

对日本大学生汉语声调发音能力的进一步探讨

—从对声调间“融合发音”的探索谈起—

丁 雷

要旨

丁雷(2015)の考察を踏まえ、声調と声調を続けて発音する過程における様々な誤りや訛りは、声調間における「溶け合い」のような「発音能力」(融合)との一定な結びつきがあると結論づけた。本研究は、中国語母語話者や大学における第2外国語・中国語の学習者を研究対象にして、多数の音声実験の考察を踏まえ、この声調間における「溶け合い」過程が発音活動にどのような影響を与えるかを解明し、母語話者の声調に対する発音メカニズムを構想する。また、実際の会話における学習者の声調誤りや訛りの矯正活動に、最新の音声合成技術(Manipulation)を導入し、誤りや訛りの見えない発音過程も具体化することも提案する。

キーワード：日本汉语教育 汉语声调 发音能力

1. 对于汉语声调发音能力的认识

近年来，我们一直非常关注日本汉语学习者的声调习得问题。一方面是希望更好地指导学生学好汉语发音，另一方面也是想探索母语为非声调语言的人群，在学习声调语言时会产生哪些能力方面的变化。这些变化或许能反映出人类对声调的把握与“语言获得机制”(Language Acquisition Device, LAD)¹⁾之间的关系。以这两点为初衷，围绕发音中的“声调难”问题，我们以日本大学生为对象，主持并开展了各种各样(单音节、双音节、三音节等)的语音实验。在丁雷(2012a、2013)里，笔者对于这些实验研究的成果做了分析和总结，研究的结果显示：“声调难”的根本原因在于音节和音节之间的“融合发音”难。

对于“融合发音”的感性认识，笔者在丁雷(2013)指出：在语音实验时，我们可以明显地感受到在学生进行调与调之间“融合发音”的过程中，总有一种“无形力量”在影响着学生的发音质量，尤其是在双音节和三音节测试中，这一现象更加明显。笔者将这种“无形力量”理解为声调的“发音能力”，并提出了“声调的发音能力=认调能力+拼调能力”这个假设。对于这个假设，笔者在丁雷(2013)的基础上做了进一步的实验探索(针对“大融合”和“小融合”的“聚合→分离”和“分离→聚合”实验)。实验的结果让我们感受到“认调能力”更多依靠发音者有意识的控制，而“拼调能力”则倾向于自动化，只能在反复练习中“潜移默化”地被动习得，难以直接进行有针对性的训练(丁雷2015,p195)。

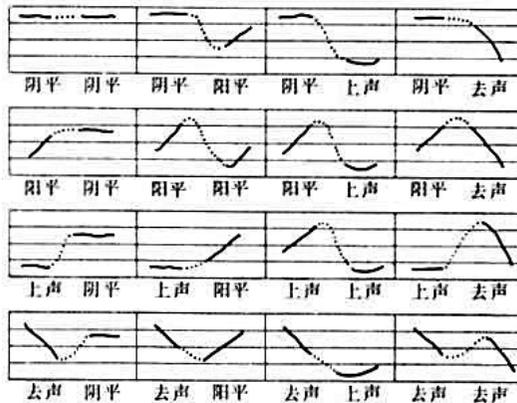
研究至丁雷(2015)为止，我们还没能拿出一个解决“声调难”问题的具体方案，这主要是因为目前对于“融合发音”的认识还不够充分。即便我们曾提出“融合发音”与学生的“认调能力”和“拼调能力”有关的假设，但这样的认识无论对于教学还是对于人类声调能力的探索都没有明显的帮助。在音节和音节进行“融合发音”的时候，在那几微秒的“融合”过程中到底发生了什么？人究竟是如何控制声调发音的？这种控制能力有什么样的特征？

这些疑问的解答都有待我们对“融合发音”的过程做进一步的研究探索。

2. 对于声调间“融合发音”的既有认识

笔者在丁雷(2015)的研究中曾提到:狭义上的“融合发音”指的是音节和音节之间的“小融合”(音节+音节),广义上的“融合发音”指的是几种不同音节之间的“大融合”(单音节+双音节、双音节+三音节)(丁雷 2015,p195)。在本文中,我们主要以狭义上的“融合发音”为研究目标。为了方便理解狭义上的“融合发音”,我们首先见图1。图1是汉语普通话双音节连读(不含轻声)时的16种²⁾调型组合模式。图中可见,每种组合模式中的实线部分调型是连读过程中各单音节声调的调型,而两条实线中间虚线部分所覆盖的区域就是我们这里要探讨的“融合发音”过程。

图1 汉语普通话双音节连读调型示意图



对于“融合发音”过程的认识,目前还没有专项的研究,在一些汉语语音学专著中也仅在介绍“二字连读变调调型”(比如吴宗济等 1992;曹文 2012)或“汉语连读变调与语调研究”(比如林焘 2010)的时候才有所涉及。其中,吴宗济先生的观点比较具有代表性:连读变调的一个总趋势就是前后两调的升降起伏相连接时,尽量使它过渡得平滑些(吴宗济等 1992,p162)。对于这种“过渡平滑”的作用,吴宗济先生举了具体的例子:

如前调的尾高而后调的头低,则前调的尾被拉下一些;如前调的尾低而后调的头高,则后调的头被压低一些;这种前后两相迁就的流程还带着一些宁低勿高的趋势,如两阳平相连时,后一阳平调竟变成低升;而两去声相连时,前一去声固然成了“半去”,而后一去声也成了低降;至于两阴平相连时,后一阴平也常常变为半高平调(吴宗济等 1992,p162)。从图1上我们不难看出吴宗济先生所说的这种“过渡平滑”使得两调相连更加自然,整体呈现起伏明显的波浪状。

