## SUNDAY, October 25th - Program at a Glance

<table>
<thead>
<tr>
<th>TIME</th>
<th>Room 1</th>
<th>Room 2</th>
<th>Room 3</th>
<th>Room 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>16:30 - 18:00</td>
<td><strong>Tutorial 1</strong>&lt;br&gt;Efficient and flexible implementation of machine learning for ASR and MT</td>
<td><strong>Tutorial 2</strong>&lt;br&gt;Spoken dialogue for social robots</td>
<td><strong>Tutorial 3</strong>&lt;br&gt;Meta learning and its applications to human language processing</td>
<td><strong>Tutorial 4</strong>&lt;br&gt;Intelligibility evaluation and speech enhancement based on deep learning</td>
</tr>
<tr>
<td>18:15 - 19:45</td>
<td><strong>Tutorial 1</strong>&lt;br&gt;contd.</td>
<td><strong>Tutorial 2</strong>&lt;br&gt;contd.</td>
<td><strong>Tutorial 3</strong>&lt;br&gt;contd.</td>
<td><strong>Tutorial 4</strong>&lt;br&gt;contd.</td>
</tr>
<tr>
<td>20:00 - 21:30</td>
<td><strong>Tutorial 5</strong>&lt;br&gt;&quot;Speech 101&quot; - What everyone working on spoken language processing needs to know about spoken language</td>
<td><strong>Tutorial 6</strong>&lt;br&gt;Neural approaches to conversational information retrieval</td>
<td><strong>Tutorial 7</strong>&lt;br&gt;Neural models for speaker diarization in the context of speech recognition</td>
<td><strong>Tutorial 8</strong>&lt;br&gt;Spoken language processing for language learning and assessment</td>
</tr>
<tr>
<td>21:45 - 23:15</td>
<td><strong>Tutorial 5</strong>&lt;br&gt;contd.</td>
<td><strong>Tutorial 6</strong>&lt;br&gt;contd.</td>
<td><strong>Tutorial 7</strong>&lt;br&gt;contd.</td>
<td><strong>Tutorial 8</strong>&lt;br&gt;contd.</td>
</tr>
</tbody>
</table>

Coffee break 18:00 - 18:15

Coffee break 19:45 - 20:00

Coffee break 21:30 - 21:45
### MONDAY, October 26th - Program at a Glance

<table>
<thead>
<tr>
<th>TIME</th>
<th>Room 1</th>
<th>Room 2</th>
<th>Room 3</th>
<th>Room 4</th>
<th>Room 5</th>
<th>Room 6</th>
<th>Room 7</th>
<th>Room 8</th>
<th>Room 9</th>
<th>Room 10</th>
<th>Room 11</th>
<th>Room 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:00</td>
<td>Opening session</td>
<td>Keynote 1: Janet B. Pierrehumbert, The cognitive status of simple and complex models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:15</td>
<td>Mon-1-1</td>
<td>Mon-1-2</td>
<td>Mon-1-3</td>
<td>Mon-1-4</td>
<td>Mon-1-5</td>
<td>Mon-1-6</td>
<td>Mon-1-7</td>
<td>Mon-1-8</td>
<td>Mon-1-9</td>
<td>Mon-1-10</td>
<td>Mon-1-11</td>
<td>Mon-S&amp;T 1</td>
</tr>
<tr>
<td></td>
<td>ASR neural network architectures</td>
<td>Multi-channel speech enhancement</td>
<td>Speech processing in the brain</td>
<td>Speech Signal Representation</td>
<td>Speech Synthesis: Neural Waveform Generation</td>
<td>Automatic Speech Recognition for Non-Native Children's Speech</td>
<td>Speaker Diarization</td>
<td>Noise robust and distant speech recognition</td>
<td>Speech in Multimodality (MULTIMODAL)</td>
<td>Speech, Language, and Multimodal Resources</td>
<td>Language Recognition</td>
<td></td>
</tr>
<tr>
<td>20:15</td>
<td>Mon-1-12</td>
<td>Mon-1-13</td>
<td>Mon-1-14</td>
<td>Mon-1-15</td>
<td>Mon-1-16</td>
<td>Mon-1-17</td>
<td>Mon-1-18</td>
<td>Mon-1-19</td>
<td>Mon-1-20</td>
<td>Mon-1-21</td>
<td>Mon-1-22</td>
<td>Mon-S&amp;T 1</td>
</tr>
<tr>
<td></td>
<td>ASR model training and strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Speech processing and analysis</td>
<td></td>
</tr>
<tr>
<td>21:00</td>
<td>Mon-S&amp;T 1</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
<td>Mon-S&amp;T 2</td>
</tr>
<tr>
<td>22:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Coffee Break
- 19:00 - 19:15
- 20:15 - 20:30
- 21:30 - 21:45

