VPS VIBRATIONS IN PHYSICAL SYSTEMS

DOI: 10.21008/j.0860-6897.2023.2.16
Vibrations in Physical Systems, 2023, 34(2), 2023216

An iterative approach to sound source localization based on spherical beamforming

Adam SZWAJCOWSKI[1], Teresa MAKUCH[1], Weronika CELNIAK[2]

[1]AGH University of Science and Technology, Faculty of Mechanical Engineering and Robotics, al. A. Mickiewicza 30, 30-059 Kraków, Poland;
[2]AGH University of Science and Technology, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, al. A. Mickiewicza 30, 30-059 Kraków, Poland

Corresponding author: Teresa MAKUCH, email: ter.mak@outlook.com

Abstract Precise and efficient localization of sound sources is essential in many applications. Traditionally, methods that use beamforming tend to scan the entire space with fixed level of precision. Although effective, this approach is inefficient when searching for a single source. In this paper we propose an iterative algorithm for localizing a single sound source utilizing signals from a 4th order ambisonic microphone array. Two beamformers were implemented: one based on signals in A-format, incorporating delay-and-sum method, commonly used for sound source localization, and the second one based on B-format, operating in the spherical harmonic domain. By utilizing an iterative algorithm, we have significantly decreased the number of points to be evaluated to localize the sound source. For the delay-and-sum beamformer, the best outcome was obtained by using all 32 channels in every iteration. For the spherical-harmonics-based beamformer, the best strategy was to use first-order harmonics in the initial iteration and fourth-order harmonics in subsequent iterations.

Keywords: source localization, spherical beamforming, optimization.

1. Introduction

The localization of sound sources is crucial in many acoustic applications, such as designing hearing aids and cochlear implants [1], speech recognition systems [2], surveillance and security systems [3], robotics [4] and even wildlife and environmental monitoring [5]. A category of techniques used for this purpose includes Steered Response Power (SRP) methods, beamforming-based methods and Multiple Signal Classification (MUSIC) method. The Steered Response Power method involves the calculation of Generalized Cross-Correlation (GCC) between every pair of microphones in the array [6]. The key idea behind MUSIC is that the signal subspace, which is the subspace spanned by the signal eigenvectors, is orthogonal to the noise subspace, which is the subspace spanned by the noise eigenvectors. By projecting the received signals onto the signal subspace, it is possible to estimate the direction of arrival of the signals [7]. In beamforming-based approaches, the signals from the microphone array are processed to obtain directional information. A microphone matrix is a collection of microphone capsules arranged in specific locations with established geometric relationships between them. The most commonly used shapes for microphone arrays include linear, planar, circular and spherical, which offer one-, two- and three-dimensional processing respectively.

One of arrays that can be used for this purpose is the mh acoustics Eigenmike® em32, which can accurately capture the sound field using 32 raw microphone signals. It is a 4th order ambisonic microphone array, meaning that the microphones are spaced so that an accurate transformation into spherical harmonics domain is possible. The signals after such transformation become ambisonic signals (also called 'eigenbeams'), which can be combined to create microphone beam patterns [8]. This beamforming process allows for steering the beam in any direction in 3D space and focusing on specific directions of sound arrival.

Output of this microphone array can also be used directly for traditional beamforming methods, such as for example delay-and-sum (DAS). There have been attempts to find the best beamforming method for spherical arrays, focusing mostly on their spatial efficiency [9-11]. However, to use the beamformer for signal amplification (or suppressing noise), one needs to identify the direction of the sound source. If a beamformer is used for that, oftentimes the whole space is searched -- the beam is steered towards each direction, with given resolution, and a map of acoustic energy in these directions is created, called the pseudospectrum (whenever the term pseudospectrum is used within the scope of this article, it means the

above-mentioned acoustic energy map). This method guarantees finding the global maximum of energy, but requires a lot of calculations. For one dominant sound source, an optimization algorithm might work better – performing a general scan of space to roughly identify the source direction and then scan more thoroughly only this subspace.

The field of beamformer scanning is still an active area of research, with various optimization algorithms being investigated. There have been some attempts to apply techniques such as gradient descent [12] and space subpartition [13–15]. The latter, also referred to as hierarchical search, has been widely used, as it greatly reduces the required number of directions to evaluate; however, the algorithm is relatively simple and unflexible and thus further research is needed to find the most effective methods.

In this work, we propose an iterative algorithm to find the sound source using less computational resources than brute search, by using beamformers of different widths. For this work two beamformers for the spherical em32 microphone array were implemented: the simple DAS and spherical-harmonic-based (SHB) beamformer. The beamformers are firstly compared in terms of localization precision and then their usability in the iterative algorithm is tested.

