

· 研究简报 ·

面向多声源的压缩感知麦克风阵列的波达方向估计

章宇栋, 黄惠祥, 童 峰*

(厦门大学 海洋与地球学院, 水声通信与海洋信息技术教育部重点实验室, 福建 厦门 361102)

摘要: 在语音识别、说话人识别等语音交互应用领域中, 麦克风阵列常常工作于多声源工作场景, 因而需要更高的波达方向(DOA)估计分辨性能. 压缩感知(CS)的 DOA 估计算法可将声源定位的问题转化成稀疏信号的重构问题, 进而提高在高混响、低信噪比环境下的 DOA 估计性能. 基于这一思想, 将 CS 方法应用于多声源方位估计. 考虑到传统的基于 CS 的 DOA 估计算法利用实测声源传输响应作为混合矩阵时, 会因噪声的存在而导致多声源条件下的匹配程度下降, 提出了利用基于阵列各阵元之间时延关系所生成的不同方位的声源传输响应来构造 CS 混合矩阵, 即构造房间冲激响应 CS (CRR-CS) 的 DOA 估计算法, 从而实现多声源的 DOA 稀疏恢复. 通过实验验证了该方法优于传统方法, 能更好地实现定位.

关键词: 压缩感知; 麦克风阵列; 多声源; 波达方向

中图分类号: TN 912. 3

文献标志码: A

文章编号: 0438-0479(2018)02-0291-06

麦克风阵列技术如今已广泛应用于智能机器人、视频会议、可穿戴设备、语音增强等领域. 麦克风阵列的语音增强功能主要通过先对声源定位, 再对其方位进行波束增强, 抑制旁瓣来实现. 实际应用中如会议等具有多说话人的场景, 语音增强设备需不断重新定位语音增强方向, 在此类场景下, 如何分辨多人同时说话以及进行定位语音增强成为了新的挑战.

麦克风阵列声源定位技术主要分为 3 大类: 1) 基于最大输出功率的可控波束形成的声源定位技术^[1]在实际使用过程中需要进行全局搜索, 运算量较大, 还会影响到定位的实时性, 同时还需要提前得到信号源频谱特性与环境噪声先验知识的关系, 因此不易实现; 2) 基于高分辨率谱估计的声源定位技术^[2]主要用于窄带信号, 对于语音信号这样的宽带信号, 会导致算法运算量增加, 同样不利于实现; 3) 基于互相关时延估计的声源定位技术^[3]计算量较小、易实现, 但在室内混响严重及低信噪比环境下性能下降严重.

相位变换加权的可控响应功率(SRP-PHAT)定位算法^[4]结合了可控响应功率和相位变换加权的优

点, 比基于时延估计的定位算法具有更好的性能, 但在混响和噪声较强的环境下, 该声源定位算法性能骤降. 此外传统的麦克风阵列针对多声源定位的方法, 如时延估计的 L 型麦克风阵列进行多声源波达方向(DOA)估计^[5]以及利用 L 型的麦克风阵列获得多声源的频率及到达角的联合估计方法^[6], 均是采用 L 型麦克风阵列来获得到达角的空间位置及时延关系, 从而对多声源进行定位, 而本研究则基于压缩感知(CS)理论对多声源进行定位与分辨.

CS 理论最重要的意义在于可以极大地减轻信号采集端的复杂度, 在采集端低采样率的情况下, 信号的接收端能以比较大的概率重构出原始信号. 在水声信道估计中, CS 信道估计可利用信道稀疏特性提高估计性能^[7]. CS 理论突破了原有的奈奎斯特采样定理的束缚, 当信号满足稀疏信号的条件时, 可线性投影到低维空间上, 若 CS 矩阵满足约束等距性(RIP)条件, 则可根据低维的压缩采样信号实现高概率的准确重构^[8].

赵小燕等^[9]基于 CS 的声源定位算法, 通过将麦

收稿日期: 2017-10-17 录用日期: 2017-12-26

基金项目: 福建省高校产学研合作项目(2015H6019); 福建省中青年骨干教师教育科研项目(JAS170012)

*通信作者: ftong@xmu.edu.cn

引文格式: 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.

