ICS 35.240.01

DB34

安　　徽　　省　　地　　方　　标　　准

DB 34/ XXXXX—XXXX

儿童语音（识别）测试集技术规范

Children's Speech (recognition) test set Technical specification

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

安徽省市场监督监管局　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
3.1 语音交互 speech interaction 1
3.2 语音识别 speech recognition 1
3.3 语音合成 speech synthesis 1
3.4 命令字识别 Command word recognition 1
3.5 声纹 Voiceprint 1
3.6 语音唤醒 speech wakeup;voice trigger 2
3.7 误唤醒 fake wakeup 2
3.8 语音打断 speech interruption 2
3.9 儿童陪伴机器人 Child companion robot 2
3.10 近场 near field 2
4 测试集要求 2
4.1 测试集内容 2
4.2 测试集构建方法 2
5 测试环境条件 2
5.1 设备要求 2
5.2 测试环境要求 3
6 测试方法 4
6.1 测试指标 4
6.2 测试方法 5

前　言

本标准按照GB/T 1.1-2009给出的规则起草。

本标准由安徽淘云科技有限公司提出。

本标准由安徽省信息技术标准化技术委员会提出并归口。

本标准起草单位：安徽淘云科技有限公司

本标准主要起草人：刘庆升

儿童语音（识别）测试集技术规范

1 范围

本标准规定了儿童陪伴机器人领域语音交互系统的术语、系统框架、能力要求、评价指标要求和测试规程。

本标准适用于儿童智能产品,可包括儿童陪伴机器人、早教故事机、学习平板、点读机等类别产品。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是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 36464.2-2018 信息技术 智能语音交互系统 第2部分：智能家居

GB/T 36464.4-2018 信息技术 智能语音交互系统 第2部分：移动终端

GB/T 21023-2007 中文语音识别系统通用技术规范

GB/T 21024 中文语音合成系统通用技术规范

SJ/T 11380 自动声纹识别（说话人识别）技术规范

3 术语和定义

3.1 语音交互 speech interaction

人类和功能单元之间通过语音进行的信息传递和交流活动

[GB/T 36464.2-2018,定义3.1]

3.2 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

[GB/T 21023-2007,定义 3.1]

3.3 语音合成 speech synthesis

通过机械、电子的方法合成人类语言的过程

[GB/T 21024-2007,定义 3.2]

3.4 命令字识别 Command word recognition

一种基于语音识别语法的语音识别方式，是在语音识别语法规则限定的范围内，对于给定的语音输入，语音识别引擎给出语音识别语法覆盖范围内的文本或拒识作为识别结果

[GB/T 34083-2017,定义 3.3]

3.5 声纹 Voiceprint

对语音中所蕴含的、能表征和标识说话人的语音特征，以及基于这些特征（参数）所建立的语音模型的

总称
[SJ/T 11380-2008,定义 3.1.1]

3.6 语音唤醒 speech wakeup;voice trigger

处于音频流监听状态的语音交互系统，在检测特定的特征或事件出现后，切换到命令字识别、连续语音是被等其他处理状态的过程
[GB/T 36464.2-2018,定义 3.13]

3.7 误唤醒 fake wakeup

语音唤醒过程中出现的，无音频流或者音频流中没有出现唤醒所需的特征或事件时，语音唤醒系统被唤醒的现象
[GB/T 36464.2-2018,定义 3.14]

3.8 语音打断 speech interruption

播放声音过程中，当语音采集设备检测到有效语音输入时，中断播放声音，转到语音识别等其他处理过程
[GB/T 36464.2-2018,定义 3.18]

3.9 儿童陪伴机器人 Child companion robot

自动执行工作的机器装置。它既可以接受人类指挥，又可以运行预先编排的程序，也可以根据以人工智能技术制定的原则纲领行动。作用主要以陪伴为主，第一类是面向幼儿，其中的功能以早教为主；第二类是面向儿童，功能主要以教育、陪伴为主；第三类是面向年龄稍大的儿童，通过机器人教授编程知识。

