

混响和噪声环境下基于角度谱的多声源定位与计数算法*

房玉琢, 许志勇, 赵 兆

(南京理工大学 电子工程与光电技术学院, 江苏 南京 210094)

摘要:针对混响及噪声环境下的多声源定位与计数问题,引入局部信噪比追踪及相关性检测模块,提取出传统广义互相关角度谱中受噪声及声源互扰影响较小的时频支撑域;同时引入双宽度匹配追踪方法替代传统的峰值搜索,改进后续定位与计数的精确度。仿真研究验证了综合应用滤波后的角度谱及双宽度匹配追踪的多声源定位与计数算法相比传统算法在较低信噪比、较强混响以及较多声源数的环境中更加精确及稳健。

关键词:麦克风阵列;角度谱;时频域滤波;匹配追踪;定位与计数

中图分类号: TN911.72 **文献标志码:** A **文章编号:** 1001-2486(2018)03-113-07

Multiple sound source localization and counting algorithm based on angular spectrum under noisy and reverberant environment

FANG Yuzhuo, XU Zhiyong, ZHAO Zhao

(School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: For multiple sound source localization and counting under noisy and reverberant environment, two modules, local signal-to-noise ratio tracking and coherence test, were introduced to extract the time-frequency bins less affected by noise and mutual interference of the sound sources in classical generalized cross-correlation angular spectrum; then the traditional peak search method was replaced by the dual-width matching pursuit to improve the following localization and counting accuracy. Through the simulation results, the proposed algorithm of multiple sound source localization and counting using both the filtered angular spectrum and dual-width matching pursuit method is proved to be more accurate and robust than the traditional one, especially under the environment with lower signal-to-noise ratio, stronger reverberation and more sound sources.

Key words: microphone array; angular spectrum; time-frequency filtering; matching pursuit; localization and counting

人们日益关注的室内会议、楼宇安防及虚拟现实等应用中,多个声频信号的定位与计数是阵列信号处理领域的核心问题,它通常受到环境噪声、混响以及声源间互扰等因素的综合影响^[1]。为实现多个声源的精确定位与计数,众多学者提出了很多富有启发意义的算法,其中较为主流的是基于角度谱的分支^[2],它可划分为角度谱构建、定位与计数两个部分。

角度谱构建部分:角度谱通过将各时频(Time-Frequency, TF)支撑域处与阵列到达方向角(Direction Of Arrival, DOA)相关的穷值枚举谱函数^[2]进行累积而得到。经典的相位变换广义互相关^[3-5](Generalized Cross-Correlation PHase Transform, GCC-PHAT, 后文简称为GCC)法由于相位加权因子的作用,具有一定的抗噪声性能且

对于空域混叠具有良好的适应性^[2,5],然而基于理想单一声源传播模型的局限性使其在较强的混响环境下性能出现恶化^[4];近似核密度估计^[6](Kernel Density Estimator, KDE)算法中的非线性核函数在较少声源数时相较于GCC抗混响性能更好^[7],且与频率相关的加权因子能够一定程度上抑制空域混叠,然而其抑制作用受到核函数带宽的影响^[8]。

定位与计数部分:通过角度谱中参数极值的判定获取声源的DOA及相应数量。传统的峰值搜索^[9](Peak Search, PS)通过高于截止阈值的谱峰个数来实现声源的定位与计数,文献[10]对其中的阈值作了自适应改进。由于较强混响环境下第一部分所得的角度谱失真较为严重,基于单点谱峰幅度的PS在DOA估计时可能出现较大的偏

* 收稿日期:2017-06-11

基金项目:国家自然科学基金资助项目(61171167,61401203)

作者简介:房玉琢(1987—),男,江苏南京人,博士研究生,E-mail:ramboyy1@hotmail.com;

许志勇(通信作者),男,副教授,博士,硕士生导师,E-mail:ezyxu@mail.njust.edu.cn

移值,造成性能的不稳定。迭代贡献消除^[5] (Iterative Contribution Removal, ICR)从声源贡献量的角度滤除每次迭代估计出的声源所包含的时频支撑域,再由剩余的时频域重新构建角度谱,该方法能够一定程度上削弱干扰对角度谱定位与计数的影响,然而时频域滤除时的搜索模块增加了该方法的计算成本,同时前次估计的错误结果会对后续迭代产生影响。

本文的创新点主要包括:①角度谱构建部分,通过局部信噪比(Signal-to-Noise Ratio, SNR)追踪^[5]和相干性检测^[11](Coherence Test, CT)两个时频域滤波模块提取出传统 GCC 角度谱中受噪声及声源互扰影响较小的时频支撑域,构建 GCCTF,从而有效缓解角度谱中伪峰对正确谱峰的干扰;②定位与计数部分,使用双宽度^[10]匹配追踪^[12](Matching Pursuit, MP),通过最大内积及声源贡献迭代消除的方式替代传统 PS 中基于单点谱峰幅度的比较,从而改善多声源定位与计数算法的性能。

