

基于匹配追踪的音符起始点检测

桂文明^{1,2}, 刘睿凡¹, 邵 曦³, 白光一^{1,4}

(1. 南京理工大学计算机与科学技术学院, 江苏南京 210094; 2. 南京财经大学, 江苏南京 210046;
3. 南京邮电大学, 江苏南京 210003; 4. 方舟信息(苏州)技术有限公司, 江苏苏州 215021)

摘要: 在基于内容的音乐研究中, 正确提取音符起始点信息是识别音高、节拍、节奏、段落等音乐高级特征的基础. 本文提出了基于匹配追踪(Matching Pursuit, 简称 MP)的两种新型音符起始点检测算法: 基于 MP 解释程度和基于分音变化的检测算法. 这两种算法均在 MP 分解的基础上, 分析 MP 码本, 并利用改进的峰值提取算法生成音符起始点向量. 从实验结果看, 本文提出算法的性能指标和 MIREX 2011 的最好结果相当.

关键词: 音符起始点检测; 匹配追踪; MP 解释程度; MP 分音变化

中图分类号: TP391.42 **文献标识码:** A **文章编号:** 0372-2112 (2013) 06-1225-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2013.06.029

Note Onset Detection Based on Matching Pursuit

GUI Wen-ming^{1,2}, LIU Rui-fan¹, SHAO Xi³, BAI Guang-yi^{1,4}

(1. School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China;
2. Nanjing University of Finance & Economics, Nanjing, Jiangsu 210046, China; 3. Nanjing University of Posts and Telecommunications,
Nanjing, Jiangsu 210003, China; 4. Noah Solution Co. Ltd., Suzhou, Jiangsu 215021, China)

Abstract: Note onset detection is the preliminary work for recognition of high-level musical feature, such as pitch, rhythm, tempo and paragraphs, in content-based music information retrieval. This paper proposed two novel algorithms based on matching pursuit(MP): the algorithm of MP degree of explanation and the algorithm of MP change of partial. Firstly, the musical signals were decomposed through MP, and then the code books were analyzed with the two algorithms. Finally, a modified peak-picking algorithm was applied to generate note onset vectors. The experiments showed the performance of our algorithms was nearly as good as that of 2011 MIREX.

Key words: note onset detection; matching pursuit; MP degree of explanation; MP change of partial

1 引言

从音乐信号本身出发, 通过信号处理、模式识别、机器学习等方法自动提取音乐特征信息, 成为近年来音乐信息检索(Music Information Retrieval, MIR)领域的研究热点. 研究者们试图从音乐信号中提取音乐分类、音乐结构、音乐指纹、节拍跟踪、音符起始点等特征信息^[1~3]. 其中, 音符起始点检测是 MIR 领域中的基础性研究工作, 是正确识别音高、节拍、节奏、段落等高级特征的前提.

音符起始点检测算法的一般包含三个部分^[4]: 预处理、检测函数生成和峰值提取. 预处理部分是音符起始点检测的关键部分, 是整个算法的出发点. 在这个过程中, 一般是将原音乐信号分帧, 并对每帧信号实施变换,

然后再从变换后的信号中提取特征信息. 其中, 常用的信号变换方法有: 短时傅立叶变换(STFT)、恒定 Q 变换(CQT)、小波变换(WT)等, 使用最多的是 STFT. 提取的特征信息主要有: 时域能量特征、幅度谱特征、相位谱特征、时频信息(时间尺度信息)等, 使用最多的是幅度谱特征. 检测函数生成部分一般是根据前后两帧的特征信息, 计算它们之间的距离. 计算的方式包括 L-1 范数, L-2 范数或者它们的变形. 峰值提取部分主要是对距离设定阈值, 提取音符起始点.