3.10 近场 near field

拾音设备与声源距离 1m（含）之内

4 测试集要求

4.1 测试集内容

测试集语料应覆盖音频、视频点播；闲聊；百科问答；打开设备的应用等常规的交互场景。

4.2 测试集构建方法

a)句识别率测试应至少男女各20名发音人进行录制,语音唤醒功能测试应至少由50名发音人进行录制，声纹识别测试应至少由50名发音人进行录制，具体要求参照GB/T 21023-2007中7.3执行

b）环境噪音集录制以家居环境实际噪声为主（包括客厅、卧室等环境噪声）

5 测试环境条件

5.1 设备要求

音频采样设备、传声器、回放设备的有关参数应符合表1、表2和表3的要求

表1 音频采样设备要求

设备名称	参数要求
可移动的声卡	支持44.1kHz及以上的采样频率，16bit及以上的模数转换器和数模转换器
录音软件	波形采样范围为±5000smpl~±10000smpl
计算机	应支持录音软件的安装和使用
声压计	可用于环境声压确认

表2 传声器的参数要求

符号	参数	测试条件	最小值	典型值	最大值	单位
S	灵敏度	1 kHz纯音，94 dB SPL	-45	-42	-39	dBV/P_A
SNR	信噪比	1 kHz纯音，94 dB SPL	-	59	-	dB(A)
Z_{out}	输出阻抗	1 kHz纯音，94 dB SPL	-	-	400	Ω
THD+N	总谐波失真	1 kHz纯音，100dB SPL	-	-	1	%
		1 kHz纯音，115dB SPL	-	-	10	%
-	指向性	反向衰减≥15dB，最佳接受范围为母线同咪头在传声器拾音方向中垂线呈60°夹角的圆锥内部	-	-	-	-

表3 回放设备要求

设备名称	参数要求	说明
计算机	支持音频播放软件的安装和使用	
播放器	频率相应（±2.5dB）；74Hz~18kHz 最大声压级：102dB(A)	推荐无人工嘴的条件下使用
功率放大器和人工嘴	信噪比：90Db 增益控制：0dB~25dB 频率响应:200Hz~10kHz 最大声压级：110dB(A)	推荐在测试环境内使用
仿真人体	根据音箱和人工嘴的尺寸和安装位置定制	

5.2 测试环境要求

5.2.1 被测语音交互系统

部署被测语音交互系统,应确保被测系统具有语音拾音功能,可通过对话方式对其进行控制和交互。

5.2.2 被测系统网络环境

针对儿童陪伴机器人领域的语音交互系统，应提供其所需的移动互联网服务，网络条件应满足上行带宽不低于100kbit/s、下行带宽不低于50kbit/s，应保持稳定的连通状态。

5.2.3 远场拾音距离要求

测试所描述远场拾音距离默认为3m

5.2.4 语音测试集

应按4.2要求，在家居环境场景下回放得到的测试语音文件和其对应的语料，作为语音测试集。

5.2.5 测试场景要求

测试场景采用真实家居环境噪声或模拟家居的环境噪声，分为低噪环境和高噪环境，要求噪音频谱保持稳定且噪音与命令词无类似发音，具体见表4

表4 典型的环境噪声的录音场景

场景编号	家居环境	房间门窗	电视(可选)	空调(可选)	传声器处的环境混响要求	信噪比dB	传声器处的环境噪声声压级dB(A)	备注
场景1	低噪	关	关	关	混响时间0.65	15	≤45	必备
场景2	高噪	开	开	开	混响时间0.65	10	45~60	可选

6 测试方法

6.1 测试指标

6.1.1 语音识别

基本要求包括

a) 识别引擎应支持远场音频处理，可支持近场音频处理。支持命令字识别或连续语音识别

b) 在低噪环境（声音强度在 50dB 以下）中，语音识别正确率应大于 85%

c）在高噪环境（声音强度在 50dB~70dB）中，语音识别正确率应大于 80%

6.1.2 语音合成

应支持汉语普通话，宜支持英语以及粤语或其他方言，宜支持多音色合成和个性化合成，主要要求包括：

a) 多音色，应支持青年女声和青年男声

b) 多方言，应支持汉语普通话

c) 混合语种，应支持中英文混读

d) 多语种，应支持英语

e) 平均意见得分，应大于或等于 4.0（满分 5.0）

6.1.3 语义理解

应支持语义抽取、模糊识别、语义排序：

语义抽取：抽取用户的关键意图

语义排序：语义理解结果中给出多个排过顺序的理解结果供用户确认

模糊识别：正确处理用户说的错别字、同义词、多字漏字、发音模糊的问题

6.1.4 交互成功率

控制指令应全面覆盖儿童日常学习、娱乐等日常交互行为的语义意图理解。

低噪环境下，针对童声，交互成功次数与总交互次数比例应大于 90%，高噪环境下，交互成功次数与总交互次数比例应大于 80%

6.1.5 响应时间

响应时间是指输出结果与语音输入结束的时间间隔。平均响应时间应小于 1s

6.1.6 语音唤醒

a）在低噪环境（声音强度在 50dB 以下）中，语音识别正确率应大于 80%，误唤醒频度应小于或等于 0.2 次/h
b）在高噪环境（声音强度在 50dB~70dB）中，语音识别正确率应大于 65%，误唤醒频度应小于或等于 0.1 次/h

