

MATLAB 中实现如下功能：针对本实验提供的 39 个音频文件，利用提供的 "get\_tune\_slow.p"函数截取片段，并利用短时傅里叶变换对截取的片段进行识别，在片段被慢放的情况下，判断其属于哪个音频文件。完成以下任务：

1. 结合实验 PPT，你如何设置池化窗口的大小？池化的作用是什么？
2. 结合实验 PPT，在计算音频片段频谱与原音频频谱相关性时，你的滑动距离是多少？如何设置的？
3. 介绍你的思路，并将你实现上述功能的 MATLAB 函数或脚本复制粘贴在此处。
4. 在 get\_tune\_slow 函数中，分别设置输入，如下：

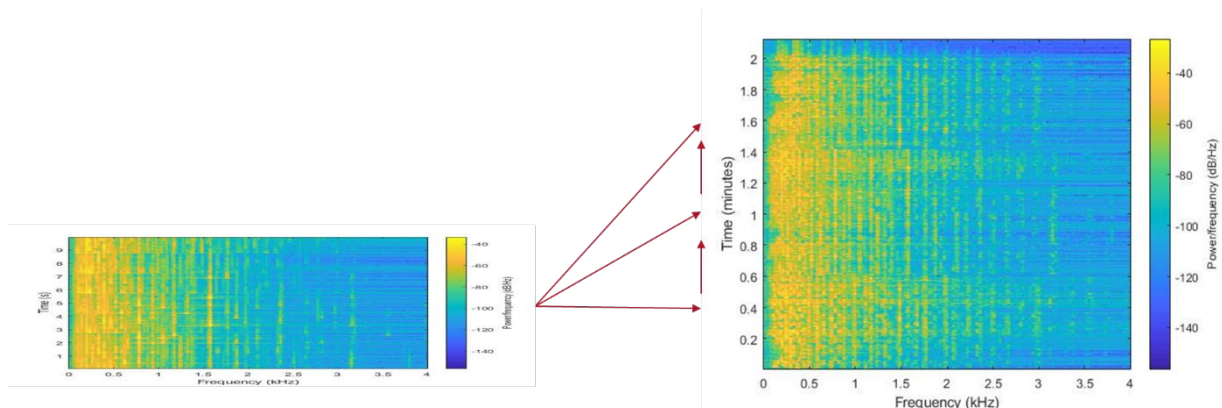
1) [clip, Fs\_clip] = get\_tune\_slow('u7654321');

2) [clip, Fs\_clip] = get\_tune\_slow('u88888888').

利用你编写的函数，判断上述两个 ID 截取的片段属于哪个原音频文件？生成对应原音频文件的频谱图，并在频谱图上分别标记这两个片段所处的大致位置。

答案：

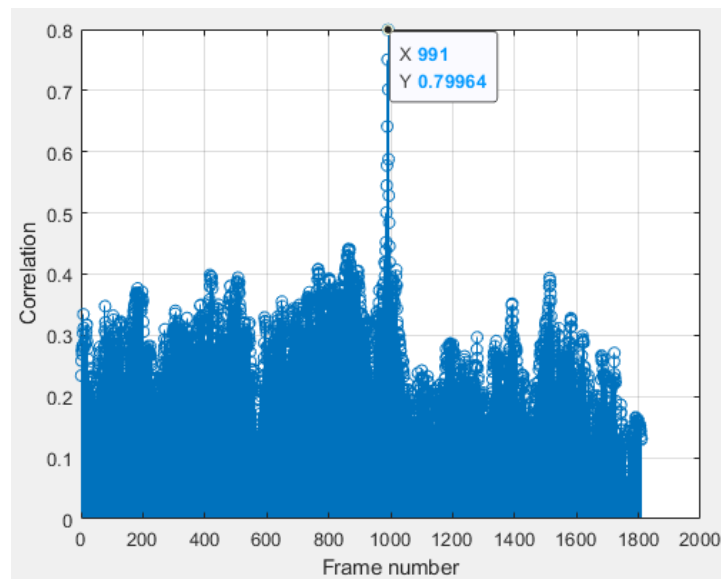
3. 实现思路：利用 STFT 对 39 个原始音频及截取片段分别生成频谱图，在每个音频频谱范围内，不断滑动片段频谱并利用 `corrcoef` 命令计算相关性，具有最大相关性的原始音频及时间区间即为该片段所述的音频及时间点。



