

· 研究进展 ·

# “视听觉信息的认知计算”重大研究计划结题综述

吴国政<sup>1\*</sup> 韩军伟<sup>1</sup> 邓方<sup>1</sup> 李建军<sup>1</sup> 张兆田<sup>1</sup>  
辛景民<sup>2</sup> 郑南宁<sup>2</sup>

(1. 国家自然科学基金委员会 信息科学部, 北京 100085;  
2. 西安交通大学, 西安 710049)

**[摘要]** 本文介绍了“视听觉信息的认知计算”重大研究计划的立项背景、总体科学目标、总体布局、实施思路及总体完成情况,并概述了该领域下一步发展的建议。

**[关键词]** 重大研究计划;视觉信息;听觉信息;自然语言;脑机接口;认知计算;综述

视听觉信息是人类获取外部信息的主要来源,对它们的机器理解与处理是人工智能的重要技术,并在社会发展、国民经济和国家安全等领域中扮演着十分重要的角色,具有重大战略意义和深远影响。由于计算机与人类大脑的结构和信息处理方式不同,尽管计算机的计算能力提升很快,但计算机的认知能力远远滞后,如计算机在类似人类对非结构化信息感知与动态环境交互方面的认知能力还不及3岁儿童。针对这一瓶颈问题,研究和探索人类视听觉的认知机理与建立可计算视听觉认知模型对人工智能的核心算法有重大启示。

## 1 立项实施情况

### 1.1 立项背景

人工智能是引领未来的战略性技术,世界主要发达国家把发展人工智能都作为提升其国家竞争力、维护国家安全的重大战略,加紧制定规划和政策。研究人脑的感知、识别、学习、联想、记忆、推理等认知功能及其神经机制一直是信息领域重大挑战性问题,而认知计算则是其中的核心基础科学问题。

探索认知计算需要信息科学与认知科学交叉融合,近二十余年世界发达国家都高度重视和大规模地支持这方面的研究。1997年日本“脑科学时代计划”将认知、运动、行为的计算原理作为重点研究领

域,2001年美国国家科学基金委和商务部共同提出将纳米技术、生物技术、信息技术和认知科学(NBIC)四大前沿技术进行融合发展的科学计划,称之为“提高人类素质的聚合技术”。2003年欧盟开始认知视觉系统与生物启发的智能信息系统研究,2004年韩国将智能机器人列入十大经济发展引擎之一。2013年美国公布了“推进创新神经技术脑研究计划”,2013年欧盟推出了由26个国家的135个合作机构参与、预期10年的“人类脑计划”。2014年日本也启动了为期10年的神经科学研究计划。特别是,2016年美国连续发布《为人工智能的未来做好准备》、《国家人工智能研究和发展战略计划》和《人工智能、自动化与经济报告》等三份人工智能发展报告,2019年1月美国商务部正式实施14类具代表性技术出口管制,其中与人工智能相关的技术就有7类。

另一方面,2006年2月国务院颁布《国家中长期科学和技术发展规划纲要(2006—2020年)》将“脑科学与认知科学”列入优先部署的八大前沿科学问题之一,2006年3月国家自然科学基金“十一五”发展规划将认知科学与智能信息处理列为13个综合交叉领域之一。2016年3月国家发布的“十三五”规划(2006—2020年)中将“脑科学与类脑研究”列为“科技创新2030重大项目”。2017年7月,国务院发布《新一代人工智能发展规划》,将人工智能

收稿日期:2019-03-07;修回日期:2019-06-05

\* 通信作者,Email:wugz@nsfc.gov.cn

上升为国家重要发展战略。

早在2000年前后,人工智能技术尚未形成当今席卷全球范围的研究热潮时,国家自然科学基金委员会(以下简称自然科学基金委)及相关专家就以敏锐的学术洞察力和战略前瞻眼光开始了相关布局。历经8年多的研讨论证和前期研究积累,自然科学基金委于2008年8月正式启动重大研究计划“视听觉信息的认知计算”,这是我国在人工智能基础研究领域发展的里程碑之一,标志着中国人工智能科学研究“国家队”的正式组建。

## 1.2 总体科学目标

与人类视听觉感知密切相关的图像、语音和文本(语言)信息在社会、经济和国家安全等领域中扮演着重要角色,并在今后一段时间内仍将迅猛增长。这类信息可被人类直接感知和理解,也可用计算机进行处理,但计算机的处理能力远逊于人类且处理效率远不能满足当今社会的发展需求。

