ICS 33.160.99
M72

团 体 标 准

T/CESA 1045-2019

智能音箱技术规范

Technology specification for smart speaker

2019-01-04 发布 2019-04-01 实施

中国电子工业标准化技术协会 发布

目　　次

前　言

本标准按照GB/T 1.1-2009《标准化工作导则 第1部分：标准的结构和编写》给出的规则起草。

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这些专利的责任。

本标准由科大讯飞股份有限公司提出。

本标准起草单位：中国电子技术标准化研究院、中国电子音响行业协会、科大讯飞股份有限公司、北京灵隆科技有限公司、北京小米科技有限责任公司、北京百度网讯科技有限公司、国光股份电器有限公司、北京理工大学、中科院声学所、海德声科贸易（上海）有限公司、国光电器股份有限公司、歌尔股份、深圳市漫步者科技股份有限公司、华为终端（东莞）有限公司、深圳三诺数字科技有限公司、北京声智科技有限公司、北京云知声信息技术有限公司、广州爱浪智能科技有限公司、上海喜日电子科技有限公司、广州番禺巨大汽车音响设备有限公司、广州长嘉电子有限公司、珠海市魅族科技有限公司、深圳摩耳声学科技有限公司、惠州超声音响有限公司、吉林航盛电子有限公司、中科睿微（宁波）电子技术有限公司、广州蓝豹智能科技有限公司、深圳市新峰龙工业有限公司、广州笙达电器有限公司、汉桑（南京）科技有限公司、北京瑞森新谱科技股份有限公司、佛山蓝旗亚数码科技有限公司、安克创新科技股份有限公司、中国华录集团有限公司、腾讯科技（北京）有限公司、南京声准科技有限公司。

本标准主要起草人：田晨燕、董桂官、刘华益、彭泓、马万钟、汤跃忠、赵群、赵立峰、谢守华、王晶、董斌、黄梳、易高雄、张金国、毕静伟、温煜、衣强、王远昌、陈孝良、郭凡、赵志扬、胡科、曾庆法、何艳、陈爱民、叶威志、熊俊、林顺达、郭杭伟、孙海原、周复元、冯明华、韩立成、张晓辉、李沫然、汪剑、车永进、翟尤、宋伟。

引　言

智能音箱系统是包括智能音箱终端、云端、手机应用、以及关联设备和资源的系统。智能音箱是指具有语音交互功能、能够访问网络内容、享受网络服务的音箱设备。智能音箱集成了人工智能处理能力，能够通过语音识别、语音合成、语义理解等技术完成语音交互，成为消费电子领域的热点产品，而现行扬声器（音箱）的国家标准及行业标准未能覆盖智能音箱的技术指标。智能音箱音质层次不齐、语音交互性能良莠不齐、内容合规及信息安全等方面存在较大风险，所以亟需加强端云一体化标准的制定；同时应该加强产品质量监管体系建设，引导智能音箱产业健康发展。

本标准是对现行扬声器（音箱）等国家标准及行业标准的有益补充，同时促进GB/T 12060.5-2011《声系统设备 第5部分：扬声器主要性能测试方法》，GB/T 14277-2013《声频组合设备通用规范》及SJ/T 11540-2015 《有源扬声器通用规范》等现行国家标准及行业标准的贯彻执行。

智能音箱技术规范

1 范围

本标准规定了智能音箱的系统架构、音频性能、语音交互等技术要求，以及智能音箱智能化要求及其测试方法。

本标准适用于智能音箱产品及其系统的研发、设计和测试。

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 3096-2008 声环境质量标准

GB/T 5271.29-2006 信息技术 词汇 第29部分:人工智能 语音识别与合成

GB/T 12060.5-2011 声系统设备 第5部分：扬声器主要性能测试方法

GB/T 12060.16-2017 声系统设备 第16部分：通过语音传输指数客观评价言语可懂度

GB/T 14277-2013 音频组合设备通用规范

GB/T 21023-2007 中文语音识别系统通用技术规范

GB/T 21024-2007 中文语音合成系统通用技术规范

GB/T 34083-2017 中文语音识别互联网服务接口规范

GB/T 34145-2017 中文语音合成互联网服务接口规范

GB/T 35273-2017 信息安全技术 个人信息安全规范

GB/T 35312-2017 中文语音识别终端服务接口规范

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

SJ/T 11380-2008 自动声纹识别（说话人识别）技术规范

SJ/T 11540-2015 有源扬声器通用规范

SJ/T 11688-2017 智能电视智能化技术评价方法

SJ/T 11712-2018 智能电视语音识别 测试方法

SJ/T 11713-2018 智能电视语音识别 通用技术要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

