维普资讯 http://www.cqvip.com

音频信息检索综述

王清亮 常 青 薛向阳

(复旦大学计算机科学与工程系,复旦大学智能信息处理开放实验室 上海200433)

摘 要 随着多媒体和Internet技术的广泛应用和深入普及,多媒体数据的数量正在飞速增长,音频数据作为多媒体数据的重要组成部分,其信息量也在不断膨胀。如何有效的对音频信息进行检索成为现代信息检索的一个重要的研究领域。然而,音频数据同传统的文本数据不同,文本检索中的相关技术不能简单地应用到音频信息检索中,因为后者往往是基于语义的,所以不可避免地要用到音频数据的特征提取和模式匹配等技术。本文将对音频数据检索相关技术和系统做一综述性的介绍。

关键词 音频信息检索,音频分类,语音识别,音乐检索

Audio Information Retrieval: an Overview

WANG Qing-Liang CHANG Qing XUE Xiang-Yang

(Dept. of Computer Science and Engineering, Fudan University, Laboratory of Intelligent Information Processing, Fudan University, Shanghai 200433)

Abstract With the rapid development and widely deployment of multimedia and Internet Technology, there has been a tremendous increase in the amount of multimedia data. This has caused the information contained in audio, which is an important part of multimedia data, to greatly expand. Audio information retrieval hence becomes a significant research field of modern information retrieval. Unlike text retrieval, however, audio information retrieval is usually semantic oriented and such technologies as feature extraction and pattern matching will be necessarily used. This article presents an overview on these technologies as well as some audio retrieval systems.

Keywords Audio information retrieval, Audio classification, Speech recognition, Music retrieval

1 引言

信息检索主要涉及信息的表示,存储,组织,和访问等[1]。传统的信息检索大多基于文本,根据用户提交的一系列关键词,找到与其相关的文本文档,由于结果文档与查询关键词的相关程度不同,还需要根据这种相关程度对结果文档进行排序,并返回这些文档的位置信息。Internet上许多著名的搜索引擎,如google(www.google.com),lycos(www.lycos.com),altavista(www.altavista.com)等都是典型的基于文本的检索系统。

音频信息检索是信息检索的一个重要分支,顾名思义,就是从众多的音频数据中定位及提取用户比较感兴趣的信息。与文本检索不同,用户的信息需求(information need)一般难以以关键词的形式提交,因为从原始数据中抽取检索项的方法并不适用于音频数据这种数字信号。在后面的论述中可以看到,用户提交的检索要求往往是对于音频数据的语义描述,从这种角度来说,音频信息检索是一种偏重于内容与语义的检索。

音频是最重要的多媒体数据之一,随着网络技术的普及,也是数量增长最快的一种数据形式,音频信息检索的重要性不言而喻,结合不同的音频处理技术,其在不同领域里的应用也愈加广泛。

音频分类是将音频数据按照不同的需求分类。音频分类对于构建音频知识系统将起到非常基础和重要的作用;同时,在非纯净音频环境中,音频分类也是进行各项特定音频处理的必要步骤。

语音识别与理解是最具有实用意义的音频处理技术,目前非特定人连续语音识别系统已经达到了实用的阶段,语音识别的技术已经相对稳定与成熟,语音的处理正转向语言的理解这一语义研究领域。应用语音识别与理解技术,计算机将会更加智能,随着语言理解研究的不断深入,语音将在人机交互中扮演着越来越重要的地位。

与语音相同,音乐也是一种非常常见的音频数据。然而受限于对人类听觉系统的认识以及音乐数据相对复杂,人们对音乐检索的研究还比较初步,虽然一些系统展示了一定的音乐检索功能,但到达实用的程度尚有距离。

另外,音频信息检索在视频处理中也起到一定的作用,比如在视频的场景分割中,以图像为基础的镜头和镜头组检测并不能很好地确定场景的边界,这时候音频信息就起到非常重要的辅助判别的作用。

本文将按照用户信息需求的不同介绍和评价各种音频信息检索方法,由于基于内容与语义的检索更加具有实用价值,因而本文将详细介绍。同其它数字信号的处理相同,特征的选择与提取在音频信息检索中起着重要的作用。由于音频信号的特征计算是音频信号处理的主要内容,本文将不作介绍,而是着重于特征的选取与使用。

