

基于差分麦克风阵列和语音稀疏性的多源方位估计方法

何赛娟 陈华伟 尹明婕 丁少为

(南京航空航天大学电子信息工程学院, 南京, 210016)

摘要: 差分麦克风阵列为实现小尺寸阵列条件下的声源定位提供了一条重要技术途径。语音信号具有稀疏性, 利用该特性可实现基于差分麦克风阵列的多声源方位估计, 其中的典型方法为直方图法。针对差分麦克风阵列, 本文提出了一种基于时频掩蔽和模糊聚类分析的短时平均复声强多声源方位估计方法。分析了不同阵列尺寸条件下时频掩蔽频带范围的选择问题。该方法具有闭式解, 在强混响噪声环境下的性能优于直方图法, 并且受阵列尺寸变化的影响较小。为了改善直方图法的性能, 基于时频掩蔽的思想, 文中还给出了一种修正的直方图方法。混响噪声环境下的仿真实验结果验证了本文所提方法的有效性。

关键词: 差分麦克风阵列; 时频稀疏性; 模糊聚类; 时频掩蔽; 多声源定位

中图分类号: TN912.3 **文献标志码:** A

Direction Finding of Multiple Sound Sources Based on Sparseness of Speech Signals and Differential Microphone Array

He Saijuan, Chen Huawei, Yin Mingjie, Ding Shaowei

(College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China)

Abstract: Differential microphone arrays have become a promising method to address multiple sound source localization. Among the differential microphone arrays, the existing typical method is the histogram approach, which utilizes the time-frequency sparseness characteristic of speech signals. A direction-finding algorithm for multiple sound sources by the short-time average complex sound intensity estimation is proposed based on time-frequency masking and fuzzy clustering. The frequency bounds for time-frequency masking under various array sizes are also discussed. The advantages of the proposed method are that it has closed-form solution, superior to the histogram approach, and also less sensitive to array size. Based on the idea of time-frequency masking, an improved histogram approach is also presented. The performance of the proposed methods is verified by simulation results under noisy and reverberant environment.

Key words: differential microphone array; time-frequency sparseness; fuzzy clustering; time-frequency masking; direction-finding of multiple sound sources

引言

基于麦克风阵列的声源定位技术是音频和语音信号处理领域的重要研究方向之一,广泛应用于视频会议系统、音频监控、助听器、人机语音交互系统^[1-3]等。经典的麦克风阵列声源定位方法主要有3类^[4-6]:时延估计法、可控波束形成法以及高分辨率空间谱估计法等。为了在混响噪声环境下获得良好的定位性能,这些经典方法往往需要较大的阵列孔径。比如时延估计方法,为有效提高时延估计精度,文献^[7]提出了同时利用多麦克风间的空域和时域预测关系进行时延估计,但该方法性能的提升是以增大阵列尺寸和通道数为代价的。而对于可控波束形成法,阵列孔径是决定其空间分辨率的关键因素。因此,为获得较好的混响噪声抑制效果,可控波束形成法也往往要求阵列尺寸较大。而对于一些实际的音频和语音处理系统,其安装平台的尺寸往往有限,这就制约了经典声源定位方法的应用。

差分麦克风阵列为小尺寸阵列条件下的音频和语音信号的采集和处理提供了一条重要技术途径,成为近年来研究的热点^[8-10]。与常规的加性麦克风阵列处理技术(如延迟求和波束形成器)相比,差分麦克风阵列具有更高的阵列增益,并且其空间响应还具有频率不变性等优点^[8]。通过差分麦克风阵列还可以估计出目标声源的声强向量,从而实现小尺寸阵列条件下的目标方位估计。虽然基本的声强向量估计法无法适用于多声源情况下的方位估计,但由于语音信号具有时频稀疏性^[11],利用该特性则可实现多语音源时的方位估计,其中典型的方法如直方图法^[12]。该方法首先计算各声源瞬时方位的直方图分布,然后通过搜索直方图的峰值位置得到目标声源的方位估计。直方图方法在低噪声混响条件下具有良好的性能,但随着环境噪声和混响的增大,其性能恶化。此外,由于没有充分利用语音信号的稀疏性条件,其对阵列尺寸的选择也较为敏感。为了有效实现混响噪声环境下的多语音声源定位,本文针对差分麦克风阵列提出了一种基于时频掩蔽和模糊聚类分析的短时平均复声强多声源方位估计方法。

1 差分阵列信号模型

考虑由4个全向性麦克风 M_1, M_2, M_3, M_4 组成的差分麦克风阵列,如图1所示。4个麦克风等间隔的分布在直径为 D 的圆周上。假设空间中有 N 个声源 $s_1(t), \dots, s_N(t)$ 入射到差分麦克风阵列,记声源 n 的方位角为 $\phi_n \in (-180^\circ, 180^\circ)$,则其中第 i 个麦克风接收到的信号可表示为

$$p_i(t) = \sum_{n=1}^N h_{ni}(t) \otimes s_n(t) + n_i(t) \quad i \in \{1, 2, 3, 4\} \quad (1)$$

