一种新型英语口语对比打分系统

一．背景

随着计算机技术的发展，越来越多的学习软件可以帮助人们更方便地学习外语。目前绝大多数计算机辅助外语学习软件主要关注文字应用能力和语言理解能力的训练，却很少关注口语发音能力训练。应用语音处理技术，可以实现英语学习中的口语发音自动打分。

当前主流的英语口语打分系统分为整体打分系统和对比打分系统两种。整体打分系统不提供标准发音，直接测试发音人的发音标准程度，因而依赖一个背景标准发音模型；对比打分系统提供标准发音，发音人跟读标准发音，系统评价发音人发音与标准发音的相似程度。

对比打分系统一般包括模板收集，语音信号预处理与特征提取，基于模板的信号对齐等三个模块。对齐方法一般采用动态规划算法，如 DTW。近来通常采用基于 HMM-MAP 的统计模型方法对模板建立简单的 HMM 模型，再通过计算测试语音在该模型上的概率值得到对比分数。本发明提出一种基于线性模型和神经网络的新型对比打分系统。

二．发明要点

本发明提出一个基于线性模型的多特性（节奏，音调，音色）对比打分系统，包括一个基于深度神经网络(DNN)的局部特征提取模块，一个基于线性预测模型的全局特征提取模，一个基于多层感知器（MLP）的打分模块。

