语音感知的特点及其解剖生理机制

陈忠敏

摘　要　本文按听觉器官以及听觉器官以上神经系统的解剖生理构造来讨论语音感知的特点以及它们的解剖生理机制。听觉器官并不是被动地、客观地接受外来输入的语音，而是对传入的语音进行了主观的改造和修饰。改造和修饰的特点及结果跟听觉器官的解剖生理构造密切相关。进入听觉器官以上的听神经系统以及大脑高级认知皮层以后，还要做更为高级的语音匹配认知过程。其中的音段切分与大脑神经的自主振荡范围有关。听者利用自己大脑运动皮层中的语音发音计划及编程进行语音解码，同时通过语音与自己储存的发音信息进行匹配，据此去除各种语音变异。

关键词　语音感知，基频消失，掩蔽效应，神经自主振荡，肌动理论

The Characteristics of Speech Perception and Their Anatomical and Physiological Mechanisms

CHEN Zhongmin

Abstract In this paper, the characteristics of speech perception and their anatomical and physiological mechanisms are discussed in terms of the anatomical and physiological structures of the auditory organs and the nervous system above the auditory organs. The auditory organs do not passively and objectively accept the incoming speech, but subjectively changes and modifies the incoming speech. The characteristics and results of modification are closely related to the anatomical and physiological structures of the auditory organs. More advanced cognitive processes of speech matching are also required when speech sound enters the nervous system above the auditory organs and the high cognitive cortex of the brain. The segmentation of speech is related to the autonomic oscillation range of the brain nerve. Listeners employ the speech planning and programming in the motor cortex of their brain to decode auditory patterns. At the same time, the auditory patterns only match the stored speech information such that various speech variations can be removed.

Keywords Speech perception, Missing fundamental, Masking effect, Autonomic oscillation of the brain nerve, Motor theory

1. 概说

言语链（见图 1）分为说者（speaker）、中间传播（transmission）和听者（listener）三方面，很好地说明了言语产出、言语传递以及言语感知三个不同阶段的各自功能。言语产出首先要在说者的大脑里做出言语规划以及言语产出的各种编程，然后通过神经元放电将发音信息传递给发音体，这部分是大脑的言语指令部分（linguistic level）；神经元放电刺激肌肉让发音体和共鸣腔做出相关动作。这是言语产生的生理发音部分（production level），随着发音体的运动和共鸣腔的变化，共鸣腔里空气粒子会有压力变化从而产生声波，这是语音的传送部分（transmission level）；声波向外传送，一路作为反馈信号进入说者的听觉接受器官，再传送到说者的大脑皮层，用来矫正后续的发音指令及动作（其实还有一路骨传导反馈图 1 没有显示）；另一路声波传送到听者的听觉器官，声波经外耳道、中耳以及内耳听觉器官的传送和修饰后，刺激听神经形成神

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

经脉冲信号，再传送到大脑听觉皮层来感知和解码，这个过程是语言的感知阶段(linguistic level)。

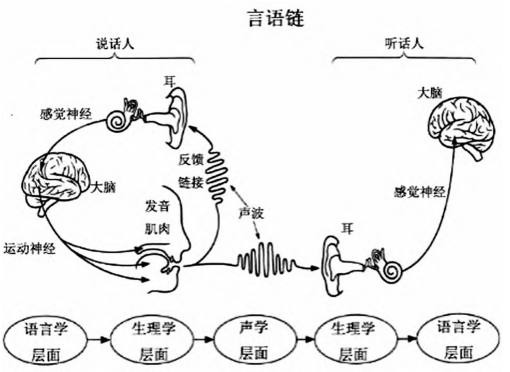


图1　言语链（引自 Denes & Pinson［2］）

根据言语发声、传递、感知的三个阶段，语音学可以分为发音语音学（Articulatory phonetics）、声学语音学（Acoustic phonetics）、听觉语音学（Auditory phonetics）。其中发音语音学的研究历史最早。达·芬奇（Leonardo da Vinci，1452－1519）曾留有一幅喉头解剖图和一幅发音器官纵面解剖图，这可能是最早的发音器官解剖图。早期研究语音产生及发音部位的大多是医生或病理语音学家，因为他们懂得发音器官的解剖知识，同时也因为有临床病理言语诊断和治疗的需要。发音语音学的集大成成果应该是国际语音学会成立（1886）后制定的国际音标表（1888年发布第一版，最新一版是2015年的）。国际音标表及标音法的不断完善也标志着发音语音学的发展和成熟。属于语音传递阶段的声学语音学研究虽然起步也较早，比如物理学家赫姆霍兹（Hermann von Helmholtz，1821－1894）发展了用数学公式推导出不同腔体的共鸣频率和人类发声的基频及各谐波频率的关系，赫姆霍兹被誉为第一位声学语音学家。但是声学语音学的全面发展不可否认是从20世纪40年代语图仪的发明开始的。半个多世纪以来，声学语音学取得长足的发展，已经相当成熟，其中的标志就是人们已基本掌握各种元辅音音段及超音段的声学特征，语音声学分析的正确性已经在语音合成领域得到印证。与发音语音学、声学语音学相比，听觉语音学研究成果相对滞后，虽然赫姆霍兹在1875年的著作中第一次提出声音感知的赫姆霍兹位置理论（the Helmholtz's place theory），但是位置理论及其特点一直到五六十年代经von Békésy的研究才得以确定［20］。听觉语音学研究相对滞后有其客观原因；观察、测试和分析人类听觉器官，特别是听觉器官以上的听神经系统的活动有技术、方法等层面的限制。进入五六十年代，这一情形有了很大的改观，特别是到了90年代，随着脑影像技术的进步，人们可以直接观察言语活动时大脑神经的激活区域，听觉语音学研究，特别是听觉器官以上的神经语音学研究水平才达到前所未有的高度。本文通过分析听觉器官和听觉器官以上的听神经解剖生理机制来讨论语音感知的特点。

2. 外耳道、中耳道的解剖生理结构及在语音传导中所起的作用

大部分哺乳动物的听觉器官从功能上是十分相似的，不过人类的听觉器官各子系统的长度及特点与其他哺乳动物不同。外周听觉器官分为外耳（outer ear）、中耳（middle ear）和内耳（inner ear）三部分。成人外耳道的长度大概是2.3厘米(0.023米)，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1/4）公式算出这一长度管子的最强的第一共振峰值：340/（4×0.023）＝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)［18］，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是外耳（outer ear）、中耳（middle ear）以及两者叠加（sum）声音敏感图［18］。

