

文章编号:1001-9081(2009)03-0768-03

基于 WAV 文件的独奏乐曲信号中的乐谱识别

曹西征,刘春红,孙 林

(河南师范大学 计算机与信息技术学院,河南 新乡 453007)

(caoxizheng@126.com)

摘 要:大多数乐谱识别工作都是在 MIDI 格式文件下进行的,而音乐数据大多采用音频格式存储,为此讨论一种从 WAV 音频文件的独奏乐曲信号中识别出乐谱的方法。采用离散余弦变换(DCT)结合能量比对的手段,并提出了一种变化音最少法则:以十六分音符对应的时值作为步长,将乐曲信号分为若干个子段;对每个子段进行一维 DCT 操作,提取基音频率;对相邻子段进行能量比对,查找音符的端点;利用变化音最少法则,生成乐谱。实验表明此方法具有良好的识别效果。

关键词:乐谱识别;WAV 文件;离散余弦变换;基音频率

中图分类号:TP391.4 **文献标志码:**A

Musical score recognition of solo music signal based on WAV file

CAO Xi-zheng, LIU Chun-hong, SUN Lin

(College of Computer and Information Technology, Henan Normal University, Xinxiang Henan 453007, China)

Abstract: Most work about musical score recognition is based on MIDI format file; however, most musical data are stored in audio frequency format. Therefore, this paper discussed a musical score recognition method from WAV format file including solo music signal, adopted the Discrete Cosine Transform (DCT) combined with energy, and presented a principle of least derived notes. The music signal was divided into several sub-sections and the length of each section was equal to the duration of one semiquaver. The base frequency was obtained using one dimension DCT operated on each section. The extreme points of the musical notes were found by comparison with the adjacent sub-sections. And then, the musical scores were got using the principle of least derived notes. The experiments show that this method is very effective.

Key words: musical score recognition; WAV file; Discrete Cosine Transform (DCT); base frequency

0 引言

一首音乐的乐谱几乎记录着该音乐的全部信息,音乐家通过阅读乐谱就可以了解音乐的全貌。乐谱自动识别是计算机音乐分析与理解中的基础工作,是指将音乐演奏的声学信号所对应的乐谱信息自动翻译出来,它是音乐分析领域中的一项十分重要同时也是十分困难的工作^[1]。

从音频文件中识别出乐谱的文献不是很多,Zhu 等提出了一种用于流行音乐和古典音乐的音调估计的新方法^[2]。Chang 等提出了一种基于循环神经网络的音高估计方法^[3],该方法减小了音高估计所需要的时间窗的长度因而可用于音高变化较快的信号,但是其计算量比较大。徐国庆等在乐音识别方面^[4-5]也做出了一些贡献。

由于 MIDI 格式音乐文件本身包含着大量的乐谱信息,所以目前基于 MIDI 格式的音乐分析工作做得比较深入。然而这仅仅是一种简单的逻辑转换,其实并不涉及音符识别的具体工作。而且它们一个共同的缺点就是对 MIDI 音乐文件格式的强烈依赖性,而大量的音乐数据采用一般音频格式存储,这制约了这些方法的应用。

本文探讨的乐谱自动识别在 WAV 格式的音频文件下进行,由于 WAV 音频文件和任何其他非 WAV 音频都可以相互转化,这就为乐谱的识别提供了更广阔的应用空间。而且

本文采用一种先分割后合并的间接时值检测方法,与其他直接时值检测方法相比,该方法相当于省略了计算音符持续时间的工作量。同时,进行了其他方法很少涉及到的拍号和调号的识别工作。

1 乐谱识别过程

一首视觉上的乐曲包含的信息主要有:拍号、调号、音符序列(包括音高和时值的流动)。乐谱识别需要完成的工作就是从音频文件中提取这些信息。概括地说,识别过程如下:

- 1) 将非 WAV 音频文件转化为 WAV 格式;
- 2) 以十六分音符对应的时值为时间步长,将乐曲分成若干个子段;
- 3) 对每个子段进行离散余弦变换(Discrete Cosine Transform, DCT),提取基音频率;
- 4) 对每个子段进行能量求和,比对相邻子段的能量,判断音符端点位置;
- 5) 比对相邻子段的基音频率,如果满足频率阈值条件,并且不是音符的端点,则合并子段,并累计时值;否则,不合并子段;
- 6) 将相邻四个子段合并,生成拍能量,进行能量比对,识别出拍号;
- 7) 根据变化音最少法则,识别出调号;

