

视 - 听注意神经机制研究综述

刘 恒, 赵 欣*

内蒙古师范大学心理学院, 内蒙古 呼和浩特

收稿日期: 2023年2月9日; 录用日期: 2023年3月15日; 发布日期: 2023年3月24日

摘 要

随着脑成像技术的快速发展, 视 - 听注意的脑功能解码已成为注意研究的热点问题。国内外学者从视觉或听觉单感觉通道到视 - 听多感觉通道等多个角度, 对视 - 听注意的神经机制进行了揭示。研究显示, 视 - 听注意既受到自上而下目标驱动加工模式的调控, 又受到自下而上刺激驱动加工模式的调控; 视 - 听刺激同步性及一致性也是注意调控的重要影响因素。未来研究可以在视 - 听刺激同步技术及多模态成像联合技术等方面寻找突破, 对视 - 听注意的脑功能解码做进一步探索。

关键词

脑功能成像技术, 视 - 听注意, 多感觉整合

A Review of the Neural Mechanism of Visual-Audio Attention

Heng Liu, Xin Zhao*

School of Psychology, Inner Mongolia Normal University, Hohhot Inner Mongolia

Received: Feb. 9th, 2023; accepted: Mar. 15th, 2023; published: Mar. 24th, 2023

Abstract

With the rapid development of functional neuroimaging technologies, the brain function decoding process of visual-audio attention has become a hot topic in attention research. The neural mechanism of visual-audio attention from visual or auditory processes to audio-visual processes has been studied by researchers at home and abroad. Researches show that visual-audio attention is regulated by both top-down (goal-driven) control and bottom-up (stimulus-driven) control. Visual-audio synchronization and consistency are also important factors affecting attention regulation.

*通讯作者。

Future researches need to look for breakthrough in audio-visual synchronization and multimodal imaging techniques to explore the brain function decoding process of visual-audio attention.

Keywords

Functional Neuroimaging Technologies, Visual-Audio Attention, Multisensory Integration

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,在特定情境下注意力维持的脑功能解码已经成为脑功能研究领域的热点问题。注意力在人类工作和生活中具有重要作用,例如车辆驾驶、航海、飞行等(Cabrall, Happee, & Winter, 2016)。注意力在特定情境中的监察和探测能力,反映了大脑进行事件加工的能力。事件越复杂,就要求能够保证事件顺利执行的注意力越集中(Jahani et al., 2016)。注意力不够集中引发的事故将造成严重损失。例如,据统计,在全球,每年因交通事故造成死亡的人数大约为 130 万人,非死亡人数大约为 200~300 万人(Goniewicz et al., 2016)。而驾驶员注意力分散是造成交通事故最普遍也是最主要的原因之一(Bealand et al., 2013)。

监测注意力的方法主要分为三种:主观量表法、行为测量法及生理测量法。主观测量法操作简单,但结果受主观影响过大,并且缺乏实时性。行为测量法相对主观量表法来说更客观一些,但是同样缺乏实时性,并且实施过程会受到一些环境因素的限制(Li et al., 2018)。生理测量法能够在注意力实验的过程中提取出与注意力变化相关的生理参数,从而实现了实时监测(龚涛, 2018)。很多脑功能成像技术已经被应用到了注意力维持的研究中,使揭示注意调控的神经机制成为了可能。

人类生活的环境极其复杂,通过不同的感官会从环境中收集到大量多种多样的信息。视觉和听觉是人类获取外部信息的主要通道,也是维持注意力的主要刺激来源(Chouhan et al., 2015)。已有研究表明,将注意力集中于某种感觉刺激(视觉、听觉或触觉)时,可以增强负责该种感觉刺激信号加工的大脑皮层的神经活动(龚涛, 2018)。那么,一种感觉刺激的增强就有可能对另一种感觉刺激产生影响。来自不同感官的刺激信号是相互促进还是相互抑制的呢?近年来,已经有很多研究对不同感觉模态下注意力加工过程的脑神经机制进行了研究,尤其是视觉和听觉。

2. 国内外研究现状及发展动态

人脑是高效信息处理系统,视觉和听觉是外来信息传入大脑的主要通道。而在人类信息处理系统中,注意力在选择和整合信息的过程中扮演着重要角色(唐晓雨等, 2021)。当人类身处特定环境中时,会有选择的将注意力偏向与当前任务相关的信息上,忽略掉不相关信息(Tootell et al., 1998)。这种偏向受到自上而下或自下而上两种模式的调控。自上而下调控模式,又叫做目标-驱动调控,即个体基于某些目的在环境中主动寻找信息(Hickey, McDonald, & Theeuwes, 2006)。自下而上调控模式又叫做刺激-驱动调控,即由于外部刺激的显著特征而吸引个体的注意(Hickey, McDonald, & Theeuwes, 2006)。目标-驱动调控是一种以条件选择变量为指导的,较慢的认知调控机制,而刺激-驱动调控是一种快速的,以场景中的显著刺激为指导的原始调控机制(Hikosaka, Miyachi, & Shimojo, 1997)。

以往研究运用 PET、EEG、fMRI 及 fNIRS 等脑功能成像技术对视听刺激引起的注意力变化的神经机制进行了探索。研究表明, 额叶及顶叶在空间注意的调控过程中扮演者重要角色。这些区域与感觉及运动相关皮质相互影响, 包括, 内侧前额叶皮质、背外侧前额叶皮质、扣带回、顶内沟、顶下小叶及顶上小叶(Brunetti et al., 2008; Grent-'t-Jong & Woldorff, 2007; Stern et al., 2007)。其中, 背外侧前额叶皮质已被认为是注意选择过程中自上而下调控模式的核心部位, 这一区域受损会导致注意及记忆损伤(Mesulam, 2000)。侧顶叶与注意力定向息息相关(Luks et al., 2008)。背外侧前额叶和侧顶叶都会对目标刺激产生活跃反应, 反应强度随着注意任务难度的增加而增强(Luks et al., 2008)。因此, 这些脑区都可能在注意的调控和维持过程中发挥着作用。

2.1. 单感觉通道研究

2.1.1. 视觉注意

人类面对复杂的视觉世界, 只有一小部分场景能够输入人脑, 用于在特定时间和环境下指导人类的行为活动。而人类为了快速有效的进行行为活动, 需要通过注意力将相关的视觉信息优先提取出来(Tsotsos, 1990)。Gola 等人通过监测 EEG 信号发现, 无论是老年人还是青年人在做视觉注意任务时, 都会在正确操作前在枕部记录到 beta 波, 而在错误操作前则不存在 beta 波。因此, 他们认为 beta 波与视觉注意有关(Gola et al., 2013)。

