文章编号:1006-0464(2016)06-0557-06

基于 GCC-PHAT 的改进近场声源定位算法

梁　音[a,b],张　华[a],陈利民[a,b*],龙学焜[b],王　伟[b],肖中奇[b]

(南昌大学 a.机电学院;b.信息工程学院,江西 南昌　330031)

摘　要:结合 GCC-PHAT 与 SRP 搜索的算法被广泛应用于声源定位系统,但其定位结果随信噪比降低而恶化严重,且难以估计窄带声源。针对这些问题,提出一种用于近场模型的改进算法。首先计算阵列信号求和功率谱的峰均比,再经对数变换和设置噪声抑制阈值等步骤,得到反映频点信噪比的加权因子,将其应用在 GCC-PHAT 中,并进一步优化 SRP 搜索策略。大量仿真结果表明:改进的算法具有强的抗噪能力,且对宽带和窄带声源信号均有良好的估计效果,很适合实时处理。

关键词:声源定位;近场;时延估计;可控功率响应;对数峰均比;宽带－窄带

中图分类号:TN912　　**文献标志码**:A

Improved acoustic source localization algorithm based on GCC-PHAT in near-field situation

LIANG Yin[a,b],ZHANG Hua[a],CHEN Limin[a,b*],LONG Xuekun[b],WANG Wei[b],XIAO Zhongqi[b]

(a. School of Mechatronics Engineering,Nanchang University,Nanchang,Jiangxi 330031,China;
b. School of Information Engineering,Nanchang University,Nanchang,Jiangxi 330031,China)

Abstract:GCC-PHAT combined with SRP searching algorithm had proved to be ineffective on narrowband sound source and the localization results deteriorated severely with the decreasing of SNR. To solve these problems,proposed an improved algorithm for near field model which introduced a weighting factor by calculating the PAPR of received signals in the frequency domain,and then make logarithmic transformation to compress the dynamic range,as well as the establishment of a threshold to suppress noise,etc.,and further optimized the sound source position searching strategy. A lot of simulation results showed that the improved algorithm has strong ability to resist noise and have good estimation effects on both broadband and narrowband sound sources,it's very suitable for real-time processing.

Key words:acoustic source localization;near-field;Time Delay Estimation (TDE);Steered Response Power (SRP);logarithmic PAPR;broadband-narrowband

麦克风阵列是声源定位的基础。根据传播距离的远近,声波传播到各麦克风的扩散形式有近场和远场两种。对大部分的智能设备如服务机器人、电脑等,其使用场景一般都在室内,声波传播属于近场传播。在同一时间,空间中可能存在不止一个声源,这些声源可以由人发出,也可以由机器产生。声源定位技术可以用来区分出不同声源在空间中所处的位置,进而对目标声源做相应的处理,如说话人跟踪、语音增强等[1－2]。

现有的声源定位算法可大致分为三类:a)基于时延估计(Time Delay Estimation,TDE)的算法;b)基于高分辨率谱估计的算法;c)基于稀疏表示的算法。其中,基于 TDE 的算法核心在于对传播时延的准确估计,一般通过对麦克风间信号做互相关处理得到。进一步获得声源位置信息,可以通过简单的延时求和[3]、几何计算[4]或是直接利用互相关结果进行可控功率响应搜索[5－9]等方法。这类算法实

收稿日期:2016-06-26。
基金项目:国家自然科学基金资助项目(51165033);江西省自然科学基金资助项目(20151BAB207052);江西省科技支撑计划项目(20142BBE50035)。
作者简介:梁音(1971－),女,讲师。 * 通信作者:陈利民(1974－),副教授。E-mail:robinkk4@gmail.com。

现相对简单,运算量小,便于实时处理,因此在实际中运用最广。基于高分辨率谱估计的算法最初被应用于窄带源的定位,后来逐渐被众多学者变换引用到宽带源定位问题中。拓展到宽带信号估计时,需要在频域将信号频率划分为多个子带,或者进行频率聚焦以转化为窄带信号处理的方式[10-11],其中代表性的如多重信号分类(Multiple Signal Classification,MUSIC)算法[11-12]。该类算法定位分辨率很高,但由于要进行宽带到窄带的转化,使得算法运算量大大增加,实际中更是因为声源个数未知以及噪声环境不满足理想的高斯白噪声条件而性能急剧下降。基于稀疏表示的算法是近年来的研究热门,它主要利用声源在空域的稀疏性来构造约束方程,通过约束范数最小来求解最优解。它的定位效果较好,且不受欠定问题限制,但运算量太大,目前难于实际应用[13-14]。

