

基于 AVS 和稀疏表示的鲁棒语者声源 DOA 估计方法

邹月娴 郭轶凡 郑炜乔

(北京大学信息工程学院 ADSPLAB/ELIP, 深圳, 518055)

摘要: 基于声学矢量传感器 (Acoustic vector sensor, AVS) 和空间声源稀疏表示理论, 进行了鲁棒的高精度语者声源到达角 (Direction of arrival, DOA) 估计方法研究。考虑混响和加性噪声影响, 本文推导了 AVS 接收信号的向量化的协方差矩阵模型, 设计了过完备字典, 依此建立声源的空间稀疏表示模型, 最终通过求解稀疏空间谱获得鲁棒的 DOA 估计。本文进行了大量的不同混响和加性噪声条件下的仿真实验和实际环境中的 DOA 估计实验, 实验结果表明, 本文提出的语者声源 DOA 估计方法在信噪比 5~30 dB 范围内可获得均方根误差 (Root mean square error, RMSE) 小于 1° 的估计精度。在实际环境中也取得了 $2\sim 10^\circ$ 误差的 DOA 估计结果。

关键词: 声学矢量传感器; 语者声源; 到达角估计 (DOA); 空间稀疏表示; 协方差矩阵

中图分类号: TN912.3 **文献标志码:** A

Robust Speaker DOA Estimation Based on AVS and Sparse Representation

Zou Yuexian, Guo Yifan, Zheng Weiqiao

(ADSPLAB/ELIP, School of Electronic and Computer Engineering, Peking University, Shenzhen, 518055, China)

Abstract: A robust high resolution speaker source direction of arrival (DOA) estimation method is proposed based on one acoustic vector sensor (AVS) and spatial sparse representation. Under the reverberation and additive noise conditions, the array covariance vector model of the received signals by AVS is first derived. Then the sparse representation model of the covariance vector is developed. Finally the robust DOA estimation is obtained by recovering the sparse vector. A large number of simulation experiments are carried out under different reverberation and additive noise conditions, and also DOA estimation experiments in the actual environment. The results show that the proposed speaker DOA estimation is able to achieve root mean square error (RMSE) of below 1° when SNR is from 5 dB to 30 dB and $2\sim 10^\circ$ error in the real scenario.

Key words: acoustic vector sensor; speaker source; direction of arrival estimation (DOA); spatial sparse representation; covariance matrix

引 言

随着信息技术的快速发展, 智慧城市与智能家居理念正改变着我们的生活, 其中, 智能服务机器人

将在未来智能家居中得到广泛应用,具有很大的应用前景和价值。视觉技术和听觉技术是智能服务机器人走向实际应用的两大关键技术,其技术发展水平备受学术界和工业界关注。语者声源到达角(Direction of arrival, DOA)估计属于听觉感知技术之一,其目标是使服务机器人在实际噪声环境下快速、鲁棒和准确地定位出空间语者声源。对于 DOA 估计技术在智能服务机器人上的应用,不仅需要考虑到 DOA 估计精度、分辨率和鲁棒性等技术参数指标,也需要考虑系统的硬件开销、体积大小等。因此,基于尺寸小和低成本的小麦克风阵列的鲁棒 DOA 估计技术是服务机器人听觉感知系统的发展方向。基于麦克风线性阵列(Unlinear array, ULA)、圆形阵列和球形阵列的 DOA 估计已有众多研究成果,但都存在阵元数多和阵列尺寸较大的特点,较难满足于智能服务机器人听觉系统的需求。研究表明,著名的基于 ULA 的 DOA 估计方法,如多重信号分类(Multiple signal classification, MUSIC)算法^[1]和旋转不变子空间类算法^[2]等,要求较多的阵元数、较多的快拍数据和较高的信噪比环境。仿真人耳的双麦克风 DOA 估计技术也同样有众多成果,如经典的基于广义互相关函数法^[3]的 DOA 估计技术,其算法简单,但易受噪声的影响,且当声源接近双麦克风连线方向时,DOA 估计精度变差;在有遮挡的情况下,两个麦克风间接收到的信号幅度差不再是线性关系,DOA 估计精度下降。此外,基于双耳声级差、双耳相位差和双耳时间差的算法^[4-5],模仿人耳对声源定位的机理,可同时对低频和低频信号实现较好的定位效果,但定位范围仅局限于半平面,这也在一定程度上限制了 DOA 估计技术的应用。

