语音合成大作业

无08 李煜彤

2020010841

目录

[1.2.1 语音预测模型 2](#_Toc110014536)

[1.2.2 语音合成模型 5](#_Toc110014537)

[1.2.3 变速不变调 6](#_Toc110014538)

[1.2.4 变调不变速 7](#_Toc110014539)

[文件清单 8](#_Toc110014540)

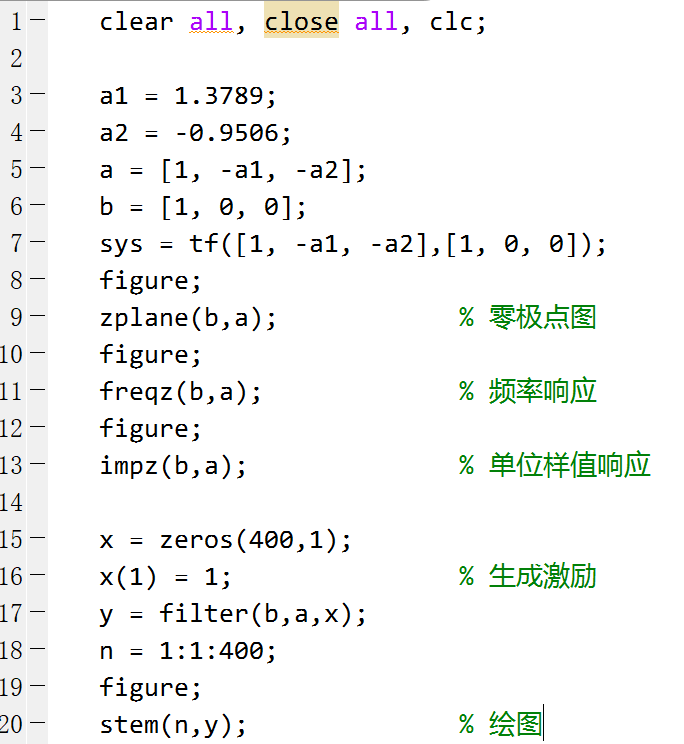