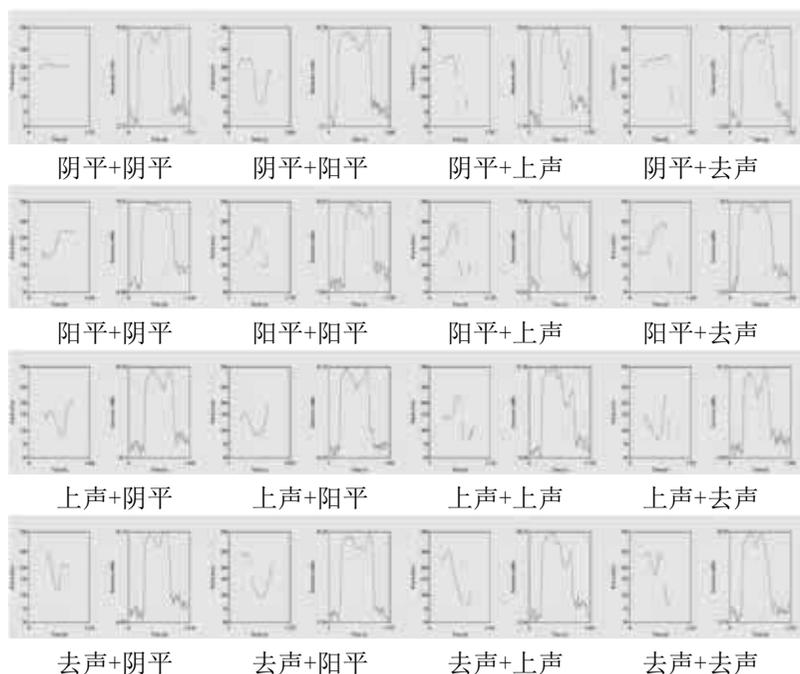
在汉语实验语音学的研究方面,研究者们对吴宗济先生的“过渡平滑”观点有这样的评价:在语音合成研究的初期,我们发现,如果按孤立音节的声调调型来规划语句的音高,听到的语音是和尚念经,平淡无味,但一旦融入这些双音节词调连读时的特点,汉语合成语音的自

然度就有明显的改进（吕士楠等 2012,p107）。研究者们从语音合成的角度再次强调了“融合发音”过程中“过渡平滑”特点的重要性。

3. 我们对于“融合发音”的认识

为了更好地理解目前对于声调间“融合发音”过程的既有认识，我们针对汉语母语者的汉语普通话双音节（最简单的“融合发音”）声调发音做了有关调型模式和强度的测算（见图 2）。图 2 中发音人（男性）的母语是汉语，有过电视台和电台的广播员经历，并且长期从事对日汉语教学工作。我们使用语音信号处理软件 Adobe Audition 和语音分析软件 Praat 绘制了发音人对 /mama/ 组合的 16 种调型模式的调型图（左图）和强度图（右图）。通过本次考察，我们对“融合发音”的过程有了以下几个新的认识。

图 2 汉语普通话双音节连读调型和强度图



3.1 “融合发音”使得前后两调之间的音强产生变化

从图 2 各调型模式的“融合”部来看，前后两调调型在图上直观地表现出波浪形的高低相连调势，在听感上也呈现出自然的效果，符合吴宗济先生的“过渡平滑”的观点。然而我们也注意到：相比较调型的“过渡平滑”而言，“融合发音”导致的音强变化更加具有代表性。由图 2 各调型的强度图（右图）可以发现，“融合发音”使得紧随前调的发音区域产生了一个强度“极速下降”的区域（形状类似“凹”）。“紧随”并不意味着两调之间最低的部分，而是前调之后第一个明显的低点（如图 2 阴平+阳平，紧随阴平后的第一个低点）。为了更加直观，我们将各调型中的“前调强度均值”、“凹点强度”和“后调强度均值”整理成表。

表1 强度比较 (dB)

调型	前调 均值	凹点	后调 均值	调型	前调 均值	凹点	后调 均值
阴平+阴平	75.4	69.9	77.6	上声+阴平	77.1	64.3	64.3
阴平+阳平	77.2	72.7	74.6	上声+阳平	75.2	65.1	65.1
阴平+上声	76.1	72.5	68.3	上声+上声	68.1	73.3	73.3
阴平+去声	78.4	73.0	80.1	上声+去声	78.1	56.6	56.6
阳平+阴平	77.2	71.6	73.8	去声+阴平	79.1	68.4	68.4
阳平+阳平	78.8	72.4	73.5	去声+阳平	76.1	69.5	69.5
阳平+上声	77.2	73.5	68.2	去声+上声	64.5	64.5	71.5
阳平+去声	76.1	70.9	75.1	去声+去声	77.0	77.0	71.4

由表1我们可以发现,这个“凹”下去的“凹点”区域和前调的发音强度的均值相比有了明显的下降。与后调比起来,有时凹点的强度还高于后调的强度均值(如后调为上声时),这主要与后调的调型有关(由表1的观察所见)。相关分析的结果显示,“凹点”强度与前调均值的相关系数为 $0.701, p < 0.05$;与后调均值的相关系数为 $-0.435, p < 0.05$ 。这说明“凹点”强度与前调的均值的相关性更高,受到前调强度变化的影响高于后调。这多少反映出Xu(2005)提出的“并行编码目标逼近模型PENTA”的“随时间轴自左向右逼近”特点(详见丁雷2015,p202)。

“融合发音”使得强度图上出现的明显“凹”形,这个特点为我们使用语音分析软件提取“融合发音”过程带来了方便。在以往的提取过程中,我们往往参考能够直观体现声调变化特征的pitch曲线,但有时候声带振动的不连续性和后字的清声母现象会使得软件常常提取不到基频,得不到pitch曲线。而“融合发音”会使得强度图上出现明显的“凹”形区域,在提取时我们可以以这个“凹点”区域的边界作为“融合发音”过程的起点和终点,结合pitch曲线来进行提取。具体而言,“融合发音”的过程是前调有声部分(前调pitch曲线连续分布部分)的最后一个强度点向后1ms到后调有声部分(后调pitch曲线连续分布部分)的第一个强度点向前1ms的强度点之间的部分。

