

COVER PAGE

Document downloaded by @DAEL

Sun Nov 15 07:19:46 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Independent Component Analysis Using Spherical Microphone Arrays, Nicolas Epain and Craig T. Jin, Acta Acustica **vol. 98** (Number 1), 2012, pp. 91-102

DOI
https://doi.org/10.3813/AAA.918495

Acta Acustica united with Acustica
Vol. 98 (2012) 91 – 102

DOI 10.3813/AAA.918495

Independent Component Analysis Using Spherical Microphone Arrays

Nicolas Epain, Craig T. Jin
Computing and Audio Research Laboratory, School of Electrical and Information Engineering, The University of Sydney, Sydney NSW 2006, Australia. craig.jin@sydney.edu.au

Summary
Spherical microphone arrays provide a new and promising tool for the spatial analysis of complex sound fields. Considerable previous work has investigated the use of spherical microphone arrays for phase-mode beamforming, which relies on the Bessel-weighted spherical harmonic transform of the sound field. In this paper, we investigate the advantages that spherical microphone arrays provide for blind separation of convolutive mixtures using independent component analysis. We demonstrate that applying a standard, linear independent component analysis model in the phase-mode domain enables one to both localize the sources and resolve the permutation problem that plagues most implementations of independent component analysis. As well, we show that the standard linear independent component analysis model can be incorporated into beamforming approaches for source localization and source separation. Simulation results indicate that this approach works in realistic scenarios that include room reverberation.

PACS no. 43.60.Fg, 43.60.Jn, 43.60.Vx

1. Introduction

Independent Component Analysis (ICA) is a statistical method developed in the 1980's [1] that separates component signals in a mixture based on statistical independence and non-Gaussianity. The principles of statistical independence and non-Gaussianity can be applied to both linear, instantaneous mixtures and convolutive mixtures. Within the ICA framework, there is generally the implicit assumption that information regarding the geometric configuration of the sensors is unavailable or not provided. On the other hand, the geometric configuration of the sensors, also referred to as the array steering vector or array manifold vector (see [2]), provides the starting point for all array and beamforming signal processing approaches.

In this paper, we explore the ICA statistical approach in the context that the array manifold vector is explicitly known and made available. In particular, we focus on the problem of blind separation of convolutive mixtures using a spherical microphone array (SMA). We should immediately clarify that although the focus of our study is the blind separation of convolutive mixtures, in this work, we only ever apply the standard linear ICA model. Nevertheless, this simple approach combined with signal processing in the phase-mode domain is still remarkably effective for convolutive mixtures. This is because the incoming plane-wave signals are convolutively mixed in the microphone domain, but linearly and instantaneously mixed in the phase-mode domain. The phase-mode domain signals can be obtained using an appropriate microphone array, such as a circular or spherical array. In the context of sound field reproduction, the phase-mode domain is commonly referred to as the HOA domain.

The application of the standard linear ICA model in the phase-mode domain was initiated by Liu [3], who refers to the approach as blind adaptive beamforming. In contrast to Liu, we focus here and in our previous work [4] on the fact that the standard linear ICA model offers more than blind source separation, but also source localization. More specifically, we demonstrate that the ICA approach combined with knowledge of the array manifold vector allows one to identify source locations and resolve the permutation problem. Both the permutation problem and the fact that the source location can remain unknown, despite achieving source separation, are a peculiarity to the ICA approach. In other words, with the standard linear ICA model there are ambiguities or indeterminacies that necessarily apply (see [1], page 154): the exact energy of the independent components cannot be determined and the order of the independent components cannot be established. In this paper, we demonstrate that these ambiguities are easily resolved when the array manifold vector is explicitly known.

The fact that these ambiguities can easily be addressed when the array manifold vector is made explicit begs the question whether statistical independence is already part of the standard beamforming repertoire. While the beamforming literature certainly makes mention of higher order statistics and statistical independence (e.g., see [5],

Received 28 February 2011,
accepted 3 December 2011.

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

section 9.8) subspace or eigenvector-based methods (the MUSIC and ESPRIT algorithms fall into this category, see [5], section 9.7 and also [2], section 9.3) and methods based on the spatial correlation matrix (see [5], section 9.6) seem to dominate the beamforming literature. In this paper, we show that the standard linear ICA model provides a viable alternative to the standard beamforming methods, with some interesting advantages.

In the first section, we briefly review the background material. This is followed by section 2, which describes the methods for implementing the standard linear ICA model using SMAs. In sections 3, 4 and 5, we present the results of increasingly realistic simulations: in section 3 and 4, the simulations are based on impulse responses simulated in an anechoic and moderately reverberant environment, respectively; in section 5, the simulations are based on impulse responses measured with an actual spherical microphone array in an office space. In the final section we present the conclusions.

2. Background

2.1. Independent Component Analysis

Within the framework of the standard linear ICA model, we assume that the Q microphone array signals consist of an instantaneous linear mixture of N underlying source signals, *i.e.* there exists a time-independent Q-by-N matrix $\mathbf{A}$ such that

$$\mathbf{d}(t) = \mathbf{A}\,\mathbf{s}(t), \tag{1}$$

where $\mathbf{d}(t)$ and $\mathbf{s}(t)$ are the vectors of the Q microphone signals and N source signals at time t, respectively,

$$\begin{aligned}\mathbf{d}(t) &= \left[d_1(t), d_2(t), \ldots, d_Q(t)\right]^T,\\ \mathbf{s}(t) &= \left[s_1(t), s_2(t), \ldots, s_N(t)\right]^T,\end{aligned} \tag{2}$$

where $d_q(t)$ and $s_n(t)$ denote the q-th microphone and n-th source signals, respectively, and $(\cdot)^T$ denotes the transpose of a vector. The matrix $\mathbf{A}$ with elements a_{qn} is referred to as the mixing matrix. In the blind source separation (BSS) problem, both the source signals and the mixing matrix are unknown; thus it is clear that the problem requires an additional constraint to obtain a solution, for otherwise the identity matrix is the simplest solution for the mixing matrix. The guiding principle of the ICA approach is to add the constraint that the source signals are statistically independent with a non-Gaussian distribution.

A difficulty with applying ICA to solve the BSS problem is that the resulting source signals are ordered and normalized in an arbitrary way because both the mixing matrix and the source signals are unknown, *e.g.*, we have

$$\mathbf{d}(t) = \begin{bmatrix} a_{13}/2 & a_{11} & a_{12} & \ldots & a_{1N} \\ a_{23}/2 & a_{21} & a_{22} & \ldots & a_{2N} \\ a_{33}/2 & a_{31} & a_{32} & \ldots & a_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{Q3}/2 & a_{Q1} & a_{Q2} & \ldots & a_{QN} \end{bmatrix} \begin{bmatrix} 2s_3(t) \\ s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{bmatrix}. \tag{3}$$

This issue is referred to as the permutation problem and causes difficulty when trying to join signals that have been processed using short and overlapping time frames.

2.2. Bessel-weighted spherical harmonic transform of a sound field

In the following we define the spherical coordinates (r, ϑ, φ) (radius, elevation angle and azimuth angle, respectively) of a point in space as

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2}, \qquad \vartheta = \operatorname{acos}\left(\frac{z}{r}\right),\\ \varphi &= \operatorname{atan2}\left(\frac{y}{x}\right), \end{aligned} \tag{4}$$

where (x, y, z) denote the cartesian coordinates and atan2 denotes the arctangent function that takes into account the correct quadrant of (x, y).