### TECHNICAL AREAS

1. Speech Perception and Production
2. Phonetics, Phonology, and Prosody
3. Paralinguistic Analysis
4. Speaker and Language Identification
5. Analysis of Speech and Audio Signals
6. Speech Coding Enhancement
7. Speech Synthesis
8. Speech Recognition II: Signal Processing
9. Speech Recognition II: Architecture
10. Speech Recognition II: New Applications
11. Spoken Dialog Systems
12. Special Sessions

---

**ISCA-SAC "2nd Mentoring"**

---

**Diversity Meeting**

21:00 - 22:00
### TUESDAY, October 27th - Program at a Glance

<table>
<thead>
<tr>
<th>TIME (Shanghai)</th>
<th>Room 1</th>
<th>Room 2</th>
<th>Room 3</th>
<th>Room 4</th>
<th>Room 5</th>
<th>Room 6</th>
<th>Room 7</th>
<th>Room 8</th>
<th>Room 9</th>
<th>Room 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>18:00</td>
<td>Keynote 2: Barbara Shinn-Cunningham, Brain networks enabling speech perception in everyday settings</td>
<td>Coffee break 19:00 - 19:15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:15 - 20:15</td>
<td>Tue-1-1 Speech Translation and multilingual/multimodal learning</td>
<td>Tue-1-2 Speaker Recognition I</td>
<td>Tue-1-3 Spoken Language Understanding II</td>
<td>Tue-1-4 Human speech processing</td>
<td>Tue-1-5 Feature extraction and distant ASR</td>
<td>Tue-SS-1-6 Voice Privacy Challenge</td>
<td>Tue-1-7 Speech Synthesis: Text Processing, Data and Evaluation</td>
<td>Tue-1-8 Search for speech recognition</td>
<td>Tue-1-9 Computational Paralinguistics I (CP I)</td>
<td>Tue-1-10 Acoustic Phonetics and Prosody</td>
</tr>
<tr>
<td>20:30 - 22:00</td>
<td>ISCA General Assembly</td>
<td>Coffee break 20:15 - 20:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### TECHNICAL AREAS

1. Speech Perception and Production
2. Phonetics, Phonology, and Prosody
3. Paralinguistic Analysis
4. Speaker and Language Identification
5. Analysis of Speech and Audio Signals
6. Speech Coding Enhancement
7. Speech Synthesis
8. Speech Recognition I: Signal Processing
9. Speech Recognition II: Architecture
10. Speech Recognition III: New Applications
11. Spoken Dialog Systems
12. Spoken Language Processing

- Special Sessions
- Show & Tell
<table>
<thead>
<tr>
<th>TIME (Shanghai)</th>
<th>Room 1</th>
<th>Room 2</th>
<th>Room 3</th>
<th>Room 4</th>
<th>Room 5</th>
<th>Room 6</th>
<th>Room 7</th>
<th>Room 8</th>
<th>Room 9</th>
<th>Room 10</th>
<th>Room 11</th>
<th>Room 12</th>
<th>Room 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>18:00</td>
<td>Keynote 3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TECHNICAL AREAS**

