

AISHELL-WakeUp-1

Specification

北京希尔贝壳科技有限公司

Beijing Shell Shell Technology Co.,Ltd

Add: Room 3-621, 6F, Zhongguancun Lifangting No. 1, Shanyuan Road, Haidian District, Beijing 100080, P.R.China

Tel: +86 10 80225006 E-mail: bd@aishelldata.com

Table of content

1 Overview.....	2
2 Database structure.....	3
2.1 File structure.....	3
2.2 Rule of the file path.....	3
3 Text script design.....	4
4 Recording design.....	5
4.1 Recording environment.....	5
4.2 Devices.....	5
4.3 Recording method.....	5
5 Transcription Specification.....	6
6 Information of speakers.....	7
7 Copyright.....	7

1 Overview

AISHELL-WakeUp-1 database contains **1,561.12** hours speech data, including 3,936,003 wake-up

words speech files.

- ⊙ Database language: Chinese and English
- ⊙ Recording area: China
- ⊙ Wake-up words for recording: “Hi mia” and the Chinese of “你好，米雅”
- ⊙ Speakers: 254 participants
- ⊙ Environment: Real home environment
- ⊙ Device setup: 7 different positions are set for recording, including:
 - 1) Six 16-channel circular microphone arrays (16kHz,16bit) for the far-field recording;
 - 2) One Hi-Fi microphone for the close-talk recording (44.1kHz,16bit).

AISHELL-WakeUp-1 database was transcribed by the professional speech annotators with high QA process, and the accuracy rate of word is 100%, which could be used in research of voiceprint recognition, wake-up words recognition and so on.

2 Database structure

2.1 File structure

File Folder	
Folder AISHELL-WAKEUP-1.pdf spk_info.xlsx └─SPEECHDATA └─speech	(Database specification) (Speaker information) (Data folder) (Wake-up words speech)

Table 2-1-1 File structure

2.2 Rule of the file path

AISHELL-WAKEUP-1\SPEECHDATA\speech\wav\<speakerID>\<wavID>.wav

e.g. AISHELL-WAKEUP-1\SPEECHDATA\speech\wav\145\145_1_recorded0_001_normal.wav

Field	Value	Field meaning
-------	-------	---------------

<speakerID>	145-398	Speaker ID
<wavID>.wav	<speakerID>_<pointID>_<micID>_<textID>_<speed>.wav	audio ID

Table 2-2-1 Rule of the file path

AISHELL-WAKEUP-1\SPEECHDATA\speech\text\<speakerID>.txt

e.g. AISHELL-WAKEUP-1\SPEECHDATA\speech\text\145.txt

Field	Value	Field meaning
<speakerID>	145-398	Speaker ID

Table 2-2-2 Rule of the file path

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

e.g. 145_1_recorded2_001_normal.wav

Field	Value	Field meaning
<pointID>	1-6: circular microphone arrays ID ; 7: Hi-Fi Mic ID	Point position ID
<micID>	recorded2-recorded17	Microphone ID of an array
<textID>	001-160	Text ID
<speed>	Slow, normal, fast	Speaking speed

Table 2-2-3 Rule of the audio file name

3 Text script design

AISHELL-WakeUp-1 database contains two kinds of wake-up words:

Wake-up word 1	你好，米雅 (Chinese of “ni hao, mi ya”)
Wake-up word 2	hi, mia

Table 3-1-1 Wake-up word

The distribution of the text script design for AISHELL-WakeUp-1 database is following:

TextID: 001 to 120 is in noisy environment

TextID: 121 to 160 is in clear environment

The noise information of each speaker is included in the speaker information form (spk_info):

TextID	Environment	Wake-up word and speed
--------	-------------	------------------------

001 to 020	Noisy environment	Wake-up word 1 normal speed
021 to 040		Wake-up word 2 normal speed
041 to 060		Wake-up word 1 fast speed
061 to 080		Wake-up word 2 fast speed
081 to 100		Wake-up word 1 slow speed
101 to 120		Wake-up word 2 slow speed
121 to 140	Clear environment	Wake-up word 1 normal speed
141 to 160		Wake-up word 2 normal speed

Table 3-1-2 Text script distribution

4 Recording design

4.1 Recording environment

1. The recording environment is real home environment, including the electrical appliance, table and chair, and so on.
2. The speaker stands in a fixed position for recording.