2 音频信息检索方法

音频信息检索系统的框架设计和音频数据的特征选取很大程度上取决于用户的查询需求[1],按照从原始数据匹配到语义查询的顺序[3],用户的查询需求可以有下述几种形式。

2.1 原始音频数据(Raw Audio Data)匹配

用户提交原始的音频数据,要求查询与此音频数据完全相同的数据,类似于在文本中精确查找字符串。尽管由于不涉及到音频数据的特征提取,所以这种方法实现起来非常简单,但对于语音数据这样一种模拟信号的数字表示进行简单的匹配与查找并不能取得好的查询结果,因此并不实用。

2.2 基于文本摘要的查询

在原始数据匹配之上是基于文本摘要的查询。用户提交的是所感兴趣的音频的文本摘要,要求查询的音频数据与此

维普资讯 http://www.cqvip.com

具有相同的摘要，同时不需要考虑音频数据的物理形式，如文件格式、采样率、量化精度以及是否压缩等等。这种方式的检索系统要求所有音频数据都被标注许多关于其所表述信息的文本摘要，不同类型的音频，摘要的形式可以不同。对于话音，这种摘要可以是说话人信息，语音内容，说话的场所和时间等；对于音乐，可以是音乐流派（古典、流行、摇滚、乡村…），作曲家，歌手，专辑名，音乐标题以及歌词（对于歌曲而言）等。一般而言，附加的标注信息越丰富，就越容易得到满意的查询结果。

基于文本摘要的查询实际是一种文本检索，所以传统的文本检索的技术，如文本文档的预处理，包括词干提取，禁用词去除和词典建立等；文本的压缩；文本的建立索引以及字符串的查找等等[1]就是这种音频信息检索系统的关键。

对于音频信息提取摘要和进行标注，可以人工进行，也可以是计算机自动进行。虽然人工标注的信息比较准确和全面，但存在一些缺点，首先，对于一些主观上的信息，不同人可能会得到不同的结论，因而会存在一定的偏差[5]。另外，人工标注要花费大量人工和时间，由于音频具有线性的特点，所以在对未知的音频数据进行标注之前，要将音频内容完整地听一遍，这一般要花费与音频同样长度的时间，考虑到音频数据正在不断增多，所要花费的人力将非常巨大。现在许多研究致力于如何用计算机自动对音频数据进行分析和分类，从中抽取用户可能感兴趣的信息，这就是下一部分要介绍的基于语义的一种方法。

在文[4,5]中介绍了一种计算机辅助的对音频数据进行标注的方法，首先要对音频数据进行特征计算、分段，根据得到的结果对音频进行简单的标注，如音乐中主要的成分（人声、钢琴、管弦乐器…），或是声学特征（柔和、低沉、急速…）等，由于这种这些标注表示的信息非常有限，而且标注的往往包含错误，相对于人工标注还有许多的不足之处。

由于基于语义的音频检索系统还未到达大规模实用的阶段，基于文本摘要的音频检索系统仍然是目前比较常用的音频信息检索方式。

2.3 基于内容和语义的检索

虽然有以上两种检索方式，但用户最关心，最具有实用价值的是基于内容的检索。在这种检索方式下，用户提交的是关于音频数据内容的一种语义上的描述。如在新闻广播的语音检索中，用户可以查询和特定事件（如伊拉克战争等）或特定人物（如萨达姆、布什等）相关的新闻播音；在音乐检索中，用户可以检索具有某种声学感知特征（如忧伤、嘈杂等），某种流派（如古典、摇滚等）音乐；用户还可以提交一段音频样本，如用户哼唱自己喜欢的某个重复的旋律[6,22]，查找包含这段旋律的音乐，由于不同用户对于同一段的旋律的哼唱是不同的，所以这种方式的检索显然不是2.1节中提到的原始音频数据查询。

下面将从音频分类、语音识别、音乐检索等几个方面对基于内容和语义的音频信息检索进行介绍，并列举了其中比较著名的系统（原型）。

2.3.1 *音频分类* 产生于音频数据不断增加和用户对于音频信息检索的需要，是将音频数据按照其感知上的特征或表述的内容等分为不同的类别。音频分类技术不仅有助于音频信息的检索[3,8]，而且在视频结构分析中也扮演着重要的角色[2,9~11]，同时音频分类也往往作为特定类型音频操作，如语音识别，音乐检索等的预处理。