式中: $h_{ni}(t)$ 表示第 n 个声源到第 i 个麦克风的房间脉冲响应; $n_i(t)$ 为第 i 个麦克风接收到的加性噪声信号;符号 \otimes 表示卷积运算。

差分麦克风阵列的输出信号包括原点处的声压信号和原点处振速的两个正交分量。其中,原点处的声压信号由各麦克风信号求均值得到,即

$$p_0(t) = \frac{1}{4} [p_1(t) + p_2(t) + p_3(t) + p_4(t)] \quad (2)$$

原点处的振速信号可以通过差分麦克风阵列的差分输出信号获得。理想流体媒质中的小振幅声波的运动方程为^[12]

$$\rho \frac{\partial v(t)}{\partial t} = -\nabla p(t) \quad (3)$$

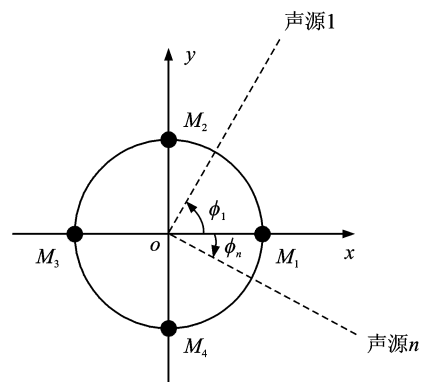


图1 阵列结构和坐标定义示意图

Fig. 1 Array structure and coordinate definition

式中: ρ 为媒质密度; $v(t)$ 为振速; $p(t)$ 为声压; ∇ 为梯度算符。由式(3)可以导出任意 r 方向上的振速分量 v_r , 可以由 r 方向上距离为 Δr 的相近两点的声压值 $p_{r_1}(t), p_{r_2}(t)$ 近似求得

$$v_r(t) = -\frac{1}{\rho} \int_{-\infty}^t \frac{\partial p(\tau)}{\partial r} d\tau \approx -\frac{1}{\rho \Delta r} \int_{-\infty}^t [p_{r_1}(\tau) - p_{r_2}(\tau)] d\tau \quad (4)$$

则由式(4)可以得到原点处 x 轴方向上的振速分量为

$$v_{ax}(t) = -\frac{1}{\rho D} \int_{-\infty}^t [p_1(\tau) - p_3(\tau)] d\tau \quad (5)$$

原点处 y 轴方向上的振速分量为

$$v_{oy}(t) = -\frac{1}{\rho D} \int_{-\infty}^t [p_2(\tau) - p_4(\tau)] d\tau \quad (6)$$

通过差分麦克风阵列测得的原点处的声压和振速信号, 可以得到声源在极点处的声强, 进而由声强的方向得到目标方位的估计值。需要指出的是, 基于声强向量估计的声源定位方法通常只适用于单声源的情况。而语音信号具有时频稀疏性特征, 利用这一特征可以实现基于差分阵列声强向量估计的多声源方位估计, 典型的方法为直方图法。

2 基于直方图法的多声源方位估计

语音信号具有短时平稳特性, 因此通常在时频域中对其进行处理。对阵列接收到的信号进行短时傅里叶变换, 则麦克风接收到的卷积混合信号就变为瞬时混合信号, 由式(3)可得

$$P_i(t, f) = \sum_{n=1}^N H_{ni}(t, f) S_n(t, f) + N_i(t, f) \quad (7)$$

式中: $P_i(t, f), S_n(t, f), H_{ni}(t, f)$ 和 $N_i(t, f)$ 分别表示麦克风接收信号 $p_i(t)$, 声源信号 $s_n(t)$, 房间脉冲响应 $h_{ni}(t)$ 以及加性噪声 $n_i(t)$ 的短时傅里叶变换。

语音信号具有典型的时频稀疏性^[13]。即不同语音声源信号在每个时频点上很少重叠, 在任何一个时频点上只有一个声源起主导作用, 其他声源的贡献可以忽略。基于时频稀疏性条件, 式(7)可进一步近似表示为

$$P_i(t, f) \approx H_{ki}(t, f) S_k(t, f) + N_i(t, f) \quad (8)$$

式中: $S_k(t, f)$ 为在时频点 (t, f) 上占主导的声源, $H_{ki}(t, f)$ 为该主导声源到第 i 个麦克风的房间脉冲响应。

多源方位估计的直方图法的主要步骤为: 首先计算各个时频点处的瞬时复声强; 然后由瞬时复声强得到各时频点所对应的声源方位角, 将所有时频点处求得的声源方位估计值以直方图的形式表现出来; 最后通过搜索直方图频数的峰值位置得到各声源的方位。对任一时频点, 其对应的声源方位估计方法如下。对式(5, 6)分别进行短时傅里叶变换得

$$V_{ax}(t, f) = \frac{j[P_1(t, f) - P_3(t, f)]}{2\pi f \rho D} \quad (9)$$

$$V_{oy}(t, f) = \frac{j[P_2(t, f) - P_4(t, f)]}{2\pi f \rho D} \quad (10)$$

式中: $V_{ax}(t, f)$ 和 $V_{oy}(t, f)$ 分别为原点处振速正交分量 $v_{ax}(t)$ 和 $v_{oy}(t)$ 的短时傅里叶变换。记式(2)中原点处声压信号 $p_0(t)$ 的短时傅里叶变换为 $P_0(t, f)$, 则各时频点的瞬时复声强的两个正交分量可表示为

$$I_{ax}(t, f) = \hat{P}_0(t, f) V_{ax}^*(t, f) \quad (11)$$

$$I_{oy}(t, f) = \hat{P}_0(t, f) V_{oy}^*(t, f) \quad (12)$$

式中: $I_{ax}(t, f)$ 和 $I_{oy}(t, f)$ 分别为声强向量的 x 轴和 y 轴分量, $\hat{P}_0(t, f)$ 为 $P_0(t, f)$ 的希尔伯特变换(因为

