

中文普通话智能家居语音数据库

AISHELL-ASR0010-[ZH-CN]

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

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.....	4
2.2 录音文本的结构设计.....	4
3 发音人信息.....	5
3.1 基本信息记录.....	5
3.2 发音人结构特征.....	5
3.2.1 性别比例.....	5
3.2.2 性别比例.....	5
3.2.3 方言区域比例.....	6
4 数据录制环境.....	6
4.1 录制现场.....	6
4.2 录制环境.....	6
4.3 录制设备.....	6
4.4 录制方法.....	7
5 语音数据转写.....	8
6 语音数据转写.....	8
6.1 目录结构.....	8
6.2 命名规则.....	9
6.2.1 目录命名规则.....	9
7 版权声明.....	9

1 产品概述

此智能家居语音数据库共 1800-2100 小时。录音文本包含主流家居场景智能控制、命令等 11 个分类。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为录音位,使用 4-5 个高保真麦克风(44.1kHz, 16bit, 300H) 与两部手机 IOS & ANDROID (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语控、命令词语、时事文本等领域的应用,语料在 11 个领域中选定。(图表 2-1)

序号	领域
1	机器人
2	空调
3	电视
4	智能音响
5	控制器
6	台灯
7	数字类
8	音乐类
9	电视、影视类
10	电台类
11	时事内容类

图表 2-1 语料池内容

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 520 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	领域	每份分配量/句
1	机器人	11
2	唤醒 1	13
3	控制器	25
4	台灯	20
5	唤醒 2-1	6
6	唤醒 2-2	6
7	唤醒 3	6
8	音响	25
9	唤醒、命令 4	50
10	收音机	5
11	空调	50
12	电视	40
13	音乐	40
14	数字	40
15	电影、视频	40
16	广域文本内容	143
合计	11 项	520 句

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、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	籍贯	场景环境
1	男	北方	A	河北	场景一

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。

口音区域：按照发言人原生语言所属区域，分为北方、南方、粤贵闽、其他。

年龄区间：A(16-25 岁)、B(26-40 岁)、C(41 岁以上)。

籍贯：记录发言人身份证所示籍贯信息。

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 性别比例

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 77 人；粤贵闽 14 人；其他 13 人。（图表 3-2-3）

区域	人数	比例
北方	542	84%
南方	77	12%
粤贵闽	14	2%
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。

4.3 录制设备

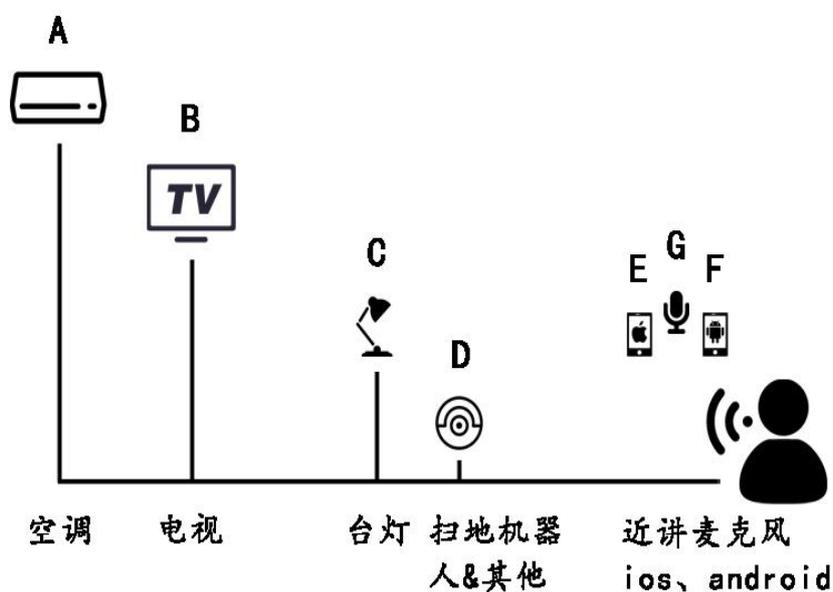
录制设备包括高保真麦克风和录音机、手机。本数据库数据存储格式为高保真录制数据 44.1kHz、16bit 单声道和手机录制数据 16kHz、16bit 单声道两种格式。

4.4 录制方法

家居环境下布置 6-7 个点位，包含 3 个近讲点位，3-4 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	
数据目录结构	
AISHELL-ASR0010-[ZH-CN]_中文普通智能家居语音数据_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-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。



希尔贝壳
A I S H E L L

copyright