

论言语发音与感知的互动机制

陈忠敏

(复旦大学, 上海 200433)

摘 要: 语言交际是一个说者发出语流串,经媒介传入听者听觉器官,再由听者大脑神经中枢重新解码的过程。语言交际涉及说和听双方。本文从听与说两方面来深入人类语言交际的机制及其特点,通过讨论发音器官与听音器官频率声能互补、低频敏感度与元音格局、音类扩散分布与感知区别增强、量子理论与语音感知、说者协同发音与听者感知补偿等五方面内容,提出说与听有相互配合、同时也有相互制约的关系,同时说明一旦这种关系失衡,就会误解,可能导致音变。这一研究成果也将对人工智能、类脑研究有所启发。

关键词: 语音发音; 语音感知; 量子理论; 协同发音; 感知补偿

On the Interactive Mechanism Between Speech Pronunciation and Perception

CHEN Zhongmin

(Fudan University, Shanghai 200433, China)

Abstract: The process of communication by means of speech involves the generation of sound by a speaker and interpretation of that sound by a listener. So language communication involves both speaking and listening. This paper explores the mechanism and characteristics of human language communication from the perspective of listening and speaking. The author illustrates the relationship between listening and speaking through the following five aspects: 1) the complementary distribution of sound energy with different frequencies between articulation organs and auditory organs; 2) low-frequency sensitivity and vowel pattern; 3) the adaptive design of speech sound systems and enhancement of perceptual differences; 4) quantal theory and speech perception; 5) speaker's co-articulation and listener's compensation. It is proposed that speaking and listening have mutual cooperation and mutual restriction in the process of communication. Once this relationship is out of balance, it can lead to misunderstandings and sound changes may occur.

Key words: speech production; speech perception; quantal theory of speech sound; co-articulation; perceptual compensation

1. 前言

语言交际涉及到说与听两方面。从说者的角度来看,说者首先要在言语认知的高级中枢——大脑运动皮层发出发音的运动命令、规划,然后主要运动皮层(primary motor cortex)再发出指令引起神经元(neuron)放电,相关神经元协调和牵动各种肌肉驱动发音体发声,共鸣体改变发出不同的音来。从传播的角度看,随着发音体与共鸣腔内的空气粒子压力发生变化,产生不同性质的声波,声波通过媒介传到听者的耳朵里。从听者的角度来看,声波传送到听者耳朵后,听音器官对外来的语音会有选择性地放大、缩小来提取声学信号再转换成神经脉冲放电,生理电信号传入大脑中

枢来重新解码,从而完成一次语言交流过程。发音、语音传播、语音解码三个阶段形成语言链中三个不同的阶段。在语音学里对应于这三个阶段分别形成三个不同的语音学分支:发音语音学(articulatory phonetics)、声学语音学(acoustic phonetics)、听觉/感知语音学(auditory phonetics/speech perception)。

语言的发音、传播、感知虽然根据语言链的前后关系分成三个不同阶段,但是它们是相互配合和制约的。本文从听与说两方面来深究人类语言交际的机制及其特点,提出说与听有相互配合、同时也相互制约的关系,一旦这种关系失衡,就会误解,再从听者发出此类音,就有可能产生音变。

2. 发音器官与听音器官频率声能互补

从发音器官来研究言语的产生,可以用“声源+共鸣滤波器”理论(source + filter)来说明。此理论由著名语音学家 Gunnar Fant 在 20 世纪 60 年代初提出(Fant 1960)。肺部出来的气流是声能,所以肺部和声门以下的器官以及里面的气流是言语发声的动力部分。气流通过两片靠拢的声带会产生布鲁力效应(Bernoulli effect)使得声带相向运动而引起振动,从而产生准周期性的脉冲(impulses)声源,声带振动的频率就是基频(Fundamental Frequency),通常写作 F_0 。另一种声源是白噪音声源(white noise),白噪音声源产生于共鸣腔内的某收紧点,收紧点变窄,急速的气流通过产生湍流般噪音,或者收紧点闭塞后突然爆破,产生爆破声源。噪音声源与声带发出的准周期性声源可以叠加,比如浊擦音;也可以分开,比如元音只有准周期声源,清辅音只有噪音声源。声源前的共鸣腔可以假设为一种特定形状的共鸣滤波器(vocal tract filter)。声源发出的声音经过特定的滤波器的修饰,再经过唇辐射放大就是我们人耳听到的语音。图 2.01 是元音产生的声源+共鸣示意图。

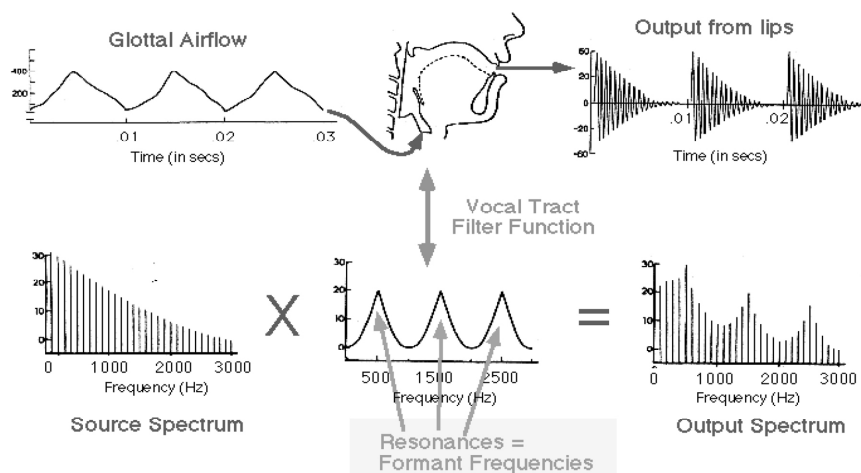


图 2.01 元音产生的声源+共鸣示意图

图 2.01 中的上图声门气流(Glottal airflow)是声门处声带振动引起的气流声波(声门波)。声门波经过共鸣腔再从嘴唇出来(output from lips)就是我们听到的声波。下图的声源谱(source spectrum)是指声门波对应的瞬间频谱,瞬间频谱的纵轴代表振幅,单位是分贝(dB),横轴是频率,单位是赫兹(Hz)。从下图的瞬间频谱图可以看出,随着频率升高,振幅急剧下降,一般来说,正常发声态(modal voice)频率每下降一个倍频程,振幅就下降 12dB。不同形状的共鸣腔具有自己特定的共鸣频率(resonances)。图 2.01 中所示三个共振峰频率(formant frequencies)是相当于均匀共

鸣腔管子,类似于发元音/ə/时的数值,第一共振峰(F1)为500赫兹,第二共振峰为1500赫兹,第三共振峰是2500赫兹。声门频谱经过共鸣频率的修饰,就产生了我们能从分析仪看到的元音共振峰的输出频谱图(output spectrum)。频谱图里能量总的特点是低频能量高,越往高频,能量越低,这个特点是发音体的生理机制所决定的。

从感知器官来看,语音感知的特性跟人类特有的听觉器官的特点有密切的关系。人耳听辨声音的频率范围虽然很广阔,覆盖16赫兹到22 000赫兹的频率范围,但是人耳频率分辨率并非线性的,对频率也是有选择性的。成人外耳道的长度大概是2.3cm(0.023m),根据一端闭一端开的管子的共振频率我们可以根据声速(每秒340米)与波长(一端开一端闭的管子是最长波长的四分之一)公式算出这一长度管子的最强的第一共振峰值: $340 / (4 \times 0.023) = 3696$ 赫兹,再加上中耳带宽的扩大效应(带宽大概是500赫兹到5000赫兹)(Rosen & Howell 2011: 267),形成低频有坡度,高频陡峭的敏感带宽区域,这种图形叫作往低延伸的带宽图(downward spread of bandwidth)。图2.02是外耳、中耳以及两者叠架声音敏感图(Rosen & Howell 2011: 268)。可以看出人耳对2 000赫兹到5 000赫兹的声音反应最为敏感,到达或超过10 000赫兹的声音敏感度会急剧下降(Johnson 2012: 86)。