2. Methods

2.1. Note on coordinate systems

Depending on the application, different spherical coordinates conventions are used. The one used in this paper is shown in Fig. 1, with the horizontal (azimuth) angle φ increasing counter-clockwise from the X axis from 0 to π and the vertical (elevation) angle ϑ equal to 0 at the horizontal plane (equator of the sphere) and π /2 and −π /2 at north and south pole, respectively; r denotes the distance from the centre of the coordinate system. Therefore, translation to the Cartesian coordinate system follows:

$$x = r\cos\theta\, cos\phi, \tag{1}$$

$$y = r\cos\theta\sin\phi, \tag{2}$$

$$z = r\sin\theta. \tag{3}$$

2.2. Delay and sum beamforming

In acoustic applications, DAS beamforming is a method used to focus sound waves in a particular direction. It involves using multiple microphones to pick up sound waves and then delaying and summing the signals from each microphone. This creates constructive interference in the desired direction and destructive interference in other directions. To determine the direction of a plane wave impinging on an array, we need to define the steering vector, which is based on the time differences between the microphones. To calculate those delays geometric relations between adjacent microphones are used. The positions of each microphone in the array are specified in EigenStudio® User Manual [16]. Delay values can be computed according to formula:

$$\tau_q = -\frac{1}{c}\left[\cos\theta\cos\phi\cdot x_q + \cos\theta\sin\phi\cdot y_q + \sin\theta\cdot z_q\right], \tag{4}$$

where x_q, y_q, z_q are Cartesian coordinates of q-th microphone.

Having calculated these values, we can define the steering vector as:

$$a(f,\theta,\phi) = \left[A_0 e^{-j2\pi f\tau_0(\theta,\phi)},\quad A_1 e^{-j2\pi f\tau_1(\theta,\phi)},\ldots,\quad A_{\mathrm{Q}-1} e^{-j2\pi f\tau_{\mathrm{Q}-1}(\theta,\phi)}\right], \tag{5}$$

where A_q is amplitude gain of signal from q-th microphone and $j = \sqrt{-1}$.

In our scenario, according to the far-field assumption, A_q for the q-th microphone was equal to 1. Therefore, the steering vector for a given direction yields:

$$a(f,\theta,\phi) = \left[\mathrm{e}^{-j2\pi f\tau_0(\theta,\phi)},\ \mathrm{e}^{-j2\pi f\tau_1(\theta,\phi)},\ \ldots,\ \mathrm{e}^{-j2\pi f\tau_{\mathrm{Q}-1}(\theta,\phi)}\right]. \tag{6}$$

To obtain the output signal, we must perform a spatial filtering operation by applying filter weights to each signal from Q microphones. The filter coefficients are determined by the values of the steering vector:

$$h(f,\theta,\phi) = \frac{1}{Q}a(f,\theta,\phi). \tag{7}$$

2.3. Spherical harmonic beamforming

Processing spherical microphone array in the spherical harmonic domain comes from solving the wave equation in the spherical coordinate system and decomposition of a plane wave on a sphere, as described e.g. in [10, 17, 18].

Using the spherical harmonic transform (SHT, also called spherical Fourier transform) for a finite maximum order N, the output of the spherical array with Q microphones can be expressed as [17]:

$$\begin{aligned} p(kr) &= \sum_{q=1}^{Q} \psi_q p\left(kr, \theta_q, \phi_q\right) w_{nm}^*\left(k, \theta_q, \phi_q\right) = \\ &= \sum_{n=0}^{N} \sum_{m=-n}^{n} p_{nm}(kr) w_{nm}^*(k) = \\ &= \sum_{n=0}^{N} \sum_{m=-n}^{n} b_n(kr) Y_n^{m*}(\theta, \phi) w_{nm}^*(k) \,. \end{aligned} \tag{8}$$

The wave number $k = \frac{2\pi f}{c}$ shows dependency on the wave frequency (f), with c denoting the speed of sound; $Y_n^m(\theta, \phi)$ are spherical harmonics of degree n and order m. Weights ψ_q for each microphone depend on their arrangement on a sphere and $b_n(kr)$ are coefficients related to the boundary conditions (rigid or open sphere). In case of the em32, the microphones are placed on a rigid sphere, on faces of the truncated icosahedron, i.e. according to a nearly uniform sampling scheme [19]. This means the weights ψ_q can be equalized while preserving orthonormality of the transformation.

The weighting function $w_{nm}^*\left(k, \theta_j, \phi_j\right)$ that appears in Eq. (8) can be generally written as [10, 17]:

$$w_{nm}^* = d_{nm}/b_n. \tag{9}$$

The choice of d_{nm} depends on the method of analysis. For the spherical beamforming [10]:

$$d_{nm} = d_n Y_n^m(\theta_l, \phi_l) \,, \tag{10}$$

where (θ_l, ϕ_l) is the array look direction, i.e. the direction of the beam. Substituting this into Eq. (8) yields a formula for the spherical beamformer output:

$$p = \sum_{n=0}^{N} d_n \sum_{m=-n}^{n} Y_n^m(\theta_l, \phi_l) Y_n^{m*}(\theta, \phi) \,. \tag{11}$$

For the regular beam pattern, $d_n = 1$ [18]. $Y_n^{m*}(\theta, \phi)$ are results of the SHT. Taking advantage from the fact that the em32 outputs B-format signals, which correspond to the real spherical harmonics, the complex conjugate can be omitted. However, for consistency, the real form of spherical harmonics needs to be used, given by [20]:

$$Y_n^m(\theta, \phi) = \sqrt{(2 - \delta_{m0}) \frac{(2n+1)}{4\pi} \frac{(n - |m|)!}{(n + |m|)!}} P_n^{|m|}(\sin\theta) y_m(\phi), \tag{12}$$

with δ_{m0} denoting the Kronecker delta, P_n^m being the associated Legendre function and:

$$y_m(\phi) = \begin{cases} \sin(|m|\phi) & m < 0, \\ 1 & m = 0, \\ \cos(m\phi) & m > 0. \end{cases} \tag{13}$$

Therefore, in practice, for N-th order processing the values of all real spherical harmonics up to this order are calculated for the beamformer look direction (θ_l, ϕ_l). They are then multiplied by each channel of the B-format signal and summed to obtain the filtered one-dimensional signal in time domain.

