

算法+数据的开源如何实践 与语音模型评测方案

卜辉 CEO

北京希尔贝壳科技有限公司

Contents

1 人工智能产业链

2 开源赋能人工智能新基建

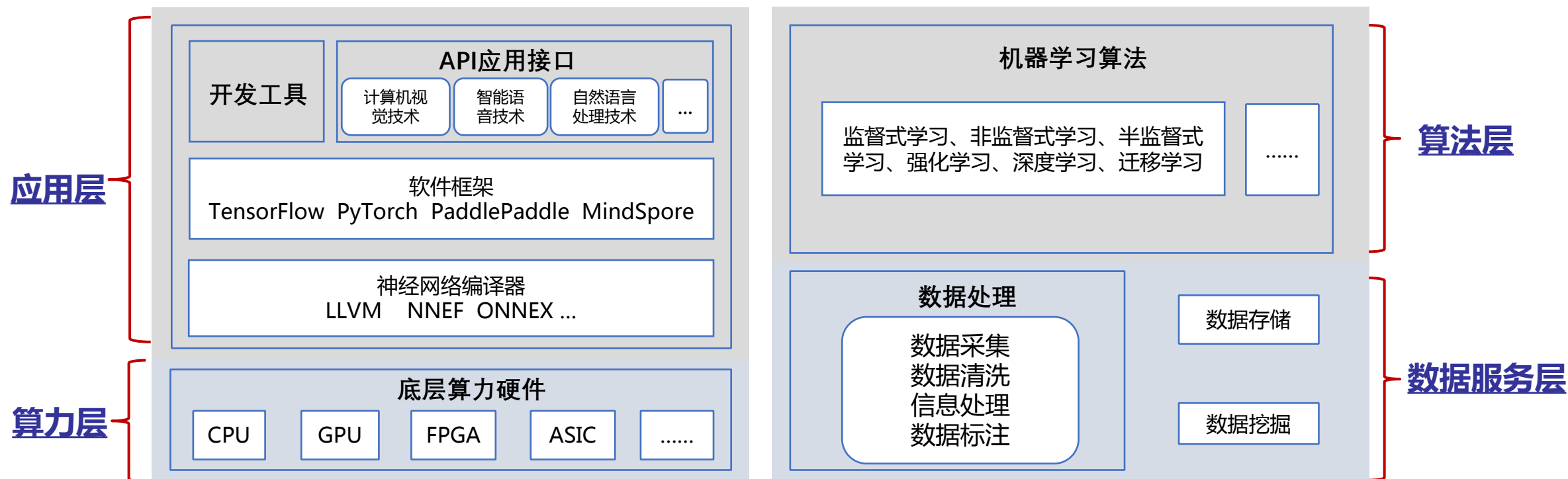
3 语音算法+数据开源方案

4 语音模型的评测方案

5 AISHELL



人工智能产业链架构示意图





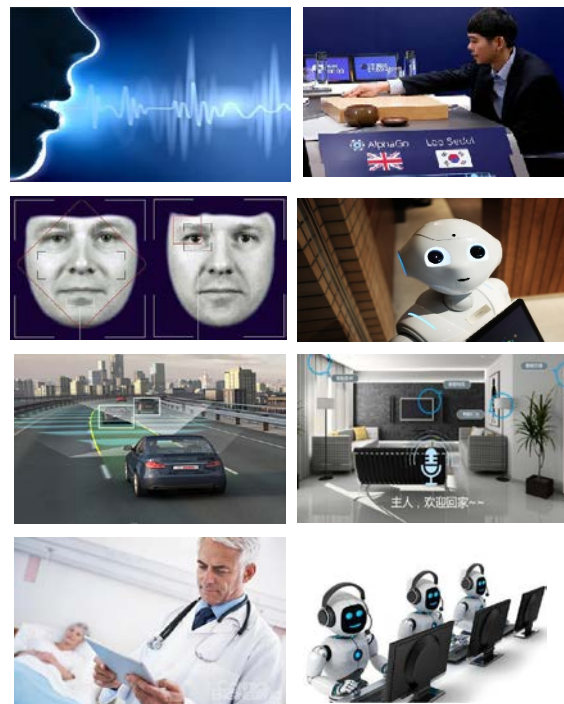
算法

算力

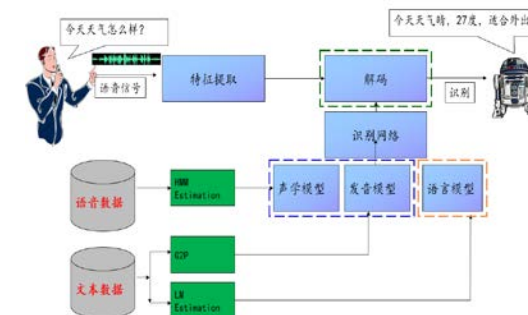
数据

人工智能的成功应用

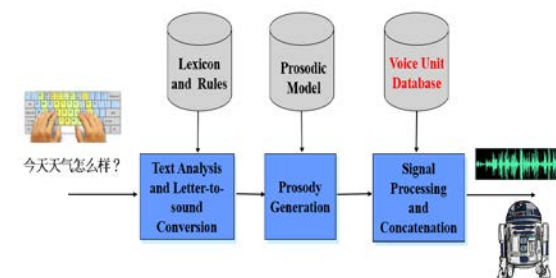
- 语音识别
- 无人驾驶
- 图像识别
- 机器翻译
- 搜索推荐
- 对话系统
- 精准营销
-



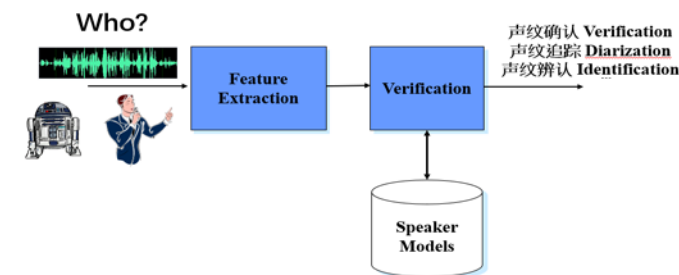
语音识别技术 (ASR)

