SHENG LI (李 勝)

Tenure-track Researcher

National Institute of Information and Communications Technology (NICT)
└──**The Universal Communication Research Institute (UCRI)**
  └──**The Advanced Speech Translation Research and Development Promotion Center (ASTREC)**
    └──**Advanced Speech Technology Laboratory (ASTL)**

E-mail: sheng.li [at] nict.go.jp / lisheng.cs [at] gmail.com / sheng.li [at] ieee.org
Address: 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan

---

Research Interest:

- multilingual speech recognition/translation/dialogue

- security-aware speech processing

Education:

- 2006.7 B.S Computer Science, Nanjing University

- 2009.7 M.E Software and Embedded Systems, Nanjing University
  (Joint Program with Chinese Academy of Sciences, Chinese University of Hong Kong)

- 2016.3 Ph.D Informatic Science, Kyoto University
  (Japanese Government MEXT, admission/tuition fee total exemption during PhD study)

Work Experiences:

- Jul.2009 - Apr.2012: Chinese Academy of Sciences [Guangdong China]

- Aprl.2016 - Jan.2017: Kyoto University, Speech and Audio Processing Lab. (ERATO researcher)

- 2017 - Present: National Institute of Information & Communications Technology (NICT)

Awards:

- **NICT**
  2022 Co-1st place in Main/OOD tracks in INTERSPEECH2022 special session: VoiceMOS challenge
  2021 3rd/4th place in constrained/unconstrained resource multilingual ASR tracks of OLR2021 challenge
  2021 R3 NICT Award: Outstanding Performance Award Excellence Award (Group)

- **Kyoto University**
  2018 IEEE Signal Processing Society Japan Student Journal Paper Award
  2016 Paper nominated as ACM/IEEE Trans. Audio, Speech and Language Process. cover
  2012-2016 Kyoto Univ. admission/tuition fee total exemption
  2012 MEXT scholarship by Japanese Government (recommanded by Kyoto Univ.)

- **Joint Lab CAS and CUHK**
  2012 Travel grant by IBM research for INTERSPEECH2012 at Portland, USA
  2011 Best Creative Project Award in Young Entrepreneur Program 2011, HK
  2011 Excellent Staff Award of Chinese Academy of Sciences

- **Nanjing University**
  2004 Encouragement Scholarship of Nanjing University
  2002 Chen Yinchuan Scholarship (Hongkong) for Excellent University New Students
  2002 Award of chemistry and biology Olympic for high school students (My wife got physics Olympic award, 1st class).

Teach/Supervise

- 2023 Superivsed PhD student of Kyoto Univ. (Qianying Liu) got IEEE-SPS grant for IEEE-ICASSP2023 oral presentation
  2022 Supervised PhD student of Univ. of Tokyo (Zhuo Gong) successfully graduated from Univ. of Tokyo
  2022 Supervised Master student of Kyoto Univ. (Wangjin Zhou and Zhengdong Yang) got Co-1st in VoiceMOS Challenge2022
  2021 Supervised PhD student of Kyoto Univ. (Soky Kak) got best student paper nomination in O-COCOSDA2021

2020 Two supervised Master students of Tianjin Univ. (Hao Shi and Yuqin Lin) got ISCA-grants in INTERSPEECH2020
2020 Supervised Master student of Kyoto Univ. (Yaowei Han) got best student paper nomination in IEEE-ICME2020

Fundings and Grants:

Multilingual Speech Recognition/Translation/Dialogue

- NICT international funding (PI): 2020-2024 (ongoing)
  Bridging Eurasia from Sea -- Multilingual Speech Recognition for Maritime Silkroad (2022-2023)
  Bridging Eurasia -- Multilingual Speech Recognition for Silkroad (2020-2022)

- Grant-in-Aid for Scientific Research (B) (Co-I): 2023-2028 (ongoing)
  意図を的確に伝える音声対話翻訳の基盤技術の創出

