Digital Signal and Image Processing

# Programming Homework #4

Qiu Yihang, 2022/04/21-04/26

## 00 Tools

使用 MATLAB 进行本次实验。所使用的 MATLAB 版本为 R2021b。

## 01 Waveform and Spectrogram of Original Audio File

本次实验使用的素材是宇多田光的《夕凪》。该曲的一些细节如下：（1）开始处有较轻的人声吟唱和环境音；（2）宇多田光的声音成分中始终都有一条较沙哑的声线；（3）原曲中接近“鼓点”的是定音鼓和钢琴重音，但两者原本都没有非常明显；（4）原曲中有贝斯的旋律，但是相对音量较轻，几乎淹没在其他乐器的声音中。

全曲的时域波形图和 STFT 时频图如下。（由于源文件是双声道，因此对两个声道都进行了可视化处理。）

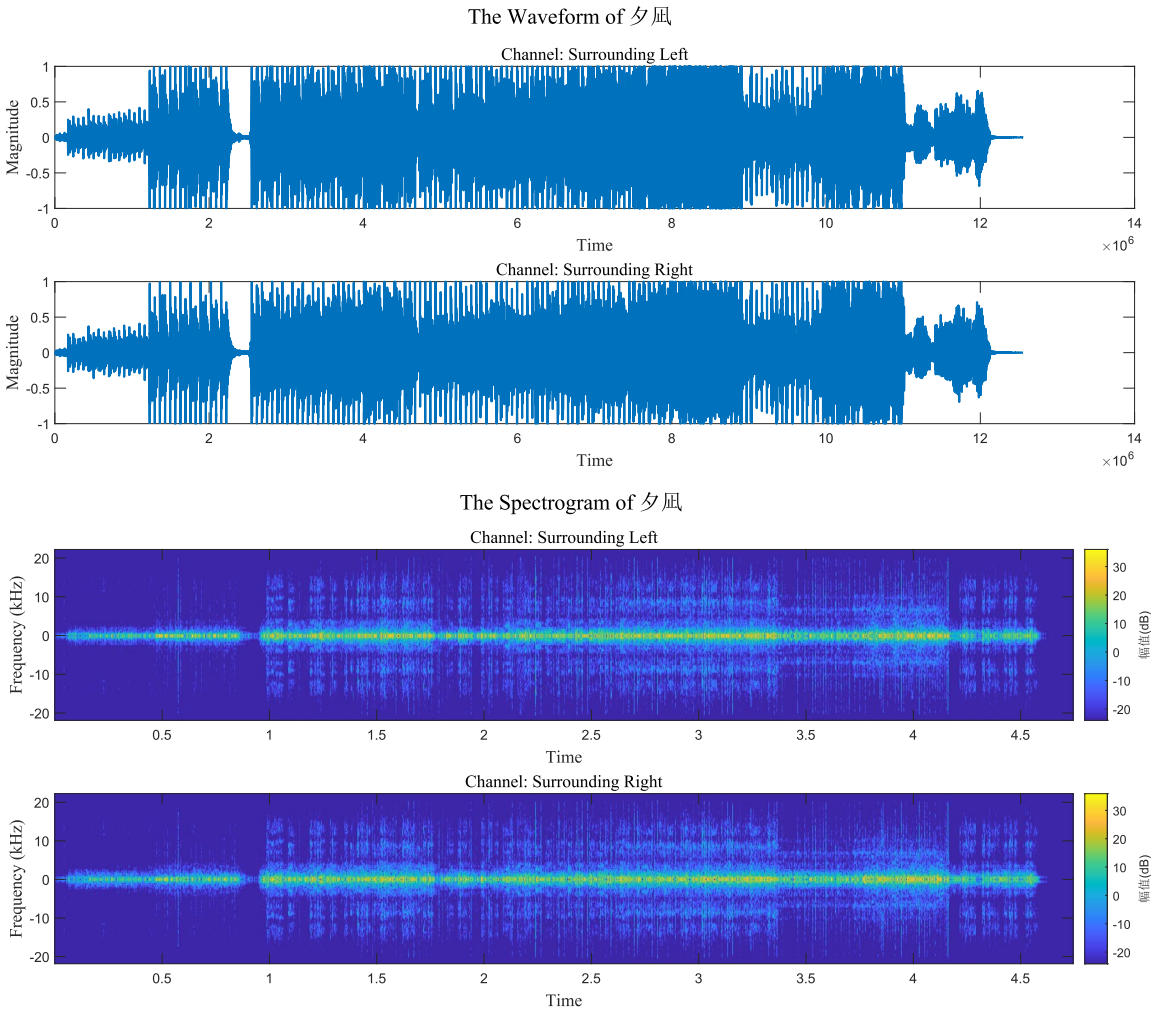


图 1 处理前音频的时域波形图与 STFT 时频图

## 02 Downsampling

考虑到原始音频的采样率与下采样的频率（5kHz、10kHz、15kHz）未必是倍数关系，因此我们在倍数关系成立时直接使用 MATLAB 自带的 downsample 函数（因为用时较短），在倍数关系不成立时使用 resample 函数采样。以上功能由笔者缩写的 downsampling 函数实现。

本次实验中原曲采样率是 44100Hz，因此最终使用 resample 进行采样。

播放 3 段下采样后的音频，可以发现高频部分的声音明显消失。如 **01** 中所说的细节（1），5kHz 下采样后的音频中吟唱部分的“ta”音几乎消失，而 10kHz 和 15kHz 下采样后的音频中这一损失并不明显；三个下采样后的音频中，人声部分均变得非常模糊。

