

一种改进的联合 SRP-PHAT 语音定位算法

袁安富¹ 孟君¹

摘要

提出了一种基于正交直线麦克风阵列的分级搜索 SRP-PHAT(可控响应功率-相位加权)语音定位算法.利用正交直线麦克风阵列将二维搜索空间缩减为2个一维空间以减少计算量,并联合使用四叉树由粗到细的分级搜索(hierarchical search)策略分别在一维空间中搜索声源位置.通过 Matlab 对联合算法进行了仿真,实现了声源定位,并与传统的全搜索 SRP-PHAT 算法和改进前的基于正交直线麦克风阵列的加速 SRP-PHAT 算法进行比较分析.实验与仿真结果显示:该联合算法大大减少了计算量和定位时间,能准确定位声源位置.

关键词

声源定位;可控响应功率-相位加权算法;正交直线麦克风阵;分级搜索

中图分类号 TH71;TG803

文献标志码 A

0 引言

如今,SRP-PHAT 算法凭借其在实际环境下良好健全的性能,已经成为最受青睐的语音定位算法,但是庞大的计算量使得实时地执行算法变得困难.因此,自 SRP-PHAT 算法被提出以来,许多研究者都尝试着减少其核心的可控响应功率搜索过程的计算量^[1-4].文献[1]提出了一种基于正交线性麦克风阵列的加速 SRP-PHAT 算法.在远场条件下,正交线性麦克风阵列把二维的搜索空间缩减成了2个成对的一维搜索空间,主要的计算过程可以分别在2个一维空间内进行.对于一个 M 元阵列,此方案需要的计算量只有常规 SRP-PHAT 算法的 $1/M$,定位精度没有损失,但是该方法只考虑了远场条件下基于正交线性麦克风阵列的情况,具有很大的局限性.文献[2]采取了一种在空间和频率上利用四叉树由粗略到细致进行网格划分的分级搜索策略,显著提高了算法的运算速度.这种策略利用了产生目标函数的信号的一些特性,比如说,语音的波长是随着定位空间的扩大而增长的.首先在一个粗略的网格中进行搜索可控响应功率(SRP)最大值,局部最大值所在区域被选中进行下一轮分级迭代搜索,但是这种方法很容易在开始搜索时选错初始网格,尤其是在调制反响环境中,此时算法会出现估计错误.仿真和实验显示,这种方案在混响时间超过 300 ms 时,不能达到和传统 SRP-PHAT 算法一样的健全性能^[5].

本文联合了上述2种算法的优势,提出了一种改进的联合 SRP-PHAT 语音定位算法,即基于正交直线麦克风阵列的分级搜索 SRP-PHAT 算法.在远场条件下,首先利用正交直线麦克风阵列将二维搜索空间削减为一对一维空间,然后分别在一维空间中执行分级搜索策略,寻找 SRP 最大值以确定声源位置,因此联合算法进一步缩减了原来的文献[1-2]所示算法的计算量,并且声源定位精确度没有明显损失.

1 正交直线麦克风阵列模型

本文采用的 18 元正交等间距直线麦克风阵列声源定位模型如图 1 所示.声源信号的波达方向 DOA(Direction of Arrival)通过坐标原点指向声源的方向向量表示,它与信号波的传播方向相反.麦克风阵元 1—9 分布在直角坐标系的 Y 轴上,其中阵元 $m_0 = 5$ 位于坐标原点即为系统参考点;麦克风阵元 10—18 分布在 Z 轴上,其中阵元 14 与阵

收稿日期 2011-03-01

资助项目 南京信息工程大学科研基金(2008-0326)

作者简介

袁安富,男,教授,硕士生导师,主要研究机电产品的动静态性能分析和优化设计.
charles-yuan@163.com

¹ 南京信息工程大学 信息与控制学院,南京,210044

元 5 位置重合. 相邻麦克风阵元间的距离为 d . 语音信号声源位置采用球坐标表示为 $s = (r, \theta, \phi)$, 其中 r 为声源与麦克风阵列参考点的距离, θ 为方位角, ϕ 为仰角. DOA 单位向量为 $\mathbf{u}_s = [\cos \phi \cos \theta, \cos \phi \sin \theta, \sin \phi]^T$, 则声源位置向量为 $\mathbf{s} = r * \mathbf{u}_s$. 麦克风阵元 m 的位置向量为 $\mathbf{r}_m = [x \ y \ z]^T$. γ 为 DOA 向量与 Z 轴正向的夹角, β 为 DOA 向量与 Y 轴正向的夹角, 可得 $\gamma = \frac{\pi}{2} - \phi$, $\cos \beta = \cos \phi \sin \theta$.

