

# 肌 动 理 论 和 语 言 认 知

陈忠敏

(复旦大学 中文系, 上海 200433)

**摘 要:** 肌动理论(Motor theory)是关于语音感知的一种理论,它认为听者是根据自己的发音动作来匹配和感知外来语音的。这一理论刚提出就在学术界充满争议。本文利用脑科学和神经科学的最新研究成果来支持语音感知的肌动理论,并进一步说明:1) 听者的语音感知机制和发音机制紧密相关;2) 与其它非语音的声音感知不同,语音感知带有很强的主观性,而这种主观性受特定语言规则的制约;3) 人类就是通过这种主观的语音感知来过滤各种语音变异,达到语言交际的目标;4) 电脑语音识别必须转变思路,要充分考虑人类语音感知的特点,这样才能提高语音识别率。

**关键词:** 肌动理论; 主要运动皮层; 主要听觉皮层; 语言感知; 语言变异

## Motor Theory and Language Cognition

CHEN Zhongmin

(Department of Chinese, Fudan University, Shanghai 200433, China)

**Abstract:** The Motor theory of speech perception proposes that the listener receives a speech sound and analyzes it by eliciting an auditory model of his own production of it. Within the academic area of speech science, despite its prominence, the theory is filled with controversy. In this paper the author shows some strong evidence of latest studies in brain science and neuroscience to support the theory, and goes a step further to claim: 1) Speech perception and production mechanisms are tightly linked. 2) Different from other non-voice sound perception, speech perception has a strong sense of subjectivity, and such a subjective matter is constrained by language-specific rules. 3) In this way human beings can normalize speech variations to achieve the objective of language communication. 4) Scientists of computer speech recognition should fully consider the characteristics of human speech perception so as to improve the speech recognition rate.

**Key words:** Motor theory; primary motor cortex; primary auditory cortex; language perception; linguistic variations

### 1.0 引言

言语发声和言语理解是非常复杂的语言认知过程。从发音这个角度来讲,首先言语认知的高级中枢——大脑运动皮层要做出发音的运动规划,这一规划本身就是一个十分复杂的过程,要各相关系统的协调和参与才能做出发音动作的规划,然后规划进入执行过程,主要运动皮层(primary motor cortex)发出指令引起神经元(neuron)放电,生理电信号通过轴突(axon)和突触(synapse)传到下一个神经元,再牵动各种肌肉来驱动发音体及共鸣体改变,从而使得腔体内的空气粒子压力发生变化,传播到听者耳朵里,再由听者的大脑分析出接受的声音,产生我们能听到的语音。其中,说者发出的声音还要经过声波和躯体运动的反馈来做出适当的调节和矫正。根据Guenther[3]语言发音的DIVA(Directions Into Velocities of Articulation)模式,语音发音的流程如图1.01。

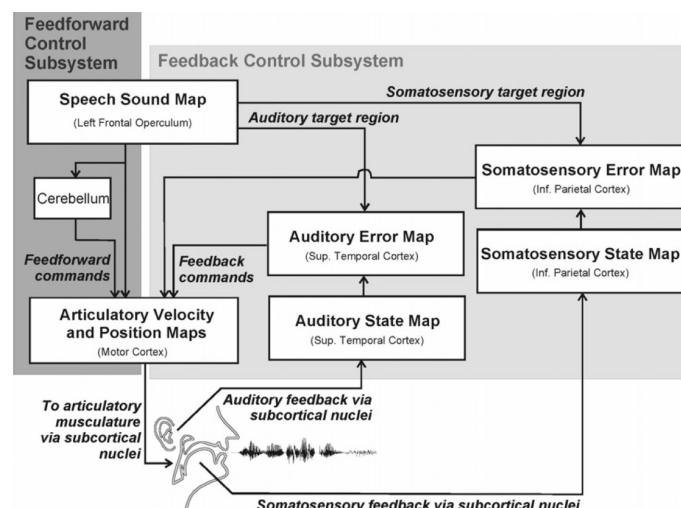


图 1.01

言语发声既有前馈控制子系统 (feedforward control subsystem) 作用,也有反馈控制子系统 (feedback control subsystem) 作用。反馈控制子系统分声波 (auditory) 反馈和躯体感觉 (somatosensory) 反馈。人类首先在左脑说话的运动皮层 (left frontal operculum) 规划发音动作,这个发音动作的规划通过前馈子系统神经元牵动发音体肌肉,使发音体作出运动,随着共鸣腔大小形状的改变,腔体内气压也产生相应变化,产生声波;与此同时左脑说话的运动皮层也向大脑颞叶 (temporal cortex) 的听觉区和顶叶 (parietal cortex) 的躯体感觉区送出这个规划音的理想目标值 (the ideal target)。经前馈子系统产生的语音会通过两种途径反馈给说话者,一种是声波,即声音通过空气粒子的振动传入自己的耳道,经过内耳再反馈到颞叶听觉皮层这一理想目标值投影区 (auditory state map);另一种是发音时的躯体感受反馈,即声带振动、共鸣腔体的动作通过躯体感觉神经反馈到顶叶躯体感觉皮层中的这一理想目标值投影区 (somatosensory state map)。如果反馈的信息跟理想目标值不同或有差异,就启动这两种反馈校对机制 (auditory error map/somatosensory error map) 对下一轮的发音作出修改,直至达到理想目标值为止。如果没有反馈,或者把反馈打乱,会直接导致发音者语无伦次。

