

改进的 SRP-PHAT 声源定位方法

谭颖 殷福亮 李细林
(大连理工大学 电子与信息工程学院 大连 116024)

摘要 基于联合可控响应功率和相位变换(SRP-PHAT)的传声器阵列声源定位方法是一种鲁棒的声源定位方法。该方法在弱噪声和适度混响的环境下,可以获得较精确的波达方向角(DOA)估计,但在低信噪比或强混响的环境下,该方法的性能并不理想,而且庞大的运算量也限制了其在实时定位系统中的应用。针对这两方面的问题,本文提出了一种改进的 SRP-PHAT 传声器阵列声源定位方法,从而使定位方法的运算量大大降低。仿真实验表明,改进的 SRP-PHAT 定位方法即使在噪声干扰较大和混响影响较严重的环境下,仍具有较高的定位精度。

关键词 声源定位, 传声器阵列, 方向角估计, SRP-PHAT 算法

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2006)07-1223-05

Sound Localization Method Using Modified SRP-PHAT Algorithm

Tan Ying Yin Fu-liang Li Xi-lin

(School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract A method, which combines Steered Response Power with PHase Transform (SRP-PHAT), is a robust speech localization approach with microphone arrays. A relatively precise direction of arrival (DOA) estimation can be achieved by using it in the environment of weak noise and moderate reverberation. However, the localization performance degrades sharply in the environment of low SNR and strong reverberation. Moreover, a real-time localization system is also difficult to be realized due to its heavy computation load. Aiming at these problems, a modified SRP-PHAT localization algorithm is proposed in order to improve the performance of speech localization system. The modified SRP-PHAT algorithm reduces computation load greatly. The results of simulation experiment indicate that the proposed method gives higher localization precision even in an environment of low SNR and strong reverberation.

Key words Sound localization, Microphone array, DOA estimation, SRP-PHAT

1 引言

传声器阵列技术是源于雷达、声纳等系统中采用的阵列信号处理技术,声源定位问题是传声器阵列中的基本问题之一。传声器阵列声源定位在通信、多媒体、人机接口、语音识别、视频会议系统、声音检测设备、移动机器人和助听器装置等许多领域,具有广泛的应用价值^[1]。例如,在视频会议系统和演播室中,可以通过传声器阵列来确定说话人的位置,以引导其波束方向和摄像机对准正在说话的人^[2]。

现有的声源定位技术主要分为 3 大类^[3],即基于最大可控响应功率的波束形成方法、采用高分辨率谱估计的方法和基于时延估计的两步定位方法。2000 年,DiBiase提出了联合可控响应功率和相位变换(SRP-PHAT)的声源定位方法^[4],该方法将可控响应功率方法固有的鲁棒性、短时分析特性与时延估计中相位变换方法对信号周围环境的不敏感性相结合,使声源定位系统具有一定的抗噪性、抗混响性和鲁棒性。但是,SRP-PHAT方法运算量庞大,并且在恶劣环境下(噪声

干扰大、混响影响严重)性能急剧下降,因此在实时处理和实际应用时变得困难。为此,本文针对SRP-PHAT算法这两方面缺点,提出了一种改进的SRP-PHAT方法。该方法去掉了相位累加和没有贡献的互功率谱成分,并将原方法中所有频段的全搜索改为分频段由粗到精的搜索,从而使得定位方法的总运算量大大降低。同时,分频段的方向角搜索方法还使定位系统的抗噪声、抗混响性能大幅提高。计算机仿真实验结果表明,改进的SRP-PHAT方法是一种更精确、更鲁棒的传声器阵列声源定位方法,可应用于对定位精度要求较高的实时定位系统。

2 传声器信号模型

假设声音传播满足线性波动方程^[5],这时声源和传声器间的声波传播通道可认为是线性系统^[6]。在实际的小型房间环境中,由于房间墙壁的反射等原因,使得语音信号在房间内进行多径传播,从而产生接收信号幅度衰减和音质变差的现象,这就是混响效应^[7]。由于混响会影响语音传声器阵列系统的性能,因此房间冲激响应模型必须考虑多径传播。

