* **강연제목:** An overview of contemporary fully neural end-to-end speech recognition and text-to-speech
* **연사 성함: 김찬우** (Chanwoo Kim)
* **연사 소속 및 직함: 삼성리서치 부사장**
* **Abstract:** (100단어 내외)

In this talk, we give an overview of the latest end-to-end Automatic Speech Recognition (ASR) and Text-To-Speech (TTS) algorithms. We also discuss optimization techniques to reduce the model size and obtain better performance. Conventional ASR and TTS systems consist of multiple handcrafted components. However, the introduction of fully neural sequence-to-sequence technologies has greatly simplified the structure while significantly improving the performance. In this talk, we give an overview of the important end-to-end ASR structures including a stack of neural network layers with a Connectionist Temporal Classification (CTC) loss, Recurrent Neural Network Transducer (RNN-T), Transformer Transducer, and Conformer Transducer (Conformer-T), and models based on Attention-based Encoder-Decoder (AED). We also describe well-known TTS models including Tacotron, Tacotron 2, and Deep Convolutional Text-To-Speech (DC-TTS).

* **Biography:**

Chanwoo Kim has been a corporate executive vice president at Samsung research leading the language and voice team. He joined Samsung research as a corporate vice president heading the speech processing Lab in Feb. 2018. He has been leading research on end-to-end speech recognition, end-to-end text-to-speech (TTS), machine translation, Natural Language Understanding (NLU), Language Modeling (LM) and Question Answering (QA), speech enhancement, key-word spotting, and so on at Samsung Research. Most of these research outcomes have been commercialized for Samsung products. He was a software engineer at the Google speech team between Feb. 2013 and Feb. 2018. He worked for acoustic modeling for speech recognition systems and enhancing noise robustness using deep learning techniques. While working for Google, he contributed to data-augmentation and acoustic modeling of Google speech recognition systems. He contributed to the commercialization of various Google AI speakers and google speech recognition systems. He was a speech scientist at Microsoft from Jan. 2011 to Jan. 2013. Dr. Kim received his Ph. D. from the Language Technologies Institute of School of Computer Science Carnegie Mellon University in Dec. 2010. He received his B.S and M.S. degrees in Electrical Engineering from Seoul National University in 1998 and 2001, respectively. Dr. Kim’s doctoral research was focused on enhancing the robustness of automatic speech recognition systems in noisy environments. Between 2003 and 2005 Dr. Kim was a Senior Research Engineer at LG Electronics, where he worked primarily on embedded signal processing and protocol stacks for multimedia systems. Prior to his employment at LG, he worked for EdumediaTek and SK Teletech as an R&D engineer.

**- 학력**

* 2005 ~ 2010 Carnegie Mellon University (CMU). School of Computer Science (SCS), Language Technologies Institute (LTI) 박사
* 1998 ~ 2001 서울대학교 전기컴퓨터 공학부 석사
* 1994 ~1998 서울대학교 전기공학부 학사

**- 주요경력 (5개 이내)**

* 2018/02 ~ 현재 Samsung Research (삼성리서치) Executive Vice President (부사장)

Head of Language & Voice Team (언어 & 음성팀장).

* 2016/06 ~ 2018/02 Google, Senior Software Engineer, Google Speech Team
* 2011/01 ~ 2013/01 Microsoft, Speech Scientist, Microsoft Speech Team
* 2003/06 ~ 2005/08 LG Electronics, Senior Research Engineer, Mobile Communication
* 2000/06 ~ 2002/07 Edumediatek, Research Engineer,

**- 상훈 (5개 이내)**

* (1st author): 2019 IEEE Signal Processing Society Best Paper Award (<https://signalprocessingsociety.org/newsletter/2020/01/2019-ieee-signal-processing-awardees>) C. Kim and R. M. Stern, “Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition”, IEEE/ACM Transactions on Audio, Speech, and Language Processing
* (1st author): 17-th Samsung Humantech Thesis Bronze Prize, C. Kim and R. M. Stern, ”Power-

Normalized cepstral coefficients for robust speech recognition”, Feb. 2011.

* (1st author): 16-th Samsung Hamantech Thesis Honour Prize, C. Kim and R. M. Stern, ”Smallpower boosting and spectral subtraction for robust speech recognition”, Feb. 2010