An Improved Two-Stage Spherical Harmonic ESPRIT-Type Algorithm

Haocheng Zhou 1,2, Zhenghong Liu 1,2,*, Liyan Luo 1,2, Mei Wang 3 and Xiyu Song 1,2

1 School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China; z290971984@163.com (H.Z.); lly1227@guet.edu.cn (L.L.); songxiyu@guet.edu.cn (X.S.)
2 Ministry of Education Key Laboratory of Cognitive Radio and Information Processing, Guilin 541004, China
3 School of Information Science & Engineering, Guilin University of Technology, Guilin 541006, China; mwang@glut.edu.cn
* Correspondence: giet.liu@163.com; Tel.: +86-18978381211

Abstract: Sensor arrays are gradually becoming a current research hotspot due to their flexible beam control, high signal gain, robustness against extreme interference, and high spatial resolution. Among them, spherical microphone arrays with complex rotational symmetry can capture more sound field information than planar arrays and can convert the collected multiple speech signals into the spherical harmonic domain for processing through spherical modal decomposition. The subspace class direction of arrival (DOA) estimation algorithm is sensitive to noise and reverberation, and its performance can be improved by introducing relative sound pressure and frequency-smoothing techniques. The introduction of the relative sound pressure can increase the difference between the eigenvalues corresponding to the signal subspace and the noise subspace, which is helpful to estimate the number of active sound sources. The eigenbeam estimation of signal parameters via the rotational invariance technique (EB-ESPRIT) is a well-known subspace-based algorithm for a spherical microphone array. The EB-ESPRIT cannot estimate the DOA when the elevation angle approaches 90°. Huang et al. proposed a two-step ESPRIT (TS-ESPRIT) algorithm to solve this problem. The TS-ESPRIT algorithm estimates the elevation and azimuth angles of the signal independently, so there is a problem with DOA parameter pairing. In this paper, the DOA parameter pairing problem of the TS-ESPRIT algorithm is solved by introducing generalized eigenvalue decomposition without increasing the computation of the algorithm. At the same time, the estimation of the elevation angle is given by the arctan function, which increases the estimation accuracy of the elevation angle of the algorithm. The robustness of the algorithm in a noisy environment is also enhanced by introducing the relative sound pressure into the algorithm. Finally, the simulation and field-testing results show that the proposed method not only solves the problem of DOA parameter pairing, but also outperforms the traditional methods in DOA estimation accuracy.

Keywords: spherical microphone array; DOA; relative sound pressure; ESPRIT; spherical harmonics domain

1. Introduction

Multisource DOA estimation based on microphone arrays is a current research hotspot [1], and the estimation results are a priori knowledge for many acoustic applications such as speech separation [2], speech enhancement [3], and multisource localization [4]. A spherical microphone array has a three-dimensional symmetric structure, which allows us to estimate the orientation of a sound source at any position in three-dimensional space [5]. The spherical microphone array enables sampling of non-confounding sound pressure to reconstruct the 3D sound field accurately. The sound pressure received by the spherical microphone array can be easily decomposed into the spherical harmonic domain for processing [6]. The frequency and angular components of the spherical harmonic signal obtained through spherical harmonic decomposition are decoupled from each other,
which is a major advantage of using the spherical harmonic domain as the space domain. Therefore, in recent years, sound source localization based on the spherical microphone array has received great attention.

Traditional DOA estimation methods mainly include: the spatial spectrum estimation algorithm based on conventional beamforming (CBF) [7]; the Capon method, also known as the minimum variance distortion response (MVDR) algorithm or the minimum variance method (MVM) [8]; the DOA estimation method based on TDOA [9]; the DOA estimation method based on steered-response power-phase transform [10]; etc. Traditional DOA estimation methods cannot break through the Rayleigh limit (Fourier limit). Subspace algorithms represented by the MUSIC algorithm and ESPRIT algorithm [11–17] have been proposed to solve this problem. Therefore, subspace algorithms are receiving great attention. However, such methods are sensitive to noise and reverberation and have large calculations. These disadvantages have a great impact on the practical applications of the algorithms and are worth in-depth consideration.

Due to the advantages of processing in the spherical harmonic domain, many DOA estimation techniques are being revisited in the spherical harmonic domain. For example, the steered-response power method manipulates the beamformer on the grid and checks the peak power output as an estimate of the DOA. In the simplest situations, a plane-wave decomposition (PWD) beamformer can be steered in multiple directions on a grid, forming a power map. Then, the peak finding on the map can be used to determine potential source DOAs. The spatial resolution of this approach is limited by the available order of the SH signals, and for low orders, the resolution is poor [18]. Another alternative is to utilize the acoustic intensity vector obtained from the first-order signals, whose temporal and spatial statistics reveal information about the presence and distribution of the sound sources [19]. There are more approaches, such as the sparse recovery approach [20], the GCC (generalized cross-correlation) technique [21], and the adaptive eigenvalue decomposition method [22]. In addition, subspace algorithms have received increasing attention due to their high resolution and noise suppression capabilities [23]. Examples include the eigenbeam multiple signal classification (EB-MUSIC) [8,11] algorithms, which exploit the orthogonality between signal and noise subspaces, and the eigenbeam estimation algorithms for signal parameters via the rotational invariance technique (EB-ESPRIT), which exploit the rotation invariance properties of subsets of array channels [12]. Among these methods, EB-ESPRIT approaches estimate DOA parameters directly using the recurrence relations of spherical harmonic functions without beam steering or a peak search, which makes them well suited for real-time speech applications. Despite these advantages, EB-ESPRIT algorithms fail when the elevation angle approaches 90°. The abovementioned subspace algorithms, including the EB-ESPRIT algorithms, also require a priori knowledge about the number of active sound sources.

Many attempts have been made to resolve the singularity problem of EB-ESPRIT. Sun et al. [12] proposed a robust sound source localization method that rotates the eigenbeam space using a Wigner-D weighting matrix. However, the method requires a priori knowledge of the region of interest to guide the rotation. Huang et al. [13] proposed a two-step EB-ESPRIT algorithm (TS-ESPRIT), which uses different recurrence relations to estimate the elevation and azimuth angles separately. However, the elevation and azimuth angles estimated with this method are not automatically paired. Byeongho Jo et al. proposed a nonsingular EB-ESPRIT approach based on two recurrence relationships. In this method, the elevation angle is parameterized with a sine function, which avoids the ill-conditioning problem in the EB-ESPRIT method [24]. However, the sine function is symmetric with respect to an elevation angle of 90°, which makes the results obtained by this method ambiguous and requires further processing to obtain the correct results. In [23], Byeongho Jo et al. used three independent recurrence relations to extract three directional parameters that uniquely specify a directional vector and solved the pairing problem between the
three directional parameters by solving a generalized eigenvalue problem to determine the common transformation matrix between the signal subspace and the spherical harmonic matrix.

