

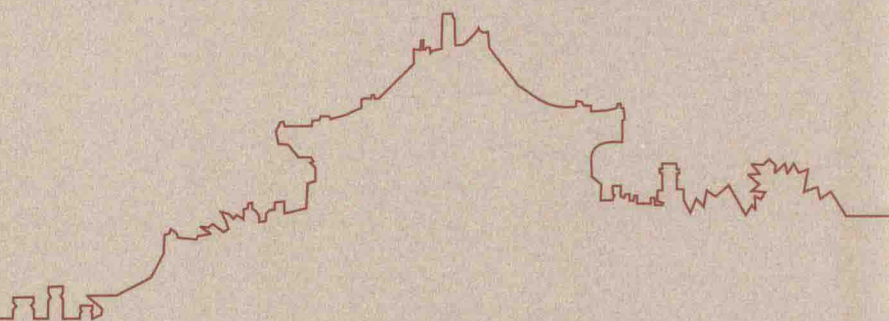
武汉大学优秀博士学位论文文库



基于双耳线索的 移动音频编码研究

Research on Binaural Cues of Spatial Audio Coding in Mobile Application

陈水仙 著



WUHAN UNIVERSITY PRESS
武汉大学出版社

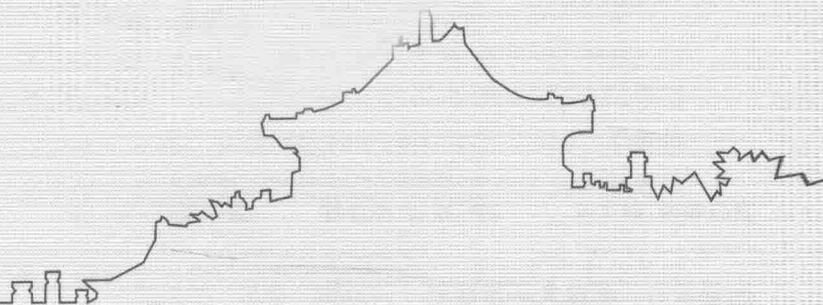
武汉大学优秀博士学位论文文库



基于双耳线索的 移动音频编码研究

Research on Binaural Cues of Spatial Audio Coding in Mobile Application

陈水仙 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

基于双耳线索的移动音频编码研究/陈水仙著. —武汉: 武汉大学出版社, 2014. 1
武汉大学优秀博士学位论文文库
ISBN 978-7-307-12318-2

I. 基… II. 陈… III. 移动通信—音频编码—研究 IV. ①TN929.5
②TN912.3

中国版本图书馆 CIP 数据核字(2013)第 312844 号

责任编辑: 谢文涛 责任校对: 鄢春梅 版式设计: 马 佳

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)
(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 湖北恒泰印务有限公司

开本: 720 × 1000 1/16 印张: 11.5 字数: 161 千字 插页: 2

版次: 2014 年 1 月第 1 版 2014 年 1 月第 1 次印刷

ISBN 978-7-307-12318-2 定价: 26.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

武汉大学
优秀博士学位论文文库
编委会

主 任 李晓红

副主任 韩 进 舒红兵 李 斐

委 员 (按姓氏笔画为序)

马费成	邓大松	边 专	刘正猷	刘耀林
杜青钢	李义天	李建成	何光存	陈 化
陈传夫	陈柏超	冻国栋	易 帆	罗以澄
周 翔	周叶中	周创兵	顾海良	徐礼华
郭齐勇	郭德银	黄从新	龚健雅	谢丹阳

总 序

创新是一个民族进步的灵魂，也是中国未来发展的核心驱动力。研究生教育作为教育的最高层次，在培养创新人才中具有决定意义，是国家核心竞争力的重要支撑，是提升国家软实力的重要依托，也是国家综合国力和科学文化水平的重要标志。

武汉大学是一所崇尚学术、自由探索、追求卓越的大学。美丽的珞珈山水不仅可以诗意栖居，更可以陶冶性情、激发灵感。更为重要的是，这里名师荟萃、英才云集，一批又一批优秀学人在这里砥砺学术、传播真理、探索新知。一流的教育资源，先进的教育制度，为优秀博士学位论文的产生提供了肥沃的土壤和适宜的气候条件。

