



语音处理的神经电生理学基础研究进展*

于海情^{1)**} 许敏鹏^{1,2)**} 万柏坤¹⁾ 明 东^{1,2)***}⁽¹⁾ 天津大学精密仪器与光电子工程学院, 天津 300072; ⁽²⁾ 天津大学医学工程与转化医学研究院, 天津 300072)

摘要 解码大脑在语音处理过程中涉及的信息加工层级结构、皮质响应机制及功能连接模式, 是神经语言学领域的研究重点. 以语音信息加工时序为依据, 可将该认知过程划分为: 初级声学信号时频编码 (spectrotemporal analysis of primary acoustic signals)、音素处理 (phonemic processing) 以及词汇-语义加工 (lexical-semantic processing) 3 个处理阶段. 目前, 研究者已对各阶段的神经机制进行了广泛且深入的研究, 但不同模型理论/假说难以整合互补, 有必要进行梳理与总结. 本文将大脑处理语音信息的 3 个阶段为主线, 以电生理学方法为侧重范式, 对各阶段下的皮质映射、神经振荡模式以及事件相关响应机制等神经基础研究现状进行总结评述, 以期为进一步了解语音信号如何在人脑中进行处理和表达等相关研究提供一定的参考.

关键词 语音处理, 神经电生理学, 初级声学信号时频编码, 音素处理, 词汇-语义加工

中图分类号 R318, Q189

DOI: 10.16476/j.pibb.2020.0050

语言 (language) 是人类独具的递归分析与组合逻辑等高级认知能力的主要表现途径之一. 其复杂精细的编解码方式、丰富的特征与信息使之具有较强的指向性、描述性以及逻辑性. 19 世纪神经语言学 (neurolinguistics) 兴起, 使语言感知、理解、表达和习得的神经机制受到广泛关注. 相关研究成果不仅在失语症、脑卒中等中枢神经系统疾病的临床诊治方面具有突出的实用意义^[1-2], 也为目前类脑学习等人工智能领域的建模与仿真开辟了新的思路^[3].

目前, 神经语言学研究常采用电生理学或功能神经影像学方法, 对不同语言任务下的大脑皮质功能活动进行记录与分析, 从而探究其神经学机制. 其中, 电生理学方法一般包括: 头皮脑电 (electroencephalography, EEG)^[4]、皮层脑电 (electrocorticography, ECoG)^[5]、立体定向脑电图 (stereoelectroencephalography, SEEG)^[6]、皮层刺激映射 (cortical stimulation mapping, CSM)^[7] 等. 功能神经影像学方法一般包括: 功能磁共振成像 (functional magnetic resonance imaging, fMRI)^[8]、脑磁图 (magnetoencephalography, MEG)^[9] 以及功能近红外光谱 (functional near-

infrared spectroscopy, fNIRS)^[10] 等. 值得一提的是, 部分研究在实验范式中引入神经调控手段, 如经颅磁刺激 (transcranial magnetic stimulation, TMS)^[11], 通过对局部皮质施加刺激以暂时阻断神经响应, 从而在确定局部脑区的部分功能.

语音的感知与加工是语言表达的基础与前提, 语言相关皮质发育前期对外源性语音刺激输入具有较强依赖性^[12]. 研究表明, 婴儿时期的声音感知水平对其语言和认知系统的发育具有决定性影响^[13]. 因此, 对于先天性耳聋等感觉剥夺疾病, 若未在患者语言发展敏感期获得有效干预或治疗, 则会导致听觉系统功能发育不足, 以及语言表达能力的严重缺损. 其病损累及程度与有效干预年龄显著相关: 干预越早, 受损越小^[14]. 因此, 语音感知与加工的潜在机制具有至关重要的研究意义.

大脑感知语音过程的神经学研究最早可追溯到 19 世纪 70 年代, 德国神经学家 Carl Wernicke 对一

* 国家自然科学基金(81601565, 81630051)资助项目.

** 并列第一作者.

*** 通讯联系人.

Tel: 13820222616, E-mail: richardming@tju.edu.cn

收稿日期: 2020-05-10, 接受日期: 2020-07-17

位具有特定病征失语症患者的大脑进行解剖后发现: 该患者所表现的语言理解能力缺失, 可能与听觉皮层后部的韦尼克区 (Wernicke's area) 受损有关^[15]. 此外, 在Lichtheim提出“概念中心”的猜想后 (无解剖学定位), Wernicke由此建立了基于声音意象中心 (韦尼克区)、语言运动中心 (布洛卡区) 以及概念中心的失语症类型划分系统. Wernicke-Lichtheim的语言处理模型首次考虑了各中心之间相联结的信息加工路径^[16]. 功能性神经

影像学发展推进了相关研究的不断深入, 并逐步建立了目前已广泛应用于智慧学习的TRACE模型^[17], cohort模型^[18]以及邻域激活模型 (neighbourhood activation models)^[19]等较完善的语音加工模型. 以非人灵长类动物为实验对象的现代神经生理学研究建立了更加精确的语音感知神经解剖学双流模型 (图1所示为语音/语言处理双流模型)^[20].

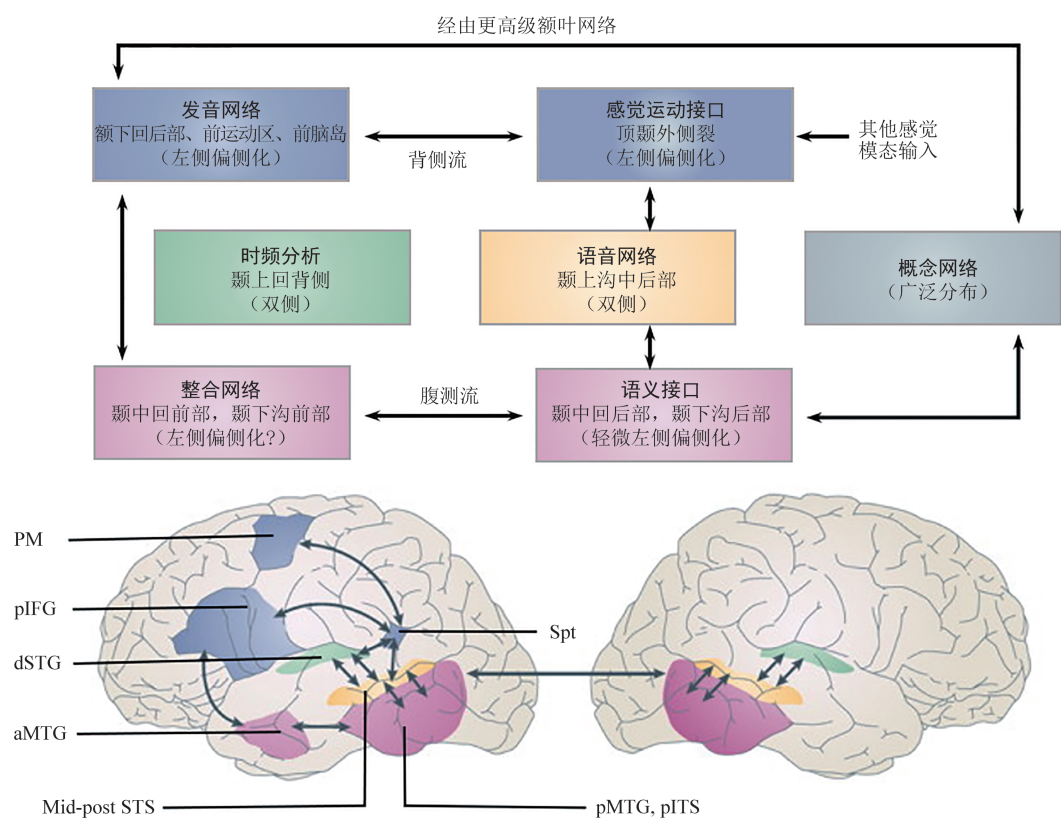


Fig. 1 The dual stream model of speech/language processing^[20]
图1 语音/语言处理双流模型^[20]