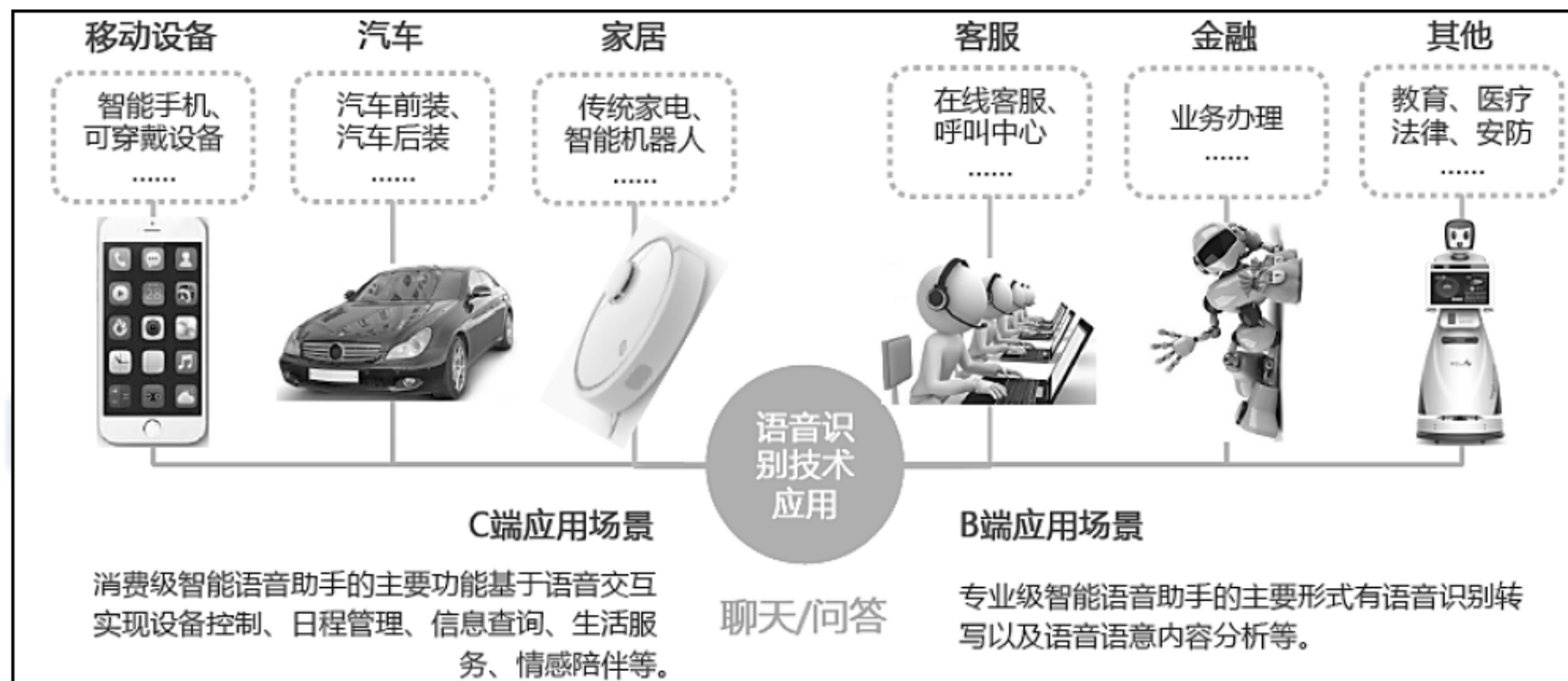
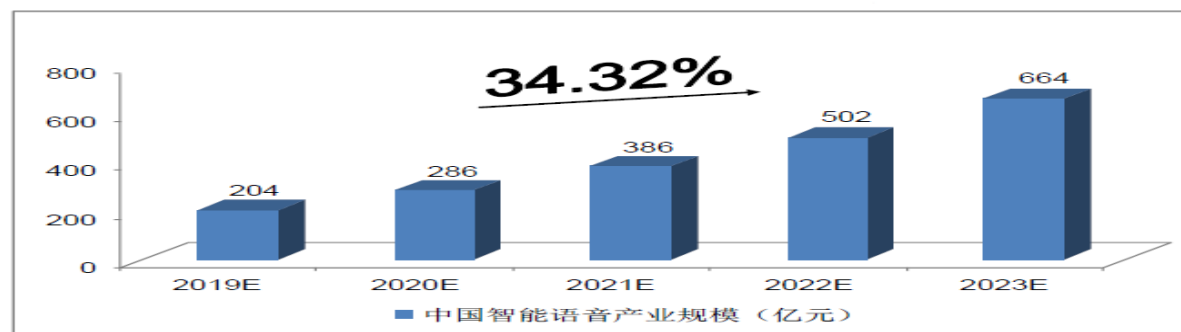


语音合成技术 (TTS)

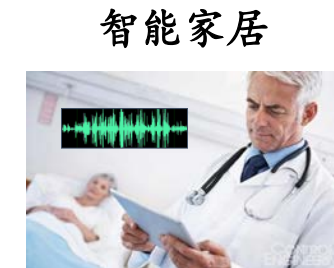


声纹识别技术 (VPR)





数据来源: 中投产业研究院

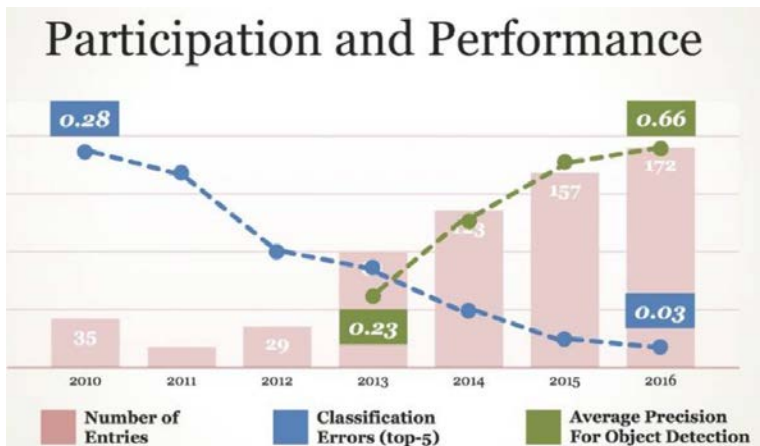


技术价值降低？人工成本降低？市场需求增加？



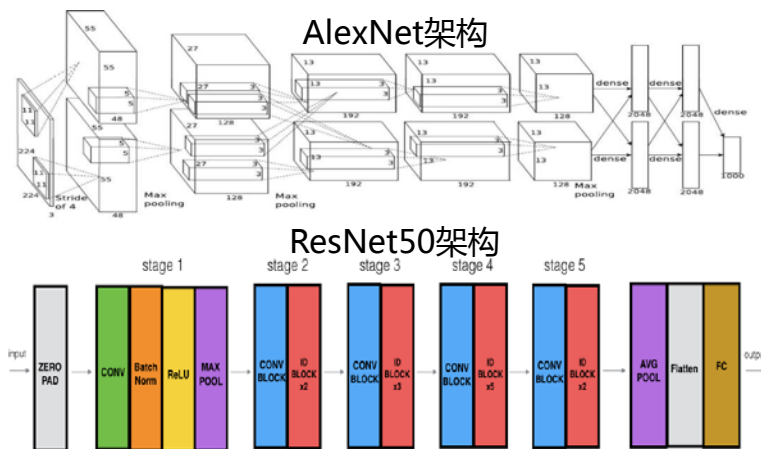
公共数据集开放提供燃料

ImageNet 中含有超过 1500 万由人手工注释的图片，包含超过 2.2 万个类别，从2010年ILSVRC第一届比赛开始，随着参赛队伍的增多，算法逐渐逼近人类识别水平



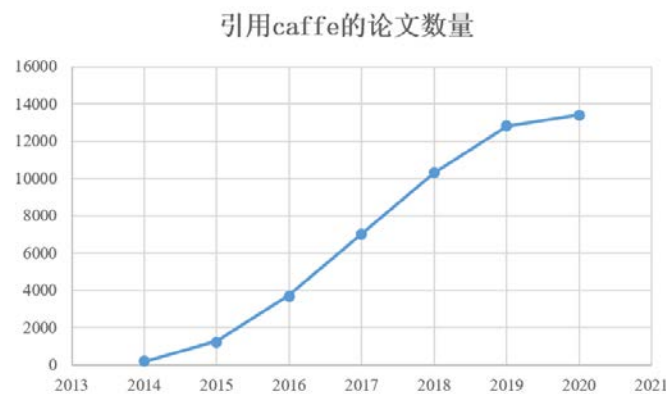
经典算法开源推动技术进步

本轮深度学习的革命兴起于2012年开源的深度卷积神经网络模型AlexNet，之后几年研究人员不断在之前开源算法的基础上进行改进和创新，从AlexNet到ResNet，再到后ResNet时代的最新架构SENet和SKNet，精度大幅提升，应用不断成熟。



