# JOINT BLIND DEREVERBERATION AND SEPARATION OF SPEECH MIXTURES

*Tariqullah Jan*† *and Wenwu Wang*‡

† Department of Electrical Engineering,
University of Engineering & Technology Peshawar, Pakistan
Email: tariqullahjan@yahoo.com
‡ Centre for Vision, Speech, and Signal Processing,
University of Surrey, UK
Email: w.wang@surrey.ac.uk

## ABSTRACT

This paper proposes a method for jointly performing blind source separation (BSS) and blind dereverberation (BD) for speech mixtures. In most of the previous studies, BSS and BD have been explored separately. It is common that the performance of the speech separation algorithms deteriorates with the increase of room reverberations. Also most of the dereverberation algorithms rely on the availability of room impulse responses (RIRs) which are not readily accessible in practice. Therefore in this work the dereverberation and separation method are combined to mitigate the effects of room reverberations on the speech mixtures and hence to improve the separation performance. As required by the dereverberation algorithm, a step for blind estimation of reverberation time (RT) is used to estimate the decay rate of reverberations directly from the reverberant speech signal (i.e., speech mixtures) by modeling the decay as a Laplacian random process modulated by a deterministic envelope. Hence the developed algorithm works in a blind manner, i.e., directly dealing with the reverberant speech signals without explicit information from the RIRs. Evaluation results in terms of signal to distortion ratio (SDR) and segmental signal to reverberation ratio (SegSRR) reveal that using this method the performance of the separation algorithm that we have developed previously can be further enhanced.

## 1. INTRODUCTION

Speech signals captured by microphones in an enclosed environment are often contaminated by room reverberations and the intereferences from the nearby sound sources. Separation of the target speech from the microphone signals is a challenging task because of the interfering speech signals, and the presence of reverberations makes it more challenging. Therefore, it is important to develop a method which can separate the target speech from the interfering sounds and can also reduce the adverse acoustic effects.

We have recently develped a source separation algorithm in [5], however its performance deteriorates in the presence of room reverberations. Therefore, we have developed recently a dereverberation algorithm to suppress the room reverberation, and in current work this dereverberation algorithm is combined with the separation algorithm to enhance the separation performance. However the dereverberation algorithms usually assume the RIRs to be known *a priori*, which in practice are not available. To address this problem, a method is proposed in this work for the blind estimation of RT and then incorporated with the dereverberation algorithm. The proposed blind RT estimation method uses the reverberant speech (i.e., mixture) directly to estimate the decay rate instead of the RIRs. In the proposed method, a Laplacian distribution based decay model for room reverberation is used along with an efficient procedure for locating the free decay in reverberant speech. Finally, the proposed RT estimation method is incorporated with the algorithms developed for separation and dereverberation to obtain a joint blind dereverberation and separation method for the speech mixtures.

The developed joint algorithm which is a two channel method has been employed in two different ways. Firstly, the available mixture signals are used to estimate blindly the RT based on a maximum-likelihood (ML) method and statistical modelling of the sound decay rate of the reverberant speech, followed by the dereverberation of the mixture signals using the method we have proposed and based on the frequency depenedent statistical model. Then the separation algorithm we have proposed in [5] is applied to these resultant mixtures so that the source speech signals can be obtained. Secondly, the separation algorithm [5] is applied first to the mixtures to segregate the speech signals, followed by the blind estimation of RT from the separated speech signal. Then dereverberation is employed to the segregated speech signals.

The rest of the paper is organized as follows. Section 3 briefly reviews the separation algorithm we have developed in [5] followed by the proposed derverberation method in Section 4. Section 5 explains briefly our proposed RT estimation algorithm, followed by Section 6 in which the proposed joint blind dereverberation and separation algorithm has been discussed. Section 7 evaluates the performance of the joint blind dereverberation and separation algorithm and reports the experimental results. Section 8 concludes the paper.