In the frequency domain, the Bessel-weighted spherical harmonic expansion of a sound field consisting of incident sound waves is given by [6]

$$p(r, \vartheta, \varphi, f) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \mathrm{i}^l\, j_l(kr)\, Y_l^m(\vartheta, \varphi)\, b_{lm}(f), \tag{5}$$

where $p(r, \vartheta, \varphi, f)$ is the acoustic pressure corresponding to the frequency f and at the point with spherical coordinates (r, ϑ, φ), k is the wave number given by $k = 2\pi f/c$ where c denotes the speed of sound, i is the imaginary unit, j_l is the spherical Bessel function of degree l, Y_l^m is the spherical harmonic function of order l and degree m and $b_{lm}(f)$ is the spherical harmonic expansion coefficient for order l, degree m.

The order-L expansion of the sound field is obtained by truncating the series to order L. It provides a good approximation of the sound field within a sphere of radius $\hat{r}$ centered on the origin (see [7, equation 44]),

$$\begin{aligned} p(r \le \hat{r}, \vartheta, \varphi, f) &\approx \sum_{l=0}^{L} \sum_{m=-l}^{l} i^l\, j_l(kr)\, Y_l^m(\vartheta, \varphi) b_{lm}(f)\\ \text{with} \qquad \hat{r} &= \frac{2L}{ek}, \end{aligned} \tag{6}$$

where e is the mathematical constant known as Euler's number.

The above equation shows that a sound field can be represented by a set of coefficients $b_{lm}(f)$ with $l \le L$. In the time domain, it is a vector of time signals, $\mathbf{b}(t)$,

$$\mathbf{b}(t) = \left[b_{00}(t),\ b_{1-1}(t),\ b_{10}(t), \ldots\ b_{LL}(t)\right]^T, \tag{7}$$

where the signals $b_{lm}(t)$ are the inverse Fourier transforms of the frequency-domain coefficients $b_{lm}(f)$. For the sake of brevity and as is common in the sound field application area, we will refer to the Bessel-weighted spherical harmonic expansion as the Higher-Order Ambisonic (HOA) expansion [8]. We will also refer to the signals $b_{lm}(t)$ as HOA signals that exist in the HOA domain.

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

2.3. Plane-wave sound fields

In the case where the sound field consists of a plane wave incoming from the angular direction (ϑ_n, φ_n), $\mathbf{b}(t)$ is given by

$$\mathbf{b}(t) = \mathbf{y}_n s_n(t), \tag{8}$$

where $s_n(t)$ is the plane-wave source signal and $\mathbf{y}_n$ is the vector of the spherical harmonic function values for direction (ϑ_n, φ_n), *i.e.*

$$\mathbf{y}_n = \left[Y_0^0(\vartheta_n, \varphi_n),\ Y_1^{-1}(\vartheta_n, \varphi_n),\ Y_1^0(\vartheta_n, \varphi_n),\right.$$
$$\left. Y_1^1(\vartheta_n, \varphi_n), \ldots,\ Y_L^L(\vartheta_n, \varphi_n)\right]^T. \tag{9}$$

In the case where the sound field consists of N plane waves, the resulting order-L HOA signals are given by

$$\mathbf{b}(t) = \mathbf{Y}\mathbf{s}(t), \tag{10}$$

where $\mathbf{s}(t)$ is defined in equation (2) and $\mathbf{Y}$ is given by

$$\mathbf{Y} = \left[\mathbf{y}_1, \mathbf{y}_2, \ \ldots\ \mathbf{y}_N\right]. \tag{11}$$

Therefore, the HOA signals form a linear, instantaneous mixture of the plane-wave signals (as expressed in equation (1)) and $\mathbf{Y}$ is the corresponding mixing matrix.

2.4. Spherical wave sound fields

In practice, a sound field rarely consists of only a sum of plane waves. In a typical sound field recording scenario, such as a music concert or a meeting, sound sources are more appropriately described as spherical sources. Spherical sources differ from plane wave sources in the way they contribute to the HOA expansion of the sound field. In particular, there is a dependence on frequency. In the case of a unique spherical source at coordinates $(r_n, \vartheta_n, \varphi_n)$, the corresponding frequency-domain HOA expansion is given by

$$\mathbf{b}(f) = s_n(f)\, \mathbf{W}(kr_n)\, \mathbf{y}_n, \tag{12}$$

where $\mathbf{W}(kr_n)$ is a diagonal matrix whose coefficients along the diagonal are the modal coefficients for source n. The elements of $\mathbf{W}(kr_n)$ are given by [9]

$$w_l(kr_n) = \mathrm{i}^{-l} \frac{h_l(kr_n)}{h_0(kr_n)}, \tag{13}$$

where h_l is the degree-l spherical Hankel function of the first kind. Figure 1 illustrates the amplitude and phase of these modal coefficients as a function of kr.

In contrast with the plane-wave case, the mixing matrix for the source signals changes with frequency. Thus, the time-domain HOA signals are no longer instantaneous mixtures of the source signal. Instead, they are obtained by convolving the source signals with filters corresponding to their modal coefficients,

$$b_{lm}(t) = \sum_{n=1}^{N} Y_l^m(\vartheta_n, \varphi_n)\, w_{ln}(t) * s_n(t), \tag{14}$$

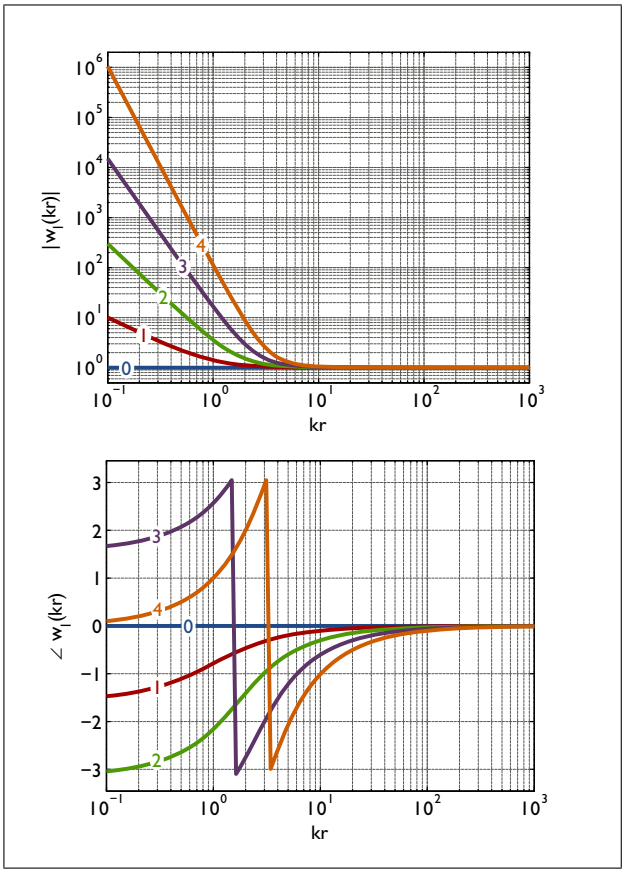


Figure 1. Amplitude (top) and phase (bottom) of the modal coefficients $w_l(kr)$ for a spherical source, as a function of kr and for orders 0 to 4.

where $*$ denotes the time-convolution operation and $w_{ln}(t)$ is the filter whose impulse response is the inverse Fourier transform of the modal coefficient $w_l(kr_n)$.

The modal coefficients asymptotically converge to 1 as kr increases, with the speed of this convergence depending on the order l: the lower the order, the quicker the convergence. In order to observe this convergence more precisely, we define the convergence error, $\epsilon_l(kr)$,

$$\epsilon_l(kr) = \left| w_l(kr) - 1 \right|. \tag{15}$$

Figure 2 shows the value of $\epsilon_l(kr)$ as a function of the order and kr value. At order 1, the convergence error is less than 10% when $kr \geq 10$ which indicates that, above this kr value, the order-1 HOA representation of a spherical source is very similar to that of a plane-wave. At order 4, the convergence error is less than 10% when $kr \geq 100$, which indicates that the source distance or frequency must be 10 times larger as compared to the order-1 case.

