

中介语声调系统的评价方法研究*

熊子瑜

中国社科院语言研究所

文 摘: 本研究基于大规模中介语语音数据的统计和分析, 提出了四项具体的参考指标来评价中介语的声调系统: 调类的区分性系数和独立性系数、调值的稳定性系数和正确性系数, 并给出了每项指标的具体测算方法。研究结果初步表明, 这些指标可为中介语声调系统提供较为客观的、可以量化的、具有操作性的评价, 将有助于全面有效地评价普通话学习者的声调系统, 揭示学习者发音中存在的声调缺陷和问题, 为学习者制定具有针对性的改进方案提供依据。

关键词: 中介语; 声调系统; 评价方法

中图分类号: H017; H116.4

1 引言

声调是指具有区别词汇意义的音高对比。普通话有阴平、阳平、上声和去声四个调类, 不同调类之间靠音高对比来区别: 阴平调具有高平特征, 标准调值为 55; 阳平调具有中升特征, 标准调值为 35; 上声调具有低降升特征, 标准调值为 214; 去声调具有高降特征, 标准调值为 51^[1]。

学习者对普通话声调系统的掌握是个朝着目标不断前进的过程, 其间发音往往不够标准, 甚至会有些错误, 这主要表现在以下几个方面: 调类独立性不强, 声调系统残缺; 调类区分性不显著, 彼此有交叉; 调值不稳定, 变化类型较多且分散; 调值偏误, 与标准调值存在显著差异。当前普通话水平测试主要是依据测评员的主观判断对发音人的声调进行打分, 这种评测方式不仅费时费力, 而且存在宽严标准不一致的潜在风险。随着语音技术的快速发展, 一些基于语音识别技术的普通话水平自动评测系统已经被研制出来。据悉, 由香港理工大学、中科院声学所、中科院软件所共同开发的香港普通话水平考试单双音节朗读自动评测系统日前通过了专家鉴定^[2]。由中科大讯飞信息科技有限公司研制的“国家普通话水平智能测试系统”可对命题说话之外的所有测试题型实现自动评测, 并可自动检测发音者存在的语音错误和缺陷。该系统已通过验收, 并在国内部分省市的普通话水平评测中得到实际应用和推广^[3]。普通话水平自动评测系统的基本工作原理是比对实际发音跟标准发音之间的相似程度, 并采取一定的统计方法计算实际发音

的正确率。这些自动评测系统在给声调打分时, 往往只关注声调调值是否准确。

本研究认为, 要全面分析学习者的声调缺陷和问题, 并为其制定具有针对性的改进方案, 不仅要考虑声调调值的正确性, 还应该综合考虑调类的独立性, 调类的区分性, 以及声调调值的稳定性等因素。比如说, 若测得某发音人的上声错误率较高, 则还需要进一步了解其上声的错误原因, 仅仅是调值错误, 还是跟其他调类混淆等。

2 基频数据提取和处理

声音的高低主要决定于基频的大小, 有研究表明, 二者在 500Hz 以下成近似线性关系^[4], 基频越大的声音听起来越高。因此有些学者将基频值换算成音高值, 并在此基础上计算声调的调值^{[5][6][7][8]}。这种方法的优点是具有可计算性, 把纯主观的听辨过程转变为具有可操作性的测算过程, 这不仅能在一定程度上排除因研究者的经验不足所造成的判别错误, 而且能够用来快速处理和分析大规模的语音数据。但根据基频数据测算声调调值实际上也是一个相当复杂的过程。其间影响声调调值测算结果的因素较多, 如: (1) 基频数据的准确性; (2) 如何确定发音人的声调音域范围; (3) 将基频值换算成音高值的方法; (4) 如何从音节的基频曲线中提取正确的音高特征点数据以确定调值; (5) 音高规整的度数; 等等。

2.1 基频数据的校准

本研究采用 Praat¹ 语音分析软件对语音材料进行标注和分析, 主要包括音节及声韵母边界的标注和基频数据的提取。采用语音分析软件直接提取出来的基

* **基金项目:** 本文研究部分得到国家 985 工程的项目资助 (项目名称: 藏族人学汉语的中介语语音数据库建设), 部分得到 Nokia 中国研究中心的项目资助 (项目名称: 汉语方言口音研究), 部分得到中国社会科学院重大项目的资助 (项目名称: 普通话语音基础数据库建设和音高模式研究)。

