

视听整合神经认知机制对听障者多模态感知研究的意义*

周爱保¹ 潘超超¹

DOI:10.3969/j.issn.1006-7299.2020.04.025

网络出版时间:2020/4/21 8:51

网络出版地址:http://kns.cnki.net/kcms/detail/42.1391.R.20200420.0935.032.html

【中图分类号】 R764.43 【文献标识码】 A 【文章编号】 1006-7299(2020)04-0457-06

社会情景中的多种刺激信息经过不同的感觉通道传递到大脑,大脑不断地接收形色各异的刺激信息,但并未将这个世界感知为支离破碎的各种形状、颜色、声音、气味等,而是整合来自多个感觉通道的线索将其知觉为一个统一、连续、稳定的整体,这个加工过程即为多感觉整合(multisensory integration)^[1]。多感觉整合能弥补单通道信息的匮乏及不确定性,有助于人们更迅速而准确的进行认知判断。因为人类所接收的信息中有 97% 来自视觉和听觉,因此,对视听整合的研究最多^[2]。

视听系统是大脑中最重要的信息处理系统。人类在获取外界信息时,往往会同时面对视觉和听觉通道的线索,并由此驱动视听信息的整合来完成信息加工;但听障人群因听力损失,更多地通过视觉系统来获取和加工信息。近年来,随着助听器、人工耳蜗等技术的不断进步及广泛应用,听障者“听不到”声音的问题得到了解决。随之而来的是更为艰巨的康复任务,如何在视觉信息的辅助下,充分利用视听

联动机制,促进听障者在听觉、言语、情绪、认知和人格等多方面的生态化发展,是当前研究工作的重心所在。因此,明确视听整合现象及其神经认知机制,对听障者多模态感知研究具有重要意义。

目前,视听整合的研究大多集中在一些与错觉相关的心理学效应中,如腹语术效应(ventriloquism effect)、麦格克效应(McGurk effect)、声音诱发闪光错觉(sound induced flash illusion)等。其中,最早发现于言语感知领域中的麦格克效应逐渐发展成为视听整合研究的一种经典实验范式——麦格克效应范式(McGurk effect paradigm),其在听障儿童、自闭症及语言学习障碍等方面的研究中均有应用^[3]。本文将在已有研究基础上,对视听整合的研究进展进行综述。

1 言语感知中的视听整合

既往言语感知一直被认为是一种纯粹的听觉加工过程,与视觉线索无关。但近年来,越来越多的学者们对该观点提出了质疑,认为言语感知是视听信息共同作用的结果。与单通道相比,视听双通道具有感知增益的研究结果为该观点提供了间接证据,言语感知过程中麦格克效应的发现更是直接证明了言语感知是视听整合的结果^[4]。在此基础上,研究

* 2019 年度国家民委民族研究项目(2019-GMD-020)

1 甘肃省行为与心理健康重点实验室 西北师范大学心理学院(兰州 730070)

通讯作者:潘超超(Email:zhoulabpapers@163.com)

35 Lei G, Zhao Z, Li Y, et al. A method to induce human cortical long-term potentiation by acoustic stimulation[J]. *Acta Otolaryngol*, 2017, 137:1069.

36 雷冠雄,邓忠,谭东辉,等. 强直声治疗慢性耳鸣及耳鸣评估方法研究[J]. *临床耳鼻咽喉头颈外科杂志*, 2019, 33:858.

37 Hobson J, Chisholm E, El Refaie A. Sound therapy (masking) in the management of tinnitus in adults[J]. *Cochrane Database Syst Rev*, 2012, 14:11.

38 Searchfield GD, Durai M, Linford T. A state-of-the-art review: personalization of tinnitus sound therapy[J]. *Front Psychol*, 2017, 8:1599.

39 Hall DA, Haider H, Szezeppek AJ, et al. Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults[J]. *Trials*, 2016, 17:270.

40 Hoare DJ, Adjamian P, Sereda M, et al. Recent technological

advances in sound-based approaches to tinnitus treatment: a review of efficacy considered against putative physiological mechanisms[J]. *Noise Health*, 2013, 15:107.