3.2 “融合发音”使得前后两调的强度产生持续时间上的不同特征

以“凹点”为参照点,我们还可以从强度的持续时间上来进一步观察图2各调型模式的“融合发音”前后,发音所需的能量在时间维度上的具体分配。为了更加直观,我们将各调型中的“前调强度时长”、“后调强度时长”和“近似比”整理成表。

由表2的“近似比”可见,由于“融合发音”过程的存在,强度在前后调的时长分配上产生了一定的分布特点:“前短后长”的占7个,“前后一样”的占5个,“前长后短”的占4个。导致这种时长分配特点的因素,我们认为这主要与前后调的调型有关(由表2的观察所见)。另外,某些声调处于前后不同位置的时候,它自身也会对强度的时长带来一定的影响:如表2中的上声处于“上声+阳平”(1:1)和“阳平+上声”(1:1.5)、“上声+去声”(1:1)和“去声+上声”(1:1.5)不同位置的时候,它会对时长的“近似比”带来一定的影响。

表 2 强度时长比较 (ms)

调型	前调 强度时长	后调 强度时长	调型	前调 强度时长	后调 强度时长
阴平+阴平	273	288	上声+阴平	376	225
近似比	1	1	近似比	1.5	1
阴平+阳平	279	418	上声+阳平	366	361
近似比	1	2	近似比	1	1
阴平+上声	248	371	上声+上声	332	499
近似比	1	1.5	近似比	1	1.5
阴平+去声	306	208	上声+去声	288	253
近似比	1.5	1	近似比	1	1
阳平+阴平	322	178	去声+阴平	283	267
近似比	2	1	近似比	1	1
阳平+阳平	256	380	去声+阳平	247	390
近似比	1	1.5	近似比	1	1.5
阳平+上声	289	406	去声+上声	254	357
近似比	1	1.5	近似比	1	1.5
阳平+去声	343	224	去声+去声	230	241
近似比	1.5	1	近似比	1	1

那么，这种强度在前后调时长分配上的特点说明了什么问题呢？回顾我们以往的语音实验经验来看，从听者的直观感受上来说，强度在前后调时长上的分配往往与声调发音的自然度有很大的关系。有的学生的声调发音虽然没有明显偏误，但是由于前后调在发音时长上的分配“不合理”，导致了其发音自然度的下降。发音虽然正确，但是不自然的情况很多。根据丁雷（2012a）的实验数据来看，“过分拉长阴平和阳平”的例子比较具有代表性。另外，我们还发现学生对于“融合发音”过程的时长控制得不好，“融合发音”过程过长的情况比较普遍，这也使得自然度大打折扣。

我们曾对“融合发音”过程的时长控制和自然度之间的关系做过如下实验：随机选取图 2 发音人的某一个调型，如“阳平+去声”，作为我们的语音合成标本。然后我们使用 Praat 软件分别提取发音人这两个声调“融合发音”部分的强度时长。之后删除原“融合发音”部分，分别加入不同长度的“融合发音”时长（原时长的 1/4、2/4、3/4 和 5/4），合成新的声调发音。使用合成后的声调发音让母语为汉语的 10 名被试（被试均为北京语言大学的学生，均为北京人）听辨，并对所听到的声调发音做出评价“1 自然，2 不自然，3 不好说”。

卡方独立性检验的结果显示，不同长度的“融合发音”过程和被试的评价之间存在显著关联： $\chi^2=26.626, df=6, p < 0.05$ 。具体来看，当“融合发音”长度为原长度的四分之一时，没有人认为“自然”；当“融合发音”时长为原长度的四分之二时，认为“自然”的有 7/10=70%；当“融合发音”时长为原长度的四分之五时，认为“自然”的有 2/10=20%；当

“融合发音”时长为原长度的四分之六时，没有人认为“自然”。由此实验可见，“融合发音”过程的时长阈值分布在原时长四分之一处至四分之六处之间，低于此值或高于此值，都会对自然度带来影响。当然，这只是粗略的估计，精确的阈值还需要更细致的实验统计结果。但这个结果不影响我们的结论。

3.3 “融合发音”使得前后两调边界音高呈现出一定的趋势特征

我们以“凹”形区域：前调有声部分的最后一个强度点向后 1ms（前调边界点）到后调有声部分的第一个强度点向前 1ms（后调边界点）两处为参照点，使用 Praat 的 Time step settings 和 pitch list 功能自动提取前调的起点、中点和边界点，后调的边界点、中点和末点的音高值整理成表。

仔细观察表 3，那么我们会发现“凹”形区域边界音高变化呈现三种情况：左高右低、左低右高、左右相差不大三类。我们将其整理成表。

表 3 “凹”形区域边界音高图 (Hz)

调型	前调			凹形区域	后调			边界差
	起点	中点	边界		边界	中点	末点	
阴平+阴平	217	212	201	融合	205	208	195	-4
阴平+阳平	237	238	237		109	93	189	128
阴平+上声	252	251	258		154	75	102	104
阴平+去声	237	239	246		272	185	81	-26
阳平+阴平	135	131	227	区域	241	237	234	-14
阳平+阳平	142	175	258		169	100	199	89
阳平+上声	140	178	283		197	75	105	86
阳平+去声	143	151	258		288	187	85	-30
上声+阴平	162	112	89	融合	159	200	212	-70
上声+阳平	157	122	88		91	107	184	-3
上声+上声	144	138	247		196	75	101	51
上声+去声	135	107	87		235	120	89	-154
去声+阴平	286	169	110	区域	195	216	190	-85
去声+阳平	295	238	126		104	101	215	22
去声+上声	298	272	150		110	75	111	40
去声+去声	306	250	168		283	259	107	-115