收稿日期:2008-09-25; **修回日期:**2008-11-06。 **基金项目:**国家自然科学基金资助项目(60873104);河南省科技攻关项目(0624220038;082102210107);河南师范大学引进博士科研启动基金(0716)。

作者简介:曹西征(1977-),男,山东莒县人,副教授,博士,主要研究方向:计算机音乐技术、多媒体技术;刘春红(1969-),女,河南新乡人,副教授,硕士,主要研究方向:网络多媒体技术;孙林(1979-),男,河南南阳人,助教,硕士,主要研究方向:多媒体数据挖掘。

8)输出乐谱。

为了书写方便,本文采用简谱形式表示乐谱。以下的介绍中都是以四分音符为一拍。下面将具体分析每个过程。

1.1 音频文件格式的转化

目前所有的音频播放软件和编辑软件都支持 WAV 格式^[6],任何其他的音频文件格式都可以转化为 WAV 格式。WAV 格式的音频文件非常灵活,对音频流的编码没有硬性规定。而且 WAV 格式编辑处理比较方便,因此在进行乐谱识别之前,可先将其他的音频文件转化为 WAV 格式。能够完成这一转化的软件非常多,例如 Cool Edit Pro、Sound forge、WaveLab 以及 AV2Wav 等。

1.2 WAV 音频的分段

音乐是一种与时间有关的媒体,乐谱中最主要的信息有两种:时值和音高,乐谱可以看成是在不同时间上出现不同音高的音符序列。因此按时间序列对音频进行分段是乐谱识别的基础工作。这里比较关键的问题是分段时如何选取时间的步长,这要归结到乐谱中时值的组织结构,也可称之为音的长短。

由于乐谱中音的长短是随时变化的,这就需要步长也应该同步变化,但这就有一个步长如何变化以及变化多少的问题,这往往是比较难以控制的。因此本文采用一种等步长的方法,取足够短的步长,将音频进行分段。如果相邻子段的音高相同,并且不是音符的端点,那就说明这是同一个音符,需要对音名进行合并,合并的相邻子段的个数决定于该音的长短,这样,就用合并等步长的尺度实现了音的长短变化。

由于音符中时值最短的是三十二分音符,因此为了使在进行 DCT 运算时不遗漏任何一个音高的变化,最佳步长应该为三十二分音符对应的时间。然而在乐谱中,三十二分音符是很少出现的,所以,为了减小计算量,本文采用十六分音符的时值作为步长,将其标记为 dT ,即时间分辨率为 dT 。如果以四分音符作为一拍,则 dT 为 $1/4$ 拍,那么对于每分钟为 100 拍的乐曲, $dT=0.15$ s。这样,乐曲就可以以 0.15 s 为步长分成若干个子段,继而进入对每段进行 DCT 的操作。

1.3 DCT 运算

设原乐曲采样序列为 $A(k)$,经过 DCT 后系数为 $A'(u)$,则变换公式^[7]为:

$$A'(0) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} A(k) \quad (1)$$

$$A'(\mu) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} A(k) \cos \left[(k+0.5) \frac{\mu\pi}{N} \right] \quad (2)$$

式(1)和(2)中, $A'(u)$ 为第 u 个余弦变换系数, u 为广义频率变量, $u=1,2,\dots,N-1$; $A(k)$ 为时域中 N 点序列, $k=0,1,2,\dots,N-1$ 。设 f 为频率变量, F_s 为采样频率,则 $f = \mu F_s / (2N)$ 。

通过对 $A(k)$ 序列进行 DCT 运算后,可以获取该区域的频率特性,提取基频,即音高对应的频率。下面说明基频提取的方法:

- 1) 从 $A'(u)$ 序列中,提取第一个极大值和极小值,即 $A'(u)_{1\max}$ 和 $A'(u)_{1\min}$,其对应的 u 分别为 $u_{1\max}$ 和 $u_{1\min}$;
- 2) 获取基频对应的广义频率,即 $u_b = (u_{1\max} + u_{1\min})/2$;
- 3) 获取基频,即 $f_b = u_b F_s / (2N)$ 。

由于在提取基音频率过程中,存在误差,因此对于同一音高,最终获取的频率可能不完全相同,为此,设置频率阈值 dF (本文在实验时取 10 Hz),当相邻两个子段的频率小于 dF 时,可视为同一音高。