致力于建设高水平的研究型大学，武汉大学素来重视研究生培养，是我国首批成立有研究生院的大学之一，不仅为国家培育了一大批高层次拔尖创新人才，而且产出了一大批高水平科研成果。近年来，学校明确将“质量是生命线”和“创新是主旋律”作为指导研究生教育工作的基本方针，在稳定研究生教育规模的同时，不断推进和深化研究生教育教学改革，使学校的研究生教育质量和知名度不断提升。

博士研究生教育位于研究生教育的最顶端，博士研究生也是学校科学研究的重要力量。一大批优秀博士研究生，在他们学术创作最激情的时期，来到珞珈山下、东湖之滨。珞珈山的浑厚，奠定了他们学术研究的坚实基础；东湖水的灵动，激发了他们学术创新的无限灵感。在每一篇优秀博士学位论文的背后，都有博士研究生们刻苦钻研的身影，更有他们的导师的辛勤汗水。年轻的学者们，犹如在海边拾贝，面对知识与真理的浩瀚海洋，他们在导师的循循善

诱下，细心找寻着、收集着一片片靓丽的贝壳，最终把它们连成一串串闪闪夺目的项链。阳光下的汗水，是他们砥砺创新的注脚；面向太阳的远方，是他们奔跑的方向；导师们的悉心指点，则是他们最值得依赖的臂膀！

博士学位论文是博士生学习活动和研究工作的主要成果，也是学校研究生教育质量的凝结，具有很强的学术性、创造性、规范性和专业性。博士学位论文是一个学者特别是年轻学者踏进学术之门的标志，很多博士学位论文开辟了学术领域的新思想、新观念、新视阈和新境界。

据统计，近几年我校博士研究生所发表的高质量论文占全校高水平论文的一半以上。至今，武汉大学已经培育出 18 篇“全国百篇优秀博士学位论文”，还有数十篇论文获“全国百篇优秀博士学位论文提名奖”，数百篇论文被评为“湖北省优秀博士学位论文”。优秀博士结出的累累硕果，无疑应该为我们好好珍藏，装入思想的宝库，供后学者慢慢汲取其养分，吸收其精华。编辑出版优秀博士学位论文文库，即是这一工作的具体表现。这项工作既是一种文化积累，又能助推这批青年学者更快地成长，更可以为后来者提供一种可资借鉴的范式亦或努力的方向，以鼓励他们勤于学习，善于思考，勇于创新，争取产生数量更多、创新性更强的博士学位论文。

武汉大学即将迎来双甲华诞，学校编辑出版该文库，不仅仅是为百廿武大增光添彩，更重要的是，当岁月无声地滑过 120 个春秋，当我们正大踏步地迈向前方时，我们有必要回首来时路，我们有必要清晰地审视我们走过的每一个脚印。因为，铭记过去，才能开拓未来。武汉大学深厚的历史底蕴，不仅仅在于珞珈山的一草一木，也不仅仅在于屋檐上那一片片琉璃瓦，更在于珞珈山下的每一位学者和学生。而本文库收录的每一篇优秀博士学位论文，无疑又给珞珈山注入了新鲜的活力。不知不觉地，你看那珞珈山上的树木，仿佛又茂盛了许多！

李晓红

2013 年 10 月于武昌珞珈山

摘 要

飞速发展的移动市场对高品质音频业务的需求日益迫切，移动环境下音频带宽不足，终端资源有限，传统音频理论和技术面临新的挑战。近年来，MPEG、ITU 和 AVS 等国内外标准化组织积极推动移动音频理论和技术研究。基于双耳线索的音频编码因其在低码率下的音质优势而备受关注，然而，为适应移动环境苛刻的应用需求，仍需克服空间参数模型的理论缺陷，突破实变换域双耳线索表示的技术瓶颈，解决空间参数码率难以下降、系统复杂度偏高等问题。本书开展基于双耳线索的移动音频编码研究，在双耳线索感知机理分析、空间感知信息度量、MDCT 域双耳线索表示和空间音频编码系统结构设计这四个方向上取得了创新性成果。

