

心理负荷的评估：基于神经人因学的视角^{*}

贾会宾 赵庆柏 周冶金

(青少年网络心理与行为教育部重点实验(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 心理负荷的评估正经历着从传统工效学向神经人因学的变革。EEG、ERPs、fMRI、fNIRS 和 TCD 等神经科学研究技术为这场变革提供了有力的工具。研究发现：在单任务情境下，随着操作者心理负荷的增加，脑电 α 波活动减弱， θ 波活动增强，并且前额皮层血流、左侧额下回血液氧合血红蛋白浓度变化均增加，大脑动脉血流速度也增快；在双任务情境下，随着操作者主任务心理负荷的增加，次任务的 N1、新异 P3 和 P3b 等 ERPs 成分波幅降低。依据这些研究成果，学者们利用人工神经网络、支持向量机等模式分类算法实现了对心理负荷的实时在线评估。但是，各种研究技术在敏感性、诊断力、主任务干扰、实施需求、可接受性和信度等方面各有优势与不足。在未来研究中，要注意促进相关技术的融合、提高其可接受性，并充分利用模式识别算法提高其诊断力和敏感性。

关键词 心理负荷；神经人因学；脑电活动；血液动力学；光学成像

分类号 B842

1 引言

1.1 心理负荷的概念

早在数十年前，心理负荷(mental workload)就引起了人们的广泛关注。1977 年北大西洋公约组织(NATO)的一些学者就针对“心理负荷的理论及测量”进行过专题研讨。学者们发现心理负荷涉及到工作要求、时间压力、操作者的能力和努力程度、行为表现等诸多因素，即心理负荷是一个多维的概念，很难对它进行简单的界定。

在实际研究中，学者们分别从主观和客观两个方面给出了心理负荷的操作定义。Rouse, Edwards 和 Hammer (1993)认为，心理负荷是一个主观的概念，是操作者知觉到的负荷水平。操作者的个体差异、动机、策略和情绪状态均会影响

到这个知觉到的负荷水平。而 O'Donnell 和 Eggemeier (1986)认为，心理负荷应该是一个客观概念，它指的是操作者完成特定工作任务时所消耗的那部分有限的信息加工资源。此外，Kramer, Sirevaag 和 Braune (1987)将心理负荷描述为执行一项任务的“成本(cost)”，而该“成本”可以依据操作者同时执行另一项同样需要认知加工资源的次任务时，能够提供给该次任务的信息加工资源总量的降低程度进行测量。这种界定对发展次任务测量等技术具有重要意义。

根据心理负荷的主观概念，研究者相继发展出了一系列的主观评定技术。而近些年来，人们越来越多地采用心理负荷的客观概念，即认为在单任务情境下，操作者的心理负荷可以依据他们完成该任务所消耗的认知资源进行评估；而在双任务情境下，主任务的心理负荷可以依据他们同时执行另一项同样需要认知加工资源的次任务时所能提供给该次任务认知加工资源总量的降低程度进行评估。

1.2 心理负荷评估技术选择与发展的标准

为了准确地对操作者心理负荷进行评估，就有必要制定心理负荷评估手段选择与发展的标

收稿日期：2012-12-19

^{*} 教育部人文社科青年项目(10YJCXLX065)；华中师范大学中央高校基本科研业务费重大培育项目(CCNU11C01005)；青少年网络心理与行为教育部重点实验室开放课题(2012B06)。

通讯作者：赵庆柏, E-mail: zqbznr@gmail.com;

周冶金, E-mail: zhouzhijin64@yahoo.com.cn

准。O'Donnell 和 Eggemeier (1986)认为此标准涉及如下几个方面: 敏感性(sensitivity)、诊断力(diagnosticity)、主任务干扰(primary task intrusion)、实施需求(implementation requirements)和可接受性(acceptability)。

敏感性指的是测量技术区分心理负荷的不同水平的能力。它包括测量技术对心理负荷变化幅度的敏感性, 以及对心理负荷变化的时间敏感性。其中, 时间敏感性极为重要, 一项时间敏感性较高的技术可以利用较短的时间内获得的数据对心理负荷进行实时或接近实时的评估。诊断力指的是测量技术区分心理负荷的不同来源, 并对不同的资源需求进行量化的能力。例如, 一项诊断力较低的测量技术可能仅为心理资源分配提供了一个总体的指标, 而一项诊断力较高的测量技术则可以敏感地区分心理负荷的不同来源。主任务干扰指的是心理负荷评估技术对操作者“主任务”绩效的影响程度。实施需求指的是一项测量技术易于实施的程度。而可接受性指的是一项测量技术不会对操作者造成繁重的负担, 容易被操作者接受的程度。

除此之外, 一项较为成熟的心理负荷评估技术应当能够在相似的实验情境下使用同质的被试得到较为一致的结论, 即具有较高的信度(Kramer, 1991)。但是在目前阶段很少有技术能够很好的符合上述所有标准。因此, 我们需要根据实际的情境和研究目的, 对心理负荷测量技术作出选择。

2 心理负荷测量技术的发展趋势

2.1 心理负荷的传统测量技术及其局限性

以认知心理学作为理论支持, 研究者相继发展出了一系列被广泛使用的心理负荷测量技术。这些传统的测量技术可以大致分为主观评定与工作测量两类。

在主观评定的过程中, 研究人员要求操作者对自己心理负荷的大小进行判断和报告。目前, 存在不少简单易行的主观评定技术, 例如 NASA 任务负荷指数(NASA Task Load Index)、主观负荷评估技术(Subjective Workload Assessment Technique)和 Cooper-Harper 量表。这些技术的优点是实施需求较低、可接受性较高。但是有研究表明它们的敏感性和诊断力存在不足, 研究者也无法完全消除其方法上的最大缺陷, 即主观性(崔凯, 孙林岩,