# 1.2.1 语音预测模型

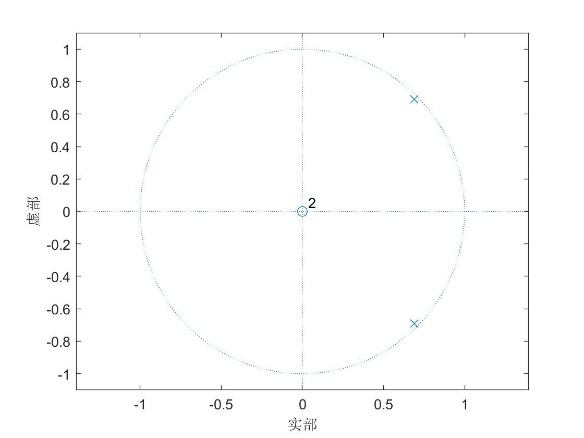
（1）

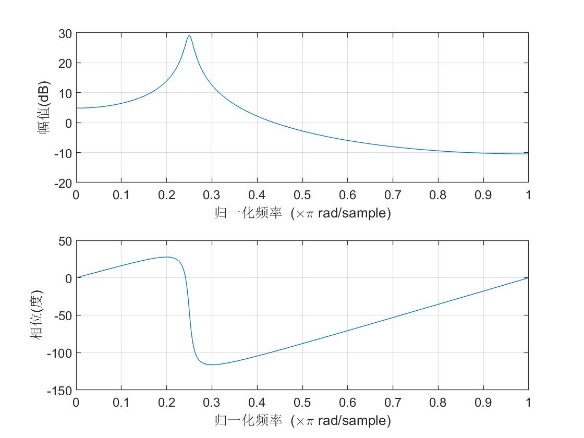
对求传递函数：

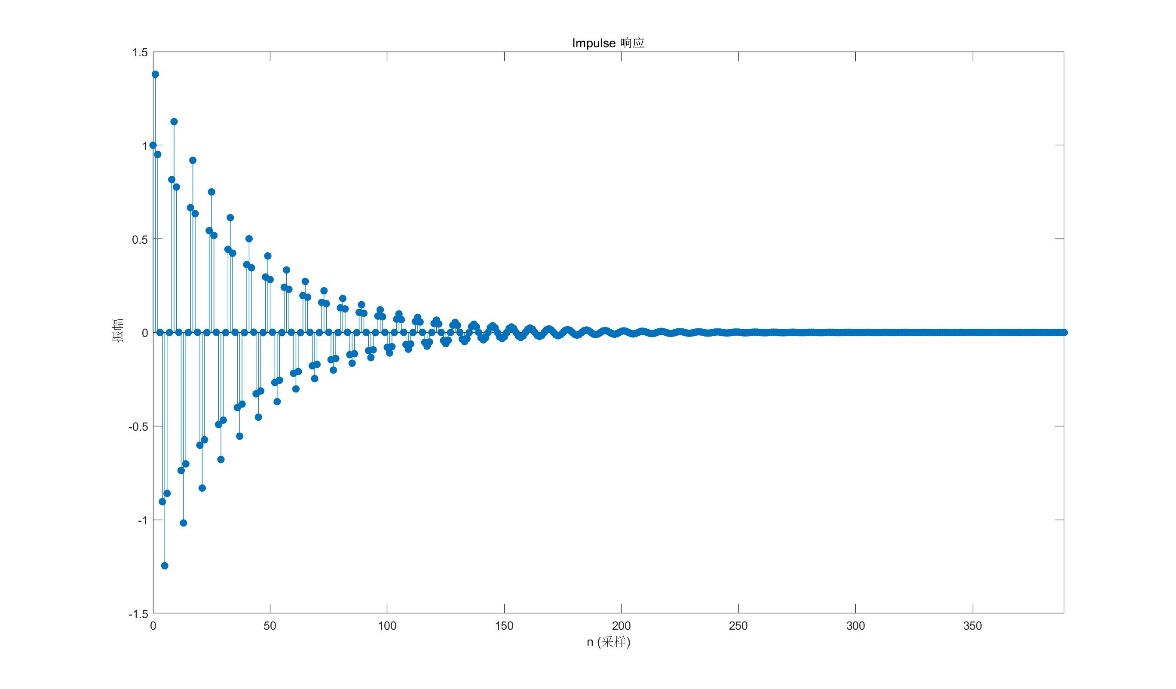
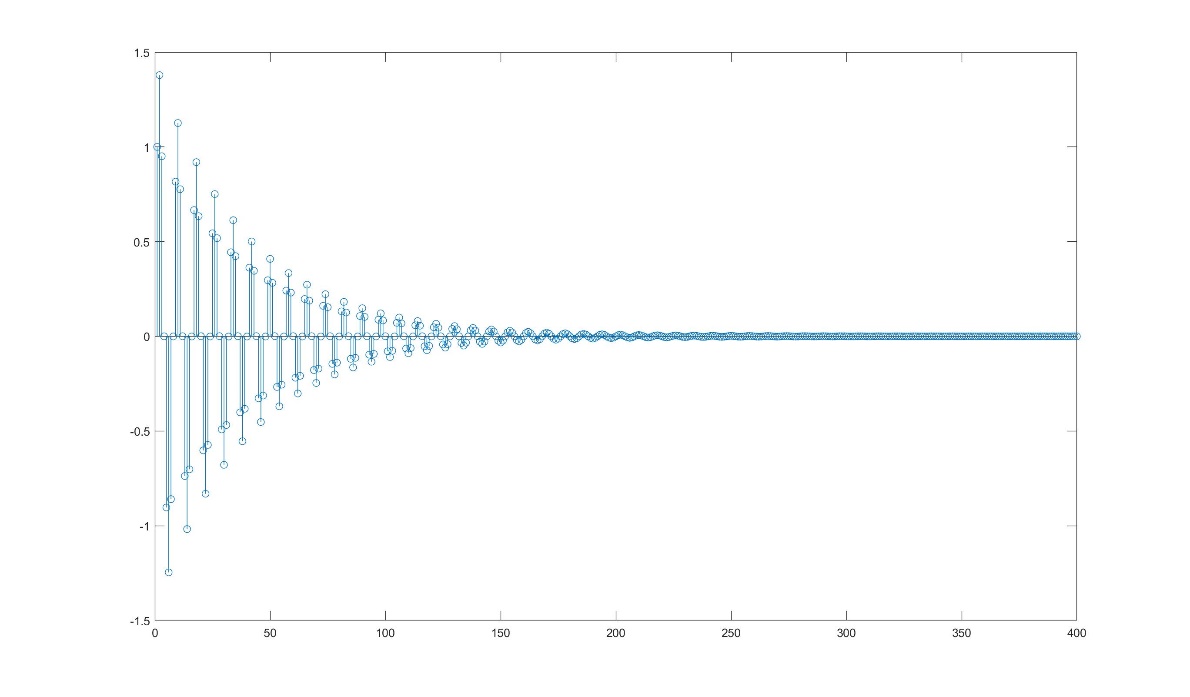
其中。