Noise and reverberation may exist in real indoor environments and are detrimental to localization algorithms. For example, high reverberation can cause spatial correlation matrix rank loss, which will cause the localization algorithm to fail in severe cases. To address this problem, Rafaely et al. [16] introduced frequency-smoothing techniques in the spherical harmonic domain MUSIC (SHD-MUSIC) algorithm to remove the correlation among correlated signal sources. Similarly, strong noise can affect the estimation of the signal subspace, leading to a decrease in the estimation accuracy of the localization algorithm. In particular, the localization algorithm in the spherical harmonic domain is more sensitive to noise, because when the modal strength of the array is compensated at low frequencies, the noise is amplified [25]. In view of the sensitivity of the localization algorithm in the spherical harmonic domain to noise and the “Bessel zero problem” [26] in the open sphere configuration, to address the problem, Yonggang Hu et al. [17] proposed a MUSIC algorithm in the spherical harmonic domain based on relative sound pressure, which improved the performance of the localization algorithm in noisy environments. By introducing relative sound pressure, the difference between the eigenvalues corresponding to the signal subspace and the eigenvalues corresponding to the noise subspace was increased under certain conditions, which was conducive to estimating the number of current active sound sources. On the other hand, to improve the performance of DOA estimation in strongly reverberant environments, an improved DPD test method [27] was proposed. The method identifies the time-frequency bins dominated by the direct sound by means of enhanced decomposition based on the direct sound. Then these time-frequency bins are used to construct the spatial spectrum. However, the DPD test requires a direct acoustic prior [28], which is often difficult to obtain in harsh sound field environments.

In this paper our contributions are the following:

- We introduce relative sound pressure and frequency-smoothing techniques into the TS-ESPRIT algorithm [14] to improve the localization accuracy of the algorithm under a low signal-to-noise ratio and high reverberation;
- By solving a generalized eigenvalue problem, we solve the pairing problem and eliminate the two eigenvalue decomposition operations for estimating DOA parameters in the original paper;
- We use the inverse tangent function to find the elevation estimation and achieve superior estimation accuracy.

The remainder of the paper is organized as follows: the definition of relative sound pressure and its corresponding signal model in the spherical harmonic domain are given in Section 2. The original TS-ESPRIT algorithm is reviewed and its improvement is given in Section 3. Extensive experimental results are given in Section 4. Finally, the paper is concluded in Section 5.

2. System Models

2.1. Definition of the Relative Sound Pressure

This paper adopts the standard Cartesian coordinate system \((x, y, z)\) and spherical coordinate system \((r, \phi, \theta)\), where the elevation angle \(\theta\) and azimuth angle \(\phi\) are measured in radians from the positive z-axis and positive x-axis, respectively. Consider a spherical array with \(M\) microphones distributed on the surface and a radius \(r\) placed in the space, and the center of the array coincides with the origin of the coordinate system; then, the coordinates of the microphones near the origin can be given as \(x_j = (r, \phi_j, \theta_j)\), \(j = 1, \ldots, M\). Assume that there are \(L\) far-field plane waves in space, located at an angle \(\Psi_j = (\phi_j, \theta_j)\),
where $i$ denotes an imaginary unit, wavenumber $k = \frac{2\pi f}{c}$; $f$ is the frequency; $c$ is the speed of sound; $\mathcal{P}(x_j, k)$ and $P(x_j, k)$ denote noisy and noise-free sound pressure, respectively; $s_i(k)$ represents the $i$th signal observed at the origin $x_o = (0, 0, 0)$; $n(x_j, k)$ is additive noise at the $j$th microphone; and the wavenumber vector is represented by $k_i = (k\cos \phi \sin \theta_i, k\sin \phi \sin \theta_i, k\cos \theta_i)^T$.

From [16], the relative sound pressure of the $j$th microphone with respect to the sound pressure at the origin $x_o = (0, 0, 0)$ is defined as

$$Q(x_j, k) = \frac{P(x_j, k)}{P(x_o, k)}.$$  \hfill (2)

In practice, because the sound pressure at the center of the rigid sphere is not easily obtained, we approximate the pressure at the array origin by averaging the pressure values of the $M$ microphones on the surface of the array as [17]

$$\mathcal{P}(x_o, k) \approx \frac{1}{M} \sum_{j=1}^{M} \mathcal{P}(x_j, k).$$  \hfill (3)

The direct calculation of the relative sound pressure using the ratio of two microphone sound pressures in a noisy environment contains nonnegligible errors, especially when the denominator in (3) is small. An alternative estimate of the relative sound pressures in the presence and absence of noise is given in [17]

$$Q(x_j, k) = \frac{S_{\mathcal{P}, \mathcal{P}_o}(k)}{S_{\mathcal{P}_o, \mathcal{P}_o}(k)} = \frac{E\{P(x_j, k) P^*(x_o, k)\}}{E\{P(x_o, k) P^*(x_o, k)\}},$$  \hfill (4)

$$\mathcal{Q}(x_j, k) = \frac{S_{\mathcal{P}, \mathcal{P}_o}(k)}{S_{\mathcal{P}_o, \mathcal{P}_o}(k)} = \frac{S_{\mathcal{P}, \mathcal{P}_o}}{S_{\mathcal{P}_o, \mathcal{P}_o} + S_{n_o, n_o}}.$$  \hfill (5)

Dividing (5) by (4), we derive the following relation between the noisy and noiseless relative sound pressure:

$$\mathcal{Q}(x_j, k) = Q(x_j, k) \gamma(k),$$  \hfill (6)

where

$$\gamma(k) = \frac{T(x_o, k)}{T(x_o, k) + 1}$$  \hfill (7)

depends only on the signal-to-noise ratio (SNR) at the origin of the array; i.e., $T(x_o, k) = \frac{S_{\mathcal{P}, \mathcal{P}_o}(k)}{S_{n_o, n_o}(k)}$.

