

音乐可视化研究综述^{*})

屈天喜 黄东军 童卡娜

(中南大学信息科学与工程学院 长沙 410083)

摘要 随着计算机技术的发展,音乐可视化日益得到研究机构和公司的重视,出现了各种音乐可视化系统。作为一种新的可视化技术,音乐可视化更多地关注音频和图形、图像的关系,是跨音乐、数字音频、图形学、图像处理、虚拟现实等领域的又一新的研究方向,在娱乐、教育、艺术和商业中具有广阔的应用前景。本文介绍了音乐可视化的概念和分类,探讨了音乐可视化的研究内容与国内外研究进展,分析了当前亟待解决的问题,并展望了未来的发展趋势。音乐可视化是一种新的概念系统,但也存在很多问题需要解决,其研究具有很强的理论和实用意义。

关键词 音乐可视化,特征提取,虚拟现实

A Survey on Music Visualization

QU Tian-Xi HUANG Dong-Jun TONG Ka-Na

(School of Information Science and Engineering, Central South University, Changsha 410083)

Abstract With the development of computer technology, music visualization is attracting more and more attention of research institutions and companies. Various music visualization systems are presented in many different applications. As a new visualization technology, music visualization focuses on the relation between audio and image. It is a new research aspect in the field of combining music, digital audio, image processing, virtual reality and so on. Potential applications of music visualization span a wide spectrum from entertainment to education, from art to commerce. This paper introduces the concept and classification of music visualization techniques, and discusses the research content in this field. In particular, this paper summarizes the current research progresses. Finally, the open research problems are also pointed out. Music visualization discussed in this paper is a novel conceptual system; and many issues need to solve. The research on this topic is with great theoretical and practical value.

Keywords Music visualization, features extraction, Virtual reality

音乐通过听觉形式表达人们的思想感情,反映社会现实生活,是一种重要的意识形态。长期以来,听音乐是非常自然的事情。但是,实验表明,视觉的刺激比听觉的刺激更强烈,更能让人加深印象。有时候音乐的意境可能是比较高深的,难以理解的,如果能用影像表达出来,达到视听结合,必定会更加形象,更能理解作者的意境。随着计算机技术的发展,音乐可视化提上议事日程。音乐可视化将听觉与视觉相结合,增强音乐的可理解性。目前的媒体播放软件,如 Winamp, Windows media player 播放音乐时已经出现了一些可视化效果,如火焰、烟雾、水波等。但是,作为一项综合性技术,音乐可视化正期待人们的深入研究。

1 音乐可视化的概念及研究内容

1.1 基本概念

可视化(Visualization)是指在人脑中形成对某物(某人)的图像,是一个心智处理过程(Mental Process),促进对事物的观察及建立概念等。在计算机技术中,可视化是一种过程呈现手段,是人机交互的工具,如可视化编程、可视化建模等。

音乐可视化(Music Visualization)是对音乐表达的一种非主观的解释和判断,是为理解、分析、比较音乐的表现力和内部结构提供的一种呈现技术^[13]。音乐可视化作为一种新

兴的可视化技术,在当前各种流行的媒体播放软件中已经产生重要的影响,其中所涉及的特征提取、图像处理、虚拟现实等技术是多媒体方面研究的热点。

音乐可视化在对音乐的特征如波形、频率、音调、音高、节奏、速度、音色等进行提取之后映射到相应的可视化效果。这种可视化效果具体形式多样,可以是烟雾、水波、火焰等的变化;特别是基于计算机图形图像变换处理而产生的、复杂的、全新的视觉效果,如虚拟角色(小动物或人的动作变化)及其虚拟场景的变幻、音乐的音符、音高、节拍等的具体形象化视觉效果等,更是动人心弦。Windows Media Player 中的可视化效果便是音乐可视化的典型代表,同时也为音乐可视化提供了最初的解决方案。音乐可视化的出现,使得“看”音乐成为可能。它是跨音乐、数字音频、图形学、图像处理、虚拟现实等领域的又一新的研究方向,在娱乐、教育、艺术和商业中具有广阔的应用前景。

1.2 分类

从技术上看,音乐可视化方法可以从以下几个角度分类:

(1) 从维数角度分,可以分为二维的可视化和三维的可视化两种。“维”是一种度量,如几何平面即二维。长、宽、高便构成“三维空间”。我们现在所生存的空间就是典型的三维空间。我们所说的 3D 效果就是三维空间的具体表现。一般

^{*})国家自然科学基金项目(90304010)。屈天喜 硕士研究生,主要研究方向为多媒体技术、虚拟现实技术;黄东军 教授,博士,主要研究方向为计算机网络、分布式多媒体系统;童卡娜 硕士研究生,主要研究方向为多媒体技术、虚拟现实技术。

说来,二维的可视化实现起来比较简单,三维的可视化实现起来比较复杂,但效果逼真,现在的许多网络游戏和虚拟现实都往三维方向发展。三维技术可以创造特殊的视觉效果,而且可以与实景拍摄结合,产生不可思议效果。文[17,34,46]中提出的系统是三维音乐可视化系统的代表。

(2)从响应时间即实时性的角度分,有实时的可视化和预处理的可视化两种。实时对响应时间有严格的要求。实时系统根据其对于实时性要求的不同,又可分为软实时和硬实时两种类型。硬实时系统指系统要确保在最坏情况下的服务时间,即对于事件的响应时间的截止期限是无论如何都必须得到满足的。比如航天飞船的控制等就是这样的系统。软实时系统就是那些从统计的角度来说,一个任务能够得到有确保的处理时间,到达系统的事件也能够在此期限到来之前得到处理,但违反截止期限并不会带来致命的错误,如实时多媒体系统就属于软实时系统。实时的音乐可视化大都属于软实时系统,如 Winamp、Cubic Player、预处理 3D Max Studio 和 Softimage^[24]等;文[17,27,34,46]中提出的系统也是实时的音乐可视化系统。

