第 38 卷 第 6 期
2019 年 11 月

应用声学
Journal of Applied Acoustics

Vol. 38, No. 6
November, 2019

◇ 研究报告 ◇

基于稀疏表示和特征加权的离格双耳声源定位*

丁建策[1,2] 厉 剑[1,2] 郑成诗[1,2†] 李晓东[1,2]

(1 中国科学院声学研究所噪声与振动重点实验室 北京 100190)

(2 中国科学院大学 北京 100049)

摘要 基于头相关传递函数数据库的传统双耳声源定位方法的定位角度往往被限定在头相关传递函数数据库的离散测量点上。当头相关传递函数数据库的测量方位角间隔较大时，这类算法的性能会显著下降，这就是典型的离格问题。该文提出了基于加权宽带稀疏贝叶斯学习的离格双耳声源定位算法。首先该算法建立离格双耳信号的稀疏表示模型，然后利用双耳相干与扩散能量比特征对各个频点进行加权以降低噪声和混响的影响，最后通过加权宽带稀疏贝叶斯学习方法估计离格声源的方位角。实验结果表明，该算法在各种复杂的声学环境下都有着较高的定位精度和鲁棒性，特别是提高了离格条件下的声源定位性能。

关键词 离格双耳声源定位，稀疏表示，双耳相干与扩散能量比，宽带稀疏贝叶斯学习

中图法分类号: TN9123 **文献标识码:** A **文章编号:** 1000-310X(2019)06-0917-09

DOI: 10.11684/j.issn.1000-310X.2019.06.002

Off-grid binaural sound source localization using sparse representation and feature weighting

DING Jiance[1,2] LI Jian[1,2] ZHENG Chengshi[1,2] LI Xiaodong[1,2]

(1 *Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing* 100190, *China*)

(2 *University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract Traditional binaural sound source localization (BSSL) techniques using measured head-related transfer function (HRTF) databases often suffer a typical off-grid problem, where their estimated azimuth angles are restricted at the measured azimuth angles of HRTF databases. When the interval of the measured azimuth angles is large, the performance of these techniques will degrade significantly. This paper proposes an off-grid BSSL algorithm based on weighted wideband sparse Bayesian learning. First, the algorithm establishes an off-grid sparse representation model. Then weighted values based on binaural coherent-to-diffuse power ratio (BCDR) for all frequency bins are calculated to reduce the impact of noise and reverberation. Finally, a weighted wideband sparse Bayesian learning algorithm is derived to solve the off-grid BSSL problem. Experimental results demonstrate that the proposed method can achieve higher localization accuracy and is more robust than the compared HRTF-based BSSL techniques in various acoustic environments, especially under the off-grid situations.

Key words Off-grid binaural sound source localization, Sparse representation, Coherent-to-diffuse power ratio, Wideband sparse Bayesian learning

2019-03-11 收稿; 2019-06-24 定稿

*国家自然科学基金项目 (61571435, 61801468)

作者简介: 丁建策 (1990-), 男, 河南南阳人, 博士研究生, 研究方向: 信号与信息处理。

†通讯作者 E-mail: cszheng@mail.ioa.ac.cn

0 引言

双耳声源定位利用耳道入口或者耳道内的传声器接收到的声信号来估计空间中声源的方位。它在虚拟声重放[1]、助听器[2]、智能音视频会议[3]等领域有着广泛的应用，研究双耳声源定位有着重要的科学意义和研究价值。

双耳声源定位算法中最常用的两种双耳特征分别为双耳时间差(Interaural time difference, ITD)和双耳声级差(Interaural level difference, ILD)。一般而言，ITD适用于中低频的声源定位，ILD适用于高频的声源定位。在Jeffress[4]提出双耳“巧合假说”模型(coincident theory)之后，研究者们提出了一系列双耳声源定位算法。常用的双耳声源定位算法有两类：一类是基于头相关传递函数(Head-related transfer function, HRTF)的双耳声源定位方法[5−6]，另一类是基于机器学习的监督式双耳声源定位方法[7−8]。基于HRTF的双耳声源定位方法的一般做法是：提取观测双耳信号的双耳特征(如ITD、ILD等)和HRTF数据库中各个离散测量方位角对应的双耳特征，然后进行匹配定位。这类方法计算量小，适用范围广，然而在低信噪比或强混响环境下其定位性能会严重下降。监督式双耳声源定位方法通过机器学习方法训练声源方位角与双耳特征之间的关系，通常有着较高的定位准确率。这类算法需要预先构建训练数据库，训练过程计算复杂度高，而且在训练条件与测试条件不匹配的情况下定位性能会严重下降。本文重点研究基于HRTF的双耳声源定位方法。

在基于HRTF的双耳声源定位方法中，声源的方位角估计结果往往被限定在HRTF数据库的离散测量点上。当声源真实方位角与HRTF数据库的测量方位角不一致时，算法的定位性能会显著下降，这就是双耳声源定位中的离格问题。HRTF数据库的测量方位角间隔一般比较大(不小于5°)，因此离格问题对基于HRTF的双耳声源定位算法的影响不可忽视。随着压缩感知技术的兴起，研究者们提出了一系列离格稀疏重建方法来解决阵列到达角(Direction of arrival, DOA)估计中的离格问题。2013年，Yang等[9]提出了稀疏贝叶斯学习(Sparse Bayesian learning, SBL)方法来解决窄带信号DOA估计中的离格问题。2017年，高阳等[10]提出了基于酉变换的实数域稀疏贝叶斯学习方法，有效降低了离格DOA估计方法的运算量。同年，胡顺仁等[11]提出了一种联合稀疏贝叶斯理论和子空间方法的近场声源定位算法，用于解决近场信号源的DOA估计问题。由于头和躯干的阴影效应的影响，双耳信号与阵列信号的声传播模型有所不同，因此上述这些离格阵列DOA估计算法都不能直接用于解决双耳声源定位中的离格问题。

为了解决双耳声源定位中的离格问题，本文提出一种基于加权宽带稀疏贝叶斯学习的离格双耳声源定位方法(Off-grid binaural sound source localization based on weighted wideband sparse Bayesian learning, WWSBL-OGBSSL)。首先建立离格双耳信号的稀疏信号模型，将离格双耳声源定位问题转化为一个凸优化问题，然后基于双耳相干与扩散能量比(Binaural coherent-to-diffuse power ratio, BCDR)特征对双耳信号的各个频点进行加权以降低噪声和混响的影响，最后利用加权宽带稀疏贝叶斯学习方法来估计离格声源的方位角。WWSBL-OGBSSL算法通过离格稀疏信号模型将声源方位角和测量方位角之间的偏离值作为估计参数进行迭代运算，有效提高了离格声源的方位角估计准确率。仿真和实际实验结果表明，和现有的基于HRTF的双耳声源定位方法相比，WWSBL-OGBSSL算法在各种复杂的声学环境下都有着更高的定位精度和更强的鲁棒性，特别是提高了离格情况下的双耳声源定位性能。

