Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2010, Article ID 127513, 5 pages
doi:10.1155/2010/127513

Research Article

Signal-Based Performance Evaluation of Dereverberation Algorithms

Patrick A. Naylor, Nikolay D. Gaubitch, and Emanuël A. P. Habets

Communications and Signal Processing Group, Department of Electrical and Electronic Engineering, Imperial College, London SW7 2AZ, UK

Correspondence should be addressed to Nikolay D. Gaubitch, nikolay.gaubitch@imperial.ac.uk

Received 7 July 2009; Accepted 13 October 2009

Academic Editor: Vesa Valimaki

We address the measurement of reverberation in terms of the (DRR) in the context of the assessment of dereverberation algorithms for which we wish to quantify the level of reverberation before and after processing. The DRR is normally calculated from the impulse response of the reverberating system. However, several important dereverberation algorithms involve nonlinear and/or time-varying processing and therefore their effect cannot conveniently be represented in terms of modifications to the impulse response of the reverberating system. In such cases, we show that a good estimate of DRR can be obtained from the input/output signals alone using the Signal-to-Reverberant Ratio (SRR) only if the source signal is spectrally white and correctly normalized. We study alternative normalization schemes and conclude by showing a least squares optimal normalization procedure for estimating DRR using signal-based SRR measurement. Simulation results illustrate the accuracy of DRR estimation using SRR.

1. Introduction

When a speech signal is acquired in an enclosed space by one or more microphones positioned at some distance from the talker, each observed signal consists of a superposition of many delayed and attenuated copies of the speech signal due to multiple reflections from the surrounding walls and other objects. These multiple reflections can number several thousands and give rise to the effect known as reverberation. The reverberation time of an enclosed space is usually measured as the time, T_{60}, taken for the free-decay of reverberation to reduce by 60 dB and is affected by the volume of the enclosed space and the acoustic properties of the reflecting surfaces [1]. Efficient schemes for modeling reverberation are widely used, for example, the source-image method [2, 3]. A general scenario comprises a source speech signal $s(n)$ which propagates through M acoustic channels, assumed Linear Time Invariant (LTI), with impulse responses $h_M(n)$ and is acquired by M microphones with output signals $x_M(n)$. The microphone signals $x_M(n)$ therefore contain reverberated versions of the source signal $s(n)$. Dereverberation algorithms operate on $x_M(n)$ and output N estimates $\hat{s}_N(n)$ of the source signal $s(n)$. We will assume that $M = N = 1$ for the purposes of this paper with $\mathbf{h} = [h(0), h(1), \dots, h(L_h - 1)]^T$, $x(n) = \mathbf{h}^T\mathbf{s}(n)$, and $\mathbf{s}(n) = [s(n), s(n-1), \dots, s(n - L_h + 1)]^T$, where $\dot{T}$ represents the transpose operator and L_h is the number of taps in the impulse response.

The development of dereverberation algorithms [4] to reduce the reverberation effects in an audio signal is a slowly maturing topic in signal processing. Early work [5] introduced a speech enhancement approach operating on the linear prediction residual and several microphone array-based approaches [6, 7] have been proposed. Blind system identification techniques have been applied [8] involving subspace decomposition [9] and adaptive filters [10]. Techniques to evaluate dereverberation algorithms are as yet not consistently defined and research is underway to address this issue. A common measure of dereverberation performance will be summarized in Section 2, where the difference between channel-based Direct-to-Reverberation Ratio (DRR) and signal-based Signal-to-Reverberation Ratio

(SSR) measures will be highlighted. The remainder of the paper will focus on signal-based measures for which normalization is not straightforward. We will justify the need for correct normalization and then briefly study alternative schemes in Section 3.

2. Measures of Reverberation

We here define the direct path as an L_h-tap impulse response $\mathbf{h}_\mathrm{d} = [h_\mathrm{d}(0), h_\mathrm{d}(1), \ldots, h_\mathrm{d}(L_h-1)]^T$ representing propagation from the talker to a microphone without reflections. We assume $\mathbf{h}_\mathrm{d}$ is known. We also define the reverberant component $\mathbf{h}_\mathrm{r} = [h_\mathrm{r}(0), h_\mathrm{r}(1), \ldots, h_\mathrm{r}(L_h - 1)]^T$ as an impulse response representing all nondirect propagation paths from talker to microphone. We therefore write

$$x(n) = \mathbf{h}^T\mathbf{s}(n) = \mathbf{h}_\mathrm{d}^T\mathbf{s}(n) + \mathbf{h}_\mathrm{r}^T\mathbf{s}(n) = s_\mathrm{d}(n) + x_\mathrm{r}(n), \qquad (1)$$

where $s_\mathrm{d}(n)$ is a delayed and scaled version of $s(n)$.