1. Speech Perception and Production
2. Phonetics, Phonology, and Prosody
3. Paralinguistic Analysis
4. Speaker and Language Identification
5. Analysis of Speech and Audio Signals
6. Speech Coding Enhancement
7. Speech Synthesis
8. Speech Recognition II: Architecture
9. Speech Recognition: Signal Processing
10. Speech Recognition: New Applications
11. Special Sessions
12. Show & Tell
13. Industry Forum

**WEDNESDAY, October 28th - Program at a Glance**

Lin-shan Lee, Doing Something we Never could with Spoken Language Technologies from early days to the era of deep learning
# THURSDAY, October 29th - Program at a Glance

<table>
<thead>
<tr>
<th>TIME (Shanghai)</th>
<th>Room 1</th>
<th>Room 2</th>
<th>Room 3</th>
<th>Room 4</th>
<th>Room 5</th>
<th>Room 6</th>
<th>Room 7</th>
<th>Room 8</th>
<th>Room 9</th>
<th>Room 10</th>
<th>Room 11</th>
<th>Room 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>18:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:15 - 20:15</td>
<td>Thu-1-1</td>
<td>Thu-1-2</td>
<td>Thu-1-3</td>
<td>Thu-1-4</td>
<td>Thu-1-5</td>
<td>Thu-SS-1-6</td>
<td>Thu-1-7</td>
<td>Thu-1-8</td>
<td>Thu-1-9</td>
<td>Thu-1-10</td>
<td>Thu-1-11</td>
<td>Industry Forum</td>
</tr>
<tr>
<td>20:30 - 21:30</td>
<td>Thu-2-1</td>
<td>Thu-2-2</td>
<td>Thu-2-3</td>
<td>Thu-2-4</td>
<td>Thu-SS-2-5</td>
<td>Thu-2-6</td>
<td>Thu-2-7</td>
<td>Thu-2-8</td>
<td>Thu-2-9</td>
<td>Thu-2-10</td>
<td>Thu-2-11</td>
<td></td>
</tr>
<tr>
<td>21:45</td>
<td>Thu-3-1</td>
<td>Thu-3-2</td>
<td>Thu-3-3</td>
<td>Thu-3-4</td>
<td>Thu-3-5</td>
<td>Thu-3-6</td>
<td>Thu-3-7</td>
<td>Thu-3-8</td>
<td>Thu-3-9</td>
<td>Thu-3-10</td>
<td>Thu-3-11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Speech in Health II (HEALTH II)</td>
<td>Speech and Audio Quality Assessment</td>
<td>Privacy and Security in Speech Communication</td>
<td>Voice Conversion and Adaptation II</td>
<td>Multilingual and code-switched ASR</td>
<td>Speech and voice disorders</td>
<td>The Zero Resource Speech Challenge 2020</td>
<td>LM adaptation, Lexicon Units and Punctuation</td>
<td>Speech in Health I (HEALTH I)</td>
<td>ASR neural network architectures II - Transformers</td>
<td>Speech and voice disorders</td>
<td></td>
</tr>
<tr>
<td>23:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Closing</td>
</tr>
</tbody>
</table>

## TECHNICAL AREAS

1. Speech Perception and Production
2. Phonetics, Phonology, and Prosody
3. Paralinguistic Analysis
4. Speaker and Language Identification
5. Analysis of Speech and Audio Signals
6. Speech Coding Enhancement
7. Speech Synthesis
8. Speech Recognition I: Signal Processing
9. Speech Recognition II: Architecture
10. Speech Recognition III: New Applications
11. Spoken Dialog Systems
12. Spoken Language Processing

## Special Sessions

- Show & Tell
- Industry Forum
**iFLY Star**

**iFLYTEK Research Institute Campus Recruitment 2021 - iFLY Star Project**

**Who are we?**

**iFLYTEK Research Institute**

iFLYTEK A.I. Research Institute was established in 2005, featuring ‘aim high with technologies, stand firm with applications; conduct useful research via correct method’. We are committed to the research of core A.I. technologies. We focus on research in following areas: intelligent speech, computer vision, natural language processing, etc. iFLYTEK boasts multiple world’s leading research achievements in the field of artificial intelligence and has won champions in global competitions for many times.