Therefore, for a given order, and depending on the frequency content of the source signals and on the source distances, it may still be possible to approximate the contribution of some spherical sources to that of plane wave sources located in the same directions. For instance, for speech signals, talkers located five meters away will be practically indistinguishable from plane-wave sources up to order 2.

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

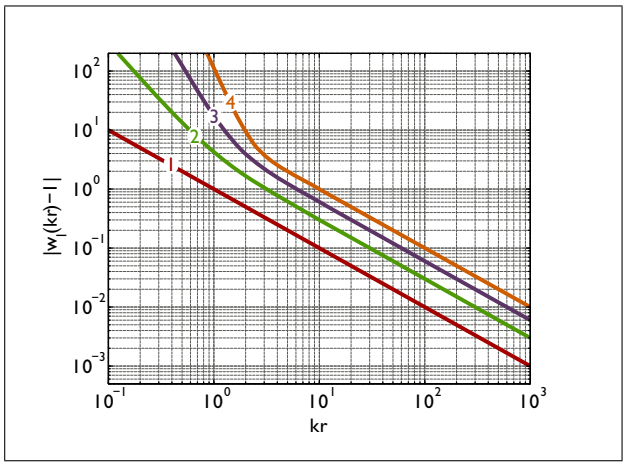


Figure 2. Magnitude of the difference between the modal coefficients $w_l(kr)$ and the value 1, as a function of kr and for orders 1 to 4.

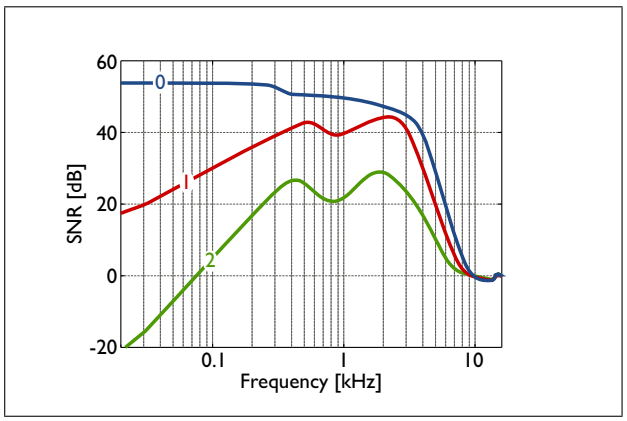


Figure 3. Signal-to-noise ratio of the HOA signals at order 0, 1 and 2 when recording with the microphone array described in section 4 (the exterior radius of which is 15 cm) in the presence of a −40 dB RMS measurement noise.

2.5. Spherical Microphone Arrays

A distinguishing property of spherical microphone arrays is that they easily allow the determination of the HOA signals corresponding to the recorded sound field. In other words, there is a transformation between the microphone array signal domain and the HOA domain,

$$\begin{aligned} \mathbf{d}(f) &= \mathbf{\Omega}(f)\mathbf{b}(f), \\ \mathbf{b}(f) &= \mathrm{pinv}\big(\mathbf{\Omega}(f)\big)\mathbf{d}(f), \end{aligned} \tag{16}$$

where $\mathbf{\Omega}(f)$ is the transfer matrix between the HOA components of the sound field and the microphone signals at frequency f, and pinv[·] is the Moore-Penrose pseudo-inverse matrix operator. In practice, the transformation is implemented in the time domain using a set of filters, referred to as HOA encoding filters. For further details regarding the transformation between the microphone array signal domain and the HOA domain, please refer to [10, 11].

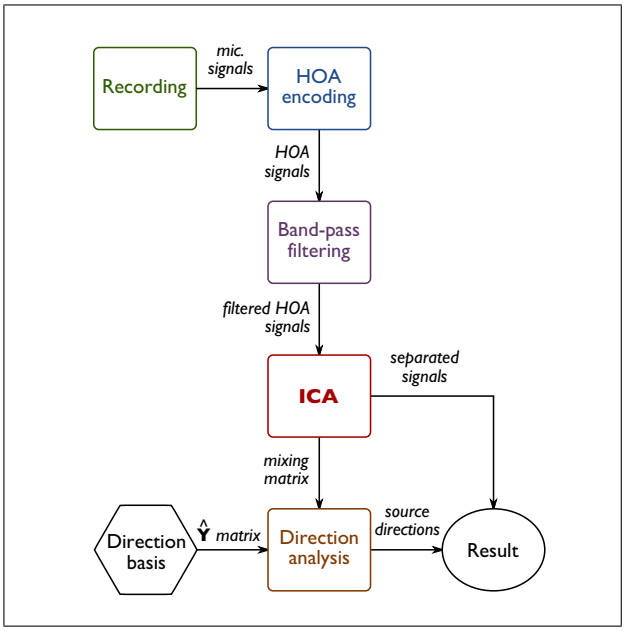


Figure 4. Flow diagram for the Blind Source Separation methods described in section 3. The microphone signals are first transformed to the HOA domain, then the HOA signals are band-pass filtered and finally ICA is done on the band-pass filtered HOA signals.

Measurement noise and spatial-aliasing limit the operational bandwidth of SMAs [12]. For a given SMA, the Signal-to-Noise Ratio (SNR) of the HOA signals depends strongly on both the frequency and the order of the HOA expansion. Figure 3 illustrates the Signal-to-Noise Ratio (SNR) of the HOA signals for a typical spherical microphone array in the case of a −40 dB RMS measurement noise (note that this array is more precisely described in section 4). Clearly, the order-2 HOA signals are accurately acquired only in the 250 to 3500 Hz frequency range. At low frequencies, the SNR is low due to the presence of measurement noise; at high frequencies, spatial aliasing pollutes the different HOA signals with information belonging to higher order spherical harmonics.

3. Implementation of the Linear ICA Model in the HOA domain

Referring to equation (10), we see that under the assumption that the sound field consists of a sum of N plane waves, the order-L HOA signals form an instantaneous mixture of the source signals and their reflections. Thus, we propose to first transform the microphone signals to the HOA domain to obtain the order-L HOA signals of the sound scene and then apply the standard linear ICA model to these signals. In the case where N is less than the number of harmonics, ICA should be able to un-mix the source signals. As well, we show below that the structure of the mixing matrix provides a means to localize the sources and resolve the ICA permutation problem.

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

In order to circumvent issues associated with spherical sources and the bandwidth limitations of SMAs, we propose to band-pass filter the HOA signals prior to running the ICA, as illustrated in the flow diagram in Figure 4. The band-pass filters should be designed primarily to accommodate the bandwidth of operation of the SMA. However, depending on the order and the expected source distances relative to their frequency content, the lower cut-off frequency of the filters can be increased.

The output of the ICA consists of a set of separated signals and the corresponding time-independent mixing and un-mixing matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{U}}$, such that

$$\begin{aligned} \hat{\mathbf{b}}(t) &= \hat{\mathbf{A}}\,\hat{\mathbf{s}}(t), \\ \hat{\mathbf{s}}(t) &= \hat{\mathbf{U}}\,\hat{\mathbf{b}}(t), \end{aligned} \tag{17}$$

where $\hat{\mathbf{b}}(t)$ denotes the vector of the band-pass filtered HOA signals and $\hat{\mathbf{s}}(t)$ denotes the vector of band-pass filtered separated signals, defined similarly to $\mathbf{b}(t)$ and $\mathbf{s}(t)$, respectively. Assuming that ICA separated the source signals perfectly and referring to equation (10), the output signals are proportional to the actual source signals. Thus, each column of $\hat{\mathbf{A}}$ is proportional to a column of matrix $\mathbf{Y}$,

$$\hat{\mathbf{A}} = \mathbf{Y}\,\mathbf{G}\,\mathbf{P}, \tag{18}$$

where $\mathbf{P}$ is a permutation matrix and $\mathbf{G}$ is a diagonal matrix whose non-zero coefficients are the proportionality constants between the actual source signals and the extracted ones. The source directions can then be estimated by comparing the columns of $\hat{\mathbf{A}}$ with a dictionary of V plane-wave direction vectors, $\hat{\mathbf{Y}}$, given by

$$\hat{\mathbf{Y}} = \left[\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ \ldots\ \hat{\mathbf{y}}_V\right], \tag{19}$$

where $\hat{\mathbf{y}}_v$ denotes the direction vector corresponding to the v-th plane-wave and is given by

$$\begin{aligned} \hat{\mathbf{y}}_v = \big[&Y_0^0(\vartheta_v, \varphi_v),\ Y_1^{-1}(\vartheta_v, \varphi_v), \\ &Y_1^0(\vartheta_v, \varphi_v), \ldots,\ Y_L^L(\vartheta_v, \varphi_v)\big]^T. \end{aligned} \tag{20}$$

