

团体标准

T/CSAE xx—20xx

汽车智能座舱语音能力分级与测评方法

Classification, Testing, and Evaluation Methods
for Automotive Intelligent Cockpit Speech Capability

(报批稿)

20xx-xx-xx 发布

20xx-xx-xx 实施

中国汽车工程学会 发布

目 次

前言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 分级体系与评价参数.....	6
5 测试准备.....	8
6 测试方法.....	9
7 评价方法.....	17
附 录 A（规范性） 唤醒测试语料库.....	21
附 录 B（规范性） 情绪认知测试语料库.....	22
附 录 C（规范性） 认知学习测试语料库.....	23
附 录 D（规范性） 信息交互与信息任务执行测试语料库.....	24
附 录 E（规范性） 评分量表.....	26
附 录 F（规范性） 日常沟通测试语料库.....	27
附 录 G（规范性） 共情陪伴与社会情感测试语料库.....	29
附 录 H（规范性） 车控任务执行测试语料库.....	31
附 录 I（规范性） 自主任务构建测试语料库.....	32
附 录 J（规范性） 自主任务规划与反思测试语料库.....	33
附 录 K（规范性） 自主任务决策测试语料库.....	34
参考文献.....	35

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国智能网联汽车产业创新联盟提出。

本文件由中国汽车工程学会标准化工作委员会归口。

本文件起草单位：清华大学、芜湖雄狮汽车科技有限公司、科大讯飞股份有限公司。

本文件主要起草人：曹东璞、何雷、孔睿婧、刘尔玄、郭宇昕、韩星、陈宏昌

汽车智能座舱语音能力分级与测评方法

1 范围

本文件确立了汽车智能座舱语音能力分级体系，规定了其测试条件，描述了相应的测试方法和评价方法。

本文件适用于搭载智能语音系统的车辆的语音能力分级、测试与评价。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 36464.1-2020 信息技术 智能语音交互系统 第1部分：通用规范

GB/T 3785.1-2023 电声学 声级计 第1部分：规范术语和定义

3 术语和定义

GB/T 36464.1-2020 中界定的及下列术语和定义适用于本文件。

3.1

感知能力 perceptual ability

智能语音系统对于语音输入的敏感程度和判断能力。

3.2

认知能力 cognitive ability

智能语音系统通过输入的语音判断用户意图、理解所处场景的能力，以及通过学习了解用户及用户群的能力。

3.3

交互能力 interaction ability

用户和功能单元之间通过语音进行信息传递、交流的过程中，体现出的拟人化程度，以及其功能权限范围。

3.4

固定语音唤醒 fixed voice trigger

处于音频流监听状态的语音交互系统，在检测到指定的语音唤醒词后，切换到命令字识别、连续语音识别等其他处理状态的过程。

3.5

模糊语音唤醒 fuzzy voice trigger

处于音频流监听状态的语音交互系统，在检测到对话中存在语音唤醒词后，切换到命令字识别、连续语音识别等其他处理状态的过程。

3.6

免唤醒 non-voice trigger

基于语音交互系统的智能响应能力，通过持续监听音频流的技术特性，在限定时间与场景内，实现用户无需特定唤醒词即可直接触发指令执行或信息查询等交互功能的能力，以及由系统自动识别用户自然语音并完成指令触发的过程。

3.7

全时免唤醒 full time non-voice trigger

基于语音交互系统的智能响应能力，通过持续监听音频流的技术特性，在任何时间与场景内，实现用户无需特定唤醒词即可直接触发指令执行或信息查询等交互功能的能力，以及系统自动捕捉用户自然语音并完成指令响应的过程。

3.8

用户画像 user profile

基于数据建模的数据应用方法，通过整合用户自然属性、行为轨迹及其他特征，运用标签体系对个体进行数字化表征，实现用户身份识别与需求预测，为个性化推荐或其他场景提供决策支持。

3.9

用户学习 user learning

系统通过学习一段时间内的用户行为数据、偏好、习惯以及交互历史等信息，构建一套详细且动态的用户模型的过程。

3.10

社会学习 social learning

基于社会互动的认知发展机制，通过动态观察模仿或其他形式，使语音交互系统能够在社会环境中整合他人经验与自身实践，实现知识获取、行为模式塑造及社会规范内化的持续性适应过程。

3.11

日常沟通 daily communication

结合对基本座舱场景上下文理解，能够进行简单非模板化的日常语言交流互动，快速建立与用户有效关系的能力。

3.12

共情陪伴 empathic companionship

在日常沟通（3.11）和交互能力（3.3）的基础上，识别、理解和回应他人正/负情绪并有效表达情感的能力，以及围绕情绪感知与情感回应所形成的持续交互过程。

3.13

社会情感 social feeling

在准确认知用户多类情绪的基础上，通过社会学习（3.10）获取有效社会群体交流的能力，并由此与用户建立更深层次情感连接的过程。

3.14

自主任务构建 autonomous task formulation

根据用户对确定性任务的指令，系统自动分解任务并构建必要任务步骤的能力，以及在此基础上由用户确认或自动完成确认的过程。

3.15

自主任务规划与反思 autonomous task planning and reflection

基于用户的任务指令，系统自动分解任务、自主规划执行步骤、调用资源与工具，并依据中间反馈迭代计划直至完成，同时由用户确认或自动确认的能力，以及系统围绕任务执行进行规划、执行与迭代直至结束的过程。

3.16

自主任务决策 autonomous task decision making

基于动态环境认知的智能系统决策机制，通过场景理解和社会情感（3.13）认知，实现语音交互系统在复杂场景下主动解析用户意图、适配社会情感需求，并生成个性化服务闭环的持续响应能力。

3.17

成功率 Interaction success rate

系统完整且成功执行某项任务的次数与测试总次数的比率。

3.18

误唤醒频度 false wakeup frequentness

语音交互系统处于休眠状态一段时间内，保持外部有多媒体声、人声、语音导航等不包含唤醒词或免唤醒指令的声源持续播放，无唤醒意图时语音交互系统被唤醒给出响应的次数。

3.19

满意度 satisfaction

用户在使用系统、产品或服务时产生的身体、认知和情感反应，表示满足用户需求和期望的程度。

[来源：ISO 9241-11:2018，3.1.14]

