

基于内容的在线音乐 检索系统调研

小组成员：

00930005 周 妍

00930029 汪丹华

00946002 刘焱远

目录

1 导言	1
1.1 选题背景.....	1
1.2 文献调研情况.....	1
1.3 基于内容的音乐检索技术综述.....	2
1.4 基于内容的音乐检索系统基本结构.....	2
1.5 研究目的与研究意义.....	3
2 研究方法和步骤.....	3
2.1 主要方法和工具.....	3
2.2 研究思路与基本步骤.....	4
3 系统选取与介绍.....	4
3.1 系统选取.....	4
3.2 系统简介.....	4
3.2.1 Folk Tune Finder	4
3.2.2 Musipedia.....	5
3.2.3 Songtapper	5
3.2.4 Themefinder.....	5
3.2.4 Themefinder.....	5
3.2.5 Midomi	5
3.2.6 Track id	6
3.2.7 爱乐搜.....	6
3.2.8 Shazam	6
4 系统调研情况.....	6
4.1 结构化音乐检索系统.....	6
4.1.1 Folk Tune Finder	6
4.1.2Musipedia.....	7
4.1.3 Songtapper	7
4.1.4 Themefinder.....	8
4.2 基于样本的音乐检索系统.....	8
4.2.1 Midomi	9
4.2.2 Track ID.....	9
4.2.3 爱乐搜.....	9
4.2.4 Shazam	9
4.3 系统比较.....	9
4.3.1 结构化音乐检索系统比较.....	9
4.3.2 基于样本的音乐检索系统比较.....	10
4.3.2 两类基于内容的音乐检索系统比较.....	11
5 基于调查结果的改进模型.....	11
5.1 基于现有系统功能的改进方案.....	11
5.1.1 结合两种音乐系统.....	11
5.1.2 结果反馈.....	11
5.1.3 情感语义搜索.....	11

5.2 可以尝试的新功能.....	12
5.2.1 上传文件检索功能.....	12
5.2.2 检索帮助服务,	12
5.2.3 搜索结果过滤改进.....	12
5.3 改进系统的模型结构.....	12
6 相关问题讨论.....	13
6.1 数据来源.....	13
6.2 匹配精度与系统评价.....	13
6.3 扩展能力.....	13
6.4 检索平台延伸.....	13
6.5 未来发展趋势.....	14
7 结语	14
附录	14
附录 1 基于内容的音乐检索系统技术简介.....	14
1.1 音高检测.....	14
1.2 位置敏感哈希函数 (LSH)	15
1.1.3 动态时间规整 (DTW)	15
1.4 音乐的存储格式及其表示.....	15
1.4.1 乐谱表示.....	15
1.4.2 MIDI 表示.....	16
1.4.3 波形表示.....	16
附录 2 小组分工.....	16
调研过程分工:	16
报告撰写分工:	16
附录 3 参考文献推荐.....	16
第一篇.....	16
文献基本信息.....	16
推荐理由.....	17
第二篇.....	17
文献基本信息.....	17
推荐理由.....	17
参考文献.....	17

1 导言

1.1 选题背景

20 世纪 80 年代以来,随着计算机技术在各方面的突破性进展,人们的信息处理需求已经不再满足于计算机的数据运算和文字处理等传统应用,而是开始重视利用计算机形象逼真地表达自然现象和社会状况。于是,研究人员开始对图形、图像、动画、音频、视频等多种不同于传统文本的媒体信息的计算机化处理进行探索,多媒体信息处理技术也由此应运而生。90 年代以后,多媒体技术日渐成熟,各种信息处理的工具、技术、载体、方法和手段等得到了进一步的综合集成。

面对用户越来越多的对图像、音频、视频等非文本信息的检索需求,研究人员逐渐发现传统的文本检索思路——“关键词+布尔逻辑”在很多情况下已无法满足用户对多媒体信息检索的实用性要求,多媒体信息检索需要依赖和使用一种更接近人类视、听觉感知能力的内容描述方法。由此便产生了针对多媒体信息的“基于内容检索”的思想。

音频信息是最重要、最具代表性的一种多媒体信息,又可分为语音信息、音乐信息、波形信息三种类型。近年来,随着音乐资源的不断增多,人们迫切需要便捷地选择和管理音乐的方法,基于内容的音乐检索因此成为近几年的研究热点之一,日益受到相关领域研究人员的重视。在这种情况下,基于内容的音乐检索技术不断发展,并产生了一批以此为基础的音乐检索系统。近年来,基于内容的音乐检索系统不断涌现,并且逐步由实验室系统走向普通搜索引擎,走向商业化应用的检索系统。检索平台也逐渐由笔记本延伸到手机。可以说,基于内容的音乐检索系统已经得到了广泛的应用。实际上, Birmingham et al¹提出了哼唱检索的几个商用场景,比如 internet 搜索音乐, mp3 检索等等。此外他们还列出一些应用场景²,比如个人曲库中的,卡拉 OK 点歌中的场景等。

尽管如此,基于内容的音乐检索仍然是一个新兴的研究领域,同时由于其检索对象和范围的复杂性和多样性,在该领域的研究还只是刚刚开始,任重而道远。通过对已经实现的基于内容的在线音乐检索系统进行调研,能够了解基于内容的音乐检索的发展现状,并有助于为该领域的进一步研究提供参考。

1.2 文献调研情况

利用中国期刊网期刊全文数据库、万方数据资源系统的全文数字化期刊数据库、INSPEC 数据库和 Google 学术搜索,可以检索到很多篇有关基于内容的音乐检索方面的文献。通过阅读这些文献可以了解到基于内容的音乐检索的基础知识、发展现状以及未来的可能趋势,以下便基于文献调研对基于内容的音乐检索的基本情况进行简述。

音乐的音频特性可以用分层描述模型来表述。该分层模型共有三层,从最底层到最高层抽象层次逐渐递增。最底层为物理样本集,主要特征有采样率、时间刻度、样本、格式、编码等;中间层为声学特征级,包括单调、音高、旋律、节奏等感知特征和能量、过零率、LPC 系数等声学特征,是音频的结构化表示;最高层为语义级,包括音乐叙事、单频对象描述、语凌识别文本、探测结果等特征。目前基于内容的音乐检索系统主要利用的音频特性位于声

¹Birmingham, W., Dannenberg, R., and Pardo, B. (2006). Query by humming with the VocalSearch system. Communication of the ACM, 49(8), 49-52.

² Birmingham, W. P., Dannenberg, R., Hu, N., Meek, C., Pardo, B., and Tzanetakis, G. (2007). A comparative evaluation of search techniques for query by humming using the MUSART testbed. Journal of the American Society for Information Science & Technology, 58(5), 687-701.

