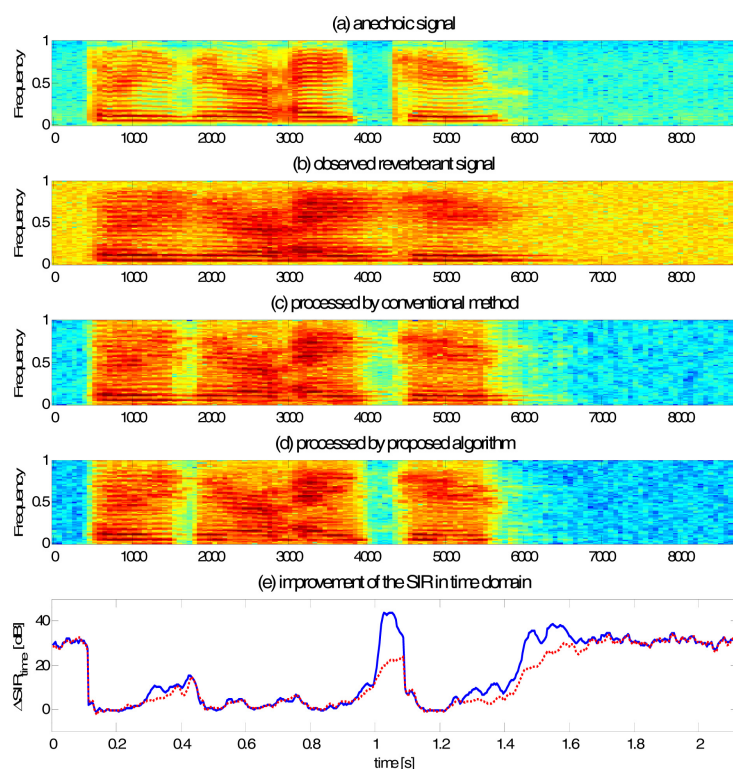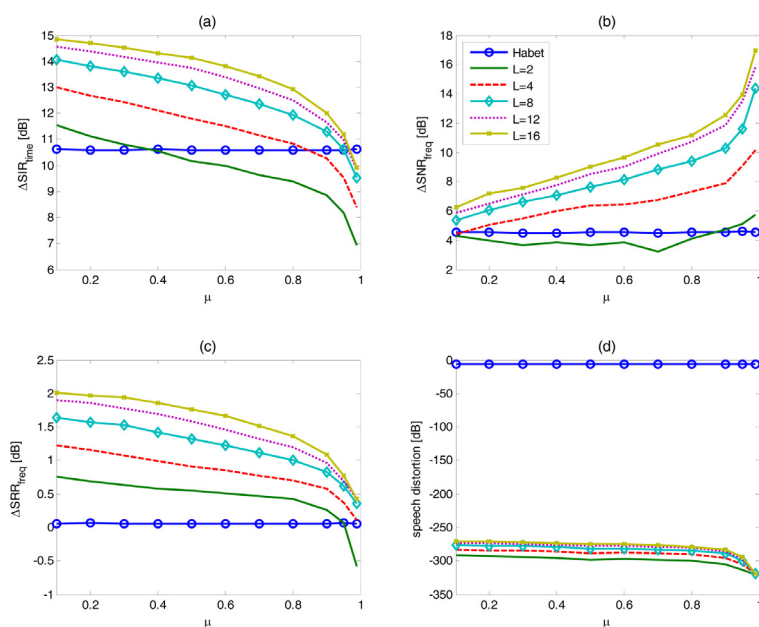
Fig. 3. The spectrograms and the improvement of the time domain SIR (white noise case). (a) anechoic speech signal, (b) observed noisy reverberant signal, (c) processed by the Lebart's method, (d) processed by the proposed algorithm, and (e) improvements of the time domain SIRs by the Habets' and the proposed enhancement approaches.

Fig. 4. Effect of the forgetting factor μ on the performance of the proposed algorithm (white noise case). (a) improvement of SIR in time domain, (b) improvement of SNR in frequency domain, (c) improvement of SRR in frequency domain, and (d) speech distortion.