本文使用匹配追踪算法^[5](Matching Pursuit, MP)进行音符起始点检测, 该算法不仅对独奏乐曲有效, 而且对现有算法很少提及的混合音乐有良好效果. 混合音乐含有多种乐器或人声, 音符起始点检测难度大. 据我们所知, 到目前为止, 还没有使用 MP 算法进行音符起始

点检测的论文和报告. 音乐信号是由谐振系统或物理冲激、或二者同时产生的. 谐振系统可产生各种频率的平稳信号, 冲激集中在某一时刻产生非平稳信号. 平稳信号使用 STFT 能准确地表示, 而 WT 在非平稳信号的表示方面有优势. MP 能结合 WT 和 STFT 的优势, 对音乐信号进行更精确地表示. MP 是在冗余字典下分解信号, 得到信号的稀疏表示. 其字典的冗余特性给了我们极大的选择空间, 我们可以在分解信号的过程中, 在字典中加入 STFT 和 WT 的基, 使得原信号得到更稀疏更精确的表示. 只要原音乐信号得到更精确的表示, 便能通过分析其分解后的原子成分得到准确的音符起始点. 混和音乐含有不同乐器或人声, 单独采用 STFT 或 WT 分解这些信号, 缺乏足够的适应性和稀疏性, 从而影响了音符起始点检测的准确率. 本文的另一个贡献是在峰值提取部分, 采用基于高斯核的光滑算法对检测函数曲线进行处理, 提升了检测的性能.

2 音乐信号的 MP 分解算法

Mallat 在 1993 年提出了 MP 算法, MP 中冗余字典的灵活性和对信号的自适应特性为 MP 带来旺盛的生命力, 此后有关 MP 的研究和应用一直方兴未艾, 它在当前热门的压缩感知领域也有着重要的角色^[6]. MP 的思路是在冗余字典中寻找和原信号最相似的原子, 并以此原子为稀疏表示的成分. 有关 MP 算法的步骤请参阅文献^[7].

在 MP 算法中, 冗余字典的选择是关键的一步, 其组成原子应尽可能和原始信号相似. 基于音乐是由丰富多彩的基音和泛音组成的特点, 文献^[8]通过 Gabor 原子的线性组合, 生成由泛音原子 (Harmonic Atoms) 构成的冗余字典. 音乐信号又可以看作由以下三部分组成的模型: 第一部分是音调部分, 可表示为幅度和频率缓慢变化的正弦函数和; 第二部分是突变部分, 为音符的起始点; 第三部分是剩余的局部平稳的白噪声成分. 针对音乐信号这个模型, 文献^[9]设计的冗余字典是改进离散余弦变换 (MDCT) 正交基和离散小波变换 (DWT) 正交基的并集, 其中 MDCT 基用来表示音调部分, 而 DWT 基表示突变部分.

本文在 MP 算法中, 冗余字典采用实值函数集 $g_r(t)$:

$$g_r(t) = \frac{K_r}{\sqrt{s}} g\left(\frac{t-u}{s}\right) \cos(\xi t + \varphi) \quad (1)$$

其中参数 $r = (s, \xi, u, \varphi)$, $s \in \mathbb{R}^+$, $\xi, u \in \mathbb{R}$, $\varphi \in [-\pi, \pi)$, 窗口函数 g 为高斯窗, K_r 使得 $\|g_r\| = 1$. 当 s 固定时, MP 类似 STFT, 有利于分析信号在平稳状态下的频率成分, 这可适应音符的音调部分; 当 ξ 固定时, MP 类似 WT, WT 可很好地描述信号的奇异性, 这正是音符起

始点的分析所需要的. 从 $g_r(t)$ 的性质分析, MP 结合了 WT 和 STFT 的优势, 可以比 WT 和 STFT 更精确地表示音乐信号.

本文的实验中, 取窗函数 $g(t) = \exp(-\frac{t^2}{2p})$, 其中 $p = 0.02$, $s = (s_i)_{i=1, \dots, 6} \in [16, 32, 64, 128, 256, 512]$, $\xi = \frac{j\pi}{s_i}$, $j = 0, \dots, (s_i - 1)$, $u = 16k$, $k = 0, 1, \dots, (\frac{512 - s_i}{16})$, $\varphi = \frac{\pi}{12}l$, $l = -12, -11, \dots, 11$, 并使用基于 MPTK 的快速 MP 算法^[7]对音乐信号进行分解, 创建的字典以 Any-wave Block 格式嵌入到 MPTK. 信号分解后, 可得到其原子成分集合, 称为 MP 码本.

3 MP 下的音符起始点检测

对音乐信号进行 MP 预处理后, 如何分析 MP 结果来检测音符起始点呢? 本文提出了基于 MP 解释程度 (EBA) 和基于分音变化 (PBA) 的两种起始点检测算法. 这两种算法分别使用不同的检测函数分析 MP 码本, 然后使用相同的峰值提取算法生成音符起始点. 图 1 为整个检测算法流程图.