为此,本重大研究计划的总体科学目标是:围绕国家重大需求,充分发挥信息科学、生命科学和数理科学的交叉优势,从人类的视听觉认知机理出发,研究并构建新的计算模型与计算方法,提高计算机对非结构化视听觉感知信息的理解能力和海量异构信息的处理效率,克服图像、语音和文本(语言)信息处理所面临的瓶颈困难,为确保国家安全与公共安全、推动信息服务及相关产业发展以及提高国民生活与健康水平做出重要贡献。具体表现为:在视听觉信息处理的基础理论研究方面取得重要进展;在视听觉信息协同计算、自然语言(汉语)理解以及

与视听觉认知相关的脑机接口等三项关键技术方面取得重大突破;集成上述相关研究成果,研制具有自然环境感知与智能行为决策能力的无人驾驶车辆验证平台,主要性能指标达到世界先进水平,从而提升我国在视听觉信息处理领域的整体研究实力,培养具有国际影响力的优秀人才与团队,为国家安全和社会发展提供相关研究环境与技术支撑。

## 1.3 总体布局和实施思路

本重大研究计划以视听觉感知相关的图像、语音和文本(语言)信息为研究对象,以解决计算机对这类非结构化复杂感知信息的理解能力和海量异构信息的处理效率为主攻目标,其的特点在于以科学任务带动视听觉信息的认知计算领域的基础研究和相关关键技术的突破,并将无人驾驶智能车辆、脑机接口等作为集中验证和体现本重大计划研究成果的物理载体。

本重大研究计划在实施过程中,始终遵循“有限”目标、稳定支持,集成升华、跨越发展”的总体思路,以解决视听觉信息的“表达”与“计算”为研究关键,围绕“感知特征提取、表达与整合”、“感知数据的机器学习与理解”和“多模态信息协同计算”等核心科学问题作为重点资助领域和方向进行项目部署。通过对上述核心科学问题的难点和要点进行深入分析和分解,确定优先研究方向,开展信息科学与生命科学、数理科学等交叉领域的前沿基础、关键技术和集成升华等创新性研究。研究方向的选择原则包括:优先支持针对核心科学问题具有创新思路的研