可以看出人耳对2000赫兹到5000赫兹的声音反应最为敏感，到达或超过1万赫兹的声音敏感度会急剧下降［9］。声音敏感图与我们常见的听力等响曲线基本是

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

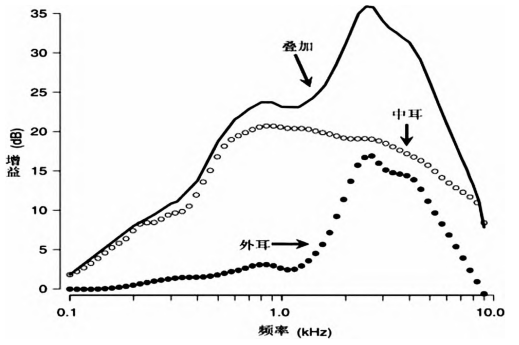


图 2　外耳、中耳以及两者叠加声音敏感图

吻合的，根据等响曲线（equal-loudness contours）图，人类听觉敏感度在 4000 赫兹处最敏感，1—2 个 dB 响度的声音就能听到，在 500 赫兹处响度要调高至 15dB 才能听到，50 赫兹则需要达到 40dB 以上才能听到。可是语言里 2000 赫兹以下的低频也是非常重要的，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 1000 赫兹以下，很多元音的第二共振峰也不到 2000 赫兹。如果 2000 赫兹以及低于 2000 赫兹的声音的音量不高，敏感度就会不高，将严重损害言语交际。低频区要达到 2000 赫兹到 5000 赫兹一样的敏感度，就要增加低频区的音量。这个缺陷正好可以由发音体来弥补。图 3 是元音产生的声源 + 共鸣原理示意图（据 Diehl［3］文章修改）。

图 3（a）声门谱（glottal spectrum）是指声门的瞬间频谱，瞬间频谱的纵轴代表振幅，单位是分贝（dB），横轴是频率，单位是千赫兹（kHz）。从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，常态发声态（modal voice），频率每下降一个倍频程，振幅就下降 12dB。图 3（b）是声道滤波响应（vocal tract filter response）。不同形状的共鸣腔具有自己特定的共鸣频率（resonances）。图 3（b）中的三个共振峰频率（formant frequencies）相当于均匀共鸣腔管子，类似发元音/ə/时的数值，第一共振峰（F1）为 500 赫兹，第二共振峰（F2）为 1500 赫兹，第三共振峰（F3）为 2500 赫兹。声门频谱经过声道的滤波，就产生了图 3（c）能从分析仪看到的元音共振峰的输出频谱图。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。在图 3 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，通过共鸣腔滤波的语音也具有这个特征，输出的第一、第二共振峰的能量远远高出第三、第四、第五等共振峰的能量，发音时具有的低频能量高，高频能量低的特点，正好抬高了人耳接受低频声音的敏感度，以保证人耳对 20—5000 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，保证言语交流有效的经典例子。5000 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 5000 赫兹以下；声调语言里男女声调的频率（基频）一般在 20—400 赫兹范围内变化；语言中塞音/k/爆破点（burst）能量集中区在 2000—4000 赫兹左右；/p/和/t/的区别是爆破点以 3000 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 6000 赫兹以上的能量区别不同擦音。所以笔者认为发音器官与听觉器官的特殊构造，造就了语音输出与接收声能可以互补，保证了语音声能能有效地传递到内耳［25］。人类能听到频率的范围是 16—2.2 万赫兹，其他非人类的哺乳动物与人类不同。狗的听觉范围介于 15—5 万赫兹，猫的听觉则在 60—6.5 万赫兹，狗熊和猕猴听觉的最高极限是 8 万赫兹，豚鼠听觉的最高极限是 10 万赫兹，鼠和鼩鼱等的听力上限可高达 9 万—12 万赫兹，尖耳鼠蝠听觉的最高极限甚至可以达到 25 万赫兹。与人类相比，其他动物的听觉频率分布范围要广得多；听觉频率的极限也高得多，不是集中低频范围。这样使得人类与其他动物在声音的感知上存在差异。图 4 是人类、鬈毛狗、老鼠三种哺乳动物的等响曲

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

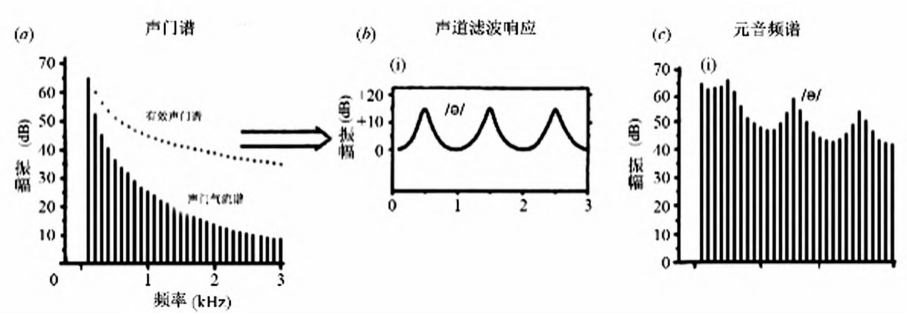


图 3　元音产生的声源 + 共鸣示意图

线图（响度级）。鬈毛狗的等响曲线比人类要宽广得多，而老鼠对声音的敏感区主要集中在 1 万赫兹以上的，反而对 1 万赫兹以下的声音不敏感［4］。

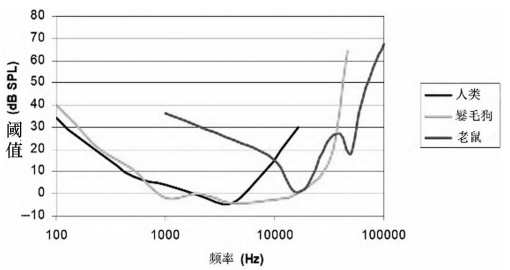


图 4　人类、鬈毛狗、老鼠等响曲线

由于不同哺乳动物的听觉器官具有不同的等响曲线，相同声音传入，就会有不同的声音感知。所以人类语言必须配备人类的听觉器官才是言语交际成功的关键。

3. 低频敏感度与语言的元音格局

外来语音经外耳和中耳修饰后传入内耳，引起内耳中基底膜（basilar membrane）的运动。成人内耳耳蜗（cochlea）的长度大约是 35 毫米，中间的基底膜约长 31 毫米，基底膜从卵形窗（oval window）到蜗顶（apex）分布着 23500 多个听觉毛细胞。基底膜的运动激发听觉毛细胞放电再刺激听神经，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是线性的，靠近耳蜗蜗顶（apex）的基底膜宽而软，用来分析低频，低频的带宽（bandwidth）窄，分辨率高；靠近蜗底（base）处的基底膜窄而硬，分析高频。高频的带宽宽，分析分辨率低。从蜗顶往蜗底基底膜频率分析是从低（≈20 赫兹）到高（≈22000 赫兹）分布的，但低处的带宽和高处的带宽不是线性分布的。图 5 是把两圈半的基底膜拉直从低频到高频分析的不同带宽（据 Johnson［9］: 89）。

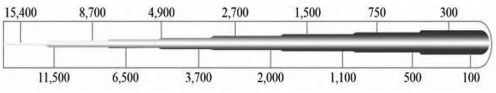


图 5　拉直基底膜频率分析示意图

相同的距离差在蜗顶低频处是 200 赫兹（300—100），放在蜗底高频处却是 3900 赫兹（15400—11500）。也即分析频率越高，对应的分析带宽也就越宽。基底膜分析带宽跟分析频率大致有这样的关系：带宽≈分析频率/8。根据此公式推导，1000 赫兹处的带宽大概是 125 赫兹，5000 赫兹处的带宽大概是 625 赫兹。换句话说，5000 赫兹处的带宽是 1000 赫兹处的 5 倍！基底膜分析频率的非线性（频率低细腻，频率高粗犷）特点保证了人类语言音类之间有分散性、区别性。一种语言里音类要保持区别性，音类之间就必须有最大的空间距离，这是 Liljencrants 和 Lindblom 在 20 世纪 70 年代时提出的语言中元音音类的 Dispersion theory 的基本思想［14］。笔者曾根据著名语音学家 Daniel Jones 所发 7 次正则元音（cardinal vowels）的第一

