

# 基于分层采样粒子滤波的麦克风阵列说话人跟踪方法

金乃高, 殷福亮, 陈

(大连理工大学电子与信息工程学院, 辽宁大连 116024)

**摘 要:** 针对噪声与混响环境下的说话人跟踪问题, 本文提出了一种基于粒子滤波的麦克风阵列声源定位与跟踪方法. 该方法在粒子滤波框架下, 将无混响影响的语音建立信号作为观测信息, 通过计算麦克风阵列波束形成器的输出能量来构建似然函数, 同时考虑语音信号不同频率成分在声源定位中的作用, 利用分层采样方法提高粒子的采样效率. 实验结果表明, 本文方法提高了说话人声源跟踪系统的抗噪声与抗混响能力.

**关键词:** 说话人跟踪; 麦克风阵列; 粒子滤波; 分层采样; 房间混响

**中图分类号:** TN911 **文献标识码:** A **文章编号:** 0372-2112 (2008) 01-0194-05

## Microphone Array Speaker Tracking Using Layered Sampling Particle Filtering

JIN Naigao, YIN Fuliang, CHEN Zhe

(School of Electronic and Information Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China)

**Abstract:** A new sound source localization and tracking method based on microphone array is presented for speaker tracking problem in noisy and reverberant environments. In the framework of particle filtering, the echo free onset signal is extracted according to the echo avoidance model of the precedence effect, and likelihood function is constructed by the output power of beamformer. Considering the effects of different frequency components on localization and tracking, layered sampling approach is used to improve the computational efficiency of particle filter. The results of simulation data and real data show that the proposed method can reliably track the speaker even in moderately reverberant rooms.

**Key words:** speaker tracking; microphone array; particle filter; layered sampling; room reverberation

### 1 引言

说话人跟踪作为人机交互中的重要环节, 在多媒体系统、人机界面、视频会议系统以及移动机器人等领域有着广泛的应用. 例如, 在视频会议系统中, 说话人跟踪可为摄像机转向控制与基于麦克风阵列的语音拾取提供方向信息. 另外, 移动机器人也需要根据说话人的空间位置进行路径规划. 说话人跟踪问题可以通过基于计算机视觉的人脸跟踪方法实现<sup>[1]</sup>, 但这类方法易受到光照条件、人脸姿态以及遮挡等因素的影响, 且跟踪范围受摄像机视角限制. 与人脸跟踪方法相比, 基于麦克风阵列的声源定位与跟踪方法具有全方位跟踪能力, 且运算量较低, 易于构建实时系统, 已成为说话人跟踪问题的主要解决方案<sup>[2]</sup>.

在实际的中小型房间环境中, 由于墙壁、地面、桌面的反射作用, 使得麦克风阵列接收到的语音信号中含有混响成分, 从而影响声源定位与跟踪系统的性能. DiBiase

提出的联合波束形成-相位变换(Steered Response PHase Transform, SRP-PHAT)声源定位方法<sup>[3]</sup>具有一定的抗噪性、抗混响性能力, 但在较恶劣的声学环境下, 虚假声源的出现会严重影响系统的定位能力. 为了克服房间混响对声源定位的影响, Huang 等借鉴人耳定位机制, 利用语音的建立信号(Onset Signal)进行声源定位, 在混响较强的条件下取得了较好的定位效果<sup>[4]</sup>. 在说话人跟踪问题中, 虚假声源的出现位置具有明显的不确定性, 而说话人的运动通常符合特定的动态模型. 据此, 文献[5, 6]通过建立合理的说话人运动模型来抑制虚假声源的影响, 分别根据时延估计与波束形成的输出能量构造似然函数, 通过粒子滤波实现说话人声源的有效跟踪.

本文综合考虑语音建立信号与说话人运动模型在抑制房间混响中的作用, 提出了一种基于分层采样粒子滤波的麦克风阵列说话人跟踪方法. 该方法在粒子滤波框架下, 将语音的建立信号作为观测信息, 根据 SRP-PHAT 波束形成器的输出能量构造粒子滤波中的似然函

数;同时考虑到语音信号不同频率成分对定位的贡献,采用分层采样方法生成采样粒子,以提高粒子的采样效率.实验结果表明,本文方法能够在复杂环境下实现说话人声源的有效跟踪.