¹ <http://www.fon.hum.uva.nl/praat/>

频数据, 难免会出现一些错误, 特别是音节基频段的起始和结束部分, 由于能量相对较弱, 周期性不够显著, 基频数据出现错误的概率会相对较大, 因此需要对音节的基频数据进行一些自动的和手工的修正。

为适应大规模的语音数据分析, 本研究在 Praat 软件的基础上设计并采用了一种基于低频谱能量和次谐波峰频率来辅助修正基频数据的方法: 通过检测低频谱能量以确定基频值的删除和添补, 通过测算低次谐波峰的频率来检校基频值。测量低频谱能量时, 其下限频率应该略低于发音人的音域下限, 上限频率应该略高于发音人的音域上限, 使低频段至少能涵盖第一谐波。通常低频段里会出现一至三个谐波, 其中频率较低的谐波往往会具有相对较高的能量, 但受共振峰或其他因素的影响, 第一谐波峰不一定总是能量最强峰。为了增加第一谐波峰被选取的概率, 本研究以低频段的下限频率为起点, 采取每倍频程递减 3 个 dB 的方法降低高次谐波峰的能量。然后根据自动检测到的数据, 逐一检校并修改 Pitch 对象中的数据, 为接下来进行手工检校基频数据提供更为准确的参考。自动检校 Pitch 对象的过程和效果如下图所示:

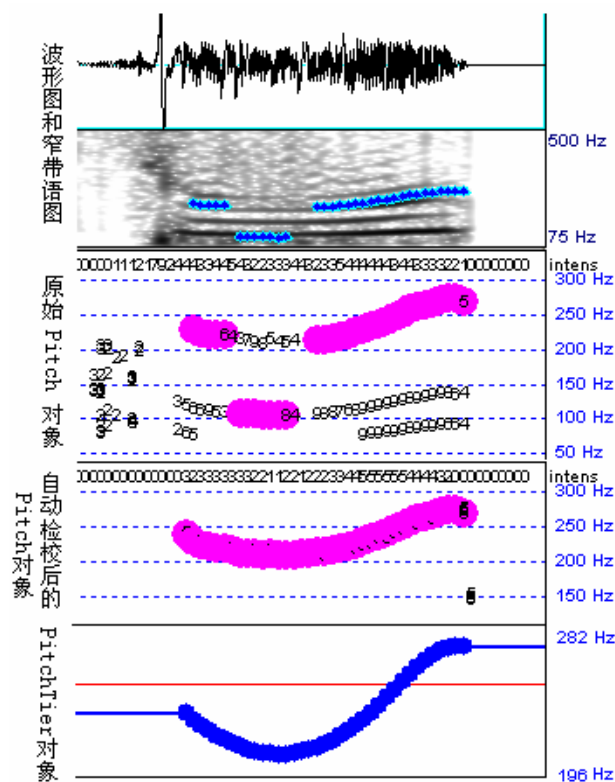


图1 Pitch 对象的自动检校过程和效果示例

数据检查结果表明, 这种方法能在一定程度上有效剔除孤立的和连续的基频散点、校正和增补基频值。但对于音节基频数据中的“弯头降尾”部分还是不能有效删除, 为了降低这些基频数据对声调调值分析结果的影响, 本研究一方面在手工检查时删除明显的弯

头降尾段基频数据, 另一方面根据声韵母边界数据, 在分析调值时, 把浊声母段的基频数据排除在外。

另外, 为了便于精确确定音高特征点, 本研究还利用 Praat 软件自身带有的插值和平滑模块对检校后的基频数据进行了处理。

2.2 测算发音人的声调音域上限和下限

计算声调调值时, 需要根据发音人的声调音域范围对其 $f0_i$ 进行规整。如何确定发音人的声调音域上限和声调音域下限, 直接关系到声调调值的测算结果。如果单取发音人的基频最大值和基频最小值为声调音域上限和声调音域下限, 很可能会受极值的影响而导致测算的调值结果不准确。为排除极值的影响, 本研究按以下方法来确定发音人的声调音域上限和声调音域下限。