(3)从交互性的角度分,可以分为交互的可视化和非交互的可视化两种。绝大部分可视化方法都是交互式的。交互,在这里指的是作为服务使用者的用户和作为服务提供者的音乐可视化系统之间的互动通信并交换信息的过程,有请求,有应答,需要双方参与。文[17,46]中提出的系统是交互式的音乐可视化系统。

(4)从音频数据的类型分,一般可以分为对自然音频的可视化和对结构性音频的可视化两种。MPEG-4 音频标准把音频分为两类:自然音频和结构性音频。自然音频分为一般性音频和语音,包括 MP3 和其它数据压缩格式的音频。结构性音频主要是指 MIDI(Musical Instrument Digital Interface)格式的音频,它主要是由一些指令和参数组成,包括音符及其起音、音高、音色、速度等信息,对于特征提取比较方便。其中文[17,34]中提出的系统即是基于结构性音频的音乐可视化系统,文[27]中提出的系统是基于自然音频的音乐可视化系统。

(5)从显示理论的角度分,有沉浸式、半沉浸式、非沉浸式三种^[37,38]。沉浸感是最近在很多虚拟现实、网络游戏等技术中反复出现的一个词。也常听到人们说沉浸在音乐中、沉浸在悲伤的旋律里、沉浸在美丽风景中之类的话。什么是沉浸呢?就是说信息接受主体完全融入到一个虚拟环境之中。当我们看一部特别吸引人的电视剧,那么很有可能看着看着就融入到那个剧情之中了,这个时候就会忘记自己的存在,自己的心情也随着电视剧主角的情节变化而变化,这就是沉浸感!显然,一部好的艺术作品需要沉浸感,一款好游戏需要营造沉浸感。同样,一个好的音乐可视化系统也要有这种沉浸感。文[46]中提出的系统是采用虚拟现实技术的沉浸式音乐可视化系统。

1.3 音乐可视化的研究内容

对音乐可视化的研究可以从以下几个方面展开:

(1)特征提取。音乐可视化离不开对音乐的特征提取。将音乐的特征提取出来之后,再把这些特征转换成可视化效果,如动作变换、场景变化等。根据对音频的内容分析^[13],音频的低层特征包括振幅、频率、相位等,高层特征包括音调、音高、节奏、速度等。好的音乐可视化效果要求对音乐的高层特征进行提取,但对自然音频的高层特征进行提取的难度很高,

目前流行的可视化效果都是基于对音乐的低层特征进行变换的。如 Windows media player, Winamp 等的可视化效果就是基于振幅和频率进行变换的。

(2)情感检测。音乐可视化需要能表达出音乐情感,这是音乐可视化的最高目标。Hevner 著名的论文通过实验研究了音乐的声音与听者的情感的关系。在实验中,当音乐播放的时候听者被要求写出他们感受的形容词。这项实验证明了一种假设,那就是音乐本身是带有情感因素的。Hevner 发现了这些描述性的形容词的存在并且把它们布置在一个圆内^[5]。她还发现在一个组中的标签总有一个相似的文化背景。Tao Li 等人^[10]在此基础上对形容词组做了一些改动,并利用支持向量机(Support Vector Machines)算法对大批量音乐文件进行情感检测分析,取得了较好的效果。

(3)可视化建模。可视化建模决定可视化效果。编程者要在头脑中形成一个可视化模型,即采用何种方式来可视化音乐。是烟雾、水波或者火焰等的形态变换,还是角色的动作变化及场景的变换;是音乐的音符、音高和音长形象化,还是可视化出一个钢琴根据音乐自行弹奏,等等。

(4)图形图像处理。可视化效果是由图形图像表达出来的。这就离不开图形图像处理,包括平移、缩放、旋转、切变、透视、波浪变换、极坐标变换、球面变换等。这些处理方法能够得到各种精彩的视觉效果,给人以极大的视觉刺激。场景中的所有对象都是可视化的主题。有些对象会在整个可视化过程中出现,有些则动态地出现在一个给定的时间内。仅仅通过一些简单的对象来创建复杂而又美观的场景是很困难的,而且对象的特征也应该为用户提供充分的可能性来表达情感,这就需要找到足够复杂的图像目标以及适当的图像处理,甚至希望通过几个有限的参数来控制这些对象及其变换^[24]。

(5)虚拟现实。虚拟现实的广义解释是由计算机生成的给人以沉浸感的虚拟环境^[37]。音乐可视化要想给人以最大程度的沉浸感,就必然要用到虚拟现实技术,使用虚拟角色和虚拟场景,根据提取的音乐特征使角色动作改变以及场景的变换。有些实时系统采用高端的可视化技术来创建实物大小的虚拟角色,从而强制性地使用户产生沉浸感。虚拟现实现在已经跟三维技术紧紧联系在一起,给人以极大的视觉享受。虚拟现实与动画最主要区别就是它的交互特性。系统可设置多种浏览模式,如行走、飞行、静物观察、摄像机动画,用户不需要定义很复杂的参数,即可实现不同方式的浏览。在运行过程中,控制模式可通过热区、热键来进行切换。行走模式中,缺省开启碰撞检测,即可实现游戏般的漫游。用户可以用鼠标、键盘、事件触发、定时触发、脚本流程来与三维场景中的物体或属性进行各种方式的互动。

(6)音乐可视化的应用。音乐可视化作为一个新兴的技术,有着非常广泛的用途:在娱乐方面,在表演现场或者酒吧装上音乐可视化系统,可以给人以强烈视觉冲击;游戏中,通过基于虚拟现实的音乐可视化技术,使用者可以通过音乐接口来控制虚拟角色达到娱乐;在教育方面,音乐学习者可以与显示音符或者乐谱的音乐可视化系统进行交互,检验自己的音乐学习效果,提高学习者的兴趣;在艺术方面,借助乐谱可视化和动画来设计舞蹈动作,从而实现虚拟表演;在商业方面,可以设计一种音乐可视化系统使得灯光组合随着不同的音乐输入而产生不同程度的亮暗变化、色彩组合的变化、图案的变化^[30];或者用于音乐爆竹烟花设计,随着不同的音乐输

