

◇ 李启虎院士八十华诞学术论文 ◇

分布式压缩感知麦克风阵列多声源方位估计*

黄惠祥 郭秋涵 童峰[†] 吴燕艺

(厦门大学 水声通信与海洋信息技术教育部重点实验室 厦门 361100)

摘要 麦克风阵列已被广泛应用于音/视频会议等人机交互领域中,多声源应用场景对声源方位估计性能提出了更高的要求。压缩感知声源定位算法将声源定位问题转化为信号的稀疏重构问题,相比传统的定位算法如联合可控响应功率和相位变换和时延累加定位能够获得较高的定位性能,但多声源的存在一定程度上降低了稀疏程度,影响了压缩感知重构性能。考虑到传统的压缩感知定位算法并未利用多个连续语音帧之间声源空间向量的共同稀疏性,提出采用分布式压缩感知理论以改善多声源的稀疏恢复估计的性能。仿真和实验结果表明,相比于传统定位算法和压缩感知-正交匹配追踪算法,分布式压缩感知-同步正交匹配追踪算法在不同信噪比和不同声源强度的环境中,对多声源的方位估计都具有更好的定位性能和定位稳健性。

关键词 麦克风阵列,多声源定位,分布式压缩感知

中图分类号: TN912.3 文献标识码: A 文章编号: 1000-310X(2019)04-0605-10

DOI: 10.11684/j.issn.1000-310X.2019.04.018

Distributed compressed sensing microphone array multi-source azimuth estimation

HUANG Huixiang GUO Qiuhan TONG Feng WU Yanyi

(Key Laboratory of Underwater Acoustic Communication and Marine Information Technology Ministry of Education, Xiamen University, Xiamen 361100, China)

Abstract Microphone arrays have been widely used in the field of human-computer interaction such as audio/video conferencing. It is necessary to make higher-resolution azimuth estimation performance for scenes with multiple sound sources in different orientations. The compressed sensing (CS) sound source localization algorithm transforms the sound source localization problem into a sparse reconstruction problem of the signal, thus achieves better estimation performance compared to traditional localization algorithms such as steered response power with the phase transform (SRP-PHAT) and time delay-sum (DS). However, the existence of multiple sound sources reduces the sparsity, to some extent degrades the performance of CS reconstruction. Considering that the traditional CS localization algorithm does not utilize the common sparsity of the sound source space vector between multiple consecutive speech frames, in this paper the distributed compressed sensing (DCS) theory is proposed to improve the performance of sparse recovery estimation of multiple sound sources. The simulation and experimental results show that compared with traditional positioning algorithm and compressed sensing-orthogonal matching pursuit (CS-OMP) algorithm, distributed compressed sensing-simultaneous orthogonal matching pursuit (DCS-SOMP) algorithm has better positioning performance and robustness for multi-sound source azimuth estimation under different SNR and different sound source intensity environments.

Key words Microphone array, Multi-source positioning, Distributed compressed sensing

2019-01-21 收稿; 2019-04-05 定稿

*国家自然科学基金项目 (11574258), 福建省高校产学研合作项目 (2015H6019), 中央高校基本科研业务费专项资金项目 (20720190102)
作者简介: 黄惠祥 (1993-), 男, 福建泉州人, 硕士研究生, 研究方向: 麦克风阵列信号处理。

[†] 通讯作者 E-mail: ftong@xmu.edu.cn

0 引言

麦克风阵列目前被广泛应用于音/视频会议系统、车载系统以及智能音箱等领域。麦克风阵列对声源方向上的语音进行增强,这就要求信号相对于麦克风阵列的位置是已知的,而基于麦克风阵列的声源定位技术为此提供了支持。传统的麦克风阵列声源定位技术主要分为三大类:基于最大输出功率的可控波束形成的定位算法^[1]、时延估计定位算法^[2]和基于高分辨率谱估计的定位算法^[3]。可控波束形成定位算法需要做全局搜索,计算量大,同时对语音信号和噪声信号的频谱特性依赖性较强。时延估计定位算法运算量较小,易于实现算法的硬件实现^[4]。但是该算法在受到噪声、混响、信号采样率等方面的影响时,麦克风阵列各个阵元获取精确的时延值难度增大,导致算法的定位性能降低。基于联合可控响应功率和相位变换(Steered response power-phase transform, SRP-PHAT)^[5]声源定位算法结合了可控响应功率和相位变换加权的优点,相对于时延估计的定位算法有更好的性能,然而在混响和噪声强的情况下,该算法可能会在错误的空间位置上产生波束峰值,定位效果下降。基于高分辨率谱估计的算法主要运用于窄带信号的处理,而对于语音这类宽带信号的处理会增大算法的运算复杂度,不适合应用于实时实现。

压缩感知(Compressive sensing, CS)理论^[6]可以通过低于奈奎斯特采样率的方式对数据采集并精确重构。声源的空间稀疏性使得该理论在声源定位研究中得到了应用。赵小燕等^[7]针对高混响低信噪比的环境提出了基于压缩感知的定位算法,将声源定位问题转化为稀疏信号的重构问题,利用正交匹配追踪(Orthogonal matching pursuit, OMP)算法对信号进行重构。针对信号在传输过程中存在多径的问题,Patel等^[8]利用信号的空间稀疏性,提出一种基于均匀线性阵列的方位估计新技术,实现多径环境下对宽带语音信号的方位估计,相比Roor-MUSIC算法和ESPRIT算法都具有更好的鲁棒性。分布式压缩感知(Distributed compressed sensing, DCS)^[9]针对多个观测信号的相关性,提出对多个信号进行联合重构,提升稀疏信号的重构性能。Baron等^[9]提出了三种联合稀疏模型(Joint sparsity model, JSM),其中JSM2模型提出各个信

