



高新科技译丛



Springer

Microphone Array Signal Processing

麦克风阵列信号处理

【加】 Jacob Benesty Jingdong Chen Yiteng Huang 著
邹霞 周彬 贾冲 张雄伟 徐晓铁 译



国防工业出版社
National Defense Industry Press



高新科技译丛

装备科技译著出版基金

麦克风阵列信号处理

Microphone Array Signal Processing

Jacob Benesty Jingdong Chen

Yiteng Huang 著

邹霞 周彬

张雄伟 徐晓铁



国防工业出版社

· 北京 ·

著作权合同登记 图字：军-2013-056 号

图书在版编目 (CIP) 数据

麦克风阵列信号处理/ (加) 贝内斯特 (Benesty, J.), 陈景东, 黄以腾著; 邹霞等译. —北京: 国防工业出版社, 2016.1

(高新科技译丛)

书名原文: Microphone Array Signal Processing

ISBN 978-7-118-10498-1

I. ①麦… II. ①贝… ②陈… ③黄… ④邹… III. ①传声器—信号处理
IV. ①TN641

中国版本图书馆 CIP 数据核字 (2015) 第 313485 号

Translation from English language edition:

Microphone Array Signal Processing

by Jacob Benesty, Jingdong Chen and Yiteng Huang

Copyright©2008 Springer Berlin Heidelberg

Springer Berlin Heidelberg is a part of Springer Science+ Business Media

All Rights Reserved

本书简体中文版由 Springer Berlin Heidelberg 授权国防工业出版社独家出版发行。
版权所有, 侵权必究。

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

国防工业出版社印刷厂印刷

新华书店经售

*

开本 710×1000 1/16 印张 13 字数 240 千字

2016 年 1 月第 1 版第 1 次印刷 印数 1—2000 册 定价 69.90 元

(本书如有印装错误, 我社负责调换)

国防书店: (010) 88540777

发行邮购: (010) 88540776

发行传真: (010) 88540755

发行业务: (010) 88540717

译者前言

近年来, 麦克风阵列技术在理论研究方面掀起了一个新的高潮, 各种新的理论和新的方法不断被提出。而且, 麦克风阵列技术已经广泛应用于电话会议、免提话音通信、人机语音交互等系统。但是, 在这些重要领域里, 竟然几乎找不到介绍其最有用的思想、概念、结论以及流行算法的成体系的书籍, 这给想从事这方面研究的研究生及工程师造成了极大的不便。

该书与传统的阵列信号处理书籍不同。传统的阵列信号处理书籍是从窄带信号处理的角度来描述, 如假定信号是窄带的、采用的信号处理算法是窄带的等。这些方法对于解决诸如天线阵列等问题来说是有效的, 但对于解决明显具有宽带信号特征的声音及语音信号的问题时则会遇到很多不可克服的困难。而本书真正从宽带信号处理角度来分析、解决问题。从声音信号的独特特性出发, 完整的阐述了麦克风阵列信号处理的基本理论、基础方法和关键技术, 使读者可以从根本上摒除传统窄带阵列信号处理的观念, 真正意义上进入麦克风阵列的宽带处理时代。

该书的作者是麦克风阵列信号处理领域的顶级专家学者。该书是作者近年来在麦克风阵列信号处理方向研究成果的一个总结。针对麦克风阵列信号的应用背景, 侧重基本概念、基本理论和基本方法的阐述, 使读者深入理解新理论和新方法提出的背景和思路, 为读者进一步开展这一方面的研究具有重要的意义。

在这里要感谢参加本书校对和译稿资料整理的孙蒙、闵刚、胡永刚、李轶南、乔林、陈栩杉、郑云飞、韩伟、李飞彬等研究生。本书出版受到了国家自然科学基金(61402519、61471394)和江苏省自然科学基金(BK20140074、BK20140071)项目的支助。