假设第 n 个传声器接收到的信号为 $x_n(t)$,它可以表示为

$$x_n(t) = s(t) * h_n(t) + v_n(t) \quad (1)$$

其中 $s(t)$ 为声源信号, $v_n(t)$ 为噪声, $h_n(t)$ 是总的冲激响应,

2004-08-13 收到, 2006-03-31 改回

国家自然科学基金(60372082, 60172073)和教育部跨世纪优秀人才基金资助课题

‘*’为卷积运算符。 $h_n(t)$ 是声源空间方向和传声器位置的函数,即房间冲激响应和传声器信道响应两级级联滤波器的输出结果。前者包括直达路径在内的所有从声源到传声器间声学通道的特性。 $v_n(t)$ 是声波多径反射干扰和环境噪声之和,而环境噪声通常比信道噪声更显著,是构成 $v_n(t)$ 的主要部分。假设 $s(t)$ 与 $v_n(t)$ 不相关,那么当关注直达路径成分的冲激响应时,式(1)修改为

$$x_n(t) = \frac{1}{r_n} s(t - \tau_n) * g_n(t) + v_n(t) \quad (2)$$

其中 r_n 表示声源与传声器间的距离, τ_n 表示直达路径的时延, $g_n(t)$ 表示修正后的冲激响应,它由原始响应减去直达路径响应构成。

3 SRP-PHAT声源定位方法^[4]

延时累加可控响应功率(SRP)方法只需较短的时间段,并对周围环境条件不敏感,但不能达到在强混响环境下应用的程度。滤波求和SRP法增加了灵活性,但在只有噪声的情况下,滤波器的设计应该适合最优化信噪比准则,并且这种方法过于依赖信号和信道的先验信息。

处理多径信道畸变的一种基本方法是通过合理选择频域相关函数的权值,使系统对混响的影响更鲁棒。相位变换(PHAT)法就是一种典型的变换方法。对于时延估计方法中的广义互相关(GCC)法,在低度乃至中度混响的条件下,已经证明PHAT加权能提供较强的鲁棒性^[8]。尽管相位变换法有效地减小了多径信道的畸变,但是同时也使低信噪比时无关的频谱成分变得更加明显,从而导致错误的定位结果。特别是在低信噪比和强混响的恶劣环境下,相位变换法性能会急剧下降。

SRP-PHAT方法有效地将声源定位中SRP法固有的鲁棒性和短时分析特性,与PHAT加权对其周围环境不敏感性相结合。SRP-PHAT方法的波束输出功率定义为^[4]

$$P(\mathbf{q}) = \sum_{k=1}^N \sum_{l=1}^N \int_{-\infty}^{\infty} \Psi_{kl}(\omega) X_k(\omega) X_l^*(\omega) e^{j\omega(\tau_l - \tau_k)} d\omega \quad (3)$$

其中 $X_k(\omega)$ 为第 k 个传声器信号 $x_k(t)$ 的加窗傅里叶变换; τ 是阵列指向声源的可控时延,用于补偿每个传声器理想信号的直接路径传播时延; \mathbf{q} 为声源的空间位置向量;多信道形式的PHAT加权系数 $\Psi_{kl}(\omega)$ 为

$$\Psi_{kl}(\omega) = \frac{1}{|X_k(\omega) X_l^*(\omega)|} \quad (4)$$

这等同于在滤波求和波束形成情况中使用单个通道滤波器^[9]。

这就是理想的SRP-PHAT滤波器,它可以通过式(3)这种频域表达形式实现。换言之,式(3)等同于所有 N 个传声器中任意两个传声器信号的广义互相关之和。这意味着,仅有两个传声器的二维阵列SRP函数,等同于两个传声器信号的广义互相关函数。因此,传声器数目的增加,使得基于传声器

对的广义互相关方法扩展到了基于传声器阵列的可控响应功率技术,但是,其计算所有广义互相关总和的运算量相当大。

假设 $R_{kl}(\tau)$ 为第 k 个和第 l 个传声器信号的PHAT加权广义互相关函数,那么SRP-PHAT方法的时域表达式为

$$P(\mathbf{q}) = 2\pi \sum_{k=1}^N \sum_{l=1}^N R_{kl}(\tau_l - \tau_k) \quad (5)$$

这是所有传声器对排列的广义互相关函数总和,是通过可控时延的不同时移得到,其中还包括 N 个零时延的广义自相关函数之和,这些项独立于可控时延,它们仅为SRP提供直流偏移量。

