

Naver, 9 papers accepted for 'Interspeech 2021'!... 'Global AI research innovation achievements'

Reported by Reporter Park Hyun-jin | Approved 2021.09.06 09:30

The most papers adopted among domestic companies... 14 papers announced by LINE combined



The global artificial intelligence (AI) research innovation ecosystem, which is expanding centered around Naver, is producing excellent AI research results and achieving results.

Naver presented nine papers, the most among domestic companies, at 'INTER_SPEECH 2021', the world's largest speech and signal processing conference held from the 30th of last month to September 3rd.

The total number of papers published by Naver's Japanese affiliate 'Line', which is part of Naver's global AI R&D ecosystem, is 14. Based on these achievements, Naver plans to actively pursue securing global AI leadership.

'Interspeech', which is in its 22nd year this year, is a place where the latest voice recognition technologies of global companies such as Google, Facebook, Amazon, and Alibaba are shared, and it is recognized as the world's largest speech and signal processing conference along with ICASSP. Naver and Line also presented 14 papers at 'ICASSP 2021' held in June of this year.

Naver's research papers accepted at 'Interspeech 2021' cover various fields related to voice and signals, such as voice recognition, voice synthesis, and dataset creation. Some of the papers are applied to Naver's actual services, providing users with a more convenient service experience.

Research on voice synthesis quality improvement technology has been applied to various voice synthesis services provided by Naver Clova, including 'Clova Dubbing', 'Clova Smart Speaker', and 'Naver AiCALL'.

Research on techniques to more accurately learn the characteristics of each speaker for 'speaker diarisation' in situations where multiple speakers are speaking simultaneously is being utilized to improve the 'Clova Note' service.

In addition, five of the nine papers published by Naver were conducted in collaboration with domestic and international companies and research institutes leading in AI technology, including EURECOM, a French research institute that is considered a leader in the fields of computer science and information systems, Carnegie Mellon University, KAIST, Yonsei University, and Line.

Naver added that its efforts to build a global AI research and innovation ecosystem are showing visible results.

Based on this achievement, Naver plans to continuously expand the global AI R&D ecosystem. Currently, in addition to Japan, which is centered on collaboration with Line, a joint research center has been established with HUST and PTIT in Vietnam, and research collaboration is ongoing with

Naver Labs Europe in France in Europe.

Last July, we also agreed to establish a joint research center with the University of Tübingen in Germany, a world leader in artificial neural networks and robotics.

Jeong Seok-geun, CEO of Naver Clova CIC, said, "Based on active investment in R&D, Naver's AI research innovation ecosystem is further expanding in size." He added, "Just as Naver, Line, and global AI researchers achieved outstanding results at Interspeech 2021, we expect that new results will emerge from the more diverse AI R&D ecosystem that will be built around Naver in the future."

For reference, the list of papers accepted by Naver INTERSPEECH 2021 is as follows.

▷ High-fidelity Parallel WaveGAN with Multi-band Harmonic-plus-Noise Model - Minje Hwang, Ryuichi Yamamoto (LINE), Eunwoo Song, Jaemin Kim (Improving the quality and stability of vocoder by applying multi-band harmonic-plus-noise model, one of the voice modeling techniques, to Parallel WaveGAN)

▷ LiteTTS: A Decoder-free Light-weight Text-to-wave Synthesizer Based on Generative Adversarial Networks - Kim Nguyen (Yonsei University), Jeong Gi-hyeok, Eom Se-yeon, Hwang Min-je, Song Eun-woo, Kang Hong-gu. (Propose a text-to-wave model that combines the TTS pipelines separated into text-to-feature and feature-to-wave)

▷ Label Embedding for Chinese Grapheme-to-Phoneme Conversion-Eunbi Choi (KAIST), Hwayeon Kim, Jonghwan Kim, Jaemin Kim (proposing a label embedding approach for the Chinese polyphone conversion problem)

▷ Look Who's Talking: Active Speaker Detection in the Wild- Kim Yu-jin, Hee-soo Heo, So-yeon Choi, Su-hwan Jeong, Yu-hwan Kwon, Bong-jin Lee, Young-ki Kwon, Jun-seon Jeong (Multimodal Active Speaker Detection Dataset Creation and Publication)

