Late Reverberation Suppression Using MMSE Modulation Spectral Estimation

Chenxi Zheng[1] *and Wai-Yip Chan*[1]

[1]Department of Electrical & Computer Engineering, Queen's University, Kingston, Canada

Abstract

The echo effect due to late reverberation can severely degrade speech quality and intelligibility. Prior attempts to reduce this degradation in the modulation domain used time-invariant filtering. In this paper, we show that performing minimum mean squared error spectral estimation in the modulation domain (MDMMSE) can significantly reduce the severity of audible reverberation and enhance the listening speech quality. Moreover, MDMMSE outperforms modulation domain spectral subtraction (SS) as well as performing MMSE spectral estimation or SS in the acoustic domain. In informal subjective listening tests, MDMMSE exhibits less residual echoes and artifacts than the other three methods.

Index Terms: Speech enhancement, dereverberation, modulation domain, MMSE, late reverberation reduction

1. Introduction

It is well known that an acoustically propagated speech signal in an enclosure is distorted by reflections from walls and objects. This reverberation distortion decreases the subjective listening quality and intelligibility of the speech signal, increasingly with the distance of the listener from the sound source. Though multiple microphone dereverberation can significantly ameliorate the reverberation degradation [1], there are situations where only one microphone is available. Thus, various single-microphone dereverberation algorithms have been proposed [2]-[8]. Some of these techniques comprise two stages of processing, with one stage ameliorating the effect of early reverberation and the other dealing with late reverberation. In all, these dereverberation algorithms can be classified into two categories: reverberation suppression and reverberation cancelation. Reverberation cancellation, also known as blind deconvolution, aims to equalize the room impulse response (RIR). Reverberation cancellation is mainly effective in cancelling early reverberations. Reverberation cancellation is reported either unable to improve speech quality [8] or can only be applied to small reverberation times (RTs) [6].

Reverberation suppression, on the other hand, views late reverberation as additive noise, and aims to suppress it as done in noise reduction. Reverberation suppression can be implemented in either the acoustic frequency domain or modulation frequency domain. Acoustic frequency domain is defined as the short time Fourier transform (STFT) domain of a speech signal, while modulation frequency domain is defined as the STFT domain of the time trajectory of an acoustic frequency domain component, for a fixed acoustic frequency bin. Acoustic frequency domain speech signal analysis, modification, and synthesis is the most common method for speech enhancement. Several recently proposed single-channel dereverberation methods, such as [3, 5], are based on this scheme. On the other hand, psychoacoustic and physiological evidence has shown a connection between modulations and speech intelligibility [9], and algorithms [10, 11] using time-invariant filtering along the time trajectory of STFT components have been proposed. These algorithms tried to restore the modulation magnitude spectra of the clean speech signal; however, little quality or intelligibility improvement was reported. However, for denoising, recent modulation domain algorithms show improved speech quality [12, 13, 14].

We observe that while previous modulation domain dereverberation algorithms [10, 11] use time-invariant filtering, many denoising ("speech enhancement") algorithms [15] track the spectral signal-to-noise ratio and hence are inherently time-varying. In this paper, we investigate whether modulation domain enhancement can provide better speech quality than acoustic domain enhancement. Specifically, we employ two speech enhancement methods, minimum mean squared error estimation (MMSE) and spectral subtraction (SS), and compare the speech quality obtained from suppressing late reverberation in the acoustic domain versus modulation domain. The rest of this paper is organized as follows: In Section 2, we establish the relative significance of (suppressing) early and late reverberation. In Section 3, we present our modulation domain MMSE dereverberation algorithm (MDMMSE). Section 4 gives the algorithm parameter selection results obtained using informal subjective listening tests. Section 5 uses objective measures and subjective listening tests to assess the speech quality obtained from the four dereverberation algorithms. Conclusions are drawn in section 6.

2. Perceptual effects of RIR components

In this section, we use informal subjective listening and an objective speech quality measure to evaluate the subjective listening effects of different RIR components, in order to establish the relative impacts of late reverberation versus early reverberation.