由于译者水平有限, 文中难免有不当之处, 敬请读者批评指正。

译者

2015年10月

前 言

在过去几年间,我们已经著述和编写了多本关于声及语音信号处理的书籍。撰稿伊始,我们发现在这些重要领域里,竟然几乎找不到介绍其最有用的思想、概念、结论以及流行算法的书籍,即使是到我们的原稿完成,准备付梓时,情况也还是大致如此。但是,大家对这些书籍却有很迫切的需求。这就是我们做这些工作的原因。值得欣慰的是,到目前为止,从读者反馈的角度来看,我们可以说做了一件很正确的事情。最近,很多学者也加入了我们的行列,围绕这些重要领域的相关问题,从他们各自的视角和观点出发,出版了许多很有意义的著作。

撰写这本《麦克风阵列信号处理》的想法来源于我们和同事以及朋友之间的讨论。经过大家的讨论,我们最终再一次印证了这样的一个结论:大家都非常热切地希望有一本详细阐述麦克风阵列理论与应用的专著出版。尽管目前市场上有很多从窄带信号处理角度(即信号为窄带信号,处理方法为窄带信号处理方法)描述的天线阵列的书籍,但是真正从宽带信号处理角度出发考虑传感器阵列问题的专著却极为匮乏。目前,许多语音信号处理应用中采用的方法,多是从窄带天线阵列信号处理算法中直接借鉴而来。这种将窄带信号处理的思想直接应用于宽带语音信号处理的朴素想法虽然简单,但可能未必合理,甚至会导致很多误解。因此,本书摒弃了这样的传统思路,从严格的宽带信号处理角度(即信号为宽带信号,处理方法为宽带信号处理方法)出发,推导和解释了麦克风阵列信号处理中最重要的基本算法。这也正是本书的主要目的。可以预见,这些全新的概念具有巨大的应用潜力,将对解决声及语音信号处理中遇到的难题带来新的曙光。

本书尤其适用于从事麦克风阵列研究的在校研究生、学者和工程师。本书对麦克风阵列处理的理论与应用方法的阐述自成体系,目的就是使读者通过阅读本书就能获得该领域完整的理论框架。我们试图以一种简洁而严格的方式来阐述麦克风阵列处理算法中最主要的思想,以便读者可以对这些算法的潜力、

机遇、挑战和局限性有很好的理解。

最后，感谢 Springer (Germany)出版社的 Christoph Baumann, Petra Jantzen, Carmen Wolf, 感谢他们对本书的准备与出版给予的重大帮助，与他们一起工作的过程是一段愉悦和美好的经历。

Jacob Benesty
Jingdong Chen
Yiteng Huang

目 录

第 1 章 绪论	1
1.1 基于麦克风阵列的信号处理技术研究	1
1.2 本书的内容与结构	5
第 2 章 经典最优滤波器	7
2.1 引言	7
2.2 维纳滤波器	7
2.3 Frost 滤波器	15
2.3.1 算法	15
2.3.2 广义旁瓣抵消结构	16
2.3.3 在线性内插中的应用	17
2.4 卡尔曼滤波	19
2.5 可替代 MSE 的准则	22
2.5.1 Pearson 相关系数	22
2.5.2 SPCC 的重要性质	23
2.5.3 由 SPCC 推导最优滤波器举例	25
2.6 总结	31
第 3 章 传统波束形成技术	32
3.1 引言	32
3.2 问题描述	33
3.3 延迟-求和技术	33
3.4 固定波束形成的设计	38
3.5 最大信噪比滤波器	41
3.6 最小方差无失真响应滤波器	43
3.7 基于参考信号的方法	44
3.8 响应不变宽带波束形成	45
3.9 陷零技术	48
3.10 麦克风阵列方向图函数	50

3.10.1	第一种信号模型	50
3.10.2	第二种信号模型	52
3.11	总结	53
第 4 章	LCMV 滤波器在室内声学环境下的应用	54
4.1	引言	54
4.2	信号模型	54
4.2.1	无回声模型	54
4.2.2	混响模型	55
4.2.3	空-时模型	55
4.3	基于无回声模型的 LCMV 滤波器	56
4.4	基于混响模型的 LCMV 滤波器	58
4.5	基于空-时模型的 LCMV 滤波	61
4.5.1	实验结果	63
4.6	频域 LCMV 滤波	65
4.7	总结	67
第 5 章	多麦克风噪声抑制: 统一处理	68
5.1	引言	68
5.2	信号模型和问题描述	68
5.3	有用的定义	70
5.4	维纳滤波	71
5.5	子空间方法	74
5.6	空-时预测方法	76
5.7	噪声完全相干的情况	78
5.8	自适应噪声消除	79
5.9	卡尔曼滤波	80
5.10	仿真	81
5.10.1	声学环境和实验设置	81
5.10.2	实验结果	83
5.11	总结	91
第 6 章	非因果(频域)最优滤波	92
6.1	引言	92
6.2	信号模型和问题描述	92
6.3	性能测度	93