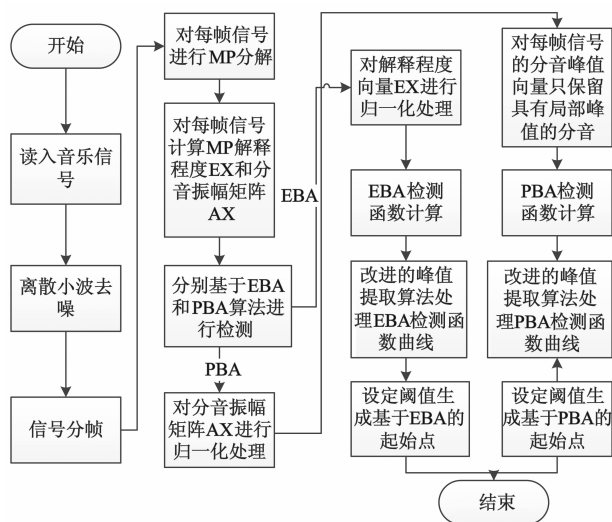


图 1 本文算法流程图

在进行 MP 分解之前, 对信号进行了离散小波变换. 采用 db4 进行一层分解后利用其近似信号进行重建, 这样可以在一定程度上对信号进行去噪, 提升检测结果性能.

3.1 EBA 的检测函数

根据 MP 分解后的能量守恒定律^[5], 有:

$$\|x\|^2 = \sum_{n=0}^{m-1} | \langle x_n, g_r \rangle |^2 + \|x_m\|^2 \quad (2)$$

其中 x 为待分解信号, D 为冗余字典, g_r 为 D 中的原子, r 为原子参数集. 信号 x 的能量被分解成 $m+1$ 个成

分,其中前 m 个部分为每次分解后的剩余信号 x_n 和最匹配的原子的内积的模平方.这前 m 个成分就是 x 在 D 下的被解释的部分,而 x_m 是剩余未解释的部分.因此定义前 m 个成分的能量和 $\sum_{n=0}^{m-1} | \langle x_n, g_{r_n} \rangle |^2$ 为信号 x 在 D 下的解释程度.

音乐信号主要是由音调和起始点两部分组成.在音调部分,频率和幅度变化缓慢,所以这部分的相邻信号帧被分解后的原子集合相对稳定,其能量和也相对稳定,即被解释程度趋于稳定.而在起始点部分,频率、幅度突变,相邻信号帧在 MP 下解释程度也会发生突变.所以可以定义基于 MP 解释程度的函数,用来检测音符起始点.

由式(2)可得 MP 分解 m 次后,解释程度 EX_m 可表示为:

$$EX_m = \sum_{n=0}^{m-1} | \langle x_n, g_{r_n} \rangle |^2 = \| x \|^2 - \| x_m \|^2 \quad (3)$$

设音乐信号 x 的相邻信号帧分别为 X_n 和 X_{n+1} ,对应的解释程度 EX 值分别为 $EX_{n,m}$ 和 $EX_{n+1,m}$.所有信号帧的 EX 值构成一个向量,首先对该向量进行归一化处理:

$$EX_{n,m} = EX_{n,m} / \max(EX_{n,m}) \quad n \in (1, \dots, N) \quad (4)$$

m 为分解次数,再定义检测函数:

$$\text{Onset}_1(n) = \text{dist}(X_n, X_{n+1}) = \text{dist}(EX_{n,m}, EX_{n+1,m}) \quad (5)$$

其中 dist 为距离函数.因为起始点都伴随着能量的增长,所以只取 $EX_{n+1,m} > EX_{n,m}$ 的距离.本文取 dist 函数为欧式距离,检测函数为:

$$\text{Onset}_1(n) = \begin{cases} (EX_{n+1,m} - EX_{n,m})^2, & \text{if } EX_{n+1,m} > EX_{n,m} \\ 0, & \text{if } EX_{n+1,m} \leq EX_{n,m} \end{cases} \quad (6)$$

3.2 PBA 的检测函数

我们听到的某时刻的音乐一般是由一个或多个音符产生的复合音组成的,复合音包含这些音符的各级分音,第一分音是基音,其他分音都是泛音^[10].在任何音符的起始点部分,其分音成分的能量都会发生不同程度的变化,因此,检测这些分音的能量变化就能检测到音符起始点.那么,又如何来检测分音的能量变化呢?每帧音乐信号经过 m 次 MP 分解后,可得到 m 个原子,如果把原子和音乐的分音通过频率参数进行匹配,那么每帧信号就可得到 m 个分音,而原子的能量就反应了分音的能量.

