

一个交互式节拍跟踪和可视化系统

Simon Dixon

奥地利人工智能研究所，维也纳 电邮：simon@oefai.at

摘要

本文介绍了BeatRoot，一个对音频或MIDI数据进行自动节拍跟踪的系统，并对数据和结果进行图形和音频表示，作为互动界面的一部分，用于纠正错误或为节拍跟踪选择其他节拍级别。图形界面显示输入的数据和计算的节拍时间，并允许用户添加、删除和调整节拍时间，然后根据用户的输入自动重新追踪剩余的数据。该系统还提供由原始输入数据组成的音频反馈，并伴随着打击乐器在计算出的节拍时间上发出的声音。该系统的核心是一个节拍跟踪算法，该算法根据成对的音符开始时间之间的各种时间长度的出现频率来估计节奏，然后使用多重假设搜索来找到最符合可能的节奏的音符开始序列。这个系统的主要应用是分析音乐表演中的节奏和时间，尽管节拍跟踪算法本身已被证明至少与其他最先进的系统一样表现良好。

1 简介

在过去十年中，在开发自动寻找音乐表演中的节拍的计算机系统方面取得了重大进展（Rowe 1992；Rosenthal 1992；Large和Kolen 1994；Large 1996；Goto和Muraoka 1995；Goto和Muraoka 1999；Scheirer 1998；Dixon 2000；Dixon和Cambouropoulos 2000；Dixon 2001a；Cemgil, Kappen, De-sain, and Honing 2000；Cemgil, Kappen, Desain, and Honing 2001）。然而，没有一个这样的系统接近于音乐家在同一任务上的能力，因此自动节拍检测，或节拍跟踪，因为它通常被称为，还没有在节拍跟踪本身以外的许多应用中被采用。我们展示了一个系统，它对音频或MIDI数据进行自动节拍跟踪，并对结果进行图形和音频表示，作为互动界面的一部分，用于纠正错误或为节拍跟踪选择其他节拍水平。图形界面显示输入数据和计算的节拍时间，并允许用户

该系统可以添加、删除和调整节拍时间，然后根据用户的输入自动重新追踪剩余的数据。该系统还提供由原始输入数据组成的音频反馈，伴随着打击乐器在计算（或调整）的节拍时间上发出的声音。

该系统的核心是一个节拍跟踪算法，它根据音符开始时间之间的各种持续时间的出现频率来估计节奏，然后使用多重假设搜索来找到最符合可能的节奏的音符开始顺序。不需要输入数据的先验知识或具体特征；所有需要的信息都是从数据中获得。通过这种方式，节拍跟踪系统能够对许多不同的音乐风格、节奏和节拍表现良好，包括表现力强的古典和爵士音乐。

这个系统的主要应用是在音乐表演中的时间分析。然而，还有其他研究和应用领域也可以使用该系统的组成部分。节拍跟踪系统在任何进行音频内容分析的系统的重要组成部分，例如在多媒体数据库和图书馆中对音频数据进行索引和基于内容的再三评价，乐谱提取和音乐的自动转录。节拍跟踪的另一个应用是设备的同步，如灯光、电子乐器、录音设备、计算机动画和视频与音乐数据的同步。这种同步可能是多媒体或人机互动表演或工作室后期制作工作所必需的。对能够以“音乐智能”方式处理数据的系统的需求越来越高，而节拍的解释是音乐智能最基本的方面之一。

图形化的前端对于可视化性能和节拍跟踪数据也很有用。该系统的低耦合性使得替代的节拍跟踪算法能够以最小的努力进行连接，从而使该系统成为不同节拍跟踪算法的理想测试和评估框架。