本文集中讨论近场情况下的声源定位技术,包括来波方向(Direction of Arrival,DOA)估计与距离估计。首先研究了利用相位变换的广义互相关(Generalized Cross Correlation-Phase Transform,GCC-PHAT)算法得到麦克风间相对传播时延信息,以及结合可控功率响应(Steered Response Power,SRP)搜索得到声源位置信息的原理,针对该算法噪声鲁棒性差以及难以定位窄带声源的缺陷,提出了一种对数峰均比加权的改进算法,将其应用于近场的声源定位问题中。仿真实验基于不同类型声源和高低信噪比,对比了改进算法与原算法的定位精度,证实了改进算法的有效性。

1 信号模型

远近场的划分可大致根据公式 $d<2D^2/\lambda$ 来判断,D 为阵列孔径,λ 为声波波长。图 1 显示了三维空间中,声音从声源传播到麦克风阵列的近场传播模型,这时声波从声源出发,以球面波形式扩散到各麦克风。

设 m 个麦克风组成的麦克风阵列分布在三维空间中,以麦克风 M_0 为参考阵元并设该点坐标为原点,各麦克风坐标为 $\vec{P}_{M_i}=(x_{M_i},y_{M_i},z_{M_i})$,$i=0,1,\cdots,m-1$。某一时间,空间中存在 n 个声源,它们的坐标为 $\vec{P}_{S_q}=(x_{S_q},y_{S_q},z_{S_q})$,$q=0,1,\cdots,n-1$,此时麦克风 i 接收到的信号可以表示为:

$$x_i(t)=\sum_{q=0}^{n-1}\alpha_{qi}\cdot s(t-\tau_{qi})+n_i(t),i=0,1,\cdots,m-1 \tag{1}$$

其中 α 为幅度衰减因子,τ_{qi} 表示第 q 个声源到麦克风 i 的传播延迟,$n_i(t)$ 为非相干噪声。τ_{qi} 可根据(2)式来计算:

$$\tau_{qi}=d_{qi}/c=|\vec{P}_{S_q}-\vec{P}_{M_i}|/c \tag{2}$$

d_{qi} 为声源 q 到麦克风的距离,c 为空气中的声速。

声源的定位过程中一般以极坐标系为参考对空域进行搜索,如声源 q 的极坐标为 (θ_q,ϕ_q,d_{q0}),极坐标与笛卡尔坐标可以按(4)式进行转换:

$$(x_{S_q},y_{S_q},z_{S_q})=d_{q0}\cdot(\sin\phi_q\cos\theta_q,\sin\phi_q\sin\theta_q,\cos\phi_q) \tag{3}$$

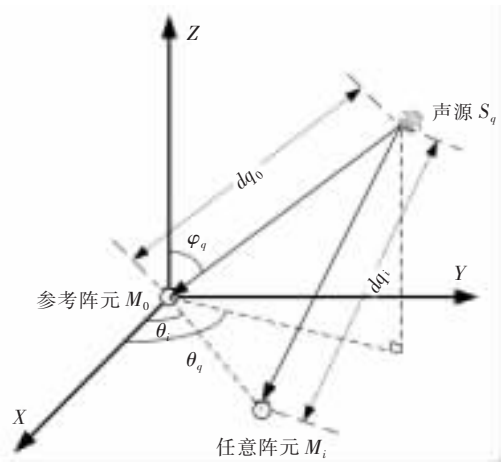


图 1 声波近场传播示意图

2 基本原理

实际做信号处理时均是在离散数字信号上进行,根据语音信号的短时平稳特性,一般以数据帧为单位进行处理。设信号的采样速率为 Fs,每帧包含 L 个采样数据,FFT 点数等于帧长。

若麦克风在三维空间呈立体结构分布,则用一组某声源到麦克风阵列的传播延迟向量 $\Gamma=[\tau_0,\tau_1,\cdots,\tau_{m-1}]$,能够唯一地确定该声源在空间中的位置。在只有阵列接收信号而声源信息未知的情况下,无法直接得到绝对传播时延 τ_i,可以通过计算两两麦克风之间的相对传播时延 $\Delta\tau_{ij}$ 来替代,$i=0,1,\cdots,m-1;j=0,1,\cdots,i-1$。

$$\Delta\tau_{ij}=Fs\cdot(|\vec{P}_{S_q}-\vec{P}_{M_i}|-|\vec{P}_{S_q}-\vec{P}_{M_j}|)/c \tag{4}$$

