测向仪是　　一种适用于音乐信息检索任务的强大声学特征

Tim Ziemer[1] *University of Hamburg, 20354 Hamburg, Germany,*

E-mail: tim.ziemer@uni-hamburg.de

arXiv:2302.01090v1 [cs.SD] 2 Feb 2023

引言

音乐信息检索（MIR）任务涵盖盲源分离、自动音乐转录以及音乐推荐。音乐学家常将声学特征和机器学习方法用作探索性工具，以分析民族音乐录音 [4, 6]，流行音乐 [2, 5] 和大型音乐合集 [2, 3]。用于音乐分析的声学特征中，使用最广泛的之一是梅尔频率倒谱系数（MFCCs），尽管它们既与音乐感知的任何方面无关，也与音乐制作和混音无关 [8]。要具有解释性和直观性，特征应当与音乐的制作方式或感知方式相关。测角仪是一个潜在的候选特征，因为音乐制作人和混音工程师会借助它监测并调整音乐混音的空间特性。本研究是一项探索性研究，旨在评估测角仪作为混合立体声音乐音乐信息检索特征的直观性和实用性。

测角仪

音乐制作人和音频工程师熟知多种录音与混音技术，可用于控制音乐制作的空间特性 [9]。许多空间特性都通过测角仪进行监测。图 1 展示了测角仪。它由两个仪表、一个形似李萨如图形或相空间图的点云，以及一个条形图组成。

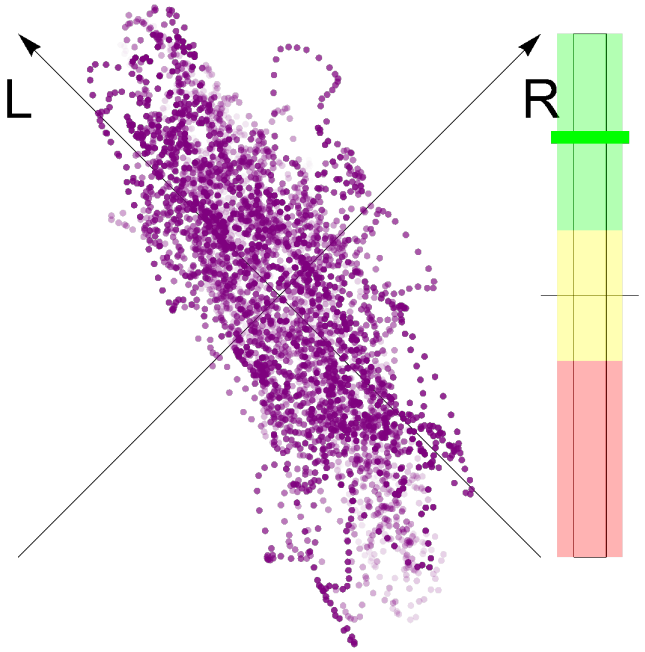

图 1： 典型的测角仪。将左声道（L）的时间信号绘制在右声道（R）的时间信号之上，形成点云（紫色）。两个信号的归一化相关性通过条形图（右侧）上的条形表示。通常，较高的相关性以绿色表示，而较高的负相关性则以红色表示。

在相空间图中，对于短时缓冲区中的所有采样点（该缓冲区可能包含 4096 个采样点），将左声道的时间序列绘制在右声道时间序列上。这个缓冲区也称为时间帧。每对采样点都表示为一个点，因此一个时间帧中的所有采样点会形成点云。这里，时间以不透明度表示：缓冲区中较早的采样点是透明的，而较晚的采样点是不透明的。由于一个简单的原因，图形往往会倾斜 $-45°$：当信号被路由到第一个立体声声道时，信号只会通过左侧扬声器播放，相应的图形看起来像一条沿 L轴分布的虚线。将单声道信号等量路由到左声道和右声道，会产生一条垂直虚线。由于存在倾斜，该图形可以粗略指示信号在立体声三角形中的声像定位。线段的长度表示声音的音量。上方和下方方格中的所有点都表示信号同相，也就是说，两个扬声器要么都正向振动，要么都负向振动。左侧和右侧方格中的点则表示信号反相，也就是说，一个扬声器正向振动，而另一个扬声器负向振动。当左声道和右声道的信号完全反相时，会产生一条穿过坐标轴原点的完美水平线。