早期音频分类研究比较著名的是 Muscle Fish 系统（http://www.musclefish.com/）[3]。Muscle Fish 是一个典型的以音频特征向量为基础对音频文件进行模式聚类和识别的系统。音频文件按照声学感知上的特征被事先分为许多类，如笑声、刺耳的声音等。音频文件特征向量的计算考虑到了音频信号的声学特征，包括响度，基音频率，亮度和带宽等，同时，由于这些特征都是短时性的，所以实际选取了这些特征在文件时间长度分布上的三个统计参数：均值，方差和自相关度作为特征向量的组成部分。

与 Muscle Fish 系统不同的，另外一些音频分类的研究侧重于音频的内容，从而将音频分为语音，音乐，环境音和静音等类型。在文[12]中，Pfeiffer 等人介绍了一种利用音频感知特征（响度、频率等）进行音频分类的理论框架和实际应用。在文[13]中，Saunders 提出了一种基于简单音频特征（如过零率和短时能量等）分类器，可以对广播进行语音与音乐的区分，当音频信号的窗口取得很足够长（2.4s）的时候，准确率可以达到98%。在文[14]中，Scheirer 等人为音频分类引入了更多特征（如4Hz modulation energy，percentage of "Low-Energy" frames，spectral rolloff point，spectral centroid 等），并被应用到不同的分类模型中（包括 GMM，BP-ANN 和 KNN），试验结果也很好（错误率为1.4%），窗口长度仍然为2.4s。

文[7]总结了上述研究的不足之处，如窗口长度过长，不利于检测的精度；能分别出的种类也比较少，基本都没有把环境音的因素考虑在内等，提出了一种改进的分类方法，图1描述了该分类系统用到的特征，分类流程和对进行音频分类后的结果。

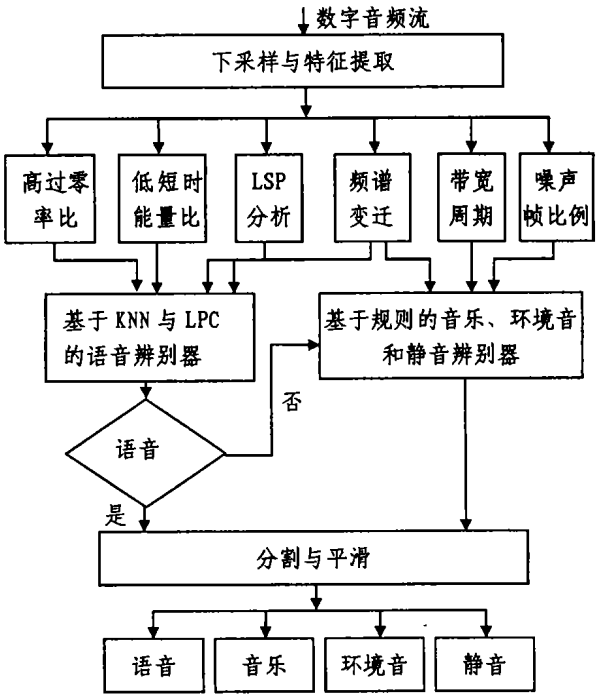

图1 音频分类基本流程

在特征选取上，该系统不仅使用音频信号分析中两个简单但又非常重要的特征：过零率和短时能量，而是考虑了这两个参数的分布情况，提出了高过零率比例（HZCRR）和低短时能量比例（LSTER）两个衍生特征。HZCRR 表示过零率高于平均值1.5倍的帧的比例，LSTER 表示短时能量值小于0.5倍平均值的帧的比例。

实际的分类操作分为两步，首先使用 KNN 分类器和 LPC 分析的方法将音频分为语音和非语音，对于非语音进一步分为静音、环境音和音乐，其中静音首先利用过零率和短时能量被检测出来，环境音和音乐可以从它们在带宽周期，频谱

维普资讯 http://www.cqvip.com

变迁(Spectrum Flux)和噪声帧比率三个参数的分布特点中进一步区分出来。

为了降低分类过程中的偶然错误,分类系统还采用了平滑技术,即忽略连续相同类型音频段中出现的某一段其他类型音频,将这一段也纠正为同一类型。

由于特征选取比较丰富,分类步骤比较细致,同时对结果进行了进一步的平滑,该分类系统的对上述四类音频的分类结果可以达到96.51%,考虑到可以识别的音频种类和分类精度,这是目前为止取得最好的分类效果。