2.4. Iterative algorithm

To compute pseudospectra, beamformers need to be directed at a large number of positions, most of which are far away from the searched sound source and thus do not carry important information. To reduce the number of redundant evaluation points, an iterative approach was proposed. The basic assumption is that

in each iteration a few beamformers are designed for such look direction and width that they cover a part of the sphere including the search direction. Then, based on power of the signal filtered by each of beamformers in a current iteration, an estimate of sound source location is returned and the searching continues around that estimate.

In the first iteration, the entire sphere needs to be covered. In order to do that, a wide beamformer is needed and the initial directions to evaluate lie on vertices of a tetrahedron inscribed in a sphere. The unit vectors representing these directions in the Cartesian coordinate system are as follows:

$$\begin{aligned} \mathbf{v}_1 &= \frac{1}{\sqrt{3}}[1,1,-1], \\ \mathbf{v}_2 &= \frac{1}{\sqrt{3}}[1,-1,1], \\ \mathbf{v}_3 &= \frac{1}{\sqrt{3}}[-1,1,1], \\ \mathbf{v}_4 &= \frac{1}{\sqrt{3}}[-1,-1,-1]. \end{aligned} \tag{14}$$

For each of the unit vectors, the power is computed:

$$P_i = \frac{1}{L}\sum_{l=0}^{L-1}(S_i[l])^2, \tag{15}$$

where S_i is the signal resulting from filtering multi-channel signal with a beamformer pointing at direction v_i. Then, an estimate of the sound source location can be derived as a weighted sum of the unit vectors:

$$\hat{\mathbf{v}} = \sum_i P_i \mathbf{v}_i. \tag{16}$$

In the following iterations, only a small part of the sphere is searched. To do that, the estimation is repeated, but a narrower beamformer is used and instead of the initial points from Eq. (14), three new points are taken lying in even intervals on a circle around the estimate defined by angle α (half the apex angle of a cone whose base is the circle). To find these points, a new coordinate system ($\mathbf{x}'$, $\mathbf{y}'$, $\mathbf{z}'$) needs to be defined. Let $\mathbf{z}'$ be pointing in the direction of the estimate:

$$\mathbf{z}' = \frac{\hat{\mathbf{v}}}{|\hat{\mathbf{v}}|}. \tag{17}$$

To find a vector orthogonal to $\mathbf{z}'$, a random vector $\mathbf{r}$ is generated and orthonormalized:

$$\mathbf{y}' = \frac{\mathbf{r} - (\mathbf{r}\cdot\mathbf{z}')\mathbf{z}'}{|\mathbf{r} - (\mathbf{r}\cdot\mathbf{z}')\mathbf{z}'|}. \tag{18}$$

Finally, the third vector $\mathbf{x}'$ has to be orthogonal to both $\mathbf{y}'$ and $\mathbf{z}'$, so it can be found as the cross product of these two:

$$\mathbf{x}' = \mathbf{y}' \times \mathbf{z}'. \tag{19}$$

Within the new coordinate system, the unit vectors for the three new directions to evaluate are given as:

$$\mathbf{v}_i = \sin\alpha\,(\mathbf{x}'\cos\beta_i + \mathbf{y}'\sin\beta_i) + \mathbf{z}'\cos\alpha, \tag{20}$$

where β_i describes the position on the circle surrounding $\hat{\mathbf{v}}$ and is equal to 0°, 120° and 240° for i = 1, 2, 3, respectively. The position of the unit vectors within the new coordinate system in relation to α and β is portrayed in Fig. 1. Value of α (controlling the circle radius) should be chosen with respect to expected current accuracy of estimation.

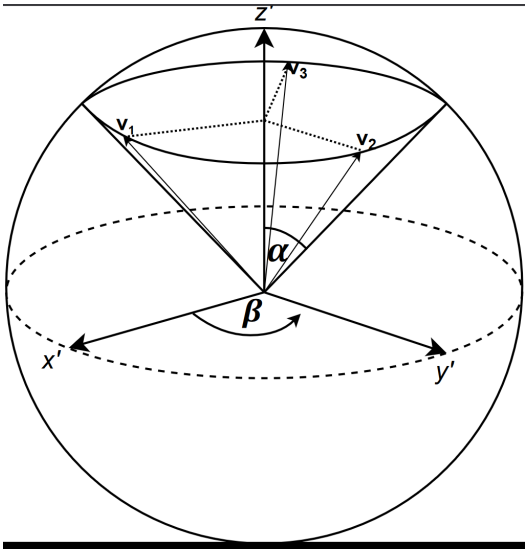

Figure 1. Updated coordinate system with new unit vectors for evaluation.

Updated estimate could be again obtained from Eq. (16); however, when values of P_i are on a similar level, the estimate will change only slightly and the convergence, although successful, will be slow. To counter this phenomenon, a modified formula for the weighted sum was implemented:

$$\hat{\mathbf{v}} = \sum_i \left(\frac{P_i - \min_i\{P_i\}}{\max_i\{P_i\}} + \gamma \right) \mathbf{v}_i \,, \tag{21}$$

where γ is a convergence coefficient. This way, new estimate is moved closer to the circle surrounding the old estimate, but still staying within its borders. The convergence only slows down when γ becomes significant compared to the differences in normalized power values, i.e. when the estimate gets very close to the maximum of the pseudospectrum. The value of γ can be fine-tuned to balance between stability (higher γ) and speed of convergence (lower γ). The optimal value of γ varies depending on the width of beam and the value of α.

