

DOI:10.19651/j.cnki.emt.2105890

基于UCA压缩感知的声源定位算法*

杨瑞峰 温斐旻 郭晨霞

(中北大学电子测试技术国家重点实验室 太原 030051)

摘要: 针对混合声源定位精度低的问题,提出了一种非迭代补全互谱矩阵的算法,利用互谱矩阵的埃尔米特特性,不经过迭代,在不牺牲定位精度的情况下提高计算效率。该方法首先建立一个稀疏模型,通过构造冗余脉冲响应(RIR)矩阵作为压缩感知测量矩阵,将源定位问题转化为压缩感知问题。然后根据多个源方向向量的空间稀疏相关性,引入投影算子,在压缩感知框架下使方位角的均方根误差保持在5%以内。均匀圆阵(UCA)环境下的仿真结果表明,与传统算法相比,该算法具有更好的估计性能。

关键词: 互谱矩阵;压缩感知;稀疏矩阵;均匀圆阵;声源定位

中图分类号: TN912.34 **文献标识码:** A **国家标准学科分类代码:** 510.3099

Sound source localization algorithm based on UCA compressed sensing

Yang Ruifeng Wen Feimin Guo Chenxia

(National Key Laboratory for Electronic Measurement Technology, North University of China, Taiyuan 030051, China)

Abstract: Aiming at the problem of low positioning accuracy of mixed sound source, a non iterative algorithm to complete cross spectral matrix is proposed, which uses the Hermitian characteristics of the cross-spectrum matrix without iteration and improves the computational efficiency without sacrificing positioning accuracy. This method first establishes a sparse model, and converts the source localization problem into a compressed sensing problem by constructing a redundant impulse response (RIR) matrix as a compressed sensing measurement matrix. Then, according to the spatial sparse correlation of multiple source direction vectors, a projection operator is introduced, and the root mean square error of azimuth is kept within 5% under the framework of compressed sensing. The simulation results in the uniform circular array (UCA) environment show that the algorithm has better estimation performance than the traditional algorithm.

Keywords: cross-spectrum matrix; compressed sensing; sparse matrix; uniform circular array; sound source localization

0 引言

麦克风阵列由特定空间模式排列的多元麦克风传感器组成,在一系列新兴应用中引起了广泛关注,如移动设备、视频会议和仿人机器人^[1-3]。在急剧增加的应用中,麦克风阵列接收的信号被普遍用于语音增强^[4]和减轻噪声^[5]等,进一步处理就需要声源定位来提供源方向。

在大多数应用中,所有可能的方向中只存在少量声源,因此声源方向矢量的估计可以转化为稀疏重构问题。近年来,压缩感知(CS)麦克风阵列声源定位算法就是从这个问题的^[6]。经典转向响应波束形成算法(SRP)^[7]通过最大化功率输出实现到达方向估计。相位变换 SRP 算法(SRP-PHAT)^[8]为了产生更窄的响应功率峰值,只考虑

了归一化的相位问题。经典的 CS 波束方向估计算法会受到大量背景噪声的影响^[9]。

本文的工作安排为:首先建立了一种稀疏模型来减轻混响的影响,利用互谱矩阵的埃尔米特特性,提出了一种补全厄米特矩阵声源定位法,实现源方向的高效估计。最后,数值仿真和实验结果验证了该方法的有效性。

1 信号模型

首先假设单个声源和由 N 个麦克风组成的麦克风阵列,以及声源和麦克风的全向特性。设 $s(n)$ 表示点声源信号,麦克风接收到的信号可以表示为:

$$x(n) = h(r_s, n) * s(n) + v(n) \quad (1)$$

其中, $h(r_s, n)$ 是从声源到麦克风的脉冲响应函数;