下采样后音频的时域波形图和 STFT 时频图如下。

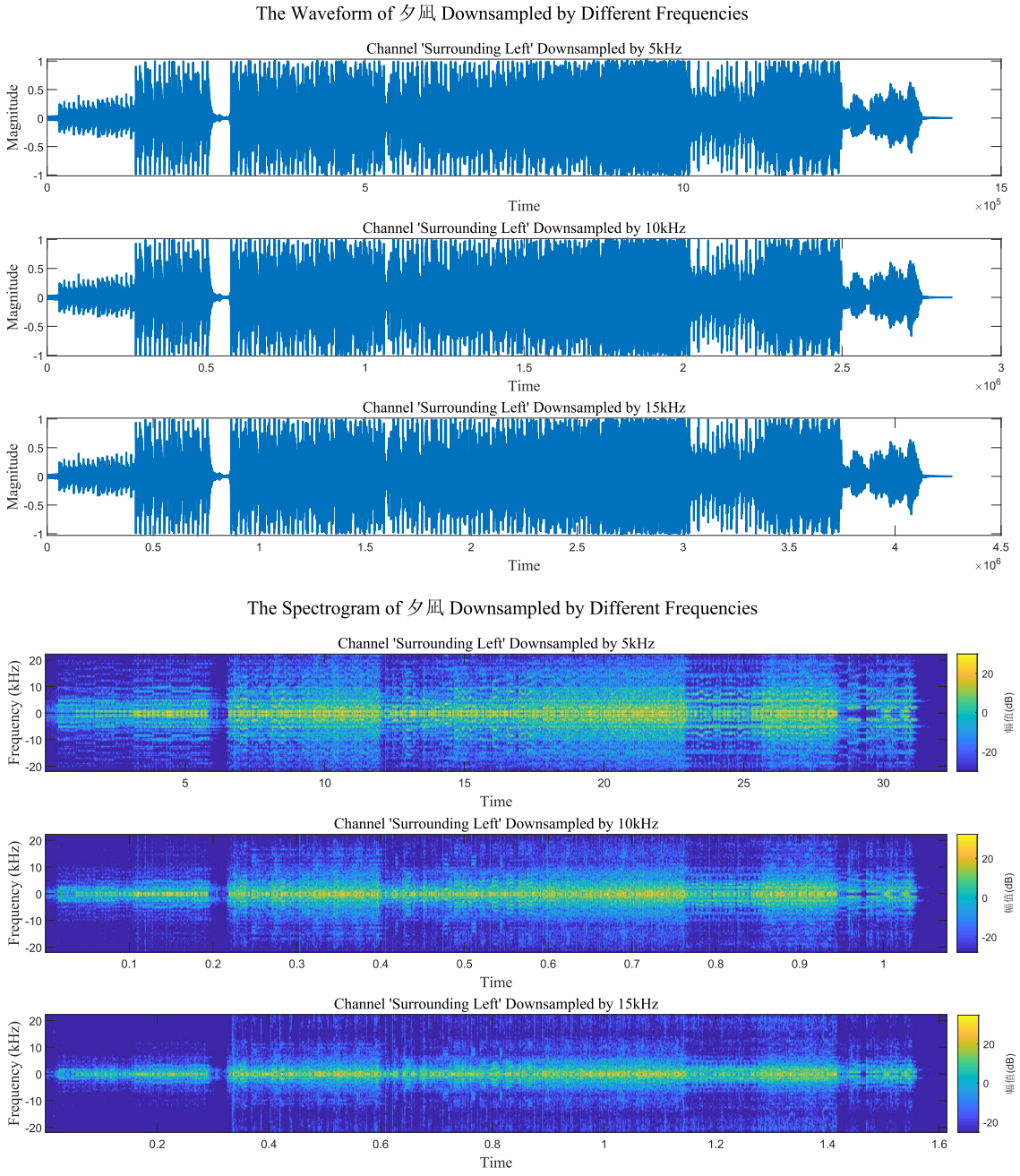


图 2 下采样后音频的时域波形图与 STFT 时频图

对比图 2 中不同频率下的 STFT 时频图可以发现，下采样频率越低，STFT 的高幅值区段在频域上的宽度越大，即在频域中的分辨率降低。由于时域波形图展示的是全曲，难以直接看出不同下采样频率下采样结果的差异，因此我们选出某一小段（原曲的 4718700~4727511 帧，即原曲的 01:47 附近），对比不同频率下采样的时域波形图。对比图如下。