SRP-PHAT定位方法是在所有可能覆盖声源位置的范围内容搜索 $P(\mathbf{q})$ 最大值,即

$$\hat{\mathbf{q}}_s = \operatorname{argmax}_{\mathbf{q}} P(\mathbf{q}) \quad (6)$$

在波束形成定位方法中,是用波束引导传声器阵列指向声源,而SRP-PHAT方法弱化了无关峰值,突出了真正峰值,从而使得SRP-PHAT定位方法对噪声和混响的灵敏度降低,即提高了鲁棒性和定位的精确性。此外,这种方法只需采用较短的分析时间段就能够实现。

尽管SRP-PHAT方法具有一定的抗噪性、抗混响性和鲁棒性,但实际上,由于较强房间混响、较强环境噪声的影响或者阵列几何位置的限制,它常常会在一些错误位置产生许多峰值,这些无关的空间极大值可以使真正的声源空间位置变得不明显,并且会使全局峰值的搜索过程变得极其复杂。因此,SRP-PHAT方法庞大的运算量和恶劣环境下急剧下降的性能,使得其应用于实时处理和实际应用系统变得困难。

4 改进的SRP-PHAT方法

SRP-PHAT方法在中度以上环境噪声和强混响下性能较差,并且其庞大的运算量限制了在实时系统中的应用。针对这些不足,我们主要进行了两方面的改进。

4.1 对运算量最大的搜索部分的改进

原方法采用在8000Hz以下所有频段方位角进行分辨率为 1° 的全搜索方式,而实际中,大部分背景噪声较集中地分布在250Hz以下这一频段,因此我们将250Hz以下的频率成分滤掉,并利用余下的噪声估计出噪声互功率谱,在信号的互功率谱中减掉噪声互功率谱。为了防止混叠,将7500Hz以上的频率成分也滤掉。于是,我们先采用带通滤波器对信号滤波,使信号频率范围为250Hz~7500Hz。在改进的SRP-PHAT方法的搜索部分,首先在语音信号能量较集中(信噪比较高)的低频段250Hz~4000Hz进行分辨率为 10° 粗搜索,再在高频段4000Hz~7500Hz进行分辨率为 0.5° 精搜索。图1和图2分别为语音信号在低频段和高频段内各角度的互功率谱加权和。由于在传声器接收到的信号中低频段信号的信噪比较大,所以采用先在低频段搜索的方式,这样会使受

噪声干扰的定位误差大大减小,从而改善了原方法及部分时延估计声源定位法都不可避免的随着信噪比的降低定位系统的性能急剧下降的缺点。

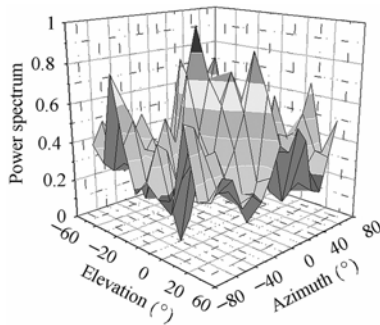


图1 低频段各角度的互功率谱加权

Fig.1 The weighed sum of cross power spectrum in low frequency band

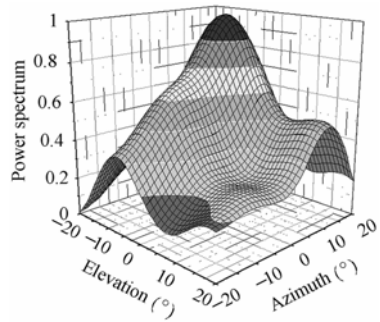


图2 高频段各角度的互功率谱加权

Fig.2 The weighed sum of cross power spectrum in high frequency band

此外,上述的改进方法也使得搜索部分的运算量大幅度降低。假设声源的偏角搜索范围为 $(\pm m)^\circ$,仰角搜索范围为 $(\pm n)^\circ$,精细搜索的角度分辨率精确到 $(1/a)^\circ$,那么改进的SRP-PHAT方法的方位角搜索次数由SRP-PHAT方法的 $(m \times a + 1)(n \times a + 1)$ 次,降为 $(m \times a / 10 + 1)(n \times a / 10 + 1) + (20 \times a + 1)^2$ 次。在实际中,通常选取 $m = 89^\circ$, $n = 55^\circ$,这样就使得方向角搜索部分的平均运算量约降为原来的9.17%。

