

基于差分全相位 MFCC 的音符起点自动检测

关 欣, 李 儒, 田洪伟

(天津大学电子信息工程学院, 天津 300072)

摘 要: 针对现有的音符起点自动检测方法难以适用于多类音乐信号, 计算复杂度较高等问题, 提出一种基于差分全相位 MFCC 的检测算法。通过全相位预处理减小频谱泄露引起的频谱模糊, 差分 Mel 频率倒谱考虑人耳对音乐不同频率响应的非线性特性和音乐信号的动态音乐特征。实验结果表明, 与公认综合检测效果好的 HFC 和 ICA 等方法相比, 该方法计算复杂度小, 适用音乐信号类型广, 具有更优的综合检测性能。

关键词: 音符起点检测; Mel 频率倒谱系数; 全相位预处理; 音乐信息检索

Note Onset Automatic Detection Based on Differential All Phase MFCC

GUAN Xin, LI Qiang, TIAN Hong-wei

(School of Electronic and Information Engineering, Tianjin University, Tianjin 300072)

【Abstract】 To reduce the limited range when used to multiple kinds of music signals and high computing complexity of existing note onset automatic detection methods, a note onset detection algorithm based on differential all phase Mel Frequency Cepstrum Coefficients(MFCC) is presented. All phase preprocessing alleviates spectrum ambiguity because of spectrum leakage. And differential Mel frequency cepstrum considered nonlinearity of human ear when responding to different frequencies and dynamic property of music signal. Experimental result demonstrates this algorithm is fit for multiple kinds of music and has perfect general detection performance with lower computing complexity compared with those of High Frequency Content(HFC) and Independent Component Analysis(ICA).

【Key words】 note onset detection; Mel Frequency Cepstrum Coefficients(MFCC); all phase preprocessing; Music Information Retrieval(MIR)

1 概述

随着互联网和数字图书馆中音乐文件的激增, 根据文件名和文本描述的方法检索音乐越来越不能满足实际需求, 因此, 基于内容的音乐信息检索(Music Information Retrieval, MIR)一经提出就成为数据挖掘和模式识别领域的研究热点。

在音乐信息检索系统中, 影响检索结果的首要因素是音符起点的精确检测, 若此步出现较大误差, 即使忽略整个乐段与输入相差的音符, 仍将使相似度较小, 大大影响检索性能。且音乐内涵的表达在很大程度上依赖于音符转换或音色变化, 这些事件的自动检测不仅是音乐信息检索中关键研究问题, 也是音乐信号内容发布、压缩、索引的基础。

目前主要有2类音符起点检测方法: (1)利用信号的幅度包络、短时能量、幅频、相位、时频表示等特征检测音符起点; (2)基于统计信号模型, 根据模型变化点检测突发信号^[1-3]。文献[1]比较了现有各类音符起始点检测算法, 发现对于有调击弦打击类乐器(如钢琴)、有调弓弦非打击类乐器(如提琴)、无调打击类乐器(如鼓)及复杂混合乐音, 加权高频成分的算法 HFC(High Frequency Content)和自适应统计模型 ICA(Independent Component Analysis)算法具有更为优良的总体检测性能^[4-5]。对于音符起点检测问题, 频谱及能量提供了主要信息。提高谱分析精度并结合人听觉系统的非线性特性, 将改善音符起点检测的精度。因此, 结合听觉滤波器组, 改进 Mel 频率倒谱系数(Mel frequency cepstrum coefficients, MFCC), 提出了全相位差分 MFCC, 由此得到的音符起点检

测函数具有适合音乐类型广、执行效率高等优点。

2 全相位差分 MFCC

与音高对应的基音和谐波分量的时间分布是确定音符起始点检测函数的基础, 而音乐信号是各组成成分的乘积或卷积信号, 更适宜采用倒谱分析方法, 分离音乐信号中乐器的激励及响应, 得到乐音的共振特征和基音频率。同时对乐音的处理还需考虑人耳听觉滤波器组的听觉掩蔽等感知特性。

在 Mel 频率倒谱分析时, 即使作加窗平滑处理, 截断输入音乐信号引起的频谱泄露, 仍然影响音乐信号基音、谐波频率的分辨率。而且, MFCC 为音乐的静态特征参数, 需改进以检测音符的动态变化。

