

# 基于河南方音的英语语音自动分析

李 楠

(河南师范大学 外国语学院,河南 新乡 453007)

**摘 要:**首先,建立基于方音的二语发音数据库,并分析发音数据的音韵特征;然后,基于隐马尔可夫理论,提出了一种识别特定方音背景的外语学习者的语音产出的可行性方案.性能试验显示,系统可以较为准确地自动识别出以河南方音为背景的学习者的外语语音产出,并形成诊断性评价,进而提升其外语语音水平.

**关键词:**基于方音的非本族语语音识别;隐马尔可夫理论;二语语音习得

**中图分类号:**TP271

**文献标志码:**A

方音背景是自动语音识别过程中造成语音采集者内部差异的重要因素之一<sup>[1]</sup>.二语语音产出的调查结果表明,母语方音较重的外语使用者所犯的发音错误相对较多<sup>[2]</sup>.这些具有方言地域性质的错误为外语语音教学在不同区域的开展提供了可靠的、有效的突破点.河南方言区方音情况较为复杂,这成为河南英语语音学习不能得到迅速提高的一大阻碍.

## 1 相关研究

自 20 世纪 50 年代 Davis 等人在 Bell 实验室实现了首个真正意义上的语音识别系统以来,语音识别研究领域又先后引进了线性预测技术(LPC)和隐马尔可夫法(HMM),使得语音识别的应用研究有了突破性进展,越来越多的应用系统问世,如 IBM 公司开发的 ViaVoice, SAPI 开发的 Microsoft 和 NaturallySpeaking of Dragon, 美国 CarnegieMellon 大学开发的 FLUENCY, 法国 Auralog 公司开发的 Auralog's Auralang, Tell Me More 和 Talk To Me 系列软件等等.国内也有《金山单词 2003》、《句霸 9000》、《新东方语音教练》等.大多国外语音平台着重语调、重音和节奏的训练.它们的系统语音库包括英语典型语调、强度和节奏模式,快速分析语调、强度和节奏并转换成可视图形,与标准发音对比.国内的语音软件对非连续语流中的音段音位的判断相对准确,而对于超音段音位层面的重读、节奏以及语调的判定较为模糊.

隐马尔可夫模型的引入使语音识别实现了由孤立字或单个词向连接字或连续语流的突破,由对特定人的依赖性强到非特定人的普适性的过渡,由小词汇量到大词汇量的飞跃.尽管语音识别的应用领域得以拓展,但依旧局限在话语产出形式有限、内容固定的一些领域<sup>[3-4]</sup>,如医院、机场自助语音服务平台等.系统的适应性能有待于进一步提高,以做到不受特定人、口音或者方言的影响<sup>[5]</sup>.目前,基于方音的自动语音识别的应用研究较少<sup>[6]</sup>,非本族语使用者的方音识别研究更少<sup>[7]</sup>.本文采用定量分析的方法对河南方音背景的外语发音数据进行分析、特征归类,制定出依赖于特定方音的外语发音词典,并建立相应的发音模型.基于 HMM 模型,提出了建立一种有河南方音的外语语音自动识别应用系统的可行性方案.

## 2 基于河南方音的英语语音识别和自动分析平台设计

### 2.1 系统结构设计

收稿日期:2014-09-04;修回日期:2014-12-10.

基金项目:教育部人文社科项目(12YJC740062);河南省哲学社科项目(2013CYY025).

作者简介:李楠(1980-),女,河南安阳人,河南师范大学讲师,研究方向为应用语言学、语料库语言学, E-mail:georgiallee1980@163.com.

基于方音的非本族语语音自动分析和识别技术,可以实现对学习者的语音输入自动识别、归类,并根据分析结果给出反馈和评价。系统构架的工作流程见图1。输入二语语音数据之后,系统进行参数分析,首先指向以河南方音为背景的二语发音数据库进行数据类型比对、匹配,再指向标准语音库进行测度估计。之后,参照词典及文法信息给出判定。输出语音识别结果后,再次将之与标准语音库中的对应语音项目内容进行测距分析,从而形成诊断性评价。系统还可以根据评价结果为学习者量身定制操练,以最大程度地提高学习者的二语语音习得效率。对于未能识别的语音数据进行人工判定,提取错误特征,将之添加到以河南方音为背景的英语学习者的发音数据库中去。在此基础上为不同方言背景和发音困难的学习者定制练习,从而增强语音学习的交互性和灵活性,提高学生参与学习的主动性和积极性,使学生更加容易和牢固地掌握语音知识。它的应用方向是以河南方音为背景的英语学习者,主要用来检测英语学习者二语产出中的错误,尤其是方音所造成的二语发音错误。系统可以识别出学习者的二语发音错误,判定错误类别,给出评价和反馈,从而指导整个语音学习过程。