人,同一种爆竹烟花可能在爆炸时产生不同的形态;或者用于控制广场内的音乐喷泉,即利用音乐的主要特征(频率、振幅、音色和节拍)控制喷水的花型组合变化、水柱高低、远近变化等^[31]。音乐可视化技术的应用前景十分美好。

2 音乐可视化研究进展

前面已经提到,音乐可视化主要研究内容为特征提取、情感检测、可视化建模、图像处理及虚拟现实技术在音乐可视化中的应用等。音乐可视化作为一种新兴的可视化技术,要求对这些方面有较成熟和多样性的解决方案。本节就目前国内外一些典型的相关研究方法阐述其研究进展。

2.1 音乐分析与特征提取

对自然音频和结构性音频的分析和特征提取方法是不一样的。为了便于对自然音频(如 wave、wma、mp3 等格式的音频)的内容分析和特征提取,应对原始音频数据进行预处理,目前出现了诸如 FFT、Wavelet、固定窗口等典型的变换方法^[13],在此基础上,再对其进行特征提取。而对于结构性音频(如 MIDI 格式的音乐),由于这类音频是由一些控制声卡发声的指令组成的^[2],很容易从中找到音色、音高、音长等信息的对应位置,然后按其格式的规则进行相应的转换,即可提取到这些特征。

李华等人^[27]利用 Windows media player 内部机制,即在播放音频文件的同时,可以实时得到当前时刻左右两声道、各 1024 个采样点的波形及频率数据,这些原始数据描述了当前音乐的特征。认为频率平均值变化幅度要大于波形平均值的变化幅度,可体现节拍之间的明显差别,故频率平均值是个合适的选择。

Martin Gasser 在文^[34]中提出,音频数据可以从时域或者频域去可视化,即将音频数据的振幅或者频率特征映射到可视化效果。时域显示出振幅情况,频域显示出声谱的能量。对符号化的音乐数据来说,如钢琴音乐,即可以被记录到计算机的钢琴音乐,符号化的信息包括音符的起音、音高和速度。在 MIDI 格式的音乐中很容易提取这些特征。但在 MIDI 格式中,响度只能从速度值中估计。但是这样的估计是不精确的,因为敲打的速度和感知的响度不是线性相关的。他指出,当前模仿响度的方法是通过在一个给定的时间内的全部弹奏的音符进行求和计算来完成的。

Taylor 等人在文^[46, 47]中指出,他们感兴趣的特征是音高、噪音的振幅、描述噪音音色质量的数据以及表演者弹奏电子琴时产生的和弦信息。对输入的噪音进行特征提取是通过 Puckette 等人在文^[28]中提出的提琴类对象来完成的。提琴类对象提取歌唱者的音高和振幅,产生了一些描述歌唱者噪音和频谱的原始峰值数据。通过测试这个和声频谱,就可以定义一个用数字表示的描述符,这个描述符显示了歌唱者的噪音振幅是否集中在某个基本频率内,或者在更高的泛音中是否存在重要的噪音振幅。还可以用数字描述歌唱者噪音的音色。Taylor 等人指出,文^[47]的系统中监视的 MIDI 事件是为了决定在电子琴键盘上弹奏了哪些和弦,并强调对旋律的感知,因为旋律才是交互虚拟角色最重要的部分。

2.2 情感检测及情感表达

音乐是可以表达情感的,音乐可视化离不开情感检测的研究。Hevner 在 1936 年通过一项实验来研究音乐情感。在这个实验里,她要求听者写出最能描述当前播放音乐的形容词。这项实验意味着音乐本身就包括情感特征。Hevner 把

这些描述性形容词列出来,在一个圆内分为 8 组^[5]。

1958 年, Farnsworth^[6]在此基础上把这些形容词重新提取出来并重新分组,分为 10 组(如表 1)。Tao Li 等人^[10]在此基础上,猜测音乐中的情感检测可以通过分析音乐信号来实现,并使用多标签分类的方法解决了这个问题。对音乐同时进行多种分类是必要的,因为同一首音乐可以由多个形容词来描述。音乐特征提取后,对这些提取出来的音乐特征使用支持向量机(Support Vector Machines)算法来定义多标签。

表 1 Farnsworth 对音乐描述的形容词分组

A	cheerful, gay, happy	F	dark, depressing
B	fanciful, light	G	sacred, spiritual
C	delicate, graceful	H	dramatic, emphatic
D	dreamy, leisurely	I	agitated, exciting
E	longing, pathetic	J	frustrated

大自然中不同的色彩变化,能使我们产生不同的观感。人们在长期的社会生活与实践中对不同的色彩也逐渐形成了某些特定的含义、感受和心理反映。比如在生活中,白色往往给人以纯洁、高雅的感觉,绿色象征生命、青春与和平,等等。这些象征性的具有丰富感情色彩的颜色观念也会随着民族传统、欣赏习惯、文化修养的不同而有不同的反映。这些情感的产生,并不是色彩本身的功能,而是人们赋予色彩的某种文化特征。使某种颜色具备某些含义和象征,久而久之,这些因素又反过来影响着人们对色彩的感受,人们在观察色彩时往往是受着这些心理、文化因素的影响和左右,使人们产生不同的情感和不同的色彩感受。因此充分认识、理解和运用色彩的这种情感象征性,对于音乐可视化的图像色彩和场景色彩的处理是很有指导意义的。