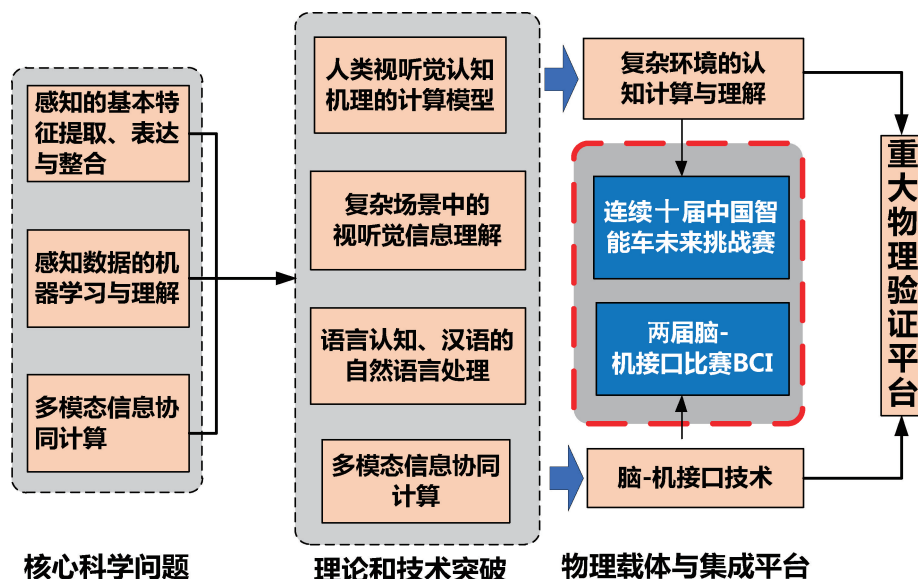


图1 重大研究计划“视听觉信息的认知计算”实施路线示意图

究;优先支持基础较好、条件较为成熟,近期可望取得突破性进展的研究;优先选择对实现本重大研究计划总体目标起决定作用的研究方向开展的跨学科集成研究。

本重大研究计划的项目立项工作主要集中分前后两个阶段。在启动的前4年间(2008年至2011年),及早部署目标需求导向的年度项目指南和研究工作,根据申报项目的创新/思想、研究价值以及对研究计划总体目标的贡献,采取“培育项目”、“重点支持项目”和“集成项目”等不同项目类型形式予以资助(项目类型体现在资助强度和实现目标上有所不同)。本重大研究计划在实施期间的第4年(2012年)进行了中期评估工作安排。根据中期评估的结果,对研究计划的实施进行了局部调整,中期评估后的4年间(2013年至2016年),项目立项选择对实现重大研究计划总体目标有决定作用的研究方向,在前期“培育项目”和“重点支持项目”成果的基础上,重点开展集成创新研究,采取“集成项目”与“重点支持项目”相结合的形式,以较大支持强度予以资助。特别是,在整个研究计划的后2年重点进行项目和成果的集成升华。

## 2 总体完成情况

十年来,在本重大研究计划指导专家组、管理工作组和承担项目科学家的共同努力下,共资助研究项目97项,其中集成项目5项、重点支持项目30项、培育项目62项。资助项目人员发表论文2255篇,其中在*PNAS*、*Neuron*、*Brain*、*IEEE T-PAMI*等知名期刊上发表论文300余篇,ESI高被引论文35篇。发表论文中SCI收录956篇,EI收录1276篇,其中20篇代表性论文WOS他引4452次,Google总引用9144次,获得知名国际期刊和会议最佳论文奖等学术奖励14项,其中包括*Brain*封面论文。出版外文专著36部,中文专著27部。申请国家发明专利532项,其中专利授权229项;申请美国等国外发明专利9项,其中专利授权6项。本重大研究计划的相关成果获国家自然科学奖二等奖5项、国家科技进步奖二等奖3项、国家技术发明二等奖1项、省部级一等奖6项、其他奖励32项。相关项目组已成为《中文语音合成系统通用技术规范》国家标准制定单位之一,并加入了国际标准组织W3C中语音合成标注语言(SSML)和情感标注语言(EMOXG)两个标准工作组。

本重大研究计划共培养博士毕业343人、硕士毕业722人,博士后出站40人。资助项目及项目团队产生的一批优秀人才,其中包括中国科学院院士4人、IEEE Fellow 9人、自然科学基金委杰出青年基金获得者6人、自然科学基金委青年基金获得者4人、教育部长江学者奖励计划教授4人、国家万人计划领军人才3人、中组部青年拔尖人才计划2人、中科院百人计划1人等。

特别是,为了进一步推动研究工作走出实验室、产生原创性重大成果,本重大研究计划创建了两个比赛平台“中国智能车未来挑战赛”和“中国脑—机接口比赛”,并组织了10届“中国智能车未来挑战赛”<sup>[1]</sup>和2届“中国脑—机接口比赛”。通过在真实的物理环境中验证理论成果,解决实际环境中复杂认知和智能行为决策等问题,改变了简单的论文汇总或实验室成果演示的传统模式,促进了应用基础研究与物理可实现系统的有机结合,有力地推动了我国智能车研发从实验室走向现场交流探索阶段,为我国在未来智能汽车技术和产业上的原始创新和发展起到了积极作用,同时也探索了新的科研项目管理模式。

总体上,本重大研究计划在认知基础理论和模型、视听觉信息处理、自然语言(汉语)理解等核心难点问题方面取得一系列标志性成果,并将无人驾驶智能车辆、脑机接口等作为集中验证该重大计划研究成果的物理载体,在脑机接口、无人驾驶平台与智能车等两个重大应用问题上已取得的战略性和前瞻性成果,在无人驾驶智能车、载人航天、医疗健康和公共安全等领域得到示范应用,在无人系统、智慧城市等国家级应用系统中发挥了重要作用。因此,本重大研究计划极大地促进了我国在信息科学等研究领域的原始创新与发展,显著提升了我国在视听觉认知机理和和视听觉信息计算领域的整体研究水平和国际影响力,有力地推动了我国智能车研发从实验室走向现场交流探索阶段,为我国在未来智能汽车技术和产业上的原始创新和发展起到了积极作用。同时,该重大研究计划培养了大批科研人才,为我国人工智能的发展奠定了坚实的基础,引领和推动了我国新一代人工智能国家战略的确立和发展。

