DOI: 10.19650/j.cnki.cjsi.J2108759

噪声混响下说话人跟踪的多特征自适应 UPF 算法*

刘望生[1],潘海鹏[1],王明环[2]
(1. 浙江理工大学机械与自动控制学院 杭州 310018; 2. 浙江工业大学特种装备制造与先进加工技术教育部重点实验室 杭州 310012)

摘 要:为了提高噪声混响环境下说话人跟踪系统的精度和稳健性,提出了一种多特征自适应无迹粒子滤波(MFAUPF)算法。该算法以语音信号的多特征作为观测信息,采用多假设和频选函数构建了时延选择机制和波束输出能量优化机制,并在两种机制融合的基础上构建了似然函数,弥补了单特征不能同时稳健噪声和混响的不足。由于说话人运动具有随机性,建立了声源跟踪的自适应 CV 模型,在此基础上将无迹卡尔曼滤波(UKF)与抗差估计理论相结合作为提议分布,提高了模型的适配能力。文中仿真和实测结果表明,在 AUPF 下,多特征算法比 SBFSRP 算法位置平均 RMSE 减少了 18% 以上,在多特征观测下,AUPF 算法比 CV 算法位置平均 RMSE 减少了 14% 以上,所提算法具有跟踪精度高和数值稳定性强的特点。

关键词: 说话人跟踪;麦克风阵列;室内混响;多特征;AUPF 算法
中图分类号: TH712 TP216 **文献标识码**: A **国家标准学科分类代码**: 510.4040

Adaptive unscented particle filter algorithm based on multi-feature for speaker tracking in noisy and reverberant environments

Liu Wangsheng[1], Pan Haipeng[1], Wang Minghuan[2]
(*1. School of Mechanical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China*;
2. Key Laboratory of Special Purpose Equipment and Advanced Processing Technology, Zhejiang University of Technology, Ministry of Education, Hangzhou 310012, China)

Abstract: To improve the accuracy and robustness of the speaker tracking system in noisy and reverberant environments, an adaptive unscented particle filter (AUPF) algorithm based on multi-feature is proposed. The multi-feature of the speech signal is regarded as the observation information in this algorithm, where the multi-hypothesis and frequency selection function is applied to the mechanisms of time delay selection and beam output energy optimization. Subsequently, the likelihood function is constructed by combining these two mechanisms, which makes up for the deficiency that noise and reverberation cannot be restrained simultaneously by a single feature. Considering the randomness of speaker motion, a new proposal distribution is utilized in the particle filter algorithm, which combines the unscented Kalman filter (UKF) and the robust estimation theory based on the adaptive constant speed model to improve the adaptability of the model. The simulation and experimental results show that based on AUPF, the position average RMSE of multi feature algorithm is reduced by more than 18% compared with that of SBFSRP, and under multi-feature observation, the position average RMSE of AUPF algorithm is reduced by more than 14% compared with that of CV algorithm. It has the characteristics of high tracking accuracy and strong numerical stability.

Keywords: speaker tracking; microphone array; room reverberation; multi-feature; AUPF algorithm

0 引 言

声源定位与跟踪技术在视频会议系统、人机交互、普适计算以及远距离语音识别等领域有着广泛的应用[1-3]。室内说话人跟踪是个典型的非线性滤波问题,其难点源于室内声学环境的不定性和说话人运动的随机性。室内说话人跟踪的关键在于如何从麦克风接收信号中最优地

收稿日期:2021-10-24 Received Date: 2022-10-24
*基金项目:国家自然科学基金(51975532)项目资助

提取说话人位置的有用信息,主要设计思想是采用空时相关法,用较少的麦克风达到较高的定位精度。

在空时相关法中,基于时延估计的定位算法精度高、实时性好,但受室内环境影响大[4]。文献[5]引入二次相关并采用最小均方自适应滤波弥补了广义互相关的不足,提高了信号相关程度和时延估计的可靠性,该方法具有较好的抗噪与抗混响性能。文献[6]提出了经验模式分解最大似然时延估计法,将传感器接收信号分解为本征模态函数并进行信号与噪声的重构,结果表明,在低信噪比和低频水声源的混响环境中取得了较好的效果。文献[7]提出了基于深度神经网络的时频掩蔽声源定位法,将在线深度回归的时频掩蔽结果代入 GCC-PHAT 中,再进行声源定位,实验结果表明了该方法的有效性。这些对时延估计的改进算法具有一定的抗噪声抗混响性能,但由于语音信号具有非平稳特征,房间混响具有累积效应,且室内信噪比随时间而变化,当环境恶化时会出现伪峰,产生虚假声源。

为了抑制虚假声源的影响,学者们对波束能量法进行了大量的研究。文献[8]提出了一种基于频率信噪比加权的可控响应功率定位算法,通过每帧阵列信号的频域协方差矩阵估计每个频率的信噪比,并将信噪比映射为可控响应功率值的加权系数,再搜索可控响应功率的最大值实现声源定位,取得了较好的鲁棒性和较高的正确率。文献[9]提出了一种基于学习的多特征估计法,采用高斯回归模型并联合距离和方向特征,取得了准确的定位效果,但当声源发生机动时,该算法跟踪性能随之下降。文献[10]提出了一种合成伪图像训练的卷积神经网络声源检测法,采用卷积神经网络与声辐射预测相结合的方法估计 DOA,进而检测声源位置,该算法具有良好的鲁棒性和抗噪声能力,并成为声源实时跟踪的一个突破。这一类改进算法是基于波束输出能量的定位算法,与基于时延估计的定位算法相比,其稳健性增强,但跟踪精度不高,并且基于学习的方法在学习数据不足或与训练数据不匹配的实际环境下可能无法获得较好的跟踪性能。因此有必要研究时延法和波束能量法的优化问题,以提高声源跟踪性能。由于说话人运动具有随机性[11],无论时延法还是波束能量法在声源发生机动时其跟踪性能均明显下降。文献[12]提出的随机行走模型基本能描述说话人行走状态,文献[13]给出了郎之万模型在跟踪说话人时的参数设置,这两种模型都有一定的适应性,但跟踪精度不高,很难兼顾声源的不同运动模式。

本文在已有研究的基础上,提出了多特征自适应无迹粒子滤波(MFAUPF)算法。利用接收信号的时间相关性和声源位置的空间相关性构建了多特征观测优化机制。针对说话人运动的随机性,建立了自适应匀速模型,并引入无迹卡尔曼滤波(UKF)和抗差估计产生提议分布来跟踪说话人的突变运动模式。仿真和实测结果验证了所提算法的有效性。

1　音频信号多特征建模

1.1　接收信号时频分析

在室内,假定单个声源信号 $s(k)$ 经多径传播后到达麦克风阵列,则第 i 个麦克风接收的音频信号可表示为:

$$x_i(k) = h_i(k) * s(k) + n_i(k) \tag{1}$$

式中:$h_i(k)$ 表示声源与第 i 个麦克风之间的房间脉冲响应;$i = 1,2,\cdots,M$,M 为麦克风数目;"$*$"是卷积运算符;$n_i(k)$ 表示第 i 路信号的高斯噪声。假设每个麦克风的噪声不相关,并且噪声与信号也不相关,则麦克风信号频域表达式为:

$$X_i(\omega) = H_i(\omega)\mathrm{S}(\omega) + N_i(\omega) \tag{2}$$

麦克风对上的广义互相关函数定义为:

$$R_{ij}(\tau) = \int_{-\infty}^{+\infty} \phi_{ij}(\omega) X_i(\omega) X_j^*(\omega) \mathrm{e}^{\mathrm{j}\omega\tau} \mathrm{d}\omega \tag{3}$$