该双流模型假定: 语音信息加工由腹侧与背侧通路协同完成. 其中背侧流自颞上皮质 (superior temporal cortex, STC) 听觉区域后部起, 经由顶叶躯体感觉皮质 (somatosensory cortex, SSC) 下部, 延伸至左侧额下回 (inferior frontal gyrus, IFG), 背侧信息通路主要负责语音定位^[21]、基于语境或音韵特征的反馈与调制、感觉运动信息整合^[22]以及对本体语言状态的监测^[23]等; 腹侧流起始于STC背侧初级听觉皮层 (primary auditory cortex, PAC), 经颞中回 (middle temporal gyrus,

MTG) 后部向颞叶前部延伸, 并连接到IFG腹侧部分, 腹侧通路主要负责音素处理^[24]、词形预期与识别^[25]、语义加工^[26]等. 语音感知过程中运动与听觉皮质的功能网络模型也是当前的研究重点之一. Liebenthal等指出3种经典听觉-运动皮质加工模型——并行^[27]、层级^[28]、交互模型^[29]——存在模型理论与神经生理学依据不一致性, 并以交互模型为理论基础, 结合语音感知时程早期 (前100 ms期间) 听觉与躯体运动区域的同步激活现象, 提出新的听觉-运动交

互模型, 即区域间功能连接与调制起始于语音处理的早期阶段, 持续至感知后期^[30-31]。

综上, 语音信息处理的神经机制已得到广泛研究, 但不同模型理论/假说的具体细节存在一定差异, 因此应进行梳理与总结。本文依据语音感知信息加工时序, 对于初级声学信号时频编码、音素处理、以及词汇-语义加工3个处理阶段的皮质映射模式、振荡特点、事件相关电位等神经语言学研究现状进行归纳评述。

1 初级声学信号时频编码神经学基础研究进展

语音信号具有高度可变性。除内容差异外, 其声学特征在形成过程中会受语境、语速以及说话者声道等多种因素影响, 致使同一音素或单词的频谱-时间 (spectral-temporal, ST) 特征发生变化。因此, 语音感知早期阶段应具有ST响应可变性与不变性^[32]: 前者表现为神经活动的ST响应随语音内容变化; 后者表现为神经活动对于ST特征的表达不随音色、重音习惯等表达差异而变化, 对相同的语音内容调用相同的处理机制, 并通过反馈进行少量特异性调整以保证早期计算的稳定进行, 即声学变化的归一化处理。这种早期感知过程提取了声学-语音特征 (acoustic-phonetic properties) 并将其预映射为音素等词汇前表达单元, 即初级声学信号的时频编码^[33], 这些表达单元具有沿多个维度对声学-语音信息进行整合和分类的功能。本节将分别介绍动物、人类编码初级声学信号时频特征的神经电生理学基础研究进展。

1.1 基于动物实验的初级声学信号时频编码研究

人类与其他哺乳动物具有相通的听觉皮质功能映射以及声信号处理机制, 因此基于动物神经生理学研究得到的清晰、准确的皮质功能映射模型, 可为研究人类对于初级声学信号的ST编码机制提供参考。沙鼠音调鉴别训练实验中, 发现其对于听觉刺激序列中的新异音调具有不同的听觉皮层激活模式^[34]。非人灵长类动物实验研究表明, 随听觉材料复杂度增加, 实验动物分别调用了听觉皮层的核心区 (core area)、带状区 (belt area) 以及旁带状区 (parabelt area) 进行信息加工^[35-36], 即不同子区域对于不同ST属性的声学信号进行选择响应^[37]。其中带状区以及旁带状区被认为主要响应同类的呼号式“沟通”以及语言行为相关的其他声学信号^[37]。非人灵长类动物的同类“语音”共振

峰感知特点与人类极为相似^[38]。O'Connell等设计了基于节律性声学刺激的选择注意任务范式, 使用单细胞记录猕猴执行任务中初级听觉皮质 (A1) 区域的神经元集群响应, 阐明了A1中多尺度神经振荡-外界节律同步化 (neural entrainment) 现象的拓扑结构。该研究表明, 神经元集群的频率调制增强, 是固有delta频段振荡具有特定于频率的反相夹带所致^[39]。

基于动物实验的听觉信息加工机制研究也存在一定的局限。首先, “语音”层级式调制相关区域在不同物种中的数量和功能仍存在争议。哺乳动物的听觉系统包括大量相互连接的核和皮层区域, 这种交互式结构阵列的复杂性给相关研究带来了一定的挑战。目前已有研究探明了大多数的皮质下核成分的功能连接模式^[35], 但是不同物种在皮质组织中存在较大的变异性。例如, 在皮质下水平, 猴子的听觉系统的结构组成类似于非灵长类动物, 其听觉核心区与黑猩猩和人类均非常相似, 但是皮质水平的结构组织是不同的。这种变异性使得从单一动物模型到人类模型的泛化过程极其艰巨。

1.2 基于人类的初级声学信号时频编码研究

语音信号在听觉皮层的表达逐渐抽象化。早期语音加工系统提取声学-语音特征进行ST表征, 其可变性一般表现为: 音素的发音模式视具体内容而定 (例如: museum与music中“u”的音长与重音是具有差异的), 通过早期计算将语音信号的时频特征抽象化为不同的前词汇表征, 并产生初步预期 (prediction)。ST响应的不变性表现为: 对于声道等发音器官、口音等因素导致的发音差异, 通过小样本训练实现校准, 保证ST特征的稳定表征^[40]。以下将从该过程的响应时序、皮质映射模型以及振荡特点进行归纳与概述。

通过比较语音序列中音素和非音素的神经响应模式差异发现, 语音起始时刻 (voice onset time, VOT) 后 50~200 ms 时程内, PAC、颞横回 (Heschl's gyri, HG)、颞平面 (planum temporale, PT)、STC以及运动功能相关皮质区域的神经活动表现出对于声学-语音特征的偏好响应^[6, 41]。首先, 该通路中双侧STC中部PAC延伸至后部HG与PT等区域实现语音信号ST特征编码, 进而在左侧颞上回 (superior temporal gyrus, STG) 形成音素类别预映射, 归一化音素表达差异, 为后续音素处理提供抽象化信息。以下将分别对上述皮质区域的神经学研究进展进行总结与归纳。

语音感知早期, 存在对于外界语音输入的感知门控 (sensory gating) 效应, 常用听觉事件相关电位 (event related potential, ERP) 成分 P50 的幅值进行表征^[42]. Godey 等^[43] 基于 MEG 以及 SEEG 技术对听觉相关早期成分 P50、N100 进行源定位, 发现 P50 定位于 HG, N100 定位于 HG 的中部、侧部以及 PT. 上述 P50 与 N100 成分的幅值受有意义语音刺激调制^[44]. 例如, 基于音素的 oddball 实验结果证明, N100 成分在音素匹配条件下呈现幅值增加^[40].