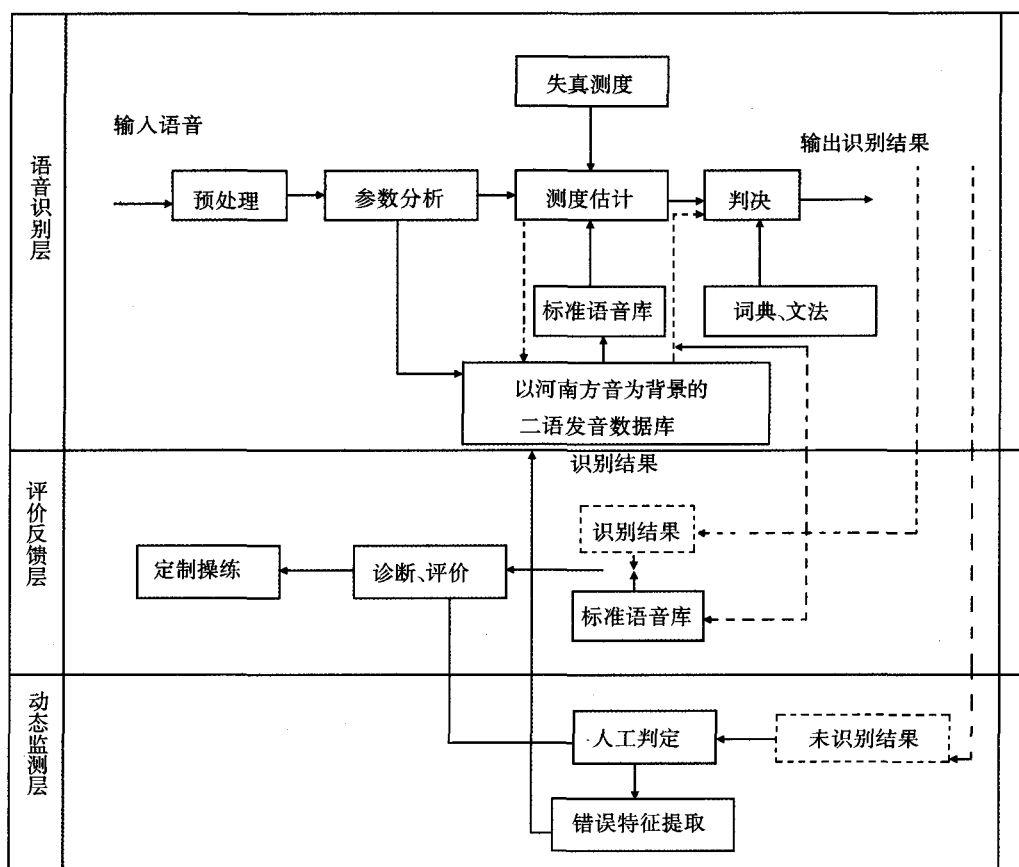


图1 基于河南方音的英语语音识别和自动分析平台技术构架图

## 2.2 基于方音的二语发音数据库的建立

本研究着眼于河南外语语音教学现状,通过依赖于方音的外语语音自动识别技术的应用提高河南外语语音教学的有效性。在建立方音数据库时,鉴于河南方言情况较为复杂,方言内部差异性较大的特点,选取方言发音特点比较突出、英语发音困难较大的典型次方言区的学习者,采集他们的发音数据。首批选取来自信阳、济源、安阳等典型次方言区的某大学一年级刚入学的本科生共计300名(男女各半)。先导试验之后从以上次方言区中各选取发音困难较大的共100名受试进行跟踪观察。数据采集过后,进行转写、归类、附码、分析,以发现数据整体特点走向,从而得出发音困难的模式,形成凸显发音困难的语音数据库,以方便与标准英语参照数据库进行比照。

先导实验显示,音段音位层面的单音发音错误是受方音影响的外语发音错误中最为典型的错误类型之

—<sup>[8-9]</sup>. 结合数据特点,进行音韵特征归类,发音错误所涉及的特征如表 1 所示,前 6 项为辅音发音错误特征,随后 3 项为元音发音错误特征,最后 3 项为元音、辅音共同的发音错误特征.

