

# 基于检测前跟踪的声源跟踪算法

戴江安, 邱天爽

(大连理工大学电子信息与电气工程学部, 辽宁 大连 116024)

**摘 要:** 针对低信噪比、强混响环境中声源跟踪误差较大的问题, 提出一种基于检测前跟踪的声源跟踪算法。该算法将一种改进的可控响应功率函数作为定位函数, 通过计算粒子所在矩形区域的可控响应功率值, 改进的可控响应功率函数可以得到更加稳健的声源位置估计; 同时, 将检测前跟踪技术应用于跟踪算法中, 可以免于重复计算处于同一区域的量测值, 从而在保证跟踪精度的前提下大幅减少跟踪算法的运算量。经仿真实验验证, 改进算法在低信噪比、强混响的环境中能够得到比传统跟踪算法更精确的跟踪结果。

**关键词:** 声源跟踪; 粒子滤波; 可控响应功率; 检测前跟踪

中图分类号: TN912

文献标识码: A

## Acoustic source tracking algorithm using track before detect

DAI Jiang-an, QIU Tian-shuang

(Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China)

**Abstract:** In order to improve tracking performance in a noisy and reverberant environment, an acoustic source tracking algorithm using track before detect was proposed. This algorithm used a modified steered response power as localization function which take into account a rectangular region to achieve more robust source location estimation than steered response power function and applied track-before-detect technology to avoid reduplicate calculation of the same rectangular region so that the algorithm could reduce the computation burden without decreasing the accuracy. The simulation results verified that the proposed algorithm can achieve more accurate tracking results than traditional tracking algorithm in a noisy and reverberant environment.

**Key words:** acoustic source tracking, particle filter, steered response power, track before detect

### 1 引言

基于麦克风的声源定位与跟踪技术即通过多个麦克风接收到的信号连续对声源的位置信息进行估计, 其在远程视频会议<sup>[1]</sup>、智能机器人<sup>[2]</sup>、智能家居<sup>[3]</sup>等领域均有广泛的应用。声源跟踪问题可以看作是一段时间内持续的声源定位问题, 即使不采用任何跟踪技术, 只要对每帧接收信号连续执行定位算法也能实现跟踪。但由于日常环境中存在各种不利于定位的因素, 如噪声和混响, 以及由于声音的不连贯产生的静音间隙, 单纯的定位算法对

目标位置的估计通常存在较大误差, 同时无法满足跟踪问题的实时性要求。跟踪算法可利用当前及过去所有时刻的观测信号对目标状态进行动态估计, 可得到比单纯定位算法更加稳健的声源位置估计。

声源跟踪问题可分为单声源跟踪<sup>[4-6]</sup>和多声源跟踪<sup>[7-9]</sup>, 本文研究的是单声源跟踪问题。单声源跟踪问题的解决方案通常都是基于贝叶斯滤波框架的, 其中, 粒子滤波是解决声源跟踪这类非线性、非高斯问题的有效方法之一。多年来, 学者们对基于粒子滤波的声源跟踪算法进行了广泛的研究。

Ward 等<sup>[4]</sup>总结了基于粒子滤波的声源跟踪算法的

收稿日期: 2016-06-10; 修回日期: 2016-11-18

通信作者: 邱天爽, qiutsh@dlut.edu.cn

基金项目: 国家自然科学基金资助项目 (No.81241059, No.61172108); 国家科技支撑计划基金资助项目 (No.2012BAJ18B06)

**Foundation Items:** The National Natural Science Foundation of China (No.81241059, No.61172108), The National Key Technology R&D Program (No.2012BAJ18B06)

一般框架,并基于该算法框架提出了 4 种声源跟踪算法,其中,SRP-PL (steered response power-pseudo likelihood) 算法的跟踪效果最佳。SRP-PL 算法使用可控响应功率 (SRP, steered response power) 作为定位函数,并基于定位函数构造一种伪似然函数 (PL, pseudo-likelihood) 来计算粒子的权值。当环境中的干扰较弱时,SRP-PL 算法性能尚可。然而在强混响和低信噪比的恶劣环境中,算法的性能会变差;同时由于静音间隙的存在,跟踪算法的误差会急剧升高,甚至造成跟踪结果发散。Lehmann 等<sup>[5]</sup>提出一种结合语音活性检测 (VAD, voice activity detection) 的声源跟踪算法,能在一定程度上解决静音间隙的问题。但是由于引入了语音活性检测模块,该算法的计算量进一步变大,很难满足实时的跟踪效果。Fallon 等<sup>[6]</sup>提出一种基于检测前跟踪 (TBD, track before detect) 的声源跟踪算法框架。在基于 TBD 的声源跟踪算法框架中,声源的状态可以直接从似然方程中进行估计,从而省去了 VAD 模块;同时,该算法将空间划分为一定的像素网格,位于同一像素网格的粒子,其权值只需计算一次,从而显著减少了跟踪算法的运算量。然而文献[6]的跟踪算法采用 SRP 作为定位函数,其在干扰环境中的跟踪效果仍然需要改善。近年来,一些学者又相继提出了结合  $\alpha$  稳定分布<sup>[10]</sup>、分布式处理技术<sup>[11,12]</sup>等方法,为解决声源跟踪问题提供了新的思路。

