BeBeC-2018-D22

A COMPARISON OF MICROPHONE ARRAY METHODS FOR THE CHARACTERIZATION OF ROTATING SOUND SOURCES

Gert Herold[1], Christof Ocker[2], Ennes Sarradj[1], Wolfram Pannert[2]
[1]Technische Universität Berlin, Einsteinufer 25, 10587 Berlin, Germany
[2]Aalen University, Beethovenstr. 1, 73430 Aalen, Germany

Abstract

The detection and quantification of rotating acoustic sources is an important challenge in many fields of application. Several approaches for compensating the rotational motion have been proposed. In this contribution, three such methods are compared: the rotating focus, the virtual rotating array, and the modal decomposition of the rotating sound field. For their evaluation, the methods are applied on two generic data sets featuring rotating sources, which were simulated with the intention to serve as benchmark for suitable array methods. Advantages and limitations of the methods are discussed, as well as possible steps in the data processing to improve the source reconstruction. Finally, the respective performance of the methods regarding the ability to detect the correct position, shape, and level of the sources is evaluated.

NOMENCLATURE

$\boldsymbol{b}$	beamformer output
f	frequency
$\boldsymbol{h}$	steering vector
L_p	sound pressure level
M	number of microphones
N	number of focus grid points
$\boldsymbol{p}$	complex sound pressures
$\boldsymbol{P}$	matrix of point spread functions
CB	Classic beamforming
CSM	Cross spectral matrix
FD	Frequency domain
MD	Modal Decomposition
PSF	Point spread function
RF	Rotating Focus
TD	Time domain
VRA	Virtual Rotating Array
SPL	Sound pressure level
SV3	Steering vector, Eq. (7)
SV4	Steering vector, Eq. (4)

1 INTRODUCTION

Microphone arrays have become a standard tool for measurements which aim at separately evaluating spatially distributed acoustic sources. The number of proposed array data processing strategies for source characterization has increased substantially in the past years. While most methods aim at reconstructing the position/extent and level of the sources as precise as possible, the degree of achieved precision depends on the chosen algorithm as well as the parameters of the posed problem [7].

Moreover, it has been shown that the performance of a given algorithm varies between individual implementations [1, 13]. In an effort to enable objective comparisons of both algorithms and their implementations, a number of research groups have set up a data base of array benchmarks, which features a number of simulated and measured test cases. The provided benchmarks differ in the complexity of their setup and sources, as well as in the explicit challenges to be handled by the array methods.

In this contribution, two simulated data sets featuring rotating point sources will be evaluated [5]. Three different methods will be tested on the data, one time-domain based approach ad two preprocessing methods enabling subsequent frequency domain beamforming.

2 THEORY

2.1 Rotational beamforming in the time domain

The basic formulation of a beamformer in the time domain (TD) for one focus point is [8]:

$$p_{\text{out}} = \sum_{m=1}^{M} h_m p_m(t + \Delta t_m) \, , \tag{1}$$

with the number of microphones M, the time delays from the microphones to the focus point $\Delta t_m = \frac{r_{t,m}}{c}$, and the steering factor

$$h_m = \frac{1}{r_{t,m} \sqrt{M \sum_{l=1}^{M} {r_{t,l}}^{-2}}} \, , \tag{2}$$

which weights the signals such that a detected maximum in a sound map coincides with the source position (SV4) [12].

If the focus point is moving, the distance from this point to the microphones is time-dependent $r_{t,m} = r_{t,m}(t)$. Thus, Δt_m and h_m have to be recalculated for each evaluated time sample according to a known trajectory. In principle, this can be applied to any moving source region. For the case at hand, only a rotation is taken into account, with all focus points rotating around a common axis. Similar approaches were used, amongst others, by Sijsma and Oerlemans [15] and Minck et al. [9].

The map resulting from Eq. (1) contains the filtered time signals. In order to compare these results with those obtained from frequency domain methods and for the application of deconvolution algorithms, sound pressure spectra are calculated using Welch's method [16].

