

# 视觉表象可视化——视觉表象研究的新途径<sup>\*</sup>

张得龙 梁碧珊 文学 黄瑞旺 刘鸣<sup>\*\*</sup>

(华南师范大学心理学院暨心理应用研究中心, 广州, 510631)

**摘要** 功能磁共振技术在表象研究中得到广泛应用是表象研究追求客观化精确性的必然趋势。本文介绍了功能磁共振多变量模式分析方法及其演变历程,探讨了借助该方法实现“视觉表象可视化”的理论依据与亟待解决的关键问题。分析指出“视觉表象可视化”将为表象研究提供全新的研究视角与方法途径。

**关键词** 表象 fMRI 多变量模式分析 解码模型 编码模型

1978年Shepard曾提出过一个“视觉表象外在化”的研究设想。亦即研制一套装置,这种装置既能检测出被试观看某种客体时的知觉象,同时也能检测出客观事物消失后出现在头脑中的记忆表象和想象表象。检测出来的视觉表象可以转换为电信号呈现在荧光屏上供研究者进行考察。毋庸置疑,实现“视觉表象外在化”需要具备多方面的条件,时至今日仍是一个大胆而富有创造性的研究设想。其中,准确客观地建立视觉表象与脑神经活动之间的对应关系是实现视觉表象外在化的基本前提。近年来兴起的功能磁共振“多变量模式分析”(multi-variate pattern analysis, MVPA)方法可以有效地建立心理活动与脑神经活动之间的对应关系,为视觉表象可视化研究奠定了基础。本文在介绍多变量模式分析起源及其演变历程的基础上,归纳分析视觉表象可视化的理论依据与亟待解决的关键问题。分析指出“视觉表象可视化”是表象研究的新途径,将促进表象实质与功能问题的探讨,有利于推动视觉表象研究的进一步发展。

## 1 表象研究的困境

众所周知,表象是心理学研究中争论颇多的主题之一。科学心理学伊始,表象就被列为心理学研究主题。但表象研究在方法与理论模型上逐渐成熟并被广泛认可肇始于20世纪70年代,此时距离表象被心理学确立为研究主题已有一个世纪之久。在此期间,表象研究由盛至衰(之间更有长达半个多世纪之久表象研究因少人问津而几近中断),继而“东山再起”重新进入心理学研究的视野。复苏的表象

研究领域更是争论不断,甚至出现了盛极一时的“表象之争”(Bartolomeo, 2002)。

表象研究之所以一步三回头,其中原委与表象具有主观复杂的特性而表象研究又缺乏客观有效性的方法之间的矛盾不无关系(刘鸣, 2004)。诚如希尔加德所言:“主观心理活动中最不容否认而又最不易通过客观途径来进行观察研究的莫过于表象”。表象研究方法的新颖独创性与客观有效性一直是研究者们竭力追求而又倍感棘手的问题。纵观表象研究进退浮沉的整个历程,心理学研究方法的取向与革新在很大程度上决定了表象研究的命运(刘鸣, 2004)。众所周知,表象不仅具有高度主观内在性,而且其表现形式与内容也非常复杂多变,较之于其他研究主题,表象研究对方法新颖独创性与客观有效性的追求更为迫切也更为执着:表象的独特性“迫使”研究者采用更加科学有效新颖巧妙的方法进行研究;而方法的发展反过来也会促进表象问题的进一步研究。因此,革新表象研究方法是表象研究的重要组成部分,表象研究期待更加科学有效方法的问世,以实现在更高层次上对表象问题进行科学研究与探讨。

## 2 多变量模式分析

功能磁共振技术具有相对较高的空间分辨率(约3毫米)可以数秒(2~3秒)内在大脑数万个不同解剖位置采集到任务加工中神经活动导致的血氧水平改变引发的信号(BOLD)(Norman et al., 2006)。目前,功能磁共振(fMRI)技术在认知神经科学研究中得到广泛应用,在心理活动神经机制的

<sup>\*</sup> 本研究得到国家自然科学基金项目(31371049)、国家自然科学基金专项基金项目(31140052)、国家社科基金“十二五”规划2011年度教育学一般课题(BBA110411)和省部共建人文社科重点研究基地项目(11JJD190003)的资助。

<sup>\*\*</sup> 通讯作者:刘鸣。E-mail: lium@scnu.edu.cn

探讨中发挥的作用日益凸显出来。然而,功能磁共振能够采集的最小单位—体素(voxel)( $3\text{mm} \times 3\text{mm} \times 3\text{mm}$ )包含约数万个神经元,而且存在血液动力学(HRF)延迟(5~10秒)现象,并且与任务有关的神经活动信号淹没在各种噪音之中(任务信号约为2~3%)。因此,如何从fMRI数据中提取出与任务相关的神经活动信号是认知神经科学的基本问题。