▷ Adapting Speaker Embeddings for Speaker Diarisation- Kwon Young-ki, Jeong Ji-won, Hee-soo, Kim Yu-jin, Lee Bong-jin, Jeong Jun-seon (Study on Speaker Embedding Enhancement Method for Improving Speaker Diarization Performance)

- ▷ Three-class Overlapped Speech Detection using a Convolutional Recurrent Neural Network- Jiwon Jeong, Heesu Heo, Yeongki Kwon, Junseon Jeong, Bongjin Lee (Proposal of Overlapped Speech Detection Technique Using 3 Classes + CRNN)
- ▷ Graph Attention Networks for Anti-Spoofing- Hemlata Tak (EURECOM), Jeong Ji-won, Jose Patino (EURECOM), Massimiliano Todisco, Nicholas Evans (A study on audio spoofing detection using GNN)
- ▷ DEMUCS-Mobile: On-device lightweight speech enhancement- Irukas, Jiyuna, Minjae Lee, Minseok Choi (Proposing a lightweight technique for CNN deep learning-based noise removal model and implementing a noise removal model that can operate in real time on a mobile on-device)
- ▷ Layer Pruning on Demand with Intermediate CTC- Jaesong Lee, Jingu Kang, Shinji Watanabe (CMU). (Learning methodology for pruning Transformer+CTC models without fine-tuning)



Reporter Park Hyun-jin nwngm@aitimes.kr

Blog & SNS

Line-Naver, 14 papers accepted by 'ICASSP', the most prestigious speech signal processing society

2021.02.26 ALL

- A total of 14 papers were adopted, including 2 joint research papers from Line and Naver
- Proven excellence in research on improving voice synthesis quality and voice recognition rate

February 26, 2021 – LINE Corporation (hereinafter referred to as LINE, Representative Director and President Shin Jung-ho and Takeshi Idezawa) and NAVER Corporation (hereinafter referred to as Han Sung-sook) announced that 14 papers from LINE and NAVER were accepted at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), the most prestigious international conference on acoustics, speech, and signal processing.

The ICASSP, now in its 46th year, is the world's largest international conference in the fields of acoustics, speech, and signal processing, hosted by the IEEE Signal Processing Society. More than 3,600 papers were registered this year, and about 1,700 of them were accepted and will be presented at 'ICASSP 2021', which will be held online in June.

Focusing on basic research in the fields of acoustics, voice, and signal processing, including voice synthesis technology, sound source separation technology, and environmental sound identification technology.

Line is working closely with Naver Clova and AI LAB in the AI R&D and technology utilization fields to accelerate the development of AI business, one of its strategic businesses. Each team in charge of data-based development, data analysis, machine learning, AI technology development, and basic research is strengthening the cycle from research, development, and commercialization by linking across their respective areas.

In addition, Line and Naver are working hard on basic research to support the development of AI-related services and functions, and are focusing on machine learning-based voice processing, language processing, and image processing. For audio, voice, and signal processing, they are continuously researching 'voice synthesis technology' that realizes fast speed and high quality by utilizing 'Parallel WaveGAN' using GPU, 'sound source separation technology' that separates

sounds occurring in the surroundings using machines.

Demonstrated excellence in research on improving voice synthesis quality and improving voice recognition rate

At this 'ICASSP 2021', a study was introduced that utilized the information of voiced and unvoiced sounds in the non-autoregressive voice generation model*² 'Parallel WaveGAN' using the 'Generative Adversarial Network (GAN)'*¹ related to voice synthesis to improve the discriminator. The existing 'Parallel WaveGAN' using a single discriminator had a problem that the quality deteriorated when applied to a corpus of multiple speakers due to the limited expressiveness and learning of the discriminator. In this study, taking note of the different phonetic properties of voiced and unvoiced sounds, an identifier suitable for the properties of each voice was designed, significantly improving the quality of the synthesized voice. In addition, although it was a sequential study, it was highly evaluated that the validity of the approach was verified by conducting a large-scale subjective evaluation experiment targeting a total of four male and female speakers.

In addition, we proposed a new method that combines Iterative Source Steering (ISS), a source separation method that does not use deep learning for source separation, and a source model estimation method using deep learning. In addition to significantly improving the voice recognition rate compared to the existing ISS, we proved the superiority of this approach in that it is a framework that can be applied regardless of the number of sound sources.

*¹ One of the classification models of machine learning, it learns using two neural networks and generates new similar data through input data and images.

*² A model that generates voices at each point in time without relying on voices generated in the past. It is computationally efficient because parallel processing is possible.

