

多媒体学习中图文加工的认知神经机制*

李松清 赵庆柏 周冶金 张 依

(青少年网络心理与行为教育部重点实验室(华中师范大学);
人的发展与心理健康湖北省重点实验室; 华中师范大学心理学院, 武汉 430079)

摘 要 多媒体学习的认知理论和图文理解整合模型指出, 多媒体学习中图片和文字在早期阶段的加工遵循双通道假设, 在晚期不同表征的信息进行整合完成知识建构。理论上的探讨得到了神经科学实证研究的支持。脑成像研究发现, 图、文加工在早期的前语义阶段存在差异, 但晚期的信息整合阶段共享了同一语义系统。事件相关电位研究也发现在早期图片诱发了独特的 N300 成分, 而晚期图、文均诱发了 N400 效应。尽管现有研究已经明确了图、文共享同一语义系统, 但是信息是如何在这一系统中进行整合的还需要进一步的研究。

关键词 多媒体学习; 语义加工; 图文整合; 脑成像; 事件相关电位

分类号 B849:G44

1 引言

伴随着计算机和视频技术的进步, 高效而丰富多彩的多媒体被广泛关注。相对于单一媒体, 多媒体是将文字、图形、声音、图片、动画等多种信息媒介进行有机整合并加以处理的综合媒体(王以宁, 王永锋, 孔得伟, 2005)。简言之, 多媒体就是用词语和画面来共同呈现材料(理查德·梅耶, 2006)。词语是以文本或言语形式呈现材料; 画面是以图片形式呈现材料, 如静态的图形或动态的图画。这个定义可以囊括多媒体的多种整合情形。多媒体技术的发展实现了词语和画面信息的集成, 使很多信息可以以视觉和听觉的形式共同呈现, 从而实现信息的多感觉通道的统一获取、存储、加工以及整合。

多媒体信息的集成性使得信息的呈现方式更加形象、生动、丰富, 相比传统的学习方式更能吸引学习者的注意力, 激发他们的学习动机和兴趣。研究表明, 多媒体环境下的图文整合有效促

进了学习者的学习成绩(Eitel, Scheiter, & Schüler, 2013)。所以, 以计算机和视频技术为基础的多媒体被教育者广泛使用, 以帮助学习者构建心理表征和理解学习材料(Mayer & Moreno, 2002)。基于上述多媒体的定义与理解, 多媒体学习可以被认为是学习者同时使用视觉通道模式和听觉通道模式加工和处理信息的过程(胡卫星, 2012)。在教育心理学和心理学研究中, 多媒体学习可进一步从两个角度加以分析: 一是信息的呈现模式角度, 主要关注信息的表征方式, 如文字或图片; 二是感觉通道角度, 它主要强调学习者用两种或两种以上的信息加工通道, 如听觉通道和视觉通道(Mayer, 2005)。

目前多媒体学习的研究主要集中于三个方面: 多媒体学习过程、学习规律和学习原则等的基本理论研究; 多媒体学习的相关理论和原则在学科教学中的具体运用与实践; 以及多媒体学习自身和在教学运用中的评价。其中, 多媒体学习的基本理论一直是研究的重点, 它是对于多媒体学习过程的模型建构, 关注的是不同呈现形式或不同感觉通道的信息的接收和建构, 是指导多媒体学习, 提高多媒体教学效果的基础(刘世清, 2013)。

2 多媒体学习中图文加工的认知理论

从上世纪 80 年代开始, 计算机多媒体技术迅

收稿日期: 2014-12-29

* 国家自然科学基金青年项目(31400875); 国家自然科学基金面上项目(31471000); 教育部人文社会科学研究规划基金项目(14YJA190014); 中央高校基本科研业务费(CCNU15A02041)。

通讯作者: 赵庆柏, E-mail: zqbznr@mail.ccnu.edu.cn

速发展并应用于教学。使用图形、动画等视觉形式呈现学习材料变得越来越普遍。研究发现,包含对应文字说明的图解能够有效地促进学习者对科学解释的理解,但当时还缺少心理学理论来对这种新的学习方式进行解释(王英豪, 2009)。

随后, Mayer 和 Moreno (2002)在双通道假设、有限容量假设以及主动加工假设基础上建立了多媒体学习认知理论,重点关注个体通过语词和图片进行学习的过程。该理论认为多媒体学习包含三个阶段、五个认知加工过程,遵循先分后总的加工模式:第一阶段是感觉注意阶段,从呈现的文本或解说中选择相关语词,从呈现的插图或视频中选择相关图片;第二阶段是工作记忆中的信息组织阶段,将所选择的语词进行组织形成连贯的言语表征,将所选择的图片进行组织以形成连贯的视觉表征;第三阶段是信息整合阶段,将语言表征和视觉表征信息与学习者的先前知识主动进行整合(Mayer, 2010)。多媒体学习认知理论不仅强调学习者在认知加工过程中对不同表征方式信息的加工,同时还关注个体对语词和图片信息的有意识的主动选择。因此,该理论不仅有利于教育者系统的认识学习者的多媒体学习过程,而且对其恰当呈现学习材料具有重要的实践指导意义。

然而,多媒体学习认知理论仅指出了言语表征和视觉表征信息以及先前知识之间的整合,缺乏对于个体是如何对不同表征信息进行加工以及如何对语词和图片信息进行整合的机制的概括和解释(文东, 王玉琴, 吴秀园, 2013)。针对这一问题, Schnotz 和 Bannert (2003)提出了面向多媒体学习的图文理解整合模型。该模型中存在两种形式的表征,一种是基于符号进行处理的描述性表征,如文字表征和命题表征;另一种是基于过程的结构进行处理的描绘性表征,如视觉表象和心理模型。在图文理解整合模型中,首先从感觉记忆中选择语词和图片进入工作记忆,并分别组成相应的文字表征和视觉表象;然后对文字表征进行语义加工,在长时记忆中的认知图式辅助下,结合视觉表象建立命题表征;同时对视觉表象进行主题选择,在长时记忆中认知图示的辅助下,结合文字表征建构心理模型;心理模型与命题表征进行交互,形成整合的知识信息(Schnotz & Kürschner, 2008; Hochpöchler et al., 2013)。在该模型中,整合发生在两个阶段,一是语词和图片信息的整合以