关键工具开源加速创新

Caffe是最早开源的深度学习框架之一，帮助研究人员屏蔽底层硬件，专注与算法的设计，极大促进了技术的创新。



算法+数据的开源如何实践 与语音模型评测方案

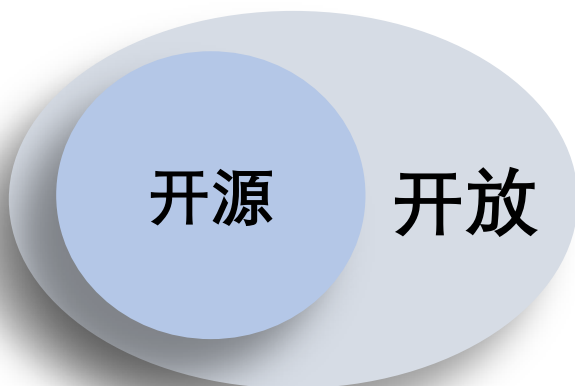
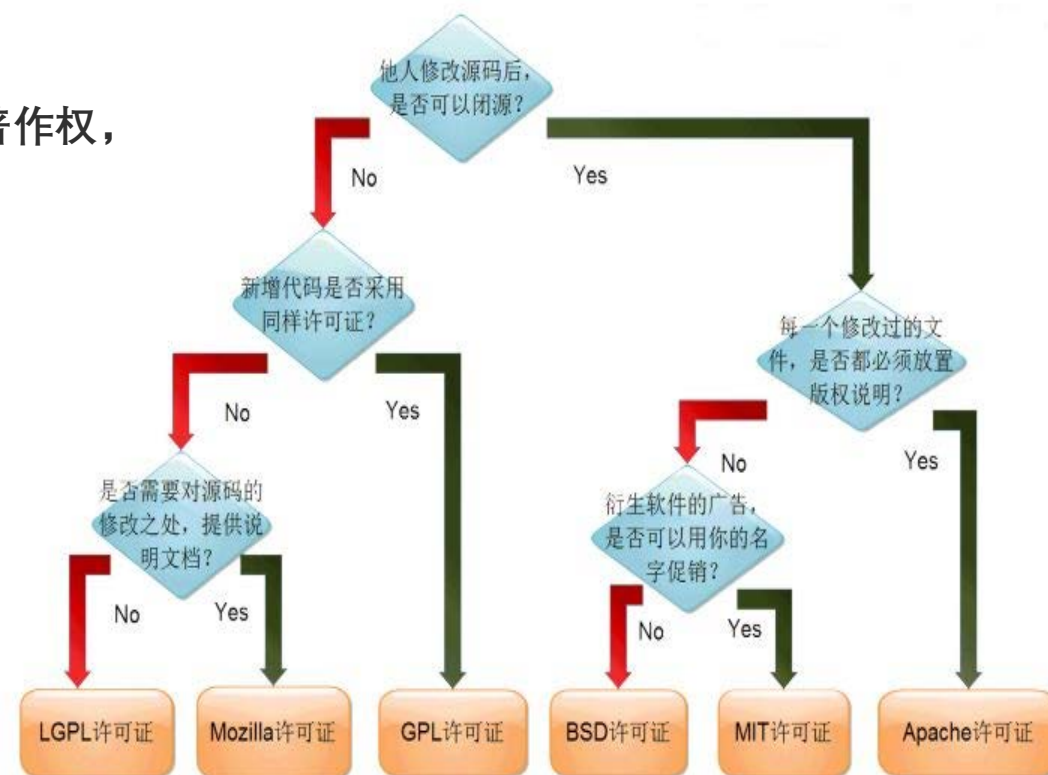


人工智能开放共享，是指个人或机构向社会和用户提供人工智能相关数据、算法、算力、工具等，并授权他人使用的一种行为。——参考盛小平等对开放科学语境下开放共享的定义，指个人或机构向社会或用户提供数据、信息或作品，并授权他人免费使用的一种行为。

企业开源？高校等公益组织？个人？

- 开放：开放不一定开源；
- 开源：开放的途径，指算法、工具等开源；

AL(Apache Licence):
鼓励代码共享和尊重原作者的著作权，
同样允许代码修改，再发布。





语音技术的开源主要是集中在海外的早期开源软件系统及算法，早期的CMU Sphinx、HTK、ISIP、Julius以及近期的Kaldi。但是在各大开源社区目前活跃最多的开发者以中国人为最多，国内的AI开发环境优于国外。目前全球最大的语音技术开发社区Kaldi在国内拥有很多开发者，其Kaldi发起者也是该系统建设者Daniel Povey博士也以首席科学家身份加入国内科技企业小米集团。国内每年由AISHELL（希尔贝壳）组织和CCF语音对话与听觉专委会协办的Kaldi线下交流会也是吸引国内众多参与者加入Kaldi社区。

Toolkit	Programming languages	Development activity	Tutorials and examples	Community	Trained models
CMU Sphinx	Java, C, Python, others	+++	+++	+++	English plus 10 other languages
Kaldi	C++, Python	+++	++	+++	Subset of English
HTK	C, Python	++	+++	++	None
Julius	C, Python	++	++	+	Japanese
ISIP	C++	++	++	+	Digits only



GitHub



聚集全球最多
开放共享开发者的社区



AISHELL-1

OpenSLR

Home Resources

Aishell

Identifier: SLR33
Summary: Mandarin data, provided by Beijing Shell Shell Technology Co.,Ltd
Category: Speech
License: Apache License v.2.0
Downloads (use a mirror closer to you):
[data_aishell.tgz](#) [15G] (speech data and transcripts) Mirrors: [\[China\]](#)
[resource_aishell.tgz](#) [1.2M] (supplementary resources, incl. lexicon, speaker info) Mirrors: [\[China\]](#)
About this resource:
Aishell is an open-source Chinese Mandarin speech corpus published by Beijing Shell Shell Technology Co.,Ltd.

<http://www.openslr.org/33/>

AISHELL-1: An open-source Mandarin speech corpus and a speech recognition baseline

Publisher: IEEE

[Cite This](#)

[PDF](#)

Hui Bu ; Jiayu Du ; Xingyu Na ; Bengu Wu ; Hao Zheng [All Authors](#)

70
Paper
Citations

740
Full
Text Views



Kaldi recipe

s5: a **speech** recognition recipe
v1: a **speaker** recognition recipe



178H
400 Speakers

- Sampling Rate: 16KHz
Sample Format: 16bit
Environment: Indoor



Sample

Netdisk

OpenSLR

License: Apache License v.2.0

<http://www.aishelltech.com/kysjcp>



AISHELL-2

1000H
1991 Speakers

- Sampling Rate: 16KHz
Sample Format: 16bit
Environment: Indoor



Sample

Netdisk

OpenSLR

License: Apache License v.2.0

http://www.aishelltech.com/aishell_2

arXiv.org > cs > arXiv:1808.10583

Search...

Help | Ad

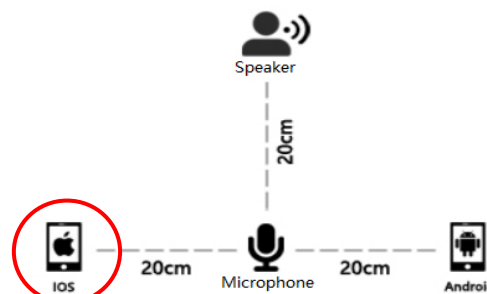
Computer Science > Computation and Language

[Submitted on 31 Aug 2018 (v1), last revised 13 Sep 2018 (this version, v2)]

AISHELL-2: Transforming Mandarin ASR Research Into Industrial Scale

Jiayu Du, Xingyu Na, Xuechen Liu, Hui Bu

AISHELL-1 is by far the largest open-source speech corpus available for Mandarin speech recognition research. It was released with a baseline system containing solid training and testing pipelines for Mandarin ASR. In AISHELL-2, 1000 hours of clean read-speech data from iOS is published, which is free for academic usage. On top of AISHELL-1-2 corpus, an improved recipe is developed and released, containing key components for industrial applications, such as Chinese word segmentation, flexible vocabulary expansion and phone set transformation etc. Pipelines support various state-of-the-art techniques, such as time-delayed neural networks and Lattice-Free MMI objective function. In addition, we also release dev and test data from other channels (Android and Mic). For research community, we hope that AISHELL-2 corpus can be a solid resource for topics like transfer learning and robust ASR. For industry, we hope AISHELL-2 recipe can be a helpful reference for building meaningful industrial systems and products.



Kaldi recipe

speech recognition recipe

[kaldi/egs/aishell2/](http://kaldi.org/egs/aishell2/)

[README.md](#)

AISHELL-2

AISHELL-2 is by far the largest free speech corpus available for Mandarin ASR research.

1. DATA

Training data

- 1000 hours of speech data (around 1 million utterances)
- 1991 speakers (845 male and 1146 female)
- clean recording environment (studio or quiet living room)
- read speech
- reading prompts from various domain: entertainment, finance, technology, sports, control command, place of interest etc.
- near field recording via 3 parallel channels (iOS, Android, Microphone).
- iOS data is free for non-commercial research and education use (e.g. universities and non-commercial institutes)

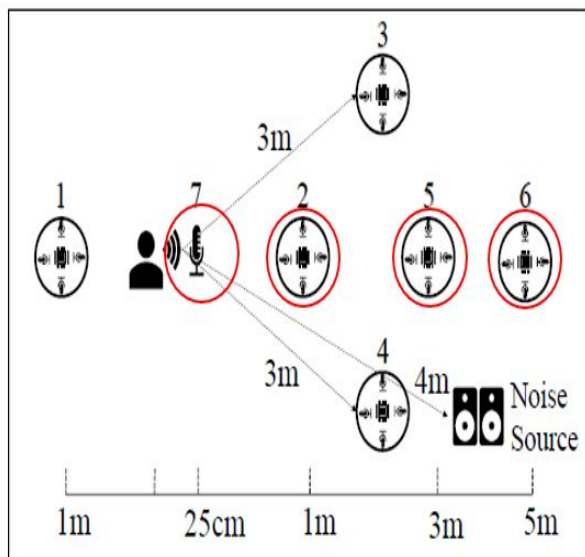




AISHELL-WakeUp-1

HI-MIA

“你好，米雅” “Hi, Mia” Open Source



AISHELL Speaker Verification Challenge 2019

HI-MIA

<http://www.openslr.org/85/>

AISHELL-WakeUp-1

http://www.aishelltech.com/wakeup_data

1561H

340 Speakers

Sampling Rate: 44.1KHz & 16KHz

Sample Format: 16bit

Environment: Indoor



Kaldi recipe

Speaker Verification

Keyword

recognition recipe

master kaldi / egs / hi_mia /

underdogliu HI-MIA update: complementing trial list format and text-independent r...

