



HAL
open science

The Maximally-Coherent Reference Technique: a solution to deal with large sets of references

M N Albezzawy, Jérôme Antoni, Quentin Leclere

► **To cite this version:**

M N Albezzawy, Jérôme Antoni, Quentin Leclere. The Maximally-Coherent Reference Technique: a solution to deal with large sets of references. Proceedings of ISMA2022, Sep 2022, Leuven, Belgium. hal-04063060

HAL Id: hal-04063060

<https://hal.science/hal-04063060>

Submitted on 8 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Maximally-Coherent Reference Technique: a solution to deal with large sets of references

M. N. Albezzawy¹, J. Antoni¹, Q. Leclère¹

¹ Univ Lyon, INSA Lyon, LVA, EA677,

69621 Villeurbanne, France.

e-mail: muhammad.albezzawy@insa-lyon.fr

Abstract

Source separation, using a number of references that largely exceeds the number of sources, is hindered by the ill-condition of the cross-spectral matrix of the references. This has been solved in literature by reducing the dimension of the reference set, for instance by a truncated eigenvalue decomposition or Gram-Schmidt orthogonalization. The main drawback of these techniques is that the resultant virtual references are not necessarily maximally correlated with the sources. This paper exclusively introduces an original solution for this problem, coined “Maximally-Coherent Reference Technique”, based on finding a set of virtual references so that they are maximally correlated with the sources. The number of sources is estimated using parametric bootstrap, which is advantageous over the asymptotic likelihood ratio test. The method is tested experimentally using numerical and real data, and is proved to solve the source separation problem while avoiding the shortcomings of the existing reference-based techniques.

1 Introduction

The characterization of sound sources is a common problem in acoustics. It aims to localize, identify, and rank sound sources as measuring directly on the source is often impractical. It is done through sensing the propagated sound on some plane on its transfer path using a microphone array. The acquired data are then used along with a propagation model to estimate the source characteristics. When this characterization process results in a map of the source field of interest, it is often called Acoustic Imaging. Acoustic Imaging techniques can be split into two main groups: Beamforming techniques and Inverse techniques. More detail on the global classification of the Acoustic Imaging techniques can be found in Reference [1]. If a composite field exists, where the contributions from incoherent sources overlap in both the spatial and the frequency domain[2], no definite single phase value can be assigned at any point on the hologram plane, and hence sources can be hardly identified by Acoustic Imaging [3]. In such cases a pre-processing source separation step is needed before retro-propagating the sound field.

If some perfect references, i.e. pure signals such that each of which is coherent with an individual source and incoherent with the rest, are available, they can be used to decompose the composite field [3]. An early application of the use of references was introduced by Hald [4] in the context of the Near-field Acoustical Holography (NAH). The reference-based techniques are split into two distinct categories based on the number of available references; if the number of references is equal to or higher than the number of incoherent sources in the system, it is an over-determined case, otherwise, it is an under-determined case. The over-determined case is the subject of this paper, especially the cases when the number of references largely exceeds the number of sources, what results in a partially correlated, and potentially noisy, set of references. This paradoxically jeopardizes the separation because the cross spectral matrix of the references becomes ill-conditioned.

1.1 The over-determined case

The number of these references must be at least equal to the number of incoherent sources in the system (the field's stochastic dimension) to be able to represent the whole sound field. When the number of references are exactly equal to the number of sources, this will correspond to a well-conditioned cross-spectral matrix (i.e. a reference cross spectral matrix with all its eigenvalues are relatively important). To achieve this condition, references are often placed as close as possible to the known physical partial sources, for more guidance on the selection and placement of reference sensors the reader is referred to the technical review by Hald [4]. The multiple coherence function, which is the sum of the ordinary coherences of each incoherent reference with the output, can be used as an index to evaluate the sufficient number of chosen references to represent the stochastic dimension of the field, and hence, offers the capability of accurately separating the total coherent output spectra from the output noise [5]. Unfortunately, perfect uncorrelated references are hardly available in practice especially when the number of references exceeds the number of sources, in this case, the multiple coherence function is not simply equal to the sum of the ordinary coherences of each incoherent reference with the output [6]. This was traditionally solved in two ways; the first was the Conditioned Spectral Analysis technique introduced by Bendat and Piersol [6], and the second was the Virtual Sources Analysis technique proposed by Price and Bernhard [7]. These two techniques enables the decomposition of the correlated MIMO system into an equivalent set of uncorrelated SIMO systems that can be then quadratically superimposed [8]. All the references-based methods can be collectively called coherence techniques.

1.1.1 Conditioned Spectral Analysis

Conditioned Spectral Analysis, which was presented by Bendat and Piersol [6] in the early 70s, is a recursive procedure that produces residual spectra from which the contributions of the reference signals are successively removed. This procedure is nothing but finding an orthogonal basis for the signals' subspace, and is equivalent to a Gram-Schmidt orthogonalization of an ordered version of Fourier spectra of the signals. This technique is also called the Partial Coherence technique. The partial coherence is defined as "the ordinary coherence between a conditioned reference and the the output" [6]. The ordering operation required before the conditioning operation is essentially problematic, as a model with arbitrary M inputs, will theoretically require analysing an $M!$ models which are conditioned by different input orders. To avoid this, one has to have a priori knowledge on the cause-and-effect relations between every pair of inputs, which is practically not guaranteed [6]. Besides, in the case of systems with large number of references (inputs), the probability of existence of perfectly correlated inputs arises, what would lead to numerical problems (singularity) during the conditioning operation [5, 7]. Solutions were proposed in the literature to tackle this challenge. The first of them was suggested by the technique's inventors in Reference [6], where they advised to reorder the inputs in a descending order according to their ordinary coherence magnitude. Another solution, for example, was causality evaluation by inspecting the negative part of the impulse response functions between inputs, or by checking the Hilbert transform pairness of the FRFs, which was found superior over ordinary coherence-based ordering [9].

1.1.2 Virtual Sources Analysis

The aforementioned difficulties, among others, led to the introduction of the principal component analysis (PCA)-based technique called Virtual Sources Analysis, which was seminally proposed by Price and Bernhard [7] under the name of Virtual Coherence. The main idea behind the technique is reducing the rank of the cross-spectral matrix of the inputs (references), so as to meet the stochastic dimension of the system using truncated singular value decomposition (TSVD). This is achieved through the diagonalization of the cross-spectral matrix, using eigenvalue decomposition, and then discarding the unimportant eigenvalues. The remaining eigenvalues represents the auto-spectra of incoherent virtual sources, which are equal in number to the real sources and are linear combinations of them. The complex amplitudes of these virtual sources as well as their individual contributions to the output are then easily computed, as clearly explained in [5, 7, 8].

