DYNAMIC RELATIVE IMPULSE RESPONSE ESTIMATION USING STRUCTURED SPARSE BAYESIAN LEARNING

Ritwik Giri, Bhaskar D. Rao

University of California, San Diego
{rgiri, brao}@ucsd.edu

Fred Mustiere, Tao Zhang

Starkey Hearing Technologies
{Fred_Mustiere, Tao_Zhang}@starkey.com

ABSTRACT

In this paper we present a novel Hierarchical Bayesian approach to estimate Relative Impulse Response (ReIR) using short, noisy and reverberant microphone recordings. The information contained in ReIRs between two microphones is useful for a wide range of multichannel speech processing applications such as speaker localization, speech enhancement, etc. It has been shown in several previous works that the Relative Transfer Function (RTF) corresponding to a given ReIR is dynamic and depends on the environment, microphone positions and target position. This acts as the main motivation of this work, as we develop a structured sparse Bayesian learning algorithm to estimate ReIR using very short recordings, which will be robust to changes in the environment. An extensive experimental study with real-world recordings has also been conducted to show the efficacy of our proposed approach over other competing approaches.

***Index Terms*—** Relative Transfer Function, Relative Impulse Response, Sparse Bayesian Learning, Reverberation.

1. INTRODUCTION

Relative Impulse Responses (ReIR) or their frequency-domain counterparts, the Relative Transfer Functions (RTF) [1] are important tools in several multichannel audio processing tasks such as speaker extraction, noise reduction, speech enhancement, source localization etc [2, 3]. For instance, RTF information can be naturally incorporated in beamforming algorithms, where the RTF is used to design the blocking matrix of an adaptive Generalized Sidelobe Canceler (GSC) [2, 4] to cancel the target signal and produce a noise reference signal. This noise reference signal is then used later for adaptive interference cancellation and post filtering to improve the speech enhancement performance. In this paper we will focus on a two-microphone setup, i.e, two channel recordings; in particular, we consider a hearing aid application where the microphones are located on left and right devices worn on a head. In other words, we aim to estimate the ReIR between these two microphones.

As discussed above, in realistic acoustic environment, reverberation has to be taken in to account in GSC to achieve satisfactory signal cancellation in the output of the blocking matrix. Following this idea, Gannot et al proposed a variant of GSC named as Transfer Function-GSC (TF-GSC) [1] that relies on estimated RTFs. The performance of TF-GSC depends strongly on the quality of the RTF estimate, which is dynamic and changes with the movements of target, head movements of the hearing aid user (i.e, movement of microphones) etc. If the RTF estimate is not updated fast enough, or if it is inaccurate, the target signal leaks through the blocking matrix and is canceled by the adaptive filtering stage, which can cause severe signal distortion at the output of GSC. This motivates us to consider a dynamic environment where the RTF is estimated every $t = 100 - 200$ ms using only recordings of duration t, which will enable us to capture any changes in the environment.

ReIRs can be easily computed in a noiseless environment using a traditional Least Squares (LS) method as shown in [4], but the LS estimate becomes unstable in presence of noise (this will be discussed below in more details). There have been many recent attempts to estimate RTFs accurately in a noisy environment [5, 6, 7, 8], but most of these solutions require a sufficiently long recording for a good estimate of RTF (i.e., significantly more than $100-200$ ms). In [1] the authors have proposed a method that exploits the non-stationarity of the target speech signal. This method assumes that the noise and the RTF are stationary, or at least much less dynamic, when compared to the target signal. However, this assumption does not hold when there is a speech interferer or if the RTF is highly non stationary. In [9] the authors propose a novel assumption that the ReIRs can be replaced by sparse filters, which regularizes the LS solution. However, in reverberant environments ReIRs will also exhibit a non-sparse decaying tail [8], which makes this approach detrimental in highly reverberant conditions. Moreover, they do not consider noisy cases. In [8] a novel approach of sparsely reconstructing time domain ReIRs from incomplete RTF measurements is proposed, where the estimation occurs only using high Signal-to-Noise Ratio (SNR) frequency bins.