## 2. PROBLEM FORMULATION AND MODELLING

In a cocktail party environment, $N$ speech signals are recorded by $M$ microphones, which can be described mathematically by a linear convolutive model

$$z_j(n) = \sum_{i=1}^{N} \sum_{p=1}^{P} h_{ji}(p) s_i(n-p+1) \qquad (j = 1, ..., M) \quad (1)$$

where $s_i$ and $z_j$ are the source and mixture signals respectively, $h_{ji}$ is a $P$-point room impulse response [1] from source $s_i$ to microphone $x_j$. In this study a two-input two-output system has been considered, i.e., $N = M = 2$. The main goal here is to separate the source signals from the mixtures $z_j(n)$ in (1) and also to remove the reverberant effects present in the mixture signals $z_j(n)$ to further enhance the separation performance. The next section will describe the method used for the separation of speech signals from the mixtures $z_j(n)$.

## 3. A MULTISTAGE APPROACH TO BLIND SEPARATION OF CONVOLUTIVE SPEECH MIXTURES

In this section we will provide a brief overview of a method we have proposed in [5] for the blind separation of convolutive speech mixtures. In this algorithm, we first apply a convolutive independent component analysis (ICA) algorithm [15] to the microphone recordings. As is common with many other existing ICA algorithms, the separated target speech from this step still contains a considerable amount of interference from other sources. The performance steadily degrades with an increase of RT. In order to reduce the interference within the target speech, we estimate the ideal binary mask (IBM) by comparing the energy of the corresponding time-frequency (T-F) units from the outputs of the convolutive ICA algorithm, and then apply the estimated IBM to the original mixtures to obtain the target speech and interfering sources. Our experimental results reveal that this process considerably improves the separation performance by reducing the interference to a much lower level. However, a typical problem with the binary T-F masking is the introduction of errors in the estimation of the masks. The errors may result in some isolated T-F units, causing fluctuating musical noise [9].

To reduce such noise the estimated IBM is further processed by using cepstral smoothing [9]. More specifically, we transform the binary mask into the cepstral domain, and smooth the transformed mask over time frames using the overlap-and-add technique. In the cepstrum domain, it is easier to distinguish between the unwanted isolated random peaks and mask patterns resulting from the spectral structure of the segregated speech. Therefore, different levels of smoothing can be applied to the binary T-F mask in different frequency ranges. The smoothed mask, after being transformed back into the T-F plane, is then applied to the outputs of the previous step in order to reduce the musical noise. Further details about the method can be found in [5].

## 4. BLIND DEREVERBERATION

To improve the separation performance of [5], we develop a dereverberation algorithm based on the frequency depenedent statistical model [2] and use it for the suppression of late reverberations. The frequency dependent RIR model used by our dereverberation method is given as

$$H_{late}(m,k) = \begin{cases} \beta(m,k)e^{-\alpha(k)mR} & \text{for } m \geq 1, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $\beta(m,k)$ is a sequence of zero-mean mutually independent and identically distributed (i.i.d.) Gaussian random variables, $m$ is the time frame index, $k$ is the frequency bin index, $R$ denotes the hop size, and $\alpha(k)$ denotes the decay rate which can be obtained from the frequency dependent reverberation time $T_{60}(k)$ as below

$$\alpha(k) \triangleq \frac{3ln(10)}{T_{60}(k)f_s} \tag{3}$$

where $f_s$ is the sampling frequency. Then $T_{60}(k)$ can be estimated using the proposed blind RT estimation method which will be discussed in the Section 5. The estimated $T_{60}(k)$ can be used then to obtain the decay rate $\alpha(k)$. Then the spectral variance of the late reverberant speech can be estimated as

$$\sigma^2_{x_{late}}(m,k) = e^{-2\alpha(k)RN_{le}} \cdot \sigma^2_x(m-N_{le},k) \tag{4}$$

where $\sigma^2_x(m,k)$ is the variance of the reverberant speech which can be estimated by recursive averaging