视觉注意究竟是目标 - 驱动调控, 刺激 - 驱动调控还是同时受两者共同调控仍然存在争议。基于脑成像技术的人类视觉皮质研究为视觉注意力调控模式的争论提供了神经层面的证据。在目标 - 驱动调控视觉注意研究方面, Clark 等人的研究证实, 只有当视觉输入经过初级视皮层 V1 区时, 才会引发自上而下注意信号的响应。当实验对象接收到视觉指导信号后, 其大脑皮层的激活强度在面对与指导信号一致的注意力任务时要高于非指导信号注意力任务, 并且随着注意任务难度的增加而增加(Clark & Hillyard, 1996)。Martínez 等人进行了空间注意的 EEG 与 fMRI 联合研究, 被试需要在干扰矩阵中搜索异常目标。EEG 结果表明, 刺激启动后最初的神活动定位于纹状皮层, 并不受到空间注意的调节。而 fMRI 结果则显示, 注意能够引起初级视觉皮层的响应(Martínez et al., 1999)。在刺激 - 驱动调控视觉注意研究方面, PET 和 fMRI 的研究均表明, 刺激信号的特征属性能够激活 V1、MT+、V5 等视觉皮质区(Liu et al., 2003; Liu & Mance, 2011)。将注意力放在整个视觉范围内时, 视觉扫过特定刺激特征时会引起与该特征相关的皮质区激活。例如, 一个运动刺激会激活大量运动相关感觉区域——MT 区的反应, 而颜色变化刺激会激活大量颜色选择区——V4/V8 区的反应(Schoenfeld et al., 2007)。

2.1.2. 听觉注意

听觉注意是人类过滤外界繁杂的感觉信息, 选择有用信息的另一个重要途径(Hernandez-Peon, 1966)。听觉注意同样受到自下而上刺激 - 驱动模式和自上而下目标 - 驱动模式的调控(Kaya & Elhilali, 2016)。自下而上听觉注意力调控模式的研究多以声谱特征中的时间、频率、振幅、音调、音色、熵值以及噪音等为实验刺激, 通过行为来反应注意力变化(Kalinli & Narayanan, 2007; Wang et al., 2015)。而自上而下听觉注意力调控模式的研究则是通过神经激活程度来反应注意力变化, 更好的体现了听觉皮质的激活是受指导信号注意力调控的(Mesgarani & Chang, 2012; Weinberger, 2011)。听觉注意调控的神经机制也通过脑成像技术被证实(Langner et al., 2012; Puschmann et al., 2017)。Petkov 等人的研究表明外侧颞叶听觉皮层的带状区在声信号刺激引发注意后会产生强烈的激活(Petkov, 2004)。Fritz 等人通过一系列的实验证实了一个关于听觉注意的重要推论, 既听觉皮质通过改变响应特征使目标信号能够更好的与干扰信号区分开。也就是说, 听觉皮层对刺激的反应与接收到的任务有关, 这符合自上而下调控模式(Fritz et al., 2007)。Alho

等人发现对连续语言刺激的选择注意调控发生在颞上回和颞上沟(Alho et al., 2014)。

揭示鸡尾酒会效应(Cocktail party effect)的神经机制是听觉注意研究的一个重要方向。鸡尾酒会效应是指人们在嘈杂的环境中将注意力集中在与某一个人的谈话之中而忽略背景中其他的对话或噪音的选择注意能力(Golumbic et al., 2013)。Ding 等人的研究显示, 在多重听觉信号中进行注意选择的神经加工位于后听觉皮层, 并且即使目标刺激和背景刺激的声信号在频谱图上完全重合时, 被试听觉皮层对二者声音的神经编码也是单独进行的(Ding & Simon, 2012b)。Golumbic 等人的研究则表明, 大脑使用低频相位、高频振幅的波跟踪语音流。在低级听觉中枢, 注意力通过增强皮质对目标语音流的跟踪来调节语音表征, 但背景语音也被表征。在高级听觉皮质中枢则表现出了更高的选择性, 因为在该脑区检测不到背景语音的表征(Golumbic et al., 2013)。Power 等人利用听觉诱导扩展频谱分析(Auditory evoked spread spectrum analysis, AESPA), 分别对两个同时出现的语音刺激进行线性脉冲反应测量, 结果证明将注意力集中到两个语音刺激中的一个, 会在左脑产生对这两个语音刺激的脉冲反应, 二者之间会有大约 200 ms 的延迟(Power et al., 2012)。O'Sullivan 等人则通过 EEG 测量提出鸡尾酒会效应神经调控的关键点在 200 ms 左右(O'Sullivan et al., 2014)。Evans 等人发现, 大脑对背景噪音和目标语言刺激的加工作用于同一通路, 注意任务完成越好的个体对左颞上回的激活越强烈(Evans et al., 2016)。Begau 等人的研究证实视觉-听混合背景下产生的 ERPs 存在年龄差异, 只有老年人的 theta 波和 beta 波在视觉和听觉条件之间存在差异, 且与年轻组相比, 年长组 beta 扰动更强烈(Begau et al., 2022)。

2.2. 多感觉通道研究

虽然有关注意加工的研究起初只关注于单一感觉通道, 但越来越多的研究开始聚焦于多感觉通道(Lee et al., 2020)。如何从复杂的多感觉模态信息网络中提取某一感觉信号, 及多个感觉通道信息如何相互整合是主要难题(Begau et al., 2022; Santangelo & Macaluso, 2012)。

2.2.1. 视觉注意与听觉注意的比较研究

一些研究结果表明, 听觉刺激相对于视觉刺激更能够引发顶枕部 α 频带信号(8~14 Hz)的激活(Trenner et al., 2008)。Li 等人的研究证实, 视觉空间注意和听觉空间注意的脑皮质功能激活区存在重叠区域, 顶叶及额叶在听觉空间注意任务和视觉空间注意任务测试过程中都会被激活。但是顶叶在视觉空间注意任务时会被激活, 在听觉空间注意任务时不会被激活(Li et al., 2007)。Grani 等人则证实了视觉和听觉组合在一起的信号刺激比二者单独存在的信号刺激更能够吸引人的注意力, 也能够引发更多的行为反应(Grani et al., 2014)。事件相关电位(ERPs)研究发现, 当用听觉刺激引发注意时, 听觉加工在 Na 成分的潜伏期(约 30 ms)以及 P1 成分期间(约 90 ms)受到调节; 而当视觉刺激引发注意时, 视觉输入在 C1 (约 60 ms)成分早期的潜伏期, P1 (80~100 ms)成分期间及 N1 成分时受到调节(Karns & Knight, 2009; Talsma & Woldorff, 2005)。近期的 fMRI 证据显示, 当利用为对话添加噪音及为画面添加马赛克的方式改编听觉刺激质量及视觉刺激质量时, 随着视觉刺激质量和听觉刺激质量的提升都会引起双侧颞上回和颞上沟激活的增强。此外, 视觉刺激质量和听觉刺激质量的下降还会引起额顶叶激活的增强, 这可能反映了注意力需求的增加。最后, 与将注意力集中在视觉刺激相比, 将注意力集中在听觉刺激能引发角回、右颞极、眶额、腹内侧前额叶和后扣带回激活的增强(Leminen et al., 2020; Zhao et al., 2021)。

2.2.2. 视觉注意与听觉注意相互转换研究

Rapela 等人运用脑电技术对视听注意快速转换间的脑电与持续性视觉注意或听觉注意的脑电的差异进行了比较研究。结果显示, 当实验对象接收到“看”的指令后, 其与注意相关的脑电活动在从听觉刺激快速转换为视觉刺激时要高于始终集中于视觉刺激时。当实验对象的注意在视觉刺激与听觉刺激间快

