

学校编码: 10384

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_

学 号: 31520101153172

UDC \_\_\_\_\_

厦 门 大 学

硕 士 学 位 论 文

哼唱自动记谱的研究与实现

Research and Implementation of Automatic Singing  
Transcription

王文涛

指导教师姓名: 冯 寅 副教授

专 业 名 称: 计 算 机 技 术

论文提交日期: 2013 年 5 月

论文答辩日期: 2013 年 月

学位授予日期: 2013 年 月

答辩委员会主席: \_\_\_\_\_

评 阅 人: \_\_\_\_\_

厦门大学博硕士学位论文摘要库

## 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为( )课题(组)的研究成果,获得( )课题(组)经费或实验室的资助,在( )实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学博硕士学位论文摘要库

## 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士学位论文摘要库

## 摘要

音乐记谱是音乐人辨识音乐旋律并寻求乐谱表达的一种智力行为，而哼唱自动记谱则是利用计算机技术将哼唱信号转换为一系列连续独立的音符，并且识别出音符的音高以及时长信息，从而来模拟人的这种记谱行为。随着多媒体技术的发展和音乐的数字化进程，在音乐传播的过程中能够自动得到音乐的乐谱信息将帮助人们更好地欣赏音乐。

目前的哼唱自动记谱研究大多基于信号处理或模式分类的方法，却往往忽略了音乐的表现和感知与音乐内容识别计算模型的关系，造成自动记谱效果一直不太理想。本文旨在通过对音乐听觉特征分析，结合信号处理方法与音乐乐理知识规则，对哼唱自动记谱展开了以下几个方面的研究：

1. 针对音高估计算法常见的倍频估算错误，提出了一个基于倍音列模型的精确音高提取算法，该算法依据快速傅里叶变换和倍音列理论，从信号变采后识别音高变化的关系出发，构建了变采情况下音高结果满足的规则约束用于估算信号的精确音高。经实验验证，算法较好地解决了音高识别中的“八度误差”问题。

2. 提出了一个多层次的音符切分算法，该算法综合考虑了音符边界产生的响度变化及音调变化原因。针对音符内不同部分的感知特征，基于音符内容判断音符稳定音区及颤音区边界；并通过 DKL 函数曲线再次寻找音符的可能边界。

3. 对 Krumhansl-Schmuckler 调式识别算法进行了分析和改进，根据哼唱数据集寻找更符合实际哼唱的调式音阶权值分布。

结合以上算法和研究，实现了一个哼唱自动记谱系统，系统具有良好的实用价值。通过实验验证了本文提出算法的有效性，哼唱自动记谱的结果更加符合哼唱者期望表达的旋律。

**关键词：**哼唱自动记谱；基频提取；多层次音符切分；调式识别

厦门大学博硕士学位论文摘要库



## Abstract

Music notation is a kind of intelligent behavior that musicians identify the melody and express them in musical scores. Automatic singing transcription refers to the analysis of a music signal in order to produce a parametric representation of the sounding notes with the pitch and duration. It is a technology to simulate the manually notation. With the development of multimedia and the procession of digital music, get the musical sores automatically over the music communication will enable people to appreciate music better.

The automatic singing transcriptions studies are mostly based on signal processing or pattern classification, but often neglected the relationship between the music perception and the notes content identification calculation models, thereby causing the result has been less than ideal. This thesis concerns the automatic transcription of pitched notes in humming audio. It aims to through the analysis of auditory features, combined with signal processing methods and music theory, and expands the study of the following aspects.

1. With regard to the harmonic errors in the existing pitch estimation algorithm, an accurate pitch extraction algorithm based on the model of overtone is proposed. It is based on FFT and overtone theory, considering the relationship between the pitches estimate from the changed-simple signals, and the detection rules is constructed. The experiment results show that the algorithm solves the problem of octave error.

