

基于改进 GCC-PHAT 声源定位方法的研究

刘力,许贵诚,农汉彪[†],黄旭就,陈广桢,曹文明

(百色学院,广西 百色 533000)

摘 要:针对广义互相关-相位变换方法(GCC-PHAT)在低信噪比条件下声源定位误差较大的问题,提出了一种新的改进算法。改进算法通过减弱噪声互谱,强化主成分的方式得到新的加权函数,以便在低信噪比条件下更准确地估计时延差,再通过麦克风阵列的时延差估计出声源的方位。仿真数据实验,验证了改进算法的有效性,并与现有算法进行比较,结果表明在低信噪比条件下改进算法具有更高的准确性,在不同角度的识别结果中,改进算法的识别平均绝对角度误差约为现有方法的 68%。

关键词:声源定位;麦克风阵列;GCC-PHAT;时延估计

中图分类号:TP206.1

文献标识码:A

Research on Improved GCC-PHAT Sound Source Localization Method

LIU Li, XU Guicheng, NONG Hanbiao[†], HUANG Xujiu, CHEN Guangzhen, CAO Wenming

(Baise University, Baise, Guangxi 533000, China)

Abstract: A new improved algorithm is proposed to address the problem of significant localization errors in the generalized cross-correlation phase transformation method (GCC-PHAT) under low SNR conditions. The improved algorithm obtains a new weighting function by weakening the noise cross-spectrum and strengthening the principal components, in order to more accurately estimate the time delay difference under low SNR conditions, and then estimate the direction of the sound source through the time delay difference displayed in the microphone. Simulation data experiments have verified the effectiveness of the improved algorithm and compared it with existing algorithms. The results show that the improved algorithm has higher accuracy under low SNR conditions. In recognition results from different angles, the average absolute angle error of the improved algorithm is about 68% of the existing methods.

Key words: sound source localization; microphone array; GCC-PHAT; time-delay estimation

在课堂直播或远程会议自动化过程中,摄像设备需要自动准确地识别发言人具体方位和距离,以便及时自动转向和调整焦距,保证摄制画面的准确性。发言人的位置识别,即声源定位,目前主要基于麦克风阵列实现,方法大致可分为三类:基于波束形成技术^[1]、基于谱图估计技术^[2]和基于时延估

计(Time-delay Estimation, TDE)方法^[3]。对于便携设备来说,由于算力有限,更多会采用基于 TDE 的算法。2017 年李猛采用时延估计精度较高的相位变换法(GCC-PHAT)来进行时延估计并对此法进行了改进,使时延估计的效果得到一定的改善^[4]。2019 年,唐浩洋等通过对不同信噪比赋予

收稿日期:2024-04-24

基金项目:广西高校中青年教师科研基础能力提升项目(2022KY0735);百色学院 2022 年广西壮族自治区级大学生创新创业训练计划项目(202210609015)

作者简介:刘力(2002—),男,重庆奉节人,本科生,研究方向:通信信号处理。

[†]通信联系人,E-mail: nonghanbiao@126.com

不同的权重,从而构造加权函数,该算法在信噪比为 0dB 以上时能有效地抑制噪声,降低时延误差^[5]。2022 年 Wang 等为了提高低信噪比条件下时延估计的精度,提出了一种基于奇异值分解和改进 GCC-PHAT 加权的时延估计优化算法 MLPT- $\rho\gamma$ 加权函数^[6]。但由于环境噪声的存在对时延估计精度有很大的影响,进而影响声源目标定位的准确性。现有方法对于高噪声情况下声源识别准确性不高。针对这一问题,在分析和比较 GCC-PHAT 算法的基础上提出了基于广义互相关法的改进算法。

1 算法原理

广义互相关法是一种基于相关分析的时延估计法,它是根据信号的相关性进行时间延迟估计^[7]。传统的广义互相关函数法(GCC)^[8]是根据两个信号时域上的互相关函数的峰值进行估计的,找出峰值所在的时间轴位置,即为时延值,为了提高估计值的准确性,可以对接收到的声音信号进行滤波和白化处理^[9],增强信噪比较高的部分,抑制噪声功率,使互相关函数取得峰值,再根据峰值查找其所在的时间轴位置,该时间值就是两个信号的相对时延。