本重大研究计划在理论和技术、重大物理验证平台、助力国家战略等三个层面,从认知计算、视觉信息处理、脑机接口、自然语言处理、多模态信

息协同计算和无人驾驶智能车验证平台等六个方面对科学问题的基础理论和关键技术开展了深入研究,所取得的主要创新性成果和实质性贡献如下。

### 2.1 突破多项核心理论和技术基础难题,为物理验证平台提供支撑

在认知机理研究上,丰富和发展了一系列脑认知基础理论,为本重大研究项目后续的脑机交互技术和集成平台提供了理论支撑。依据陈霖 global-first 的拓扑性质知觉理论,就注意过程操作的基本单元问题,认为注意操作的基本单元是知觉物体,并提出“知觉物体”的拓扑学定义,对传统视觉通路理论提出了挑战;根据知觉物体的拓扑学定义开展的一系列实验,提出注意瞬脱的拓扑学解释,可以解释目前注意瞬脱研究中许多著名难题如“lag-1 sparing”现象等<sup>[2]</sup>;发现视听觉融合也是基于特定的时间尺度,即通过跨模态地实时调整时间窗口的相位来跟踪处理自然视听觉流,首次在人类受试上找到和发现可能解决“鸡尾酒会效应”(听觉场景分析问题)的神经机制<sup>[3]</sup>。

在视听觉信息处理方面,针对视觉注意力研究,建立了多特征最优融合的条件随机场模型,实现了显著性目标检测中 Top-down 的注意力控制,并发展出视觉显著性检测研究的新方向,解决了视觉感知数据的显著目标及其精细轮廓的自动提取与语义标注等重要理论问题;针对人类初级视皮层神经元非经典感受野的非结构化建模,建立了若干视知觉加工计算模型;建立了用于定量评测显著性目标检测算法的大规模图像数据库(20 000 多幅图像),成为本领域公认的、使用最广泛的评测视觉注意力检测算法的基准数据库之一<sup>[4]</sup>;针对视觉显著度问题,建立了针对眼动的计算模型;开展了视觉显著度与眼动计算模型研究,探索了注意对复杂场景视觉信息加工的调节作用以及视觉显著图的产生机制;针对听觉感知机理与音频场景分析,提出了若干语音处理方法;通过分析视觉通路中最早阶段细胞的输出,建模自然图像的感知差异,提出了新的图像质量感知模型,并应用到医学成像和视频编码等领域<sup>[5]</sup>。面向无人驾驶车辆的交通场景退化问题,提出一系列交通标识图文识别与理解模型;针对视听觉基本特征的提取与选择,在物体识别与图像内容理解、复杂场景下运动目标的行为分析等问题上取得创新结果<sup>[6]</sup>;面对复杂多变的驾驶

场景,开发了系列在复杂环境光条件下颜色恒常知觉模型的图像增强方法;围绕着为无人车的智能行为决策提供辅助听觉信息,研究行驶中车辆内外声音的自动检测、识别和理解等方面的各项关键技术,为后续开展的无人驾驶集成平台提供强力技术支撑。

在多模态信息处理方面,借鉴新的工作记忆机制,建立了融合“自下而上”和“自上而下”的信息选择性注意模型,提出了基于超图及其相关矩阵的表示、高维异构特征稀疏选择及基于几何特点的流形学习、大规模机器学习和多模融合等高效计算模型及方法,研发跨媒体搜索引擎;在跨媒体复杂关联关系表达与建模、高维特征降维与选择、跨媒体检索、跨媒体内容挖掘等方面取得了一系列研究成果,提出了基于混合超图的复杂关系表示和非负矩阵分解的主题建模、高维特征结构性稀疏选择、基于几何观点的流形学习、跨媒体耦合字典学习与多模态哈希索引、基于跨媒体双向结构学习的排序机制、基于跨媒体关联的视频内容与文本内容对齐、以地理位置为核心的跨媒体数据挖掘与知识抽取等算法和具体方法,搭建了跨媒体垂直搜索引擎平台搭建,并在新华社和国家知识产权局等单位得到了应用;相关成果应用于“百万册数字图书馆的多媒体技术和智能服务系统”。