..	
v1	HI-MIA update: complementing trial list format and text-independent r...
w1	[egs] Add hi_mia/w1 recipe (#4120)
README.txt	[egs] Add HIMIA (openslr-85) recipe speaker verification (#4069)

README.txt

This is repository for HIMIA dataset.

base can be downloaded from openslr:

AISHELL1: <http://www.openslr.org/33/>

HI-MIA: <http://www.openslr.org/85/>

This folder contains two subfolders:

w1: a keyword spotting recipe, using aishell1 asr data for negative data
and aishell hi_mia data as positive data

v1: a speaker verification recipe. More can be found in v1/README.md



AISHELL-3



- Sampling Rate: 44.1KHz
Sample Format: 16bit
Environment: Indoor
- 85H
218 Speakers

Download-Dataset

http://www.aishelltech.com/aishell_3

Download-Paper

<https://arxiv.org/abs/2010.11567>



multi-speaker Text-to-Speech (TTS) systems

AISHELL-3 Corpus: System recipe Baseline System Samples

Arxiv: [2010.11567](https://arxiv.org/abs/2010.11567)

Github Repo: [sos1sos2Sixteen/aishell-3-baseline-fc](https://github.com/sos1sos2Sixteen/aishell-3-baseline-fc)

Dataset Download: www.aishelltech.com/aishell_3

For further questions regarding the dataset: tech@aishelldata.com

Authors

- SHI, Yao (Wuhan University, Duke-Kunshan University)
- BU, Hui (AISHELL)
- XU, Xin (AISHELL)
- ZHANG, Shaoji (AISHELL)
- LI, Ming (Duke-Kunshan University, Wuhan University) ming.li369@dukekunshan.edu

AISHELL-3 V2

INTERSPEECH 2021

The **Interspeech** 2021 Program Committee are pleased to inform you that your paper

Paper ID: 755

Title: AISHELL-3: A Multi-Speaker Mandarin TTS Corpus

has been accepted for presentation at the conference. Please read through the rest of this email carefully



AI SHELL 4

一个通过麦克风阵列实录的八通道中文普通话会议场景语音数据集



120 小时 | 120 Hours
211 场会议 | 211 Meeting Sessions
10个 会议室 | 10 Meeting Rooms
60 人 | 60 Speakers



Speech front-end processing
Speech recognition
Speaker diarization



开源系统
Open Source

该数据集共包含**211**场会议，每场会议4至8人，数据集共**120**小时左右。该数据集旨在促进实际应用场景下多说话人处理的研究。AISHELL-4数据包括了实际会议场景下各种重要特性，例如停顿、重叠、说话人轮转、噪声等。同时数据集提供了准确的音字转写文本及时间戳信息，方便研究者进行诸如[前端处理](#)、[语音识别](#)、[说话人分割](#)等单独任务，并可以进行联合优化。



120H 60 Speaker

Meeting sessions

211

Meeting rooms

10

Recording Device

8-channel

circular microphone array (16kHz, 16-bit);

财经
房产
家居
教育
科技
时尚
时政
体育
游戏
娱乐



<http://www.openslr.org/111/>

OpenSLR

[Home](#) [Resources](#)

AISHELL-4

Identifier: SLR111

Summary: A Free Mandarin Multi-channel Meeting Speech Corpus, provided by Beijing Shell Shell Technology Co.,Ltd

Category: Speech

License: CC BY-SA 4.0

Downloads (use a mirror closer to you):

[train_L.tar.gz](#) [7.0G] (Training set of large room, 8-channel microphone array speech) Mirrors: [\[China\]](#)

[train_M.tar.gz](#) [25G] (Training set of medium room, 8-channel microphone array speech) Mirrors: [\[China\]](#)

[train_S.tar.gz](#) [14G] (Training set of small room, 8-channel microphone array speech) Mirrors: [\[China\]](#)

[test.tar.gz](#) [5.2G] (Test set) Mirrors: [\[China\]](#)

About this resource:

The AISHELL-4 is a sizable real-recorded Mandarin speech dataset collected by 8-channel circular microphone array for speech process dataset consists of 211 recorded meeting sessions, each containing 4 to 8 speakers, with a total length of 120 hours. This dataset aims research on multi-speaker processing and the practical application scenario in three aspects. With real recorded meetings, AISHELL-4 provides rich natural speech characteristics in conversation such as short pause, speech overlap, quick speaker turn, noise, etc. Meanwhile, the speaker voice activity are provided for each meeting in AISHELL-4. This allows the researchers to explore different aspects in meeting processing, ranging from individual tasks such as speech front-end processing, speech recognition and speaker diarization, to multi-modality modeling and joint optimization of relevant tasks. We also release a PyTorch-based training and evaluation framework as a baseline system to promote reproducible research in this field. Generated samples are available [here](#).

You can cite the data using the following BibTeX entry:

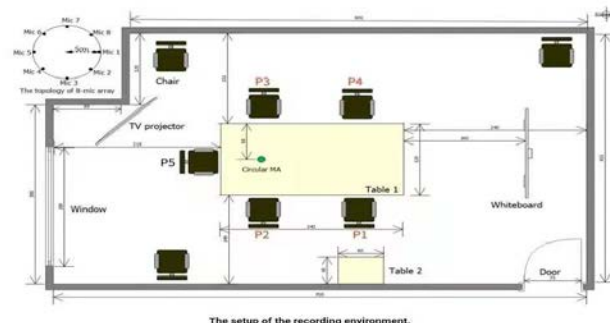
http://www.aishelltech.com/aishell_4

AISHELL-4 多通道中文会议语音数据库

AISHELL-4 Open Source Mandarin Multi-channel Meeting Speech Corpus

AISHELL-4是一个通过麦克风阵列采集的八通道中文普通话会议场景语音数据集。该数据集共包含211场会议，每场会议4至8人，数据集共120小时左右。该数据集旨在促进实际应用场景下多说话人处理的研究。AISHELL-4数据包括了实际会议场景下各种重要特性，例如停顿、重叠、说话人轮转、噪声等。同时数据集提供了准确的音字转写文本及时间戳信息，方便研究者进行诸如前端处理、语音识别、说话人分割等单独任务，并可以进行联合优化。

The AISHELL-4 is a sizable real-recorded Mandarin speech dataset collected by 8-channel circular microphone array for speech processing in conference scenario. The dataset consists of 211 recorded meeting sessions, each containing 4 to 8 speakers, with a total length of 120 hours. This dataset aims to bridge the advanced research on multi-speaker processing and the practical application scenario in three aspects. With real recorded meetings, AISHELL-4 provides realistic acoustics and rich natural speech characteristics in conversation such as short pause, speech overlap, quick speaker turn, noise, etc. Meanwhile, the accurate transcription and speaker voice activity are provided for each meeting in AISHELL-4. This allows the researchers to explore different aspects in meeting processing, ranging from individual tasks such as speech front-end processing, speech recognition and speaker diarization, to multi-modality modeling and joint optimization of relevant tasks. We also release a PyTorch-based training and evaluation framework as baseline system to promote reproducible research in this field.



