

在鸡尾酒会场景下利用去掩蔽知觉线索 提高言语识别的脑网络机制*

李 量¹ 郑英君² 吴 超³ 黎绢花² 张畅芯⁴ 陆灵犀¹

(¹北京大学心理与认知科学学院, 北京 100080) (²广州医科大学附属脑科医院
(广州市惠爱医院), 广州 510370) (³北京师范大学心理学部, 北京 100875)
(⁴华东师范大学教育学部教育康复学系, 上海 200241)

摘 要 在有多人说话的嘈杂环境下, 听者如何利用知觉线索来强化对目标言语的识别? 为解答这一鸡尾酒会问题, 研究者用脑成像的方法来考察相应的脑网络机制。研究表明, 听者对与目标言语某个特征相关联的去掩蔽线索的利用, 不但能促进听觉皮层对目标言语信号的短潜伏期反应, 而且强化了线索特异性及非特异性的注意、言语表达、抑制功能和言语运动这四类脑区的活动及其功能连接, 从而构成了信息掩蔽环境下知觉线索促进目标言语客体完好化的脑网络基础。

关键词 鸡尾酒会问题; 信息掩蔽; 去掩蔽; 言语识别; 主观空间分离; 听觉言语启动; 视觉言语启动; 脑网络; 神经成像。

分类号 B842; B845

复杂听觉场景下的“鸡尾酒会效应”是大脑的基本功能的一种反映。人和动物所面对的感觉刺激场景往往是复杂的, 大脑的基本功能之一就是在复杂刺激场景下选择性地加工有特定意义的目标信号。Cherry (1953)所提出的鸡尾酒会问题生动地反映了这一基本功能: 在嘈杂的有多人说话的“鸡尾酒会”环境下, 各种声波在外耳道中叠加混合并形成相互之间复杂的掩蔽作用(Du, Kong, Wang, & Li, 2011; Schneider, Li, & Daneman, 2007), 但听者为什么还能选择性地觉察、跟随和识别所关注的目标语句? 随着研究的深入, 人们已经意识到鸡尾酒会问题的基本内涵并不仅仅局限于对言语的识别, 而是反映了大脑从复杂场景中提取和加工目标信息的一个基本功能。迄今为止, 鸡尾酒会问题已经成为了心理学、神经科学和信息科学领域(如人工智能)中的一个基本性的科学问题; 但这个科学问题还没有得到最终的破

解。本综述旨在结合心理声学 and 脑功能网络两个视角来探讨鸡尾酒会场景下, 个体利用去掩蔽知觉线索提高目标言语识别的脑网络机制。

1 知觉线索在“鸡尾酒会”场景中的去掩蔽作用: 现状和挑战

为探索鸡尾酒会效应的心理机制和脑机制, 多年以来, 国内外的一些研究团队系统地考察了多种知觉线索的去掩蔽作用, 并证实了 Cherry (1953)的一个理论猜想: 听者可以利用某些与目标言语特征有关联的知觉线索来促进对被掩蔽的目标言语的识别(又称对目标言语的去掩蔽, unmasking)。例如, 所研究过的线索包括: 含语音精细结构(fine structure)声学信息的嗓音启动线索(Huang, Xu, Wu, & Li, 2010; Yang et al., 2007)、含语音时间包络信息的唇读启动线索(Wu, Cao, Wu, & Li, 2013; Wu et al., 2013; Wu, Zheng, Li, Wu et al., 2017; Wu, Zheng, Li, Zhang et al., 2017)、含语句语义信息的内容启动线索(Wu, Cao et al., 2012; Wu Li et al., 2012; Wu, Zheng, Li, Wu et al. 2017; Yang et al., 2007)、以及反映不同说话人之间的知觉空间位置差异信息的主观空间分离线索(Huang,

收稿日期: 2017-04-03

* 国家重大研发计划匹配(北京市科技计划)(课题编号: Z161100002616017)项目支持。

通讯作者: 李量, E-mail: liangli@pku.edu.cn

Huang, Chen, Wu, & Li, 2009; Li, Daneman, Qi, & Schneider, 2004; Li, Kong, Wu, & Li, 2013; Wu et al., 2005)。每一种知觉线索都对应于目标言语客体的某一个特征。而关键的问题是,听者为何能通过目标言语客体某个别特征的注意而提升了对目标言语的懂度(intelligibility)?更进一步,不同线索的最终作用效果都是提高了对目标言语的注意,进而产生了对目标言语的去掩蔽作用,但近期的脑成像研究结果明确地表明在不同线索的加工机制之间有诸多的不同(例如, Wu, Zheng, Li, Wu et al., 2017; Wu, Zheng, Li, Zhang et al., 2017; Zheng et al., 2016),这提示不同的去掩蔽线索之间可能既存在共有的规律,也同时具有其特异性的去掩蔽机制(线索特异性机制)。

近年来,有关鸡尾酒会效应的脑机制正在成为新的研究热点,但大多数研究主要集中在探讨掩蔽(言语竞争)场景下目标言语的中枢表达特性(例如, Ding & Simon, 2012; Mesgarani & Chang, 2012; Moon et al., 2014; Power, Foxe, Forde, Reilly, & Lalor, 2012; Scott & McGettigan, 2013),而针对知觉线索去掩蔽的脑机制的研究还很少(Wu, Zheng, Li, Wu et al., 2017; Wu, Zheng, Li, Zhang et al., 2017; Zhang, Lu, Wu, & Li, 2014; Zheng et al., 2016);同时,考察几种去掩蔽线索交互作用脑机制的研究也刚刚起步(Du, He et al., 2011)。

近期的脑成像研究工作在不同的被试中分别探索了几种不同线索去掩蔽作用的特异性脑机制(唇读启动: Wu, Zheng, Li, Zhang et al., 2017; 听觉言语启动: Wu, Zheng, Li, Wu et al., 2017; 主观空间分离: Zheng et al., 2016)。一个关键的科学问题是,在各种线索的去掩蔽作用的脑机制之间是否存在一种跨线索的共有性的基本机制模式?我们认为,只有揭示各种线索去掩蔽过程机制的共同性规律,才能最终真正地揭示鸡尾酒会效应的基本机制,也是最终破解鸡尾酒会问题的关键。到目前为止,尽管“鸡尾酒会效应”可以建立在多种去掩蔽线索作用基础上这一观点已成为共识,但不同线索的去掩蔽机制存在怎样的共性尚不清楚。

2 鸡尾酒会问题研究中无法摆脱的听觉捆绑问题

任何声波(包括言语刺激)都是在时间维度上

展开的动态信号。对言语的感知是一种同时在音素、音节、短语、语句等时间层次上的整合过程。由于信号的整合是听知觉的关键性机制,因此对言语去掩蔽过程基本机制的研究必然会涉及到另外一个重大的科学问题:捆绑问题。

