



南開大學  
Nankai University

南 開 大 學

計 算 機 學 院

軟件工程

---

## 車載系統軟件需求分析

---

董珺

年級：2022 級

專業：計算機科學與技術

指導教師：李起成

2025 年 4 月 7 日

# 目录

<b>一、 引言</b>	<b>1</b>
(一) 编写目的 . . . . .	1
(二) 项目背景 . . . . .	1
<b>二、 任务概述</b>	<b>1</b>
(一) 任务目的 . . . . .	1
(二) 用户特点 . . . . .	2
(三) 假定与约束 . . . . .	2
<b>三、 业务描述</b>	<b>3</b>
(一) 系统总业务流程图 . . . . .	3
(二) 各个子业务流程图及其描述 . . . . .	3
<b>四、 数据需求</b>	<b>5</b>
(一) 数据需求描述 . . . . .	5
(二) 数据流图 . . . . .	6
(三) 数据字典 . . . . .	6
<b>五、 功能划分</b>	<b>6</b>
<b>六、 功能描述</b>	<b>7</b>
<b>七、 性能/非功能需求</b>	<b>8</b>
<b>八、 系统运行要求</b>	<b>10</b>

## 一、引言

### （一）编写目的

本报告的目的是对车载多模态智能交互系统进行需求分析。通过深入了解该系统的功能需求、性能要求、用户需求等方面，为系统的设计与开发提供详细的参考依据。随着人工智能技术的进步，尤其是大模型的广泛应用，车载系统智能化程度不断提升，因此开发一个高效、智能、安全且人性化的交互系统变得尤为重要。本报告将为系统的需求分析、架构设计、技术实现等阶段提供系统化的指导和理论依据，确保最终系统能够满足用户的多样化需求，提升驾驶体验，并保障驾驶安全。

### （二）项目背景

随着智能汽车和自动驾驶技术的迅速发展，车载系统的功能需求不断升级。传统的车载交互系统主要依赖于单一的交互方式，如语音或触摸屏，而这些系统在实际使用中往往存在交互不自然、误操作、响应迟缓等问题。为了应对这些挑战，车载多模态智能交互系统应运而生。该系统集成了语音识别、自然语言处理、图像处理等多项人工智能技术，能够通过多种交互方式（如语音、触摸、手势等）为驾驶员和乘客提供更加智能、便捷和安全的体验。

目前，车载系统的智能化水平正在迅速提升，尤其是大模型的应用使得系统能更加精准地理解和预测用户需求，提升交互效率和用户体验。随着消费者对智能化和个性化车载体验的要求不断提高，开发一款高效、智能、安全的车载多模态交互系统已成为行业发展的必然趋势。

该项目旨在利用先进的人工智能技术，打造一款多模态智能交互系统，提升车载系统的智能化水平，并满足未来智能汽车市场对高效、便捷、安全的多模态交互体验的需求。

## 二、任务概述

### （一）任务目的

车载多模态智能交互系统的开发旨在通过整合语音识别、自然语言处理、图像处理、手势识别等多项先进的人工智能技术，为驾驶员和乘客提供更加智能、便捷和安全的车载交互体验。具体目标包括：

- **提高用户体验：**通过多模态交互方式（语音、触摸、手势等）的结合，使车载系统能根据不同场景和需求灵活适应用户的交互方式，从而提升交互的自然性和流畅性。
- **提升驾驶安全性：**系统应当能够实时监测和分析驾驶员及周围环境的情况，提供警告、辅助驾驶或紧急干预功能，确保驾驶安全。
- **实现精准识别与响应：**系统应具备高准确率的语音识别、图像识别和自然语言理解能力，能够准确识别用户指令并做出及时响应，避免误操作和延迟。
- **增强系统的智能化：**通过大模型的应用，使得车载系统能够逐渐学习并理解用户的偏好、习惯和需求，提供个性化的服务。
- **提供高效的功能整合：**系统应能够无缝整合车载导航、娱乐、控制、车况监测等多项功能，为用户提供一站式服务，减少驾驶员的操作负担。
- **确保系统的可扩展性和升级性：**系统应具有良好的扩展性，能够方便地进行软件更新和功能增强，以适应未来技术的发展。