冯泰文, 邢星, 2008; Cain, 2007)。

工作测量的方法可以大致分为主任务测量和次任务测量两类。在主任务测量中, 研究者通过操作者的工作绩效来推断其心理负荷水平; 在次任务测量中, 操作者需要同时执行主任务和次任务, 研究者依据操作者在次任务中的行为表现来推断主任务的心理负荷水平。尽管工作测量技术在心理负荷研究中获得了广泛的应用, 但是它们至少存在以下 3 点不足: (1)主任务测量对心理负荷的不同来源的区分并不敏感, 因而其诊断力不足; (2)在次任务测量技术中, 次任务会对操作者的主任务构成干扰; (3)在使用次任务测量技术前, 需要对操作者进行训练, 而这将限制其在一些现实工作中的应用(O'Donnell & Eggemeier, 1986)。

2.2 心理负荷的神经生理评估技术及其特点

早在 20 世纪 70 年代, 心率、皮肤导电性、尿儿茶酚胺、血压和瞳孔直径等主要关注自主神经系统(autonomic nervous system, ANS)的心理生理学指标已经被用于心理负荷评估。尽管生理学测量提供了相对客观且连续的指标, 但是它们存在一些固有的缺陷, 例如心率指标的诊断力不足、可重复性不高; 血压指标的敏感性不高, 并且容易受练习伪迹的影响; 呼吸指标不仅受心理负荷大小的影响, 它同样也受到体力工作负荷的影响; 而皮电活动指标的时间敏感性低于其他的生理学指标, 并且诊断力不足(Cain, 2007; Kramer, 1991)。因此在实践中需要发展新的研究技术, 以弥补上述各种技术的不足。

随着神经科学的飞速发展, 认知科学的研究重心开始转向探索大脑认知功能的神经基础, 这一趋势也对人类工效学产生了深远的影响, 并促进了神经人因学(neuroergonomics)这门交叉学科的形成与发展。神经人因学研究的是现实世界工作状态下人的大脑与行为(Liu, Wu, & Berman, 2012)。传统的心理负荷评估注重对影响人机系统效率和安全的心理与行为规律进行研究, 而神经人因学则注重对工作状态下人类大脑认知功能的神经基础进行研究。

依据研究技术的差异, 心理负荷的神经人因学评估可以大致分为两类, 即神经电生理研究和血液动力学研究。前者使用的技术主要包括脑电(Electroencephalography, EEG)、事件相关电位(Event-Related Potentials, ERPs)等技术; 而后者

使用的技术主要包括功能磁共振成像(functional Magnetic Resonance Imaging, fMRI)、功能性近红外光学成像(functional Near-Infrared Spectroscopy, fNIRS)、经颅多普勒超声(Transcranial Doppler, TCD)等技术。

上述各种脑成像技术的研究焦点是操作者的中枢神经系统(central nervous system, CNS)。它们直接测量了与心理负荷相关的大脑神经活动,有利于加深人们对相关脑加工机制的理解。因此,在一些因变量只有行为学指标的心理负荷研究中,研究者依据已有的认知神经科学研究成果和行为数据,可以推测相应大脑皮层的活动模式。此外,与前文那些主要关注自主神经系统的生理心理学方法相比,神经成像方法可以提供更为敏感、诊断力更高的指标,且大部分技术的信度较高。因此下文将较为详细地介绍研究者利用脑电、事件相关电位等技术对心理负荷进行的神经电生理研究及利用功能磁共振成像、功能性近红外光学成像、经颅多普勒超声等技术对心理负荷进行的血液动力学研究,并从敏感性、诊断力、主任务干扰、实施需求和可接受性等5个方面比较它们与其他技术的优势与不足,对未来的神经人因学视角的心理负荷研究提供一些指导。

3 心理负荷的神经电生理研究

3.1 心理负荷的脑电(EEG)研究

脑电是通过电极记录下来的大脑神经元的自发性、节律性电活动。脑电活动依据其频率,可以分为 δ 波(0.5~3.5 Hz)、 θ 波(4~7 Hz)、 α 波(8~13 Hz)、 β 波(14~25 Hz)和 γ 波(26 Hz以上)或 δ 波(0.5~3 Hz)、 θ 波(4~7 Hz)、 α 波(8~13 Hz)、中间快波(14~17 Hz)、 β 波(18~30 Hz)和 γ 波(31 Hz以上)。研究发现, θ 波和 α 波对心理负荷的变化敏感:被试的心理负荷与其脑电 θ 节律活动成正比,而与其脑电 α 节律活动成反比(Antonenko, Paas, Grabner, & van Gog, 2010)。基于这一研究结果,学者们利用 EEG 数据,并结合判别分析、人工神经网络(Artificial Neural Network, ANN)和支持向量机(Support Vector Machine, SVM)等技术建立了可以对操作者心理负荷状态进行在线或离线评估的模式分类器(Baldwin & Penaranda, 2012; Dorneich, Mathan, Ververs, & Whitlow, 2007; Zarjam, Epps, & Chen, 2011)。这些模式分类器达到了较高的分类

