

基于压缩感知的双麦克风混响多声源定位算法

张奕, 李娟, 张敏

(大连大学信息工程学院, 辽宁 大连 116622)

摘要: 针对混响条件下现有声源定位技术中麦克风数量必须大于声源数量的现状, 提出了一种基于压缩感知的双麦克风混响多声源 (至少 3 个声源) 定位算法。将多声源定位问题看作是块稀疏信号的重构问题, 在频域将全房间冲激响应归一化来构造压缩观测矩阵, 重构的块稀疏信号中非零块的位置即对应了空间中实际声源的位置。仿真实验表明, 与基于子带可控响应功率 (SRP-sub) 的多声源定位方法相比, 在双麦克风混响条件下定位多声源, 基于压缩感知的多声源定位算法的定位性能更高, 在混响时间为 0.6 s 时, 仅采用 40 个频点值, 定位 3 个声源的成功率可以达到 80%。

关键词: 压缩感知; 混响; 麦克风阵列; 块稀疏; 多声源定位

中图分类号: TN912.3

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2019011

Reverberation multi-source localization algorithm based on compressed sensing with dual microphones

ZHANG Yi, LI Juan, ZHANG Min

College of Information Engineering, Dalian University, Dalian 116622, China

Abstract: In traditional multi-source localization field, it is necessary to guarantee that the number of microphone is more than the number of source. To overcome this constraint, a dual-microphone multi-source localization algorithm based on CS was proposed, where the number of sound source localized successfully was more than 3. The multi-source localization was regarded as the block sparse signal reconstruction in this algorithm, and the full room impulse responses normalized were exploited to construct the compressed observation matrix in frequency domain. In reconstructed block sparse signal, the positions of non-zero blocks were corresponded to the positions of sound sources in space. The simulation shows that compared with the SRP-sub algorithm, in reverberation time 0.6s with dual-microphone, the proposed multi-source localization algorithm based on compressed sensing has higher capability which can reach 80% success rate by using 40 frequency points to localize 3 sound sources.

Key words: compressed sensing, reverberation, microphones array, block sparse, multi-source localization

1 引言

目前, 在视频会议、智能机器人、车载电话、语音增强等领域中, 基于麦克风阵列的声源定位技术现已得到广泛应用^[1]。传统的声源定位算法主要基于以下 3 种: 1) 基于波束形成的定位算法, 该算法采用波束形成技术, 通过调节麦克风阵列的指向

对整个接收空间进行搜索, 使阵列波束对准声源而获得最大的输出功率^[2]; 2) 基于高分辨率谱估计的定位算法, 通过求解阵列接收信号的协方差矩阵获得空间谱函数, 再由此信息进行声源定位, 使用较广泛的算法有 MUSIC (multiple signal classification)、ESPRIT (estimating signal parameter via rotational invariance techniques)、子空间拟合等^[3];

收稿日期: 2018-05-14; 修回日期: 2018-11-07

基金项目: 国家自然科学基金资助项目 (No.61201420)

Foundation Item: The National Natural Science Foundation of China (No.61201420)

3) 基于到达时延差的方法, 此为当前应用最广泛的一种定位方法, 该算法根据声源信号到达不同位置麦克风的时间差, 计算出声源位置相对于多组麦克风所在的多个双曲面, 其交点就是声源的位置^[4]。然而传统的麦克风阵列声源定位技术都无法摆脱三角定位的限制, 在混响环境下受到更大挑战, 即使采用信道盲分离技术, 可恢复的空间声源数目也无法超过麦克风数量, 这使得现有的具有实用价值的麦克风阵列系统都是海量阵列。

压缩感知 (CS, compressed sensing) 理论提出以后, 很多学者将其引进到语音信号处理的多个领域中。例如, Asaei 利用结构性稀疏的压缩感知框架来进行多说话人语音识别, 实现了 4 个麦克风以 95% 的正确率识别 9 个说话人^[5]。Wabnitz 利用压缩感知框架进行三维空间声场重构, 极大地减少了数据集的处理量, 并扩大了听觉甜点的范围^[6]。Benichoux 利用压缩感知技术, 突破了全房间冲激响应估计中麦克风数量大于等于声源数目的限制^[7]。Mignot 利用压缩感知技术通过低频插值来恢复房间的全冲激响应^[8], 仅采用 120 个麦克风的采样点即可恢复全房间冲激响应, 极大地减少了麦克风的使用数量。Chardon 采用 MMV (multi-measurement vector) 框架, 利用随机稀疏字典构造广义 MUSIC 算法进行多声源识别和定位^[9]。广义 MUSIC 算法在较少麦克风情况下仍可取得理想定位效果, 但该算法基于理想模型假设, 只适用于线性阵列。

文献[10]基于压缩感知框架, 提出了使用麦克风阵列定位单个声源的定位算法。该算法假设麦克风阵列和声源处于同一水平面, 将声源可能存在的空间离散成 I 个位置, 并以这 I 个位置的房间冲激响应来构建字典, 再采用 OMP (orthogonal matching pursuit) 算法恢复出声源信号矢量, 则矢量中最大元素对应的位置即为实际声源的位置。相位变换加权的可控响应功率 (SRP-PHAT, steered response power-phase transform)^[11] 定位算法通过搜索空间中使 SRP (steered response power) 函数最大的点作为声源的估计位置, 该类算法对混响的顽健性较强, 但在多声源条件下会受到各声源之间的相互影响, 使传统的 SRP-PHAT 算法的多声源定位性能并不高。文献[12]对此做出改进, 提出了基于子带可控响应功率的多声源定位算法 (SRP-sub, sub-band steered response power), 该算法对语音信号频谱进行划分之后, 在每个子带寻求 SRP 的最大值以定位声源。以

