

# 言语知觉中的感觉运动整合机制

梁柏燊<sup>1,2</sup>, 杜忆<sup>1,2</sup>

1. 中国科学院心理研究所; 中国科学院行为科学重点实验室, 北京 100101

2. 中国科学院大学心理学系, 北京 100049

**摘要** 大脑如何感知和理解言语是心理学和脑科学研究的重要问题之一。近年来, 言语产出相关的皮层言语运动系统(SMS)通常被认为以代偿调节方式参与有噪音干扰和/或信息缺失情况下的言语知觉中。本研究采用功能性核磁共振成像(fMRI)技术考查在噪音干扰下的音节辨识任务中的感觉运动整合机制。研究1发现, SMS的参与程度随语音辨识难度增加而增加、在噪音干扰下比听皮层具有更高的音位特异性编码和表征能力, 并且在难度适中时对听皮层编码代偿作用最大。研究2发现, 听力正常的老年人在噪音干扰下存在音节辨识困难, 且比年轻人更加依赖SMS的活动增加以及较少受到老化和噪音影响的言语运动表征以代偿听觉加工能力的下降。本研究不仅加深了对言语知觉中跨通道感觉运动信息整合机制及其毕生发展机制的认识, 还为帮助听力受损或正常的老年人设计和开展言语交流康复训练方案提供了新的视角和启发。

**关键词** 言语知觉; 噪音掩蔽; 感觉运动整合; 功能性磁共振成像; 老年化

言语交流是人类理解世界和社会交往的基石, 准确理解他人的言语表达并做出适应性反应是人类大脑重要的本质功能。自然情景下, 言语交流通常不是在理想条件下发生, 而是伴随着噪音干扰和/或信息缺失, 然而人脑却有着惊人的在嘈杂、不确定情景下识别言语的能力。深入解析人在自然交流条件下感知和理解言语的大脑运作机理, 特别是如何从多个感觉通道获取和整合信息, 不仅具有极其重要的基础研究价值, 还可提升类脑计算和人工智能的自然语言处理, 同时促进言语加工障碍或退化等脑疾病的早期诊断和干预。

言语知觉研究领域存在一个令人困扰的问题: 言语信号缺少声学语音学不变量。即相同的音位在不同语境中的声学特征存在巨大的差异。由于协同发音, 辅音的声学模式受到前后元音的影响, 例如, 音位/d/的声学模式在音节/di/和音节/du/中非常不同。此外, 说话人的嗓音、语速和情绪等因素也会影响音位的声学特征。另一方面, 面对高度变异的语音信号, 人对其的心理表征和知觉过程却是恒常的。为了解释人如何在变化的语音当中形成不变的音位知觉, Liberman等<sup>[1]</sup>提出了言语知觉的运动理论(motor theory of speech perception)。该理论认为, 在言语知觉过程中, 听者并非从声学上对语音进行解码分析, 而是还原说话者的发音姿势(motor gestures), 从而获得恒常的编码(例如/d/的发音必须包含舌和齿的特定运动)。但是, 运动理论在提出之后, 便受到了许多质疑。例如, 处理知觉不变性不一定需要范畴表征的参与。即便没有运动表征, 语境依赖的语音变异也可以直接通过听

觉加工来加以区分<sup>[2]</sup>。

让运动理论重新受到重视的是镜像神经元的发现。运用单细胞记录技术, Pellegrino等<sup>[3]</sup>发现, 在恒河猴的前运动皮层(premotor cortex, PMC)(F5区)存在一些特殊的神经元。这些神经元不仅会在恒河猴执行目标导向的运动(如往嘴里送食物)时激活, 在观察其他猴子执行相同的运动时也会被激活。这类神经元如镜子般反映他者的运动, 因此被称为“镜像神经元”。随后的研究在猴脑的其他区域发现了相似的神元, 这些神经元共同构成了镜像神经元系统, 包括顶下小叶(inferior parietal lobule, IPL)喙部、中央前回(precentral gyrus, preCG)底部以及额下回尾部(posterior inferior frontal gyrus, pIFG)<sup>[4]</sup>。从进化学的角度可以推知, 人类大脑中也可能存在着镜像神经元系统。细胞结构学的研究表明, 猴脑的F5区与人脑的布洛卡区(Broca's area)中的BA44区具有高度同源性<sup>[5]</sup>。研究显示, 布洛卡区损伤的病人在运动观察和理解中都有缺陷<sup>[6]</sup>, 意味着布洛卡区具有镜像神经元系统的功能。布洛卡区是言语产出的核心脑区, 此处可能存在联结了言语产出和言语理解的认知过程的镜像神经元<sup>[7]</sup>。具体而言, 在言语知觉的过程中, 听者的镜像神经元系统可能通过运动模拟还原说话者的声道运动, 从而理解语音。

近年来, 大量的人类脑成像研究证实, 言语产出和言语知觉具有高度重合的神经网络, 包括听皮层、运动和前运动皮层, 以及布洛卡区<sup>[8]</sup>。言语知觉中的运动贡献更直接的证据来源于经颅磁刺激(transcranial magnetic stimulation, TMS)

收稿日期: 2017-07-10; 修回日期: 2017-09-18

作者简介: 梁柏燊, 博士研究生, 研究方向为言语感知和理解的脑机制, 电子邮箱: liangbs@psych.ac.cn; 杜忆(通信作者), 研究员, 研究方向为人类听觉与言语认知的脑机制, 电子邮箱: duy@psych.ac.cn

引用格式: 梁柏燊, 杜忆. 言语知觉中的感觉运动整合机制[J]. 科技导报, 2017, 35(19): 21-28; doi: 10.3981/j.issn.1000-7857.2017.19.002