本文提出一种新的声源跟踪算法,简称 MSRP-TBD (modified steered response power-track before detect)。该算法使用一种改进的可控功率响应函数<sup>[13]</sup>作为定位函数,并将其运用于 TBD 的声源跟踪算法框架中。在强混响和低信噪比的恶劣环境中,本文算法能够获得比传统的 SRP-PL 算法更好的跟踪效果,仿真实验验证了本文算法的有效性。

## 2 声源跟踪问题的基本模型

### 2.1 信号模型

本文研究封闭环境中的单声源二维跟踪问题。假设声源信号为  $S(t)$ , 环境中共有  $M$  个麦克风,第  $m$  个麦克风 ( $m=1,2,\dots,M$ ) 接收到的信号  $X_m(t)$  为

$$X_m(t) = h_m(t) * S(t) + N_m(t) \quad (1)$$

其中,  $h_m(t)$  为声源与第  $m$  个麦克风之间的冲激响应,  $N_m(t)$  为第  $m$  个麦克风接收信号中的噪声,  $*$  表示卷积运算。

### 2.2 贝叶斯滤波框架

针对本文情况,声源在  $t$  时刻状态可表示为

$$\boldsymbol{\alpha}_t = [x_t, y_t, \dot{x}_t, \dot{y}_t]^T \quad (2)$$

其中,“T”表示转置运算,  $\boldsymbol{\ell}_t = [x_t, y_t]^T$  表示声源  $x$  轴和  $y$  轴的坐标矢量,  $[\dot{x}_t, \dot{y}_t]^T$  表示声源沿  $x$  轴和  $y$  轴的速度分量。假设系统的状态转移是一阶马尔可夫过程,则声源跟踪问题可用以下 2 个方程来描述。

$$\boldsymbol{\alpha}_t = T(\boldsymbol{\alpha}_{t-1}, u_t) \quad (3)$$

$$Y_t = O(\boldsymbol{\alpha}_t, v_t) \quad (4)$$

其中,式(3)称为状态转移方程,式(4)称为观测方程,  $Y_t$  表示目标状态的观测值,通过定位函数转换得到。  $u_t$ 、 $v_t$  表示噪声项。  $T(\cdot)$ 、 $O(\cdot)$  分别表示转移函数和观测函数。贝叶斯滤波理论是声源跟踪问题的常用解决方法,其基本思想是根据目标的先验信息和观测信息估计出当前时刻目标状态的后验概率密度函数。根据贝叶斯滤波理论,声源跟踪问题可归纳为以下的递推形式。

$$p(\boldsymbol{\alpha}_t | Y_{1:t-1}) = \int p(\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t-1}) p(\boldsymbol{\alpha}_{t-1} | Y_{1:t-1}) d\boldsymbol{\alpha}_{t-1} \quad (5)$$

$$p(\boldsymbol{\alpha}_t | Y_{1:t}) \propto p(Y_t | \boldsymbol{\alpha}_t) p(\boldsymbol{\alpha}_t | Y_{1:t-1}) \quad (6)$$

其中,  $Y_{1:t} = \{Y_1, \dots, Y_t\}$  表示从初始时刻到  $t$  时刻的观测值,  $p(\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t-1})$  表示状态转移概率密度,  $p(\boldsymbol{\alpha}_t | Y_{1:t-1})$  表示状态预测概率密度,  $p(Y_t | \boldsymbol{\alpha}_t)$  表示似然函数。式(5)称为预测,式(6)称为更新。假设  $t-1$  时刻的后验概率密度  $p(\boldsymbol{\alpha}_{t-1} | Y_{1:t-1})$  已知,则通过式(5)和式(6)的递推贝叶斯滤波可以得到  $t$  时刻状态的后验概率密度。

## 3 SRP-PL 声源跟踪算法

贝叶斯滤波框架看似为跟踪问题提供了完美的解决方案,然而由于声源跟踪问题模型式(3)和式(4)中的  $T(\cdot)$ 、 $O(\cdot)$  通常是非线性函数,且噪声项  $u_t$ 、 $v_t$  可能是非高斯噪声,因此,式(5)和式(6)通常不存在闭式解。而粒子滤波正是解决声源跟踪这一类非线性非高斯问题的有力工具。粒子滤波的基本思想是使用一组具有相应权值的随机样本来表示状态的后验概率分布。基于粒子滤波的声源跟踪算法框架参考文献[4],文献[4]提出的 SRP-PL 算法性能优于文中其他 3 种算法,下面对 SRP-PL 算法做简要介绍。