对于之后的预测模型，其采样频率为8kHz，即采样周期。则共振峰频率

为了绘制出零极点图、频率响应和单位样值相应，编写下列代码：

 得到零极点图：

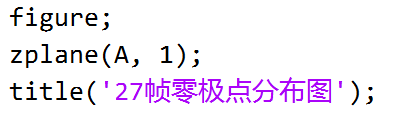
 频率响应：

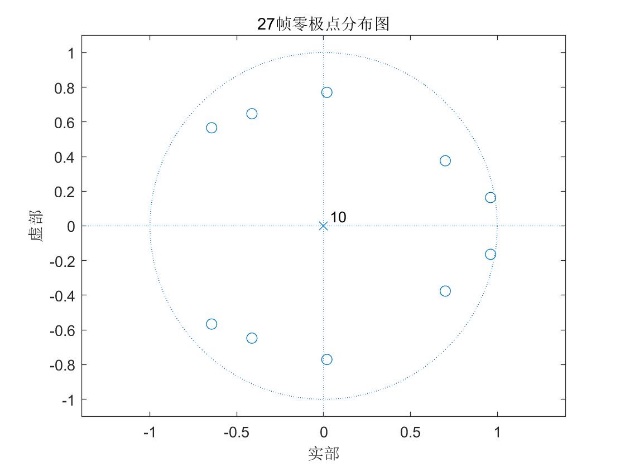
两种方法绘制出来的单位样值响应

可以看出这两种方法的结果是一样的。

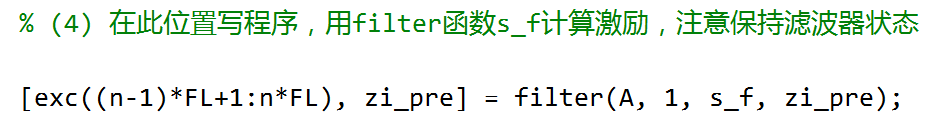
（2）阅读程序。略。

（3）观察预测系统的零极点图。

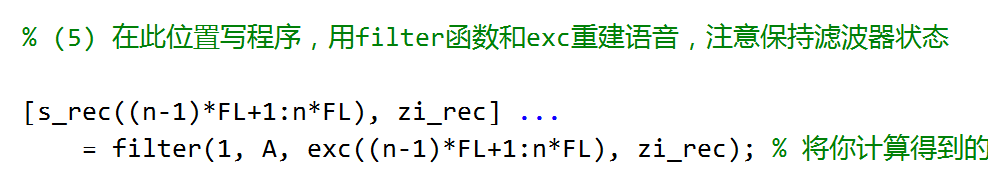
 编写代码如下：

 作出图象如下：

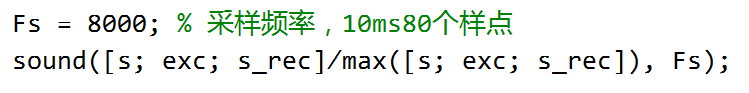
可以看到零点均关于实轴对称。

（4）

将zi\_pre作为初始状态送入并捕捉，可以满足保持滤波器状态的要求。

（5）

和第（4）问作类似的处理。

（6）

起初直接使用sound，发现声音嘈杂；而sound应该将大小控制在[-1,1]之间。

听取声音，发现s较为清晰；exc不甚动听，仿佛蒙在鼓里；s\_rec亦很清晰，与s别无而致。其原因为：s是原语音，理应清晰；s\_rec是重建的语音，也应当清晰；而exc是为了适应模型从而得到最佳结果的激励，因此会与最后结果听起来不太相仿。