41 Makar SK, Mukundan G, Gore G. Treatment of tinnitus: a scoping review[J]. *Int Tinnitus J*, 2017, 21:144.

42 许轶,王铭歆,任飞,等. 声治疗结合心理咨询治疗慢性耳鸣[J]. *听力学及言语疾病杂志*, 2015, 23:358.

43 Li J, Jin J, Xi S, et al. Clinical efficacy of cognitive behavioral therapy for chronic subjective tinnitus[J]. *Am J Otolaryngol*, 2019, 40:253.

44 江洋,华清泉,杨琨,等. 声治疗与药物治疗慢性耳鸣疗效的 Meta 分析[J]. *听力学及言语疾病杂志*, 2017, 25:651.

(2019-07-31 收稿)

(本文编辑 雷培香)

者们对言语感知过程中视听整合的现象及其神经认知机制进行了系统的理论探讨。

1.1 言语感知中的视听整合现象 在视听言语整合的研究中,最经典的便是 McGurk 等^[5]的研究,其将剪辑后视听同步却不匹配的视频刺激呈献给受试者,要求受试者仅对听觉刺激进行复述;结果发现,当呈现的听觉刺激为/ba/、视觉刺激(指唇部发音动作)为/ga/时,98%的受试者报告他们听到的是/da/,研究者将该反应定义为“融合反应”;反之,当呈现的听觉刺激为/ga/、视觉刺激为/ba/时,大部分受试者报告他们听到的是/gabga/或/bagba/,研究者将该反应定义为“组合反应”;但在没有冲突的视觉信息影响的条件下,受试者对听觉信息的报告正确率高达 99%。上述现象被研究者称之为“麦格克效应”或“麦格克错觉”(McGurk illusion),其本质是当视听刺激同时出现但又彼此冲突时,人们对声音的识别出现感知偏差的现象。

1.2 言语感知中视听整合的神经机制

许多神经生理学研究为理解视听言语整合的神经结构和神经加工的时间进程提供了新的解释。其中,脑磁图(magnetoencephalogram, MEG)、功能性磁共振成像(functional magnetic resonance image, fMRI)和颅内电生理研究(intracranial electrophysiological study)都一致发现,正常人在无声唇读期间激活了听觉相关皮层的活动,这些区域主要包括颞上沟(superior temporal sulcus, STS)、颞上回(superior temporal gyrus, STG)、颞平面(planum temporale, PT)和颞横回(Heschl gyrus, HG)等^[6],表明说话者的唇部发音动作即可导致“听”的主观感知。相对而言,先天性聋患者在无声唇读时没有系统激活这些区域,但后天致聋的听障者在接受视觉刺激时,能激活听觉皮层。在此基础上, Schall 等^[7]进一步探讨了视听言语整合的神经机制,结果发现,视觉和听觉信息能够相互影响,两者进行整合的关键脑区在左侧后颞上沟(posterior superior temporal sulcus, pSTS)。Beauchamp 等^[8]的研究也发现,与单通道呈现刺激相比,在一致和不一致的视听言语感知期间, pSTS 和颞中回(middle temporal gyrus, MTG)显示出超加性的血液动力学反应,因此认为视觉语音线索和听觉语音线索被整合在这个大脑区域内。经颅磁刺激(transcranial magnetic stimulation, TMS)研究也发现,向 STS 施加 TMS 可以调制言语的视听整合,说明在言语感知的视听整合中 STS 起着关键作用。近年来, Komeilipoor 等^[9]使用事件相关电位(event-related potentials, ERP)技术来探讨皮层震荡(cortical oscillations)对

视听言语整合的影响,其源定位结果也显示, STS 可能是视听言语整合的关键脑区。尽管上述各研究在方法上存在很大差异,但一致发现,大脑左侧 pSTS/STS 参与了视听语音信号的整合^[10]。随后,该脑区再通过反馈投射调制特定单感官功能区的活动。

