《Matlab信号处理及仿真》第八次作业报告

撰写人：邱楚寒　学号：2020209023026　班级：录音工程

本次实验前阅读了**《An Automatic Singing Skill Evaluation Method for Unknown Melodies Using Pitch Interval Accuracy and Vibrato Features》**及基于其的改进文章**《Perceptual Evaluation of Singing Quality》**中的颤音检测部分，受益匪浅。

> **《An Automatic Singing Skill Evaluation Method for Unknown Melodies Using Pitch Interval Accuracy and Vibrato Features》** ***Nakano et al***

本篇文章对Vibrato的检测和特征提取算法进行了细致的描述，主要思路是对提取的基频曲线再次进行STFT（原文对基频曲线使用32点长度的Hanning窗进行分帧），变换得到功率谱密度函数$X(f,t)$，接下来对其沿频率轴进行归一化：

$$\hat{X}(f,t)=\frac{X(f,t)}{\int X(f,t)df} \tag{1}$$

基于归一化的功率谱密度函数，根据下式得到功率函数$\Psi_v(t)$和锐度函数$S_v(t)$：

$$\Psi_v(t)=\int_{F_L}^{F_H}\hat{X}(f,t)df \tag{2}$$

$$S_v(t)=\int_{F_L}^{F_H}|\frac{\partial\hat{X}(f,t)}{\partial f}|df \tag{3}$$

其中，F_H,F_L表示震荡频率范围，文献中定义为5～8Hz。由$\Psi_v(t)$和$S_v(t)$即可得到Vibrato似然值：

$$P_v(t)=S_v(t)\Psi_v(t) \tag{4}$$

根据每一帧的似然值大小并与设定的阈值比较，可以判断出最符合Vibrato颤音特征的时间戳，进而对对应的时间帧进行波动频率(rate)和深度(extent)的分析：

$$\frac{1}{rate}=\frac{1}{N}\sum_{n=1}^{N}R_n \tag{5}$$

$$extent=\frac{1}{2N}\sum_{n=1}^{N}E_n \tag{6}$$

> **《Perceptual Evaluation of Singing Quality》** ***Haizhou et al***

本篇文章的Vibrato算法部分是基于前面那篇文章的算法的改进版本，主要改动部分在于提出了一种修正Vibrato似然值的算法：

$$P_{v_{mod}}(t)=\frac{\int_{F_L}^{F_H}X(f,t)df}{\int X(f,t)df} \tag{7}$$

修正似然值变成了在振荡频率范围内的基频功率谱能量和总功率谱能量之比，这使得最终可以得到一个在0～1范围内的概率值，这样一来设置阈值就变得非常方便了，解决了原先算法中对阈值定义不明确的缺点。

对于本次实验，主要利用上述论文中的算法对得到的基频曲线进行分析，进而分析颤音Vibrato的rate和extent；同时，还对比了颤音和非颤音音频的Shimmer和Jitter，并利用这两个参数比较颤音在这两个参数维度上的特征。由于时间所限，对时域峰值振幅曲线影响的Tremolo颤音没有进行针对性的分析。

在进行实际的实验过程中，发现了原文献中的一些参数并不适用于所有的音频文件，对于不同采样率的音频文件Tremolo颤音的阈值以及必然会有区别，因为这一特性，我在原始算法的基础上改变了一些参数，比如最后的振荡频率范围和阈值设定，这样一来才得以得到一些满足颤音条件的帧。

因为文献中的算法是针对基频曲线的短时傅里叶变换，文献中所谓“5～8 Hz”和原始音频的傅里叶变换频域分辨率不同，需要根据第一次分帧加窗时使用的帧长和帧移进行换算，根据推导，得出基频曲线的采样率为：

