音乐中的两个重要元素，音乐构成和音乐表现。音乐构成关注的是音乐本身，它的指的是能够定义一首歌曲的音符。然而这对音乐家而言只是第一步。这些乐谱如何被演奏家演奏，这才是音乐工作的灵魂。因为演奏者演奏者演奏得各不相同，所以我们用音乐风格来描述个人化的音乐演奏。

1. 音乐风格定义

音乐风格很难定义，因为不能把音乐风格像音高一样被参数化。如果你曾听过很多经典的钢琴曲，那么你会发现一个新手和一个资深的钢琴家会奏出大不相同的强弱力度（dynamics），这里指的是音乐响度的变化。一个音符的响度可以通过控制敲击琴键的轻重程度来实现。在音乐符号中这些不同级别的强弱力度一般用意大利字母表示。这些字母被称为情感符。不同的人会有不同的感觉，所以这里的特定的强弱集合都有着各自的情感表现。

1. 强弱力度的应用

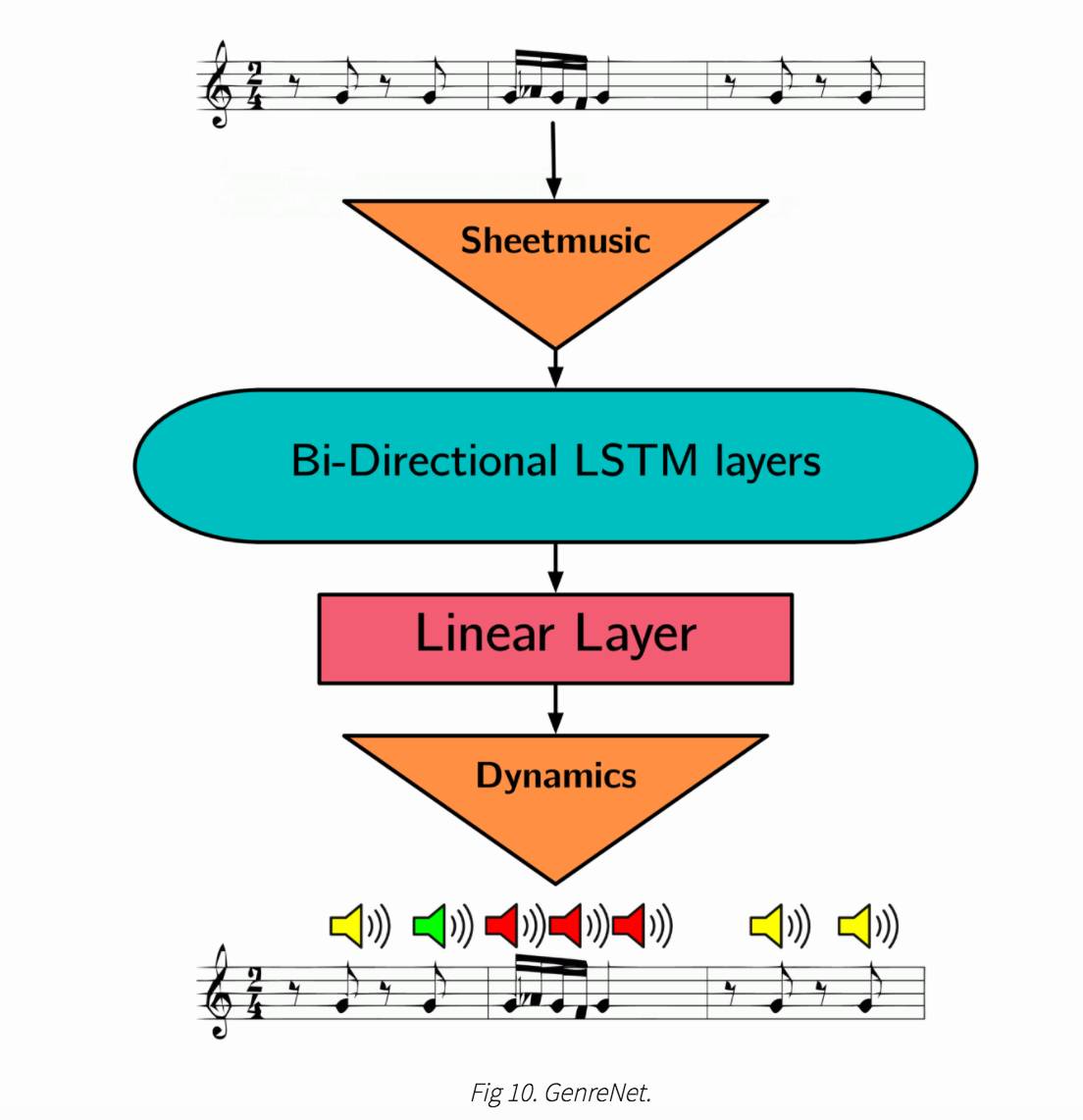
因此情感表现就意味着一组独一无二的强弱力度，而强弱力度也就成了风格的重要特征。在一些特定的音乐风格中是存在一些规则的，因此人们可以通过强弱力度来识别音乐风格。这也意味着人么可以用流派来归类音乐风格。

