乐谱跟踪

乐谱跟踪，即实时判断乐曲演奏对应于乐谱中的位置，接收到数据时就实时处理，认为其后续演奏不可知。常应用于自动乐谱翻页、自动伴奏等。2003年Shazam公司提出了一种工业级的基于音乐指纹的音频检索算法[1]，在实际应用中有较高的准确率和较强的鲁棒性。2011年Müller M 分析了基于音乐指纹的乐谱跟踪算法的结构和性能，指出音乐指纹难以处理节奏变化，以及乐器本身的音调、音量差异等问题[2]。2012年Arzt A等提出了一种新的用于乐谱跟踪的音频指纹构造方法[3]，该方法改善了音乐指纹算法对节奏敏感的问题，但是只是考虑了节奏的线性变化。在实际钢琴练习中，节奏非线性变化是很常见的。因此，现有的基于音乐指纹的乐谱跟踪算法均没有完善地解决演奏中的节奏变化问题。

我们基于多音调检测结果和动态时间规整技术进行乐谱跟踪。乐谱跟踪模块实时接收多音调检测结果，并记录各个音符被连续检测到的时长。约定只有当某音符被连续检测到至少一定时长，才认为该音符被演奏了，比如约束钢琴演奏中音符的最短时长为60ms。此外，认为onset间隔小于上述最短时长的音符是应该同时被演奏的，只是需要同时演奏多个音符时，演奏者的多个手指按下钢琴键有微小的时差。每当乐谱跟踪模块监测到新演奏了音符时，就计算新演奏的音符对应于乐谱中的位置。

下面以乐谱为为例，介绍乐谱和演奏中可能出现的各种情形以及对应的乐谱跟踪处理方法。其中，序号①、②、③、④标识乐谱中依次的位置，一列数字表示一次演奏的音符，若一列有多个数字则表示多个音符同时演奏，不同的数字代表不同的音符。

若乐谱跟踪模块确定上一演奏对应于乐谱中第i个位置，表1列举了当检测到新演奏了音符时可能出现的各种情形。

表 1 确定上一定位时可能出现的情形

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 一 | 二 | 三 | 四 | 五 |
| 乐谱中第i个位置是否包含所有新演奏的音符 | √ | √ | √ | × | × |
| 乐谱中第i个位置上新演奏的音符是否已经被演奏过 | × | √ | √ |  |  |
| 乐谱中第i+1个位置是否包含所有新演奏的音符 |  | × | √ | √ | × |
| 新演奏对应于乐谱中的位置 | i | i | 待定 | i+1 | 待定 |
| 演奏示例 |  |  |  |  |  |

下面结合演奏示例对各种情形分别进行具体说明：

1. 乐谱中第i个位置为多个音符同时演奏，而演奏者演奏各音符时有较大的时差。
2. 重复演奏乐谱中第i个位置。
3. 乐谱中第i、i+1个位置包含相同的音符，当再次检测到两个位置均包含的音符被演奏时，无法确定第2次演奏的音符是重复演奏乐谱中第i个位置，还是继续演奏乐谱中第i+1个位置，需要等待更多的演奏信息来确定定位。如演奏为，
4. 若继续演奏乐谱中第i个位置有而第i+1个位置没有的音符，如演奏为，则可以确定是重复演奏；
5. 若继续演奏乐谱中第i+1个位置有而第i个位置没有的音符，如演奏为，则可以确定是继续演奏；
6. 若继续演奏乐谱中第i+1个位置，如演奏为，则判定第2次演奏的音符1对应于乐谱中第①或第②个位置均可。

一种更特殊的情形是乐谱中存在相同音符连续演奏，如示例乐谱中第②、③个位置。

1. 顺序演奏乐谱中下一个位置。
2. 演奏错误或跳跃，需要等待更多的演奏信息来确定定位。

不确定定位或者不确定上一演奏的定位时，计算演奏的音符与乐谱中各可能位置的匹配程度。优先考虑是继续演奏，若确定上一演奏对应于乐谱中第i个位置，则认为该次演奏可能对应于第i、i+1、i+2个位置；若不确定上一演奏的定位，则认为该次演奏可能是上次演奏可能对应的位置及其后面1、2个位置。一定时间后，若不确定定位的演奏与第i个位置后续的乐谱差异较大，则全谱搜索，重新计算与乐谱中所有位置的匹配程度。