4. 1) 片段“u7654321”来自于“rudenko\_26.mp4”，95-105 秒。

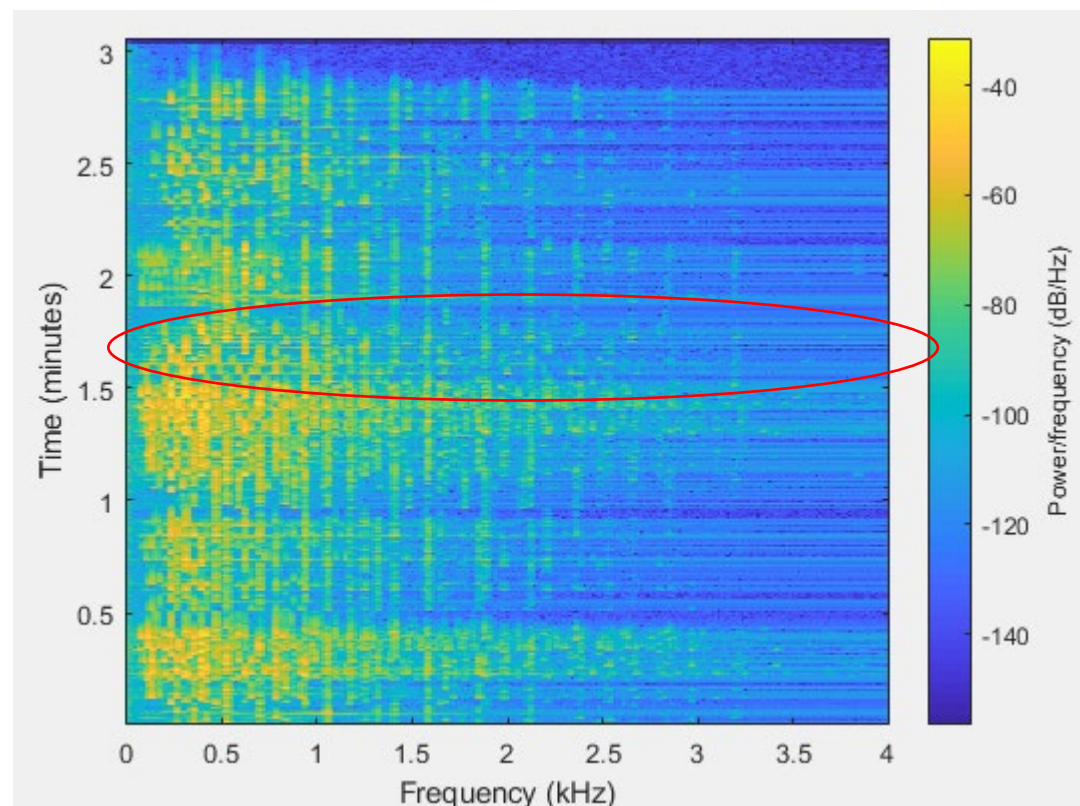
注：最大相关性时间点如何确定？

3X3 池化窗口，滑动步长为 1，所得相关性结果：

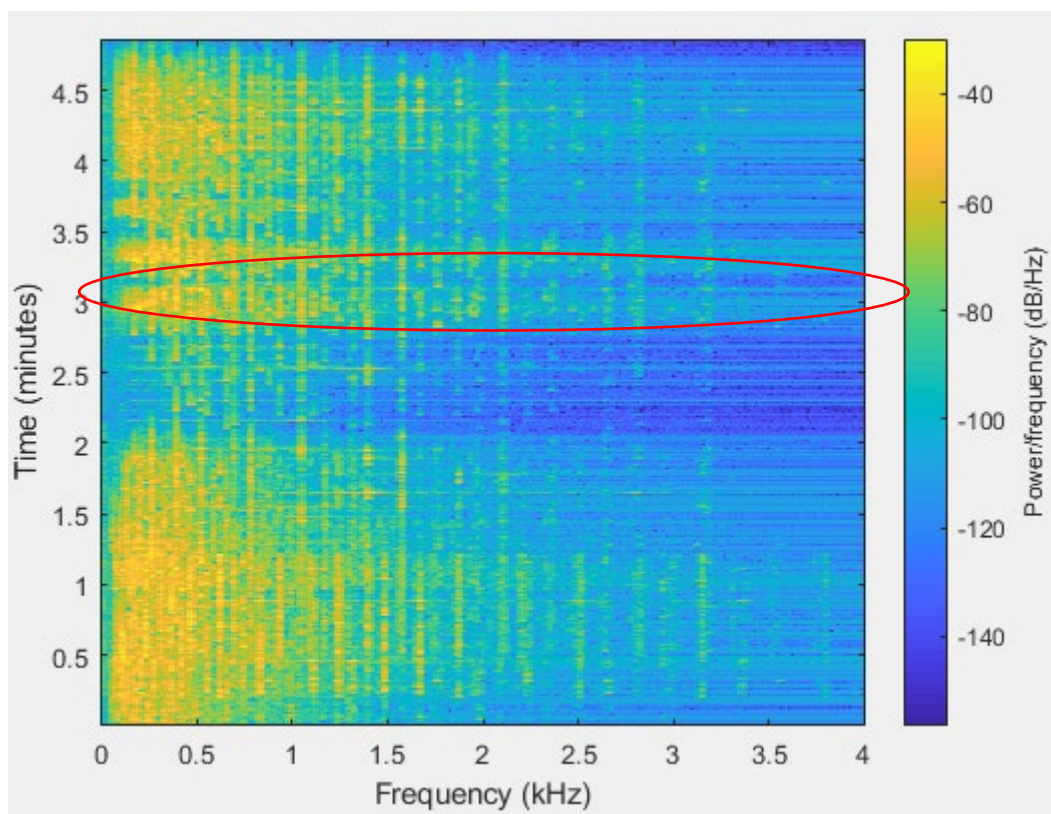
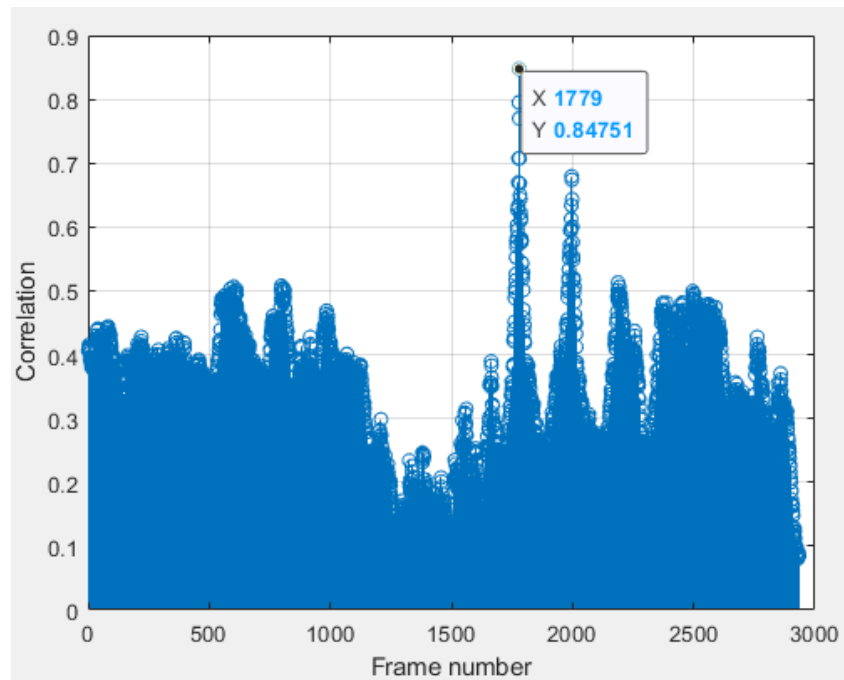


STFT 窗长 64ms，重叠部分 32ms；经过 3X3 的池化，窗长 192ms，重叠部分 96ms。滑动步长为 1，则每滑动一次，时间向后推 96ms。

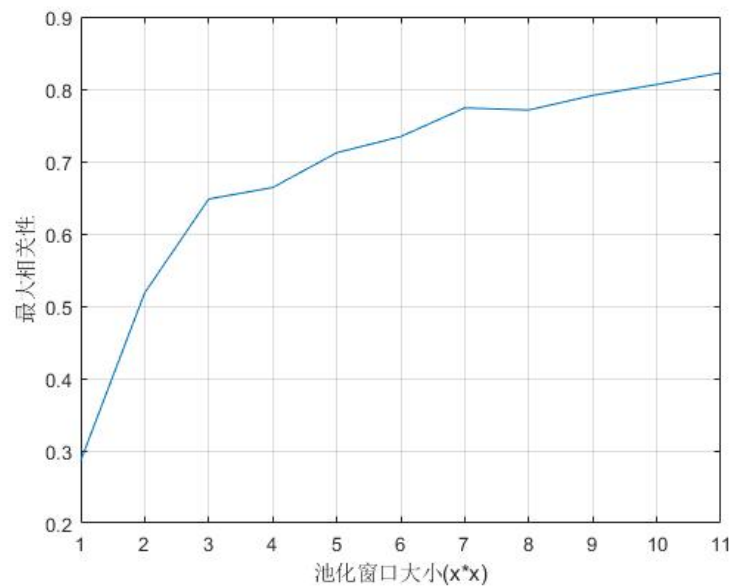
最大相关时间点： $96 \times 10^{-3} \times 991 = 95.136\text{s}$



2) 片段“u8888888”来自于“rudenko\_08.mp4”，171-181 秒。



1. 池化的作用：消除随机变量对特征的影响（对神经网络模型来说，提高泛化性能，避免过拟合）。



保持滑动步长不变，一定程度的增大池化窗口大小，可在一定程度上降低随机变量（噪声）对频谱相关性的影响。

2. 较小的滑动步长表示，对频谱上更多相互存在重叠的区域计算相关性。在此实验中，较小的滑动步长表示对时间进行更加细微的分割，从而最终计算得到的匹配的时间点更加准确，然而计算量更高。