SRR_{freq} increases.

Figure 5 and Figure 6 depict the time domain SIR performances of the conventional and the proposed algorithm in noisy (white noise and destroyer-engine noise) reverberant environment as a function of the input SNR. Each sub-figure depicts results of the proposed algorithm and the Habet's method for different reverberation time RT_{60}. The results show that the SIR improvement of both surveyed algorithms monotonically decrease as the input SNR increases. Especially, the results in relatively less-reverberant environment such as case of $RT_{60} = 0.6$ s degrades faster than the others. As shown in the figure, the proposed system outperforms the conventional one in every environments studied in this simulation. The superiority of the proposed algorithm is appeared when the environment is strongly reverberant with large RT_{60} value. It is interesting that the differences

between the results of both algorithms remains same regardless of changes of the input SNR.

Figure 7 and Figure 8 depict the same results in Figure 5 and Figure 6, respectively, after reorganizing with different axis. It shows the time domain SIR performances of the surveyed algorithms in noisy reverberant environment as a function of the reverberation time RT_{60}. Each sub-figure depicts results for different input SNR. In Figure 7, we observe that ΔSIR_{time} values of both algorithms monotonically increase as the reverberation time increases, except for the result of the Habet's method for 5 dB input SNR. In Figure 8, ΔSIR_{time} values of the conventional algorithm decrease more rapidly than the proposed one as the reverberation time increases, especially for the result under 20 dB input SNR. It is clearly noticable that the proposed algorithm outperforms the conventional one especially when the reverberation time
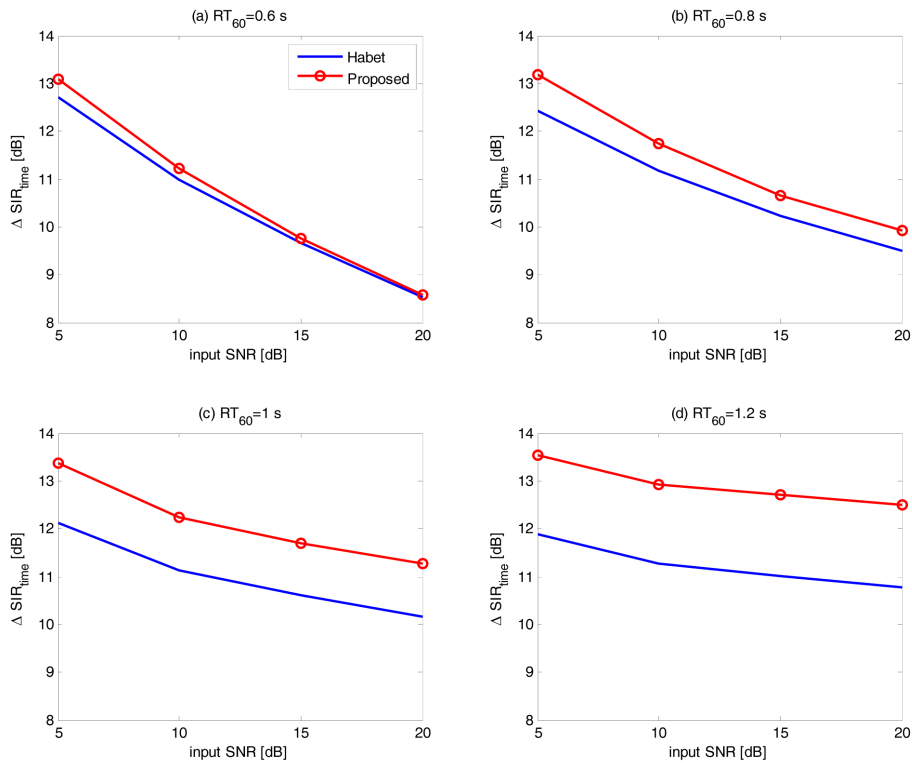

