摘 要

本课题研究基于古琴空弦和敲击音的音频数据，提出了一种基于频域特征的音频处理流程，结合信号处理手段和机器学习方法，我们能尝试量化古琴空弦和敲击音的预估关系。本研究的方案充分考虑了**空弦音和敲击音对于琴声共鸣体特征的揭示作用**。研究主要分为三个部分：第一部分是对于空弦音的提取，我们应用了一种比较先进的“多倍频法”作为滤波器提取出空弦音的主频音；第二个部分我们用空弦主频音的截至频率和主频音数量作为指导，通过巴特沃斯滤波器处理了敲击音的音频数据，并基于 K-means 聚类算法提取出敲击音的主频音；最后，作为本文的主要贡献之一，我们提出了一种基于余弦相似性的音频相似性量化方法，充分比较了空弦音和敲击音的预估关系。并针对提供的数据辨识出其中比较好的能反映空弦音的敲击音。

This study is based on the audio data of the empty string and percussion of the guqin. A frequency-domain feature-based audio processing flow is proposed. Combined with signal processing methods and machine learning methods, we can try to quantify the prediction relationship between the empty string and percussion of the guqin. The scheme of this study fully considers the revealing effect of the empty string and percussion on the resonance characteristics of the sound body. The research is mainly divided into three parts: the first part is the extraction of the empty string sound. We applied a more advanced ”multiple frequency method” as a filter to extract the main frequency sound of the empty string; In the second part, we used the cut-off frequency of the main frequency sound of the empty string and the number of main frequency sounds as a guide, and processed the audio data of the percussion with a Butterworth filter, and extracted the main frequency sound of the percussion based on the K-means clustering algorithm; Finally, as one of the main contributions of this article, we propose an audio similarity quantification method based on cosine similarity, which fully compares the prediction relationship between the empty string and the percussion. And identify the better percussion that can reflect the empty string sound among the provided data.

目 录

1 引言和项目背景

1.1 背景概述

中国民族乐器中的弦乐器具有悠久历史，这类乐器通过弹拨琴弦产生振动作为音源，琴弦振动通过乐器配件的传导（例如琴码）传递到琴体上，通过琴体面、底板以及内腔空气的共鸣作用，对相应频段的激励信号进行响应，从而扩大音量、美化音色，满足乐器的功能性和审美性需要。

因此，弦乐器的演奏音质可以看作决定于共鸣体（琴板）对于不同琴弦分段振动产生的泛音列（即谐波，范围可以从几十 Hz 到数千 Hz）激励信号的响应特性的主观感受和宏观评价。在本文的研究中，我们以古琴为研究对象，尤其通过采集到的空弦弹奏音和敲击音信号的频域音频特征，来研究空弦音和敲击音对于琴声共鸣体特征的揭示作用。

1.2 古琴历史，制作和结构

古琴是中国传统拨弦乐器，有 3000 年以上历史。它与中国文明初期的帝王有关，如伏羲、神农、舜。在《诗经》等古籍中有记载。古琴在不同朝代，尤其是宋朝，非常流行。

古琴的制作过程分为多个步骤，包括选择材料、造型、槽腹、合琴、灰胎、研磨、擦光、定徽、安足、上弦等。每个步骤都非常重要，影响着古琴的质量和音色。古琴的正面，从上到下分别是琴额、承露、岳山、1～7 弦、项、1～13 徽、肩、腰、冠角、龙龈；古琴的底面，从上到下分别是护轸、轸池、弦眼、龙池、雁足、凤沼、龈托；古琴的内部，从上到下分别是舌穴、项实、声池、槽腹、天柱（圆）、纳音、地柱（方）、足池、韵沼；古琴的侧面，从上到下分别是护轸、琴轸、低头、弦、雁足。下图展示了古琴的结构示意图。