本文面向智能服务机器人语者声源定位需求,采用一种新型的同位传感器阵列——声学矢量传感器(Acoustic vector sensor, AVS)^[6]实现鲁棒的高精度语者声源 DOA 估计。与传统的麦克风阵列相比,AVS 的尺寸非常小(仅 $1\sim 2\text{ cm}^3$),且所含阵元数少,大大降低硬件开销,减少成本,AVS 具有二维指向性,同时其近似空间同位、各通道间无时间差且存在三角函数关系等特殊优点,单个 AVS 就能估计出语者声源的三维 DOA 信息,AVS 的自身特性使其成为实现面向服务机器人的语者声源 DOA 估计以及空间干扰噪声抑制的研究热点^[7-11]。

基于 AVS 的独特优点,本实验室团队已开展了相关研究工作。在空间稀疏表示框架下,充分利用 AVS 阵列接收数据模型与其子阵列流形矩阵的关系,提出了一种采用较低复杂度获得较高 DOA 估计精度的算法^[9]。但该方法采用了 8 个间隔半波长的 AVS 阵列,体积较大,限制了其在服务机器人上的应用。为了克服该缺点,作者提出了一种基于单个 AVS 和语音的时频稀疏特性的 DOA 估计新算法^[10]。该方法推导了时频域 AVS 的不同传感器之间的数值比(Inter-sensor data ratio, ISDR)与声源 DOA 之间的一一对应关系,通过对 ISDR 值进行聚类及聚类中心,获得 DOA 的估计。然而,实际环境下,计算出的 ISDR 值会受到噪声和混响干扰,因此,DOA 估计算法的噪声鲁棒性尚不能满足实际需求。

本文借鉴文献[12]提出的基于阵列接收信号的互相关矩阵稀疏表示模型实现 DOA 估计的思想,推导单 AVS 接收信号的向量化协方差矩阵稀疏表示模型,将 DOA 估计问题转换为稀疏向量求解问题。通过大量的实验,验证了本文所提方法的噪声鲁棒性和 DOA 估计有效性。

1 AVS 数据模型

基于单个 AVS 的声源定位系统设置如图 1 所示,其中 AVS 位于直角坐标系原点。单个 AVS 是由 1 个全向传感器及 3 个正交指向的指向性传感器组成,全向传感器被称为 o 分量,指向 x, y 及 z 轴的指向性传感器分别称作 u 分量、 v 分量及 w 分量^[6]。本文假设空间存在单个目标语者声源 $s(t)$ (服务机器人应用场景),该目标声源的 DOA 表示为 (θ_s, ϕ_s) 。因此,AVS 对目标声源的流形矢量为

$$\mathbf{a}(\theta_s, \phi_s) = [u_s, v_s, w_s, 1]^T \quad \mathbf{a} \in \mathbf{R}^{4 \times 1} \quad (1)$$

式中

$$u_s = \sin\theta_s \cos\phi_s, v_s = \sin\theta_s \sin\phi_s, w_s = \cos\theta_s \quad (2)$$

被称为声源的 x, y 及 z 方向余弦,且设定俯仰角 $\theta_s \in (0^\circ, 180^\circ)$,方位角 $\phi_s \in [0^\circ, 360^\circ)$ 。考虑混响和背

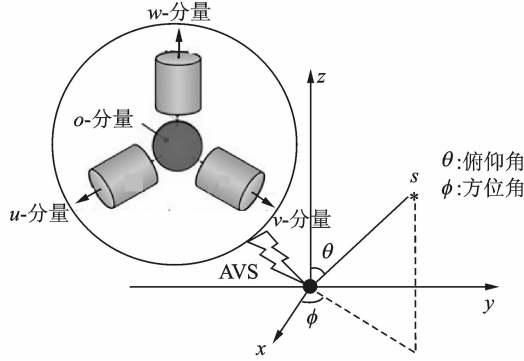


图1 声学矢量传感器结构图
Fig.1 AVS's structure flow

景噪声, AVS 在时刻 t 接收到的数据可表示为

$$\mathbf{x}(t) = \mathbf{a}(\theta_s, \phi_s) s(t) * h(t) + \mathbf{n}(t) \quad (3)$$

