COMBINED FREQUENCY-DOMAIN DEREVERBERATION AND NOISE REDUCTION TECHNIQUE FOR MULTI-MICROPHONE SPEECH ENHANCEMENT

Simon Doclo, Marc Moonen

Katholieke Universiteit Leuven, Dept. of Electrical Engineering (ESAT - SISTA),
Kasteelpark Arenberg 10, B-3001 Heverlee (Leuven), Belgium
{simon.doclo,marc.moonen}@esat.kuleuven.ac.be

ABSTRACT

In this paper a frequency-domain technique is described for estimating the acoustic transfer functions, when reverberated speech signals are corrupted by spatially coloured noise. This technique is an extension of the frequency-domain procedure of [1], which is only optimal in the case of spatially white noise. Using the estimated acoustic transfer functions, dereverberation can be performed with a matched filtering approach.
Also a GSVD-based noise reduction technique is discussed, which computes an optimal (MMSE) estimate of the speech components in each frequency bin. It is shown that the optimal estimate of the clean dereverberated speech signal is obtained by dereverberating the optimal estimate of these speech components. Since the same decomposition is required for dereverberation and noise reduction, both procedures can easily be combined.

1. INTRODUCTION

In many speech communication applications, like hands-free mobile telephony, hearing aids and voice-controlled systems, the recorded speech signals are of poor quality, because of background noise and reverberation (e.g. reflections against walls). Background noise and reverberation cause a signal degradation which can lead to total unintelligibility of the speech and which decrease the performance of speech recognition systems. Therefore efficient noise reduction and dereverberation algorithms are called for.
The objective of these algorithms can be either dereverberation (not caring about residual noise), noise reduction (not caring about residual reverberation), or combined dereverberation and noise reduction. This problem formulation is elaborated in section 2.
Section 3 describes a frequency-domain technique for estimating the transfer functions of noisy reverberated speech signals using a Generalised Singular Value Decomposition (GSVD). Using the estimated transfer functions, the dereverberated clean speech signal is computed with a matched filtering approach. This algorithm is in fact an extension of the technique presented in [1].

Simon Doclo is a Research Assistant supported by the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). This research was carried out at the ESAT laboratory of the KULeuven, in the frame of the Interuniversity Attraction Pole IUAP P4-02, *Modeling, Identification, Simulation and Control of Complex Systems*, the Concerted Research Action *Mathematical Engineering Techniques for Information and Communication Systems* (GOA-MEFISTO-666) and the IT-project *Multi-microphone Signal Enhancement Techniques for hands-free telephony and voice controlled systems (MUSETTE-2)* of the I.W.T.

In section 4 a frequency-domain noise reduction algorithm is described which computes an optimal MMSE estimate of the speech components of the microphone signals in each frequency bin. This algorithm is a frequency-domain version of the GSVD-based noise reduction algorithm of [2].
A combined dereverberation and noise reduction technique is presented in section 5. It is shown that the optimal estimate of the dereverberated clean speech signal is obtained by dereverberating the optimal estimate of the speech components. Since the same decomposition is required for both algorithms, they can easily be combined.
Section 6 discusses some practical implementation issues. Since essentially a fast convolution in the frequency-domain is performed, the time-domain filters need to be constrained in order to avoid circular convolutions [3].
Section 7 describes the simulation results, showing that the GSVD-based noise reduction technique produces the best signal-to-noise ratio (SNR), the GSVD-based dereverberation algorithm has the best dereverberation performance, while the combined noise reduction and dereverberation algorithm provides a trade-off between both objectives.

2. PROBLEM FORMULATION

Figure 1 depicts a microphone array with M microphones which records a desired speech source $s[k]$ and an undesired background noise source $v[k]$. Each microphone signal $y_m[k], m = 0 \dots M-1$, therefore consists of a filtered version of the clean speech signal and additive noise,

$$y_m[k] = x_m[k] + v_m[k] = g_m[k] \otimes s[k] + v_m[k] \; , \qquad (1)$$

with $g_m[k]$ the (unknown) impulse response of length K between the speech source and the mth microphone. The objective of dereverberation is to undo the filtering by the acoustic impulse responses and to compute dereverberation filters $f_m[k], m = 0 \dots M-1$, of length N such that the total transfer function $h[k]$ between $s[k]$ and the output $z[k]$,

$$z[k] = \sum_{m=0}^{M-1} \underbrace{(f_m[k] \otimes g_m[k])}_{h[k]} \otimes s[k] + \underbrace{\sum_{m=0}^{M-1} f_m[k] \otimes v_m[k]}_{r_v[k]}$$

is equal to 1 (or more realistically a delay), i.e.

