ICS 33.050
M 30

团 体 标 准

T/TAF 041-2019

智能产品语音识别测评方法——第一部分 车载语音交互系统

Testing Methods for Speech Recognition of Intelligent Products

——The First Part Speech Interaction System in Vehicle

2019-06-17 发布 2019-06-17 实施

电信终端产业协会 发布

目录

前　　言

本标准按照 GB/T 1.1-2009给出的规则编写。

本标准由电信终端产业协会提出并归口。

本标准起草单位：中国信息通信研究院。

本标准主要起草人：张文奇、马治国、赵兴龙、胡景楠、朱红叶、孙悦、卢玉平、李俊林

引　　言

近年来，语音识别技术广泛应用于终端的人机界面和应用输入中。特别是对于车载交互系统来说，驾车过程中语音识别技术使得驾驶员可以在双手不离开方向盘，视线不离开路面的情况下，完成对汽车辅助功能的操控，提高了安全性和快捷性。语音识别技术可以使传统繁琐的手动控制变为语音控制，极大提高了驾驶员操作的方便性。作为最人性化的人机交互方式，语音识别技术在车载中的应用得到了广泛关注。大量已上市和开发中的车载交互系统产品都以语音识别作为重要功能。

由于目前车载交互系统的语音识别性能参差不齐，没有一个通用可参照的标准对车载设备的性能进行评估，无法准确衡量其性能，因此有必要对车载交互系统的语音识别提出相应的技术要求和测试方法。

智能产品语音识别测评方法——第一部分车载语音交互系统

1 范围

本标准从影响车载交互系统的语音识别性能的各技术角度出发，制定相应的测试方法和技术要求。

本标准适用于车载终端设备配置的中文语音识别系统。本标准的制定和实施主要用于指导前装的车载语音交互系统。后装的车载语音交互系统可参考本标准。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注明日期的引用文件，其随后所有的修改版（不包括勘误的内容）或修订版均不适用于本标准，然而，鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件，其最新版本适用于本标准。

[GB/T 21023]（2007） 中文语音识别系统通用技术规范

[ITU-T P.56]（12/2011） 激活语音电平的客观测量（Objective measurement of active speech level）

[ITU-T P.581]（2000） HATS在免提终端测试中的使用(Use of head and torso simulator (HATS) for hands-free terminal testing)

[ITU-T P.851] 基于口语对话系统的电话服务的主观质量评价（Subjective quality evaluation of telephoneservices based on spoken dialogue systems）

3 术语、定义和缩略语

3.1 术语和定义

下列术语和定义适用于本标准。

3.1.1 语音识别 speech recognition

将人类的声音信号转化为文字或者指令的过程。

3.1.2 语音识别系统 speech recognition system

具有语音识别功能的开发工具、软件、装置或应用。

3.1.3 车载系统 in-vehicle system

能产生人类智能行为的车载计算机系统，可为用户提供车辆控制、安全、信息、娱乐等方面的功能或服务。

3.1.4 车载语音交互系统 in-vehicle speech interaction system

实现人类与车载系统间语音交互的系统，以将人类的语音输入转化为车载系统可识别的控制指令，同时将车载系统的反馈信息通过语音或文字输出为主要目的。

3.1.5 识别决策 recognition strategies

根据客观的可能性，以已知的信息和知识为基础，借助一定的方法对识别目标的诸多可能情况进行分析、计算和选优后，做出的行动决定。

3.1.6 语音唤醒 speech wakeup

车载语音识别系统在睡眠模式下自动检测背景语音中的唤醒词，在成功匹配的情况下转入正常工作模式。

3.1.7 响应时间

对于特定的语音识别任务，若语音输入的结束时刻为t_e；车载语音识别系统的开始响应时刻为t_r。则：响应时间= t_r-t_e

3.2 缩略语

下列缩略语适用于本标准。

HATS HATS Head And Torso Simulator 头和躯干模拟器

4 车载语音交互系统分类及表述

4.1 概述

车载语音交互系统根据基本属性如应用场景、词汇量、应用人群、工作模式、应用环境等进行分类。从用户感受的角度触发，把车载语音交互系统当作黑匣子，车载语音交互系统的性能指标仅基于系统的应用场景属性提出。

