成都启英泰伦科技有限公司标准

Q/QYTL 001-2018

本地语音模组识别效果及性能测试标准

The Recognition Effect and Function Performance Test Specification for Local Voice Module of Speech Recognition

Chipintelli Confidential

2018-08-27 发布 2018-08-27 实施

成都启英泰伦科技有限公司发布

目 录

前 言.....1
1 范围.....2
2 规范性引用文件.....2
3 术语和定义.....2
3.1 人工智能 artificial intelligence.....2
3.2 语音识别 speech recognition.....2
3.3 语音指令 voice command.....3
3.4 人工嘴 artificial mouth.....3
3.5 识别率 recognition rate.....3
3.6 误识别次数 mis-recognition times.....3
3.7 误唤醒 false wakeup.....3
3.8 信噪比 signal-noise ratio.....3
3.9 家居环境 house environment.....3
3.10 车载环境 vehicle environment.....3
3.11 安静环境 quiet environment.....3
3.12 中度噪声环境 moderate noise environment.....4
3.13 强噪声环境 strong noise environment.....4
3.14 语音模组机械噪声 machinery noise.....4
3.15 嘈杂背景噪声 background noise.....4
3.16 回声噪声 echo noise.....4
3.17 混响噪声 reverberation noise.....4
3.18 环境噪声 environmental noise.....4
3.19 测试指令集 test audio data.....4
3.20 噪声集 noise audio data.....5
3.21 运转 in operation.....5
3.22 非运转 not in operation.....5
3.23 播报 broadcasting.....5
3.24 非播报 not broadcasting.....5
3.25 唤醒词集 wake-up words.....5
3.26 指令词集 command words.....5
3.27 单麦/多麦 single microphone/ microphones.....5
3.28 语音识别设备 speech recognition equipment.....5

ChipIntelli Confidential

4 测试说明.....5
4.1 语音识别测试项目、内容.....5
4.1.1 识别率/唤醒率测试.....5
4.1.2 误唤醒测试.....6
4.1.3 响应时间测试.....6
4.1.4 稳定性测试.....6
4.2 语音识别测试说明.....6
4.2.1 语音识别测试环境说明.....6
4.2.2 测试语音模组麦克风说明.....7
4.2.2 测试语言要求.....7
4.2.3 测试语速要求.....7
4.3 语音识别测试技术要求、指标.....7
4.3.1 识别率/唤醒率测试指标.....7
4.3.2 误唤醒测试指标.....9
4.3.3 响应时间测试指标.....9
4.3.4 稳定性测试指标.....9
4.4 语音识别测试音频集采集及音频标准化.....9
4.4.1 唤醒词音频集.....9
4.4.2 指令词音频集.....9
4.4.3 测试音频集标准化.....10
4.5 语音识别测试设备.....10
4.5.1 人工嘴.....10
4.5.2 精密噪音计.....11
4.5.3 噪声源监听音箱.....11
4.6 语音识别测试环境.....12
5 语音识别测试方法及步骤.....12
5.1 识别率/唤醒率测试方法与步骤.....12
5.3 误唤醒测试方法与步骤.....13
5.4 响应时间测试方法与步骤.....13
5.5 稳定性测试方法与步骤.....13
6 特别说明.....14
附录 A 部分噪声集.....15

ChipIntelli Confidential

前 言

由于目前国家或行业没有针对人工智能本地语音识别模组产品语音识别及性能的测试及验收标准。一方面为保证人工智能本地语音识别模组产品的质量，需要对其测试，另一方面为统一产品的开发方和产品用户方对人工智能本地语音识别模组产品的测试活动，并便于实现测试结果的相互认可和可重复性，特制定本标准。本标准是针对本地语音模组识别效果及性能测试提供规范。本标准可以作为人工智能本地语音识别模组识别效果及性能测试、验收的依据。

本标准的编写格式遵循 GB/T 1.1-2009《标准化工作导则 第 1 部分：标准的结构和编写》。

本标准由成都启英泰伦科技有限公司提出。

本标准于 2018 年首次发布。

Chipintelli Confidential

本地语音模组识别效果及性能测试标准

1 范围

本标准规定了本地语音模块组识别效果及性能测试的术语、定义、测试相关说明（包括测试技术要求、测试指标、测试项目、测试内容、测试设备，测试环境）、测试方法、步骤、以及测试结果报告及可追溯性。

本标准适用于所有人工智能本地语音模组的识别效果测试。

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，仅所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 1.1-2009 标准化工作导则 第 1 部分：标准的结构和编写