1 多声源传播模型及 GCC 角度谱

以最基本的双元麦克风阵列为例,相应的多声源信号传播模型如图 1 所示。声场空间中存在数个独立声源 s_1, \dots, s_N 及一对间距 d_{mic} 的全向麦克风 m_1, m_2 。其中声源的到达方向角 $\theta_1, \dots, \theta_N$ 及数量 N 均未知。在近似远场的条件下,其传播模型可表示为

$$x_m(\rho) = \sum_{n=1}^N \mathbf{h}_{m,n}^T s_n(\rho) + x_m^{(W)}(\rho) \quad (1)$$

式中: $x_m(\rho)$ ($m = 1, 2$)表示第 m 个麦克风的接收信号; $\mathbf{h}_{m,n} = [h_{m,n}(0), \dots, h_{m,n}(L_h - 1)]^T$ 对应第 n 个声源与第 m 个麦克风间长度为 L_h 的线性时不变冲激响应; $s_n(\rho) = [s_n(\rho), \dots, s_n(\rho - L_n +$

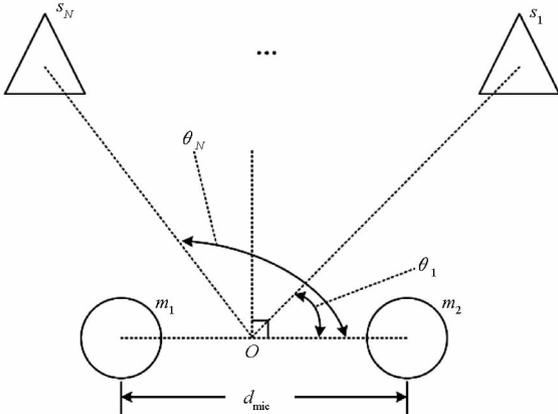


图 1 多声源信号传播模型

Fig. 1 Propagation model of multiple sound source signals

$]^T$ 表示第 n 个声源的离散时间信号矢量,采样率为 f_s ; $x_m^{(W)}(\rho)$ 表示与声源信号及冲激响应不相关的第 m 个麦克风的加性高斯白噪声。将式(1)通过 L_{FFT} 点短时傅里叶变换(Short Time Fourier Transform, STFT)变换到离散时频域,可得

$$X_m(r, k) = \sum_{n=1}^N H_{m,n}(k) S_n(r, k) + X_m^{(W)}(r, k) \quad (2)$$

式中: $X_m(r, k)$ 和 $S_n(r, k)$ 分别表示与接收信号 $x_m(\rho)$ 和声源信号 $s_n(\rho)$ 相对应的第 r 个时间帧、第 k 个频率段处的 STFT 系数; $X_m^{(W)}(r, k)$ 表示加性复噪声; $H_{m,n}(k)$ 表示第 n 个声源与第 m 个麦克风间的传递函数,在空间扩散混响的假设下^[13],可将其分解为直达波 $H_{m,n}^{(D)}(k)$ 及混响 $H_{m,n}^{(R)}(k)$ 两个部分,即

$$H_{m,n}(k) = H_{m,n}^{(D)}(k) + H_{m,n}^{(R)}(k) = \alpha_{m,n} \exp(-j2\pi f_k T_{m,n}) + H_{m,n}^{(R)}(k) \quad (3)$$

其中, $\alpha_{m,n}$ 和 $T_{m,n}$ 分别表示第 n 个声源与第 m 个麦克风间直达波的传播衰减和时间, $f_k = kf_s / L_{FFT}$ 表示第 k 个频率段处的中心频率。理想情况下,混响和噪声均不存在,且声频信号的时频稀疏和短时正交(W-Disjoint Orthogonality, WDO)假设^[14]成立,各时频支撑域处至多只有单个声源能量占优,其序号记作 $\eta(r, k)$ 。此时,式(2)可简化为

$$X_m(r, k) = (\alpha_{m,\eta(r,k)} S_{\eta(r,k)}(r, k)) e^{-j2\pi f_k T_{m,\eta(r,k)}} \quad (4)$$

由式(4)中的简化传播模型可得各时频支撑域处关于声源 DOA 的局部 GCC 角度谱^[2]

$$\begin{aligned} \phi_{GCC}(r, k, \theta) &= \text{Re} \left(\frac{X_1(r, k) X_2^*(r, k)}{|X_1(r, k) X_2(r, k)|} e^{-j2\pi f_k \tau(\theta)} \right) \\ &= \text{Re} (e^{-j2\pi f_k [\tau(\theta) - \tau_{\eta(r,k)}]}) \end{aligned} \quad (5)$$

式中, $\tau_{\eta(r,k)} = T_{1,\eta(r,k)} - T_{2,\eta(r,k)}$ 表示各支撑域处主导声源到麦克风对的到达时间差(Time Difference Of Arrival, TDOA), $\tau(\theta) = -d_{mic} \cos(\theta) / c$ (c 表示大气声速), $\text{Re}(\cdot)$ 表示取复数实部运算符,* 表示复共轭运算符。将式(5)在所有时频支撑域组成的集合 $\Omega_{TF}^{(total)}$ 中进行累积可得到 GCC 角度谱

$$\Phi_{GCC}(\theta) = \sum_{(r,k) \in \Omega_{TF}^{(total)}} \phi_{GCC}(r, k, \theta), \theta \in \Omega_{\theta}^{(total)} \quad (6)$$