$$\sigma^2_x(m,k) = e^{-2\alpha(k)R}[\tau \cdot \sigma^2_x(m-1,k) + (1-\tau) \cdot |\, X(m,k)\,|^2] \tag{5}$$

where $\tau \in [0,1]$ is a forgetting factor and $X(m,k)$ is the T-F representation of the reverberant speech $x(n)$. Note that $N_{le}$ is the number of hops after which the late reverberation begins and $e^{-2\alpha(k)R}$ measures the reverberation decay rate. We can then estimate the *posteriori* signal-to-distortion ratio (SDR) [6] as follows

$$\varphi(m,k) = \frac{|\, X(m,k)\,|^2}{\sigma^2_{x_{late}}(m,k)} \tag{6}$$

To reduce the late reverberations, we apply the following spectral subtraction mask [6] to $X(m,k)$

$$\tilde{G}_{late}(m,k) = 1 - \frac{1}{\sqrt{\varphi(m,k)}} \tag{7}$$

In order to avoid over-estimation of $\sigma^2_{x_{late}}(m,k)$, a lower bound $\tilde{G}^{min}_{late}$ is applied to all the weighting gains in the mask. The musical noise induced by spectral mask estimation error is further reduced by smoothing (Further details can be found in [3]).

## 5. BLIND RT ESTIMATION

The key parameter that needs to be estimated for the dereverberation algorithm is $T_{60}(k)$, which in blind separation, has to be estimated from the microphone signal. Here we model the reverberation decay as a Laplacian random process modulated by a deterministic envelope, and use an efficient procedure for locating free decay from the reverberant speech. Then the RT is estimated from the free decays by a maximum likelihood (ML) estimator. The method was motivated by our observation that the distribution pattern for temporal decay of the reverberant hand clap is much closer to the Laplace distribution [4].

Therefore, the reverberant tail of a decaying sound is modeled using a sequence of random variables with Laplace distribution $\mathscr{L}(\theta,\beta)$, where $\theta$ is the mean considered as zero here and $\beta$ is the variance of the Laplace distribution. The model is based on the assumption that the reverberation tail of a decaying sound denoted here as $y$ is the product of a fine structure denoted as $r$ that is a random process, and an envelop $a$ that is deterministic. Suppose $r(n)$ is a random sequence for $n \geq 0$, of i.i.d. random variables having Laplace distribution with zero mean and variance $\beta$, $\mathscr{L}(0,\beta)$. Similarly for each $n$ a deterministic sequence is defined as $a(n) > 0$. As a result, the model for the room decay $y$ is represented as $y(n) = a(n)r(n)$ [13]. As $a(n)$ is a time varying term, $y(n)$ are independent but not identically distributed, with probability density function $\mathscr{L}(0,\beta a(n))$.

Suppose that a single decay rate $\rho$ defines the damping of the sound envelop during the regions of free decay (i.e.,


Figure 1: Block diagram showing the first scheme for the proposed joint blind dereverberation and separation algorithm. $z_1(n)$ and $z_2(n)$ are the available mixtures (microphone signals).

the period following the sharp offset of a speech sound) instead of those regions where the sound is actually ongoing, onset, or gradually declining speech offsets. As a result the sequence $a(n)$ is determined by [13]

$$a(n) = exp(-n/\rho) \tag{8}$$

Hence, the $N$-dimensional parameter $a(n)$ can be replaced by a single scalar parameter $a$ which is denoted by $\rho$ as

$$a = exp(-1/\rho) \tag{9}$$

As a result Equation (8) can be written as

$$a(n) = a^n \tag{10}$$

The log likelihood function used by the proposed blind RT estimation method is given as

$$lnL(a_j;\mathbf{y}) = -Nln(2) - \sum_{n=0}^{N-1} ln(a_j^n \cdot \beta) - \frac{1}{\beta}\sum_{n=0}^{N-1} a_j^{-n} \mid y(n) \mid \tag{11}$$

whereas the range of $a \in [0,1)$ be quantized into $Q$ values, so that $a_j$ is obtained with $j = 1, ..., Q$. Then, for each $a_j$, the log-likelihood given by Equation (11) is calculated. The ML criterion can be used to select the best estimate of $a$ given as

$$\hat{a} = argmax_{a_j}\{lnL(a_j;\mathbf{y})\} \tag{12}$$

then the decay rate $\rho$ in (9) is estimated from $\hat{a}$ obtained in (12) which results in the estimation of RT.