4.2 基于应用场景的分类

车载语音交互系统基于应用场景可以分为：文字输入类、声音检测识别类和对话类三类系统，或兼容三种应用场景。

4.2.1 文字输入类

以文字录入为主要目的，要求把语音转化成文字的系统，如短消息功能。

4.2.2 声音检测识别类

是指根据用户语音中发出的特定命令或者关键词，完成特定操作和业务的应用场景。如拨打电话、音乐命令控制等。

4.2.3 口语对话类

是指接受用户以对话形式发出的自然的口头语言，明白及理解用户意图及想要获取的信息，并将以各种形式与用户进行反馈，以将对话继续进行的应用。如地图查询，天气查询等。

4.3 车载语音交互系统的结构表述

语音交互系统可分为前端语音预处理模块、语音识别模块、语义理解模块、交互决策模块和语音合成模块。前端语音预处理模块负责将语音输入转化为语音流，作为语音识别模块的输入。语音识别模块负责将语音流转换为人类可识别的文本信息直接输出到相关的应用模块，或转换为计算机可识别的字符串输出到语义理解模块。语义理解模块负责对语音识别模块的识别结果做语义解析。识别决策模块负责根据语义理解模块“理解的”的结果制定识别决策，并依此向相关应用模块下达控制指令及获取反馈信息。语音合成模块负责将识别决策模块或应用模块提供的计算机可识别的文本信息转换为语音信号输出。某些子模块可选择在本地、云端或融合实现。