原子和分音通过频率如何匹配呢?第一种方法是原子频率等间隔划分.这种方式下分音的频率和原子频率不是一一对应的,因为分音的频率分布不是均匀的,而是呈指数增长的.第二种方法,也是本文采用的

方法,原子频率按十二平均律的频率间隔划分,这种划分方法类似 CQT,使原子和分音的频率对应具备了听觉特性,更符合音乐自身的特性.

本文参照钢琴十二平均律^[10]的频率分布规律匹配原子和分音,分音的能量即是该分音频段内的各原子的能量和.频率范围取 27.5Hz 到 3520Hz,在音乐中,超出这个范围的频率的音一般不用.整个频段包含了 7 个八度,把每个半音的频率间隔划分成 p 个部分(我们称 p 为半音的分辨率),相邻两个频率间隔的增长率为 $12\sqrt[12]{2} - 1$.这样就把分音的频率分成了 $84p$ 个小段,在这些小频段内分别收集匹配原子的能量和(用振幅值).因此,通过分析 MP 码本,可得到一个分音振幅矩阵 $AX(k, n)$,其中 $k \in [1, 84p]$ 为分音序号, $n \in [1, \dots, N]$ 为帧序号.在进行检测函数计算之前,和 EBA 算法一样,需要对分音振幅矩阵进行归一化处理.

$$AX(k, n) = AX(k, n) / \max(AX(k, n)) \\ , k \in [1, \dots, 84p], n \in [1, \dots, N] \quad (7)$$

对每个 n ,再利用式(8)对 $AX(k, n)$ 进行处理,只保留具有局部峰值的分音.这一步可以去除分音噪声,只保留主要分音.

$$AX(k, n) = \begin{cases} AX(k, n), & \text{if } AX(k, n) > AX(k-1, n) \\ & \text{and } AX(k, n) > AX(k+1, n) \\ 0, & \text{if } AX(k, n) \leq AX(k-1, n) \\ & \text{or } AX(k, n) \leq AX(k+1, n) \end{cases} \quad (8)$$

然后定义检测函数:

$$\text{Onset}_2(n) = \text{dist}(X_n, X_{n+1}) \\ = \text{dist}(AX_n, AX_{n+1}) \\ = \sum_{k=1}^{84p} \text{dist}(AX_{k,n}, AX_{k,n+1}) \quad (9)$$

PBA 的检测函数值即是各分音距离之和.因为起始点一般都伴随着分音振幅的增大,所以只取 $AX_{k,n+1} > AX_{k,n}$ 的距离,分音距离函数定义如下:

$$\text{dist}(AX_{k,n}, AX_{k,n+1}) = \begin{cases} (AX_{k,n+1} - AX_{k,n})^2, & \text{if } AX_{k,n+1} > AX_{k,n} \\ 0, & \text{if } AX_{k,n+1} \leq AX_{k,n} \end{cases} \quad (10)$$

3.3 改进的峰值提取算法

在设定阈值前,对检测函数曲线进行光滑处理是峰值提取的重要步骤.音符起始点检测的光滑处理应该有三个目标:第一使曲线足够光滑;第二尽可能使峰值突出;第三保证峰值移位尽可能小.最简单的光滑方法是取相邻 m 个点的算数平均值,算术平均光滑虽然能达到曲线光滑效果,但会使峰值平坦化,易造成音符起始点的重复检测现象.本文在峰值提取算法中,提出了基于高斯核的非线性光滑处理算法,该算法能很好

地满足光滑处理的三个目标,有助于提升检测性能.

选定原点对称的概率密度函数 $K(\cdot)$ ($\int K(u) du = 1$) 为核函数,将 $\text{onset}(n)$ 做如下变换:

$$\widehat{\text{onset}}(k) = \sum_{j=1}^n w_{nj}(k) \text{onset}(j), \quad k = 1, \dots, n \quad (11)$$

其中

$$w_{nj}(k) = K_h(k-j) / \sum_{i=1}^n K_h(k-i) \quad (12)$$

这里 $K_h(u) = h^{-1}K(h^{-1}u)$ 是概率密度函数, h 为窗宽. 在实验中,选取了高斯函数 $K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$ 为核函数,并取窗宽 $h = 1$.