式(3)取最大值的自变量就是时延估计值[4]。当权函数为 $\phi_{ij}(\omega) = 1/|X_i(\omega)X_j^*(\omega)|$ 时,GCC 称作 GCC-PHAT[6]。为了提高 GCC-PHAT 对噪声的鲁棒性,文献[14]利用频选函数选取信噪比大的频率用于时延估计,其中信噪比定义为:

$$SNR(\omega) = 10\log \frac{P_s(\omega)}{\sigma^2(\omega)} \tag{4}$$

式中:

$$P_s(\omega) = \frac{2}{M^2 - M} \sum_{i=2}^{M} \sum_{j=1}^{i-1} |\boldsymbol{R}_\omega(i,j)| \tag{5}$$

$$\sigma^2(\omega) = \frac{1}{M}[\mathrm{tr}(\boldsymbol{R}_\omega) - MP_s(\omega)] \tag{6}$$

式中:$P_s(\omega)$、$\sigma^2(\omega)$ 分别为信号和噪声的功率谱密度;$\boldsymbol{R}_\omega$ 为频率协方差矩阵;$\mathrm{tr}(\cdot)$ 表示求迹运算符。且有:

$$\boldsymbol{R}_\omega = E\{\boldsymbol{X}(\omega)\boldsymbol{X}^{\mathrm{H}}(\omega)\} = P_s(\omega)\boldsymbol{H}(\omega)\boldsymbol{H}^{\mathrm{H}}(\omega) + \sigma^2(\omega)\boldsymbol{I} \tag{7}$$

频选函数定义为:

$$\varphi(\omega) = \begin{cases} 1, & SNR(\omega) \geqslant \eta \\ 0, & SNR(\omega) < \eta \end{cases} \tag{8}$$

式中:η 为信噪比阈值。

采用频选函数能在一定程度上改善 GCC-PHAT 时延估计的精度和稳健性,但在低信噪比或强混响下其性能随之下降。

1.2　波束能量加权方法

由波束输出能量构建定位似然函数的方法主要有 SBF 法和 SRP-PHAT 法。假定声源位置为 l,则 SBF 的表达式如式(9)所示[15]。

$$P_{\mathrm{SBF}}(l)=\int_{\Omega}\left|\sum_{m=1}^{M}X_m(\omega)\exp(\mathrm{j}\omega\|\boldsymbol{l}-\boldsymbol{l}_m\|c^{-1})\right|^2\mathrm{d}\omega \tag{9}$$

式中：$X_m(\omega)$ 表示第 m 个麦克风接收信号的频谱；$\boldsymbol{l}_m=[x_m\ \ y_m]^{\mathrm{T}}$ 表示第 m 个麦克风的位置；c 为声速。SRP-PHAT 的输出为[8]：

$$P_{\mathrm{SRP}}(l)=\sum_{i=1}^{M}\sum_{j=i+1}^{M}G_{ij}(\tau_{ij}) \tag{10}$$

$$G_{ij}(\tau_{ij})=\int_{\Omega}\phi_{ij}(\omega)X_i(\omega)X_j^{*}(\omega)\exp(\mathrm{j}\omega(\tau_i-\tau_j))\mathrm{d}\omega \tag{11}$$

式中：$G_{ij}(\tau_{ij})$ 是第 i 个麦克风与第 j 个麦克风接收信号的广义互相关函数；$\phi_{ij}(\omega)=1/|X_i(\omega)X_j^{*}(\omega)|$，$X_i(\omega)$ 为第 i 个麦克风信号 $x_i(k)$ 的加窗傅里叶变换；τ_i 与 τ_j 是阵列的可控时延。

在混响环境中，SBF 的空间谱呈现出多峰特性，定位性能差，但 SBF 对背景噪声有较强的鲁棒性。SRP-PHAT 空间谱峰比较尖锐，对混响有较强的抑制能力，但 SRP-PHAT 对噪声敏感。为了充分利用 SBF 和 SRP-PHAT 两特征的互补性，本文定义特征向量距离函数为：

$$v_f=P_f(\boldsymbol{l})-P_{f,\tau}(\boldsymbol{l}) \tag{12}$$

式中：$P_f(\boldsymbol{l})$ 表示 $P_{\mathrm{SBF}}(\boldsymbol{l})$ 或 $P_{\mathrm{SRP}}(\boldsymbol{l})$ 两特征，$P_{f,\tau}(\boldsymbol{l})$ 表示 $P_f(\boldsymbol{l})$ 与时延估计下特征函数的融合。此时波束输出能量表达式为：

$$w_k^{(l)}=\mu P_{\mathrm{SBF}}(\boldsymbol{l})+(1-\mu)P_{\mathrm{SRP}}(\boldsymbol{l}) \tag{13}$$

$$\mu=\frac{v_{\mathrm{SRP}}^2}{v_{\mathrm{SBF}}^2+v_{\mathrm{SRP}}^2} \tag{14}$$

式中：μ 为加权因子，实际应用时 μ 值可通过空时相关和迭代滤波优化得到，此时式(13)能同时抑制噪声和混响的影响。

2 多特征优化机制

为了提高时延估计的可靠性并改善单一特征定位性能的局限性，本文以时延估计和波束输出能量构建了多特征观测量。为了获取最佳多特征观测量，本文建立了多假设时延估计模型，利用接收信号的空时相关并结合迭代粒子滤波对多假设下多特征观测量进行了优化。

2.1 声源跟踪建模

在迭代粒子滤波跟踪系统中，说话人的状态空间模型可描述为：

$$\boldsymbol{\chi}_k=f_k(\boldsymbol{\chi}_{k-1})+\boldsymbol{v}_{k-1} \tag{15}$$

$$\boldsymbol{y}_k=g_k(\boldsymbol{\chi}_k)+\boldsymbol{n}_k \tag{16}$$

式中：$\boldsymbol{\chi}_k$ 为系统状态向量；$\boldsymbol{y}_k$ 为量测向量；$\boldsymbol{v}_{k-1}$ 和 $\boldsymbol{n}_k$ 是独立同分布零均值系统噪声和观测噪声；$f_k(\cdot)$ 为状态转移函数；$g_k(\cdot)$ 为观测函数。麦克风接收信号的观测量一般包括到达时延、输出波束能量等特征[13]，将不同观测特征代入式(16)便可形成不同的观测方程。给定量测向量 $\boldsymbol{y}_k$，则 $\boldsymbol{\chi}_k$ 在最小方差意义下的最优估计由下式条件均值给出[16]：

$$\hat{\boldsymbol{\chi}}_k=E[\boldsymbol{\chi}_k|\boldsymbol{y}_{1:k}]=\int\boldsymbol{\chi}_k p(\boldsymbol{\chi}_k|\boldsymbol{y}_{1:k})\mathrm{d}\boldsymbol{\chi}_k \tag{17}$$

根据贝叶斯理论，后验概率密度(PDF)构成序贯估计问题完全解。通过递推计算 PDF，可获得系统状态均值、方差和峰值等估计。在粒子滤波中重要采样表示为：

$$\boldsymbol{\chi}_k^{(p)}\sim \mathrm{P}_r(\boldsymbol{\chi}_k^{(p)}|\boldsymbol{\chi}_{k-1}^{(p)}) \tag{18}$$

权值更新为：

$$w_k^{(p)}\propto w_{k-1}^{(p)}\mathrm{P}_r(\boldsymbol{y}_k|\boldsymbol{\chi}_{k-1}^{(p)})=w_{k-1}^{(p)}\mathrm{P}_r(\boldsymbol{y}_k|\boldsymbol{\chi}_k^{(p)}) \tag{19}$$

式中：p 为粒子索引，$1\leqslant p\leqslant P$，P 为粒子个数。

假定说话人状态为 $\boldsymbol{\chi}_k=[x_k\dot{x}_k y_k\dot{y}_k]^{\mathrm{T}}$，$(x_k,y_k)$ 和 $(\dot{x}_k,\dot{y}_k)$ 分别表示说话人的位置和速度，则说话人运动的状态模型可表示为：

$$\boldsymbol{\chi}_k=\boldsymbol{A}\boldsymbol{\chi}_{k-1}+\boldsymbol{u}_k \tag{20}$$