表 1 方音数据库中发音错误的特征类型及赋值

音韵及其他特征	基数	特征值
发音方式	8	空、无、爆破音、擦音、破擦音、鼻音、边音、近似音
发音部位	9	空、无、双唇音、齿音、唇齿音、齿龈音、齿龈后音、硬腭音、软腭音
清浊	3	空、清、浊
送气	3	空、是、否
辅音特征	4	空、无、辅音、辅音加音
音段	3	空、有、无
舌前	8	空、无、前、中、后、前一中、后一中、中一前
舌高	8	空、无、高、中、低、中一高、低一高、高一中
音长	3	空、长、短
圆唇	5	空、圆、扁、圆一扁、扁一圆
卷舌	3	空、卷、展

基于方音的二语发音数据库不仅包含了方音发音错误类型及其特征<sup>[10-12]</sup>,还包含了发音错误出现的位置、条件、音节环境等信息.另外,我们还计算了每种错误特征值在同类错误中所占的比例等权重信息,以便于在语音识别中对错误权重、排序,增加识别和错误判定的速率.实验也表明,有些发音错误是共有的,有些是某方音特有的,如信阳学生出现的/l/和/n/替换错误.

### 2.3 隐马尔科夫算法

单词  $W$  的发音涉及呈线性排列的许多音素的产出.就音韵特征而言,发音包括一系列音韵特征状态的过渡.第  $k$  个音韵特征的第  $j$  个状态标记为  $S_{k_j}$ ,音韵特征状态  $S_{k_j}$  持续  $O_{k_j}$  帧.那么,发音过程就可以表示为以下成对序列,

$$\{(S_{k_1}, O_{k_1}), (S_{k_2}, O_{k_2}), \dots, (S_{k_N}, O_{k_N})\}. \quad (1)$$

整个发音过程是  $N$  个状态的延续.状态持续对列的线性排列过程为一个马尔可夫过程,见图 2.

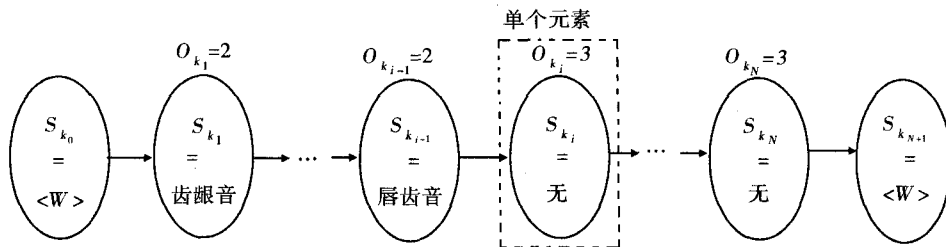


图2 带有方音的二语发音的马尔可夫过程图

示例单词:nevermore.

示例音韵特征:发音部位.

分别用模型状态和状态过渡表示发音器官移动的持续性和特点.为了模拟开始和结束马尔可夫过程,我们将有序的对列  $(S_{k_0}, 0)$  和  $(S_{k_{N+1}}, 0)$  界定为开始状态和结束状态.正如图 2 所示,开始和结束状态的值恒为  $\langle w \rangle$  和  $\langle /w \rangle$ .并且,开始和结束状态持续的时间均为 0.加入开始状态和结束状态,状态持续序列就可拓展为:

$$\{(S_{k_0}, 0), (S_{k_1}, O_{k_1}), (S_{k_2}, O_{k_2})\}, \dots, (S_{k_N}, O_{k_N}), (S_{k_{N+1}}, 0)\}, \quad (2)$$

第  $k$  个音韵特征状态和状态持续的联合概率即为:

$$\Lambda(W_k) = p_w(O_{k_0}, \dots, O_{k_{N+1}}, S_{k_0}, \dots, S_{k_N}, S_{k_{N+1}}) = p_w(O_k, S_k), \quad (3)$$