语音反馈刺激对语音发声具有的作用在聋哑人语言教学中得到充分体现。中国俗语有“十聋九哑”之说,就是说绝大多数耳聋的都是不能说话的哑巴。其实哑巴的发音器官没有任何缺陷,不会说话是因为耳聋,听不到语音,没有言语的刺激。19 世纪中叶 Alexander Melville Bell [1] 了解到这一点,为聋哑人设计一套“可视语言” (Visible Speech),他根据发音的解剖生理构造把每个音 (元音或辅音) 画成可视的发音图,让聋哑人看图发音,可是 Bell 没有认识到语音反馈在语音发音中的重要性,所以这种对聋哑人的发音教学是失败的。即使是正常人,如果在没有言语环境中生长,过了言语习得的最佳时机 (critical period/sensitive period),以后说话的能力就会大打折扣。道理很简单:因为听不见,所以说不出。可见,言语发声与言语感知有十分密切的关系。既然语言中说和听有这么密切的关系,那么,是不是可以反之亦然,即:因为说不出,所以听不见? 这个命题使人想起语音感知的“肌动理论” (motor theories)。

肌动理论的最早版本叫 analysis-by-synthesis theory,上世纪 50 年代末由麻省理工学院著名教授 Kenneth N. Stevens 和 Morris Halle [5] 提出。他们认为听者是根据自己的发音来分析和感知外来的语音信号的。Haskins 实验室的科学家 Alvin M. Liberman 等人 [6] 也在差不多的时间正式提出

Motor theory 这个名称,以后 Liberman 等人 [7]对 Motor theory 又做了重大修改。基于语音感知和声学特征缺乏一致性( lack of invariance)、语音声学信号的不可切割性( nonsegmentability) ,以及语音传送信息的相对高效性( the relatively high efficiency of transmission of information) 他们也认为听者“雇佣”自己声腔大小的信息以及自己的发音运动特性,直接对外来语音作出分析和解码。Stevens 和 Halle 提出的 analysis-by-synthesis theory 和 Liberman 等人提出的 Motor theory 都强调听者是根据自己的发音动作来对外来语音解码和感知的。

这一假说当时由于缺乏脑科学及神经解剖科学的证明,曾遭受很多科学家、语言学家的责难,如 Ohala [10] ,以至于 Stevens [13]本人后来也放弃自己提出的假说,转而支持语言感知的听觉理论( Auditory Theories) 。Liberman 等人的肌动理论版本虽经过多次的修改,但是也没有阐述此一理论的神经解剖学机制,所以也处在理论的假说阶段。

上世纪 90 年代后,随着人脑科学及神经科学的发展,特别是脑功能成像技术的应用,人脑的语言认知研究及相关理论进入活跃期。基于神经科学的语言认知研究越来越受到语言学界和神经科学界的重视,这些研究极大地促进了人类对自己语言认知的认识。其中 Watkins 等人 [9; 12; 14; 15] 的一组研究成果从神经科学的角度揭示了语言感知和语言发声的紧密关系,从根本上奠定了肌动理论的神经生理机制。Watkins 等人的研究虽然只涉及语音感知的脑神经机制等相关问题,但是其研究结果对整个语言学将产生重大影响。本文将着力介绍他们的研究成果,在此基础上进一步发挥,来解释语音变异与语言认知的相关问题。

2.0 言语感知与发声关系的三个实验

Watkins 等人的文章 [14]采用经颅磁刺激( Transcranial magnetic stimulation ,TMS) 技术,并配合脑磁图记录( magnetoencephalography) 所显示的运动诱发电位( motor-evoked potentials ,MEP) 变化来观测听者接受不同声音( 语言声音、非语言声音) 刺激时大脑主要运动皮层中部动作的反应。实验的条件分为四类: 1) 言语( Speech) 条件,即在有视觉干扰时听连续的散文; 2) 非语言( Nonverbal) 条件,即在有视觉干扰的同时听非语言的声音( 如玻璃破裂声、铃声、枪射击声等); 3) 唇动( Lips) 条件,即在有白噪音( white noise) 的时候观察与言语相关的嘴唇动作; 4) 眼部( Eyes) 条件,即在有白噪音的时候观看眼睛和眉毛的动作。实验结果表明听者在条件 1) 接受言语声音,在条件 3) 观察与言语相关的嘴唇部动作的状态下,左半球主要运动皮层有更高的运动诱发电位值。在跟言语无关的声音和动作,即 2) 和 4) 状态下运动诱发电位值变化不显著;而在右半球,这四种状态下主要运动皮层的运动诱发电位值变化无明显区别,变化也不显著。如表 2.01 所示:

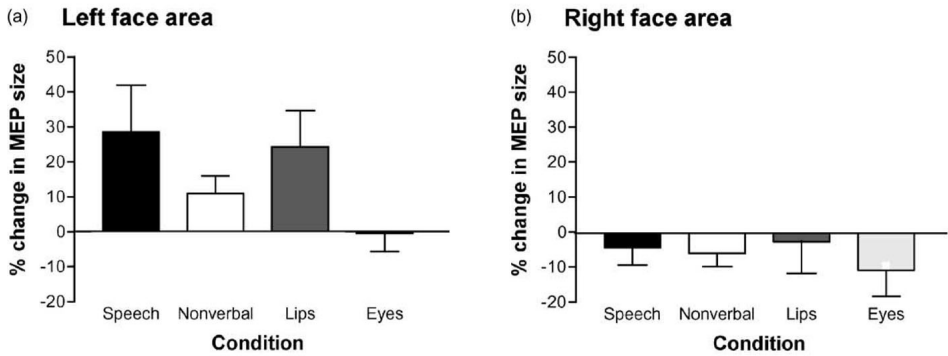


表 2.01

这项实验说明,听者在接受语音信号或跟语音信号相关的嘴唇发音动作时会激发他自己左半

脑管发音动作大脑皮层的反应。

经颅磁刺激只能在已知或目标固定的大脑皮层区域测试诱发运动电位的变化,语音的感知很可能是整个大脑皮层协调得以完成的,所以在未知哪个区域协调工作的情形下,经颅磁刺激这一手段就显其局限性。鉴于此,Watkins 等人在次年又发表相关的文章[15],把经颅磁刺激和正电子发射断层扫描(positron emission tomography, PET)技术结合起来,测试在言语感知过程中大脑哪些区域一起协调运动。正电子发射断层扫描根据血液中葡萄糖消耗程度来追踪大脑皮层各区域血流量(CBF)变化,从而来判断大脑不同皮层运动兴奋的程度。实验的条件分为三类,即:1)言语(Speech)条件;2)唇动(Lips)条件(与言语相关的嘴唇动作);3)眼动(Eyes)条件。

在言语、唇动及眼动三种状态下,血流量在大脑皮层的激增区是不同的,唇动和眼动血流量激增区重叠在大脑两侧枕颞相交区域(occipitotemporal regions);而言语状态下,虽然大脑两侧的颞上回(superior temporal gyrus)都显示血流量激增,但是左半球激增区更大。特别是在接受言语和模仿言语的唇动状态下,除了颞上回、颞叶前部钩回(the uncus in the anterior temporal lobe)外,左半球的额下回(left inferior frontal gyrus),也即布罗卡(Broca)区,血流量激增明显。作者再根据运动诱发电位数值(MEP)及布罗卡区脑血流量激增多元回归分析指出,听者在接受言语信息时,位于左半球额下回颞盖部(the inferior frontal operculum),即44区,以及顶叶,血流量和运动诱发电位数呈正相关性同步增长,见图2.01。

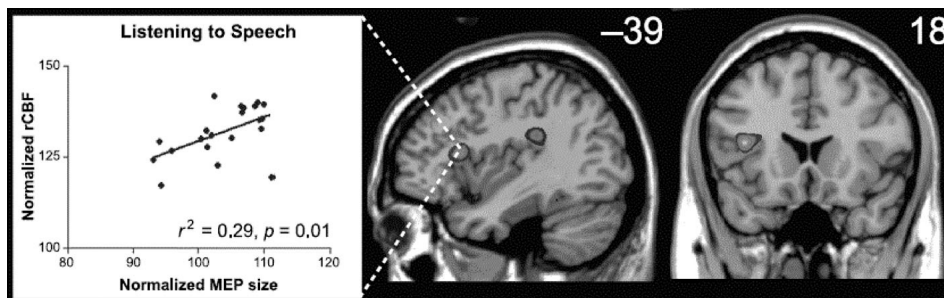


图 2.01

左半球44区正好跟管发音动作的主要运动皮层区域接壤。由此可以认为,外来语音信息首先在左上颞叶主要听觉皮层被接受,然后转入顶叶缘上回(the supramarginal gyrus),再投射到左半球44区。44区处在调制发音运动的皮层上,它可以通过直接或者间接方式连接腹侧前运动皮层(the ventral premotor cortex),来匹配自己的发音动作,从而对外来语音进行感知。换句话说听者必须“雇用”自己的发音“动作”来感知外来语音。