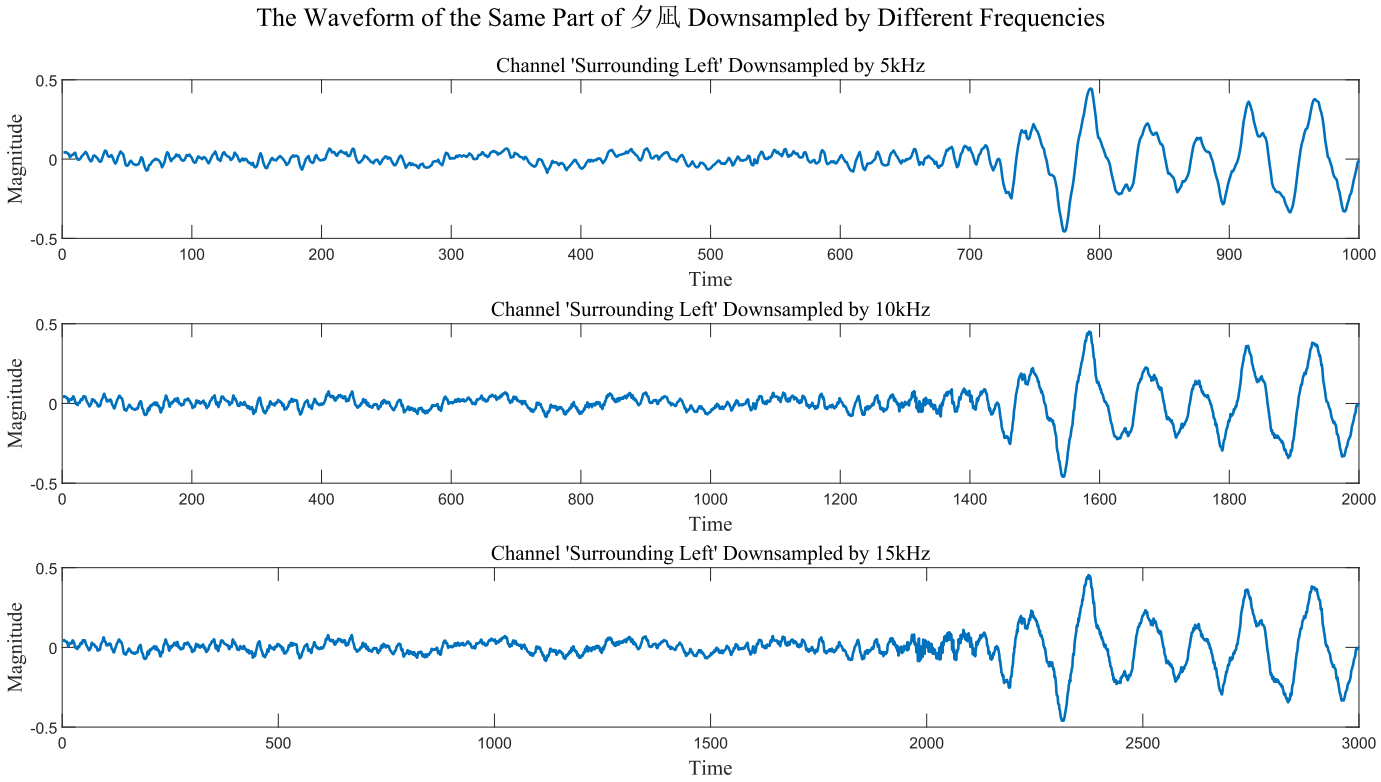


图 3 不同频率下对同一片段的下采样结果对比

对比三种频率下的同一段落（如 5kHz 下第 600~700 帧的段落，即 10kHz 下的第 1200~1400 帧、15kHz 下的第 1800~2100 帧），可以发现下采样频率越低，损失的信息越多。

