Reducing Musical Noise in Blind Source Separation by Time-Domain Sparse Filters and Split Bregman Method

Wenye Ma[1], *Meng Yu*[2], *Jack Xin*[2], *Stanley Osher*[1]

[1]Department of Mathematics, University of California, Los Angeles, USA
[2]Department of Mathematics, University of California, Irvine, USA

mawenye@math.ucla.edu, myu3@uci.edu, jxin@math.uci.edu, sjo@math.ucla.edu

Abstract

Musical noise often arises in the outputs of time-frequency binary mask based blind source separation approaches. Post-processing is desired to enhance the separation quality. An efficient musical noise reduction method by time-domain sparse filters is presented using convex optimization. The sparse filters are sought by l_1 regularization and the split Bregman method. The proposed musical noise reduction method is evaluated by both synthetic and room recorded speech and music data, and found to outperform existing musical noise reduction methods in terms of the objective and subjective measures.

Index Terms: Musical noise, time-frequency mask, time-domain sparse filters, split Bregman method.

1. Introduction

Blind source separation (BSS) is a major area of research in speech and music signal processing for recovering source signals from their mixtures without detailed knowledge of the mixing process. The time-frequency (TF) binary mask approaches to the BSS were widely studied ([1, 2] among others), which have the advantage in fast computation speed and handling the underdetermined problems where sources outnumber the microphones. The TF binary mask approaches rely on the sparseness of signals in the TF domain. It is assumed that speech signals are sufficiently sparse, and therefore at most one source signal is dominant at each TF point of their mixtures, i.e. the sources rarely overlap. These approaches extract source signals by applying binary masks to the observed mixtures. However, the nonlinear distortion (musical noise) exists in the outputs due to the winner-take-all property of the binary mask. It may cause too many discontinuous zero-paddings in the extracted signals, which often suffer from the musical noise. This is even worse for BSS of music sources and large number of sources since the assumption of sparseness is not quite satisfied.

In order to suppress the musical noise, a few methods were proposed recently [3, 4]. The main ingredients of these methods are: (1) employing the overlap-add method for reconstructing the waveform outputs from estimated spectrograms of source signals; (2) using a finer shift of Hanning window while taking short time Fourier transform (STFT); (3) adopting non-binary masks either based on a sigmoid function, where the mask of k-th source at TF point (f,τ) is defined by $\mathcal{M}_k(f,\tau) = 1/[1+\exp(g(d_k(f,\tau)-\theta_k))]$ (θ_k and g are parameters deciding the shape of the sigmoid function, $d_k(f,\tau)$ is the distance between cluster members and their centroids), or based on Bayesian inference by $\mathcal{M}_k(f,\tau) = P(C_k|\boldsymbol{X}(f,\tau))$ where C_k is the k-th cluster and $\boldsymbol{X}(f,\tau)$ are the spectrograms of mixtures. In brief, the noise reduction methods above were approached from either a gradual change of the spectrogram or non-binary masks.

In this paper, we propose a fast and efficient time domain method to suppress the musical noise in the output of TF mask based BSS. A convex optimization problem is formulated for seeking sparse filters to re-estimate the source signals in the time domain. The sparse filters are computed by l_1 norm regularization and the split Bregman method for which fast convergence was recently studied [6]. The paper is organized as follows. In section 2, we review the TF binary mask based BSS [1] and propose a way to adapt mask generation and minimize fuzzy points in the feature space for extending another TF binary mask based BSS [2] in the case that the microphone spacing exceeds the effective range [1, 2]. In section 3, an efficient musical noise suppression model is introduced based on a convex optimization problem with l_1 norm regularization. In section 4, computational framework by the split Bregman method is shown. In section 5, evaluations of the proposed method demonstrate its merits in comparison with existing methods. Even in the case of large and unknown microphone spacing, the proposed masking and musical noise reduction method enhances the recovered speech and music signals significantly. The concluding remarks are in section 6.