研究。数个TMS研究表明,言语运动系统参与到了没有运动任务的言语知觉任务中,而且还具有特异性的拓扑表征<sup>[9-10]</sup>。进一步的研究表明,使用TMS抑制言语运动脑区(如前运动皮层)会导致音位分辨或切分能力下降(如导致反应时延长),而且实验效应在语音信号存在信息缺失和噪音干扰时更明显<sup>[11-12]</sup>。因此,言语运动脑区参与了言语知觉过程,并以一种调节和代偿的方式发挥作用。

Hickok等<sup>[13]</sup>在整合了言语产出和言语知觉的研究后,提出言语知觉系统参与到言语产出过程、言语运动系统(SMS)参与到言语知觉过程的观点。他们针对言语产出提出了状态-反馈控制(state-feedback control)模型,并认为该模型同样可以解释言语知觉过程,因为在早期言语学习中,模仿说话者的发音动作是必要的。具体而言,在言语知觉中,听者的SMS能够对说话者的发音动作进行前向预测(forward prediction),并自上而下地调节听皮层的神经活动,使得对于特定刺激的知觉加工更加敏感;而听皮层能够自下而上地将所加工的语音信息反馈至SMS进行比对和修正<sup>[13]</sup>。因此,SMS以反馈回路的形式参与到言语知觉当中。言语的生成分析假设(analysis-by-synthesis)则更好地整合了言语知觉中运动代偿的情景依赖性<sup>[14-15]</sup>。该假设认为,只有在需要时,即嘈杂、信息不确定的情景下,言语运动系统才会生成言语运动预期以限定和辅助语音信号的解析。在神经环路层面,言语理解的皮层双通路模型(dual-stream model)认为,腹侧通路负责语音至语义的转换,而背侧通路负责语音的感觉运动整合(sensorimotor integration)<sup>[16]</sup>。对于背侧通路而言,听皮层的信息经由处于感觉运动界面(sensorimotor interface)的颞顶外侧裂(sylvian parietal temporal, Spt)传至包括额下回尾部的布洛卡区、前运动皮层和脑岛前部的言语运动系统(SMS);同时,SMS生成发音动作预期,该预期下行投射至感觉运动界面的Spt,与上行而来的语音听觉表征进行匹配以限定和辅助语音加工。

虽然越来越多的研究已表明,SMS以预测-反馈回路的形式调节和代偿言语知觉,但SMS的具体功能及言语感觉运动整合的神经基础仍不清楚。特别是,SMS与听皮层对语音信号的编码和表征,以及两个系统间的信息整合如何受到语音信号的模糊性(如操纵噪音强度)的调节,仍然缺少系统研究。另一方面,老年人(即使是听力正常的老年人)在嘈杂环境中也存在不同程度的言语理解困难。研究发现,老年人比年轻人在识别语音时更能从多模态的视听整合中获益<sup>[17]</sup>。并且,相对于年轻人,老年人在有噪音干扰或是语音受损模式下识别言语时激活了更强的前额叶脑区,这被认为与听觉注意、言语工作记忆、执行控制等功能对受损听觉加工的代偿有关<sup>[18-19]</sup>。这种老年化相关的前额叶活动增加是否与言语运动代偿功能增强有关,言语知觉中的感觉运动整合机制是否以及如何受到老年化的影响都需要进一步探索。本研究运用fMRI技术研究年轻被试(研究1<sup>[20]</sup>)和老年被试(研究2<sup>[21]</sup>)在不同噪音强度下的音节辨识任务中的行为和脑血流量变

化,通过对fMRI数据进行单体素的血氧水平依赖(blood oxygenation level dependent, BOLD)分析和多体素模式分析(multi-voxel pattern analysis, MVPA)分析,试图回答这些问题。

## 1 研究1:言语感觉运动整合对噪音掩蔽的动态代偿机制

### 1.1 研究设计

16名听力正常的成年人(21~34岁,女性8名)参与本实验。实验材料选取4个由一名女性自然说出的英语音节(/ba/,/ma/,/da/和/ta/),每个音节长度为500 ms,声压水平一致。选择元音/a/是由于其频谱分布能够较好地抵抗MRI的噪音,4个辅音的选取考虑到发声部位的平衡(双唇音/b/和/m/以及齿舌音/d/和/t/)。在有噪音掩蔽的条件下,500 ms的白噪音(4 kHz低通滤波,10 ms上升/下降包络)与音节同步出现和消失。声音通过MRI兼容的耳机播放。刺激声强保持在85 dB SPL,调节噪音声强使信噪比(signal-to-noise ratio, SNR)达到-12、-9、-6、-2和8 dB。此外,还有一个没有噪音掩蔽的条件。

在MRI扫描之前,音节在无噪音中播放(每个音节播放4次),被试通过按4键反应盒相对应的键进行辨音反应,准确率为94%以上则进入实验。实验在MRI扫描中进行,分为5个听音小节和1个默读小节。在每个听音小节中,80个噪音掩蔽的音节(每种SNR下每个音节呈现4次)和20个没有噪音掩蔽的音节(每个音节呈现5次)以随机方式呈现,刺激的间隔均值为4 s(2~6 s)。被试需要认真听音,并且迅速而准确地按键分辨。在默读小节中,被试依据视觉刺激默读/ba/,/ma/,/da/,/ta/4个音节。

