



中华人民共和国国家标准

GB/T 41813.1—2022

信息技术 智能语音交互测试方法 第 1 部分：语音识别

Information technology—Intelligent speech interaction testing method—
Part 1: Speech recognition

2022-10-12 发布

2023-05-01 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 概述	2
5 测试准备和执行	2
5.1 测试数据集	2
5.2 测试工具	3
5.3 测试设备	3
5.4 测试环境	4
5.5 测试执行	4
5.6 测试结果	4
6 功能测试方法	4
6.1 语音信号采集	4
6.2 语音转文字	5
6.3 语音唤醒	5
6.4 前端信号处理	5
6.5 说话人分离	5
6.6 语言信息识别	6
6.7 语音识别后处理	6
7 性能测试方法	6
7.1 语音识别效果	6
7.2 语音识别效率	7
7.3 语音唤醒效果	8
7.4 前端信号处理效果	9
7.5 说话人分离效果	10
7.6 语言信息识别效果	10
7.7 系统稳定性	11
参考文献	12

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。

本文件是 GB/T 41813《信息技术 智能语音交互测试方法》的第 1 部分。GB/T 41813 已经发布了以下部分：

- 第 1 部分：语音识别；
- 第 2 部分：语义理解。

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。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位：中国电子技术标准化研究院、科大讯飞股份有限公司、小米通讯技术有限公司、华为终端有限公司、深圳市优必选科技股份有限公司、中国电信集团有限公司、思必驰科技股份有限公司、中国科学院自动化研究所、中国医学科学院生物医学工程研究所、哈尔滨工业大学、海信视像科技股份有限公司、马上消费金融股份有限公司、腾讯科技(北京)有限公司、沈阳新松机器人自动化股份有限公司、深圳市人马互动科技有限公司、平安科技(深圳)有限公司、安徽咪鼠科技有限公司、涇丰科技(深圳)有限公司、北京捷通华声科技股份有限公司、北京百度网讯科技有限公司、深圳市北科瑞声科技股份有限公司、阿里云计算有限公司、云从科技集团股份有限公司、网易(杭州)网络有限公司、南京云问网络技术有限公司、联想(北京)有限公司、福州数据技术研究院有限公司、国家网络软件产品质量监督检验中心(济南)、中汽研(天津)汽车工程研究院有限公司、华南理工大学、山东省计算中心(国家超级计算济南中心)、中科极限元(杭州)智能科技股份有限公司、神思电子技术股份有限公司、郑州中业科技股份有限公司、中汽数据(天津)有限公司、中国电器科学研究院有限公司、上海计算机软件技术开发中心、北京爱数智慧科技有限公司。

本文件主要起草人：董建、徐洋、吴国纲、马万钟、朱亚军、贾一君、周立君、宋文林、袁杰、杨震、田定书、钱彦旻、陶建华、花云飞、蒲江波、刘斌、李海峰、王峰、杨春勇、苏丹、张锋、冯海洪、刘国涛、任军民、陈楠、刑启洲、魏韬、李笑如、黄石磊、汪淼淼、李军、胡光龙、杨萌、孟宪明、温正棋、鹿飞、方斌、王岳、井焜、李介、张莹、蔡立志、徐向民、高永超、张晴晴。

引 言

智能语音交互在智能家居、智能客服、移动终端、车载终端以及智慧教育、智慧医疗、智能办公、服务机器人等诸多领域应用广泛,已成为当前人机交互的重要方式之一。随着智能语音交互越来越深入到生产生活的方方面面,需要对智能语音交互的系统参考框架、基础技术要求、互联网接口要求等进行统一规范。在这方面,国家已制定了支撑智能语音交互系统的基础性国家标准。在此基础上,也需要用统一的测试方法和评价标准来对智能语音交互系统的能力进行评测,为智能语音交互相关的产品和服务提供评测的基础方法和依据。

GB/T 41813《信息技术 智能语音交互测试方法》为 GB/T 36464(所有部分)《信息技术 智能语音交互系统》提供基础通用的测试方法。智能语音交互包括语音识别、语义理解和语音合成三个基本环节,各环节所涉及的测试对象、测试项目、测试环境和测试方法均有所不同。GB/T 41813《信息技术 智能语音交互测试方法》旨在确立和描述适用于智能语音交互各环节的通用测试项和通用测试方法,拟由三个部分构成。

- 第 1 部分:语音识别。目的在于为智能语音交互应用中的语音识别环节提供通用测试项和通用测试方法。
- 第 2 部分:语义理解。目的在于为智能语音交互应用中的语义理解环节提供通用测试项和通用测试方法。
- 第 3 部分:语音合成。目的在于为智能语音交互应用中的语音合成环节提供通用测试项和通用测试方法。

信息技术 智能语音交互测试方法

第1部分：语音识别

1 范围

本文件描述了智能语音交互测试中语音识别系统的通用测试项和通用测试方法。

本文件适用于智能语音服务提供商、用户和第三方检测机构对智能语音交互应用的语音识别系统测试的设计和实施。

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 21023 中文语音识别系统通用技术规范

GB/T 36464(所有部分) 信息技术 智能语音交互系统

3 术语和定义

GB/T 36464(所有部分)界定的以及下列术语和定义适用于本文件。

3.1

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。

[来源：GB/T 36464.1—2020,3.7]

3.2

说话人分离 speaker diarization

对包含有效语音信号的音频流中的多个说话人进行说话人分割和说话人聚类的过程。

注：说话人分离的目的一般是对空间中存在的多个说话人进行分类和追踪。

3.3

说话人分割 speaker segmentation

在多个说话人中找出说话人改变的时间边界，并根据这些边界将音频流切分成多个语音片段。

3.4

说话人聚类 speaker clustering

将属于同一个说话人的一个或多个语音片段进行归类。

3.5

语音编码 speech coding; speech encoding

语音波形编码 speech waveform coding

按照一组能合理重构语音信号的规划，由经数字化的语音信号到离散的数据元序列的转换。

注：语音数字化可与用于语音压缩的某一编码相结合。因此，“语音编码”这一术语常指这种组合运算。

[来源：GB/T 5271.29—2006,29.01.23]