在感知觉的早期加工阶段,一个知觉客体的各种特征由不同的脑区、环路和神经元群体来特异性地表达。但观察者为什么感知到的并不是一堆离散的特征,而是各个特征被“捆绑”成完整而连贯的知觉客体?这个捆绑问题(binding problem)也是心理学和神经科学领域中的一个重大科学问题。Treisman和Gelade(1980)所提出的特征整合理论(the feature integration theory, FIT)为这个问题作了经典性的阐述。大量的研究证实了知觉特征捆绑类型的多样性和多层次性。例如,知觉捆绑不仅可以扩展到不同的感觉通道之间,而且可以扩展到记忆的保持、语义的表达以及运动加工等方面(Burwick, 2014; Feldman, 2013; Spence, 2011; Velik, 2012; von der Malsburg, 1999)。因此,大脑的一个基本机能特性是:在一个知觉客体的各个特征之间存在着某种内在的捆绑(整合)机制。与鸡尾酒会效应密切关联的听觉捆绑有以下两种:

2.1 包络的神经表达和精细结构的神经表达之间的捆绑

听觉外周器官耳蜗的功能之一就是将一个复杂声音在频率维度上分解为一系列中心频率不同的窄带信号。听觉中枢系统因此具有了神经元按频率分布的构筑模式(tonotopic organization)。在有频谱重叠的多声源的“鸡尾酒会”环境下,来自同一声源的不同频率的窄带信号之间可以形成选择性的知觉捆绑进而使听者能感受到一个个音色完整的听觉客体。

然而,实际情况还更加复杂。声音信号在频谱维度上被分解后,分解过程并未结束。每个频道的窄带信号都由希尔伯特变换(Hilbert transform)(Hilbert, 1912)的形式进一步分解成在时间波动节奏上较慢的包络(envelope)成份和波动较快的时间精细结构(temporal fine structure, TFS)成份,而每个频道的输出信号是这两个成份的乘积(综述见 Moore, 2008)。对于语音来说,包络成份包括了时间调幅信号成份,而时间精细结构成份包括了谐波结构(harmonic structures)和调频信号(frequency modulation)以及周期性出现的类噪音的辅音

(periodic occurrences of noise-like consonants)等声学信号结构(Huang et al., 2011)。Smith, Delgutte 和 Oxenham (2002)认为, 言语的这两种成份有不同的作用: 精细结构主要对语音的音调识别和声源定向起重要作用, 而包络主要对言语的内容识别起重要作用, 进而它们分别形成了听觉皮层的空间(when)环路与内容(what)环路的声学基础(另见 Zeng et al., 2004)。在频谱和时间都有重叠的多声源的“鸡尾酒会”环境下, 来自同一声源的包络信号和精细结构信号之间可以特异地在知觉层次上捆绑在一起而形成相应于该声源的完整而连贯的有语义、声调和空间位置的听觉客体, 这种捆绑对鸡尾酒会效应的形成有重要的意义。

2.2 言语知觉客体的不同特征(嗓音、语音节奏、语义内容、以及空间等)之间的知觉捆绑

因为一个言语知觉客体具有多种特征, 其表达过程也是对多维度信息的同时性传送。例如, 一个人在说话时, 其语义内容、嗓音(身份)、空间位置以及情绪色彩等信息可以通过时变性(time varying)语音序列的动态模式来同时表达。有意思的是, 在有多个言语源的“鸡尾酒会”场景下, 只有隶属于同一说话人的各种言语特征才可以被特异地捆绑而成为一个特定的言语知觉客体。

我们的一个基本假设是, 当一个目标言语客体中的某一个特征(如语义、嗓音、时间韵律或空间)的中枢表达被注意强化时, 该特征的中枢表达与该目标言语客体的其它特征的中枢表达之间的知觉捆绑也会被强化, 以及目标言语客体的音素、词素、音节等成份的时间整合(捆绑)的加强, 从而造成该目标言语知觉客体的整体显现度的提高以及对掩蔽刺激信号的抑制。这对鸡尾酒会效应的形成也有重要的意义。

2.3 来自反映捆绑过程的特征捕捉理论的启发

由此看来, 捆绑问题与鸡尾酒会问题之间一定有密切的关联, 对其中一个问题的破解会带动对另一个问题的破解。这里介绍一个很有启发性的、涉及到这两个科学问题的特征捕捉理论: 一个有多人说话以及反射声(混响)的嘈杂的室内“鸡尾酒会”环境是一种对言语识别的掩蔽效果最强的听觉环境。在这种环境中, 一个言语信号会受到同时来自各个掩蔽声源的直达声波和其反射声波的强烈掩蔽作用。大脑为应对这一复杂的刺激场景, 将某个声源的直达声信号与该声源的反射

波信号之间形成知觉整合(捆绑), 即可以将某一声源的直达声波和与其有高相关的反射波群一起知觉为来自该声源的空间位置处的一个单独的融合声像。这就是所谓的听觉优先效应(precedence effect, Huang et al., 2011; 综述见 Li & Yue, 2002)。反映这种知觉融合倾向性的行为指标是回声阈限(echo threshold), 即将反射声听成一个未被融合的分离声像时该反射声与直达声之间的最小时间延迟(Li & Yue, 2002)。只有当延迟小于回声阈限时, 被试才能听到一个融合的声像。有意思的是, 语音的回声阈限要远大于非语音声音的回声阈限(Huang et al., 2011), 这表明人类大脑对语音信号有特殊的敏感性和选择性。

一系列的研究工作证实(Huang et al., 2008, 2009, 2011; Li et al., 2004, 2013; Wu et al., 2005), 在模拟的嘈杂而有混响的环境下, 目标信号的反射波与该目标信号直达波之间的知觉融合以及掩蔽信号的反射波与掩蔽信号直达波之间的知觉融合不但可以使听者感知不到来自不同方向的反射声(听觉客体的数量减少), 而且形成了目标信号和掩蔽信号之间的主观空间分离, 促进了对目标言语信号的选择性注意, 在知觉层次上抑制了掩蔽刺激的干扰作用, 进而促进了对目标言语的识别(详细介绍见 4.1)。

为解释听觉优先效应的机制, 研究者们提出了一个特征捕捉(attribute capture)理论(Li, Qi, He, Alain, & Schneider, 2005): 在混响环境下, 由于发自某声源的直达声波与落后的反射声波之间有高度的时间动态相关性, 反射波信号的知觉表达被其直达声波信号的知觉表达所“捕捉”, 形成了在原初的声源处的一个直达声和反射声之间融合的声像, 而同声源的直达声与反射声的融合也同时产生了混响条件下不同声源之间的知觉分离。这一理论的核心是, 有时间动态相关性的不同神经表达之间要进行某种形式的“捕捉”, 进而相互“捆绑”为更高级的整合性的表达。因此, 知觉捕捉是知觉捆绑的一种形式, 深入研究特征捕捉过程可能是破解捆绑问题以及鸡尾酒会问题的一个突破口。

由此看来, 鸡尾酒会问题也可以从听知觉捕捉和特征捆绑的角度来表述: 在一个多嗓音而且有混响的“鸡尾酒会”环境下, 某一被掩蔽的说话人的言语声波中所包含的时变性的频谱特征、精

细结构特征、包络特征、音节序列特征等成份以及空间特征,如何通过知觉捕捉过程而被捆绑成完整的言语听觉客体进而被听者所识别?

