ISSN 1009-3044
Computer Knowledge and Technology 电脑知识与技术
Vol.6,No.12, April 2010, pp.3029-3031

E-mail: eduf@cccc.net.cn
http://www.dnzs.net.cn
Tel:+86-551-5690963 5690964

哼唱音符划分技术研究

钟声声[1], 冯寅[2]

(1. 厦门大学 计算机科学系,福建 厦门 361005;2. 厦门大学 智能科学系,福建 厦门 361005)

摘要:哼唱音乐一般是一种波形文件,这样的格式并不利于检索和查找。在使用哼唱音乐检索音乐内容时,需要将哼唱文件转换为音高和时值的形式,作为检索关键字。这些步骤都建立在哼唱已经被按音符切分的基础上。论文采用一种基于振幅能量的多层次音符切分方法,实现对哼唱文件的快速切分。基于能量的划分方法具有简便快速的特点。分层次的划分方法能够针对各种不同音符情况,采用最合适的方法切分。论文还讨论了一种基于音高识别技术的音符划分方法。

关键词:哼唱;音符;划分;音块;检索;振幅

中图分类号:TP311 **文献标识码**:A **文章编号**:1009-3044(2010)12-3029-03

Study of Humming Divide Technology According to Note

ZHONG Sheng-sheng[1], FENG Yin[2]

(1.Department of Computer Sinence of Xiamen University, Xiamen 361005, China;2. Department of Intelligence Science of Xiamen University, Xiamen 361005, China)

Abstract: Humming music is commonly a wave file, whose format is not conducive to be searched and found. When retrieving musical content by Humming music, humming documents need to be transformed to the format of pitch and time value, as the search keyword. These steps are built on that the singing has been cut by note basis points. This paper presents a multi-level note segmentation methods based on the amplitude of the energy, achieving rapid humming file segmentation. Energy-based partition method is characterized for simplicity and rapidity. Hierarchical classification approach can adopt the most appropriate method of segmentation due to the situation for a variety of different notes. This paper also discusses a note division method based on pitch recognition technology.

Key words: humming; note; divide; note piece; retrieve; swing

随着数字化技术的发展,日常生活与计算机的结合越发紧密,越来越多的事物需要在计算机中予以表示。从古至今,音乐都是人们最为喜爱的娱乐活动之一,随着 WAV 和 Mp3 等音乐格式的诞生,音乐已经能够储存于计算机之中,并实现了数字化。但是,只有这样的格式并不利于人们查找和检索音乐,特别是当用户以自己的哼唱来检索[1-2]音乐时,往往难以实现高效而准确的检索。

当采用哼唱检索音乐时,必须先将哼唱转换为计算机可识别的检索关键字——音高和时值序列。但是要区分音高和获得时值,就必须先对哼唱文件进行音符切分。文献[3-4]从乐理的角度介绍了哼唱音符的划分,文献[5]提出了利用倒谱峰值曲线划分哼唱的方法。本文受到文献[5]中的启发,设计了一种基于能量的多层次切分方法。

1 音符划分技术概述

为方便描述音符划分方法,本文引入了音块的定义。

1.1 音块的定义

在哼唱波形中,同时满足以下两个条件的波形片段,称为一个音块:

1) 波形片段内有且仅有一种音高存在;

2) 波形片段内有且仅有一个汉字。该汉字来自哼唱的歌词。

在实际系统应用中,音块的定义可以适当放松,只要人耳认为某个波形片段大致满足这两个条件即可。在某些哼唱者哼唱不准确的情况下,还可以人工的掠过哼唱错误的音。

2.1 音块切分技术

所谓音块切分技术, 即将音频文件按照不同音块切割开来,可表现为在音频文件波形图之上,画上罗干垂直切割线,以区分不同音块,见图 1。

图 1 是对钢琴弹奏的"Do Re Mi Fa So"录制成的音频文件进行音块切割的结果。其中的每个音块都有且仅有一个音高。

由于乐器发音规范,变化较简单,因此其切分的实现也简单。

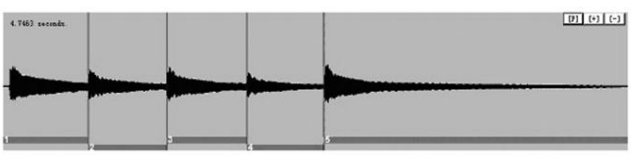


图 1 音块切分示例

收稿日期:2010-02-22

作者简介:钟声声(1986-),男(畲),福建霞浦人,厦门大学计算机科学系学生, 硕士,主要研究方向为基于内容的音乐检索;冯寅(1963-),男,福建福州人,副教授,博士,主要研究方向为算法作曲,计算机音乐和自然语言处理。

未经专业训练的人是不能像乐器那样规范发声的,他们哼唱时,可能声音会抖动,不平稳,音量大小也控制不当,再加上录音环境和设备可能带来的噪音,给哼唱音块划分带来很多困难。