这里的振速和声压相位相差 90° , 采用希尔伯特变换以消除相位差), 符号 $*$ 表示共轭运算。由式(11, 12), 可得各个时频点处的方位角估计值为

$$\bar{\phi}(t, f) = \arctan \left\{ \frac{\operatorname{Re} [I_{oy}(t, f)]}{\operatorname{Re} [I_{ox}(t, f)]} \right\} \quad (13)$$

式中: $\operatorname{Re}\{\cdot\}$ 表示取实部, 因为只有声强的实部(即有功声强)含有声源的方位信息^[14]。

图 2 给出了不同环境下直方图方法对两语音声源的方位估计结果。两个声源的实际方位角分别为 -30° 和 30° , 差分阵列的尺寸 $D=0.04$ m。其中, 图 2(a) 为理想条件下的仿真结果; 图 2(b) 为信噪比 20 dB, 混响时间 RT60 为 0.16 s 时的仿真结果。从仿真结果可以看出, 混响噪声对直方图的影响较大, 使得频数的峰值减弱。在混响较严重时, 直方图法的性能急剧恶化。另外, 从仿真结果还可以看出, 除真实方位附近存在频数峰值外, 还有一些伪峰出现, 将对声源方位估计造成影响。为了改善直方图方法在混响噪声环境下的性能, 本文结合时频掩蔽思想, 还给出了一种修正的直方图方法。

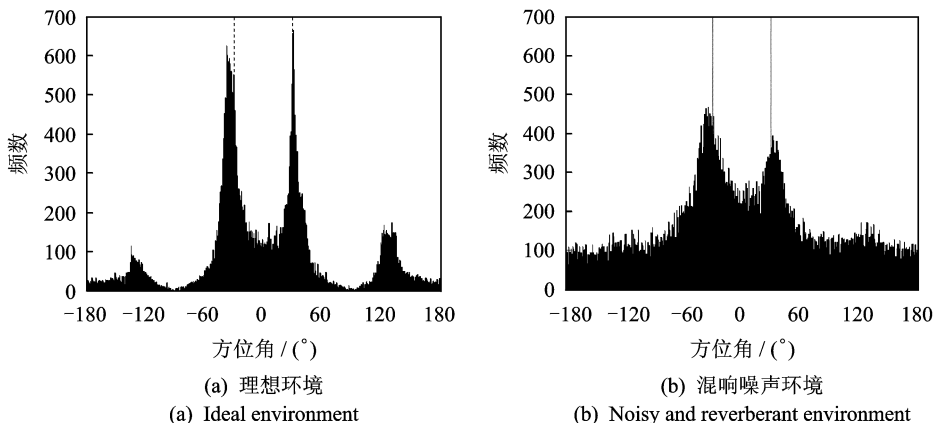


图 2 直方图法的声源方位估计结果
Fig. 2 Results of the histogram method

3 基于时频掩蔽和模糊聚类分析的短时平均复声强多声源方位估计

为了提高混响噪声环境下差分阵列的多源方位估计性能, 本文提出了一种基于时频掩蔽和模糊聚类分析的短时平均复声强多源方位估计方法, 如图 3 所示。该方法的实现流程包括 3 部分: (1) 为减小混响噪声的影响, 在时频域内对差分阵列的输出信号进行掩蔽, 并根据每个时频点处的瞬时复声强估计出该时频点所对应的声源瞬时方位; (2) 利用语音信号在时频域所具有的稀疏特性, 采用模糊聚类分析对所有时频点进行分离; (3) 对每个声源所属的时频点进行短时平均复声强处理, 以估计出相应的方位值。

3.1 基于时频掩蔽的瞬时方位估计

通过前文对直方图法的分析, 发现瞬时方位直方图的频数存在一些伪峰, 这些伪峰的出现将直接影响声源方位的估计性能。伪峰出现的主要原因之一是由于语音信号的稀疏性, 对于那些无语音信号能量分布或者很弱的时频点, 由瞬时复声强估计出的声源方位值将出现异常(即对应于非真实声源方位)。本文的研究表明, 存在异常瞬时方位估计的时频分布范围具有明显的规律性, 与所采用的差分阵列的尺寸有关。因此, 为了有效提高声源方位的估计性能, 需要在瞬时方位估计前进行时频掩蔽处理, 以消除异常时频点的影响。