An application example in NAH can be found in Reference [10]. Two major challenges facing the effectiveness of this technique can be stated. The first one, is that discarding some of the virtual sources based only on their low eigenvalue may lead to losing information concerning a source of interest (SOI), because it may end up with a number of virtual sources lower than the number of actual incoherent sources (i.e. lower than the stochastic dimension of the field). Hence, for this technique to be useful, the stochastic dimension of the field should be known beforehand. The second challenge is that these virtual sources lack physical significance with respect to the description of the output (and so the actual sources), i.e. they can not be attributed to specific physical phenomena. Actually, all that they represent is nothing but the oriented energy distribution of the inputs (reference signals). Leclère et al. [11] decided on the significance of a partial (or virtual) reference based on its degree of coherence with the output. Hence, they introduced the so called Threshold Multiple Coherence which is the summation of the individual partial (or virtual) coherences, conditioned to their significance. The significance of a reference's coherence with the output is judged using a predefined threshold. The limitation of this method is that it is only formulated for MISO systems.

This paper exclusively introduces an original solution for dealing with these challenges based on finding a set of virtual references, which are necessarily linear combinations of the measured ones, so that they are maximally correlated with the outputs and so with the actual sources. This technique is coined “The Maximally-Coherent Reference (MCR) technique” and it is formulated to deal with the complex, more general MIMO model. When tested against both the least-squares solution (LSS) and the TSVD-based solution, using simulated partially correlated and noisy sets of references, the MCR solution proved to be the most accurate among them. Similar to the other techniques, the proposed technique needs the number of sources to be a priori known. The paper also introduces a parametric bootstrap algorithm for estimating the number of sources. It is advantageous over the asymptotic likelihood ratio test, which faces some limitations in the cases of dealing with factor analysis models.

2 Problem statement

Let $\mathbf{Y} \in \mathbb{C}^{M \times I}$ denote a set of complex Fourier coefficients of measurements returned by M output sensors and recorded for I independent snapshots. The data \mathbf{Y} are supposed to be produced by S sources, say $\mathbf{S} \in \mathbb{C}^{S \times I}$, whose contributions are noted $\mathbf{X} = \mathbf{H}\mathbf{S} \in \mathbb{C}^{M \times I}$ for some linear but unknown operator $\mathbf{H} \in \mathbb{C}^{M \times S}$, and corrupted by additive disturbances $\mathbf{N} \in \mathbb{C}^{M \times I}$ uncorrelated with \mathbf{X} , such that

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}. \quad (1)$$

It is assumed that the number of output sensors M exceeds or is equal to the number of sources, i.e. $M \geq S$. The data \mathbf{Y} comes with a set of R references, $\mathbf{R} \in \mathbb{C}^{R \times I}$, which are supposed to be perfectly correlated with the S sources, in the sense that there exists a linear but unknown operator \mathbf{L} such that

$$\mathbf{R} = \mathbf{L}\mathbf{S}. \quad (2)$$

In practice, a small amount of additive noise may be present too in the references, such that

$$\mathbf{R} = \mathbf{L}\mathbf{S} + \epsilon\mathbf{E}, \quad (3)$$

with $\epsilon \geq 0$ arbitrarily small, and where the noise covariance matrix on the references, $\mathbf{C}_{EE} = \mathbb{E}\{\mathbf{E}\mathbf{E}^H\}$, is not necessarily proportional to the identity.

It is further assumed that the number of references exceeds or is equal to the number of sources, i.e. $R \geq S$. The aim is to predict the contributions \mathbf{X} from the references \mathbf{R} , i.e. to find an operator $\mathbf{G} \in \mathbb{C}^{M \times R}$ such that

$$\hat{\mathbf{X}} = \mathbf{G}\mathbf{R} \quad (4)$$

is an estimate of \mathbf{X} .

3 Solution based on optimal linear combinations of references

A simple solution to the above problem that is often used in practice is provided by the least square estimate

$$\mathbf{G} = \underset{\mathbf{A}}{\text{Arg min}} \|\mathbf{Y} - \mathbf{A}\mathbf{R}\|^2 = \mathbf{S}_{YR}\mathbf{S}_{RR}^{-1}, \quad (5)$$

with $\mathbf{S}_{RR} = \mathbf{R}\mathbf{R}^H/I$ and $\mathbf{S}_{YR} = \mathbf{Y}\mathbf{R}^H/I$.

Unfortunately, the matrix \mathbf{S}_{RR} might be badly conditioned when $R > S$ and ϵ is small; it tends to be singular when $\epsilon \rightarrow 0$ and $R > S$.

A popular way to deal with the aforementioned situation is to replace the set of R references \mathbf{R} by their first S principal components, as returned by PCA. Although this transformation returns a pseudo inverse, there is no guarantee that the so selected S components are the best ones to predict the contributions \mathbf{X} or, in other words, are the most correlated with the sources \mathbf{S} .

In light of this discussion, a better choice is to select S linear combinations of references \mathbf{R} that best predict the data \mathbf{Y} . This amounts to finding a vector \mathbf{b}_1 such that $\mathbf{z}_1 = \mathbf{b}_1^H \mathbf{R}$ is maximally correlated with \mathbf{Y} , then to remove the contribution of \mathbf{z}_1 from the data and to find the next linear combination of references defined by vector \mathbf{b}_2 such that $\mathbf{z}_2 = \mathbf{b}_2^H \mathbf{R}$ is maximally correlated with the residual, etc. The proposed algorithm reads

- Set $\mathbf{Y}_0 = \mathbf{Y}$
- FOR $i = 1$ to S
 - Find \mathbf{b}_i : $\max \text{corr}(\mathbf{Y}_i, \mathbf{b}_i^H \mathbf{R})$
 - Set $\mathbf{z}_i = \mathbf{b}_i^H \mathbf{R}$
 - Find \mathbf{c}_i : $\min_{\mathbf{c}_i} \|\mathbf{Y}_i - \mathbf{c}_i \mathbf{z}_i\|^2$
 - Set $\mathbf{Y}_i = \mathbf{Y}_{i-1} - \mathbf{c}_i \mathbf{z}_i$
- $i \leftarrow i + 1$
- END if $i > S$

Let denote $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_S]$ and $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_S]$ the values obtained from the above algorithm. Then, an estimate of the contribution \mathbf{X} in the data \mathbf{Y} is returned by

$$\hat{\mathbf{X}} = \sum_{s=1}^S \mathbf{c}_s \mathbf{b}_s^H \mathbf{R} = \mathbf{C}\mathbf{B}^H \mathbf{R}. \quad (6)$$

It is not hard to show that this algorithms is equivalent to the generalized eigenvalue decomposition (GEVD) of the pair of matrices $(\mathbf{S}_{RY}\mathbf{\Gamma}\mathbf{S}_{YR}, \mathbf{S}_{RR})$, where only the eigenvectors associated to the S largest eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_S \geq 0$ are conserved, for some positive definite matrix $\mathbf{\Gamma} \in \mathbb{C}^{M \times M}$ which can be chosen to be the identity matrix \mathbf{I} or the output cross-spectral matrix $\mathbf{S}_{Y\mathbf{Y}}^{-1}$. It is clear by looking at the first matrix of this generalized eigenvalue problem that, if $M < R$, then the rank of that first matrix is M , and there is no hope to search for more than M sources. This means that one must have $M \geq S$ if the effects of all sources are to be recovered, as initially assumed.