在双耳线索感知机理分析方向，本书针对现有系统只能去除声道间数据客观冗余的不足，以巴克频带为感知分析单元，开展了对双耳线索临界感知 JND 的分析研究，提出基于感知的双耳线索参数表示模型，有效去除了声道间数据的主观冗余信息，下降 4.28% ~ 11.78% 的空间参数码率，主观测试 MUSHRA 得分平均提高 2.6 分，客观测试 ODG 得分平均提高 1.1 分。

在空间感知信息度量方向，本书针对建立在单耳听音模型上的感知熵理论无法度量双耳空间感知信息量的问题，结合双耳线索的感知机理，建立了基于双耳线索生理感知的 BCPPM 模型来模拟双耳对空间信息的感知过程，提出基于模型的空间感知熵 SPE 的计算方法，解决了双耳空间感知信息量的度量问题，丰富和发展了空间音频编码技术的理论和实践。

在 MDCT 域双耳线索表示方向，本书针对 MDCT 实变换的相位信息缺失及能量抖动造成双耳线索表示不准确的问题，采用基函

数“和分离”技术，提出了基于共轭窗函数的 MDCT 频谱复数化扩展方法，并针对新方法引入 MDST 带来复杂度增加的问题，提出了基于转换矩阵稀疏性的 MDCT-MDST 快速转换算法，同时利用转换矩阵的正交性和循环性，将复杂度从 $O(N^2)$ 下降到 $O(N)$ ，主观测试 MUSHRA 得分平均提高 2.07 分，客观 ODG 得分平均提高 0.4281 分。

在空间音频编码系统结构设计方向，本书针对空间信息和下混信息分离处理模式导致的系统复杂度高和延时长的问题，采用滤波复用和并行处理的技术，提出并行滤波共享的分析综合 ABS 编码结构，去除了空间参数编码与下混声道编码中的冗余操作，缩短了 5% 的系统延时，减少了 38% ~ 86% 的运算复杂度。

本书在基础理论和关键技术方面的研究成果可望成为国内移动音频标准的支撑技术，增强我国在高速成长的全球移动音频产业中的核心竞争力，为全面参与移动音频编解码领域的国际竞争和标准化工作奠定坚实的研究基础。

关键词：移动音频 空间音频编码 感知熵 双耳线索 修正离散余弦变换

Abstract

The mobile communication industry, with its unprecedented pace of development, expects ever higher audio quality. Due to the bandwidth and computing power limits, audio under mobile environment meets a number of great challenges. In recent years international and national standardization organizations such as MPEG, ITU and AVS actively promote mobile audio researches. Among them, a novel coding scheme named spatial audio coding attracts enormous research and application interests for its significant quality advantage at low bitrate. This dissertation contributes at 4 aspects: perceptible binaural cues analysis, perceptible spatial information estimation, binaural cues representation in transformed domain, and coding architecture design.

On perceptible binaural cues analysis, the dissertation studies the frequent dependency of the binaural cues threshold perception on the full audible bark band, proposes a frequency dependent spatial parameter representation model to remove redundancy, cutting down 4.28% ~ 11.78% of the parameter bitrate while increasing the audio quality: MUSHRA by 2.6 points and ODG by 1.1 grades averagely.

On perceptible spatial information estimation, the dissertation builds the Binaural Cue Physiological Perception Model (BCPPM) based on spatial hearing theory, overcoming defects of perception entropy theory which based on monaural hearing model. Then proposes Spatial Perception Entropy (SPE) and its computation method, which advancing the theory and practice of spatial audio coding technology.

On binaural cues representation, the dissertation proposes a com-

plex extension of MDCT based on conjugate window function through the basis function “sum split” technique, resolving the difficulty to obtain energy and phase information in MDCT. Then provides a fast MDCT-MDST conversion algorithm for the high efficiency representation based on sparse transition matrices and complexity reduced from $O(N^2)$ to $O(N)$ while increasing the audio quality; MUSHRA by 2.07 points and ODG by 0.4281 grades averagely.

