一种稳健的室内无模糊多声源 TDOA 估计算法

房玉琢 许志勇*
(南京理工大学电子工程与光电技术学院 南京 210094)

摘 要：该文针对室内环境下的宽间距多声源到达时间差(TDOA)估计问题，研究了一种基于近似核密度估计(KDE)的无模糊算法。根据声频信号的短时频谱稀疏性，利用相关性检测(CT)提取单个声源能量占优的时频支撑域，进而将观测信号的归一化互功率谱(NCS)所构建的近似核函数通过累加平均削弱室内混响的干扰，同时引入多阶段(MS)分频带处理有效解决宽间距时的空域模糊。理论推导及仿真研究验证了该算法是一种稳健的室内无模糊多声源 TDOA 估计算法。

关键词：语音信号处理；麦克风阵列；归一化互功率谱；相关性检测；近似核密度函数；无模糊到达时间差估计
中图分类号： TN912.3 **文献标识码**： A **文章编号**：1009-5896(2016)05-1143-08
DOI: 10.11999/JEIT150824

A Robust Algorithm for Unambiguous TDOA Estimation of Multiple Sound Sources under Indoor Environment

FANG Yuzhuo XU Zhiyong
(*School of Electronic and Optical Engineering*, *Nanjing University of Science and Technology*, *Nanjing* 210094, *China*)

Abstract: For Time Difference Of Arrival (TDOA) estimation of multiple sound sources with wide spacing under indoor environment, an unambiguous algorithm based on approximated Kernel Destiny Estimator (KDE) is studied. According to the short-time spectral sparseness of audio signals, the time-frequency bin with energy dominance of a single source is extracted from Coherence Test (CT), then an approximated kernel function constructed of Normalized Cross-Spectrum (NCS) of obtained signals is used to weaken the interference of indoor reverberation with cumulative average, while adding Multi-Stage (MS) to divide the frequency band, the spatial ambiguity with wide spacing can be solved effectively. This algorithm is verified as an unambiguous TDOA estimation algorithm of multi-source under indoor environment by both theoretical derivation and simulation results.

Key words: Speech signal processing; Microphone array; Normalized Cross-power Spectrum (NCS); Coherence Test (CT); Approximate kernel density function; Unambiguous Time Difference Of Arrival (TDOA) estimation

1 引言

通过麦克风阵列中各对阵元间的空域多样性估计多个声源的到达时间差(Time Difference of Arrival, TDOA)是近期研究的热点。它可对后续声源的测向、定位、动态追踪以及频域盲源分离(Blind Source Separation, BSS)中的转置模糊修正起到关键性的作用，广泛应用于视频会议、电视游戏、人机交互等领域。

室内环境下，基于一对麦克风阵元的 TDOA 估计通常面临混响效应和空域模糊两个问题的综合影响。其中混响效应将回声以及频谱衰减引入观察信号；由于窄阵元间距时较低的 TDOA 域分辨率难以获得精确的估计，有必要提高间距，然而突破信号波长之半限制的宽间距将高频混叠引入估计函数。这两者都将对 TDOA 估计产生影响，此时，经典的基于理想传播模型的相位变换广义互相关(Generalized Cross-Correlation PHAse Transform, GCC-PHAT)法[1]性能出现恶化[2]。

近年来，BSS 领域兴起了两类 TDOA 估计算法。第 1 类方法基于独立变量分析(Independent Component Analysis, ICA)[3]，如平均方向图(Average Directivity Pattern, ADP)[4]、状态相关变换(State Coherence Transform, SCT)[5]等，文献[6]提出了近似核密度估计(Kernel Density Estimator,

收稿日期：2015-07-09；改回日期：2015-12-18；网络出版：2016-02-19
*通信作者：许志勇 ezyxu@mail.njust.edu.cn
基金项目：国家自然科学基金(61171167, 61401203)，江苏省自然科学基金(BK20130776)
Foundation Items: The National Natural Science Foundation of China (61171167, 61401203), Natural Science Foundation of Jiangsu Province (BK20130776)

KDE)算法对文献[5]中的非线性函数进行改进，使得谱峰更加尖锐。上述算法大多没有考虑空域混叠造成的模糊问题，因此，文献[7]引入多阶段(Multi Stage, MS)处理环节对文献[6]中方法作出改进，分频带抑制空域模糊。第 1 类算法对多个统计独立的声源所构成的卷积混合信号进行频域 ICA，再由分离矩阵提取声源的 TDOA 信息，因此矩阵收敛所需的观察数据量将对算法的实际运用产生限制。

第 2 类方法基于声频信号的短时频谱稀疏特性，如互谱加权直方图法[8]、归一化时频聚类[9]等，文献[10]使用归一化互功率谱(Normalized Cross-power Spectrum, NCS)替代文献[6]中 ICA 环节，从而避免了矩阵收敛问题，便于算法的实时处理。上述方法均是基于每个时频支撑域单个声源能量占优的近似假设，实际中该条件并不能够严格满足，因此，文献[11]提出相关性检测(Coherence Test, CT)去除上述假设的限制。以上所列方法同样没有考虑空域模糊问题。我们通过对文献[8]中方法的改进，利用 TDOA 模糊周期的频变特性，迭代求取互谱直方图，从而解决空域模糊问题[12]，然而该方法基于理想传播模型，混响环境下，性能将出现恶化。