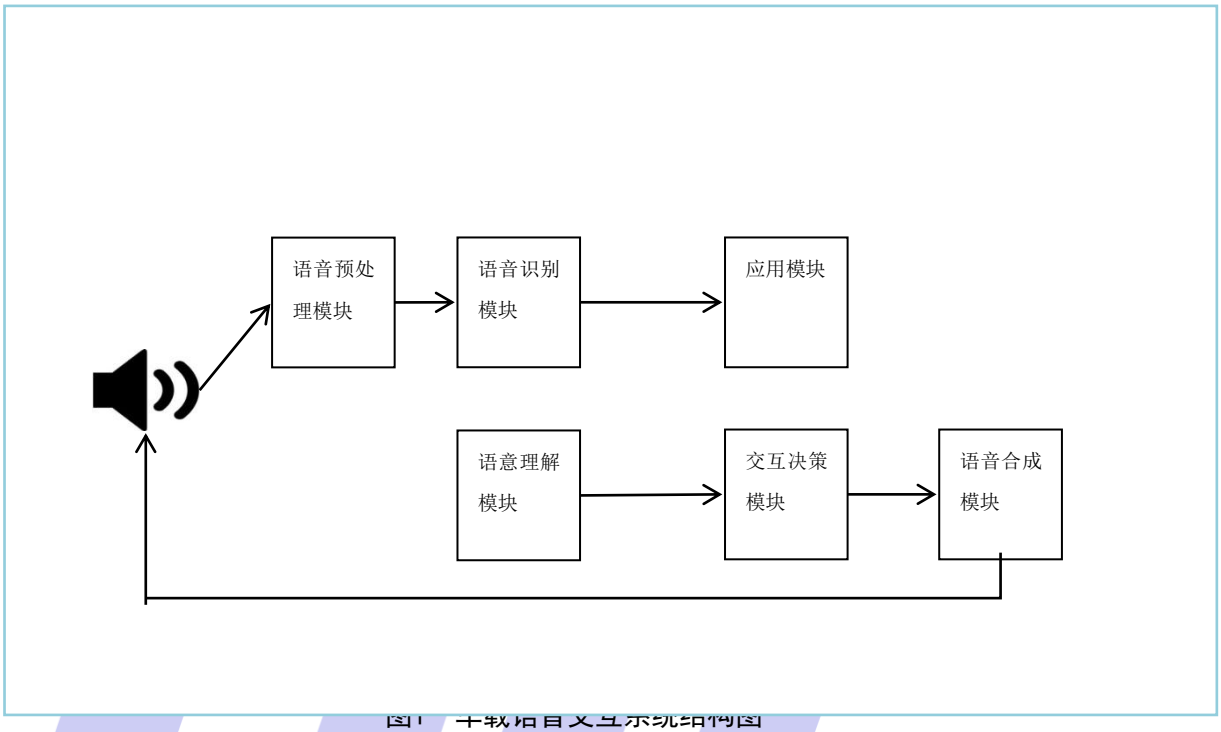

图1 车载语音交互系统结构图

5 车载语音交互系统的性能指标要求

5.1 概述

车载语音交互系统的性能需满足以下指标要求，其中唤醒率和误唤醒率仅针对支持语音唤醒功能的车载语音交互系统。这些要求与系统的用途有关，多用途的需求分别满足多指标的要求。系统给出的具体指标应明确在何种条件下成立。

5.2 字准确率

字准确率的性能指标定义详见 GB/T 21023 中 5.2.1 的内容。

该指标用于评价车载语音交互系统输出的人类可识别的文本信息的正确率。对于中文普通话车载语音交互系统，该项指标的评价分类如表 1 所示。

表1 中文普通话感受效果分类

识别率	评价效果
≥90%	优秀

＜90%且＞70%	可接受
≤70%	不可接受

5.3 识别成功率

若车载语音交互系统在既定的识别轮数内完成了语音识别任务，则此次语音识别成功。语音识别成功与否应兼顾车载系统动作的可靠性问题。若车载语音识别系统共进行了 R 次特定的语音识别任务，其中 SR 次识别成功，FR 次识别出现误操作（包括未在既定的识别轮数内完成的识别、未完成识别前退出、识别无响应和错误识别）。则：

识别成功率=SR/R×100%；

误操作率=FR/R×100%；

识别成功率+误操作率=1。

该指标用于评价车载语音交互系统对语音识别任务的正确响应情况。对于中文普通话车载语音交互系统，该项指标的评价分类如表 1 所示。

在如表 2 所述的测试环境下，具体要求如下：场景 1 的识别成功率应≥80%；场景 2 的识别成功率应≥75%；场景 3 的识别成功率应≥70%。

5.4 平均响应时间

该指标用于评价车载语音交互系统对语音识别任务的响应速度。离线车载语音交互系统的平均响应时间应≤2s；在线车载语音交互系统的平均响应时间应≤3s。

5.5 唤醒率

若车载语音交互系统共进行了W次唤醒，其中SW次成功唤醒。则：

唤醒率=SW/W×100%。

该指标用于评价车载语音识别系统在叠加背景音乐的情况下，对语音唤醒操作的正确响应情况。在表2所述的测试环境下叠加低档背景音乐，场景1的唤醒率应≥95%；场景2的唤醒率应≥88%。

5.6 误唤醒率

若车载语音交互系统在T小时内出现FW次误唤醒。则：

误唤醒率=FW/T 次/小时。

该指标用于评价车载语音交互系统在叠加背景语音的情况下，对语音唤醒操作的错误响应情况。在表 2 所述的测试环境下叠加低档背景音乐，误唤醒率应≤0.5 次/小时。

6 车载语音交互系统测试方法

6.1 概述

为保证车载语音交互系统测试的可重复性和性，应采用基于车载语音识别系统测试标准库的测试方法。语音识别标准库的建立应按照 GB/T 21023 中的要求进行。车载语音识别系统测试标准库应同其开发库独立同分布。测试语料的设计与测试语音/环境噪声的录制应保证与实际应用场景的一致性。

