第 33 卷 第 2 期 计 算 机 工 程 2007 年 1 月
Vol.33 No.2 Computer Engineering January 2007
·人工智能及识别技术· 文章编号：1000—3428(2007)02—0165—03 文献标识码：A 中图分类号：TP391.42

基于音轨特征量的多音轨 MIDI 主旋律抽取方法

赵 芳，吴亚栋，宿继奎
(上海交通大学电子信息学院，上海 200030)

摘 要：在基于内容的数字音乐检索研究中，其音乐库大都直接使用复合音音乐数据文件。然而这种直接采用复合音音乐数据的索引必将给检索处理带来巨大的计算量以及复杂的匹配算法。该文提出了一种基于音轨特征量的多音轨 MIDI 主旋律信息音轨抽取方法。通过与人工标注结果的实验比较，表明该文实现的抽取方法可有效地从多音轨 MIDI 演奏数据文件中提取出主旋律音轨。

关键词：哼唱检索；复合音音乐；MIDI 格式；主旋律提取

Melody Extraction Method from Polyphonic MIDI Based on Melodic Features

ZHAO Fang, WU Yadong, SU Jikui
(School of Electronic Information Engineering, Shanghai Jiaotong University, Shanghai 200030)

【Abstract】 Much of the work on content based music query focuses on polyphonic music files. However the polyphonic format causes large calculation and complex algorithms for matching. This paper represents a model for auto melody extraction method for multi-track MIDI files based on melodic features. The results are compared with manual output and show that the implemented method can extract melody track from multi-track MIDI file effectively.

【Key words】 QBH; Polyphonic music; MIDI format; Melody extraction

1 概述

一般来说，QBH系统都是基于乐曲的旋律内容特征来进行乐曲检索的，除了在小规模乐库有采用单音轨MID数据格式之外[1]，大都直接采用复合音音乐数据格式文件(音频波形文件wav、mp3 或演奏数据文件多音轨MIDI)。这主要是由于从复合音音乐中提取主旋律特征信息具有很大的难度，一些研究者放弃主旋律提取处理手段而直接以复合音音乐文件为整体处理对象，如文献[2]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，以提供检索索引；文献[3]针对复合音音乐的音乐特征提取和音乐数据库的管理与检索，提出了一种基于特征字典来抽取复合音音乐中的重复模式的方法。这种直接采用复合音音乐数据格式作为基于内容的大规模音乐检索用索引将会给检索处理带来巨大的计算量以及复杂的匹配算法。

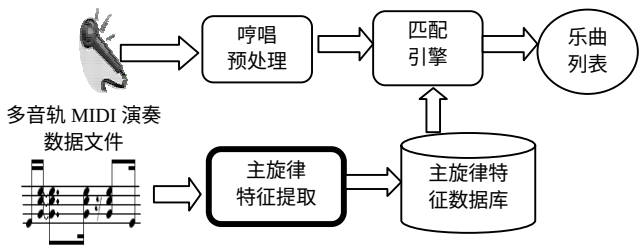


图 1 哼唱检索系统工作流程

针对多音轨 MIDI 演奏数据文件，如何提取出其中的主旋律音轨，文献[5]提出了以下方法：(1)音轨元数据法：如果记录了表征旋律音轨的名称，那么就直接抽取出该音轨数据；(2)多音轨保留法：保留所有音轨数据，检索时对所有音轨进行相同的匹配处理；(3)多音轨合并法：将所有音轨数据合并到一条音轨中。然而，在实际应用中，上述方法并不十分理想。

图 1 给出了我们的基于多音轨 MIDI 旋律提取的 QBH 实验系统的工作流程。

2 与旋律特征信息相关的音轨特征量

标准 MIDI 文件(SMF)是用于存储 MIDI 演奏数据的文件格式。对于主旋律和伴奏的研究已经长达多年，但是这方面的研究成果很少能被应用到对数字音乐的特征提取处理中，例如对多音轨 MIDI 音乐的旋律信息提取。为探讨如何从多音轨 MIDI 演奏数据中有效地提取出与主旋律相关的音轨信息，以下就一些与旋律特征信息相关的、并可从 MIDI 演奏数据中获取的音轨特征量进行分析讨论。