$$f0_{Amax} = f0_k$$

$$\text{令 } k = \text{round}(97.5\% * \text{allNum})$$

其中, $f0_{Amax}$ 表示发音人的声调音域上限频率, $f0_k$ 表示按自小到大的顺序排列后该发音人的第 k 个基频值, allNum 为该发音人提取的基频值总个数, $\text{round}()$ 函数用以求整。

$$f0_{Amin} = f0_k$$

$$\text{令 } k = \text{round}(2.5\% * \text{allNum})$$

其中, $f0_{Amin}$ 表示发音人的声调音域下限频率, $f0_k$ 表示按自小到大的顺序排列后该发音人的第 k 个基频值, allNum 为该发音人提取的基频值总个数, $\text{round}()$ 函数用以求整。

设定 97.5% 和 2.5% 的取值范围, 目的在于排除 2.5% 的极大值和 2.5% 的极小值, 以取得相对稳定的声调音域上限频率和声调音域下限频率。

2.3 基于声调音域上下限调整音节基频数据

因存在 2.5% 的极大值和 2.5% 的极小值超出了发音人的声调音域范围, 故需要对这些基频值进行重新调整, 使得调整后的基频数据都处于发音人声调音域范围之内。本研究根据发音人的声调音域上限 $f0_{Amax}$ 和声调音域下限 $f0_{Amin}$, 对每个音节的基频数据进行检查, 并采取以下操作:

(1) 如果音节的基频最大值 $f0_{max}$ 和音节的基频最小值 $f0_{min}$ 都大于发音人的声调音域上限 $f0_{Amax}$, 则直接将该音节的全部基频值都设定为声调音域上限 $f0_{Amax}$ 。

(2) 如果音节的基频最大值 $f0_{max}$ 和音节的基频最小值 $f0_{min}$ 都小于发音人的声调音域下限

$f0_{Amin}$ ，则直接将该音节的全部基频值都设定为声调音域下限 $f0_{Amin}$ 。

(3) 如果音节的基频最大值 $f0_{max}$ 大于发音人的声调音域上限 $f0_{Amax}$ ，则将该音节的全部基频数据进行如下换算：

$$f0_i = \frac{(f0_i - f0_{min}) * (f0_{Amax} - f0_{min})}{(f0_{max} - f0_{min})} + f0_{min}$$

其中 $f0_{min}$ 表示该音节的基频最小值。换算之后，使得该音节的基频最大值压缩至发音人的声调音域上限，并保持其基频最小值不变，同时使得介于基频最大值和基频最小值之间的基频值按一定比例压缩，且维持调形基本不变。

(4) 如果音节的基频最小值 $f0_{min}$ 小于发音人的声调音域下限 $f0_{Amin}$ ，则将该音节的全部基频数据进行如下换算：

$$f0_i = \frac{(f0_i - f0_{min}) * (f0_{max} - f0_{Amin})}{(f0_{max} - f0_{min})} + f0_{Amin}$$

换算之后，使得该音节的基频最小值上抬至发音人的声调音域下限，并保持其基频最大值不变，同时使得介于基频最小值和基频最大值之间的基频值按一定比例压缩，且维持调形基本不变。

(3) 和 (4) 的操作结果图示如下：

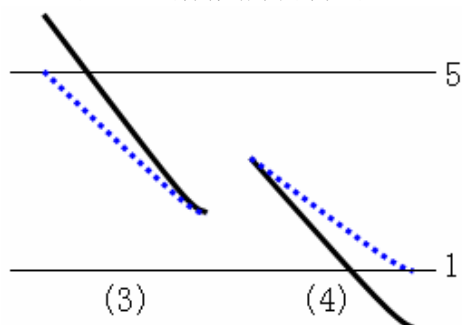


图2 (3) 和 (4) 的操作结果示意图

上图 (2) 中的黑色实线表示原始基频数据曲线，蓝色虚线表示调整后的基频数据曲线。这样做的目的就是把所有单音节的基频数据都调整到发音人的声调音域范围之内，以保证最后规整出来的音高数据不会大于 5 和小于 1。从图示可看出，调整前后的单字调形保持基本不变。而且对于一个发音人的数据来说，只有不到 3% 的音节需要进行此类基频数据调整。