Citation: ZHANG Y D, HUANG H X, TONG F. Direction of arrival estimation of compressed sensing microphone arrays for multiple sound sources[J]. J Xiamen Univ Nat Sci, 2018, 57(2): 291-296. (in Chinese)



<http://jxmu.xmu.edu.cn>

克风接收的信号转换至频域,将声源可能存在的空间位置所对应的房间频域响应定为特征字典,再利用正交匹配追踪(OMP)算法重构出信号,加强声源的位置信息,从而获得更为鲁棒的声源位置估计.在高混响低信噪比的仿真实验条件下,CS-OMP算法的定位性能要明显优于SRP-PHAT算法.但该算法需要事先测量各声源可能存在的空间位置的房间冲激响应,在实际应用背景下仍有较大的不便性^[10].

在CS-OMP算法的基础上,本研究在高混响低信噪比环境下直接利用麦克风阵列阵元间的时延关系构造房间冲激响应,并进行了在此环境下的多说话人声源方位估计的实验.对SRP-PHAT、DS、CS-OMP以及本研究基于构造房间冲激响应CS(CRR-CS)的DOA估计算法在不同信噪比等多种条件下对多声源的分辨能力进行对比和评估.

1 基于CRR-CS的DOA估计算法

1.1 信号模型

对于一个处在室内环境中的M元线性麦克风阵列,第m个麦克风接收到的信号为

$$x_m(n) = h_m(r_s, n) * s(n) + \omega_m(n), \quad m = 1, 2, \dots, M, \quad (1)$$

其中,*表示卷积运算,n为离散时间序列,s(n)为声源信号,ω_m(n)为第m个麦克风接收到的噪声,h_m(r_s,n)为声源位置r_s到第m个麦克风的房间冲激响应.

1.2 CS理论

假定N×1维复矢量Y=[Y₁,Y₂,...,Y_N]^T可用基矩阵Ψ=[Ψ₁,Ψ₂,...,Ψ_N]^T线性表示为

$$Y = \Psi S = \sum_{i=1}^N S_i \Psi_i, \quad (2)$$

其中,S=[S₁,S₂,...,S_N]^T表示N×1维的系数矢量.

若矢量S中非零的个数||S||₀满足

$$||S||_0 = P \ll N, \quad (3)$$

则称信号Y为基矩阵Ψ上的稀疏信号.式中,||·||为l₀范数,P为Y的稀疏度,Ψ为稀疏基.

在CS理论中,若Y在某已知基矩阵Ψ上的系数是稀疏的,则稀疏的信号Y可以线性投影到低维空间上,利用低维的压缩采样信号可高概率地无损重构出高维的原始信号.通常用一个M×N维观测矩阵Φ对信号Y进行线性变换,得到M×1维的观测矢量

$$X = \Phi Y = \Phi \Psi S = \Theta S. \quad (4)$$

其中,X=[X₁,X₂,...,X_M]^T(M≪N)为观测矢量,Θ=ΦΨ为CS矩阵.

当随机观测矩阵Φ满足M≥CPlog(N/P)时(C是一个与恢复精度有关的常数),CS矩阵Θ能够以较大概率满足RIP条件.若CS矩阵Θ满足RIP条件,则可实现高概率准确的信号重构^[11].在实际情况下,噪声往往无法避免,需对欠定方程(4)引入加性噪声项,即

$$X = \Theta S + W, \quad (5)$$

其中,W为有界噪声.

在Y为稀疏信号的前提条件下,式(5)可以通过l₀范数最小化方式求解,即

$$\begin{aligned} \hat{S} &= \operatorname{argmin} ||S||_0, \\ \text{s.t. } ||X - \Theta S||_2 &\leq \epsilon, \end{aligned} \quad (6)$$

其中,ε是与噪声相关的常量.

CS理论的核心问题是信号的重构,目前已有的重构算法有OMP算法、基追踪(BP)算法等,本研究采用OMP算法.