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

共振峰（F1）、第二共振峰（F2）平均数据画出两幅声学元音图。表 1 是 Daniel Jones 发 7 次正则元音 F1、F2 的平均值。

表 1　　Daniel Jones 所发正则元音 F1 和 F2 平均数值

	Vowel symbol	Mean F1（Hz）	Mean F2（Hz）
1	i	266	2581
2	e	376	2213
3	ɛ	588	1910
4	a	929	1688
5	ɑ	650	940
6	ɔ	522	932
7	o	354	724
8	u	248	490

笔者用了两种计算方法来画出元音声学舌位图。第一种方法的结果是图 6a。从 F1 和 F2 的刻度可以看出，F1 从 500 赫兹到 2000 赫兹，F2 从 500 赫兹到 2500 赫兹的距离是线性等量排列的，第二种方法的结果是图 6b。根据人的内耳基底膜分析频率的非线性特点（带宽≈分析频率/8）重新排列 F1 和 F2 的刻度，F1 从 200 赫兹到 1000 赫兹，F2 从 500 赫兹到 2500 赫兹，可以看出它们的距离是非线性，也不是等量排列的。

图 6a 是 F1 × F2 线性频率声学元音图，音与音的距离近，也跟我们常见的元音舌位图不配；图 6b 根据内耳低频到高频分析的不同带宽所作的非线性频率声学元音图，音与音的距离远，与我们常见的元音舌位图吻合。不同的音类之间形成最大的距离感，也就把音类之间的混淆度降到最低，这是符合人类言语交际和传递具有最佳清晰度的要求。人类内耳基底膜非线性频率分析有其独特性，跟其他哺乳类动物不同。图 7 是不同频率刺激物在人类、猕猴、猫三类哺乳动物内耳基底膜的延迟反应时间［10］。

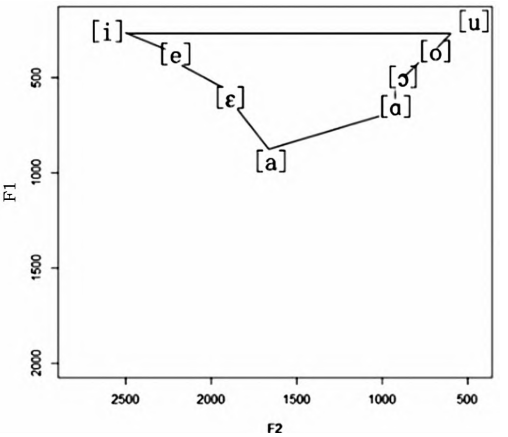


图 6a　F1 × F2 线性频率声学元音图

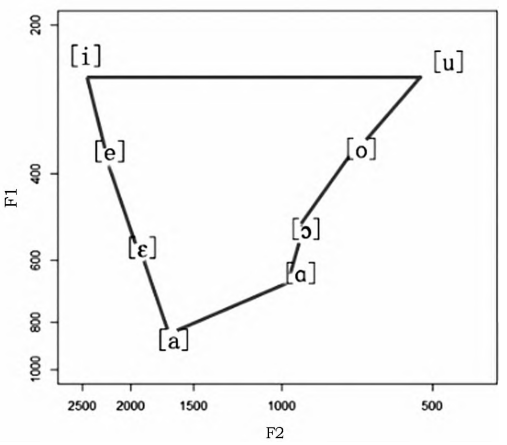


图 6b　F1 × F2 非线性频率声学元音图

这种延迟反应的时间差异可以换算出基底膜不同频段的分析带宽。从图 7 可以看出不同频段人类、猕猴、猫比较，人类都是延迟时间最长，猕猴次之，猫最短。也就是说明：第一，人类内耳基底膜不同频段的分析带宽与其他动物的是不同的；第二，人类内耳基底膜不同频段的分析带宽都要比猕猴、猫的窄。如果分析频率的带宽不同哺乳动物有差异，那就是相同的语音输入，不同的哺乳动物会有不同的语音感知，这也再次证明了人类的语言感知是独特的。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

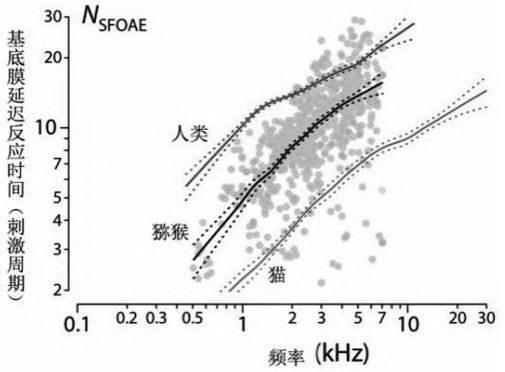


图 7　不同频率刺激物在人类、猕猴、猫三类哺乳动物内耳基底膜的延迟反应时间

4. 基频丢失与音高感知

基频（fundamental frequency）是一个物理量，指周期性复合波中的最低频率。音高（pitch）是一种感知量，指人对周期性复合声波所产生的一种感知。周期性的复合波里每个谐波（harmonic）与基频有整数倍关系。人类的音高感知是对复合波整合后的感知，而不是对每个谐波分离感知。比如图 8a、8b 是笔者用 Praat 软件合成两列谐波数不等的复合波，但是最长的周期都是 5 毫秒（ms）振动一次，其频率，也叫做基频，就都是 200 赫兹（ =1/0.005 秒）。

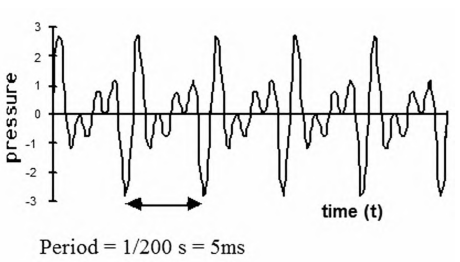


图 8a　周期为 5ms 的复合波 a

尽管波形不同，但是基频一致，都是 200 赫兹，因此人们认为它们有相同的音高，可见基频与音高有十分密切的关系。

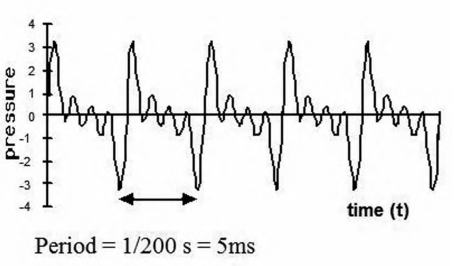


图 8b　周期为 5ms 的复合波 b

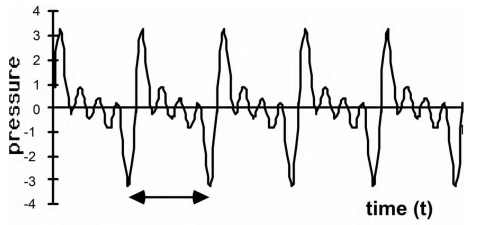
在语言里音高的变化和不同有十分重要的意义。说话者的音高不同可以辨别童声、女声、男声等；音高也负载语调、重音等语言信息；在声调语言里音高的不同可以区别词义。