影响车载语音识别性能的因素包括不同用户、语言种类、口音、发音、语速、词汇量、语境、噪声环境。测试时应充分考虑车载语音交互系统的设计目标和各因素下对性能指标影响。

6.2 测试语料设计

车载语音交互系统测试语料设计原则应符合语音识别系统输入准则：

a) 中文车载语音交互系统应支持汉语通用语,从信息社会发展角度支持推广普通话。

b) 语音输入标准语速为 180-300 字/min，时长一般不超过 30s，特殊情况下不超过 60s。

c) 发音单元的持续时间不应小于 0.2s，发音单元间的间隔不超过 2s；停顿时间超过 2s，则认为一次语音输入结束。

d) 对于文本中的分汉字字符，包括数字、电话号码、标点符号，其可参照 GB/T 21023-2007 附录 A 所定义的方法朗读。

测试语料应从词汇量覆盖、开放业务覆盖、音节覆盖，以及常用性角度加以设计，设计要求如下：

e) 对于命令词识别车载语音识别系统，测试语料应覆盖被测系统的所有词汇。建议测试语料的规模不小于 200 句。

f) 对于连续语音识别车载语音识别系统，测试语料应尽量覆盖被测系统的词汇，对于每种开放业务从音节覆盖和常用性角度挑选典型语料。建议每种开放业务测试语料的规模为 500 句。

6.3 测试语音/环境噪声录制

测试语音/环境噪声录制要求如下：

g) 对于非特定人群车载语音识别系统,特别强调对含有一定口音的汉语语音适应能力和汉语声调问题。

h) 测试语音至少应由男女各 30 名以上的发音人录制，用于语音唤醒功能的测试语音需要 50-100 名以上的发音人录制。应在符合系统对识别人群限制的条件下，尽量选择具有代表性和统计分布规律的发音人，特别是应考虑不同口音、不同的普通话等级、不同年龄、不同语速、不同教育背景、不同说话韵律等因素。对于命令词识别车载语音识别系统，应尽量保证由各发音人分别录制全部测试语料。对于连续语音识别车载语音识别系统，每组测试语料不应超过 100 句，尽量保证由不同的发音人录制。

i) 测试语音可以是发音人的语音或通过高保真设备回放的测试语音。测试语音文件的录制格式为.wav，纯净测试语音的录制应保证 44.1kHz 的采样频率和 16bit 的量化间隔，发音人和麦克风间应保持一定距离（如 15cm,避免风噪的影响），并确保波形采样范围为±5000—±10000smpl；录制过程至少应包括录音、标注和确认三个步骤，以保证测试标准库的正确性。录制完成后需按测试语料完成测试语音文件的切分，进入测试系统前需进行语音处理，保证电平符合要求。

j) 背景噪声的录制在真车内进行,使用一个放置在靠近车载免提麦克风处的测量传声器来录制背景噪声，如果条件允许，也可以使用车载免提麦克风来直接录制。一般由测试实验室（与生产厂家）来共同决定使用背景噪声的类型。环境噪声文件的录制格式为.wav，应保证 44.1kHz 的采样频率和 16bit 的量化间隔，并记录噪声幅值，以便重放。环境噪声应考虑行车环境、车速、前车窗和空调的使用情况。典型的环境噪声的录制场景如表 2 所示。

表2 典型的环境噪声的场景

场景编号	行车环境	车速	车窗	空调	麦克风处的环境噪声声压级（仅供参考，以实际录制结果为准）	备注
1	安静	0km/h	关	关	45-50dB（A）	必选
2	闹市	40-60km/h	关	低档	50-65dB（A）	必选
3	高速	60-120km/h	关	中档	65-75dB（A）	必选
4	闹市	40-60km/h	半开	关	65-75dB（A）	可选
5	高速	60-120km/h	半开	关	70-85dB（A）	可选

6.4 基于语音标准库的测试方法

6.4.1 概况

测试需要在可重复的、模拟真实场景下进行。测试声场景应模拟行车使用环境，在真实汽车车壳或真实车辆内进行测试。推荐使用符合ITU-T P.581规定的HATS进行声音信号的重现与采集，使用前对HATS进行校准和均衡。

将语音标准库中预先录制好的语音输入待测系统，并统计系统输出结果。车载系统的响应可以录像的方式记录下来，作为测试结果之一。

6.4.2 背景噪声重放

使用四个中音扬声器和一个低音扬声器组成的阵列来模拟行车噪声场景，如图2所示。使用扬声器阵列来重放放背景噪声时，应首先经过均衡和校准，使得免提麦克风位置处的声功率谱密度和录音信号一致。均衡既可以使用测量麦克风，也可以使用录制背景噪声时用过的车载免提麦克风。比较录音信号和模拟背景噪声信号，两者的最大A计权声压级偏差应不超过±1dB，在100Hz～10kHz频率范围内的1/3倍频程功率谱密度偏差应不超过±3dB。