由于fMRI数据信噪比(signal noise ratio, SNR)较低,传统数据处理需要把兴趣区(ROI)内体素激活值(voxel activation value)进行空间平滑(smoothing)在统计分析中将兴趣区内体素激活值进行平均叠加(aggregation)通过比较不同实验条件下兴趣区激活强度的差异来推断不同脑区在任务加工中的功能(Friston et al., 1995; Mur et al., 2009; Worsley & Friston, 1995)。因此,传统数据处理方法也被称之为“激活为基础”(activation basis)的方法(Norman et al., 2006)。从本质上看,传统数据处理主要检测单个体素在不同实验条件下激活强度是否存在显著差异,属于单变量数据分析方法(voxel-wise analysis)。其基本假设是:兴趣内各个体素在同一实验条件下反应趋势是一致的,相互之间在激活强度上可以叠加。虽然传统数据处理方法在一定程度上可以提高信噪比,但数据处理过程只考虑单一决策变量,故而对fMRI数据信号的统计检测力不足。主要体现在如下三个方面(Haynes & Rees, 2006; Norman et al., 2006; Pereira et al., 2009):(1)传统数据处理以不同实验条件下兴趣区内的平均激活强度为指标进行数据分析,仅考虑了较强信号的体素,但实际上反应微弱的体素也携带任务相关信息(Haxby et al., 2001);(2)传统数据处理假设兴趣内各个体素对同一任务的反应趋势一致,从而相互之间可以进行叠加平均。但实际上单个体素内包含数万个神经元,不同体素内神经元在任务加工中反应趋势往往不一致,将兴趣内的体素进行叠加操作无疑会丢失信息(Kriegeskorte, Goebel, & Bandettini, 2006);(3)传统数据处理只考虑单一体素的信息,而大脑体素间的协同活动也携带任务相关信息,但传统数据处理并没将其考虑在内(Cox & Savoy, 2003)。

借鉴统计学习理论,研究者们发展出“多变量模式分析”(MVPA)(Haynes & Rees, 2006; Kriegeskorte & Bandettini, 2007; Norman et al., 2006)。与传统的单变量数据分析方法不同,MVPA将兴趣区内各个体素的信号同时考察,将兴趣区看作一个多

维变量,即空间模式(spatial pattern)来分析大脑激活数据与任务相关的信息(Mur et al., 2009; Norman et al., 2006)。其基本假设是:兴趣区内不同体素携带与任务相关的不同维度信息,体素构成的激活模式能够有效地将对应的信息内容与其他激活模式对应的信息内容区分开,也就是大脑认知状态与神经激活模式之间存在对应关系。MVPA充分考虑了体素特异性以及体素之间的关系,从而有效提高了数据处理的敏感性(Haynes et al., 2007; Haynes & Rees, 2005; Polyn et al., 2005; Raizada et al., 2010)。借助MVPA方法,研究者可以成功从大脑激活模式中“读取”与心理活动状态相关的“信息内容”,而非仅仅关注脑区在任务加工中的激活程度(Mur, Bandettini, & Kriegeskorte, 2009)。MVPA被研究者称之为“读心术”,其根本原因也在于此。

## 2.1 解码模型与编码模型

MVPA数据处理一般包括4个步骤(Norman et al., 2006):特征选择(feature selection),模式聚类(pattern assemble)模型训练(model training),以及模型泛化测试(model test)。其中,模型训练,亦即建立实验条件与脑神经激活模式之间的函数关系,是多变量模式分析的核心。目前,模型主要有两类,分别为:编码模型与解码模型(Naselaris, Kay, Nishimoto, & Gallant, 2011)。

解码模型是MVPA首先发展起来的方法。传统功能磁共振研究一般遵循“实验条件(experiment condition) - 大脑反应(response) - 激活(activation)”的关系链,数据处理关注不同实验条件下兴趣区内脑神经激活强度的差异。既然存在“实验条件 - 大脑反应 - 激活”关系链,那么数据处理能否反其道而行之,能否从大脑的神经激活中反推出与之对应的实验条件类型?研究者借鉴机器学习理论发展出以解码模型为基础的MVPA方法(Cox & Savoy, 2003; Haxby et al., 2001)。解码模型通过创建判决函数来“学习”实验条件类型与脑神经激活模式之间的关系。经过训练的分类器可以从大脑的激活模式中“解码”其表征的信息内容。这一数据处理思路的转变看似简单,实则意义重大。首先,传统数据处理面临研究效度问题,数据处理需要多重比较校正,实验研究也需要进行多模态、多变式、多指标、多程序的交互验证。而解码模型将研究效度转换为从脑激活模式中预测实验条件的准确性问题,避免了研究效度问题。其次,传统数据处理方法的结果不具有统计预测力。通过建立脑神经激活模