2.2 Frequency domain beamforming

In the frequency domain (FD), the beamforming result (i.e., the squared sound pressures at one of the N focus points $\boldsymbol{x}_t$) is calculated via

$$b(\boldsymbol{x}_t) = \boldsymbol{h}^{\mathrm{H}}(\boldsymbol{x}_t)\, \boldsymbol{C}\, \boldsymbol{h}(\boldsymbol{x}_t)\,, \qquad t = 1 \ldots N\,, \tag{3}$$

with the steering vector $\boldsymbol{h}$, which contains the phase shift and the amplitude correction according to the sound travel time from a focus point to all microphones. Equivalent to Eq. (2), its entries are:

$$h_m = \frac{1}{r_{t,m}\sqrt{M \sum_{l=1}^{M} r_{t,l}{}^{-2}}}\, \mathrm{e}^{-\mathrm{j}k(r_{t,m}-r_{t,0})}\,, \qquad m = 1 \ldots M\,. \tag{4}$$

$\boldsymbol{C}$ is the cross-spectral matrix (CSM) of the microphone signals, approximated again using Welch's method [16].

As the CSM is based on stationary data, classic frequency domain beamforming can only be done with non-moving sources. In this contribution, two methods for calculating a CSM in the rotating reference frame will be compared.

2.3 Virtual rotating array

The motion compensation with a virtual rotating array (VRA) follows the description of Herold and Sarradj [6]: The angular position of the rotating object is tracked for the whole measurement time and the signals of a virtual array "rotating" at the same rate as the sources are interpolated from neighboring microphones. For the distances between focus point and virtual microphones, it has to be taken into account and compensated that in the rotating reference frame, the medium between microphones rotates itself.

This method is applicable for constant or varying rotational motion.

2.4 Modal decomposition

For the modal decomposition (MD) as described by Pannert and Maier [11] and Ocker and Pannert [10], the time signals are Fourier-transformed with a high-resolution FFT. The complex sound pressures at the equidistant microphones are then phase-shifted according to a constant rotational rate. Subsequently, the sound pressures in the rotating reference frame are cross-correlated, and averaging over a number of neighboring frequency bands yields the CSM.

As with the VRA, the rotational motion of the medium in the rotating reference frame has to be compensated. This is done via a modal decomposed Green's function, which describes the radiation from a rotating sound source in the rotating reference frame.

This method is only applicable for a constant rotational rate. For the evaluation of data with varying rotation, the time data has to be divided into small enough segments during which the rotation can be considered constant, and each segment has to be evaluated separately. Another possibility to deal with a varying rotation (not investigated in this contribution) is to resample the time data to a constant/average rotational rate as presented by Dougherty and Walker [4].

2.5 Deconvolution

For a higher spatial resolution of the final sound map, deconvolution algorithms can be applied.

A computationally efficient method is the CLEAN-SC algorithm [14]. It is based on the ideas that sidelobes caused by a source in a map are coherent to this source, and that the level at an actual source position is usually higher than at its sidelobes. Therefore, it iteratively searches for the maxima in a map, stores their value (partly) in a new map and removes coherent portions from the original map. However, for the identification of coherent sources, the algorithm necessitates the CSM; it can therefore only used with frequency domain beamforming.

For the deconvolution of maps originating from time domain beamforming, algorithms based on the evaluation of the theoretic point spread function (PSF) can be used. A commonly used method is the DAMAS algorithm [2]. This method reconstructs the correct source level distribution $\boldsymbol{b'}$ by solving the system of equations

$$\boldsymbol{b} = \boldsymbol{P}\boldsymbol{b'} \,, \tag{5}$$

where $\boldsymbol{b}$ is a vector with N entries, containing the output of Eq. (3) or (after FFT/Welch) Eq. (1). $\boldsymbol{P}$ contains the point spread functions from each focus point to all other focus points. Depending of the number of chosen focus points, the resulting system of equations can become quite large. An efficient way of solving it is with a non-negative least squares (NNLS) solver, which solves the minimization problem:

$$\arg\min_{\boldsymbol{b'}} \|\boldsymbol{b} - \boldsymbol{P}\boldsymbol{b'}\|_2 \,, \qquad \boldsymbol{b'} \geq 0 \,. \tag{6}$$

3 TEST SETUP

The evaluated data sets are part of a number of benchmark test cases, which have been put together to provide an evaluation standard for array method algorithms and their individual implementations. The data of this benchmark comprises two sub-cases, each featuring simulated rotating monopole sources [5]. The rotational rate is tracked by a tacho signal (1 trigger / revolution). For "measuring" the sound pressure time data, a circular array with 64 microphones is used. The array center is axially aligned with the axis of rotation of the sources, the distance to the source plane is $z = 0.5\,\mathrm{m}$, the array aperture is $d_{\mathrm{array}} = 1\,\mathrm{m}$.