收稿日期:2021-03-04

* 基金项目:山西省重点研发计划(201903D121060)、山西省回国留学人员科研项目(2020-111)资助

$v(n)$ 是附加噪声,假设不相关; r_s, n 分别为声源和时间指数。这里 $h(r_s, n)$ 可以进一步分解为延时响应和混合响应,则:

$$x(n) = h_d(r_s, n) * h_r(r_s, n) * s(n) + v(n) \quad (2)$$

其中,延时响应由阵列集合形状决定,不同的麦克风单元对应不同的延时,混合响应由墙壁的反射决定。那么相应频域可以等价表示为:

$$X(k) = H_d(r_s, k) H_r(r_s, k) \mathbf{S}(k) + V(k) \quad (3)$$

其中, k 是频率指数; $X(k), H_d(r_s, k), H_r(r_s, k), \mathbf{S}(k)$ 和 $V(k)$ 分别表示 $x(n), h_d(r_s, n), h_r(r_s, n), s(n)$ 和 $v(n)$ 的光谱。假设单个声源可能的方向集合为 $\{r_1, \dots, r_I\}$, I 为方向集合的个数,其远大于声源的个数。

$$X(k) = \mathbf{D}(k) \mathbf{S}(k) + V(k) \quad (4)$$

$$\mathbf{D}(k) = [H(r_1, k), \dots, H(r_I, k)] \quad (5)$$

其中, $\mathbf{S}(k) = [s_1(k), \dots, s_I(k)]$ 是声源方向向量。虽然脉冲响应函数矩阵包含混响,但是它的性能在高混响下回减弱,所以通过分解脉冲响应函数的延时项和混响项,剩余冗余频域脉冲响应矩阵只包含延时项,有 $\mathbf{D}_d(k) = [H_d(r_1, k), \dots, H_d(r_I, k)]$, 所以去除混响的影响后式(4)可以表达为:

$$X(k) = \mathbf{D}_d(k) \overline{\mathbf{S}}_r(k) + V(k) \quad (6)$$

$\overline{\mathbf{S}}_r(k) = [H_r(r_1, k) s_1(k), \dots, H_r(r_I, k) s_I(k)]^T$ 是混响源方向向量。 $[\cdot]^T$ 为转置运算符。

2 声定位算法和具体步骤

2.1 UCA 压缩感知算法

首先假设具有 M 个麦克风组成的阵列模型从位置 1 移动到位置 L , 如图 1 所示。在位置 i 处, 麦克风平面与声音平面之间的距离为 h , 麦克风的坐标为 $\mathbf{P}^i = [p_1^i, \dots, p_M^i]^T$, 麦克风所测声压为 $\mathbf{Y}^i = [y_1^i, \dots, y_M^i]^T$ 。声源平面由 $N = m \times n$ 个等距网格组成, 声源平面位置为 $\mathbf{Q} = [q_1, \dots, q_N]$, N 个网格节点处的声压为 $\mathbf{b} = [b_1, \dots, b_N]^T$ 。两平面声压的关系表达为:

$$\mathbf{Y}^i = [\mathbf{W}_1^i, \dots, \mathbf{W}_N^i] \mathbf{b} = \mathbf{W}^i \mathbf{b} \quad (7)$$

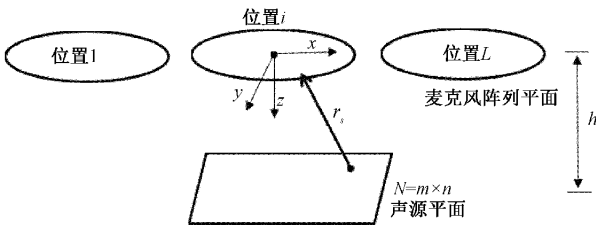


图1 声信号传播模型

其中 $\mathbf{W}_N^i = [w_1^i, \dots, w_M^i]^T$ 为权重函数, $\mathbf{W}^i \in \mathbf{C}^{M \times N}$ 为传递函数。根据自由场中的格林函数:

$$w_M^i = \frac{e^{-jkr_M^i}}{4\pi r_M^i} \quad (8)$$

在位置 i 处厄米特矩阵定义为:

$$\mathbf{S}^i = E\{\mathbf{Y}^i(t) \mathbf{Y}^{iH}(t)\} \quad (9)$$

$\mathbf{S}^i \in \mathbf{C}^{ML \times ML}$, $E\{\cdot\}$ 为数学期望, $\mathbf{Y}^i(t) \in \mathbf{C}^{ML \times 1}$ 是快拍数为 t 时麦克风所测声压。网络节点处的声音功率为:

$$\mathbf{B} = \frac{\mathbf{W}^H \mathbf{S}^i \mathbf{W}}{\|\mathbf{W}\|_2^4} \quad (10)$$

然后对厄米特矩阵 \mathbf{S}^i 进行特征值分解^[10]:

$$\mathbf{S}^i = [\mathbf{u}_1, \dots, \mathbf{u}_M] \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_M \end{pmatrix} [\mathbf{u}_1, \dots, \mathbf{u}_M]^H \quad (11)$$

其中特征值 $\lambda_1 > \dots > \lambda_M$, \mathbf{u}_M 是其对应的特征向量。令 \mathbf{S}^i 正负惯性指数分别为 p, q , $\mathbf{U}_i = [\mathbf{u}_1, \dots, \mathbf{u}_M]$, 当满足 $\mathbf{u}_i = \sqrt{|\lambda_i|} \times \mathbf{u}_i$ 时, \mathbf{S}^i 可以转化成标准矩阵 \mathbf{D} :

$$\mathbf{D} = \mathbf{U}^{-1} \mathbf{S}^i \mathbf{U}^{H-1} = \begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix} \quad (12)$$

定义一个矩阵:

$$\Phi = \mathbf{U} \mathbf{S}^i \mathbf{U}^H = \begin{pmatrix} \mathbf{D}_1 & \dots & \mathbf{X}_{1L} \\ \vdots & \ddots & \vdots \\ \mathbf{X}_{1L}^H & \dots & \mathbf{D}_L \end{pmatrix} \quad (13)$$

$$\mathbf{X}_{1L} = \begin{pmatrix} I_{p_{1L}} & & \\ & -I_{q_{1L}} & \\ & & 0 \end{pmatrix} \quad (14)$$

根据希尔韦斯特惯性定律^[11], 矩阵 \mathbf{S}^i 和 Φ 的秩相等。因此互谱矩阵 \mathbf{S}^i 可以通过如下迭代方程获得。

$$\mathbf{S}^i = \mathbf{U}^{-1} \Phi \mathbf{U}^{H-1} \quad (15)$$

为保证声场的空间连续性, 麦克风的位置数据也应该包含在互谱矩阵中^[12]。在这里引入投影算子^[13] $\Psi = \theta \theta^{\dagger}$, $(\cdot)^{\dagger}$ 是伪逆运算符, 满足如下方程:

$$\hat{\mathbf{S}} = \Psi \mathbf{S}^i \Psi^H \quad (16)$$

θ 可以通过麦克风阵列与声源平面之间的传播函数 $\mathbf{G} \in \mathbf{C}^{ML \times N}$ 进行奇异值分解获得。

$$\mathbf{G} = \mathbf{U} \Sigma \mathbf{V}^H \quad (17)$$

其中, Σ 为奇异值, $\mathbf{U} \in \mathbf{C}^{ML \times ML}$ 和 $\mathbf{V} \in \mathbf{C}^{N \times N}$ 为左奇异向量和右奇异向量, 分别代表声场和声源。然后就可以计算得到 θ , 运算符 $Ext(\Delta)$ 表示提取 Δ 的第 1 列。

$$\theta = Ext(\mathbf{U}, M\sqrt{L}) \quad (18)$$

由于在不同位置测得得声压 \mathbf{Y}^i 和 \mathbf{Y}^j 不是同时测量的, 其得到的厄米特矩阵 \mathbf{S}^i 是不完整的, 所以需要用到对角块将其填充为 \mathbf{S}^{ii} 。当矩阵 \mathbf{S}^i 的秩最小时, 通过下面方程得到完整的 \mathbf{S}^{ii}

$$\min rank(\mathbf{S}^i) \Rightarrow \mathbf{S}^i = \mathbf{S}^{ii} \quad (19)$$

