相关学术研究与技术分析

2211392 郭笑语

为进一步理解车载多模态智能交互系统背后的核心技术，我选择了两篇具有代表性的学术研究论文进行综述。一篇侧重于AI大模型在智能座舱多模态交互中的技术框架和应用，另一篇聚焦于多模态乘客意图识别这一具体问题。通过对这两项研究的总结，我们可以从理论与实验角度审视前文提到的多模态大模型和多模态意图识别技术如何提升车载交互效果，以及当前仍面临的技术挑战。

研究论文一：基于AI大模型的智能座舱多模态交互技术综述

这篇论文综述了AI大模型赋能汽车智能座舱多模态交互的原理和应用案例 。作者首先指出，智能座舱是汽车智能化的重要组成，而多模态交互是智能座舱的核心功能之一 。随着大模型（尤其是大语言模型）在人工智能领域的突破，将其引入座舱有望显著提升人机交互体验。文中分析了一个完整的多模态交互技术框架，涵盖感知层、理解层、控制层和表达层，从底层传感器数据处理到高层决策反馈 。其中，AI大模型主要发挥作用于“理解层”，利用其强大的学习与泛化能力，对多模态数据进行融合理解和推理 。论文调研了国内外主要科技公司在这一领域的探索，例如百度、华为、腾讯、科大讯飞都已推出面向座舱交互的大模型解决方案，将其用于语音、图像、手势、面部识别等场景。

文章详细比较了几家公司的大模型应用成效。例如，百度研发了文心系列通用预训练模型（文心一言等），并将其集成到Apollo智能驾驶和小度车载系统中，实现了语音、手势、面部表情、情绪等多模态交互功能，可为驾驶者提供个性化导航、娱乐、安全服务 。华为则推出了盘古大模型，融合于HarmonyOS座舱中，在语音助手和视觉识别方面取得了接近百度的效果，在图像描述和问答任务上准确率仅次于百度。腾讯的混元大模型也用于座舱语音和视觉交互，在图像生成质量上表现较好，略逊于百度但优于科大讯飞 。科大讯飞推出了自己的大语言模型和大对话系统，强调多模态融合能力。其模型可以整合语音、图像、视频、手势、生理信号等多模态输入，构建复杂语境并生成最优动作序列，实现对车辆和外部环境的智能控制 。科大讯飞基于该能力开发了智能导航和娱乐系统，使座舱能够根据用户多种输入需求提供更精准个性化服务 。不过，论文的实验评估也指出，不同公司的大模型在具体任务上仍有差异：在作者比较的图像描述、视觉问答（VQA）、图像生成三项任务中，百度模型在图像描述准确度和生成质量上均为最佳，华为次之，腾讯稍逊，科大讯飞模型的图像相关任务指标相对较低，显示在视觉领域还有提升空间。这可能与科大讯飞模型侧重语音和多模态融合、而视觉训练数据较少有关，需要后续增强。

通过大量案例分析和实验结果，论文得出结论：AI大模型在座舱多模态交互中的应用显著提高了任务处理效率与准确性，增强了人机交互体验 。具体体现为：语音识别理解更准确了，多模态意图判断更可靠了，响应更加智能多样了。这与我们在前文评估两大平台时的结论相呼应。例如，大模型让语音助手能听懂更复杂的话，也能结合手势眼神“读懂”人的真正需求。尤其是在开放域对话、复杂问答等过去车载系统较薄弱的环节，引入大模型后实现了质的飞跃。作者也表明了当前面临的一系列挑战。首先是训练和部署挑战：训练一个车载多模态大模型需要大量语音、图像、视频等数据，数据获取处理非常复杂 ；而模型部署在车端受限于算力和存储，需要进行模型优化和剪裁。这就需要在数据和算法两端共同努力，设计高效模型结构、制定模型更新迁移策略，以适应不断变化的新交互场景 。另一大挑战是模型的可解释性和可信赖性。大模型作为黑箱，如果决策过程不透明，用户难以信任。未来必须让模型决策更可解释，对外部效果进行充分评估，确保输出符合预期，才能赢得用户信赖 。此外，作者也提到多模态协同还存在诸如异步性、不一致性等技术问题，需要继续研究解决。最后总结展望了AI大模型赋能智能座舱的未来前景——随着计算平台的发展和算法的创新，多模态大模型有望成为下一代座舱的中枢大脑，实现真正以人为本、情感自然的智能交互体验。但同时，数据隐私、安全、伦理等方面也需要建立标准，以规避风险。总的来说，这篇综述全面概括了当前车载多模态大模型技术的发展状况：既让我们看到了它带来的性能提升和丰富应用，也提醒我们仍有不少挑战有待科研和工程界去攻克。