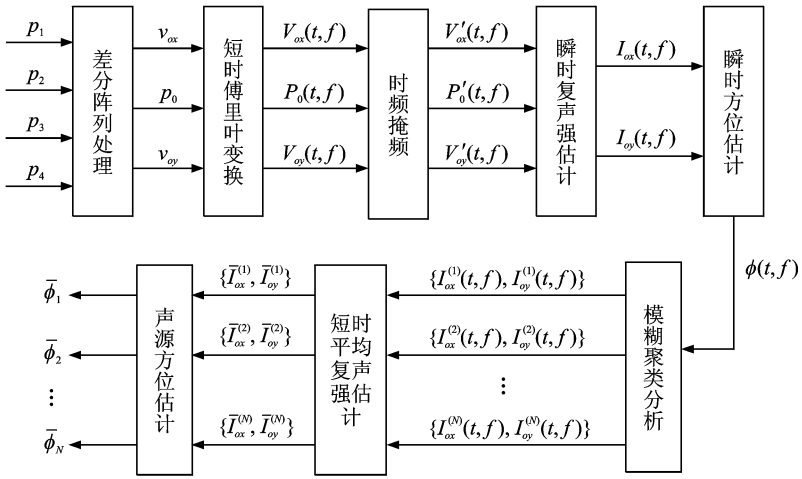


图3 本文方法的流程图

Fig. 3 Flow chart of method proposed in this paper

考虑尺寸 D 分别为 0.06 m 和 0.08 m 的差分麦克风阵列, 语音声源的入射方位为 30° 。图 4 给出了相应的瞬时方位估计误差大于 50° 的异常值的时频分布情况。图 4(a, c) 为理想环境下的结果; 图 4(b, d) 为信噪比 20 dB, 混响时间 0.16 s 时的结果。由图 4 可以看出, 瞬时方位估计的异常值主要出现在零频附近和低频区域, 并且随着差分阵列尺寸的增大, 瞬时方位估计异常值出现的时频域亦增大。在零频附近和低频区域出现异常值的主要原因是语音信号在该频段范围的能量很弱; 而当差分阵列尺寸增大时, 由于高频部分出现空间混叠^[15], 从而导致异常值出现的时频区域增大。这些瞬时方位估计异常值是导致直方图法出现伪峰的直接原因。为了提高差分麦克风阵列的方位估计性能, 本文采用掩蔽方法以减小异常值的影响, 设定掩蔽频率范围为 $(0, f_L) \cup (f_H, f_s/2)$, 其中 f_s 为信号采样频率(负频率掩蔽范围与正频率对称)。考虑零频附近语音信号能量较弱, 易出现异常值, 本文设定掩蔽频率下限为 $f_L = 200 \sim 300$ Hz。对于掩蔽频率上限 f_H , 则与差分阵列的尺寸有关, 不同阵列尺寸所对应的掩蔽频率上限如表 1 所示。

表 1 不同差分阵列尺寸时的掩蔽频率上限选择

Table 1 Upper frequency bounds of time-frequency masking for various array sizes

阵列尺寸 D/m	0.02	0.04	0.06	0.08	0.10	0.12
掩蔽频率上限 f_H/Hz	$<4\ 000$	$<3\ 600$	$<3\ 300$	$<2\ 600$	$<2\ 000$	<1700

结合以上对瞬时方位估计进行时频掩蔽的思想, 为减小异常值的影响, 可以对直方图法进行改进, 即对时频掩蔽后的瞬时方位估计结果作直方图处理。为表示方便, 改进后的方法称之为修正直方图法。图 5 给出了与图 2 仿真条件相同时的修正直方图结果。对比可以看出, 修正直方图方法有效抑制了伪峰的出现, 在混响噪声混响下具有更好的方位估计性能。

3.2 多源瞬时方位估计的聚类分析

根据语音信号的时频稀疏性特征, 可以通过聚类分析对时频掩蔽后的瞬时方位估计值进行分离。本文采用模糊聚类方法^[16]。记声源的个数为 N , 各声源瞬时方位估计值所对应的时频点集合分别为 C_1, C_2, \dots, C_N 。通过最小化以下代价函数可实现多源方位估计的聚类分离