位于 HG 后部的 PT 则表现出对声学-语音变化以及复杂声音信号的敏感性. 研究表明, PT 的声学-语音特征响应水平受刺激材料的声学复杂度调制, 但不受语义影响^[45-46]. Humphries 等^[47] 基于功能网络分析对该区域的功能性神经学基础进行研究, 结果表明 PT 在功能上衔接了 PAC 与上颞沟 (superior temporal sulcus, STS), 其中, PAC 对于语音 ST 特征具有适应性响应特点, STS 主要参与词汇前表征过程.

左侧 STG 对于不同类别音素具有一定的早期响应差异. 以 N100 成分为信号源建立的辅音映射皮质图中, 左侧 STG 对于辅音/d/和/t/的响应位点比对/g/和/k/的响应位点更偏向前部大约 7 mm^[41]. 韦尼克氏失语症 (Wernicke's aphasia, WA) 患者病理性研究证实, 左侧 STG 参与初级声学信号时频编码: WA 患者左侧 STG 病变导致听觉腹侧流损伤, 听觉网络中声学-语音加工受到破坏, 语义层级的输入由抽象化表征映射降级为初级声学刺激, 从而影响到语义加工效果^[48]. 此外, Penfield 等^[49] 建立语音时长感受野分布图表明: HG、PT、与 STG 分别对于 200、300 以及 600 ms 的声音刺激具有更强响应. 验证了上述皮质在语音感知期间的响应时序.

在语音感知功能网络研究中, 运动相关皮质的引入具有至关重要的意义. 研究者结合 fMRI 与 EEG 技术, 采用失真语音与线性调频音调的鉴别任务范式有效分离了初级听觉处理与高级语音处理过程. 发现听觉-运动网络的激活起始于语音感知任务早期: VOT 后 95~230 ms 期间双侧 STC、左半球下顶叶皮质 (inferior parietal cortex, IPC) 和腹侧中央沟呈现持续激活^[50]. 此外, Liebenthal 等^[30] 提出了另一种听觉-运动网络的早期激活模式: VOT 后 50~150 ms 期间, 发音相关躯体运动区域与

STC 中部产生同步激活, 即运动相关皮质的参与起始于语音感知任务中的时频编码阶段. Sidiropoulos 等^[51] 进行传导性失语症患者病理研究发现, 顶-颞功能连接可能是时频编码处理的关键组成部分.

从声音映射到有意义的表征涉及在不同时间尺度上集成信息, 即进行多时间分辨率的处理. 音节水平的信息呈现时间较长, 大约为 150~300 ms. 这种较长的信息大致与语音的包络相称, 包含音节边界和音节速率的提示等信息. 因此, 在语音感知初始阶段, 特别是连续声音流的分割期间, 听觉皮层中的低频 (delta, theta) 神经元振荡起着至关重要的作用, 以右侧脑区占优的低频采样网络是音节水平信息抽象表征的前提^[52-54]. 研究表明, 听觉皮层中的低频神经元活动参与追踪语音中缓慢的声学能量波动^[54]、编码声学特征^[55]、解析音节边界以及在复杂的感知环境中选择特定感觉信息^[56]. 然而, 要确定构成语言结构的片段序列的处理机制, 需要在更窄的时间窗口 (大约 20~50 ms) 中编码信息. 因此, 高频带的采样网络用于片段水平信息的明确表征^[20]. 多分辨率处理的神经组织模式目前尚存争议: 有研究认为, 该处理具有半球非对称性, 右半球显示出对长期积分的选择性偏好, 左半球则对不同积分时间尺度的反应呈现较低的选择性^[57]. 但 Marc 等^[58] 则发现左半球在处理 (快速) 时间信息方面占主导地位, 右半球在处理频谱信息方面占主导地位.

在语音处理期间, 神经活动与声学语音输入中缓慢时间波动 (包络) 的对准被称为语音神经夹带 (speech-brain entrainment). 主要表现在低频神经活动的波形 (<8 Hz) 和 high gamma (>40 Hz) 活动的功率包络中^[55]. 如 delta 夹带与声韵律感知有关, 并且可以表征声音属性以及抽象句法的编码, 但是其语音特异性仍存在争议; theta 夹带编码语音清晰度相关特征^[52], 但是, 也有研究表明 theta 夹带可能主要反映了语音加工过程中的早期声学分析; 该过程还存在与 beta 频带相关的效应, 如 beta 频带参与听觉-运动交互, theta-低 beta 跨频耦合 (cross-frequency coupling, CFC) 参与记忆处理, 高 beta 参与语音声学波动的追踪^[59]; 这种同步化响应不仅与声刺激的物理性质联系紧密^[60], 还受到自上而下的认知功能 (例如注意力) 的强烈调制^[61]. 此外, theta 和高频振荡之间的相位-幅度耦合可以提供一种相关的编码策略^[62]. 值得注意的

是,低频和高频神经振荡夹带语音之间的差异机制尚存在争论.例如,不同振荡源是否存在层级式耦合?若存在这种耦合,theta振荡是否在语言内容编码过程中作为一种“主振荡”机制?若跨频耦合机制是在不同的大脑区域之间起作用,那么较低频率的振荡源是否持续限制较高频率的振荡源活动?未来的研究将明确关于不同的大脑振荡源对语音的反应,它们如何相互作用以及它们对语音理解的预期影响.此外,进一步理解语音夹带的功能相关性,需要严格的分离不同语音成分以及信息加工过程,才能明确皮层活动同步化编码的语音处理过程.

综上所述,基于相关研究结果,语音加工阶段早期主要通过PAC、HG、PT、STC以及运动功能相关皮质区域实现初级声学信号时频编码加工,delta、theta、beta以及gamma频段振荡活动均对其具有重要作用.有研究认为,这一信息处理过程具有微弱的语言优势半球偏侧化效应,但是,由于语音处理的神经组织相关研究具有显著的任务相关性,因此也有研究者认为将声音信号转换为与心理词典相关表达形式的计算集的过程,更依赖于双边组织的腹侧通路(尽管两个半球之间存在重要的计算差异).(皮质映射归纳见表1,ERP成分归纳见表2).

2 初级声学信号时频编码神经学基础研究进展

音素是语言上有意义的最小单位,因此,对抽象化音素表征进行类别划分,对于语音感知而言至关重要.大脑中包含音素类别的信息库建立于认知系统发育初期^[63],随年龄增长,信息库更新灵活性逐渐降低,仅能进行少量适应性调整.音素处理期间,信息流开始分别传入大脑皮层腹侧与背侧通路:其中,前者自STG背侧起,经颞中回(middle temporal gyrus, MTG)后部向前部延伸,将抽象化信息进行音素加工并映射为词汇的概念表征,后者则传至SSC下部及运动相关皮质,产生语音所包含的发音运动信息^[20].