应当指出,尽管目前的音频分类研究取得了很好的结果,但试验数据仍然过于纯净,即音频数据中很少或者没有各种成分混叠的情况,如在噪音比较强的背景下的音乐和语音等。因而对于真实环境的音频数据,最终的分类结果将会受到一定的影响,在这种情况下,可以考虑使用各种音频处理技术,如降噪,语音和音乐的增强等方法来增加分类系统的适应性和可扩展性,从而在实际应用中取得更好的分类效果。

2.3.2 *语音识别* 语音在音频数据占据了非常大的比重,因此对于语音的处理理所当然地成为音频处理的一个重要研究领域。将语音识别成文本,再利用文本检索技术即可以初步实现应用对语音内容的检索。

相对于其他音频处理领域,如2.3.3节中介绍的音乐分析等,语音识别的研究经历了更长的时间,技术发展也相对成熟,已经达到实用的阶段,如IBM公司推出的ViaVoice就是一个基于大词汇量,非特定人的连续语音识别系统[15]。

语音识别技术涉及很多领域,包括发声机理和听觉机理、信号处理、模式识别、概率论和信息论、人工智能等等。本部分将只对语音识别做一系统性的介绍,图2是语音识别系统的基本框架。

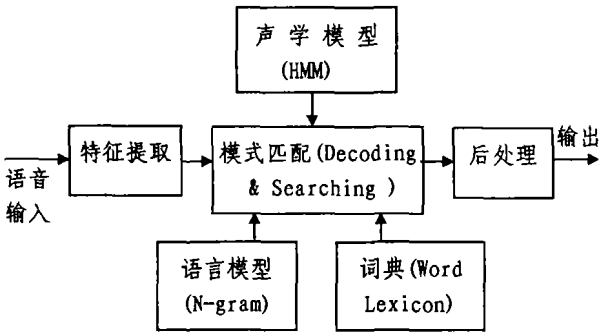

图2 语音识别系统基本框架

语音信号特征的提取是一个信息大幅度压缩的过程,在进行特征分析的时候,信号的窗长比较短,一般为20～25ms[16],常用的特征有LPC系数、MFCC等。

声学模型主要表征语音到音节概率的计算,基本上所有的语音识别系统都采用隐马尔可夫模型(HMM)[17]作为声学模型。马尔可夫模型是一个离散时域的有限状态自动机,隐马尔可夫模型是指这一马尔可夫模型的内部状态外界不可见,外界只能看到各个时刻的输出值。对于语音识别来说,一个HMM就是一个语音单位的统计上的表示,这个语音单位可以是音素,也可以是一个字,多个音素的HMM可以串行起来构成一个字的HMM,而整个连续语音识别的模型简单的所就是词和静音组合起来的HMM。

构词法(Lexicon)主要处理如何将识别出来的音节进一步组合成为单词(字),在有些系统中,这一部分是省略的,语音被直接识别为单词(字)。

由于从语音到单词(字)的识别是基于概率模型的,所以一段语音可能就应与多个候选单词(字),为了增加识别的精度,就需要考虑不同单词(字)之间的组合概率,由此引入了语音模型的概念。

语言模型就是根据已经出现的字/词来确定一个字/词出现在下一位置的概率。语言模型可以分为基于规则的模型和基于统计的模型。统计模型主要通过概率统计的方法揭示语言单位内在的统计规律,是主要采用语言模型,N-Gram[18]就是其中的一种。N-Gram假定一个词的出现至于其前面N－1个词相关,而与其他任何词不相关,实际使用中考虑到计算的复杂度,N取值都不大,常为2(Bi-Gram)或3(Tri-Gram)。由于对于有些情况,N值固定取得的效果并不十分理想,因此文[19～21]等对于N取变化值的语言模型进行了研究。

上述模型是对声音以及语言在统计上的一种静态的描述,在语音识别时,这些信息被用来计算出最可能的单词序列,这一过程就是解码与查找。解码器一般使用Viterbi算法搜索可能的候选路径,并选择最佳路径作为识别结果。

后处理用来拒绝可能的误识和非词表词,进一步提高整个系统的性能。图3显示了一些情况下语音识别的错误率。

CORPUS	TYPE	VOCABULARY SIZE	WORD ERROR RATE
Connected Digit Strings--TI Database	Spontaneous	11 (zero-nine, oh)	0.3%
Connected Digit Strings--Mall Recordings	Spontaneous	11 (zero-nine, oh)	2.0%
Connected Digits Strings--HMIHY	Conversational	11 (zero-nine, oh)	5.0%
RM (Resource Management)	Read Speech	1000	2.0%
ATIS(Airline Travel Information System)	Spontaneous	2500	2.5%
NAB (North American Business)	Read Text	64,000	6.6%
Broadcast News	News Show	210,000	13-17%
Switchboard	Conversational Telephone	45,000	25-29%
Call Home	Conversational Telephone	28,000	40%