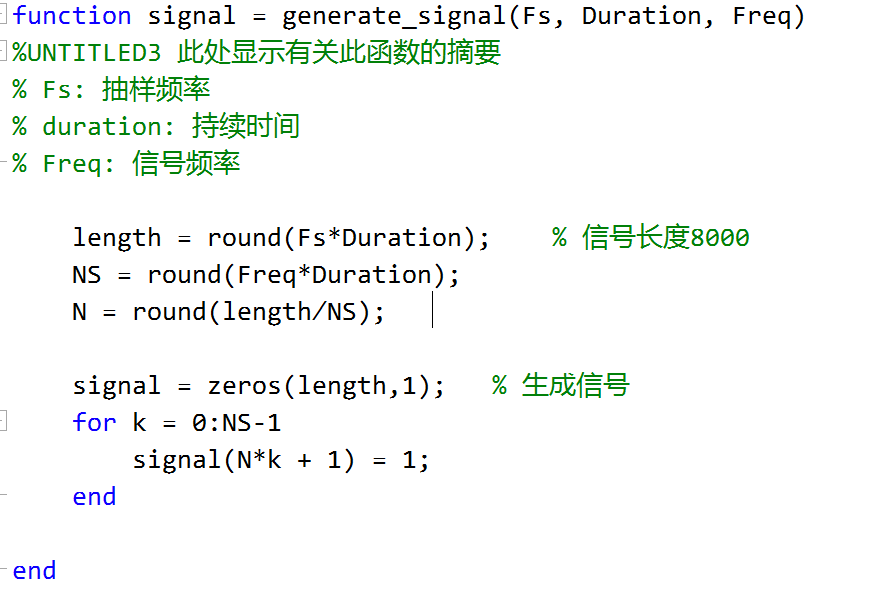
# 1.2.2 语音合成模型

（7）

对于该数字信号，持续时间为1s且抽样频率为8kHz，故该信号应有8000个抽样点。由于最后形成频率为200Hz的单位样值串，故该8000个抽样点中应有200个1，也即每40个点中有一个1。

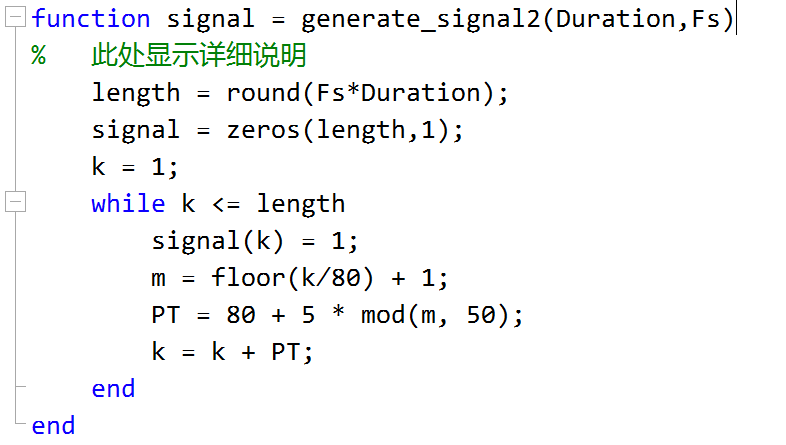
则对于式

其中的，。

 将生成信号的过程封装为函数，编写代码如下：

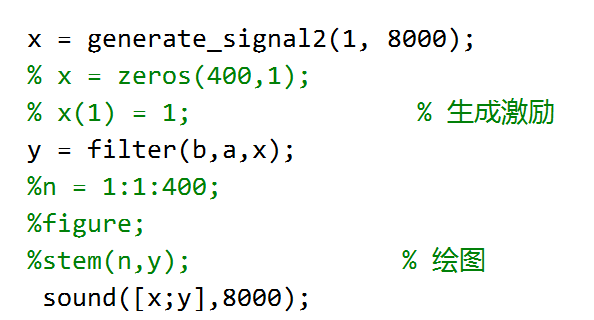
试听Freq=200和Freq=300的信号，可以发现一个声音较低，一个声音较高。