形成命题表征和心理模型,二是命题表征和心理模型的交互整合以完成知识信息的整体建构。

尽管多媒体学习的认知理论和图文理解整合模型强调的重点不同,但两种理论都认为多媒体学习早期阶段图、文信息通过不同的通道进行获取和加工,晚期阶段完成不同表征信息和先前知识的整合。然而,先前对于多媒体学习过程中图文加工过程的研究多是基于理论的探讨。借助认知神经科学的力量,对多媒体学习中图文信息整合的内部心理机制进行深入剖析,将成为多媒体学习认知理论未来关注的重点(郑旭东, 吴秀圆, 王美倩, 2013)。

3 多媒体学习中图文加工的神经机制

个体在多媒体学习过程中,需要整合图文信息以完成知识建构。根据多媒体学习的认知理论和图文理解整合模型,图片和文字在早期加工阶段的认知机制存在差异,而在后期的语义加工阶段进行整合。神经成像和电生理技术的发展使直接观察多媒体学习过程中大脑信息加工成为可能。高空间分辨率的正电子发射断层扫描(Positive Emission Tomography, PET)和功能核磁共振成像(Functional Magnetic Resonance Imaging, fMRI)技术可以帮助我们定位多媒体学习过程中图文信息加工和整合的相关脑区,高时间分辨率的脑电图(Electroencephalograph, EEG)和事件相关电位技术(Event-related Potential, ERP)有利于我们区分多媒体学习的认知加工阶段。

然而,现实的多媒体学习很难进行直接的神经成像研究。原因有三:首先,多媒体学习往往是在语篇水平上,包含了信息识别、选择、整合和记忆等多个认知过程,这些认知过程往往在时间上存在交叉,很难实现具体子过程的独立分割;其次,从脑成像技术上考虑,可靠的分析需要通过相当数量的、可重复观察的同类心理事件进行,而多媒体学习材料比较复杂,很难保证材料的同质性,也正是因此行为研究多采用少数的几个案例来进行多媒体学习研究,如“闪电的形成过程与原理”(段朝辉, 颜志强, 王福兴, 周宗奎, 2013; 龚少英, 段婷, 王福兴, 周宗奎, 卢春晓, 2014);最后,多媒体学习受到学习者知识经验、认知风格、学习动机和情绪等多种因素影响(Hede, 2002; 康诚, 周爱保, 2010),这也增大了对多媒

体学习进行直接神经成像研究的难度。因此,目前多媒体学习的研究主要采用行为研究方法,即便是多媒体学习研究权威 Richard E. Mayer 也较少进行神经成像研究,仅仅使用 fMRI 研究了人们采用不同学习策略时的脑区差异(Miller, Donovan, Bennett, Aminoff, & Mayer, 2012)。

尽管如此,依然可以通过实验室研究对多媒体学习过程进行概括和简化,对其核心的图文加工的神经机制进行探讨。正如物理学中,不管机械运动多么复杂,总脱离不开牛顿三大定律。同样,不管多媒体学习多么复杂,其核心过程都是图文信息加工与整合。认知神经科学关于图文加工的基础研究,为了解多媒体学习的认知神经机制提供了窗口。目前对图文加工的神经机制研究主要围绕图文信息语义加工系统争论和图文加工过程比较两方面展开。

3.1 图文信息语义加工系统争论

认知心理学与神经心理学对图文信息语义加工系统的争论持续已久(Bright, Moss, & Tyler, 2004; Devereux, Clarke, Marouchos, & Tyler, 2013)。研究多在概念表征水平展开,争论的焦点在于图、文加工是否共享同一语义系统。

早期对脑损伤病人的研究支持了图、文加工的多重语义加工系统观点,认为言语(文字)和视觉(图片)输入形态下的信息加工具有不同的概念表征系统,大脑对不同输入方式信息的加工和存储的对象是不同的(Warrington & Shallice, 1984)。Warrington (1975)对两位视觉失认病人的研究发现,其中一位患者可以识别视觉呈现的物体,但是不能对其命名;另一位患者可以再认物体的名称,但没有视觉表征。视觉失认症具有的不同表现形式,支持了视觉和言语系统的独立性。随后一项对进行性非流利性失语患者的研究(Saffran, Coslett, Martin, & Boronat, 2003)发现,病人在一系列的语义任务中,对词汇的识别加工明显不如图片刺激。研究者认为这一现象支持了分布式的语义系统,知识分布在不同的独立的子系统中,因此不同脑区的损伤导致不同的语义行为缺陷。此外,一系列基于语义判断的图片与词汇的分类关系实验结果也都支持了具体通道的语义信息表征独立性(Warrington & McCarthy, 1994)。