2.2 具体步骤

1) 获取 L 个麦克风阵列测量的声压数据, 并分为 $(M \times L \times T)$ 份, T 为快拍数;

2) 将声源平面离散为 N 个等距网格, 设定声源平面

与麦克风阵列之间的距离 h , 构造权重函数 $W_u, u = 1, \dots, N$;

3) 将步骤 1) 的数据进行傅里叶变换, 然后根据所需的频率选取数据;

4) 分别计算矩阵 S^i, U_i 和 X_{iL} ;

5) 通过式 (11) 和 (12) 构造矩阵 Φ , 计算矩阵 S^i ;

6) 引入投影算子 Ψ , 然后计算矩阵 \hat{S} ;

7) 令 $S = \hat{S}$, 通过式 (10) 计算声源功率 $B, u = 1, \dots, N$, 得到声源位置。

3 数值模拟

首先模拟一个 $30 \text{ m} \times 20 \text{ m} \times 5 \text{ m}$ 的混响室进行数值分析。由 6 个麦克风组成的均匀圆阵 (UCA) 位于房间中, UCA 的半径是 7 cm 。声源平面是 $2 \text{ m} \times 2 \text{ m}$ 的矩形平面, 若干相关声源在声源平面任意位置产生声音信号。 z 轴方向距离 UCA 平面 1 m 。UCA 的中心位置依次移动到不同的位置。仿真实验研究的重点是方位估计, 与常规声定位 (SRP-PHAT) 算法、延时求和 (DS) 算法^[14]、经典压缩感知 (OMP) 算法^[15] 和本文所提算法进行性能评估和比较。采用波达方向的均方根误差 (RMSE) 和信噪比 (SNR) 作为性能评估的度量, 定义为:

$$RMSE = \sqrt{\frac{1}{KM} \sum_{k=1}^K (\theta - \hat{\theta})^2} \quad (20)$$

$$SNR = 10 \lg \left(\frac{1}{\sigma^2} \right) (29 \times 10^1) \quad (21)$$

其中, K 是声源个数, M 为实验次数, $\hat{\theta}$ 是声源波达方向的估计值, θ 为真实值。

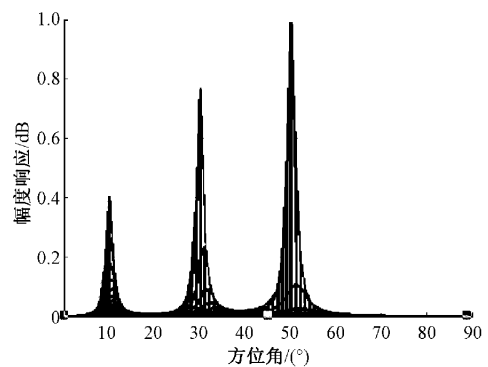
3.1 验证算法有效

在这一部分中, 主要采用 3 个相干声源, 位置分别在 $(0.32 \text{ m}, 1 \text{ m})$ 、 $(1 \text{ m}, 1.32 \text{ m})$ 和 $(1.32 \text{ m}, 0.32 \text{ m})$ 。仿真实验将在两种情况下进行, 一种是混响系数为 0.4, 信噪比 5 和 10 dB 的条件下进行, 另一种是信噪比为 8, 混响系数分别为 0.2 和 0.6 dB 的情况下进行。网格上声源的真实方位角 θ 分别为 10° 、 30° 和 50° 。快拍数 T 设置为 1 000, 采样频率为 32 ksp。