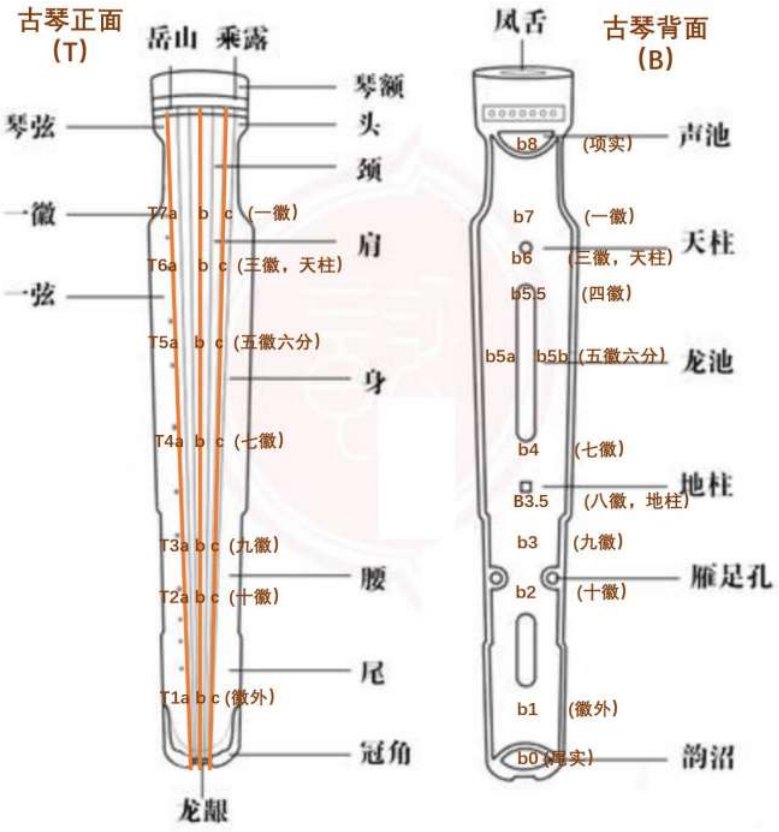


图 1: 古琴结构示意图

1.3 古琴的发声原理和项目的可行性分析

理解古琴的发声原理，尤其是其中反映的空弦音和敲击音对琴身共鸣体的特征揭示作用对于我们理解项目的合理性和可行性至关重要。古琴的发声过程包括激发、弦振、传导和辐射四个阶段。琴弦的振动通过岳山、龙龈等部件传导至琴体，产生共鸣。琴体的结构和材料对音色有重要影响。

激发（Excitation）：这是发音的起始阶段，演奏者通过手指作为牵引力或推动力的来源，使琴弦脱离静止位置。这一阶段可以细分为散音（无左手接触，即弦乐器的空弦音）、泛音（左手轻触琴弦但不按实，突出某个分段振动的发音）和按音（将琴弦按压于面板，使琴弦的某一段完全振动发出声音）。

弦振（String Vibration）：琴弦是古琴的原始振动体，其声学特征主要来源于弦。琴弦被激发后，由于张力作用，会在静止位置两侧进行往复运动。这种振动随着能量的逐渐耗损而逐渐衰减，最终回到静止状态。琴弦的机械振动能量转换为声音能量，主要通过部件如岳山和龙龈将振动传导至乐器共鸣体完成。同时，琴弦的运动还会挤压周围空气，产生声波。

传导（Conduction）：这一阶段的目的是使乐器共鸣系统和槽腹内的空气产生充分共振。琴弦振动作用于岳山与龙龈，这些部件再将振动传导至琴的面板。面板通过槽腹内的天地柱（部分琴具有）将振动传导至底板，从而使整个古琴参与振动。为了快速传导振动能量，岳山和龙龈等部件通常使用质地坚硬、传导性能良好的材料制成。

辐射（Radiation）：古琴声音的辐射能力与振动部件的振幅大小及振动体参与振动的面积有关。为了增大音量，共鸣系统需要通过增大单位面积和时间内输出的声音能量。古琴的面板和背板是完成声音辐射的主要部件，同时槽腹内的空气共振也会通过背板两个音孔向外辐射声波。辐射过程中，由于振动体与共鸣体之间的耦合作用，会使原始振动中的某些频率成分得到增强或衰减，从而进一步改善音高、音色和辐射指向性等声学特性

由此可见，古琴的**振动特性依赖于琴弦弹奏传导到琴身引发的共振效应**。因此，从信号处理和系统辨识的角度分析，如果把**古琴的空弦输入的响应音频视作琴身系统的正弦响应，底板和面板的敲击输入响应视作琴身的冲激响应**，我们就能联系两种音频并用它们建模系统。另外，由于两种输入都反应了琴身的共振特性，我们也可以**寻找量化两种输入相似性的方案**提出预估空弦音的敲击音的方法，这正是研究的动机所在。