式中：$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{B}&0\\0&\boldsymbol{B}\end{bmatrix}$ 为状态转移矩阵，$\boldsymbol{B}=\begin{bmatrix}1&T\\0&1\end{bmatrix}$，$T$ 是采样时间。$\boldsymbol{u}_k$ 为高斯白噪声过程，其方差为 $\boldsymbol{\Gamma}$，且 $\boldsymbol{\Gamma}=q_0\boldsymbol{I}$，$\boldsymbol{I}=\mathrm{diag}(1,1,1,1)$，$q_0$ 用来描述系统噪声方差的强度。基于时延估计的观测模型为：

$$\boldsymbol{\tau}_k(\boldsymbol{\chi}_k)=\frac{(\boldsymbol{\chi}_k-\boldsymbol{m}_1^{(i)}-\boldsymbol{\chi}_k-\boldsymbol{m}_2^{(i)})}{c} \tag{21}$$

式中：$i=1,2,\cdots,D$，D 为麦克风对的组数；$\boldsymbol{m}_1^{(i)}$、$\boldsymbol{m}_2^{(i)}$ 表示第 i 组麦克风对的坐标，时延量测方差为 $\boldsymbol{R}$。

2.2 多特征优化方法

为了提取最佳多特征观测量，本文建立了多假设时延估计模型，设 k 时刻麦克风对接收信号的时延估计为：

$$\tilde{\boldsymbol{y}}_k=[(\tilde{\boldsymbol{y}}_k^{(1)})^{\mathrm{T}},(\tilde{\boldsymbol{y}}(2)_k)^{\mathrm{T}},\cdots,(\tilde{\boldsymbol{y}}(d)_k)^{\mathrm{T}}]^{\mathrm{T}} \tag{22}$$

式中：$\tilde{\boldsymbol{y}}(d)_k$ 表示麦克风对 d 上的所有候选时延值，$1\leqslant d\leqslant D$。且有：

$$\tilde{\boldsymbol{y}}_k^{(d)}=[(\tilde{\boldsymbol{y}}_k^{(d,1)})^{\mathrm{T}},(\tilde{\boldsymbol{y}}(d,2)_k)^{\mathrm{T}},\cdots,(\tilde{\boldsymbol{y}}(d,n)_k)^{\mathrm{T}}]^{\mathrm{T}} \tag{23}$$

式中：$\tilde{\boldsymbol{y}}(d,n)_k$ 为第 d 组麦克风对的第 n 个候选时延值，$1\leqslant n\leqslant N$，N 是候选时延值个数。采用 $H_{\bar{n}}$ 表示 D 组麦克风对上候选直达波的假设，$1\leqslant\bar{n}\leqslant N^D$，则 $H_{\bar{n}}$ 对应麦克风对上的候选时延值为：

$$\bar{\boldsymbol{y}}_k^{(\bar{n})}=[(\tilde{\boldsymbol{y}}_k^{(1,\tilde{n}_1)})^{\mathrm{T}},(\tilde{\boldsymbol{y}}(2,\tilde{n}_2)_k)^{\mathrm{T}},\cdots,(\tilde{\boldsymbol{y}}(D,\tilde{n}_D)_k)^{\mathrm{T}}]^{\mathrm{T}} \tag{24}$$

式中：$\tilde{n}_d$ 表示第 d 组麦克风对上的候选时延值的索引，且 $1\leqslant\tilde{n}_d\leqslant N$。将候选直达波分别作为时延观测方程，联合式(13)进行粒子滤波，其滤波过程为：

$$\boldsymbol{\chi}_{k|k-1}^{(p)}=f_k(\boldsymbol{\chi}_{k-1}^{(p)})+\boldsymbol{v}_{k-1} \tag{25}$$

$$w_k^{(p,\bar{n})}=w_k^{(p)}\cdot P_r(\bar{\boldsymbol{y}}_k^{(\bar{n})}|\boldsymbol{\chi}^{(p)}{}_{k|k-1},H_{\bar{n}})\cdot w_k^{(l)} \tag{26}$$

$$\hat{\boldsymbol{\chi}}_k = \sum_{p=1}^{P} w_k^{(p,\bar{n})} \Big/ \sum w_k^{(p,\bar{n})} \cdot \boldsymbol{\chi}_{k|k-1}^{(p)} \tag{27}$$

式中:$P_r(\bar{\boldsymbol{y}}_k^{(\bar{n})} | \boldsymbol{\chi}^{(p)}{}_{k|k-1}, H_{\bar{n}})$ 表示假设 $H_{\bar{n}}$ 下粒子滤波似然函数,$\hat{\boldsymbol{\chi}}_k$ 为 k 时刻粒子滤波估计值。

在迭代粒子滤波中,首先根据式(21)和(25)求得时延预测值 $\hat{\boldsymbol{\tau}}_{k|k-1}$,从 N^D 个假设直达波中预选 N_k 组具有最大似然函数的时延估计作为候选测量集,记 $\hat{n}$ 为对应候选直达波的索引,$1 \leqslant \hat{n} \leqslant N_k$,其中似然函数为:

$$P_r(\bar{\boldsymbol{y}}_k^{(\bar{n})} | \boldsymbol{\chi}_{k-1}^{(p)}, H_n) = \mathcal{N}(\bar{\boldsymbol{y}}_k^{(\bar{n})}; \hat{\boldsymbol{\tau}}_{k|k-1}, \boldsymbol{R}) \tag{28}$$

以 N_k 组候选测量集建立时延观测方程,按式(25)~(27)进行滤波。根据每组候选测量集后验估计函数优化时延估计,后验估计函数定义为:

$$P_r(\hat{\boldsymbol{\chi}}_k | \bar{\boldsymbol{y}}_k^{(\hat{n})}, H_{\hat{n}}) = \mathcal{N}(\hat{\boldsymbol{\chi}}_k; \boldsymbol{\chi}_{k|k-1}^{(p)}, \boldsymbol{Q}) \tag{29}$$

假定后验估计函数取最大值对应的时延估计观测值为:

$$\bar{\boldsymbol{y}}_k = \bar{\boldsymbol{y}}_k^{(\hat{n}_k)} = [(\tilde{\boldsymbol{y}}_k^{(1,\bar{n}_1)})^{\mathrm{T}}, \cdots, (\tilde{\boldsymbol{y}}(D, \bar{n}_D)_k)^{\mathrm{T}}]^{\mathrm{T}} \tag{30}$$

式中:$\hat{n}_k = \underset{\hat{n} \in \{1,2,\cdots,N_k\}}{\arg\max} \{P_r(\hat{\boldsymbol{\chi}}_k | \bar{\boldsymbol{y}}_k^{(\hat{n})}, H_{\hat{n}})\}$,$\hat{n}_k$ 为 $\bar{\boldsymbol{y}}_k$ 在候选测量 $\bar{\boldsymbol{y}}_k^{(\hat{n})}$ 中的索引,$\bar{n}_d$ 为 $\tilde{\boldsymbol{y}}_k^{(d,\bar{n}_d)}$ 在麦克风对 d 上候选时延中的索引。将 $\bar{\boldsymbol{y}}_k$ 代入式(28)求得似然函数,按式(14)优化波束能量加权因子 μ,其中 $P_{f,\tau}(l) = P_f(l) \cdot P_r(\bar{\boldsymbol{y}}_k | \hat{\boldsymbol{\chi}}_k)$。将优化出的多特征观测量构建似然函数代入式(25)~(27)再次进行粒子滤波。

