

分布式声传感器阵列校准方法综述

王 睿, 陈 喆, 殷福亮

(大连理工大学信息与通信工程学院, 辽宁大连 116024)

摘 要: 分布式声传感器阵列在军事侦察、公共安全监控和人机交互等领域应用广泛. 但时钟异步、频率响应失配和节点位置未知问题严重影响后续基于分布式声传感器阵列的语音处理算法的性能, 因此近年来学者们热衷于研究校准这三类阵列失配问题的方法. 首先, 本文介绍了导致分布式声传感器网络各类失配问题的原因和三类问题解决的先后顺序. 然后, 将现有的解决三类问题的方法进行了综述. 其中, 时钟同步方法主要包括基于时间戳交换和短时傅里叶变换域线性相位漂移的方法, 频率响应校准方法主要包括基于声源导向矢量和自适应滤波器的方法, 节点位置/位姿校准方法则主要包括基于节点间几何构型和深度学习的方法. 最后, 给出了本领域未来研究的方向.

关键词: 分布式声传感器阵列; 阵列失配; 阵列校准; 位置校准; 频率响应校准; 时钟同步

中图分类号: TN912.2 **文献标识码:** A **文章编号:** 0372-2112(2021)12-2468-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20201088

Review on Calibration of Distributed Acoustic Sensor Arrays

WANG Rui, CHEN Zhe, YIN Fu-liang

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

Abstract: Distributed acoustic sensor arrays have been widely employed in many applications, such as military reconnaissance, security monitoring, man-machine interaction, etc. However, the problems of clock asynchrony, frequency response mismatch and unknown node location seriously affect the performance of subsequent speech processing algorithms based on distributed acoustic sensor arrays. Therefore, scholars have been keen to study calibration methods for these three problems in recent years. This article first introduces the causes of these mismatch problems and the sequence of solving the three problems. Then, the existing methods for these three problems are reviewed. Specifically, clock synchronization methods mainly include methods based on timestamp exchange and linear phase shift in short-time Fourier transform domain, frequency response calibration methods mainly include methods based on steering vector of sound source and adaptive filter, and node position/pose calibration methods mainly include methods based on geometric configuration among nodes and deep learning. Finally, the future research directions are provided.

Key words: distributed acoustic sensor array; array mismatch; array calibration; position calibration; frequency response calibration; clock synchronization

1 引言

近年来随着移动科技的发展, 出现了一批以智能手机为代表的便携设备, 如手机、笔记本电脑、平板电脑(如 iPad)、PDA、智能可穿戴设备(如 iWatch)等, 这些设备常配备有音频输入输出接口. 利用这些便携设备, 很容易构成 Ad Hoc 分布式声传感器网络^[1]. 此外, 分布式声传感器阵列在现代军事侦察、公共安全监控以及人机交互等领域也有广泛的应用前景^[2]. 分布式

声传感器阵列在以下几方面已经表现出优异性能:

(1) 阵列建立容易, 不必预先考虑和设计阵列阵元的几何位置坐标和拓扑结构; (2) 覆盖空间范围大; (3) 具有分布式计算能力; (4) 容错性较强. 但也正因为分布式声传感器阵列的以上特点, 为其相关算法的研究带来了挑战^[3]: 例如, (1) 分布式声传感器网络中节点位置一般为随机布放, 节点之间距离往往大于音频信号波长, 不满足采样定理, 常规声传感器阵列相关

收稿日期: 2020-09-30; 修回日期: 2020-12-17; 责任编辑: 覃怀银

基金项目: 国家自然科学基金(No.61771091, No.61871066); 国家 863 高技术研究发展计划(No.2015AA016306); 辽宁省自然科学基金(No.20170540159); 中央高校基本科研业务费(No.DUT17LAB04)

的算法对其并不适用;(2)各个节点具有独立的时钟系统,时钟异步问题会严重影响基于时间信息的后续信号处理算法的性能;(3)很多用于分布式声传感器阵列的算法都依赖于节点的空间位置,而分布式声传感器网络中的节点位置在阵列布置前通常未知;(4)各节点中声传感器的内部电路很可能各不相同,声传感器之间的频率响应失配问题会严重影响后续基于各声传感器输出信号幅度和相位特征的信号处理算法的性能.因此,在将分布式声传感器网络用于语音增强、识别、定位和跟踪之前,需要对其节点间的时钟异步、频率响应失配和节点位置进行校准,以避免由于这三类失配问题导致的后续算法处理效果不佳等问题^[4].从时钟异步、频率响应失配和位置校准三个问题的产生机理来看,在无法同时进行校准时,应按照时钟同步、频率响应校准及位置校准的顺序依次进行更为合理.

本文综述近年来分布式声传感器阵列校准相关研究的最新进展与成果,直观描述了分布式声传感器阵列中的各类失配问题,详细阐述了现有基于分布式声传感器阵列的时钟同步、节点位置校准以及频率响应校准方法的基本原理、主要特点以及彼此间的区别与联系,分析了分布式声传感器阵列校准相关技术面临的挑战与未来发展方向,希望为相关领域的后续研究提供有益参考.

2 分布式声传感器阵列时钟同步方法

时钟同步的目的就是对各个节点的本地时钟提供一个统一的时间标尺^[5,6].现有分布式声传感器阵列的时间同步算法主要分两类,一类是利用分布式节点之间的时间戳传输来校准采样率差异^[7-10],另一类为采样率偏差盲估计方法^[11-18].根据原理,盲估计方法可分为两类,第一类采用短时傅里叶变换(Short Time Fourier Transform, STFT)域中的各声传感器信号间的线性相位漂移对采样率失配问题进行建模^[11-17],第二类利用时域互相关之间的偏差盲估计采样率偏差^[18].

2.1 基于时间戳传输的时钟同步方法

张九宾等^[7]对时间统一基准性和时钟频率的校准进行重点研究,主控单元通过GPS秒脉冲沿触发并无线发送授时命令,测试单元采用软硬件结合的机制实现时间统一基准信号的产生,并根据基准信号对时钟频率进行校准,从而达到对随机事件触发的同步采样.在文献[8]中,一种双路时间信息交换方法被提出.Schmalenstroer等^[9]基于时间戳传递和卡尔曼滤波器来估计节点间的时钟抖动,该方法需额外附加中心处理器.因此,该方法后续被改进为采用Gossip策略分布式地同步各节点的采样率^[10].然而,这些基于

时间戳的算法的准确性通常与定时消息传输的数量成正比^[19],这将引入较大的节点间通信链路负担.因此,基于各声传感器接收信号的各类采样率盲估计方法被提出.