Since we are considering a dynamic environment RTF estimation, existing frequency domain approaches (described in more details in next section) give a biased estimate because of the inaccuracy of the power spectral density estimate which must be approximated by a finite average [10]. This is the main motivation of focusing on a time domain solution. However, conversely a traditional time domain LS approach produces ineffective and unstable estimates due to the presence of noise and finite amount of samples in the deconvolution problem [11]. To circumvent this, we propose a regularized LS approach where the regularization has been incorporated by exploiting a model for the prior structure of a ReIR. Specifically, unified treatment of sparse early reflection and exponentially decaying reverberation tail in a prior distribution using a hierarchical Bayesian framework is the main novelty of our work.

The rest of the paper is organized as follows: In Section 2 we introduce the problem and Section 3 presents the popular existing solutions to that problem which will be used as our baseline. We present our proposed solution along with the inference procedure in Section 4. Extensive experimental results over real world recordings are presented in Section 5 and finally Section 6 concludes the paper and discusses some future directions of this work.

2. PROBLEM FORMULATION

Consider a two channel noisy recording of a target in a diffuse noise environment, whose position is fixed for a certain time interval. This situation can be represented as:

$$x_L[n] = (h_L \star s)[n] + \epsilon_L[n] \tag{1}$$

$$x_R[n] = (h_R \star s)[n] + \epsilon_R[n] \approx (h_{rel} \star x_L)[n] + \epsilon_R[n] \tag{2}$$

Where h_L and h_R denote the impulse response between the target and the two microphones, $s[n]$ denotes the target speech, $\epsilon_L[n]$ and $\epsilon_R[n]$ denote the noise components. The main problem is to estimate h_{rel}, which denotes the ReIR between the left and right microphone. The oracle solution of this problem in the time domain is, $h_{rel} = h_R \star h_L^{-1}$. To ensure that the solution is causal, a fixed delay of a few milliseconds can be introduced [12, 13], i.e., $h_{rel} = h_R \star h_L^{-1} \star \delta(n-d)$ where d is the delay in samples. The oracle RTF, denoted as H_{RTF} which is a Fourier Transform of h_{rel} can also be written as, $H_{RTF}(\theta) = \frac{H_R(\theta)}{H_L(\theta)}$.

3. EXISTING SOLUTIONS

In this section we recapitulate some existing approaches to solve the problem presented above, which will be included in our experimental results section for comparison as baseline.

3.1. Traditional Least Square Solution

In a noise-free condition the size-L ReIR vector $\mathbf{h}$ can be estimated using a Least Square approach, i.e:

$$\hat{\mathbf{h}}_{LS} = \arg\min_{\mathbf{h}} \|\mathbf{x_R} - \mathbf{X_L}\mathbf{h}\|_2^2 \tag{3}$$

where $\mathbf{X_L}$ is the convolution matrix of size $N \times L$ which has been constructed using $x_L[n]$. The solution of Equation (3) can be easily found by taking the pseudo-inverse:

$$\hat{\mathbf{h}}_{LS} = (\mathbf{X_L}^T\mathbf{X_L})^{-1}\mathbf{X_L}^T\mathbf{x_R} \tag{4}$$

Unfortunately, in the presence of noise the LS solution becomes unstable, which gives rise to a fluctuating ReIR estimate.

A workaround to the ill-conditioning problem above is to use diagonal loading to make the matrix $\mathbf{X_L}^T\mathbf{X_L}$ well conditioned. The solution then becomes:

$$\hat{\mathbf{h}}_{RLS} = (\mathbf{X_L}^T\mathbf{X_L} + \alpha I)^{-1}\mathbf{X_L}^T\mathbf{x_R} \tag{5}$$

We can show that $\hat{\mathbf{h}}_{RLS}$ is actually the solution of the following optimization problem:

$$\hat{\mathbf{h}}_{RLS} = \arg\min_{\mathbf{h}} \|\mathbf{x_R} - \mathbf{X_L}\mathbf{h}\|_2^2 + \alpha\|\mathbf{h}\|_2^2 \tag{6}$$