## 03 Interpolation

使用 griddedInterpolant 对下采样后的音频进行插值恢复。我们尝试 linear、spline 两种插值方式，对比其恢复结果。

使用线性方法（linear）进行插值恢复后的音频时域波形图和 STFT 时频图如下。

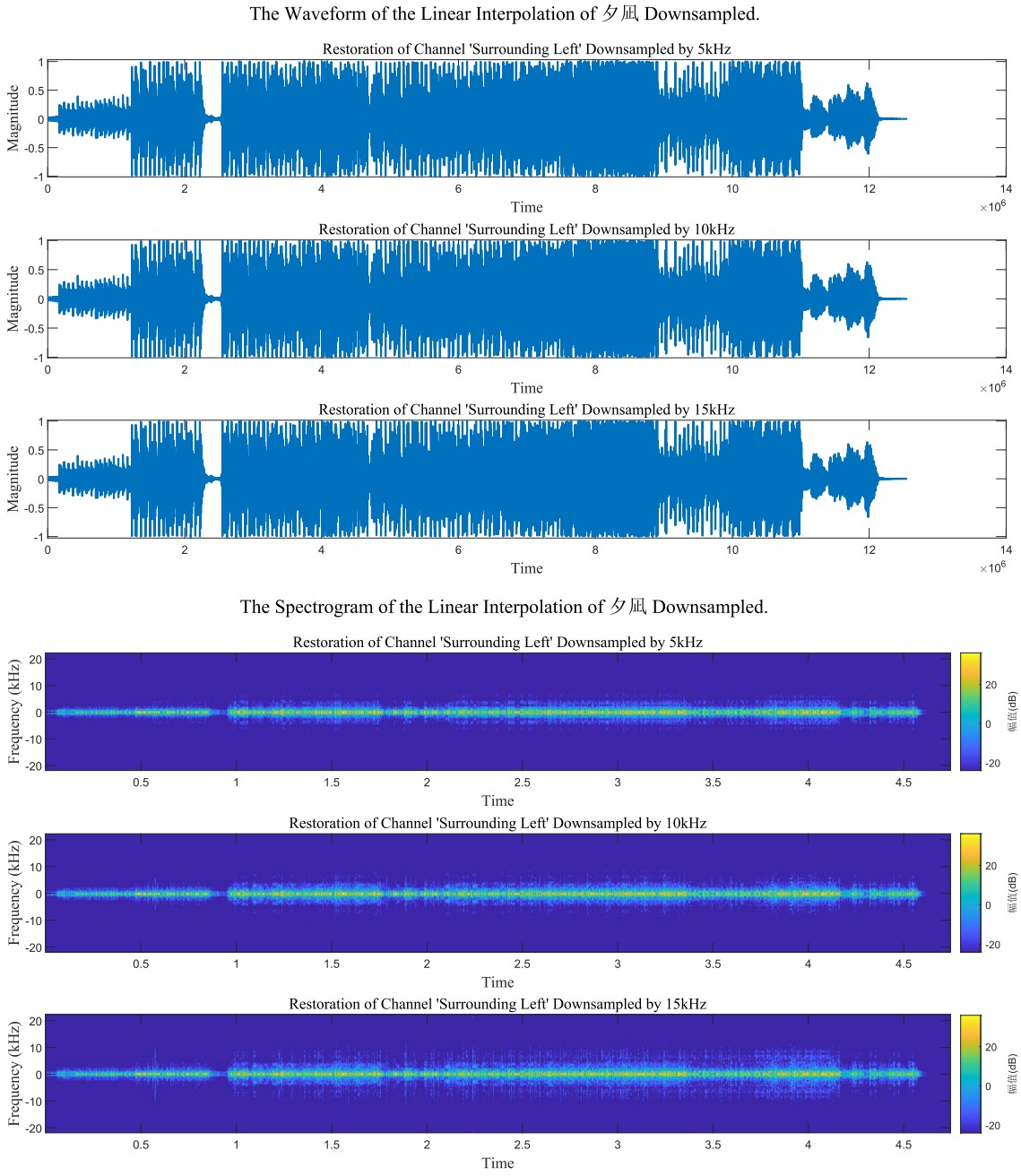


图 4 线性插值方法恢复的下采样音频的时域波形图与 STFT 时频图

可以发现，线性插值方法提高了音频在频域上的分辨率，在频率轴上的宽度变小了。但是与图 1 中的原始信号 STFT 时频图对比可以发现，原本许多时间点上 STFT 在较宽频率上均有较高幅值，但在下采样和插值后有较高幅值的频率宽度也被大幅衰减。

实际上，对比线性插值恢复后的音频与原始音频可以发现，虽然线性插值的恢复让下采样后的音频从“模糊”的状态变得清楚了一些，但是也产生了非常明显的类似于“电音”的切片效果。此外，下采样频率越高，线性插值恢复后的音频越接近原始音频。但即使是 15kHz 下采样音频线性插值后的结果也存在明显的“切片感”和噪音。

