

家居场景语控词语音数据库

AISHELL-ASR0039

产品说明书

目录

1 产品概述.....	3
2 录音人信息.....	3
2.1 性别比例.....	3
2.2 年龄比例.....	3
2.3 口音分布.....	3
3 录制环境.....	4
3.1 录制场地.....	4
3.2 录制设备.....	4
3.3 录制方法.....	4
4 语音数据转写.....	5
5 数据文件目录.....	5
5.1 目录结构.....	5
5.2 命名规则.....	5
6 版权声明.....	6

1 产品概述

AISHELL-ASR0039 家居场景语控词语音数据库，共邀请 399 人在安静环境下通过高保真麦克风进行录制。拾音点位共 7 个，每个点位拾音 80.78 小时，共 565 小时，采样率 48kHz，比特率 16bit。

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 录音人信息

2.1 性别比例

数据库总人数为 399 人，男 209 人，女 190 人。

性别	男性	女性	合计
比例	52%	48%	100%

图表 2-1-1 录音人性别比例

2.2 年龄比例

录音人年龄覆盖 16~50 岁，具体分布如下。

	年龄段	人数	比例
A	16-25 岁	149	37.3%
B	26-40 岁	185	46.4%
C	41-50 岁	65	16.3%
合计		399	100%

图表 2-2-1 年龄分布

2.3 口音分布

录音人在 7 个地区均匀分布，具体如下：

序号	地区	人数
1	福建	57

2	广东	57
3	广西	57
4	江苏	57
5	上海	57
6	四川	57
7	浙江	57
合计		399

图表 2-3-1 口音分布

3 录制环境

3.1 录制场地

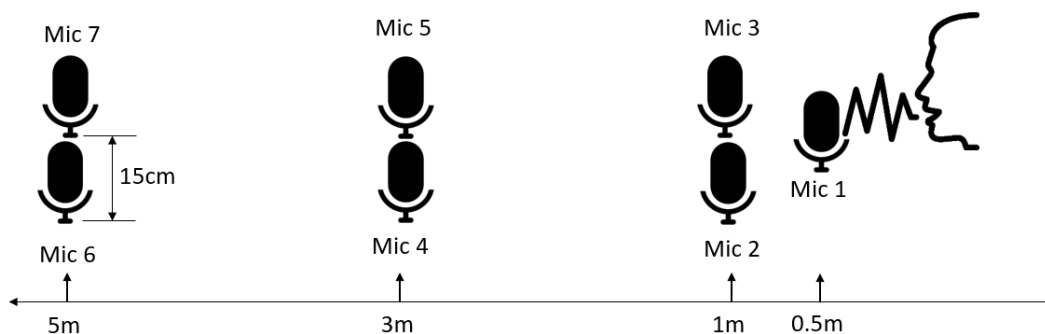
广东、广西、四川、福建、江苏、浙江和上海共七个地方城市，录制环境为安静室内。

3.2 录制设备

录制设备为高保真麦克风和录音机。本数据库数据存储格式为 48kHz、16bit 单声道。

3.3 录制方法

安静室内布置 7 个点位，包含 1 个 0.5 米近讲点位，6 个功能点位。功能点位固定在 1、3、5 米，每个功能点位 2 个麦克风，麦克风之间的间距为 15cm，如下图所示：



图表 3-3-1 录制点位示意图

4 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) 数字要转写为中文汉字形式，如“一二三”，而不是“123”。
- 3) 英文按单词小写，拼读发音需写成大写字母加空格，如，“W W W dot Google dot com”，“CEO”。
- 4) 标注内容的完整性要与实际发音一致，不得删减。

5 数据文件目录

5.1 目录结构

数据目录树	
数据目录结构	
AISHELL-ASR0039.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	(录音人信息表)
└─SPEECHDATA	(音频文件夹)
├─M0001	(录音人编号)
├─M0001.txt	(转录文本)
├─wav	(音频文件夹)
├─├─ M0001P01W0001.wav	(音频文件)

图表 5-1-1 目录结构

5.2 命名规则

CORPUS/USAGE/SPEAKER_ID/AUDIO_DIR/SPEECH_ID

e.g. AISHELL-ASR0039/SPEECHDATA/M0001/wav/M0001P01W0001.wav

目录名称	内容	备注
------	----	----

CORPUS	AISHELL-ASR0039	语音数据库编号
USAGE	SPEECHDATA	音频文件夹
SPEAKER_ID	M0001	录音人编号
AUDIO_DIR	wav	音频文件夹
SPEECH_ID	M0001P01W0001.wav	音频文件

图表 5-2-1 命名规则

6 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。



Copyright