

中文普通话智能车载语音数据库

AISHELL-ASR0011-[ZN-CN]

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

目录

| | |
|--------------------|---|
| 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..... | 3 |
| 2.2 录音文本的结构设计..... | 3 |
| 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..... | 4 |
| 3.2.1 性别比例..... | 4 |
| 3.2.2 性别比例..... | 4 |
| 3.2.3 方言区域比例..... | 4 |
| 4 数据录制环境..... | 5 |
| 4.1 录制环境..... | 5 |
| 4.2 录制设备..... | 6 |
| 4.3 录制方法..... | 6 |
| 5 语音数据转写..... | 7 |
| 6 数据文件目录..... | 8 |
| 6.1 目录结构..... | 8 |
| 6.2 命名规则..... | 8 |
| 6.2.1 目录命名规则..... | 8 |
| 6.2.2 文件命名规则..... | 9 |
| 7 版权声明..... | 9 |

1 产品概述

此智能车载语音数据库共 150 小时。录音文本包含主流车载场景智能控制、命令等 10 个分类。以中国北方口音区域为主邀请 306 名发音人参与录制。录制过程在真实车载环境中，模拟自动驾驶点位，使用 4 个高保真麦克风（44.1kHz，16bit，150H）与一部手机（16kHz，16bit，150H）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车载、无人驾驶领域的应用，语料在 10 个领域中选定。（图表 2-1-1）

| 序号 | 领域 |
|----|---------|
| 1 | 唤醒类 |
| 2 | 高速公路信息 |
| 3 | 行政地理位置 |
| 4 | 命令、控制语句 |
| 5 | 数字串 |
| 6 | 音乐 |
| 7 | 收音机 |
| 8 | 空调 |
| 9 | 地图坐标位置 |
| 10 | 广域内容文本 |

图表 2-1-1 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。
- 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，此数据库录音文本采用每份 508 句的分配方式设计，从语料池中抽取，结构如下。（图表 2-2）

| 序号 | 领域 | 每份分配量/句 |
|-----------|-------------|--------------|
| 1 | 综合 | 50 |
| 2 | 唤醒类 | 48 |
| 3 | 高速公路信息 | 20 |
| 4 | 行政地理位置 | 30 |
| 5 | 命令、控制语句 | 35 |
| 6 | 数字串 | 40 |
| 7 | 音乐 | 30 |
| 8 | 收音机 | 30 |
| 9 | 空调 | 35 |
| 10 | 地图坐标位置 | 40 |
| 11 | 广域内容文本 | 50 |
| 合计 | 10 项 | 508 句 |

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、年龄区间、籍贯。（图表 3-1）

| 任务编号 | 性别 | 口音区域 | 年龄区间 | 籍贯 | 场景环境 | 车型 |
|------|----|------|------|----|------|----|
| 1 | 男 | 北方 | A | 河北 | 场景一 | 小型 |

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(16-25 岁)、B(26-40 岁)、C(41 岁以上)。

籍贯：记录发言人身份证所示籍贯信息。

3.2 发音人结构特征

3.2.1 性别比例

数据库总人数为 306 人，男 155 人，女 151 人。（图表 3-2-1）

| 性别 | 男性 | 女性 | 合计 |
|----|-----|-----|------|
| 比例 | 50% | 50% | 100% |

图表 3-2-1

3.2.2 性别比例

A (16-25 岁) 300 人；B (26-40 岁) 6 人；C (41 岁以上) 0 人。（图表 3-2-2）

| 年龄区间 | 年龄段 | 人数 | 比例 | 男性 | 女性 |
|------|---------|-----|------|-----|-----|
| A | 16-25 岁 | 300 | 98% | 151 | 149 |
| B | 26-40 岁 | 6 | 2% | 4 | 2 |
| C | >41 岁 | 0 | 0% | 0 | 0% |
| 合计 | | 306 | 100% | 155 | 151 |

图表 3-2-2

3.2.3 方言区域比例

北方 294 人；南方 10 人；其它 2 人。（图表 3-2-3）

| 区域 | 人数 | 比例 |
|----|-----|------|
| 北方 | 294 | 96% |
| 南方 | 10 | 3% |
| 其它 | 2 | 1% |
| 合计 | 300 | 100% |

图表 3-2-3

4 数据录制环境

4.1 录制环境

在车载环境中，设计 7 种不同场景，依次录制。车型包含 3 辆小型车、2 辆中型车、1 辆大型车。（图表 4-1-1）

| 场景环境编号 | 车速 |
|--------|------------------------------------|
| A | 关闭引擎-> 市区慢速 0-40km/h |
| B | 市区中速 40-80km/h 打开 1/3 副驾驶车窗 |
| C | 高速 80-120km/h 关窗 |
| D | 关闭引擎 打开 1/3 副驾驶车窗 -> 市区慢速 0-40km/h |
| E | 关闭引擎 打开 1/3 驾驶员车窗 -> 市区慢速 0-40km/h |
| F | 市区中速 40-80km/h 打开 1/3 副驾驶车窗 |
| G | 市区中速 40-80km/h 打开 1/3 驾驶员车窗 |