上算法虽然都能在混响环境下达到声源定位的目的, 但要取得显著的定位效果, 所采用的麦克风数目必须远多于空间中声源的数目。

本文基于压缩感知框架, 结合房间混响所带来的空间信息来构建室内混响环境下采用双麦克风进行多声源 (至少 3 个声源) 定位的模型。该模型将混响环境下的多声源定位描述为多点声源经由房间冲激响应模板压缩到麦克风的压缩感知问题, 从而提出了一种基于压缩感知的双麦克风混响多声源定位算法。由于在本文的研究框架下, 多点声源相对房间的整个空频域而言呈现块稀疏特性, 而块稀疏所需的观测量 (麦克风数目) 要比普通稀疏的观测量少得多, 故对块稀疏信号的求解使得使用欠定数量 (麦克风数目小于声源数目) 的麦克风阵列定位多声源成为可能。

2 混响多声源问题描述

室内混响环境下, 声源数目未知的多声源多信道的时域模型为

$$x_m(t) = \sum_1^N h_{mn}(t) * s_n(t), \quad m=1, \dots, M \quad (1)$$

其中, x_m 表示第 m 个麦克风接收的信号, $s_n(t)$ 表示第 n 个声源信号, $h_{mn}(t)$ 表示第 n 个声源与第 m 个麦克风之间的房间冲激响应, $*$ 为卷积符号, M 为麦克风的数量^[13] (本文中 $M=2$)。经短时傅里叶变换得到式(2)。

$$X_m(f, T) = \sum_{n=1}^N H_{mn} S_n(f, T), \quad m=1, \dots, M \quad (2)$$

3 基于压缩感知的双麦克风混响多声源定位算法介绍

3.1 压缩感知理论

假设 $N \times 1$ 维的矢量 $\mathbf{Z} = \{Z_1, Z_2, \dots, Z_N\}^T$ 在稀疏基 $\boldsymbol{\psi} = \{\psi_1, \psi_2, \dots, \psi_N\}$ 上有如式(3)的表示。

$$\mathbf{Z} = \boldsymbol{\psi} \mathbf{S} = \sum_{i=1}^N S_i \psi_i \quad (3)$$

则称在稀疏基 $\boldsymbol{\psi}$ 上, 信号 \mathbf{Z} 是稀疏的。其中 $\mathbf{S} = [S_1, S_2, \dots, S_N]^T$, 矢量 \mathbf{S} 中非零值的个数为 k , 也被称为稀疏度, 且 $k \ll N$ 。

压缩感知理论指出,若矢量 \mathbf{Z} 本身稀疏或者其在某一已知基 $\boldsymbol{\psi}$ 上的系数呈现稀疏性,则可通过一个 $M \times N$ 维 ($M \ll N$) 的压缩观测矩阵 $\boldsymbol{\Phi}$ 将矢量 \mathbf{Z} 线性投影到一个低维的观测矢量 \mathbf{X} 中^[14],其中 $\mathbf{X} = [X_1, X_2, \dots, X_M]^T$, 该线性投影可以表示为 $\mathbf{X} = \boldsymbol{\Phi}\mathbf{Z} = \boldsymbol{\Phi}\boldsymbol{\psi}\mathbf{S} = \mathbf{A}\mathbf{S}$, 其中 \mathbf{A} 被称为 CS 矩阵, $\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\psi}$ 。

当 CS 矩阵 \mathbf{A} 满足有限等距约束时,通过低维的观测数据可以高概率重构出原始信号^[15]。信号 \mathbf{Z} 稀疏或 \mathbf{Z} 在某个基上稀疏的情况下,欠定方程组 $\mathbf{X} = \mathbf{A}\mathbf{S}$ 可以通过 l_0 范数最小化约束来求解^[16],即

$$\hat{\mathbf{S}} = \operatorname{argmin} \|\mathbf{S}\|_0 \quad \text{s.t. } \mathbf{X} = \mathbf{A}\mathbf{S} \quad (4)$$

由于 l_0 范数求解属于非凸问题,一般采用 l_p ($0 < p \leq 1$) 范数近似地对 \mathbf{S} 的稀疏性进行约束而取得更好的重构效果。

根据 CS 理论,信号重构可以看作是在一定条件下的优化问题。现有的重构算法主要基于以下三类^[17]: 第一类为贪婪迭代算法,主要有匹配追踪 (MP, matching pursuit) 算法、正交匹配追踪 (OMP, orthogonal matching pursuit) 算法等^[18]; 第二类为凸优化算法,常用的有基追踪 (BP, basic pursuit) 算法^[19]; 第三类是基于贝叶斯框架提出的算法^[20]。