速转换时, 脑电信号会随着指令与视觉刺激间间隔的增加而减弱。脑区视觉注意的衰退甚至能持续到注意已经转移到听觉刺激上, 表明视觉与听觉注意间存在某种交叉作用, 能够提高注意水平, 并且不仅作用于单一感觉注意, 还作用于任务执行过程中涉及到的其他感觉注意(Rapela et al., 2012)。陈雪莹等人也进行了视觉和听觉跨通道注意转移的 ERPs 研究。结果证实, 视觉和听觉线索目标刺激的 N1, P1 和 P3 成分皆出现通道转移的主效应。听觉目标刺激在 N1 成分上在有通道转移时引发的波幅更大, 在 P1 和 P3 成分上在无通道转移时引发更大的波幅。视觉目标刺激在 N1, P1 和 P3 成分上在有通道转移时引发的波幅都要大于在无通道转移时引发的波幅。在视觉和听觉中, 有无通道转移引发的脑电活动的差异主要集中在额区(陈雪莹, 吕勇, 2013)。视 - 听注意的跨通道转换不仅体现在通道间转变, 也存在于通道间传递。例如, 在情绪对注意调控的研究中, 发现无论视听情绪刺激是否一致, 在 200~300 ms 的时间窗内, 视 - 听刺激同步呈现时提取的听觉 ERPs 都明显大于单独呈现听觉刺激时的 ERPs, 说明一个通道的激活会传递给另一个通道(Chen et al., 2022)。无论将视觉还是听觉设置为任务相关, 在有消极情绪呈现的情况下, 情绪一致刺激呈现 180 ms 后的 ERPs 均比情绪不一致刺激呈现 180 ms 后的 ERPs 显著活跃。相对于情绪一致刺激, 情绪不一致刺激在 400~550 ms 内出现的 N400 成分显现出了更高的负面影响(Foecker, 2019)。表明与情绪相关的视 - 听感觉通道间的信息是自动交互的。又如, 在空间选择注意的研究中发现, 听觉目标刺激呈现 200~400 ms 间出现的 Nd 成分并不会因周围视觉刺激是否受到关注而改变, 但当听觉刺激单独呈现时会减少(Zhao et al., 2021)。

2.2.3. 注意的多感觉通道整合研究

有关视 - 听注意多感觉整合(Multisensory integration, MSI)的研究也受到了人们的广泛关注。多感觉整合是个体将不同感觉通道(如视觉、听觉、触觉)的信息整合为统一、连贯、稳定的有意义知觉过程(王爱君等, 2020; Feng, 2022)。能够整合来自多种感官的信息是我们认知周围世界的基本要求。通过注意的调控, 多个感觉通道输入的信息被整合到不同的加工阶段, 最终形成整体认知(Badamdorj et al., 2021; Ding & Simon, 2012a; Shamma, Elhilali, & Michey, 2011)。如, 在听一个人说话时, 我们的感觉系统能够将我们耳朵听到的声音和眼睛看到的说话者的手部动作和唇部运动整合起来, 最终让我们对说话者叙述的事件形成一个整体的认识(Schutz & Silverman, 2011)。神经影像研究已经显示, 多感觉通道信息整合涉及多个脑皮层区域。例如, 颞上区和内侧前额叶都参与了语言和唇部运动的整合(Ruesseler et al., 2017)。再如, 听觉刺激与视觉刺激语义一致会引起右侧颞中回和颞上回显著激活(Plank et al., 2012)。与单独的听觉刺激相比, 只有当同步呈现的视觉刺激与听觉刺激指示相同时, 在视觉皮层和听觉皮层进行的语义追踪加工及听觉注意编码才会增强(Fu, Wu, & Chen, 2019; Fu & Chen, 2020)。颞顶叶连接区和颞前叶也被普遍认为在跨通道语义表征中起着关键而特殊的作用(Mirman, Landrigan, & Britt, 2017; Ralph et al., 2016; Schwartz et al., 2011)。电生理的研究显示, 选择注意的调控机制不会局限于一个单一的感觉模态, 可以包含或跨越多个感觉系统(Busse et al., 2005), 来自多个感官系统的信息也不会单独加工, 而是会在多重情境下整合到一个多感觉认知系统中进行加工(Fort et al., 2002a, 2002b)。注意的多感觉通道整合还受到时间同步性的调控, 例如 ERP 研究证实, 仅当视 - 听刺激同步时, 枕叶早期跨感觉通道相关 P195 成分和顶叶晚期跨感觉通道相关 N440 成分才显著(Feng et al., 2022)。

注意与多感觉整合的相互作用关系复杂又具有争议。一方面, 是因为注意与多感觉整合的相互影响具有多个水平(Rohe, Ehrlis, & Noppeney, 2019); 另一方面, 无论注意还是多感觉整合都是复杂、多层面的过程, 他们作用于感觉加工的调控, 并最终影响行为(Macaluso et al., 2016; Mole, 2020)。Giard 等人 and Molholm 等人提出了视听融合研究的理论模型。他们认为, 视听整合就是同时接受视觉与听觉刺激后的 ERPs (Multisensory audiovisual, AV)与单独接受视觉刺激的 ERPs (Unisensory visual, V)和单独接受听觉刺

激的 ERPs (Unisensory auditory, A)之和(A + V)的差值(Giard & Peronnet, 1999; Molholm et al., 2002)。随后, Durk 等人基于这一理论对视听融合与注意的相互作用关系进行了研究。结果显示, 只有当视觉和听觉同时被注意时, 才会在 P50 成分发现视听融合效应, 即 P50 成分的视听融合 ERPs 比单独的视觉和听觉 ERPs 之和小。他们还发现, 当被试在视听共同刺激任务中将注意力集中于视觉刺激时, 会发生一个延迟的前额叶负波增强, 而这种负波增强与听觉刺激注意加工负波极其相似, 暗示着以视觉刺激为主的视听注意加工在后期已经从视觉注意加工传递到了听觉注意加工(Durk, Doty, & Woldorff, 2007)。