- Grant-in-Aid for Scientific Research (C) (PI): 2023-2026 (ongoing)
  M3OLR: Towards Effective Multilingual, Multimodal and Multitask Oriental Low-resourced Language Speech Recognition

- Grant-in-Aid for Research Activity Start-up (PI): 2019-2021
  Next generation multilingual End-to-End speech recognition (from G30 to G200)

- NICT tenure-track start-up funding (PI): 2020-2022
  Advanced Multilingual End-to-End Speech Recognition

Speech Security

- ICT Virtual Organization of ASEAN Institutes and NICT (ASEAN IVO) 2023 accepted project:(Co-I): 2023 (ongoing)
  Spoof Detection for Automatic Speaker Verification

- Grant-in-Aid for Young Scientists (PI): 2021-2023
  Phantom in the Opera -- the Vulnerabilities of Speech Interface for Robotic Dialogue System

- NII Open Collaborative Research (collaborator): 2020-2021
  Speaker De-identification with Provable Privacy in Speech Data Release

---

Publications (Selected)

GoogleScholar | ResearchGate | DBLP | Researchmap | SemanticScholar

Ph.D Thesis:

- Sheng Li (supervised by Prof. Tatsuya Kawahara).
  Speech Recognition Enhanced by Lightly-supervised and Semi-supervised Acoustic Model Training.
  Ph.D. Thesis, Kyoto University, Feb, 2016.

Book Chapter:

- **S. Li**, Bridging Eurasia: Multilingual Speech Recognition for Silkroad, ISBN: 978-4-904020-29-6, 2023.

- **S. Li**, Voices of the Himalayas: Investigation of Speech Recognition Technology for the Tibetan Language, ISBN: 978-4-904020-28-9, 2022.

- **S. Li**, Phantom in the Opera: The Vulnerabilities of Speech-based Artificial Intelligence Systems, ISBN: 978-4-904020-26-5, 2022.

- X. Lu, **S. Li**, M. Fujimoto, Speech-to-Speech Translation, pp. 21-38, Springer Singapore, 2020.
  S. Li, Chapter: From Shallow to Deep and Very Deep.
  S. Li, Chapter: End-to-End and CTC models.

Invited Talks:

- Phoneme-level articulatory animation in pronunciation training using EMA data,
  2012, Speech Synthesis Lab., Tsinghua University, host: Prof. Zhiyong Wu.

- Lightly-supervised training and confidence estimation by using CRF classifiers,
  2014, Speech and Cognition Lab., Tianjin University, host: Prof. Jianwu Dang and Prof. Kiyoshi Honda.

- End-to-End Speech Recognition,
  2019, University of Tokyo.

- Towards Security-awared Speech Recognition System,
  2023, NECTEC-NICT joint seminar, Thailand.

Journals (Peer reviewed):

- S. Qin, L. Wang, **S. Li** (corresponding), J. Dang and L. Pan. Improving Low-resource Tibetan End-to-end ASR by Multilingual and Multi-level Unit Modeling. EURASIP Journal on Audio, Speech and Music Processing. (EURASIP JASMP), No.2 , 2022.

- P. Shen, X. Lu, **S. Li**(patent co-inventor), H. Kawai. Knowledge Distillation-based Representation Learning for Short-Utterance Spoken Language Identification. IEEE Trans. Audio, Speech \& Language Process. (TASLP), vol. 28, pp. 2674--2683, 2020.

- **S. Li**, Y.Akita, and T.Kawahara. Semi-supervised acoustic model training by discriminative data selection from multiple ASR systems' hypotheses.
  IEEE Trans. Audio, Speech \& Language Process. (TASLP), Vol.24, No.9, pp.1520--1530, 2016.
  (Cover Paper, IEEE Signal Processing Society Japan Student Journal Paper Award)

- L. Wang, H. Chen, **S. Li**, and H. Meng. Phoneme-level articulatory animation in pronunciation training, Speech Communication (SPEECH COMMUN), Vol. 54, Issue 7, Sept. pp. 845--856, 2012.