学特征级，包括旋律、节奏、乐谱等等。

基于哼唱的人声输入和基于键盘敲击或鼠标点击的手工输入。旋律的表示方法分为绝对音高序列和相对音高序列两种，其中相对音高序列主要以旋律轮廓的方式呈现，如用 U 表示 Up、D 表示 Down、S 表示 Same，以 U、D、S 的组合代表旋律轮廓。还有一部分文献讨论的是基于语义的音乐检索，主要是基于情感的音乐检索。

在论述基于内容的音乐检索系统的文献中，大部分文献都是提出系统结构、模型或设计方案等，以及对关键技术的研究。也有一部分文献是基于对内部实现的实验系统的分析。

音乐检索系统的发展包括三个阶段。第一代音乐检索系统指的是早期的基于人工标注乐曲名、作者等外部特征提供检索入口的音乐检索系统，这一阶段的系统基于传统的文本检索思路——“关键词+布尔逻辑”。第二代音乐检索系统主要指的是基于旋律哼唱或手工输入的音乐检索系统，目前基于声学特征级音频特性的音乐检索系统即处于这一阶段，这也是基于内容的音乐检索系统的发展现状。音乐检索系统的未来发展趋势是向语义级层次发展，即通过音乐传达的情感或叙事特征来实现检索，这样的系统是第三代音乐检索系统。

1.3 基于内容的音乐检索技术综述

音乐属于声音世界的一部分，是一种在时间维度里组织声音的艺术³。常见的基于内容的音乐检索技术，无外乎是从音高、音程、音色、力度等方面考察音乐的。

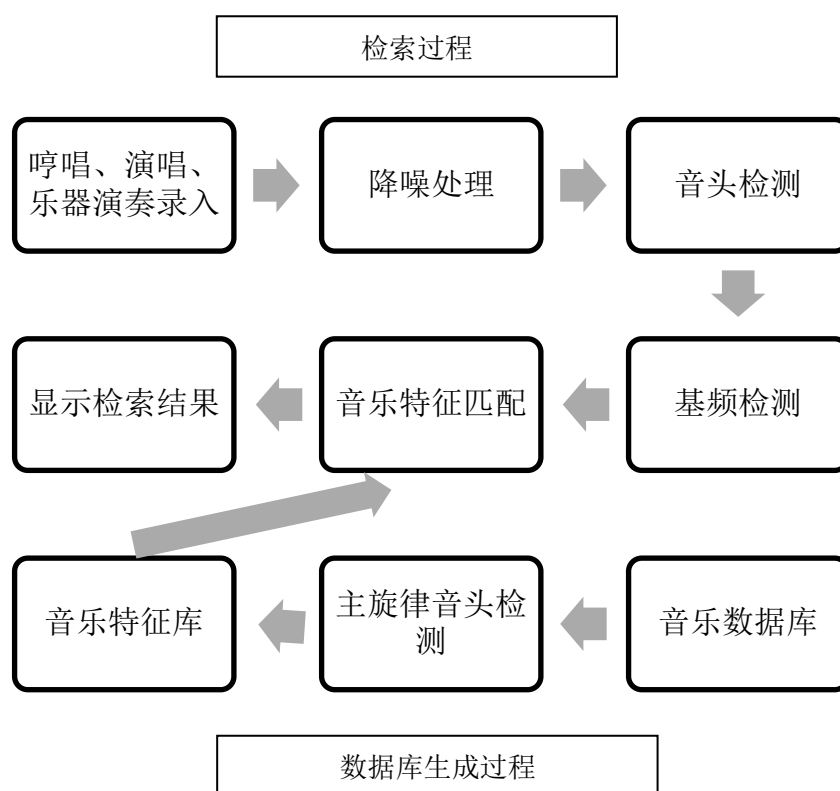
基于内容的音乐检索技术，主要有两大类。一是结构化音乐检索系统所用的匹配技术。这些技术往往是将歌曲结构化的存储在数据库中，使用如 MIDI、ABC、GUI-DO、Musedata、Lilypond 等格式。结构化的格式，往往基于乐谱，更容易抽取可检特征。同时，检索接口也是结构化的。比如输入调式、节奏、几个音符等等。由于本身的数据库是结构化存储的，所以匹配技术本身并不复杂。

另一类是基于样本的音乐检索技术，实际上，更多的音乐格式是以 PCM 现行编码为基础的，比如在网络上最为流行的 MP3，CD 存储使用的 WAV 格式等。由于这些格式本身是基于声波特征进行存储的，所以相对难以提取可检特征。一般通过能量分布、音乐轮廓等进行提取。这类系统，用户的输入接口一般是哼唱、演唱或录音产生的样本。服务器将用户的样本与特征库进行匹配，将交集不为空的乐曲按照一定的度量算法进行排序。由此将匹配程度较高的作为结果输出给用户。

1.4 基于内容的音乐检索系统基本结构

基于内容的音乐检索系统主要由音乐数据库、音乐特征库组成。主要的功能，一是进行检索，而是进行数据库的生成，即添加新的乐曲进入数据库，具体的工作流程如下页图。

³(美)罗杰凯密恩著，王美珠等译. 听音乐-音乐欣赏教程. 世界图书出版公司，2008.



1.5 研究目的与研究意义

目前基于内容的在线音乐检索系统主要通过分析音乐的旋律、节奏等特征来进行检索，不同的系统侧重的检索基础不同。我们希望通过对所选取的音乐检索系统的试用和调研，了解各个系统的基本情况和检索功能，并对检索结果进行评价。

在熟悉各个系统的基础上，我们希望对系统进行比较分析，并总结各个系统的设计和技术特色，进而找到其适用的领域，分门别类地为不同需求的音乐检索提供系统选择指导。

同时，结合试用的切身感受，我们希望能在此基础上建立一个优化的基于内容的音乐检索系统模型。该模型结合了从技术和用户需求与用户体验两个角度提出的改进建议。此外，我们还会对基于内容的音乐检索系统所涉及到的相关问题进行讨论，提出我们自己的观点和见解。

在本次调研基础上得到的调研结果、优化模型以及对相关问题的见解，对于现有的各基于内容的音乐检索系统而言，具有指导其进一步改进、完善，提升用户体验的意义；对于新系统的建设，则有来自真实用户的建议参考、启发价值；而对于普通用户，本次调研成果也不失为一份评价各检索系统的参照指南。

总之，本次作业通过对现有的具有代表性的基于内容的在线音乐检索系统的调研，有助于掌握基于内容的音乐检索系统的发展现状，并对基于内容的音乐检索的发展提出更好的方向和思路，为音乐检索系统的进一步研究提供参考和启发。

2 研究方法和步骤

2.1 主要方法和工具

本次作业的主要方法为调查法，重点体现在对各个系统的检索实践。使用的工具包括浏览器、必要的插件以及手机软件。在具体操作上，通过一些乐曲的特征输入，对检索系统进

行实际的使用，从而进一步分析检索结果，以对系统进行调查和评价。

2.2 研究思路与基本步骤

首先通过文献调研了解基于内容的音乐检索的基本情况和现状，然后利用期刊数据库和网络搜索引擎等工具收集具有代表性的基于内容的在线音乐检索系统。在调查系统基本情况的基础上，对各个系统进行具体试用——针对系统不同的检索功能构造相应的提问式或进行相应的输入，来进行具体的检索实践，并对得到的检索结果进行分析评价。接着将各个系统进行比较和总结。然后在此基础上提出一个更加满足用户需求与体验的且在技术上加以改进的基于内容的音乐检索系统的优化模型。最后进行相关问题讨论，提出自己的观点和见解。