## 2 基于粒子滤波的说话人跟踪方法

### 2.1 SRP-PHAT 声源定位算法

声源信号  $s(t)$  经多径传播后到达麦克风阵列,第  $i$  个麦克风接收到的信号  $m_i(t)$  可以表示为

$$m_i(t) = s(t) * h_i(t) + v_i(t) \quad (1)$$

其中  $v_i(t)$  为噪声,  $h_i(t)$  是声源与第  $i$  个麦克风之间的冲激响应,“\*”为卷积运算符.

DiBiase 提出的 SRP-PHAT 声源定位算法将波束形成方法的鲁棒性、短时分析特性与相位变换方法对环境的不敏感性相结合,使声源定位系统对噪声和混响的灵敏度降低,提高了系统的鲁棒性和定位的精确性<sup>[3]</sup>.由于只需采用较短的数据进行分析,因此这种方法适于对运动的说话人进行实时跟踪.

设声源的空间位置向量为  $x$ , 语音信号  $m_i(t)$  的傅里叶变换为  $M_i(\omega)$ , 麦克风数目为  $M$ , 直达路径中声源信号至第  $i$  个麦克风的传播时延为  $\tau_i$ , SRP-PHAT 声源定位算法的波束输出能量  $P(x)$  定义为

$$P(x) = \sum_{i=1}^M \sum_{j=1}^M \int_{-\infty}^{\infty} \frac{M_i(\omega) M_j^*(\omega)}{|M_i(\omega) M_j^*(\omega)|} e^{j\omega(\tau_i - \tau_j)} d\omega \quad (2)$$

SRP-PHAT 声源定位算法在整个空间范围内搜索波束输出能量的最大值,从而确定声源位置.

尽管 SRP-PHAT 声源定位算法具有一定的抗噪性、抗混响性和鲁棒性,但在较恶劣的声学环境下,以及麦克风阵列的位置误差较大时,它常常会产生一些虚峰,这就增大了说话人声源所对应峰值的搜索难度,难以满足实际应用的要求.

### 2.2 语音建立信号的检测

在说话人刚刚发出声音时,语音中的混响信号总是比直达信号延迟一段时间到达,语音建立信号便是指这段先于混响信号到达的语音信号<sup>[4]</sup>.人耳优先效应(Precedence Effect)实验表明,在房间混响较强的情况下,人耳可以利用未被混响污染的语音建立信号准确地判断出声源方向.本文借鉴人耳的定位机制,从麦克风阵列获取的多路语音中提取出无混响影响的建立信号,将其作为观测信息来增强声源定位系统的抗混响能力.

房间冲激响应可建模为指数衰减形式,即

$$R_m(t) = \begin{cases} 1, & t = 0 \\ 0, & 0 < t < \tau_e \\ e^{-\alpha(t - \tau_e)/\tau} & t \geq \tau_e \end{cases} \quad (3)$$

其中第一项表示幅度归一化后的直达信号;第二项与第三项为接收信号中的混响成分,记为  $I_{echo}(t)$ , 这里  $\alpha_e$  和  $\tau_e$  分别为首次到达的混响信号幅度和时延,  $\tau$  为描述混响衰减的时间常数.

语音建立信号的检测过程如图 1 所示.设语音信号通过窄带滤波器后输出的包络为  $s(t)$ , 则  $t$  时刻最大混响信号可估计为

$$e_{echo}(t) = \max\{s(t - t_p) I_{echo}(t_p)\}, 0 \leq t_p < \infty \quad (4)$$

在语音包络信号的初始段,信号的幅度明显高于混响的幅度,可将此段语音视为无混响影响的建立信号.

