

DOI: 10.19650/j.cnki.cjsi.J2413576

基于二次互相关 SRP-PHAT 算法声源定位研究

席旭刚¹, 王晨¹, 李文国², 丁一², 马秉宇²

(1. 杭州电子科技大学智能控制与机器人研究所 杭州 310018; 2. 咸亨国际(杭州)电气制造有限公司 杭州 310009)

摘要:可控响应功率相位变换算法(SRP-PHAT)因其在混响环境下的较强鲁棒性被广泛应用于声源定位。然而,传统的SRP-PHAT算法在多传声器阵列声源定位场景下的定位精度不足且计算量较大,不能满足高精度实时声源定位的需求。针对上述问题,提出一种基于二次互相关的SRP-PHAT算法,将阵列中两组通道信号间自相关和互相关的结果进行二次互相关运算,基于广义互相关相位变换函数(GCC-PHAT)进一步计算得到改进的SRP-PHAT函数,对其进行峰值搜索实现声源定位,以提高定位精度;在计算方面,通过将传声器阵列划分为参考通道和声源通道,仅在两组通道间进行互相关运算,避免了传统算法在全通道之间逐一计算带来的冗余,极大地减少了运算量。将传统的SRP-PHAT算法与基于二次互相关的SRP-PHAT算法在自研的128阵元多螺旋臂阵列声源定位系统进行移植,并在室内进行4种声源频率下(10~25 kHz)的声源定位实验。实验结果显示改进后的算法对4种不同声源频率下的声源定位时的方位角估计误差平均降低2.5°,俯仰角估计误差平均降低2°,定位的空间分辨率平均提升45.78%。改进后的算法相较于原算法在提高定位精度的同时大幅降低了计算量,为SRP-PHAT算法在多传声器阵列的实时声源定位提供了有效解决方案。

关键词: 声源定位; 麦克风阵列; 相位变换; 二次互相关

中图分类号: TH89 TB52 **文献标识码:** A **国家标准学科分类代码:** 510.40

Research on sound source localization based on the second-order cross-correlation SRP-PHAT algorithm

Xi Xugang¹, Wang Chen¹, Li Wenguo², Ding Yi², Ma Bingyu²

(1. Institute of Intelligent Control and Robotics, Hangzhou Dianzi University, Hangzhou 310018, China;

2. Xianheng International Technology Co., Ltd., Hangzhou 310009, China)

Abstract: The steered response power with phase transform (SRP-PHAT) algorithm is widely used for sound source localization due to its strong robustness in reverberant environments. However, the traditional SRP-PHAT algorithm has insufficient localization accuracy and large computational volume in multi-microphone array source localization scenarios, which cannot meet the demand for high-precision real-time source localization. Aiming at the above problems, a SRP-PHAT algorithm based on quadratic cross-correlation is proposed. In this approach, the autocorrelation and cross-correlation between the signals of two groups of channels in the array are subjected to quadratic cross-correlation. The generalized cross-correlation phase transform function (GCC-PHAT) is used as the basis for further calculation to obtain the improved SRP-PHAT function, followed by peak detection to achieve sound source localization with improved accuracy. To reduce computational load, the microphone array is divided into reference channels and sound source channels, and correlation operations are performed only between these groups. This avoids the redundancy of traditional pairwise computations across all channels, significantly lowering the computational cost. The traditional SRP-PHAT algorithm and the SRP-PHAT algorithm based on quadratic inter-correlation are transplanted in the self-developed 128-array multi-spiral-arm array source localization system, and the experiments of source localization are conducted indoors at four source frequencies (10~25 kHz). The experimental results show that the improved algorithm reduces the azimuth estimation error by an average of 2.5°, the pitch estimation error by an average of 2°, and the spatial resolution of the localization by an average of 45.78% for the localization of sound sources at four different source frequencies. Compared with the conventional algorithm, the improved algorithm improves the localization accuracy and reduces the computation

amount significantly, which provides an effective solution for the SRP-PHAT algorithm for real-time sound source localization in multi-microphone arrays.

Keywords: sound source localization; microphone array; phase transform; secondary cross-correlation

0 引言

基于传声器阵列的声成像技术是通过测量二维全息面上的声压,运用各种重构算法重构声源表面的声场并将声场以图形或动画的形式显示出来^[1-2]。其基本原理是通过接收由传声器阵列采集的声音信号,采用适当的信号处理算法计算得出声源的位置及参数信息并融合成声源定位图像。该技术能够将声源在真实环境中的具体位置直观的进行显示。作为现代声学信号处理领域的一个关键研究方向,该技术对于工业巡检、语音识别、人机交互等多个领域具有重要的应用价值^[3-4]。在实际应用中,如何在低信噪比和复杂环境中快速准确地定位声源成为了一个亟待解决的问题。

基于相位变换加权的可控响应功率算法 (steered response power-phase transform, SRP-PHAT),在理论上具有较高的定位精度,被广泛应用于声源定位^[4]。但在实际应用中该算法往往受限于环境噪声和混响的影响,导致定位性能下降,其在计算复杂度上也存在一定的局限性,例如在多传声器下的计算量过大,难以应用于需要实时处理的场合^[5-6]。近年来,学者们提出了多种改进算法以提升 SRP-PHAT 算法的定位性能。在提升定位精度方面,Wang 等^[7]提出子空间可控响应功率 (subspace-steered response power, Sub-SRP) 算法,利用一系列的滤波和谱减法技术滤除大部分的非扬声器的噪声源,这一方法使得 SRP-PHAT 算法定位精度提升 10%,在强干扰条件下,提升达到了 17%;在针对 SRP-PHAT 算法进行峰值搜索时峰值丢失的问题,文献[8]提出一种迭代重加权可控响应功率 (iterative reweighted-steered response power, IR-SRP) 方法,利用通道间相位差信息,找到主导声源的显著 SRP 峰值,并消除这些声源所贡献的时频单元,使得剩余的声源就可以在后续迭代中在剩余的时频单元中被找到;刘祥楼等^[9]提出基于增强型递归最小平均功率算法 (recursive logarithmic least mean power, RLLMP) 和倒谱预滤波的二次互相关算法。通过引入增强型递归最小平均功率算法和倒谱法进行预滤波,提高时延估计精度;在减少计算量方面,Cobos 等^[10]提出了一种粗网格搜索策略,考虑了空间网格离散位置周围的体积并对 SRP-PHAT 函数进行扩展,在采样空间进行全面探索增加了算法的鲁棒性,并减少了计算量;Yin 等^[11]从使用 SRP-PHAT 特征的三维卷积神经网络 (cross-dimensional 3D convolutional neural network, Cross3D) 模型