The matrix \mathbf{B} then will be the matrix consisting of the eigenvectors as its columns, and $\mathbf{C} = \mathbf{S}_{YR}\mathbf{B}$

Equation (6) then becomes

$$\hat{\mathbf{X}} = \mathbf{C}\mathbf{B}^H \mathbf{R} = \mathbf{S}_{YR}\mathbf{B}\mathbf{B}^H \mathbf{R}. \quad (7)$$

As compared to Equation (5), the above result replaces the (possibly unstable or non-existing) inverse operator \mathbf{S}_{RR}^{-1} by the projector $\mathbf{B}\mathbf{B}^H$.

3.1 Generalized Eigenvalue problem with rank deficient second matrix

When there are no noises on the references (i.e $\epsilon = 0$ in Eq.(3)), \mathbf{S}_{RR} is rank deficient, and the generalized eigenvalue problem becomes ill-posed as it does not possess a unique solution [12]. Melzer [13] referred to a

solution to this problem through a reduced rank simultaneous diagonalization algorithm.

The idea is to find a nonsingular transformation $\mathbf{T} \in \mathbb{C}^{R \times r}$, where $r = \text{rank}(\mathbf{S}_{RR})$, such that

$$\mathbf{T}^H \mathbf{S}_{RY} \mathbf{\Gamma} \mathbf{S}_{YR} \mathbf{T} = \mathbf{\Lambda} \quad (8)$$

$$\mathbf{T}^H \mathbf{S}_{RR} \mathbf{T} = \mathbf{I} \quad (9)$$

so that $\mathbf{\Lambda}$ and \mathbf{T} are the eigenvalues and eigen-vectors of the generalized eigenvalue decomposition $(\mathbf{S}_{RY} \mathbf{\Gamma} \mathbf{S}_{YR}, \mathbf{S}_{RR})$. The simultaneous diagonalization is done by finding an intermediate transformation \mathbf{T}' that transforms \mathbf{S}_{RR} into the identity matrix (a whitening step). \mathbf{T}' can be found from the eigenvalue decomposition of $\mathbf{S}_{RR} = \mathbf{F} \mathbf{\Theta} \mathbf{F}^H$, such that $\mathbf{T}' = \mathbf{F}_r \mathbf{\Theta}_r^{-1/2}$, taking only the first $r = \text{rank}(\mathbf{S}_{RR})$ eigenvalues and eigen vectors, to get rid of the singularity problem by restricting the diagonalization to the column space of \mathbf{S}_{RR} . The second step is to find the orthogonal transformation \mathbf{T}'' that diagonalizes the matrix $\mathbf{A} = (\mathbf{T}')^H \mathbf{S}_{RY} \mathbf{\Gamma} \mathbf{S}_{YR} \mathbf{T}'$. This can also be achieved through the eigenvalue decomposition of the matrix $\mathbf{A} = \mathbf{Q} \mathbf{\Theta}' \mathbf{Q}^H$, such that $\mathbf{T}'' = \mathbf{Q}$, hence $\mathbf{T} = \mathbf{T}' \mathbf{T}''$ and $\mathbf{\Lambda} = \mathbf{\Theta}'$, and due to the orthonormality of \mathbf{T}'' it will not affect the whitening effect that \mathbf{T}' has on \mathbf{S}_{RR} .

3.2 The expected supremacy of the technique (references with correlated noises)

When noises, that are inter-correlated among each other but uncorrelated with the sources, exist on the references (i.e. $\epsilon > 0$ in Eq.(3))

$$\mathbf{S}_{RR} = \tilde{\mathbf{S}}_{RR} + \mathbf{S}_{EE} \quad (10)$$

where $\tilde{\mathbf{S}}_{RR}$ represents the noise free part, and \mathbf{S}_{EE} is the cross-spectral matrix of the noises on the references, in a matrix form

$$\mathbf{S}_{RR} = \begin{pmatrix} s_{1,1} + \sigma_{1,1}^2 & s_{1,2} + \sigma_{1,2}^2 & \cdots & s_{1,R} + \sigma_{1,R}^2 \\ s_{2,1} + \sigma_{2,1}^2 & s_{2,2} + \sigma_{2,2}^2 & \cdots & s_{2,R} + \sigma_{2,R}^2 \\ \vdots & \vdots & \ddots & \vdots \\ s_{R,1} + \sigma_{R,1}^2 & s_{R,2} + \sigma_{R,2}^2 & \cdots & s_{R,R} + \sigma_{R,R}^2 \end{pmatrix} \quad (11)$$

Replacing $\tilde{\mathbf{S}}_{RR}$ with $\mathbf{L} \mathbf{S}_{SS} \mathbf{L}^H$, where \mathbf{S}_{SS} is the sources' cross-spectral matrix, in Equation (10) one gets

$$\mathbf{S}_{RR} = \mathbf{L} \mathbf{S}_{SS} \mathbf{L}^H + \mathbf{S}_{EE} \quad (12)$$

keeping in mind that the sources are uncorrelated, therefore \mathbf{S}_{SS} is diagonal. Moreover, even though the diagonal elements would be of different values, they can be absorbed into the transfer function \mathbf{L} ; hence, and without any loss of generality, \mathbf{S}_{SS} can be taken as the identity matrix. Then,

$$\mathbf{S}_{RR} = \mathbf{L} \mathbf{L}^H + \mathbf{S}_{EE} \quad (13)$$

Now let's consider the very special case when \mathbf{S}_{EE} is proportional to the identity, i.e. $\mathbf{S}_{EE} = \sigma_E^2 \mathbf{I}$, hereafter referred to as "scalar noise", then the previous equation becomes

$$\mathbf{S}_{RR} = \mathbf{L} \mathbf{L}^H + \sigma_E^2 \mathbf{I} \quad (14)$$

Replacing \mathbf{L} with its singular value decomposition $\mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$, where $\mathbf{U} \mathbf{U}^H = \mathbf{I}$, $\mathbf{V}^H \mathbf{V} = \mathbf{I}$, and $\mathbf{\Sigma}$ is diagonal. Also, replacing \mathbf{I} in the second term with $\mathbf{U} \mathbf{U}^H$, one gets

$$\begin{aligned}
\mathbf{S}_{RR} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H\mathbf{V}\mathbf{\Sigma}\mathbf{U}^H + \sigma_E^2\mathbf{U}\mathbf{U}^H \\
&= \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^H + \sigma_E^2\mathbf{U}\mathbf{U}^H \\
&= \mathbf{U}(\mathbf{\Sigma}^2 + \sigma_E^2\mathbf{I})\mathbf{U}^H
\end{aligned} \tag{15}$$

Which is recognized as the eigenvalue decomposition of \mathbf{S}_{RR} , where the variation among the eigenvalues is coming only from the noise free term $\tilde{\mathbf{S}}_{RR}$, thus only from the sources. That is why the traditional PCA (the truncated singular value decomposition (TSVD)) exclusively works fine in this very special situation of $\mathbf{S}_{EE} = \sigma_E^2\mathbf{I}$.