To further improve the computational efficiency it would be helpful to capture the free sound decay regions first in the reverberant speech signal so that only the detected sound decay regions can be used for the ML estimation of the decay rate. Lollmann *et al.* [8] devised an estimation procedure which can be used for this purpose. Such a procedure also has the advantage in reducing the effects of the outliers on the estimated RT value. We have used this efficient procedure in our proposed RT estimation method to improve the ML estimation of the Laplacian parameters.

Now to estimate $T_{60}(k)$ in (3) in each frequency band, we used the blind RT estimation method discussed above for the filtered reverberant signal. First, we pass the reverberant signal $x(n)$ through a Gammatone filter-bank to get sub-band signals $x(p,n)$, where $p$ is the sub-band index. Subsequently, $x(p,n)$ were analyzed using Schroeder's method [14] to estimate the reverberation time $\check{T}_{60}(p)$ in each sub-band $p$. Since this filterbank (indexed by $p$) is different from the one used in the above section (indexed by $k$), the $\check{T}_{60}(p)$ values need to be inter- and extra-polated to obtain the estimate of $T_{60}(k)$ in each frequency bin $k$.

First we apply interpolation to $\check{T}_{60}(p)$ so that $\check{T}_{60}(p)$ from each sub-band $p$ is mapped to $\tilde{T}_{60}(f)$, where $f \in [f_c - \frac{bw}{2}, f_c + \frac{bw}{2}]$ denotes the frequency range (in Hz) of sub-band $p$, $f_c$ and $bw$ are the centre frequency and the bandwidth of this sub-band respectively. Then we apply smoothing across the overlapped regions between the neighbouring sub-bands

$$\tilde{T}_{60}(f) = \tilde{T}_{60}(f_1) + \frac{\tilde{T}_{60}(f_2) - \tilde{T}_{60}(f_1)}{f_2 - f_1}(f - f_1) \tag{13}$$

where $f_1$ and $f_2$ are the frequency points of the neighbouring sub-bands at which their overlap begins and ends respectively. $\tilde{T}_{60}(f_1)$ and $\tilde{T}_{60}(f_2)$ are the reverberation times at frequency points $f_1$ and $f_2$ respectively. For non-overlapped regions, no such interpolation as (13) is required for $\tilde{T}_{60}(f)$. Finally, $\tilde{T}_{60}(f)$ is then mapped to the STFT sub-bands by an extrapolation method as

$$T_{60}(k) = \sum_{f=(k-1)\frac{F}{K}+1}^{k\cdot\frac{F}{K}} \tilde{T}_{60}(f)/(F/K-1) \tag{14}$$

Note that, $f = 1, 2, ..., F$, where $F$ is the whole frequency range and $K$ denotes the number of frequency bins (indexed by $k$).

## 6. JOINT BLIND DEREVERBERATION AND SEPARATION

This section presents results of joint blind dereverberation and separation algorithm for speech mixtures proposed in this paper. The proposed method is assessed in three different ways. In the first scheme, mixture signals are employed to estimate the RT blindly using the proposed blind RT estimation method followed by the blind dereverberation using frequency dependent statistical models employing the RT obtained from the previous step to estimate the spectral variance of room reverberation and then the spectral subtraction mask and the smoothed mask which is used to dereverberate the mixtures. Next the separation algorithm developed in [5] (called as Multistage algorithm hereafter) is applied to the dereverberated mixtures in order to segregate the speech signals. A block diagram is given in Figure 1 explaining the structure of this scheme.