2.2 基于STFT域线性相位漂移的采样率校准方法

节点间采样率偏差会在STFT域呈现线性相位漂移,因此通过估计该漂移能够校准节点间的采样率偏差.现有基于STFT域线性相位漂移的采样率偏差盲估计方法大致分为两类:(1)基于相关系数的方法^[11-13],(2)基于瞬时相关系数的方法^[14-17].

2.2.1 基于相关系数的采样率校准方法

由于节点间的采样率偏差会在STFT域呈现线性相位漂移,节点接收信号之间的相关系数或协方差函数存在一个固定的相位偏差.Wang等^[12]提出一种基于最大相关系数的采样率校准方法,该方法通过在采样率偏差可能范围内穷尽搜索能够使两节点接收信号之间的相关系数取得最大值的解作为采样率偏差的估计值.文献[11]中的方法与Wang等的方法^[12]类似,但代价函数表示方法不同^[12].Cherkassky等^[13]通过最大协方差函数来估计采样率偏差,该方法通过梯度上升法迭代搜索能够使节点接收信号之间的协方差函数最大的解.该类方法的采样率校准精度与搜索步长有关,步长较小时其精度很高.但该类方法需要在预定义的采样率偏差范围内进行搜索,并且在搜索过程中不断计算信号之间的相关系数,导致计算复杂度较大.此外,该方法往往需要已知帧数索引值 l ,而 l 在对信号“盲”的条件下不易获取.

2.2.2 基于瞬时相关系数的采样率校准方法

文献[14~17]假设声传感器接收信号的前后帧之间具有较强的稳定性,则前后帧之间的瞬时相关系数之比的相位与采样率偏差呈正比.因此,直接估计前后帧瞬时相关系数之比的相位,可获得瞬时采样率偏差估计值^[16].文献[14,17]对多帧信号的瞬时采样率偏差估计结果求平均,以此提高采样率偏差的估计精度.Bahari等^[15]先通过瞬时采样率偏差估计结果排除多帧估计结果中的异常值,然后对剩余结果求平均.该方法不需要任何关于信号的先验信息,但由于接收信号前后帧之间的稳定性较差,因此该类方法的采样率偏差估计精度较基于相关系数的方法稍差.

2.3 基于双重互相关性的采样率校准方法

与STFT域采样率校准方法不同,文献[18]提出一种基于二次互相关的时域采样率偏差估计方法.节点间的采样率偏差会在STFT域呈现线性相位漂移,那么在时域上表现为随时间累积而线性增加的时间漂移.因此,该方法首先求取两两节点接收信号之间的瞬时互相关,然后对相邻帧的互相关结果再次进行互相关,

该结果与采样率偏差成正比。最终,对各帧互相关的结果在时间上求平均获取采样率偏差估计值。实际应用中,该方法的精度较基于相关系数的方法稍差但其运算量较低。

2.4 采样率偏差补偿方法

重采样就是将采样频率为 f_s 的信号按照参考采样率 f_s^{ref} 进行插值。文献[14, 20]采用时域四阶 Lagrange 插值算法来进行重采样,首先对信号进行四倍插值得到插值后的信号,然后重采样信号计算为原采样点附近的四个插值时间点对应信号的加权平均。文献[21]中,直接采用 sinc 函数对各声传感器的输出信号插值补偿采样率偏差。文献[22]通过实验对比发现只要保证有足够长的窗尺寸, sinc 函数插值方法的补偿精度很高。然而,大的窗函数也对应着大的计算量和较长的运算时间,因此 sinc 方法不适合实时补偿采样率的场景。

3 分布式声传感器阵列频率响应校准方法

频率响应校准的目的是通过信号处理方法来弥补由于声传感器内部电路导致的声传感器输出信号之间幅度和相位的差异。现有的声传感器频率响应校准方法可以分为三类,一类通过估计各声传感器相对于参考声传感器的增益校正因子来校准声传感器增益失配^[23-26],第二类通过分析频率响应偏差对声源导向矢量幅度和相位的影响估计频率响应失配^[27],第三类由声传感器输出信号之间最小均方误差(Minimum Mean Square Error, MMSE)获得各声传感器的频率响应校准滤波器^[28, 29]。

3.1 声传感器增益校准方法

增益失配实际上是频率响应失配问题的一种简化形式,增益校准能够快速地对声传感器增益之间的失配,但忽略了相位失配的影响。

3.1.1 基于时间平均功率谱的增益校准方法

Madhu 和 Martin^[23]通过最小化各声传感器与参考声传感器输出信号的功率谱之差来估计每个声传感器对应的增益校准因子 g_i ,通过令代价函数对 g_i 求偏导并令其等于零,得到最小均方误差方法的维纳解。增益校准后,在各声传感器的输出信号前乘上对应的校准因子得到校准后的输出信号。

3.1.2 基于能量的增益校准方法

Tashev^[25]发现在到达方向(Direction Of Arrival, DOA)线上可通过线性插值来获得各声传感器输出信号能量的估计值,因此实际能量值与该估计值之比可作为声传感器增益的校准因子,也可以结合前后帧增益来进行实时校正。但该方法需要事先已知声源相对于节点的 DOA。Gaubitch 等^[24]利用各声传感器与参考

声传感器接收信号之间的能量比计算相对增益,则各声传感器的增益校准因子为相对增益的倒数。增益校准时,在各声传感器的输出信号前乘上对应的校准因子得到校准后的输出信号。但该类方法需要事先获取各声传感器与各声源之间的距离。

3.1.3 基于长周期平均功率的增益校准方法

Hua 等^[26]在长周期上对功率进行平均,得到增益校准因子。校准增益时,各声传感器的输出为原输出信号与校准因子之间的 Hadamard 乘积。该方法在估计声传感器输出信号之间的功率偏差时,在时间上对信号的功率进行了平均,因此它较前面所述方法具有更好的增益校准性能。