3.1 Subcase a: One source, constant rotation

The test data consists of one monopole source, rotating clockwise at a constant rate of 1500 rpm at a radius of $r_1 = 0.25\,\mathrm{m}$. The source emits white noise, the sound pressure level at 1 m distance from the source is 94 dB. Figure 1 depicts the general setup of the data simulation.

This case is mostly intended to be used as test for whether an algorithm is capable of detecting a rotating source in principle, i.e. if the motion compensation works correctly.

3.2 Subcase b: Three sources, varying rotation

Three uncorrelated monopole sources (white noise signals) rotate clockwise at the same rate, which varies slightly over time (≈1500 rpm). The sources feature different relative positions and

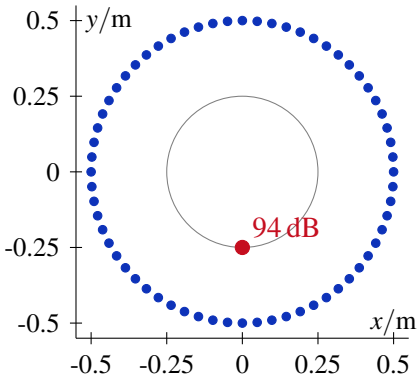

Figure 1: Subcase a, schematic of the simulated setup, microphones in blue.

levels:

- $r_1 = r_2 = 0.25\,\mathrm{m}, r_3 = 0.125\,\mathrm{m}$
- $\varphi_1 = \varphi_3 = \varphi_2 - 40°$
- $L_{p,1} = L_{p,2} + 3\,\mathrm{dB} = L_{p,3} + 6\,\mathrm{dB}$

Figure 2 schematically shows the simulated setup as well as the variation of the rotation.

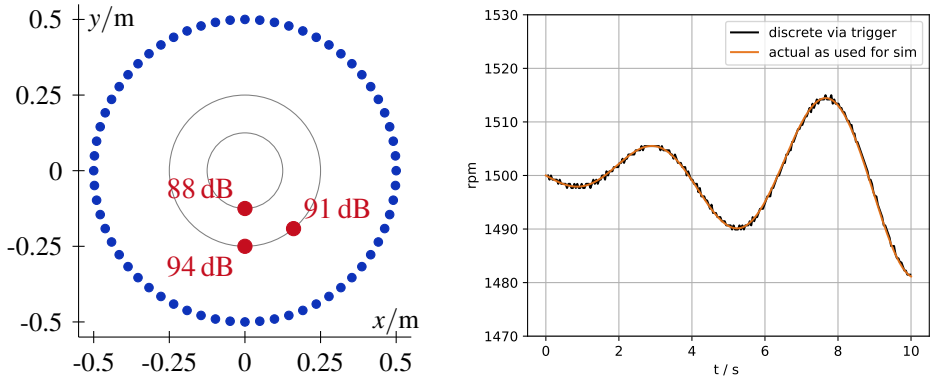

Figure 2: Subcase b. Left: schematic of the simulated setup, microphones in blue. The depicted levels represent the SPL of each source at 1 m distance of the respective source. Right: variation of the rotational rate.

The challenge here is to handle the varying rotation as well as multiple sources with different levels.

3.3 Evaluation setup

Three different motion compensation strategies are compared in this contribution: Rotating Focus (RF), Virtual Rotating Array (VRA), and Modal Decomposition (MD). The evaluation parameters are summarized in Table 1.

Table 1: Data processing parameters.

Number of microphones	64	Array to focus plane	0.5 m
Array diameter	1 m	Focus grid resolution	0.01 m
Evaluated time	10 s	Number of grid points	61×61 (RF, VRA)
Sampling rate	48 kHz		60×60 (MD)
FFT block size	1024 samples	CLEAN-SC iterations	500
FFT window	von Hann	CLEAN-SC damping	0.9
	50% overlap	DAMAS solver	NNLS

4 RESULTS

4.1 Computational effort

A direct comparison between the computational cost of the three methods proves difficult, since not all implementations were optimized for speed, and, in case of the modal decomposition method, were implemented in completely independent frameworks and run on different hardware. A rough estimate would put the calculation time of the MD method between that of the VRA and the RF method.

The rotating focus method in its current form is by far the most time-consuming, as the distances between all focus points and the microphones have to be continuously recalculated. For the calculations of the classic beamforming, the runtime of RF was about 350 times that of VRA. Furthermore, calculating a DAMAS-NNLS result (including the calculation of 47 GB worth of PSFs) took about 150 times as long as a CLEAN-SC calculation.