2. Initial Source Estimation

We briefly review the TF binary mask method DUET [1] which will be used as the initial separation. The standard mixing model for two receivers and multiple sources is $x_j(t) = \sum_{k=1}^{N} h_{jk} * s_k$, where $j = 1, 2$, $*$ is the convolution and h_{jk} represents the impulse response from source s_k to sensor j. The time-domain signals $x_j(t)$, $j = 1, 2$, sampled at frequency f_s are first converted into frequency-domain time-series signals $X_j(f,\tau)$ with STFT.

To group TF points into N clusters such that the points within each cluster are dominated by a single source signal, the feature parameters associated with each TF point are defined as $a(f,\tau) = |R(f,\tau)|$ and $\delta(f,\tau) = \frac{-1}{f}\angle R(f,\tau)$, where the ratio $R(f,\tau) = \frac{X_2(f,\tau)}{X_1(f,\tau)}$, $|\cdot|$ denotes the magnitude and $\angle\cdot$ denotes the phase angle of a complex number. Sufficient values of $a(f,\tau)$ and $\delta(f,\tau)$ generate a smooth two dimensional histogram. The K-means clustering algorithm finds the N most prominent peaks in the histogram. Each peak corresponds to one source in the mixture and the value for $a(f,\tau)$ and $\delta(f,\tau)$ at that peak are the feature parameters for that source. Once the

The authors M. Yu and W. Ma contributed equally to this work. M. Yu and J. Xin were partially supported by NSF DMS-0911277 and DMS-0712881; W. Ma and S. Osher were partially supported by NSF DMS-0914561, NIH G54RR021813 and the Department of Defense.

feature parameters for each source have been estimated, DUET assigns the energy in each TF point to the source whose peak lies closest to that point in the feature space of a and δ. The individual separated signal spectrogram $Y_k(f,\tau)$ is estimated based on the clustering result. TF binary mask for the k-th source signal is:

$$\mathcal{M}_k(f,\tau) = \begin{cases} 1 & (f,\tau) \in \text{cluster } C_k \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Then $Y_k(f,\tau) = \mathcal{M}_k(f,\tau)X_J(f,\tau)$, where $k = 1, ..., N$ and J is a selected sensor index. Finally, inverse STFT (iSTFT) is applied to $Y_k(f,\tau)$ with overlap-add method [4] to recover the waveform $y_k(t)$.

Accurate estimation of the feature parameters is critical to a successful source separation. Single source dominance at each TF point may not be valid with the increase of source number N and reverberation time (convolution length). In order to alleviate clustering error, a stricter criterion is introduced below. At each TF point (f,τ), the confidence coefficient of $(f,\tau) \in C_k$ is defined by $CC(f,\tau) = \frac{d_k}{\min_{j \neq k} d_j}$, where d_j is the distance between the value of a and δ at (f,τ) and that at j-th peak. The mask is redefined for some $\rho > 0$ as

$$\mathcal{M}_k(f,\tau) = \begin{cases} 1 & (f,\tau) \in C_k \;\&\; CC(f,\tau) \leq \rho \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The motivation for the refined mask is to eliminate the fuzzy feature points which have nearly equal distances to at least two cluster centers. The refined mask also applies to the situation where an unknown receiver spacing is not small enough and phase aliasing errors appear [1, 2]. In this case, based on another TF binary mask BSS method [2], we modify the feature $\mathbf{\Theta}(f,\tau)$ (defined in [2]) by dropping the directions of arrival (DOA) part yet keeping the distance part so that $\mathbf{\Theta}(f,\tau) = \left[\frac{|X_1(f,\tau)|}{|X(f,\tau)|}, ..., \frac{|X_M(f,\tau)|}{|X(f,\tau)|}\right]$, where $|X(f,\tau)|$ is a normalization and M is the number of receivers (sensors). This feature theoretically works for clustering if the number of sources N is not large and the ratio of source to microphones' distances varies from one source to another. However, the quality of recovered signals may not be good because in practice the distance feature does not distinguish the sources well [2]. Then the mask (2) helps to improve the separation quality and set a better stage for the subsequent time-domain noise reduction and quality enhancement of recovered source signals.