6.1.7 声纹识别

应可根据声纹识别结果，实现对不同身份用户的差异化反馈。声纹识别错误率应小于或等于 10%，错误接受率应小于或等于 5%

6.1.8 语音打断

应支持交互过程中的语音打断，实现交互速度与自然度的提高
在语音交互过程中，$P_i=N_i/N*100\%$
式中：
P_i——语音打断成功率
N——交互内容中需要执行打断操作的次数
N_i——被语音交互系统正确响应的次数

6.2 测试方法

6.2.1 语音识别测试

在表 4 测试环境场景下，将智能设备调至待命状态，在远场距离使用回放设备播放语音识别测试语料，当智能设备传声器的语音声压级为 55dB(A)时，记录低噪环境（SNR=15dB）及高噪环境（SNR=10dB）下智能设备的识别结果，并与预期结果进行比对，统计结果并给出句识别率
使用以上测试方法，测试验证是否满足 6.1.1 的要求。

6.2.2 回音消除

使用同一首歌曲，分别用待测设备播放和非待测设备播放，对比两种情况下的唤醒率（信噪比=0dB），若待测设备播放情况下唤醒率高于非待测设备播放，说明回声消除生效。
使用以上测试方法，测试验证是否满足 6.1.6 的要求

6.2.3 语音唤醒测试

语音唤醒测试包括唤醒正确率和误唤醒频度测试：

a)唤醒测试：在表4测试环境场景下，将智能设备调至待命状态，使用回放设备在远场距离播放唤醒测试语料，当智能设备传声器的语音声压级为55dB(A)时，记录低噪环境（SNR=20dB）及高噪环境（SNR=-15dB）下智能家居是否给出正确响应，分别统计低噪环境和高噪环境下智能设备唤醒正确率

b)误唤醒频度测试：在表4测试环境场景下，将智能家居调至待命状态6h，记录低噪环境及高噪环境下智能家居被误唤醒频度

使用以上测试方法，测试验证是否满足 6.1.6 的要求

6.2.4 声纹识别测试

声纹识别测试包括声纹识别错误拒绝率和声纹识别错误接受率测试：

a)声纹识别错误拒绝率：使用智能家居进行语音注册，测试人数50人（男女各25人），注册完毕后，在远场距离下让测试人使用本人注册语句进行声纹验证，共验证50条，统计结果并给出错误拒绝率

b)声纹识别错误接受率：使用声纹识别语料对智能家居进行语音注册，注册人数50人（男女各25人），注册完毕后，在远场距离下使用回放设备播放非本人同性别3人的1句话进行冒认，当智能家居传声器的语音声压级为55dB(A)时，累积150条冒认，统计结果并给出错误接受率

c)具体人员选择要求参照GB/T 21023-2007中的7.3

使用以上测试方法，测试验证是否满足 6.1.7 的要求

6.2.5 语音合成测试

选取10个体验人员，男女各5人，通过对智能家居人为的唤醒或识别命令反馈，测听合成语音同真人语音在音质、可懂度和自然度等方面的差异，并以平均意见得分对主管测评进行量化，记录平均结果。

使用以上测试方法，测试验证是否满足6.1.2的要求

6.2.6 交互成功率

根据以上6.2.1、6.2.2、6.2.3、6.2.4的测试结果对产品的基本交互功能进行统计分析，给出整体交互成功率

使用以上测试方法，测试验证是否满足6.1.2的要求

6.2.7 响应时间

根据以上6.2.1、6.2.2、6.2.3、6.2.4的测试结果对产品的基本功能进行统计分析，给出整体交互成功率

使用以上测试方法，测试验证是否满足6.1.5的要求

6.2.8 语音打断成功率

选取已定制的命令词，在语音交互的过程中，通过回放设备进行播放，记录在1小时内被语音交互系统正确响应及播放的次数，根据6.1.8所给出的公式进行统计分析，给出该命令词的语音打断成功率。