In other words, for the conventional PCA technique to work well, the first S eigenvalues of \mathbf{S}_{RR} has to reflect only the variances due to the sources. This requires two conditions to be met:

- All the off-diagonal noise elements $\sigma_{i,j}^2 = 0$ for all $i \neq j$.
- A constant noise power on all the references, i.e. all the noise diagonal elements are equal, i.e. $\sigma_{i,i}^2 = C$ for all i .

The first condition can be theoretically met, if it is assumed that the noises on all the references are completely uncorrelated with each other and with the sources, and that an infinite amount of averaging is conducted, the off-diagonal noise-related cross-spectral terms vanish. However, this can not be guaranteed in practice. Moreover, if it is assumed in certain situation that the first condition is met, the second condition is even harder to meet. The situation even gets worse as the noises level increases (i.e. $\epsilon \gg 0$). Hence, the proposed method provides a good alternative to the conventional PCA technique through finding an alternative reference set, which is a linear combinations of the measured one, that maximally correlates to the output. In this way, the first S eigenvalues of the GEVD ($\mathbf{S}_{RY}\mathbf{\Gamma}\mathbf{S}_{YR}, \mathbf{S}_{RR}$) approximates the variances due to the sources as close as possible.

4 Numerical validation

Two simulated experiments were carried out to verify the effectiveness of the proposed MCR method. In the two experiments, the weight matrix was set as $\mathbf{\Gamma} = \mathbf{S}_{YY}^{-1}$.

4.1 Simulation 1 (references with scalar noise)

The first experiment was simulated to compare the results of the MCR method to the results of both the LSS specified in Eq. 5, which uses the full \mathbf{S}_{RR} matrix, and the TSVD, which decomposes the \mathbf{S}_{RR} and then keep only the important eigenvalues (and their corresponding eigenvectors), then substitute for \mathbf{S}_{RR} in Eq. 5. For this experiment the number of the preserved general eigenvalues in the MCR method and the number of preserved eigenvalues in the TSVD both were equally set to the number of simulated sources (S). The relative error in the estimated coherent output defined as $e = \|\hat{\mathbf{X}} - \mathbf{X}\|_F^2 / \|\mathbf{X}\|_F^2$ was used in comparing the results of the different methods. In the simulation of the first experiment, the number of sources was $S = 3$, the number of references $R = 20$, the number of the output sensors was $M = 100$, and the number of snapshots I was set as 10,000. The sources were simulated as an iid complex Gaussian random processes with zero mean and unit variance, and were mapped to the output sensors and to the references using a complex-valued transfer matrix $\mathbf{H} \in \mathbb{C}^{(R+M) \times N}$, whose elements was randomly drawn from an iid complex Gaussian random processes with zero mean and unit variance. Uncorrelated noises were added to the output sensors in such a way that the SNR is the same for all the output sensors, which means that the noise variance was different from a sensor to the other. Contrarily, uncorrelated noises were added to the references in such a way that the noise variance was the same on each reference, which means that the SNR varied from one reference to the other. The uncorrelated noises on the output sensors were simulated as random complex Gaussian variables with zero mean, and their variances were specified so as to give output sensors' signal to noise ratio (SNR) of 0 dB, $\text{SNR}_{out} = 0$ dB, on each output signal, what means that 50 % of each output

signal is noise. The simulation was iterated for reference SNRs values that approximately varied in the range $\text{SNR}_{ref} = [-48.7 \pm 4.6, 151.3 \pm 4.6]$ dB, for each iteration the noises on the references was uncorrelated random complex Gaussian variables with zero mean, and their variances were specified so as to give the reference SNR corresponding to the ongoing iteration.

Figure 1.a shows the mean relative error in the estimated coherent outputs among all output sensors as a function of the average SNR of the references. Obviously, the LLS gives inaccurate estimation results when SNR_{ref} is high, because \mathbf{S}_{RR} becomes very ill-conditioned in this range, while the TSVD yields considerably more accurate results. When SNR_{ref} decreases, both methods returns the same estimation accuracy with higher error value. It is easily noticed that the results of the proposed MCR method are nearly identical to that of the TSVD. It is concluded that the proposed MCR method represents a robust solution to the separation problem whatever the value of SNR of the references was, in the sense that it is guaranteed to provide as good results as the best of the other two methods. Another conclusion is that, despite its robustness, the MCR has not yet provided an improvement over the TSVD method in the case of scalar noise on the references, as nearly there is no difference between their estimation errors along the SNR_{ref} range; and hence, the TSVD suffices. This conclusion can also be drawn by looking at Fig. 1.b, which shows the estimated SNR of the outputs using each of the three method along with the simulated true $\text{SNR}_{out} = 0$ dB.

4.2 Simulation 2 (references with correlated noise)

In the second experiment, all the simulation parameters are kept as they were in the first experiment except from the noise on the references. Correlated noise with cross-correlation of 0.9 have been imposed among the noises on the references, i.e. the references' noise cross-spectral matrix is no more diagonal. In contrary to the first experiment, these noises were added to the references in such a way that the SNR is the same for all the references, thus the variance of the noise was allowed to vary from one reference to the other. The simulation was iterated for reference SNRs values that varied in the range $\text{SNR}_{ref} = [-50, 150]$ dB, for each iteration the variances of the noise on the references were specified so as to give the reference SNR corresponding to the ongoing iteration.

As seen in Fig. 2.a, although both the TSVD and the MCR methods equivalently fix the flaw in the LSS caused by the ill-condition of \mathbf{S}_{RR} at the high SNR_{ref} range, the TSVD returns very high estimation error for approximately $\text{SNR}_{ref} < 30$ dB. This failure of the TSVD is due to the cross-correlation between the noises on the references as discussed in Section 3.2. On the other hand, the MCR method shows robust results that are better than those of the two other methods with estimation error that does not exceed 0.01 in the worst case. By taking a look at Fig. 2.b, one finds out that by using the MCR method, very good separation results were achieved in the approximate range of $\text{SNR}_{ref} = [5, 30]$ dB, where the TSVD was not valid at all, this with keeping on being able to fix the ill-condition problem of \mathbf{S}_{RR} , at the high SNR_{ref} range, as good as the TSVD does.

5 Selection of the number of sources

Prior knowledge of the number of of true sources (S) is needed for both the proposed MCR method and the TSVD, because in both methods only the largest S eigenvalues are preserved. In practice, the number of sources is rarely known and must be estimated from the data. One candidate to solve this problem could be the likelihood ratio test (LRT), which is usually performed to compare the fit of nested models based on Wilks' theorem. However, Wilks' theorem, which states that the LRT statistic asymptotically follows a χ^2 -distribution with degrees of freedom equal to the difference in the number of free parameters between the two nested models under comparison, showed often not to hold for models with latent variables such as factor analysis [14]. An alternative candidate to solve this problem is parametric bootstrap, where a large number of data sets are bootstrapped under the restricted model, and the corresponding LRT statistics is computed, yielding a distribution for the LRT statistic. The parametric bootstrap have the merit of better approximation for finite sample data than the asymptotic distributions. Hence, the parametric bootstrap approach is adopted in this work, and is detailed in the following subsections.