Since vectors from Eq. (14) add up to 0, they are invariant to multiplying by or adding constant values; thus, the formula (21) can be applied from the first iteration, since it returns the same results as Eq. (16). The iterative algorithm can be summed up in the following steps:

1) Compute power of the signal filtered with a wide beamformer at the directions defined by unit vectors from Eq. (14).
2) Find new estimate by taking weighted sum of the unit vectors using Eq. (21).
3) Find three points around the estimate by following Eqs. (17) – (20).
4) Compute power of the signal filtered with a narrow beamformer at the directions defined by the points around the estimate.
5) Repeat steps 2-4 until estimate accuracy is satisfactory.

2.5. Experiment design

Both beamformers and the iterative algorithm described in Section 2 were implemented in Python 3.10 using mainly functions from the numpy and scipy packages. To test the implemented beamformers, recordings with known source localization were used.

Two groups of recordings were used, collected using the Eigenmike® em32 microphone. All recordings were acquired with sample rate of 48 000 Hz, either in A-format and B-format simultaneously, or in A-format and later converted to B-format, in both cases utilizing the capabilities of the EigenStudio® software [16]. The source positions were set and written down but not measured precisely.

The first group of recordings, for initial algorithm tests, were impulses obtained by a hand clap in a room with acoustic adaptation and reverberation time of 0.2 s (denoted as '6E'). There were four source positions around the microphone: in front of it ($\phi \approx 0°$), to the left ($\phi \approx 105°$), behind ($\phi \approx 180°$), and to the right ($\phi \approx 270°$). The height of the clapping hands was approximately equal to the height of the centre of the microphone ($\theta \approx 0°$). Each recording was cut to create an 8192-sample impulse response (i.e. approximately 170 ms).

The second group consisted of speech recordings acquired in the same room, with the same source positions as mentioned above and microphone placed in the same place in the room. The speaker was female, the microphone height was approximately aligned with her head centre. In this case, signals of 2 lengths were analysed: a one-word recordings of 120 – 241 ms and the initial 4096 samples (approx. 85 ms) of each of them.

The third group were again hand claps recorded in a 5.6 m × 10.5 m conference room with tables removed (denoted as '103'). Approximate source positions, including source-microphone distance d, were:

$$\begin{gathered} S1(\phi \approx 0°, \theta \approx 0°, d = 1.5 \text{ m}), \\ S2(\phi \approx 300°, \theta \approx 0°, d = 3 \text{ m}), \\ S3(\phi \approx 90°, \theta \approx -45°, d = 1.5 \text{ m}), \\ S4(\phi \approx 60°, \theta \approx 0°, d = 4 \text{ m}). \end{gathered} \tag{22}$$

In this case also "full-length" signals (20372 samples on average, approximately 424 ms) and their 4096-sample excerpts were analysed.

To evaluate the performance, first, pseudospectra for both beamformers were computed. For DAS, using either all 32 microphone channels and also using only every third microphone in an attempt to widen the beam. For SHB, either all 25 channels of B-format were utilized (4th order ambisonics) or only first four (1st order) to obtain narrower and wider beamformers, respectively. Then, both designs were incorporated in the iterative algorithm, with its parameters fine-tuned depending on the information obtained by plotting pseudospectra. Angular error is calculated after each iteration, following the formula:

$$\epsilon = \cos^{-1}\left(\sin\hat{\theta}\sin\theta_c + \cos\hat{\theta}\cos\theta_C\cos\left(\hat{\phi} - \phi_C\right)\right), \tag{23}$$

where $\left(\hat{\phi},\ \hat{\theta}\right)$ are estimated sound source coordinates, while (ϕ_C, θ_C) are coordinates to which the algorithm should converge.

3. Results

3.1. Part 1 – 4092-sample impulses

Pseudospectra were computed and plotted for both designs and for two preset numbers of channels each; they are shown in Fig. 2 and 3 for DAS and SHB, respectively. 32-microphone DAS appeared to have a clean characteristic, but also a relatively wide beam. 11-microphone DAS beamformer was supposed to yield even wider beam, but instead the most notable change are distortions in the spatial characteristic. The artifacts are likely caused by the directivity of microphones – DAS is supposed to be used with omnidirectional microphones of uniform sensitivity. Eigenmike® em32 is designed so that the 32 microphones together cover the entire sphere uniformly, but it is no longer true for an arbitrary selected subset of channels.

In case of the pseudospectra plotted for SHB, there is a clear difference in beam width between 1st and 4th order ambisonics. The latter is much narrower and so is expected to yield better precision although the pseudospectrum features multiple local maxima outside of the main lobe, which could potentially cause problems with convergence.

The sound source location is estimated as maximum of the pseudospectra for 32-microphone DAS and 4th order SHB, with 1 degree precision (Tab. 1). It can be noticed that, even though the results are obtained from the same signal (only preprocessed to different formats), the estimates are slightly different, up to over 10° in both azimuth and elevation for the third recording (S3). Although we expect SHB to be more accurate, it is not entirely clear what is the true sound source location. Nevertheless, in order to compare the results of the iterative algorithm with pseudospectrum approach, the maxima from Table 1 are considered ground truth.

