



Self-Consistent MUSIC algorithm to localize multiple sources in acoustic imaging

4TH BERLIN BEAMFORMING CONFERENCE

Forooz Shahbazi Avarvand^{1,4}, Andreas Ziehe², Guido Nolte³

¹Fraunhofer Institute FIRST, Kekuléstrasse 7, 12489 Berlin, Germany

²Technische Universität Berlin, Machine Learning Group, Franklinstr. 28/29, 10587 Berlin, Germany

³Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg -Eppendorf
Martinistr. 52, 20246 Hamburg, Germany

⁴Humboldt University of Berlin, Computer Science Department, Rudower Chaussee 25, 12489 Berlin, Germany

01.02.2012

ABSTRACT

The Multiple Signal Classification (MUSIC) algorithm is a standard method to localize acoustic sounds as well as brain sources in electroencephalography (EEG) and magnetoencephalography (MEG). In one of its variants used for EEG/MEG source analysis, called RAP-MUSIC, sequential projections were proposed to properly identify local maxima of the gain function as origins of true sources. The purpose of this paper is twofold: a) we introduce the concept of RAP-MUSIC to the acoustic community, and b) we extend the RAP-MUSIC approach to a fully recursive algorithm. The latter highly suppresses distortions of the estimate of each source due to the presence of other sources. The method is demonstrated for a measurement of the sounds of two loudspeakers placed on a table in a reverberant room. We presented the localization results for Delay and Sum (DAS) beamformer, RAP-MUSIC and SC-MUSIC. In contrast to both DAS beamformer and RAP-MUSIC the new method correctly localizes and separates four sources: the two loudspeakers and the two respective echoes.

1 INTRODUCTION

Locating sound sources in reverberant real-world environments is a challenging problem with many applications including e.g. speaker detection, machine fault analysis and monitoring. One of the most widely used source localization algorithms is Delay and Sum (DAS) beamformer. DAS delays each signal in the sensor space corresponding to the distance of the potential source position located on a virtual plane and sums all the delayed signals from different sensors [2].

The grid point on the plane with the highest estimated power is the estimated source position. One limitation of this method is that sources may be masked by other simultaneously present sources of higher amplitude. The other limitation of DAS beamformer is that the resolution of the algorithm depends highly on the frequency of the signal eventually giving rise to substantial sidelobes for narrow band signals.

Another class of techniques for source localization in array signal processing are methods based on the signal subspace analysis, which are methods for localizing multiple sources by exploiting the structures of the cross-spectra assuming that relatively few sources are active. The general idea behind the subspace methods is to divide the vector space of the data into a signal subspace and a noise-only subspace which is orthogonal to the signal subspace. Among the more frequently used subspace methods is the MULTiple SIGNAL Classification (MUSIC) algorithm [10] which is used for acoustic imaging [4, 5] but also for spectral estimation from single channel data [10, 11] and for the analysis of electrophysiological recordings of brain activity [1, 8]. The MUSIC algorithm finds the source locations as those for which the principle angle between the noise subspace and the forward model of the source is maximum, or, equivalently, for which the principle angle between the signal subspace and the forward model is minimal. In a nutshell, the MUSIC algorithm scans all possible source locations and estimates whether a source at each location is consistent with the measured data explicitly including the possibility that several sources are simultaneously active and in general not independent of each other.

We here focus on a specific variant with the long full name 'Recursively Applied and Projected-MULTiple SIGNAL Classification' (RAP-MUSIC) algorithm. This algorithm is prominent in EEG/MEG source analysis [7]. Its goal is to remove apparent weaknesses of the 'classical' MUSIC algorithm: first, errors in estimating the signal subspace could cause errors in differentiating between true and false maxima. Second, finding several local maxima becomes difficult when the number of sources increases.

While RAP-MUSIC is very effective if sources are uncorrelated, the performance degrades substantially if correlations are not negligible. This is the case in acoustic imaging when echoes are present.

We here propose an extension of the RAP-MUSIC algorithm to address the problem of highly correlated sources. In section 2 we first recall in short the theory behind the MUSIC and the RAP-MUSIC algorithms and then describe the proposed variant. In section 4 we demonstrate the performance of the proposed algorithm for a real measurement of two sources with two strong echoes and reverberant noise in comparison to RAP-MUSIC and DAS beamformer.