式中: $\mathbf{x}(t) = [x_u(t), x_v(t), x_w(t), x_o(t)]^T$, 分别为 u, v, w, o 传感器的输出, $*$ 为线性卷积操作, $h(t)$ 代表房间的冲击响应, $\mathbf{n}(t) = [n_u(t), n_v(t), n_w(t), n_o(t)]^T$ 为 u, v, w, o 传感器上的加性噪声, 假设它们互不相关, 且与目标语音信号不相关。

2 DOA 估计算法

2.1 协方差矩阵稀疏表示模型

声源在空域具有稀疏性, 因此, 首先将整个空间的方位角和俯仰角等间隔划分为 N_1 和 N_2 个空间网格, 形成 $M = N_1 \times N_2$ 个空间网格角度集, 可表示为 $\Theta = \{(\theta_i, \phi_j) | i=1, \dots, N_1; j=1, \dots, N_2\}$, 且 $N_1 \gg 1, N_2 \gg 1$ 。如果空间声源位于网格 (θ_i, ϕ_j) , 则其流形矢量为 $\mathbf{a}(\theta_i, \phi_j)$, 而相应的与空间网格角度集 Θ 对应的流形矩阵如下

$$\mathbf{A} = [\mathbf{a}(\theta_1, \phi_1), \dots, \mathbf{a}(\theta_i, \phi_j), \dots, \mathbf{a}(\theta_{N_1}, \phi_{N_2})] \quad \mathbf{A} \in \mathbf{R}^{L \times M} \quad (4)$$

相对于单个目标声源的 DOA 估计, 式(4)被称为过完备字典(或稀疏基矩阵)。显然, 对于单声源 $s(t)$, 由式(3,4), AVS 接收数据 $\mathbf{x}(t)$ 的等效稀疏表示模型为

$$\mathbf{x}(t) = \mathbf{A}\mathbf{z}(t) + \mathbf{n}(t) \quad (5)$$

式中: $\mathbf{z}(t) = [z_1(t), \dots, z_M(t)]^T$ 为 M 维稀疏矢量, 即当声源来自方向 (θ_i, ϕ_j) 时, 只有与之对应的 $\mathbf{z}(t)$ 中的第 idx 个元素为非零值, 且 $z_{idx}(t) = s(t) * h(t)$ 。其中

$$idx = (i-1) \times N_2 + j \quad (6)$$

根据式(5), 可以将 DOA 求解问题转换为稀疏矢量 $\mathbf{z}(t)$ 的求解问题。文献[13]指出阵列协方差矩阵在稀疏基矩阵表示下存在稀疏解。

根据式(5), 推导 $\mathbf{x}(t)$ 的协方差矩阵如下

$$\begin{aligned} \mathbf{R} = E(\mathbf{x}(t)\mathbf{x}^H(t)) &= \mathbf{A}[E(\mathbf{z}(t)\mathbf{z}^H(t))]\mathbf{A}^H + [E(\mathbf{n}(t)\mathbf{n}^H(t) + \mathbf{A}\mathbf{z}(t)\mathbf{n}^H(t) + \mathbf{n}(t)\mathbf{z}^H(t)\mathbf{A}^H)] = \\ &= \mathbf{A}\mathbf{R}_z\mathbf{A}^H + [\mathbf{R}_n + \mathbf{A}\mathbf{z}(t)\mathbf{n}^H(t) + \mathbf{n}(t)\mathbf{z}^H(t)\mathbf{A}^H] \end{aligned} \quad (7)$$