### 1.2 fMRI数据采集和分析

成像数据在西门子Trio-3T磁共振成像仪(Siemens Trio 3T magnet)上通过12通道相控阵头线圈采集。全脑功能像由T2\*连续回波平面成像(Echo Planar Imaging, EPI)序列获得,30层,层厚=5 mm,扫描矩阵64×64,重复时间(TR)=2000 ms,回波时间(TE)=30 ms,偏转角(flip angle)=70°,扫描视野(FOV)=200 mm,体素大小=3.125 mm×3.125 mm×5 mm。3次功能成像扫描后,结构像由T1加权扰相梯度回波成像(Spoiled Gradient Recalled Echo, SPGR)序列获得(160层,层厚=1 mm,TR=2000 ms,TE=2.6 ms,FOV=256 mm)。

fMRI数据经过前处理(通过AFNI 2.56a软件进行,包括生理运动校正、切片时间校正和按照参考图像进行对准)后,分别采用一般线性模型(general linear model, GLM)分析基于单体素的脑区BOLD激活水平,和基于探照灯的多体素模式相似性分析方法MVPA考查不同脑区对音位的刺激特异性表征和分辨能力。MVPA是一种基于多体素的fMRI数据处理方法<sup>[22]</sup>,核心思想是选取与认知任务相关的一系列体素,再运用机器学习的方法探究体素组对特定刺激的相对稳定的激活模式,进而获得与某一认知维度相关的特征分类器。对

比于传统的单像素 fMRI 数据处理方法,该方法具有以下优点:1) 对信噪比低的体素包容性更强,因而能探测出更多传统方法认为“任务无关”的脑区;2) 能同时对同一认知维度上的多种信号进行对比。因此,采用4个音节(/ba/,/ma/,/da/,/ta/)作为实验材料,通过MVPA可以找出与音位范畴知觉相关的脑区。本研究通过对每个探照灯内的多体素激活数据建立音位诱发反应的相似性矩阵,计算范畴内与范畴间激活模式相关性(相似性)的平均差值获得音位特异性指数(phoneme-specificity index, PSI)以表示该处的音位范畴表征特异性程度。

### 1.3 结果

#### 1.3.1 行为结果、无噪音水平辨音及默读任务的脑区激活

行为结果表明(图1(a)和(b))<sup>[20]</sup>,每种音节的分辨任务都出现一致的准确率随SNR降低而降低、反应时随SNR降低而延长的结果,噪音控制和材料选取有效。

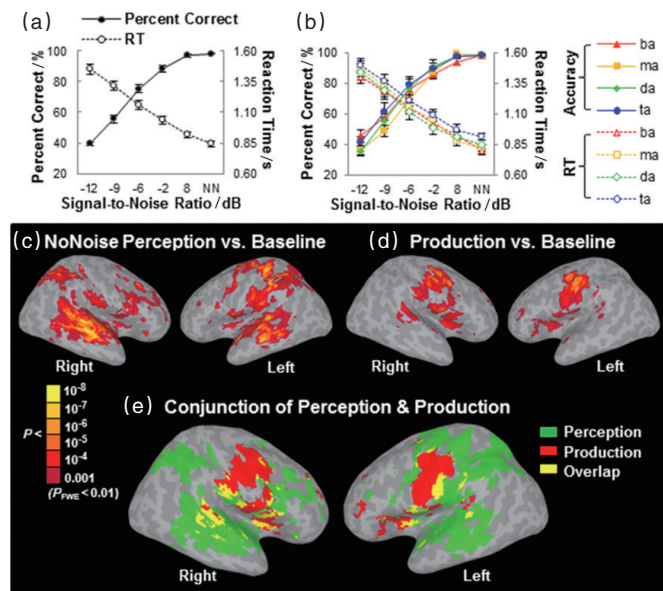


图1 行为结果及言语知觉和默读任务的激活脑区

Fig. 1 Behavioral results and brain activation patterns in speech perception and subvocal production tasks

GLM分析表明,无噪音条件下音节分辨任务(与静息基线相减)激活的脑区包括:双侧听皮层(包含颞横回, Heschl's gyrus(HG)和颞上回(superior temporal gyrus(STG))、脑岛前部(anterior insula)、腹侧前运动皮层(ventral PMC, PMv)、布洛卡区的脑盖部(pars opercularis, POP)、额叶和顶叶小叶(frontal and parietal lobules),以及左脑背外侧初级运动皮层(M1)和躯体感觉皮层(S1)(图1(c))。默读任务(与静息基线相减)激活的脑区包括:双侧背外侧初级运动皮层和前运动皮层、脑岛前部与相邻的腹侧前运动皮层和布洛卡区,以及颞上回后部(图1(d))。无噪音条件下音节分辨任务与默读任务共同激活的脑区包括:双侧脑岛前部、布洛卡区的脑盖

部、背外侧初级运动皮层、前辅助运动区、颞上回后部和豆状核,以及左脑腹侧躯体感觉皮层和顶下小叶(图1(e))。

#### 1.3.2 脑区激活水平与任务难度的关系

听觉腹侧和背侧通路脑区的激活水平受到SNR相关的任务难度的动态双向调节。位于双侧颞叶的听觉腹侧通路脑区的激活水平与SNR正相关(即与任务难度负相关),而位于额顶叶的听觉背侧通路(包括SMS和感觉运动界面)脑区的激活水平与SNR负相关(即与任务难度正相关)(图2)<sup>[20]</sup>。激活水平与难度负相关的脑区包括:双侧颞上回喙部(aSTG)、颞中回喙/尾部(a/p MTG)、扣带回皮层前部、顶上小叶;激活水平与难度正相关的脑区包括:双侧前部脑岛和布洛卡区(Ins/Broca)、前辅助运动区、丘脑,左脑腹侧前运动皮层(PMv)、颞上回尾部(pSTG)、顶下小叶(IPL)和右脑额中回。SMS激活程度与任务难度之间的正相关说明SMS适应性地代偿噪音干扰下的言语知觉困难。