### 3.1 定位函数

定位函数的功能是将麦克风接收到的原始语音信号转换为目标状态的观测值。SRP-PL 算法采用 SRP 作为定位函数, SRP 函数的表达式为

$$\begin{aligned} Y_i(\boldsymbol{\ell}) &= f_{\text{SRP}}(\boldsymbol{\ell}) = \sum_{\omega \in \Omega} \left| \sum_{m=1}^M W_m(\omega) X_m(\omega) e^{j\omega c^{-1} \|\boldsymbol{\ell} - \boldsymbol{\ell}_m\|} \right|^2 \\ &= \sum_{\omega \in \Omega} \left( \sum_{k=1}^M \sum_{l=1}^M W_k(\omega) X_k(\omega) \overline{W_l(\omega) X_l(\omega)} e^{j\omega \tau_{kl}(\boldsymbol{\ell})} \right) \\ &= \sum_{k=1}^M \sum_{l=1}^M R_{kl}[\tau_{kl}(\boldsymbol{\ell})] \end{aligned} \quad (7)$$

其中,  $\boldsymbol{\ell}$  代表声源的坐标矢量,  $\boldsymbol{\ell}_m$  代表第  $m$  个麦克风的坐标矢量,  $\|\cdot\|$  代表矢量的 2 范数,  $X_m(\omega)$  代表第  $m$  个麦克风接收到的信号  $X_m(t)$  的傅里叶变换。

$W_m(\omega)$  代表权函数, 声源跟踪算法通常取相位变换 (PHAT, phase transform) 权函数  $W_m(\omega) = \frac{1}{|X_m(\omega)|}$ 。

$\Omega$  代表频率取值范围, 根据语音频率范围取  $\Omega = [0, 8]$  kHz。  $c$  代表声速, 取 343 m/s。  $\cdot$  代表取共轭复数。  $R_{kl}[\tau_{kl}(\boldsymbol{\ell})]$  代表第  $k$  个和第  $l$  个麦克风的接收信号的广义互相关函数 (GCC, generalized cross-correlation),  $\tau_{kl}(\boldsymbol{\ell}) = \frac{(\|\boldsymbol{\ell} - \boldsymbol{\ell}_k\| - \|\boldsymbol{\ell} - \boldsymbol{\ell}_l\|)}{c}$  代表声源

传播到第  $k$  个和第  $l$  个麦克风的时延值。由式(7)可以看出, SRP 函数值可以看作是各麦克风对接收信号的 GCC 值之和。

### 3.2 似然函数

似然函数的作用是根据粒子状态来评估粒子权值。SRP-PL 算法通过式(8)似然函数来计算粒子权值为

$$F(Y_r, \boldsymbol{\alpha}) = \max \{Y_i(\boldsymbol{\ell}), 0\}^r \quad (8)$$

其中,  $r \in \mathbb{R}^+$ , 其作用是使似然函数更尖锐, 从而使靠近声源的粒子获得更大的权值。

## 4 基于 TBD 和 MSRP 的声源跟踪算法

### 4.1 MSRP 定位函数

以 SRP-PL 为代表的声源跟踪算法在干扰较弱的情况下跟踪效果尚可, 但当环境中的干扰较强时, 该类算法效果变差。由于噪声和混响的影响, SRP 的谱峰会变得模糊, 同时可能会偏离声源的真实位置。

本文提出的声源跟踪算法使用一种改进的 SRP

(MSRP, modified steered response power)<sup>[13]</sup> 函数作为定位函数。传统的 SRP 函数是单个离散点的 GCC 值之和, 而 MSRP 则通过累加正方形区域的 GCC 值来计算该像素区域的 SRP 函数值, 可以获得比 SRP 函数更加顽健的定位效果, 其表达式为

$$f_{\text{MSRP}}(\boldsymbol{\ell}) = \sum_{k=1}^M \sum_{l=k+1}^M \sum_{\tau=L_{kl_1}(\boldsymbol{\ell})}^{L_{kl_2}(\boldsymbol{\ell})} R_{kl}(\tau) \quad (9)$$

其中,  $L_{kl_2}(\boldsymbol{\ell})$  和  $L_{kl_1}(\boldsymbol{\ell})$  分别代表粒子所处区域时延值的上限和下限, 其计算表达式为

$$L_{kl_1}(\boldsymbol{\ell}) = \tau_{kl}(\boldsymbol{\ell}) - \|\nabla \tau_{kl}(\boldsymbol{\ell})\| d \quad (10)$$

$$L_{kl_2}(\boldsymbol{\ell}) = \tau_{kl}(\boldsymbol{\ell}) + \|\nabla \tau_{kl}(\boldsymbol{\ell})\| d \quad (11)$$

其中,  $\tau_{kl}(\boldsymbol{\ell})$  代表像素中心的时延值,  $d$  代表从像素中心沿时延梯度方向到边界的最短距离,  $\nabla \tau_{kl}(\boldsymbol{\ell})$  代表时延函数的梯度值。