为了使得扬声器、免提麦克风和HATS的声传输路径间的干扰最小，应仔细选择扬声器的放置位置。低音扬声器放置在座位后面正中后备箱上面，后排两个扬声器分别放置在后座靠枕与后窗玻璃之间靠近车壳的支架处，前面两个扬声器分别放在仪表面板的上部两边。

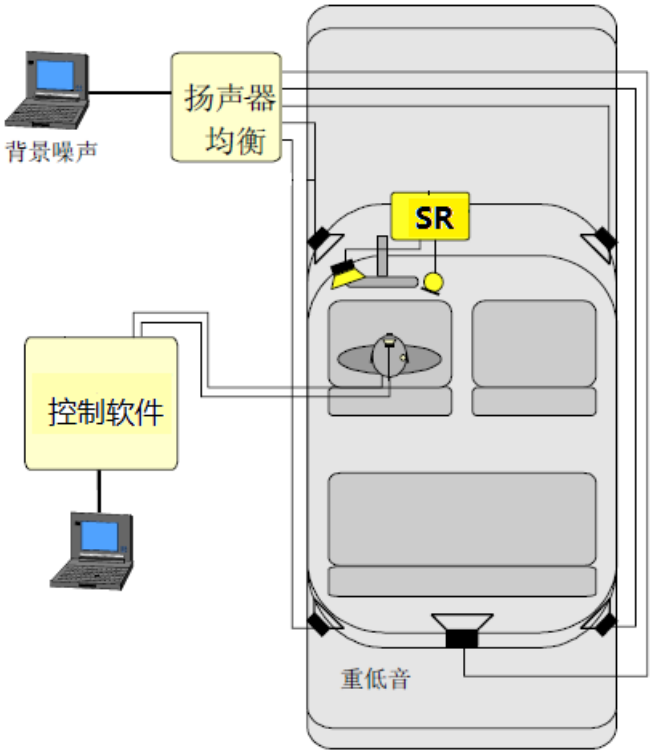

图2　语音识别测量及背景噪声重放设置

6.4.3　车内测试配置

6.4.3.1　HATS 的放置

一般由厂家来指定HATS的放置位置（包括仿真嘴和仿真耳分别相对于麦克风和扬声器的距离）。如没有特别指定，测试时HATS应放置在驾驶员的座位上，该具体位置应和多数人的驾驶习惯位置一致，并且定义仿真嘴到麦克风的距离。测试报告应包含位置信息。

为了保证每次车内测试时HATS的位置相对固定，可以通过在车内进行标注的方法来帮助定位（比如在车内中标出HATS相对于某一位置的距离，如左、右车门和车顶等固定物体。

注1：如有特殊测试要求，也可以放置在副驾驶位和乘客位。

6.4.3.2　仿真嘴

仿真嘴应符合ITU-T P.58的规定，并依据ITU-T P.340在MRP处进行均衡。

在MRP校准的声压级为-4.7dBPa。

对于扬声式车载免提终端，应在HATS-HFRP（HATS 免提参考点）处进行声压校准，使得HATS-HFRP处的平均声压级为-25.7 dBPa，此时MRP处的声压为发送方向源信号大小。以上过程的具体操作步骤见ITU-T P.581的相关部分。

测试行车噪声环境下时，由于“伦巴效应”，仿真嘴的输出电平会增大。

$$I(N)=\begin{cases}0 & for \quad N<50\\ 0.3(N-50) & for \quad 50\le N<77\\ 8.0 \quad for & N\ge 77\end{cases}$$

其中 I=仿真嘴输出电平增大值

N=靠近驾驶员头部位置的长时A记权噪声大小

ITU-T P.340中规定，在免提装置的发送测试中，0.3倍的语音电平增长应单独计算。

6.4.3.3 仿真耳

对于扬声免提终端，HATS左右耳的声信号均被使用。HATS应进行自由场或扩散场均衡，具体可参考ITU-T P.581。

对于头戴免提终端，使用的耳型和佩戴位置见ITU-T P.380。

6.4.3.4 测试信号和电平

测试可以使用提前录制的真人语音信号。

所有的测试信号电平都是指测试信号的激活语音电平（详见ITU-T P.56）。语音识别测试在发送方向使用的是非限带信号。

测试信号的平均大小规定如下：

——发送方向(MRP)：-4.7 dBPa（典型的讲话平均声压级，相当于 HATS-HFRP 处声压级大小为-28.7 dBPa），这一水平适用于耳机的免提终端；