$$h[k] = \sum_{m=0}^{M-1} f_m[k] \otimes g_m[k] = \delta(k - \Delta) \; . \qquad (2)$$

The residual noise $r_v[k]$ is not constrained in any way, such that the noise can even be amplified by the filters $f_m[k]$.

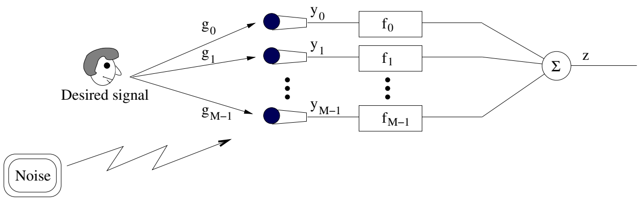

Figure 1: Problem formulation for dereverberation and noise reduction

The goal of combined dereverberation and noise reduction is to compute filters $f_m[k]$ such that the output $z[k]$ is an optimal estimate for the clean speech signal $s[k]$, thereby both reducing reverberation effects and background noise.

3. SPEECH DEREVERBERATION

Different methods for multi-microphone speech dereverberation have been proposed, e.g. in the time-domain [4], in the frequency-domain [1] or using cepstral processing.
In the frequency-domain the signal model (1) for each frequency $f = 0 \dots N-1$ and frame n becomes

$$\begin{aligned} \mathbf{Y}_n(f) &= \begin{bmatrix} G_0^n(f) \\ G_1^n(f) \\ \vdots \\ G_{M-1}^n(f) \end{bmatrix} S_n(f) + \begin{bmatrix} V_0^n(f) \\ V_1^n(f) \\ \vdots \\ V_{M-1}^n(f) \end{bmatrix} \quad (3) \\ &= \mathbf{G}_n(f)S_n(f) + \mathbf{V}_n(f) = \mathbf{X}_n(f) + \mathbf{V}_n(f) \,, \end{aligned}$$

with e.g.

$$S_n(f) = \sum_{l=0}^{N-1} s[nN+l]\, e^{-\frac{j2\pi l f}{N}} \,, \quad (4)$$

for the time being considering no overlap between frames (see section 6). We will assume slow time-variations of the transfer functions $G_m(f)$, such that $\mathbf{G}_n(f) \simeq \mathbf{G}(f)$.
The objective of dereverberation is to compute filters $\mathbf{F}(f)$ such that $H(f) = \mathbf{F}^H(f)\mathbf{G}(f) = 1$. The matched filter $\mathbf{F}(f) = \mathbf{G}(f)/\|\mathbf{G}(f)\|^2$ clearly is a solution to this equation.
The $M \times M$ noisy (spatial) correlation matrix $\mathbf{R}_{yy}(f)$ is

$$\begin{aligned} \mathbf{R}_{yy}(f) &= \mathcal{E}\{\mathbf{Y}_n(f)\mathbf{Y}_n^H(f)\} \quad (5) \\ &= \underbrace{\mathbf{G}(f)\mathcal{E}\{|S_n(f)|^2\}\mathbf{G}^H(f)}_{\mathbf{R}_{xx}(f)} + \underbrace{\mathcal{E}\{\mathbf{V}_n(f)\mathbf{V}_n^H(f)\}}_{\mathbf{R}_{vv}(f)} \,, \end{aligned}$$

assuming that the speech and noise components $X_m^n(f)$ and $V_m^n(f)$ are uncorrelated. The speech correlation matrix $\mathbf{R}_{xx}(f)$ is a rank-1 matrix. The noise correlation matrix $\mathbf{R}_{vv}(f)$ can be measured during speech pauses and reduces to $\sigma_v^2(f)I_M$ for spatially white noise.
The transfer function vector $\mathbf{G}(f)$ can be estimated using the Generalised Eigenvalue Decomposition (GEVD) of the correlation matrices $\mathbf{R}_{yy}(f)$ and $\mathbf{R}_{vv}(f)$ [5],