3. Time Domain Noise Reduction

Let us first consider the determined case of mixing model with 2 sensors and 2 sources. The output of the TF domain mask based BSS are $y_k(t)$, $k = 1, 2$. We seek a pair of filters u_{jk}, $j = 1, 2$, for each source k such that

$$u_{1k} * x_1 - u_{2k} * x_2 \approx y_k, \quad k = 1, 2. \tag{3}$$

In general, BSS output y_k's may differ from s_k's by a convolution [7]. Cross multiplication and subtraction of the two equations in the mixing model implies a family of solutions to (3) of the form: $u_{1k} = g_k * h_{2\neg k}$ and $u_{2k} = g_k * h_{1\neg k}$, where g_k is an un-determined filter, and $\neg k$ denotes complementary index of k, e.g. if $k = 1$, $\neg k = 2$. The solutions u_{jk} may differ from the room impulse responses (RIRs) or the $h_{1\neg k}$ and $h_{2\neg k}$, by a convolution g_k. The optimal choice of g_k is the de-reverberation filter which minimizes the length (support) of $g_k * h_{1\neg k}$ and $g_k * h_{2\neg k}$. Without knowledge of RIRs however, we shall use l_1 norm regularization of u_{1k} and u_{2k} to achieve this goal indirectly as follows.

Let us consider a duration D of $y_k(t)$, and seek a pair of sparse filters u_{jk}, $j = 1, 2$ to minimize the energy (l_2 norm) of $u_{1k} * x_1 - u_{2k} * x_2 - y_k$ subject to l_1-norm regularization. The resulting convex optimization problem for $t \in D$ is:

$$\begin{aligned}(u_{1k}^*, u_{2k}^*) = \arg \min_{(u_{1k}, u_{2k})} & \frac{1}{2}||u_{1k} * x_1 - u_{2k} * x_2 - y_k||_2^2 \\ & + \mu(||u_{1k}||_1 + ||u_{2k}||_1).\end{aligned} \tag{4}$$

Denote the length of signal in D as L_D and the length of solution as L. In practice, D can be as short as several seconds, which makes the proposed method efficient on data usage. Since u_{jk}'s are l_1-regularized, we essentially recover minimal-length solutions of (4). In matrix form, the convex objective (4) becomes:

$$u_k^* = \arg \min_{u_k} \frac{1}{2}||Au_k - y_k||_2^2 + \mu||u_k||_1 \tag{5}$$

where u_k is formed by stacking up u_{1k} and u_{2k}, and $L_D \times 2L$ matrix A is (T is transpose):

$$A = \begin{pmatrix} x_1(1) & x_1(2) & \ldots & \ldots & x_1(L_D{-}1) & x_1(L_D) \\ & x_1(1) & \ldots & \ldots & x_1(L_D{-}2) & x_1(L_D{-}1) \\ & & \ddots & & & \vdots \\ & & & x_1(1) & \ldots & x_1(L_D{-}L{+}1) \\ -x_2(1) & -x_2(2) & \ldots & \ldots & -x_2(L_D{-}1) & -x_2(L_D) \\ & -x_2(1) & \ldots & \ldots & -x_2(L_D{-}2) & -x_2(L_D{-}1) \\ & & \ddots & & & \vdots \\ & & & -x_2(1) & \ldots & -x_2(L_D{-}L{+}1) \end{pmatrix}^T$$

Once u_{1k} and u_{2k} are found, the cross multiplication and subtraction $u_{1k}^* * x_1 - u_{2k}^* * x_2$ is a better approximation of s_k for human ear with muscial noise reduced. If the acoustic environment does not change much, the estimation during $t \in D$ still applies when $t \notin D$. Otherwise, an adaptive estimation can be repeated at a suitable time interval later. The objective (4) takes the same form as that in image denoising [6].