智能音箱 smart speaker

是指具有语音交互功能、能够访问网络内容、享受网络服务的音箱设备。

注1：智能音箱可关联扩展其他设备、实现内容接入。

注2：智能音箱利用云端或者本地的人工智能处理能力，能够通过语音识别、语音合成、自然语言理解等技术完成语音交互。

注3：智能音箱可以对关联设备进行控制。

3.2

智能音箱系统 smart speaker system

是包括智能音箱终端、云端、手机应用、以及关联设备和资源的系统。

3.3

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]

3.4

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]

3.5

语音合成 speech synthesis

将给定的文本转换成与之对应的语音的过程。
[GB/T 34145—2017，定义3.1]

3.6

自然语言理解 natural language understanding

让计算机能够读懂自然语言文本中蕴含的含义及意图的过程。

3.7

语音唤醒 speech wake-up;voice trigger

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

3.8

误唤醒 false wake-up

音箱处于音频流监听状态，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。

[改写GB/T 36464.2-2018，定义3.14]

3.9

噪声 noise

语音采集过程中，采集到的能干扰对目标语音信号的识别、理解或处理的信号。

3.10

声纹 voiceprint

指语音中所蕴含的、能表征和标识特定说话人的独有的特性或特征。
[SJ/T 11380—2008，定义3.1.1]

3.11

声纹识别 voiceprint recognition

根据待识别语音的声纹特征识别该段语音所对应的说话人的过程。
[SJ/T 11380—2008，定义3.1.6]

3.12

麦克风阵列 microphone array

由具有确定空间拓扑结构的多个麦克风组成的，对信号的空间特性进行采样并处理的系统。

3.13

语音打断 speech interruption

语音交互系统在播放声音的过程中，当语音采集设备检测到有效语音输入时，终端播放声音，转到语音识别等其他处理过程。

[GB/T 36464.2-2018，定义3.18]

3.14

隐私标签 privacy label

由厂商或者开放平台应用定义的涉及使用者私密信息的数据，对该类型数据加以标识的标签。

4 系统逻辑结构

智能音箱系统分为输入、处理和输出三个模块，可选择在本地、云端或融合实现，其中：

a) 输入模块包括麦克风阵列、语音采集、语音唤醒和声纹识别，负责将语音输入转化为语音流，作为处理的输入。其中：

1) 麦克风阵列负责对音频信号进行定向采集；
2) 语音采集包括对麦克风阵列拾取到的音频进行降噪、去混响、回声消除等处理；
3) 语音唤醒负责音频流监听，并在检测到特定的特征或事件出现后，切换到语音识别状态；
4) 声纹识别（可选支持）负责对发音人声纹进行获取、分析并输出反馈结果。

b) 处理模块包括语音识别、自然语言理解、业务逻辑。其中：

1) 语音识别负责将语音流转换为人类可识别的文本信息并直接输出，或转换为计算机可识别的文本信息并输出到自然语言理解；
2) 自然语言理解负责对语音识别提供的文本信息做自然语言解析；
3) 业务逻辑负责根据自然语言理解的结果，映射到相应的业务线，并依此向相关应用下达指令并提供反馈信息。

c) 输出模块包含语音合成和资源调用。其中：

1) 语音合成模块负责将业务逻辑反馈的计算机可识别的文本信息转换为语音流的输出；
2) 资源调用负责将业务逻辑反馈的信息与对应的应用资源进行匹配，并对外提供应用与服务的输出；
3) 语音合成和资源调用相互关联对应，共同作为输出结果。