1.4 频率与音名的映射

音高与频率是有直接联系的,为了将基频映射到音名中,需要建立映射表。以第一国际音高 $a^1 = 440$ Hz 为基准,根据相邻半音之间的频率之比为 $f_2/f_1 = 2^{1/12} = 1.06$,可建立基频和音名的映射,如表 1。例如, a^1 音与 $\sharp a^1$ 音相差半音,因此 $f(\sharp a^1) = f(a^1) * 2^{1/12} = 440 * 2^{1/12} = 466.2$ Hz。

表 1 中只列了常用的部分音名的基频,其他音名的基频可以导出。将利用 DCT 方法提取的基频与表 1 比对,就可以确定该基频对应的音名。

1.5 子段能量比对和合并

通过上面操作,并将音名翻译出来后,可以获取步长为 dT 的音名序列。如果相邻子段的音名不同,则此处一定为音符的间断;如果音名相同,则可能间断,也可能是同一音符。那么怎样判断同音名是否为断点,这里采用能量对比和合并的方法,下面以例子说明能量比对和合并的必要性。

表 1 基频和音名的映射表

音名	基频/Hz	音名	基频/Hz	音名	基频/Hz	音名	基频/Hz
c	130.8	a	220.0	$\sharp f^1$	370.0	$\sharp d^2$	622.3
$\sharp c$	138.6	$\sharp a$	233.1	g^1	392.0	e^2	659.3
d	146.8	b	246.9	$\sharp g^1$	415.3	f^2	698.5
$\sharp d$	155.6	c^1	261.6	a^1	440.0	$\sharp f^2$	740.0
e	164.8	$\sharp c^1$	277.2	$\sharp a^1$	466.2	g^2	784.0
f	174.6	d^1	293.7	b^1	493.9	$\sharp g^2$	830.6
$\sharp f$	185.0	$\sharp d^1$	311.1	c^2	523.3	a^2	880.0
g	196.0	e^1	329.6	$\sharp c^2$	554.4	$\sharp a^2$	932.3
$\sharp g$	207.7	f^1	349.2	d^2	587.3	b^2	987.8

例如翻译后的音名序列为表 2,其中 i 为子段变量。此序列中时值 $T=8dT$,即 8 个十六分音符,2 拍。 $i=6$ 和 7 之间一定为音符的间断。而其他的序列可以有多种解释,下面以调号为 $1=F$ 为例,说明此序列的几种解释。

表 2 音名序列

i	音名	i	音名	i	音名	i	音名
1	f^1	3	f^1	5	f^1	7	g^1
2	f^1	4	f^1	6	f^1	8	g^1

- 1) 不作合并处理,则简谱为 1111 1122;
- 2) 合并 1~6,7~8,则简谱为 1·2;
- 3) 合并 1~4,5~6,7~8,则简谱为 1 1 2;
- 4) 合并 1~2,3~4,5~6,7~8,则简谱为 1 1 1 2;
- 5) 合并 1~2,3~6,7~8,则简谱为 1 1 2。

当然,还可以做出更多的解释,但这种多解性往往不是我们希望看到的。因此,为了寻找唯一解,我们采用了能量的对比法。

一般来说,乐器弹奏一个音符后,乐音是一个衰减的过程。该音符的时值越长,则衰减的幅度越大;当弹奏新的音符时,乐音又达到极大值,然后又开始衰减。图 1 为某几个音符衰减特性的波形图,此波形为钢琴弹奏 321 的波形(1=C)。

从图 1 可以看出,新的音符产生时,振动幅度达到极大,然后衰减到极小。为了描述的方便,定义 dT 时间内的子段能量为式(3):

$$E(k) = \sum_{i=k}^{k+n} |A(i)| \quad (3)$$

式(3)中 $E(k)$ 是第 k 个子段的能量, n 为 dT 时间内的采样个

数, $n = dT * F_s$ 。

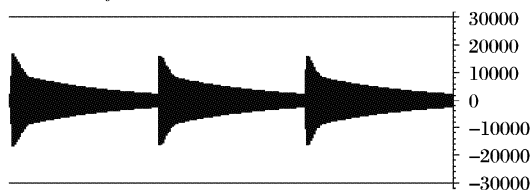


图1 波形衰减图

音符的间隔也就是 $E(k)$ 出现拐点的位置。设 $E(k)$ 和 $E(k+1)$ 为相邻的子段能量, 对它们进行比对, 则可通过以下的过程达到子段音符合并的目的。