2.1 音素处理响应时序研究

目前,通常采用基于特定音素的声学连续体对音素处理响应时序进行研究,如辅音-元音(consonant-vowel, CV)音节.结果表明,VOT后约150 ms,STC开始编码音素特征,并在200 ms

左右进行音素分类映射^[64-65].Humphries等^[47]建立了语音感知期间听觉皮质层级响应模型,其中ST特性信息由STG背侧传至腹侧前部,进行更复杂抽象的音素表征.Liebenthal等^[66]认为,单音素特征表达有可能表现为VOT后180~230 ms的颞叶皮质响应.但此类范式存在刺激材料局限,即受试者对刺激完全熟悉后,音素特异响应减弱甚至消失.Mesgarani等^[64]采用500句自然连续语音材料进行实验,验证了诱发模式稳定的单音素特征响应时程.

2.2 音素处理加工机制研究

音素处理机制研究主要涉及音素与非音素、辅音与元音、清辅音与浊辅音的响应模式差异.音素与非音素的感知差异主要定位于STG,且部分研究认为存在偏侧化效应.Humphries等^[47]采用音素及具有类似ST特征的非音素刺激(如声谱反转语音)判别实验研究证明,与非音素刺激相比,双侧STG腹侧对音素刺激响应更强,右侧PT部分区域响应模式相反.DeWitt和Rauschecker^[32]通过对100例功能成像实验进行meta分析,验证了人类听觉腹侧流对复杂声音的层级式偏好,其中与短时间尺度模式(音素)处理相关的激活位于左侧STG背侧,将音素整合为复杂时间模式(单词)的激活位于左前STG.

听觉皮质对辅音、元音音素处理具有偏差效应.普遍认为辅音加工与词汇处理更相关,而元音加工与韵律/句法处理更相关.分工假说(division of labor hypothesis)中假定辅音在词汇处理(学习和识别单词时)中比元音更重要,并且辅音处理具有更强的明确性,这种辅音偏差(consonant bias, C-bias)得到了许多以婴、幼儿及成年人为研究对象的实验支持^[67-68].研究发现,婴儿一般在8~11个月开始产生稳定的C-bias,双母语婴儿则提前至6~8个月.并且,这种偏差可能与初级声学信号时频编码加工能力习得有关,如连续语音的分段.但是,部分研究发现在产生稳健的C-bias之前,具有相反的元音偏差(vowel bias, V-bias)效应^[69].例如Bouchon等^[70]对于5个月的婴儿进行自我姓名的单音素误差感知实验,发现婴儿对于元音变化的敏感性更强.并将其归因于出生前的声学-语音特征感知经验.此外,元/辅音偏差的发展具有语种特异性:法语使用者C-bias的产生早于英语使用者,丹麦语使用者则始终呈现稳定的V-bias^[71].

皮质分布式调制可能是上述 C/V-bias 效应的潜在机制. 一项 ECoG 研究结果证实, 左侧 STG 中、后部对于不同英语音素具有选择性调制, 即分布式集群响应 (distributed population response)^[64]. fMRI 研究发现, 元音或辅音序列的失匹配刺激均会诱发双侧 STG 背侧响应. 其中, 元音失匹配响应聚类位于 STG 中部的较小皮质区域, 辅音失匹配响应聚类沿 STG 背侧延伸至 HG. 除 STG 外, 其他皮质区域也观察到元辅音响应差异: 如右侧枕外侧沟存在辅音特征响应, 左侧 MTG 和双侧颞中央沟 (middle temporal sulcus, MTS) 中存在元音特征响应. 研究认为, 这种皮质偏好表征了大脑对于 ST 特征快速转变的刺激 (辅音) 以及 ST 特征呈现稳态的刺激 (元音) 具有不同加工模式^[47].

大脑对于清辅音与浊辅音也呈现出经验调制的感知差异. 研究者基于成对的清、浊辅音 (如/p/、/d/) 设定声学共振峰梯度变化的声学刺激, 发现其并未诱发梯度性的感知差异, 而是存在明显的感知类别边界. 类似的, 音素判别任务中, 即使清、浊辅音目标刺激与靶刺激具有相等的音素距离, 但目标刺激与靶刺激为不同类别时, 判别准确率较高^[72-73]. 且音素判别任务行为学表现水平与 IFC 后部的音素响应强度呈正相关^[30]. 进一步研究辅音/元音不对称处理的可能性因素是非常必要的, 例如所涉及音素的声学/语音特性, 辅音作为词汇访问主要线索的潜在加工机制, 以及范式、年龄等因素对于辅音/元音加工的调制作用等.

近年兴起的语音感知运动理论 (motor theory of speech perception) 模型重申了运动相关皮质在音素处理中的作用. Rogers 等^[74]采用持续 theta 脉冲磁刺激 (continuous theta burst stimulation, cTBS) 作用于受试者唇部运动皮层 20~35 min, 发现语音感知任务中包含唇部发音音节的跨音素类别刺激识别水平较其他组别呈现显著下降. 该团队通过进一步研究, 验证了听觉-运动交互模型中, 语音信息自听觉皮质背侧传递至顶叶 SSC 下部, 并延伸至额叶运动区域这一假设^[22].

音素处理阶段实现了抽象语音特征的音素类划分, 并在词汇前水平产生语音内容的有效预期. 该阶段不同功能皮质对于不同音素具有偏好响应, 并且产生了更加明显的功能皮质左侧化效应 (皮质映射归纳见表 1, ERP 成分归纳见表 2).

3 词汇-语义加工神经学基础研究进展

在沟通与交流中, 听众需要快速地感知语音、理解语义, 进一步对语音内容进行合理预期, 并在有限的时间限制内准备相关回应. 更高层级的功能皮质通过整合音素信息及词汇预期信息形成独立词汇, 并基于词汇、语法、语境等信息, 实现语义理解, 即词汇-语义加工过程. 本节将从典型响应时序以及加工机制综述这一认知过程的研究进展.

3.1 词汇-语义加工响应时序研究

按照响应时序, 词汇-语义加工过程可划分为词汇加工与语义加工. 以下将分别对词汇加工阶段中的 N200、早期左前部负波 (early left anterior negativity, ELAN), 以及语义加工阶段中的 N400、语音失匹配负波 (phonological mismatch negativity, PMN)、N400-P600 以及 P600 等 ERP 成分相关研究进展进行总结与归纳.

3.1.1 词汇加工响应时序研究

词汇加工阶段中主要涉及 N200 与 ELAN 成分. 相关研究提出, 与早期词汇选择过程相关的 N200 成分一般出现在刺激后 200~310 ms 左右 (注: 该成分的响应模式与基于图片的 go/no go 范式诱发的 N200 不同), 定位于前扣带皮质 (anterior cingulate cortex, ACC)^[75], 一般表征词汇识别的开始. 该成分也与语法处理有关, 语义和/或语法错误均会引起该成分幅值增大.

一般认为, 早期阶段的时间窗口中, 短语结构错误会诱发 ELAN 成分, 响应时序一般滞后于 N200 成分, 主要响应区域为左侧顶颞区. 这一成分被认为表征了与语义约束和任务要求无关的语音初始解析过程^[76]. 此外, 词汇重音分布等音节结构差异, 也会导致词汇加工差异. 研究发现, 与第一音节非重音的单词相比, 第一音节为重音的单词在处理中会产生一定的延迟效应, 主要表现为左后部电极 300~600 ms 期间事件相关负波峰值潜伏期延长. 重音分布模式不一致所诱发的失匹配响应 (mismatch response) 也在这一时程内, 特征皮质主要定位于双侧 IFG^[40].

