实验三：基于 PPG 的语音转换系统

(语音信号数字处理课程报告)

姓　　名：肖文韬

学　　号：2020214245

二〇二一年一月四日

目 录

目录.....I
插图清单.....II
第 1 章 任务一: 提取 PPG 与声学参数 (15").....1
1.1 任务介绍.....1
1.2 提取音素后验概率 PPG （4"）.....1
第 2 章 任务二: 训练并测试特定目标说话人的语音转换模型（40"）.....2
第 3 章 任务三: 探究残差网络对转换性能的影响（15"）.....3
第 4 章 任务四: 增加说话人嵌入网络，实现多目标说话人的语音转换（20"）.. 4
参考文献.....5

插图清单

图 1.1　PPG 提取流程图 …… 1

第 1 章　任务一: 提取 PPG 与声学参数 (15")

1.1　任务介绍

为了进行语音转换，我们首先需要使用 ASR 系统将源音频转换为一种中间特征（在本实验中就是音素序列 PPG[1]），对每一帧的 MFCC 特征 X_t 我们可以得到所有音素（音素集 S）的后验概率 $\{p(s|X_t)|s \in S\}$。同时，我们还可以将原始波形序列加窗得到语音帧，对语音帧进行离散傅里叶变换后，计算各频率分量的能量后可以得到语谱图（线性谱）。而我们知道人类对低频成分更加敏感，而对高频不敏感，所以我们取对数后可以得到对应的 Mel 谱。本任务就是使用预训练模型得到音频的 PPG，同时还需要计算得到基频 F_0，线性谱，Mel 谱等声学参数。

接下来的小节就是回答问题啦。

1.2　提取音素后验概率 PPG （4"）

(1) 简要说明 PPG 提取器 (ppg_extractor) 的网络结构，给出网络的基本结构图。

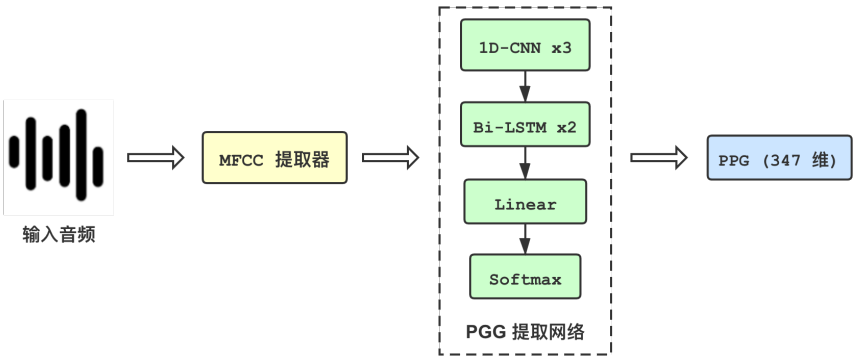

图 1.1　PPG 提取流程图

答: PPG 提取器网络由卷积层、LSTM 和线性层组成，具体组成如图所示。

第 2 章　任务二: 训练并测试特定目标说话人的语音转换模型（40"）

第 3 章　任务三: 探究残差网络对转换性能的影响（15"）

第 4 章　任务四: 增加说话人嵌入网络，实现多目标说话人的语音转换（20"）

参考文献

[1] Sun L, Li K, Wang H, et al. Phonetic posteriorgrams for many-to-one voice conversion without parallel data training[C/OL]// 2016 IEEE International Conference on Multimedia and Expo (ICME). 2016: 1-6. DOI: 10.1109/ICME.2016.7552917.