3.2 块稀疏表示

为简化问题,本文将房间看作二维平面 (俯视图),将该平面区域分割成 G 个网格^[21],每个声源最多只占据一个网格。假设房间中存在 K 个声源 ($K \ll G$),且在时间观测窗内声源位置不变,在每个网格 g 处选取 F 个频点,即 $\mathbf{S}_g \in C^{F \times 1}, g \in \{1, \dots, G\}$,并将所有网格点的频点数据按列堆叠成一维向量 $\mathbf{S} = [\mathbf{S}_1^T, \dots, \mathbf{S}_G^T]^T \in C^{FG \times 1}$ 。该方法同样适用于三维空间,只需要将三维空间划分的网格平铺到一维即可。当空间中只存在 K 个声源时,理论上有声源的网格对应的 $\mathbf{S}_i (i \in [1, G])$ 在频点处就会出现非零值。如图 1 所示,房间被划分成 G 个网格,黑色网格表示声源所在位置,其频域表示如图 1 右侧所示,而其他没有声源的网格在各频点的值均为 0。这使得多个声源信号的空-频域表达 \mathbf{S} 具有某种结构稀疏性,由于在这一结构中,非零元素成块出现,故也称声源信号在空-频域上具有块稀疏性,即有几个说话人 (声源) 就对应几个非零块,块的大小等于选取的频点数。

3.3 构造压缩观测矩阵

假设已知房间为有限阻抗墙壁组成的矩形封闭结构,则可利用 Image 方法^[22]构造房间中任意空间点声源到达某一空间位置麦克风的房间冲激响应。Image 方法又叫镜像法,麦克风接收到的声源信号包括了直达声和经多次反射后的信号,Image

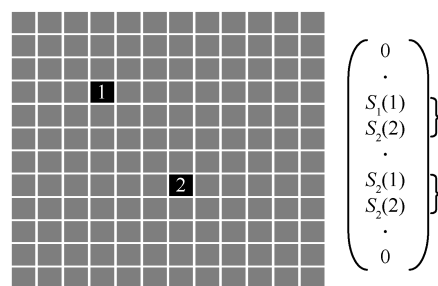


图 1 具有块稀疏结构的声源位置分布的空-频域示意

方法与光的镜面反射特性相似,将反射的信号看成是墙壁另一侧的虚声源传播到麦克风的信号。考虑到多径效应,则网格 g 处到第 m 个麦克风的投影 $\xi_{V_g^r \rightarrow \mu_m}^f$ 可采用 Green 函数表示,如式(5)所示。

$$\xi_{V_g^r \rightarrow \mu_m}^f : X(f, \tau) = \sum_{r=0}^R \frac{l^r}{\|\boldsymbol{\mu}_m - V_g^r\|^\alpha} e^{-j\frac{\boldsymbol{\mu}_m - V_g^r}{c}} S(f, \tau) \quad (5)$$

其中, $j = \sqrt{-1}$; f 表示频率; $\boldsymbol{\mu}_m$ 和 V_g^r 分别表示第 m 个麦克风的位置和第 g 个网格点的第 r 个虚声源的位置; l^r 为第 r 次反射的反射系数; α 为衰减常数,在球面传播模型中, $\alpha = 1$; c 表示声速^[23]。采用式(5)的声源-麦克风投影构造 F 个频点的频域投影矩阵为

$$\Xi_{V_g \rightarrow \mu_m} = \operatorname{diag}(\xi_{V_g^1 \rightarrow \mu_m}^1, \dots, \xi_{V_g^F \rightarrow \mu_m}^F) \quad (6)$$

则所有网格点经由房间投影到第 m 个麦克风对应的压缩观测模板为

$$\boldsymbol{\Phi}'_m = [\Xi_{V_1 \rightarrow \mu_m} \dots \Xi_{V_g \rightarrow \mu_m} \dots \Xi_{V_G \rightarrow \mu_m}] \quad (7)$$

所以麦克风阵列对应的压缩观测模板为

$$\boldsymbol{\Phi}' = [\boldsymbol{\Phi}'_1, \dots, \boldsymbol{\Phi}'_M]^T \quad (8)$$

将 $\boldsymbol{\Phi}'$ 按列归一化后得到压缩观测矩阵如式(9)所示。

$$\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_M]^T \quad (9)$$

3.4 本文提出的定位算法

基于以上描述,本文算法整体流程如下。

1) 采用 Image 方法产生全房间冲激响应,并根据

式(6)~式(9)完成压缩观测矩阵 Φ 的构造。

2) 利用压缩感知框架描述语音定位问题, 并采用块稀疏优化算法恢复出声源信号的估计 \hat{S} 。

3) 将 \hat{S} 重新按频点数分块得到声源信号矢

$$\begin{bmatrix} X_1(1) \\ \vdots \\ X_1(F) \\ X_2(1) \\ \vdots \\ X_2(F) \end{bmatrix} = \begin{bmatrix} \left(\begin{array}{ccc} \Phi_{11}(1) & & \\ & \ddots & \\ & & \Phi_{11}(F) \end{array} \right) \left(\begin{array}{c} \Phi_{21}(1) \\ \vdots \\ \Phi_{21}(F) \end{array} \right) \\ \vdots \\ \left(\begin{array}{ccc} \Phi_{12}(1) & & \\ & \ddots & \\ & & \Phi_{12}(F) \end{array} \right) \left(\begin{array}{c} \Phi_{22}(1) \\ \vdots \\ \Phi_{22}(F) \end{array} \right) \end{bmatrix} \begin{bmatrix} \left(\begin{array}{ccc} \Phi_{G1}(1) & & \\ & \ddots & \\ & & \Phi_{G1}(F) \end{array} \right) \\ \vdots \\ \left(\begin{array}{ccc} \Phi_{G2}(1) & & \\ & \ddots & \\ & & \Phi_{G2}(F) \end{array} \right) \end{bmatrix} \begin{bmatrix} S_1(1) \\ \vdots \\ S_1(F) \\ \vdots \\ S_G(1) \\ \vdots \\ S_G(F) \end{bmatrix} \quad (10)$$