The correlation, Ψ_{nv}, between the n-th column of the mixing matrix and the v-th column of $\hat{\mathbf{Y}}$ is given by

$$\Psi_{nv} = \frac{\hat{\mathbf{a}}_n^T \hat{\mathbf{y}}_v}{\|\hat{\mathbf{a}}_n\|\,\|\hat{\mathbf{y}}_v\|}, \tag{21}$$

where $\hat{\mathbf{a}}_n$ denotes the n-th column of $\hat{\mathbf{A}}$. The estimated n-th source direction, $(\hat{\vartheta_n}, \hat{\varphi_n})$, is then chosen as the dictionary direction for which the correlation is maximum, *i.e.*

$$\begin{aligned} \left(\hat{\vartheta_n}, \hat{\varphi_n}\right) &= (\vartheta_{v'}, \varphi_{v'}) \\ \text{where} \qquad \Psi_{nv'} &= \max\left\{\Psi_{n1},\ \Psi_{n2},\ \ldots,\ \Psi_{nV}\right\}. \end{aligned} \tag{22}$$

Determining the source directions solves the permutation problem because the different sources can be ordered based on their directions and tracked over time. Furthermore, the value of $\Psi_{nv'}$ can be used to determine whether or not the n-th extracted signal corresponds to an actual source. Real sources should show a large correlation with at least one direction in space, provided the dictionary is dense enough. Therefore, signals whose corresponding spatial correlation values are below a certain threshold (typically 0.95) can be considered as residuals of the ICA and discarded.

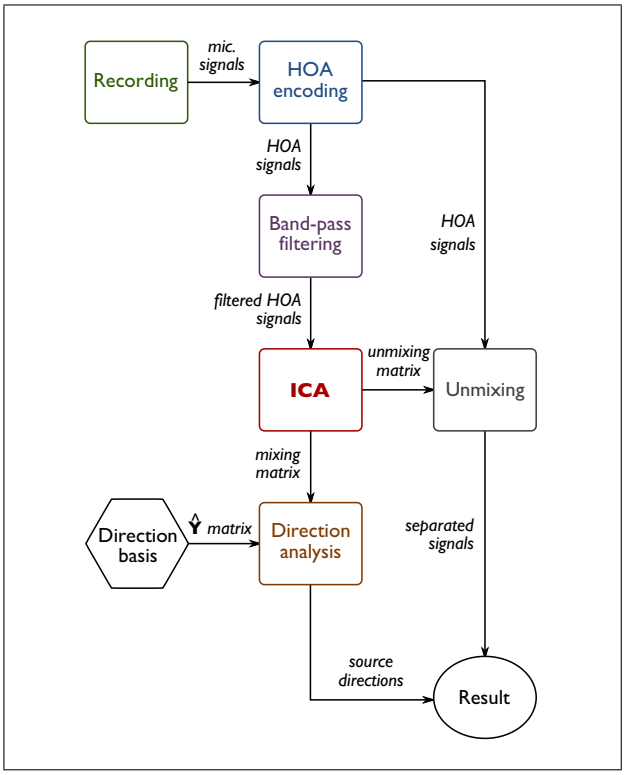


Figure 5. Flow diagram for the BSS methods described in section 3.1. The un-mixing matrix provided by the ICA is applied on the wide-band HOA signals.

3.1. Extending the bandwidth of the separated signals

A simple way of obtaining wide-band separated signals is to apply the un-mixing matrix resulting from the band-limited ICA to the full-band HOA signals $\mathbf{b}(t)$,

$$\tilde{\mathbf{s}}(t) = \hat{\mathbf{U}}\,\mathbf{b}(t), \tag{23}$$

where $\tilde{\mathbf{s}}(t)$ denotes the vector of full-band separated signals and is defined similarly to $\mathbf{s}(t)$. The flow diagram of this second method is illustrated in Figure 5.

Of course, this approach will not separate the source signals effectively at the low and high frequencies, for which equation (10) does not hold due to the spherical nature of the source signals and due to the presence of noise in the HOA signals. Nevertheless, the perceived quality of the output signals may improve because of the extended bandwidth.

3.2. Source separation in the microphone signal domain

We now propose to extract the source signals directly from the microphone signals, as illustrated in Figure 6. In this

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

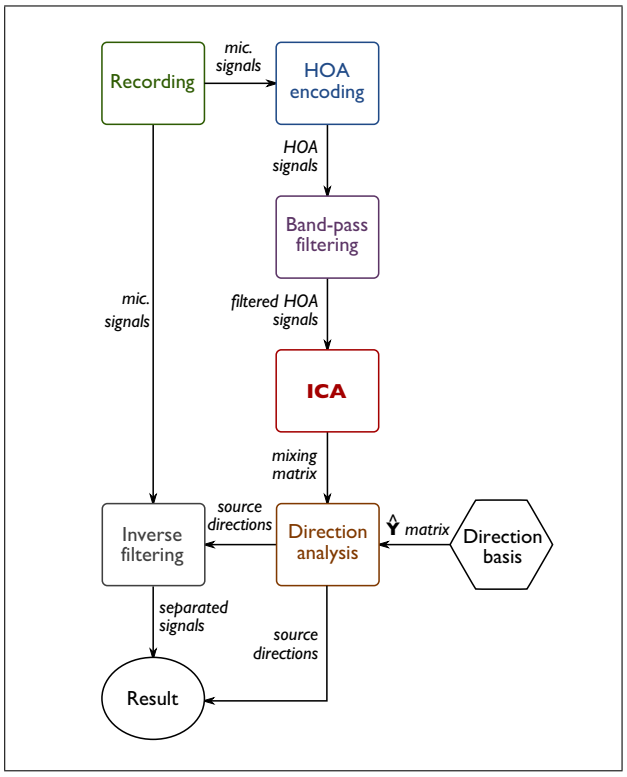


Figure 6. Flow diagram for the BSS methods described in section 3.2. The sound source directions found via ICA are used to generate a set of inverse filters, then these filters are used to separate the microphone signals directly.