然而,上述研究其实并不能完全的说明图、文加工具有多重语义加工系统,因为脑损伤病人

在图、文任务中的差异也可能是发生在图、文识别的前语义加工阶段。脑成像的研究支持了这一说法,发现因为图片和文字刺激形式的不同,图文早期的前语义加工阶段激活的脑区存在差异。Vandenberghe, Price, Wise, Josephs 和 Frackowiak (1996)利用 PET 技术比较了健康被试在三种图文概念匹配任务(语义关联匹配;视觉语义匹配;物理大小匹配)下的图文加工过程,探讨前语义加工过程中图片和文字语义加工过程的独特性。所有图片刺激条件共同激活的脑区被认为图片刺激的前语义加工脑区,所有文字刺激条件共同激活的脑区被认为文字刺激的前语义加工激活脑区。结果显示,图片刺激的前语义加工激活右侧枕中回,而文字刺激的前语义加工激活左侧顶下叶。Moore 和 Price (1999)在概念水平的图片和文字命名和浏览(命名时不出声;浏览时默念“Okay”)任务的 PET 研究同样支持了图片和文字的前语义加工激活脑区的差异。研究发现,在语义加工过程中文字刺激更多地激活了与语音加工有关的左侧颞上回(BA 22/41) (包括缘上回);而图片刺激更多的激活了与物体识别有关的腹侧颞枕皮质(BA 19) (包括外侧枕叶皮质)。由于外侧枕叶皮质与表征的概念水平加工无关(Grill-Spector, Kourtzi, & Kanwisher, 2001),因此研究者认为这些不同的激活脑区或与前语义加工阶段相关。

尽管图片和文字在早期前语义加工阶段存在差异,但在后期言语和视觉输入形态下的语义加工都使用共同的概念表征系统,系统中的信息以共享的形式进行存储和加工,并不会追溯到其输入形式(Federmeier & Kutas, 2001)。Bright 等(2004)采用词汇判断任务及语义类别判断任务,探究在语义表征水平,图片和文字的加工是否具有通道独特性。通过对四类语义任务的 PET 数据进行了联合元分析,发现图片与文字输入形态下都强烈激活的区域主要是左半球的额下回、颞上回、颞中回、梭状回以及海马旁回。与具体通道相关的脑区中,文字激活左侧颞上回、海马旁回以及右侧的颞上回;图片加工激活右侧枕上回和双侧梭状回,其中枕上回的激活与图片加工相关,反映视觉加工过程的前语义阶段。研究支持了图文语义加工系统的一致性。Shinkareva, Malave, Mason, Mitchell 和 Just (2011)采用了 fMRI 技术比较了不同呈现方式刺激的语义类别解码脑区。结果发现,

图文概念的加工具有明显的左半球优势,图文共同激活了左侧颞上回、颞中回、颞下回、梭状回、顶下叶、缘上回以及枕下回和舌回等区域。文字独特激活的脑区主要是言语及负责执行功能的脑区,图片独特激活的脑区主要是视空间及视觉功能区域。结果同样支持了图、文前语义加工阶段存在差异,但图片和文字在语义加工阶段共享同一语义系统。

然而, van Doren, Dupont, Grauwe, Peeters 和 Vandenberghe (2010)认为在比较同一信息的图片与文字呈现形式的语义加工激活脑区时,不仅要排除前语义加工过程激活脑区的差异,还应排除通达概念表征系统时的语义执行控制所诱发的脑活动,这样才能明确探究图片与文字的语义概念系统是否一致。他们将图片与文字刺激的呈现设置在被试的识别阈限水平,探究当个体在刚刚识别不同通道输入刺激时的大脑加工差异,以排除语义执行需求对刺激加工活动的影响。当被试能够刚刚识别刺激时,图片与文字刺激共同激活的区域主要有左半球的颞枕沟(包括梭状回)、顶内沟、前额连接以及额下沟中间的三分之一(左侧额下回)。相比先前有关图文语义加工共同激活的脑区,该研究在前额叶区域的激活很少,说明额叶与语义执行控制需求有关。在将语义执行控制降至最低时,图文的语义加工仍有大量的脑区重叠,包括左侧额下回、颞枕沟和梭状回,结果仍支持图文语义具有单一的语义加工系统。

除了概念表征水平,也有研究比较了命题和语篇水平的图文加工激活脑区,从不同水平为图文具有共享的语义加工系统提供支持。Gates 和 Yoon (2005)在字母、词汇以及句子水平上比较了图片与文字理解过程中的语义加工脑活动。在句子理解和图片(与句子具有相同的语义信息)语义加工过程中发现了大脑中的分离区域。句子理解独特的激活了左右舌脑回的部分区域,左侧梭状回,左右枕下回,右侧楔叶和左侧枕叶脑回;虽然图片语义的知觉刺激独特激活也涉及舌回、梭状回和楔叶,但是具体区域与句子理解并无重叠。在句子理解以及图片知觉中颞枕联合的 80% 区域作为图文加工共享区域被激活,在单个物体的图片和对应文字的知觉中颞枕联合区共享区域占 96%。这些结果说明,图片语义加工与句子理解之间的共享脑区会随着心智表征的复杂性增加

而减少。Li, Zhong 和 Lu (2011)通过眼动和 fMRI 技术探究了语篇水平上图片与文本的语义加工是否传达相同的信息含义以及是否具有共享的神经系统。眼动研究结果发现,图片与文本两种形式的信息呈现方式在阅读效率上没有显著差异。fMRI 数据的联合分析显示,图片与文本在语义加工过程中存在很多共享的加工网络:额叶、顶叶、颞叶和边缘系统,眼动及神经机制研究的结果都支持图片和文本加工共享同一语义系统。尽管对图文语义加工系统争论的研究在命题水平、语篇水平的研究并不多,但其研究结果与概念表征水平的研究结果一致,都支持图文共享语义加工系统的观点。