In the second arrangement, the Multistage algorithm is applied first to the mixtures to obtain the separated speech signals. Then using the proposed blind RT estimation method, RT is estimated blindly from the separated speech

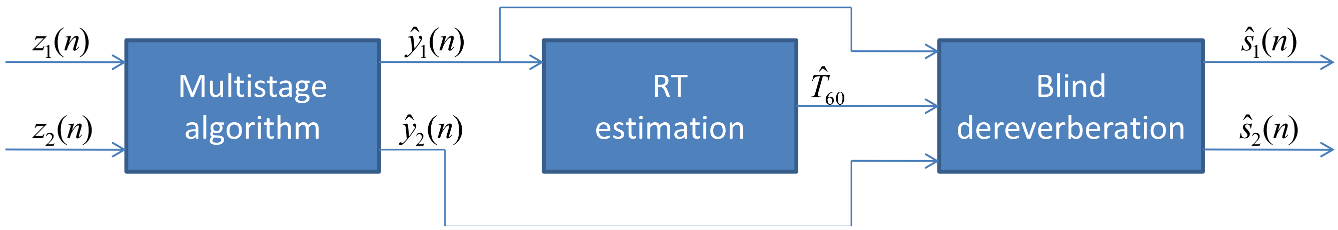


Figure 2: Block diagram showing the second scheme for the proposed joint blind dereverberation and separation algorithm. $z_1(n)$ and $z_2(n)$ are the available mixtures (microphone signals).

followed by the frequency dependent statistical model employing the estimated RT from the previous step to estimate the spectral variance of room reverberations and then spectral subtraction mask and the smoothed mask which is used to dereverberate the separated signals. A block diagram is given in Figure 2 describing the second scheme.

## 7. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of the proposed joint blind dereverberation and separation method has been evaluated using simulated RIRs from the image model [1] and real room recordings that were obtained in [12]. A pool of ten different speech signals from the TIMIT database, uttered by 5 male and 5 female speakers and all sampled at 16 KHz, has been used in the experiments to generate the reverberant mixtures. A system with two inputs and two outputs is considered here in this work. The size of the room used in the case of simulated RIRs is 6.5 x 7 x 8 ($m^3$). The position matrices of two sources and two sensors (microphones) are set as, $[1\ \ 1\ \ 3; 3\ \ 1\ \ 3]$, and $[2\ \ 3\ \ 3; 3\ \ 3\ \ 3]$ respectively. Performance indices used in the evaluations are the segmental signal to reverberation ratio (SegSRR) [7], and the signal to distortion ratio (SDR) [10, 11]. Notations $\Delta SegSRR$ and $\Delta SDR$ are used in the evaluations, where $\Delta SegSRR = mSegSRR_o - mSegSRR_i$ and $\Delta SDR = mSDR_o - mSDR_i$. $SegSRR_i$ and $SDR_i$ are the input $SegSRR$ and $SDR$ respectively. Similarly $SegSRR_o$ and $SDR_o$ are the output $SegSRR$ and $SDR$ respectively. Note that $mSegSRR_o$, $mSegSRR_i$, $mSDR_o$, and $mSDR_i$ are the average results for fifty random tests. The performance of the methods proposed in this paper are compared with that of the Multistage algorithm which we proposed in [5].

First the simulated room model [1] is used to generate the reverberant mixture signals from the pool of the clean speech signals described above, at different reverberation times, i.e., $T_{60} = \{200, 250, 300, 350, 400, 450, 500\}$ ms to evaluate and compare the performance of the proposed method at different RTs. For each $T_{60}$, ten anechoic signals from the pool have been used to generate different reverberant mixtures, with each consisting of two speech sources randomly picked up from the pool. In total 50 random tests have been carried out for each $T_{60}$, and hence in total 350 different reverberant mixtures have been used here in the evaluation. Table 1 and 2 show for each $T_{60}$, the results averaged over the 50 random tests for the first and second scheme of the proposed method respectively in comparison to the Multistage algorithm.