3 系统选取与介绍

3.1 系统选取

对于本次调研来说，系统的选择正确与否、是否具有代表性，对调研的可信度、后续研究的开展、系统的优化改进建议等都具有重要意义。因此，本调研基于以下原则进行系统选取：

- (1) 所选系统应类型全面，既要包括研究性、实验性的检索系统，又要包括商业性检索系统；既要包括结构化音乐检索系统，又要包括基于样本的音乐检索系统。
- (2) 所选系统需正式投入使用，并已取得一定的基础。
- (3) 所选系统需提供典型的基于音乐的声学特征级音频特性的检索功能。
- (4) 所选系统需是在线的网络音乐检索系统。
- (5) 所选系统需被较为广泛地使用，具备一定的关注度。

基于上述原则，我们选择了以下八个基于内容的音乐检索系统进行调研（其中前四个是结构化音乐检索系统，后四个是基于样本的音乐检索系统）：

检索系统名称	来源
1. Folk Tune Finder	http://www.folktunefinder.com/
2. Musipedia	http://www.musipedia.org/
3. Songtapper	http://www.bored.com/songtapper/
4. Themefinder	http://www.themefinder.org
5. Midomi	http://www.midomi.com
6. Track ID	手机软件
7. 爱乐搜	手机软件（中国国产）
8. shazam	http://www.shazam.com

3.2 系统简介

3.2.1 Folk Tune Finder

Folk Tune Finder 音乐检索系统建立于 2008 年，是由英国牛津大学的软件专家 Joe Wass 开发的实验性系统。该系统建立后更新过两次，目前提供在线使用的版本是 Joe 在第一、二版的基础上根据用户意见修改并于 2010 年 3 月起推出的第三版。该系统采用网页抓取的方式收录了以 ABC 格式录制的超过 20 万首民歌。由于一首民歌可能存在不同版本，系统收录

的总民歌数超过 40 万首。该系统还允许用户通过注册 FolkTuneFinder Blog 的方式为系统添加乐曲。系统以钢琴键盘作为主要输入接口，也提供曲名检索功能。在检索结果中不仅呈现结果乐曲的五线谱，还提供了 MIDI 格式音乐文件下载和来源网址等信息。

3.2.2 Musipedia

Musipedia 建立于 1997 年，是由奥地利学者 Rainer Typke 创立的商用音乐检索系统。2006 年后，Musipedia 可以实现万维网范围内 MIDI 文件的检索。该系统中融合了一些开源软件，如 Lilypond、Typo3、MySQL 等。截至 2006 年 5 月，该系统采用网页抓取和用户添加方式收录了超过 3 万首乐曲，还特别收录了部分国家的国歌。系统提供了钢琴键盘、键盘敲击、轮廓检索、文本检索、哼唱等多种输入接口，在检索结果中呈现乐曲名称、作者、乐谱、版本、音乐类型等信息，支持在线收听功能，并提供相关外部链接。

3.2.3 Songtapper

Songtapper 音乐检索系统是由加拿大温哥华西蒙弗雷泽大学开发的实验性系统。该系统需要使用开启 Javascript 的网页浏览器和 Macromedia Flash 播放器 6.0 或更高版本。系统采用自建数据库和用户添加的方式收录乐曲，截至 2011 年 12 月共收录乐曲 88438 首。Songtapper 除了提供传统的曲名检索和作者检索两种文本式检索功能外，只提供节奏检索，在其结果中呈现乐曲名称和作者信息，并提供试听链接。

3.2.4 Themefinder

Themefinder 是由美国斯坦福大学计算机辅助人文研究中心和俄亥俄州立大学认知和系统音乐学实验室合作开发的音乐检索系统。该系统自行建立了包含 35000 首乐曲的数据库，包括古典音乐、民歌、文艺复兴音乐 3 种类型。为增加收录乐曲的数量，系统允许用户添加。系统仅提供文本式检索接口，但提供了音高、音程、音级、轮廓、音调、韵律等多种输入维度的组配检索，同时采用精确匹配的方法。检索结果包括作者、曲名、五线谱、试听链接和下载链接等。

3.2.4 Themefinder

Themefinder 是由美国斯坦福大学计算机辅助人文研究中心和俄亥俄州立大学认知和系统音乐学实验室合作开发的音乐检索系统。该系统自行建立了包含 35000 首乐曲的数据库，包括古典音乐、民歌、文艺复兴音乐 3 种类型。为增加收录乐曲的数量，系统允许用户添加。系统仅提供文本式检索接口，但提供了音高、音程、音级、轮廓、音调、韵律等多种输入维度的组配检索，同时采用精确匹配的方法。检索结果包括作者、曲名、五线谱、试听链接和下载链接等。

3.2.5 Midomi

Midomi 基于唱歌、哼歌或者吹口哨搜索。就可以立即找到你喜爱的乐曲，并连接到与你的音乐兴趣相同的社区（比如 facebook）。在搜索过程中用户不需要输入类似歌曲作者、演唱者或者歌曲名等任何关于歌曲的文字信息，而只需要把歌曲哼唱出来，就可以通过 Midomi 的数据库找到匹配的歌。

Midomi 的数据库基于云思想。Midomi 允许用户注册并上传自己演唱、哼唱或者吹口哨的录音，用户把自己的录音上传并加上标签标明标题、坐着等信息。Midomi 正是基于这种强大的云数据库，在搜索时，用户上传约 10 秒的录音，Midomi 的服务器与那些已经上传的录音进行匹配，从而得出用户搜索的音乐的具体信息。

Midomi 以网站和软件两种方式并行运行。用户可以通过 www.midomi.com 网站进行搜

索，也可以在自己的手机上安装 Midomi 的 Soundhound 软件，并进行搜索。Midomi 几乎是免费的，使用网站进行搜索不需要付任何附加费用，而 Soundhound 软件，在 Android、Windows Phone 平台以及 Symbian 平台都是完全免费的。在 iOS 平台上，Soundhound 有免费版和收费版，其主要区别在于免费版有广告。

3.2.6 Track id

TrackID™ 功能常见于 Sony Ericsson 手机中。用户可以通过录制一小段录音，上传到服务器上，服务器根据一定的算法就可以返回相应音乐的标题、作者等信息。虽然 Track ID 服务是免费的，但是由于 Track ID 以一款软件的形式存在，并且只能运行在 Sony Ericsson 手机上，因而其应用范围较窄。

Track ID 服务的数据库源自全球最大的音乐数据公司 Grace Note。Grace Note 涵盖了世界上所有正式出版发行的音乐。因为 Track ID 的检索数据库非常的完备和精确。

3.2.7 爱乐搜

爱乐搜是中国电信推出的一款基于语音搜索和哼唱搜索的手机软件。用户只需对手机说出或唱出喜爱歌手、歌曲，心爱音乐随即从爱音乐曲库中搜出呈现，并可在线试听、歌曲下载以及彩铃订购。可以与歌曲下载及彩铃相结合，是国产软件爱乐搜的特点。