2.1 音轨特征量

2.1.1 音轨名称(F_1)

对于 MIDI 数据文件中的每一个音轨，通常都会有一个音轨名称标注字段，但并没有规定音乐作者必须为自己的 MIDI 作品的每一个音轨都给出其名称。即有的作者会在该音轨的音轨名称字段给出标注。因此该音轨名称标注字段的内容为提取主旋律音轨提供了有益的信息，并可作为表征旋律音轨的特征量来使用。音轨名称可由以下事件格式中获取“FF 03 length text”。该事件格式中的“text”给出了音序器或是音轨的名称，通常被放置在音轨的开头。

基金项目：国家自然科学基金资助项目(60473041)；上海科委重大攻关项目(045115016)
作者简介：赵 芳(198－)，女，硕士生，主研方向：语音信息处理，旋律提取；吴亚栋，博士、副教授；宿继奎，硕士生
收稿日期：2006-01-25 **E-mail：**aka9103@sjtu.edu.cn

在获取了音轨名称之后，就可以使用一个子串搜索来查看该名称并判断出该音轨是否为主旋律音规。我们通过实验观察建立了 2 个名称列表，其中一个列表是用来表征主旋律名称的，而另一个则用来表征通常伴奏的。举例来说，在主旋律名称列表中会有“ MELODIES ”,“ VOCAL ”, “ SING ”, “ SOLO ”,“ LEAD ”,“ VOICE ”等关键词；而在伴奏列表中，则会有“ ACCU ”,“ DRUM ”,“ BASS ”,“ PERCUSSION ”, “ COMPANION ”,“ BACK ”等关键词。由于作曲者也可以使用其它的名称来标记每个音轨的功能，因此这 2 张列表的关键词将随着 MIDI 数据的大量采集而不断更新。

2.1.2 通道号(F_2)

所有的音轨消息中都带有一个比特来记录将本音轨的演奏数据送往哪个演奏通道。MIDI有 16 个通道，在多音轨MIDI演奏数据文件中，多个通道将被用于同时演奏。由于第 10 号通道一般保留用作打击乐器或是鼓，因此可以将第 10 号通道的音轨从主旋律的候选音轨名单上去除[5]。通道号特征量的检测比较简单，所以可以将它和音轨名称特征量组成一个基于音轨元数据的过滤器。

2.1.3 左右声道平衡度(X_1)

每个音轨都记录了一个名为Pan的特征值，用来表示该音轨的音符在左右声道中音量大小的比例。通常可以观察到旋律音轨的发音基本处在左右声道音量平衡位置。因此，根据Pan特征值偏离平衡位置值的大小，大致上可以判断该音轨是否为主旋律音轨。Pan值可由消息“ 0xBn 0A value ”获取。该事件中的“ n ”表示通道号，并使用“ value ”来记录每一个音轨的Pan值，取值范围从 0~127，其中 64 代表左右声道平衡，如果该值相对较高或较低都不太可能是旋律音轨。因此，可以用以下公式来定义音轨k的左右声道平衡度$X_1(k)$：

$$X_1(k)=Pan(k)-64/127，k=1, 2, …, 16 \quad (1)$$

2.1.4 平均力度(X_2)