GB/T 5271.28-2001 信息技术 词汇 第 28 部分：人工智能 基本概念与专家系统

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 5271.31-2006 信息技术 词汇 第 31 部分：人工智能 机器学习

GB/T 5271.34-2006 信息技术 词汇 第 34 部分：人工智能 神经网络

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

GB50096-2011 住宅设计规范

3 术语和定义

下列术语和定义适用于本标准。

3.1 人工智能 artificial intelligence

AI（缩略语），一门交叉学科，通常视为计算机科学的分支，研究表现出与人类智能（如推理和学习）相关的各种功能的模型和系统。

3.2 语音识别 speech recognition

自动语音识别 automatic speech recognition

ChipIntelli Confidential

ASR（缩略语）

利用功能单元进行的，从语音信号到语音内容的某一标识的转换。

3.3 语音指令 voice command

语音模组可识别的声音指令。

3.4 人工嘴 artificial mouth

人工嘴或称仿真嘴：高保真放音设备，播放语音指令，代替人工发声，作为标准测试声源。

3.5 识别率 recognition rate

播报语音指令对语音模组进行测试后，正确识别的指令数占总指令数的百分比。

3.6 误识别次数 mis-recognition times

在模拟语音模组实际使用的生活环境里，一段时间内语音模组发生误识别的次数。

3.7 误唤醒 false 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。

3.8 信噪比 signal-noise ratio

SNR 或 S/N（缩略语）

语音指令的功率与环境噪声功率的比值，单位是分贝。

3.9 家居环境 house environment

语音模组所处工作环境为家居，包含卧室环境、客厅环境、厨房环境、卫浴环境、阳台环境等。

3.10 车载环境 vehicle environment

语音模组所处工作环境为车内空间，包含车辆行驶和熄火状态、开窗及关闭状态等。

3.11 安静环境 quiet environment

语音模组所处工作环境噪声强度介于 25dB-45dB，定义为安静环境。

ChipIntelli Confidential

3.12 中度噪声环境 moderate noise environment

语音模组所处工作环境噪声强度介于 45dB-60dB，定义为中度噪声环境。

3.13 强噪声环境 strong noise environment

语音模组所处工作环境噪声强度介于 60dB-80dB，定义为强噪声环境。

3.14 语音模组机械噪声 machinery noise

语音模组机械噪声，指的是由于语音模组机械（也包含语音识别设备系统中集成的机械部件）运转时，部件间的摩擦力、撞击力或非平衡力，使机械部件和壳体产生振动而辐射噪声。机械噪声按声源的不同可分为 3 类：空气动力性噪声、机械性噪声、电磁性噪声。

3.15 嘈杂背景噪声 background noise

嘈杂背景噪声，指的是背景人声或类人声（如会场、卖场环境下的嘈杂人声）或语音模组之外的其他音响设备所播放的干扰声音，如播放音乐、新闻、电视、电影发出的声音。

3.16 回声噪声 echo noise

回声噪声，指的是语音模组通过自带喇叭播放的声音，对语音识别结果形成干扰。

3.17 混响噪声 reverberation noise

目标说话人的声音经光滑表面（如墙面或物体表面）反射后被语音模组接收的声音。

3.18 环境噪声 environmental noise

语音模组所处的环境包含的背景噪声及混响，其中背景噪声往往包含一个或多个噪声源。如厨房环境同时存在油烟机、炒菜等声音；卫浴环境同时存在浴霸风噪、淋浴水声及光滑墙面反射的人声混响等声音；客厅环境同时存在人声、电视等声音；阳台环境同时存在风噪、室外噪声（如车辆喇叭人声等）；车载环境同时存在发动机噪声、路噪等。

3.19 测试指令集 test audio data

用于语音测试的非训练集音频指令集。

ChipIntelli Confidential

3.20 噪声集 noise audio data

用于语音测试的噪声音频集。

3.21 运转 in operation

语音模组处于功能工作中。

3.22 非运转 not in operation

语音模组没有处于功能工作中。

3.23 播报 broadcasting

语音模组处于自身语音播报中。

3.24 非播报 not broadcasting

语音模组没有在进行语音播报。

3.25 唤醒词集 wake-up words

包含唤醒词以及无需唤醒就能直接控制的指令词的语料集。

3.26 指令词集 command words

包含唤醒词和其他所有指令词的语料集。

3.27 单麦/多麦 single microphone/ microphones

单麦：语音模组采用单个麦克风采集语音数据。

多麦：语音模组采用多个麦克风（两个及以上）采集多路语音数据。按麦克风数量可分为双麦阵列、四麦阵列、六麦阵列、八麦阵列等等；按麦克风排列形式可分为线性麦克风阵列、环形麦克风阵列。