### 4.2 MSRP-TBD 的声源跟踪算法

MSRP 函数虽然是一种比 SRP 函数更加顽健的定位函数, 然而由于其需要计算像素区域的多个 GCC 值。若直接将 MSRP 运用于文献[4]的算法框架中 (文中简称 MSRP-PL 算法), 则其运算量相比于 SRP-PL 会大幅提升。本文提出的声源跟踪算法结合了检测前跟踪的声源跟踪算法框架<sup>[9]</sup>, 能大幅减少因使用 MSRP 而大幅增加的运算量, 从而满足跟踪算法的实时性要求。

本文算法结合了 MSRP 和 TBD 技术, 简称 MSRP-TBD。与 SRP-PL 算法计算每个离散点的观测值不同, MSRP-TBD 算法是基于空间像素区域进行观测值计算的。将观测区域划分成  $I \times J$  的像素网格, 对像素  $(i, j)$  取其几何中心的 MSRP 函数值  $z_{ij}(k)$  作为观测值, 则区域的观测集可表示为

$$Z(k) = \{z_{ij}(k) : i = 1, \dots, I, j = 1, \dots, J\} \quad (12)$$

假设声源只对其所在像素的观测值有影响, 故似然函数可表示为

$$p(Z|\boldsymbol{\alpha}) = \prod_{i,j} p(z_{ij}|\boldsymbol{\alpha}) = \prod_{i,j \in C(\boldsymbol{\alpha})} p_{S+N}(z_{ij}) \prod_{i,j \notin C(\boldsymbol{\alpha})} p_N(z_{ij}) \quad (13)$$

$$C(\boldsymbol{\alpha}) = \left\{ (i, j) : \left| i\Delta + \frac{\Delta}{2} - x \right| < \frac{\Delta}{2}, \left| j\Delta + \frac{\Delta}{2} - y \right| < \frac{\Delta}{2} \right\} \quad (14)$$

其中,  $C(\boldsymbol{\alpha})$  表示声源所在像素的下标,  $\boldsymbol{\alpha}$  表示目标状态矢量,  $\Delta$  表示正方形像素的边长,  $p_N(z_{ij})$  代

表只存在噪声时观测值的似然函数,  $p_{S+N}(z_{ij})$  代表噪声与声源并存时观测值的似然函数。由于粒子的权值只是相对的比例因子, 将式(13)除以  $\prod_{i,j} p_N(z_{ij})$ , 可得似然函数为

$$q(Z|\alpha) \propto \prod_{i,j \in C(\alpha)} l(z_{ij}) \quad (15)$$

$$l(z_{ij}) = \frac{p_{S+N}(z_{ij})}{p_N(z_{ij})} \quad (16)$$

式(15)的似然函数只取决于声源所在像素的观测值, 因此, 处于同一像素的粒子权值相等。而采用上述方法时, 对处于同一像素的多个粒子, 其权值只需计算一次。相比于传统粒子滤波方法需计算所有粒子的权值, 上述方法的计算量会显著减少。然而由于粒子分布的随机性以及MSRP函数累加上下限的不确定性, 使其运算量难以量化分析, 本文不做深入讨论。

MSRP-TBD 算法的观测值通过式(9)的MSRP函数进行计算。然而此时  $p_{S+N}(z_{ij})$  并不满足正态分布这一条件<sup>[14]</sup>, 无法直接应用于TBD的算法框架中, 因此, 对观测值  $z_{ij}$  进行以下的非线性映射操作

$$z'_{ij} = \Phi(z_{ij}; \bar{Z}, \sigma_z^2) \quad (17)$$

其中,  $\Phi$  表示以  $\bar{Z}$  为均值, 以  $\sigma_z^2$  为方差的正态分布的累积分布函数。 $\bar{Z}$  和  $\sigma_z^2$  的大小与MSRP值的分布情况相关。经非线性映射操作后, 将式(17)代入式(16)可近似得

$$l(z'_{ij}) = \frac{p_{S+N}(z'_{ij})}{p_N(z'_{ij})} = \frac{c_{S+N} \mathcal{N}(z'_{ij}; 1, \sigma_{S+N}^2)}{c_N \mathcal{N}(z'_{ij}; 0, \sigma_N^2)} \quad (18)$$

其中,  $\sigma_{S+N} = \sigma_N = 0.5$ , 则  $c_{S+N} = c_N$ , 式(18)可简化为

$$l(z'_{ij}) = \exp\left[\frac{2z'_{ij} - 1}{2\sigma_N^2}\right] \quad (19)$$

为区分有声和无声段, 在状态变量中添加语音活性指示子  $\lambda_t \in \{0, 1\}$ , 其中, “1”代表处于有声段, “0”代表处于无声段, 则声源在  $t$  时刻状态可改写为

$$\alpha_t = [x_t, y_t, \dot{x}_t, \dot{y}_t, \lambda_t]^T \quad (20)$$

其中, 指示子  $\lambda_t$  根据给定的转移概率进行一阶马尔科夫状态转移。

由状态  $a$  转变为  $b$  的概率表达式为

$$\text{Prob}\{\lambda_{k+1} = a | \lambda_k = b\} = \Pi_{a,b} \quad (21)$$

其中,  $a, b \in \{0, 1\}$ ,  $\Pi_{1,0}$  称为声源出现概率,  $\Pi_{0,1}$  称为声源消失概率, 其取值将在后续的仿真实验中具体讨论。