$$SampleRate = \frac{sr - winlength}{noverlap} + 1 \tag{8}$$

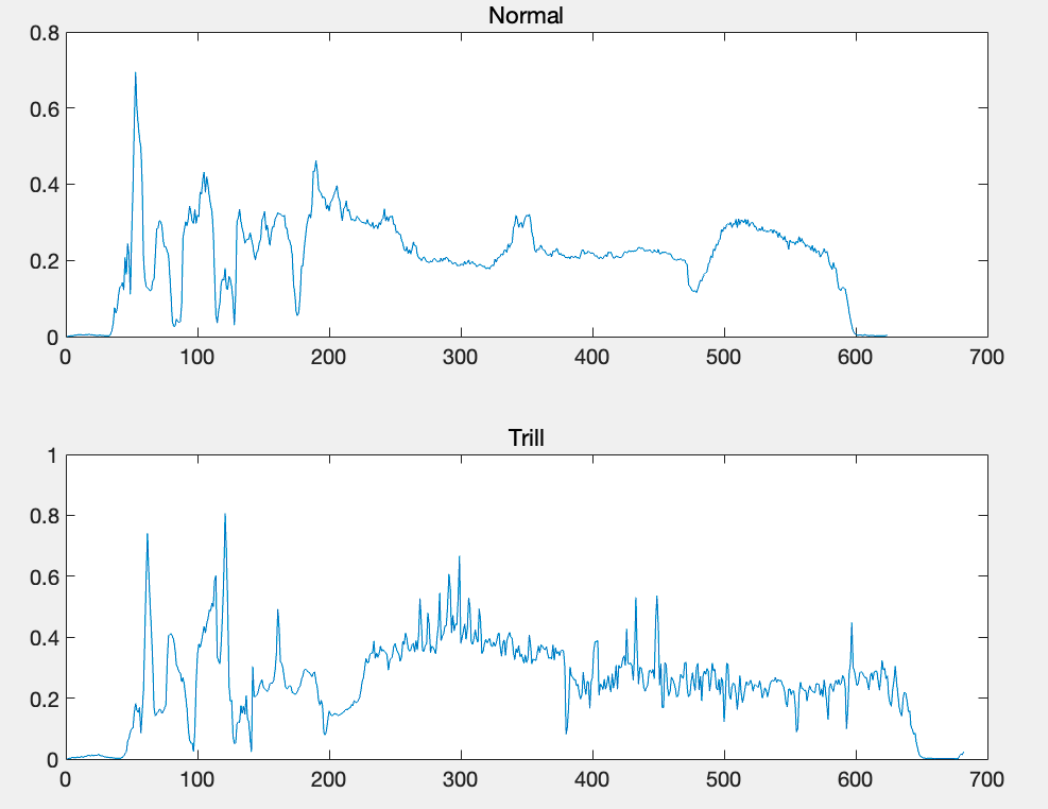
其中，$winlength, noverlap$ 分别代表对原始音频进行分帧时的帧长和帧移，sr 代表原始音频的采样率大小。