在多假设时延估计模型下,通过空时相关并进行迭代滤波,引导粒子向高似然区移动,能从接收信号中最优地提取多特征观测量,以改善粒子滤波抗噪声抗混响性能。

2.3　多特征优化实现

将多特征观测信息与声源运动模型相结合,采用迭代粒子滤波,可优化出最佳多特征观测量,步骤如下。

1) 粒子集初始化,$k=0$,采样 $\boldsymbol{\chi}_0^{(p)} \sim P(\boldsymbol{\chi}_0)$,即根据 $P(\boldsymbol{\chi}_0)$ 分布采样得到 $\boldsymbol{\chi}_0^{(p)}, p = 1, \cdots, P$。

2) 滤波更新,k 时刻,通过状态转移方程预测新的粒子集,$\boldsymbol{\chi}_{k|k-1}^{(p)} = \boldsymbol{A}\boldsymbol{\chi}_{k-1}^{(p)} + \boldsymbol{u}_k$。将状态预测值代入式(21)得到时延预测值 $\hat{\boldsymbol{\tau}}_{k|k-1}$,在多假设模型中预选 N_k 组使式(28)最大的时延估计作为候选测量集。

3) 以 N_k 组时延估计作为观测方程,在式(13)的基础上,求使式(29)具有最大值的时延估计 $\bar{\boldsymbol{y}}_k$。将 $\bar{\boldsymbol{y}}_k$ 代入式(28)求出似然函数,按式(14)优化出波束能量融合因子 μ_{opt},其中 $P_{f,\tau}(l) = P_f(l) \cdot P_r(\bar{\boldsymbol{y}}_k | \hat{\boldsymbol{\chi}}_k)$。此时波束能量融合值为 $w_{k,\mathrm{opt}}^{(l)} = \mu_{\mathrm{opt}} P_{\mathrm{SBF}}(l) + (1 - \mu_{\mathrm{opt}}) P_{\mathrm{SRP}}(l)$。

4) 采用 $\bar{\boldsymbol{y}}_k$ 和式(13)计算粒子权重,有 $w_k^{(p)} = w_k^{(p)} \cdot P_r(\bar{\boldsymbol{y}}_k | \boldsymbol{\chi}_{k-1}^{(p)}) \cdot w_{k,\mathrm{opt}}^{(l)}$,$\tilde{w}_k^{(p)} = \dfrac{w_k^{(p)}}{\sum_{p \in P} w_k^{(p)}}$。

5) 粒子滤波中,采用文献[16]的方法进行重采样,设得到粒子集为 $\bar{\boldsymbol{\chi}}_k^{(p)}$,对应归一化权值为 $\bar{w}_k^p$。

6) 状态更新,$\hat{\boldsymbol{\chi}}_k = \sum_{p=1}^{P} \bar{w}_k^{(p)} \cdot \bar{\boldsymbol{\chi}}_k^{(p)}$。

7) 递归:$k=k+1$ 时,重复步骤 2)~步骤 7)。

3　自适应无迹粒子滤波(AUPF)算法

在获取最佳多特征观测量后,为了克服说话人运动的随机性,建立了声源运动的自适应 CV 模型。由于标准粒子滤波采用先验分布,没有考虑最新观测信息,其滤波精度低且鲁棒性弱,为了更好地逼近后验概率密度分布,本文采用抗差 UKF 产生提议分布,使粒子滤波中每个粒子的每个 sigma 点用抗差 UKF 算法来更新,可形成 AUPF 算法。

3.1　自适应 CV 模型

声源机动可看作是随机状态噪声的激励,由此可建立声源跟踪的自适应 CV 模型。在式(20)中,状态噪声方差实时估计可表示为:$\boldsymbol{P}_{k|k} = \mathrm{diag}(q_x \boldsymbol{I}_2, q_y \boldsymbol{I}_2)$,$\boldsymbol{I}_2 = \mathrm{diag}(1,1)$,其中 q_x、q_y 为说话人在 X、Y 轴上机动的强度。利用速度估计偏差可以近似表示机动强度的大小,即:

$$q_x = C | \dot{x}_{k|k} - \dot{x}_{k|k-1} | / T \tag{31}$$

式中:C 为比例系数;T 为采样时间;$\dot{x}_{k|k}$ 为当前速度估计值;$\dot{x}_{k|k-1}$ 为速度一步预测值。q_y 的计算与 q_x 类似。

由式(31)可知,$\dot{x}_{k|k-1}$ 没有考虑速度变化率扰动量的影响,$\dot{x}_{k|k}$ 包含了速度变化率扰动量对观测值的影响,系统噪声方差 $\boldsymbol{P}_{k|k}$ 可用两者的偏差估计。通过实时调整噪声方差,使粒子采样空间发生改变,提高了最新观测信息下高似然值粒子的比重,改善了模型对突变状态的跟踪性能,并保持了对非机动或弱机动的跟踪精度。

3.2　抗差 UKF 提议分布

为了更好地逼近后验分布,根据残差理论,当模型预测值和传感器测量值之间的残差增大时,应相应增加滤波增益以提高估计精度。抗差 UKF 采用自适应调节因子来改变滤波增益,使系统状态预报值协方差更合理,能更好逼近粒子后验概率密度分布。本文采用预测残差作为调节因子判别统计量,即[17]:

$$\tilde{\boldsymbol{\gamma}}_k = g_k(\hat{\boldsymbol{\chi}}_k) - \boldsymbol{y}_k \tag{32}$$

$$\mu_k = | \mathrm{tr}(\boldsymbol{P}_{\tilde{\boldsymbol{y}}_k}) / \tilde{\boldsymbol{\gamma}}_k^{\mathrm{T}} \tilde{\boldsymbol{\gamma}}_k |^{1/2} \tag{33}$$

调节因子为:

$$\mu_k = \begin{cases} 1, & \mu_k \geqslant 1 \\ \mu_k, & \mu_k < 1 \end{cases} \tag{34}$$

设 k 时刻抗差 UKF 滤波解为：

$$\hat{\boldsymbol{\chi}}_k = \hat{\boldsymbol{\chi}}_k^- + \boldsymbol{K}_k(\boldsymbol{y}_k - \hat{\boldsymbol{y}}_k^-) \tag{35}$$

式中：$\hat{\boldsymbol{\chi}}_k^-$ 和 $\hat{\boldsymbol{y}}_k^-$ 分别表示状态和测量的一步预测值；$\boldsymbol{K}_k$ 为增益矩阵；此时 $\hat{\boldsymbol{\chi}}_k$ 对应的协方差矩阵可表示为：

$$\boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k} = \mu_k \boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k^-} - \boldsymbol{K}_k \boldsymbol{P}_{\hat{\boldsymbol{y}}_k^-} \boldsymbol{K}_k^{\mathrm{T}} \tag{36}$$

$$\boldsymbol{K}_k = \mu_k \boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k^- \hat{\boldsymbol{y}}_k^-} \boldsymbol{P}_{\hat{\boldsymbol{y}}_k^-}^{-1} \tag{37}$$

相应协方差矩阵分别为：

$$\boldsymbol{P}_{\hat{\boldsymbol{y}}_k} = \mu_k(\boldsymbol{P}_{\hat{\boldsymbol{y}}_k^-} - \boldsymbol{R}_{\boldsymbol{y}_k}) + \boldsymbol{R}_{\boldsymbol{y}_k} \tag{38}$$

$$\boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k^-} = \sum_{i=0}^{2L_a} w_i^c [\boldsymbol{x}_{k\mid k-1,i}^x - \hat{\boldsymbol{\chi}}_k^-][\boldsymbol{x}_{k\mid k-1,i}^x - \hat{\boldsymbol{\chi}}_k^-]^{\mathrm{T}} + \boldsymbol{Q}_k \tag{39}$$

$$\boldsymbol{P}_{\hat{\boldsymbol{y}}_k^-} = \sum_{i=0}^{2L_a} w_i^c [\boldsymbol{y}_{k\mid k-1,i} - \hat{\boldsymbol{y}}_k^-][\boldsymbol{y}_{k\mid k-1,i} - \hat{\boldsymbol{y}}_k^-]^{\mathrm{T}} \tag{40}$$

$$\boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k^- \hat{\boldsymbol{y}}_k^-} = \sum_{i=0}^{2L_a} w_i^c [\boldsymbol{x}_{k\mid k-1,i}^x - \hat{\boldsymbol{\chi}}_k^-][\boldsymbol{y}_{k\mid k-1,i} - \hat{\boldsymbol{y}}_k^-]^{\mathrm{T}} \tag{41}$$