3 去掩蔽知觉线索提高言语识别的脑网络机制

3.1 主观空间分离去掩蔽线索强化了目标言语信号在听觉皮层中的表达

目标声音与掩蔽声音之间(实际的)客观空间分离形成的去掩蔽效应,有以下三种不同的机制:1)基于头影作用的提高信噪比的声学效应、2)基于双耳加工作用的强化对目标信号的神经反应的神经生理学效应、3)基于提高对目标信号选择性空间注意的心理学(知觉)效应。第三种效应可以是基于听觉优先效应的主观空间分离所产生的知觉层次上的去掩蔽效应(Li et al., 2004)。具体来说,在我们以往研究主观空间分离去掩蔽的实验中,当空间对称的左、右两个扬声器都发出目标声音和掩蔽声音时,对目标声音来讲,右扬声器提前 3 ms 播放(右、左扬声器分别模拟目标声音的直达和反射声源),而对掩蔽声音来讲,左扬声器提前 3 ms 播放(左、右扬声器分别代表掩蔽声音的直达和反射声源)。由于听觉优先效应的特征捕捉作用,两个扬声器之间的目标声音和两个扬声器之间的掩蔽声音分别形成知觉融合(即落后的目标声和掩蔽声分别被领先的目标声和掩蔽声所知觉捕捉),其融合后的目标和掩蔽声像分别在右、左扬声器处。这样,目标声音和掩蔽声音被知觉为来自不同的空间位置从而具有不同的主观空间特征,即出现主观空间分离(perceived spatial separation)。目标客体的所有特征的主观空间属性与掩蔽客体的所有特征的主观空间属性之间的不同有效地提高了听者在掩蔽条件下的对目标客体某些特征的选择性注意,进而促进了目标客体的知觉完好化(例如减少了目标言语的音节与掩蔽言语的音节之间的混淆)。目标客体的知觉完好化也同时加大了对掩蔽语句的忽视,进而减少了掩蔽语句对目标语句的信息掩蔽作用(Li et al., 2004; Wu et al., 2005; Rakerd, Aaronson, & Hartmann, 2006; Freyman, Balakrishnan, & Helfer, 2008; Huang et al., 2009; Huang, Wu, & Li, 2009)。

需要一提的是,由于目标和掩蔽声音都被每个扬声器所播放,并没有客观上的空间分离。这

样,建立在听觉优先效应基础上的目标语句与掩蔽语句之间的主观空间分离并不改变目标声波与掩蔽声波之间的信噪比以及各自的知觉密度。因此,主观空间分离的作用是消除了头影效应和双耳加工效应之外的一种心理/知觉的效应。因为目标言语客体的所有特征在主观空间分离条件下与掩蔽刺激客体的所有特征都形成了知觉分离,这种分离所产生的去掩蔽效果最显著。

近期张畅芯等人的头皮脑电研究发现,在言语信息掩蔽的条件下,从被动听觉条件向主动听觉条件的转换可以提高对目标音节的事件相关电位(Event-Related Potentials, ERP)反应中的 N1/P2 的成份(Zhang et al., 2014, 2016)。而在主动听觉的条件下,从目标音节与言语掩蔽刺激之间的主观空间重叠向主观空间分离的转换可以进一步强化该 N1/P2 成份(Zhang et al., 2014)。这些头皮脑电记录的结果与杜忆等人的脑磁记录研究的结果相符(Du, He et al., 2011)。这表明主观空间分离线索强化了目标言语信号在早期皮层加工中的表达,即在初级听觉皮层中的表达。

近期郑英君等人在国际上首次报告了使用功能核磁共振(functional Magnetic Resonance Imaging, fMRI)的研究方法所得到主观空间分离去掩蔽的神经机制的研究结果(Zheng et al., 2016)。具体来说,空间分离去掩蔽线索主要强化了包括顶叶的“where”皮层通路(背侧通路)的顶上小叶和楔前叶,以及额下回、额中回、前扣带回和尾核等脑区。这表明,主观空间分离去掩蔽过程涉及到了听觉系统以外的额叶、顶叶、扣带回和基底神经节等多个脑网络环路的协同活动。其中,在主观分离的条件下,顶上小叶的激活程度与对目标的识别成绩(识别的阈限值)之间有显著的相关(左侧顶上小叶: $r = -0.550, p = 0.027$; 右侧顶上小叶: $r = -0.538, p = 0.021$)。已有研究报道,顶上小叶参与了在“鸡尾酒会”环境下对某一说话人的注意(Hill & Miller, 2010)以及自觉性的注意的转换。因此,在对目标言语的信息掩蔽被主观空间分离所削弱的条件下,顶上小叶参与了对目标言语信号的空间注意的维持,直接参与了对目标言语的识别过程。

此外,由主观空间分离所产生的与抑制和注意选择加工有关的左侧前扣带回激活程度的提高与正常被试对目标语句识别行为成绩的提高之间

有显著的正相关($r = 0.568, p = 0.009$)。已有研究表明,前扣带回有对干扰刺激信号进行抑制的功能(Fornito, Yoon, Zalesky, Bullmore, & Carter, 2011; Schulz, Bédard, Czarnecki, & Fan, 2011; Shenhav, Botvinick, & Cohen, 2013)。因此,主观空间分离所产生的目标言语识别成绩的提高在一定程度上是对掩蔽刺激信号的抑制的结果。此外,顶上小叶、额下回、额中回、以及尾核也都有抑制干扰刺激信号的功能。

主观空间分离还强化了两侧的顶上小叶与右侧额下回(三角部),右侧尾核以及左侧丘脑之间的功能连接。已有研究表明,右侧额下回不但参与了对言语信号的觉察(Vouloumanos, Kiehl, Werker, & Liddle, 2001),也具有言语产生的功能,如语词决策(Carreiras, Mechelli, Estévez, & Price, 2007)以及汉语语义声调的形成(Liu et al., 2006)。因为在言语识别中,言语的运动系统活动与加工负载有关,特别是与“鸡尾酒会”环境下的言语知觉有关(综述见 Wu, Chen, Wu, & Li, 2014),因此顶上小叶与额下回的功能连接的加强表明了“鸡尾酒会”环境下的言语识别过程和言语运动系统的激活之间可能存在密切关系。