号间具有相同的稀疏支撑集但是非零系数不同,该模型主要被应用于水声MIMO通信和阵列信号处理与定位两个方面。

在实际应用的场景比如多人会议中,麦克风阵列设备往往需要不断重新定位或者定位多人说话方向,这类场景对于麦克风阵列如何定位多说话人方位提出了新的挑战。金光明等^[10]提出用L型麦克风阵列,对多声源的频域与到达角进行联合估计实现多声源的定位挑战,该算法能够得到5%以下误差的定位精度,但是对于频率相同的情况仍需要进一步优化,同时该算法也难以适用于宽带语音信号。章宇栋等^[11]在传统的CS算法的基础上,以阵列时延关系构造房间冲激响应CS矩阵,实现多声源的波达方向(Direction of arrival, DOA)稀疏恢复,该算法相对于传统CS算法具有更好的多声源定位性能。Carabias等^[12]针对多声源定位的问题,将非负矩阵分解(Non-negative matrix factorization, NMF)算法优化为复数非负矩阵分解(Complex non-negative matrix factorization, CNMF)算法,用于求解空间协方差模型(Spatial covariance model, SCM),最高可实现达6个声源的同时定位,对于声源数目的变化具有很好的鲁棒性。然而该方法的前提为多声源的信号需要高度相关的,这使得算法在实际应用中出现了一定的局限性。

对于压缩感知声源定位算法,多声源场景一定程度上降低了声源方向向量的空间稀疏程度,因而将导致多声源条件下压缩感知声源定位算法的性能下降。针对这一问题,考虑到对声源方向向量在多个连续语音帧中具有稀疏相关性,本文以分布式压缩感知的JSM2模型为框架,将多声源定位问题转化为稀疏信号的联合重构问题。以同步正交匹配追踪(Simultaneous orthogonal matching pursuit, SOMP)算法对信号进行联合稀疏重构,进而估计声源方位。在低信噪比的环境和不同声源强度的场景下,对比传统的时延累加定位(Delay-and-sum, DS)算法、SRP-PHAT算法、压缩感知-正交匹配追踪(Compressed sensing-orthogonal matching pursuit, CS-OMP)算法和本文提出的分布式压缩感知-同步正交匹配追踪(Distributed compressed sensing-simultaneous orthogonal matching pursuit, DCS-SOMP)算法对多声源的定位能力。

1 算法

1.1 CS理论

考虑有限长一维离散时间信号 \mathbf{Y} , 可以看作 $N \times 1$ 维的列向量, $\mathbf{Y} = [Y_1, Y_2, \dots, Y_N]^T$ 可用基矩阵 $\Psi = [\Psi_1, \Psi_2, \dots, \Psi_N]^T$ 线性表示为

$$\mathbf{Y} = \mathbf{S}\Psi = \sum_{i=1}^N S_i \Psi_i, \quad (1)$$

其中, \mathbf{S} 系数矢量为 $N \times 1$ 维的列向量, \mathbf{S} 和 \mathbf{Y} 是同一个信号的等价表示, \mathbf{Y} 是信号的时域表示, 而 \mathbf{S} 则是信号的 Ψ 域表示。若矢量 \mathbf{S} 中的非零个数 $\|\mathbf{S}\|_0$ 满足

$$\|\mathbf{S}\|_0 = K \ll N, \quad (2)$$

那么认为信号 \mathbf{Y} 在基矩阵 Ψ 上为 K 稀疏信号, K 为 \mathbf{Y} 信号的稀疏度, 基矩阵 Ψ 为稀疏基。

在CS理论中, 若信号在基矩阵上是稀疏的, 那么稀疏信号可以线性投影到低维空间, 利用低维空间的采样信号, 能够高概率地无损重构出高维原始信号。通常利用一个平稳的、与基矩阵 Ψ 不相关的 $M \times N (M \ll N)$ 维观测矩阵 Φ 对信号 \mathbf{Y} 进行线性变换, 得到 $M \times 1$ 维的观测矢量 $\mathbf{X} = [X_1, X_2, \dots, X_M]^T$, 可表示为

$$\mathbf{X} = \Phi \mathbf{Y} = \Phi \Psi \mathbf{S} = \Theta \mathbf{S}, \quad (3)$$

其中, $\Theta = \Phi \Psi$ 称作CS矩阵。当随机观测矩阵 Φ 满足 $M \geq CK \lg(N/K)$ 时 (C 是一个与恢复精度有关的常数), Θ 有较大概率满足有限等距性质 (Restricted isometry property, RIP) 条件, 可以实现高概率的信号重构^[13]。通过 l_0 范数最小化求解 \mathbf{S} 的精确或者近似逼近 \mathbf{S} , 即

$$\begin{aligned} \hat{\mathbf{S}} &= \operatorname{argmin} \|\mathbf{S}\|_0 \\ \text{s.t. } &\|\mathbf{X} - \Theta \mathbf{S}\|_2 \leq \varepsilon, \end{aligned} \quad (4)$$

其中, ε 为与噪声有关的常量。重构后的系数 \mathbf{S} 在基矩阵 Ψ 上的表示最稀疏。CS理论的核心问题是信号的重构, 目前的重构算法有贪婪追踪算法、凸松弛法、组合算法等三大类^[14]。

1.2 CS算法的DOA估计模型

处于室内远场环境的 M 元麦克风阵列, 第 m 个麦克风接收到的信号 $\mathbf{x}_m(n) (m = 1, 2, \dots, M)$ 为

$$\mathbf{x}_m(n) = \mathbf{h}_m(r_s, n) * \mathbf{s}(n) + \boldsymbol{\omega}_m(n), \quad (5)$$

其中, n 为离散时间序列, $\mathbf{h}_m(r_s, n)$ 是声源位置 r_s 到第 m 个麦克风的房间冲激响应, $\mathbf{s}(n)$ 为声源信号, $\boldsymbol{\omega}_m(n)$ 为第 m 个麦克风接收到的噪声信号, $*$ 为卷积运算。对 $\mathbf{x}_m(n)$ 加窗后经离散傅里叶变换得到频域上的麦克风阵列信号可以表示为

$$\mathbf{X}_m(k) = \mathbf{H}_m(r_s, k) \mathbf{S}(k) + \mathbf{W}_m(k), \quad (6)$$

