

## MLM-based Misrecognized Word Correction for Speech Recognition

Yonghun Jang\*, Jung Min Lim\*\*, Seong-Guk Nam\*\*\*, Minhyung Ryu\*\*\*, Eunjin Yoo\*\*\*,  
Myung-Sub Lee\*\*\*\*, Jong Wook Kwak\*\*\*\*\*

\*Director, R&D Center, NEARNETWORKS Co., Ltd., Daegu, Korea

\*\*Researcher, Dept. of Computer Engineering, Yeungnam University, Gyeongsan, Korea

\*\*\*Researcher, R&D Center, NEARNETWORKS Co., Ltd., Daegu, Korea

\*\*\*\*Professor, Dept. of Software Convergence, Yeungnam University College, Daegu, Korea

\*\*\*\*\*Professor, Dept. of Computer Engineering, Yeungnam University, Gyeongsan, Korea

## [Abstract]

In this study, we propose an integrated approach to improving the accuracy of Korean speech recognition by addressing phonetic similarity-induced misrecognitions. The proposed system combines three key components: (1) enhancing the signal-to-noise ratio through frequency-domain noise reduction using Minimum Mean Square Error (MMSE)-based log-spectral estimation and a high-pass emphasis filter, (2) detecting contextually inappropriate words using KoBERT-based Masked Language Modeling (MLM), and (3) selecting the final correction word using Jamo-level Levenshtein Distance, which reflects the phonetic characteristics of the Korean language. In an experiment conducted on 1,000 Korean sentences containing misrecognized words, the proposed method reduced the Word Error Rate (WER) from 9.2% to 4.7% compared to the baseline. In addition, the proposed method achieved a maximum detection accuracy of 96.4% for misrecognized words. In conclusion, the proposed method was verified to significantly improve the performance of real-world speech recognition systems.

▶ **Key words:** Speech Recognition, Speech to Text, Error Correction, Language Model, Korean NLP

## [요 약]

본 연구에서는 음성인식 결과의 정확도를 향상하기 위해 발음 유사성에 기인한 오인식 단어를 효과적으로 교정할 수 있는 전후처리 기법을 제안한다. 제안 시스템은 세 단계로 구성된다. 첫째, MMSE(Minimum Mean Square Error) 기반 로그 스펙트럼 추정과 고주파 강조 필터를 적용한 주파수 영역 잡음 제거를 통해 신호대잡음비를 향상한다. 둘째, KoBERT 기반의 MLM(Masked Language Modeling)을 활용하여 문맥적으로 부적절한 단어를 탐지한다. 셋째, 한국어 음운 특성을 반영한 자모 단위 Levenshtein Distance 알고리즘을 이용하여 최종 교정 단어를 선택한다. 오인식 단어가 포함된 1,000개의 한국어 문장을 대상으로 한 실험에서, 제안 기법은 기존기법과 비교해 WER(Word Error Rate)을 9.2%에서 4.7%로 감소시켰다. 또한, 오인식 단어 탐지에서는 최대 96.4%의 탐지 정확도를 보였다. 결론적으로, 제안 기법을 통해 실제 음성인식 시스템의 품질을 유의미하게 향상할 수 있음을 입증한다.

▶ **주제어:** 음성인식, 음성텍스트변환, 오류교정, 언어모델, 한국어처리

- First Author: Yonghun Jang, Corresponding Author: Jong Wook Kwak
- \*Yonghun Jang (cto@aaf.co.kr), R&D Center, NEARNETWORKS Co., Ltd.
- \*\*Jung Min Lim (toygoon@yu.ac.kr), Dept. of Computer Engineering, Yeungnam University
- \*\*\*Seong-Guk Nam (nsg@aaf.co.kr), R&D Center, NEARNETWORKS Co., Ltd.
- \*\*\*Minhyung Ryu (tkj06039@aaf.co.kr), R&D Center, NEARNETWORKS Co., Ltd.
- \*\*\*Eunjin Yoo (ej.yoo@aaf.co.kr), R&D Center, NEARNETWORKS Co., Ltd.
- \*\*\*\*Myung-Sub Lee (skydream@ync.ac.kr), Dept. of Software Convergence, Yeungnam University College
- \*\*\*\*\*Jong Wook Kwak (kwak@yu.ac.kr), Dept. of Computer Engineering, Yeungnam University
- Received: 2025. 09. 22, Revised: 2025. 10. 03, Accepted: 2025. 10. 24.