相位。通常会避免这种情况，因为在典型的聆听环境中，也就是立体声三角形布局下，这会导致低音频率抵消。此外，对信号进行单声道混音会彻底抵消反相信号。至少在两种典型的聆听场景中，单声道混音仍然相当常见。第一种是在夜总会，扬声器彼此相距很远时。此时，许多混音工程师会选择让两个立体声声道至少部分混合。这样可以避免硬声像定位乐器只在舞池的一个角落里听得见。第二种是通过手机的内置扬声器播放音乐时。在过去，祖母那台老式便携式收音机完全是单声道的；而如今，年轻一代有时会通过手机的单声道高音扬声器，或通过立体声声道都装在一个小小箱体中的便携式蓝牙音箱听音乐。相位空间中与单一线条截然相反的情况，是整个区域都被点占满。这只有在每个声道都输入各自独立的白噪声时才会出现。只有将两路白噪声都归一化至最大声压级，整个空间才会被点填满。在其他所有情况下，场中的某些区域都不会被占据。音乐制作人和

混音工程师了解多种录音技术和数字音频效果，可以拓宽点的分布，让乐器和合奏听起来更加宽广，并使聆听体验更具沉浸感、包围感、空间感和鲜活感 [7]。与此同时，也有一些技术可以收窄分布，例如增强单个乐器的存在感和低音，并使其更加清晰、更容易定位。总之，点云的宽度表明整体混音的空间感有多强。点的大小代表音量。分布的整体方向表示声像定位。在所绘示的示例中，左侧扬声器发出的声音更响，因为一个声像定位偏左的响亮乐器主导了整体混音。音乐制作人和混音工程师对于点云应呈现何种形态都有自己的理念，例如它应有多宽、多大、波动多明显、倾斜程度如何，这取决于他们正在混音的音乐素材。

测角仪的第二个图是一张简单的条形图，用来显示两个缓冲立体声声道的归一化相关性。因此，声道相关性的取值范围为 -1 至 1。有些测角仪还会绘制短期历史记录，即过去几秒内声道相关性的最高值和最低值，或迄今测得的声道相关性的最小值和最大值。通常，颜色表示立体声混音是否完全兼容单声道（绿色）、基本兼容单声道（黄色）或不兼容单声道（红色）。同样，音乐制作人和混音工程师对于条形图应处于哪个区域、可以接受的最低位置，以及他们希望条形图位置呈现怎样的动态变化，都各有看法，具体取决于他们混音的音乐类型。

测角仪特征

在实际应用中，音乐制作人和混音工程师会借助测角仪观察混音的空间特性，并监控录音和混音过程中所做的修改会对其产生怎样的影响。测角仪是一种音乐测量与监控工具。这一过程更多是定性的，而非定量的；也就是说，从业者会评估测角仪两个图表的外观。要将测角仪的输出作为声学特征使用，就必须对其进行量化。

为了量化点云的外观，我们采用一种简单的盒计数法。如图 1 所示，将表示所有可能坐标对的正方形划分为 20×20 个小方格，共得到 400 个小方格。然后，只需统计每个时间帧中被一个或多个点占据的小方格数量。将该数量除以 400，即可得到一个百分比。对于每个分析的时间帧，都将此百分比存储在一个一维特征向量中。在 Python 中，可以针对每个时间帧 fs 使用以下代码高效地完成计算：

```
import numpy as np
bc = []#空盒计数特征向量 for i in range(fs):
    box[i]=[int(round(l[i]*10)),int(round(r[i]*10))]
bc.append(len(np.unique(box))/400)
```