Substituting the far-field plane wave model into Equation (2) yields

$$Q(x_j, k) = \frac{\sum_{l=1}^{L} s_l(k) e^{ik^T x_j}}{\sum_{l=1}^{L} s_l(k) e^{ik^T x_o}} = \frac{\sum_{l=1}^{L} s_l(k) e^{ik^T x_j}}{\sum_{l=1}^{L} s_l(k)},$$  \hfill (8)

where

$$s_l(k) = \frac{s_l(k)}{\sum_{l=1}^{L} s_l(k)}.$$  \hfill (9)
2.2. Spherical Harmonic Decomposition and Signal Model in the Spherical Harmonic Domain

From [25], the unit amplitude plane wave can be expanded into the following form:

\[ e^{ikr}\tau_j = \sum_{n=0}^{N} \sum_{m=-n}^{n} b_n(kr) [Y^m_n(\phi_j, \theta_j)]^* Y^m_n(\phi_j, \theta_j), \]  

(10)

where \( Y^m_n(\phi, \theta) \) denotes the spherical harmonics of order \( n \) and degree \( m \) as follows:

\[ Y^m_n(\phi, \theta) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P^m_n(\cos \theta)e^{im\phi}. \]  

(11)

In Equation (10), \( N \) denotes the highest order of spherical harmonic decomposition, \( 0 \leq n \leq N \), \(-n \leq m \leq n\), \( P^m_n \) denotes the associated Legendre functions of order \( n \) and degree \( m \), \( b_n(kr) \) indicates the far-field mode strength associated with the array configuration, and the commonly used array configurations are the open sphere and rigid sphere,

\[ b_n(kr) = \begin{cases} 4\pi i^m j_n(ka) & \text{open sphere} \\ 4\pi i^m j_n(ka) - \frac{\phi'_m(ka)}{h^2_m(ka)} h^2_m(ka) & \text{rigid sphere}, \end{cases} \]  

(12)

where \( j_n \) and \( h^2_m \) are, respectively, the \( n \)-th order spherical Bessel function and the \( n \)-th order second-class spherical Hankel function of class II, and \( j'_m(ka) \) and \( h^2_m(ka) \) represent the first derivatives of the corresponding functions, respectively.

According to Equations (10) and (8), we can expand the noise-free relative sound pressure at the \( j \)-th microphone into the following form:

\[ Q(x_j, k) = \sum_{l=1}^{L} \sum_{m=-n}^{n} \sum_{l=-n}^{n} \tilde{\tau}_l(k) b_n(kr) [Y^m_n(\phi_l, \theta_l)]^* Y^m_n(\phi_j, \theta_j). \]  

(13)

Applying the discrete spherical Fourier transform to Equation (13) yields

\[ q_{nm}(k) = \sum_{j=1}^{M} a_j Q(x_j, k) [Y^m_n(\phi_j, \theta_j)]^*, \]  

(14)

where \( a_j \) is the sampling weight determined by the sampling schemes; for a uniform sampling scheme, \( a_j = \frac{4\pi}{M} \) [29]. The derivation of Equation (14) uses the following relation:

\[ \sum_{j=1}^{M} a_j Y^m_n(\phi_j, \theta_j) Y^m_n(\phi_j, \theta_j) = \delta_{nm}\delta_{mm'}, \]  

(15)

where \( \delta_{nm} \) is equal to 1 when \( n = n' \); otherwise, it is equal to 0. Equation (14) can be written in the following matrix form:

\[ \bar{q}(k) = B(kr) \bar{y}(k), \]  

(16)

where \( \bar{q}(k) \) is the \((N+1)^2 \times 1\) dimensional noise-free spherical harmonic coefficient vector

\[ \bar{q}(k) = \begin{bmatrix} q_{00}(k), q_{11}(k), q_{10}(k), q_{11}(k), \ldots, q_{N-N}(k), \ldots, q_{NN}(k) \end{bmatrix}^T, \]  

(17)

and

\[ B(kr) = \text{diag} \begin{bmatrix} b_0(kr), b_1(kr), b_2(kr), b_3(kr), \ldots, b_N(kr), \ldots, b_N(kr) \end{bmatrix}. \]  

(18)
Moreover, \( Y = [a(\Psi_1), \ldots, a(\Psi_L)]^T \) is a \( L \times (N + 1)^2 \) dimensional spherical harmonic matrix, where
\[
a(\Psi_l) = \left[ Y_0^0(\phi_l, \theta_l), Y_1^{-1}(\phi_l, \theta_l), Y_1^0(\phi_l, \theta_l), Y_1^1(\phi_l, \theta_l), \ldots, Y_N^{-N}(\phi_l, \theta_l), \ldots, Y_N^N(\phi_l, \theta_l) \right]^T.
\] (19)

Finally, \( \bar{s}(k) = [\bar{s}_1(k), \ldots, \bar{s}_L(k)]^T \) is the signal vector.

We assume that the configuration of the spherical array is known; therefore, we can simply divide Equation (14) by \( B_n(kr) \) or multiply Equation (16) from the left by \( B^{-1}(kr) \) to get
\[
\tilde{q}(k) = A\bar{s}(k),
\] (20)

where \( \tilde{q}(k) = B^{-1}(kr)\bar{q}(k) \) and \( A = \Psi^H(\Psi) \). Combining Equations (6) and (20), we can express the spherical harmonic coefficient of the relative sound pressure with noise as the following form:
\[
q(k) = \tilde{q}(k)\gamma(k).
\] (21)

Then, the correlation matrix of the spherical harmonic coefficients can be written as
\[
S(k) = E\{q(k)q^H(k)\} = AR_s(k)A^H,
\] (22)

where \( R_s(k) \) is the covariance matrix, which is assumed to be a full-rank matrix in the subspace approach. However, in practice, this assumption does not always hold because multiple signals may be correlated when there is reverberation in the space. For this reason, we can implement a frequency-smoothing technique to decorrelate multiple correlated signals, and the smoothed covariance matrix is the average of the covariance matrix over multiple frequency bins [16]:
\[
\bar{S} = \frac{1}{K} \sum_{k=1}^{K} S(k) = A\frac{1}{K} \sum_{k=1}^{K} R_s(k)A^H.
\] (23)

The signal subspace and noise subspace can be obtained by performing the eigenvalue decomposition of \( \bar{S} \):
\[
\bar{S} = [U_s \ U_n] \left[ \begin{array}{cc} \Sigma_s & 0 \\ 0 & 0 \end{array} \right] \left[ \begin{array}{c} U_s^H \\ U_n^H \end{array} \right],
\] (24)

where \( \Sigma_s = diag(\sigma_1, \ldots, \sigma_L) \) contains \( L \) nonzero eigenvalues, and \( U_s \) and \( U_n \) correspond to the signal subspace and noise subspace, respectively. From this, we can see that ideally, after introducing the relative sound pressure, the eigenvalues corresponding to the noise subspace are zero, and the difference between them and the eigenvalues corresponding to the signal subspace becomes larger. This difference makes it easier to estimate the number of sound sources from the sorted eigenvalues so that a priori knowledge about the number of sound sources is no longer needed.