2.1. Time domain reverberation model

In a reverberant room, the reverberated speech signal $z(n)$ is modeled as produced by the convolution of the clean speech signal $s(n)$ and RIR $h(n)$

$$z(n) = \sum_{i=0}^{Q-1} h(i)s(n-i), \tag{1}$$

where Q is the length of $h(n)$. The RIR depends on the acoustic conditions of the room and contains all the directly propagated and reflected pulses from the sound source to the reception point. For analyzing the listening effects, the RIR can be partitioned into three components: the direct signal, early reflections, and late reflections. The direct signal travels along the direct path from the speech source to the listener. The early reflections are the pulses that arrive within 50 ms after arrival of the

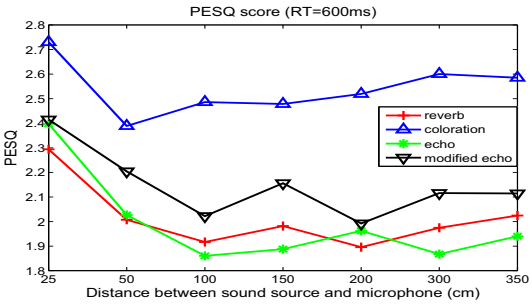

Figure 1: Average PESQ score for a reverberated speech dataset with variable source microphone distance and fixed RT=600 ms.

direct signal. Early reflections are known to cause short-term reverberation or "coloration" effects. Late reflections, which arrive at times greater than 50 ms past the direct pulse, smear the speech spectrum and may severely reduce speech quality and intelligibility. Late reflections cause the so-called long-term reverberation or echo effect.

It is mentioned in [16] that early reflections can effectively boost signal energy as well as emphasize modulation frequency content around 4 Hz, thus they do not degrade intelligibility. Indeed, several dereverberation algorithms [2, 3, 5] only suppress late reverberation. However, other dereverberation algorithms suppress both early and late reverberation, such as [6, 7]. Below, we seek to clarify the subjective effects of early and late reflections, and whether early reverberation suppression is needed. An experiment is designed to assess the subjective listening effects of different RIR components.

2.2. Speech quality of different RIR components

Three modified and the original RIRs are used in this experiment. The first RIR comprises the direct path impulse and early reflections. The corresponding convolved speech signal is called coloration corrupted signal. The second RIR is composed of the direct path impulse and late reflections. The corresponding convolved speech is named echo corrupted signal. The third RIR is composed of a modified direct path impulse and the late reflections. The modified direct path impulse has the energy of the sum of the original direct path impulse and early reflections. The corresponding convolved signal is called modified echo corrupted signal. A reverberated speech dataset is created by convolving image method [17] generated RIRs (and their modifications) with Wall Street Journal November 92 (WSJ) clean speech data set. The WSJ data set consists of 330 sentences, uttered by eight different speakers, both male and female, and is sampled at 16 kHz. Perceptual evaluation of speech quality (PESQ) score [18] is used as an objective measure (on 8 kHz downsampled speech), and the average PESQ score over the whole dataset is plotted in Fig. 1 for increasing source-microphone distance d_{sm}, when RT is 600 ms. Fig. 1 shows that late reverberation degrades speech quality more severely than early reverberation. The fact that echo corrupted speech performs somewhat poorer than reverberated speech when source-microphone is larger than 50 cm shows the role of early reverberation in boosting reverberated speech quality. The gap between the modified echo corrupted speech and the echo corrupted speech provides an upper bound on the performance of doing early reverberation equalization. Informal subjective listening supports the PESQ objective test results in Fig. 1, specifically the quality ranking of the four RIRs and the trends as a function of d_{sm}.

On one hand, it is shown that echo corrupted speech has worse quality than reverberated speech for relatively large d_{sm}, which not only corroborates the claim in [16], but also demonstrates that just removing coloration and not suppressing late reverberation actually will decrease speech quality. Compared with reverberated speech, the maximum quality improvement given by the modified-echo upper bound is modest, in comparison with the potential maximum quality gain that could be obtained from suppressing late reverberation. Without the echo part, the coloration part will decrease speech quality to some extent. The above results suggest that only when the degradation due to late reverberation is sufficiently ameliorated, equalization of early reverberation becomes gainful. Since late reverberation is the crucial part degrading speech quality and its dominance increases with RT, we propose a RT adaptive late reverberation reduction algorithm.

3. Modulation domain dereverberation

In this section, we lay out the processing steps of our MD-MMSE dereverberation algorithm. Our modulation domain spectral subtraction (MDSS) dereverberation algorithm is described in [19].