Here in iFLYTEK. Appreciate the world's leading A.I. technology. Enjoy a simple and sincere culture. Explore the profound mystery of A.I. with scientists.

**iFLY Star Project**

“iFLY Star Project” is a global recruitment looking for for highly-motivated graduates of leading universities to join iFLYTEK Research. Through the tailored growth plans, we will help you pave your way to rising stars in the A.I. field.

**You will get**

1. Luxury offer package: Competitive salary + long-term incentives.
2. Tailored growth plan: An integrated system of production, education and research; broad participation in industrialization and application; “Ace Pilot” Elite training program; Adequate guidance provided by top tutors of executives/scientists; Systematic training for infinite possibilities; Interdisciplinary research opportunities.
3. Opportunities and abilities to change the world together: Work side by side with the world’s top research teams; Create a better world with artificial intelligence.

**Recruitment Positions**

**Researcher**

Responsible for the research of core A.I. algorithms, including intelligent speech / computer vision / natural language processing, etc. Solve technical problems in application and industrialization of core algorithms.

**Algorithm Engineer**

Responsible for the R&D engineering, industrialization, application and promotion of core A.I. algorithms, including intelligent speech / computer vision / natural language processing, etc.

**Locations**

Hefei / Shanghai / Suzhou / Beijing / Guangzhou / Xi’an / Changchun

**We need you**

2021 undergraduates, postgraduates and doctoral graduates

Apply if you meet one of the following criteria:

- Rising academic stars with publications in top international academic conferences or journals.
- Legendary champions of college entrance examinations, gifted youths in special classes, and TOP 20% students in well-known universities at home and abroad.
- Medal winners of ACM–ICPC regional or global competitions.
- Candidates with rich internship experience in top companies or extensive project practices on campus are preferred, and those with unique understanding or breakthroughs in a certain technical field are preferred.

**How to submit your resume:**

Scan the QR code below via WeChat

Enable machines to listen, speak, understand and think.
Let us join hands,
Create a better world with artificial intelligence.

For more information, please visit the official website of campus recruitment (campus.iflytek.com)
**WHO ARE WE?**

Magic Data Technology, an AI data service provider. We are committed to providing a wide range of customized data services in the fields of Automatic Speech Recognition, Text To Speech, Computer Vision Recognition and Natural Language Processing. With human-in-the-loop data processing, we significantly improved the efficiency and quality of AI data labeling. Established in 2016, we have collected more than 100,000 hours of standard multilingual speech corpora under various scenarios. We help our clients gain easy and timely access to data with accuracy up to 99% and also provide them customized solutions. Magic Data employs a vast team of skilled data specialists and has a wide network of consultants around the world to assist with specific data needs.

**MAGIC DATA PROVIDES**

- Quality: Pre-packaged datasets available for immediate application in AI system development.
- Scalability: Capable to cover 50+ languages and dialects for audio recording and collecting services.
- Compliance: Fully legitimate and traceable data, prepared under strict data encryption, in supervised system, promises its reliability and security.
- Expertise: Strict quality management system, ensuring a continuous output of high-quality data products.

**DATASETS**

We provide valuable and reliable training data to empower your AI models. You can find datasets in different languages, styles, and solutions. Our datasets can improve your AI models’ performance, thus accelerating the commercialization of AI initiatives.