周海宏以联觉关系为突破口,用心理学的实证方法证明了与音乐听觉相关的 6 种联觉对应关系规律:与音高(声音频率的高低)相关的联觉、与音强(声音的强弱)相关的联觉、与时间(声音的长短、快慢和疏密)相关的联觉、与时间变化率(起音速度的快慢)相关的联觉、与紧张度相关的联觉和与新异性体验相关的联觉^[7]。联觉,即从一种感觉引起另一种感觉的心理活动。以与音高相关的联觉为例说明这些联觉对应关系,如表 2。此研究对音乐可视化的视觉表达、情感表达以及其它方面的感觉表达有比较实用的指导意义。

表 2 与音高相关的联觉

声音频率	高	低
视觉亮度	亮、明、白	暗、昏、黑
情感体验	兴奋、快乐、积极、向上	抑制、哀伤、消沉、沉深
空间知觉	上、浅、高	下、深、低
物体质量	轻、飘	重、沉
物体形状体积	小	大
物体运动体态	敏捷、灵巧	迟钝、笨拙

Cooke 通过对大量经典音乐作品的研究,在文^[8]中提出了音乐和情感之间的更为一般性的关系。他已经了解到某些特征就是大量音乐片断中显著的情感特征,比如在十二平均律的音阶中,如果一个旋律中包含很多小三度(minor third)的实例,那么这个情感将是难以接受的或者悲伤的。相反,大三度(major third)则常常表示和谐或者喜悦,等等。

Taylor 等人在文^[47]中指出,情感表达可以采取多种形式。角色可以从一个它们发现快乐的特别的旋律中产生高兴的感受。

觉,或者在感知到发现威胁的一系列和弦之后进入到一个恐惧的状态。角色可以不喜欢某种调式的乐曲,或者对某种调式表达羡慕之情。Taylor 等人应用了 Cooke 在文[8]中提出的那些规则,把歌唱的音高联系到 Cooke 研究的情感内容,这可以用于修改角色的内部状态和触发相应的行为。采用的 ANIMUS 框架使用了驱动机制来控制角色的行为,即当一个指定的角色的特征的级别达到一个目标最大值或者最小值时就触发相应的行为。例如,角色可能有一个“快乐驱动”机制,当歌唱者的一个旋律中包含了很多大三度实例时对应的级别值就增大,当出现小三度时级别值就减小。还有其它多种驱动,如“专心驱动”、“行走驱动”等。其它的特征如和弦、音色等也可以联系到驱动系统中,影响角色的反应。

2.3 虚拟现实与音乐可视化

随着计算机图形学技术的进步和相关项目的推动,虚拟现实近 40 年来从无到有,快速发展。现在的虚拟现实技术已经广泛应用于建筑行业、室内设计、工业、文物保护、交通、游戏、军事等领域^[37]。简言之,只要是用到计算机三维图形的地方,都有虚拟现实的用武之地。因此,音乐可视化要有极强的三维视觉效果和沉浸感,虚拟现实是一个必然的选择。

由 Oliver 等人^[39]创建的“歌唱树”(The Singing Tree)系统,可以使用户沉浸在这个艺术空间里。这个环境对用户的声音会产生视觉和听觉的反应。“歌唱树”为用户提供了声音可视化的反馈,让他们在一个稳定的音高上保持一个被延长的音符,从而提高他们的歌唱水平。

Jack Ox^[40]建立了一个使用三面墙式的沉浸感的洞穴状可视化系统,用来探究音乐输入的和声结构。通过建立用户可以随意探测和操纵的三维场景,系统的“色彩化管风琴”(Color Organ)使得用户可以看到音乐片段中形象化的和声关系。

由 Golan Levin 等人创建的可视化系统“Messa di voce”^[22],将表演者的物理特性和他们表演的虚拟空间的物理特性联系起来。该系统形象化地表示了现场歌唱者的发声法,系统中虚拟场景的大小让表演者感觉自己也是表演所在的虚拟空间的一部分。

Robyn Taylor 等人在文[46]中描述了一个用户交互作用虚拟角色的沉浸感的音乐可视化系统。为了提取音乐特征和触发虚拟角色的行为,该系统参数化了现场音乐的表演(通过麦克风和电子琴)。系统的虚拟角色显示在一个很大的立体投影屏幕上,并且角色是三维实体大小的,从而使得用户产生极强的沉浸感。在此基础上,Robyn Taylor 等人在文[47]中又提出,将人性化的情感因素加入到虚拟角色的行为中,从而使得该音乐可视化系统更加趋于完美。

2.4 音乐可视化的实现技术

李华等人^[27]采用 COM 插件技术与 Windows Media Player 结合,实现了一种基于实时渲染技术的水彩画风格的音乐可视化效果(如图 1)。主要应用基于 alpha 通道的水彩画效果实现技术,基于优化物理模型、逐段拼接以及组合反走样等方法的动态场景仿真技术,借助 Direct3D 实现动画场景。结果显示,该方法可以在较低的 CPU 占用率和较小内存请求下达到良好效果。

Elaine Chew 等人^[15~18]开发了 MuSA. RT(Music on the Spiral Array. RealTime)系统(如图 2)。该系统是 MuSA 系统^[18]的扩展,通过对现场表演的 MIDI 输入实时处理,分析和映射到三维螺旋模型上,反映出音乐的音高(pitch)、和弦

(chord)和主音(tonal center)等相关音乐特征。如在螺旋模型中(如图 3),音高实体是用空间上的不同颜色的圆球来表示的,和弦也包含在螺旋模型中以三个音高实体为顶点的三角形内,对大度(major key)和小度(minor key)之间的距离也有一定的反映。该系统的用户可以通过游戏设备和导航器来操纵螺旋模型的视图及感兴趣的部位等。



图 1 具有水彩画风格的音乐可视化效果



图 2 MuSA. RT 系统

MuSA. RT 系统不但是交互式的音乐可视化,而且是实时的音乐分析,在音乐分析、信息检索、表演分析及声调合成等方面有着广泛的应用。Elaine Chew 还指出期望通过不同的形式和艺术方法来交互作用该三维空间。