2.1 GCC-PHAT 计算传播时延

互相关算法常被用来做时延估计,表示为:

$$\Delta\hat{\tau}\tau_{ij}=\arg\max_{\tau}C_{ij}(\tau) \tag{5}$$

C_{ij} 为通道 i 与通道 j 的信号在同一帧内的互相

关,而直接在时域对信号做互相关的计算量很大,根据维纳－辛钦定理,时域互相关等价于频域互功率谱的傅里叶逆变换,由此推导出广义互相关的方法。两个通道采样信号的 GCC 计算式定义为:

$$C_{ij}^{(g)}(\tau)=\sum_{k=0}^{L-1}X_i(k)\cdot X_j^{*}(k)\cdot e^{\frac{j2\pi k\tau}{L}} \tag{6}$$

$X_i(k)$ 为通道 i 的接收信号在频域的表示。在声源定位问题中,通常对上式进行相位变换得到 GCC－PHAT:

$$C_{ij}^{(g-p)}(\tau)=\sum_{k=0}^{L-1}\frac{X_i(k)\cdot X_j^{*}(k)}{|X_i(k)|\cdot|X_j(k)|}\cdot e^{\frac{j2\pi k\tau}{L}} \tag{7}$$

上式去掉了互功率谱的幅度信息而只保留了相位信息。这样做有两个好处,一是压缩了互相关结果的动态范围,在多个声源存在时不易发生高能量信号对低能量信号的掩蔽;二是能使得互相关的峰值更加尖锐。

2.2　SRP 搜索声源位置

依据系统对估计精度要求,声源所有可能位置可简化为测量空间内三维网格的格点。每空间格点根据式(4)计算得相对延时向量 $\Gamma=[\Delta\tau_{01},\Delta\tau_{02},\cdots,\Delta\tau_{0m-1},\cdots,\Delta\tau_{m-2m-1}]$。理论上,将 Γ 代入(7)并求和,其结果作为每个格点的能量输出,若某处存在声源,该距离最近的格点处必会形成一个能量峰值。如此遍历整个空间格点,对峰值进行搜索即可得到声源位置:

$$(\hat{\theta},\hat{\phi},\hat{d})=\arg\max_{(\theta,\phi,d)}\sum_{i=0}^{M}\sum_{j=1}^{i-1}C_{ij}^{(g-p)}[\Delta\tau_{ij}(\theta,\phi,d)] \tag{8}$$

3　改进的算法

观察式(7)可发现,由于完全去除了频域的幅度信息,使得噪声频点获得了与有效频点相同的权重,正确的相关峰值容易被噪声淹没。对于窄带声源,其频点分布很少,被其它声源或噪声所掩蔽也就变得更加容易。

3.1　对数峰均比加权的 GCC-PHAT

为改善上述 GCC-PHAT 的缺陷同时又能保留其固有优势,我们提出了下面几点改进思路:

(1) 频点幅值一定程度上反映了信噪比的大小,适当保留其有关信息可以提高噪声鲁棒性;

(2) 窄带声源因频点分布少而难以检测,可以对这些频点进行特殊加强来增强它对结果的影响;

(3) 适当缩小不同频点间权重的差别,避免破坏白化带来的优势。

首先对各麦克风信号的功率谱相加得求和功率谱:

$$A_{sum}(k)=\sum_{i=1}^{m-1}|X_i(k)| \tag{9}$$

以此抑制非相干噪声,增强有效频点。计算求和功率谱的峰均比:

$$P(k)=\frac{[A_{sum}(k)]^2}{\sum_{k=1}^{L}[A_{sum}(k)]^2/L} \tag{10}$$

利用上式计算的结果可能会出现部分频点过大而掩蔽其它频点的情况,因此将其变换到对数域,这能压缩频点间的差异但同时又不影响相互间的大小关系:

$$P_{\log}(k)=\log_{10}P(k) \tag{11}$$

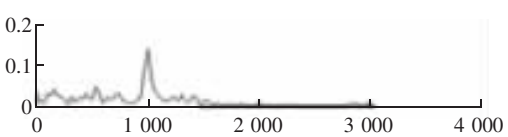
这就得到了整个麦克风阵列在频域的联合对数峰均比,其每个频点的大小和该频点处的信噪比有着密切关联。最终利用它构建出如下的加权系数:

$$w(k)=\begin{cases}1 & P_{\log}(k)<a\\ b\cdot[P_{\log}(k)-a]+c & P_{\log(k)}\geqslant a\end{cases} \tag{12}$$