式与实验条件类型之间的函数关系而形成的映射模型(解码模型)则可以根据当前的脑神经激活模式“预测”与其所对应的最大可能的实验条件类型(Cox & Savoy, 2003; Haxby et al., 2001)。再次, MVPA 考虑单个体素的特异性以及体素间的相互关系, 明显提高了数据处理的敏感性。数据处理可以实现从脑神经激活模式中“追踪”出单一实验条件的信息内容, 为认知神经科学实时的研究人类心理活动的神经机制提供了可能(Norman et al., 2006)。

虽然解码模型存在上述优势, 但不足之处也很明显。解码模型的泛化能力较差, 对实验条件类型的区分不具普遍性(Kay, Naselaris, Prenger, & Gallant, 2008)。根据脑神经激活模式预测实验条件类型只能局限在参与分类器训练的实验条件类型范围内, 不能泛化到其他实验条件类型。为克服解码模型的这一局限, 研究者发展出编码模型(Nevado et al., 2004; Wu et al., 2006)。如前所述, 解码模型通过兴趣区内体素激活模式来区分实验条件的类型, 那么能否使用实验条件包含刺激的不同特征(例如, 视觉特征)模式反过来预测大脑的激活? 这是编码模型数据处理的基本出发点。解码模型泛化能力差的一个主要原因是只能通过刺激类别标签描述刺激。与之不同, 编码模型通过刺激特征模式来预测大脑的激活, 其对刺激的描述通过刺激的特征进行, 而刺激的特征在刺激之间是具有普遍性的, 从而克服了解码模型泛化能力差的局限(Kay et al., 2008)。毫无疑问, 编码模型是在解码模型基础上对传统数据处理方式的又一次超越。在数据处理路线上, 编码模型与传统的数据处理路线是一致的, 都是遵循“实验条件-大脑反应-激活”关系链。但是编码模型与传统数据处理有本质区别(Naselaris et al., 2011)。我们知道传统数据处理需要估计一个模型, 属于模型驱动的数据处理方法。例如, SPM 数据处理方法用一般线性模型(GLM)拟合(fit)每个体素激活, 一般线性模型的参数估计完成后, 单个体素的参数显著性被评估并在兴趣区内进行叠加(aggregate)(Worsley & Friston, 1995)。基于单个体素的多变量模式分析的编码模型也需要建立外界刺激的特征模式和大脑单个体素激活之间的函数关系(也是基于一般线性模型)(Naselaris et al., 2011)。但与传统的数据处理方式不同, 编码模型通过一般线性模型拟合的 $\beta$ 值描述刺激事件, 进而建立外界刺激事件与脑神经激活事件之间的概率分布关系

(Kay et al., 2008; Naselaris et al., 2011)。借助编码模型, 研究者可以通过外界刺激特征模式“预测”大脑激活模式, 以此探讨外界刺激信息在大脑内表征的问题(Naselaris et al., 2011)。

总之, 解码模型和编码模型是多变量模式分析两种主要的数据处理方法, 在认知神经科学研究中发挥重要作用(Kay & Gallant, 2009)。从解码模型的角度, 研究者试图探讨从观测到的脑激活模式中“解码”出何种信息; 而从编码模型的角度, 研究者尝试探讨当外界刺激变化时, 大脑激活是如何随之改变的(Naselaris et al., 2011)。目前, 借助于编码模型或解码模型, 研究者可以有效地研究不同类型刺激特征在大脑同一个脑区是如何表征的, 同一类型的刺激特征在大脑不同脑区是如何表征的(Kriegeskorte et al., 2006; Pereira et al., 2009); 编码模型可以很好地探讨哪些刺激特征可以完全“解释”一个脑区的激活(Naselaris et al., 2009; Wu et al., 2006); 解码模型可以有效建立脑激活模式和行为反应之间的关系(Raizada et al., 2010; Walther et al., 2009; Williams et al., 2007)。

## 2.2 多变量模式分析的应用

目前, 多变量数据分析方法被广泛应用于脑神经科学研究领域, 包含多种类型: 功能连接(functional connectivity)、效应连接(effective connectivity)(Friston, 2011)、动态因果模型(dynamic causal modeling, DCM)、偏最小二乘法(partial least square)、主成分分析等。虽然都属于多变量数据分析方法, 多变量模式分析与其他多变量数据分析方法存在显著差别。其他多变量数据分析方法主要关注不同脑区之间的功能关系, 而多变量模式分析主要关注从脑激活模式中“读取”出与大脑状态相关的“信息内容”(Haynes & Rees, 2006; Norman et al., 2006)。目前, 借助多变量模式分析方法的认知神经科学研究大致可以归结为3个层次。