1. 架构

使用了双向-长短期记忆网络（Bi-Directional LSTM）。分别用两个网络实现对音乐流派和风格的分析，分别叫做 GenreNet 和 StyleNet。

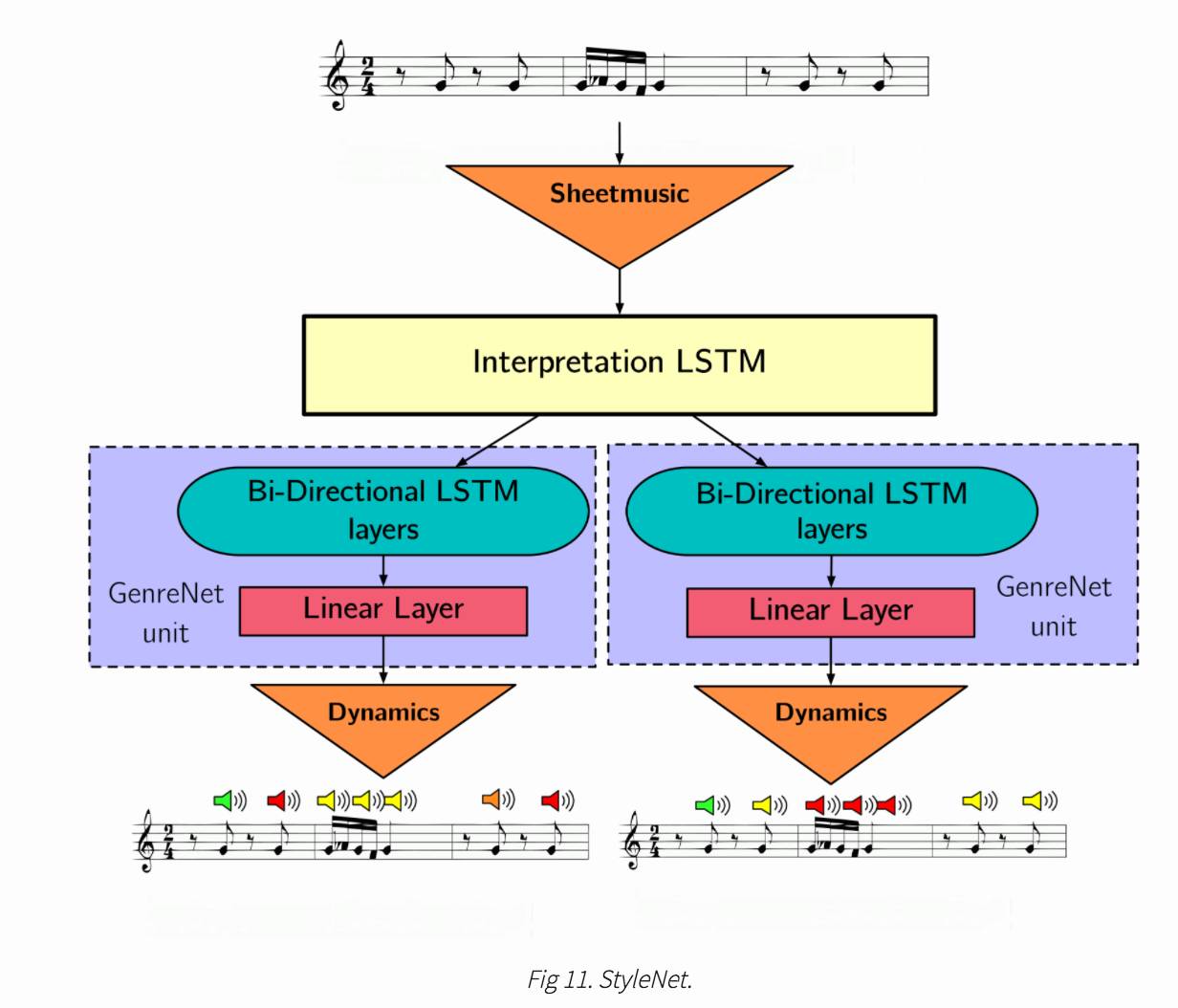
* GenreNet

模型中有两个主要的层：双向 LSTM（Bi-Directional LSTM）和线性层。双向 LSTM 层结合了 LSTM 网络的优点，它对学习相关的依赖提供了记忆，双向结构还使得模型在学习的时候同时考虑了未来的信息。这使得模型的输出可以作为另一层网络的输入。线性层就是被用来把双向 LSTM 的输出值的范围从 [-1,1] 变得更大。



