

## 实时跟踪音乐节拍<sup>1</sup>

Paul E. Allen和Roger B.

Dannenberg计算机科学学院

卡内基梅隆大学

匹兹堡，宾夕法尼亚州15213，美国

电子邮件。Paul.Allen@cs.cmu.edu, Roger.Dannenberg@cs.cmu.edu

### ABSTRACT:

识别下拍子的时间位置是一项基本的音乐技能。观察到以前对这一过程进行自动化的尝试被限制在对节拍时间和位置的单一当前概念上，我们发现它们将无法预测节拍，并且在第一次犯错的时候不能恢复。我们提出一个新的模型，使用波束搜索来考虑对表演的多种解释。在任何时候，对节拍时间和位置的预测都是根据所考虑的许多解释中最可信的一种来进行的。

**简介。**识别下行节拍的时间位置是一项基本的音乐技能。即使是没有受过音乐训练的人，在听到音乐表演时也能随着节拍拍打脚。即使面对不寻常的节奏、音乐表现力和不精确的表演，人类通常也能轻松而准确地完成这项任务。一个完全通用的、自动的节拍跟踪器在许多任务中都会有很大的价值，比如人机即兴表演、音乐转录、音乐编辑和同步，以及（民族）音乐学研究。

节拍跟踪与Dannenberg[1]和Vercoc[2]的记分有区别。

[2]因为没有乐谱。此外，大多数节拍跟踪器不考虑音高，将输入限制在音符的起始时间和初始速度上。

**早期的节拍跟踪模型。**以前关于节拍跟踪的大部分工作都是针对不同音乐理解问题的研究的副产品，如转录和节奏、节拍和旋律的感知。Longuet-Higgins[3, 4, 5]描述了可能是第一次尝试在面对变化的节奏时跟踪表演的节拍。他对下行节拍的可变性使用了一个静态的容忍限度，并根据一个音符的开始时间是比预期的晚还是早来调整感知的节奏。这项工作是Longuet-Higgins提出的西方古典音乐旋律感知的计算心理学模型的一部分。

Chafe、Mont-Reynaud和Rush[6]在一个面向转录的任务中使用了与Longuet-Higgins类似的方法。Desain和Honing[7]报告了一种使用放松技术将音程量化为音符值的方法，尽管这种方法似乎不适合于实时应用。作为实时爵士乐即兴演奏跟踪任务的一部分，Mont-Reynaud[8]描述了一种新的实时节拍跟踪方法，该方法结合了信心的概念和历史机制，使用以前感知的节奏的加权平均来计算当前感知的节奏。该方法依靠少量的状态来表示节拍跟踪器的状态，并有一个简单的方法来更新该状态，当

---

<sup>1</sup>发表于。Allen和Dannenberg，"'实时跟踪音乐节拍'，在1990年国际计算机音乐会议上，国际计算机音乐协会（1990年9月），第140-143页。

收到新的音符数据。

我们选择通过实施Mont-Reynaud的实时方法开始我们的节拍跟踪实验。我们对它的性能不满意；特别是，它很迟钝，对节奏变化不够敏感，导致错误。为了使这个方法对节奏变化的反应更加灵敏，我们对历史和信心机制进行了参数化，期望能够找到对特定音乐风格或表演有良好效果的参数值。最初的历史机制使用指数衰减来计算以前节奏的权重。衰减率对节拍跟踪器的反应能力和稳定性有直接影响。在一个极端，即100%的衰减，跟踪器会忽略所有的历史，只对最新的音符数据做出反应，允许快速的节奏波动，但使跟踪器非常不稳定。在另一个极端，即0%的衰减，跟踪器将平等地考虑每一个历史位和当前状态，因此对最新的音符数据的考虑是轻微的，反应也是缓慢的。衰减率是我们研究的一个参数。

第二个参数化适用于置信度的计算。置信度的作用是表明节拍跟踪器认为一个给定的音符发生在下行节拍上的确定性，因此，间接地，节奏应该改变，以响应该开始的量。原始的信心函数是一个锯齿形的函数，在预测的下行节拍处的最大值为1.0，在两个相邻的预测下行节拍之间的一半处的最小值为0.0；其斜率为1.0或-1.0。我们把这个原始的置信度计算提高到一个指数参数（因此在参数值为1.0的情况下，我们达到的置信度计算与原始方法相同）。这使我们能够调节一个音符的起音对新节奏的影响，作为起音与预测的下拍的距离的函数。在这里，我们正在寻找一种折中的方法，即对不精确的下拍音符的容忍度和对正确演奏的非下拍音符的拒绝。