4.2 对互功率谱加权求和部分的改进

将原方法中对所有排列组合得到的传声器对的归一化互功率谱总和,改变为相同的传声器对只取一次互功率谱进行加权求和,去掉对相位累加和没有贡献的自相关功率谱求和部分,于是有

$$P'(q) = \sum_{k=1}^N \sum_{l=k+1}^N \int_{-\infty}^{\infty} \Psi_{kl}(\omega) X_k(\omega) X_l^*(\omega) e^{j\omega(\tau_l - \tau_k)} d\omega \quad (7)$$

这样,改进的SRP-PHAT方法的互功率谱加权求和部分的运算量,由SRP-PHAT方法此部分所需要的 N^2 次降低为 $N(N-1)/2$ 次,平均运算量仅为原来的50%。显然,上述两方面的改进使得改进后的SRP-PHAT方法的总运算量大大降低。

虽然传声器阵列声源定位中常常采用基于时延估计的两步定位方法,但是由于时延估计方法中第一步得到的时延差估计值只能为整数点样本,这就导致了第二步方向角度估

计值也只有整数点样本间隔,而改进的SRP-PHAT方法实现了时延估计方法难以达到的分数级分辨率,在混响、噪声和信号分析帧长相同的情况下,把时延估计方法中由于整数级时延差所限而导致的整数级估计度数的分辨率,精确到了 0.5° 这个分数级度数的分辨率。这里的分辨率还可以精确到 0.1° 甚至更小,但是考虑到运算量的限制, 0.5° 的精度已经能够满足实际定位系统的性能要求。改进的SRP-PHAT方法继承了基于可控响应功率的波束形成方法可以应用于多声源定位系统的优势,并且在多声源的情况下不影响各声源的定位精度。

综上所述,改进后的SRP-PHAT方法提高了声源定位系统的抗噪性能,大幅度降低了总体运算量,是一种可以应用于传声器阵列多声源系统中的鲁棒定位方法。

5 计算机仿真实验

为了验证本文方法的有效性,我们在不同信噪比和混响环境下进行了仿真实验。实验模拟了实际中小型会议室的环境情况,其中房间混响的冲激响应函数由IMAGE模型^[10]产生,房间的长宽为 $7\text{m} \times 5\text{m}$,通过调整房间墙壁的反射系数来改变对应的混响时间。实验中的噪声由两部分组成,即噪声源经过房间冲激响应卷积后的散射噪声和传声器自身噪声;噪声源使用MATLAB中randn()产生的高斯白噪声,传声器自身噪声使用CoolEdit软件产生的粉红噪声。

输入语音信号首先经过250Hz到7500Hz的带通滤波器变为带限信号,然后以32KHz的采样率进行采样,并将采样值以16位二进制数表示。语音数据帧长为32ms(1024采样点),相邻两帧重叠3/4,窗函数为汉明窗(Hamming)。每1s为一次定位判决周期。测试的语音长度为10s,每次实验可估计角度约10次。纯净语音信号波形、受噪声和混响干扰的语音信号波形如图3所示。实验选取DOA角度的搜索范围是:偏角 $[-75^\circ, 75^\circ]$ 和仰角 $[-55^\circ, 55^\circ]$ 。

当声学环境较理想且传声器摆放较合理时,使用适当数量的传声器就可以实现声源的准确定位。传声器阵列摆放模式如图4所示。组成传声器阵列的5个传声器位置分别为 $m_0 = [0, 0, 0]$, $m_1 = [0, 0.2, 0]$, $m_2 = [0, 0.4, 0]$, $m_3 = [0,$

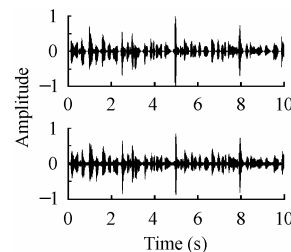


图3 分别为持续10s的原始纯净语音信号波形和混有噪声、混响的单通道语音信号波形

Fig.3 Waveforms of the clean speech signal and the speech signal corrupted by noise and reverberation