大脑的默认网络(default mode network, DMN)在不同的认知加工负载下呈现出不同的激活程度(Andrews-Hanna, 2012; Raichle et al., 2001; Whitfield-Gabrieli et al., 2009; Zhang & Li, 2012)。主观空间分离相对于主观空间重叠可以降低认知加工的负载,并增强了楔前叶以及腹侧前扣带回这两个默认网络成员的活动(Zheng et al., 2016)。

综上所述,主观空间分离会产生以下四种与言语识别成绩提高所相关的脑成像结果,而这些结果反映了去掩蔽线索的作用基于四类脑网络的协同活动: 1)对于言语客体特征的注意被强化并伴随着加工负载的下降,因此有注意环路以及默认网络的激活; 2)目标言语客体信号在听觉中枢的表达被强化; 3)掩蔽信号的中枢表达被抑制,因此涉及某些抑制机能的脑区(如尾状核)被激活; 4)符合言语认知的运动理论(the motor theory, 综述见 Wu et al., 2014),言语去掩蔽过程伴随着一些言语运动中枢(如额下回)的激活。

进一步,从捆绑加工的角度来看,主观空间分离所产生的去掩蔽效应是基于四种脑网络环路之间协调活动的功能捆绑过程,即注意环路、言

语表达环路、抑制功能环路和言语运动环路之间的捆绑。主观空间分离减弱了目标言语客体的特征与位于不同的知觉空间位置上的掩蔽言语客体特征之间的错误性捆绑,进而促进了(位于同一知觉空间位置上的)目标客体内特征之间的捆绑以及目标客体的知觉完好性,最终提高了个体对目标言语的可懂度。

3.2 听觉言语启动线索的作用及其脑机制

在“鸡尾酒会”环境下,当一位听者没有完全听清楚目标说话人所讲的内容时,该听者往往会请目标说话人“再说一遍”。由于该听者对目标说话人第一遍的表达所提供的目标语言的噪音、节奏、嘴唇运动、语义内容等方面信息的工作记忆还在保持,这些信息的工作记忆对目标说话人再说一遍的内容的识别就有显著的启动作用,即可以让听者更快地将选择性注意投放到被复述的目标语流上,促进听者对目标语流的注意追随。下面先介绍一下有关听觉言语启动线索的作用及其脑网络机制的研究工作。

在以往的研究听觉言语启动线索去掩蔽的实验中,一个目标说话人所朗诵的语句为语法正确但在主、谓、宾语之间无语义关联的无意义语句(nonsense sentences),例如在“那个笼具可以退休他的海洋”一句中,下划线的主、谓、宾词为句子的关键词(Freyman et al., 2008; Li et al., 2004; Wu et al., 2005; Yang et al., 2007)。因为关键词之间不存在有正常意义的语义关系,被试无法利用所听清的关键词来推断和猜测所未能被听清的关键词。在实验中,这些目标语句因与另外两个不同噪音和内容的掩蔽语句同时播放而受到能量掩蔽和信息掩蔽。如果在目标和掩蔽语句同时播放之前呈现一个由该目标说话人的噪音所朗读的目标句子的前面的一部分时,(正常年轻)被试可以根据对目标噪音或内容的工作记忆来显著地提高对(掩蔽条件下的)目标句子中(未在启动句中呈现的)最后一个关键词的识别(Gao et al., 2014; Freyman, Balakrishnan, & Helfer, 2004; Helfer & Freyman 2009; Huang et al., 2010; Yang et al., 2007)。

我们认为,听觉言语启动作用的心理学机制涉及到了一个复杂的链性加工过程,包括: 1)对启动噪音的精细结构信息所进行的感觉分析(Belin, Bestelmeyer, Latinus, & Watson, 2011); 2)形成基于统计学特性的对启动噪音知觉特征的表达; 3)形

成对启动噪音知觉特征的工作记忆、对目标噪音的预期以及启动句内容的工作记忆; 4) 目标噪音出现后, 对目标噪音的精细结构进行感觉分析并形成目标噪音的知觉特征; 5) 由于启动噪音的知觉特征与目标噪音的知觉特征高度相关(匹配)以及启动句内容与目标句的前部分一致, 对目标言语的知觉特征的选择性注意被强化, 进而造成目标言语客体的显现度(saliency)的提高; 6) 目标噪音特征与目标言语的其它特征(如包络特征等)之间的时间动态捆绑被加强; 7) 目标语句客体的知觉完好性提高, 进而加大了目标客体与掩蔽客体之间的知觉分离, 并提高了目标语句的可懂度。

近期吴超等人使用 fMRI 的方法在国际上首次报告了听觉言语启动线索去掩蔽的神经机制的研究结果(Wu, Zheng, Li, Wu et al., 2017)。相对于听觉非言语启动线索的刺激条件, 听觉言语启动线索的呈现显著地提高双侧颞上回、双侧颞中回、左侧颞中回后部、左侧的额下回、以及左侧的壳核的活动。进一步, 以左侧的颞上回/颞中回为种子点脑区, 听觉言语启动线索的呈现还强化了左侧颞上回/颞中回与左侧额下回、左侧岛叶、以及左侧尾核之间的功能连接。其中左侧颞上回与左侧额下回三角部之间功能连接的强度与在言语掩蔽下的目标言语的识别成绩之间有显著的正相关($r = 0.512, p = 0.048$)。

颞上回被认为是早期言语识别加工的皮层区, 具有言语声音识别的功能(Ahveninen et al., 2006; Boatman, 2004; Friederici, Rueschemeyer, Hahne, & Fiebach, 2003; Hickok & Poeppel, 2004; Rauschecker & Scott, 2009, 2015; Scott & Wise, 2003)。左侧的颞中回后部具有从长时记忆中提取词汇-句法信息的功能(Snijders et al., 2009), 而它与左侧额下回的功能连接涉及句子结构的加工以及言语的语法性表达的构建(Papoutsis, Stamatakis, Griffiths, Marslen-Wilson, & Tyler, 2011; Tyler, Wright, Randall, Marslen-Wilson, & Stamatakis, 2010; Tyler, Cheung, Devereux, & Clarke, 2013)。此外, 左侧的颞中回后部还涉及到对中文言语中句子层次上的韵律加工(Tong et al., 2005)。由于启动去掩蔽要基于对启动信息的工作记忆保持, 因此也应当有涉及到工作记忆的脑区被启动语句所激活。有研究表明, 左侧颞上回的喙侧部有工作记忆的功能(Arnott, Grady, Hevenor, Graham, &

Alain, 2005), 而颞中回的后部也有储存词汇表达的功能(Lau, Phillips & Poeppel, 2008)。这些记忆功能对启动线索去掩蔽有重要的意义。因此, 听觉言语启动线索对左侧颞上回/颞中回的激活也反映了对目标言语信号的皮层表达的强化。