第一个层次, 类别解码(class decoding)。解码模型可以有效地建立刺激类别与脑激活模式之间的函数关系。借助解码模型, 研究者可以从给定的脑激活模式中读取其表征刺激的“类别”信息(Carlson et al., 2003; Cox & Savoy, 2003; Haxby et al., 2001; Haynes & Rees, 2005; Kamitani & Tong, 2005)。目前, 解码模型被广泛应用于认知神经科学研究中, 研究问题涉及到认知活动的各个方面, 主要包括: 视觉领域的研究(Brouwer & Heeger, 2009; Macevoy & Epstein, 2009; Peelen et al., 2009), 体

感 (Beauchamp et al., 2009; Rolls et al., 2009) 嗅觉 (Howard, Plailly, Grueschow, Haynes, & Gottfried, 2009) 听觉 (Ethofer, Van De Ville, Scherer, & Vuilleumier, 2009), 运动 (Dinstein, Gardner, Jazayeri, & Heeger, 2008) 注意 (Kamitani & Tong, 2006) 意识 (Schurger, Pereira, Treisman, & Cohen, 2010) 记忆 (Harrison & Tong, 2009; Johnson et al., 2009) 意图 (Haynes et al., 2007) 认知控制 (Esterman, Chiu, Tamber - Rosenau, & Yantis, 2009) 决策 (Soon, Brass, Heinze, & Haynes, 2008) 表象 (Reddy, Tsuchiya, & Serre, 2010)。

第二个层次, 刺激鉴别 (stimulus identification)。编码模型可以根据脑激活模式“鉴别”给定刺激集合中哪一个具体刺激最大可能的激活了当前观测到的脑激活模式。Kay 等人 (2008) 在研究中发现编码模型可以根据脑激活模式“鉴别”自然场景图片。自然场景图片 (对应着以像素为维度的数字矩阵) 可以通过非线性变换转换为以视觉特征 (如, Gabor wavelet) 为维度的视觉特征模式。编码模型可以建立视觉特征模式与大脑单个体素激活值之间的函数关系, 从而建立图像视觉特征模式与脑激活模式之间的事件概率模型。研究者通过该模型可以有效地“鉴别”给定的图像集合中哪个图像在概率上最大可能的激活了当前观测到的脑激活模式。不仅具体的图片刺激可以根据脑激活模式“鉴别”出来, 较为抽象的名词同样也可以实现。Mitchell 等人将 25 个动词和具体名词在文本搭配使用的频率作为描述名词的特征, 通过编码模型同样实现了根据脑激活模式“鉴别”具体名词的目的 (Mitchell et al., 2008)。需要注意, 具体刺激能够从脑激活模式中“鉴别”出来, 其前提条件是刺激集合是预先给定的, 编码模型只能从给定的集合中推测哪个个体刺激在概率上最大可能的激活了当前观测到的激活模式。

第三个层次, 形象重建 (image reconstruction)。物体图片形象和视觉特征存在对应关系 (两者之前的关系是非线性), 而物体视觉特征和脑激活模式之间也可以建立关系 (大量研究证实两者之间可以通过线性关系描述) (DiCarlo & Cox, 2007; Naselaris et al., 2011)。那么能否可以根据当前观测到的大脑激活模式将引起该激活模式的刺激图片的“形象” (对应灰度值矩阵值) “还原”出来。也就是通过大脑激活模式“再造成”引起该激活模式的“刺激” (Kay & Gallant, 2009; Miyawaki et al., 2008; Naselaris et al., 2009; Thirion et al., 2006)。当然, 在数

学上通过脑激活模式“反推”刺激特征不存在“唯一解” (只存在“最优解”), 因此将真正的“实验刺激”从激活模式中“还原”出来是不可能的, 我们只能通过先验概率推测出在概率上最大可能引起当前激活模式的刺激的“副本” (Naselaris, Prenger, Kay, Oliver, & Gallant, 2009)。现有研究业已证实编码模型结合贝叶斯模型 (Bayesian model) 可以实现从大脑激活模式中“重建”外在刺激形象的目的。其基本框架为: 首先将刺激转换到视觉特征 (如, Gabor wavelet), 然后使用编码模型建立特征模式和兴趣区单个体素激活之间的关系, 进而建立刺激特征模式和大脑兴趣区激活模式之间的事件概率分布 (先验概率) (这一步同于“刺激鉴别”操作), 借助贝叶斯模型推测引起当前脑激活模式的刺激特征的模式, 进而“还原”刺激形象 (Ma, Beck, Latham, & Pouget, 2006)。除了视觉特征, 刺激其他维度的信息 (如, 语义特征) 也可以通过编码模型建立“刺激”和“脑激活模式”之间的对应关系 (Naselaris et al., 2009)。