1. 基于线性预测模型的多特性对比打分系统框架

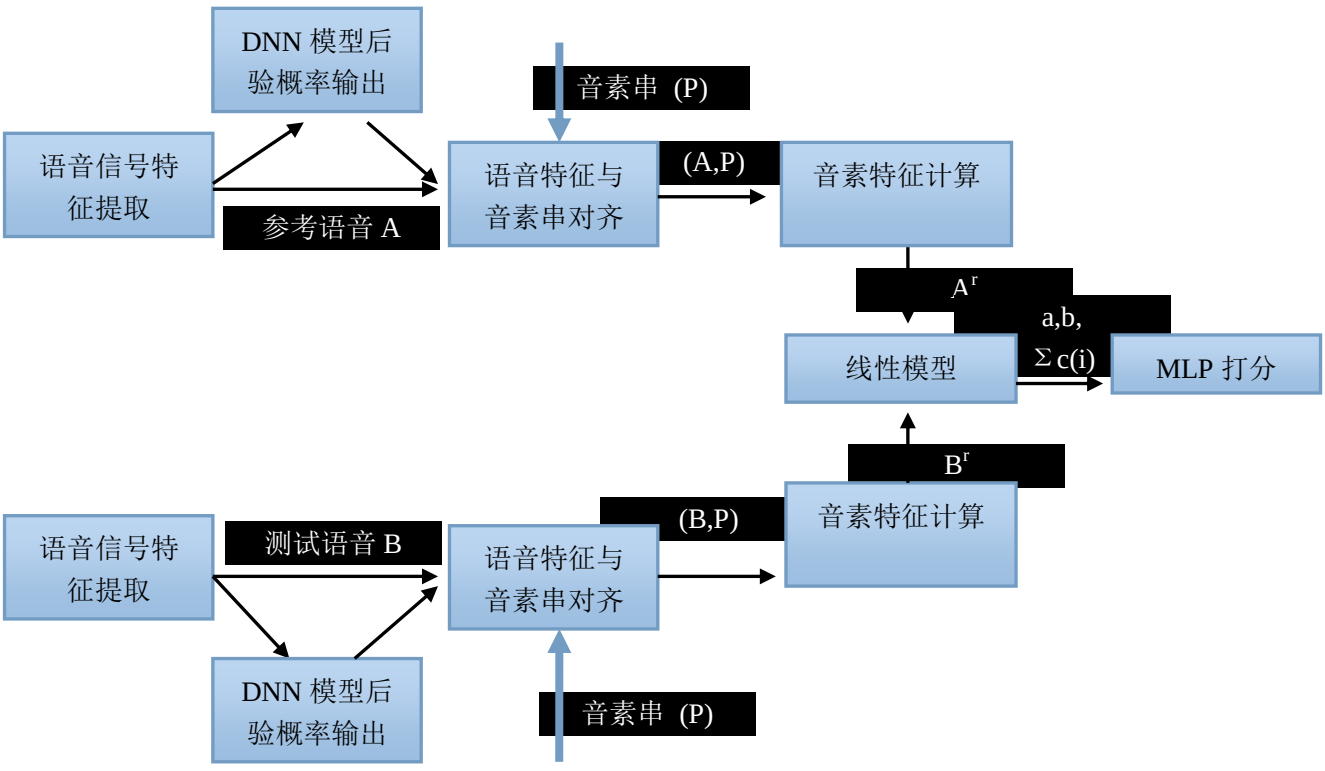

图 1: 基于线性模型的口语对比打分系统流程图

图(1)给出该方法的流程图。我们假设参考语音 A 和测试语音 B，打分的标准是计算语音 B 对语音 A 的相似性。假设语音 A 和语音 B 对应的音素串 P 一致并已知。首先用训练好的 DNN 模型产生帧后验概率作为语音的原始特征。根据音素串 P，依最大后验概率准则将语音帧与音素串对齐。将语音 A 与 P 对齐，得到对齐结果 L(A,P); 将语音 B 与 P 对齐，得到对齐结果 L(B,P)。依 L(A,P)和 L(B,P)计算语音 A 和语音 B 在某一语音特性 r 上的特征向量，记为 $A^r=[A^r(1), A^r(2), \ldots, A^r(N)]$和 $B^r=[B^r(1), B^r(2), \ldots, B^r(N)]$，其中 N 为音素串 P 中的音素个数。$A^r$ 和 B^r 描述语音中各音素段内的特征，称为**局部特征**。依 A^r 和 B^r 进行线性模型建模，得到模型参数，将该模型参数作为对比语音 A 和 B 的**区分性全局特征**，并作为 MLP 模型的输入，输出即为 A 和 B 关于特性 r 的区分性对比打分结果。

2. 基于线性预测模型的区分性全局特征提取

设句子 A 和 B 关于特性 r 的局部特征向量为 A^r 和 B^r，可建立线性预测模型：

$$A^r(i)=a\,B^r(i) + b + c(i)$$

其中 a,b 分别代表线性模型的比例量和偏移量，$\Sigma c(i)$对应 A^r 和 B^r 之间的相关系数。这三者共同描述了 A^r 和 B^r 的相似程度，作为对比语音 A 和 B 的区分性全局特征。

3. 基于 MLP 的区分性打分

将这区分性全局特征(a,b, $\Sigma c(i)$) 作为特征输入到 MLP 模型中，即可得到对应于特性 r 的区分性对比打分结果。如图 2 所示。

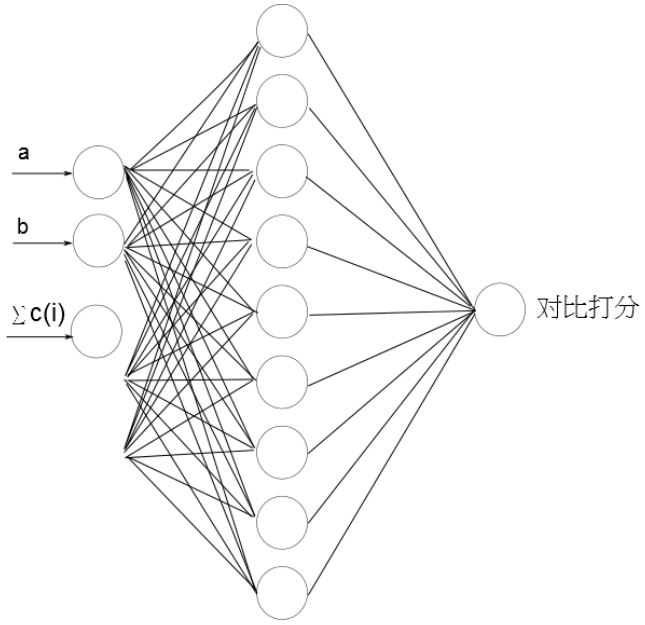

图 2: MLP 对比打分

4. 多特性打分

基于上述线性预测模型和 MLP 打分模型，本发明对测试语音与参考语音在节奏、音调、音色三个特性上进行打分，分别叙述如下。

节奏打分

节奏描述发音过程中的发音长度变化，即每个音素发音长短的相关性。我们采用线性模型来描述两句话之间的节奏关系。在前述线性模型中，我们计算 $A^r(i)$ 与 $B^r(i)$分别为 A 和 B 中第 i 个音素的时长，这可以通过音素对齐(A,P)与(B,P)直接得到。

音调打分

音调描述发音的基频随时间变化的规律。我们采用同节奏打分相似的方法，通过音素对齐(A,P)与(B,P)得到每个音素的基频 $A^r(i)$与 $B^r(i)$，求解线性模型，得到模型参数，送入 MLP 得到音调打分。

音色打分

音色是描述说话人发音特质的变量。音色具有模糊性，广义的音色包括节奏和音调，狭义的音色与声音中共振峰的分布相关性更强。所以，我们通过计算共振峰的相性得到音色打分。具体而言，通过音素对齐(A,P)和(B,P)，计算每一音素的前 10 个共振峰频率位置作为 $A^r(i)$与 $B^r(i)$，送入线性模型。注意这里的 $A^r(i)$与 $B^r(i)$为向量，因此该模型是多元性线模型。求解该模型得到模型参数，送入 MLP 得到音色打分。

三. 方案优势

1. DNN 模型是区分性模型，因此基于 DNN 的原始特征提取方法相对 MFCC 传统特征可以有效区分语音音素和噪音，因而对噪音具有较强的鲁棒性。
2. 基于线性预测模型的区分性全局特征提取方法可以有效描述对比语音段的静态相似性和动态相似性，简单高效，抗干扰能力强。
3. 基于 MLP 模型的区分性打分方法，以对比语音的相似性作为模型训练目标，因此打分具有更强的区分性。