大脑半球优势支持上述观点。大脑在解剖学上的结构不对称是其功能不对称的基础,人脑结构和认知的主要特征是大脑半球的一侧优势效应,即大脑的某一功能向一侧半球集中的现象。右半球主要与情感感知、音乐欣赏等活动相关;而左半球主要负责语言处理、语法表达等,是语言的优势半球所在。其次,从脑区的功能特异性来讲, STS 是主要的“高阶”多感觉整合区之一,它不仅接收来自初级听觉和视觉皮层的刺激输入,而且该脑区还包含多感觉神经元^[11]。经由听觉关联皮层的听觉输入和通过视觉区的视觉刺激会整合在 STS,以感知整合后的言语信息。但近年来,随着研究逐渐深入,有研究者对该观点提出了质疑,他们认为,大脑中可能还存在一条更直接的皮层通路,即视觉皮层通过皮质-皮质间(cortico-cortical)的输入直接影响听觉皮层的活动,并不受高阶多感觉整合区的调节^[12],而这种较低水平的视觉皮层和听觉皮层间的功能连通可能是通过丘脑(thalamus)实现的。丘脑是皮层下的中继中心,用于各感觉通道的信号传输,可能有助于多感觉信号的加工。基于此,研究者们提出了视听言语感知的双路径模型,但在视听整合过程中,如何区分直接的皮质-皮质间通路和反馈通路的贡献,还有待进一步的研究和探讨。

2 情绪感知中的视听整合

对情绪的正确识别是社会交往所必须掌握的技能^[13]。情绪识别主要包括面孔情绪识别和语音情绪识别,通过整合面部表情、声乐韵律和其他线索的信号,大脑可以对他人的情绪作出判断。但事实上,大多数研究都只关注面部的情绪表达,而关于情感韵律(affective prosody)的研究较少,只有极少数的研究会关注这两者的整合。de Gelder 等^[14]以面部表情和声音语调作为实验材料来考察情绪感知的研究发现,视觉情绪检测过程和听觉情绪检测过程间存在很强的双向链接。

2.1 情绪感知中视听整合的现象 de Gelder 等^[14]使用经典的麦格克效应范式,利用情绪面孔照片和声音语调之间可能存在不同程度的冲突创建一种视听双通道的感知情境,在这种情境下,要求受试者报告其感知到的情绪体验;结果发现,在情绪识别过程中,受试者可以有效的整合面部表情和声音语

调的信息,对体验到的情绪作出独立判断。Piwek 等^[15]认为,现有情绪感知中视听整合的研究大多是通过呈现一个单独的人物来进行考察,但在现实情景中,人们通常面临更加复杂的社交场合;因此,在他们的研究中,实验材料由两个正在互动的人所产生的自然运动和语音对话组成,以此来考察在较简单的社会情景中所产生的视听整合效应能否延伸到更复杂和更生态的情景中去;在该实验中,听觉刺激指两人间的语音对话,视觉刺激指以光点(point-light display)形式呈现的两人互动的侧视图,视听双通道刺激则由视觉光点与语音对话结合而产生,视听刺激间可以是匹配的(光点刺激与语音对话表达同样的情绪),也可以是不匹配的(光点刺激与语音对话表达不同的情绪),刺激呈现后,要求受试者判断此次互动是高兴的还是愤怒的;结果发现,在视听结合的刺激条件下受试者对情绪的正确识别显著高于单通道条件。

2.2 情绪感知中视听整合的神经机制 受试者上述经典研究从行为层面描述了人脑在加工双通道情绪信息时存在的交互影响,近年来的认知神经科学研究则从神经电生理层面观察了双通道情绪信息整合加工的脑机制。Kreifelts 等^[16]通过神经成像技术探讨情绪信息整合的神经机制,结果发现,与单通道条件相比,在视听双通道条件下,双侧 pSTG 及右侧丘脑表现出更强的激活,且这些脑区的血氧水平依赖性反应强度(blood oxygenation level-dependent responses, BOLD)与面孔-语音情绪整合所带来的行为增益呈线性相关,说明这些脑区可能与情绪信息的整合加工有关。同样的, Watson 等^[17]使用功能磁共振适应技术(functional magnetic resonance adaptation, fMR-A),以动态的视听情绪刺激为实验材料来探讨情绪信息整合的神经机制,结果发现,在大脑右侧 pSTS 中视听情绪信息会产生彼此适应的现象,说明该脑区中可能存在同时从两种通道接收刺激输入的多感觉神经元。Thompson 等^[18]的研究结果也支持该观点,即面孔-声音情绪信息的整合加工发生在大脑右侧 pSTS。上述研究表明,双通道情绪信息的识别也存在半球优势;但与言语感知不同的是,感知情绪、识别情绪和产生情绪反应是大脑右半球的功能所在。此外,情绪信息的整合研究进一步证实了 STS 在视听整合中的重要作用。

