基于分层次聚类的 MIDI 音乐主旋律提取方法

冯国杰,王吉军
FENG Guo-jie,WANG Ji-jun

大连大学 辽宁省智能处理重点实验室,辽宁 大连 116622
Liaoning Key Laboratory of Intelligent Information Processing,Dalian University,Dalian,Liaoning 116622,China
E-mail:fguojie@gmail.com

FENG Guo-jie,WANG Ji-jun.Melody extraction method of MIDI music files based on layer clustering.Computer Engineering and Applications,2009,45(26):233-235.

Abstract: In order to extract melody of polyphonic MIDI music accurately,and reduce the extraction errorless in the case that melody distributed on the instrument channel or located on lowest pitch notes synchronously,this paper proposes an approach to extract melody of polyphonic MIDI music based on layer clustering.First,this paper analyzes the MIDI file;and then eliminates the channels those do not contain melodic information and control notes in each channel;consequently,picks up the melody through the note set which with the feature of pitch histogram summed up in a notepad file.The results are compared with manual output and show that the implemented method can extract melody accurately.

Key words: MIDI analysis;polyphonic music;melody extraction;layer clustering

摘 要:为了准确提取多音轨 MIDI 主旋律,同时减小主旋律分布在乐器音轨或音高较弱部分所产生的提取误差,提出了基于分层次聚类的多音轨 MIDI 主旋律提取方法。首先解析 MIDI 音乐文件,然后去除每一音轨中的控制音符和不包含旋律信息的音轨,通过归并到文该文件中的具有音高柱状图特征的音符集,从而提取出主旋律。通过与人工标识结果的实验进行比较,表明该提取主旋律方法的准确性。

关键词:MIDI 解析;多音轨音乐;主旋律提取;分层次聚集法
DOI:10.3778/j.issn.1002-8331.2009.26.070 **文章编号:**1002-8331(2009)26-0233-03 **文献标识码:**A **中图分类号:**TP391

1 概述

当今多媒体技术的蓬勃发展,使音乐艺术和信息科学相互结合产物的计算机音乐成为“计算机科学形成过程中最重要的推动力之一”[1]。音乐特征识别技术是感知音乐的前提性工作,而主题旋律的提取研究对于音乐特征的识别又具有至关重要的作用。音乐主旋律的提取是基于内容的音乐信息组织和检索的前提。影响旋律的因素主要有音高和音长。由于人感知到的旋律只是一种有意义的轮廓,它远远超过人对于单纯音高的感知[2],因此多采用记录相对音高的方法[3-5](如小二度为“1”)。从多音轨音乐中提取主旋律特征信息具有很大的难度,故一些研究者直接提取单音轨音乐文件,由于现有大部分的 MIDI 均为多音轨文件,主要针对的也正是研究多音轨 MIDI 的处理。

常规的主旋律提取方法,如文献[6]从复合音文件的片断中计算出相似矩阵,然后对这些矩阵采用自下而上的方法来提取音乐特征模式;如文献[7]针对多音轨音乐中与旋律特征相关的音轨特征量研究出发,建立了旋律音轨的模型。这些方法在提取主旋律过程中,多数都忽视了主旋律分布在乐器音轨(除第 10 号通道分配给打击乐器使用外,主音轨多半不会占用此通道)或音高较弱部分的情况。提出了一种基于分层次聚类的 MIDI 音乐主旋律提取方法,首先解析 MIDI 音乐文件,再去除每一音轨中的控制音符和不包含旋律信息的音轨,然后归并具有音高柱状图特征的音符集进文该文件。根据轮廓算法(Skyline Algorithm)的思想确保聚集在一个音轨中的音符都具有音高柱状图特征。最后选择音符序列最高的音高线视为主旋律。在实验与评价部分,图示了所提取出的主旋律的准确性。本研究具有一定的创新性,对于提升数字媒体和数字娱乐产品的情感交互能力、推进情感化人机交互的研究工作具有重要的意义。