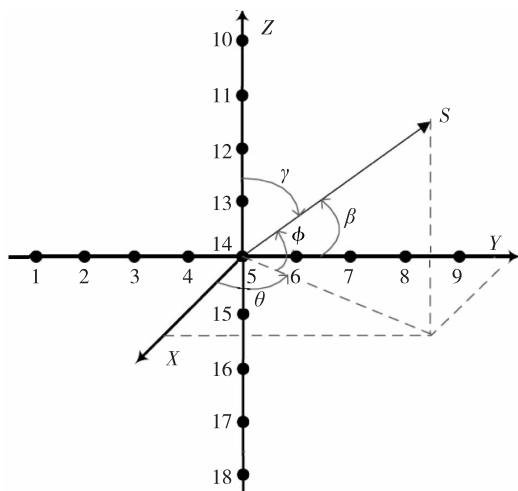


图 1 正交直线麦克风阵列

Fig. 1 Orthogonal linear array of microphones

当麦克风阵列系统应用于一般会议室等室内环境时,若声源和麦克风阵列的距离小于 500 cm,麦克风阵列信号接收模型应采用精确的近场模型,信号波波前为球面波波前;若声源与麦克风阵列相距较远,即 $r \gg d$ 时,则满足远场条件,可以采用近似的平面波前模型. 这时声源的位置向量实际上只是一个方向而已,与距离 r 无关^[6]. 本文研究远场条件下的声源定位算法,因此算法中声源位置的确定过程归结为方位角 θ 和仰角 ϕ 的估测.

2 信号模型

在一个室内混响环境中,某一位置处的信号可以认为与声源信号 $s(t)$ 是线性相关的. 一个 M 元麦克风阵列中第 m 个麦克风接收到的信号可表示为

$$x_m(t) = h_m(t) * s(t) + b_m(t). \quad (1)$$

其中 $m = 1, 2, \dots, M$, $*$ 表示卷积. $h_m(t)$ 为声源与第 m 个麦克风之间的房间脉冲响应 (RIR), 描述了声源与麦克风位置之间所有声道的特性, 包括声源直接到达麦克风的直达声道. RIR 高度依赖于声源位

置和麦克风位置,并且受温度、湿度等环境因素影响,还会因为家具等房间内的物体位置移动而变化. 当这些因素变化显著时,需要假设其在短时间内保持恒定,因此,当声源和麦克风位置确定时,可认为 RIR 在短时间内是时不变的. $b_m(t)$ 为第 m 个麦克风接收到的噪声,包括麦克风系统中的通道噪声和环境噪声 2 部分,其中环境噪声为主导成分. 因此可认为各个麦克风所接收的噪声之间以及噪声与声源信号之间均是独立不相关的. 时延估计算法的目的是计算 2 个麦克风接收信号的直达时间差别,因此 RIR 可以分解成 2 部分,一部分为直达声道的响应,剩余的另一部分设为 $h'_m(t)$. 因此式(1)改写为

$$x_m(t) = \alpha_m s(t - \tau_m) + h'_m(t) * s(t) + b_m(t). \quad (2)$$

其中 $m = 1, 2, \dots, M$, $\alpha_m = r_m^{-1}$ 是声音在大气中传播的幅度衰减因子, $\tau_m = r_m/c$ 是信号从声源位置直接到达第 m 个麦克风的传播时延,这里 r_m 是声源到第 m 个麦克风的距离, c 为声音在空气中的传播速度. 这时麦克风接收到的信号即为声源信号的延迟形式加上声源信号与剩余 RIR 部分的卷积^[7]. 信号到达麦克风 m 和麦克风 n 的时延差别记为 $\tau_{mn} = \tau_n - \tau_m$.

