语音特征提取实验报告

王松宸 2024201594

2025 年 12 月 19 日

1 实验目的

本实验旨在通过使用 Librosa 库，完成音频信号从时域到频域、时频域以及感知域的全面分析。实验将计算并分析幅度包络以理解信号能量随时间的变化，同时，利用快速傅里叶变换（FFT）进行频域分析，绘制幅度谱以观察信号的频率成分。进一步地，通过短时傅里叶变换（STFT）生成语谱图，并结合听觉感知基础绘制 Mel 语谱图。最后，提取梅尔频率倒谱系数（MFCC），掌握这一在语音识别与处理中至关重要的特征提取方法。

2 Task 1：振幅包络

振幅包络反映了音频信号能量随时间的轮廓变化，对于检测语音的起始点以及分析信号的动态特性具有重要意义。

在实现上，我将音频信号按照设定的窗口大小（`frame_size`）进行分帧，并以一定的步长（`hop_length`）在时间轴上滑动。对于每一个帧，取其信号绝对值的最大值作为该帧的包络值。这种方法能够有效地平滑信号细节，保留整体的能量走势。

下图展示了音频的原始时域波形以及提取出的振幅包络，可以看到包络线很好地描绘了波形的外部轮廓。

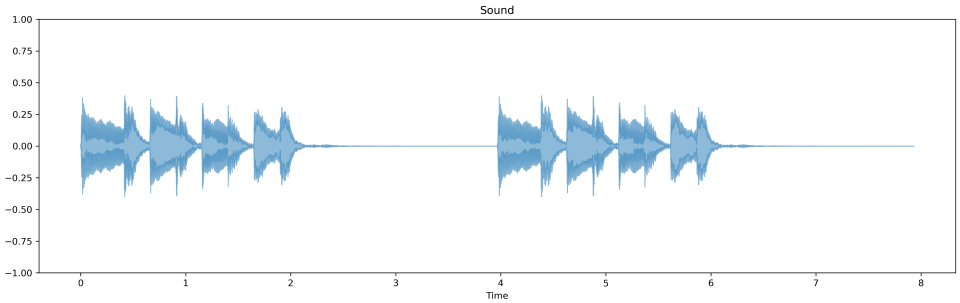

图 1: 音频时域波形

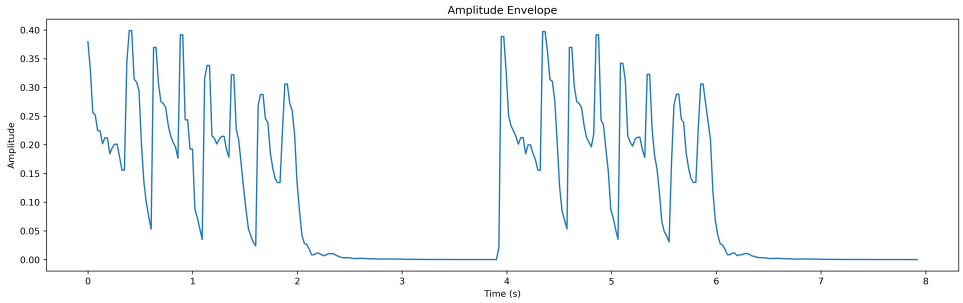

图 2: 振幅包络（frame size=1024, hop=512）

3 Task 2：幅度谱

幅度谱用于展示信号在频率域上的能量分布，帮助我观察信号的主频成分及其谐波结构。

实现过程中，我对整段音频信号执行实数快速傅里叶变换（`rFFT`）。由于实数信号的频谱是共轭对称的，只需关注正频率部分。计算变换结果的幅度，并绘制频率-幅度曲线。