在自然语言理解方面,从自然语言的语义表示与计算方法到复杂场景下的语音识别、口语理解和人机交互,以及多媒体信息融合,提出了一系列创新性的方法和模型。开发相关的汉语语言理解、标准数据库和语言交互系统,为公共安全、医疗健康提供技术支撑;面向大规模中文语言知识库建设,制定了汉语篇章标注规范,提出了多层次主题模型,从词汇、语句和篇章三个层次进行了语言建模;深入研究了多源混响场景下语音增强和分离方法,提出了复杂多变声学环境下说话人和语音识别的新方法;以语音识别技术、人脸表情跟踪技术、基于语音与表情的双模态情感识别技术、多模态对话管理技术和多模态数字虚拟人表达技术为基础,构建了多模态人机交互系统;设计制定了汉语篇章标注规范,并标注了一万余篇汉语篇章语料;参与制定了《中文语音合成系统通用技术规范》国家标准;开展多模态信息协同的高效计算模型和方法研究,并在国家知识产权局检索与服务系统示范应用。

### 2.2 构建两项重大集成验证平台,为理论和技术验证提供支撑

认知机理的研究成果可以通过生物学与认知心理学实验来验证,而受认知机理和神经科学启发、抽象和简化得到的可计算模型只能通过构建人工系统进行验证。为此,本重大研究计划构建了无人驾驶集成创新验证平台和脑机接口集成创新验证平台,验证了复杂环境的感知和认知理解、适应性控制与决策等一系列视听觉信息处理的新模型和算法,促使我国无人驾驶智能车实现从简单封闭道路到真实开放道路交通环境的自主行驶。

在脑机接口方面,开创性地提出了高维编码的ERP字符拼写范式,比传统的P300范式输入速度

提高20%以上,正确率90%;设计了基于三维编码的64字符和125字符两种P300-Speller,达到了每分钟54比特的信号传输率;提出并实现了伪随机编码调制视觉诱发电位脑机接口技术,实现了32目标、信息传输率最高达到123比特/分<sup>[7]</sup>;实现了“模拟阅读”脑机字符录入范式;提出仿射聚类的脑功能多体素分析方法等脑电信号处理理论和相关算法<sup>[8]</sup>;利用静息fMRI数据构建了全脑静息功能连接模式,研究表明脑网络多变量模式分析有助于揭示脑认知的神经机制、全脑静息功能连接网络可作为脑认知状态解码的潜在生物学标记<sup>[9]</sup>;实现了多模态脑机接口与多自由度控制,成功开发了脑机接口网络浏览器与邮件系统;建立了基于稀疏表示的