3 联合 SRP-PHAT 算法

3.1 基于正交直线麦克风阵列的分级搜索算法

SRP-PHAT 算法联合了相位变换加权和可控响应功率,通过在所有可能覆盖声源位置的范围内搜索可控响应功率 SRP 的最大值来确定声源位置. 由文献[1]可知,可控响应功率是 M 元麦克风阵列中所有麦克风对的广义互相关函数总和,即

$$P^{\text{PHAT}}(s) = \int \left| \sum_{m=1}^M \frac{X_m(\omega) e^{-j\omega\tau_{m_0}(s)}}{|X_m(\omega)|} \right|^2 d\omega. \quad (3)$$

其中, $X_m(\omega)$ 是 $x_m(t)$ 的傅里叶变换形式, $s = (r, \theta, \phi)$ 是语音信号声源位置的球坐标, $\tau_{m_0}(s)$ 是可控波束形成过程中麦克风 m 处的可控时延,即信号直达麦克风 m 的真实时延与信号直达参考麦克风 m_0 时延的相对时延. 当声源位于远场时,信号波波前可以近似认为是平面波,因此可控时延与声源的距离无关.

$$\tau_{m_0}(s) = \mathbf{u}_s \cdot (\mathbf{r}_m - \mathbf{r}_{m_0})/c. \quad (4)$$

其中 \mathbf{u}_s 是声源位置的单位方向矢量, \mathbf{r}_m 和 \mathbf{r}_{m_0} 分别是麦克风 m 和参考麦克风 m_0 在直角坐标系中的位置向量, c 是声音传播速度. 在执行算法时,麦克风接收的信号都经过采样,信号处理过程处理的对象是离散信号

$$x_m(n) = x_m(nT_s), \quad m = 1, 2, \dots, M. \quad (5)$$

其中 $T_s = f_s^{-1}$ 是采样周期, f_s 是采样频率. 在 SRP-PHAT 算法中, 声源位置可以通过麦克风阵列信号的一帧数据估算, 仍用 $x_m(n)$ 表示一帧数据, 数据帧的离散傅里叶变换记作 $X_m(k)$. 把式(3)中的傅里叶变换替换成各自的数据帧离散傅里叶变换形式, 则相位变换可控响应可以表示为

$$\hat{Y}^{\text{PHAT}}(k, s) = \sum_{m=1}^M \frac{X_m(k)}{|X_m(k)|} e^{-j\omega\tau_{mm_0}(s)}, \quad (6)$$

其中 ω 仍是连续信号的频率. 把总共 K 个离散帧的可控响应相加, 则可以得到基于相位变换加权的可控响应功率

$$\hat{P}^{\text{PHAT}}(s) = \sum_{k=0}^{K-1} \hat{Y}^{\text{PHAT}}(k, s) \hat{Y}^{\text{PHAT}*}(k, s), \quad (7)$$

其中 * 表示复共轭.

图 1 所示直角坐标系中, 麦克风阵元 m 的位置为

$$\mathbf{r}_m = \begin{cases} [0 \quad (m-5)d \quad 0]^T, & m = 1, 2, \dots, 9, \\ [0 \quad 0 \quad (14-m)d]^T, & m = 10, 11, \dots, 18. \end{cases} \quad (8)$$

参考麦克风 m_0 的位置向量为 $\mathbf{r}_{m_0} = [0 \ 0 \ 0]^T$. 已知在远场条件下, 声源位置与声源距离无关, 由方位角 θ 和仰角 ϕ 确定. 把以上位置向量代入 SRP-PHAT 算法的可控时延式(4), 并且由上文可知, 声源单位方向向量 $\mathbf{u}_s = [\cos \phi \cos \theta, \cos \phi \sin \theta, \sin \phi]^T$, DOA 向量与 Z 轴正向的夹角 γ 和 DOA 向量与 Y 轴正向的夹角 β 分别满足 $\gamma = \frac{\pi}{2} - \phi$, $\cos \beta = \cos \phi \sin \theta$, 则麦克风阵列中第 m 个麦克风的可控时延为

$$\tau_{mm_0}(s) = \begin{cases} (m-5)d \cos \beta / c, & m = 1, 2, \dots, 9, \\ (14-m)d \cos \gamma / c, & m = 10, 11, \dots, 18. \end{cases} \quad (9)$$