2 相关工作

2.1 古琴的共振特性研究

在本研究之前，上海交通大学的卢艺老师在他的著作《琴书密码》，一书中比较完整地阐述了古琴的历史，制作，结构和文化意象等。同时也创新性地把信号处理的手段应用到古琴的研究中。作为他工作的延续之一，后来的一项工作 (赵逸凡，佟庆月等，2023) 创新性地利用” 最大倍频和法”和“多倍频法”提取出了古琴空弦音的主频音，本文的工作在处理空弦音的部分受此启发。本文的工作可以视作两个方法的一个简单的拓展，**通过引入机器学习方法和考虑敲击音和空弦主频音的相关关系**，我们尝试提出一种预估两种音的量化方案。

2.2 乐器合成建模方法

乐器合成是指通过数学模型或者物理模型来模拟乐器的声音，从而实现乐器的电子合成。在 ICML 等国际音乐领域会议的论文中，有很多关于乐器合成的研究，其中有一些研究也使用了信号

处理的手段。例如，(Bittner et al., 2023) 提出了一种基于神经网络的乐器合成方法，通过训练神经网络来学习乐器的声音特征，从而实现乐器的合成。

显而易见的是，基于古琴的共振特性研究的工作和乐器合成的工作有很多相似之处，从乐器合成的建模方法中我们也能得到一些启发，本文的机器学习方法也是受此启发。

3 研究方法

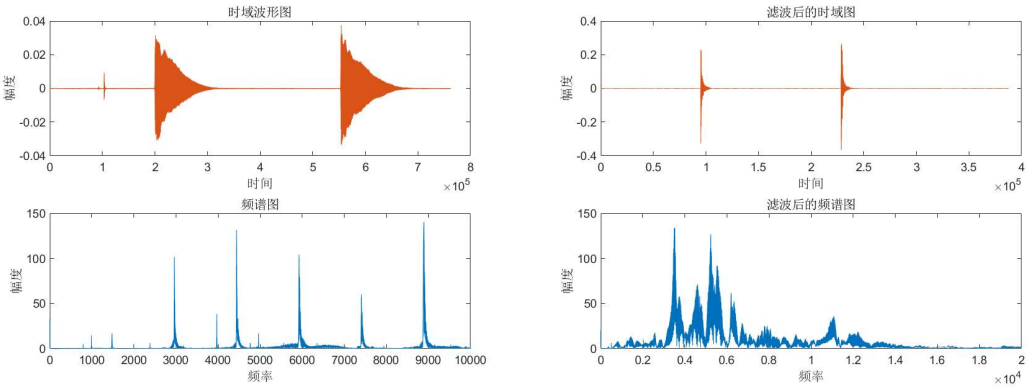
3.1 研究思路概述

本文的研究思路如下图所示，我们首先通过信号处理的手段提取出空弦音和敲击音的频域特征，然后通过机器学习的方法提取出空弦音和敲击音的特征向量，最后通过比较两种音频的特征向量来量化两种音频的相似性。

在频域特征提取的部分，我们首先通过“多倍频法”提取出空弦音的主频音，然后通过巴特沃斯滤波器提取出敲击音的主频音。在特征向量提取的部分，我们首先通过 K-means 聚类算法提取出敲击音的主频音。最后在相似性量化阶段，我们通过余弦相似性量化两种音频的相似性。

3.2 音频录入和频域特征的初步提取

下左图展示了录入的空弦音频的时域和频域波形，下右图展示了录入的敲击音频的时域和频域波形。（详细的涉及全部数据的图片放在了附属的文件夹中）。



(a) 空弦音时域频域波形　　(b) 敲击音时域频域波形

图 2: 空弦音和敲击音时域和频域波形比较

我们可以看到，空弦音频的主频音比较明显，**呈现出明显的正弦输入响应的特征，而敲击音频的主频音比较模糊，呈现出明显的冲激输入响应的特征。而处理两种音频的方法也应该根据这两种特征来设计。**

