第 48 卷第 5 期
2023 年 9 月

声 学 学 报
ACTA ACUSTICA

Vol. 48, No. 5
Sept., 2023

利用时频相关性的球谐波阶数感知鲁棒伪声强多声源定位*

高伟霞[1,2] 陈华伟[1†]

(1 南京航空航天大学 电子信息工程学院 南京 211106)
(2 安徽科技学院 电气与电子工程学院 蚌埠 233030)

2022 年 9 月 19 日收到
2022 年 11 月 22 日定稿

摘要 针对伪声强多声源定位方法对环境噪声和混响敏感的问题, 利用时频相关性构造球谐波阶数感知因子, 提出了一种鲁棒的伪声强声源定位方法。与现有的球谐波阶数感知因子不同, 所构造的球谐波阶数感知因子充分利用了相邻时频点中属于同一声源的特征波束间存在的相关性。理论分析表明, 所构造的球谐波阶数感知因子对环境噪声和混响的抑制能力更强。仿真结果显示, 与现有的阶数感知方法相比, 在信噪比为 10 dB、混响时间为 0.4~1.0 s 时, 所提方法的定位精度提升了 1.3°~1.9°, 同时计算复杂度减小了 25%。最后通过实测实验进一步验证了所提方法在实际声场环境中的有效性。

关键词 球面传声器阵列, 声源定位, 阶数感知, 球谐波, 时频相关性

PACS 数 43.60, 43.72 **DOI:** 10.12395/0371-0025.2022085

Localization of multiple speakers in the spherical harmonic domain by robust order-aware pseudo-intensity vectors exploiting time-frequency correlation

GAO Weixia[1,2] CHEN Huawei[1†]

(1 *College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics* Nanjing 211106)
(2 *College of Electrical and Electronic Engineering, Anhui Science and Technology University* Bengbu 233030)

Received Sept. 19, 2022
Revised Nov. 22, 2022

Abstract Multiple sound source localization using pseudo-intensity vector is known sensitive to noise and room reverberation. To deal with the problem, a robust sound localization approach is proposed using pseudo-intensity vector via constructing an order-aware factor of spherical harmonics (OAFSH) by employing signal correlation property in the time-frequency domain. Unlike its existing counterpart, the proposed OAFSH fully takes advantage of the correlation of the eigenbeams that belong to the same source between adjacent time-frequency bins. Theoretical analysis shows that the proposed OAFSH performs better than the existing counterpart in suppressing noise and room reverberation. Simulation results demonstrate that the sound source localization accuracy of the proposed approach is 1.3° ~ 1.9° higher than that of the existing approach under the condition of a 10 dB signal to noise ratio and reverberation time of 0.4 ~ 1.0 s. Moreover, the computational complexity of the proposed approach is also reduced by 25%. Finally, the effectiveness of the proposed approach is further verified by the real-world experiments in practical room environment.

Keywords Acoustic spherical array, Speaker localization, Order-aware, Spherical harmonics, Time-frequency correlation

引言

声源定位是音频信号处理中重要的研究方向之一, 经常被应用于空间滤波、声源分离、去混响、语音增强、视频会议、机器人听觉等领域中[1-5]。目前声源波达方位 (DOA) 主要利用声源定位方法从传声器阵列采集的声场信息中进行估计, 其中球面传声

* 国家自然科学基金项目 (61971219, 61471190) 和安徽高校自然科学研究重点项目 (KJ2020A0076) 资助
† 通讯作者: 陈华伟, hwchen@nuaa.edu.cn

器阵列可以对三维声场进行等分辨率分析, 且便于将采集的声场信息转换到球谐波域中进行处理而受到人们的关注[6-10]。在球谐波域中导向矢量与阵列分布无关, 并且可以将导向矢量与频率解耦[7-8], 这为声源的定位带来了便利。

在球谐波域中常用的声源定位方法有: 可控波束响应方法[10]、子空间类方法[11-12]、最大似然估计方法[13-14]、伪声强 (PIV) 类方法[15-23] 等。其中 PIV[15] 声源定位方法具有声源定位的闭式解, 计算量较低, 适用于实时的声源定位场合, 但是 PIV 方法只采用了 0 阶和 1 阶的声场球谐波特征波束, 使得其空间分辨率较低, 在高混响和多声源环境下定位性能较差[16-17]。为了弥补 PIV 方法的缺陷提出了一些改进方法, 如将声场高阶特征波束引入到 PIV 方法中提升其空间分辨率的增广伪声强 (AIV) 方法[18-19]、局部扫描可控波束响应方法[20]、子空间伪声强方法[21] 和直接路径占优 (DPD) 检测伪声强 (DPD-PIV) 方法[22] 等。

将高阶特征波束引入到 PIV 方法中提升了定位性能, 但是相应地增大了计算量, 且低频段的高阶特征波束易受噪声污染[7-8], 限制了定位方法性能的提升。为了减小该问题的影响常用方法是将低频特征波束舍去[11-22], 但是噪声的随机性使得保留下来的高阶特征波束中仍存在被噪声严重污染的特征波束, 并且若舍去的低频信息较多将会影响能量主要位于低频段的声源的定位。针对该问题, 文献 [23] 提出了阶数感知 (OA) 定位方法, 利用各阶特征波束因噪声产生的差异感知每阶特征波束受噪声污染的程度, 从中筛选出受噪声影响较小的特征波束用于伪声强声源定位中, 提升了定位方法的抗噪能力。并且该方法采用分阶处理思想, 能够将低频段受噪声影响较小的低阶特征波束筛选出来, 从而保留了声源的低频信息, 提升了伪声强定位方法的鲁棒性。

现有阶数感知因子只利用了单个时频点的信息, 限制了其对于混响和多声源干扰程度的鉴别能力, 以及在高频段各阶特征波束因噪声产生的差异减小引起的噪声感知能力下降。为解决现有阶数感知存在的问题, 本文将相邻多个时频点的信息应用到阶数感知中, 利用相邻时频点中源自同一声源的特征波束之间存在的相关性, 构造了局部时频相关的阶数感知因子, 将受噪声污染较小的单声源直达声信号占优的特征波束筛选出来用于伪声强声源定位方法中, 提升定位方法的抗噪和抗混响能力。通过理论分析、仿真实验和实测实验验证, 所提方法相比已有的伪声强声源定位方法在降低计算量的同时具有更高的定位精度和鲁棒性, 且随信噪比的降低、混响时间的增大、声源间距的减小优势提升得越明显。

1 理论基础

1.1 信号模型

考虑由 Q 个传声器构成的半径为 r_a 的球面阵列, 以阵列中心为球坐标系原点, 第 q 个传声器的坐标为 $(r_a,\Omega_q)=(r_a,\theta_q,\varphi_q)$, 其中 θ_q 为俯仰角、φ_q 为方位角。$p(\tau,k,r_a,\Omega_q)$ 为第 q 个传声器采集到的声压信号, 其中 k 为波数, τ 为帧序号。对声压信号进行离散球傅里叶变换 (DSFT) 得到声场的球谐波系数[8]:

$$p_{nm}(\tau,k,r_a)\approx\sum_{q=1}^{Q}\alpha_q p(\tau,k,r_a,\Omega_q)\mathrm{Y}_n^{m*}(\Omega_q), \tag{1}$$

其中, α_q 为离散采样权值, 若阵列为均匀采样分布[8,24] $\alpha_q=4\pi/Q$, $(\cdot)^*$ 表示复共轭, $\mathrm{Y}_n^m(\Omega)$ 为 n 阶 m 维球谐波函数, $-n\leqslant m\leqslant n$, 其定义为[8,24]

$$\mathrm{Y}_n^m(\Omega)=\sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}\mathrm{P}_n^m(\cos\theta)\mathrm{e}^{\mathrm{i}m\phi}, \tag{2}$$

其中, $(\cdot)!$ 为阶乘, $\mathrm{P}_n^m(\cdot)$ 为 n 阶 m 维连带勒让德函数。若需要的最大球谐波阶数为 N, 为避免式 (1) 离散球傅里叶变换出现空间混叠, 要求 N 和传声器个数 Q 之间满足 $(N+1)^2\leqslant Q$ 以及 N 和波数 k 之间满足 $kr<N$ [7-8]。球谐波函数在离散采样点处满足正交性[24]: $\sum_{q=1}^{Q}\alpha_q\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_q)=\delta_{n-n'}\delta_{m-m'}$, $n\leqslant N$, 其中 $\delta_{n-n'}$ 与 $\delta_{m-m'}$ 为克罗内克 δ 函数。

对于由 L 个声源辐射出的平面波在球谐波域中可以表示为[10-15]

$$p_{nm}(\tau,k,r_a)=b_n(kr_a)\sum_{l=1}^{L}s_l(\tau,k)\mathrm{Y}_n^{m*}(\Theta_l)+n_{nm}(\tau,k), \tag{3}$$

其中, $n_{nm}(\tau,k)=\sum_{q=1}^{Q}\alpha_q n_q(\tau,k)\mathrm{Y}_n^{m*}(\Omega_q)$ 为加性噪声的球谐波系数, $n_q(\tau,k)$ 为第 q 个传声器的噪声幅度; $s_l(\tau,k)$ 和 $\Theta_l=(\theta_l,\phi_l)$ 分别为第 l 个声源的幅度和 DOA; $b_n(kr_a)$ 为模式强度, 与波数 k、球阵半径 r_a 以及球阵的配置有关。对于本文采用的远场刚性球其表达式为[24]

$$b_n(kr_a)=4\pi\mathrm{i}^n\left[\mathrm{j}_n(kr_a)-\frac{\mathrm{j}'_n(kr_a)}{\mathrm{h}_n^{(2)\prime}(kr_a)}\mathrm{h}_n^{(2)}(kr_a)\right], \tag{4}$$

其中, i 为虚数符号, $\mathrm{j}_n(\cdot)$ 为第一类球贝塞尔函数, $\mathrm{h}_n^{(2)}(\cdot)$ 为第二类球汉克尔函数, $(\cdot)'$ 表示一阶微分。

对式 (3) 进行模式强度补偿, 即式 (3) 除以 $b_n(kr_a)$, 可以得到特征波束[6-8]:

$$a_{nm}(\tau,k)=\sum_{l=1}^{L}s_l(\tau,k)\mathrm{Y}_n^{m*}(\Theta_l)+n_{nm}(\tau,k)/b_n(kr_a). \tag{5}$$

将n阶所有特征波束表示为向量的形式$\boldsymbol{a_n}(\tau,k)=[a_{n(-n)}(\tau,k),\cdots,a_{n0}(\tau,k),\cdots,a_{nn}(\tau,k)]^{\mathrm{T}}$, $[\cdot]^{\mathrm{T}}$为转置。

1.2　伪声强定位方法

1.2.1　PIV 方法

设声场在时频点(τ,k)的伪声强向量为$\boldsymbol{U}(\tau,k)=[U_x(\tau,k),U_y(\tau,k),U_z(\tau,k)]^{\mathrm{T}}$, 其中任一项的表达式为[15]

$$U_\varsigma(\tau,k)=0.5\mathrm{Re}[a_{00}{}^*(\tau,k)D_{-\varsigma}(\tau,k)],\tag{6}$$

其中, $\mathrm{Re}(\cdot)$表示取实部, $\varsigma\in\{x,y,z\}$。$D_{-\varsigma}(\tau,k)$的表达式为[15]

$$D_{-\varsigma}(\tau,k)=\sum_{m=-1}^{1}\mathrm{Y}_{1,m}(\Psi_{-\varsigma})a_{1m}(\tau,k),\tag{7}$$

式中, $\Psi_{-x}=(\pi/2,\pi)$, $\Psi_{-y}=(\pi/2,-\pi/2)$, $\Psi_{-z}=(\pi,0)$。则估计的 DOA 由下式给出[15]:

$$\boldsymbol{u}(\tau,k)=-\boldsymbol{U}(\tau,k)/\|\boldsymbol{U}(\tau,k)\|,\tag{8}$$

其中$\|\cdot\|$为二范数。从式 (6)—式 (8) 中可以看出 PIV 方法估计声源 DOA 具有闭式解, 计算量较低, 但只利用 0, 1 阶特征波束使其空间分辨率较差。为改善该问题, 出现了一些应用高阶特征波束进行 PIV 定位的方法, 增广伪声强 (AIV)[18-19] 为其中典型的一种方法。

1.2.2　AIV 方法

0 阶模式强度$b_0(kr)$的模值比较大, 进行模式补偿得到的 0 阶特征波束$a_{00}(\tau,k)$相对受噪声影响较小, 可以用$a_{00}(\tau,k)$得到声场幅度的近似值, $s(\tau,k)\simeq\sqrt{4\pi}a_{00}(\tau,k)$, 构造$\widetilde{a}_{nm}(\tau,k,\Theta)=\sqrt{4\pi}a_{00}(\tau,k)\mathrm{Y}_n^{m*}(\Theta)$来逼近$a_{nm}(\tau,k)$, 定义代价函数[18-19]:

$$\Pi(\tau,k,\Theta)=\sum_{n=0}^{N}\sum_{m=-n}^{n}|a_{nm}(\tau,k)-\sqrt{4\pi}a_{00}(\tau,k)\mathrm{Y}_n^{m*}(\Theta)|^2,\tag{9}$$