爱乐搜的官方介绍中，并没有详细的介绍其所使用的技术和数据库，因而其背后的技术支持、来源都不得而知。爱乐搜只能运行在 Android 平台上，其余平台不能运行。

3.2.8 Shazam

Shazam 是儿童用的俚语，其意思是指通过一种咒语是某些东西突然出现或者消失。Shazam 的功能如其名字，用户通过上传一小段录音，可以检索到相应的歌曲信息，并连接到如 Youtube 的视频连接、Facebook 分享等。

Shazam 作为一款软件的形式出现，在 iOS、Android、Windows Phone、Blackberry、Symbian 甚至 Java 平台上都可以运行。在每一个平台上，Shazam 都有免费版和收费的 Encore 版本，主要区别在于广告以及 Encore 版本的更美观的界面和更快的检索速度。在 iOS 平台上还有收费的 Red 版，界面是红色，价格与 Encore 版本持平，均为 4.99 美元，但是 Shazam 公司会用 Red 版的收入捐献给非洲，以援助当地的医疗事业。

Shazam 是最早出现的基于样本录音的音乐检索系统。经过近十年的发展，全球有约 1.6 亿人的手机上安装了 Shazam。Shazam 的数据库有超过一千万的歌曲，最早至 20 世纪 50 年代。Shazam 的数据库主要由公司自己构建，可以说是常年累月不断积累下的财富。

4 系统调研情况

4.1 结构化音乐检索系统

4.1.1 Folk Tune Finder

Folk Tune Finder 提供曲名检索和旋律检索两项检索功能。曲名检索是传统的文本式检索功能。旋律检索的输入接口是钢琴键盘的形式，用户点击琴键表示旋律，钢琴下方会显示相应的五线谱作为对照，以此构造检索提问式。除此以外，系统还提供浏览式查询和随机查询两种查询方式。使用浏览式查询，系统会列出所有收录的民歌供用户选择。使用随机查询，系统每次会随机显示一首乐曲。

提问式构造完成后，点击“搜索”按钮，经过短暂的响应时间，即会显示检索结果。初

始检索结果列表包括曲名和乐谱片段信息，并提供在线收听功能。点击曲名可以查看每条结果的详细信息，包括曲名、作曲者、完整乐谱、来源网址链接等信息，还提供在线收听和 MIDI 格式下载功能。

针对该系统的收录范围，我们输入了 The Londonderry Air、Green Sleeves、Scarborough Fair、Last Rose of Summer、Skye Boat Song 等一系列具有代表性的英国民歌的旋律片段进行检索，但由于该系统基于模糊匹配和相对音高序列，因此检索结果中除了目标乐曲以外，还包括一些与提问式的旋律片段具有类似旋律轮廓的乐曲。我们还输入了一些随机的旋律片段进行检索尝试，除个别情况下没有检索结果外，大部分时候系统都会提供包含类似旋律片段的乐曲。

该系统的主要特色是其钢琴键盘的输入接口易于使用，并于下方的五线谱显示形成清晰的对照。不足之处在于系统不支持曲名和旋律的组配检索，无法实现传统文本检索与基于旋律检索的结合。

4.1.2 Musipedia

Musipedia 提供多种输入接口和检索功能。在进行旋律检索时，用户可以限定检索范围（musipedia 或万维网）和旋律出现的位置（初始处、初始附近、任何位置），并与关键词检索相结合。用户可以利用系统提供的虚拟钢琴键盘（JavaScript Piano 和 Flash Piano 两种），采用键盘输入或鼠标点击的方式输入旋律（用户可以在线收听自己输入的旋律片段），也可以采用哼唱的方式输入旋律。系统还提供旋律轮廓检索和节奏检索功能。其中旋律轮廓检索基于 Parsons Code，以 U 表示 Up、D 表示 Down、R 表示 Repeat，用户通过输入 U、D、R 的组合代表旋律轮廓（初始状态用*代表）。在进行旋律轮廓检索时，还可以同时对音乐类型进行限制，这些音乐类型包括古典音乐（classical）、流行音乐（popular）、民歌（folk songs）、圣歌（hymns&carols）和国歌（national anthems）。

初始检索结果列表提供乐曲名称、作者、乐谱、版本、相关外部链接等信息并支持在线收听功能。每条检索结果的详细信息中还包括乐曲所属的音乐类型和乐曲的 Parsons Code（以 U、D、R 表示的旋律轮廓）。

我们输入了 The Londonderry Air、Green Sleeves、Scarborough Fair、Last Rose of Summer、Skye Boat Song、Hotel California、Don't cry for me Argentina、美国国歌等多种类型多首乐曲的旋律片段对系统进行具体的试用。同样由于系统基于模糊匹配和相对音高序列的原因，检索结果中除了目标乐曲以外，还包括一些与提问式的旋律片段具有类似旋律轮廓的乐曲。我们还输入了一些随机的旋律片段进行检索尝试，系统基本都会提供包含类似旋律片段的乐曲。

该系统的特色在于输入接口种类齐全、检索功能丰富，而且支持关键词和旋律的组配检索，在一定程度上实现了传统文本检索与基于内容检索的结合。

4.1.3 Songtapper

Songtapper 除了提供曲名检索和作者检索两种传统的文本式检索功能外，只提供节奏检索功能。用户通过敲击键盘来表示乐曲的节奏，系统依此提问式进行检索。不过，所输入的节奏片段不能太短，至少要敲击十下，否则无法进行检索。此外，系统还提供浏览式查询方式，分别为用户列出被检索次数最多的乐曲（most tapped songs）、最新乐曲（the latest songs）和系统所有收录乐曲的列表（list of all songs）供用户选择。

初始检索结果列表中提供曲名和作者信息。每条结果的详细信息包括曲名、作者和歌词链接，并提供在线收听和下载功能。

我们输入了 Jingle Bells、Happy Birthday To You、Super Mario、Smoke On The Water、Mary Had A Little Lamb、We Will Rock You、Stairway to Heaven、Hey Jude、We Wish You A Merry Christmas 等乐曲的节奏片段进行检索，顺利地检索出了目标曲目，且目标曲目大都排在检

索结果的第一位，效果很好。同时也检索出了部分具有类似节奏的乐曲。

该系统的特色在于其集中的节奏检索，尽管功能相对单一，但是通过系统试用发现该功能效果比较好，具有进一步研究的价值。此外，节奏检索与旋律检索相比对专业知识要求更低，输入也更加方便，易于用户使用。

4.1.4 Themefinder

Themefinder 仅提供文本式检索接口，但支持多种基于声学特征级音频特性的输入维度的组配检索。具体的输入维度包括 Repertory（音乐类型）、Pitch（音高）、Interval（音程）、Scale degree（音级）、Gross contour（粗略轮廓）、Refined contour（精细轮廓）、Location（出现位置）、Key（音调）和 Meter（韵律）。系统提供详细的在线帮助文档以指导用户正确构造提问式。此外，系统还提供作曲家索引（按字顺排列），支持用户查询特定作曲家的乐曲。

我们选取了贝多芬的《命运交响曲》、《致爱丽丝》、《月光奏鸣曲》，莫扎特的《费加罗的婚礼》、《魔笛》，巴赫的《b 小调弥撒曲》、《勃兰登堡协奏曲》，柴可夫斯基的《四小天鹅舞曲》、《如歌的行板》等交响乐曲。利用不同的输入维度构造相应的提问式进行检索，在提问式包含信息较为充分的情况下，检索结果十分准确。