$W_k$  是单词  $W$  的第  $k$  个音韵特征序列,并且

$$O_k = (O_{k_0}, O_{k_1}, \dots, O_{k_N}, O_{k_{N+1}}), \quad (4)$$

和

$$S_k = (S_{k_0}, S_{k_1}, \dots, S_{k_N}, S_{k_{N+1}}) \quad (5)$$

代表一系列的状态持续和状态占有. 使用条件概率, 公式(3) 可以写为,

$$\Lambda(W_k) = p_w(O_k | S_k) p_w(S_k). \quad (6)$$

假定各个状态之间是相互独立的, 并且只依赖于他们各自的状态本身, 那么公式(6) 的第一部分可以写为:

$$p_w(O_k | S_k) = \prod_{i=0}^{N+1} p_w(O_{k_i} | S_{k_i}), \quad (7)$$

$p_w(O_{k_i} | S_{k_i})$  是单词  $W$  在状态为  $S_{k_i}$  的条件下, 状态持续为  $O_{k_i}$  帧的概率. 这里需要注意的是,  $p_w(O_{k_0} | S_{k_0}) = p_w(O_{k_{N+1}} | S_{k_{N+1}}) = 1$ . 如果假定状态过渡为一个马尔可夫过程, 也就是说状态过渡到  $S_{k_i}$  依赖于前一个状态  $S_{k_{i-1}}$  本身, 那么公式(6) 的第 2 部分可以写为:

$$p_w(S_k) = \prod_{i=1}^{N+1} p_w(S_{k_i} | S_{k_{i-1}}). \quad (8)$$

根据马尔可夫过程假设,  $p_w(S_{k_i} | S_{k_{i-1}})$  为过渡概率. 结合公式(7) 和公式(8), 公式(3) 中的联合概率可写为:

$$\Lambda(W_k) = \prod_{i=1}^{N+1} p_w(O_{k_i} | S_{k_i}) p_w(S_{k_i} | S_{k_{i-1}}). \quad (9)$$

换言之, 出现第  $k$  个音韵特征状态和状态持续的对数似然比可以写为:

$$\lg(\Lambda(W_k)) = \sum_{i=1}^{N+1} \lg(p_w(O_{k_i} | S_{k_i}) p_w(S_{k_i} | S_{k_{i-1}})). \quad (10)$$

假定  $p_w(S_{k_i} | S_{k_{i-1}})$  分布为  $\Gamma$  分布, 即为,

$$p_w(O_{k_i} | S_{k_i}) = (O_{k_i})^{K_{uS_{k_i}} - 1} \frac{\exp\left(-\frac{O_{k_i}}{\theta_{uS_{k_i}}}\right)}{(\theta_{uS_{k_i}})^{K_{uS_{k_i}}} \Gamma(K_{uS_{k_i}})}, \quad (11)$$

其中,  $K_{uS_{k_i}}$  和  $\theta_{uS_{k_i}}$  是  $S_{k_i}$  的形状和规模参数, 反映的分别是音韵特征状态和单词  $W$ . 在带有方音的二语发音模型中, 我们使用直接由数据中计算得来的分布参数的最大概率估算. 同样, 过渡概率  $p_w(S_{k_i} | S_{k_{i-1}})$  的最大概率估算也是直接由数据决定的. 所以, 对每一个词  $W$  和音韵特征状态, 状态持续密度  $p_w(O_{k_i} | S_{k_i})$  和状态过渡密度  $p_w(S_{k_i} | S_{k_{i-1}})$  的数量为  $V$  和  $V^2$ . 其中,  $V$  是第  $k$  个音韵特征状态可取的值的个数. 音韵特征状态可取的值的集合见表 1. 上述推算就形成了我们提出的带有方音的二语发音模型. 为了直接对比不同方音的发音特点, 我们参照英语和汉语发音特征, 基于马尔可夫过程, 继而提出了差异模型.