另外，在研究的初期，我们还针对录入的波形计算一定的频域特征量辅助我们研究，比如：

平均幅度谱（Mean Amplitude）：平均幅度是指信号的平均幅度，是信号的幅度的一种度量。平均幅度越大，信号的幅度越大。平均幅度的计算公式如下：

$$A_{mean} = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{1}$$

均方根能量（Root Mean Square Energy）：均方根能量是指信号的均方根值，是信号的能量的

一种度量。均方根能量越大，信号的能量越大，信号的幅度越大。均方根能量的计算公式如下：

$$E_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{2}$$

其中，x_i 是信号的第 i 个采样值，N 是信号的采样点数。

频谱质心（Spectral Centroid）：频谱质心是指信号频谱的质心，是信号频谱的中心频率，是信号频谱的一阶矩。频谱质心越大，信号的频谱越向高频偏移，频谱质心越小，信号的频谱越向低频偏移。频谱质心的计算公式如下：

$$C_s = \frac{\sum_{i=1}^{N} f_i X_i}{\sum_{i=1}^{N} X_i} \tag{3}$$

其中，f_i 是信号的第 i 个频率，X_i 是信号的第 i 个频率的幅度，N 是信号的频率点数。

下面的表格展示了上图展示的两种音频的频域特征量的计算结果。(我们在这里仅展示 r1a 空弦音和 b1 底板敲击音的计算结果)

表 1: 空弦音和敲击音频域特征量比较

特征量	空弦音	敲击音
平均幅度谱	0.000047	0.000033
均方根能量	0.004860	0.004055
频谱质心	5110.303	9187.357

从两种音频的频域特征量比较中，我们可以看到，空弦音的频谱质心比较小，而敲击音的频谱质心比较大，同时空弦音的平均幅度比敲击音大，但是均方根能量相似，说明空弦音更稳定，而敲击音更具有冲击的突变性。

3.3 空弦音频的主频音提取

在本研究中，我们采用了“多倍频法”作为滤波器提取空弦音的主频音。这种方法的原理是，通过对空弦音频的频谱进行分析，我们可以发现空弦音频的频谱中有明显的主频音，而且主频音的倍频音也比较明显，因此我们可以通过提取主频音的倍频音来提取主频音。参考研究中的“多倍频法”提取空弦音的伪代码如下: 算法1。下图展示了一个我们采用” 多倍频法” 提取空弦音的例子，我们可以看到，空弦音的主频音和倍频音都被提取出来了。**而那些无关紧要的次要频率则被滤除了**。我们把这个方法也作为我们的空弦音频的滤波器设计。

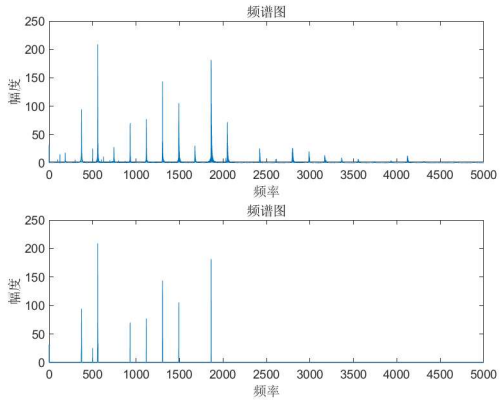


图 3: 空弦音频的主频音提取

Algorithm 1 多倍频法提取主频音

Require: Signal $signal$, Frequency list $freqs$
Ensure: Filtered signal new_signal

$new\_signal \leftarrow signal$
$n \leftarrow 50$ ▷ 设定窗口 50Hz 的数据点数
$sum \leftarrow signal[1:n]$ ▷ 用于计算的信号均值
$mul \leftarrow 3$ ▷ 倍数阈值确定异常数据点的倍数
$tmp \leftarrow 0$ ▷ 临时变量，用于记住峰值索引
for $i = 0$ TO $\text{len}(signal)$ **do**
  $sum \leftarrow sum + signal[i]$
  if $signal[i] < \max(signal)/15$ **then**
    $new\_signal[i] \leftarrow 0$
    continue
  end if
  if $i < n$ **then**
    $mean \leftarrow \frac{sum+signal[n+i]}{n+i+1}$
  else if $i < \text{len}(signal) - n$ **then**
    $mean \leftarrow \frac{sum+signal[n+i]-signal[i-n-1]}{2*n+i}$
  else
    $mean \leftarrow \frac{sum-signal[i-n-1]}{\text{len}(signal)-i+n}$
  end if
  if $signal[i] > mean * mul$ **then**
    if $freqs[i] - freqs[tmp] < 20$ **then** ▷ 对间隔小于 20 Hz 的峰进行筛选
      if $signal[tmp] < signal[i]$ **then**
        $new\_signal[tmp] \leftarrow 0$
        $tmp \leftarrow i$
      end if
    else
      $new\_signal[i] \leftarrow 0$
    end if
  end if
end for
return new_signal ▷ 输出结果