3.2 基于声源导向矢量的频率响应校准方法

实际上,一些声传感器阵列,例如助听器中的超指向阵列或差分阵列,对相位失配非常敏感^[30],因此不能忽略对声传感器相位失配的校准。

对于近场环境下的单个声源,彭启宗^[27]等将声传感器的相位和幅度频率响应偏差估计为声源实际导向矢量与采用多信号分类(Multiple Signal Classification, MUSIC)算法中信号子空间的特征矢量表示的导向矢量在相位和幅度上的偏差矢量。该方法适用于在一个平面上以任意几何形状布放的固定声传感器阵列,但需要精确已知声源在各处摆放的位置以及参考声传感器的位置。

3.3 自适应频率响应校准方法

呼德等^[28]和王睿等^[29]分别提出采用均衡器和频域自适应滤波器来校准各声传感器的频率响应。首先,将各声传感器彼此贴近放置,确保它们的接收信号相似。然后,两种方法分别将各声传感器输出信号作为均衡器和自适应滤波器的输入。最后,将代价函数分别构造为各声传感器输出信号与它们的均值之间的 MMSE 和成对自适应滤波器输出信号的 STFT 之间的联合 MMSE,前者直接求取代价函数的维纳解,而后者则通过梯度下降法迭代求解。该类方法需要保证各声传感器接收信号相似,可以通过聚拢放置声传感器来满足这一要求。但在校准大量声传感器时,即使是聚拢放置也难以保证此类方法的校准精度,因此该类方法更适合于离线校准少量声传感器频率响应的场景。

4 分布式声传感器节点位置/位姿校准方法

分布式声传感器阵列各节点的位置/位姿通常无法作为先验知识来使用^[31]。现有分布式声传感器阵列节点位置/位姿的校准方法主要分为两类,一类是基于节点间几何构型代价函数的方法^[24, 32-57],另一类是基于深度学习的方法^[58]。其中,第一类方法主要包含两类方案,一类是基于优化算法求解关于节点位置/位姿的代

价函数^[32-47];另一类是基于矩阵降维算法求取节点位置/位姿封闭解^[24,48-57].

法的观测量、约束条件、是否定位声源以及核心算法等信息,各文献中的具体内容在下面章节进行展开讨论.

表 1 中总结了大部分现有节点位置/位姿校准方

表 1 现有节点位置/位姿校准方法总结

文献	位置/位姿	观测量	同步/异步	声源	约束	方法			
[32]	位置	TDOA	异步	√	-	理论和测量 TDOA 的 MMSE	基于辅助函数的优化方法		
[33]	位姿	PD, TDOA, α			-		PD+MDS 获取初始位置+ITRR		
[34]	位置	PD, TDOA			TDOA 估计 PD 和 δ				
[42]	位姿	DOA, TDOA	同步	×	声源到各节点的距离已知	理论和测量 TOA 的 MMSE	差分进化算法		
[35]	位置	TOA, r_s			节点间同步且声源位置可测		Simplex 优化方法		
[36,37]	位姿	DOA	异步	√	说话人需在多处停留并发声	理论和测量 DOA 的 ML	牛顿寻根法		
[43]					×	已知声源位置和节点内拓扑结构	理论和测量 DOA 的 MMSE	人工蜂群算法	
[38,39]					√	代价函数求解易陷入局部极小值	二维 DOA 与位姿关系的代价函数	牛顿寻根法	
[45]					√	需已知摄像机位置和安装角度	视听联合系统方程	三维平移旋转关系	
[46]					×	摄像机可获取声源的视觉 DOA			
[40]					DOA, r_s	√	附加摄像机,已知 d	理论和测量声源位置的 ML	BFGS 优化方法
[41]					DOA		计算量高	理论和测量声源位置的 ML	
[44]					$[x_s^L, y_s^L, z_s^L]$		要求参考麦克风和三个可控声源	三角定位算法	
[47]	位置	TDOA, d	同步	√	声源与参考节点重叠	距离平方差矩阵	SVD 降维	-	
[48,49]	TOF	异步					声源信号可控	最小范数+SVD 和 QR 分解	
[25,50]	TOA						天线位置及声源集拓扑结构已知	MMSE+ACD 求相对位置	平移旋转关系+Kabsch
[52]	位置	TOF, PD	同步	√	-	与 SVD 降维矩阵的 MMSE	时间信息矩阵的 STLS		
[57]	位姿	TOA					异步	×	弥漫噪声场且节点内拓扑结构已知
[53]	位置	PD	大部分 PD 可通过相干性拟合获得	OPTSPACE 算法补全距离矩阵	MDS				
[55]	位置	分散度+R 矢量	已知相对直达传递函数	非负距离矩阵分解算法补全距离矩阵	MDS				
[56]			节点内拓扑结构已知	理论与测量相干矩阵拟合获得 PD 矩阵+MDS					
[58]	位置	分散度+R 矢量	节点内拓扑结构已知	分散度+R 矢量作为 DNN 输入求声源-节点间距+DOA 定位方法					

注: α :节点相对于全局坐标系的旋转角度; \mathbf{r}_s :声源位置矢量; $[x_s^L, y_s^L, z_s^L]$:节点局部坐标系内声源的坐标; d :声源到节点的距离; δ :声传感器的内部延迟.