Subscript RLS denotes Regularized Least Square, which is essentially a ridge regression framework [14]. We will use the RLS method as one of our baseline methods with $\alpha = \frac{0.1}{L} \times$ trace$(\mathbf{X_L}^T\mathbf{X_L})$. (Heuristic Choice)

3.2. Frequency Domain Estimation (FD)

In the Short-Time Fourier Transform (STFT) domain, assuming noiseless recordings we can rewrite Equation (2) as:

$$X_R(\theta, k) = H_{RTF}(\theta)X_L(\theta, k) \tag{7}$$

Where θ denotes the frequency bin and k denotes the frame index. A straightforward estimate of the RTF can be found using:

$$\hat{H}_{RTF}(\theta) = \frac{\sum_k X_L^\star(\theta, k)X_R(\theta, k)}{\sum_k |X_L(\theta, k)|^2} \tag{8}$$

The numerator is a sample estimate of the cross Power-Spectral Density (PSD), and the denominator is a sample estimate of the auto PSD. As discussed in [1] this method produces a biased estimate. In future discussions we will refer to this method by FD and include it in our comparative experiments as another baseline method.

3.3. Non-Stationarity based FD Estimation (NSFD) [1]

This method depends on the assumption that the noise signals are stationary, or at least "less dynamic" when compared to the target speech signal. Again in the STFT domain we can represent the model as:

$$X_R(\theta, k) = H_{RTF}(\theta)X_L(\theta, k) + E(\theta, k) \tag{9}$$

Where E denotes the environmental noise. If we consider that H_{RTF} is static for a specific interval and divide that interval into P frames, then for the p^{th} frame:

$$\Phi^p_{X_RX_L}(\theta) = H_{RTF}(\theta)\Phi^p_{X_LX_L}(\theta) + \Phi^p_{EX_L}(\theta) \tag{10}$$

Where, $\Phi^p_{AB}(\theta)$ denotes the (cross) power spectral density between A and B during the pth frame. Since the noise is stationary, we can write $\Phi^p_{EX_L} = \Phi_{EX_L}$ and solve the overdetermined set of equations for $p = 1...P$, to estimate H_{RTF}. As in the FD case, in practice the PSDs in the above set of equations are replaced by their sample based estimates.

4. STRUCTURED SPARSE BAYESIAN LEARNING (S-SBL)

In presence of noise, in order to make the LS solution stable we will use a novel regularization strategy by incorporating the structure information of ReIRs as a prior in a Bayesian framework. The main difference of our work from [9] is that we consider both the sparse early reflections and the reverberation tail in a unified framework. Moreover, we do not need any *a priori* knowledge of SNR since the noise variance is also estimated within the proposed framework.

4.1. Model

Consider the following model, $\mathbf{x_R} = \mathbf{X_L}\mathbf{h} + \epsilon$, along with the Gaussian Likelihood assumption i.e, $p(\mathbf{x_R}|\mathbf{h}) \sim N(\mathbf{X_L}\mathbf{h}, \sigma^2)$.

The prior distribution over $\mathbf{h}$ is proposed to follow:

$$p(\mathbf{h}|\gamma_i, c_1, c_2) \sim N(0, \Gamma) \tag{11}$$

With:

$$\Gamma = diag\big[\gamma_1, ..., \gamma_P, c_1e^{-c_2}, ..., c_1e^{-c_2m}, ..., c_1e^{-c_2M}\big] \tag{12}$$

Where:

- γ_p corresponds to p^{th} early reflection
- $c_1e^{-c_2m}$ corresponds to m^{th} tap out of the M exponentially decaying reverberation tail components

Note that the proposed approach follows a Relevance Vector Machine (RVM)/Sparse Bayesian Learning (SBL) [15] framework to incorporate the sparse regularization. In this variant of SBL we have also incorporated the reverberation tail regularization by tying the last M diagonal elements of Γ in an exponentially decaying tail.