该系统的特色在于其提供了支持多种基于声学特征级音频特性的输入维度组配的文本式检索功能，并有详细的帮助文档作为辅助。系统采用精确匹配的方式，检索结果准确率高。不足之处在于对用户的专业知识要求较高，而且容错性较差。

4.2 基于样本的音乐检索系统

由于四个系统都是基于哼唱或演唱样本检索系统，因而，我们以《费加罗的婚礼》、蔡琴的《渡口》、Eminem 的《Mocking Bird》、老鹰乐队的《加州旅馆》为测评曲目对以演唱样本为接口的检索功能进行调研。另一方面，使用《世界末日》、《思念是一种病》、《如烟》、《双节棍》四首快慢有致的歌曲，对检索系统进行使用和调研。

录音样本测试结果如下：

音乐	爱乐搜	Soundhound	Track ID	Shazam
渡口	√	√	√	√
Mocking Bird	√	√	√	√
Hotel California	√	√	× √	√
费加罗的婚礼	× × ×	× × ×	√	× × ×

哼唱样本测试结果如下：

音乐	爱乐搜	Soundhound	Track ID	Shazam
世界末日	√	√	× × ×	√
思念是一种病	√	× √	× × ×	× × ×
如烟	√	√	× × √	× √
双节棍	× × ×	× × ×	× × ×	× × ×

注：其中√表示正确、×表示错误、每个歌曲最多试验三次，因而×××表示三次均为错误、×√表示第一次错误、第二次正确。

以下对各系统的特色特点及不足进行分析。

4.2.1 Midomi

Midomi 以网站和软件两个形式存在。无论是网站还是软件，其检索过程基本一致，但 Soundhound 软件最终显示的是单一的匹配结果，而网站则会显示一些匹配结果，按照匹配程度进行排序。由于 Midomi 是基于云思想，从用户数据进行采集的。因而检索结果来看，《费加罗的婚礼》这种交响乐就很难有用户上传，同时也就很难被检索到。

在哼唱样本测试方面，在《思念是一种病》中出现小差错，未能检出《双节棍》。总体合格。

4.2.2 Track ID

Track ID 的最大的特点就是:在录音样本测试阶段，Track ID 一次成功的检测出了《费加罗的婚礼》。这体现出 Track ID 在数据库方面的强大和完善。

在哼唱部分，由于 Track ID 软件声明中写清了不支持哼唱或演唱样本输入，尽管如此，Track ID 仍然检出了《如烟》。但其余三首完全检测失败，可以说是 Track ID 的最大缺陷。

4.2.3 爱乐搜

与前几款软件不同，爱乐搜的最大特点在于，把哼唱搜索与录音搜索分离开来。即哼唱搜索与录音搜索采用了不同的引擎。从检索结果来看，未能检出《费加罗的婚礼》与《双节棍》。总体合格。

4.2.4 Shazam

与 Track ID 相类似，Shazam 的官方说明里也阐明并不支持哼唱输入，然而其在哼唱部分的实际检索结果尚可，也具有一定的实用性。同时 Shazam 也是所有系统中检索速度最快的。

4.3 系统比较

4.3.1 结构化音乐检索系统比较

四个结构化音乐检索系统的主要特征比较如下：

		Folk Tune Finder	Musipeida	Songtapper	Themefinder
匹配方式	模糊	○	○	○	
	精确				○
数据采集方式	自建数据库			○	○
	网页抓取	○	○		
	用户添加	○	○	○	○
数据库规模（首）		>200000	>30000	88438	35000

输入接口类型	钢琴键盘	○	○		
	键盘敲击		○	○	
	文本	○	○	○	○
	哼唱		○		
检索功能	文本	○	○	○	○
	旋律	○	○		○
	节奏		○	○	

上表显示了四个系统在匹配方式、数据采集方式、输入接口类型、检索功能等方面的相同与不同之处。需要说明的是，上表只能反映出各结构化音乐检索系统的主要特征，每个系统更为细节的设计在 4.1 中有介绍。

四个系统共同存在的主要不足在于无法保留检索提问式和提供二次检索功能，不便于用户在原有基础上改进检索或进行缩减。此外，系统也不提供简谱输入，对于那些只识简谱、不识钢琴键盘和五线谱的用户十分不利。另外，有的系统输入模块中提供的“清除提问式”功能会一次性地将用户此前输入的所有音符清除，而不是逐个清除，用户一旦输入有误则要从头开始重新输入，这给用户造成了极大的不便，降低了系统对用户的友好性。

4.3.2 基于样本的音乐检索系统比较

四个基于样本的音乐检索系统的主要特征比较如下：

		Midomi	Track ID	爱乐搜	Shazam
数据采集方式	自建数据库	○	○(GraceNote)	○	○
	网页抓取			○	○
	用户添加	○			
数据库规模（首）		未知	>1000000000	未知	>10000000
输入接口	文本	○		○	
	哼唱	○		○	
	录音	○	○	○	○
	语音			○	
检索结果	试听链接	○	○	○	○
	购买链接	○	○		○
	图文信息	○	○	○	○

上表体现了四个检索系统在数据采集方式、数据库规模、输入接口与检索结果上的不同。值得一提的是，Midomi 中的不同用户上传的同一首歌，由于是不同的人演唱，因而在数据库中也算作两首，所以 Midomi 的数据库总体曲目数量与后三者不具有可比性。

Track ID 的 GraceNote 数据库是最为全面的。在接口上，对录音的支持均较好，而哼唱方面水平参差不齐。Track ID 和 Shazam 都是只支持录音的系统。而爱乐搜有一语音搜索的特色。另外，midomi 公司的另一产品：hound。就是一款只支持语音搜索的系统。

检索结果方面基本水平一致。虽然爱乐搜未能提供购买链接，但多数歌曲都提供了完整的免费下载链接。

4.3.2 两类基于内容的音乐检索系统比较

显然，接口上的不同是两类系统的最大不同。另一方面，数据库规模上的差距也比较明显。究其原因，商业化与实用化更加突出的基于样本的检索系统，无论在商业上获得投入，还是用户的支持上都比较领先，因而数据库相应的也较大。

其次，对于古典音乐，结构化的系统就更为适合，检索结果也更加令人满意，而新兴的基于样本的音乐检索系统则更适合自唱片业发达以来产生的各种现代音乐。

此外，结构化音乐检索系统通常要求用户具有一定程度的专业知识，如能够识别五线谱、会使用钢琴键盘等；而基于样本的音乐检索系统一般不需要用户具备专业知识，用户只需通过实时哼唱或上传录音输入旋律即可。

5 基于调查结果的改进模型

5.1 基于现有系统功能的改进方案

5.1.1 结合两种音乐系统

音乐类型的不同，决定了两种音乐检索系统在检索不同的音乐上各具优劣。比如一首莫扎特的《费加罗的婚礼》，无数的交响乐团都演奏过，也出过唱片，在各大数据库上都存在。各个版本区别较小，难以区分，几乎无法使用基于样本的检索系统，即使是 Track ID，也会在每次检索时，返回不同交响乐团演奏的同一首歌曲。其他三类样本检索系统正是出于这样的考虑，才没有收录更多的交响乐。然而在这一方面，结构化的搜索引擎就有较多的优势，非常便于用户进行检索，但其接口却更为专业化，一般用户难以使用。