对英语本族语使用者 ( $L_1$ ) 和母语为汉语的二语发音群体 ( $L_2$ ), 单词  $W$  的基于马尔可夫过程的方音模型分别用  $\Lambda_{L_1}(W_k)$  和  $\Lambda_{L_2}(W_k)$  来表示. 使用公式(10),  $\Lambda_{L_1}(W_k)$  和  $\Lambda_{L_2}(W_k)$  可以写为:

$$\lg(\Lambda_{L_1}(W_k)) = \sum_{i=0}^{N+1} \lg(p_w^{L_1}(O_{k_i} | S_{k_i}) p_w^{L_1}(S_{k_i} | S_{k_{i-1}})) = \sum_{i=1}^{N+1} \lg(\Lambda_{L_1}, O_{k_i}). \quad (12)$$

和

$$\lg(\Lambda_{L_2}(W_k)) = \sum_{i=0}^{N+1} \lg(p_w^{L_2}(O_{k_i} | S_{k_i}) p_w^{L_2}(S_{k_i} | S_{k_{i-1}})) = \sum_{i=1}^{N+1} \lg(\Lambda_{L_2}, O_{k_i}). \quad (13)$$

为了建立差异模型, 要看图 2 所示的马尔可夫链的第  $i$  个单个元素. 从公式(12) 和(13) 可知, 第  $i$  个单个元素对整体  $L_1$  和  $L_2$  对数似然比的贡献表示为  $\lg(\Lambda_{L_1}(S_{k_i}, O_{k_i}))$  和  $\lg(\Lambda_{L_2}(S_{k_i}, O_{k_i}))$ . 对于一个单个元素  $(S_{k_i}, O_{k_i})$ , 将归一化的  $\delta$  对数似然比界定为:

$$\Delta_\delta(S_{k_i}, O_{k_i}) = \frac{\lg(\Lambda_{L_1}(S_{k_i}, O_{k_i}) + 1) - \lg(\Lambda_{L_2}(S_{k_i}, O_{k_i}) + 1)}{\lg(\Lambda_{L_1}(S_{k_i}, O_{k_i}) + 1) + \lg(\Lambda_{L_2}(S_{k_i}, O_{k_i}) + 1)}, \quad (14)$$

其中,  $-1 \leq \Delta_\delta(S_i, O_i) \leq +1$ . 从公式(14) 可以看出,  $\Delta_\delta(S_i, O_i) \rightarrow -1$  显示发音倾向于母语, 即二语发音中母语方音较浓厚. 而  $\Delta_\delta(S_i, O_i) \rightarrow +1$  则更接近于本族语使用者的水平. 那么, 整个单词  $W$  的  $\delta$  似然比值就为单个状态持续  $\delta$  似然比的平均值,

$$\Delta_\delta(W_k) = \frac{1}{N} \sum_{i=1}^N \Delta_\delta(S_{k_i}, O_{k_i}), \quad (15)$$

其中,  $N$  为状态的总数. 显然,  $-1 \leq \Delta_\delta(W_k) \leq +1$ . 由于  $\Delta_\delta(W_k)$  是被约束的, 就可以使用它对不同说话者和不同的单词进行对比. 特别是, 通过这种方式定位了一个词, 也就可以将不同方音群体的不同个体的发音在  $[-1, +1]$  的维度内进行排序. 排序的标准就是方音的浓厚程度. 值得注意的是, 公式(15) 中的  $\Delta_\delta(W_k)$  表示的是单词  $W$  的第  $k$  个音韵特征类型的方音分数.  $\Delta_\delta(W_k)$  的方音总值可以由求  $\Delta_\delta(W_k)$  所有音韵特征类型的平均值计算得出,

$$\Delta_{\delta}(W_k) = \frac{1}{M} \sum_{k=1}^M \Delta_{\delta}(W_k), \quad (16)$$

其中,  $M$  是所有音韵特征类型的总数.

上述基于马尔可夫过程的方音二语识别模型可以检测不同发音群体之间的主要发音差异,也可以用来检测发音个体的特定的发音特征. 故该模型可提高二语习得效率并消除发音中的方音影响.

### 3 性能实验

性能实验选来自信阳和安阳方言区的 20 名外语学习者,进行了为期一年的跟踪调查. 期间,受试持续使用基于方音的二语发音学习系统. 在前测和后测中,要求受试者朗读涵盖所有英语音素及各种音节环境的单词和段落.

#### 3.1 人工评分与系统评分结果

在实验中,要求 3 位本族语使用者对上述 20 位受试的发音评分. 评分者给出的分值范围在 0 到 100 之间,反映方音浓厚程度. 0 表示方音较为浓厚,100 表示无方音影响,达到本族语使用者水平.