智能音箱系统逻辑结构见图1。

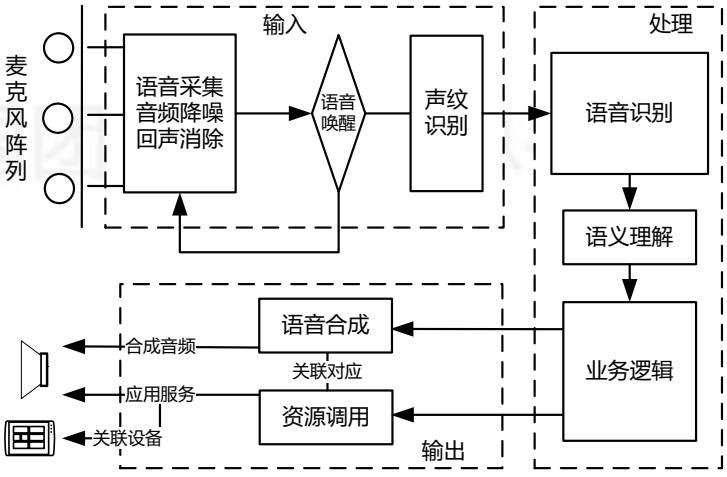

图1 智能音箱系统逻辑结构

5 基本技术要求

5.1 概述

智能音箱应具备语音交互功能；
智能音箱应可关联一个以上的扩展设备；
智能音箱应提供多种基于网络的应用和服务。

5.2 声学性能要求

产品的声性能参数及要求遵照SJ/T 11540-2015 表2的要求。

6 语音交互技术要求

6.1 概述

智能音箱的语音交互技术要求包括语音技术要求及交互技术要求,智能音箱指标级别分为A级和B级两类，具体见7.2。

6.2 语音技术要求

6.2.1 语音采集

智能音箱应能够通过麦克风或麦克风阵列等具备语音采集能力的硬件设备对语音进行采集。

6.2.2 语音识别

语音识别基本要求包括:
a) 识别引擎应支持连续语音识别；
b) 在低噪环境中,语音识别字准确率应≥85%(B级)或90%(A级)；
c) 在高噪环境中,语音识别字准确率应≥80%(B级)或85%(A级)。

6.2.3 声纹识别

智能音箱可具备声纹识别功能，实现对不同身份用户的差异化反馈，如：系统访问权限、系统响应内容等。

6.2.4 语音打断

智能音箱应具备交互过程中的语音打断功能，实现交互速度与自然度的提高。
语音打断成功率的计算方法见公式（1）。

错误!未找到引用源。 .....(1)

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。
在声源距离待测样品1 m距离处语音打断成功率应≥85%(B级)或90%(A级)；

6.2.5 语音合成

应支持汉语普通话,宜支持多音色、混合语种和多语种，宜支持个性化合成，MOS评分应≥3.5（B级）或4.2（A级）(满分5.0)。主要要求包括:
a) 多音色,应支持青年女声和青年男声；
c) 混合语种,应支持中英文混读；
d) 多语种,应支持英语。

6.2.6 语音唤醒

智能音箱应具备唤醒功能（语音唤醒或硬件按键唤醒），为了区分音箱发声与否状态下的语音唤醒功能，本标准中“语音打断”指音箱发声状态下的语音唤醒，“语音唤醒”特指音箱不发声状态下的语音唤醒。不同噪声环境中的语音唤醒能力应满足表1要求。

误唤醒频度应小于等于5次/24小时。

表1　不同噪声环境下的唤醒能力要求

声环境功能区类别	环境噪声等效声级 dB(A)	唤醒正确率 %
1 类	低噪	≥90（B 级）或≥95（A 级）
2 类	高噪	≥85（B 级）或≥90（A 级）

其中，唤醒正确率是指计算方法见公式（2）。

$$P_r = \frac{N_{sw}}{N_w} \times 100\% \quad \text{.....(2)}$$

式中：

P_r**错误!未找到引用源。**——唤醒正确率；

N_{sw}——正确唤醒次数；

N_w——总唤醒次数。

6.3 交互技术要求

6.3.1 交互方式

智能音箱宜实现多轮语音交互，即通过两轮及两轮以上的对话完成一个任务。

6.3.2 交互拒识率

交互拒识率是指在智能音箱在语音交互过程中，交互目的不能够在既定交互轮次内完成，被判定为交互失败的测试比率，其计算方式如公式（3）所示。

错误!未找到引用源。错误!未找到引用源。.....(3)