## (二) 用户特点

车载多模态智能交互系统的用户主要包括：

1. **驾驶员：**作为主要的用户群体，驾驶员使用车载系统时需要关注安全性、操作便捷性以及驾驶辅助功能的需求。驾驶员通常在驾驶过程中有较高的认知负荷，因此车载系统需要简化操作过程，并尽量减少视觉和手动输入的干扰。
  - 需求：便捷的语音交互、自动导航、智能安全提醒、驾驶状态监控等。
  - 特点：驾驶员在使用过程中希望系统提供直观、精准的反馈，同时避免过多的干扰或复杂操作，确保驾驶时的集中注意力。
2. **乘客：**乘客对车载系统的需求更多集中在娱乐和信息服务方面。乘客希望能够通过语音或触摸方式轻松控制车载娱乐系统、导航以及其他车内服务。
  - 需求：车内娱乐、信息查询、车内环境控制（如空调、座椅加热等）、语音助手等。
  - 特点：乘客的使用场景多变，交互方式应当灵活，操作简便，并具备一定的个性化推荐能力。
3. **车主与系统维护人员：**车主和维护人员通常需要系统的后台管理功能、车况监测、远程诊断等。
  - 需求：车辆状态监控、远程控制、系统升级等。
  - 特点：车主希望通过手机或其他设备与车载系统进行互动，而维护人员则需要便于查看车辆健康状况的系统功能。

## (三) 假定与约束

1. **硬件约束：**车载系统需要在车内硬件环境中运行，受限于车辆的计算资源（如处理器、内存等）以及车载传感器和摄像头的性能。因此，系统设计需要考虑硬件的性能限制，确保能够在现有硬件上高效运行。
2. **实时性要求：**车载多模态智能交互系统需要处理大量实时数据（如语音指令、图像数据等），并能够做出快速响应。这要求系统必须具备高性能的处理能力，同时保证低延迟和高稳定性。
3. **安全性和隐私保护：**由于车载系统涉及到用户的个人信息、车主数据以及驾驶行为等敏感内容，系统必须遵守严格的数据安全和隐私保护规范，确保数据的加密存储和传输。
4. **兼容性：**系统需要与不同型号、不同品牌的车辆兼容，并且能够与车载硬件、其他设备（如智能手机、智能家居等）进行无缝连接，形成一个智能生态系统。
5. **法律和合规性：**系统开发和运行必须遵循相关的法律法规，特别是在数据保护、车载系统标准、安全性等方面的要求。
6. **用户接受度：**虽然车载多模态智能交互系统能够提升智能化体验，但也需要考虑用户的技术接受度。对于一些不太熟悉新技术的用户，系统的操作应尽量简化，并提供必要的培训和帮助文档。
7. **市场竞争：**市场上已有一些车载智能系统，新的系统在功能上需要有创新性，同时在性能、稳定性、用户体验等方面要具有明显优势，才能在市场中脱颖而出。