From Equations (23) and (24), we obtain \( \text{span}\{A\} = \text{span}\{U_s\} \). Thus, the signal subspace and the steering matrix can be related by a nonsingular matrix \( T \):
\[
U_s = AT.
\] (25)

3. Method

The TS-ESPRIT algorithm uses two different recurrence relations about the spherical harmonic to estimate the elevation and azimuth angles, and three types of TS-ESPRIT algorithms are given in [14] based on different signal models. The improvements in this
paper are valid for all three types of algorithms. Here, we present only the TS-ESPRIT algorithm based on the complex spherical harmonic model for simplicity.

To estimate the elevation angle, the TS-ESPRIT algorithm uses the following recurrence relation [15]:

$$
cos \theta Y_n^m = K_n^m Y_{n-1}^m + K_{n+1}^m Y_{n+1}^m
$$

(26)

where $$K_n^m = \left(\frac{(n+m)(n-m)}{(2n+1)(2n-1)}\right)^{1/2}$$ [30]. Next, the terms satisfying the above relationship are selected from each column of $$A$$, corresponding to the three submatrices $$\tilde{A}^0 = O^\dagger A$$, where $$O^\dagger$$ is the selection matrix and $$q = 0, 1, 2$$. The $$l$$th column of $$A$$ in Equation (19) can be expressed as $$a_l = [a_l^0, a_l^1, a_l^2]^T$$, where $$a_l^q = [Y_l^0, Y_l^1, Y_l^2]^T$$. To clearly illustrate the selection rules of the submatrices, we define $$n_0 = 0, \ldots, N-2$$, $$n_1 = 1, \ldots, N-1$$, and $$n_2 = 2, \ldots, N$$. Thus, $$O^0$$ selects all elements in $$a_l^0$$, $$O^1$$ selects elements from the second row to the $$2n_1$$th row in $$a_l^1$$, and $$O^2$$ selects $$2n_2 - 3$$ elements from the third row to the $$(2n_2 - 1)$$th row in $$a_l^2$$. Then, Equation (26) can be written in the matrix form as follows:

$$
\tilde{A}^1 \Gamma_e = A^0 \tilde{A}^0 + A^1 \tilde{A}^2,
$$

(27)

where

$$
\Gamma_e = \text{diag}(\cos \theta_1, \ldots, \cos \theta_L),
$$

(28)

$$
A^0 = \text{diag}\left(\frac{K_0^0}{K_2^0}, K_2^0, \ldots, K_{N-2}^0, K_{N-2}^0, \ldots, K_N^0\right),
$$

(29)

and

$$
A^1 = \text{diag}\left(\frac{K_2^0}{K_2^0}, K_2^0, K_2^0, \ldots, K_{N-2}^0, K_{N-2}^0, \ldots, K_N^0\right).
$$

(30)

Similarly, we select three submatrices from the signal subspace in the same way as above; that is, $$\tilde{Z}^l = O^\dagger U_l = O^\dagger AT = \tilde{A}^l T$$, and then the following is obtained,

$$
\tilde{Z} \Phi_e = A^0 \tilde{Z}^0 + A^1 \tilde{Z}^2,
$$

(31)

where $$\Phi_e = T^{-1} \Gamma_e T$$. By taking the pseudoinverse, we can obtain the minimum mean square estimate, as follows:

$$
\Phi_e = \left(\tilde{Z}^l\right)^+ \left(A^0 \tilde{Z}^0 + A^1 \tilde{Z}^2\right).
$$

(32)

In Equation (32), $$\tilde{Z}^l \in \mathbb{C}^{(N-1)^2 \times L}$$ is a $$(N - 1)^2 \times L$$ dimensional matrix, so we can estimate up to $$(N - 1)^2$$ elevation angles.

The following recurrence relation on the spherical harmonic is used to estimate the azimuth angle [15]:

$$
e^{-i\phi} \sin \theta Y_n^m = -j_n^m Y_{n-1}^{m-1} + j_{n+1}^{m+1} Y_{n+1}^{m-1},
$$

(33)

where $$j_n^m = \left(\frac{(n+m)(n-m-1)}{(2n+1)(2n-1)}\right)^{1/2}$$ [30]. Similarly, the items satisfying Equation (33) are selected from each column of $$A$$ to form three submatrices: $$\tilde{A} = N^\dagger A$$, $$p = -1, 0, 1$$, and $$N^\dagger$$ from the selection matrix. Similar to $$O^\dagger$$, $$N^\dagger$$ selects all elements in $$a_l^0$$, $$N^0$$ selects $$2n_1 - 1$$ elements from the third row to the $$(2n_1 + 1)$$th row in $$a_l^1$$, and $$N^1$$ selects $$2n_2 - 3$$ elements from the third row to the $$(2n_2 - 1)$$th row in $$a_l^2$$. Then, Equation (33) can be written in the matrix form as follows:

$$
\tilde{A} \Gamma_a = -j_0 \tilde{A}^{0-1} + j_1 \tilde{A}^{1},
$$

(34)

where

$$
\Gamma_a = \text{diag}\left(e^{-i\Phi_1 \sin \theta_1}, \ldots, e^{-i\Phi_L \sin \theta_L}\right),
$$

(35)
where \( H \) and \( J \) are the right generalized eigenvector and the generalized eigenvalue of the given problem, respectively. When \( H \) is nonsingular, the GEV can be transformed into an eigenvalue decomposition problem of matrix \( \Phi_e^{-1} \Phi_c \), whose eigenvalues are given by \( e^{-i\theta} \tan \theta \). Therefore, we find a unique transformation matrix \( T_y = H^{-1} \) that can simultaneously diagonalize the matrices \( \Phi_c \) and \( \Phi_a \). Even when the matrix \( H \) is singular, GEV can be solved using the QZ algorithm [32]. The QZ algorithm is used to iteratively compute the eigenvalues of two matrices simultaneously in ascending order using a closed-form solution, and the eigenvectors are computed byproducts of the accumulated unitary matrices. When using the QZ algorithm to solve the generalized eigenvalue problem, the inverse of the matrix does not need to be computed, so the QZ algorithm can be used to solve the above generalized eigenvalue problem regardless of \( H \) and obtain matrix \( E \) containing the generalized eigenvalues.