本文常用的符号如下：$\bar{\boldsymbol{x}}$、$\boldsymbol{x}^{\mathrm{T}}$和$\boldsymbol{x}^{\mathrm{H}}$分别表示$\boldsymbol{x}$的共轭、转置和共轭转置；$\boldsymbol{A}^{P\times Q}$表示一个$P\times Q$的矩阵，$\boldsymbol{0}^{P\times Q}$表示$P\times Q$的全零矩阵，$\boldsymbol{I}_P$表示一个$P\times P$的单位矩阵，$\mathrm{diag}(\boldsymbol{x})$表示一个对角矩阵，其对角线的元素与向量$\boldsymbol{x}$的元素相同；$\mathrm{tr}(\boldsymbol{A})$表示矩阵$\boldsymbol{A}$的迹，$(\boldsymbol{A})^{ij}$表示矩阵$\boldsymbol{A}$中的$(i,j)$元素值。$\|\boldsymbol{A}^{P\times Q}\|_1$和$\|\boldsymbol{A}^{P\times Q}\|_2^2$分别表示$\boldsymbol{A}^{P\times Q}$的L1范数和L2范数；$\mathbf{C}$表示复数集。

1 信号模型

1.1 离格双耳信号的稀疏表示模型

假设$s(n)$为点声源，$x_l(n)$和$x_r(n)$分别为左右耳传声器采集到的声信号。研究表明，声源到双耳内传声器的房间传递函数与声源到传声器的距离、声源的方位角和俯仰角密切相关[12]。本文只考虑远场声源水平面方位角定位问题，此时双耳信号

$x_{l|r}(n)$ ($x_l(n)$ 或 $x_r(n)$) 可表示为

$$x_{l|r}(n) = h_{l|r}(\theta) * s(n) + v_{l|r}(n), \quad (1)$$

其中，“$*$”为卷积运算符，$h_{l|r}(\theta)$ 为声源到达左右耳 (左耳或右耳) 传声器的房间脉冲响应，θ 为声源的方位角，$v_{l|r}(n)$ 为左右耳传声器接收到的环境噪声。在频域中，式 (1) 可表示为

$$X_{l|r,k}(m) = H_{l|r,k}(\theta) S_k(m) + V_{l|r,k}(m), \quad (2)$$

其中，$X_{l|r,k}(m)$、$H_{l|r,k}(\theta)$、$S_k(m)$、$V_{l|r,k}(m)$ 分别为 $x_{l|r}(n)$、$h_{l|r}(\theta)$、$s(n)$、$v_{l|r}(n)$ 第 m 帧 N_{STFT} 点短时傅里叶变换 (Short-time Fourier transform, STFT) 第 k 个频率分量，$k \in \{0, 1, \cdots, K-1\}$，$K$ 为频点总数。

声源方位角 θ 对应的导向矢量可定义为 $\boldsymbol{a}_k(\theta) = [H_{l,k}(\theta)/H_{r,k}(\theta), 1]^{\mathrm{T}}$，那么式 (2) 可近似为

$$\boldsymbol{X}_k(m) \approx \boldsymbol{a}_k(\theta) X_{r,k}(m) + \boldsymbol{V}_k(m), \quad (3)$$

其中，$\boldsymbol{X}_k(m) = [X_{l,k}(m), X_{r,k}(m)]^{\mathrm{T}}$，$\boldsymbol{V}_k(m) = [V_{l,k}(m), V_{r,k}(m)]^{\mathrm{T}}$。

假设 HRTF 数据库在人工头前半水平面内包含 J 个等间隔分布的测量方位角，为 $\tilde{\boldsymbol{\theta}} = \{\tilde{\theta}_1, \cdots, \tilde{\theta}_j, \cdots, \tilde{\theta}_J\}$，方位角间隔为 δ。若声源方位角 θ 满足 $\theta = \tilde{\theta}_q$ 且 $\tilde{\theta}_q \in \{\tilde{\theta}_1, \cdots, \tilde{\theta}_j, \cdots, \tilde{\theta}_J\}$，那么该声源为在格声源，对应的双耳信号为在格双耳信号；若声源方位角 θ 位于测量方位角之间，即 $\theta \notin \{\tilde{\theta}_1, \cdots, \tilde{\theta}_j, \cdots, \tilde{\theta}_J\}$，那么该声源为离格声源，对应的双耳信号为离格双耳信号。利用 HRTF 数据库中的头相关脉冲响应 (Head-related impulse responses, HRIRs) 可计算出每个测量方位角 $\tilde{\theta}_j$ 对应的导向矢量 $\boldsymbol{a}_k(\tilde{\theta}_j)$，由此双耳声源定位中的字典矩阵可表示为 $\boldsymbol{A}_k^{2\times J} = \left[\boldsymbol{a}_k(\tilde{\theta}_1), \cdots, \boldsymbol{a}_k(\tilde{\theta}_j), \cdots, \boldsymbol{a}_k(\tilde{\theta}_J)\right]$。

如图 1 所示，若声源方位角 $\theta \notin \{\tilde{\theta}_1, \cdots, \tilde{\theta}_j, \cdots, \tilde{\theta}_J\}$，那么声源方位角 θ 对应的导向矢量 $\boldsymbol{a}_k(\theta)$ 不在字典矩阵 $\boldsymbol{A}_k^{2\times J}$ 中，这会导致字典不匹配问题。假设 $\tilde{\theta}_p \in \{\tilde{\theta}_1, \cdots, \tilde{\theta}_j, \cdots, \tilde{\theta}_J\}$ 且 $\tilde{\theta}_p$ 为 $\tilde{\boldsymbol{\theta}}$ 中距离 θ 最近的测量方位角，利用一阶泰勒展开，$\boldsymbol{a}_k(\theta)$ 可表示为

$$\boldsymbol{a}_k(\theta) \approx \boldsymbol{a}_k(\tilde{\theta}_p) + \boldsymbol{b}_k(\tilde{\theta}_p)(\theta - \tilde{\theta}_p), \quad (4)$$

其中，$\boldsymbol{b}_k(\tilde{\theta}_p)$ 为 $\boldsymbol{a}_k(\tilde{\theta}_p)$ 的一阶偏导数。定义一个偏导数矩阵 $\boldsymbol{B}_k^{2\times J} = \left[\boldsymbol{b}_k(\tilde{\theta}_1), \cdots, \boldsymbol{b}_k(\tilde{\theta}_j), \cdots, \boldsymbol{b}_k(\tilde{\theta}_J)\right]$ 和一个偏移矢量 $\boldsymbol{\beta} = \left[\beta_1, \cdots, \beta_j, \cdots, \beta_J\right]^{\mathrm{T}}$，其中，

$$\beta_j = \begin{cases} \tilde{\theta}_p - \theta, & j = p, \\ 0, & \text{其他}, \end{cases} \quad (5)$$