图 3 显示了人工打分的方音分值和使用基于方音的二语语音学习系统的时间长短之间的关系. 如图 3 所示,在前测和后测中人工评分方音平均值分别为 53 和 62. 数据显示,随着使用系统时间的增长,受试二语发音中的方音影响变小. 受试方音总分值和使用系统时间长短之间的相关性为 0.2.

在系统自动方音评分中,系统使用公式(16)计算出每一位受试的所有单词的所有音韵特征的归一  $\Delta$  似然比值的平均值,最终结果就是系统给出的每位受试二语发音中的方音分值. 在图 4 中,我们也分别计算这 20 名受试在前测和后测中的平均  $\Delta_{\delta}$  值. 系统评分结果显示,随着使用系统时间的增长受试的二语发音水平也随之增长,二者之间的相关性为 0.38 ( $p < 0.5$ ).

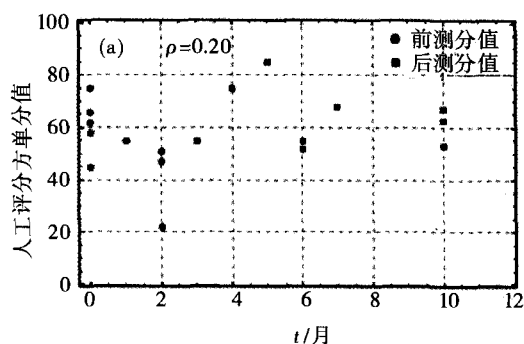


图3 人工评分方音分值和使用系统时间的关系

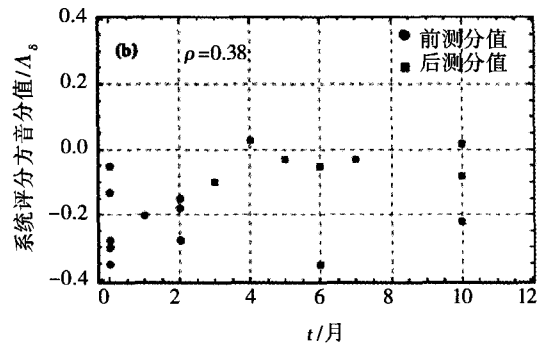


图4 系统评分方音分值和使用系统时间的关系

最后,我们计算了前测和后测中系统自动评分分值和人工评分分值之间的相关性,二者之间的相关性为 0.7 ( $p < 0.0001$ ). 结果显示,系统自动评分分值和人工评分分值之间的相关性较高,数据内部一致性较强. 此外,图 5 还显示了系统评分分值和人工评分分值的散点分布. 数据点显示的为受试在前测和后测中的方音分值. 结果显示系统评分分值和人工评分分值之间存在一定的线性关系.

#### 3.2 状态持续—过渡差异分析

在上述方音模型中,我们区别了发音过程的状态过渡和状态持续,并给出了他们的音韵特征类型和不同的赋值. 下面,我们分析受试发音过程中的状态过渡和状态持续的方音分值. 使用公式(15),分别计算了状态过渡和状态持续的每一个音韵特征的方音分值. 我们主要聚焦于发音方式、发音部位、舌前、舌高 4 个音韵特征的系统评分分值. 这是因为,发音方式和发音部位是定位一个辅音最重要的两个音韵特征,而舌前和舌高是定位一个元音最重要的两个音韵特征<sup>[13]</sup>. 此外,我们还观察了上述 4 个音韵特征的方音分值和使用系统时间长短之间的关系.

通过对比发音过程中的状态持续和状态过渡的相关系数,发现状态持续发音水平较状态过渡发音水平提高快. 总的来讲,状态持续的方音分值和使用系统时间之间的相关系数(0.35)要高于状态过渡的方音分值

和使用系统时间之间的相关系数(0.11). 在辅音方音分析中,发音部位这个音韵特征的状态持续与使用系统时间的相关系数(0.36),明显要高于状态过渡与使用系统时间的相关系数(0.03). 发音方式的状态持续水平较状态过渡水平也提高较快(状态持续系数为 0.19,状态过渡系数为 0.28).

舌高指的是舌头抬起的高度,决定了开口度的大小. 分析结构显示,舌高是随着使用系统时间的增长,在状态持续和状态过渡这两个发音过程中提高最大的音韵特征(状态过渡相关系数为 0.62;状态持续相关系数为 0.37). 相对而言,舌前(即舌头抬起最高点的位置)这个音韵特征提高较缓慢.