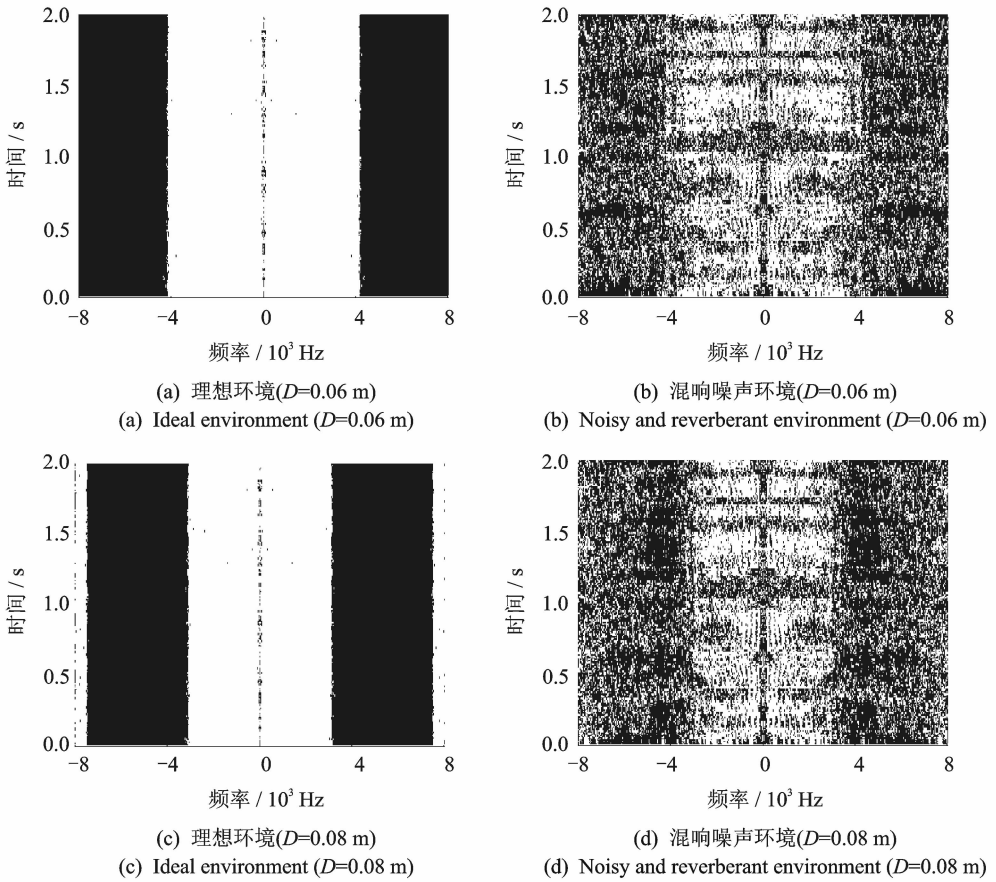


图4 不同阵列尺寸下,瞬时方位估计异常值的分布

Fig. 4 Outliers distribution of azimuth estimation with different array sizes

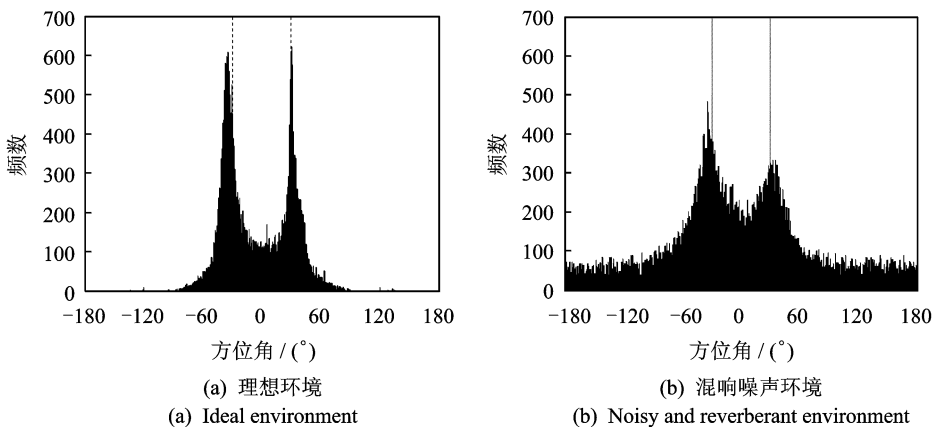


图5 修正直方图法的声源方位估计结果

Fig. 5 Results of the improved histogram method

$$J = \sum_{k=1}^N \sum_{\phi(t, f) \in C_c} u_k^q(t, f) \|\phi(t, f) - c_k\|^2 \quad (14)$$

式中: $\|\cdot\|$ 表示 2 范数; $q > 1$ 为控制聚类平滑度的参数。 $u_k(t, f)$ 表示某个时频点在第 k 类的隶属度, 表征了该时频点属于第 k 类的可能性大小

$$u_k(t, f) = \left[\sum_{i=1}^N \left(\frac{\|\phi(t, f) - c_k\|}{\|\phi(t, f) - c_i\|} \right)^{2/(q-1)} \right]^{-1} \quad (15)$$

式中: $0 \leq u_k(t, f) \leq 1$ 。特别地, 如 $u_k(t, f)$ 取 0 或 1, 则属于硬性划分聚类, 即退化为 k -均值聚类。与 k -均值聚类相比, 模糊聚类的优点是可以更加精确地描述各时频点归属为每个类别的可能性大小。定义将时频点归属为隶属度大于 0.5 的相应类别。相反地, 如果某个时频点在所有类中的隶属度都小于 0.5, 就舍弃该时频点。式(15)中, c_k 表示第 k 类的聚类中心可表示为

$$c_k = \frac{\sum_{\phi(t, f) \in C_k} u_k^q(t, f) \phi(t, f)}{\sum_{\phi(t, f) \in C_k} u_k^q(t, f)} \quad (16)$$

为了避免每次聚类次序产生混乱, 令 $c_1 < c_2 < \dots < c_N$ 分别对应逐渐增大的方位角, 类别序号越大, 对应的声源方位角越大。为了实现聚类, 需要对每个时频点的隶属度进行初始化, 可以采用一组满足在所有类中的隶属度之和为 1 的数作为初始隶属度; 然后通过迭代聚类中心 c_k 和隶属度矩阵 $u_k(t, f)$, 直到代价函数的变化量小于某个阈值为止, 即得到最终的聚类分析结果。