研究论文二：多模态乘客意图识别方法研究

第二篇论文聚焦于车内多模态意图识别这一具体技术点，探讨如何结合语音和视觉线索更好地理解乘客的指令意图。

论文的背景是：在高度自动驾驶车辆（AV）内，乘客可能通过语音向车载代理（如智能助手）发出各种指令，希望车辆执行相应操作。例如，让车辆导航到某地、变更路线、加减速、停车、开门等 。单纯依靠语音，有时难以准确判断意图，比如乘客说“在那里停车”，如果不知道乘客视线所指，就无法确定具体地点。为此，该研究提出利用车内摄像头视觉信息（乘客的手势、视线）以及车辆外部摄像头（乘客指向的车外物体）等非语言线索，与语音文本一同输入对话系统，以提升意图识别的准确率。

作者构建了一个叫AMIE（自动驾驶车辆多模态座舱体验）的对话代理原型。他们采用“Wizard-of-Oz”方法收集了一套多模态车内对话数据集：在模拟的无人驾驶车厢环境中，让乘客执行寻宝游戏任务，通过与系统对话来完成一系列指令。真人操作员在幕后控制车辆行为（所以乘客以为在和AI对话），同时记录下乘客的语音语料以及车内视频（捕捉乘客的目光、手势）和车外视频（乘客所指示的目标物）。由此获得了包含多轮对话、多模态同步数据的训练集 。然后，他们定义了一组*意图*和*槽位*用于评估，如意图包括“设定目的地”“改变路线”“加快速度”“停车”“靠边”“开门”“下车”等10类，槽位

包括地点、方向、对象、手势/视线目标等 。传统方法仅基于语音的文本输入来预测乘客意图和填充相关槽位，而作者实验了将音频特征和视频特征一并输入模型的方法 。

具体而言，他们采用了RNN（循环神经网络）架构的模型，其中文本模态利用Bi-LSTM对ASR识别的语句进行意图分类和序列标注（槽位提取），作为baseline。在此基础上，加入语音模态：提取乘客语音的声学嵌入（Speech2Vec向量）作为辅助输入 ；加入视觉模态：提取乘客在说话时的眼神凝视方向、手部动作以及乘客所指向的车外物体等特征 。模型将这三种模态的特征融合，在每个用户语句上输出对应的意图类别和槽位。实验结果非常鼓舞：相比仅使用文本，融合了语音+视觉后的多模态模型在意图识别准确率和槽位填充正确率两方面都有明显提升 。多模态模型成功超越了文本基线模型的性能，证明了语音音频和视觉信息对理解乘客语义有帮助。例如，对于“靠边停车”这类命令，光看文本可能与一般“停车”混淆，但结合视觉看到乘客在指向路边某位置，模型更有把握判定是“pull over”而非一般的“stop”。又如乘客说“在那里下车”，文本并未指明具体位置，但模型利用乘客视线锁定了某个建筑物正门，因而正确将槽位位置填充为该建筑的入口。这些例子体现了多模态融合助力精细意图理解的优势。

论文研究结合语言、声音和视觉的多模态理解可以显著提高车内语音指令的意图识别和槽位提取性能，从而使人与自动驾驶车辆的交互更加自然高效。研究也表明，在带有丰富上下文的场景中，依赖单一语音模态会有一定局限，而多模态模型能够弥补这方面不足，让系统“听得到也看得懂”。虽然该实验是在模拟环境下进行，但它揭示的原理在现实中正变为可能。随着近年深度学习和大模型的发展，这种多模态意图识别不再需要繁琐的规则，而可以通过端到端模型来实现。比如，我们前文介绍的Apollo小度和讯飞Spark系统，其实已在部分场景应用了类似思想：用驾驶员的目光/手势来辅助语音理解，从而达到更高意图识别率。另外，科大讯飞Spark座舱所强调的“多模态语境模型”，本质上也是希望解决像论文中乘客指令那样复杂的语义理解问题。可以说，这篇论文的成果为车载AI助手如何融合多模态来更好地服务乘客提供了实验依据，其提出的方法在当前高端智能座舱中正逐步得到验证和应用。