1) 如果 $E(k) > E(k+1)$, 则 k 子段和 $k+1$ 子段为同一个音符, 需要将其子段对应的音名合并;

2) 如果 $E(k) < E(k+1)$, 则 $k+1$ 子段为新的音符, 不作合并处理。

这样, 就可以判断相邻子段中相同音高的两个音是否处在同一个音符, 即查到了音符的间断。

1.6 拍号的识别

将子段分组, 按照从前往后的顺序, 每四个子段为一组, 如表 3。

表 3 子段能量和拍能量

子段能量	拍能量
...	
$E(n)$	
$E(n+1)$	$E_b(n)$
$E(n+2)$	
$E(n+3)$	
$E(n+4)$	

表 3 中, n 是 4 的整数倍。第 n 拍能量 $E_b(n)$ 可表示为:

$$E_b(n) = \sum_{i=1}^4 E(n+i) \quad (4)$$

拍号的识别是基于强拍和弱拍之间的交替关系规律。例如, 对于 2/4 拍, 强弱规律为“强、弱”; 对于 4/4 拍, 强弱规律为“强、弱、次强、弱”; 对于 3/4 拍, 强弱规律为“强、弱、弱”, 还有 6/8 拍等等。为了找到这种强弱关系, 先寻找强拍, 由基本乐理^[8]可知, 在连续的 6 拍中, 必有一个是强拍。因此拍号的识别可由以下步骤进行。

1) 任意选定连续的 6 拍, 依次计算各拍的能量 $E_b(n)$;

2) 比较这 6 拍的能量, 确定能量最大值 E_{\max} 、极大值 E_{\max}' 和最小值 E_{\min} , 分别对应强拍、次强拍和弱拍;

3) 根据强拍和弱拍出现的位置, 从而确定拍号。

1.7 调号的识别

为了识别调号, 这里根据乐理知识提出一种变化音最少法则, 即如果选定的调号使得乐谱中变化音(升降音)的个数最少, 那么此调号即为该乐曲的调号。下面举例验证。

对于调号为 1 = E 的简谱音符序列 $N(i)$: $\underline{6} \underline{7} \underline{6} \underline{3} \underline{4} \underline{5} \underline{6} \underline{5} \underline{3} - -$, 此序列的变化音个数为 0; 如果改为调号为 1 = F, 则为了保证绝对音高不变, 简谱序列为: $\underline{5} \underline{6} \underline{5} \underline{2} \underline{3} \underline{4} \underline{5} \underline{6} \underline{5} \underline{3} - -$, 此序列的变化音个数为 8; 如果改为调号为 1 = G, 则序列为: $\underline{4} \underline{5} \underline{4} \underline{1} \underline{2} \underline{3} \underline{4} \underline{1} \underline{3} \underline{1} - -$, 变化音个数为 6。因此只有在对应的调号 1 = E 的情况下, 变化音的个数最少。

在计算机进行识别时, 可进行这样的算法过程:

1) 任意选取一定的调号, 在此调号下计算变化音的个数 n ;

2) 如果 $n = 0$, 则结束; 如果 $n \neq 0$, 重复 1), 直到 n 最小。

2 实验

下面以钢琴独奏曲《化蝶. mp3》的某一微小片段为例, 说明乐谱识别过程, 操作均在 VC++ 下进行。

1) 利用 AV2Wav 软件, 将其转化为《化蝶. wav》, 格式为: 采样频率 44.1 kHz, 16 位, 双声道。利用 CoolEdit 截取某几个音符, 一个声道的波形如图 2 所示, 该微小片段的时间为 $t = 2.5$ s。

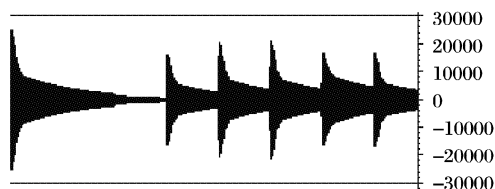


图2 一个声道的波形图

2) 以十六分音符的时间步长 $dT = 0.15$ s, 将乐句分为 17 个子段。

3) 利用 DCT 测量得各子段的基音频率结果以及相对误差和音名如表 4。其中第 1 子段的 DCT 频谱(取样本个数为 2000)如图 3, 基音频率为 587 Hz。