听觉言语启动线索的引入还加强了左侧颞上回/颞中回与左侧尾核、左侧额下回、以及左侧岛叶的功能连接。这些功能连接的增强表明以下几种神经过程的增强: 因为尾状核涉及到言语抑制以及更加广泛性的抑制过程(Ali, Green, Kherif, Devlin, & Price, 2010; Ketteler, Kastrau, Vohn, & Huber, 2008; Li, Yan, Sinha, & Lee, 2008; Menon, Adleman, White, Glover, & Reiss, 2001), 而左侧额下回也具有在竞争性刺激源中选择信号的功能(Thompson-Schill, Bedny, & Goldberg, 2005), 因此听觉言语启动线索去掩蔽的过程包括了选择性注意神经过程以及对干扰信号进行抑制的神经过程的增强。此外, 左侧额下回有多种言语加工的功能, 包括通过与左侧颞中回后部的交互作用而实现句法统一、言语含义的选择、句子的再解释、以及言语的产生等(Herholz et al., 1996; Papathanassiou et al., 2000; Paulesu et al., 1997; Rodd, Johnsrude, & Davis, 2012; Schuhmann, Schiller, Geobel, & Sack, 2009; Snijders et al., 2009)。如前所述, 因为言语运动系统的功能对在“鸡尾酒会”听觉场景下的言语知觉有重要的作用(Wu et al., 2014), 因此左侧颞上回/颞中回后部与左侧额下回的功能连接通过强化对掩蔽信号的抑制过程、对目标信号的表达过程以及言语的运动过程来实现对目前信号的去掩蔽。

3.3 (视觉言语)唇读启动线索的作用及其脑机制

同样, 在以往研究(视觉言语, visual speech)唇读启动线索去掩蔽的实验中, 所用的言语刺激也是无意义语句。如果在同时播放目标语句和掩蔽语句之前播放一个朗诵该目标语句的无声的嘴唇运动的录像, 被试可以利用该嘴唇运动的工作记忆来促进在有信息掩蔽的“鸡尾酒会”环境下对该目标语句的识别(Wu, Cao et al., 2013; Wu, Li et al., 2013; Wu, Zheng, Li, Zhang et al., 2017)。已有研究证实, 说话时嘴唇等脸部部位的运动含有丰富的时间节奏的信息, 而这些信息对言语的声音信息来说既有冗余又有互补的关系(Summerfield, 1979)。唇读线索中的嘴唇形状变化的时间动态信

号与言语声学成份中的时间包络信息之间是高度相关的。因此,我们认为,在知觉层次上,目标语音包络信息被视觉唇读启动包络信息所“捕捉”。这种视-听的时间同步性(temporal synchronization)使得对目标言语中包络信息的加工得到强化。随后,对目标言语流的节奏信号的选择性注意强化促进了对目标言语其它特征的时间动态知觉捆绑,继而目标客体知觉完好性的提高进一步加深了目标客体与掩蔽客体之间的知觉分离。

近期吴超等人在国际上首次报告了使用 fMRI 的方法所得到视觉言语唇读启动线索去掩蔽的神经机制的研究结果(Wu, Zheng, Li, Zhang et al., 2017)。相对于视觉非言语启动条件,视觉言语启动刺激(对目标言语的唇读刺激)的呈现激活了两侧的颞下回后部以及两侧的梭状回。而只有左侧颞下回后部的激活程度与视觉言语唇读启动线索所产生的对目标言语的去掩蔽程度之间有显著的正相关($r = 0.611, p = 0.012$)。因此,只有左侧颞下回后部被认为是唇读去掩蔽的相关脑区。

从理论上分析,唇读线索去掩蔽建立在以下四种神经功能过程的基础上:(1)对唇读中的言语信号的加工,(2)对唇读动态信息的工作记忆的保持,(3)对唇读中的视觉言语信息与听觉言语信息之间的整合加工,(4)对目标言语的选择性注意以及对掩蔽刺激信号的抑制。已有充分的前人研究工作表明,左侧颞下回是唯一符合这四项神经功能的脑区:(1)唇读刺激可以有效地激活颞下回(Campbell et al., 2001; Ludman et al., 2000; Xu, Gannon, Emmorey, Smith, & Braun, 2009);(2)颞下回具有视觉工作记忆的功能(Ranganath, 2006; Ranganath, Cohen, Dam, & D'Esposito, 2004; Woloszyn & Sheinberg, 2009);(3)左侧颞下回后部参与了跨通道言语的语义加工(Giraud & Truy, 2002; Mummery et al., 1999; Vandenberghe, Price, Wise, Josephs, & Frackowiak, 1996; Wise et al., 1991);(4)颞下回也参与了对目标信号的选择性注意以及对干扰信号的抑制(Chelazzi, Duncan, Miller, & Desimone, 1998; Chelazzi, Miller, Duncan, & Desimone, 1993; Zhang et al., 2011)。值得指出的是,左侧的颞下回,而不是右侧的颞下回,可以被以下的刺激或任务所激活:短暂的言语刺激(Alain et al., 2005)、言语声音的辨别 discrimination of speech sounds (Ikeda et al.,

2010)、对语义不明确的言语的加工(Rodd et al., 2012)、听觉与视觉信号的整合(Romanski, 2012)、以及对言语信号的理解(Giraud & Truy, 2002)。因此,尽管右侧的颞下回与左侧的颞下回有一些相同的功能,但它不参与更加特异、复杂以及更高层的言语加工。

更进一步,视觉言语启动刺激的呈现还加强了左侧颞下回后部与以下脑区的功能连接:双侧的颞上回、双侧的内侧额上回、双侧小脑、左侧前中央皮质、左侧后中央皮质、额下回岛盖部、左侧岛叶,左侧上缘板和左侧辅助运动区,以及右侧的额中回。

如上所述,颞上回被认为是早期言语识别加工的皮层区(Ahveninen et al., 2006; Hickok & Poeppel, 2004; Scott & McGeehan, 2013)。脑成像研究(Ahveninen et al., 2006; Friederici et al., 2003)和功能损伤的研究(Boatman, 2004)都证明了颞上回在语音、语义和句法等几个层次上参与了语音识别。因此,左侧颞下回与左侧颞上回之间的功能连接反映了唇读启动去掩蔽的过程是基于视觉言语信号的加工与听觉言语信号的加工之间的功能整合。

额下回不仅参与了言语信号的识别(Vouloumanos et al., 2001)也参与了言语的产生过程(Carreiras et al., 2007; Liu et al., 2006)。对唇读启动线索的利用所形成的颞下回与额下回之间的功能连接的强化反映了在“鸡尾酒会”场景下的言语识别过程与言语产生过程的同时性强化,这对目标言语的去掩蔽有重要的意义。此外,内侧颞上回参与了目标指向性行为(Dosenbach et al., 2007; Rushworth, Walton, Kennerley, & Bannerman, 2004),辅助运动区在计划、准备、控制和实施等执行功能中起了重要的作用(Nachev, Kennard, & Husain, 2008; Price, 2012),而额中回涉及到对无关信号的抑制以及在竞争条件下的目标选择(Jeurissen, Sack, Roebroek, Russ, & Pascual-Leone, 2014; Lesh, Niendam, Minzenberg, & Carter, 2011; Sokol-Hessner, Hutcherson, Hare, & Rangel, 2012; Zheng et al., 2016)。此外,岛叶皮层参与了抑制过程(Menon et al., 2001)。因此,呈现唇读去掩蔽启动线索不仅可以强化目标言语信号的中枢表达,而且同时强化了对目标信号的注意、对掩蔽信号的抑制、视觉言语信号与听觉言语信号之