通过寻找使代价函数$\Pi(\tau,k,\Theta)$最小的Θ作为声源定位结果[18-19]:

$$\Theta_{\mathrm{aiv}}(\tau,k)=\arg\min_{\Theta}\Pi(\tau,k,\Theta).\tag{10}$$

AIV 方法通常采用局部扫描寻找式 (10) 最小值, 将 PIV 定位结果作为扫描初始值, 在初始值附近一定范围内进行搜索。用$a_{00}(\tau,k)$近似声场幅度在噪声、混响的影响下会产生偏差, 且高阶特征波束在低频段易受噪声影响, 使得 AIV 方法的定位性能在低信噪比和高混响环境下会有所下降。

2　时频相关球谐波阶数感知鲁棒声源定位方法

针对低信噪比和高混响环境下 PIV 和 AIV 方法定位性能下降的问题, 本节提出局部时频相关的阶数感知 (TFCOA) 因子以及基于该因子的 TFCOA-PIV 和 TFCOA-AIV 定位方法。本节首先介绍时频相关球谐波阶数感知因子的构造, 并对其在噪声和混响环境下的表现进行理论分析, 最后介绍 TFCOA-PIV 和 TFCOA-AIV 定位方法流程。

2.1　时频相关球谐域阶数感知因子的构造

根据语音信号的稀疏性[25], 若由相邻时频点构成的时频块中的信号主要来自于同一方位, 在噪声影响较小时它们的特征波束之间存在相关性, 可以利用这种相关性将受噪声、混响影响较小的单声源直达声信号占优的时频点块筛选出来用于定位, 提升定位方法的性能。为表达的简洁, 在下文中与时频点有关的变量除第一次出现外将省略(τ,k)。

考虑以时频点(τ,k)作为右下角参考点的时频块, 包含连续J_τ帧, 每一帧连续J_k个频点, 共$J_\tau J_k$个时频点。将其中任一时频点$(\tau-j_\tau,k+j_k)$的n阶特征波束向量$\boldsymbol{a_n}(\tau-j_\tau,k+j_k)$用$\boldsymbol{a_n}^{j_\tau,j_k}(\tau,k)$表示, 其中的特征波束和信号幅度分别用$a_{nm}^{j_\tau,j_k}(\tau,k)$, $s_l^{j_\tau,j_k}(\tau,k)$表示, $j_\tau\in[0,J_\tau-1]$, $j_k\in[0,J_k-1]$。定义:

$$P_n^{j_\tau,j_k}(\tau,k)=\Gamma_n\left\|\boldsymbol{a_n}^{j_\tau,j_k}(\tau,k)\right\|^2,\tag{11}$$

其中, $\Gamma_n=4\pi/(2n+1)$。为衡量参考时频点(τ,k)的n阶特征波束向量$\boldsymbol{a_n}^{0,0}$与$\boldsymbol{a_n}^{j_\tau,j_k}$的相关程度, 定义:

$$H_n^{j_\tau,j_k}(\tau,k)=\Gamma_n\langle\boldsymbol{a_n}^{0,0}(\tau,k),\boldsymbol{a_n}^{j_\tau,j_k}(\tau,k)\rangle,\tag{12}$$

其中, $\langle\boldsymbol{a},\boldsymbol{b}\rangle$为向量$\boldsymbol{a}$与$\boldsymbol{b}$的内积。根据柯西–施瓦茨不等式有$|H_n^{j_\tau,j_k}|\leqslant\Gamma_n\|\boldsymbol{a_n}^{0,0}\|\|\boldsymbol{a_n}^{j_\tau,j_k}\|$, 即$|H_n^{j_\tau,j_k}|\leqslant\sqrt{P_n^{0,0}P_n^{j_\tau,j_k}}$, 当$\boldsymbol{a_n}^{j_\tau,j_k}=\beta^{j_\tau,j_k}\boldsymbol{a_n}^{0,0}$时等号成立, β^{j_τ,j_k}为常数, 即$|H_n^{j_\tau,j_k}|$取值与$\boldsymbol{a_n}^{0,0}$和$\boldsymbol{a_n}^{j_\tau,j_k}$的的相关程度有关。

直达声中含有声源的真实 DOA 信息, 是声源定位的重要信息, 现对只包含直达声信号的情况进行分析。当时频点(τ,k)和$(\tau-j_\tau,k+j_k)$中只含有声源l的直达声信号时, 根据式 (5) 有$a_{nm}^{0,0}=s_l^{0,0}\mathrm{Y}_n^{m*}(\Theta_l)$, $a_{nm}^{j_\tau,j_k}=s_l^{j_\tau,j_k}\mathrm{Y}_n^{m*}(\Theta_l)$, 此时满足条件$\boldsymbol{a_n}^{j_\tau,j_k}=\beta^{j_\tau,j_k}\boldsymbol{a_n}^{0,0}$, $\beta^{j_\tau,j_k}=s_l^{j_\tau,j_k}/s_l^{0,0}$。根据式 (11) 和球谐波函数求和公式[8] 有$P_n^{0,0}=|s_l^{0,0}|^2$, $P_n^{j_\tau,j_k}=|\beta^{j_\tau,j_k}|^2|s_l^{0,0}|^2$, $P_n^{0,0}$, $P_n^{j_\tau,j_k}$与阶数n无关, 即各阶$P_n^{0,0}$, $P_n^{j_\tau,j_k}$相等。根据式 (12) 有$H_n^{j_\tau,j_k}=\beta^{j_\tau,j_k*}|s_l^{0,0}|^2$, 同样与阶数$n$无关, 因此$H_n^{j_\tau,j_k}H_{n-1}^{j_\tau,j_k*}=|\beta^{j_\tau,j_k}|^2|s_l^{0,0}|^4$为与阶数$n$无关的实数, 且$H_n^{j_\tau,j_k}H_{n-1}^{j_\tau,j_k*}=P_n^{0,0}P_n^{j_\tau,j_k}=P_{n-1}^{0,0}P_{n-1}^{j_\tau,j_k}$。因此在$(\tau,k)$和$(\tau-j_\tau,k+j_k)$中只含有单声源直达声信号时, 各阶$P_n^{0,0}$, $P_n^{j_\tau,j_k}$相等, 各阶$H_n^{j_\tau,j_k}H_{n-1}^{j_\tau,j_k*}$相等且等于最大值$P_n^{0,0}P_n^{j_\tau,j_k}$或$P_{n-1}^{0,0}P_{n-1}^{j_\tau,j_k}$。

若时频点(τ,k)和$(\tau-j_\tau,k+j_k)$中含有多个声源的直达声信号时, 不妨假设包含声源l和l'的直达声信

号,则$a_{nm}^{0,0}=s_l^{0,0}\mathrm{Y}_n^{m*}(\Theta_l)+s_{l'}^{0,0}\mathrm{Y}_n^{m*}(\Theta_{l'})$, $a_{nm}^{j_\tau,j_k}=s_l^{j_\tau,j_k}\mathrm{Y}_n^{m*}(\Theta_l)+s_{l'}^{j_\tau,j_k}\mathrm{Y}_n^{m*}(\Theta_{l'})$,若$\Theta_l\neq\Theta_{l'}$且两个声源辐射的语音信号不同时,则由各$\mathrm{Y}_n^{m*}(\Theta_l)$与$\mathrm{Y}_n^{m*}(\Theta_{l'})$之间的差异,以及语音信号随机性导致的$s_l^{j_\tau,j_k}/s_l^{0,0}\neq s_{l'}^{j_\tau,j_k}/s_{l'}^{0,0}$,使得$\boldsymbol{a}_{\boldsymbol{n}}^{j_\tau,j_k}\neq\beta^{j_\tau,j_k}\boldsymbol{a}_{\boldsymbol{n}}^{0,0}$。此时$P_n^{j_\tau,j_k}$和$H_n^{j_\tau,j_k}$可分别表示为

$$P_n^{j_\tau,j_k}=\left|s_l^{j_\tau,j_k}\right|^2+\left|s_{l'}^{j_\tau,j_k}\right|^2+2\mathrm{Re}\left(s_l^{j_\tau,j_k}s_{l'}^{j_\tau,j_k*}\right)\mathrm{P}_n(\cos(\Theta_{ll'})),\tag{13}$$

$$H_n^{j_\tau,j_k}=s_l^{0,0}s_l^{j_\tau,j_k*}+s_{l'}^{0,0}s_{l'}^{j_\tau,j_k*}+(s_l^{0,0}s_{l'}^{j_\tau,j_k*}+s_{l'}^{0,0}s_l^{j_\tau,j_k*})\cdot\mathrm{P}_n(\cos(\Theta_{ll'})),\tag{14}$$

其中,$\mathrm{P}_n(\cdot)$为n阶勒让德多项式,$\Theta_{ll'}$为声源l和l'的角度间距。从式(13)和式(14)中可以看出,此时各阶$P_n^{j_\tau,j_k}$, $H_n^{j_\tau,j_k}$的值与$\mathrm{P}_n(\cos(\Theta_{ll'}))$有关,而当$\Theta_l\neq\Theta_{l'}$时各阶$\mathrm{P}_n(\cos(\Theta_{ll'}))$的值存在差异,使得各阶$P_n^{j_\tau,j_k}$, $H_n^{j_\tau,j_k}$不再相等,同理各阶$P_n^{0,0}$也不再相等。可以推广到声源个数大于2的情况。

另外在实际室内环境中存在噪声和混响,在噪声的随机性、各阶$b_n(kr_a)$的差异、不同混响信号传输路径的不同、语音信号的随机性等因素的影响下,综合导致$\boldsymbol{a}_{\boldsymbol{n}}^{j_\tau,j_k}\neq\beta^{j_\tau,j_k}\boldsymbol{a}_{\boldsymbol{n}}^{0,0}$,各阶$P_n^{0,0}$, $P_n^{j_\tau,j_k}$, $H_n^{j_\tau,j_k}$, $H_{n-1}^{j_\tau,j_k}$不再相等。因此在多声源混响噪声环境下,$H_n^{j_\tau,j_k}H_{n-1}^{j_\tau,j_k*}$实部与最大值之间的差距,$P_n^{0,0}$, $P_n^{j_\tau,j_k}$与$P_{n-1}^{0,0}$, $P_{n-1}^{j_\tau,j_k}$之间的差异,反映了$\boldsymbol{a}_{\boldsymbol{n}}^{0,0}$, $\boldsymbol{a}_{\boldsymbol{n-1}}^{0,0}$, $\boldsymbol{a}_{\boldsymbol{n}}^{j_\tau,j_k}$, $\boldsymbol{a}_{\boldsymbol{n-1}}^{j_\tau,j_k}$受噪声、混响以及多个声源间相互干扰的影响程度。

若时频点(τ,k)和$(\tau-j_\tau,k+j_k)$中包含的信号以经同一方位反射的混响信号为主,则可能会出现误选情况,为减小这种可能性,考虑时频块中每个时频点的特征波束向量与参考时频点特征波束向量之间的相关程度。将除$(j_\tau,j_k)=(0,0)$外的$H_n^{j_\tau,j_k}$和$P_n^{j_\tau,j_k}$表示为向量形式,$\boldsymbol{H_n}(\tau,k)=[H_n^{0,1},\cdots,H_n^{j_\tau,j_k},\cdots,H_n^{J_\tau-1,J_k-1}]^{\mathrm{T}}$, $P_n(\tau,k)=[P_n^{0,1},\cdots,P_n^{j_\tau,j_k},\cdots,P_n^{J_\tau-1,J_k-1}]^{\mathrm{T}}$,同理可以得到$\boldsymbol{H_{n-1}}(\tau,k)$和$\boldsymbol{P_{n-1}}(\tau,k)$。定义时频相关阶数感知因子(TFCOA)为

$$K_n(\tau,k)=\frac{2\mathrm{Re}(\langle\boldsymbol{H_n}(\tau,k),\boldsymbol{H_{n-1}}(\tau,k)\rangle)}{\sum\left(P_n^{0,0}(\tau,k)\boldsymbol{P_n}(\tau,k)+P_{n-1}^{0,0}(\tau,k)\boldsymbol{P_{n-1}}(\tau,k)\right)}.\tag{15}$$

将式(15)中分子的内积展开,由前面的分析可知,分子中的任一项都小于分母中对应的项,即$2\mathrm{Re}[H_n^{j_\tau,j_k}H_{n-1}^{j_\tau,j_k*}]\leqslant2\sqrt{P_n^{0,0}P_n^{j_\tau,j_k}}\sqrt{P_{n-1}^{0,0}P_{n-1}^{j_\tau,j_k}}$,而$2\sqrt{P_n^{0,0}P_n^{j_\tau,j_k}}\sqrt{P_{n-1}^{0,0}P_{n-1}^{j_\tau,j_k}}\leqslant P_n^{0,0}P_n^{j_\tau,j_k}+P_{n-1}^{0,0}P_{n-1}^{j_\tau,j_k}$,因此$|K_n|\leqslant1$, K_n反映了时频块中其他时频点的特征波束向量与参考时频点(τ,k)的特征波束向量的相关程度。只有当时频块中只包含源自同一方位的信号时,每一时频点都有$P_n^{0,0}P_n^{j_\tau,j_k}=P_{n-1}^{0,0}P_{n-1}^{j_\tau,j_k}$, $\boldsymbol{a}_{\boldsymbol{n}}^{j_\tau,j_k}=\beta^{j_\tau,j_k}\boldsymbol{a}_{\boldsymbol{n}}^{0,0}$,此时$K_n=1$。