我们在实际的代码中适当简化了这一过程，同时也调整了需要被滤掉的信号阈值下限，以适应我们的数据。

同时经过“多倍频法”的提取我们的信号简化为相互靠近的几个峰值，但相邻的数据仍然有较大的混淆，为进一步简化数据，我们用一个 20 步长的过滤窗进一步筛选频率，最后得到了一个比较干净的主频音一信号。

为适当简化研究，我们把三根弦的数据分别进行了合并，得到了三个主频音信号，下图展示了三个主频音信号的频谱图。

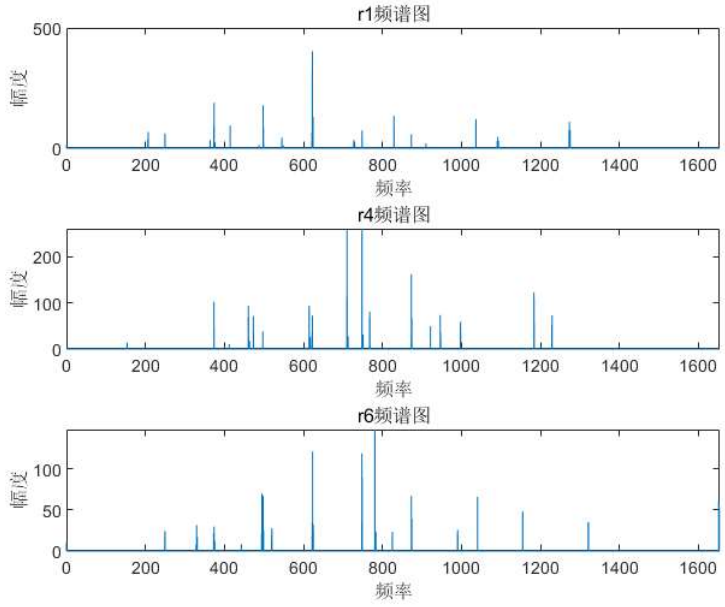
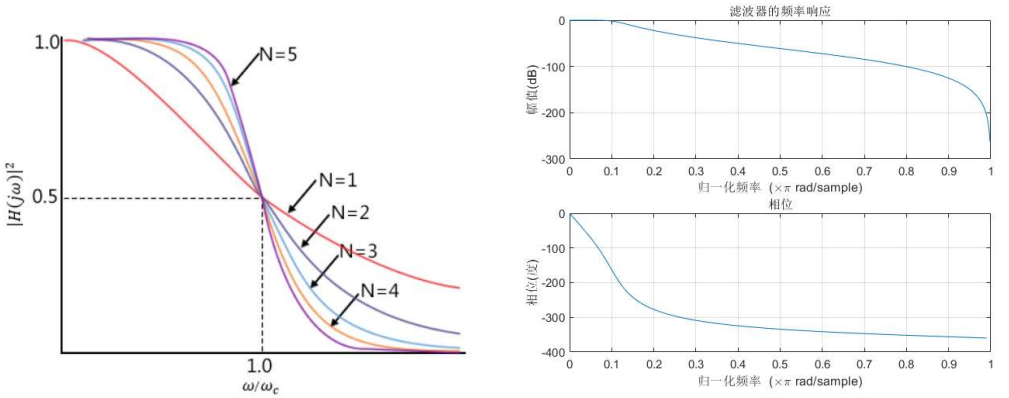


图 4: 三根弦“多倍频法”处理的主频音信号的频谱图

3.4 敲击音频的主频音提取

3.4.1 空弦音截至频率启发性的巴特沃斯低通滤波器设计

考虑到未经处理的敲击音频谱图存在包含多个成分的复杂性，我们仅用根据“多倍频法”提取的空弦主频音的**最高截至频率**作为敲击音频的巴特沃斯低通滤波器的通带截至频率，以在保留**潜在的和空弦主频音相关的**敲击音频的主频音的同时，**滤除敲击音频的次要频率**。下左图展示了一个理论的归一化巴特沃斯低通滤波器的频率响应曲线，下右图展示了一个我们设计的巴特沃斯低通滤波器的频率响应曲线。