式(3)写成矩阵的形式为

$$X = \Phi S \quad (11)$$

式(10)中, 等式左侧项表示 2 个麦克风接收的信号在不同频点的值, 其中, 下标 1、2 为麦克风编号, 即第一个麦克风和第二个麦克风。等式右侧第一项为根据式(6)~(9)选取的 F 个频点构造的压缩观测矩阵, 第二项为各个网格点的信号在频域中表示, 每个网格各取 F 个频点。理论上, 有声源的网格对应的频点值存在非零值, 而无声源的网格处,

$$\begin{bmatrix} X_1(1) \\ X_1(2) \\ X_2(1) \\ X_2(2) \end{bmatrix} = \begin{bmatrix} \left(\begin{array}{cc} \Phi_{11}(1) & \\ & \Phi_{11}(2) \end{array} \right) \left(\begin{array}{c} \Phi_{21}(1) \\ \Phi_{21}(2) \end{array} \right) \\ \left(\begin{array}{cc} \Phi_{12}(1) & \\ & \Phi_{12}(2) \end{array} \right) \left(\begin{array}{c} \Phi_{22}(1) \\ \Phi_{22}(2) \end{array} \right) \\ \left(\begin{array}{cc} \Phi_{31}(1) & \\ & \Phi_{31}(2) \end{array} \right) \left(\begin{array}{c} \Phi_{41}(1) \\ \Phi_{41}(2) \end{array} \right) \\ \left(\begin{array}{cc} \Phi_{32}(1) & \\ & \Phi_{32}(2) \end{array} \right) \left(\begin{array}{c} \Phi_{41}(1) \\ \Phi_{41}(2) \end{array} \right) \end{bmatrix} \begin{bmatrix} S_1(1) \\ S_1(2) \\ 0 \\ 0 \\ S_3(1) \\ S_3(2) \\ S_4(1) \\ S_4(2) \end{bmatrix} \quad (12)$$

对于求解如式(12)所示的欠定线性方程组, 如果仅从普通稀疏优化的角度, 即基于 l_0 范数优化的角度, 信号 S 的维度 $n=8$, 非零变量个数为 6 (稀疏度 $k=6$), 则至少需要 $m=6$ 个观测数据才能恢复; 如果采用基于 l_1 范数的优化算法, 则需要满足更严格的 RIP (restricted isometry property) 约束条件, 所需的观测数据量 $m \approx k \text{lb} \left(\frac{n}{k} \right)$ 。因此, 式(12)的 4 个观测数据采用普通稀疏优化时, 无法唯一恢复信号, 即采用普通稀疏算法时, 无法仅使用 2 个麦克风定位 3 个声源。然而, 如

量 S , 将其非零块与房间网格进行对应, 求出声源位置。

综上, 使用双麦克风对多个声源进行定位的问题可表示为求解线性方程组, 如式(10)所示。

其所有频点值对应为零值; 即空间中声源的位置和式(11)信号 S 中的非零块所在的位置之间存在一一对应的关系, 故估计空间中多声源的位置问题可转换为确定非零块在 S 中的位置问题。

要确定 S 中的非零块需要先通过求解式(10)的欠定线性方程组来获得声源信号的估计 \hat{S} 。为简化问题描述, 在本文的双麦克风定位研究中, 假设使用频点数为 2, 声源数目为 3, 麦克风数为 2, 并将房间划分为 4 个网格, 则式(10)可以简化为式(12)。

3.2 节所述, 由于该声源信号在空-频域具有结构稀疏性, 所以, 当对其进行块结构约束时, 对于信号 S 的 $N=4$ 个块中, 非零块 $K=3$, 每个非零块大小 $d=2$, 则采用块稀疏优化算法 (基于 l_0 范数优化) 只需 3 个观测数据即可得到该 3 个非零块的全部数据; 若采用基于 l_1 范数的优化算法, 则需满足 block RIP, 即观测数据量最少满足 $m \approx K \text{lb} \left(\frac{N}{K} \right) = \frac{k}{d} \text{lb} \left(\frac{n}{k} \right)$ 。可见块稀疏所需的观测数据量是普通稀疏数据量的 $\frac{1}{d}$, 此时 2 个麦克风可以

定位 3 个声源, 即采用块稀疏优化时, 式(12)可解。所以, 式(10)的欠定线性方程组可采用块稀疏约束求解。

本文采用 ReGOMP(regularized group orthogonal matching pursuit)^[24]块稀疏重构算法求解, 这是 Majumdar 提出的一种结合了正则化正交匹配追踪 (ROMP, regularized orthogonal matching pursuit) 算法^[25]和块正交匹配追踪 (BOMP, block orthogonal matching pursuit) 算法^[26]的针对块稀疏信号的重构算法。该算法在每次迭代时首先挑选出与残差最匹配的 K 个块, 然后在正则化过程中筛选出其能量介于最大能量与最大能量的一半之间的块集合作为候选支撑块集合。正则化过程提高了筛选的准确性, 同时相比 BOMP 算法每次只选择一个支撑块, ReGOMP 算法每次选取多个支撑块, 加快了收敛速率, 解决了 ROMP 算法不能解决的块稀疏问题; 其缺点是块稀疏度必须先验已知。