长期以来音高的感知生理机制有两种理论，一种是基底膜位置（place representation）理论，另一种是放电时间（temporal representation）理论。基底膜位置理论其实还分两部分内容。第一部分内容是说不同频率引起基底膜的不同位置的运动，由此激发相应位置的听觉神经兴奋。第二部分内容是由第一部分内容推导得出。即：基频引起基底膜的运动从而激发听神经兴奋。位置理论第一部分的内容早已得到证明，不过第二部分内容存在争议，因为在复合波里即使基频不存在，人们仍能感知到基频。图 9 是笔者用 Praat 软件合成的复合波，上边是波形图，下边是它的谐波分析。

可以看出这列波是每隔 5 毫秒（ms）做一次大振动，大振动中有很多频率高的小振动，所以下边的频谱分析可以看出最低的一个谐波（harmonic），又叫做基频（fundamental）是 200 赫兹，第二、第三、第四谐波分别是 400 赫兹、600 赫兹、800 赫兹等，谐波之间相隔 200 赫兹。这列波我们可以感知到 200 赫兹的音高。在图 10 里，在上述复合波里滤掉 200 赫兹的基频，人们仍能感知到 200 赫兹的音高。

这种现象叫做基频丢失（missing fundamental）现象。基频丢失但是人们仍能感知这个丢失的基频。事实上如果用滤波抹去这个声音的所有谐波，只保留 1400

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net



Period = 1/200 s = 5ms

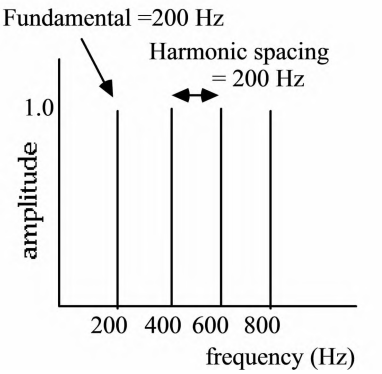
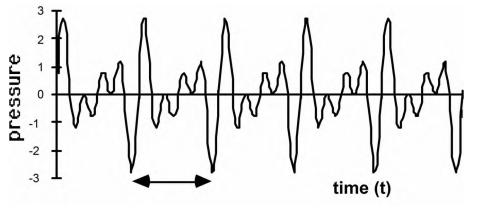


图 9　有 200 赫兹基频的复合波



Period = 1/200 s = 5ms

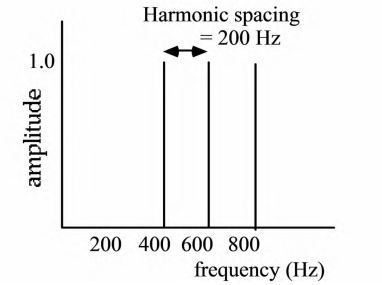


图 10　无 200 赫兹基频的复合波

赫兹、1600 赫兹、1800 赫兹、2000 赫兹中段谐波，人们仍然能感知到 200 赫兹的音高。显然音高的感知并不是基频引起基底膜的运动激发相应位置的听神经兴奋。于是放电时间理论就应运而生。放电时间理论基于以下假设：音高的感知与引起神经兴奋的模式，也即与时间有关。神经兴奋往往出现在基底膜运动的特定相位（锁相，phase locking）上，神经兴奋连续发放的时间间隔就是波形周期的整数倍，人们就是根据这个整数倍来换算出基频的，从而再感知为特定的音高［17］。图 11 是英国伦敦大学学院所设计的内耳基底膜振动模型。

在基底膜 1230 赫兹位置上做最大振动的正弦波其神经兴奋每次间隔时长都是 1/1230 秒，在 14 毫秒的时长里可以清晰看出有这样 16 个间隔，人们就是根据神经兴奋的间隔时间来换算出基频的。根据连续神经兴奋间隔来换算基频必须要有一段时间，如果时间不够，换算就会不正确，时间与频率成反比关系，时长长频率就低，反之则频率高，所以这种换算一般是发生在低中频段（50—4000 赫兹）。图 12 是频率差别阈值（difference limen for frequency，DLF）与中心频率的函数（对数坐标）关系［15］。

DLF 阈值越小，表示中心频率的分辨能力越高，反之，阈值越高，中心频率的分辨能力越差。每根实线上的数值是表示脉冲声的间隔时长，单位为毫秒（分别是 6.25 ms、12.5 ms、25 ms、50 ms、100 ms、200 ms）。可以看出脉冲声越间隔时长长（如 100 ms、200 ms），分辨能力越高。同时也可以看出不管什么时长，在 1000—2000 赫兹这一频段分辨能力最好，但是到了 5000 赫兹左右，DLF 阈值急剧升高，分辨率就差。所以从图 12 可以看出人们对音高感知的灵敏度有两个因素决定，一个是频率，另一个是时长。频率因素是由基底膜作最大振动位置所决定；而时长因素是指在基底膜的特定位置上引起神经兴奋的时间。从图 12 还可以得出这样的结论：要有最佳的音高感知，最好是 2000 赫兹以下，特别是在 1000—2000 赫兹的位置上引起神经兴奋，且这种神经兴奋具有一定的时长。