3.1. Modulation domain dereverberation scheme

The processing steps applied to reverberated speech $z(n)$ are shown in Fig. 2. First, $z(n)$ is windowed by an acoustic analysis window. A K-point DFT is then taken and the outputs are acoustic domain spectral magnitude $Z_{p,k} = |z_{p,k}|$ ($p = 1...P, k = 1...K$) and phase $\angle Z_{p,k}$. Here p indexes the acoustic domain windowed speech frame and k the DFT coefficients. Along p, the $Z_{p,k}$'s are then windowed by an modulation analysis window. An M-point DFT is taken and the modulation domain spectral magnitude $Z_{j,k,m}$ is input to a late reverberation spectral variance (LRSV) estimation module. Here j indexes the modulation analysis window frames, k the acoustic domain DFT coefficients, and m the modulation domain DFT coefficients. The estimated LRSV $\sigma^2_{l_{j,k,m}}$ and $Z_{j,k,m}$ are input to a MMSE estimator to estimate the modulation magnitude spectral component of the early reverberated speech $\widehat{Z}_{e_{j,k,m}} = |\widehat{z}_{e_{j,k,m}}|$. Then, $\widehat{Z}_{e_{j,k,m}}$ and the unmodified modulation spectral phase of the reverberated speech $\angle z_{j,k,m}$ are input to an M-point IDFT and further windowed by a modulation synthesis window. Overlap-and-add is used to reconstruct the enhanced acoustic magnitude spectrum $\widehat{Z}_{e_{p,k}}$. $\widehat{Z}_{e_{p,k}}$ and the unmodified acoustic spectral phase $\angle Z_{p,k}$ are then input to a K-point IDFT before windowed by an acoustic synthesis window. Overlap-and-add is used to reconstruct the final enhanced signal $\widehat{z}_e(n)$, an estimate of the early reverberated speech signal. In our experiments, a square-root Hann window is used throughout.

3.2. MMSE estimation in the modulation domain

We treat the MMSE estimation of the magnitude spectrum in the modulation domain similarly to the MMSE spectral estimation problem in [20], except that a second frequency dimension (modulation frequency dimension) is introduced. As such, the estimate of the magnitude spectrum in the modulation domain, $\hat{Z}_{e_{j,k,m}}$ can be computed as:

$$\hat{Z}_{e_{j,k,m}} = G_{j,k,m} Z_{j,k,m}. \quad (2)$$

Here, $G_{j,k,m}$ is the gain function in the modulation domain, and can be computed as:

$$G_{j,k,m} = \frac{\sqrt{\pi}}{2} \frac{\sqrt{v_{j,k,m}}}{\gamma_{j,k,m}} \Lambda[v_{j,k,m}], \quad (3)$$

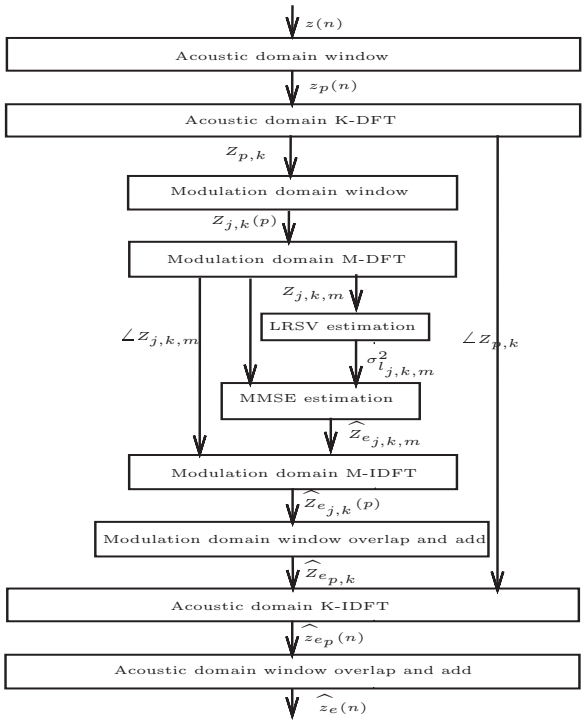

Figure 2: Modulation domain dereverberation scheme.