2.1 全相位预处理

有限的音乐信号采样值、噪声、截断音乐采样数据的随意性、相邻采样值中信息的缺失, 使音乐信号的频率结构特征、乐器的真实响应函数变坏, 甚至不再能反映音乐信号固有的频率结构特征, 所以对频率成分和结构异常丰富的音乐信号来说, 采用简单的加窗处理不能得到好的倒谱系数。从统计学的角度, 为提高倒谱分析算法的频率分辨率, 可以利

基金项目: 国家自然科学基金资助项目(60802049)

作者简介: 关 欣(1977—), 女, 讲师、博士, 主研方向: 音乐、生物信号处理, 统计学习, 盲信号处理; 李 儒, 副教授、博士; 田洪伟, 硕士研究生

收稿日期: 2010-01-10 **E-mail:** guanxin@tju.edu.cn

用包含各相位样本点的时间序列作为倒谱分析的输入信号。全相位预处理方法考虑了包含输入样本点 $x(n)$ 的所有长度为 N 的时间序列^[6]。 x_i 是包含数据 $x(n)$ 的各时间序列集合, 为了避免时序错乱, 先周期延拓, 同时, 考虑到馈入系统的数据向量和叠加后的数据矢量应保持为 N , 作截断求和处理。以 $N=3$ 为例, 如图 1 所示。

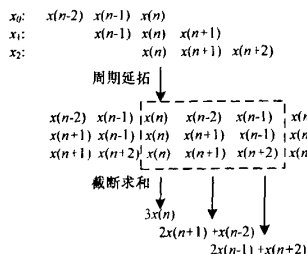


图 1 全相位预处理

2.2 全相位预处理对频谱分析精度的影响

单频复指数序列 $\{x(n) = e^{j\omega_0 n}, -N+1 \leq n \leq N-1\}$ 的全相位 FFT 谱表达式为 $Y(k) = W_c(e^{j\omega_0 - \omega_k})$, 其中, W_c 为加权系数。与传统不加窗的 DFT 谱 $X_N(k)$ 有如下关系:

$$Y(k) = e^{j\theta} W_c(e^{j(\omega_0 - \omega_k)}) = e^{j\theta} |R_N(e^{j(\omega_0 - \omega_k)})|^2 = e^{j\theta} |X_N(k)|^2 \quad (1)$$

全相位 FFT 振幅谱与传统 FFT 振幅谱之间的平方关系使全相位 FFT 与传统 FFT 相比能更好地抑制频谱泄漏, 减弱各基音、谐波频率旁瓣之间的相互影响, 更适用于原始音乐信号的倒谱分析, 能更真实地反映音乐信号本来的频率成分和结构。

2.3 全相位差分 MFCC

全相位差分 MFCC 计算步骤如下:

(1) 对每帧序列进行全相位预加重处理, 经离散 FFT 变换, 取模的平方 $S(n)$ 。

(2) 采用 M 个滤波器的听觉滤波器组 $H_m(n)$, 计算 $S(n)$ 经此滤波器的取值, 将频率转换到 Mel 频率, 得到 N 个参数 p_m , $m=1, 2, \dots, N$ 。

$$p_m = H_m(n) S(n) \quad (2)$$

(3) 计算 p_m 的自然对数, 得到 m_j , $j=1, 2, \dots, N$ 。

$$m_j = \ln(p_m) \quad (3)$$

(4) 对 m_j 进行离散余弦变换, 得到 $APMFCC_i$ 参数, $i=1, 2, \dots, N$ 。

$$APMFCC_i = \sqrt{\frac{2}{N} \sum_{j=1}^N m_j \cos\left(\frac{\pi j}{N} (j-0.5)\right)} \quad (4)$$

(5) 计算差分全相位 MFCC 参数

$$DAPMFCC_k = (APMFCC_{k+1} - APMFCC_k) / \Delta t \quad (5)$$

其中, $APMFCC_k$ 为第 k 帧参数向量。

Mel 域滤波利用与人耳听觉特性相似的滤波器组对音乐信号的幅度平方谱进行平滑。对数操作压缩音乐频谱的动态范围, 并将频域中的乘性成分变成对数谱域中的加性成分, 滤除乘性噪声。离散余弦变换主要用来对不同频段的频谱成份进行去相关处理, 使各向量之间相互独立, 并将 $APMFCC$ 压缩至 20 个。Mel 域取 40 个频段相等的滤波器以覆盖音乐 20 Hz~20 kHz 的频率范围。