复合波音高的感知主要由分辨谐波

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

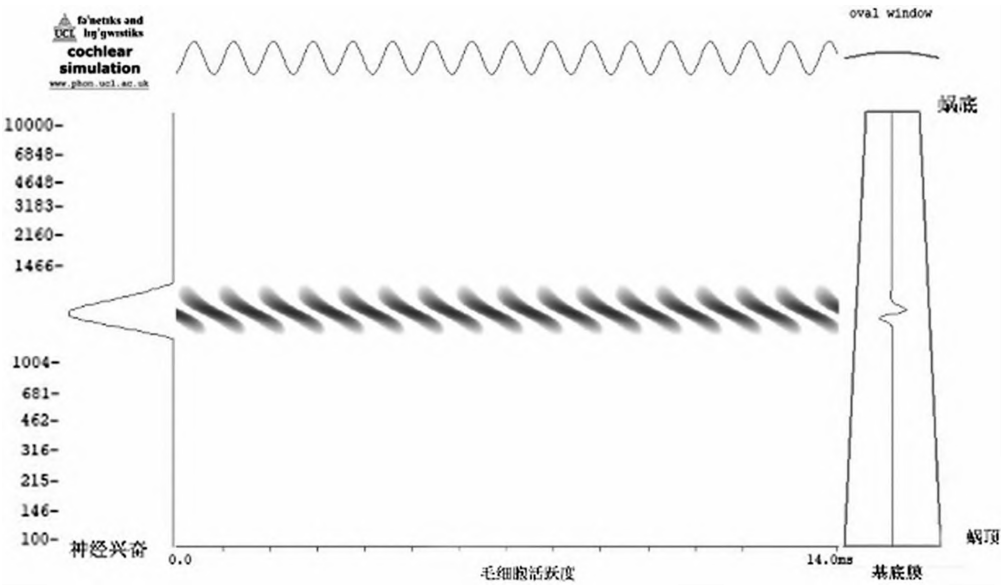


图 11　内耳基底膜振动模型

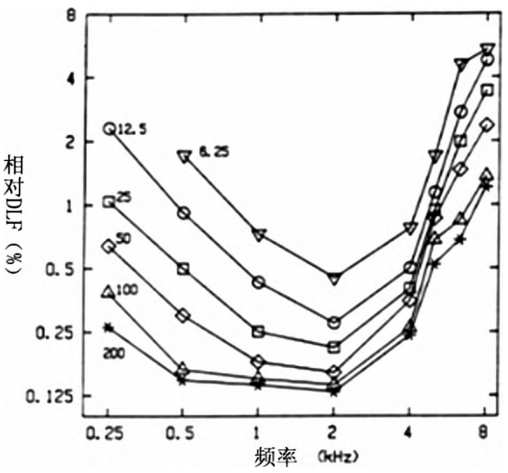


图 12　频率差别阈值与中心频率的关系

(resolved harmonics) 范围内神经兴奋的时间间隔决定。这也是人类语言音高感知最为重要的音征。根据语音音高对应的基频范围以及基底膜带宽的特性，能分辨基频的谐波大致是第四到第八谐波范围内。根据基底膜带宽的特性，到第八谐波 (= 1600 赫兹) 都是分辨的谐波，所以每个谐波通过基底膜带宽滤波后的结果就是产生 200 赫兹正弦波。这就是所说的能分辨基频的谐波。第八谐波后由于基底膜的带宽变粗以至于无法分辨单个 200 赫兹的谐波，通过每个滤波带宽出来的结果就不是一个个谐波，而是一群谐波，所以叫做不能分辨谐波 (unresolved harmonics) 。如果成人男性的平均基频是 125 赫兹，第四谐波到第八谐波的范围就是 500—1000 赫兹；成人女性的平均基频大概是 250 赫兹，那么第四谐波到第八谐波的范围就是 1000—2000 赫兹。所以内耳基底膜 500—2000 赫兹是人类语言音高的最为重要的分辨区。500—2000 赫兹的范围其实是语言绝大部分元音第一、第二共振峰的范围。共振峰就是提升此处频率的能量，看来音高的感知还跟能量的提高有关。在真实的言语交际环境里，低频噪音往往会掩蔽基频，但是人们仍能感知音高，人们的音高感知主要是靠在 500—2000 赫兹频段内的神经兴奋时间间隔来获取。如果 500—2000 赫兹区域外毛细胞或内毛细胞受损必将对音高的感知产生非常不利的后果。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

5. 掩蔽效应与频率的选择特点

掩蔽效应是指一种频率的声音掩蔽另一种频率的声音，从而带来频率的选择性。听觉器官感知声音的重要特点。掩蔽效应与频率的选择有直接关联。声音的掩蔽有同时掩蔽（simultaneous masking）与非同时掩蔽（nonsimultaneous masking）两种，其中同时掩蔽效应对一种语言的元音选择有十分重大的意义。图 13 是同时掩蔽效应（据 Stevens［19］: 232 改画）的图示。

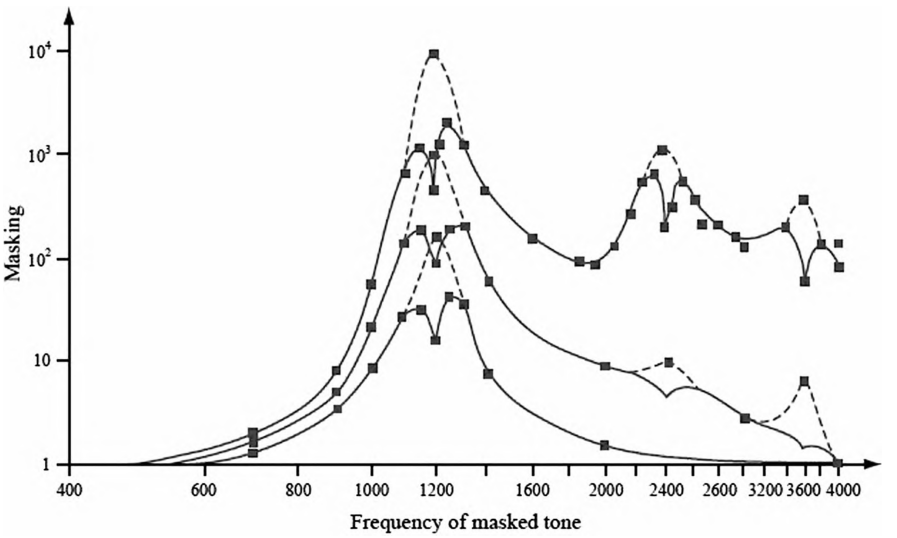


图 13　同时掩蔽效应示意图

图 13 有三列音量不等的掩蔽音（虚线代表），它们都是周期性复合波，第一谐波是 1200 赫兹，第二谐波是 2400 赫兹，第三谐波是 3600 赫兹。掩蔽音由于音量大，可以掩蔽与它频率接近但音量等级较低的两个音（即图 13 中虚线两边的实线）。现在知道这是基底膜外毛细胞的作用。图 14 是笔者根据掩蔽原理所作的外毛细胞作用于内毛细胞的抑制影响层级示意图。

图 14 上图横轴是代表基底膜蜗顶（低频）往蜗底（高频）方向延伸的位置，纵轴表示振幅的大小。基底膜振动带动外毛细胞（图 14 中间插有纤毛的一组圆筒代表外毛细胞）振动，传入基底膜的频谱振幅是三个圆润且坡度较为平缓的“山峦”，如图 14 上半部分。外毛细胞的机械运动转变为内毛细胞的兴奋放电中间有个掩蔽效应，音量大的掩蔽邻近音量较

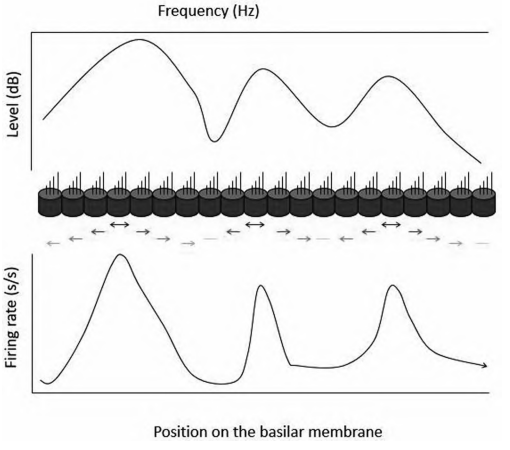


图 14　外毛细胞作用于内毛细胞的掩蔽层级示意图

弱的频率，外毛细胞下的箭头表示掩蔽效应的方向。掩蔽效应从三个共振峰的峰值

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

沿左右两边下降，离峰顶越远掩蔽效应越弱。外毛细胞掩蔽效应起作用就会产生图 14 下半部分三个尖锐且坡度陡峭的“尖峰”。图 14 下半部分是内毛细胞兴奋放电的示意图，横轴与图 14 上半部分一样是表示基底膜从蜗顶往蜗底方向的位置，纵轴则是内毛细胞兴奋放电的激发率（firing rate）。掩蔽效应在语音上使得所选择的频率的振幅峰值尖而敏锐，所选择的目标频率共振峰峰谷比增大，就能屏蔽无用的声音信息或背景噪音，更敏锐地锁定目标声源。这样就能在有背景噪音的环境里，提高目标语音的抗噪能力。元音/i/、/u/、/a/是一个语言里三个最为重要的元音，Stevens 称它们为“量子元音”（quantal vowels）［19］。量子元音声学特点是音量相对大的邻近共振峰（F1、F2 或者 F3）会形成双峰，两个共振峰就会在音强（amplitude）上产生相互叠加效应，增强彼此的音量。图 15 是元音/u/、/i/、/a/的 LPC 和窄带频谱示意图（根据 Diehl［3］Figure 1 增删加工）。