根据表 4 的分布特点，我们将“左高右低”称为“降势”，将“左低右高”称为“升势”，将“左右相差不大”称为“平势”。由此一来，“融合发音”就可以分成三类：“降势融合”、“升势融合”和“平势融合”。当然，发音因人而异，并不是每个人的音高边界差都一定会像图 2 发音人在表 4 中表现出来的那样。在后文介绍的针对汉语母语者的双音节声调组合发音测试中，我们发现一些发音人在“去声+上声”和“阳平+去声”中也会出现左右相差不大的现象；

“左高右低”和“左低右高”也不一定像图2的发音人那样各7组。不过，这种边界差的幅度不同并不影响“降势”“升势”和“平势”的趋势特征，也不影响我们对“融合发音”的分类。

表4 音高的边界差分布特点（按边界差的绝对值从大到小排列）

左高右低	左低右高	左右相差不大
阴平+阳平	上声+去声	阴平+阴平
阴平+上声	去声+去声	上声+阳平
阳平+阳平	去声+阴平	
阳平+上声	上声+阴平	
上声+上声	阳平+去声	
去声+上声	阴平+去声	
去声+阳平	阳平+阴平	

4. 声调的发音能力在“融合发音”过程中的具体体现

在以上针对“融合发音”过程的研究中，我们在传统的“过渡平滑”的基础上，通过实验语音学的技术性分析，有了对“融合发音”过程的一些新认识。那么接下来，我们希望通过进一步的实验，探索发音活动是如何控制和调节“融合发音”过程的。

4.1 理论探索

首先我们需要研究的一个问题是：人究竟是如何发出声调的？我们知道，声调主要是音节中的音高变化，虽然它跟音长、音色也有些关系，不过主导方面还是音高（吴宗济等1992,p137）。而音高，即声音的高低，从声学的观点来看，是由一定时间里音波数或发音体颤动数的多少来区辨的，多的则音高，少的则音低（罗常培、王均1981,p36）。而人类的声带振动就是产生音高的源头，同一个人而发音有高有低，是因为发音时，绷紧声带或放松声带从而导致的声调高低变化。

近年，朱晓农先生的研究对声调发生的过程提出了更加精确的描述：声调发生是指两个过程，一个是音高模式的形成，另一个是这个音高模式的音法化（或称语法化、范畴化）；而对于音高模式的形成，则又分成“音高差异出现”（与发声态有关）和“音高差异的显著化”两个阶段（朱晓农2012,p379）。将吴宗济等（1992），罗常培、王均（1981）和朱晓农（2012）的三个观点结合起来之后，我们认为在声调的发生过程中，人的作用主要体现在“绷紧声带”、“放松声带”以及“连续控制”三个方面。

为什么这样说呢？首先，“绷紧声带”和“放松声带”会产生音高差异的出现。比如我们在理解单音节阴平声调特征的基础上，尝试发不同音高的阴平33（低）、44（中）、55（高）。音高不同的阴平起点是我们对声带绷紧的控制程度不同所导致的结果。从起点来看，高阴平55的5相对于中阴平44的4和低阴平33的3而言，声带的紧张度逐次降低；反之，低阴平33的3相对于中阴平44的4和高阴平55的5而言，声带的紧张度相对逐次增高。声带紧张度的不同导致发音起点的音高不同就是音高差异的出现阶段。

其次，当音高差异出现后，声调的调型要求发音进入到“音高差异的显著化”阶段，这

样一来就必须依靠“连续控制”来实现音高差异的显著化。具体来看这种显著化有三种模式：第一种是平调的“持平”模式，如单音节阴平 55 从起点 5 开始随着发音时间的增加而维持 5 的强度的连续发音。第二种是升调的“升趋势”模式，如单音节阳平 35 从起点 3 开始随着发音时间的增加，声带开始逐步绷紧，音高开始逐步增高，最后到达最高点 5。第三种是降调的“降趋势”模式，如单音节去声 51 从起点 5 开始随着发音时间的增加，声带开始快速放松，音高迅速降低，最后到达最低点 1。单音节上声是“降趋势”和“升趋势”连续作用的结果，不单独构成一个“连续控制”的新模式。

以上探讨使得我们对于“人是如何发出单音节声调的？”这一问题有了一定的了解。那么在声调组合的“融合发音”过程中，人又是如何控制发音的？下面我们通过语音实验来对这两个问题做进一步的探索。

4.2 实验探索

4.2.1 实验设计

本实验的对象是母语为汉语的 14 位在校大学生（首都师范大学中文系，均为北京本地生源）。实验考察的内容是：无意义的双音节组合 /mama/ 的 16 种声调搭配（不含轻声）。实验时，笔者在录音室里采用电脑、投影仪和录音设备（Sony Ecm-PCV80U 话筒、ICD-UX71 录音笔、Adobe Audition 语音信号处理软件），以 PowerPoint 幻灯片的方式在屏幕上呈现实验的内容。和我们之前针对日本学生所做的实验不同，我们对发音人的每一种声调搭配分别设置了 A、B、C、D 四个音高（pitch 值）观测点，见图 3（阴平+阳平，发音人 1 的发音）。

图 3 观测点设置示意图

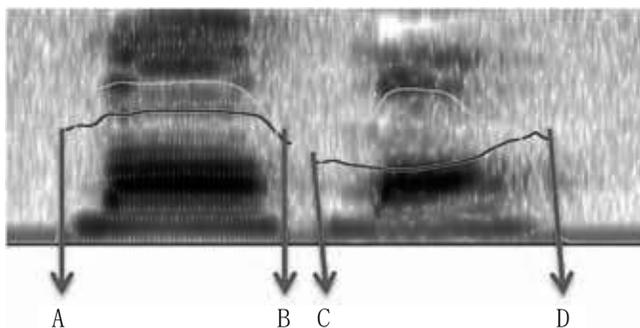


图 3 中，细线是强度曲线，因为能量是连续分布的，所以细线是不会断开的。粗线是音高曲线，由于声带的振动并不一定连续，尤其是在“融合发音”过程中，所以粗线有断开的时候。