Finally, using the generalized eigenvalue matrix \( E \), the estimates of the elevation and azimuth angles can be found with the following equation:

\[
\phi = \arg(\text{diag}(E)),
\]

\[
\hat{\theta} = \text{atan} \left( \frac{\text{real}(\text{diag}(E))}{\text{real}(e^{-i\phi})} \right),
\]
where ./ denotes the division of the elements in the corresponding positions of the matrix or vector on the left and right sides, \( \phi = [\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_L] \) and \( \theta = [\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_L] \) contain estimates of the azimuth and elevation angles of the sound source, respectively, and the direction parameters at their corresponding positions are automatically paired. Thus, we solve the problem of DOA parameter pairing in TS-ESPRIT. Here, the estimate of the elevation angle is given by the arctan function, which has the steepest rate of change with respect to the angle among the trigonometric functions. Therefore, the accuracy of the conversion regarding the elevation angle in this paper is more accurate than that of the original TS-ESPRIT algorithm (see A in Section 4.1.2).

4. Experiment and Analysis

In this section, simulation examples are given to illustrate the superiority of the proposed algorithm compared with the TS-ESPRIT algorithm [14] and the sine-based ESPRIT algorithm [24]. Firstly, we compare the TS-ESPRIT algorithm proposed in this paper without the pairing problem (referred to as NPTS-ESPRIT) with the traditional TS-ESPRIT algorithm and the sine-based ESPRIT algorithm under a far-field signal to verify the effectiveness of the proposed algorithm. Secondly, we compare the NPTS-ESPRIT algorithm with the introduction of relative sound pressure (referred to as RNPTS-ESPRIT) and the NPTS-ESPRIT algorithm under a far-field wideband signal to verify the influence of the introduction of relative sound pressure on the anti-noise performance and the estimation of the number of active sound sources of the proposed algorithm. Finally, we verify the effectiveness of the proposed algorithm using field measurement data collected by real rigid spherical arrays.

To fairly evaluate the algorithms, we use the mean absolute estimated error (MAEE) between their estimated and original source DOAs to measure the performance. MAEE is defined as follows:

\[
\text{MAEE} = \frac{1}{2L} \left( \sum_{i=1}^{L} \left| \theta_{ori}^{i}(l) - \theta_{est}^{i}(l) \right| + \left| \phi_{ori}^{i} - \phi_{est}^{i} \right| \right),
\]

which characterizes the average numerical accuracy of \( I \) independent tests.

4.1. Comparison with TS-ESPRIT and Sine-Based ESPRIT

4.1.1. Free-Field Simulation: Setup

In the following simulation, a rigid spherical array with a radius of 4.2 cm and 32 sensors uniformly distributed on the surface is used, which can decompose sound pressure into spherical harmonic coefficients up to the fourth order [5]. We assume that the narrowband sound source is located in the far field relative to the microphone array and emits uncorrelated Gaussian white noise with equal power. Considering the modal compensation and aliasing error of the spherical microphone array, we set \( kr = 2 \). The signal-to-noise ratio (SNR) is defined as the ratio of the total noise power (\( \sigma_n^2 \)) to the total power of the source signals \( \text{SNR}(dB) = 10\log_{10} \frac{\sum s^2}{\sum \sigma_n^2} \). The self-noise of the microphone is uncorrelated Gaussian white noise.

To simulate spatial aliasing, the pressure data of each microphone on the rigid sphere is generated using spherical harmonics up to a sufficient order (\( N = 20 \)). Then the spherical Fourier transform and mode-strength compensation is applied to the microphone signals to obtain the aliased harmonic coefficients up to order \( N = 4 \). Various numbers of measurement snapshots of harmonic coefficients are used to construct the covariance matrix.

4.1.2. Free-Field Simulation: Results

A. Errors for Various Elevation Angles

In this example, we inspect the MAEEs along the elevation and azimuth separately for a sound source with a fixed azimuth angle \( \phi = 90^\circ \) while the elevation angle is varied from \( \theta = 1^\circ \sim 90^\circ \). The aliasing error is ignored. We construct the covariance matrix
At the same time, we can also see that when the elevation angle is close to 0, the TS-ESRIT algorithm and sine-based ESPRIT algorithm is seriously affected by directional parameters under both high and low SNRs. The reason is that the sine function used for the directional parameter. Because the rate of change in the cosine function is small near \( \theta = 0 \) and \( \theta = 20 \) dB, respectively. As can be seen from the figures, the DOA estimation of the TS-ESPRIT algorithm and sine-based ESPRIT algorithm is seriously affected by directional parameters under both high and low SNRs. The reason is that the sine function used for the directional parameter (see Equation (42)), which always shows a tangent function as the directional parameter.

In the proposed algorithm, we use the tangent function as the directional parameter (see Equation (42)), which always shows a very stable performance and has the minimum estimation error. From Figure 1c,d, we can see that the proposed method can estimate the azimuth well in all possible elevation regions without singularity. The azimuth estimation error of the proposed method is almost the same as that of the TS-ESPRIT algorithm because they use the same directional parameter. At the same time, we can also see that when the elevation angle is close to 0, the estimation accuracy of the azimuth angle of the three algorithms is relatively low, especially in the case of low SNR. The main reason is that when the elevation angle of the sound source is close to 0, the positioning algorithm is very sensitive to the azimuth angle, and even a slight error will lead to a large difference between the estimated value of the azimuth angle and the true value. However, this situation weakens rapidly with an increase in the elevation angle, and when the elevation angle increases to 10, the estimation of the azimuth angle almost returns to normal. In practice, we can try to avoid the occurrence of the above situation by placing the array in an appropriate position.