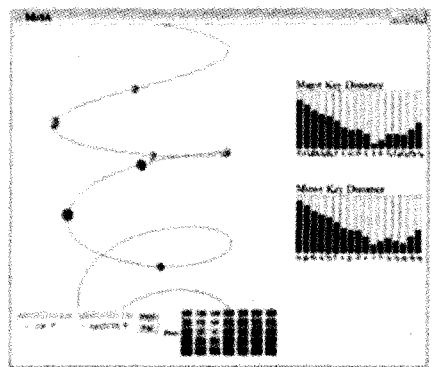


图 3 MuSA 显示面板的一个案例

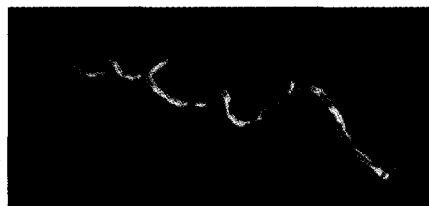


图 4 三维表演蠕虫的可视化

Martin Gasser 在文[34]中提出了一个表达钢琴表演的

三维蠕虫的可视化系统(如图4)。表演蠕虫所在的空间是时间、速度和响度构成的三维空间。这种三维蠕虫是由反映速度和响度关系的二维蠕虫扩展而来的。三维蠕虫可视化是一些连续的空间曲线,而原来的二维蠕虫可视化只是速度/响度空间的一系列点。

这种三维表演蠕虫可视化系统的基本用户接口是建立在钢琴卷的观点基础之上的。钢琴卷是一个已经得到广泛应用的很强大的可视化方法,它是在时间、音高、速度决定的三维空间里的,除音符的起音外,音高信息和清晰度(与两个连续音符的重叠量有关)都可以从钢琴卷上清楚地看到(如图5)。因此,在三维蠕虫可视化系统中,根据平滑的清晰度值,曲调的清晰度就可以通过蠕虫的形状来显示。



图5 三维的钢琴卷

Robyn Taylor 等人^[46,47]建立了一个基于 ANIMUS 框架^[19]的采用虚拟现实技术的音乐可视化系统。在这个系统中,音乐家可以交互作用于在虚拟环境下的有生命的虚拟角色(如图6)。这些虚拟角色“倾听”音乐家的表演并且同时改变它们的行为。这个系统使用 Max/MSP^[43]参数化音乐表演乐器和音乐家表演的音乐,一些有用的音乐特征如音高、振幅、和弦、音色等从现场表演中实时提取出来,然后这些提取出来的音乐特征映射到角色的行为。即一些特定的音乐特征被预先定义到一个触发器里触发特定的行为回应,因此当音乐家表演的时候就可以观看角色的回应。这种系统可以有多种用途。用户可以通过音乐表演来控制虚拟角色以达到娱乐目的。教育方面,虚拟角色的回应可以激励用户的音乐表演,从而提高音乐表演热情。艺术方面,乐谱和动画可以用来设计舞蹈动作从而实现虚拟表演。

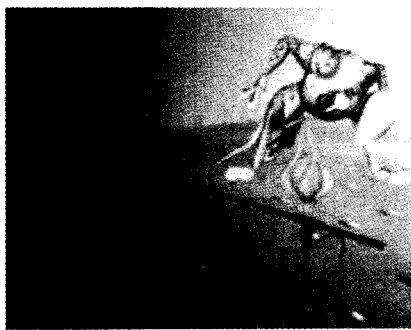


图6 正在交互作用虚拟角色的用户

3 音乐可视化亟待解决的问题

3.1 音乐特征的提取

音乐中有各种各样的特征,如频率、振幅、音色、音调、音高、音长、和弦、速度、响度、节拍、旋律等,以及近来提出的一些新的特征^[13],如能量、过零率、频谱矩、频谱流、带宽、带周期、噪音帧率等。音乐中有如此多的特征,对每种特征都进行提取是很耗费时间的,也是很现实和没有必要的。因此,应该根据音乐可视化的应用方面的要求,提取出音乐中的那些

必要的特征,忽略或者去除那些不必要的特征,从而缩短程序运行时间,达到实时性等方面的要求。在交互式的音乐可视化系统中,对于不同输入设备(如麦克风和 MIDI 输入设备)得到的数据,对其特征提取的方法也应该是不一样的。此外,音乐的风格也是多样化的^[1],最大的分类是严肃音乐和流行音乐,而流行音乐下有乡村音乐、民谣、爵士乐、摇滚乐、流行美声等分类,如果能在对音乐的特征提取过程中得到音乐的分类,然后采用不同的可视化风格来反映这些不同的音乐风格,效果一定会更好一些。因此,在特征提取过程中得到音乐风格的特征,是音乐可视化进一步的追求目标。

3.2 对音乐的情感检测与情感表达

音乐既是一种娱乐方式,也是抒发作曲者和歌手情感的手段。情感因素是音乐中的一个重要的特征,在音乐可视化研究中越来越受到重视。因此,必须建立一个完善的情感检测以及情感表达机制。这种情感检测机制应当是综合性的,包括对音乐风格的分析、对音乐与情感关系的准确刻画、联觉中的各音乐特征与情感的对应。情感表达机制还应扩充表达情感的角色动作集,尽可能准确地表达出尽可能多的情感。

3.3 以用户为中心的系统设计

应用系统的设计应该以用户为中心,也就是说要有灵活的图形用户界面(GUI),实时性地反映用户的操作,用户的操作和输入的方式要尽可能地简单,而输出的结果要尽可能完美,等等。这就要做好用户需求调查和分析。在设计音乐可视化系统之前,必须要对用户的需求进行调查和分析,以此来确定系统的用户类型和应用类型,如教育方面的音乐教师和学生,商业娱乐方面的大众人群等,从而设计相应的可视化系统。