The above derivation generalizes to M sensors and N sources ($M \geq 3$ and $N = M$) case. When $t \in D$, then for proper value of $\mu > 0$, we minimize:

$$\frac{1}{2}||\sum_{j=1}^{M} u_{jk} * x_j - y_k||_2^2 + \mu \sum_{j=1}^{M} ||u_{jk}||_1,$$

and estimate s_k by $\hat{s}_k = \sum\limits_{j=1}^{M} u_{jk} * x_j$. Though 2 sensors are enough for DUET, the remaining $M-2$ sensors are also used here for reducing the musical noise.

4. Split Bregman Method

The split Bregman method was introduced and analyzed in [6] as an efficient tool for solving optimization problems arising from l_1 regularization based models. It aims to solve the unconstrained problem: $\min_u J(\Phi u) + H(u)$, where J is convex but not necessarily differentiable such as the l_1 norm, H is convex and differentiable, and Φ is linear operator. The key idea of the split Bregman method is to introduce an auxiliary variable $d = \Phi u$, and try to solve the constrained problem:

$$\min_{d,u} J(d) + H(u), \text{ s.t. } d = \Phi u$$

In [5, 6], it is proved that this kind problem can be solved by the following iterations:

$$\begin{aligned}(u^{n+1}, d^{n+1}) &= \arg\min_{u,d} J(d) + H(u) - \langle p_d^n, d - d^n \rangle \\ &\quad - \langle p_u^n, u - u^n \rangle + \frac{\lambda}{2}\|d - \Phi u\|_2^2 \\ p_d^{n+1} &= p_d^n - \lambda(d^{n+1} - \Phi u^{n+1}) \\ p_u^{n+1} &= p_u^n - \lambda \Phi^T(\Phi u^{n+1} - d^{n+1})\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the inner product. For simplicity, we introduce a new variable $b^n = p_d^n/\lambda$, and notice that $p_d^n = \lambda b^n$ and $p_u^n = -\lambda \Phi^T b^n$. The iterates d^{n+1} and u^{n+1} can be updated alternatively. The general split Bregman iteration scheme is:

$$\begin{aligned}d^{n+1} &= \arg\min_{d} \frac{1}{\lambda} J(d) - \langle b^n, d - d^n \rangle + \frac{1}{2}\|d - \Phi u^n\|_2^2 \\ u^{n+1} &= \arg\min_{u} \frac{1}{\lambda} H(u) + \langle b^n, \Phi(u - u^n) \rangle \\ &\quad + \frac{1}{2}\|d^{n+1} - \Phi u\|_2^2 \\ b^{n+1} &= b^n - (d^{n+1} - \Phi u^{n+1})\end{aligned}$$

In the case of (5), $J(u_k) = \mu \|u_k\|_1$, $\Phi = I$, and $H(u_k) = \frac{1}{2}\|Au_k - y_k\|_2^2$. Then the iterations are:

$$\begin{aligned}d^{n+1} &= \arg\min_{d} \frac{\mu}{\lambda}\|d\|_1 - \langle b^n, d - d^n \rangle + \frac{1}{2}\|d - u_k^n\|_2^2 \quad (6) \\ u_k^{n+1} &= \arg\min_{u_k} \frac{1}{2\lambda}\|Au_k - y_k\|_2^2 + \langle b^n, u_k - u_k^n \rangle \\ &\quad + \frac{1}{2}\|d^{n+1} - u_k\|_2^2 \quad (7) \\ b^{n+1} &= b^n - (d^{n+1} - u_k^{n+1}) \quad (8)\end{aligned}$$