一次演奏与乐谱中一个位置的匹配程度的计算方法为



其中，MatchRate为匹配程度，nMatch为匹配的音符的个数，即新演奏的音符乐谱中该位置有的个数，nNewNotes为新演奏的音符的总个数，nIgnoreOctave为忽略倍频后匹配的音符的个数。

通过动态时间归整（DTW，Dynamic Time Warping）算法寻找演奏最可能对应于乐谱中的位置，即匹配程度最高。

评价

由于未找到通用的评价钢琴演奏乐谱跟踪算法的数据集，人工制作测试集。邀请了1位钢琴学习者用MIDI键盘多次演奏《G大调小步舞曲》，要求其自然地演奏出钢琴练习过程中可能出现的所有情况，如应该同时演奏多个音符而演奏时有较大时差、重复、漏弹、多弹、错弹、跳跃、颤音等。直接导出演奏的MIDI文件，并用音色库合成WAV文件，乐谱为MIDI文件或musicXML文件。选取包含了上述所有的演奏错误类型的12次演奏。人工制作ground truth（正确结果）。

为了分别评测乐谱跟踪算法和实际系统中根据多音调检测结果进行乐谱跟踪的准确性，分别用演奏的MIDI文件和WAV文件进行乐谱跟踪，分别称为midi-to-midi乐谱跟踪和wav-to-midi乐谱跟踪。

midi-to-midi乐谱跟踪即对于MIDI文件中的各个音符，认为其后续演奏不可知，实时判断其对应于乐谱中的位置。准确率的计算公式为：



其中，precision为准确率，nCorr为判断正确的个数，nAll为MIDI文件中音符的总个数。测试结果见表2。

表 2 midi-to-midi乐谱跟踪准确率

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 文件序号 | nCorr | nAll | precision | 文件序号 | nCorr | nAll | precision |
| 1 | 113 | 113 | 1 | 7 | 104 | 108 | 0.963 |
| 2 | 92 | 96 | 0.958 | 8 | 97 | 113 | 0.858 |
| 3 | 93 | 103 | 0.903 | 9 | 131 | 147 | 0.891 |
| 4 | 108 | 115 | 0.939 | 10 | 109 | 119 | 0.916 |
| 5 | 108 | 117 | 0.923 | 11 | 106 | 109 | 0.972 |
| 6 | 115 | 115 | 1 | 12 | 114 | 122 | 0.934 |
|  |  |  |  | 合计 | 1290 | 1377 | 0.937 |

wav-to-midi乐谱跟踪，将定位结果表示为WAV中每10ms对应于乐谱中的位置。对于10ms的一帧，若其前后100ms对应于的正确结果中包含其定位结果，则认为定位正确。准确率的计算公式为：



其中，precision为准确率，nCorr为定位正确的帧数，nAll为音频的总帧数。测试结果见表3。

表 3 wav-to-midi 乐谱跟踪准确率

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 文件序号 | nCorr | nAll | precision | 文件序号 | nCorr | nAll | precision |
| 1 | 2345 | 2357 | 0.995 | 7 | 4629 | 5029 | 0.921 |
| 2 | 2358 | 2437 | 0.968 | 8 | 4690 | 5166 | 0.908 |
| 3 | 2284 | 2785 | 0.820 | 9 | 2932 | 3312 | 0.885 |
| 4 | 2379 | 2413 | 0.986 | 10 | 3808 | 3808 | 1 |
| 5 | 4173 | 4173 | 1 | 11 | 2358 | 2472 | 0.954 |
| 6 | 5375 | 5386 | 0.998 | 12 | 3484 | 3502 | 0.995 |
|  |  |  |  | 合计 | 40815 | 42840 | 0.953 |

[1] Wang A. An Industrial Strength Audio Search Algorithm[C]//ISMIR. 2003, 2003: 7-13.

[2] Müller M. New developments in music information retrieval[C]//Audio Engineering Society Conference: 42nd International Conference: Semantic Audio. Audio Engineering Society, 2011.

[3] Arzt A, Widmer G, Böck S, et al. Towards a complete classical music companion[C]//Proceedings of the 20th European Conference on Artificial Intelligence. IOS Press, 2012: 67-72.