4.1 基于节点间几何构型代价函数的方法

4.1.1 基于优化方法的节点位置/位姿校准方法

4.1.1.1 基于理论与测量值之间代价函数优化的方法

(1) 基于到达时间差的方法

节点之间异步时,各声传感器内部系统延迟彼此不同且未知.声源和节点位置的联合估计问题可建模为声源到各声传感器的到达时间差(Time Difference Of Arrival, TDOA)理论值与测量值之间的最小范数问题^[32],其中未知量包含声源和各声传感器的位置以及各声传感器内部系统延迟. Ono等^[32]采用基于辅助函数的优化方法来求解该代价函数,该方法与最大期望算法原理上类似,能够获得比基于梯度的优化算法更好的收敛特性. Hennecke等^[33]和 Pertilä等^[34]建立的代价函数与 Ono等^[32]相似,但前者假设每个节点内都附加有扬声器和地磁罗盘且节点内部扬声器到声传感器的间距已知.则节点内置扬声器的位置可采用地磁罗盘测量到的节点旋转角度、间距和节点内置麦克风位置来表示.此时,剩余未知量只有节点的位置和内部系统延迟. Hennecke等^[33]先采用节点收发信号之间的TDOA来获取成对节点之间的距离(Pairwise Distance, PD)并采用多维标度分析(Multiple Dimensional Scaling, MDS)算法^[59]粗略估计节点初始位置,再利用迭代信赖域反射(Iterative Trust-Region-Reflective, ITRR)算法^[60]求解该代价函数,避免陷入局部极值点. Pertilä等^[34]在每对声传感器连线的两端先后放置声源,并测量出两声源到达两声传感器之间的时间差,以此求解出两两声传感器之间的内部系统延迟偏差^[61]和距离^[62],并采用MDS方法求解出各声传感器的初始位置.然后,采用主成分分析得到各声源对应的TDOA值,并将其代入代价函数中,采用ITRR算法^[60]求解节点的精确位置.

Joshua等^[35]测量多个声源到各声传感器的传输时间和声源的位置,并将声传感器位置的代价函数建立为声源到声传感器间距与到达时间(Time Of Arrival, TOA)之间关系的MMSE,然后采用simplex方法^[63]求解该代价函数.但声源到各声传感器的传输时间往往不易精确估计,因此影响该方法的估计精度.

(2) 基于DOA方法

Kemper等^[38]和Schmalenstroer等^[39]根据笛卡尔坐标系下声源相对于节点内部坐标系的DOA与声源和节点位置以及节点旋转角度之间的关系,先后建立两种声传感器二维位姿的代价函数.第一种^[38]代价函数的非线性程度较高,严重影响牛顿法求解的收敛性能^[39].第二种^[39]的稳定性较第一种有所提高,但其在 $[-\pi, \pi]$ 范围内仍存在三个最小值,求解过程陷入局部极小值

的可能性仍然存在^[36].因此, Jacob等^[36,37]建立关于节点的位置和角度以及声源的位置的最大似然函数.该代价函数在 $[-\pi, \pi]$ 范围内仅存在一个最优解,成功避免了陷入局部极小值的情况^[36].混响时间 $T_{60} = 50 \sim 450\text{ms}$ 时,该方法获得了 $0.05 \sim 0.29\text{m}$ 的位置估计偏差.但说话人需要在多个位置处停留并说话以获取足够的观测DOA.

Plinge等^[40,41]采用Broyden-Fletcher-Goldfarb-Shanno (BFGS)算法^[64]最小化声源到节点矢量表达式两侧函数的 L_2 范数误差来获取节点的二维位姿.文献[41]中采用基于DOA的说话人跟踪方法^[65]来获取至少三个声源的位置并代入该代价函数中,然后求解节点位姿.为了减少依赖于声信号求解的未知数个数,文献[40]中根据摄像机所在位置与拍摄角度之间的关系获得说话人的位置.

(3) 基于TDOA和DOA的方法

当假设各节点时钟同步时,相当于各声传感器内部系统延迟已知. Plinge等^[42]利用声源到各节点的矢量表达式将声源位置矢量表示为各表达式的累加,然后通过声源位置矢量表达式、声源相对于各节点的DOA测量值和TDOA测量值建立关于节点位姿的代价函数.最后,采用差分进化算法^[66]快速求解该代价函数.该方法融合了分布式声传感器网络中各节点的TDOA和DOA测量值,有效提升了节点位姿的校准精度.

(4) 三维空间内节点的位姿校准方法

三维空间内的节点位姿校准问题更为严峻,由于未知数数量的增多,对观测量的需求也更多,并且测量误差使得节点校准精度更难保障.

王睿等^[43]根据节点内局部坐标系与全局坐标系之间的旋转平移关系建模声源的理论DOA,通过节点内三个声传感器之间的TDOA测量值获取测量DOA,将三维空间内节点位姿校准的代价函数建立为理论与测量DOA之间的MMSE,并采用人工蜂群算法来对其进行求解.该方法位姿校准精度高但算法迭代所需时间较长并且需要三个已知位置的声源,因此更适用于离线校准的场景.

Valente等^[44]在两节点坐标系内分别测量一个声源的位置,并利用节点内局部坐标系之间的旋转平移关系建立两个位置之间的关系.声源位置的测量值和理论值之间的偏差服从正态分布,因此可将三维节点位姿问题表示为对应的似然函数^[44],搜索令该似然函数最大的解,即为节点相对位姿.但由于搜索维度过大, Valente等^[44]先将节点位姿固定,搜索声源位置及其误差方差,再反向估计节点位姿,并重复该迭代过程.

4.1.1.2 基于视听联合系统方程的方法

为了获得节点位置且不必进行节点间的时钟同步, Jacob 等^[45,46]提出两种基于视听相关性的全局声传感器位置校准方法. 第一种方法^[45,46]将声传感器节点估计出的说话人轨迹与视频系统估计出的说话人轨迹相映射, 已知视频系统的安装位置及拍摄角度, 根据三维平移旋转关系得到节点的位姿. 第二种方法^[46]将视觉 DOA 作为视听联合系统方程的输入并解出声传感器的位置. 实验证明第二种方法的性能优于第一种方法, 即使在混响环境中也可实现 0.20m 的校准误差. 但是因为摄像机的引入, 使节点位姿校准的成本飙升.

4.1.1.3 基于三角定位的方法

文献[47]中假设在已知位置放置参考声传感器, 根据各声传感器所在位置的声信号能量可粗略计算声源与各声传感器的间距及其对应的 TDOA. 然后, 采用 GCC 函数在该 TDOA 附近快速搜索各声传感器与参考声传感器之间的精细 TDOA 及其对应的距离. 最后, 根据三角定位算法获取各声传感器所在位置. 然而, 环境条件恶劣时 GCC 函数往往无法精确估计 TDOA, 三角定位算法也会因此呈现较差的估计结果.