* **StyleNet**

这个网络被用来学习一些在 Genre 网络上无法训练得到的更加复杂的风格信息。StyleNet 中有 GenreNet 的子网络，它们被用在 StyleNet 中来学习特定流派的风格。在 StyleNet 中，有一个解释层，它可以被 GenreNet 子网络共享。这大大减少了网络需要学习的参数。并且 StyleNet 就像是一个能够将音乐转换为不同风格的转换工具。如图 11 所示，这是一个多任务学习模型。



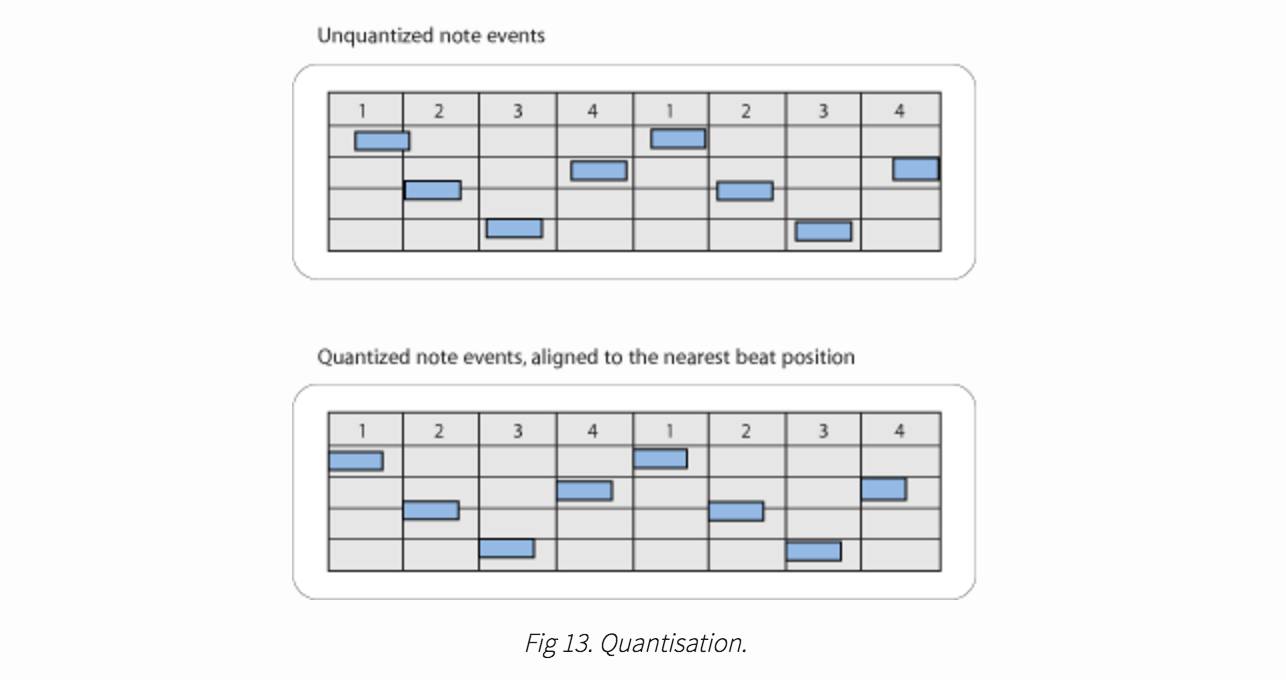
* 数据

使用了 MIDI 格式的音乐文件，因为这种格式的文件包含了音乐属性。有一个叫做速率（velocity）的参数来存储强弱力度。它类似于音量，但是取值范围在 0~127 之间。作者在本文中用速率来检测强弱力度。