其中, $\mathbf{X}_m(k)$ 、 $\mathbf{H}_m(r_s, k)$ 、 $\mathbf{S}(k)$ 、 $\mathbf{W}_m(k)$ 分别为 $\mathbf{x}_m(n)$ 、 $\mathbf{h}_m(r_s, n)$ 、 $\mathbf{s}(n)$ 、 $\boldsymbol{\omega}_m(n)$ 的离散傅里叶变换。将其用矢量表示为

$$\begin{aligned} \mathbf{X}(k) &= \{x_1(k), x_2(k), \dots, x_M(k)\}^T, \\ \mathbf{H}(r_s, k) &= \{H_1(r_s, k), H_2(r_s, k), \dots, H_M(r_s, k)\}^T, \\ \mathbf{W}(k) &= \{W_1(k), W_2(k), \dots, W_M(k)\}^T, \end{aligned}$$

则式(6)可表示为

$$\mathbf{X}(k) = \mathbf{H}(r_s, k) \mathbf{S}(k) + \mathbf{W}(k), \quad (7)$$

其中, $\mathbf{X}(k)$ 为频域的麦克风接收信号矢量, $\mathbf{H}(r_s, k)$ 为声源位置 r_s 处的房间频域响应矢量, $\mathbf{W}(k)$ 为频域的噪声矢量。

令声源的空间位置集为 $\{r_1, r_2, \dots, r_I\}$, 其中空间位置集个数 I 远大于目标声源个数, 则冗余的房间频域响应矩阵为

$$\mathbf{D}(k) = \{\mathbf{H}(r_1, k), \mathbf{H}(r_2, k), \dots, \mathbf{H}(r_I, k)\}, \quad (8)$$

称冗余的房间频域响应矩阵 $\mathbf{D}(k)$ 为字典, 在字典扩展下的声源频域的信号矢量可表示为 $\bar{\mathbf{S}}(k) = \{\bar{\mathbf{S}}_1(k), \bar{\mathbf{S}}_2(k), \dots, \bar{\mathbf{S}}_I(k)\}^T$, 则式(7)可以扩展为

$$\mathbf{X}(k) = \mathbf{D}(k) \bar{\mathbf{S}}(k) + \mathbf{W}(k). \quad (9)$$

由于目标声源的个数小于空间位置集个数 I , 因此 $\bar{\mathbf{S}}(k)$ 在空间位置集下具有稀疏性, $\bar{\mathbf{S}}(k)$ 中的非零元素与可能存在的声源位置 r_s 一一对应。但是在实际应用中, 由于房间冲激响应的长度远大于窗函数的长度, 因此 $\bar{\mathbf{S}}(k)$ 中的非零个数大于声源的实际个数。为了获得更准确的声源位置, 往往综合利用多个频点的信息^[7], 离散傅里叶变换长度设置为 K , 则声源的频域信号矢量 $\mathbf{X}(k)$ 包含 $K/2$ 个频点, 设置阈值 $\eta (0 \leq \eta \leq 100)$ 选取能量较高的 $(\eta K/2)\%$ 个频点叠加获得 $\bar{\mathbf{S}} = \sum_{(\eta K/2)\%} \bar{\mathbf{S}}(k)$, 则叠加后的 $\bar{\mathbf{S}}$ 中最大元素所对应的位置为声源信号的空间位置。

在各类信号重构算法中,OMP算法^[15]的计算复杂度较低,该算法的主要过程为计算观测矩阵和压缩采样信号的最大相关位置,通过求解最小二乘问题得到信号的近似解。在迭代次数小于稀疏度的情况下,重复上述过程,输出最大相关的索引集和重构后的信号。

1.3 DCS算法的DOA估计模型

分布式压缩感知算法在经典压缩感知算法的基础上,考虑多组信号之间的共同稀疏性来提高重构性能。对于阵列各个阵元接收信号,可以看作同一个信号经过傅里叶变换,但是由信号传播而导致的幅度衰减和相位偏移^[9]后获得,符合DCS理论中的JSM2模型。在理想环境下,如果声源相对于麦克风阵列的位置保持固定,则可认为此时声源在空间域的稀疏性是固定的,即不同时刻从同一个声源传到麦克风阵列处的信号在空间稀疏性是保持一致的,这种同一个声源连续语音之间具有共同稀疏性的特性,可以在DCS理论中被用于提高稀疏信号的无损恢复概率,从而进一步提高对声源方向的压缩感知估计能力。

在JSM2模型中,第*i*个数据块的接收信号 \mathbf{Y}_i 在基矩阵下可线性表示为

$$\mathbf{Y}_i = \mathbf{S}_i \boldsymbol{\Psi}^T + \mathbf{W}_i, \quad i = 1, 2, \dots, I, \quad (10)$$

其中,*I*表示数据块个数。建立如下优化问题:

$$\hat{\mathbf{S}} = \operatorname{argmin} \sum_{i=1}^N \|\mathbf{S}_i\|_0, \quad (11)$$

其中,*N*表示联合*N*个数据块进行联合稀疏重构。CS矩阵 $\boldsymbol{\Theta}$ 在满足RIP的条件下,对系数 \mathbf{S} 联合重构后的结果在基矩阵 $\boldsymbol{\Psi}$ 上的表示最稀疏。对 \mathbf{S} 的联合重构问题可以采用SOMP算法进行求解。

联合稀疏重构实现声源定位的算法流程如下:

输入:*N*个接收数据块 $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N]$,观测矩阵 $\boldsymbol{\Phi}$,稀疏度*K*,残差误差门限 *thres*。

输出:重构信号 $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_N]$,稀疏位置集 Ω ,残差 \mathbf{r} 。

(1) 初始化:残差 $\mathbf{r}_i^t = \mathbf{Y}_i, (i = 1, 2, \dots, N)$,上标*t*表示迭代次数,初始化*t* = 0,下标表示第*i*个数据块;稀疏位置集 $\Omega = \emptyset$;原子集 $\mathbf{Phit}_i = \emptyset$;第*i*个数据块对应的重构信号 $\mathbf{S}_i = \emptyset, (i = 1, 2, \dots, N)$ 。

(2) 选取观测矩阵 $\boldsymbol{\Phi}$ 中的原子 \mathbf{X}_i 与残差 \mathbf{r}_i^t 做内积,并求出*N*个数据块对应的内积和,计算内积