此时, 相位变换可控响应(6)可以改写成

$$\hat{Y}^{\text{PHAT}}(k, s) = \sum_{m=1}^M \frac{X_m(k)}{|X_m(k)|} e^{-j\omega\tau_{mm_0}(s)} = \sum_{m=1}^9 \frac{X_m(k)}{|X_m(k)|} e^{-j\omega(m-5)d \cos \beta / c} + \sum_{m=10}^{18} \frac{X_m(k)}{|X_m(k)|} e^{-j\omega(14-m)d \cos \gamma / c} = \hat{Y}_1^{\text{PHAT}}(k, \beta) + \hat{Y}_2^{\text{PHAT}}(k, \gamma). \quad (10)$$

通过式(10), 可以将 DOA 元素分离开来考虑, 使得二维的搜索空间被削减成为一对一维空间. 具体地说, 就是 $\hat{Y}_1^{\text{PHAT}}(k, \beta)$ 和 $\hat{Y}_2^{\text{PHAT}}(k, \gamma)$ 可以分别在一个一维空间中计算, 因此算法总的计算量会大大地减少. 在一维搜索空间中寻找 SRP 最大值时, 首先需要进行横向寻优, 在一维空间 $\beta \in [0, \pi]$ 中进行角度搜

索, 计算每一个候选 β 处的可控响应功率部分; 然后如上进行纵向寻优, 在一维空间 $\gamma \in [0, \pi]$ 中进行角度搜索, 计算每个候选 γ 处的可控响应功率部分; 最后由式(7)、(10)计算出基于相位变换加权的可控响应功率. 同样地, 可控响应功率最大值对应的 (β, γ) 即为算法的执行结果, 再通过角度转化, 便可以得到声源位置 (θ, ϕ) .

这里的角度搜索过程都是全角度搜索, 因此仍需要处理相当大的计算量. 若是在横向角度搜索和纵向角度搜索过程中应用文献[2]的分级搜索策略, 利用四叉树把一维角度搜索空间划分为多个子区, 四叉树的所有节点都可以被继续分割成 4 个子节点进行迭代搜索; 在构建的能量分布图中, 局部极值对应的子区被选中进入下一轮算法的分级搜索; 当四叉树不断分叉直至达到期望的级数和定位精确度, 即达到设定的最小子区大小时, 迭代过程终止. 在分级搜索的最后一级, 能量最大的子区的中心角度作为声源定位的结果. 假设麦克风阵列位于房间的一面墙上, 那么声源位置都将位于麦克风阵列的前方. 首先建立一个 1 级的四叉树, 一维搜索空间总共被粗级划分成了 2^{21} 个子区, 若存在多个声源时, 这些子区的大小需要足够小, 使得不会有 2 个声源位于同一子区中. 需要注意的是, DOA 空间中的子区在房间坐标体系中并不是矩形的, 需要先由已知的房间大小估计出声源位置距离麦克风阵可能的最远距离, 参照此距离把 DOA 划分网格以 (x, y, z) 的坐标形式重新绘制成坐标系网格, 再通过坐标系网格来选择截止频率和初始网格大小. 相对于全角度搜索, 分级搜索策略的优势已经在文献[2]中得以证明, 因此联合算法进一步缩减了原来的单一利用正交直线麦克风阵列结构优势或分级搜索策略的算法的计算量, 集合了这 2 种算法的核心优势, 并且声源定位精确度没有明显损失.

3.2 算法执行

本文提出的基于正交直线麦克风阵列的分级搜索算法在执行时, 首先利用正交直线麦克风阵列的几何特性, 把二维的 DOA 空间划分成 2 个一维 DOA 空间; 然后分别在横向角度和纵向角度, 利用四叉树将一维搜索空间分级划分为多个子区, 计算所有子区中心处的可控波束形成输出的 SRP; 能量最大的子区被选中为最有可能的声源位置, 进入下一轮分级迭代搜索; 四叉树不断分叉达到某一层度, 即达到设定的最小子区大小时, 迭代过程终止. 在分级搜索

的最后一级,能量最大的子区的中心角度作为算法执行的结果,即声源位置.具体步骤如下.

第1步.构建混响声学房间系统.对于 M 元麦克风阵列,算法先由式(2)、(5)计算出 M 个麦克风接收到的离散信号 $x_m(n)$,并求其离散傅里叶变换 $X_m(k)$.

第2步.按式(9)计算出正交直线麦克风阵列中信号到达各个麦克风处的可控时延 $\tau_{mq}(s)$.