上述对图文语义加工的研究多是对语义系统脑区进行识别,主要关注的是共享的语义加工激活脑区,而非脑区的具体运算表现。Devereux 等 (2013)在以往研究基础上,利用多体素模式分析法对图、文分类任务中语义理解过程激活的脑区进行分析,语义类别效应显示了图、文语义加工的左半球效应,图片和文字的语义加工在左半球存在较多的重叠,主要集中在左后侧颞中回、左侧角回及左侧顶内沟。尽管左后侧颞中回和左侧顶内沟都参与了语义加工,但在跨区域和跨通道的表征内容差异性分析中,只有左侧顶内沟在不同的视觉输入形式下的神经活动没有显著差异,而左后侧颞中回在加工文字与物体时的神经活动并不相同,说明不同语义表征形式的信息在同一脑区的加工计算模型可能不同。Devereux 等人的研究支持了图文理解整合模型,即多媒体学习过程中图文信息的加工并非形成单一模式的知识信息,而是同时构建命题表征和心理模型,二者虽然表征的是同一知识信息,但计算模型并不相同。

总结已有的研究,无论呈现的方式,图文具有单一的语义表征激活,但是系统中某些部分的激活是特定于刺激的输入类型的(Plaut, 2002)。具体表现在(如图 1 所示),图片加工的前语义阶段激活与物体识别有关的右外侧枕叶,文字刺激的前语义加工阶段主要激活与语音加工和语义存储有关的左侧顶下叶、颞上回(包括缘上回)和顶内沟;而后期的语义加工过程图片和文字共同激活的脑区主要包括左侧额下回、颞叶皮质、梭状回,具有明显的左半球优势效应。结果基本上支持了多媒体学习的认知理论以及图文理解整合模型对图

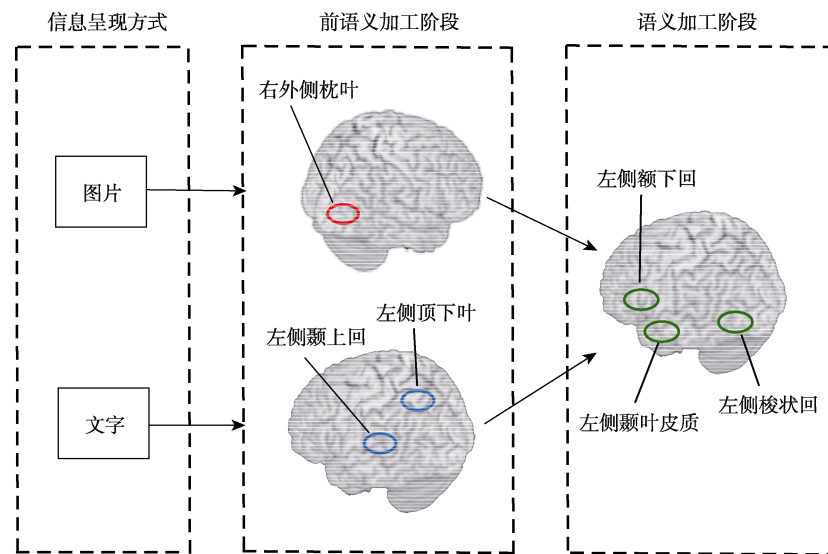


图 1 图、文加工过程及脑区示意图

片和文字的语义加工过程的阐述。

3.2 图、文语义加工过程比较

PET 和 fMRI 技术从空间脑区定位的角度对图文语义加工系统进行了探讨,但由于 PET 和 fMRI 技术时间分辨率较低,对于激活脑区参与的具体加工阶段大多是基于理论的推测,并无明确的证据。ERP 技术以其较高的时间分辨率为了解图文加工的动态过程提供了新的视角。

N400 成分是言语研究中最常见的成分 (Stafura & Perfetti, 2014), 为揭示不同呈现方式信息的语义加工的神经机制提供了有效指标。Nigam, Hoffman 和 Simons (1992) 在命题水平上通过操纵句尾图片或文字与句子的语义一致性, 比较了图片与文字在句意理解过程中诱发的 N400 成分。无论图片还是文字条件, 正常句意加工不会诱发 N400 成分, 但语义违反会诱发出一个清晰明显的 N400 成分, 且此成分在文字与图片刺激之间没有显著差异。说明 N400 与语义的表征形式无关, 反映的是深层的语义加工机制。N400 不仅仅是对言语系统的反映, 而是对图片和文字共用的概念系统反映。进一步, Khateb, Pegna, Landis, Mouthon 和 Annoni (2010) 通过分析 3 种匹配任务 (语义任务; 语音判断任务; 图片类别判断任务) 诱发的 N400 成分特点, 发现 N400 效应不仅独立于信息表征方式, 而且具有跨任务类型的一致性。

尽管图文诱发的 N400 的潜伏期与波幅没有差异 (Willems, Özyürek, & Hagoort, 2008), 但图片刺激条件下的语义加工与文字刺激的加工并不同, 图片刺激的加工会诱发出一个比 N400 更靠前的独特成分—N300。在图片刺激的相关性判断任务中, 不相关比相关任务在额区有一个 300ms 左右的峰值 (N300), 随后出现一个分布更加广泛的 N400 (Holcomb & McPherson, 1994), 在单纯的言语刺激相关性任务中并没有观察到 N300 的出现, 认为 N300 是对图片刺激独特的反映。West 和 Holcomb (2002) 通过控制图片刺激的复杂性, 降低对图片刺激自主命名的加工过程, 发现额区分布的 N300 成分及更广泛分布的 N400 成分都有出现, 说明对于图片刺激而言, N300 的出现并不是由于言语操作引起的, 而直接反映了图片语义信息的加工机制。由于 N300 出现在 N400 之前, 因此 N300 反映的可能是知觉输入图片与长时记忆中已存储的图片表征的匹配加工 (Schendan & Maher, 2009), 而 N400 成分与跨通道的语义信息整合相关 (Wu & Coulson, 2011)。也有研究者在词汇加工过程中发现类 N300 效应。Cristescu 和 Nobre (2008) 发现具体词汇如工具或动物的词汇也会诱发一个广泛分布的 N300 成分, 该成分可能反映的是具体词汇的视觉表象加工, 与视知觉存在认知系统重叠, 结果仍然支持 N300 成分对图片刺激加工的独特性。