2.2 时频相关阶数感知因子抗噪抗混响能力分析

2.2.1 抗混响能力分析

为简化分析和突出时频相关阶数感知因子在多声源和混响条件下的表现,本节只考虑多声源和混响,忽略噪声的影响。

先考虑K_n分子中的任一项$\mathrm{Re}[H_n^{j_\tau,j_k}H_{n-1}^{j_\tau,j_k*}]$。在多声源和混响环境下,令$a_{nm}^{0,0}=\sum_l s_l^{0,0}\mathrm{Y}_n^{m*}(\Theta_l)$, $a_{nm}^{j_\tau,j_k}=\sum_d s_d^{j_\tau,j_k}\mathrm{Y}_n^{m*}(\Theta_d)$。根据式(12)和球谐波函数求和公式,可得

$$\mathrm{Re}[H_n^{j_\tau,j_k}H_{n-1}^{j_\tau,j_k*}]=\sum_l\sum_d\sum_{l'}\sum_{d'}\mathrm{Re}(s_l^{0,0}s_d^{j_\tau,j_k*}s_{l'}^{0,0*}s_{d'}^{j_\tau,j_k})\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l',d'},\tag{16}$$

其中,$\mathrm{P}_n^{l,d}=\mathrm{P}_n(\cos(\Theta_{ld}))$, $\mathrm{P}_{n-1}^{l',d'}=\mathrm{P}_{n-1}(\cos(\Theta_{l'd'}))$, Θ_{ld}为声源l和声源d的DOA角度间距。式(16)中除$l=l'$, $d=d'$以外的每一项都存在一个对称项$\mathrm{Re}(s_{l'}^{0,0}s_{d'}^{j_\tau,j_k*}s_l^{0,0*}s_d^{j_\tau,j_k})\mathrm{P}_{n-1}^{l,d}\mathrm{P}_n^{l',d'}$,将两项合并在一起,$\mathrm{Re}(s_l^{0,0}s_d^{j_\tau,j_k*}s_{l'}^{0,0*}s_{d'}^{j_\tau,j_k})(\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l',d'}+\mathrm{P}_{n-1}^{l,d}\mathrm{P}_n^{l',d'})$。即$\mathrm{Re}[H_n^{j_\tau,j_k}H_{n-1}^{j_\tau,j_k*}]$由$(\tau,k)$中声源和$(\tau-j_\tau,k+j_k)$中声源的DOA角度间距的函数$\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l',d'}+\mathrm{P}_{n-1}^{l,d}\mathrm{P}_n^{l',d'}$和每个声源的幅度共同决定,而$\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l',d'}+\mathrm{P}_{n-1}^{l,d}\mathrm{P}_n^{l',d'}$随声源DOA角度间距$\Theta_{ld}$, $\Theta_{l'd'}$的变化情况如图1所示。

从图1中可以看出,只有当Θ_{ld}, $\Theta_{l'd'}$为0或2π时,即声源l和声源d、声源l'和声源d'的DOA相同时,$\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l',d'}+\mathrm{P}_{n-1}^{l,d}\mathrm{P}_n^{l',d'}$取最大值2,之后随着声源角度间距$\Theta_{ld}$, $\Theta_{l'd'}$的增大$\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l',d'}+\mathrm{P}_{n-1}^{l,d}\mathrm{P}_n^{l',d'}$逐渐减小,阶数$n$越大减小的速度越快,虽然$n>1$时$\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l',d'}+\mathrm{P}_{n-1}^{l,d}\mathrm{P}_n^{l',d'}$下降到0附近开始波动,但是波动的幅度较小,在$\Theta_{ld}$, $\Theta_{l'd'}$增大到π时达到最小值−2。对于式(16)中的$l=l'$, $d=d'$项,此时Θ_{ld}, $\Theta_{l'd'}$, $\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l,d}$按图1中对角线的趋势变化,幅度为其一半,波动幅度更小。从整体上看$\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l',d'}+\mathrm{P}_{n-1}^{l,d}\mathrm{P}_n^{l',d'}$随声源角度间距的增大是减小的趋势。即当$(\tau,k)$和$(\tau-j_\tau,k+j_k)$中声源DOA角度间距增大时,$\mathrm{Re}[H_n^{j_\tau,j_k}H_{n-1}^{j_\tau,j_k*}]$的取值呈减小趋势,反映了$(\tau,k)$和$(\tau-j_\tau,k+j_k)$中声源信息的相似程度。

考虑一种最坏情况,时频块中任一时频点都包含相同方位的直达声和混响信号,且任一时频点$(\tau-j_\tau,k+j_k)$满足条件$\boldsymbol{a}_{\boldsymbol{n}}^{j_\tau,j_k}=\beta^{j_\tau,j_k}\boldsymbol{a}_{\boldsymbol{n}}^{0,0}$。根据式(11)和式(12)有$P_n^{j_\tau,j_k}=|\beta^{j_\tau,j_k}|^2P_n^{0,0}$, $H_n^{j_\tau,j_k}=\beta^{j_\tau,j_k*}P_n^{0,0}$, $P_{n-1}^{j_\tau,j_k}=|\beta^{j_\tau,j_k}|^2P_{n-1}^{0,0}$, $H_{n-1}^{j_\tau,j_k}=\beta^{j_\tau,j_k*}P_{n-1}^{0,0}$,代入式(15),有

$$O_n^{n-1}(\tau,k)=\frac{2P_n^{0,0}P_{n-1}^{0,0}}{(P_n^{0,0})^2+(P_{n-1}^{0,0})^2}.\tag{17}$$

从式(17)可以看出,K_n在最坏情况下退化为单时频点的阶数感知(OA)检测因子[23] O_n^{n-1},此时O_n^{n-1}的值由$P_n^{0,0}$和$P_{n-1}^{0,0}$的差距决定。由于语音信号具有随

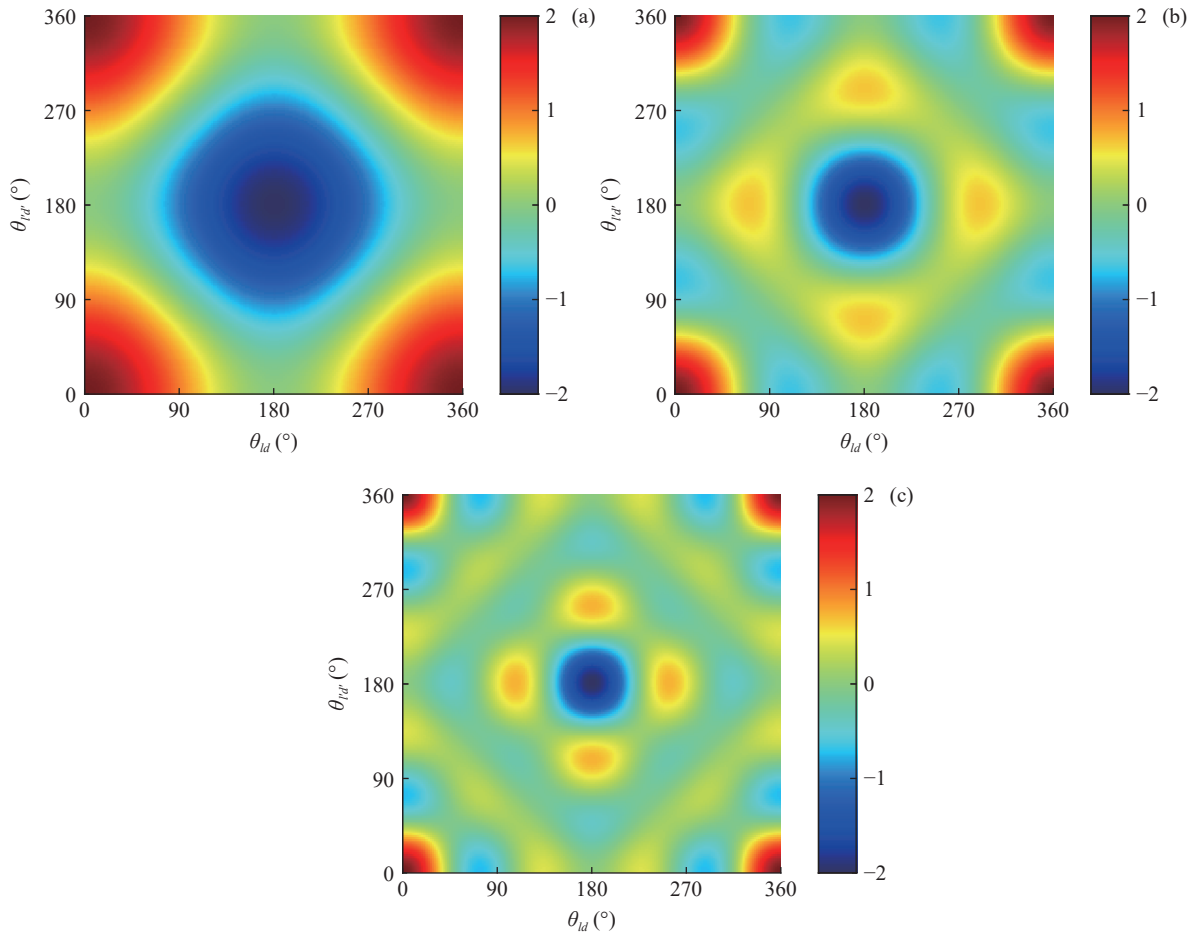

图 1　各阶 $\mathrm{P}_n^{l,d}\mathrm{P}_{n-1}^{l',d'}+\mathrm{P}_{n-1}^{l,d}\mathrm{P}_n^{l',d'}$ 随声源 DOA 角度间距变化情况 (a) $n=1$; (b) $n=2$; (c) $n=3$

机性, 要使每个时频点都满足条件 $\boldsymbol{a}_n^{j_\tau,j_k}=\beta^{j_\tau,j_k}\boldsymbol{a}_n^{0,0}$ 几乎是不可能的, 除非时频块中每个时频点只含有来自同一声源的直达声信号, 此时 $P_n^{0,0}=P_{n-1}^{0,0}$, $K_n=O_{n,n-1}=1$。因此多声源混响环境下 $K_n<O_{n,n-1}$, 即 TFCOA 检测因子对多个声源和混响的区分能力优于 OA 检测因子, 当 K_n 值较大时说明时频块中的信号主要为来自同一声源的直达声信号。

2.2.2　时频相关阶数感知因子抗噪声能力分析

本节分析时频相关阶数感知因子在噪声环境下的表现, 为简化分析只考虑噪声, 忽略语音和混响信号。设噪声的均值为 μ, 方差为 σ^2, 此时 K_n 为

$$K_n=\frac{\dfrac{2\mathrm{Re}\left(\sum\limits_{j_\tau,j_k}\Gamma_n\Gamma_{n-1}\langle\boldsymbol{n}_n^{0,0},\boldsymbol{n}_n^{j_\tau,j_k}\rangle\langle\boldsymbol{n}_{n-1}^{0,0},\boldsymbol{n}_{n-1}^{j_\tau,j_k}\rangle\right)}{|b_n(kr_a)|^2|b_{n-1}(kr_a)|^2}}{\dfrac{\sum\limits_{j_\tau,j_k}\Gamma_n^2\|\boldsymbol{n}_n^{0,0}\|^2\|\boldsymbol{n}_n^{j_\tau,j_k}\|^2}{|b_n(kr_a)|^4}+\dfrac{\sum\limits_{j_\tau,j_k}\Gamma_{n-1}^2\|\boldsymbol{n}_{n-1}^{0,0}\|^2\|\boldsymbol{n}_{n-1}^{j_\tau,j_k}\|^2}{|b_{n-1}(kr_a)|^4}},\tag{18}$$

其中, $\boldsymbol{n}_n^{0,0}$ 和 $\boldsymbol{n}_n^{j_\tau,j_k}$ 分别为时频点 (τ,k) 和 $(\tau-j_\tau,k+j_k)$ 的第 n 阶噪声球谐波系数构成的向量。从上式可以看出, 当时频点块中只包含噪声时, K_n 的值由两方面决定, 一方面由模式强度补偿 $1/(|b_n(kr_a)|^2|b_{n-1}(kr_a)|^2)$ 与 $1/|b_n(kr_a)|^4$, $1/|b_{n-1}(kr_a)|^4$ 的差距决定, 这与 OA 方法是相同的, 在本文中不再讨论; 另一方面由参考时频点 (τ,k) 的 $\boldsymbol{n}_n^{0,0}$, $\boldsymbol{n}_{n-1}^{0,0}$ 与时频点 $(\tau-j_\tau,k+j_k)$ 的 $\boldsymbol{n}_n^{j_\tau,j_k}$, $\boldsymbol{n}_{n-1}^{j_\tau,j_k}$ 的内积决定。由于噪声是随机的, 因此采用统计平均观察 K_n 在纯噪声环境下的取值情况, 为分析方便, 将 K_n 的分子分母分开讨论。

先考虑 K_n 的分子中的任一时频点 $(\tau-j_\tau,k+j_k)$ 对应的项, 由于 $b_n(kr_a)$ 为确定函数, 省略 $b_n(kr_a)$ 不影响对均值的讨论。当球阵的传声器分布为均匀分布、高斯分布和等角度分布时, 有

$$2\mathrm{Re}[\Gamma_n\Gamma_{n-1}\mathrm{E}(\langle\boldsymbol{n}_n^{0,0},\boldsymbol{n}_n^{j_\tau,j_k}\rangle\langle\boldsymbol{n}_{n-1}^{0,0},\boldsymbol{n}_{n-1}^{j_\tau,j_k}\rangle)]=0.\tag{19}$$

