Automatic choice of microphone array processing methods for acoustic testing

Ennes SARRADJ, Gert HEROLD, Simon JEKOSCH

Technische Universität Berlin, Germany

ABSTRACT

If microphone arrays are to be used in acoustic testing, a signal processing method must be applied to produce measurement results such as spectra and location of sources from the data. It is well know that different processing methods may produce different results, so the practical question arises which method to choose. We propose some strategies that allow for the a-priori choice of the most suitable processing method from the raw measured data. One method uses the eigenvalue spectrum of the measured cross spectral matrix while another is based on the classical beamformer output and a convolutional neural network. After estimating the apparent number of sources and the dynamic range, the best method is looked up based on the statistical analysis of a large number of synthetic test cases. The procedure is demonstrated using a practical example.
Keywords: Sound, Insulation, Transmission

1 INTRODUCTION

If microphone arrays are to be used in acoustic testing, a signal processing method must be applied to produce measurement results such as spectra and location of sources from the data. However, since many different signal processing methods are available, thought must be given also to which array data processing method to choose. Among those methods are such that are derived from beamforming, but more interestingly also such that use deconvolution approaches or that can be formulated as inverse methods (see [1] for an overview and a slightly different classification). These methods are so numerous and may have very different properties that it is not easy to understand which method(s) would be the best to use in a particular case. However, it is apparent from both theoretical study and from practical application that different methods may behave different and may yield different results. Consequently, the question which one to choose deserves some attention and was already addressed in past. Ehrenfried and Koop compared different deconvolution approaches [2] using an example test case. Data from practical array wind tunnel measurements was used for comparison of different methods in [3] and also in [4]. Comparisons based on synthesized data can be found in [5], in [6] and elsewhere. An international benchmarking effort offered multiple test cases with synthesized and measured data sets which were analyzed with different methods [7, 8]. It turns out that the quality of the results from a certain method do not just depend on the method, but also on the input data (i.e. the specific test case). This was addressed in a monte-carlo-analysis with 12600 different test cases [9]. It was found that among others the number of sources, the dynamic range, and the distance between sources are important factors. Based on this approach, the present paper discusses some strategies to decide which microphone array method to choose for a given data set from a measurement.

2 METHODS AND DATA

The basis for any decision are the 12600 synthetic test cases [9] covering different number of sources, different distances between the sources and different level differences (dynamic range) of the sources. For the present analysis all of the following methods were applied to this data: DAMAS with modified Gauss-Seidel solver [10], CLEAN-SC [11], Orthogonal Beamforming [12], CMF with LassoLars solver [13] and Bayesian information criterion (BIC), CMF with NNLS solver, CMF with OMP solver and cross validation.The open source Acoular package [14] was used to perform the computations. The results in each case were stored for 13 third-octave frequency bands, leading to around 1.5 million results in form of sound source mappings.

All of these results were compared to the original data set of source information from the synthetic test cases and the estimation error was stored for each case. This way, a database with 1.8 million data sets was established that allows for the assessment of the method-specific performance as a function of the frequency, the number of sources, the dynamic range, and the distances between the sources. The performance is given in form of some error measures defined in [9]:

- The overall level error $\Delta L_{p,e,o}$ is the difference between the power of the sum of all apparent sources in the map and the sum of the power of the actual sources.
- The specific level error $\Delta L_{p,e,s,i}$ is the difference between the power in the map at the location of the source i and the actual power of that source.
- The inverse level error $\Delta L_{p,e,i}$ is the difference between sum of the power at all source location and the sum of the power of the actual sources.

The latter serves as a measure for the correct location while the former two characterize the ability of the method to estimate correct levels.

Inspecting the database, it turns out that none of the methods performs always better than the other methods. Instead, one general conclusion is that sometimes the inverse methods CMF-NNLS und CMF-LassoLars give reliable results, sometimes the variants of DAMAS. In other cases however CLEAN-SC or orthogonal beamforming are better suited. If it is assumed that one and the same method shows a similar performance if the number of sources, the source distances and the dynamic range are similar, the database can be used to look up the performance of each of the methods for a certain case at hand and to decide which method to use for optimum result reliability.