图表 4-1-1

场景内环境结合语料再分配为音乐开关、空调开关等辅助环境。（图表 4-1-2）

| 场景环境编号 | 车速 | 辅助环境 |
|--------|------------------------------------|-------|
| A | 关闭引擎-> 市区慢速 0-40km/h | 音乐/空调 |
| B | 市区中速 40-80km/h 打开 1/3 副驾驶车窗 | 音乐/空调 |
| C | 高速 80-120km/h 关窗 | 音乐/空调 |
| D | 关闭引擎 打开 1/3 副驾驶车窗 -> 市区慢速 0-40km/h | 音乐/空调 |
| E | 关闭引擎 打开 1/3 驾驶员车窗 -> 市区慢速 0-40km/h | 音乐/空调 |
| F | 市区中速 40-80km/h 打开 1/3 副驾驶车窗 | 音乐/空调 |
| G | 市区中速 40-80km/h 打开 1/3 驾驶员车窗 | 音乐/空调 |

图表 4-1-2

音乐和空调状态以组合形式再分为 6 组分配到语料中。（图表 4-1-3）

| 辅助环境 | |
|-------|------|
| 空调状态 | 音乐状态 |
| 空调关闭 | 音乐关闭 |
| 空调关闭 | 音乐打开 |
| 空调打开低 | 音乐关闭 |
| 空调打开低 | 音乐打开 |
| 空调打开高 | 音乐关闭 |
| 空调打开高 | 音乐打开 |

图表 4-1-3

4.2 录制设备

录制设备包括高保真麦克风和录音机、手机。本数据库数据存储格式为高保真录制数据 44.1kHz、16bit 单声道和手机录制数据 16kHz、16bit 单声道两种格式。

4.3 录制方法

车内环境下布置 5 个点位，包含 2 个近讲点位，3 个功能点位。功能点位固定在驾驶员正前方、后视镜、中控台；近讲点位与发言人距离 20 厘米，包含一部高保真麦克风和一部手机。发音人以讲话正常音量，正常语速，朗读录音文本。（图表 4-3-1、图表 4-3-2）

| 编号 | 位置 | 描述 |
|----|---------------|--------|
| A | 驾驶员（副驾驶模拟）正前方 | 高保真麦克风 |
| B | 车顶 | 高保真麦克风 |
| C | 中控台 | 高保真麦克风 |
| D | 后视镜 | 高保真麦克风 |
| F | 录音人前方 | 手机 |

图表 4-3-1



图表 4-3-2

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 数据文件目录

6.1 目录结构

| 数据目录树 | |
|---|----------|
| 数据目录结构 | |
| AISHELL-ASR0011-[ZH-CN]_中文普通智能车载语音数据_产品说明书.docx | (数据库简介) |
| └─DOC | (文本说明文件) |
| ├─all_wav_list.txt | (音频列表) |
| ├─content.txt | (转写内容列表) |
| ├─readme.txt | (目录说明文件) |
| ├─spkrinfo.xlsx | (录音人信息) |
| └─SPEECHDATA | (数据文件夹) |
| ├─0001 | (录音人文件夹) |
| ├─CC0001A | (录音人文件夹) |
| ├─CC0001A0001.wav | (音频文件) |
| └─CC0001A0001.txt | (语音内容文本) |

6.2 命名规则

6.2.1 目录命名规则（图表 6-2-1）

/<CORPUS>/<USAGE>/<FILE_ID>/<DRIVING CONDITIONS>/<SPEECH_ID>

e. g. AISHELL-ASR0011/SPEECHDATA/0001/A/CC0001A0001. wav

| 目录 | 内容 | 备注 |
|---------------------------|-----------------|----------|
| CORPUS | AISHELL-ASR0011 | 数据库名称编号 |
| USAGE | SPEECHDATA | 文件夹名称 |
| DRIVING CONDITIONS | A\B\C\D\E\F\G | 行车环境 |
| FILE_ID | 0001 | 录音人文件夹名称 |
| POINT_ID | CC0001A | 录音位 ID |
| SENTENCE_ID | CC0001A0001.txt | TXT 文件 |
| SPEECH_ID | CC0001A0001.wav | WAV 文件 |

图表 6-2-1

6.2.2 文件命名规则（图表 6-2-2）

<CORPUS_ID><SPEAKER_IC><WAV_NUM>

e. g. CC0001A0001.wav

| 文件 | 内容 | 备注 |
|--------------------|------------------|--------|
| CORPUS_ID | CC (Chinese\Car) | 数据库名称 |
| SPEAKER_NUM | 0001 | 录音人 ID |
| POINT_ID | A\B\C\D\F | 录音位 ID |
| WAV_NUM | 0001 | WAV 编号 |

图表 6-2-2

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。