Fig. 5. Time domain SIR performances of the conventional and the proposed algorithm in noisy reverberant environment as a function of the input SNR (white noise case).
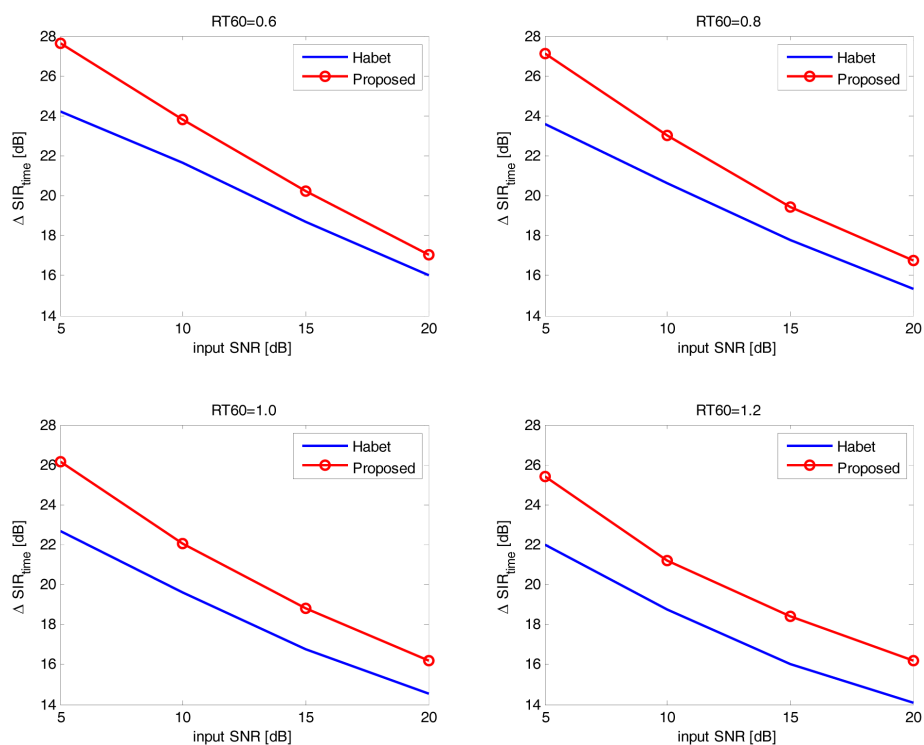
Fig. 6. Time domain SIR performances of the conventional and the proposed algorithm in noisy reverberant environment as a function of the input SNR (destroyer-engine noise case).
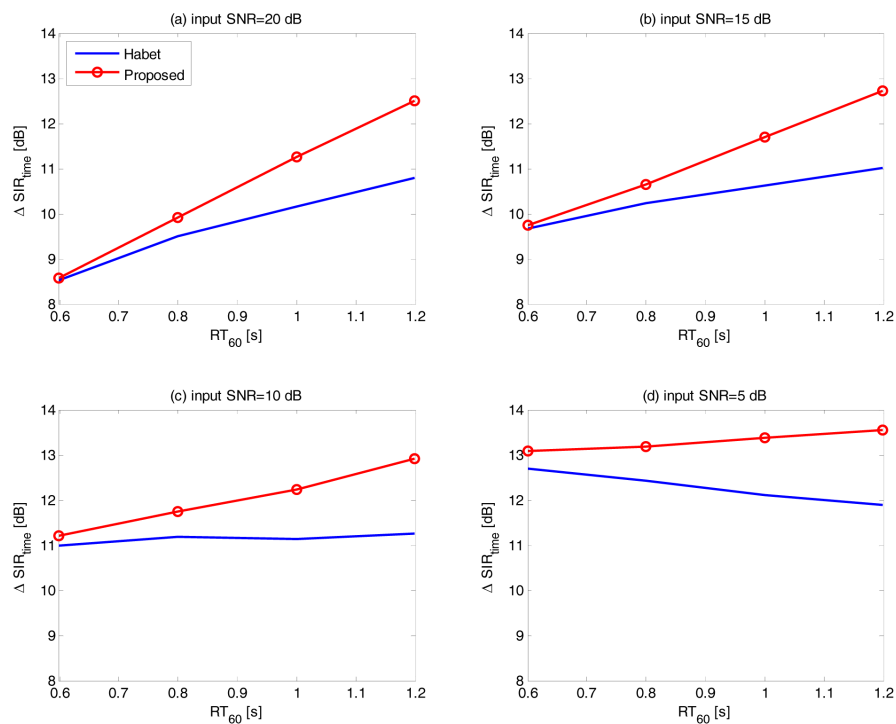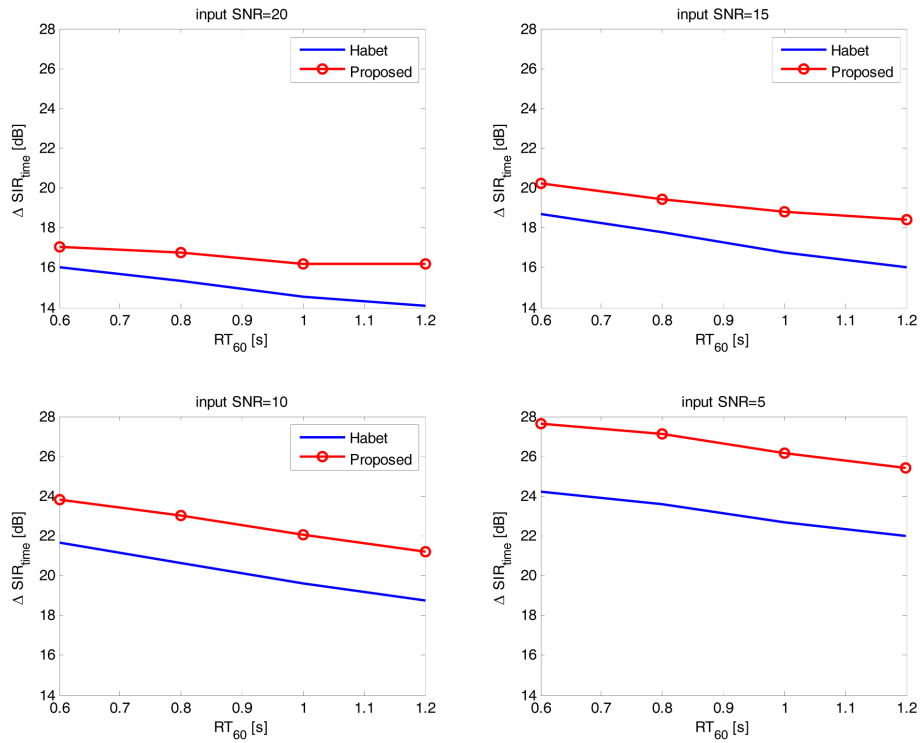