关于式 (19) 的证明见附录 A。即式 (18) 分子中任一时频点对应项的均值均为 0, 将式 (18) 中 K_n 的分子和分母同除以总项数 $(J_\tau J_k-1)$, 则分子是 $(J_\tau J_k-1)$ 项随机变量的平均, 接近于随机变量平均值 0, 即纯噪声时 K_n 的分子取值在 0 附近。

对于式 (18) 中的分母, 其均值不为 0, 证明见附录 B。因此当只有噪声时, K_n 取值在 0 附近, 且与频

率无关, 因此 K_n 在整个频段都具有衡量噪声大小的能力, 弥补了 OA 检测因子在高频段衡量噪声大小能力下降的问题。在信号与噪声共存时, 若 K_n 值较大则说明该时频块中语音信号占优, 若 K_n 值较小则说明噪声占优, 可以舍去。

2.3 球谐域时频相关阶数感知伪声强声源定位方法

将时频相关阶数感知因子 (TFCOA) 与 PIV 和 AIV 方法相结合, 形成鲁棒的伪声强声源定位方法, 分别记为 TFCOA-PIV 方法、TFCOA-AIV 方法, 相应的声源定位流程图如图 2 所示, 具体步骤如下:

(1) 将阵列采集的声压信号转换到球谐域。把采集的声压信号 $\boldsymbol{p}_{0,0}(\tau,k)$, $\cdots$, $\boldsymbol{p}_{J_\tau,J_k}(\tau,k)$ 代入到式 (1) 中进行离散球傅里叶变换, 再进行模式强度补偿得到特征波束向量 $\boldsymbol{a}_{\boldsymbol{nm}}^{0,0}(\tau,k)$, $\cdots$, $\boldsymbol{a}_{\boldsymbol{nm}}^{J_\tau,J_k}(\tau,k)$。其中, $\boldsymbol{p}_{j_\tau,j_k}(\tau,k)=[p_1^{j_\tau,j_k},p_2^{j_\tau,j_k},\cdots,p_Q^{j_\tau,j_k}]^{\mathrm{T}}$, 为 Q 个传声器采集声压信号组成的向量, $p_q^{j_\tau,j_k}=p_q(\tau-j_\tau,k+j_k)$ 为第 q 个传声器采集的声压信号; $\boldsymbol{a}_{\boldsymbol{nm}}^{j_\tau,j_k}$ 为时频点 $(\tau-j_\tau,k+j_k)$ 所有特征波束构成的向量, $\boldsymbol{a}_{\boldsymbol{nm}}^{j_\tau,j_k}=[a_{00}^{j_\tau,j_k},a_{1(-1)}^{j_\tau,j_k},\cdots,a_{NN}^{j_\tau,j_k}]^{\mathrm{T}}$。

(2) 计算 $K_n(\tau,k)$ 并进行时频点和可靠特征波束筛选。将时频块各时频点的 N 阶和 $N-1$ 阶的特征波束向量代入到式 (15) 中计算 $K_N(\tau,k)$, 并与阈值 T_N 比较, 超过 T_N 则可靠特征波束阶数 $\kappa=N$; 否则计算次高阶的 $K_{N-1}(\tau,k)$, 判断是否超过阈值 T_{N-1}, 超过则 $\kappa=N-1$; 否则处理下一阶, 以此类推, 直到 $K_2(\tau,k)$, 若 $K_2(\tau,k)$ 超过阈值 T_2 则 $\kappa=2$, 否则丢弃该时频点。

(3) 计算局部时频相关矩阵并进行特征值分解。由筛选出的时频块的 κ 阶可靠特征波束构成局部时频相关矩阵[11,21,22]:

$$\boldsymbol{R}_a(\tau,k)=\frac{1}{J_\tau J_k}\sum_{j_k=0}^{J_k-1}\sum_{j_\tau=0}^{J_\tau-1}\boldsymbol{a}_{\boldsymbol{\kappa m}}^{j_\tau,j_k}(\boldsymbol{a}_{\boldsymbol{\kappa m}}^{j_\tau,j_k})^{\mathrm{H}}, \tag{20}$$

其中, $\boldsymbol{a}_{\boldsymbol{\kappa m}}^{j_\tau,j_k}=[a_{00}^{j_\tau,j_k},a_{1(-1)}^{j_\tau,j_k},\cdots,a_{\kappa\kappa}^{j_\tau,j_k}]^{\mathrm{T}}$。对 $\boldsymbol{R}_a(\tau,k)$ 进行特征值分解, 可以表示为[21-22]

$$\boldsymbol{R}_a(\tau,k)=\boldsymbol{U}_S\boldsymbol{\Sigma}_S\boldsymbol{U}_S{}^{\mathrm{H}}+\boldsymbol{U}_V\boldsymbol{\Sigma}_V\boldsymbol{U}_V{}^{\mathrm{H}}, \tag{21}$$

其中, $\boldsymbol{U}_S$ 为信号特征向量构成的信号子空间矩阵, $\boldsymbol{U}_V$ 为噪声特征向量构成的噪声子空间矩阵, $\boldsymbol{\Sigma}_S$ 和 $\boldsymbol{\Sigma}_V$ 为由信号和噪声特征值构成的对角矩阵。

(4) PIV 定位和 AIV 定位。选取 $\boldsymbol{U}_S$ 中最大特征值对应的特征向量 $\widehat{\boldsymbol{a}}_{\boldsymbol{\kappa m}}(\tau,k)=[\widehat{a}_{00}(\tau,k),\widehat{a}_{1(-1)}(\tau,k),\cdots,\widehat{a}_{\kappa\kappa}(\tau,k)]^{\mathrm{T}}$, 将其中的 0 阶和 1 阶代入到式 (6)—式 (8) 中得到 TFCOA-PIV 定位结果 $\Theta_{\mathrm{P}}(\tau,k)$; 然后把 $\widehat{\boldsymbol{a}}_{\boldsymbol{\kappa m}}(\tau,k)$ 代入到式 (9) 中, 以 $\Theta_{\mathrm{P}}(\tau,k)$ 作为空间扫描初始值, 根据式 (10) 得到 TFCOA-AIV 定位结果 $\Theta_{\mathrm{A}}(\tau,k)$。

(5) 由平滑处理过的直方图给出声源定位结果。将筛选出的每个时频块的 TFCOA-PIV 或 TFCOA-AIV 的定位结果构成直方图, 并进行平滑处理[18,22], 将其最高 L 个峰对应的角度坐标作为 L 个声源的定位结果。

2.4 计算量分析

本节对所提方法的计算量进行分析, 并与 PIV[15]、AIV[18-19]、OA-AIV[23]、DPD-AIV[11,18] 方法的计算量进行对比。其中 DPD-AIV 方法是 DPD 与 AIV 方法的结合, 将通过 DPD 检测的最大特征值对应的特征向量代入到 AIV 方法中进行定位。计算量由定位每个时频点需要的实数乘法次数进行衡量, 其中复数乘法按照 4 次实数乘法计算, $|\cdot|^2$ 按照两次实数乘法计算。

各对比方法定位一个时频点所需要的实数乘法次数如表 1 所示。其中 PIV、AIV、OA-AIV、DPD-AIV 方法的计算量参考了文献 [18, 23]。TFCOA 检测因子和 OA 检测因子的实数乘法次数按照计算全部 $(N-1)$ 个因子所需的计算量统计, 其中 P_{OA}, P_{DPD}, P_{TFC} 分别表示 OA、DPD 和 TFCOA 检测因子选中该时频点的概率; G_{L} 表示局部扫描的平均栅格点数, $\bar{\kappa}$ 和 $\bar{\kappa}_{\mathrm{OA}}$ 分别为 TFCOA 和 OA 检测因子筛选出的可靠特征波束的平均阶数, 其中 $\bar{\kappa}=\sum_{n=2}^{N}nP_{\mathrm{TFC}}^{n}$, $\bar{\kappa}_{\mathrm{OA}}=\sum_{n=2}^{N}nP_{\mathrm{OA}}^{n}$, P_{TFC}^{n} 和 P_{OA}^{n} 分别为 TFCOA 和 OA 检测因子筛选出的最大阶数为 n 的可靠特征波束百分比。

接下来测试各对比方法定位1 s时长信号所需的运行时间, 如表 2 所示。运行时间为各对比定位方法运行 100 次的平均值, 实验参数设置与仿真实验一致。实验所用计算机的处理器为英特尔 i7-10875H, 内存为 24 G。

从表 1 和表 2 中可以看出, PIV 方法具有闭式解, 其计算量是最低的, 运行时间最短; 而 TFCOA-PIV 需要对通过 TFCOA 检测的时频块进行子空间

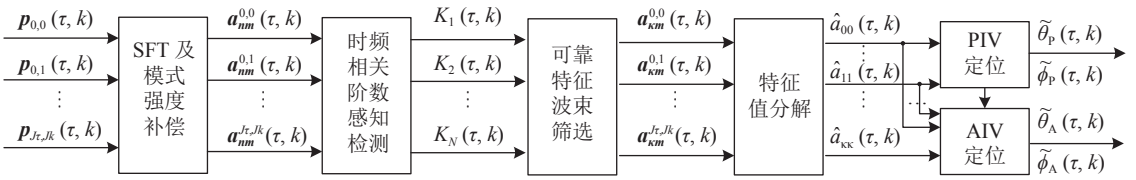

图 2 所提定位方法流程图

表 1　各对比方法定位一个时频点所需的实数乘法次数

对比方法	分阶筛选	计算相关矩阵	特征值分解	预定位	空间扫描
PIV	0	0	0	48	0
TFCOA-PIV	$[6(N+1)^2+8N]J_\tau J_k$	$4(\bar{\kappa}+1)^2 J_\tau J_k P_{\mathrm{TFC}}$	$3(\bar{\kappa}+1)^6 P_{\mathrm{TFC}}$	$48P_{\mathrm{TFC}}$	0
AIV	0	0	0	48	$8(N+1)^2 G_{\mathrm{L}}$
OA-AIV	$4(N+1)^2+6$	0	0	$48P_{\mathrm{OA}}$	$8(\bar{\kappa}_{OA}+1)^2 G_{\mathrm{L}} P_{\mathrm{OA}}$
DPD-AIV	0	$4(N+1)^2 J_\tau J_k$	$3(N+1)^6$	$48P_{\mathrm{DPD}}$	$8(N+1)^2 G_{\mathrm{L}} P_{\mathrm{DPD}}$
TFCOA-AIV	$[6(N+1)^2+8N]J_\tau J_k$	$4(\bar{\kappa}+1)^2 J_\tau J_k P_{\mathrm{TFC}}$	$3(\bar{\kappa}+1)^6 P_{\mathrm{TFC}}$	$48P_{\mathrm{TFC}}$	$8(\bar{\kappa}+1)^2 G_{\mathrm{L}} P_{\mathrm{TFC}}$

表 2　各对比方法定位 1s 信号所需的平均运行时间

对比方法	PIV	TFCOA-PIV	AIV	OA-AIV	DPD-AIV	TFCOA-AIV
运行时间 (s)	0.2375	0.5779	1.3149	0.8265	0.7675	0.6097

分解, 因此计算量和运行时间高于 PIV 方法。对于采用高阶特征波束定位的方法, 其中 AIV 方法需要对每个时频点进行局部空间扫描搜索, 而 OA-AIV 方法则只需要对通过 OA 检测的时频点进行 AIV 定位, 其计算量约是 AIV 方法的 P_{OA}; TFCOA-AIV 方法虽然需要对通过 TFCOA 检测的时频块进行相关矩阵计算和特征值分解, 但是其筛选时频点的概率 P_{TFC} 比 P_{OA} 低许多, 使得其计算量低于 OA-AIV 方法, 定位 1 s 时长信号所需时间比 OA-AIV 方法减小了约 25%; 而 DPD-AIV 方法需要对每个时频块进行子空间分解, 因此在 $P_{\mathrm{DPD}}=P_{\mathrm{TFC}}$ 时其计算量高于 TFCOA-AIV 方法, 如表 1 和表 2 所示。

3　仿真实验

本节通过仿真实验测试所提声源定位方法的性能, 测试多声源环境下定位性能随信噪比、混响时间以及声源角度间距变化的情况, 并与 PIV、AIV、OA-AIV、DPD-AIV 方法进行对比, 主要测试低信噪比、高混响条件下的性能表现, 以及阵列存在增益、相位误差时对定位性能的影响。

3.1　仿真参数设置

房间的尺寸设置为 6 m × 5 m × 3 m, 阵列中心位于 4.4 m × 2.5 m × 1.5 m, 阵列中心到每个声源的距离均为 1 m。阵列为刚性球面阵, 半径 4.2 cm, 表面均匀分布放置 32 个传声器。声源个数为 3, 放置在 48 个不同的位置进行测试, 其中声源 1 的方位角以 30°递增分别放置于 $\{0°,30°,\cdots,300°,330°\}$, 俯仰角位于 $\{60°,80°,100°,120°\}$, 共 48 种组合; 声源 2 和声源 3 的俯仰角与声源 1 相同, 相互之间方位角间隔 30°。每个位置测试 10 组语音信号, 共进行 $48\times10=480$ 次蒙特卡罗实验。这 10 组语音信号从 TIMIT 语音库中随机挑选, 每组信号由 3 个不同人的语音组成, 从中截取 1 s 作为测试信号, 信号采样频率为 16 kHz。短时傅里叶变换的帧长为 256, 帧移为 50%, 采用汉宁窗函数进行分帧。最大球谐波阶数 $N=3$, 为减小球傅里叶变换的空间混叠和模式强度补偿对低频噪声过度放大的影响, 将参与定位的信号频率范围设置为 400~5000 Hz。房间混响由基于像−源法的刚性球室内声学单位冲激响应[26] 产生。噪声为高斯白噪声, 不同阵元之间噪声互不相关。