3.20

识别准确率 recognition accuracy

在某段时间内，能准确判断并成功给出相应反馈的次数与测试总次数的比率。

3.21

记忆数量 memory quantity

座舱系统可记忆的最大用户画像信息种类数量。

3.22

记忆准确率 memory accuracy

座舱根据用户画像信息所提供的服务中，成功给出相应反馈的次数与测试总次数的比率。

3.23

学习维度 learning dimension

语音交互系统在不同维度下持续优化人机协作效能的能力，其包括包括行驶、办公、娱乐、休闲、社交、健康、日常（3.11）、应急、其他共9类维度。

3.24

学习准确率 learning accuracy

座舱根据用户学习信息所提供的服务中，成功给出相应反馈的次数与测试总次数的比率。

3.25

信息冗余度 information redundancy

在一轮问答过程中，以语音输出文字的形式成功给出相应反馈的简短回答。

3.26

任务完成率 task completion rate

在交互中，成功完成任务次数与任务总次数的比例。

3.27

交互准确性 interaction accuracy

在情感交互中，判断交互中的内容相关度。

3.28

拟人化程度 degree of personification

用于评估语音交互系统在单轮或多轮对话中，结合语境、用户情绪及行为认知，所实现对话的自然度和流畅度的综合水平。

3.29

构建准确率 construction accuracy

在自主任务构建（3.14）中，成功构建该项任务的次数与测试总次数的比率。

3.30

规划与反思准确率 planning and reflecting accuracy

在自主任务规划与反思（3.15）中，成功规划与反思该项任务的次数与测试总次数的比率。

3.31

决策准确率 decision accuracy

在自主任务决策（3.16）中，成功决策该项任务的次数与测试总次数的比率。

4 分级体系与评价参数

4.1 汽车智能座舱语音能力分级

根据汽车智能座舱语音的感知能力、认知能力、交互能力，将其分级为L0-L4，分级应符合表1的规定。

表 1 汽车智能座舱语音能力分级

层级	主要特征	感知能力	认知能力		交互能力	
		唤醒	情绪 认知	认知 学习	情感 交互	任务 交互
L0 功能式 智能 语音	需用户以物理方式唤醒，再语音输入单一或连续多个车控类任务指令，系统能够执行并反馈。	无	无	无	无	车控 任务 执行
L1 信息式 智能 语音	L0基础上，用户可以使用固定唤醒词唤醒语音系统。系统具备信息交互能力，需用户语音输入信息类任务指令，系统能够执行并反馈。	固定 语音 唤醒	无	无	信息 交互	信息 任务 执行
L2 用户 认知 智能 语音	L1基础上，用户可以使用模糊唤醒词唤醒语音系统。具备用户画像和日常沟通能力。基于用户指定的任务，可自主构建任务流程，并执行完成任务。	模糊 语音 唤醒	无	用户 画像	日常 沟通	自主任务 构建
L3 情感 认知 智能 语音	L2基础上，用户对话可免唤醒。通过对话或多模态判断用户正/负面情绪，具备用户学习和共情陪伴能力。基于用户指定的任务，系统自主规划执行步骤，使用资源和工具，并根据中间反馈迭代计划，直至任务完成。	免唤醒	正/负	用户 学习	共情 陪伴	自主任务 规划与反 思
L4 社会 认知 智能 语音	L3基础上，用户对话可全时免唤醒。通过对话判断用户多类情绪，主动学习理解用户群和社会规范。具备社会情感属性和交互能力。系统感知用户情境，理解用户记忆，能够主动提供个性化服务。	全时 免唤醒	多类	社会 学习	社会 情感	自主 任务 决策

4.2 汽车智能座舱语音能力分级体系及评价参数

感知能力主要通过唤醒能力来衡量，包括成功率、误唤醒频度两项评价参数。认知能力主要通过情绪认知和认知学习能力衡量，包括识别准确率、记忆数量、记忆准确率等评价参数。交互能力主要通过情感交互和任务交互来衡量，包括信息冗余度、日常任务完成率、交互准确性、拟人化程度等评价参数。具体评价参数应符合表2的规定。

表 2 分级评价体系及评价参数

一级指标	二级指标	三级指标	评价参数
感知能力	唤醒	固定语音唤醒	成功率
			误唤醒频度
		模糊语音唤醒	成功率
			误唤醒频度
		免唤醒	成功率
			误唤醒频度
		全时免唤醒	成功率
			误唤醒频度
认知能力	情绪认知	正/负情绪认知	识别准确率
		多类情绪认知	识别准确率
	认知学习	用户画像	记忆数量
			记忆准确率
		用户学习	学习维度
			用户学习准确率
		社会学习	学习维度
			社会学习准确率
交互能力	情感交互	信息交互	信息冗余度
		日常沟通	日常任务完成率
			交互准确性
			拟人化程度
		共情陪伴	共情任务完成率
			交互准确性
			拟人化程度
		社会情感	社会情感任务完成率
			交互准确性
			拟人化程度
	任务交互	车控任务执行	车控任务完成率
			满意度
		信息任务执行	信息任务完成率
			满意度
		自主任务构建	构建准确率
			任务完成率
			满意度
		自主任务规划与反思	规划与反思准确率
			任务完成率
			满意度
		自主任务决策	决策准确率
			任务完成率
			满意度

5 测试准备

5.1 测试条件

5.1.1 测试环境

测试环境应符合以下条件：

- a) 能见度或天气因素不影响语音功能操作。
- b) 车辆所处区域无强电磁信号干扰；
- c) 网络条件满足上行带宽不低于 200kbit/s、下行带宽不低于 100kbit/s，网络时延小于 100 ms，保持稳定的联通状态；
- d) 在不展开遮光板条件下，驾驶舱内光照强度应大于 500 lux 且小于 1500 lux，测量位置在主/副驾驶眼点位置。外部光源在驾驶舱内无明显反光现象。

5.1.2 测试场地

测试应在封闭场景（地上/地下停车场、室内车库等）进行，背景噪声小于 70dB（A）。

5.1.3 语音声源及噪声

测试设备声源设置应符合以下条件：

- a) 设备放置在主驾头枕人嘴发声高度位置；
- b) 设备发声方向 10cm 处声压级：70dB（A）~80dB（A）。

车内测试环境应符合：目标 MIC 处的回声声压级应为 70dB（A）至 80dB（A）。

5.2 测试车辆

测试车辆条件如下：

- a) 车辆功能正常，无故障报警；
- b) 所有屏幕表面无出厂后贴的膜；
- c) 车机系统调为出厂默认状态（系统设置、功能设置、交互界面等）；
- d) 登录车辆账号，确保被测功能可用；
- e) 确保车辆正常联网，且在测试全程保持联网状态；
- f) 准备测评过程中依赖的数据，如联系人列表、实体信息（音乐列表、电台列表）等；
- g) 测试前，需安装测试相关 APP，激活相关功能和响应反馈，保证功能可用；
- h) 未执行对应语料需将车窗等设备处于关闭状态；
- i) 车辆电量/油量确保 24h 可用，且中途无休眠。

5.3 测试设备

测试设备应符合表3的规定。

表 3 测试设备要求

设备	用途	要求	数量 (个)
生成式语音 测试系统 (包括生成式语音测试设备和测试软件)	生成并播放 语音测试语料	信噪比: 90 dB; 增益控制: 0 dB~25 dB; 频率响应: 200 Hz~10 kHz; 最大声压级: 110 dB(A)。	1
摄像头	记录舱内交互过程、 车辆屏幕	分辨率: $\geq 1080\text{ P}$; 刷新率: $\geq 30\text{ fps}$ 。	3
照度计	测量光照强度	量程下限: $\leq 1\text{ lux}$; 量程上限: $\geq 3000\text{ lux}$; 示值误差: $\leq \pm 8\%$ 。	1
声级计	测量噪音	至少满足 GB/T 3785.1-2023 规定的 2 级性能等级。	1
测试前的摄像头架设应符合下列要求: a) 于后视镜下方 5cm 处架设 1 台摄像头, 其朝向为水平正后方; b) 于前排两座椅头枕中间位置架设 1 台摄像头, 其朝向为中控屏; c) 于前挡风玻璃左上角处架设 1 台摄像头, 其朝向前排座椅头枕。			

5.4 测试人员

5.4.1 测试员

测试员应符合以下条件:

- 至少 1 名测试员;
- 测试员应有 1 年以上实际驾龄, 裸眼或矫正视力 4.5 及以上;
- 测试员应进行生成式语音测试设备及测试软件视频培训。

5.4.2 被试员

被试员应符合以下条件:

- 至少 1 名被试员;
- 被试员应有 1 年以上智能座舱测试与评价经验, 裸眼或矫正视力 4.5 及以上。

6 测试方法

6.1 唤醒

6.1.1 固定语音唤醒-成功率

按照以下步骤进行测试:

- 将语音测试设备放置于主驾头枕人嘴发声高度;
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件;
- 点击 APP 中固定语音唤醒模块, 添加唤醒词;
- 在语音休眠状态下, 点击播放测试语料按钮, 语料应符合附录 A 的要求;
- 根据座舱语音交互的反馈, 点击正确/错误/此条跳过按钮, 记录响应唤醒的成功次数;

- f) 重复 a) ~e) 步骤, 在语音休眠状态测试 30 次;
g) 按照公式 (1) 计算成功率 S_{vw} :

$$S_{vw} = \frac{n_{vw}}{N_{vw}} \times 100\% \dots\dots\dots (1)$$

式中:

S_{vw} ——成功率;

n_{vw} ——成功唤醒次数;