1.3 DOA估计算法的描述及信号重构

麦克风阵元接收到的信号x_m(n)加窗后,经离散傅里叶变换得到频域信号X_m(k).假如房间冲激响应的长度远小于窗函数的长度,再将其用矢量描述为

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

其中:S(k)为声源信号的频域矢量;X(k)为麦克风接收信号的频域矢量,X(k)=[X₁(k),X₂(k),...,X_M(k)]^T;H(r_s,k)为声源r_s处的房间频域响应矢量,H(r_s,k)=[H₁(r_s,k),H₂(r_s,k),...,H_M(r_s,k)]^T;W(k)为噪声的频域矢量,W(k)=[W₁(k),W₂(k),...,W_M(k)]^T.

声源可能存在的空间位置集为{r₁,r₂,...,r_I},则r_i对应的房间频域响应矢量为H(r_i,k),假定空域离散位置的个数即稀疏度I,当其远大于目标声源的个数时,则冗余的房间频域响应矩阵为

$$D(k) = [H(r_1, k), H(r_2, k), \dots, H(r_I, k)]. \quad (8)$$

将式(8)中冗余房间频域响应矩阵D(k)称为字典,在此条件下,可将式(7)改写为

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

其中,̄S(k)=[̄S₁(k),̄S₂(k),...,̄S_I(k)]^T表示字典D(k)下扩展的频域声源信号矢量.由于空域离散位置的个数I远大于目标声源的个数Ī,因此,̄S(k)在空域下具有稀疏性,且̄S(k)中非零项与可能存在的声源位置r'_i一一对应.

在OMP重构算法中,首先根据信号压缩采样与

CS 矩阵列向量的相关性,来确定稀疏表示中非零系数的位置,再通过求解最小二乘问题,获得稀疏表示中的非零系数^[12],同时 OMP 算法复杂度较低,故用此算法来求解 $\bar{\mathbf{S}}(k)$ 。

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, $\mathbf{X}(k)$ 与 $\mathbf{D}(k)\bar{\mathbf{S}}(k) + \mathbf{W}(k)$ 只是近似的关系,式(9)实际为

$$\mathbf{X}(k) \approx \mathbf{D}(k)\bar{\mathbf{S}}(k) + \mathbf{W}(k), \quad (10)$$

从而导致了根据 OMP 重构算法获得的 $\bar{\mathbf{S}}(k)$ 的非零项个数大于声源的个数。为获得更加准确的声源位置,文献[9]综合利用多个频点信息,将离散傅里叶变换长度设置为 K ,即接收矢量 $\mathbf{X}(k)$ 为 $K/2$ 个频点。由于人的语言信号主要集中于低频段,能量较高,从而可以仅计算这一部分能量较高的频点信息。通过 OMP 算法重构,获得 $\eta K/2$ 个 $\bar{\mathbf{S}}(k)$,叠加获得

$$\bar{\mathbf{S}} = \sum_{k \in \Omega} \bar{\mathbf{S}}(k), \quad (11)$$

其中, Ω 为信号能量从大到小排列的前 $100\eta\%$ ($0 \leq \eta \leq 1$) 的频点,则 I 维矢量 $\bar{\mathbf{S}} = \{\bar{\mathbf{S}}_1, \bar{\mathbf{S}}_2, \dots, \bar{\mathbf{S}}_I\}^T$ 中最大元素对应的空间位置即为估计的声源位置,即

$$\hat{\mathbf{r}}_s = \mathbf{r}_{\arg\max_i \bar{\mathbf{S}}_i}. \quad (12)$$

1.4 时延关系下的 CRR

由于式(10)中噪声项 $\mathbf{W}(k)$ 的存在,多声源场景下根据重构算法获得的 $\bar{\mathbf{S}}(k)$ 中非零元素大于声源个数,即存在非零值对应的空间位置处没有声源;同时,房间冲激响应 $h_m(\mathbf{r}_s, n)$ 还需事先通过测量获得,不便于实际应用。基于此,文中利用麦克风阵列阵元之间的时延关系构造房间冲激响应。考虑阵元间距为 d 的 M 元线性麦克风阵列,有

$$h_m(\mathbf{r}_s, n) = \delta(n, t_m), \quad (13)$$

其中,

$$t_m = \frac{(m-1)d \cos \theta}{c} \times f_s, m = 1, 2, \dots, M, \quad (14)$$