On coding architecture design, the dissertation proposes a parallel filter bank-share Analysis-by-Synthesis (ABS) structure, connecting more closely the spatial parameter coder and the core coder, reducing 5% of system delay, 38% ~ 86% of operation complexity.

The theoretical and technological advancements of this dissertation may hopefully back the national mobile audio development; enhance China's core competence in the fast growing global mobile audio industry and the related international standardization process.

Keyword: Mobile Audio; Spatial Audio Coding; Perceptual Entropy; Binaural Cues; Modified Discrete Cosine Transform



武汉大学优秀博士学位论文文库

已出版:

- 基于双耳线索的移动音频编码研究 / 陈水仙 著
- 多帧影像超分辨率复原重建关键技术研究 / 谢伟 著
- Copula函数理论在多变量水文分析计算中的应用研究 / 陈璐 著
- 大型地下洞室群地震响应与结构面控制型围岩稳定研究 / 张雨霆 著
- 迷走神经诱发心房颤动的电生理和离子通道基础研究 / 赵庆彦 著
- 心房颤动的自主神经机制研究 / 鲁志兵 著
- 氧化应激状态下维持黑素小体蛋白低免疫原性的分子机制研究 / 刘小明 著
- 实流形在复流形中的全纯不变量 / 尹万科 著
- MITA介导的细胞抗病毒反应信号转导及其调节机制 / 钟波 著
- 图书馆数字资源选择标准研究 / 唐琼 著
- 年龄结构变动与经济增长: 理论模型与政策建议 / 李魁 著
- 积极一般预防理论研究 / 陈金林 著
- 海洋石油开发环境污染法律救济机制研究 / 高翔 著
——以美国墨西哥湾漏油事故和我国渤海湾漏油事故为视角
- 中国共产党人政治忠诚观研究 / 徐霞 著
- 现代汉语属性名词语义特征研究 / 许艳平 著
- 论马克思的时间概念 / 熊进 著
- 晚明江南诗学研究 / 张清河 著

目 录

第1章 绪论	1
1.1 研究背景和意义	1
1.2 国内外研究现状	4
1.2.1 冗余信息去除技术	6
1.2.2 感知信息度量方法	9
1.3 面临的关键问题	12
1.3.1 空间参数主观冗余难以去除	12
1.3.2 MDCT 域空间参数分析困难	12
1.4 主要工作和全文组织结构	13
1.4.1 主要工作	13
1.4.2 全文组织结构	15
第2章 双耳线索的特征分析与表示	17
2.1 引言	17
2.2 双耳线索的特征分析实验系统	19
2.2.1 特征分析实验系统设计	20
2.2.2 双耳线索临界感知的频率依赖性	22
2.2.3 双耳线索的统计分布特性	28
2.3 编码系统中的双耳线索编码实现	32
2.3.1 基于频率感知的参数集选取	32
2.3.2 基于频率感知的参数量化和熵编码	34
2.4 实验与性能分析	38
2.4.1 参数码率分析与比较	38
2.4.2 主观性能分析与比较	40

2.4.3 客观性能分析与比较	41
2.5 本章小结	42
第3章 双耳线索感知信息的度量	44
3.1 引言	44
3.2 基于双耳听音的生理感知模型构建	47
3.2.1 空间信息的物理层表示	48
3.2.2 空间信息的生理感知	52
3.2.3 双耳线索生理感知模型 BCPPM	55
3.3 基于 BCPPM 模型的空间感知熵计算	59
3.3.1 空间感知熵的定义	60
3.3.2 临界频带滤波单元	61
3.3.3 双耳线索计算单元	62
3.3.4 有效感知量计算单元	63
3.4 实验与性能分析	65
3.4.1 立体声序列的信息量估算方法对比	66
3.4.2 不同序列 SPE 数值的差异分析	68
3.4.3 同一序列 SPE 随时间的变化分析	69
3.5 本章小结	70
第4章 MDCT 域双耳线索的分析与合成	72
4.1 引言	72
4.2 基于共轭窗函数的 MDCT 复数化扩展	74
4.2.1 双耳线索的形成机理	75
4.2.2 MDCT 频谱对双耳线索的影响	79
4.2.3 基于共轭窗函数的 MDCT 复数化扩展	88
4.2.4 实验与性能分析	95
4.3 基于 C2S 转换的双耳线索分析	98
4.3.1 MDCT-MDST 快速转换	98
4.3.2 C2S 转换矩阵的稀疏化近似	103
4.3.3 CMCLT 域的双耳线索表示	108