2 基于能量的音块划分

本论文采用的划分方法主要分为两个层次处理。第一个层次考虑到哼唱中的停顿和音块相邻区域的特点作出初级音块划分;第二层次针对初级音块划分未考虑到的情况,实现了弥补作用的二级音块划分。

2.1 初级音块划分

声音在图像上的表现是波形。通过观察可知,在声音发出的位置,波形上有一个波峰;对于哼唱和单声部音乐而言,歌词汉字以及音符的开始处,一般都有一处波峰的出现,而波峰的两侧是波谷。考虑到这个特点,可通过设置一个阈值,来截取出波峰的位置,作为一个音块的初步划分。由于这种阈值的划分像一个筛子,可以筛出我们需要的特征,因此,可称初步音块划分为筛法划分。

筛法划分的具体算法如下:

1) 读取波形文件,对每个振幅能量做绝对值运算,保证为正数。

2) 以 1600 为步长,对所有振幅能量按窗口划分。每个窗口中计算出一个平均能量值,所有的平均能量值保存为数组 AverSwing。

3) 求出最大振幅能量,以最大振幅能量和阈值系数 0.07 的乘积,作为划分阈值 Threshold_0,进行噪音过滤。如果 AverSwing[i] >= Threshold_0,AverSwing[i−1] < Threshold_0,AverSwing[i+1] < Threshold_0,说明该音块的延续太短,认为是一个噪音。此时,令 AverSwing[i] = AverSwing[i+1],以过滤噪音。

4) 对 AverSwing 按照以下公式锐化:

AverSwing[i] = AverSwing[i] * (1.0 − C + C * AverSwing[i] / Max)

其中 C 是锐化因子,一般取 0.5 左右。

5) 在 AverSwing 基础上,重新计算最大能量 Max,以最大能量和阈值系数的乘积,作为新的划分阈值 Threshold_1。

6) 以阈值 Threshold_1 为界限,遍寻 AverSwing,如果满足以下条件,则认为存在一个音块。

AverSwing[i],…,AverSwing[i + k] > Threshold_1　其中 k>3

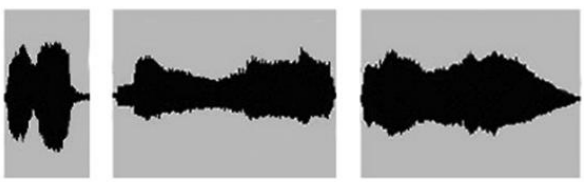
该音块的开始位置为 AverSwing[i],结束位置为 AverSwing[i + k]。

其余处,由于能量太小或持续太短,暂时认为无音块存在。此情况后续处理见 2.2.3 小节。

7) 筛法划分结束

2.2 二级音块划分

初级音块划分已经能够实现如图 1 那样的乐器音声划分。但是,由于人声哼唱的复杂性,会出现很多初级划分无法解决的情况。本文考虑到这其中的一种常见情况,如图 2。



(a)　(b)　(c)

图 2　几种音块连接情况

这三幅波形截图有一个共同点,就是波形中间下陷,但又没有陷到足够区分为两个音块。这种情况下,大多是两个音符连贯得唱出。如此的情况,如果要用筛法划分,需要调大阈值;可是,如果阈值过大,又会影响到其他音块的划分,可能导致许多音符被忽略掉。对此,本文专门设计了二级音块划分算法。设计了第二层的阈值 Threshold_2,来解决这个问题。

二级音块划分详细算法如下:

1) 取出已经初步获得的音块,逐个处理(以下是循环结构)

2) 令 AverSwing[St]为音块第一个平均能量值,满足如下条件时,St 自增:AverSwing[St + 1]> AverSwing[St] ‖ AverSwing[St + 2] > AverSwing[St] ‖ AverSwing[St + 3]> AverSwing[St]

3) 令 AverSwing[Ed]为音块第一个平均能量值,满足如下条件时,Ed 自减:AverSwing[Ed + 1]> AverSwing[Ed] ‖ AverSwing[Ed + 2]> AverSwing[Ed] ‖ AverSwing[Ed + 3]> AverSwing[Ed]

4) 寻找 K,使得 AverSwing[K] = Min(AverSwing[St],…, AverSwing[Ed])

5) 如果满足以下两个条件,则以 K 为切分位置,将当前音块分为两个。

① 存在 a ∈ [St,K) 使得 AverSwing[K] < AverSwing[a] * 0.6

② 存在 b ∈ (K,Ed] 使得 AverSwing[K] < AverSwing[b] * 0.6

6) 完成所有音块的处理后,二级分割即结束。

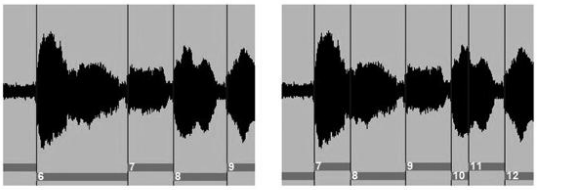
2.3 整理音块

由于前两步的处理,使得音块之间可能不是连续的,这可能导致忽略掉部分音量较小的音符或给后期划分和音高的手工调整带来麻烦。因此,可以将第一个的音块的开始位置固定调整的哼唱的开始位置,将最后一个音块的结束位置固定调整到哼唱的结束位置,将除去最后一个音块的所有音块的结束位置调整为后继音块的开始位置。如此处理,音块可连续。