式中: $(\cdot)^H$ 为向量或矩阵的共轭转置; $\mathbf{R}_z = E(\mathbf{z}(t)\mathbf{z}^H(t))$ 为稀疏矢量的自相关矩阵。 \mathbf{R}_z 中第 idx 个对角元素 $\mathbf{R}_z(idx, idx)$ 为来自 (θ_i, ϕ_j) 方向的声源信号功率, 定义为 p_{idx} 。实际操作中, 需要基于 T 个快拍数据估计 \mathbf{R} , 式(7)中的 $\mathbf{x}(t)$ 的协方差矩阵估计如下

$$\begin{aligned} \bar{\mathbf{R}} &= \frac{1}{T} \mathbf{A} [\mathbf{z}(t) \mathbf{z}^H(t)] \mathbf{A}^H + \\ &\frac{1}{T} [\mathbf{n}(t) \mathbf{n}^H(t) + \mathbf{A} \mathbf{z}(t) \mathbf{n}^H(t) + \mathbf{n}(t) \mathbf{z}^H(t) \mathbf{A}^H] \end{aligned} \quad (8)$$

因为 $\mathbf{n}(t)$ 是加性零均值高斯白噪声, 且根据假设信号和噪声不相关, 当 $T \rightarrow \infty$, 式(8)的右半部分近似于 $\sigma^2 \mathbf{I}$, \mathbf{I} 是单位矩阵, 因此式(8)可以化简如下

$$\bar{\mathbf{R}} = \mathbf{A} \bar{\mathbf{R}}_z \mathbf{A}^H + \sigma^2 \mathbf{I} \quad (9)$$

将式(9)展开, 可改写为

$$\bar{\mathbf{R}} = \sum_{idx=1}^M p_{idx} \mathbf{a}(\theta_i, \phi_j) \mathbf{a}^H(\theta_i, \phi_j) + \sigma^2 \mathbf{I} \quad (10)$$

定义空间谱 \mathbf{p} 为

$$\mathbf{p} = \text{diag}(\mathbf{R}_z) = [p_1, p_2, \dots, p_M]^T \quad (11)$$

定义 $\mathbf{E}_v = \text{vec}(\sigma^2 \mathbf{I})$, $\mathbf{v}_v(\theta_i, \phi_j) = \text{vec}(\mathbf{a}(\theta_i, \phi_j) \mathbf{a}^H(\theta_i, \phi_j))$, 其中 $\text{vec}(\cdot)$ 操作为矩阵的向量化操作, 依据式(4)定义新的过完备字典 Ψ 为

$$\Psi = [\mathbf{v}_v(\theta_1, \phi_1), \dots, \mathbf{v}_v(\theta_i, \phi_j), \dots, \mathbf{v}_v(\theta_{N_1}, \phi_{N_1})] \quad (12)$$

对式(9)两边进行向量化, 代入式(10,11)可得

$$\mathbf{R}_v = \Psi \mathbf{p} + \mathbf{E}_v \quad (13)$$

式中: $\mathbf{R}_v = \text{vec}(\mathbf{R})$ 。

如前所述, M 远远大于声源个数, 因此空间谱 \mathbf{p} 是稀疏矢量, 称为稀疏空间谱, 其非零行所在位置则对应语者声源的 DOA (θ_i, ϕ_i) 。因此, 本文将式(13)称为接收信号协方差矩阵的 DOA 估计稀疏表示模型。 Ψ 具有非平凡的零空间, 可以由 \mathbf{p} 的候选稀疏进行重构 \mathbf{R}_v 。为了区别于其他算法, 本文称其为基于 AVS 的协方差稀疏空间谱(AVS-CSS)模型。

2.2 DOA 估计方法

基于语者 DOA 估计稀疏表示模型, DOA 估计问题被转换为求解稀疏空间谱 \mathbf{p} 的问题, 对式(13)求稀疏解, 获得稀疏矢量 \mathbf{p} , 寻找其非零位置, 即可以估计出声源所在 DOA (θ_i, ϕ_i) 。

式(13)中的稀疏空间谱 \mathbf{p} 可以通过在稀疏约束条件下最小化重构误差优化问题求解, 具体表示为

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\text{argmin}} \|\mathbf{R}_v - \Psi \mathbf{p}\|_2^2 + \lambda \|\mathbf{p}\|_1 \quad (14)$$

式中: l_2 模项控制稀疏重构的精确性, 而 l_1 模项则用来控制 \mathbf{p} 的行稀疏性, λ 为根据经验值设定的正则化参数, 是在稀疏性和重构误差之间进行权衡折衷的参数。考虑到计算效率和对噪声的鲁棒性, 本文选用 l_1 -SVD^[14] 方法, 利用奇异值分解技术获取信号子空间, 以降低重构问题的维度, 降低后续重构的运算复杂度。对 \mathbf{p} 的重构, 本文选用斯坦福大学推出的 CVX 工具包^[15]。