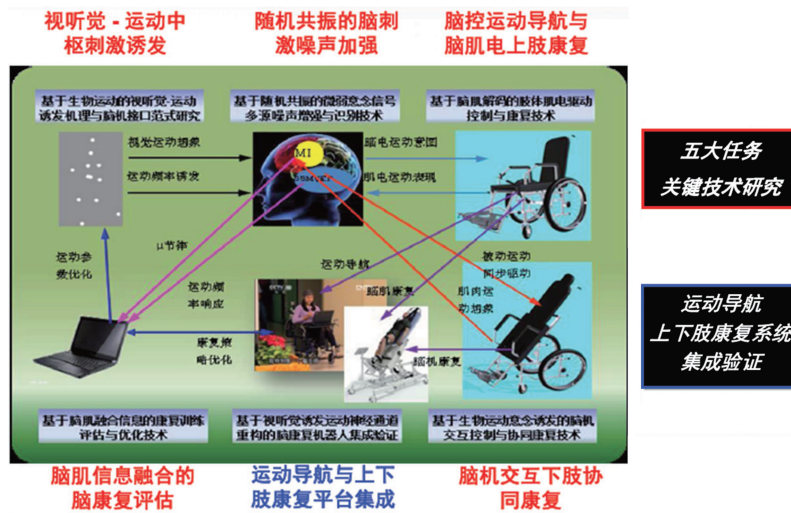


图2 基于视听觉诱导运动神经通道重建的脑控康复机器人系统



图3 我国无人驾驶智能车技术发展历程

脑信号特征选择算法;探索了人脸识别中的视听觉信息整合对人脸语义特征的神经表征的作用;提出多种脑机接口范式,用于意识障碍患者的意识检测与辅助诊断<sup>[10]</sup>;在面向高性能人机交互系统的脑机交互、肌机交互、脑肌交互等,开展了复杂神经信息检测识别与外部机器设备反馈调控的系统研究;设计高效编解码范式与方法,开发新型主动式、反应式脑机接口技术;结合神经肌骨模型算法,开发了机电助行康复与稳定性评价新方法;开展基于脑功能网络、脑肌电相干性的可塑性变化检测与分析研究,建立新型康复评价的客观标准;建立了面向任务、基于跨感觉诱发事件相关电位信号的在线实验平台,并尝试用于残障病人的肢体康复;建立了双向多维脑机接口实验范式和可实现双向信息交互与范式整合的侵入式脑机接口系统研究平台;研制了直流耦合式16通道USB信号采集模块和直流耦合式8通道无线信号采集模块,开发了一套完整的便携式脑机接口系统,系统关键技术指标达到国际先进商用产品水平;创办并举办2届“中国脑机接口比赛”,实现国际上首次实时脑机接口比赛,促进了团队间的学术交流和技术融合。

在无人驾驶车辆集成验证平台方面,以视听觉认知的基础理论和模型的基础上,基于视听觉信息认知机理的模型和理论突破,集成多传感器跨模态跨尺度信息融合、交通环境理解及运动决策等多项关键技术,构建了面向高速公路、城市道路、乡村道路和多车交互协调的多个无人驾驶车辆验证平台,为视听觉信息处理理论和方法提供了集成验证平台,复杂环境的感知和认知理解、适应性控制与决策以及智能系统硬件与架构等一系列视听觉信息处理的新模型和算法推进了无人驾驶技术的进步,实现了我国无人驾驶智能车从早期的简单封闭道路到真实开放的道路交通环境的自主行驶;建立了离线/在线无人车认知及行为能力测评体系及综合测试环境,突破了环境感知、行为规划和智能控制等关键技术,主要性能指标接近或达到了国际先进水平。无人车集成验证平台实现了自然环境中连续行驶逾200公里,人工干预里程比例 $<4\%$ ;在高速公路环境中连续行驶逾2000公里,人工干预里程比例 $<3\%$ ;创办并连续组织了10届“中国智能车未来挑战赛”,实现了应用基础研究与物理验证系统有机结合,催生了一批满足国家重大需求的理论与技术成果,在国防、智能辅助安全驾驶等相关领域得到应用

并产生积极重要影响,为国家安全与国民经济发展做出了重大贡献。

### 2.3 助力国家战略和科技发展,为产业化、医疗、国防安全、人才培养提供支撑

我国新一代人工智能发展规划的一些重要内容是本重大计划研究方向的延伸和发展。2017年7月,国务院发布《新一代人工智能发展规划》,将我国人工智能技术与产业的发展上升为国家战略,标志着我国人工智能研发战略的全面启动。2017年11月,科技部公布了国家“新一代人工智能发展规划”首批四个开放创新平台,其中包括自动驾驶、城市大脑、医疗影像、智能语音等4大人工智能核心技术,而本重大研究计划的研究内容涵盖了其中3个平台。本重大研究计划早在2008年立项之初,即已确立了以“自动驾驶”和“脑机接口”技术为两大示范集成平台的总体研究思路,开国内“自动驾驶”领域研发之先河,为本领域的产业引领和人才储备做出了实质性贡献。在“智能语音”及“康复医疗”等领域,本重大计划也实现了基础理论、核心技术及示范应用上的突破,有力提升了我国在相关研究领域的国际声誉。通过本重大研究计划10年的实施,有力引领和推动了我国人工智能国家战略的确立和人工智能产业的良性发展。

在重大研究计划启动初期,我国没有一家无人车领域的企业。2018年,我国无人驾驶技术创新创业企业已经在神州大陆多地开花,中国无人驾驶技术发展的背后是重大研究计划物理验证平台与人才培养的厚积薄发。在本重大计划的推动下,经过十年发展,如今一大批国内汽车制造公司(如一汽、上汽、北汽、东风、长安、吉利、奇瑞、宇通等)、信息科技公司(如华为、百度、阿里巴巴、腾讯等)以及创业公司(如驭势科技、武汉光庭等)纷纷积极致力于无人驾驶、自动驾驶等智能车及相关关键技术的研发创新。不少国产汽车制造企业(如广汽、吉利、奇瑞、上汽、比亚迪、江淮、长安、安科等)纷纷与本计划的一些课题组开展无人车技术合作研发,且2015年在江苏省常熟市建立了我国首个智能车综合技术研发与测试中心,推动了工信部2016年启动的5个智能汽车与智慧交通应用示范区建设,有力促进我国智能车、车联网等相关技术的发展。因此从某种意义上说,本重大计划的实施促进了中国无人车产业的创立和发展,对中国汽车产业升级和健康良性发展起到实质性的重要引领