Table 1. Estimated directions for 32-microphone DAS and 4th order SHB.

Source position	DAS (ϕ, θ)	SHB (ϕ, θ)
S1	(358, -4)	(1, -1)
S2	(108, -8)	(107, -1)
S3	(172, -19)	(183, -7)
S4	(273, -7)	(274, -5)

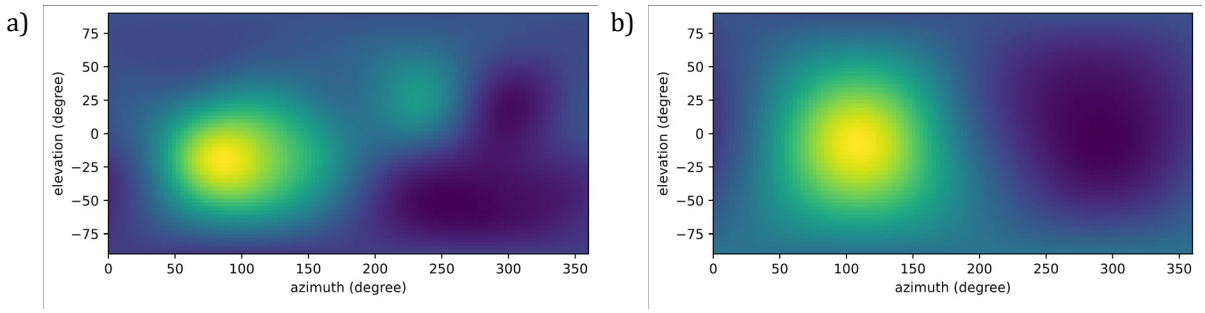

Figure 2. Pseudospectra for DAS: a) 32-microphone, b) 11-microphone.

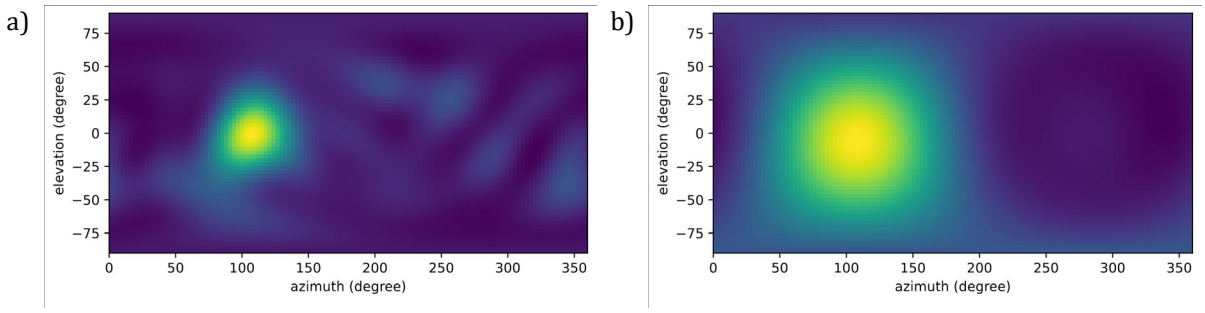

Figure 3. Pseudospectra for SHB: a) 4th order (25 channels), b) 1st order (4 channels).

Since the beam for 32-microphone DAS is quite wide and reducing the number of microphones only introduces undesired artifacts, all channels are used in the algorithm when incorporating DAS design. For the SH beamforming, however, first iteration employs only four channels (1st order ambisonics) in order to cover the entire sphere, and in all the following iterations, the narrower beamformer is used (25 channels 4th order). To account for the difference in the width for the two designs, values of α were set to 20° and 10° for DAS and SHB, respectively. In both cases, manual tests showed that $\gamma = 0.001$ was effective for a quick and stable convergence below 2°, which was considered satisfactory given limited accuracy of the beamformers.

The iterative algorithm was run with random seed in numpy.random set to 1, although using default seeding yielded similar and repeatable performance (random number generator is used only for orientation of β). The results are gathered in Tables 2 and 3 for DAS and SHB, respectively.

Table 2. Angular error (in degrees) after each iteration for DAS beamformer.

	Source position			
Iteration	S1	S2	S3	S4
1	1.16	6.26	7.28	5.26
2		1.41	1.31	0.61

Table 3. Angular error (in degrees) after each iteration for SHB.

	Source position			
Iteration	S1	S2	S3	S4
1	2.70	7.40	9.00	13.40
2	2.47	4.03	6.46	10.45
3	0.74	0.77	3.51	5.94
4			1.13	3.22
5				0.52

The algorithm successfully converged for both designs, although the results for DAS are better both in terms of initial estimation and speed of convergence. However, since the sound source location was estimated only based on a wide beamforming, the end results are less reliable than for SHB. 32-microphone DAS had slightly wider beam than 1st order SHB, which might have resulted in better initial estimation. The estimation for SHB is still quite accurate, but it might be improved by designing wider beam by utilizing higher order ambisonics in the first iteration. 32-microphone DAS appeared to be very well-fitted in that regard, in one case even stopping the algorithm after the first iteration, since the initial estimation was already accurate enough.

The importance of proper beam width can be showcased when SHB is run with the narrow beamformer from the start (Tab. 4). The error after first iteration is then much higher (25° to 45°) and thus up to 16 iterations were needed to converge. Moreover, although for all four samples the algorithm successfully found the global maximum, very rough initial estimation causes risk of stopping at a local maximum. An exemplary convergence is showed in Fig. 4.