120 小时 | 120 Hours
211 场会议 | 211 Meeting Sessions
10 个会议室 | 10 Meeting Rooms
60 人 | 60 Speakers



Speech front-end processing
Speech recognition
Speaker diarization



开源系统
Open Source



Open Source



arxiv



Sample

Netdisk

OpenSLR



Readme

Speaker Information



aishell4 System

License: CC BY-SA 4.0



<https://github.com/felixfuyihui/AISHELL-4>

AISHELL-4

This project is associated with the recently-released AISHELL-4 dataset for speech enhancement, separation, recognition and speaker diarization in conference scenario. The project, served as baseline, is divided into five parts, named *data_preparation*, *front_end*, *asr* and *sd*. The Speaker Independent (SI) task only evaluates the ability of front end (FE) and ASR models, while the Speaker Dependent (SD) task evaluates the joint ability of speaker diarization, front end and ASR models. The goal of this project is to simplify the training and evaluation procedure and make it easy and flexible for researchers to carry out experiments and verify neural network based methods.

Setup

```
git clone https://github.com/felixfuyihui/AISHELL-4.git
pip install -r requirements.txt
```

Introduction

- **Data Preparation:** Prepare the training and evaluation data.
- **Front End:** Train and evaluate the front end model.
- **ASR:** Train and evaluate the asr model.
- **Speaker Diarization:** Generate the speaker diarization results.
- **Evaluation:** Evaluate the results of models above and generate the CERs for Speaker Independent and Speaker Dependent tasks respectively.

General steps

1. Generate training data for fe and asr model and evaluation data for Speaker Independent task.
2. Do speaker diarization to generate rttm which includes vad and speaker diarization information.
3. Generate evaluation data for Speaker Dependent task with the results from step 2.
4. Train FE and ASR model respectively.
5. Generate the FE results of evaluation data for Speaker Independent and Speaker Dependent tasks respectively.
6. Generate the ASR results of evaluation data for Speaker Independent and Speaker Dependent tasks respectively with the results from step 2 and 3 for No FE results.
7. Generate the ASR results of evaluation data for Speaker Independent and Speaker Dependent tasks respectively with the results from step 5 for FE results.
8. Generate CER results for Speaker Independent and Speaker Dependent tasks of (No) FE with the results from step 6 and 7 respectively.

Contributors



AISHELL联合西北工业大学、中国科学技术大学、微软合著的论文

INTERSPEECH 2021接收

Cornell University

arXiv.org > cs > arXiv:2104.03603

Computer Science > Sound

[Submitted on 8 Apr 2021 (v1), last revised 18 Jun 2021 (this version, v2)]

AISHELL-4: An Open Source Dataset for Speech Enhancement, Separation, Recognition and Speaker Diarization in Conference Scenario

Yihui Fu, Luyao Cheng, Shubo Lv, Yukai Jv, Yuxiang Kong, Zhuo Chen, Yanxin Hu, Lei Xie, Jian Wu, Hui Bu, Xin Xu, Jun Du, Jingdong Chen

In this paper, we present AISHELL-4, a sizable real-recorded Mandarin speech dataset collected by 8-channel circular microphone array for speech processing in conference scenario. The dataset consists of 211 recorded meeting sessions, each containing 4 to 8 speakers, with a total length of 118 hours. This dataset aims to bridge the advanced research on multi-speaker processing and the practical application scenario in three aspects. With real recorded meetings, AISHELL-4 provides realistic acoustics and rich natural speech characteristics in conversation such as short pause, speech overlap, quick speaker turn, noise, etc. Meanwhile, the accurate transcription and speaker voice activity are provided for each meeting in AISHELL-4. This allows the researchers to explore different aspects in meeting processing, ranging from individual tasks such as speech front-end processing, speech recognition and speaker diarization, to multi-modality modeling and joint optimization of relevant tasks. Given most open source dataset for multi-speaker tasks are in English, AISHELL-4 is the only Mandarin dataset for conversation speech, providing additional value for data diversity in speech community.

Comments: Accepted by Interspeech 2021

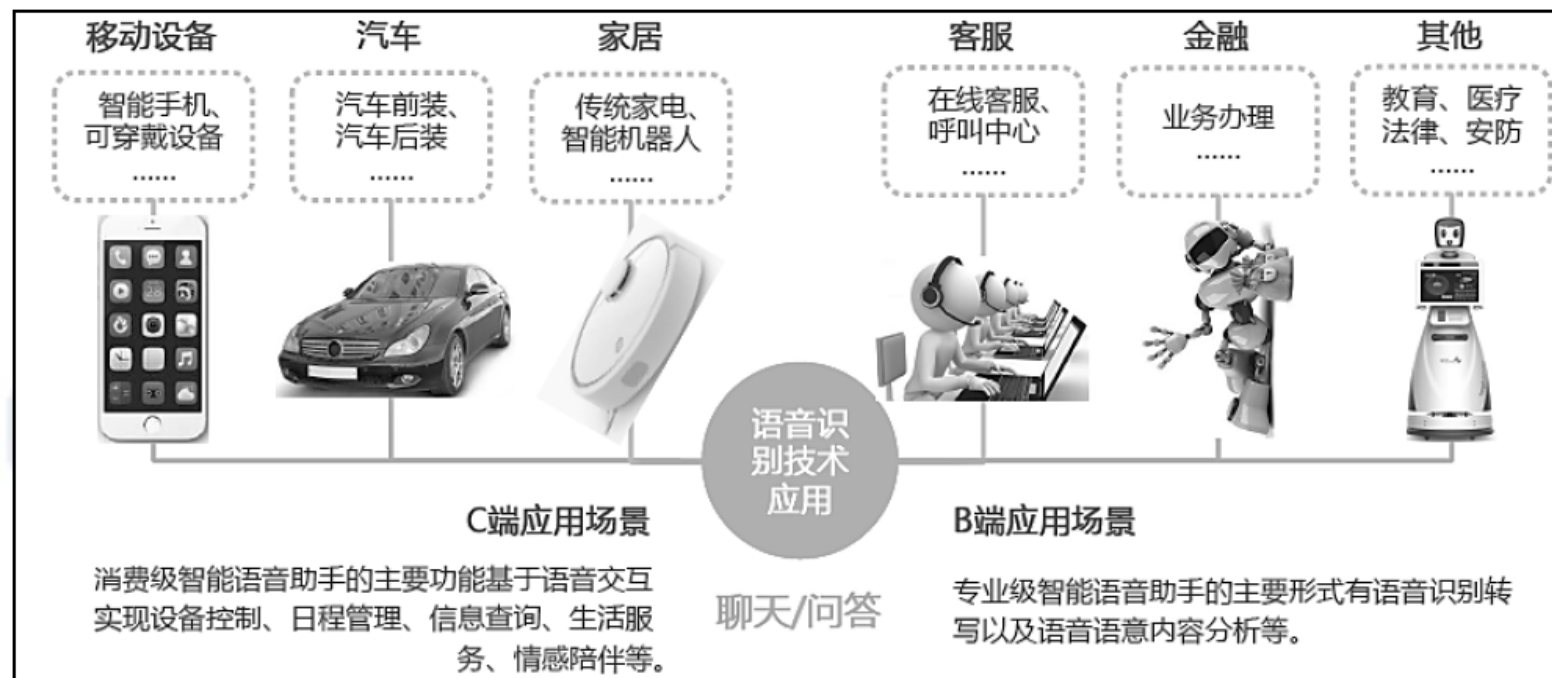
Subjects: **Sound (cs.SD)**; Audio and Speech Processing (eess.AS)

Cite as: **arXiv:2104.03603 [cs.SD]**
(or **arXiv:2104.03603v2 [cs.SD]** for this version)

语音之家@Speech Home



语音识别（Automatic Speech Recognition, ASR）技术的算法模型和高性能硬件环境的结合，技术到产品落地变得越来越可便捷实现。随着整体AI算法环境的提升，ASR做为单一技术也在向多技术融合方案发展，听觉+视觉+自然语言处理已经成为更智能化的趋势。ASR的模型也从单一场景拓展到多场景建模化，ASR的模型由大到高定制再到强匹配需要科学有效的评估，面向场景ASR模型的测试评估数据集成为了必要的基础和重要的评估方式。





语音识别的准确率通过数据集的评估可以体现具体的性能指标，例如句准、字准。面向场景语音识别模型的评测数据集评估规范建设可以推动智能语音技术的规范化，对智能语音产业及用户体验的提升都起到了促进作用。