声源信号用 $s(t)$ 表示,假设有两个麦克风,它们接收到的信号分别以 $x_1(t)$ 和 $x_2(t)$ 表示, $n_1(t)$ 和 $n_2(t)$ 表示白噪声, Δ 表示信号从声源处传播到两个麦克风阵列的传播时间,这两个麦克风之间的时间延迟为 τ , $x_1(t)$ 和 $x_2(t)$ 可以表示为:

$$x_1(t) = s(t) + n_1(t) \quad (1)$$

$$x_2(t) = s(t - \Delta) + n_2(t) \quad (2)$$

$x_1(t)$ 和 $x_2(t)$ 的互相关函数可以表示为:

$$R_{12}(\tau) = E[x_1(t)x_2(t - \tau)] \quad (3)$$

假设两个麦克风接收到的语音信号互不相关,语音信号和噪声也互不相关,则 $x_1(t)$ 和 $x_2(t)$ 的互相关函数可以变化成:

$$R_{12}(\tau) = \int_0^\pi G_{12}(\omega) e^{-j\omega\tau} d\omega \quad (4)$$

其中, $G_{12}(\omega)$ 是声源信号与时延后的信号的互相关功率谱,由公式(5)给出。

$$G_{12} = X_1(\omega)X_2^*(\omega) \quad (5)$$

其中, $X_1(\omega)$ 为 $x_1(t)$ 傅里叶变换, $X_2^*(\omega)$ 为 $x_2(t)$ 傅里叶变换后的共轭

使用 GCC 方法已经可以得到时延值,但考虑到实际的信号的互相关函数的峰值不够明显,容易

在判断峰值的时候造成误差^[10]。为了得到具有更陡峭极值的互相关函数,一般在频域使用加权函数来白化输入信号,即 GCC-PHAT 算法。此时的加权函数为:

$$A(\omega) = \frac{1}{|G_{12}|} \quad (6)$$

理想加权输出结果为:

$$R_{12}(\tau) = a \cdot \delta(\tau - \Delta) \quad (7)$$

Δ 表示信号从声源处传播到两个麦克风的传播时间, τ 表示两个麦克风之间的时间延迟, $\delta(\tau - \Delta)$ 表示延迟 Δ 后的冲激函数。此时, R_{12} 的峰值越大信号越可靠,质量越高。

GCC-PHAT 算法通过舍弃频域幅度,得到在时域的冲激函数,从而可以较好地确定时延值,但是在低信噪比时性能大为减弱。

为了减小噪声影响,经过分析比较,将加权函数改为:

$$H(\omega) = \left| \frac{G_{12}(\omega)G_{11}^{1/2}(\omega)}{G_{22}^2(\omega) - G_{12}^2(\omega)} \right| \quad (8)$$

其中, $G_{11}(\omega)$ 、 $G_{22}(\omega)$ 是声源信号和时延后的信号自相关功率谱。理想加权输出结果为:

$$R_{12}(\tau) = \int_0^\pi H(\omega)G_{12}e^{-j\omega\tau} d\omega \quad (9)$$

通过上述算法得到时延,根据时延和声速的关系,计算声源相对于麦克风阵列的角度 θ_i :

$$\theta_i = \arccos\left(\frac{\tau \cdot c}{R \cdot 2}\right) \cdot \frac{180}{\pi} \pm \varphi_i \quad (i = 1, 2, 3) \quad (10)$$

其中, θ_i 表示声源由观测点相对于麦克风阵列正方向的夹角, τ 表示时延, R 表示观测点之间的距离。 φ_i 为所需的相位叠加角,方向由麦克风相对于声源的声程差决定。采用欧氏距离公式计算,使用任意两个麦克风的位置坐标和声源的位置坐标,可以算出声源到每个麦克风的声程 d_i 。即:

$$d_i = \sqrt{(x_i - s_x)^2 + (y_i - s_y)^2} \quad (11)$$

其中, (s_x, s_y) 为声源坐标, (x_i, y_i) 为麦克风 M_i 坐标。声源到两个观测麦克风之间的传声时差 Δ_i 可以表示为:

$$\Delta_i = \frac{|(d_i - d_{i+3})|}{c}, (i = 1, 2, 3) \quad (12)$$

其中, c 为声速, d_i 为麦克风 M_i 到声源的声程。

当遇到复数时需将实部与虚部相加,并加上对应的相位叠加角即可得到此时相对于麦克风阵列正方向的夹角。以 $x_1(t)$, $x_2(t)$ 为例,本文算法流程如下:

Input: $x_1(t)$
 $x_2(t)$
 $s(t)$
 (s_x, s_y)
 (x_1, y_1)
 (x_2, y_2)
利用公式(5)得到 $G_{11}(\omega)$
 $G_{22}(\omega)$
 $G_{12}(\omega)$ 。
利用公式(8)得到本文算法加权函数 $H(\omega)$ 。
利用公式(9)输出估计时延 τ 。
利用公式(10)输出估计角度 θ 。
利用公式(11)得到 d_1
和 d_2
if: $d_1 > d_2$
 $\theta = \theta$
endif
elseif: $d_1 < d_2$
 $\theta = 180 - \theta$
endif
Output: θ 。