third method the result of the ICA performed on the band-pass filtered signals is used only for localizing the sources. These locations are then used to calculate a set of separating FIR filters. These filters are calculated in the frequency domain, using the following formula:

$$\Xi(f) = \mathrm{pinv}\big(\mathbf{\Gamma}(f)\big), \tag{24}$$

where $\Xi(f)$ is the matrix of the inverse filter frequency responses at frequency f, $\mathbf{\Gamma}(f)$ denotes the transformation matrix for the given frequency between each microphone and each of the determined plane-wave source directions, and pinv $(\cdot)$ here denotes the Tikhonov-regularised pseudo-inversion [13], *i.e.*

$$\Xi(f) = \left(\mathbf{\Gamma}^H(f)\mathbf{\Gamma}(f) + \beta\,\mathbf{I}_{N\times N}\right)^{-1}\mathbf{\Gamma}^H(f), \tag{25}$$

where $(\cdot)^H$ denotes the conjugate transpose, β is the regularization coefficient and $\mathbf{I}_{N\times N}$ denotes the identity matrix of dimensions $N \times N$. In the case where the microphone array consists of omnidirectional microphones distributed around a rigid sphere with radius R, the elements of $\mathbf{\Gamma}(f)$ are given by (see [14], equations 1 and 2)

$$\gamma_{pn}(f) = \sum_{l=0}^{\Lambda}\left[\mathrm{i}^l\left(j_l(kr_p) - \frac{j_l'(kR)}{h_l^{(2)\prime}(kR)}h_l^{(2)}(kr_p)\right)\cdot\sum_{m=-l}^{l}Y_l^m(\vartheta_p,\varphi_p)Y_l^m(\vartheta_n,\varphi_n)\right], \tag{26}$$

where $\gamma_{pn}(f)$ denotes the transfer function between the plane-wave source incoming from the direction (ϑ_n, φ_n) and the microphone with spherical coordinates $(r_p, \vartheta_p, \varphi_p)$ at frequency f, j_l' denotes the derivative of the spherical Bessel function of degree l and $h_l^{(2)}$ and $h_l^{(2)\prime}$ denote the degree-l spherical Hankel function of the second kind and its derivative, respectively. In equation (26), the maximum order Λ must be chosen sufficiently large for the sum to converge. The order required for convergence depends mostly on the size of the microphone array and on the maximum frequency at which the transfer functions are to be calculated: the greater kr, the greater Λ.

Once the transfer functions have been calculated, the inverse Fourier transform is used to obtain a matrix of time-domain FIR filters, $\Xi(t)$. Finally, the separated signals are calculated by convolving the vector of microphone signals with the matrix of FIR filters,

$$\check{\mathbf{s}}(t) = \Xi(t) \circledast \mathbf{d}(t), \tag{27}$$

where $\check{\mathbf{s}}(t)$ denotes the vector of the inverse-filter separated signals, defined similarly to $\mathbf{s}(t)$, and $\circledast$ denotes the convolution of a vector of signals by a matrix of filters, *i.e.*:

$$\check{s}_n(t) = \sum_{q=1}^{Q}\Xi_{nq}(t) * d_q(t), \tag{28}$$

where $\Xi_{nq}(t)$ is the inverse filter corresponding to the n-th source and the q-th microphone. In the following, we refer to this method as the "inverse filtering" method.

4. Anechoic simulation

4.1. Simulation setup

In the simulation, eight talkers are located at a distance of two meters around the microphone array. The different angular positions of the talkers are listed in Table I. Each talker is modeled as an omnidirectional spherical source. The talker signals are male or female speech signals recorded in free field conditions.

The microphone array has been designed to provide high quality 3D, order 2, HOA signals from about 300 to 3500 Hz, which approximately corresponds to the narrow-band frequency range of a telephone. Its design is similar to that presented in [15]. It consists of two concentric arrays of 12 omnidirectional microphones. There are 12 microphones located on the surface of a rigid sphere with a radius of 3 cm; the other 12 microphones are located on the surface of an open sphere with a radius of 15 cm. For both arrays, the angular positions of the microphones correspond to the corners of an icosahedron. The HOA encoding filters are of length 512 and are calculated to maximize the SNR of the HOA signals for every frequency value. The resulting SNR is presented in Figure 3.

As we focus on the separation of speech sources, a sampling frequency of 16 kHz is used to reduce the computational complexity of the simulation. The sound wave propagation between the sources and the sensors is modeled

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

using an order-31 HOA expansion, which simulates the diffractive effect of the rigid sphere accurately up to the Nyquist frequency, 8 kHz. The transfer functions between the spherical sources and the microphone array sensors are calculated in the frequency domain using

$$\tau_{pn}(f) = \sum_{l=0}^{31} \left[\mathrm{i}^l \left(j_l(kr_p) - \frac{j_l'(kR)}{h_l^{(2)\prime}(kR)} h_l^{(2)}(kr_p) \right) \cdot w_l(kr_n) \sum_{m=-l}^{l} Y_l^m(\vartheta_p, \varphi_p) Y_l^m(\vartheta_n, \varphi_n) \right], \quad (29)$$

where $\tau_{pn}(f)$ denotes the transfer function between the spherical source with spherical coordinates $(r_n, \vartheta_n, \varphi_n)$ and the microphone with spherical coordinates $(r_p, \vartheta_p, \varphi_p)$ at frequency f and $w_l(kr_n)$ is defined in equation (13). Note that this equation is the same as in the case of a plane-wave source (see equation (26)) except that the spherical harmonic function values corresponding to the source direction are weighted by the coefficients $w_l(kr_n)$.

In addition, the effect of measurement noise is modeled by adding a −40 dB RMS uncorrelated white noise to the microphone signals. The microphone signals are then filtered using the HOA encoding filters described previously. The resulting HOA signals are band-pass filtered so that only the 500 to 3500 Hz frequency range is used for the analysis. The standard linear ICA model is then applied to the narrow-band HOA signals using FastICA [16], an Independent Component Analysis package for the MATLAB environment. The FastICA algorithm is called within MATLAB using the following options: approach: 'symm' ; stabilization: 'on' ; nonlinearity: 'tanh'. Finally, the source directions are determined using a plane-wave dictionary consisting of directions that are regularly distributed in angle with a constant step of 1° for both the azimuth and elevation angles.

For the purposes of comparison, we also apply a standard narrow-band MUSIC analysis as described in section 9.7.1 of [5] to the band-pass filtered HOA signals. The results of the MUSIC analysis are accumulated across the overlapping time frames that comprise the entire signal.

4.2. Simulation results

Table I compares the results of the source localization analysis performed on the mixing matrix, as described in section 3, with a standard narrow-band MUSIC analysis. The resulting error for the source directions using the linear ICA model is of the order of 1°, which is on the order of the accuracy of the direction vector dictionary that was used. The output of the ICA contains 9 signals, which corresponds to the number of input HOA signals. However, it was easy to eliminate one of them based on the low correlation score between the corresponding mixing matrix column and all of the direction basis vectors. Compared with the MUSIC source localization algorithm, the linear ICA model is more accurate and does not miss any source directions.

Table I. The true and estimated source directions for the eight simultaneous talkers are shown for the anechoic simulation. If there was no estimated direction corresponding to one of the true target directions, then a '·' placeholder is used.

Talker	(ϑ, φ) True	(ϑ, φ) ICA	(ϑ, φ) MUSIC
1	(90, 0)	(91, 0)	(94, 13)
2	(100, −30)	(99, −30)	(112, −43)
3	(90, 40)	(89, 40)	·
4	(90, −70)	(91, −70)	(90, −75)
5	(80, 60)	(81, 59)	(93, 65)
6	(90, −90)	(90, −90)	·
7	(110, 90)	(110, 89)	(150, 126)
8	(120, −140)	(120, −141)	(119, −138)

Table II. PESQ scores for talkers 1 to 8 and average PESQ score obtained in the anechoic simulation with, from top to bottom: an order-2 spherical beamformer steered in the talker directions; the proposed narrow-band ICA method; the proposed wide-band ICA method; the proposed inverse filtering method.

Method	Talker 1	2	3	4	5	6	7	8	avg.
Beam.	1.5	1.1	1.5	1.7	1.7	1.1	1.6	1.8	1.5
ICA 1	2.2	2.1	2.0	1.9	2.2	1.9	2.5	2.3	2.1
ICA 2	2.2	2.0	2.1	2.0	2.5	1.8	2.9	3.0	2.3
ICA 3	2.3	2.5	2.4	2.5	2.8	2.2	2.9	3.0	2.6

In order to evaluate the quality of the source separation, the PESQ (Perceptual Evaluation of Speech Quality) scores [17] have been calculated for the output signals of the three proposed ICA methods for every talker. The calculation was performed using the MATLAB PESQ package included in [18]. The results are presented in Table II and are compared with the PESQ scores obtained when using an order-2 spherical beamformer [19] steered in the actual talker directions. Note that, in the BSS paradigm, the true talker directions would not actually be available and would thus have to be estimated. The ICA methods demonstrate a much larger PESQ score than the beamformer, suggesting that the sources are better separated. On the other hand, the scores obtained by the ICA methods are low in absolute terms, which is partly caused by the high level of noise in the signals. This noise results from the HOA encoding of noisy microphone signals.

