International Journal of Engineering & Technology, 7 (4) (2018) 2848-2852

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.18771
Research paper

Bayesian probabilistic approach by blind source separation for instantaneous mixtures

Pallavi Agrawal [1] *, Madhu Shandilya [1]

[1] *Department of Electronics and Communication MANIT, Bhopal, India*
**Corresponding author E-mail: pallaviagrawal4@gmail.com*

Abstract

In this work, the novel method of blind source separation using Bayesian Probabilistic approach is discussed for instantaneous mixtures. This work demonstrates the source separation problem which is well suited for the Bayesian approach. This work also provides a natural and logically consistent method in which prior knowledge can be incorporated to estimate the most probable solution. The distri-butions of the coefficients of the sources in the basis are modeled by a generalized Gaussian distribution (GGD) which is dependent on the sparsity parameter q. This method also utilizes prior distribution of the appropriate sparsity parameter of sources present in the mixture. Once, the prior distribution for each parameter (like mixing matrix, source matrix, sparsity parameter and error or noise covariance matrix) are defined, the Bayesian a posterior probabilistic approach using Markov chain Monte Carlo (MCMC) method is exploited in estimation of a posterior distribution of mixing matrix, source matrix, sparsity parameter and covariance matrix of error. The blind source separation provides the results in the form of signal to distortion ratio (SDR), signal to artifacts ratio (SAR) and signal to interference ratio (SIR) at different SNR.

Keywords: *Gaussian Distribution; Markov Chain Monte Carlo; Noise Covariance; Signal Distortion Ratio; Signal To Interference Ratio.*

1. Introduction

Bayesian sound source separation [1], [2] method is used to separate the various sources from the combined mixture of signals recorded by the microphone. The best example of mixed sound sources is a cocktail party organized by a small group of people. In this party there is one speaker who are speaking as a near end signal using one microphone at a particular instant of time. All the conversation by the speakers are recorded over the microphone. In this way, all speakers speak subsequently using the microphone. For example, suppose there are four microphone and three speakers are speaking using one of these microphone. Then the conversation of this speaker is being recorded by the rest of other three microphone. A similar situation occurs with rest of the speakers who are using the microphone. In this way microphone records the speech of all the speakers and sound is mixed together inside the microphone and it is difficult to separate these sources. In this work, the blind source separation [4] for over-determined mixtures are described using Bayesian a posterior approach [3]. The Bayesian inference on the unknown parameters is conducted using a Markov chain Monte Carlo (MCMC) methods [5], [6]. To promote sparsity in the sources, present in the mixture, the source matrix uses the generalized Gaussian distribution (GGD) [7], [8]. The GGD has a sparsity parameter which has Gamma distribution [9] as a prior distribution. The prior distribution for mixing matrix and error covariance matrix is taken in such a way that the solution for the equivalent posterior distribution in a closed form can be achieved. In such cases, the Gibbs sampling can be exploited for obtaining the samples for mixing matrix and error covariance matrix. On the other hand, the Metropolis Hasting algorithm is applied for obtaining the samples from the a posterior distribution of source matrix and sparsity parameter because the closed form is not achieved for the equivalent posterior distributions. The results are presented in the form of SDR, SIR and SAR at different SNRs.

The rest of the paper is organized as follows: Section 2 describes the Bayesian model followed by analysis of the estimation of its Parameters using Posterior Conditionals. Section 3 covers the performance evaluation followed by the result and analysis in section 4 and 5. Finally, section 6 presents a brief conclusion and future scope.

2. Research method

2.1. Bayesian model

In this section, the Bayesian analysis associated with the mixing model defined in Equation 1 is described. This Bayesian analysis is defined in terms of source separation likelihood function and priors defined for each parameter present in the mixing model.

2.1.1. Source separation likelihood

In this work, the Bayesian approach is used to obtain statistical inference via available prior information either from subjective expert experience or prior experiments. This prior information yields progressively less influence in the final results as the sample size increases. The source separation model can be written as

$$(X|S,W) = SW' + N \quad (1)$$

Where, X, S, W and N are mixed signal matrix, source signal matrix, mixing matrix and noise matrix respectively. The dimension of S, W, N and X is n × m, p × m, n × p and n × p. Here, X = $(x_1, x_2, \cdots, x_n)^T$.T here corresponds to the transpose operation. Each

x_i, is a p-dimensional observed vector and i∈ (1, 2, · · ·, n) is a time increment. Also, m and p correspond to number of sources and number of sensors and p > m. Similarly, S is defined as S = $(s_1, s_2, \cdots, s_n)^T$ and each s_i is m dimensional true unobservable source vector. The ensuing section deals with the Bayesian statistical model which includes likelihood function and prior distribution for each parameter.