3.3 基于短时平均复声强的多源方位估计

通过对多源瞬时方位估计值的时频点聚类分离, 可以得到各声源对应的时频点的声强信息。采用图 3 中的符号定义, 对于第 k 个声源, 其短时平均复声强的两个正交分量可以表示为

$$\bar{I}_{ar}^{(k)} = \frac{1}{N_{C_k}} \sum_{\phi(t, f) \in C_k} I_{ar}(t, f), \bar{I}_{oy}^{(k)} = \frac{1}{N_{C_k}} \sum_{\phi(t, f) \in C_k} I_{oy}(t, f)$$

式中: N_{C_k} 表示声源 k 所属的类 C_k 包含的时频点数。则声源 k 的方位估计值可表示为

$$\bar{\phi}_k = \arctan \left(\frac{\operatorname{Re}\{\bar{I}_{ar}^{(k)}\}}{\operatorname{Re}\{\bar{I}_{oy}^{(k)}\}} \right) = \arctan \left(\frac{\operatorname{Re}\left\{ \sum_{\phi(t, f) \in C_k} I_{oy}(t, f) \right\}}{\operatorname{Re}\left\{ \sum_{\phi(t, f) \in C_k} I_{ar}(t, f) \right\}} \right) \quad (17)$$

与直方图方法相比, 式(17)具有闭式解, 避免了谱峰搜索。另外, 直方图方法只采用瞬时复声强得到声源方位估计, 每个时频点的处理是孤立的。而式(17)则利用了复声强估计原理, 将各个时频点的瞬时复声强进行联合处理, 因而对不相关加性噪声具有更好的抑制能力。

3.4 算法步骤

本文提出的基于时频掩蔽和模糊聚类分析的短时平均复声强多声源方位估计方法包含以下步骤:

(1) 对差分阵列测得的原点处声压和振速的正交分量进行短时傅里叶变换, 得到 $P_o(t, f)$, $V_{ar}(t, f)$ 和 $V_{oy}(t, f)$;

(2) 根据 3.1 节的分析, 由差分阵列的尺寸选择对应的时频掩蔽范围, 得到掩蔽后的声压和振速分量 $P'_o(t, f)$, $V'_{ar}(t, f)$ 和 $V'_{oy}(t, f)$;

(3) 由式(11, 12) 计算瞬时复声强的正交分量 $I_{ar}(t, f)$ 和 $I_{oy}(t, f)$, 并进而得到各时频点处的瞬时方位估计值 $\phi(t, f)$;

(4) 利用模糊聚类方法对各时频点处的瞬时方位估计值 $\phi(t, f)$ 进行聚类分析, 得到各声源瞬时方位估计值所对应的时频点集合 C_1, C_2, \dots, C_N ;

(5) 根据聚类结果, 对每一个声源进行短时平均复声强估计, 并由式(17)得到各声源的方位估计值。

4 仿真实验

通过混响噪声环境下的仿真实验研究本文提出方法的性能,并与直方图法和修正直方图法进行比较。分别从不同混响噪声环境、不同声源方位、以及不同阵列尺寸 3 个角度进行分析和比较。考虑房间的尺寸为 $6\text{ m} \times 6\text{ m} \times 4\text{ m}$,差分麦克风阵列的中心位于 $(3\text{ m}, 3\text{ m}, 1\text{ m})$ 处,且语音声源到阵列中心的距离为 2 m 。麦克风加性噪声为互不相关的白噪声。空气声速为 340 m/s ,信号的采样频率为 16 kHz 。短时傅里叶变换的帧长为 512 点,相邻帧间的重叠为帧长的一半,即 256 点。声源与麦克风之间的房间脉冲响应利用虚源法产生^[17]。对于模糊聚类分析,设定聚类平滑度控制参数 $q=2$,聚类迭代终止的阈值取为 1×10^{-5} 。

4.1 不同混响噪声环境下的性能

设两语音声源的真实方位为 -90° 和 90° 。所采用的差分阵列的尺寸为 0.04 m 。当信噪比分别为 20, 10 和 5 dB,混响时间分别为 0.16, 0.28 和 0.4 s 时,100 次蒙特卡洛实验对应的 3 种方法的声源方位估计的均方根误差如表 2 所示(无数据的情况表示均方根误差大于 20°)。

表 2 不同混响噪声条件下 3 种方法的方位估计均方根误差

Table 2 Root mean square error of azimuth estimation for the three methods with different environments