4.1.2 基于矩阵降维算法的节点位置校准方法

关于节点位姿的代价函数往往非线性较高, 未知数维度较高, 导致求解过程常常陷入局部极小值. 因此, 相关学者提出将代价函数表示为对应的矩阵形式, 并采用矩阵降维方法, 降低未知数维度, 从而简化代价函数的求解.

(1) 基于距离平方差矩阵降维的方法

Crocco 等^[48,49]假设一个声源与参考节点在全局坐标系的原点处重叠且声源到各节点的传输时间可测, 则各声源到各声传感器的间距和到参考声传感器的间距平方之差构成距离矩阵, 对该矩阵奇异值分解 (Singular Value Decomposition, SVD) 和正交三角 (Quadrature Rectangle, QR) 分解减少矩阵内未知数个数. 最终, 将估计声源和节点位置问题转化为求解各分解矩阵之间线性 MMSE 问题. 该方法能够有效避免基于梯度的优化方法陷入局部极小值, 但该方法隐含着已知各声信号发射时间以及各声传感器的内部延迟.

因此, Gaubitch 等^[24,50]提出两阶段声源节点联合定位的方法, 第一阶段解决节点间的时钟同步问题 (即估计声信号发射时间和声传感器内部系统延迟), 仍采用矩阵降维方法; 第二阶段采用文献[48,49]中方法解决声源节点定位问题. Wang 等^[51]同样需要先求解时间信息, 但在构建代价函数时采用结构总体最小平方 (Structured Total Least-Square, STLS) 构架, 采用 Gauss-Newton

算法能够有效求解该代价函数. 在获取时间信息后, 声源和节点定位问题仍可采用文献[48,49]中方法进行描述, 并对距离矩阵进行奇异值分解. 不同的是, 该方法不需要声源与参考节点重叠, 而是假设参考节点位于原点且参考声源在横坐标上. 因此, 未知量包括混合矩阵中的元素和参考声源位置. 利用距离矩阵与时间矩阵之间线性 MMSE 构建代价函数并采用 Gauss-Newton 算法求解出各未知量.

Le 等^[52]提出基于传输时间 (Time Of Flight, TOF) 获得距离矩阵, 并通过距离矩阵、声源位置矩阵和节点位置矩阵之间的最小 Frobenius 范数获得声源传感器联合定位代数封闭解. 但为了获得精确的 TOF, 该方法要求每个声源和所有传感器之间均完美同步、声源可控并且各声源发出的信号在时间上或频率上都不能重叠.

(2) 基于接收噪声相干性测距的方法

Hennecke 等^[53]提出一种无监督的节点位置校准方法, 该方法将节点校准过程分为两步, 第一步假设噪声场为弥散噪声场, 两声传感器的间距可以通过拟合测量相干性与理论相干性来获取, 然后通过 MDS 算法校准节点内声传感器间的几何结构. 第二步校准节点间的相对位置, 测量 N 次声源在两节点局部坐标系内位置, 并对它们形成矩阵对应的分散矩阵^[35]求 SVD, 通过 SVD 矩阵可以表示两节点坐标系之间的平移矢量和旋转矩阵. 该方法中的步骤是在完全无监督状态下进行的, 但需要满足弥散噪声场的假设.

Schwartz 等^[54]利用相干性矩阵与直达声和混响功率谱密度之间的关系迭代估计相干性矩阵. 然后, 将其与理论相干性矩阵进行拟合, 得到两两声传感器之间的距离. 最后采用 MDS 算法校准各声传感器的位置. 该方法假设相对直达传递函数已知, 混响会影响该方法的校准精度.

(3) 基于距离信息补充的方法

Taghizadeh 等^[55,56]同样通过拟合测量相干性与理论相干性来获取两节点之间的距离, 但这种方法常常出现异常观测值. 因此, 在采用 MDS 算法构造欧氏距离矩阵时往往需要先将这些异常观测值剔除, 而缺失的距离信息也会使 MDS 算法求解节点位置的性能下降. 所以, Taghizadeh 等^[55,56]先后采用基于 OPTSPACE 算法^[67]和非负距离矩阵分解算法^[68]来补充欧氏距离矩阵中缺失的信息.

(4) 基于特殊场景的方法

Bouley 等^[57]假设一种特殊的应用场景, 一个大型声传感器阵列正对着天线装置, 并且空间内分布着多个声源集. 采用多维展开 (Multi-Dimensional Unfolding, MDU)^[69], 搜寻声源、声传感器和天线位置, 使其对

应的欧氏距离矩阵与观测距离矩阵最相近,并采用交替坐标下降(Alternating Coordinate Descent, ACD)算法^[70]分布式求解该代价函数,但各节点的绝对角度和位置仍未知.通过MDU算法求解出锚点的相对位置,则求解旋转矩阵的代价函数建立为旋转矩阵乘以相对位置与全局位置之间的RMSE,再利用Kabsch算法^[71,72]基于相对位置和全局位置的互协方差矩阵的SVD来求解该代价函数.该方法可获得0.4cm~0.8cm的坐标估计误差和0.1度~0.7度的旋转角度估计误差.但由于需要天线装置和特殊几何构型的声源集的参与,因此限制了该方法的应用范围.

4.2 基于深度学习的节点位置校准方法

近些年,深度学习技术常应用于语音识别、语音增强、人机交互等领域.Gburrek等^[58]将深度神经网络(Deep Neural Network, DNN)应用到节点位置校准问题中,采用分散度和R矢量^[73]作为DNN的输入特征,训练时输出端为声源与节点之间的距离,并结合DOA测量值获取节点位置.文献^[58]中采用两类DNN,即多层感知网络和卷积递归神经网络,区别在于对于时间信息的使用.后者利用了每帧信号的分散度中包含的时间信息,而前者对分散度在时间上求平均,舍弃了时间信息.该方法需要粗略地同步各传感器节点,节点内部必须包含多个声传感器组成的阵列.

5 未来挑战与展望

分布式声传感器阵列校准是音频信号处理领域的重要研究课题.现有分布式声传感器阵列校准研究已取得极大进展,但仍面临很多挑战:(1)节点位置/位姿实时校准;(2)仅采用节点内信息来校准节点的位置/位姿避免节点间信息传输;(3)利用校准信号的特征校准各节点之间的采样率偏差,提高采样率补偿方法对环境干扰的鲁棒性;(4)研究频率响应失配对分布式阵列信号处理技术的影响,并对频率响应失配在线校准.