本文采用短时傅立叶变换在频域提取语音建立信号.设第  $l$  帧语音信号  $m^{(l)}(t)$  的加窗短时傅立叶变换为  $M^{(l)}(k)$ , 混响衰减系数为  $\lambda$ , 根据混响的指数衰减特性,第  $l$  帧、第  $k$  频带处的最大混响幅度  $M_{echo}^{(l)}(k)$  可近似估计为

$$M_{echo}^{(l)}(k) = \max\{\lambda^n |M^{(l-n)}(k)|\}, n = 1, 2, \dots, l-1, 0 < \lambda < 1 \quad (5)$$

$M_{echo}^{(l)}(k)$  的递推计算公式为

$$M_{echo}^{(l)}(k) = \max\{M_{echo}^{(l-n)}(k), \lambda |M^{(l-1)}(k)|\} \quad (6)$$

当  $M^{(l)}(k)$  与  $M_{echo}^{(l)}(k)$  的比值大于指定阈值  $Th$  时,便可将  $M^{(l)}(k)$  视为无混响影响的建立信号.提取出的语音建立信号后,采用 SRP-PHAT 算法计算波束输出能量,并据此构造粒子滤波中的似然函数来计算粒子权值.

### 2.3 分层采样粒子滤波在说话人跟踪中的应用

粒子滤波将贝叶斯理论与蒙特卡罗方法相结合,使用非参数化的序贯蒙特卡罗方法实现递推贝叶斯滤波,其核心思想是利用一系列随机粒子及其对应权值来表示后验概率密度或滤波概率密度.

设从重要性采样函数  $\pi(x_t | y_{1:t})$  中抽样获取的粒子集为  $\{x_t^{(i)}, w_t^{(i)}, i = 1, \dots, N\}$ , 其中  $w_t^{(i)}$  为第  $i$  个粒子  $x_t^{(i)}$  的权值,则滤波概率密度  $p(x_t | y_{1:t})$  可以表示为

$$p(x_t | y_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta(x_t - x_t^{(i)}) \quad (7)$$

其中

$$w_t^{(i)} \propto p(x_t^{(i)} | y_{1:t}) [\pi(x_t^{(i)} | y_{1:t})]^{-1} \quad (8)$$

关于粒子滤波的详细介绍见文献<sup>[7]</sup>.

当存在多种不同类型的观测量或可从一种观测量

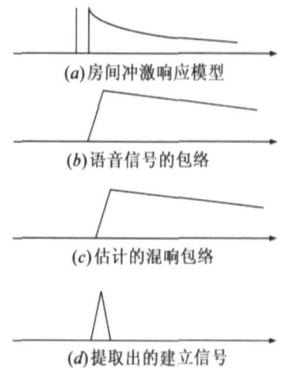


图 1 语音建立信号检测示意图

中提取出多尺度的观测信息时, 分层采样可为重要性采样函数的选择提供一种有效的方法<sup>[8]</sup>. 分层采样方法的示意图如图 2 所示, 其中圆圈代表粒子集, “~”表示重采样操作, 箭头表示从相应的概率分布中生成采样粒子,  $f_M$  为最精确的似然函数,  $f_{m-1}$  为  $f_m$  的近似逼近. 在图 2 中, 第一行描述了一般的似然加权-重采样粒子生成过程, 虚线框内则为分层采样方法的具体实现过程. 分层采样方法按照量测精度从低至高的顺序, 依次使用不同精度的似然函数计算粒子权值, 并根据权值对粒子进行重采样, 将粒子引导至高似然区域, 从而改善粒子在状态空间中的分布, 提高粒子的采样效率.

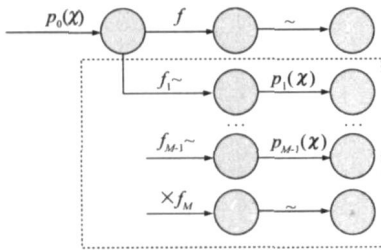


图 2 分层采样示意图

采用粒子滤波方法进行说话人跟踪时, 需要建立描述说话人运动模型的动态方程. 设  $t$  时刻说话人的坐标为  $X_t = (x_t, y_t, z_t)^T$ , 速度  $\dot{X}_t = (\dot{x}_t, \dot{y}_t, \dot{z}_t)^T$ . 本文使用 Langevin 过程建立说话人的运动模型. 假设说话人在  $x$  轴与  $y$  轴方向上的运动相互独立, 且说话人声源的高度  $z_k$  为固定值. 在  $x$  轴方向上, 说话人运动模型的状态方程可以描述为