in which

$$\Lambda[\theta] = exp(-\frac{\theta}{2})[(1+\theta)I_0(\frac{\theta}{2}) + \theta I_1(\frac{\theta}{2})], \quad (4)$$

and

$$v_{j,k,m} = \frac{\xi_{j,k,m}}{1+\xi_{j,k,m}}\gamma_{j,k,m}. \quad (5)$$

Here, $I_0(.)$ and $I_1(.)$ are the zeroth and first order modified Bessel functions. $\xi_{j,k,m}$ and $\gamma_{j,k,m}$ are *a priori* and *a posteriori* SNRs, which are defined as:

$$\xi_{j,k,m} = \frac{E[|Z_{e_{j,k,m}}|^2]}{E[|Z_{l_{j,k,m}}|^2]}, \quad (6)$$

and

$$\gamma_{j,k,m} = \frac{|Z_{j,k,m}|^2}{E[|Z_{l_{j,k,m}}|^2]}, \quad (7)$$

where $E[|Z_{l_{j,k,m}}|^2]$ is the noise variance we call LRSV. An estimate of LRSV, $\sigma^2_{l_{j,k,m}}$ is obtained using the method described in Sec. 3.3 below. The *a posteriori* SNR can be estimated as:

$$\widehat{\gamma}_{j,k,m} = \frac{|Z_{j,k,m}|^2}{\sigma^2_{l_{j,k,m}}}. \quad (8)$$

The *a priori* SNR $\xi_{j,k,m}$ is estimated by the "decision-directed" method as:

$$\xi'_{j,k,m} = \alpha\frac{|\hat{Z}_{e_{j-1,k,m}}|^2]}{\sigma^2_{l_{j-1,k,m}}} + (1-\alpha)max[\hat{\gamma}_{j,k,m} - 1, 0]. \quad (9)$$

Here, α is a tradeoff parameter between late reverberation reduction and transient distortion [20]. A lower bound is used to prevent the *a priori* SNR from falling below a prescribed value:

$$\hat{\xi}_{j,k,m} = max[\xi'_{j,k,m}, \xi_{min}]. \quad (10)$$

3.3. Model based estimation of LRSV

In [2], a statistical model for the RIR is proposed and used to derive a LRSV estimator. Subsequently, [5] derived a refined LRSV estimator which reduces the estimation error when the source-to-microphone distance is less than the critical distance, or equivalently, the direct-to-reverberation ratio (DRR) greater than 0 dB. The estimators of [2] and [5] operate on acoustic spectral data of the reverberated signal. In this work, we adopt the approach of [2] and [5] to the modulation domain, i.e., modulation spectral samples of reverberated speech $Z_{j,k,m}$ are modeled as resultant from the convolution of a modulation-subband impulse response $h_{j,k,m}$ and the modulation spectral samples of anechoic speech $s_{j,k,m}$:

$$z_{j,k,m} = \sum_{j'} s_{j',k,m}h_{j-j',k,m} = s_{j,k,m} * h_{j,k,m}. \quad (11)$$

Similar to the acoustic frequency domain RIR model [5], modulation domain RIR $h_{j,k,m}$ is modeled as:

$$h_{j,k,m} = \begin{cases} B_{d_{k,m}} & j = 0; \\ B_{r_{j,k,m}}e^{-v_{k,m}jT_{MFS}} & j > 0. \end{cases} \quad (12)$$

Here, $B_{d_{k,m}}$ and $B_{r_{j,k,m}}$ are assumed to be zero mean independent Gaussian random variables, with variance $\sigma^2_{d_{k,m}} = E\{|B_{d_{k,m}}|^2\}$ and $\sigma^2_{r_{k,m}} = E\{|B_{r_{j,k,m}}|^2\}$. T_{MFS} is the modulation frame shift in seconds. $v_{k,m}$, the decay rate, depends on the RT $T_{60}(k,m)$ (seconds) in acoustic frequency bin k and modulation frequency bin m as:

$$\Delta_{k,m} = \frac{3ln10}{T_{60}(k,m)} \quad (13)$$

We assume the distance between the source and microphone is larger than the critical distance and the energy of the direct path impulse is small compared to the total energy of the other reflections. Thus, we can assume $\sigma^2_{r_{k,m}} = \sigma^2_{d_{k,m}}$. Similar to the LRSV estimation in the acoustic frequency domain [2], the relation between spectral variance of the reverberated speech $\sigma^2_{j,k,m} = E\{|z_{j,k,m}|^2\}$ and that of late reverberation speech $\sigma^2_{l_{j,k,m}} = E\{|z_{l_{j,k,m}}|^2\}$ in the modulation domain is given as:

$$\sigma^2_{l_{j,k,m}} = e^{-2\Delta_{k,m}T_l}\sigma^2_{j-J_l,k,m}. \quad (14)$$

Here, T_l is the time interval that separates early reflections from late reflections, and is 50 ms in this experiment. J_l is the frame number corresponding to T_l and is equal to $int(\frac{T_l}{T_{MFS}} + 0.5)$.

The spectral variance of reverberated speech $\sigma^2_{j,k,m}$ is obtained recursively from $Z^2_{j,k,m}$ as:

$$\sigma^2_{j,k,m} = \kappa\sigma^2_{j-1,k,m} + (1-\kappa)Z^2_{j,k,m}. \quad (15)$$

The value of the smoothing factor κ is optimized using subjective listening, as described below. For LRSV estimation, we assume that the required RIR parameters can be blindly estimated. [21] proposes blind estimators (assessed in [22]) of RT using modulation spectral features of reverberated speech. Hence, these estimators can exploit the modulation analysis computation in the proposed dereverberation algorithm. For the rest of this paper, the acoustic domain RT estimated from the RIR is used for all the subband RT parameters.

4. Algorithm parameter tuning

We used informal subjective listening to optimize the MD-MMSE algorithm parameters, which include the modulation domain frame length T_{MFL}, modulation frame shift T_{MFS}, decision-directed smoothing parameter α, *a priori* SNR lower bound ξ_{min}, and smoothing parameter κ. Our listening experiments show that $T_{MFL} = 128$ ms, $T_{MFS} = 8$ ms, $\alpha = 0.998$, $\xi_{min} = -15$ dB, and $\kappa = 0$ give the best subjective listening quality. When $T_{MFL} > 128$ ms, the speech becomes

more slurred. When $T_{MFL} < 128$ ms, the speech tends to have more distortion, and this distortion becomes obvious when $T_{MFL} = 32$ ms. T_{MFS} is selected as 8 ms, as that can give a relatively accurate estimate of LRSV in (14), in which $J_l = 6$. Different from acoustic domain LRSV estimation [5], $\kappa = 0$ gives the best subjective listening quality. When $\kappa > 0$ is used, more residual reverberation is heard. This is likely due to the smoothing effect of the longer frame used in the modulation domain processing.

5. Experiments

5.1. Databases

Two clean speech datasets and RIR datasets are used in our experiments. The first clean speech dataset consists of 128 clean speech files, spoken by two male and two female subjects. Reverberated speech datasets are generated by corrupting the clean speech dataset using the SImulation of REal ACoustics (SIREAC) tool [23], with RT values ranging from 0.5-2 s provided by the tool. The speech files originally sampled at 16 kHz are downsampled to 8 kHz due to PESQ algorithm restriction. The second clean speech dataset is the afore-mentioned WSJ data set. Reverberated speech data sets are generated by convolving clean speech files with the Aachen RIR data set using an ITU-T G.191 tool. The Aachen data set [24] comprises binaural RIRs measured in various enclosures, including an office, lecture room, stairway, and auditorium (Aula Caronila). The first channel of the binaural RIR is used in our experiment. RT values are measured using the Schroeder method [25]. Source-microphone distances are all larger than critical distances in these two RIR data sets. All the clean and reverberated speech files are level-normalized to -26 dBov using the P.56 voltmeter [26]. PESQ score is used as objective measure to compare acoustic domain spectral subtraction (ADSS) [2], acoustic domain MMSE (ADMMSE) [7], MDSS [19], and the proposed MDMMSE algorithm. For ADSS, ADMMSE, and MDSS, we use the same parameters as documented in the literature. Informal subjective listening is used to validate objective measurement.

5.2. Experiment result

Fig. 3 shows the average PESQ scores of reverberated speech for the first dataset and the dereverberated speech for the four dereverberation schemes. It is shown that the modulation domain dereverberation algorithms generally perform better than their acoustic domain counterparts. This improvement becomes more obvious when RT becomes larger, and this is confirmed by our subjective listening. However, our informal subjective listening shows MDMMSE to sound better than MDSS, which contradicts the PESQ test results.