### 3.3 目标发音—实际发音对分析

下面,通过分析每个单词的音韵特征,可以发现来自不同方言区的受试在哪些单词或哪些音素上发音不同. 每个词都可看作是一个发音任务. 每个发音任务都要求说话者经历一系列的规范的发音状态. 这一系列的规范受到必要的状态持续和状态过渡的约束. 我们将规范发音称作目标发音,而受试的发音称作实际发音. 实际发音和目标发音之间的距离越大,发音过程所涉及的状态持续和状态过渡的音韵特征对学习来说就越困难. 使用公式(16),可以对目标发音和实际发音进行测距分析. 首先,计算所有目标和实际发音对的归一  $\Delta$  对数似然比值,并计算每一对的平均值. 那么,  $\Delta_s$  值最大就表明目标发音的难度最大.  $\Delta_s$  值大表明目标发音和实际发音的不匹配程度强. 此外,还需要测量不匹配现象的出现频数,实际发音和目标发音对列出现的频数,也就是某目标发音被发成某一实际发音的次数.

表 2 以音韵特征发音部位、发音方式、舌前、舌高为维度,显示了受试感到最难的发音任务. 此外,我们还给出了发生错误的具体单词、目标发音和实际发音中所涉及的音韵特征的赋值情况,计算了前测和后测中所涉及的实际发音和目标发音对列的  $\Delta_s$  平均值,以及每一个实际发音和目标发音对列出现的比例. 计算方法为

$$D = \frac{F_E}{E_T} \cdot F_E$$

$F_E$  表示试图发某一目标发音时发成某一实际发音的原始频数,  $E_T$  表示某一目标发音发错的总频数.

表 2 基于方言的二语发音错误的音韵特征类型及比例

音韵特征	目标发音	实际发音	单词	前测 $\Delta_s$ 平均值	后测 $\Delta_s$ 平均值	比例/%
发音方式	浊齿龈鼻音	浊齿龈边音	night, knife	-0.231	-0.125	62.33
	浊齿龈边音	浊齿龈鼻音	light, line	-0.189	-0.111	70.36
	浊腭近似音	浊齿龈边音	red, right	-0.128	-0.087	74.31
	浊齿龈边音	浊腭近似音	glove, blow	-0.122	-0.092	70.36
发音部位	浊齿擦音	浊齿龈擦音	they, bathe	-0.325	-0.146	98.26
	浊唇齿擦音	浊双唇近似音	five, thrive	-0.311	-0.167	88.73
	清齿擦音	清齿龈擦音	thank, bath	-0.298	-0.225	53.31
	清后齿龈擦音	清腭擦音	sheep, she	-0.223	-0.095	28.57
	清齿龈擦音	清后齿龈擦音	bus, see	-0.189	-0.092	42.31
	清后齿龈擦音	清齿龈擦音	she, fish	-0.097	-0.043	40.68
舌前	中低扁唇短音	后低扁唇长音	cut, but	-0.096	-0.038	28.89
	中中扁唇短音	前高扁唇短音	separate	-0.088	-0.041	24.53
舌高	前高扁唇短音	前中—高扁唇短音	feet, seat	-0.098	-0.025	66.27
	中中扁唇短音	前高扁唇短音	separate	-0.067	-0.023	24.53

如表 2 所示,就发音方式这个音韵特征而言,对河南方言区学习者难度最大的为鼻音、边音和近似音. 这几类错误主要出现在词首,尤其是辅音连缀中. 其中,鼻音和边音替换占错误总数的 43%. 而所有的错误都涉及边音. 就发音部位而言,齿龈音和后齿龈音替换占错误总数的 38%. 而大部分错误都是发音部位后移造成的. 从前测和后测的  $\Delta_s$  平均值可以看出,随着系统使用时间的增长,发音部位的准确性得到了较大的提