随着研究不断细化,人类认知与脑科学家们提出了情绪的效价假说(valence hypothesis),即大脑的激活区域与情绪效价直接相关^[19]。Pourtois 等^[20]采用正电子发射断层扫描技术(positron emis-

sion tomography, PET)研究发现,积极情绪和消极情绪在大脑中具有单侧化效应;高兴的视听刺激所激活的脑区主要位于大脑左半球前部,而恐惧的视听刺激更多激活大脑右半球脑区。在此研究基础上, Park 等^[21]将中性刺激纳入实验条件,进一步将情绪刺激(如高兴和恐惧)与中性刺激进行比较,结果发现,恐惧信息整合加工的特异性脑区主要位于后扣带回(posterior cingulate)、梭状回(fusiform gyrus)和小脑(cerebellum),而高兴信息的整合加工涉及广泛的神经网络,主要包括 MTG、海马旁回(parahippocampal gyrus)、海马(hippocampus)、屏状核(claustrum)、顶下小叶(inferior parietal lobule, IPL)、楔叶(cuneus)、额中回(middle frontal gyrus, MFG)、额下回(inferior frontal gyrus, IFG)和前扣带回(anterior cingulate)。这些情绪加工的特异性脑区表明,各类情绪信息的整合既存在共同的神经机制,又有独特的加工网络。对于该研究结果,生物进化的观点认为,积极情绪和消极情绪在个体进化过程中具有不同的适应意义,进化的适应性使积极情绪和消极情绪具有不同的神经机制。恐惧、愤怒等消极情绪反映个体的生物性需求(如安全和食物等)受到威胁的状态,其解决的是生存问题,此时大脑要对外界输入的威胁信息进行整合加工,更多的是在右脑中进行;而高兴、兴奋等积极情绪反映人类在进化中形成的高级需要(如审美和自我实现等)得到满足时的状态,解决的是个体成长和发展问题,其更多的与语言和交际行为相关,而这正是大脑左半球的任务。

3 物体感知中的视听整合

人类对物体的感知是一项非常复杂的涉及多感官参与的心理过程。例如,在购买产品时,人们不仅会通过眼睛去看、用手去摸,有时还会敲击它来听它所产生的声音,通过这种“多通道”的方法来获得关于物体的可靠信息。然而,关于物体信息的跨通道整合研究却屈指可数。

3.1 物体感知中的视听整合现象 尽管很少有研究集中于物体感知中的视听整合现象,但是一些研究已经解决了与物体属性相关信息的跨通道整合问题。在此研究基础上, Fujisaki 等^[22]采用经典的麦格克效应范式,即把一种物体的视觉外观与另一种物体的声音相结合,要求受试者报告其感知到的物体类别,以此对物体感知中的视听整合现象进行研究。该研究通过计算机生成手拿棍子敲打物体(如玻璃、金属等)的场景视频作为视觉刺激,将隔音室内用木槌敲击真实物体(如陶瓷、塑料等)而产生的声音作为听觉刺激,在此基础上,将视觉刺激和听觉

刺激进行合并创建出新的场景视频作为视听刺激。要求受试者对视听刺激条件下所感知到的物体类别的可能性进行评分。结果发现,当“玻璃”的视觉外观和“玻璃”的听觉声音组合时受试者感知到的物体是“玻璃”;但当“玻璃”的视觉外观和“彩椒”的听觉声音组合时,受试者倾向于将感知到的物体认为是“塑料”;同样,当“木头”的视觉外观和听觉声音相匹配时,受试者感知到的物体是“木头”;但当“木头”的听觉声音与“玻璃”的视觉外观组合时,受试者倾向于感知到的物体是“塑料”。在其他视听刺激条件下,也观察到与之类似的实验结果。这表明,在物体类别感知过程中出现了视听整合现象,视觉图像和听觉声音之间存在很强的交互作用。