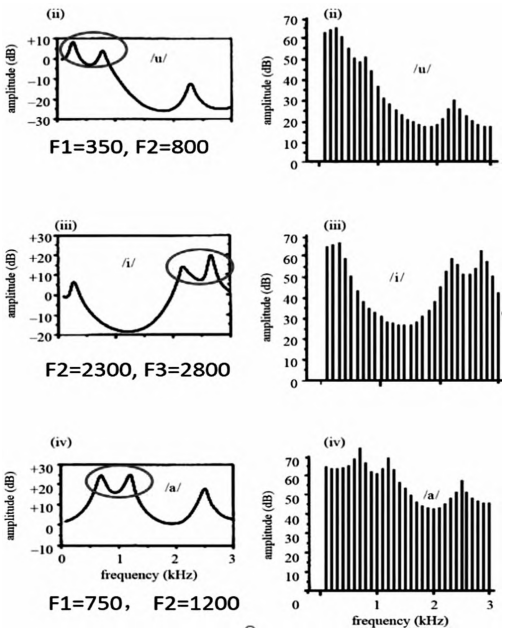


图 15　元音/u/、/i/、/a/的 LPC 和窄带频谱示意图

元音/u/的 F1 和 F2 彼此接近，形成双峰；元音/i/的 F2 和 F3 接近，形成双峰；元音/a/的 F1 和 F2 彼此接近，形成双峰。从图 15 可以看出，形成双峰固然提升了共振峰的峰值，但是也同时抬高了谷底。此时内耳基底膜中的外毛细胞掩蔽效应就起作用，掩蔽效应的作用使得上述元音双峰之间的峰谷比增大，两个相邻近峰的斜率变得陡峭。这样才能保证量子元音更具有区别性，才能保证语音信息在嘈杂的环境里做有效的传播。外毛细胞掩蔽效应增强语音清晰度的重要作用可以从调节人工耳蜗语音算法得到证明。有人在人工耳蜗里通过算法来增加输入信号中元音的峰谷比（20 句句子中的元音），其结果是大大增进了人工耳蜗携带者在有背景噪音时的听音清晰度［1］。图 16 是德语/o:/（成年女性发音）的三个共振峰峰谷比增加的例子。

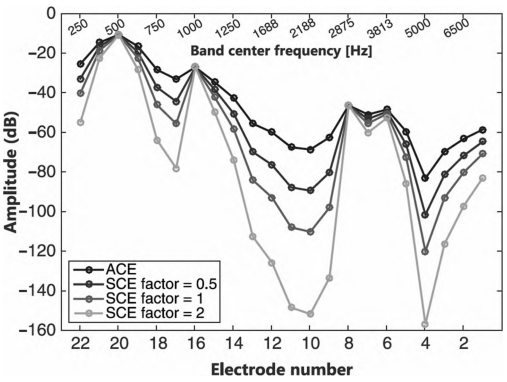


图 16　德语/o:/共振峰峰谷比增强策略效果

图 16 中/o:/三个共振峰的数值是 F1 = 500赫兹，F2 = 1000 赫兹，F3 = 2875 赫兹（女性声音）。ACE 曲线是未增加峰谷比的原声，SCE（spectral contrast enhancement）factor = 0.5、factor = 1、factor = 2 分别为经算法加工增加原声峰谷比 0.5 倍、1 倍、2 倍的曲线。实验表明峰谷比越大（如 SCE factor = 2），人工耳蜗携带者的听音清晰度越高。人工耳蜗是用电极直接刺激听觉神经，并没有正常人内耳基底膜外毛细胞对内毛细胞的掩蔽效应，人工耳蜗如果能仿照正常外毛细胞的掩蔽效

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

应功能，必将能大大提升嘈杂环境内听音的清晰度，从另一方面也证明内耳基底膜中外毛细胞的掩蔽效应在提升言语交际时的语音清晰度有重要作用。

6. 听神经放电特点与语流分段

内耳中内毛细胞兴奋放电刺激听神经，语音感知分析就进入听觉器官以上的神经系统。神经系统是语音感知的高级阶段，语音感知的加工也进入高级阶段。对语言感知来说，有两大挑战性的问题必须得到解释。

第一个挑战问题是要解释语音单位的切分机制。人们说话时连贯的，尤其在正常的语流中，一句话声波的振幅有高低不等，前后的音段、音节等都会有协同发音（coarticulation），发音时的协同发音会磨灭相邻音段、音节的界限，所以一句话是连续的语流信号。那么人类是根据怎么样的机制把连续语音信号切分为不同的离散的音段、音节等单位？现在发现连续语音信号被切分为不同的离散单位跟人类大脑神经振荡（neuronal oscillations）追踪有关。人类大脑神经系统有自主的神经振荡，左右两半自主神经振荡的频率是不同的，在大脑右半球颞上皮层（superior temporal cortices）连着主要听觉皮层（primary auditory cortex）自主振荡的主要频率是 3—6 赫兹，又称 Theta 振荡（Theta oscillator）。对侧左半球相同的地方自主振荡的重要频率是 28—40 赫兹［7］，又称 Gamma 振荡（Gamma oscillator）。

根据大脑神经振荡的语音感知理论可以解释大脑神经系统如何将输入的连续语音信号转为离散语音单位的［6］：第一步，当大脑的听觉皮层接收到传来的语音，自主振荡会重新设定振荡起点与语流“同步共舞”，这一步叫做“相位重新设定”（phase reset）。第二步，在重新设定振荡起点后，Theta 振荡在自己的频率范围内（3—6 赫兹）追踪语音包络（speech envelope）单位（音节），这一过程叫做刺激包络追踪（stimulus envelope tracking）。第三步，在新的 Theta 振荡内套合频率较高的 Gamma 振荡（25—35 赫兹），相当于音节内嵌入音段。本来自主振荡时这种套合信号是弱的，现在由于外加了语音信号，增强了振荡幅度，使得套合信号增强。第四步，神经元调节使兴奋增强后，输出离散的语音单位（音节、音段等）。第五步，神经兴奋振荡与声学语音结构对齐。捕捉到的这些离散单位及语音结构必须跟大脑原先储存的离散信号及结构匹配，才能对这些信号解码（decoding process）。最后根据这些分析再去捕捉和匹配下一段语音信号。Theta 和 Gamma 振荡不仅是切分解码语音单位的神经机制，同时也是语言理解的关键因素。有研究表明当语速变快或变慢，超出或不到 Theta 振荡频率的许可范围，语言理解就会产生困难［5］。