TFCOA 因子的筛选阈值按照筛选出总时频点的 10% 设置, 其中三阶阈值为固定值 $T_3=0.3$, 若三阶筛选出的时频点不足总时频点的 10% 时, 则按照 K_2 从大到小的顺序从剩余时频点中筛选补足, 因此二阶阈值 T_2 为动态的。DPD-AIV 方法的阈值也是按照筛选出总时频点的 10% 设置[11,22]。OA-AIV[23] 方法的三阶阈值为 0.75, 二阶阈值为 0.8, 与文献 [23] 一致。所有对比方法的定位结果由每个时频点定位结果形成的直方图经高斯平滑处理[18,22] 后的最高 3 个峰的角度坐标给出, 其中高斯函数的方差 $\sigma=4$。AIV、OA-AIV 和 TFCOA-AIV 的区域扫描半径为 10°, 扫描栅格的俯仰角和方位角间距均为 1°。在参考文献 [11, 18, 21, 22] 中关于 J_τ, J_k 设置和大量仿真测试的基础上, 将 J_τ 设置为 3, J_k 设置为 7。

声源定位误差用估计角度单位矢量与实际角度单位矢量之间的夹角计算, 若估计的声源 DOA 为 $\{\widehat{\theta},\widehat{\phi}\}$, 其单位矢量为 $\widehat{\boldsymbol{u}}=[\sin\widehat{\theta}\cos\widehat{\phi},\sin\widehat{\theta}\sin\widehat{\phi},\cos\widehat{\theta}]^{\mathrm{T}}$; 声源真实 DOA 为 $\{\theta,\phi\}$, 其单位矢量 $\boldsymbol{u}=[\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta]^{\mathrm{T}}$; $\boldsymbol{u}$ 和 $\widehat{\boldsymbol{u}}$ 之间的夹角为[8]

$$\varepsilon=\arccos(\boldsymbol{u}^{\mathrm{T}}\widehat{\boldsymbol{u}}). \tag{22}$$

3.2　性能指标

为衡量定位方法的性能采用了两种指标: 平均

声源检测个数[11,21,22]用来衡量定位方法的鲁棒性, 均方根误差用来衡量定位方法的精度。

(1) 平均声源检测个数: 将每次仿真实验的声源定位结果代入到式 (22) 中计算每个声源的估计误差, 若估计误差小于 15°, 则认为检测出声源, 否则认为没有检测出声源。平均声源检测个数由每次实验检测出的声源个数的均值给出, 定义为

$$\bar{I}_{\mathrm{D}} = \frac{1}{J_{\mathrm{mon}}} \sum_{j=1}^{J_{\mathrm{mon}}} L_{\mathrm{D}}^{j}, \tag{23}$$

其中, J_{mon} 为总的蒙特卡罗实验次数, L_{D}^{j} 为第 j 次实验检测出的声源个数。

(2) 均方根误差: 将成功检测出声源的定位结果代入到式 (22) 中计算估计误差, 根据估计误差可以得到均方根误差, 有

$$\mathrm{RMSE} = \sqrt{\frac{1}{F_{\mathrm{D}}} \sum_{j=1}^{J_{\mathrm{mon}}} \sum_{l=1}^{L_{D}^{j}} (\arccos \boldsymbol{u}_{j,l}^{\mathrm{T}} \widehat{\boldsymbol{u}}_{j,l})^2}, \tag{24}$$

其中, $F_{\mathrm{D}} = \bar{I}_{\mathrm{D}} J_{\mathrm{mon}}$ 为成功检测出的声源总数, $\widehat{\boldsymbol{u}}_{j,l}$ 为第 j 次实验检测出的第 l 个声源的估计 DOA 单位矢量, $\boldsymbol{u}_{j,l}$ 为其对应的真实 DOA 单位矢量。

3.3 信噪比对所提方法定位性能的影响

本节测试所提方法在不同信噪比条件下的性能表现, 信噪比设置为 $\mathrm{SNR} = \{10, 15, 20, 25, 30\}$ dB, 混响时间设置为 $T_{60} = 0.7$ s保持不变。各对比声源定位方法的均方根误差和平均声源检测个数随信噪比变化情况如图 3 所示。

从图 3 中可以看出在不同信噪比条件下, TFCOA-AIV 方法的均方根误差和平均声源检测个数在所有对比方法中的表现是最好的, 且信噪比越低优势越明显, 反映出 TFCOA-AIV 方法相比其他对比方法具有更强的抗噪能力。其中 TFCOA-AIV 方法与 AIV 和 OA-AIV 方法相比在均方根误差上分别有 1.3°~2.1°和 1.1°~1.6°的改善, 定位精度提升比较明显。在平均声源检测个数上, TFCOA-AIV 在 10 dB 时平均能检测出 2.99 个声源, 相比 AIV、OA-AIV 方法分别提升了 8% 和 4.5%。TFCOA-AIV 在信噪比增大到 20 dB 以后能够检测出全部声源, 而 AIV 和 OA-AIV 方法在信噪比为 30 dB 时也无法检测出全部声源, TFCOA-AIV 具有更高的鲁棒性。TFCOA-AIV 与 DPD-AIV 方法相比, 在均方根误差上约有 0.3°~0.5°降低, 在信噪比为 10 dB 时平均声源检测个数提升了约 2%。

TFCOA-PIV 方法与 PIV 方法相比, 在均方根误差上有 2.7°~2.8°的降低, 甚至在信噪比为 10 dB 时均方根误差低于采用高阶特征波束定位的 AIV 方法; 在平均声源检测个数上, 仅次于 TFCOA-AIV 方法, 优于其他对比方法, 相比 PIV 有 26%~21% 的提升。综合可以看出 TFCOA 因子具有较强的抗噪能力, 能够将受噪声影响较小的可靠特征波束筛选出来。

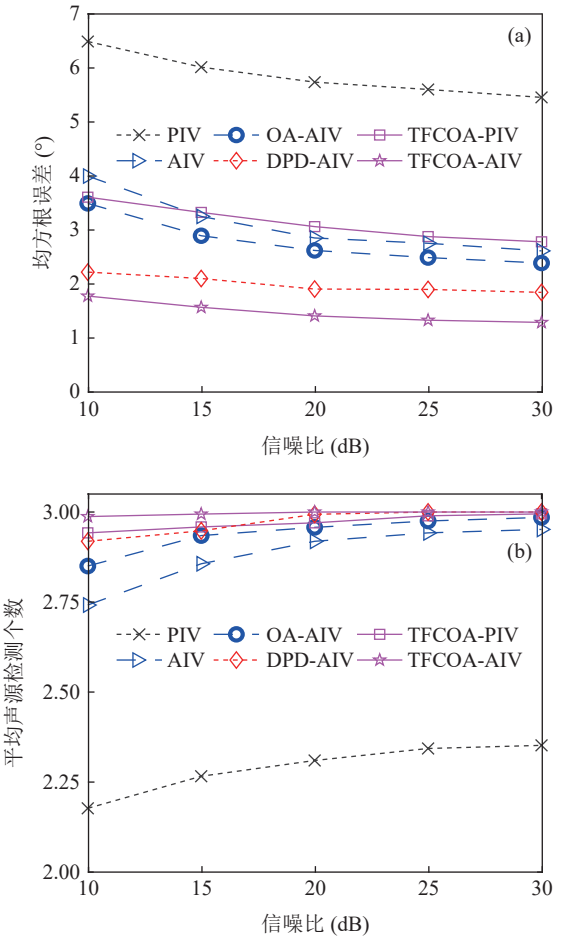

图 3 各对比方法在不同信噪比条件下的均方根误差和平均声源检测个数 (a) 均方根误差; (b) 平均声源检测个数

TFCOA 因子的时频点筛选和定位效果如图 4 所示。图 4 为信噪比为 10 dB 时一次仿真实验的 OA-AIV、DPD-AIV、TFCOA-AIV 方法筛选出的各声源时频点分布及纯净语音语谱图。图中按照筛选出的时频点的定位结果归属到每个声源, 若时频点的定位结果与某声源真实 DOA 的误差在 15°以内则认为属于该声源。图中用不同颜色表示不同声源的时频点, 其中蓝色为声源 1、紫色为声源 2、红色为声源 3, 绿色为不属于任何声源的误选时频点, 其中声源 2 能量主要位于低频部分。从图 4(a) 中可以看出, OA-AIV 筛选出的时频点中误选时频点较多, 主要位于中高频部分, 从中反映出 OA 方法主要针对低频噪声, 对于高频噪声、混响和多声源相互干扰的鉴别能力有限。从图 4(b) 和图 4(c) 中可以看出, TFCOA-

AIV 和 DPD-AIV 方法的误选时频点很少, 反映出 TFCOA 检测因子和 DPD 检测因子能够有效地将受噪声影响较小和单声源直达声占优的时频点筛选出来用于定位。但是从图 4(b) 中可以看出, DPD 检测因子筛选出的时频点中属于声源 2 的时频点较少, 原因是声源 2 主要位于低频段, 其高阶特征波束易受噪声影响, 很难通过 DPD 检测。而 TFCOA 检测因子采用分阶处理, 在低频段无法筛选出高阶特征波束的情况下, 将受噪声污染程度较小的低阶特征波束筛选出来用于声源定位, 尽可能地保留了声源 2 的信息, 如图 4(c) 中紫色时频点所示, 增加了声源 2 被定位出的可能性, 这也是低信噪比环境下 TFCOA-PIV 和 TFCOA-AIV 方法的平均声源检测个数优于 DPD-AIV 方法的主要原因, 即 TFCOA 检测因子具有更强的鲁棒性。

3.4　混响对所提方法定位性能的影响

本节测试所提方法定位性能在不同混响条件下的表现, 其中信噪比为 SNR = 10 dB保持不变, 混响时间为 $T_{60} = \{0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ s。各对比声源定位方法的均方根误差和平均声源检测个数随混响时间的变化情况如图 5 所示。

从图 5 可以看出各对比方法的定位性能均随混响时间的增大而有所降低, 其中性能最优的 TFCOA-AIV 方法的变化趋势最小, 反映出 TFCOA-AIV 方法与其他对比方法相比具有更强的抗混响能力。随着混响时间的增大, TFCOA-AIV 方法的均方根误差从 1.69°增大到 1.88°, 只增大了 0.2°, 平均声源检测个数基本没变; 而 DPD-AIV 方法的均方根误差从 1.99°增大到了 2.26°, 约增大了 0.27°, 平均声源检测个数也基本没变; 而 AIV 和 OA-AIV 方法的均方根误差分别从 3.5°增大到 4.2°和从 3.03°增大到 3.78°, 均增大了 0.7° 左右, 平均声源检测个数分别下降了约 4% 和 2.5%。

对于只利用 0, 1 阶特征波束定位的方法, TFCOA-PIV 方法的均方根误差从 3.29°增大到 3.93°,

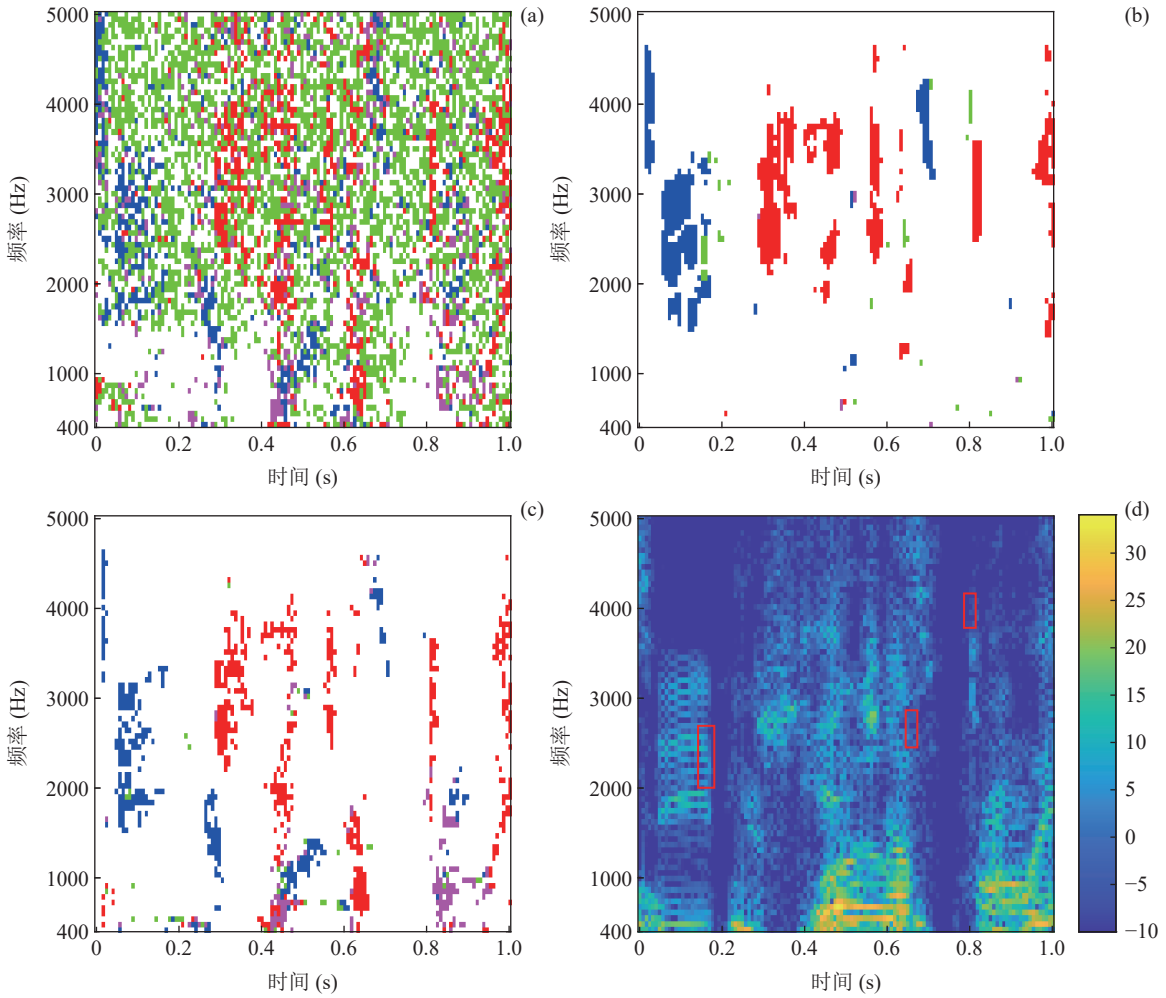

图 4　OA-AIV、DPD-AIV、TFCOA-AIV 方法筛选出的各声源时频点分布及纯净语音语谱图 (a) OA-AIV; (b) DPD-AIV; (c) TFCOA-AIV; (d) 纯净语音语谱图

