DOI:10.3969/j.issn.1673-4785.201201003
网络出版地址:http://www.cnki.net/kcms/detail/23.1538.TP.20120219.2224.001.html

机器人听觉声源定位研究综述

李晓飞[1],刘宏[1,2]
(1.北京大学深圳研究生院 集成微系统科学工程与应用重点实验室,广东 深圳 518055;2.北京大学 机器感知与智能教育部重点实验室,北京 100871)

摘 要:声源定位技术定位出外界声源相对于机器人的方向和位置,机器人听觉声源定位系统可以极大地提高机器人与外界交互的能力.总结和分析面向机器人听觉的声源定位技术对智能机器人技术的发展有着重要的意义.首先总结了面向机器人听觉的声源定位系统的特点,综述了机器人听觉声源定位的关键技术,包括到达时间差、可控波束形成、高分辨率谱估计、双耳听觉、主动听觉和视听融合技术.其次对麦克风阵列模型进行了分类,比较了基于三维麦克风阵列、二维麦克风阵列和双耳的7个典型系统的性能.最后总结了机器人听觉声源定位系统的应用,并分析了存在的问题和未来的发展趋势.

关键词:机器人;机器人听觉;声源定位;麦克风阵列
中图分类号:TP242.6;TN912.3 **文献标志码**:A **文章编号**:1673-4785(2012)01-0009-12

A survey of sound source localization for robot audition

LI Xiaofei[1], LIU Hong[1,2]
(1. Key Laboratory of Integrated Microsystems, Shenzhen Graduate School of Peking University, Shenzhen 518055, China; 2. Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China)

Abstract: The technology of sound source localization can localize the direction and position of a sound source relative to a robot. Sound source localization system for robot audition can greatly improve the ability of a robot to interact with external conditions. The summary and analysis of sound source localization for robot audition can significantly promote the development of intelligent robots. In this work, first, the characteristics of sound source localization for robot audition were summarized. The key technologies were summarized, including the time delay of arrival, steered beamforming, high resolution spectral estimation, binaural, active audition, and audio-visual. Then, the models of a microphone array were classified, and the performances of seven typical systems based on a 3-D microphone array, 2-D microphone array, and binaural were compared. Finally, the applications of a sound source localization system of robot audition were summarized. Several issues that sound source localization systems face as well as development trends were analyzed.

Keywords: robot; robot audition; sound source localization; microphone array

机器人听觉系统是一种自然、方便、有效、智能的机器人与外界系统交互的方式.由于声音信号的衍射性能,听觉具有全向性,相较于视觉、激光等其他的传感信号听觉不需要直线视野,在有视野遮蔽障碍物的情况下依然可以有效地工作.一般来讲机器人听觉包括声源信号的定位与分离、自动语音识别、说话人识别等.机器人听觉声源定位是指机器人利用搭载在机器人上或者外部设备上的麦克风阵列定位出声源的相对位置.随着信息技术、电子科学技术、计算机科学技术和智能科学的迅速发展,自20世纪90年代中期始,人们对机器人听觉声源定位技术进行了深入而广泛的研究,并取得了重要的进展.

声源的位置信息包括轴向角、仰角和距离,其中

收稿日期:2012-01-10. **网络出版时间**:2012-02-19.
基金项目:国家"863"计划资助项目(2006AA04Z247);国家自然科学基金资助项目(60675025,60875050);深圳市科技计划及基础研究计划资助项目(JC20090316039).
通信作者:刘宏. E-mail:hongliu@pku.edu.cn.

轴向角可以确定声源的二维方向,轴向角和仰角可以确定声源的三维方向,轴向角、仰角和距离可以确定声源的三维位置.在噪声环境下,利用少量的麦克风实时地定位声源的三维位置是一个实用的机器人听觉声源定位系统的目标.虽然机器人听觉声源定位技术的研究取得了很多成果,但是该技术的实际应用还面临很多问题.

1 机器人听觉系统声源定位的特点

相较于一般的声源定位系统,机器人听觉声源定位具有以下特点:

1)麦克风阵列易搭载:搭载在机器人平台上的麦克风阵列应该尽量小,麦克风阵列的小型化可以通过减少麦克风的数量和优化阵列拓扑来实现.

2)机器人运动:搭载在机器人平台上的麦克风阵列的运动改变了听觉场景,给声源定位带来了困难.但另一方面可以通过麦克风阵列的主动运动,丰富麦克风阵列的拓扑,提高定位能力.

3)声源移动:在大多数机器人听觉声源定位系统应用中,声源是移动的,需要进行移动声源的定位与跟踪.

4)实时性高:机器人的运动和声源的移动造成机器人和声源相对位置的即时变化,要求定位具有较高的实时性.机器人与外界交互的实时性是机器人友好性和安全性的保障,是评价交互性能的重要指标,因此声源定位系统的实时性是极其必要的.

5)抗混响和噪声:机器人工作在真实环境中,信号混响和噪声是难以避免的,因此声源定位系统的抗混响和抗噪声能力在很大程度上影响定位性能.

2 机器人听觉系统声源定位方法

1995年Irie第1次将声源定位技术用于智能机器人[1],利用短时时域、频域特征和神经网络技术区分摄像头视角内的左中右3个声源方向.其后,基于麦克风阵列的到达时间差技术(time delay of arrival, TDOA)、基于最大输出功率的可控波束形成技术(steered beamforming, BS)、高分辨率谱估计技术(high resolution spectral estimation)、双耳听觉(binaural)、机器学习(machine learning)、主动听觉技术(active audition)、视听融合(audio-visual)等方法被用于机器人听觉声源定位.

2.1 到达时间差技术

基于TDOA的定位技术是一种2步定位方法,首先估计出声源信号到达各个麦克风之间的时间延迟,然后利用几何定位方法求出声源位置.

稳健的时间延迟估计是精确声源定位的基础,常用的时延估计算法包括广义互相关(generalized cross correlation, GCC)[2]、互功率谱相位法(cross-power spectrum phase, CSP)[3]、特征值分解[4]、声学传递函数比[5]等.获取TDOA以后,乘以声速便可以得到距离差,这样就可以通过声源与麦克风的几何关系得到声源位置.主要的几何定位方法包括最大似然估计(maximum likelihood estimator)[6]和最小均方估计(least square estimator)[7-8].TDOA方法计算量小,可实时实现,但双步估计带来累积误差,是一种次最优估计,为了取得较高的分辨率,对信号采样率要求较高,适用于单声源定位.

1997年Huang等利用3个麦克风组成平面三角阵列定位声源的全向轴向角[9].根据声音的优先效应,通过无回响起点检测算法(echo-free onset detection)检测出无回响的声音段,利用过零点(zero-crossing point)检测时延,然后根据几何关系定位声源轴向角.2002年他们利用如图1所示的三维麦克风阵列进行声源轴向角和仰角的定位[10],互相关函数和互功率谱相位差分别被用于时延估计.识别阶段,6个时间差组成时间差序列:$\Delta \boldsymbol{t}_m = (\Delta t_{12}, \Delta t_{13}, \Delta t_{14}, \Delta t_{23}, \Delta t_{24}, \Delta t_{34})$,时间差序列误差为$e(\theta,\varphi) = \| \Delta \boldsymbol{t}(\theta,\varphi) - \Delta \boldsymbol{t}_m \|$,其中$\Delta \boldsymbol{t}(\theta,\varphi)$为理论时间差,轴向角$\theta$和仰角$\varphi$取使$e(\theta,\varphi)$最小化的值.2007年文献[11]对于多个声源,利用6个互相关函数的几何平均:

$$P(\theta,\varphi) = \{\prod C_{ij}(\Delta t_{ij}(\theta,\varphi))\}^{1/6}$$

表示一个声源位置存在声源的概率,概率越大则存在声源的可能性越大.