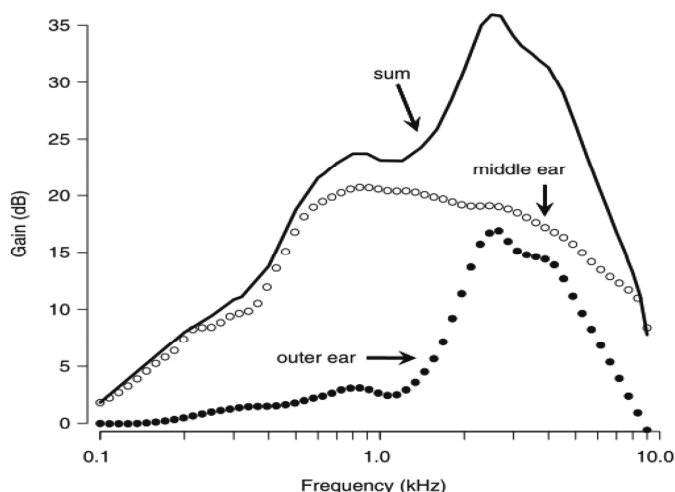


图 2.02 外耳、中耳以及两者叠架声音敏感图

可是在语言里2 000赫兹以下的低频对语言来讲也是非常重要的,不同元音的区别主要靠第一、第二共振峰的不同,第一共振峰一般在1 000赫兹以下,很多元音的第二共振峰也不到2 000赫兹(见图3.02、3.03),如果2 000赫兹以及低于2 000赫兹的声音敏感度不高,将大大损害言语交际。低频区如果要达到2 000赫兹到5 000赫兹一样的敏感度,就要增加低频区的音量。在图2.01里我们知道声门的瞬间频谱图的振幅是低频处高,越往高频,振幅越低,通过共鸣腔修饰的语音也具有这个特征,输出的第一、第二共振峰的音量远远高出第三、第四、第五等共振峰的音量,发音时具有的低频音量高,高频音量低的特点,正好拉高了人耳接受低频声音的敏感度,以保证人耳20赫兹到5 000赫兹范围内的声音都具有最佳敏感度。这是人类发音器官与听音器官相互作用,互相补充,保证语言交流有效的经典例子。5 000赫兹以下的敏感度对人类言语交际有极为重要的意义。因为人类语音的最重要音征都集中在低频处。比如最低的三个共振峰(简称F1、F2、F3)频率决定元音的音色,一般都在5 000赫兹以下;声调语言里男女声调的频率(基频)一般在20赫兹至400赫兹范围内变化;语言中塞音/k/爆破点(burst)能量集中区在2 000-4 000赫兹左右;/p/和/t/的

区别是爆破点以 3 000 赫兹左右为界,前者是能量往下降,后者是上升的;某些辅音的能量集中区虽然处在较高的频率段里,比如清辅音 s,但是没有一种语言凭借 6 000 赫兹以上的能量区别不同擦音。

3. 低频敏感度与元音格局

人类对低频的敏感不仅体现在强度的感知上,还体现在音类分辨精度上。这种听辨的特性直接制约着人类语言语音的发声及语音格局、音变的方向。成人内耳耳蜗(cochlea) 大约长 35 毫米,中间的基底膜(basilar membrane) 约长 31 毫米,基底膜从卵形窗(oval window) 到蜗孔尖端(apex) 分布着 23 500 多个听觉毛细胞,以此来感知从高到低不同频段的声音,不过不同频段的感知精度在基底膜上并非线性的,低频感知的带宽(bandwidth) 窄,分析声音精密;高频感知的带宽阔,分析声音粗犷。图 3.01 是内耳基底膜伸张时声音感知频率分布图(取自 Johnson 2012: 89 Figure 4.5 (b))

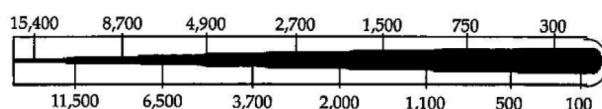


图 3.01 内耳基底膜伸张时声音感知频率分布图

粗段处是靠近蜗孔尖端部分,对低频起反应;细段处是近卵形窗部分,对高频起反应。低频段往高频段等距离所对应的频率反应范围是不同的,相同的间距,低频处 $300 - 100 = 200$ 赫兹,到了高频处是 $15\,400 - 11\,500 = 3\,900$ 赫兹。换句话说低频处的带宽分辨率细,高频处带宽分辨率粗。比如 1 000 赫兹处的有效感知带宽是 130 赫兹,而 5 000 赫兹处的感知带宽是 650 赫兹,所以人耳对低频端的声音分辨率高,高频段声音则分辨率低。感知频率的这种特点,直接影响着人类语言的发音特点与语音格局。图 3.02、3.03 分别是美式英语(Peterson & Barney 1952) 和普通话(据吴宗济、林茂灿主编、鲍怀翘、林茂灿增订 2014: 113 的数据制作) 单元音共振峰舌位图:

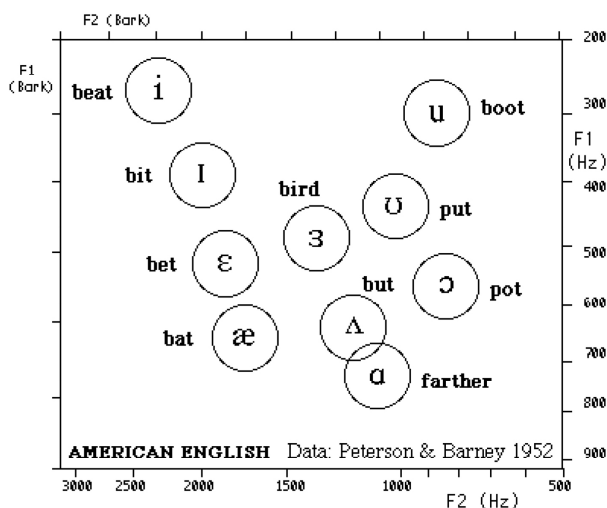


图 3.02 美式英语声学元音图

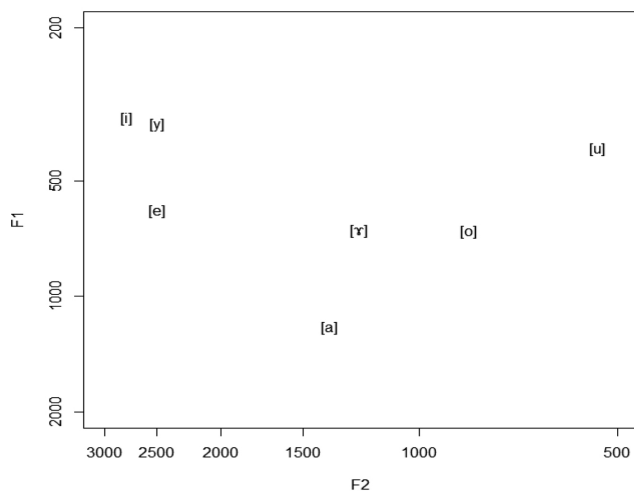


图 3.03 普通话声学元音图

图 3.02、3.03 里 y 轴是发音时共鸣腔最低的共振峰(第一共振峰,简称 F1)的刻度,第一共振峰的数值跟发音时舌位的高低有关,一般来说数值越高,舌位越低。x 轴则是共鸣腔次低的共振峰(第二共振峰,简称 F2)的刻度,第二共振峰的数值跟舌位的前后有关,一般来说数值越高,舌位越前。从赫兹数来看,y 轴上的每个刻度的间距都是 100 赫兹,但是仔细比较,可以看出 200 赫兹到 300 赫兹的间距与 800 赫兹到 900 赫兹并非等距离,前者的间距大,后者的间距小;y 轴上虽然每个刻度都相差 100 赫兹,但是刻度的间距是非等距的,赫兹数值越大,刻度的间距越小。这一点在 x 轴上更为明显,同样相差 500 赫兹,1000 至 1500 赫兹的间距远远大于 2500 到 3000 赫兹的间距。再比较 y 轴与 x 轴的赫兹数值,y 轴上 800 至 900 赫兹的间距与 x 轴上 2500 至 3000 赫兹的间距大致相当,但是前者这一间距的只相差 100 赫兹,后者同样的间距则要相差 500 赫兹。低频处分辨精细,高频处分辨相对粗狂,正好也能说明人类语言里元音舌位高低的层级分辨要比舌位前后的层级分辨来得重要。表现在语言的元音舌位图里就是元音舌位的高低分布数量要多于元音舌位的前后分布数量。图 3.04 是世界多数语言元音舌位分布图(Liljencrants & Lindblom 1972),可以看出除了三元音构建成元音舌位倒三角外,其他五元音、七元音的,舌位高低的层级数都要多于舌位前后的层级数。

| | | | | | |
|---------------------------------------|---|--|---|-------------------------------------|---|
| i | u | i | u | i | u |
| | | e | o | e | o |
| | | ɛ | ɔ | | |
| a | | a | | a | |
| Arabic, Nyangumata, Aleut, etc. | | Spanish, Swahili, Cherokee, etc. | | Italian, Yoruba, Tunica, etc. | |

图 3.04 不同语言元音音位分布图

4. 音类扩散分布与感知区别增强

元音舌位图还有一个特点就是世界多数语言里前元音更多的是不圆唇,后元音更多的是圆唇。表 3.04 所列的三种类型元音舌位图里,除了最低的元音外,无论是三元音系统、五元音系统,还是七元音系统,前元音都是不圆唇,后元音都是圆唇。在 Maddieson(1984)所调查的语言里,94%的前元音是不圆唇的,93.5%的后元音是圆唇的。从发音的角度来看,舌的前伸后缩和唇的圆唇不

圆唇是两个相对独立的发音机制 相互间没有牵制关系 ,也即舌位往前 ,双唇不必展开 ,舌位往后 ,双唇也不必一定要圆唇。但是前元音不圆唇、后元音圆唇则是最大限度拉大了音类的声学距离 (此处是指第二共振峰 F2 的声学数据) 。如图 4.01 所示:

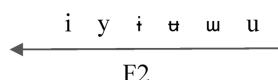


图 4.01 元音第二共振峰 F2 的分布图

尽管语音声学信号与语音感知的关系错综复杂 ,但是不可否认语音感知的主要依据还是声学信号 ,换句话说尽量大的声学区别 ,对听者来说是增强了音类间的区别感 ,音类间的区别清晰则更有利于言语交际。Liljencrants 和 Lindblom 把这种元音分布的特点叫做元音适应性扩散分布理论 (Theory of Adaptive Dispersion) 。其中心思想就是一个语言里的音位与音位之间的距离应该尽量拉开 ,分布要扩散。拉大音类之间的距离也就提高了言语交际的清晰度 ,使得言语交际不模糊 ,而扩散的距离必须要适应这种语言的音系格局 (Liljencrants & Lindblom 1972) 。所以一个语言不管是 3 元音、4 元音 ,还是 5 元音、6 元音 ,甚至 7 元音系统 ,前后元音最重要的声学区别是 F2 的差异 ,前元音 F2 大 ,后元音 F2 小 ,前元音伴随不圆唇特征和后元音伴随圆唇特征进一步拉大了 F2 的差异 ,使得前后元音的对立更为突显 ,所以前元音不圆唇特征和后元音圆唇特征并非主要的区别特征 ,而是一种伴随特征 ,这种伴随特征起到了增强区别性的作用。这是发音和声学迁就感知的例子。

与感知有关的 ,在发音上唇和舌有互动作用的另外一个例子就是卷舌辅音。无论在汉语里还是在其他语言里 ,发卷舌音往往会伴随双唇圆唇或突出 (protrusion) 动作。图 4.02 是美式英语发 /r/ 时的情形 (取自 Gick , Wilson , & Derrick 2013: 213 Figure 11.2)

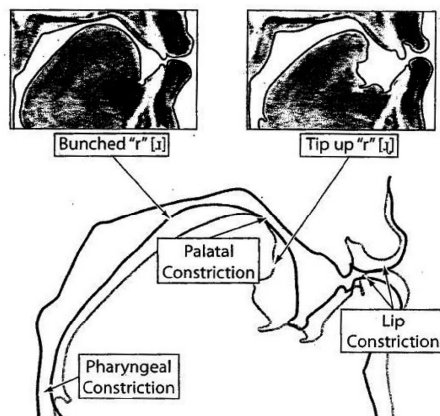


图 4.02 美式英语发 /r/ 时三个收紧点

从图 4.02 可以看出尽管卷舌的舌尖位置有变异 ,但是三个收紧点是明显的: 双唇 (lip constriction) 、硬腭处 (palatal constriction) 、咽上壁 (pharyngeal constriction) 。卷舌音在汉语普通话及方言中发音有双唇突出状也是屡见不鲜的 (甚至有人记为有唇突出的舌叶音 ,以及带有伴随性舌尖后圆唇成分 ʁ 的辅音) 。舌尖与唇都是相对独立的发音体 ,从发音的角度看不存在相互牵制和关联的关系 ,不过从感知的角度看 ,这种发音的姿态能起到音类适应性扩散分布的目的。在语言里有舌尖辅音的一般是舌尖前辅音 (如 ts ts^h s 之类) ,如果另有一套卷舌辅音 (如 tʂ tʂ^h ʂ 之类) ,两者就要尽量拉开音类感知距离 ,以提高感知区别度 ,舌尖前辅音的特点是爆破点或擦音持续噪音的能量集中区处于高频 ,如果语言里再有一套舌尖音 ,最佳的区别空间是能量集中区处于低频 ,双唇收紧

或者突出可以很好到达这个效果。波兰语也有类似的情况,波兰语有三套舌齿擦音 /s/、/ç/、/ʃ/ , 卷舌的 /ʃ/ 发音时有圆唇和向外突出的姿态 (Puppel, Nawarocka-Fisiak & Krassowska, 1977: 157) 。/ʃ/ 的舌尖卷舌姿态已经使得收紧点比发 /s/、/ç/ 更靠后,前腔的共鸣频率降低,双唇突出和圆唇又使前腔拉长,使得前腔的共鸣频率更低,以有利于区别具有前腔高频的 /s/、/ç/。根据物理学共鸣腔频率的微扰理论(perturbation theory) ,共鸣腔体中的波腹(antinode) 处收紧,会降低此共振的频率。图 4.03 就是口腔作为均匀管子时第一(R1)、第二(R2)、第三(R3) 共振波形示意图,粗线处是发卷舌音的三个收紧点(唇、硬腭中部、咽上壁),这正是卷舌音发音时的三个收紧点。