$$\begin{cases} \mathbf{R}_{yy}(f) = \mathbf{Q}(f)\mathbf{\Sigma}_y(f)\mathbf{Q}^H(f) \\ \mathbf{R}_{vv}(f) = \mathbf{Q}(f)\mathbf{\Sigma}_v(f)\mathbf{Q}^H(f) \,, \end{cases} \quad (6)$$

with $\mathbf{Q}(f)$ and invertible, but not necessarily orthogonal matrix. Since the speech correlation matrix

$$\mathbf{R}_{xx}(f) = \mathbf{R}_{yy}(f) - \mathbf{R}_{vv}(f) = \mathbf{Q}(f)\big(\mathbf{\Sigma}_y(f) - \mathbf{\Sigma}_v(f)\big)\mathbf{Q}^H(f)$$

has rank 1, it is equal to $\mathbf{R}_{xx}(f) = \sigma_x^2(f)\mathbf{q}_1(f)\mathbf{q}_1^H(f)$, with $\mathbf{q}_1(f)$ the M-dimensional principal generalised eigenvector, corresponding to the largest generalised eigenvalue. Since

$$\mathbf{R}_{xx}(f) = \sigma_x^2(f)\mathbf{q}_1(f)\mathbf{q}_1^H(f) = \mathcal{E}\{|S_n(f)|^2\}\mathbf{G}(f)\mathbf{G}^H(f) \,,$$

$\mathbf{G}(f)$ can be estimated up to a phase shift $e^{j\alpha(f)}$ as

$$\mathbf{q}_1(f) = \frac{\|\mathbf{q}_1(f)\|}{\|\mathbf{G}(f)\|}\mathbf{G}(f)e^{j\alpha(f)} \,. \quad (7)$$

We will assume that the human auditory system is not very sensitive to this phase shift. The dereverberated speech signal $\tilde{S}_n(f)$ can now be computed using the matched filter

$$\mathbf{F}(f) = \frac{\mathbf{G}(f)}{\|\mathbf{G}(f)\|^2} = \frac{\mathbf{q}_1(f)}{\|\mathbf{q}_1(f)\|\|\mathbf{G}(f)\|} \,, \quad (8)$$

such that

$$\tilde{S}_n(f) = \mathbf{F}^H(f)\mathbf{Y}_n(f) = S_n(f) + \frac{\mathbf{q}_1^H(f)}{\|\mathbf{q}_1(f)\|\|\mathbf{G}(f)\|}\mathbf{V}_n(f) \,.$$

which corresponds to the time-domain signal $\tilde{s}[k]$.
In practice, the transfer function vector $\mathbf{G}(f)$ is estimated using the Generalised Singular Value Decomposition (GSVD) of the speech and noise data matrices $\mathcal{Y}(f)$ and $\mathcal{V}(f)$,

$$\mathcal{Y}(f) = \begin{bmatrix} \mathbf{Y}_n^H(f) \\ \mathbf{Y}_{n-1}^H(f) \\ \vdots \\ \mathbf{Y}_{n-p+1}^H(f) \end{bmatrix} \quad \mathcal{V}(f) = \begin{bmatrix} \mathbf{V}_{n'}^H(f) \\ \mathbf{V}_{n'-1}^H(f) \\ \vdots \\ \mathbf{V}_{n'-q+1}^H(f) \end{bmatrix} . \quad (9)$$

As can be seen from (8), the norm $\|\mathbf{G}(f)\|$ needs to be known for computing $\mathbf{F}(f)$. As indicated in [1], $\|\mathbf{G}(f)\|$ is less affected by small speaker movements than the individual transfer functions and will be assumed to be known, which is a disadvantage of this dereverberation algorithm.
In [1] the transfer function vector $\mathbf{G}(f)$ is calculated using an adaptive subspace tracking procedure, which estimates the principal singular vector of $\mathcal{Y}(f)$. However, this procedure is only optimal for spatially white noise, while for spatially coloured noise the principal generalised singular vector has to be computed. Using recursive algorithms it is also possible to update the GSVD [6].