我们选择前调音高的起点 A 作为第一个观察点，它的设置目的是为了观测发音人对声带的控制会产生什么样的音高初始值，这个值一般与对音高的控制熟练度有关。我们根据之前对强度曲线中“凹”形区域的认识，将 B 点设置为进入“凹”形区域（也就是“融合发音”过程）之前 1ms 的音高。B 点的设置目的是为了与其后 C 点做音高差对比，B 点和 C 点间的音高差能够印证之前我们提出的三种不同种类的“融合发音”过程。

我们将 C 点设置为“凹”形区域结束后 1ms 的音高。C 点设置的也有两个，第一个

之前提过的与 B 点音高作对比，从而观察“融合发音”过程内部的音高变化趋势；另一个目的是为了与 A 点做对比。A 点和 C 点一个是前调的起点音高，另一个是后调的起点音高。C 点的发音是不是也像 A 点那样，仅仅只是受到发音人对声带控制的影响？还有没有其他的影响因素？这是我们了解“融合发音”对后调的影响，以及研究“拼调能力”的关键点。我们将 D 点设置为后音节音高结束的位置，在本次实验中，D 点只作为参考。我们选取其中一位发音人的四点统计表，整理如下：

表 5 发音人 1 的四点统计 (Hz)

发音人1	前调		后调	
	A	B	C	D
阴平+阴平	300	315	322	313
阴平+阳平	304	300	221	262
阴平+上声	300	330	256	197
阴平+去声	299	296	354	204
阳平+阴平	249	266	306	300
阳平+阳平	237	272	223	283
阳平+上声	241	301	258	189
阳平+去声	232	266	344	210
上声+阴平	230	170	281	278
上声+阳平	219	163	179	280
上声+上声	242	300	259	188
上声+去声	232	182	279	211
去声+阴平	321	189	265	284
去声+阳平	357	194	200	259
去声+上声	371	191	181	192
去声+去声	365	202	364	195

4.2.2 讨论

通过本次实验，我们在数据分析和整理的基础上，对 14 位汉语母语者的双音节声调组合发音有了以下一些认识：

(1) 纵观整个实验过程，正如我们在对图 2 发音人的测试中所表现出来的一样：“融合发音”过程使得前后两调之间的音强产生了明显变化，出现了“凹”形区域。

(2) 正是因为“凹”形区域的存在导致音强的不连续分布，使得“融合发音”在客观上造成了前后调发音时长分配上的不均衡。当然，对于这种“不均衡”的理解，我们只是强调有差异。在本次实验中我们发现，发音人的前后调发音时长近似比并不都像图 2 发音人那样形成“前短后长”、“前后一样”和“前长后短”的 7:5:4 这样的比例。实验中，我们观察到时长近似比分布并没有形成固定的特征。另外，在实际的语流输出中声母和韵母的自身的

时长以及语调也会对音节整体的发音时长带来一定的影响，所以目前我们还不能对这种时长分配上的不均衡做更进一步的类别化处理。

(3) 在“融合发音”对前后两调边界值的影响度方面，由B点和C点的观察可知（参考表5），B点和C点的音高差异呈现出之前我们提到的“升势”“降势”和“平势”。因而我们之前对于“融合发音”的“降势融合”、“升势融合”和“平势融合”的分类也证明是站得住脚的。但是“融合发音”过程中的“升势”“降势”“平势”和之前我们提到的声调发音时“连续控制”所导致的“升趋势”“降趋势”“持平”是什么样的关系？实验数据本身给不出这样的答案。我们目前倾向于将这两者区别对待。理由有三个。第一，“融合发音”是调间音高变化，主要为了调节前后调，达到“过渡平滑”的目的；与此相对，“连续控制”是调中音高变化，主要为了实现“音高差异的显著化”。两者功能不同，所起的作用也不一样。第二，“连续控制”伴随单音节声调的出现而出现，其出现次数等于单音节声调的出现次数。而“融合发音”是伴随音节间发音的出现而出现的，其出现次数少于音节的出现数，即 $\boxed{\text{“融合发音”次数} = \text{音节数} - 1}$ 。

图4 A点和C点的同调型音高差异（离散度）

阴平	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A点SD	2.2	10.3	11.1	3.4	8.8	1.5	5.0	2.9	1.7	1.3	3.9	4.7	4.5	2.4
C点SD	25.4	28.7	33.7	11.2	21.1	13.1	13.8	9.1	8.9	10.1	9.1	9.8	12.3	13.8

阳平	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A点SD	7.2	7.7	3.7	5.0	7.1	3.3	7.0	5.6	3.3	2.6	4.7	7.1	1.7	4.7
C点SD	20.6	15.1	11.1	11.9	12.3	9.2	12.2	15.8	8.1	12.3	17.1	17.5	10.9	11.1

上声	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A点SD	7.0	11.3	6.1	2.0	9.1	13.1	5.9	5.2	4.0	3.5	4.9	4.0	2.1	2.8
C点SD	38.4	22.2	24.6	9.1	15.7	30.8	37.6	13.2	12.4	16.7	11.0	19.1	15.0	12.3

去声	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A点SD	18.4	9.0	7.1	6.0	11.4	3.0	4.6	6.2	4.3	4.9	3.5	4.6	4.4	4.6
C点SD	38.4	15.4	34.5	11.2	28.5	22.5	15.3	11.7	18.3	15.5	9.4	17.5	12.6	22.6