The error or number of noise samples observations are supposed to be independent to each other and they are normally distributed with zero mean and full positive definite covariance matrix C. Thus, the observation matrix X is multivariate normally distributed given W as a mixing matrix, S as the source matrix and the C as noise covariance matrix. The likelihood of X is expressed as:

$$p(X|S,W,C) \propto |C|^{-n/2}\, e^{-\frac{1}{2}tr(X-SW')C^{-1}(X-SW')'} \quad (2)$$

Here, tr (.) is the trace operator. Thus the objective is to estimate the source S which is present in the mixture. And also to obtain the knowledge about the way the mixing process occurred by the speakers, the estimation of W as mixing matrix and noise covariance matrix of C are required. The ensuing section describe the distribution for prior model parameters required for estimating the corresponding a posterior distribution.

2.1.2. Prior and conjugate priors

In this section, the priors and conjugate priors are defined for each model parameters. In source separation scenario using Bayesian approach, the prior distribution is used to quantify the information available regarding the values of model parameters. In this work, the conjugate prior are used for mixing the source matrix S and noise covariance matrix of C and normal priors are used for source matrix S and sparsity parameter. The joint prior distribution for the model parameters which are source matrix S, mixing matrix W, the error or noise covariance matrix C and the sparsity parameter of the source matrix q is given by

$$(S,W,C,q) = p(S|q) * p(q) * p(W|C) * P(C) \quad (3)$$

Where p (S|q), p(q), p(W|C) and p(C) are defined as

$$P(S|q) \propto \frac{q}{\Gamma(1/q)} e^{|S|^q} \quad (4)$$

$$p(q) \propto \text{Gamma}(a,b) \quad (5)$$

$$p(W|C) \propto |D|^{-p/2}\, |C|^{-m/2} e^{-\frac{1}{2}trC^{-1}(W-W_0)D^{-1}(W-W_0)'} \quad (6)$$

$$p(C) \propto |C|^{-\frac{v}{2}} e^{-\frac{1}{2trC^{-1}Q}} \quad (7)$$

Here, positive definite matrices are C, Q and D. Also W_0, v, Q, a, b and D are hyper-parameters need to be assessed. Once, these hyper-parameters are assessed completely, and then joint prior distribution can be determined. Thus, by using Bayes rule, the unknown parameters of joint posterior distribution is proportional to the product of the likelihood function and joint prior distribution as shown below [10].

The joint posterior distribution will now be used to obtain model parameters S (the source matrix), the sparsity parameter q, the mixing matrix W and the error or noise covariance matrix C.

2.2. Estimation of model parameters using posterior conditionals

In this section, the mixing model parameters are estimated using Posterior conditionals separately.

2.2.1. Estimation of source matrix s

The source matrix S for the conditional posterior distribution can be found by seeing only those terms that involves the joint posterior distribution which in turn, involves only S. Hence, conditional Posterior distribution for S is given by

$$p(S,W.C.q|X) \propto p(X|S,W,C) * p(S/q) * p(q) * p(W|C) * P(C) \quad (8)$$

$$P(S|q,W,C,X) \propto P(S|q) * P(X|S,W,C) \quad (9)$$

$$P(S|q,W,C,X) \propto \frac{q}{\Gamma(1/q)} e^{-|S|^q * |C|^{-\frac{n}{2}}} e^{-\frac{1}{2}tr(X-SW')C^{-1}(X-XW')'} \quad (10)$$

It can be seen from Equation 9 that we cannot simulate directly from these Posterior distributions because they are not in a closed form. It is also difficult to get closed form for Equation 9. To solve this problem, the Metropolis-Hasting (M-H) algorithm is utilized [11]. In M-H algorithm, the proposal distribution for S is taken to be normal, N (μt−1, ζ2). Here, μt−1 is the mean for t-1 iteration and ζ2 is the variance. To increase the sensitivity of the random walk sampler, the scale of the random walk was chosen to be ζ2 = 0.01 [9]. The M-H algorithm used for sampling is explained in the appendix section.