However, this process is not straightforward because the necessary information is usually unknown for measured data. One solution for this problem could be to first apply one of the methods and to estimate the necessary information from the result. Another solution that was realized here is to use some other approach to estimate the number of sources and the dynamic range a priori before applying any array processing method.

Two different approaches are proposed here to do this a-priori assessment. One is based on the eigenvalue spectrum of the cross spectral matrix of microphone signals. The other approach uses a standard beamforming map and machine learning based on convolutional neural networks (CNN) to decide which processing method is preferable.

2.1 Eigenvalue method

The cross spectral matrix of the microphone signals can be assumed to be of full rank and positive semidefinite. In practice, only an estimate of this matrix is known, but these properties still hold. The eigenvalues of the cross spectral matrix can be generally divided into two groups: those belonging to signals (the associated eigenvectors span the signal subspace) and those belonging to noise. Thus, the number of sources corresponds to the number of eigenvalues in the signal subspace.

In a practical estimate of the cross spectral matrix from measurements it is not easy to decide which eigenvalues belong to the signal subspace. However, as the noise subspace eigenvalues have to be the same, this property can be used as a basis for an a-priori estimator. One simple implementation that is used here is to split the spectrum of the matrix in two equal sums. One consists of the higher-valued signal eigenvalues and the other of the noise eigenvalues and possibly some of the lower signal eigenvalues. This estimator is biased towards an underestimation of number of sources but is not sensitive to small changes in the cross spectral matrix.

Likewise, the dynamic range is also estimated from the ratio of the largest and the smallest signal eigenvalue. This estimator is also biased. However, for the present application, this seems to be acceptable. Figure 1 shows the results for all 12600 cases from the database. The estimator for the source number shows a considerable dependence on the Helmholtz number (reduced frequency) that is defined as

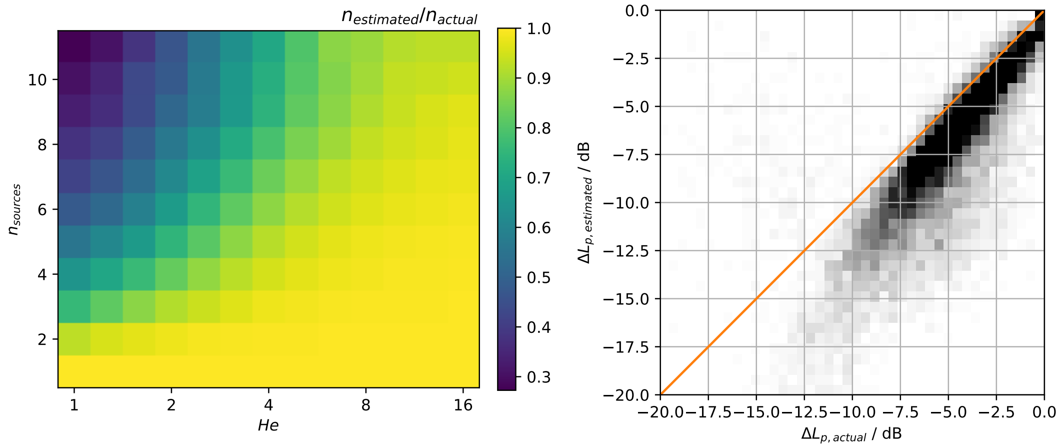


Figure 1. left: average ratio of estimated number of sources and true number of sources as a function of the true number of sources and the Helmholtz number, right: histogram of estimated dynamic range over the actual dynamic range from the 12600 data sets (absolute number of cases per bin).

$He = fd/c$ with the frequency f, the array aperture d and the speed of sound c. For low He and large number of sources the estimator tends to underestimate the number sources considerably.