如果说Watkins等人上述的两篇文章是从正面来阐述“语言感知过程中听者需要雇用自己的发音动作来解码”这一理论的话,那么他们2009年的文章[9]则是从反面来说明听者无法发音会影响对这类语音的感知。作者用重复经颅磁刺激(repetitive transcranial magnetic stimulation, rTMS)短暂抑制管唇部发音动作和管手动作的左半球运动皮层活动,来测试听者对特定辅音的类别感知(categorical perception)的反应。塞音的类别感知是人类语音认知的一个重要特点,即人类对不同类别的辅音有非常敏感的认知反应,但是对属于同类辅音中的各种差异则反应不敏感。该文的作者合成四组辅音连续刺激音(continuum stimuli),分别是:

(1) ba-da; (2) ka-ga; (3) pa-ta; (4) da-ga

每组连续刺激音分八步(eight-step continua),逐渐变化声学参数,在短暂抑制左脑运动皮层中的唇部代表点(the lip representation)和手部代表点(the hand representation),看听者对这些辅音的

认定(identification)和区别(discrimination)能力。参加者共30位,英语是他们的母语。表2.02是听者短暂抑制左脑运动皮层某些部位前(Pre-rTMS)和后(Post-rTMS)测试数据的对比统计:

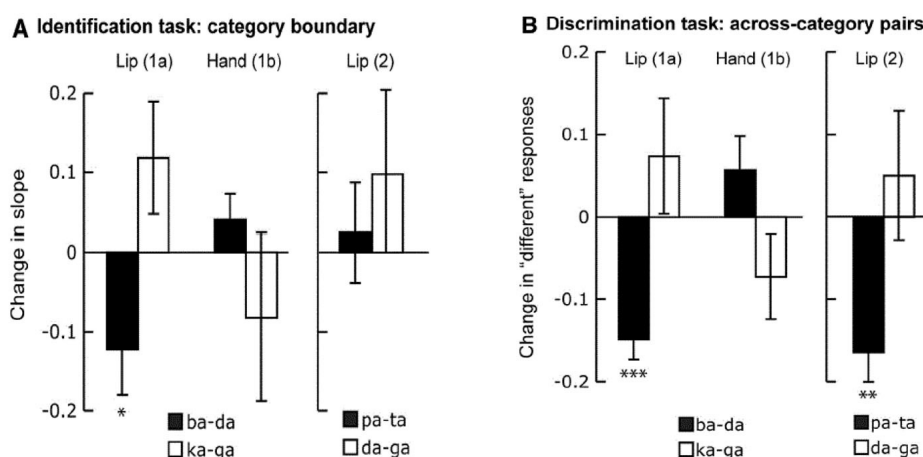


表 2.02

表中的\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$

表 2.02A 分别是四组塞音认定统计表,根据统计数据可以总结如下:

- 1) 在 Lip(1a) 组里,重复经颅磁刺激(rTMS)扰乱左半脑主要运动皮层中的唇部代表点会引起 ba-da 连续体感知斜率的降低,而对 ka-ga 连续体感知不起变化。
- 2) 在 Hand(1b) 组里,重复经颅磁刺激(rTMS)扰乱左半脑主要运动皮层中的手部代表点对听者 ba-da 和 ka-ga 连续体感知都不受任何影响。
- 3) 无论是扰乱唇部代表点还是手部代表点对音类界限位置的感知都不受影响。
- 4) 在 Lip(1b) 组里,重复经颅磁刺激(rTMS)扰乱左半脑主要运动皮层中的唇部代表点对 pa-ta 或 da-ga 连续体类别感知的界限不起变化。

表 2.02B 分别是四组塞音区别统计表,根据统计数据可以总结如下:

- 1) 重复经颅磁刺激(rTMS)扰乱左半脑主要运动皮层中的唇部代表点会破坏 ba-da 和 pa-ta 连续体感 ba 与 da、pa 与 ta 的区别精确度,而 ka-ga 和 da-ga 连续体中 ka 与 ga、da 与 ga 的区别精确度不受影响。
- 2) 每组音类内的区别度则不受重复经颅磁刺激(rTMS)扰乱的影响。
- 3) 重复经颅磁刺激(rTMS)扰乱左半脑主要运动皮层中的手部代表点对每组的区别精度都不产生作用。

重复经颅磁刺激扰乱左半脑主要运动皮层唇部代表点对唇音的认定感知和区别感知都产生影响,虽然这种影响并不足以取消不同音类的类别感知特点,但是会改变听者对音类感知的的时间以及感知错误率的增加。显然对左半脑主要运动皮层发音动作代表点的干扰会对听者的语音感知产生影响。这再一次从不同的角度证明听者左半脑管语音发音动作的运动皮层的活动对感知相应的语音起着十分重要的作用。

### 3.0 肌动理论的修正及其意义

上述 Watkins 等人的四篇研究论文从神经机制这一角度弥补了语音感知肌动理论的缺陷,使肌动理论从假说成为真理。我们可以把语音感知的过程总结如下:

外来的语音首先到达大脑的主要听觉皮层(primary auditory cortex),被分解为不同频段的声

音再转送到附近的维尼克区( Wernicke's area) 进行理解;与此同时,语音信息通过白质弓状纤维束( arcuate fasciculus) 传送到左脑主要运动皮层( left primary motor cortex) 做出相应的“动作”,即通过听者自己的“发音”特性来匹配外来的语音,从而对外来的语音解码。

语音感知的肌动理论认为,听者是通过自己发音诱导出听觉模式来分析和理解外来语音的。学者们也通过许多语音感知实验来证明他们的理论,例如,在/sa/音节中用静音填入/s/与/a/之间,逐渐拉大静音的间距,以至于听者感知为/sta/,如图3.01所示:

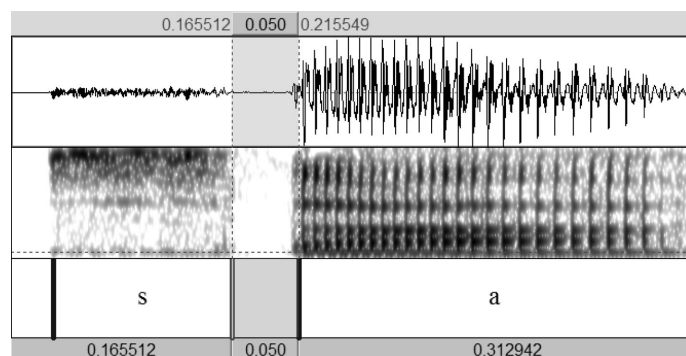


图 3.01

图3.01里/s/与/a/之间的间距到达50毫秒(=0.050秒)左右,听者才会感知为/sta/。中间的间隙之所以被感知为/t/而不是其他的/p/、/k/等塞音,是因为被感知的音段跟前面的音段是相同发音部位,根据协同发音特点,/s/后的塞音音征就应该是同部位的塞音/t/。而这个50毫秒左右的空隙正是相当于人类发/t/这个动作的最短时长。换句话说如果/s/和/a/中间的静音段不到50毫秒,也即做这个动作的时间不够,我们就不会感知这个动作,从而就不能感知这个/t/。

肌动理论假设听者是雇用他们自己的发音体及共鸣腔的知识来对外来语音解码,但是他们没有对听者雇用自己发音体及共鸣腔知识的细节及步骤做出说明。随着神经科学及语言认知科学的发展,现在人们对大脑如何指挥语音发声的认识越来越深入,所以可以对肌动理论略作修正,把语言发音和感知的过程细化,研究语言感知与语言发音哪些阶段有密切的联系。我们认为听者不必通过自己的发音,而是依靠自己的发音信息来解码的。即布罗卡区储存着个人发音动作的信息,当外来的语音信息跟这些既有的信息相匹配,听者才会正确地感知外来的语音。整个语音感知过程在语言运动计划( language motor planning) 和语言运动的编程( language motor programming) 阶段完成,而不是在语言规划的执行( language motor execution) 过程中完成。

语言感知机制与语言发音的机制紧密相关,而语言发音也是一个非常复杂的过程。要发音,发什么样的音是语言运动计划的事情,这个工作由大脑主要运动皮层先做出规划。语言运动的编程则是指大脑运动皮层中发某个语音所需的各种协调运动的编程。根据 Guenther 等人[4]提出的语言发音 DIVA 模式,整个编程既跟前后音段的协调( co-articulation)、制约( constraint)、语音反馈有关,也跟特定语言的音系规则( phonological rules) 等因素有关,最终这些因素的信息通过大脑各功能区再汇总到主要运动皮层,完成语言发音的计划及语言运动的编程两个阶段。语言运动规划的执行是指经过语言运动计划和语言运动编程,神经元激发肌肉收缩和运动改变发音体和共鸣腔,引发发音腔体内的气压变化,产生语音声波的整套运动。显然语音感知跟听者大脑运动皮层中的发音计划、发音编程密切相关,跟语言运动规则的执行无关。也就是说语音感知跟语言发音“肌动系统”( motor system) 的早期阶段紧密关联,如果没有听者语音发音、编程的参与,语音感知就会失败。