式中, $\Omega_{\theta}^{(total)}$ 表示 $[0^\circ, 180^\circ]$ 内 DOA 单位网格宽度为 θ_{min} 的线性空间, θ_{min} 为一兼顾计算准确度与复杂度的经验值,本文取 0.5° ^[10]。

2 基于角度谱的多声源定位与计数

2.1 角度谱的时频域滤波

实际中,噪声不可避免,且在声源数较多的混响环境下,WDO 假设无法严格满足。因此本节引入局

部 SNR 追踪、相关性检测模块提取出 $\Omega_{TF}^{(total)}$ 中受噪声影响较小且单个声源能量占优的时频支撑域。

2.1.1 局部 SNR 追踪

各时频支撑域处的局部 SNR 可表示为

$$\gamma_{ST}(r, k) = \min \left\{ \log_{10} \left(\frac{|X_m(r, k)|^2}{|X_m^{(W)}(r, k)|^2} - 1 \right) \mid m = 1, 2 \right\} \quad (7)$$

式中, $\min(\cdot)$ 表示求集合中的最小值。假设噪声时不变且前 L_w 段为纯噪声^[5], 则有

$$|X_m^{(W)}(r, k)|^2 = |X_m^{(W)}(k)|^2 = \frac{1}{L_w} \sum_{r'=1}^{L_w} |X_m(r, k)|^2 \quad (8)$$

此时可以根据应用环境设置经验阈值 Γ_{ST} 将所有受噪声影响较小的时频支撑域提取出来, 从而得到满足局部 SNR 追踪的时频支撑域集合:

$$\Omega_{TF}^{(ST)} = \{ (r, k) \mid \gamma_{ST}(r, k) > \Gamma_{ST} \} \quad (9)$$

2.1.2 相干性检测

各时频支撑域处的相干性参数^[11]可表示为

$$\gamma_{CT}(r, k) = \left| \frac{E(X_1(r, k)X_2^*(r, k))}{\sqrt{E(X_1(r, k)X_1^*(r, k))} \sqrt{E(X_2(r, k)X_2^*(r, k))}} \right| \quad (10)$$

式中, $E(\cdot)$ 表示对 $2C + 1$ 个连续时间帧作平均的近似数学期望, 即

$$E(X_m(r, k)X_m^*(r, k)) = \frac{1}{2C + 1} \sum_{r'=r-C}^{r+C} X_m(r', k)X_m^*(r', k) \quad (11)$$

$m \in \{1, 2\}, m' \in \{1, 2\}$

若该支撑域的 $\gamma_{CT}(r, k)$ 高于给定经验阈值 Γ_{CT} , 则认为仅包含单个主导声源, 由此满足相干性检测的所有时频支撑域的集合可被表示为

$$\Omega_{TF}^{(CT)} = \{ (r, k) \mid \gamma_{CT}(r, k) > \Gamma_{CT} \} \quad (12)$$

通过局部 SNR 追踪及相干性检测两个模块, 结合式(6), 可得 GCCTF 的角度谱

$$\Phi_{GCCTF}(\theta) = \sum_{(r, k) \in \Omega_{TF}} \phi_{GCC}(r, k, \tau(\theta)) \quad (13)$$

式中, $\Omega_{TF} = \Omega_{TF}^{(ST)} \cap \Omega_{TF}^{(CT)}$ 为经过时频域滤波后的所有支撑域组成的集合。

2.2 双宽度匹配追踪定位与计数

当第1节或第2.1小节中的角度谱被构建后, 为进一步实现多个声源的定位与计数, 采用角度谱矢量

$$\Phi_{str} = [\Phi_{str}(0^\circ), \dots, \Phi_{str}(z_{\theta_{min}}), \dots, \Phi_{str}(180^\circ)] \quad (14)$$

其长度为 $\text{card}(\Omega_{\theta}^{(total)})$, $z \in \mathbb{N}$, 下标 str 代表字符串“GCC”或“GCCTF”, 以下为表述方便, 不失一般性, Φ_{str} 统一表述为 Φ 。

峰值搜索通过谱峰幅度与截止阈值 Γ_{PS} 的比较实现定位与计数。混响与噪声环境下, 角度谱失真较为严重, 此时通过内积最大值的匹配追踪^[12]具有更高的精确度。

首先构建基本原子组成的集合

$$\Omega_u = \{ \mathbf{u}^{>>q} \mid 0 \leq q \leq \text{card}(\Omega_{\theta}^{(total)}) - 1, q \in \mathbb{N} \} \quad (15)$$

式中, “ $>>$ ”表示循环右移运算符, $\mathbf{u}^{>>q}$ 表示

$$\mathbf{u} = \mathbf{u}^{>>0} = \frac{\mathbf{v}^{>>(-Q)}}{\|\mathbf{v}^{>>(-Q)}\|} \quad (16)$$