——发送方向（MRP）：-1.7 dBPa 免提扬声器终端（典型的平均说话声压级）（相当于 HATS-HFRP 处声压级大小为-25.7 dBPa)。

注：背景噪声测试中要考虑“伦巴效应”(由于高背景噪声而增加说话者的说话声压级)。

6.5 测试步骤

6.5.1 字准确率

1) 人工头按照 6.4.3.1 配置,背景噪声系统按照 6.4.2 配置并均衡。
2) 测试信号为根据 6.2、6.3 录制的语音信号。由人工嘴产生测试信号的频谱在嘴参考点(MRP)处在自由声场的条件下进行校准。测试信号幅值见 6.4.3.4。
3) 同步播放测试信号和相应背景噪声场景。
4) 记录字准确率。

注2：安静情况下，不需播放背景噪声。

6.5.2 识别成功率

1) 人工头按照 6.4.3.1 配置,背景噪声系统按照 6.4.2 配置并均衡。
2) 测试信号为根据 6.2、6.3 录制的语音信号，即相应命令词语。由人工嘴产生测试信号的频谱在嘴参考点(MRP)处在自由声场的条件下进行校准。测试信号幅值见 6.4.3.4。
3) 同步播放测试信号和相应背景噪声场景。
4) 记录识别结果。

5） 完成所有命令词测试后，统计不同场景下语音识别率。

6.5.3 平均响应时间

1） 人工头按照 6.4.3.1 配置。

2） 测试信号为根据 6.2、6.3 录制的语音信号，即相应命令词语。由人工嘴产生测试信号的频谱在嘴参考点(MRP)处在自由声场的条件下进行校准。测试信号幅值见 6.4.3.4。

3） 播放测试信号，记录语音输入的结束时刻为t_e；车载语音识别系统的开始响应时刻为t_r。

4） 计算响应时间= t_r-t_e。

注3：从标准库中选择 5 段语音，分别进行本地语音识别和在线语音识别。如是在线语音识别，请保持网络良好。

6.5.4 唤醒率

1） 人工头按照 6.4.3.1 配置,背景噪声系统按照 6.4.2 配置并均衡。

2） 测试信号为根据 6.2、6.3 录制的语音信号，即相应唤醒词。由人工嘴产生测试信号的频谱在嘴参考点(MRP)处在自由声场的条件下进行校准。测试信号幅值见 6.4.3.4。

3） 同步播放测试信号和相应背景噪声场景。

4） 记录唤醒测试结果。

5） 重复唤醒测试，统计不同场景下唤醒率。

6.5.5 误唤醒率

1） 人工头按照 6.4.3.1 配置,背景噪声系统按照 6.4.2 配置并均衡。

2） 测试信号为根据 6.2、6.3 录制的语音信号，测试信号不能是唤醒词。由人工嘴产生测试信号的频谱在嘴参考点(MRP)处在自由声场的条件下进行校准。测试信号幅值见 6.4.3.4。

3） 同步播放测试信号和相应背景噪声场景。

4） 记录 T 小时内，车载语音识别系统被唤醒的次数。

5） 分别在不同背景噪声场景下重复测试，统计不同场景下误唤醒率。

6.6 测试报告内容

报告应至少包含下面内容：

1） 被测系统的完整属性描述。

2） 测试数据的语音属性；测试词汇以及测试说话人的选择及确定情况。

3） 语音识别系统输出结果的统计：每个人识别各项指标以及平均识别指标。

4） 测试过程的情况纪录，采用的测试方法及运行过程的流畅性。

5） 被测系统的配置情况。

6） 测试结果记录所使用噪声和声级。

附 录 A
（规范性附录）
标准修订历史

修订时间	修订后版本号	修订内容

附　录　B
（资料性附录）
附录

参 考 文 献