加入指示子  $\lambda_t$  后, 式(15)的似然函数可改写为

$$q'(Z|\alpha) \propto \prod_{\lambda=1, (i,j) \in C(\alpha)} l(z'_{ij}) \quad (22)$$

当处于有声段时, 粒子将按式(22)计算权值。可以看出, 若粒子与声源处在同一像素网格, 则式(22)的函数值会远大于 1; 若粒子位于声源所在像素网格之外的区域, 则式(22)的函数值会远小于 1。当处于无声段, 粒子的权值将设为 1。通过这样的操作, 在重采样阶段, 处于正确状态的粒子总会具有较大的权值, 可被筛选出来。

## 5 仿真实验和结果分析

为验证本文提出的MSRP-TBD算法的有效性, 本文进行了多组仿真实验。实验中模拟了小型房间的声学环境, 房间尺寸为  $6\text{ m} \times 4\text{ m} \times 3.5\text{ m}$ , 在高度为  $1\text{ m}$  的水平空间分布有 6 个麦克风, 其分布如图 1 所示。房间冲激响应通过 Image 方法<sup>[15]</sup>产生, 仿真实验模拟了从坐标为(2,1,1)到坐标(4,3,1)的一段直线运动轨迹。

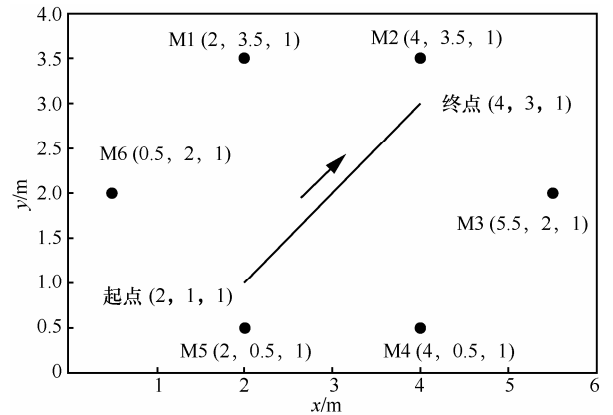


图1 仿真实验中麦克风分布情况

跟踪算法效果用均方根误差 (RMSE, root mean square error) 来评价, 其表达式为

$$E_{RMSE} = \sqrt{\frac{1}{N_S} \sum_{t=1}^{N_S} \|\ell_t - \hat{\ell}_t\|^2} \quad (23)$$

其中,  $N_S$  代表信号帧数,  $\ell_t$  代表第  $t$  帧时声源位置的真实值,  $\hat{\ell}_t$  代表第  $t$  帧时声源位置的估计值。

实验语音为 IEEE-Harvard 语音库的一段 3.09 s 左右的语句，语音的内容为 “The birch canoe slid on the smooth planks”，其波形如图 2 所示。其中，语音的采样率为 25 kHz，划分为 151 帧。FFT 变换长度为 512 点，每帧信号长度为 512 点，帧间不重叠，选用汉明窗函数。3 种算法均采用 100 个粒子，MSRP-TBD 算法的正方形像素边长为 0.05 m，仿真实验 1~仿真实验 3 中声源出现/消失概率取 0.05。仿真实验中的背景噪声为高斯白噪声，环境中的混响强弱通过声音在环境中传播声强减弱 60 dB 所消耗时间 ( $Rt_{60}$ ) 来描述。

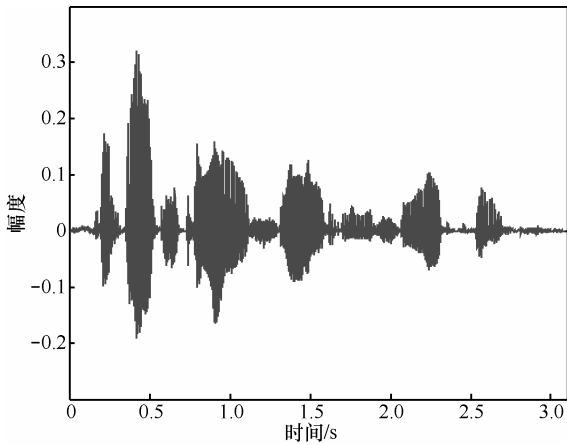


图 2 仿真实验采用的语音波形

1) 仿真实验 1

对比单纯噪声和单纯混响的环境下 SRP-PL、MSRP-PL 和 MSRP-TBD 这 3 种算法的  $RMSE$  值，每种算法均进行 100 次蒙特卡洛仿真。图 3 为 3 种算法在不同强度的单纯噪声环境下的  $RMSE$  值，图 4 为 3 种算法在不同强度的单纯混响环境下的  $RMSE$  值。