通常情况下，在一个音乐作品中，主旋律声部的全频域平均能量要比其它声部要高。因此，可以为每个音轨定义一个名为平均力度的特征量X_2来比较各个音轨的能量大小。

$$X_2(k)=\sum_{i=1}^{N} Vel(k,i)/N,k=1,2,…,16 \quad (2)$$

式中 k 表示音轨号，N 表示第 k 音轨的音符数，Vel(k, i)表示第 k 音轨中第 i 音符的按键力度值(Velocity)，其取值范围从 0~127。然而，在统计时，只有那些力度值为非零的“ 音符开 ”事件被计算在内并进行排序。在同一音轨，如出现同时发音的多个音符，则计算时取音高最高音符的力度值而删除其余同时发音音符的力度值。表 1 给出了同时发音音符的例子。

表 1 同时发音和交叉发音的音符例

音符序号	音高	开始 Tick	结束 Tick	持续时间	力度值
1	63	2 499	2 518	19	70
2	65	2 499	2 518	19	66
3	63	2 518	2 537	19	70
4	65	2 529	2 547	18	66

2.1.5 主音量(X_3)

基于和 2.1.4 节同样的原因，每个音轨都记录了一个名为主音量的特征值。假设除了主音量特征，其它的特征值都相同，那么较好的办法是选取主音量较高的音轨作为主旋律。主音量值可由“ 0xBn 07 size ”获取。在该事件格式中，“ n ”表示通道号码，“ size ”记录该通道的主音量值，其范围在 0~127 之间。

2.1.6 发音时间(X_4)

音轨音声消息通过在音符开/关之间相互切换来控制 MIDI 乐器的声音，并且还可以改变响度和音高等。该事件的格式如下所示：(1)音符开“ 0x9n note velocity ”；(2)音符关“ 0x8n note velocity ”。

在多音轨 MIDI 演奏数据文件中，每当多个音符在同一音轨中同时发音时，只计算其中一个音高较高的音符的发音时间。这是因为人们通常在同一时刻对音高较高的音符较为容易感知，而该音符通常是在主旋律线上的音符。

对于表 1 中同时发音的音符 1 和音符 2，只计算一次发音时间。而交叉发音的音符 3 和音符 4，有几种不同的方法来处理总发音时间。在本方法中，采用从后一个音符的结束时间减去前一音符的开始时间来表示总发音时间，而不是直接将两个音符的持续时间相加来计算，如式(3)所示：

$$X_4(k)=\sum_{i=1}^{N-1} [note(i+1).end-note(i).start],k=1,2,…,16 \quad (3)$$

式中，k表示音轨号，N表示第k音轨的音符数，note(i)表示音符i，note(i).end表示该音符的关时刻，note(i).start表示该音符的开时刻。

2.1.7 发音面积(X_5)

在选取旋律特征时，不仅音符的持续时间(音长)起到重要作用，而且音高信息往往也是关键的因素。这样，可以将每个音轨的这两个因素结合在一起考虑，即用发音面积的概念来表征音高和音长的综合效果。音符发音面积可用音高与音长的乘积来定义，而音轨的发音面积则可以将其定义成每个音符的发音面积之和，如式(4)所示：

$$X_6(k)=\sum_{i=1}^{N} [note(i).pitch*note(i).duration] \ k=1,2,…,16 \quad (4)$$

式中，note(i)表示音符 i，note(i).pitch 表示该音符的音高值(半音值)，而 note(i).duration 表示该音符的音长(此处为本文定义音长)。同时发音的音符，可以取音高较高的值乘以发音时间来计算发音面积。但是对于交叉发音的音符，如表 1 中的音符 3 和音符 4，处理方法有所不同。

通过人们的经验和前期研究，认为相对于音符的结束时间，人们对于音符的开始时间更敏感。所以，在实际系统中，表 1 中的交叉发音的音符 3 和音符 4 面积的计算如下：

发音面积 3＝note(3).pitch*(note(4).start -note(3).start)

发音面积 4＝note(4).pitch*(note(4).duration)