国外情况：

1. CHiME(Computational Hearing in Multisource Environments) 属于国际语音识别评测中具有高难度挑战性的赛事，其场景复杂、音频源质量不一的数据集来评估语音识别模型的准确率。
2. NIST美国国家标准与技术研究院的评测标准。
3. AWS 的基于字词准确率的评测标准规范。

国内情况：

1. 科大讯飞的语音识别技术准确率参考。
2. SPEECHIO 语音评测结果参考。
3. 2020年中国通信标准化协会立项了行业标准《面向人工智能的数据集质量评估规范》系列规范，包括总体要求、语音数据集、人脸图像数据三个部分，该系列标准旨在支持各个领域人工智能数据的规范化使用，保证数据的高质量需求，引导和规范数据的安全、合规使用，推动人工智能在各个领域与行业的健康发展。



AISHELL-2018A-EVAL

TEST DATA: 5000 utterances from 10 speakers

DEV DATA: 2500 utterances from 5 speakers



TEST 10 speaker
DEV 5 speaker

DATA Information

- Sampling Rate : 16kHz
- Sample Format : 16bit
- Environment : Indoor
- Speech Data Type : PCM
- Channel Number : 1
- Recording Equipment : iOS / Android / High Fidelity Microphone



语音识别实验评测
声纹实验评测

Speech & Speaker Recognition evaluation



TEST&DEV DATA

License: Apache License v.2.0

AISHELL-2019A-EVAL

家居环境近远讲同步语音数据库



50 speakers in the recording

DATA Information

- Total Time : 24.3 Hours
- Sampling Rate : 44.1kHz & 16kHz
- Sample Format : 16bit
- Environment : Indoor
- Speech Data Type : wav
- Recording Equipment : High Fidelity Microphone



语音识别实验评测
声纹实验评测

Speech & Speaker Recognition Evaluation



Sample



Readme



Database

License: Apache License v.2.0

AISHELL-2019B-EVAL

语音唤醒词数据库



86 speakers in the recording

DATA Information

- Total Time : 437.67 Hours
- Sampling Rate : 44.1kHz & 16kHz
- Sample Format : 16bit
- Environment : Indoor
- Speech Data Type : wav
- Recording Equipment : Microphone-Array Board/ High Fidelity Microphone



语音识别实验评测
声纹实验评测

Speech & Speaker Recognition Evaluation



Sample



Readme



Database



Paper

License: Apache License v.2.0

AISHELL-2019C-EVAL

翻译机录制语音数据库



12 speakers in the recording

DATA Information

- Total Time : 31.2 Hours
- Sampling Rate : 44.1kHz & 16kHz
- Sample Format : 16bit
- Environment : Indoor
- Speech Data Type : wav
- Recording Equipment : JoneR Translator/ High Fidelity Microphone



Sample



Readme



Database

License: Apache License v.2.0

AISHELL-2021A-EVAL

电动车载环境智能语音交互数据库



2066 speakers in the recording

DATA Information

- Total Time : 7.2 Hours
- Sampling Rate : 16kHz
- Sample Format : 16bit
- Environment : Car
- Speech Data Type : wav
- Recording Equipment : High Fidelity Microphone



Sample



Readme



Database

License: Apache License v.2.0



语音识别实验评测

Speech Recognition Evaluation

AISHELL-2021C-EVAL

低幼儿童智能语音产品交互数据库



1117 speakers in the recording

DATA Information

- Total Time : 2.18 Hours
- Sampling Rate : 16kHz
- Sample Format : 16bit
- Environment : Indoor
- Speech Data Type : wav
- Recording Equipment : Mobile Phone



Sample



Readme



Database

License: Apache License v.2.0



语音识别实验评测
声纹实验评测

Speech & Speaker Recognition Evaluation

AISHELL-2021B-EVAL

安静+噪声家居环境近远讲同步语音数据库



30 speakers in the recording

DATA Information

- Total Time : 15.8 Hours
- Sampling Rate : 16kHz
- Sample Format : 16bit
- Environment : Indoor
- Speech Data Type : wav
- Recording Equipment : Android Mobile Phone