在本文的其余三节中，我们描述了节拍跟踪算法，然后是用户界面，最后讨论了该系统目前的主要应用--性能计时分析。

2 节拍跟踪算法

本节简要描述了系统是如何进行节拍跟踪的。关于节拍跟踪算法的完整描述，见Dixon (2001a)。对输入数据进行离线处理，以检测突出的节奏事件，并对这些事件的相对时间进行分析，以产生不同节拍水平的节奏假设。基于这些节奏假设，一个多假设搜索找到了与节奏事件最匹配的节拍时间序列。尽管该系统不是实时运行的，但目前正在扩展以提供实时能力。该系统已经足够高效：CD质量的音频数据的处理时间不到数据长度的1/5（一首5分钟的歌曲的处理时间不到1分钟），而MIDI数据的处理速度是2到10倍，取决于音符密度。

输入数据可以是数字音频或符号表示，如MIDI。音频数据由一个时域起音检测算法进行预处理，该算法根据振幅包络的斜率峰值计算起音时间。起始时间由每个峰值的振幅包络的大小来加权。MIDI数据需要较少的处理，因为起音时间可以直接从数据中提取。然而，要进行一些预处理：音符的起始集被分组为和弦（节奏事件），并根据组成音符的音高、速度和持续时间对其进行加权。这些加权是对节奏事件的感知显著性的粗略估计，这已被证明是影响节拍跟踪系统性能的一个重要因素（Dixon和Cambouropoulos 2000）。

处理的第一个主要阶段使用了一个关于发作间期的聚类算法，即不一定是连续事件的发作间期的时间长度。这些集群代表了重要的节拍单位，通过研究集群之间的关系，产生了一套节奏混合模式。这些信息被送入系统的第二阶段，即跟踪阶段，该阶段试图将所执行的事件与代表每个节奏假设的灵活的事件网格相一致。这个阶段使用一个多代理架构，其中每个代理代表一个关于节拍的节奏和对准（相位）的假设，而代理根据他们的假设与数据的拟合度来评估他们的性能。被评估为最佳的代理将其节拍跟踪解决方案作为节拍跟踪系统的输出。

节拍跟踪系统已经在大型的数据库中进行了测试。在几乎所有的情况下，由系统第一阶段提出的节奏假设集包含了正确的（每个形成的）节奏，尽管排名最高的假设并不总是对应于记谱的节拍水平。第二阶段的处理，即计算节拍时间和选择最佳代理，不太稳健，但当错误发生时

关于节拍对齐（例如，当系统跟踪相对于真实节拍的半拍时），系统通常可以迅速恢复，恢复正确的节拍跟踪，而不需要求助于任何高水平的音乐知识。在最近的一项比较研究中（Dixon 2001b），该系统被证明比Cemgil等人（2001）的最先进的系统在使用他们的测试数据时表现得略好。

3 音频-图形用户界面

这个系统最独特的特点是它的界面，它提供了可视化、声波化、数据编辑、文件管理和节拍跟踪系统的控制。图形显示以钢琴卷标显示MIDI数据（图1），以平滑的振幅包络显示音频数据，并在显示中标明检测到的起始点（图2）。音频数据的显示也可以包含一个可选的规格图（未显示）。估计的节拍时间显示为垂直线，节拍间的时间间隔在显示的顶部。每种类型的数据都有颜色编码，这样显示就很容易阅读。一个滚动条控制数据的可见时间窗口，一个缩放功能允许调整时间分辨率。控制面板可以方便地访问系统的所有功能，如加载和保存输入数据和结果、音频播放和节拍跟踪本身。

节拍时间可以直接通过拖动代表节拍的垂直线来编辑（使用鼠标按键1）。其他的鼠标按键用于在光标位置插入和删除节拍。通过点击显示屏底部的时间轴，可以选择数据的某一部分进行回放、删除或重新追踪节拍。通过这种方式，用户可以在自动和手动节拍跟踪之间交替进行，以获得准确的表演节奏跟踪。

音频播放功能产生一个音频或MIDI轨道（与输入数据的类型相对应），包含一个用户可选择的打击乐器，在每个确定的节拍时间发出声音。这与输入数据相结合，并通过声卡进行播放（MIDI输出通过软件合成器发送）。合并的数据也可以保存到文件中，以便在其他时间播放。

