

语音唤醒词数据库说明书

AISHELL-2019B-EVAL

目录

1 产品概述.....	3
2 产品目录结构.....	3
2.1 目录结构.....	3
2.2 命名规则.....	3
3 文本设计.....	4
4 录制环境.....	4
4.1 录制环境.....	4
4.2 录制设备.....	5
4.3 录制方法.....	5
5 标注转写规范.....	7
6 发言人信息.....	7
7 版权声明.....	7

1 产品概述

AISHELL-2019B-EVAL 语音数据库共 1334720 条 437.67 小时。录音语言，中文；录音地区，中国。录音文本为“你好，米雅” “hi, mia” 唤醒词。邀请 86 名发言人参与录制。录制过程在真实家居环境中，设置 7 个录音位，每个麦克风拾音语音时长 4.51 小时。包括 6 个圆形 16 路 PDM 麦克风阵列录音板做远讲拾音(16kHz, 16bit)、1 个高保真麦克风用做近讲拾音(44.1kHz, 16bit)。

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，字正确率 100%。可用于远讲声纹识别、语音唤醒识别等研究使用。

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-2019B-EVAL.pdf	(数据库简介)
spk_info.xlsx	(录音人信息)
└─SPEECHDATA	(数据文件夹)
─wave	(音频文件夹)
─text	(文本文件夹)

图表 2-1-1 目录结构

2.2 命名规则

AISHELL-2019B-EVAL\SPEECHDATA\wav

e.g. AISHELL-2019B-EVAL\SPEECHDATA\wav\001\001_1_recorded0_001_normal.wav

名称	内容	备注
<speakerID>	001-089	发言人编号
<wavID>.wav	<speakerID>_<pointID>_<micID>_<textID>_<speed>.wav	音频编号

图表 2-2-1 路径命名规则

<speakerID>_<pointID>_<micID>_<textID>_<speed>.wav

e.g. 001_1_recorded0_001_normal.wav

名称	内容	备注
<pointID>	1-6: 录音板; 7: 近讲高保真麦克风	点位
<micID>	recorded0-recorded15	阵列麦克风编号
<textID>	001-160	文本编号
<speed>	slow 慢速, normal 中速, fast 快速	语速

图表 2-2-2 音频文件命名规则

3 文本设计

本数据集定义为唤醒词语音数据集，唤醒词有两种，分别为：

唤醒词 1	你好，米雅
唤醒词 2	hi, mia

图表 3-1-1 唤醒词

文本设计分布如下表所示。textID001 到 120 为安静环境，121 到 160 为噪音环境，每个发言人噪音环境的噪音内容详见发言人信息表 spk_info。

textID	环境	内容
001 到 020	安静	唤醒词 1 中速
021 到 040		唤醒词 2 中速
041 到 060		唤醒词 1 快速
061 到 080		唤醒词 2 快速
081 到 100		唤醒词 1 慢速
101 到 120		唤醒词 2 慢速
121 到 140	噪音	唤醒词 1 中速
141 到 160		40 人音乐 40 人电视