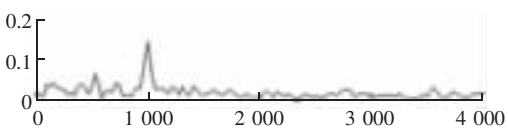
a 是为抑制噪声而设置的门限,对数峰均比小于这一值时表明该对应频点能量不突出,判为噪声。b 是增益系数,它配合着偏置系数 c 来取值,$c\geqslant 1$,它们用于调节有效频点间加权系数的差异。为避免 GCC 结果的动态范围过大,在使用 $w(k)$ 前最好对其做归一化处理。基于对数峰均比加权的 GCC-PHAT 可最终表示为:

$$C_{ij}^{(W)}(\tau)=\sum_{k=0}^{L-1}\frac{w(k)\cdot X_i(k)\cdot X_j^{*}(k)}{|X_i(k)|\cdot|X_j(k)|}\cdot e^{\frac{j2\pi k\tau}{L}} \tag{13}$$

图 2 显示了信噪比为 0 dB 时,宽窄带混合声源下 $w(k)$ 的获取过程。麦克风阵列取均匀圆阵,阵元个数为 20。共有 3 个声源,其中一个窄带单音位于 1 KHz 频点处,2 个宽带声源取纯净说话人语音,单音幅度取为宽带声源最大幅度的 0.5 倍。从图中看到,



(a) 干净幅度谱



(b) SNR = 0 dB 带噪幅度谱

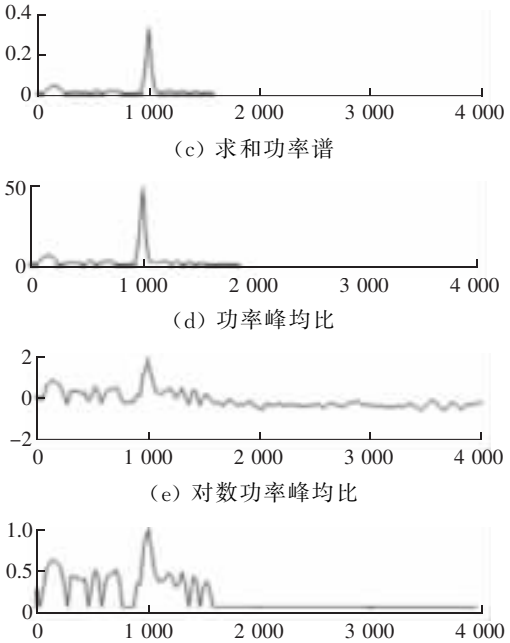


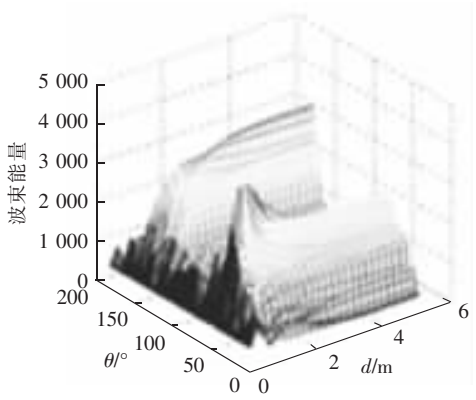
图 2 加权系数 $w(k)$ 的获取过程

最终得到的加权系数中,有效声源频点有更大的权值,从而抑制了噪声干扰,且单音频点也因能量更为集中而突显出来。

3.2 近场搜索策略

为简化讨论,下面假设声源和麦克风阵列均处于同一平面,即麦克风阵列为平面阵,声源的定位问题限于对方位角 θ 和距离 d 的估计。

图3展示了相同外部环境下,用GCC-PATH(上)和对数峰均比加权的 GCC-PHAT(下) 对方位角和距离进行全局 SRP 遍历搜索的能量输出,左右分别为同一次搜索的三维网格图和二维网格图。空间中同时存在 3 个声源,分别位于 $\theta=[40,90,140]$,$d=[0.7,2,3.5]$ 米处。从搜索结果看,某一方位角上的声源将会在同一方位角的不同距离上产生很长的波峰纵伸,但最大的峰值还是位于声源处。距离麦克风