其中, θ 为信号入射方向与阵列夹角, c 为声速, f_s 为采样率^[13]。

2 实验与分析

2.1 实验设置

实验在空间尺寸约为 $30 \text{ m} \times 20 \text{ m} \times 6 \text{ m}$ 的厦门大学艺术学院音乐厅进行。麦克风阵列为阵元间距 13.5 cm 的 7 元均匀线阵,语音信号为 TIMIT 标准语音信号^[14],通过 Marshall 蓝牙音箱播放,采样率为 16 kHz 。将 2 个声源分别放置在阵列前方 $30^\circ, 0^\circ$ 两个方位

上进行播放,对比不同算法性能的多声源 DOA 估计性能。此外利用 Marshall 蓝牙音箱播放噪声进行实录以用于叠加构成不同信噪比场景。

麦克风阵列的波束方向性函数为^[15]:

$$S(\psi, \theta) = \frac{\sin[N\pi fd(\cos \psi - \cos \theta)/c]}{N \sin[\pi fd(\cos \psi - \cos \theta)/c]}, \quad (15)$$

其中, N 为麦克风数, f 为语音段频率, ψ 为定位角。波束宽度为主极大值到理论上出现零时的角度,假定入射角 $\theta = 90^\circ$, $f = 2 \text{ kHz}$, $d = 13.5 \text{ cm}$,可以求得理论波束宽度约为 21.9° ,故实验中将扫描精度设置为 15° ,略小于理论波束宽度。由于是对多声源进行定位与分辨,双声源若设置间隔太远,则易于分辨;若设置角度间隔小于理论波束宽度,则会导致双声源的信号强度都最大而无法判定是否准确分辨。故双声源角度间隔设置为略大于理论波束宽度的 30° 。

原始语音信号信噪比为 15.66 dB ,通过叠加实录的噪声来构造不同信噪比的测试信号,用来对 SRP-PHAT、DS,以及 CRR-CS 算法在不同信噪比条件下对多声源定位估计的分辨性能。其中,通过实测获得的各空间位置到麦克风阵列的冲激响应用于 CS-OMP 算法构造稀疏恢复方程。算法实验参数设置见表 1。

表 1 实验参数设置

Tab 1 The parameters of experiment

信号处理帧长/帧	加窗	OMP 算法频点阈值
128	汉明窗	0.10, 0.15, 0.20

2.2 实验结果及分析

2.2.1 不同信噪比下各算法对多声源的分辨性能

当 $I = 2$, $\eta = 0.15$ 时,不同信噪比条件下各算法对多声源的分辨能力如图 1 所示。

从图 1 可以看出,在不同信噪比条件下 CRR-CS 算法曲线在不同声源方向都有尖锐的指向性,可见 CRR-CS 算法对两个声源的方位估计具有很强的分辨能力,且定位效果很好;DS 算法能够大致估计出声源的位置,但无法分辨两个不同的声源;SRP-PHAT 算法也能够分辨出两个不同的声源位置,但指向性尖锐程度及分辨性能都要低于 CRR-CS 算法;CS-OMP 算法分辨两个声源的能力较差。同时在图中可以发现随着信噪比的下降,各算法对于多声源的分辨能力也在下降,但 CRR-CS 算法仍明显优于其他算法。