用 CVX 工具包稀疏求解 \mathbf{p} , 假定用 i_p 来表示稀疏空间谱 \mathbf{p} 的峰值, 即重构的稀疏矢量 \mathbf{p} 的主导非零行, 则根据式(6)可以由 i_p 索引到声源 DOA 所在的网格 (i, j) , 得到 DOA 估计的结果为

$$\hat{\theta}_s = \theta_i \quad \hat{\phi}_s = \phi_j \quad \theta_i, \phi_j \in \Theta \quad (15)$$

本文提出的 AVS-CSS DOA 估计算法流程总结如下:

- (1) 利用 $x_u(t)$, $x_v(t)$, $x_w(t)$ 和 $x_o(t)$ 计算协方差矩阵向量化后的 \mathbf{R}_v ;
- (2) 构造式(12)中的过完备字典 Ψ ;
- (3) 利用 l_1 -SVD 方法稀疏求解式(13), 获得稀疏空间谱 \mathbf{p} ;
- (4) 由稀疏空间谱 \mathbf{p} 寻找非零行的序号 i_p , 利用式(6)计算 i_p 对应的语者声源所在网格 (i, j) , 得出声源 DOA 信息。

从上述算法可见, N_1 和 N_2 的取值会影响 DOA 估计精度。 N_1 或 N_2 越大, 网格间距越小, 语者声

源 DOA 与 Θ 中预定义的角度就越接近,精度越高,但相应的计算复杂度会增加。如果通过减小 N_1 和 N_2 来降低计算复杂度,则空间信源不落在候选网格上的概率会增加,导致 DOA 估计精度下降。为了折衷算法的复杂度和 DOA 估计精度,文献[14]提出一种网格细化的策略,首先对空间进行较粗糙的划分并执行方法, N_1 和 N_2 的取值较小,以较低的复杂度实现初步的 DOA 估计,随后仅仅对选定的 DOA 区域进行再次精定位。

3 实验与分析

为了评估本文提出的 AVS-CSS DOA 估计算法的性能,本文进行了 3 个仿真实验和 1 个实际环境中的 DOA 估计实验。选择 GMDA-Laplace 算法^[16]进行性能比较。仿真实验的语者声源为男性,长度为 3 s,采样率为 16 kHz。对本文提出的 AVS-CSS 算法,相关参数设置为 $\theta_s \in (0^\circ, 180^\circ)$, $\phi_s \in [0^\circ, 180^\circ]$, $N_1 = N_2 = 180$,正则化参数 $\lambda = 20$, $T = 200$ (200 个快拍数据)。对 GMDA-Laplace 算法,参数设置为:语音信号分帧长度为 30,20 ms 重叠,加窗函数采用汉明窗,窗长为 30 ms,STFT 长度为 512 点;两个麦克风沿着 z 轴放置,间隔 8 cm。需要注意的是,GMDA-Laplace 算法每次只能估计声源的方位角或俯仰角,如果需要同时估计方位角和俯仰角,则需分别进行 2 次计算。选用绝对值误差(Absolute error, AER)和均方根误差(Root mean square error, RMSE)作为 DOA 估计性能的标准,定义如下

$$AER = \frac{(|\hat{\theta} - \theta| + |\hat{\phi} - \phi|)}{2} \tag{16}$$

$$RMSE = 0.5 \sqrt{\frac{\sum_{i=1}^{N_T} ((\hat{\theta}_i - \theta)^2 + (\hat{\phi}_i - \phi)^2)}{N_T}} \tag{17}$$

式中: N_T 为独立试验的次数。

3.1 算法的 DOA 估计精度

该实验旨在评估本文提出的 AVS-CSS 算法在单个语者位于不同方位角的 DOA 估计精度。信噪比 SNR 为 20 dB,不考虑混响, $\theta = 60^\circ$, ϕ 在每次试验中从 $0 \sim 180^\circ$ 随机生成,共进行 100 次,实验结果如图 2 所示。从实验结果得出结论:在该实验条件下,两个算法的 DOA 估计绝对值误差大部分低于 0.5° ; 本文提出算法在 ϕ 为 $0 \sim 20^\circ$ 和 $160 \sim 180^\circ$ 的范围内,性能优于对比算法。