4.2. Bayesian Inference

We will follow a Type II likelihood/Evidence maximization [16, 17] procedure to estimate the ReIR. For estimating $\mathbf{h}$ we will compute the posterior:

$$p(\mathbf{h}|\mathbf{x_r};\gamma,c_1,c_2) = N(\mathbf{h};\mu,\Sigma) \tag{13}$$

Where

$$\mu = \sigma^{-2}\Sigma\mathbf{X}_L^T\mathbf{x_R} \tag{14}$$

$$\Sigma = (\sigma^{-2}\mathbf{X}_L^T\mathbf{X}_L + \Gamma^{-1})^{-1} \tag{15}$$

Hence, we approximate the true posterior by a Gaussian distribution whose mean and covariance depends on the estimated hyperparameters. We can use $\hat{\mathbf{h}} = \mu$ as the point estimate of the relative impulse response.

For the estimation of the hyperparameters we will use an evidence maximization approach, i.e:

$$\hat{\Gamma},\hat{c_1},\hat{c_2} = \arg\max p(\mathbf{x_R}|\gamma_i,c_1,c_2) \tag{16}$$

We employ the Expectation-Maximization (EM) algorithm to solve the above optimization because of its monotonic convergence property.

To estimate the previously discussed hyperparameters we treat the ReIR $\mathbf{h}$ as a hidden variable. In the E step, for iteration t we only need to compute the following conditional expectation for all taps $i \in \{1,\ldots,P+M\}$:

$$< h_i^2 >= E_{\mathbf{h}|\mathbf{x_R};\gamma^t,c_1^t,c_2^t,\sigma^2}[h_i^2] = \Sigma_{(i,i)} + \mu_i^2 \tag{17}$$

where $\Sigma_{(i,i)}$ is the i^{th} diagonal element of Σ. We use this E step to compute the Q-function:

$$Q(\gamma,c_1,c_2,\sigma^2) = E_{\mathbf{h}|\mathbf{x_R};\gamma^t,c_1^t,c_2^t,\sigma^2}[\log(p(\mathbf{x_R}|\mathbf{h};\sigma^2)p(\mathbf{h}|\gamma,c_1,c_2))] \tag{18}$$

In the M step, maximizing this Q-function with respect to the hyperparameters i.e, γ, c_1, c_2 and σ^2, we get:

$$\gamma_p = \Sigma_{(p,p)} + \mu_p^2 \quad \text{for } p = 1\ldots P \tag{19}$$

$$c_1 = \frac{1}{M}\sum_{m=1}^{M} e^{c_2 m} < h_{m+P}^2 > \tag{20}$$

$$\sum_{m=1}^{M} m e^{c_2 m} < h_{m+P}^2 > -c_1\frac{M(M+1)}{2} = 0 \tag{21}$$

$$\sigma^2 = \frac{\|\mathbf{x_R} - \mathbf{X_L}\mathbf{h}\|^2}{N-(M+P)+\sum_{i=1}^{M+P}\Sigma_{(i,i)}/\Gamma_i} \tag{22}$$

In Equation (20) we will use the estimate of c_2 from the previous iteration. We also need to solve Equation (21) to get the closed form update rule of c_2. Representing it as a polynomial of $v = e^{c_2}$, we can show using Descartes' sign rule that there is only one positive root $\hat{v}$ of (21). Therefore we can update c_2 using $c_2 = \log\hat{v}$. Hence, every iteration we will update all the hyperparameters using the update rules shown above, and we can compute the point estimate $\hat{h}$ substituting the updated hyperparameters in Equation (14). In the following iteration we will start with the updated μ and Σ, and recompute all the hyperparameters. In practice, 10 to 15 iterations of the above S-SBL procedure yields a converged relative impulse response estimate $\mathbf{h}$.

Before moving on to experimental validation, in the next subsection we show the connection between S-SBL and the RLS methodology.