## I. Introduction

STT(Speech-To-Text) 기술은 입력된 음성신호로부터 문자 기반 텍스트를 자동으로 생성하는 음성인식 기술의 핵심 구성 요소로, 사람의 음성 명령을 시스템이 이해 가능한 형태로 변환하는 데 활용된다. 해당 기술은 자동 음성인식(Automatic Speech Recognition, ASR) 시스템의 중심 기능으로, 음향 신호 처리, 발음 단위 인식, 언어모델 기반 텍스트 예측 등의 복잡한 처리 과정을 포함한다. 최근에는 딥러닝 기반의 End-to-End 음성인식 모델이 도입되면서, STT 시스템의 성능과 구조적 단순성이 동시에 개선되었으며, 이러한 기술은 음성 기반 정보 검색, 개인화 음성 인터페이스 등 다양한 응용 분야에서 실용적으로 활용되고 있다[1-3].

기존의 STT 시스템은 음성신호로부터 텍스트를 추출하는 인식 정확도 향상을 위해 음향 모델, 발음 모델, 언어모델 등 각 요소를 정교화하는 방향으로 연구가 진행되었다. 특히 CNN, LSTM, DNN 기반의 딥러닝 모델은 다양한 잡음 환경과 발화자의 음색 변이에 대한 강인성을 높이는 데 기여하였다[4-5]. 그러나, 이러한 구조는 낮은 신호 대 잡음비(Signal to Noise Ratio, SNR) 환경에서 오인식 문제를 빈번히 일으키며, 생성된 텍스트의 문법적 일관성이나 의미 해석의 정확성을 충분히 보장하지 못하는 한계를 가진다.

특히 한국어는 언어적 특수성으로 인해 STT 시스템에서의 인식 오류가 더욱 빈번하게 나타난다. 한국어는 문법 구조상 어미와 조사의 변화가 복잡하고, 띄어쓰기 기준이 명확하지 않으며, 초성-중성-종성의 복합적인 글자 구조를 가지고 있어 영어나 다른 언어와 비교하였을 때 음성인식의 난이도가 더 높다. 이러한 특성은 단순한 음성-텍스트 변환만으로는 사용자 수준에서 이해가 가능한 문장을 구성하는데 한계가 있으며, STT 출력 텍스트의 품질 저하로 이어진다. 그 결과, 정확한 발음이 입력되었음에도 불구하고 모델이 사전에 학습되지 않은 단어 또는 문맥 상 부적절한 단어를 출력하는 경우가 빈번히 발생한다.

이를 보완하기 위해, STT 시스템의 출력 결과를 대상으로 한 사후 교정 접근 방식이 연구되었다. 해당 방식은 음성신호 처리 단계에서 발생한 오인식 오류를 음성 자체가 아닌, 인식된 텍스트의 문맥 정보를 기반으로 보정함으로써 결과물의 문법적 일관성과 의미 해석 가능성을 높이는 것을 목표로 한다.

문맥 기반 단어 보정을 위한 연구는 기존에도 다수 제안되었으며, 최근에는 사전학습 언어모델(Pretrained Language Model)을 활용한 접근이 시도되고 있다. 이러한 방

법은 인식된 텍스트 내 단어의 문맥 적합성을 평가하고, 가장 적절한 교정 단어를 선택함으로써 STT 결과의 자연스러움을 향상하는데 이바지한다. 그러나 대부분의 기존 연구는 정형화된 문장 구조나 제한된 도메인 내에서 학습되어, 구어체, 신조어, 일상 표현, 노이즈 환경에서 발화 등 비정형 STT 출력에 대한 대응력은 여전히 제한적이라는 한계를 가진다.