Xi 等人针对注意调控视 - 听整合的神经机制进行了一系列有价值的研究。首先, 他们应用 fMRI 研究证实了视 - 听整合注意能够引起前额叶、额顶联合区以及额顶叶的激活, 并且前额叶激活程度最高, 表明视听整合注意加工是一个需要多个皮层网络联合加工的复杂过程, 前额叶在其中发挥着重要作用(Xi, Li, Gao, He, & Tang, 2019)。然后, 他们对比了注意状态和非注意状态下视 - 听整合的神经机制, 结果表明, 初级视觉皮层、初级听觉皮层及颞上回、颞中回是注意调控视 - 听整合的重要脑区, 且注意状态下的脑区激活都要显著高于非注意状态下的脑区激活(Xi, Zhang, Gao, Li, Liu, & Li, 2019)。在最新的研究中, 他们通过添加听觉干扰和视觉干扰排除了视 - 听整合中的基本刺激整合, 提取出了更高层次的语义整合。结果显示, 在注意状态下, 在 220~240 ms 的双侧枕颞区和 560~600 ms 的额区分别出现了两次语义整合效应; 在非注意状态下, 只在 340~360 ms 的中央前额区发现一次语义整合效应。说明, 在注意状态下, 视 - 听语义信息首先进行一次整合, 再针对当前任务进行一次再分析, 符合目标驱动的自上而下注意调控模式; 在非注意状态下, 没有进行注意分配, 只进行了一次无意识的自动语义整合(Xi et al., 2020)。Rohe 等人结合心理物理学、fMRI 及最大似然估计法, 发现听觉皮层和视觉皮层分别主要编码听觉刺激和视觉刺激中的空间定位信息, 而顶内沟则是对视 - 听整合后的信息进行空间表征。最重要的是, 顶内沟对于视听整合信号内的空间信息表征是受到注意的调控的, 也就是受到自上而下模式的调控(Rohe & Noppeney, 2018)。来自 ERPs 的证据显示, 早期的视 - 听整合发生在 40~160 ms 的注意相关脑区, 视 - 听整合信号空间及语义加工发生在 200~280 ms 的与注意目标相关的脑区, 后期的视 - 听整合发生在 320~520 ms 的决策相关脑区。并且在视 - 听拼写任务中神经网络信息传递的能力和效率在 40~160 ms、200~280 ms 及 320~520 ms 都要高于单一视觉拼写任务(Lu et al., 2020)。

腹语效应是视 - 听整合发生扭曲的现象, 声信号的空间定位被视觉信号输入扭曲, 以至于观察者误以为声音发自被控制的木偶, 而不是控制木偶的人(Jack & Thurlow, 1973)。通过 EEG 研究验证空间注意是否能够调节腹语效应, 结果显示自上而下的注意抑制了对侧颞电极上的 alpha 波(10 Hz)的波幅, 而自下而上的注意将脑电相位转移到了 theta 波(7 Hz), 而不是抑制振幅。因此, 两种不同的空间注意对于腹语效应进行调节的神经机制也是不同的(Kumagai & Mizuhara, 2016)。

2.3. 认知障碍人群视 - 听注意研究

认知障碍人群对视觉、听觉的感知和对注意的加工都可能会存在某种异常, 因此, 越来越多的研究者开始探索该类人群视 - 听注意加工的差异(Daibert-Nido et al., 2021)。例如, Kang 等人通过对阅读障碍(Dyslexia)儿童在目标搜索任务中事件相关电位的研究, 发现阅读障碍儿童在视觉搜索任务中诱发的 P300 波幅显著低于非阅读障碍儿童, 而 P300 成分通常被认为与视觉空间注意及注意的分配和持续加工有关(Kang et al., 2016)。Fan 等人发现, 高功能孤独症患者(High-functioning autism, HFA)在进行视觉注意判断任务过程中, 虽然其左右半球的上前扣带回、右缘上回、梭状回等会发生明显激活, 但与正常发育个体间并不存在显著差异, 因此认为该类人群小脑和顶叶并不存在异常(Fan et al., 2012)。对注意力缺陷多动障碍(Attention deficit/hyperactivity disorder, ADHD)儿童的研究发现, 在视觉注意定向任务中, 他们与无关信息自主加工的 P2 成分会显著增强, 而与工作记忆相关的 CNV 波幅会明显降低(Ortega et al., 2013)。

对注意力缺陷多动障碍视-听整合注意研究发现, 注意力缺陷多动障碍患者无论是对单独视觉刺激、单独听觉刺激还是视-听同步刺激的反应都要快于正常对照组。EEG 分析结果显示, 无论是注意力缺陷多动障碍患者还是正常对照组都存在多感觉整合早期加工, 并且注意力缺陷多动障碍患者在刺激后的早期潜伏期 110~130 ms 时, 在顶枕区会进行更多的多感觉整合加工(McCracken et al., 2019; McCracken et al., 2020)。在述情障碍患者中则发现, 无论视-听情绪刺激是否一致, 高述情障碍被试与低述情障碍被试相比, P2 成分振幅都更高, 情绪识别能力都更差。贝叶斯统计显示, 述情障碍患者情绪的多通道整合情况与以往研究中发现的多通道整合 ERP 成分(N1 和 N2)无差异(Wang et al., 2021)。说明述情障碍患者虽然在视觉或听觉单通道上存在情绪识别障碍, 但是在多通道整合上是没有缺陷的。

3. 小结和展望

人类的感知是通过多个感觉通道不断输入的信息所控制的。本文从单一的视觉或听觉得到视-听多感觉整合对视-听注意的神经机制进行了系统的梳理。众多研究结果表明, 视听注意既受到自上而下目标驱动加工模式的调控, 又受到自下而上刺激驱动加工模式的调控。这是因为人类在感知周围环境时即受到环境中一些视觉及听觉刺激的特定特征的影响, 又受到自身行为目标、先前经验、短时记忆等内在因素的影响。自上而下调控和自下而上调控对稳定、准确的认知加工至关重要, 探索二者的神经网络组成与联合也一直是认知科学的研究重点。进一步的神经层面的研究显示, 同步的且一致的多感觉刺激对注意具有增进的作用。视觉和听觉同步的信号刺激比单独的视觉信号刺激或听觉信号刺激更能够引起人的注意, 也能够引发更多的行为反应, 并且只有听觉刺激与视觉刺激语义一致时才会引起大脑皮层一些区域的显著激活。这可能是由于视觉注意与听觉注意都占用着同一注意资源, 而不一致的听觉与视觉信号的输入会降低目标信号所分配到的注意资源, 从而对于目标刺激的神经加工也会降低。对于多感觉加工的方式, 虽然一些证据表明, 不同感觉通道输入的信息是以平行的方式处理的, 而且只在高级中枢进行整合加工, 但也有证据表明, 一些多感觉整合在初级感觉中枢就存在了。

脑功能成像技术短时间内的迅猛发展, 助力了揭示视-听注意神经机制的研究。然而, 与许多快速发展中的科学领域一样, 视-听注意的脑功能解码在探索过程中也遇到了很多困难, 发现了一些问题, 总结了一些经验。

3.1. 视-听多感觉通道整合是研究趋势

上文的研究均表明, 听觉注意与视觉注意无论是在神经层面, 还是在行为层面都不是相互独立的。实际上, 人类从环境中接受的信息基本不会是纯视觉或者纯听觉的, 进行视-听多感觉通道整合研究, 探索视-听刺激对注意影响的神经机制更加具有现实意义。单一感觉刺激对注意的影响与视-听同步刺激对注意的影响是否存在差异? 促进注意与视听融合通道相互影响的神经网络是什么样的? 最能解释注意与视听融合通道相互影响的模型是哪个? 这些问题尚需进一步探索。