若能把二者结合起来，则会有更好的结果。一方面利用基于样本的音乐检索系统的技术，使得用户通过哼唱或演唱输入一段样本，而后系统根据样本，分析和转变成曲谱式的旋律，而后与结构化的曲谱类数据库进行匹配分析，再将结果返回给用户。这样一种模式就非常适合古典音乐。

5.1.2 结果反馈

多数系统都不具备反馈结果的功能，这对于系统的改进比较不利。Midomi 在这方面做得较好，由于结合了云思想，Midomi 在用户未能检索到相应音乐时，会引导用户上传自己的录音，并为其标明，这样就为以后的用户提供了方便。

5.1.3 情感语义搜索

无论是哪种系统，哪类平台，都对音乐中旋律的强弱，高低起伏所产生的情感因素难以提取出可检特征。但对于音乐艺术来说，情感语义是其灵魂所在，不能提供这方面的可检特征就不能是终极的音乐搜索引擎，在这方面还是有探索和改进的空间。

5.2 可以尝试的新功能

5.2.1 上传文件检索功能

应当允许用户上传一些信息不全的音乐片段，可能是 MIDI、可能是 MP3 等格式。通过这种接口，可以更方便的为用户提供曲目识别服务。音乐检索系统应当不只局限于提供最终的乐曲，还应当提供识别、添加标签等功能。调查发现，苹果公司的 iTunes 软件已经提供了一种 iTunes Match 服务，可以为用户识别其乐曲库中标题作者等信息不完整的乐曲，并为其完善信息。这一功能是目前其他系统所不具备的，也是未来检索系统应当提供的一种服务。

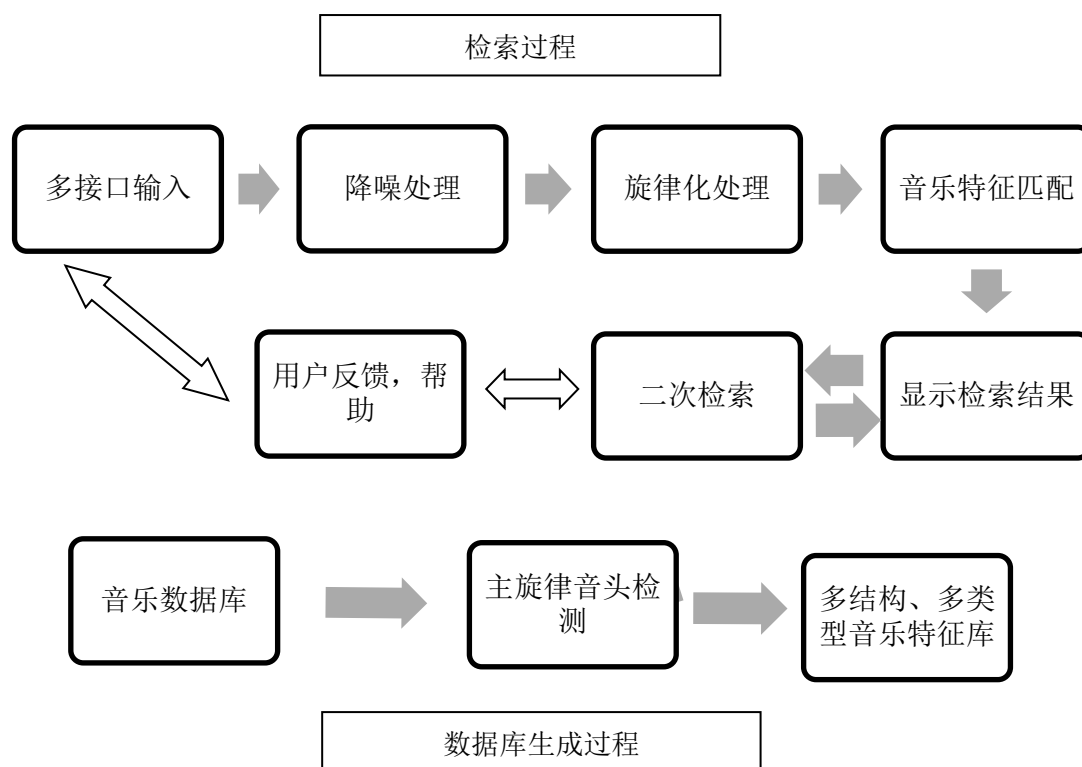
5.2.2 检索帮助服务，

这对于专业的检索系统更有用一些。提供更完善的帮助文档是必要的，会使用户的检索能力更强、检索结果更令人满意。另一方面，提供完善的检索求助服务也是非常重要的，主要分为：交互检索平台和实时在线客服服务两类。交互式检索平台是指用户可以留言给工作人员，用文本的方式提出自己的问题，工作人员看到后再进行解答，并返回相应的结果。而实时在线客服服务平台是指用户可以与工作人员实时的进行一对一的交流。工作人员直接解答用户的使用问题。

5.2.3 搜索结果过滤改进

对于模糊匹配的检索系统，在检索结果中搜索等更深入的检索方式可以带来更好的检索结果。同时应当为用户提供一些过滤功能，这会使得检索结果更完善。

5.3 改进系统的模型结构



新的模型有如下特点：

- 1、多接口输入、一站式的搜索服务。无论是专业的钢琴键盘输入、还是哼唱、录音输

入乃至上传文件输入，都一站式处理。

2、多结构、多类型的音乐特征库。对于不同类型的音乐采用不同结构的特征库。比如古典音乐不区分版本，摇滚音乐更注重节奏等。

3、提供完善的用户反馈和帮助功能。用户在检索过程中可以得到应有的帮助使检索能力和检索效率得到提高。

4、提供完善的二次检索和过滤的功能。这使得用户可以不断精简使得搜索结果逐步变为所求。

6 相关问题讨论

6.1 数据来源

音乐资源的选择与采集是音乐检索系统的起点和基础。一个设计再好的系统，如果收录的乐曲范围不足，也难以满足用户的需求。自建数据库的数据采集方式成本较高，因此采用网页抓取和用户添加相结合的方式是比较合适的选择。不过，用户添加的音乐资源可能在著录、格式等方面存在不规范的问题，应当有专门的系统维护人员对用户添加的音乐资源进行规范化处理。

6.2 匹配精度与系统评价

目前大部分基于内容的在线音乐检索系统采用的都是模糊匹配方式，只有少数如 Themefinder 等系统采用的是精确匹配方式。模糊匹配方式的容错性较好，并且可以提高检全率，尤其适合那些希望检索到若干首具有类似旋律轮廓的乐曲的用户（例如正在寻找灵感的作曲家）。模糊匹配方式的不足在于检准率较低，对于那些希望检索出特定目标曲目的用户来说有效性不足。模糊匹配方式也不利于检全率、检准率等指标的衡量，从而给系统评价带来很大困难。如何建立一套适用于基于内容的音乐检索系统的评价指标体系，是有待进一步研究的课题。