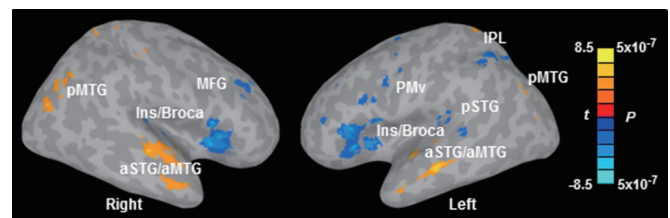


图2 言语知觉脑区激活水平与SNR的关系

Fig. 2 Relationship between SNR and activation level of regions involved in speech perception

#### 1.3.3 音位范畴表征特异性程度及其与任务难度的关系

使用基于探照灯的多体素模式相似性分析方法(MVPA),选取音位特异性指数(PSI)作为各脑区音位范畴表征特异性程度的指标。在无噪音条件下,具有显著音位特异性范畴表征(即PSI显著大于0)的脑区包括双侧听觉区的颞上回,以及左侧背外侧运动和前运动皮层、额下回和额中回等区域(图3(a))<sup>[20]</sup>。

在噪音干扰下,各脑区的PSI随着噪音强度的增大而下降(图3(b))。最明显的变化是双侧听觉区不再具有音位特异性范畴表征,而SMS(如前运动皮层和额下回)的激活模式依然可以分辨不同的音位。其中,PSI与SNR呈正相关的脑区包括:左脑额下回(IFG,包括BA45和BA46)、腹侧前运动皮层(PMv)、背侧中央后回和楔前叶,右脑顶下小叶和脑岛(图3(c))。具体看,SMS(左脑IFG和PMv)的PSI随着SNR的降低而逐步下降,在-6 dB之后接近于0(图3(d));但是听觉区(双侧STG)只在无噪音条件中PSI显著大于0(图3(e))。这说明,在噪音干扰下,SMS比听觉区具有更强的音位特异性编码和表征能力,而这恰好构成了噪音下言语运动代偿的神经基础。

进一步的分析显示,只在中等噪音强度下(8 dB>SNR>-6 dB),SMS的左侧脑岛和布洛卡区(Ins/Broca)处的PSI才显著高于左侧听觉区(pSTG)的PSI,即言语感觉运动整合的代

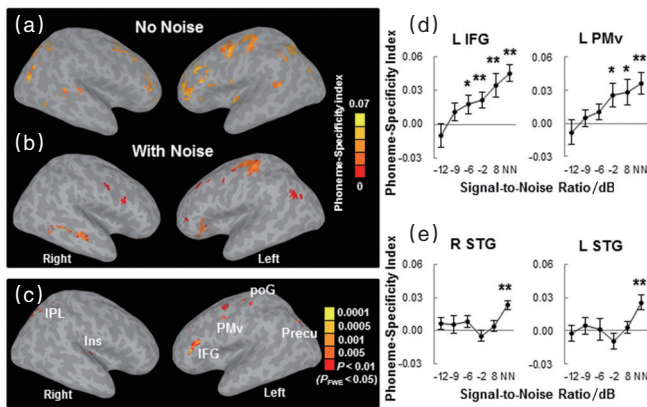


图3 言语相关脑区的音位范畴表征特异性及其与SNR的关系  
Fig. 3 Relationship between SNR and phoneme-specificity in speech-related regions

偿效应最大。整体而言,言语感觉运动代偿效应显示出与SNR的凸函数关系(图4)<sup>[20]</sup>。

#### 1.4 讨论

研究1显示,在辨音任务中,随着噪音的强度增大,SMS的参与程度增加;噪音干扰下,SMS比听觉区具有更高的音位范畴表征特异性;SMS相对听觉区的音位编码和表征优势与SNR呈现出凸函数的关系,在噪音强度中等的情况下言语运动代偿的效应最大。

本研究同时验证了言语知觉的皮层双通路模型<sup>[6]</sup>,发现听觉腹侧通路脑区的激活程度随着难度增大而下降(即随着语音可懂度的下降而下降),证实听觉腹侧通路与语音理解和语义加工密切相关;相反,听觉背侧通路脑区的激活程度随着难度增大而增大,表明背侧通路可能与感觉运动整合对言语知觉的代偿功能有关。然而,背侧通路中的SMS区域(如布洛卡区和腹侧前运动皮层)的激活增加可能意味着前向预测发音的内在模型的激活增强(预测-反馈回路的迭代),也可能只是反映了注意、工作记忆或者决策的需求增加。尽管无法完全排除第二种可能性,但MVPA分析支持了第一种可能性。

MVPA分析表明SMS和听皮层都有音位范畴特异性编码和表征能力,但SMS比听皮层的音位特异性表征受噪声的干扰更少。因此,在噪音中,SMS可能通过激活增加及较好保持的音位表征特异性去代偿听皮层受噪音削弱的音位分辨能力。通过噪音强度的大范围参数化设计,本研究还首次系统地考查了言语运动代偿能力受任务难度的动态调节特征。研究发现,当噪音太强(SNR<-6 dB)时SMS的发音预测准确性较低,因而无法代偿言语知觉;而噪音太弱(SNR>8 dB)时听皮层自身的音位表征特异性程度很高,不需要SMS的运动代偿参与;只有当噪音适中(8 dB>SNR>-6 dB)的时候,SMS产生的发音预期及较强的音位表征能力才会对听皮层损失的音位表征能力进行自上而下的代偿。