准确率。例如,为了评估军队指挥官在严峻战场环境下的心理负荷,Dorneich 等人(2007)利用6个头皮电极的 EEG 及 ECG 数据建立了能够对心理负荷水平进行实时在线评估的支持向量机模型。这个模式分类器对高低心理负荷任务的分类准确率可达到95%。

从心理负荷评估手段选择与发展的几个标准来看,EEG 技术对心理负荷变化的敏感性较高。Dussault, Jouanin, Philippe 和 Guezennec (2005)在一个模拟飞行驾驶任务中探讨了脑电、心率及主观评定对心理负荷变化的敏感性。研究结果发现,随着心理负荷的增加,脑电 θ 频段的活动增强。然而,当被试的心理负荷不同时,其平均心率和主观量表评分并没有显著性差异。

虽然脑电节律(例如 α 和 θ)为唤醒状态的总体水平提供了一个相对可靠的指标,但它们对不同来源的认知需求并不能很好地区分开来,因此将脑电节律用于心理负荷评估时诊断力较低。此外,大量针对 α 和 θ 节律的 EEG 频谱分析结果表明,EEG 技术具有较高的信度(Kramer, 1991)。而且与传统的主观评定及主任务测量法相比,EEG 技术的主任务干扰程度较低。EEG 技术对测量场所的要求较低,能够在自然环境下进行研究,实施需求较低;但是为操作者戴上电极帽经常会消耗很长的时间,这将限制其在实际工作场所的应用,就此方面而言其可接受性亟待提高。

3.2 心理负荷的事件相关电位(ERPs)研究

ERPs 是对 EEG 数据进行深入处理而获得的事件相关电活动,因此与各种心理功能具有更准确的对应关系。在众多的 ERPs 成分中,用于研究人类心理负荷最多的是 P3b。P3b 成分是一个正向的 ERPs 成分,平均潜伏期大约在刺激出现后的 300 ms。P3b 成分的波幅较大,有时可以从较少的试次得到。

有研究发现,P3b 成分的潜伏期为知觉和分类的时间提供了一个相对纯粹的指标,其波幅与分配给该任务的注意资源有关(Luck, 2005; Polich, 2007)。因此在双任务情境下,随着被试主任务难度的增加,其心理负荷增加、可分配给次任务的认知加工资源降低,这使得次任务刺激所诱发的 P3b 成分的波幅降低、潜伏期延长。

基于 P3b 的实验室研究结果已用于模拟器或真实世界条件下,以实现心理负荷水平的在线

或离线评估。J. S. Song, Zhuang, G. L. Song, & Wanyan (2011)研究了飞行驾驶任务中 P3b 成分对心理负荷变化的敏感性。在实验中, 被试需要在完成一个模拟飞行任务(主任务)的同时, 执行一个视觉 oddball 任务(次任务)。在主任务中, 研究者通过设置平视显示器中呈现的目标信息的参数来对心理负荷的大小进行操纵。实验结果发现, 当主任务的难度增加时, 被试的心理负荷增加, 主任务准确率均降低、反应时均延长。与此同时, 由于被试分配给次任务的认知加工资源降低, 次任务刺激所诱发的 P3b 的波幅降低、潜伏期延长。为了对各种心理负荷评估手段的敏感性进行比较, Baldwin, Freeman 和 Coyne (2004)要求被试在模拟驾驶的同时执行一个颜色辨别的视觉 oddball 任务。结果发现, 在可见度较低(有雾)的情况下, oddball 任务诱发的 P3b 成分的波幅显著降低, 但被试 oddball 任务的反应时和正确率无显著变化; 当交通拥挤程度增加时, 行为指标的敏感性则较高, 但是 P3b 波幅却对交通流量的变化不敏感。这表明, 要实现对不同情境下心理负荷的准确评估, 需要综合应用多种评估手段。

此外, P3b 成分已经被成功地用于建立心理负荷的计算模型。例如, Wu, Liu 和 Quinn-Walsh (2008)基于产生 P300 成分的神经生理机制, 采用一个包含知觉子网络(perceptual subnetwork)、认知子网络(cognitive subnetwork)、运动子网络(motor subnetwork)的排队网络架构, 准确地解释了人类行为和双任务情境下 P300 成分波幅的变化。这表明该排队网络模型不仅具有其生理学合理性, 而且具有在实时情境下对心理负荷进行建模和预测的能力。

除 P3b 以外, 还有一些其它 ERPs 成分对心理负荷的大小和注意资源的分配敏感。Miller, Rietschel, McDonald 和 Hatfield (2011)发现, 随着被试心理负荷的增加, 每间隔 6s 至 30s 呈现的新异听觉刺激所诱发的 N1、P2 和 LPP 成分波幅降低。在 Schultheis 和 Jameson (2004)的研究中, 被试的主任务是超文本阅读任务, 次任务使用的是一个听觉 3-stimulus oddball 范式。他们发现, 随着被试主任务难度的增加, 其心理负荷水平增加, 新异听觉刺激所诱发的新异 P3 波幅降低。新异 P3 是一个类 P3b 成分, 它由意外而不经常出现的新异刺激所诱发。但是如果将新异 P3 用于实际工