3.28 语音识别设备 speech recognition equipment

集成语音识别模组的电器设备。

4 测试说明

4.1 语音识别测试项目、内容

4.1.1 识别率/唤醒率测试

测试语音指令在安静和噪声环境的识别率。

测试唤醒词在安静和噪声环境的识别率。

ChipIntelli Confidential

4.1.2 误唤醒测试

测试语音模组在安静和噪声环境被非唤醒词（不能包含与唤醒词发音相同或难以区分的语音）唤醒的次数。

4.1.3 响应时间测试

测试语音模组在安静和噪声环境下从接收语音指令结束到给出正确识别结果的时间。

4.1.4 稳定性测试

测试语音模组的语音识别稳定性。

4.2 语音识别测试说明

4.2.1 语音识别测试环境说明

语音识别测试环境需能模拟语音识别设备常规应用时所处的真实环境及工况。通常家居情况下，还包含设备不工作时的安静环境。因此，所使用的测试集为语音模组的指令语料集，所使用的噪声集则为对应的环境噪声及设备工作时的机械噪声。如厨房电器，使用厨房噪声，卫浴电器，使用卫浴噪声。

如语音识别设备为音响设备或需要长时播报语音时，还需要测试设备在进行播放过程中的识别情况。

表 1 语音识别测试环境说明

环境	应用场景	环境噪声（dB）	混响（s）	最小距离（m）	最大距离（m）	应用场景参考面积（m2）	适用语音设备
安静环境	不限	35-45	0.45-0.55	1	5	15-35	所有语音识别设备
工况环境	厨房	55-60	0.65-0.75	1	2	5-10	厨电语音识别设备（如微波炉，抽油烟机，电饭煲等）
	卫生间	55-60	0.65-0.75	1	2	5-10	卫浴语音识别设备（如浴霸，风暖，马桶等）
	阳台	55-60	NA	1	2	5-10	阳台语音识别设备（如洗衣机，晾衣机，阳台灯等）
	起居室（厅）	55-60	0.45-0.55	1	5	15-35	客厅语音识别设备（如空调，中控，遥控器，茶具，制氧机，客厅灯，电视等）
	卧室	55-60	0.45-0.55	1	5	10-20	卧室语音识别设备（如空调，遥控器，台灯，电视等）
	强噪声	65-75	NA	0.5	2	5-10	强噪声设备或环境（如烟机高风档工作时）

注：表 1 中应用场景参考面积符合《GB50096-2011 住宅设计规范》中“5 套内空间”对厨房、卫生间、阳台、起居室（厅）以及卧室套内空间面积的规范。

ChipIntelli Confidential

4.2.2 测试语音模组麦克风说明

语音模组采集语音，可采用单麦克风及麦克风阵列的方式（麦克风阵列的分布具有一定的几何尺寸和结构，如圆形阵列，线性阵列等）。图 1、图 2、图 3 分别为单麦、双麦及四麦的结构示意图。