4.3. Connection between S-SBL and RLS

Simplifying Equation (14) we get,

$$\mu = (\mathbf{X}_L^T\mathbf{X}_L + \sigma^2\Gamma^{-1})^{-1}\mathbf{X}_L^T\mathbf{x_R} \tag{23}$$

Comparing this with Solution (5) we see that S-SBL can be viewed as an iterative reweighted ridge regression/reweighted ℓ_2 norm minimization algorithm, where the penalty weight factor α is not the same for all taps, and where the penalty weights are estimated every iteration through γ_i, c_1, c_2 and σ^2 which enforces the desired ReIR structure through regularization. A similar connection between SBL and reweighted ℓ_2 minimization approach can be found in [18].

5. EXPERIMENTAL VALIDATION

In this section we present the detailed experimental results to evaluate several competing algorithms in term of their target signal blocking ability.

5.1. Experimental Settings

We follow the experimental setting described in [8] and use the publicly available database of measured impulse responses [19] to generate the reverberant recordings. The signal for the target source has been taken from the task of the online Signal Separation Campaign (SISEC) 2013 [20]. All other details are summarized below in Table 1.

Table 1. Experimental Settings

Parameters	Values
Sampling Frequency	8 kHz
SNR_{in}	0 dB
Target Angle	0°
Directional Noise Angle	-60°
Microphone Pair	[3 4] (3 cm)
Distance between source and mic	2 m
T_{60}	360 ms

The testing utterance (female talker) is 10 s long, which we divide into intervals of 1024 samples, i.e., 128 ms at 8 kHz (Total 78 segments). Experiments are conducted on each interval independently. The average Attenuation Rate (described in the next subsection) is been reported over the intervals where speech is present. For all our experiments we use $P = 30$ for S-SBL, although we have found out that our algorithm is not very sensitive to different choices for P.

5.2. Performance Metric

To quantitatively evaluate the competing algorithms, we use a well-known and widely used performance metric called the Attenuation Rate.

The Attenuation Rate (ATR) can be evaluated as the ratio between SNR_{out} and SNR_{in} in dB scale, where:

$$SNR_{in} = \frac{\sum_{i=L,R}\sum_n[(h_i \star s)(n)]^2}{\sum_{i=L,R}\sum_n[\epsilon_i(n)]^2} \tag{24}$$

and,

$$SNR_{out} = \frac{\sum_n[(\hat{h}_{rel} \star s_L)(n) - s_R(n)]^2}{\sum_n[(\hat{h}_{rel} \star \epsilon_L)(n) - \epsilon_R(n)]^2} \tag{25}$$

The numerator of SNR_{out} measures the leakage of the target signal whereas the denominator measures the attenuation of the noise signal. Overall, the more negative the value of ATR is, the better is the blocking performance. A low ATR indicates a good noise reference signal for further processing (such as single-channel postfiltering).

5.3. Results

In this section, we present results for the diffuse noise case (white and babble) and directional noise case (white and interfering talker).

5.3.1. Diffuse noise

In Table 2 we show the average ATR obtained using all competing algorithms for two diffuse noise cases. In first case the target speech is contaminated by stationary Gaussian noise which has been generated independently for each channel and in the second case we have used omnidirectional babble noise to contaminate the target signal. As expected, all the algorithms perform better in presence of white noise compared to the babble noise case. Next, the proposed S-SBL approach achieves the best attenuation rate for both cases, most significantly so in the babble noise case. Informal subjective listening exercises to the output of the blocking matrix also consistently show noticeable differences.