5.1 Calculation of the log-likelihood ratio

One assumes the presence of K sources and wants to test whether a model with $K + 1$ is more likely. When the data \mathbf{Y} follows a multivariate complex Gaussian with mean $\hat{\mathbf{X}}(K)$ (resp. $\hat{\mathbf{X}}(K + 1)$), as returned by Eq. (6) with \hat{S} components involved, and covariance matrix $\mathbf{\Omega}(K)$ (resp. $\mathbf{\Omega}(K + 1)$) under the first (resp. second) model, then the log-likelihood ratio of the second model (of $K + 1$ factors) to the first model (of K factors) reads

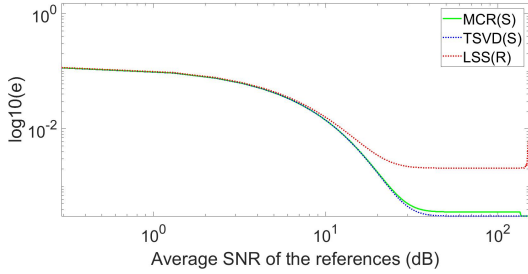
$$\Lambda(K) = -\|\mathbf{Y} - \hat{\mathbf{X}}(K + 1)\|_{\mathbf{\Omega}(K+1)}^2 + \|\mathbf{Y} - \hat{\mathbf{X}}(K)\|_{\mathbf{\Omega}(K)}^2 - I \ln |\mathbf{\Omega}(K + 1)| + I \ln |\mathbf{\Omega}(K)| \quad (16)$$

where $\|\mathbf{A}\|_{\mathbf{\Omega}}^2 = \text{tr}\{\mathbf{A}^H \mathbf{\Omega}^{-1} \mathbf{A}\}$. The expression of the above log-likelihood ratio could be further simplified, assuming a full covariance matrix. In this case, the covariance matrix entering Eq. (16) is taken as the estimate

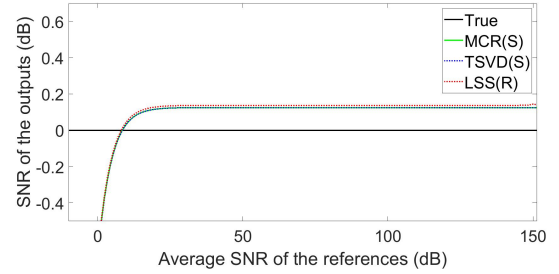
$$\mathbf{\Omega}(K) = (\mathbf{Y} - \hat{\mathbf{X}}(K))(\mathbf{Y} - \hat{\mathbf{X}}(K))^H / I. \quad (17)$$

Then, it follows that

$$\Lambda(K) = I(\ln |\mathbf{\Omega}(K)| - \ln |\mathbf{\Omega}(K + 1)|). \quad (18)$$



(a) Mean relative error among all output sensors vs. SNR_{ref} .



(b) MCR-estimated SNR for a single output sensor vs. SNR_{ref} .

Figure 1: Results of simulation 1 (references with uncorrelated noises).

The above expression might still be difficult to evaluate numerically, because the computation of the logarithm of the determinant of a large matrix tends to be unstable. This can be fixed by noting that

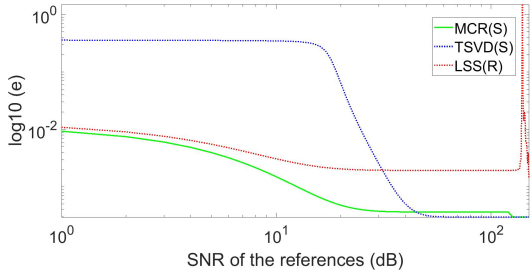
$$\mathbf{\Omega}(K + 1) = \mathbf{S}_{YY} - \sum_{k=1}^{K+1} \mathbf{S}_{YR} \mathbf{b}_k \mathbf{b}_k^H \mathbf{S}_{RY} = \mathbf{\Omega}(K) - \mathbf{S}_{YR} \mathbf{b}_{K+1} \mathbf{b}_{K+1}^H \mathbf{S}_{RY}. \quad (19)$$

Setting $\mathbf{t} = \mathbf{S}_{YR} \mathbf{b}_{K+1}$ and using the determinant lemma,

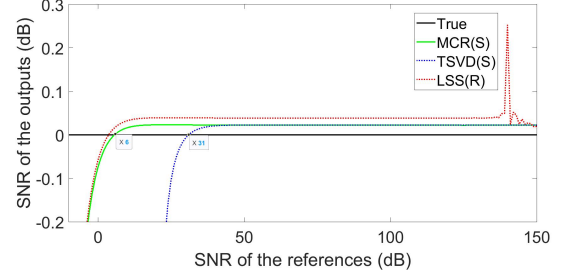
$$|\mathbf{\Omega}(K + 1)| = |\mathbf{\Omega}(K)| \cdot (1 - \mathbf{t}^H \mathbf{\Omega}(K)^{-1} \mathbf{t}) \quad (20)$$

and so

$$\Lambda(K) = -I \ln(1 - \mathbf{t}^H \mathbf{\Omega}(K)^{-1} \mathbf{t}). \quad (21)$$



(a) Mean relative error among all output sensors vs. SNR_{ref} .



(b) MCR-estimated SNR for a single output sensor vs. SNR_{ref} .

Figure 2: Results of simulation 2 (references with correlated noises).

5.2 The bootstrap algorithm

The above quantity has a reference distribution, say $D(K)$, that one can obtain using parametric bootstrap under the restricted model (i.e. the one with K sources), by estimating $\mathbf{\Omega}(K)$ as in Eq. 17, and then synthesis of I independent vectors of M complex random variables, with zero means and imposed covariance matrix $\mathbf{\Omega}(K)$. This is done by computing the singular value decomposition of $\mathbf{\Omega}(K) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, then compute the new complex noise signals as $\mathbf{N}_{new} = \mathbf{U}\mathbf{\Sigma}^{1/2}\mathbf{N}_1$, where the elements of \mathbf{N}_1 is drawn from a complex Gaussian distribution with zero mean and a unit variance. A new set of references \mathbf{R}_{new} is also synthesized in the same way using the \mathbf{S}_{RR} matrix. Finally the new output data matrix is computed as $\mathbf{Y}_n = \mathbf{C}(K)\mathbf{B}(K)^H\mathbf{R}_{new} + \mathbf{N}_{new}$. Hence, one accepts the new model with $K + 1$ sources if $\Lambda(K) > D(K)_{1-\alpha}$, with $D(K)_{1-\alpha}$ the percentile of the bootstrapped distribution associated with probability $1 - \alpha$. If so, the next model with $K + 2$ is tested; if not, K sources are selected.