(信噪比/dB,混响/s)	声源 1/ $(^\circ)$			声源 2/ $(^\circ)$		
	直方图法	修正	本文方法	直方图法	修正	本文方法
		直方图法			直方图法	
(20,0.16)	3.113 3	1.160 3	0.161 4	0.553 7	0.553 3	0.201 9
(10,0.16)	6.493 6	7.005 1	1.040 0	3.858 1	0.998 9	1.164 4
(5,0.16)	4.636 1	7.638 1	2.788 4	10.249 4	7.758 8	2.981 6
(20,0.28)	1.696 3	0.527 0	0.369 8	0.598 5	1.638 5	0.421 7
(10,0.28)	1.910 6	1.574 0	2.228 3	9.307 2	3.432 0	2.364 3
(5,0.28)	14.625 2	9.437 3	4.915 8	5.307 2	5.475 5	5.103 8
(20,0.4)	4.133 0	0.630 8	0.616 8	5.696 0	1.214 5	0.595 8
(10,0.4)	12.904 3	10.441 2	3.552 4	6.200 1	3.546 8	3.253 9
(5,0.4)		12.380 5	7.189 4		8.342 5	6.630 5

由仿真结果可以看出,在信噪比较低、混响较大的情况下,直方图法性能变差,均方根误差超过了 20° 。而修正直方图法在低信噪比高混响下的性能有所改善。本文提出的方法在整体上要优于直方图法以及修正直方图法。特别地,在低噪声和高混响条件下的方位估计性能的优势更为显著。因此,相对于直方图法和修正直方图法,本文提出的方法对混响和噪声具有更好的鲁棒性。

4.2 不同声源方位时的性能

设差分麦克风阵列尺寸为 0.04 m ,房间混响时间为 0.4 s ,信噪比为 20 dB。两语音声源的方位角与 x 轴的夹角由 20° 到 160° 以 20° 间隔递增。对每一组声源方位情况,均作 100 次蒙特卡洛实验,3 种方法在不同声源方向时的均方根误差如表 3 所示(无数据的情况表示均方根误差大于 20°)。

由仿真结果可以看出,当两个声源的方位较接近时,如声源方位为 $(-20^\circ, 20^\circ)$ 和 $(-160^\circ, 160^\circ)$ 的情况,直方图法基本失效。此时由于混响噪声的存在,声源对应的直方图频数的峰值被其他角度所湮没,无法有效实现方位估计。相比直方图法,修正直方图法在声源方位较近时,性能一定程度上有所改善,如声源方位为 $(-160^\circ, 160^\circ)$ 的情况,但对于声源方位为 $(-20^\circ, 20^\circ)$ 情况,依然无法有效估计。比较

而言,本文提出的方法受声源方位变化的影响较小,比直方图法具有更高的角度分辨率,能够实现相近的两个声源目标的方位估计。

表 3 不同声源方向时 3 种方法的方位估计均方根误差 (°)

Table 3 Root mean square error of azimuth estimation for the three methods with different source locations

(声源 1 方位, 声源 2 方位)	声源 1			声源 2		
	直方图法	修正 直方图法	本文方法	直方图法	修正 直方图法	本文方法
(-20,20)		7.775 1	2.412 1			9.949 5
(-40,40)	7.033 1	2.028 3	7.323 9		12.549 6	6.040 7
(-60,60)	5.164 9	8.747 4	0.900 3	6.084 8	2.514 0	4.081 8
(-80,80)	6.068 3	4.972 6	0.893 2	0.941 8	6.629 4	2.263 2
(-100,100)	3.136 5	3.764 2	2.368 7	4.548 1	2.859 3	0.751 4
(-120,120)	7.412 6	3.120 9	0.329 9	2.087 5	4.058 1	2.773 9
(-140,140)	6.092 8	7.632 6	9.689 4		1.159 1	4.766 5
(-160,160)		4.211 7	7.438 1		15.139 1	10.758 8

4.3 不同阵列尺寸时的性能

考虑两个语音声源方位分别位于 -30° 和 30° 的情况,信噪比设为 10 dB,混响时间为 0.28 s。这里,本文研究阵列尺寸从 0.02 m 变到 0.12 m 时 3 种方位估计方法的性能。针对每一种阵列尺寸分别进行 100 次蒙特卡洛实验,各种方法的声源方位估计均方根误差如表 4 所示(无数据的情况表示均方根误差大于 20°)。

表 4 不同阵列尺寸条件下 3 种方法的方位估计均方根误差

Table 4 Root mean square error of azimuth estimation for the three methods with different array sizes

尺寸 D/m	声源 1/(°)			声源 2/(°)		
	直方图法	修正 直方图法	本文方法	直方图法	修正 直方图法	本文方法
0.02	3.858 1	2.257 0	0.662 3		10.249 4	5.729 0
0.04	4.012 0	2.119 5	1.242 8	16.101 8	8.230 3	4.982 1
0.06	2.340 7	2.289 3	1.937 2	12.306 9	6.766 8	2.430 5
0.08	2.484 0	2.449 3	1.665 3	9.984 7	4.317 6	2.282 8
0.10	4.611 1	2.795 6	2.719 1	8.625 5	5.729 0	3.111 0
0.12	5.802 4	2.463 2	2.922 0	7.618 4	4.633 1	3.535 1

3.1 节指出直方图法的性能受差分阵列尺寸的影响。差分阵列尺寸变化时,声源瞬时方位估计的异常值表现差异较大。因而,由表 4 可看出,直方图法的性能对阵列尺寸较敏感,不同阵列尺寸下的估计精度波动较大。而修正直方图法则考虑到了这一因素,采用时频掩蔽思想,有效抑制了阵列尺寸变大时所产生的瞬时方位估计异常值,因而比直方图具有更好的方位估计性能,受阵列尺寸的影响较小。比较而言,本文提出的方法受阵列尺寸的影响较小,声源方位估计性能在整体上要优于直方图法和修正直方图法。