效学研究, 需注意如下问题: 随着新异刺激的重复出现, 新异 P3 的波幅将逐渐降低, 而这将不利于心理负荷的准确评估。

由于 ERPs 是从 EEG 中提取出来的“事件相关电位”, 因此它的很多优势和局限性与 EEG 相同。该技术的敏感性和信度均较高, 且实施需求较低。Humphrey 和 Kramer (1994)发现, 当用于叠加平均的试次数目小于 10 时, P3b 成分对高低心理负荷的分类已经达到了很高的精确率(90%), 而如此少的试次数是可以在很短的时间内获得的。此外, 由于不同的 ERPs 成分涉及不同的认知加工机制, 因此 ERPs 技术具有相对较高的诊断力。研究者发现, 增加任务中认知过程的难度可以显著增加 P3b 的潜伏期, 但是增加反应选择阶段的难度并不能增加 P3b 的潜伏期(Magliero, Bashore, Coles, & Donchin, 1984)。

然而, 在将 ERPs 技术应用于心理负荷评估时广泛使用的双任务范式会对操作者的主任务构成干扰, 在这方面传统的主观评定及主任务测量技术均要更好一些; ERPs 技术的信噪比较低, 这限制了波幅值相对较低的 ERPs 成分在心理负荷的实时评估中的应用, 因为它们需要较多的试次才能对操作者的状态进行准确地评估。

4 心理负荷的大脑血液动力学研究

尽管 EEG 技术和 ERPs 技术敏感性较高、实施需求较低, 但是它们的空间分辨率较低, 头皮电极记录的数据被认为是大脑神经发生源的源信号在头皮混叠的结果, 因此其诊断力受到了一定程度的限制。而 fMRI、fNIRS 和 TCD 等脑成像技术恰可以弥补这些缺点, 并且它们提供的大脑血液动力学信息为研究者探索现实工作情境下人的脑机制提供了一个新的视角。

4.1 心理负荷的功能磁共振成像(fMRI)研究

随着现代神经影像学技术的发展, 以 fMRI 为代表的测量大脑血流和新陈代谢活动指标的脑成像技术已经被广泛用于研究人类的认知活动。在神经人因学研究中, fMRI 也被用于研究一些实际工作情境下心理过程的神经机制, 例如警戒(Lim et al., 2010)、心理负荷(Parasuraman & Wilson, 2008)和空间巡航(Just, Keller, & Cynkar, 2008)等。

当被试执行单一任务时, 心理负荷水平的增

加导致他们前额皮层(prefrontal cortex, PFC)及扣带前回(anterior cingulate cortex, ACC)等脑区新陈代谢活动更强。Parasuraman 和 Wilson (2008)通过对前人研究的总结发现:当操作者的心理负荷增加时,前额皮层的血流增加,因此他们指出前额皮层的血流可以作为心理负荷大小的一个指标。而另一些研究者特别强调扣带前回的作用(Onton, Delorme, & Makeig, 2005)。Badre 和 Wagner (2004)以及 Barch 等人(1997)均发现与某一个特定的认知过程相比,扣带前回似乎与一般性的认知控制和任务难度联系更为紧密,而认知控制需求和任务难度的增加将引起被试心理负荷增加。

需要注意的是,与执行单一任务相比,操作者执行由该任务和另一任务组成的双任务并不一定会导致前额皮层和扣带前回血流增加。为了探讨与模拟驾驶任务同时进行的言语理解任务对大脑活动的影响,Just, Keller 和 Cynkar (2008)要求被试在两种实验条件下完成任务。在第一种实验条件下,被试只需完成模拟驾驶任务;在第二种实验条件下,被试需要在完成上述模拟驾驶任务的同时完成一个听觉言语理解任务。实验结果发现,与单任务条件相比,在双任务条件下被试的顶叶活动减弱、颞叶活动增强,但是被试的前额皮层和扣带前回的活動并没有显著变化。他们指出,在双任务中需要对被试必须在两项任务间快速地相互转换还是同时执行进行区分。当这两项任务存在快速地相互转换时,由于执行控制需求的增加,在双任务条件下操作者前额皮层与扣带前回的血流会增加;但是当两项任务需要同时执行时,由于它们对神经加工资源的竞争,使得一些脑区的血流会降低。

由于不同的认知活动有不同的大脑激活模式,因此与其他技术相比,fMRI 技术最大的优势在于它具有相对较高的诊断力。Tomasi, Chang, Caparelli 和 Ernst(2007)比较了视觉工作记忆任务和视觉注意任务这两种不同类型的认知任务中高低心理负荷的脑激活网络的差异。研究发现:在高心理负荷的条件下,顶叶(parietal cortex)、枕叶(occipital cortex)、丘脑(thalamus)和小脑(cerebellum)在两种任务中均激活更强;前额皮层的活动随着视觉工作记忆任务心理负荷的增加而增强,而视觉注意任务心理负荷的增加并不会引起前额皮层活动的增强;中央后回(postcentral

gyri, PostCG)和枕上回(superior occipital gyri, SOG)的活动随着视觉注意任务心理负荷的增加而增强,而视觉工作记忆任务心理负荷的增加并不会引起这些脑区活动的增强。在当前复杂的现实工作中,操作者的心理负荷往往来自很多不同的方面,这便需要一项诊断力较高的评估技术,而在这方面 fMRI 技术体现出了自己的优势。另一方面,尽管目前缺乏对其信度的正式评估,但是相关的实验结果在很多研究中得到了重复验证,这表明它是具有一定信度的(Kramer, 1991)。