下图为该音频信号的幅度谱（展示前 10% 的频率范围），从中可以清晰地分辨出信号的主要频率分量。

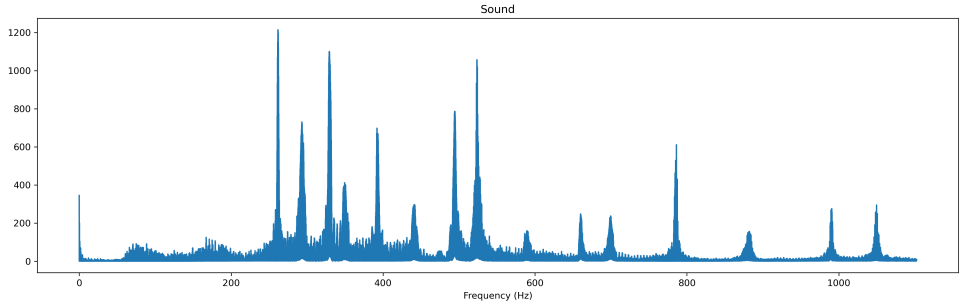

图 3: 幅度谱（rFFT，展示 10% 频率范围）

4 Task 3：语谱图与 Mel 语谱图

语谱图结合了时域和频域的信息，展示了信号频率随时间的演变过程。

我通过短时傅里叶变换（STFT）实现这一分析：在时间轴上对信号分帧，并对每一帧进行 FFT 变换，从而获得时间-频率-能量的分布矩阵。接着我分别绘制了线性频率刻度和对数分贝刻度的语谱图。

通过对比可以发现，线性频率刻度的语谱图往往难以直观展示低频部分的细节，因为大部分能量集中在低频段，而线性轴将高频部分的空白区域拉得过长，导致关键信息被压缩，视觉上“基本啥也看不到”。相比之下，对数分贝刻度的语谱图通过对能量取对数，显著增强了弱信号的可视性，使得频谱结构更加清晰。

此外，由于人耳对频率的感知是非线性的（对低频更敏感），程序还给出了绘制 Mel 语谱图的代码。它基于 Mel 滤波器组将线性频率映射到感知域，能够更好地模拟人类听觉系统对声音的响应。从结果来看，Mel 语谱图在视觉上最为直观，它剔除了人耳不敏感的高频冗余信息，重点突出了语音信号在低频段的共振峰结构。