2 定义和相关工作

2.1 MIDI 音乐文件概述

MIDI(Music Instrument Digital Interface)是音乐信号在电子乐器之间传输的标准,包括硬件接口标准以及电子音乐信号在不同硬件之间的异步串行传输协议。由于 MIDI 格式的音乐文件记录了音乐的全部乐谱和演奏的全过程,很多音乐的基本特征可以直接提取出来。图 1 所示的是一个 MIDI 文件的格式。

基金项目:辽宁省教育厅科研项目(the Scientific Research Project of Educational Office of Liaoning Province under Grant No.20060040);大连市科技基金项目(the Science Foundation Project of Dalian City under Grant No.2004-166-1)。
作者简介:冯国杰(1983-),男,硕士,研究方向是人工智能;王吉军(1964-),男,教授,硕士生导师,研究方向是人工智能、计算机动画。
收稿日期:2008-05-20 **修回日期:**2008-08-07

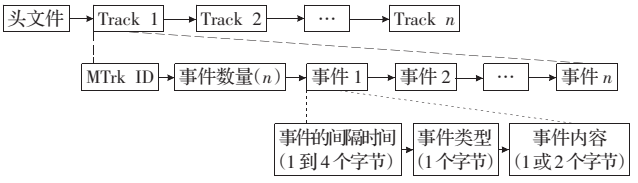

图 1 一个 MIDI 的文件格式

按照音乐理论[8],音乐中使用的、有固定音高的音的总和,叫做乐音体系。乐音体系中的各音叫做音级。两个音级在音高上的相互关系叫做音程。在 MIDI 文件中,每一类特征都是用 0~127 之间的数字描述。例如对于音高的编码,60 代表 256 Hz 的 C,音高相差小二度则编码相差 1。

假定 M 是一首 MIDI 音乐文件的音轨数。则有:$M=\{C_1,C_2,\cdots,C_i\}$,这里 $1\leqslant i\leqslant 16$;对于每个音轨,C_i 是一个集合,包含 k 个音符。则有:$C_i=\{n_{i1},n_{i2},\cdots,n_{ik}\}$;在音乐认知领域,音符是最小的认知单位。采用一个三元组来定义 MIDI 中的每个音符。

定义(音符) 包含音高、时值、音强三个属性。则有三元组:$n_{ij}=\{p_{ij},s_{ij},e_{ij}\}$,这里下标 i 表示第 i 个音轨,下标 j 表示第 j 个音符,p_{ij},s_{ij} 和 e_{ij} 分别代表音高、音强和时值,且有 $1\leqslant p_{ij}\leqslant 128$,音强和时值采用可变长表示方法,其定义域为实数。MIDI 音符是根据时值挑选出来的。因此,$\forall n_{ij},n_{i(j+1)}\in c_i;s_{ij}\leqslant s_{i(j+1)}$。然而,音乐这样表示并没有暴露出受持续时间约束。如果一个音符的持续时间 e_{ij} 是大于连续音符的开始时间,那么多音就产生了。复音的正式定义如下:$\exists n_{ij},n_{i(j+1)}\in c_i,e_{ij}>s_{i(j+1)}$。

2.2 音程统计

旋律是音乐的灵魂,而音程则是旋律运动最基本的要素[9]。音程利用音与音之间高低上的差异创造变幻莫测的旋律组合,使人们感知到不同的音乐形象与思想情感。进行主旋律的第一步就是对乐曲的每一个音轨进行旋律音程的统计,如图 2 所示的是所做的音程统计的一个实例,以中国古典名曲“花好月圆”为统计对象。

对第 5 个音轨进行统计:

Interval	0	1	2	3	4	5	6	7
Count	990	29	151	168	302	136	28	38
Percent	51.16%	1.50%	7.80%	8.68%	15.61%	7.03%	1.45%	1.96%