图3 各种情况下语音识别系统的错误率

可以发现,尽管在一些比较特定的条件下(如背景噪音很低,说话人语音很清晰、标准等),语音识别系统可以得到很好的结果,如NAB,但对于大词汇量,自然情况下的语音识别,其效果还不是十分理想,如最后三种应用。因此在提高准确率和适用性方面,语音识别还有待于进一步的提高。

由于语音识别技术发展相对成熟,目前语音方面的研究更多的致力于自然语言的理解。然而同任何人工智能领域一样,人们对于语义的研究仍然进展缓慢。但随着关键技术的突破,人们必将能使计算机理解人们的语言,从而使语言成为人类与计算机交流的主要途径。

2.3.3 *音乐检索* 音乐作为音频的另一种重要的组成部分,其编码与解码相对容易,体积一般也比较小,所以制作、复制与传播都很方便,因而也是数量增加的最快的一种音频数据。相对于语音来说,音乐在基于语义检索的技术方面还有相当的差距,目前还处在比较初级的阶段。这是因为一方面人们对于人类的听觉系统的认识有限,人类的很多听觉特征还在研究之中;另一方面音乐是组成最复杂的音频类型,往往是各种声音,包括不同乐器、人声、以及自然界的声音等的复合,而且不同类型的音乐,其各种成分所占的比重也有所相同,这就为音乐信号的特征选取与计算增加了很大难度。

很显然,从音乐到文本的转换要失去大量的语义信息,因此在音乐检索中,用户提交文本作为查询关键词并不实用,最直接的查询方式是哼唱[6,22],字串匹配是在这种检索方式中

维普资讯 http://www.cqvip.com

最常使用的技术[28~30]，原理就是音乐可以被视为一串音符的序列，这些音符可以被进一步映射成字符。实际的系统中，往往将相邻音符之间的差异映射为字符，因为首先，用户哼唱的检索样本的音高可能与被检索的数据不同，但相邻音符之间的差异则与这种音高的差异无关；另一方面，当用户的哼唱存在一定错误的话，这种方式可以降低查询的错误率[23]，如可以将所有音符的差异分为三种：高(U)，低(D)和相同(S)，那么即使用户哼唱的音调与实际有一定出入，只要满足相邻音符差异不变即可保证查询结果的正确性，但同时这将会增加返回无关结果的可能性[22]。

由于上述自然音乐的复杂性，很多的音乐检索的研究只是针对 MIDI(其规范的详细介绍，请访问 www. midi. org.)音乐进行的[22~27]。MIDI 音乐是一种合成音乐，MIDI 文件实际是一种描述语言，存放各音轨所使用的乐器和播放的内容，MIDI 音乐播放器用于解释这种语言，并最终播放出声音。由于 MIDI 音乐的这种特性，其特征可以比较直接的从 MIDI 数据提取出来。

在字串匹配哼唱检索系统中使用的特征一般是旋律轮廓，文[22]利用一种改进的自相关方法对用户的哼唱数据计算这一特征。

为了支持多种检索模式，另一基于哼唱的检索系统 Musart 综合使用了马尔可夫模型，旋律轮廓和语音流等技术[6]。

虽然 MIDI 音乐检索使音乐检索技术有了很大的进展，但 MIDI 音乐在音乐中的比重比自然音乐小很多，所以对于自然音乐检索技术的研究更具有实际意义。文[31]对自动音乐分类进行了研究，提出了 Musical Surface 这一特征对音乐进行分类，并使用了9维向量(mean-Centroid，mean-Rolloff，mean-Flux，mean-ZeroCrossings，std-Centroid std-Rolloff，std-Flux，std-ZeroCrossings，LowEnegry)对这一特征进行描述。如前所述，目前音乐分类大都是人工完成的，所需工作量巨大，因此对于音乐自动分类的研究，在音频数据不断迅速增加的情况下，其重要性自是不言而喻。

为了进一步对音乐进行基于语义的组织，Islands of Music (http://www. ocfai. at/~clias/music)系统介绍了一种基于感知相似度对音乐库进行可视化组织和浏览的方法。在这个系统中，音乐作品被分布在一个类似地图的二维平面上，具有感知相似性的音乐会分布在一起，如图4所示。