和的最大值,保存最大值位置 λ_t 的索引和对应的原子 \mathbf{X}_{i,λ_t} ,也就是 \mathbf{X}_i 对应的第 λ_t 列。公式表示为

$$\lambda_t = \operatorname{argmax} \sum_{i=1}^N |\langle \mathbf{X}_i, \mathbf{r}_i^t \rangle|,$$

$$\Omega = \Omega \cup \lambda_t,$$

$$\mathbf{Phit}_i = \mathbf{Phit}_i \cup \mathbf{X}_{i,\lambda_t}. \quad (12)$$

(3) 采用最小二乘法求出信号的近似解:

$$\beta_i = [(\mathbf{X}_{i,\lambda_t})^T \mathbf{X}_{i,\lambda_t}]^{-1} \mathbf{X}_{i,\lambda_t}^T \mathbf{Y}_i, \quad i = 1, 2, \dots, N. \quad (13)$$

保存各个数据块的信号近似解: $\mathbf{S}_i = \mathbf{S}_i \cup \beta_i, (i = 1, 2, \dots, N)$,并求残差:

$$\mathbf{r}_i^t = \mathbf{Y}_i - \mathbf{Phit}_i \mathbf{S}_i, \quad i = 1, 2, \dots, N. \quad (14)$$

(4) 收敛判断:如果残差 $\mathbf{r}_i^t < \text{thres}$,或者迭代次数*t* < *K*,则返回第(2)步继续迭代,否则停止迭代并输出结果。

在上述迭代的过程中可以看出,联合稀疏重构算法在OMP基础上,利用每个数据块具有相同稀疏位置的特性,对数据块之间相同的稀疏部分进行叠加,以进一步提高找出信号稀疏位置的能力。特别的,当*N* = 1时,联合稀疏重构的SOMP算法退化经典的OMP算法。

在理想环境的多声源场景下,声源数量和信号稀疏度*K*是对应的。但是实际环境下由于房间冲激响应长度大于窗函数的长度,算法重构后的信号 \mathbf{S} 的稀疏度*K*大于声源个数,即存在非零值处的空间位置没有声源。DCS理论同样需要综合利用多个频点信息,声源信号经过长度为*L*的离散傅里叶变换后,保留接收信号的*L*/2个频点信息,选择其中能量最高的($\eta L/2$)%个频点做叠加估计。而对于人的语音信号,低频段的能量较高,从而可以直接选择低频部分能量叠加得到 $\mathbf{S}(k) = \{\bar{\mathbf{S}}_1(k), \bar{\mathbf{S}}_2(k), \dots, \bar{\mathbf{S}}_I(k)\}^T$,对于*I*维矢量 $\mathbf{S}(k)$ 中最大的*K*个元素所对应的索引即为估计的多声源的空间位置。

2 数值仿真

2.1 仿真设置

为了评估DCS-SOMP算法对于多声源的定位性能,本文首先进行了仿真评估分析。仿真实验麦克风阵列的直径为65 mm,阵列拓扑结构为圆形阵

列,阵元数为圆周6个阵元。将声源可能存在的空间按照间隔15°划分为24个空间域集。语音信号选取TIMIT标准语音信号,信号时间长度为5 s,语音信号采样率为32 kHz,处理帧长为128个点,每帧信号采用汉明窗加窗处理,做128点离散傅里叶变换。实验通过对语音信号叠加不同程度的高斯白噪声后形成不同信噪比的语音信号,信噪比的计算公式为

$$SNR = 10 \lg \frac{E_{\text{signal}}}{E_{\text{noise}}}, \quad (15)$$

其中, E_{signal} 为多个声源叠加后的信号能量, E_{noise} 为噪声信号的能量。通过房间混响产生工具^[16]获取不同混响程度下的房间冲激响应与语音信号进行卷积形成不同房间冲激响应情况下的语音信号。

考虑实验中设置算法的分辨率为15°,因此采用样条插值法进行均方根误差(Root mean square error, RMSE)做性能评估^[17],计算公式为

$$RMSE = \sqrt{\frac{1}{N} \sum_{K=1}^N (\theta_K - \bar{\theta}_K)^2}, \quad (16)$$

其中, N 为目标声源的方位估计次数, θ_K 为算法估计的目标声源角度, $\bar{\theta}_K$ 为目标声源的实际角度。多声源定位效果的评估方法为先对算法的定位结果做定性分析,在算法对多个声源的定位结果均为正确的情况下,再计算每个声源单独的RMSE结果。

2.2 仿真结果与分析

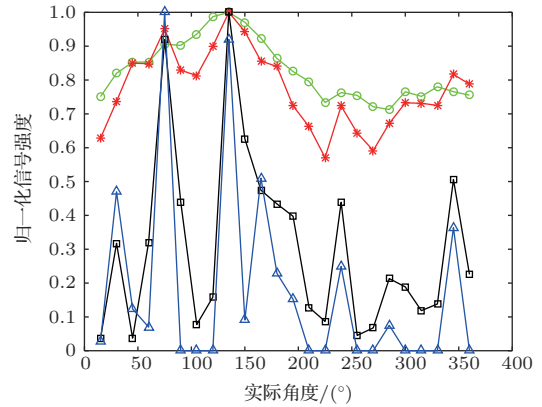
仿真条件下设置声源数为2,分别位于75°和135°的位置,在房间反射强度为0.4,信噪比分别为12.2 dB、8.2 dB、5.7 dB的环境下,几类算法对于多声源的定位效果和RMSE分别如图1和表1所示,由于DS算法没有明显估计出多个声源的方位,

表1 不同信噪比下各算法的多声源分辨的RMSE(仿真)

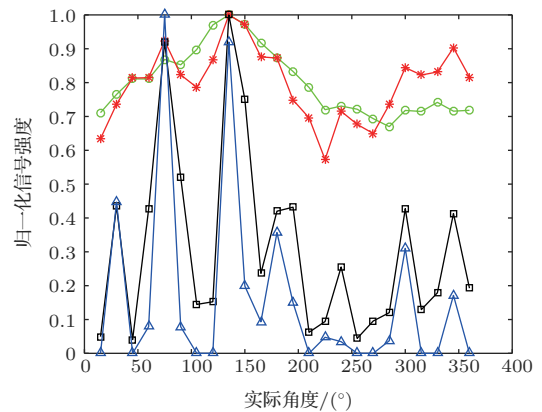
Table 1 Multiple source-resolved RMSE of each algorithm under different SNR (simulation)