4. NOISE REDUCTION

In [2] a GSVD-based optimal filtering technique has been described in the time-domain for noise reduction in multi-microphone speech signals. This technique produces an optimal MMSE estimate of the speech components in all microphone signals (but achieves no dereverberation at all).
When translating this technique to the frequency-domain, the objective is to compute an $M \times M$ filter matrix $\mathbf{W}(f)$, such that the enhanced speech vector

$$\hat{\mathbf{X}}_n(f) = \mathbf{W}(f)\mathbf{Y}_n(f) \quad (10)$$

is the optimal MMSE estimate of the (reverberated) speech component $\mathbf{X}_n(f)$ in the microphone signals. The optimal filter matrix is the multi-dimensional Wiener filter $\mathbf{W}_{WF}(f)$,

$$\begin{aligned} \mathbf{W}_{WF}(f) &= \mathbf{R}_{xy}(f)\mathbf{R}_{yy}^{-1}(f) = \mathbf{R}_{xx}(f)\mathbf{R}_{yy}^{-1}(f) \quad (11) \\ &= \big(\mathbf{R}_{yy}(f) - \mathbf{R}_{vv}(f)\big)\mathbf{R}_{yy}^{-1}(f) \,. \quad (12) \end{aligned}$$

Using the rank-1 definition of $\mathbf{R}_{xx}(f)$ in (5),

$$\hat{\mathbf{X}}_n(f) = \mathbf{G}(f)\mathcal{E}\{|S_n(f)|^2\}\mathbf{G}^H(f)\mathbf{R}_{yy}^{-1}(f)\mathbf{Y}_n(f) \,, \quad (13)$$

which corresponds to the time-domain signals $\hat{x}_m[k]$.
The filter matrix $\mathbf{W}_{WF}(f)$ can be computed using the GEVD of $\mathbf{R}_{yy}(f)$ and $\mathbf{R}_{vv}(f)$ or the GSVD of $\mathcal{Y}(f)$ and $\mathcal{V}(f)$ as

$$\begin{aligned} \mathbf{W}_{WF}(f) &= \mathbf{Q}(f)\big(\mathbf{\Sigma}_y(f) - \mathbf{\Sigma}_v(f)\big)\mathbf{\Sigma}_y^{-1}(f)\mathbf{Q}^{-1}(f) \\ &= \frac{\sigma_x^2(f)}{\sigma_{y1}^2(f)}\mathbf{q}_1(f)\bar{\mathbf{q}}_1^H(f) \,, \quad (14) \end{aligned}$$

with $\sigma_{y1}^2(f)$ the principal generalised eigenvalue and $\bar{\mathbf{q}}_1^H(f)$ the corresponding row of $\mathbf{Q}^{-1}(f)$.

5. COMBINED DEREVERBERATION AND NOISE REDUCTION

The objective of combined dereverberation and noise reduction is to compute filters $\mathbf{C}(f)$ such that

$$\hat{S}_n(f) = \mathbf{C}^H(f)\mathbf{Y}_n(f) \quad (15)$$

is the optimal MMSE estimate of the clean speech signal $S_n(f)$, thereby taking into account both dereverberation and noise reduction. The optimal filter $\mathbf{C}(f)$ is equal to

$$\mathbf{C}(f) = \mathbf{R}_{yy}^{-1}(f)\mathbf{r}_{ys}(f) \ , \quad (16)$$

such that the optimal MMSE estimate is

$$\hat{S}_n(f) = \mathcal{E}\{|S_n(f)|^2\}\mathbf{G}^H(f)\mathbf{R}_{yy}^{-1}(f)\mathbf{Y}_n(f) \ , \quad (17)$$

which corresponds to the time-domain signal $\hat{s}[k]$.
Comparing (13) and (17), we notice that

$$\hat{\mathbf{X}}_n(f) = \mathbf{G}(f)\hat{S}_n(f) \ , \quad (18)$$

which implies that the optimal estimate $\hat{S}_n(f)$ of the clean speech signal can be obtained by performing the dereverberation technique of section 3 to the optimal estimate of the speech components $\hat{\mathbf{X}}_n(f)$. Since the same decomposition is used for both dereverberation and noise reduction, the two procedures can easily be combined.
Using (8) and (14), the filter $\mathbf{C}^H(f)$ can be written as