2.2.2. Sparsity parameter q estimation

The sparsity parameter q for the conditional posterior distribution can be found by seeing only the terms of q in the joint posterior distribution. Hence, the conditional Posterior distribution for q is given by

$$P(q|S,W,C,X) \propto P(q) * P(S|q) * P(X|S,W,C) \quad (11)$$

$$P(S|q,W,C,X) \propto \Gamma(a,b) * \frac{q}{\Gamma(1/q)} e^{-|S|^q * |C|^{-\frac{n}{2}}} e^{-\frac{1}{2}tr(X-SW')C^{-1}(X-SW')'} \quad (12)$$

From Equation 12, it can be noted that similar to Equation 9, we cannot simulate directly from these Posterior distribution because they are not in a closed form. So in this case also, the M-H algorithm is used with normal distribution as the proposal distribution. The prior distribution for q is taken to be gamma distribution as explained in detail in [10].

2.2.3. Estimation of mixing matrix

The mixing matrix W for the conditional posterior distribution is obtained by seeing only those terms of W only in the posterior distribution. Hence, the conditional Posterior distribution for W is given by

$$p(W|S,C,X) \propto p(W|C)p(X|S,W,C) \quad (13)$$

Here, the new variable $\widehat{W}$ corresponds to the posterior conditional mean and thus defined as

$$p(W|S,C,X) \propto e^{-\frac{1}{2}trC^{-1}(W-\widehat{W})(D^{-1}+S'S)(W-W')'} \quad (14)$$

$$\widehat{W} = (W_0 D^{-1} + X'S)(D^{-1} + S'S)^{-1} \quad (15)$$

Since, the form shown in Equation 14 is in a closed form, so Gibbs sampling [11] can be used to obtain samples from this Posterior distribution. The posterior conditional distribution for the matrix of W as the mixing coefficients given the matrix of sources S, the error covariance matrix C, and the data matrix X is Matrix normally distributed.

2.2.4. Error covariance matrix estimation

In estimation of conditional Posterior error covariance matrix estimation C, the conditional Posterior distribution for C consist of only those terms which involves only C. Thus, it is given by

$$p(C|S,W,X) \propto p(W|C) * p(C) * p(X|S,W,C) \quad (16)$$

$$P(C|S,W,X) \propto |C|^{\frac{-(m+v+n)}{2}} e^{-1/2 trC^{-1}\hat{C}} \quad (17)$$

Here, the variable C can be defined as

$$\hat{C} = (X - SW')'(X - SW') + (W - W_0)D^{-1}(W - W_0)' + Q \quad (18)$$

The posterior conditional distribution of the observation error covariance matrix given S is the matrix of sources, the W is the matrix of mixing coefficients, and X is the data which is defined as Inverted Wishart.

3. Performance evaluation

Different evaluation tools such as signal to distortion ratio (SDR), signal to interference ration (SIR) and signal to artifacts ratio (SAR) [12] measures the overall performance of the proposed work of blind source separation Experiments are performed using GRID corpus database which consists of high quality of audio and video signals recordings of 1000 sentences spoken by each of 34 talkers (18 male, 16 female). Sentences are of the form "put red at G9 now". The corpus, together with transcriptions, is freely available for research use. In the experiment, 158 different mixtures of speakers are utilized which speaks different types of sentences. These mixtures are then mixed together with AWGN noise at different SNRs. The number of sources present in the mixture is 2. The number of microphones used in this work is 4. It may be noted that all the model parameters used in the mixture are randomly initialized and iterated up to 1000 times to achieve corresponding optimal values. In 1000 iterations, 500 are burn in so the last 500 iterations are used to achieve model parameters. The hyperparameter used in the assignment are described in the ensuing section.