3.2 物体感知中视听整合的神经机制 与行为研究相比,只有少数神经生理学研究采用常见物体作为实验刺激^[6,23]。Kourtzi 等^[24]的研究发现,侧枕叶(lateral occipital complex, LOC)中的神经元在看到物体时会产生强烈活动,说明视觉物体信息主要在 LOC 中加工。Beauchamp 等^[23]通过视觉呈现物体的 fMRI 研究发现,视觉物体刺激主要激活的脑区集中在枕叶(occipital)、腹侧颞叶(ventral temporal)和后外侧颞叶(posterior lateral temporal cortex)。但不同类别的物体刺激激活颞叶皮层的不同区域,腹侧颞叶皮质主要对物体的形状、颜色和质感进行反应,而外侧颞叶皮质对物体运动更敏感^[8,25]。听觉通道中物体信息加工的研究则主要基于动物的解剖学数据,非人灵长类动物的解剖学研究、单细胞记录研究以及人类的 fMRI 研究表明,被 belt 和 parabelt 区包围的 core 区专门用于加工听觉刺激,而该区主要位于颞横回,但沿着颞平面向前和向后延伸,覆盖了颞上回的大部分区域并扩展到了下额叶皮层(inferior frontal cortex)^[26,27]。在单通道物体感知研究基础上,研究者们对物体视听信息整合的神经机制进行了探讨,较为一致的结论是外侧颞叶皮质(尤其是颞上沟)、腹侧颞叶皮质和额叶皮质可能参与物体视听信息的整合^[6,8]。Beauchamp 等^[8]的研究也发现,与单通道感知相比,当要求受试者整合物体的视听信息时,pSTS 或 MTG、背外侧前额叶皮层(dorsolateral prefrontal cortex, DLPFC)和腹侧颞叶皮质的激活程度更高。此外,ERP 研究中的溯源分析和最近的 TMS 研究均表明,腹侧视觉流,特别是侧枕叶,可能参与了物体视听信息的整合加工^[28,29]。

从上述研究结果来看,物体视听信息的整合加工主要涉及视觉的腹侧枕-颞通路(ventral occipitotemporal pathway)(即“*What*”通路)。Cichy

等^[30]的研究表明,物体信息所诱发的神经活动最早(大约在 50~80 ms)出现在枕叶,然后沿着腹侧和背侧向前迅速、渐进地传递,最终抵达颞下脑区或后顶叶区域,这个加工过程也是信息不断整合的过程。其中,背侧通路的主要功能是空间位置和运动的识别;而腹侧通路的主要功能是物体识别,该观点与当前的研究结论相一致。此外,物体信息的视听整合与言语、情绪信息的整合相一致,也证明了颞上沟是重要的多感觉整合区之一;同时,物体信息的整合加工又独特地激活了枕叶皮层,尤其是侧枕叶,表明侧枕叶中可能存在只对物体信息敏感的多感觉神经元,是物体信息整合加工的特异性脑区。

4 社会认知中的视听整合

对人身份进行识别是社会认知的重要方面,仅基于说话者的面孔或声音便能提取到对方的个人信息(如性别、年龄和人格特质等)。但在现实场景中,人们在听他人说话时也会看到他的脸,因此,大脑同时要对面孔和声音进行加工。近年来的行为和神经生理学研究表明,在身份识别过程中,面孔-声音信息会整合加工且相互影响。

4.1 社会认知中的视听整合现象 Peynircioglu 等^[31]的研究利用声音音色和面孔信息间的不匹配证明身份识别中也存在视听整合现象。该实验选用面孔性别与声音性别一致或不一致的视觉刺激作为实验材料,要求受试者对视频中的音域进行判断——女高音、女低音、男高音或男低音。其中,女高音和女低音的判断被认为是“女性”反应,男高音和男低音的判断被认为是“男性”反应(即采用间接推理的方法判断声音属于男性还是女性)。结果发现,在面孔-声音相匹配的条件下,正确率为 100%,但在不匹配的条件下,正确率仅为 31%。这表明,在识别声音的性别身份时存在视听整合现象,换言之,当听觉和视觉信息导致识别冲突时,人们倾向于将“听到”的音色报告成与视觉信息所暗示的音色一致。