向右循环移动 q 位的行矢量。其中, $\|\cdot\|$ 表示矢量的 l_2 范数运算符, $\mathbf{v}^{>>(-Q)}$ 表示

$$\mathbf{v} = \mathbf{v}^{>>0} = [\mathbf{b}, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{\text{card}(\Omega_{\theta}^{(total)}) - (2Q + 1)}, \mathbf{0}]^T \quad (17)$$

向左循环移动 Q 位的行矢量, 其中, \mathbf{b} 表示长度为 $2Q + 1$ 的 Blackman 窗, 半窗宽度 Q 是正整数。

式(17)中窗宽的选择是一个需要折中的问题: 过宽, 距离较近的 DOA 间估计精确度降低; 过窄, 获得所有 DOA 估计的迭代次数增加, 效率降低。因此引入双宽度结构^[10], 由窗宽较窄及较宽的原子分别进行位置判断及迭代处理。

将 $Q, \mathbf{u}, \mathbf{v}, \mathbf{b}$ 加上下标“1”“2”分别对应较窄及较宽的窗宽, 相应的双宽度 MP 定位与计数方法原理框图如图2所示, 其中循环体第 i 次迭代的角谱谱矢量设为 $\Phi(i)$, 初始值 $\Phi(1) = \Phi$ 。

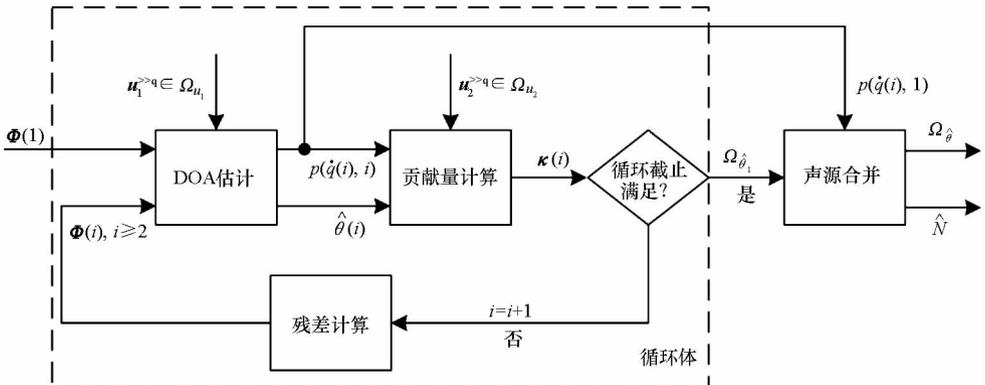


图2 双宽度 MP 原理框图

Fig. 2 Block diagram of dual-width MP method

1) DOA 估计: 计算集合 Ω_n 中各窄原子与第 i 次角度谱矢量的内积

$$p(q, i) = \langle \mathbf{u}_1^{>q}, \boldsymbol{\Phi}(i) \rangle \quad (18)$$

式中, $\langle \cdot \rangle$ 表示求矢量内积运算符。进而通过搜索内积最大的位置得到第 i 次循环的 DOA 估计值

$$\begin{aligned} \hat{\theta}(i) &= \theta_{\min} \cdot \arg \max_q p(q, i) \\ &= \theta_{\min} \cdot \hat{q}(i) \end{aligned} \quad (19)$$

式中, $\hat{q}(i)$ 对应第 i 次循环中使得内积最大的 \mathbf{u}_1 循环右移的位数。

2) 贡献量计算: 由式(18)中的最大内积及相应宽原子得到第 i 次循环估计声源的贡献矢量

$$\boldsymbol{\kappa}(i) = p(\hat{q}(i), i) \cdot \mathbf{u}_2^{>\hat{q}(i)} \quad (20)$$

3) 循环截止判断: 当

$$\begin{cases} \frac{\kappa_{\text{sum}}(i)}{\kappa_{\text{sum}}(1)} < \Gamma_{\text{MP}} \\ i > I_{\text{max}} \end{cases} \quad (21)$$

中两个条件表达式满足其一, 循环截止。其中, Γ_{MP} 为截止阈值, I_{max} 为最大循环次数,

$$\kappa_{\text{sum}}(i) = \text{sum}(\boldsymbol{\kappa}(i)) \quad (22)$$

表示贡献矢量 $\boldsymbol{\kappa}(i)$ 所对应的累积标量值, $\text{sum}(\cdot)$ 表示矢量中元素求和运算符。

4) 残差计算: 在 $\boldsymbol{\Phi}(i)$ 的基础上去除式(20)中的声源贡献矢量 $\boldsymbol{\kappa}(i)$, 所得残差

$$\boldsymbol{\Phi}(i+1) = \boldsymbol{\Phi}(i) - \boldsymbol{\kappa}(i) \quad (23)$$