MIDI编码方案

* 量化

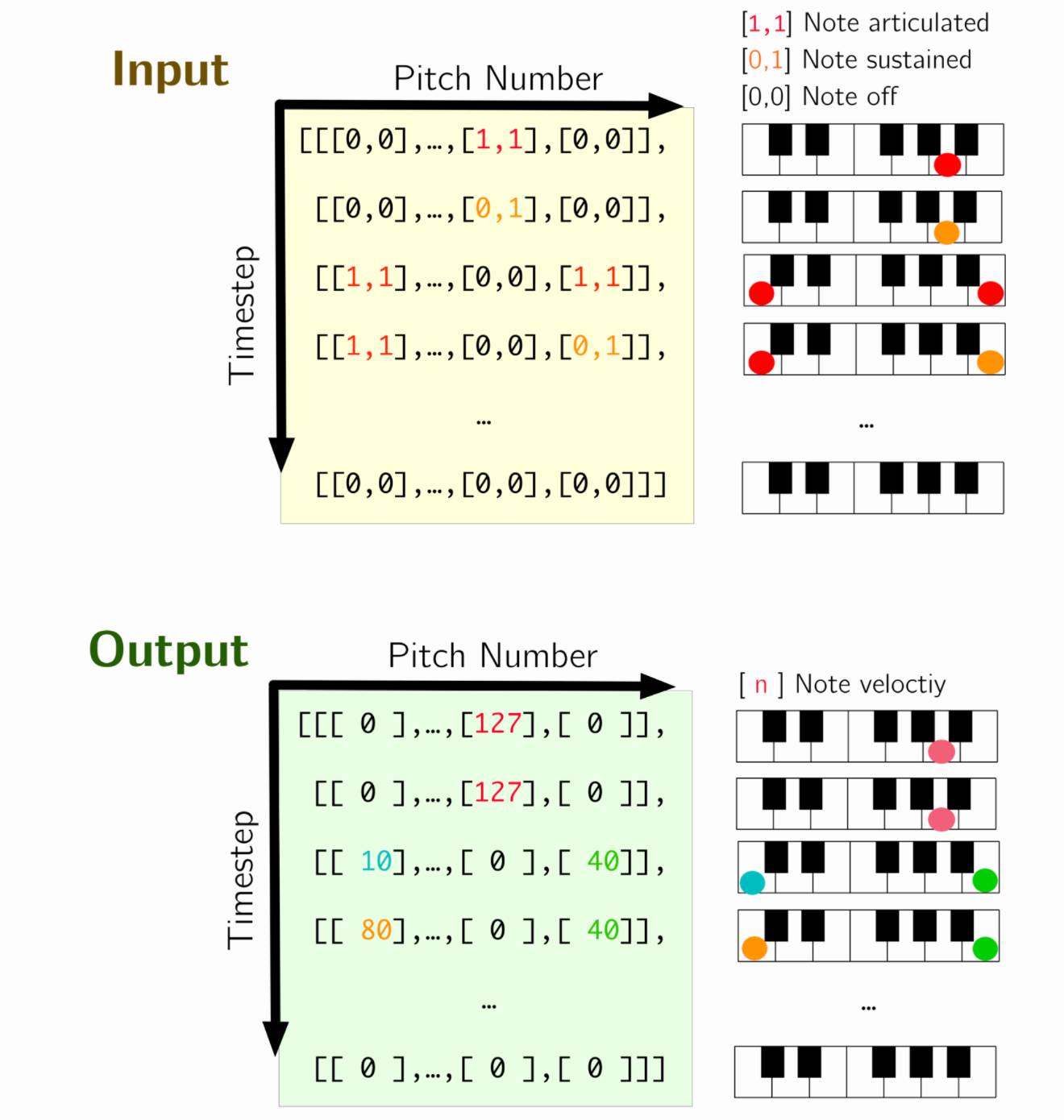
首先，数据集应该被量化，这使得作者能够拿矩阵来表征乐谱。但是但是这样做会导致一个结果，那就是我们会丢失乐谱相关的定时信息。文章中作者将乐谱的时间戳近似到了最接近的第 1/16 个音符。然后我们呢就可以捕获乐谱。图 13 展示了量化之前和量化之后的乐谱表征的不同。



* 输入矩阵表征

输入将会携带有关乐谱音高、开始时间以及结束时间等相关信息。乐谱总共有三个状态，所以作者使用二进制向量来表示这三个状态。「on」被编码为 [1,1]，「sustained」被编码为 [0,1]，「off」被编码为 [0,0]。

乐谱音高也需要被编码，我们会创建一个矩阵，其中第一维是关于 MIDI 音高数量的值，第二维是关于 1/16 音符量化为时间步骤的值。这些在图 14 中可以看到。



* 输出矩阵的表征

输出矩阵携带着输入的速率。如图 14 所示，矩阵的列也表示音高，行表示时间步骤。音高维度只有 88 个音符，因为我们这次只需要表示速率。然后我们以最大的速率 127 把数据进行分割，最后输出速率被解码为一个 MIDI 文件。

节选自 <http://www.sohu.com/a/159288996_465975>