Explicitly solving (6) and (7) gives the simple algorithm

```
Initialize $u_k^0 = d^0 = b^0 = 0$
While $\|u_k^{n+1} - u_k^n\|_2 / \|u_k^{n+1}\|_2 > \epsilon$
  (i) $d^{n+1} = \text{shrink}(u_k^n + b^n, \frac{\mu}{\lambda})$
  (ii) $u_k^{n+1} = (\lambda I + A^T A)^{-1}(A^T y_k + \lambda(d^{n+1} - b^n))$
  (iii) $b^{n+1} = b^n - d^{n+1} + u_k^{n+1}$
end While
```

Here shrink is the soft threshold function defined by $\text{shrink}(v, t) = (\tau_t(v_1), \tau_t(v_2), \cdots, \tau_t(v_{NL}))$ with $\tau_t(x) = \text{sign}(x)\max\{|x| - t, 0\}$. Noting that the matrix A is fixed, we can precalculate $(\lambda I + A^T A)^{-1}$, then the iterations only involve matrix multiplication and are fast as a result. Since the size of matrix $\lambda I + A^T A$ is $NL \times NL$, where N is the number of sources, the computational cost for matrix inverse is not high. The above algorithm runs fast for the purpose of the proposed musical noise suppression model, averagely within 5 seconds' processing for each source channel. The entire algorithm is:

Input: Acoustic mixing signals, $x_j, j = 1, ..., M \geq 2$
Output: Estimated sources with musical noise
suppressed, $\hat{s}_k, k = 1, ..., N$ $(N = M)$
Initial separation: Extract signals $y_k, k = 1, ..., N$ by TF mask approaches with a proper ρ
Fitler estimation: Apply the **split Bregman** method to obtain the filters $u_{jk}, j = 1, ..., M$ for each source k
Musical Noise Suppression: $\hat{s}_k = \sum_{j=1}^{M} u_{jk} * x_j$

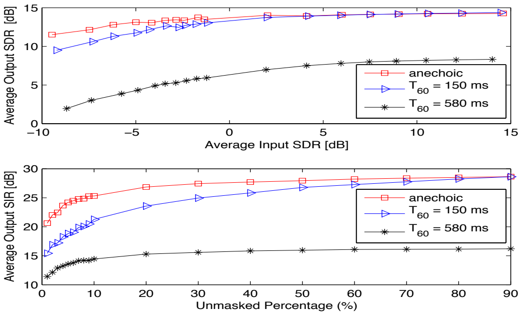

Figure 1: *Upper panel: average signal to distortion (SDR) ratios of input and output signals in synthetic test. Lower panel: signal to interference ratio (SIR) vs. unmasked percentage (percentage of 1's) in the mask. The data points in the upper panel have the same unmasked percentage as those in the lower panel.*

5. Evaluation and Comparison

The parameters for the proposed method are chosen as $\mu = \epsilon = 10^{-3}$, $\eta = 1$, $\lambda = 2\mu$, D contains 4 seconds' duration, and $L = 1000$ taps. As suggested in [4], the STFT frame size is 512 and frame shift is 512/8. For simplicity, we denote DUET [1] by BM1 and the extension of binary mask BSS [2] with redefined feature $\boldsymbol{\Theta}(f, \tau)$ in section 2 by BM2.

To test the musical noise reduction portion of our method, synthetic mixture data are used to recover a spectrogram-masked source signal where energy loss due to binary mask is simulated. The masked signal plays the role of BSS output y_k in section 3. Measured binaural RIRs (h_{jk}, $j, k = 1, 2$) are used to generate mixtures x_1 and x_2. For the spectrogram $S_{11} = STFT(h_{11} * s_1)$ of $h_{11} * s_1$, a mask $\mathcal{M}$ of the same size as S_{11} is defined to contain a certain percentage of entries equal to 1 and the rest equal to 0. The mask is entry-wise multiplied to S_{11} to produce a distorted waveform signal $s_d = iSTFT(S_{11} \circ \mathcal{M})$. We recover $h_{11} * s_1$ from the two mixture signals x_1, x_2 and s_d (in place of y_1) with the Bregman iterations in section 4. The test is repeated under different reverberation times (anechoic, 150 ms, 580 ms). Though a little interference from s_2 is introduced, i.e. a little decrease of signal to interference ratio (SIR), the gain in signal to distortion ratio (SDR) is found to be significant in low input SDR regime (Fig. 1). This phenomenon is observed in processing room recorded data as well.