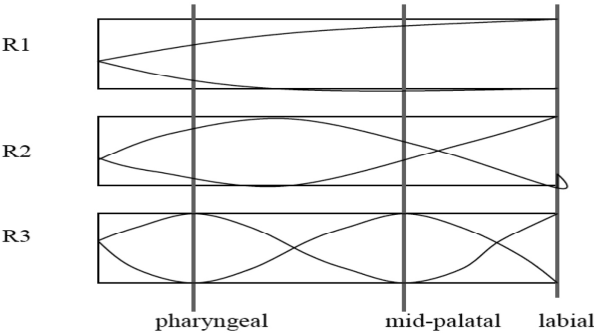


图 4.03 均匀管子第一(R1)、第二(R2)、第三(R3) 共振波形示意图

表 4.01 浙江吴语某些开口二等字有圆唇介音

| | 牙 | 马 | 嫁 | 窗 | 梗 | 瞎 |
|----|------------------|------------------|------------------|----------------------------------|-------------------|------------------|
| 松阳 | ɲuΛ ² | muΛ ⁵ | kuΛ ⁵ | | kuaŋ ³ | |
| 龙游 | ɲua ² | | kua ⁵ | | | |
| 汤溪 | ɦua ² | | kua ⁵ | | | hua ⁷ |
| 兰溪 | ɦua ² | | kua ⁵ | tɕ ^h yaŋ ¹ | kuæ ³ | hua ⁷ |
| 永康 | ɲua ² | mua ⁵ | kua ⁵ | tɕ ^h yaŋ ¹ | kuai | |
| 金华 | ɲua ² | | kuɐ ⁵ | tɕ ^h yaŋ ¹ | kuaŋ ³ | huɐ ⁵ |
| 武义 | ɲua ² | mua ⁵ | kuɔ ⁵ | | kua ³ | hua ⁷ |
| 建德 | ɲuɔ ² | | kuɐ ⁵ | | kuæ ³ | |
| 寿昌 | ɲuɔ ² | muΛ ⁴ | kua ⁵ | tɕ ^h yaŋ ¹ | kuã ³ | xuΛ ³ |

在唇部突出或者收紧都是在三个共振峰的波腹处,所以降低整个腔体的共鸣频率。特别是第三共振峰,发卷舌音的三个收紧处都是在波腹上,所以卷舌音如果有圆唇或者唇突出发音姿态,则第三共振峰一定是相当低的,压低第三共振峰也间接压低了第一、第二共振峰,如果是清塞擦音、擦音,则能量集中区都会落在很低的频率范围内,这样就跟能量集中区处于较高频率的舌尖前塞擦音、擦音(ts ts^h s 之类) 有效区别开来。所以,发卷舌辅音时的双唇圆唇或突出,虽然不是卷舌音的区别性发音特征,两者也没有发音机制上的联系,但是两者结合在一起起到有效拉大与舌尖前音的感知空间。Stevens 等(1986) 把这些非区别性语音特征命名为音类增强性(enhancement) 特征。这种增强性特征作为音位的冗余特征会伴随区别性特征而存在,它加强了区别性特征的声学效果,起到了让听者更容易感知音类间差异的作用。如果知道卷舌音圆唇化特性后,就能解释很多看似困惑的音变。如上古汉语二等介音为 r,它跟知、庄组声母相配,发展为中古的卷舌音声母

(李方桂 1971)。二等介音 r 如果圆唇特征突显 就会变为圆唇介音。表 4.01 所列浙江吴语开口二等字有圆唇介音 u(y), 可以解释为早先的 r 介音圆唇特征突显的结果。

5. 音类量子理论与语音感知

Stevens 和 House(1961) 基于 Fant “声源 + 共鸣滤波器”的理论模式, 提出用三个参数, 它们分别是收紧点位置、唇的突展度、以及共鸣声道的截面积等参数可以估算出某时某刻共鸣腔的共振峰频率。虽然如此, 发音体在共鸣腔中运动形成的不同声学效果并不是等量的。沿着唇往里直到喉部会发现某些发音部位的移动对声学参数的改变不是十分敏感, 但是在另外一些地方移动则会引起较大的声学参数改变, Stevens 把这一现象叫作语音发声与声学产出的量子理论(quantal theory of speech sound, Stevens 1972, 1989)。如图 5.01 所示(取自 Stevens 1989):

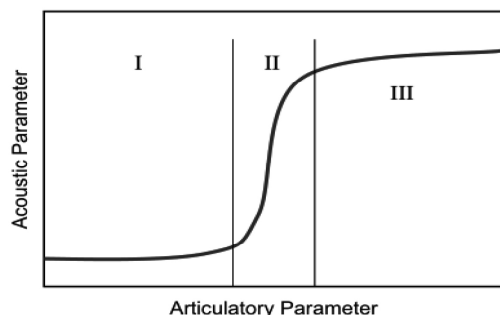


图 5.01 发音部位参数与对应的声学参数示意图

图 5.01 是代表发音体与声学产出对应图, 横轴代表发音体的变化参数, 纵轴则代表对应的声学产出参数, I 部位和 III 部位发音体参数的改变不会引起相应声学参数的大的变动, 也即声学参数对之不敏感; II 部位发音体参数的改变则会引起声学参数剧烈的变化, 也即声学参数对之非常敏感。Stevens 把 I 部位和 III 部位叫作稳定段(stable regions), 把 II 部位叫作不稳定段(transitional regions)。Stevens 认为稳定段与不稳定段交替形成的音段音位是一个语言里音位系统的最佳序列(optimization of the phonemic inventory)。Stevens 从发音部位及声学参数两方面来考察稳定段的音到底是哪些, 他认为对于元音来说收紧点后的腔体(后腔)形成的最低共振, 即元音的第二共振峰(F2), 如果与邻近的共振峰(F1 或 F3)形成双高峰, 那就是量子元音(quantal vowels)的必备条件。元音 /u/、/i/、/a/、/u/ 正好符合这些条件。请看下列发元音 /i/、/a/、/u/ 共鸣腔收紧点位置和对应的共振峰曲线图(据 Stevens 1989 重新排列)。

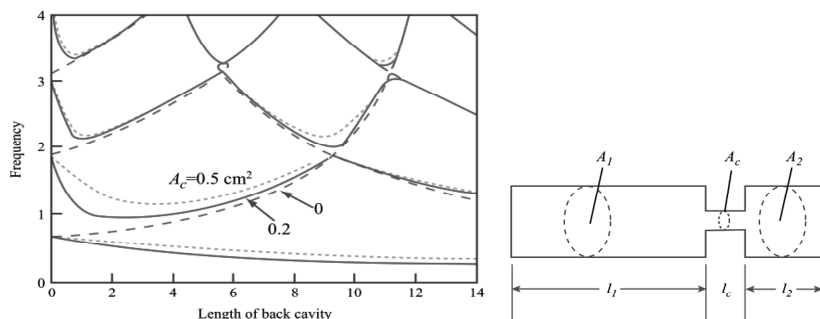


图 5.02 发 /i/ 时共振峰曲线及共鸣腔收紧点示意图

发元音 /i/ 从喉往外 10 cm 处发 /i/ 的收紧点,也正好是第二共振峰与第三共振峰最为接近点。
从喉往外 8 cm 处是发 /a/ 的收紧点,也正好是第二共振峰与第一共振峰最为接近处。