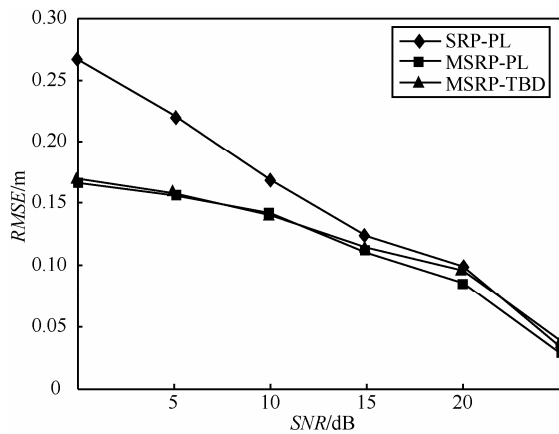


图 3 单纯噪声环境下算法性能对比

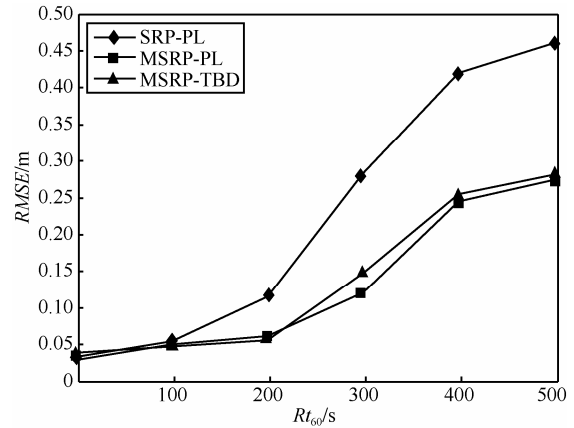


图 4 单纯混响环境下算法性能对比

从图 3 和图 4 的实验结果可以看出，在噪声和混响较弱的环境下，3 种算法  $RMSE$  值均较小；随着环境干扰因素逐渐加强，本文提出的 MSRP-TBD 算法的  $RMSE$  值与 MSRP-PL 算法接近，但 2 种算法的  $RMSE$  值均要显著小于 SRP-PL 算法。

2) 仿真实验 2

对比 SRP-PL、MSRP-PL 和 MSRP-TBD 这 3 种算法的运算时间。每种算法在  $SNR = 5 \text{ dB}$ ， $Rt_{60} = 250 \text{ ms}$  的声学环境中进行 100 次蒙特卡洛仿真。其中，SRP-PL 算法执行一次平均用时约为 2.5 s；MSRP-PL 执行一次平均用时约为 7 s；而 MSRP-TBD 算法平均每次用时约为 3 s。可以看出，MSRP-PL 算法相比于 SRP-PL 算法运算量会大幅增加；而 MSRP-TBD 算法运算时间与 SRP-PL 算法相近，不会显著提升运算量。

3) 仿真实验 3

在  $SNR = 20 \text{ dB}$ 、 $Rt_{60} = 50 \text{ ms}$  和  $SNR = 5 \text{ dB}$ 、 $Rt_{60} = 250 \text{ ms}$  这 2 种典型声学环境中，对 SRP-PL、MSRP-PL 和 MSRP-TBD 这 3 种算法中特定 7 帧语音的跟踪误差值进行对比，实验结果如表 1 和表 2 所示。

表 1  $SNR=20 \text{ dB}$ 、 $Rt_{60}=50 \text{ ms}$  环境下的跟踪误差 (单位: m)

帧号	SRP-PL	MSRP-PL	MSRP-TBD
20	0.039 1	0.031 6	0.044 5
40	0.023 6	0.022 7	0.021 6
60	0.023 7	0.026 3	0.019 5
80	0.032 7	0.028 1	0.026 0
100	0.042 5	0.018 6	0.014 6
120	0.195 7	0.155 4	0.174 9
140	0.168 2	0.163 5	0.146 9

表 2 SNR=5 dB、 $Rt_{60}=250$  ms 环境下的跟踪误差 (单位: m)

帧号	SRP-PL	MSRP-PL	MSRP-TBD
20	0.145 5	0.102 3	0.092 1
40	0.359 9	0.068 4	0.053 2
60	0.341 0	0.075 8	0.083 3
80	0.171 7	0.089 4	0.098 7
100	0.317 6	0.246 9	0.242 3
120	0.462 1	0.106 1	0.085 3
140	0.573 1	0.303 2	0.277 6

从表 1 可以看出, 当噪声和混响较弱时, 3 种算法的跟踪误差比较接近。从表 2 可以看出, 当环境干扰较强时, MSRP-TBD 算法和 MSRP-PL 算法的跟踪误差要显著小于 SRP-PL 算法。另外, 需要说明的是, 第 120 帧和第 140 帧的跟踪误差相对其他帧较大, 原因是其处于语音的静音间隙和语音结尾处的无声段。