In general, the measurement of the level of reverberation in a signal requires a comparison of the energy due to the direct path propagation and the energy due to the reverberant paths. This may be characterized as the DRR which will be discussed below. Evaluation of the performance of a dereverberation algorithm can classified into two approaches: channel based and signal based.

2.1. Channel-Based Measure. Channel-based measures are appropriate when the effect of the dereverberation algorithm on the reverberating system impulse response $\mathbf{h}$ is known or can be deduced. The DRR can be found straightforwardly from the corresponding impulse response coefficients [1] as

$$\mathrm{DRR} = 20\log_{10}\left(\frac{\|\mathbf{h}_\mathrm{d}\|_2}{\|\mathbf{h} - \mathbf{h}_\mathrm{d}\|_2}\right)\mathrm{dB}. \qquad (2)$$

If the direct path propagation time corresponds to an integer number of sampling periods then $\mathbf{h}_\mathrm{d}$ may be an impulse; otherwise it has the form of a sinc function [3]. Comparison of DRR before and after processing leads to a measure of improvement in DRR. We note that, in contrast to the evaluation of dereverberation using improvement in DRR, evaluation of system identification performance is usually done in terms of the Normalized Projection Misalignment [11].

2.2. Signal-Based Measure. Signal-based measures are needed when the effect of a dereverberation algorithm cannot be characterized in terms of an impulse response, such as [5, 6, 12], where the processing is not LTI. In such cases it is necessary to determine the SRR only from the signals before and after processing. The SRR can be written

$$\mathrm{SRR} = 20\log_{10}\left(\frac{\|\mathbf{s}_\mathrm{d}\|_2}{\|\hat{\mathbf{s}} - \mathbf{s}_\mathrm{d}\|_2}\right)\mathrm{dB}, \qquad (3)$$

where $\mathbf{s}_\mathrm{d} = [s_\mathrm{d}(0), s_\mathrm{d}(1), \ldots, s_\mathrm{d}(L_s - 1)]^T$, $s_\mathrm{d}(n) = \mathbf{h}_\mathrm{d}^T\mathbf{s}(n)$, and $\hat{\mathbf{s}} = (\hat{\mathbf{s}}_\mathrm{d} + \hat{\mathbf{x}}_\mathrm{r})$ is the reverberant signal to be measured of length L_s samples, for example, at the input and the output of a dereverberation algorithm in order to measure the improvement in DRR achieved. The SRR is an intrusive measure that requires both the original and the processed speech signals. In addition, knowledge of the direct path component of the true impulse response is assumed in our approach such that the speech signals can be time-aligned correctly.

2.3. Relationship between DRR and SRR. Subject to correct level normalization as will be discussed below, the SRR is equivalent to the DRR when the source $s(n)$ is spectrally white. In the case when $\hat{\mathbf{s}}_\mathrm{d} = \mathbf{s}_\mathrm{d}$ and evoking Parseval's theorem, in the frequency domain we have $\sum_k |S(k)|^2|H_\mathrm{d}(k)|^2/\sum_k |S(k)|^2|H_\mathrm{r}(k)|^2$. When $S(k) = S$, independent of k, $|S|^2$ can be taken outside the summation in both numerator and denominator and cancelled. An illustrative example is when $s(n) = \delta(n)$, so that $S(k) = 1$ for all k, in which case (3) reduces directly to the formulation of the DRR in (2). In practice, when speech signals are considered, a prewhitening filter can be employed [13] as will be shown below.

These effects are illustrated in Figure 1 which shows a comparison of DRR and SRR for a room of dimensions 6 × 5 × 4 m simulated using the source-image method [2, 3] (left) and for real measured room impulse responses from MARDY [14] (right). The SRR calculated for a white noise input is shown in curve (a) and is seen to correspond almost exactly to DRR. Curve (b) shows SRR calculated for five sentences of male speech, sampled at 20 kHz from the APLAWD database [15]. Lastly the results with prewhitened speech are shown in curve (c). The prewhitening filters were computed over all five sentences using a 10th order linear predictor; separate filters were obtained for $\mathbf{s}_\mathrm{d}$ and $\hat{\mathbf{s}}$ and were applied to each of the signals, respectively. It is clear that whitening the speech signal has a significant effect.