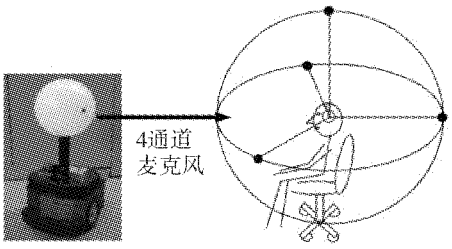


图1 4通道阵列

Fig.1 4-ch array

2002年Sekmen等提出一种自然的人机交互方式,把人作为一个被动的用户,不用通过键盘、鼠标等人工的方式与机器人进行交互[12].机器人只是人的运动的一个直接物理再现,利用声源定位和红外运动跟踪,为人脸跟踪系统提供候选区域和机器人的注意力.2个麦克风摆放在一个开放的空间,头部传输函数不用考虑.假设声源位于仿人机器人的前方,利用互相关法估计时延,通过远场近似几何方法便可定位远场声源.

2003年Valind等放置8个麦克风在长方体支架的顶点[13],如图2所示.该麦克风阵列搭载在Pi-

oneer 2 机器人上,用来进行声源轴向角和仰角定位. 利用谱加权 GCC-PHAT 方法提取时间差,给信噪比大的频带赋予更大的权值可以有效地抑制窄带噪声的影响. 然后利用远场几何定位方法定位声源的轴向角和仰角.

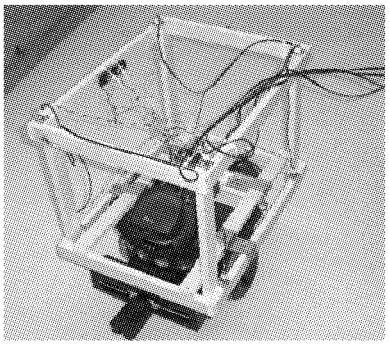


图 2　8 通道立方体阵列

Fig. 2　8-ch cube array

2007 年 Rudzyn 利用与图 1 相似的麦克风阵列定位声源三维位置[14],包括距离、轴向角和仰角. 利用加权互相关函数(weighted cross correlation, WCC)估计时延:$f_{wcc}=f_{gcc}/(f_{amdf}+\delta)$,其中 f_{amdf} 为平均幅度差函数(average magnitude difference function),用于增强 GCC 的性能. 同样使用近场几何定位方法来定位三维声源.

2008 年 Kwak 等利用平面正三角形阵列定位声源[15]. 语音信号的声门激励信息被用于时延估计,首先求出语音信号线性预测残差表示声门激励信号,然后线性预测残差的希尔伯特包络(Hilbert envelop)信号被用于基于 GCC-PHAT 的声源估计,再通过一种可靠的几何定位方法定位出声源轴向角. 该系统成本低廉、实时性好,可用于家庭服务机器人.

2009 年 Hu 等利用基于特征结构(eigen structure)的 GCC 方法估计多个声源的时延[16]. 多声源情况下麦克风接受信号的频域表示为

$$X_m = \sum_{d=1}^{D} a_{md} S e^{-j\omega\tau_{md}} + N_m.$$

式中:D 为声源个数. 接收信号互相关矩阵的特征分解为

$$R_{xx}(\omega) = \left(\sum_{k=1}^{K} \boldsymbol{X}(\omega,k)\boldsymbol{X}^{\mathrm{T}}(\omega,k)\right)/K = \sum_{i=1}^{M} \lambda_i(\omega)\boldsymbol{V}_i(\omega)\boldsymbol{V}_i^{\mathrm{T}}(\omega).$$

式中:λ 为特征值,$\boldsymbol{V}$ 为特征向量. 与前 $\tilde{D}$ 个最大特征值对应的向量表示声源向量,利用声源向量的 GCC 方法进行时延估计. 文献[17]利用声速的限制求出声源个数 $\tilde{D}$,定位阶段,利用最小均方估计求解超定线性方程组定位多个声源,近场情况下求解声源三维直角系坐标,远场情况下求解声源轴向角. 图 3 为该系统搭载在移动机器人平台上的 8 通道麦克风阵列.

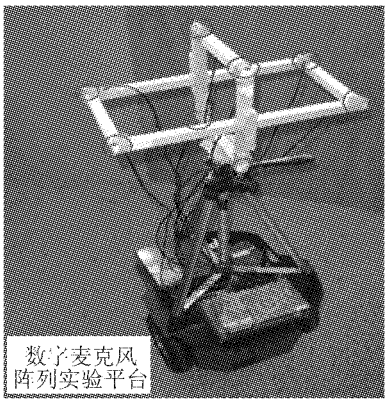


图 3　8 通道麦克风阵列

Fig. 3　8-ch microphone array

2010 年 Lee 等利用远场情况下声源角度和到达时延的几何关系建立了 angle-TDOA 图[18]. 融合该图和互相关函数得到 Cross-Angle-Correlation 函数 $R(\theta)$,该函数在声源方向取较大的值. 对于多个声源,竞争 K-means 算法被用于基于 Cross-Angle-Correlation 函数的声源角度聚类,该系统利用正三角形麦克风阵列定位声源轴向角.

2.2　基于最大输出功率的可控波束形成技术

该方法对麦克风接受到的声源信号滤波并加权求和形成波束,按照某种搜索策略全局搜索可能的声源位置来引导波束,波束输出功率最大的位置即为声源位置[19-20]. 延迟和波束形成算法(delay-and-sum beamforming, DSB)[21]通过对麦克风接受信号采用时间移位以补偿声源到达各麦克风的传播延迟,并通过相加求平均来形成波束. 滤波和波束形成算法(filter-and-sum beamforming, FSB)[22]在时间移位的同时进行滤波,然后相加求平均形成波束.

可控波束形成算法的定位性能取决于麦克风阵列方向图的主瓣和旁瓣的分布. 主瓣能量越大,宽度越窄,则形成波束的分辨率越高. 通常该算法要求大量的麦克风以取得较好的方向图. 该算法本质上是一种最大似然估计,需要声源和噪声的先验信息,但通常这些信息在实际应用中不易获得. 最大似然估计是一种非线性优化问题,传统搜索算法容易陷入局部最小点,而遍历式的搜索方法的运算量极大[23].

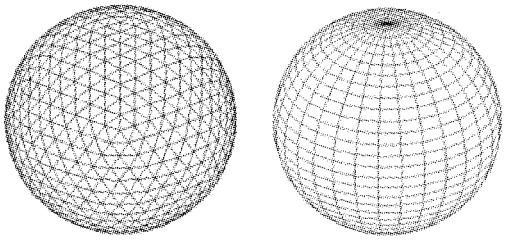
1999 年 Matsui 等研制出一种办公室接待机器人 Jijo-2,它可在办公室环境下引导客人参观[24]. 该机器人视觉声源定位系统基于波束形成算法,利用平均分布于半圆弧的平面 8 通道麦克风阵列定位声源的轴向角.