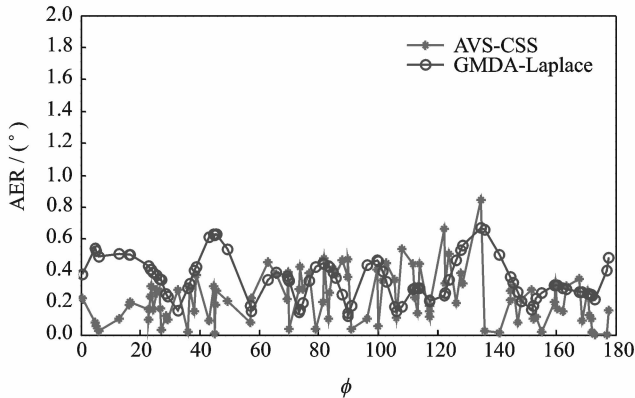


图 2 不同声源 DOA 处的绝对值误差 AER

Fig. 2 Absolute error versus different source DOA

3.2 算法的噪声鲁棒性

该实验的目的是评估 AVS-CSS 算法在无混响条件下对加性噪声的鲁棒性。实验条件为:目标语者声源位于 $(60^\circ, 45^\circ)$, 信噪比 SNR 从 5 dB 变化到 30 dB, 每个 SNR 进行 100 次独立试验, RMSE 实验结果如图 3 所示。由图 3 可知, 本文提出的算法的 RMSE 性能在所有 SNR 条件下都要小于 GMDA-Laplace 算法。当 SNR 为 5 dB 时, AVS-CSS 算法的 RMSE 约为 1.1° , 而对比算法的 RMSE 为 2.2° , 当 $\text{SNR} \geq 10$ dB 时, AVS-CSS 算法的 RMSE 均小于 0.5° 。此结果表明, 本文提出的 AVS-CSS 算法对加性噪声不敏感。

3.3 算法的混响鲁棒性

本实验评估 AVS-CSS 算法在不同混响条件下的 DOA 估计性能。实验设置为:采用文献[17]所述的 Image 方法产生混响信号, 在 $10 \text{ m} \times 5 \text{ m} \times 4 \text{ m}$ 的房间中, AVS 位于房间正中, 语者声源在距 AVS 为 2 m 的 $(60^\circ, 45^\circ)$ 方向, SNR 设为 10 dB。采用 5 个不同的混响时间 (RT_{60}) 设置, 如表 1 所示。在每个混响条件下进行 100 次独立试验后平均获得 RMSE 性能曲线, 如图 4 所示。从图中清楚地看到 AVS-CSS 算法的 RMSE 曲线基本不随混响时间而变化, 即该算法对混响具有很好的鲁棒性, 这对于实际 DOA 估计应用有价值。而 GMDA-Laplace 算法的性能随着混响时间的增加急剧下降, 未显示出对混响的鲁棒性。