那么包含离格偏移参数的离格字典矩阵可表示为

$$\boldsymbol{\Phi}_k^{2\times J}(\boldsymbol{\beta}) = \boldsymbol{A}_k^{2\times J} + \boldsymbol{B}_k^{2\times J} \operatorname{diag}(\boldsymbol{\beta}). \quad (6)$$

进一步定义一个稀疏系数向量 $\boldsymbol{Y}_k(m) = \left[Y_k(1,m), \cdots, Y_k(j,m), \cdots, Y_k(J,m)\right]^{\mathrm{T}}$，其中，

$$Y_k(j,m) = \begin{cases} X_{r,k}(m), & j = p, \\ 0, & \text{其他}, \end{cases} \quad (7)$$

那么，离格双耳信号的稀疏表示模型为

$$\boldsymbol{X}_k(m) \approx \boldsymbol{\Phi}_k^{2\times J}(\boldsymbol{\beta}) \boldsymbol{Y}_k(m) + \boldsymbol{V}_k(m). \quad (8)$$

由于声源方位角 θ 为未知量，因此稀疏系数向量 $\boldsymbol{Y}_k(m)$ 和方位角偏移矢量 $\boldsymbol{\beta}$ 都是未知量。基于声源的空间稀疏性，可将式 (8) 中的离格声源方位角估计问题转化为一个凸优化问题，并通过稀疏重建方法 [13] 估计 $\boldsymbol{Y}_k(m)$ 和 $\boldsymbol{\beta}$。离格声源方位角估计问题可简化为

$$\arg\min_{\boldsymbol{Y}_k(m), \boldsymbol{\beta}} \sum_{k=1}^{K} \left\{ \left\| \boldsymbol{X}_k(m) - \boldsymbol{\Phi}_k^{2\times J}(\boldsymbol{\beta}) \boldsymbol{Y}_k(m) \right\|_2^2 + \lambda \left\| \boldsymbol{Y}_k(m) \right\|_1 \right\}, \quad (9)$$

其中，λ 为常量，表示拉格朗日乘子。

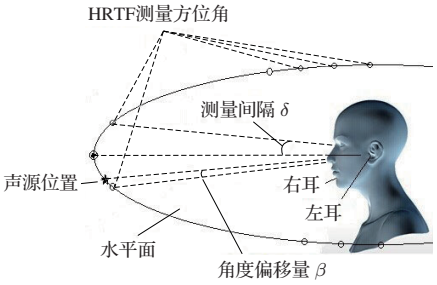


图 1　离格双耳声源定位示意图

Fig. 1　Off-grid binaural sound source localization

1.2　导向矢量模型

由于头和躯干阴影效应的影响，测量方位角 $\tilde{\theta}_j$ 处的 $h_{l|r}(\tilde{\theta}_j)$ 难以用 $\tilde{\theta}_j$ 的线性函数表示出来，因此 $\tilde{\theta}_j$ 处的导向矢量 $\boldsymbol{a}_k(\tilde{\theta}_j)$ 和其导数矢量 $\boldsymbol{b}_k(\tilde{\theta}_j)$ 难以通过 HRTF 数据库直接计算得出。本文提出利用 ILD 和 ITD 的参数模型 [5] 计算测量方位角 $\tilde{\theta}_j$ 处的相对传递函数 (Relative transfer functions, RTFs)，进而获取 $\boldsymbol{a}_k(\tilde{\theta}_j)$ 和 $\boldsymbol{b}_k(\tilde{\theta}_j)$。

自由场环境下，每个测量方位角 $\tilde{\theta}_j$ 对应的双耳信号的 ILD 和 ITD 可直接通过 HRTF 数据库中的 HRIRs 计算获取，其计算方法如文献 [5] 所示。假设测量方位角 $\tilde{\theta}_j$ 处双耳信号第 k 个频率分量的

ILD 和 ITD 分别为 $L_k(\tilde{\theta}_j)$ 和 $T_k(\tilde{\theta}_j)$，那么 $\boldsymbol{a}_k(\tilde{\theta}_j)$ 可近似为

$$\boldsymbol{a}_k(\tilde{\theta}_j)\approx\left[10^{\frac{L_k(\tilde{\theta}_j)}{20}}\exp\left(-\mathrm{i}\frac{\pi k f_s}{K}T_k(\tilde{\theta}_j)\right),1\right]^{\mathrm{T}},\quad(10)$$

其中，f_s 为信号采样率，$\mathrm{i}=\sqrt{-1}$ 为虚数单位。根据文献 [5]，$L_k(\tilde{\theta}_j)$ 和 $T_k(\tilde{\theta}_j)$ 与测量方位角 $\tilde{\theta}_j$ 之间的关系可近似为以下参数模型：

$$L_k\left(\tilde{\theta}_j\right)\approx\xi_k\sin\tilde{\theta}_j,\quad(11)$$

$$T_k\left(\tilde{\theta}_j\right)\approx\varsigma_k\frac{d(\tilde{\theta}_j+\sin\tilde{\theta}_j)}{2c},\quad(12)$$

其中，ξ_k 和 ς_k 为与频率有关的参数，d 为双耳内传声器之间的距离，c 为空气中的声速。ξ_k 和 ς_k 可通过最小均方 (Least square, LS) 算法计算得出 [5]。将式 (11) 和式 (12) 代入到式 (10) 中，并对其求导，可得

$$\boldsymbol{b}_k\left(\tilde{\theta}_j\right)=\left[\frac{\ln 10}{20}\xi_k\cos\tilde{\theta}_j\cdot\Theta_k(\tilde{\theta}_j)-\mathrm{i}\frac{d\varsigma_k(1+\cos\tilde{\theta}_j)}{2c}\cdot\Theta_k(\tilde{\theta}_j),0\right]^{\mathrm{T}},\quad(13)$$

其中，$\Theta_k(\tilde{\theta}_j)=10^{\frac{L_k(\tilde{\theta}_j)}{20}}\exp\left(-\mathrm{i}\frac{\pi k f_s}{K}T_k(\tilde{\theta}_j)\right)$。

1.3 基于双耳相干与扩散能量比的加权方法

与现有的基于 HRTF 的双耳声源定位算法类似，若直接采用稀疏重构方法求解式 (9) 中的离格双耳声源定位问题，那么算法的性能在噪声或混响条件下会显著下降 [5−6]。研究表明，利用双耳相干与扩散能量比 (BCDR) 特征对各频点双耳特征进行加权处理可以明显提高噪声或混响条件下双耳声源定位算法的性能 [14−15]。因此本文进一步提出利用 BCDR 特征对各个频点的双耳信号进行加权以降低噪声和混响的影响。