在使用 ReGOMP 算法求得声源信号的估计 \hat{S} 后, 根据选取的频点数 F 重新整合分块。如式(13)所示, 对每 F 个频点求解其 l_2 范数, 则非零范数 (非零块) 在 \hat{S} 中的位置对应的就是空间中声源所在网格点的位置。

$$\sqrt{\sum_{f=1}^F |\hat{S}_g(f)|^2} > 0, g \in [1, G] \quad (13)$$

4 仿真实验与分析

为进一步验证上述所提出的算法的有效性, 本文使用 Matlab 2014b 进行软件仿真实验。首先对不同说话人同时位于房间不同位置时的情况进行双麦克风定位模拟, 然后分别就混响时间 RT_{60} 对定位成功率的影响, 频点数 F 的变化对定位成功率的影响以及可定位的声源数目增加时算法的定位性能进行讨论, 并分别和文献[12]的多声源定位算法 SRP-sub 进行仿真实验对比, 最后从平均运行时间的角度分析了本文算法的复杂度。

4.1 实验设置

仿真使用的矩形房间为 3.3 m×4.4 m×2.5 m (长×宽×高)。分别以 0.3 m×0.4 m 为间隔将房间二维平面划分为 10×10 的网格, 2 个全向麦克风位置坐标分别为 [1.6, 1.6, 1.7] m 和 [1.4, 1.9, 1.7] m。假设说话人

与麦克风阵列处于同一水平面, 采样频率 $F_s=16$ kHz, 声速 $c=343$ m/s, 墙壁反射系数为 0.7, 采用 Image 方法产生全房间冲激响应, 长度约为 2 500 点, 统一截取前 1 000 个频点的值, 在频域进行 2 048 点的傅里叶变换, 并等间隔选取 F 个频点构造压缩观测矩阵。

图 2 模拟了不同说话人 (声源) 同时位于房间不同位置时, 采用本文算法和文献[12]算法的定位情况。混响时间 $RT_{60}=0.3$ s, 频点数 $F=80$ 。

从图 2 可以看出本文算法对多声源定位的优势。SRP-sub 算法虽然采用子带划分的思想来避免多个声源距离较近时出现易混淆的现象, 但在麦克风数量少于声源数目时, 表现依然乏力。所以, 在一定的误差条件下, 使用双麦克风在混响条件下进行多声源定位是可行的。

4.2 混响时间 RT_{60} 的影响

本实验中, 混响时间 RT_{60} 分别取 0.2 s、0.3 s、0.4 s、0.5 s、0.6 s 产生全房间冲激响应, 等间隔取 $F=40$ 个频点构造压缩观测矩阵, 并分别采用本文算法和 SRP-sub 算法进行多声源定位, 两者在不同混响时间下的定位结果如图 3 所示。由图 3 可知, 当混响时间增加时, 本文算法的定位能力不但没有降低反而略有上升的趋势, 可见本文算法对混响时间的顽健性。而 SRP-sub 算法在多声源情况下本身定位性能不高, 高混响更是严重影响了其对多声源定位的能力。

4.3 频点数 F 对定位性能的影响

参考 4.2 节的仿真结果, 取混响时间 $RT_{60}=0.3$ s, 统计频点数 F 取值为 20、40、60、80、100、120 时分别采用本文算法和 SRP-sub 算法对多声源的定位成功率, 其结果如图 4 所示。

由图 4 可知, 随频点数的增加, 本文算法的定位成功率逐渐提升, 但超过某一频点数时, 其上升趋势渐趋缓慢。这是因为本文算法在信号重构之后需要再整合 F 个频点的能量信息来进行分块, 当频点数达到一定数值时, 由于这些频点已经包含了整个信号的绝大部分能量, 所以继续增加频点数目, 对算法的定位性能提升较少。而对于 SRP-sub 算法, 由于其每个子带的 SRP 函数也依赖于各频点的值, 所以频点数的增加也使 SRP-sub 算法的定位性能得到提升, 但整体而言, 本文算法在双麦克风条件下的多声源定位性能远高于 SRP-sub 算法。