作为第 $i+1$ 次循环的角度谱矢量。

将循环截止时所历经的迭代次数设为 I_{loop} , 则经过循环体 1) ~ 4) 迭代后得到的 DOA 估计值的集合可表示为

$$\Omega_{\hat{\theta}_i} = \{ \hat{\theta}(i) \mid 1 \leq i \leq I_{\text{loop}} - 1, i \in \mathbb{N} \} \quad (24)$$

由于原子宽度有限, 实际中每次迭代不一定能完全去除某个声源的贡献量值, 式(24)中可能包含重复估计的 DOA, 为避免该情况的发生, 在循环体外引入声源合并模块。

5) 声源合并: 当第 i 次与第 i' 次估计的 DOA 之间角度差的绝对值小于 A_{min} 时, 即

$$| \hat{\theta}(i') - \hat{\theta}(i) | < A_{\text{min}} \quad (25)$$

根据各次对应窄原子与初始角度谱矢量的内积大小将其重新赋值, 具体表达式为

$$\begin{cases} \hat{\theta}(i) = \hat{\theta}(i'), p(\hat{q}(i), 1) > p(\hat{q}(i'), 1) \\ \hat{\theta}(i') = \hat{\theta}(i), \text{其他} \end{cases} \quad (26)$$

其中, A_{min} 是一经验值, 通常取 10° ^[5]。将式(26)处理后的元素去重即可得到最终的定位和计数结果

$$\begin{cases} \Omega_{\hat{\theta}} = \{ \hat{\theta}_n \mid 1 \leq n \leq \hat{N}, i \in \mathbb{N} \} \\ \hat{N} = \text{card}(\Omega_{\hat{\theta}}) \end{cases} \quad (27)$$

式中, 集合 $\Omega_{\hat{\theta}}$ 中的元素 $\hat{\theta}_n$ 表示第 n 个声源对应

DOA 的估计值, \hat{N} 为声源数量的估计值。

3 计算机仿真与统计分析

为验证所提算法的定位与计数性能, 使用镜像法^[15]来产生观察数据。其中房间尺寸为 $8 \text{ m} \times 7 \text{ m} \times 3 \text{ m}$, 声速 c 为 344 m/s 。一对平行于 x 轴的全向麦克风 m_1 、 m_2 位于房间中心 o 附近, 阵元间距 d_{mic} 为 0.8 m ; 以 o 为圆心、间距 d_{ms} 为 2 m 的圆弧上包含 N 个独立声源, 当 N 取 2、4、6 时其真实 DOA 的分布如表 1 所示, 各声源处的声频信号摘录于 8 个男声、8 个女声组成的纯净语音数据库, 采样率 f_s 为 16 kHz 。每次仿真从中随机取出长度为 1.024 s 的 N 个信号段, 通过预处理使其平均功率相同。

表 1 DOA 随声源数的分布

Tab. 1 Number of sources versus DOA

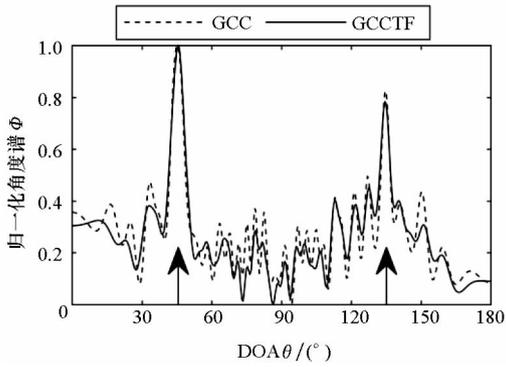
声源数目 N	真实 DOA/(°)
2	45, 135
4	15, 45, 135, 165
6	15, 45, 75, 105, 135, 165

仿真中的其他参数设置如下: 角度谱构建部分, 将帧长 512、帧移 50% 的观察数据作 512 点汉宁窗加权的 STFT; 局部 SNR 追踪模块中, $L_w = 2$, $\Gamma_{\text{ST}} = 5 \text{ dB}$; 相干性检测模块中, $C = 2$, $\Gamma_{\text{CT}} = 0.8$; 在使用 MP 及 PS 方法时, 窄、宽半原子宽度分别取 $Q_1 = 4$ 以及 $Q_2 = 8$, 最大循环数 $I_{\text{max}} = 10$, 截止参数 $\Gamma_{\text{PS}} = 0.6$, $\Gamma_{\text{MP}} = 0.55$ 。

图 3(a) ~ (b) 分别给出了 $N = 2$, $\text{SNR} = 10 \text{ dB}$ 时, 在混响时间 (Reverberation Time of 60 dB, RT_{60}) 为 0.2 s 、 0.5 s 时单次仿真的归一化角度谱, 各子图中的箭头标识出了正确的 DOA 位置。由图 3 可以看出, 当 $\text{RT}_{60} = 0.2 \text{ s}$ 时, GCC 和 GCCTF 在正确 DOA 位置处的谱峰幅度基本持平, 其中 GCCTF 中的噪声抑制模块能够有效降低角度谱基底, 使得偏离 DOA 位置的伪峰幅度基本低于 GCC 的; 当 $\text{RT}_{60} = 0.5 \text{ s}$ 时, 混响程度的加强使得角度谱的失真程度加剧, GCC 中 135° 附近的谱峰幅度出现了下降, 且出现了由椭圆标识出的接近其幅度的伪峰, 因此有产生虚警的趋势, 而 GCCTF 在正确位置的谱峰幅度没有明显变化, 且谱基底仍然低于 GCC, 相比 GCC 区分度更好。