6.4	非因果维纳滤波	96
6.5	参数化维纳滤波	99
6.6	推广到多通道的情况	101
6.6.1	信号模型	101
6.6.2	定义	102
6.6.3	多通道维纳滤波	104
6.6.4	空间最大 SNR 滤波	106
6.6.5	最小方差无失真响应滤波	108
6.6.6	无失真多通道维纳滤波	109
6.7	总结	110
第 7 章	多输入多输出视角下的麦克风阵列	111
7.1	引言	111
7.2	信号模型与问题描述	111
7.2.1	SISO 模型	112
7.2.2	SIMO 模型	113
7.2.3	MISO 模型	113
7.2.4	MIMO 模型	114
7.2.5	问题描述	115
7.3	二元麦克风阵列	116
7.3.1	最小二乘法	116
7.3.2	Frost 算法	117
7.3.3	广义旁瓣消除结构	118
7.4	N 元麦克风阵列	120
7.4.1	最小二乘和 MINT 方法	120
7.4.2	Frost 算法	122
7.4.3	广义旁瓣消除结构	123
7.4.4	最小方差无失真响应方法	125
7.5	仿真	125
7.5.1	声环境与实验设置	125
7.6	总结	130
第 8 章	逐级分离和去混响：两级方法	132
8.1	引言	132

8.2	信号模型与问题描述	132
8.3	源分离	134
8.3.1	2×3 MIMO 系统	134
8.3.2	$M \times N$ MIMO 系统	138
8.4	语音去混响	140
8.4.1	直接反向法	142
8.4.2	最小均方误差和最小二乘法	142
8.4.3	MINT 方法	143
8.5	总结	145
第 9 章	到达方向和到达时间差估计	146
9.1	引言	146
9.2	问题描述和信号模型	148
9.2.1	单源自由场模型	148
9.2.2	多源自由场模型	149
9.2.3	单源混响模型	150
9.2.4	多源混响模型	151
9.3	互相关方法	152
9.4	广义互相关方法族	153
9.4.1	经典互相关	154
9.4.2	平滑相干变换	154
9.4.3	相位变换	155
9.5	空间线性预测方法	155
9.6	多通道互相关系数算法	158
9.7	基于特征矢量的技术	162
9.7.1	窄带 MUSIC	162
9.7.2	宽带 MUSIC	163
9.8	最小熵方法	165
9.8.1	高斯源信号	165
9.8.2	语音源信号	166
9.9	自适应特征值分解算法	167
9.10	基于自适应多通道盲辨识的方法	168
9.11	多源 TDOA 估计	170
9.12	总结	173

第 10 章 未涉及的问题 174

10.1 引言 174

10.2 语音源数目估计 174

10.3 鸡尾酒会效应和盲源分离 175

10.4 盲 MIMO 辨识 176

10.5 总结 178

参考文献 179

第1章 绪 论

1.1 基于麦克风阵列的信号处理技术研究

麦克风阵列是以特定方式排列,从而能够准确获取空间信息的一组麦克风。类比于无线通信的空间分集,这里讨论麦克风阵列的空间多样性。这种多样性通常由辐射源到传感器之间的声脉冲响应来表示,并且可以通过不同的方式加以理解和应用,本书将对此进行详细阐释。这些采用有限冲激响应(Finite Impulse Response, FIR)滤波器建模的声信道通常是非唯一的,尤其是当 FIR 滤波器具有共同的零点时问题最大。

对于空间多样性所提供的丰富信息需要做进一步处理。因此,麦克风阵列信号处理的主要目的就是:根据应用的不同,利用麦克风阵列输出信号中包含的空域-时域(也可能是频域)信息,估计某些参数或提取感兴趣的信号。尽管本书中也涉及单麦克风系统的特殊情况(第2章中的最优滤波器,第6章中的频域噪声抑制),但本书的重点主要集中在多传感器系统的应用方面,其原因在于多传感系统更加灵活,因而能够更好地解决许多重要的实际问题。