3.1. Hyperparameter assignment

In this section, the different hyperparameter used in calculation of Posterior distribution of individual model parameters are defined. Firstly, the score is obtained after applying principal component analysis on the mixture matrix X.

$$\text{Score=princomp(X)} \quad (19)$$

The S_0 is then obtained by considering all the rows of score for total number of sources as shown below.

$$S_0 = \text{score}(:,1:\text{Number of Sources}) \quad (20)$$

The hyperparameter D is obtained as

$$D = (S_0' * S_0)^{-1} \quad (21)$$

The hyperparameter n is taken equal to length of the greatest speech signals present in the mixture. The hyperparameter W_0 is taken as

$$W_0 = X * S_0 * D \quad (22)$$

The scale matrix Q for the error covariance matrix was assessed from

$$Q = (X - S_0 * W_0')' * (X - S_0 * W_0') \quad (23)$$

The hyperparameter a and b used in the work are kept fixed at each iteration and taken equal to 0.2 and 2 [9] respectively.

4. Results

MATLAB toolbox of BSS Eval is used to determine the performance of (blind) source separation algorithms. The original source signals are available as ground truth [12]. In this work, the SDR, SIR and SAR are obtained for proposed method at different SNRs. The SNR used in this work are 50 dB and 10 db. The mean SDR, SIR and SAR are obtained from all the mixtures after averaging the 1000 iterations.

5. Analysis

The Figure 1 correspond to the mean or average SDR, SIR and SAR at SNR equal to 50 dB. The variation of SDR, SIR and SAR with total mixture is shown in Figure 2, 3 and 4 respectively. It can be seen from these Figures that the variation in SDR, SIR and SAR is not much among the different mixtures. This indicates the significance of the proposed method for blind source separation problem. Moreover, the standard deviation is also within 6 dB for all the three SDR, SIR and SAR. Similar explanation can be obtained for SDR, SIR and SAR variation for different mixture at SNR equal to 10 dB in Figure 6, 7 and 8 respectively. Moreover, the average SDR, SIR and SAR in figure 5 at 10 dB SNR indicates that the proposed method have still significance performance for BSS problem. Although the performance have been decreased with decrease in SNR but still the SDR, SIR and SAR values shows reasonable improvement in BSS at this SNR.

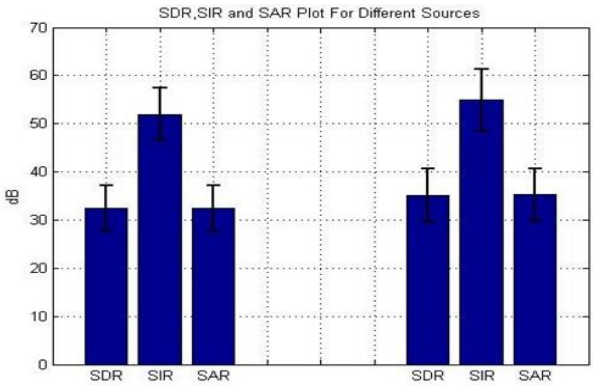

Fig. 1: A Bar Graph Illustrating Mean Value of SDR, SIR AND SAR at DB 50 SNR.

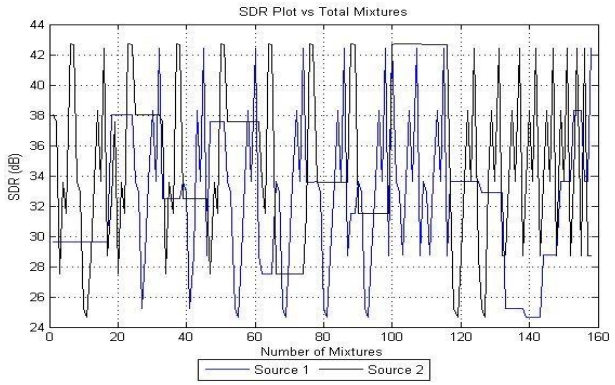

Fig. 2: A Plot Illustrating the Variation of SDR with Different Mixtures at 50 DB SNR.

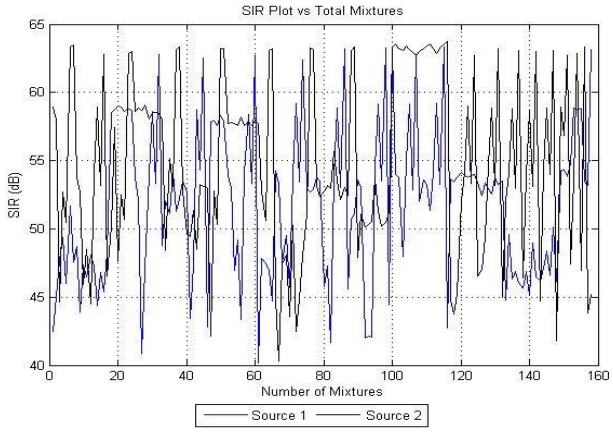

Fig. 3: A Plot Illustrating the Variation of SIR with Different Mixtures at 50 DB SNR.