2.2 基于旋律特征关联性的音轨特征量的贡献度评估

为了考察上述音轨特征量(X_1～X_5)用于描述旋律特征时的有效性，按如下假设并采用统计方法定义音轨特征量的对旋律特征描写的贡献度。对同一首多音轨MIDI乐曲中的 16 个音轨的同一名称音轨特征量Xm(k)的最大值所对应的音轨为主旋律音轨K_m，即

$$K_m = \text{argmax}_{1<=k<=16}\{X_m(k)\} \quad (5)$$

并通过多音轨 MIDI 文件数为 T 的测试集获得该测试集中符合该假定的 MIDI 文件数 C，然后用 C/T 的百分比值来定义该音轨特征量 Xm 对旋律特征的实际贡献度。即用 W(m)表示 Xm 的贡献度，则 W(m)如式(6)所示：

$$W(m)=C/T,m=1,2,…,5 \quad (6)$$

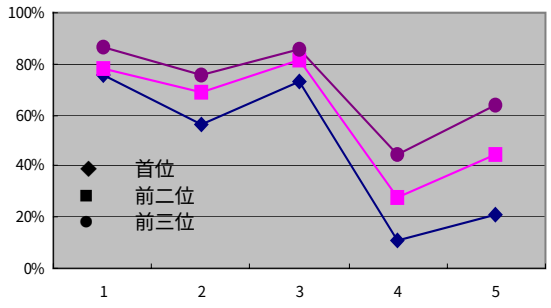
若定义由旋律抽取假设得到的旋律音轨号K^j_m(j=1,2,…,T; m=1,2,…6)与人工标注旋律音轨号R_j的(j=1,2…,T)比较函数：

$$Com(K_m,Rj)=\begin{cases}1 & K^j_m=Rj\\ 0 & K^j_m\neq Rj\end{cases}$$

则一个音轨特征量的贡献度 W(m)就由下式来表示：

$$W(m)=\sum_{j=1}^{T} Com\,(K^j_m, Rj)\ /T$$

W(m)依赖于测试集的大小，测试集一般尽可能取大。图 2 给出了上述 5 个音轨特征量(X_1~X_5)的贡献度测试实例，该测试集T的大小为 310 首多音轨MIDI乐曲。如图 2 所示，如果取音轨排序后的第 1 位音轨作为旋律音轨，那么特征量X_4和X_5的贡献度不是很高，但是从结果可以看到，若取音轨排序后的前 2 位或前 3 位音轨作为候补结果的话，这两个特征量还起到了比较明显的作用。



(横轴为音轨特征量序号，纵轴为贡献度)

图 2 音轨特征量X_1～X_5的贡献度

3 基于音轨特征量线性加权的旋律音轨提取模型

为综合利用音轨的各特征量来描述该音轨为旋律音轨的可能性程度，定义同一 MIDI 第 k 音轨的得分函数 Y(k)如下：

$$Y(k)=\sum_{m=1}^{5} \alpha_m X_m(k) \qquad k=1,2,\ldots,16 \qquad (7)$$

式中，α_m为由式(8)给出的音轨特征量X_m的权重，$X_m(k)$为由式(9)给出的规整化后的音轨特征量。

$$\alpha_m= W(m)/\sum_{m=1}^{5} W(m) \qquad (8)$$

$$X_m(k)=X'_m(k)/X'_m(max)，k=1,2,\ldots,16 \ ；m=1,2,\ldots,5 \qquad (9)$$

$X'_m(k)$：原始第k音轨中的第m特征量

$X'_m(max)$：同一MIDI文件所有音轨(16)中第m特征量的最大值。表 2 给出了一首著名的乐曲“ Carmen.mid ”主旋律抽取实例。

表 2 示例乐曲的音轨特征量

音轨 / 特征量	1	2	3	4	5	6
名称F_1	Bass	Drums	Piano	Guitar	Strings	Melody
通道号F_2	2	10	8	5	6	4
PanX_1	0	0	0	0	0	0
平均力度X_2	0.73	0.77	0.82	0.70	0.36	1
主音量X_3	0.71	0.71	0.63	0.55	0.63	0.87
发音时间X_4	0.89	0. 50	0.92	0.35	0.58	1
发音面积X_5	0.28	0.15	0.59	0.17	0.29	1
音轨得分 Y(m)	0.069	-0.021	0.35	0.27	0.24	0.74