得到基频曲线的采样率后才能得到其短时傅里叶变换后的频率分辨率，进而得到震荡频率范围内的频率下标。

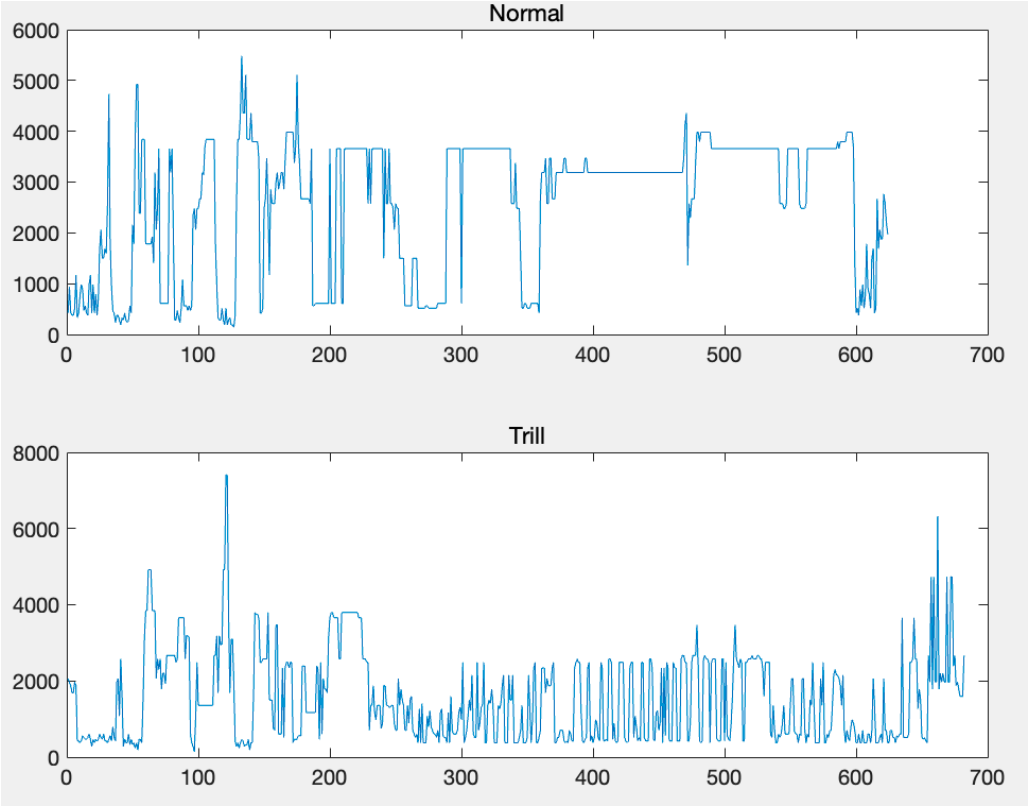
首先对每个函数文件进行介绍：

> pre_processing.m：对音频进行预处理，主要是进行取单声道和STFT操作，并根据帧长和帧移由式(8)返回后续基频曲线的采样率**f_sr**。
>
> time_domain_curve.m：对分帧后的音频取峰值振幅曲线，得到时域上音频振幅的时变特性。
>
> frequency_domain_curve.m：对分帧后的信号求取基频曲线，同样利用findpeaks函数并设定一些参数，然后峰值间距取平均值得到近似的基频大小，这里还需要将得到的下标和频域分辨率进行相乘，得到以Hz为单位的基频大小，在这个过程中因为限制了峰值最低高度，所以可能出现找不到基频的情况，为了后续的计算不出错误，我这里选择去掉不存在基频的帧。
>
> Tremolo.m：利用PPT里的公式，分别计算出Shim和ShdB参数，因为ShdB的求取方式是比值求对数，可能存在帧内峰值为零的情况，所以要将这种情况考虑在内，计算ShdB前设置前置判断条件，满足**amp_ratio**不为空或无限时才进行计算。
>
> Vibrato.m：利用PPT里的公式，计算出Jitter，计算方法与前面的Shim和ShdB基本一致。
>
> Vibrato_analyse.m：根据前文提到过的文献中的算法，对Vibrato颤音的检测和特征提取算法进行复现，这里我认为主要的难点首先是理解对基频曲线分帧加窗的思路，并找出初次分帧加窗时的窗长和帧移与基频曲线之间的关联；以及在检测出满足修正Vibrato似然值条件的帧后如何提取extent参数。前者解决方案如式(8)所示，后者我的思路是先单独提取出满足似然值条件的帧并放在一个矩阵中，然后遍历每个帧，将帧整体减去其平均值从而使其中心移动到零点，进而利用过零率的办法计算出其震荡周期，将此振荡周期作为extent参数进行计算。

这次作业使用了三个音频文件，其中**normal.wav**和**trill.wav**是找女朋友录的两个分别不颤和颤音的音频，为了更好提取颤音特征，还专门从网上的Freesound找了一段质量较高的女声音频，命名为**singing.wav**。由上述文件所得到的时域时变幅度峰值曲线如下：



基频曲线如下（纵轴为频率，横轴为时间）：



因为没有做中值滤波所以存在一部分野点，但依然能很明显的看出，颤音因为无论是时域还是基频，都比非颤音信号多很多幅度和频率上的震荡，这一显著的特征也是颤音与非颤音信号区分的最主要特点，后续对颤音的检索和特征提取也是基于颤音的这一特性而做的。

然后计算出两者的Shim、ShdB和Jitter参数如下：

Shim_norm	0.0638
Shim_trill	0.1265
ShdB_norm	0.8712
ShdB_trill	1.3456
Jitter_norm	277.2622
Jitter_trill	515.3497

对比以上三者，可以很明显的看出无论是在哪个维度，非颤音信号三者参数都要远远低于颤音信号，因此Shim、ShdB和Jitter作为判定信号是否为颤音信号的标准是有很大可行性的。正因如此，后面即将提到的Vibrato实际上由Jitter来衡量的（或者说两者成正相关），而Tremolo则是由Shimmer来衡量颤音的“颤动程度”大小。其中，上课提到Shim参数与嗓音“病变”程度成正相关，根据这一知识前提可以看出颤音信号同时也对应着更加”病变“的声音。

接下来就是本次作业实现真正的难点所在，即上文中引用文献算法的复现。

如果想尽可能精确的得到基频曲线在5～8 Hz的能量占比，就应该让基频曲线FFT时能有尽可能高的频域分辨率。然而，两篇文献中使用的窗长均为32点，我也不能为了提高频域分辨率无端修改（并且经过实测，把此处的窗长提高所得结果也并不理想）。这里有过多局限性所在导致我不得不修改了很多参数，首先因时频域存在着“测不准原理”，无法在保证时域分辨率的情况下提高频域分辨率，在此前提下不得不对帧长和帧移做出取舍，第一次分帧的帧长为1024点，帧移512点（音频采样率为48 KHz），此时所得基频采样率为92.75 Hz，32点的窗长最终所得频域分辨率为2.8984 Hz，最终落在5～8 Hz范围内的频域点仅下标为*2* 的单个分量而已，此时计算结果远远达不到预期，甚至对于明显的颤音信号都没有一个时间戳能够满足原文0.4阈值的要求。

在大量、反复调试和修改了第一次分帧帧长、帧移与基频曲线的分帧帧长、帧移后，均无法达到预期目标，而且在修改这些参数的过程中还会影响到其他参数如Jitter、Shim的正常计算，所以不得不修改了似然值阈值要求和震荡范围，最终才获得了还算合理的结果。修改参数为：

> $F_H = 10Hz -> 20Hz$
>
> $F_L = 5Hz$
>
> $threshold = 0.4 -> 0.2$

最终对**singing.wav**文件提取的颤音特征参数为：

rate	3.8709
extent	2

本次作业仍然遗留的问题：

> 1、如何改进Vibrato检测和特征提取的算法，在不修改参数的前提下完成对颤音的频率和深度进行提取；
>
> 2、基于Tremolo的extent和rate的提取算法，还没来得及查阅相关文献。