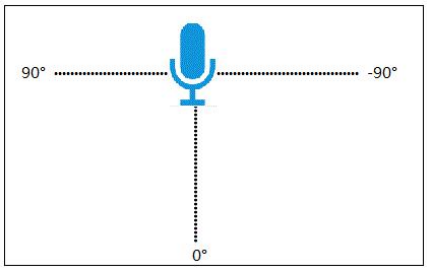


图 1 单麦角度示意图

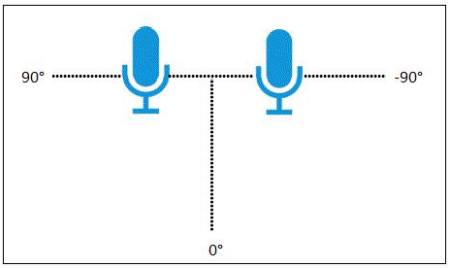


图 2 双麦角度示意图

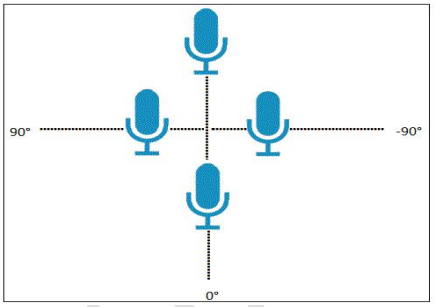


图 3 四麦角度示意图

4.2.2 测试语言要求

测试语音指令使用标准官方语言，中文要求标准普通话二级乙等及以上。

4.2.3 测试语速要求

正常说话语速，中文普通话要求 170-240 字/分钟。

4.3 语音识别测试技术要求、指标

4.3.1 识别率/唤醒率测试指标

ChipIntelli Confidential

表 2 识别率/唤醒率测试指标

测试项	环境	语音设备工作状态	信噪情况	噪声集	测试集	指标	适用性说明
识别率/唤醒率测试	安静环境	非运转/非播报	人声：60dB~70dB 噪声：35dB~45dB	NA	唤醒词集 指令词集	最小距离[1]：≥97% 最大距离[2]：≥95%	适用于所有语音识别设备
	工况环境	运转	人声：65dB~75dB 噪声[4]：55~60dB	环境噪声[3]+语音设备机械噪声	唤醒词集 指令词集	最小距离：≥92% 最大距离：≥85%	适用于能产生机械噪声的语音识别设备
		非运转/非播报	人声：65dB~75dB 噪声：55~60dB	环境噪声	唤醒词集 指令词集	最小距离：≥92% 最大距离：≥88%	适用于工况环境噪声为中度以上的语音识别设备
		播报	人声：65dB~75dB 噪声：55~60dB	环境噪声+回声噪声	唤醒词集 指令词集	最小距离：≥92% 最大距离：≥85%	适用于长播报及音频播放的语音识别设备
		运转（强噪声）	人声：65dB~75dB 噪声：65~75dB	环境噪声+语音设备机械噪声	唤醒词集 指令词集	最小距离：≥90% 最大距离：≥75%	适用于强噪声设备或环境（如烟机高风档工作时）

[1] 最小距离，根据“环境”，“应用场景”参考“表 1"确定具体距离。

[2] 最大距离，根据“环境”，“应用场景”参考"表 1"确定具体距离。

[3] 参考本标准 3.18 项。

[4] 产生强机械噪声的语音设备（如烟机），噪声将达到 70+/-5dB。

ChipIntelli Confidential

4.3.2 误唤醒测试指标

表 3 误唤醒测试指标

测试项	噪声集	指标	误唤醒噪声集说明
误唤醒测试	误唤醒噪声集	<= 3 次/24H	1、误唤醒噪声集：为 24 小时时长噪声语料集，包括：4 小时的电视噪声集（带人声）+4 小时的音乐（纯音乐或歌曲）+8 小时的环境噪声集（设备所处环境）+8 小时的安静环境； 2、误唤醒噪声集不出现唤醒词语音，噪声分贝为 55~60dB。

4.3.3 响应时间测试指标

响应时间：人工嘴近距离（<50cm）播放完语音指令开始到语音识别模组将识别到的指令推送到设备控制或通信端口的时间间隔。响应时间<1.0s。

4.3.4 稳定性测试指标

语音识别模组在环境噪声下，分唤醒和非唤醒状态下的识别稳定性测试。

唤醒状态下识别稳定性测试：每隔 1 秒播放一次唤醒词，运行 72 小时，无死机无重启现象，能正常识别。

非唤醒状态下识别稳定性测试：每隔 T_wakeup_time 秒播放一次唤醒词，运行 72 小时，无死机无重启现象，能正常识别。

T_wakeup_time 等于唤醒后到退出唤醒状态的时间加 1 秒。

4.4 语音识别测试音频集采集及音频标准化

4.4.1 唤醒词音频集

唤醒词集包含唤醒词以及无需唤醒就能直接控制的指令词的语料集。5 男 5 女共 10 成人次对该唤醒词集进行朗读，并采用高保真录音设备进行录音。语音采样率为 44.1KHz，环境噪声<30dB，混响<0.3s，说话人距离麦克风 20-30cm，词与词之间间隔 2 至 3 秒，使用标准官方语言进行朗读；中文标准普通话要求在二级乙等及以上，命令词的朗读速度为 170-240 字/分。

4.4.2 指令词音频集

包含唤醒词和其他所有指令词的语料集。5 男 5 女共 10 成人次对该唤醒词

ChipIntelli Confidential

集进行朗读，并采用高保真录音设备进行录音。语音采样率为 44.1KHz，环境噪声<30dB，混响<0.3s，说话人距离麦克风 20-30cm，词与词之间间隔 2 至 3 秒，使用标准官方语言进行朗读；中文标准普通话要求在二级乙等及以上，命令词的朗读速度为 170-240 字/分。