3 音符起点检测

音符起点检测算法一般可以分为预处理、信号约简、峰

值提取 3 个阶段。其中, 预处理部分将原始信号的不同方面予以加强或减弱, 一般包括子带分割和暂稳态分离处理。预处理后的原始信号经过某种变换约简得到的检测函数表示出原始信号中的暂态变化, 这是音符起点检测的研究重点。检测函数应包含较少噪声, 尽可能反映不同种类乐器信号音符起始特点。最后提取检测函数的峰值, 即为音符的起始点。

4 DAPMFCC, HFC, ICA 音符起点检测函数比较

仿真实验比较了 DAPMFCC 与文献[1]中目前综合检测效果较佳的 HFC 和 ICA 算法的性能。实验中采用了具有典型意义的两类音乐信号: 有调击弦打击类乐器音乐片段帕赫贝尔《卡农》中的前 4 个小节、有调弓弦非打击类乐器音乐片段巴赫《无伴奏小提琴组曲三号第一乐章》中前 2 个小节, 采样率均为 44.1 kHz。

为客观地比较 3 种算法, 采用了相同的预处理方法、帧长和帧间重叠, 且峰值提取均采用二次插值算法。图 2(a)、图 3(a)中的起点根据 DAPMFCC 检测函数自动标识, 图 2(c)、图 3(c)中 HFC 检测函数动态范围较大, 故取对数。前两类信号还给出了相应的五线谱。

从图 2 可以看出, 对于有调打击类音乐 DAPMFCC, HFC, ICA 方法的峰值有区分度。ICA 检测函数包含较多噪声, HFC 检测函数相对平滑。而 DAPMFCC 方法在音符起点处有显著峰值, 同时噪声水平较低。图 2 中 ICA, DAPMFCC 方法正确地检测和标识了 8 个音符的起点。

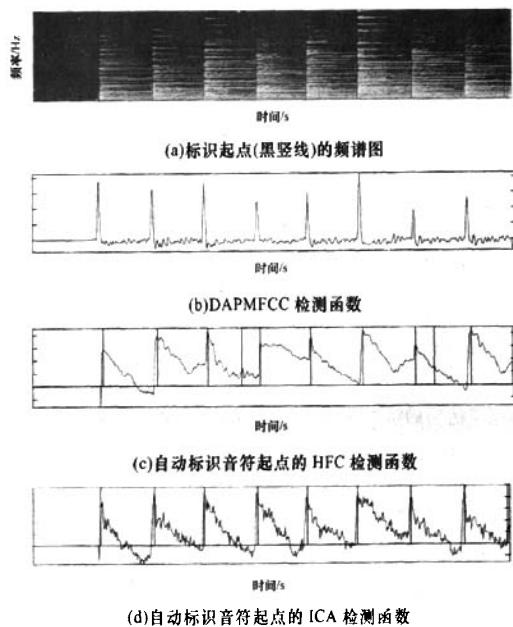


图 2 有调击弦打击类

与打击类音乐信号相比, 有调非打击类音乐起点能量变化缓慢, 检测相对困难。图 3 所示的 3 种方法中, 只有 DAPMFCC 方法给出了起处的显著变化。此时 HFC 对高频成分的加权不足以区分各个音符, 影响了检测函数的精度和准确度。

而 ICA 中, 噪声的影响愈加明显, 影响了后续的峰值提取和起点定位。若增加去噪环节改善 ICA 方法的性能, 又可能影响峰值的显著性。

(下转第 29 页)

份特征识别，3 s 时的数据有较好的效果。

4.3 特定波段的脑电密码计算

单侧肢体运动和想象运动也可以激活初级感觉运动区，在对侧大脑半球产生事件相关去同步现象，在同侧大脑半球产生事件相关同步化^[5-6]，而运动想象主要是以和节律的 ERD 和 ERS 为基础的，因此，分别以 μ (8 Hz~13 Hz)， β (14 Hz~20 Hz)， γ (21 Hz~30 Hz) 节律等频段分别进行身份识别，识别结果如图 7 所示。

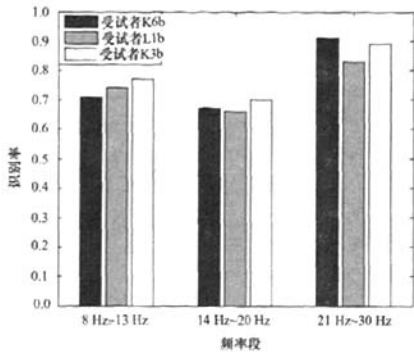


图 7 3 个不同频段的数据识别率

可以看出，特征波段如 21 Hz~30 Hz 波段对受试者的运动识别率贡献最大，几乎接近所有频段的识别率。同时也表明，对特定的运动想象或刺激，与认知心理相关的某些频段可能对身份识别起到重要作用。