$$\begin{aligned} \mathbf{C}^H(f) &= \mathbf{F}^H(f)\mathbf{W}_{WF}(f) & (19)\\ &= \frac{\mathbf{q}_1^H(f)}{\|\mathbf{q}_1(f)\|\|\mathbf{G}(f)\|}\frac{\sigma_x^2(f)}{\sigma_{y1}^2(f)}\mathbf{q}_1(f)\bar{\mathbf{q}}_1^H(f) & (20)\\ &= \frac{\|\mathbf{q}_1(f)\|}{\|\mathbf{G}(f)\|}\frac{\sigma_x^2(f)}{\sigma_{y1}^2(f)}\bar{\mathbf{q}}_1^H(f) & (21)\end{aligned}$$

In the case of spatially white noise, the EVD of $\mathbf{R}_{yy}(f)$ is used and the matrix $\mathbf{Q}(f)$ is orthogonal, such that

$$\mathbf{C}_w^H(f) = \frac{\sigma_x^2(f)}{\sigma_{y1}^2(f)}\frac{\mathbf{q}_1^H(f)}{\|\mathbf{G}(f)\|} \ , \quad (22)$$

which is equal to the spatially white dereverberation filter $\mathbf{F}_w^H(f)$, up to the spectral weighting term $\sigma_x^2(f)/\sigma_{y1}^2(f)$.

6. PRACTICAL IMPLEMENTATION

In (4) we have assumed non-overlapping frames. However, in practice we will use frames of length N with an overlap of $N-L$ samples for computing the filters and for filtering the microphone signals. For this block-processing scheme it is well known that the underlying fast convolution in the frequency-domain should be constrained to be linear. Therefore, in order to avoid circular convolutions, we put the last $L-1$ taps of the time-domain filters to zero and only keep the last L samples of the time-domain filtered microphone signals in an overlap-save procedure [3].
E.g., for combined dereverberation and noise reduction, the microphone signals $\mathbf{Y}_n(f)$ in the frequency-domain are computed as the FFT of $\left[\, y_m[nL] \quad \ldots \quad y_m[nL+N-1] \,\right]$, $m = 0\ldots M-1$. The M-dimensional frequency-domain filters $\mathbf{C}(f), f = 0\ldots N-1$, are computed using (21). The N-dimensional time-domain filters $\mathbf{C}_m[k], m = 0\ldots M-1$, are obtained as the IFFT of $\mathbf{C}(f)$. These time-domain filters are constrained by putting the last $L-1$ taps to zero and are transformed to the constrained frequency-domain filters $\bar{\mathbf{C}}(f), f = 0\ldots N-1$. The enhanced speech signal is computed as $\hat{S}_n(f) = \bar{\mathbf{C}}^H(f)\mathbf{Y}_n(f)$. From the IFFT of $\hat{S}_n(f)$ the last L samples $\left[\, \hat{s}[(n-1)L+N] \ \ldots \ \hat{s}[nL+N-1] \,\right]$ are kept in an overlap-save procedure.

7. SIMULATIONS

In our simulations we have filtered a 16kHz speech signal and a white noise signal with two room impulse responses ($K = 1000$), constructed with the image method [9]. The room dimensions are 3 m × 3 m × 4 m, the speech source position is $\left[\, 1 \ 2 \ 2 \,\right]$ and the noise source position is $\left[\, 0.5 \ 1 \ 1 \,\right]$. We have used an array of $M = 4$ omni-directional microphones and the distance between adjacent microphones is 2 cm. The positions of the microphones are $\left[\, 1 \ 1 \ 1 \,\right]$, $\left[\, 1.02 \ 1 \ 1 \,\right]$, $\left[\, 1.04 \ 1 \ 1 \,\right]$, and $\left[\, 1.06 \ 1 \ 1 \,\right]$. The reverberation time T_{60} of the room is 400 msec. The smaller the microphone distance and the smaller the reverberation time, the more the frequency-domain signals are spatially correlated. The unbiased SNR of the noisy first microphone signal $y_0[k]$ is 0 dB. In all the algorithms we have used a frame length (FFT-size) $N = 1024$ and overlap $L = 16$.
As objective measures for the noise reduction performance we use the unbiased signal-to-noise ratio SNR_u,