出发,提出低复杂度 SRP 特征 (low complexity-steered response power-edge, LC-SRP-Edge) 和轻量化 Cross3D 神经网络 (Cross3D-Edge) 算法优化 SRP 特征表示和神经网络,降低硬件开销,在多种 SRP 分辨率下减少计算复杂度和神经网络权重,显著提升了计算效率。

根据现有研究,SRP-PHAT 算法的改进主要集中在改进互相关函数来提升定位精度以及通过优化 SRP-PHAT 函数的搜索过程来降低计算量。针对提升定位精度方面,本研究提出了一种基于二次互相关的 SRP-PHAT 算法。将两组通道信号自相关和互相关运算后的结果进行二次互相关运算,基于相位变换的广义互相关函数 (generalized cross-correlation with phase transform, GCC-PHAT) 进一步计算得到改进的 SRP-PHAT 函数,从而发挥二次互相关带来的高精度和抗混响能力。在减少计算量方面,将信号采集通道分为声源通道和参考通道,仅在这两组通道之间进行互相关运算,避免了传统算法在全通道之间逐一计算带来的冗余,大大降低了算法计算量。采用 128 通道进行数据采集,对改进后的算法进行了仿真分析,并在室内进行了单声源定位实验。实验结果表明,与传统的 SRP-PHAT 算法相比,二次互相关的 SRP-PHAT 算法在定位精度和空间分辨率上具有明显优势。

1 算法原理

1.1 SRP-PHAT 算法原理

广义互相关相位变换算法 (GCC-PHAT) 是一种基于频域互功率谱加权的声源定位方法,能有效抑制噪声和混响的影响^[12]。在此基础上发展而来的可控响应功率相位变换算法 (SRP-PHAT) 由美国科学家 DiBiase^[13]于 2000 年首次提出。该算法的核心思想是通过假设声源位置,计算所有麦克风对信号的 GCC-PHAT 函数之和,在整个声源空间寻找 SRP-PHAT 函数最大值来估计声源位置。具体实现步骤为:

首先对麦克风阵列接收到的信号进行分帧处理,对于每一帧信号:假设第 m 个麦克风与第 l 个麦克风接收到的信号分别为 $x_m(n)$ 和 $x_l(n)$,将其进行快速傅里叶变换 (fast Fourier transform, FFT) 转换至频域得到 2 个麦克风的频域信号 $X_m(n)$ 和 $X_l(n)$ 。计算两组麦克风接收信号的互相关功率 $X_{ml}(n)$,即:

$$X_{ml}(n) = X_m(n) X_l^*(n) \quad (1)$$

计算信号到达第 m 个麦克风和 l 个麦克风的时延 τ_{ml} ,

假设 $\mathbf{q}$ 为声源位置的直角坐标矢量, $\mathbf{r}_m$ 和 $\mathbf{r}_l$ 为麦克风 m 和麦克风 l 的位置直角坐标, 则时延 τ_{ml} 计算表达式为:

$$\tau_{ml}(\mathbf{q}) = \frac{(\|\mathbf{q} - \mathbf{r}_m\| - \|\mathbf{q} - \mathbf{r}_l\|)}{c} \quad (2)$$

其中, c 为声音在空气中传播的速度。

计算第 l 个麦克风和第 m 个麦克风接收信号的 GCC-PHAT 函数 $R_{ml}(\tau)$ 。

$$R_{ml}(\tau) = \frac{1}{K} \sum_{n=0}^K \frac{X_{ml}(n)}{|X_{ml}(n)|} e^{j\omega\tau} \quad (3)$$

其中, K 为傅里叶变换的长度, ω 是模拟角频率。

利用式 (2) 和 (3) 中计算得到时延 τ_{ml} 以及 GCC-PHAT 函数 $R_{ml}(\tau)$ 求出这一帧信号的 SRP-PHAT 函数。

$$\hat{P}(\mathbf{q}) = \sum_{l=1}^M \sum_{m=1}^M \hat{R}_{ml}[\tau_{ml}(\mathbf{q})] \quad (4)$$

其中, M 为麦克风阵列的总麦克风数, $\mathbf{q}$ 为声源位置的直角坐标矢量。通过搜索 SRP-PHAT 函数的最大值来估计声源的位置, 真实声源的位置可以表示为:

$$\hat{\mathbf{q}} = \text{argmax}[\hat{P}(\mathbf{q})] \quad (5)$$

该算法可实现较高精度的声源定位, 具有一定的抗混响能力^[14]。但在低信噪比和复杂噪声环境下该算法的定位性能会下降, 且计算量较大, 不适用于实时声源定位场景。

1.2 倒谱预滤波的二次互相关算法

GCC-PHAT 算法通过对两路信号在频域上进行互相关得到互功率谱, 在经过相位变换 (phase transform, PHAT) 加权函数^[15]的加权处理检索互相关函数峰值得到时延估计。基于倒谱预滤波的二次互相关算法在此基础上对两路信号全通分量频谱求取自相关和互相关功率谱后所得的结果进行二次互相关运算, 得到二次互相关功率谱, 并在此基础上进行时延估计。算法具体步骤为:

假设第 m 个和第 l 个麦克风接收到的信号分别为 $\mathbf{x}_m(n)$ 和 $\mathbf{x}_l(n)$, 对其进行 FFT 转换至频域得到 2 个麦克风的频域信号 $\mathbf{X}_m(n)$ 和 $\mathbf{X}_l(n)$ 。

计算两组信号的倒谱 $\mathbf{c}_m(n)$ 和 $\mathbf{c}_l(m)$: 将 $\mathbf{X}_m(n)$ 和 $\mathbf{X}_l(n)$ 的绝对值取对数, 然后做傅里叶逆变换 (inverse fast Fourier transform, IFFT)。

将倒谱 $\mathbf{c}_m(n)$ 和 $\mathbf{c}_l(m)$ 通过同态滤波器得到信号的最小相位分量的倒谱 $\mathbf{c}_m^{\min}(n)$ 和 $\mathbf{c}_l^{\min}(n)$ 。

对最小相位分量倒谱 $\mathbf{c}_m^{\min}(n)$ 和 $\mathbf{c}_l^{\min}(n)$ 的傅里叶变换做指数运算得到两路信号的最小相位分量的频谱 $\mathbf{X}_m^{\min}(n)$ 和 $\mathbf{X}_l^{\min}(n)$ 。计算信号的全通分量频谱, 即:

$$\mathbf{X}_m^{\text{all}}(f) = \frac{\mathbf{X}_m(f)}{\mathbf{X}_m^{\min}(f)} \quad (6)$$

$$\mathbf{X}_l^{\text{all}}(f) = \frac{\mathbf{X}_l(f)}{\mathbf{X}_l^{\min}(f)} \quad (7)$$

对信号 $\mathbf{X}_m^{\text{all}}(f)$ 求自相关功率谱, 即:

$$\mathbf{X}_{mm}^{\text{all}}(f) = \mathbf{X}_m^{\text{all}}(f) \mathbf{X}_m^{\text{all}*}(f) \quad (8)$$

对每两个信号求互相关功率谱, 即:

$$\mathbf{X}_{ml}^{\text{all}}(f) = \mathbf{X}_m^{\text{all}}(f) \mathbf{X}_l^{\text{all}*}(f) \quad (9)$$

对 $\mathbf{X}_{mm}^{\text{all}}(f)$ 和 $\mathbf{X}_{ml}^{\text{all}}(f)$ 求二次互相关功率谱, 即:

$$\mathbf{Y}(f) = \mathbf{X}_{mm}^{\text{all}}(f) \mathbf{X}_{ml}^{\text{all}*}(f) \quad (10)$$

利用式 (10) 的二次互相关功率谱估计时延。该方法较于 GCC-PHAT 算法具有更高精度的时延估计。

1.3 128 阵元螺旋麦克风阵列模型

常见的麦克风阵列排布有线阵、十字阵、圆阵等。麦克风数量、间距等参数对最终的声源系统定位精度有着很大的影响^[16-18]。多螺旋臂阵列在采集音频信号时具有较好的抑制旁瓣性能。采用 128 阵元多螺旋臂麦克风阵列采集算法音频信号, 麦克风阵列参数如表 1 所示。

表 1 阵列参数
Table 1 array parameters

指标	参数
阵列增益	21 dB
孔径	118 mm
阵元平均间隔	11.39 mm

阵列的分布如图 1 所示。

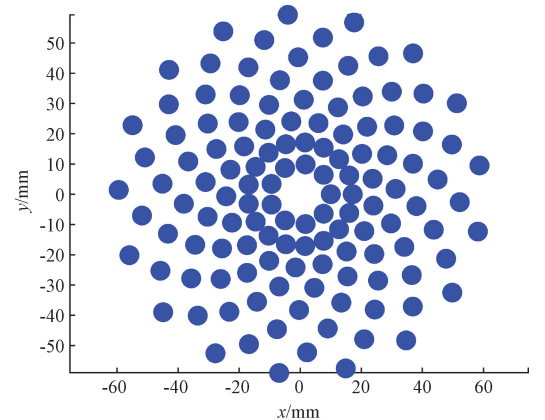


图 1 麦克风位置分布

Fig. 1 Microphone location distribution

阵列方向图可以反映麦克风阵列对空间中不同方向上信号的接收能力。在 MATLAB 上对上述 128 阵元多螺旋臂阵列的方向图进行仿真, 结果如图 2 所示。

由图 2 可以看出, 该阵列能够有效的抑制旁瓣水平, 仅在靠近主瓣周围附近分布着一圈均匀的旁瓣, 对波束方向具有较好的指向性。后续的研究将围绕此麦克风阵列展开。

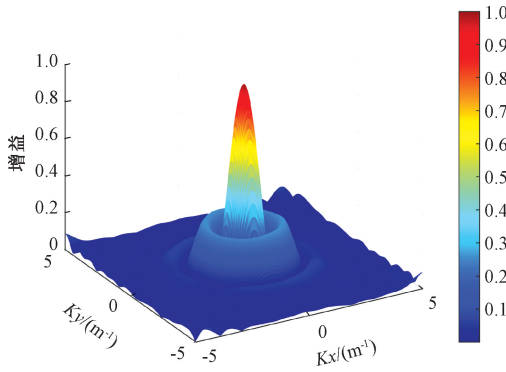


图2 阵列方向图

Fig. 2 Array orientation diagram

1.4 基于二次互相关的可控响应功率相位变换算法

可控响应功率相位变换算法 (SRP-PHAT) 在低信噪比和多传声器阵列的情况下, 定位下性能较差, 且缺乏运算实时性, 难以满足实际噪声源定位系统的实时性需求。而基于倒谱预滤波的广义二次互相关算法较 GCC-PHAT 算法有更高的时延估计精度且在抑制噪声和抗混响方面具有一定优势。基于二次互相关的可控响应功率相位变换算法结合了两种算法的优势, 将两组通道的信号自相关和互相关的结果进行二次互相关运算, 代入 GCC-PHAT 函数进一步计算得到改进的 SRP-PHAT 函数, 对其进行峰值搜索得到声源位置。传统的 SRP-PHAT 算法对麦克风阵列的所有通道两两之间进行互相关运算, 而二次互相关算法的计算量较于原算法略高, 为降低计算复杂度, 将麦克风阵列划分为参考通道和声源通道, 仅在这两组通道间进行组合运算, 大幅减少算法计算量。改进后算法的具体步骤为:

设第 m 个麦克风在 t 时刻的输入信号为 $\mathbf{x}_m(t)$, 将输入信号进行分帧处理, 对每一帧信号使用汉明窗 (Hamming window) 进行加窗处理, 得到帧信号 $\mathbf{x}_{um}(t)$ 。

$$\mathbf{x}_{um}(t) = \mathbf{x}_m(t) \times \text{Wind}(t) \quad (11)$$

其中, $\text{Wind}(t)$ 表示窗函数值, $\mathbf{x}_m(t)$ 为第 m 个麦克风对应的输入信号。分帧加窗可以将长时域信号分成短时域帧, 以在频域上减少频谱泄漏、提高频谱分析的准确性。对每一帧的信号进行傅里叶变换, 得到频域信号 $\mathbf{X}_{um}(f)$ 。

$$\mathbf{X}_{um}(f) = \int_{-\infty}^{\infty} \mathbf{x}_{um}(t) e^{-j2\pi ft} dt \quad (12)$$

将时域信号转换在频域上, 可以得到每一帧信号的频谱信息, 包括频率成分、能量分布等^[19]。

计算声源中心点的声音强度 (通常用 dB 值表示) 和信号的信噪比 (signal-to-noise ratio, SNR) 用于评估输入信号的质量和性能。信号能量和信噪比的计算方法为:

$$E_{db} = 20 \lg \left(\sqrt{\frac{1}{2f_s} \sum_{k=f_1}^{f_2} |\mathbf{X}_{um}(f)|^2} \right) + 120 \quad (13)$$

$$\text{SNR} = 10 \lg \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right) \quad (14)$$

其中, E_{db} 表示信号能量, f_s 为信号的帧长, 即一帧信号的长度。 $\mathbf{X}_{um}(f)$ 表示输入信号的频谱。 SNR 为信噪比, P_{signal} 表示信号功率, P_{noise} 表示噪声功率。

对输入信号保留特定频率范围内的信号:

$$\mathbf{Y}(f) = \begin{cases} \mathbf{X}_{um}(f), & [f_1, f_2] \\ 0, & \text{其他} \end{cases} \quad (15)$$

式 (15) 保留了特定频率范围 $[f_1, f_2]$ 内的信号频谱。将 128 个麦克风分成参考通道和声源通道, 这 2 个通道的信号分别用 $\mathbf{Y}_s(f)$ 和 $\mathbf{Y}_r(f)$ 表示。麦克风参考通道和声源通道的划分如图 3 所示。

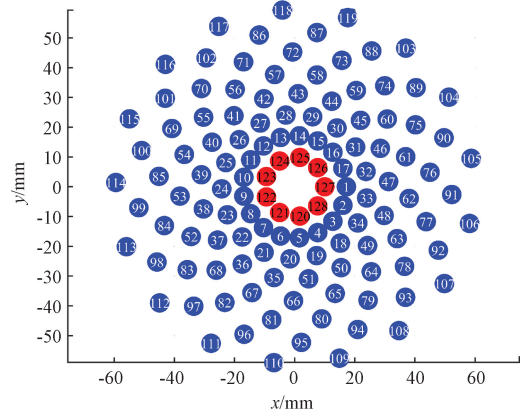


图3 通道分布

Fig. 3 Channel distribution graph

图 3 中, 编号 120~128 为参考通道麦克风通道, 编号 1~119 为声源麦克风通道。

对参考通道的信号进行自相关运算, 即:

$$\mathbf{Y}_{ss}(f) = \mathbf{Y}_s(f) \mathbf{Y}_s^*(f) \quad (16)$$

对参考通道和声源通道进行互相关运算, 即:

$$\mathbf{Y}_{sr}(f) = \mathbf{Y}_s(f) \mathbf{Y}_r^*(f) \quad (17)$$

将两组结果进行互相关运算得到二次互相关函数, 即:

$$\mathbf{Y}_{ssr}(f) = \mathbf{Y}_{sr}(f) \mathbf{Y}_{ss}^*(f) \quad (18)$$

利用二次互相关函数计算改进的 GCC-PHAT 函数, 即:

$$\mathbf{Z}(p, t) = \frac{\mathbf{Y}_{ssr}(f)}{|\mathbf{Y}_{ssr}(f)| + \varepsilon} \times \mathbf{W} \quad (19)$$

其中, ε 为一个很小的数, 防止分母为 0, $\mathbf{W}$ 表示声源通道的权重。将 GCC-PHAT 函数进行傅里叶逆变换得到时域函数, 即:

$$\mathbf{R}(p, t) = \text{ifft}(\mathbf{Z}(p, t)) \quad (20)$$

其中, $\mathbf{R}(p, t)$ 表示第 p 组通道组合在时刻 t 的时域函数。

对时域函数进行加权求和得到改进后的 SRP-PHAT 加权函数, 即:

$$\mathbf{P}_n(l_x, l_y) = \sum_{p=1}^N \mathbf{R}(p, \tau(p, l_x, l_y)) \quad (21)$$

其中, $\tau(p, l_x, l_y)$ 为麦克风组合的时延, l_x 和 l_y 分别表示声源到达角度的取值范围长度。其决定了对声源到

达角度的离散化取样密度,时延的计算方法为:

$$\tau(p) = \frac{\tau_s - \tau_r}{c}$$

(22)