5 结束语

本文研究了混响噪声环境下基于差分麦克风阵列的多声源方位估计问题。分析了直方图法存在的

缺点,结合时频掩蔽思想,提出了修正直方图方法。该方法有效提高了混响噪声环境下直方图法的方位估计性能,并且改善了直方图方法对于差分阵列尺寸的敏感度。为进一步提高混响噪声环境下差分阵列的声源定位性能,本文提出了一种基于时频掩蔽和模糊聚类分析的短时平均复声强多声源方位估计方法。与直方图法相比,该方法具有闭式解,无需峰值搜索。仿真结果表明,本文提出的方法在低信噪比和高混响条件下的性能要优于直方图类方法,并且具有较高的空间分辨能力、对阵列尺寸变化的敏感度小等优点。

参考文献:

- [1] 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.
Ju Tailiang. Research on speech source localization methods based on microphone arrays[D]. Chengdu: University of Electronic Science and Technology of China, 2006.
- [2] 林志斌,徐柏龄. 基于传声器阵列的声源定位[J]. 电声技术,2004(5):19-23.
Lin Zhibin, Xu Bailing. Sound source localization based on microphone array[J]. Audio Engineering, 2004(5):19-23.
- [3] 李晓飞,刘宏. 机器人听觉声源定位研究综述[J]. 智能系统学报,2012,7(1):9-20.
Li Xiaofei, Liu Hong. A survey of sound source localization for robot audition[J]. CAAI Transactions on Intelligent Systems, 2012,7(1):9-20.
- [4] DiBiase J H, Silverman H F, Brandstein M S. Microphone arrays: Signal processing techniques and applications[M]. Berlin: Springer-Verlag, 2001:157-180.
- [5] 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, Lin Xilin. Sound localization method using modified SRP-PHAT algorithm[J]. Journal of Electronics and Information Technology, 2006,28(7):1223-1227.
- [6] 周康辉,董万胜,刘恒毅. 利用二次相关改进的广义互相关时延估计算法[J]. 数据采集与处理,2013,28(6):801-806.
Zhou Kanghui, Dong Wansheng, Liu Hengyi. An improved generalized cross correlation method for estimation of time delay by using second correlation[J]. Journal of Data Acquisition and Processing, 2013,28(6):801-806.
- [7] He H, Wu L, Lu J, et al. Time difference of arrival estimation exploiting multichannel spatial-temporal prediction[J]. IEEE Trans Audio, Speech, Lang Process, 2013,21(3):463-475.
- [8] Benesty J, Chen J. Study and design of differential microphone arrays[M]. Berlin: Springer-Verlag, 2012.
- [9] Günel B, Hacıhabiboglu H, Kondoç A M. Acoustic source separation of convolutive mixtures based on intensity vector statistics[J]. IEEE Trans Audio, Speech, Lang Process, 2008,16(4):748-756.
- [10] Gore A, Fazel A, Chakrabartty S. Far-field acoustic source localization and bearing estimation using sigma-delta learners[J]. IEEE Trans Circuits and Systems I: Regular Papers, 2010,57(4):783-792.
- [11] Yilmaz O, Rickard S. Blind separation of speech mixtures via time-frequency masking[J]. IEEE Transactions on Signal Processing, 2004,52(7):1830-1847.
- [12] 惠俊英,惠娟. 矢量声信号处理基础[M]. 北京:国防工业出版社,2009.
Hui Junying, Hui Juan. The basic acoustic signal processing[M]. Beijing: National Defence Industry Press, 2009.
- [13] Zhong X, Chen X, Wang W, et al. Acoustic vector sensor based reverberant speech separation with probabilistic time-frequency masking[C]//Proc 21st European Signal Processing Conference (EUSIPCO). [S.l.]: IEEE, 2013:1-5.
- [14] Fahy F J. Sound intensity[M]. 2nd Ed. London: E & FN Spon, 1995.
- [15] Derkx R M M, Janse K. Theoretical analysis of a first-order azimuth-steerable superdirective microphone array[J]. IEEE Trans Audio, Speech, Lang Process, 2009,17(1):150-162.
- [16] Kühne M, Togneri R, Nordholm S. Robust source localization in reverberant environments based on weighted fuzzy clustering[J]. IEEE Signal Processing Letters, 2009,16(2):85-88.
- [17] Allen J B, Berkley D A. Image method for efficiently simulating small-room acoustics[J]. Journal of the Acoustical Society of America, 1979,65(4):943-950.

作者简介:何赛娟(1990-),女,硕士研究生,研究方向:音频与语音信号处理,E-mail:hesaijuan@nuaa.edu.cn;陈华伟(1977-),男,教授,研究方向:音频与语音信号处理、阵列信号及统计信号处理;尹明婕(1991-),女,硕士研究生,研究方向:阵列信号及统计信号处理;丁少为(1992-),男,硕士研究生,研究方向:音频与语音信号处理。