Fig. 7. Time domain SIR performances of the conventional and the proposed algorithm in noisy reverberant environment as a function of the reverberation time RT_{60} (white noise case).

Fig. 8. Time domain SIR performances of the conventional and the proposed algorithm in noisy reverberant environment as a function of the reverberation time RT_{60} (destroyer-engine noise case).
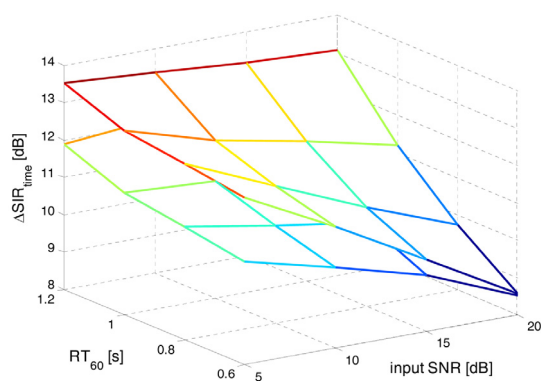

Fig. 9. Time domain SIR performances of the conventional and the proposed algorithm in noisy reverberant environment according to both of the input SNR and the reverberation time RT_{60} (white noise case).
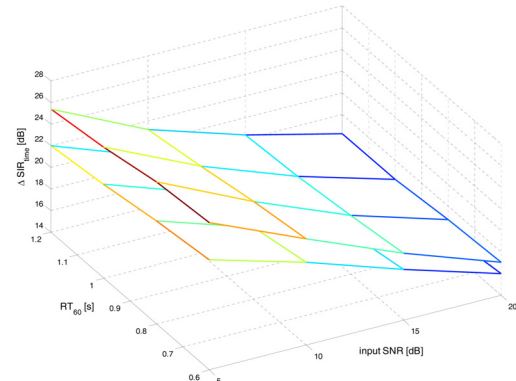