其中, τ_s 和 τ_r 分别表示信号到达参考麦克风通道以及声源麦克风通道的时延,参考麦克风通道 s 和声源麦克风通道 r 构成组合通道 p 。

对 SRP-PHAT 加权函数进行空间谱峰值检测。计算每组通道组合“的 SRP-PHAT 函数”应该是 SRP-PHAT 函数的峰值,找到最大峰值的位置,即为声源的定位角度。

$$[P_{nx}, P_{ny}] = \max(P_n)$$

(23)

2 仿真分析

在 MATLAB 中对改进前后的算法进行仿真,仿真参数如表 2 所示。

表 2 仿真参数

Table 2 Simulation parameters

指标	参数
信号频率	10 kHz
信噪比	12 dB
采样率	96 kHz
帧长	1 024
声源真实位置 (θ, φ)	(45°, 8°)

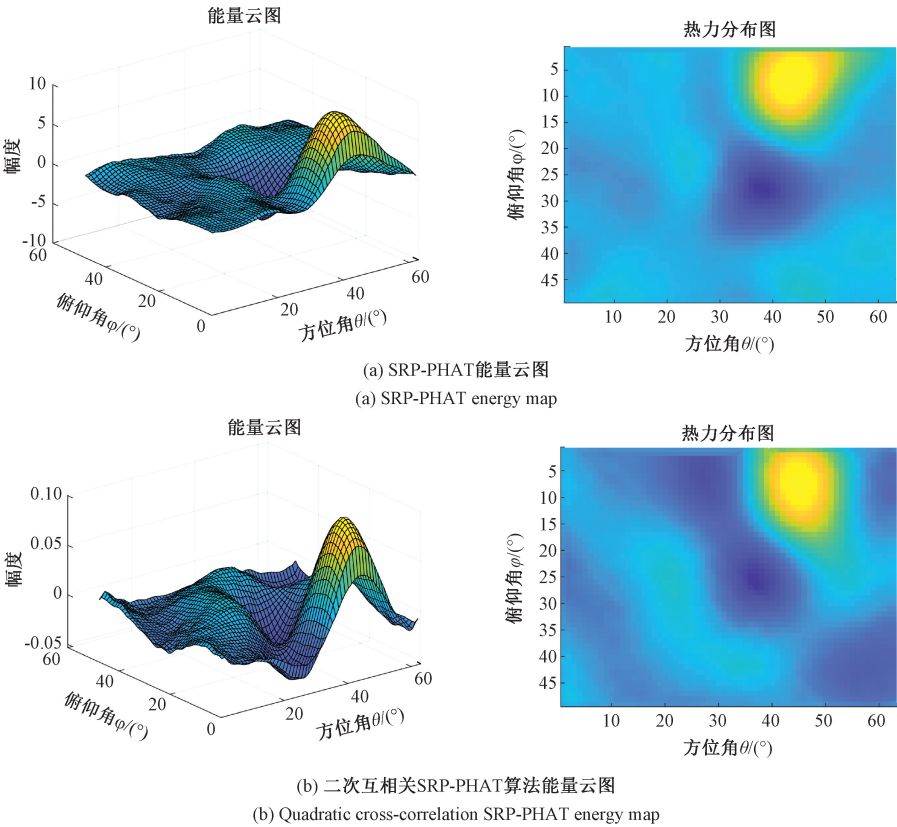


图 5 声源定位图

Fig. 5 Sound source localization map

信号帧长设定为 1 024, 信号被分为 49 帧, 其中第 1 帧信号的波形以及频谱如图 4 所示。

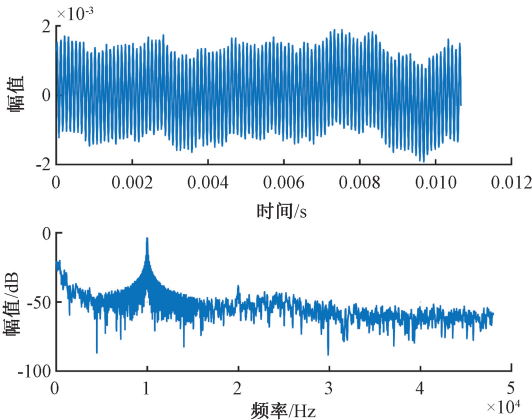


图 4 第 1 帧信号波形

Fig. 4 First frame signal waveform

根据算法原理,计算每一帧信号的能量函数,并对其进行峰值搜索。峰值位置对应声源的位置,在水平方向上对这些峰值进行投影,可得到声源的方位角和俯仰角。为了更好地比较两种算法的定位性能,仿真中对两种算法进行计算时均进行了通道划分处理。图 5 展示了 SRP-PHAT 算法和改进的二次互相关 SRP-PHAT 算法在第 1 帧信号下的能量云图^[19]。

图 5(a) 为传统 SRP-PHAT 算法计算得到的能量云图,图 5(b) 为改进后算法的能量云图。从仿真结果可见,传统的 SRP-PHAT 算法和改进的二次互相关 SRP-PHAT 算法均能有效定位声源位置。改进后的算法能量云图具有更高的主瓣和更窄的主瓣宽度,这表明其定位性能优于传统算法,这一趋势在后续帧中也得到了保持。传统的 SRP-PHAT 算法需要对所有麦克风通道进行两两互相关运算,计算量随麦克风阵列规模呈平方增长,难以满足实时性要求。通道划分策略(参考通道与声源通道的组合)极大地减少了算法运算量。通道划分前后的传统 SRP-PHAT 算法和二次互相关 SRP-PHAT 算法的计算时间(计算一帧云图的平均时间)如表 3 所示。

表 3 计算时间对比
Table 3 Calculate time (s)