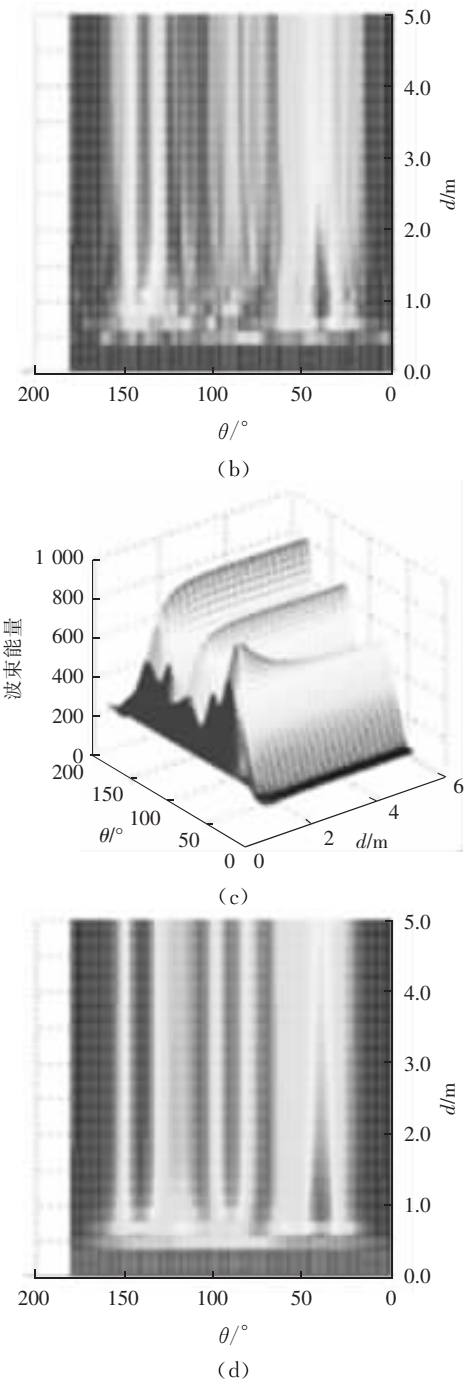


图 3 SNR = 20 dB 时方位角和距离估计的 SRP 搜索

阵列越远的声源其纵伸越明显,因其声波传播到麦克风阵列更接近于远场模式。

根据对全局 SRP 搜索结果的分析,提出以下优化的近场搜索策略:在特定应用场景,优先选取距离原点 $d=x$ 米处的格点进行方位角遍历,搜索出最大的 n 个波峰所在 $\hat{\theta}'$(滤除波峰附近 $\pm5°$ 范围内的抖动) 作为粗定位结果,然后在这 n 个 $\hat{\theta}'$ 范围内进行距离上的遍历,进一步确定出 n 个声源的精确方位角 $\hat{\theta}$ 和距离 $\hat{d}$。

3.3　算法小结

下面对改进算法的处理过程进行描述,并分析主要步骤的运算复杂度,记录于表 1。表中"—"表示运算量不固定,由具体数据和计算方法决定。从表中看到,提出的改进算法只在原算法基础上增加了步骤 2 ～ 6,运算量均很少,而优化的近场搜索策略相比全局搜索减少了倍数级的运算量。整个算法没有平方级的运算步骤,一般性能的数字信号处理器便能满足实时处理要求。

表 1　改进算法步骤分析

算法步骤描述	乘法运算量	加法运算量
(1) 各麦克风数据帧求 *FFT* 变换到频域得到 $X_i(k)$	$O(2L \cdot \log_2 2L)$	$O(2L \cdot \log_2 2L)$
(2) 求 $X_i(k)$ 得功率谱 $\lvert X_i(k)\rvert^2$	$O(2\ \mathrm{mL})$	$O(\mathrm{mL})$
(3) 各麦克风功率谱求和得 $A_{sum}(k)$		$O((m-1)L)$
(4) 计算 $A_{sum}(k)$ 峰均比得到 $P(k)$	$O(L+1)$	$O(L-1)$
(5) $P(k)$ 取 10 为底的对数得到 $P_{\log}(k)$	—	—
(6) 计算加权系数 $w(k)$	$)O(L)$	$O(2L)$
(7) 依据式(13) 做 IFFT,求麦克风间 GCC	$O(m(m-1) \cdot L \cdot (\log_2 2L+2))$	$O(m(m-1) \cdot L \cdot \log_2 2L)$
(8) 遍历搜索 $d=1$ 米处的 θ,范围取 $-179° : 1° : 180°$	$O(180m(m-1))$	$O(1260m(m-1))$
(9) 从 8 中结果搜索出 n 个峰值	—	—
(10) 距离遍历搜索,范围取 $0.1 : 0.1 : 5\ m$	$O(225n \cdot m(m-1))$	$O(1575n \cdot m(m-1))$
(11) 从 10 结果中搜索出 n 个峰值	—	—