2 Theory

2.1 The MUSIC Algorithm

In acoustic imaging the MUSIC algorithm is an analysis of cross-spectral matrices $C(f)$ defined as

$$C(f) = \langle x(f)x^\dagger(f) \rangle \quad (1)$$

where $x(f)$ is a column vector of length N for N sensors and denotes the Fourier transforms at frequency f of the measured signals in a specific segment, and $\langle \cdot \rangle$ denotes expectation value which is approximated by an average over all segments. In its simplest form, MUSIC assumes that only P omni-directional sources are active where $P \ll N$. In that case the rank of $C(f)$ is P

and $C(f)$ spans a P -dimensional subspace of the N -dimensional sensor space.

If a source is placed at a location with distance d_k to the k .th sensor and the activity of the source is $s(f)$ then this source induces a sound pressure

$$x_k(f) = \frac{\exp(-i2\pi f d_k/c)}{d_k} s(f) \equiv v_k(f) s(f) \quad (2)$$

in the k .th sensor, where $i = \sqrt{-1}$ and c is the speed of sound.

For multiple source, $s_p(f)$ with $p = 1, \dots, P$, the k .th sensor reads

$$x_k(f) = \sum_p \frac{\exp(-i2\pi f d_{kp}/c)}{d_{kp}} s_p(f) \equiv \sum_p V_{kp}(f) s_p(f) \quad (3)$$

The columns of the matrix $V(f)$ span the same space as $C(f)$, and the MUSIC algorithm now scans a set of locations on a predefined grid, calculates for each grid point the corresponding forward vector $v(f)$ and measures the deviation of $v(f)$ to the subspace spanned by $C(f)$. In practice, $C(f)$ has always full rank and one considers the space spanned by the eigenvectors of $C(f)$ of the P largest eigenvalues as the signal space. Let U be the $N \times P$ matrix consisting of these eigenvectors and $w(l) = v/\|v\|$ be the normalized forward vector for l .th grid point, then

$$\lambda(l) = \|U^\dagger w\|^2 = \cos^2 \Theta \quad (4)$$

is a measure of the angle Θ between $w(l)$ and the space spanned by U . If the l .th grid point coincides with one of the source locations, then in the ideal situation $\lambda(l) = 1$. In practice one searches for maxima of the gain function

$$h(l) = \frac{1}{1 - \lambda(l)} \quad (5)$$

Up to this point, all calculations were made independently for all frequencies f , i.e. specifically $h(l) = h(l, f)$. To combine different frequencies it was suggested to multiply $h(l, f)$ over all frequencies. We here suggest a slightly different approach, namely to combine frequencies as a weighted sum

$$H(l) = \sum_f \frac{h(l, f)}{q(f)} \quad (6)$$

with $q(f) = \sum_l h(l, f)$. The weighting is necessary because at low frequencies phase variations are smaller and topographies are closer to the respective signal subspace such that the low frequencies dominate the estimates. We found that this approach results in somewhat sharper images. However, the results presented below are not substantially different compared to the product approach.

2.2 The RAP-MUSIC Algorithm

A limitation of the MUSIC algorithm is the detection of multiple sources in less than ideal situations. In noisy, reverberant background and for an imperfect forward model strong signals eventually induce local maxima (called 'ghosts') of the gain function (Eq. 5) which are larger

in magnitude than true locations of weaker sources. To distinguish ghosts from true sources Mosher et al.[7] proposed the Recursively Applied and Projected MUSIC (RAP-MUSIC) algorithm in the context of EEG/MEG source imaging. In spite of its name, that algorithm is sequential rather than recursive. In order to apply this idea to acoustic imaging, one needs to sequentially project out topographies $v(f)$ of previously identified source locations both from data and from the forward model. The algorithm starts with a MUSIC scan without projection as described in the previous section. The maximum of the gain function defines one topography $v(f)$ to be projected out in the next iteration.