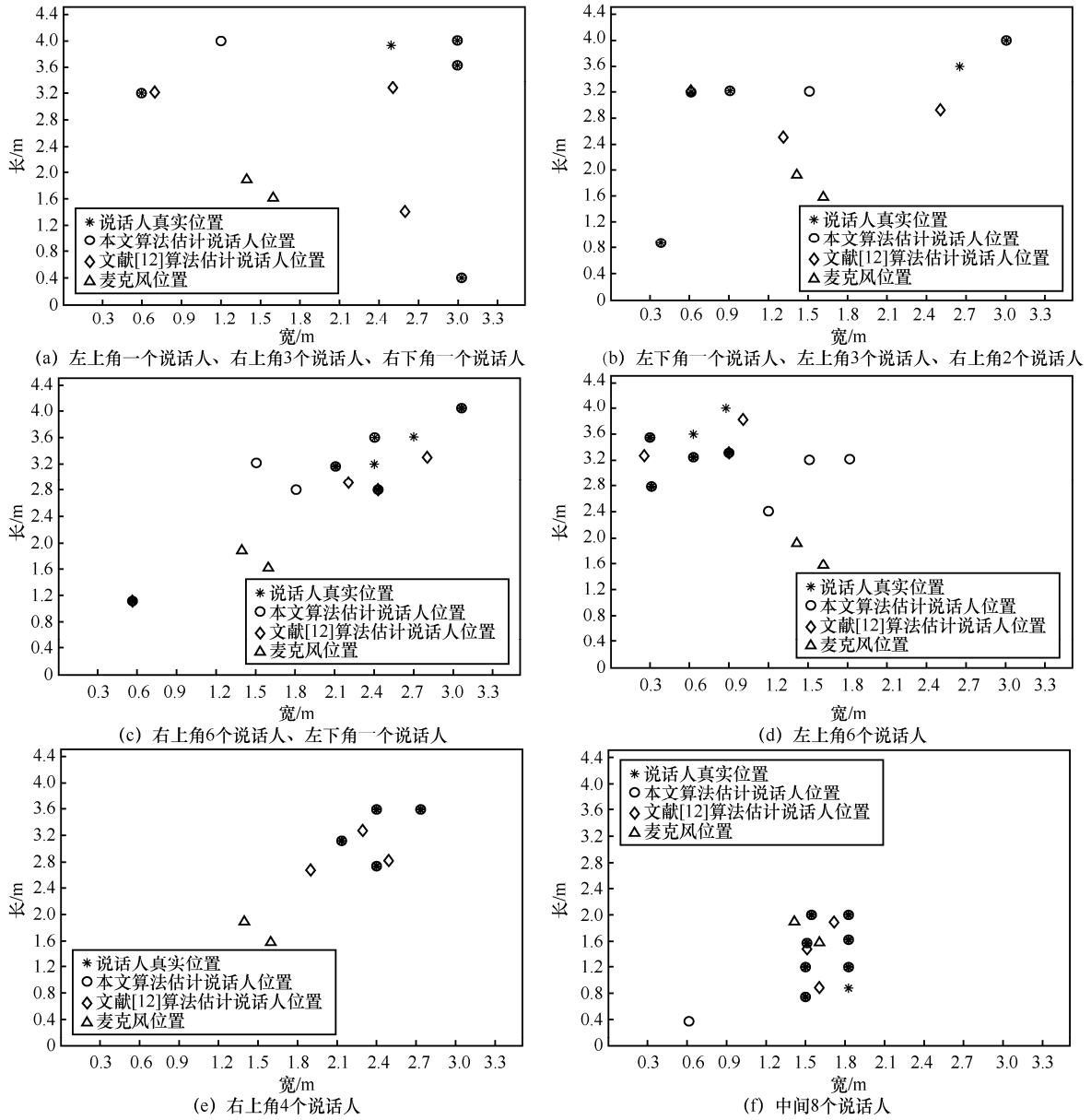


图 2 不同说话人同时位于房间的不同位置

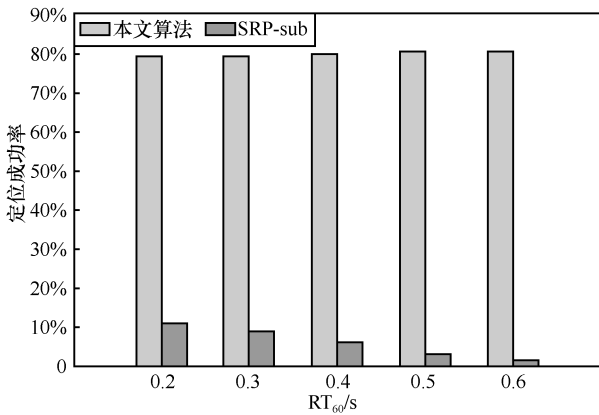


图 3 不同混响时间下的定位成功率

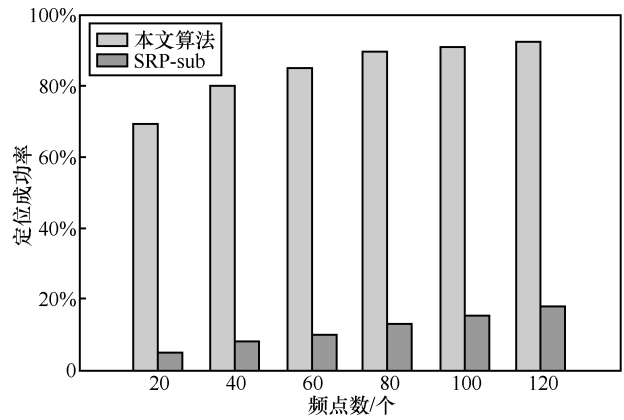


图 4 频点数对定位成功率的影响

4.4 可定位的声源数目增加时的定位情况

参考 4.2 节与 4.3 节的仿真结果, 设置混响时间 $RT_{60}=0.3\text{ s}$, 频点数 $F=60$, 在可定位的声源数目分别为 3、4、5、6、7 时, 分别采用 2 种算法进行定位仿真实验, 其定位成功率统计结果如图 5 所示。