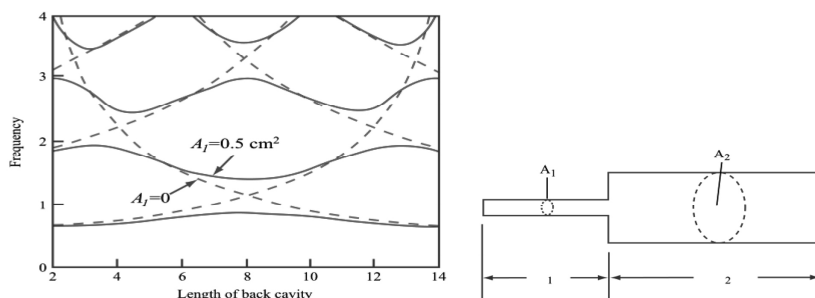


图 5.03 发 /a/ 时的共振峰曲线与共鸣腔示意图

从喉往外 4-6 cm 处和 12 cm 处是发 /u/ 时的收紧点(双收紧点)。4-6 cm 处、12 cm 处也正好分别是第二共振峰与第三共振峰、第二共振峰与第一共振峰最为接近处。

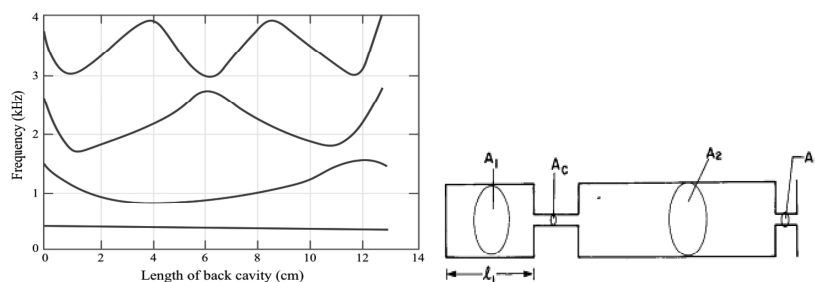


图 5.04 发元音 /u/ 时的共振峰曲线与共鸣腔示意图

任何的两个共振峰数值接近都会起到增强彼此的能量的作用,从而使得这些元音的声学音征显赫。

量子理论的提出不仅是发音与声学对应关系的重大发现,而且还可以解释语音感知及语音变化很多现象。

第一,处于声学参数不明感的发音部位是双唇、硬腭、软腭、以及咽腔上部,这些部位收紧是音段的稳定部位,元音 /i/、/u/、/a/ 的发音收紧点正是坐落在这些稳定部位,这就解释了为什么世界绝大多数语言都有着三个最为重要的周边元音(peripheral vowels)。

第二,除了三个周边元音外,其他重要的语音特征也可以在这些稳定部位得以落实。这些不同的稳定段所提供的声学参数正是人类感知语音区别性特征的重要依据,Stevens 本人就以此为依据来建立自己的语音区别特征(Stevens 2006)。

第三,发音动作与声学参数不是一一线性对应的。在 I 和 III 稳定段范围内,发音动作不必十分到位和精确,因为在这个稳定段内所产生的声学参数是差不多的。换句话说,在一种语言里,落在稳定段的音类,发音动作的变异不会引起声学参数的剧烈变化,人们对这些变异的感知也是不敏感的。处在发音稳定段的不同音类,如图 5.01 的 I 部位和 III 部位,也是语音感知空间突然跳跃的类别,也是切分不同音段发声(不同的发音部位)、声学(不同的声学表现)、以及感知(不同的感知类别)的重要依据。

第四,语音音类感知空间的突然跳跃与人类特有的共鸣腔特点具有密不可分的关系。喉头下沉,形成人类特有的与口腔大致成直角和等距离的咽腔;颌骨由于进化而退缩,与此同时,舌骨、舌

体后退,舌则可以更自如在咽腔处收紧以及在共鸣腔内灵活运动。其中舌体能在口腔与鼻咽腔成90度直角处隆起或收紧是产生 /i//u//a/ 三个距离最大元音的关键(Lieberman 1984: 276 - 280)。

第五,它还可以解释许多音变现象。两个相邻的音段 AB 在发音时必定会产生协同发音(co-articulation) 不过到底是 A 影响 B 还是 B 影响 A? 量子理论有一个较为合理的解释。一般来说收紧点较紧的那个声学参数易变,收紧点较松的那个声学参数较稳定。这就解释了在语言里,辅音 + 元音的组合里,往往是辅音不稳定,变化大,而元音的变化小,较稳定的现象。如我们常见的腭化音变:

$$k \ k^h \ g > t_e \ t_e^h \ d_z \ (t_f \ t_f^h \ d_3) / \text{---} (i, y)$$

腭化音变是前高元音使得前面的辅音往硬腭方向变,而前高元音本身并不发生改变。用量子理论来解释就是因为辅音收紧点通常比元音的紧,不稳定,易变;元音收紧点松,较为稳定,更何況高元音 /i//y/ 也正好处于量子理论的稳定段,在音变的过程中是起主导作用的。

量子理论所涉及的内容主要是发音体运动和对应的声学参数的非线性(nonlinearity) 关系,但是这种关系其实也制约着听者的音类感知,一个语言里选择量子元音与其说是基于说者,还不如说是基于听者和说者双方配合默契的选择。

6. 说者协同发音与听者补偿效应

语言交际是高速而有效的。从说话者的角度看,相邻的音段在正常的语流中,发音动作并非线性运作而完成,而是部分或全部重合(overlap),这时候发音动作必须相互协调、迁就,从而形成协同发音(coarticulation)。图 6. 01 是英语 palm /pa: m/ ([p^ha: m]) 的发音动作(Browman & Goldstein 1992)。

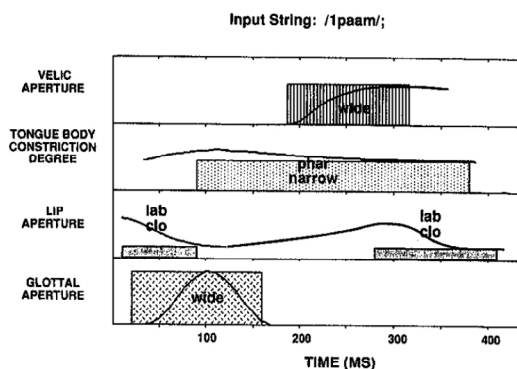


图 6.01 英语 palm /pa: m/ ([p^ha: m]) 发音动作

在发送气双唇塞音 p^h 时舌体已经在为发 a 做准备,舌体后缩向咽腔收紧(tongue body constriction degree) 的动作在双唇爆破不到 50 毫秒时已经开始(tongue body constriction degree 中的黑线)。元音开始则在爆破不到 100 毫秒处,此时双唇塞音送气段尚未结束,送气段一直延伸到 150 毫秒左右,换句话说元音前部分,大概 50 多毫秒是气噪音发声态。而在发元音时,鼻咽通道在 180 毫秒处已经开始开启(velic aperture),并逐渐增大,一直贯穿后面的整个元音发音阶段,换句话说 180 毫秒以后的元音都是鼻化的。而 m 的双唇闭塞的动作是从 275 毫秒处就开始了,此时发元音 a 时舌根向咽腔收紧的动作尚未结束。从图 6. 01 可以看出双唇收紧发 m 与舌根向后收缩发 a 的动作在 275 毫秒到 375 毫秒处是重合的。可见 [p^ha: m] 语流串里音段与音段间都是部分重合的。音段与音段重合必定会相互影响,如果两个音发音动作差距较大,就会相互妥协、协同。

不过发音上的协同是要受其他因素制约的,比如受一个语言音系格局的制约。Manuel 曾指出在元音与元音间(V-to-V)的协同发音,不同音系格局协同发音的程度是不同的,元音数量多的语言协同发音程度小,元音数量少的语言协同发音程度大(Manuel 1990)。见图 6.02。