4.2 One rotating source

The general effect of neglecting or compensating the rotational motion of a source is illustrated in Fig. 3, which shows exemplary sound maps for the one-third octave band around 5 kHz, calculated with different data processing strategies.

In Figures 3a and 3b, the data processing chain is exactly the same but for the virtual rotating array being applied before further processing in Fig. 3b. With no motion compensation, the point source is blurred along the circumference of its rotation, its energy evenly distributed on the trajectory. Applying the VRA method, the source is well-focused; the visible sidelobes correspond to the expected point spread function of the array geometry.

Time domain beamforming and a focus grid rotating synchronously with the source yields a similar result. As can be seen in Fig. 3c, the sidelobe pattern resembles the one calculated with the VRA method and frequency domain beamforming. However, the overall appearance is not as smooth here, and additional artifacts are visible. This is due to the PSF changing depending on the relative orientation of the array microphones to the focus grid. The rotation-induced "Airy pattern" can be modeled and used as an approximation for the PSF, e.g. for the application of deconvolution algorithms like DAMAS [3]. However, since with a rotational-axis-aligned circular microphone arrangement the PSF varies only little (depending on the frequency), for the deconvolutions calculated here, the PSF of the rotating-focus result is assumed to be the same as for a stationary focus grid.

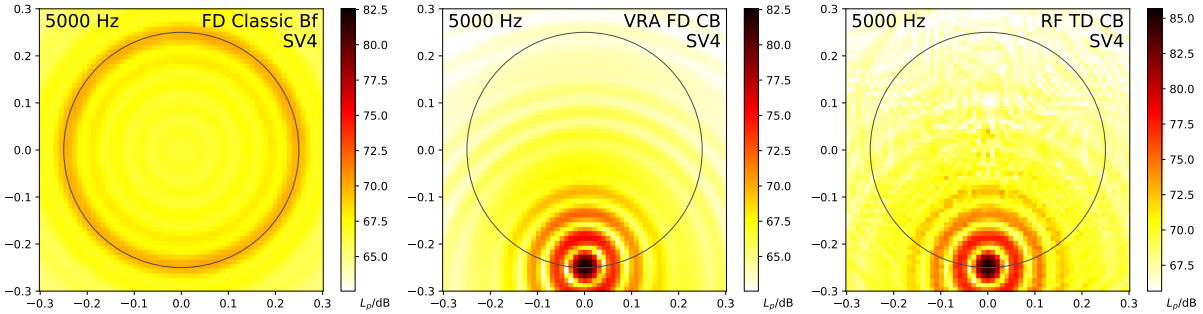

Figure 3: Classic beamforming sound maps for the 5 kHz one-third octave band and different data processing: (a) Frequency domain classic beamforming, no motion compensation. (b) Frequency domain classic beamforming, Virtual rotating array method. (c) Time domain beamforming with rotating focus.