Fig. 10. Time domain SIR performances of the conventional and the proposed algorithm in noisy reverberant environment according to both of the input SNR and the reverberation time RT_{60} (destroyer-engine noise case).

is large. In other words, the proposed algorithm using multiple consecutive STFT fames improves the dereverberation performances and is far better than the conventional algorithm under strongly reverberant environments.

Figure 9 and Figure 10 represent the improvement of the SIR by the surveyed algorithms against the input SNR and the reverberation time. In these figures, it is shown that the results of the Habet's method always stays under the results surface of the proposed
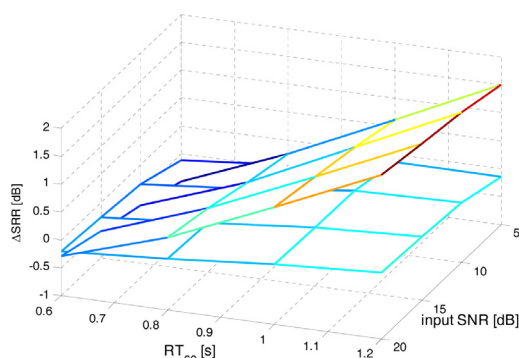
Fig. 11. SRR performances of the conventional and the proposed algorithm in noisy reverberant environment as a function of the input SNR and the reverberation time RT_{60} (white noise case).
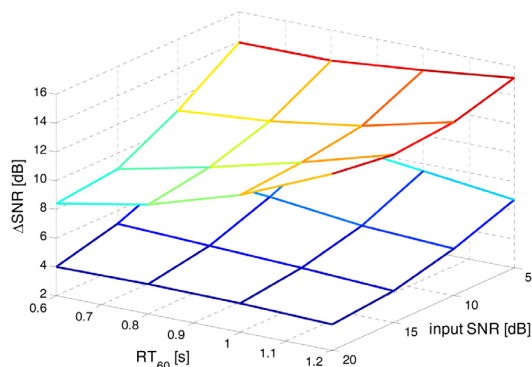

Fig. 12. SNR performances of the conventional and the proposed algorithm in noisy reverberant environment according to the input SNR and the reverberation time RT_{60} (white noise case).

algorithm. The differences between the results under the environment with low input SNR and large reverberation time is much larger than one with high input SNR and small reverberation time. The results illustrates that the proposed system has its superiority for the severely noisy and strongly reverberant environment.

Figure 11 represents the SRR performances of the conventional and the proposed algorithms as a function of the input SNR and the reverberation time RT_{60}. The SRR values are computed with Eq. (59). The upper surface represents the results of the proposed algorithm and the lower is ones of the conventional method. The results of both algorithms monotonically

increase as the reverberation time increases, while those remain regardless to the change of the input SNR. This shows that the SRR has dependency to the RT_{60} and is independent to the input SNR.

It is shown that the proposed algorithm is superior to the Habet's method, especially in the environment with strong reverberation. Because the proposed method utilizes additional information of the correlated components from the subsequent frames, it results in dynamic suppression of the late reverberation and thus, much more improvement of the SRR is attained under highly reverberant environment, such as RT_{60} = 1.2 s.

Figure 12 depicts the SNR performances of both studied algorithms against the input SNR and the reverberation time. From this figure, it is shown that the SNRs of the proposed algorithm is always bigger than those of the Habet's method. It is interesting that the SNR values of both algorithms has dependency not only to change of the input SNR but also to alteration of the reverberation time. The noise reduction capacity of the proposed algorithm is improved for strongly reverberant environment, while the performance of the conventional one is not enhanced or rather counteracts.

We also conducted informal PESQ (Perceptual Evaluation of Speech Quality) measurement results. The results show that the proposed algorithm slightly outperforms to all the reference approaches. However, we do not include the detailed scores here because there is a clarification issue whether the PESQ score is suitable measure for measuring qualities in reverberant environment.