3.2. 视-听刺激同步是挑战

视-听多通道整合是感觉-注意调控的神经机制研究的发展方向, 相应的, 视-听刺激同步就成为了该类研究的技术关键。因为视-听刺激同时呈现是此类研究的主要环节, 只有视-听刺激同步达到一定水平, 才能保证研究结果的准确性。人类具有很强的辨别视觉信号与听觉信号是否同步的能力, 能够在视频中注意到画面与音频非常微小的不匹配, 因此, 在对人类视-听注意进行研究时就更需要保证视-听刺激的同步性。目前, 参考已被广泛接受的将单一感觉刺激融合成多感觉刺激的方法(Vroomen & Keetels, 2010; Werner & Noppeney, 2011), 实现视-听刺激同步的做法, 是将纯音刺激的频率与视觉刺激

的脉冲频率调整到同一频率(Keitel & Müller, 2015)。但是, 这种方法只实现了视觉刺激与听觉刺激的时间上的同步, 却不能保证空间同步, 还需要更加复杂和精确的技术及模型实现视觉刺激与听觉刺激的时间-空间同步。

3.3. 多模态成像研究是新手段

采用不同脑成像技术对大脑数据进行共同采集被称为多模态成像研究(叶佩霞等, 2017)。PET、EEG、fMRI 及 fNIRS 等多种脑信号技术均可用于注意的神经机制研究。然而, 每种技术都有各自的优点和局限性: 从时间分辨率来说, EEG 要高于 fNIRS 和 fMRI, 最低的是 PET; 从空间分辨率来说, fMRI 要高于 PET, 其次是 fNIRS, 最低的是 EEG; PET 需要将标记有放射性核素的蛋白质、核酸等物质注入人体才能进行脑信号测量; fMRI 虽为非侵入式的技术, 但是设备费用高昂且携带不便, 无法进行较灵活的实验; EEG 和 fNIRS 即属于非侵入式技术, 携带上也相对来说比较便利, 但是又各自存在空间分辨率或时间分辨率低的缺点(熊馨等, 2018)。多模态成像技术集合两种及以上成像技术的优点, 抵消缺点, 既可以对同一研究问题的不同模态数据进行相互比较, 又可以提供较全面神经信号信息, 能够更加准确揭示认知加工神经机制。

基金项目

本研究得到内蒙古自治区自然科学基金联合项目(2020LH03014)、内蒙古师范大学高层次人才科研启动项目(2019YJRC015, 2020YJRC016)、内蒙古师范大学基本科研业务费项目(2022JBQN119, 2022JBQN116)的资助。

参考文献

- 陈雪莹, 吕勇(2013). 视觉和听觉跨通道注意转移的 ERP 研究. 见 *第十六届全国心理学学术会议*(pp. 410-411).
- 龚涛(2018). *利用功能磁共振成像技术探讨注意力调控感觉信息处理的相关脑功能区*. 硕士学位论文, 天津: 天津医科大学.
- 唐晓雨, 佟佳庚, 于宏, 王爱君(2021). 内外源性空间注意对多感觉整合的影响. *心理学报*, 53(11), 1173-1188. <https://doi.org/10.3724/SP.J.1041.2021.01173>
- 王爱君, 黄杰, 陆菲菲, 何嘉滢, 唐晓雨, 张明(2020). 多感觉整合中的声音诱发闪光错觉效应. *心理科学进展*, 28(10), 1662-1677. <https://doi.org/10.3724/SP.J.1042.2020.01662>
- 熊馨, 伏云发, 张夏冰, 李松, 徐保磊, 尹旭贤(2018). 一种多模态脑电和近红外光谱联合采集头盔设计及实验研究. *生物医学工程学杂志*, 35(2), 290-296. <https://doi.org/10.7507/1001-5515.201611025>
- 叶佩霞, 朱睿达, 唐红红, 买晓琴, 刘超(2017). 近红外光学成像在社会认知神经科学中的应用. *心理科学进展*, 25(5), 731-741. <https://doi.org/10.3724/SP.J.1042.2017.00731>
- Alho, K., Rinne, T., Herron, T. J., & Woods, D. L. (2014). Stimulus-Dependent Activations and Attention-Related Modulations in the Auditory Cortex: A Meta-Analysis of fMRI Studies. *Hearing Research*, 307, 29-41. <https://doi.org/10.1016/j.heares.2013.08.001>
- Badamdorj, T., Rochan, M., Yang, W., & Li, C. (2021). Joint Visual and Audio Learning for Video Highlight Detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 8107-8117). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/ICCV48922.2021.00802>
- Beanland, V., Fitzharris, M., Young, K. L., & Lenné, M. G. (2013). Driver Inattention and Driver Distraction in Serious Casualty Crashes: Data from the Australian National Crash in-Depth Study. *Accident Analysis and Prevention*, 54, 99-107. <https://doi.org/10.1016/j.aap.2012.12.043>
- Begau, A., Klatt, L. I., Schneider, D., Wascher, E., & Getzmann, S. (2022). The Role of Informational Content of Visual Speech in an Audiovisual Cocktail Party: Evidence from Cortical Oscillations in Young and Old Participants. *The European Journal of Neuroscience*, 56, 5215-5234. <https://doi.org/10.1111/ejn.15811>
- Brunetti, M., Della Penna, S., Ferretti, A., Gratta, C., Cianflone, F., Belardinelli, P., & Romani, G.-L. (2008). A Frontoparietal Network for Spatial Attention Reorienting in the Auditory Domain: A Human fMRI/Meg Study of Functional and