当然，作者也指出了进一步优化方向。例如，当多模态信息不一致时如何裁决？某些情况下视觉线索可能有干扰（比如乘客无意中看向一处并非指令相关），模型需要学会甄别真正有用的模态信号。这涉及更智能的模态权重分配和上下文理解，可能需要引入更先进的深度模型（例如Transformer架构的大模型）才能更好解决。此外，数据采集成本和标注也是挑战，该论文的数据是半仿真的，如何获取真实车辆中的大规模多模态交互数据是一大难题。不过，随着自动驾驶测试逐渐普及和车内摄像头标配，未来获取这类数据将相对容易，模型效果也会进一步提升。

第二篇研究为我们深入剖析了多模态意图识别这一关键技术如何在车载场景下发挥作用，验证了融合语音与视觉能极大提高系统理解乘客意图的可靠性。这一结论为当前多模态智能座舱的开发提供了学术支撑，解释了为何各大厂商都在给车内助手增加摄

像头、传感器来结合语音交互——目的就是提高复杂场景下的人机交互正确率，让汽车更聪明地领会乘员的用意。

结论与展望

通过上述产业分析和学术梳理，我们可以清晰地看到：中国车载多模态智能交互系统正在大模型与多模态融合的驱动下快速演进。一方面，百度Apollo小度和科大讯飞Spark座舱OS等主流平台已率先将AI大模型引入量产座舱，在功能性和用户体验上实现了飞跃：车辆可以“听会说”“眉目传情”，甚至成为拥有专业知识和情感关怀的出行伙伴。这些系统在ISO 9126质量维度上表现优异，功能全面且可靠易用，为驾乘者带来前所未有的智能体验。另一方面，学术研究也在为行业提供源源不断的创新动能。从多模态大模型框架的理论完善，到具体意图识别算法的提升，都在推动产品技术迭代。例如，多模态大模型让座舱交互有了更强的理解和生成能力，多模态意图识别算法则切实解决了嘈杂环境、含糊指令下的人机误解问题 。可以预见，未来的智能座舱将进一步朝着“以人为本”的方向发展：具备更深的场景理解和逻辑推理能力，真正读懂乘员的所思所想 ；拥有共情交互能力，成为情感伙伴而非冰冷机器；在效率上，通过更先进的车规芯片和模型优化，实现接近零延迟的响应 。同时，随着生态融合，大模型座舱或将成为车企构建差异化服务的新平台——智能座舱有望进化为连接车内外、虚实世界的超级数字生态。

然而大模型的引入也带来挑战和问题——对车载算力和成本带来压力 .....如何在有限资源上平衡性能与成本、云端与端侧，是产业需要持续攻关的课题。其次，多模态数据的获取和标注成本高，且涉及用户隐私，需要在数据利用和隐私保护间取得平衡。再次，AI助手的决策可信度和安全需要确保，不能因为智能化而引入新的驾驶风险——这需要更严格的验证和解释机制 。最后，统一的接口标准和生态协同也很重要，大模型座舱涉及整车多个域，只有实现软硬件抽象和标准化，才能降低维护复杂度，方便OTA升级和跨车型移植。

总体而言，车载多模态智能交互系统正处在从“感知智能”向“认知智能”跃升的关键时期。中国的科技公司和车企在这一领域已经走在前列，通过产学研结合不断突破。从目前的行业实践和研究成果来看，多模态大模型无疑是未来座舱人机交互的核心引擎，其应用将拓展出更加丰富的人性化服务场景，提升汽车的智能化品级。在不久的将来，消费者或许将习惯于这样的场景：坐进汽车，AI助手通过目光接触和语音问候与你交流，理解你的情绪和需求，主动安排路线、调节氛围甚至讲一个专属笑话逗你开心——汽车真正成为“懂你”的智慧出行伙伴。可以相信，随着技术的成熟和标准的完善，多模态智能座舱将成为汽车产业下一轮创新浪潮的焦点，为我们开启更安全、高效、愉悦的出行新体验。