2 实验数据分析

2.1 实验仿真模型

实验模型为一个水平放置的麦克风阵列,阵列由六个麦克风均匀分布在半径为 $R=40\text{ mm}$ 同平面圆周上,如图 1 所示。 $M_1 \sim M_6$ 为阵列麦克风观测点,声源位置 S 位于麦克风阵列所在的平面内。假设测试系统阵列内所有麦克风进行同步采样,采样频率为 51.2 kHz 。设阵列中心 O 与麦克风 M_1 的连线 $\overrightarrow{OP}$ 为模型系统坐标的正方向,即声源 S 方向位置由夹角 $\angle SOP$ 确定。

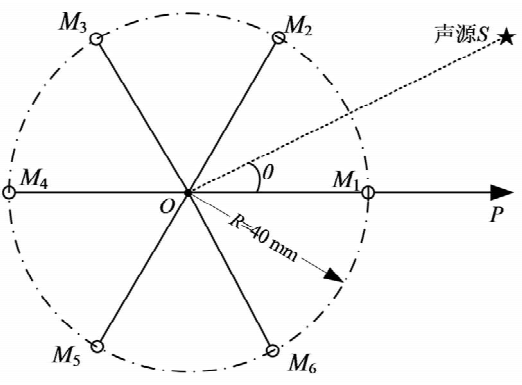


图 1 仿真系统模型图

2.2 仿真数据与结果

为了验证本文所述的识别方法,首先对麦克风接收到的声音信号进行预处理,即将接收到的信号通过一个截止频率为 8000 Hz 的十阶巴特沃斯滤

波器。假设声源位置距圆点的距离为 3540 mm 且与圆阵正方向的夹角为 60° ;观测点采用 M_1 和 M_4 ,则声源位置到 M_1 、 M_4 之间的时延差为 0.011761 s 。声源信号采用英文单词“Hello”的发音,波形图如图 2 所示,同时为了模拟实际环境,对声源信号添加了与声源的信噪比为 30 dB 的加性高斯白噪声,其波形如图 3 所示。

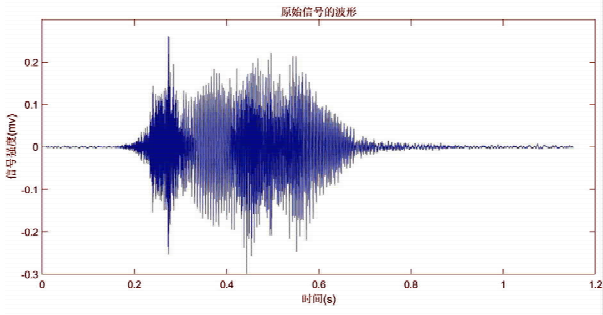


图 2 原始信号的波形

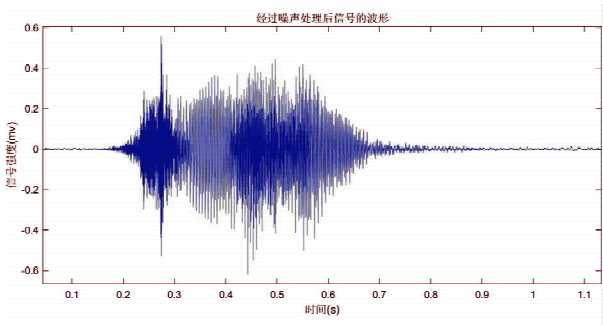


图 3 含噪信号的波形

为了比较本文算法与现有算法 GCC-PHAT 之间的性能,具体实验流程如图 4 所示。

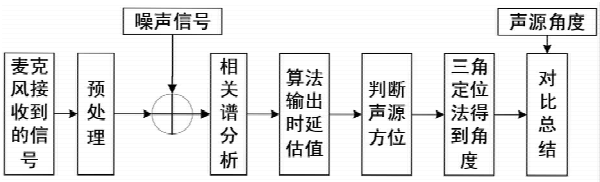


图 4 实验流程图

当声源位置不在 M_1 、 M_4 连线的中线上,则观测点 M_1 、 M_4 收到的声音信号存在一定时延,如图 5 所示。

由式(6)~式(9),可以得出此时的观测点对应的时延估计值。图 6 为 GCC-PHAT 算法所得到的相关图谱,其时延估计为 0.0117791 s ,可以观察到由于经过预处理的原因,在零点处有一定的冲激。图 7 为本文算法得到的相关图谱,其时延估计为 0.011773 s ,可以观察到新的加权函数有效抑制