Fig. 4 shows the objective measurement results of the second data set. The highest objective score for each RIR is boldfaced. Comparing the PESQ scores of reverberated speech for the four enclosures, we see the PESQ scores drop as RT and d_{sm} increase. Despite that the auditorium has a much larger RT and the speech sounds much more reverberant than the ones for the office with the same $d_{sm} = 200$ cm, their PESQ scores are almost identical. The PESQ scores across different enclosures do not correctly reflect the actual and relative subjective listening quality. When we compare the scores across the rows of the table, we see quality improvement for all four dereverberation algorithms and this is corroborated by the reduction of reverberation we hear in our subjective listening tests. PESQ scores

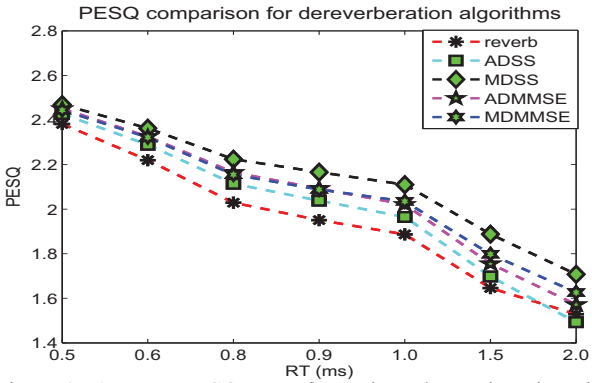

Figure 3: Average PESQ score for various dereverberation algorithms and RTs.

Enclosure(d_{sm}(cm))	RT(ms)	Reverb	ADSS	MDSS	ADMMSE	MDMMSE
Office(200)	440	2.1743	2.2702	2.2949	**2.3091**	2.2750
Office(300)	480	2.0859	2.1273	**2.1949**	2.1728	2.1679
Lecture(400)	720	2.0731	2.1579	**2.2874**	2.2153	2.2132
Lecture(868)	810	1.9771	2.0699	**2.2198**	2.1423	2.1215
Stairway(300)	980	1.9564	2.0562	**2.2064**	2.1243	2.0886
AulaCarolina(200)	4910	2.1252	2.1603	2.3194	2.2718	**2.3300**

Figure 4: Average PESQ score comparison of the dereverberation algorithms using RIRs in the Aachen data set [24].

also show that the modulation domain algorithms perform better than the acoustic domain algorithms, especially for MDSS. However, in our subjective listening tests, MDMMSE performs consistently better than the other algorithms, even MDSS. For enclosures with small RT, such as the office in [24], MDMMSE performs slightly better than the other algorithms; MDMMSE exhibits less residual reverberation and introduces less distortion. This improvement becomes more pronounced for enclosures with large RTs, such as the stairway and Aula Carolina in [24]. One visible evidence in the waveforms is that MDSS dereverberated speech has longer residual reverberation tails at the ends of talkspurts. ADMMSE dereverberated speech, on the other hand, exhibits more audible fluctuating distortions than MDMMSE dereverberated speech. This is probably partly due to the larger modulation frame length (128ms) used in MDMMSE than the 32 ms acoustic frame length in ADMMSE.

6. Conclusion

Prior works on modulation domain dereverberation of speech rely on time-invariant filtering. In this paper, we take a spectral enhancement approach to suppressing late reverberation in the modulation domain. Through informal subjective listening, we found that our MDMMSE algorithm significantly reduces the severity of audible reverberation and enhances the listening speech quality. Moreover, MDMMSE provides better dereverberated speech quality than MDSS, as well as MMSE and SS dereverberation in the acoustic domain. We also found that PESQ scores fail in various ways to provide reliable rating and ranking of (de)reverberated speech quality. PESQ scores fail to reflect the severity of audible late (residual) reverberation as well as distortions induced by dereverberation algorithm processing.

7. References

[1] K. Eneman and M. Moonen, "Multimicrophone speech dereverberation: Experimental validation," *EURASIP J. Audio, Speech, Music Process.*, 2007, 19 pages.