选项框提供了对节拍跟踪系统的所有内部参数的访问。这些参数都有默认值，但可以被覆盖，例如调整系统对节奏变化的敏感度。

节拍跟踪系统是用C++编写的（大约10000行代码），界面是用Java编写的（大约1000行），在Linux操作系统上运行。该软件可从<http://www.oefai.at/>下载。

~simon/beatRoot用于非商业用途。

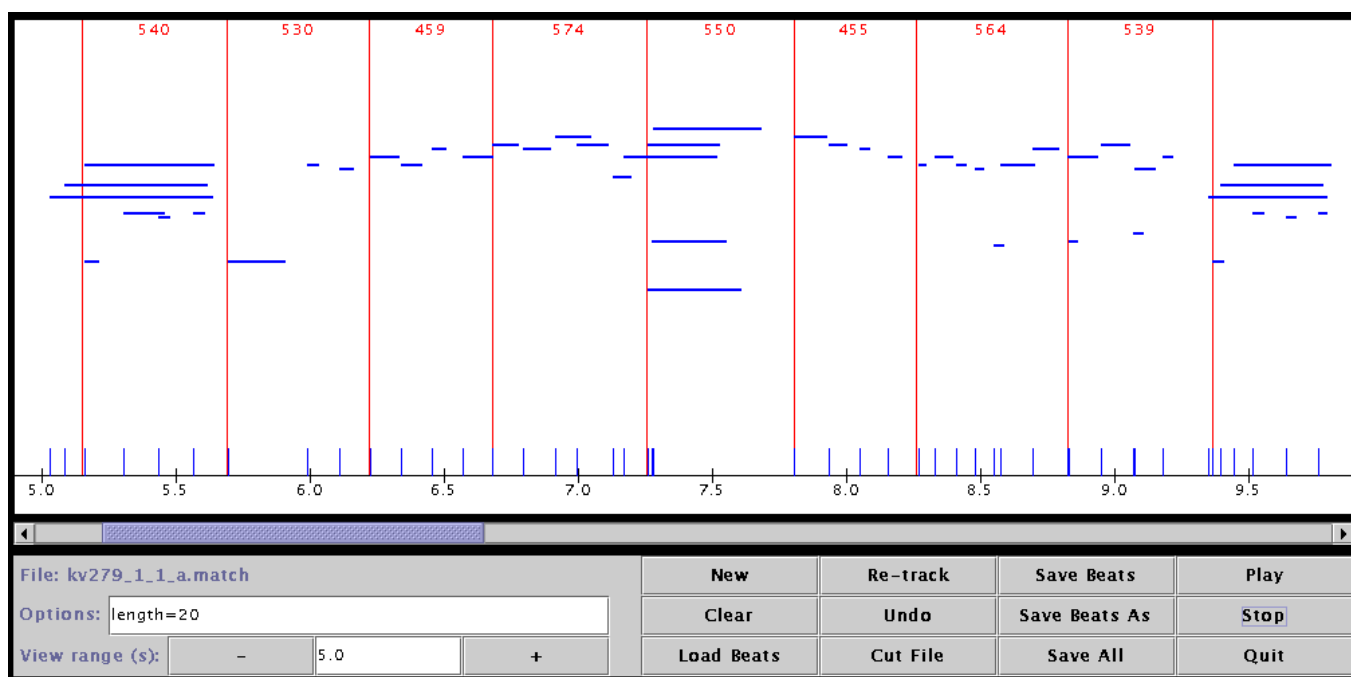


图1：使用MIDI输入数据的节拍跟踪器屏幕截图。MIDI数据以钢琴卷轴符号显示（蓝色），节拍时间以垂直线叠加（红色）。

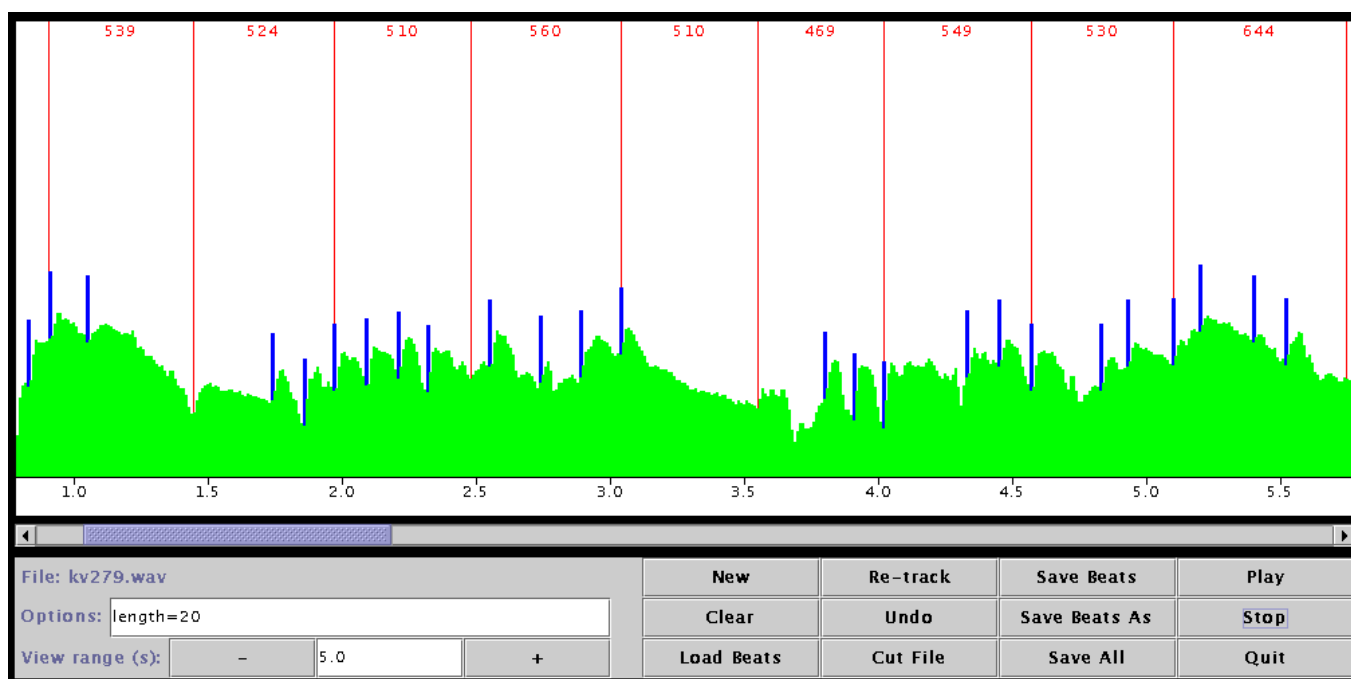


图2：使用音频输入数据的节拍跟踪器屏幕截图。振幅包络显示在显示屏的底部（绿色），系统检测到的起始时间用从振幅包络延伸出来的短垂直线标记（蓝色），节拍时间用长垂直线标记（红色）。

4 性能时间分析

在音乐表演的研究中,表演者对音乐作品的解释被调查,例如,他们对节奏和表达时机的选择。这些参数对于向听众传达结构和情感的形成非常重要(Clarke 1999)。通过找到音乐节拍的时间,我们可以自动计算出节奏和表演中的节奏变化。这就加速了分析过程,从而可以进行更广泛的研究。

作为一个自动分析音乐表现的工具,人们期望一个自动节拍跟踪系统是非常有用的,因为,例如,它可以直接从表演数据中生成节奏曲线。但是到目前为止,这种系统还没有被证明是实用的,因为它们不能避免繁琐的手工编辑计时数据。这种系统需要的一个必要的特点是,它要提供一个简单的编辑输出的方法,例如,通过一个交互式的图形用户界面。