2. A multi-level note segmentation algorithm is proposed. It concerned the diffident note boundary causes of change in pitch and increase in intensity. The algorithm extracts the perceptually relevant boundaries based on the note content, and detects the potential onsets using DKL function.

3. The analysis and improvement is implemented to the Krumhansl-Schmuckler key finding algorithm. A better tonal scale weight distribution which is more in line with the actual singing is found from the humming dataset.

A singing transcription system is presented based the algorithms proposed in this thesis, which has good practical value. The effectiveness of the proposed algorithms is verified by experiments, and the results are more in line with the hum desired.

**Key words:** automatic singing transcription; pitch estimation; multi-level note segmentation; key estimation

厦门大学博硕士学位论文摘要库

<b>摘要</b> .....	<b>I</b>
<b>Abstract</b> .....	<b>III</b>
<b>第一章 绪论</b> .....	<b>1</b>
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.3 论文的主要内容.....	4
1.4 论文结构.....	5
<b>第二章 音乐的表示与感知</b> .....	<b>7</b>
2.1 音乐术语.....	7
2.2 哼唱的表现与感知.....	9
2.2.1 哼唱的表现.....	9
2.2.2 音乐的感知.....	11
2.3 音乐信号的表示与特征分析.....	12
2.3.1 音乐信号的表示.....	12
2.3.2 音乐信号的特征分析.....	14
2.4 本章小结.....	17
<b>第三章 音乐基频特征提取</b> .....	<b>19</b>
3.1 YIN 算法.....	19
3.1.1 自相关法.....	19
3.1.2 YIN 算法的实现.....	20
3.2 基于倍音列模型的精确音高识别算法.....	24
3.2.1 快速傅里叶变换.....	25
3.2.2 基于倍音列模型的基频提取.....	26
3.2.3 变采音高识别及倍音替换.....	28
3.2.4 基于变采音高关系的精确音高识别.....	30

3.3 本章小结.....	33
<b>第四章 音符切分算法 .....</b>	<b>35</b>
4.1 基于音符内容的音符切分算法.....	35
4.1.1 有声/无声的识别 .....	36
4.1.2 基于音符内容的音符初步切分 .....	38
4.2 基于能量特征距离的音符切分算法.....	40
4.2.1 检测函数生成 .....	40
4.2.2 峰值提取.....	42
4.2.3 多层次音符边界的确定.....	43
4.3 基于音程的音符音高估算.....	46
4.4 基于调式识别的音高规整.....	48
4.4.1 Krumhansl-Schmuckler 调式识别算法.....	48
4.4.2 基于实际哼唱的调式识别 .....	50
4.4.3 音高规整.....	51
4.5 本章小结.....	52
<b>第五章 哼唱自动记谱系统的实现与实验评估 .....</b>	<b>53</b>
5.1 系统设计 .....	53
5.2 实验评估.....	54
5.2.1 基频提取算法评估.....	54
5.2.2 音符切分算法评估.....	56
5.2.3 调式识别算法评估.....	57
5.3 本章小结.....	58
<b>第六章 总结与展望 .....</b>	<b>59</b>
<b>参考文献.....</b>	<b>61</b>
<b>攻读硕士学位期间发表的论文.....</b>	<b>65</b>
<b>致谢.....</b>	<b>67</b>