使用三次样条函数方法（spline）进行插值恢复后的音频时域图和 STFT 时频图如下。

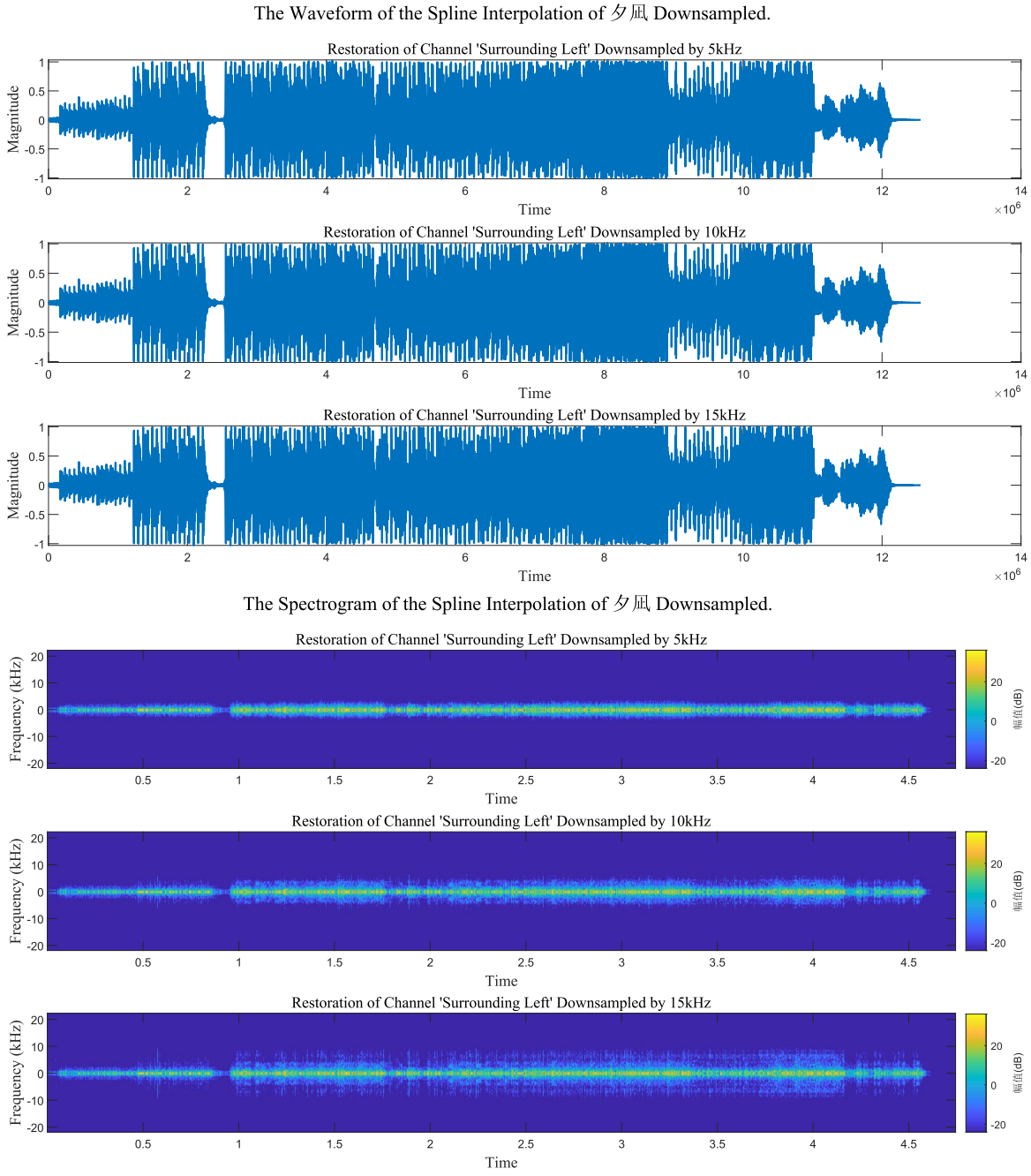


图 5　Spline 插值方法恢复的下采样音频的时域波形图与 STFT 时频图

对比图 5 和图 4 可以发现两种插值方法的效果均提升了频域上的分辨率，但原始音频的许多时间点处 STFT 在较宽频率上均有较高幅值，但在下采样和插值后有较高幅值的频率宽度也被大幅衰减。

播放插值恢复后的音频，虽然 spline 插值恢复的音频没有线性插值带来的切片感和电音感，但噪声相对较多，音频依然比较模糊。与线性插值类似，下采样频率越高，线性插值恢复后的音频越接近原始音频。相对来说，15kHz 下采样音频 spline 插值后的结果几乎与原始音频相差无几，插值恢复的效果要优于线性插值的恢复结果。

# 04 Equalizer Design

## 4.1 Vocal & Percussion Enhancement and Attenuation

首先我们设计了一些特殊的滤波器，用于增强（或减弱）人声和增强（或渐弱）鼓点与低频音。详情见 `DSIP_04_TASK_04.m` 中的 Vocal_Enhance 和 Percussion_Enhance 函数。