ERP研究从时间进程上揭示了图片和文字的早期加工阶段确实存在差异,图片刺激独特的诱发了N300成分,但后期的语义加工阶段图片和文字均会诱发N400成分。N400成分的波幅及潜伏期的相似性预示着它是对图片和文字系统的共同活动的反映,支持了图、文共享同一语义加工系统的观点。

3.3 图、文语义整合的神经机制

在多媒体学习环境下,图片和文字常常同时呈现,需要学习者对图片和文字的语义信息进行整合。多媒体学习的认知理论指出,图、文信息经过不同的通路进行加工,并最终与先前知识进行整合;而图文理解整合模型进一步提出,在长时记忆中认知图示的辅助下图、文信息整合最终形成相应的命题表征和心理模型。然而,图片和文字信息在大脑中整合的神经机制并不清楚,目前尚没有研究对这一问题进行直接探讨。对图、文语义加工系统的争论及图、文语义加工过程的探讨,结果支持了图片和文字具有共同的语义加工系统,因此可以推测,不同的表征形式的信息整合也发生在共享的语义加工系统中。本文试图从图、文共享的语义系统出发,结合以往语义整合的相关研究,分析图文语义整合的关键脑区。

语言作为一种高级的认知加工活动,涉及额下皮层、颞上回、颞中回和顶叶等大范围区域,这些脑区在语言加工过程中具有不同的功能,共同构成了语言加工的神经网络(Vigneau et al., 2006)。朱祖德、王穗苹、冯刚毅和刘颖(2011)通过分析42篇命题水平的语义整合相关文献,发现语义整合过程主要激活了额下回、颞中回、颞上回后部、颞上回前部等脑区。Huang, Wang, Jia 和 Chen (2012)利用ERP结合事件相关光信号(Event-related Optical Signals)技术,发现句子语义整合中激活的大脑关键区域包括左侧额下回及颞叶前部。通过对图、文语义系统争论研究的总结,发现图片和文字在语义加工过程中共同激活的脑区主要包括左侧额下回、颞叶皮质和梭状回。

3.3.1 梭状回

左侧梭状回对正词法规则具有加工偏好,因此被称为“视觉词形区”(Reinke, Fernandes, Schwindt, O'Craven, & Grady, 2008; Yeatman, Rauschecker, & Wandell, 2013)。然而,有研究对梭状回对文字刺激加工的机理提出了质疑。Price和Devlin (2003)

通过综述相关文献,发现左侧梭状回在非文字实验条件下也能产生激活,如物体知觉、图片加工等,而不仅仅只针对视觉呈现的文字刺激。在实证研究中,Price (2010)同样发现左侧梭状回中部不仅只在视觉词形加工任务中激活,还在颜色和图片命名、盲文阅读和重复听到的单词时也存在显著激活,认为左侧梭状回中部对多种模态文字刺激均具有响应。Kassuba等人(2011)进一步研究发现在对物体的多重感觉信息加工过程中,左侧梭状回具有统一多重通道信息对物体进行表征的关键作用。

目前,左侧梭状回在文字和物体加工中的具体作用仍无定论(Barton, Fox, Sekunova, & Iaria, 2010)。不过从现有文献来看,左侧梭状回对文字并没有特异性加工。梭状回对所有刺激都具有一般外部特征加工,实现把刺激的视觉综合信息向更高级加工中枢(如语音、语义中枢)传递的功能(任静, 2010)。因此,在多媒体学习中,图、文信息加工中梭状回的激活反映的是对视觉刺激特征的一般性加工,并不涉及语义整合。

3.3.2 左侧额下回与颞叶皮质

图、文共同激活的脑区除了与视觉刺激特征加工有关的梭状回,还包括左侧额下回和颞叶皮层,二者均被认为与语义整合有关。

左侧颞叶皮层是典型的语言加工脑区,与语义信息通达和提取密切相关(Lau, Phillips, & Poeppel, 2008)。Baron和Osherson (2011)以及Baron, Thompson-Schill, Weber和Osherson (2010)在复杂概念整合过程的研究中观察到了左侧颞叶前部的激活,但左侧额下回未见激活。研究者认为颞叶前部储存着一般语义关系,对概念间的整合具有重要作用,而左侧额下回一般在比较复杂的语义表征整合时才会被激活。Peelen和Caramazza (2012)以及Hoffman, Evans和Lambon Ralph (2014)通过fMRI实验也发现,颞叶前部储存着所有物体的属性概念表征,这种表征将具体通道的属性信息抽象出来。结果同样支持了左侧颞叶前部区域储存着语义概念,其主要功能是将不同的信息源整合成新颖概念。因此,左侧颞叶前部被认为是语义整合的关键脑区(Lau et al., 2008),负责了捆绑分布式网络间的概念特征以及言语中的整合操作(Westerlund & Pykkänen, 2014),而左侧额下回只是负责执行认知控制功能。