（8）

 与上一问相同，先生成具有8000个抽样点的信号。思路为：对于每一个有值的信号，考察其基音周期，在相应基音周期后的采样点继续考察。编写代码如下：

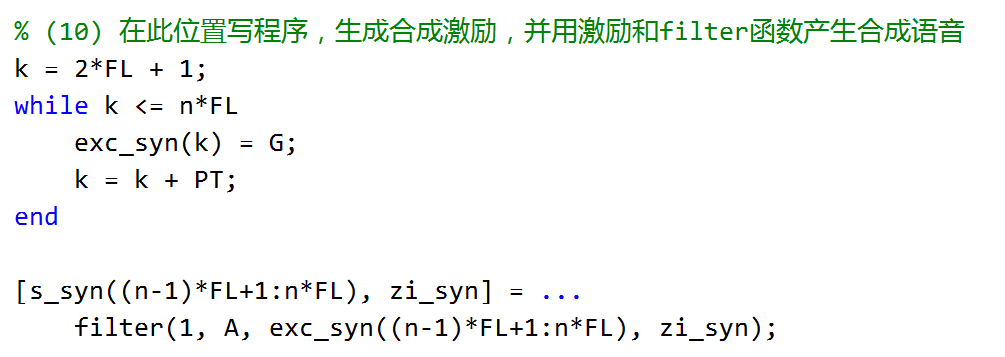
生成的声音效果和憋着嗓子挤出来的声音有点像。

（9）

 编写代码如下：

试听可以发现，经过滤波器的声音好像更自由开放一些，不像原始的声音有一种沉郁的紧绷感。

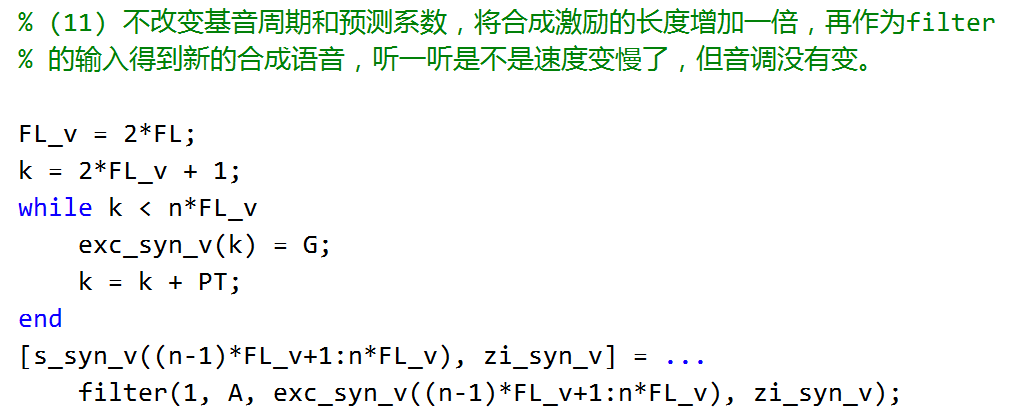
（10）

 编写代码如下：

试听合成语音，可以发现该语音和原语音比较类似，在实际生活中会被认为是同一个人，直观上会觉得只是音质的区别。

# 1.2.3 变速不变调

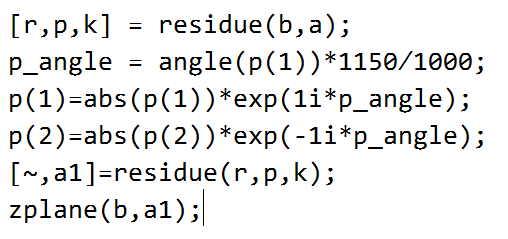
（11）

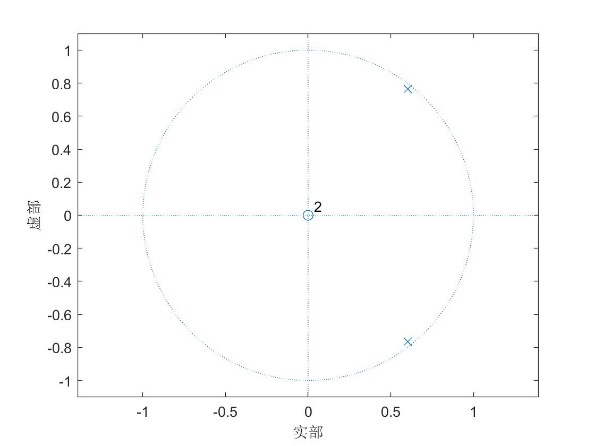
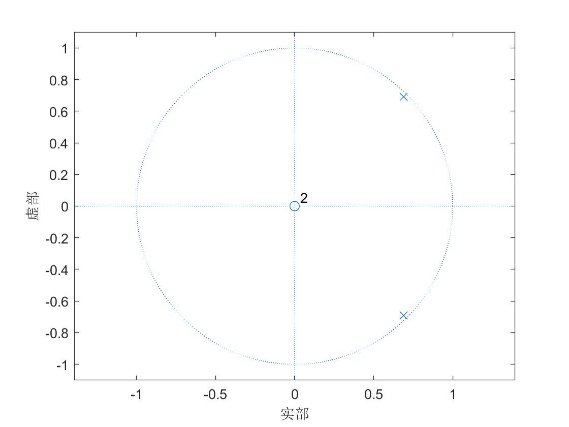