参考文献：

汽车公社, 菠萝蜜 (2024). *大模型赋能智能座舱，中国军团迎接新挑战*. 盖世汽车资讯 (大模型赋能智能座舱，中国军团迎接新挑战-盖世汽车资讯) (大模型赋能智能座舱，

中国军团迎接新挑战-盖世汽车资讯) (大模型赋能智能座舱，中国军团迎接新挑战-盖世汽车资讯) (大模型赋能智能座舱，中国军团迎接新挑战-盖世汽车资讯)

Auto-Testing.net, 中汽数据 (2025). *中汽智联首创车载AI大模型多维体验评价体系 主流车型横评结果出炉* (中汽智联首创车载AI大模型多维体验评价体系 主流车型横评结果出炉_测试行业动态__汽车测试网) (中汽智联首创车载AI大模型多维体验评价体系 主流车型横评结果出炉_测试行业动态__汽车测试网) (中汽智联首创车载AI大模型多维体验评价体系 主流车型横评结果出炉_测试行业动态__汽车测试网)

电子工程专辑, 智能汽车设计 (2025). *基于AI大模型的新能源汽车智能座舱多模态交互技术研究综述* (基于AI大模型的新能源汽车智能座舱多模态交互技术研究综述-电子工程专辑) (基于AI大模型的新能源汽车智能座舱多模态交互技术研究综述-电子工程专辑) (基于AI大模型的新能源汽车智能座舱多模态交互技术研究综述-电子工程专辑) (基于AI大模型的新能源汽车智能座舱多模态交互技术研究综述-电子工程专辑) (基于AI大模型的新能源汽车智能座舱多模态交互技术研究综述-电子工程专辑) (基于AI大模型的新能源汽车智能座舱多模态交互技术研究综述-电子工程专辑)

InfoQ China (2018). *Apollo小度车载系统打造更舒心的出行* (Apollo小度车载系统打造更舒心的出行 - 专知) (Apollo小度车载系统打造更舒心的出行 - 专知) (Apollo小度车载系统打造更舒心的出行 - 专知)

Power Systems Design China (2024). *商汤绝影多模态大模型以人为本，引领智能汽车交互革新* (大模型赋能智能座舱，中国军团迎接新挑战-盖世汽车资讯) (大模型赋能智能座舱，中国军团迎接新挑战-盖世汽车资讯) (大模型赋能智能座舱，中国军团迎接新挑战-盖世汽车资讯)

IFAL汽车照明论坛 (2023). *多模态大模型会是未来人机交互的方向吗？* (〖行业动态〗多模态大模型会是未来人机交互的方向吗？ - IFAL汽车照明论坛)

Gasgoo汽车资讯 (2023). *iFLYTEK introduces 'Spark Desk' AI-driven cognitive large model* (
iFLYTEK introduces 'Spark Desk' AI-driven cognitive large model - Gasgoo

) (
iFLYTEK introduces 'Spark Desk' AI-driven cognitive large model - Gasgoo

) (
iFLYTEK introduces 'Spark Desk' AI-driven cognitive large model - Gasgoo

)

Gasgoo China News (2024). *LIVAN Automobile, iFLYTEK to cooperate on automotive intelligence development* (

LIVAN Automobile, iFLYTEK to cooperate on automotive intelligence development - Gasgoo

) (
LIVAN Automobile, iFLYTEK to cooperate on automotive intelligence development - Gasgoo

)

arXiv (2019). *Towards Multimodal Understanding of Passenger-Vehicle Interactions in Autonomous Vehicles: Intent/Slot Recognition Utilizing Audio-Visual Data* ([1909.13714] Towards Multimodal Understanding of Passenger-Vehicle Interactions in Autonomous Vehicles: Intent/Slot Recognition Utilizing Audio-Visual Data) ([1909.13714] Towards Multimodal Understanding of Passenger-Vehicle Interactions in Autonomous Vehicles: Intent/Slot Recognition Utilizing Audio-Visual Data)