STT의 출력 결과 텍스트는 대부분 기계 번역, 질의응답, 요약, 감정 분석 등 다양한 자연어 처리(Natural Language Processing) 시스템의 입력으로 이어지기 때문에, 그 정확도는 전체 처리 시스템의 성능을 결정짓는 핵심 요소로 간주된다. 특히 음성 기반 자동 번역 시스템의 경우, STT 결과의 오류는 번역 품질 저하로 직결되며, 문장의 종결 인식 실패, 조사 누락, 고유명사 오인식 등은 문맥 파악 실패 및 문법적 비문을 유발한다. 이와 같은 오류는 기계 번역 시스템의 사용자 이해도 및 서비스 품질의 저하로 이어진다. 따라서, STT 시스템의 출력 품질을 향상하는 작업은 단순한 인식 성능 향상을 넘어 번역 기술과 결합한 복합 시스템 구성에서 필수적인 고려 요소이다.

최근에는 Whisper 모델과 같은 오픈소스 기반의 고성능 STT 모델이 도입되면서, 다양한 언어에서도 비교적 높은 인식의 음성인식 정확도를 확보할 수 있게 되었다. Whisper 모델은 다양한 발화 조건에서도 높은 정확도를 보이는 LLM(Large Language Model)으로, 기존 시스템 대비 구조적으로 단순하면서도 강력한 성능을 제공하는 것으로 평가된다[6]. 그러나, Whisper는 한국어와 같이 문장 구조가 복잡하고 조사 및 어미변화가 잦은 언어에 대해선 인식 오류가 여전히 빈번히 발생한다.

본 연구의 목적은 Whisper 모델과 같은 고성능 STT 모델을 기반으로, 음성입력에 대한 전처리 및 텍스트 출력에 대한 후처리 기법을 통합 적용함으로써 음성인식 결과의 품질을 향상시키고, 이를 통해 기계 번역 시스템에서 요구되는 고품질 입력 텍스트를 생성하는 데 있다. 특히, 주파수 영역 기반의 음성 전처리와 문장 단위 절제, 문맥 기반 단어 교정으로 구성된 후처리 과정을 통해 번역 성능 향상에 실질적으로 기여할 수 있는 STT-번역 연계 처리 구조를 제안한다. 본 제안을 통해 Whisper 모델의 성능을 활용하여, MMSE 기반 전처리와 BERT 기반 후처리를 통합적으로 적용함으로써 음성인식 결과의 품질을 향상시키고, 결과적으로 기계 번역 등 후속 자연어 처리 시스템의 입력으로 적합한 고품질 텍스트 생성을 가능하게 한다. 이를 통해 단순한 음성인식 정확도 뿐만 아니라, 결과 텍스트가 기계 번역 시스템의 성능까지도 향상된다.

## II. Related Work

### 1. Speech Preprocessing & Noise Reduction

음성인식 시스템의 입력 신호는 다양한 잡음 환경에 노출될 수 있으며, 이는 인식 정확도를 저하시킬 수 있는 주요 요인 중 하나이다. 잡음 제거에는 MMSE 기반 로그 스펙트럼 추정 기법이 효과적이며 많은 분야에서 널리 사용되고 있다. 이 기법은 잡음이 포함된 관측 신호로부터 원래 음성신호의 스펙트럼 크기를 추정하는 방법으로, 평균 제곱 오차(Mean Square Error)를 최소화하는 조건 하에서 복원 계수를 계산한다. 특히 로그 스펙트럼 도메인에서의 추정은 인간 청각 시스템의 감각과 유사한 감쇠 곡선을 제공하기 때문에, 인식 정확도 향상에 유리하다.