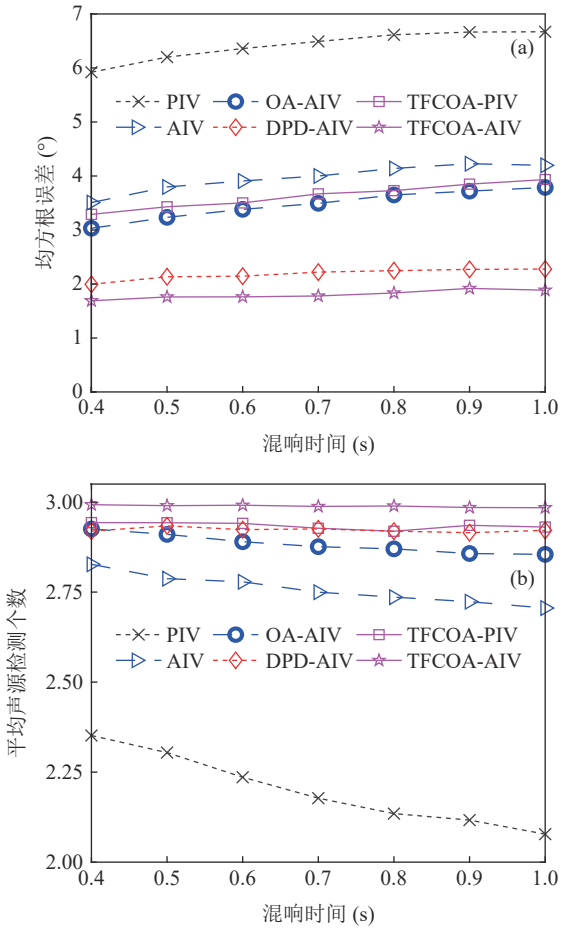

图 5 各对比方法在不同混响时间条件下的均方根误差和平均声源检测个数 (a) 均方根误差; (b) 平均声源检测个数

增大了约 0.64°, 增大趋势低于 OA-AIV 和 AIV 方法, 平均声源检测个数基本没变; PIV 方法的均方根误差和平均声源检测个数随混响时间的增大变化是最大的, 说明其抗混响能力是最差的。

通过以上分析可以看出 TFCOA-AIV 和 TFCOA-PIV 方法的平均声源检测个数基本不随混响时间的增大而变化, 均方根误差变化也小于或与其他方法持平, 反映出 TFCOA 因子具有较强的抗混响能力, 这一点可以从图 4 中进行分析。对比图 4(c) 中 TFCOA-AIV 方法筛选出的时频点在图 4(d) 语谱图中的位置可以看出, 筛选出的时频点主要位于语音信号的起始位置, 根据优先效应可知这部分信号主要是经直接路径传输的直达声信号, 受混响的影响较小, 因此 TFCOA 因子具有较强的抗混响能力。DPD 检测因子筛选出的时频点中有一部分位于语音信号的末尾, 这部分信号以混响为主, 即 DPD 检测因子会将部分混响占优的信号误选出来, 产生错误的定位结果, 影响了其抗混响能力, 如图 4(d) 语谱图中红色方框的位置。虽然 TFCOA 检测因子也会出现误选情况, 但是数量要少于 DPD 检测因子, 且比较分散, 对定位结果的影响较小, 因此具有更强的抗混响能力。

3.5 声源角度间距对所提方法定位性能的影响

接下来测试不同声源角度间距对各对比方法的定位性能影响, 主要测试空间分辨能力。将信噪比设置为 10 dB, 混响时间为 0.7 s, 保持仿真设置中各声源的俯仰角不变, 声源之间的方位角间距进行变化, 方位角间距取值为 {30°, 45°, 60°, 75°, 90°}。不同声源角度间距条件下各对比方法的均方根误差和平均声源检测个数的变化情况如图 6 所示。

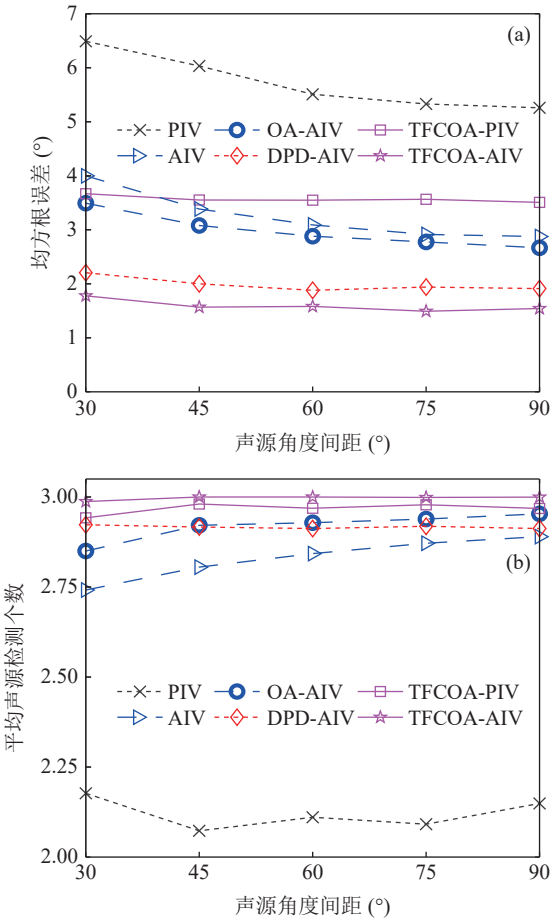

图 6 各对比方法在不同声源角度间距条件下的均方根误差和平均声源检测个数 (a) 均方根误差; (b) 平均声源检测个数

从图 6 中可以看出各对比方法的定位性能随声源角度间距的增大而有所提升, 其中 TFCOA-AIV、TFCOA-PIV 和 DPD-AIV 方法的均方根误差和平均声源检测个数随声源角度间距的增大变化得比较缓慢, 说明这三种方法受声源角度间距的影响较小, 具有较高的空间分辨率。原因是 TFCOA 检测因子和 DPD 检测因子筛选出的时频点绝大部分可以清晰地归属到每个声源, 误选时频点较少, 在形成直方图时每个声源对应的峰比较尖锐, 即使在声源角度间距较小时相互之间的干扰也比较小, 因此随角度间距

的增大定位性能变化不大。对于 PIV、AIV 和 OA-AIV 方法, 由于采用全部时频点或误选时频点较多导致形成直方图时相邻声源的峰之间会产生相互干扰、淹没等现象, 这种现象会随着声源角度间距的增大而改善, 因此随声源角度间距的增大定位性能有较大程度提高。

3.6　阵列增益相位误差对所提方法定位性能的影响

接下来测试传声器阵列存在增益相位误差时各对比方法的定位性能表现。为突出增益相位误差对定位性能的影响, 将信噪比设置为 20 dB, 混响时间为 0.3 s, 相位误差和增益误差按均匀分布随机产生, 其中相位误差的取值范围为 ±5°, 增益误差设置了 5 种取值范围 {±0.02, ±0.04, ±0.06, ±0.08, ±0.10}。不同增益相位误差条件下各对比方法的均方根误差和平均声源检测个数的变化情况如图 7 所示。

从图 7 中可以看出, 各对比方法的定位性能均随传声器阵列增益误差的增大出现不同程度的下降。其中 TFCOA-AIV 方法的定位性能仍然是最优的, 其均方根误差随着阵列增益误差的增大约增加了 0.2°, 平均声源检测个数几乎没变; 与定位性能次优的 OA-AIV 方法相比在均方根误差上降低了 0.5°~0.7°, 在平均声源检测个数上两者几乎相同; 而 TFCOA-PIV 方法在阵列增益相位误差的影响下, 其均方根误差和平均声源检测个数均略优于 AIV 方法。对于 DPD-AIV 方法, 在没有增益相位误差时其均方根误差和平均声源检测个数分别为 1.47°和 3, 仅次于 TFCOA-AIV 方法的 1.22°和 3, 而当存在增益相位误差时其定位性能迅速恶化, 在增益误差取值范围增大到 ±0.1时, 其均方根误差增大到 2.8°, 平均声源检测个数减小到 2.91, 定位性能仅优于 PIV 方法。从中可以看出, TFCOA 检测因子与 DPD 检测因子相比具有更强的抗传声器阵列增益相位误差的能力。

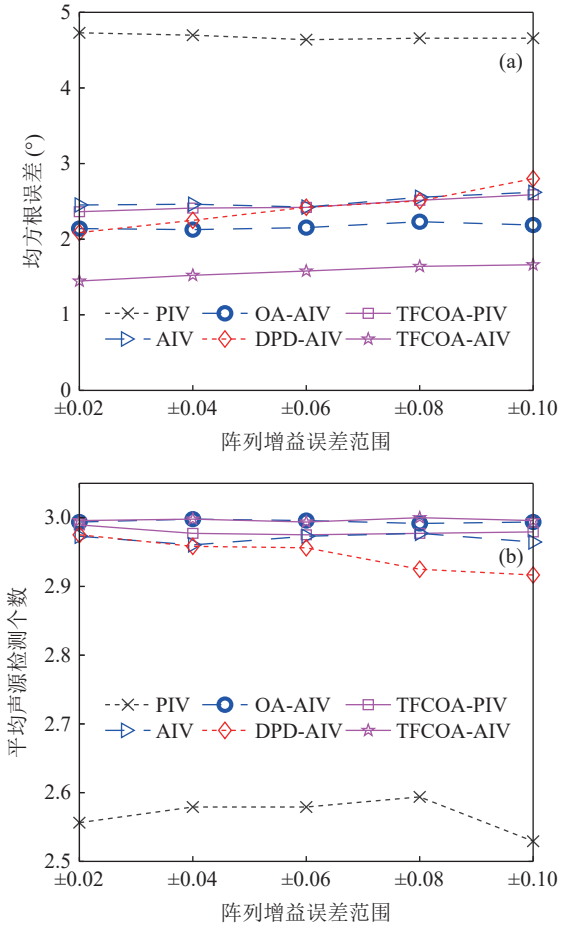

图 7　各对比方法在不同增益误差条件下的均方根误差和平均声源检测个数 (a) 均方根误差; (b) 平均声源检测个数

4　实测实验

本节通过实测实验对所提方法的定位性能进行测试。实测实验环境及球面传声器阵列如图 8 所示。实验的房间尺寸为 9 m × 7 m × 3 m, 房间的混响时间约为 0.35 s。阵列采用 MH 声学公司的 em32 Eigenmike 刚性球面传声器阵列, 半径为 4.2 cm, 阵列表面按均匀分布放置了 32 个全指向性传声器。阵列的采样频率为 44.1 kHz, 为了与仿真实验保持一致, 将采样频率降到 16 kHz。阵列中心位于房间的 4.5 m × 3.5 m × 1.5 m,阵列与 3 个声源的距离均为 1.5 m, 共测试了 3 个位置, 每个位置的声源角度坐标如表 3 所示 (俯仰角在前)。每个位置测试了 10 组从 TIMIT 语音库中随机选取的语音信号, 从阵列采集的信号中截取 1 s 信号进行性能测试。各对比方法的参数设置同仿真实验。各对比定位方法的均方根误差和平均声源检测个数如表 4 所示。

图 8　实测实验环境和球面传声器阵列

从表 4 中可以看出, 定位性能表现最好的是 TFCOA-AIV 方法, 均方根误差和平均声源检测个数均优于其他对比方法; 而 TFCOA-PIV 方法由于只采

表 3 实测实验声源角度坐标

声源	测试位置1	测试位置2	测试位置3
声源1	(108°, −60°)	(108°, −20°)	(108°, −100°)
声源2	(78°, 40°)	(78°, 40°)	(78°, −20°)
声源3	(98°, 140°)	(98°, 100°)	(98°, 60°)