图 2 音程统计的一个实例

其软件设计命名中,Pitches 是从 MIDI 文件中提取的音高序列。和弦在 MIDI 文件中由一个音长不为 0 和多个长度为 0 的音共同表示。遇到这种情况,取和弦中的最高音参与统计。Interval 是相邻两个音符的音程差的绝对值,根据如下公式计算获取:

$$Interval\_ToneSat_i=\begin{cases}|Interval_i|, & |Interval_i|<25\\ 25, & |Interval_i|\geqslant 25\end{cases}$$

设数组 StatisticArray 存放音程统计结果,共有 26 个单元,可统计的音程跨度为两个八度,即音高差为 0~24,超过两个八度的音程都记入下标为 25 的数组单元。由前述的音程统计特性可知,进行这样的处理不会影响后继的主旋律定位判断。

2.3 修订的轮廓线算法

众多研究者实验证明,从复调音乐中提取主旋律的一个主要方法就是轮廓线算法(Skyline Algorithm)[10]。算法的主要思想是聚集所有音符到一个音轨并选择最高音高的音符。虽然轮廓线能得出很好的结果,但有人声称,首先,对音符的持续时间的操作可能改变旋律。接着,聚集所有音符到一个音轨中并除去那些可能继续隐藏在旋律中的无声间隔。第三,伴奏音符可能高频出现。修订了轮廓线法,提议保存音符的原始持续时间作为轮廓算法的一个恢复参数,如的算法 1 所示。

算法 1

1. 赋 $i=1;j=1$;
2. 对于每一个 $n_{ij}\in M$ 就有
3. $k=j+1$;
4. while($s_{ij}==s_{ik}$)
5. if($p_{ij}<p_{ik}$)
6. 去除掉 p_{ij}
7. $j=k$;
8. else
9. 去除掉 p_{ik}
10. $k=k+1$;
11. if $e_{ij}>s_{ik}$ then $e_{ij}=s_{ik}$;
12. $j=k$;
13. 重复语句 2

假设多数音符拥有相同的开始时间,音符保存着最大频度音高,尽管剩余音符都舍去(见语句 4~10)。第二,它可能缩短音符的持续时间。在 11~12 语句,当一个新的音符在不同的开始时间出现时,现存的持续音符将被缩短。从音符 n_{ij} 开始,伴随最小的 p_{ij},每一个音符,都用来测试它对复音的贡献度。如果至少有 50%的在 s_{ij} 和 e_{ij} 持续时间是复音,然后 n_{ij} 将从集合中去除。

3 MIDI 音乐主旋律的提取

音乐的多复调特性决定了音符的同时发音。许多研究者提出打击乐器是从来不作为用于主旋律的。在一般情况下,他们去除打击乐器的音轨不仅能增强搜索的相关性,而且加快了提取时间。运用常规方法也能获得每一 MIDI 音轨的轮廓线输出,并计算出平均音频度和每一音轨的平均信息量,再把拥有最大音调平均值或平均信息量的音轨视为主旋律音轨。但在文献[11]中指出,这些音轨选择算法极大地依赖音乐的风格。如果察觉到的主旋律是分布在各音轨中,选择其中一个主旋律音轨将导致旋律信息的丢失。提出基于选择所有包含旋律的音轨的方法来提取主旋律,称之为基于分层次聚类算法的选择最佳 K 旋律音轨。

3.1 预先操作

欲选择 k 个音轨来提取主旋律,这里 $k\geqslant 1$。为了决定一个合适的 k 值,利用 MIDI 的每一个音轨的音高柱状图特征。作为音高柱状图的结果,能计算出每一音轨和完整集合之间的乐谱不同点,从而聚集在 MIDI 音轨中。提出的 k 就等于聚集的全部总合。由于每一 MIDI 和它的音轨都有它特有的柱状图特性,全部被选的 k 是不确定的。执行三个音高柱状图的预先操作处理[12]。首先,去掉打击乐器的音轨 c_{10}。第二,对所有音轨应用轮廓线算法。第三,每一 MIDI 音符都分成十二个半音音程的律制,称十二平均律。为了做到这三点,计算音高以 12 为模的等价物。结果,音轨的音高柱状图成为一个 12 维空间中的点。