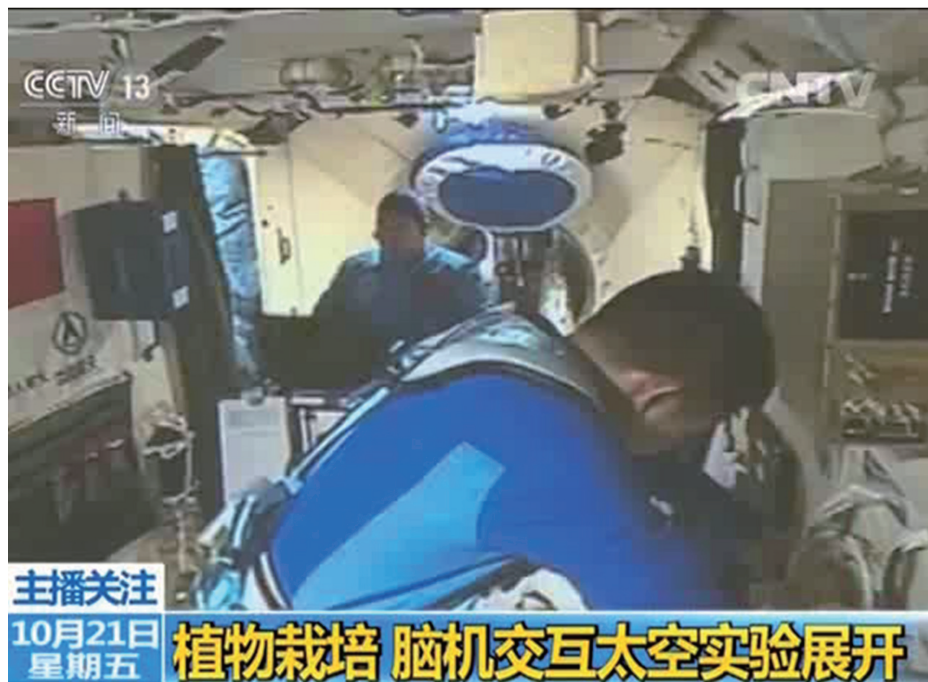


图4 人类首次太空脑机交互实验

带动作用。

该重大计划构建的集成创新验证平台不仅支持对相关重点或培育项目取得的研究成果进行验证,同时也为相关领域的实验室理论成果走向实际应用提供了可验证的物理载体。特别是,该重大计划连续举办了10届“中国智能车未来挑战赛”,该赛事已成为我国规模最大的无人驾驶比赛,受到国内外媒体的高度关注。在国际上首次提出无人驾驶技术测试理论和标准,建立了无人驾驶环境认知基础能力测评体系,促进测试环境平台基础性技术发展,从而推动了我国无人驾驶技术实现跨越式发展。此外,该重大计划所形成的无人驾驶集成验证平台不仅是无人驾驶的物理验证载体,也是未来互联网、车联网、智能交通、智能传感器以及新型计算机器的综合验证载体。同时,在“中国智能车未来挑战赛”的示范带动和影响下,陆军装备部2014年创办了地面无人平台挑战赛“跨越险阻”,多个来自本重大计划资助团队的车队获得奖励,有效对接了国防重大需求,促进了军民融合。

此外,研制了面向载人航天的世界首套脑机交互在轨试验平台,系统自重481克,开展地基实验100余人次,测试23项,入选2015年“天宫二号”的16项航天医学实验之一,并在2016年10月应用于天宫二号与神舟十一号载人飞行任务,成功完成人类首次太空脑机交互实验,促进脑机接口在国防领