 将FL\*2.编写代码如下：

试听后，语音同预期一样，语速慢了一倍，但音调基本没有变化。

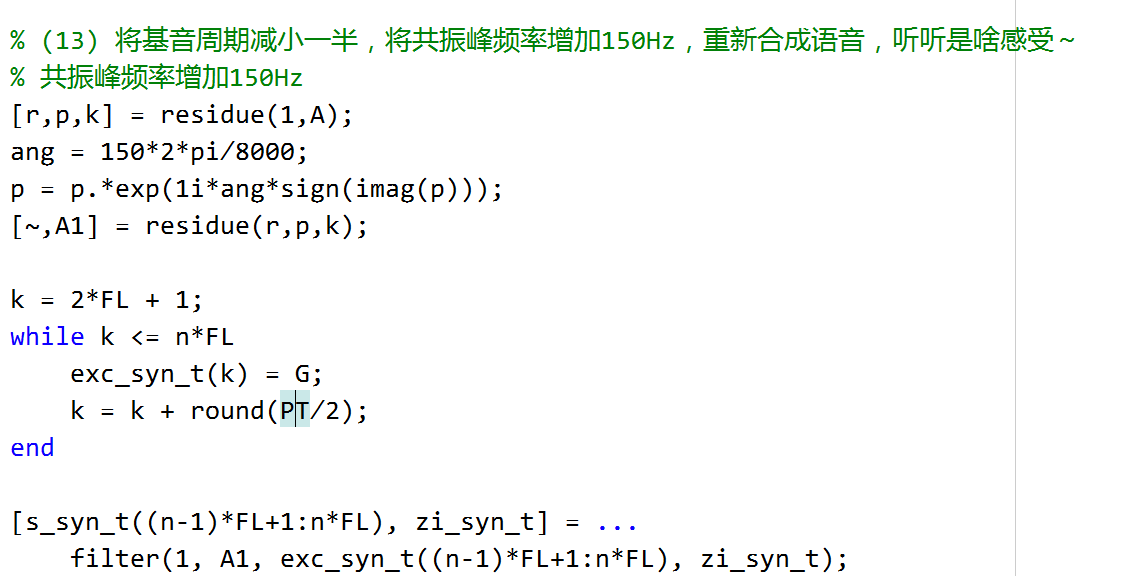
# 1.2.4 变调不变速

（12）

原系统共振峰频率为1kHz，增加150Hz后即为1150Hz。注意到，共振峰频率只取决于极点的幅角，则编写代码如下：

 对比前后零极点图的差别：

（13）

 与（12）做法类似，改变系统每个极点的幅角（注意其在上or下半平面），并在增量时变为PT/2。编写代码如下：

试听生成的声音，可以发现声音变高了，但速度没有变，好像一个男的在学女的说话。

# 文件清单

jpg文件：作业文档中需要作的图，在上述报告中出现过。

pcm文件：所需或生成的文件。

generate\_signal1.m:生成第7问所需信号

generate\_signal2.m:生成第8问所需信号

VoiceSynthesisAppend.m:除了主要代码，其余的代码部分

speechproc.m:主要代码文件