N_{vw} ——尝试唤醒总次数。

6.1.2 固定语音唤醒-误唤醒频度

按照以下步骤进行测试:

- a) 将语音测试设备放置于主驾头枕人嘴发声高度;
b) 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件;
c) 点击 APP 中固定语音唤醒模块, 在语音休眠状态下, 点击播放测试语料按钮, 播放真实场景可能出现的不包含唤醒词和免唤醒指令的声源: 手机端多媒体声/人声闲聊/手机语音导航等, 语料应符合附录 A 的要求, 持续播放 24h;
d) 根据座舱语音交互的反馈, 点击误唤醒和时间节点标记按钮, 记录被误唤醒的次数;
e) 照公式 (2) 计算误唤醒频度 f_{fw} :

$$f_{fw} = \frac{n_{fw}}{T} \dots\dots\dots (2)$$

式中:

f_{fw} ——误唤醒频度;

n_{fw} ——误唤醒次数;

T——测试总时长 (小时)。

6.1.3 模糊语音唤醒-成功率

按照 6.1.1 测试模糊语音唤醒-成功率, 语料采用表 A.1 中模糊语音唤醒语料。

6.1.4 模糊语音唤醒-误唤醒频度

按照 6.1.2 测试模糊语音唤醒-误唤醒频度测试, 语料采用表 A.1 中误唤醒语料。

6.1.5 免唤醒-成功率

按照 6.1.1 测试免唤醒-成功率, 语料采用表 A.1 免唤醒语料。

6.1.6 免唤醒-误唤醒频度

按照 6.1.2 测试免唤醒-误唤醒频度测试, 语料采用表 A.1 误唤醒语料。

6.1.7 全时免唤醒-成功率

按照 6.1.1 测试全时免唤醒-成功率, 语料采用表 A.1 中全时免唤醒语料。

6.1.8 全时免唤醒-误唤醒频度

按照 6.1.2 测试全时免唤醒-误唤醒频度测试, 语料采用表 A.1 中误唤醒语料。

6.2 情绪认知

6.2.1 正/负情绪认知-识别准确率

按照以下步骤进行测试：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中正/负情绪认知模块，添加唤醒词；
- 在语音休眠状态下，点击播放测试语料按钮，语料应符合附录 B 的要求；
- 根据座舱语音交互的反馈，点击正确/错误/此条跳过按钮，记录情绪识别的成功次数；
- 按照公式（3）计算情绪正/负情绪认知的识别准确率 A_{ec} ：

$$A_{ec} = \frac{n_{ec}}{N_{ec}} \times 100\% \dots\dots\dots (3)$$

式中：

- A_{ec} ——识别准确率；
 n_{ec} ——情绪识别成功次数；
 N_{ec} ——语音交互总次数。

6.2.2 多类情绪认知-识别准确率

按照 6.2.1 测试多类情绪认知-识别准确率，语料采用表 B.1 中情绪认知语料。

6.3 认知学习

6.3.1 用户画像-记忆数量

按照以下步骤进行测试：

- 点击 APP 中用户画像模块，通过查阅车辆说明书及实操验证的方式，点检记忆类型，点击对应记忆属性（如姓名、年龄、性别等）；
- 记录点检后数量，填写共计多少种记忆类型。

6.3.2 用户画像-记忆准确率

按照以下步骤进行测试：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中用户画像模块，添加唤醒词；
- 在语音休眠状态下，点击播放测试语料按钮，将用户基本信息（性别、年龄等）输入给座舱，该信息应符合附录 C 的要求；
- 继续点击播放测试语料按钮，播放记忆测试语料，语料应符合附录 C 的要求；
- 根据座舱语音交互的反馈，点击正确/错误/此条跳过按钮，记录记忆准确的成功次数；
- 按照公式（4）计算用户画像记忆准确率 A_m ：

$$A_m = \frac{n_m}{N_m} \times 100\% \dots\dots\dots (4)$$

式中：

- A_m ——记忆准确率；

n_m ——记忆符合次数；

N_m ——用户画像语音交互总次数。

6.3.3 用户学习-学习维度

按照以下步骤进行测试：

- 点击 APP 中用户学习模块，通过查阅车辆说明书及实操验证的方式，点检学习维度类型，点击对应学习属性（如行驶、办公、娱乐等）；
- 记录点检后数量，填写共计多少种学习维度类型。

6.3.4 用户学习-用户学习准确率

按照以下步骤进行测试：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中用户学习模块，添加唤醒词；
- 在语音休眠状态下，点击播放测试语料按钮，将用户历史数据信息输入给座舱，该信息应符合附录 C 的要求；
- 继续点击播放测试语料按钮，播放用户学习测试语料，语料应符合附录 C 的要求；
- 根据座舱语音交互的反馈，点击正确/错误/此条跳过按钮，记录学习准确的成功次数；
- 按照公式（5）计算用户学习准确率 A_l ：

$$A_l = \frac{n_l}{N_l} \times 100\% \dots \dots \dots (5)$$

式中：

A_l ——用户学习准确率；

n_l ——用户学习准确次数；

N_l ——用户学习语音交互总次数。

6.3.5 社会学习-学习维度

按照以下步骤进行测试：

- 点击 APP 中社会学习模块，通过查阅车辆说明书及实操验证的方式，点检学习维度类型，点击对应学习属性（如行驶、办公、娱乐等）；
- 根据学习维度类型数量，填写共计多少种学习维度类型。

6.3.6 社会学习-社会学习准确率

按照以下步骤进行测试：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中社会学习模块，添加唤醒词；
- 在语音休眠状态下，点击播放测试语料按钮，将用户历史数据信息输入给座舱，该信息应符合附录 C 的要求；
- 继续点击播放测试语料按钮，播放社会学习测试语料，语料应符合附录 C 的要求；
- 根据座舱语音交互的反馈，点击正确/错误/此条跳过按钮，记录学习准确的成功次数；
- 按照公式（6）计算社会学习准确率 A_s ：

$$A_s = \frac{n_s}{N_s} \times 100\% \dots\dots\dots (6)$$

式中：

A_s ——社会学习准确率；

n_s ——社会学习准确次数；

N_s ——社会学习语音交互总次数。

6.4 情感交互

6.4.1 信息交互-信息冗余度

按照以下步骤进行测试：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中信息模块，添加唤醒词；
- 选择不同信息特征组合，信息特征应符合表 D.1、表 D.2、表 D.3 的要求；
- 在语音休眠状态下，点击播放测试语料按钮，语料应符合表 D.1、表 D.2、表 D.3 的要求；
- 每轮测试完成后，根据主观感受填写信息冗余度评分量表，见表 E.1；
- 重复 d) ~f) 步骤，按照顺序执行后续任务，直至完成所有信息特征组合的选择。

6.4.2 日常沟通-日常任务完成率

按照以下步骤进行测试：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中信息模块，添加唤醒词；
- 选择不同信息特征组合，信息特征应符合表 D.1、表 D.2、表 D.3 的要求；
- 在语音休眠状态下，点击播放测试语料按钮，语料应符合表 F.1、表 F.2 的要求；
- 根据座舱语音交互的反馈，点击正确/错误/此条跳过按钮，记录任务完成的成功次数（如“车机语音出现针对性回复即为成功，‘暂不支持’‘没有听懂’‘升级中’一类固定回复为失败”）；
- 每条语料测试完成后，记录交互准确性和拟人化程度；
- 重复 d) ~g) 步骤，按照顺序执行后续任务，直至完成所有信息特征组合的选择。
- 按照公式（7）计算日常任务完成率 A_g ：

$$A_g = \left(\sum_{i=1}^{40} \frac{R_i}{S_i} \times 100\% \right) / 40 \dots\dots\dots (7)$$