6 结束语

阵列校准是分布式声传感器阵列信号处理领域的一项重要研究课题,在语音增强识别、声源跟踪定位、盲源分离等领域具有广阔应用前景.本文综述了分布式声传感器阵列校准的研究进展,阐述了现有的分布式声传感器阵列时钟同步、频率响应校准以及位置校准的各类方法,并对性能突出的方法进行了详细的说明,总结了各方法的优缺点.最后,分别介绍了现有分布式声传感器阵列校准方法的局限性,指出了分布式声传感器阵列校准研究所面临的主要挑战和未来发展方向.

参考文献

- [1] Liu H, McLachlan D, Wang D. Overview of wireless microphones—Part I: System and technologies [J]. IEEE Transactions on Broadcasting, 2015, 61(3): 494–504.
- [2] Bertrand A. Applications and trends in wireless acoustic sensor networks: A signal processing perspective [A]. IEEE Symposium on Communications and Vehicular Technology [C]. Ghent, Belgium: IEEE Vehicular Technology/Communications Society, 2011. 1–6.
- [3] Thiergart O, Galdo G D, Taseska M, Habets E A P. Geometry-based spatial sound acquisition using distributed microphone arrays [J]. IEEE Transactions on Audio Speech and Language Processing, 2013, 21(12): 2583–2594.
- [4] Doclo S, Moonen M. Superdirective beamforming robust against microphone mismatch [J]. IEEE Transactions on Audio Speech and Language Processing, 2007, 15(2): 617–631.
- [5] Simeone O, Spagnolini U, Scutari G, Bar-Ness Y. Physical-layer distributed synchronization in wireless networks and applications [J]. Physical Communication, 2008, 1(1): 67–83.
- [6] Zhong X, Hopgood J R. A time-frequency masking based random finite set particle filtering method for multiple acoustic source detection and tracking [J]. IEEE/ACM Transactions on Audio Speech, and Language Processing, 2015, 23(12): 2356–2370.
- [7] 张九宾, 张丕状, 杜坤坤. 无线分布式测试系统时间统一技术的研究[J]. 核电子学与探测技术, 2010, 30(3): 380–384.
Zhang Jiubin, Zhang Pizhuang, Du Kunkun. The research of wireless distributional test system time synchronization technology[J]. Nuclear Electronics and Detection Technology, 2010, 30(3): 380–384. (in Chinese)
- [8] Chaudhari Q M. A simple and robust clock synchronization scheme [J]. IEEE Transactions on Communications, 2012, 60(2): 328–332.
- [9] Schmalenstroer J, Haeb-Umbach R. Sampling rate synchronization in acoustic sensor networks with a pre-trained clock skew error model [A]. European Signal Processing Conference [C]. Morocco: IEEE Signal Processing Society, 2013. 1–5.
- [10] Schmalenstroer J, Jebramcik P, Haeb-Umbach R. A gossiping approach to sampling clock synchronization in wireless acoustic sensor networks [A]. IEEE International Conference on Acoustics, Speech and Signal Processing

- [C]. Florence: IEEE Signal Processing Society, 2014. 7575 – 7579.
- [11] Miyabe S, Ono N, Makino S. Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain[A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Vancouver, Canada: IEEE Signal Processing Society, 2013. 674 – 678.
- [12] Wang L, Doclo S. Correlation maximization-based sampling rate offset estimation for distributed microphone arrays [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(3): 571 – 582.
- [13] Cherkassky D, Markovich-Golan S, Gannot S. Performance analysis of MVDR beamformer in WASN with sampling rate offsets and blind synchronization[A]. European Signal Processing Conference [C]. Nice, France: IEEE Signal Processing Society, 2015. 245 – 249.
- [14] Markovich G S, Gannot S, Cohen I. Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming[A]. International Workshop on Acoustic Signal Enhancement [C]. Aachen, Germany: IEEE Signal Processing Society, 2012. 1 – 4.
- [15] Bahari M H, Bertrand A, Moonen M. Blind sampling rate offset estimation based on coherence drift in wireless acoustic sensor networks[A]. 2015 European Signal Processing Conference [C]. Nice, France: IEEE Signal Processing Society, 2015. 2281 – 2285.
- [16] Bahari M H, Bertrand A, Moonen M. Blind sampling rate offset estimation for wireless acoustic sensor networks through weighted least-squares coherence drift estimation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(3): 674 – 686.
- [17] Schmalenstroer J, Heymann J, Drude L, Boeddecker C, Haeb-Umbach R. Multi-stage coherence drift based sampling rate synchronization for acoustic beamforming [A]. IEEE International Workshop on Multimedia Signal Processing [C]. Luton, UK: IEEE Press, 2017. 1 – 6.
- [18] Chinaev A, Thüne P, Enzner G. A double-cross-correlation processor for blind sampling rate offset estimation in acoustic sensor networks[A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C], Brighton, UK: IEEE Signal Processing Society, 2019. 641 – 645.
- [19] Zeng Y, Hendriks R C, Gaubitch N D. On clock synchronization for multi-microphone speech processing in wireless acoustic sensor networks[A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Brisbane: IEEE Signal Processing Society, 2015. 231 – 235.
- [20] Erup L, Gardner F M, Harris R. A. Interpolation in digital modems. II. Implementation and performance [J]. IEEE Transactions on Communication, 1993, 41 (6) : 998 – 1008.
- [21] Robledo-Arnuncio E, Wada T S, Juang B. On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation [A]. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics [C]. New Paltz, USA: IEEE Press, 2007. 34 – 37.
- [22] Schmalenstroer J, Haeb-Umbach R. Efficient sampling rate offset compensation—An overlap-save based approach[A]. European Signal Processing Conference [C]. Rome, Italy: IEEE Signal Processing Society, 2018. 504 – 508.
- [23] Madhu N, Martin R. Low-complexity, robust algorithm for sensor anomaly detection and self-calibration of microphone arrays[J]. IET Signal Processing, 2011, 5(1): 97 – 103.
- [24] Gaubitch N D, Kleijn W B, Heusdens R. Calibration of distributed sound acquisition systems using TOA measurements from a moving acoustic source [A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Florence, Italy: IEEE Signal Processing Society, 2014. 7455 – 7459.
- [25] Tashev I. Gain self-calibration procedure for microphone arrays[A]. IEEE International Conference on Multimedia and Expo [C]. Taipei, China: IEEE Press, 2004. 983 – 986.
- [26] Hua T P, Sugiyama A, Faucon G. A new self-calibration technique for adaptive microphone arrays [J]. IEICE Technical Report Signal Processing, 2005, 105(1): 237 – 240.
- [27] Xiao H, Shao H, Peng Q. A new calibration method for microphone array with gain, phase, and position errors [J]. Journal of Electronic science and Technology of China, 2007, 5(3): 248 – 251.
- [28] Hu D, Chen Z, Yin F. Frequency response calibration using multi-channel Wiener filters for microphone arrays [J]. IEEE Sensors Journal, 2019, 19(17): 7507 – 7514.
- [29] Wang R, Chen Z, Yin F. Adaptive frequency response calibration method for microphone arrays [J]. IEEE Sensors Journal, 2020, 20(13): 7118 – 7128.