算法	未采用通道划分策略	采用通道划分策略
传统 SRP-PHAT	47.78	0.71
二次互相关 SRP-PHAT	51.29	1.32

3 实验验证

实验采用基于瑞芯微 RK3568 平台自主研发的声源定位系统。该系统采用前文介绍的 128 阵元多螺旋臂阵列采集外界音频信号。音频信号采集完成后,通过 FPGA(field-programmable gate array)进行信号采样,采样后的信号会暂时缓存在这里,等待 CPU(central processing unit)发送通信指令后传输至处理器进行定位运算。阵列正中心为采集外界现实图像的摄像头。在声源定位过程中,音频信号经采样后在 CPU 中完成算法运算生成声场能量矩阵,同时摄像头采集的现实图像信号被同步发送至处理器中。GPU(graphics processing unit)将能量矩阵与现实图像信号进行渲染、融合生成声源定位图。图 6 为用于音频信号采集的麦克风阵列。

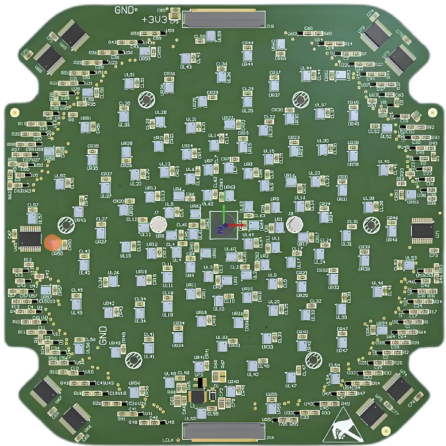


图 6 麦克风阵列

Fig. 6 Microphone array

由摄像头采集的视频信号还会在数据处理模块中进行解码和预处理,同时对麦克风采集的音频信号经 FPGA 采样后进行分帧处理。对每一帧音频信号采用声源定位算法进行声源定位得到声源的相对位置(声源相对麦克风阵列平面的方位角与俯仰角),将方位角与俯仰角转换为平面坐标,最后将麦克风阵列平面坐标系与摄像头的像素坐标系进行映射统一,并将声源定位能量云图与与现实图像进行融合形成声源定位图。声源定位系统整体原理如图 7 所示。

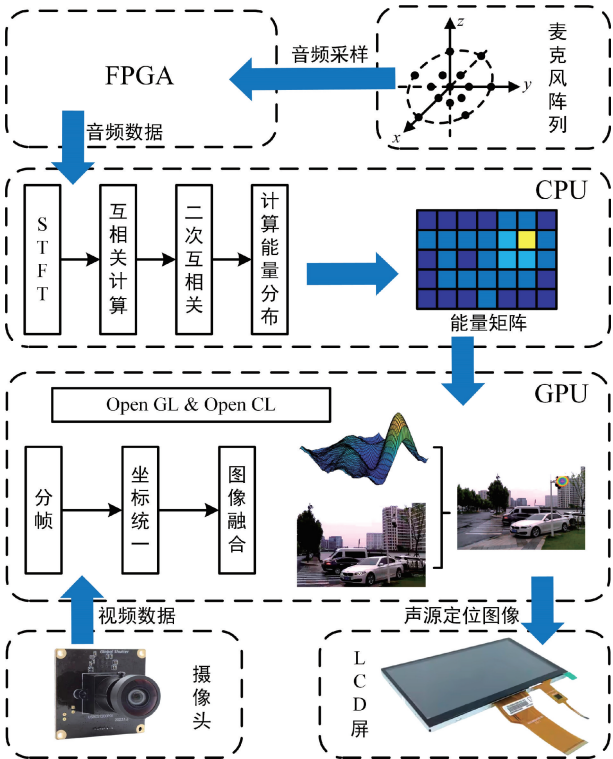


图 7 系统原理

Fig. 7 System schematic diagram

算法的主要计算在 CPU 中完成,对 FPGA 采集的音频信号做快速傅里叶变换转换到频域,对信号进行分帧加窗处理后对每组麦克风通道的信号进行互相关和二次互相关计算得到能量函数,对其进行峰值搜索得到声源位置。声源位置将会以能量分布矩阵的方式传递到 GPU 中与视频帧信号进行图像融合得到最终的声源定位图像,并在 LCD 屏中进行显示。系统定位热力图由能量云图主瓣向下截断 1 dB 能量值得到。

将改进后的算法与传统算法分别在系统中进行移植,并进行声源定位实验,检验两种算法的定位效果。图 8 为实验采用的声源定位系统,该系统集成了麦克风阵列、摄像头以及显示屏,可以将定位结果实时进行输出。



图 8 声学成像系统

Fig. 8 Acoustic imaging system

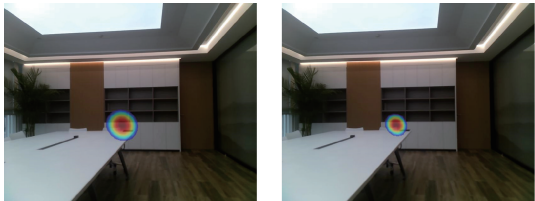
实验声源频率分别为 10、15、20、25 kHz,由扬声器依次播放,实验声源放置在距离采集阵列 5 m 处。实验环境如图 9 所示。



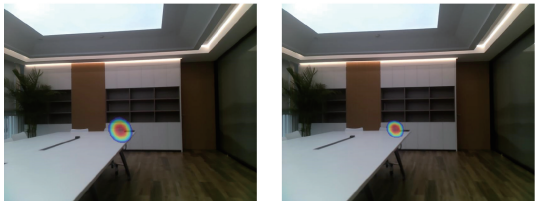
图 9 实验环境

Fig. 9 Experimental environment

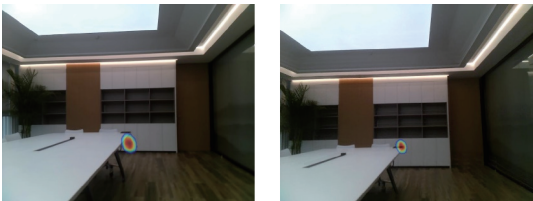
麦克风阵列对扬声器播放的 10、15、20、25 kHz 进行采集,采样率 96 kHz,声源真实位置 $(\theta, \varphi) = (90^\circ, 85^\circ)$ 。阵列中心为采集视频的摄像头,将音频信号分别采用传统的 SRP-PHAT 算法和改进后的算法进行声源定位,定位后的能量云图进行图像处理与摄像头采集的视频帧进行融合生成声源定位点图。实验结果如图 10 所示。