考虑到实验设置的分辨率为 15° ,故采用样条插值法进行均方根误差 (RMSE, ϵ_{RMSE}) 性能评估^[16]。各

<http://jxmu.xmu.edu.cn>

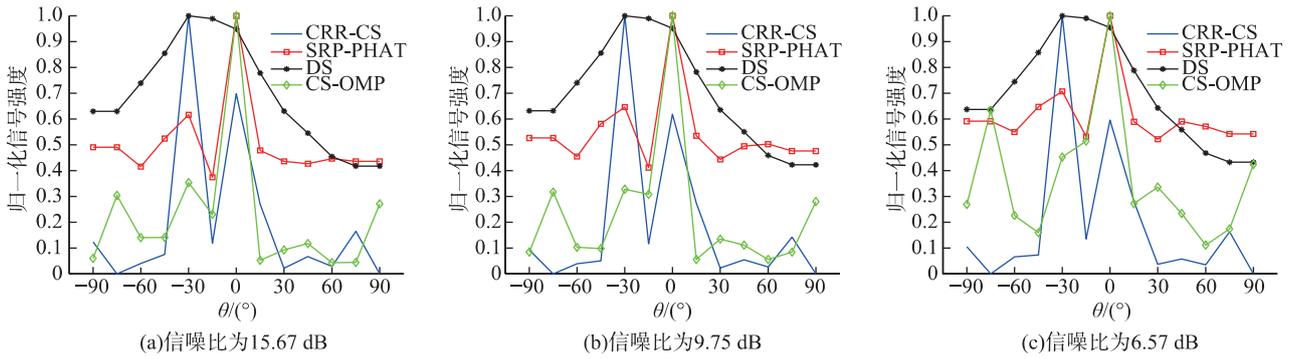


图 1 不同信噪比下各算法对多声源的分辨性能

Fig. 1 Resolving performance of multiple sources for each algorithm at different signal noise ratios

算法信号强度最高的两处分别通过样条插值法获得预测角度,真值方向为声源所在的角度,通过式(16)计算各算法的 RMSE,考虑到 DS 算法无法分辨出 2 个不同的声源,故不参与比较.

$$\epsilon_{\text{RMSE}} = \sqrt{\sum_{i=1}^N (\theta_i - \hat{\theta}_i)^2}, \quad (16)$$

其中, θ_i 为预测角度, $\hat{\theta}_i$ 为其所对应的真值角度, N 为 DOA 算法分辨的角度数. 各算法的 RMSE 如表 2 所示.

从表 2 可以看出各算法随着信噪比的下降, RMSE

表 2 不同信噪比下各算法 DOA 估计的 RMSE

Tab 2 RMSE of DOA estimation for different algorithms at different signal noise ratios

信噪比/dB	$\epsilon_{\text{RMSE}} / (^\circ)$		
	CRR-CS	SRP-PHAT	CS-OMP
15.67	1.12	2.15	5.57
9.75	1.43	4.39	6.41
6.57	1.43	3.73	45.42

总体上有增加的趋势,但 CRR-CS 的结果不但误差较小,且增长小于其余两种算法,而 CS-OMP 算法在 6.57 dB 信噪比环境下的预测角度出现了明显的错误.

2.2.2 不同频点阈值下各算法对多声源的分辨性能

当 $I=2$, 信噪比为 15.67 dB 时,不同算法频点阈值条件下各算法对多声源的分辨能力如图 2 所示.

由于 DS 算法与 SRP-PHAT 算法中无需设置频点阈值,故此部分没有参数发生变化,故 DS 算法与 SRP-PHAT 算法的结果在此处仅作参考.从图 2 中可以看出随着频点阈值 η 的增大, CS-OMP 算法与 CRR-CS 算法的分辨能力都有所提升,但 CRR-CS 算法性能要明显优于 CS-OMP 算法.

2.2.3 不同稀疏度下各算法对多声源的分辨性能

当信噪比为 15.67 dB, $\eta=0.15$ 时,不同声源数 I 条件下各算法对多声源的分辨能力如图 3 所示.

考虑到实际应用场景中往往无法准确获知声源准确数量,文中对设置不同声源数时传统 OMP 算法与 CRR-CS 算法的性能变化信息评估.在实验中分别设 I 为 1,2,3,对比 CS-OMP 和 CRR-CS 算法在不同