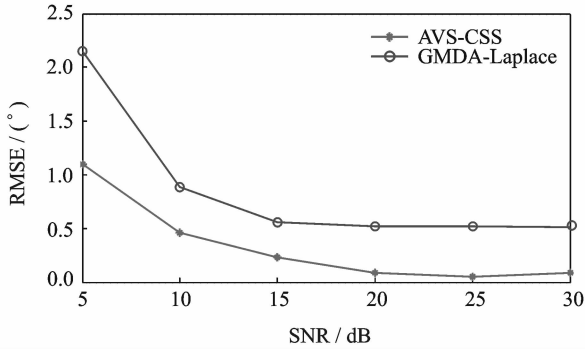


图 3 不同信噪比 SNR 下的 DOA 估计 RMSE

Fig. 3 RMSE versus different signal-to-noise ratio levels

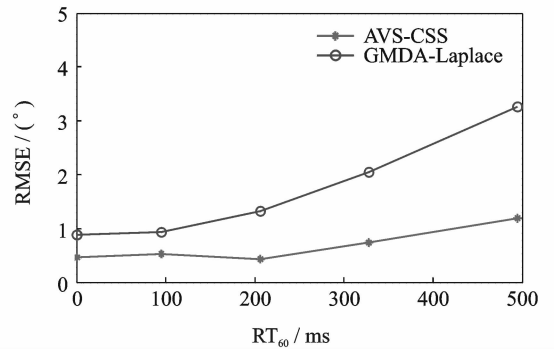


图 4 不同混响条件下的 DOA 估计 RMSE

Fig. 4 RMSE versus different reverberation levels

表 1 RT_{60} 和相应的反射系数

Table 1 RT_{60} and corresponding reflection coefficients

RT_{60}/ms	反射系数		
	墙壁	房顶	地板
0	0.0	0.0	0.0
95	0.3	0.2	0.1
206	0.6	0.5	0.3
328	0.8	0.6	0.4
494	0.9	0.7	0.7

3.4 实际环境中的 DOA 估计

通过仿真实验可见,本文提出的 DOA 估计算法具有对加性噪声和混响的鲁棒性。为了进一步验证该算法在实际 DOA 估计应用中的性能,本文开展了实际环境 DOA 估计实验。本文采用的是本实验室自制的 AVS 录制数据,AVS 安装在智能小车上方,DOA 定位结果可通过实测定位界面可视化。实验空间约 8.5 m×3 m×5 m 的房间,非静音、非抗混响环境(自然实验室环境),同时存在背景噪声和混响。SNR 测量值约为 20 dB,数据采集系统采样率为 44.1 kHz。声源 DOA: $\theta_s=90^\circ$,方位角 ϕ_s 分别设置为 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 和 180° 。目标语者和 AVS 的距离分别设置为 0.5, 1.0 和 1.5 m(如图 5)。DOA 估计结果见表 2。由表 2 可以看出,在相同响度的语者声源条件下,本算法 AVS-CSS 并不会因为声源与 AVS 的距离增加而使 DOA 估计结果变差。最好的 DOA 估计误差只有 3° 。这些初步的实验结果进一步验证了本文所提出算法的有效性。需要说明的是,由于本实验设置依靠人工进行角度划分,以及 AVS 硬件前端的元件失配会产生系统误差,因此对比图 3 和表 2 可见,实测平均 DOA 误差大于仿真平均 DOA 误差,其误差范围在 $2\sim 10^\circ$,从一定程度上验证了算法的有效性和鲁棒性。

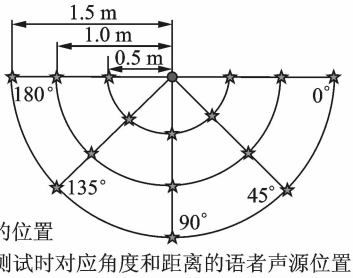


图 5 实际环境实验设置

Fig. 5 Experimental setup in real scenario

表 2 实际环境下的 DOA 估计结果

Table 2 DOA estimation results in real scenario ($^\circ$)

(θ_s, ϕ_s)	距离/m		
	0.5	1.0	1.5
(90,0)	(96,4)	(96,2)	(95,2)
(90,45)	(97,44)	(98,44)	(95,42)
(90,90)	(94,93)	(94,92)	(94,91)
(90,135)	(89,141)	(89,143)	(88,144)
(90,180)	(87,180)	(86,180)	(94,176)

4 结束语

本文面向智能服务机器人语者声源定位需求,采用了单个声学矢量传感器 AVS 获取空间声源信息,AVS 具有尺寸小和空间同位的三维结构,相比传统麦克风阵列更适合用于服务机器人听觉系统。本文推导出基于单个 AVS 接收数据的协方差矩阵稀疏表示模型,并通过求解稀疏解获得鲁棒的语者声源 DOA 估计(称为 AVS-CSS 算法)。仿真实验结果表明,本文提出的 AVS-CSS 算法能同时估计俯仰角和方位角,具有高精度、对噪声和混响不敏感的优点。实际测试结果表明,AVS-CSS 算法在实际噪声和混响环境下也能获得 $2\sim 10^\circ$ 的 DOA 估计精度,且对语者与 AVS 的距离不敏感。上述结果表明,AVS-CSS 算法有望为解决服务机器人声源 DOA 估计提供解决方案。研发系统误差小的 AVS 硬件系统是下一步的研究重点,期望获得与仿真结果一致的实际测试结果。

参考文献:

[1] Schmidt R O. Multiple emitter location and signal parameter estimation [J]. Antennas and Propagation, IEEE Transactions on, 1986, 34(3): 276-280.

