

# 智能家居语音数据库

**AISHELL-ASR0020**

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 产品目录结构.....	3
2.1 目录结构.....	3
2.2 命名规则.....	4
3 文本设计.....	4
3.1 语料制作.....	4
3.2 文本结构.....	4
4 录制环境.....	5
4.1 录制现场.....	5
4.2 录制设备.....	5
4.3 录制方法.....	5
5 标注转写规范.....	5
6 发音人信息.....	6
6.1 基本信息记录.....	6
6.2 发音人结构特征.....	7
6.2.1 性别比例.....	7
6.2.2 年龄比例.....	7
7 版权声明.....	7

# 1 产品概述

AISHELL-ASR0020 智能家居语音数据库共 318.78 小时。录音语言，英语；录音地区，美国。录音文本包含主流家居场景智能控制、影音娱乐、唤醒词、自由文本。邀请 200 名美国本土发音人参与录制。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为录音位，使用 4 个高保真麦克风（44.1kHz，16bit）同时进行录制。

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 产品目录结构

### 2.1 目录结构

数据目录结构	
数据目录结构	
AISHELL-ASR0020.pdf	(数据库简介)
└─DOC	(文本说明文件)
─wav_list.txt	(音频列表)
─content.txt	(转写内容列表)
─spk_info.xlsx	(录音人信息)
└─SPEECHDATA	(数据文件夹)
─0001	(录音人文件夹)
H0001A	(音频文件夹)
H0001A0001.txt	(语音内容文本)
H0001A0001.wav	(音频文件)
H0001B	(音频文件夹)
H0001B0001.txt	(语音内容文本)
H0001B0001.wav	(音频文件)

图表 2-1-1 数据目录结构

## 2.2 命名规则

CORPUS/CORPUS\_ID/SPEAKER\_NUM/SPEECH\_ID

e.g. AISHELL-ASR0020/SPEECHDATA/0001/H0001A/ H0001A0001.wav

目录名称	内容	备注
CORPUS	AISHELL-ASR0020	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 2-2-1 命名规则

## 3 文本设计

### 3.1 语料制作

语料包含主流家居场景智能控制、影音娱乐、唤醒词、自由文本。按照规则处理生成语料池。

- 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。
- 删除含有英文以外语言的内容。
- 统一格式。

### 3.2 文本结构

考虑到语音覆盖及音素平衡，此数据库录音文本采用每份 500 句的分配方式设计，从语料池中抽取，结构如下。

序号	内容	每份分配量/句
1	智能控制	10
2	影音娱乐	90
3	唤醒词	50
4	自由文本	350
合计	4 项	500 句

图表 3-2-1 文本结构

## 4 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. 录制现场录音位分为 0.3m,1m,3m,5m 四种距离。
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 录制设备

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

### 4.3 录制方法

家居环境下布置 4 个点位，包含 1 个近讲点位，3 个功能点位。近讲点位与发言人距离 30 厘米；功能点位分别固定在 1m，3m，5m 的位置，模拟各种家电可能出现的位置。发音人以讲话正常音量，正常语速，朗读录音文本。

序号	内容
A	近讲点位：0.3m
B	功能点位：1m
C	功能点位：3m
D	功能点位：5m

图表 4-3-1 录音点位

## 5 标注转写规范

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) 数字要转写为英文形式, 如“one two three”, 而不是“123”。按照实际的发音如实转写。如“one hundred twenty three”、“twelve and three”。
- 3) 句中包含的符号, 按实际发言人发音转写。如“triple W dot Google dot com”。没有发音的符号, 需要删掉。品牌名称, 专有名称等按照实际惯用格式转写, 如“iPhone”、“PayPal”。
- 4) 按拼读朗读的字母, 需转写成大写字母加空格的形式。如, “NBA”、“DNA”。
- 5) 标注内容的完整性要与实际发音一致, 不得删减。

## 6 发音人信息

### 6.1 基本信息记录

发音人信息记录内容包括任务编号、性别、年龄、场景。

任务编号	性别	年龄区间	场景
0001	M	B	场景 1

图表 6-1-1 基本信息表

**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:** 录制场景共 4 个, 房间信息如下表。

场景	人数	房间
场景 1	27	15m 长*20m 宽*3.5m 高
场景 2	20	6m 长*4m 宽*3m 高
场景 3	142	5.5m 长*4m 宽*2.8m 高
场景 4	11	5.2m 长*3.5m 宽*3.2m 高

图表 6-1-2 场景列表

## 6.2 发音人结构特征

### 6.2.1 性别比例

数据库总人数为 200 人，男 128 人，女 72 人。

性别	男性	女性	合计
比例	64%	36%	100%

图表 6-2-1 性别比例

### 6.2.2 年龄比例

A（16-25 岁）80 人；B（26-40 岁）80 人；C（41 岁以上）40 人。

年龄区间	年龄段	人数	比例	男性	女性
A	16-25 岁	80	40%	40	40
B	26-40 岁	80	40%	68	12
C	>41 岁	40	20%	20	20
合计		200	100%	128	72

图表 6-2-2 年龄比例

## 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。



**希尔贝壳**  
A I S H E L L

**Copyright**