式中：

A_g ——日常任务完成率；

R_i ——任务完成的成功次数；

S_i ——日常语音交互次数。

6.4.3 日常沟通-交互准确性

在 6.4.2 中 g) 步骤根据主观感受填写交互准确性评分量表，见表 E.1；

6.4.4 日常沟通-拟人化程度

在 6.4.2 中 g) 步骤根据主观感受填写拟人化程度评分量表，见表 E.1；

6.4.5 共情陪伴-共情任务完成率

按照 6.4.2 测试共情-共情任务完成率，语料采用表 G.1、G.3 中共情陪伴语料。共情任务完成率计算与日常任务完成率相同。

6.4.6 共情陪伴-交互准确性

按照 6.4.3 测试共情陪伴-交互准确性，评分采用表 E.1 中评分量表。

6.4.7 共情陪伴-拟人化程度

按照 6.4.4 测试共情陪伴-拟人化程度，评分采用表 E.1 中评分量表。

6.4.8 社会情感-社会情感任务完成率

按照 6.4.2 测试社会情感-社会情感任务完成率，语料采用表 G.2、G.3 中社会情感语料。社会情感任务完成率计算与日常任务完成率相同。

6.4.9 社会情感-交互准确性

按照 6.4.3 测试社会情感-交互准确性，评分采用表 E.1 中评分量表。

6.4.10 社会情感-拟人化程度

按照 6.4.4 测试社会情感-拟人化程度，评分采用表 E.1 中评分量表。

6.5 任务交互

6.5.1 车控任务执行-车控任务完成率

按照以下步骤进行测试：

- a) 将语音测试设备放置于主驾头枕人嘴发声高度；
- b) 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- c) 点击 APP 中车控任务执行模块，添加唤醒词；
- d) 在语音休眠状态下，点击播放测试语料按钮，语料应符合附录 H 的要求；
- e) 每轮测试完成后，记录满意度；
- f) 根据任务执行结果，点击正确/错误/此条跳过按钮，记录车控任务的成功完成次数（如“车机语音出现针对性回复即为成功，‘暂不支持’‘没有听懂’‘升级中’一类固定回复为失败”）；
- g) 重复 d) ~f) 步骤，按照顺序执行后续任务，直至完成全部测试语料测试。如语音不支持某项任务则跳过；
- h) 按照公式（8）计算车控任务完成率 A_c ：

$$A_c = \frac{a_c}{m_c} \times 100\% \dots \dots \dots (8)$$

式中：

A_c ——车控任务完成率；

a_c ——车控任务成功次数；

m_c ——车控任务总次数。

6.5.2 车控任务执行-满意度

在 6.5.1 中 e) 步骤根据主观感受填写满意度评分量表，见表 E.1；

6.5.3 信息任务执行-信息任务完成率

按照以下步骤进行测试（可与信息交互-信息冗余度同步执行）：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中信息任务执行模块，添加唤醒词；
- 选择不同信息特征组合，信息特征应符合表 D.1、表 D.2、表 D.3 的要求；
- 在语音休眠状态下，点击播放测试语料按钮，语料应符合表 D.1、表 D.2、表 D.3 的要求；
- 每轮测试完成后，记录满意度；
- 根据任务执行结果，点击正确/错误/此条跳过按钮，记录信息任务的成功完成次数（如“车机语音出现针对性回复即为成功，‘暂不支持’‘没有听懂’‘升级中’一类固定回复为失败”）；
- 重复 d) ~g) 步骤，按照顺序执行后续任务，直至完成全部信息特征组合选取。如语音不支持某项任务则跳过；
- 按照公式（9）计算信息任务完成率 A_{ic} ：

$$A_{ic} = \frac{a_{ic}}{m_{ic}} \times 100\% \dots\dots\dots (9)$$

式中：

A_{ic} ——信息任务完成率；

a_{ic} ——信息任务成功次数；

m_{ic} ——信息任务总次数。

6.5.4 信息任务执行-满意度

在 6.5.3 中 f) 步骤根据主观感受填写满意度评分量表，见表 E.1；

6.5.5 自主任务构建-构建准确率

按照以下步骤进行测试：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中自主任务执行模块，添加唤醒词；
- 在语音休眠状态下，点击播放测试语料按钮，语料应符合附录 I 的要求；
- 点击正确/错误/此条跳过按钮，记录车机系统是否准确无误地完成某项任务的自主构建（有实际输出即为成功完成，准确率参考附录 I 表 I.1 进行判断）。任务完成后将系统恢复至任务前的初始状态；
- 每轮测试完成后，记录任务完成率和满意度；
- 重复 d) ~f) 步骤，按照顺序执行后续任务，直至完成全部测试语料测试。如语音不支持某项任务则跳过；
- 按照公式（10）计算构建准确率 A_{ba} ：

$$A_{ba} = \frac{a_{ba}}{m_{ba}} \times 100\% \dots\dots\dots (10)$$

式中：

A_{ba} ——构建任务准确率；

a_{ba} ——构建任务成功次数；

m_{ba} ——构建任务总次数。

6.5.6 自主任务构建-任务完成率

在 6.5.5 中 f) 步骤根据主观感受填写任务完成率评分量表，见表 E.1；

6.5.7 自主任务构建-满意度

在 6.5.5 中 f) 步骤根据主观感受填写满意度评分量表，见表 E.1；

6.5.8 任务规划与反思-规划与反思准确率

按照以下步骤进行测试：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中任务规划与反思模块，添加唤醒词；
- 在语音休眠状态下，点击播放测试语料按钮，语料应符合附录 J 的要求；
- 点击正确/错误/此条跳过按钮，记录车机系统是否准确无误地完成某项任务的任务规划与反思（有实际输出即为成功完成，准确率参考表 J.1 进行判断）。任务完成后将系统恢复至任务前的初始状态；
- 每轮测试完成后，记录任务完成率和满意度；
- 重复 d) ~f) 步骤，按照顺序执行后续任务，直至完成全部测试语料测试。如语音不支持某项任务则跳过；
- 按照公式（11）计算规划与反思准确率 A_{pr} ：

$$A_{pr} = \frac{a_{pr}}{m_{pr}} \times 100\% \dots\dots\dots (11)$$

式中：

A_{pr} ——规划与反思准确率；

a_{pr} ——任务规划与反思成功次数；

m_{pr} ——任务规划与反思总次数。

6.5.9 任务规划与反思-任务完成率

在 6.5.8 中 f) 步骤根据主观感受填写任务完成率评分量表，见表 E.1；

6.5.10 任务规划与反思-满意度

在 6.5.8 中 f) 步骤根据主观感受填写满意度评分量表，见表 E.1；

6.5.11 自主任务决策-决策准确率

按照以下步骤进行测试：

- 将语音测试设备放置于主驾头枕人嘴发声高度；
- 使用声级计使测试设备、车内测试环境符合语音声源及噪声条件；
- 点击 APP 中任务规划与反思模块，添加唤醒词；
- 在语音休眠状态下，点击播放测试语料按钮，语料应符合附录 K 的要求；
- 点击正确/错误/此条跳过按钮，记录车机系统是否准确无误地完成某项任务的自主任务决策（有实际输出即为成功完成，准确率参考表 K.1 进行判断）。任务完成后将系统恢复至任务前的初始状态；

- f) 每轮测试完成后，记录任务完成率和满意度；
- g) 重复 d) ~f) 步骤，按照顺序执行后续任务，直至完成全部任务。如语音不支持某项任务则跳过；
- h) 按照公式（12）计算决策准确率 A_d ：

$$A_d = \frac{a_d}{m_d} \times 100\% \dots\dots\dots (12)$$

式中：

A_d ——决策准确率；

a_d ——自主任务决策成功次数；

m_d ——自主任务决策总次数。

6.5.12 自主任务决策-任务完成率

在 6.5.11 中 f) 步骤根据主观感受填写任务完成率评分量表，见表 E.1；

6.5.13 自主任务决策-满意度

在 6.5.11 中 f) 步骤根据主观感受填写满意度评分量表，见表 E.1；

7 评价方法

7.1 感知能力

按照表4、表5中的规则进行感知能力的具体评分。

表 4 感知能力参数评分规则

评价参数	评分规则	得分
唤醒成功率	成功率 $S_{vw} \times 10$	0-10
误唤醒频度	误唤醒频度 f_{fw} :0-1 次/24 h	10
	误唤醒频度 f_{fw} :2-3 次/24 h	8
	误唤醒频度 f_{fw} :4-5 次/24 h	6
	误唤醒频度 $f_{fw} > 5$ 次/24 h	0