In another set of experiments real room recordings have been used that were obtained in [12]. The real recordings were made in a reverberant room with $T_{60}$ = 400 ms. Two omnidirectional microphones vertically placed and closely spaced are used for the recordings. Different loudspeaker positions are used to measure the room impulse responses. The room dimensions are 5.2 x 7.9 x 3.5 ($m^3$), and the distance between the microphones and the loudspeakers is 2 m. Further details about the recordings can be found in [12]. Clean speech signals from the pool of ten speakers were convolved with the room impulses to generate the source signals. The average results of $\Delta SDR$ and $\Delta SegSRR$ over the 50 different random tests are given in Table 3 and 4 for the first and second scheme of the proposed method respectively.

Table 1: $\Delta SDR$ and $\Delta SegSRR$ For Simulated Data under Different $T_{60}s$

| $T_{60}$ (ms) | $\Delta SDR$ (dB) | | $\Delta SegSRR$ (dB) | |
|---|---|---|---|---|
| | Proposed method (scheme 1) | Multistage method | Proposed method (scheme 1) | Multistage method |
| 200 | 4.52 | 3.61 | 2.15 | 1.45 |
| 250 | 3.73 | 2.91 | 1.88 | 1.14 |
| 300 | 3.22 | 2.45 | 1.66 | 0.94 |
| 350 | 2.88 | 2.18 | 1.48 | 0.82 |
| 400 | 2.68 | 1.96 | 1.35 | 0.75 |
| 450 | 2.50 | 1.77 | 1.23 | 0.68 |
| 500 | 2.37 | 1.62 | 1.12 | 0.63 |

Now if the results obtained for both simulated and real data are observed in a sequence of the different schemes, it can be found that the proposed method implemented in the first scheme consistently giving better results both in terms of SDR and SegSRR than the Multistage algorithm. For the real recordings, the proposed method in scheme 1 achieves approximately 1.5 dB gain for both SDR and SegSRR over the Multistage algorithm. It is observed that in the first scheme blind dereverberation applied to the reverberant mixtures prior to separation helps in improving the separation performance. Similarly it can be found that the proposed method in

Table 2: $\Delta SDR$ and $\Delta SegSRR$ For Simulated Data under Different $T_{60}s$

| $T_{60}$ (ms) | $\Delta SDR$ (dB) | | $\Delta SegSRR$ (dB) | |
|---|---|---|---|---|
| | Proposed method (scheme 2) | Multistage method | Proposed method (scheme 2) | Multistage method |
| 200 | 4.49 | 3.61 | 2.06 | 1.45 |
| 250 | 3.73 | 2.91 | 1.78 | 1.14 |
| 300 | 3.20 | 2.45 | 1.55 | 0.94 |
| 350 | 2.88 | 2.18 | 1.37 | 0.82 |
| 400 | 2.63 | 1.96 | 1.22 | 0.75 |
| 450 | 2.42 | 1.77 | 1.10 | 0.68 |
| 500 | 2.27 | 1.62 | 1.01 | 0.63 |

Table 3: $\Delta SDR$ and $\Delta SegSRR$ For the Real Data

| Algorithm | $\Delta SDR$ (dB) | $\Delta SegSRR$ (dB) |
|---|---|---|
| Proposed method (scheme 1) | 6.40 | 3.55 |
| Multistage method | 4.74 | 2.01 |

Table 4: $\Delta SDR$ and $\Delta SegSRR$ For the Real Data

| Algorithm | $\Delta SDR$ (dB) | $\Delta SegSRR$ (dB) |
|---|---|---|
| Proposed method (scheme 2) | 4.85 | 2.54 |
| Multistage method | 4.74 | 2.01 |

the second scheme also performs better than the Multistage algorithm for both simulated and real data. However, it can be noticed that in the second scheme of the proposed method improvement is less than the improvement achieved in the first scheme especially for real recordings. This is because in the second scheme, the separation algorithm is applied first and hence the enhancement performance is not as good as in the first scheme due to the reverberant effects in the mixture at the time of separation. Therefore, it is concluded that the proposed blind dereverberation and separation algorithm implemented in the first scheme provides better results in comparison to the implementation of the second scheme.