4.4 实验与性能分析	113
4.4.1 主观性能分析与比较	113
4.4.2 客观性能分析与比较	116
4.5 本章小结	118
第5章 基于合成分析的空间音频编码结构	120
5.1 引言	120
5.2 空间音频编码系统的基础操作模块	122
5.2.1 声源分离模块	123
5.2.2 下混/上混模块	125
5.2.3 空间参数分析模块	128
5.2.4 系统结构的优缺点	129
5.3 并行滤波共享的 ABS 编码结构	130
5.3.1 频带划分对空间信息分析的影响	130
5.3.2 ABS 编码结构设计	134
5.3.3 并行结构下的滤波共享	138
5.4 实验与性能分析	143
5.4.1 系统延时分析	144
5.4.2 计算复杂度分析	145
5.4.3 系统性能分析	146
5.5 本章小结	147
第6章 总结与展望	149
6.1 研究工作总结	149
6.2 研究展望	151
参考文献	153
致谢	165
附录	166

第 1 章 绪 论

1.1 研究背景和意义

当前，我国的移动通信产业迎来了前所未有的发展机遇。从中国移动发布的 2007 年报可以看到：中国移动用户总数截至 2007 年已高达 3.6934 亿，庞大的移动通信用户群带动了移动音频增值业务市场规模的快速增长，移动音乐点播、手机电视和移动应急通信等增值业务收入也在逐年提升，如图 1-1 所示。

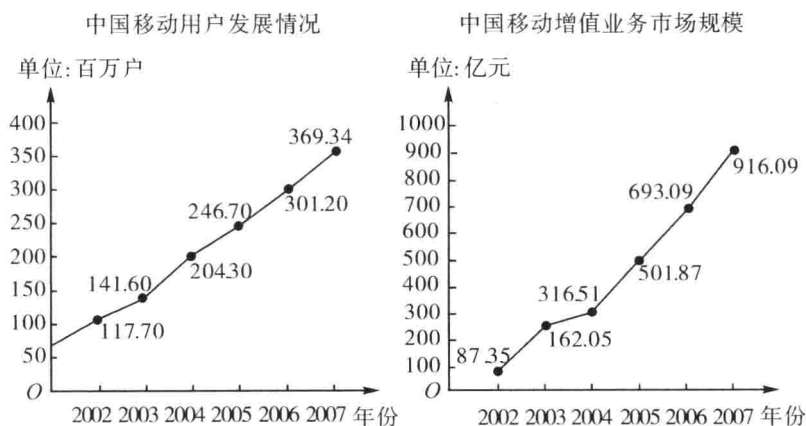


图 1-1 用户发展及增值业务市场规模(摘自中国移动 2007 年报)

本论文得到国家自然科学基金重点项目(60832002)和国家 863 高新技术研究发展计划基金子项目(2004AA119010)的资助。

面对如此巨大的产业,国内企业的利润空间相对较小,在总体技术上难以超越发达国家。究其原因,主要是由于我国音视频产品中具有自主知识产权的专利技术少,数字音视频编码标准长期以来一直由欧美、日本等跨国公司所制定的 MPEG 系列标准垄断。MPEG 系列标准是数字电视、高密度数字存储媒体、宽带网络多媒体、移动多媒体通信等产业所需要的共性关键技术,掌握核心技术标准就意味着取得了巨大的竞争优势。因此,积极开展数字音视频编解码关键技术研究,将有利于推动移动通信产业的发展,增强我国在移动音视频编解码领域的国际竞争力。