Given estimates for the number of sources and the dynamic range, similar cases are looked up in the database. Then, the performance of the different methods is measured using a combined error metric

$$L_{p,g} = \sum_{n=0}^{N} \left|\Delta L_{p,e,s,n}\right| + \alpha \cdot \left|\Delta L_{p,e,i}\right|. \tag{1}$$

This includes both the specific level error and the inverse level error. In the present analysis, the weighting factor α is chosen to be 0.01, giving emphasis to the specific level error.

2.2 CNN classifier method

For many cases of practical interest, standard beamforming is not capable of producing reliable results. However, the map that can be produced without too much computational effort implicitly contains information on the number of sources and the dynamic range. Decoding this information is not straightforward because no clear criterion is known to estimate the number of sources just by looking at the map. The same is true for the dynamic range. The approach proposed here is to use image processing by a CNN that ingests the map and acts as a classifier that decides which method to use. Thus, in contrast to the eigenvalue method no lookup procedure is necessary.

A convolutional neural network processes the input (often in form of an image) in a number of consecutive layers that applied different convolution-like operations. In this analysis a network model developed by He et al. [15] is applied. This Residual Neural Network (ResNet) is a CNN which is build up of similar blocks (residual layers), that can be daisy-chained as often as needed. The network used here consists of 22 layers with about 725000 variables having an influence on the processing. The structure of the whole network is shown in Table 1. The network is implemented in Tensorflow [16]. To limit the considerable training effort, only four of the microphone array methods were included: DAMAS (Gauss-Seidel), CMF-LLBIC, orthogonal Beamforming and CleanSC.

Prior to the use of the CNN as a classifier, its variables have to be estimated. In the training process the CNN sees data for which the correct answer is known. For this process, the database was used and split

Table 1. Structure of residual neural network ResNet-v2 22 layers

Processing Block	Dimension input	Dimension output	No. kernels	Size
ConvLayer	$51 \times 51 \times 1$	$51 \times 51 \times 26$	26	3×3
Residual Layer 1	$51 \times 51 \times 26$	$26 \times 26 \times 26$	$\begin{bmatrix}26\\26\end{bmatrix} \times 3$	$\begin{bmatrix}3\times 3\\3\times 3\end{bmatrix} \times 3$
Residual Layer 2	$26 \times 26 \times 26$	$13 \times 13 \times 52$	$\begin{bmatrix}52\\52\end{bmatrix} \times 3$	$\begin{bmatrix}3\times 3\\3\times 3\end{bmatrix} \times 3$
Residual Layer 3	$13 \times 13 \times 52$	$7 \times 7 \times 104$	$\begin{bmatrix}104\\104\end{bmatrix} \times 3$	$\begin{bmatrix}3\times 3\\3\times 3\end{bmatrix} \times 3$
AvgPoolLayer	$7 \times 7 \times 104$	104×1	1	7×7
Regression Layer	104×1	64×1	-	64 nodes
Regression Layer	64×1	64×1	-	64 nodes
Output Layer	64×1	$4 \times 1(a)$	-	4 nodes

into three parts - 80% were used for training, 10% for cross validation and 10% for evaluation after the training. The training was realized using the Adam-optimizer [17]. The learning rate was set to 0.001, and the parameters were set to standard values ($\beta_1 = 0.9, \beta_2 = 0.999$ und $\varepsilon = 10^{-8}$). The softmax-crossentropy function was chosen as the loss function.

The learning result after 6000 epochs of training is shown in Figure 2. The confusion matrix for the learning data shows that in most cases, the classifier gives the correct answer. However, using the evaluation data, the result are no that good and orthogonal beamforming is often getting confused with CleanSC.