Table 4. Angular error after 1st iteration using 4th order SHB.

	Source position			
	S1	S2	S3	S4
Angular error after 1st iteration	25.35	27.81	38.36	44.70
No. of iterations to converge	9	9	12	16

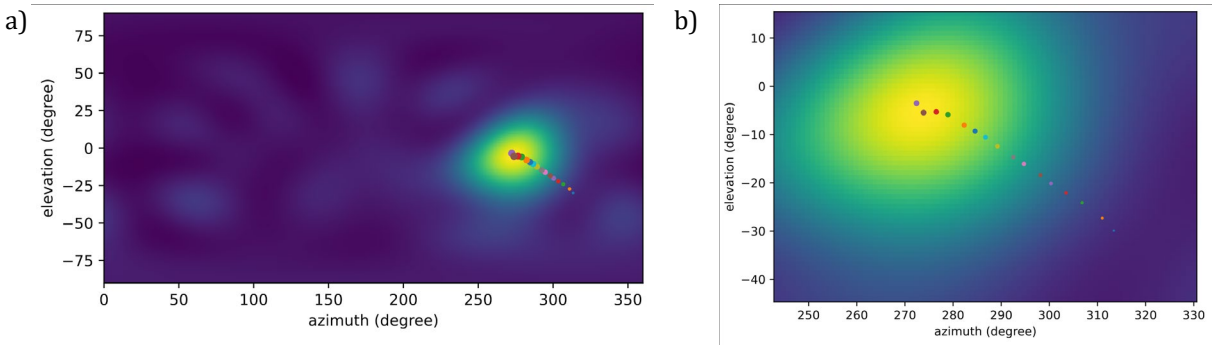

Figure 4. Convergence of results when using 4th order SHB in all iterations: a) full picture, b) zoomed in on the region of interest. Light blue point (far right-bottom) marks the initial estimate while the dark brown one marks the final sound source location estimation.

3.2. Part 2 – longer impulses and speech signals

The analysis described in the previous section would be more valuable if the same method could be used not only for short parts of an impulse but also for real-life continuous signals. Source localization was estimated for 4096-samples excerpts of the second set of impulses and speech signals, using firstly pseudospectrum with 1° resolution, followed by the iterative algorithm. As presented above, localization estimated using pseudospectrum of those short excerpts was taken as a reference for the algorithm. Figure 5 shows angular error after each iteration for all cases, including those presented in the previous section. Description of rooms and source positions is given in Sect. 2.5.

In case of DAS, for 'long' samples, the pseudospectrum maximum from short excerpts was used as a reference direction. This was accurate for impulse samples, but for speech, in one case, the algorithm did not converge (source position S1), meaning the maximum was shifted more than 2° related to the short excerpt.

In most cases, the algorithm using DAS, with a median number of iterations of 3, converges faster than SHB, which needs 6 iterations. The angular error for DAS usually decreases over the iterations, although e.g. for S3 in room 6E it increased after 3rd iteration, doubling the number of iterations to converge. Also, for S4 in room 103 the estimation went away in the wrong direction, before coming down to the aimed position. For the SHB, in all cases but one the angular error was decreasing over iterations. However, for S2 in room 103 the number of iterations to converge was much higher than in other cases (16–17 related to a median of 6). Hence it is difficult to define the number of iterations needed to localize the source.

A more detailed study on the algorithm parameters (γ, α and beam width) would be needed to use it for source localization without preliminary cues.

The algorithm converges similarly regardless of the signal length, which is especially important for DAS, which uses the Fourier transform in its implementation and so increasing signal length impacts the performance dramatically. Comparing average angular error after the first iteration (Tabs. 5 and 6) it can be noticed that the SHB starts off further from the desired direction and, because of a narrower band, it takes more iterations to converge; however, at the same time, this makes the estimate more precise.