### 三、 业务描述

#### (一) 系统总业务流程图

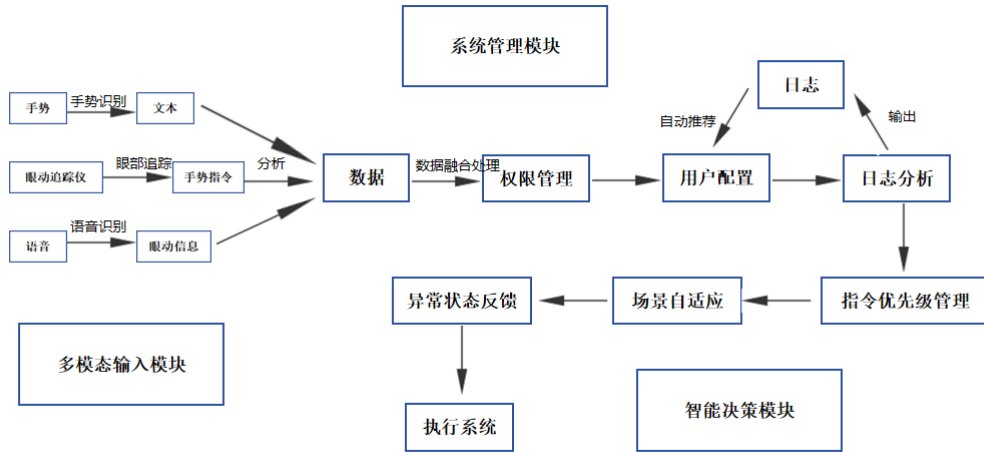


图 1: 系统总业务流程图

#### (二) 各个子业务流程图及其描述

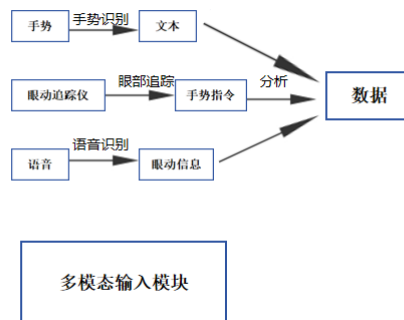


图 2: 多模态输入模块业务流程图

#### 1. 多模态输入融合层

##### 1.1 语音控制模块

- 语音采集（车载麦克风阵列）
- 语音预处理（降噪、回声消除）
- 语音识别（ASR, Automatic Speech Recognition）
- 文本解析（NLU, Natural Language Understanding）
- 意图识别（识别指令，如“开窗”）
- 发送到数据融合处理中心

### 1.2 手势交互模块

- 手势图像采集（车内摄像头）
- 图像预处理（去噪、增强）
- 手势识别（关键点提取 + 深度学习识别）
- 手势意图解析（如挥手表示拒接电话）
- 发送到数据融合处理中心

### 1.3 视觉交互模块

- 视觉图像采集
- 驾驶员关注区检测（面部关键点跟踪）
- 头部姿态估计（Pitch, Yaw, Roll 角度提取）
- 眼动追踪（注视点检测）
- 数据预处理（稳定追踪、异常剔除）
- 视觉行为解析
- 判断是否分心、疲劳驾驶
- 发送到数据融合处理中心

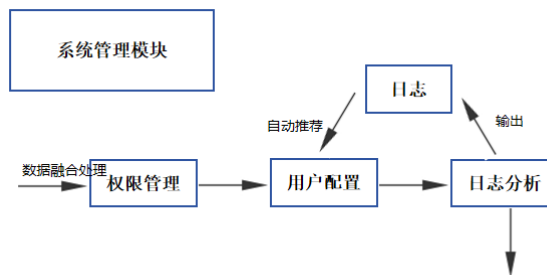


图 3: 系统管理模块业务流程图

## 2. 系统管理功能层

### 2.1 用户个性化配置

- 用户偏好学习（语音喜好、常用指令、常用手势）
- 模型自适应（不同驾驶员切换自适应）

### 2.2 多模态交互日志记录与分析

- 实时记录交互指令（时间戳、来源模态、指令内容）
- 行为分析（识别异常，如频繁取消指令）

### 2.3 系统权限管理

- 用户登录验证（人脸识别、声纹识别）
- 不同用户不同权限（如驾驶员 vs 乘客）

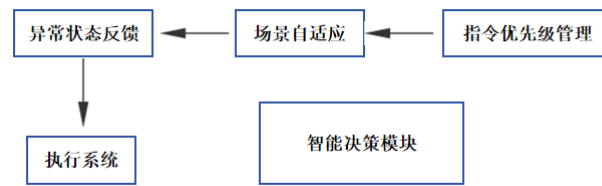


图 4: 智能决策模块业务流程图

### 3. 智能决策模块

#### 3.1 多模态指令优先级管理

- 设定不同模态优先级（如驾驶状态下手势 > 语音）
- 动态调整优先级（根据环境噪音、驾驶状态）

#### 3.2 驾驶场景自适应

- 城市道路 vs 高速公路 vs 停车场景切换
- 场景推断（基于视觉数据 + 车载传感器数据）
- 自动调整交互策略（如高速上减少交互复杂度）

#### 3.3 异常状态反馈

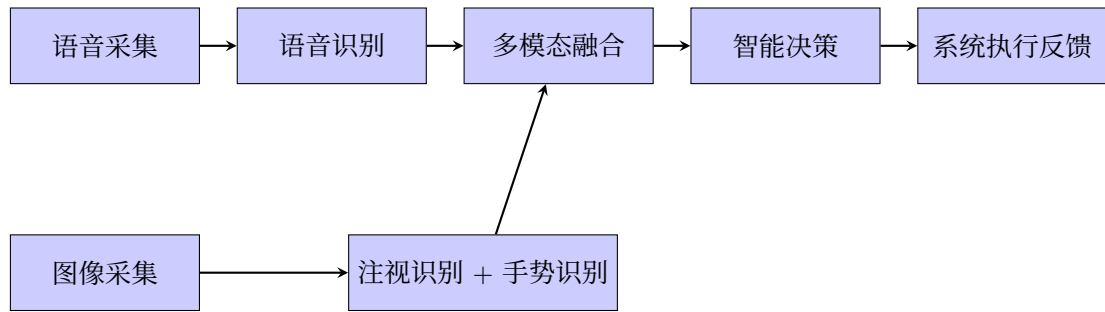
- 异常检测（如驾驶员疲劳、突发病症）
- 自动报警（本地警报 + 远程通知）
- 应急处理（如自动减速、靠边停车）