The algorithm for the selection of the number of sources reads:

- compute \mathbf{S}_{YY} , \mathbf{S}_{RR} and \mathbf{S}_{YR}
- Set $K = 0$
- WHILE $K \leq R$
 - IF $K = R$, $S \leftarrow R$,
 - ELSE
 - * Set $itr = 1$
 - * WHILE $itr \leq NOI$ (Number of Iteration)
 - estimate $\mathbf{C}(K)\mathbf{B}(K)^H$
 - estimate $\hat{\mathbf{X}}(K)$ and $\mathbf{\Omega}(K)$
 - Draw \mathbf{R}_{new} using \mathbf{S}_{RR}
 - Draw \mathbf{N}_{new} using $\mathbf{\Omega}(K)$
 - compute $\mathbf{Y}_n = \mathbf{C}(K)\mathbf{B}(K)^H\mathbf{R}_{new} + \mathbf{N}_{new}$
 - estimate $\mathbf{B}(K + 1)^H$
 - compute $\mathbf{t}(K + 1) = \mathbf{S}_{YR}\mathbf{b}(K + 1)_{K+1}$
 - compute $\Lambda(K)_{new}$
 - Set $D(K)^{(itr)} = \Lambda(K)_{new}$
 - * IF $\Lambda(K) < D(K)_{1-\alpha}$ STOP
 - * $S \leftarrow K$
 - * ELSE $K \leftarrow K + 1$

6 Experimental validation

Data measured on an electric motor from SIEMENS were used for validating the proposed algorithm. Along with, numerical simulations that mimicked the experimental configuration.

6.1 Description of SIEMENS' experimental setup and the data

The sound emitted from the electric motor was measured using a moving array with 50 microphones. The complete hologram, at 2 cm from the engine sides, was totally captured by nonsynchronously placing the array at eight positions (2/each side). Another two holograms were captured by repeating the experiment at distances 4 and 6 cm. In this analysis, only position 3 is considered, which is a part of the hologram at 2 cm from the engine sides, as shown in Fig. 3.a. Five fixed microphones were used as references. The sampling frequency of the microphone signals was 40960 Hz. Data from two tri-axial accelerometers (6 tracks) were also used as references; hence, there exists a total of 11 references. The sampling frequency of the acceleration tracks was 10240 Hz, which is only one-fourth the sampling frequency of the microphone signals, hence they needed to be up-sampled by factor of 4. White gaussian noise was added to every acceleration signal with SNR of 50 dB to fill the acceleration spectra after its initial Nyquist frequency (i.e. its Nyquist frequency prior to up-sampling = 5120 Hz). Guided by the Tacho signal shown in Fig. 3.b, the non-stationary parts at the beginning and end of the records were removed. The trimmed time records have a signal length of approximately 14.6 second, which corresponds to 600,000 samples. To enable incorporating the acceleration signals into the analysis along with the microphone (pressure) signals, the amplitude spectra of all the signals were normalized through dividing by the corresponding RMS values. Frequency resolutions of 2.5 Hz were obtained using window lengths of 2^{14} samples, when computing the power spectra with Welch's averaged periodogram method, which corresponds to an effective number of snapshots of 74.

6.2 Simulation 3: numerical verification

In order to verify the algorithm before application on the real SIEMENS data, a numerical simulation was performed so as to mimic the measurement configuration of the data, i.e. using the same number of sensors (microphones), $M = 50$, the same number of references, $R = 11$, and the same number of snapshots, $I = 74$. The number of sources S was needed for simulating the data, therefore it was assumed that $S = 4$. As in the numerical simulations in Section 4, the sources and the transfer functions were randomly drawn from an iid complex Gaussian random processes with zero mean and unit variance. The noises were simulated as random complex Gaussian variables with zero mean, and their variances were specified so as to give output sensors' $\text{SNR}_{out} = 0$ dB, on each output signal, and to give references' $\text{SNR}_{ref} = 30$ dB, on each reference signal. The simulation was repeated twice, the first time using uncorrelated noises on the references, and the second time using 90% cross-correlated noises on the references. A 99% confidence level was used in the statistical test, i.e. $\alpha = 0.01$. The results obtained while using uncorrelated and correlated noises on the references turned out to be very similar; hence, only one of them is shown in Fig. 4. The figure consists of ten sub-figures, each one is representing a statistical test, for $K = 0, \dots, 10$. For example, Fig. 4.a displays the bootstrapped distribution of the log-likelihood ratio $D(K = 0)$ obtained under the restricted model with $K = 0$ source; hence, it can be used for comparing a model with $K + 1 = 1$ source to a model without sources $K = 0$, i.e. a model which describes data that contains only noise, what is indicated in the figure captions by $Model(1)/Model(0)$. Similarly, Fig. 4.b displays $D(1)$ and can be used to compare a model with $K = 2$ sources, $Model(2)$, to a model with only $K = 1$ source, $Model(1)$, ...etc. The vertical red dotted line is the current value of $\Lambda(K)$ to be tested, such that, one accepts the new model with $K + 1$ sources if $\Lambda(K) > D(K)_{1-\alpha}$, and then the next model with $K + 2$ is tested. It is obvious that, $\Lambda(K) \gg D(K)_{1-\alpha}$ up to Fig. 4.d; which represents the results of testing a model with 4 sources against the one with only 3 sources. However, as seen in Fig. 4.e that, the value of $\Lambda(4)$ is approximately in the middle of the distribution $D(4)$, i.e. $\Lambda(4) \ll D(4)_{1-\alpha}$. Hence, a model with 5 sources is very unlikely. Based on the previous discussion a model with 5 sources is rejected, therefore the estimated number of sources $\hat{S} = 4$. Consequently, the algorithm proved to be effective as it could correctly estimate the simulated true number of sources, i.e. it

returned $\hat{S} = S = 4$. It also showed a very good amount of accuracy as we can see from the figures that we could have got the same correct results using a lower confidence level, say 95%. Bearing in mind that these excellent results were obtained using a very limited number of snapshots (74 snapshots) compared to the number of sensors (50 output sensors).

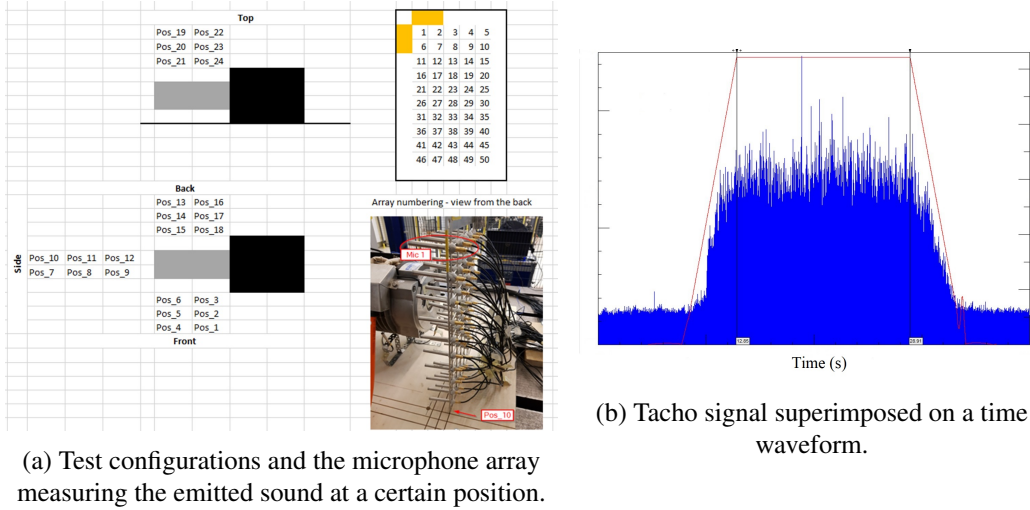


Figure 3: SIEMENS e-motor setup and data.

