

题目: Evaluation and combination of pitch estimation methods for melody extraction in symphonic classical music

作者: Juan J. Bosch, Ricard Marxer & Emilia Gómez

领域: 旋律提取

核心:

实现方法:

一、数据集建立

数据集主要来自交响乐、芭蕾舞的音乐录音。

- 为了验证摘录中包含一个清晰的旋律并确定笔记的确切顺序,我们通过记录演唱旋律的人物来收集人物注释
- 最终的数据集合包含 64 个音频摘录及其对应的 MIDI 格式的旋律注释
- 在执行提取之前,将文件转换为单声道合并左和右声道,以确保所有算法使用完全相同的材料
- 使用 MIDI 工具箱,并建立了旋律描述符: 音符密度, 音高极差, 音域范围等

- Density: amount of notes per second.
- Range: difference in semitones between highest and lowest note pitch.
- Tessitura: melodic tessitura based on pitch deviation from median pitch height (Von Hippel, 2000).
- Complexity (pitch, rhythm, mixed): expectancy-based model of melodic complexity (Eerola & North, 2000) based either on pitch or rhythm-related components, or on a combination of them together.
- Melodiousness: 'suavitatis gradus' proposed by Euler, which is related to the degree of softness of a melody, and is a function of the prime factors of musical intervals (Leman, 1995).
- Originality: Different measurement of melodic complexity, based on tone-transition probabilities (Simonton, 1984).

- 录音会议, 32 个音乐相关人士进行哼唱录音, 关注于音乐信息
- 分析唱歌旋律, 并选择音乐片段, 音乐片段满足有四个被试者哼唱的音符相同
- 选择音乐片段后, 我们手动转录参与者唱的音符, 调整音频的开销和偏移量。由于声道音域范围与播放主旋律的乐器的范围不同, 因此音符被转换为匹配音频。对于不同八度的乐器同时演奏旋律音符的节选, 我们通过最大化旋律轮廓平滑度(最小化音符之间的跳跃)来解决了歧义。

二、评估算法

- 十一种评估算法
- Multipitch 方法

最初计算音调显著函数,然后执行细化或跟踪以平滑俯仰轨迹。例如, Duan 估计以最大似然(ML)方法存在的音调,假设谐波位置的频谱峰值和其他地方的能量较低。然后,他们采用邻域优化方法在帧附近创建一个音调直方图,以消除瞬态估计,以及改进多项式估计。我们评估了这种方法的两个变体,一个是精细化(MP-DUA-Ref),另一个是非精细化的(MP-DUA)。在这两种情况下,我们没有使用多项式估计,因此这两种算法都会输出所有估计的音高。

- 提出一种混合方法 RNSCOMB, 结合了多个音高突出功能的输出, 然后执行峰值检测和基于邻域的细化。主要的假设是, 如果几个算法对“旋

律”音调的估计达成一致，那么估计更有可能是正确的。

d) RNSCOMB 方法是基于 COMB 方法的改进，COMB 方法是：

组合原始突出功能，使用高斯函数估计旋律间距之间的小差异。由于每个算法具有不同的音高显著范围，所以我们在组合它们之前对其进行归一化，使得给定帧中所有频率块的显著性之和等于 1（在概率原理之后）。最后，我们将每个方法(M)的显著值乘以不同的值($\alpha_M \in [0, 1]$)，允许加权组合。因此 $\alpha_M = 0$ 的值等同于不包括组合中的方法。图 7 给出了显著函数的组合的示例，其中具有相同权重($\alpha_{MAR}, \alpha_{DUR}, \alpha_{CAN} = 1$)的三个显著函数对于在 MIDI 音符 75 和 87 周围的音调的估计是一致的，而其中只有一个估计围绕 MIDI 音符 74 和 77 的间距。这给 75 附近的音高（组合）最大显著性，这对应于带注释的旋律音高。在加法之后，我们提取 N 个最高峰，它们之间的四分之一音调的最小差值。

论文的主要特点是数据集好，提出的旋律提取算法好，并设置了一套旋律算法的评估指标（有些复杂，没在读书笔记中呈现，论文第四部分）。