SNR/dB	SRP-PHAT		CS-OMP		DCS-SOMP	
	75°	135°	75°	135°	75°	135°
12.2	0.03	2.640	0.24	2.895	0.045	0.045
8.2	0.465	3.765	0.315	3.84	0.135	0.375
5.7	0.885	4.62	0.825	5.58	0.24	0.87

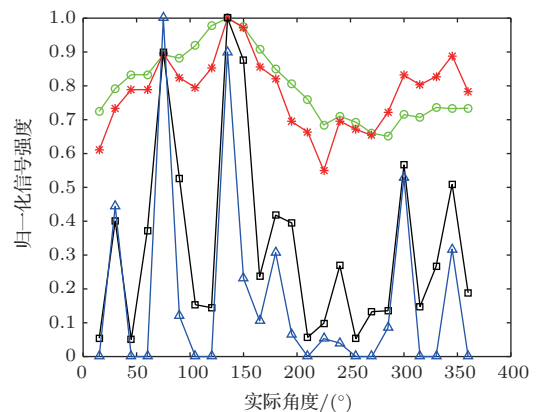
因此不对其做RMSE的比较。其中,CS-OMP算法和DCS-SOMP算法选择的能量最高的频点阈值 $\eta = 35\%$,DCS-SOMP算法对128个数据块进行联合稀疏估计,两类算法的稀疏度均为2。



(a) 信噪比12.2 dB



(b) 信噪比8.2 dB



(c) 信噪比5.7 dB

—○— DS —★— SRP-PHAT —□— CS-OMP —▲— DCS-SOMP

图1 不同信噪比下各算法的多声源分辨性能(仿真)

Fig. 1 Multi-sound source resolution performance of each algorithm under different SNR (simulation)

由于DS算法具有较大的旁瓣能量,在两个声源角度间隔较近的情况下,该算法对多个声源的方位估计能力较低。而SRP-PHAT算法具有相对于DS算法更窄的波束,因此从图1中可以看出SRP-PHAT算法能够估计出两个声源的位置,但是算法的波束指向性不明显,在没有声源处的能量接近于实际声源位置处的能量,导致其RMSE值较大。CS-OMP算法具有更尖锐的波束指向性,相对于前两种算法具有更好的多声源定位性能和更低的RMSE值。CS-OMP算法在信噪比降低的时候旁瓣的波束宽度增大,而DCS-SOMP算法在考虑连续语音帧之间的共同稀疏性的情况下,提高了对多声源方位估计的能力,在信噪比降低后仍然能够保持较好的波束尖锐程度和较低的RMSE值。同时在信噪比较高时,相对于其他几种算法都具有最低的RMSE值,说明DCS-SOMP不仅能够准确估计多个声源的实际方位,同时也具有更好的方位估计性能。

为对比在不同房间反射强度下几类算法的定位性能,在声源数目和位置不变的条件下,设定环境的信噪比为12.2 dB,在房间反射强度分别为0.4、0.6、0.8的情况下,几类不同算法对多个声源的定位效果和RMSE值分别如图2和表2所示。其中,CS-OMP算法和DCS-SOMP算法的参数设置与上文相同。

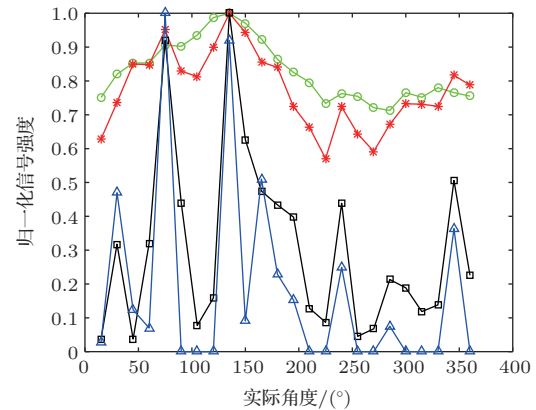
表2 不同房间反射强度下各算法的多声源分辨的RMSE(仿真)

Table 2 Multiple source-resolved RMSE for each algorithm under different room reflection intensities (simulation)

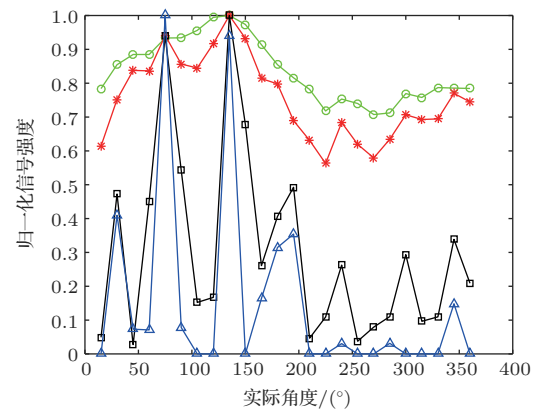
房间反射强度	SRP-PHAT		CS-OMP		DCS-SOMP	
	75°	135°	75°	135°	75°	135°
0.4	0.03	2.640	0.24	2.895	0.045	0.045
0.6	1.5	4.215	4.68	2.865	0.165	0.045
0.8	3.06	4.5	7.77	4.875	0.24	0.93

从图2中可以看出,DS算法由于波束的旁瓣能量较大,因此在不同房间反射强度下,均不能对多个声源的位置有较好的判断。SRP-PHAT算法尽管能够判断出两个声源的位置,但是波束尖锐程度较差,且在房间反射强度增大后,其他角度的能量逐渐接近于实际声源位置的能量,对声源位置的分辨能力下降。CS-OMP算法的波束尖锐程度更