了预处理的影响。通过对比图谱,改进的算法通过抑制预处理的影响,提高了时延精度。

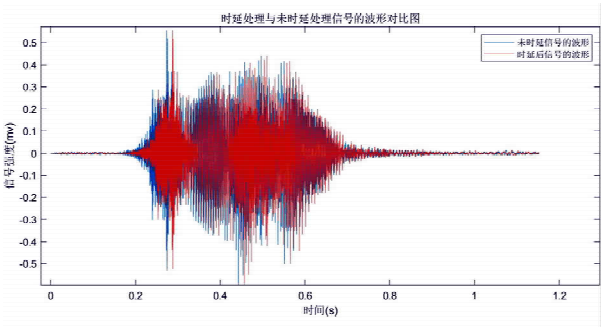


图 5 时延处理与未时延处理信号的波形对比图

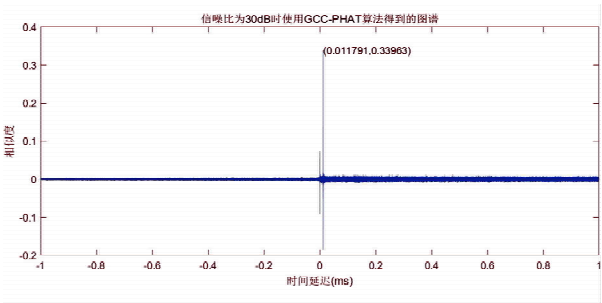


图 6 信噪比为 30 dB 时使用现有算法得到的图谱

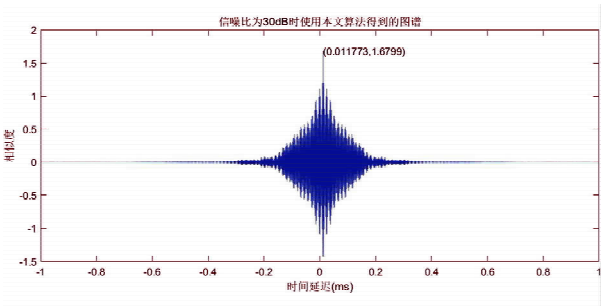


图 7 信噪比为 30 dB 时使用本文算法得到的图谱

当信噪比为−5 dB 时,两种方法所得到的相关图谱如图 8、图 9 所示。

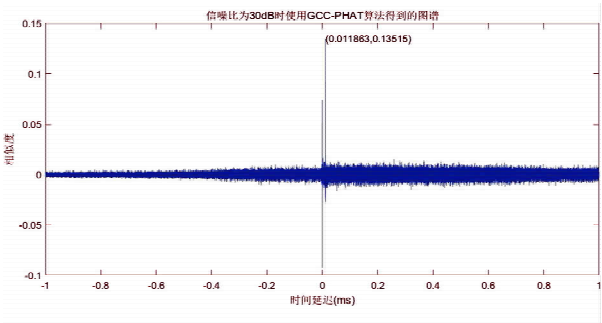


图 8 信噪比为−5 dB 时使用现有算法得到的图谱

图 8 为 GCC-PHAT 算法所得到的相关图谱,其时延估计为 0.011863 s,可以观察到由于预处理的原因和信噪比降低,图谱中出现了伪峰值,这对时延精度会产生一定的影响。图 9 为本文算法得到的相关图谱,其时延估计为 0.011773 s,可以看出本文算法在不同信噪比条件下,时延精度较稳定,表明本文算法普适性强。同理,可以计算其余两组观测点在信噪比为−5 dB 时,相对于声源位置的时延,结果如表 1 所示。从数值上看,本文方法的结果误差相比 GCC-PHAT 方法更小。

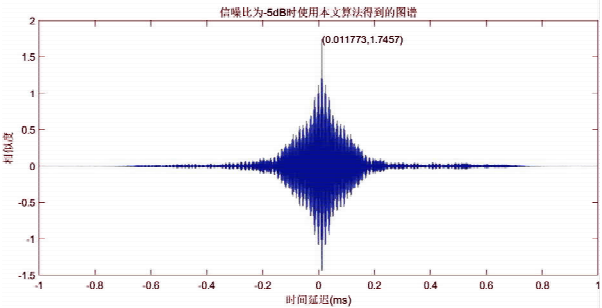


图 9 信噪比为−5 dB 时使用本文算法得到的图谱

项目	时延/s		
	M_1,M_4	M_2,M_5	M_3,M_6
理论时延	0.011704	0.023329	0.011704
GCC-PHAT	0.011863	0.023476	0.011863
与理论误差	0.000159	0.000147	0.000159
本文算法	0.011773	0.023419	0.011773
与理论误差	0.000069	0.000009	0.000069