$$\mathrm{SNR}_u = 10\log_{10}\frac{\sum \bar{x}^2[k]}{\sum \bar{v}^2[k]} \ , \quad (23)$$

with $\bar{x}[k]$ and $\bar{v}[k]$ the speech and noise component of the considered signal, the frequency weighted signal-to-noise ratio SNR_w [7] and the speech intelligibility index (SII) [8], which are both weighted subband SNRs. As an objective measure for dereverberation we use a dereverberation index (DI), which is defined as the standard deviation of the amplitude response of the total filter $h[k]$ between $s[k]$ and the speech component in the output signal $z[k]$.
Table 1 gives an overview of the objective dereverberation and noise reduction performance measures for the different algorithms. Figure 2 plots the noisy microphone signal $y_0[k]$ and the enhanced microphone signal $\hat{x}_0[k]$ using the GSVD-based noise reduction technique. As can be seen in table 1, this technique produces the highest SNR_w and SII, but does not achieve any dereverberation, since the DI of $\hat{x}_0[k]$ and $y_0[k]$ are almost equal (DI $\simeq$ 4.7).
Figures 3a and 3b show the amplitude responses of the total transfer function $H(f)$ between $s[k]$ and the speech component in the output signal for 2 dereverberation algorithms. Figures 4a and 4b depict the time-domain output signals $\tilde{s}_w[k]$ and $\tilde{s}[k]$ for these algorithms. As can be seen, the filter $\mathbf{F}(f)$ computed using the GEVD of $\mathbf{R}_{yy}[k]$ and $\mathbf{R}_{vv}[k]$

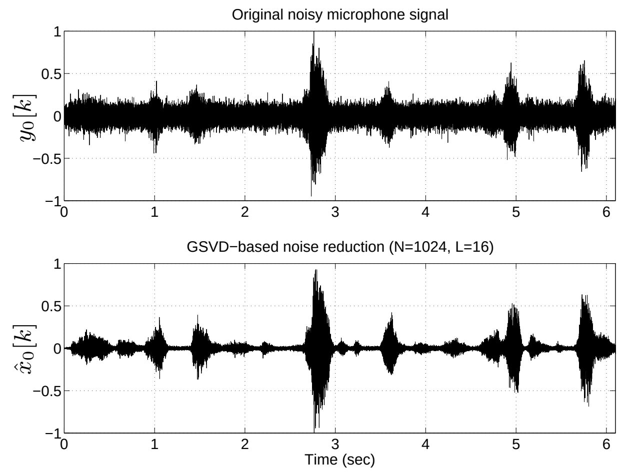

Figure 2: (**a**) Noisy microphone signal $y_0[k]$, (**b**) Enhanced microphone signal $\hat{x}_0[k]$ with GSVD-based noise reduction technique ($N = 1024, L = 16$)

	$\mathbf{SNR_u}$ (dB)	$\mathbf{SNR_w}$ (dB)	**SII**	**DI** (dB)
Noisy microphone signal $y_0[k]$	0	2.88	0.55	4.74
GSVD-based noise reduction $\hat{x}_0[k]$	17.81	**16.82**	**0.94**	4.73
SVD-based dereverberation $\tilde{s}_w[k]$	11.99	-0.30	0.46	1.86
GSVD-based dereverberation $\tilde{s}[k]$	15.10	2.30	0.54	**0.86**
Dereverberation and noise reduction $\hat{s}[k]$	**20.15**	10.12	0.79	1.35

Table 1: Dereverberation and noise reduction performance measures for the different algorithms ($N = 1024, L = 16$)