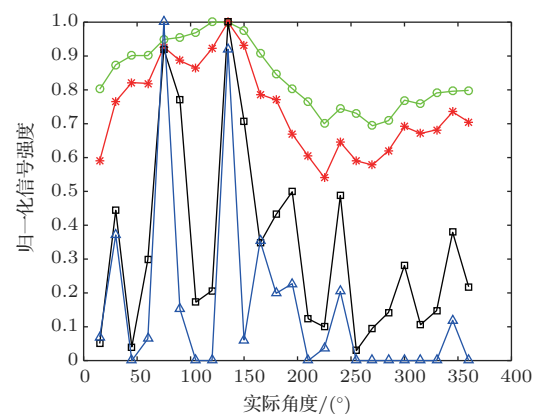
好,随着房间反射强度的增强波束也逐渐变宽。而DCS-SOMP算法则具有最好的波束尖锐程度,同时,在不同房间反射强度的情况下,其波束的尖锐基本



(a) 房间反射强度 0.4



(b) 房间反射强度 0.6



(c) 房间反射强度 0.8

—○—DS —*—SRP-PHAT —□—CS-OMP —△—DCS-SOMP

图2 不同房间反射强度下各算法的多声源分辨性能(仿真)

Fig. 2 Multi-sound source resolution performance of each algorithm under different room reflection intensities (simulation)

上不发生改变。说明DCS-SOMP算法在面对不同的房间反射强度时,其定位效果具有更好的鲁棒性。从表2中可以看出,SRP-PHAT算法、CS-OMP算法和DCS-SOMP算法都能够估计出声源的位置,在房间反射强度增加的情况下,三类算法的RMSE值都有增加,但是DCS-SOMP算法的RMSE值在几类算法中都是最低的,具有更好的方位估计分辨率。

当声源个数继续增多时,声源空间稀疏程度将进一步下降,影响了CS重构性能。为了评估DCS理论应用在多声源场景下的方位估计能力,仿真实验中设置声源数为3,分别位于 75° 、 135° 、 195° ,在信噪比为12.2 dB和房间反射强度为0.4的环境下,分别采用几类算法对声源进行方位估计。设置CS-OMP算法和DCS-SOMP算法的稀疏度为3,其他参数设置与上文相同。几类算法的方位估计结果如图3所示。可以看出,在声源数目增加到3个后,DS算法和SRP-PHAT算法已经基本无法对声源数和声源位置进行准确估计,而CS-OMP算法能够对声源的方位进行准确的估计,但是在 30° 的位置也出现了一个信号强度较高的波束。分析认为该现象的原因在于声源的空间稀疏程度下降后,原本没有声源的位置不再是稀疏的,因此算法重构信号后,在稀疏的位置上出现了较高的波束能量。而DCS-SOMP算法不仅能够估计出三个声源的位置,非声源位置处的信号强度相对于CS-OMP算法也出现了明显的抑制,在多声源的方位估计能力上比CS-OMP算法有了一定的提高。

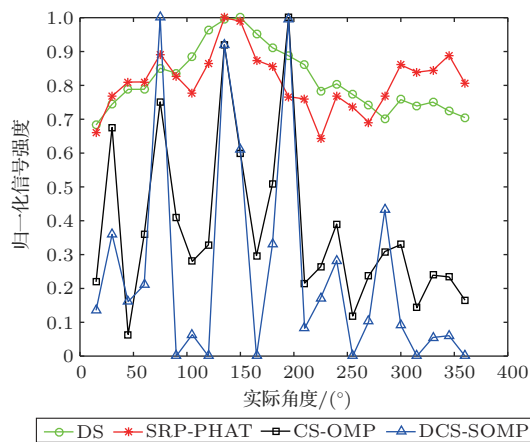


图3 各算法的多声源分辨性能(仿真,声源数=3)
Fig. 3 Multi-sound source resolution performance of each algorithm (simulation, number of sound sources = 3)

3 实验与分析

3.1 实验设置

实验地点为厦门大学艺术学院音乐厅,房间尺寸为 $30\text{ m}\times 20\text{ m}\times 6\text{ m}$,结构为正常音乐厅结构,采用吸声设计。房间的传输响应^[18]可以分为3个部分:直达声的传输响应;在直达信号到达之后的50 ms内到达的早期混响信号的传输响应,该部分严重影响信号处理的性能;以及后期的混响信号的传输响应。房间的早期混响时间通过实测分析^[19]约为30 ms。实验采用六元麦克风圆形阵列采集板进行语音信号的采集,阵列直径为70 mm,语音信号采集采样率为32 kHz。实验采集的语音信号为TIMIT标准语音信号,播放设备为Marshall Kilburn移动式无线蓝牙音箱。设置两个声源分别放置于阵列前方 75° 和 135° 两个方位。实验中播放的语音信号为三段音量不同、内容相同的语音信号,麦克风阵列采集到的三段语音信号信噪比为分别为11.30 dB、6.55 dB、 -0.23 dB 。实验比较了DS算法、SRP-PHAT算法、CS-OMP算法和本文的DCS-SOMP算法对多个声源的定位性能。实验中用到的参数如表3所示。

表3 实验参数表

Table 3 Experimental parameter table

实验变量	参数值
采样率	16 kHz
FFT长度	128点
加窗	汉明窗
联合估计数据块个数	128个
频点阈值(η)(%)	35

3.2 实验结果与讨论

在不同的信噪比下,几类不同算法的定位结果和RMSE分别如图4和表4所示。与仿真实验中一致的是,DS算法并不具有多个声源的方位估计能力,因此在RMSE比较中不考虑DS算法。SRP-PHAT算法能够准确分辨出两个声源的位置,但是指向性尖锐程度和分辨率都较低。CS-OMP算法相对于前两种算法在不同的信噪比下都表现出了更好的指向性尖锐程度,能更明显地估计出两个声源的方位,但是在信噪比降低到 -0.23 dB 的环境