肌动理论的提出以及以后神经科学对它的证明至少给语言学研究带来三点重要的启示。

第一,语言认知(感知)具有主观性,任何感知都带有主观性,语音感知也不例外。但是这种主观感知也不是随意的,任凭听者随心所欲去主观发挥;它必须受特定语言规则的制约,所以在一个言语社团,或者在一种语言里这种制约是有强制性的,这样才能保证言语交际顺利地进行。Jeff Elman 和 Jay McClelland [2]曾做过一个很著名的感知实验,充分说明语音感知的主观性以及这种主观感知受特定语言规则制约的特点。这个实验分两个步骤实施,整个实验虽然有点复杂,但是设计十分巧妙。第一步用 /af/ 和 /as/ 放在 /da/-/ga/ 的连续体(continuum)环境前,让英语为母语的被试来测试听到 /da/ 或是 /ga/ 的比率,这时感知对协同发音有补偿作用(compensation for coarticulation),被试会在 /as/ 而不是 /af/ 的环境里对 /ga/ 的感知更多。显然主观的感知补偿在起作用。第二步,把自然英语“abolish”和“progress”两个单词分别抹去最后的擦音 [ʃ] 和 [s],代之以相同的宽波段噪音(broadband noise),所有被试人都把前者认为 [ʃ],后者认为 [s],也即仍被感知为“abolish”和“progress”两个词;然后他们又把带有词尾噪音的这两个词再次放入 /da/-/ga/ 的连续体环境前,看看能否可以像真的“abolish”和“progress”那样,会对后面的 /da/-/ga/ 连续体产生如第一步那样的感知补偿效应,测试的结果证明是发生的。说明以英语为母语的人相同的噪音放入不同的环境中,会随词汇的不同而感知为不同的音,而这个噪音会像“幽灵”那样在不同的环境里变换角色对其他音产生影响。在这里,相同的噪音感知为不同的音类并非是客观的,而是主观的感知;但是这种主观感知受英语词语读音规则的支配。因为英语里“abolish”词末不会出现清擦音 [s],在“progress”词末也不会出现清擦音 [ʃ]。当这个幽灵进入具体词的环境里,被人们感知为“真的”语音,它也会如真的 [s] 或 [ʃ] 那样对邻近的语音产生补偿感知效应。这一研究充分体现了语音感知的主观性,以及这种主观感知受具体语言规则制约的特点。

第二,人类的语言发音以及语音感知是独特的,跟其他动物发出的声音和感知不同。人类语言发音和语音感知的独特性可以从以下几个方面来阐述:其一,其他动物的发音体和共鸣器官与人类的具有本质上的差异,它们的神经机制以及大脑结构也跟人类的显示出巨大差异。比如,最接近人类的大猩猩共鸣腔里会厌与软腭是相连的,所以发任何音都会引起鼻腔共鸣,而人类会厌与软腭分离,软腭就有可能根据需要上抬和下垂,发出更为丰富多彩的音来。比起其他动物来,人脑的体积更大,具有更多的神经元和皮质功能。管运动区、感觉区大脑皮质回路是人类特有功能增加的投射,其中嘴的运动投射皮层有不成比例的增大,这跟人类用嘴说话、呼吸、吃等运动量大、动作频繁等有直接的关系。所以,动物对它们的“语言”也有感知机制,只是因动物发音体、共鸣腔、脑结构等跟人类存在巨大的差异,“语言”感知的特点就会有很大的区别。其二,Mann 和 Repp 的文章[8]指出语音感知的补偿机制,即在连续的音段和语句中人类发音会发生音段间的协同发音(coarticulation)现象,听者则用感知补偿机制来克服由于协同发音产生的混淆。这也是人类语言发音和感知的特点,此现象并没有在任何动物“语言”里发现。其三,语音的感知具有主观性,这种主观性受特定语言规则的制约,语言规则是指特定语言里的音系结构规则、构词法特点、语义特点、句法规则等,对非语言的动物声音来说,既然不具备人类语言的这些特征,自然就不会有人类语言那样的发音及感知特点。

第三,根据前述语音感知的肌动理论,听者是雇用了自己大脑运动皮层中的语音发音计划及编程对外来的语音解码;换句话说,外来的语音只有跟自己储存的发音信息相匹配,听者才会认同这个(些)外来的语音。这一理论很好地解决了人类语音发音变异与语音感知之间的困惑。正如 Weinreich, Labov and Herzog [16]指出,语言变异是人类语言的本质,没有变异,语言就失去了它的功能,因为许多重要的社会功能以及语言的演变都是建立在语言有序变异(orderly heterogeneous)的基础上的,可以说没有变异的语言是不可想象的。每个人发音器官存在着差异,男人与女人有

差异(成年女性的共鸣腔只有成年男性的70%左右)老人与小孩有差异,发音体形状不同,共鸣声学数据自然不同;即使是同一个人,在不同的时间点发同样的音,由于语速、语言环境以及其他非语言因素差异,声学参数也会有很大的变异。图3.02是 Peterson 和 Barney 1951 年 [11] 针对 76 位美式英语发音人十个元音第一共振峰对应第二共振峰所画的声学元音舌位图:

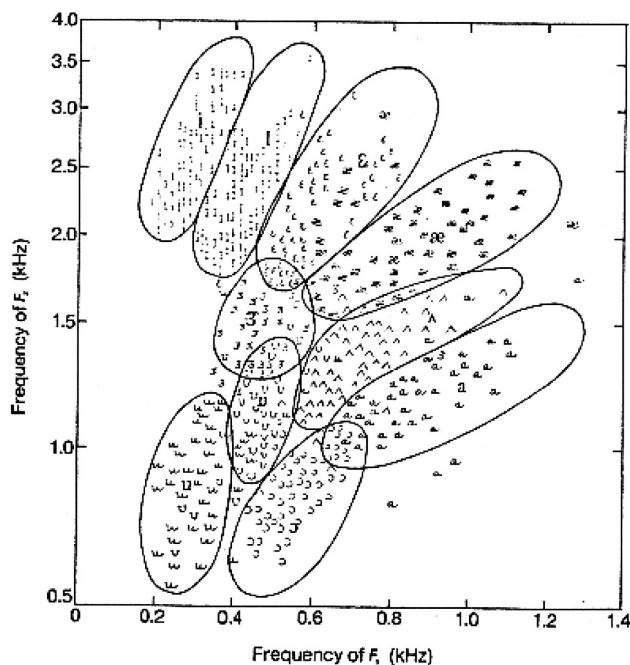


图 3.02

由图可以看出,同样的元音在不同人的口里是有变异的,有的音声学数据的变异还相当大,有的时候不同元音的声学特点会重叠在一起。图3.03则是两位标准的德语发音人(男67岁,女68岁)发相同的100个句子,从中抽取四个松元音的第一共振峰、第二共振峰的声学数据,画成的声学舌位图(左为男性发音人,右为女性发音人。此图为笔者根据 the Kiel Corpus of Read Speech 数据库提取数据用 R 统计软件画成)。

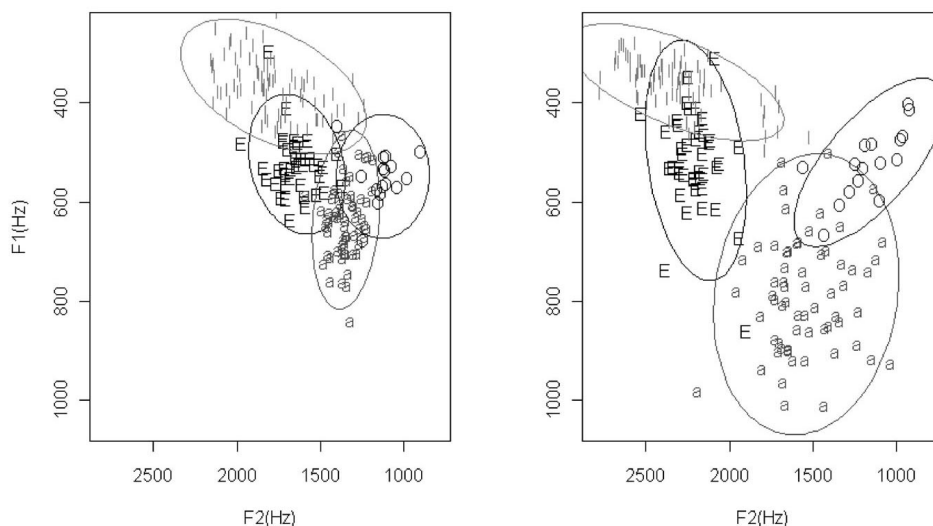




图 3.03

从图 3.03 可以看出,即使是同一个人,具有相同的发音体,但是在不同的语言环境里以及不同的语速中,发同样的音还是有相当大的变异。对上面两位发音人的第一、第二共振峰取平均值,然后叠加成图 3.04:

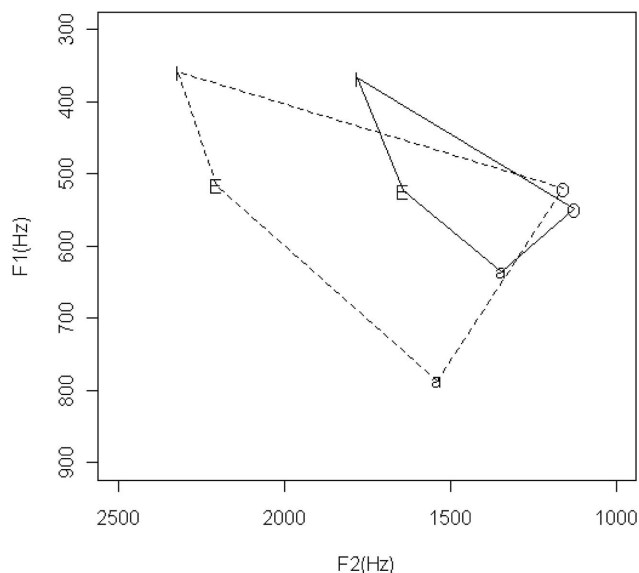


图 3.04

图 3.04 中实线、虚线分别为男性、女性发音人四个松元音第一共振峰、第二共振峰的声学数值距离。可以看出女性元音间的声学数值距离更大,而男性四个元音的声学图只占女性的一个角,仅凭声学数据无法找到男女间对应的元音。图 3.02、图 3.03 和图 3.04 很明确地告诉我们,面对如此复杂的变异,如果听者全凭客观的声波特点(声学数值)来感知或识别外来的语音,语言交际一定会崩溃。语音感知的肌动理论告诉我们,听者是启动了自己的发音动作来匹配外来语音,也可以说是根据自己的发音计划、发音编码以及特定语言的音系规则来对外来的语音作归一化(normalization)处理,根据自己的发音信息就可以过滤掉由各种因素造成的语音变异,从充满变异的声学信号里提取一致的、符合语言规则的音位信息,来感知语音,从而达到语言交际的目的。