### Selected Language Datasets

**ASR**

- Read Speech Corpus
  - American English Speech Corpus
  - French Speech Corpus
  - Peninsular Spanish Speech Corpus
  - Korean Speech Corpus
  - Japanese Speech Corpus
  - Bahasa Indonesian Speech Corpus
  - Thai Speech Corpus
  - Mandarin Chinese Speech Corpus
  - Shanghai Dialect Speech Corpus
  - Guangzhou Cantonese Speech Corpus
  - Minnan Dialect Speech Corpus
  - Wuhan Dialect Speech Corpus
  - Shanxi Dialect Speech Corpus
  - Guangzhou Cantonese In-Vehicle Speech Corpus
  - Chinese-English Code-Mixing Speech Corpus

- Conversational Speech Corpus
  - English Conversational Telephone Speech Corpus
  - Japanese English Conversational Speech Corpus
  - Korean English Conversational Speech Corpus
  - Japanese Conversational Speech Corpus
  - Korean Conversational Speech Corpus
  - Bahasa Indonesian Conversational Speech Corpus
  - Turkish Conversational Speech Corpus
  - Malay Conversational Speech Corpus
  - Mandarin Chinese Conversational Speech Corpus
  - Mandarin Chinese Conversational Telephone Speech Corpus
  - Guangzhou Cantonese Conversational Speech Corpus
  - Shanghai Dialect Conversational Speech Corpus
  - Sichuan Dialect Conversational Speech Corpus
  - Uygur Conversational Speech Corpus
  - Hangzhou Dialect Conversational Speech Corpus

- **TTS**
  - Chinese female voice emotion TTS dataset
  - Chinese female customer service TTS dataset

- **Popular Launched**
  - ASR
    - German Conversational Speech Corpus
    - French Conversational Speech Corpus
    - Peninsular Spanish Conversational Speech Corpus

  - ASR
    - Italian Conversational Speech Corpus
    - Brazilian Portuguese Conversational Speech Corpus
    - Peninsular Arabic Conversational Speech Corpus

**Scan QR code for more information**

For more dataset products, please consult customer service: 400-900-5251

Email: business@magicdatatech.com

Website: www.magicdatatech.com
英特尔® 至强® 可扩展平台
高效实践人工智能

深度学习加速，显著提升性能；创新存储技术，优化批量训练；广泛硬件组合，满足多样推理；强大软件工具，加速开发部署。

打破理论与现实的壁垒

加速AI实践，请访问www.intel.cn/ai
Intelligent Speech Interaction Team
@ Alibaba Group

Our mission is to enable speech & NLP to become one of the technical infrastructures supporting Alibaba’s business platforms and the small businesses in our ecosystem.

We are building state-of-the-art technologies in speech recognition, speech synthesis, speaker identification and verification, emotion detection, natural language understanding, dialogue system, QA system. We are also building large-scale deep learning infrastructure, and creating scalable speech & NLP services on top of Alibaba Cloud computing infrastructure. With all these technologies and platforms, we are the hub of natural user interface between human and Alibaba services.
The mission of Meituan is “We help people eat better, live better”. As China’s leading e-commerce platform for services, Meituan operates well-known mobile apps in China, including Meituan, Dianping, Meituan Waimai, and others. Meituan offers over 200 service categories, including catering, on-demand delivery, car-hailing, bike-sharing, hotel, and travel booking, movie ticketing, and other entertainment and lifestyle services, which covers 2800 cities and counties across China.

Meituan Speech Interaction

Meituan Speech Interaction Department is established in 2017, our mission is to develop advanced speech interaction technology to meet the needs in Meituan's commercial ecosystem. We focus on speech interaction technologies such as speech signal processing, speech recognition, speech synthesis, speaker recognition, natural language understanding, dialog and knowledge graph. We are looking for talents (both full time and internship) who are interested in speech interaction technology.

We Need You – Speech Engineer

Responsibilities
• Architect, design and develop ASR/TTS/speech processing algorithms.
• Improve the system performance in production with cutting edge technologies.