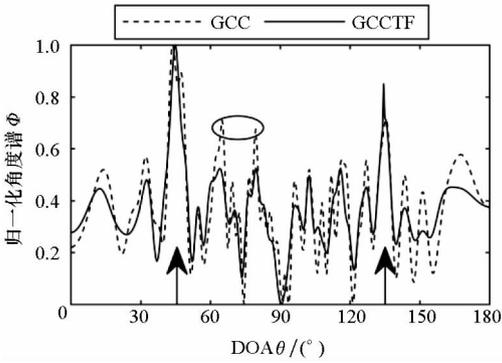
当 $\text{SNR} = 0 \text{ dB}$ 时, 图 4 给出了相应的归一化角度谱。相比图 3, SNR 的降低使得谱基底进一

步升高,对同一 RT_{60} ,伪峰与正确位置间谱峰幅度的差距缩小。图4(a)中 $RT_{60} = 0.2$ s时的情形与图3(a)类似;而当 $RT_{60} = 0.5$ s,图4(b)中GCC中的谱峰相对 135° 的正确位置有所偏离(角度差小于 A_{\min} ,故仍视作正确的谱峰),且角度谱中的伪峰数目变多,GCCTF中也出现了高于正确位置的伪峰。整体而言,GCCTF的区分度仍明显好于GCC。



(a) $RT_{60} = 0.2$ s时的归一化角度谱

(a) Normalized angular spectrum when $RT_{60} = 0.2$ s

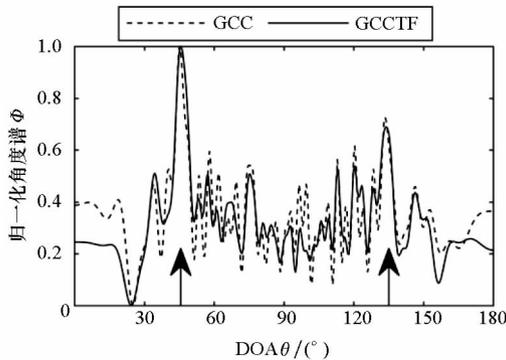


(b) $RT_{60} = 0.5$ s时的归一化角度谱

(b) Normalized angular spectrum when $RT_{60} = 0.5$ s

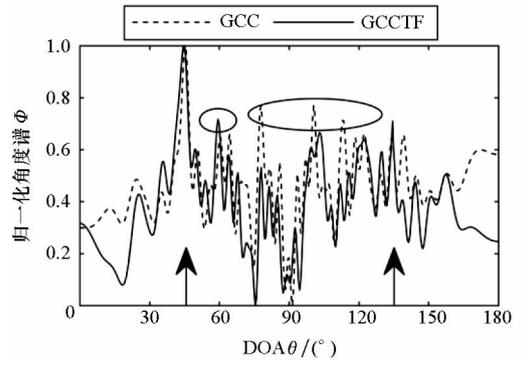
图3 SNR = 10 dB时不同混响环境下的归一化角度谱

Fig.3 Normalized angular spectrum under different reverberant environment when SNR = 10 dB



(a) $RT_{60} = 0.2$ s时的归一化角度谱

(a) Normalized angular spectrum when $RT_{60} = 0.2$ s



(b) $RT_{60} = 0.5$ s时的归一化角度谱

(b) Normalized angular spectrum when $RT_{60} = 0.5$ s

图4 SNR = 0 dB时不同混响环境下的归一化角度谱

Fig.4 Normalized angular spectrum under different reverberant environment when SNR = 0 dB

进一步地将GCC、GCCTF两种角度谱与PS、MP两种方法相结合,得到4种定位与计数算法——GCC-PS、GCCTF-PS、GCC-MP以及GCCTF-MP。使用平均绝对估计误差^[10](Mean Absolute Estimated Error, MAEE)来考察 RT_{60} 为0.2 s、0.5 s, SNR为0~20 dB(变化间隔5 dB),声源数 N 为2、4和6的仿真环境下,当仿真次数 $I_{\text{sim}} = 100$ 时的统计平均定位与计数性能,其表达式为

$$MAEE = \frac{1}{I_{\text{sim}}} \sum_{i=1}^{I_{\text{sim}}} \frac{1}{N_{\text{max}}(i)} \sum_{n=1}^{N_{\text{max}}(i)} |\hat{\theta}_n(i) - \theta_n(i)| \quad (28)$$