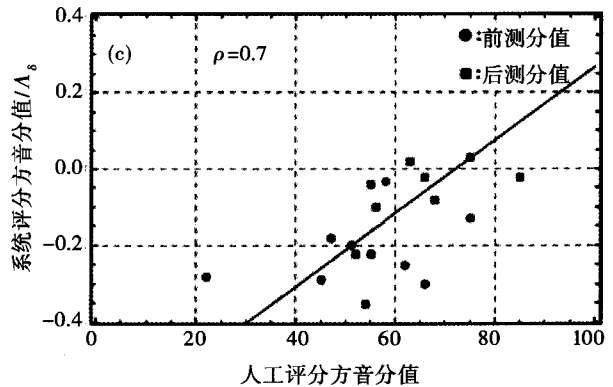


图5 人工评分方音分值和系统评分方音分值

高。在元音发音错误中,就舌前特征而言,中元音是发音难度最大的。其中, $\Lambda$ - $\alpha$ :和 $\ominus$ -I替换最多,但提高也较快。就舌高而言,中-高元音替换误用是最常见的错误类型, $i$ :- $\ominus$ I和 $\ominus$ -I替换出现最多,但前测和后测的 $\Delta$ <sub>s</sub>平均值显示,这种错误提高也较快。

## 4 总 结

依赖于方音的非本族语语音识别技术应用前景广阔。从算法研发和软件开发的角讲,以方言为背景的语音识别为现有语音识别技术增加了一个新的研究层面。从语音习得的角度讲,以方音为背景的机助语音学习系统的建立对二语语音习得研究和教学有着重大的现实意义。以河南方音为背景的二语语音自动识别不仅能为河南省外语语音学习提供有效的辅助,还可以为其它方言区的英语语音学习系统的建立提供一个模型。

## 参 考 文 献

- [1] Huang C, Chen T, Chang E. Accent issues in large vocabulary continuous speech recognition[J]. *International Journal of Speech Technology*, 2004, (2/3): 141-153.
- [2] Flege J E, Schirru C, MacKay I R A. Interaction between the native and second language phonetic subsystems[J]. *Speech Communication*, 2003, 40: 467-491.
- [3] Mangold H. Speech Technology in reality- applications, their challenges and solutions[C]. 4<sup>th</sup> International Conference-Text, Speech and Dialogue, Zelezná Ruda, 2001.
- [4] Callaway E C, Clifford F S, Eliot S, et al. Speech recognition interface to a hospital information system using a self-designed visual basic program: initial experience[J]. *Journal of Digital Imaging*, 2002(1): 43-53.
- [5] Munro M J, Derwing T M. Modelling perceptions of the accentedness and comprehensibility of L2 speech[J]. *Studies in Second Language Acquisition*, 2001, 23: 451-468.
- [6] Yusnita M A, Paulraj M P, Sazali Y, et al. Acoustic analysis of formants across genders and ethnical accents in Malaysian English using ANOVA[J]. *Procedia Engineering*, 2013, 64: 385-394.
- [7] Amino K, Osanai T. Native vs. non-native accent identification using Japanese spoken telephone numbers[J]. *Speech Communication*, 2014, 56: 70-81.
- [8] Goronzy S. Robust Adaptation to Non-Native Accents in Automatic Speech Recognition[M]. Berlin: Springer-Verlag, 2002.
- [9] Derwing T M, Munro M J. Accent, intelligibility, and comprehensibility: evidence from four L1s[J]. *Studies in Second Language Acquisition*, 1997, 20: 1-16.
- [10] Arslan L M, Hansen J H L. A study of temporal features and frequency characteristics in American English foreign accent[J]. *Journal of the Acoustical Society of America*, 1997(1): 28-40.
- [11] Angkititrakul P, Hansen J H L. Use of trajectory models for automatic accent classification[C]. Eurospeech Geneva, Switzerland, 2003.
- [12] Bartkova K, Jovet D. On using units trained on foreign data for improved multiple accent speech recognition[J]. *Speech Communication*, 2007, 49: 836-846.
- [13] Sangwan A, Hansen J H L. Automatic analysis of Mandarin accented English using phonological features[J]. *Speech Communication*, 2012, 54: 40-54.

## Automatic Analysis of Henan-accent-dependent English Pronunciation

LI Nan

(Faculty of International Studies, Henan Normal University, Xinxiang 453007, China)

**Abstract:** Firstly, the accented second language phonetic corpus is built and its phonological features are analyzed. Then based on HMM, this paper has proposed a feasible scheme for the automatic speech recognition of the foreign language learners' speech production from specific accent background in Henan. The result of performance testing shows that the system may recognize Henan-accented English, evaluate both acceptable and unacceptable productions, and promote their foreign language pronunciation.

**Keywords:** accented non-native speech recognition; HMM; second language phonetic acquisition