

# 智能家居语音数据库

**AISHELL-ASR0012**

产品说明书

北京希尔贝壳科技有限公司

**Beijing Shell Shell Technology Co.,Ltd**

Add: Room 3-621, 6F, Zhongguancun Lifangting No. 1, Shanyuan Road, Haidian District, Beijing 100080, P.R.China

Tel: +86 10 80225006 E-mail: [bd@aishelldata.com](mailto:bd@aishelldata.com)

## 目录

1 产品概述.....	3
2 录音语料.....	3
2.1 语料池的制作.....	3
2.1.1 语料池内容.....	3
2.1.2 语料池处理.....	3
2.2 录音文本的结构设计.....	3
3 发音人信息.....	4
3.1 基本信息记录.....	4
3.2 发音人结构特征.....	4
3.2.1 性别比例.....	4
3.2.2 年龄比例.....	4
3.2.3 方言区域比例.....	5
4 数据录制环境.....	5
4.1 录制现场.....	5
4.2 录制环境.....	5
4.3 录制设备.....	5
4.4 录制方法.....	6
5 语音数据转写.....	6
6 语音数据转写.....	7
6.1 目录结构.....	7
6.2 命名规则.....	8
6.2.1 目录命名规则（图表 6-2-1）.....	8
7 版权声明.....	8

# 1 产品概述

AISHELL-ASR0012 智能机器人语音数据库共 22 小时。录音语言，中文；录音地区，中国。以中国北方口音区域为主邀请 646 名发音人参与录制。录制过程在真实家居环境中，模拟智能家居电器及应用产品使用情况，设置 7 个录音位。其中 5 个为高保真麦克风(44.1kHz, 16bit, 5\*300H)、1 个为 iOS 系统手机(16kHz, 16bit, 300H)与 1 个 Android 系统手机(16kHz, 16bit, 300H)进行录制。

此数据库经过专业语音校对人员转写标注，并通过严格质量检验，文本正确率在 95% 以上。

## 2 录音语料

### 2.1 语料池的制作

#### 2.1.1 语料池内容

考虑到语音识别在主流家居场景下的智能设备的应用，文本设计选用机器人领域文本。

#### 2.1.2 语料池处理

- 脱敏处理。删除政治敏感、个人隐私、色情暴力等内容。
- 删除 <, >, [, ], ~, /, \, = 等符号。
- 删除含有中文和英文以外语言的内容。
- 删除单句含有 25 字以上的内容。
- 统一格式。

### 2.2 录音文本的结构设计

考虑到语音覆盖及音素平衡，此数据库录音文本采用每份 11 句的分配方式设计，从语料池中抽取。

## 3 发音人信息

### 3.1 基本信息记录

发音人信息记录内容包括任务编号、性别、口音区域、年龄区间、籍贯。（图表 3-1）

任务编号	性别	口音区域	年龄区间	场景环境
0001	男	北方	A	场景一

图表 3-1-1

任务编号：每个发言人领取 1 个任务编号，每个任务编号对应 1 份录音文本。每个发言人只能参加一次录制。

性别：男性，女性。

口音区域：按照发言人原生语言所属区域，分为北方、南方、其他。

年龄区间：A(16-25 岁)、B(26-40 岁)、C(41 岁以上)。

场景环境：录制场景共十二个（图表 3-1-2），场景信息按编号记录在附件 AISHELL-ASR0010-ROOM-INFO 中。

场景	编号	场景	编号
场景一	bedroom_1	场景七	diningroom_1
场景二	bedroom_2	场景八	drawingroom_1
场景三	bedroom_3	场景九	drawingroom_2
场景四	bedroom_4	场景十	drawingroom_3
场景五	bedroom_5	场景十一	drawingroom_4
场景六	bedroom_6	场景十二	drawingroom_5

图表 3-1-2

### 3.2 发音人结构特征

#### 3.2.1 性别比例

数据库总人数为 646 人，男 277 人，女 369 人。（图表 3-2-1）

性别	男性	女性	合计
比例	43%	57%	100%

图表 3-2-1

#### 3.2.2 年龄比例

A（16-25 岁）631 人；B（26-40 岁）12 人；C（41 岁以上）3 人。（图表 3-2-2）

年龄区间	年龄段	人数	比例	男性	女性
A	16-25 岁	631	97.5%	270	361
B	26-40 岁	12	2%	6	6

C	>41 岁	3	0.5%	1	2
合计		646	100%	277	369

图表 3-2-2

### 3.2.3 方言区域比例

北方 542 人；南方 91 人；其他 13 人。（图表 3-2-3）

区域	人数	比例
北方	542	84%
南方	91	14%
其它	13	2%
合计	646	100%

图表 3-2-3

## 4 数据录制环境

### 4.1 录制现场

1. 录制场景为真实家居场景，场景内包括基本家居用品、家电桌椅。
2. 录制现场录音位由现场具体环境来确定。（涉及到录音位的距离、高度等）  
录制距离由实际场景左右会出现 0.5m,1.5m,2m,2.5m,3m,3.5m,4m,5m 的距离。
3. 在确定好录音场地录音位后，调试设备、实录等准备工作。保证录音设备正常、稳定、满足录制要求。
4. 录音人位置在指定位置录制语音数据。