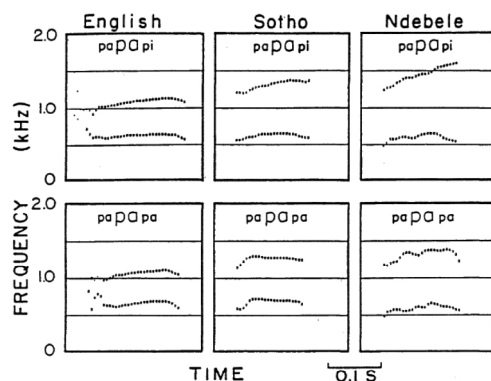


图 6.02 三种语言协同发音比较

Sotho 语和 Ndebele 语属于南部班图语系(Bantu language family)。Sotho 语元音有七个/i u e o ε ɔ a/, Ndebele 语元音只有五个/i u e o a/。英语的元音是较多的,根据方言的不同一般在 13 至 15 个之间,其元音数量远远高于 Sotho 语和 Ndebele 语。图 6.02 上三图是三种语言发/papapi/时中间音节元音/a/的共振峰 F1、F2 的走向,比较下三图三种语言发/papapa/时中间音节元音/a/的 F2 走向,可以看出英语的变化不大,也即最后一个音节的/i/对中间的元音/a/的共振峰走向影响不大,而 Sotho 语和 Ndebele 语里最后一个音节的/i/对中间音节中的元音/a/有明显的协同发音效果,前高元音/i/高的 F2 会使前一音节中的元音趋同,所以前一元音的 F2 的走向是往上翘的(数值渐大),这种上翘的走向在 Ndebele 语里更为显著。Manuel(1990)的解释是英语里由于元音多,元音之间的声学 and 感知空间就拥挤,英语里/a/与/æ/是不同的音位,所以/a/如果要受最后一个音节/pi/的元音/i/协同,舌位无法作明显的升高,因为上面有/æ/,所以变异的空间就小,协同发音效果就无法彰显,否则会泯灭元音间的感知空间,音类的区别就会混淆。Sotho 语,特别是 Ndebele 语,元音少,元音间的声学 and 感知空间就宽裕,Sotho 语/a/上面的元音是/ε/,所以受最后一个音节/pi/的元音/i/协同发音,舌位可上升的空间比英语的大。Ndebele 语/a/上面的元音是/e/,受最后一个音节/pi/的元音/i/协同发音,舌位可上升的空间就更大,协同发音的效果更能彰显出来。不仅语言的音位系统制约着协同发音的程度,音节结构的复杂程度也对协同发音的程度有制约作用,一般来讲具有复杂音节(如 CCCVCC 结构)的语言比只有简单音节(如 CV 结构)的语言元音协同发音的程度要高,因为音节结构复杂,元音在其中的负担就轻,就可以有更多的变异空间(Mok 2010)。

协同发音本来是发音器官掌控的,是发音的问题,但是协同发音的程度显然是受听者感知的制约,协同发音所引起的语音变异不能损害不同音类的感知。所以保持音类的区别性会倒逼发音时协同发音的范围和程度。

协同发音与感知补偿(perceptual compensation)这一对概念更是体现人类语言发音与感知的互动关系。感知补偿是指听觉对协同发音的一种补偿机制。发音时线性相连的音段会产生协同发音,相邻的音段某些特征会趋同;而听者接受声音时会有能力把这种音段间的趋同特征剥离,分辨出它们的差异来。这种听者的矫正机制就叫做感知补偿。Linblom 和 Studdert-Kennedy(1967)首先提出感知补偿这一概念。他们用语音合成器合成 /I ~ ʊ/ 持续刺激音(continuum stimulant)分别放在 w_w 和 j_j 的环境里测试,人们在 w_w 环境里比 j_j 环境里更多听成 /I/。从发音的角度看,刺

激音在 w_w 的环境里会受前后 /w/ 第二共振峰(F2) 低的影响而产生协同发音效应 ,两边语音 F2 低的特征也会影响中间的语音 ,使得它的 F2 也低 ,而 F2 低正是 /O/ 的特征。但是听者的感知结果恰恰相反 ,更多的听成 /I/ 而不是 /O/! 该文作者对这一现象作了解释 ,他们认为听者在 w_w 的环境里已经预测到前后音 F2 低的特征会给中间的音 ,也即发生协同发音 ,但是听者会认为这并非中间那个音本身的音征 ,应该在感知中把 F2 低这个特征剥离掉 ,这样中间的那个音就会凸现高的 F2 ,那么听感上更容易听成具有 F2 高特征的 /I/。他们认为这是听者感知补偿机制起了作用 ,也即: 用听者的感知补偿自动矫正了发音上的协同发音。感知补偿效应在以后的各种研究中不断被证实。Mann 和 Repp (1980) 的文章也是这一领域的经典之作。文章用语音合成器合成 [s] 到 [ʃ] 的七个持续刺激音 ,放在合成音 [a] 或 [u] 前 ,组成 CV 结构 ,然后做听辨实验 ,测试听者在 [a] 和 [u] 环境里听到 [s] 和 [ʃ] 的百分比。笔者用 Praat 语音分析软件从自然声音截取 [s] 和 [ʃ] (从 see 中截取 [s] ,从 shoe 中截取 [ʃ]) ,然后组成 [s] 逐渐到 [ʃ] 七个均等的连续体。如图 6.03:

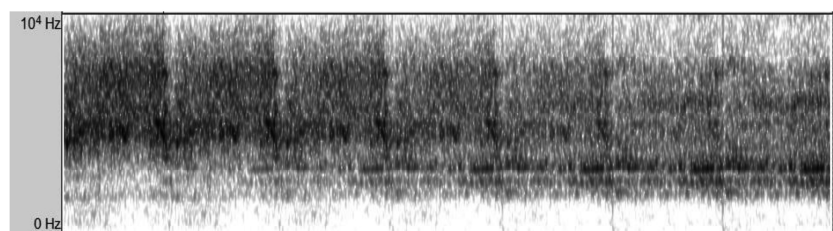


图 6.03 [s] 到 [ʃ] 七个均等的连续体频谱

[s] 到 [ʃ] 噪音的能量集中区的频率范围逐渐下移 , [s] 主要的能量分布在 4000Hz 以上 , [ʃ] 主要能量分布则介于 1700 – 5000Hz 之间。通常我们会觉得 [s] – [ʃ] 持续刺激音放在 [u] 前被听为 [s] 的比例会比较高。由于有协同发音作用 ,发前面的辅音要为后面的 [u] 做准备 ,即发辅音时已撮起双唇 ,噪音生源前的腔体就拉长 ,腔体长共振频率就低 ,所以这个噪音的能量集中频率也会较低 ; [a] 没有双唇收紧动作 ,所以对前面的辅音的影响不会像 [u] 那样。根据上述推理 ,可以想象在 [u] 前的刺激音听成 [ʃ] 的比例相对会比较高 ,但是听辨实验的结果恰恰相反 ,在 [u] 前的刺激音听成 [ʃ] 的比例相对较低! 图 6.04 是笔者根据 Mann 和 Repp(1980) 的分析重新绘制而成。

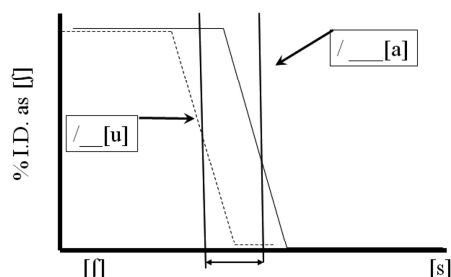


图 6.04 在 u 和 a 前 [s] – [ʃ] 持续刺激音听成 [ʃ] 的百分比

图 6.04 纵轴 “% I. D. as [ʃ]” 表示听为 [ʃ] 的百分比 ,横轴是从 [ʃ] 到 [s] 七个连续刺激音 ,虚线代表在 “/__ [u]” 环境里听成 [ʃ] 的百分比 ,实线代表在 “/__ [a]” 环境里听成 [ʃ] 的百分比。这个表格显示在 “/__ [u]” 环境里听成 [ʃ] 的百分比要比在 “/__ [a]” 环境里听成 [ʃ] 的百分比少。作者认为这是听者的感知补偿在起作用。也即 ,从发音角度来看 ,“擦音 + [u]” 的 CV 组合后面的后高圆唇 [u] 一定会对前面的擦音产生协同发音作用 ,使擦音能量集中区的频率范围降低 ,但从人对声