我们的系统提供了输入数据、生成的节拍时间和节拍间的间隔(瞬时节奏)的图形可视化,并允许使用图形界面对节拍时间进行编辑(插入、删除和调整),以及在进行了修正后重新启动自动节拍跟踪。这为用户提供了一个完整的工具,用于从有表现力的表演数据中生成节奏曲线。该系统目前正被用于研究莫扎特钢琴奏鸣曲的CD表演中的节奏(Goebl和Dixon 2001),并且已经进行了初步的研究,分析了系统的分辨率和各种形式的反馈对用户节拍感知的影响(Dixon, Goebl, and Cambouropoulos 2001)。

在目前的工作中,正在改进音频起始点检测算法。在进一步的工作中,我们打算在两个主要方向上扩展该系统。第一个方向是允许系统使用乐谱信息来指导其节拍跟踪;这将大大减少乐谱可用时节拍跟踪的错误率,并大大加快对同一作品的多种解释的研究。第二个计划的扩展是创建一个乐谱提取系统,通过增加诸如量化、节拍识别、音符拼写、部分分离以及乐谱显示和编辑等功能,从而可以从表演数据中交互式地生成一个符号乐谱。

鸣谢

这项研究是START计划Y99-INF的一部分,由奥地利联邦教育、科学和文化部(BMWK)资助。BMWK还为奥地利人工智能研究所提供财政支持。感谢Emilios Cambouropoulos、Gerhard Widmer和Werner Goebl对这项工作的贡献。

参考文献

- Cemgil, A., B. Kappen, P. Desain, and H. Honing (2000). 关于节奏跟踪。Tempogram表示和Kalmanteling.在2000年国际计算机音乐会议论文集中,第352-355页。国际计算机音乐协会。
- Cemgil, A., B. Kappen, P. Desain, and H. Honing (2001). 关于节奏跟踪。Tempogram表示和Kalmanteling. *Journal of New Music Research*.即将出版。
- Clarke, E. (1999). 音乐中的节奏和时间。In D. Deutsch (Ed.), *The Psychology of Music*, pp.473-500.学术出版社。
- Dixon, S. (2000). 一个轻量级的多代理音乐节拍跟踪系统。In *PRICAI 2000: 环太平洋国际人工智能会议论文集*, 第778-788页。Springer。
- Dixon, S. (2001a). 从表现性表演中自动提取节奏和节拍。 *Journal of New Music Research* 30(1).即将出版。
- Dixon, S. (2001b). 节奏跟踪器的经验性比较。在第八届巴西计算机音乐研讨会论文集中。将会出现。
- Dixon, S. and E. Cambouropoulos (2000). 用音乐知识追踪节拍。In *ECAI 2000: 第14届欧洲人工智能会议论文集*, 第626-630页。IOS出版社。
- Dixon, S., W. Goebl, and E. Cambouropoulos (2001). 从有表现力的音乐表演中提取节拍。在2001年音乐感知和认知学会会议(SMPC2001)上,安大略省金斯顿。将出现。
- Goebl, W. and S. Dixon (2001). 莫扎特奏鸣曲中的节奏等级分析。In *Proceedings of the VII International Symposium on Systematic and Comparative Musicology, Jyväskylä, Finland*.待发表。
- Goto, M. and Y. Muraoka (1995). 一个音频信号的实时节拍跟踪系统。In *Proceedings of the International Computer Music Conference*, pp.计算机音乐协会,加利福尼亚州旧金山。
- Goto, M. and Y. Muraoka (1999). 无鼓音频信号的实时节拍跟踪。 *Speech Communication* 27(3-4), 331-335。
- Large, E. (1996). 用非线性oscillator建立节拍感知模型。在认知科学协会第18届年会的会议记录中。
- Large, E. and J. Kolen (1994). 共振和音乐节拍的感知。 *连接科学* 6, 177-208。
- Rosenthal, D. (1992). 对人类节奏感知的模仿。 *计算机音乐杂志* 16 (1), 64-76。
- Rowe, R. (1992). 机器听觉和用cypher作曲。 *计算机音乐杂志* 16(1), 43-63。
- Scheirer, E. (1998). 声学音乐信号的节奏和节拍分析。 *美国声学学会*

杂志103 (1) , 588-601。