图2是检测函数曲线的处理对比例图,从这个图的对比例看,经高斯核光滑处理过的曲线显然既光滑峰值又突出.

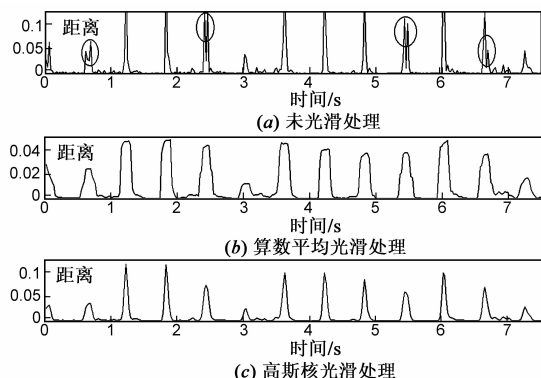


图2 检测函数曲线光滑处理对比

光滑后的检测函数再经过自适应阈值设定即可生成音符起始点. 本文在自适应阈值部分采用文献[11]的方法. 实验中,移动窗口的宽度为2s内的检测函数值个数.

4 实验

4.1 评估方法和数据集

本文采用 MIREX (Music Information Retrieval Evaluation eXchange) 的评估方法^[12] 评估实验结果. 准确率 P (Precision) 和召回率 R (Recall) 是两个互相制约的指标, F 值 (F-measure) 是两种指标的综合,提高 F 值是本文检测算法的目标.

本文使用 Matlab 对检测算法进行了编码实现. 测试音乐数据集分成 4 类: 固定音高打击乐 (Pitched Percussion, PP 类), 例如钢琴和吉他乐曲; 固定音高非打击乐 (Pitched Non-percussion, PN 类), 例如小提琴和风琴; 非固定音高打击乐 (Non-pitched Percussion, NP 类), 例如鼓; 混合音乐 (Complex Mixtures, CM 类), 可同时包含人声和乐器, 例如爵士乐和摇滚等. 测试数据共包含 18 首

乐曲, 其中一部分来自 Goto(2003) RWC 数据库, 一部分来自文献[13]提供的乐曲, 还有 3 首下载自互联网, 测试乐曲的事实起始点都经过音乐专业人员使用音频软件标注. 这些乐器覆盖了多种演奏乐器和音乐类型, 具有广泛的代表性. 事实起始点数量共 2050 个, CM、NP、PN、PP 类型的音乐各含 1156、165、263、466 个. 其中 CM 类样本数量较多, 含爵士乐、摇滚、戏剧、黑人乐等多种乐曲, 为的是验证本算法对混合音乐的检测效果. 4.2 节的实验结果均是基于此数据集, 选取最佳阈值得到的.

4.2 实验结果

本文数据集中的所有音乐信号的抽样率均是 22.05kHz, 实验中每帧的时长为 23.22ms, 每帧长度为 512, 帧重叠率为 3/4. PBA 和 EBA 两种检测算法的最优检测结果见表 1. 在评估结果时, 我们只考察检测结果的 F 值.

表1 PBA 和 EBA 算法的检测结果

算法	CM	NP	PN	PP	总平均
PBA	0.82841	0.80145	0.78346	0.89252	0.83067
EBA	0.79741	0.74749	0.76119	0.84647	0.79396

两种方法中性能较好的是 PBA 算法, 比文献[14]和文献[15]报告的最高值 82.1% 和 82.6% 分别高出 0.967% 和 0.467%. 和 MIREX 2011 的最佳结果 83.10% 仅相差 0.033%. 其中 CM 类, 比 MIREX 2011 最佳结果高出 6.471% (MIREX 为 76.37%), 证实了基于 MP 的算法在混合音乐的音符起始点检测中有优势. PP 类的检测结果最好, PN 类的检测结果较差, 因为 PN 类含软起始点多, 分音变化时能量变化不显著, 检测难度大.