为了有效解决室内环境下的多声源 TDOA 估计问题，本文基于声频信号的短时频谱稀疏特性提出 NCS-KDE-MS 算法。引入 CT 环节[11]去除每个时频支撑域内单个声源能量占优的假设限制，增强算法的普适性；进而通过文献[10]中的 NCS-KDE 算法构建统计近似核函数模型，累加平均削弱室内混响造成的干扰；同时使用较宽的阵元间距克服窄间距时出现的估计不精确问题，并针对由此产生的高频混叠，引入 MS 分频带处理环节进行抑制，从而实现稳健的无模糊 TDOA 估计。本文算法针对室内宽间距 TDOA 估计，具有抗混响、解空域模糊等优良特性。

本文具体安排如下：第 2 节由短时频谱稀疏特性通过建立近似核函数介绍 NCS-KDE 算法的基本原理；第 3 节分别引入 CT 判决以及 MS 分频带处理两个环节对 NCS-KDE 进行改进，进而提出 NCS-KDE-MS 算法，并给出其具体步骤和流程图；计算机仿真、性能分析与结论分别于第 4～6 节中给出。

2 NCS-KDE 算法原理

假设发射端 $\{s_n(t), n=1,2,\cdots,N\}$ 表示 N 个声源所产生的声源信号，接收端 $\{x_m(t), m=1,2,\cdots,M\}$ 表示 M 个阵元所组成的麦克风阵列观测到的混合信号，信号采样率为 f_s，通过 N_{FFT} 点短时傅里叶变换(Short-Time Fourier Transform, STFT)将上述时域信号转换到离散时频域，则变换后的声源信号及观察信号分别由 $\boldsymbol{S}(r,k)=[S_1(r,k), S_2(r,k),\cdots,S_N(r,k)]^{\mathrm{T}}$，$\boldsymbol{X}(r,k)=[X_1(r,k),X_2(r,k),\cdots,X_M(r,k)]^{\mathrm{T}}$ 表示，其中 $S_n(r,k)$, $X_m(r,k)$表示第 k 个离散频率、第 r 帧处的复值系数，$r\in\{1,2,\cdots,N_r\}$，$k\in\{N_{kL},\cdots,N_{kH}\}$，其中 N_r 表示观察时间内的信号总帧数，N_{kL} 与 N_{kH} 分别表示所考虑的频率范围内最低和最高频率的序号。于是系统的传播模型为

$$\boldsymbol{X}(r,k)=\boldsymbol{H}(k)\boldsymbol{S}(r,k) \tag{1}$$

式中 $\boldsymbol{H}(k)$表示第 k 个频率处声源与麦克风阵元间的传递函数矩阵，其表达式可写为

$$\boldsymbol{H}(k)=\left[\left|H_{mn}(k)\right|\exp(-\mathrm{j}2\pi f_k T_{mn}(k))\right]_{m,n} \tag{2}$$

式中 $\left|H_{mn}(k)\right|$ 表示第 n 个声源与第 m 个阵元间频率响应的幅度，$T_{mn}(k)$表示与之相应的声源传播时间，$f_k=kf_s/N_{\mathrm{FFT}}$ 表示第 k 个频率段中心所对应的实际频率。

假设每个时频支撑域内最多只可能有一个声源能量占主导地位，由此可得

$$\boldsymbol{X}(r,k)=\begin{bmatrix}\left|H_{1n_d(r,k)}(k)\right|\exp\left(-\mathrm{j}2\pi f_k T_{1n_d(r,k)}\right)\\ \vdots\\ \left|H_{Mn_d(r,k)}(k)\right|\exp\left(-\mathrm{j}2\pi f_k T_{Mn_d(r,k)}\right)\end{bmatrix}\cdot s_{n_d(r,k)}(r,k) \tag{3}$$

其中 $n_d(r,k)$表示在任一时频支撑域(r,k)处占主导地位的声源序号。考虑阵列中任意一对麦克风阵元 $(a,b)\in[(a,b)\,|\,(1\le a<b\le M)]$，在$(r,k)$处接收信号的归一化互功率谱 NCS 可写为

$$\mathrm{NCS}_{ab}(r,k)=\frac{X_a(r,k)X_b^*(r,k)}{\left|X_a(r,k)X_b^*(r,k)\right|} \tag{4}$$

其中“*”表示复共轭运算，将式(3)代入式(4)中，可得

$$\begin{aligned}&\mathrm{NCS}_{ab}(r,k)\\&=\exp\left(-\mathrm{j}2\pi f_k\left(T_{a,n_d(r,k)}(k)-T_{b,n_d(r,k)}(k)\right)\right)\\&=\exp\left(-\mathrm{j}2\pi f_k\tau_{(ab)}^{n_d(r,k)}(k)\right)\end{aligned} \tag{5}$$

式中 $\tau_{(ab)}^{n_d(r,k)}(k)$ 表示(r,k)处声源 $n_d(r,k)$到麦克风对(a,b)的直达波 TDOA，在无混响的情况下，该参数只和麦克风与声源间的位置有关，与频率无关，因而可写成 $\tau_{(ab)}^{n_d(r,k)}$。实际室内环境中混响的干扰不可避免，不同方向入射角的反射波在接收阵元处混合叠加，式(5)不能够严格满足，此时我们采用文献[13]中所提出的近似混合模型，将混响看作一种空间均匀散布的噪声，通过如下的 KDE 算法[6]建立统计模型求取声源的 TDOA。

将所需估计的声源直达波 TDOA 看作一种满足给定概率分布的随机变量，用 τ 表示，首先使用

式(6)所示的非线性变换

$$f(\tau)=\frac{\exp\left(-\left|\exp(-\mathrm{j}2\pi f_k\tau)-\mathrm{NCS}_{ab}(r,k)\right|^2\Big/\left(2B_k^2\right)\right)}{2\pi f_k} \tag{6}$$