4　仿真分析

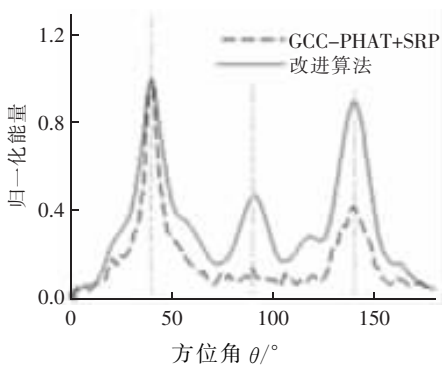
XY 平面上设一均匀圆阵,阵元数为 20,阵列孔径为 0.6 m。同一平面空间中同时存在三个声源,两个等幅宽带语音位于 $\theta=[40,140]$,$d=[0.7,2]$ 米处,另一个窄带单音(1 kHz) 位于 $\theta=90°$,$d=1.5$ 米处,单音幅度为两个宽带语音最大幅度的 0.5 倍。

4.1　SRP 搜索曲线

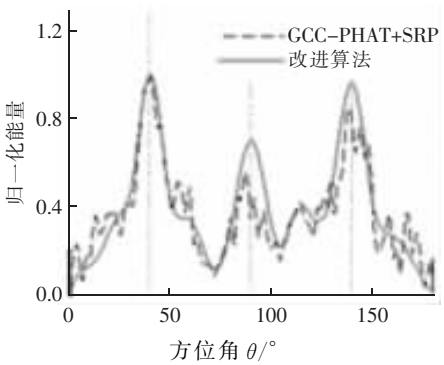
图 4 展示了高低两种信噪比下,距离 d 米处的声源方位角搜索结果。从图中看到,原定位算法的性能随着信噪比的降低下降很多,产生过大的定位模糊与伪峰,且它对窄带单音的定位效果很不理想,这种不理想在高信噪比时表现得愈加明显。而我们提出的改进算法在低信噪比下依然能保持良好的搜索曲线,窄带声源与宽带声源的能量峰一样得到了增强,能够被检测出来。

4.2　算法定位性能分析

设置估计结果的均方根误差(RMSE) 作为指标



(a) SNR = 30 dB,d = 1 m



(b) SNR = 0 dB,d = 1 m

图 4　不同信噪比下的能量搜索曲线

来衡量算法的定位性能。方位角和距离估计的 RMSE 可用式(14) 和(15) 计算:

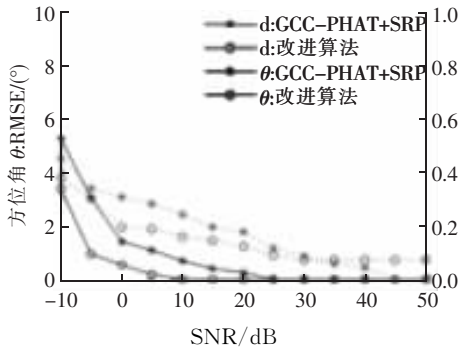
$$\theta^{RMSE}=\sqrt{\sum_{n=1}^{N}\{\min[(\theta_n-\hat{\theta}_n],10\}^2/N} \tag{14}$$

$$d^{RMSE}=\sqrt{\sum_{n=1}^{N}\{\min[(d_n-\hat{\theta}_n],0.5\}^2/N} \tag{15}$$

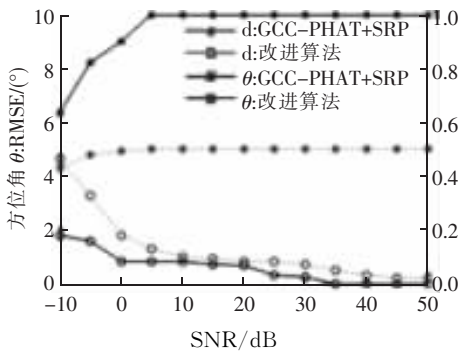
其中 N 为重复试验的次数,实验中取值 200。为体现不同方法在宽、窄带声源估计上的差异,将两个宽带声源和窄带声源分开进行统计,统计结果如图 5 所示。