Comparison of several musical noise suppression methods is carried out on room recorded data. The set-up is shown in Fig. 2. In case of 2 sources, their locations are at S_1 and S_2 in Fig. 2, and the sensors Mic_1 and Mic_2 provide data for separation and noise suppression. In case of 4 sources, all the loudspeakers and microphones contribute to the musical noise reduction but only Mic_2 and Mic_3 are used for separation. Table 1 lists results of different musical noise suppression methods discussed in section 1. Compared with BM1, sigmoid mask and Bayesian mask methods, our method leads in the overall quality PESQ [8], and with a significant margin in SDR [dB] and SAR [dB] (signal to artifact ratio). The SIR improvement is however not uniformly better. In case of 4 sources, SIR improvement lags the other methods. When the number of sources increases, ρ in the mask (2) should increase accordingly to control the growth of zero-paddings.

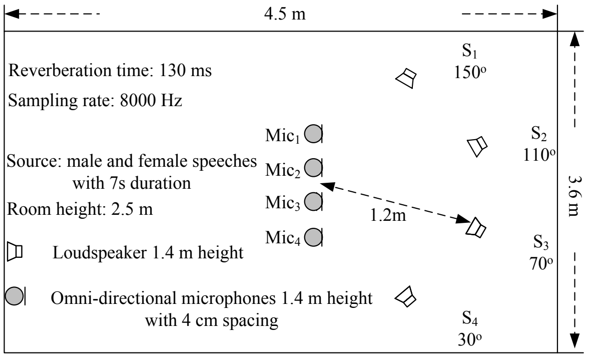

Figure 2: *Configuration and parameters of the room recording.*

Next we remove Mic_1 and Mic_4 from the set-up of Fig. 2, so only 2 microphones Mic_2 and Mic_3 are active. The unknown microphone spacing is in $[15, 20]$ cm outside the effective range of binary mask BSS methods [1, 2]. Two sources (one speech and one music, 8000Hz and 5s' duration) are with azimuth 0° and 60°. We use BM2 and the refined mask (2) with a nearly optimal value of $\rho = 0.5$ as the initial separation for our method. Discussed in section 2, since BM2 may not work well, eliminating the fuzzy feature points by a proper value of ρ helps to gain a good SIR but sacrifice the signal quality. However, as seen in Table 2, the overall quality is improved significantly by both Bayesian mask and our method without losing SIR.

Furthermore, we conduct a subjective test on five listeners with normal hearing to evaluate the reduction of musical noise. The paired comparison test requires each listener to rank the four methods according to the performance of musical noise reduction in the experiments conducted in Tables 1 and 2. The percentage of our method's superiority over other three methods in musical noise reduction is shown in Table 3. Since the initial estimated music sources contain more musical noise, the contrasts between these methods on the music channel are more pronounced.

Table 1: *Comparison of musical noise reduction methods on room recorded speech data. Average evaluation results are shown as 2 sources/4 sources. BM1 with conventional mask (1); SM (Sigmoid mask); BYM (Bayesian mask). The initial separation for our method employs BM1 with refined mask (2), where* $\rho = 0.50, 0.25, 0.10, 0.05$