International Conferences (Peer reviewed):

- **S. Li**, J. Li, Correction while Recognition: Combining Pretrained Language Model for Taiwan-accented Speech Recognition. in Proc. International Conference on Artificial Neural Networks (ICANN), 2023.

- Z. Yang, S. Shimizu, W. Zhou, **S. Li**, C. Chu, The Kyoto Speech-to-Speech Translation System for IWSLT 2023. in Proc. International Conference on Spoken Language Translation (IWSLT), 2023.

- L. Yang, J. Li, **S. Li**, T. Shinozaki, Multi-Domain Dialogue State Tracking with Disentangled Domain-Slot Attention. in Proc. ACL, (findings), 2023.

- S. Shimizu, C. Chu, **S. Li**, S. Kurohashi, Towards Speech Dialogue Translation Mediating Speakers of Different Languages. in Proc. ACL, (findings), 2023.

- C. Tan, Y. Cao, **S. Li**, M. Yoshikawa, General or Specific? Investigating Effective Privacy Protection in Federated Learning for Speech Emotion Recognition.
  in Proc. IEEE-ICASSP, 2023.

- K. Wang, Y. Yang, H. Huang, Y. Hu, **S. Li**, SpeakerAugment: Data Augmentation for Generalizable Source Separation via Speaker Parameter Manipulation. in Proc. IEEE-ICASSP, 2023.

- Y. Yang, H. Xu, H. Huang, E.S. Chng, **S. Li**, Speech-Text Based Multi-Modal Training with Bidirectional Attention for Improved Speech Recognition. in Proc. IEEE-ICASSP, 2023.

- K. Soky, **S. Li**, C. Chu, T. Kawahara, Domain and Language Adaptation Using Heterogeneous Datasets for Wav2vec2.0-Based Speech Recognition of Low-Resource Language. in Proc. IEEE-ICASSP, 2023.

- Q. Liu, Z. Gong, Z. Yang, Y. Yang, **S. Li**, Ding C. Chen, N. Minematsu, H. Huang, F. Cheng, C. Chu, S. Kurohashi, Hierarchical Softmax for End-To-End Low-Resource Multilingual Speech Recognition. in Proc. IEEE-ICASSP, 2023. (Travel Granted by IEEE-SPS)

- L. Yang, J. Li, **S. Li** and T. Shinozaki, Multi-Domain Dialogue State Tracking with Top-k Slot Self Attention. in Proc. SIGdial Meeting Discourse \& Dialogue, pp. 231--236, 2022.

- K. Soky, **S. Li**, M. Mimura, C. Chu and T. Kawahara, Leveraging Simultaneous Translation for Enhancing Transcription of Low-resource Language via Cross Attention Mechanism. in Proc. INTERSPEECH, pp. 1362--1366, 2022.

- L. Yang, W. Wei, **S. Li**, J. Li and T. Shinozaki, Augmented Adversarial Self-Supervised Learning for Early-Stage Alzheimer's Speech Detection. in Proc. INTERSPEECH, pp. 541--545, 2022.

- K. Li, **S. Li (corresponding)**, X. Lu, M. Akagi, M. Liu, L. Zhang, C. Zeng, L. Wang, J. Dang and M. Unoki, Data Augmentation Using McAdams-Coefficient-Based Speaker Anonymization for Fake Audio Detection. in Proc. INTERSPEECH, pp. 664--668, 2022.

- Z. Yang, W. Zhou, C. Chu, **S. Li**, R. Dabre, R. Rubino and Y. Zhao, Fusion of Self-supervised Learned Models for MOS Prediction. in Proc. INTERSPEECH, pp. 5443--5447, 2022.

- S. Qin, L. Wang, **S. Li**, Y. Lin and J. Dang, Finer-grained Modeling units-based Meta-Learning for Low-resource Tibetan Speech Recognition. in Proc. INTERSPEECH, pp. 2133--2137, 2022.

