

# 中文普通话口语语音数据库

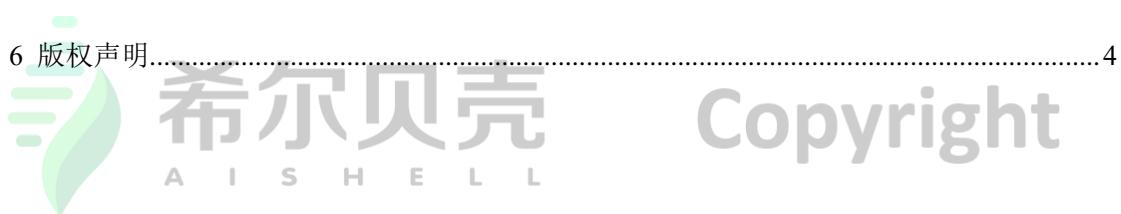
AISHELL-ASR0016



Copyright

## 目录

1 产品概述.....	2
2 发音人信息.....	2
2.1 性别比例 .....	2
2.2 年龄比例 .....	2
3 数据录制环境.....	2
3.1 录制环境.....	2
3.2 录制设备.....	3
3.3 录制方法.....	3
4 语音数据转写.....	3
5 数据文件目录.....	4
6 版权声明.....	4



# 1 产品概述

AISHELL-ASR0016 中文普通话口语语音数据库共 566 小时，单通道 188.7 小时，录音语言为中文。共邀请 219 名来自中国北方地区发音人参与录制。录制过程为录音棚，录音人通过高保真麦克风，以自由说的形式进行录制。

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 96% 以上。此数据库可用于声纹识别、语音识别等研究。

## 2 发音人信息

### 2.1 性别比例

数据库总人数为 219 人，男 115 人，女 104 人。

性别	男性	女性	合计
比例	52.3	47.7	100%

图表 2-1-1

### 2.2 年龄比例

录音人年龄覆盖 16-58 岁，具体分布如下所示：A (16-25 岁)；B (26-40 岁)；C (41 岁及以上)。

年龄段	人数	比例 (%)
16-25	127	58.3
26-40	56	25.2
>41	36	16.5
合计	219	100

图表 2-2-1

## 3 数据录制环境

### 3.1 录制环境

录音棚内，无其他人说话。

## 3.2 录制设备

高保真麦克风。

## 3.3 录制方法

录音人在录音棚内通过高保真麦克风录制，麦克风距离为嘴边(P03)、1米(P04)以及2米(P05)同步录制。文本不限，自由说话，每个人不低于10分钟，不超过1小时。

# 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。
- 5) 一句中有个别字听不清的标注\*，几个字标注一个\*；整句听不清占比超过50%的句子丢弃。
- 6) 异常噪声，笑声，咳嗽等人声噪声在发声位置进行符号标记，符号如下所示：

序号	符号	说明
1	<\$>	笑声
2	<%>	咳嗽
3	<#>	打哈欠等
4	<@>	手机铃声

5	<->	动椅子声
6	<_>	动桌子声
7	<*>	听不清
8	<fy>	个别字存在方言

图表 4-1-1

## 5 数据文件目录

数据目录树	
<b>数据目录结构</b>	
AISHELL-ASR0016.pdf	(数据库简介)
└ DOC	(文本说明文件夹)
content.txt	(音频列表)
spkrinfo.xlsx	(录音人信息)
└ SPEECHDATA	(数据文件夹)
0001	(录音人编号)
P03	(点位)
00001.wav	(音频文件)
P04	(点位)

## 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