其中，$\boldsymbol{Q}_k = q\boldsymbol{I}$，$q$ 如式(31)；L_a 为扩维向量维数；且 $\hat{\boldsymbol{\chi}}_k^- = \sum_{i=0}^{2L_a} w_i^m \boldsymbol{x}_{k\mid k-1,i}^x$，$w_i^m$ 为均值权值，w_i^c 为方差权值；$w_i^c = w_i^m = 1/(2(L_a+\lambda))$，$i = 1, \cdots, 2L_a$，$w_0^m = \lambda/(L_a+\lambda)$，$w_0^c = w_0^m + (1-\alpha^2+\beta^2)$。粒子滤波中每个粒子的 sigma 点集为[18]：

$$\boldsymbol{x}_{k\mid k-1,i}^a = \hat{\boldsymbol{\chi}}_k^{a-} \tag{42}$$

$$\boldsymbol{x}_{k\mid k-1,i}^a = \hat{\boldsymbol{\chi}}_k^{a-} + (\sqrt{(L_a+\lambda)\boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k^-}^a})_i \tag{43}$$

$$\boldsymbol{x}_{k\mid k-1,i}^a = \hat{\boldsymbol{\chi}}_k^{a-} - (\sqrt{(L_a+\lambda)\boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k^-}^a})_i \tag{44}$$

式中：$\boldsymbol{x}_{k\mid k-1,i}^a$ 为 sigma 点集；$\hat{\boldsymbol{\chi}}_k^{a-}$ 为状态 $\hat{\boldsymbol{\chi}}_k^-$ 的扩维向量；$\lambda = \alpha^2(L_a+\kappa) - L_a$ 为合成比例参数，其中 κ、α 决定 sigma 点以均值为原点的分散程度，一般选取 $\kappa = 0$，$\alpha = 1$。式(42) ~ (44) 中 i 取值分别为 0，(1, …, L_a)，(n_a+1, …, $2L_a$)。

采用抗差 UKF 滤波，一方面通过加权采样粒子提高了非线性滤波精度，另一方面当说话人发生状态突变时，提高了粒子对突变机动的跟踪性能。

3.3 AUPF 算法实现

将自适应 CV 模型和抗差 UKF 相结合构成的提议分布应用于粒子滤波便构成 AUPF 算法，其实现步骤如下。

1）粒子集初始化，$k=0$，采样 $\boldsymbol{\chi}_0^{(p)} \sim P(\boldsymbol{\chi}_0)$，即根据 $P(\boldsymbol{\chi}_0)$ 分布采样得到 $\boldsymbol{\chi}_0^{(p)}$，$p = 1, \cdots, P$。

2）滤波更新，k 时刻，对每个粒子的每个 sigma 点，采用 CV 模型和抗差 UKF 计算其调节因子、估值及方差，并重新生成新的粒子：$\hat{\boldsymbol{\chi}}_k^i = \hat{\boldsymbol{\chi}}_k^{i-} + \boldsymbol{K}_k^i(\boldsymbol{y}_k - \hat{\boldsymbol{y}}_k^{i-})$，$\hat{\boldsymbol{\chi}}_k^i$ 对应的协方差矩阵为 $\boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k}^i$。将 $\hat{\boldsymbol{\chi}}_k^i$ 和 $\boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k}^i$ 分别作为正态分布均值和方差，则 k 时刻粒子集为：$\boldsymbol{x}_k^i \sim q(\boldsymbol{x}_k^i \mid \boldsymbol{x}_{0:k-1}^i, \boldsymbol{y}_{1:k}) = \mathcal{N}(\hat{\boldsymbol{\chi}}_k^i, \boldsymbol{P}_{\hat{\boldsymbol{\chi}}_k}^i)$。

3）权值计算及归一化，$w_k^i = w_{k-1}^i \dfrac{p(\boldsymbol{y}_k \mid \boldsymbol{x}_k^i) p(\boldsymbol{x}_k^i \mid \boldsymbol{x}_{k-1}^i)}{q(\boldsymbol{x}_k^i \mid \boldsymbol{x}_{k-1}^i, \boldsymbol{y}_k)}$，$\tilde{w}_k^i = w_k^i \Big/ \sum_{i=1}^{P} w_k^i$。

4）粒子滤波中，重采样参考文献[16]，设得到粒子集为 $\bar{\boldsymbol{x}}_k^{(p)}$，对应归一化权值为 $\bar{w}_k^{(p)}$。

5）状态更新，状态估计为 $\hat{\boldsymbol{\chi}}_k = \sum_{p=1}^{P} \bar{w}_k^{(p)} \cdot \bar{\boldsymbol{x}}_k^{(p)}$，方差估计为 $\boldsymbol{P}_k = \sum_{p=1}^{P} \bar{w}_k^{(p)} (\bar{\boldsymbol{x}}_k^{(p)} - \hat{\boldsymbol{x}}_k)(\bar{\boldsymbol{x}}_k^{(p)} - \hat{\boldsymbol{x}}_k)^{\mathrm{T}}$。

6）递归，$k=k+1$ 时，重复步骤 2)~6)。

4 仿真与实测分析

为了验证本文所提算法的有效性，在不同的声学环境和跟踪路径下进行了一系列仿真和实测实验。不失一般性，仿真实验模拟了普通会议室的声学环境，房间大小为 6 m×5 m×2.7 m，墙面平均反射系数为 0.92，全向麦克风为 8 个，成对布置于 4 个墙面上，每个麦克风特性相同且严格同步，麦克风距地面和墙壁分别为 1.5、0.12 m，麦克风阵列的摆放如图 1 所示。

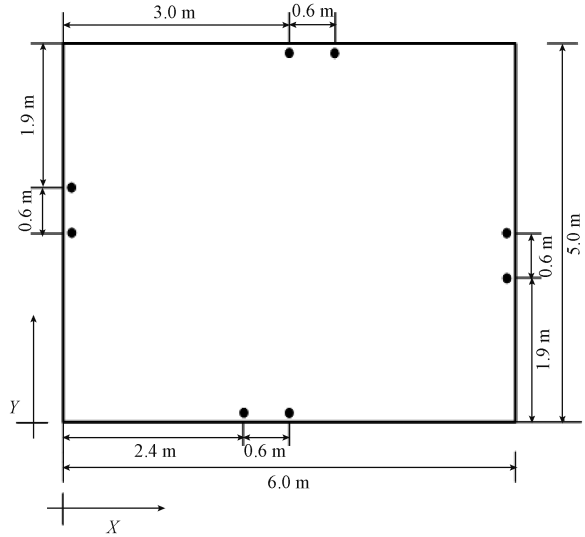

图 1 室内麦克风布置

Fig. 1 Indoor microphone layout

只考虑二维跟踪问题，即假定声源的高度与麦克风的高度相同。声源运动的过程噪声标准差为 0.07，TDOA 观测噪声标准差为 0.031 6。粒子滤波中粒子数 $N=300$，采用残差重采样以降低粒子退化现象的影响。仿真实验中的硬件平台为 Intel(R) Core(TM) I7-7700 CPU@3.60 GHz 处理器，16 G 内存，120 G 硬盘，软件平台为 Windows 7 操作系统和 MATLAB2017a。为了对不同滤波算法的估计精度作定量比较，采用位置均方根误差(RMSE)和平均位置均方根误差[19]来描述不同算法的跟踪性能。