## Contents

<b>Abstract in Chinese</b> .....	<b>I</b>
<b>Abstract</b> .....	<b>III</b>
<b>Chapter 1 Introduction</b> .....	<b>1</b>
<b>1.1 Research Background and Motivation</b> .....	<b>1</b>
<b>1.2 Related Research</b> .....	<b>2</b>
<b>1.3 Main Work of the Thesis</b> .....	<b>4</b>
<b>1.4 Organization of the Thesis</b> .....	<b>5</b>
<b>Chapter 2 Music Representation and Perception</b> .....	<b>7</b>
<b>2.1 Music Terminology</b> .....	<b>7</b>
<b>2.2 Singing Representation and Perception</b> .....	<b>9</b>
2.2.1 Singing Technique.....	9
2.3.2 Aural Singing Perception .....	11
<b>2.3 Audio Format and Feature Analysis</b> .....	<b>12</b>
2.3.1 Audio Format.....	12
2.3.2 Audio Feature Analysis .....	14
<b>2.4 Conclusion</b> .....	<b>17</b>
<b>Chapter 3 Pitch Extraction</b> .....	<b>19</b>
<b>3.1 The YIN Algorithm</b> .....	<b>19</b>
3.1.1 Autocorrelation Method .....	19
3.1.2 The YIN Pitch Estimator .....	20
<b>3.2 Accurate Pitch Extraction Algorithm based on Overtone Series</b> .....	<b>24</b>
3.2.1 FFT .....	25
3.2.2 Pitch Extraction Algorithm based on Overtone Series.....	26
3.2.3 Pitch Extraction based on Sample-Changes and Overtone Replace.....	28
3.2.4 Accurate Pitch Extraction based on Pitches Relationship.....	30
<b>3.3 Conclusion</b> .....	<b>33</b>

<b>Chapter 4 Note segmentation .....</b>	<b>35</b>
<b>4.1 Content-based Note Segmentation Algorithm.....</b>	<b>35</b>
4.1.1 Breathy Voice Detection.....	36
4.1.2 Preliminary Note Segmentation based on Note Content.....	38
<b>4.2 Note Segmentation Algorithm based on Amplitude Distance.....</b>	<b>40</b>
4.2.1 Detection Fuction.....	40
4.2.2 Peak Picking .....	42
4.2.3 Multi-level Boundary Determination.....	43
<b>4.3 Note Pitch Detection based on Interval.....</b>	<b>46</b>
<b>4.4 Key Estimation and Note Sequence Adjustment .....</b>	<b>48</b>
4.4.1 Key Estimation based on Krumhansl-Schmuckler.....	48
4.4.2 Key Estimation based on Actual Humming .....	50
4.4.3 Note Sequence Adjustment .....	51
<b>4.5 Conclusion.....</b>	<b>52</b>
<b>Chapter 5 Automatic Singing Transcription System and Evaluation</b>	<b>53</b>
<b>5.1 Automatic Singing Transcription System .....</b>	<b>53</b>
<b>5.2 Evaluation .....</b>	<b>54</b>
5.2.1 Pitch Extraction Algorithm Evaluation .....	54
5.2.2 Note Segmentation Algorithm Evaluation.....	56
5.2.3 Key Estimation Algorithm Evaluation .....	57
<b>5.3 Conclusion.....</b>	<b>58</b>
<b>Chapter 6 Conclusion and Future Work .....</b>	<b>59</b>
<b>Bibliography .....</b>	<b>61</b>
<b>Published Papers.....</b>	<b>65</b>
<b>Acknowledgement.....</b>	<b>67</b>

## 第一章 绪论

### 1.1 研究背景与意义

音乐是时间的艺术，但是时间稍纵即逝，美好的音乐一去不回。为了保留声音，人们发明了乐谱，试图通过记谱使音乐长久流传。传统的音乐记谱是音乐人聆听一段音乐并记录下所包含的音乐音高、节奏、调式等信息，一般采用五线谱或简谱形式。随着多媒体技术的发展和音乐的数字化进程，数字化的声音和声音处理技术极大丰富了音乐效果的表现力，使音乐的质量和音乐构造能力表现出一个前所未有的巨大飞跃。在音乐传播的过程中，能够自动得到音乐相应的乐谱信息将帮助人们更好地欣赏音乐，从而充分利用丰富的计算机音乐素材。