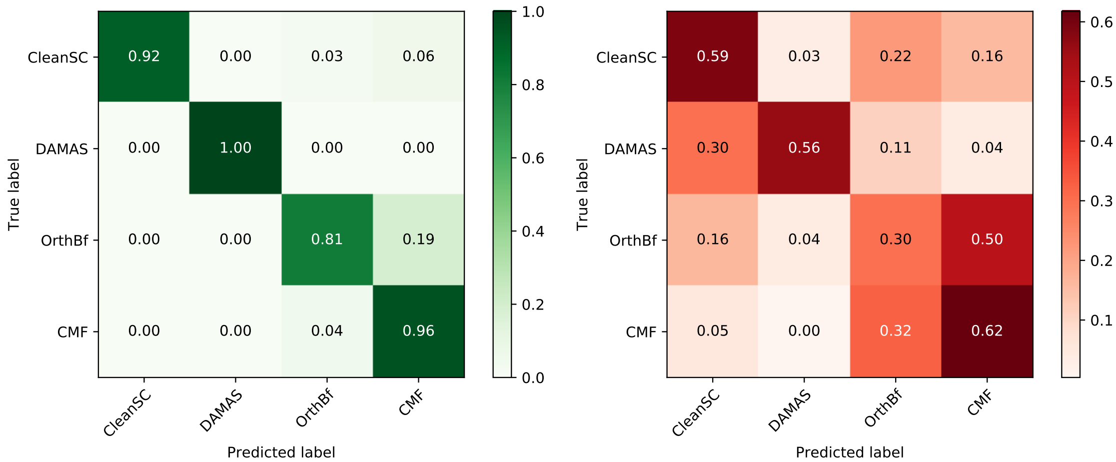


Figure 2. Confusion matrix after epoch 6000 between predicted labels and true labels for training data (left) and evaluation data (right)

3 Application example

An airfoil noise measurement in an aeroacoustic wind tunnel is used here as an application example. The setup is the same as used in [18] and consists of a NACA0012 airfoil of 28 cm span and in an open jet wind tunnel with 20 cm nozzle diameter and a flow speed of 50 m/s. The airfoil is tripped on both sides to ensure a turbulent boundary layer. The microphone array has an aperture of 1.3 m and is placed outside

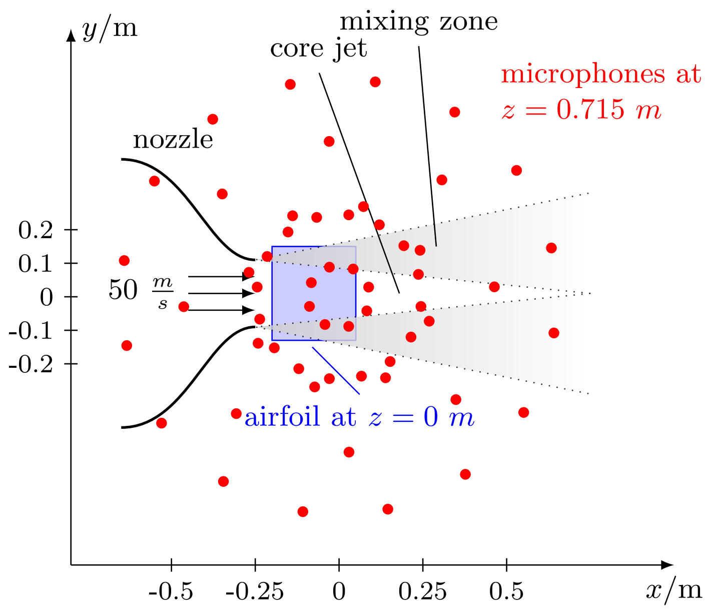


Figure 3. Schematic overview of the setup for the NACA0012 airfoil in the wind tunnel

the flow (Figure 3).

The eigenvalue method estimates the number of sources and the dynamic range as shown in Figure 4. It is interesting to note that the estimated number of sources increases with frequency, while the dynamic range stay more or less the same over the whole frequency range considered. Taking these values and look up the performance for six different methods in similar cases, scores can be computed for each of the methods. Figure 5 shows the result. Depending on frequency (Helmholtz number), different methods get the best score. For lower frequencies, mostly CMF with LassoLars and BIC is to be preferred, while at medium to high frequencies the best performance is predicted for CMF with NNLS. At very low and very high frequencies, the methods that are recommended are DAMAS and orthogonal beamforming, respectively.