尽管左侧额下回在加工语义过程中是负责语义整合功能还是伴随任务的认知控制尚无定论,但近年来相关研究却为左侧额下回在语义整合过程中的作用提供了支持。Zhu 等人(2012)以及 Huang, Zhu 等(2012)使用 ERP 和 fMRI 技术探究在内隐与外显的语义句子理解过程中的语义整合过程,联合分析结果发现语义整合过程激活的脑区主要在左侧额下回的前部。跨任务的比较显示,外显比内隐任务在双侧前扣带回/背外侧前额叶皮层,左侧额中回以及右侧额下回显示了更加广泛的激活,作者认为这些激活脑区是由于语义加工的一般过程所诱发的,如认知控制。随后 Zhu, Zhang, Li, Li 和 Wang (2013)利用 fMRI 技术对此进行了更加深入的探究,他们让被试完成内隐与外显的语义判断任务,同时增加一项与认知控制相关的 stroop 任务。对内隐和外显任务的联合分析发现,在左侧额下回有两个区域与语义整合的相关:一个是左侧额下回前部,一个是左侧额下回后部;而语义判断任务与 stroop 任务在左侧额下回后部具有更多的激活重叠。该结果支持左侧额下回后部对于一般领域的认知控制比较重要,而左侧额下回前部对于语义整合过程具有重要的作用。

王穗苹和黄健(2012)总结语义整合的相关文献,认为左侧额下回和颞叶皮层都参与了语义整合,只不过参与的整合过程不同。因为语义整合加工的研究多使用语义违反范式,语义正常句通常是自动进行加工的,非常迅速;语义违反句需要更多注意加工资源,因此整合的时间会比较晚。有研究发现,当采用合理的短语跟假词短语进行对比,左侧颞叶前部会在很早的时间内出现持续时间很短的激活,约发生在 184 至 255 ms 这短短的 70 ms 左右(Bemis & Pykkänen, 2011);而语义违反句子跟正常句子的对比可以观察到左侧额下回出现激活,但这一区域激活的时间发生在 N400 时窗内(Hagoort, Hald, Bastiaansen, & Petersson, 2004)。由于颞叶前部的激活与正常句的语义加工关系更为密切,因此较早的被激活;而左侧额下回的激活与语义违反的句子加工有关,其激活时间比颞叶前部晚(王穗苹,黄健,2012)。

不过,左侧额下回和颞叶皮层在语义整合作用的主要差异并非时间进程上的差异,而是体现在不同类型的语义整合上。Hagoort, Baggio 和

Willems (2009)将语义整合的形式分为两类,一类是小的信息块组合成已有的记忆表征,由于该记忆表征已经存在,因此实际上该过程是再认过程,主要由颞叶负责;另一类是将不同的信息结合在一起,形成了一个新的、原有记忆中不存在的意义表征,主要由额叶负责。由于在第二类的语义整合中,左侧额上回和颞中回也会激活(Willems, Ozyürek, & Hagoort, 2007; Willems et al., 2008),因此研究者认为左侧额下回和左侧额中回/颞上回的交互作用共同调节语义整合过程。

根据多媒体学习的认知理论,多媒体学习同时包含了两类语义整合,首先图、文信息经过不同的通道分别形成视觉表征和言语表征,这个过程涉及片段语义信息的组合,因此激活了左侧颞叶皮层;随后视觉表征和言语表征与先前知识进行整合形成新的、原有记忆中不存在的意义表征,因此激活了左侧额下回。由于学习的本质在于新知识的获得,因此左侧额下回在多媒体学习中的作用更为重要。

4 总结与展望

多媒体学习中,文字和图片的共同呈现实现了信息的多通道统一获取、存储、加工以及整合。多媒体学习的认知理论认为,文字和图片在早期阶段的加工遵循双通道假设,分别形成言语表征和视觉表征,并最终与先前知识进行整合。图文理解整合模型进一步指出,图文整合的结果是在长时记忆中认知图示的辅助下形成相应的命题表征和心理模型。理论上的探讨得到了神经科学实证研究的支持。脑成像研究发现,图、文加工在前语义阶段存在差异,图片激活了与物体识别有关的右外侧枕叶,文字则激活了与语音及拼字法有关的左侧颞顶皮层;图片和文字共同激活了与语义信息整合有关的左侧额下回和颞叶皮层,说明图、文在信息整合阶段共享了同一语义系统。事件相关电位研究也发现了一致的结果,在信息加工早期图片独特的诱发了一个 N300 成分,而晚期图片和文字均诱发了 N400 效应。

尽管现有研究已经明确了图、文共享同一语义系统,但是信息是如何在这一系统中进行整合的还需要进一步的研究。目前图文语义整合相关的文献多是使用眼动追踪技术的行为研究(Richardson & Matlock, 2007; van Genuchten,

Scheiter, & Schüler, 2012), 对图文整合的神经机制方面的研究还不足。而对于语义整合的研究则多采用文字材料, 发现在语义加工过程中颞叶前部负责语义的加工和提取, 而同时左侧额叶负责语义加工整合过程中的调控。这一结论是否能够推广到图文信息整合中, 尚不明确。今后的研究可采用眼动追踪和 fMRI 联动的方式同时记录行为和神经活动, 眼动数据为 fMRI 数据的分析提供精准的图文整合的时间信息, 进而准确地锁定图文整合的关键脑区。

共享同一语义系统并不代表图文整合的神经加工模式是单一的。已有研究证实, 图文共同的语义加工系统对两种表征方式信息的加工计算模型是不同的(Devereux et al., 2013)。根据图文理解整合模型, 多媒体学习中的图文整合包含视觉表象辅助文字表征形成命题表征, 文字表征辅助视觉表象形成心理模型, 以及命题表征和心理模型进行整合完成知识建构, 这些不同类型整合的神经加工模式应该是不同的。在同一语义系统中, 三者的差别可能体现在相关脑区的不同协同模式上。大脑连通性分析方法为揭示脑区之间的协同模式提供了有力的工具(Friston, Frith, & Frackowiak, 1993)。

在语义加工过程的研究中, 通常采用 N400 作为语义加工整合的指标。但 N400 是在语义违反句式中诱发, 是一个人造的成分, 并不能反映正常的语义整合过程。今后的研究可考虑将晚期正成分(Late Positive Component, LPC)作为语义整合的指标, 因为它可能更能反映正常句子中的语义整合(Spotorno, Cheylus, van Der Henst, & Noveck, 2013)。