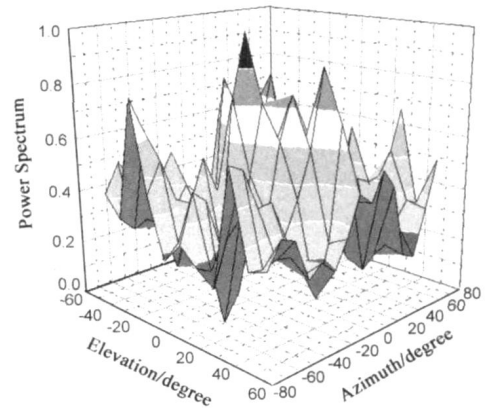
$$\begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} \begin{bmatrix} 1 & \Delta T \\ 0 & \alpha_x \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \dot{x}_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ b_x \end{bmatrix} u_x \quad (9)$$

其中  $\alpha_x = e^{-\beta_x \Delta T}$ ,  $\beta_x$  为常数;  $b_x = u_x \sqrt{1 - \alpha_x^2}$ ,  $u_x$  为稳态均方根速度;  $u_x$  为单位方差的高斯白噪声.

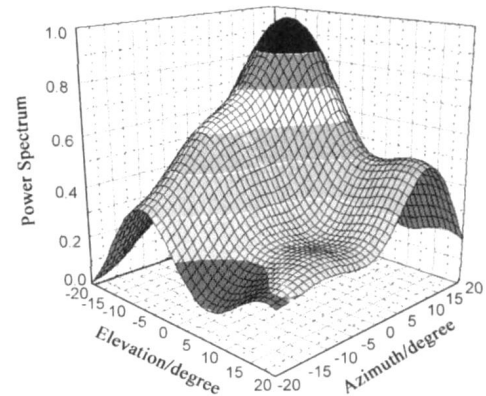
考虑到语音信号不同频率成分对定位的贡献, 在频域中采用由粗到细的搜索方法, 可以有效提高声源定位系统的抗噪声、抗混响能力<sup>[9]</sup>. 在麦克风接收到的语音信号中, 通常低频段信号的信噪比较高, 利用低频段信息进行粗略的声源搜索, 可以减小噪声干扰对定位误差的影响; 大多数材料(包括传媒空气)对高频声波吸收能力较强, 这使得语音信号中高频成分受混响的影响相对较小. 图 3 为一组实测语音信号在低频段(250Hz~3400Hz)与高频段(3400Hz~6800Hz)的空间谱图. 从图中可以看出, 与语音低频段信息所对应的空间谱相比, 高频段空间谱的平滑性较好, 故适于精确定位.

据此, 本文采用分层采样方法有效地生成采样粒子, 从而减少逼近滤波概率密度函数所需的粒子数. 分层采样方法首先根据说话人的运动模型, 从状态转移

概率密度函数  $p(x_t | x_{t-1})$  中生成采样粒子  $\{\tilde{\alpha}_t^{(i)}\}_{i=1}^N$ , 将语音建立信号低频段所对应的波束能量  $P_L$  作为似然函数, 计算粒子权值  $\{w_t^{(i)}\}_{i=1}^N$ ; 接着根据粒子权值对前一时刻的粒子进行重采样, 更新粒子集  $\{\tilde{\alpha}_t^{(i)}\}_{i=1}^N$ , 引导粒子至高似然区域; 最后计算语音建立信号高频段所对应的波束能量  $P_H$ , 将其作为似然函数重新计算粒子权值  $\{w_t^{(i)}\}_{i=1}^N$ , 从而实现说话人声源的精确定位.