4.4.3 测试音频集标准化

测试音频集录制完成后，对所有音频进行音量标准化处理。

4.5 语音识别测试设备

语音识别测试中用到的设备、设备型号如下表所示（供参考），这里给出主要设备的参数。

表 4 语音识别测试设备

序号	类别	设备	设备型号	设备品牌	作用
01	电脑	台式机/笔记本	不限	不限	监测语音模块反馈是否准确输出测试结果
02	声音源	人工嘴	4227-A	Brüel & Kjær	播放音频信号
03	噪声监测	精密噪音计（声级计）	1357	TES	测试到达麦克风的声压
04	噪声源	音箱/电视	监听音箱建议型号：FX8	Fluid Audio	播放噪声、模拟外部干扰音
05	音频收集	高保真录音设备	R44	逻兰/Roland	音频的录制

4.5.1 人工嘴

型号：4227-A

性能指标：

a）额定输出声压 SPL：

●200Hz - 2kHz ----- 110dB

●100Hz - 8kHz ----- 100dB

b）失真（@94dB）：

●200Hz - 250Hz ----- <2%

●>250Hz ----- <1%

●阻抗 ----- 4Ω

ChipIntelli Confidential

●最大承受 ----- 10W

●瞬间承受功率 ----- 50W

●嘴径 ----- 20mm

4.5.2 精密噪音计

型号：TES 1357

性能指标：

a）0.1dB 分辨率

b）测量范围 30 到 130dB

c）1/1,1/3,1/6,1/12,1/24 倍频程频谱分析软件(可选)

d）准确度 +/- 1.5dB (ref 94dB @1KHz)

e）加权测量范围 30dB to 130dB

f）C 加权测量范围 35dB ~ 130dB

g）量测档位 30-80dB, 50-100dB, 60-110dB, 80-130dB

h）频率响应 31.5 Hz to 8KHz

i）数字显示 4 位数 LCD , 0.1dB resolution, updated every 0.5s

j）AC / DC 信号输出 2Vrms/每档满刻度，10mV/dB

4.5.3 噪声源监听音箱

型号：Fluid Audio FX8

性能指标：

a) 频率响应：35Hz - 22kHz(+/-3dB)

b) 交叉频率：2.4kHz

c) 低频放大器功率：80 watts

d) 高频放大器功率：50 watts

e) 信号噪声：> 100dB（typical A-weighted）

f) 极性：正信号+输入时产生一个向外的低频位移

g) 输入阻抗：20 千欧（平衡式），10 千欧（不平衡式）

h) 输入灵敏度：当音量控制设置为最大值（102dB 的最大声压）时

i) 输入 85 毫伏的粉红噪声会产生 95dBA 的输出声压

j) 电源：115V ~50/60 Hz 或 230V~50/60 Hz（用户可进行切换）

ChipIntelli Confidential

k) 保护装置：射频干扰，输出电流限制，过温保护，瞬态开启/关

l) 保护，超低音滤波器，外部电源保险丝

m) 箱体：乙烯基层压中密度纤维板

n) 尺寸（单个监听音箱）：340 毫米（高）x254 毫米（宽）x270 毫米（长）

o) 重量（单个监听音箱））：9.8 千克

4.6 语音识别测试环境

如图 4、图 5 所示，人工嘴（声音源）位于语音模块麦克风正前方[1]，水平直线距离 L 米[2]。人工嘴（声音源）距离地面 120-150cm；噪声源（监听音箱/电视）[3]、语音模块和精密噪音计位于同一平面处（距离地面 80-100cm）；噪声源（监听音箱/电视）与语音模块麦克风距离≥150cm，精密噪音计与语音模块麦克风尽量靠近（两者之间距离≤5cm），但不能与语音模块麦克风接触。