The input of the $k + 1$.th iteration are the k topographies identified in previous iterations and combined in the $N \times k$ matrix V containing these topographies as columns. The matrix

$$P_V = id - V \left(V^\dagger V \right)^{-1} V^\dagger \quad (7)$$

with id being the $N \times N$ identity matrix is a projector orthogonal to the space spanned by V . Let $U_V = P_V U$ be the $N \times P$ matrix after projecting out the previously found forward models from the subspace. Then for the l .th grid location with topography $v(l)$

$$\lambda(l) = \frac{\|U_V^\dagger P_V v(l)\|^2}{\|P_V v(l)\|^2} \quad (8)$$

defines the angle now between projected forward model and projected signal space. The gain function $h(l)$ can again be calculated with Eq. 5, and the topography corresponding to its maximum will be included in the projection of the next iteration.

3 Self-Consistent MUSIC(SC-MUSIC)

The main point of this paper is to extend the RAP-MUSIC to a truly recursive algorithm. If RAP-MUSIC indeed found all source locations this can easily be tested as e.g. the original maximum of the first iteration should remain unchanged if all topographies except the first are projected out. In practice, however, this will rarely be the case because the presence of other sources in general biases also the location estimate of the strongest source. This bias can be suppressed by taking out the influence of all other sources as good as possible by projection. This procedure can be repeated M times until all estimated source locations converge. We refer to this strategy as SC-MUSIC (Self-Consistent MUSIC).

Since the original RAP-MUSIC was formulated for EEG/MEG data analysis where the forward model is independent of frequencies, it must be specified how these are combined for the suggested extension to be applied in acoustic imaging. There are two possibilities: either, the entire algorithm is formulated for each frequency separately leading in general to different source locations for each frequency. Although we found that results for individual frequencies appear sharper, the interpretation of the different locations is not trivial. Especially, we encounter a permutation problem, as the order of the sources is in general not identical across frequencies. Or, alternatively, we always estimate a source location from the entire gain function defined in Eq. 6. Although it is conceivable that for different data different approaches are preferable we here proceed with the second approach.

The mathematical details of this algorithm are already given in the previous section. The

methods only differ in what is projected out and when the algorithm terminates. We can hence summarize the SC-MUSIC algorithm as follows

1. Perform a RAP-MUSIC scan for P sources and identify P corresponding topographies.
2. Perform a sequence of sweeps. Each sweep consists of a sequence of P RAP-MUSIC source estimates. To estimate the p .th source project out the topographies of all other previously found sources and maximize the gain function Eq. 6 calculated from Eq. 5 and Eq. 8.
3. The program terminates when a sweep results in the same estimated source locations as the previous sweep.

We found that the algorithm converges rapidly typically requiring less than 10 sweeps. For 50,000 data points in as many as 120 channels and combining 26 frequencies the calculation required 23 seconds for 61×61 grid points on a standard laptop.

4 Example

We conducted an experiment in a realistic environment. To perform the measurement, two loudspeakers were placed on a desk with a distance of approx. 1.50 m from a microphone array. The distance between the two speakers was 50 cm. The acoustic device, manufactured by Gfai tech GmbH company, is shown in Fig.1. It is a spherical acoustically transparent array with 120 omnidirectional electret microphones and a built-in video camera. The measurement was performed in an undamped office room of size $5m \times 4m \times 3m$ with a reverberation time of 660 ms.



Figure 1: A sensor array consisting of 120 microphones placed regularly on a sphere of radius 60cm is combined with a standard CCD web-cam for acquiring an acoustical and an optical image simultaneously [3].

A speech signal saying 'Good Morning' was played simultaneously from both speakers. The reason that we played the same signal from both speakers is to have correlated sources, which is a problematic case for the MUSIC algorithm. The measurement was done with 48 KHz sampling rate and lasted two seconds. The desk acts as a mirror for the acoustic signals and produces two echo sources. So, in total four individual correlated sources including the two

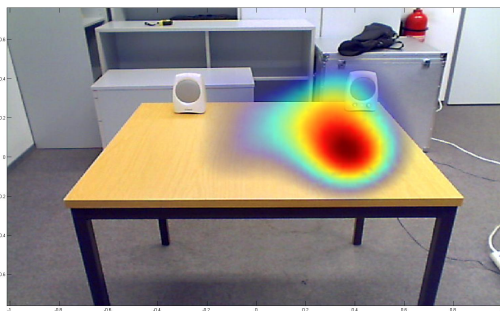


Figure 2: Delay and Sum beamformer applied to the real data. The estimated location of the source doesn't represent any of the source locations.

speakers and the reflections of the sources from the table are expected to be localized. Echoes from wall reflections are fully included in the data.