3.4 场景对象和动作建模

Windows media player 中的可视化效果是音乐可视化的雏形,并为音乐可视化提供了一个有用的参照。但是它的可视化效果过于简单和单调,使得用户感觉到这些可视化模式与所播放的音乐之间并没有什么关系,只是有一些微小的变化重复出现而已^[51]。这种面向大众的音乐可视化的娱乐应用,应该要另辟蹊径,采用场景以及角色对象来实现,角色对象可以是各种拟人化的动物、花草树木、自定义的动画角色甚至人本身,而场景对象可以是山河湖泊、花草树木等。为了避免角色的动作种类太少或者场景对象的种类太少而产生的单调性,以及表达音乐中各种复杂的情感因素,就必须尽量扩充角色的动作集^[24,47],增加场景对象的种类。

3.5 虚拟现实技术的硬件和软件支持

音乐可视化要有好的沉浸感,就要用到虚拟现实技术。这就需要解决相关的硬件设备和软件技术方面的支持问题。硬件设备方面,三维立体声设备、大屏幕(主要有洞穴状、圆柱型或者由矩形块拼接成的大屏幕)投影设备等^[37,38]是必不可少的。文^[46,47]的系统中的大屏幕的立体显示、实体大小的虚拟角色以及用户输入接口的最自然的方法(如图6)都可以帮助减少现实和虚拟的距离,让用户体验与虚拟角色之间的真正的交互过程,从而产生强制性的沉浸感。软件技术方面,如碰撞检测算法,以此来确保虚拟角色的行走路径和活动范围的合理性;丰富的特效,如模拟火焰、水流、烟雾、雨雪等,光照、阴影等,花草树木随风摆动、实时水波等;触发机制和动作定义,由特定的音乐特征触发角色相应的动作,通过给定的几个参数就可以实现各种动作等。

3.6 网络多机并行处理问题

由于视觉效果是对音乐播放的补充,因此,不能由于处理可视化效果而占用了大量的 CPU 处理时间而使得音乐本身播放的不连续或者可视化效果的不连续。因此,一个效果好的音乐可视化系统需要性能优化的算法以及高性能的计算机(也包括高性能的显卡和声卡)等方面的支持。

如果要实现的算法和功能足够复杂,如三维的虚拟现实技术,则需要将多台计算机组成一个网络,进行多机并行处理操作,由不同的计算机处理不同的功能,在尽量短的时间内实现总体功能,如文[36,46]中都用到网络多机并行处理。

结论 音乐可视化作为一种新兴的可视化技术,关注音乐的特征提取、情感检测、图像处理等技术,在娱乐、教育、艺术和商业中具有广阔的应用前景。本文中,我们介绍了音乐可视化的概念、分类和研究内容,国内外的研究进展和亟待解决的问题,展望了进一步的研究方向,期待本文的工作能够推动同行学者对这一新兴技术的关注和研究。