Figures 7, 8 and 9 illustrate the differences in the quality of the output signals obtained with the ICA methods, as compared with the output of the spherical beamformer. Figure 7 shows the spectrogram of the original signal corresponding to the source located in direction (100°, −30°). This spectrogram can be compared to the spectrograms of the corresponding outputs obtained with the spherical beamformer, shown in Figure 8, and the three proposed ICA-based methods, shown in Figure 9. Despite the significant amount of noise, the output of the narrow-band ICA method is clearly less polluted by the other talkers than

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

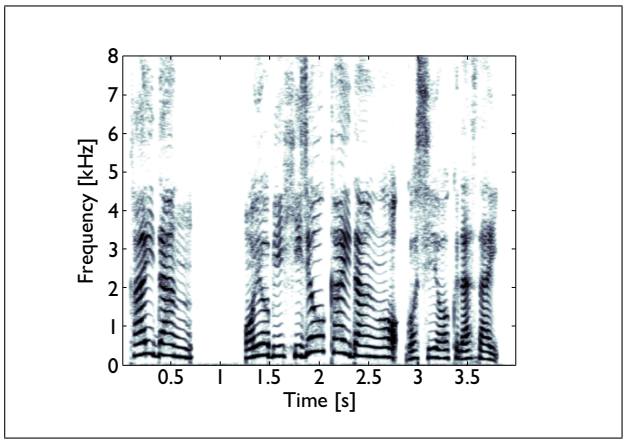


Figure 7. Spectrogram of the signal corresponding to the source located in direction (100°,−30°).

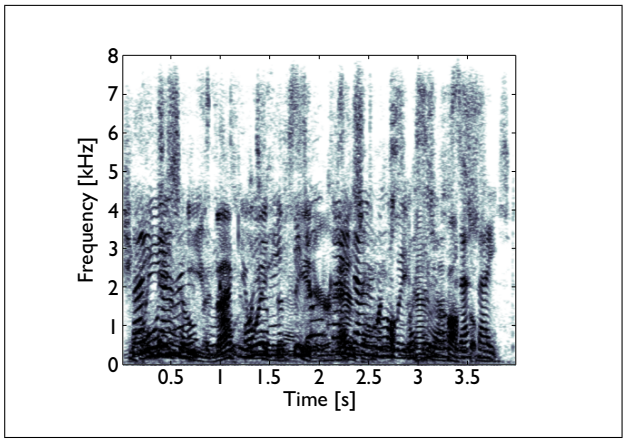


Figure 8. Spectrogram of the output of a second-order spherical beamformer steered to the direction (100°,−30°).

the spherical beamformer output. On the other hand, due to band-pass filtering, it has no energy apart from the 500 to 3500 Hz frequency band. The wide-band ICA method provides good source separation in the frequency band 350 to 4000 Hz. However, there is a large amount of noise because the signal has been obtained by un-mixing noisy HOA signals. Nonetheless, the output still seems cleaner than that of the spherical beamformer. Finally, the output of the inverse filtering method is clearly the most similar to the original source signal, with a relatively low noise level compared to the outputs of the other methods. The source signal seems accurately separated from the others across the entire frequency range, except below 350 Hz where the microphone array is too small for the signals to be effectively unmixed.

5. Reverberant simulation

Microphone arrays are typically used in rooms, where sound reverberation occurs. The reflections of sound on the walls can be roughly modeled as the presence of image sources outside the room. This causes a problem as the

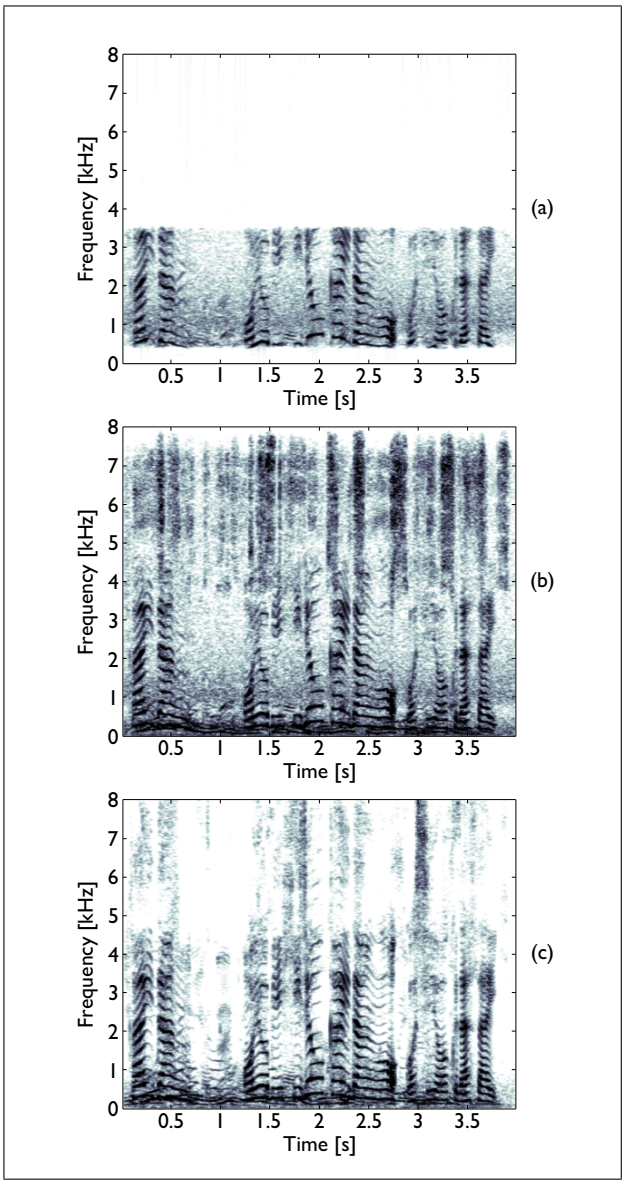


Figure 9. Spectrograms of the outputs of the three proposed methods, corresponding to the source located in direction (100°,−30°). From top to bottom: (a) narrow-band ICA method; (b) full-band ICA method; (c) inverse filtering method.

standard linear ICA model cannot separate more sources than there are channels in the analyzed mixture. Depending on the scenario, the amount of reverberation can be moderate (typical office room) to extreme (concert hall or cathedral). The proposed method may be ineffective if the direct sound sources are not "dominant" enough relative to the reflected waves, which is the case in highly reverberant environments. In order to assess the influence of reverberation on the performance of the proposed methods, we simulated a sound scene in a reverberant environment.

5.1. Simulation setup

The setup for the reverberant simulation is similar to that of the free-field simulation: five talkers are located at a distance of two meters around the microphone array. The

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

Table III. The absorption coefficients for the room modeled in the reverberant simulation are shown.

Walls	Frequency (Hz) 125	250	500	1000	2000	4000
Side Walls	0.2	0.3	0.3	0.45	0.5	0.6
Floor	0.4	0.5	0.5	0.45	0.65	0.8
Ceiling	0.3	0.4	0.3	0.45	0.6	0.7

Table IV. The true and estimated source directions for the five simultaneous talkers are shown for the reverberant simulation. Other details as in Table I.