### 3 视觉表象可视化的理论依据

视知觉加工的图片形象之所以可以根据脑激活模式“再造”出来, 根本原因在于编码模型可以有效建立图片的视觉特征与脑激活模式之间的对应关系。而视觉表象是在缺乏外界刺激的情况下生成的感性形象, 根本不具备这一基本条件, 因而知觉形象可视化的途径难以直接应用于视觉表象。

传统观点认为初级视觉皮层的激活主要受刺激的低层视觉特征 (明暗、纹理变化等) 驱动 (Hsieh et al., 2010), 研究也证实低层视觉特征与脑区激活之间的“拟合”关系在初级视觉皮层也较为一致 (Naselaris et al., 2009)。目前, 根据脑激活模式重建知觉图片形象主要通过编码模型建立低层视觉特征 (Gabor 特征) 与初级视觉皮层 (V1, V2, V3) 的脑激活模式之间的关系实现 (Naselaris et al., 2009)。但大量研究也证实即使是初级视觉皮层的激活也并非完全由外部刺激驱动, 大脑自上而下的调控机制发挥重要作用 (Diekhof et al., 2011; Friston et al., 2006; Hsieh et al., 2010)。现有研究单纯通过建立图片视觉特征 (语义特征) 与脑激活模式之间的关系来重建的知觉图片形象质量比较低 (Naselaris et al., 2009), 这应该是其中一个比较重要的原因。其实, 早在十九世纪后期, 威廉 - 詹姆斯在《心理学原理》中曾预言 (Raichle, 2010): “我们所知觉到的内

容一部分来自于我们眼前的事物,而更大的一部分来自于我们头脑内部”。这一预言很精辟的概括了自十九世纪起关于大脑功能的两种针锋相对的观点:以谢林顿为代表的脑功能反映论(reflexive),主张大脑活动主要由环境刺激驱动。与此不同,格雷汉姆-布朗等人主张大脑活动是自主的(intrinsic),大脑获得(acquisition)和保持(maintenance)的信息用于理解环境刺激(environmental demands),作用于甚至可以提前预测环境刺激。根据Kosslyn等人知觉预期理论的观点(Kosslyn, 2005),表象神经机制很大程度上涉及到大脑自上而下调控知觉加工的机制。大量功能磁共振研究也证实知觉过程与表象过程激活的脑区存在重叠现象。例如,即使闭眼的情况下,想象一个视觉场景同样激活视觉皮层(Kosslyn, Thompson, Kim, & Alpert, 1995);而且视觉皮层激活降低会降低表象的质量(Kosslyn et al., 1999);表象的激活模式与知觉刺激一样也存在范畴特异性(Mechelli, Price, Friston, & Ishai, 2004)。基于上述的发现,研究者倾向于将表象看作是知觉加工的特例(special case)(Kosslyn, Ganis, & Thompson, 2001):知觉由外界视觉刺激驱动产生,而表象是由高级认知加工中心,如前额皮层(Mechelli et al., 2004)的信号激发。但需要注意,主观想象的事物形象与我们真实看到事物肯定是不同的,表象过程与知觉过程脑区激活的重叠现象并不足以证明两者是完全相同的心理过程。已有的功能磁共振研究也证实大脑活动在表象和知觉中的差别。例如单细胞追踪技术发现表象与知觉的反应潜伏期存在差别(Kreiman, Koch, & Fried, 2000);有些现象(如听觉皮层的钝化现象)在表象与知觉过程中并不一致(Amedi, Malach, & Pascual-Leone, 2005);表象与知觉激活模式重叠现象在不同脑区也存在差异(Reddy et al., 2010)。但根据知觉预期理论,表象参与大脑自上而下调控知觉过程的机制,表象与知觉仍存在密切关系。

以多变量模式分析方法探讨表象神经激活模式与知觉神经激活模式之间的关系,是实现“视觉表象可视化”的重要途径。既然知觉形象可以根据脑激活模式“重建”出来,而表象神经激活模式与知觉神经激活模式存在关系,如果我们能够揭示视觉表象神经激活模式与知觉神经激活模式之间的关系,视觉表象应该也可以实现可视化。而表象神经激活模式与知觉神经激活模式之间存在关系的根源是表象参与大脑自上而下调控知觉的机制,因此通过多变