(a) SRP-PHAT(10 kHz) (b) 改进后(10 kHz)
(b) Improved(10 kHz)



(c) SRP-PHAT(15 kHz) (d) 改进后(15 kHz)
(d) Improved(15 kHz)



(e) SRP-PHAT(20 kHz) (f) 改进后(20 kHz)
(f) Improved(20 kHz)



(g) SRP-PHAT(25 kHz) (h) 改进后(25 kHz)
(h) Improved(25 kHz)

图 10 实验结果

Fig. 10 Experimental results

由图 10 可以看出,随着声源频率增大,两种算法的定位精度都在提升,而在所有结果(10、15、20 以及 25 kHz)中改进后的算法定位精度始终高于原算法。可见二次互相关提高了算法的定位精度,达到更好的成像和定位效果。将两种算法对 4 个频段声源进行定位的结果进行统计。表 4 记录了两种算法在 4 种频率声源下的定位结果。

表 4 实验定位结果

Table 4 Localization results

频率/kHz	SRP-PHAT DOA 估计	二次互相关 SRP-PHAT DOA 估计
10	(85°, 88°)	(88°, 86°)
15	(85°, 87°)	(89°, 84°)
20	(88°, 81°)	(90°, 85°)
25	(89°, 87°)	(90°, 84°)

定量评估两种算法在不同频率下的定位性能,对 SRP-PHAT 算法以及改进后的算法在声源定位实验中的角度估计进行统计。此外引入单位空间分辨率指标 (angular resolution width, ARW),该指标定义为主瓣峰值下降 1 dB 处的角度范围,用于表征算法的角度分辨能力,角度误差估计与 ARW 估计如表 5 所示。

对方向角和俯仰角的平均误差和 ARW 指标进行统计,传统的 SRP-PHAT 算法与改进的二次互相关的 SRP-PHAT 算法的平均角度误差和 ARW 柱状分布如图 11 所示。

表 5 角度误差估计与 ARW 统计

Table 5 Angle error estimation and ARW statistics

频率 /kHz	SRP-PHAT DOA 误差	改进后算法 DOA 误差	SRP-PHAT ARW	改进后算法 ARW
10	(5°,3°)	(2°,1°)	7.9°	5.1°
15	(5°,2°)	(1°,1°)	7.5°	4.0°
20	(2°,4°)	(0°,0°)	6.8°	3.2°
25	(1°,2°)	(0°,1°)	5.2°	2.7°

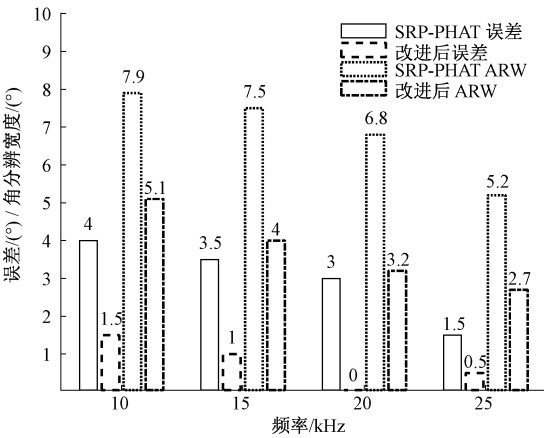


图 11 平均角度误差

Fig. 11 Average angular error

从表 5 和图 11 中可以得出,在对 4 组频率声源进行定位时,改进后的算法的平均角度误差和 ARW 始终低于原算法。

相比于传统的 SRP-PHAT 算法,改进后的算法方向角估计误差平均降低 2.5°,俯仰角估计误差平均降低 2°,空间分辨率指标 ARW 平均降低 3.1°,空间分辨率在不同频段声源定位下平均提升 45.78%。

4 结 论

基于二次互相关的可控响应功率相位变换 (SRP-PHAT) 算法将 SRP-PHAT 算法和基于倒谱预滤波的二次互相关算法在定位精度和抗混响等方面的优势相结合。在计算互相关功率谱时引入二次互相关计算提高算法的定位精度。将用于采集音频信号的麦克风阵列通道分为参考通道和声源通道,并仅在这两组通道之间进行互相关运算,提高了算法的实时性。在 MATLAB 中对两种算法进行了仿真,结果显示,改进算法在声源定位时的定位精度等指标上优于传统的 SRP-PHAT 算法。将原算法与改进后的算法在 128 阵元多螺旋臂麦克风阵列声学成像系统中进行了实验验证,对 10、15、20、25 kHz 的单点声源进行了声源定位成像。实验结果表明,改进后的

算法在定位精度上优于原算法,方向角估计误差平均降低 2.5°,俯仰角估计误差平均降低 2°,空间分辨率指标 ARW 平均降低 3.1°。在不同频段声源定位下,空间分辨率平均提升 45.78%。此外,通过将信号通道进行划分,减少了互相关运算的组数,大大降低了算法的计算量。

参考文献

[1] 陆利冬,曹永刚,郑慧峰,等. 基于球壳聚焦阵列的超声成像检测技术研究[J]. 仪器仪表学报,2019,40(4):61-68.

LU L D, CAO Y G, ZHENG H F, et al. Research on ultrasonic imaging detection technology based on spherical shell focused array [J]. Chinese Journal of Scientific Instrument, 2019, 40(4): 61-68.

[2] 赵慎,李伟,覃业梅,等. 传声器阵列函数反卷积声源成像算法研究[J]. 仪器仪表学报,2023,44(10):112-119.