[2] Roy R, Kailath T. ESPRIT-estimation of signal parameters via rotational invariance techniques [J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1989, 37(7): 984-995.

[3] Brandstein M S, Adcock J E, Silverman H F. A practical time-delay estimator for localizing speech sources with a microphone array [J]. Computer Speech & Language, 1995, 9(2): 153-169.

[4] Handzel A A, Krishnaprasad P S. Biomimetic sound-source localization [J]. Sensors Journal, IEEE, 2002, 2(6): 607-616.

[5] Raspaud M, Viste H, Evangelista G. Binaural source localization by joint estimation of ILD and ITD [J]. Audio, Speech,

and Language Processing, *IEEE Transactions on*, 2010, 18(1): 68-77.

- [6] Hawkes M, Nehorai A. Acoustic vector-sensor beamforming and Capon direction estimation [J]. *Signal Processing, IEEE Transactions on*, 1998, 46(9): 2291-2304.
- [7] Lockwood M E, Jones D L. Beamformer performance with acoustic vector sensors in air [J]. *The Journal of the Acoustical Society of America*, 2006, 119(1): 608-619.
- [8] Shujau M, Ritz C H, Burnett I S. Designing acoustic vector sensors for localization of sound sources in air [C] //17th European Signal Processing Conference (EUSIPCO 2009). Scotland: EURASIP, 2009:849-853.
- [9] Li B, Zou Y X. Improved DOA estimation with acoustic vector sensor arrays using spatial sparsity and subarray manifold [C] //Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. Kyoto, Japan: IEEE, 2012: 2557-2560.
- [10] Zou Y X, Shi W, Li B, et al. Multisource DOA estimation based on time-frequency sparsity and joint inter-sensor data ratio with single acoustic vector sensor [C] //Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on. Vancouver, Canada:IEEE, 2013:4011-4015.
- [11] Zou Y X, Guo Y F, Zheng W Q, et al. An effective DOA estimation by exploring the spatial sparse representation of the inter-sensor data ratio model [C] //Signal and Information Processing (China SIP), 2014 IEEE China Summit & International Conference on. Xi'an, China:IEEE, 2014: 42-46.
- [12] Zheng J, Kaveh M. Data-adaptive regularization for DOA estimation using sparse spectrum fitting [C] //Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. Vancouver, Canada:IEEE, 2013:3957-3961.
- [13] Yin J, Chen T. Direction-of-arrival estimation using a sparse representation of array covariance vectors[J]. *Signal Processing, IEEE Transactions on*, 2011, 59(8): 4489-4493.
- [14] Malioutov D, Cetin M, Willsky A. A sparse signal reconstruction perspective for source localization with sensor arrays [J]. *Signal Processing, IEEE Transactions on*, 2005, 53(8): 3010-3022.
- [15] Grant M, Boyd S. CVX: Matlab software for disciplined convex programming [EB/OL]. <http://cvxr.com/>, 26th Jan 2015.
- [16] Zhang W, Rao B D. A two microphone-based approach for source localization of multiple speech sources [J]. *Audio, Speech, and Language Processing, IEEE Transactions on*, 2010, 18(8):1913-1928.
- [17] Allen J B, Berkley D A. Image method for efficiently simulating small room acoustics [J]. *The Journal of the Acoustical Society of America*, 1979, 65(4):943-950.

作者简介: 邹月娴(1964-), 女, 教授, 研究方向: 多媒体信息处理技术、模式识别和机器学习, E-mail: zouyx@pkusz.edu.cn; 郭轶凡(1989-), 女, 硕士研究生, 研究方向: 多媒体信息处理技术; 郑炜乔(1991-), 男, 硕士研究生, 研究方向: 多媒体信息处理技术。