音节内部结构的切分和感知还跟听神经兴奋放电的特性有关。当大脑中枢根据外来语音重新设定振荡起点，听神经就会按照新设定的 theta 振荡的分段兴奋放电。开始时神经兴奋放电率（firing rate）非常大，这段叫做起始反应（onset response）。起始反应时间短，大概是 5—10 毫秒。以后放电率马上回落到较小的适应段（adaptation）。此适应段时间较长，大概有 50—60 毫秒。放电率会再次大幅降低，就进入下一个适应段，这段时间大概也有 50—60 毫秒。从起首放电到最后适应段结束，整个时长可以超过 100 毫秒。图 17 是听神经对噪音脉冲（burst of noise）滞后放电直方图（PST histogram of auditory nerve unit）［11］。

当神经兴奋放电率处在这两个适应阶段时，放电率会维持一段平稳时间，特别是在第二段适应阶段，神经兴奋反应阈值（response threshold）相当高，任何声音刺激都只维持很低的兴奋反应。声波包络刺激听神经，经听神经改造后的神经元放电模式可以想象应该如图 18 所示（笔者根据图 17 所作）：神经元兴奋放电的这种特点是人类语言音节结构的生理机制。神经元兴奋放电率高就是说明此处语音的分辨率敏感；反之则迟钝。人类语言里音节的最为普遍的结构是“辅音 + 元音 + 辅音”

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

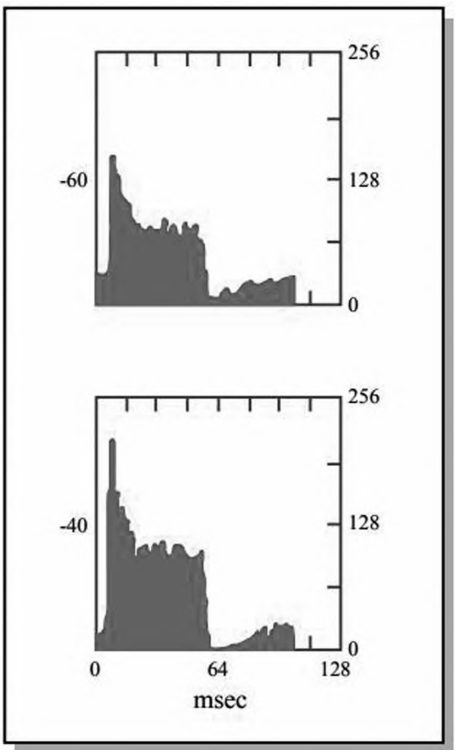


图 17　听神经对噪音脉冲的放电率

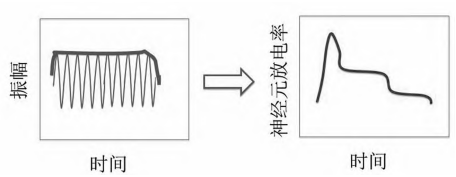


图 18　声波包络刺激听神经引起听神经放电示意图

或“辅音 + 元音”。音节的起首辅音不论种类还是数量都比结尾辅音的复杂和多样，因为音节起首位置神经元兴奋放电率高，分辨率高。在音节结尾神经元兴奋放电率低，对音段的分辨率也就低。

7. 肌动理论与语音感知

言语交际中充满变异，这些变异包括同一个说话人不同语速造成的变异，不同语境造成的变异，不同人由于性别、年龄、地区差异造成的变异等。所以语音感知的第二大挑战问题是听者如何去除说者的各种变异来匹配自己大脑里的相对一致的语音单位？其中的机制是什么？

20 世纪五六十年代 Liberman 等人提出的语音感知的“肌动理论”（motor theory）就是想回答这个问题。由于语音感知和声学特征缺乏不变量（lack of invariance）、语音声学信号的不可切割性（non-segmentability），以及语音传送信息的相对高效性（the relatively high efficiency of transmission of information），肌动理论认为听者“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。听者是通过自己发音所诱导出的听觉模式来分析和理解外来传入的语音的。比如听者听到“beat you”[biʧʰə]，他大脑里的神经元就会指挥和协调发音肌肉来发出这个语音；如果发出的这个语音模式和听到的语音模式相吻合，听者就会认同这个语音感知，从而语言得到正确无误的传递。换句话说，听者根据自己的发音器官的特征，以及牵动这些器官的肌肉运动（motor）“发音”来对外来的语音进行解码［12，13］。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。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。

20 世纪 90 年代后，随着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［26］。其中 Watkins 等人的一组研究成果从神经科学的角度揭示了言语感知和言语发声的紧密关系，从根本上奠定了肌动理论的神经生理机制［16，21，22］。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。下面介绍他们的研究成果，在此基础上进一步发挥，来解释语言变异与言语认

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

知的相关问题。

Watkins 等人的文章采用经颅磁刺激(Transcranial Magnetic Stimulation, TMS)技术并配合脑磁图记录(magnetoencephalography)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)刺激时大脑初级运动皮层(primary motor cortex)中管脸部的反应。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)非言语(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 和条件 3 情况下,左半脑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;而在右半脑,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。

经颅磁刺激只能在已知或目标固定大脑皮层区域测试诱发运动电位变化,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,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)言语条件;(2)唇动条件(与言语相关的嘴唇动作);(3)眼动条件。

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,但是左半脑激增区更大。特别在接受言语和模仿言语的唇动状态下,除了颞上回、颞叶前部沟回(the uncus in the anterior temporal lobe)外,左半脑的额下回(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,位于左半脑额下回鳃盖部(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。左半脑 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),再投射到左半脑 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 Transcranical Magnetic Stimulation, rTMS)短暂抑制管唇部发音动作和管手动作的左半脑运动皮层活动,来测试听者对特定辅音的范畴感知(categorical perception)的反应。塞音的范畴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

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

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辨认（identification）和区分（discrimination）能力。参加者共30位，英语是他们的母语。图19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。

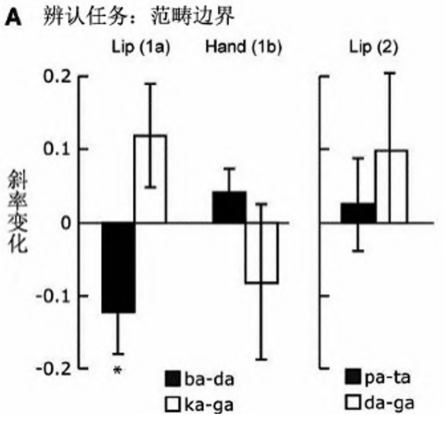


图19a 听者短暂抑制左脑运动皮层某些部位前（Pre-rTMS）测试数据的对比统计（* $p<0.05$， $p<0.01$，*** $pP<0.001$）**

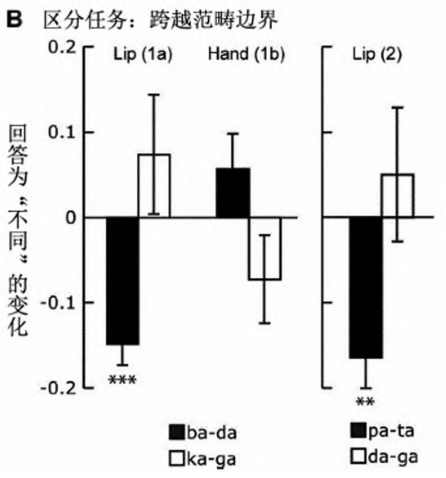


图19b 听者短暂抑制左脑运动皮层某些部位后（Post-rTMS）测试数据的对比统计（* $p<0.05$， $p<0.01$，*** $p<0.001$）**

图19a分别是四组塞音认定统计图，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范畴边界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连续体范畴感知的边界不起变化。