假设 h_i 表示音高柱状图的 C_i 的类,然后柱状图集 H 表示如下:$H=\{h_1,h_2,\cdots,h_{16}\}$;自从定义十二平均律的音乐,每一 h_i

也形式如下：$h_i=\{h_{i1},h_{i2},\cdots,h_{i12}\}$。设 T 是音轨数，它至少包含一个音符，然后第 i 维的标准值 $\bar{h}_i$ 定义如下：$\bar{h}_i=\frac{\sum_{k=1}^{16}h_{ki}}{T}$。相对地，完整音轨的平均柱状图 $\bar{h}$ 将定义为：$\bar{h}=\{\bar{h}_1,\bar{h}_2,\cdots,\bar{h}_{12}\}$。

在该方法中，计算每一音轨和 $\bar{h}_i$ 之间的不同点，设定一个函数 $D(h_i,\bar{h}_i)$ 用来计算 h_i 与 $\bar{h}_i$ 之间的欧氏距离或不同点，且 $d_i=D(h_i,\bar{h}_i)$。目标是保存音轨距离在一个新的集合中，形式有：$DS=\{d_1,d_2,\cdots,d_{16}\}$。而且已知更短的 d_i 导致一个音轨同整个 MIDI 音乐更相似。

3.2　分层次聚类算法

算法提出是基于音轨距离建立的一个音轨的聚集。因此，每一聚集包含特殊柱状图和度量距离。随后，此法从保存基本音高信息和距离特征的聚集中选择一个最佳音轨。分层次聚集法能够选择包含长音程距离的主旋律音轨，同时，它保存那些短音程的主旋律音轨，它集合最近两音阶的相邻聚集。

定义如下：

(1)类中心间距：$d_1=\|M_i-M_j\|$，其中 $M_i=\frac{1}{n_i}\sum_{Z\in C_i}Z$，$n_i$ 是属于 C_i 的样本数。

(2)靠得最近的样本：$d_2=\min_{Z_i\in C_i,Z_j\in C_j}\|Z_i-Z_j\|$

(3)类间平均距离：$d_3=\frac{1}{n_in_j}\sum_{Z_i\in C_i}\sum_{Z_j\in C_j}\|Z_i-Z_j\|$

设 $\overline{h_w}$ 是一个音乐 M 的平均音调柱状图的加权平均值，有 $\overline{h_w}=(\overline{h_{w1}},\overline{h_{w2}},\cdots,\overline{h_{w128}})$，这里 $\overline{h_{wi}}=\sum_{k=1}^{16}h_{ki}f_i$。时值 t 是由音高柱状图的加权平均值 $\bar{h}$ 和平均音调的加权平均值 $\overline{h_w}$ 的距离和计算得来的，有：$t=\frac{d(\overline{h_w},\overline{h})}{2}$。

在音轨聚集完之后，发现每一聚集的最佳音轨。实现此目标，引入一种新的从文本中结合选择的音轨选择算法[13]，它能得出更好的结果。假设，a_i 和 b_i 分别是音高平均值和 C_i 的预测熵。结合选择计算 x_i 标准有：$x_i=a_i+b_i\times128$。因此，拥有最大 x_i 的音轨，可以看成是主旋律。音高平均值 a_i 范围在 1~128，而预测熵 b_i 的范围在 0~1 之间。还是以“花好月圆”为实例阐述运用分层次聚类算法，见表 1 所示。表的第 1 栏显示按距离栏 d_i 排序的音轨 IDs；第 2 栏显示此音轨是否包含旋律信息；在第 4 栏记录着当前和先前音轨适宜距离的不同。如果连续行之间有长的阶跃，它将认为是一个新的聚集的边界。因此，在这个例子中的音轨将分解成三个基本的聚集。表 1 中的 ID 聚集栏显示通过分层次聚集方法生成的聚集。所有音轨的 x_i 值是由结合选择法来计算的，由于 C_6，C_2 和 C_{13} 包含最大 x_i 值在各自的聚集中，最后将它们作为一个典型音轨也选择上了。