量模式分析探讨视觉表象神经激活模式如何调控知觉神经激活模式是探讨视觉表象激活模式与知觉激活模式之间关系的关键。已有研究证实在大脑形状特异性脑区(LOC),字母“X”和“O”的知觉激活模式与对应的表象激活模式相似(Stokes, Thompson, Cusack, & Duncan, 2009)。也就是,表象在LOC脑区对应的神经激活模式与知觉对应的神经激活模式类似,而表象在LOC神经激活模式应该是其自上而下调控知觉神经刺激模式的一种体现。在物体识别领域,研究发现在LOC后部(LO)内场景的典型物体的知觉激活模式可以线性“合成”场景的知觉表征模式(MacEvoy & Epstein, 2011)。由于该研究涉及的场景内的典型物体并没在场景中出现,而且排除了注意的影响,这一研究结果可能意味着:形状特异性脑区后部表征物体的个别属性(property)而非整体属性,而且表征主要在大脑已有“知识”自上而下的调控下进行的。该结论的得出主要基于两个原因:(1)大脑自上而下生成的表象在该脑区激活的模式与知觉过程大脑激活模式相似;(2)在场景中没有出现的物体知觉表征模式可以线性“组合”为场景的表征模式。既然物体的知觉神经表征模式可以“组合”为场景的神经表征模式,那么相应的物体表象激活模式在该脑区应该也可以“组合”形成场景激活模式,这应该是物体知觉激活模式可以“组合”成场景表征模式的根源所在。亦即,在知觉过程中可能不存在“纯粹”由外界刺激驱动的知觉加工,知觉过程是在表象参与调控下进行的(注意等其他心理过程也会参与调控)(Diekhof et al., 2011; Gazzaley et al., 2005; Kamitani & Tong, 2005)。其中,大脑形状特异性脑区应该是大脑自上而下调控知觉表征的核心脑区,已有研究已经证实LOC在调控大脑初级视觉皮层知觉表征中发挥重要作用。如此一来,在大脑视觉皮层的不同脑区(存在等级结构)探讨表象激活模式与知觉激活模式之间的关系既是揭示知觉表征的方法,也是探究表象实质的重要途径。毋庸置疑,该问题的研究既有利于丰富当前从脑激活模式中重建知觉形象的研究,也为“视觉表象可视化”提供理论依据。

#### 4 “视觉表象可视化”亟待解决的问题

如前所述,表象调控知觉激活模式是视觉表象激活模式与知觉激活模式存在关系的根源。虽然现有研究已经证实形状特异性脑区是表象调控知觉激活模式的关键脑区,但是在视觉皮层(包括形状特异性

脑区以及初级视觉皮层) 知觉激活模式信息与表象参与的自上而下调控信息之间的关系尚不清楚。该问题是“视觉表象可视化”的核心,至少涉及如下两个方面:

(1) 表象激活模式在调控知觉表征过程中的作用。如果传统的脑成像研究已经证实表象的实质涉及到大脑自上而下调控知觉表征的神经机制,但表象激活模式是如何调控知觉的神经机制尚不清楚。

(2) 表象表征模式和知觉表征模式之间关系的数学模型。表象可以如知觉形象一样从脑激活模式中重建出来的基本前提是我们可以数量化的描述表象激活模式和知觉表征模式之间的关系。如何数量化的描述两者之间的关系目前仍是个难题。可能的一个途径是借鉴编码模型的思想,用知觉过程中的视觉特征(语义特征)模式来拟合表象过程中特定脑区的激活模式,进而建立知觉过程刺激的视觉特征模式事件与表象过程脑激活模式事件之间的概率分布,然后借助贝叶斯模型推测出与表象激活模式对应的视觉特征模式,进而实现“视觉表象的可视化”。

## 5 挑战与展望

目前根据脑激活模式重建知觉形象尚处于起步阶段,实现“视觉表象可视化”更是“任重而道远”,视觉表象活动与我们的心理活动接近,“视觉表象可视化”的实现将为实现真正意义上的“读心术”奠定基础,这应该是认知神经科学未来研究的一个热点。但作为一个新兴领域,“视觉表象可视化”的最终实现需要如下三个方面相关工作的开展。

### 5.1 磁共振数据获取与数据处理技术的发展

(1) 如何更好地借助功能磁共振技术获取大脑活动的信号。例如提高功能磁共振磁场的场强,提高功能磁共振信号采集序列的空间分辨率或时间分辨率等都是未来该领域的研究需要改进的方向(Naseleris et al., 2009; Op de Beeck et al., 2008)。

(2) 如何更好地优化现有的数学模型。读心术的实现很大程度上基于数学模型构建“刺激特征”与脑激活模式之间的关系实现的。如何优化现有的数学模型以更准确有效地建立“刺激特征”与脑神经活动之间的关系应该未来的一个重要研究方向(MacEvoy & Epstein, 2011)。