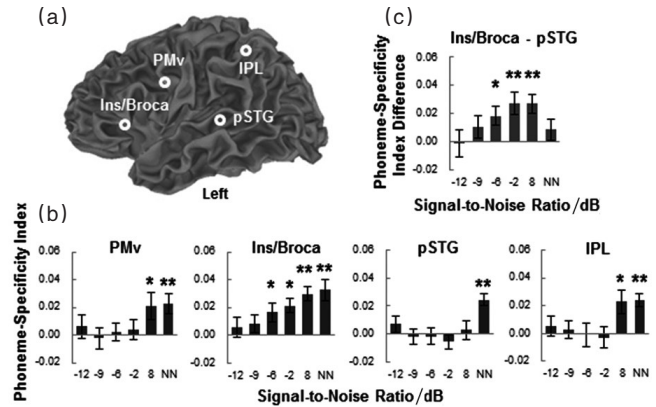


图4 言语感觉运动代偿功能与SNR的关系  
Fig. 4 Relationship between SNR and compensation function of sensorimotor integration

## 2 研究2:言语感觉运动整合代偿老化引发的言语感知下降

### 2.1 研究设计

16名听力正常的年轻人(21~34岁,女性8名)和16名听力正常(在250~4000 Hz频率区间纯音听阈<25 dB)的老年人(65~75岁,女性9名)参与本实验。研究设计与研究1相同,但研究2不包含默读任务。

### 2.2 fMRI数据采集和分析

fMRI数据的采集、预处理和GLM分析与研究1相同。不同的是,本研究使用基于解剖兴趣区(region-of-interest, ROI)的多体素模式分类分析<sup>[21]</sup>(anatomical ROI-based Multi-voxel Pattern Classification)而非研究1中使用的“探照灯”式MVPA考查语音加工相关脑区对音位的刺激特异性表征和分辨能力。这样做的好处是可以划分空间相邻却具有不同语音表征特异性的脑区的边界(比如腹侧前运动皮层与颞上回),也可以大大减少机器学习的计算量。

具体而言,首先对预处理后的数据进行最小平方和回归(Least Square Sum regression)估计每个试次激活的 $\beta$ 系数。接着使用FreeSurfer 5.3软件对被试T1图像进行自动解剖区域标注(aparc2009<sup>[23]</sup>)获得148个皮层和皮层下脑区结构,并利用Neurosynth网站的元分析工具设置关键词“speech”找到与语音加工相关的脑区(<http://neurosynth.org/features/speech/><sup>[24]</sup>),将两种方法获得的脑区重叠后得到38个与语音加工相关的解剖加功能ROI(19个左侧和19个右侧相对应脑区)。然后,针对这38个ROI内的所有体素在各SNR条件下的 $\beta$ 系数使用R程序中的“sda”语言包(shrinkage discriminant analysis<sup>[25]</sup>)训练分类器分辨4种音位,再用新的数据测试分类成绩,交叉验证的方式为leave-one-out。音位分类成绩用曲线下面积(area-under-the-curve, AUC)表示,因为该指数较好地平衡了分类器输出概率和离散类别关系。

## 2.3 结果与分析

### 2.3.1 老年人额叶运动区激活增加代偿行为成绩

在各噪音强度下,老年人的音节分辨正确率都比年轻人显著低,而老年人的行为正确率与其纯音听阈值成反比。这证实了,即使是听力正常的老年人在噪音干扰下也存在言语识别困难,而该困难可被外周听力下降部分解释。

对无噪音条件下的fMRI进行GLM分析显示:对于年轻人,无噪音辨音任务(减去静息基线)激活了双侧颞叶、顶叶和额叶的广泛区域(图5(a))<sup>[21]</sup>;对于老年人,相同的任务激活了相似的脑区,但是激活的程度更大,特别是双侧颞叶和感觉运动区以及左脑额叶(图5(b)和(c))。在有噪音的条件中,对年轻人和老年人的BOLD反应按照行为准确率进行配对(年轻人SNR=-6, -2dB对比老年人SNR=-2, 8 dB)比较发现,为了达到相同的正确率,老年人比年轻人激活程度更高的脑区有:左脑脑盖部和邻近的腹侧前运动皮层(PMv)、双侧颞上回(STG)和颞中回(MTG)喙部和中部、背侧中央前回(preCG,包含运动皮层和前运动皮层)和中央后回(postCG)、顶上小叶、额内侧回和丘脑(图5(d))。将这些脑区BOLD激活水平与行为准确率进行相关分析发现,老年人言语运动相关的左脑脑盖部和中央前后回(pre/postCG)的激活水平与行为准确率呈显著正相关,而老年人的双侧听觉区域(STG)以及年轻人的运动区域和听觉区域的激活水平与行为准确率相关不显著(图5(e))。这提示,老年人在加工噪音干扰下的语音时出现额叶言语运动区激活程度增加,而言语运动区的

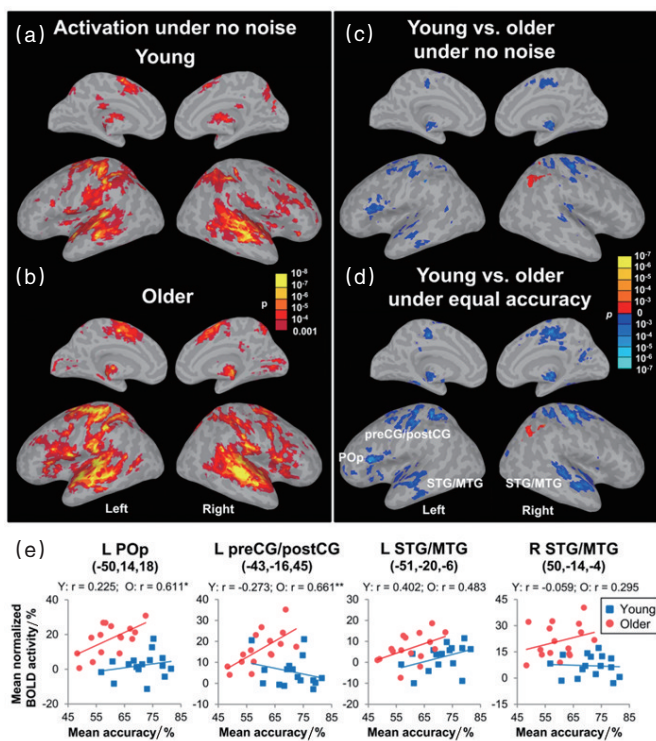


图5 年轻人与老年人在音节分辨任务中的脑区激活模式及差异  
Fig. 5 Different brain activation patterns in syllable identification tasks in young and older adults