(a) 低频信号的空间谱



(b) 高频信号的空间谱

图 3 语音信号的空间谱

综上所述, 基于分层采样粒子滤波的麦克风阵列说话人跟踪方法的具体步骤如下:

(1) 粒子集初始化: 令  $k=0$ , 根据初始概率密度函数生成粒子集  $\{\alpha_0^{(i)}, w_0^{(i)}\}_{i=1}^N$ , 令所有粒子的权值为  $1/N$ ;

For  $t=1, 2, \dots$

(2) 根据说话人的运动模型产生粒子  $\{\tilde{\alpha}_t^{(i)}\}_{i=1}^N$ ;

(3) 采用 2.2 节中的式(6)提取语音的建立信号;

(4) 计算语音建立信号低频段所对应的波束能量  $P_L$ , 将其作为似然函数计算粒子权值  $\{w_t^{(i)}\}_{i=1}^N$ ;

(5) 根据粒子权值  $\{w_t^{(i)}\}_{i=1}^N$  对粒子  $\{\tilde{\alpha}_t^{(i)}\}_{i=1}^N$  进行重采样, 以生成新的粒子  $\{\alpha_t^{(i)}\}_{i=1}^N$ ;

(6) 计算语音建立信号高频段所对应的波束能量  $P_H$ , 将其作为似然函数计算粒子  $\{\alpha_t^{(i)}\}_{i=1}^N$  对应的权值

$$\{w_t^{(i)}\}_{i=1}^N;$$

(7) 采用最小均方误差 (MMSE) 准则估计说话人声源的当前位置  $x_t$ ;

(8) 根据权值对粒子进行重采样, 更新粒子集;

END

### 3 实验与分析

#### 3.1 仿真实验结果

为了验证本文方法的有效性, 我们在不同的信噪比与混响环境下, 与文献[6]方法进行了对比仿真实验。仿真实验模拟普通会议室的声学环境, 其中房间的大小为  $5\text{m} \times 7\text{m} \times 3\text{m}$ , 房间混响的冲激响应函数由 IMAGE 模型<sup>[10]</sup>产生。本文采用的噪声类型为高斯白噪声, 麦克风阵列获取的语音信号以  $f_s = 44.1\text{kHz}$  的采样率进行 16 比特采样, FFT 变换长度  $L = 1024$ , 窗函数为汉明窗, 声源的高度为

固定值。麦克风阵列的摆放如图 4 所示。在图 4 中, 中心点的坐标位置为  $[0.50, 0.00, 1.00]$ , 水平麦克风之间的距离与垂直麦克风之间的距离都为  $0.24\text{m}$ 。在语音建立信号的提取过程中, 混响衰减系数  $\lambda = 0.9$ ; 阈值  $Th = 1.7$ 。

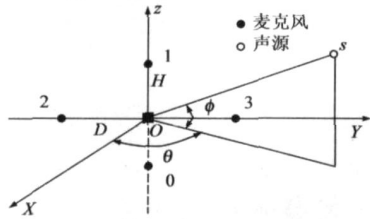


图 4 麦克风阵列摆放示意图

仿真实验 1: 在两种典型的声学环境下, 比较两种方法的仿真实验结果。在仿真实验中, 声源的起始位置为  $x_0 = [3.00, 0.30, 1.49]$ , 以  $0.1\text{m/s}$  的速度沿着  $y$  轴正向匀速运动。在粒子滤波算法中, 粒子数  $N = 500$ , 从均值为  $x_0$ 、方差为 1 的均匀分布中生成采样粒子, 采用残差重采样 (Residual Resampling) 方法降低退化现象的影响。在  $\text{SNR} = 25\text{dB}$ 、混响时间  $T_{60} = 50\text{ms}$  与  $\text{SNR} = 5\text{dB}$ 、混响时间  $T_{60} = 200\text{ms}$  两种典型声学环境下, 采用文献[6]方法和本文方法对一段含有噪声与混响的说话人语音进行连续定位, 实验结果分别如图 5 和图 6 所示。

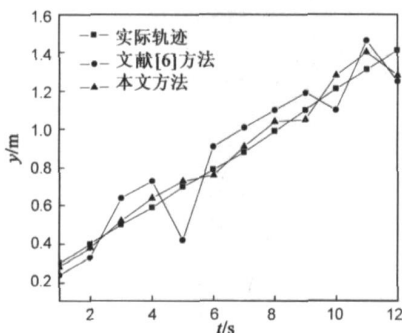


图 5  $\text{SNR} = 25\text{dB}$ ,  $T_{60} = 50\text{ms}$  环境下的测试结果 (粒子数  $N = 500$ )