式中：

P_f——交互拒识率；

S ——交互成功的次数；

F ——交互失败的次数。

智能音箱在低噪环境等级下，拒识率应小于B级15%或A级10%。

6.3.3 响应时间

响应时间包含响应时间及实时系数，实时系数衡量指标遵照GB/T 21023-2007的5.3。

响应时间是指智能音箱在语音交互过程中，不同网络环境下的从语音输入结束时刻起到输出结果开始时刻之间的时间，其计算方法如公式（4）。

$$T_{ack} = t_r - t_e \quad \text{.....(4)}$$

式中：

T_{ack}——响应时间；

t_r——给出结果时刻；

t_e——语音输入结束的时刻。

注：如语音交互系统支持识别结果分多次返回，t_e应为第一部分识别结果返回的时刻。

最大响应时间是指智能音箱在语音交互过程中，厂商规定的在不同网络环境下的从语音输入结束时刻起到输出结果开始时刻之间的时间，最大响应时间应符合厂商的产品质量要求。

智能音箱宜支持不同类型网络接入方式；在典型应用场景对话交谈中，其平均响应时间应满足B级≤2.50 s或A级≤2.00 s。

智能音箱在不同网络接入方式中，网络条件应满足上行带宽不低于100 kbit/s、下行带宽不低于50 kbit/s,应保持稳定的连通状态。

6.3.4 休眠要求

智能音箱上应具备休眠键,并且明确提示用户音箱是否处于休眠状态。在休眠状态下，音箱应停止拾音。

6.4 安全性要求

6.4.1 概述

智能音箱在语音交互过程中涉及的信息安全和隐私应符合GB/T 35273-2017的要求。

6.4.2 数据加密

在局域网和互联网中，同服务端和手机客户端，进行数据交互，应使用加密算法。

6.4.3 网络端口

除正常功能需要的网络端口之外，应关闭其他服务端口。

6.4.4 设备调试

量产版本的音箱，应关闭调试模式。

6.4.5 隐私标签

支持开放平台应用时，应支持隐私标签。

7 智能化及音质性能程度等级

7.1 概述

根据智能音箱智能化程度及音质性能的不同，将智能音箱分为I、II、III三个不同的程度等级。

7.2 智能化及音质性能程度等级

智能音箱指标级别及其要求应符合表2的要求，智能音箱智能化程度等级分类如表3所示。

表2　指标级别分类表

指标级别	语音识别字准确率 %		语音打断成功率 %	语音合成MOS评分	语音唤醒正确率 %		交互拒识率 %	平均响应时间 s
A级	高噪	≥85	≥90	≥4.2	高噪	90	小于10	≤2.0
	低噪	≥90			低噪	95		
B级	高噪	≥80	≥85	≥3.5	高噪	85	小于15	≤2.5
	低噪	≥85			低噪	90		

表3　智能化程度等级分类表

级别分类	智能化及音质功能及性能要求
III级	1）具备II级所有功能； 2）满足100%A级别指标要求； 3）扬声器性能达到GB/T 14277-2013的A类及以上。

表 3　智能化程度等级分类表 （续）

II级	1）具备I级所有功能； 2）满足100%B级别指标要求； 3）具备远场识别功能（不小于5 m）； 4）具备降噪功能； 5）支持智能家居控制协议； 6）具备其他更智能化功能及性能； 7）扬声器性能达到GB/T 14277-2013的B类及以上。
I级	1）具备语音交互功能，满足50%B级别指标要求； 2）可连接云端音频媒体库； 3）至少具备Wi-Fi无线传输功能 。