激活程度可预测老年人的语音分辨成绩,即老年人更多地调用了SMS以代偿言语知觉过程。

### 2.3.2 老年化引起语音表征同质化

在无噪音条件下,年轻被试具有显著的音位表征分类能力(AUC显著大于随机水平)的脑区包括:双侧颞横回(HG)、颞上回(STG)、缘上回、中央后回(postCG)和中央前回(preCG),左脑颞平面(planum temporal, PT)和布洛卡区;老年被试具有显著的音位表征分类能力的脑区包括:左脑中央后回(postCG)、中央前回(preCG)和布洛卡区的脑盖部(POp)(图6(a))。可见,在无噪音干扰下,老年人的听皮层已经无法有效地编码和分辨不同的音位了而言语运动脑区(如POp)的音位分辨能力依旧完好并保持到了一定的噪音强度(SNR=8 dB)(图6(b))<sup>[21]</sup>。因此,老年化引起全脑广泛的语音表征同质化(即可分辨性降低),但言语运动区域比听皮层的音位范畴表征特异性更少受到老年化和噪音的负性影响。

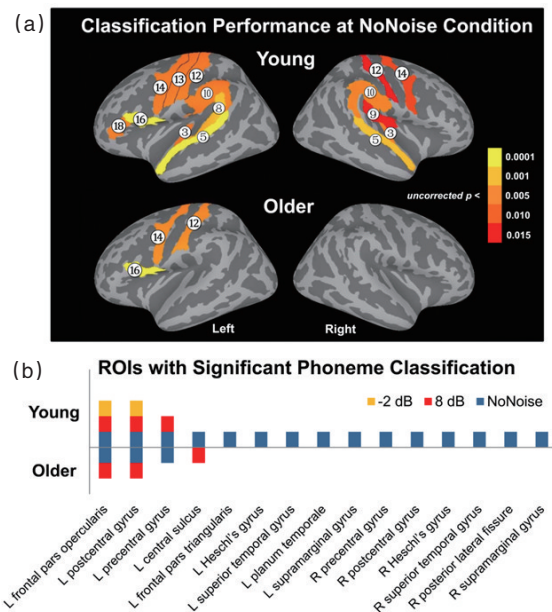


图6 各噪音条件下年轻人与老年人各脑区的音位表征分类成绩

Fig. 6 Phoneme classification performance in ROIs at various noise levels in young and older adults

### 2.3.3 老年人言语运动代偿向低难度偏移

无论是年轻人还是老年人,在噪音干扰下,言语运动区的音位分辨能力都高于听觉区和感觉运动界面(图6(b)),这表明在噪音中分辨语音可能受到自上而下的感觉运动代偿。为量化言语运动区对听觉区的代偿如何受信噪比和年龄的影响,将每个年龄段每个信噪比下的左脑布洛卡区脑盖部(POp)的AUC得分分别与左脑颞横回(HG)、颞上回(STG)和颞平面(PT)的AUC得分进行相减,作为言语感觉运动代偿的指标(图7)<sup>[21]</sup>。结果显示,年轻人的言语运动代偿与SNR呈凸函数关系,但老年人的运动代偿峰值向低噪音处偏移。因此,老年人在更低认知难度时就调用了言语运动代偿功能。

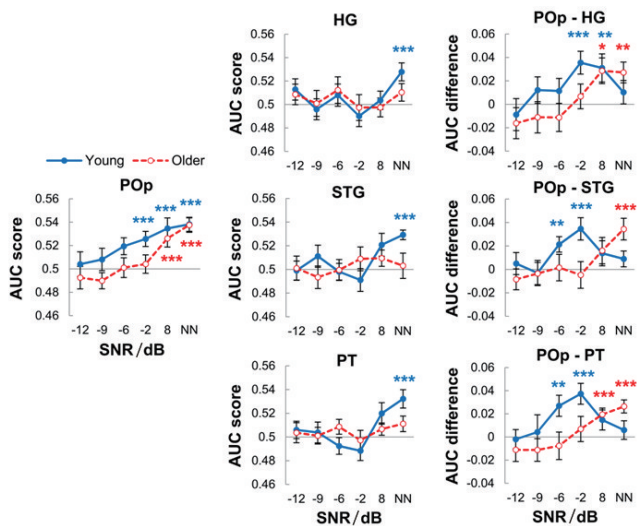


图7 年轻人和老年人的言语感觉运动代偿与SNR的关系

Fig. 7 Relationship between SNR and sensorimotor compensation function in young and older adults

### 2.3.4 讨论

研究2发现,在噪音干扰的音节分辨任务中,相比于年轻人,老年人SMS和听觉区激活增加,其中SMS的激活程度与准确率正相关;老年化引发广泛的语音表征同质化,但运动区比听觉区较少受到老化和噪音影响;老年人在更低认知难度时就调用了运动代偿,而且相比年轻人,老年人运动代偿的峰值向低噪音水平移动。