由各个时频支撑域(r,k)处$\exp(-\mathrm{j}2\pi f_k\tau)$与式(6)中$\mathrm{NCS}_{ab}(r,k)$的绝对误差值构建近似高斯核密度函数[7]，式中$B_k=\tau_{\max}/B$为该核函数的带宽，其中$\tau_{\max}=d/c$为最大可能的 TDOA($d$为阵元间距，$c$为声传播速度)，$B$为影响 TDOA 域分辨率的因子，然后将所得函数时域累加并求平均得到各个频率处的窄带谱。

$$\Phi_k(\tau)=\frac{1}{N_r}\sum_{r=1}^{N_r}f(\tau) \tag{7}$$

最后将所得窄带谱按频率叠加并求平均，得到关于的 NCS-KDE 密度谱函数[11]。

$$\Phi_{\text{NCS-KDE}}(\tau)=\frac{1}{N_{kH}-N_{kL}+1}\sum_{k=N_{kL}}^{N_{kH}}\Phi_k(\tau) \tag{8}$$

由于混响被视作空间散布的噪声，通过式(6)中近似核函数的非线性特性以及式(7)和式(8)的统计累加平均效应可以弱化混响对 TDOA 估计结果的影响，从而在正确的 TDOA 处形成峰值，因此上述 NCS-KDE 算法能够提供较好的抗混响性能[6,10]。

3　NCS-KDE-MS 算法原理

3.1 相干检测(Coherence Test, CT)判决

实际中，对于每个时频支撑域(r, k)，单个声源能量占优的假设条件并不能够严格满足，因此我们将 CT 判决[11]引入 NCS-KDE，去除算法中该假设条件的限制，具体实现方式如下，首先求得 a,b 阵元间时频域观察信号的协方差矩阵：

$$\begin{aligned}\mathrm{COV}_{ab}(r,k)&=E\left[\boldsymbol{X}_{ab}(r,k)\boldsymbol{X}_{ab}^{\mathrm{H}}(r,k)\right]\\&\approx\frac{1}{C}\sum_{l=r-C+1}^{r}\boldsymbol{X}_{ab}(l,k)\boldsymbol{X}_{ab}^{\mathrm{H}}(l,k)\end{aligned} \tag{9}$$

式中$\boldsymbol{X}_{ab}(r,k)=[X_a(r,k),X_b(r,k)]^{\mathrm{T}}$，$C$表示近似协方差矩阵所使用的帧数。当上述矩阵的秩为 1 时，可以认为在此矩阵所对应的支撑域(r, k)，只有单个声源起主导作用，判断矩阵秩为 1 的方法有多种，本文通过对上述矩阵作特征值分解，求出其中较大及较小的特征值分别对应$\sigma_{\max}^2$和$\sigma_{\min}^2$，使用信干噪比(Signal-Interference-Noise Ratio, SINR)即$(\sigma_{\max}^2-\sigma_{\min}^2)/\sigma_{\min}^2$作为秩是否为 1 的判据，理想情况下，当矩阵秩为 1 时，SINR 的值无穷大，实际中，由于噪声及弱声源干扰的存在，SINR 为一非负实数，因此，我们可以根据应用的实际环境设置经验门限 Thd 并使用如式(10)的判决公式：

$$f(\tau)\big|_{(r,k)}=\begin{cases}f(\tau), & \mathrm{SINR}\big|_{(r,k)}\geq\mathrm{Thd}\\0, & \text{其他}\end{cases} \tag{10}$$

通过式(10)将所有秩为 1 的矩阵所对应的(r,k)提取出来，使得所有满足条件的(r, k)参与核密度函数的累加，而其他(r,k)不参与累加。

3.2 多阶段(Multi-Stage, MS)分频带处理

基于一对麦克风阵元的 TDOA 估计中，较窄的阵元间距 d带来的低分辨率会使 TDOA 算法的估计精确度较差，为了获得较高的分辨力，需要提高 d，然而 d 的变宽势必会打破最小信号波长$\lambda_{\min}$之半的限制，此时高频混叠所引起的相位绕卷模糊会对 TDOA 的正确估计产生干扰。虽然近似核密度函数引入的与频率相关的$1/(2\pi f_k)$加权因子一定程度上能够抑制高频混叠对累加结果的影响，但这种影响并不能完全消除[8]。

因此，引入如下的 MS 处理环节[7]。假设所考虑的频率区间为$[f_L,f_H]$，其中 f_L和 f_H分别为所考虑频率范围内的最低及最高频率，由空域 Nyquist 采样定理，可得最大不模糊频率：

$$f_{\mathrm{UA}}=c/(2d) \tag{11}$$

则f_{UA}将整个频率区间划分为$[f_L,f_{\mathrm{UA}}]$和$[f_{\mathrm{UA}},f_H]$两个子频带，其中第 1 个子频带并不存在高频混叠的影响，当 $f_H>2f_{\mathrm{UA}}$时，第 2 个子频带又可分为$[f_{\mathrm{UA}},2f_{\mathrm{UA}}]$和$[2f_{\mathrm{UA}}, f_H]$两个部分，以此类推，整个频率区间将可拆分成

$$P=\left\lceil\frac{f_H-f_L}{f_{\mathrm{UA}}}\right\rceil \tag{12}$$

个子频带，式中$\lceil\bullet\rceil$表示向上取整运算符。

我们首先求得各个子频带内的谱密度函数：

$$\Phi^{(p)}(\tau)=\sum_{k=N_{pL}}^{N_{pH}}\Phi_k(\tau),\quad p=1,2,\cdots,P \tag{13}$$