Table 2. ATR measure in diffuse noise scenario

Algorithms	Diffused White Noise	Omni Babble Noise
	ATR (dB)	ATR (dB)
FD	-6.18	-3.68
NSFD	-11.24	-5.18
RLS	-7.36	-4.35
S-SBL (proposed)	**-12.05**	**-7.49**

5.3.2. Directional Noise

In Table 3 we present the average ATR obtained using all competing algorithms in directional noise. Specifically, in the first case the target speech is contaminated by directional Gaussian noise generated following the experimental setting discussed above, and in the second case we have used a male speaking interferer. This situation is more challenging compared to diffuse noise, even more so when the directional noise is a speech interferer. The performance of all the algorithms is reduced in directional white noise when compared with diffuse white noise. In Figure 1, 2 and 3 we show the spectrograms of the clean speech and the noise reference signal obtained using S-SBL and NSFD, respectively, in the case of directional white noise. It is evident from Figure 3 that dominant low-frequency speech harmonic structure is still present in the NSFD noise reference estimate. For a speech interferer, when there is no Voice Activity Detection (VAD) all algorithms struggle; particularly the FD and NSFD struggle (each producing positive ATR). The main reason behind this result is that the RTF estimate could be that of the speech interferer, since there is no way to distinguish who is the desired target. We also present results assuming that an Oracle VAD is available and see a significant improvement (as expected). In real life scenarios, an oracle VAD can be substituted by a VAD operating on a close talk microphone recording, or a phone microphone recording. We have conducted such experiments using the database presented in [21] and the results are encouraging.

6. CONCLUSION

We proposed a novel approach of estimating relative impulse response in a dynamic environment, i.e, using very short, noisy, reverberant recordings. Our proposed time domain solution benefits from exploiting the prior relative impulse response structure during estimation. Detailed experimental results also show consistent improvement of our proposed approach over competing algorithms. Incorporating this relative impulse response estimation technique in a generalized sidelobe canceller structure to improve the binaural noise suppression performance will be considered in our future works.

Table 3. ATR measure in presence of directional noise

Algorithms	White	Talker (Oracle VAD)	Talker (No VAD)
	ATR (dB)	ATR (dB)	ATR (dB)
FD	-3.98	-0.86	2.41
NSFD	-10.37	-9.63	1.62
RLS	-7.25	-11.40	-0.81
S-SBL	**-10.79**	**-15.72**	**-1.38**

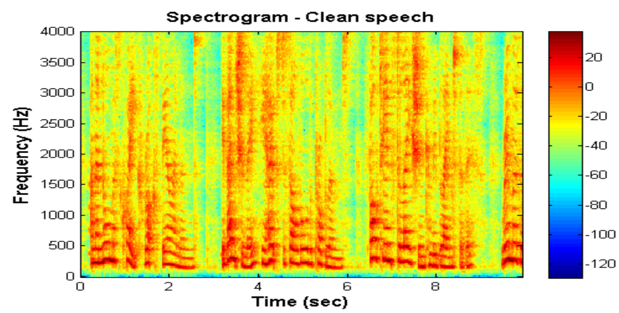

Fig. 1. Spectrogram of clean utterance recorded at left mic

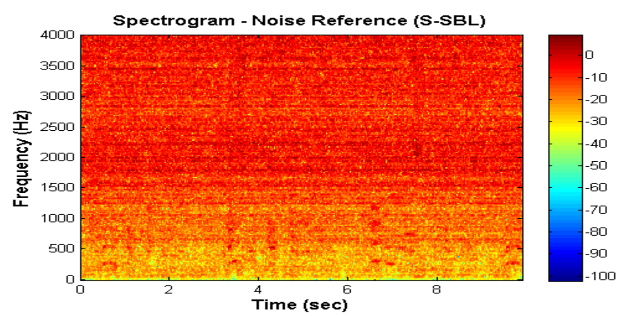

Fig. 2. Spectrogram of the noise reference signal obtained using S-SBL (Directional white noise)

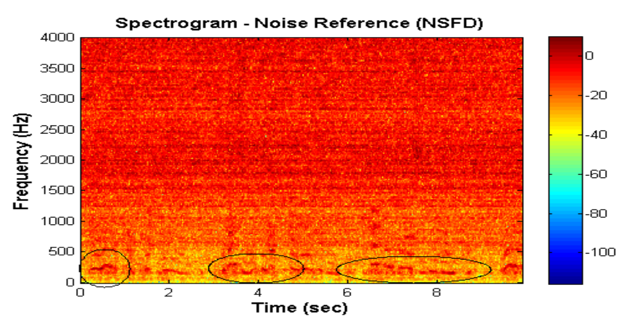

Fig. 3. Spectrogram of the noise reference signal obtained using NSFD (Directional white noise)

7. REFERENCES

[1] Sharon Gannot, David Burshtein, and Ehud Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, 2001.