老年化的衰退—代偿理论(decline-compensation hypothesis)认为,作为一个动态的生物学适应过程,健康的老年化也伴随一些代偿性的大脑认知资源重组和认知策略转变,以应对无法避免的感知觉功能下降<sup>[26]</sup>。与衰退—代偿理论一致,本研究中,老年人前额叶言语运动区活动增强与自上而下的言语运动调节增加有关,可用以抵消听觉加工能力的衰退进而代偿语音分辨成绩。

MVPA分析显示,老化削弱了各脑区的音位范畴表征特异性程度,这可能导致了老年人的音节分辨成绩下降。但是,老年化引发的音位表征同质化程度在全脑并非均匀分布,SMS比听觉区受到的影响更少,因此老年人可能更依赖于SMS的发音预期以对受损的语音听觉编码进行运动代偿。本研究还发现当噪音非常小甚至没有时,老年人的言语运动代偿最为明显。这说明老年人在更低任务难度中就调用了言语感觉运动代偿功能,这与神经回路的补偿相关使用假设(compensation-related utilization of neural circuits hypothesis, CRUNCH)<sup>[27]</sup>一致。

总之,老年人在听觉系统无法有效提取和编码目标言语,也难以从其他知觉线索获益的情况下,可能比年轻人更加依赖较少受到老年化和噪音干扰的言语运动预期和感觉运动整合作为一种优势线索和重要的代偿机制。这种由数

据驱动的信息加工模式向自上而下的预测加工(predictive coding)模式的转变,由单通道向多通道信息整合的转变不仅反映了大脑言语加工系统应对听觉掩蔽和听觉老年化双重挑战的适应性,也提示了一条对老年人进行言语康复训练的重要途径。

## 3 结论与展望

大脑的基本功能之一就是应对各种复杂的知觉场景,而多通道信息整合是大脑处理复杂问题的重要一般性策略。本研究对嘈杂环境下言语知觉过程中的感觉运动整合功能的探索可能成为对诸如“鸡尾酒会问题”等听觉问题以及更广泛的感知觉问题和跨通道信息整合问题的研究突破口。同时,老年化极大地减弱了听者在嘈杂环境中识别言语的能力,但研究者对其影响机理的认识还很大程度上局限于听觉加工内部和高级认知调节层面。本研究将老年化对言语知觉过程的作用靶点放到跨通道的言语感觉运动整合功能上,开启了新的研究领域。

本研究的发现还可能具有一定应用价值。在言语康复训练方面,如果老年人在嘈杂环境中识别言语时更依赖SMS的发音预期和感觉运动整合机制,那么针对广大的老年人群可以设计和开展或许更为有效的言语康复训练方案(如针对SMS的靶向增强训练),并将之扩展到更多听力受损的儿童和成人,增进他们的社会交往和生活质量。此外,在工业转化应用方面,声学场景的复杂性不仅增加了人-机交互中计算机语音识别的难度,也降低了助听器和人工耳蜗等听力辅助设备在患者中的使用效果。本研究对言语知觉中跨通道感觉运动信息整合机制的研究可在一定程度上为计算机语音识别算法及听力辅助设备的研制提供神经生物学指导。

但是,本研究也存在一些局限性。例如,为测量感觉运动代偿,采用MVPA方法得出SMS和听觉区在不同难度中的音位范畴表征特异性程度并进行相减,这种从SMS与听觉区范畴表征程度的差异推断出言语感觉运动整合代偿的方式较为间接。未来需要采用诸如有效性连接(effective connectivity)等分析方法探究SMS和听皮层两个区域间的信息传递和整合方式,从而给出更直接的证据。其次,由于研究采用的fMRI时间分辨精度较低,SMS与听觉区的信息整合和代偿效应的时程动态变化特性仍未被探明。未来可使用兼具时间和空间分辨精度的脑磁成像技术(magnetoencephalography, MEG)研究该问题,还可以同时考查言语感觉运动整合的神经振荡机制。再次,研究多通道知觉运动训练方式(如音乐训练)对言语感觉运动整合的影响,以及言语障碍背后的感觉运动代偿机制的变化也是未来的研究方向。

综上所述,本研究采用功能性核磁共振成像(functional magnetic resonance imaging, fMRI)技术考察了在噪音干扰下的音节辨识任务中的感觉运动整合机制:

1) 皮层言语运动系统(speech motor system, SMS)的参

与程度随语音辨识难度增加而增加。

2) SMS在噪音干扰下比听皮层具有更高的音位特异性编码和表征能力。

3) SMS在难度适中时对听皮层编码代偿作用最大。

4) 听力正常的老年人在噪音干扰下音节辨识有困难。

5) 听力正常的老年人比年轻人更加依赖SMS的活动增加以及较少受到老化和噪音影响的言语运动表征以代偿听觉加工能力的下降。

### 参考文献 (References)

- [1] Liberman A M, Cooper F S, Shankweiler D P, et al. Perception of the speech code[J]. *Psychological Review*, 1967, 74(6): 431.
- [2] Holt L L. Temporally nonadjacent nonlinguistic sounds affect speech categorization[J]. *Psychological Science*, 2005, 16(4): 305-312.
- [3] Pellegrino G, Fadiga L, Fogassi L, et al. Understanding motor events: a neurophysiological study[J]. *Experimental Brain Research*, 1992, 91(1): 176-180.
- [4] Gallese V, Fadiga L, Fogassi L, et al. Action recognition in the premotor cortex[J]. *Brain*, 1996, 119(2): 593-609.
- [5] Fogassi L, Ferrari P F. Mirror neurons and the evolution of embodied language[J]. *Current Directions in Psychological Science*, 2007, 16(3): 136-141.
- [6] Fazio P, Cantagallo A, Craighero L, et al. Encoding of human action in Broca's area[J]. *Brain*, 2009, 132(7): 1980-1988.
- [7] Fadiga L, Craighero L, D'Ausilio A. Broca's area in language, action, and music[J]. *Annals of the New York Academy of Sciences*, 2009, 1169(1): 448-458.
- [8] Skipper J I, Devlin J T, Lametti D R. The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception[J]. *Brain and Language*, 2017, 164: 77-105.
- [9] D'Ausilio A, Pulvermüller F, Salmas P, et al. The motor somatotopy of speech perception[J]. *Current Biology*, 2009, 19(5): 381-385.
- [10] Fadiga L, Craighero L, Buccino G, et al. Speech listening specifically modulates the excitability of tongue muscles: A TMS study[J]. *European Journal of Neuroscience*, 2002, 15(2): 399-402.
- [11] Meister I G, Wilson S M, Deblieck C, et al. The essential role of premotor cortex in speech perception[J]. *Current Biology*, 2007, 17(19): 1692-1696.
- [12] Sato M, Tremblay P, Gracco V L. A mediating role of the premotor cortex in phoneme segmentation[J]. *Brain and Language*, 2009, 111(1): 1-7.
- [13] Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization[J]. *Neuron*, 2011, 69(3): 407-422.
- [14] Bever T G, Poeppel D. Analysis by synthesis: A (re-) emerging program of research for language and vision[J]. *Biolinguistics*, 2010, 4(2-3): 174-200.
- [15] Poeppel D, Monahan P J. Feedforward and feedback in speech perception: Revisiting analysis by synthesis[J]. *Language and Cognitive Processes*, 2011, 26(7): 935-951.
- [16] Hickok G, Poeppel D. The cortical organization of speech processing [J]. *Nature Reviews Neuroscience*, 2007, 8(5): 393-402.
- [17] Maguinness C, Setti A, Burke K, et al. The effect of combined sensory and semantic components on audio-visual speech perception in older adults[J]. *Frontiers in Aging Neuroscience*, 2011, 3: 19.
- [18] Erb J, Obleser J. Upregulation of cognitive control networks in older adults' speech comprehension[J]. *Frontiers in Systems Neuroscience*, 2013, 7: 116.
- [19] Wong P C M, Jin J X, Gunasekera G M, et al. Aging and cortical mechanisms of speech perception in noise[J]. *Neuropsychologia*, 2009, 47(3): 693-703.
- [20] Du Y, Buchsbaum B R, Grady C L, et al. Noise differentially impacts phoneme representations in the auditory and speech motor systems[J]. *PNAS*, 2014, 111(19): 7126-7131.
- [21] Du Y, Buchsbaum B R, Grady C L, et al. Increased activity in frontal motor cortex compensates impaired speech perception in older adults [J]. *Nature Communications*, 2016: 7: 12241.
- [22] Norman K A, Polyn S M, Detre G J, et al. Beyond mind-reading: multi-voxel pattern analysis of fMRI data[J]. *Trends in Cognitive Sciences*, 2006, 10(9): 424-430.
- [23] Destrieux C, Fischl B, Dale A, et al. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature[J]. *Neuroimage*, 2010, 53(1): 1-15.
- [24] Yarkoni T, Poldrack R A, Nichols T E, et al. Large-scale automated synthesis of human functional neuroimaging data[J]. *Nature Methods*, 2011, 8(8): 665-670.
- [25] Ahdesmäki M, Strimmer K. Feature selection in omics prediction problems using cat scores and false nondiscovery rate control[J]. *Annals of Applied Statistics*, 2010, 4(1): 503-519.
- [26] Cabeza R, Anderson N D, Locantore J K, et al. Aging gracefully: Compensatory brain activity in high-performing older adults[J]. *Neuroimage*, 2002, 17(3): 1394-1402.
- [27] Reuter-Lorenz P A, Cappell K A. Neurocognitive aging and the compensation hypothesis[J]. *Current Directions in Psychological Science*, 2008, 17(3): 177-182.

## Mechanisms of sensorimotor integration in speech perception

LIANG Baishen<sup>1,2</sup>, DU Yi<sup>1,2</sup>

1. Key Laboratory of Behavioral Science; Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

2. Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract** The way our brains perceive and comprehend speech remains an essential issue in psychology and brain science. In recent years, consensus has gradually been made on that the speech motor system (SMS) contributes to speech perception in a modulatory and compensatory manner, especially under noise masking or speech degradation circumstances. Using functional magnetic resonance imaging, the current work investigates neural mechanisms of sensorimotor integration and compensation during syllable identification task in noisy conditions. Study 1 reveals an increment of the SMS involvement as the task difficulty increases. In addition, the SMS shows more robust specificity of phoneme representations than auditory regions in noisy conditions, as such a superiority climaxes at moderate noise levels. Study 2 finds that older adults with normal hearing have difficulty in syllable-in-noise identification. Compared with younger counterparts, the elders rely more on increased recruitment of the SMS and preserved speech motoric representations, which are less affected by noise and aging effects than auditory ones, to compensate for deficient auditory encoding. Our findings deepen the knowledge of neural mechanisms underlying cross-modal sensorimotor integration in speech perception and its life-span developmental changes. Moreover, new insights could be given on designing and implementing rehabilitative and training regimens for speech comprehension (e.g., SMS-targeted brain enhancement) in older adults with or without hearing loss.

**Keywords** speech perception; noise masking; sensorimotor integration; functional magnetic resonance imaging; aging

(责任编辑 祝叶华)