由图 5 可以看到,对宽带声源的定位中,低信噪比时改进算法比原算法平均高出 1° 以上的方位角精度,同时在距离定位的效果上也要更优。对窄带声源的定位中,原算法的角度和距离估计 RMSE 均接近设计的极限值,几乎无法对窄带声源进行定位,而改进算法对窄带声源甚至是单音的定位效果依然能保持和宽带声源相当的水平。综合来看,我们提出的

改进算法对外部声源环境变化的适应能力大大增强,定位效果优于原算法。



(a) 宽带声源定位性能分析



(b) 窄带声源定位性能分析

图 5 2 种算法对宽 / 窄带声源的定位性能分析

5 结束语

针对近场多目标声源定位问题,本文提出了一种基于对数功率峰均比加权的 GCC-PHAT 时延估计方法,并结合优化后的 SRP 搜索策略,得到改进的声源定位算法。该算法相比原算法增强了抗非相干噪声性能,解决了原算法对窄带声源位置估计困难的问题。改进算法仅增加了几个低计算复杂度的处理步骤,且计算均在单帧内进行,加之优化的 SRP 搜索策略相比全局搜索减少了倍数级的计算量,使其非常适合于实时处理。仿真结果也表明,面对环境噪声的变化和声源的多样性,改进算法在低信噪比下也均能取得接近 1°的角度估计精度和 0.15米的距离估计精度,具有很强的鲁棒性。

参考文献:

[1] KWAK K,KIM S. Sound Source Localization with the Aid of Excitation Source Information in Home Robot Environments[J]. IEEE Trans on Consumer Electronics,2008,54(2):852-856.

[2] 崔玮玮. 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,2009.

[3] CIGADA A,RIPAMONTI F,VANALI M. The Delay & Sum Algorithm Applied to Microphone Array Measurements: Numerical Analysis and Experimental Validation[J]. Mechanical Systems and Signal Processing,2007,21(6):2645-2664.

[4] ALAMEDA X,HORAUD R. A Geometric Approach to Sound Source Localization from Time-Delay Estimates[J]. IEEE/ACM Trans on Audio, Speech, and Language Processing,2014,22(6):1082-1095.

[5] RAYKAR C, YEGNANARAYANA B, PRASANNA S, et al. Speaker Localization Using Excitation Source Information in Speech[J]. IEEE Trans on Speech and Audio Processing,2005,13(5):751-761.

[6] OMOLOGO M,SVAIZER P. Use of the Crosspower-spectrum Phase in Acoustic Event Location[J]. IEEE Trans on Speech and Audio Processing,1997,5(3):288-292.

[7] DOCLO S,MOONEN M. Design of Broadband Beamformers Robust Against Gain and Phase Errors in the Microphone Array Characteristics[J]. IEEE Transactions on Signal Processing,2003,51(10):2511-2526.

[8] COBOS M, MARTI A, LOPEZ J. A Modified SRP-PHAT Functional for Robust Real-Time Sound Source Localization With Scalable Spatial Sampling[J]. IEEE Signal Processing Letters,2011,18(1):71-74.

[9] NISHIURA T,YAMADA T,NAKAMURA S,et al. Localization of Multiple Sound Sources Based on a CSP Analysis with a Microphone Array[C]. IEEE International Conference on Acoustics,Speech,and Signal Processing,2000,2:1053-1056.

[10] ARGENTIERI S,DANES P. Broadband Variations of the MUSIC High-resolution Method for Sound Source Localization in Robotics[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems,2007.

[11] 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.

[12] 刘志明,周辉林,谭思浩,等. TR-MUSIC 与图像熵相结合的室内多目标穿墙雷达图像重构[J]. 南昌大学学报:理科版,2015,39(4):338-341.

[13] XENAKI A,GERSTOFT P,MOSEGAARD K. Compressive Beamforming[J]. The Journal of the Acoustical Society of America,2014,136(1):260-271.

[14] YARDIBI T,LI J,STOTCA P,et al. Sparse Representations and Sphere Decoding for Array Signal Processing[J]. Digital Signal Processing,2012,22(2):253-262.