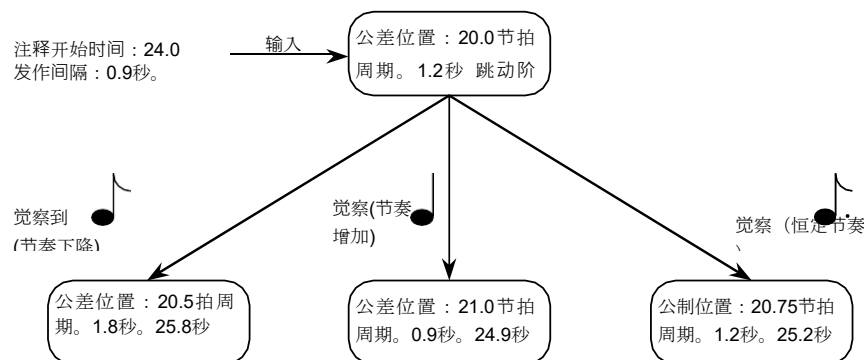
我们的希望是，通过搜索参数空间，我们可以找到能够实现良好性能的参数值。在实践中，我们发现，虽然我们经常能够找到大范围的参数值，以有效地跟踪更直接的表演，但同样的参数值范围对具有合理的节奏变化和/或更复杂的音乐结构的表演来说，结果相当杂乱无章。

**一种新的节拍跟踪方法。**我们将"节拍"的音乐概念扩展到两个方面，即周期和相位。在音乐表演中，我们会说一个节拍的周期是两个连续节拍之间的时间长度（节奏的倒数），而一个节拍的相位决定了一个节拍相对于表演时间的发生位置。一个节拍跟踪器必须包含状态信息，至少包括当前感知的节拍相位和周期。

前面讨论的方法的困难之一是，节拍跟踪器需要决定一个新的状态，代表它对当前节拍周期和阶段的信念。如果它在某一点上犯了错误，它很可能永远无法恢复，并且无法预测它犯错误的那一点之后的节拍。

我们建立了一种使用实时波束搜索的方法，使节拍跟踪器能够同时考虑几种可能的状态。该方法使用了与之前的方法类似的历史机制，但不再仅仅依靠置信度来指导它。取而代之的是，使用一个可信度衡量标准，以便在任何时候都有一组活动状态，代表迄今为止遇到的性能的最可信的解释。每个状态都有一个由启发式评价函数决定的可信度值，该函数衡量该状态所代表的特定解释的可信度。描述最可信解释的状态被用作进行节拍阶段和周期预测的基础，直到一个新的最可信状态出现。

发现。当演奏中遇到一个新的音符起音时，每个活动状态被扩展成几个新的状态，这些状态代表了由起音和活动状态描述的解释所产生的多种可能解释的子集（见图1）。那些没有音乐意义的解释会被立即丢弃。剩下的一组状态成为下一次扩展的新的活动状态。



因为这是一种实时方法，而且解释的数量是无穷无尽的，所以我们必须自由地应用剪枝技术来减少搜索的规模（见图2）。第一种修剪方法已经提到过了--

我们放弃那些对音乐没有意义的解释的状态。第二种技术涉及到状态的可信度和扩展的顺序：因为我们只想在新的输入准备好之前进行扩展，所以我们根据状态的可信度来扩展，那些可信度最高的状态会先被扩展。因此在扩展过程中，如果我们发现新的输入已经准备好被处理，我们可以立即停止扩展活动状态，因为我们已经扩展了最可信的状态。（当然，我们假设输入之间有足够的时间来扩展至少几个最可信的状态）。第三种修剪方法是将代表类似解释的多个状态合并为代表单一解释的一个状态。第四种也是最后一种修剪搜索的技术，通过限制单一活动状态的扩展可能产生的新解释的数量来限制分支因子。在这一点上，第一种修剪方法已经发生了一些修剪，但是由于给定一个状态和一个新的音符起始点，可能有无限多的解释，所以从一个状态中得到的扩展数量被一个任意的静态截止点限制。在这一点上，从单一状态的扩展是以这样的顺序进行的，即产生的解释是那些具有最小的节奏变化的解释。

这种节奏连贯性的概念--

节奏不会从一个音符到另一个音符发生剧烈的变化（如果它确实发生了重大变化，它应该总是在短时间内增加或减少）--

也被用于启发式评价功能。一个状态的启发式可信度值是三个参数的函数：（1）该状态与其父状态之间的节奏变化的大小（有利于小的变化而不是大的变化），（2）其父状态的启发式价值（对父状态有利的可信度值导致对子状态有利的可信度值），以及（3）代表音乐上可能但不可能的解释的静态惩罚（有利于简单、普通的音乐结构而不是复杂的结构）。