- Z. Gong, D. Saito, L. Yang, T. Shinozaki, **S. Li**, H. Kawai and N. Minematsu, Self-Adaptive Multilingual ASR Rescoring with Language Identification and Unified Language Model. In Proc. ISCA-Odyssey (The Speaker and Language Recognition Workshop), pp. 415--420, 2022.

- **S. Li**, J. Li, Q. Liu, Z. Gong, Adversarial Speech Generation and Natural Speech Recovery for Speech Content Protection. in Proc. LREC (Language Resources and Evaluation Conference), pp. 7291--7297, 2022.

- Y. Lv, L. Wang, M. Ge, **S. Li (corresponding)**, C. Ding, L. Pan, Y. Wang, J. Dang, K. Honda, Compressing Transformer-based ASR Model by Task-driven Loss and Attention-based Multi-level Feature Distillation. in Proc. IEEE-ICASSP, pp. 7992--7996, 2022.

- K. Soky, M. Mimura, T. Kawahara, **S. Li**, C. Ding, C. Chu, and S. Sam. Khmer Speech Translation Corpus of the Extraordinary Chambers in the Courts of Cambodia (ECCC).In Proc. O-COCOSDA, pp. 122--127, 2021. (Best student paper nomination, invited as fast tracking journal paper)

- D. Wang, S. Ye, X. Hu, **S. Li**, and X. Xu, An End-to-End Dialect Identification System with Transfer Learning from a Multilingual Automatic Speech Recognition Model. in Proc. INTERSPEECH, pp. 3266--3270, 2021.

- S. Chen, X. Hu, **S. Li**, and X. Xu, An investigation of using hybrid modeling units for improving End-to-End speech recognition systems. in Proc. IEEE-ICASSP, pp. 6743--6747, 2021.

- Y. Lin, L. Wang, **S. Li**, J. Dang, and C. Ding. Staged Knowledge Distillation for End-to-End Dysarthric Speech Recognition and Speech Attribute Transcription. In Proc. INTERSPEECH, pp. 4791--4795, 2020 (Travel Granted by ISCA).

- H. Shi, L. Wang, **S. Li (corresponding)**, C. Ding, M. Ge, N. Li, J. Dang, and H. Seki. Singing Voice Extraction with Attention based Spectrograms Fusion. In Proc. INTERSPEECH, pp. 2412--2416, 2020 (Travel Granted by ISCA).

- **S. Li**, X. Lu, R. Dabre, P. Shen, and H. Kawai. Joint Training End-to-End Speech Recognition Systems with Speaker Attributes. In Proc. ISCA-Odyssey (The Speaker and Language Recognition Workshop), pp. 385--390, 2020.

- Y. Han, **S. Li (corresponding)**, Y. Cao, Q. Ma, and M. Yoshikawa. Voice-Indistinguishability: Protecting Voiceprint in Privacy Preserving Speech Data Release. In Proc. IEEE-ICME, pp. 1--6, 2020. (Best student paper nomination, select as fast tracking journal paper of IEEE Trans. Multimedia (TMM))

- Y. Lin, L. Wang, J. Dang, **S. Li (corresponding)**, and C. Ding. End-To-End Articulatory Modeling for Dysarthria Articulatory Attribute Detection. In Proc. IEEE-ICASSP, pp. 7349--7353, 2020.

- H. Shi, L. Wang, M. Ge, **S. Li (corresponding)**, and J. Dang. Spectrograms Fusion with Minimum Difference Masks Estimation for Monaural Speech Dereverberation. In Proc. IEEE-ICASSP, pp. 7544--7548, 2020.

- **S. Li**, X. Lu, C. Ding, P. Shen, T. Kawahara, and H. Kawai. Investigating Radical-based End-to-End Speech Recognition Systems for Chinese Dialects and Japanese. In Proc. INTERSPEECH, pp. 2200--2204, 2019.

- **S. Li**, C. Ding, X. Lu, P. Shen, T. Kawahara, and H. Kawai. End-to-End Articulatory Attribute Modeling for Low-resource Multilingual Speech Recognition. In Proc. INTERSPEECH, pp. 2145--2149, 2019.