VI. Conclusion

In this paper, an efficient single-channel dereverberation algorithm to suppress the late reverberation from the noisy reverberant signal. The non-causal MVDR filter was proposed to attenuate the reverberant-

plus-noise interference while minimizing speech distortion. It is interesting that the derived final filter equation is equivalent to one for noise-free environment, in spite of the additional interference (i.e. background noise). It apparently shows that the proposed MVDR filter is an algorithm robust to the noise, since it is primarily designed to minimize every components uncorrelated with desired speech.

An efficient method to estimate the correlation parameter was derived based on a statistical reverberant model and it is practically implemented. By adopting the correlation of late reverberation $\gamma_R(k, m, l)$ and that of noise signal $\gamma_V(k, m, l)$, the proposed method can control the aggressiveness of suppression of the interferences by estimating the changes which may occur at the subsequent frames.

Evaluation was conducted to verify the performance of the proposed algorithm by comparing with the conventional algorithm. The evaluation analysis was performed separately for each interference (the reverberant interference and the noise interference). The results showed that the proposed algorithm always outperformed the conventional one in various noisy reverberant environments. The performance improvement of the proposed algorithm was increased at the region that the speech ended, as it aggressively reduced the late reverberation at the region of the speech tail. The proposed algorithm maintained the minimal speech distortion and improved the SIR, the SNR, and the SRR performances in all studied conditions. The proposed system showed its superiority especially for the severely noisy and strongly reverberant environment.

References

1. E.A.P. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," 2007.
2. A.K. N´abˇelek, T.R. Letowski, and F.M. Tucker, "Reverberant overlap-and self-masking in consonant identification," *J. Acoust. Soc. Am.*, vol. 86, no. 4, pp. 1259-65, 1989.
3. L.L. Beranek, *Concert and opera halls: how they sound*, Published for the Acoustical Society of America through the American Institute of Physics, 1996.
4. L.E. Kinsler, A.R. Frey, A.B. Coppens, and J.V. Sanders, *Fundamentals of Acoustics, 4th Edition*, pp. 560. ISBN 0-471-84789-5. Wiley-VCH, December 1999.
5. H. Kuttruff, *Room acoustics*, Taylor & Francis, 2000.
6. M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," , *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 2, pp. 145-152, 1988.
7. J. Mourjopoulos, P. Clarkson, and J. Hammond, "A comparative study of leastsquares and homomorphic techniques for the inversion of mixed phase signals," *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'82.*, vol. 7, pp. 1858-1861, 1982.
8. S. Treitel and EA Robinson, "The design of high-resolution digital filters," *IEEE Transactions on Geoscience Electronics*, vol. 4, no. 1, pp. 25-38, 1966.
9. D. Bees, M. Blostein, and P. Kabal, "Reverberant speech enhancement using cepstral processing," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP-91.*, pp. 977-980, 1991.
10. RA Kennedy and BD Radlovic, "Iterative cepstrum-based approach for speech dereverberation," *Proceedings of the Fifth International Symposium on Signal Processing and Its Applications, ISSPA'99.*, vol. 1, pp. 55-58, 1999.
11. A.P. Petropulu and S. Subramaniam, "Cepstrum based deconvolution for speech dereverberation," *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-94.*, vol. 1, pp. I-9, 1994.
12. S. Subramaniam, A.P. Petropulu, and C. Wendt, "Cepstrum-based deconvolution for speech dereverberation," *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 5, pp. 392-396, 1996.
13. B. Yegnanarayana, P. Satyanarayana Murthy, C. Avendano, and H. Hermansky, "Enhancement of reverberant speech using lp residual," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 405-408, 1998.
14. B. Yegnanarayana and P.S. Murthy, "Enhancement of reverberant speech using lp residual signal," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 267- 281, 2000.
15. T. Nakatani, M. Miyoshi, and K. Kinoshita, "Implementation and effects of single channel dereverberation based on the harmonic structure of speech," *In Proc.*

IWAENC2003. Citeseer, 2003.