The CNN based method does not predict a score for each method, but chooses one out of the available four methods that is recommended. Figure 6 shows the results. Similar to the eigenvalue method, CMF is recommended for lower and mid to high frequencies. For high frequencies, DAMAS ist recommended. This is different to what the eigenvalue method predicts.

While space constraints do not allow to show all maps for all methods, Figure 7 shows the maps for six different methods at one third octave band. It seems reasonable to follow the recommendation of both the eigenvalue method and the CNN based method and use CMF with LassoLars solver and BIC.

4 CONCLUSION

Using a database of 12600 precomputed cases, two methods were proposed recommend the most suitable array signal processing method a priori from raw data. The first method uses the eigenvalue spectrum of the cross spectral matrix to estimate the number of sources and the dynamic range. Using this estimates similar cases are looked up in the database and from this results the most suitable method is derived. It was demonstrated that this approach works in principle, but the estimates of source number and dynamic range from the cross spectral matrix are biased and may lead to wrong lookups. The second method uses the map from standard beamforming and relies on a convolutional neural network (CNN) to choose the most appropriate processing method. The CNN was trained using the database. It turned out that the

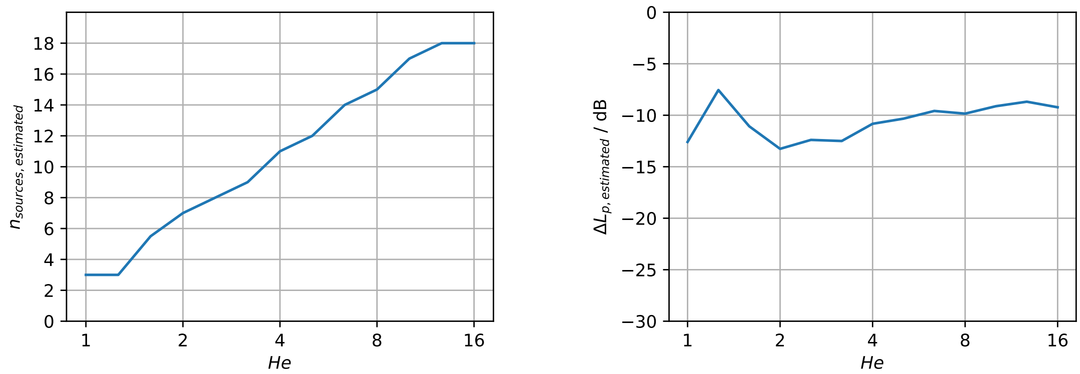


Figure 4. Estimates from the eigenvalue spectrum of the cross spectral matrix for the NACA0012 airfoil measurement and Helmholtz numbers between 1 and 16: number of sources (left) and dynamic range (right)

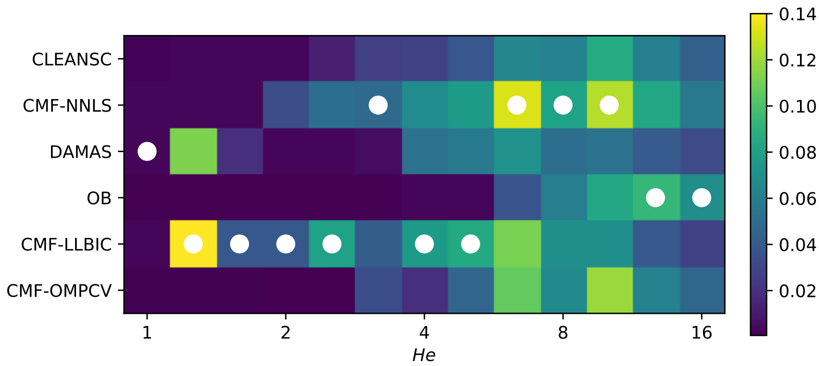


Figure 5. Estimated frequency-dependent score $(1/L_{p,g})$ from the eigenvalue based method for the NACA0012 airfoil example and for six different methods. White circles mark the best choice per Helmholtz number.

performance is good but not perfect, because sometimes it does not recommend the truly best method, but a different method. A general conclusion is that the implementation of both methods reported here works in principle acts as a proof of concept. It can be assumed that a larger database and more refined lookup and training methods will lead to an even more reliable prediction of the best method to choose.