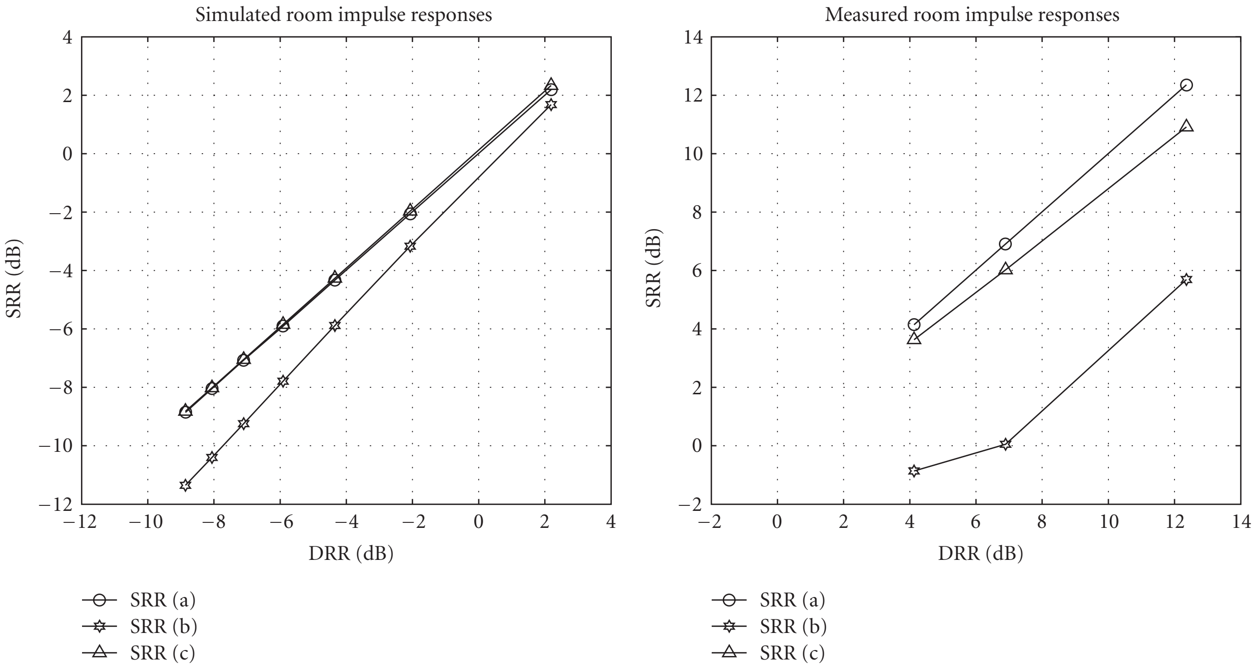
3. Level Normalization

A dereverberation algorithm aims to attenuate the level of reverberation and may affect either or both of the direct path signal $s_\mathrm{d}(n)$ or the reverberant component $x_\mathrm{r}(n)$ in order to improve the SRR. Therefore we can write that

$$\hat{s}(n) = \alpha s_\mathrm{d}(n) + \overline{x}_\mathrm{r}(n), \qquad (4)$$

where $\overline{x}_\mathrm{r}(n)$ is the reverberant component remaining after dereverberation processing and α is a scalar assumed stationary over the duration of the measurement. We also assume that any processing delay has been appropriately compensated as is generally assumed in other measurements such as the SNR.

We propose that the measurement of the reverberant component's energy and the assessment of its impact on the speech signal must be done relative to the energy of the direct path component. This can be conveniently accomplished by normalization in order to match the level of the direct path component before and after processing. The aim of this normalization is to adjust the magnitude of $\hat{\mathbf{s}}$ such that the direct path signal energy is unchanged by the dereverberation



(a) (b)

Figure 1: Comparison of DRR and SRR for (a) white Gaussian noise input, (b) speech input, and (c) prewhitened speech input, with simulated impulse responses (left) and mesured impulse responses (right).

algorithm. This can be achieved by determining α. Our motivation comes from the observation that signal-based measures are not, in general, scale independent as can be seen in the case of (3) and therefore misleading results can be obtained unless the scaling is correctly normalized.

We formulate this problem as a search for a scalar $\hat{\alpha}$ such that the Normalized Signal-to-Reverberation Ratio (NSRR)

$$\text{NSRR} = 20\log_{10}\left(\frac{\|\mathbf{s}_\text{d}\|_2}{\|(1/\hat{\alpha})\hat{\mathbf{s}} - \mathbf{s}_\text{d}\|_2}\right)\text{dB} \tag{5}$$

is a good estimate of DRR.