4) 仿真实验 4

在不同的声源出现/消失概率下, 分析本文提出的 MSRP-TBD 算法的跟踪精度和平均计算时间。每组均在  $SNR = 5$  dB、 $Rt_{60} = 250$  ms 的声学环境中进行 100 次蒙特卡洛仿真。

首先, 考虑声源出现概率与消失概率相等的情况, 实验结果如图 5 和图 6 所示。

从图 5 可以看出, 当声源出现/消失概率取值过大或过小时,  $RMSE$  值都较大; 概率取 0.025 到 0.1 之间的值时  $RMSE$  值较小。从图 6 可以看出, 随着出现/消失概率变大, 平均计算时间会略微减少, 这主要是因为当判断粒子为静音态时, 不需要计算相对复杂的 MSRP 值, 计算量较小。

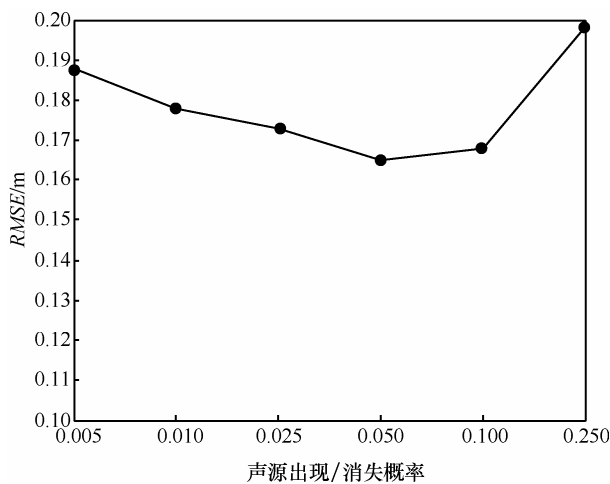


图 5 不同声源出现/消失概率下的  $RMSE$  值

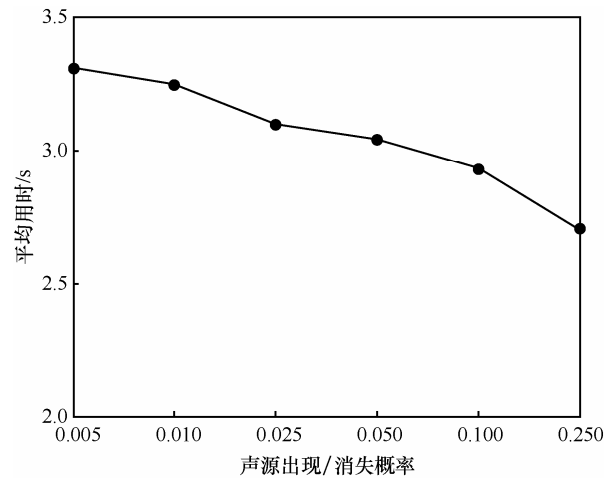


图 6 不同声源出现/消失概率下的平均计算时间

然后, 使出现概率、消失概率其中一个固定为 0.05, 另外一个概率值在 0.01 到 0.1 之间变化, 观察跟踪结果  $RMSE$  值的变化情况, 实验结果如图 7 和图 8 所示。