REFERENCES

[1] Q. Leclere, A. Pereira, C. Bailly, J. Antoni, and C. Picard, "A Unified Formalism for Acoustic Imaging Techniques : Illustrations in the Frame of a Didactic Numerical Benchmark," in *Proceedings of the 6th Berlin Beamforming Conference*, pp. 1–17, 2016.

[2] K. Ehrenfried and L. Koop, "Comparison of Iterative Deconvolution Algorithms for the Mapping of Acoustic Sources," *AIAA Journal*, vol. 45, no. 7, pp. 1584–1595, 2007.

[3] E. Sarradj, "Aeroacoustic source characterisation using subspace and deconvolution techniques," in *Noise and Vibration: Emerging Methods*, 2009.

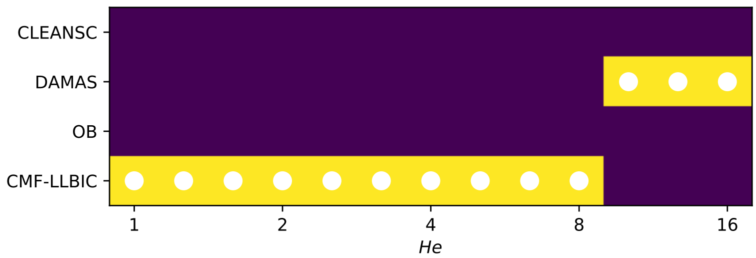
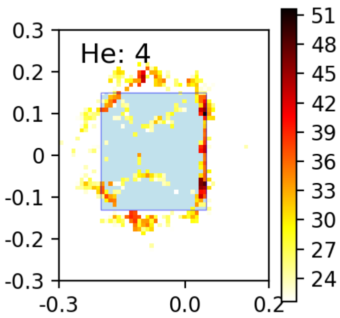
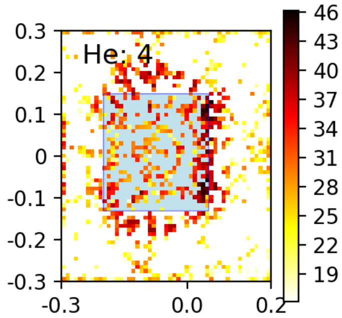


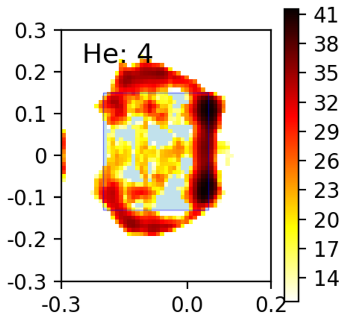
Figure 6. Estimated frequency-dependent recommendation from the CNN based method for the NACA0012 airfoil example and for four different methods. White circles mark the best choice per Helmholtz number.



