**AISHELL-2中文语音数据库**

[1产品概述 2](#_Toc513118581)

[2录音语料 2](#_Toc513118582)

[2.1语料池的制作 2](#_Toc513118583)

[2.1.1语料池内容 2](#_Toc513118584)

[2.1.2语料池处理 3](#_Toc513118585)

[3录音人信息 3](#_Toc513118586)

[3.1 基本信息记录 3](#_Toc513118587)

[3.2 录音人信息结构特征 3](#_Toc513118588)

[3.2.1性别比例 3](#_Toc513118589)

[3.2.2年龄比例 4](#_Toc513118590)

[3.2.3方言区域比例 4](#_Toc513118591)

[4数据录制环境 4](#_Toc513118592)

[4.1 录制环境 4](#_Toc513118593)

[4.2 录制设备 4](#_Toc513118594)

[4.3 录制方法 5](#_Toc513118595)

[5语音数据转写 5](#_Toc513118596)

[6数据文件目录 6](#_Toc513118597)

[6.1 目录结构 6](#_Toc513118598)

[6.2 命名规则 6](#_Toc513118599)

[6.2.1 目录命名规则 6](#_Toc513118600)

[6.2.2 文件命名规则 7](#_Toc513118601)

#

# 1产品概述

1.
2.
3.
4.
5.
6.

希尔贝壳中文普通话语音数据库AISHELL-2的语音时长为1000小时，其中718小时来自AISHELL-ASR0009-[ZH-CN]，282小时来自AISHELL-ASR0010-[ZH-CN]。录音文本涉及唤醒词、语音控制词、智能家居、无人驾驶、工业生产等12个领域。

录制过程在安静室内环境中， 同时使用3种不同设备： 高保真麦克风（44.1kHz，16bit）；Android系统手机（16kHz，16bit）；iOS系统手机（16kHz，16bit）。AISHELL-2采用iOS系统手机录制的语音数据。

1991名来自中国不同口音区域的发言人参与录制。经过专业语音校对人员转写标注，并通过严格质量检验，此数据库文本正确率在96%以上。

# 2录音语料

## 2.1语料池的制作

### 2.1.1语料池内容

考虑到语音识别在智能家居、无人驾驶、工业生产等领域的应用，语料在12个领域中选定。

|  |  |
| --- | --- |
| **序号** | **领域** |
| **1** | 家居家电名称和控制命令 |
| **2** | POI（地理信息） |
| **3** | 音乐、收音机类（语控） |
| **4** | 数字类（语控） |
| **5** | 电视、电影名称  |
| **6** | 财经 |
| **7** | 科技 |
| **8** | 体育 |
| **9** | 娱乐 |
| **10** | 时事新闻 |
| **11** | 英文拼读 |
| **12** | 广域文本 |

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。
* 统一格式。

# 3录音人信息

## 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** | **年龄区间** | **性别** | **口音区域** |
| **C0002** | **B** | **男** | **North** |

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、其他。

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

## 3.2 录音人信息结构特征

### 3.2.1性别比例

 数据库总人数为1991人，男842人，女1149人。（图表3-2-1）

|  |  |  |  |
| --- | --- | --- | --- |
| **性别** | **男性** | **女性** | **合计** |
| **比例** | **42.3%** | **57.7%** | **100%** |

图表3-2-1

### 3.2.2年龄比例

 A(16岁以下) 16人；B(16-25岁)1707人；C(26-40岁)201人；D(41岁以上)67人。（图表3-2-2）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **年龄段** | **人数** | **比例** | **男性** | **女性** |
| **A** | **16岁以下** | 16 | 0.8% | 5 | 11 |
| **B** | **16-25岁** | 1707 | 85.7% | 721 | 986 |
| **C** | **26-40岁** | 201 | 10.1% | 88 | 113 |
| **D** | **> 41岁** | 67 | 3.4% | 28 | 39 |
| **合计** |  | **1991** | **100%** | **842** | **1149** |

图表3-2-2

### 3.2.3方言区域比例

 北方1279人；南方686人；其他26人。（图表3-2-3）

|  |  |  |
| --- | --- | --- |
| **区域** | **人数** | **比例** |
| **北方** | 1279 | 64.2% |
| **南方** | 686 | 34.5% |
| **其他** | 26 | 1.3% |
| **合计** | **1991** | **100%** |

图表3-2-3

# 4数据录制环境

## 4.1 录制环境

安静室内包括真实家居场景，不包括明显的其他人说话声音及其他噪音，无回音。发言人按照正常语速，朗读录音文本。

## 4.2 录制设备

录制设备包括高保真麦克风和录音机、iOS系统手机、Android系统手机。AISHELL-2数据均来自iOS设备。

## 4.3 录制方法

 发音人距离高保真麦克风20厘米，以讲话正常音量，正常语速，朗读录音文本。Android系统手机与iOS系统手机分别与麦克风间隔20厘米布置。（图表4-3）



图表4-3

# 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：C E O、C C T V 等。

4）标注内容的完整性要与实际发音一致，不得删减。

# 6数据文件目录

## 6.1 目录结构

|  |
| --- |
| **数据目录树** |
| **数据目录结构****│ AISHELL2-Data-Specification .docx （数据库说明）****│ README.txt （概述）****│ ChangeLog （数据版本信息）** **└─ data （数据文件夹）****│├─wav.scp （音频文件列表）****│├─trans.txt （转写文本列表）****│├─spk\_info.txt （录音人信息列表）****│├─wav （音频文件夹）****│├─D2166 （录音人文件夹）** **││ ││ ID2166W0001.wav （音频文件）** |

## 6.2 命名规则

### 6.2.1 目录命名规则

**/<CORPUS>/<DATA1>/<DATA2>/<USAGE>/<FILE\_ID>/<SPEECH\_ID>**

e.g.AISHELL-2/iOS/data/wav/D2166/ID2166W0001.wav

|  |  |  |
| --- | --- | --- |
| **目录** | **内容** | **备注** |
| **CORPUS** | AISHELL-2 | 数据库 |
| **DATA1** | iOS | iOS |
| **DATA2** | data & RADME & ChangeLog | 数据信息 |
| **USAGE** | wav | 数据文件夹 |
| **FILE\_ID** | D2166 | 录音人文件夹 |
| **SPEECH\_ID** | ID2166W0001.wav | WAV文件 |

图表6-2-1

### 6.2.2 文件命名规则

**<CORPUS\_ID><SPEAKER\_IC><WAV\_NUM>**

e.g. ID2166W0001.wav

|  |  |  |
| --- | --- | --- |
| **文件** | **内容** | **备注** |
| **CORPUS\_ID** | I | iOS |
| **SPEAKER\_NUM** | D2166 | 录音人ID |
| **WAV\_NUM** | W0001 | WAV 编号 |

图表6-2-2