根据麦克风应用场合不同,麦克风阵列可以采用不同的几何排列,麦克风阵列的几何排列形状对处理算法的构建具有重要影响。例如,在声源定位中,阵列的几何排列形状必须是已知的,由此才能正确地对声源进行定位。在某些情况下,适当的规则排列可以简化估计问题,这就是均匀线性阵列和圆形阵列得到广泛应用的原因^[148]。尽管目前这两种排列形状的阵列占据了大部分的市场,但同时也出现了一些能够更好地获取声场信息的高复杂度三维球状阵列^[163, 164]。然而,在另外一些关键问题,例如噪声抑制或者源分离应用中,阵列几何形状对算法的影响较小。在这种情况下,可以将系统称为多麦克风系统,而不称为麦克风阵列。二者的区别在具体场合下是显而易见的,因此不需要对这两种情况严格区分。

麦克风阵列中遇到的许多问题看似容易解决,因为类似的问题在窄带天线阵列中已经得到了相当长时间的研究。事实却并非如此。实际上,麦克风阵列与天线阵列(如雷达和声纳)在实际应用中的工作方式是不同的,其原因如下^[105, 215]:

- (1) 语音信号是宽带信号。
- (2) 室内混响(或者多径)效应显著。

- (3) 环境特性和信号均高度非平稳。
- (4) 噪声与目标语音信号可能具有相同的频谱特性。
- (5) 传感器的数目通常有限。

(6) 人耳听力具有极宽的动态范围(听力正常情况下高达 120dB),并且对通道冲激响应的弱拖尾极其敏感。因此,滤波器模型的长度很长(抽头数目数以千计的情况并不少见)。

正是由于这些问题的存在,现有的许多算法在一些实际问题的解决中性能并不好。

许多麦克风阵列处理算法都是借用窄带阵列处理算法或者对其进行简单推广后得到的^[51]。这种策略的优点在于,能够简便地对数十年来形成的天线阵列算法进行扩展。但是,缺点在于,这些算法并不是针对实际的声学环境而设计,其性能往往受到很大限制。简单来说,麦克风阵列需要的是宽带处理。总的来说,这也正是本书所采用的研究方法。

通过麦克风阵列能够解决的实际问题主要包括:

- (1) 噪声抑制。
- (2) 回声抑制。
- (3) 去混响。
- (4) 单声源定位。
- (5) 声源数目估计。
- (6) 多声源定位。
- (7) 源分离。
- (8) 鸡尾酒会效应。

图 1.1 对上述的大多数问题进行了描述,其中,所有麦克风拾取得到的信号被送给一组给定的滤波器,而这些滤波器需要根据上述实际问题进行优化设计^[20]。

噪声抑制算法的目的就是要从受到加性噪声影响的观测信号中估计出目标语音信号。对于这一问题,已经存在许多基于单麦克风的技术^[16,154,156]。然而,所有单麦克风噪声抑制算法的主要问题在于它会使语音信号失真^[41,42]。尽管语音质量得到改善,但可懂度却降低了。然而,采用麦克风阵列,却可以在不太影响语音信号的条件抑制噪声(至少在理论上可以实现)。

在免提通信中,由于扬声器与麦克风之间的声场耦合以及系统延时的存在,导致回声效应的产生,使得实时交谈变得非常困难^[10,29,84,98,99,121]。并且,声学系统可能变得极不稳定。利用麦克风阵列技术,使阵列指向目标声源,同时在扬声器方向陷零,将能够显著降低回声强度。尽管这种思路简单直接且极具吸引力,但是在实际中并没有得到应用。对于这一问题,当前最好的解决方案仍然是采用声学回声抵消^[10]。