基于分层次聚类法提取主旋律工作分为如下 6 个基本步骤：

(1)对所有音轨运用轮廓线算法，实现每一音轨在去除控制音符之后的聚集，是具有音高柱状图特征的音符集。

(2)$\forall C_i\in M$，计算 a_i，b_i 和 x_i；

(3)用十二平均律来描述音乐音符，计算音高柱状图集 H。

(4)基于音程来实现聚集的集合。

(5)选择旋律音轨并去除其余的音轨聚集。

(6)再次应用轮廓线算法，得出最高的音高轮廓线。

表 1　“花好月圆”重要音轨的距离特征

音轨号（IDs）	包含旋律	欧氏距离（d_i）	乐器名称	聚集号（ID）	x_i	选择为旋律音轨
C_1	否	0.089 1	八孔竖笛	1	49.3	否
C_6	是	0.093 5	原声吉他	1	137.4	是
C_5	否	0.097 7	原声吉他	1	112.5	否
C_7	否	0.207 1	原声吉他	1	119.1	否
C_2	是	0.113 0	长笛（横笛）	1	98.6	是
C_{13}	是	0.100 9	大提琴	2	104.8	是
C_9	否	0.088 5	定音鼓	3	81.4	否
C_4	否	0.215 4	古筝	3	99.2	否
C_{11}	否	0.177 3	弦乐合奏	3	76.3	否

4　实例分析与评价

为了验证该方法的有效性，在测试评估实验中使用了一个从网上随机得到的多音轨 MIDI 文件的乐曲库，乐曲包括古典、流行、乡村音乐、爵士乐等来自不同国家的风格音乐，重点包含同人类情感紧密相关的治疗音乐，共有 257 首。图 3 为基于分层次聚类算法提取出的“花好月圆.mid”的主旋律。

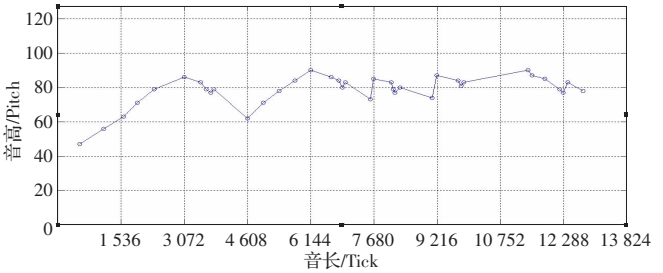

图 3　“花好月圆”主旋律图片段

其中横坐标为音长，单位为 Tick；纵坐标为音高信息。得出的主旋律图与乐曲实际主旋律相似度高达 97%，误差主要发生在 12 音轨的合成弦乐合奏部分，这也为今后的工作提供了一个方向。在提取过程中，发生了多处多音轨旋律合成的琶音，例如在 10 773Tick（大约 1′53″左右），就是多个旋律音轨（原声吉他、长笛（横笛）、大提琴）共同作用的。

在人工标定评价的实验中，通过解析当前乐曲的主旋律，与实际的轮廓线进行比较相似度，得出当乐曲主音轨为两条时的成功率达 96%，当主旋律分布在三条音轨时的成功率达到 98%。同普通的旋律提取算法相比较，具有明显的优越性，准确度也有所提高。

5　总结

提出了一种基于分层次聚类算法的多音轨 MIDI 音乐主旋律提取方法，并用数学模型解析计算了旋律音轨的聚类间距离，提出了主旋律提取方法流程。与目前现有的提取方法研究相比，结合软件实现的创新中，本系统在减小主旋律分布在乐器音轨或音高较弱部分所产生的提取误差的同时，对提取准确性上有所改进。在建立音符聚集方面，采用软件实现算法提取工作，尽可能不遗漏地筛选出包含具有音高柱状图特征的音符

（下转 239 页）