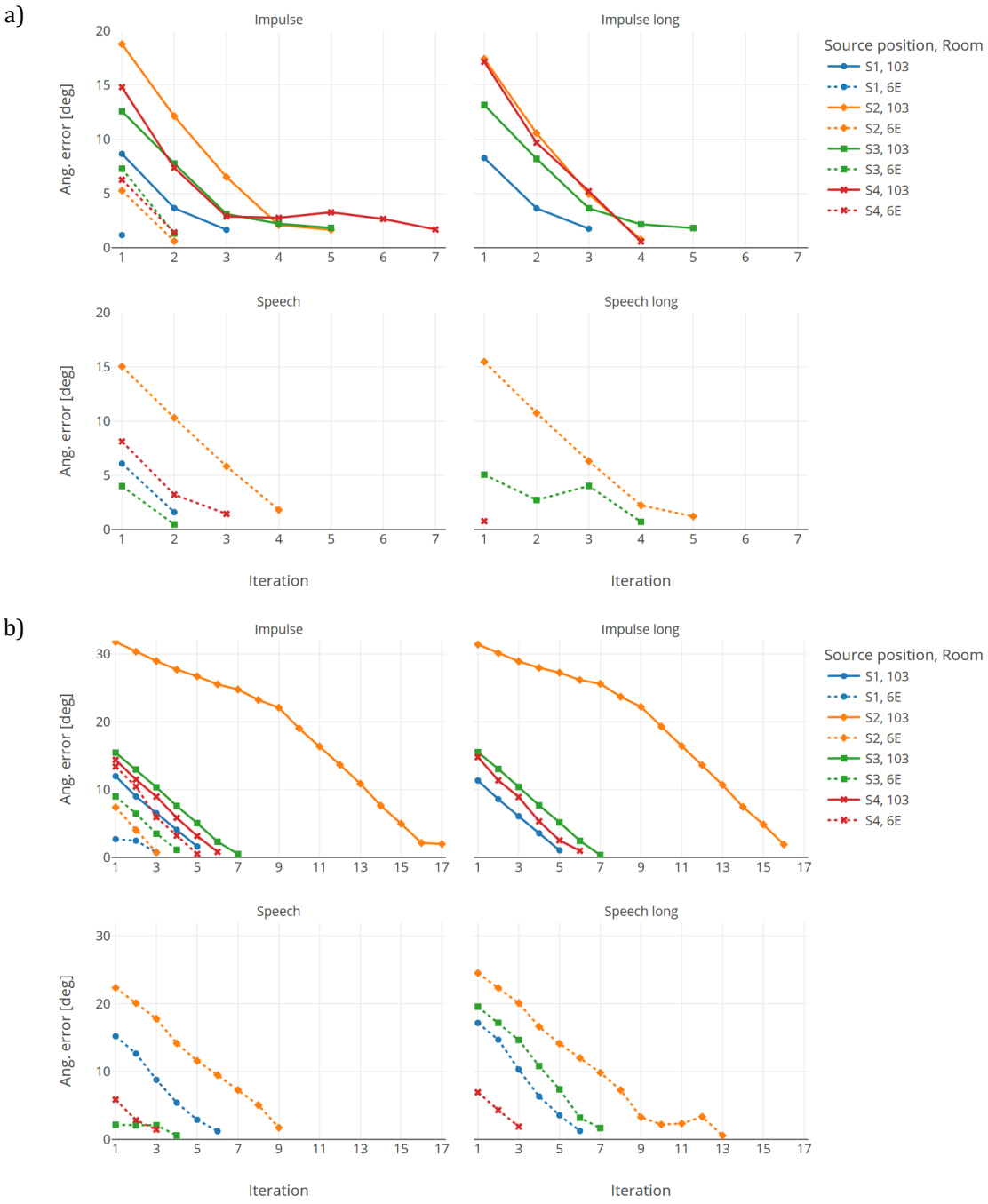

Figure 5. Angular error in subsequent iterations for different signals when using: a) DAS, b) SHB. Single points denote cases where algorithm converged after one iteration.

Table 5. Average angular error after 1st and last iteration for DAS.

	Average	Std. deviation
Angular error after 1st iteration	9.75	5.62
Angular error after last iteration	1.27	0.48
Median number of iterations to converge: 3		

Table 6. Average angular error after 1st and last iteration for SHB (1st order in the 1st iteration, 4th order in subsequent ones).

	Average	Std. deviation
Angular error after 1st iteration	14.65	8.29
Angular error after last iteration	1.14	0.53
Median number of iterations to converge: 6		

4. Discussion

For all cases but one, the iterative algorithm successfully found the global maximum with 2° precision requiring only 1 to 7 iterations, yielding 4 to 22 total evaluations (4 in the first iteration and 3 in each subsequent one). For comparison, to obtain comparable precision with a standard approach of computing pseudospectrum on an equiangular grid, about 3° resolution would be needed (maximum possible error would be then equal to $\frac{\sqrt{2}}{2} \cdot 3° \approx 2.12°$), which yields a total of 21482 directions to evaluate. Even when employing more efficient equidistant grids, it would still require hundreds times more points than our iterative algorithm to achieve comparable precision.

When using hierarchical search, reaching the 2° precision in 3D far-field subspace requires about 25 evaluations, which is also a low number, although still higher than the results for our algorithm in most of cases. Furthermore, the precision increase for the hierarchical search is fixed, which means that the initial estimates are very rough, while our solution provides about 10 – 15° precision already after the first iteration. Finally, the algorithm can be fine-tuned by adjusting the beam width and the convergence parameters, which could further improve its performance. Obtaining satisfactory results even with somewhat simplistic setup shows that the proposed method has applicatory potential.

Pseudospectrum provides full information on the distribution of power around the microphone. The algorithm is intended to only find the location of a single, most prominent sound source around the microphone array. It is thus a more specialized approach, much more efficient in its niche, but less universal overall. Furthermore, the iterative algorithm was designed to work in environments where there is only one dominant sound source. If there are multiple sound sources of similar power, the initial estimation will be less accurate and thus more prone to land near a minor local maximum. It is possible to adjust parameters of the algorithm (γ, α and beam width) to make it more robust at the expense of the speed of convergence. This should be considered in scope of further research in this area.

DAS needs less iterations to converge, but also the convergence threshold was relative to its own pseudospectrum estimate, which is less precise due to larger beam width. SHB, on the other hand, is more reliable in that regard and also enables straightforward control on the beam width. This potentially allows for designing more robust algorithm that utilizes continuous updates to beam width rather than switching between only two preset widths.