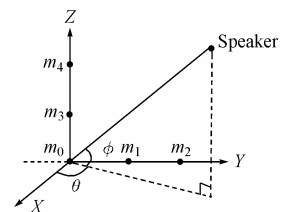


图4 实验中传声器摆放坐标示意图

Fig.4 Diagram of microphones placement

0, 0.2], $m_4 = [0, 0, 0.4]$, 水平传声器之间和垂直传声器之间的距离都是 20cm。三个水平传声器用来定声源的偏角 θ , 三个垂直传声器用来定声源的仰角 ϕ 。在前 3 组仿真实验中, 声源的位置为 $X_S = [3.5, 4.8, 2.9]$ 。仿真实验表明, 这种传声器阵列的位置摆放结构使方向角估计精度较高, 角度估计值能够较稳定地出现在某个固定值周围, 估计角度误差的方差值有所减小, 即系统具有一定的鲁棒性。

实验 1 在两种典型的声学环境下, 对实际语音进行连续定位的实验结果 在 SNR=25dB、混响为 50ms 的声学环境下和在 SNR=0dB、混响为 250ms 的声学环境下, 我们用 SRP-PHAT 方法和改进后的 SRP-PHAT 定位方法对一段实际语音分别进行连续定位, 实验结果分别如表 1 和表 2 所示。其中实际声源方向角为 $(52.8^\circ, 31.7^\circ)$, $\hat{\theta}$ 表示声源的偏角估计值, $\hat{\phi}$ 表示声源的仰角估计值, A 表示由于当前帧混响或噪声过大等原因而放弃此帧估计。

从表 1 可以看出, 在信噪比较大、混响影响较弱的环境下, 原 SRP-PHAT 方法的第 5 帧对可以估计出的一帧放弃估计, 而第 6 帧在高频段受到短时噪声严重干扰时没有放弃估计, 导致出现了明显错误的估计值。相比之下, 改进的 SRP-PHAT 方法在连续定位的 12 帧中, 没有出现较大误差的角度估计值。由此可见, 改进后方法的抗噪能力有所提高。

从表 2 可以看出, 由于噪声干扰较大和混响影响较严重(强噪声、强混响), 不但许多帧都放弃了估计, 而且原 SRP-PHAT 方法在第 8 帧和第 12 帧处都出现了明显错误的估

计值。与原 SRP-PHAT 方法相比, 改进的 SRP-PHAT 定位方法在恶劣的背景噪声和房间影响混响下, 错误估计的次数有较明显的减少。

实验 2 在不同声学环境下, 对声源位置错误估计百分比的实验结果 测试语音长度为 10s (1250 帧), 除去非语音帧和放弃估计帧, 语音帧约为 700 帧。在实验中, 若偏角值或仰角值偏离实际值 5° 以上, 则认为是错误估计值。先将信噪比固定在 10dB, 混响时间以 50ms 为步长从 0ms 到 300ms 依次取值; 再将混响时间固定在 50ms, 信噪比以 5dB 为步长从 30dB 变化到 -10dB。我们进行了 10 组实验, 统计出错误估计帧数占总语音估计帧数的百分比, 如表 3 和表 4 所示。

从表 3 和表 4 可以看出, 当信噪比较好、混响影响不明显时, 改进的 SRP-PHAT 定位方法与原方法的错误估计率都较低, 均有着良好的性能。然而, 随着噪声干扰变大和混响影响更严重时, 两种方法的性能都有着不同程度的下降, 改进的 SRP-PHAT 方法好于原 SRP-PHAT 方法。这说明改进的 SRP-PHAT 方法抗噪声抗混响能力都有所提高, 并且在恶劣条件下有较强鲁棒性。

实验 3 DOA 估计值随信噪比和混响时间变化的均方根误差曲线 采用均方根误差(RMSE)来比较两种方法的性能。均方根误差的定义式为

$$E_{\text{RMSE}}(\hat{\theta}, \hat{\phi}) = \sqrt{(\hat{\theta} - \theta)^2 + (\hat{\phi} - \phi)^2}$$

表 1 SNR=25dB, 混响为 50ms 的环境下, 一组连续定位的测试结果(单位: $^\circ$)

Tab.1 A set of localization testing results under the conditions of SNR=25dB and RT60=50ms (unit:degree)