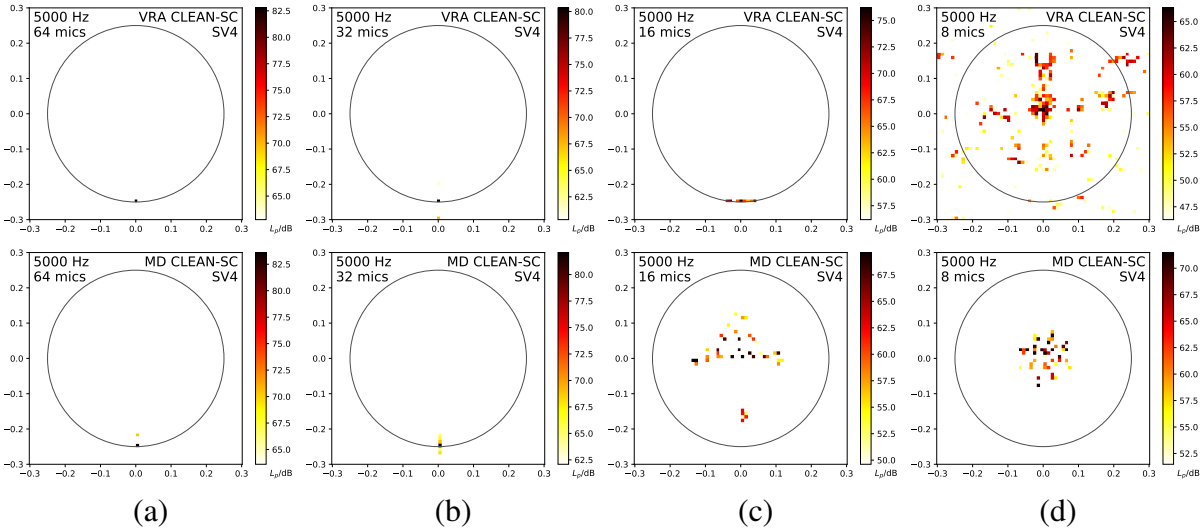

Figure 4: Sound maps for different numbers of microphones (same data and method parameters, different array subsets with equally-spaced microphones): (a) 64 microphones , (b) 32 mics (every second microphone used), (c) 16 (every fourth), and (d) 8 (every eighth).

The number of microphones used on a given radius, and with this the distance between neighboring microphones, limits the upper frequency at which meaningful results can be calculated. This is investigated in Figure 4, which shows 5 kHz CLEAN-SC sound maps calculated with the virtual rotating array and the modal decomposition method – applied on the same data and with the same parameters, but using subsets of the full array geometry.

With the full number of available microphones (Fig. 4a), the source is identified as a point at the correct position. Using only every other microphone for the calculations (Fig. 4b), the source is still detected at the expected focus point, however, additional source levels are calculated radially apart from the correct position. With the VRA method, the “false” sources are farther apart from the correct one, generating separated sources. The calculation with MD leads to an

increased extent of the source in radial direction.

With a subset of only 16 microphones (Figure 4c), the source is calculated at the correct position with the VRA method, if somewhat distributed azimuthally. MD calculates the azimuthal source position correctly, but the radial position is too small. Moreover, a number of false sources are calculated in a broad focus area opposite of the actual source.

Further decreasing the number of microphones, as shown in Figure 4d (8 mics), leads to both methods failing to reconstruct the source position at this frequency. VRA calculates a number of false sources over the whole map area, while with MD, sources are mostly found in the region of the rotation center.

4.3 Three rotating sources

For the evaluation of the qualitative reconstruction capabilities of the three investigated methods, exemplary sound maps shall be discussed. In Figure 5, maps for the DAMAS-NNLS deconvolution of the RF TD result and the VRA FD result are shown alongside the CLEAN-SC deconvolutions of the VRA and MD results. As CLEAN-SC is based on the evaluation of the CSM, it can not be used with RF method, which operates only in time domain. The VRA result is processed using DAMAS-NNLS as well as CLEAN-SC. This allows the attribution of reconstruction-limiting effects to either the motion-compensating processing or the deconvolution method. Furthermore, the manifestation of reconstruction errors due to the motion compensation can be different depending on the deconvolution strategy.

For each method, the sound maps for the one-third-octave bands around 800 Hz, 5 kHz, and 12.5 kHz are plotted. At 5 kHz, all methods calculate three distinct sources at the correct positions. With the DAMAS-NNLS deconvolutions (Fig. 5a und 5b), the extent of the sources is very slightly larger than with CLEAN-SC. The modal decomposition CLEAN-SC result features a minor false source between the strongest and the weakest source (Fig. 5d, 5000 Hz), which is not found at the VRA CLEAN-SC (Fig. 5c). However, the CLEAN-SC implementations for MD and VRA are completely independent, thus leaving a definite attribution of this artifact to the MD method somewhat open.

At 800 Hz, both CLEAN-SC methods feature similar errors in the localization of the sources. The VRA DAMAS-NNLS method, on the other hand, still renders the source positions correctly (Fig. 5b, 800 Hz). This indicates a failure of the CLEAN-SC algorithm at this frequency. This observation is in line with earlier studies of the performance of the CLEAN-SC algorithm [7]. DAMAS-NNLS applied on the RF results (Fig. 5a, 800 Hz), leads to the strongest source being reconstructed as elongated source region. In light of the corresponding good result in the frequency domain, this can be attributed to the time domain data processing.

The sound maps calculated for the 12.5 kHz band all feature artifacts. Since these are similar to those observed in Figure 4b, it can be concluded that the number of available microphones should be higher to resolve sources at this frequency. Aside from that, a more detailed investigation of the PSF / Airy pattern for the RF method at this frequency could be considered and might prove beneficial for the RF DAMAS-NNLS result (Fig. 5a, 12 500 Hz).