Precision of the SHB can also be improved by processing the signal in bands – the magnitude of higher order harmonics ($N > 1$) is frequency dependent, meaning that in lower frequencies the output of the SHT is noisy. On the other hand, above certain frequencies, spatial aliasing occurs, where spherical harmonics of higher order than the array is able to process reach significant magnitude. Both frequency limits depend on the geometry of the array (radius and spacing between microphones) [8]. Filtering the signals would result in a more precise beam. However, bandwidth of the beamformer could also be improved for the DAS, therefore such processing was excluded from the project.

5. Conclusions

We have implemented two beamformers for the spherical microphone array, DAS and SHB. By utilizing an iterative algorithm, we were able to greatly decrease the number of points necessary for localization of the sound source with each beamformer. For DAS the optimal results were achieved by utilizing all 32 channels

in every iteration. The angular error below 2 degrees was obtained in no more than 3 iterations for half of the 20 tested cases. For the spherical-harmonics-based beamformer, the most effective approach involved using first-order spherical harmonics in the initial iteration and fourth-order harmonics in all subsequent iterations.

The number of iterations needed to achieve a satisfactory outcome varied between 3 and 5 for a single impulse sound source placed in one of the main directions (front-left-rear-right) and reached up to 17 for other cases. Further investigation is necessary to confirm effectiveness of the method for other directions – preferably with more precise measurement of actual source direction, to also verify estimation accuracy of both beamformers. Additional research is necessary to evaluate and improve the performance for localization of multiple sources.

Additional information

The authors declare: no competing financial interests and that all material taken from other sources (including their own published works) is clearly cited and that appropriate permits are obtained.

References

1. S.R. Anderson, R. Jocewicz, A. Kan, J. Zhu, S. Tzeng, R.Y. Litovsky; Sound source localization patterns and bilateral cochlear implants: Age at onset of deafness effects; PLOS ONE, 2022, 17(2), e0263516; DOI: https://doi.org/10.1371/journal.pone.0263516
2. F. Asano, M. Goto, K. Itou, H. Asoh; Real-time sound source localization and separation system and its application to automatic speech recognition; In: 7th European Conference on Speech Communication and Technology (Eurospeech 2001), 2001; DOI: https://doi.org/10.21437/eurospeech.2001-291
3. J. Stachurski, L. Netsch, R. Cole; Sound source localization for video surveillance camera; 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, 2013; DOI: https://doi.org/10.1109/avss.2013.6636622
4. C. Rascon, I. Meza; Localization of sound sources in robotics: A review; Robotics and Autonomous Systems, 2017, 96, 184–210; DOI: https://doi.org/10.1016/j.robot.2017.07.011
5. T.A. Rhinehart, L.M. Chronister, T. Devlin, J. Kitzes; Acoustic localization of terrestrial wildlife: Current practices and future opportunities; Ecology and Evolution, 2020, 10(13), 6794–6818; DOI: https://doi.org/10.1002/ece3.6216
6. M.V.S. Lima, W.A. Martins, L.O. Nunes, L.W.P. Biscainho, T.N. Ferreira, M.V.M. Costa, B. Lee; Efficient steered-response power methods for sound source localization using microphone arrays; arXiv:1407.2351 (preprint), 2014
7. R. Schmidt; Multiple emitter location and signal parameter estimation; IEEE transactions on antennas and propagation, 1986, 34(3), 276–280
8. J. Meyer, G. Elko; A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield; In: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002, 2, II-1781-II-1784
9. H. Sun, S. Yan, U.P. Svensson; Space domain optimal beamforming for spherical microphone arrays; In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 2010, 117–120
10. B. Rafaely; Phase-mode versus delay-and-sum spherical microphone array processing; IEEE Signal Processing Letters, 2005, 12(10), 713–716
11. R. Hu, Q. Huang; Source localization using constrained Kalman beamforming in spherical harmonics domain; In: 2013 IEEE International Conference of IEEE Region 10 (TENCON 2013), 2013, 1–4
12. S. Gombots, J. Nowak, M. Kaltenbacher; Sound source localization – state of the art and new inverse scheme; Elektrotechnik und Informationstechnik e & i, 2021, 138(3), 229–243
13. R. Duraiswami, D. Zotkin, L.S. Davis; Active speech source localization by a dual coarse-to-fine search; In: Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001, 5, 3309–3312
14. L.O. Nunes, W.A. Martins, M.V. Lima, L. W. Biscanho, M. V. Costa, F. M. Goncalves, A. Said, B. Lee; A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays; IEEE Transactions on Signal Processing, 2014, 62(19), 5171–5183
15. A. Marti, M. Cobos, J.J. Lopez, J. Escolano; A steered response power iterative method for high-accuracy acoustic source localization; J. Acoust. Soc. Am., 2013, 134(4), 2627–2630
16. EigenStudio® User Manual; 2019
17. B. Rafaely; Analysis and Design of Spherical Microphone Arrays; IEEE Transactions on Speech and Audio Processing, 2005, 13(1), 135–143

18. B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, E. Fisher; Spherical Microphone Array Beamforming; In: Speech Processing in Modern Communication: Challenges and Perspectives; I. Cohen, J. Benesty, S. Gannot, Springer Berlin Heidelberg, 2010, 281–305
19. S. Moreau, J. Daniel, S. Bertet; 3D sound field recording with higher order ambisonics – objective measurements and validation of spherical microphone; In: Proceedings of the 120th Audio Engineering Society Convention, 2006
20. A. Politis; Microphone array processing for parametric spatial audio techniques; PhD Thesis; Aalto University, Helsinki, 2016