当前帧数		1	2	3	4	5	6	7	8	9	10	11	12
SRP- PHAT	$\hat{\theta}$	A	51.0	51.0	A	A	31.0	49.0	50.5	50.5	51.0	A	50.5
	$\hat{\phi}$	A	31.5	31.5	A	A	10.0	31.0	31.0	31.0	31.5	A	31.0
改进 SRP- PHAT	$\hat{\theta}$	A	51.0	51.0	A	51.0	A	49.5	49.5	50.5	51.0	A	50.5
	$\hat{\phi}$	A	31.5	31.5	A	31.5	A	31.0	31.0	31.0	31.5	A	31.0

表 2 SNR=0dB, 混响为 250ms 的环境下, 一组连续定位的测试结果(单位: $^\circ$)

Tab.2 A set of localization testing results under the conditions of SNR=0dB and RT60=250ms (unit: degree)

当前帧数		1	2	3	4	5	6	7	8	9	10	11	12
SRP- PHAT	$\hat{\theta}$	A	32.5	A	A	A	A	50.0	33.5	31.0	A	A	44.5
	$\hat{\phi}$	A	10.5	A	A	A	A	31.5	11.0	10.0	A	A	30.0
改进 SRP- PHAT	$\hat{\theta}$	A	50.5	A	A	A	A	50.0	31.0	50.0	A	A	62.5
	$\hat{\phi}$	A	32.0	A	A	A	A	31.5	10.0	31.5	A	A	30.0

表 3 SNR=10dB, 混响时间变化的实验结果

Tab.3 A set of testing results under the conditions of SNR=10dB and different reverberation time

混响时间(ms)		0	50	100	150	200	250	300
DOA 错误估计百分比(%)	SRP-PHAT	1.43	2.43	3.86	9.28	19.86	30.14	47.86
	改进的 SRP-PHAT	0.71	1.14	2.85	6.00	12.29	21.57	40.43

表 4 混响时间 50ms, 信噪比变化的实验结果

Tab.4 A set of testing results under the conditions of RT60=50ms and different SNR levels

SNR (dB)		30	25	20	15	10	5	0	-5	-10
DOA 错误估计百分比 (%)	SRP-PHAT	0.71	0.71	0.86	1.29	2.43	7.28	15.43	44.71	67.00
	改进的 SRP-PHAT	0	0.28	0.28	0.57	1.14	2.14	9.43	21.71	37.43

其中 θ 和 ϕ 是真正的偏角值和仰角值, $\hat{\theta}$ 和 $\hat{\phi}$ 分别是它们的估计值。

测试语音长度为 10s。DOA 估计值的 RMSE 随信噪比变化的曲线如图 5 所示。DOA 估计值的 RMSE 随混响变化的曲线如图 6 所示。

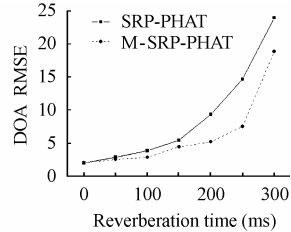
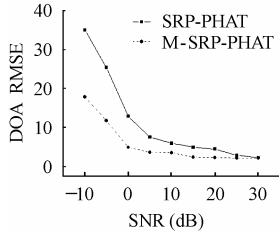


图5 混响时间 50ms, DOA 估计值随信噪比变化的 RMSE 曲线

图6 信噪比 25dB, DOA 估计值随混响时间变化的 RMSE 曲线

Fig.5 RMSE curve of DOA estimate with different SNR levels (RT60=50ms)

Fig.6 RMSE curve of DOA estimate with different RT60s(SNR=25dB)

随着信噪比的减小和混响影响的增强, SRP-PHAT 方法会出现毫不相干的较大偏差的估计值, 这主要是错误位置产生的许多空间极大值使真正声源位置的空间峰值变得不明显, 从而使全局峰值的搜索出错。从图 5 和图 6 可以明显地看到, 改进后的方法无论在抗噪性能还是在抗混响能力上均得到较好的改善。

实验 4 对不同声源空间位置的方向角估计实验 选取 SNR=25, 混响为 50ms。改变声源的空间位置, 使其方向角分别为: (1) (52.8°, 31.7°); (2) (25.3°, 15.3°); (3) (16.0°, -10.0°); (4) (0°, -5.6°); (5) (-33.0°, -44.5°) 这 5 种情况。分别进行 10 组方向角估计实验。除去其中因干扰次峰太大而会放弃估计的次数, 计算其余所有估计出的声源偏角、仰角的平均值与实际声源偏角、仰角的偏差, 如表 5 所示。