Method	PESQ	SIR	SDR	SAR
Input	1.37/1.10	0.04/-4.49	0.02/-4.51	46.48/26.54
BM1[1]	2.24/1.89	13.24/9.39	6.44/3.79	9.37/6.57
SM[3]	2.17/1.71	11.38/8.22	6.52/2.21	9.14/5.40
BYM[3]	2.33/1.83	13.30/8.21	7.20/3.34	10.20/6.65
Our-0.50	2.18/1.90	9.47/6.30	8.58/5.85	17.74/19.06
Our-0.25	2.21/**1.91**	10.07/**6.35**	9.26/**5.89**	17.97/**18.93**
Our-0.10	2.22/1.84	10.18/5.63	9.51/5.23	18.94/18.76
Our-0.05	**2.40**/1.75	**13.41**/5.36	**12.18**/4.79	**19.05**/16.86

6. Conclusions

We proposed and evaluated an efficient time domain method for reducing musical noise in the output of TF mask based BSS methods. By a more selective TF mask, we reduced percentage of fuzzy points on TF domain to improve separation quality. We employed fast Bregman iterations to compute sparse time-domain filters while minimizing a convex l_1 norm regularized objective. Both synthetic and recorded data showed that the filters reduced musical noise and enhanced the overall quality in music and speech sources effectively in terms of objective and subjective measures.

Table 2: *Comparison of musical noise reduction methods on speech/music mixtures with unknown large microphone spacing. Refined mask (2) with* $\rho = 0.5$ *is employed.*

Method	PESQ	SIR	SDR	SAR
Input	1.50	1.90	1.85	33.16
BM2	1.63	16.87	3.58	4.01
Sigmoid mask [3]	2.07	22.10	8.86	9.10
Bayesian mask[3]	2.52	16.54	11.66	14.50
Ours	2.45	16.52	12.81	16.32

Table 3: *Subjective evaluation on musical noise reduction.* $>$ *(*$<$*) means the output of our method is perceived with less (more) musical noise, while* $\approx$ *means "hard to distinguish". Binary Mask is BM1 (BM2) for Table 1 (2).*

Method	Test Category		$>$	$\approx$	$<$
Ours vs Binary mask	Table 1	2 sources	80%	20%	-
		4 sources	75%	5%	20%
	Table 2	Speech	98%	2%	-
		Music	99%	-	1%
Ours vs Sigmoid mask[3]	Table 1	2 sources	70%	20%	10%
		4 sources	70%	15%	15%
	Table 2	Speech	33%	57%	10%
		Music	94%	-	6%
Ours vs Bayesian mask[3]	Table 1	2 sources	50%	20%	30%
		4 sources	55%	30%	15%
	Table 2	Speech	15%	50%	35%
		Music	89%	-	11%

7. References

[1] O. Yilmaz and S.Rickard, Blind separation of speech mixtures via time-frequency masking, IEEE Trans. Signal Processing, vol. 52, no. 7, pp. 18301847, July 2004.

[2] S. Araki, H. Sawada, R. Mukai and S. Makino,"Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors." Signal Processing, 87, 1833-1847, 2007.

[3] S. Araki, H. Sawada, R. Mukai, and S. Makino, Blind sparse source separation with spatially smoothed time-frequency masking, in IWAENC, Paris, France, 2006.

[4] S. Araki, S. Makino, H. Sawada, and R. Mukai, Reducing musical noise by a fine-shift overlap-add method applied to source separation using a time-frequency mask, in Proc. ICASSP2005, Mar. 2005, vol. III, pp. 8184.

[5] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for compressed sensing and related problems. SIAM J. Imaging Sciences 1(1):143-168, 2008.

[6] T. Goldstein and S. Osher, The Split Bregman Algorithm for L1 Regularized Problems, SIAM J. Imaging Sci. 2:323-343, 2009.

[7] J. Liu, J. Xin, Y. Qi, F-G Zeng, "A Time Domain Algorithm for Blind Separation of Convolutive Sound Mixtures and l_1 Constrained Minimization of Cross Correlations", Comm Math Sciences, vol. 7, No.1, pp 109-128, 2009.

[8] ITU-T Rec. P. 862, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, International Telecommunication Union, Geneva, 2001.