We will increase the value of our services through active AI basic research.

LINE Clova and Naver Clova, joint AI brands of LINE and Naver, aim to improve the quality of life of users and create a more convenient and prosperous world by solving the inconveniences of daily life and business through various AI technologies and services. Currently, they are providing solutions utilizing LINE and Naver's voice recognition technology 'CLOVA Speech' and voice synthesis technology 'CLOVA Voice'.

'AiCall', a representative solution of Line Clova and Naver Clova, is a service in which AI naturally responds to user requests by combining Clova Speech, Clova Voice, and a conversation control system. It is utilized in administrative agencies and restaurants. In addition, 'CLOVA Note', launched last year, is a service that recognizes conversations in various situations such as interviews or business meetings and records and manages the content. It implements a voice

for teachers who are conducting remote classes due to COVID-19 to produce video materials. Voice recognition and synthesis technologies were also applied to 'Clova Lamp', an AI light that reads books.

Going forward, Line and Naver plan to further improve the quality of existing services and introduce new functions and services through active basic AI research.

Accepted paper (jointly by Line and Naver)

- PARALLEL WAVEFORM SYNTHESIS BASED ON GENERATIVE ADVERSARIAL NETWORKS WITH VOICING-AWARE CONDITIONAL DISCRIMINATORS / R. Yamamoto, E. Song, M. Hwang, and J. Kim
- TTS-BY-TTS: TTS-DRIVEN DATA AUGMENTATION FOR FAST AND HIGH-QUALITY SPEECH SYNTHESIS / M. Hwang, R. Yamamoto, E. Song, and J. Kim

Accepted paper (line)

- END TO END LEARNING FOR CONVOLUTIVE MULTI-CHANNEL WIENER FILTERING / M. Togami
- DISENTANGLED SPEAKER AND LANGUAGE REPRESENTATIONS USING MUTUAL INFORMATION MINIMIZATION AND DOMAIN ADAPTATION FOR CROSS-LINGUAL TTS / D. Xin, T. Komatsu, S. Takamichi, H. Saruwatari
- SURROGATE SOURCE MODEL LEARNING FOR DETERMINED SOURCE SEPARATION / R. Scheibler, M. Togami
- REFINEMENT OF DIRECTION OF ARRIVAL ESTIMATORS BY MAJORIZATION-MINIMIZATION OPTIMIZATION ON THE ARRAY MANIFOLD / R. Scheibler, M. Togami
- JOINT DEREVERBERATION AND SEPARATION WITH ITERATIVE SOURCE STEERING / T. Nakashima, R. Scheibler, M. Togami, N. Ono

Adopted paper (Naver)

- NN-KOG2P: A NOVEL GRAPHEM-TO-PHONEME MODEL FOR KOREAN LANGUAGE / Kim Hwa-yeon, Kim Jong-hwan, Kim Jae-min

-
- PLAYING A PART: SPEAKER VERIFICATION AT THE MOVIES / Andrew Brown, Jae-seong Huh, Arsha Nagrani, Jun-seon Jeong, Andrew Zisserman
 - GRAPH ATTENTION NETWORKS FOR SPEAKER VERIFICATION / Jeong Ji-won, Hee-soo Heo, Ha-jin Yoo, Jun-seon Jeong
 - INTERMEDIATE LOSS REGULARIZATION FOR CTC-BASED SPEECH RECOGNITION / Lee Jae-song, Shinji Watanabe
 - TWO-STAGE TEXTUAL KNOWLEDGE DISTILLATION TO SPEECH ENCODER FOR Spoken Language Understanding / Kim Seong-bin, Kim Gyu-wan, Shin Seong-jin, Lee Sang-min
 - ST-BERT: CROSS-MODAL LANGUAGE MODEL PRE-TRAINING FOR END-TO-END SPOKEN LANGUAGE UNDERSTANDING / Kim Min-jung, Kim Gyu-wan, Lee Sang-woo, Ha Jung-woo

■ Introduction to Line Corporation

LINE Corporation is building a global smart portal that brings together people, information, content, and services with the mission of “Closing The Distance.” Since launching the LINE messenger service in June 2011, it has expanded into various fields such as AI technology and fintech, creating a global mobile ecosystem.

List

[privacy policy](#) [LINE Partners](#) [Contact](#)

©LINE Plus Corporation