

实验二: 按键音识别

概述



按键手机按下每个键，会发出相应的声音

按键声音实则由高频正弦信号+低频正弦信号叠加而成

叠加方式如下表:

	1209 Hz	1336 Hz	1477 Hz
697 Hz	1	2	3
770 Hz	4	5	6
852 Hz	7	8	9
941 Hz	*	0	#

任务

- 输入
 - 十秒的音频，采样率为48000Hz
 - 每段声音由数个叠加高斯噪声的按键音组成
 - 按键音的时长是不固定的
 - 测试样例:
 - 在测试样例test.wav中，按键1持续了1秒，静默2秒，按键5持续3秒，静默3秒，按键#持续0.5秒，静默0.5秒

- 输出:
 - 以1/64秒为一帧，输出每帧对应的按键，如该帧静默，则输出-1
 - 示例格式:1 1 1 ... 1(重复64次) -1 -1 -1(重复128次) 5 ... 5(重复192次)...
 - 每帧输出之间加一个空格
 - 具体样例输出见test.txt
 - 将通过5个测试点进行测试
-

运行环境

- Python 3.6
 - Numpy >= 1.19
 - Librosa==0.9.2
 - 我们将提供音频读取部分的代码，其余部分需要各位同学实现
 - 运行方式 `python main.py --audio_file test.wav`
 - [Numpy有fft, rfft等实现好的快速傅里叶变换方法](#)
-

方法介绍

- 第一步: 判断当前音频是否静默
 - 计算一帧的平均能量(root-mean-square)
- 第二步: 判断非静默帧的按键
 - 计算当前帧的短时傅里叶变换(STFT)
 - 什么是STFT? 将长时间信号分成数个较短的等长信号，然后再分别计算每个较短片段的傅里叶变换,可同时描绘频域与时域的变化
 - 根据短时傅里叶变换，判断当前帧的按键