根据声源到观测点的距离差异,确定声源所在的象限,再由公式(10)计算出声源相对于正方向的夹角。三组观测角度计算结果如表 2 所示。

观测点	理论角度	信噪比为−5dB 相对于正方向的夹角	
		GCC-PHAT	本文方法
M1,M4	60°	59.3430°	59.6840°
M2,M5	60°	56.9078°	57.8087°
M3,M6	60°	60.5493°	60.3159°
均值	60°	58.9334°	59.2695°
绝对误差	0°	1.0666°	0.7305°

从表中数值看出,本文方法所得到的角度与真实角度的差值更小。另外,为了验证识别方法对于不同角度的普适性,在圆周角上随机生成 7 个不同

角度的模拟声源,在信噪比为-5 dB 条件下,采用两种不同方法对每个声源方向角进行计算,得到结果与理论角度的绝对误差如图 10 所示。

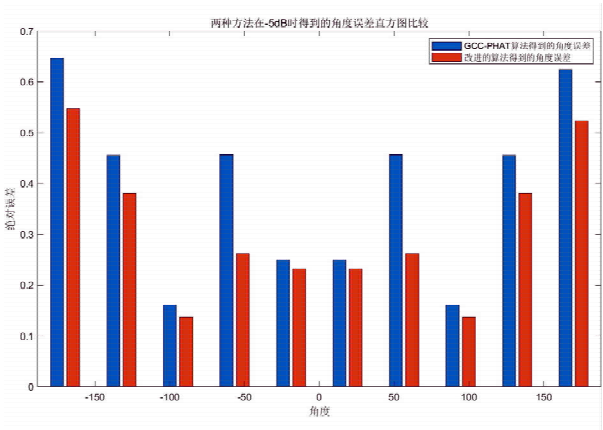


图 10 不同方法结果对比图

通过直方图可以看出,本文算法得到的角度误差小于 GCC-PHAT 算法的误差,表明该算法在信噪比较低的情况下,比现有方法识别性能更好。在不同角度的声源识别过程中,改进方法的平均绝对误差约为现有方法平均绝对误差的 68%。

3 结 论

通过对仿真数据实验结果的分析 and 对比,验证了本文算法在噪声环境下对于声源定位的准确度高于现有的 GCC-PHAT 方法。即表明,该算法在声源定位任务中表现出色,对于实际应用中的声源定位和时延估计具有重要意义,为改进传感器网络和声学信号处理提供了有力支持。未来的研究方向可能包括进一步优化算法参数,探索更广泛的应用场景,并结合实际工程问题进行验证,以进一步提高声源定位系统的性能和稳定性。

参考文献

[1] 潘超,黄公平,陈景东. 面向语音通信与交互的麦克风阵列波束形成方法[J]. 信号处理,2020,36(6):804—815.

[2] 徐进,赵益波,郭业才. 一种新的麦克风阵列自适应语音增强方法[J]. 应用科学学报, 2015, 33 (2): 187—193.

[3] 余萍,杨乘,王紫薇,等. 基于联合加权的广义二次相关时延估计算法[J]. 计算机仿真,2023, 40(3):400—404.

[4] 李猛. 基于改进 GCC-PHAT 的麦克风阵列声源定位技术[D]. 阜新:辽宁工程技术大学,2017.

[5] 唐浩洋,陈子为,黄维. 基于麦克风阵的 GCC 时延估计算法分析[J]. 计算机系统应用,2019,22(12):140—145.

[6] WANG S Z, LI Z J, WANG P B, et al. Optimization algorithm for delay estimation based on singular value decomposition and improved GCC-PHAT weighting [J]. Sensors. 2022; 22(19):7254.

[7] 行鸿彦,唐娟. 时延估计方法的分析[J]. 声学技术,2008,27 (1):110—114.

[8] 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.

[9] SILVERMAN H F, PATTERSON W R. Visualizing the performance of large-aperture microphone arrays[C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99, 1999(2), 969—972.

[10] 茅惠达,张玲华. 声源定位中广义互相关时延估计算法的研究[J]. 计算机工程与应用,2016,52(22):138—142.