2004 年 Valin 等利用 DSB 定位多声源位置,预求出所有对的麦克风信号频域的互相关[25]:

$$R_{ij}(\tau) = \sum_{k=0}^{L-1} X_i(k) X_j^*(k) e^{j2\pi k\tau/L}.$$

每个波束输出功率可以通过 $N(N-1)/2$ 个互相关

累积和求得. 谱加权在互相关求解中给信噪比大的频带赋予更大的权值,有效地抑制了窄带噪声的影响. 另外为了避免声源的错误检测,一个基于声源存在概率的后处理算法被提出. 2009 年 Badali 和 Valin 等利用如图 2 所示的麦克风阵列测试了可控响应功率(steered response power)和其他算法的性能[26]. 运用谱加权用于抑制噪声,方向优化算法是在 DSB 算法定位的结果临近范围内应用高分辨率方法,从而更精确地定位声源. 上述 2 篇文献的麦克风阵列如图 2 所示,图 4 显示了 2 种球形搜索网格. 文献[26]的实验结果显示三角网格声源搜索策略相较于矩形网格更有效,三角网格共2 562个搜索点,每个搜索区域覆盖 2.5°.

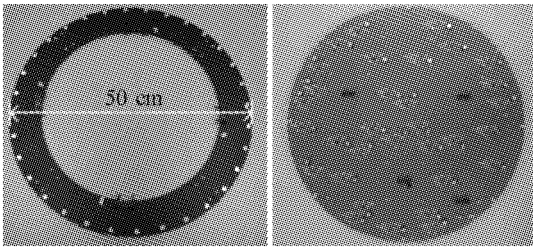


(a)三角网格 (b)矩形网格

图 4 球形搜索网格

Fig. 4 Spherical search grids

2004 年 Tamai 等利用搭载在 Nomad 机器人上的平面圆形 32 通道麦克风阵列定位 1 ~ 4 个声源的水平方向和垂直方向[27]. 由于麦克风数量较多,DSB 算法可以很好地抑制环境噪声和机器人机体噪声. 文献[28]提出了一种 3 个圆形阵列组成的 32 通道阵列,相较于一个圆形阵列具有更好的波束方向图分布. 以上 2 种阵列如图 5 所示.



(a)单个圆形阵列 (b)3个圆形阵列

图 5 32 通道二维麦克风阵列

Fig. 5 32-ch 2-D microphone array

2005 年 Nakadai 等利用 64 通道分布式麦克风阵列在电视等噪声环境中检测真实语音信号[29],并定位声源的平面二维位置. 图 6 为麦克风阵列,麦克风分布在 1.2 m 高度的墙壁和高度为 0.7 m 的桌面上. 加权 DBS 用于求解每个可能方向的方向性模式(directivity pattern),方向性模式用于检测麦克风接收信号是否为真实的语音信号,并定位声源. 2006 年他们在文献[30]中基于 MUSIC 方法利用搭载在 ASIMO 机器人头部的 8 通道麦克风阵列定位多声源,并利用粒子滤波(particle filter)方法融合房间麦克风阵列和机器人麦克风阵列的定位结果,跟踪多个声源.

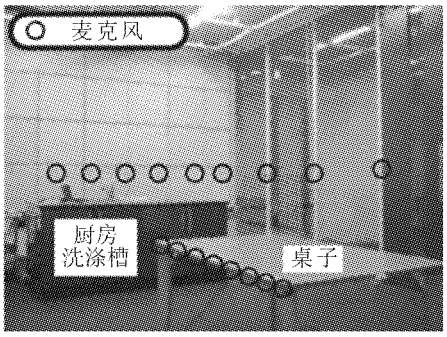
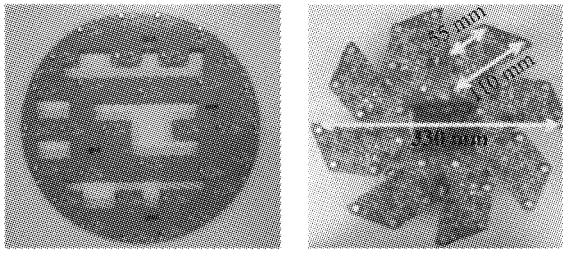


图 6 分布式麦克风阵列

Fig. 6 Distributed microphone array

2006 年 Sasaki 等利用 32 通道 3 同心圆阵列通过机器人的运动定位多声源的二维位置[31]. 首先利用基本 DSB 算法减弱噪声,然后通过频带选择算法(frequency band selection)消除剩余噪声并定位出多声源的水平方向,最后根据运动的机器人可以在不同的位置检测同一个声源的方向,通过三角定位方法和 RANSAC 算法(random sample consensus)定位出声源的精确位置. 2007 年他们通过主瓣消除算法(main-lobe canceling)从 DSB 算法得出的空间谱中逐个检测声源的位置[32]. 每次检测出当前具有最大能量的方向作为当前声源的方向,然后减除该方向的主瓣继续检测下一个声源. 主瓣消除算法需要阵列方向图具有较小的旁瓣. 图 7 显示了同心圆阵列和八边形 32 通道麦克风阵列,八边形阵列在700 ~ 2 500 Hz的频率范围内旁瓣能量较小. Kagami 等利用文献[32]中的声源方向定位和粒子滤波方法,通过机器人的运动定位静止声源的精确位置[33]. 2010 年 Sasaki 等综合上述的声源定位功能,并进行短时声音信号识别以标定声源[34],通过跟踪多个声源,画出声源图并定位机器人的位置.



(a)同心圆阵列 (b)八边形阵列

图 7 同心圆阵列和八边形阵列

Fig. 7 Concentric array and octagonal array

2.3 高分辨率谱估计技术

该方法来源于现代高分辨率谱估计技术,如自回归模型(autoregressive)[35]、多重信号分类(multiple signal classification, MUSIC)[36]等方法,利用特

征值分解(eigenvalue decomposition)将麦克风信号的协方差矩阵分解为信号子空间和噪声子空间,然后找出与噪声子空间正交的方向矢量来进行声源的方向估计.

基于高分辨率谱估计的定位方法是一种超分辨率的估计技术,其空间分辨率不会受到信号采样频率的限制,并且在一定条件下可以达到任意定位精度[37].然而,该类方法也存在一定的不足,主要表现在:1)易受空间相关噪声的干扰,当方向性噪声的能量与声源信号能量相当时,该算法容易定位到噪声方向;2)房间的反射作用使信号和干扰之间有一定的相关性,从而降低了该方法的有效性;3)需要对整个空间进行搜索来确定声源的位置,且其估计精度与空间的细分程度有关,计算复杂度偏高.

1999 年 Asano 等利用搭载在办公室机器人 Jijo-2 上的平均分布于半圆弧的平面 8 通道麦克风阵列定位多个声源的轴向角[38].扩展的 MUSIC 算法被用于近场定位,近场方向向量为

$$\boldsymbol{a}(r,\theta) = [\mathrm{e}^{-\mathrm{j}\omega\tau_1(r,\theta)} \quad \mathrm{e}^{-\mathrm{j}\omega\tau_2(r,\theta)} \quad \cdots \quad \mathrm{e}^{-\mathrm{j}\omega\tau_M(r,\theta)}]^{\mathrm{T}}.$$

式中:r 和 θ 分别为声源的水平距离和轴向角.