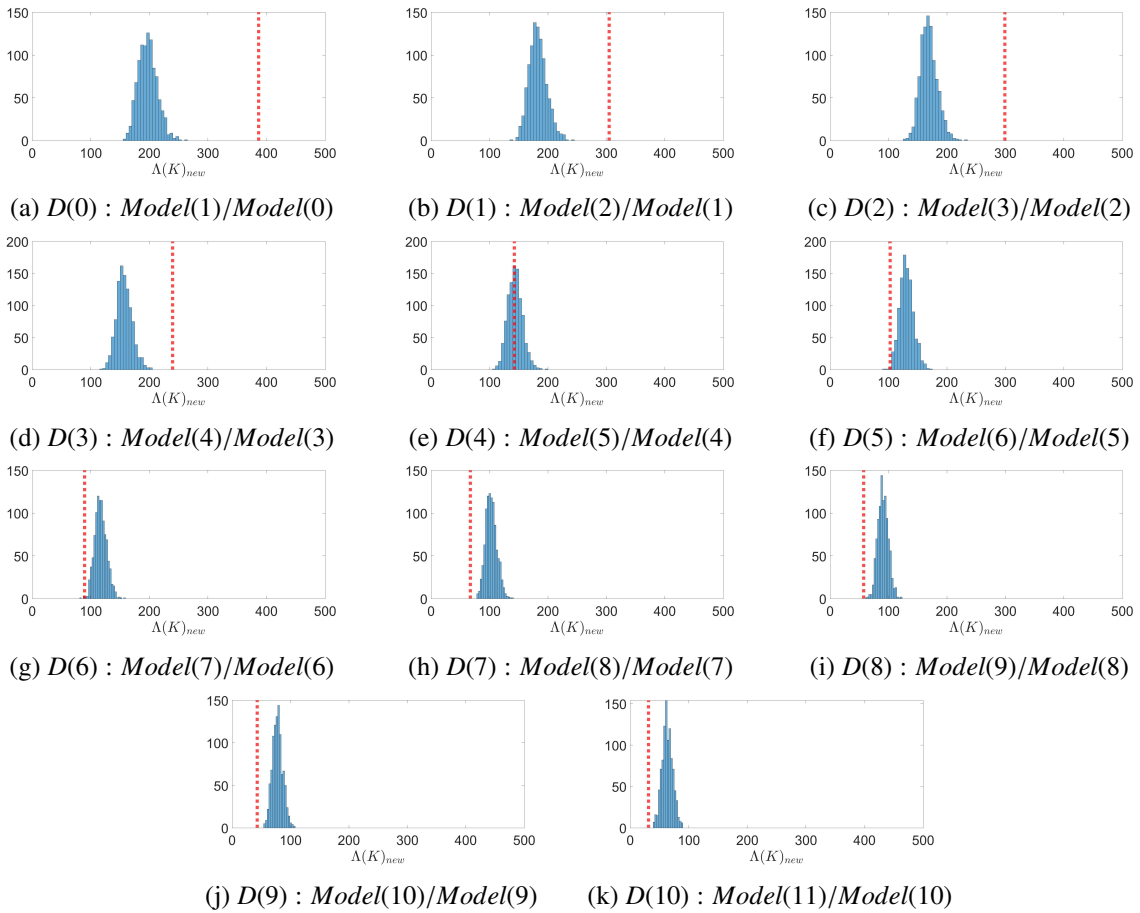
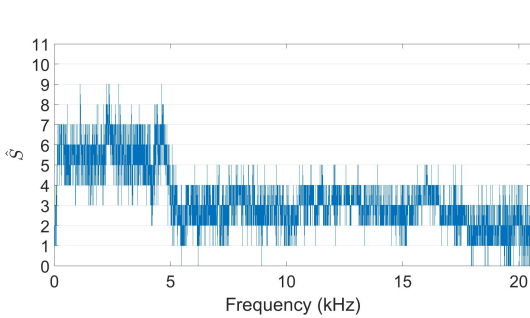


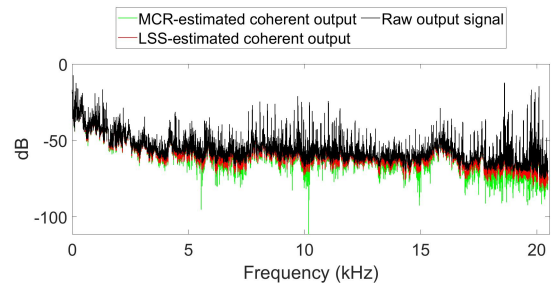
Figure 4: Results of simulation 3 (references with correlated noises); Comparing $\Lambda(K)$ (red dotted line) to the bootstrapped distribution $D(K)$ (blue bar-chart) for the 11 virtual references. The labels $Model(K + 1)/Model(K)$ indicate the two models in comparison.

6.3 Application to the real data

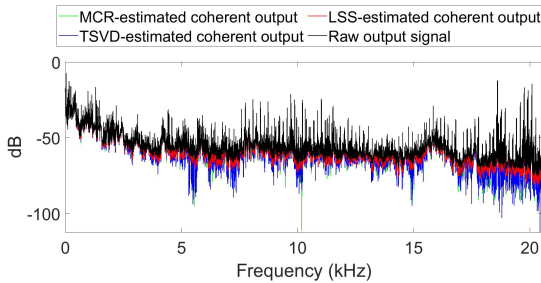
The parametric bootstrap algorithm in section 5.2 is now illustrated on the SIEMENS e-motor data described in section 6.1. The estimated number of sources \hat{S} is displayed in Fig. 5.a as a function of frequency. From the first sight, the effectiveness of the proposed method is obvious, as one easily notices the sudden drop in \hat{S} above around 5 kHz, which is the initial Nyquist frequency (i.e. before upsampling) of the acceleration tracks ($10240/2 = 5120$ Hz). This sudden drop is due to the fact that, above this frequency the acceleration tracks no longer contain information, but only numerically added white noise, which is by definition uncorrelated with the true sources. Another evidence of the effectiveness of the proposed method is that \hat{S} is systematically less than the number of physical references ($R = 11$): $\hat{S} \leq 9$ in the frequency range below 5 kHz, and $\hat{S} \leq 5$ in the frequency range above 5 kHz, bearing in mind that in theory \hat{S} can not exceed 5 in this frequency range (i.e. the 5 microphones). The few frequencies, that have $\hat{S} = 0$, correspond to very low output SNR, as was revealed by investigating the power spectral densities (PSD) of the output of the microphone array. Figure 5.b compares the PSD of the separation result of the MCR method (in green) to that of the LSS method (in red), both are superimposed on the PSD of the raw output signal (in black), the results are shown only for the output microphone 1 of the array position 3. It is clearly seen in this figure that the LSS method, as expected, estimates coherent power that is higher than that is estimated using the MCR method, this might be because of the tendency of LSS to overestimate the coherent output in the case of an ill-conditioned reference cross-spectral matrix. The result of the TSVD is added to the comparison in Fig. 5.c. By careful inspection, it is witnessed that, for most frequencies, the TSVD-estimated coherent output (in blue) is lower in power than the MCR-estimated coherent output (in green). A closer look at this figure is depicted in Fig. 5.d, where it is easier to see that the MCR solution lies between the LSS and the TSVD solution, what conforms with the simulation results depicted in Fig. 2.b. This conformity further confirm the effectiveness of the proposed solution in the case of real data.