其中 N_{pL}, N_{pH}分别为第 p 个子频带所对应的最低频率及最高频率的序号。对于每个声源，第 1 个子频带对应的谱密度函数$\Phi^{(1)}(\tau)$不存在高频混叠的影响，$\Phi^{(p+1)}(\tau)$将比$\Phi^{(p)}(\tau)$ 至多多出一个由于混叠所产生的伪峰。将$\Phi^{(2)}(\tau)$与$\Phi^{(1)}(\tau)$相乘，将能有效抑制第 2 个子频带中伪峰的影响，此时将受到加权抑制混叠的第 2 个子频带的谱密度$\Phi^{(2)}(\tau)\Phi^{(1)}(\tau)$作为下一个频带的加权因子，以此类推，直至整个讨论的频率范围。

3.3 NCS-KDE-MS 算法

将 3.1 节中所提 CT 判决以及 3.2 节中所提 MS 分频带处理环节同时运用到 NCS-KDE 算法中，可以得到本文所提出的 NCS-KDE-MS 算法，该方法去除了单个声源能量占优假设条件的限制，包含

NCS-KDE 算法较强的抗混响特性，同时吸收了 MS 环节解空域模糊的优点，其相应的谱密度函数为

$$\Phi_{\text{NCS-KDE-MS}}(\tau)=\prod_{p=1}^{P}\Phi^{(p)}(\tau) \tag{14}$$

对比式(8)和式(14)，可以看出相比于 NCS-KDE，该算法的计算量并没有改变，均约为 $(N_{\text{CF}}\times N_{\tau}\times N_{r}\times(N_{kH}-N_{kL}+1))$，其中，$N_{\text{CF}}$ 表示计算 $f(\tau)$ 所需的运算次数，N 表示给定分辨率的情况下 τ 的离散采样个数。

我们将所得谱密度函数归一化，从而由函数的谱峰位置求得多源 TDOA 估计，具体表达式为

$$\hat{\tau}_n=\arg\max_{\tau\in\{\tau_n\}}\frac{\Phi_{\text{NCS-KDE-MS}}(\tau)}{\max(\Phi_{\text{NCS-KDE-MS}}(\tau))} \tag{15}$$

式中 $\{\tau_n\}$，$n=1,2,\cdots,N$ 表示第 n 个声源所对应的 TDOA 参数空间。

综上，对于接收端一对麦克风观察信号 $x_m(t)$, $m=a,b$, NCS-KDE-MS 算法的流程图如图 1 所示，相应的处理步骤归纳如下：

(1)将观察信号作 STFT，得到其离散时频域的表示式 $X_a(r,k)$, $X_b(r,k)$；

(2)由式(4)求得各个离散时频支撑域(r,k)处的归一化互功率谱 $\text{NCS}_{ab}(r,k)$；

(3)将每个频率处的 NCS 按式(6)构建近似核函数，由式(10)通过 CT 判决检测出秩为 1 的协方差矩阵所对应的(r,k)，并在时域累加平均，得到窄带谱密度函数 $\Phi_k(\tau)$ ；

(4)根据式(13)将每个子频带中所有频率处的谱密度函数求和，得到 $\Phi^{(p)}(\tau)$, $p=1,2,\cdots,P$ ；

(5)由式(14)将所有子频带所对应的谱密度函数联立连乘，求得 $\Phi_{\text{NCS-KDE-MS}}(\tau)$，再通过式(15)的归一化密度谱峰值求得各个声源的 TDOA 估计。

4 计算机仿真

仿真中采用的声源信号来自由 18 个男声、17 个女声所构成的纯净语音数据库，长度约 2 s(平均功率相同)，采样率 $f_s=16$ kHz。室内信道冲激响应由镜像法[14]产生，房间尺寸为 6 m×5 m×3 m。图 2 给出了麦克风阵元与声源位置关系的室内 2 维平面示意图，其中两个声源 S_1,S_2 位置向量分别为$[1.9, 2.9, 1.5]^{\text{T}}$和$[3.4, 2.9, 1.5]^{\text{T}}$，两个全向麦克风 a, b 分别为$[2.4, 1.6, 1.5]^{\text{T}}$和$[2.4+d, 1.6, 1.5]^{\text{T}}$，位置向量的单位均为 m。室内环境中声速 c 取 344 m/s，实验中，我们将接收端 SNR 为 20 dB 的观察信号通过 $f_L=200$ Hz 到 $f_H=4000$ Hz 的带通滤波器，在此频率范围内，B 取常数 20 足够精确[6]，此时，信号的最小波长 $\lambda_{\min}=c/f_H=0.086$ m。我们将滤波后的观察信号做 1024 点(合 64 ms)汉宁窗加权的 STFT，帧移 25%(合 16 ms)。CT 检测中，$C=5$, Thd=8 dB。为了比较的公平性，对 3 种算法，我们均取其中前 60 帧(约合 1.024 s)观察数据做函数密度谱的时域累加。