第3步.进行横向寻优,在一维空间 $\beta \in [0, \pi]$ 中进行二叉树分级角度搜索.先在频域对信号进行低通滤波处理,截止频率为1 kHz,初始子区划分大小为 4×4 .因为角度搜索精度为 1° ,所以二叉树分叉阶次为6,即子区边界大小依次为 $\pi/4, \pi/8, \pi/16, \pi/32, \pi/64, \pi/128$.在每一轮能量局部极值搜索过程中,按式(10)依次计算每一个子区候选角度 β 处的可控响应 $\hat{Y}_1^{\text{PHAT}}(k, \beta)$,取得最大值的子区被选中进入下一轮迭代搜索.

第4步.如上进行纵向寻优,在一维空间 $\gamma \in [0, \pi]$ 中进行二叉树分级角度搜索,按式(10)计算每个子区候选 γ 处的可控响应 $\hat{Y}_2^{\text{PHAT}}(k, \gamma)$.

第5步.由式(7)、(10)计算出基于相位变换加权的可控响应功率 $\hat{P}^{\text{PHAT}}(s)$,可控响应功率最大值对应的 (β, γ) 即为算法的执行结果,通过角度转化,便可以得到声源位置 (θ, ϕ) .

4 算法仿真

本文基于正交直线麦克风阵列,利用 Matlab 构建了一个存在混响和噪声的声学房间环境,对提出的基于正交直线麦克风阵列的分级搜索 SRP-PHAT 算法(以下简称为联合 SRP-PHAT 算法)的性能进行了仿真,并与传统的全搜索 SRP-PHAT 算法、文献[1]的基于正交直线麦克风阵列的加速 SRP-PHAT 算法(以下简称为阵列加速 SRP-PHAT 算法)的定位性能进行比较和分析.正交直线麦克风阵列如图1所示,麦克风之间距离均为0.04 m.混响声学房间的仿真利用了镜像模型^[8],由此可以计算出房间脉冲响应 RIR.本文中 RIR 计算到了第8级反射,混响时间 T_{60} 变化范围为50~200 ms.将声源信号与各个麦克风位置的 RIR 进行卷积,可以得到各个麦克风输出的仿真信号.仿真出的房间是一个由平面反射界面(墙壁、天花板、地板)围成的矩形空间,各界面的反射系数与声源信号的频率无关,且不随声源信号的入射角度的改变而变化.本文仿真实验中,以房间

地面的一角为坐标原点,构建如图2所示房间坐标系,房间三维尺寸为(4 m, 4 m, 3 m).天花板的反射系数为0.9,地板的反射系数为0.3,墙壁 ZOY 与对面墙壁反射系数分别为0.7和0.25,墙壁 ZOX 与对面墙壁反射系数分别为0.25和0.85.麦克风阵列位于墙壁 ZOY 上,阵列中心即图1所示坐标系的原点位于房间坐标系的(0, 2, 1.5)处.声音在空气中的传播速度为342 m/s.声源信号的内容为男声英文数字“one, two, three, four...”,数字间有短暂停顿,时长10 s,由计算机通过放置于说话人嘴边的麦克风采集,信号采样频率为8 kHz,采样精度为16 bit,单声道,PCM 格式.数据帧的大小为256个采样点(32 ms),窗函数为汉宁窗.仿真时去除了语音信号中空白停顿的频谱,共选出100个没有重叠的语音数据帧用来执行算法.算法执行得出的结果为估计声源位置的方位角 θ 和仰角 ϕ ,若其中任意一个估计值与实际声源位置方位角或仰角偏差超过 5° ,则认为是异常估计.通过衡量非异常估计的均方根误差(RMSE)来对比分析本文提出的联合 SRP-PHAT 算法(算法3)、传统的全搜索 SRP-PHAT 算法(算法1)以及阵列加速 SRP-PHAT 算法(算法2)的性能.均方根误差公式为^[1]

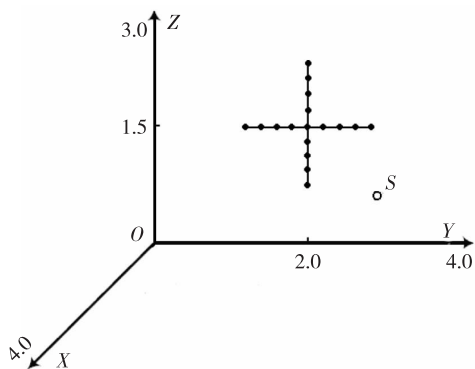


图2 混响房间坐标系(单位:m)