大多数机器人听觉声源定位系统接收的声源信号是宽带信号,原始的 MUSIC 算法只能定位窄带信号.2007 年 Argentieri 等给出 MUSIC 算法的宽带声源扩展[39],近场 MUSIC 空间谱为

$$h(r,\theta) = 1/\boldsymbol{V}^{\mathrm{T}}(r,\theta)\hat{\prod}_N \boldsymbol{V}(v,\theta).$$

式中:$\boldsymbol{V}$ 为可能声源位置的方向向量,$\hat{\prod}$ 为噪声子空间.令空间谱最大的方向向量对应于声源位置,一种朴素的宽带扩展方法为

$$h_{\mathrm{nalve}}(r,\theta) = \sum_{b=1}^{B} h_b(r,\theta)/B.$$

式中:b 为信号频点数,B 为频带宽度.实验证明该宽带扩展方法性能很好,但计算量太大.波束空间算法利用频率和范围不变的波束形成聚焦频点,生成一个对所有兴趣频点有效的空间谱.

2009 年 Nakamura 等利用广义特征值分解抑制空间相关噪声的影响[40],在静音段估计出噪声的空间互相关矩阵,对带噪声源信号的互相关矩阵和噪声的互相关矩阵进行广义特征值分解,生成一个完全抑制噪声的空间谱.2011 年他们联合视觉跟踪算法,利用粒子滤波进行说话人的跟踪[41].

2009 年 Ishi 测试了 MUSIC 方法在办公室环境和室外环境下定位轴向角的性能[42].办公室环境存在空调噪声和机器人机体噪声,室外环境存在背景音乐噪声.他们分别测试了信号分帧长度对方向估计性能和实时性的影响,宽带 MUSIC 频带宽度和声源个数对方向估计的影响.另外还提出了一种确定声源个数的方法,对每个频率采用固定声源数,并设置宽带 MUSIC 的声源个数上限,实验证明这种方法与已知声源个数情况下的定位性能差不多.图 8 显示了该机器人平台和 14 通道的稀疏麦克风阵列.

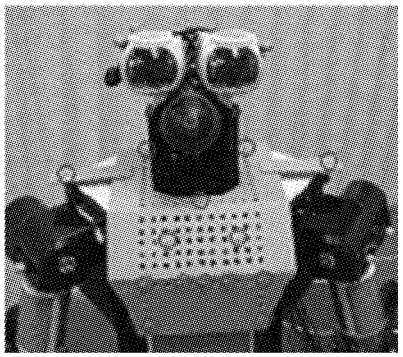


图 8　稀疏麦克风阵列

Fig. 8　Sparse microphone array

2.4　双耳听觉

人可以通过双耳定位 3-D 空间声源方向,双耳时间差(interaural time difference, ITD)和双耳强度差(interaural intensity difference, IID)用于定位声源轴向角,由耳廓衍射和散射效应带来的声谱特性(spectral cue)用于定位声源仰角[43].声音信号从声源位置传播至人耳鼓膜处的传输函数被称为头部相关传递函数(head-related transfer functions, HRTFs)[44],影响 HRTFs 的因素有耳廓、头部、耳道、肩膀和躯体等.基于双耳的声源定位方法对于仿人机器人是一种自然、有效的方式,利用人工头和人工耳廓可以有效地模仿人的听觉定位能力[45].

Nakadai 等基于仿人机器人 SIG 的双耳听觉定位声源轴向角[46-48].由立体视觉扩展的听觉 Epipolar 几何可以数学化地估计出特定声源方向的 IPD:$\Delta\varphi_e = 2\pi fr(\theta + \sin\theta)/v$,其中 f、r、θ 和 v 分别为信号频率、头部半径、声源角度和声速,一般 f 小于 1 500 Hz.可能声源方向和实测信号的 IPD 之差最小的为声源方向.Epipolar 几何很难确定出精确的 IID,只能通过频率大于1 500 Hz的频带确定出声源的大概方向.利用物理学中的散射理论(scattering theory)也可以数学化地估计 IPD $\Delta\varphi_s(\theta,f)$ 和 IID $\Delta\rho_s(\theta,f)$,同样分别采用小于和大于1 500 Hz的频带,相较于 Epipolar 几何散射理论的 IPD 估计误差更小,并且可以较精确地估计出 IID.利用 Dempster-Shafe 理论联合 IPD 和 IID 信息,联合概率取最大的可能位置为声源位置.

2005 年 Kumon 等根据声波在耳廓中反射决定的声谱特性(spectral cue)设计了一个人工耳廓[49].耳廓形状如图 9 所示,耳廓必须关于声源仰角非对称以保证可以区分不同仰角的声源信号.该耳廓对于仰角大于 90°的声源具有较明显的谱峰(spectral peak).2006 年 Shimoda 等改进了文献[49]中设计

的人工耳廓的仰角定位算法[50]. 由于机器人头部运动是连续的,所以声谱特性变化也是连续的,即相邻时刻的声谱特性不会产生突变. 根据此特性对长时间检测的声谱特性进行聚类,得到更精确的声谱特性,一定程度上抑制了噪声的干扰.

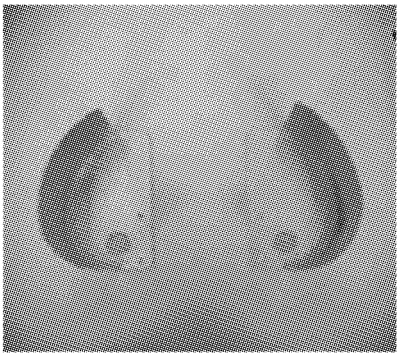


图9 人工耳廓

Fig. 9 Artificial pinnae

2006年Hornsteind等利用人工耳廓和人工头模拟人的听觉定位[51]. 人工头模型如图10所示,通过ITD、IID和谱谷(spectral notches)定位声源的轴向角和仰角以控制头部转向声源.

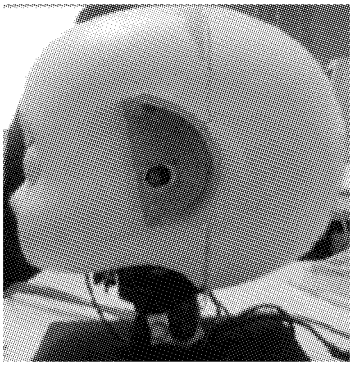


图10 人工头部和耳廓

Fig. 10 Artificial head and pinnae

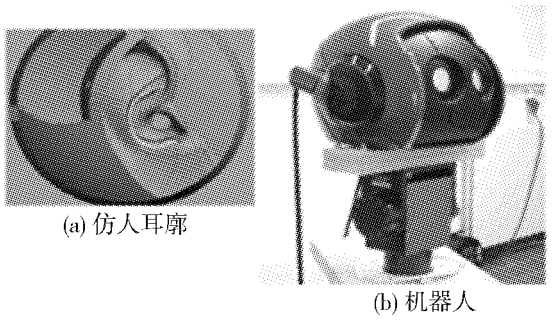
2006年Keyrouz等利用人工头和人工双耳同时分离和定位2个声源的轴向角和仰角[52],一种时域的盲源分离算法被用于分离2个独立且相距不太近的声源. 令第1个声源到第2个分离信号的冲激响应为c_{12},第2个声源到第1个分离信号的冲激响应为c_{21},则声源到麦克风的冲激响应h需满足:

$$c_{12} = h_{11}w_{12} + h_{12}w_{22} = 0, \quad (1)$$

$$c_{21} = h_{21}w_{11} + h_{22}w_{21} = 0. \quad (2)$$

式中:w为解混冲激响应. 通过式(1)、(2)可以分别求出2个声源方向的HRTFs,进一步可以定位声源的全向轴向角和仰角.