**结论。**我们对我们的光束搜索方法的初步结果相当满意。我们方法的一个实时原型正在运行，并在最初的测试中给出了良好的结果。它跟随一个业余键盘手演奏的简单旋律的节奏变化，只需很少的训练。我们已经采取了实际的、恒定节奏的人类表演，并且人为地

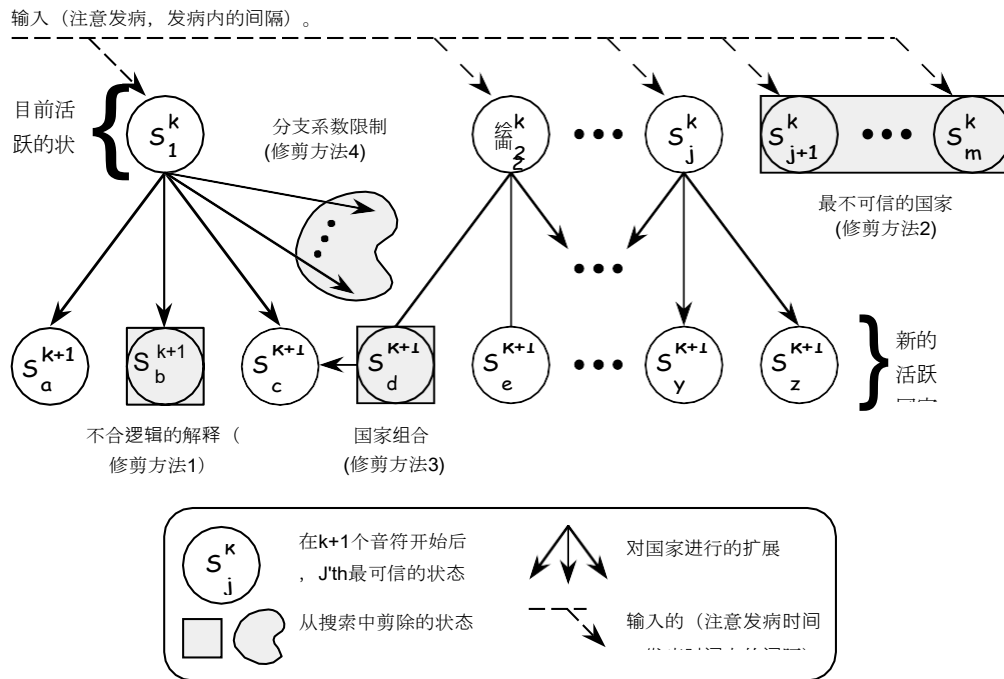


图2：波束搜索的一个周期与剪枝应用实例。

增加了速度变化的幅度（每拍超过50%的变化），以至于完全没有音乐感，但早期的非实时原型却成功地遵循了这种变化。

我们的方法的一个附加价值是，在对音乐表演进行成功的节拍跟踪后，可以从最终最可信状态的历史中恢复表演的转录。这个转录来自于从搜索树的根到最终最可信的音符的路径上产生的每个扩展的音符值。

#### 参考文献

1. Dannenberg, R. B. An On-Line Algorithm for Real-Time Accompaniment. Proceedings the International Computer Music Perception 1984年会议，国际计算机 Music Association, 1984, pp.193-198.
2. Vercoc, B. and M. Puckette. Synthetic Digital 排练。训练合成表演者。国际计算机会议记录 1985年音乐会议，国际 Honing, 计算机音乐协会，1985年，第。 275-278.
3. Longuet-Higgins, H. C. "感知的问题 旋律"。自然263 (1976)，646-653。
4. Longuet-Higgins, H. C. "The Perception of Music". 跨学科科学评论 32 (1978), 148-156.
5. Longuet-Higgins, H. C. and C. S. Lee. "The Perception of Musical Rhythms". of 11 (1982)，115-128.
6. Chafe, Chris, Bernard Mont-Reynaud, and Loren Rush. "迈向智能编辑的 Audio:音乐 的识别 构造"。计算机音乐杂志 6, (1982年春季)，30-41。
7. Desain, Peter and Henkjan "音乐时间的量化：一个 连接主义方法"。计算机音乐 杂志13, 3 (1989)，56-66。
8. Dannenberg, R. B. and B. Mont-Reynaud. Following an Improvisation in Real Time. 1987年国际会议的会议记录 计算机音乐会议，国际计算机音乐协 会，旧金山，1987，241-248页。