4.2 社会认知中视听整合的神经机制 在行为研究基础上,神经成像研究解决了面孔-声音加工的脑机制问题。关于面孔识别的 fMRI 研究发现,面孔加工是基于广泛分布的神经网络,但主要由梭状回面孔区(fusiform face area, FFA)、枕部面孔区(occipital face area)和颞上沟组成了面孔识别的核心体系^[32~34]。此外,研究还进一步发现,不同的面孔特征激活的大脑区域不同:恒定的面孔特征及面孔身份识别主要激活 FFA,而动态的面孔特征(如眼动和面部表情)主要激活颞上沟^[33]。相比之下,声音身份识别的研究较少,但研究结论较为一致:大脑右侧颞上沟/G 前部是声

音身份识别的相关脑区^[35,36]。在单通道面孔/声音识别的基础上,研究者们对于大脑如何整合面孔一声音信息以进行身份识别提出了两种截然不同的观点,传统观点认为,面孔和声音信息分别加工,在一个超通道节点(即 person identity nodes, PINs)上视听信息整合,身份识别才得以实现^[37,38]。Joassin 等^[37]的研究支持该观点, PINs 所对应的脑区可能是右侧海马(right hippocampus)、左侧角回(angular)和左下顶区(left inferior parietal regions)。尽管该观点在解释身份识别的研究结果方面具有一定的影响力,但最近的一些脑科学研究对其提出质疑,他们认为面孔与声音加工之间存在直接的联系,而并非各自独立加工^[33,36,39,40]。Blank 等^[39]采用磁共振扩散追踪成像技术(probabilistic tractography)证明,面孔敏感区 FFA 与语音敏感区 STS/G 直接相连。这种连接可能为单通道感觉区域间的信息交流提供直接途径,并能促进 FFA 的反应^[36,41]。Von Kriegstein 等^[42]的研究结果也证明,在身份识别过程中,右侧 FFA 与右侧颞上沟的连接显著增强,表明多通道身份识别不一定涉及超通道皮层,而是由于单通道视觉和听觉区域间直接共享信息^[37]。其中,神经震荡(neutral oscillatory)活动的同步,尤其是 gamma 频段(>30 Hz),可能是单通道脑区间直接连接的神经基础^[43]。

5 对听障者多模态感知研究的意义

综上所述,最初发现于言语感知中的视听整合现象也普遍存在于情绪感知、物体感知和社会认知等更复杂的感知加工中,几乎参与了人类所有的感知活动。而神经系统在加工外界刺激时,通常是以多模态信息整合的方式进行。

多模态感知是人类智能的重要能力。人类的感知能力是多元的,任何感知能力的缺失或退化都有可能造成能力异常。听障者因听觉障碍,不仅降低了其对单纯的语音物理刺激的感知,而且也丧失了由语音信息所承载的语义、情感、情绪、逻辑、态度等内部心理特征的感知,导致其在言语、情绪、认知、人格和个性发展等方面都表现出异于常人的特点;即使通过助听器等电子设备进行适当的听力补偿或重建,听障患者仍难以得到完整的听觉信息,需依赖于视觉线索的辅助作用。因此,从视觉、听觉等单模态感知延伸到视听双模态,甚至多模态信息的联合感知对听障者的感知加工具有重要意义。

视觉、听觉等信息间交互的多模态现象在日常生态环境中普遍存在,听障者应善于利用多模态信息间的互补性,将多模信息结合并消除冗余,以此来增强综合感知能力,并有效地补偿患者的听力损失。近年来,越来越多的研究者致力于言语感知中视听

模式的研究,结果显示,视听结合的语音识别效果要显著好于单纯视或单纯听条件下的识别效果,表明视听条件对听障患者语言信息的获取、加工、理解具有重要意义。基于此,研究者们提出了对听障者采用视听结合的训练方法,在残余听力的基础上,或人工听力补偿的条件下,充分发挥视觉等其他线索所提供大量信息的辅助作用,广泛应用“残余听力+助听设备+视觉信息+其他感觉通道”的多模态感知模式作为日常接收外界信息的手段,以此来提高其在日常生活环境下的综合交流能力。