音的感知角度来看,人们会把前面擦音段低频的特点归结为是后面的元音的影响所致,从而把擦音中这个低频的特征剥离开来,恢复原来擦音的面貌。可以用简单的公式作如下的表示:

说者 听者
/su/ > [ɸu] [ɸu] > /su/

由于有协同发音的作用,说者把/su/组合实际发成[ɸu]信号,听者接收的信号是[ɸu],此时听者的感知补偿起作用,仍然感知为/su/,恢复跟说者相同的音段序列。

从发音这个角度看,在快速的语流中为了要达到最短时间段内发出最多言语信息的目的,必定要协调发音动作,产生协同发音,相连的音段会趋同;但是从听的角度来看,相连的音段趋同,会模糊音与音的区别,从而导致言语交际的困难。如果听者完全凭借特定音段的信息来解码,就会误解读音,言语交际会失败。现在听者感知补偿机制启动,就会对传来的声学信息作自动矫正,恢复说者要想传递的信息,言语信息才能正确无误传递。

感知补偿效应对音变的解释也具有革命性的意义。我们把感知补偿效应的环境称之为补偿产生作用的条件(predictor),简称补偿条件。在其他条件不变的情况下,只要补偿条件存在,就不会发生音变。笔者曾对此做过研究(陈忠敏 2014)。比如北京话“钢”/kaŋ/中间的元音其实是有鼻化的,但是一般人不会觉得这是一个鼻化元音,因为补偿条件/ŋ/存在;一旦补偿条件失去,元音就会被感知为鼻化。比如上海话同样的“钢”,就是因为韵尾没有鼻音作补偿条件,元音就被感知为鼻化元音。“女”字早期的韵母读音在北部吴语读如麻韵,类似于[o]这样的读音(潘悟云 1995)。上海话“囡”其实是“女儿”/no + n/的合音,平行的例子是麻韵的“虾”,早期上海话“虾儿”读音是/ho + ŋ/,以后“女儿”“虾儿”都发生了相同的音变,首先跟自称音节的词尾儿语素合并成一个音节,然后才是鼻韵尾脱落,同时主元音改变:

女儿 /no + ŋ/ > /non/ > /nø/
虾儿 /ho + ŋ/ > /hon/ > /hø/

现在要解释韵母/on/ > /ø/的音理。/o/音是+grave特征,/n/是舌尖鼻音,具有+acute特征。根据协同发音的道理,在/on/阶段,/o/受/n/的影响,必定舌位前移也带有+acute特征,可以表述为:/on/[ø̃n]。此时第一个音段虽然受后一音段[n]影响,舌位前移到[ø̃],带有+acute特征,但是由于有补偿条件[n]的存在,人们不会觉得发生了音变,所以仍然感知为/on/。只有鼻韵尾失去,没有了感知补偿条件所凭借,人们才会觉得元音不一样了,认为元音变化了,所以标为/ø/。

英语及其他日尔曼语语音(umlaut)音变也可以用感知补偿条件丢失引起音变来得到解释。见表6.01。

表 6.01 英语名词复数曲音(umlaut)音变

| 不发生 umlaut 音变(单数) | | 单数 + 复数后缀[i],发生 umlaut 音变 |
|-------------------|---------|--|
| mu: s | ‘mouse’ | * mu: s + i > mu: si > my: s > mi: s > majs ‘mice’ |
| fo: t | ‘foot’ | * fo: t + i > fo: ti > fø: t > fe: t > fi: t ‘feet’ |
| to: θ | ‘tooth’ | * to: θ + i > to: θi > tø: θ > te: θ > ti: θ ‘teeth’ |
| go: s | ‘goose’ | * go: s + i > go: si > gø: s > ge: s > gi: s ‘geese’ |

单数 + 后缀语素[i]就是早期英语名词复数,以后后缀[i]丢失,词干主元音也同时发生了从后元音变为前元音的变化,μ > y, o > ø。笔者认为在后缀[i]尚未丢失之前,由于受后面[i]的影响,其实词干元音已经发生了舌位前移的变化,即[my: si]、[fø: ti]、[tø: θi]、[gø: si],只是补偿条件[i]

还在,所以没有音变的感知。一旦词缀[i]丢失,补偿条件不存在了,人们才会有词干元音发生变化的感觉。

当声母还有清浊区别的时候,我们古人对声调调类的感知是平上去入四个声调,但是有理由相信,浊声母字的调值应该低一点,只是清浊声母这一感知补偿条件存在,同调类声调的高低并不凸现。一旦这一补偿条件失去,即浊声母清化,阴阳调的差异才会被人们感知。

不过作为感知补偿条件的因素是有限制的,从音理上它必须能激发音变。藏语的音变就能很好说明这个问题。见表 6.02:

表 6.02 古藏语与现代拉萨话音变

| 七世纪书面藏文 | 今拉萨话 | 词义 |
|-------------------|-------------------|----|
| goŋ | qõ: | 价钱 |
| nub | nu: | 西方 |
| drug | t ^h u: | 六 |
| t ^h og | t ^h o: | 屋顶 |
| | | |
| lus | ly: | 身体 |
| bod | p ^h ø: | 西藏 |
| ston | tõ: | 秋天 |
| spos | pø: | 香 |

/ŋ/、/b/、/g/ 都不具备使前面的元音舌位前移的条件,所以当它们脱落以后,主元音仍保留 o 和 u,同时主元音作补偿性拉长(compensatory vowel lengthening)。不过/on/也可以认为是[õŋ],感知补偿条件/ŋ/还在的时候,元音不会被认为是鼻化的。当鼻韵尾丢失,原来的元音鼻化就凸现了,所以才会/q^hõ:/。/us/ > /y:/、/os/ > /ø:/、/od/ > /ø:/、/on/ > /õ:/很明显是经历了下面的阶段:

/us/ [ys] > /y:/
 /os/ [øs] > /ø:/
 /od/ [ød] > /ø:/
 /on/ [õn] > /õ:/

其中/s/、/d/、/n/都是这些音的感知补偿条件,这些音都具有 +acute 特征,基于协同发音,前面的音都会舌位前移。但是一旦这些补偿条件失去,感知补偿就无所依托,功能失效,音变就产生了,与此同时元音作补偿性拉长。

7. 结语

交际是语言的最大功能。一次言语交际过程包括发声、传播、感知三个流程,涉及到说者与听者双方。成功的言语交际具有两大要求,第一大要求是要清晰。Jakobson 等人(1952)曾说过“我们说是为了让别人听和懂”(we speak in order to be heard, in order to be understood)。为了让别人听和懂,就要求说者说话时要说得清楚和明白。听者的这种要求会倒逼说者发音要具备清晰的发音特征,增加或增强语音的非区别性发音特征就起到这一作用。如前元音增加或增强不圆唇性特

征,后元音增加或增强圆唇性特征成为最常见元音格局。卷舌音常伴随圆唇特征也是属于这种性质。成功交际的另外一大要求是语言传播要高效,单位时间内传播的信息量要大,所以语速就要快。说者语速快,必定会产生协同发音,听者就启动感知补偿效应来自动矫正传来的模糊语音,恢复说者心目中想说的话,从而达到交际成功。从音变的角度看,发音清晰是音变的一个潜在的目标和方向。说者和听者毕竟不是同一个人,语音声波通过不同环境的媒介传送也会有损耗、扭曲,有时说者与听者的配合不是那么默契和有效,从而导致听者误解,产生音变。

