

# 中文普通话儿童语音数据库

## 数据产品说明书

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

|                    |   |
|--------------------|---|
| 1 产品概述.....        | 2 |
| 2 录音语料.....        | 2 |
| 2.1 语料池的制作.....    | 2 |
| 2.1.1 语料池内容 .....  | 2 |
| 2.1.2 语料池处理 .....  | 2 |
| 2.2 录音文本的结构设计..... | 2 |
| 3 发音人信息.....       | 3 |
| 3.1 基本信息记录.....    | 3 |
| 3.2 发音人结构特征.....   | 3 |
| 3.2.1 性别比例 .....   | 3 |
| 3.2.2 年龄比例 .....   | 3 |
| 3.2.3 方言区域比例 ..... | 4 |
| 4 数据录制环境.....      | 4 |
| 4.1 录制环境.....      | 4 |
| 4.2 录制设备.....      | 4 |
| 4.3 录制方法.....      | 4 |
| 5 语音数据转写.....      | 4 |
| 6 版权声明.....        | 5 |

# 1 产品概述

此中文普通话语音数据库共邀请 1000 名来自中国不同口音区域发音人参与录制。录制过程在安静室内环境中，使用手机（16kHz，16bit）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

考虑到语音识别在智能家居、教育、工业生产等领域的应用，语料设计包含诗歌、童话、食物名称、数字串等内容。

#### 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，此数据库录音文本设计分为 A 级（3-6 岁）、B 级（7-10 岁）两类。A 级文本每份 200 句，B 级文本每份 275 句，从语料池中抽取，结构如下。（图 2-2）

|      |    | A 级 | B 级 |
|------|----|-----|-----|
| 语控词  | CO | 38  | 63  |
| 食物名称 | FD | 25  | 19  |
| 诗歌   | PT | 30  | 30  |

|        |    |     |     |
|--------|----|-----|-----|
| 数字串    | NU | 36  | 51  |
| 童话     | TA | 27  | 11  |
| 平衡句    | SE | 25  | 45  |
| 十万个为什么 | QU | 19  | 56  |
| 合计     |    | 200 | 275 |

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。（图表 3-1）

| 任务编号  | 性别 | 口音区域 |
|-------|----|------|
| A0001 | 男  | 北方   |

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。任务编号已文本等级开头，A 级(3-6 岁)、B 级（7-10 岁）。

口音区域：按照发言人原生语言所属区域，分为北方、南方。

### 3.2 发音人结构特征

#### 3.2.1 性别比例

| 性别 | 男性  | 女性  | 合计   |
|----|-----|-----|------|
| 比例 | 43% | 57% | 100% |

图表 3-2-1

#### 3.2.2 年龄比例

A

|   | 年龄段   | 比例  |
|---|-------|-----|
| A | 3-6 岁 | 50% |

|    |       |      |
|----|-------|------|
| B  | 7-10岁 | 18%  |
| 合计 |       | 100% |

图表 3-2-2

### 3.2.3 方言区域比例

| 区域 | 比例   |
|----|------|
| 北方 | 50%  |
| 南方 | 50%  |
| 合计 | 100% |

图表 3-2-3

## 4 数据录制环境

### 4.1 录制环境

安静室内，不包括明显的其他人说话声音及其他噪音，无回音。由领读人朗读文本，发言人按照正常语速，重复文本。

### 4.2 录制设备

录制设备为手机。

### 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，重复朗读人所述录音文本。

## 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 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。