#### 4.0 结语

肌动理论的提出是语音感知研究里程碑式的成就,随着脑科学及神经科学的进展,人类对语言认知的特点也会越来越深入。我们相信肌动理论是正确的;听者语音感知的机制和听者自己的发音机制是紧密相连的;人类语言的发音和语音的感知是独特的,跟其他动物的发声和感知是不同的;听者利用自己的语言发音机制来过滤外来的语言变异,以达到语言交际的目的;目前电脑语音识别的设计并没有认识到这一点,所以为了提高电脑语音识别率,必须改变思路,充分考虑人类语音感知的特点,这样电脑语音识别才会走出困境。

#### 参考文献:

- [1] Bell, A. M. *Visible speech: The science of universal alphabets; or self-interpreting physiological letters for the printing and writing of all languages in one alphabet; elucidated by theoretical explanations, tables, diagrams,*

and examples [M]. London: Simpkin, Marshall, 1867.

- [2] Elman, J. L. & J. L. McClelland. Cognitive penetration of the mechanisms of perception: Compensation for co-articulation of lexically restored phonemes [J]. *Journal of Memory and Language*, 1988, 27: 143 – 65.
- [3] Guenther, F. H. Cortical interactions underlying the production of speech sounds [J]. *Journal of Communication Disorders*, 2006, 39: 360 – 365.
- [4] Guenther, F. H., Ghosh, S. S. & J. A. Tourville. Neural modeling and imaging of cortical interactions underlying syllable production [J]. *Brain & Language*, 2006, 96(3): 280 – 30.
- [5] Halle, M. & K. N. Stevens. Analysis by synthesis [A]. W. Wathen-Dunn and L. E. Woods. *Proceedings of the Seminar on Speech Compression and Processing* [C]. AFRC-TR – 59 – 198, December 1959, vol. II, paper D7.
- [6] Liberman, A. M. et al. Perception of the speech code [J]. *Psychological Review*, 1967, 74: 431 – 461.
- [7] Liberman, A. M. & I. G. Mattingly. The motor theory of speech perception revised [J]. *Cognition*, 1985, 21: 1 – 36.
- [8] Mann, V. A. & B. H. Repp. Influence of vocal context on perception of the [f] – [s] distinction [J]. *Perception and Psychophysics*, 1980, 28: 213 – 228.
- [9] Moöttöinen, R. & Kate E. Watkins. Motor Representations of Articulators Contribute to Categorical Perception of Speech Sounds [J]. *Journal of Neuroscience*, 2009, 29(31): 9819 – 9825.
- [10] Ohala, J. Listeners hear sounds not tongues [J]. *Journal of Acoustic Society of America*, 1996, 99: 1718 – 1728.
- [11] Peterson, G. E. & H. L. Barney. Control Methods Used in a Study of the Vowels [J]. *Journal of Acoustic Society of America*, 24(2): 175 – 184.
- [12] Moöttöinen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J. & M. Sams. Perceiving identical sounds as speech or nonspeech modulates activity in the left posterior superior temporal sulcus [J]. *Neuroimage*, 2006, 30: 563 – 569.
- [13] Stevens, K. N. & S. E. Blumstein. Invariant cues for places of articulation in stop consonants [J]. *Journal of Acoustic Society of America*, 1978, 64: 1358 – 1368.
- [14] Watkins, K. E., Strafella, A. P. & T. Paus. Seeing and hearing speech excites the motor system involved in speech production [J]. *Neuropsychologia*, 2003, 48: 989 – 994.
- [15] Watkins, K. E. & T. Paus. Modulation of Motor Excitability during Speech Perception: The Role of Broca's Area [J]. *Journal of Cognitive Neuroscience*, 2004, 16(6): 978 – 987.
- [16] Weinreich, U., Labov, W. & M. Herzog. Empirical foundations for a theory of language change [A]. W. Lehmann and Y. Malkel. *Directions for historical linguistics* [C]. Austin: University of Texas Press, 1968. 95 – 188.

基金项目：国家社科基金重大项目“汉语方言自然口语变异有声数据库建设”(批准号 12&ZD177)。

收稿日期：2014 – 05 – 12

作者简介：陈忠敏(1962 – )，男，上海人。美国加州大学伯克莱校区博士，教授。研究方向：实验语音学、历史语言学、汉语方言学。