(5) 如果同时存在上述 (3) 和 (4) 两种情况，则先操作 (3)，再操作 (4)；

2.4 提取音节的音高特征点基频数据

本研究基于声韵母边界的标注结果，将每个单音节的调型段在时间维度上等分为 $n-1$ 个基频段，并自其起点至末点等时长间隔地提取 n 个基频值 $f0_i$ ，其中 $i = 1, 2, \dots, n$ 。

为了防止 $f0_i$ 过大或过小，先计算每个基频段的基频均值 $f0Mean_i$ 和标准差 $f0Stdv_i$ ，其中 $i = 1, 2, \dots, n-1$ ，并以此作为参考剔除 $f0_i$ 中可能出现的奇异值：

(1) 如果 $f0_i > f0Mean_k + 2 * f0Stdv_k$ ，则令 $f0_i = f0Mean_k + 2 * f0Stdv_k$ ；

(2) 如果 $f0_i < f0Mean_k - 2 * f0Stdv_k$ ，则令 $f0_i = f0Mean_k - 2 * f0Stdv_k$ ；

其中，当 $i < n$ 时，令 $k = i$ ；

当 $i = n$ 时，则 $k = i-1$ 。

在本研究中， n 值设定为 5，即对每个单音节的调型段等时长间隔地提取 5 个基频数据。

采取这种方法提取出来的音高特征点基频数据能够较好地反映调型段的基频走势，其缺点是随着 n 值增大会大幅增加调值类型的数量。

为了使提取出来的音高特征点基频数据能够较好地体现普通话的声调调值特性，本研究还采取了另一种方案，只提取音节调型段的起点基频值、末点基频值、基频最大值和基频最小值，然后根据这些关键基频值的时间间距以及音高间距等因素来确定音节的音高特征点基频数据。采取这种方法对每个调型段最多会提取出 4 个特征点基频数据，最少会提取出 2 个特征点基频数据。

2.5 音高数据规整和调值测算

对不同发音人间的音高数据进行规整的方案有很多，朱晓农^[9]细致地比较了多种规整方案在消除人际音高差异方面的效度，并提出用“标准指数”和“离散系数”两个指标来衡量各类规整方法。本文所谓的音高规整是指基于音高值测算调值的过程。因为在规整过程中针对不同的发音人，采用了不同的声调音域上下限频率，因此能在一定程度上消除发音人的音域差异和音阶差异。采取如下公式将音高特征点基频数据换算成声调调值：

(1) 线性规整公式：

$$X_i = n * \frac{f0_i - f0_{Amin}}{f0_{Amax} - f0_{Amin}}$$

(2) 非线性规整公式:

$$X_i = n * \frac{\log_2(f0_i / f0_{Amin})}{\log_2(f0_{Amax} / f0_{Amin})}$$

其中, $f0_{Amax}$ 和 $f0_{Amin}$ 分别为发音人的声调音域上限和下限, n 为规整的度数, 经过规整之后得到的 X_i 取值范围在 0 至 n 之间, 再对规整后的数据采取以下方法进行求整:

$$X_i = \text{floor}(X_i) + 1$$

如果 $X_i > n$, 则令 $X_i = n$ 。

其中, $\text{floor}(X_i)$ 函数的功能是取小于或等于 X_i 的最大整数。如 X_i 等于 4 或 4.1 或 4.9, 经过 $\text{floor}(X_i)$ 函数求整后, 其返回值都等于 4。

本研究采取非线性规整公式, 并将 n 值设定为 5, 即进行音高 5 度值的非线性规整。

如前文所述, 根据音节的基频数据计算其声调调值是一个比较复杂的过程。如何使计算出来的声调调值既能较好地反映音节的音高走势, 又能符合语音听感, 并具有语言学意义? 这远非本文所能解决的, 还有待学界进一步研究。本研究采用的音高特征点基频数据提取及规整方法是否有效, 也还有待进一步检验。

3 中介语声调系统的评价方法

分析某种语言或方言的声调系统时, 一般包括调类、调值以及连读变调等方面。本研究仅涉及单音节数据, 所以暂不考虑连读变调问题, 只从调类和调值两个方面来分析和评价中介语的声调系统。其中, 对调类系统的评价又分为调类的独立性和区分性两个方面, 对调值系统的评价又分为调值的稳定性和正确性两个方面。