我们设计的人声增强滤波器、低频增强滤波器的频率响应如下。（前者带通，后者低通）

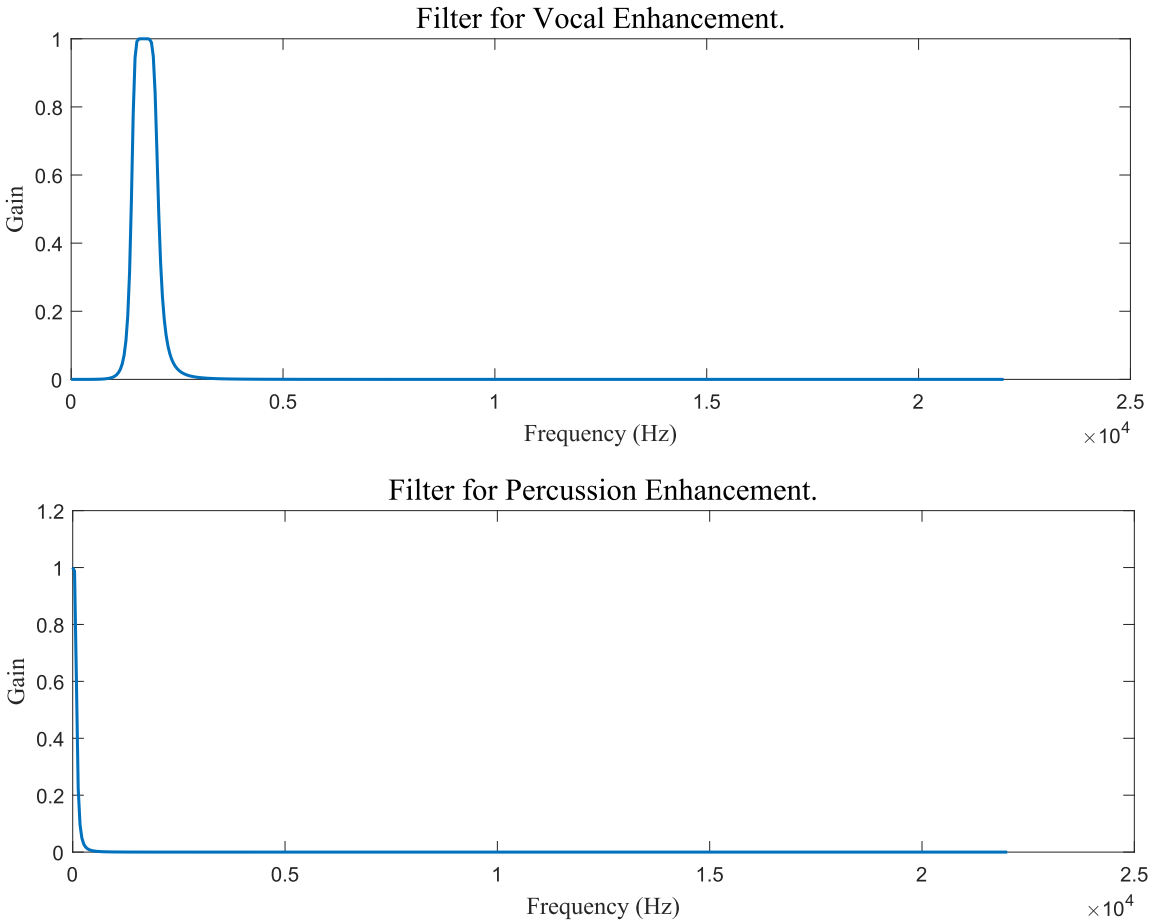


图 6 人声增强滤波器和低频增强滤波器的频率响应

经过人声增强滤波器和低频增强滤波器处理后的音频（夕凪_Vocal.wav 和夕凪_Perc.wav）有如下特点：（1）前者的人声旋律响度得到增强，高音部分音色相对更通透，但同时也增强了开头原始音频所采样的环境音，并引入了较多噪音；（2）后者的低频音被大幅增强，钢琴重音和定音鼓的音量被增强，低频的贝斯旋律被放大，打击乐的震动感被增强。