当多音轨文件转成单音轨文件时，还要做一些额外的工作。不可以直接删除除了旋律音轨的其它所有音轨，因为对于格式为多音轨的 MIDI 文件来说，像节拍和时间标记等信息是存放在通常不发音的第一个音轨中的。在删除之前，所有这些有用的信息都要转移到旋律音轨中去，这样才不会缺失演奏的细节。

4 实验结果

为了验证本方法的有效性，我们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，乐曲类型包括流行、摇滚、爵士、民歌、乡村音乐等来自不同国家的各种音乐，共 310 首。图 4 给出了音轨排序后的结果取至前 3 位为止的、前 N 位(按音轨得分排序)音轨中含有主旋律音轨的百分比率。显然，随着前 N 位的候选数不断增大，抽取的成功率也就越高。

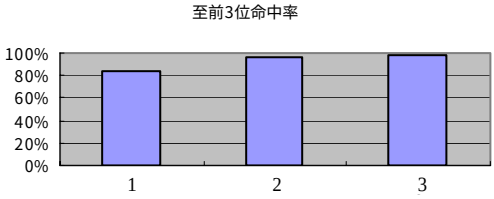


图 4 前 3 位抽取成功率

如图 4 所示，在本实验中，候选数为一条音轨作为旋律音轨时的成功率为 83%，当候选数为两条音轨的成功率达到 96%，而前 3 位的成功率为 98%。

5 结论

本文提出了一种基于音轨特征量的多音轨MIDI音乐主旋律的抽取方法，并在此基础上实现了一个多音轨MIDI音乐的主旋律提取系统原型。与已有的研究相比[5]，本系统在音轨特征量选取及各特征量的综合利用方式上均有所不同。在音轨特征量选取方面，加入了尽可能多的与主旋律特征信息相关的特征量。在建立音轨评估得分模型时，通过合理选取各特征量的权重，使得每个音轨得分结果可以反映出该音轨作为主旋律音轨的相对重要性，从而通过得分排序可以找出候选旋律音轨。实验结果是较为令人满意的，这表明了本方法的有效性及其应用于大型数字音乐检索的可行性。

参考文献

1 Wu Yadong, Li Yang, LIU Baolong. A New Method for Approximate Melody Matching[C]. Proceedings of IEEE International Conference on Machine Learning and Cybernetics, Xi'an, 2003-11: 2687-2691.

2 Meudic B, St-James E. Automatic Extraction of Approximate Repetitions in Polyphonic Midi Files Based on Perceptive Criteria[C]. Proc. of Internation Symposium on Computer Music Modeling and Retreval, 2003.

3 Shih H H, Narayanan S S, Kuo C C J. Automatic Main Melody Extraction from Midi Files with A Modified Lempel-ZIV Algorithm [C]. Proc. of International Symposium on Intelligent Multimedia, Video & Speech Processing, 2001-05.

4 Lu Lie, You Hong, Zhang Hongjiang: A New Approach to Query by Humming in Music Retrieval[C]. Proc. of IEEE International Confernce on Multimedia and Expo., 2001.

5 Tang M, Yip C L, Kao B. Selection of Melody Lines for Music Databases[C]. Proceedings of the 24th IEEE Annual International Computer Software and Applications Conference, 2000: 243-248.

6 Nagano H, Kashino K, Murase H. Fast Music Retrieval Using Polyphonic Binary Feature Vectors[C]. Proc. of IEEE International Conference on Multimedia and Expo., 2002: 101-104.

7 李　扬，吴亚栋，刘宝龙. 一种新的近似旋律匹配方法及其在哼唱检索系统中的应用[J]. 计算机研究与发展，2003，40(11): 1554-1560.