在传统的音乐检索系统中，对于某一个音乐作品的查询只是返回艺术家，标题，或是歌词等中包含检索项的结果，与此不同，Island of Music 系统对于音乐的分类是基于感知特性的，所以特别适合于“听起来像 xx 作品”和“具有 xx 感情色彩的作品”等类似的偏向语义查询请求。

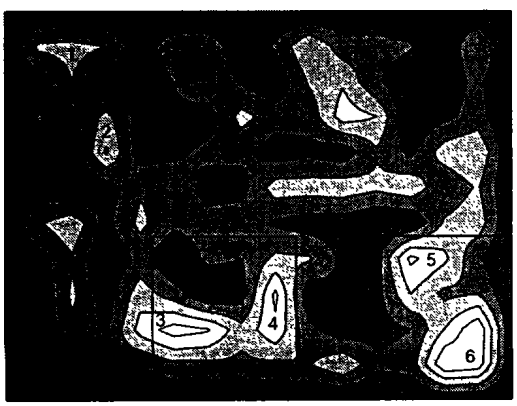

图4　具有六个聚类中心的音乐分布图

在实现上，Island of Music 使用了韵律模式这一特征，该特征包含了音乐在某一特定频段内节拍的强弱和快慢，是一个时不变参数。图5是特征提取的流程。

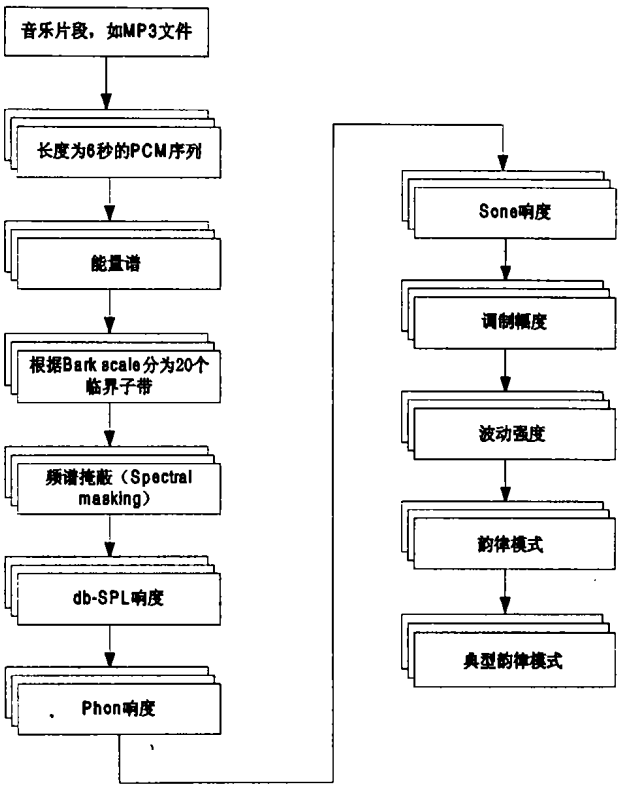

图5　韵律模式(Rhythm Pattern)特征提取流程

经过上述步骤得到一个20×60的矩阵，经过降维最终得到一个80维的向量。为了对特征向量进行分类，采用了 SOM (Self-Organizing Map)这一对多维向量特别有效的数据分析算法。经过分类的数据再利用平滑数据直方图(SDH)[34]方法，被映射到二维的图上，最终得到图4所示的分布图。

与以往的基于旋律轮廓检索的系统相比，Island of Music 在特征选取方面更接近语义层，同时，由于采用了 SOM 方法，对音乐的分类不需依赖于已有的关于音乐分类的知识，所有的类别自动产生，所有的分类操作自动完成。但另一方面，由于该系统的分类完全基于听觉特征，缺少先验知识，因而会产生一些不符合实际的类别，如将少量的古典音乐和摇滚乐分在一个单独的类别里，这是因为这些音乐中都含有很多管弦乐器的成分，鼓声相对较少，所以没有被分到各自的类别中。如果能够结合一定的先验知识，将会在一定程度上弥补这种缺陷。尽管改系统比较原始，分类的结果也并不十分理想，但为音乐数据的浏览与检索提供了一个可视化的途径，在一定程度上体现了音频检索的发展方向。