(4) 在发音能力如何影响“融合发音”这方面，我们通过A点和C点的同调型音高差异（离散度）比较来进行探讨，见图4。由图4内各表我们不难看出，对于同一个调型而言，C点音高的分布比A点的范围更广（离散度更大）。同一个发音人，又是同一个调型，为什么起点的音高值会出现这样分布上的差异呢？对比A点和C点，我们认为发音人对于A点的音高控制主要在发声态，也就是对声带的紧张度控制方面。而对于C点音高而言，除了对声带的紧张度控制以外，“融合发音”过程本身也会对C点的音高带来影响。尤其是在发音过程中具有一定升力或降力的“升势”和“降势”。我们认为这些或许都是“拉高”或者“压低”C点音高分布范围的发音因素。C点音高离散度所表现出来的不同于A点的发音特征，或许一定程度上也反映出了“拼调能力”的特征。

图 5 双音节声调组合的发音过程示意图

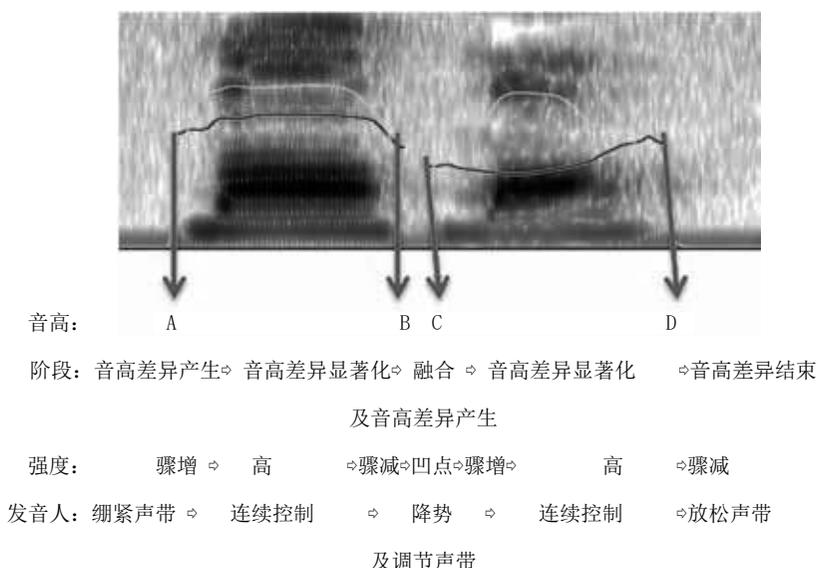
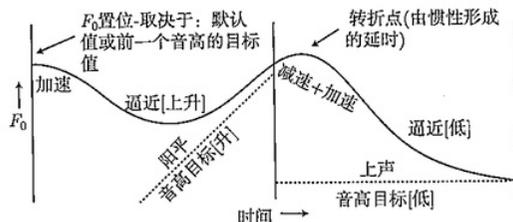


图 6 PENTA 模型的目标逼近原理图

(转引 Prom-on et al.,2009 见吕士楠等 2012,p102)



(5) 我们通过对图 2 发音人以及 14 位汉语为母语的发音人所做的双音节声调发音组合实验, 对含有“融合过程”的双音节声调组合的发音过程有了一定的了解。下面我们将这部分的研究结果, 以双音节声调组合的发音过程示意图的方式展现出来, 见图 5。同时, 我们将韵律合成模型: PENTA (parallel encoding and target approximation) 的声调目标逼近原理图(见图 6, 图 6 是一个阳平+上声的示意图)与我们的图 5 结合起来分析。

PENTA 模型给予前调 F_0 一个起始值(也就是我们在图 5 中的 A 点), 这个起始值被设定为从前面音节传递下来的值。PENTA 认为如果是话语的开始, 可以设定一个初始值。而我们认为这个初始值的设定与发音人控制声带的紧张度有关, 这个值可以通过对发音人的单音节声调的初始值研究得到(本文不做展开)。而在实际的语流中, 我们认为这个起始值与 D 点的音高关系密切, D 点的音高值或许是研究多音节组合时不可忽略的一个要素, 这将作为今后研究的一个主攻方向。另外, PENTA 在图 6 两调之间呈现出一个由前调阳平末端音高抬升导致的“减速+加速”发音过程, 它认为这是一个由惯性形成的“过冲”阶段。这从音高

的速度特征上反映出了“升势融合”的特征。“惯性”的产生是因为音高目标由阳平的“升”转为上声的“降”所导致，我们认为这个“惯性”的解释是可取的。PENTA 更给出了这个惯性的计算公式： $\bar{X}(t)=mt+b$ 。其中 b 和 m 分别是应高目标的起始高度和音高曲线的斜率（吕士楠等 2012,p102）。

4.2.3 小结

在双音节这个最简单的声调组合中，汉语母语者的发音能力（发音模式）具体体现为对“绷紧声带→连续控制→融合运动及调节声带→连续控制”这一过程的控制力。如果控制力不够成熟，在任何一个环节，一个细小的失误都会使得整个发音链出现“走调”现象。这一点在初学者身上体现得更加淋漓尽致。

从我们以往对日本大学生所进行的声调实验（丁雷 2012a）来看，由于对声带的音高控制不成熟，导致前调音高起始值容易出现异值。比如前音节阴平发成“低平”就是最有代表性的。另外，对音高“连续控制”的不熟练，尤其是“升趋势”和“降趋势”的不明显，常常会导致学生发出“高不上来，低不下去”的“尴尬调”。比如前音节阳平和去声都发成“平调”是比较具有代表性的。而最严重的问题往往出现在“融合过程”后的 C 点的音高控制上。由于 C 点自身的音高是“融合运动”和声带调节后的共同产物（拼调的结果，惯性的产物），所以对发音技巧的要求特别高，出现偏误的几率也就比较大。比如有的把后音节阴平发成“低平”，有的把后音节阳平发成“高升”，有的把后音节去声发成“短快”就是比较有代表性的例子。