从图 5 可以看出, 在信噪比较大、混响影响较弱的

环境下, 文献[6]方法在第 5 秒时出现了较大的估计误差。相比之下, 本文方法在连续跟踪的 12 秒内, 没有出现误差较大的估计值。从图 6 可以看出, 当噪声干扰和混响影响较大时, 两种方法都出现了较大的跟踪误差。与文献[6]方法相比, 本文方法的跟踪误差相对较小, 仍然能够较好地跟踪说话人的运动轨迹。

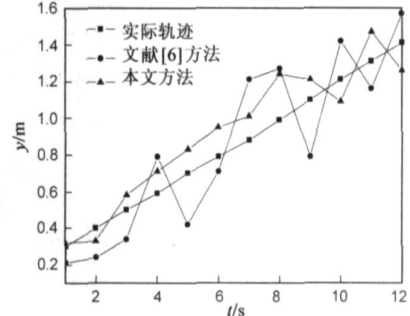


图 6  $\text{SNR} = 5\text{dB}$ ,  $T_{60} = 200\text{ms}$  环境下的测试结果 (粒子数  $N = 500$ )

仿真实验 2: 在不同的声学环境下, 比较两种方法的仿真实验结果。在本实验中, 声源的位置固定在  $x_s = [4.50, 3.40, 1.49]$ 。在粒子滤波算法中, 将说话人的运动模型建模为随机游走模型, 粒子数为 500, 从均值为  $x_s$ 、方差为 1 的均匀分布中生成采样粒子, 采用残差重采样方法降低退化现象的影响。先将混响时间设定在  $100\text{ms}$ , 信噪比以  $-5\text{dB}$  为步长从  $20\text{dB}$  变化到  $-5\text{dB}$ ; 再将信噪比设定在  $10\text{dB}$ , 混响时间以  $50\text{ms}$  为间隔从  $0\text{ms}$  到  $300\text{ms}$  依次取值。两种方法分别进行 50 次蒙特卡罗实验, 均方根误差 (Root Mean Square Error, RMSE) 的统计结果如表 1 和表 2 所示。

表 1 不同信噪比下的 RMSE 实验结果 ( $T_{60} = 100\text{ms}$ )

信噪比 (dB)	20	15	10	5	0	-5
文献[6]方法	0.21	0.31	0.36	0.39	0.41	0.47
本文方法	0.16	0.18	0.25	0.32	0.38	0.43

表 2 不同混响时间下的 RMSE 实验结果 ( $\text{SNR} = 10\text{dB}$ )

混响时间 (ms)	0	50	100	150	200	250	300
文献[6]方法	0.21	0.37	0.39	0.46	0.48	0.51	0.55
本文方法	0.18	0.25	0.27	0.32	0.28	0.30	0.29

从表 1 和表 2 可以看出, 当信噪比较大、混响时间较短时, 本文方法与文献[6]方法的均方根误差较低, 均可保持良好的跟踪性能。随着噪声与混响的增大, 两种方法的性能都有所下降, 与文献[6]方法相比, 本文方法在复杂声学环境下的抗噪声与抗混响能力有所增强。

两组仿真实验表明, 在房间混响较强的条件下, 混响产生的虚声源使说话人声源对应的峰值变得不明显。此时, 文献[6]方法虽然对混响具有一定的抑制作用, 但有时仍会出现较大的定位误差。本文方法采用语音的建立信号进行说话人跟踪, 可以有效抑制房间混响对系统定位性能的影响, 使得说话人跟踪系统在较

恶劣的声学环境中仍可保持一定的鲁棒性。

### 3.2 实测结果

本实验进一步比较本文方法与文献[6]方法在实际环境中的跟踪能力。跟踪系统的硬件设备包括:4个全指向麦克风、信号放大器、多通道A/D转换器与多通道声卡。房间大小为 $4.4\text{m} \times 6.8\text{m} \times 3.3\text{m}$ ,说话人的起始位置为 $X_0 = [3.00, -3.00, 1.49]$ ,以 $0.1\text{m/s}$ 的速度沿着 $y$ 轴正向匀速运动。文献[6]方法与本文方法采用的粒子数分别为250与100。在说话人运动模型中选取 $\beta_y = 0.5$ ,  $\nu = 0.4\text{m/s}$ ,每22帧(约0.5秒)给出一次跟踪结果。