8 测试方法

8.1 测试准备

8.1.1 测试语料要求

测试语料应覆盖被测系统的核心词汇，并从被测系统词汇量覆盖、业务覆盖、音节覆盖，以及常用性角度进行设计，具体要求应按GB/T 21023-2007执行。

8.1.2 语音测试集要求

语音测试集应符合以下要求：

a）语音识别准确率测试应至少由男女各 20 名发音人进行录制，语音唤醒功能测试应至少由 50 名发音人录制，具体要求应按 GB/T 21023-2007 执行；

b）声纹识别测试应至少由 50 名发音人录制验证，具体要求应按 GB/T 21023-2007 执行。

8.1.3 环境噪声要求

表4　语音识别测试环境要求见下表

家居环境	房间门窗	电视（可选）	抽油烟机（可选）	空调（可选）	待测音箱位置处的环境混响要求 s	信噪比 dB	待测音箱位置处的环境噪声声压级 dB(A)	备注
低噪	关	关	关	关	混响时间 0.2～0.3	≥20	≤45	必选
高噪	开	开	开	开	混响时间 0.2～0.3	≥10	55	必选

8.1.4 测试设备要求

测试设备要求如下：

a）　声音重放设备：由信号发生器、功率放大器和扬声器组成。应满足以下条件：

——功率放大器和扬声器产生的声源幅度非线性影响值应足够小；

——声音重放设备产生的本底噪声应足够小。

b）　声压测试设备：声级计。

c）　识别时间测试设备：宜采用示波器或高速相机测试识别时间，或者开发自动化软件进行测试。

8.1.5 拾音距离要求

测试所描述的拾音距离为通常为1 m、3 m和5 m，使用的测试距离应在测试报告中说明。

8.1.9 测试布置

推荐按照图2布置测试，推荐在A=90° 且B=150° 、A=60° 且B=150° 下测试，使用的空间布置应在报告中说明。

当声源与待测样品的空间布置（包括但不限于角度、高度、摆放位置等）对测试结果有影响时，应改变空间布置重复测试，并提供不同布置下的测试结果。

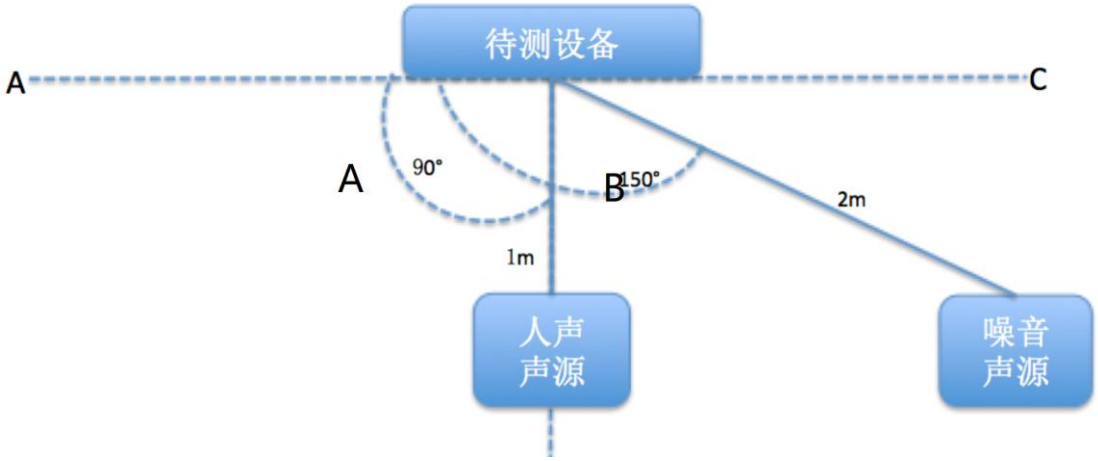

图2　测试布置图

8.2 基本功能和性能测试

智能音箱的声学性能参照SJ/T 11540-2015、GB/T 12060.5-2011、GB/T 14277-2013测试。

8.3 语音交互测试

8.3.1 语音识别

分别在低噪、高噪测试场景下，如下设置播测音源的音量：距离待测设备1 m处，播放唤醒语料或识别语料，在待测设备麦克风处测得平均声压级为65 dB(A)，以此为基准音量。

将智能音箱被测系统调至待命状态，在拾音距离内使用回放设备播放语音识别测试语料，记录低噪、高噪环境下智能音箱被测系统的识别结果，并与预期结果进行比对，统计结果并按照6.2.2条要求计算字准确率。