音乐记谱是音乐人辨识音乐旋律并寻求乐谱表达的一种智力行为，而自动音乐记谱（Automatic Music Transcription）则是利用计算机技术将音乐旋律信号转换为符号化、标准化的乐谱，从而模拟人的这种记谱行为。根据输入旋律片段对象的不同，自动音乐记谱可以分为器乐记谱和声乐记谱。器乐记谱是对器乐发出的标准音乐信号记谱转化的过程，由于其旋律的音高和时值比较标准，器乐音乐自动记谱的实现过程比较简单，且转化效率也比较高；声乐音乐记谱旋律输入则为人声哼唱信号，由于人在哼唱表演时往往加入个人有表现力的改动和个性化的感情因素，而且哼唱过程中对音乐不能很好的把握，音高误唱和节奏错误在哼唱过程中也时有发生，因此声乐音乐记谱的实现相对较为困难，但有更广的应用价值。

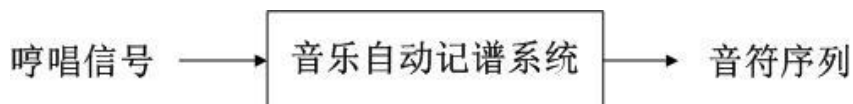


图 1.1 哼唱自动记谱系统

哼唱自动记谱就是将人声哼唱信号转换为一系列连续的独立的音符，并且识别出音符的音高以及时长信息。也就是把一段连续的音乐理解为一系列具有不同

音高不同持续时间的音符序列。图 1.1 显示了哼唱自动记谱系统的流程。

哼唱自动记谱的研究方法及应用对计算机音乐相关领域具有相当大的推动作用。哼唱自动记谱的实现是基于内容音乐检索的研究中最为关键的部分，其对音乐特征的识别，如旋律、和声、调式能够为音乐检索系统提供多样化、标准化、适用于检索的特征，音乐识别的精度会直接影响音乐查询系统的性能；自动记谱技术还广泛应用于音乐自动伴奏、音乐情感计算、音乐的知识表达和标注以及电子音乐设备的输入等领域。另外，哼唱自动记谱与音乐描述和音乐分析密切相关，它的研究对传统音乐学习与创作起到促进作用。自动音乐记谱可以极大的方便音乐人自动化的音乐分析、乐谱标记及音乐特征的统计，帮助作曲家在哼唱识别乐谱的基础上进行创作；自动记谱系统的输出结果还可以对音乐歌唱或乐曲演奏的效果进行合理的评价，这必然有助于提高人们对自己音乐水平的认识，也会促进其在音乐表演方面的进步。

本文旨在通过对音乐听觉特征分析，结合信号处理方法、音乐乐理知识规则，研究对哼唱音乐信号进行基频提取以及音符边界识别，并实现了一个哼唱自动记谱系统，最后通过实验分析验证了方法的可行性。

## 1.2 国内外研究现状

从 20 世纪 70 年代开始，国内外从事自动记谱研究的学者不同程度的做了各种有益的探索，并取得了不错的成绩。

Dressler<sup>[1]</sup>采用短时傅立叶变换和正弦建模来分析频谱，选取频谱峰值并计算瞬时频率，然后基于谐波特征估计基音，在后处理阶段使用平滑规则快速提取出旋律；Fujihara 和 Goto 等人<sup>[2]</sup>通过统计嗓音模型和 Viterbi 算法实现旋律提取，他们首先基于多基音估计来完成音高估计，然后计算每一个候选音高的谐波结构概率，接下来使用高斯混合模型来判别音高，最后使用 Viterbi 算法实现最优音高序列查找；Ryynanen<sup>[3,4]</sup>和 Krige<sup>[5,6]</sup>基于 HMM 模型，采用多基频估计方法，结合声学模型和音乐学模型以及音符过渡模型进行了歌曲旋律的提取；Poliner<sup>[7]</sup>运用了 SVM 方法进行旋律提取，通过支持向量机分类器将音符分开，并将这种方法扩展到学习音乐结构和风格。



Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士学位论文摘要库