我将这一特征称为相位范围 或盒计数。变量 l 和 r 是长度为 fs 的数组，包含左声道和右声道的时间序列。

为了量化条形图，将每个时间帧的幅值存储在一个一维特征向量中。在 Python 中，可以轻松地通过以下方式实现，例如：

```
import numpy as np corr=[]#相关性特征向量的空数组
corr.append(np.min(np.corrcoef(l,r)))
```

我将这一特征称为声道相关性。

用于提取该特征并将特征向量写入 CSV 文件的代码已公开发布在我的 GitHub 仓库 [1] 中。

在 [2] 中，我们已经证明，这两种测角仪特征与其他录音室特征（例如峰值因数和均方根能量）结合使用时，可以预测哪位唱片骑师会播放哪首歌曲。在本研究中，我以探索性方式单独评估测角仪，揭示测角仪如何作为一种特征用于线性和非线性音乐分类与聚类任务。

方法

为了评估测角仪特征的实用性，我采用了探索性研究方法。我分析了数百首歌曲。首先，我导入立体声 MP3 文件，并将其降采样至 22050 赫兹。然后，从每首乐曲的中间部分截取 500 帧。每帧长 2048 个采样点，彼此不重叠，因此每个文件分析约 46 秒。显然，分析的只有立体声文件，因为测角仪体现的是立体声的空间特性。为了了解测角仪特征是否直观且具有解释力，我比较了若干风格迥异的音乐类型。第一次比较选取了两张专辑中的歌曲：美国作曲家 David Miles Huber 的冥想专辑中的 16 首歌曲，以及 Jimmy-J & Cru-L-T 的 15 首歌曲；后者是英国二人组合，在20世纪90年代制作鼓与低音音乐，也称为碎拍硬核。第二次比较则是广义上的 90 首古典音乐作品与 93 首嘻哈作品。我通过目视检查两种测角仪特征的散点图，寻找线性规律。如果无法在线性层面区分这两种音乐类型，我便使用测角仪训练自组织映射，以了解测角仪在非线性聚类方法中的解释力。分析结果时，我考察了音乐制作人／混音工程师的创作意图或声音理想与测角仪数值之间的因果关系。

结果与讨论

第一次比较的分析结果绘于图 2。该散点图展示了特征空间。在这里，鼓打贝斯音乐与冥想音乐相距甚远，可以用线性方式将二者分开。任何聚类算法，例如 K 均值或自组织映射，都应该能够

将它们聚类；即使是单层感知机这类线性分类器，也应该能够以完美准确率将它们分类。请注意，在本例中，相位范围已足以区分二者，而声道相关性则不足以做到这一点。这幅清晰的图景揭示了音乐制作人的审美理想和创作意图。Jimmy-J & Cru-L-T 的鼓打贝斯音乐响亮且经过压缩，并具有较高的声道相关性，以保持低音频率同相，从而获得厚实、有冲击力的鼓声。另一方面，David Miles Huber 创作的音乐动态范围大得多，分布也更宽、相关性更低。当然，这对音乐专家来说显而易见。测角仪特征反映了这种创作意图，这一事实表明这些特征可能具有很高的解释价值，因为它们有助于发现音乐与制作人意图之间的因果关系。

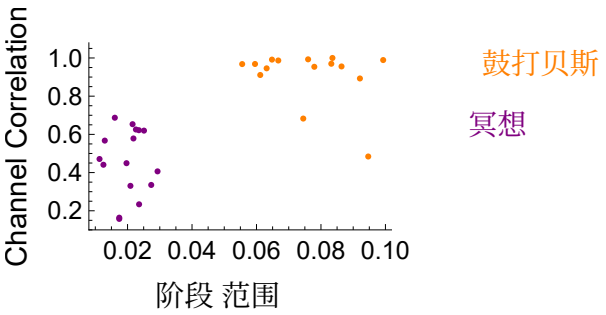

图2：鼓与低音（橙色）以及中介音乐（紫色）的平均相位范围与平均声道相关性。利用这两个测角仪特征可以将两者完美区分。

古典音乐和嘻哈音乐的散点图见图3。可以看到与之前类似的趋势，即嘻哈音乐位于右上角，古典音乐位于左下角。但两个维度上都有大得多的重叠。仅凭测角仪特征无法将这两种音乐线性分离。