Sample



Readme



Database

License: Apache License v.2.0



语音识别实验评测
声纹实验评测

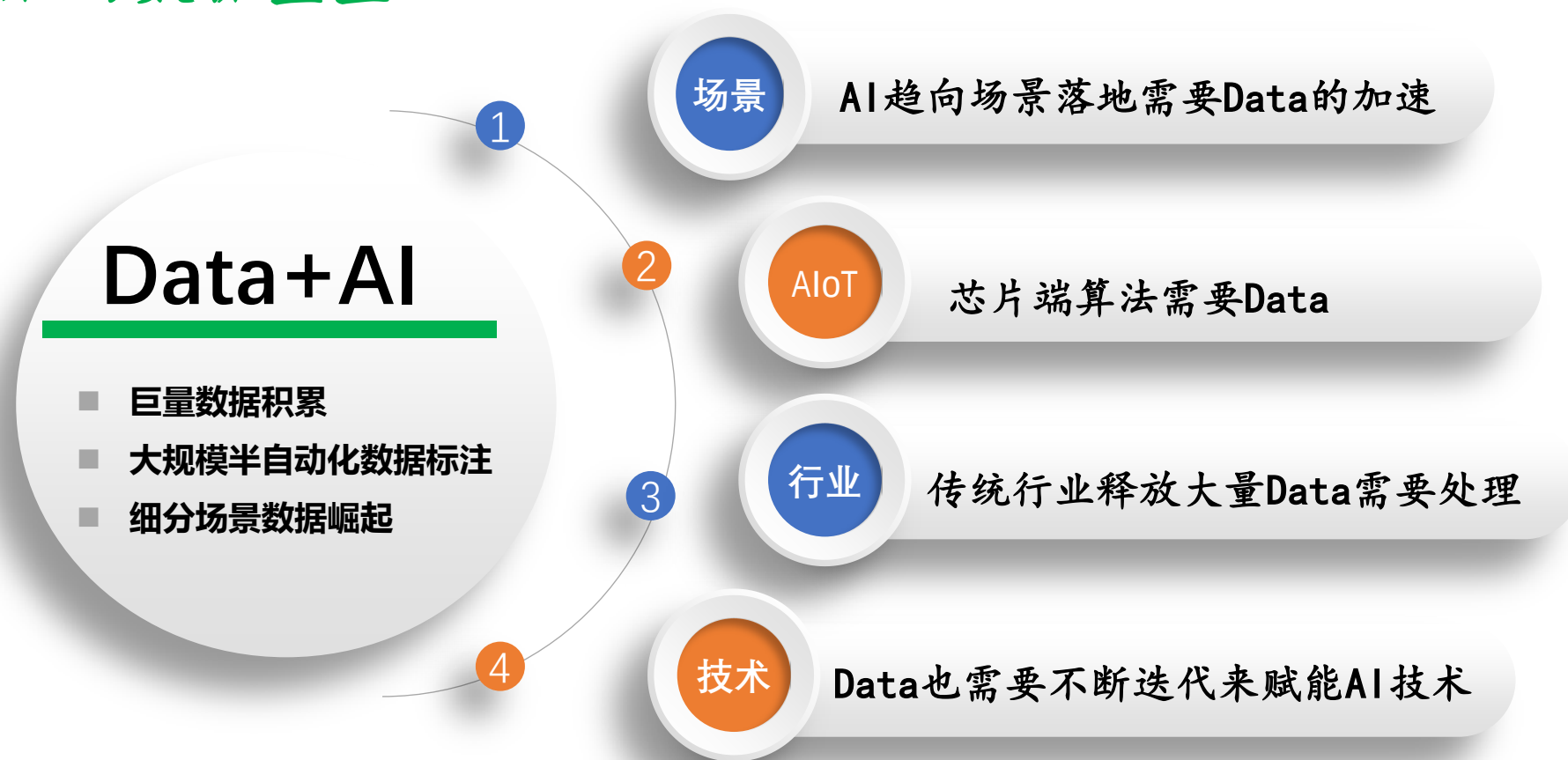
Speech & Speaker Recognition Evaluation

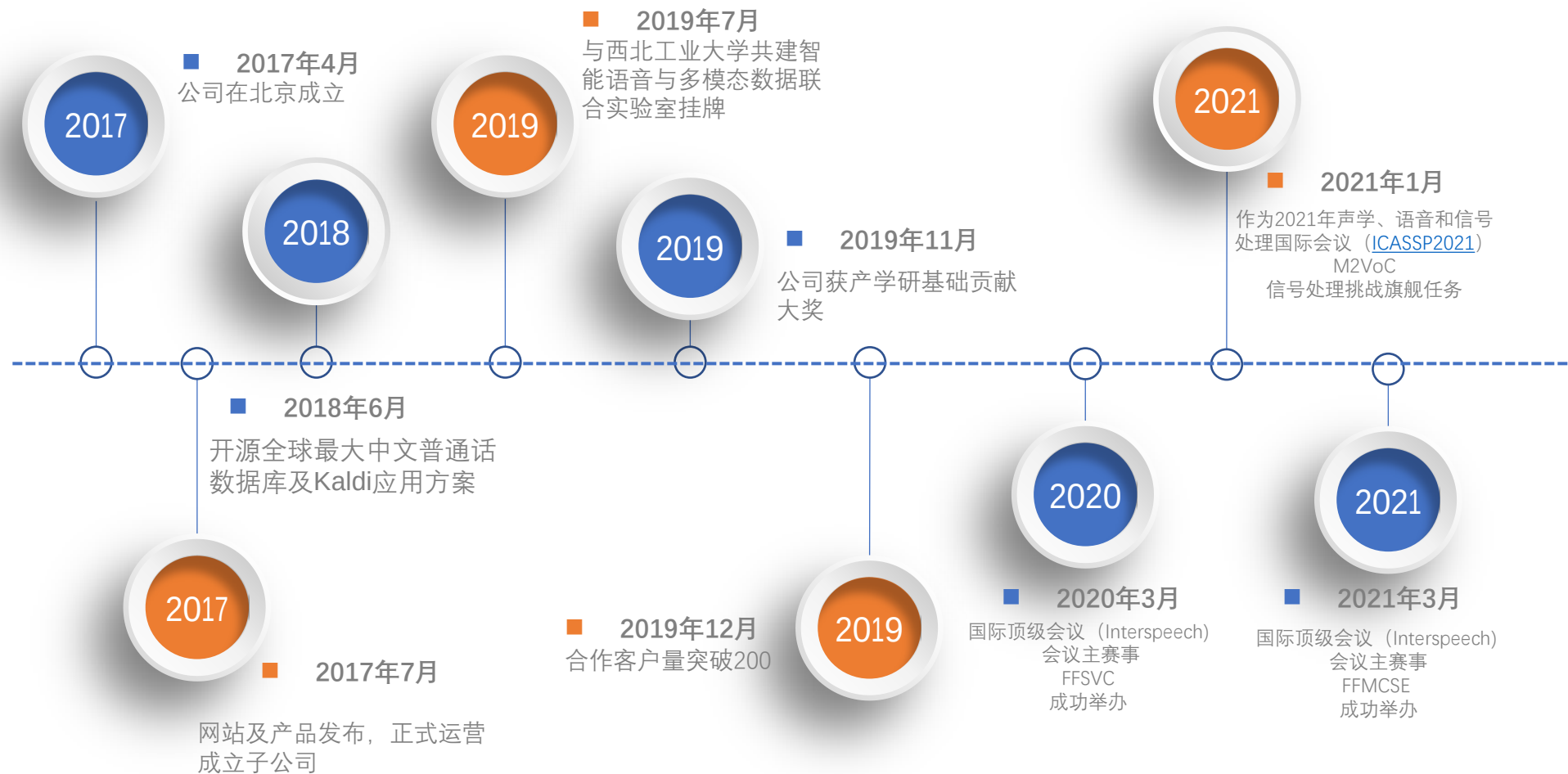


解决智能语音技术背后的数据壁垒

北京希尔贝壳科技有限公司是一家专注人工智能大数据的创新公司。

聚焦人工智能场景化数据服务，利用机器学习平台，在数据评测、辅助标注、数据分析等场景业务建立了领先的核心技术体系。希尔贝壳基于多年的人工智能基础技术研究，在行业必须的训练数据集结构分析和建设上有着不断创新的能力。







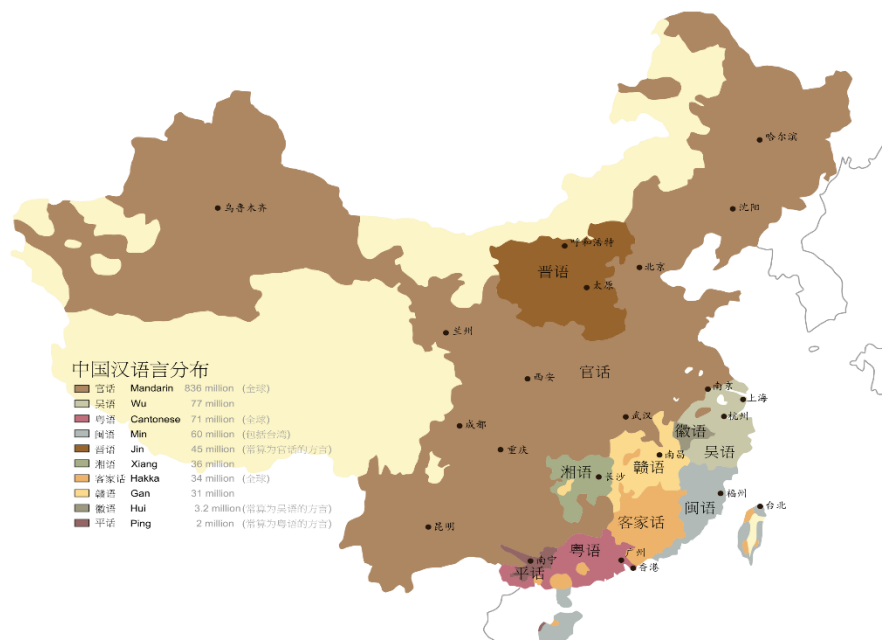
■ 研发建设不完善的语言数据

方言

民族语言

多语并行

等等



■ 结合图像、感知等的数据来形成多模态智能语音数据



图像识别的痛点：光线、动作、属性等

语音识别的痛点：语速、口音、噪音等

语义理解的痛点：知识不足、模糊处理等

单模态向多模态转变，优点互补优化高智能系统



数据的开源、开放

从用国外的数据到国外用我们的数据

558

200

AISHELL-1 & AISHELL-2

AISHELL-1 Kaldi recipe
s5: a **speech** recognition recipe
v1: a **speaker** recognition recipe
<http://www.openslr.org/33/>
178H
400 Speakers
● Sampling Rate: 16KHz
Sample Format: 16bit
Environment: Indoor

Serial Number	Domain
1	Smart Home Voice Control
2	POI (Geographic Information)
3	Music (Voice Control)
4	Digital Sequence (Voice Control)
5	TV Play and Film Names
6	Finance
7	Science and Technology
8	Sports
9	Entertainments
10	News
11	English Spelling
12	Wide Area Text

AISHELL-2 Kaldi recipe
speech recognition recipe
kaldi/egs/aishell2/
http://www.aishelltech.com/aishell_2
1000H
1991 Speakers
● Sampling Rate: 16KHz
Sample Format: 16bit
Environment: Indoor

Serial Number	Domain
1	Smart Home Voice Control
2	POI (Geographic Information)
3	Music (Voice Control)
4	Digital Sequence (Voice Control)
5	TV Play and Film Names
6	Finance
7	Science and Technology
8	Sports
9	Entertainments
10	News
11	English Spelling
12	Wide Area Text

AISHELL-3

AISHELL-3 multi-speaker Text-to-Speech (TTS) systems

AISHELL-3 Corpus:
Baseline System Samples

Arxiv: 2010.11567
Github Repo: [sos1sos2sixteen/aishell-3-baseline-tts](https://github.com/sos1sos2sixteen/aishell-3-baseline-tts)
Dataset Download: www.aishelltech.com/aishell_3
For further questions regarding the dataset: tech@aishelldata.com

● Sampling Rate: 44.1KHz
Sample Format: 16bit
Environment: Indoor

85H
218 Speakers

System recipe
<https://sos1sos2sixteen.github.io/aishell3/>

Download-Dataset
http://www.aishelltech.com/aishell_3

Download-Paper
<https://arxiv.org/abs/2010.11567>

Authors

- Shi Yao (Wuhan University, Duke Kunshan University)
- BU, Hu (AISHELL)
- XU, Xin (AISHELL)
- ZHANG, Shaoji (AISHELL)
- Li, Ming (Duke Kunshan University, Wuhan University) — ming.1369@dukekunshan.edu