(3) 如何实现“重建”动态心理活动。心理活动的一个重要特点是动态性,动态的心理活动能否从脑激活模式中“还原”?较之于静态的图片刺激,动

态的刺激信息(如,电影)从脑激活模式中重建出来的难度较大,其中如何解决动态变化的刺激与大脑血液动力学函数延迟之间的矛盾是关键(Nishimoto et al., 2011)。

(4) 能否实时的再现心理活动。目前,功能磁共振研究中实时技术是个热点问题。而实时技术与读心术的结合应该可以实现心理活动实时再现。

### 5.2 心理活动实质的认识

(1) 物体知觉信息的神经表征模式。“视觉表象可视化”的实质在于将大脑激活模式所表征的“信息”以外在化的方式“重建”出来,因此大脑皮层的神经活动以何种方式表征物体信息无疑是“视觉表象可视化”的基本问题。我们知道多变量模式分析方法与分布式重叠表征观点一脉相承,相互印证(Cox & Savoy, 2003; Haxby et al., 2001)。对于大脑知觉表征问题,除了分布式重叠表征观点外,还存在其他比较有影响力的表征模型,如模块化表征模型,过程特异性表征模型(Haxby et al., 2001)。各个模型对知觉表征问题都有一定程度的解释力,在分布式重叠表征观的基础上如何整合不同的表征模型,并以此探究神经表征的实质是实现“视觉表象可视化”的关键问题(O'Toole, Jiang, Abdi, & Haxby 2005; Spiridon & Kanwisher, 2002)。

(2) 大脑自上而下调控知觉表征的神经机制。知觉加工是在大脑自上而下的调控下进行的,大脑内部不存在单纯“自下而上”驱动的知觉加工。尽管目前的研究已经证实即使初级视觉皮层的知觉表征都受到大脑高级视觉皮层的调控(Hsieh, Vul, & Kanwisher, 2010),然而目前尚无研究从“脑激活模式信息内容”的角度揭示高级视觉皮层调控初级视觉皮层的神经机制。对于该问题的探讨包括在该过程中涉及的脑区的定位,这些脑区激活模式之间的关系,以及何种类型的编码模型(如,视觉特征或语义特征)可以有效地描述不同脑区的激活模式(Naseleris et al., 2009)。对于这些问题的回答关系到编码模型对于脑激活模式描述的有效性,不仅对于“视觉表象可视化”问题,即使对单纯的“视觉形象可视化”问题也非常关键。

### 5.3 “视觉表象可视化”与现实生活需求的结合

(1) “视觉表象可视化”目前仍处于可行性探讨阶段,但现实生活的需要无疑是“视觉表象可视化”发展的最根本推动力。而且,“视觉表象可视化”本身就具有巨大的应用前景,既可以应用于临床实践中特殊群体的研究,如植物人(Monti et al., 2010);

也可以应用于正常人的日常心理活动的研究之中,如梦境的研究(Naselaris et al., 2009)。因此,如何将“视觉表象可视化”的研究与现实生活的需要结合起来以推动“视觉表象可视化”问题的发展并满足社会生活的需要是研究者面临的一个重要问题。

(2) 伦理道德规范建设。由于涉及到伦理、道德方面的问题,如何规范该领域的相关研究是研究者面临的一个重要问题。

### 参考文献

- 刘鸣. (2004). 表象研究方法论. *心理科学*, 27(2), 258-260
- Amedi, A., Malach, R., & Pascual-Leone, A. (2005). Negative BOLD differentiates visual imagery and perception. *Neuron*, 48(5), 859-872.
- Bartolomeo, P. (2002). The relationship between visual perception and visual mental imagery: A reappraisal of the neuropsychological evidence. *Cortex*, 38(3), 357-378.
- Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, 19(2), 261-270.
- Dinstein, I., Gardner, J. L., Jazayeri, M., & Heeger, D. J. (2008). Executed and observed movements have different distributed representations in human aIPS. *Journal of Neuroscience*, 28(44), 11231-11239.
- Esterman, M., Chiu, Y. C., Tamber-Rosenau, B. J., & Yantis, S. (2009). Decoding cognitive control in human parietal cortex. *PNAS*, 106(42), 17974-17979.
- Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology*, 19(12), 1028-1033.
- Friston, K. J. (2011). Functional and effective connectivity: A review. *Brain Connectivity*, 1(1), 13-36.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425-2430.
- Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current Biology*, 17(4), 323-328.
- Howard, J. D., Plailly, J., Grueschow, M., Haynes, J. D., & Gottfried, J. A. (2009). Odor quality coding and categorization in human posterior piriform cortex. *Nature Neuroscience*, 12(7), 932-938.
- Hsieh, P. J., Vul, E., & Kanwisher, N. (2010). Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex. *Journal of Neurophysiology*, 103(3), 1501-1507.
- Kamitani, Y., & Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Current Biology*, 16(11), 1096-1102.
- Kay, K. N., & Gallant, J. L. (2009). I can see what you see. *Nature Neuroscience*, 12(3), 245-251.
- Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, 452(7185), 352-355.
- Kosslyn, S. M. (2005). Mental images and the Brain. *Cognitive Neuropsychology*, 22(3), 333-347.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Neuroscience*, 2(9), 635-642.
- Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., et al. (1999). The role of area 17 in visual imagery: Convergent evidence from PET and rTMS. *Science*, 284(5411), 167-170.
- Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, 378(6556), 496-498.
- Kreiman, G., Koch, C., & Fried, I. (2000). Imagery neurons in the human brain. *Nature*, 408(6810), 357-361.
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *PNAS*, 103(10), 3863-3868.
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11), 1432-1438.
- MacEvoy, S. P., & Epstein, R. A. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nature Neuroscience*, 14(10), 1323-1329.
- Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: Neuronal interactions during perception and imagery. *Cerebral Cortex*, 14(11), 1256-1265.
- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880), 1191-1195.
- Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., et al. (2010). Willful modulation of brain activity in disorders of consciousness. *New England Journal of Medicine*, 362(7), 579-589.
- Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI: An introductory guide. *Social Cognitive and Affective Neuroscience*, 4(1), 101-109.
- Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, 56(2), 400-410.
- Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, 63(6), 902-915.
- O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, 17(4), 580-590.
- Raichle, M. E. (2010). Two views of brain function. *Trends in Cognitive Sciences*, 14(4), 180-190.
- Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's