然而就目前阶段而言,fMRI 技术存在诸多不足之处。fMRI 技术的时间分辨率较低,因此它对心理负荷变化的时间敏感性不足;测量过程中,要求被试或操作者身体不动,而这对于大多数的现实工作是不可能的,并且扫描过程中会产生较大的噪音,会对操作者的主任务构成严重干扰;fMRI 价格昂贵并且对研究场所的要求很高,因此其实施需求较高;fMRI 并非是完全无创的,不适合对操作者进行日常的连续测量,可接受性较低。

4.2 心理负荷的功能性近红外光学成像(fNIRS)研究

fNIRS 技术利用血液的主要成分对近红外波段的光(680~1000 nm)具有良好的散射性,将光源置于头部某一部位,并在距其 2~7 cm 处放置近红外光的探测器,光源发射的光子除被大脑组织吸收外,仍会有一部分光子经过多次散射到达头皮,而探测器就可以接收到这些光信号。在近红外波段,由于氧合血红蛋白(oxyhemoglobin, oxy-Hb)和脱氧血红蛋白(deoxyhemoglobin, deoxy-Hb)的吸收光谱不同,通过测定某一脑区的光学信号变化,可以推知氧合血红蛋白和脱氧血红蛋白的变化。因此,fNIRS 技术测量的常用指标主要有 3 种:氧合血红蛋白浓度、脱氧血红蛋白浓度和总体血红蛋白指数。

目前 fNIRS 技术已经被广泛应用于人类复杂认知功能的研究。fNIRS 技术的优点在于它的时间分辨率和空间分辨率均较高,并较为便携,允许被试在测量过程中运动,设备价格也较 fMRI 低廉。这些优势使得它成为神经人因学的一项重要技术。近些年来,研究者利用 fNIRS 技术对工效学中的心理负荷问题进行了很多研究(Ayaz et al., 2010, 2012; Helton et al., 2010; Molteni, Butti,

Bianchi, & Reni, 2008), 显示出了它在现实情境中应用的潜在价值。在 Ayaz 等人(2010)的研究中, 被试需要在执行空中交通管理任务的同时, 完成一个具有 4 个心理负荷水平的 n-back 任务。他们发现, 随着心理负荷水平的增加, 被试空中交通管理任务的反应速度和准确率均单调下降, 同时他们左侧额下回的血液氧合血红蛋白浓度变化单调递增。研究者同时比较了数据通信和语音通信的 fNIRS 指标的差异。他们发现数据通信引起的平均血氧变化远小于语音通信。这个结果与他们的预期相一致, 即语音通信需要消耗更多的认知资源。

利用 fNIRS 技术提供的血液动力学信息, 研究者结合人工神经网络(ANN)和支持向量机(SVM)等分类算法建立了可以在线或离线对操作者心理负荷水平进行评估的模式分类器(Abibullaev, An, & Moon, 2011; Abibullaev, Kang, Lee, & An, 2010; Girouard, Solvovey, & Jacob, in press)。这些模式分类器达到了相对较高的分类准确率。例如 Girouard 等人(in press)报告了一个由他们开发的、可以对操作者的情绪与工作负荷状态进行在线评估的软件系统, 该软件系统对心理负荷水平进行在线评估的准确率可达到 85%以上。

为了实现对心理负荷的准确评估, 研究者需要清楚各种典型的操作者行为对 fNIRS 指标测量的影响。Solovey 等人(2009)研究了各种典型的人机交互(Human-Computer Interaction, HCI)行为对 fNIRS 测量指标的影响。他们发现, HCI 实验中的鼠标点击、键盘输入和眼动并不会影响 fNIRS 测量指标, 但是较大的头动和前额运动会对 fNIRS 指标测量构成较大影响, 因此需要努力避免或控制这些行为。同时, 尽管较小的头动、呼吸、心跳会产生一定的伪迹, 但这些伪迹可以通过滤波加以矫正。

有研究表明, fNIRS 是一项具有相对较高敏感性的技术。例如, Hirshfield 等人(2009)通过同时记录操作者的 EEG 数据和 fNIRS 数据探讨了这两项技术对不同心理负荷水平的人机交互任务进行分类的敏感性。他们发现 fNIRS 技术对不同心理负荷水平进行分类的准确率显著高于 EEG 技术。

尽管目前缺乏对 fNIRS 诊断力的正式评估, 但是由于不同类型的心理负荷有着不同的血液动

力学活动模式, 因此多通道 fNIRS 技术可能具有较高的诊断力。Abibullaev, Kang, Lee 和 An (2010)使用 fNIRS 技术记录了被试进行 4 种认知活动(心理旋转、乘法运算、单词拼写和书信构思)时前额皮层的血氧浓度, 并利用小波变换进行了特征提取。通过利用提取出的小波系数, 研究者借助支持向量机模型实现了对 4 种认知任务较为准确地分类(平均正确率 92%)。

此外, fNIRS 技术还具有如下优点: fNIRS 对操作者的运动不是特别敏感, 不会对其主任务构成干扰, 因此适合日常工作情境下的神经人因学研究; fNIRS 价格较 fMRI 低廉、较为便携, 实施需求较低; fNIRS 无创伤, 容易被操作者接受; fNIRS 技术具有相当较高的信度(Ayaz et al., 2012)。