- Temporal Dynamics. *Cerebral Cortex*, 18, 1139-1147. <https://doi.org/10.1093/cercor/bhm145>
- Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The Spread of Attention across Modalities and Space in a Multisensory Object. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 18751-18756. <https://doi.org/10.1073/pnas.0507704102>
- Cabrall, C. D. D., Happee, R., & Winter, J. C. F. D. (2016). From Mackworth's Clock to the Open Road: A Literature Review on Driver Vigilance Task Operationalization. *Transportation Research, Part F. Traffic Psychology and Behaviour*, 40, 169-189. <https://doi.org/10.1016/j.trf.2016.04.001>
- Chen, M., Zhao, S., Yu, J., Leng, X., Zhai, M., Feng, C., & Feng, W. (2022). Audiovisual Emotional Congruency Modulates the Stimulus-Driven Cross-Modal Spread of Attention. *Brain Sciences*, 12, 1229. <https://doi.org/10.3390/brainsci12091229>
- Chouhan, T., Panse, A., Smitha, K. G., & Vinod, A. P. (2015). A Comparative Study on the Effect of Audio and Visual Stimuli for Enhancing Attention and Memory in Brain Computer Interface System. In *The IEEE International Conference on Systems* (pp. 3104-3109). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/SMC.2015.539>
- Clark, V., & Hillyard, S. (1996). Spatial Selective Attention Affects Early Extrastriate but Not Striate Components of the Visual Evoked Potential. *Journal of Cognitive Neuroscience*, 8, 387-402. <https://doi.org/10.1162/jocn.1996.8.5.387>
- Daibert-Nido, M., Pyatova, Y., Markowitz, M., & Markowitz, S. N. (2021). Visual Outcomes of Audio-Luminous Biofeedback Training for a Child with Idiopathic Nystagmus. *Arquivos Brasileiros de Oftalmologia (AHEAD)*, 84, 179-182. <https://doi.org/10.5935/0004-2749.20210026>
- Ding, N., & Simon, J. Z. (2012a). Neural Coding of Continuous Speech in Auditory Cortex during Monaural and Dichotic Listening. *Journal of Neurophysiology*, 107, 78-89. <https://doi.org/10.1152/jn.00297.2011>
- Ding, N., & Simon, J. Z. (2012b). Emergence of Neural Encoding of Auditory Objects While Listening to Competing Speakers. *Proceedings of the National Academy of Sciences*, 109, 11854-11859. <https://doi.org/10.1073/pnas.1205381109>
- Durk, T., Doty, T. J., & Woldorff, M. G. (2007). Selective Attention and Audiovisual Integration: Is Attending to Both Modalities a Prerequisite for Early Integration? *Cerebral Cortex*, 17, 679-690. <https://doi.org/10.1093/cercor/bhk016>
- Evans, S., Mcgettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the Cocktail Party Started: Masking Effects in Speech Perception. *Journal of Cognitive Neuroscience*, 28, 483-500. https://doi.org/10.1162/jocn_a_00913
- Fan, J., Bernardi, S., Dam, N. T. V., Anagnostou, E., Gu, X., Martin, L., & Soorya, L. (2012). Functional Deficits of the Attentional Networks in Autism. *Brain and Behavior*, 2, 647-660. <https://doi.org/10.1002/brb3.90>
- Feng, C., Feng, W., Zhao, S., Wang, C., & Wang, Y. (2022). The Interplay between Audiovisual Temporal Synchrony and Semantic Congruency in the Cross-Modal Boost of the Visual Target Discrimination during the Attentional Blink. *Human Brain Mapping*, 43, 2478-2494. <https://doi.org/10.1002/hbm.25797>
- Feng, W. (2022). Audiovisual Emotional Congruency Modulates the Stimulus-Driven Cross-Modal Spread of Attention. *Brain Sciences*, 12, 1229. <https://doi.org/10.3390/brainsci12091229>
- Foecker, J. (2019). Event-Related Potentials Reveal Evidence for Late Integration of Emotional Prosody and Facial Expression in Dynamic Stimuli: An ERP Study. *Multisensory Research*, 32, 473-497. <https://doi.org/10.1163/22134808-20191332>
- Fort, A., Delpuech, C., Pernier, J., & Marie-Hélène, G. (2002a). Dynamics of Cortico-Subcortical Cross-Modal Operations Involved in Audio-Visual Object Detection in Humans. *Cerebral Cortex*, 12, 1031-1039. <https://doi.org/10.1093/cercor/12.10.1031>
- Fort, A., Delpuech, C., Pernier, J., & Marie-Hélène, G. (2002b). Early Auditory-Visual Interactions in Human Cortex during Nonredundant Target Identification. *Cognitive Brain Research*, 14, 20-30. <https://doi.org/10.1093/cercor/12.10.1031>
- Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory Attention-Focusing the Searchlight on Sound. *Current Opinion in Neurobiology*, 17, 437-455. <https://doi.org/10.1016/j.conb.2007.07.011>
- Fu, Z., & Chen, J. (2020). Congruent Audiovisual Speech Enhances Cortical Envelope Tracking during Auditory Selective Attention. In *The Interspeech 2020* (pp. 116-120). <https://doi.org/10.21437/Interspeech.2020-1957>
- Fu, Z., Wu, X., & Chen, J. (2019). Congruent Audiovisual Speech Enhances Auditory Attention Decoding with EEG. *Journal of Neural Engineering*, 16, Article ID: 066033. <https://doi.org/10.1088/1741-2552/ab4340>
- Giard, M. H., & Peronnet, F. (1999). Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. *Journal of Cognitive Neuroscience*, 11, 473-490. <https://doi.org/10.1162/089892999563544>
- Gola, M., Magnuski, M., Szumska, I., & Wróbel, A. (2013). EEG Beta Band Activity Is Related to Attention and Attentional Deficits in the Visual Performance of Elderly Subjects. *International Journal of Psychophysiology*, 89, 334-341. <https://doi.org/10.1016/j.ijpsycho.2013.05.007>
- Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., & Schroeder, C. E. (2013). Me-