MMSE 로그 스펙트럼 기반 잡음 제거는 일반적으로 다음과 같은 처리 절차를 따른다. 우선 입력 신호를 짧은 프레임 단위로 나누고, 각 프레임에 대해 윈도우 함수를 적용한 후 단시간 푸리에 변환(Short-Time Fourier Transform, STFT)을 수행한다. 초기 프레임을 활용하여 잡음의 파워 스펙트럼을 통계적으로 추정한 후, 음성신호의 파워 스펙트럼을 로그 도메인에서 베이지안 추론 기반으로 계산한다. 이후, 주파수별 MMSE 추정치를 통해 필터를 구성하고, 이를 잡음이 포함된 스펙트럼에 적용함으로써 음성신호의 왜곡을 최소화한다. 마지막으로 푸리에 역변환(Inverse STFT)과 Overlap-Add 방식으로 시간 도메인 신호를 재구성한다.

실제 대화 환경에서 수집된 자연스러운 음성 데이터를 기반으로 다단계 STFT 기반 전처리, 주파수 필터링, 정규화 기법을 조합하여 입력 신호의 품질을 안정화하고, 감정 인식 성능을 크게 향상한 기법이 제안된 바 있다[22]. 이 기법은 다양한 단위로 변환된 STFT 결과를 병렬처리하고, CNN 기반 모델에 입력함으로써 다중 해상도 특성을 효과적으로 학습하였다. 결과적으로 복잡한 잡음 환경에서도 안정적인 감정 분류 정확도를 보였으며, 음성 전처리 단계에서의 스펙트럼 정보 보존과 잡음 완화가 전체 음성 처리 파이프라인의 성능에 큰 영향을 미친다는 점을 실험적으로 검증하였다.

이와 별도로, 극한의 잡음 환경에서도 높은 음성인식 성능을 유지하기 위한 다른 접근으로, Wavelet 변환 기반의 전처리 방식과 CNN 기반 모델을 결합한 구조에 관한 연구가 제안되었다[23]. 해당 기법은 음성 입력 신호를 DWT(Discrete Wavelet Transform) 변환으로 시간-주파수 영역에서 세분화하고, 이를 다채널 CNN 구조에 입력하여 잡음 제거와 음성 복원을 동시에 수행하였다. 이처럼

변환 기반 전처리와 딥러닝 모델의 결합은 비정상 신호와 같은 다양한 환경에서 유효한 잡음 처리 기술로 평가된다.

또한, 스펙트럼 감산 기반 기법과 DWT 변환을 결합하여 하이브리드 접근을 통해 잡음 제거 성능을 개선한 연구가 제안되었다[24]. 해당 방식은 주파수 대역별로 잡음을 정밀하게 억제하여 높은 SNR 개선 효과를 보였으며, 다양한 환경에서의 적용 가능성을 확인하였다.

### 2. STT Systems

STT 시스템은 입력된 연속 음성신호를 대응하는 문자 기반 텍스트로 변환하는 음성인식 기술로, 음성 인터페이스, 디지털 비서, 음성 기반 번역 시스템 등 다양한 응용 분야에서 핵심적으로 사용된다. 그러나 전통적인 STT 시스템은 각 모듈 간의 오류 전파 문제와 모듈 간의 통합 학습이 불가능하다는 구조적 한계가 존재하였다[9]. 이러한 문제를 해결하기 위해, End-to-End 구조의 STT 시스템이 연구되었다. End-to-End STT 시스템은 음성 입력부터 텍스트 출력까지 전체 변환 과정을 하나의 통합된 신경망 모델로 처리하는 방식으로, 전통적인 파이프라인을 구성하는 각 요소들을 학습 가능한 단일 아키텍처로 통합한다. 대표적인 접근 방식으로는 CTC(Connectionist Temporal Classification) 기반 프레임 정렬 모델, RNN-Transducer 기법 등이 있으며, 이를 기반으로 한 다양한 개선 연구가 지속적으로 제안되었다[10-11].

최근에는 이러한 End-to-End STT 시스템이 LLM 기반으로 확장되어 Whisper 모델과 같은 대규모 사전학습 기반의 다국어 음성인식 모델이 활용되고 있다. Whisper 모델은 Encoder-Decoder 구조의 자기회귀 방식의 텍스트 디코더를 기반으로 하며, 다양한 언어와 환경을 학습한 거대 음성-텍스트 병렬 데이터 세트를 바탕으로 범용성과 안정성을 확보한 모델로 평가된다. 또한, Whisper 모델을 한국어와 영어의 혼합 발화 환경에서 확장한 연구도 보고되었다[13]. 해당 연구는 코드 스위칭 환경에서 Whisper 모델의 다국어 인식 성능을 향상시키기 위해, 영어 중심의 음향 모델에 한국어 데이터를 추가 학습시키는 구조적 확장 방식을 제안하였다. 이는 Whisper 모델의 구조적 확장 가능성을 실험적으로 입증하였다는 의미가 있다.