4.3 心理负荷的经颅多普勒超声(TCD)研究

经颅多普勒超声(TCD)技术由 Aaslid, Markwalder 和 Nornes (1982)首创, 他们指出通过 TCD 技术可以无侵入性地连续测量大脑动脉血管的血流速度。TCD 技术通过使用一个沿着颞骨放置在颞骨弓上方的 2MHz 脉冲多普勒探头来测量动脉血流。主试对超声波脉冲的深度进行调节, 直到分离出来想要的颅内动脉血管。该探头发射的超声波经颅内血管反射后被该探头接收。研究者依据多普勒效应(Doppler Effect), 通过测量超声波频率的改变便可以得出相应动脉血管的血流速度。由于大脑血流速度改变与神经活动具有密切的耦合关系, 大脑某一区域血流速度的相对变化将为该脑区的新陈代谢活动提供重要的指标。

近些年来, 有研究者利用 TCD 技术研究了工效学中的心理负荷问题。Shaw, Guagliardo, de Visser 和 Parasuraman (2010)使用 TCD 技术研究了模拟空中指挥和控制任务中的心理负荷问题。在该模拟任务中, 心理负荷这一变量共有两个水平, 即低心理负荷(敌军威胁出现的速率恒定)、高心理负荷(敌军威胁出现的速率一直在变化)。在高心理负荷水平下, 敌军威胁出现的速率有两个增长长期。结果发现, 在高心理负荷水平的第一个增长长期转折点, 大脑血流速度(CBFV)增加, 但是在高心理负荷水平的第二个增长长期转折点, 大脑血流速度(CBFV)并没有增加。该研究表明, TCD 技术可以在一定程度上监控不可预期情境下操作者心

理负荷的动态变化。另外一些研究也表明警戒任务的心理负荷水平对大脑血流速度(CBFV)具有一定的调节作用(Hitchcock et al., 2003; Mayleben, 1998)。

Zinni 和 Parasuraman (2004)使用工作记忆任务和视觉运动任务检验了大脑血流速度是否可以敏感地区分不同水平的心理负荷。但是令人失望的是,大脑血流速度并没有随心理负荷的变化而系统地发生变化。在视觉运动任务中,只有低心理负荷条件下的大脑血流速度快于基线条件,并且中等心理负荷和高心理负荷条件下的大脑血流速度慢于低心理负荷条件。

由此看来,TCD 技术提供的大脑血流速度(CBFV)并不总是一个敏感且可靠的指标。但是正如 Cosenzo (2007)指出的那样,虽然它可能并不能为评估复杂认知任务的绩效提供一个最为直接的神经生理学指标,但考虑到该技术较为便携、价格低廉,且上述两个心理负荷研究中使用的被试数量较少,研究者依旧需要对该技术的可用性进行再探讨,例如可尝试对特征提取及模式分类算法的优化提高其敏感性。

从心理负荷评估手段选择与发展的几个标准来看,TCD 技术的时间敏感性相对较高,但是它似乎不能敏感地区分心理负荷的不同水平;目前缺乏对 TCD 技术诊断力的正式评估,但是它的空间分辨率较低,这无疑将限制其对心理负荷不同来源的诊断力;TCD 技术对被试是否运动没有严格限制,不会对操作者的主任务构成干扰;TCD 技术的实施需求较低;TCD 技术可以无创地监控大脑血流速度,主任务干扰较低,容易被操作者接受;尽管 Parasuraman 和 Rizzo (2008)认为该技术具有相对较高的信度,但是正如前文所述很多研究却提供了不利的证据。

5 研究展望

上述各种研究技术为心理负荷的评估提供了相对客观、连续的指标,但是在敏感性、诊断力、主任务干扰、实施需求、可接受性和信度等方面,它们各有优势与不足。迄今为止,并没有一项技术能够达到最为理想的标准。

在未来,为了对心理负荷进行准确地在线或离线评估,研究者应注意如下问题:

首先,鉴于上述各种神经科学研究技术各有

优缺点,研究者需要发展它们的融合技术,以期同时获得两项优势互补的神经生理学技术的数据,例如 EEG-fMRI、EEG-fNIRS。尽管近年来 EEG-fMRI 融合研究获得了很多成果,但是依旧存在很多未克服的技术难题,例如心冲击伪影(ballistocardiographic artifact, BCGa)和 MRI 脉冲伪影(MRI pulse artifact)。与之相比,EEG-fNIRS 融合技术并没有很大的技术难题,通过该技术研究者可以同时获得操作者的 EEG 数据和各个脑区的血液动力学信息。

其次,神经人因学要求各种认知神经科学实验技术在实际的工作场所评估心理负荷时的主任务干扰和实施需求较低、易为操作者所接受。因此,研究者需要对上述各种技术进一步改进,以提高其在野外环境的可推广性,例如发展无线遥感技术、干电极脑电信号记录技术,并在保证其性能的前提下,降低造价。

再次,为了提高各种研究技术的敏感性和诊断力,研究者需要对神经生理信号的特征提取和模式分类算法进一步优化,并提高其运算速度。例如,在未来利用 ERPs 技术对心理负荷进行实时监控时,研究者可以充分利用独立成分分析(independent component analysis, ICA)算法提高该技术的敏感性。Xu 等人(2004)的研究表明与利用头皮电极记录的 ERPs 数据相比,利用与 P3b 成分关联程度较高的独立成分 ERPs 数据进行 P3b 成分检测能获得更为准确的结果。而利用近些年出现的多体素模式分析(multi-voxel pattern analysis, MVPA)将有望提高 fMRI 的诊断力。MVPA 方法能够基于多个体素信号的空间模式对个体的认知活动进行“解码”,实现由大脑活动模式反推个体在这段时间内的认知活动。已有研究表明,将该数据分析方法用于心理负荷评估,可以大大提高 fMRI 的敏感性和诊断力(Bååth, 2009)。