图19b分别是四组塞音区别统计图，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连续体中ba与da、pa与ta的区分精确度，而ka-ga和da-ga连续体中ka与ga、da与ga的区分精确度不受影响；

（2）每组范畴内的区分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分精度都不产生作用。

重复经颅磁刺激（rTMS）扰乱左半脑主要运动皮层唇部代表点对唇音的辨认感知和区分感知都产生影响，虽然这种影响并不足以取消不同音类的范畴感知特点，但是会改变听者对范畴感知的时间以及增感知错误率的增加。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。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。

Watkins等人上述的三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。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

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初级听觉皮层(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)传送到左脑初级运动皮层(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。

肌动理论假设听者是"雇用"他们自己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,把言语发音和感知的过程细化,研究言语感知与言语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,而是依靠听者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)过程中完成的[24]。

Hickok 和 Poeppel 提出的言语加工处理的双流理论(dual-stream model of speech processing)也认为语音感知要通过左脑背侧听觉区和运动区这条"流"的整合才能达到目的[8]。

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[24]。

第一,言语认知(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,或者在一个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。

第二,人类的言语发音以及语音感知是独特的,跟其他动物的发出的声音和感知不同。人类言语发音和语音感知的独特性可以从以下几个方面来阐述。(1)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的差异。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,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,即使动物对它们的"语言"也有感知机制,也因动物发音体、共鸣腔、脑结构等跟人类存在巨大的差异,"语言"感知的特点也会有很大的区别。(2)语音感知有补偿机制,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(co-articulation)现象,听者则用感知补偿机制来克服由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。(3)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来对外来的语音解码;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 和 Herzog[23]指出,语言变异是人类语言的本质,没有变异,语言就失去它的功能,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,

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

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的差异，声学参数会有很大的变异。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，言语交际一定会崩溃。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，从而达到言语交际的目的。

8. 结语

以上我们从听觉器官和听觉器官以上听神经系统的解剖生理系统来分析语音感知的特点。可以看出，语音进入听觉器官，听觉器官并不是被动的、客观的接收，而是对传入的声音进行了主观的改造和修饰。改造和修饰的特点及结果是跟听觉器官的解剖生理构造密切相关。由于人类听觉器官与其他哺乳类动物不同，所以相同的语音输入经由不同哺乳动物听觉器官改造和修饰所产生的特点和结果也应该是不同的。从而也说明人类语言的独特性。当语音进入听觉器官以上的听神经系统以及大脑高级认知皮层，还要做更为高级的语音匹配认知过程。其中的音段切分（segmentation）与大脑神经的自主振荡范围有关。大脑神经在自主振荡的范围内做刺激包络追踪，以此增强振荡幅度，从而将连续的音段切分为离散的单位。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来对外来的语音解码。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和矛盾。

语音感知还跟语言中的音系结构、语义、句法以及邻近词语或语音的疏密程度等因素有关，随着神经认知科学，脑影像技术的进步，语言神经认知研究也将取得越来越多的成果，语言认知的神秘面纱也将会被揭开。

9. 致谢

本文写作受国家社科基金重大项目“上海城市方言现状与历史研究及数据库建设”（批准号19ZDA303）的资助，特此鸣谢。

参考文献

[1] Boghdady, N. E., Langner, F., Gaudrain, E., Başkent, D., Nogueira, W. 2021. Effect of spectral contrast enhancement on speech-on-speech intelligibility and voice cue sensitivity in cochlear implant users. *Ear & Hearing* 42, 271–289.

[2] Denes, P., Pinson, E. 1993. *The Speech Chain*. New York: W. H. Freeman and Co.

[3] Diehl, R. L. 2008. Acoustic and auditory phonetics: the adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society B*. 363, 965–978.

[4] Fay, R., Popper, A. N. 1994. *Comparative Hearing: Mammals*. New York: Springer–Verlag.

[5] Ghitza, O., Greenberg, S. 2009. On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66, 113–126.

[6] Giraud, A., Poeppel, D. 2012. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience* 15 (4), 511–517.

[7] Giraud, A. et al. 2007. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56, 1127–1134. [PubMed: 18093532]

[8] Hickok, G., Poeppel, D. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8, 393–402.

[9] Johnson, K. 2012. *Acoustic and Auditory Phonetics. 3rd edition*. Wiley-Blackwell.

[10] Joris, P. X., Bergevin, C., Kalluri, R., Laughlin, M., Mc Michelet, P., van der Heijden, M., Shera, C. A. 2011. Frequency selectivity in

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Old-World monkeys corroborates sharp cochlear tuning in humans. *Proc. National Academy of Sciences of the United States of America* 108, 17516 – 17520.

[11] Kiang, N. Y-S., Watanabe, T., Thomas, E. C., Clark, Louise F. C. 1965. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*. Cambridge, MA: MIT Press.

[12] Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review* 74, 431 – 461.

[13] Liberman, A. M., Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition* 21, 1 – 36.

[14] Liljencrants, J., Lindblom, B. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48, 839 – 862.

[15] Moore, B. C. J. 2013. *An Introduction to the Psychology of Hearing. 6th edition*. Boston: Brill.

[16] Möttönen, R., Watkins, K. E. 2009. Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience* 29 (31), 9819 – 9825.

[17] Oxenham, A. J. 2012. Pitch perception. *Journal of Neuroscience* 32 (39), 13335 – 13338.

[18] Rosen, S., Howell, P. 2010. *Signals and Systems for Speech and Hearing. Ind edition*. Emerald Group Publishing Limited.

[19] Stevens, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17, 3 – 45.

[20] Von Békésy, G. 1960. *Experiments in Hearing*. New York: McGraw-Hill.

[21] Watkins, K. E., Strafella, A. P., Paus, T. 2003. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 48, 989 – 994.

[22] Watkins, K. E., Paus, T. 2004. Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience* 16 (6), 978 – 987.

[23] Weinreich, U., Labov, W., Herzog, M. 1968. Empirical foundations for a theory of language change. In: Lehmann, W., Malkel, Y. (eds.), *Directions for Historical Linguistics*. Austin: University of Texas Press, 95 – 188.

[24] 陈忠敏 《肌动理论和语言认知》, 《外国语》2015 年第 38 卷第 2 期, 第 15—24 页。

[25] 陈忠敏 《论言语发音与感知的互动机制》, 《外国语》2019 年第 42 卷第 6 期, 第 2—17 页。

[26] 张高燕、党建武 《言语信息处理的脑神经机制研究进展》, 《中国语音学报》2018 年第 2 辑, 第 13—21 页。

陈忠敏 复旦大学中国语言文学系、现代语言学研究院教授, 博士, 主要研究领域为神经语言学、历史语言学、实验语音学。E-mail: zhongminchen@ fudan. edu. cn

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net