5 结束语

本文对 8 Hz~13 Hz，14 Hz~20 Hz，21 Hz~30 Hz 波段的脑电运用小波分解和人工神经网络的方法进行了计算，得出了不同受试者不同的识别率。该结果有助于深化对于脑电信号特性的了解，下一步的工作重点是改善利用脑电信号进行密码计算的可靠性和准确性。

参考文献

[1] Huang P, Harris C, Nixon M. Human Gait Recognition in Canonical Space Using Temporal Templates[J]. Vision Image and Signal Processing, 1999, 146(2): 93-100.

[2] 张佳佳, 盘宏斌, 黄辉先. 基于小波对比度和神经网络的图像隐写方法[J]. 计算机工程, 2010, 3(5): 154-158.

[3] Bao Xuecai, Hu Jianfeng. Phase Synchronization for Classification of Motor Imagery EEG[J]. Journal of Information & Computational Science, 2008, 5(2): 949-956.

[4] 尹晶海, 蒋德荣, 穆振东, 等. 基于运动想象的脑机接口残疾人游戏辅助平台的设计与应用[J]. 中国组织工程研究与临床康复, 2008, 12(35): 6839-6843.

[5] Poulos M. On the Use of EEG Features Towards Person Identification via Neural Networks[J]. Medical Informatics & the Internet in Medicine, 2001, 26(1): 35-48.

[6] Pfurtscheller G. Motor Imagery and Direct Brain-computer Communication[J]. Proceedings of the IEEE, 2001, 89(7): 1123-1134.

编辑 金胡考

(上接第 26 页)

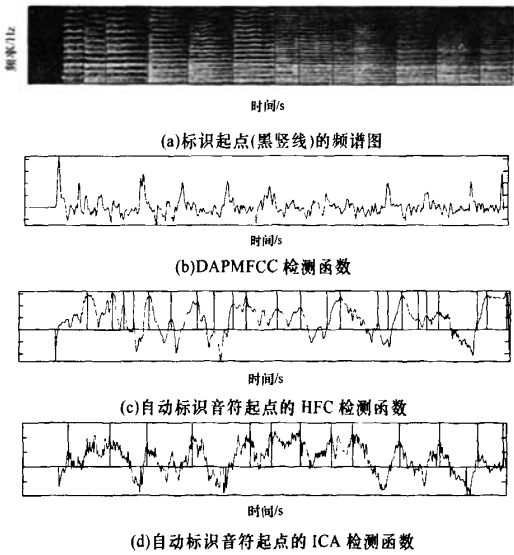


图 3 有调弓弦非打击类

在更多音乐样本上的实验发现，DAPMFCC 方法对 2 类音乐信号具有最优的综合检测性能，ICA 方法次之，但由于模型训练和参数拟合的时间、空间复杂度较高，计算时间约为 DAPMFCC 方法的 2 倍。HFC 方法简单、运算时间短，但综合检测性能较差。音符起点的自动检测还与峰值提取算法相关，改进峰值提取算法，可进一步改善音符起点检测性能。

5 结束语

基于某种简单参数的音符起点检测算法很难适于各类音乐。本文提出的算法结合了高精度频谱分析和人类听觉系统特性，提高了音符起点检测的精度和准确率，适用于更多音乐类型。若采用更复杂有效的预处理，利用各种音符音长的统计特性等先验知识，针对不同类型音乐，改进峰值提取算法，都将进一步改善音符起点检测的准确度。此外，模仿人类感知系统工作原理，结合不同算法优点构造分级混合音符起点检测方案也有望提高音符起点检测的准确度。

参考文献

[1] Bello J P, Daudet L, Abdallah S, et al. A Tutorial on Onset Detection in Music Signals[J]. IEEE Trans. on Speech and Audio Processing, 2005, 13(5): 1035-1047.

[2] 马志欣, 周利华. 哼唱检索中一种新颖有效的哼唱信息处理方法[J]. 计算机科学, 2006, 33(5): 169-173.