### 3. STT Errors and Post-Correction

STT 시스템은 음성 입력으로부터 자동으로 결과 텍스트를 생성하지만, 다양한 음성 조건과 모델 한계로 인해 출력 결과에는 오류가 빈번히 발생한다. 이러한 오류는 시스템 성능뿐만 아니라 후속 자연어 처리 작업의 품질에도

직접적인 영향을 미친다. 따라서, STT 시스템의 출력 결과 텍스트에 대한 후처리는 전체 음성 기반 처리 파이프라인의 품질을 유지하기 위해 필수적인 요소이다.

STT의 오류는 일반적으로 음소 유사성 기반 단어 대체, 단어의 삽입 및 생략, 문장 구조의 비일관성 등으로 구분된다[12]. 첫째, 음소 유사성 기반 단어 대체는 발음이 유사한 단어 간의 혼동으로 인해 발생하며, 특히 고유명사나 신조어에 대해 자주 나타난다. 둘째, 단어의 삽입 및 생략 오류는 모델의 디코딩 과정에서 문맥 예측 실패 또는 불확실한 확률 분포에 의해 유도된다. 셋째, 문장 구조의 비일관성은 조사, 어미 등 문법 단위의 탈락이나 변형을 포함하며, 특히 한국어와 같이 어순과 문법적 종결이 중요하게 작용하는 언어에서 해석 오류로 이어질 수 있다.

이러한 오류는 주로 음향 모델 또는 디코더가 발화자의 억양, 속도, 방언, 주변 잡음 등 다양한 요인에 대해 충분히 대응하지 못하거나, 모델이 학습하지 못한 어휘 및 표현이 입력될 경우 발생한다. 이러한 문제를 보완하기 위한 후처리 정정 기법은 STT 출력 텍스트에 대해 문법적, 의미적 일관성을 보장하는 것을 주요 목표로 한다. 일반적인 접근 방식으로는 정규 표현식이나 사전 기반 규칙(rule-based), 철자 교정(spelling correction), 통계 기반 언어모델, 사전학습 언어모델을 활용한 문맥 기반 정정이 있다. 특히 사후 정정은 음성신호에 직접 접근하지 않고도 STT 결과의 품질을 개선할 수 있어, 다양한 도메인에 적용 가능한 경량화된 보정 전략으로 주목받고 있다.

최근에는 한국어의 형태적, 음운적 특성을 반영하기 위한 사전학습 언어모델 연구가 활발히 이루어지고 있다. KoBERT 이후에도 다양한 한국어 특화 BERT 계열 모델이 제안되었으며, KRongBERT 기법은 형태소 기반의 임베딩 구조를 통해 한국어 문맥 이해 성능을 개선하였다[16]. 또한, KoGEC 기법은 사전학습 번역모델을 기반으로 한국어 문법 오류를 자동으로 교정하는 프레임워크를 제안하였으며[17], 조사, 어미, 띄어쓰기 등 한국어 고유의 문법적 특성을 반영하였다. 그리고 KcBERT를 이용한 연구에서는 MLM 기반의 교정기능을 활용하여 문맥에 맞는 언어로 교정하여 정확도를 향상시켰다[18]. 이러한 선행연구는 STT 후처리 단계의 교정 정확도 향상에 참고할 수 있는 한국어 기반 오류 정정 접근으로, 본 연구의 KoBERT 기반 문맥 평가 및 교정 과정의 타당성을 뒷받침한다. 그러나 BERT의 교정기법만을 활용하는 것은 문맥에 기반한 교정만을 반영하기 때문에 교정 후 실제 말하지 않은 용어가 등장할 수 있다는 문제가 존재한다.