3.1 调类的独立性系数

调类的独立性系数用来度量某个调类跟声调系统中其他调类的综合区分程度。一般说来, 如果某个调类使用的调值类型是其所特有的, 不跟或很少跟其他调类的调值类型发生交叉, 那么该调类就具有相对较高的独立性。相反, 如果某调类的调值类型跟其他调类的调值类型发生较多交叉, 则该调类的独立性就会较低。例如, 古汉语中的入声字在北京话里已被分派到了平、上、去三个调类之中了, 一部分古入声字读为阴平, 一部分古入声字读为阳平, 一部分古入声字读为上声, 还有一部分古入声字读为去声^[10]。可见古入声字在北京话里已没有了其特有的调值类型, 因此在北京话里已经失去了其独立性。

本研究采用以下方法计算调类独立性系数:

$$I = \sum_{i=1}^n \left(\frac{(\text{sylNsTone}V_i^{TD} / \text{sylNsTone}T^{TD})^2}{\sum_{k=1}^t (\text{sylNsTone}V_i^{Tk} / \text{sylNsTone}T^{Tk})} \right)$$

其中, n 表示 TD 类声调所具有的调值类型总数目, i 的取值范围为 1 至 n ; $\text{sylNsTone}V_i^{TD}$ 表示 TD 类声调中读为第 i 类调值类型的音节数目, $\text{sylNsTone}T^{TD}$ 表示 TD 类声调的音节总数目; t 表示声调系统中的调类数目, k 的取值范围为 1 至 t ; $\text{sylNsTone}V_i^{Tk}$ 表示第 i 类调值类型下 T_k 类声调的音节数目, $\text{sylNsTone}T^{Tk}$ 表示发音人读 T_k 类声调的单音节总数目。

I 值得分越大, 表明 TD 类声调的独立性越强, 跟其他调类的综合区分度就越好。下表给出了五个地区发音人的调类独立性系数, 从中可以看出: 藏区中介语的阴平和上声的独立性系数都相对较低, 另外安多藏方言区阴平调的调类独立性系数也比较低; 而汉方言区(广州和厦门)中介语的各调类独立性系数都相对较高。

表 1 五类中介语²的调类独立性系数

	综合	阴平	阳平	上声	去声
安多	2.22	0.55	0.39	0.55	0.73
拉萨	2.98	0.87	0.57	0.60	0.94
康巴	3.16	0.94	0.61	0.63	0.98
广州	3.73	0.97	0.92	0.89	0.95
厦门	3.77	0.96	0.97	0.91	0.93

调类的独立性系数可以用来分析发音人的调类系统, 如果各调类的独立性系数都比较高, 则说明该发音人习得的调类系统比较完整。如果某个或某些调类的独立性系数较低, 则说明该发音人可能存在调类区分不好、甚至调类缺失的问题。

3.2 调类的区分性系数

基于调类的独立性系数, 可以确定发音人是否存在调类区分不好或者调类缺失的问题, 但还不能确定具体是哪些调类之间存在混淆。如上表 1 数据显示, 安多藏方言区的阴平、阳平和上声的独立性系数都较低, 那么它们之间到底谁跟谁容易混淆呢? 要回答

² 本研究使用的语音材料包括从粤方言地区(广州人)和闽方言地区(厦门人)采集到的普通话中介语单音节数据, 还包括从藏族地区(安多方言区的西宁人、卫藏方言区的拉萨人和康巴方言区的甘孜人)采集到的普通话中介语单音节数据。每个调查点约选取了 30 位发音人, 男女各半, 发音人基本上都是在读本科生, 他们的普通话水平主要分布在二乙及以下水平。每位发音人的发音材料中大约有 150 个单音节, 四类声调均衡覆盖。

这个问题,则需要进一步分析调类的区分性系数。

调类的区分性系数用来度量发音人对不同调类的区分程度。如果发音人朗读两个不同调类的例字时,使用的调值类型交叉程度很高,则会导致这两个调类的区分性较低。例如,某发音人读 20 个阴平字,都是 55 调,读 20 个阳平字,都是 35 调,则其阴平调和阳平调在调值类型上没有交叉,因此这两个调类间的区分性较高;如果其中有 10 个阴平字被读成了 35 调,则其阴平调和阳平调在调值类型上有一定的交叉,这两个调类之间的区分性就会降低。