For a quantitative analysis of the source reconstruction performance of the methods, source spectra were calculated by integrating sound maps over the focus points contained in a circular area around each of the three sources. The diameter of these integration areas is 0.05 m, which corresponds to the diameter of the circles representing the source positions in Fig. 2.

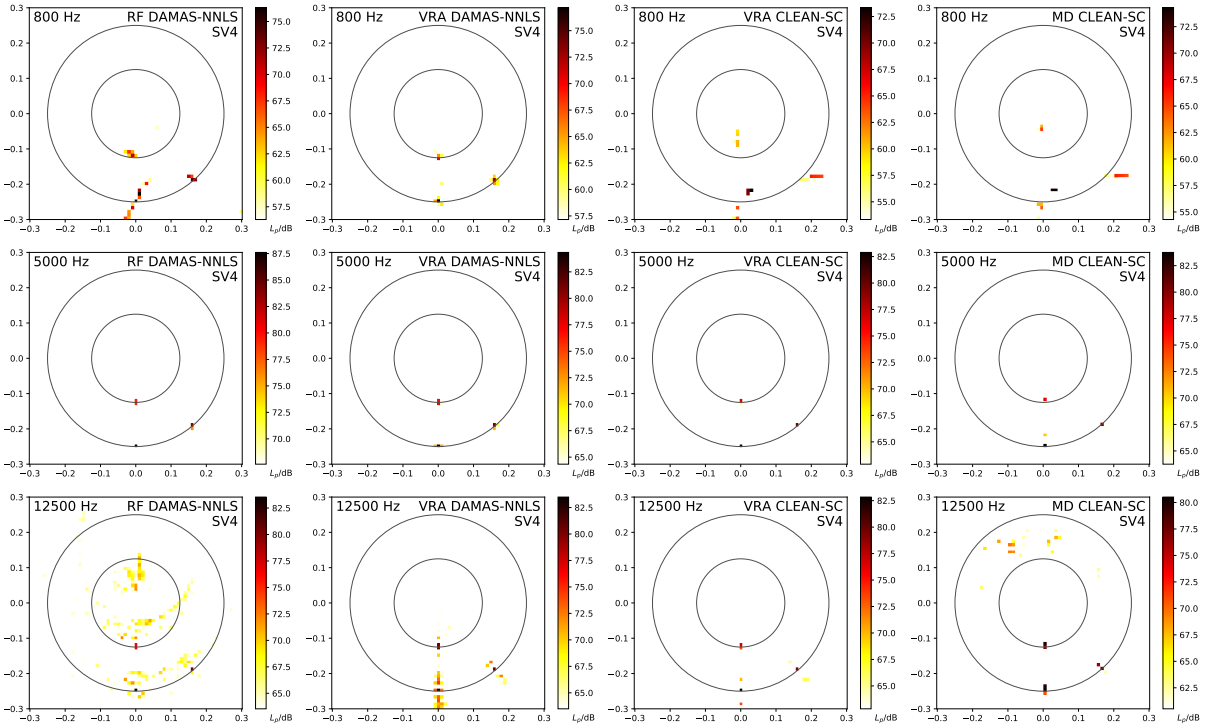

Figure 5: Sound maps for different methods and frequencies. (a): Rotating focus, DAMAS-NNLS deconvolution. (b): Virtual rotating array, DAMAS-NNLS. (c): Virtual rotating array, CLEAN-SC deconvolution. (d): Modal decomposition, CLEAN-SC.

In addition to the methods used for the calculations in Fig. 5, the spectra of the separate simulated sources as they would be measured by an array-centered microphone are plotted, as is the VRA CLEAN-SC method with a different steering vector formulation (SV3):

$$h_m = \frac{1}{r_{t,0} r_{t,m} \sum_{l=1}^{M} r_{t,l}^{-2}} \mathrm{e}^{-\mathrm{j}k(r_{t,m}-r_{t,0})} , \qquad m = 1 \ldots M . \tag{7}$$

This formulation has been shown theoretically to yield the correct source level [12].

Figures 6a-c show the integrated spectra for the individual sources. The simulated "correct" result as would be expected is plotted as a dashed black line. There are several trends that can be observed for all sources: The spectra of all methods is parallel to the correct one between 1250 Hz and 6300 Hz. Above and below, most methods yield lower levels, indicating erroneous source reconstructions at these frequencies. The RF DAMAS-NNLS method tends to generally overestimate the levels by 2 to 3 dB. On the other hand, VRA CLEAN-SC SV4 and MD CLEAN-SC SV4 reconstruct levels that are below the correct ones, with MD CLEAN-SC levels being closer to the expected ones.