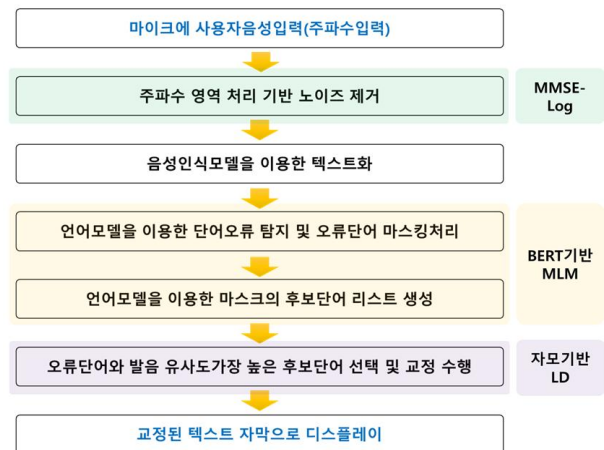


Fig. 1. Process for Improving Speech Recognition

### III. The Proposed Scheme

본 장에서는 음성인식 시스템 개선을 위한 제안구조와 처리 프로세스의 각 단계를 설명한다. Fig. 1은 음성인식 개선을 위한 프로세스를 나타낸 것으로, 각 처리단계는 노이즈 제거, 음성인식 및 단어오류 탐지, 마스크 후보단어 리스트 생성, 발음 유사도 기반 후보단어 선택으로 구성되어 있다. 기존 연구인 BERT를 이용한 오인식 단어 교정 방법에서 자모기반LD 유사도를 이용한 단어교정절차가 추가되어 문맥반영 뿐만 아니라 실제 발화에 근거한 교정이 수행된다는 강점을 가진다[18]. 또한 주파수 필터링이 추가되어 잡음에도 효과가 있다.

#### 1. Frequency-Domain Noise Reduction

사용자 음성을 입력받은 후 이루어지는 첫 번째 작업은 주파수 영역 처리 기반 노이즈 제거 단계이다. 이 단계에서는 음성신호의 품질을 향상시키기 위해 MMSE 기반 로그 스펙트럼 추정법과 고주파 강조 필터(High Pass Filter)를 사용한다. 이 기법을 통해 음성신호는 강화하고 노이즈를 약화하여 신호대잡음비(SNR)를 개선할 수 있다. 이러한 과정으로 노이즈가 많은 환경에서 Whisper 모델이 더욱 정확한 인식 결과를 생성한다. MMSE 기반 로그 스펙트럼 향상 기법은 노이즈에 의해 변형된 음성신호의 스펙트럼을 로그 도메인에서 복원하는 방식으로, 음성신호의 통계적 특성과 노이즈의 분포를 고려하여 최적의 추정값을 계산한다. 이 방식은 배경 노이즈가 지속적이거나, 신호대잡음비가 낮은 입력에서도 과도한 신호 왜곡 없이 음성 성분을 보존할 수 있다는 점이 장점이다. 고주파 강조 필터는 음성신호의 고주파 대역을 선택적으로 증폭하