在一般声场中，双耳信号可以分为相干信号成分 $x_{l|r,ss}(n)$ 和扩散信号成分 $x_{l|r,dd}(n)$。在 BSSL 中，声源角度是未知量，因此本文根据双耳信号 $x_{l|r}(n)$ 的相干函数 $\Gamma_{xx,k}(m)$ 和扩散信号 $x_{l|r,dd}(n)$ 的相干函数 $\Gamma_{dd,k}(m)$ 来估计 BCDR。$x_{l|r}(n)$ 的相干函数 $\Gamma_{xx,k}(m)$ 可以表示为

$$\Gamma_{xx,k}(m)=\frac{\mathrm{E}\left\{(X_{l,k}(m))^*(X_{r,k}(m))\right\}}{\sqrt{\mathrm{E}\{|X_{l,k}(m)|^2\}\mathrm{E}\{|X_{r,k}(m)|^2\}}},\quad(14)$$

其中，$(\cdot)^*$ 表示取共轭，$\mathrm{E}\{\cdot\}$ 为期望函数。Lindevald 等 [16] 的研究表明，远场条件下房间内 $x_{l|r,dd}(n)$ 的相干函数 $\Gamma_{dd,k}(m)$ 与帧数 m 无关，可以近似为

$$\Gamma_{dd,k}(m)\equiv\Gamma_{dd,k}\approx\frac{1}{\sqrt{1+(\pi f_k d/c)^4}}\frac{\sin(4.4\pi f_k d/c)}{4.4\pi f_k d/c},\quad(15)$$

其中，$f_k=k\cdot f_s/2(K-1)$。那么 BCDR 的无偏估计公式为 [17]

$$\mathrm{BCDR}_k(m)=\frac{\Gamma_{dd,k}\Re\{\Gamma_{xx,k}(m)\}-|\Gamma_{xx,k}(m)|^2-\sqrt{(\Gamma_{dd,k})^2\Re\{|\Gamma_{xx,k}(m)|^2\}-|\Gamma_{xx,k}(m)|^2+\left(\Gamma_{dd,k}-\Re\{\Gamma_{xx,k}(m)\}\right)^2}}{|\Gamma_{xx,k}(m)|^2-1},\quad(16)$$

其中，$\Re\{\cdot\}$ 表示取实部。定义加权系数 $W_k(m)$ 为

$$W_k(m)=\begin{cases}1, & \mathrm{BCDR}_k(m)>1,\\ 0, & \text{其他.}\end{cases}\quad(17)$$

令 $\tilde{\boldsymbol{X}}_k(m)=W_k(m)\cdot\boldsymbol{X}_k(m)$，那么经过特征加权的离格声源方位角估计问题可表示为

$$\underset{\boldsymbol{Y}_k(m),\boldsymbol{\beta}}{\arg\min}\sum_{k=1}^{K}\left\{\left\|\tilde{\boldsymbol{X}}_k(m)-\boldsymbol{\Phi}_k^{2\times J}(\boldsymbol{\beta})\boldsymbol{Y}_k(m)\right\|_2^2+\lambda\left\|\boldsymbol{Y}_k(m)\right\|_1\right\}.\quad(18)$$

2 基于加权宽带稀疏贝叶斯学习的离格双耳声源定位算法

在多测量矢量 (Multiple measurement vector, MMV) 模型下，本节推导加权宽带稀疏贝叶斯学习算法 (WWSBL)，用于解决式 (18) 的凸优化问题，从而实现离格声源方位角的估计。

在 MMV 模型下，经过特征加权后的离格双耳信号的稀疏表示模型为

$$\tilde{\boldsymbol{X}}_k^{2\times M}\approx\boldsymbol{\Phi}_k^{2\times J}(\boldsymbol{\beta})\boldsymbol{Y}_k^{J\times M}+\boldsymbol{V}_k^{2\times M},\quad(19)$$

其中，$\tilde{\boldsymbol{X}}_k^{2\times M}=\left[\tilde{\boldsymbol{X}}_k(1),\cdots,\tilde{\boldsymbol{X}}_k(M)\right]$，$\boldsymbol{Y}_k^{J\times M}=\left[\boldsymbol{Y}_k(1),\boldsymbol{Y}_k(2),\cdots,\boldsymbol{Y}_k(M)\right]$，$\boldsymbol{V}_k^{2\times M}=\left[\boldsymbol{V}_k(1),\cdots,\boldsymbol{V}_k(M)\right]$，$M$ 为 MMV 模型中的快拍数。

本文中，$\left\{\tilde{\boldsymbol{X}}_k^{2\times M}\right\}$、$\left\{\boldsymbol{Y}_k^{J\times M}\right\}$ 和 $\left\{\boldsymbol{V}_k^{2\times M}\right\}$ 分别表示所有频点的 $\tilde{\boldsymbol{X}}_k^{2\times M}$、$\boldsymbol{Y}_k^{J\times M}$ 和 $\boldsymbol{V}_k^{2\times M}$ 的集合。假设双耳内的传声器接收到的环境噪声 $v_{l|r}(n)$ 为高斯白噪声，则可进一步假设每帧信号各个频率分量的噪声信号 $\boldsymbol{V}_k(m)=[V_{l,k}(m),V_{r,k}(m)]^{\mathrm{T}}$ 相互独立且符合同一高斯分布，即 $\boldsymbol{V}_k(m)\sim$

$\mathcal{CN}\left(\mathbf{0}^{2\times1},\varepsilon\boldsymbol{I}_2\right)$，其中$\varepsilon$为$V_{l,k}(m)$的方差。由于$\boldsymbol{V}_k(m)$相互独立，$\left\{\boldsymbol{V}_k^{2\times M}\right\}$的概率密度函数可表示为

$$p(\{\boldsymbol{V}_k^{2\times M}\}|\varepsilon)=\prod_{m=1}^{M}\prod_{k=1}^{K}\mathcal{CN}(\boldsymbol{V}_k(m)|\mathbf{0}^{2\times1},\varepsilon\boldsymbol{I}_2). \tag{20}$$