- [30] Oak P, Kellermann W. A calibration algorithm for robust generalized sidelobe cancelling beamformers [A]. 2005 IEEE International Workshop on Acoustic Signal Enhancement [C]. Eindhoven, Netherlands: IEEE Signal Processing Society, 2005. 1 – 4.
- [31] Bertrand A, Doclo S, Gannot S, Ono N, Waterschoot T. Special issue on wireless acoustic sensor networks and ad hoc microphone arrays [J]. *Signal Processing*, 2015, 107(2): 1 – 3.
- [32] Ono N, Kohno H, Ito N, Sagayama S. Blind alignment of asynchronously recorded signals for distributed microphone array [A]. *IEEE Workshop Application Signal Processing Audio, Acoustics* [C]. New Paltz, USA: IEEE Press, 2009. 161 – 164.
- [33] Hennecke M H, Fink G A. Towards acoustic self-localization of ad hoc smartphone arrays [A]. *Joint Workshop on Hands-free Speech Communication and Microphone Arrays* [C]. Edinburgh, UK: IEEE Press, 2011. 127 – 132.
- [34] Parviainen M, Pertilä P, Hämäläinen M S. Self-localization of wireless acoustic sensors in meeting rooms [A]. *Joint Workshop on Hands-free Speech Communication and Microphone Arrays* [C]. Villers-les-Nancy, France: IEEE Press, 2014. 152 – 156.
- [35] Sachar J M, Silverman H F, Patterson W R. Microphone position and gain calibration for a large-aperture microphone array [J]. *IEEE Transactions on Speech and Audio Processing*, 2005, 13(1): 42 – 52.
- [36] Jacob F, Schmalenstroer J, Haeb-Umbach R. Microphone array position self-calibration from reverberant speech input [A]. *International Workshop on Acoustic Signal Enhancement* [C]. Aachen, Germany: IEEE Signal Processing Society, 2012. 1 – 4.
- [37] Jacob F, Schmalenstroer J, Haeb-Umbach R. DOA-based microphone array position self-calibration using circular statistics [A]. *IEEE International Conference on Acoustics, Speech and Signal Processing* [C]. Vancouver, Canada: IEEE Signal Processing Society, 2013. 116 – 120.
- [38] Kemper J, Walter M, Linde H. Human-assisted calibration of an angulation based indoor location system [A]. *International Conference on Sensor Technologies and Applications* [C]. Cap Esterel, France: IEEE Computer Society, 2008. 196 – 201.
- [39] Schmalenstroer J, Jacob F, Haeb-Umbach R, Hennecke M, Fink G A. Unsupervised geometry calibration of acoustic sensor networks using source correspondences [A]. *Annual Conference of the International Speech Communication Association* [C]. Florence, Italy: ISCA Press, 2011. 1 – 4.
- [40] Plinge A, Fink G A. Geometry calibration of distributed microphone arrays exploiting audio-visual correspondences [A]. *European Signal Processing Conference* [C]. Lisbon, Portugal: IEEE Signal Processing Society, 2014. 116 – 120.
- [41] Plinge A, Fink G A. Geometry calibration of multiple microphone arrays in highly reverberant environments [A]. *International Workshop on Acoustic Signal Enhancement* [C]. Juan les Pins: IEEE Press, 2014. 243 – 247.
- [42] Plinge A, Fink G A, Gannot S. Passive online geometry calibration of acoustic sensor networks [J]. *IEEE Signal Processing Letters*, 2017, 24(3): 324 – 328.
- [43] Wang R, Chen Z, Yin F. DOA-based three-dimensional node geometry calibration and Cramér-Rao bound analysis in acoustic sensor networks [J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019, 27(9): 1455 – 1468.
- [44] Valente S D, Tagliasacchi M, Antonacci F, Bestagini P, Sarti A, Tubaro S, Milano P. Geometric calibration of distributed microphone arrays from acoustic source correspondences [A]. *IEEE Workshop Multimedia Signal Processing* [C]. Saint Malo, France: IEEE Signal Processing Society, 2010. 13 – 18.
- [45] Jacob F, Haeb-Umbach R. Coordinate mapping between an acoustic and visual sensor network in the shape domain for a joint self-calibrating speaker tracking [A]. *ITG Symposium on Speech Communication* [C]. Erlangen, Germany: VDE, 2014. 1 – 4.
- [46] Jacob F, Haeb-Umbach R. Absolute geometry calibration of distributed microphone arrays in an audio-visual sensor network [J]. *Computer Science*, 2015, 4(3): 128 – 132.
- [47] Chen Z, Li Z, Wang S, Yin F. A microphone position calibration method based on combination of acoustic energy decay model and TDOA for distributed microphone array [J]. *Applied Acoustics*, 2015, 95(8): 13 – 19.
- [48] Crocco M, Bue A D, Bustreo M, Murino V. A closed form solution to the microphone position self-calibration problem [A]. *IEEE International Conference on Acoustics, Speech and Signal Processing* [C]. Kyoto, Japan: IEEE Signal Processing Society, 2012. 2597 – 2600.
- [49] Crocco M, Bue A D, Murino V. A bilinear approach to the position self-calibration of multiple sensors [J]. *IEEE*