3.1.2 语义加工响应时序研究

目前普遍认为语音感知过程中的晚期 ERP 成分表征语义处理相关的神经响应, 如 N400、PMN、N400-P600 以及 P600 等. 最典型的 N400 受词义整合及语义相关信息调制: 波形在 VOT 后约

400 ms 达到峰值, 且当前词汇不符合语义连贯性预期或在语义上不正确时, 幅值显著增加^[77]. 同时, 该成分受任务要求调控, 并在特定范式下响应句法错误^[76]. 此外, Song 与 Iverson^[78] 研究认为, N400 可以表征不利语言环境中语音感知系统的补偿机制. 例如母语使用者 (native speaker, NS) 与非母语使用者 (non-native speaker, nNS) 语音感知实验结果表明, NS 在感知由 NS 或 nNS 录制的语音材料期间, 会对后者的语音包络产生更强的神经夹带, 并且表现出 N400 成分幅值增加; 类似的结果也呈现在 nNS 中. 这一现象可能表征了异常语音环境下感知与加工的补偿机制. N400 成分也常用于失语症病理研究. 结果表明, WA 患者在词图匹配 (word-picture matching, WPM) 处理任务中呈现较低幅度的 N400, 但是词图匹配与失匹配差异, 在正常人对照组与患者组中诱发了类似的 N400 响应模式. 由于 WA 患者脑损伤常伴随不同程度的特异性扩散, 如下颞叶、顶叶甚至额叶区域, 因而引入 PMN 进行 WPM 范式响应机制的进一步分析. 该成分一般在 VOT 后 160~260 ms 时程内进行响应, 特征皮质位于额叶. 研究发现, PMN 可以对任务操作表现进行预测, 并且 WA 患者额叶中 PMN 幅值降低^[48].

呈现在句末后约 600 ms 的 P600 成分对句法相关信息加工异常敏感. 该成分通常在句法复杂性偏高或者句意整合困难时呈现幅值增加^[79]. 研究认为 P600 可以表征经过反复分析后的句意理解水平^[80]. 此外, Sheppard 等^[81] 基于表达性失语症 (Broca's aphasia, BA) 患者的研究提出, N400-P600 复合波成分可以表征句意加工能力, 即句意歧义范式诱发的神经响应模式中, BA 组存在 N400 成分的响应迟滞, 且无 P600 成分, 并揭示该类患者存在韵律信息与词汇语义信息整合功能障碍.

3.2 词汇-语义加工机制研究

词汇-语义加工功能皮质主要涉及前颞叶 (anterior temporal lobe, ATL)、后颞叶 (posterior temporal lobe, PTL), STG, MTG 以及 IFG^[20]. 腹侧流被认为是语义信息的主要加工途径. 短时间尺度语音信息传入颞叶前部进行整合, 经颞下沟 (inferior temporal sulcus, ITS) 后部以及 MTG 进行词汇水平加工, 产生的词汇信息传递至 IFG 腹侧进行语义解码, PTL 实现了词汇语义信息与语法信息的融合. 此外, 部分研究表明左侧角回也涉及词

汇检索以及词汇概念的表征^[82]. 值得注意的是, 经典双流模型中还提及了顶颞边界的外侧裂 (sylvian fissure at the parietotemporal boundary, Spt) 区域, 视其为感觉运动交互区域, 该区域在在语音感知处理期间, 负责接收来自其他感官的信息, 并进行有效整合, 以促进语音解码^[83]. 复杂或失真语音信号的处理, 一般涉及 IFG 后部 (BA 44) 与 STG / STS 后部, 前者负责非相邻语音元素层级处理结构的构建, 后者负责在复杂句子处理中整合语义和句法信息.

在汉语等声调语言中, 具有不同音高轮廓的词汇声调是提供词义信息的一个关键特征. Si 等^[84] 基于 ECoG 进行了汉语的音调-语义加工研究. 通过不同区域神经响应潜伏期差异建立了由 STG 传至运动皮质, 再扩散到 MTG 的信息加工时序, 验证了 STG 与 MTG 在语义映射中的层级处理模式, 以及对于中文使用者而言, ATL 作为一种语音处理的神经生理学标志物, 对于音素表征与语义的媒介功能.

高频活动 (>40 Hz) 是感知语音的可靠指标, 主要参与词汇记忆检索或语法类别处理, 并且可能反映了语音分析的结果^[85]. 值得注意的是, 低频神经夹带语音声音动力学对语义的理解而言也是至关重要的, 并且在听觉受损情况下, 语音包络状经颅电刺激可用于调控语音理解水平^[86], 但是无法明确主要影响因素为电刺激的频率特征或形状特征, 包络频率范围内的正弦 tACS 是否会提供更加有效的调制作用也未可知. 此外, 该过程存在不同脑区神经夹带的相互作用. 研究表明, STG 左前部的 delta 节律同步化受眶额区的 beta 功率调制, 可能反映了预期在听觉编码过程中自上而下的调制作用. STG 右后部的 delta 节律同步化调制了顶区的 theta 功率, 可能表征语义加工过程中记忆的参与.

综上所述, 词汇-语义加工是语音感知中最为复杂的认知过程. 根据激活时序, N200、ELAN、PMN、N400、P600 以及 N400-P600 复合波分别对这一时程的不同特征进行了表征; 对于皮质加工模式, 形成了从 PTL 到 ATL 并最终延伸至 IFG 的层级式偏侧化神经网络组织, 实现了词汇以及语义的高效处理. 在词汇-语义加工过程中, 功能皮质左侧化现象显著 (皮质映射归纳见表 1, ERP 成分归纳见表 2).

Table 1 Summary of the cortical mapping of speech perception and processing

表1 语音感知与加工皮质映射的研究归纳

脑区	主要结果	所属语音加工过程	文献
STG 中部	对所接收声学信号的时频处理处理; 响应音素的匹配与失匹; 对不同类音素差异响应	初级声学信号时频编码; 音素处理	[41]
双侧STG腹侧	对于音素类刺激具有更强的响应; 进行更复杂更抽象的音素表征	初级声学信号时频编码; 音素处理	[47]
左前STG	对复杂声音的层级式偏好, 音素处理	初级声学信号时频编码; 音素处理	[32]
STS	前词汇信息的处理具有一定的敏感性	初级声学信号时频编码	[87]
HG外侧部分	响应广泛的听觉刺激	初级声学信号时频编码	[44]
PT	对于声学-语音变化以及复杂频谱的敏感性	初级声学信号时频编码	[45]
PT	连接PAC与STS	初级声学信号时频编码	[47]
STG背侧	元音与辅音的响应模式差异	音素处理	[47]
左侧STS	单个音素的特征响应	音素处理	[66]
顶叶SSC下部	与听觉皮层进行交互, 表征发音特征	音素处理	[22]
腹侧中央沟	与听觉皮层进行交互, 表征发音特征	音素处理	[50]
右侧枕外侧沟	辅音加工的特征响应皮质	音素处理	[47]
左侧MTG	元音的特征响应皮质	音素处理	[47]
双侧MTS	元音的特征响应皮质	音素处理	[47]
唇部运动皮层	皮层磁刺激使包含唇部发音的音素鉴别能力下降	音素处理	[74]
顶颞边界的外侧裂Spt	感觉运动信息整合	词汇-语义加工	[82]
左侧角回	词汇检索以及词汇概念的表征	词汇-语义加工	[82]
ATL	对抽象化的音素信息进行整合	词汇-语义加工	[88]
PTL	音素和语义信息的连接	词汇-语义加工	[89]
MTG	音素和语义信息的连接	词汇-语义加工	[88]
IFC后部	音素类别选择性响应水平与对语音刺激进行分类的行为学表现呈正相关	音素处理; 词汇-语义加工	[30]
左侧IFC	句意复杂的语音信号处理, 负责非相邻元素层级结构的构建	词汇-语义加工	[90]
右侧额颞区	韵律感知	词汇-语义加工	[91]