式中, $\theta_n(i)$ 和 $\hat{\theta}_n(i)$ 分别表示第 i 次仿真中第 n 个声源真实的和估计的DOA,

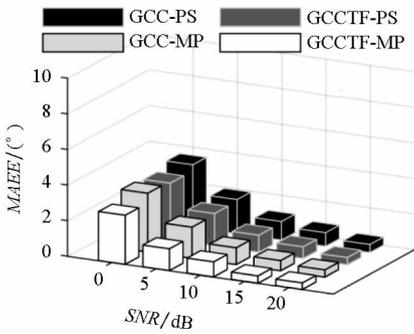
$$N_{\text{max}}(i) = \max(N(i), \hat{N}(i)) \quad (29)$$

其中 $N(i)$ 和 $\hat{N}(i)$ 分别表示第 i 次仿真中真实的和估计的声源数量。引入 $N_c(i)$ 表示第 i 次仿真中估计正确的声源数量,正确的标准是真实值和估计值之间的角度间隔小于 A_{\min} 。式(28)中,当 $n > N_c(i)$ 时, $|\hat{\theta}_n(i) - \theta_n(i)| = A_{\min}$ 。

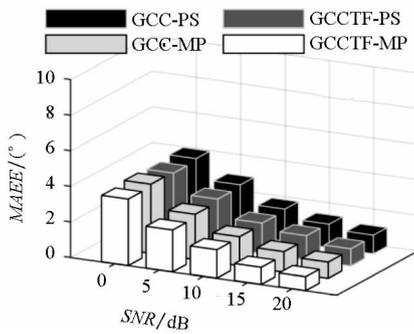
由图5给出的仿真结果可以看出, $RT_{60} = 0.2$ s时, MAEE的取值约为 $0.3^\circ \sim 4.7^\circ$ 。当SNR及声源数 N 相同时, GCC-PS和GCCTF-MP在4种算法中分别取得最高和最低的MAEE,对应当前环境下最差和最好的定位与计数性能。同一声源数下,图5(a)、图5(b)中4种算法的MAEE随SNR的增强而逐渐降低;当SNR相同,声源数 N 增加时,由于声源间相互干扰的加剧,4种算法的MAEE随之升高。当 $SNR > 0$ 时, GCC-MP的MAEE比GCCTF-PS的更低,而当 $SNR = 0$ 时,

GCCTF-PS 的相对更低,可见 GCCTF 中的局部 SNR 追踪模块在低 SNR 时能够有效提取出受噪声影响较小的时频支撑域,从而提供更好的性能。

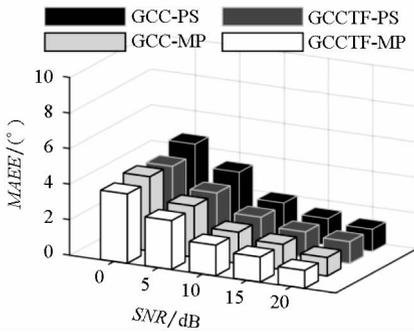
够有效地提取出单声源能量占优的时频支撑域,且声源数越多,作用越明显,在局部 SNR 追踪模块的辅助作用下,获得了相对更好的性能。



(a) $N=2$ 时随 SNR 变化的 MAEE
(a) MAEE versus SNR when $N=2$



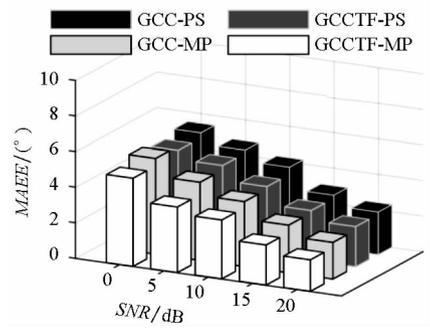
(b) $N=4$ 时随 SNR 变化的 MAEE
(b) MAEE versus SNR when $N=4$



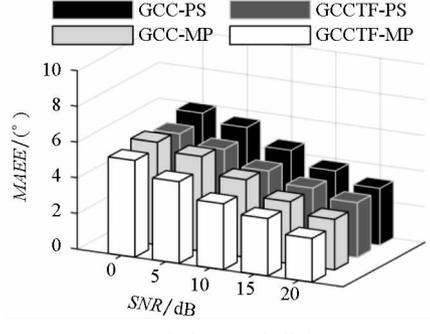
(c) $N=6$ 时随 SNR 变化的 MAEE
(c) MAEE versus SNR when $N=6$

图 5 $RT_{60}=0.2$ s 时,随 SNR 变化的 MAEE
Fig. 5 MAEE versus SNR when $RT_{60}=0.2$ s

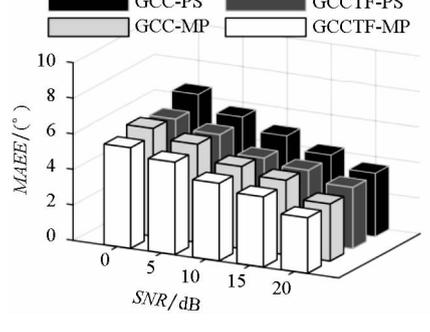
当 $RT_{60}=0.5$ s 时,由于混响程度的加剧,图 6 中整体 MAEE 相比 $RT_{60}=0.2$ s 时有所升高,取值约为 $1.7^\circ \sim 6.6^\circ$ 。与 $RT_{60}=0.2$ s 时的情况类似,4 种算法中 GCCTF-MP 仿真性能最好,且当 $N=2$ 时,GCCTF-PS 在 $SNR=0$ 时的性能优于 GCC-MP;而当 N 为 4 和 6 时,GCCTF-PS 分别在 $SNR \leq 10$ dB 和 $SNR \leq 15$ dB 时获得更优的性能。可见在较强的混响环境下,随着声源数的增加,WDO 假设越发难以满足,此时相干性检测模块能