- eye: Decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825.
- Schurger, A., Pereira, F., Treisman, A., & Cohen, J. D. (2010). Reproducibility distinguishes conscious from nonconscious neural representations. *Science*, *327*(5961), 97–99.
- Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*(5), 543–545.
- Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron*, *35*(6), 1157–1165.
- Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of Neuroscience*, *29*(5), 1565–1572.
- Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited – again. *NeuroImage*, *2*(3), 173–181.

## Visualization of the Mental Imagery — A New Approach of Mental Imagery Study

Zhang Delong, Liang Bishan, Wen Xue, Huang Ruiwang, Liu Ming

(School of psychology, Center for Studies of Psychological Application, South China Normal University, Guangzhou, 510631)

**Abstract** Over the past decade, fMRI researchers have developed increasingly sensitive techniques for analyzing the information represented in BOLD activity. The current understanding of the goal of many fMRI studies has been achieved by extracting representational information from analyzing fMRI data in some specific region of the brain rather than just comparing the activation difference among experiment conditions. Multivariate pattern analysis is one of the methods which have recently emerged as a promising computational technique in neuroimaging studies. Recently, the multivariate pattern analysis method based on fMRI technology has been widely used in the field of the neuroscience, which is significantly changing the related question and methods of the studies of cognition. Specifically, the multivariate pattern analysis method will be applied to the study of mental imagery with the aim to improve the research of the essence and its related function of mental imagery processing. In this study, we introduced the principle and the history of the multivariate pattern analysis method; we showed the advantage of the multivariate pattern analysis method when compared with that of the traditional method such as the fMRI analysis based on a general linear model. Then, we introduced the different stages of the multivariate pattern analysis in the neuroimaging studies including the classification, identification, and reconstruction. In the aspect of the classification, the researchers focused on analyzing data from the decoding perspective and attempted to determine how much could be learned about the sensory stimuli, cognitive state, movement and so on, where the linear classifier such as the support vector machine and linear fisher classifier was widely used. In the aspect of the stimuli identification, the researchers were interested in understanding how activity varies in different brain regions when there is concurrent variation in the world by analyzing fMRI data with the encoding model such as the general linear model. In the aspect of the image reconstruction, the corresponding relationship between the stimuli properties and the neural activity patterns was established, and then the Bayesian model was applied to reconstruct the visual image of the stimuli from the neural activity pattern in the brain. Taken together, we indicated the developing process and the logic of the three stages. Based on this, we introduced the dilemma of the mental imagery study, and the necessity of the application of the multivariate pattern analysis method in the studies of mental imagery. Then, we analyzed the advantage of the application of multivariate pattern analysis in mental imagery study. Specifically, we discussed the basis of the approach of visualization of the mental imagery from the fMRI brain activity pattern. In this article, we introduced the basic claims of the perception prediction theory in the field of mental imagery that mental imagery serves to activate most of the same brain regions involved in the visual perception to modulate the visual stimuli processing, where the lateral occipital complex region is the critical region. The study proposed that the realization of the visualization of the mental imagery will contribute to providing the new viewpoints and methods for the study of mental imagery.

**Key words** mental imagery, fMRI, multivariate pattern analysis, decoding model, encoding model