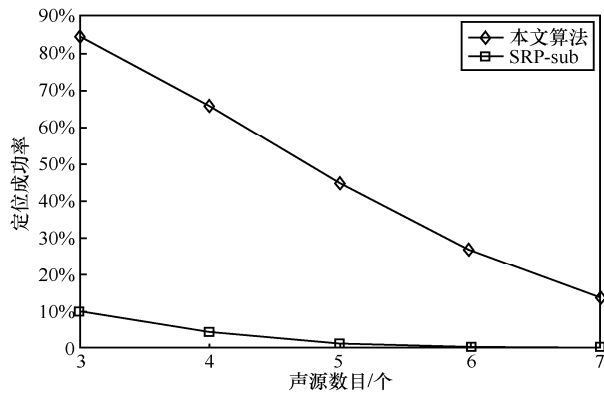


图 5 可定位的声源数目与对应的定位成功率

由图 5 可知, 在以上混响时间和频点数条件下, 使用本文算法定位 3~4 个声源的结果是可以接受的, 而 SRP-sub 算法在同等条件下表现较差。尽管对于更多声源的情况, 本文算法未能取得很好的效果, 但整体性能依然优于 SRP-sub 算法。同时本文的算法很好地说明了混响环境下欠定数量 (麦克风数量小于声源数量) 的麦克风阵列多声源定位的可行性和潜力, 这对体积受限的通信设备来说具有极大的研究价值。

4.5 算法复杂度分析

本文的定位算法中, 假设全房间冲激响应已知, 故不考虑计算压缩观测矩阵的时间, 这样算法的运行时间仅是利用 ReGOMP 算法恢复块稀疏信号的时间与对该信号整合分块的时间之和。图 6 记录了在不同频点数和块稀疏度时, 使用双麦克风定位 1 000 次的平均运行时间。由于本文的块稀疏优化算法的稀疏度为先验已知的, 实验中块稀疏度等于声源数目, 所以在不同块稀疏度时的运行时间可认为是不同声源数目时的运行时间。

由图 6 可知, 频点数取值在 20~80 时, 块稀疏度 (声源数目) 增加, 本文算法的平均运行时间只是略微增加; 频点数取值 100~120 时, 本文算法的平均运行时间随块稀疏度的增加增幅稍大。图 6 也反映了频点数对算法的运行时间的影响比块稀疏度对运行时间的影响更大。但通过 4.3 节的实验也说明了当频点数超过一定数值时, 算法的定位性能提升幅度较小, 所以可酌情选择频点数目, 使得在

保证一定的定位成功率的情况下, 运行时间不至于很大。

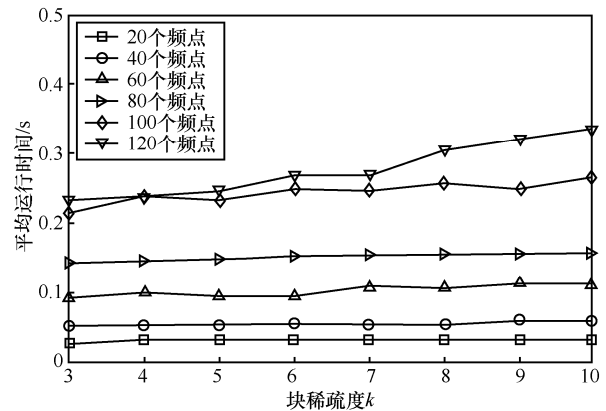


图 6 算法在不同块稀疏度和频点数时的平均运行时间

5 结束语

基于麦克风阵列的声源定位技术是阵列信号处理领域的一项研究热点, 而压缩感知的出现又为该领域的发展提供了一个崭新的视角。本文提出的基于压缩感知的双麦克风阵列混响多声源定位算法以归一化的全房间冲激响应为特征, 构建压缩观测矩阵, 将多声源定位问题转化为块稀疏信号的重构问题, 通过重构信号的块与空间声源的对应关系得到实际声源的位置。

该算法不仅能解决室内混响环境, 如会议室、小型会议场等的定位问题, 且麦克风数量的减少可大大减轻系统的运算负担和硬件成本。本文仅使用 2 个麦克风就能对混响空间中至少 3 个声源进行定位, 其中声源数目为 3~4 个时, 定位效果较好。虽然本文算法对于更多声源的定位情况不是很理想, 但依然摆脱了传统声源定位算法中, 麦克风数量必须大于声源数目的限制。这对于室内混响空间中小型化的便捷语音设备而言很有研究价值。

参考文献:

- [1] 龙学焜. 基于麦克风阵列的近场多声源目标定位与跟踪[D]. 南昌: 南昌大学, 2016.
LONG X K. Multi-target sound source localization and tracking based on microphone array in near-field environment[D]. NanChang: Nan-chang University, 2016.
- [2] 刘超. 基于麦克风阵列的声源定位算法研究[D]. 南京: 南京大学, 2015.
LIU C. Study of sound source localization algorithm based on microphone array[D]. Nanjing: Nanjing University, 2015.
- [3] 石婷. 麦克风阵列声源定位算法研究综述[J]. 科技视界, 2016(27): 274.