4.1　多特征自适应定位性能仿真分析

将一段时长为 4.75 s 的纯净语音用作声源,在不同信噪比和不同混响时间下进行仿真实验,模拟声源从点(1.5,1.5)出发,以 0.8 m/s 的速度沿 X 轴正向匀速运动,以 0.5 m/s 的速度沿 Y 轴正向匀速运动。将声源沿着说话人的运动轨迹分成 147 帧,帧长为 32 ms。采用能量衰减 IMAGE 模型模拟房间冲激响应[20]。麦克风接收信号是将每帧语音信号与相应的房间冲激响应做卷积,再加上不同信噪比的背景噪声和标准差 $\sigma=0.02$ 的高斯白噪声获取,其中高斯白噪声用来模拟说话人行走时的转角变化。采样频率为 8 kHz,快速傅里叶变换(FFT)变换长度 $L=512$,相邻两帧重叠 50%,窗函数为汉明窗。

采用 4 种不同观测特征量构建似然函数进行滤波。第 1 种采用带扩散掩蔽的 GCC-PHAT 时延估计(TDE)法[21],第 2 种采用基于 SBF-SRP 的定位算法[15],第 3 种采用 SRP-TDE 定位算法[22],第 4 种采用本文所提多特征自适应(MFA)算法。滤波过程中均采用匀速模型以客观评价不同观测特征的定位性能。为了全面比较 4 种不同特征算法对房间混响和噪声的抑制性能,做了在不同混响时间和不同信噪比下的仿真,每种情形做 30 次,取其 RMSE 的平均值,结果如图 2 和 3 所示。

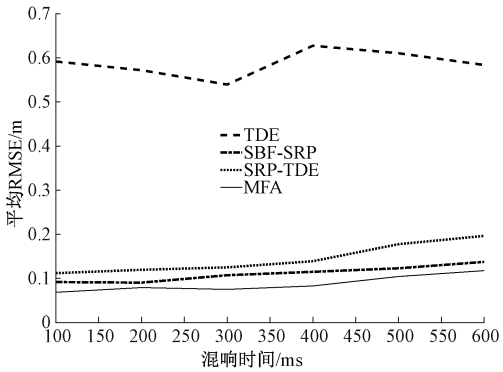

图 2　平均 RMSE 与混响时间的关系

Fig. 2　Average RMSE versus reverberation time

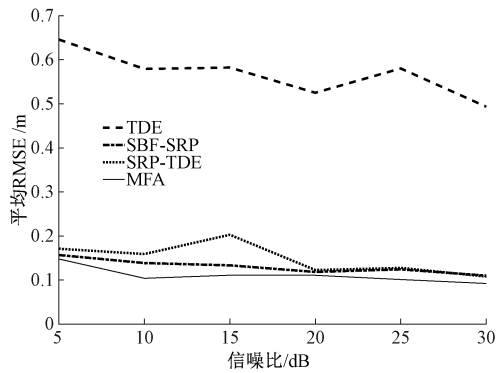

图 3　平均 RMSE 与信噪比的关系

Fig. 3　Average RMSE versus SNR

图 2 是信噪比为 10 dB 时,位置平均 RMSE 与混响时间的关系。由图 2 可知,在不同混响时间条件下,MFA 的跟踪精度均高于其他 3 种算法,这说明多特征观测机制对室内混响具有较好的鲁棒性。SBF-SRP 跟踪性能好于 SRP-TDE,主要是因为受背景噪声和混响的影响,TDE 估计精度较低,随着混响的增强,SRP-TDE 跟踪性能进一步降低。仅采用单特征的 TDE 跟踪效果明显低于其他 3 种算法。MFA 算法在强噪声强混响中性能下降缓慢,表明了本文所提算法抗混响性能优于其他算法。

图 3 是混响时间为 600 ms 时,位置平均 RMSE 与信噪比(SNR)的关系。随着信噪比的提高,4 种算法跟踪性能的总趋势变好,其中 MFA 平均 RMSE 均低于其他算法,这是由于多特征观测机制更加可靠,取得了比其他算法更高的估计精度。随着信噪比逐步增加,麦克风接收信号受噪声干扰变小,时延估计精度随之提高,SRP-TDE 跟踪性能逐步接近并超过 SBF-SRP。TDE 受室内环境影响大,跟踪性能最差。由图 3 可知,MFA 抗噪声性能好于其他 3 种算法。

4.2　AUPF 跟踪精度仿真分析

1)跟踪路径 1,声源由点(1.0,1.0)开始做折线运动,分别在点(1.0,2.6)和点(5.0,2.6)作 90°转向,终止于点(5.0,1.0)。采用一段时长为 7.28 s 的纯净语音用作声源,沿着声源运动轨迹将语音信号分成 226 帧,帧长为 32 ms。在 $SNR=20$ dB,混响时间 $T_{60}=300$ ms 下,利用 4.1 节方法获得麦克风仿真数据。采用 4 种不同模型粒子滤波算法对说话人进行跟踪。第 1 种模型采用自适应郎之万(MFAL)模型,参数设置参考文献[13],第 2 种采用文献[12]提出的自适应随机行走(MFARW)模型,第 3 种采用自适应匀速(MFACV)模型,第 4 种采用 MFAUPF 算法,观测特征量均采用 MFA。仿真结果如图 4 所示。

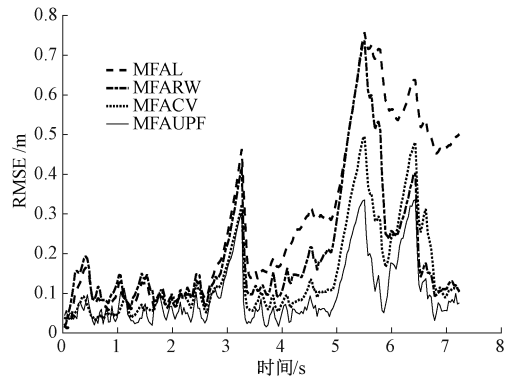

图 4　转向机动位置估计 RMSE

Fig. 4　RMSE of location estimation for steering

2)跟踪路径 2,声源由点(1.0,2.5)开始作逆时针半圆形轨迹运动,停止于点(5.0,2.5)。其余设置同跟踪路径 1。仿真结果如图 5 所示。表 1 为跟踪路径 1 和跟踪路径 2 对应的位置平均 RMSE,表 2 为 4 种算法对应的复杂度和运行时间,其中运行时间为单次仿真时间。

从图 4 可看出,转向机动时 MFACV 和 MFAUPF 跟踪性能优于 MFAL 和 MFARW。随着时间推移,房间多径累积效应影响变大,MFAL 和 MFARW 跟踪性能下降较快,MFACV 和 MFAUPF 均保持了较好的跟踪精度。由于 MFAUPF 采用了抗差 UKF 作为后验概率分布,融入了当前观测信息进而提高了粒子产生的有效性,其收敛速度快于 MFACV。表 1 结果表明,相对于 MFACV,转向机动时 MFAUPF 位置平均 RMSE 减少了 33.49% 左右。图 5 为圆弧机动时 4 种算法的位置估计 RMSE,从图 5 可以看出,MFACV 和 MFAUPF 在起始 1.2 s 内跟踪效果稍差,随后跟踪过程中 MFACV 和 MFAUPF 跟踪效果均好于其他两种算法,其中 MFAUPF 收敛速度和跟踪精度优于 MFACV。由表 1 可知,圆弧机动时 MFAUPF 位置平均 RMSE 比 MFACV 位置平均 RMSE 降低了 36.7% 左右,表明了抗差 UKF 的有效性。