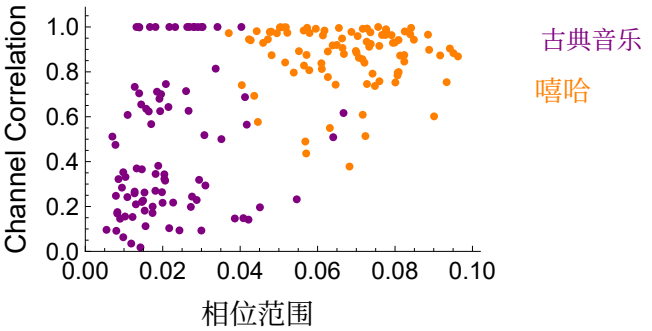

图3：只需绘制测角仪均值的分布，就能看出明显趋势：古典音乐和嘻哈音乐在流派上无法线性分离。

为评估测角仪特征的实用性，我们使用这些特征训练了一个自组织映射（SOM） [6]。SOM 是用于无监督学习（即聚类）的神经网络。我们将一个包含 $23 \times 23 = 529$ 个神经元的网格拟合到训练数据上，进行了 500 次迭代，学习率为 $\eta = 0.025$，初始邻域大小为 $\nu = 25$，并允许其线性递减至 1。图 4 以统一距离矩阵的形式绘出了结果。其中，深蓝色区域代表簇，黄色区域代表边界。每个标注的红点代表一首乐曲。可以看出，基于测角仪特征训练的 SOM 清晰地区分了嘻哈与古典音乐（红色边界），甚至成功将圣彼得堡男童合唱团的一张专辑聚为一簇（橙色椭圆标记）；在线性特征空间中无法观察到这一点，见图 3。SOM 显示，测角仪特征能够对所分析的两类音乐进行非线性分离，甚至识别出圣彼得堡男童合唱团形成的一个簇，这可能是因为整张专辑是在同一场录音中录制的，使用了固定的立体声麦克风配置，因此具有均一的相位范围和声道相关性。

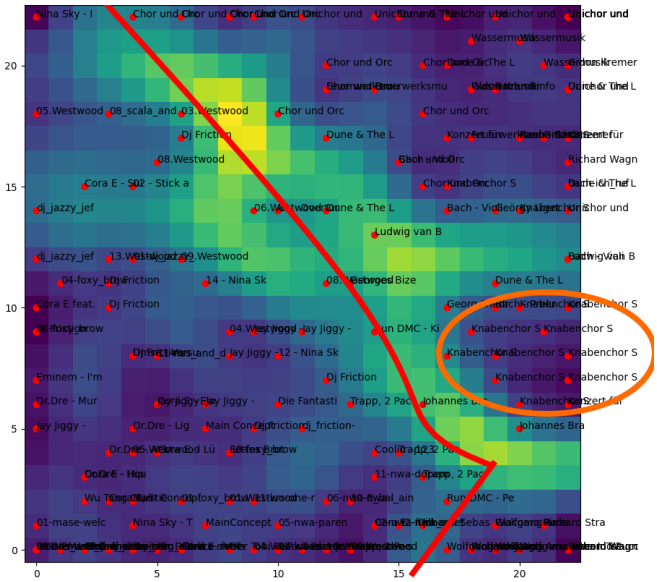

图4：以测角仪输出的均值进行训练后，自组织映射能够完美地区分嘻哈音乐与古典音乐（红线），甚至还能识别出一张专辑（橙色椭圆）。

两个分量矩阵如图5所示。它们展示了每个特征对自组织映射中各场的贡献程度。可以看出，相位范围的权重自上而下排列，而声道相关性的权重则自右向左排列。这种正交性表明二者彼此正交，因此能够完美互补。正交特征对机器学习非常重要，因为冗余会降低算法的准确性，而冗余减少方法在计算上可能代价高昂。

