

儿童语音数据库

AISHELL-ASR0018



希尔贝壳
A I S H E L L

产品说明书

Copyright

目录

1 产品概述.....	2
2 产品目录结构.....	2
2.1 目录结构.....	2
2.2 命名规则.....	2
3 文本设计.....	3
3.1 语料制作.....	3
3.2 文本结构.....	3
4 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 标注转写规范.....	4
6 发音人信息.....	5
6.1 基本信息记录.....	5
6.2 发音人结构特征.....	5
6.2.1 性别比例.....	5
6.2.2 年龄比例.....	5
6.2.3 方言区域比例.....	6
7 版权声明.....	6

1 产品概述

AISHELL-ASR0018 儿童语音数据库共 171 小时。录音语言，中文；录音地区，中国。邀请 1117 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 产品目录结构

1.1 目录结构

数据目录结构	
数据目录结构	
AISHELL-ASR0018.pdf	(数据库简介)
└─DOC	(文本说明文件)
─wav_list.txt	(音频列表)
─content.txt	(语音文本内容)
─spk_info.xlsx	(录音人信息)
└─SPEECHDATA	(数据文件夹)
─A0001	(录音人文件夹)
A0001CO0001.wav	(音频文件)

图表 2-1-1 数据目录结构

1.2 命名规则

CORPUS/USAGE/SPEAKER_NUM/POINT_ID/SPEECH_ID

e. g. AISHELL-ASR0018/SPEECHDATA/A0001/A0001CO0001.wav

目录名称	内容	备注
CORPUS	AISHELL-ASR0018	语音数据库编号
USAGE	SPEECHDATA	文件夹名称
SPEAKER_NUM	A0001	录音人文件夹名称

SPEECH_ID	A0001CO0001.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。按照规则处理语料池。

- 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，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。

内容	内容编号	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 3-2-1 文本结构

4 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 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、性别、口音区域。

任务编号	性别	口音区域
A0001	男	北方

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。任务编号已文本等级开头，A 级（3-6 岁）、B 级（7-10 岁）。

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。

6.2 发音人结构特征

6.2.1 性别比例

数据库总人数为 1117 人，男 491 人，女 626 人。

性别	男性	女性	合计
比例	44%	56%	100%

图表 6-2-1 性别比例

6.2.2 年龄比例

A（3-6 岁）583 人；B（7-10 岁）534 人。

编号	年龄段	比例
A	3-6 岁	51%
B	7-10 岁	49%
合计		100%

图表 3-2-2 年龄比例

6.2.3 方言区域比例

北方 453 人；南方 664 人；其他 0 人。

区域	比例
北方	40.6%
南方	59.4%
其他	0%
合计	100%

图表 6-2-3 方言区域比例

7 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。



希尔贝壳
A I S H E L L

Copyright