(a) $N=2$ 时随 SNR 变化的 MAEE
(a) MAEE versus SNR when $N=2$



(b) $N=4$ 时随 SNR 变化的 MAEE
(b) MAEE versus SNR when $N=4$



(c) $N=6$ 时随 SNR 变化的 MAEE
(c) MAEE versus SNR when $N=6$

图 6 $RT_{60}=0.5$ s 时,随 SNR 变化的 MAEE
Fig. 6 MAEE versus SNR when $RT_{60}=0.5$ s

综上,可以得出结论:定位与计数方法 MP 相比 PS 更加精确;角度谱 GCCTF 相比传统 GCC 在较强混响、较低 SNR 以及声源数较多的情况下,可为后续定位计数环节提供更加稳健可靠的前端处理结果。同时采用角度谱 GCCTF 及 MP 方法的 GCCTF-MP 是讨论的 4 种算法中鲁棒性最好、精确度最高的定位与计数算法。

4 结论

本文提出了一种混响及噪声环境下基于角度

谱的定位及计数算法 GCCTF-MP。在角度谱构建部分,通过局部 SNR 追踪及相干性检测模块对 GCC 角度谱进行时频域滤波;在定位与计数部分,使用双宽度 MP 通过声源贡献迭代去除的方式替代 PS 中的单点谱峰比较。仿真实验证实了 GCCTF-MP 是一种在较低 SNR、较强混响以及声源数较多的情况下,相比于 GCC-PS 更加精确、稳健的定位与计数算法。实验中还发现 MP 中双宽度的选择会对最终的结果产生影响,后续研究将围绕该问题展开。

参考文献 (References)

- [1] Benesty J, Chen J D, Huang Y T. Microphone array signal processing [M]. Berlin: Springer Berlin Heidelberg, 2008.
- [2] Blandin C, Ozerov A, Vincent E. Multi-source TDOA estimation in reverberant audio using angular spectra and clustering [J]. Signal Processing, 2012, 92(8): 1950 - 1960.
- [3] Knapp C H, Carter G C. The generalized correlation method for estimation of time delay [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1976, 24(4): 320 - 327.
- [4] Champagne B, Bedard S, Stephenne A. Performance of time-delay estimation in the presence of room reverberation [J]. IEEE Transactions on Speech & Audio Processing, 1996, 4(2): 148 - 152.
- [5] Wang L, Hon T K, Reiss J D, et al. An iterative approach to source counting and localization using two distant microphones [J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2016, 24(6): 1079 - 1093.
- [6] Nesta F, Omologo M. Generalized state coherence transform for multidimensional TDOA estimation of multiple sources [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 246 - 260.
- [7] 房玉琢, 许志勇. 一种稳健的室内无模糊多声源 TDOA 估计算法 [J]. 电子与信息学报, 2016, 38(5): 1143 - 1150.
FANG Yuzhuo, XU Zhiyong. A robust algorithm for unambiguous TDOA estimation of multiple sound sources under indoor environment [J]. Journal of Electronics & Information Technology, 2016, 38(5): 1143 - 1150. (in Chinese)
- [8] Reddy V V, Khong W H, Ng B P. Unambiguous speech DOA estimation under spatial aliasing conditions [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(12): 2133 - 2145.
- [9] Loesch B, Yang B. Source number estimation and clustering for underdetermined blind source separation [C]// Proceedings of International Workshop for Acoustic Echo and Noise Control (IWAENC), 2008.
- [10] Pavlidi D, Griffin A, Puigt M, et al. Real-time multiple sound source localization and counting using a circular microphone array [J]. IEEE Transactions on Audio Speech & Language Processing, 2013, 21(10): 2193 - 2206.
- [11] Mohan S, Lockwood M E, Kramer M L, et al. Localization of multiple acoustic sources with small arrays using a coherence test [J]. Journal of the Acoustical Society of America, 2008, 123(4): 2136 - 2147.
- [12] Mallat S G, Zhang Z F. Matching pursuits with time-frequency dictionaries [J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397 - 3415.
- [13] Gustaffson T, Rao B D, Trivedi M. Source localization in reverberant environments; modeling and statistical analysis [J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(6): 791 - 803.
- [14] Yilmaz O, Rickard S. Blind separation of speech mixtures via time-frequency masking [J]. IEEE Transactions on Signal Processing, 2004, 52(7): 1830 - 1847.
- [15] Lehmann E, Johansson A. Prediction of energy decay in room impulse responses simulated with an image-source model [J]. Journal of the Acoustical Society of America, 2008, 124(1): 269 - 277.