

多人对话语音数据库

AISHELL-ASR1056



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 数据录制环境.....	2
2.1 录制环境.....	2
2.2 录制设备.....	2
2.3 录制方法.....	2
3 语音数据转写.....	2
4 数据文件目录.....	3
5 版权声明.....	4



希尔贝壳
A I S H E L L

Copyright

1 产品概述

AISHELL-ASR1056 多人对话语音数据库共 267 小时，录音语言为中文，语料形式为说明会、总结会、专题会等。录制过程为安静室内环境，录音人通过线阵等设备，以自然对话形式进行录制。此数据库可用于声纹识别、语音识别等研究。

2 数据录制环境

2.1 录制环境

安静室内，无回音，无其他人说话。

2.2 录制设备

16 麦线阵、手机、16 麦圆阵、头戴麦克风等。数据存储格式为 16kHz 16bit wav 文件。

2.3 录制方法

录音人在安静室内，给定主题，多人自然对话。

3 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。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”。注：汉字间不要有空格。中英混合情况，英文之间要加空格。比如字母和字母间，字母和单词间，单词和单词间都要加空格。汉字和英文间不加空格。

- 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) 重叠音，不同的说话人内容标注在对应说话人层，例如有四个说话人的会议，就有四个说话人层，每个说话人层只标注对应的说话人的说话内容，即便有说话重叠，也只标注这个说话人的，重叠说话人的声音标注在对应的说话人层。

细分场景参照如下：

- 多个人说话时，在重叠处将两个人的声音都进行标注并用 & 符号进行区分；
- 多人重叠，其中一人整句无法听清时，则该人当背景音忽略不标；
- 多人重叠，其中一人说整句话但一部分无法听清时，则只标注听清的部分；
- 多人重复，非主说话人的语气词。语气词：直接标注嗯，啊，呵等，{针对主说话人，出现了非主说话人的清楚语气词需要标注，格式为&嗯&、&啊&；若主说话人说的语气词，直接正常标注就可以，无需加“&”}[语气词都要有口字旁,除了诶]。

- 7) 特殊符号标注，详见下表：

序号	标签	说明
1	<sil>	说话人犹豫、拖延时间、停顿、拖长音
2	<->	发出半音卡壳
3	<\$>	笑声
4	<_>	切音，开头或者结尾句子截断
5	<%>	咳嗽
6	<#>	人声杂音，如吸气、打哈欠
7	&	重叠音

4 数据文件目录

数据目录树	
数据目录结构	(数据库简介)
AISHELL-ASR1056.pdf	

└─TextGrid	(文本说明文件夹)
─20200813_S_R002S05L01.TextGrid	(标注文本)
└─SPEECHDATA	(数据文件夹)
─20200813_S_R002S05L01.wav	(音频文件)

5 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



Copyright