Fig. 2 Coordinate system of the reverberant room(unit:m)

$$E_{\text{RMSE}} = \sqrt{N_{\text{na}}^{-1} \sum_i [|\hat{\theta}_i - \theta_s|^2 + |\hat{\phi}_i - \phi_s|^2]}. \quad (11)$$

其中 $\hat{\theta}_i$ 和 $\hat{\phi}_i$ 是用数据帧 i 执行算法时得到的声源方位角及仰角的估计值, θ_s 和 ϕ_s 是声源的实际方位角及仰角, N_{na} 是所有非异常估计的个数.根据不同的混响时间、不同的信噪比(SNR)以及不同的声源位置,本文共做了3个仿真实验,实验中由 Matlab 在麦克风处对声源信号添加适量的高斯白噪声,以获得所需信噪比.

实验 1. 实际声源位置为 $S_1 = (2 \text{ m}, 37^\circ, 11^\circ)$, 信噪比设为 30 dB, 混响时间 T_{60} 分别为 50、100、150、200 ms. 仿真结果如表 1 所示.

实验 2. 实际声源位置为 $S_1 = (2 \text{ m}, 37^\circ, 11^\circ)$, 混响时间 T_{60} 为 100 ms, 信噪比分别为 20、25、30 dB. 仿真结果如表 2 所示.

实验 3. 信噪比设为 30 dB, 混响时间 T_{60} 为 100 ms, 实际声源位置分别为 $S_1 = (2 \text{ m}, 37^\circ, 11^\circ)$ 、 $S_2 = (0.5 \text{ m}, -60^\circ, 30^\circ)$ 、 $S_3 = (2 \text{ m}, 4^\circ, -3^\circ)$. 仿真结果如表 3 所示.

从实验 1、2 可以发现, 当信噪比较好、混响影响不明显时, 改进的 SRP-PHAT 定位方法与原全搜索算法的错误估计率都较低, 均有着良好的性能, 但随着混响时间增加, 噪声干扰变大时, 3 种算法的性能都有着不同程度的下降, 全搜索算法在恶劣条件下有着较强鲁棒性. 联合算法的性能与另 2 种算法性能的差距并不明显, 在保证良好的抗噪性能和抗混响能

力的同时, 极大地减少了算法计算量和定位时间.

实验 3 说明, 在声源位置的方向角和仰角较大时, 算法定位的误差也随之变大. 该实验结果符合实际人耳定位特性^[9-10].

5 结论

在保证 SRP-PHAT 声源定位方法的抗噪声、抗混响能力的前提下, 为了降低其庞大计算量以提高实时处理速度, 本文提出了改进的联合 SRP-PHAT 声源定位方法. 基于正交直线麦克风阵列, 首先利用缩减二维搜索空间为一对一维搜索空间的加速策略, 减少了 SRP-PHAT 算法的主要计算量, 然后分别在一维空间中按由粗到细的分级搜索策略搜索能量峰值. 改进的 SRP-PHAT 方法计算量再次大幅度降低, 声源定位精确度没有明显损失, 能准确定位声源位置, 并且在噪声和混响环境中能保持良好的鲁棒性.

表 1 算法定位性能(信噪比 = 30 dB, 混响时间变化)

Table 1 Performance of algorithms with SNR = 30 dB and variable reverberant time

算法	T_{60}/ms							
	50		100		150		200	
	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$
算法 1	90	2.012 5	87	2.270 6	87	2.483 2	85	2.505 5
算法 2	88	2.125 7	85	2.317 3	81	2.476 6	80	2.549 7
算法 3	84	2.363 1	82	2.479 2	80	2.629 0	77	2.825 0

表 2 算法定位性能(混响 100 ms, 信噪比变化)

Table 2 Performance of algorithms with $T_{60} = 100 \text{ ms}$ and variable SNR

算法	信噪比/dB					
	20		25		30	
	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$
算法 1	84	2.338 9	87	2.321 1	87	2.270 6
算法 2	80	2.408 2	83	2.670 0	85	2.317 3
算法 3	75	2.886 7	80	2.822 9	82	2.479 2