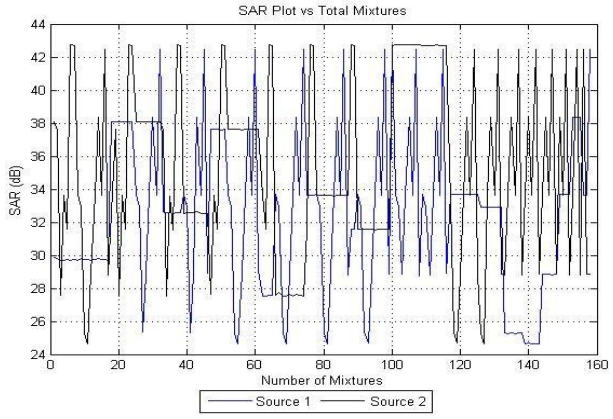

Fig. 4: A Plot Illustrating the Variation of SAR with Different Mixtures At 50 DB SNR.

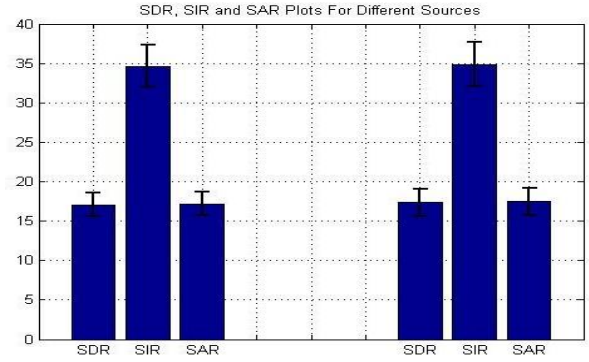

Fig. 5: A Bar Graph Illustrating the Mean Value of SDR, SIR and SAR at 10 DB SNR.

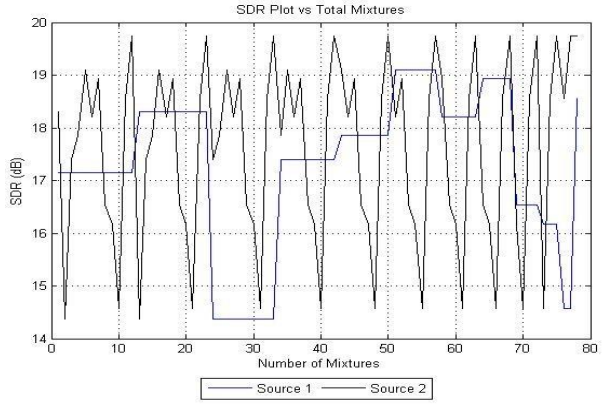

Fig. 6: A Plot Illustrating the Variation of SDR with Different Mixtures at 10 DB SNR.

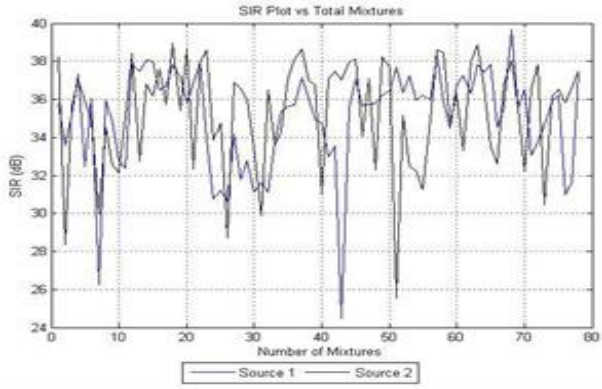

Fig. 7: A Plot Illustrating the Variation of SIR with Different Mixtures at 10 DB SNR.