2008年Rodemann等利用仿人耳蜗和双麦克风进行声源的3-D方向定位[53],耳蜗和机器人如图11所示. 在提取ITD、IID和spectral cue前先进行双耳信号的同步谱减去噪. 为了消除声源信号特性对声谱特性的影响,用左右耳对数谱之差表示声谱特性:$\bar{S}(k) = \lg(\bar{s}_r(k)) - \lg(\bar{s}_l(k))$. 2010年他们在文献[54]中联合声音幅度、谱幅度、ITD和IID定位声源的距离.

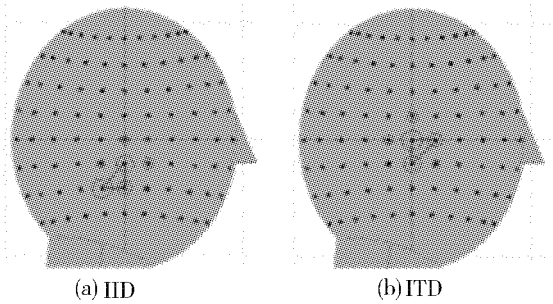


(a) 仿人耳廓 (b) 机器人

图11 仿人耳廓和机器人

Fig. 11 Humanoid pinnae and robot

2011年Kim等为了降低基于信号相关的时延估计算法的信号采样率对定位分辨率的影响,利用最大似然方法找出最大化互功率谱之和的声源轴向角,分辨率达到1°[55]. 另外考虑机器人球形头部带来的多径效应,一个基于front-back的多径补偿因子被用来修正时延估计. 2011年Skaf等[56]测试了放置在一个椭球人工头上的88对对称双耳的定位性能,IID和ITD被分别测试,实验结果显示,综合IID和ITD性能时双耳放置在人工头的后下方性能最优. 人工头及双耳位置如图12所示.



(a) IID (b) ITD

图12 最优双耳位置

Fig. 12 Optimal position of two ears

2.5 机器学习

Saxena等利用单麦克风和人工耳廓基于机器学习方法定位声源方向[57]. 不同声源方向到麦克风的传输函数不同,用隐马尔可夫模型表示时变的麦克风信号Y_t,则声源方向可以通过式(3)估计,式(3)可以通过前向-后向算法求解,以15°的步长遍历轴向角求解$\hat{\theta}$.

$$\hat{\theta} = \arg\max_{\theta} P(Y_1^2, Y_2^2, \cdots, Y_T^2 \mid \theta). \quad (3)$$

2.6 主动听觉

文献[58]指出机器人的感知能力应该是主动的,可以通过机器人的移动和传感器参数的控制获得更好的感知环境. 该文基于SIG人形机器人的头

部转动建立了主动听觉系统,通过头部的转动可以调节双耳麦克风垂直于声源方向以取得更好的定位性能. 机器人头部和摄像机的马达转动、齿轮、传送带和滚珠会带来内部噪声,由于离麦克风较近,所以会极大影响声源定位性能,因此自适应滤波器被用于抑制内部噪声.

文献[59]提出感知-马达(sensory-motor)融合的概念:感知信息指导马达的运动和导航,通过机器人的运动消除双耳声源定位算法的前后向混淆.

2011 年 Martinson 等用3 台 Pioneer3-AT 机器人分别搭载 2、1 和 1 个麦克风组成动态可重置的麦克风阵列[60],如图 13 所示. 对于给定的兴趣区域,吸引/排斥模型可以动态优化各麦克风位置以获得更好的声源定位性能.

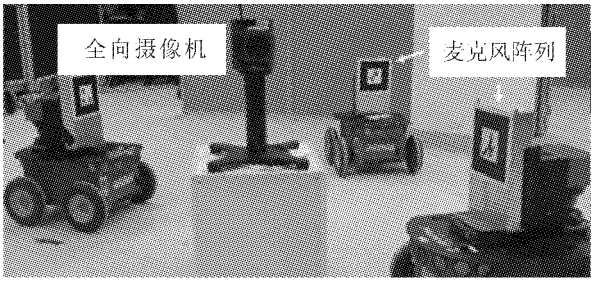
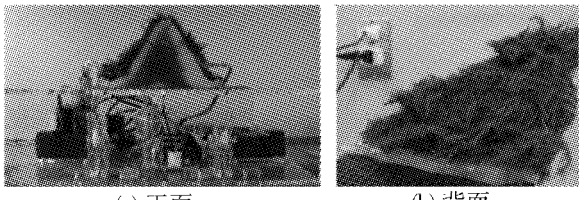


图 13　动态麦克风阵列

Fig. 13　Dynamic microphone array

Portello 等建立了一个动态双耳听觉模型[61],麦克风和声源相对运动的动态 ITD 模型给无味卡尔曼滤波器提供了一个 ground credible 等式,以确定声源的距离和轴向角的定位,该算法不适用于声源和传感器之间高速相对运动的情况.

Kumon 提出一种主动软耳廓[62],软耳廓由具有弹性的硅橡胶制成,背面覆盖一层皮毛,以保证耳廓的单向性. 耳廓可以旋转和变形以提供主动听觉声源定位能力,软耳廓如图 14 所示.



(a) 正面　　(b) 背面

图 14　软耳廓

Fig. 14　Soft pinnae

2.7　视听融合技术

基于视觉的物体检测与跟踪在光照条件好、视野无遮挡的情况下具有更好的性能. 融合听觉信息与视觉信息可以极大提高各传感器单独的感知能力. Okuno 和 Nakadai 等融合听觉事件形成的听觉流与视觉事件形成的视觉流生成联合流,以控制 SIG 机器人注意力的转移[47,63-64],其中听觉事件为声源方向估计,视觉事件为多人人脸检测. Lv 等利用视觉物体检测修正听觉声源定位结果[65]. Lee 等利用视觉信息在多个人中区分出真正的说话者[66].

3　机器人听觉系统分析

3.1　麦克风阵列类型

声源定位系统的麦克风数量和拓扑主要取决于声源定位方法,一般情况下 TDOA 方法、高分辨率方法和波束形成方法需要的麦克风数量依次增多. 麦克风阵列类型如表 1 所示.

表 1　麦克风阵列类型

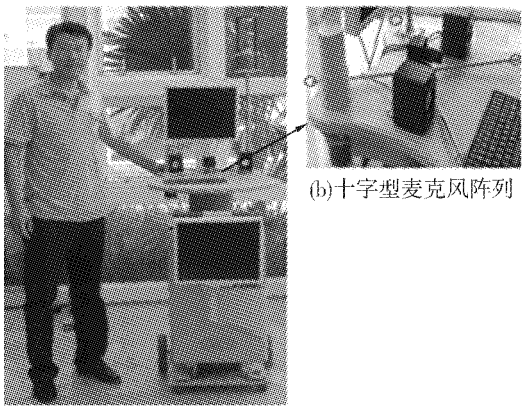
Table 1　Types of microphone array

麦克风阵列类型	阵列举例
三维阵列	图 1 ~ 3
二维阵列	图 5、7
稀疏阵列	图 8
分布式阵列	图 6
动态阵列	图 13
双麦克风	图 9 ~ 11
单麦克风	文献[57]

二维和三维阵列一般为规则拓扑麦克风阵列,如线性、三角形、多边形、多面体阵列等,分别具有二维平面和三维空间声源定位能力. 面向机器人听觉的声源定位的麦克风阵列应该易搭载在机器人平台上,通常要求阵列的小型化,包括麦克风数量的减少和阵列尺寸的减小. 实时性是人机交互的重要特点,因此实时的机器人听觉系统声源定位要求选取计算复杂度低的定位方法,一般来讲双耳定位和基于到达时间差的定位具有较小的计算复杂度,其次是基于高分辨率定位方法,基于波束形成方法的定位复杂度较高. 双麦克风模拟人耳听觉,通常需要借助人工头和耳廓的辅助,并且精确的头部相关传递函数较难获取.