最后,由于上述各种认知神经科学研究技术存在一些固有的不足,研究者不应将自己的视野局限于它们。尽管目前使用分子遗传学、生物工程和纳米技术进行心理负荷的神经人因学研究并不是很多,但是它们必将在未来的研究中发挥越来越大的作用(Parasuraman & Rizzo, 2008)。例如,针对认知能力个体差异的分子遗传学结果可以为人员的选拔、培训以及人机交互界面设计提供重要依据(Parasuraman, 2009)。

参考文献

- 崔凯, 孙林岩, 冯泰文, 邢星. (2008). 脑力负荷度量方法的新进展述评. *工业工程*, 11(5), 1-5.
- Aaslid, R., Markwalder, T. -M., & Nornes, H. (1982). Noninvasive transcranial doppler ultrasound recording of flow velocity in basal cerebral arteries. *Journal of Neurosurgery*, 57, 769-774.
- Abibullaev, B., An, J., & Moon, J.-I. (2011). Neural network classification of brain hemodynamic responses from four mental tasks. *International Journal of Optomechatronics*, 5(4), 340-359.
- Abibullaev, B., Kang, W.-S., Lee, S. H., & An, J. (2010, April). Functional near infrared spectroscopy based cognitive task classification using support vector machines. *Paper presented at the meeting of 2010 5th International Symposium on Health Informatics and Bioinformatics (HIBIT)* (pp. 7-12). Antalya, Turkey.
- Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, 22, 425-438.
- Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage*, 59, 36-47.
- Ayaz, H., Willems, B., Bunce, S., Shewokis, P. A., Izzetoglu, K., Hah, S., Onaral, B. (2010). Cognitive workload assessment of air traffic controllers using optical brain imaging sensors. In T. Marek, W. Karwowski, & V. Rice (Eds.), *Advances in understanding human performance: Neuroergonomics, human factors design, and special populations* (pp. 21-31). Boca Raton, FL: CRC Press, Taylor & Francis Group.
- Bååth, R. A. (2009). *Predicting working memory load from fMRI data*. Unpublished master's thesis, Lund University.
- Badre, D., & Wagner, A. D. (2004). Selection, integration, and conflict monitoring: Assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron*, 41, 473-487.
- Baldwin, C. L., Freeman, F. G., & Coyne, J. T. (2004). Mental workload as a function of road type and visibility: Comparison of neurophysiological, behavioral, and subjective indices. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, 2309-2313.
- Baldwin, C. L., & Penaranda, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *Neuroimage*, 59, 48-56.
- Barch, D. M., Braver, T. S., Nystrom, L. E., Forman, S. D., Noll, D. C., & Cohen, J. D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia*, 35, 1373-1380.
- Cain, B. (2007). *A review of the mental workload literature* (Tech. Rep. No. RTO-TR-HFM-121-Part-II). Toronto, Canada: Defence Research and Development.
- Cosenzo, K. A. (2007). The effects of multitasking on cerebral hemodynamics and task performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 51, 239-243.
- Dorneich, M. C., Mathan, S., Ververs, P. M., & Whitlow, S. D. (2007). An evaluation of real-time cognitive state classification in a harsh operational environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 51, 146-150.
- Dussault, C., Jouanin, J. -C., Philippe, M., & Guezennec, C. -Y. (2005). EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviation, Space, and Environmental Medicine*, 76, 344-351.
- Girouard, A., Solovey, E. T., & Jacob, R. J. K. (in press). Designing a passive brain computer interface using real time classification of functional near-infrared spectroscopy. *International Journal of Autonomous and Adaptive Communications Systems*.
- Helton, W. S., Warm, J. S., Tripp, L. D., Matthews, G., Parasuraman, R., & Hancock, P. A. (2010). Cerebral lateralization of vigilance: A function of task difficulty. *Neuropsychologia*, 48, 1683-1688.
- Hirshfield, L. M., Chauncey, K., Gulotta, R., Girouard, A., Solovey, E. T., Jacob, R. J. K., Fantini, S. (2009, July). Combining Electroencephalograph and functional Near Infrared Spectroscopy to explore users' mental workload. *Paper presented at the 13th International Conference on Human-Computer Interaction* (pp. 239-247). San Diego, CA.
- Hitchcock, E. M., Warm, J. S., Matthews, G., Dember, W. N., Shear, P. K., Tripp, L. D., Parasuraman, R. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science*, 4, 89-112.
- Humphrey, D. G., & Kramer, A. F. (1994). Toward a psychophysiological assessment of dynamic changes in mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36(1), 3-26.
- Just, M. A., Keller, T. A., & Cynkar, J. (2008). A decrease in brain activation associated with driving when listening to someone speak. *Brain Research*, 1205, 70-80.
- Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 279-328). London: Taylor & Francis.