- SHI T. Research on the algorithms of sound source localization based on microphone array[J]. Science & Technology Vision, 2016(27): 274-274.
- [4] 赵圣. 基于麦克风阵列的仿人智能机器人声源定位技术的研究[D]. 哈尔滨: 哈尔滨工业大学, 2013.
- ZHAO S. Research on sound source localization technology of humanoid intelligent robot based on microphone array[D]. Harbin: Harbin Institute of Technology, 2013.
- [5] ASAEI A, BOURLARD H, CEVHER V. Model-based compressive sensing for multi-party distant speech recognition[C]//2011 IEEE International Conference on Acoustics, Speech and Signal Processing. 2011:4600-4603.
- [6] WABNITZ A, EPAIN N, SCHAIK A V, et al. Time domain reconstruction of spatial sound fields using compressed sensing[C]//2011 IEEE International Conference on Acoustics, Speech and Signal Processing. 2011: 465-468.
- [7] BENICHOUX A, VINCENT E, GRIBONVAL R. A compressed sensing approach to the simultaneous recording of multiple room impulse responses[C]//2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. 2011: 285-288.
- [8] MIGNOT R, CHARDON G, DAUDET L. Low frequency interpolation of room impulse responses using compressed sensing[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2014, 22(1): 205-216.
- [9] CHARDON G. A block-sparse music algorithm for the localization and the identification of directive sources[C]//2014 IEEE International Conference on Acoustics, Speech and Signal Processing. 2014: 3953-3957.
- [10] 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 X Y, ZHOU L, WU Z Y. Compressed sensing-based sound source localization algorithm for microphone array[J]. Journal of Southeast University(Natural Science Edition), 2015, 45(2):203-207.
- [11] DO H, SILVERMAN H F. A method for locating multiple sources from a frame of a large-aperture microphone array data without tracking[C]//IEEE International Conference on Acoustics, Speech, and Signal Process. 2008: 301-304.
- [12] 倪志莲, 蔡卫平, 张怡典. 基于子带可控响应功率的多声源定位方法[J]. 计算机工程与应用, 2013, 49(24):205-209.
- NI Z L, CAI W P, ZHANG Y D. Method for multiple speech source localization based on sub-band steered response power[J]. Computer Engineering and Applications, 2013, 49(24): 205-209.
- [13] ASAEI A, BOURLARD H, TAGHIZADEH M J, et al. Computational methods for underdetermined convolutive speech localization and separation via model-based sparse component analysis[J]. Speech Communication, 2016, 76(C):201-217.
- [14] 石光明, 刘丹华, 高大化, 等. 压缩感知理论及其研究进展[J]. 电子学报, 2009, 37(5):1070-1081.
- SHI G M, LIU D H, GAO D H, et al. Advances in theory and application of compressed sensing[J]. Chinese Journal of Electronics, 2009, 37(5): 1070-1081.
- [15] ZHAO R, LIN W, LI H, et al. Reconstruction algorithm for compressive sensing based on smoothed ℓ_1 -norm and revised newton method[J]. Journal of Computer-Aided Design & Computer Graphics, 2012, 24(4):478-484.
- [16] 李坤, 马彩文, 李艳, 等. 压缩感知重构算法综述[J]. 红外与激光工程, 2013, 42(s1):225-232.
- LI K, MA C W, LI Y, et al. Survey on reconstruction algorithm based on compressive sensing[J]. Infrared and Laser Engineering, 2013, 42(s1):225-232.
- [17] YAGHOUBI M, WU D, DAVIES M E. Fast non-negative orthogonal matching pursuit[J]. IEEE Signal Processing Letters, 2015, 22(9): 1229-1233.
- [18] ZHOU Y, ZENG F Z, GU Y C. A gradient descent sparse adaptive matching pursuit algorithm based on compressive sensing[C]//2017 IEEE International Conference on Machine Learning and Cybernetics. 2017:2883-2890.
- [19] ZHONG J, WEN G, MA C, et al. Radar signal reconstruction algorithm based on complex block sparse bayesian learning[C]//2015 IEEE International Conference on Signal Processing. 2015:1930-1933.
- [20] HAMARUNNISA P, SHERIKH K K. An efficient structured sparsity model for reverberant speech separation[J]. International Journal of Advanced Trends in Computer Science & Engineering, 2015, 4(10): 14658-14662.
- [21] ASAEI A, GOLBABAEE M, BOURLARD H, et al. Structured Sparsity models for reverberant speech separation[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2014, 22(3): 620-633.
- [22] ALLEN J B, BERKLEY D A. Image method for efficiently simulating small-room acoustics[J]. Journal of the Acoustical Society of America, 1979, 65(S1):943-950.
- [23] ASAEI A, BOURLARD H, TAGHIZADEH M J, et al. Model-based sparse component analysis for reverberant speech localization[C]//2014 IEEE International Conference on Acoustics, Speech and Signal Processing 2014:1439-1443.
- [24] MAJUMDAR A. Greedy algorithms promoting group sparsity[Z]. Math Works, 2009.
- [25] NEEDELL D, VERSHYNIN R. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit[J]. Foundations of Computational Mathematics, 2009, 9(3):317-334.
- [26] ELDAR Y C, KUPPINGER P, BOLCSKEI H. block-sparse signal: uncertainty relations and efficient recovery[J]. IEEE Transactions on Signal Processing, 2010, 58(6):3042-3054.

[作者简介]



张奕(1978-), 男, 山东济南人, 博士, 大连大学副教授, 主要研究方向为语音信号处理、阵列信号处理、进化算法和压缩感知。



李娟(1993-), 女, 甘肃会宁人, 大连大学硕士生, 主要研究方向为阵列信号处理。

张敏(1966-), 女, 辽宁大连人, 博士, 大连大学副教授, 主要研究方向为大数据挖掘、生物信息、智能算法。