6 参考文献

- 1 文小辉, 李国强, 刘强. 视听整合加工及其神经机制[J]. 心理科学进展, 2011, 19:976.
- 2 欧阳丹, 欧阳文娟. 视听整合的研究进展[J]. 教育研究, 2017, 4:176.
- 3 石涯, 王永华, 李文婧. 唇读对听障儿童语音识别的帮助作用[J]. 听力学及言语疾病杂志, 2016, 24:482.
- 4 罗霄骁, 康冠兰, 周晓林. McGurk 效应的影响因素与神经基础[J]. 心理科学进展, 2018, 26:1935.
- 5 McGurk H, MacDonald J. Hearing lips and seeing voices[J]. Nature, 1976, 263: 47.
- 6 Amedi A, von Kriegstein K, van Atteveldt NM, et al. Functional imaging of human crossmodal identification and object recognition[J]. Experimental Brain Research, 2005, 166: 559.
- 7 Schall S, von Kriegstein K. Functional connectivity between face—movement and speech—intelligibility areas during auditory—only speech perception[J]. PLoS One, 2014,9:1.
- 8 Beauchamp MS, Lee KE, Haxby JV, et al. Parallel visual motion processing streams for manipulable objects and human movements[J]. Neuron, 2002, 34:149.
- 9 Komeilipoor N, Cesari P, Daffertshofer A. Involvement of superior temporal areas in audiovisual and audiomotor speech integration[J]. Neuroscience, 2017,343:276.
- 10 Maguinness C, von Kriegstein K. Cross—modal processing of voices and faces in developmental prosopagnosia and developmental phonagnosia[J]. Visual Cognition, 2017,25: 644.
- 11 Shinozaki J, Hiroe N, Sato M, et al. Impact of language on functional connectivity for audiovisual speech integration[J]. Scientific Reports, 2016, 6:1.
- 12 Arnal LH, Morillon B, Kell CA, et al. Dual neural routing of visual facilitation in speech processing[J]. Journal of Neuroscience, 2009,29:13445.
- 13 王苹, 潘治辉, 张立洁, 等. 动态面孔和语音情绪信息的整合加工及神经生理机制[J]. 心理科学进展, 2015,23:1109.
- 14 de Gelder B, Vroomen J. The perception of emotions by ear and by eye[J]. Cognition and Emotion, 2000, 14:289.
- 15 Piwek L, Pollick F, Petrini K. Audiovisual integration of emotional signals from others' social interactions[J]. Frontiers in Psychology, 2015, 6:1.
- 16 Kreifelts B, Ethofer T, Grodd W, et al. Audiovisual integration of emotional signals in voice and face: an event—related