## 四、 数据需求

### （一） 数据需求描述

1. 语音数据流实时采集驾驶员语音指令，格式为压缩音频（如 AAC），进行语音识别转文本。
2. 视觉数据流摄像头持续采集驾驶员图像、注视区域、手势动作，格式为压缩视频帧（如 H.264），用于图像识别和姿态解析。
3. 传感器数据流眼动仪、头部姿态传感器数据（如坐标、角度、速度），以二进制小数据包格式发送。
4. 融合数据流将多源数据（文本指令、注视区域、手势动作）在中间件中统一标准格式（JSON 或自定义二进制协议），供智能决策模块使用。
5. 日志数据流包括用户交互记录、指令执行记录、异常状态反馈，格式为本地 JSON 文件或上传至云端数据库。

(二) 数据流图



(三) 数据字典

数据项	说明	类型	备注
VoiceCommand	驾驶员语音指令文本	字符串 (UTF-8)	例如: "打开车窗"
GazeArea	驾驶员注视区域编号	整数	0= 中控屏, 1= 空调区域等
GestureType	手势动作类型编号	整数	1= 滑动, 2= 点击, 3= 圈选等
HeadPose	头部朝向数据 (pitch/yaw/roll)	浮点数组	单位: 度 (°)
EyeMovementData	眼动追踪数据	坐标数组	X、Y 轴位置点
FusionCommand	融合后的最终指令	JSON 对象	包含来源、置信度、动作
SystemLogEntry	系统日志记录	JSON 对象	时间戳、事件、结果等信息

表 1: 数据字典 (关键数据项)

## 五、 功能划分

系统功能主要划分为以下几个大模块:

### 1. 多模态输入融合模块

- 语音交互处理
- 手势交互处理
- 视觉交互处理 (图像识别、头部姿态、眼动追踪)
- 多模态数据融合处理

### 2. 智能决策与响应模块

- 多模态指令优先级管理
- 驾驶场景自适应
- 异常状态反馈机制

### 3. 系统管理模块

- 用户个性化配置管理
- 多模态交互日志记录与分析



- 系统权限管理与用户身份验证

#### 4. 创新功能模块

- 三模态深度融合指令生成（语音 + 手势 + 眼动数据共同作用）
- 智能推荐与预测交互（基于大模型的上下文理解和行为预测）
- 智能异常预警系统（基于多模态感知的安全性监测）