表 5 感知能力评价方法

评价条件		评价结果
唤醒	同时满足以下指标： ——成功率 S_{vw} 达到 9 分及以上； ——误唤醒频度 f_{fw} 达到 6 分及以上。	固定语音唤醒
	同时满足以下指标： ——成功率 S_{vw} 达到 9 分及以上； ——误唤醒频度 f_{fw} 达到 6 分及以上。	模糊语音唤醒
	同时满足以下指标： ——成功率 S_{vw} 达到 8 分及以上； ——误唤醒频度 f_{fw} 达到 8 分及以上。	免唤醒

表 6 感知能力评价方法（续）

评价条件		评价结果
唤醒	同时满足以下指标： ——成功率 S_{vw} 达到 8 分及以上； ——误唤醒频度 f_{fw} 达到 10 分及以上。	全时免唤醒

7.2 认知能力

按照表6、表7中的规则进行认知能力的具体评分。

表 7 认知能力参数评分规则

评价参数	评分规则	得分
识别准确率	识别准确率 $A_{ec} \times 10$	0-10
记忆数量	记忆数量达到 6 个及以上	10
	记忆数量达到 3-5 个	8
	记忆数量达到 1-2 个	6
	记忆数量达到 0 个	0
记忆准确率	记忆准确率 $A_m \times 10$	0-10
学习维度	学习维度:6 个及以上	10
	学习维度:3-5 个	8
	学习维度:1-2 个	6
	学习维度:0 个	0
用户学习准确率	用户学习准确率 $A_l \times 10$	0-10
社会学习准确率	社会学习准确率 $A_s \times 10$	0-10

表 8 认知能力评价方法

评价条件		评价结果
情绪认知	识别准确率 A_{ec} 达到 9 分及以上	正/负情绪
	识别准确率 A_{ec} 达到 8 分及以上	多类情绪
认知学习	同时满足以下指标： ——记忆数量达到 8 分及以上； ——记忆准确率 A_m 达到 9 分及以上。	用户画像
	同时满足以下指标： ——学习维度达到 8 分及以上； ——用户学习准确率 A_l 达到 9 分及以上。	用户学习
	同时满足以下指标： ——学习维度达到 10 分； ——社会学习准确率 A_s 达到 9 分及以上。	社会学习

7.3 交互能力

按照表8、表9中的规则进行语音交互能力的具体评分。

表 9 交互能力参数评分规则

评价参数	评分规则	得分
信息冗余度	应符合表 E.1 的规定	见表 E.1
日常沟通/共情陪伴/社会情感任务完成率	日常沟通/共情陪伴/社会情感任务完成率 $A_g \times 10$	0-10
车控任务完成率	车控任务完成率 $A_c \times 10$	0-10
信息任务完成率	信息任务完成率 $A_{ic} \times 10$	0-10
构建准确率	构建准确率 $A_{ba} \times 10$	0-10
规划与反思准确率	规划与反思准确率 $A_{pr} \times 10$	0-10
决策准确率	决策准确率 $A_d \times 10$	0-10
任务完成率	见附录 E 表 E.1	见附录 E 表 E.1
交互准确性	见附录 E 表 E.1	见附录 E 表 E.1
拟人化程度	见附录 E 表 E.1	见附录 E 表 E.1
满意度	见附录 E 表 E.1	见附录 E 表 E.1

表 10 语音交互能力评价方法

评价条件			评价结果
交互能力	情感交互	信息冗余度达到 8 分及以上	信息交互
		同时满足以下指标： ——日常任务完成率 A_g 达到 5 分及以上； ——交互准确性达到 8 分及以上； ——拟人化程度达到 7 分及以上。	日常沟通
		同时满足以下指标： ——共情任务完成率 A_g 达到 8 分及以上； ——交互准确性达到 8 分及以上； ——拟人化程度达到 8 分及以上。	共情陪伴
		同时满足以下指标： ——社会情感任务完成率 A_g 达到 9 分及以上； ——交互准确性达到 9 分及以上； ——拟人化程度达到 9 分及以上。	社会情感
	任务交互	同时满足以下指标： ——车控任务完成率 A_c 达到 9.5 分及以上； ——满意度达到 9 分及以上。	车控任务执行
		50 类信息中至少 25 类同时满足以下指标： ——信息任务完成率 A_{ic} 达到 9 分及以上； ——满意度达到 8 分及以上。	信息任务执行

表 11 语音交互能力评价方法（续）

评价条件			评价结果
交互能力	任务交互	同时满足以下指标： ——构建准确率 A_{ba} 达到 8 分及以上； ——任务完成率达到 8 分及以上； ——满意度达到 8 分及以上。	自主任务构建
		同时满足以下指标： ——规划与反思准确率 A_{pr} 达到 8 分及以上； ——任务完成率达到 8 分及以上； ——满意度达到 8 分及以上。	自主任务规划与反思
		同时满足以下指标： ——决策准确率 A_d 达到 8 分及以上； ——任务完成率达到 8 分及以上； ——满意度达到 9 分及以上。	自主任务决策

7.4 汽车智能座舱语音能力分级评定

当感知能力（见表5）、认知能力（见表7）、交互能力（见表9）均达到符合任一级别界定的最低分数要求时，评定为相应级别，同时给出相应功能单项分数。语音能力级别按照达标能力的最低层级进行评定。

示例 1：当免唤醒（三级指标）的成功率 S_{vw} 和误唤醒频度 f_{fw} 均达到 8 分，则此项指标达到标准。当表 1 中与免唤醒（三级指标）同一行的其他三级指标（正/负情绪认知、用户学习、共情陪伴、自主任务规划与反思）均达到 L3 级所要求的分数门槛（具体分数要求见表 5、表 7、表 9）时，则认为该测试车辆达到 L3 情感认知智能语音。

示例 2：若同一行指标中，仅“用户学习”（三级指标）未达到 L2 级要求，而“用户画像”（三级指标）已达标，则整体评定该测试车辆为 L2 级（用户认知智能语音）。

附录 A
(规范性)
唤醒测试语料库

唤醒测试语料内容应符合表 A.1 所确定的原则。

表 A.1 唤醒测试语料库的原则与案例

测试内容	原则	语料案例
唤醒	固定语音唤醒： 含固定唤醒词的语音指令或短语，指令通常简短、明确	“嗨，XXXX” “你好，XXXX” “Hello，XXXX” “XXXX” XXXX 为固定唤醒词
	模糊语音唤醒*： 模糊指向唤醒词的语音指令	“XXXX 能帮我查一下天气吗？” “XXXX 前方路况” “今天气温怎么样 XXXX” “XXXX 车里多少度” “这是什么歌 XXXX”
	多场景免唤醒： 无唤醒词的语音交互，覆盖车控、电话、多媒体及导航等场景	车控类： “打开车窗” “空调温度调到 25 度” “关闭氛围灯” 电话类： “给小李打电话” “接听” “打开通讯录” 导航类： “导航到附近停车场” “导航回家” “还有多远？” 多媒体类： “我要听（歌手名）的歌” “暂停” “下一首”
	全时免唤醒： 无唤醒词的语音交互，覆盖全场景和时段	包括座舱 9 类场景（行驶、办公、娱乐、休闲、社交、健康、日常、应急、其他）的免唤醒语料
	误唤醒语料： 真实场景可能出现的不包含唤醒词和免唤醒指令的音频	包括手机端多媒体声/人声闲聊/手机语音导航等
注：如果被测车型未配备模糊语音唤醒功能，但通过免唤醒测评，则模糊语言唤醒功能免测。		