有必要时可以在其他信噪比下执行测试，并在报告中说明具体的测试安排。

8.3.2 语音合成

选取20个体验人员，男女各10人，通过对智能音箱被测系统语音唤醒或语音识别命令的反馈，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（MOS分）量化进行主观测评，记录平均结果。

使用以上测试方法，测试验证是否满足6.2.5的要求。

8.3.3 交互拒识率

交互拒识率测试方法如下：

a) 将智能音箱被测系统调至待命状态，使用回放设备在拾音距离内播放语音识别测试语料，记录当次语音交互会话是否成功和有效；
b) 分别在低噪、高噪条件下按上述步骤完成测试，计算各测试场景下的语音交互拒识率。

8.3.4 响应时间

响应时间测试方法如下：

a) 准备智能音箱测试设备及其网络环境，开启被测系统拾音功能，用回放设备在拾音距离内播放语音识别测试语料，记录当次成功的语音交互会话测试录音输入完成的时刻 t_e和返回服务结果的时刻 t_r，计算当次语音交互会话的响应时间；

b) 分别在表 2 所示网络环境下，按上述步骤完成测试，然后计算平均识别时间、平均实时系数和最大响应时间。

实时系数测试遵照GB/T 21023中5.3要求。

最大响应时间测试遵照厂商产品质量要求。

8.3.5 语音唤醒

语音唤醒测试包括唤醒正确率测试和误唤醒频度测试，方法如下：

a) 唤醒测试：分别在低噪、高噪测试场景下，将智能音箱被测系统调至待命状态，使用回放设备播放唤醒测试语料，记录被测系统是否给出正确响应，统计各场景下的唤醒正确率，其计算方法见公式（2）；
b) 误唤醒测试：将智能音箱被测系统调至待命状态，测试 24 h，记录被测系统被误唤醒次数，统计误唤醒频度。

8.3.6 语音打断

设定人在距离待测设备1m处发声，统一在待测设备的麦克风处测量得到人声音量70 dB，待测设备的内噪声85 dB，信回比-15 dB，并以此为音量基准在不同距离重复进行语音打断测试。

使用定制的语音打断的唤醒词或命令词集，在待测设备播放状态或者语音交互状态中，使用统一的内噪声素材，进行语音打断，按公式（1）计算语音打断成功率。

8.3.7 声纹识别

根据产品使用说明设置声纹并验证其能正常使用。

8.3.9 安全性测试

8.3.9.1 数据加密

通过加密有效性测试,判断被测系统是否正确使用了加密技术。

使用一台安装无线网卡的电脑，开启无线热点，启动抓包工具（如Wireshark），将音箱连接到这个无线网络中。在联网状态下，正常使用音箱，分析抓包工具生成的数据，判断数据是否已加密。

8.3.9.2 网络端口

将音箱联网，在网络设备上，查询音箱的IP地址，使用一台电脑，接入局域网，启动端口扫描工具（如Nmap）对音箱的IP地址进行扫描，根据扫描结果，判断常见服务端口是否已关闭。

8.3.9.3 设备调试

检查音箱端口，将端口同电脑USB端口连接，检查电脑是否弹出发现设备等提示，使用调试工具（如adb）测试向音箱发送调试指令，判断调试模式是否已关闭。

附录 A
（资料性附录）
语音测试集构建

A.1 输入输出要求

智能音箱在语音交互过程中的输入应满足以下要求：

a) 应支持中文普通话输入，宜支持英语；
b) 可处理语音输入为（180 ～300）字/分的语速，单次语音输入时长不应超过 30 s，特殊情况下不应超过 60 s；
c) 发音单元的持续时间应不小于 0.2 s，发音单元间的间隔不超过 2 s；若停顿时间超过 2 s，则认为一次语音输入结束。

A.2 测试集构建方法

从噪声、回声、人声，空间、待测设备这几个维度组合构建语音唤醒测试集和语音误唤醒测试集，尽量覆盖各种声学场景，模拟用户真实使用环境。

语音唤醒测试集通过专业录音麦克风在安静环境下组织录制人员录制待测设备的唤醒词。参与录制的人员，需考虑性别、口音、年龄等维度。

误唤醒测试集的构成，主要考虑实际应用场景中引起待测设备误唤醒的噪声来源。例如，家居环境下音箱的误唤醒主要来源于电视、人声谈话等，所以此时选择的误唤醒语料，每24小时包含6小时电视节目，6小时新闻节目，6小时人声对话（可选择谈话节目模拟），6小时音乐播放。