- chanisms Underlying Selective Neuronal Tracking of Attended Speech at a “Cocktail Party”. *Neuron*, 77, 980-991. <https://doi.org/10.1016/j.neuron.2012.12.037>
- Goniewicz, K., Goniewicz, M., Pawłowski, W., & Fiedor, P. (2016). Road Accident Rates: Strategies and Programmes for Improving Road Traffic Safety. *European Journal of Trauma and Emergency Surgery*, 42, 433-438. <https://doi.org/10.1007/s00068-015-0544-6>
- Grani, F., Argelaguet, F., Gouranton, V., Badawi, M., Gagne, R., Serafin, S., & Lécuyer, A. (2014). Audio-Visual Attractors for Capturing Attention to the Screens When Walking in Cave Systems. In *The 2014 IEEE Virtual Reality (VR)* (pp. 75-76). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/VR.2014.6802058>
- Greit-‘t-Jong, T., & Woldorff, M. (2007). Timing and Sequence of Brain Activity in Top-Down Control of Visual-Spatial Attention. *PLOS Biology*, 5, e12. <https://doi.org/10.1371/journal.pbio.0050012>
- Hernandez-Peon, R. (1966). Physiological Mechanisms in Attention. In R. W. Russell (Ed.), *Frontiers in Physiological Psychology* (pp. 121-144). Academic Press.
- Hickey, C., McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological Evidence of the Capture of Visual Attention. *Journal of Cognitive Neuroscience*, 18, 604. <https://doi.org/10.1162/jocn.2006.18.4.604>
- Hikosaka, O., Miyachi, S., & Shimojo, S. (1997). Orienting of Spatial Attention—Its Reflexive, Compensatory, and Voluntary Mechanisms. *Cognitive Brain Research*, 5, 1-9. [https://doi.org/10.1016/S0926-6410\(96\)00036-5](https://doi.org/10.1016/S0926-6410(96)00036-5)
- Jack, C. E., & Thurlow, W. R. (1973). Effects of Degree of Visual Association and Angle of Displacement on the “Ventriloquism” Effect. *Perceptual and Motor Skills*, 37, 967-979. <https://doi.org/10.2466/pms.1973.37.3.967>
- Jahani, S., Berivanlou, N. H., Rahimpour, A., & Setarehdan, S. K. (2016). Attention Level Quantification during a Modified Stroop Color Word Experiment: An fNIRS Based Study. In *The Biomedical Engineering* (pp. 99-103). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/ICBME.2015.7404124>
- Kalinli, O., & Narayanan, S. S. (2007). A Saliency-Based Auditory Attention Model with Applications to Unsupervised Prominent Syllable Detection in Speech. In *The Interspeech 2007* (pp. 1941-1944). <https://doi.org/10.21437/Interspeech.2007-44>
- Kang, J. G., Lee, S. H., Park, E. J., & Leem, H. S. (2016). Event-Related Potential Patterns Reflect Reversed Hemispheric Activity during Visual Attention Processing in Children with Dyslexia: A Preliminary Study. *Clinical Psychopharmacology and Neuroscience*, 14, 33-42. <https://doi.org/10.9758/cpn.2016.14.1.33>
- Karns, C. M., & Knight, R. T. (2009). *Intermodal Auditory, Visual, and Tactile Attention Modulates Early Stages of Neural Processing*. MIT Press. <https://doi.org/10.1162/jocn.2009.21037>
- Kaya, E. M., & Elhilali, M. (2016). Modelling Auditory Attention. *Philosophical Transactions of the Royal Society of London*, 372, Article ID: 20160101. <https://doi.org/10.1098/rstb.2016.0101>
- Keitel, C., & Müller, M. M. (2015). Audio-Visual Synchrony and Feature-Selective Attention Co-Amplify Early Visual Processing. *Experimental Brain Research*, 234, 1221-1231. <https://doi.org/10.1007/s00221-015-4392-8>
- Kumagai, T., & Mizuhara, H. (2016). Top-Down and Bottom-Up Attention Cause the Ventriloquism Effect with Distinct Electroencephalography Modulations. *Neuroreport*, 27, 647-651. <https://doi.org/10.1097/WNR.0000000000000590>
- Langner, R., Kellermann, T., Boers, F., Sturm, W., Willmes, K., & Eickhoff, S. B. (2012). Staying Responsive to the World: Modality-Specific and -Nonspecific Contributions to Speeded Auditory, Tactile, and Visual Stimulus Detection. *Human Brain Mapping*, 33, 398-418. <https://doi.org/10.1002/hbm.21220>
- Lee, Y. H., Jang, D. W., Kim, J. B., Park, R. H., & Park, H. M. (2020). Audio-Visual Speech Recognition Based on Dual Cross-Modality Attentions with the Transformer Model. *Applied Sciences*, 10, 7263. <https://doi.org/10.3390/app10207263>
- Leminen, A., Alho, K., Verwoert, M., Moisala, M., & Wikman, P. (2020). Modulation of Brain Activity by Selective Attention to Audiovisual Dialogues. *Frontiers in Neuroscience*, 14, Article No. 436. <https://doi.org/10.3389/fnins.2020.00436>
- Li, C., Kochiyama, T., Wu, J., Chui, D., Tsuge, T., & Osaka, K. (2007). Attention Systems and Neural Responses to Visual and Auditory Stimuli: An fMRI Study. In *The 2007 IEEE/ICME International Conference on Complex Medical Engineering* (pp. 1494-1498). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/ICCME.2007.4381995>
- Li, T., Lin, Y., Gao, Y., & Zhong, F. (2018). Longtime Driving Induced Cerebral Hemodynamic Elevation and Behavior Degradation as Assessed by fNIRS and a Voluntary Attention Test. *Journal of Biophotonics*, 11, e201800160. <https://doi.org/10.1002/jbio.201800160>
- Liu, T., & Mance, I. (2011). Constant Spread of Feature-Based Attention across the Visual Field. *Vision Research*, 51, 26-33. <https://doi.org/10.1016/j.visres.2010.09.023>
- Liu, T., Slotnick, S. D., Serences, J. T., & Yantis, S. (2003). Cortical Mechanisms of Feature-Based Attentional Control. *Cerebral Cortex*, 13, 1334-1343. <https://doi.org/10.1093/cercor/bhg080>
- Lu, Z., Li, Q., Gao, N., & Yang, J. (2020). Time-Varying Networks of ERPS in P300-Speller Paradigms Based on Spatially and Semantically Congruent Audiovisual Bimodality. *Journal of Neural Engineering*, 17, Article ID: 046015.

- <https://doi.org/10.1088/1741-2552/aba07f>
- Luks, T. L., Sun, F. T., Dale, C. L., Miller, W. L., & Simpson, G. V. (2008). Transient and Sustained Brain Activity during Anticipatory Visuospatial Attention. *Neuroreport*, *19*, 155-159. <https://doi.org/10.1097/WNR.0b013e3282f4a14f>
- Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J. et al. (2016). The Curious Incident of Attention in Multisensory Integration: Bottom-Up vs. Top-Down. *Multisensory Research*, *29*, 557-583. <https://doi.org/10.1163/22134808-00002528>
- Martínez, A.-V. et al. (1999). Involvement of Striate and Extra-Striate Visual Areas in Spatial Attention. *Nature Neuroscience*, *2*, 364-369. <https://doi.org/10.1038/7274>
- McCracken, H. S., Murphy, B. A., Burkitt, J. J., Glazebrook, C. M., & Yelder, P. C. (2020). Audiovisual Multisensory Processing in Young Adults with Attention-Deficit/Hyperactivity Disorder. *Multisensory Research*, *33*, 1-25. <https://doi.org/10.1163/22134808-20191472>
- McCracken, H. S., Murphy, B. A., Glazebrook, C. M., Burkitt, J. J., Karellas, A. M., & Yelder, P. C. (2019). Audiovisual Multisensory Integration and Evoked Potentials in Young Adults with and without Attention-Deficit/Hyperactivity Disorder. *Frontiers in human neuroscience*, *13*, 95. <https://doi.org/10.3389/fnhum.2019.00095>
- Mesgarani, N., & Chang, E. F. (2012). Selective Cortical Representation of Attended Speaker in Multi-Talker Speech Perception. *Nature*, *485*, 233-236. <https://doi.org/10.1038/nature11020>
- Mesulam, M. M. (2000). Attentional Networks, Confusional States, and Neglect Syndromes. In M. M. Mesulam (Ed.), *Principles of Behavioral and Cognitive Neurology* (pp. 174-256). Oxford University Press.
- Mirman, D., Landrigan, J. F., & Britt, A. E. (2017). Taxonomic and Thematic Semantic Systems. *Psychological Bulletin*, *14*, 499-520. <https://doi.org/10.1037/bul0000092>
- Mole, C. (2020). The Role of Attention in Multisensory Integration. *Multisensory Research*, *34*, 337-349. <https://doi.org/10.1163/22134808-bja10025>
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., & Foxe, J. J. (2002). Multisensory Auditory-Visual Interactions during Early Sensory Processing in Humans: A High-Density Electrical Mapping Study. *Cognitive Brain Research*, *14*, 115-128. [https://doi.org/10.1016/S0926-6410\(02\)00066-6](https://doi.org/10.1016/S0926-6410(02)00066-6)
- Ortega, R., López, V., Carrasco, X., Anllo-Vento, L., & Aboitiz, F. (2013). Exogenous Orienting of Visual-Spatial Attention in ADHD Children. *Brain Research*, *1493*, 68-79. <https://doi.org/10.1016/j.brainres.2012.11.036>
- O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., & Lalor, E. C. (2014). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, *25*, 1697-1706. <https://doi.org/10.1093/cercor/bht355>
- Petkov. (2004). Attentional Modulation of Human Auditory Cortex. *Nature Neuroscience*, *7*, 658-663. <https://doi.org/10.1038/nn1256>
- Plank, T. et al. (2012). Neural Correlates of Audio-Visual Object Recognition: Effects of Implicit Spatial Congruency. *Human Brain Mapping*, *33*, 797-811. <https://doi.org/10.1002/hbm.21254>
- 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. <https://doi.org/10.1111/j.1460-9568.2012.08060.x>
- Puschmann, S., Steinkamp, S., Gillich, I., Mirkovic, B., Debener, S., & Thiel, C. M. (2017). The Right Temporoparietal Junction Supports Speech Tracking during Selective Listening: Evidence from Concurrent EEG-fMRI. *Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*, 11505-11516. <https://doi.org/10.1523/JNEUROSCI.1007-17.2017>
- Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2016). The Neural and Computational Bases of Semantic Cognition. *Nature Reviews Neuroscience*, *18*, 42-55. <https://doi.org/10.1038/nrn.2016.150>
- Rapela, J., Gramann, K., Westerfield, M., Townsend, J., & Makeig, S. (2012). Brain Oscillations in Switching vs. Focusing Audio-Visual Attention. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 352-355). The Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/EMBC.2012.6345941>
- Rohe, T., & Noppeney, U. (2018). Reliability-Weighted Integration of Audiovisual Signals Can Be Modulated by Top-Down Control. *ENEURO*, *5*, e0315-0317. <https://doi.org/10.1523/ENEURO.0315-17.2018>
- Rohe, T., Ehlis, A. C., & Noppeney, U. (2019). The Neural Dynamics of Hierarchical Bayesian Causal Inference in Multisensory Perception. *Nature Communications*, *10*, 1907. <https://doi.org/10.1038/s41467-019-09664-2>
- Ruesseler, J., Gerth, I., Munte, T. et al. (2017). Audiovisual Speech Integration in the Superior Temporal Region Is Dysfunctional in Dyslexia. *Neuroscience*, *356*, 1-10.
- Santangelo, V., & Macaluso, E. (2012). Spatial Attention and Audiovisual Processing. In B. E. Stein (Ed.), *The New Handbook of Multisensory Processing* (pp. 359-370). MIT Press. <https://doi.org/10.7551/mitpress/8466.003.0031>
- Schoenfeld, M., Hopf, J.-M., Martinez, A., Mai, H., Sattler, C., Gasde, A., & Hillyard, S. (2007). Spatio-Temporal Analysis