3 其它划分方法探讨

基于能量的划分虽然能够快速而有效的解决大多数音块划分的情况,但是总有一些特殊情况,是能量无法划分的。比如多个音符之间连续,哼唱过程不换气,且音符切换之间无明显音量降低或音量降低的幅度不够。如图 3 的(a)所示,音块 6 和音块 8 是具有

两个音符的,可是用前两步的音块划分方法,只能得到图(a)的划分结果。因此本文提供了一种基于以已经识别音高的划分方法——音高切分方法,解决这种问题。音高切分方法,是基于第三章的音高识别技术的基础上,提出的一种修补性音块划分方法,用于解决音块划分不够细化的情况。对于(a),音高切分后的结果即为图 3 的(b)。



(a) (b)

图 3 音高切分示例

音高切分详细算法如下:

1) 对现成音块逐个处理(以下是循环结构)

2) 取音高切割系数 NoteDivideCoef=10,与音块中 FFT 窗口数 N 比较

3) 如果 N<NoteDivideCoef 不做切分。

4) 如果 N<=2*NoteDivideCoef,将音块平分为两子音块分别求出音高 Note1 和 Note2,如果 Note1 等于 Note2,则不切分。否则切分。

5) 如果 N<=3* NoteDivideCoef,将音块均分为三子音块分别求出音高 Note1、Note2 和 Note3,如果三者相等,则不切分。否则,如果 Note1 等于 Note2 或者 Note2 等于 Note3,则将原音块重新按照第 4 步方法,均分两块判断,以决定是否平分切分。如果上述两个条件都不能满足,则将音块均分为三块。

6) 如果 N<=4* NoteDivideCoef,将音块均分为四个子音块。分别求出音高。如果相邻的两个音高相同,则将他们对应的子音块合并成一个子音块。如此合并,直到不存在相邻音高。则将原音块替换为新生成的子音块。

7) 对于其它情况,按 NoteDivideCoef 长度平分为罗干个子块判断。采用类似第 6 步的方法,合并连续子音块。最后用全部子音块,替换原音块。

8) 循环结束,则结束音高切割。

音高切分方法可以一定程度上解决音块划分不准确的问题,但是这种方法也可能将现有的正确音块划分破坏掉。如果哼唱者声音不稳定,在分成子音块识别音高时,可能会造成子音块之间并非同一个音高,而音高切分方法会将这个音块切分为多个,这反而不符合方法设计的初衷——使音块划分更加准确。

4 结论

本文介绍了哼唱文件的音块切分技术,提出并详细叙述了基于能量的多层次音块划分方法,提出和探讨了一种基于已识别音高的音块进一步分割的方法。前者方法,在实际使用中,迅速有效,但是容易忽略部分音符连接的特殊情况,而且抗噪声能力不够。后者可以完成前者不能完成的切分,但可能过度切分。两种方法各有优劣,实际应用中应综合考虑。

参考文献:

[1] Foote. Content-Based Retrieval of Music and Audio[J].Multimedia Storage and Archiving systems II,Proe of SPIE,1997, 3229:138-147.
[2] 薛锋,杨宗英,郑巧英,黄敏.基于内容的音乐检索[J].大学图书馆学报,1999,(4):28-30.
[3] 赵宋光.音乐教育心理学概论[M].上海:上海音乐出版社,2003.
[4] 李玫缪.中国传统律学[M].福建:福建教育出版社,2008.
[5] 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.
[6] 郭红波.基于内容的音乐检索关键技术的研究[D].西安:西北大学,2007.
[7] 许文豪,高名扬,张智星. 直觉式哼唱输入音乐搜索引擎[C].台湾:第五届人工智能与应用研讨会,2000:734-740.
[8] 杨俊,蔡宣平,颜飞翔.数字音频技术及其应用与发展(二)[J].电声技术,2001,6:12.

(上接第 3007 页)

综上所述,在制作多媒体课件时,要采用适当的方法,使得多媒体课件的色彩、文字、页面以及背景的构设上既平衡又有变化,看起来美观。同时要使课件中具有表现力和艺术感染力,我们可将色彩与文字、图形、图像、动画、视频有机的结合在一起,这样不仅可以给学习者提供良好的视觉美感和愉悦感,而且还可以为他们提供分析、理解教学内容的学习条件,这是做好多媒体课件中色彩搭配的最终目的。

参考文献:

[1] 赵子江.多媒体技术应用教程[M].北京:机械工业出版社,2007.1.
[2] 陈静.办公软件应用教程工作任务汇编[M].北京:化学工业出版社,2010.1.
[3] 王涛.多媒体课件的页面构设[J].教育技术研究,2002.2.
[4]《冷暖对比》[DB/OL]. http://www.apoints.com/colour/scjc/scdb05.htm,2003.5.
[5] 杨治良. 实验心理学[M].浙江:浙江教育出版社,1998.12.