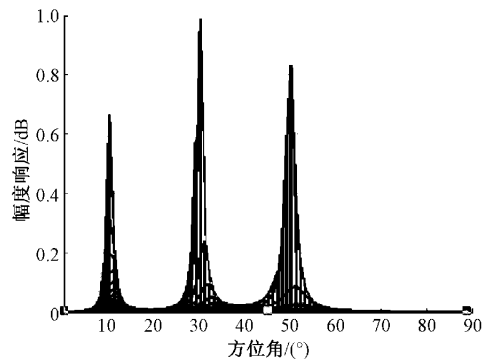
如图 2 和 3 所示, 分别显示了 3 个相关光源在网格上不同情况的定位结果。可以看出所提出的算法成功定位了这 3 个声音信号, 并且在图中可以直观地了解信号的幅度响应。由此证明所提出的方法是有效可行的。

3.2 定位性能分析

为了评估在不同噪声和混响背景下的性能, 数值模拟是通过向 UCA 的每个麦克风通道添加独立的高斯白噪声 (WGN) 实现不同的信噪比, 此次仿真实验的信噪比分别为 14.12、6.9 和 3.77。将模型中不同的混响系数都设置为 0.2、0.4 和 0.6。快拍数 T 设置为 1 000, 采样频率为 32 ksp。

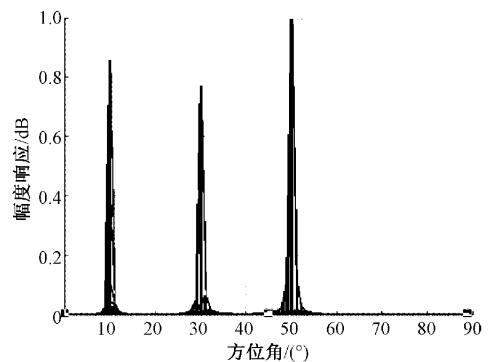


(a) SNR=5 dB

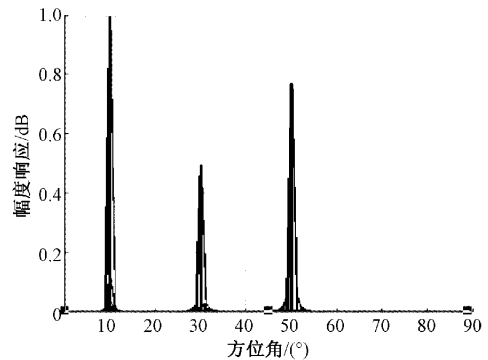


(b) SNR=10 dB

图 2 混响系数为 0.4 的定位结果



(a) 混响系数为 0.2 dB



(b) 混响系数为 0.6 dB

图 3 SNR=8 的定位结果

如图 4 和 5 所示, 分别呈现了在不同信噪比和混响水平下用不同算法获得的 RMSE 结果。如两个图中所示, 很

明显,虽然 DS 算法和 OMP 算法通常实现了比 SRP-PHAT 算法观察到的误差小,但与本文所提方法相比,所提出的方法表现出更好的性能,并且方位估计值的均方根误差均保持在 5% 以内。这是因为本文利用源方向稀疏相关性减轻了噪声和混响的影响。