6.3 扩展能力

随着用户需求的不断提高，“一站式获取”势在必行。在线音乐检索系统除了显示结果乐曲外，还应该为用户提供乐曲的在线收听、下载、分享、歌词链接等一系列服务。这样扩展用户一次使用检索所能得到的信息，这样检索的价值就会更高，检索的效果就会更好。

6.4 检索平台延伸

从 SONY 的 Walkman 到苹果的 iPod，音乐已经无处不在，不能仅仅局限于计算机上了。手机、各种移动音乐平台，都应当有相应的检索功能，检索平台也应当不仅停留于联网检索，更应当有本地检索。比如应当在 iTunes 当中提供一种哼唱搜索本地歌曲的功能。

从目前的形势来看，手机上已经有了数款基于哼唱的检索系统，便于用户使用，这是未来发展的大趋势，同时也是音乐检索新的商业价值开发所在。

另一方面，本地化的检索功能也是值得开发和延伸的检索平台功能。

6.5 未来发展趋势

尽管目前基于内容的在线音乐检索系统还处于利用声学特征级音频特性的层面,但是未来的音乐检索会越来越关注音乐最本质的情感、叙事等语义特征,第三代音乐检索系统——基于语义的音乐检索系统具有广阔的发展和应用前景,也是一个令人振奋的有待研究的课题。

7 结语

本次调研选取了 8 个基于内容的在线音乐检索系统: Folk Tune Finder、Musipedia、Songtapper、Themefinder、Midomi、track id、爱乐搜和 shazam。其中,前四个系统是结构化音乐检索系统,后四个系统是基于样本的音乐检索系统。我们在对每个系统进行较为全面的检索测试的基础上,指出各系统的特色和不足,并对每类中的不同系统和两类系统进行比较。接着我们根据调研情况,并结合该领域最新的理论和技术,提出了一个基于内容的音乐检索系统的改进模型。

基于内容的音乐检索与传统的文本音乐检索相比,能更好地满足用户检索音乐的需要,是音乐检索的一个必然的发展趋势。近年来出现了不少利用声学特征级音频特性的基于内容的音乐检索系统,其中不乏性能较为出色的代表性系统。基于语义的音乐检索也得到越来越多的研究,不少研究者已经提出了一些相关的技术、模型、结构及实现方案,但是目前尚不存在面向广大互联网用户开放的正式投入运行的基于语义的音乐检索系统,这一具有广阔发展前景的课题还需要广大研究者持续不断的深入研究。总的来说,基于内容的音乐检索目前还是一个比较新的研究领域,其检索对象和范围具有较高的复杂性和多样性,在该领域的研究还只是刚刚开始,可谓任重而道远。

通过本次具体的调研实践,我们对基于内容的音乐检索理论和技术现状、发展前景与趋势以及业已实现的系统有了更为全面和深入的了解和掌握,获益匪浅。然而,由于时间和精力有限,本次调研也存在不足之处,如对各个系统的具体技术与算法不够明确,也未能建立一套基于内容的音乐检索系统的评价指标体系。这是需要我们进一步改进、补充与完善的地方。

附录

附录 1 基于内容的音乐检索系统技术简介

1.1 音高检测

第一类方法是时域方法。

音高检测的最直接方法是零点交叉法,如给定一个正弦音,只需要计算信号过零点的频率即可,这种方法实现简单效率高,然而噪声敏感性比较大,所以难以处理较为复杂的声音信号,不适合做音乐信号的提取方法

自相关方法(auto-correlation)是另一种时域方法,主要思想基于信号延迟后于原信号的相似程度。

与自相关方法相似的另一类方法是平均幅度差函数法(average magnitude difference function AMDF),与自相关方法计算信号分段与其位移新号段的积不同的是,AMDF 计算的是差值。

第二类方法是频域方法。

傅里叶变化是最为熟知的频域方法，它允许将一个函数分解成数个周期震荡函数序列，尤其傅里叶级数可以让任意周期函数表示为不同频率、幅度的正弦函数之和，且这些正弦函数满足一个简单的关系：其频率为周期函数基本频率的整数倍，换句话说，这些正弦函数是调和的。在数字信号处理领域，处理的离散变化，其对应的工具就是离散傅里叶变换。在工程实践中，则运用快速傅里叶变换较多。

频域方法中还包括 Q 变换和 Hartley 变换。

1.2 位置敏感哈希函数（LSH）

位置敏感哈希函数，即 Location Sensitive Hash，LSH⁴，是一种近似紧邻查找技术，与普通哈希函数不同的是，它具有位置敏感性，即空间相邻的点在散列后，其相邻性具有任意给定的概率保证。LSH 广泛应用于高维空间的近邻查询，特别在图像、视频、音频和文本等含有丰富特征信息的领域中的应用越来越多。

1.1.3 动态时间规整（DTW⁵）

动态时间规整是用以在两个时间序列之间找到最优对齐的经典算法，直观来看，序列可以以非线性的方式曲折对齐。最初，DTW 在语音自动识别中用以比较两种不同的语音模式，随后被应用到其他领域，如数据挖掘和信息检索领域中用以自动处理依时数据中时间维度上的性比纳和速度不一带来的匹配问题。

1.4 音乐的存储格式及其表示

音乐有其相应的存储格式，不同的格式对应着不同的处理技术。

1.4.1 乐谱表示

乐谱就是五线谱技术，以五线谱的形式对音乐进行存储。可以记录包括音高、节奏在内的信息，MusicXML⁶就是其中一种格式。已经得到了广泛的应用，他用文本的方式表示印刷本曲谱中的、音符、节拍、五线谱等，下面的一段代码，就代表了贝多芬命运交响曲中的中央 E。

```
<note>
  <pitch>
    <step>E</step>
    <alter>-1</alter>
    <octave>4</octave>
  </pitch>
  <duration>2</duration>
  <type>half</type>
</note>
```



⁴A. Andoni and P. Indyk Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In 47th Annual IEEE Symposium on Foundation of Computer Science(FOCS'06) 2006, 459-468.

⁵L R. Rabiner and B, HJuang, Fundamentals of Speech Recognition, Prentice Hall Signal Process Series, 1993.

⁶MusicXML, A universal translator for common western musical notation.Recordare, <http://www.recordare.com/xml.html>, 2006.