Talker	(ϑ, φ) True	(ϑ, φ) ICA	(ϑ, φ) MUSIC
1	(90, 0)	(90, 0)	(91, 0)
2	(100, −30)	(97, −30)	(92, −28)
3	(90, −70)	(88, −75)	·
4	(80, 60)	(80, 58)	(82, 57)
5	(90, 120)	(91, 122)	(90, 121)

Table V. The PESQ scores for talkers 1 to 5 and the average PESQ score obtained in the reverberant simulation are shown. Other details are as described in Table II.

Method	Talker 1	2	3	4	5	avg.
Beam.	1.7	1.1	2.1	2.0	1.9	1.8
ICA 1	1.8	1.5	1.9	2.0	2.2	1.9
ICA 2	1.7	1.0	2.2	2.1	2.2	1.9
ICA 3	2.0	1.1	2.1	2.3	2.1	1.9

different angular positions of the talkers are shown in Table IV. The microphone array has the same configuration as previously.

In contrast with the free-field simulation, the talkers and microphone array are now located in a moderately reverberant room. The reverberant impulse responses between the sources and the microphone array sensors have been calculated using MCRoomSim [20, 21], a room simulation package for the MATLAB environment. The room characteristics are those of a common office room, with dimensions 14×10×3 m. The absorption coefficients for the walls are listed in Table III. The T30 reverberation time of the room is approximately 450 ms.

5.2. Simulation results

Table IV shows the results of the source localization analysis performed on the mixing matrix obtained in the reverberant condition. The resulting error for the source directions is slightly higher compared to the free-field condition, and on the order of 3°. The results are again arguably better than those obtained using the MUSIC analysis, which fails to identify one of the source directions. As was the case in the anechoic simulation, the output of the ICA contains 9 signals, which corresponds to the number of input HOA signals. Four of them have been identified as residual signals because they mostly contained a mixture of reverberated sounds.

We calculated the PESQ scores of the separated signals obtained via the three proposed methods and compared them to the scores obtained using an order-2 spherical beamformer steered in the identified source directions. The results are given in Table V. The PESQ scores obtained with the different methods are not sufficiently different to allow any conclusions to be made on the performances of the various techniques. However we believe perceptual tests are warranted as the perceptual quality of the sounds seems to vary for the different methods.

6. Simulation based on measured impulse responses

In the two previous sections, we presented the results of simulations based on simulated microphone impulse responses. Although some random noise was added to the microphone signals, the model used for the microphone array was not as realistic as possible because some sources of error in the HOA encoding of the sound field were neglected. Among the assumptions made, the sensors were assumed to be located at their ideal positions, which is never the case in practice. In addition, gain and phase mismatches between the sensors were not taken into account. In order to evaluate the performance of the proposed algorithms in real-world conditions, we simulated a sound scene using *measured* microphone impulse responses.

6.1. Simulation setup

The simulation is similar to the ones presented in the sections above, except that the impulse responses from the talkers to the microphone array sensors were measured with an actual microphone array [15, 10]. A picture of the measurement setup is shown in Figure 10. The microphone array consists of 64 microphones distributed on the surface of two concentric spheres: 32 microphones are located on a rigid sphere of radius 16.3 mm, and the 32 others are located on an open sphere of radius 60 mm. The impulse responses were measured using a Tannoy V6 loudspeaker that was moved to different locations in a room. Note that the loudspeaker was positioned using a tape measure and simple geometry principles, which resulted in a placement accuracy on the order of a few centimeters. The room was an office space with approximate dimensions 14×8×3 m and its average reverberation time (T30) was approximately 0.5 s.

In the simulation scenario, five simultaneous talkers are present around the microphone array. The microphone signals were calculated using the impulse responses measured with the loudspeaker located in the directions indicated in Table VI. The distance between the loudspeaker and the center of the microphone array was approximately 1.80 m. A −40 dB RMS uncorrelated white noise was added to the microphone signals in order to simulate the presence of background noise.

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only



Figure 10. A photo is shown of the dual-radius spherical microphone array being measured in an office at the University of Sydney.

Compared to the microphone array used in the previous simulations, this SMA has many more microphones. Therefore, it can be used to record the HOA signals up to the fourth order. However, as the overall size of this SMA is smaller than the SMA used for the previous simulations, the SNR of the HOA signals is rather poor at low frequencies. In order to circumvent this issue, as well as to make this simulation more comparable with the ones presented in the previous sections, the microphone signals were encoded to HOA signals only up to order 2. In addition, due to the size of the SMA, the HOA signals were high-pass filtered with a cutoff frequency of 750 Hz prior to running the ICA. On the other hand, no low-pass filtering was required as the microphone array's spatial aliasing frequency is approximately 16 kHz, which is twice the Nyquist frequency in this simulation.

6.2. Simulation results

Table VI shows the talker directions estimated from the output of the linear ICA model. The five talkers were localized within approximately 5 degrees on average, which is clearly less precise than what was obtained in the previous simulations. This can be explained by the poorer quality of the HOA signals obtained with the actual microphone array, due to mismatches between the sensors and errors in the modeling of the SMA. In addition, the HOA signals were high-pass filtered with a cutoff frequency of 750 Hz prior to running the ICA, which means that important features of the speech signals were not available in the analysis. As was the case previously, the ICA source localization seems to outperform the MUSIC source localization, which failed to identify two of the source directions.

Table VI. The true and estimated source directions for the five simultaneous talkers are shown for the simulation based on the measured impulse responses. Other details are as in Table I.

Talker	(ϑ, φ) True	(ϑ, φ) ICA	(ϑ, φ) MUSIC
1	(90, 45)	(87, 48)	(84, 56)
2	(90, −45)	(88, −43)	·
3	(90, 90)	(90, 97)	(82, 94)
4	(90, −90)	(88, −93)	(85, −95)
5	(90, 180)	(88, 180)	·

Table VII. The PESQ scores for talkers 1 to 5 and the average PESQ score obtained in the simulation based on the measured impulse responses are shown. Other details as described in Table II.

Method	Talker 1	2	3	4	5	avg.
Beam.	1.8	1.8	1.7	1.7	0.8	1.6
ICA 1	1.3	1.6	1.7	1.6	1.4	1.5
ICA 2	1.0	1.6	1.7	1.6	1.4	1.4
ICA 3	0.5	1.6	1.2	1.1	1.2	1.1

In order to assess the performance of the source separation, we calculated the PESQ scores for the signals separated using the three proposed algorithms. We then compared these scores to the ones obtained with a second-order spherical beamformer steered in the actual talker directions. The results are shown in Table VII. Due to the greater amount of reverberation in the room, the PESQ scores are lower than the ones obtained in the previous simulations for all four methods.

Unlike the previous simulations the beamformer performed better than the three proposed algorithms on average. The reason for this is that the minimum angular distance between the talkers (45°) was greater than in the previous simulations: 20° and 30° in the anechoic and reverberant conditions, respectively. In other words, in the previous simulations, the beam pattern of the order-2 spherical beamformer was too wide to separate the talkers located very close to each other, and the corresponding PESQ scores were low.

Interestingly, it can be noted that the inverse filtering algorithm performed rather poorly in this simulation. The explanation for this is that the inverse filters are calculated to keep the target direction undistorted, while canceling the sounds from the other incoming talker directions. For other directions, however, the filters are not constrained and may not cancel interfering sources or even amplify them. In the previous simulations, because the reverberation was moderate or not present, the talkers were then

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

Table VIII. The Signal to Interference Ratio Improvements (SIRI) corresponding to the numerical experiments described in sections 4–6 are shown. The values have been averaged for every combination of target signal and microphone. Other details are as described in Table II.

Method	Condition		
	Anechoic	Reverberant	Measured
Beam.	9.8	11.8	10.7
ICA 1	20.0	15.5	12.9
ICA 2	10.4	12.9	12.4
ICA 3	23.6	15.7	5.8

the main interferers and the inverse filters could separate the signals effectively. In contrast, the amount of reverberation was much greater for this simulation resulting in a large portion of the noise consisting of reverberated waves which the inverse filters failed to cancel.