AISHELL-DMASH

DMASH Distributed Microphone Arrays in Smart Home (DMASH) Dataset

Open Source
The INTERSPEECH 2020 Far-Field Speaker Verification Challenge
http://www.aishelltech.com/DMASH_Dataset

50000H
Sampling Rate: 44.1KHz & 16KHz
Sample Format: 16bit
Environment: Indoor

500 Speakers

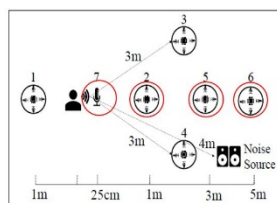
Authors

- SHI Yao (Wuhan University, Duke Kunshan University)
- BU, Hu (AISHELL)
- XU, Xin (AISHELL)
- ZHANG, Shaoji (AISHELL)
- Li, Ming (Duke Kunshan University, Wuhan University) — ming.1369@dukekunshan.edu

AISHELL-WakeUp-1

HI-MIA

“你好，米雅” “Hi, Mia” Open Source



AISHELL Speaker Verification Challenge 2019

HI-MIA
<http://www.openslr.org/85/>

AISHELL-WakeUp-1
http://www.aishelltech.com/wakeup_data

1561H
340 Speakers
Sampling Rate: 44.1KHz & 16KHz
Sample Format: 16bit
Environment: Indoor

AISHELL-4: An Open Source Dataset for Speech Enhancement, Separation, Recognition and Speaker Diarization in Conference Scenario

Yihui Fu^{1,*}, Luyao Cheng^{1,*}, Shubo Lv¹, Yukai Jv¹, Yuxiang Kong¹, Zhuo Chen³
Yanxin Hu¹, Lei Xie¹, Jian Wu³, Hui Bu², Xin Xu², Jun Du⁴, Jingdong Chen¹

¹Northwestern Polytechnical University, Xi'an, China

²Beijing Shell Shell Technology Co., Ltd., Beijing, China

³Microsoft Corporation

⁴University of Science and Technology of China, Hefei, China

{yhfu, lycheng, shblv, ykju, yxkong, yxhu}@npu-aslp.org, lxie@nwpu.edu.cn, {zhuc, wujian}@microsoft.com, {buhui, xuxin}@aishelldata.com, jundu@ustc.edu.cn, jingdongchen@ieee.org



高价值、前沿的赛事 AI语音技术的第一靶场

主办方: AISHELL AISHELL FOUNDATION
昆山杜克大学 DUKE KUNSHAN UNIVERSITY 中国计算机学会 语音对话与听觉专家组

AISHELL 2019 Speaker Verification Challenge

2019 9月6日 火爆开赛

竞赛简介:
未来智能语音交互的设备都将具备声纹识别功能,借此确认用户身份,只有自己的声纹才可以启动购物、签字、控制等。万物联网所带来的智能化时代,语音助手、安防等领域有着广泛的应用场景,声纹识别技术将面临很多挑战。本届大赛以智能家居场景为假设,从近场注册远场测试和远场注册远场测试两个技术点出发设计赛题。通过赛事发现参赛技术方的优秀创新成果,引领声纹识别的未来。

赛事任务: 多通道远场文本相关声纹识别

TRACK 1 近场数据注册, 远场数据测试
TRACK 2 远场数据注册, 远场数据测试

指定数据:
赛事数据来自希尔贝壳的AISHELL-Wakelife-1唤醒数据库,录音内容为高保真近讲Mic, 1m, 3m, 5m的中文内容“你好, 米雅”的数据

奖项设置:
第一名奖金6000元 高品质数据, 定制纪念品
第二名奖金4000元 高品质数据, 定制纪念品
第三名奖金3000元 高品质数据, 定制纪念品

PS:
前三名还可获得企业内推广与入职通道, 颁发纸质、电子证书, 一份神秘大礼

日程安排:
09月04日: 开放注册, 开放训练集与开发集
09月28日: 报名截止
10月10日: 开放测试集, 开启结果提交通道
10月20日: 关闭提交结果通道
10月28日: 在kaigi第四届中国语音技术大会上公布结果并颁奖

报名方式: 扫描二维码报名 或登陆网站: challenge.aishelltech.com



INTERSPEECH 2020
OCTOBER 25-29/ SHANGHAI, CHINA
SHANGHAI INTERNATIONAL CONVENTION CENTER

FFSVC 2020
Far-Field Speaker
Verification Challenge



Distributed Microphone Arrays in Smart Home Database.
(DMASH)

Team Paper

184

5+

<http://2020.ffsvc.org/>



ConferencingSpeech 2021

Far-field Multi-Channel Speech Enhancement Challenge for Video Conferencing

Far-field Multi-Channel Speech Enhancement Challenge for Video Conferencing (ConferencingSpeech 2021)

The ConferencingSpeech 2021 challenge is proposed to stimulate research in multi-channel speech enhancement and aims for processing the far-field speech from microphone arrays in the video conferencing rooms. Targeting the real video conferencing room application, the ConferencingSpeech 2021 challenge database is recorded from real speakers. The number of speakers and distances between speakers and microphone arrays vary according to the sizes of meeting rooms. Multiple microphone arrays from three different types of geometric topology are allocated in each recording environment.



Multi-Speaker Multi-Style Voice Cloning Challenge (M2VoC)

Text-to-speech (TTS) or speech synthesis has witnessed significant performance improvement with the help of deep learning. The latest advances in end-to-end text-to-speech paradigm and neural vocoder have enabled us to produce very realistic and natural-sounding synthetic speech reaching almost human-parity performance. But this amazing ability is still limited to the ideal scenarios with a large single-speaker less-expressive training set. The speech quality, target similarity, expressiveness and robustness are still not satisfied for synthetic speech with different speakers and various styles, especially in real-world low-resourced conditions, e.g., each speaker only has a few samples at hand. The current open solutions are also not robust enough to unseen speakers. We call this challenging task as multi-speaker multi-style voice cloning (M2VoC).



人才社区的开源、开放 AI人才的建设应该是AI新基建最应该夯实的



Speech home 首页 声浪 活动 知识 数据分享 招聘 [发表](#)



傅里叶变换取代Transformer自注意力层，谷歌这项研究GPU上快7倍、TPU上快2倍



「语音识别」算法常规赛
SPEECH RECOGNITION REGULAR ALGORITHM COMPETITION
4月8日-6月30日 火热进行中 >>

语音之家 - speechhome
助力AI语音开发者的社区

问题反馈:
010--80225006
邮箱:
Jack@speechhome.com

热门主题

英伟达CPU问世: ARM 架构, 对比x86实现十...

27赞 · 0评论

微软拟斥资 160 亿美元收购全球最大语音识别公...

23赞 · 0评论

ImageNet验证集6%的标签都是错的, MIT: 十...

22赞 · 0评论

语音小管家 · [转载](#) 1天前

傅里叶变换取代Transformer自注意力层，谷歌这项研究GPU上快7倍、TPU上快2倍

👍 0 | 💬 0 NLP

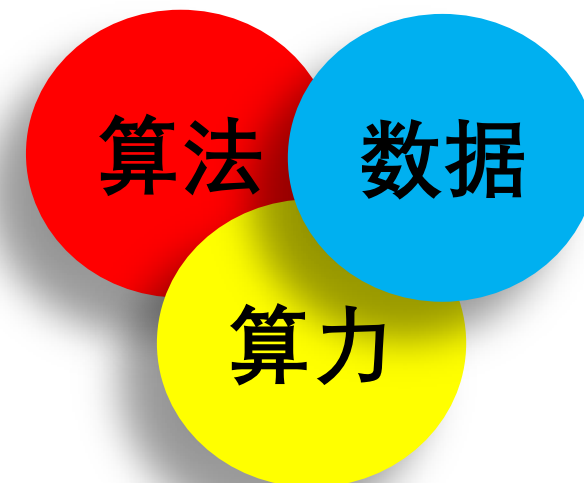
语音小管家 · [转载](#) 5天前

TensorFlow 助力: AI 语音降噪打造 QQ 音视频通话新体验

👍 2 | 💬 0 语音增强

案例分享
Powered by TensorFlow







THANKS