An ASR Framework for Robust Recognition in Low-Bitrate Environments

Wootaek Lim, Sooyoung Park, Inseon Jang

Electronics and Telecommunications Research Institute (ETRI), Korea

wtlim@etri.re.kr, sooyoung@etri.re.kr, jinsn@etri.re.kr

Abstract— **In this paper, we propose an automatic speech recognition (ASR) framework for robust recognition in low-bitrate environments, developed within the context of machine listening. The framework integrates a pre-trained voice activity detector for silence removal, a generalized neural codec for efficient low-bitrate compression, and a neural post-enhancement module to refine the decoded signal prior to recognition by the ASR model. Unlike feature-based transmission systems tailored to specific tasks, this approach leverages generalized codecs to ensure broad applicability and compatibility with existing systems. The post-enhancement model, trained on paired clean and decoded speech, focuses on reducing spectral distortion and suppressing coding noise without relying on perceptual losses. Experiments on the LibriSpeech corpus demonstrate that the proposed framework achieves an average bitrate reduction of 12.53% while maintaining or improving recognition accuracy compared to baseline codec performance, achieving word error rates of 7.18% with Lyra and 6.65% with DAC. These results confirm the effectiveness of the proposed framework for machine-oriented speech communication under low-bitrate conditions.**

Keyword— **Automatic speech recognition, low-bitrate speech coding, voice activity detection, neural codec, speech enhancement**

Wootaek Lim received the Ph.D. degree in the Graduate School of Culture Technology at the Korea Advanced Institute of Science and Technology (KAIST), South Korea, and the B.S. and M.S. degrees in Electronic Engineering from Kwangwoon University, Seoul, South Korea, in 2010 and 2012, respectively.
Since 2012, he has been a Senior Researcher at the Electronics and Telecommunications Research Institute (ETRI). His research interests include audio signal processing, machine learning, and deep learning.

Sooyoung Park received the B.S. degree in Industrial and Systems Engineering and Electrical Engineering (double major) from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2016, and the M.S. degree in Electrical Engineering from KAIST in 2018. Since 2018, he has been a Researcher at the Electronics and Telecommunications Research Institute (ETRI), South Korea. His research interests include multimodal deep learning and neural audio codecs.

Inseon Jang received the B.S. degree in Electrical and Electronic Engineering from Chungbuk National University, Cheongju, South Korea, in 2001, the M.S. degree in Computer Science and Engineering from POSTECH, Pohang, South Korea, in 2004, and the Ph.D. degree in Electronic Engineering from Chungnam National University, Daejeon, South Korea, in 2018. Since 2004, she has been a Principal Researcher at the Electronics and Telecommunications Research Institute (ETRI). Her research interests include audio coding and audio signal processing