表 4 实测实验各对比方法定位结果的均方根误差和平均声源检测个数

	PIV	DPD-AIV	TFCOA-PIV	AIV	OA-AIV	TFCOA-AIV
均方根误差 (°)	4.71	2.35	**3.36**	2.54	2.23	**2.02**
平均声源检测个数	2.30	2.70	**2.97**	2.77	2.87	**3.00**

用了 0 阶和 1 阶特征波束, 使得其均方根误差高于 OA-AIV、DPD-AIV 和 AIV 方法, 但是其平均声源检测个数均优于这三种方法; 对于 DPD-AIV 方法, 其平均声源检测个数仅高于 PIV 方法, 均方根误差高于 OA-AIV 方法, 与仿真实验中阵列增益相位误差影响下的性能表现相似, 说明 DPD 检测因子在实际应用中由于阵列增益相位误差的存在使其筛选直接路径占优时频点的能力下降。PIV 方法的定位性能最差。通过实测实验可以看出, TFCOA 检测因子在实际环境中的定位表现与仿真实验中一致, 具有较高的定位精度和鲁棒性, 由于实测环境的混响较低、信噪比较高, 使得 TFCOA-AIV 和 TFCOA-PIV 方法与其他方法相比优势并不像仿真实验中那么明显。

图 9 为一次实测实验各对比声源定位方法的归一化直方图。该次实验中声源的真实方位为表 3 中的测试位置 1, 其中声源 2 的音量较低, 且有语音的时长约为 0.4 s, 可视为一个弱声源。在图 9 中符号“+”为声源的真实方位, 符号“o”为各对比方法估计出的声源方位。从图 9 中可以看出 TFCOA-AIV、TFCOA-PIV 和 OA-AIV 方法成功定位出了 3 个声源的方位, 而 PIV、AIV 和 DPD-AIV 在定位声源 2 时出现了较大的误差。其中 PIV、AIV 方法由于没有经过时频点筛选, 其归一化直方图中存在许多由混响噪声等因素产生的虚假峰, 这些虚假峰会影响对声源 DOA 的估计。PIV 方法在定位声源 2 时就错误定位在最高虚假峰的方位 (93°, 139°); AIV 方法虽然在声源 2 的真实方位处形成明显的峰, 但是其归一化高度 0.090 低于最高虚假峰的归一化高度 0.105, 声源 2 被错误定位在 (70°, 154°)。而 OA-AIV 方法由于采用 OA 检测因子进行可靠特征波束和时频点筛选, 对噪声和混响进行了一定程度的抑制, 使得最高虚假峰的归一化高度降低到 0.071, 低于声源 2 的峰

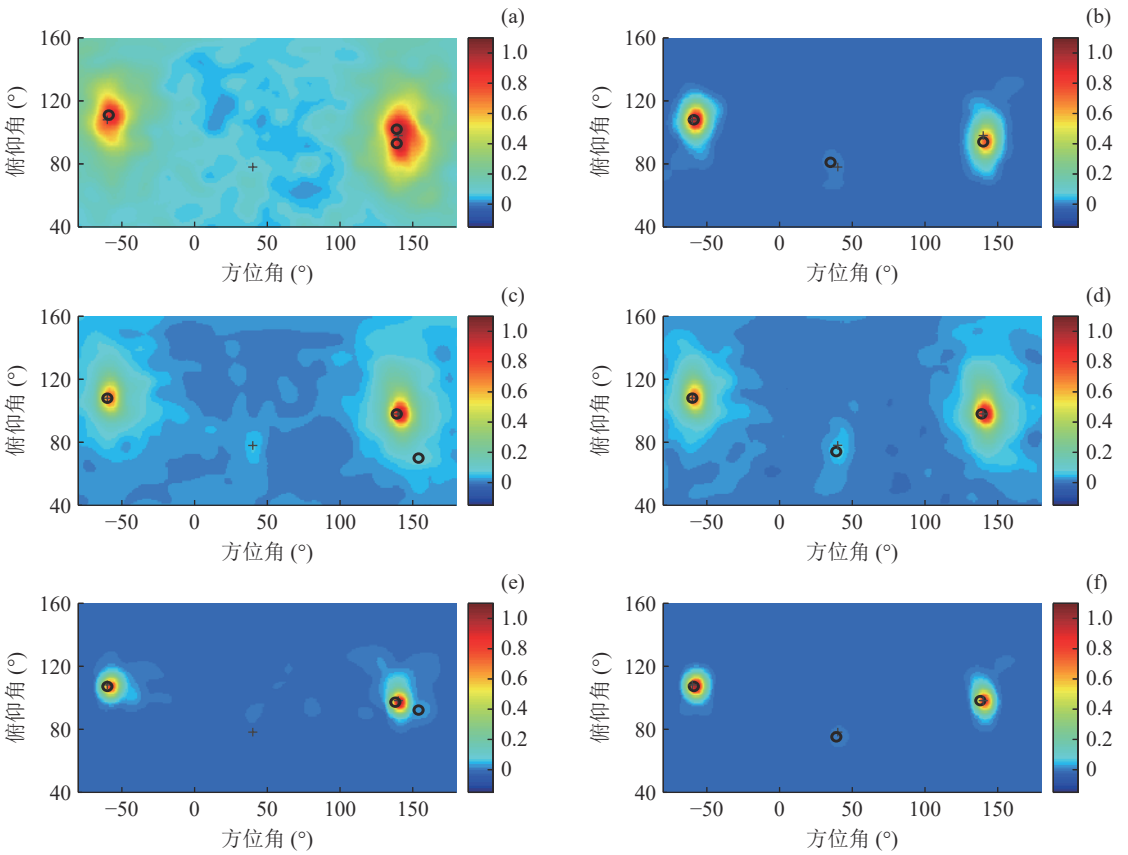

图 9 一次实测实验各对比声源定位方法归一化平滑直方图 (a) PIV; (b) TFCOA-PIV; (c) AIV; (d) OA-AIV; (e) DPD-AIV; (f) TFCOA-AIV

的高度 0.078, 因此将声源 2 成功检测出来。

对于 DPD-AIV 方法, 其归一化直方图相比 PIV、AIV、OA-AIV 方法要干净许多, 但是在定位声源 2 时也出现了错误, 定位在 (92°, 154°), 原因是 DPD 检测因子在阵列增益相位误差的影响下筛选出许多位于低频段的误选时频点, 影响了 DPD-AIV 方法的鲁棒性, 如图 10(a) 所示。图 10 为该次实测实验 DPD-AIV、TFCOA-AIV 方法筛选出的各声源时频点分布图。从图 10(a) 中可以看出 DPD 检测因子误选的绿色时频点数量远远大于声源 2 的紫色时频点, 并且分布比较集中, 使得在图 9 中形成的伪峰高于声源 2 的峰, 产生了错误的定位结果。在图 9 中 TFCOA-AIV、TFCOA-PIV 方法的归一化直方图与其他方法相比非常清晰干净, 原因是 TFCOA 因子筛选出的时频点可以清晰地归类到每个声源, 如图 10(b) 所示, 因此形成的 3 个声源的峰非常尖锐, 具有较高的空间分辨率。另外误选时频点很少, 虽然数量与声源 2 紫色时频点数量相近, 但是分布较为分散, 形成的伪峰低于声源 2 的峰, 因此能够将声源 2 定位出来, 具有较高的鲁棒性。

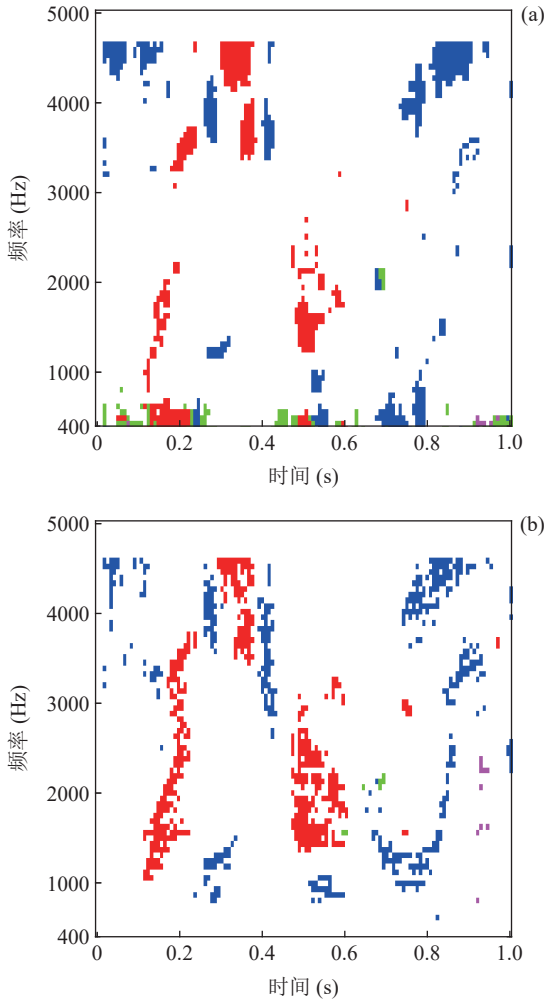

图 10　一次实测实验 DPD-AIV、TFCOA-AIV 方法筛选出的各声源时频点分布图 (a) DPD-AIV; (b) TFCOA-AIV

TFCOA 检测因子在实际环境中同样能够将受噪声和混响影响较小的单声源直达声占优的时频点和特征波束筛选出来, 具有较强的抗噪和抗混响能力, 使得 TFCOA-PIV 和 TFCOA-AIV 方法在实际环境中也具有较高的定位精度和鲁棒性。

5 结论

本文提出一种利用时频相关的球谐域特征波束阶数感知因子以及基于该因子的伪声强声源定位方法, 理论分析了该阶数感知因子相比已有的阶数感知具有更强的抗噪和抗混响能力, 并通过仿真和实测实验测试了所提声源定位方法在混响噪声环境下的定位性能。针对球谐域高阶特征波束易受噪声污染的问题, 利用相邻多个时频点特征波束之间的相关性进行阶数感知的自适应分阶处理, 在降低计算量的同时提升了定位方法的抗噪能力和抗混响能力。仿真实验和实测实验数据处理结果显示, 利用时频相关性的球谐域阶数感知因子能够将受噪声影响较小的和主要位于语音信号起始位置的特征波束筛选出来, 减小了噪声和混响的影响, 同时对阵列的增益和相位误差也具有较强的抵抗能力, 将其应用于伪声强声源定位中提升了定位的性能, 并且随着信噪比降低、混响时间增大、声源角度间距减小, 性能提升得越明显。

参　考　文　献

1　厉剑, 彭任华, 郑成诗, 等. 球谐域自适应混响抵消与声源定位算法. 声学学报, 2019; **44**(5): 874—886

2　Tervo S, Politis A. Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2015; **23**(10): 1539—1551

3　Pezzoli M, Cobos M, Antonacci F, *et al*. Sparsity-based sound field separation in the spherical harmonics domain. 47th IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, Singapore, 2022: 1051—1055

4　Li Z, Duraiswami R. Flexible and optimal design of spherical microphone arrays for beamforming. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2007; **15**(2): 702—714

5　Sun H, Mabande E, Kowalczyk K, *et al*. Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing. *J. Acoust. Soc. Am.*, 2012; **131**(4): 2828—2840

6　高玥, 卢铃, 吴鸣, 等. 采用球谐分解的二范数广义逆波束形成. 应用声学, 2022; **41**(1): 12—20

7　Rafaely B. Analysis and design of spherical microphone arrays. *IEEE Trans. Speech Audio Process.*, 2005; **13**(1): 135—143

8　Rafaely B. Fundamentals of spherical array processing. Berlin Heidelberg: Springer, 2015

9 刘月婵, 何元安, 商德江, 等. 高精度球面阵聚焦声源定位方法研究. 声学学报, 2013; **38**(5): 533—540

10 Rafaely B. Phase-mode versus delay-and-sum spherical microphone array processing. *IEEE Signal Process. Lett.*, 2005; **12**(10): 713—716

11 Nadiri O, Rafaely B. Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2014; **22**(10): 1494—1505

12 Choi J W, Zotter F, Jo B, *et al*. Multiarray eigenbeam-ESPRIT for 3D sound source localization with multiple spherical microphone arrays. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2022; **30**(6): 2310—2325

13 Hu Y, Lu J, Qiu X. A maximum likelihood direction of arrival estimation method for open-sphere microphone arrays in the spherical harmonic domain. *J. Acoust. Soc. Am.*, 2015; **138**(2): 791—794

14 Madmoni L, Rafaely B. Improved direct-path dominance test for speaker localization in reverberant environments. 26th European Signal Processing Conference, EURASIP, Rome, Italy, 2018: 2438—2442

15 Jarrett D P, Habets E, Naylor P. 3D source localization in the spherical harmonic domain using a pseudo-intensity vector. 18th European Signal Processing Conference, EURASIP, Aalborg, Denmark, 2010: 442—446

16 Hafezi S, Moore A H, Naylor P A. 3D acoustic source localization in the spherical harmonic domain based on optimized grid search. 41st IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE, Shanghai, China, 2016: 415—419

17 Pavlidi D, Delikaris-Manias S, Pulkki V, *et al*. 3D localization of multiple sound sources with intensity vector estimates in single source zones. 23rd European Signal Processing Conference, EURASIP, Nice, France, 2015: 1556—1560

18 Hafezi S, Moore A H, Naylor P A. Augmented intensity vectors for direction of arrival estimation in the spherical harmonic domain. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2017; **25**(10): 1956—1968