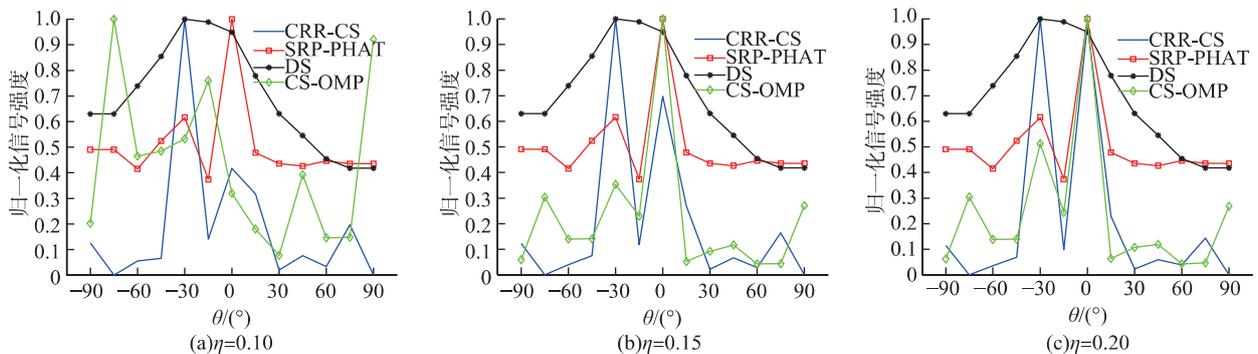


图 2 不同频点阈值下各算法对多声源的分辨性能

Fig. 2 Resolving performance of multiple sources for each algorithm at different frequency threshold

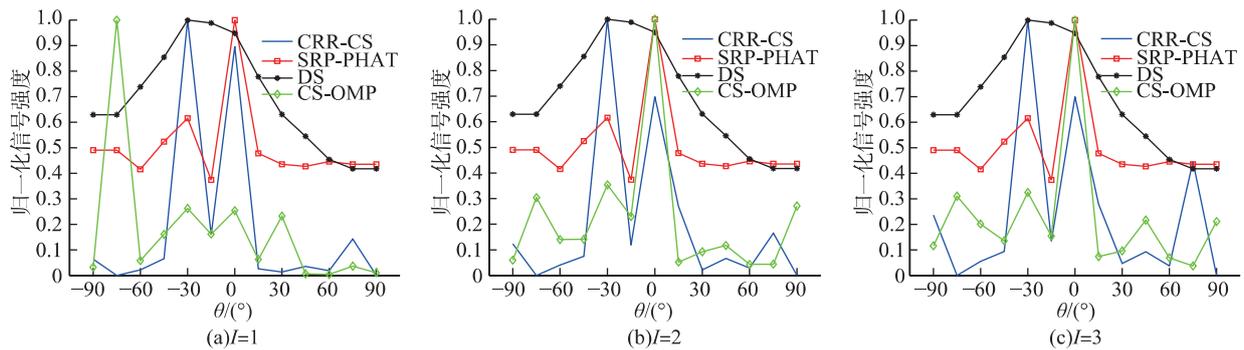


图 3 不同稀疏度下各算法对多声源的分辨性能

Fig. 3 Resolving performance of multiple sources for each algorithm at different sparseness

稀疏度下的分辨性能。 $I=1$ 时, CS-OMP 算法无法成功定位出声源位置且不能分辨多声源, I 为 2 和 3 时能够定位准确但分辨效果差; 尽管 CRR-CS 算法随着稀疏度的增加分辨性能有所下降, 但仍旧远优于 CS-OMP 算法, 可见 CRR-CS 算法对于设置不同声源数的情况下, 其估计结果仍具有一定的稳健性。

3 结 论

本研究将基于 CS 的麦克风阵列定位算法应用于多声源环境当中, 在 CS-OMP 算法的基础上, 考虑到多声源场景, 通过利用阵元间时延关系直接产生的房间冲激响应进行混合矩阵构造, 结果表明, 本文中提出的 CRR-CS 算法对多声源的方位估计能力相比 SRP-PHAT 和 DS 算法更优秀, 且在不同信噪比环境中均对多声源有更好的分辨能力。同时, 在声源数未知及不同频点阈值的情况下, CRR-CS 算法的分辨及定位性能要比 CS-OMP 算法强, 可见: 文中利用麦克风阵列阵元之间的时延关系构造房间冲激响应既减少了需要进行实测的房间冲激响应的步骤, 又提高了多声源分辨能力。

参考文献:

- [1] WAX M, KAILATH T. Optimum localization of multiple sources by passive arrays [J]. IEEE Transaction on Acoustics, Speech, and Signal Processing, 1983, 31 (5): 1210-1217.
- [2] GUSTAFSSON T, RAO B D, TRIVEDI M. Source localization in reverberant environments: modeling and statistical analysis [J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(6): 791-803.
- [3] HUANG L, WU S J, ZHANG L R. A novel MUSIC algorithm for direction-of-arrival estimation without the estimate of covariance matrix and its eigendecomposition [C] // Proceedings of IEEE International Conference on Vehicular Technology. Stockholm: IEEE, 2005: 16-19.
- [4] ZHAO X Y, TANG J, ZHOU L, et al. Accelerated steered response power method for sound source localization via clustering search [J]. Science China Physics, Mechanics and Astronomy, 2013, 56(7): 1329-1338.
- [5] 张艳娜. 基于麦克风阵列的多声源定位算法研究 [D]. 沈阳: 沈阳航空航天大学, 2014: 11-15
- [6] 付金山, 李秀坤. 声矢量阵 DOA 估计的稀疏分解理论研究 [J]. 哈尔滨工程大学学报, 2013, 34(3): 281-286.
- [7] 伍飞云, 童峰. 块稀疏水声信道的改进压缩感知估计 [J]. 声学学报, 2017, 42(1): 27-36.
- [8] 金光明. 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.
- [9] 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.
- [10] 李剑汶, 章宇栋, 童峰. 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.
- [11] CANTLES E, ROMBERG J. Uncertainty principles: exact signal reconstruction from highly incomplete frequency information [J]. IEEE Transactions on Information Theory, 2006, 52(2): 489-509.
- [12] TROPP J A, GILBERT A C. Signal recovery from random measurements via orthogonal matching pursuit [J]. IEEE Transactions on Information Theory, 2007, 53(12): 4655-4666.
- [13] 李芳兰, 周跃海, 童峰. 采用可调波束形成器的 GSC 麦克风阵列语音增强方法 [J]. 厦门大学学报(自然科学版), 2013, 52(2): 186-189.
- [14] GAROFOLO J S, LAMEL L F, FISHER W M, et al. TIMIT acoustic - phonetic continuous speech corpus

- LDC93S1.[DB/OL].[2017-06-22].<https://catalog.ldc.upenn.edu/LDC93S1>.
- [15] BENESTY J, CHEN J, HUANG Y. Microphone array signal processing [M]. Berlin Heidelberg: Springer Science & Business Media, 2008: 43-46.
- [16] GANGNLY A, REDDY C, HAO Y, et al. Improving sound localization for hearing aid devices using smart-phone assisted technology[C]//2016 IEEE International Workshop on Signal Processing Systems (SiPS). Dallas: IEEE, 2016: 165-170.

Direction of Arrival Estimation of Compressed Sensing Microphone Arrays for Multiple Sound Sources

ZHANG Yudong, HUANG Huixiang, TONG Feng*

(Key Laboratory of Underwater Acoustic Communication and Marine Information

Technology Ministry of Education, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China)

Abstract: In applications of voice recognition, speaker recognition and other voice interaction, microphone arrays often work in multiple sound sources scenes. Therefore microphone arrays need higher resolution performance of direction of arrival (DOA) estimation. Compressed sensing—direction of arrival (CS-DOA) transforms the problems of sound source localization into the reconstruction of sparse signals, thus can improve the performance of DOA estimation in high reverberation and low SNR environment. In this paper, the compressed sensing is applied to location estimation of multiple sound sources, the traditional CS-DOA method using the measured sound sources transmission response as a hybrid matrix, which will result in a decrease in the degree of matching under multiple sound sources due to the noise. This article constructs the compression-sensing hybrid matrix based on the different azimuth sound sources transmission response generated by the array-element delay relation, which constructs the room impulse response (constructed room response, CRR), so as to achieve multiple sound sources DOA sparse recovery. Finally experimental results are given to verify the effectiveness of the proposed method.

Key words: compressed sensing (CS); microphone array; multiple sound sources; direction of arrival (DOA)