域的实验应用。

### 3 展望及未来发展趋势

伴随着互联网、大数据、云计算和新型传感等技术的发展,视听觉信息的认知计算正引发可产生链式反应的科学突破,催生一批颠覆性技术、培育经济发展新动能、塑造新型产业体系,加速新一轮科技革命和产业变革。视听觉信息的认知计算的这种性能飞跃一直伴随着用于基本操作的硬件技术的显著进步,比如感觉、感知和目标识别。数据驱动型产品的平台和新市场,以及发现新产品和新市场的经济激励机制,也都促进了人工智能驱动型技术的问世。

(1) 基于神经形态计算的视听觉信息的认知计算理论与方法。传统计算机执行计算的冯·诺依曼模型,它分离了输入/输出、指令处理和存储器模块。随着深度神经网络在一系列任务中的成功,制造商正在积极追求计算的替代模型—特别是那些受到生物神经网络所启发的一为了提高硬件的效率和计算系统的稳定性的模型。

(2) 增强视听觉信息的认知计算系统的感知能力。感知是智能系统面对世界的窗口,感知来自传感器数据,具有多种形态和形式,如系统本身的状态或环境的相关信息。传感器数据常与先验知识和模型一起进行处理和整合,以提取与人工智能系统任

务相关的信息,如几何特征、属性、位置和速度。来自感知的综合数据形成环境感知,为视听觉信息的认知计算系统提供综合知识和世界状态模型,有助于有效、安全地规划和执行任务。

(3) 人在回路的人机协同混合增强智能。由于人类面临的许多问题具有不确定性、脆弱性和开放性,任何智能程度的机器都无法完全取代人类,这就需要将人的作用或人的认知模型引入到视听觉信息的认知计算的智能系统中,形成混合增强智能的形态,这种形态是人工智能或机器智能的可行的、重要的成长模式。

### 参 考 文 献

- [1] Li L, Xiao Wang, Wang KF, et al. Parallel testing of vehicle intelligence via virtual-real interaction. *Science Robotics*, 2019, 4(28): eaaw4106.
- [2] Zhou K, Luo H, Zhou T, et al. Topological change disturbs object continuity in attentive tracking. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(50): 21920—21924.
- [3] Luo H, Liu Z, Poeppel D, et al. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, 2010, 8(8): e1000445.
- [4] Liu T, Yuan ZJ, Sun J, et al. Learning to detect a salient object. *IEEE transactions on pattern analysis and machine intelligence*, 2011, 33(2): 353—367.
- [5] Zhang L, Zhang L, Mou X, et al. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*. 2011, 20 ( 8 ), 2378—2386.
- [6] Liu GH, Yang JY. Content-based image retrieval using color difference histogram. *Pattern Recognition*, 2013, 46 (1), 188—198.
- [7] Lin Y, Wu W, Wu C, et al. Extraction of mismatch negativity using a resampling-based spatial filtering method. *Journal of Neural Engineering*, 2013, 10(2): 026015.
- [8] Chang X, Shen H, Wang L, et al. Altered default mode and fronto-parietal network subsystems in patients with schizophrenia and their unaffected siblings. *Brain Research*, 2014, 1562: 87—99.
- [9] Zhang Z, Liao W, Chen H, et al. Altered functional-structural coupling of large-scale brain networks in idiopathic generalized epilepsy. *Brain*, 2011, 134(10): 2912—2928.
- [10] Li Y, Pan J, Long J, et al. Multimodal BCIs: Target detection, multidimensional control, and awareness evaluation in patients with disorder of consciousness. *Proceedings of the IEEE*, 2016, 104(2): 332—352.

## Review of Major Research Plan on “Cognitive Computing of Visual and Auditory Information”

Wu Guozheng<sup>1</sup> Han Junwei<sup>1</sup> Deng Fang<sup>1</sup> Li Jianjun<sup>1</sup> Zhang Zhaotian<sup>1</sup>  
Xin Jingmin<sup>2</sup> Zheng Nanning<sup>2</sup>

(1. *Department of Information Sciences, National Science Foundation of China, Beijing 100085;*

2. *Xi'an Jiaotong University, Xi'an 710049*)

**Abstract** In this paper, the background, scientific objectives, layout, implementation and academic management, as well as the overall outcome of the Major Research Plan “Cognitive Computing of Visual and Auditory Information” are reviewed. The suggestions of future development in the relative fields are provided.

**Key words** Major Research Plan; visual information; auditory information; natural language; brain-computer interface; cognitive computing; review