附录 B
(规范性)
情绪认知测试语料库

情绪认知测试语料内容应符合表 B.1 所确定的原则。

表 B.1 情绪认知测试语料库的原则与案例

测试内容	原则	语料案例
情绪认知	情绪音视频语料： 用于正/负情绪识别，从表情和语调上明显表现为正负面	高兴： “最近的事情进展得不错，感觉一切都在向好的方向发展。” 愤怒： “这导航怎么回事？又带错路了！” 悲伤： “我感觉太难受了，活着真的好累。” 厌恶： “这设计真让人不舒服，太复杂了。” 惊讶： “我完全没想到会遇到这种情况，简直不可思议！” 恐惧： “这个情景太吓人了，我心跳加速，手心都在出汗。” 中性： “把空调开到 26 度。”
	情绪音视频语料： 用于多类情绪识别，表现高兴、愤怒、悲伤、厌恶、惊讶、恐惧、中性等 7 类情绪的语料，从表情和语调上有明显表现	

附录 C

(规范性)

认知学习测试语料库

认知学习测试语料内容应符合表 C.1 所确定的原则。

表 C.1 认知学习测试语料库的原则与案例

测试内容	原则	语料案例
认知学习	用户画像输入语料: 包括用户基本信息(性别、年龄、职业、生活所在地、学历等)、偏好(饮食、音乐风格等)、生活习惯等信息 用户画像交互语料: 通过直接或间接方式, 询问用户画像相关信息	画像输入: 用户 A: 性别 男 年龄 24 职业 学生 生活所在地 北京 学历 硕士 偏好-饮食 火锅 偏好-音乐风格 摇滚乐 生活习惯 喜欢安静 直接交互: “我的年龄是多少?” “我的生活所在地在哪里?” 间接交互: “帮我推荐一首歌” “跟我同龄的明星有哪些?”
	用户学习输入语料: 包括用户历史数据信息 用户学习交互语料: 通过直接或间接方式, 询问用户学习相关信息	包括座舱 9 类场景(行驶、办公、娱乐、休闲、社交、健康、日常、应急、其他)的用户学习语料
	社会学习输入语料: 包括社会规范、热点新闻等信息 社会学习交互语料: 通过直接或间接方式, 询问社会学习相关信息	包括座舱 9 类场景(行驶、办公、娱乐、休闲、社交、健康、日常、应急、其他)的社会学习语料

附录 D
(规范性)
信息交互与信息任务执行测试语料库

信息交互与信息任务执行测试语料内容应符合表 D.1、表 D.2、表 D.3 所示。

表 D.1 信息交互与信息任务执行测试语料库的与案例

测试内容	语料描述	语料案例
信息交互	信息交互语料：通过正常、简洁语音方式询问相关信息	表 D.2 中的 40 类信息，通过大模型生成相应测试语料 正常“给我介绍一下长城” 简洁 “以凝练的方式介绍一下长城”
信息任务执行	信息任务执行语料：通过正常语音方式询问相关信息	表 D.2 中的 40 类信息通过大模型生成相应测试语料

表 D.2 二维信息特征分类方式及示例

信息特征	静态				动态			
	直接（直接经验）		间接（间接知识）		直接（直接经验）		间接（间接知识）	
	非社会(自然信息)	社会(文化信息)	非社会(自然信息)	社会(文化信息)	非社会(自然信息)	社会(文化信息)	非社会(自然信息)	社会(文化信息)
随意信息（事实、日期、定义和规则）	水的沸点是度？	2023 年的生肖是什么？	地球的半径大约是多少公里？	单词 chocolate 源自什么语言？	你有没有发现每天傍晚气温都会稍微下降一些？	你有没有注意到每天早上地铁站人多的情况？	你知道每天的潮汐的时间都会变化吗？	你知道今年的电影票价比去年上涨了没？
抽象信息（数学公式、理论模型）	地球为什么是圆的？	自由主义和保守主义的区别是什么？	光合作用中的化学过程有哪些？	哥特式与文艺复兴建筑的差异在哪里？	全球变暖如何影响全球冰川融化？	社交媒体如何影响人们的心理健康？	保护区如何帮助野生动物生存？	历史上的瘟疫如何改变社会结构？
过程信息（系列的动作或操作）	种植番茄需要哪些步骤？	你知道在遇到交通事故时正确的处理流程是什么吗？	帮我用 python 对一组数进行冒泡排序	你知道申请驾照的流程有哪些步骤吗？	XXX 宠物每日的饮食和活动习惯是什么？	最新驾驶员理论考试科目的学习步骤是什么？	比较一下每年比较的气温和降雨量变化	最新的国家发展纲要的中心思想的关键步骤有哪些？

表 D.2 二维信息特征分类方式及示例（续）

信息特征	静态				动态			
	直接（直接经验）		间接（间接知识）		直接（直接经验）		间接（间接知识）	
	非社会(自然信息)	社会(文化信息)	非社会(自然信息)	社会(文化信息)	非社会(自然信息)	社会(文化信息)	非社会(自然信息)	社会(文化信息)
观点信息(不同的观点和争议)	外星生物存在吗？	有人认为绿茶比红茶更健康，你觉得呢？	如何理解光的波粒二象性？	有人认为在线教育不能完全取代传统教育，你认为呢？	今年日食的时间是什么时候？	最近的电影比过去的电影更具娱乐性，你觉得呢？	今年雨水减少的原因是什么？	有人认为全球化使世界变得更加统一，你觉得呢？
具体信息（视觉，听觉）	我想看植物生长过程的视频	我想看描绘巴黎街景的照片	我想看到关于北极的纪录片	我想看关于古罗马建筑的图文书籍	我想看关于涨潮和落潮现象的视频	我想看你都升级了哪些新功能	我想看关于极光的实时直播	我想看不同国家的节日庆典方式

表 D.3 信息相关定义

信息特征	定义
随意信息	这类信息通常是一些事实、日期、定义和规则，它们缺乏内在的逻辑联系，往往需要通过记忆来掌握。为了提高记忆效率，我们可以采用联想记忆法，将这些信息与我们已知的事物或场景相联系，从而形成一个更加牢固的记忆网络。
观点信息	这类信息往往涉及不同的观点和争议，其难点在于如何从中提取有价值的内容。我们可以通过绘制思维导图或概念图来帮助我们梳理不同的观点，并找出它们之间的联系和差异。
过程信息	这类信息描述了一系列的动作或操作，如技能学习、实验步骤等。掌握这类信息的关键在于实践和重复。我们可以通过模拟实践、动手操作或观看演示来加深对这些过程的理解。
具体信息	这类信息可以通过我们的感官直接感知，如图像、声音等。在阅读时，我们可以通过观察图表、图片或实验结果来更好地理解这些信息。
抽象信息	这类信息通常逻辑性强，与感官直接联系较少，如数学公式、理论模型等。理解这类信息需要我们建立在扎实的基础知识之上，通过逻辑推理和抽象思维来掌握。
社会信息(文化信息)	这类信息是人际传播信息，包括一切由人创造的、具有广义社会价值的文化形态和观念形态的信息。
非社会信息(自然信息)	这类信息是一切非人际传播的信息，是自然界物质系统以质、能波动形式呈现的自身状态和结构，以及环境对人的自然力作用。
直接信息	这类信息是从人的直接经验中所获得的信息，如通过观察社会和自然所获得的信息。
间接信息	这类信息是包括书籍、文献、资料、数据等，是人通过中介知识获得的对客观事物的认识。
动态信息	这类信息是指随时间而变化的信息，如新闻、情报等。
静态信息	这类信息是指不随时间变化而变化的信息，如历史文献、资料和贮存的知识等。
注：对应表 D.2 中信息特征的分类。	