5. 对话流中“融合发音”的初步探索

5.1 研究初衷

我们认为针对日本大学生的汉语语音研究，尤其是声调研究一定要区分“偏误”和“口音”。偏误是口音的低级状态，偏误研究可以描述偏误分布、归纳偏误种类、研究偏误倾向，但是不能完全依靠偏误研究的成果对学生的汉语声调习得下总结性的结论。在汉语学习的初级阶段，发音中的确存在大量偏误。但是如果看看中级甚至高级阶段学生的声调发音就会发现“偏误”骤然剧减，取而代之的是不同程度的“口音”（“洋强洋调”）。下面我们通过语音实验，从促成“口音”产生的要素方面再来探讨一下语流中的“融合发音”。

5.2 实验探索

5.2.1 被试

参加本次实验的被试为中级水平的汉语学习者 12 人（岛根大学汉语辅修专业的学生，实际课时数约为 218 小时）。

5.2.2 实验内容

录音材料为 13 个不同长度的句子，具体见图 7。让被试朗读录音材料的同时，通过耳机向被试呈现听觉信号从而控制发音时长。开始是两个敲击音，最后是一个时长为 100 毫秒、频率为 500Hz 的纯音，后一个信号音和纯音之间的间隔为 2000 毫秒。要求被试在听到第一个敲击声的时候做准备，听到第二个敲击声立即开始朗读，尽量在听到纯音时能恰好读完整整个句子。正式录音之前被试被要求进行多次没有旁人干扰的练习，正式录制时每个句子重复

3次。录音是在录音教室完成，通过语音信号处理软件 Adobe Audition 直接录制在电脑上，采样频率为 44.1kHz，16bit 的立体声文件。

5.2.3 实验结果及讨论

我们将实验结果总结成图 7。通过实验，我们对语流中的“融合发音”与“口音”之间的关系有了一定的认识：

图 7 语流中的口音感位置图

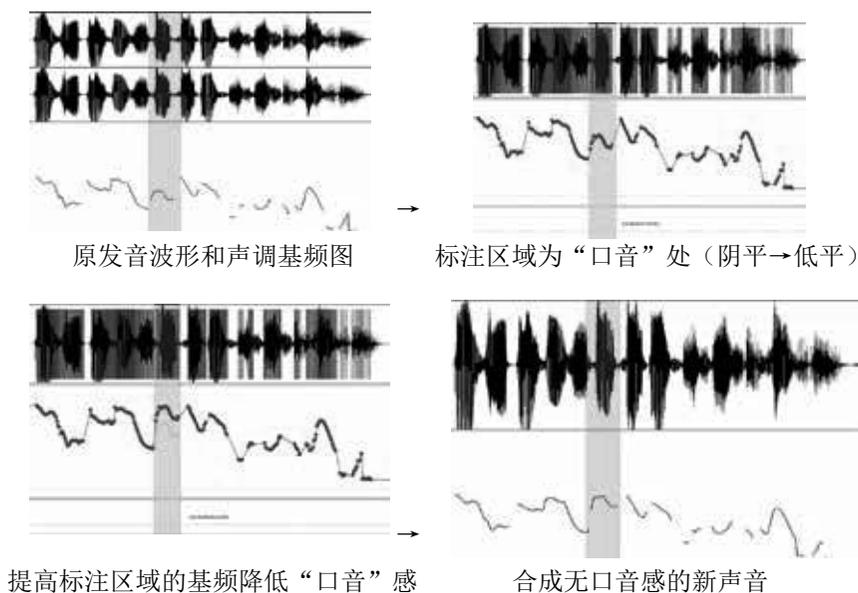
口音感的位置	具体感受
田 _{Δ1} 家中家有五 _{Δ2} 口人	Δ1: 阳平变平调 Δ2: 升调 (变调后) 变平调
他爸爸妈 _{Δ1} 妈一 _{Δ2} 个姐 _{Δ3} 姐一个 _{Δ4} 哥哥和 _{Δ5} 他	Δ1: 阴平变降调 Δ2 Δ4 Δ5: 阳平变平调 Δ3: 上声变平调或升调
他姐 _{Δ1} 姐前 _{Δ2} 年大 _{Δ3} 学毕 _{Δ4} 业了	Δ1: 上声变平调或升调 Δ2: 升调 (变调后) 变平调 Δ3: 去声变平调 Δ4: 去声变升调
现在在一 _{Δ1} 家贸易公司工作	Δ1: 去声变升调
他哥 _{Δ1} 哥现在是在大 _{Δ2} 学三年级 _{Δ3} 的学生	Δ1: 阴平变升调 Δ2: 去声变平调 Δ3: 上声变平调
田 _{Δ1} 中比 _{Δ2} 他哥哥小 _{Δ3} 两 _{Δ4} 岁	Δ1: 阳平变平调 Δ2: 上声变平调 Δ3: 上声下降幅度不够低 Δ4: 上声变升调
他现在是大大 _{Δ1} 学一年级 _{Δ2} 的学生	Δ1: 去声变平调 Δ2: 去声变平调
他和他哥哥都在大大 _{Δ1} 学学习汉语 _{Δ2}	Δ1: 去声变平调 Δ2: 上声变升调
他哥哥很 _{Δ1} 喜 _{Δ2} 欢汉语	Δ1: 升调 (变调后) 变平调 Δ2: 上声变平调
他汉语 _{Δ1} 说得比田 _{Δ2} 中 _{Δ3} 好	Δ1: 上声变升调 Δ2: 阳平变平调 Δ3: 阳平变平调
他希 _{Δ1} 望大 _{Δ2} 学毕 _{Δ3} 业后到中国去工作	Δ1: 阴平变低平 Δ2: 去声变平调 Δ3: 去声变升调
田 _{Δ1} 中也很 _{Δ2} 喜 _{Δ3} 欢汉语	Δ1: 阳平变平调 Δ2: 升调 (变调后) 变平调 Δ3: 上声变平调
他希 _{Δ1} 望明 _{Δ2} 年夏 _{Δ3} 天能到中国去留学	Δ1: 阴平变低平 Δ2: 阳平变平调 Δ3: 去声变平调

