

智能车载语音数据库

AISHELL-ASR0011

产品说明书

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 录制现场.....	4
4.2 录制设备.....	5
4.3 录制方法.....	6
5 标注转写规范.....	7
6 发音人信息.....	8
6.1 基本信息记录.....	8
6.2 发音人结构特征.....	8
6.2.1 性别比例.....	8
6.2.2 年龄比例.....	8
6.2.3 方言区域比例.....	9
7 版权声明.....	9

1 产品概述

AISHELL-ASR0011 智能车载语音数据库共 613 小时。录音语言，中文；录音地区，中国。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 300 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，122 小时）与一部手机（16kHz，16bit，124 小时）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-ASR0011.docx	(数据库简介)
└─DOC	(文本说明文件)
├─wav_list.txt	(音频列表)
├─content.txt	(转写内容列表)
├─spk_info.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)

图表 2-1-1 数据目录结构

1.2 命名规则

<CORPUS_ID><SPEAKER_IC><WAV_NUM>

e.g. CC0001A0001.wav

目录名称	内容	备注
CORPUS_ID	CC (Chinese in 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 6-2-1 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车载、无人驾驶领域的应用，语料在 10 个领域中选定。按照规则处理语料池。

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 3-1-1 语料池内容

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

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 3-2-1 文本结构

4 录制环境

4.1 录制现场

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。

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 场景列表

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。

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 辅助环境列表

音乐和空调状态以组合形式再分为 6 组分配到语料中。

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
低风空调打开	音乐关闭
低风空调打开	音乐打开
高风空调打开	音乐关闭
高风空调打开	音乐打开

图表 4-1-3 辅助环境

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。

舒伯乐 superlux ECM999 测试话筒



信息项	技术参数
音头	背驻极
指向性	无指向
频率响应	20-20000Hz
灵敏度	-37 dBV/Pa (14 mV)
阻抗	200 Ω
负载	1000 Ω
信噪比	70 dB 等效噪音 SPL:24 dB
声压	SPL: 132dB
动态范围	108 dB
动态范围	模数: 100dB, 数模: 104dB

图表 4-2-1 高保真麦克风



信息项	技术参数
通道	4
数据类型	WAV/BWF
采样精度	16/24-bit
采样频率	44.1kHz/48kHz/88.2kHz/96kHz/192kHz
输入阻抗	XLR:4K 欧姆或更高 (平衡), TRS:6K 欧姆或更高 (平衡)
输入	+24dBu (输入灵敏度旋钮: +4dBu)
输出阻抗	线路输出: 600 欧姆
噪音	线路输出: -100dBu (输入灵敏度: +4dBu, 输入音量: 中间)
频率响应	20Hz-40kHz (0/-3dB)
动态范围	模数: 100dB, 数模: 104dB

图表 4-2-2 录音机

手机系统	手机型号
Android	三星 NOTE 4、三星 S6、OPPO A33、OPPO A59s、华为荣耀畅玩 6X

图表 4-2-3 手机型号

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。

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 录音点位



图表 4-3-2 录制示意图

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、年龄区间、籍贯。

任务编号	性别	口音区域	年龄区间	场景环境
1	男	北方	A	场景一

图表 6-1-1 基本信息表

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。

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、其他。

年龄区间：A(16-25 岁)、B(26-40 岁)、C(41 岁以上)。

场景环境：

6.2 发音人结构特征

6.2.1 性别比例

数据库总人数为 300 人，男 151 人，女 149 人。

性别	男性	女性	合计
比例	50%	50%	100%

图表 6-2-1 性别比例

6.2.2 年龄比例

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

年龄区间	年龄段	人数	比例	男性	女性
A	16-25 岁	296	98.7%	149	147
B	26-40 岁	4	1.3%	2	2
C	>41 岁	0	0%	0	0%
合计		300	100%	151	149

图表 6-2-2

6.2.3 方言区域比例

北方 294 人；南方 8 人；其它 2 人。

区域	人数	比例
北方	294	98.0%
南方	8	2.7%
其它	2	0.7%
合计	300	100%

图表 6-2-3

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