- of Feature-Based Attention. *Cerebral Cortex*, 17, 2468-2477. <https://doi.org/10.1093/cercor/bhl154>
- Schutz, M., & Silverman, L. (2011). The Role of Attention in Audio-Visual Integration. *Canadian Acoustics*, 39, 108-109.
- Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., & Coslett, H. B. (2011). Neuroanatomical Dissociation for Taxonomic and Thematic Knowledge in the Human Brain. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 8520-8524. <https://doi.org/10.1073/pnas.1014935108>
- Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal Coherence and Attention in Auditory Scene Analysis. *Trends in Neurosciences*, 34, 114-123. <https://doi.org/10.1016/j.tins.2010.11.002>
- Stern, E. R., Wager, T. D., Egner, T., Hirsch, J., & Mangels, J. A. (2007). Preparatory Neural Activity Predicts Performance on a Conflict Task. *Brain Research*, 1176, 92-102. <https://doi.org/10.1016/j.brainres.2007.07.060>
- Talsma, D., & Woldorff, M. G. (2005). Selective Attention and Multisensory Integration: Multiple Phases of Effects on the Evoked Brain Activity. *Journal of Cognitive Neuroscience*, 17, 1098-1114. <https://doi.org/10.1162/0898929054475172>
- Tootell, R. B., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From Retinotopy to Recognition: fMRI in Human Visual Cortex. *Trends in Cognitive Sciences*, 2, 174-183. [https://doi.org/10.1016/S1364-6613\(98\)01171-1](https://doi.org/10.1016/S1364-6613(98)01171-1)
- Trenner, M., Heekeren, H., Bauer, M., Rössner, K., Wenzel, R., Villringer, A., & Fahle, M. (2008). What Happens in Between? Human Oscillatory Brain Activity Related to Crossmodal Spatial Cueing. *PLOS ONE*, 3, e1467. [https://doi.org/10.1016/S1364-6613\(98\)01171-1](https://doi.org/10.1016/S1364-6613(98)01171-1)
- Tsotsos, J. K. (1990). Analyzing Vision at the Complexity Level. *Behavioral and Brain Sciences*, 13, 423-445. <https://doi.org/10.1017/S0140525X00079577>
- Vroomen, J., & Keetels, M. (2010). Perception of Intersensory Synchrony: A Tutorial Review. *Attention Perception and Psychophysics*, 72, 871-884. <https://doi.org/10.3758/APP.72.4.871>
- Wang, J., Zhang, K., Madani, K., & Sabourin, C. (2015). Salient Environmental Sound Detection Framework for Machine Awareness. *Neurocomputing*, 152, 444-454. <https://doi.org/10.1016/j.neucom.2014.09.046>
- Wang, Z., Chen, M., Goerlich, K. S., Aleman, A., Xu, P., & Luo, Y. (2021). Deficient Auditory Emotion Processing but Intact Emotional Multisensory Integration in Alexithymia. *Psychophysiology*, 58, e13806. <https://doi.org/10.1111/psyp.13806>
- Weinberger, N. M. (2011). Reconceptualizing the Primary Auditory Cortex: Learning, Memory and Specific Plasticity. In J. A. Winer, & C. E. Schreiner (Eds.), *The Auditory Cortex* (pp. 465-491). Springer. https://doi.org/10.1007/978-1-4419-0074-6_22
- Werner, S., & Noppeney, U. (2011). The Contributions of Transient and Sustained Response Codes to Audiovisual Integration. *Cerebral Cortex*, 21, 920-931. <https://doi.org/10.1093/cercor/bhq161>
- Xi, Y., Li, C., Gao, N., He, S., & Tang, X. (2019). Cortical Network Underlying Audiovisual Semantic Integration and Modulation of Attention: An fMRI and Graph-Based Study. *PLOS ONE*, 14, e0221185. <https://doi.org/10.1371/journal.pone.0221185>
- Xi, Y., Li, Q., Gao, N., Li, G., Lin, W., & Wu, J. (2020). Co-Stimulation-Removed Audiovisual Semantic Integration and Modulation of Attention: An Event-Related Potential Study. *International Journal of Psychophysiology*, 151, 7-17. <https://doi.org/10.1016/j.ijpsycho.2020.02.009>
- Xi, Y., Zhang, M., Gao, N., Li, Y., Liu, L., & Li, Q. (2019). The Neural Mechanism of Audiovisual Integration Modulated by Attention: A Dcm Study. In J.-S. Pan, A. Ito, P.-W. Tsai, & L. C. Jain (Eds.), *Recent Advances in Intelligent Information Hiding and Multimedia Signal Processing* (pp. 162-170). Springer International Publishing. https://doi.org/10.1007/978-3-030-03745-1_20
- Zhao, S., Li, Y., Wang, C., Feng, C., & Feng, W. (2021). Updating the Dual-Mechanism Model for Cross-Sensory Attentional Spreading: The Influence of Space-Based Visual Selective Attention. *Human Brain Mapping*, 42, 6038-6052. <https://doi.org/10.1002/hbm.25668>