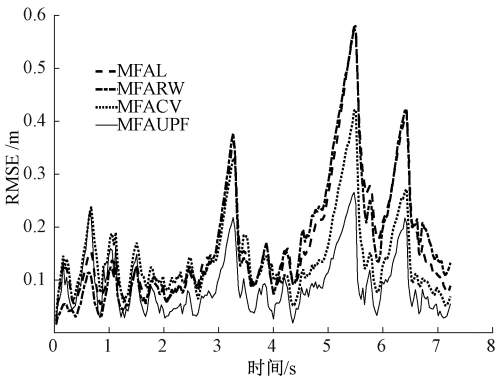

图 5 圆弧机动位置估计 RMSE

Fig. 5 RMSE of location estimation for circular arc

表 1 平均均方根误差比较

Table 1 Comparison of average root mean square error (m)

算 法	转向机动 位置平均 RMSE	圆弧机动 位置平均 RMSE
MFAL	0.284 7	0.171 4
MFARW	0.196 5	0.168 3
MFACV	0.144 8	0.141 5
MFAUPF	0.096 3	0.089 5

表 2 为 4 种算法的复杂度和运行时间,从表中可看出,MFAL、MFARW 和 MFACV 空间时间复杂度相等,运行时间相当。MFAUPF 由于引入当前观测信息并进行无迹卡尔曼滤波,其空间时间复杂度高,所需存储容量大。

表 2 算法性能比较

Table 2 Comparison of algorithm performance

算 法	空间 复杂度	时间 复杂度	运行 时间/s
MFAL	$O(n^4 n_a+p)$	$O(n^4 n_a+p)$	2.640 3
MFARW	$O(n^4 n_a+p)$	$O(n^4 n_a+p)$	2.610 0
MFACV	$O(n^4 n_a+p)$	$O(n^4 n_a+p)$	2.638 6
MFAUPF	$O(n^4 n_a+L_a{}^2 p)$	$O(n^4 n_a+L_a{}^4 p)$	14.030

4.3 实测分析

为了进一步验证本文所提算法的有效性,在一个典型的办公室环境下进行了实测实验。测试房间大小为 6.74 m×5.69 m×4.05 m,其中一块 6 m×5 m 矩形区域用来录制语音,房间混响时间约为 T_{60}=819 ms,背景噪声级约为 41 dBA。为了便于比较,采用与 4.1 相同的麦克风布置方式(图 1),四组麦克风对的中心坐标分别为(3.0 m,0 m)、(5.4 m,2.7 m)、(2.4 m,4.8 m)、(0 m,2.1 m)。自由场传声器采用 MPA436A,频响范围为 20 Hz~10 kHz (±2 dB),满足基本的灵敏度和相位一致性要求,采用 DEWE-51-PCI-16 数据采集设备直接读取传声器信息,其余设置与仿真模拟相同。

为了验证 AUPF 算法的有效性,实测过程选取说话人两种典型运动轨迹作为验证方案,说话人按预定轨迹运动并直接发声,其运动轨迹如下:1)说话人从点(2.1 m,4.18 m)处出发,以 0.7 m/s 的速度沿 Y 轴负向匀速运动,在点(2.1 m,1.4 m)处向左转向,再以 0.7 m/s 的速度沿 X 轴正向匀速运动,在点(4.2 m,1.4 m)处停止运动;2)说话人由点(3.6 m,4.2 m)开始作逆时针半圆形轨迹运动,停止于点(3.2 m,0.65 m)。

根据仿真结果,从不同观测特征和不同跟踪模型中分别选取了性能较好的观测特征和跟踪模型来对实测数据进行跟踪分析,第 1 种算法采用 MFARW 算法,第 2 种采用 MFACV 算法,第 3 种采用 SBF-SRP 作为观测特征,AUPF 作为滤波算法,称作 SBFSRPAUPF 算法,第 4 种采用本文 MFAUPF 算法。算法跟踪结果如图 6~9 所示,位置平均均方根误差如表 3 所示。

从图 6、7 及表 3 可知,采用随机行走模型的 MFARW 算法跟踪转向轨迹性能差,不能描述强混响下说话人的真实运动状态,采用 CV 模型的 MFACV 算法在说话人转向机动时收敛速度慢、跟踪误差较大,SBFSRPAUPF 算法尽管采用了 AUPF,但由于波束能量提取的位置信息精度不高,因此跟踪性能不佳,MFAUPF 算法跟踪轨迹最

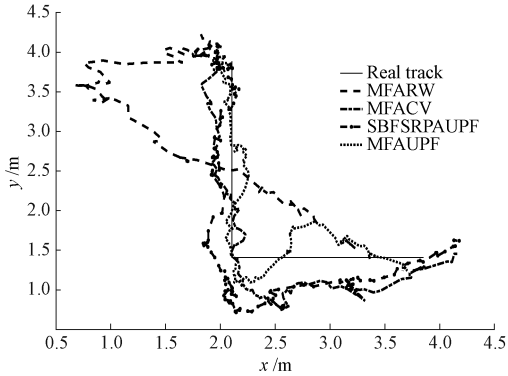

图 6　转向轨迹

Fig. 6　Steering track

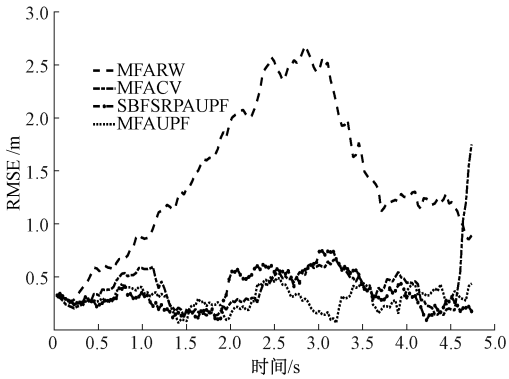

图 7　转向轨迹位置估计 RMSE

Fig. 7　RMSE of location estimation for steering track

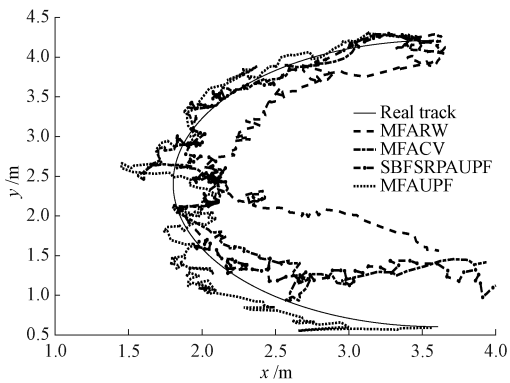

图 8　圆弧轨迹

Fig. 8　Circular arc track

接近真实轨迹,其位置平均均方根误差最小,跟踪性能受混响累积效应影响小。

由图 8、9 及表 3 可知,圆弧运动时,受室内混响和说话人行走偏离预定轨迹的影响,MFARW 算法跟踪误差最大,MFAUPF 算法位置平均均方根误差最小,随着时间推移,MFARW 算法、MFACV 算法和 SBFSRPAUPF 算

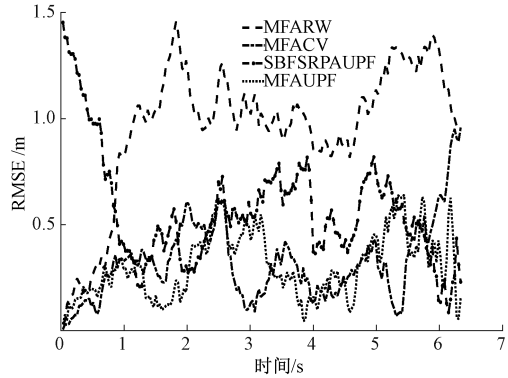

图 9　圆弧轨迹位置估计 RMSE

Fig. 9　RMSE of location estimation for circular arc

表 3　实际轨迹平均均方根误差比较表

Table 3

(m)

算法	转向机动 位置平均 RMSE	圆弧机动 位置平均 RMSE
MFARW	1. 432 5	0. 965 1
MFACV	0. 404 2	0. 362 8
SBFSRPAUPF	0. 358 7	0. 556 6
MFAUPF	0. 292 7	0. 310 5