- Transactions on Signal Processing, 2012, 60(2): 660 – 673.
- [50] Gaubitch N D, Kleijn W B, Heusdens R. Auto-localization in ad-hoc microphone arrays[A]. IEEE International Conference Acoustics, Speech, and Signal Processing [C]. Vancouver, Canada: IEEE Signal Processing Society, 2013. 106 – 110.
- [51] Wang L, Hon T, Reiss J D, Cavallaro A. Self-localization of ad-hoc arrays using time difference of arrivals[J]. IEEE Transactions on Signal Processing, 2016, 64(4): 1018 – 1033.
- [52] Le T, Ho K C. Algebraic complete solution for joint source and sensor localization using time of flight measurements[J]. IEEE Transactions on Signal Processing, 2020, 68: 1853 – 1869.
- [53] Hennecke M, Plötz T, Fink G A, Schmalenströer J, Häb-Umbach R. A hierarchical approach to unsupervised shape calibration of microphone array networks [A]. IEEE/SP Workshop on Statistical Signal Processing [C]. Cardiff, UK: IEEE Signal Processing Society, 2009. 257 – 260.
- [54] Schwartz O, Plinge A, Habets E A P, Gannot S. Blind microphone geometry calibration using one reverberant speech event [A]. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics [C]. New Paltz, New York: IEEE Signal Processing Society, 2017. 131 – 135.
- [55] Taghizadeh M J, Parhizkar R, Garner P N, Bourlard H, Asaei A. Ad hoc microphone array calibration: Euclidean distance matrix completion algorithm and theoretical guarantees [J]. Signal Processing, 2014, 107(2): 123 – 140.
- [56] Asaei A, Mohammadiha N, Taghizadeh M J, Doclo S, Bourlard H. On application of non-negative matrix factorization for ad hoc microphone array calibration from incomplete noisy distances[A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. South Brisbane: IEEE Signal Processing Society, 2015. 2694 – 2698.
- [57] Bouley S, Vanwynsberghe C, Magueresse T L, Antoni J, Outrequin A. Microphone array positioning technique with Euclidean distance geometry [J]. Applied Acoustics, 2020, 167: 1 – 11.
- [58] Gburrek T, Schmalenstroeer J, Brendel A, Kellermann W, Haeb-Umbach R. Deep neural network based distance estimation for geometry calibration in acoustic sensor networks[A]. European Signal Processing Conference [C]. Amsterdam, the Netherlands: IEEE Signal Processing Society, 2020. 1 – 5.
- [59] Cox T F, Cox M A. Multidimensional Scaling[M]. London: Chapman and Hall Press, 2001.
- [60] Coleman T, Li Y. An interior trust region approach for nonlinear minimization subject to bounds[J]. Mathematica, 1996, 6(2): 418 – 445.
- [61] Pertilä P, Hämäläinen M S, Mieskolainen M. Passive temporal offset estimation of multichannel recordings of an ad hoc microphone array [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(11): 2393 – 2402.
- [62] Pertilä P, Mieskolainen M, Hämäläinen M S. Passive self-localization of microphones using ambient sounds [A]. European Signal Processing Conference [C]. Bucharest, Romania: IEEE Signal Processing Society, 2012. 1314 – 1318.
- [63] Press W H, Flannery B P, Teukolsky S A, Vetterling W T. Numerical Recipes in C [M]. Cambridge, UK: Cambridge University Press, 1988.
- [64] Byrd R H, Lu P, Nocedal J. A limited memory algorithm for bound constrained optimization[J]. SIAM Journal on Scientific and Statistical Computing, 1995, 16: 1190 – 1208.
- [65] Plinge A, Fink G A. Online multi-speaker tracking using multiple microphone arrays informed by auditory scene analysis [A]. European Signal Processing Conference [C]. Marrakesh, Morocco: IEEE Signal Processing Society, 2013. 1 – 5.
- [66] Storn R, Price K. Differential evolution-A simple and efficient heuristic for global optimization over continuous spaces [J]. Journal of Global Optimization, 1997, 11: 341 – 359.
- [67] Keshavan R H, Montanari A, Oh S. Matrix completion from noisy entries [J]. Journal of Machine Learning Research, 2010, 11: 2057 – 2078.
- [68] Lee D D, Seung H S. Algorithms for nonnegative matrix factorization [J]. Advances in Neural Information Processing Systems, 2001, 13: 556 – 562.
- [69] Schönemann P H. On metric multidimensional unfolding [J]. Psychometrika, 1970, 35(3): 349 – 66.
- [70] Parhizkar R. Euclidean Distance Matrices: Properties, Algorithms and Applications[D]. Swiss: École Polytechnique Fédérale de Lausanne, 2013.
- [71] Kabsch W. A solution for the best rotation to relate two

sets of vectors [J]. Acta Crystallographica Section A, 1976, 32(5): 922 – 923.

- [72] Kabsch W. A discussion of the solution for the best rotation to relate two sets of vectors[J]. Acta Crystallographica Section A, 1978, 34(5): 827 – 828.
- [73] Khokhlov Y, Zatvornitskiy A, Medennikov I, Sorokin

I, Prisyach T, Romanenko A, Mitrofanov A, Bataev V, Andrusenko A, Korenevskaya M, Petrov O. R-vectors: New technique for adaptation to room acoustics[A]. Annual Conference of the International Speech Communication Association [C]. Graz, Austria: ISCA Press, 2019. 1243 – 1247.

参考文献



王睿女, 1989年生于辽宁省阜新市. 现为大连理工大学信息与通信工程学院博士研究生. 主要研究方向为阵列信号处理.
E-mail: wangr1989@mail.dlut.edu.cn



陈喆男, 1975年生于黑龙江省泰来县. 现为大连理工大学信息与通信工程学院教授. 主要研究方向为音频信号处理、图像处理和宽带无线通信.
E-mail: zhechen@dlut.edu.cn



殷福亮男, 1962年11月生于辽宁省抚顺市. 现为大连理工大学信息与通信工程学院教授. 研究方向为音频信号处理、图像处理和宽带无线通信.
E-mail: flyin@dlut.edu.cn