16. T. Nakatani and M. Miyoshi, "Blind dereverberation of single channel speech signal based on harmonic structure," *In Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, vol. 1, pp. I-92, 2003.

17. T. Nakatani, K. Kinoshita, and M. Miyoshi, "Harmonicity-based blind dereverberation for single-channel speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. v15, no. 1, pp. 80-95, 2007.

18. K. Lebart, J.M. Boucher, and PN Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica united with Acustica*, vol. 87, no. 3, pp.359-366, 2001.

19. E.A.P. Habets, N.D. Gaubitch, and P.A. Naylor, "Temporal selective dereverberation of noisy speech using one microphone," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008.*, pp. 4577-4580, 2008.

20. E.A.P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Processing Letters*, vo. 16, no. 9, pp. 770-773, 2009.

21. H.W. Lollmann and P. Vary, "A blind speech enhancement algorithm for the suppression of late reverberation and noise," IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009., pp. 3989-3992, 2009.

22. E.A.P. Habets, N.D. Gaubitch, and P.A. Naylor. Temporal selective dereverberation of noisy speech using one microphone. In Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, pages 4577-4580. IEEE, 2008.

23. J.S. Erkelens and R. Heusdens, "Single-microphone late-reverberation suppression in noisy speech by exploiting long-term correlation in the dft domain," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009.*, pp. 3997-4000, 2009.

24. J.S. Erkelens and R. Heusdens, "Noise and late-reverberation suppression in time-varying acoustical environments," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 4706-4709, 2010.

25. Myung-Suk Song and Hong-Goo Kang, "Single-Channel Dereverberation using a Non-Causal MVDR Filter," *Journal of the Acoustical Society of America Express Letter*, vol. 132, no. 1, pp. 29-35, 2012.

26. Benesty, J. and Chen, J, *Optimal Time-Domain Noise Reduction Filters: A Theoretical Study*, vol. 1. Springer-Verlag New York Inc, 2011.

27. Accardi, A.J. and Cox, R.V, "A modular approach to speech enhancement with an application to speech

coding," *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. ICASSP'99.*, pp. 201-204, 1999.

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

29. Schroeder, M.R, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, pp. 409, 1965.

30. R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 2, pp. 137-145, 1980.

31. I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal processing*, vol. 81, no. 11, pp. 2403-2418, 2001.

32. D. Malah, R.V. Cox, and A.J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments," *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'99.*, vol. 2, pp. 789-792, 1999.

33. R. Martin, Spectral subtraction based on minimum statistics, *EUSIPCO*, pp. 6-8, 1994.

34. R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, 2001.

35. I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12-15, 2002.

36. I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466-475, 2003.

37. Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109-1121, 1984.

38. J. Benesty, J. Chen, and E.A.P. Habets, *Speech Enhancement in the STFT Domain*, Springer Verlag, 2011.

39. J. Benesty and Y. Huang, "A single-channel noise reduction mvdr filter," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 273-276, 2011.

40. A. Varga, H.J.M. 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 Commun.*, vol. 12, no. 3, pp. 247-251, 1993.

Profile

▸ Myung—Suk Song

1998 ~ 2005: Dept. of Electrical Engineering, Yonsei University (B.S.)
2005 ~ 2007: Dept. of Electrical Engineering, Yonsei University (M.S.)
2007 ~ Present: Dept. of Electrical Engineering, Yonsei University (Ph.D.)
Research: Digital Signal Processing, Speech / Audio Signal Processing, Speech Enhancement, Speech Recognition, 3-D Audio

▸ Hong—Goo Kang

1985 ~ 1989: Dept. of Electrical Engineering, Yonsei University (B.S.)
1989 ~ 1991: Dept. of Electrical Engineering, Yonsei University (M.S.)
1991 ~ 1995: Dept. of Electrical Engineering, Yonsei University (Ph.D.)
1996 ~ 2002: Senior Technical Staff Member, AT&T Labs-Research
2002 ~ 2005: Assistant Professor, Yonsei University
2005 ~ Present: Associated Professor, Yonsei University
〈Research〉 Speech Signal Processing, Adaptive Digital Signal Processing and Real-time Implementation