间的整合,以及早期视觉加工与运动执行功能之间的整合。

4 总结

鸡尾酒会问题和捆绑问题都是心理学和脑科学中的重大理论问题,不但反映了大脑基本的知觉加工策略,而且反映了在有大量感觉输入的环境压力下,大脑如何利用有限的加工资源来有效地形成知觉客体以及获得有意义的目标信号。因此,探讨“鸡尾酒会”场景下利用去掩蔽线索提高对目标言语识别的脑网络机制,其意义不局限于听觉和言语认知领域,而是有助于建立一个初步的在特征捕捉理论基础上的鸡尾酒会问题与捆绑问题统一化的理论、从而推动对大脑反映客观世界的基本规律的探索。

根据已有的脑成像的研究工作,我们认为,目标言语客体的不同去掩蔽线索特征之间的捆绑是形成鸡尾酒会效应的核心机制。从理论上分析,去掩蔽过程包括以下三个主要加工阶段:1)去掩蔽线索与目标言语客体中相应的特征之间因捕捉而出现整合,进而强化了对该特征的注意加工;2)由于该特征的知觉加工被强化,该特征与其它目标言语特征之间的知觉捆绑被强化,进而突出了掩蔽环境下的目标言语客体的知觉显现度;3)目标言语客体的显现度的提高使得对该客体的选择性注意被进一步强化,进而抑制了掩蔽客体的表达并提高了目标言语的可懂度。

本文综述了鸡尾酒会问题中利用线索去掩蔽的脑机制及其与捆绑问题的统一化的研究,提出了多种去掩蔽线索的“特异性”加工和“共有性”加工是探讨鸡尾酒会问题的重要突破口。已往的研究工作通常只在一项研究中考察一个线索的去掩蔽作用,所发现的脑机制往往只针对某个特定的去掩蔽线索,这就需要在未来的研究中,同时探讨“线索依赖性”和“线索独立性”的神经机制,即将不同知觉线索在去掩蔽过程中建立的“特异性-共有性”机制的探讨定位在(1)注意环路、(2)言语表达环路、(3)抑制功能环路和(4)言语运动环路之间的动态协同活动的基础上。

参考文献

- Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., Belliveau, J. W. (2006). Task-modulated “what” and “where” pathways in human auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America of the United States of America*, *103*, 14608–14613.
- Alain, C., Reinke, K., McDonald, K. L., Chau, W., Tam, F., Pacurar, A., & Graham, S. (2005). Left thalamo-cortical network implicated in successful speech separation and identification. *NeuroImage*, *26*, 592–599.
- Ali, N., Green, D. W., Kherif, F., Devlin, J. T., & Price, C. J. (2010). The role of the left head of caudate in suppressing irrelevant words. *Journal of Cognitive Neuroscience*, *22*, 2369–1386.
- Andrews-Hanna, J. R. (2012). The brain’s default network and its adaptive role in internal mentation. *The Neuroscientist*, *18*, 251–270.
- Arnott, S. R., Grady, C. L., Hevenor, S. J., Graham, S., & Alain, C. (2005). The functional organization of auditory working memory as revealed by fMRI. *Journal of Cognitive Neuroscience*, *17*, 819–831.
- Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*, 711–725.
- Boatman, D. (2004). Cortical bases of speech perception: evidence from functional lesion studies. *Cognition*, *92*, 47–65.
- Burwick, T. (2014). The binding problem. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 305–315.
- Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., David, A. S. (2001). Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cognitive Brain Research*, *12*, 233–243.
- Carreiras, M., Mechelli, A., Estévez, A., & Price, C. J. (2007). Brain activation for lexical decision and reading aloud: Two sides of the same coin? *Journal of Cognitive Neuroscience*, *19*, 433–444.
- Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*, 2918–2940.
- Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*, 345–347.
- Cherry, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *The Journal of the Acoustical Society of America*, *25*, 975–979.
- Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 11854–11859.
- Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., Petersen, S. E. (2007). Distinct brain networks for adaptive and stable

- task control in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 11073–11078.
- Du, Y., Kong, L. Z., Wang, Q., Wu, X. H., & Li, L. (2011). Auditory frequency-following response: A neurophysiological measure for studying the “cocktail-party problem”. *Neuroscience & Biobehavioral Reviews*, 35, 2046–2057.
- Du, Y., He, Y., Ross, B., Bardouille, T., Wu, X. H., Li, L., & Alain, C. (2011). Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. *Cerebral Cortex*, 21, 698–707.
- Feldman, J. (2013). The neural binding problem (s). *Cognitive Neurodynamics*, 7, 1–11.
- Fornito, A., Yoon, J., Zalesky, A., Bullmore, E. T., & Carter, C. S. (2011). General and specific functional connectivity disturbances in first-episode schizophrenia during cognitive control performance. *Biological Psychiatry*, 70, 64–72.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 115, 2246–2256.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2008). Spatial release from masking with noise-vocoded speech. *The Journal of the Acoustical Society of America*, 124, 1627–1637.
- Friederici, A. D., Rüschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex*, 13, 170–177.
- Gao, Y. Y., Cao, S. Y., Qu, T. S., Wu, X. H., Li, H. F., Zhang, J. S., & Li, L. (2014). Voice-associated static face image releases speech from informational masking. *PsyCh Journal*, 3, 113–120.
- Giraud, A. L., & Truy, E. (2002). The contribution of visual areas to speech comprehension: A PET study in cochlear implants patients and normal-hearing subjects. *Neuropsychologia*, 40, 1562–1569.
- Helfer, K. S., & Freyman, R. L. (2009). Lexical and indexical cues in masking by competing speech. *The Journal of the Acoustical Society of America*, 125, 447–456.
- Herholz, K., Thiel, A., Wienhard, K., Pietrzyk, U., Von Stockhausen, H. M., Karbe, H., Heiss, W. D. (1996). Individual functional anatomy of verb generation. *NeuroImage*, 3, 185–194.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.
- Hilbert, D. (1912). *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*. Leipzig, Berlin: B. G. Teubner.
- Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, 20, 538–590.
- Huang, Y., Huang, Q., Chen, X., Qu, T. S., Wu, X. H., & Li, L. (2008). Perceptual integration between target speech and target-speech reflection reduces masking for target-speech recognition in younger adults and older adults. *Hearing Research*, 244, 51–65.
- Huang, Y., Huang, Q., Chen, X., Wu, X. H., & Li, L. (2009). Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1618–1628.
- Huang, Y., Li, J. Y., Zou, X. F., Qu, T. S., Wu, X. H., Mao, L. H., Li, L. (2011). Perceptual fusion tendency of speech sounds. *Journal of Cognitive Neuroscience*, 23, 1003–1014.
- Huang, Y., Wu, X. H., & Li, L. (2009). Detection of the break in interaural correlation is affected by interaural delay, aging, and center frequency. *The Journal of the Acoustical Society of America*, 126, 300–309.
- Huang, Y., Xu, L. J., Wu, X. H., & Li, L. (2010). The effect of voice cuing on releasing speech from informational masking disappears in older adults. *Ear and Hearing*, 31, 579–583.
- Ikeda, Y., Yahata, N., Takahashi, H., Koeda, M., Asai, K., Okubo, Y., & Suzuki, H. (2010). Cerebral activation associated with speech sound discrimination during the diotic listening task: An fMRI study. *Neuroscience Research*, 67, 65–71.
- Jeurissen, D., Sack, A. T., Roebroek, A., Russ, B. E., & Pascual-Leone, A. (2014). TMS affects moral judgment, showing the role of DLPFC and TPJ in cognitive and emotional processing. *Frontiers in Neuroscience*, 8, 18.
- Ketteler, D., Kastrau, F., Vohn, R., & Huber, W. (2008). The subcortical role of language processing. High level linguistic features such as ambiguity-resolution and the human brain; an fMRI study. *NeuroImage*, 39, 2002–2009.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De) constructing the N400. *Nature Reviews Neuroscience*, 9, 920–933.
- Lesh, T. A., Niendam, T. A., Minzenberg, M. J., & Carter, C. S. (2011). Cognitive control deficits in schizophrenia: Mechanisms and meaning. *Neuropsychopharmacology*, 36, 316–338.
- Li, C. S. R., Yan, P. S., Sinha, R., & Lee, T. W. (2008). Subcortical processes of motor response inhibition during a stop signal task. *NeuroImage*, 41, 1352–1363.
- Li, H. H., Kong, L. Z., Wu, X. H., & Li, L. (2013). Primitive auditory memory is correlated with spatial unmasking that is based on direct-reflection integration. *PLoS One*, 8, e63106.
- Li, L., Daneman, M., Qi, J. G., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *Journal of Experimental Psychology:*