4.2 Devices

The recording devices includes Hi-Fi microphone and PDM microphone array.

4.3 Recording method

The recordings of each speaker could be cataloged into three subset according to the speaking speed (normal speed, fast speed and slow speed). We simulated real smart home scenes by adding noise sources such as TV, music, and background noises to the room. The room setting is shown in speaker information form (spk_info). The Hi-Fi microphone is 25 cm away from the speaker. The circular microphone arrays are placed around the person with a distance including 1m, 3m and 5m from the person. The noise source is randomly placed close to one of the microphone arrays for each speaker. The room setting and position information is shown below:

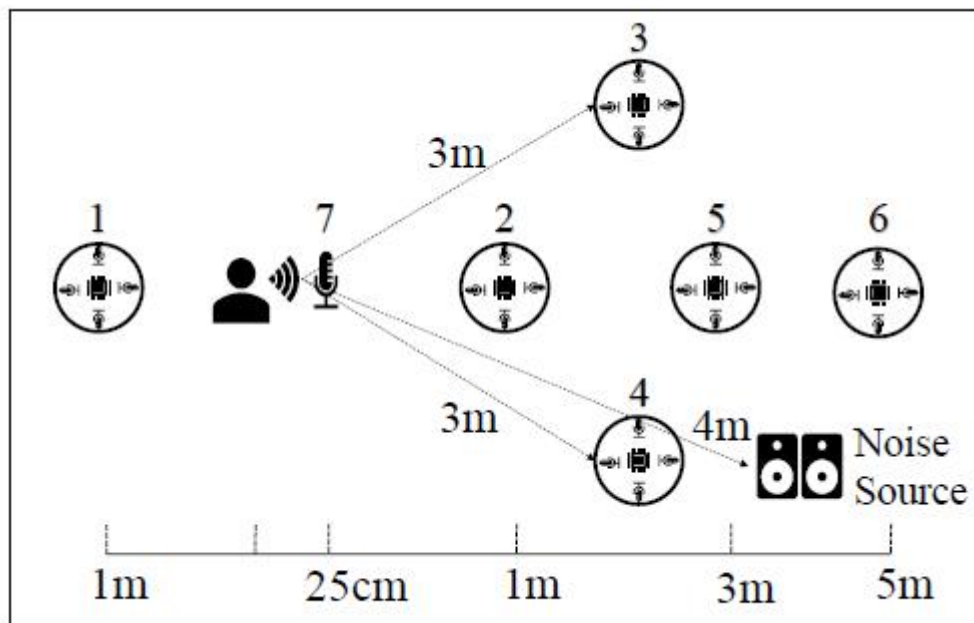


Table 4-3-1 The setup of the recording environment

Mic Position ID	Distance (m)	Height(m)	Device
1	1.5	1	Microphone array
2	1	1	Microphone array
3	3	1	Microphone array
4	3	1	Microphone array
5	3	1	Microphone array
6	5	1	Microphone array
7	0.25	1	Hi-Fi Microphone

Table 4-3-2 Recording position

5 Transcription Specification

1. The content of the transcription text is 100% the same as the audio content, redundant, lacking or wrong text is not allowed
2. If there is English word in the text, it shall be transcribed according to the pronunciation, e.g. "hi, mia".
3. The completeness for the transcription and real pronunciation in the audio should be the same.

6 Information of speakers

The information of each speaker, which is recorded in spk_info file, contains mission ID, Gender, Age, Noise type, Noise position, see below example:

Mission ID	Gender	Age	Noise type	Noise position
145	Male	25	Cartoon	Position 2

Table 4-3-4 information of speaker

Mission ID: Each speaker has only 1 mission ID, and each mission contains a unique script package for recording. A speaker can only complete one mission.

Gender: Male/Female

Age: The speaker's age at recording time

Noise type: Singing, pure music, TV play, cartoon, fan's noise, sport rebroadcast and variety show.

Noise position: Randomly placed close to one of the microphone arrays for each speaker.

There are 254 speakers in this database, and all of them comes from the northern China (northern Chinese accent).

7 Copyright

All copyright of this document is reserved by AISHELL (Beijing Shell Shell Technology Co., Ltd.). AISHELL owns all the right of the amendment, the update, and final explanation for this document.



Copyright