参考文献

- 段朝辉, 颜志强, 王福兴, 周宗奎. (2013). 动画呈现速度对多媒体学习效果影响的眼动研究. *心理发展与教育*, 29(1), 46–53.
- 龚少英, 段婷, 王福兴, 周宗奎, 卢春晓. (2014). 装饰图片影响多媒体学习的眼动研究. *心理发展与教育*, 30(4), 403–410.
- 胡卫星. (2012). *动画情境下多媒体学习的实验研究* (博士学位论文). 辽宁师范大学, 大连.
- 康诚, 周爱保. (2010). 信息呈现方式与认知风格对多媒体环境下学习效果的影响. *心理科学*, (6), 1397–1400.
- 刘世清. (2013). 多媒体学习与研究的基本问题——中美学者的对话. *教育研究*, (4), 113–117.
- [美]理查德·梅耶著, 牛勇, 邱香译. (2006). *多媒体学习*. 北京: 商务印书馆:3.
- 任静. (2010). *文字和物体图形视觉加工脑成像研究* (硕士学位论文). 广州大学.
- 王穗苹, 黄健. (2012). 句子理解中的语义加工回路——研究热点与未来的展望. *华南师范大学学报: 社会科学版*, (6), 47–55.
- 王以宁, 王永锋, 孔得伟. (2005). 多媒体学习中的认知心理学因素考察——来自梅耶 (Mayer) 的研究和实践. *开放教育研究*, 11(3), 34–37.
- 王英豪. (2009). *梅耶多媒体学习研究探析* (硕士学位论文). 华东师范大学, 上海.
- 文东, 王玉琴, 吴秀园. (2013). 基于认知神经科学视角的多媒体学习认知理论创新. *现代远程教育研究*, (3), 40–49.
- 郑旭东, 吴秀圆, 王美倩. (2013). 多媒体学习研究的未来: 基础, 挑战与趋势. *现代远程教育研究*, (6), 17–23.
- 朱祖德, 王穗苹, 冯刚毅, 刘颖. (2011). 左侧额下回在句子语义整合加工中的作用. *心理科学进展*, 19(8), 1147–1157.
- Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *NeuroImage*, 55(4), 1847–1852.
- Baron, S. G., Thompson-Schill, S. L., Weber, M., & Osherson, D. (2010). An early stage of conceptual combination: Superimposition of constituent concepts in left anterolateral temporal lobe. *Cognitive Neuroscience*, 1(1), 44–51.
- Barton, J. J. S., Fox, C. J., Sekunova, A., & Iaria, G. (2010). Encoding in the visual word form area: An fMRI adaptation study of words versus handwriting. *Journal of Cognitive Neuroscience*, 22(8), 1649–1661.
- Bemis, D. K., & Pykkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience*, 31(8), 2801–2814.
- Bright, P., Moss, H., & Tyler, L. K. (2004). Unitary vs multiple semantics: PET studies of word and picture processing. *Brain and Language*, 89(3), 417–432.
- Cristescu, T. C., & Nobre, A. C. (2008). Differential modulation of word recognition by semantic and spatial orienting of attention. *Journal of Cognitive Neuroscience*, 20(5), 787–801.
- Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *The Journal of Neuroscience*, 33(48), 18906–18916.
- Eitel, A., Scheiter, K., & Schüler, A. (2013). How inspecting a picture affects processing of text in multimedia learning. *Applied Cognitive Psychology*, 27(4), 451–461.
- Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 202–224.

- Friston, K. J., Frith, C. D., & Frackowiak, R. S. J. (1993). Time-dependent changes in effective connectivity measured with PET. *Human Brain Mapping, 1*(1), 69–79.
- Gates, L., & Yoon, M. G. (2005). Distinct and shared cortical regions of the human brain activated by pictorial depictions versus verbal descriptions: An fMRI study. *NeuroImage, 24*(2), 473–486.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*(10–11), 1409–1422.
- Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 819–836). Cambridge, MA: MIT Press.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438–441.
- Hede, A. (2002). Integrated model of multimedia effects on learning. *Journal of Educational Multimedia and Hypermedia, 11*(2), 177–191.
- Hochpöchler, U., Schnotz, W., Rasch, T., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2013). Dynamics of mental model construction from text and graphics. *European Journal of Psychology of Education, 28*(4), 1105–1126.
- Hoffman, P., Evans, G. A. L., & Lambon Ralph, M. A. (2014). The anterior temporal lobes are critically involved in acquiring new conceptual knowledge: Evidence for impaired feature integration in semantic dementia. *Cortex, 50*, 19–31.
- Holcomb, P. J., & Mcpherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition, 24*(2), 259–276.
- Huang, J., Zhu, Z. D., Zhang, J. X., Wu, M. X., Chen, H.-C., & Wang, S. P. (2012). The role of left inferior frontal gyrus in explicit and implicit semantic processing. *Brain Research, 1440*, 56–64.
- Huang, J., Wang, S., Jia, S., Chen, H.-C. (2012). *The dynamics of brain activation in semantic integration: A study combining ERP and Event-Related Optical Signal*. Poster presented on the 19th Annual Cognitive Neuroscience Society Meeting. Chicago, USA.
- Kassuba, T., Klinge, C., Hölig, C., Menz, M. M., Ptito, M., Röder, B., & Siebner, H. R. (2011). The left fusiform gyrus hosts trisensory representations of manipulable objects. *NeuroImage, 56*(3), 1566–1577.
- Khateb, A., Pegna, A. J., Landis, T., Mouthon, M. S., & Annoni, J.-M. (2010). On the origin of the N400 effects: An ERP waveform and source localization analysis in three matching tasks. *Brain Topography, 23*(3), 311–320.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920–933.
- Li, M., Zhong, N., & Lu, S. F. (2011). *Shared and separate semantic systems for text and figure: Converging evidence from eye-tracking and brain imaging*. Paper presented at the meeting of 2011 IEEE/ICME International Conference on Complex Medical Engineering (CME).
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge: Cambridge University Press.
- Mayer, R. E. (2010). Applying the science of learning to medical education. *Medical Education, 44*(6), 543–549.
- Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107–119.
- Miller, M. B., Donovan, C.-L., Bennett, C. M., Aminoff, E. M., & Mayer, R. E. (2012). Individual differences in cognitive style and strategy predict similarities in the patterns of brain activity between individuals. *NeuroImage, 59*(1), 83–93.
- Moore, C. J., & Price, C. J. (1999). Three distinct ventral occipitotemporal regions for reading and object naming. *NeuroImage, 10*(2), 181–192.
- Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4*(1), 15–22.
- Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *The Journal of Neuroscience, 32*(45), 15728–15736.
- Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology, 19*(7), 603–639.
- Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences, 1191*(1), 62–88.
- Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage, 19*(3), 473–481.
- Reinke, K., Fernandes, M., Schwindt, G., O'Craven, K., & Grady, C. L. (2008). Functional specificity of the visual word form area: General activation for words and symbols but specific network activation for words. *Brain and Language, 104*(2), 180–189.
- Richardson, D., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition, 102*(1), 129–138.
- Saffran, E. M., Coslett, H., Martin, N., & Boronat, C. B. (2003). Access to knowledge from pictures but not words in a patient with progressive fluent aphasia. *Language and Cognitive Processes, 18*(5–6), 725–757.
- Schendan, H. E., & Maher, S. M. (2009). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *NeuroImage, 44*(4), 1423–1438.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*(2), 141–156.
- Schnotz, W., & Kürschner, C. (2008). External and internal

