

# 美国英语智能家居语音数据库

**AISHELL-ASR0020-[US-EN]**

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
4 数据录制环境.....	5
4.1 录制现场.....	5
4.2 录制设备.....	5
4.3 录制方法.....	5
5 语音数据转写.....	6
6 语音数据转写.....	6
6.1 目录结构.....	6
6.2 命名规则.....	7
6.2.1 目录命名规则.....	7
7 版权声明.....	7

# 1 产品概述

此智能家居语音数据库共 400 小时。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)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

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。

#### 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 500 句的分配方式设计，从语料池中抽取，结构如下。（图表 2-2）

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
<b>合计</b>		<b>500 句</b>

图表 2-2

## 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、年龄、场景。（图表 3-1）

任务编号	性别	年龄	场景
<b>0001</b>	M	29	场景 1

图表 3-1

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。

#### 场景信息：

场景 1(27 人) : 15m 长\*20m 宽\*3.5m 高

场景 2(20 人) : 6m 长\*4m 宽\*3m 高

场景 3-(142 人) : 5.5m 长\*4m 宽\*2.8m 高

场景 4-(11 人) : 5.2m 长\*3.5m 宽\*3.2m 高

### 3.2 发音人结构特征

#### 3.2.1 性别比例

数据库总人数为 203 人，男 128 人，女 72 人。（图表 3-2-1）

性别	男性	女性	合计
<b>比例</b>	<b>62%</b>	<b>38%</b>	<b>100%</b>

图表 3-2-1

## 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 0.3m,1m,3m,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；功能点位固定在空调、电视、台灯（符合实际家居场景的距离）。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）

序号	内容
A	录音位 1: 近讲
B	录音位 2: 1 米
C	录音位 3: 3 米
D	录音位 4: 5 米

图表 4-4-1

## 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) 数字要转写为大写或实际发音形式，如“12”，转写为“one two”。如有连读的数字例如“123”，发音为“one hundred twenty three”。
- 3) 网址中包含的符号，需按发音转写。例如：发音内容为“www.abc.com”，应转写为“triple W dot A B D dot 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-ASR0020-[US-EN]_美国英语智能家居语音数据_产品说明书.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-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 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。