3.2　机器人听觉声源定位系统

笔者利用搭载在移动机器人平台上的二维平面 4 通道十字型麦克风阵列定位说话人的轴向角和距离,以进行友好、有效的人机交互. 文献[67]提出指导性谱时定位方法(guided ST position algorithm),通过粗定位结果估计的声场条件进行二次精确定位,可以有效地消除混响的影响. 文献[68]提出一种基于时间差特征的空间栅格匹配(spatial grid matching)算法,找到与待定位声源的时间差特征最匹配的栅格作为声源位置. 该方法可以有效地避免几何定位方法的非线性方程组求解问题,复杂度较低,并且合理的麦克风阵列拓扑可以避免几何定位方法可能陷入局部最优点的问题. 移动机器人和麦克风阵列如图 15 所示.



(b)十字型麦克风阵列

(a)移动机器人

图15 移动机器人和麦克风阵列

Fig. 15 Mobile robot and microphone array

首先,利用谱加权GCC-PHAT方法求出各个麦克风对之间的信号时间差,6个时间差组成时间差特征序列:$\tau=(\tau_{12},\tau_{13},\tau_{14},\tau_{23},\tau_{24},\tau_{34})$. 可以证明,时间差特征与声源位置是一一对应的,即一个特定的时间差对应一个特定位置,反之亦然;另外2个声源位置之间的时间差特征的差与声源的位置之差成正比,即2个声源距离越远,另外2个位置的时间差特征的差越大. 根据这2个特点,可以把二维平面按照某种方式分割成栅格,每个栅格内的声源看作同一类声源,平面栅格如图16所示.

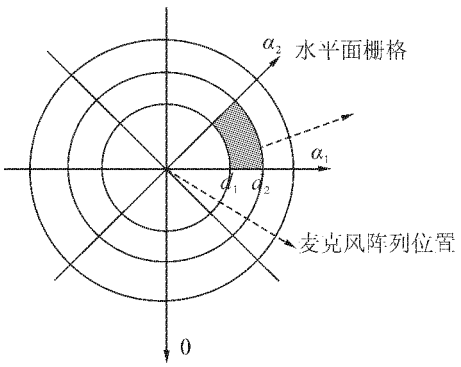


图16 平面栅格

Fig. 16 Horizontal grid

然后基于时间差特征,利用蒙特卡洛方法为每个栅格训练一个混合高斯模型,该模型表示平均分布于栅格内的时间差特征. 定位阶段,声源定位的问题可以表示为

$$G_s \propto \arg\max_G P(\tau \mid G).$$

式中:G表示栅格,G_s表示声源栅格. 计算出未知声源的时间差特征相对于所有栅格的似然值,似然值最大的栅格被定位为声源栅格. 另外有效特征检测算法利用信号时间差之间的约束移除错误的时间差,提高了定位性能. 并且决策树提供了一种由粗到细的定位方式,极大减少了未知声源的时间差与栅格的匹配次数.

实验测试了4 m以内的2 016组数据,轴向角栅格精度为1°,距离分为0~1.5 m、1~2 m和1.5~4 m 3个栅格. 轴向角测量误差小于5°的定位率超过95%,距离定位率超过90%,可以有效定位说话人的方位和说话人是否处于人机交互的安全距离. 而且听觉声源定位结果控制机器人转向说话人,使说话人在摄像头的视野范围之内,基于视觉的人体检测技术被用于更精确的目标人定位,以进行进一步的人机交互.

3.3 机器人听觉声源定位系统分析

一个机器人听觉声源定位系统可以从麦克风阵列拓扑、麦克风数量、声源定位能力、声源个数、抗噪声和混响能力、定位性能等方面来评价,其中定位能力指是否能进行声源轴向角、仰角和距离的定位. 表2列出了基于三维麦克风阵列、二维麦克风阵列和双耳的7个典型声源定位系统,其中声源个数只是列出了相关文献中实验测试的声源个数,不能完全反映该声源定位系统的能力. 因为机器人听觉声源定位算法发展的时间较短,并没有公共的测试实验数据库或实验平台;所以不同系统的实验场景和性能测量标准不同,本文只列出了相关文献中公布的定位性能.

表2 典型机器人听觉声源定位系统的分析

Table 2 Analysis of typical sound source localization system for robot audition

作者及文献	麦克风阵列	定位方法	声源数	定位能力	抗噪声和混响	定位性能
J. M. Valin[13]	8通道、三维阵列	TDOA	1	轴向角	抗噪声	精度:3 m之外、声源3°
H. Li[11]	4通道、三维阵列	TDOA	1~2	轴向角、仰角	抗混响	误差:单声源小于5°
Y. TAMAI[28]	32通道、二维阵列	DSB	1~2	轴向角、仰角、距离	抗噪声	误差:轴向角小于5°、仰角小于6°、距离小于300 mm
Y. Sasaki[34]	32通道、二维阵列	DSB	多个	轴向角、距离	—	误差:平均282 mm
K. Nakamura[40]	8通道	GEVD	1~2	轴向角	抗噪声	定位率:100%
K. Nakadai[48]	双耳、头部	Binaural	1	180°轴向角	—	精度:10°
F. Keyrouz[52]	双耳、头部、耳廓	Binaural	2	轴向角、仰角	—	精度:轴向角5°、仰角10°

4　总结与展望

机器人听觉声源定位系统的应用场景主要有家庭环境、公共场所、危险环境和一些其他特定场景中,面向的声源有人的语音和其他各种声源,主要包括以下几类应用:

1)服务机器人:声源定位系统提供了一种自然、高效的人机交互方式,主要应用在家庭、商场等环境. 服务机器人定位的声源通常为人的语音,并且面临复杂的噪声.

2)接待机器人:在办公室或家庭等场所接待客人,引导客人的行动,一般具有一定的语音识别能力,如文献[24].

3)军用机器人:战场声源的定位,如文献[69]在城市环境基于军用无人车定位枪声、尖叫声.

4)救援机器人:危险环境中救援任务的声源定位,如文献[70]. 在危险环境中,由于对人来说工作环境较为恶劣,因此机器人可以发挥较大的作用,比如救援、事故检测等.

5)助残机器人:引导残疾人,特别是盲人的活动,如文献[71]. 与机器人的语音交互和机器人的引导可以极大地提高盲人的活动能力.

自 1995 年,经过十几年的研究与探索,面向机器人听觉的声源定位技术取得了一定的成果,但系统的实用化还面临着一些问题,这些问题引导了未来的发展趋势:

1)机器人的运动. 机器人运动带来的麦克风阵列的运动是机器人听觉与传统声源定位技术主要的差别所在,运动的麦克风阵列会面临即时变化的声学环境,要求声源定位系统具有较高的实时性. 现在大多数声源定位系统的传感器数量较多,导致算法计算复杂度较高. 少量的麦克风和低复杂度的定位算法有待进一步探索.

2)复杂的声学环境. 几乎所有的实用声源定位系统必然面临着复杂的声学环境,存在各种类型的噪声. 现有的抗噪声技术大多只是针对某类或某几类噪声有效,一种鲁棒的、对各种噪声广泛适用的抗噪声技术或方案也还有待进一步研究.