Table 2 Summary of the ERPs of speech perception and processing

表2 语音感知与加工ERPs的研究归纳

成分	响应时程 范围/ms	响应脑区	主要结果	所属语音加工过程	文献
P50	50~100	HG外侧部分	响应于类型广泛的听觉刺激, 但其潜伏期以及响应位点受到有意义语音刺激的调制	初级声学信号时频编码	[44]
N100	100~150	HG外侧部分, STG	响应语音的声学-语音特征, 潜伏期以及响应位点受到音素差异调制	初级声学信号时频编码; 音素处理	[41]
P150	150~250	STC	开始对语音的音素特征进行编码	音素处理	[65]
N200	180~230	左侧STS	音素分类的神经处理	音素处理	[66]
N200	200~310	ACC	早期词汇选择; 语法错误引起幅值增加	词汇-语义加工	[75]
ELAN	N200之后	顶颞区	句意有效性	词汇-语义加工	[76]
负向成分	300~600	左后部	词汇中的音节结构差异, 例如重音分布	词汇-语义加工	[40]
PMN	160~260	额叶	对受试者对于词图匹配任务的操作表现进行预测	词汇-语义加工	[48]
N400	300~500	颞叶, 顶叶, 额叶	词汇后语义整合	词汇-语义加工	[77]
N400-P600复合波	300~600	颞叶, 顶叶, 额叶	整合韵律线索与词汇语义信息	词汇-语义加工	[81]
P600	600	颞叶, 顶叶, 额叶	句法相关信息的加工异常敏感; 表征经过反复分析后的句意理解水平	词汇-语义加工	[80]

4 总 结

归纳前文语音处理过程中3个主要阶段的皮质映射、神经振荡模式以及事件相关响应机制等神经语言学研究现状,可以得出以下结论:语音早期加工阶段的前词汇处理主要是通过PAC延伸至后部HG与PT等区域对于语音信号进行ST特征编码,并将其映射为可能的音素类别,实现相同音素表达差异的归一化,为后续的音素处理提供抽象信息.音素处理阶段,信息流开始分别流向腹侧与背侧通路.其中,前者向颞叶前部延伸,将抽象化的信息映射为词汇的概念表征,后者与运动相关皮质进行交互作用,产生语音相关的发音运动信息.最后自左侧PTL到ATL并最终延伸至IFG的偏侧化脑网络组织,对于词汇以及语义进行了高效的层级式处理.但是,语音处理的神经学基础研究中仍然存在一些亟待解决的难题:例如,各频率成分的潜在作用机制以及响应模型存在争议,尚未建立精确和完善的功能定位,听觉流内组织计算操作细节颇具争论,失语症等病理学补偿机制存在一定的特异性,不同理论模型尚存在与神经生理学现象的不一致等局限.以上问题均有待深入研究.

整合不同领域的研究成果可能是该领域进一步研究工作的方向之一.神经计算模型相关研究对语音感知系统进行了概念化表征,但是仅该类研究不能产生语音处理的神经解剖模型,但它可以通过提供理论框架来指导神经科学研究.基于认知减法范式的功能性神经影像学广泛地定义了语音处理的神经解剖学基础,同时神经影像数据的多变量分析技术有可能更详细地研究语音的时频编码和抽象化过程.已有研究表明^[92],使用神经数据作为输入来训练语言处理的计算模型是可行的.因此,整合来自认知科学,神经生理学以及功能成像的相关进展,可能进一步理解皮质中语音信号的处理与表达.另外,单一神经成像技术在语音处理相关研究中具有一定的局限.例如,神经振荡的电生理学基础对于感知对象形成过程而言具有至关重要的作用,振荡网络中的同步性变化也是听觉和语音感知的标志^[93].但依赖于代谢测量(例如fMRI)的功能性神经成像方法在研究振荡网络方面具有一定的局限性.因此,将fMRI研究在空间解剖领域的结果,与EEG、MEG等高时间分辨率技术进行结合,是目前语音处理相关研究进一步发展的可能方向之一.

该领域研究的进一步深入和完善,将为失语症等中枢神经系统疾病的临床诊治建立更加可靠的干预体系,也将进一步为当代类脑智能等人工智能领域发展提供重要的生物学基础.

参 考 文 献

- [1] Leung J H, Purdy S C, Tippet L J, *et al.* Affective speech prosody perception and production in stroke patients with left-hemispheric damage and healthy controls. *Brain & Language*, 2017, **166**:19-28
- [2] Peach R K. Cognitive Approaches to aphasia treatment: application of the cognition of language to aphasia intervention. *Seminars in Speech & Language*, 2017, **38**(01):3-4
- [3] Jaeger H. Artificial intelligence: Deep neural reasoning. *Nature*, 2016, **538**(7626):467-468
- [4] Nguyen C H, Karavas G, Artemiadis P. Inferring imagined speech using EEG signals: a new approach using Riemannian Manifold features. *Journal of Neural Engineering*, 2017, **15**(1):016002
- [5] Martin S, Brunner P, Iturrate I, *et al.* Corrigendum: Word pair classification during imagined speech using direct brain recordings. *Scientific Reports*, 2016, **6**(1):25803
- [6] Giroud J, Trébuchon A, Schön D, *et al.* Asymmetric sampling in human auditory cortex reveals spectral processing hierarchy. *PLoS Biol*, 2020, **18**(3):e3000207
- [7] Chang E F, Breshears J D, Raygor K P, *et al.* Stereotactic probability and variability of speech arrest and anomia sites during stimulation mapping of the language dominant hemisphere. *Journal of Neurosurgery*, 2017, **126**(1):114-121
- [8] Babajani-Feremi A, Narayana S, Rezaie R, *et al.* Language mapping using high gamma electrocorticography, fMRI, and TMS versus electrocortical stimulation. *Clinical Neurophysiology*, 2016, **127**(3):1822-1836
- [9] Esti B E, Karen E, Liina P N. Language switching decomposed through MEG and evidence from bimodal bilinguals. *Proceedings of the National Academy of Sciences*, 2018, **115**(39):9708-9713
- [10] Altvater-MN, Grossmann T. The role of left inferior frontal cortex during audiovisual speech perception in infants. *Neuroimage*, 2016, **133**:14-20
- [11] Dietrich S, Müller-Dahlhaus F, Ziemann U, *et al.* The role of pre-SMA for time-critical speech perception-a transcranial magnetic stimulation (TMS) study//Wolters M, Livingstone J, Beattie B, *et al.* *Proceedings of ICPhS 2015*. London: International Phonetic Association: Neurophonetics, 2015
- [12] Knudsen E I. Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience*, 2004, **16**(8): 1412-1425
- [13] Perszyk D R, Waxman S R. Infants' advances in speech perception shape their earliest links between language and cognition. *Scientific Reports*, 2019, **9**(1):3293
- [14] Sharma A, Campbell J, Cardon G. Developmental and cross-modal plasticity in deafness: evidence from the P1 and N1 event related potentials in cochlear implanted children. *International*