[2] Sharon Gannot and Israel Cohen, "Speech enhancement based on the general transfer function gsc and postfiltering," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 6, pp. 561–571, 2004.

[3] Bracha Laufer, Ronen Talmon, and Sharon Gannot, "Relative transfer function modeling for supervised source localization," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013*. IEEE, 2013, pp. 1–4.

[4] Alexander Krueger, Ernst Warsitz, and Reinhold Haeb-Umbach, "Speech enhancement with a gsc-like structure employing eigenvector-based transfer function ratios estimation," *IEEE Transactions on Audio, Speech, and Language Processing* , vol. 19, no. 1, pp. 206–219, 2011.

[5] Shmulik Markovich, Sharon Gannot, and Israel Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1071–1086, 2009.

[6] M Schwab, P Noll, and T Sikora, "Noise robust relative transfer function estimation," in *IEEE 14th European Signal Processing Conference, 2006*. 2006, pp. 1–5.

[7] Ronen Talmon, Israel Cohen, and Sharon Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 546–555, 2009.

[8] Zbynek Koldovsky, Jiri Malek, and Sharon Gannot, "Spatial source subtraction based on incomplete measurements of relative transfer function," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 8, pp. 1335–1347, 2015.

[9] Jiri Malek and Zbynek Koldovsky, "Sparse target cancellation filters with application to semi-blind noise extraction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*. IEEE, 2014, pp. 2109–2113.

[10] Israel Cohen, Sharon Gannot, and Baruch Berdugo, "Real-time tf-gsc in nonstationary noise environments," *Israel Institute of Technology*, pp. 1–4, 2003.

[11] Yuanqing Lin and Daniel D Lee, "Bayesian regularization and nonnegative deconvolution for room impulse response estimation," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 839–847, 2006.

[12] Yuanqing Lin, Jingdong Chen, Youngmoo Kim, and Daniel D Lee, "Blind channel identification for speech dereverberation using l1-norm sparse learning," in *Advances in Neural Information Processing Systems*, 2007, pp. 921–928.

[13] Zbynek Koldovsky, Petr Tichavsky, and David Botka, "Noise reduction in dual-microphone mobile phones using a bank of pre-measured target-cancellation filters," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013*. IEEE, 2013, pp. 679–683.

[14] Donald W Marquardt and Ronald D Snee, "Ridge regression in practice," *The American Statistician*, vol. 29, no. 1, pp. 3–20, 1975.

[15] Michael E Tipping, "Sparse bayesian learning and the relevance vector machine," *The journal of machine learning research*, vol. 1, pp. 211–244, 2001.

[16] Ritwik Giri and Bhaskar D Rao, "Type i and type ii bayesian methods for sparse signal recovery using scale mixtures," *arXiv preprint arXiv:1507.05087*, 2015.

[17] David P Wipf and Bhaskar D Rao, "Sparse bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[18] David Wipf and Srikantan Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 317–329, 2010.

[19] Elior Hadad, Florian Heese, Peter Vary, and Sharon Gannot, "Multichannel audio database in various acoustic environments," in *14th International Workshop on Acoustic Signal Enhancement (IWAENC), 2014*. IEEE, 2014, pp. 313–317.

[20] Nobutaka Ono, Zbynek Koldovsky, Shigeki Miyabe, and Noboru Ito, "The 2013 signal separation evaluation campaign," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2013*. IEEE, 2013, pp. 1–6.

[21] William S. Woods, Elior Hadad, Ivo Merks, Buye Xu, Sharon Gannot, and Tao Zhang, "A real-world recording database for ad hoc microphone arrays," in *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2015.