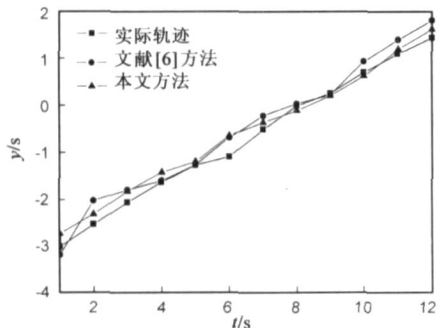


图7 两种方法的跟踪结果比较

两种方法的跟踪结果如图7所示。由图7中看出,文献[6]方法与本文方法均能大致估计出说话人运动轨迹,但文献[6]方法的跟踪误差有时偏大。本文将生成采样粒子语音的建立信号作为观测信息,并利用分层采样方法用较少的采样粒子取得了优于文献[6]的跟踪效果。这表明了本文方法在降低计算量的同时,提高了说话人声源定位与跟踪系统的抗噪声、抗混响能力。

### 4 结论

本文提出了一种基于麦克风阵列的说话人声源跟踪新方法,其特点在于采用分层采样粒子滤波与语音建立信号联合抑制房间混响对说话人跟踪系统的影响。仿真实验与实测结果表明,本文方法提高了粒子的采样效率,可为实际环境中的说话人跟踪问题提供一种有效的解决方案。考虑到说话人运动模型在跟踪中的重要作用,下一步拟采用交互多模型方法来描述说话人的复杂运动,以进一步改进说话人跟踪系统性能。

#### 参考文献:

- [1] R C Vemba, C Schmid, K Mikolajczyk. Face detection and tracking in a video by propagating detection probabilities[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(10): 1215-1228.
- [2] 杜江,朱柯. 智能麦克风阵列语音分离和说话人跟踪技术研究[J]. 电子学报, 2005, 33(2): 382-384.  
Du Jiang, Zhu Ke. Smart microphone arrays for speech sources

separation and speaker tracking [J]. Acta Electronica Sinica, 2005, 33(2): 382-384. (in Chinese)

- [3] J DiBiase. A high accuracy, low latency technique for talker localization in reverberant environments[D]. Ph D Dissertation, Brown University, Providence RI, USA, 2000.
- [4] J Huang, N Ohnishi, N Sugie. Sound localization in reverberant environment based on the model of the precedence effect[J]. IEEE Transactions on Instrumentation and Measurement, 1997, 46(4): 842-846.
- [5] J Vemaak, A Blake. Nonlinear filtering for speaker tracking in noisy and reverberant environments[A]. IEEE International Conference on Acoustic, Speech, Signal Processing [C]. Salt Lake City, USA, 2001. 3021-3024.
- [6] D B Ward, R C Williamson. Particle filter beamforming for acoustic source location in a reverberant environment[A]. IEEE International Conference on Acoustic, Speech, Signal Processing [C]. Orlando, USA, 2002. 1777-1780.
- [7] M Arulampalam, S Maskell, N Gordon, T Clapp. A tutorial on particle filters for on line nonlinear/non Gaussian Bayesian tracking[J]. IEEE Transactions on Signal Processing, 2002, 50(2): 174-188.
- [8] J Sullivan, A Blake, M Isard, J Maccormick. Bayesian object localisation in images[J]. International Journal of Computer Vision, 2001, 44(2): 111-135.
- [9] R Duraiswami, D Zotkin, L Davis. Active speech source localization by a dual coarse to fine search[A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Salt Lake City, USA, 2001. 3309-3312.
- [10] J B Allen, D A Berkley. Image method for efficiently simulating small room acoustics[J]. Journal of Acoustical Society of America, 1979, 65(4): 943-950.

#### 作者简介:



金乃高 男,1977年生于辽宁盖州,大连理工大学博士研究生,主要研究方向为语音信号处理和信息融合。E-mail: jin\_naigao@hotmail.com



殷福亮 男,1962年生于辽宁抚顺,大连理工大学教授、博士生导师,主要研究方向为语音信号处理、阵列信号处理与现代通信技术。E-mail: flyin@dlut.edu.cn