## 8. CONCLUSION

In this paper a method has been developed to perform blind dereverberation and separation of convolutive speech mixtures jointly. The method has been evaluated in two different schemes. In the first scheme, the mixture signal is used to estimate RT followed by blind dereverberation and then the separation algorithm is applied to the dereverberated mixture to obtain the segregated speech signals. In the second scheme, the separation algorithm is applied first to the mixtures. Then the obtained separated signal is used to estimate the RT blindly followed by the blind dereverberation. Our experiments indicate that the proposed method implemented in the first scheme performs better than the second scheme. Both schemes outperform our old approach

**Acknowledgment**

T. Jan was supported by the UET Peshawar Pakistan, and W. Wang was supported in part by the EPSRC of the UK (Grant EP/H050000/1 and EP/H012842/1).

## REFERENCES

[1] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, 1979.

[2] E. A. P. Habets, S. Gannot and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Letters*, vol. 16, pp. 770–773, Sep. 2009.

[3] T. Jan and W. Wang, "Frequency dependent statistical model for the suppression of late reverberations," *IEEE Statistical Signal Process. Workshop*, submitted Feb. 2012.

[4] T. Jan and W. Wang, "Blind reverberation time estimation based on Laplace distribution," *20th European Signal Processing Conference (EUSIPCO 2012)*, submitted Mar. 2012.

[5] T. Jan, W. Wang and D. L. Wang, "A multistage approach to blind separation of convolutive speech mixtures," *Speech Communication*, vol. 53, pp. 524–539, 2011.

[6] M. Jeub, M. Schafer, T. Esch and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 18, pp. 1732–1745, Sep. 2010.

[7] P. Krishnamoorthy and S. R. Mahadeva Prasanna, "Reverberant speech enhancement by temporal and spectral processing," *IEEE Trans. Audio, Speech, and Lang. Process.*, pp. 253–266, vol. 17, Feb. 2009.

[8] H. W. Lollmann, E. Yilmaz, M. Jeub and P. Vary, "An improved algorithm for blind reverberation time estimation," *Proc. Int. Workshop Acoust. Echo and Noise Control (IWAENC)*, Aug. 2010, Tel Aviv, Israel.

[9] N. Madhu, C. Breithaupt and R. Martin, "Temporal smoothing of spectral masks in the cepstral domain for speech separation," *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, pp. 45–48, Las Vegas USA, 2008.

[10] M. I. Mandel, R. J. Weiss and P. W. Ellis, "Model-based expectation maximization source separation and localization," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 18, pp. 382–394, Feb. 2010.

[11] P. A. Naylor, N. D. Gaubitch and E. A. P. Habets, "Signal-based performance evaluation of dereverberation algorithms," *Journal of Electrical and Computer Engineering*, Vol. 2010, Article ID 127513, 5 pages, doi:10.1155/2010/127513, 2010.

[12] M. S. Pedersen, D. L. Wang, J. Larsen and U. Kjems, "Two-microphone separation of speech mixtures," *IEEE Trans. Neural Networks*, vol. 19, pp. 475–492, Mar. 2008.

[13] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. OBrien, C. R. Lansing and A. S. Feng, "Blind estimation of reverberation time," *J. Acoust. Soc. Am.*, vol. 114, pp. 2877–2892, Nov. 2003.

[14] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. America*, vol. 37, pp. 409–412, 1965.

[15] W. Wang, S. Sanei and J. A. Chambers, "Penalty function-based joint diagnolization approach for convolutive blind separation of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 53, pp. 1654–1669, May 2005.