- Journal of Psychophysiology, 2015, **95**(2):135-144
- [15] Wernicke C. Der Aphasische Symptomenkomplex. Berlin Heidelberg: Springer-Verlag, 1974:18-25
- [16] Lichtheim L. On aphasia. Brain, 1885, **4**(7):433-484
- [17] McClelland J L, Elman J L. The TRACE model of speech perception. Cognitive Psychology, 1986, **18**(1):1-86
- [18] Marslen-Wilson W D. Functional parallelism in spoken word-recognition. Cognition, 1987, **25**(1):71-102
- [19] Luce P A, Pisoni D B. Recognizing spoken words: the neighborhood activation model. Ear & Hearing, 1998, **19**(1):1-36
- [20] Hickok G, Poeppel D. The cortical organization of speech processing. Nature Reviews Neuroscience, 2007, **8**(5):393-402
- [21] Looijestijn J, Diederik K. The auditory dorsal stream plays a crucial role in projecting hallucinated voices into external space. Schizophrenia Research, 2013, **146**(1-3):314-319
- [22] Rothwell J. Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. Cerebral Cortex, 2015, **25**(10):3690-3698
- [23] Cummine J, Boliek C A. Understanding white matter integrity stability for bilinguals on language status and reading performance. Brain Structure and Function, 2013, **218**(2):595-601
- [24] Molholm S, Mercier M R, Liebenthal E, *et al.* Mapping phonemic processing zones along human perisylvian cortex: an electro-corticographic investigation. Brain Structure & Function, 2014, **219**(4):1369-1383
- [25] Franck-Emmanuel R, Krasimir M, Jean-Baptiste D, *et al.* Electrostimulation mapping of comprehension of auditory and visual words. Cortex, 2015, **71**:398-408
- [26] Mckinnon E T, Fridriksson J, Glenn G R, *et al.* Structural plasticity of the ventral stream and aphasia recovery. Annals of Neurology, 2017, **82**(1):213-220
- [27] Venezia J H, Saberi K, Chubb C, *et al.* Response bias modulates the speech motor system during syllable discrimination. Frontiers in Psychology, 2012, **3**:157
- [28] Rauschecker J P, Scott S K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nature Neuroscience, 2009, **12**(6):718-724
- [29] Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron, 2011, **69**(3):407-422
- [30] Liebenthal E, Möttönen R. An interactive model of auditory-motor speech perception. Brain & Language, 2018, **187**:33-40
- [31] 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. Brain & Language, 2017, **164**:77-105
- [32] DeWittl, Rauschecker J P. Phoneme and word recognition in the auditory ventral stream. Proceedings of the National Academy of Sciences of the United States of America, 2012, **109**(8):505-514
- [33] Monahan J. Philip D P. The interface between neuroscience and psychological science Speech perception: cognitive foundations and cortical implementation. Current Directions in Psychological Science, 2008, **17**(2):80-85
- [34] Ohl F W, Scheich H, Freeman W J. Change in pattern of ongoing cortical activity with auditory category learning. Nature, 2001, **412**(6848):733-736
- [35] Kaas J H, Hackett T A. Subdivisions of auditory cortex and processing streams in primates. Proceedings of the National Academy of Sciences of the United States of America, 2000, **97**(22):11793-11799
- [36] Fritz J, Shamma S M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nature Neuroscience, 2003, **6**(11):1216-1223
- [37] Petkov C I, Kayser C, Steudel T, *et al.* A voice region in the monkey brain. 2008, **11**(3):367-374
- [38] Fitch W T, Boer B D, Mathur N, *et al.* Monkey vocal tracts are speech-ready. Science Advances, 2016, **2**(12):e1600723
- [39] O'Connell M N, Barczak A, Schroeder C E, *et al.* Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. Journal of Neuroscience, 2014, **34**(49): 16496-16508
- [40] Schild U, Becker A B C, Friedrich C K. Phoneme-free prosodic representations are involved in pre-lexical and lexical neurobiological mechanisms underlying spoken word processing. Brain & Language, 2014, **136**(100):31-43
- [41] Jonas O, Scott S K, Carsten E. Now you hear it, now you don't: transient traces of consonants and their nonspeech analogues in the human brain. Cerebral Cortex, 2006, **16**(8):1069-1076
- [42] Kurthen M, Trautner P, Rosburg T, *et al.* Towards a functional topography of sensory gating areas: invasive P50 recording and electrical stimulation mapping in epilepsy surgery candidates. Psychiatry Research: Neuroimaging, 2007, **155**(2):121-133
- [43] Godey B, Schwartz D, De Graaf J B, *et al.* Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. Clinical neurophysiology, 2001, **112**(10):1850-1859
- [44] Pinheiro A P, Vasconcelos M, Dias M, *et al.* The music of language: An ERP investigation of the effects of musical training on emotional prosody processing. Brain & Language, 2015, **140**:24-34
- [45] Jonas O, Jonas Z, John V M, *et al.* Multiple stages of auditory speech perception reflected in event-related fMRI. Cerebral Cortex, 2007, **17**(10):2251-2257
- [46] Elmer S, Kühnis J, Rauch P, *et al.* Functional connectivity in the dorsal stream and between bilateral auditory-related cortical areas differentially contribute to speech decoding depending on spectro-temporal signal integrity and performance. Neuropsychologia, 2017, **106**:398-406
- [47] Humphries C, Sabri M, Lewis K, *et al.* Hierarchical organization of speech perception in human auditory cortex. Frontiers in Neuroscience, 2014, **8**(4):406
- [48] Robson H, Pilkington E, Evans L, *et al.* Phonological and semantic processing during comprehension in Wernicke's aphasia: An N400 and phonological mapping negativity study. Neuropsychologia, 2017, **100**:144-154