![Graphs](image1.png)

**Figure 1.** MAEEs for various source elevation angles and their SNRs. (a) Elevation errors at SNR = 0 dB; (b) Elevation errors at SNR = 20 dB; (c) Azimuth errors at SNR = 0 dB; (d) Azimuth errors at SNR = 20 dB.
In addition, it can also be concluded from the figure that the DOA estimation performance of the three algorithms differs greatly at both high and low SNRs (the difference between the two traditional algorithms is particularly large), so it is necessary to introduce additional techniques to enhance the anti-noise performance of the algorithm, for which the relative sound pressure technique is introduced in this paper. In the subsequent part, we give the performance comparison of the proposed algorithm before and after the introduction of relative sound pressure.

B. Errors for Various SNRs

In this simulation, the simultaneous detection performance for two uncorrelated sources is compared. The MAEE of three methods for elevation angle and azimuth angle are compared at different SNRs (0−20 dB). Two incoherent signals are incident on the spherical array from (46°, 43°) and (69°, 135°), respectively. The number of snapshots is set to 300 and the aliasing error is not considered. The corresponding MAEEs of 200 independent experiments performed at different SNRs are shown in Figure 2a,b. We can see that the two traditional algorithms are superior to each other in estimating elevation angle and azimuth angle, and the proposed algorithm has the lowest estimation error at almost all SNRs.

![Figure 2. MAEEs of two incoherent sound sources at different SNRs. (a) Azimuth errors at different SNRs; (b) Elevation errors at different SNRs.](image)

C. Errors for Various Snapshots

The effect of the number of independent snapshots in the construction of covariance matrices is examined by computing MAEEs for a varying number of snapshots (100–500). The two uncorrelated signals are located at (−96°, 46°) and (−103°, 142°), respectively. The SNR is set at 10 dB and the aliasing error is ignored. The MAEE for the azimuth angle and elevation angle are shown in Figure 3a,b, respectively. The MAEEs from all methods are inversely proportional to the number of snapshots. Similar to the results given in B (Section 4.1.2), it is obvious that the proposed algorithm has a lower DOA estimation error under the same conditions.

D. Number of Detectable Sources

In this case, we compare the maximum number of sources that can be detected simultaneously by three different methods. Table 1 shows the theoretical upper limit of the number of acoustic sources that can be estimated simultaneously by the three algorithms. Although the maximum number of acoustic sources that can be estimated by the proposed algorithm is smaller than that of the sine-based algorithm, when N = 4, the number of acoustic sources that can be estimated by the proposed algorithm is nine, which is fully sufficient in the general indoor sound source location scenario. Moreover, the proposed algorithm provides nonsingular and ambiguity-free estimation with higher accuracy, which
can be much more useful in practice. In addition, the computational complexity of the EB-ESPRIT algorithm is mainly dominated by the eigenvalue decomposition of the covariance matrix. Therefore, the computational complexity of the proposed algorithm is similar to that of other methods using complex spherical harmonics. Figure 4 shows the results of nine uncorrelated signals from \(120°, 26°\), \(28°, 20°\), \(146°, 52°\), \(-104°, 65°\), \(131°, 73°\), \(61°, 83°\), \(-24°, 107°\), \(48°, 140°\), and \(29°, 177°\). Among them, snapshots are set to 300, without considering aliasing and noise, and we can see that the proposed algorithm successfully locates all the sound sources.

![Figure 3. MAEEs of two incoherent sound sources at different snapshots. (a) Azimuth errors at different snapshots; (b) Elevation errors at different snapshots.](image)

![Figure 4. Hash map of NPTS–ESPRIT (each red circle denotes a theoretical DOA, and each blue × denotes an estimated DOA of NPTS–ESPRIT).](image)

### Table 1. Maximum number of sources that can be detected theoretically.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sine-based [33]</td>
<td>(N^2 + \lfloor N/2 \rfloor)</td>
</tr>
<tr>
<td>TS-ESPRIT [13]</td>
<td>((N - 1)^2)</td>
</tr>
<tr>
<td>Proposed</td>
<td>((N - 1)^2)</td>
</tr>
</tbody>
</table>

E. Comparison of the Computational Efficiency

In a real scene, the sound source may be in constant motion, which requires the algorithm to be able to update the position of the sound source in real time. In this case, we compare the average time required for the three algorithms to run once. As the three algorithms have the same preprocessing process of the signal (the process from the signal received by the microphone to obtain the signal subspace), we count only the time taken by the three algorithms to extract the DOA of the sound source from the signal subspace. Refer to B (Section 4.1.2) for the setting of the simulation environment. The signal-to-noise
The ratio is fixed at 0 dB. The algorithm is run in MATLAB with version R2021b in Windows 10. An Intel Core i7-6700 processor with 3.40 GHz speed and 32 GB RAM is used for running the simulation. Table 2 shows the average time required for the three algorithms to run once. From Table 2, we can see that the running time of the three algorithms is on the order of $10^{-4}$ s, which is far less than the order of magnitude of speech frame length ($10^{-2}$ s). Therefore, the proposed algorithm is competent for the sound source localization problem in real-time scenarios.

Table 2. Average time to run the algorithm once.

<table>
<thead>
<tr>
<th>Method</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sine-based [33]</td>
<td>$3.6366 \times 10^{-4}$</td>
</tr>
<tr>
<td>TS-ESPRIT [13]</td>
<td>$4.2575 \times 10^{-4}$</td>
</tr>
<tr>
<td>Proposed</td>
<td>$4.8659 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

4.1.3. Room Simulation: Setup

To evaluate the estimation performance for source signals in a reflective environment, we simulated one source in a reverberant room. We used a rigid sphere room impulse response generator [10] (which takes into account the scattering effect of a rigid sphere) to generate the room impulse response from the source to the microphones. The virtual spherical microphone array with 32 microphones uniformly distributed over a rigid sphere of 4.2 cm radius was positioned at [2.8898 m, 2.9349 m, 3.3934 m] in a simulated room of size $8 \times 7 \times 6$ m$^3$. The reverberation time of the room was set to $R_{T60} = 0.3$ s and 0.6 s, respectively. A Gaussian white noise signal with a duration of 0.64 s (sampling rate of 16 kHz) was selected as the sound source. Because the sine-based ESPRIT algorithm has an ambiguity problem [24], for convenience, we selected 24 DOAs (total of 24) with elevation angles less than 90° from the 48 spherical t-design directions [34] as the true direction of the sound source, and tested independently on each direction. The sound source was 2 m away from the center of the sphere. The signal received by the virtual microphone was generated by convoluting the source signal with the generated room impulse response and adding 15 dB microphone self-noise. Then we framed and windowed the time domain record and transformed it into the frequency domain using an FFT with a length of 1024 samples. A Hann window with a length of 512 samples was used, and the overlap between frames was 50%. Finally, the spherical Fourier transform ($N = 4$) and modal compensation were applied to each time-frequency bin to obtain the spherical harmonic coefficients used to construct the covariance matrix.