- Human Perception and Performance*, 30, 1077–1091.
- Li, L., Qi, J. G., He, Y., Alain, C., & Schneider, B. A. (2005). Attribute capture in the precedence effect for long-duration noise sounds. *Hearing Research*, 202, 235–247.
- Li, L., & Yue, Q. (2002). Auditory gating processes and binaural inhibition in the inferior colliculus. *Hearing Research*, 168, 98–109.
- Liu, L., Peng, D. L., Ding, G. S., Jin, Z., Zhang, L., Li, K., & Chen, C. S. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *NeuroImage*, 29, 515–523.
- Ludman, C. N., Lecturer, S., Summerfield, A. Q., Hall, D., Elliott, M., Foster, ... Morris, P. G. (2000). Lip-reading ability and patterns of cortical activation studied using fMRI. *British Journal of Audiology*, 34, 225–230.
- Menon, V., Adelman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping*, 12, 131–143.
- Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485, 233–236.
- Moon, I. J., Won, J. H., Park, M. H., Ives, D. T., Nie, K. B., Heinz, M. G., ... Rubinstein, J. T. (2014). Optimal combination of neural temporal envelope and fine structure cues to explain speech identification in background noise. *Journal of Neuroscience*, 34, 12145–12154.
- Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, 9, 399–406.
- Mummery, C. J., Patterson, K., Wise, R. J. S., Vandenberg, R., Price, C. J., & Hodges, J. R. (1999). Disrupted temporal lobe connections in semantic dementia. *Brain*, 122, 61–73.
- Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, 9, 856–869.
- Papathanassiou, D., Etard, O., Mellet, E., Zago, L., Mazoyer, B., & Tzourio-Mazoyer, N. (2000). A common language network for comprehension and production: A contribution to the definition of language epicenters with PET. *NeuroImage*, 11, 347–357.
- Papoutsis, M., Stamatakis, E. A., Griffiths, J., Marslen-Wilson, W. D., & Tyler, L. K. (2011). Is left fronto-temporal connectivity essential for syntax? Effective connectivity, tractography and performance in left-hemisphere damaged patients. *NeuroImage*, 58, 656–664.
- Paulesu, E., Goldacre, B., Scifo, P., Cappa, S. F., Gilardi, M. C., Castiglioni, I., ... Fazio, F. (1997). Functional heterogeneity of left inferior frontal cortex as revealed by fMRI. *Neuroreport*, 8, 2011–2016.
- Power, A. J., Foxe, J. J., Forde, E. J., Reilly, R. B., & Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *European Journal of Neuroscience*, 35, 1497–1503.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62, 816–847.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 676–682.
- Rakerd, B., Aaronson, N. L., & Hartmann, W. M. (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *The Journal of the Acoustical Society of America*, 119, 1597–1605.
- Ranganath, C. (2006). Working memory for visual objects: Complementary roles of inferior temporal, medial temporal, and prefrontal cortex. *Neuroscience*, 139, 277–289.
- Ranganath, C., Cohen, M. X., Dam, C., & D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *Journal of Neuroscience*, 24, 3917–3925.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12, 718–724.
- Rauschecker, J. P., & Scott, S. K. (2015). Pathways and streams in the auditory cortex: An update on how work in nonhuman primates has contributed to our understanding of human speech processing. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 287–298). Oxford: Elsevier.
- Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cerebral Cortex*, 22, 1761–1773.
- Romanski, L. M. (2012). Integration of faces and vocalizations in ventral prefrontal cortex: Implications for the evolution of audiovisual speech. *Proceedings of the National Academy of Sciences of the United States of America*, 109(Suppl. 1), 10717–10724.
- Rushworth, M. F. S., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences*, 8, 410–417.
- Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, 18, 559–572.
- Schuhmann, T., Schiller, N. O., Goebel, R., & Sack, A. T. (2009). The temporal characteristics of functional activation in Broca's area during overt picture naming.