### 4.2 录制环境

在家居环境中，设计客厅、卧室、餐厅三种不同场景，模拟智能电器及语控应用产品录音位。录制场景点位设置，请参见附件 AISHELL-ASR0010-ROOM-INFO。

### 4.3 录制设备

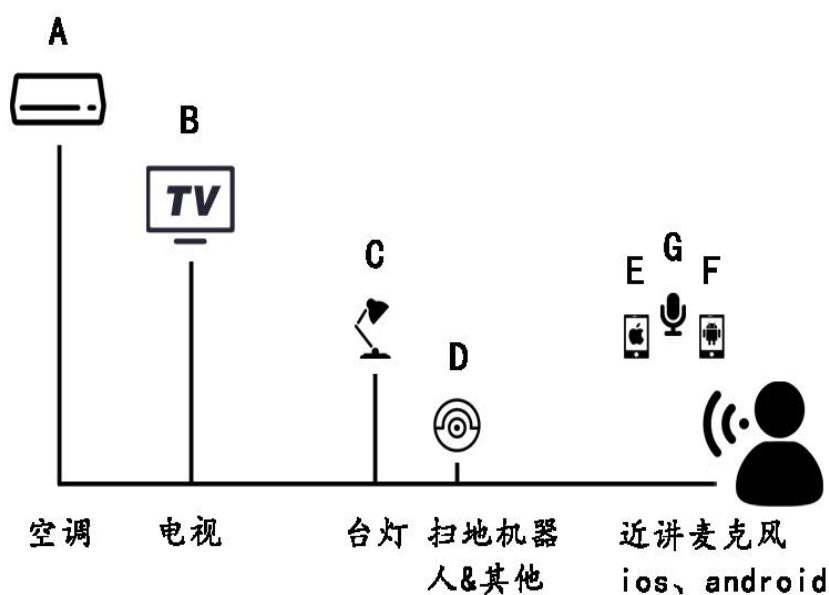
录制设备包括高保真麦克风和录音机、手机。本数据库数据存储格式为高保真录制数据 44.1kHz、16bit 单声道和手机录制数据 16kHz、16bit 单声道两种格式。

## 4.4 录制方法

家居环境下布置 7 个点位，包含 3 个近讲点位，4 个功能点位。功能点位固定在空调、电视、台灯、扫地机器人；近讲点位与发言人距离 30 厘米，包含一部高保真麦克风和两部手机(iOS & Android)。发音人以讲话正常音量，正常语速，朗读录音文本。（图表 4-4-1、图表 4-4-2）

序号	内容
A	录音位 1: 空调
B	录音位 2: 电视
C	录音位 3: 台灯
D	录音位 4: 扫地机器人&其它
E	录音位 5: 近讲 iOS
F	录音位 6 近讲 Android
G	录音位 7: 近讲麦克风

图表 4-4-1



图表 4-4-2

## 5 语音数据转写

数据转写人员根据所听到的音频写出内容，力求使文本内容与音频发音内容保持一致。

一般准则如下：

- 1) 转写的内容必须和听到的语音完全一致，不能多字、少字、错字。
- 2) 数字要转写为汉字形式，如“一二三”，而不是“123”。注意区分“一”和“幺”，“二”和“两”。
- 3) 音频中有英文发音的应写成相应的汉字或英文。具体分为以下几种情况：
  - 网址中包含的所有的字母或单词，均为大写。例如：发音内容为“www.abc.com”，应转写为“三 W 点 A B C 点 com”
  - 发音中包含的英文单词，转写时全部为小写。
  - 发音中包含的英文字母，转写时全部为大写。
  - 对于一些大写专有名词，或者一些英文缩写全部大写加空格，例如：CEO、CCTV 等。
- 4) 标注内容的完整性要与实际发音一致，不得删减。

## 6 语音数据转写

### 6.1 目录结构

数据目录树	
<b>数据目录结构</b>	
AISHELL-ASR0010.docx	( 数据库简介 )
└─DOC	( 文本说明文件 )
├─all_wav_list.txt	( 音频列表 )
├─content.txt	( 转写内容列表 )
├─readme.txt	( 目录说明文件 )
├─spkrinfo.xlsx	( 录音人信息 )
└─SPEECHDATA	( 数据文件夹 )
├─0001	( 录音人文件夹 )
H0001A	( 音频文件 )
H0001A0001.txt	( 语音内容文本 )
H0001A0001.wav	( 音频文件 )
H0001B	( 音频文件 )
H0001B0001.txt	( 语音内容文本 )
H0001B0001.wav	( 音频文件 )

## 6.2 命名规则

### 6.2.1 目录命名规则（图表 6-2-1）

CORPUS/CORPUS\_ID/SPEAKER\_NUM/SPEECH\_ID  
e.g. AISHELL-ASR0010/SPEECHDATA/0001/H0001A/ H0001A0001.wav

目录名称	内容	备注
CORPUS	AISHELL-ASR0010	语音数据库编号
USAGE	SPEECHDATA	文件夹名称
FILE_ID	0001	录音人文件夹名称
POINT_ID	H0001A	录音位编号
SENTENCE_ID	H0001A0001.txt	TXT 文件
SPEECH_ID	H0001A0001.wav	WAV 文件

图表 6-2-1

## 7 版权声明

本文内容禁止转载，AISHELL(北京希尔贝壳科技有限公司)对本文拥有修改权、更新权及最终解释权。