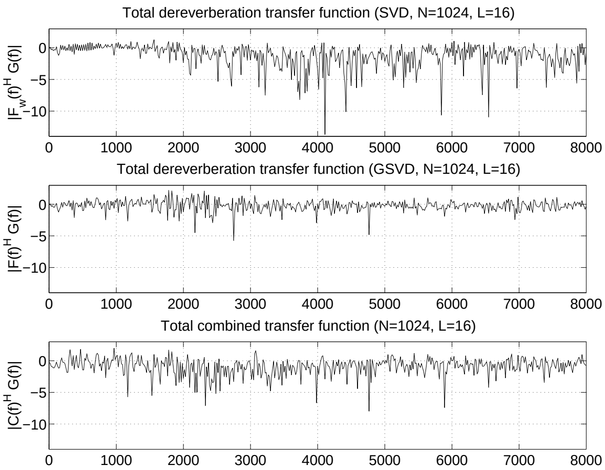

Figure 3: Total transfer function computed with (**a**) SVD-based dereverberation technique, (**b**) GSVD-based dereverberation technique, (**c**) combined noise reduction and dereverberation technique ($N = 1024, L = 16$)

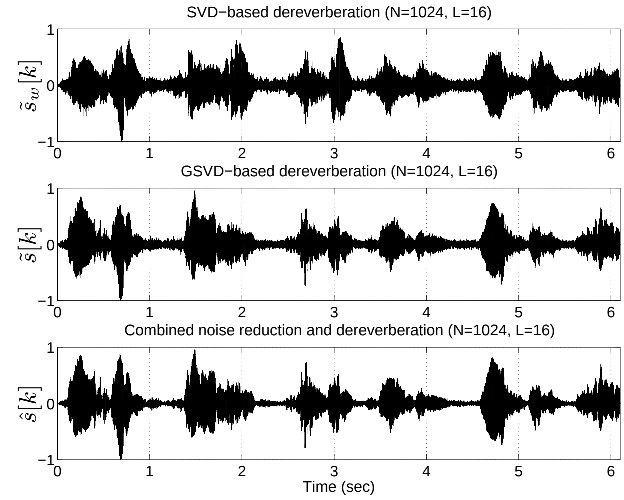

Figure 4: (**a**) SVD-based dereverberated signal $\tilde{s}_w[k]$, (**b**) GSVD-based dereverberated signal $\tilde{s}[k]$, (**b**) Enhanced signal $\hat{s}[k]$ with combined dereverberation and noise reduction technique ($N = 1024, L = 16$)

produces the flattest amplitude response (DI = 0.86), while the filter $\mathbf{F}_w(f)$ computed using the EVD of $\mathbf{R}_{yy}[k]$ (as in [1]) has a worse dereverberation performance (DI = 1.86). However, as can be seen in table 1, the noise reduction performance of both algorithms is quite bad, since the SNR_w and SII are smaller than for the noisy microphone signal. Figures 3c and 4c show the amplitude response of the total transfer function $H(f)$ and the time-domain output signal $\hat{s}[k]$ for the combined dereverberation and noise reduction technique. From table 1 it can be seen that its dereverberation performance is not as good as for the GSVD-based dereverberation technique (but it has better SNR values), while its noise reduction performance is not as good as for the GSVD-based noise reduction algorithm (but is has a better DI). It is therefore clear that the combined technique makes a trade-off between optimal dereverberation and noise reduction.

8. CONCLUSION

In this paper we have presented GSVD-based frequency-domain signal enhancement techniques for noise reduction and dereverberation. It has been shown that the optimal MMSE estimate of the clean speech signal can be obtained by dereverberating the optimal estimate of the speech components in the microphone signals. By simulations it has been shown that the combined algorithm makes a trade-off between the dereverberation and noise reduction objectives.

9. REFERENCES

[1] S. Affes and Y. Grenier, "A Signal Subspace Tracking Algorithm for Microphone Array Processing of Speech," *IEEE Trans. Speech, Audio Processing*, vol. 5, no. 5, pp. 425–437, Sept. 1997.

[2] S. Doclo and M. Moonen, *GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement*, chapter 6 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B.), pp. 111–132, Springer-Verlag, May 2001.

[3] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, pp. 15–37, Jan. 1992.

[4] S. Gannot and M. Moonen, "Subspace methods for multi-microphone speech dereverberation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Germany, Sept. 2001.

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, MD : John Hopkins University Press, Baltimore, 3rd edition, 1996.

[6] M. Moonen, P. Van Dooren, and J. Vandewalle, "A systolic algorithm for QSVD updating," *Signal Processing*, vol. 25, pp. 203–213, 1991.

[7] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[8] Acoustical Society of America, "ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index," June 1997.

[9] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1527–1529, 1986.