- [49] Penfield W, Roberts L. *Speech and Brain Mechanisms*. USA: Princeton University Press, 2014
- [50] Liebenthal E, Sabri M, Beardsley S A, *et al.* Neural dynamics of phonological processing in the dorsal auditory stream. *Journal of Neuroscience the Official Journal of the Society for Neuroscience*, 2013, **33**(39):15414-15424
- [51] Sidiropoulos K, De-Bleser R H, Preilowski B. Pre-lexical disorders in repetition conduction aphasia. *Neuropsychologia*, 2008, **46**(14):3225-3238
- [52] Ding N, Simon J Z. Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in Human Neuroscience*, 2014, **8**:311
- [53] Keitel C, Quigley C, Ruhnu P. Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response?. *Journal of Neuroscience*, 2014, **34**(31):10137-10140
- [54] Ding N, Melloni L, Zhang H, *et al.* Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 2016, **19**(1):158-164
- [55] Pasley B N, David S V, Mesgarani N, *et al.* Reconstructing speech from human auditory cortex. *PLoS Biology*, 2012, **10**(1):e1001251
- [56] Rimmele J M, Golumbic E Z, Schröger E, *et al.* The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex*, 2015, **68**:144-154
- [57] Boemio A, Fromm S, Braun A, *et al.* Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 2005, **8**(3):389-395
- [58] Nwiesner M S, Rübsamen R, Von Cramon D Y. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur J Neurosci*, 2005, **22**(6):1521-1528
- [59] Mai G, Minett J W, Wang W S Y. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *Neuroimage*, 2016, **133**:516-528
- [60] Steinschneider M, Nourski K V, Fishman Y I. Representation of speech in human auditory cortex: is it special?. *Hearing Research*, 2013, **305**:57-73
- [61] Golumbic E M Z, Ding N, Bickel S, *et al.* Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, 2013, **77**(5):980-991
- [62] Hyafil A, Giraud A, Fontolan L, *et al.* Neural cross-frequency coupling: connecting architectures, mechanisms, and functions. *Trends in Neurosciences*, 2015, **38**(11):725-740
- [63] Werker J F, Tees R C. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. 1984, **1**(7):49-63
- [64] Mesgarani N, Cheung C, Johnson K, *et al.* Phonetic feature encoding in human superior temporal gyrus. *Science*, 2014, **343**(6174):1006-1010
- [65] Maritz R G, Juan S P, Lindsay K, *et al.* Principal component analyses and scalp distribution of the auditory P150-250 and N250-550 to speech contrasts in Mexican and American infants. *Developmental Neuropsychology*, 2007, **31**(3):363-378
- [66] Liebenthal E, Desai R, Ellingson M. Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex*, 2010, **12**(20):2958-2970
- [67] Luche C D, Poltrock S, Goslin J, *et al.* Differential processing of consonants and vowels in the auditory modality: A cross-linguistic study. *Journal of Memory & Language*, 2014, **72**(72):1-15
- [68] Mélanie H, Josette S, Thierry N. A consonant/vowel asymmetry in word-for processing: evidence in childhood and in adulthood. *Language & Speech*, 2014, **57**(2):254-281
- [69] Nishibayashi L L, Nazzi T. Vowels, then consonants: Early bias switch in recognizing segmented word forms. *Cognition*, 2016, **155**:188-203
- [70] Bouchon C, Floccia C, Fux T, *et al.* Call me Alix, not Elix: vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, 2015, **18**(4):587-598
- [71] Højen A, Nazzi T. Vowel bias in Danish word-learning: Processing biases are language-specific. *Developmental Science*, 2016, **19**(1):41-49
- [72] Zhang Q, Hu X, Luo H, *et al.* Deciphering phonemes from syllables in blood oxygenation level-dependent signals in human superior temporal gyrus. *European Journal of Neuroscience*, 2016, **43**(6):773-781
- [73] Nourski K V, Steinschneider M, Rhone A E, *et al.* Sound identification in human auditory cortex: Differential contribution of local field potentials and high gamma power as revealed by direct intracranial recordings. *Brain and language*, 2015, **148**:37-50
- [74] Rogers J C, Möttönen R, Boyles R, *et al.* Discrimination of speech and non-speech sounds following theta-burst stimulation of the motor cortex. *Frontiers in Psychology*, 2014, **5**(4):754
- [75] Britton D. *Semantic Attention: Effects of Modality, Lexicality and Semantic Content[D]*. New York: The Graduate Center, City University of New York, 2017
- [76] Hahne A, Friederici A D. Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, 2002, **13**(3):339-356
- [77] Maess B, Mamashli F, Obleser J, *et al.* Prediction signatures in the brain: semantic pre-activation during language comprehension. *Frontiers in Human Neuroscience*, 2016, **10**:591
- [78] Song J, Iverson P. Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition*, 2018, **179**:163
- [79] Friederici A D. The brain basis of language processing: from structure to function. *Physiological Reviews*, 2011, **91**(4):1357-1392
- [80] Key-Delyria S E. Sentence processing in traumatic brain injury: evidence from the P600. *Journal of Speech Language & Hearing Research*, 2016, **59**(4):759-771
- [81] Sheppard S M, Love T, Midgley K J, *et al.* Electrophysiology of prosodic and lexical-semantic processing during sentence comprehension in aphasia. *Neuropsychologia*, 2017, **107**:9-24

-
- [82] Thames A D, Sayegh P, Terashima K, *et al.* Increased subcortical neural activity among HIV+ individuals during a lexical retrieval task. *Neurobiology of Disease*, 2016, **92**(Pt B):175-182
- [83] He Y, Nagels A, Schlesewsky M, *et al.* The role of gamma oscillations during integration of metaphoric gestures and abstract speech. *Frontiers in Psychology*, 2018, **9**:1348
- [84] Si X, Zhou W, Hong B. Cooperative cortical network for categorical processing of Chinese lexical tone. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, **114**(46):12303-12308
- [85] KoSem A, van Wassenhove V. Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Language, Cognition and Neuroscience*, 2016, **32**(5):536-544
- [86] Riecke L, Formisano E, Sorger B, *et al.* Neural entrainment to speech modulates speech intelligibility. *Current Biology*, 2018, **28**(2): 161-169
- [87] Liebenthal E, Desai R H, Humphries C, *et al.* The functional organization of the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults. *Frontiers in Neuroscience*, 2014, **8**:289
- [88] Sarubbo S, De B A, Merler S, *et al.* Structural and functional integration between dorsal and ventral language streams as revealed by blunt dissection and direct electrical stimulation. *Human Brain Mapping*, 2016, **37**(11):3858-3872
- [89] Friederici A D. *Neurobiology of Language*. California: Academic Press, 2016:349-356
- [90] Grodzinsky Y, Friederici A D. Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology*, 2006, **16**(2):240-246
- [91] Sammler D, Grosbras M H, Anwander A, *et al.* Dorsal and ventral pathways for prosody. *Current Biology*, 2015, **25**(23):3079-3085
- [92] Mitchell T M, Shinkareva S V, Carlson A, *et al.* Predicting human brain activity associated with the meanings of nouns. *Science*, 2008, **320**(5880):1191-1195
- [93] Huan L, David P. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, **2007**(54): 1001-1010

The Progress of Research on Basis of Neuroelectrophysiology in Speech Processing*

YU Hai-Qing^{1)**}, XU Min-Peng^{1,2)**}, WAN Bai-Kun¹⁾, MING Dong^{1,2)***}

⁽¹⁾*School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China;*

⁽²⁾*Institute of Medical Engineering & Translational Medicine, Tianjin University, Tianjin 300072, China)*

Abstract The investigation into neural basis of speech processing, including distributed hierarchical organization, neural mechanisms, and functional connectivity of brain, stands in the central of neurolinguistic research. In line with the sequence of speech processing, it can be subdivided into three processing stages: spectrotemporal analysis of primary acoustic signals, phonemic processing, and lexical-semantic processing. Despite underlying neural mechanisms of these stages have been investigated extensively and intensively, it appears that there are still inconsistencies and controversies between different model theories/hypotheses. Consequently, it is crucial to sort out and summarize them. The present review takes the three speech processing stages of the brain as the main line and reviews the research status of neurological mechanism under each stage, covers cortical mapping, neural oscillations, and event-related potential characteristics. Of note, this review focuses more on observations based on the electrophysiological approaches. We aiming to synthesize the latest findings of this field and provide new insights for understanding how speech signals are represented and processed in the human brain.

Key words speech processing, neuroelectrophysiology, spectrotemporal analysis of primary acoustic signals, phonemic processing, lexical-semantic processing

DOI: 10.16476/j.pibb.2020.0050

* This work was supported by grants from The National Natural Science Foundation of China (81601565, 81630051).

** These authors contributed equally to this work.

*** Corresponding author.

Tel: 86-13820222616, E-mail: richardming@tju.edu.cn

Received: May 10, 2020 Accepted: July 17, 2020