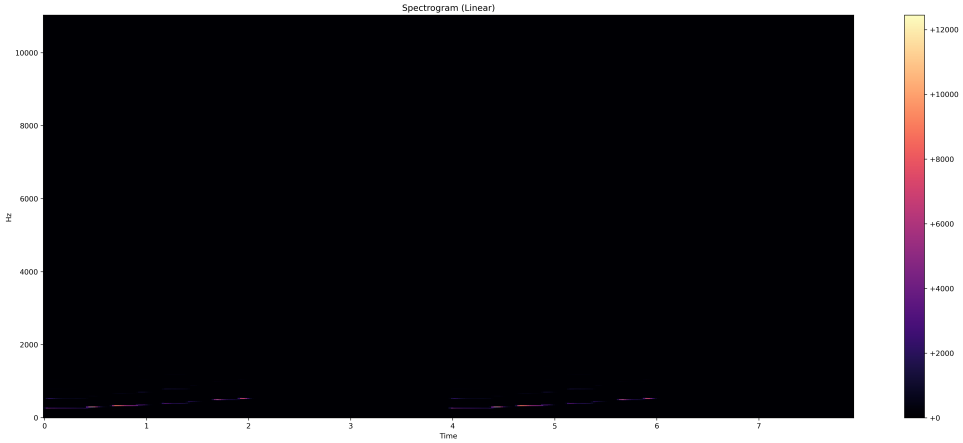

图 4: 语谱图（Linear dB）

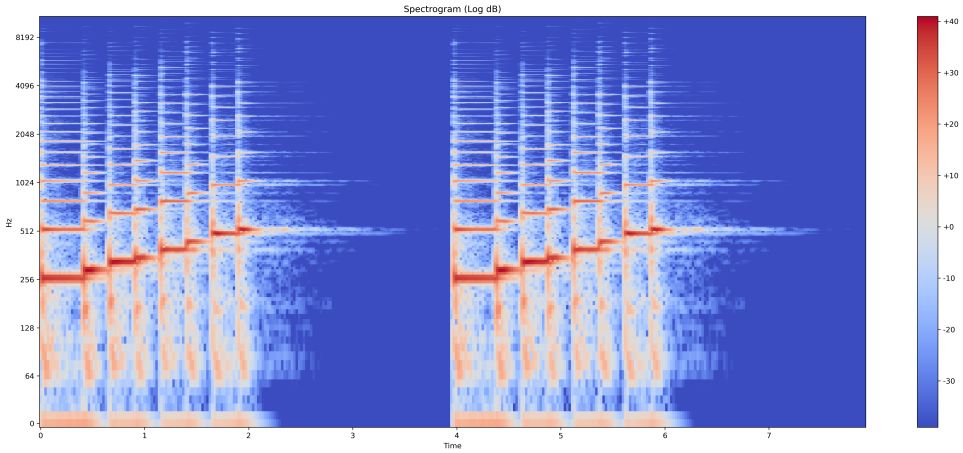

图 5: 语谱图（Log dB）

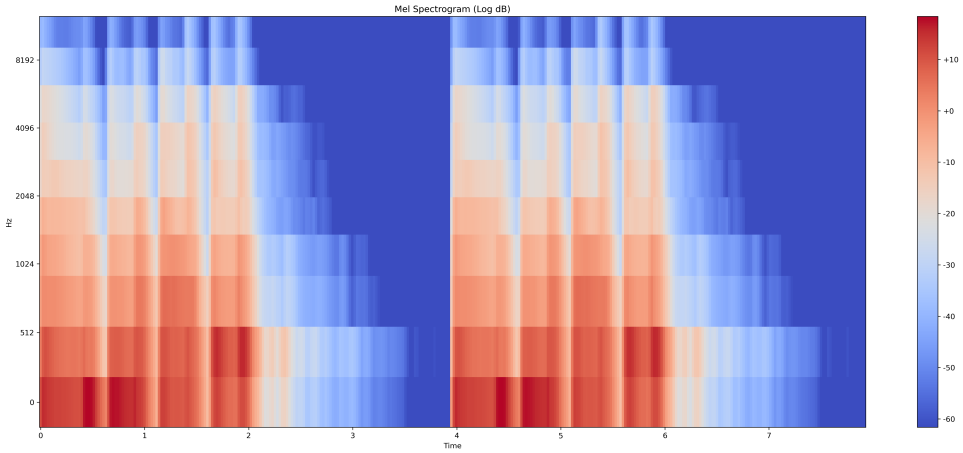

图 6: Mel 语谱图（Log dB）

5 Task 4：MFCC 特征提取

梅尔频率倒谱系数（MFCC）是语音识别和音色分析中最经典的特征之一。它在 Mel 能量谱的基础上，通过对数变换和离散余弦变换（DCT）得到。

MFCC 能够提取出频谱的包络信息，去除高频的细节噪声，得到一组低维且稳定的特征系数。

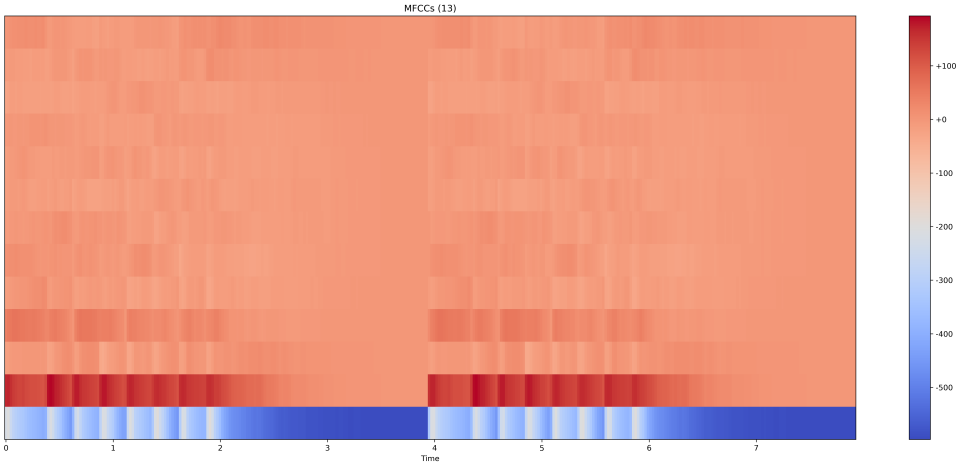

图 7: MFCCs（n_mfcc=13）

6 总结

通过本次实验，我深刻体会到了语音信号处理实际上是一个不断提炼核心信息、去除冗余噪声的过程。从最直观的时域波形到频域的幅度谱，再到引入时间维度的语谱图，最后到模拟人耳感知的 Mel 语谱图和 MFCC，每一步变换都有其特定的物理意义和应用价值。

在代码层面，本实验不仅让我初步掌握了 Librosa 库的使用，更建立起了从信号处理理论到实际代码实现的桥梁。