In this simulation, we set the forgetting factor to 0.9 to update the covariance matrix online [35]. DOAs were estimated in the frequency range [1296.875, 5203.125] Hz, and estimation errors ($\Delta \Psi(t, f) = \cos^{-1}\{\cos \theta \cos \hat{\theta} + \cos (\phi - \hat{\phi}) \sin \theta \sin \hat{\theta}\}$) were calculated for each time-frequency bin. The mean angular error across the TF bins is defined as

$$\bar{\Delta \Psi} = \frac{1}{2I_{n_{TF}}} \sum_{i=1}^{I} \sum_{n,f \in s_{TF}} \Delta \Psi_i(t, f), \quad (44)$$

where $s_{TF}$ is a set of TF bins whose energy is higher than $-30$ dB from the maximum energy, $n_{TF}$ represents the number of valid TF bins, $(\hat{\theta}, \hat{\phi})$ represents the estimated direction of the algorithm, subscript $i$ represents the $ith$ experiment, and $I$ represents the total number of experiments.

4.1.4. Room Simulation: Results

The mean angular error is shown in Figure 5. It can be seen from the figure that compared with the aforementioned free-field simulation, the estimation error under the room simulation increases. The main reason is that the early reflection under the broadband signal has a greater impact on the DOA estimation of the real sound source. In this case, we carry out DOA estimation on only a single TF bin, and do not apply anti-reverberating
techniques such as frequency smoothing [16] (we combine this technique in the following simulation and field test). The estimation error of the proposed algorithm under the two reverberation environments is far less than that of the two traditional algorithms, with a better estimation accuracy. The performance of the sine-based ESPRIT algorithm is the worst in this simulation, and the main reason is that most of the candidate sound sources are closer to the equator (the elevation angle of 12 directions is greater than 60°, and the elevation angle of 3 directions is 86.4480°). This is consistent with the simulation results given in A (Section 4.1.2).

![Figure 5. Mean angular error under two different reverberation times.](image)

4.2. Comparison before and after the Introduction of Relative Sound Pressure

From the previous section, we can see that all three algorithms are sensitive to noise, which is caused by subspace swap [36]. In this section, we try to introduce relative sound pressure technology [17] into the proposed algorithm (RNPTS-ESPRIT), and compare the RNPTS-ESPRIT algorithm with the NPTS-ESPRIT algorithm to evaluate the performance improvement of the proposed algorithm before and after the introduction of relative sound pressure. Meanwhile, the simulation in the following part simulates an indoor environment full of reverberation, so we introduce frequency-smoothing technology [16] into the two algorithms mentioned above to enhance the anti-reverberating performance of the algorithm.

4.2.1. Room Simulation: Setup

The simulation settings are referred to the previous section, and we give only the differences here. Firstly, the virtual array was placed in the center of a room with dimensions of $6 \times 6 \times 6$ m$^3$. Speech signals with a duration of 4 s (sampling rate of 16 kHz) were selected from TIMIT corpus as the signal source. For STFT, a Hamming window of 256 samples was used, which was shifted with 50% overlap for each time bin, and 256-points FFT was applied. The localization algorithm used 21 frequency bins ranging from 2000 Hz to 3250 Hz, which accurately measured the third-order sound field. When calculating the relative sound pressure, we assumed that the speech signal remains stationary within 0.1 s, and used the Welch method to estimate the PSD and CPSD of the signal, with a window length of 0.016 s and an overlap of 50%.

4.2.2. Room Simulation: Results

A. Estimation of the Number of Active Sources

In the example, it is verified that the introduction of the relative sound pressure can increase the difference in the eigenvalues corresponding to the signal and noise subspaces, where a larger difference is more beneficial for the estimation of the number of active sources [37]. In this example, we set the reverberation time of the room to $RT_{60} = 0.5$ s,
and three sources of sound from \((150^\circ, 145^\circ), (110^\circ, 75^\circ),\) and \((10^\circ, 25^\circ)\) are initialized simultaneously during the test. The eigenvalue decomposition of the covariance matrix is performed separately for the SNRs in the range of 0–25 dB, and the corresponding normalized eigenvalues are shown in Figure 6a–e. From the figures, we can see that the introduction of the relative sound pressure does not increase the difference between the normalized eigenvalues corresponding to the signal and noise subspaces (the difference between the third and fourth eigenvalues) for signal-to-noise ratios less than 5 dB (including 5 dB). However, when the signal-to-noise ratio is greater than 5 dB, we can clearly see that the introduction of the relative sound pressure can significantly increase the difference between the third and fourth normalized eigenvalues. The underlying reason for the above phenomenon is that the estimation of the relative sound pressure (Equation (6)) may not hold at low SNRs, which is also verified by the next estimation error of the algorithm, where different SNRs are applied to the DOA of the sound source.

![Figure 6](image-url)

Figure 6. Normalized eigenvalues obtained via a singular value decomposition of the source signal covariance matrix (room reverberations \(T_{60} = 0.3\) s): (a) SNR = 0 dB; (b) SNR = 5 dB; (c) SNR = 10 dB; (d) SNR = 15 dB; (e) SNR = 20 dB.