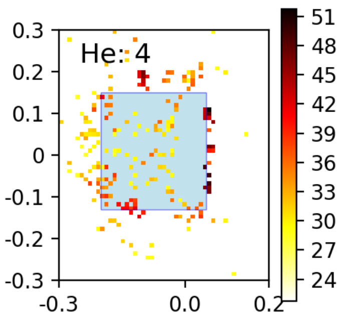
(a) Clean SC



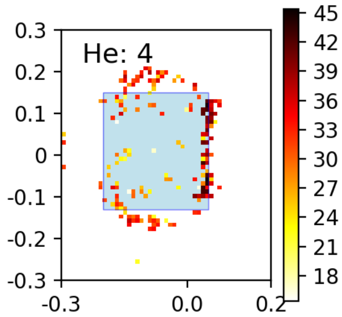
(b) CMF NNLS



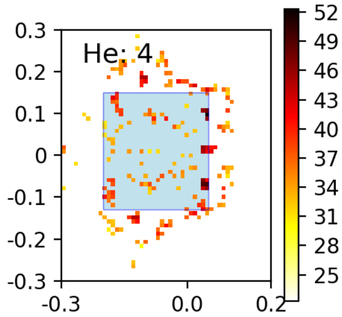
(c) DAMAS



(d) Orthogonal Beamforming



(e) CMF LassoLars BIC



(f) CMF OMP CV

Figure 7. Third-octave band sound maps for $He = 4$ for six different methods

[4] T. Yardibi, J. Li, P. Stoica, N. Zawodny, and L. Cattafesta III, “A covariance fitting approach for correlated acoustic source mapping,” *The Journal of the Acoustical Society of America*, vol. 127, p. 2920, 2010.

[5] G. Herold and E. Sarradj, “Preliminary benchmarking of microphone array methods,” in *Fifth Berlin Beamforming Conference*, 2014.

[6] Z. Chu and Y. Yang, “Comparison of deconvolution methods for the visualization of acoustic sources based on cross-spectral imaging function beamforming,” *Mechanical Systems and Signal Processing*, vol. 48, no. 1-2, pp. 404–422, 2014.

[7] C. J. Bahr, W. M. Humphreys, D. Ernst, T. Ahlefeldt, C. Spehr, A. Pereira, Q. Leclere, C. Picard, R. Porteous, D. Moreau, *et al.*, “A comparison of microphone phased array methods applied to the study of airframe noise in wind tunnel testing,” in *23rd AIAA/CEAS Aeroacoustics Conference, AIAA-paper 2017-3718*, 2017.

[8] E. Sarradj, G. Herold, P. Sijtsma, R. Merino-Martinez, T. F. Geyer, C. J. Bahr, R. Porteous, D. Moreau, and C. J. Doolan, “A microphone array method benchmarking exercise using synthesized input data,” in *23rd AIAA/CEAS Aeroacoustics Conference, AIAA-paper 3719*, 2017.

[9] G. Herold and E. Sarradj, “Performance analysis of microphone array methods,” *Journal of Sound and Vibration*, vol. 401, pp. 152–168, 2017.

[10] T. F. Brooks and W. M. Humphreys, “A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays,” AIAA paper 2004-2954, 2004.

[11] P. Sijtsma, “CLEAN based on spatial source coherence,” *International Journal of Aeroacoustics*, vol. 6, pp. 357–374, 2007.

[12] E. Sarradj, “A fast signal subspace approach for the determination of absolute levels from phased microphone array measurements,” *J. Sound Vib.*, vol. 329, pp. 1553–1569, April 2010.

[13] T. Yardibi, J. Li, P. Stoica, and L. Cattafesta III, “Sparsity constrained deconvolution approaches for acoustic source mapping,” *The Journal of the Acoustical Society of America*, vol. 123, p. 2631, 2008.

[14] E. Sarradj and G. Herold, “A python framework for microphone array data processing,” *Applied Acoustics*, vol. 116, pp. 50–58, 2017.

[15] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9908 LNCS, pp. 630–645, 2016.

[16] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, and G. Brain, “TensorFlow: A System for Large-Scale Machine Learning TensorFlow: A system for large-scale machine learning,” *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI ’16)*, pp. 265–284, 2016.

[17] J. B. Diederik P. Kingma, “Adam: A Method for Stochastic Optimization,” in *3rd International Conference for Learning Representations, San Diego, 2015*, 2015.

[18] G. Herold, T. Geyer, and E. Sarradj, “Comparison of inverse deconvolution algorithms for high-resolution aeroacoustic source characterization,” *23rd AIAA/CEAS Aeroacoustics Conference, 2017*, pp. 1–8, 2017.