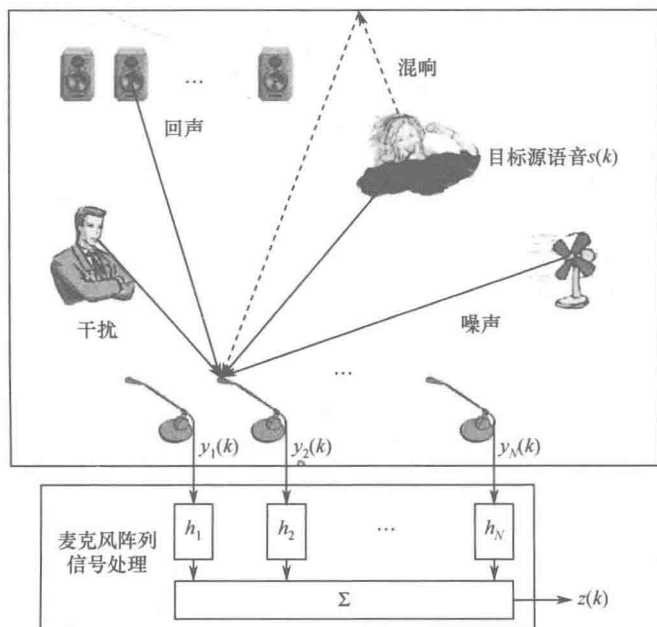


图 1.1 麦克风阵列信号处理

在房间内的免提通信场景下，麦克风采集到的语音信号不仅包含直接路径传播的信号，而且包含了由于房间边界以及房间内物体的反射而产生的延迟衰减信号。这种多径传播效应在观测信号中引入了回声和谱失真，称为混响，这可能会严重影响源信号，造成信号质量和可懂度下降。因此，去除混响对提高语音信号的可懂度有很大的作用^[125]。四十多年来，为了利用麦克风阵列，研究者做了大量的工作寻求切实可行的解决方案。

在声学环境下，声源位置信息对于许多实际应用发挥着重要作用，例如视频会议中的自动摄像机跟踪、噪声抑制和去混响中波束方向的调整。信号源位置的估计，通常也称为声源定位问题，数十年来受到了极大的关注^[26,117,175,222]。要实现对声源的入射角度或者其在直角坐标系中位置的估计，需要采用二维或三维的麦克风阵列。对声源数目的估计和对多个声源的定位是两个相互联系的问题，对于窄带信号源已有一些相关的算法，但是，这些问题在宽带信号源中的研究才刚刚起步。

在多麦克风声源分离中，需要同时对来自不同方向的信号进行分离。由于不能直接获取声通道和声源信号的信息，因此所有这些方法本质上都属于盲处理。独立成分分析（Independent Component Analysis, ICA）是盲源分离（Blind Source Separation, BSS）问题中使用最为广泛的工具^[127]，它充分利用了源信号的独立性。在信号瞬时混合的情况下，多数基于 ICA 的算法都能够获得良好的性能，然而，在混响（卷积）环境下，此类算法的性能并不是很好。对于这一问题，近年来的研究取得了很大进展，但如何将其应用于语音和声学环境，

以及可实用化的程度并不明确。有关 ICA 方面的文献已经相当丰富^[159] (参见文献[182]), 因此本书不再从这一角度讨论 BSS。

众所周知, 人类具有在嘈杂的谈话环境或背景噪声中专注于某一特定的话音或声音的能力。这一有趣的心理声学现象称为鸡尾酒会效应^[45,46]。从文献[45,46]中的心理声学实验中可以观察到一个重要的现象, 即空间听觉发挥着重要作用。人类对语音的感知很大程度上得益于空间听觉能力, 而这种能力又主要归因于双耳的作用。这一点可以通过日常经验直观地验证, 同时还可以观察单耳和双耳在多个说话人同时存在的密闭空间环境中对声音的理解力差异来进一步说明^[125]。具有正常听力和大脑处理能力的人都可以毫不费力地有效解决这种鸡尾酒会问题, 但对于麦克风阵列信号处理仍然是一个极其复杂的问题。这一问题成为该领域所面临的诸多挑战的根源所在, 直到今天仍然没有一个明确的解决思路。

无论麦克风阵列的尺寸、几何形状或阵元数目如何, 上述问题都是极难解决的。某些情况下, 采用特定几何形状的阵列或增加麦克风的数目可以更准确地解决问题。然而, 这种改善可能非常有限, 甚至可以忽略不计。因此, 出现了一些根本性问题: 如何利用空间信息? 对于一个具体问题, 能够解决到何种程度? 合理的模型是怎样的? 可以达到的极限是什么及其理由? 除了空间信息还可以利用其他信息吗? 如果可以, 如何利用?