- **S. Li**, R. Dabre, X. Lu, P. Shen, T. Kawahara, and H. Kawai. Improving Transformer-based Speech Recognition Systems with Compressed Structure and Speech Attributes Augmentation. In Proc. INTERSPEECH, pp. 4400--4404, 2019.

- **S. Li**, X.Lu, R.Takashima, P.Shen, T.Kawahara, and H.Kawai. Improving very deep time-delay neural network with vertical-attention for effectively training CTC-based ASR systems. In Proc. IEEE Spoken Language Technology Workshop (IEEE-SLT), pp. 77--83, 2018.

- **S. Li**, X.Lu, R.Takashima, P.Shen, T.Kawahara, and H.Kawai. Improving CTC-based Acoustic Model with Very Deep Residual Time-delay Neural Networks. In Proc. INTERSPEECH, pp. 3708--3712, 2018.

- **S. Li**, X.Lu, P.Shen, R.Takashima, T.Kawahara, and H.Kawai. Incremental training and constructing the very deep convolutional residual network acoustic models. In Proc. IEEE Workshop Automatic Speech Recognition \& Understanding (IEEE-ASRU), pp. 222--227, 2017.

- P. Shen, X. Lu, **S. Li**, and H. Kawai. Conditional Generative Adversarial Nets Classifier for Spoken Language Identification. In Proc. INTERSPEECH, pp. 2814--2818, 2017.

- **S. Li**, X.Lu, S.Sakai, M.Mimura, and T.Kawahara. Semi-supervised ensemble DNN acoustic model training. In Proc. IEEE-ICASSP, pp. 5270--5274, 2017.

- **S. Li**, Y.Akita, and T.Kawahara. Data selection from multiple ASR systems' hypotheses for unsupervised acoustic model training. In Proc. IEEE-ICASSP, pp. 5875--5879, 2016.

- **S. Li**, Y.Akita, and T.Kawahara. Discriminative data selection for lightly supervised training of acoustic model using closed caption texts. In Proc. INTERSPEECH, pp. 3526--3530, 2015.(oral)

- **S. Li**, X.Lu, Y.Akita, and T.Kawahara.Ensemble speaker modeling using speaker adaptive training deep neural network for speaker adaptation. In Proc. INTERSPEECH, pp. 2892--2896, 2015.

- **S. Li** and L. Wang.Cross Linguistic Comparison of Mandarin and English EMA Articulatory Data, In Proc. INTERSPEECH, pp. 903--906, 2012. (Travel granted by IBM research)

Challenges・Demo (selected):

- **S. Li**, R. Dabre, R. Raphael, W. Zhou, Z. Yang, C. Chu, Y. Zhao.
The System Description for VoiceMOS Challenge 2022 (KK team, main/ood tasks). (co-top1)

- D. Wang, S. Ye, X. Hu, **S. Li**.
The RoyalFlush-NICT System Description for AP21-OLR Challenge (Silk-road team, full tasks).
In OLR2021 (oriental language recognition challenge), 2021. (top3)

- Y. Han, Y. Cao, **S. Li**, Q. Ma, and M. Yoshikawa.
  Voice-Indistinguishability: Protecting Voiceprint with Differential Privacy under an Untrusted Server.
  ACM conference on Computer and Communications Security (CCS), demo, pp. 2125--2127, 2020.

- H. Zhang, **S. Li**, X. Ma, Y. Zhao, Y. Cao, and T. Kawahara,
  Phantom in the Opera: Effective Adversarial Music Attack on Keyword Spotting Systems.
  in Proc. IEEE-SLT, 2021 (demo session, introduction).

Patents:

- 発明者:**李勝**、ルーシュガン、高島遼一、沈鵬、河井恒・出願人/特許権者：国立研究開発法人情報通信研究機構,
  学習方法・特願2017-236626・特開2019-105899・特許番号「6979203」,
  出願2017年12月11日・公開2019年6月21日, 特許登録日2021年11月17日.