表 A.1 测试集构建方法示例

噪声	噪声来源	平稳噪声（家居环境噪声等）
		非平稳噪声（电视噪声等）
		交通工具
		自然声音
		其他
	噪声类型	点声源干扰
		散射噪声
	到待测设备距离	0.3 m
		1 m
		3 m
		5 m
	与待测设备角度	0°、45°、90°、180°、其他
	信噪比	原始
		(-5～ 15) dB ,步长5 dB
回声	内容类型	音乐、有声节目、听声音Skill、TTS等
	信回比	原始
		(-35 ～ 0) dB，步长5 dB

表 A.1　测试集构建方法示例（续）

待测空间	空间类型	马路
		家居
		办公
	待测空间混响（500Hz）	T60 =（300±30）ms
		T60 =（500±30）ms
		T60 =（800±30）ms
待测设备	设备类型	表明被测设备类型，如小米AI音箱
	位置	一面靠墙 < 0.1 m，三面开阔 > 1 m
		两面靠墙均 < 0.1 m,两面开阔 > 1 m
		一面离墙 0.4 m，三面开阔 >1 m
		两面离墙 0.4 m，两面开阔 > 1 m
		四面离墙均 >1 m
	高低	离地 0.4 m
		离地 0.4 m
	设备音量	例如，AI音箱0 dB，30 dB，50 dB，90 dB，100 dB
	设备编号	从1-10
目标声源	性别	男
		女
		儿童
	口音	普通话
		地区性方言
	语速	正常（0.85～1.5）s
		较快（0.65～0.85）s
	与待测设备距离	1 m
		3 m
		5 m
	与待测设备角度	0°、45、°90°、180°、其他
	发声位置	站姿：嘴离地面约 （1.5～1.62）m
		坐姿：嘴离地面约 0.8 m
		躺姿：嘴离地面约 0.4 m
	语料内容	唤醒词
		提问句
注：以上主要考虑家居和办公场景		

在线测试时，信噪比/信回比通过改变噪声源和纯语音段音量以及待测设备音量和纯语音段音量来获得。

离线测试语料中，按信噪比/信回比合成测试语料的方法：采用段信噪比计算方法，即纯语音段能量与混合时间段内的噪声/回声能量对比；实际语料合成时，整段噪声/回声语料设置同一增益来获得目标信噪比/信回比，但一段噪声/回声语料中混合多段唤醒词的时候，由于噪声、回声的能量实时在变化，每段唤醒词的信噪比/信回比不可能完全相同，应允许±1dB的误差。

音量设置需根据被测设备的音量范围和实现机制做定制化设计。

A.3 测试场景设置

测试所描述的场景应满足以下条件：

——环境：温度(23～26) ℃，相对湿度(25～75)%，大气压(95～101.3) kPa；

——高度：(1.1 ± 0.01) m；

——半径：距离被测中心(1.5 ± 0.02) m；

——角度：45°。

表 A.2 测试的音量、距离、角度、噪音类型设置

测试环境	播放语料的音箱		播放噪音的音箱		音量	
	角度	距离	角度	距离	人声（1m 基准）	噪音
安静	60°	1 m	—	—	70 dB(A)	—
	90°	1 m	—	—	70 dB(A)	—
	90°	3 m	—	—	70 dB(A)	—
	90°	5 m	—	—	70 dB(A)	—
电视噪音	60°	1 m	150°	2 m	70 dB(A)	60 dB(A)
	90°	1 m	150°	2 m	70 dB(A)	60 dB(A)
	90°	3 m	150°	2 m	70 dB(A)	60 dB(A)
	90°	5 m	150°	2 m	70 dB(A)	60 dB(A)
家庭聊天噪音	60°	1 m	150°	2 m	70 dB(A)	60 dB(A)
	90°	1 m	150°	2 m	70 dB(A)	60 dB(A)
	90°	3 m	150°	2 m	70 dB(A)	60 dB(A)
	90°	5 m	150°	2 m	70 dB(A)	60 dB(A)