Better agreement with the actual levels can be achieved using VRA CLEAN-SC and the steering vector formulation from Eq. (7). VRA DAMAS-NNLS exhibits the best overall fit to the expected levels. Moreover, it is the only evaluated algorithm following the expected trend for

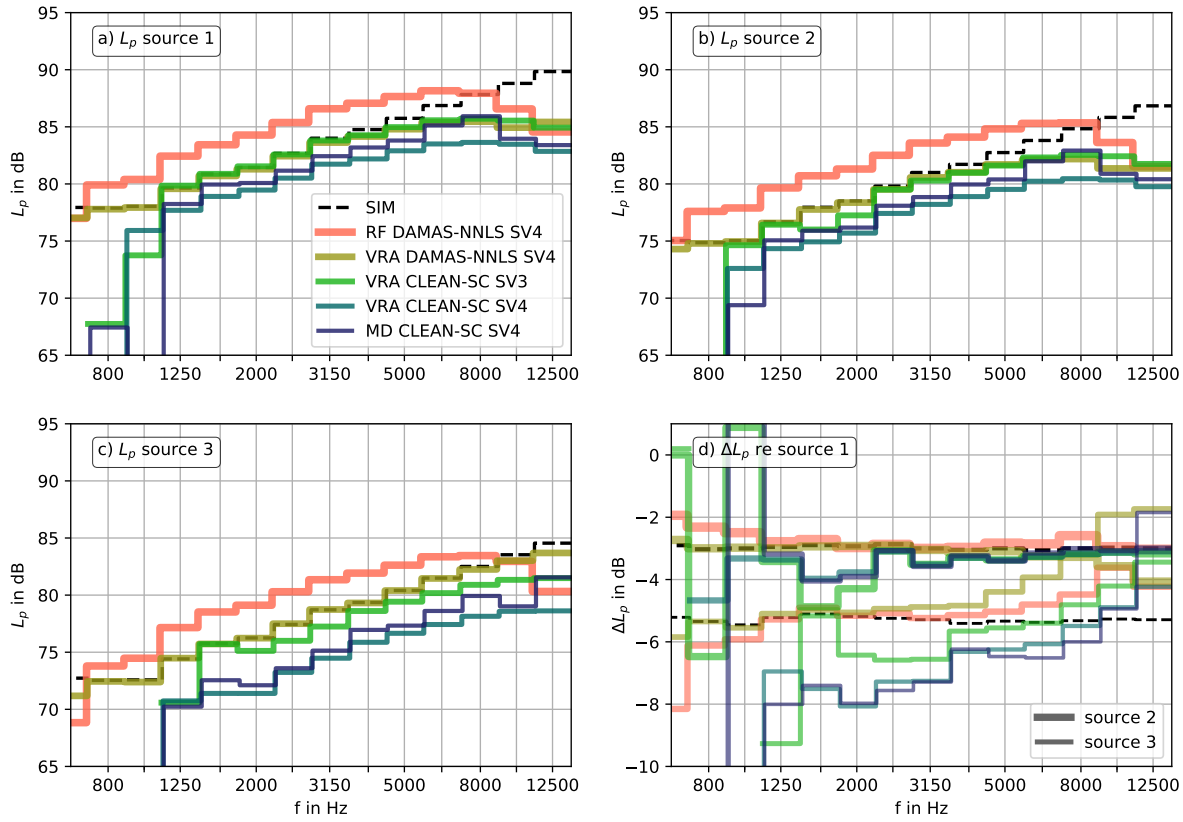

Figure 6: Integrated spectra for different methods and sources. a) Source 1. b) Source 2. c) Source 3. d) Level differences of sources 2 and 3 to source 1.

frequencies below 1250 Hz.

While correctly reconstructing the absolute level of one or more sound sources is desirable, a correct reconstruction of the *relative* levels of all sound sources is often equally or even more important. Figure 6d shows the level differences of the two weaker sources with respect to the strongest source. As can be seen, VRA CLEAN-SC SV4 and MD CLEAN-SC SV4 systematically underestimate secondary sources. Here, the VRA CLEAN-SC SV3 performance is not better; the level differences fluctuate considerably with the frequency. All CLEAN-SC methods suffer from false relative levels below 1250 Hz.

The best performance at low frequencies is again delivered by VRA DAMAS-NNLS. However, at frequencies abaove 3150 Hz, the weakest source level gets overestimated more and more, even exceeding the level of the second-loudest source at 10 kHz.