- Cortex*, 45, 1111–1116.
- Schulz, K. P., Bédard, A. C. V., Czarnecki, R., & Fan, J. (2011). Preparatory activity and connectivity in dorsal anterior cingulate cortex for cognitive control. *NeuroImage*, 57, 242–250.
- Scott, S. K., & McGettigan, C. (2013). The neural processing of masked speech. *Hearing Research*, 303, 58–66.
- Scott, S. K., & Wise, R. J. (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication*, 41, 23–34.
- Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, 79, 217–240.
- Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416, 87–90.
- Snijders, T. M., Vosse, T., Kempen, G., van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, 19, 1493–1503.
- Sokol-Hessner, P., Hutcherson, C., Hare, T., & Rangel, A. (2012). Decision value computation in DLPFC and VMPFC adjusts to the available decision time. *European Journal of Neuroscience*, 35, 1065–1074.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73, 971–995.
- Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, 36, 314–331.
- Thompson-Schill, S. L., Bedny, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology*, 15, 219–224.
- Tong, Y. X., Gandour, J., Talavage, T., Wong, D., Dziedzic, M., Xu, Y. S., ... Lowe, M. (2005). Neural circuitry underlying sentence-level linguistic prosody. *NeuroImage*, 28, 417–428.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Tyler, L. K., Cheung, T. P. L., Devereux, B. J., & Clarke, A. (2013). Syntactic computations in the language network: Characterizing dynamic network properties using representational similarity analysis. *Frontiers in Psychology*, 4, 271.
- Tyler, L. K., Wright, P., Randall, B., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Reorganization of syntactic processing following left-hemisphere brain damage: Does right-hemisphere activity preserve function? *Brain*, 133, 3396–3408.
- Vandenbergh, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, 383, 254–256.
- Velik, R. (2012). From simple receptors to complex multimodal percepts: A first global picture on the mechanisms involved in perceptual binding. *Frontiers in Psychology*, 3, 259.
- von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron*, 24, 95–104.
- Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, 13, 994–1005.
- Whitfield-Gabrieli, S., Thermenos, H. W., Milanovic, S., Tsuang, M. T., Faraone, S. V., McCarley, R. W., ... Seidman, L. J. (2009). Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 1279–1284.
- Wise, R., Chollet, F., Hadar, U. R. I., Friston, K., Hoffner, E., & Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain*, 114, 1803–1817.
- Woloszyn, L., & Sheinberg, D. L. (2009). Neural dynamics in inferior temporal cortex during a visual working memory task. *Journal of Neuroscience*, 29, 5494–5507.
- Wu, C., Cao, S. Y., Wu, X. H., & Li, L. (2013). Temporally pre-presented lipreading cues release speech from informational masking. *The Journal of the Acoustical Society of America*, 133, EL281–EL285.
- Wu, C., Cao, S. Y., Zhou, F. C., Wang, C. Y., Wu, X. H., & Li, L. (2012a). Masking of speech in people with first-episode schizophrenia and people with chronic schizophrenia. *Schizophrenia Research*, 134, 33–41.
- Wu, C., Li, H. H., Tian, Q., Wu, X. H., Wang, C. Y., & Li, L. (2013). Disappearance of the unmasking effect of temporally pre-presented lipreading cues on speech recognition in people with chronic schizophrenia. *Schizophrenia Research*, 150, 594–595.
- Wu, C., Zheng, Y., Li, J., Wu, H., She, S., Liu, S., Ning, Y., & Li, L. (2017). Brain substrates underlying auditory speech priming in healthy listeners and listeners with schizophrenia. *Psychological Medicine*, 47, 837–852.
- Wu, C., Zheng, Y. J., Li, J. H., Zhang, B., Li, R. K., Wu, H. B., ... Li, L. (2017). Activation and Functional Connectivity of the Left Inferior Temporal Gyrus during Visual Speech Priming in Healthy Listeners and Listeners with Schizophrenia. *Frontiers in Neuroscience*, 11, 107.
- Wu, M. H., Li, H. H., Gao, Y. Y., Lei, M., Teng, X. B., Wu, X. H., ... Li, L. (2012). Adding irrelevant information to the content prime reduces the prime-induced unmasking effect on speech recognition. *Hearing Research*, 283, 136–143.
- Wu, X. H., Wang, C., Chen, J., Qu, H. W., Li, W. R., Wu, Y. H., ... Li, L. (2005). The effect of perceived spatial separation on informational masking of Chinese speech.

- Hearing Research*, 199, 1–10.
- Wu, Z. M., Chen, M. L., Wu, X. H., & Li, L. (2014). Interaction between auditory and motor systems in speech perception. *Neuroscience Bulletin*, 30, 490–496.
- Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 20664–20669.
- Yang, Z. G., Chen, J., Huang, Q., Wu, X. H., Wu, Y. H., Schneider, B. A., & Li, L. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Communication*, 49, 892–904.
- Zeng, F. G., Nie, K. B., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., & Chen, H. B. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues (L). *The Journal of the Acoustical Society of America*, 116, 1351–1354.
- Zhang, C. X., Arnott, S. R., Rabaglia, C., Avivi-Reich, M., Qi, J., Wu, X. H., ... Schneider, B. A. (2016). Attentional modulation of informational masking on early cortical representations of speech signals. *Hearing Research*, 331, 119–130.
- Zhang, C. X., Lu, L. X., Wu, X. H., & Li, L. (2014). Attentional modulation of the early cortical representation of speech signals in informational or energetic masking. *Brain and Language*, 135, 85–95.
- Zhang, S., & Li, C. S. R. (2012). Functional connectivity mapping of the human precuneus by resting state fMRI. *NeuroImage*, 59, 3548–3562.
- Zhang, Y., Meyers, E. M., Bichot, N. P., Serre, T., Poggio, T. A., & Desimone, R. (2011). Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 8850–8855.
- Zheng, Y., Wu, C., Li, J., Wu, H., She, S., Liu, S., ... Li, L. (2016). Brain substrates of perceived spatial separation between speech sources under simulated reverberant listening conditions in schizophrenia. *Psychological Medicine*, 46, 477–491.

The brain network mechanisms underlying perceptual unmasking cue-induced improvement of speech recognition under cocktail-party listening conditions

LI Liang¹; ZHENG Yingjun²; WU Chao³; LI Juanhua²;
ZHANG Changxin⁴; LU Lingxi¹

⁽¹⁾ School of Psychological and Cognitive Sciences, Peking University, Beijing 100080, China

⁽²⁾ The Affiliated Brain Hospital of Guangzhou Medical University (Guangzhou Huiai Hospital), Guangzhou 510370, China

⁽³⁾ School of Psychology, Beijing Normal University, Beijing 100875, China

⁽⁴⁾ School of Education, China East Normal University, Shanghai 200241, China

Abstract: In a noisy environment with multiple-people talking, how does a listener use certain perceptual cues to facilitate recognition of target speech? To answer this “cocktail-party” problem, researchers have used brain-image methods to examine the underlying brain-network mechanisms. Recent studies have shown that using a unmasking cue, which is associated with a feature of target speech, is able to not only shorten the latency of the auditory cortical responses to a target-speech signal, but also facilitate activation and functional connection of the brain regions that are associated with both cue-specific and cue-nonspecific attention, speech expression, inhibitory functions, and speech motoring, thereby forming the brain-network foundation for the unmasking-cue-induced improvement of the target speech perception against speech informational masking.

Key words: cocktail-party problem; informational masking; unmasking; speech recognition; perceived spatial separation; auditory speech priming; visual speech priming; brain network; neuroimaging.