为了估计稀疏系数矩阵$\boldsymbol{Y}_k^{J\times M}$和方位角偏移矢量$\boldsymbol{\beta}$，需要已知二者的先验概率分布。在高斯混合模型下，假设每帧信号各个频率分量对应的稀疏系数向量$\boldsymbol{Y}_k(m)$相互独立，且符合同一复高斯分布$\boldsymbol{Y}_k(m)\sim\mathcal{CN}\left(\mathbf{0}^{J\times1},\boldsymbol{\Lambda}^{J\times J}\right)$，其中，协方差矩阵$\boldsymbol{\Lambda}^{J\times J}=\mathrm{diag}\left(\boldsymbol{\alpha}\right)$为一个对角矩阵，$\boldsymbol{\alpha}=[\alpha_1,\cdots,\alpha_j,\cdots,\alpha_J]$为$\boldsymbol{Y}_k(m)$中各个元素的方差。根据高斯分布的性质，$\alpha_j$的先验概率分布可假设为独立同分布的Gamma分布。稀疏系数矩阵$\left\{\boldsymbol{Y}_k^{J\times M}\right\}$以及$\boldsymbol{\alpha}$的先验概率密度函数可表示为

$$\begin{aligned}&p\left(\left\{\boldsymbol{Y}_k^{J\times M}\right\}|\boldsymbol{\alpha}\right)\\&=\prod_{m=1}^{M}\prod_{k=1}^{K}\mathcal{CN}\left(\boldsymbol{Y}_k(m)\middle|\mathbf{0}^{J\times1},\boldsymbol{\Lambda}^{J\times J}\right),\end{aligned} \tag{21}$$

$$p\left(\boldsymbol{\alpha}\right)=\prod_{j=1}^{J}\Gamma\left(\alpha_j|1,\gamma\right), \tag{22}$$

其中，γ为Gamma分布的参数。方位角偏移矢量$\boldsymbol{\beta}$中各个元素的先验分布可假设为相互独立的均匀分布，那么$\boldsymbol{\beta}$的先验概率分布可表示为

$$\boldsymbol{\beta}\sim U\left(\left[-\frac{\delta}{2},\frac{\delta}{2}\right]^J\right). \tag{23}$$

综上，WWSBL算法中的待估参数如下：

$$\Delta=\{\varepsilon,\boldsymbol{\alpha},\boldsymbol{\beta}\}. \tag{24}$$

根据文献[9]，式(24)中的模型参数可通过期望最大化(Expectation maximization, EM)算法进行求解。WWSBL中的EM算法将稀疏系数矩阵$\left\{\boldsymbol{Y}_k^{J\times M}\right\}$作为一个隐含变量处理，即优化过程中不再出现$\left\{\boldsymbol{Y}_k^{J\times M}\right\}$，而将参数$\boldsymbol{\alpha}$和偏移向量$\boldsymbol{\beta}$作为优化对象，通过最大化

$$\mathrm{E}\left\{\lg\left[p\left(\left\{\tilde{\boldsymbol{X}}_k^{2\times M}\right\},\left\{\boldsymbol{Y}_k^{J\times M}\right\},\varepsilon,\boldsymbol{\alpha},\boldsymbol{\beta}\right)\right]\right\}$$

来估计各个参数的最佳值。各个参数的迭代更新公式如下：

$$\begin{aligned}\varepsilon^{\mathrm{new}}=&\sum_{m=1}^{M}\sum_{k=1}^{K}\left\|\tilde{\boldsymbol{X}}_k(m)-\boldsymbol{\Phi}_k^{2\times J}(\boldsymbol{\beta})\,\boldsymbol{\mu}_k(m)\right\|_2^2\\&+\frac{1}{2K}\sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{\Sigma}_k^{J\times J}\left(\boldsymbol{\Phi}_k^{2\times J}(\boldsymbol{\beta})\right)^{\mathrm{H}}\boldsymbol{\Phi}_k^{2\times J}(\boldsymbol{\beta})\right),\end{aligned} \tag{25}$$

$$\alpha_j^{\mathrm{new}}=\frac{1}{KM}\sum_{k=1}^{K}\sum_{m=1}^{M}\left(\boldsymbol{\Sigma}_k^{J\times J}+\boldsymbol{\mu}_k(m)\left(\boldsymbol{\mu}_k(m)\right)^{\mathrm{H}}\right)^{jj}, \tag{26}$$

其中，$\boldsymbol{\mu}_k(m)$和$\boldsymbol{\Sigma}_k^{J\times J}$分别为$\boldsymbol{Y}_k(m)$后验概率分布的均值和方差，可通过高斯混合模型推导计算出来。每次迭代中，更新了参数ε和$\boldsymbol{\alpha}$之后，再更新偏移向量$\boldsymbol{\beta}$。假设$\boldsymbol{\alpha}^{\mathrm{new}}$的第$j_{\mathrm{opt}}$个元素为$\boldsymbol{\alpha}^{\mathrm{new}}$的最大值，那么只更新$\boldsymbol{\beta}$的第$j_{\mathrm{opt}}$个元素，其他元素保持不变。偏移向量$\boldsymbol{\beta}$的更新公式如下：

$$\begin{aligned}\boldsymbol{\beta}^{\mathrm{new}}=&\mathop{\arg\min}_{\boldsymbol{\beta}\in[-\delta/2,\delta/2]^J}\mathrm{E}\Bigg\{\sum_{m=1}^{M}\sum_{k=1}^{K}\big\|\tilde{\boldsymbol{X}}_k(m)\\&-\boldsymbol{\Phi}_k^{2\times J}(\boldsymbol{\beta})\,\boldsymbol{\mu}_k(m)\big\|_2^2\Bigg\}.\end{aligned} \tag{27}$$

偏移向量$\boldsymbol{\beta}$的更新公式无法用显式表达，可以通过遍历法得到最优解。

当$\|\boldsymbol{\alpha}^{\mathrm{new}}-\boldsymbol{\alpha}\|_2^2/\|\boldsymbol{\alpha}\|_2^2\leqslant0.001$或者迭代次数超过1000次时，停止迭代，得到模型中各个参数的最优解$\tilde{\varepsilon}$、$\tilde{\boldsymbol{\alpha}}$和$\tilde{\boldsymbol{\beta}}$。假设$\tilde{\boldsymbol{\alpha}}$的最大元素值为$\tilde{\alpha}_{j_{\mathrm{opt}}}$，那么离格声源的方位角估计值为

$$\hat{\theta}=\tilde{\theta}_{j_{\mathrm{opt}}}+\tilde{\beta}_{j_{\mathrm{opt}}}. \tag{28}$$

3　实验结果及分析

本文分别在仿真和实际声学环境下对WWSBL-OGBSSL算法的性能进行了测试。3.1节测试了本文算法在自由场环境下的双耳声源方位角估计性能，3.2节测试了本文算法在噪声环境下的方位角估计性能，3.3节测试了本文算法在混响环境下的方位角估计性能，3.4节测试了本文算法在实际环境下的方位角估计性能。