3)阵列的小型化. 机器人搭载平台要求麦克风的数量尽量少,阵列尺寸尽量小,并且通常麦克风数量的减少会有效降低运算量. 现有的麦克风阵列大多需要专门的搭载平台,甚至需要辅助设备,实用化比较差. 双耳声源定位的发展提供了更接近于人的定位方式和能力,但特制的人工头和耳廓,以及它们的数学模型的建立都带来了诸多不便.

4)友好、智能的交互方式. 人机交互中人应该是被动的,即不用通过某种不方便的主动方式与机器人交互. 这就要求机器人可以主动、透明地与人交互,因此,智能声源定位技术的应用还与其他相关技术息息相关,并且一定程度上受到它们的制约,比如声音的检测与识别等.

本文主要依据定位算法综述了机器人听觉声源定位技术,不同于传统的声源定位技术,智能机器人带来了一些新的问题,比如机器人平台对麦克风阵列结构的限制、机器人运动给声源定位带来的诸多问题、人机交互对实时性的要求、机器人特定的工作场景等. 依据机器人技术的声源定位系统仍然有待进一步地总结与分析. 总之,实时、精确的机器人系统与外界系统的交互是机器人听觉声源定位技术追求的目标. 声源定位技术与机器人技术的融合带来了很多新的挑战,但更重要的是两者会互相促进对方的发展.

参考文献:

[1] IRIE R E. Robust sound localization: an application of an auditory perception system for a humanoid robot[D]. Cambridge, USA: Department of Electrical Engineering and Computer Science, MIT, 1995.

[2] 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.

[3] OMOLOGO M, SVAIZER P. Acoustic source location in noisy and reverberant environment using CSP analysis[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. Atlanta, USA, 1996: 921-924.

[4] BENESTY J. Adaptive eigenvalue decomposition algorithm for passive acoustic source localization[J]. Journal of Acoustical Society of America, 2000, 107(1): 384-391.

[5] DVORKIND T G, GANNOT S. Time difference of arrival estimation of speech source in a noisy and reverberant environment[J]. IEEE Transactions on Signal Processing, 2005, 85(1): 177-204.

[6] HAHN W, TRETTER S. Optimum processing for delay-vector estimation in passive signal arrays[J]. IEEE Transactions on Information Theory, 1973, 19(5): 608-614.

[7] WANG H, CHU P. Voice source localization for automatic camera pointing system in videoconferencing[C]//IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, USA, 1997: 187-190.

[8] SCHAU H, ROBINSON A. Passive source localization employing intersection spherical surfaces from time-of-arrival difference[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1987, 35(8): 1223-1225.

[9]HUANG Jie, SUPAONGPRAPA T, TERAKURA I, et al. Mobile robot and sound localization[C]//IEEE/RSJ International Conference on Intelligent Robots and System. Grenoble, France, 1997: 683-689.

[10]HUANG Jie, KUME K, SAJI A, et al. Robotic spatial sound localization and its 3-D sound human interface [C]//First International Symposium on Cyber Worlds (CW 2002). Tokyo, Japan, 2002: 191-197.

[11]LI H K, YOSIARA T, ZHAO Q F. A spatial sound localization system for mobile robots[C]//IEEE Instrumentation and Measurement Technology Conference. Warsaw, Poland, 2007: 1-6.

[12]SEKMEN A S, WIKES M, KAWAMURA K. An application of passive human-robot interaction: human tracking based on attention distraction[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 2002, 32(2): 248-259.

[13]VALIN J M, MICHAUD F, ROUAT J, et al. Robust sound source localization using a microphone array on a mobile robot[C]//IEEE/RSJ International Conference on Intelligent Robots and System. Las Vegas, USA, 2003: 1228-1233.

[14]RUDZYN B, KADOUS W, SAMMUT C. Real time robot audition system incorporating both 3D sound source localisation and voice characterization[C]//IEEE International Conference on Robotics and Automation. Roma, Italy, 2007: 4733-4738.

[15]KWAK K C, KIM S S. Sound source localization with the aid of excitation source information in home robot environments[J]. IEEE Transactions on Consumer Electronics, 2008, 54(2): 852-856.

[16]HU J S, CHAN C Y, WANG C K, et al. Simultaneous localization of mobile robot and multiple sound sources using microphone array[C]//IEEE International Conference on Robotics and Automation. Kobe, Japan, 2009: 29-34.

[17]HU J S, YANG C H, WANG C K. Estimation of sound source number and directions under a multi-source environment[C]//IEEE/RSJ International Conference on Intelligent Robots and System. Louis, USA, 2009: 181-186.

[18]LEE B, CHOI J S. Multi-source sound localization using the competitive K-means clustering[C]//IEEE Conference on Emerging Technologies and Factory Automation. Bilbao, Spain, 2010: 1-7.

[19]HAHN W R. Optimum signal processing for passive sonar range and bearing estimation[J]. Journal of Acoustical Society of America, 1975, 58(1): 201-207.

[20]CARTER G. Variance bounds for passively locating an acoustic source with a symmetric line array[J]. Journal of Acoustical Society of America, 1977, 62(4): 922-926.

[21]RAMOS L L, HOLM S, GUDYANGEN S, et al. Delay-and-sum beamforming for direction of arrival estimation applied to gunshot acoustics[C]//Proceedings of SPIE Defense, Security, and Sensing. Orlando, USA, 2011.

[22]KAJALA M. Filter-and-sum beamformer with adjustable filter characteristics[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City, USA, 2001: 2917-2920.

[23]曹玮玮.基于麦克风阵列的声源定位与语音增强方法研究[D].北京:清华大学, 2008.
CAO Weiwei. Study on methods of microphone array based sound source localization and speech enhancement[D]. Beijing: Tsinghua University, 2008.

[24]MATSUI T, ASOH H, FRY J, et al. Integrated natural spoken dialogue system of Jijo-2 mobile robot for office services[C]//Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Conference on Innovative Applications of Artificial Intelligence. Menlo Park, USA, 1999: 621-627.

[25]VALIN J M, MICHAUD F, HADJOU B. Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach [C]// IEEE International Conference on Robotics and Automation. New Orleans, USA, 2004: 1033-1038.

[26]BADALI A, VALIN J M, MICHAUD F. Evaluating real-time audio localization algorithms for artificial audition in robotics[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, USA, 2009: 2033-2038.

[27]TAMAI Y, KAGAMI S, AMEMIYA Y, et al. Circular microphone array for robot's audition[C]//IEEE International Conference on Sensors. Vienna, Austria, 2004: 565-570.

[28]TAMAI Y, SASAKI Y, KAGAMI S. Three ring microphone array for 3D sound localization and separation for mobile robot audition[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Edmonton, Canada, 2005: 4172-4177.

[29]NAKADAI K, NAKAJIMA H, YAMADA K, et al. Sound source tracking with directivity pattern estimation using a 64 ch microphone array [C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Edmonton, Canada, 2005: 1690-1696.

[30]NAKADAI K, NAKAJIMA H, MURASE M, et al. Robust tracking of multiple sound sources by spatial integration of room and robot microphone arrays[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. Toulouse, France, 2006: 929-932.

[31]SASAKI Y, KAGAMI S, MIZOGUCHI H. Multiple sound source mapping for a mobile robot by self-motion triangulation[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing, China, 2006: 380-385.

[32]SASAKI Y, KAGAMI S, MIZOGUCHI H. Main-lobe canceling method for multiple sound sources localization on mobile robot[C]//IEEE/ASME International Conference on Advanced Intelligent Mechatronics. Zurich, Switzerland, 2007: 1-6.