To estimate cross-spectral matrices the data were divided into segments consisting of 200 trials corresponding to 4.2 msec duration resulting in frequency resolution of 240 Hz. Each segment was windowed with a Hanning function and Fourier transformed to calculate the cross spectral matrices $C(f)$ with Eq. 1. For each frequency we chose $P=10$ eigenvectors corresponding to the 10 largest eigenvalues of $C(f)$. The RAP/SC-MUSIC and DAS beamformer scans were performed up to 5 KHz. As source space we chose a grid of size $1.2 \text{ m} \times 1.2 \text{ m}$ at a distance of 1.5 m from the acoustic array. Finally, we searched for 4 sources both for RAP- and SC-MUSIC.

DAS beamformer, RAP-MUSIC and SC-MUSIC were applied to the data and the power of the signal at each location for DAS and the value of the gain function (Eq. 6) in RAP/SC-MUSIC for each grid point is visualized as a heat map superimposed on the camera image.

In Fig.2 the result of DAS beamformer is shown. The estimated source location resembles to none of the true source locations, since all of the sources are localized simultaneously and that causes the stronger sources mask the other sources. Sometimes the echos have higher power than the sources themselves, which causes the DAS beamformer fail in localizing the true source locations.

In RAP-MUSIC (Fig.3), the first source is localized without projection, and in all following iterations, the previous sources are projected out from the data and from the forward model. We observe that the first scan, corresponding to a 'classical' MUSIC scan, shows a source not corresponding to any of the source locations but rather corresponding roughly to an average location. The following two source estimates are apparently superpositions of the true sources. Only the last source estimate properly identifies one of the sources.

In Fig.4 we show the results for SC-MUSIC. Each panel corresponds to a MUSIC scan after projecting out the sources corresponding to the maxima of all other panels. We notice that the sources are much sharper now and are always located at expected locations. Additionally and most importantly, each source estimate only contains one of the true sources.

We found that neither RAP-MUSIC nor SC-MUSIC is very sensitive to any of the above parameters. E.g., choosing a grid distance of 3 m, a frequency resolution of 480 Hz, and 5 eigenvectors resulted in an almost identical solution. For this measurement, the number of

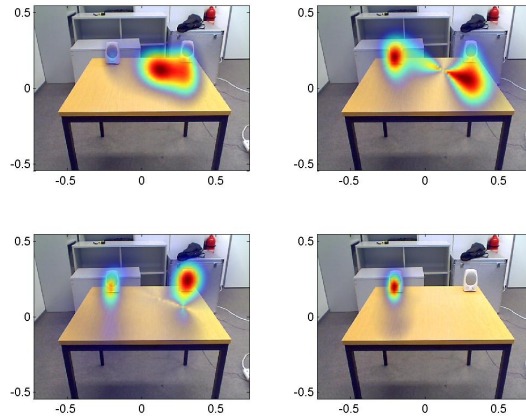


Figure 3: Results for RAP-MUSIC. Top left: MUSIC scan without projection. Top right: Scan after projecting out the source corresponding to the maximum of top left scan. Bottom left: scan after projecting out previous two sources. Bottom right: scan after projecting out previous three sources.