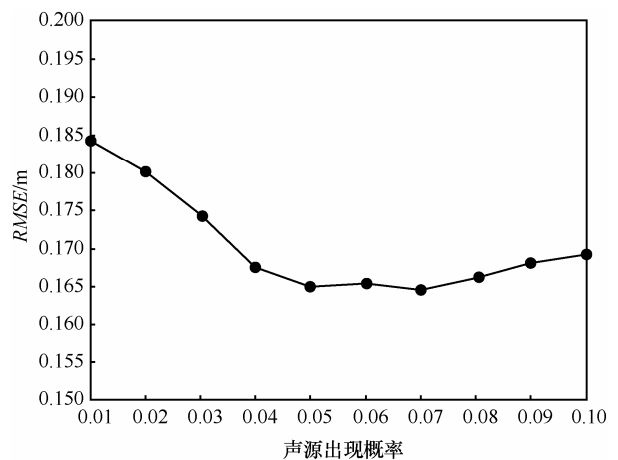


图 7 声源出现概率对跟踪  $RMSE$  值的影响

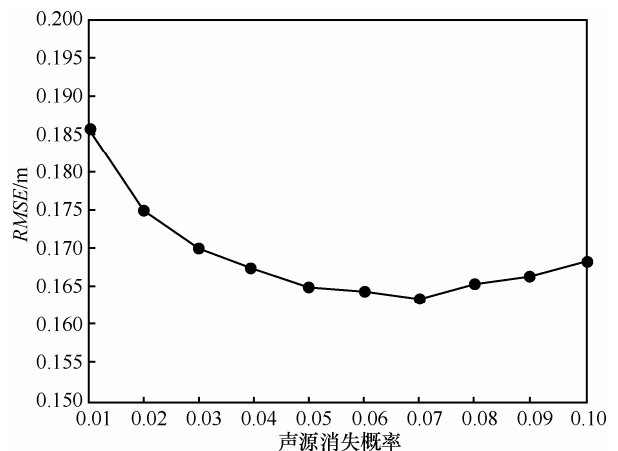


图 8 声源消失概率对跟踪  $RMSE$  值的影响

从图7可以看出,当声源消失概率固定为0.05、声源出现概率取0.04到0.08之间的值时,跟踪结果 $RMSE$ 较小。从图8可以看出,当声源出现概率固定为0.05、声源消失概率取0.05到0.09之间的值时,跟踪结果 $RMSE$ 较小。

从以上实验结果可以看出,针对本文情况,声源出现/消失概率取0.025到0.1之间的值时比较合适,本文其他仿真实验中出现/消失概率取0.05。

## 6 结束语

以SRP-PL算法为代表的传统声源跟踪算法在低信噪比和强混响的干扰环境中性能会急剧恶化。本文提出一种基于检测前跟踪的声源跟踪算法。该算法使用MSRP作为定位函数,由于MSRP函数计算区域的而非单点的GCC值,可以使跟踪算法获得更好的顽健性。同时,MSRP-TBD算法兼具TBD算法框架运算量小的优点,能够在不显著提高运算量的情况下,使跟踪算法的性能得到较大的改善。仿真实验验证了MSRP-TBD算法的有效性。

## 参考文献:

- [1] BILL K, JENKIN M R M, EVANGELOS M. Audio-visual localization of multiple speakers in a video teleconferencing setting[J]. International Journal of Imaging Systems and Technology, 2003, 13(1):95-105.
- [2] MARKOVIĆ I, PETROVIĆ I. Speaker localization and tracking with a microphone array on a mobile robot using von Mises distribution and particle filtering[J]. Robotics and Autonomous Systems, 2010, 58(11): 1185-1196.
- [3] 杨轶星, 刘永敬, 焦学军, 等. 智能家居系统中人体定位技术研究进展[J]. 中国生物医学工程学报, 2013, 32(6):716-722.
- YNAG Y X, LIU Y J, JIAO X J, et al. Research progress on human body localization in smart home system[J]. Chinese Journal of Biomedical Engineering, 2013, 32(6):716-722.
- [4] WARD D B, LEHMANN E A, WILLIAMSON R C. Particle filtering algorithms for tracking an acoustic source in a reverberant environment[J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(6): 826-836.
- [5] LEHMANN E A, JOHANSSON A M. Particle filter with integrated voice activity detection for acoustic source tracking[J]. EURASIP Journal on Advances in Signal Processing, 2007, 2007(1):1-11.
- [6] FALLON M F, GODSILL S. Acoustic source localization and tracking using track before detect[J]. IEEE Transactions on Audio Speech

and Language Processing, 2010, 18(6):1228-1242.

- [7] ZHONG X, HOPGOOD J R. Particle filtering for TDOA based acoustic source tracking: nonconcurrent multiple talkers[J]. Signal Processing, 2014, 96(5):382-394.
- [8] SCHWARTZ O, GANNOT S. Speaker tracking using recursive EM algorithms[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2014, 22(2):392-402.
- [9] ZHONG X, HOPGOOD J R. A time frequency masking based random finite set particle filtering method for multiple acoustic source detection and tracking[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 23(12):2356-2370.
- [10] ZHONG X, PREMKUMAR A B, MADHUKUMAR A S. Particle filtering for acoustic source tracking in impulsive noise with alpha-stable process[J]. IEEE Sensors Journal, 2013, 13(2):589-600.
- [11] ZHANG Q L, CHEN Z, YIN F L. Global coherence field and distributed particle filter-based speaker tracking in distributed microphone networks[J]. Journal of Computational Acoustics, 2015, 23(3).
- [12] TIAN Y, CHEN Z, YIN F. Distributed IMM-unscented Kalman filter for speaker tracking in microphone array networks[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 23(10): 1637-1647.
- [13] COBOS M, MARTI A, LOPEZ J J. A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling[J]. IEEE Signal Processing Letters, 2011, 18(1):71-74.
- [14] SALMOND D J, BIRCH H. A particle filter for track-before-detect[C]// Proceedings of the American Control Conference, 2001, 5:3755-3760.
- [15] 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, 123(11): 4654-4666.

## 作者简介:



戴江安(1991-),男,江西抚州人,大连理工大学博士生,主要研究方向为语音信号处理、雷达信号处理和无线电信号处理。



邱天爽(1954-),男,江苏海门人,博士,大连理工大学教授、博士生导师,主要研究方向为统计信号处理、射频与通信信号处理等。