B. Errors for Various SNRs

In this example, we compare the performance of RNPTS-ESPRIT and NPTS-ESPRIT for SNRs in the range of 0–20 dB, where the reverberation time \(R_{T_{60}}\) is 0.5 s. Two incoherent signals are incident to the spherical array from \((90^\circ, 165^\circ)\) and \((-70^\circ, 135^\circ)\). Two hundred independent Monte Carlo experiments correspond to the overall MAEE regarding the elevation and azimuth angles, as shown in Figure 7. We can see that the localization accuracy of RNPTS-ESPRIT is worse than that of NPTS-ESPRIT when the signal-to-noise ratio is less than 10 dB, and the difference between the localization accuracy of RNPTS-ESPRIT and NPTS-ESPRIT is larger when the SNR is lower than 10 dB; however, the localization accuracy of RNPTS-ESPRIT is better than that of NPTS-ESPRIT when the SNR is higher than 10 dB, and the localization accuracy of the former is higher when the SNR is higher, while that of the latter is almost invariant.
C. Errors for Various SNRs in Multiple Sources

In the case of SNRs of 5–20 dB, the overall MAEE of elevation angles and azimuth angles corresponding to 200 independent Monte Carlo experiments is shown in Figure 8. It can be seen that, compared with Figure 7 in the case of multiple sound sources, the performance of both algorithms deteriorates significantly. However, when SNR is greater than or equal to 10 dB, the RNPTS-ESPRIT algorithm improves the MAEE by more than 1.5° compared with the NPTS-ESPRIT algorithm, and this value increases with the increase in SNR, to reach 2.16° when SNR is 20 dB. Through verification in different scenarios, we can conclude that the introduction of relative sound pressure can enhance the accuracy of the DOA estimation of the algorithm and increase the difference between the corresponding eigenvalues of signal subspace and noise subspace in the environment of high SNR (greater than or equal to 10 dB). However, when SNR is less than 10 dB, relative sound pressure is no longer effective. Therefore, the proposed algorithm for low SNR and strong reverberation environment needs to be improved.

4.3. Verification Using Real Recordings

In this section, we used real recordings to verify the effectiveness of the NPTS-ESPRIT algorithm proposed in this paper. The experiment was completed in a conference room of the Guilin University of Electronic Technology. In this experiment, we used a rigid spherical array with a sampling rate of 48,000 Hz and 32 microphones uniformly distributed on the surface to record 4 s long recordings each time. Then we carried out STFT analysis on the
collected time-domain records. The parameters of the STFT are as follows: Hanning window length of 1024 samples, 50% overlap between adjacent frames, and FFT of 1024 points. The localization algorithm uses 21 frequency points ranging from 2250 Hz to 3187.5 Hz. The spherical microphone array and experimental site we used are shown in Figure 9.

![Figure 9. Field testing. (a) Spherical microphone array; (b) Conference room.](image)

Firstly, we conducted the test in the single-source scenario. In this scenario, there is a speaker located at $(175°, 75°)$ relative to the center of the spherical array. Table 3 shows the estimation results of the NPTS-ESPRIT algorithm and TS-ESPRIT algorithm in this scenario and the angular error between the estimated value and the true value, respectively.

<table>
<thead>
<tr>
<th>Ground Truth DOA</th>
<th>NPTS-ESPRIT/Angular Error</th>
<th>TS-ESPRIT/Angular Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(175°, 75°)$</td>
<td>$(174.9062°, 74.9743°)/0.0942°$</td>
<td>$(174.9062°, 77.5281°)/2.5297°$</td>
</tr>
</tbody>
</table>

Then, we conducted the test in the dual-source scenario. In this scenario, there are two simultaneous speakers located at $(235°, 85°)$ and $(125°, 75°)$ relative to the array center. Table 3 shows the estimation results of the NPTS-ESPRIT algorithm and TS-ESPRIT algorithm in this scenario and the angle error between the estimated value and the true value, respectively. From Tables 3 and 4, we can see that both algorithms successfully detect the location of the sound source in different scenarios and that the estimation error of the NPTS-ESPRIT algorithm is smaller than that of the TS-ESPRIT algorithm.

<table>
<thead>
<tr>
<th>Ground Truth DOA</th>
<th>NPTS-ESPRIT/Angular Error</th>
<th>TS-ESPRIT/Angular Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(175°, 75°)$</td>
<td>$(174.9062°, 74.9743°)/0.0942°$</td>
<td>$(174.9062°, 77.5281°)/2.5297°$</td>
</tr>
<tr>
<td>$(125°, 75°)$</td>
<td>$(126.8516°, 74.8148°)/1.7973°$</td>
<td>$(126.5452°, 77.1942°)/2.6578°$</td>
</tr>
</tbody>
</table>

5. Discussion

In this paper, we write the two eigenvalue decompositions in the TS-ESPRIT algorithm to determine the DOA parameter as a generalized eigenvalue problem, which avoids the two eigenvalue decomposition operations. By using the QZ algorithm to solve the generalized eigenvalue problem, the problem of DOA parameter pairing in TS-ESPRIT is solved without increasing the computational complexity of the algorithm. At the same time, due to the use of the tangent function as the directional parameter, the proposed method has higher and more stable estimation accuracy for the elevation angle compared with traditional algorithms. We also further improve the positioning accuracy of the algorithm in noisy and multi-source scenes by introducing the concept of relative sound pressure. Finally,
by introducing frequency-smoothing technology to remove the correlation between signals, our algorithm can be applied to indoor sound source positioning scenes with reverberation.

Through rich numerical simulations and field measurements, we demonstrate that the proposed method solves the problem of DOA parameter pairing in TS-ESPRIT techniques and has outstanding DOA estimation accuracy compared with traditional algorithms. In the future, we will further optimize the proposed algorithm to make it better able to complete the positioning and tracking of multiple sound sources in low-SNR and high-reverberation environments.

Author Contributions: Conceptualization, Z.L.; Data curation, H.Z.; Formal analysis, H.Z.; Funding acquisition, Z.L.; Investigation, H.Z., Z.L. and L.L.; Methodology, Z.L. and L.L.; Project administration, M.W.; Resources, M.W. and X.S.; Software, H.Z.; Supervision, Z.L.; Validation, H.Z.; Visualization, H.Z.; Writing—original draft, H.Z.; Writing—review and editing, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant numbers 62071135 and 62201163, and Project (CRKL200111) from the Key Laboratory of Cognitive Radio and Information Processing, Ministry of Education (Guilin University of Electronic Technology).

Data Availability Statement: Publicly available datasets were analyzed in this study. The data can be found on https://en.wikipedia.org/wiki/TIMIT (accessed on 1 January 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

22. Famoriji, O.J.; Shongwe, T. Critical Review of Basic Methods on DoA Estimation of EM Waves Impinging a Spherical Antenna Array. Electronics 2022, 11, 208. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.