[2] K. Lebart, J. M. Boucher and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica*, vol. 87, pp. 359-366, 2001.

[3] H. W. Lollmann and P. Vary, "A blind speech enhancement algorithm for the suppression of late reverberation and noise," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3989-3992, 2009.

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

[5] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Processing Letters*, pp. 770–773, 2009.

[6] M. Wu and D. Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Trans. on Audio, Speech, Lang. Process.*, pp. 774–784, May 2006.

[7] E. A. P. Habets, N. D. Gaubitch, and P. A. Naylor, "Temporal selective dereverberation of noisy speech using one microphone, " in *International Conference on Acoustics, Speech and Signal Processing*, pp. 4577–4580, 2008.

[8] S. Mosayyebpour, H. Sheikhzadeh, T. A. Gulliver, and M. Esmaeili, "Single-Microphone LP Residual Skewness-Based Inverse Filtering of the Room Impulse Response," *IEEE Trans. on Audio, Speech, Lang. Process.*, pp. 1617–1632, July 2012.

[9] N. Mesgarani and S. Shamma, "Speech enhancement based on filtering the spectrotemporal modulations," in *Proc. ICASSP*, vol. 1, pp. 1105-1108, Mar 2005.

[10] C. Avendano and H. Hermansky, "Study on the dereverberation of speech based on temporal envelope filtering," in *Proc. Fourth Int. Conf. Spoken Language*, vol. 2, pp. 889-892, 1996.

[11] A. Kusumoto, T. Arai, K. Kinoshita, N. Hodoshima, and Nancy Vaughan, "Modulation enhancement of speech by a preprocessing algorithm for improving intelligibility in reverberant environments," *Speech Comm.*, pp. 101-113, 2005.

[12] K. Paliwal, K. Wojcicki, and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," *Speech Comm.*, 52 (5), pp. 450-475, 2010.

[13] K. Paliwal, B. Schwerin, and K. Wojcicki, "Speech enhancement using a minimum mean- square error short-time spectralmodulation magnitude estimator," *Speech Comm.*, 54(2), pp. 282–305, 2012

[14] T. H. Falk, S. Stadler, W. B. Kleijn, and W.-Y. Chan, "Noise suppression based on extending a speech-dominated modulation band," in *Proc. ISCA Conf. Internat. Speech Commun. Assoc. (INTERSPEECH)*, pp. 970-973, 2007

[15] P. Loizou, "Speech Enhancement: Theory and Practice," Taylor and Francis, Boca Raton, FL, 2007.

[16] T. H. Falk, C. Zheng, and W.-Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Trans. on Audio, Speech and Language Processing*, pp. 1766-1774, 2010.

[17] B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J Acoust. Soc. Amer.*, vol. 65, pp. 943-950, Apr, 1978.

[18] ITU-T Rec. P.862, Perceptual evaluation of speech quality (PESQ), ITU, Geneva, 2001.

[19] C. Zheng, "Single-Microphone Speech Dereverberation: Modulation Domain Processing and Quality Assessment," Queen's Univeristy M.Sc. thesis, July 2011. Available at http://hdl.handle.net/1974/6609.

[20] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[21] T. H. Falk and W.-Y. Chan, "Temporal Dynamics for Blind Measurement of Room Acoustical Parameters," *IEEE Trans. Instrum. Meas.*, Vol. 59, No. 4, pp. 978-989, April 2010.

[22] N. D. Gaubitch, H. W. Löllmann, M. Jeub, T. H. Falk, P. A. Naylor, P. Vary, and M. Brookes, "Performance Comparison of Algorithms for Blind Reverberation Time Estimation from Speech," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aachen, Germany, 2012.

[23] H. Hirsch and H. Finster, "The simulation of realistic acoustic input scenarios for speech recognition systems," in *Proc. Interspeech*, pp. 2697-2700, 2005.

[24] M. Jeub, M. Schäfer, and P. Vary, "A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms," in *Proceedings of International Conference on Digital Signal Processing (DSP)*, pp. 1-4, July 2009.

[25] M. R. Schroeder, "New method of measureing reverberation time," J. Acoust. Soc. Am., vol. 37, no. 3, pp. 409-412, 1965.

[26] ITU-T P.56, "Ojective Measurement of Active Speech Level," Int. Telecom. Union, 1993.