我们讨论当混响时间 T_{60} 分别为 250 ms, 500 ms，阵元 a,b 在间距 d 分别为 $0.5\lambda_{\min}$ (0.043 m), $2.5\lambda_{\min}$ (0.215 m), $4\lambda_{\min}$ (0.344 m)时，对两个不同声源 S_1,S_2 的 TDOA 估计性能，同时与经典的 GCC-PHAT 及 NCS-KDE[10]作比较。所用声源信号及其时频谱分别如图 3 和图 4 所示，图 3(a), 3(b)分别给出了 S_1, S_2 的时域波形，图 4(a), 4(b)分别表示相应的短时频谱图。由图 4 可以看出两个语音信号的短时频谱稀疏性明显，每个声源的能量都集中在比例很小的时频支撑域内，我们使用 CT 处理环节保证统计累加时每个时频支撑域内最多只有一个声源能量占主导地位。图 5～图 7 给出了不同间距 d 时的 3 组 TDOA 估计结果，其中标示为 AL_1(下三角)，AL_2(圆圈)，AL_3(上三角)的实线分别对应 GCC-PHAT, NCS-KDE, NCS-KDE-MS，左右两个箭头分别标示出 S_1, S_2 正确的 TDOA 位置，图中的横坐标以最大可能的 TDOA $\tau_{\max}$ 为基准进行了归一化处理。