表 4 测量结果与相对误差

子段	基频测 量值/Hz	基频期 望值/Hz	相对 误差/%	音 名	子段	基频测 量值/Hz	基频期 望值/Hz	相对 误差/%	音 名
1	587	587.3	-0.05	d ²	10	662	659.3	0.41	e ²
2	585	587.3	-0.39	d ²	11	585	587.3	-0.39	d ²
3	589	587.3	0.29	d ²	12	591	587.3	0.63	d ²
4	591	587.3	0.63	d ²	13	490	493.9	-0.79	b ¹
5	583	587.3	-0.73	d ²	14	496	493.9	0.43	b ¹
6	586	587.3	-0.22	d ²	15	591	587.3	0.63	d ²
7	782	784.0	-0.26	g ²	16	583	587.3	-0.73	d ²
8	792	784.0	1.02	g ²	17	570	587.3	-2.95	d ²
9	665	659.3	0.86	e ²					

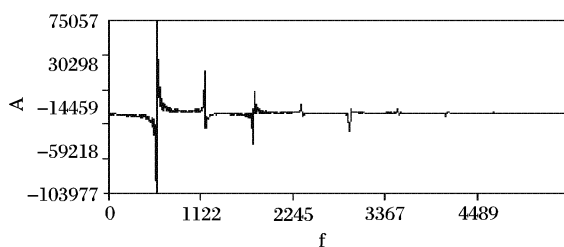


图3 第 1 子段的频谱

由表 4 知, 第 17 子段的相对误差较大, 这可能是由于边界效应产生的, 是下一个音符和第 17 子段对应的音符综合作用的结果, 为此, 可将此子段剔除。这样, 相对误差范围在 -0.79% 和 1.02% 之间。

4) 利用式(3)并进行子段能量比对, 测得音符端点在第 6、8、10、12、14 子段处, 因此可合并 1~6、7~8、9~10、11~12、13~14、15~16 子段, 得到准乐谱序列为:

$$|d^2 \cdot g^2 \cdot e^2 \cdot d^2 \cdot b^1 \cdot d^2|$$

5) 利用式(4)并进行拍能量比对, 测得强弱关系为: “强、弱、次强、弱”, 因此拍号为 4/4。

6) 根据最少变化音法则, 通过第 1.7 节的算法过程, 可知调号为 1 = G, 最终乐谱的音符序列为:

$$|5 \cdot \underline{1} \underline{6} \underline{5} \underline{3} \underline{5}|。$$

(下转第 788 页)

快速的人脸定位准则。

参考文献:

- [1] GOLDMANN L, MÖNICH U J. Components and their topology for robust face detection in the presence of partial occlusions[J]. IEEE Transactions on Information and Forensics Security, 2007, 2(3): 559 – 569.
- [2] MCKENNA S, GONG S, RAJA Y. Modelling facial colour and identity with gaussian mixtures[J]. Pattern Recognition, 1998, 31(12): 1883 – 1892.
- [3] WU JIANXIN, MATTHEW B S. Fast asymmetric learning for cascade face detection[J]. IEEE Transactions on Pattern and Machine Intelligence, 2008, 30(3): 1 – 14.
- [4] WANG P, JI Q. Multi-view face and eye detection using discriminant features[J]. Computer Vision and Image Understanding, 2007, 105(2): 90 – 111.
- [5] HSU R L, ABDEL-MOTTALEB M, JAIN A K. Face detection in color images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 696 – 706.
- [6] ROWLEY H A, BALUJA S, KANADE T. Neural network-based face detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(1): 23 – 38.
- [7] PHIMOLTARES S, LURSINSAP C, CHAMNONGTHAI K. Face detection and facial feature localization without considering the appearance of image context[J]. Image and Vision Computing, 2007, 24(5): 741 – 753.
- [8] CHEN JIE, WANG RUIPING, YAN SHENGYE. Enhancing human face detection by resampling examples through manifolds[J]. IEEE Transactions on System, Man, and Cybernetics – Part A: System and Humans, 2007, 37(6): 1017 – 1028.
- [9] WU H Y, CHEN Q, YACHIDA M. Face detection from color images using a fuzzy pattern matching method[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(6): 557 – 563.
- [10] CHIANG C C, HUANG C J. A robust method for detection arbitrarily titled human faces in color images[J]. Pattern Recognition Letters, 2005, 26(16): 2518 – 2536.
- [11] YANG J, LU W, WAIBEL A. Skin-color modeling and adaptation [C]// ACCV98, LNCS 1352. Berlin: Springer, 1997: 687 – 694.
- [12] YOO T W, OH I S. A fast algorithm for tracking human faces based on chromatic histograms[J]. Pattern Recognition Letters, 1999, 20(10): 967 – 978.
- [13] JONES M J, REHG J M. Statistical color models with application to skin detection[J]. Computer Vision, 2002, 46(1): 81 – 96.
- [14] LIN C. Face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network[J]. Pattern Recognition Letters, 2007, 28(16): 2190 – 2200.
- [15] YANG M H, AHHJA N. Gaussian mixture model for human skin color and its application in images and video databases[C]// Proceedings of SPIE: Conference on Storage and Retrieval for Image and Video Database. San Jose, CA: SPIE, 1999: 458 – 466.
- [16] SORIANO M, HUOVINEN S. Using the skin lotus to cope with changing illumination conditions in color-based face tracking[C]// IEEE Nordic Signal Processing Symposium. Kolmarden, Sweden: IEEE, 2000: 383 – 386.
- [17] CHIANG C C, TAI W K. A novel method for detecting lips, eyes and faces in real time[J]. Real-Time Imaging, 2003, 9(4): 277 – 287.