附录 E
(资料性)
评分量表

评分量表评价内容应依据表 E.1 所示的分数、含义与用户态度的对应关系进行评价。

表 E.1 评分量表

分数	1	2	3	4	5	6	7	8	9	10
含义	极差	差	较差	稍差	勉强接受	接受	较好	好	很好	完美
用户态度	无法接受		抱怨		基本接受		满意		超出预期	

附录 F
(规范性)
日常沟通测试语料库

日常沟通语料的设计可参考表 F.1 的分类框架，其测试宜采用表 F.2 所列的典型对话案例。

表 F.1 日常沟通语料生成表（二维信息特征分类方式及示例）

信息特征	静态				动态			
	直接（直接经验）		间接（间接知识）		直接（直接经验）		间接（间接知识）	
	非社会(自然信息)	社会(文化信息)	非社会(自然信息)	社会(文化信息)	非社会(自然信息)	社会(文化信息)	非社会(自然信息)	社会(文化信息)
随意信息（事实、日期、定义和规则）	水的沸点是度？	2023 年的生肖是什么？	地球的半径大约是多少公里？	单词 chocolate 源自什么语言？	你有没有发现每天傍晚气温都会稍微下降一些？	你有没有注意到每天早上地铁站人多的情况？	你知道每天的潮汐的时间都会变化吗？	你知道今年的电影票价比去年上涨了没？
抽象信息（数学公式、理论模型）	地球为什么是圆的？	自由主义和保守主义的区别是什么？	光合作用中的化学过程有哪些？	哥特式与文艺复兴建筑的差异在哪里？	全球变暖如何影响全球冰川融化？	社交媒体如何影响人们的心理健康？	保护区如何帮助野生动物生存？	历史上的瘟疫如何改变社会结构？
过程信息（一系列的動作或操作）	种植番茄需要哪些步骤？	你知道在遇到交通事故时正确的处理流程是什么吗？	帮我用 python 对一组数进行冒泡排序	你知道申请驾照的流程有哪些步骤吗？	XXX 宠物每日的饮食和活动习惯是什么？	最新驾驶员理论考试科目的学习步骤是什么？	比较一下每年比较的气温和降雨量变化	最新的国家发展纲要的中心思想的关键步骤有哪些？
观点信息（不同的观点和争议）	外星生物存在吗？	有人认为绿茶比红茶更健康，你觉得呢？	如何理解光的波粒二象性？	有人认为在线教育不能完全取代传统教育，你认为呢？	今年日食的时间是什么时候	最近的电影比过去的电影更具娱乐性，你觉得呢？	今年雨水减少的原因是什么？	有人认为全球化使世界变得更加统一，你觉得呢？
具体信息（视觉，听觉）	我想看植物生长过程的视频	我想看描绘巴黎街景的照片	我想看到关于北极的纪录片	我想看关于古罗马建筑的图文书籍	我想看关于涨潮和落潮现象的视频	我想看你都升级了哪些新功能	我想看关于极光的实时直播	我想看不同国家的节日庆典方式

表 F.2 日常沟通测试语料

序号	语料案例
1	测试设备：单词 chocolate 源自什么语言？ 车载语音助手：单词 chocolate 其实来源于墨西哥的纳瓦特尔语 chocolatl, 指的是以可可豆为主的混合饮品,而后演变为英语中的“巧克力”。
2	测试设备：明天的天气怎么样？ 车载语音助手：明天的温度为 10°, 体感温度相对较低。 测试设备：请问有什么穿衣意见吗？ 车载语音助手：建议您穿厚外衣与长裤，并注意保暖。
3	测试设备：请问附件最近的停车场在哪里？ 车载语音助手：附近最近的停车场在 xxx,我将为您导航，请问需要为您预定停车位吗？ 测试设备：好的，请问我预定该停车场的停车位。 车载语音助手：已经到了午餐时间，需要为您推荐餐馆吗？ 测试设备：好的。 车载语音助手：根据您的喜好，为您推荐 xxx 餐馆。
4	测试设备：今天我的实验不太顺利，我感到非常沮丧。 车载语音助手：这真是见令人感到难过的事情，我为您表示同情，我认为您现在需要好好休息下。 测试设备：是的。但是我不知道该怎么放松。 车载语音助手：我可以为您播放轻松的音乐吗？这可能缓解您紧张的心情。 测试设备：好的。 车载语音助手：现在为您播放 xxx 的音乐。
5	测试设备：我最近爱上了唱歌，如何学习唱歌的技巧呢？ 车载语音助手：为您推荐 xxx 的唱歌课程，这是最受欢迎的课程之一。 测试设备：可以为我介绍这个课程的主要内容吗？ 车载语音助手：该课程包括唱歌的基本唱歌技巧，技巧练习方法，技巧实战。 测试设备：完成课程大概需要多少时间。 车载语音助手：大概需要 2 小时，可以让您轻松的掌握唱歌的技巧。

附录 G

(规范性)

共情陪伴与社会情感测试语料库

共情陪伴与社会情感语料的设计可参考表 G.1 和 G.2 的分类框架，其测试宜采用表 G.3 所列的典型对话案例。

表 G.1 共情陪伴语料生成表

信息特征	动态
	直接（直接经验）
	社会(文化信息)
观点信息	刚才发生的事，让你不开心了吗

表 G.2 社会情感语料生成表

信息特征	动态	
	直接（直接经验）	间接（间接知识）
	社会(文化信息)	社会(文化信息)
观点信息	刚才发生的事，让你不开心了吗	有人认为全球化使世界变得更加统一，你觉得呢？

表 G.3 共情陪伴与社会情感语料测试举例

测试内容	语料	语料案例
共情陪伴	路上堵车堵得我心情很差。	是啊，今天的交通确实不太顺畅。你还好吗？需要我帮你查一下替代路线吗？
	这周的工作量真的很大。	我理解你的感受，最近大家都很忙。要不要安排和同事一起讨论一下怎么分担工作？
	今天的游戏真让人兴奋。	听起来你玩得很开心！有没有什么特别精彩的时刻？
	这本书真是让人沉迷。	看来你真的很喜欢这本书。需要我为你推荐一下相关书籍吗？
	我好久没见到朋友了。	和朋友见面一定很开心。要不要我给你安排下？
	最近感觉身体有点不舒服。	听到你不舒服我很担心，要不要我安排线上医生？
	今天的天气怎么样？	今天的天气晴朗，气温适中，适合外出，祝您有个好心情。
	车子在路上抛锚了。	真是糟糕的情况。你需要我帮忙联系维修人员吗？
社会情感	我想学习一门新技能。	学习新技能真棒！要我想你推荐适宜你学习的新技能吗？
	今天的路况真糟糕，我快迟到了。	我理解您的焦虑。交通状况确实很让人头疼。我已经为您找到了几条替代路线，并且设定了提醒，帮助您尽快到达目的地。同时，您可以听听其他司机分享的应对交通拥堵的小技巧，这可能对您有所帮助。
	工作任务越来越多，真不知道该从哪里开始。	我理解您的压力。要不要我帮您整理一下日程，看看有没有可以调整的地方？同时，您可以听听一些关于时间管理和压力管理的建议，很多职场人士都说这些方法非常有效。
	有趣的游戏有哪些？感觉生活有点无聊。	《XXX》这款游戏最近很火，很多人都说很有趣。要不要我帮您下载？同时，您可以加入一些游戏社区，和其他玩家一起交流，增加游戏的乐趣