MP 分解次数对检测结果有重要的影响. 为考察 MP 分解次数对检测结果的影响, 在同等条件下 (64、128、256 和 512 四尺度分解, 半音分辨率 p 为 20), 使分解次数从 15 到 250 变化, PBA 算法的检测结果见图 3. 从图中看出, 分解次数从 15 到 35 变化时, 各类音乐的 F 值迅速提高, 这是因为分解次数和提取的分音数量相关, 当分解次数逐渐提高时, 提取的分音数量逐渐增加, 对音乐信号的解析逐渐准确, 所以 F 值得到提高. 但当分解次数达到一定值后, 随着信号剩余能量下降, 外部噪

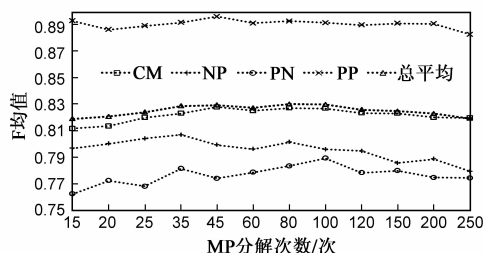


图3 分解次数对PBA算法的影响

声和内部组成信号之间的干扰越强,相似原子的选择(即分音)变得不稳定,分解次数增多并不能提升 F 值,故各类音乐的 F 值变化趋缓甚至出现下降趋势.从图中看出,PP类和CM类在分解次数大于45时呈缓慢下降趋势,NP类在分解次数大于35时下降趋势显著,PN类在分解次数大于100时呈缓慢下降趋势.总平均 F 值在分解次数为80时达到最高值,随后呈下降趋势.

EBA算法的检测结果见图4.和PBA算法类似,各分类音乐在分解次数达到一定值时,F值达到峰值,随后便趋于平稳或者下降.这是由于分解达到一定次数后,再分解时, $l < x_n, g_{r_n} > l^2$ 的值占 EX 值比重减小到不足以影响 EX 值的变化.

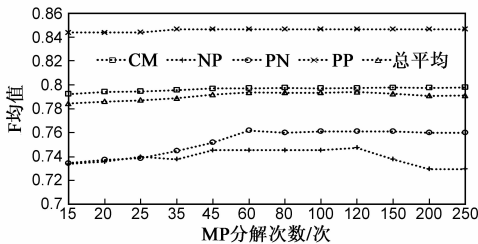


图4 分解次数对EBA算法的影响

在PBA算法中,半音的分辨率 p 对检测结果有无影响呢?图5为其他条件不变的情况下(尺度为64、128、256、512四尺度,分解次数为80次), p 变化对检测结果的影响。 p 越大表示半音的分辨率越高,那么PBA算法性能越好.但当 p 增长到一定值时,因为分辨率已经足以准确定位分音,所以PBA算法性能会达到峰值.通过图5可看出,各类音乐在半音分辨率超过6,PBA的性能接近相同.

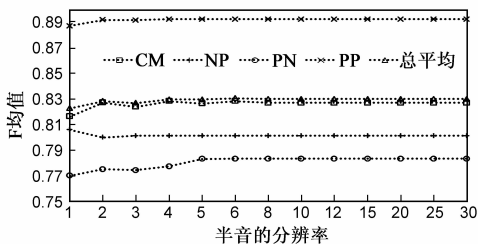


图5 半音的分辨率对PBA算法的影响

MP字典的分解尺度数对算法性能也有着影响,图6为其他条件不变的情况下(半音分辨率 p 为20,分解

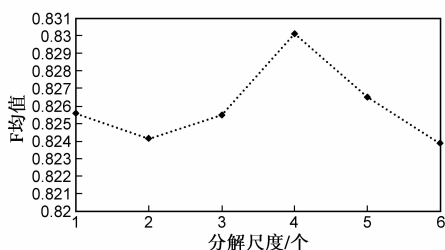


图6 分解尺度对算法影响

次数为80次)分解尺度个数从1变化到6时PBA算法的性能变化情况.在分解尺度数为4时(即64、128、256、512四尺度,此时的检测结果为表1的最优结果),算法性能最好.

5 结论

本文提出了基于MP的两种音符起始点检测算法,有一定的理论和实践价值.对现实生活中占最大比重的混合音乐类型来说,PBA算法相对其他算法有着显著的优势.本文提出的改进的峰值提取算法可提升音符起始点算法的性能,在语音起始点检测和其他信号奇异点检测算法中也有借鉴意义.

MP的最大优势是对信号的自适应性和灵活性,而这一点体现在冗余字典的构成上.在后续的研究中,将尝试寻找更适合音符起始点检测的冗余字典来提升检测性能.MP算法的缺陷是算法复杂度仍然较高,在音符起始点检测应用中,如果对检测速度要求高则不宜使用.