3.1. RMS and Peak Normalization. It is necessary to estimate α from the available signals and, for baseline comparison purposes, we have initially considered straightforward approaches to determining α using

$$\alpha_\text{norm} = \frac{\|W\{\hat{\mathbf{s}}\}\|_\text{norm}}{\|W\{\mathbf{s}_\text{d}\}\|_\text{norm}} \tag{6}$$

corresponding to RMS and peak matching for norm $=$ 2 and norm $= \infty$, respectively, and employing uniform and A-weighting [1] for $W\{\cdot\}$ representing a corresponding weighting filter. These approaches lead to incorrect calculation of SRR as will be shown below.

3.2. Least Squares Optimal Normalization. We propose that a good solution to the normalization problem can be obtained using α_ls from the least squares minimization

$$\alpha_\text{ls} = \arg\min_{\hat{\alpha}} \|\hat{\mathbf{s}} - \hat{\alpha}\mathbf{s}_\text{d}\|_2^2. \tag{7}$$

The solution for α_ls is found by minimizing $J = E\{\|\hat{\mathbf{s}} - \hat{\alpha}\mathbf{s}_\text{d}\|_2^2\}$ arising from (7), where $E\{\cdot\}$ denotes mathematical expectation.

To minimize J, we differentiate it with respect to $\hat{\alpha}$ and set the result to zero, which gives

$$\frac{\partial J}{\partial \hat{\alpha}} = -2E\left\{\mathbf{s}_\text{d}^T[\hat{\mathbf{s}} - \hat{\alpha}\mathbf{s}_\text{d}]\right\} = 0. \tag{8}$$

The final step is to approximate expectations with sample averages giving α_ls to be the value of $\hat{\alpha}$ satisfying (8) as

$$\alpha_\text{ls} = \frac{\mathbf{s}_\text{d}^T\hat{\mathbf{s}}}{\mathbf{s}_\text{d}^T\mathbf{s}_\text{d}}, \tag{9}$$

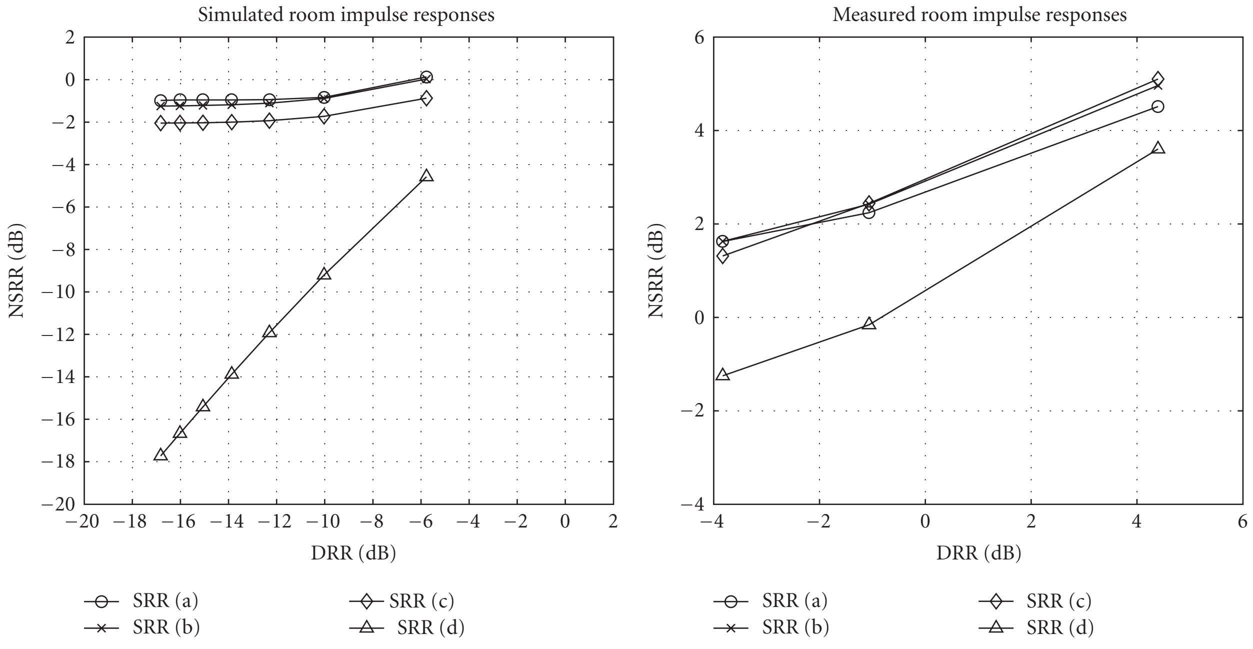
which is a projection of $\hat{\mathbf{s}}$ onto the direct component $\mathbf{s}_\text{d}$.