(a) 理论归一化巴特沃斯低通滤波器的频率响应 (b) 我们设计巴特沃斯低通滤波器的频率响应

图 5: 巴特沃斯低通滤波器的频率响应曲线比较

能观察到，在 dB 域中，我们的低通滤波器在所需的截至频率后衰减至零轴以下，符合低通滤波器的设计要求。

3.4.2 基于 K-means 聚类算法的敲击音频的主频音提取

回顾上面我们提取出的过滤后的敲击频域波形图，我们发现敲击音的频域在极大的频域范围内都有分布，并且分布复杂不具有较强的规律性。

为了进一步提取出主要的频率成分，我们采用了 **K-means 聚类算法**，通过聚类算法的手段，我们可以把敲击音频的频域分布聚类成若干个类，从而提取出敲击音频的主频音。比较合理的一种考虑是，我们**要以三根弦的空弦主频音的条数作为我们期望的聚类组数**，这样我们提取的敲击音频的主频音就可以和空弦音频的主频音在**维度上一一对应起来**，从而方便我们后续对它们的预估关系进行量化处理。

作为一种基于距离的聚类算法，K-means 聚类算法的原理是，通过迭代的方式，把数据集中的数据点分成若干个类，使得类内的数据点的距离尽可能的小，而类间的数据点的距离尽可能的大。K-means 聚类算法的伪代码如下: 算法2。

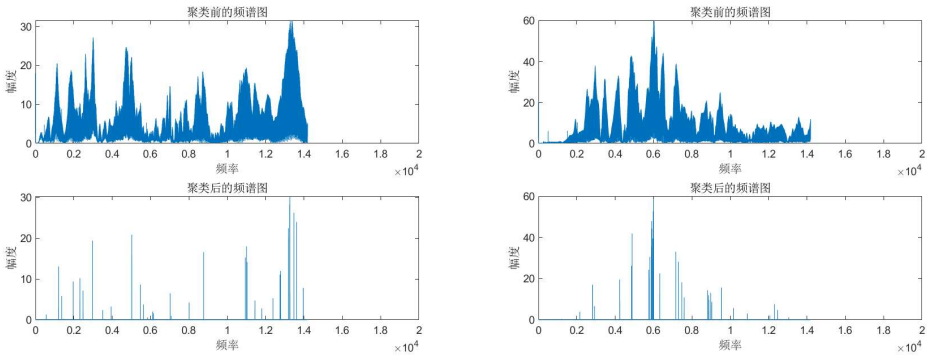
Algorithm 2 K-means 聚类算法

Require: 数据集 X, 聚类数 k
Ensure: 聚类中心 C, 聚类标签 L
$C \leftarrow$ 从 X 中随机选择 k 个样本 ▷ 初始化聚类中心
$L \leftarrow$ 全零向量 ▷ 初始化聚类标签
while 未收敛 **do**
  for $i = 0$ TO $\text{len}(X)$ **do**
    $L[i] \leftarrow \text{argmin}_j \text{dist}(X[i], C[j])$ ▷ 分配聚类标签
  end for
  for $j = 0$ TO k **do**
    $C[j] \leftarrow \frac{1}{\text{len}(X[L==j])} \sum_{i=0}^{\text{len}(X)} X[i] \cdot [L[i] == j]$ ▷ 更新聚类中心
  end for
end while
return C, L ▷ 输出结果

在经过一定的迭代后，K-means 可以求出样本的 K 个聚类中心，从而把样本分成 K 个类，同时也可以求出每个样本的类标签，从而把样本分成 K 个类。在此之后：我们进行以下两个操作：

1. **求出和样本中心最近的样本值保留;**
2. **把保留的样本值之外的值设置为零。**

如此我们便从敲击音频的频域波形图中提取出了和空弦主频音条数相同的敲击音频的主频音，下图展示了我们提取出的敲击音频的主频音的频谱图。



(a) 底板敲击音频的频谱图 (b) 面板敲击音频的频谱图

图 6: 敲击音频的主频音频谱图

可以看到不管是底板敲击音频还是面板敲击音频，我们都提取出了和空弦主频音条数相同的敲击音频的主频音，而且主频音的条数和空弦主频音的条数一一对应起来了。并且这种基于**机器学习的聚类离散化技巧也很好地保留了原本频谱的分布结构**。更多的结果我们都保存在了附录文件中。

3.5 基于余弦相似性的两种音频的预估关系量化

本工作的核心目的旨在**量化空弦音和敲击音的预估关系**。而经过之前的工作我们已经得到了空弦音和敲击音的**稀疏分布的同维度离散主频音**。充分考虑到频域分布的稀疏性和随机性，我们在这里不再采用一般信号处理的相似性计算方法。而是采用了一种“向量化”的余弦相似性量化两者的预估关系：