参 考 文 献

- 李重光. 音乐理论基础. 北京:人民音乐出版社,2000
- 谢朝宗. 音乐设备数字接口(MIDI)与计算机音乐. 北京:北京希望电子出版社,1995
- Chai W, Vercoe B. Using User Models in Music Information Retrieval Systems. In: Proceedings International Symposium on Music Information Retrieval, 2000 (8):201~205
- Deutsch D, Feroe J. The Internal Representation of Pitch Sequences in Tonal Music. *Psychological Review*, 1981; 503~522
- Hevner K. Experimental studies of the elements of expression in music. *American Journal of Psychology*, 1936 (48):246~268
- Farnsworth P R. The social psychology of music. New York: The Dryden Press, 1958
- 周海宏. 音乐与其表现的世界. 北京:中央音乐学院出版社,2004. 58~99
- Cooke. The Language of Music. New York: Oxford University Press, 1959. 89~99
- Wieczorkowska A, Synak P, Lewis R, et al. Extracting Emotions from Music Data. 2005. <http://www.springerlink.com/index/LRC0995XXL5M12X4.pdf>
- Li T, Ogihara M. Detecting Emotion in Music. 2001. <http://ismir2003.ismir.net/papers/Li.PDF>
- Cardle M, Barthe L, Brooks S, et al. Music-Driven Motion Editing: Local Motion Transformations Guided by Music Analysis. In: Proceedings Eurographics UK, Leicester, 2002. 38~44 <http://www.cardle.info/lab/publications/EGUK2003.htm>
- Zhu J J, Lu L. Perceptual Visualization of A Music Collection. In: Proc of IEEE International Conference on Multimedia and Expo (ICME05), 2005. 1058~1061
- 白亮,陈剑赞,老松杨,吴玲达. 音频内容分析研究. 见:中国计算机大会, 2003. 492~496
- Brougher K, Strick J. Visual Music. In: Synaesthesia in Art and Music Since 1900. New York: Thames and Hudson, 2005. 15~20
- Chew E. Modeling Tonality: Applications to Music Cognition. In: Proc 23rd Annual Meeting of the Cognitive Science Society, Edinburgh, Scotland, 2001
- Chew E, Chen YC. The Spiral Array: An Algorithm For Determining Key Boundaries. Special issue of the LNCS /LNAI on Music and Artificial Intelligence. Heidelberg, Germany: Springer, 2002
- Chew E, Francois A. MuSA. RT: Music on the Spiral Array . RealTime. 2005. http://cappriccio.usc.edu/MuSA_RT
- Chew E, Chen Y C. Mapping midi to the spiral array: Disambiguating pitch spellings. In: Proc 8th INFORMS Computer Society Conference, Chandler, AZ, 2003. 259~275
- Foote J. Visualizing Music and Audio using Self-Similarity. In: Proceedings of ACM Multimedia, Orlando, Florida, 1999. 77~80
- Morchen F, Ultsch A, Thies M. MusicMiner: Visualizing timbre distances of music as topographical maps; [Technical Report]. Dept. of Mathematics and Computer Science, University of Marburg, Germany, 2005 (47). <http://www.mybytes.de/papers/moerchen05musicminer.pdf>
- Misra A, Wang G, Perry R C. Sndtools: real-time audio DSP and 3D visualization. In: Proceedings of the 2005 International Computer Music Conference, 2005. <http://www.cs.princeton.edu/sound/publications/sndtools-icmc2005.pdf>
- Levin G, Lieberman Z. In-situ Speech Visualization in Real-Time Interactive Installation and Performance. In: Proceedings of the 3rd International Symposium on Non-Photorealistic Animation and Rendering. New York: ACM Press, 2004. 7~14
- Jain V, Lawrence K. Exploratory analysis and visualization of speech and music by locally linear embedding. In: Acoustics, Speech, and Signal Processing, 2004. 984~987. <http://ieeexplore.ieee.org/iel5/9248/29345/01326712.pdf>
- Kubelka O. Interactive music visualization. 2000. <http://www.cescg.org/CESCG-2000/OKubelka/paper.pdf>
- Collopy F. Visual Music in a Visual Programming Language. In: Proceedings of the IEEE Symposium on Visual Languages, Washington DC: IEEE Computer Society, 1999. 111~118
- Lund H. Visual Music. 2004. <http://visual-music.net/files/VisualMusicAdRdF.pdf>
- 李华,胡春晖,顾明. 基于实时渲染技术具有水彩画风格的音乐可视化方法. *计算机应用*, 2005, 25(3):729~731
- Puckette M, Apel T, Zicarelli D. Real-Time Audio Analysis Tools for Pd and MSP. In: Proceedings of the International Computer Music Conference, International Computer Music Association, 1998. 109~112
- Pampalk E. Islands of Music Analysis, Organization, and Visualization of Music Archives. Master's thesis, Vienna University of Technology, 2001. <http://citeseer.ist.psu.edu/481449.html>
- 王仰昭,谭瑞勇. 视听彩色音乐王. 五邑大学学报, 1999, 13(1). <http://www.wanfangdata.com.cn/qikan/periodical/Articles/wydxxb/wydx99/wydx9901/990109.htm>
- 蔡彬. 可编程控制器在音乐灯光喷泉中的应用. *农机化研究*, 2005 (4). <http://engine.cqvip.com/content/tp/97903x/2005/000/004/gc45...tp1-15620193.pdf>
- Langner J, Goebel W. Visualizing Expressive Performance in Tempo-Loudness Space. *Computer Music Journal*, 2006, 27(4): 69~83. <http://www.scf.usc.edu/~ise575/b/reports/langner-cheng.pdf>
- Maloney K. Sounding Images and Imaging Sounds - Audiovisual Interactivity in Performance. In: Sight Lines Thesis Projects, Graduate Program in Visual criticism, CCA, 2005. <http://sites.cca.edu/sightlines/pdf/05kmaloney.pdf>
- Gasser M. Interactive Visualization of Expressive Piano Performance. 2005. <http://www.cp.jku.at/research/papers/Gasser-diplomarbeit.pdf>
- Lee W K, Leung B, Lai K W. Seeds of harmony - edutainment platform on music creativity. 2006. <http://neuron.sd.polyu.edu.hk/seeds-of-harmony/login-resource/www2006-alpha-ee.pdf>
- Wang G, Misra A. Building collaborative graphical interfaces in the audicle. In: Proceedings of the 2006 conference on New interfaces for musical expression, Paris: IRCAM - Centre Pompidou, 2006. 49~52
- 潘志庚. 虚拟现实和数字化文化遗产. 2005. <http://www.86vr.com/files/2005-07/6029/slides-zgpan.pdf>
- 戴国忠. 虚拟现实中的人机交互. 2005. <http://www.cad.zju.edu.cn/chinagraph/chinese/hotreport/CG-Symposium05-DaiGuozhong-Interactive.pdf>
- Oliver W, Yu J, Metois E. The Singing Tree: Design of an Interactive Musical Interface. In: Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques. New York: ACM Press, 1997. 261~264
- Ox J. Two Performances in the 21st Century Virtual Color Or-

- gan. In: Proceedings of the Fourth Conference on Creativity and Cognition. New York: ACM Press, 2002. 20~24
- 41 Taylor R M, Hudson T C, Seeger A, et al. VRPN: A Device-Independent, Network Transparent VR Peripheral System. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, New York: ACM Press, 2001. 55~61
- 42 Torres D, Boulanger P. A Perception and Selective Attention System for Synthetic Creatures. In: Proceedings of the Third International Symposium On Smart Graphics, 2003. 141~150
- 43 Cycling '74. Max/MSP. <http://www.cycling74.com/products/maxmsp>
- 44 Torres D, Boulanger P. The ANIMUS Project: A Framework for the Creation of Interactive Creatures in Immersed Environments. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, New York: ACM Press, 2003. 91~99
- 45 Torres D. The ANIMUS Project: A Framework for the Creation of Emotional and Rational Social Agents. <http://www.cs.ualberta.ca/~dtorres/projects/animus/docs/proposal.pdf>
- 46 Taylor R, Torres D, Boulanger P. Using Music to Interact with a Virtual Character. In: Proceedings of New Interfaces for Musical Expression, Vancouver, BC, Canada, 2005. 220~223
- 47 Taylor R, Boulanger P, Torres D. Visualizing Emotion in Musical Performance Using a Virtual Character. In: 5th International Symposium on Smart Graphics, Munich, Germany, 2005. 13~24
- 48 McDonnell M. Visual Music. 2003. http://www.soundingvisual.com/visualmusic/visualmusic2003_2004.pdf
- 49 Baumann S. Visualization for Music IR. 2005. <http://ismir2005.ismir.net/documents/Baumann-ISMIR05Tutorial.pdf>
- 50 Chew E. Foreword to special issue on music visualization. Computers in Entertainment, New York: ACM Press, 2005, 3(4): 1~3
- 51 Winzenried R. Hear the Colors, See the Music. 2006. <http://www.symphony.org/news/room/06jfl-articles/currents.pdf>