## 4.2 Equalizer

均衡器是一系列在不同频率波段产生效果的带通滤波器。

我们设计的均衡器中共计十个带通滤波器，它们的频率响应如下。（为能够更好地显示每个滤波器的效果，第二张图中频率轴使用了对数轴。）

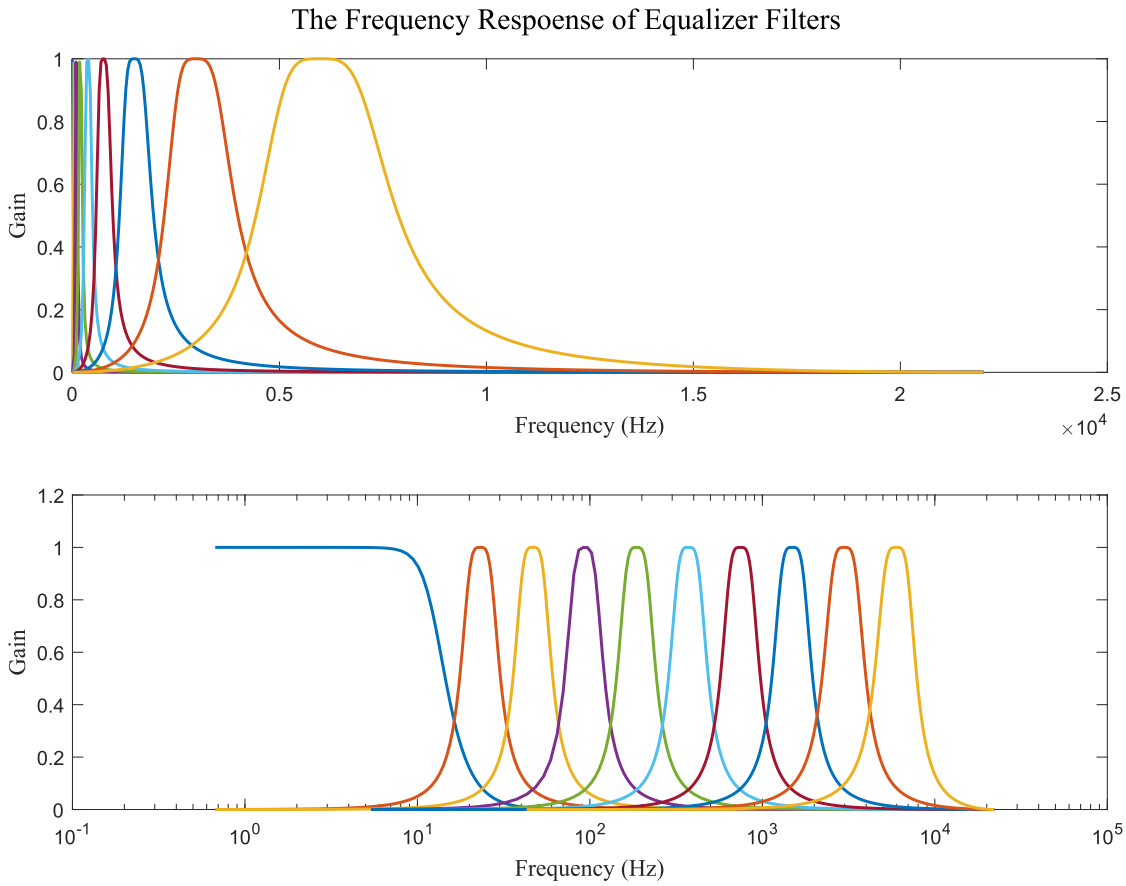


图 7 均衡器中滤波器的频率响应

使用时，只要调整每个滤波器生成的分量的增强程度（实际上我们的实现方式是调节各分量在生成音频中的占比）即可。压缩文件中的“夕凪_equalizer.wav”是各个滤波器增强程度设置为[1.0, 0.8, 0.5, 0.3, -1, -3, 2, -1, -2, 3]后生成的音频，对比原音频可以发现，原本附带的沙哑声线的存在感被减弱，人声相对更加清晰、明亮；同时鼓点和贝斯旋律也被放大。