由图 5 可以看出，当 $d=0.5\lambda_{\min}$ 时，虽然不存在高频混叠造成的空域模糊问题，然而过低的分辨

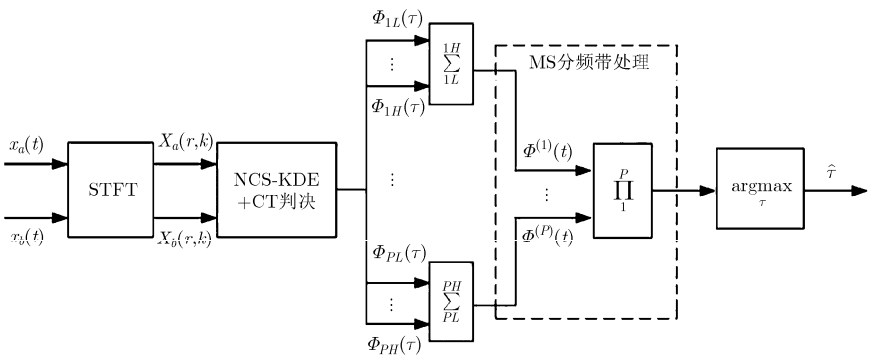

图 1 NCS-KDE-MS 算法流程图

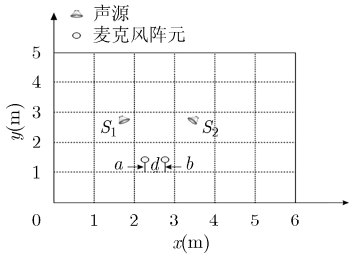

图2 室内环境平面示意图

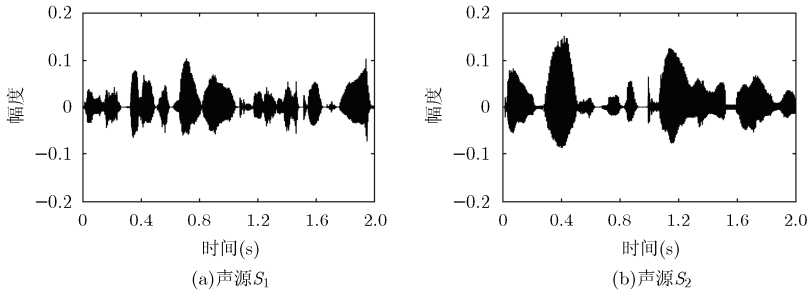

图3 时域波形图

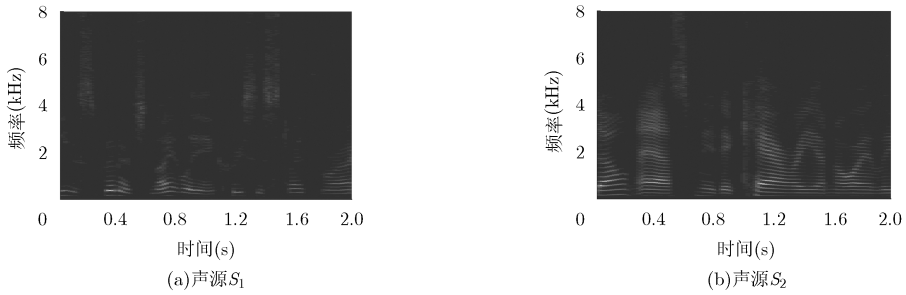

图 4 短时频谱图

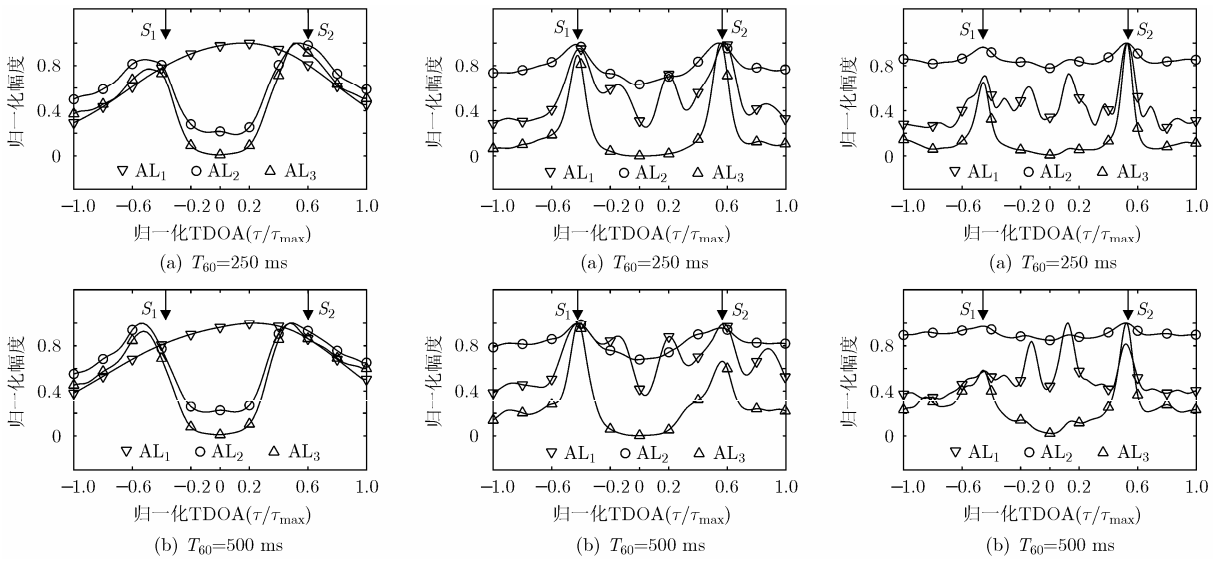

图 5　d =0.5 $\lambda_{\min}$ 时的TDOA估计结果　　图 6　d =2.5 $\lambda_{\min}$ 时的TDOA估计结果　　图 7　d =4.0 $\lambda_{\min}$ 时的TDOA估计结果

率使得各个算法的主瓣过胖，GCC-PHAT 算法仅有一个谱峰，而两类 KDE 算法虽然形成了两个谱峰，然而峰值位置与正确的 TDOA 位置存在不小的误差。可见，较低的分辨率会对算法的准确估计产生影响，因此需要加宽 d 来改善算法的性能。

图 6 中，当 $d = 2.5\,\lambda_{\min}$ 时，分辨率的提高使得 3 种算法均能够在正确的 TDOA 位置处形成最高的两个峰值。然而同时引入的高频混叠使得 GCC-PHAT 的密度谱中出现了伪峰，且对比图 6(a)和图 6(b)，可以看出，混响的增强，使得 GCC-PHAT 谱的伪峰高度变大，接近正确位置的峰值，因此产生错误估计的风险较大。而两类 KDE 算法由于加权因子对高频混叠的抑制，波形相对平坦，产生的伪峰数量较少。同时，我们注意到，NCS-KDE 的谱基底(0.7, 0.8 左右)相比 GCC-PHAT(0.3, 0.4 左右)以及 NCS-KDE-MS(0, 0.2 左右)大得多，估计的辨识度较差。由此可见加权因子虽然可以抑制混叠，但最终的频域累加使得密度谱基底过高，辨识度变差，而 MS 的分频带处理基本消除了高频混叠对谱基底的累加干扰，带来了更好的辨识效果。

随着 d 的变宽，空域混叠的增强使得图 7 中GCC-PHAT 谱中的伪峰数量进一步增多，在较低混响时，GCC-PHAT 谱中的伪峰高度就已经超过了正确位置，造成了错误的 TDOA 估计。同时我们可以发现两类 KDE 算法的虽然波形相对平坦，但空域混叠的增强使得 NCS-KDE 的谱基底进一步提高，TDOA 估计的辨识度进一步降低，此时，伪峰的高度与两个正确位置的峰值非常接近；而 NCS-KDE-MS 的谱基底几乎不受 d 变宽的影响，只在混响增强时有了些许提高，这体现了该算法抗空域模糊的良好特性。

5 性能分析与评估

通过第 4 节单组声源的计算机仿真，我们初步认识了算法在较宽间距时的解模糊特性，为更充分地分析与评估算法的性能，我们将 $0.5\lambda_{\min}\sim5\lambda_{\min}$ 范围内间隔 $0.5\lambda_{\min}$ 的 10 组 d 均做 1000 次仿真，每次仿真从数据库 35 个声源中顺次取出 2 个不同的声源(取共计 1000 组)，仿真所用的其他参数设置与第 4 节相同，使用式(16)～式(18)所示的均方根误差(Root Mean Square Error, RMSE)、TDOA 检测百分比(Percentage of TDOA, PTDOA)[6]、以及准确率(Precision)[7,15]3 个统计指标来量化比较。

$$\mathrm{RMSE}=\sqrt{\sum_{u=1}^{N_u}e(u)^2/N_u} \tag{16}$$

式中，单次误差$e(u)$为两个声源TDOA相对误差的平均值，N_u表示仿真次数(此处为1000)。

$$\mathrm{PTDOA}=N_{\mathrm{CO}}/(N_uN_S) \tag{17}$$

表示检测出声源TDOA位置的正确率百分比，其中N_{CO}表示能够检测出正确TDOA位置的次数，以TDOA估计位置与实际TDOA位置的相对误差不大于5%为基准判断是否正确[6]，N_S表示声源个数。

$$\text{准确率}=\sum_{u=1}^{N_u}\left(\widehat{N}_{\mathrm{SP}}(u)/N_{\mathrm{TP}}(u)\right)\Big/N_u \tag{18}$$

是一个表征TDOA密度谱辨识度的量化指标，其定义与文献[7,15]中角度谱的情况类似，其中$\widehat{N}_{\mathrm{SP}}(u)$表示单次仿真归一化密度谱中所估计出的正确的TDOA谱峰的个数(正确与否的标准同PTDOA)，$N_{\mathrm{TP}}(u)$表示幅值大于0.2的谱峰个数。