总结　作为多媒体检索的重要组成部分，音频信息检索在现实生活中具有重要的意义。本文对相关的技术和系统进行了综述性的介绍。随着语音识别技术的成熟，人们已经更注重于自然语言的理解与合成，使得计算机能通过语言与人进行交流，使语音成为人机交互的主要界面。在音乐的分析与检索方面，由于音乐本身的复杂性与人们对人类听觉认识的有限，相关的研究还处于起步的阶段，对于大规模音乐数据(如数据库包括各种不同类型音乐，音乐文件包括复杂声部等)的分类与检索还有待于进一步的增强。结合对人类听觉的深入研究，征服音乐分析与检索将会是不十分遥远的事情。

参 考 文 献

1　Baeza-Yates R, Ribeiro-Neto B. Modern Information Retrieval.

维普资讯 http://www.cqvip.com

ACM Press, Jan. 1999
2 Zhang T, Kuo C-C J. Video Content Parsing Based on Combined Audio and Visual Information. SPIE 1999, 1999, IV
3 Wold E, Blum T, Wheaton J. Content-base classification, search, and retrieval of audio. IEEE Multimedia, 1996, 3(3): 27～36
4 Tzanetakis G, Cook P. Multifeature audio segmentation for browsing and annotation. In: Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA99, New Paltz, NY, 1999
5 Tzanetakis G, Cook P. Experiments in computer-assisted annotation of audio. In: Proc. Int. Conf on Auditory Display, ICAD, 2000
6 Birmingham W, Dannenbert R, et al. Musart: Music Retrieval Via Aural Queries. In: Proc of the Annual Symposium on Music Information Retrieval(ISMIR 2001), Bloomington, ID, Oct. 2001
7 Lu Lie, Jiang Hao, Zhang Hongjiang. A robust audio classification and segmentation method. in ACM Multimedia, 2001
8 Foote J. Content-based retrieval of music and audio. In: C. C. J. Kuo et al., eds. Multimedia Storage and Archiving Systems Ⅰ. In: Proc. of SPIE, volume 3229, 1997. 138～147
9 Srinivasan S, Petkovic D, Ponceleon D. Towards robust features for classifying audio in the CueVideo System. In: Proc. of the seventh ACM intl. conf. on Multimeida, 1999. 393～400
10 Liu Z, Wang Y, Chen T. Audio feature extraction and analysis for scene segmentation and classification. Journal of VLSI Signal Processing Systems, June 1998
11 Boreczky J S, Wilcox L D. A Hidden Markov Model framework for video segmentation using audio and image features. In: Proc. of ICASSP'98, Seattle, May 1998. 3741～3744
12 Pfeiffer S, Fischer S, Effelsberg W. Automatic audio content analysis. In: Proc. of the fourth ACM intl. conf. on Multimedia, 1996. 21～30
13 Saunders J. Real-time discrimination of broadcast speech/music. In: Proc. ICASSP96, vol. II, Atlanta, May, 1996. 993～996
14 Scheirer E, Slaney M. Construction and evaluation of a robust multifeature music/speech discriminator. In: Proc. ICASSP 97, vol. II, IEEE, April 1997. 1331～1334
15 http://www-3.ibm.com/software/speech/windows/index.shtml
16 James G. Droppo Ⅱ. Time-Frequency features for speech recognition: [PhD thesis]. University of Washington
17 Rabiner L R. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE, 1989, 77(2): 257～286
18 Jelinek F, Mercer R, Roukos S. Principles of lexical language modeling for speech recognition. in Readings in Speech Recognition, ed. A Waibel and K. F. Lee, Morgan Kaufmann Publishers, 1990
19 Riccardi G, Bocchieri E, Pieraccini P. Non-deterministic stochastic language models for speech recognition. In: Proc. ICASSP'95, Detroit, May 1995. 237～240
20 Ron D, Singer S, Tishby N. The power of amnesia. In: J. Cowan et al., eds. Advances in Nueral Information Processing Systems, Morgan Kauffmann, 1996, 6
21 Teuhola J, Raita T. Application of a finite-state model to text compression. The Computer Journal, 1993, 36: 607～614
22 Ghias A, Logan J, Camberlin D, Smith B C. Query by humming: Musical information retrieval in audio database. In: Proc. ACM Int. Conf. on Multimedia, San Fancisco, CA, ACM, 1995. 231～236
23 Kosugi N, et al. A practical query-by-humming system for a large music database. In: Proc. ACM Int. Conf. on Multimedia, Los Angeles, CA, 2000. 333～342
24 Pampalk E, Rauber A, Merkl D. Using smoothed data histograms for cluster visualization in self-organizing maps. In: Proc of the Int. Conf on Neural Networks, 2002
25 Bainbridge D. The role of music IR in the New Zealand Digital Music Library project. In: Proc. of the Intl. Symposium on Music Information Retrieval, 2000
26 Tseng Y H. Content-based retrieval for music collections. in SIGIR, 1999, ACM
27 Rolland P Y, Raskinis G, Ganasicia J G. Musical content based retrieval: an overview of the Melodiscov approach and system. in Multimedia, Orlando, FL: ACM, 1999
28 Blackburn S, DeRoure D. A Tool for Content Based Navigation of Music. In: Proc. ACM Multimedia 98, 1998
29 McNab R J, Smith L A, Bainbridge D, Witten I H. The New Zealand Digital Library MELody inDEX. http://www.dlib.org/dlib/may97/meldex/05written.html, May 1997
30 Uitdenbogerd A, Zobel J. Melodic Matching Techniques for Large Music Database. In: Proc. ACM Multimedia 99, Nov. 1999. 57～66
31 Tzanetakis G, Essl G, Cook P. Automatic musical genre classification of audio signals. In: Proc. Int. Smposium on Music Information Retrieval(ISMIR), 2001
32 Pampalk E. Islands of music: analysis, organization, and visualization of music archives. Master's thesis, Vienna University of Technology, 2001
33 Pampalk E, Rauber A, Merkl D. Content-based organization and visualization of music archives. In: Proc of ACM Multimedia 2002, Juan-les-Pins, France, ACM, 2002
34 Rauber A, Pampalk E, Merk D. Using psycho-acoustic models and self-organizing maps o create a hierarchical structuring of music by sound similarities. In: Porc. Int. Symposium on Music Information Retrieval (ISMIR), Paris, France, 2002