本研究采用以下方法来计算两个调类之间的区分性系数:

$$D = \frac{\sum_{i=1}^n \left| \frac{\text{sylNsTone}V_i^{T1} / \text{sylNsTone}T^{T1}}{\text{sylNsTone}V_i^{T2} / \text{sylNsTone}T^{T2}} \right|}{2}$$

其中, $\text{sylNsTone}T^{T1}$ 和 $\text{sylNsTone}T^{T2}$ 分别表示要计算区分性系数的两个调类下的单音节数目, n 表示这两个调类所具有的调值类型变体总数目, $\text{sylNsTone}V_i^{T1}$ 和 $\text{sylNsTone}V_i^{T2}$ 分别表示这两类声调中具有同一调值变体类型的音节数目,其中 $i = 1, 2, \dots, n$ 。

如上所述, D 值的大小跟两个调类之间调值类型的交叉程度相关, D 值得分越大,表明这两个调类的区分性越好。下表 2 给出了五个地区发音人的调类区分性系数,从中可以看出:方言区学习者的调类区分性要整体好于藏语区的学习者,这跟调类的独立性系数是一致的。对于三个藏方言区的学习者来说,阳平和上声这两个调类的区分性系数都比较低。另外,由于安多藏方言区的阴平和阳平的区分性也比较差,导致该方言区的阴平调类的独立性系数也相对较低。比较而言,安多藏区学习者的调类区分性最差,这可能跟安多藏语无声调有关^[11]。

表 2 五类中介语的调类区分性系数

	综合	阴- 阳	阴- 上	阴- 去	阳- 上	阳- 去	上- 去
安多	4.11	0.55	0.77	0.76	0.41	0.79	0.84
拉萨	5.16	0.90	0.94	0.95	0.42	0.98	0.98
康巴	5.39	0.96	0.98	0.98	0.49	0.99	0.99
广州	5.82	0.99	1.00	0.98	0.90	1.00	0.96
厦门	5.84	0.99	0.98	0.98	0.97	1.00	0.92

需要说明的是,调类的独立性系数和区分性系数主要是用来分析发音人的调类系统,以考察其调类系统中是否存在调类混淆或调类缺失的问题。基于调类的独立性系数和区分性系数并不能直接得出发音人对

哪个或哪些调类掌握得较好结论来,因为调类上的独立性和区分性并不能保证其调值的正确性。但对于那些独立性较差或区分性较低的调类,可提醒学习者加以对比学习和训练,以获得更好的独立性和区分性。

3.3 调值的稳定性

一个调类的调值类型变体越少,分布就越集中,稳定性越高,可正确识别率也就越高。影响调值稳定性的因素很多,一方面受其自身的音高特征影响,相对低调而言,高调可能具有更多的调值变体,这是因为发音人的声调音域下限通常会达到其生理音域的下限位置,而发音人的生理音域上限要比其声调音域上限高的多,这给高调的实现留下了更多的活动空间。另一方面还受研究者所采取的调值测算方法所影响,对调值描写的越精细,通常会导致调值类型的变体增多,从而会降低调值的稳定性。此外,调值的稳定性还跟发音人对声调的掌握程度和控制能力密切相关。发音人对声调调值掌握的越熟练,其调值会实现得越精准,稳定性就会越高,否则就会较差。

本研究采取以下方法分析调值稳定性系数:

$$S = \frac{\sqrt{\sum_{i=1}^n (\text{sylNsTone}V_i^{TD} * \text{sylNsTone}V_i^{TD})}}{\text{sylNsTone}T^{TD}}$$

其中, n 表示发音人朗读某类声调所产生的调值类型变体的总数目, i 的取值范围为 1 至 n , $\text{sylNsTone}V_i^{TD}$ 表示发音人朗读的 TD 类声调中具有第 i 类调值的音节数目, $\text{sylNsTone}T^{TD}$ 表示发音人朗读 TD 类声调的单音节总数目。

某调类的 S 值大小与其调值类型变体的数目相关,减少调值类型变体的数目,可以提升该调类的稳定性, S 值也就会越大; S 值的大小还跟该调类的调值类型的使用概率有关,提高该调类的某调值类型的使用概率,会提升该调类的稳定性, S 值也就会越大。