众所周知,音频编码的根本目的是在降低码率的同时提供高质量的听觉感受,感知音频编码思路正是从人类的听觉特性出发,通过引入不可察觉的失真提高压缩比,实现在保持听觉质量的条件下降低码率。人耳听觉门限、噪音和纯音的频率掩蔽模型,以及时域超前、滞后掩蔽模型等一系列与人类听觉特性相关的研究成果^[1,2,3,4]已经广泛应用于当前各种感知音频编码器中,实现了48~64kbps/ch下的透明音质编码,满足数字音频广播、音频内容存储、基于固定网络的音频内容分发等领域的业务需求。但是,48~64kbps/ch的码率对移动等内存容量小、带宽资源受限的应用环境来说仍然过高,传统感知音频信号在压缩理论和编码技术上难以满足新兴移动音频增值业务对低码率的要求。

近年来,空间心理声学、脑科学、认知心理学等学科对人类听觉的研究逐渐深入,随之出现了利用人类空间方位感知参数展开的音频压缩技术研究,称之为空间音频编码技术(Spatial Audio Coding, SAC)。该技术将左右耳输入信号通过一个下混声道和若干个描述声场信息的空间感知参数来表征,如图1-2所示。通常空间感知参数只消耗2~8kbps/ch的码率^[5,6],克服了传统技术中码率随声道数近似线性增长的缺陷,在满足人们对高音质追求的同时,大大降低了对传输带宽和存储空间的需求。

空间音频编码技术中反映人耳对空间方位感知的参数称为空间线索。双耳线索作为当前主要的空间线索,其准确提取、表示和编码压缩是空间编码技术研究的重点。

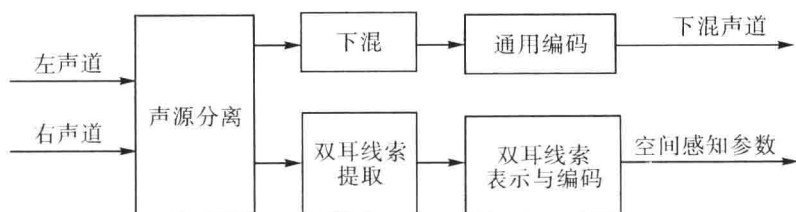


图 1-2 空间音频编码系统主要技术模块(编码端)

1. 在双耳线索提取方面

耳间时间差 (Interaural Time Differences, ITD)、耳间强度差 (Interaural Level Differences, ILD) 和耳间相关度 (Interaural Coherence, IC) 是研究的主要对象。考虑到 ILD 和 ITD 都可独立产生人耳对音源的空间方位感知, 且在音箱回放条件下 ITD 的听觉重要性较低, 当前以 MPEG-4 音频标准为代表的空间音频编码系统 Eaac+, 在全频带只提取 ILD 和 IC 两个参数, 降低了 27% 的空间参数码率^[7]。然而, ITD 参数的缺失导致了信噪比下降, 重建的声像位置发生偏移, 在耳机回放条件下尤为明显^[8]。若增加 ITD 参数的提取, 参数码率将随之增加, 难以满足移动应用中对低码率需求。因此, 在保证编码音质的前提下, 降低空间参数码率是当前空间音频编码研究的热点问题之一, 亟待新的理论和方法创新。

2. 在双耳线索表示和编码方面

在空间音频编码系统主要技术模块中, 声源分离是双耳线索提取、表示和编码的基础。由于独立声源分离技术复杂度高, 通常采用时频分析技术替代。主要的时频分析包括人工耳蜗滤波器组^[9] (Cochlear Filter Bank, CFB)、短时 DFT^[10] 和混合正交镜像滤波器组^[11] (Hybrid Quadrature Mirror Filterbank, HQMF) 等。这些时频分析工具独立于通用编码器中的时频分析, 引入了滤波延时, 增加了系统复杂度, 难以满足移动应用苛刻的复杂度要求。因此, 从时频分析工具和编解码结构两个方面着手, 研究具有低复杂度的变换域参数表示方法及其编解码结构空间音频编码的另一个研究热点。

为此, 本文在国家自然科学基金重点项目“移动音频编解码基