

# 中文普通话语音合成数据库



AISHELL-TTS0006  
希尔贝壳  
产品说明书  
A I S H E L L

Copyright

## 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
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.....	3
3.2.1 性别比例.....	3
4 数据录制环境.....	4
4.1 录制环境.....	4
4.2 录制设备.....	4
4.3 录制方法.....	4
5 标注转写规范.....	5
6 产品目录结构.....	6
6.1 目录结构.....	6
6.2 命名规则.....	6
7 版权声明.....	6

# 1 产品概述

AISHELL-TTS0006 中文普通话语音合成数据库共 654 小时 65.7 万条。录音文本涉及智能家居、无人驾驶、工业生产等 15 个领域。邀请 1928 名来自中国不同口音区域发音人参与录制。录制过程在安静室内环境中，使用高保真麦克风（44.1kHz，16bit）。AISHELL-TTS0006 的音频数据来自 AISHELL-ASR0009 和 AISHELL-ASR0010 数据库。

此数据库经过专业语音校对人员对数据做了拼音和声调的转写标注，并通过严格质量检验，文本准确率在 97% 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家居、无人驾驶、工业生产等领域的应用，语料在 15 个领域中选定分别以三个不同内容采集。

内容一		内容二		内容三	
序号	领域	序号	领域	序号	领域
1	家居家电名称和控制命令	1	家居	1	机器人
2	POI (地理信息)	2	POI	2	空调
3	音乐类 (语控)	3	音乐类 (语控)	3	电视
4	数字类 (语控)	4	数字类 (语控)	4	智能音响
5	电视、电影名称	5	电视名称	5	控制器
6	财经	6	财经-政策	6	台灯
7	科技	7	财经-地产	7	数字类
8	体育	8	科技-国际	8	音乐类
9	娱乐	9	科技-国内	9	电视、影视类
10	时事新闻	10	体育	10	电台类
11	英文拼读	11	娱乐-欧美	11	时事内容类
		12	娱乐-港台	12	领域
		13	时事新闻	13	机器人
		14	综合	14	空调

		15	英文拼读	15	电视
--	--	----	------	----	----

图表 2-1 语料池内容

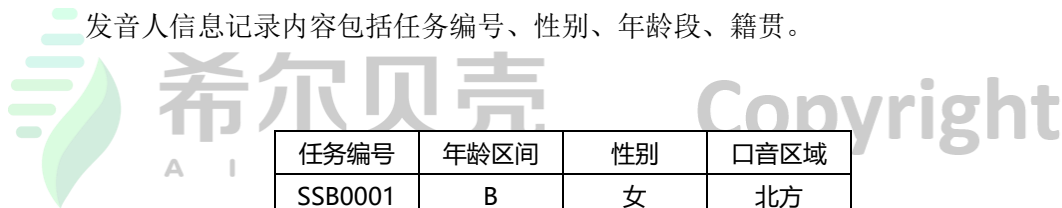
## 2.1.2 语料池处理

- 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。
- 统一格式。

# 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、年龄段、籍贯。



任务编号	年龄区间	性别	口音区域
SSB0001	B	女	北方

图表 3-1 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。

**年龄区间：**A(18 岁以下)、B(18-25 岁)、C(26-45 岁)、D(45 岁以上)。

**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。

## 3.2 发音人结构特征

### 3.2.1 性别比例

数据库总人数为 1928 人，男 828 人，女 1100 人。

性别	男性	女性	合计
比例	43%	57%	100%

图表 3-2-1 性别分布

## 4 数据录制环境

### 4.1 录制环境

安静室内，不包括明显的其他人说话声音及其他噪音，无回音。发言人按照正常语速，朗读录音文本。

### 4.2 录制设备

录制设备包括高保真麦克风，本数据库数据存储格式为 44.1kHz、16bit 单声道。

### 4.3 录制方法

发音人距离高保真麦克风 20 厘米，以讲话正常音量，正常语速，朗读录音文本。Android 系统手机与 iOS 系统手机分别与麦克风间隔 20 厘米布置。



图表 4-3 录制示意图

## 5 标注转写规范

文本层数据转写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。

拼音层数据转写根据听到的实际发音来标注拼音和声调，准则如下。

- 1) 1, 2, 3, 4 分别代表拼音声调，5 代表轻声。例如：  
SSB00610358.wav 到 dao4 时 shi2 我 wo3 会 hui4 联 lian2 系 xi4 你 ni3  
SSB00610502.wav 天 tian1 赐 ci4 良 liang2 缘 yuan2 家 jia1 具 ju4 商 shang1 场 chang3

2) 两个三声和三个三声的变调。如果是两个三声连在一起的情况，多数第一个音调会降为二声，例如：你好吗 ni2 hao3 ma5、洗澡 xi2 zao3、缓缓 huan2 huan3 但不是所有的音调都会降调，所以标注的内容都要以录音为准，不能以平时自己的读音做判断，需要根据实际语音读的来标注，听到的语音是几声就标几声。

- 3) “一” “不” 变调；

“一”：本调为一声，在单念时即本调一声。

“不”：本调为四声，在单念或在非四声音节处念本调四声。

“一” “不” 在四声前读二声，例如：

“一” 在四声前：一 yi2 半 ban4、一 yi2 道 dao4、一 yi2 律 lv4；

“不” 在四声前：不 bu2 必 bi4、不 bu2 但 dan4、不 bu2 错 cuo4；

“一” 在非四声前读四声的情况，例如：

“一”在一声前：一 yi4 般 ban1、一 yi4 些 xie1、一 yi4 朝 zhao1

“一”在二声前：一 yi4 齐 qi2、一 yi4 旁 pang2、一 yi4 同 tong2

“一”在三声前：一 yi4 举 ju3、一 yi4 手 shou3、一 yi4 体 ti3

## 6 产品目录结构

### 6.1 目录结构

数据目录结构	
数据目录结构	
AISHELL-TTS0006.pdf	(数据库简介)
└─DOC	(文本说明文件)
─wav_list.txt	(音频列表)
─content.txt	(语音文本内容)
─spk_info.xlsx	(录音人信息)
└─SPEECHDATA	(数据文件夹)
─SSB0001	(录音人文件夹)
SSB00010001.wav	(音频文件)

图表 6-1-1 数据目录结构

### 6.2 命名规则

**CORPUS/USAGE/SPEAKER\_NUM/SPEECH\_ID**

e. g. AISHELL-TTS0006/SPEECHDATA/SSB0001/SSB00010001.wav

目录名称	内容	备注
<b>CORPUS</b>	AISHELL-TTS0006	语音数据库编号
<b>USAGE</b>	SPEECHDATA	文件夹名称
<b>SPEAKER_NUM</b>	SSB0001	录音人文件夹
<b>SPEECH_ID</b>	SSB00010001.wav	WAV 文件

图表 6-2-1 命名规则

## 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



Copyright