[3] 秦 静, 周明全, 王醒策, 等. 基于动态分割和加权综合匹配的音乐检索算法[J]. 计算机工程, 2007, 33(13): 194-197.

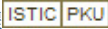
[4] Masri P. Computer Modeling of Sound for Transformation and Synthesis of Musical Signal[D]. Bristol, UK: University of Bristol, 1996.

[5] Abdallah S A, Plumbley M D. Probability as Metadata: Event Detection in Music Using ICA as a Conditional Density Model[C]//Proc. of ICA'03. Nara, Japan: [s. n.], 2003: 233-238.

[6] 黄翔冬. 全相位数字信号处理[D]. 天津: 天津大学, 2006.

编辑 金胡考

基于差分全相位MFCC的音符起点自动检测

作者: [关欣](#), [李镛](#), [田洪伟](#), [GUAN Xin](#), [LI Qiang](#), [TIAN Hong-wei](#)
作者单位: [天津大学电子信息工程学院, 天津, 300072](#)
刊名: [计算机工程](#) 
英文刊名: [COMPUTER ENGINEERING](#)
年, 卷(期): 2010, 36(11)
被引用次数: 1次

参考文献(6条)

1. [Bello J P;Daudet L;Abdallah S](#) [A Tutorial on Onset Detection in Music Signals](#)[外文期刊] 2005(05)
2. [马志欣;周利华](#) [哼唱检索中一种新颖有效的哼唱信息处理方法](#)[期刊论文]-[计算机科学](#) 2006(05)
3. [秦静;周明全;王醒策](#) [基于动态分割和加权综合匹配的音乐检索算法](#)[期刊论文]-[计算机工程](#) 2007(13)
4. [Masri P](#) [Computer Modeling of Sound for Transformation and Synthesis of Musical Signal](#) 1996
5. [Abdallah S A;Plumbley M D](#) [Probability as Metadata:Event Detection in Music Using ICA as a Conditional Density Model](#) 2003
6. [黄翔冬](#) [全相位数字信号处理](#) 2006

本文读者也读过(10条)

1. [桂文明](#), [邵曦](#), [刘睿凡](#), [GUI Wenming](#), [SHAO Xi](#), [LIU Ruifan](#) [一种新型音符起始点检测算法](#)[期刊论文]-[计算机工程与应用](#)2011, 47(24)
2. [刘若伦](#), [K. Feurer](#), [N. Griffith](#), [J. Walker](#), [LIU Ruo-lun](#), [K. Feurer](#), [N. Griffith](#), [J. Walker](#) [结合多通道处理的智能音符起点检测方法](#)[期刊论文]-[声学技术](#)2006, 25(6)
3. [兰帆](#), [黄明和](#), [LAN Fan](#), [HUANG Ming-he](#) [一种改进旋律匹配算法在MIDI演奏系统中的应用](#)[期刊论文]-[计算机与现代化](#)2009(6)
4. [曹大平](#), [陈正想](#), [李绍周](#) [器件型三轴磁通门磁传感器及其数字化技术](#)[会议论文]-2001
5. [马文婷](#) [基于MEGA64单片机的智能搬运机器人设计](#)[期刊论文]-[科技信息](#)2011(30)
6. [于迎霞](#) [基于语音学知识的鲁棒性两级语音起点检测方法](#)[期刊论文]-[电声技术](#)2004(5)
7. [张伟伟](#), [杨鼎才](#), [Zhang Weiwei](#), [Yang Dingcai](#) [用于说话人识别的MFCC的改进算法](#)[期刊论文]-[电子测量技术](#)2009, 32(8)
8. [赵荣阳](#), [杨祥](#), [张远翼](#), [Zhao Rongyang](#), [Yang Xiang](#), [Zhang Yuanyi](#) [家庭智能控制系统设计与实现](#)[期刊论文]-[电脑与电信](#)2007(6)
9. [张文希](#), [胡志刚](#), [ZHANG Wen-xi](#), [HU Zhi-gang](#) [基于FPGA的语音MFCC参数提取](#)[期刊论文]-[制造业自动化](#)2010, 32(8)
10. [张秋月](#) [基于相位特征的音乐音符起点检测](#)[学位论文]2010

引证文献(1条)

1. [郭奕](#), [徐红兵](#), [唐继勇](#), [何俐](#) [自动音乐标注系统综述](#)[期刊论文]-[计算机应用研究](#) 2011(4)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_jsjgc201011009.aspx