表1，表2分别给出了$T_{60}=250$ ms, 500 ms时3种算法的RMSE，其中标示下划线的数据表示大于5%的数值。由两表可以看出，$d\le1.5\lambda_{\min}$时，较低的TDOA域分辨率使得算法的RMSE较大，可见，较窄间距时的低分辨率不适用于多源测向。由表1，$d>1.5\lambda_{\min}$时，随着d的变宽，分辨率的提高使两类KDE算法的RMSE有降低的趋势，可见提高算法的分辨率对于降低算法误差的重要性，而GCC-PHAT由于高频混叠的影响，RMSE仍然很大。表2中，较强混响环境下，$d\ge3\lambda_{\min}$时，NCS-KDE的RMSE均高于5%，究其原因，空域模糊的增强使NCS-KDE算法的谱基底提升明显，此时伪峰高度高于正确位置的情况出现，估计出的TDOA值偏离正确位置较多，因而产生了较大的误差。而NCS-KDE-MS由于MS分频带处理环节的作用，谱基底相对低得多，因此能够更好地抑制伪峰高度的增长，从而有效控制估计性能的恶化，将RMSE值始终控制在一个合理的范围，因而较NCS-KDE更加稳健。

表3，表4分别给出了$T_{60}=250$ ms, 500 ms时3种算法的PTDOA。综合两表可以看出，$d\le1.5\lambda_{\min}$时，3种算法的PTDOA较低，其中GCC-PHAT均为0，无法估计出正确的TDOA。$d\ge2\lambda_{\min}$时，表3中随着d的变宽，GCC-PHAT的PTDOA仍然较低，而两种KDE类算法均在80%以上，准确率较高，且NCS-KDE-MS略好于NCS-KDE；混响增强时，由于谱基底的提高，伪峰高度超过正确位置峰值的风险加大，表4中NCS-KDE的PTDOA降低较为明显，而NCS-KDE-MS保持在75%以上，体现了该算法在较强混响与宽间距综合作用的环境中稳健的估计性能。

图 8 给出了 3 种算法的准确率指标，图中实线的标示含义与图 5～图 7 相同。由图 8(a)可以看出，$d\le1.5\lambda_{\min}$时，3 种算法的准确率较差，随着分辨率的提高，准确率有增大的趋势，$2\lambda_{\min}\le d\le3.5\lambda_{\min}$时，NCS-KDE-MS 的准确率稳定在 1 附近，远高于 NCS-KDE 以及 GCC-PHAT，当 d 进一步变宽时，高频混叠的增强使得算法的准确率降低。由图 8(b)，当混响增强时，准确率相比于低混响时有所降低，然而 NCS-KDE-MS 仍明显好于其他两种算法。

综上，GCC-PHAT 在较窄的 d 时面临 TDOA 分辨率过低，估计误差过大的问题，且 d 变宽时所带来的空域模糊也会对算法的性能产生不可忽视的影响。而两种 KDE 类算法相比于 GCC-PHAT 有更好的抑制混叠的特性，其中 NCS-KDE-MS 算法几乎不受到空域模糊的影响，在较强混响时能够有效地控制 NCS-KDE 算法中谱基底过高引起的性能恶化，从而给出较为精确的 TDOA 估计，因而算法的 RMSE 和 PTDOA 两个指标均好于 NCS-KDE。同时 NCS-KDE-MS 在准确率这一指标上要明显好于 NCS-KDE，说明该算法具有良好的 TDOA 估计辨识度，为后续的其他信号处理环节提供了更加稳健良好的前端处理结果。

表 1　T_{60} = 250 ms 时的 RMSE(%)

阵元间距 $d(\times\lambda_{\min})$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
GCC-PHAT	单峰	25.41	22.67	15.43	53.15	51.62	53.40	52.57	51.35	46.25
NCS-KDE	17.26	13.67	8.01	3.48	3.42	3.45	2.74	2.38	1.37	1.24
NCS-KDE-MS	14.13	9.15	4.63	2.34	1.67	2.59	2.42	1.62	1.26	1.17

表 2　T_{60} = 500 ms 时的 RMSE(%)

阵元间距 $d(\times\lambda_{\min})$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
GCC-PHAT	单峰	50.74	43.57	36.14	62.15	57.14	71.42	69.99	64.71	58.45
NCS-KDE	24.83	21.05	17.25	5.95	4.46	6.94	6.37	6.37	12.14	12.06
NCS-KDE-MS	21.48	15.74	10.32	4.83	3.64	4.62	4.54	4.36	5.59	5.70

表 3　T_{60} = 250 ms 时的 PTDOA(%)

阵元间距 $d(\times\lambda_{\min})$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
GCC-PHAT	0	0	0	74.90	49.55	57.65	51.85	60.30	58.85	59.15
NCS-KDE	24.75	33.10	58.45	84.25	89.85	85.40	92.60	94.35	100	100
NCS-KDE-MS	32.40	40.10	75.45	94.65	100	93.15	93.75	100	100	100

表 4　T_{60} = 500 ms 时的 PTDOA(%)

阵元间距 $d(\times\lambda_{\min})$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
GCC-PHAT	0	0	0	0	24.75	31.15	23.05	25.45	25.85	30.65
NCS-KDE	22.65	30.45	33.25	67.75	78.45	60.70	62.15	62.45	51.45	52.15
NCS-KDE-MS	26.45	33.35	38.95	78.85	83.80	75.65	81.10	83.55	75.25	75.10