7. Conclusions

In this paper, we described the application of a linear ICA model to the HOA domain signals recorded by a spherical microphone array. One of the reasons for exploring the linear ICA model is that, in the HOA domain, plane-wave sources and their reflections form a linear, instantaneous mixture. We have shown that spherical sources break the assumption of instantaneous mixing in the HOA domain, but that depending on the source distance and frequency, the plane-wave approximation may still hold.

One of the main, new considerations of this paper are the advantages that arise when applying ICA to a set of microphone signals for which the array steering vector is explicitly known and made available. We have shown that projection of the ICA mixing matrix onto a plane-wave dictionary based on the array steering vector provides a method for localizing the source directions and resolving the permutation problem. Typically, one naturally associates array steering vectors with beamforming. Nonetheless, the application of statistical independence and non-Gaussianity to separate and localize sources has generally not formed a dominant part of the beamforming literature. Instead, the literature more often focuses on subspace algorithms and/or the spatial correlation matrix. Using simulated and measured impulse responses for dual-concentric SMAs, in both anechoic and reverberant conditions, we have shown that source localization based on the standard, linear ICA model seems to outperform a standard, narrow-band MUSIC algorithm. It is important to note that both the ICA model and the MUSIC algorithm were applied to the same narrow-band HOA signals. These results suggest that higher-order statistics may have advantages for source localization with beamformers compared with standard techniques. Sun *et al.* in [22] have applied the ESPRIT algorithm for reverberant source localization in the phase-mode domain using a rigid SMA. While it is difficu lt to directly compare results, the ESPRIT algorith m employs second-order statistics similar to MUSIC, and likely performs similarly. Furthermore, the linear ICA model makes no *a priori* assumptions on the number of sources, as is required for the MUSIC and ESPRIT algorithms, and can be implemented efficiently.

With regard to source separation, the application of blind source separation techniques to a beamformer, such as the SMAs described in this paper, is usually considered as adaptive beamforming [3, 23]. Our approach differs, however, from typical blind source separation techniques in that we first determine the source directions from the ICA mixing matrix. This provides additional options for source separation. In other words, we compared the ICA source separation results with a standard, second-order spherical beamformer and also with a direct inverse-filtering approach based on the determined source directions. Taking a speech enhancement viewpoint, we presented our results in terms of the PESQ scores. Table VIII, on the other hand, provides a summary of the results using the Signal to Interference Ratio Improvement (SIRI), which is a signal quality measure more commonly used in the source separation literature (see for instance [23], page 221). The SIRIs confirm the PESQ results: when the reverberation is moderate, the inverse-filtering approach (method ICA 3) separates the sources more effectively; however when more reverberation is present, as with the measured impulse responses, the performance of the inverse-filtering method drops dramatically. As well, the SIRIs confirm that when the source directions are too close for the beamformer to resolve, *e.g.* in the anechoic condition, the standard ICA source separation (method ICA 1) outperforms the spherical beamformer.

The quality of the HOA signals is clearly a significant factor for the proposed methods and highlights the importance of the microphone array design. For instance, being able to separate up to 8 talkers requires high quality HOA signals up to order 2 for a relatively large frequency range. These performance characteristics can be achieved using a dual, concentric array design, such as that described in [24].

In this work, we only considered a simple, linear ICA model. In future work, we intend to investigate convolutive ICA models applied in a similar manner to SMAs. As well, in this study, we assessed the quality of the separated signals via PESQ scores. While the scores obtained in the simulations are nearly equal to each other, the perceived quality of these signals is diverse. Thus, a perceptual assessment of the separated signals is required in order to evaluate the performance of the three proposed methods. The different sounds used and generated in the simulations are fully downloadable from the CARLab website [25].

References

[1] A. Hyvärinen, J. Karhunen, E. Oja: Independent component analysis. Wiley Interscience, New York, USA, 2001.

[2] H. V. Trees: Optimum array processing. Wiley, New York, 2002, 1155–1194.

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only

[3] W. Liu: Blind adaptive beamforming for wideband circular arrays. Proceedings of the ICASSP 2009, Taiwan, April 2009.

[4] N. Epain, C. Jin, A. van Schaik: Blind source separation using independent component analysis in the spherical harmonic domain. Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics, Paris, May 2010.

[5] J. Benesty, J. Chen, Y. Huang: Microphone array signal processing. Springer-Verlag, Berlin, 2008.

[6] E. Williams: Fourier acoustics: Sound radiation and near-field acoustic holography. Academic Press, London, UK, 1999, 218.

[7] R. Kennedy, P. Sadeghi, T. Abhayapala, H. Jones: Intrinsic limits of dimensionality and richness in random multipath fields. IEEE Transactions on signal processing **55** (June 2007) 2542–2556.

[8] J. Daniel: Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. Dissertation. Université Paris 6, Paris, France, 2000.

[9] J. Daniel: Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. Proceedings of the AES 23rd International Conference, Copenhagen, May 2003.

[10] A. Parthy, C. Jin, A. van Schaik: Evaluation of a concentric rigid and open spherical microphone array for sound reproduction. Proceedings of the 1st International Symposium on Ambisonics and Spherical Acoustics, Graz, June 2009.

[11] N. Epain, J. Daniel: Improving spherical microphone arrays. Proceedings of the 124th AES Convention, Amsterdam, May 2008.

[12] B. Rafaely: Analysis and design of spherical microphone arrays. IEEE Transactions on speech and audio processing **13** (January 2005) 135–143.

[13] H. Engl, M. Hanke, A. Neubauer: Regularization of inverse problems. Kluwer Academic Publishers, 2000, Ch. 5.

[14] E. Fisher, B. Rafaely: Near-field spherical microphone array processing with radial filtering. IEEE Transactions on audio, speech and language processing **19** (February 2011) 256–265.

[15] A. Parthy, C. Jin, A. van Schaik: Acoustic holography with a concentric rigid and open spherical microphone array. Proceedings of the ICASSP 2009, Taiwan, April 2009.

[16] H. Gävert, J. Hurri, J. Särelä, A. Hyvärinen: The FastICA package for MATLAB. http://www.cis.hut.fi/projects/ica/fastica/.

[17] ITU-T: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. ITU-T Recommendation P.862, 2001.

[18] P. Loizou: Speech enhancement: theory and practice. CRC Press, Boca Raton, USA, 2007.

[19] J. Meyer, T. Agnello: Spherical microphone array for spatial sound recording. Proceedings of the 115th AES convention, New York, USA, October 2003.

[20] A. Wabnitz, N. Epain, C. Jin, A. van Schaik: Room acoustics simulation for multichannel microphone arrays. Proceedings of the International Symposium on Room Acoustics, ISRA 2010, Melbourne, Australia, August 2010.

[21] A. Wabnitz, N. Epain: The MCROOMSIM package for MATLAB. http://www.ee.usyd.edu.au/carlab/mcroomsim.htm.

[22] H. Sun, H. Teutsch, E. Mabande, W. Kellermann: Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays. Proceedings of the ICASSP 2011, May 2011, 117–120.

[23] S. Makino: Blind source separation of convolutive mixtures of speech. – In: Adaptive Signal Processing. J. Benesty, Y. Huang (eds.). Springer-Verlag, Berlin, 2003, Kap. 7.

[24] A. Parthy, C. Jin, A. van Schaik: Optimisation of co-centred rigid and open spherical microphone arrays. Proceedings of the AES 120th Convention, Paris, France, May 2006.

[25] N. Epain, C. Jin: The accompanying sounds for this paper are available at the following url. http://www.ee.usyd.edu.au/carlab/smaica.htm.

Document downloaded by @DAEL, 2026-11-15 07:19:49 - personal use only