参考文献

- [1] Michael A. Casey Remco Veltkamp, Mastaka Goto, Marc Leman, Christophe Rhodes, Malcolm Slaney. Content-based music information retrieval: current directions and future challenges [J]. Proceedings of the IEEE, 2008, 96(4): 668 - 696.
- [2] 刘亚多,李伟,李晓强,等.压缩域鲁棒音乐指纹算法研究[J].电子学报,2010,38(5):1172 - 1176.
LIU Ya-duo, LI Wei, LI Xiao-qiang, et al. A robust compressed-domain music fingerprinting technique based on MDCT spectral entropy[J]. Acta Electronica Sinica, 2010, 38(5): 1172 - 1176. (in Chinese)
- [3] 陈哲,许洁萍.基于内容的音乐节拍跟踪[J].电子学报,2009,37(4A):156 - 160.
CHEN Zhe, XU Jie-ping. Content music based beat tracking [J]. Acta Electronica Sinica, 2009, 37(4A): 156 - 160. (in Chinese)
- [4] Bello J P, Daudet L, Abdallah S, Duxbury C, Davies M, Sandler M B. A tutorial on onset detection in music signals[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 1035 - 1047.
- [5] Mallat S G, Zhifeng Zhang. Matching pursuits with time-frequency dictionaries[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397 - 3415.
- [6] 石光明,刘丹华,高大化,等.压缩感知理论及其研究进展[J].电子学报,2009,37(5):1070 - 1081.
SHI Guang-ming, LIU Dan-hua, GAO Da-hua, et al. Advances in theory and application of compressed sensing[J]. Acta Electronica Sinica, 2009, 37(5): 1070 - 1081. (in Chinese)

- [7] Krstulovic S, Gribonval R. Mptk: Matching pursuit made tractable[A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing[C]. USA: IEEE Press, 2006. III.
- [8] Gribonval R, Bacry E. Harmonic decomposition of audio signals with matching pursuit[J]. IEEE Transactions on Signal Processing, 2003, 51(1): 101 – 111.
- [9] Daudet L. Sparse and structured decompositions of signals with the molecular matching pursuit[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14(5): 1808 – 1816.
- [10] 周复三. 音乐基础理论教程(修订版)[M]. 山东济南: 山东大学出版社, 2005.
Zhou Fu-san. Basic Music Theory Tutorial(revised)[M]. Jinan, Shandong: Shandong University Press, 2005. (in Chinese)
- [11] 桂文明, 刘睿凡, 邵曦, 等. 一种新型音符起始点检测算法[J]. 计算机工程与应用, 2011, 47(24): 9 – 11.
GUI Wen-ming, LIU Rui-fan, SHAO Xi, et al. A new algorithm for note onset detection[J]. Computer Engineering and Applications, 2011, 47(24): 9 – 11. (in Chinese)
- [12] MIREX. 2011 Audio Onset Detection[OL]. [http://www.music-ir.org/mirex/wiki/2011: Audio_ Onset_ Detection](http://www.music-ir.org/mirex/wiki/2011:Audio_Onset_Detection), 2011.
- [13] Pertusa Antonio. Computationally Efficient Methods for Polyphonic Music Transcription[D]. Alicante, Spain: University of Alicante, 2010.
- [14] Holzapfel A, Stylianou Y, Gedik A C, et al. Three dimensions

of pitched instrument onset detection[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(6): 1517 – 1527.

- [15] Benetos E, Stylianou Y. Auditory spectrum-based pitched instrument onset detection[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(8): 1968 – 1977.

作者简介



桂文明 男, 1974 年 8 月出生, 江西鹰潭人. 高级工程师, 南京理工大学在读博士, 研究方向为音乐信号处理、模式识别.

E-mail: guiwenming@126.com



刘睿凡 男, 1977 年 11 月出生, 江苏南京人. 在读博士, 研究方向为音乐信号处理, 图像分割.

E-mail: work_jerry@126.com

邵曦 男, 1976 年 11 月出生, 江苏南京人. 副教授, 博士毕业于新加坡国立大学, 研究方向为音乐信号处理、机器学习.

白光一 男, 1963 年 9 月出生, 吉林长春人. 研究员、博导, 博士毕业于日本九州大学, 研究方向为嵌入式数据库、音乐信号处理.