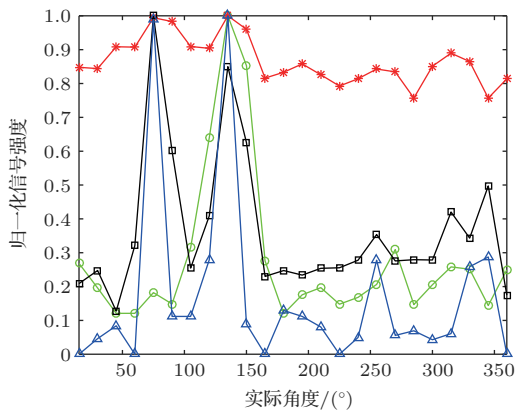
下, CS-OMP算法对声源的估计角度出现了偏差, RMSE相对于SRP-PHAT算法更高。

而本文的算法在压缩感知的基础上, 考虑到连续接收信号具有共同稀疏性, 针对压缩感知算法进一步优化, 对多声源的估计结果中, 具有比CS-OMP算法更好的指向性尖锐程度, 对声源的

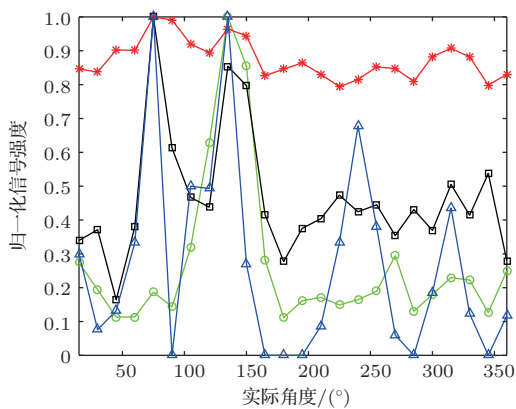
表4 不同信噪比下各算法的多声源分辨的RMSE(实验)

Table 4 Multiple source-resolved RMSE of each algorithm under different SNR (experiment)

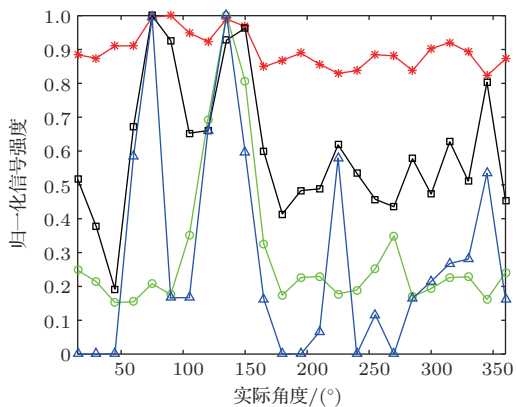
SNR/dB	SRP-PHAT		CS-OMP		DCS-SOMP	
	75°	135°	75°	135°	75°	135°
11.30	5.67	4.365	1.575	2.49	0.3	0.51
6.55	5.82	5.715	1.96	5.685	1.395	0.33
-0.23	7.425	5.535	4.365	9.075	2.445	0.63



(a) 信噪比11.30 dB



(b) 信噪比6.55 dB



(c) 信噪比-0.23 dB

—○— DS —★— SRP-PHAT —■— CS-OMP —▲— DCS-SOMP

图4 不同信噪比下各算法的多声源分辨性能(实验)

Fig. 4 Multi-sound source resolution performance of each algorithm under different SNR (experiment)

位置具有更好的估计效果。同时不同的信噪比下都具有最低的RMSE, 在各类算法中具有更好的分辨率。另外, 对信号强度做归一化以后, 可以看出CS-OMP算法尽管估计出双声源的角度, 但是对声源的信号强度却没有正确的估计结果, 而DCS-SOMP算法在不同信噪比下对于信号强度都有正确的估计结果, 相对于CS-OMP算法对多个声源的方位估计能力有了明显的提升。

为分析算法在声源强度不同的场景下对多声源的定位能力, 实验设置两声源位置和播放内容不变, 录制声音强度比值分别为1:0.75和1:0.5的标准语音信号。几类不同算法的定位结果和RMSE分别如图5和表5所示。可知DS算法对于不同声源强度仍然不具备多声源定位能力。尽管在不同声源强度的情况下, SRP-PHAT算法能够估计多声源方位, 但是指向性尖锐程度和分辨率都很低。同时可以看出在声源强度下降后, SRP-PHAT算法已经出现了对强声源估计角度的偏差, 对声源的指向性效果较差。

CS-OMP算法在声源强度下降后仍然保持较尖锐的波束指向性, 相比于SRP-PHAT算法的RMSE更低。但是在声源强度下降为1:0.5的时候, 对于强度较低的声源信号, 已经出现了估计角度偏差。而DCS-SOMP算法在这几种算法当中具有最好的波束指向性和最低的RMSE, 定位性能最好。同时在声源强度比值下降后, 由于算法考虑多个数据块共同稀疏性, 提升了对弱信号的稀疏重构性能, 因此相对于CS-OMP算法依然能够准确估计两个声源的方位。可见在不同信噪比或者不同信号强度的情况下, 本文的算法对于多声源定位都具有更好的稳健性。

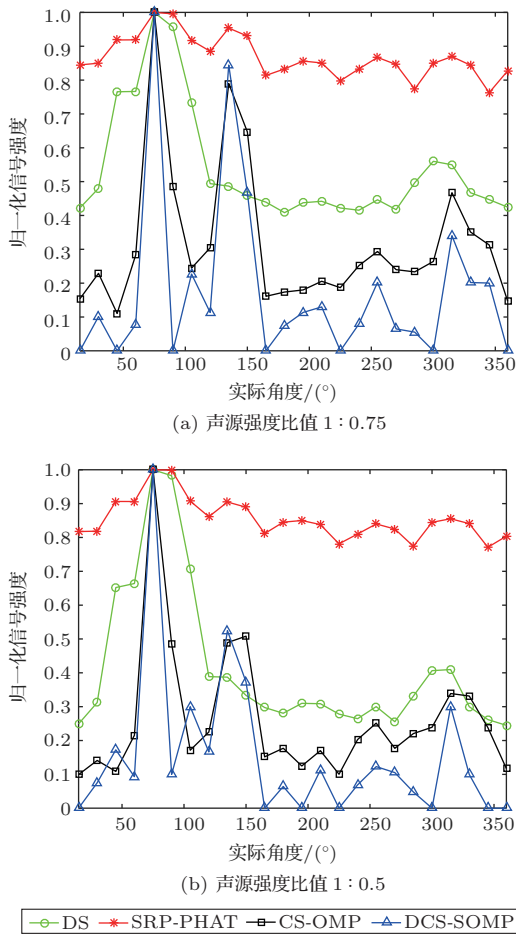


图5 不同声源强度下各算法的多声源分辨性能
Fig. 5 Multi-sound source resolution performance of each algorithm under different sound source intensities