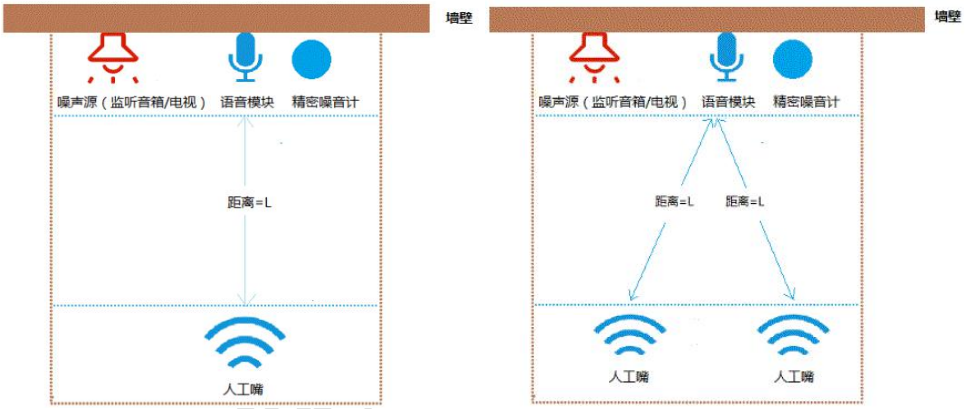


图 4 测试房间布置（定点测试）　　图 5 测试房间布置（非定点测试）

[1] 人工嘴的位置与角度可以根据客户的实际的场景而定；

[2] L 根据实际场景而定；

[3] 噪声源可以通过监听音箱或电视播放，噪声源的位置与角度可以根据客户的实际的情况而定。

5 语音识别测试方法及步骤

5.1 识别率/唤醒率测试方法与步骤

根据测试要求，改变人工嘴距离语音模组的位置与角度，构建不同声学场景，噪声源（监听音箱/电视）播放噪声集，人工嘴播放对应的测试集，记录测试的数据。

计算方法：

识别率 =（正确识别指令数/输入指令总数）*100%

唤醒率 =（正确唤醒率次数/输入指令总数）*100%

步骤：

1、使用噪声源（监听音箱/电视）连续播放噪声集；人工嘴按一定的时间间隔对测试集中的指令进行逐条播放；

2、记录测试数据；

3、统计、计算测试结果。

5.3 误唤醒测试方法与步骤

根据测试要求，改变人工嘴距离语音模块的位置与角度，构建不同声学场景，噪声源（监听音箱/电视）播放噪声集，人工嘴播放对应的测试集，统计误唤醒次数。

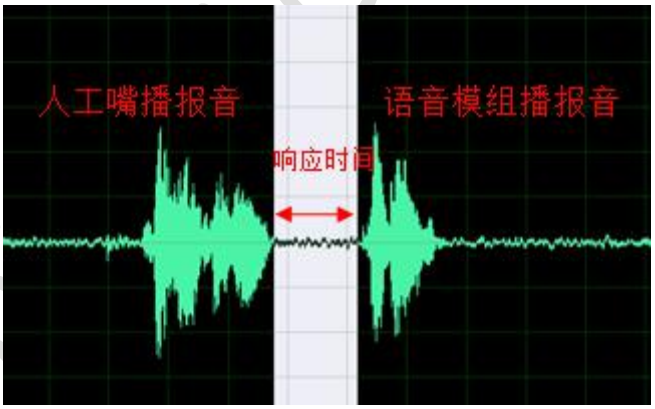
步骤：

1、使用噪声源（监听音箱/电视）连续播放噪声集，使用人工嘴播放测试集；

2、统计误唤醒次数。

5.4 响应时间测试方法与步骤

搭建好测试环境，打开语音录制工具，播放测试集，播报完成后，利用语音录制工具计算出语音指令与播报之间的时间间隔即为响应时间。



步骤：

1、使用人工嘴播放测试集；

2、记录测试数据；

3、计算响应时间。

5.5 稳定性测试方法与步骤

搭建好测试环境，噪声源（监听音箱/电视）播放不同种类噪声，人工嘴播放测试集，测试语音模组正常运行 168h，无重启记录，响应时间<1.0S。

ChipIntelli Confidential

步骤：

1、使用人工嘴播放测试集

2、记录测试数据

6 特别说明

本标准作为通用语音识别设备测试的参考标准和方法，可根据实际应用场景及条件调整。

Chipintelli Confidential

附录 A 部分噪声集

语音识别设备	对应的语音设备机械噪声	对应的环境噪声
油烟机	烟机噪声	厨房环境噪声
洗碗机	洗碗机噪声	厨房环境噪声
电饭煲	无	厨房环境噪声
微波炉	微波炉噪声	厨房环境噪声
豆浆机	豆浆机噪声	厨房环境噪声
咖啡机	咖啡机噪声	厨房环境噪声
电冰箱	电冰箱噪声	厨房环境噪声
空调	空调噪声	客厅环境噪声
电风扇	电风扇噪声	客厅环境噪声
吸尘器	吸尘器噪声	客厅环境噪声
加湿器	无	客厅环境噪声

注：噪声可以根据终端设备的实际应用场景来收集。

Chipintelli Confidential