结论

在这项工作中，我提出将测角仪作为一种便于解释的声学特征，用于音乐信息检索任务。我定义了一种方法，用以量化音乐制作人和混音工程师对测角仪的定性认识。对不同类型音乐的分析表明，测角仪能够给出直观且合理的结果，可依据音乐录音与制作中的空间和动态方面，对不同音乐风格进行聚类。

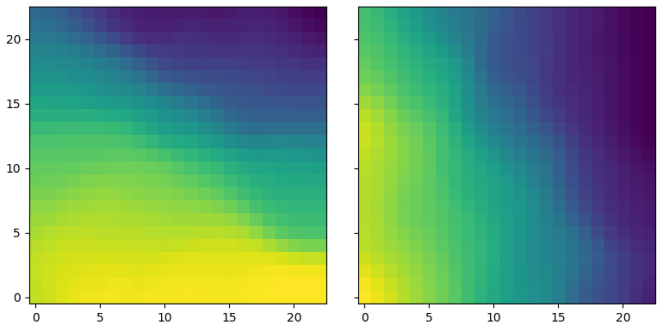

图5： 相位范围（左）和声道相关性（右）的分量矩阵表明，测角仪的这两个方面在线性上相互独立，即一个沿垂直方向排列，另一个沿水平方向排列。

过程。相位范围和声道相关性是相互正交、完美互补的特征。与梅尔频率倒谱系数相比，提取测角仪特征所需的计算资源极少，而且这些特征与音乐制作人的意图存在因果联系。进一步的分析将揭示，测角仪特征是否也适用于探索其他立体声音频文件，例如播客、广播剧和电影声音；以及乐器声辐射特性的录音、建筑声学的双耳录音和声景录音。

参考文献

[1]Ziemer, T.，goniometerfeature，GitHub 代码库，网址：https://github.com/TimZiemer/ goniometerfeature

[2]Ziemer, T.、Kiattipadungkul, P. 和 Karuchit, T.：《用于音乐信息检索任务的录音室声学特征》。声学会议论文集 42（2020），DOI：10.1121/2.0001363

[3] Ziemer, T.、Yu, Y. 和 Tang, S.：《在音乐声音分析中使用心理声学模型》。FIRE '16：第八届信息检索评估论坛年度会议论文集（2016），DOI：10.1145/3015157.3015158

[4]Bader, R., Blaß, M. & Franke, J.：对缅甸北部克钦族音乐与新疆维吾尔族音乐进行计算音色与调性系统相似性分析。DOI：10.48550/arXiv.2103.08203

[5] Bader, R., Zielke, A. & Franke, J：基于音色的机器学习：对中国嘻哈音乐与西方嘻哈音乐进行聚类。AES Convention 150（2021），URL：http: // www.aes.org/e-lib/browse.cfm?elib=21066

[6]Blaß, M. & Bader, R.：《民族音乐档案的基于内容的音乐检索与可视化系统》。R. Bader（编）：《计算录音档案》，施普林格，Cham（2019），145-173, DOI：10.1007/978-3-030-02695-0 7

[7]Ziemer, T.：《音乐制作中的声源宽度：立体声、全景声与波场合成方法》。A. Schneider（编）：《音乐声学与心理声学研究》，施普林格，Cham（2017），399-440，DOI：10.1007/978-3-319-47292-8 10

[8]Alluri V. & Toiviainen P.：探索复调音色的感知与声学相关特征。《音乐感知跨学科期刊》27(3)（2009），223–242，DOI：10.1525/mp.2010.27.3.223

[9] Stirnat, C. & Ziemer, T.：《音乐中的空间感：音响工程师的意图与听众的感知》，KLG 2017。klingt gut! 2017——国际声音研讨会（2017），DOI：10.29007/nv93