在自由场环境和噪声环境下，实验中的双耳信号是由HRTF数据库中的HRIRs卷积纯净语音信号生成。本文选用的HRTF数据库为MIT HRTF数据库[18]，纯净语音信号选自TIMIT数据库[19]。由于只考虑声源水平角的估计，因此本文算法只采用了HRTF数据库中前半水平面的HRIRs数据。本文将生成的双耳信号分帧加窗后，提取ILD、ITD等双耳特征。双耳信号的采样率为16 kHz，帧长为32 ms，帧移为16 ms，窗函数采用汉明窗。由于MIT HRTF数据库使用的KEMAR人工头半径为7.6 cm，因此本文将ITD特征的取值范围限定为$[-1,1]$ ms，同时将ILD特征的取值范围设定为$[-40,40]$ dB。空气中的声速为343 m/s。

本文选取两种现有的基于HRTF的双耳声源定位方法与WWSBL-OGBSSL算法作对比，分别为Finger等[6]提出的在线校准(Online calibration, OC)算法和Liu等[20]提出的双耳匹配滤波器(Interaural matching filter, IMF)定位算法。本文中声源方位角估计的均方根误差(Root mean square error, RMSE)定义如下：

$$\mathrm{RMSE}=\sqrt{\frac{1}{L}\sum_{l=1}^{L}|\hat{\theta}_l-\theta_l|^2}, \tag{29}$$

其中，L为双耳信号数据段总数，θ_l为第l个数据段的声源真实方位角，$\hat{\theta}_l$为对应的声源方位角估计值。声源方位角的估计准确率(Accuracy, Acc)定义如下：

$$\mathrm{Acc}=L_{|\hat{\theta}_l-\theta_l|\leqslant 10^\circ}/L\times 100\%, \tag{30}$$

其中，$L_{|\hat{\theta}_l-\theta_l|\leqslant 10^\circ}$为方位角估计误差不大于10°的数据段总数。

3.1 自由场环境下的双耳声源定位实验

本小节通过仿真实验测试WWSBL-OGBSSL算法在自由场环境下的方位角估计性能。在前半水平面内，MIT HRTF数据库包含37个方位角的HRIRs，分别为$\{-90^\circ,-85^\circ,\cdots,85^\circ,90^\circ\}$，方位角间隔为5°。为了仿真在格声源和离格声源的情况，可假设$\tilde{\boldsymbol{\theta}}=\{-90^\circ,-80^\circ,\cdots,80^\circ,90^\circ\}$为所有的测量方位角，这些方位角对应的HRIRs数据用于生成在格双耳信号，其余18个方位角的HRIRs数据用于生成离格双耳信号。测量方位角共有19个，方位角间隔$\delta=10^\circ$。在每个方位角下，随机选取TIMIT数据库的400句语音信号仿真生成400句自由场环境下的双耳信号。首先将每个双耳信号分成时长为1 s的数据段，并对每段双耳信号分帧，然后基于语音端点检测(Voice activity detection, VAD)算法去除非语音帧数据。分别采用OC算法、IMF算法和WWSBL-OGBSSL算法对每段信号进行声源方位角估计。每段信号中语音帧的总数即为WWSBL-OGBSSL算法中MMV 模型的快拍数。自由场环境下的加权系数$W_k(m)$恒为1。图2给出了自由场环境下三种算法对在格声源和离格声源的方位角估计均方根误差(RMSE)曲线图，“on-grid”和“off-grid”分别表示在格声源和离格声源，“Proposed”表示WWSBL-OGBSSL算法。

从图2中可以看出，WWSBL-OGBSSL算法对在格声源的定位性能稍优于OC算法和IMF算法，对离格声源的定位性能明显优于OC算法和IMF算法。这是因为OC算法和IMF算法的方位角估计结果被限定在了离散测量方位角上，而WWSBL-OGBSSL算法通过迭代估计出离声源真实方位角最近的测量方位角和二者之间的偏移量，估计结果可能为声源真实方位角附近的任意值。OC算法和IMF算法对在格声源的方位角估计误差为$\{0^\circ,10^\circ,20^\circ,\cdots\}$，对离格声源的方位角估计误差为$\{5^\circ,15^\circ,25^\circ,\cdots\}$，最小估计误差为5°；而WWSBL-OGBSSL对在格声源和离格声源的方位角估计误差都可以为任意小的值，在理想情况下误差可以降至0°，因此WWSBL-OGBSSL算法可以显著提高离格条件下的双耳声源方位角估计性能。

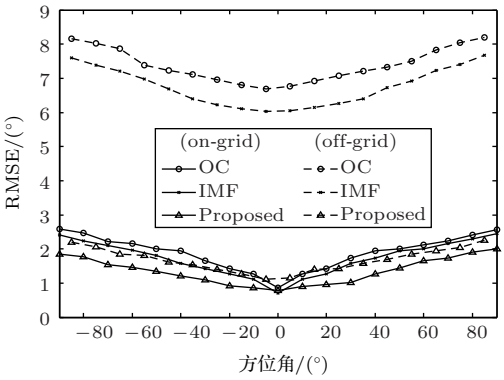


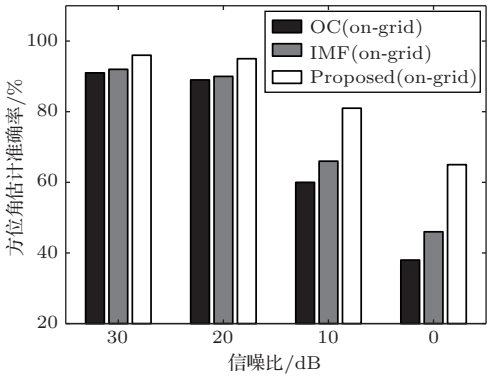
图2 自由场环境下三种算法对在格声源和离格声源的方位角估计的RMSE曲线

Fig. 2 The RMSE of azimuth estimation of the OC, the IMF, the WWSBL-OGBSSL methods for the on-grid sound sources and the off-grid sound sources

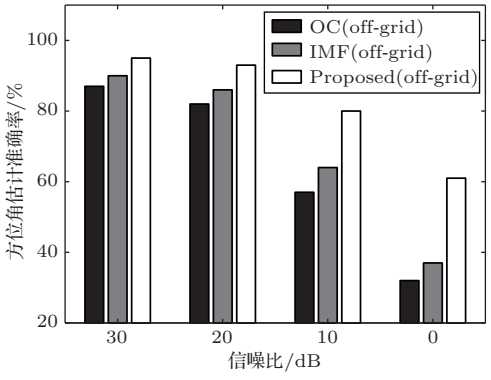
3.2 不同噪声环境下的双耳声源定位实验

本实验主要测试了WWSBL-OGBSSL算法在不同信噪比下的方位角估计性能。为了模拟噪声环境，本实验在3.1节自由场环境下生成的双耳信号中加入扩散场噪声，生成带噪双耳信号。本实验中在格双耳信号的声源方位角为−30°，离格双耳信号的声源方位角为25°，其他实验条件与3.1节相同。扩散场噪声是由MIT HRTF数据库中72个水平面方位角的HRIRs卷积高斯白噪声后叠加生成的。带噪双耳信号的信噪比(Signal-to-noise ratio, SNR)设定为0 dB到30 dB，间隔为10 dB。本实验采用方位角估计准确率指标来衡量三种算法对在格声源和离格声源的方位角估计性能。图3给出了不同信噪比下三种算法对在格声源和离格声源的方位角估计准确率。