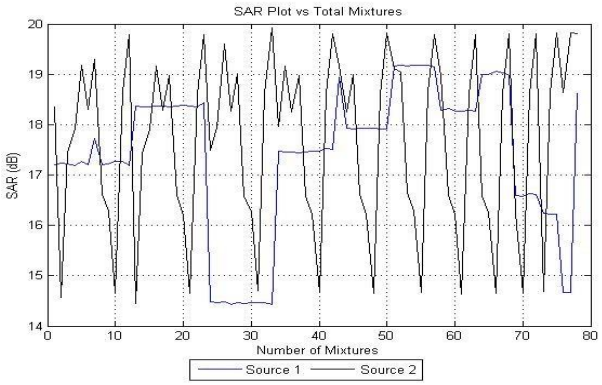

Fig. 8: A Plot Illustrating the Variation of SAR with Different Mixtures at 10 DB SNR.

6. Conclusion

In this work, the Bayesian probabilistic approach is discussed for instantaneous mixture for BSS application. This work estimates separately the optimal conditional Posterior of source matrix, mixing matrix, sparsity parameter and error or noise covariance matrix iteratively. The sparsity parameter is promoted by using generalized Gaussian distribution for source matrix. Once the prior for each model parameter is defined along with the hyper-parameters, the conditional Posterior for each model parameter are obtained. The performance of the proposed method shows significant improvement in terms of source separation and noise cancellation obtained at different SNR.

References

[1] V. P. Minotto, C. R. Jung, and B. Lee, "Multimodal on-line speaker diarization using sensor fusion through SVM," *IEEE Trans. Multimedia*, vol. 17, no. 10, pp. 1694–1705, Oct. 2015. https://doi.org/10.1109/TMM.2015.2463722.

[2] N. Sarafianos, T. Giannakopoulos, and S. Petridis, "Audio-visual speaker diarization using Fisher linear semi-discriminant analysis,*"Multimedia Tools Appl.,* vol. 75, no. 1, pp. 115–130, 2016 https://doi.org/10.1007/s11042-014-2274-x.

[3] I. Kapsouras, A. Tefas, N. Nikolaidis, G. Peeters, L. Benaroya, and I. Pitas, "Multimodal speaker clustering in full length movies,"*Multimedia Tools Appl.*, pp. 1–20, 2016.

[4] I. D. Gebru, S. Ba, G. Evangelidis, and R. Horaud, "Tracking the active speaker based on a joint audio-visual observation model," *in Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 15–21.

[5] A. Deleforge, R. Horaud, Y. Y. Schechner, and L. Girin, "Colocalization of audio sources in images using binaural features and locally-linear regression*," IEEE Trans. Audio Speech Language Process., vol.* 23, no. 4, pp. 718–731, Apr. 2015. https://doi.org/10.1109/TASLP.2015.2405475.

[6] I. D. Gebru, X. Alameda-Pineda, *et al.*, "EM algorithms for weighted-data clustering with application to audio-visual scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 12,

pp. 2402–2415, Dec. 2016. https://doi.org/10.1109/TPAMI.2016.2522425.

[7] P. Agrawal, and M. Shandilya. "Model-Based Method for Acoustic Echo Cancelation and Near-End Speaker Extraction: Non-negative Matrix Factorization" *Journal of Telecommunications & Information Technology*, 2 (2018).

[8] G. Skantze, A. Hjalmarsson, and C. Oertel, "Turn-taking, feedback and joint attention in situated human–robot interaction," *Speech Commun.*, vol. 65, pp. 50–66, 2014. https://doi.org/10.1016/j.specom.2014.05.005.

[9] I. D. Gebru, S. Ba, X. Li, and R. Horaud, "Audio-Visual Speaker Diarization Based on Spatiotemporal Bayesian Fusion*,"IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 5, may 2018. https://doi.org/10.1109/TPAMI.2017.2648793.

[10] L. Bourdev and J. Malik, "Poselets: Body part detectors trained using 3d human pose annotations," *in Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 1365–1372.

[11] X. Li, L. Girin, R. Horaud, and S. Gannot, "Estimation of relative transfer function in the presence of stationary noise based on segmental power spectral density matrix subtraction," *in Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 320–324.

[12] X. Li, R. Horaud, L. Girin, and S. Gannot, "Local relative transfer function for sound source localization," *in Proc. Eur. Signal Process. Conf.*, Aug. 2015.