

智能车载交互语音数据库（测试集）



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 813, Building No. 4, Shangdi 10th Street, Haidian District, Beijing 100080, P.R.China

Tel: 010-80225006 E-mail: bd@aishelldata.com

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.....	5
5 标注转写规范.....	5
6 发音人信息.....	6
6.1 基本信息记录.....	6
6.2 发音人结构特征.....	7
6.2.1 性别比例.....	7
6.2.2 年龄比例.....	7
6.2.3 方言区域比例.....	7
7 版权声明.....	7

1 产品概述

AISHELL-2021A-EVAL 是 AISHELL-ASR0031 的子库，共 7.2 小时。AISHELL-ASR0031 智能车载语音数据库共 10903 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、命令词等 5 个分类交。文本设计包含正常语速和快速语速。邀请了来自中国不同口音区域的 2078 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 7 个高保真麦克风（44.1kHz，16bit，1211H*7 个）、安卓系统手机 1 部（16kHz，16bit，1211H）、苹果系统手机 1 部（16kHz，16bit，1211H）进行录制。

AISHELL-2019A-EVAL 随机抽取 2066 人，每人从点位 2、5 和 7（高保真 44.1kHz，16bit）各抽取一句，覆盖所有车载环境与语速。

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，文本正确率达到 100%。

2 产品目录结构

2.1 目录结构

数据目录结构	
数据目录结构	
AISHELL-2021A-EVAL.pdf	(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)
└─SPEECHDATA	(数据文件夹)
└─P2	(录音通道编号)
V0001P20021.wav	(音频文件)

图表 2-1-1 数据目录结构

2.2 命名规则

<CORPUS>/<USAGE>/<POINT_ID>/<WAV_ID>

e.g. AISHELL-20021A-EVAL/SPEECHDATA/P2/V0001P20021.wav

目录	内容	备注
CORPUS	AISHELL-2021A-EVAL	数据库名称编号
USAGE	SPEECHDATA	文件夹名称
POINT_ID	P2	录音位 ID
WAV_ID	V0001P20021.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车载、无人驾驶领域的应用，语料在 5 个领域中选定。按照规则处理语料池。

序号	领域
1	语音控制
2	通讯命令词
3	多媒体
4	POI
5	平衡语句

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。

内容	分配量 (句数)	句子编号
语音控制	35	0021~0055
语音控制 (快语速)	35	0056~0090
通讯命令词	30	0091~0120
通讯命令词 (快语速)	30	0121~0150
多媒体	35	0151~0185
多媒体 (快语速)	35	0186~0220
通讯命令词	25	0221~0245
POI	36	0246~0281
POI (快语速)	36	0282~0317
平衡语句	200	0318~0517
合计	497	

图表 3-2-1 文本结构

AISHELL-2021A-EVAL 数据集文本随机抽取自 0021 到 0150 句，且不重复。

4 录制环境

4.1 录制现场

按照 3 种车速（低速、中速、高速），3 种开车窗类型（关闭、开 1/3 副驾、开 1/3 驾驶员），2 种音乐（开、关），2 种空调环境（关、开）均衡设计。共 28 个场景。高速情况下不含开窗场景。

低速：车速 < 40km/h

中速：40km/h ≤ 车速 < 80km/h

高速：80km/h ≤ 车速 < 120km/h

场景 1	低速-窗关-音关-调关	场景 13	中速-窗关-音关-调关	场景 25	高速-音关-调关
场景 2	低速-窗关-音关-调开	场景 14	中速-窗关-音关-调开	场景 26	高速-音关-调开
场景 3	低速-窗关-音开-调关	场景 15	中速-窗关-音开-调关	场景 27	高速-音开-调关
场景 4	低速-窗关-音开-调开	场景 16	中速-窗关-音开-调开	场景 28	高速-音开-调开
场景 5	低速-副开-音关-调关	场景 17	中速-副开-音关-调关		
场景 6	低速-副开-音关-调开	场景 18	中速-副开-音关-调开		
场景 7	低速-副开-音开-调关	场景 19	中速-副开-音开-调关		
场景 8	低速-副开-音开-调开	场景 20	中速-副开-音开-调开		

场景 9	低速-驾开-音关-调关	场景 21	中速-驾开-音关-调关		
场景 10	低速-驾开-音关-调开	场景 22	中速-驾开-音关-调开		
场景 11	低速-驾开-音开-调关	场景 23	中速-驾开-音开-调关		
场景 12	低速-驾开-音开-调开	场景 24	中速-驾开-音开-调开		

图表 4-1-1 场景列表

4.2 录制设备

录制设备包括高保真麦克风和录音机、安卓手机、苹果手机。高保真录音机存储格式为 44.1kHz、16bit 单声道，安卓手机和苹果手机存储格式为 16kHz、16bit 单声道。AISHELL-2021A-EVAL 抽取的为高保真数据。

4.3 录制方法

车内环境下布置 7 个点位，包含 1 个近讲点位，6 个功能点位。功能点位固定在驾驶员正前方、后视镜、中控台等；近讲点位与发言人距离 25 厘米。发音人以讲话正常音量，分别按照提示，使用正常语速，快速语速，朗读录音文本。

点位 1 与 2 平行相距 5 厘米；点位 3 与 4 平行相距 5 厘米。AISHELL-2021A-EVAL 数据集共抽取 2、5 和 7 三个点位数据。

编号	位置	描述
点位 1	后视镜	高保真麦克风
抽取点位 2	后视镜	高保真麦克风
点位 3	副驾驶正上方把手	高保真麦克风
点位 4	副驾驶正上方把手	高保真麦克风
抽取点位 5	中控大屏正上方	高保真麦克风
点位 6	副驾驶正前方	高保真麦克风
抽取点位 7	头戴	头戴高保真麦克风
点位 8	距发言人 25 厘米，副驾驶右前方	苹果手机
点位 9	距发言人 25 厘米，副驾驶正前方	安卓手机

图表 4-3-1 录音点位分布

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，“NBA”、“UFO”。
- 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) AISHELL-ASR0031 中，快语速的文本前后用符号“@@@”标记。如，“@@@ 烦请把温度调为十九度 @@@”。录音环境设定的文本前后用“#####”标记。如，“##### 音乐开空调开#####”，录音环境设定文本为选读语句。

6 发音人信息

Copyright

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。

录音人编号	性别	口音区域	年龄区间	场景环境
V0001	男	北方	A	低速-窗关-音关-调关

图表 6-1-1 基本信息表例

录音人编号：每个发言人只能参加一次录制。

性别：男-男性；女-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(25-35 岁)、B(36-45 岁)。

场景环境：记录录制场景环境。场景列表见图表 4-1-1。

6.2 发音人结构特征

6.2.1 性别比例

数据库总人数为 2078 人，男 1469 人，女 609 人。AISHELL-2021A-EVAL 总人数为 2066 人，其中男 1464 人，女 602 人。

性别	男性	女性	合计
比例	70.7%	29.3%	100%

图表 6-2-1 性别比例

6.2.2 年龄比例

A (25-35 岁) 1512 人；B (36-46 岁) 566 人。

年龄区间	年龄段	人数	比例	男性	女性
A	25-35 岁	1512	73%	1099	413
B	36-45 岁	566	27%	370	196
合计		2078	100%	1469	609

图表 6-2-2 年龄比例

6.2.3 方言区域比例

北方 1450 人；南方 541 人；其它 9 人。

区域	人数	比例
北方	1461	70.3%
南方	608	29.3%
其它	9	0.4%
合计	2078	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。



Copyright