(上接第53页)

于自己的业务，不必关心具体物理设备的大小、类型、位置及特性等物理参数，增加了用户的投资回报(ROI)。同时，采用存储虚拟化技术，用户已经购买的NAS产品和SAN设备都可以有效地利用起来，这样就保护了用户的已有投资，减少了总体拥有成本(TCO)。在一个集中存储池中能更充分地利用存储空间，使得存储可以被共享在一个异构服务器的网络环境集中管理。采用非对称的结构来实现虚拟网络存储系统，由于数据传输与控制信息采用不同的通道，因此系统I/O性能得到了显著的提高。

同时，在本文设计的非对称结构模型中，由于元数据服务器负责整个系统的全局存储管理，NAS和SAN的任何一个客户端的访问请求都要经过元数据存储服务器的统一管理和分配，这样就容易产生单点失效的问题，而且系统性能也会受到不同程度的影响，所以下一步工作，有待于开发冗余结构配置，研究冗余服务器配置环境下负载平衡和高效的数据锁机制算法等问题。

参 考 文 献

1 Arunkundram R S, et al. Special Edition Using Storage Area Networks. Que, 2001. 11
2 Liu Zhaobin, Xie Changsheng, Fu Xianglin, Cao Qiang. A high scalable and performance storage architecture for multimedia applications. In: Proc. of SPIE, v 4861, 2002. 116～120
3 Anderson D C, et al. Interposed Request Routing for Scalable Network Storage. ACM Transactions on Computer Systems, 2002, 20(1): 25～48
4 Blunden M, Berx-Debeys M, Sim D. Storage Networking Virtualization. Redbooks Publications (IBM), Dec. 2000
5 Phillips B. Have Storage Area Networks Come of Age? IEEE Computers, 1998. 7
6 Clark C T. The Virtualization of Storage. TidalWare white paper
7 Heath J R, Yakutis P J. High speed storage area networks using a fibre channel arbitrated loop. interconnect IEEE Network, 2000, 14(2): 51～56
8 Milligan C, Selkirk S. Online storage virtualization: the key to managing the data explosion. System Sciences, 2002. HICSS. In: Proc. of the 35th Annual Hawaii Intl. Conf. 2002. 2905～2913
9 Milanovic S, Petrovic Z. Building the enterprise-wide storage area network. EUROCON'2001, Trends in Communications, International Conference, 2001, 1(1)136～139
10 Griswold R. Storage topologies. Computer, 2002, 35(12): 56～63