S 的取值下限为 $\sqrt{1/\text{sylNsTone}T^{TD}}$, 取值上限为 1。这表明其下限取值跟 $\text{sylNsTone}T^{TD}$ 的大小有关。若要消除音节数目差异对值造成的影响,可对 S 值进行如下规整:

$$S = \frac{S - \sqrt{1/\text{sylNsTone}T^{TD}}}{1 - \sqrt{1/\text{sylNsTone}T^{TD}}}$$

规整后的 S 值,取值下限为 0, 上限为 1。 S 值越大,表明该调类的调值类型分布越集中,调值系统越稳定。比如说,一个发音人读 20 个阴平字,都是 55 调,则其阴平调的调值稳定性系数为 1; 如果其中有

10个读为55调,有10个读为44调,则其阴平调的调值稳定性系数为0.623,稳定性系数明显降低。

在计算调值的稳定性系数过程中,可以同时获得各调类的调值类型分布数据,可用来回答某调类主要采用的调值类型是什么之类的问题,详见本次会议的有关论文^{[12][13]}。

3.4 调值的正确性

调类的区分性系数和独立性系数以及调值的稳定性系数这三个指标主要是用来度量中介语声调系统是否完整和稳定,如果这三个指标都能获得较高的分值,则表明中介语的声调系统是完整和稳定的。但即便如此,仍然不能保证中介语声调调值的正确性。例如,一个发音人把阳平全部读成214调,把上声全部读成35调,这种系统性错误并不会影响上述三项指标的得分。因此,对中介语声调系统的评价必需充分考虑调值的正确性系数,而且这是一项最为重要的评价指标。由于计算调值的正确性系数时需要以目标语的声调调值系统为参考,所以首先需要计算出目标语各调类的调值类型的出现概率,计算方法如下:

$$V_i^{Tk} = \text{sylNsTone}V_i^{Tk} / \text{sylNsTone}T^{Tk}$$

其中, V_i^{Tk} 表示 T_k 类声调出现第*i*种调值类型的概率, $\text{sylNsTone}V_i^{Tk}$ 表示 T_k 类声调出现第*i*种调值类型的音节数目, $\text{sylNsTone}T^{Tk}$ 表示 T_k 类声调的音节总数目。

如果中介语某类声调中出现的调值类型在目标语的同类声调下有着较高的出现概率,则该调值类型的正确率就会较高,如果在目标语的同类声调下没有出现,则该调值类型的正确率就会较低。本研究采取以下方法计算中介语声调调值的综合正确率:

$$R^{Tk} = \sum_{i=1}^n (\text{sylNsTone}V_i^{Tk} / \text{sylNsTone}T^{Tk} * V_i^{Tk})$$

其中, n 表示发音人朗读某类声调所产生的调值类型变体数目, i 的取值范围为1至 n , $\text{sylNsTone}V_i^{Tk}$ 表示发音人朗读的 T_k 类声调中具有第*i*类调值的音节数目, $\text{sylNsTone}T^{Tk}$ 表示发音人朗读 T_k 类声调的单音节总数目, V_i^{Tk} 表示在目标语中 T_k 类声调出现第*i*种调值类型的概率。

R^{Tk} 值越大,表明该调类的调值正确率越高。

4 评价参数的意义和应用

对于声调系统的研究和评价来说,上述四项指标

各有侧重。

研究中介语的声调系统时,调值的正确性系数和稳定性系数是较为重要的评价指标,不仅可用来衡量学习者对目标语声调系统的整体掌握程度,还可用来分析学习者对各个调类的具体掌握程度,以及调类的错误类型和概率等。调类的独立性系数和区分性系数反映了学习者区分调类的能力,如果独立性系数和区分性系数较低,则表明该学习者应加强调类间的对比训练。这些指标不仅可以反映出发音人的声调掌握水平和存在的问题,还可以为制定有针对性的学习计划和改进方案提供数据支持。

而在研究某个语言或方言的声调系统时,只要找到合适的发音人,就可以不考虑正确性系数这项指标,因为对于一个未知声调系统的方言研究来说,事先无法确定其标准调值类型。而其他三项指标显得尤为重要,调值的稳定性系数能够在一定程度上反映出发音人对声调产生的控制能力,如果稳定性系数较差,同一调类下的调值类型变体较多,则表明发音人不太适合担任发音任务。调类的区分性系数反映了调类之间的区分程度,如果两个调类间的区分性系数较低,则表明这两个调类可能需要进行归并。调类的独立性系数反映了调类的独立程度,如果某个调类的独立性系数较低,则表明该调类中的例字已被分派到其他调类之中了,该调类可能已经消失。因此综合参考上述指标,不仅可以考察方言的声调系统(如存在哪些调类,各调类的常用调值等),还可以进一步考察古汉语声调在该方言声调系统中的分化演变规律。