- 発明者:高島遼一、**李勝**、河井恒・出願人/特許権者：国立研究開発法人情報通信研究機構,
  時系列情報の学習システム、方法および ニューラルネットワークモデル・特願2018-044134・特開2019-159654・特許番号「7070894」,
  出願2018年03月12日・公開2018年03月12日・特許登録日2022年5月10日.

- 発明者:**李勝**、ルーシュガン、高島遼一、沈鵬、河井恒・出願人/特許権者：国立研究開発法人情報通信研究機構,
  音声認識システム、音声認識方法、学習済モデル・特願2018-044491・特開2019-159058・特許番号「7070894」,
  出願 2018年03月12日・公開2018年3月12日, 特許登録日2022年7月22日.

- 発明者:**李勝**、ルーシュガン、高島遼一、沈鵬、河井恒・出願人/特許権者：国立研究開発法人情報通信研究機構,
  識別器、学習済モデル、学習方法・特願2018-142418・特開2020-020872・特許番号「7209330」,
  出願2018年07月30日・公開2018年07月30日, 特許登録日2023年1月12日.

- 発明者:沈鵬、ルーシュガン、**李勝**、河井恒・出願人/特許権者：国立研究開発法人情報通信研究機構,
  言語識別モデルの訓練方法及び装置、並びにそのためのコンピュータプログラム・特願2019-086005・特開2020-038343・特許番号「7282363」,
  出願2019年04月26日・公開2019年04月26日, 特許登録日2023年5月19日.

Academic Services:

Academic Membership:

IEEE/IEEE-SPS (Signal Processing Society),
ISCA (International Speech Communication Association),
ASJ (Acoustical Society of Japan),
SIG-CSLP (Chinese Spoken Language Processing),
APSIPA (Asia Pacific Signal and Information Processing Association)

Chairing and organizing:

[1] Session Chair of INTERSPEECH2020 session: Topics of ASR I
[2] Co-organizing INTERSPEECH2020 workshop: Spoken Language Interaction for Mobile Transportation System (SLIMTS)
[3] Session Chair of Speaker Odyssey2022 session: Evaluation and Benchmarking (EB)
[4] Co-organizing Coling2022 workshop: when creative ai meets conversational ai (cai + cai = cai^2)
[5] Co-organizing ACM Multimedia Asia 2023 workshop: M3Oriental
[6] Area Chair of APSIPA 2023
[7] Area Chair of EMNLP 2023

Reviewer/Program committee:
Journal:

[1] IEEE/ACM Trans. Audio, Speech \& Language Process.
[2] Computer Speech and Language
[3] Speech Communication
[4] IEICE transactions, letters
[5] APSIPA transactions
[6] Applied Acoustics
[7] Transactions on Asian and Low-Resource Language Information Processing (TALLIP)

Conferences:

[1] ICASSP-2021/2022/2023, INTERSPEECH-2015/2018/2019/2020/2021/2022/2023, SLT-2022, ASRU-2023
[2] APSIPA-2019/2020/2021/2022/2023, IJCNN-2023, ICONIP2023
[3] BC_VCC-2020 (Blizzard Challenge and Voice Conversion Challenge 2020)
[4] ACL-2017/2018/2020, EACL-2020/2022, NAACL-HLT-2016/2018/2019/2021
[5] IJCNLP-2017, EMNLP-IJCNLP-2019, EMNLP-2020/2021/2022, AACL-IJCNLP-2020/2022/2023, COLING-2018/2022
[6] NLP-2022/2023
[7] AAAI-2019, ICLR-2021, NeurIPS-2022/2023, ICML-2023
[8] IROS-2019, Ubiquitous Robots (UR)-2020, IEEE-ROMAN 2023
[9] ICME-2020/2021/2022/2023(main+workshop), ACM Multimedia 2021/2022/2023, ACM Multimedia Asia 2021/2022/2023
[10] PAKDD-2023

last update: 8/14/2023