本书的目的并不是为了针对上述问题提出不同解决方案, 也不是为了描述最新的研究成果, 而是为了提出总体框架、介绍重要工具以及信号模型, 以帮助读者更加直观而精确地理解多麦克风信号的处理。

最后, 这里简要列出一些麦克风阵列的典型应用:

- (1) 电话会议;
- (2) 多方通信;
- (3) 免提人机声音交互;
- (4) 对话系统;
- (5) 计算机游戏;
- (6) 指挥和控制接口;
- (7) 听写系统;
- (8) 高品质录音;
- (9) 声学监视;
- (10) 听觉场景分析;
- (11) 传感器网络技术。

可以看出, 麦克风阵列的应用范围很广并且发展迅猛。显然, 优秀的麦克风阵列技术及其解决方案符合市场需求, 必将得到广泛应用。

1.2 本书的内容与结构

本书共 10 章（包括本章）。后面的 9 章尝试从一个崭新的视角对麦克风阵列信号处理所涉及的重要问题进行研究与探索。每章都从基本概念逐步深入，使读者可以顺着从基本理论到实际应用的思路进行学习研究。各章之间虽然存在连贯性，但又是自包含的，可以独立地进行学习。

线性最优滤波器在麦克风阵列信号处理等许多应用中都具有根本性的作用。最优滤波的概念本身容易理解，对本书的其他章节也非常重要。第 2 章主要研究维纳滤波器、Frost 滤波器和卡尔曼滤波器。本章还提出了 Pearson 相关系数的概念，并将其作为均方误差的替代准则。这一研究得出了许多有意义的成果。

传统的应用于窄带天线阵列空间滤波的波束形成方法已经相当成熟。第 3 章以信号增强为应用背景，通过采用一个简单信号传播模型，讨论了这一方面的相关典型技术。第 3 章还介绍了宽带波束形成的理念，这一理念对于语音信号具有重要意义。

线性约束最小方差（Linear Constrained Minimum Variance, LCMV）滤波器在天线阵列中的应用非常广泛。这种最优滤波器非常强大，其原因在于它可以将所有的约束与其代价函数联系起来。第 4 章通过使用三种不同的信号模型，介绍了在室内声学环境下如何使用 LCMV 滤波器进行噪声抑制和混响去除。

第 5 章专门介绍了多麦克风噪声抑制问题，在一个统一的框架内推导了几种经典的多通道处理方法。此外，本章还对语音增强涉及的重要内容进行了讨论，包括噪声抑制能力和语音失真大小。

第 6 章介绍了非因果（频域）维纳滤波及其在降噪方面的应用，分为单通道和多通道两种情况。为帮助读者更好地理解频域算法的工作原理，本章还介绍了有关语音增强方面的许多基本内容，尤其是针对多个麦克风的情况。

第 7 章以目标信号和干扰信号作为输入，麦克风信号作为输出，从而将麦克风阵列视为一个多输入多输出（Multiple Input Multiple Output, MIMO）系统，并基于 MIMO 通道冲激响应建立一个总体框架，用于分析波束形成在信号提取、混响去除及干扰抑制方面的性能。

作为第 7 章的延续，第 8 章指明了如何采用分两步处理的策略，将干扰和混响这两个问题分离开，分别采用不同的方法来解决。并且，清楚地说明了实现分离的前提条件。这种分离的实现能够有助于更好地理解源分离和去混响之间的相互关系。

第9章研究了波达方向(Direction of Arrival, DOA)估计和波达时间差(Time Difference of Arrival, TDOA)估计这一重要问题。由于DOA估计问题在本质上与TDOA估计问题是相同的,因此本章更多地将研究重点放在TDOA估计算法上面,并介绍了许多具体算法:既介绍了经典算法,如互相关方法等,也介绍了许多新颖算法,如最小熵技术等。本章还讨论了多源情况下TDOA估计的原理。

最后,第10章对本书中未涉及的问题进行了讨论。