首先我们定义**“主频分布特征向量”**如下：

$$\vec{v} = \begin{bmatrix} f_1 & f_2 & \cdots & f_n & H_1 & H_2 & \cdots & H_n \end{bmatrix}^T \tag{4}$$

其中，f_i 是第 i 个主频音的频率，H_i 是第 i 个主频音的幅度，n 是主频音的条数序号。

一旦定义了“主频分布特征向量”，我们就可以定义“主频分布特征向量”的余弦相似性：

$$\cos(\vec{v_1}, \vec{v_2}) = \frac{\vec{v_1} \cdot \vec{v_2}}{|\vec{v_1}| \cdot |\vec{v_2}|} \tag{5}$$

其中，$\vec{v_1}$ 和 $\vec{v_2}$ 是两个“主频分布特征向量”。这里一般其中一个向量是空弦音的“主频分布特征向量”，另一个向量是敲击音的“主频分布特征向量”。

由此我们便可以量化空弦音和敲击音特征向量的相似性，从而量化空弦音和敲击音的预估关系。

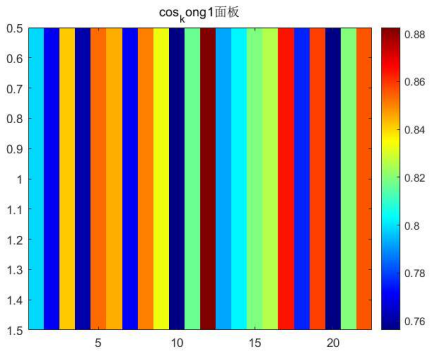
但是我们必须要指出，这种量化方式并不是唯一的。同时由于在我们强制拆开了频域分布点和幅度分布点的耦合关系，仅从排序的“相同位置”相同大小上度量它们的相似性，因此我们的量化方式也存在一定的缺陷。

4 研究结果

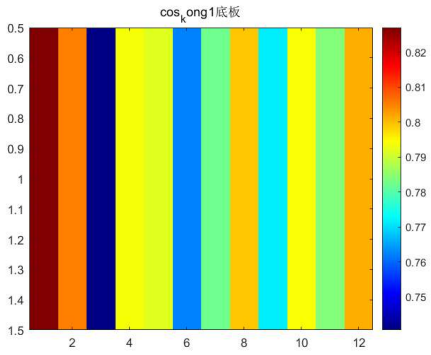
4.1 空弦音和敲击音的余弦相似性量化结果

下图我们用热力图的形式分别展示了基于余弦相似性计算出的 1,4,7 三个空弦音和敲击音的余弦相似性量化结果。左边是面板敲击音的余弦相似性量化结果，右边是底板敲击音的余弦相似性量化结果。

进而，通过读取其中符合最高颜色的数值（即最高相似性）的位置，在我们的热力图里为深红色。并把处理中的数据序号对应到具体的敲击位置，我们就能初步判断出空弦音和敲击音的预估关系。