法都出现了不同程度的偏离,MFAUPF 算法受混响累积效应影响小,跟踪性能基本不变。实测结果同仿真结果以及理论分析相吻合。

实测过程中造成跟踪误差较大的主要原因是房间混响大、麦克风定位存在误差以及说话人行走偏离预定轨迹且说话人发声具有一定的指向性等。

4.4　结果分析

仿真和实测结果表明,多特征观测机制降低了混响和噪声对滤波性能的影响,AUPF 算法提高了系统的适应性能,在低信噪比和强混响下,本文所提算法具有较高的估计精度和较强的鲁棒性。

5　结　　论

本文提出了一种 MFAUPF 算法,利用多特征优化机制从麦克风接收信号中有效地提取了最佳多特征观测量,解决了噪声混响环境下声源定位精度低的问题。针对说话人运动的随机性,建立了自适应匀速模型,引入当前观测信息和实时调节因子来获取后验分布,提高了声源突变运动跟踪时的鲁棒性。仿真和实测结果表明,本文所提算法在低信噪比强混响下均取得了良好的跟踪效果,并能有效地处理说话人的不同运动模式。

参考文献

[1] EVERS C, LOELLMANN H, MELLMANN H, et al. The LOCATA challenge: Acoustic source localization and tracking [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 1620-1643.

[2] CHAKRABARTY S, HABETS E A P. Multi-speaker DOA estimation using deep convolutional networks trained with noise signals[J]. IEEE Journal of Selected Topics in Signal Processing, 2019, 13(1): 8-21.

[3] VERA-DIAZ J M, PIZARRO D, MACIAS-GUARASA J. Acoustic source localization with deep generalized cross correlations[J]. Signal Processing, 2021, 187(2): 1-22.

[4] 李保伟, 张兴敢. 基于广义互相关改进的麦克风阵列声源定位方法[J]. 南京大学学报(自然科学), 2020, 56(6): 917-922.
LI B W, ZHANG X G. Improved microphone array sound source localization method based on generalized cross correlation[J]. Journal of Nanjing University (Natural Sciences), 2020, 56(6): 917-922.

[5] 程方晓, 刘璐, 姚清华, 等. 基于改进时延估计的声源定位算法[J]. 吉林大学学报(理学版), 2018, 56(3): 681-687.
CHENG F X, LIU L, YAO Q H, et al. Acoustic source localization algorithm based on improved time delay estimation [J]. Journal of Jilin University (Science Edition), 2018, 56(3): 681-687.

[6] MARXIM R B B, MOHANTY A R. Time delay estimation in reverberant and low SNR environment by EMD based maximum likelihood method [J]. Measurement, 2019, 137: 655-663.

[7] PERTILA P, PARVIAINEN M. Time difference of arrival estimation of speech signals using deep neural networks with integrated time-frequency masking [C]. IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), 2019: 436-440.

[8] 赵小燕, 陈书文, 周琳. 基于频率信噪比加权的麦克风阵列声源定位算法[J]. 信号处理, 2020, 36(3): 449-456.
ZHAO X Y, CHEN SH W, ZHOU L. Sound source localization using SNR-based frequency weighting with microphone array [J]. Journal of Signal Processing, 2020, 36(3): 449-456.

[9] BRENDEL A, ALTMANN I, KELLERMANN W. Acoustic source position estimation based on multi-feature Gaussian processes [C]. European Signal Processing Conference. Coruna: EUSIPCO, 2019: 1-5.

[10] KWAK Y, KIM D, HAM H, et al. Convolutional neural network trained with synthetic pseudo-images for detecting an acoustic source [J]. Applied Acoustics, 2021, 179(6): 1-7.

[11] JOHANSSON A M, LEHMANN E A. Evolutionary optimization of dynamics models in sequential Monte Carlo target tracking [J]. IEEE Transactions on Evolutionary Computation, 2009, 13(4): 879-894.

[12] LEVY A, GANNOT S, HABETS E A P. Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(6): 1540-1555.

[13] 曹洁, 李军, 李伟, 等. 基于自适应有限差分粒子滤波的说话人跟踪[J]. 兰州理工大学学报, 2012, 38(5): 93-97.
CAO J, LI J, LI W, et al. Speaker tracking based on adaptive finite-difference particle filteration[J]. Journal of Lanzhou University of Technology, 2012, 38(5): 93-97.

[14] 万新旺, 吴镇扬. 基于自适应频率选择的鲁棒时延估计算法[J]. 东南大学学报(自然科学版), 2010, 40(5): 890-894.
WAN X W, WU ZH Y. Robust time delay estimation algorithm based on adaptive frequency selection [J]. Journal of Southeast University (Natural Science Edition), 2010, 40(5): 890-894.

[15] 蔡卫平, 吴镇扬. 一种基于粒子滤波的鲁棒声源跟踪算法[J]. 电子测量与仪器学报, 2010, 24(5): 407-413.
CAI W P, WU ZH Y. Robust acoustic source tracking algorithm based on particle filtering [J]. Journal of Electronic Measurement and Instrument, 2010, 24(5): 407-413.

[16] ARULAMPALAM M S, MASKELL S, GORDON N, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesial tracking[J]. IEEE Transactions Signal Processing, 2002, 50(2): 174-188.

[17] YANG Y X, GAO W G. A new learning statistic for adaptive filter based on predicted residuals[J]. Progress in Natural Science, 2006, 16(8): 833-837.

[18] VAN DER MERWE R, DOUCET A, DE FREITAS N, et al. The unscented particle filter[D]. Cambridge: Cambridge University Engineering Department, 2000.

[19] TIAN Y, CHEN Z, YIN F L. Distributed Kalman filter-based speaker tracking in microphone array networks[J]. Applied Acoustics, 2015, 89(3): 71-77.

[20] LEHMANN E A, JOHANSSON A M. 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.

[21] LEE R, AN B, KANG M S, et al. Sound source localization based on GCC-PHAT with diffuseness mask in noisy and reverberant environments[J]. IEEE Access, 2020(8): 7373-7382.

[22] TRANSFELD P, MARTENS U, BINDER H, et al. Acoustic event source localization for surveillance in reverberant environments supported by an event onset detection [C]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015: 2629-2633.

作者简介

刘望生,2003 年于长春工业大学获得硕士学位,2008 年于中国舰船研究院获得博士学位,现为浙江理工大学副教授,主要研究方向为声源定位与跟踪、信号检测与处理。
E-mail: lwsh22@ hotmail. com

Liu Wangsheng received his M. Sc. degree from Changchun University of Technology in 2003, and received his Ph. D. degree from China Ship Research Institute in 2008. He is currently an associate professor at Zhejiang Sci-Tech University. His main research interests include sound source location and tracking, signal detection and processing.

潘海鹏,1992 年于天津大学获得硕士学位,现为浙江理工大学教授,主要研究方向为智能检测与控制、机器人控制等。
E-mail:pan13989896598@ 163. com

Pan Haipeng received his M. Sc. degree from Tianjin university in 1992. He is currently a professor at Zhejiang Sci-Tech University. His main research interests include intelligent detection and control, robot control, etc.

王明环,2000 年于甘肃工业大学获得学士学位,2003 年于长春工业大学获得硕士学位,2007 年于南京航空航天大学获得博士学位,现为浙江工业大学副教授,主要研究方向为机械制造及其自动化、智能检测与控制。
E-mail:wangmh@ zjut. edu. cn

Wang Minghuan received her B. Sc. degree from Gansu University of Technology in 2000, received her M. Sc. degree from Changchun University of Technology in 2003, and received her Ph. D. degree from Nanjing University of Aeronautics and Astronautics in 2007. She is currently an associate professor at Zhejiang University of Technology. Her main research interests include machinery manufacturing and automation, intelligent detection and control.