ZHAO SH, LI W, QIN Y M, et al. Functional deconvolutional approach for the mapping of acoustic sources algorithm of microphone array [J]. Chinese Journal of Scientific Instrument, 2023, 44(10): 112-119.

[3] JEKATERYŃCZUK G, PIOTROWSKI Z. A survey of sound source localization and detection methods and their applications[J]. Sensors, 2023, 24(1): 68.

[4] WANG Y, DENG ZH, ZHAO J X, et al. Progress in beamforming acoustic imaging based on phased microphone arrays: Algorithms and applications [J]. Measurement, 2025, 242: 116100.

[5] TENGAN E, DIETZEN T, ELVANDER F, et al. Multi-source direction-of-arrival estimation using steered response power and group-sparse optimization[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2024, 32: 3517-3531.

[6] 史文旭,陈华伟. 融合共点声强特征的掩蔽加权声源定位方法[J]. 声学学报,2024,49(6):1242-1254.

SHI W X, CHEN H W. A masking weighted sound source localization method integrating the characteristics of coincident point sound intensity [J]. Acta Acustica, 2024, 49(6): 1242-1254.

[7] WANG Z R, WAN W, YE B Y, et al. Tailored for vehicle horn: A novel sound source capture method[J]. Journal of Vibration and Control, 2024: 10775463241283867.

[8] DANG X D, ZHU H Y. An iteratively reweighted steered response power approach to multisource localization using a distributed microphone network[J]. The Journal of the Acoustical Society of America, 2024, 155(2): 1182-1197.

- [9] 刘祥楼, 安梦琪, 刘昭廷. 基于 RLLMP 与倒谱预滤波的二次相关时延估计算法[J]. 自动化与仪器仪表, 2023(5):18-21.
- LIU X L, AN M Q, LIU ZH T. A quadratic correlation delay estimation algorithm based on RLLMP with inverse spectral prefiltering [J]. Automation and Instrumentation, 2023(5):18-21.
- [10] 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, 2010, 18(1): 71-74.
- [11] YIN J, VERHELST M. CNN-based robust sound source localization with SRP-PHAT for the extreme edge[J]. ACM Transactions on Embedded Computing Systems, 2023, 22(3): 1-27.
- [12] LIU H T, ZHANG X L, LI P G, et al. Time delay estimation for sound source localization using CNN-based multi-GCC feature fusion[J]. IEEE Access, 2023, 11: 140789-140800.
- [13] DIBIASE J H. A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays[M]. Brown University, 2000.
- [14] KWON B, PARK Y, PARK Y S. Analysis of the GCC-PHAT technique for multiple sources[C]. ICCAS 2010, 2010: 2070-2073.
- [15] 孙建红, 张涛, 焦琛. 麦克风数量与阵型对声源定位性能的影响[J]. 电子测量与仪器学报, 2019, 33(11): 14-21.
- SUN J H, ZAHNG T, JIAO CH. The impact of microphone count and array configuration on sound source localization performance [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(11): 14-21.
- [16] 王全增, 封皓, 沙洲, 等. 基于 TDOA 的水下泄漏气泡定向方法[J]. 仪器仪表学报, 2024, 45(8): 103-111.
- WANG Q Z, FENG H, SHA ZH, et al. Underwater leakage bubble orientation method based on TDOA[J]. Chinese Journal of Scientific Instrument, 2024, 45(8): 103-111.
- [17] DO H, SILVERMAN H F. SRP-PHAT methods of locating simultaneous multiple talkers using a frame of microphone array data [C]. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 2010: 125-128.

- [18] 刘恺忻, 付进, 邹男, 等. 利用协方差矩阵拟合的阵列孔径扩展方法[J]. 声学学报, 2023, 48(5): 911-919.
- LIU K X, FU J, ZOU N, et al. An array aperture expansion method using covariance matrix fitting [J]. Acta Acustica, 2023, 48(5): 911-919.
- [19] BIESHEUVEL J, TUINSTRA M, DE SANTANA L D, et al. Acoustic lucky imaging for microphone phased arrays[J]. Journal of Sound and Vibration, 2023, 544: 117357.

作者简介



席旭刚 (通信作者), 分别在 1999 年、2002 年、2019 年于杭州电子科技大学获学士、硕士、博士学位, 现为杭州电子科技大学教授, 主要研究方向为信号处理, 人机交互和康复机器人

E-mail: xixugang@hdu.edu.cn

Xi Xugang (Corresponding author) received his B. Sc., M. Sc and Ph. D degrees all from Hangzhou Dianzi University in 1999, 2002 and 2019, respectively. He is currently a professor with Hangzhou Dianzi University, His research interests include signal processing, human-robot interaction, and rehabilitation Robot.



王晨, 2022 年于皖西学院获得学士学位, 现为杭州电子科技大学自动化学院研究生, 主要研究方向为阵列声成像。

E-mail: wangchen6920@163.com

Wang Chen received his B. Sc. degree from West Anhui University in 2022. Now he is a postgraduate student majoring in the School of Automation at Hangzhou Dianzi University. His main research interest is array acoustic imaging.



李文国, 2005 年于杭州电子科技大学获得学士学位, 2009 年于杭州电子科技大学获得硕士学位, 2022 年于杭州电子科技大学获得博士学位, 现为威亨国际(杭州)电气制造有限公司正高级工程师, 主要研究方向为生物电信号处理、多模态信息融合。

E-mail: wgli_hz@163.com

Li Wenguo received his B. Sc. degree from Hangzhou Dianzi University in 2005, received his M. Sc. degree from Hangzhou Dianzi University in 2009, received his Ph. D. degree from Hangzhou Dianzi University in 2022. Now he is a senior technical engineer in Xianheng International (Hangzhou) Electrical Manufacturing Co., Ltd. His main research interests include bioelectrical signal processing and multimodal information fusion.