表 3 算法定位性能(混响 100 ms, 信噪比 = 30 dB, 声源位置改变)

Table 3 Performance of algorithms for different locations with $T_{60} = 100 \text{ ms}$ and SNR = 30 dB

算法	声源位置					
	$(37^\circ, 11^\circ)$		$(-60^\circ, 30^\circ)$		$(4^\circ, -3^\circ)$	
	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$	非异常率/%	$E_{\text{RMSE}}/(\text{^\circ})$
算法 1	87	2.270 6	85	2.295 0	91	2.167 9
算法 2	85	2.317 3	82	2.353 4	88	2.279 8
算法 3	82	2.479 2	79	2.596 0	85	2.386 0

参考文献

References

- [1] Cai W P, Wang S K, Wu Z Y. Accelerated steered response power method for sound source localization using orthogonal linear array[J]. *Applied Acoustics*, 2010, 71 (2): 134-139
- [2] Zotkin D N, Duraiswami R. Accelerated speech source localization via a hierarchical search of steered response power[J]. *IEEE Transactions on Speech and Audio Processing*, 2004, 12(5): 499-508
- [3] Dmochowski J P, Benesty J, Affes S. A generalized steered response power method for computationally viable source localization[J]. *IEEE Transactions on Audio, Speech and Language Processing*, 2007, 15 (8): 2510-2526
- [4] Do H, Silverman H F, Yu Y. A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array [C] // *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, 2007: I-121-I-124
- [5] Perterson J, Kyriakakis C. Analysis of fast localization algorithms for acoustical environments[C] // *Proceedings of the 39th Asilomar Conference on Signals, System and Computer*, 2005: 1385-1389
- [6] 居太亮. 基于麦克风阵列的声源定位算法研究[D]. 成都: 电子科技大学通信与信息工程学院, 2006
JU Tai liang. Research on speech source localization methods based on microphone arrays[D]. Chengdu: University of Electronic Science and Technology of China, 2006
- [7] DiBiase J H, Silverman H F, Brandstein M S. Robust localization in reverberant rooms[M] // *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001: 157-180
- [8] Allen J B, Berkley D A. Image method for efficiently simulating small-room acoustics[J]. *J Acoust Soc Am*, 1979, 65(4): 943-950
- [9] 谭颖, 殷福亮, 李细林. 改进的 SRP-PHAT 声源定位方法[J]. *电子与信息学报*, 2006, 28(7): 1223-1227
TAN Ying, YIN Fuliang, LI Xilin. Sound localization method using modified SRP-PHAT algorithm[J]. *Journal of Electronics & Information Technology*, 2006, 28(7): 1223-1227
- [10] 孟君. 基于 DSP 的数字助听器多通道响度补偿方案[J]. *南京信息工程大学学报: 自然科学版*, 2010, 2(5): 420-425
MENG Jun. Multi-channel loudness compensation method for digital hearing aids based on DSP[J]. *Journal of Nanjing University of Information Science & Technology: Natural Science Edition*, 2010, 2(5): 420-425

A combined SRP-PHAT method for sound source localization

YUAN Anfu¹ MENG Jun¹

¹ School of Information and Control, Nanjing University of Information Science & Technology, Nanjing 210044

Abstract In microphone arrays application, it is difficult to localize sound source accurately and quickly in a noisy and reverberant environment. In order to solve this problem, many researchers have presented different approaches. The steered response power-phase transform weighted (SRP-PHAT) source localization algorithm has been proved robust, however, it requires high computation cost for searching the peak of steered response power in a large location space. Thus, an improved SRP-PHAT method using an orthogonal linear array was presented in this paper, which reduced a two-dimension searching space to a couple of one-dimension ones. Then the parameters of direction of arrival (DOA) were separated, and computed respectively in the one-dimension searching space with coarse-fine strategies of hierarchical search. This algorithm was based on the observation that the wavelengths of the sound from a speech source are comparable to the dimensions of the space being searched and that the source was broadband. A systematic series of comparisons with previous algorithms were made basing on Matlab. Simulations show that the main computational load of SRP-PHAT algorithms has been greatly cut down, and there is no loss in accuracy in the proposed method. The performance of the algorithm can be further improved by using constraints from computer vision.

Key words sound source localization; SRP-PHAT; orthogonal linear array; hierarchical search