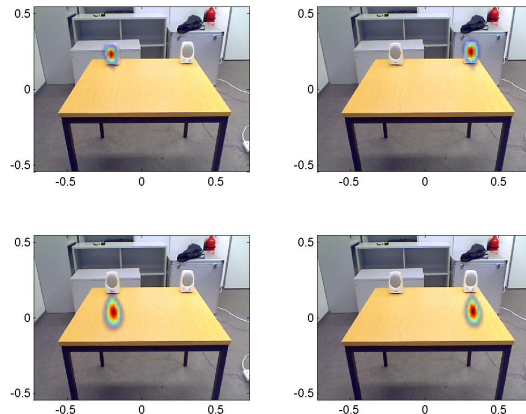


Figure 4: Results for SC-MUSIC. Each panel corresponds to a MUSIC scan after projecting out the sources corresponding to the maxima of all other panels.

estimated sources is more critical, but also for 6 sources the true source locations can easily be identified in 4 maps while the other two show unidentifiable patterns. However, a detailed presentation of the dependence on parameter settings is beyond the scope of this paper.

5 Conclusion

We proposed to extend the iterative RAP-MUSIC algorithm to a fully recursive algorithm to estimate a specified number of sources. The guiding principle was self-consistency. For two sources: if projecting out the first removes the bias in the estimate of the second, then projecting

out the so found second source should also remove the bias in estimating the first source. This method is potentially applicable also to sources which are perfectly coherent, even though it is not guaranteed that the method converges to a global optimum. Under which conditions this is the case and what can be done to avoid local optima will be studied in future research.

We demonstrated the method for a real measurement in reverberant room using loudspeakers with essentially identical output and deliberately strong echoes. In contrast to RAP-MUSIC and DAS beamformer we found that SC-MUSIC properly localizes and separates all four sources. Especially, the clear separation indicates the potential of this approach to recover the original sound sources, i.e. to perform dereverberation.

Finally, we recall that RAP-MUSIC is a prominent method in the EEG/MEG community, and SC-MUSIC can also be formulated for applications in brain research. Here, one major challenge is the separation of brain signals to study interconnected brain structures, while avoiding misinterpretations of superimposed sources as real interactions [6, 9]. Consequently, we will also focus on formulating and testing the method on electrophysiological measurements of brain activity.

References

- [1] G. Crevecoeur, H. Hallez, P. V. Hese, Y. D'Asseler, L. Dupre, and R. V. de Walle. "A hybrid algorithm for solving the EEG inverse problem from spatio-temporal EEG data." *Med Biol Eng Comput*, 46(8), 767–77, 2008.
- [2] L. Griffiths and C. Jim. "An alternative approach to linearly constrained adaptive beamforming." *IEEE Transactions on Antennas and Propagation*, 30(1), 27–43, 1982.
- [3] G. Heilmann, A. Meyer, and D. Döbler. "Time-domain beamforming using 3d-microphone arrays." BeBeC-2008-20, 2008.
- [4] N. Ito, E. Vincent, N. Ono, R. Gribonval, and S. Sagayama. "Crystal-MUSIC: Accurate localization of multiple sources in diffuse noise environments using crystal-shaped microphone arrays." In *LVA/ICA*, pages 81–88. 2010.
- [5] K. Lo. "Adaptive array processing for wide-band active sonars." *IEEE Oceanic Engineering*, 29(3), 837 – 846, 2004. ISSN 0364-9059. doi:10.1109/JOE.2004.833096.
- [6] L. Marzetti, C. Del Gratta, and G. Nolte. "Understanding brain connectivity from EEG data by identifying systems composed of interacting sources." *Neuroimage*, 42(1), 87–98, 2008.
- [7] J. Mosher and R. Leahy. "Source localization using recursively applied and projected (RAP) MUSIC." *Signal Processing, IEEE Transactions on*, 47(2), 332 –340, 1999. ISSN 1053-587X. doi:10.1109/78.740118.
- [8] J. Mosher, P. Lewis, and R. Leahy. "Multiple dipole modeling and localization from spatio-temporal MEG data." *IEEE Trans Biomed Eng.*, 39(6), 541–557, 1992.

- [9] G. Nolte, A. Ziehe, V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, and M. KR. “Robustly estimating the flow direction of information in complex physical systems.” *Phys Rev Lett.*, 100(23):234101, 2008.
- [10] R. Schmidt. “Multiple emitter location and signal parameter estimation.” *IEEE Transactions on Antennas and Propagation*, 34, 276–280, 1986.
- [11] P. Stoica and R. Moses. *Introduction To Spectral Analysis*. Prentice-Hall, 1997.