- representations in the acquisition and use of knowledge: Visualization effects on mental model construction. *Instructional Science*, 36(3), 175–190.
- Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., & Just, M. A. (2011). Commonality of neural representations of words and pictures. *NeuroImage*, 54(3), 2418–2425.
- Spotorno, N., Cheylus, A., van Der Henst, J.-B., & Noveck, I. A. (2013). What's behind a p600? Integration operations during irony processing. *PloS One*, 8(6), e66839.
- Stafura, J. Z., & Perfetti, C. A. (2014). Word-to-text integration: Message level and lexical level influences in ERPs. *Neuropsychologia*, 64, 41–53.
- van Doren, L., Dupont, P., De Grauwe, S., Peeters, R., & Vandenberghe, R. (2010). The amodal system for conscious word and picture identification in the absence of a semantic task. *NeuroImage*, 49(4), 3295–3307.
- van Genuchten, E., Scheiter, K., & Schüler, A. (2012). Examining learning from text and pictures for different task types: Does the multimedia effect differ for conceptual, causal, and procedural tasks? *Computers in Human Behavior*, 28(6), 2209–2218.
- Vandenberghe, 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(6597), 254–256.
- Vigneau, M., Beaucousin, V., Herve, P.-Y., Duffau, H., Crivello, F., Houdé, O., ... Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, 30(4), 1414–1432.
- Warrington, E. K. (1975). The selective impairment of semantic memory. *The Quarterly journal of Experimental Psychology*, 27(4), 635–657.
- Warrington, E. K., & McCarthy, R. A. (1994). Multiple meaning systems in the brain: A case for visual semantics. *Neuropsychologia*, 32(12), 1465–1473.
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107(3), 829–853.
- West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, 13(3), 363–375.
- Westerlund, M., & Pykkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, 57, 59–70.
- Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, 17(10), 2322–2333.
- Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, 20(7), 1235–1249.
- Wu, Y. C., & Coulson, S. (2011). Are depictive gestures like pictures? Commonalities and differences in semantic processing. *Brain and Language*, 119(3), 184–195.
- Yeatman, J. D., Rauschecker, A. M., & Wandell, B. A. (2013). Anatomy of the visual word form area: Adjacent cortical circuits and long-range white matter connections. *Brain and Language*, 125(2), 146–155.
- Zhu, Z. D., Feng, G. Y., Zhang, J. X., Li, G. C., Li, H., & Wang, S. P. (2013). The role of the left prefrontal cortex in sentence-level semantic integration. *NeuroImage*, 76, 325–331.
- Zhu, Z. D., Hagoort, P., Zhang, J. X., Feng, G. Y., Chen, H.-C., Bastiaansen, M., & Wang, S. P. (2012). The anterior left inferior frontal gyrus contributes to semantic unification. *NeuroImage*, 60(4), 2230–2237.

The Cognitive and Neural Mechanism of Text and Picture Processing in Multimedia Learning

LI Songqing; ZHAO Qingbai; ZHOU Zhijin; ZHANG Yi

(Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education;
Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology,
Central China Normal University, Wuhan 430079, China)

Abstract: According to the cognitive theory of multimedia learning and the integrated theory of text and picture comprehension, text and picture are processed through distinct channels at the early stage, and then integrated into a whole. The theories are supported by empirical researches in neuroscience. Brain imaging studies have shown that the processing of text and picture is different in the semantic pre-process, but shares the same semantic neural system in information integration later. Event-related potential studies have found that only picture evokes a N300 in the early stage, while both picture and text induce the N400 effect later. Although existing studies have provided clear evidence that text and picture share the same semantic system, further research is needed to explore how the information is integrated in this system.

Key words: multimedia learning; semantic processing; text and picture integration; brain imaging; event-related potential