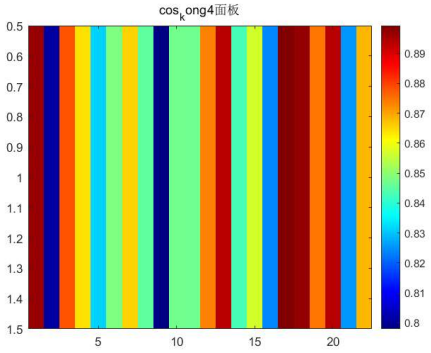


(a) 面板敲击音的余弦相似性量化结果

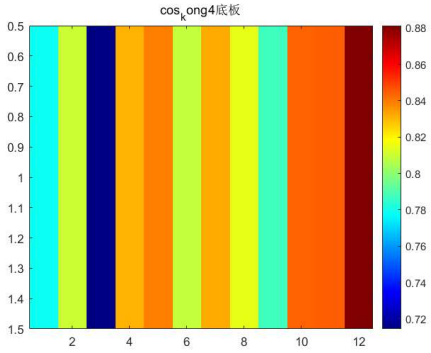


(b) 底板敲击音的余弦相似性量化结果

图 7: 空弦音 1 和敲击音的余弦相似性量化结果

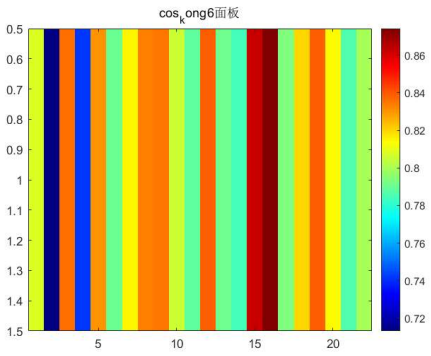


(a) 面板敲击音的余弦相似性量化结果

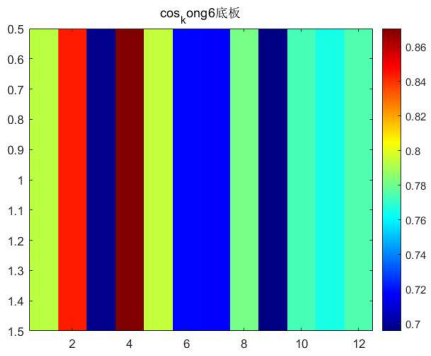


(b) 底板敲击音的余弦相似性量化结果

图 8: 空弦音 4 和敲击音的余弦相似性量化结果



(a) 面板敲击音的余弦相似性量化结果



(b) 底板敲击音的余弦相似性量化结果

图 9: 空弦音 7 和敲击音的余弦相似性量化结果

4.2 尝试找到最好的敲击位置

以下的表格统计了六个空弦音和敲击音的余弦相似性量化结果中，最高相似性的位置，相似性数值和对应的实际敲击位置。

结果 | 空弦音对应的弦	空弦 1	空弦 1	空弦 4	空弦 4	空弦 7	空弦 7
最高相似性数值	0 .8826	0 .8270	0 .8992	0 .8813	0 .8742	0 .8705
最高相似性的位置	12	1	17	12	16	4
对应的实际敲击位置	面板 t4b 轻	底板 b0	面板 t6a	底板 b7	面板 t5c	底板 b3.5

表 2: 六个空弦音和敲击音的余弦相似性量化结果

4.3 研究结论和合理性分析

通过以上的最值对比我们能得到以下重要结论：

- **面板敲击音普遍比底板敲击音与空弦音的相似性高；**
- **面板的敲击音相似性较大的位置集中在面板中央位置；**
- **底板敲击音相似性较大的位置则分布在底板的各个位置，并且有向边缘靠拢的趋势；**
- **相较于空弦 1,7 等分布在边缘位置的琴弦，中心位置的空弦 4 与敲击音的相似度较高。**

这些结论都与我们的实际经验一定程度上相符，考虑到研究开头的介绍，古琴振动的本质是琴弦振动引起的空气振动传导至琴身引起的共鸣现象。

相较于琴背，琴面能更快更近距离地接受琴弦引起的振动，同时中心位置也能最平均地接收到各个方向的振动，因此面板敲击音的相似性普遍比底板敲击音的相似性高。

而仅对琴背而言，由于琴弦的连接位置最靠近边缘，同时由于空气的接触首先产生在边缘，因此边缘部分相较于琴背位置能产生更大的振动相关关系。

总的来说，基于我们的方法，如果我们希望通过敲击音频特征比较恰当地反映一定的空弦音和敲击音的预估关系，我们应该**在面板中心位置敲击**，同时辅助地**在底板边缘位置敲击**。

5 研究总结和感悟

5.1 研究总结

本研究通过对古琴的空弦音和敲击音的频域特征进行分析，提出了一种基于余弦相似性的量化方法，用于量化空弦音和敲击音的预估关系。我们的方法主要分为对空弦音和敲击音的频域特征提取和对空弦音和敲击音的预估关系量化两个部分。在主频提取阶段，我们采用“多倍频法”提取空弦音的主频音，之后在空弦主频音的指导下采用 K-means 聚类算法提取敲击音的主频音。最后在预估关系量化阶段，我们采用余弦相似性量化空弦音和敲击音的预估关系，并得到了具有一定合理性的结果，**如果我们希望通过敲击音频特征比较恰当地反映一定的空弦音和敲击音的预估关系，则更需要面板中心部分的敲击音频和底板边缘部分的敲击音频。**

当然我们的方法也还存在很多问题，比如主频的提取从满秩频率数据降维到稀疏数据的过程我们也许忽略了较多的重要信息，并且余弦相似性的相似度估量可能也缺乏了对频域分布点和其幅度的耦合关系。

5.2 研究感悟

《礼记 • 乐记》中说：**“琴者，乐之枢纽也。”**古琴作为中国传统文化的重要组成部分，其研究价值不言而喻。但是乐器系统研究的一个难点就在于“量化”过程的复杂性和难解释性。