家居环境近远讲同步语音数据库说明书

AISHELL-2019A-EVAL

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 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.....5
4 录制环境.....5
4.1 录制现场.....5
4.2 录制设备.....6
4.3 录制方法.....6
5 标注转写规范.....7
6 发音人信息.....8
6.1 基本信息记录.....8
6.2 发音人结构特征.....8
6.2.1 性别比例.....8
6.2.2 年龄比例.....9
6.2.3 方言区域比例.....9
7 版权声明.....9

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

1 产品概述

AISHELL-2019A-EVAL 是 AISHELL-ASR0010 的子库，共 24.3 小时。AISHELL-ASR0010 智能家居语音数据库共 1854 小时。录音语言，中文；录音地区，中国。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，每个麦克风录制时长在 218~284 小时之间)、1 个 iOS 系统手机(16kHz，16bit，283 小时)与 1 个 Android 系统手机(16kHz，16bit，281 小时)进行录制。

AISHELL-2019A-EVAL 随机抽取 50 个发音人。每人从位置 A(高保真 44.1kHz，16bit)与位置 F(Android 系统手机 16kHz，16bit)中，各选取 232 句到 237 句。

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 100%。

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-2019A-EVAL.pdf	**(数据库简介)**
| spk_info.xlsx	**(录音人信息)**
| room_info.xlsx	**(场景信息)**
|	
└─Android	**(设备文件夹)**
|├─data	**(数据文件夹)**
|| || wav.scp	**(音频列表)**
|| || trans.txt	**(标注结果)**
|| || wav	**(音频文件夹)**
|| || 0010	**(发言人文件夹)**
|| || H0010F0163.wav	**(音频文件)**

图表 2-1-1 数据目录结构

2.2命名规则

CORPUS/EQUIPMENT/data/wav/SPEAKER_NUM/SPEECH_ID

e.g.AISHELL-2019A-EVAL/Android/data/wav/0013/H0013F0108.wav

目录名称	内容	备注
CORPUS	AISHELL-2019A-EVAL	语音数据库编号
EQUIPMENT	MIC/Andriod	设备
data/wav	data/wav	音频文件夹
SPEAKER_NUM	0013	录音人文件夹名称
SPEECH_ID	H0013F0108.wav	WAV 文件

图表 2-2-1 命名规则

3 文本设计

3.1 语料制作

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。按照规则处理语料池。

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 3-1-1 语料池内容

- 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。

3.2 文本结构

考虑到语音覆盖及音素平衡，AISHELL-ASR-0010 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。

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 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. 录制现场录音位由现场具体环境来确定。（涉及到录音位的距离、高度等）
3. 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。

4. 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。
5. 录音人位置在指定位置录制语音数据。

4.2 录制设备

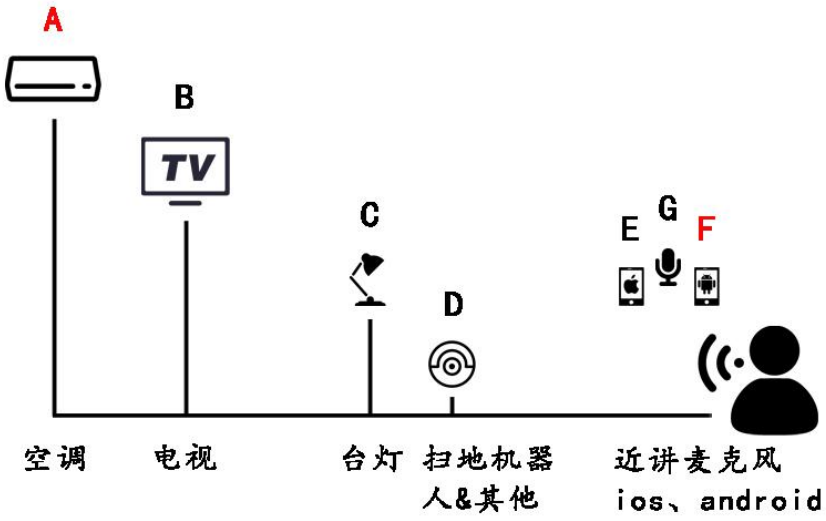
录制设备包括高保真麦克风和录音机、手机。本数据库数据存储格式为高保真录制数据44.1kHz、16bit 单声道和手机录制数据 16kHz、16bit 单声道两种格式。

4.3 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。

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 4-3-1 录音点位



图表 4-3-2 录制示意图

AISHELL-2019A-EVAL 数据集来自位置 A 与位置 F。

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。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）句中出现的英文按照发音写出单词，如“thank you”。按拼读朗读的字母，需转写成大写字母加空格的形式。如，“N B A”、“U F O”。

4）句中包含的符号，按实际发言人发音转写。如“三 W 点 百度 点 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“QQ 空间”、“iPhone”、“喜马拉雅”。

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、年龄区间、性别、口音区域、场景环境。

任务编号	年龄区间	性别	口音区域	场景环境
0013	B	男	北方	场景一

图表 6-1-1 基本信息表例

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。

年龄区间：A(15 岁以下)、B(16-25 岁)、C(26-40 岁)、D(41 岁以上)。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。

场景环境：录制场景共十二个，场景信息按编号记录在附件 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 6-1-2 场景列表

6.2 发音人结构特征

6.2.1 性别比例

AISHELL-2019A-EVAL 人数为 50 人，男 11 人，女 39 人。

性别	男性	女性	合计
比例	22%	78%	100%

图表 6-2-1 性别比例

6.2.2 年龄比例

A（15 岁以下）0 人；B（16-25 岁）49 人；C（26-40 岁）1 人；D（41 岁以上）0 人。

年龄区间	年龄段	人数	比例	男性	女性
A	＜15 岁	0	0	0	0
B	16-25 岁	49	98%	11	38
C	26-40 岁	1	2%	0	1
D	＞41 岁	0	0	0	0
合计		**50**	**100%**	**11**	**39**

图表 6-2-2 年龄比例

6.2.3 方言区域比例

北方 45 人；南方 5 人；其他 0 人。

区域	人数	比例
北方	45	90%
南方	5	10%
其它	0	0%
合计	**50**	**100%**

图表 6-2-3 方言区域比例

7 版权声明