图表 3-1-2 文本分布

4 录制环境

4.1 录制环境

1. 录制场景为真实家居场景，场景内包括基本家居用品、家电桌椅。

2. 录音人站在指定位置。

4.2 录制设备

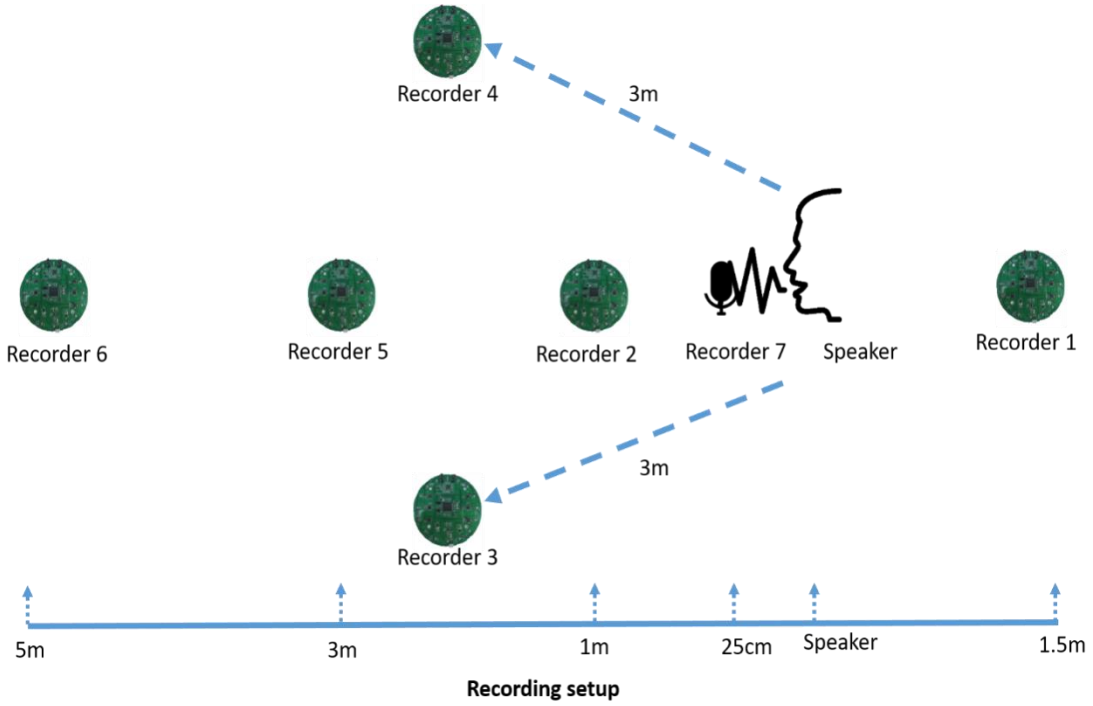
录制设备包括高保真麦克风和 16 路 PDM 麦克风录音开发板。

4.3 录制方法

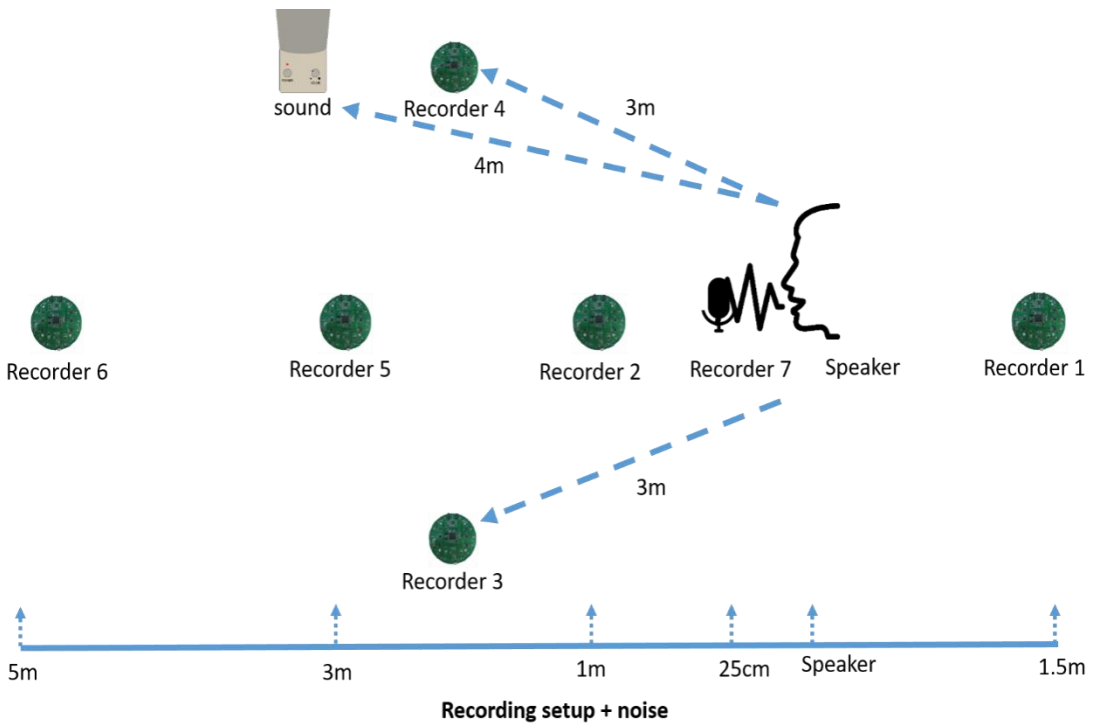
在固定场景环境中分别在 7 个不同位置（相对录音人）放置不同的拾音设备。唤醒词的录制语速设计为三种，正常、慢速、快速，分别体现在每一位录音人的数据中。以电视、音乐两种噪音源进行模拟场景的加噪音工作。录音点位的信息如下表：

序号	距离 (m)	高度 (m)	设备
1	1.5	1	录音板
2	1	1	录音板
3	3	1	录音板
4	3	1	录音板
5	3	1	录音板
6	5	1	录音板
7	0.25	1	高保真麦克风

图表 4-3-1 录音点位



图表 4-3-2 不加噪的拾音示意图



图表 4-3-3 加噪的拾音示意图

5 标注转写规范

数据标注内容与音频一致，要求如下：

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. 句中出现的英文按照发音写出单词，如“hi, mia”。
3. 标注内容的完整性要与实际发音一致，不得删减。

6 发言人信息

发言人信息 `spk_info` 记录内容包括任务编号、性别、年龄、噪音内容。如图表所示：

任务编号	性别	年龄	噪音内容
001	男	20	音乐

图表 4-3-4 发言人信息表示例

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。

性别：男，女。

年龄：发言人参加录制时的年龄。

噪音内容：电视，音乐。

本数据集总录音人数为 86 人，口音区域均属于北方。

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。