由表 5 可以看出, 改进的 SRP-PHAT 方法比 SRP-PHAT 方法的估计偏差小。但是两者在偏角和仰角较大时, 估计方向角的误差也随之变大。该实验结果与实际中人耳定位特性相似, 理论分析也是如此。

表 5 两种定位算法的方向角估计测试结果

Tab.5 Testing results of DOA estimate for the tow localization methods

方向角	SRP-PHAT 法 估计偏差		改进 SRP-PHAT 法 估计偏差	
	$\hat{\theta} - \theta$	$\hat{\phi} - \phi$	$\hat{\theta} - \theta$	$\hat{\phi} - \phi$
(1)	-2.8565	-0.4093	-2.1235	-0.4000
(2)	-0.3750	0.3913	-0.2852	-0.2650
(3)	0.3913	-0.2905	-0.3644	-0.1740
(4)	0.0500	0.1650	0.0500	0.1145
(5)	1.9110	-3.0319	1.9056	-2.7908

从以上仿真实验结果可以看出, 改进的 SRP-PHAT 方法在多种声学环境下, 始终具有较低的错误估计率和较小的均方根误差值, 因此用于传声器阵列声源定位系统中具有较高的 DOA 估计精度, 其性能优于 SRP-PHAT 方法。此外, 对于实际系统而言, 改进的 SRP-PHAT 方法计算量较低, 适于进行实时处理。

6 结束语

为了提高 SRP-PHAT 声源定位方法的抗噪声、抗混响能力, 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 SRP-PHAT 声源定位方法。该方法在强噪声和强混响环境下明显地弱化了干扰峰值, 突出了真正峰值, 对噪声和混响的灵敏度降低, 使定位系统具有较强的鲁棒性。此外, 改进的 SRP-PHAT 方法计算量大幅度降低, 可用于对定位精度要求较高的实时定位系统。

参考文献

- [1] Wang Qinghua, Ivanov T, Aarabi P. Acoustic Robot Navigation using Distributed Microphone Arrays. *Information Fusion*, 2004, Vol.5: 131-140.
- [2] Potamitis I, Chen Huimin, Tremoulis G. Tracking of Multiple Moving Speakers with Multiple Microphone Arrays. *IEEE Trans. on speech and audio processing*, 2004, 12(5): 520-529.
- [3] Brandstein M. A framework for speech source localization using sensor arrays. Doctoral Dissertation, Brown University, May 1995.
- [4] DiBiase J. A high-accuracy, low-latency technique for talker localization in reverberant environments. PhD thesis, Brown University, Providence RI, USA, May 2000.
- [5] Kinsler L, Frey A, Coppens A, Sanders J. *Fundamentals of Acoustics*. Third edition, New York: John Wiley & Sons, 1982: 104-131.
- [6] Ziomaek L. *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing*. Boca Raton, FL: CRC Press, 1995: 210-248.
- [7] Huang Jie, Ohnishi N, Sugie N. Modeling the precedence effect for sound localization in reverberant environment. *IEEE Instrumentation and Measurement Technology Conference*, Brussels, June 1996: 633-636.
- [8] Svaizer P, Matassoni M, Omologo M. Acoustic source location in a three-dimensional space using crosspower spectrum phase. in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, April 1997:231-234.
- [9] Hahn W, Tretter S. Optimum processing for delay vector estimation in passive signal arrays. *IEEE Trans. on Information Theory*, 1973, 19 (1): 608-614.
- [10] Allen J B, Berkley D A. Image method for efficiently simulating small room acoustics. *Journal of Acoustical Society of America*, 1979, 65(4): 943-950.
- [11] Brandstein M, Adcock J, Silverman H. Microphone array localization error estimation with application to sensor placement. *J. Acoust. Soc. Am.*, April 1997, 99(6): 251-254.

谭颖: 女, 1979年生, 硕士生, 从事语音处理和阵列信号处理的理论与应用研究工作。
 殷福亮: 男, 1962年生, 教授, 博士生导师, 主要从事语音处理和阵列信号处理的理论与应用研究工作。
 李细林: 男, 1980年生, 硕士生, 从事语音处理和阵列信号处理的理论与应用研究工作。