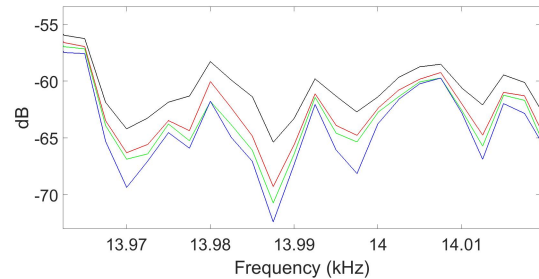
(a) The estimated number of sources in SIEMENS data.



(b) PSD of the estimated coherent output using the MCR method and the LSS method, superimposed on the PSD of raw signal.



(c) PSD of the estimated coherent output using the MCR method, the LSS method and the TSVD, superimposed on the PSD of raw signal.



(d) The results in Fig. 5.c zoomed on

Figure 5: Source separation results of output microphone 1 in array position 3 of SIEMENS data.

7 Conclusions

Reference-based source separation techniques require a number of references at least equal to the actual number of sources, and strictly larger if the references contain noise. However, when the number of references largely exceeds the number of sources, it paradoxically jeopardizes the separation because the reference cross spectral matrix becomes ill-conditioned. This is usually solved in the literature by reducing the dimension of the reference set, for instance using the virtual coherence or the partial coherence techniques. The main difficulty of these techniques is that the resultant virtual references are not necessarily maximally correlated with the sources, and thus not optimal to solve the problem. This paper introduces a solution for these problems based on finding a set of virtual references that are maximally correlated with the measurements, and hence are maximally correlated with the actual sources. A bootstrap algorithm for the selection of the number of sources is also introduced. The method is tested using numerical and real data, and it proves to solve the source separation problem while avoiding the shortcomings of the existing techniques.

Acknowledgements

We gratefully acknowledge the European commission for its support of the Marie Skłodowska Curie program through the H2020 ETN ECO DRIVE project (GA 858018). We also would like to express our special thanks to SIEMENS for providing us with the e-motor data.

References

- [1] Q. Leclère, A. Pereira, C. Bailly, J. Antoni, and C. Picard, “A unified formalism for acoustic imaging based on microphone array measurements,” *International Journal of Aeroacoustics*, vol. 16, no. 4-5, pp. 431–456, 2017. [Online]. Available: <https://doi.org/10.1177/1475472X17718883>
- [2] B. Dong, “Spatial separation of sound sources,” Theses, INSA de Lyon, Apr. 2014. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-01175498>
- [3] D. Halman and J. Bolton, “Multi-reference nearfield acoustical holography,” in *PrOC, of Inter-Noise*, vol. 92, 1992, pp. 1165–1170.
- [4] J. Hald, “STSF - a unique technique for scan-based near-field acoustic holography without restrictions on coherence,” Brüel & Kjser, Tech. Rep., 01 1989.
- [5] D. Otte, “Development and evaluation of singular value analysis methodologies for studying multivariate noise and vibration problems,” Ph.D. dissertation, KUL, 05 1994. [Online]. Available: <http://dx.doi.org/10.13140/RG.2.1.2140.2486>
- [6] J. Bendat and A. Piersol, *Random Data: Analysis and Measurement Procedures*, ser. Wiley Series in Probability and Statistics. Wiley, 2011. [Online]. Available: <https://books.google.fr/books?id=qYSViFRNMIwC>
- [7] S. Price and R. Bernhard, “Virtual coherence: A digital signal processing technique for incoherent source identification,” in *Proceedings of IMAC*, vol. 4, 1986, pp. 3–6.
- [8] Q. Leclère, “Multi-channel spectral analysis of multi-pass acquisition measurements,” *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1415–1422, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0888327008003142>
- [9] B.-K. Bae and K.-J. Kim, “A Hilbert transform approach in source identification via multiple-input single-output modeling for correlated inputs,” *Mechanical Systems and Signal Processing*, vol. 12, no. 4, pp. 501–513, 1998. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0888327097901527>

- [10] J.-F. Li, J.-C. Pascal, and C. Carles, "Reconstruction of partially coherent sources by use of principal component analysis," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 1995, no. 1. Institute of Noise Control Engineering, 1995, pp. 1355–1358.
- [11] Q. Leclère, A. Dinsenymer, J. Antoni, E. Julliard, and A. Pintado-Peño, "Thresholded multiple coherence as a tool for source separation and denoising: Theory and aeroacoustic applications," *Applied Acoustics*, vol. 178, p. 108021, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0003682X21001146>
- [12] G. Golub and C. Van Loan, *Matrix Computations*, ser. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013.
- [13] T. Melzer, "SVD and its application to generalized eigenvalue problems," University of Technology, Vienna, Austria, 01 2004.
- [14] Y. Chen, I. Moustaki, and H. Zhang, "A note on likelihood ratio tests for models with latent variables," 2020. [Online]. Available: <https://arxiv.org/abs/2008.03971>

Appendix

A Nomenclature

Y	Output measurements
S	Set of sources
X	True sources' contributions
H	Transfer matrix between the sources and the measurements
N	Noises on the outputs
R	Set of references
L	Transfer matrix between the sources and the references
$\epsilon\mathbf{E}$	Noises on the references, where ϵ is a constant
$\hat{\mathbf{X}}$	Estimated sources' contributions (Coherent outputs)
G	Transfer matrix between the references and the outputs
\mathbf{S}_{YY}	Cross-spectral matrix of the outputs
\mathbf{S}_{YR}	Cross-spectral matrix between the outputs and the references
\mathbf{S}_{RY}	Cross-spectral matrix between the references and the outputs
\mathbf{S}_{RR}	Cross-spectral matrix of the references
\mathbf{S}_{SS}	Cross-spectral matrix of the sources
\mathbf{S}_{EE}	Cross-spectral matrix of the noise on the references
$\mathbf{\Omega}$	Output noise covariance matrix
K	Assumed number of sources of a model
λ	Log-likelihood ratio
$D(K)_{1-\alpha}$	Bootstrapped distribution of the log-likelihood ratio