从图3中可以看出，在不同的噪声环境下，WWSBL-OGBSSL算法对在格声源的方位角估计准确率比OC算法和IMF算法高出3%~15%，对离格声源的方位角估计准确率比OC 算法和IMF算法高出5%~20%。特别是在信噪比为0 dB时，WWSBL-OGBSSL算法对在格声源和离格声源的方位角估计准确率分别比OC算法和IMF算法高出15%和20%左右。值得注意的是，OC算法和IMF算法对离格声源的最小估计误差为5°，WWSBL-OGBSSL算法对离格声源的最小估计误差理论上可达到0°。WWSBL-OGBSSL算法基于BCDR特征对各个频点进行加权，将扩散场噪声占主要成分的频点去除，降低扩散噪声对方位角估计性能的影响，有效提高了双耳声源方位角估计准确率；同时WWSBL-OGBSSL算法基于各频点信号能量优化模型参数，能量强的频点具有更大的权重，从而进一步提高了噪声环境下的方位角估计准确率。



(a) 在格声源



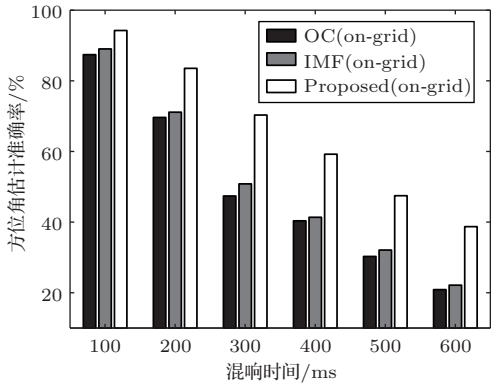
(b) 离格声源

图3　不同信噪比下OC算法、IMF算法和WWSBL-OGBSSL算法的方位角估计准确率

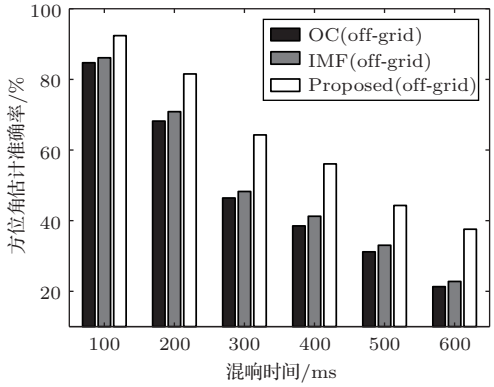
Fig. 3 The azimuth estimation accuracies of the OC, the IMF, the WWSBL-OGBSSL methods under different SNRs

3.3 不同混响环境下的双耳声源定位实验

本小节通过仿真实验测试了WWSBL-OGBSSL算法在不同混响条件下的方位角估计性能。双耳信号是由不同混响条件下的双耳房间脉冲响应(Binaural room impulse responses, BRIRs)卷积纯净语音信号生成。不同混响条件下的BRIRs是由MIT HRTF数据库中的HRIRs经镜像法[21]模拟生成。混响时间分别设定为100 ms到600 ms，间隔为100 ms。本实验中在格双耳信号的声源方位角为−30°，离格双耳信号的声源方位角为25°。图4给出了不同混响条件下三种算法对在格声源和离格声源的方位角估计准确率。



(a) 在格声源



(b) 离格声源

图4　不同混响条件下OC算法、IMF算法和WWSBL-OGBSSL算法的方位角估计准确率

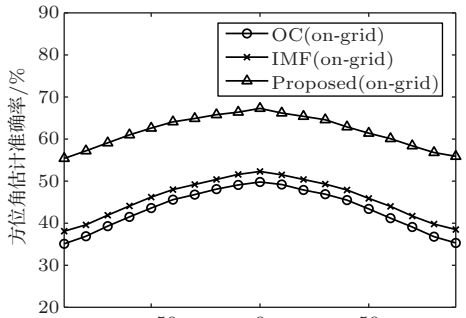
Fig. 4 The azimuth estimation accuracies of the OC, the IMF, the WWSBL-OGBSSL methods under different reverberation times

从图4中可以看出，在不同混响条件下，WWSBL-OGBSSL算法对在格声源和离格声源的方位角估计准确率比OC算法和IMF算法高出2%~15%。随着混响增大，OC算法和IMF 算法的方位角估计性能都会严重恶化，这是因为混响情

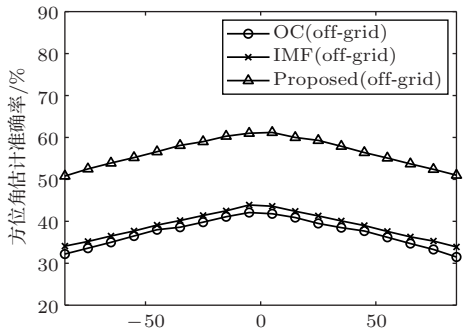
况下，由于房间反射，在格双耳信号和离格双耳信号的双耳特征(ILD、ITD等)与自由场环境测量的HRTF数据库中提取的双耳特征严重不匹配。WWSBL-OGBSSL算法基于BCDR对各个频点的双耳信号进行加权，将混响占主要成分的频点去除，有效降低了混响对方位角估计性能的影响；而且WWSBL-OGBSSL算法基于各频点的能量对各个参数迭代更新，能量强的频点会有更大的权重，因此在混响条件下WWSBL-OGBSSL算法的方位角估计性能更优。

3.4 实际环境下的双耳声源定位实验

本小节通过实际实验测试了WWSBL-OGBSSL算法在实际环境下的方位角估计性能。本文在一个铺设有吸声材料的房间内采用B&K 4128人工头采集双耳信号。房间的大小约为6.4 m×4.8 m×2.8 m，混响时间约为$T_{60}\approx 350$ ms，混响半径$r_0\approx 1.60$ m。声源位于人工头的水平面上，距人工头的距离为1.80 m，真实方位角分别为$\{-90°,-85°,\cdots,85°,90°\}$，



(a) 在格声源



(b) 离格声源

图5 实际环境下OC算法、IMF算法和WWSBL-OGBSSL算法的方位角估计准确率

Fig. 5 The azimuth estimation accuracies of the OC, the IMF, the WWSBL-OGBSSL methods in real environments