- fMRI study[J]. *NeuroImage*, 2007, 37:1445.
- 17 Watson R, Latinus M, Noguchi T, et al. Crossmodal adaptation in right posterior superior temporal sulcus during face—voice emotional integration[J]. *Journal of Neuroscience*, 2014, 34: 6813.
 - 18 Thompson JD, Elli G, Rezk M, et al. Hierarchical brain network for face and voice integration of emotion expression[J]. *Cerebral Cortex*, 2017, 11:1.
 - 19 何良玉, 何湛情. 情绪性信息诱发的脑偏侧化认知神经机制研究进展[J]. *心理科学进展*, 2016, 6:1231.
 - 20 Pourtois G, de Gelder B, Bol A, et al. Perception of facial expressions and voices and of their combination in the human brain[J]. *Cortex*, 2005, 41: 49.
 - 21 Park JY, Gu BM, Kang DH, et al. Integration of cross—modal emotional information in the human brain: an fMRI study[J]. *Cortex*, 2010, 46:161.
 - 22 Fujisaki W, Goda N, Motoyoshi I, et al. Audiovisual integration in the human perception of materials[J]. *Journal of Vision*, 2014, 14:1.
 - 23 Beauchamp MS, Lee KE, Argall BD, et al. Integration of auditory and visual information about objects in superior temporal sulcus[J]. *Neuron*, 2004, 41: 809.
 - 24 Kourtzi Z, Kanwisher NG. Cortical regions involved in perceiving object shape[J]. *Journal of Neuroscience*, 2000, 20: 3310.
 - 25 Martin A, Chao LL. Semantic memory and the brain: structure and processes[J]. *Current Opinion in Neurobiology*, 2001, 11:194.
 - 26 Belin P, Zatorre RJ, Lafaille P, et al. Voice—selective areas in human auditory cortex[J]. *Nature*, 2000, 403:309.
 - 27 Zatorre RJ, Bouffard M, Belin P. Sensitivity to auditory object features in human temporal neocortex[J]. *Journal of Neuroscience*, 2004, 24:3637.
 - 28 Molholm S, Ritter W, Javitt DC, et al. Multisensory visual—auditory object recognition in humans: a high—density electrical mapping study[J]. *Cerebral Cortex*, 2004, 14:452.
 - 29 Giovannelli F, Giganti F, Righi S, et al. Audio—visual integration effect in lateral occipital cortex during an object recognition task: an interference pilot study[J]. *Brain Stimulation*, 2016, 9:574.
 - 30 Cichy RM, Pantazis D, Oliva A. Similarity—based fusion of meg and fmri reveals spatio—temporal dynamics in human cortex during visual object recognition[J]. *Cerebral Cortex*, 2016, 26: 3563.
 - 31 Peynircioglu ZF, Brent W, Tatz JR, et al. McGurk effect in gender identification: vision trumps audition in voice judgments[J]. *The Journal of General Psychology*, 2017, 144: 59.
 - 32 Axelrod V, Yovel G. Successful decoding of famous faces in the fusiform face area[J]. *PLoS One*, 2015, 10:1.
 - 33 Bühlhoff I, Newell FN. Crossmodal priming of unfamiliar faces supports early interactions between voices and faces in person perception[J]. *Visual Cognition*, 2017, 25:611.
 - 34 Weibert K, Andrews TG. Activity in the right fusiform face area predicts the behavioural advantage for the perception of familiar faces[J]. *Neuropsychologia*, 2015, 75:588.
 - 35 Maguinness C, Roswandowitz C, von Kriegstein K. Understanding the mechanisms of familiar voice—identity recognition in the human brain[J]. *Neuropsychologia*, 2018, 116: 179.
 - 36 Maguinness C, von Kriegstein K. Cross—modal processing of voices and faces in developmental prosopagnosia and developmental phonagnosia[J]. *Visual Cognition*, 2017, 25:644.
 - 37 Joassin F, Maurage P, Campanella S. The neural network sustaining the crossmodal processing of human gender from faces and voices: An fMRI study[J]. *Neuroimage*, 2010, 54: 1654.
 - 38 Joassin F, Maurage P, Bruyer R, et al. When audition alters vision: an event—related potential study of the cross—modal interactions between faces and voices[J]. *Neuroscience Letter*, 2004, 369:132.
 - 39 Blank H, Anwender A, von Kriegstein K. Direct structural connections between voice—and face—recognition areas[J]. *The Journal of Neuroscience*, 2011, 31:12906.
 - 40 Hasan BAS, Valdes—Sosa M, Gross J, et al. “Hearing faces and seeing voices”: a modal coding of person identity in the human brain[J]. *Scientific Reports*, 2016, 6.
 - 41 Schall S, Kiebel SJ, Maess B, et al. Early auditory sensory processing of voices is facilitated by visual mechanisms[J]. *NeuroImage*, 2013, 77:237.
 - 42 von Kriegstein K, Dogan Ö, Grüter M, et al. Simulation of talking faces in the human brain improves auditory speech recognition[J]. *Proceedings of the National Academy of Sciences*, 2008, 105:6747.
 - 43 Senkowski D, Schneider TR, Foxe JJ, et al. Crossmodal binding through neural coherence: implications for multisensory processing[J]. *Trends in Neurosciences*, 2008, 31:401.

(2019—06—25 收稿)

(本文编辑 周涛)