19 Hafezi S, Moore A H, Naylor P A. Multiple source localization in the spherical harmonic domain using augmented intensity vectors based on grid search. 24th European Signal Processing Conference, EURASIP, Budapest, Hungary, 2016: 602—606

20 Pavlidi D, Delikaris-Manias S, Pulkki V, *et al*. 3D DOA estimation of multiple sound sources based on spatially constrained beamforming driven by intensity vectors. 41st IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE, Shanghai, China, 2016: 96—100

21 Moore A H, Evers C. Direction of arrival estimation in the spherical harmonic domain using subspace pseudo-intensity vectors. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2017; **25**(1): 178—192

22 Moore A H, Evers C, Naylor P A, *et al*. Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test. 23rd European Signal Processing Conference, EURASIP, Nice, France, 2015: 2296—2300

23 Gao W, Chen H. An order-aware scheme for robust direction of arrival estimation in the spherical harmonic domain. *J. Acoust. Soc. Am.*, 2019; **146**(6): 4883—4897

24 鄢社锋, 侯朝焕, 马晓川. 从阵元域到模态域阵列信号处理. 声学学报, 2011; **36**(5): 461—468

25 Rickard S, Dietrich F. DOA estimation of many W-disjoint orthogonal sources from two mixtures using duet. 10th IEEE Workshop Statistical Signal Array Processing, IEEE, Pocono Manor, Pennsylvania, America, 2000: 311—314

26 Jarrett D P, Habets E A P, Thomas M R P, *et al*. Rigid sphere room impulse response simulation: Algorithm and applications. *J. Acoust. Soc. Am.*, 2012; **132**(3): 1462—1472

附录 A

将式 (19) 中的内积展开, 并根据不同时频点噪声之间互不相关, 可得

$$2\mathrm{Re}[\Gamma_n\Gamma_{n'}\mathrm{E}(\langle \boldsymbol{n}_{\boldsymbol{n}}^{0,0},\boldsymbol{n}_{\boldsymbol{n}}^{j_\tau,j_k}\rangle\langle \boldsymbol{n}_{\boldsymbol{n}'}^{0,0},\boldsymbol{n}_{\boldsymbol{n}'}^{j_\tau,j_k}\rangle)] = 2\mathrm{Re}\left[\Gamma_n\Gamma_{n'}\sum_{m=-n}^{n}\sum_{m'=-n'}^{n'}\mathrm{E}(n_{nm}^{0,0}n_{n'm'}^{0,0*})\mathrm{E}(n_{nm}^{j_\tau,j_k*}n_{n'm'}^{j_\tau,j_k})\right], \quad (A1)$$

其中, $n' = n-1$。

考查其中的一项 $\mathrm{E}(n_{nm}^{j_\tau,j_k*}n_{n'm'}^{j_\tau,j_k})$, 根据式 (3) 将 $n_{nm}^{j_\tau,j_k}$, $n_{n'm'}^{j_\tau,j_k}$ 展开, 可得

$$\mathrm{E}(n_{nm}^{j_\tau,j_k*}n_{n'm'}^{j_\tau,j_k}) = \sum_{q=1}^{Q}\sum_{q'=1}^{Q}\alpha_q\alpha_{q'}\mathrm{E}(n_q^{j_\tau,j_k*}n_{q'}^{j_\tau,j_k})\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_{q'}), \quad (A2)$$

其中, $n_q^{j_\tau,j_k}$ 为第 q 个阵元的 $(\tau-j_\tau,k+j_k)$ 时频点的噪声幅度。将上式的求和分成两部分: $q=q'$ 和 $q\neq q'$, 根据不同阵元噪声互不相关, 有

$$\mathrm{E}(n_{nm}^{j_\tau,j_k*}n_{n'm'}^{j_\tau,j_k}) = \mu^2\sum_{q=1}^{Q}\sum_{q'=1,q'\neq q}^{Q}\alpha_q\alpha_{q'}\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_{q'}) + \mathrm{E}(|n^{j_\tau,j_k}|^2)\sum_{q=1}^{Q}\alpha_q^2\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_q). \quad (A3)$$

在式 (A3) 等号右侧第 1 项中添加 $\mu^2\sum_{q=1}^{Q}\alpha_q{}^2\cdot\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_q)$, 并在第 2 项中减去该添加项, 可得

$$\mathrm{E}(n_{nm}^{j_\tau,j_k*}n_{n'm'}^{j_\tau,j_k}) = \mu^2\sum_{q=1}^{Q}\alpha_q\mathrm{Y}_n^m(\Omega_q)\sum_{q'=1}^{Q}\alpha_{q'}\mathrm{Y}_{n'}^{m'*}(\Omega_{q'}) + \sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_q). \quad (A4)$$

由于 $\mathrm{Y}_0^0(\Theta) = 1/\sqrt{4\pi}$, 则式 (A4) 等号右侧第 1 项可以改写为

$$\mu^2\sum_{q=1}^{Q}\alpha_q\mathrm{Y}_n^m(\Omega_q)\sum_{q'=1}^{Q}\alpha_{q'}\mathrm{Y}_{n'}^{m'*}(\Omega_{q'}) = 4\pi\mu^2\sum_{q=1}^{Q}\alpha_q\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_0^{0*}(\Omega_q)\sum_{q'=1}^{Q}\alpha_{q'}\mathrm{Y}_n^{m*}(\Omega_{q'})\mathrm{Y}_0^0(\Omega_{q'}). \quad (A5)$$

根据离散采样球谐波函数的正交性可知, 当 $n=0$, $m=0$, $n'=0$, $m'=0$时式 (A5) 才不为 0, 而 $n\neq n'$, 因此式 (A5) 等于 0。

对于式 (A4) 第 2 项 $\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_q)$, 若采样点为均匀分布时有 $\alpha_q=4\pi/Q$, 同样由于 $n\neq n'$ 有 $\alpha_q\sigma^2\sum_{q=1}^{Q}\alpha_q\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_q)=0$。若采样点为高斯分布或等角度分布时, 根据其分布特点: 同俯仰角的采样点权值相同、同俯仰角采样点的方位角是等角度间距分布, 由式 (2) 可知 $\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_q)$中包含虚指数信号 $\mathrm{e}^{\mathrm{j}(m-m')\varphi_q}$, 因此同俯仰角的采样点的 $\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_q)$相加时, 只有当 $m=m'$时不为 0, 其他情况为 0; 下面分析 $m=m'$时, 由于等角度和高斯分布的采样点关于阵列赤道对称且对称采样点权值相同, 对称采样点 Ω_q 和 $\breve{\Omega}_q$ 的俯仰角存在关系 $\theta_q=\pi-\breve{\theta}_q$, 有 $\mathrm{Y}_n^m\left(\breve{\Omega}_q\right)\mathrm{Y}_{n'}^{m^*}\left(\breve{\Omega}_q\right)=(-1)^{n+2m+n'}\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m^*}(\Omega_q)$, 由于 $n+n'=2n-1$, 因此 $(-1)^{n+2m+n'}=-1$, 使得 $\mathrm{Y}_n^m\left(\breve{\Omega}_q\right)\mathrm{Y}_{n'}^{m^*}\left(\breve{\Omega}_q\right)+\mathrm{Y}_n^m\left(\breve{\Omega}_q\right)\mathrm{Y}_{n'}^{m^*}\left(\breve{\Omega}_q\right)=0$, 因此等角度和高斯分布时 $\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\mathrm{Y}_n^m(\Omega_q)\mathrm{Y}_{n'}^{m'*}(\Omega_q)=0$。综合以上分析当采样点为均匀分布、高斯分布和等角度分布时有 $\mathrm{E}(n_{nm}^{j_\tau,j_k*}n_{n'm'}^{j_\tau,j_k})=0$, 结论同样适用于 $\mathrm{E}(n_{nm}^{0,0*}n_{n'm'}^{0,0})$, 因此:

$$2\mathrm{Re}[\Gamma_n\Gamma_{n-1}\mathrm{E}(\langle\boldsymbol{n}_{\boldsymbol{n}}^{0,0},\boldsymbol{n}_{\boldsymbol{n}}^{j_\tau,j_k}\rangle\langle\boldsymbol{n}_{\boldsymbol{n}-1}^{0,0},\boldsymbol{n}_{\boldsymbol{n}-1}^{j_\tau,j_k}\rangle)]=0. \tag{A6}$$

附录 B

考察式 (18) $K_n(\tau,k)$的分母中的任一时频点中阶数为 n 的项, 省略 $1/|b_n(kr_a)|^4$, 有

$$\mathrm{E}(\Gamma_n^2\|\boldsymbol{n}_{\boldsymbol{n}}^{0,0}\|^2\|\boldsymbol{n}_{\boldsymbol{n}}^{j_\tau,j_k}\|^2)=\mathrm{E}\left(\Gamma_n^2\sum_{m=-n}^{n}n_{nm}^{0,0}n_{nm}^{0,0*}\sum_{m=-n}^{n}n_{nm}^{j_\tau,j_k}n_{nm}^{j_\tau,j_k*}\right), \tag{B1}$$

其中, $n_{nm}^{0,0}$, $n_{n'm'}^{0,0*}$ 与 $n_{nm}^{j_\tau,j_k*}$, $n_{n'm'}^{j_\tau,j_k}$属于不同时频点, 是统计独立的, 因此上式可改写为

$$\mathrm{E}\left(\Gamma_n{}^2\sum_{m=-n}^{n}n_{nm}^{0,0}n_{nm}^{0,0*}\sum_{m=-n}^{n}n_{nm}^{j_\tau,j_k}n_{nm}^{j_\tau,j_k*}\right)=\Gamma_n^2\sum_{m=-n}^{n}\mathrm{E}(n_{nm}^{0,0}n_{nm}^{0,0*})\sum_{m=-n}^{n}\mathrm{E}(n_{nm}^{j_\tau,j_k}n_{nm}^{j_\tau,j_k*}). \tag{B2}$$

考查其中一项:

$$\Gamma_n\sum_{m=-n}^{n}\mathrm{E}(n_{nm}^{j_\tau,j_k}n_{nm}^{j_\tau,j_k*})=\Gamma_n\sum_{m=-n}^{n}\sum_{q=1}^{Q}\sum_{q'=1}^{Q}\alpha_q\alpha_{q'}\mathrm{E}(n_q^{j_\tau,j_k}n_{q'}^{j_\tau,j_k}{}^*)\mathrm{Y}_n^{m*}(\Omega_q)\mathrm{Y}_n^m(\Omega_{q'}). \tag{B3}$$

将上式的求和分成两部分: $q=q'$ 和 $q\neq q'$, 根据不同阵元噪声互不相关, 有

$$\Gamma_n\sum_{m=-n}^{n}\sum_{q=1}^{Q}\sum_{q'=1}^{Q}\alpha_q\alpha_{q'}\mathrm{E}(n_q^{j_\tau,j_k}n_{q'}^{j_\tau,j_k}{}^*)\mathrm{Y}_n^{m*}(\Omega_q)\mathrm{Y}_n^m(\Omega_{q'})=\mu^2\Gamma_n\sum_{m=-n}^{n}\sum_{q=1}^{Q}\alpha_q\mathrm{Y}_n^{m*}(\Omega_q)\sum_{q'=1}^{Q}\alpha_{q'}\mathrm{Y}_n^m(\Omega_{q'})+\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2. \tag{B4}$$

由于 $n>0$, 根据式 (A4) 可知上式第 1 项为 0, 因此 $\Gamma_n\sum_{m=-n}^{n}\mathrm{E}(n_{nm}^{j_\tau,j_k}n_{nm}^{j_\tau,j_k*})=\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2$; 同理 $\Gamma_n\sum_{m=-n}^{n}\mathrm{E}(n_{nm}^{0,0}n_{nm}^{0,0*})=\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2$。因此:

$$\mathrm{E}(\Gamma_n^2\|\boldsymbol{n}_{\boldsymbol{n}}^{0,0}\|^2\|\boldsymbol{n}_{\boldsymbol{n}}^{j_\tau,j_k}\|^2)=\left(\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\right)^2. \tag{B5}$$

对于 $\mathrm{E}(\Gamma_{n-1}^2\|\boldsymbol{n}_{\boldsymbol{n}-1}^{0,0}\|^2\|\boldsymbol{n}_{\boldsymbol{n}-1}^{j_\tau,j_k}\|^2)$, 由于 $n-1$可以取 0, 则 $n=1$ 时式 (B4) 中的第 1 项不为 0。因此:

$$\mathrm{E}(\Gamma_{n-1}^2\|\boldsymbol{n}_{\boldsymbol{n}-1}^{0,0}\|^2\|\boldsymbol{n}_{\boldsymbol{n}-1}^{j_\tau,j_k}\|^2)=\begin{cases}\left[(4\pi\mu)^2+\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\right]^2, & n=1,\\ \left(\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\right)^2, & n>1.\end{cases} \tag{B6}$$

应用到整个分母, 因此分母的均值为

$$\begin{cases}(J_kJ_\tau-1)\left(\dfrac{\left[(4\pi\mu)^2+\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\right]^2}{|b_0(kr_a)|^4}+\dfrac{\left(\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\right)^2}{|b_1(kr_a)|^4}\right), & n=1,\\ (J_kJ_\tau-1)\left(\dfrac{\left(\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\right)^2}{|b_{n-1}(kr_a)|^4}+\dfrac{\left(\sigma^2\sum_{q=1}^{Q}\alpha_q{}^2\right)^2}{|b_n(kr_a)|^4}\right), & n>1.\end{cases} \tag{B7}$$