Basic Qualifications
• Advanced knowledge of software engineer, familiar with C++, python.
• Knowledge of speech signal processing, machine learning or deep learning.
• Experience in developing online speech service.

Preferred Qualifications
• Expert knowledge of ASR, TTS or other related technologies.
• Authors of papers in Interspeech, ICASSP or other related top tier journals/conferences.
• Experience with distributed training system, GPU optimization.

Contact us: meituan oi@meituan.com (Titled with “Interspeech2020” )
与AI同行——2021思必驰校招进行中

如果你对人工智能极具热情
如果你具备优秀的学习能力
如果你善于沟通与协
思必驰以诚挚之心邀请你的加入
我们一起 可以做更酷的事情！

招聘岗位

<table>
<thead>
<tr>
<th>岗位</th>
<th>苏州</th>
<th>上海</th>
<th>北京</th>
<th>深圳</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLP算法工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>自然语言处理研究员</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>语言技术研究员</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>语音算法工程师（识别/唤醒/合成）</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>语音信号处理算法工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAVA研发工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>安全工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>硬件工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C++研发工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>嵌入式开发工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>助理测试工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>安卓开发工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>产品经理</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>产品运营</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>后端开发工程师</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

岗位详情，请扫描下方二维码

简历投递

简历投递邮箱（邮件正文请备注，消息来自INTERSPEECH）
— xiaozhao@aispeech.com
Sogou Vocational Avatars: Digital human-based Multimodal Human AI Interaction system

**Multimodal recognition (ASR+lip reading)**
- More than 1.4 billion speech recognition requests per day, including multilingual & dialect recognition
- Initiated personalized speech recognition, which greatly improved personalized speech recognition experience
- First issued Chinese lip recognition system, command word recognition accuracy can reach 90% in vehicle, smart home and other scenes
  - Based on modality attention, initiated multimodal recognition technology, which can significantly improve the recognition effect in noisy environment

**Speech synthesis**
- Leading neural network modeling technologies, Self-developed StyleTTS system
- Supporting multilingual, multi-tone and multi-style diversified broadcast
- First issued personalized speech synthesis and emotion transfer in the industry
- Won the first place in the 2 subtasks of Blizzard Challenge 2018, the most authoritative speech synthesis competition in the world

**Voice conversion**
- Research of innovative technologies based on representation learning and transfer learning
- Industry’s first voice conversion product which can support real-time any-to-one voice conversion.
- With high similarity and sound quality, it has supported dozens of speakers including stars, anime, games and more types.

**Digital human**
- Integrating Sogou Multimodal HCI technology including core technologies of speech, computer vision and natural language processing.
- The industry’s first commercial application level of multi-mode audio and video generation technology with broadcasting effect similar to human.
- Has been applied in the media, customer service, justice, education, entertainment and other fields, and has gone abroad.
Multimodal recognition

The greatest voice input method in China
- More than 1.4 billion daily speech recognition requests
- Language and Dialect
- Speech enhancement
- Personalized interaction

Sogou AI Hardware

Multimodal representation

Speech synthesis
Personalized speech synthesis
- Upload 10 sentences about 3 to 5 minutes of audio
- The timbre is close to human, customizing the proprietary sound archive.
- Featuring audio companionship, you can tell stories, read private book lists, etc

Voice conversion
Pioneering in the industry, double breakthrough of representational learning and transfer learning.
- High similarity
- Free to convert your voice
- Multi-representation transfer

AI news anchor
As the world's first synthetic news anchor, Sogou Vocational Avatars have been used in media.
- Vivid image
- Mature solutions
- Extremely low customization costs
- Quick customization

Machine translation

Sogou simultaneous interpretation
Ranked first in IWSLT 2018.

Sogou Translator

Please visit Sogou AI Open Platform: ai.sogou.com

Join us
Please send your resume to zhaopin@sogou-inc.com

Learn More