(上接第 770 页)

7) 该识别方法与其他相关方法的比较如下:

①与傅里叶变换方法相比,该方法计算简单,而且相对误差较小,基本上在 1% 以内;

②与 MIDI 下的乐谱识别相比,由于该识别方法针对的是广泛存在的音频文件,因此具有更普遍的实用价值;

③与其他同类识别方法相比,该识别方法是以十六分音符的时值作为步长,是从乐理的本身出发,具有艺术层面和信息层面的双层含义。

④其他方法往往只进行基频的提取和绝对音高的映射,而该方法最终的识别结果是含有拍号和调号等信息的相对音高序列,这与音乐美学中更关心相对音高是相吻合的。

8) 管类乐曲的特殊处理。

对于管类乐曲,由于吹奏一个音符时形成的振幅包络线比较平坦,衰减比较小,因此第 1.5 节的方法不十分有效。为此,可将式(3)进行线性变换,即 $E'(k) = aE(k) + b$,目的是为了进行能量对比度的提升,一般情况下,取 $a > 3$, $b > 0$ 即可进行比较有效的能量比对。

3 结语

本文以十六分音符作为时间分辨率进行音频的分割,解决了短时值音符的遗漏问题。利用一维离散余弦变换提取了基音频率,样本个数越多,误差就越小,但计算量也增大,取 2000 个样本可基本满足要求。利用十二平均律完成了音频到音名的映射。通过子段能量的比对,判定了音符的端点,解决了相邻同音高的归属问题。利用拍能量的大小关系,识别

出了拍号。利用提出的变化音最少法则,识别出了调号。最终完成了乐谱的识别。

考虑到管类乐曲具有振幅包络线比较平坦,衰减比较小的特性,需要对其进行特殊处理,本文采用了线性变换,以提升能量高低的对比度,使其有效地完成能量的比对,从而使该识别方法具有更普遍意义。

参考文献:

- [1] 张一彬,周杰,边肇祺,等. 基于内容的音频与音乐分析综述[J]. 计算机学报, 2007, 30(5): 712 – 729.
- [2] ZHU Y-W, KANKANHALLI M S, GAO S. Music key detection for musical audio[C]// Proceedings of the 11th International Multimedia Modeling Conference. Washington, DC: IEEE Computer Society, 2005: 30 – 37.
- [3] CHANG W-C, SU A W Y. A novel recurrent network based pitch detection technique for quasi-periodic/pitch-varying signals[C]// Proceedings of the IEEE International Joint Conference on Neural Networks. Honolulu, USA: IEEE, 2002, 1: 816 – 821.
- [4] 徐国庆,杨丹,王彬洁,等. 乐音识别方法及应用[J]. 计算机应用, 2005, 25(4): 968 – 970.
- [5] 徐国庆,张彦铎,王海晖,等. 乐音旋律识别研究[J]. 武汉大学学报, 2007, 29(2): 60 – 63.
- [6] 郑阿奇,刘毅. 多媒体实用教程[M]. 北京: 电子工业出版社, 2007: 21.
- [7] 韩纪庆,冯涛,郑贵滨,等. 音频信息处理技术[M]. 北京: 清华大学出版社, 2007: 37.
- [8] 李重光. 基本乐理通用教材[M]. 北京: 高等教育出版社, 2004: 41 – 45.