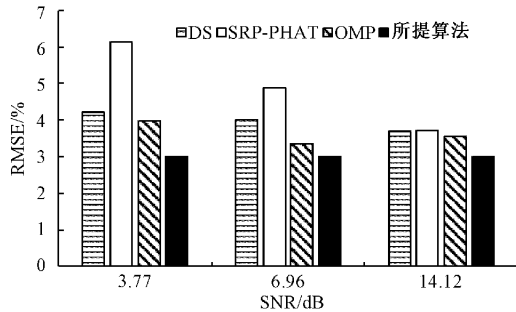


图4 不同信噪比下的 RMSE

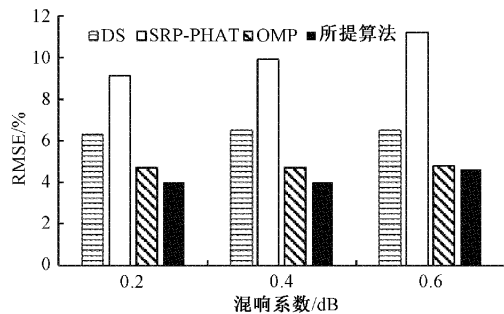


图5 不同混响系数下的 RMSE

4 结 论

为了提高混合声源定位精度,首先将脉冲响应分解为两个独立的项后建立一个稀疏模型,以便于减轻混响的影响。此外,利用投影算子方法找到完整的厄米特矩阵,提出了基于UCA压缩感知的声源定位方法。数值仿真以及与传统算法的对比实验结果验证了该方法的有效性和性能提升,具有应用于实际声源定位的潜力。下一步的工作将集中在解决不规则阵列声源定位问题。

参考文献

- [1] 郑广宁,陈廷森,崔旭. 电视会议终端故障快速定位及处置方法研究[J]. 电力信息与通信技术, 2015, 13(10):49-51.
- [2] 刘振,陈万米,熊奇. 移动机器人测程法定位误差分析与校正[J]. 电子测量技术, 2017, 40(12):75-81.
- [3] WAN X, WU Z. Sound source localization based on discrimination of cross-correlation functions [J]. Applied Acoustics, 2013, 74(1):28-37.

- [4] 周正仙,邹翔,袁扬胜,等. 干涉型光纤语音传感器及语音降噪方法研究[J]. 仪器仪表学报, 2017, 38(11): 2715-2724.
- [5] 王义,王江云,宋晓,等. 基于模糊转换的图像椒盐噪声检测和去除[J]. 电子测量与仪器学报, 2017, 31(4): 537-542.
- [6] GURBUZ A C, CEVHER V, MCCLELLAN J H. Bearing estimation via spatial sparsity using compressive sensing [J]. IEEE Transactions on Aerospace and Electronic Systems, 2012, 48(2): 1358-1369.
- [7] 陶震. 基于导向矢量失配估计的鲁棒波束域自适应波束形成算法[J]. 国外电子测量技术, 2017, 36(7): 23-26.
- [8] 钱宇宁,陈亚伟,孙俊. 基于提升Keystone变换的声呐宽带自适应波束形成方法[J]. 电子与信息学报, 2019, 41(2):324-331.
- [9] 杨云仙,梁卫文. 基于混合测量模式的压缩感知研究[J]. 湖北工业职业技术学院学报, 2020, 33(3): 73-76.
- [10] 李伟,李开宇. 基于特征提取矩阵的稀疏系数求解算法[J]. 电子测量技术, 2017, 40(9):146-150.
- [11] 孔祥阳,徐保根,周杰. 基于多线性权重核范数最小化的遥感图像去噪[J]. 机床与液压, 2020, 48(12):184-190,208.
- [12] 段云场. 基于谱矩阵分解重构的波束形成声源识别方法[D]. 重庆:重庆大学, 2016.
- [13] WANG Y, YE G. The key technology of QR code image recognition[J]. Journal of Image and Signal Processing, 2015, 4(3): 53-66.
- [14] 初萍,司伟建. 基于延时相关预处理的MUSIC算法[J]. 沈阳工业大学学报, 2013, 35(1):93-98.
- [15] MERT B C, HUSEYIN H. Acoustic source separation using rigid spherical microphone arrays via spatially weighted orthogonal matching pursuit [C]. International Workshop on Acoustic Signal Enhancement, Tokyo, Japan, 2018, 8(2):158-171.

作者简介

杨瑞峰,教授,博士生导师,主要研究方向为自动化测试与控制技术。

E-mail: yangruifeng@nuc.edu.cn

温斐旻(通信作者),硕士研究生,主要研究方向为仪器科学与技术。

E-mail: 1394959453@qq.com