表5 不同声源强度下各算法的多声源分辨的RMSE

Table 5 Multiple source-resolved RMSE of each algorithm under different sound source intensities

比值	SRP-PHAT		CS-OMP		DCS-SOMP	
	75°	135°	75°	135°	75°	135°
1 : 0.75	6.6	5.295	0.84	4.47	0.435	2.52
1 : 0.5	7.125	5.565	1.245	8.61	0.9	3.87

4 结论

本文将分布式压缩感知理论引入到对麦克风阵列的多声源定位场景中, 该理论在CS理论的基础上考虑到接收信号的多个连续语音帧之间具有

共同稀疏性, 对接收信号进行联合稀疏重构进而实现声源定位。仿真和实验结果表明, DCS-SOMP算法在不同的信噪比和不同的房间反射强度的环境下, 相对于传统的DS算法、SRP-PHAT算法和CS-OMP算法, 对多声源的方位估计都具有更好的分辨能力和准确度。同时, 在多个声源的强度不一致的情况下, 本文算法考虑接收信号的共同稀疏性, 在保证对强声源定位精度的同时, 提高了对弱声源的定位能力, 在几类算法中具有更强的稳健性。

致谢 本文在数据采集、测试中得到了厦门大学海洋与地球学院教学实验中心刘志鑫的大力协助, 特此致谢。

参 考 文 献

- [1] Wax M, Kailath T. Optimum localization of multiple sources by passive arrays[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1983, 31(5): 1210-1217.
- [2] Bechler D, Kroschel K. Reliability criteria evaluation for TDOA estimates in a variety of real environments[C]// ICASSP (05). IEEE International Conference on Acoustics, Speech, and Signal Processing. USA: IEEE, 2005, 4: iv/985-iv/988.
- [3] Krim B H, Viberg M. Two decades of array signal processing research: the parametric approach[J]. IEEE Signal Processing Magazine, 1996, 13(4): 67-94.
- [4] Silverman H F, Kirtman S E. A two-stage algorithm for determining talker location from linear microphone array data[J]. Computer, Speech, and Language, 1992, 6(2): 129-152.
- [5] Dibiase J H. A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays[J]. European Journal of Biochemistry, 2000, 216(1): 281-291.
- [6] Donoho D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.
- [7] 赵小燕, 周琳, 吴镇扬. 基于压缩感知的麦克风阵列声源定位算法[J]. 东南大学学报(自然科学版), 2015, 45(2): 203-207. Zhao Xiaoyan, Zhou Lin, Wu Zhenyang. Compressed sensing-based sound source localization algorithm for microphone array[J]. Journal of Southeast University (Natural Science), 2015, 45(2): 203-207.
- [8] Patel K K, Fowler M L. Direction of arrival (DOA) estimation of a wideband acoustic source in multipath environment using spatial sparsity[C]//2017 IEEE Sensors Applications Symposium (SAS). IEEE, 2017: 1-5.
- [9] Baron D, Wakin M B, Duarte M F, et al. Distributed compressed sensing[J]. Preprint, 2012, 21(10): 2729-2732.
- [10] 金光明, 谢植, 张传义. 基于麦克风阵列多声源定位的新方法[J]. 东北大学学报(自然科学版), 2012, 33(6): 769-773.

- Jin Guangming, Xie Zhi, Zhang Chuanyi. A new localization method of multiple sound sources based on microphone array[J]. Journal of Northeastern University (Natural Science), 2012, 33(6): 769-773.
- [11] 章宇栋, 黄惠祥, 童峰. 面向多声源的压缩感知麦克风阵列的波达方向估计[J]. 厦门大学学报(自然科学版), 2018, 57(2): 291-296.
- Zhang Yudong, Huang Huixiang, Tong Feng. Direction of arrival estimation of compressed sensing microphone arrays for multiple sound sources[J]. Journal of Xiamen University (Natural Science), 2018, 57(2): 291-296.
- [12] Carabias-Orti J J, Cabanas-Molero P, Vera-Candeas P, et al. Multi-source localization using a DOA kernel based spatial covariance model and complex nonnegative matrix factorization[C]//2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM). IEEE, 2018: 440-444.
- [13] Candes E, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information[J]. IEEE Transactions on Information Theory, 2006, 52(2): 489-509.
- [14] Needell D, Tropp J A. CoSaMP: iterative signal recovery from incomplete and inaccurate samples[J]. Applied and Computational Harmonic Analysis, 2009, 26(3): 301-321.
- [15] Tropp J A, Gilbert A C. Signal recovery from partial information by orthogonal matching pursuit[EB/OL]. April 2005, www-personal.umich.edu/~jtropp/papers/TG05-Signal-Recovery.pdf.
- [16] Allen J, Berkley D. Image method for efficiently simulating small-room acoustics[J]. Journal of the Acoustical Society of America, 1979, 65(4): 943-950.
- [17] Ganguly A, Reddy C, Hao Y, et al. Improving Sound localization for hearing aid devices using smartphone assisted technology[C]//IEEE International Workshop on Signal Processing Systems. IEEE, 2016: 165-170.
- [18] Yoshioka T, Sehr A, Delcroix M, et al. Making machines understand us in reverberant rooms: robustness against reverberation for automatic speech recognition[J]. IEEE Signal Processing Magazine, 2012, 6(29): 114-126.
- [19] 李剑汶, 章宇栋, 童峰, 等. 一种采用旁瓣增强的麦克风阵列抗混响算法[J]. 厦门大学学报(自然科学版), 2017, 56(5): 711-717.
- Li Jianwen, Zhang Yudong, Tong Feng, et al. A microphone-array de-reverberation method using sidelobe enhancement[J]. Journal of Xiamen University(Natural Science), 2017, 56(5): 711-717.