(上接第 15 页)

- 2 Bar-Yehuda R, Halldórsson M, Naor J, et al. Scheduling split intervals. In: Proc. Thirteenth Annu. ACM-SIAM Symp. On Discrete Algorithms, 2002. 732~741
- 3 Crescenzi P, Kann A V. Compendium of NP Optimization Problems. <http://www.nada.kth.se/viggo/problemist/compendium.html>, 1998
- 4 Arkin E M, Hassin R. On local search for weighted k-set packing. In: Proc. 5th Ann. European Symp. on Algorithms, Lecture Notes in Comput. Sci. 1284, Springer-Verlag, 1997. 13~22
- 5 Cesati M. Compendium of Parameterized Problems. <http://citeseer.ist.psu.edu/cesati01compendium.html>, 2001
- 6 Chen J. Parameterized computation and complexity: a new approach dealing with NP-hardness. Journal of Computer Science and Technology, 2005, 20: 18~37
- 7 Wigderson A. Improving the performance guarantee for approximate graph coloring. Journal of the Association for Computing Machinery, 1983, 30(4): 729~735
- 8 Bar-Yehuda R, Moran S. On approximation problems related to the independent set and vertex cover problems. Discrete Applied Mathematics, 1984, 9: 1~10
- 9 Boppana R, Halldórsson M M. Approximating maximum independent sets by excluding subgraphs. Bit 32, 1992. 180~196
- 10 Hastad J. Clique is hard to approximate within $n^{1-\epsilon}$. Acta Math, 1999, 182(1): 105~142
- 11 Feige U, Goldwasser S, Lov'asz L, et al. Interactive proofs and the hardness of approximating cliques. J ACM, 1996, 43(2): 268~292
- 12 Arora S, Safra S. Probabilistic checking of proofs: A new characterization of NP. Journal of the ACM, 1998, 45: 70~122
- 13 Arora S, Lund C, Motwani R, et al. Proof verification and intractability of approximation problems. Journal of the ACM, 1998, 45: 501~555
- 14 Tarjan R. Finding a maximum clique: [Technical Report]. 72-123. Ithaca, NY: Computer Sci. Dept, Cornell Univ, 1972
- 15 Tarjan R, Trojanowski A. Finding a maximum independent set. SIAM Journal on Computing, 1997, 6(3): 537~546
- 16 Jian T. An $O(20.304n)$ algorithm for solving maximum independent set problem. IEEE Transactions on Computers, 1986, 35(9): 847~851
- 17 Robson J M. Algorithms for maximum independent sets. Journal of Algorithms, 1986, 7(3): 425~440
- 18 Fomin F V, Grandoni F, Kratsch D. Measure and conquer: a simple $O(2^{0.288n})$ independent set algorithm. In: Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithm, 2006. 18~25
- 19 Gabow H N. An efficient implementation of Edmonds' algorithm for maximum matching on graphs. Journal of ACM, 1976, 23: 221~234
- 20 Hurkens C A J, Schrijver A. On the size of systems of sets every t of which have an sdr, with an application to the worst-case ratio of heuristics for packing problems. SIAM Journal Discrete Math, 1989, 2: 68~72
- 21 Trevisan L. Non-approximability results for optimization problems on bounded degree instances. In: Proc. 33rd ACM Symp. on Theory of Computing, 2001
- 22 Hazan E, Safra S, Schwartz O. On the Complexity of Approximating k-Set Packing. Computational Complexity, 2006. 20~39
- 23 Arkin E, Hassin R. Approximating weighted set packing by local search. Math Oper Res, 1998. 640~648
- 24 Bafna V, Narayan B, Ravi R. Nonoverlapping local alignments (Weighted independent sets of axis-parallel rectangles). Discrete Appl Math, 1996: 41~53
- 25 Chandra B, Halldórsson M M. Greedy Local Improvement and Weighted Set Packing Approximation. Journal of Algorithms, 2001. 223~240
- 26 Berman P. A $d/2$ approximation for maximum weight independent set in d -claw free graphs. In: Proceedings, Seventh Scandinavian Workshop on Algebraic Theory (SWAT), Lecture Notes in Computer Science, Vol 1851. New York/Berlin, Springer-Verlag, July 2000. 214~219
- 27 Jia W, Zhang C, Chen J. An efficient parameterized algorithm for m -set packing. J. Algorithms, 2004, 50: 106~117
- 28 Chen J, Friesen D K, et al. Using nondeterminism to design deterministic algorithms. Algorithmica, 2004, 40: 83~97
- 29 Koutis I. A faster parameterized algorithm for set packing. Information Processing Letters, 2005, 94: 7~9
- 30 Fellows M R, Knauer C, Nishimura N, et al. Faster fixed-parameter tractable algorithms for matching and packing problems. Lecture Notes in Computer Science 3221 (ESA 2004), 2004: 311~322
- 31 Kneis J, Molle D, Richter S, et al. Divide-and-color. In: Proceedings of the 32nd International Workshop on Graph-Theoretic Concepts in Computer Science (WG 2006), LNCS 4271. Springer, 2006
- 32 Chen J, Lu S, Sze S-H, et al. Improved algorithms for path, matching, and packing problems. In: Proc. 18th ACM-SIAM Symposium on Discrete Algorithms (SODA 2007), 2007
- 33 Alon N, Yuster R, Zwick U. Color-coding. Journal of the ACM, 1995, 42: 844~856
- 34 Liu Y, Lu S, et al. Greedy localization and color-coding: improved matching and packing algorithms. In: Proc. 2nd International Workshop on Parameterized and Exact Computation, Lecture Notes in Computer Science 4169, 2006. 84~95
- 35 Chen J, Kanj I, Meng J, et al. On the effective enumerability of NP problems. In: Proc. 2nd International Workshop on Parameterized and Exact Computation (IWPEC '06), Lecture Notes in Computer Science 4169, 2006. 215~226