[33]KAGAMI S, THOMPSON S, SASAKI Y, et al. 2D sound source mapping from mobile robot using beamforming and particle filtering[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. Taipei, China, 2009: 3689-3692.

[34]SASAKI Y, THOMPSON S, KANEYOSHI M, et al. Map-generation and identification of multiple sound sources from robot in motion[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, China, 2010: 437-443.

[35]SCHMIDT R O. Multiple emitter location and signal parameter estimation[J]. IEEE Transactions on Antennas and Propagation, 1986, 34(33): 276-280.

[36]WANG H, KAVEH M. Coherent signal subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1985, 33(4): 823-831.

[37]居太亮.基于麦克风阵列的声源定位算法研究[D].成都:电子科技大学, 2006.
JU Tailiang. Research on speech source localization methods based on microphone arrays[D]. Chengdu: University of Electronic Science and Technology of China, 2006.

[38]ASANO F, ASOH H, MATSUI T. Sound source localization and signal separation for office robot "Jijo-2"[C]//IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. Taipei, China, 1999: 243-248.

[39]ARGENTIERI S. Broadband variations of the MUSIC high-resolution method for sound source localization in robotics[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. San Diego, USA, 2007: 2009-2014.

[40]NAKAMURA K, NAKADAI K, ASANO F, et al. Intelligent sound source localization for dynamic environments[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, USA, 2009: 664-669.

[41]NAKAMURA K, NAKADAI K, ASANO F, et al. Intelligent sound source localization and its application to multimodal human tracking[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA, 2011: 143-148.

[42]ISHI C T, CHATOT O, ISHIGURO H, et al. Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, USA, 2009: 2027-2032.

[43]LYON R F. A computational model of binaural localization and separation[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. Boston, USA, 1983: 1148-1151.

[44]ALGAZU V R, DUDA R O, MORRISON R P, et al. Structural composition and decomposition of HRTFs[C]//IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New York, USA, 2001: 103-106.

[45]HANDZEL A A, KRISHNAPRASAD P S. Biomimetic sound-source localization[J]. IEEE Journal on Sensors, 2002, 2(6): 607-616.

[46]NAKADAI K, OKUNOT H G, KITANO H. Epipolar geometry based sound localization and extraction for humanoid audition[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Maui, USA, 2001: 1395-1401.

[47]NAKADAI K, HIDAI K, MIZOGUCHI H, et al. Real-time auditory and visual multiple-object tracking for humanoids[C]//Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence. Seattle, USA, 2001: 1425-1436.

[48]NAKADAI K, MATSUURA D, OKUNO H G, et al. Applying scattering theory to robot audition system: robust sound source localization and extraction[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, USA, 2003: 1147-1152.

[49]KUMON M, SHIMODA T, KOHZAWA R. Audio servo for robotic systems with pinnae[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Edmonton, Canada, 2005: 1881-1886.

[50]SHIMODA T, NAKASHIMA T, KUMON M, et al. Spectral cues for robust sound localization with pinnae[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing, China, 2006: 386-391.

[51]HOMSTEIN J, LOPES M, SANTOS-VICTOR J, et al. Sound localization for humanoid robots-building audio-motor maps based on the HRTF[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing, China, 2006: 1170-1176.

[52]KEYROUZ F, MAIER W, DIEPOLD K. A novel humanoid binaural 3D sound localization and separation algorithm[C]//IEEE-RAS International Conference on Humanoid Robot. Genova, Italy, 2006: 296-301.

[53]RODEMANN T, INCE G, JOUBLIN F, et al. Using binaural and spectral cues for azimuth and elevation localization[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Nice, France, 2008: 2185-2190.

[54]RODEMANN T. A study on distance estimation in binaural sound localization[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, China,

2010: 425-430.

[55] KIM U H, MIZUMOTO T, OGATA T, et al. Improvement of speaker localization by considering multipath interference of sound wave for binaural robot audition[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA, 2011: 2910-2915.

[56] SKAF A. Optimal positioning of a binaural sensor on a humanoid head for sound source localization[C]//IEEE-RAS International Conference on Humanoid Robot. Bled, Slovenia, 2011: 165-170.

[57] SAXENA A, NG A Y. Learning sound location from a single microphone [C]//IEEE International Conference on Robotics and Automation. Kobe, Japan, 2009: 1737-1742.

[58] NAKADAI K, LAURENS T, OKUNO H G, et al. Active audition for humanoid[C]//Proceedings of the 17th National Conference on Artificial Intelligence. Austin, USA, 2000: 832-839.

[59] ANDERSSON S B, HANDZEL A A, SHAH V, et al. Robot phonotaxis with dynamic sound-source localization [C]//IEEE International Conference on Robotics and Automation. Barcelona, Spain, 2004: 4833-4838.

[60] MARTINSON E, APKER T, BUGAJSKA M. Optimizing a reconfigurable robotic microphone array[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA, 2011: 125-130.

[61] PORTELLO A. Acoustic models and Kalman filtering strategies for active binaural sound localization[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA, 2011: 137-142.

[62] KUMON M, NODA Y. Active soft pinnae for robots[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA, 2011: 112-117.

[63] OKUNO H G, NAKADAI K, HIDAI K, et al. Human-robot interaction through real-time auditory and visual multiple-talker tracking[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Maui, USA, 2001: 1402-1409.

[64] OKUNO H G, NAKADAI K, KITANO K. Social interaction of humanoid robot based on audio-visual tracking [C]//International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert System. Cairns, Australia, 2002: 1-10.

[65] LV Xiaoling, ZHANG Minglu. Sound source localization based on robot hearing and vision[C]//International Conference on Computer Science and Information Technology. Singapore, 2008: 942-946.

[66] LEE B, CHOI J S, KIM D, et al. Sound source localization in reverberant environment using visual information [C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, China, 2010: 3542-3547.

[67] LIU Hong, SHEN Miao. Continuous sound source localization based on microphone array for mobile robots[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, China, 2010: 4332-4339.

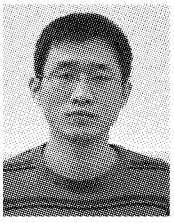
[68] LI Xiaofei, LIU Hong, YANG Xuesong. Sound source localization for mobile robot based on time difference feature and space grid matching[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA, 2011: 2879-2886.

[69] YOUNG S H, SCANLON M V. Detection and localization with an acoustic array on a small robotic platform in urban environments, technical report ADA410432[R]. Adelphi, USA: U.S. Army Research Laboratory, 2003.

[70] SUN Hao, YANG Peng, LIU Zuojun, et al. Microphone array based auditory localization for rescue robot[C]//Chinese Control and Decision Conference. Taiyuan, China, 2011: 606-609.

[71] LUO R C, HUANG C H, LIN T T. Human tracking and following using sound source localization for multisensor based mobile assistive companion robot[C]//IEEE Conference on Industrial Electronics Society. Glendale, USA, 2010: 1552-1557.

作者简介:



李晓飞,男,1987年生,博士研究生,主要研究方向为语音识别、声源定位.

刘宏,男,1967年生,教授,博士生导师,中国人工智能学会常务理事、副秘书长、青年工作委员会主任,主要研究方向为智能机器人、计算机视听觉.先后承担国家自然科学基金项目7项,国家“863”、“973”计划课题项目5项,曾获国家航天科技进步奖.发表学术论文100余篇,其中60余篇被SCI、EI检索.