方位角间隔为5°。在每个方位角处，随机选取TIMIT数据库的200句语音信号作为声源信号，采集200句双耳信号。假设$\tilde{\boldsymbol{\theta}}=\{-90°,-80°,\cdots,80°,90°\}$为HRTF数据库中所有的测量方位角，那么当声源真实方位角$\theta\in\{-90°,-80°,\cdots,80°,90°\}$时，声源为在格声源，当声源真实方位角$\theta\in\{-85°,-75°,\cdots,75°,85°\}$时，声源为离格声源。将每个双耳信号划分为时长为1 s的双耳信号数据段，然后分别采用OC算法、IMF算法和WWSBL-OGBSSL算法估计每段信号的方位角。图5给出了实际环境下三种算法对在格声源和离格声源的方位角估计准确率。

从图5中可以看出，在实际环境下WWSBL-OGBSSL算法对在格声源和离格声源的方位角估计准确率比OC算法和IMF算法高出15%左右。这是因为WWSBL-OGBSSL算法基于BCDR对各个频点进行加权，去除了受混响影响比较严重的频点的双耳信号，有效降低了混响的影响；而且WWSBL-OGBSSL算法中能量高的频点有着更大的权重。另外，从图5中可以看出三种算法对人工头正前方声源的方位角估计性能明显优于人工头两侧声源的方位角估计性能，这是因为人工头正前方声源方位角的变化对双耳信号双耳特征的影响更显著，因此定位性能更好。

4 结论

针对双耳声源定位中的离格问题，提出了基于加权宽带稀疏贝叶斯学习的离格双耳声源定位算法(WWSBL-OGBSSL)。首先，该算法基于压缩感知理论建立了离格稀疏双耳信号模型，将离格双耳声源定位问题简化为一个凸优化问题，并采用双耳相干与扩散能量比特征对各个频点进行加权以降低噪声和混响的影响，然后通过加权宽带稀疏贝叶斯学习方法来估计模型参数，最终实现离格声源方位角估计。与现有的离格阵列DOA估计算法相比，离格双耳声源定位算法既考虑了离格问题的影响，也考虑了头和躯干的阴影效应的影响。仿真和实际实验结果表明，本文算法在各种声学环境下都有着更高的定位精度和更强的鲁棒性，特别是提高了离格条件下的双耳声源方位角估计性能。

参 考 文 献

[1] 李军锋, 徐华兴, 夏日升, 等. 基于听觉感知特性的双耳音频处理技术 [J]. 应用声学, 2018, 37(5): 706–716.

Li Junfeng, Xu Huaxing, Xia Risheng, et al. Binaural audio technologies based on human auditory perception[J]. Journal of Applied Acoustics, 2018, 37(5): 706–716.

[2] Farmani M, Pedersen M S, Tan Z, et al. Maximum likelihood approach to "informed" sound source localization for hearing aid applications[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2017, 25(3): 611–623.

[3] Deleforge A, Horaud R, Schechner Y Y, et al. Co-localization of audio sources in images using binaural features and locally-linear regression[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2015, 23(4): 718–731.

[4] Jeffress L A. A place theory of sound localization[J]. Journal of Comparative and Physiological Psychology, 1948, 61: 468–486.

[5] Raspaud M, Viste H, Evangelista G. Binaural source localization by joint estimation of ILD and ITD[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(1): 68–77.

[6] Finger H, Ruvolo P, Liu S C, et al. Approaches and databases for online calibration of binaural sound localization for robotic head[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, 2010: 4340–4345.

[7] Deleforge A, Forbes F, Horaud R. Acoustic space learning for sound source separation and localization on binaural manifolds[J]. International Journal of Neural Systems, 2015, 25(1): 1–20.

[8] Ma N, May T, Brown G J. Exploiting deep neural networks and head movements for robust binaural localization of multiple sources in reverberant environments[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2017, 25(12): 2444–2453.

[9] Yang Z, Xie L H, Zhang C S. Off-grid direction of arrival estimation using sparse Bayesian inference[J]. IEEE Transactions on Signal Processing, 2013, 61(1): 38–43.

[10] 高阳, 陈俊丽, 杨广立. 基于酉变换和稀疏贝叶斯学习的离格 DOA 估计 [J]. 通信学报, 2017, 38(6): 177–182.
Gao Yang, Chen Junli, Yang Guangli. Off-grid DOA estimation based on unitary transform and sparse Bayesian learning[J]. Journal on Communications, 2017, 38(6): 177–182.

[11] 胡顺仁, 刘骁, 李双. 联合稀疏贝叶斯学习与子空间的近场信号源定位 [J]. 信号处理, 2017, 33(3A): 27–32.
Hu Shunren, Liu Xiao, Li Shuang. Source localization of joint subspace method and sparse Bayesian learning in the near-field[J]. Journal of Signal Processing, 2017, 33(3A): 27–32.

[12] Schroeder M. The statistical of frequency responses in large room[J]. Acustica, 1954, 4: 594–600.

[13] 孙洪, 张智林, 余磊. 从稀疏到结构化稀疏: 贝叶斯方法 [J]. 信号处理, 2012, 28(6): 759–773.
Sun Hong, Zhang Zhilin, Yu Lei. From sparsity to structured sparsity: Bayesian perspective[J]. Journal of Signal Processing, 2012, 28(6): 759–773.

[14] 丁建策, 郑成诗, 李晓东. 双耳相干混响比加权的声源定位算法 [C] //2017 年全国声学学术会议, 2017.

[15] Ding J, Wang J, Zheng C, et al. Analysis of binaural features for supervised localization in reverberant environments[C]. 141st Audio Engineering Society Convention, 2016: 1–9.

[16] Lindevald I M, Benade A H. Two-ear correlation in the statistical sound field of rooms[J]. Journal of the Acoustical Society of America, 1986, 80(2): 661–664.

[17] Schwarz A, Kellermann W. Coherent-to-diffuse power ratio estimation for dereverberation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2015, 23(6): 1006–1018.

[18] Gardner B, Martin K. HRTF measurements of a kemar dummy-head microphone[R]. MIT Media Lab. Perceptual Computing-Technical Report, 1994: 1–7.

[19] Garofolo J S. DAPRA TIMIT acoustic-phonetic speech database[DB]. National Institute of Standards and Technology(NIST), 1988.

[20] Liu H, Zhang J, Fu Z. A new hierarchical binaural sound source localization method based on interaural matching filter[C]. IEEE International Conference on Robotics and Automation, 2014: 1598–1605.

[21] Allen J B, Berkley D A. Image method for efficiently simulating small-room acoustics[J]. Journal of the Acoustical Society of America, 1979, 65(4): 943–950.