RF DAMAS-NNLS performs comparably well, with correct relative levels between 1250 Hz and 5 kHz.

5 CONCLUSIONS

Three different array methods for the characterization of rotating sources have been tested on a simulated benchmark case: Rotating Focus (RF), Virtual Rotating Array (VRA), and Modal Decomposition (MD). In general, all methods are able to successfully detect the sources. The frequency range, in which a successful reconstruction is possible, is limited – at high frequencies due to the number of microphones, at low frequencies by the capabilities of the deconvolution algorithms. For a better reconstruction at low frequencies, the DAMAS-NNLS algorithm is suitable. However, the computational cost is considerably higher as for the CLEAN-SC algorithm. The same is true for the RF in comparison with the VRA method, where the RF does not exhibit any advantage over VRA with the current test setup. Whether this can be changed by better modeling the mapping properties of the RF merits further investigation.

REFERENCES

[1] C. J. Bahr, W. M. Humphreys Jr., D. Ernst, T. Ahlefeldt, C. Spehr, A. Pereira, Q. Leclere, C. Picard, R. Porteous, D. J. Moreau, J. Fischer, and C. J. Doolan. "A Comparison of Microphone Phased Array Methods Applied to the Study of Airframe Noise in Wind Tunnel Testing." In *23rd AIAA/CEAS Aeroacoustics Conference*. 2017.

[2] T. F. Brooks and W. M. Humphreys. "A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays." *Journal of Sound and Vibration*, 294(4-5), 856–879, 2006. doi:10.1016/j.jsv.2005.12.046.

[3] M. Debrouwere and D. Angland. "Airy pattern approximation of a phased microphone array response to a rotating point source." *The Journal of the Acoustical Society of America*, 141(2), 1009–1018, 2017. doi:10.1121/1.4976068.

[4] R. P. Dougherty and B. E. Walker. "Virtual Rotating Microphone Imaging of Broadband Fan Noise." In *Proceedings of the 15th AIAA/CEAS Aeroacoustics Conference*. 2009.

[5] G. Herold. "b11: Rotating Point Sources.", 2017-07-28. URL `https://www.b-tu.de/fg-akustik/lehre/aktuelles/arraybenchmark`.

[6] G. Herold and E. Sarradj. "Microphone array method for the characterization of rotating sound sources in axial fans." *Noise Control Engineering Journal*, 63(6), 546–551, 2015.

[7] G. Herold and E. Sarradj. "Performance analysis of microphone array methods." *Journal of Sound and Vibration*, 401, 152–168, 2017. doi:10.1016/j.jsv.2017.04.030.

[8] D. H. Johnson and D. E. Dudgeon. *Array Signal Processing: Concepts and Techniques*. Prentice Hall, Englewood Cliffs, 1993. ISBN 9780130485137.

[9] O. Minck, N. Binder, O. Cherrier, L. Lamotte, and V. Budinger. "Fan noise analysis using a microphone array." In *Fan 2012 - International Conference on Fan Noise*, pages 18–20. 2012.

[10] C. Ocker and W. Pannert. "Calculation of the cross spectral matrix with Daniell's method and application to acoustical beamforming." *Applied Acoustics*, 120, 59–69, 2017. doi: 10.1016/j.apacoust.2017.01.011.

[11] W. Pannert and C. Maier. "Rotating beamforming – motion-compensation in the frequency domain and application of high-resolution beamforming algorithms." *Journal of Sound and Vibration*, 333(7), 1899–1912, 2014. doi:10.1016/j.jsv.2013.11.031.

[12] E. Sarradj. "Three-Dimensional Acoustic Source Mapping with Different Beamforming Steering Vector Formulations." *Advances in Acoustics and Vibration*, 2012, 1–12, 2012. doi:10.1155/2012/292695.

[13] 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*, June. 2017. doi:10.2514/6.2017-3719.

[14] P. Sijtsma. "CLEAN based on spatial source coherence." *International Journal of Aeroacoustics*, 6(4), 357–374, 2007. doi:10.1260/147547207783359459.

[15] P. Sijtsma, S. Oerlemans, and H. Holthusen. "Location of rotating sources by phased array measurements." In *Proceedings of the 7th AIAA/CEAS Aeroacoustics Conference*, pages 1–11. 2001.

[16] P. D. Welch. "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms." *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70–73, 1967. doi:10.1109/TAU.1967.1161901.