人类的语言是精妙绝伦的,既然是交际的工具,说和听双方必定是相互配合和制约的。接受语言的听觉外延器官以及听觉神经中枢会有选择性地捕捉有用的语音信号,屏蔽或抑制与此无关的非语音信号;从发音角度来说,发出的音要调试到听者最容易接受和最容易辨别的音类范围内,才是达到最有效的语言交际目的,换句话说,说者的发音也受到来自听者感知器官的掣肘。语言研究必须同时考虑说、传播、听三者的机制以及它们之间的关系,从说者和听者的角度同时研究,才能全面认识人类语言的奥秘。

参考文献:

- [1] Browman, C. P., & L. Goldstein. Articulatory Phonology: An Overview [J]. *Phonetic*, 1992, 49: 155 – 180.
- [2] Flege, J. E. & W. Eefting. Cross-language switching in stop consonant perception and production by Dutch speakers of English [J]. *Speech Communication*, 1987, 6: 185 – 202.
- [3] Gick, B., Wilson, I. & D. Derrick. *Articulatory Phonetics* [M]. Wiley-Blackwell, 2013.
- [4] Johnson, Keith. *Acoustic and Auditory Phonetics*. Third edition [M]. Wiley-Blackwell, 2012.
- [5] Jakobson, R., Fant, G. & M. Halle. *Preliminaries to Speech Analysis* [M]. Cambridge, MA: MIT Press, 1952.
- [6] Lieberman, Philip. *The Biology and Evolution of Language* [M]. Harvard University Press, 1984.
- [7] Liljencrants, Johan & Bjorn Lindblom. Numerical simulation of vowel quality systems: the role of perceptual contrast [J]. *Language*, 1972, 48(4): 839 – 862.
- [8] Lisker, L. & A. S. Abramson. A cross-language study of voicing in initial stops: acoustical measurements [J]. *Word*, 1964, 20: 384 – 422.
- [9] Lindblom, B. & M. Studdert-Kennedy. On the role of formant transitions in vowel recognition [J]. *Journal of the Acoustical Society of America*, 1967, 42: 830 – 843.
- [10] Fant, G. *Acoustic Theory of Speech Production*. 2nd edition [M]. The Hague: Mouton and Co, 1970.
- [11] Mann, V. A. & B. H. Repp. Influence of vocalic context on the [s]—[ʃ] distinction [J]. *Perception and Psychophysics*, 1980, 28: 213 – 228.
- [12] Manuel, S. The role of contrast in limiting vowel-to-vowel coarticulation in different languages [J]. *Journal of the Acoustical Society of America*, 1990, 88: 1286 – 1298.
- [13] Mielke, J., Baker, A. & Archangeli, D. Variability and Homogeneity in American English /r/ Allophony and /s/ Retraction [C]// C. Fougerson, B. Kuhnert, M. d'Imperio, and N. Vallee (eds). *Laboratory phonology 10* (Phonology and Phonetics 4 – 4). Berlin: de Gruyter Mouton, 2010: 699 – 730.
- [14] Mok, P. P. K. Language-specific realizations of syllable structure and vowel-to-vowel coarticulation [J]. *Journal of the Acoustical Society of America*, 2010, 128 (3): 1346 – 56.
- [15] Moore, B. C. J. & B. R. Glasberg. Masking patterns of synthetic vowels in simultaneous and forward masking [J]. *Journal of the Acoustical Society of America*, 1983, 79: 906 – 917.
- [16] Ohala, John. The phonetics and phonology of aspects of assimilation [C]// Kingston, J. & M. Beckman. *Papers in Laboratory Phonology, vol. 1: Between the grammar and the physics of speech*. Cambridge: Cambridge University Press, 1990. 258 – 275.

- [17] Peterson , G. E. , & H. L. Barney. Control methods used in a study of the vowels [J]. *J. Acoust. Soc. Am* , 1952 , 24: 175 – 184.
- [18] Puppel , S. , Nawrocka-Fisiak , J. & H. Krassowska. *A Handbook of Polish Pronunciation for English Learners*. [M]. Warsaw: Panstwowe Wydawnictwo Naukowe , 1977.
- [19] Rosen , Stuart & Peter Howell. *Signals and Systems for Speech and Hearing , Second Edition* [M]. Emerald Group Publishing Limited , 2011.
- [20] Stevens , N. Kenneth. The quantal nature of speech: Evidence from articulatory-acoustic data [C]// Denes , P. B. & E. E. David Jr. *Human Communications: A unified view*. New York: McGraw Hill , 1972. 51 – 66.
- [21] Stevens , N. Kenneth. On the quantal nature of speech [J]. *Journal of Phonetics* , 1989 , 17: 3 – 46.
- [22] Stevens , N. Kenneth. *Acoustic phonetics* [M]. MIT Press , 1999.
- [23] Stevens , N. Kenneth. Features in speech perception and lexical access. [C]// Pisoni , David B. & Robert E. Remez. *The Handbook of Speech Perception*. Blackwell Publishing , 2006. 125 – 155.
- [24] Stevens , K. N. & A. S. , House. An acoustical theory of vowel production and some of its implications [J]. *J. Speech Hear. Res.* , 1961 , 4: 303 – 320.
- [25] Stevens , K. N. , Keyser , S. J. & Haruko Kawasaki. Toward a phonetic and phonological theory of redundant features. [C]// Perkell , Joseph S. & Dennis H. Klatt. *Invariance and Variability in Speech Processes*. Hillsdale: Lawrence Erlbaum , 1986. 426 – 449.
- [26] 陈忠敏. 音变的原因探索——从发音生理、声学解释到感知解释 [C]// 潘悟云. 高山流水——郑张尚芳教授八十寿诞庆祝文集. 上海: 上海教育出版社 2014. 71 – 83.
- [27] 潘悟云. “因”所反映的吴语历史层次 [J]. *语言研究* , 1995 , (1) : 2 – 14.
- [28] 吴宗济、林茂灿主编. 鲍怀翘、林茂灿增订. 实验语音学概要 [M]. 北京: 北京大学出版社 2014.

基金项目: 国家社科基金重点项目“上海市方言地图集”(15AYY005)

收稿日期: 2018 – 05 – 21

作者简介: 陈忠敏 教授 博士生导师。研究方向: 语音学。

“教师专业学习共同体(PLC) 研究丛书”书讯

近年来 随着教师专业化的深入 教师专业发展的途径及策略研究成为大家关注的热点。在此背景下 教师专业学习共同体(Professional Learning Communities 简称 PLC) 作为教师专业发展的新途径 引起了广大学者及教育工作者的重视。当下 国内外关于教师专业学习共同体的研究成果也不断涌现。

专业学习共同体是作为学习者的教师以自愿为前提 以分享、合作为核心 以促进学生学习的愿景为纽带 相互交流和共同学习的一种组织形式 是教师获得可持续专业成长的有效途径 为教师间、师生间的共同进步和发展提供了良好平台 对于促进学校的文化变革和教师工作环境的改善意义深远。

为了更好地方便我国读者研习教师专业发展的途径和策略 上海外语教育出版社特从 Solution Tree 精心遴选引进了“教师专业学习共同体(PLC) 研究丛书” 从不同层面对教师专业发展、教师管理和培养进行指导 对教师专业学习共同体的常见问题进行解答 希望丛书的引进出版能为我国教师专业的可持续发展、为构建中国特色教师专业学习共同体提供切实可行的理论指导与实践借鉴。