- Kramer, A. F., Sirevaag, E. J., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(2), 145–160.
- Lim, J., Wu, W. -C., Wang, J. J., Detre, J. A., Dinges, D. F., & Rao, H. Y. (2010). Imaging brain fatigue from sustained mental workload: An ASL perfusion study of the time-on-task effect. *Neuroimage*, 49, 3426–3435.
- Liu, Y. L., Wu, C. X., & Berman, M. G. (2012). Computational neuroergonomics. *Neuroimage*, 59, 109–116.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, 21, 171–186.
- Mayleben, D. W. (1998). *Cerebral blood flow velocity during sustained attention*. Unpublished doctoral dissertation, University of Cincinnati.
- Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, 80(1), 75–78.
- Molteni, E., Butti, M., Bianchi, A. M., & Reni, G. (2008, August). Activation of the prefrontal cortex during a visual n-back working memory task with varying memory load: A near infrared spectroscopy study. *Paper presented at the meeting of 30th Annual International IEEE EMBS Conference* (pp. 4024–4027). Vancouver, British Columbia, Canada.
- O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance*, Vol. 2: *Cognitive Processes and Performance* (chap. 42, pp. 1–49). New York: Wiley-Interscience.
- Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *Neuroimage*, 27, 341–356.
- Parasuraman, R. (2009). Assaying individual differences in cognition with molecular genetics: Theory and application. *Theoretical Issues in Ergonomics Science*, 10, 399–416.
- Parasuraman, R., & Rizzo, M. (2008). *Neuroergonomics: The brain at work*. New York: Oxford University Press.
- Parasuraman, R., & Wilson, G. F. (2008). Putting the brain to work: Neuroergonomics past, present, and future. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50, 468–474.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical neurophysiology*, 118, 2128–2148.
- Rouse, W. B., Edwards, S. L., & Hammer, J. M. (1993). Modeling the dynamics of mental workload and human performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23, 1662–1671.
- Schultheis, H., & Jameson, A. (2004). Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods. In P. DeBra & W. Nejdl (Eds.), *Adaptive hypermedia and adaptive web-based systems* (Vol. 3137, pp. 225–234). Berlin: Springer.
- Shaw, T. H., Guagliardo, L., de Visser, E., & Parasuraman, R. (2010). Using transcranial doppler sonography to measure cognitive load in a command and control task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54, 249–253.
- Solovey, E. T., Girouard, A., Chauncey, K., Hirshfield, L. M., Sassaroli, A., Zheng, F., Jacob, R. J. K. (2009, October). Using fNIRS brain sensing in realistic HCI settings: Experiments and guidelines. *Paper presented at the meeting of Proceedings of the 22nd annual ACM symposium on User interface software and technology Victoria* (pp. 157–166). British Columbia, Canada.
- Song, J. S., Zhuang, D. M., Song, G. L., & Wanyan, X. R. (2011, October). Pilot mental workload measurement and evaluation under dual task. *Paper presented at the meeting of 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)* (pp. 809–812). Shanghai, China.
- Tomasi, D., Chang, L., Caparelli, E. C., & Ernst, T. (2007). Different activation patterns for working memory load and visual attention load. *Brain Research*, 1132, 158–165.
- Wu, C. X., Liu, Y. L., & Quinn-Walsh, C. M. (2008). Queuing network modeling of a real-time psychophysiological index of mental workload—P300 in event-related potential (ERP). *IEEE Transactions on Systems, Man, and Cybernetics*, 38, 1068–1084.
- Xu, N., Gao, X. C., Hong, B., Miao, X. B., Gao, S. K., & Yang, F. S. (2004). BCI competition 2003—data set IIb: Enhancing P300 wave detection using ICA-based subspace projections for BCI applications. *IEEE Transactions on Biomedical Engineering*, 51(6), 1067–1072.
- Zarjam, P., Epps, J., & Chen, F. (2011, August–September). Characterizing working memory load using EEG delta activity. *Paper presented at the meeting of 19th European Signal Processing Conference (EUSIPCO 2011)* (pp. 1554–1558). Barcelona, Spain.
- Zinni, M., & Parasuraman, R. (2004). The effects of task load on performance and cerebral blood flow velocity in a working memory and a visuomotor task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, 1890–1894.

Mental Workload Assessment: From the Perspective of Neuroergonomics

JIA Huibin; ZHAO Qingbai; ZHOU Zhijin

(¹ Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan 430079, China)

(² Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology,
Central China Normal University, Wuhan 430079, China)

Abstract: The assessment of mental workload is undergoing transition from traditional ergonomics to neuroergonomics. The neuroimaging techniques such as electroencephalography (EEG), event-related potentials (ERPs), functional magnetic resonance imaging (fMRI), functional Near-infrared spectroscopy (fNIRS) and transcranial doppler (TCD) provide the strong supports for this revolution. It is found that under single-task condition, with the increase of mental workload level, alpha-band power decreases, theta-band power increases. In addition, the cerebral blood flow in the regions of the prefrontal cortex (PFC), the average oxygenation changes in the regions of left inferior frontal gyrus (LIFG) and the cerebral blood flow velocity (CBFV) also increase. Under dual-task condition, with the increase of mental workload level, the amplitudes of several ERPs components elicited by the secondary task stimuli (e.g., N1, Novelty P3, P3b, etc) decrease. Based on these findings, researchers have achieved the real-time online evaluation of mental workload by pattern classification algorithms (e.g., artificial neural network, support vector machine). But all of these techniques have their own advantages and disadvantages on several aspects (i.e., sensitivity, diagnosticity, primary task intrusion, implementation requirements, acceptability and reliability). In the future, researchers should promote the combination of these neuroimaging technologies, improve their acceptability and enhance their sensitivity and diagnosticity via pattern recognition algorithms.

Key words: mental workload; neuroergonomics; brain electrical activity; hemodynamics; optical imaging