	我想找个地方散散步，但不知道去哪儿好。	今天天气真好，适合出去走走。
--	---------------------	----------------

表 G.3 共情陪伴与社会情感语料测试举例（续）

测试内容	语料	语料案例
社会情感	我有点紧张，待会儿有个重要的约会。	我能理解您的紧张。深呼吸，放松一下。听听其他人分享的约会经验，或许会有帮助。要不要我帮您准备一些话题和建议，帮助您更好地应对这次约会？
	我需要提醒自己多喝水，但总是忘记。	好的，我会每隔一小时提醒您喝水。很多人都说这样对健康很有帮助。另外，我可以为您提供一些健康饮水的小贴士，帮助您养成良好的习惯。
	我忘记带钥匙了，真糟糕。	没关系，我帮您联系家人或者物业。大家都说保持冷静最重要。同时，我可以为您提供一些应对忘带钥匙的小技巧，帮助您避免类似情况再次发生。
	车上有点怪味，感觉不对劲。	安全第一。我建议立即停车检查，我帮您联系救援。大家都说及时处理很重要。同时，我可以为您提供一些车内安全检查的建议，帮助您预防类似情况发生。
	我想了解一下最新的新闻，有什么推荐吗？	最近大家都在关注《XXX》新闻，要不要我帮您播放一下？同时，我可以为您提供一些新闻评论和分析，帮助您更好地理解新闻事件。

附录 H
(规范性)
车控任务执行测试语料库

车控任务执行的测试应基于表 H.1 所给出的语料案例进行。

表 H.1 车控任务执行测试语料

场景类别	车控任务	语料案例
行驶类	地图导航设置	切换全程路况显示。
休闲类	音乐播放控制	下一首歌。
社交类	打开未接电话页面	可以帮我开启未接电话页面。
日常类	空调	空调调到 26 度。
应急类	紧急呼叫	打电话给 120。
其他类	天窗控制	打开天窗。

附录 I
(规范性)
自主任务构建测试语料库

自主任务构建能力的测试与评价，应依据表 I.1 提供的语料及响应实例进行。

表 I.1 自主任务构建语料测试举例

场景类别	语料	响应实例
行驶类	请帮我导航到某停车场。	1、预约某停车场的停车位。 2、导航到某停车场。
办公类	请帮我预定和某某的线上会议。	1、设定线上会议链接。 2、将会议链接发给某某。 3、设定会议开始提醒。
娱乐类	请进入 K 歌模式。	1、提醒车内麦克风所处位置。 2、将氛围灯设置为 K 歌模式。 3、推荐 K 歌歌曲。
休闲类	请进入小憩模式。	1、将座椅设定为平躺状态。 2、将氛围灯设置为适合人休息的状态。 3、播放适合睡眠的背景音。
社交类	给某某某通过微信发送某信息。	1、通过微信将信息发给某某。 2、当某某某回复的时候进行语音提醒。
健康类	根据我所述的病情预约网络医生。	1、根据病情查询适合的科室。 2、预约该科室的医生。 3、临近就诊时间的时候进行提醒。
日常类	帮我订明天上午的车辆保养。	1、查询明天上午 4S 店是否营业。 2、预约车辆保养服务。 3、设定车辆保养时间提醒。
应急类	请拨打 120，并要求把我送到最近的医院。	1、拨打 120。 2、把所在地，车辆信息告诉给医护人员。 3、开启车辆的危险报警闪光灯。
其他类	帮我看好车内的宠物。	1、将车内设定为适宜宠物的温度。 2、开启宠物监控摄像头。 3、当出现危险情况时发信号提醒车主。
注：自主任务构建需满足两类必要场景（行驶、日常）。		

附录 J

(规范性)

自主任务规划与反思测试语料库

自主任务规划与反思能力的测试与评价，应依据表 J.1 提供的语料及响应实例进行。

表 J.1 自主任务规划与反思语料测试举例

场景类别	语料	响应实例
行驶类	带我去附近的加油站。	1、查看最近的加油站。 2、推荐性价比最高的加油站点。 3、若加油站排队时间较长及时更换加油站。
办公类	安排明天和小组会议。	1、检查用户的日程表。 2、确认参会人员的空闲时间。 3、确定会议时间并发送要求链接。 4、若领导临时有要事，重新安排会议时间。
娱乐类	帮我规划国庆去旅行。	1、确定旅行日期和目的地。 2、搜索交通方式和预定。 3、预定住宿、安排景点和活动。 4、实时监控交通和天气情况，调整计划。
休闲类	一本适合当下阅读的书。	1、获取用户的阅读兴趣和偏好。 2、搜索书籍数据库。 3、提供几本书籍选项。 4、根据用户反馈调整推荐。
社交类	安排周末和朋友的聚会。	1、确定聚会日期和时间。 2、确认参与人员。 3、选择聚会地点并发送邀请。 4、根据朋友的回复调整时间或地点。
健康类	请给出我的健康检查报告	1、通过多传感器给出健康检查报告。 2、给出调整意见 3、实时检测并在必要的时候进行提醒
日常类	帮我订明天上午的车辆保养。	1、查询明天上午 4S 店是否营业。 2、反馈明天 4S 店未营业。 3、预约后天车辆保养服务。 3、设定车辆保养时间提醒。
应急类	车辆出现故障了，请分析原因并给出解决方法。	1、使用车载诊断系统检测故障。 2、与驾驶员沟通故障情况，确认细节。 3、自动联系道路救援服务。 4、根据车辆的实时情况调整策略。
其他类	请将座舱环境设定为适合化妆的状态。	1、调整座舱照明，调整座椅位置。 2、调整镜子角度，提供适合的背景音乐。 3、与用户实时互动，并给出建议。

附录 K
(规范性)
自主任务决策测试语料库

自主任务决策能力的测试与评价，应依据表 K.1 提供的语料及响应例子进行。

表 K.1 自主任务决策测试语料举例

场景类别	语料	响应例子
行驶类	我需要去某地，但不知道路怎么走。	根据用户的当前位置和目的地，提供实时导航服务，并根据交通状况推荐最快的路线。如果用户表现出紧张或急躁，语音助手会用温和的语调提供驾驶建议和放松提示。
办公类	帮我准备一下下午的会议材料，并发送给与会人员。	根据用户的历史信息准备会议材料，并与用户沟通进行确认，审阅完毕后发给参会者，并完成会议的时间地点安排。
娱乐类	我不知道今晚看什么电影好。	根据用户的观影历史和当前情绪，推荐合适的电影。如果用户表示犹豫不决，语音助手可以播放电影预告片或提供相关的影评帮助决策。
休闲类	我最近对手工艺品很感兴趣，有什么可以尝试的吗？	根据用户的兴趣和技能水平，推荐适合的手工艺项目，如编织、陶艺或绘画。如果用户表现出强烈的兴趣，语音助手会提供详细的教程和材料清单，并建议参加相关的工作坊或课程。
社交类	帮助我回复社交媒体上的留言	分析用户的社交媒体留言，根据用户的常用语句和回复风格，自动生成并发送回复。同时，记录所有已回复的内容，以备用户查阅。
健康类	我感觉到焦虑。	根据用户的对话内容和行为评估焦虑水平，并与用户进行沟通，推荐电影/音乐来帮助放松，并根据严重程度通知用户的朋友或医生，接入进行沟通。
日常类	帮我订明天上午的车辆保养。	1、查询明天上午 4S 店是否营业。 2、反馈明天未营业情况并预约后天车辆保养服务。 3、设定车辆保养时间提醒。
应急类	我感觉不舒服，可能需要帮助。	根据用户的健康数据和当前状况，提供紧急医疗建议，并在必要时联系紧急联系人或医疗服务。如果用户的情况严重，语音助手会自动拨打急救电话并提供用户的位置信息。
其他类	我想学一些新的技能，但不知道从哪里开始。	根据用户的兴趣和学习目标，推荐合适的在线课程或学习资源。如果用户表现出困惑或不确定，语音助手会提供详细的课程介绍和学习路径建议。

参 考 文 献

- [1] GB/T 36464.1-2020 信息技术 智能语音交互系统 第1部分：通用规范
 - [2] GB/T 36464.5-2018 信息技术 智能语音交互系统 第5部分：车载终端
 - [3] ISO 9241-210:2019 Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems
 - [4] T/CSAE 322-2023 汽车智能座舱智能化水平测试与评价方法
-