The effect of $\hat{\alpha}$ is seen by substituting (4) into J to obtain

$$\begin{aligned} J &= E\left\{\|\alpha\mathbf{s}_\text{d} + \overline{\mathbf{x}}_\text{r} - \hat{\alpha}\mathbf{s}_\text{d}\|_2^2\right\} \\ &= E\left\{(\alpha - \hat{\alpha})^2\|\mathbf{s}_\text{d}\|_2^2\right\} + E\left\{2(\alpha - \hat{\alpha})\mathbf{s}_\text{d}^T\overline{\mathbf{x}}_\text{r}\right\} + E\left\{\|\overline{\mathbf{x}}_\text{r}\|_2^2\right\}. \end{aligned} \tag{10}$$

Clearly, J is minimized when $\alpha = \hat{\alpha}$. Although the normalization constant has been considered stationary, it could also be applied in a frame-based manner as, for example, in Segmental SNR.

3.3. Results. Figure 2 shows a comparison of DRR with NSRR computed from (5) with $\hat{\alpha}$ obtained using four different level normalization schemes. These results were



(a)

(b)

FIGURE 2: Comparison with DRR and NSRR calculated using (a) peak normalization, (b) RMS normalization, (c) A-weighted RMS normalization, and (d) least squares optimal normalization.

obtained for the same experimental setup as in Section 2.3. The test signal $\hat{s}$ was generated as in (4) with α chosen arbitrarily and $\bar{x}_r(n) = x_r(n)$. The speech signals were prewhitened with prewhitening filters computed from $\mathbf{s}_d$ and $(1/\hat{\alpha})\hat{\mathbf{s}}$ and applied, after the level normalization, to each of the signals, respectively. Curves (a), (b), and (c) show SRR with the normalization factor α from (6) with peak normalization, RMS normalization, and A-weighted RMS normalization, respectively. Curve (d) shows SRR with least squares optimal normalization. It can be seen that the match between DRR and least squares optimal normalized SRR is much smaller over a wide range of DRRs; whereas other normalization schemes substationally overestimate and offer little discrimination between different values of DRR. These discrepancies are more severe at lower DRR values.

4. Discussion and Conclusions

An important class of dereverberation algorithms employ nonlinear and/or time-varying processing such that the effect of their processing on the reverberation cannot be characterized in terms of an impulse response. In such cases, the improvement in DRR cannot be measured directly. Accordingly, it is necessary to estimate the DRR values at the input and output of the dereverberation algorithm using SRR.

We have shown that two effects require consideration. First, the signal characteristics affect the SRR calculation such that good estimates of DRR are obtained when the signal is white. Prewhitening of speech with a 10th-order predictor has been seen to be sufficient for the cases studied here. Second, the level of the signals must be correctly normalized. We have shown that level normalization using RMS, A-weighted RMS, and peak matching are not appropriate. We have formulated a least squares optimal normalization scheme and shown that this can be expressed as a projection of the signal onto the direct path component. Simulation results confirm that the least squares optimal level normalization and prewhitening enable DRR to be estimated without the requirement for impulse response measurements.

References

[1] H. Kuttruff, *Room Acoustics*, Taylor & Frances, Boca Raton, Fla, USA, 4th edition, 2000.

[2] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[3] 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.

[4] P. A. Naylor and N. D. Gaubitch, "Speech dereverberation," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '05)*, Eindhoven, The Netherlands, September 2005.

[5] B. Yegnanarayana and P. Satyanarayana Murthy, "Enhancement of reverberant speech using LP residual signal," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 267–281, 2000.

[6] S. M. Griebel and M. S. Brandstein, "Wavelet transform extrema clustering for multi-channel speech dereverberation,"

in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '99)*, Pocono Manor, Pa, USA, September 1999.

[7] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, Berlin, Germany, 2001.

[8] G. Xu, H. Liu, L. Tong, and T. Kailath, "Least-squares approach to blind channel identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, 1995.

[9] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1074–1090, 2003.

[10] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 882–895, 2005.

[11] D. R. Morgan, J. Benesty, and M. Mohan Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 174–176, 1998.

[12] N. Gaubitch, P. A. Naylor, and D. B. Ward, "On the use of linear prediction for dereverberation of speech," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '03)*, pp. 99–102, Kyoto, Japan, September 2003.

[13] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1978.

[14] J. Wen, N. D. Gaubitch, E. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '06)*, Paris, France, September 2006.

[15] G. Lindsey, A. Breen, and S. Nevard, "SPAR's archivable actual-word databases," Tech. Rep., University College London, London, UK, June 1987.