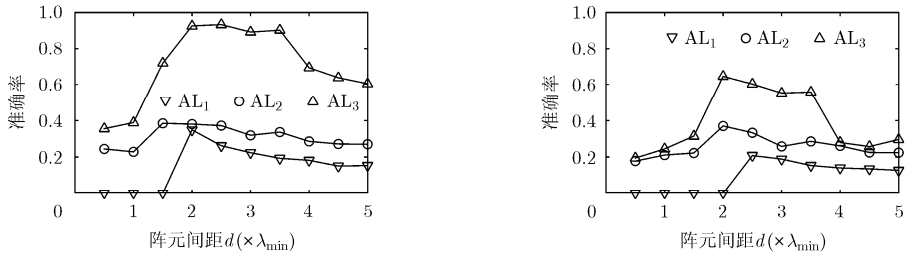

图 8 准确率仿真结果

6　结束语

针对室内环境下的多声源 TDOA 估计问题，本文提出的 NCS-KDE-MS 采用较宽的间距以提高算法的 TDOA 估计精确度，使用 CT 判决去除单声源能量占优这一假设条件的限制，进而将 NCS 替代 ICA 以避免实际应用中矩阵收敛的问题，再通过近似核密度函数时域累加平均从而降低混响造成的干扰，最后将窄带密度函数分频带累加，所得结果联立求积，以克服累加结果谱基底过高所造成的低辨识度问题，从而有效地抑制了宽间距时空域混叠所引起的估计模糊问题。通过 RMSE, PTDOA 以及量化表征辨识度的准确率这 3 个统计指标充分说明了该方法不仅具有良好的抗混响特性，而且有效地消除了高频混叠造成的空域模糊，相较于 GCC-PHAT, NCS-KDE-MS[10]是一种室内环境下稳健的 TDOA 估计算法。然而当声源信号频谱成分集中在较高的频率时，阵元间距过宽将导致低频无模糊段谱能量非常微弱(如鸟声)，此时，性能的恶化将限制间距的进一步变宽，进而影响最终结果，因此本文方法适用于语音等包含较多低频能量的声信号。

参 考 文 献

[1] KNAPP C H and 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.

[2] TSIAMI A, KATSAMANIS A, MARAGOS P, *et al*. Experiments in acoustic source localization using sparse arrays in adverse indoors environments[C]. Proceedings of 2014 European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, 2014: 2390–2394.

[3] 张超, 吴小培, 吕钊. 基于独立分量分析的运动目标检测算法中对通道数选择和观测向量生成方式的实验和分析[J]. 电子与信息学报, 2015, 37(1): 137–142. doi: 10.11999/JEIT140197.

ZHANG Chao, WU Xiaopei, and LÜ Zhao. Experiments and analysis on observation vector generation and channel number selection in motion detection algorithm based on independent component analysis[J]. *Journal of Electronics & Information Technology*, 2015, 37(1): 137–142. doi: 10.11999/JEIT140197.

[4] LOMBARD A, ZHENG Y, BUCHNER H, *et al*. TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, 19(6): 1490–1503.

[5] NESTA F, SVAIZER P, and OMOLOGO M. Cumulative state coherence transform for a robust two-channel multiple source localization[C]. Proceedings of the 8th International Conference on Independent Component Analysis and Signal Separation (ICA), Berlin, Germany, 2009: 290–297.

[6] NESTA F and 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] REDDY V V, KHONG W H, and NG B P. Unambiguous speech DOA estimation under spatial aliasing conditions[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2014, 22(12): 2133–2145.

[8] YILMAZ O and RICKARD S. Blind separation of speech mixtures via time-frequency masking[J]. *IEEE Transactions on Signal Processing*, 2004, 52(7): 1830–1847.

[9] ARAKI S, SAWADA H, MUKAI R, *et al*. DOA estimation for multiple sparse sources with normalized observation vector clustering[C]. Proceedings of 2006 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2006), Toulouse, France, 2006: 33–36.

[10] BRUTTI A and NESTA F. Tracking of multidimensional TDOA for multiple sources with distributed microphone pairs[J]. *Computer Speech & Language*, 2013, 27(3): 660–682.

[11] THO N T N, ZHAO Shengkui, and JONES D L. Robust DOA estimation of multiple speech sources[C]. Proceedings of 2014 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Florence, Italy, 2014: 2287–2291.

[12] 许志勇, 赵兆, 刘明. 宽间距麦克风阵列实时无模糊多声源被动测向[J]. 电子与信息学报, 2011, 33(9): 2056–2061. doi: 10.3724/SP.J.1146.2010.01273.

XU Zhiyong, ZHAO Zhao, and LIU Ming. Real-time unambiguious passive direction finding for multiple sound sources with widely spaced microphone array[J]. *Journal of Electronics & Information Technology*, 2011, 33(9): 2056–2061. doi: 10.3724/SP.J.1146.2010.01273.

[13] GUSTAFFSON T, RAO B D, and 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] LEHMANN E and JOHANSSON A. Prediction of energy decay in room impulse responses simulated with an image-source model[J]. *Acoustical Society of America*, 2008, 124(1): 269–277.

[15] BLANDIN C, OZEROV A, and VINCENT E. Multi-source TDOA estimation in reverberant audio using angular spectra and clustering[J]. *Signal Processing*, 2012, 92(8): 1950–1960.

房玉琢： 男，1987 年生，博士，研究方向为阵列信号处理、声学探测、盲信道辨识等.

许志勇： 男，1968 年生，博士，副教授，研究方向为阵列信号处理、声学探测、雷达技术等.