1.4.2 MIDI 表示

MIDI 是 Musical Instrument Digital Interface 的缩写，它是乐器数字化接口标准，创建于上世纪 80 年代，主要由三部分组成：1、消息规范；2、传输规范；3、文件格式规范。

1.4.3 波形表示

波形是从本质上表示声音的一种方法。声音是由频率来表示音调的，振幅来表示音量的。最常见的波形格式是 WAV。

WAV 是微软和 IBM 联合开发的用于音频数字存储的标准，符合 RIFF 文件规范，用于保存 Windows 平台的信息资源，被 Windows 平台广泛支持 44.1K 采样率，16 位编码，和 CD 格式完全一样，可以说质量相差无几。

波形表示较之于前两种表示，其主要缺点在于存储空间占用大，以及目前单个 WAV 文件的 2G 大小限制。

附录 2 小组分工

调研过程分工：

文献收集与整理：周妍、汪丹华

系统收集与选取：周妍、刘垚远

结构化音乐检索系统调研：周妍

基于样本的音乐检索系统调研：刘垚远

改进模型的提出：刘垚远从技术角度提出改进思路，周妍从用户需求与用户体验角度提出改进思路

报告撰写分工：

导言、结语两部分由周妍和刘垚远共同完成；研究方法和步骤、相关问题讨论两部分主要由周妍完成；基于调查结果的改进模型、附录 1 两部分主要由刘垚远完成；系统选取与介绍、系统调研情况两部分由周妍和刘垚远共同完成，其中周妍主要负责结构化音乐检索系统的相关部分，刘垚远主要负责基于样本的音乐检索系统的相关部分；周妍和刘垚远分别对对方完成的部分进行修改和补充。

附录 3 参考文献推荐

第一篇

《Content-Based Music Information Retrieval: Current Directions and Future Challenges》

文献基本信息

作者：Micheal A. Casey, RemcoVelkamp, MasatakeGoto, Marc Leman, Christophe Rhodes, and Malcolm Slaney.

来源出版物：Proceedings of the IEEE

卷 96 页 0018-9219 DOI: 10.1109/JPROC.2008.916370 出版年：2008 年 3 月

推荐理由

本文详细的介绍了目前基于内容的音乐检索领域的相关情况，不仅对当前的技术做了总结，并且对未来基于内容的音乐发展方向给予了指引。同时文章还分析了未来的挑战和发展前景。

本文刊载在 *Proceedings of IEEE* 上，被引次数 32 次，IEEE 是美国电气和电子工程师协会，是世界上最大的专业技术组织之一，到 2005 年为止拥有 36 万会员。本文能在 *Proceedings of IEEE* 上发表，说明具有较强的前瞻性和前沿性，同时具有非常好的指导作用。

第二篇

《Content-Based Retrieval of Music and Audio》

文献基本信息

作者: Foote, JT (Foote, JT)

编者: Kuo, CCJ; Chang, SF; Gudivada, VN

来源出版物: MULTIMEDIA STORAGE AND ARCHIVING SYSTEMS II

丛书: PROCEEDINGS OF THE SOCIETY OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS (SPIE)

卷: 3229 页: 138-147 DOI: 10.1117/12.290336 出版年: 1997

推荐理由

本文被引次数 98 次（在 wos 上），从技术角度详细分析了音乐乃至音频系统 CBR 技术的根本。无论是从算法还是思想、技术还是理论角度上都写的非常清晰合理，读者通过这篇文章，可以全面的了解如隐马尔科夫链、动态时间规整等理论。

读者通过读这篇文章，可以就基本技术有一定的了解，对想要开发相应系统，或者想要对相应的技术理论等有所了解的读者，非常有帮助。

参考文献

- [1] 孟宪巍. 基于内容的音乐信息检索研究与系统实现[D]. 北京邮电大学, 2009.
- [2] 赵静. 基于内容的音乐检索系统研究[D]. 西北大学, 2005.
- [3] 郭贺平. 基于内容的音乐检索的算法研究及系统实现[D]. 东北师范大学, 2007.
- [4] 魏文华. 哼唱式音乐检索系统 [D]. 华中科技大学, 2007.
- [5] 支林. 基于哼唱的歌曲检索系统的研究与设计[D]. 电子科技大学, 2011.
- [6] 李聪. 基于哼唱的音乐检索系统[D]. 清华大学, 2009.
- [7] 吴长君. 基于旋律的音乐检索系统设计与实现[D]. 北京邮电大学, 2011.
- [8] 李江涛. 基于旋律特征提取的哼唱式音乐检索[D]. 哈尔滨工业大学, 2009.
- [9] 张静朱悦心. 采用人声输入的网络音乐检索系统[J]. 微电子学与计算机, 2006 年第 5 期.
- [10] 夏小亮周明全王学松李鹏李娟. 基于分句匹配的哼唱音乐检索系统的算法研究与实现[J]. 北京师范大学学报（自然科学版）, 2011 年第 4 期.
- [11] 鞠源. 基于哼唱的音乐检索系统的研究与探索[J]. 情报杂志, 2010 年 6 月.
- [12] 王长宇张蕊菽刘蔚王启贤罗颖琳. 基于内容的结构化音乐信息检索系统调查[J]. 情报探索, 2010 年第 8 期.
- [13] 薛锋杨宗英郑巧英黄敏. 基于内容的音乐检索[J]. 大学图书馆学报, 1999 年第 4 期.
- [14] 林小兰王晓光王晖. 基于内容的音乐检索关键技术研究[J]. 中国传媒大学学报（自然科

学版), 2010 年第 4 期.

[15] 马希荣梁景莲. 基于情感音乐模板的音乐检索系统研究[J]. 计算机科学, 2009 年第 1 期.

[16] 王小凤耿国华刘晓宁孙霞. 基于相对特征的音乐哼唱多句检索算法[J]. 计算机应用研究, 2011 年第 3 期.

[17] 张宝华张品. 基于旋律的音乐检索系统[J]. 电声技术, 2005 年第 12 期.

[18] 焦玉英周华敏. 基于音频内容的交互渐进式音乐检索系统的设计[J]. 情报科学, 2004 年第 12 期.

[19] 傅鹏孙坚付春雷. 基于语义的音乐检索系统[J]. 重庆理工大学学报(自然科学), 2011 年第 1 期.

[20] 黄磊冯寅. 简析基于内容的音乐检索[J]. 福建电脑, 2008 年第 2 期.

[21] 鲁帆王民. 一个基于哼唱的音乐检索系统[J]. 中国西部科技, 2011 年第 4 期.

[22] 徐霁袁旭海王让定. 一个音乐检索系统的研究与实现[J]. 宁波大学学报(理工版), 2007 年第 3 期.

[23] 陈晓颖高峰陆汝占. 一种新型的音乐信息检索索引方法及其应用[J]. 计算机工程与应用, 2007 年第 14 期.

[24] 王昉. 音乐检索现状及发展趋势研究[J]. 科技广场, 2008 年第 3 期.

[25] 赵丹群编著. 现代信息检索: 原理、技术与方法[M]. 北京: 北京大学出版社, 2008.

[26] Remco C. Veltkamp, Frans Wiering, Rainer Typke. Content Based Music Retrieval. Encyclopedia of Multimedia, Springer, 2006.

[27] Michael A. Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, Malcolm Slaney. Content-Based Music Information Retrieval: Current Directions and Future Challenges. Proceedings of the IEEE, 2008.

[28] Bokyoung Sung, Kwanghyo Koo, Jungsoo Kim, Myung-Bum Jung, Jinman Kwon, Ilju Ko. Effective Digital Music Retrieval System through Content-based Features.

<http://www.waset.ac.nz/journals/waset/v50/v50-122.pdf>.

[29] Foote, J.T. Content-Based Retrieval of Music and Audio. PROCEEDINGS OF THE SOCIETY OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS, 1997.