## 六、 功能描述

### 2.1 多模态输入融合模块

#### 2.1.1 语音交互处理

- 语音采集 → 语音识别（ASR）→ 文本转化
- 调用自然语言理解（NLU）模块，提取意图、关键词
- 匹配对应的交互命令或调用功能

#### 2.1.2 手势交互处理

- 摄像头采集手势图像流
- 手势识别算法识别特定手势（如滑动、握拳、点击）
- 映射为交互指令，如控制音乐、接电话等

#### 2.1.3 视觉交互处理

- 识别驾驶员视线关注区域
- 解析头部姿态数据（前倾、侧头、低头等）
- 眼动追踪技术跟踪注视点
- 结合场景判断意图（例如注视中控屏时自动放大信息）

#### 2.1.4 多模态数据融合处理

- 时间同步不同来源数据
- 语义层面融合（如语音“打开这个” + 注视导航图标 → 打开导航）
- 解决冲突与不一致（如语音和手势同时指向不同操作）

### 2.2 智能决策与响应模块

#### 2.2.1 多模态指令优先级管理

- 根据上下文环境、交互信号强度、用户意图置信度确定指令优先级
- 高优先级指令优先执行，低优先级指令排队或取消

#### 2.2.2 驾驶场景自适应

- 依据环境（高速、城区、夜晚、恶劣天气）动态调整交互策略
- 在高速时减少交互复杂度、优先保障驾驶安全

- 在停车状态下解锁更多复杂交互（如多任务操作）

### 2.2.3 异常状态反馈机制

- 检测驾驶员疲劳、分神等异常状态
- 检测系统内部故障或异常（例如语音识别失败）
- 提供实时提示、声音警报、必要时采取安全措施（如切换辅助驾驶模式）

## 2.3 系统管理模块

### 2.3.1 用户个性化配置管理

- 支持多用户登录
- 保存个性化配置，如常用导航地点、喜欢的音乐风格、常用手势动作
- 配置可以导出/导入

### 2.3.2 多模态交互日志记录与分析

- 实时记录每次交互的类型、结果、时间戳等
- 支持后续数据分析，用于系统优化和用户体验提升

### 2.3.3 系统权限管理与用户身份验证

- 驾驶员与乘客角色分离
- 不同权限控制（如车辆设置仅驾驶员可修改）
- 支持指纹识别、人脸识别或密码认证等方式登录系统

## 2.4 创新功能模块

### 2.4.1 三模态深度融合指令生成

- 融合语音命令、手势动作、眼动注视，实现更精准高效的意图识别
- 例如：“打开” + 手势指向空调 + 眼神注视空调区域 → 自动打开并调整空调温度

### 2.4.2 智能推荐与预测交互

- 基于大模型（如 LLM）分析用户行为习惯
- 主动推荐功能（如下雨时弹出雨刷控制、晚上弹出灯光控制）

### 2.4.3 智能异常预警系统

- 基于多模态融合（视觉 + 声音 + 车辆传感器）实时监控驾驶安全状态
- 提前预警潜在风险（例如疲劳驾驶预警、分神驾驶提醒）

# 七、性能/非功能需求

## 1. 识别准确性

- 语音识别准确率 95%

- 注视识别准确率 92%
- 手势识别准确率 90%
- 保障交互可靠性

## 2. 响应及时性

- 语音响应 500ms
- 注视响应 300ms
- 手势响应 400ms
- 多模态融合处理延迟 800ms
- 指令到执行反馈时间 1 秒

## 3. 系统可扩展性

- 支持新增交互模态
- 支持设备无缝接入
- 支持远程 OTA 更新
- 便于后期功能迭代

## 4. 易用性

- 用户交互 3 步完成
- 提供语音引导和视觉界面反馈
- 提升驾驶员操作便捷性

## 5. 易维护性

- 模块化设计
- 支持独立更新与自动恢复
- 减少故障修复时间和成本

## 6. 标准合规性

- 符合 ISO 26262 安全标准及 ASIL 等级要求
- 兼容主流车载通信协议（如 CAN、Ethernet）

## 7. 先进性

- 采用轻量级深度学习技术
- 支持小型大模型推理与边云协同
- 保持系统技术领先

## 八、 系统运行要求

### 1. 硬件配置要求

#### 1.1 主控芯片

- 需搭载车规级 AI 处理器，如 NVIDIA Orin 或高通 SA8295P
- 支持深度学习推理与图像处理

#### 1.2 内存与存储

- 内存 8GB
- 存储空间 128GB
- 支持高速读写，满足多模态数据处理需求

#### 1.3 多模态传感器

- 集成高清摄像头（驾驶员监控/手势识别）
- 阵列麦克风（语音输入）
- 红外或 ToF 传感器（精准手势捕捉）

#### 1.4 通信模块

- 支持 CAN 总线、Ethernet 通信
- 具备 4G/5G 远程连接能力

#### 1.5 显示与反馈设备

- 配备高清中控屏（10 寸）
- 支持触摸及视觉反馈显示

### 2. 软件配置要求

#### 2.1 操作系统

- 基于车载 Linux（如 QNX、Yocto）或 Android Automotive 平台
- 实时性高，安全性强

#### 2.2 AI 推理框架

- 支持 TensorRT、ONNX Runtime 等轻量化深度学习推理库

#### 2.3 中间件

- 采用 ROS 2.0 或自研车载中间件
- 保障模块间通信高效可靠

#### 2.4 数据安全性与隐私保护

- 支持本地数据加密、身份认证机制
- 符合 GDPR 及 ISO/SAE 21434 网络安全标准

#### 2.5 远程运维支持

- 系统需支持 FOTA (Firmware Over-the-Air) 远程升级与故障监测

NIKU