5. 讨论

如前文所述,要全面有效地评价发音人的声调系统,所涉及的因素和过程很复杂。首先是声调调值的测算,如何基于音节的基频数据测算其声调的调值,并使之既能较好地反映音节的音高走势,又能符合语音听感,并且具有一定的语言学意义?其次是各评价指标的计算方法,如何基于声调的调类和调值数据来评价发音人的声调系统,使自动评价的结果跟人工评价的结果具有较高的一致性?再其次是对发音人声调系统的概括,如何基于计算出来的评价指标来正确判定发音人的声调系统存在哪些缺陷或问题?这个过程中某一环节出问题都会导致评价结果无意义。

本研究对此工作还只是个尝试,初步提出了一些供参考的评价指标和测算方法,但这些评价指标的测算过程和方法也许还存在这样或那样的问题,有待进一步检验和改进。例如,研究中我们看到,在标准普通话里绝大部分的去声都读为51调,因此当中介语中

去声读为 51 调时的得分率也就相应较高，而有一位发音人由于其上声的低点相对较低，导致其大部分去声调值测算为 52 调，由于 52 调在标准普通话的去声里出现率和得分率比较低，从而导致这位发音人的总体调值正确性系数降低，但在听感上该发音人的去声 52 调仍然可以认为是正确的。这表明评价指标的测算结果和实际听感之间出现了不一致。因此需要重新考虑评价指标的测算方法，如进一步考虑同一调类下不同调值类型间的相似距离来计算调值的正确性系数，或者重新考虑声调值的提取及规整方法，以提高评价结果跟人工评测结果之间的一致性。

参 考 文 献

- [1] 赵元任著，吕叔湘译《汉语口语语法》，北京：商务印书馆 1979 年版。
- [2] 引自新华网 07 年 5 月 22 日新科技频道的科技新闻栏目，网址：
http://news.xinhuanet.com/tech/2007-05/22/content_6136381.htm
- [3] 引自讯飞公司网站：
<http://www.iflytek.com/Html/ywkd/xfdt/143327849.html>
- [4] 引自吴宗济、林茂灿主编《实验语音学概要》，P71，高等教育出版社 1989 年版。
- [5] 刘复，《四声实验录》，1924 年，上海：上海群益出版社。
- [6] 林茂灿，音高显示器与普通话声调音高特性，《声学学报》第 2 卷第 1 期，1963 年。
- [7] 沈炯，北京话声调的音域和语调，《北京语音实验录》（林焘、王理嘉等著），1985，北京：北京大学出版社。
- [8] 石锋，《语音从稿》，北京：北京语言学院出版社，1994。
- [9] 朱晓农，基频归一化——如何处理声调的随机差异，徐州：《语言科学》2004 年第 2 期。
- [10] 林焘、王理嘉，《语音学教程》，P131，北京：北京大学出版社 1992 年版。
- [11] 胡坦，《藏语研究文论》，P165，中国藏学出版社 2002 年版。
- [12] 段海凤等，藏族人学普通话的单字调系统研究，本会论文集。
- [13] 贾媛等，甘孜藏汉中介语声调系统研究，本会论文集。

Methodology of Evaluating Tonal System of Inter-language

XIONG Ziyu

Institute of Linguistics, Chinese Academy of Social Sciences

Abstract: Based on the statistical analysis of large-scale inter-language data, this study proposes four specific parameters to access the inter-language tonal system, specifically, it contains the coefficient of independence and distinction for tonal type and the coefficient of stability and distinction for tonal values, besides that, this study further states the concrete calculating method. The preliminary results demonstrate that these parameters can give the inter-language tonal system an objective, quantifiable, operational evaluation which is helpful to access the tonal system of Mandarin learners and rectify the defects and problems in the pronunciation of the tones so as to provide the improved program for the learners.

Key words: Inter-language; tonal system; evaluation method