(1) 由图 7 我们可以观察到在多音节“融合发音”的过程中，“趋平”和“趋升”是“口音感”比较强的两种表现。其中“趋平”（发成平调）是最有代表性的。我们在前文的研究中提到，发音人对声调的作用体现在“绷紧”或“放松”声带和“连续控制”方面。我们认为“趋平”的问题主要产生在“连续控制”上，正是由于发音人对音高“连续控制”不熟练（也可能是来自母语的某种趋平意识所致，丁雷 2015 中对这点有所涉及），尤其是“升趋势”和“降趋势”不明显，因而使得升调升幅不够（如阳平→平调），降调降幅不够（如半上或去声→平调），在听感上就容易让人产生“平”的直观感受。而“趋升”的原因主要体现在变调控制方面，本次实验中比较有代表性的是上声的连续变调（如很喜欢）和“一”的变调上。

(2) 实验中我们也发现重复出现的某些“口音字”往往也会使得“口音感”在发音过程中持续延长。如“田”“大”“姐”“一”等字因为后文也会出现，一旦前文读错，这个错音

就会一直带入后文并影响后面的发音质量。相对于单音节和双音节而言，由多音节组合起来的“融合发音”中，这种有点儿“遗传”味道的“口音”现象比较明显。对于这种问题，我们一般的做法是在读错处停下，改正好“口音字”的发音之后再让学生继续读。但是这样做依旧还是不直观，学生只是知道自己的发音跟老师不一样，具体哪里不一样？如何才能跟老师的发音一样？这始终是我们纠正语流发音问题时的短板。因此我们在这里考虑使用 Manipulation 语音合成技术，修改学生原有的发音错误，用学生自己的声音合成出没有“口音感”的发音，详见图 8。这样做最直接的效果是建立起了学生发音时的自信心，让自己的发音和标准发音之间不在遥不可及。这项技术目前还没有应用于汉语语音教学的先例，其教学应用前景还有待我们的后续研究。

图 8 Manipulation 的处理过程示意图



6. 结语

本文是丁雷（2015）的后续研究，主要课题还是探讨日本大学生汉语声调的发音能力。文中我们从声调与声调之间的“融合发音”问题谈起，通过语音实验探索了双音节声调组合中的“融合发音”过程，也搞清了“融合发音”对音强、时长以及前后调之间的音高边界值所带来的影响，并给“融合发音”进行了分类。同时，我们还通过对汉语母语者的双音节声调组合测试，探讨了人在汉语声调发音过程中能够发挥的主要作用、构拟了发音模式。并从这个角度上尝试解释了导致日本大学生声调发音出现偏误的主要原因。

另外，我们又通过对语流中“融合发音”的观察，探讨了学生在语流“融合发音”过程中容易出现的发音偏误倾向。并且尝试采用最新的语音合成技术向学生展现声调偏误的具体细节和产生过程，辅助纠音教学。我们一直主张：声调习得研究是“能力研究”不是“现象研究”。语音实验研究不应该只是记录和观测偏误现象，还应该利用语音技术将人无法直接观测到的

“能力成长过程”构拟出来，让语言的习得过程变得更加通俗易懂。今后我们还要通过更新的技术手段（比如3D虚拟现实技术和发声过程的IMAX技术）进一步探讨有关人类声调发音能力的课题，希望本文的一些小成果能够对今后的汉语声调教学带来些许帮助。

注

- 1) 乔姆斯基在《句法理论的若干问题》(Aspect of the Theory of Syntax,1965)的研究中发现，尽管儿童智力还不十分发达，学习诸如数学、物理等方面的知识还相当困难，但他们却能在五六岁时很容易地掌握了自己的母语。乔姆斯基声称儿童天生具有一个“语言获得机制LAD”，尽管语言的规则非常抽象，但由于有了这样一个“语言获得机制”，儿童只要稍许接触语言材料就能在很短的时间内掌握自己的母语。
- 2) 我们知道上声与上声相连时，传统上视前一个上声与阳平同调，将上声相连处理为与阳平+上声同一调型，这使得普通话双音节连读只有15种调型。近年，张本楠(2012)的研究指出，上声连调并不等同于阳平+上声，而应该是“后半上+上声”。笔者也比较接受这个观点，所以双音节连读调型模式还应该是16种。不过这个观点与本文的讨论范围无关，特此补充说明。

参考文献

日本語文献

1. 丁雷 2012a. 『日本人学習者の中国語の声調誤用の分析と指導方法について 日本の大学における第2外国語としての中国語教育を例にして』: 1-145页，广岛大学博士论文

中国語文献

1. 曹文 2010. 『现代汉语语音问答』: 30-47页，北京：北京大学出版社
2. 丁雷 2012b. 谈谈汉语教学种如何看待汉语单音节声调与双音节声调的关系，『島根大学外国語教育センタージャーナル』第8号：69-76页，島根大学
3. 丁雷 2013. 对日本大学生汉语声调发音能力的探讨，『第十一届国际汉语教学学术研讨会论文集』: 48-54页，四川：四川出版集团巴蜀书社
4. 丁雷 2015. 对日本大学生汉语发音中“声调难”问题的探索 - 以初修汉语的教学对象为例 -，『中国語教育』第13号：186-204页，中国语教育学会
5. 林焘 2010. 『中国语音学史』: 533-549页，北京：语文出版社
6. 吕士楠、初敏、许洁萍、贺琳 2012. 『汉语语音合成 - 原理和技术』: 86-125页，北京：科学出版社
7. 罗常培、王均 1981. 『普通语音学纲要』: 132-170页，北京：商务印书馆
8. 吴宗济、赵金铭、朱竹、刘骥 1992. 『现代汉语语音概要』: 162-164页，北京：华语教学出版社
9. 朱晓农 2013. 『语音学』: 272-297页，北京：商务印书馆