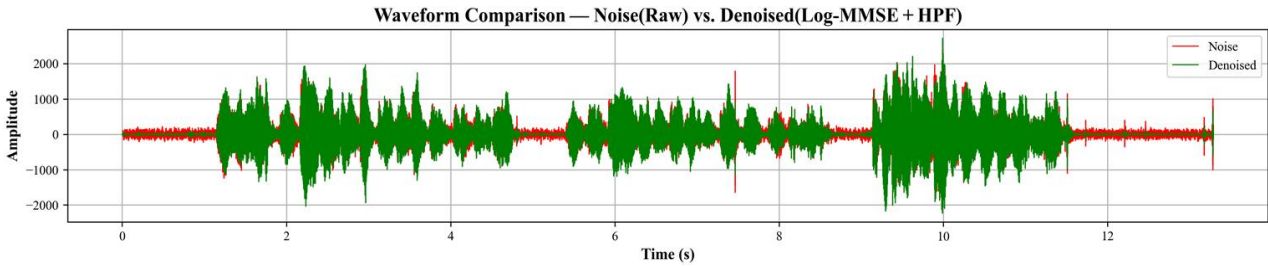


Fig. 2. Comparison of Original Frequency and Frequency After Processing

여 노이즈로 처리가 어려운 쉬운 자음 계열의 특성을 보완하며, 음성-비음성 구간의 분리 정확도를 향상한다. 이를 통해 개선된 음성신호는 Whisper 기반 음성인식 모델이 더욱 안정적으로 음성을 해석할 수 있도록 하며, 특히 노이즈 환경에서의 단어오류를 절감에 실질적인 효과를 보인다. Fig. 2는 MMSE 기반 로그 스펙트럼 추정법과 고주파 강조 필터를 적용했을 때 원본 주파수와 비교하여 나타난 것이다. 노이즈제거 기법을 적용한 결과(초록), 저주파에 해당하는 잡음이 크게 줄어들었고 불필요한 배경 소리도 함께 감소하였다. 동시에 음성에서 중요한 고주파 성분은 원래 신호 수준으로 복원되거나 오히려 더 강조되어, 발화의 특징이 더욱 뚜렷하게 드러났다. 즉, 처리 후 신호는 잡음은 줄고 음성은 더 선명해져 전체적으로 명확도가 높아진 것을 확인할 수 있다.

2. Word Error Detection via LLM

음성인식 결과에는 발음 습관이나 문맥적 혼동으로 인한 단어 오류가 존재할 수 있다. 이러한 오류 교정을 위해 최근에는 LLM을 활용한 STT 출력의 오류를 탐지, 교정,

검증하는 단계적 접근이 제안되었다[14]. 해당 선행 연구에서는 ASR 후처리 단계에 LLM을 적용한 3단계 구조를 사용하며, 본 연구에서는 이러한 절차적 흐름에 착안하여 KoBERT 기반 문맥 평가를 통한 한국어 특유 문장 구성과 형태 정보를 반영한 오류 탐지 과정을 설계하였다.

노이즈가 제거된 음성신호는 음성인식 모델에 입력되어 최종적으로 텍스트 형태로 변환된다. 그러나 음질 개선이 이루어졌다고 하더라도, 여전히 화자의 발음 습관, 속도, 억양 변화, 발화 길이, 연음 및 축약 등 다양한 요인으로 인해 음성인식 결과에는 문맥상 부적절한 단어 혹은 발음 유사 오인식 단어가 포함될 수 있다.

이러한 오인식 오류는 일반적으로 문장의 흐름과 의미를 어긋나게 만들며, 텍스트 교정, 음성 명령 해석,질의 응답 등과 같은 후속 처리 단계에서 예기치 않은 문제를 발생시킬 수 있다. 따라서 변환된 텍스트로부터 오인식되거나 또는 의미적, 문맥적으로 맞지 않는 단어를 자동으로 탐지하는 기법이 요구된다.

본 절에서는 이러한 오류 단어를 탐지하기 위한 마스킹 처리(Masking) 알고리즘에 대해 설명한다. 특히, 문맥 기반 단어 적합성 평가를 위해 BERT 기반 한국어 사전학습 모델인 KoBERT를 활용하며, 이를 통해 한국어 문장의 어순, 조사, 맥락 정보를 반영한 정확한 평가를 가능하게 한다[7-8].

만약, “내일까지 리포트를 제출하세요”라는 문장이 음성인식 시스템에 의해 “내일까지 니포트를 제출하세요”로 잘못 변환되었을 경우, “니포트”는 발음상 “리포트”와 유사하나 문맥적으로는 부적절한 단어이다. 이러한 오인식 단어를 식별하기 위해 KoBERT 기반의 토큰 마스킹(Masked Token Evaluation) 기법을 적용한다.

Fig. 3은 전체 오류 단어 탐지 및 마스킹 처리 과정을 시각적으로 나타낸다. 입력된 문장은 KoBERT의 토큰라이저(Tokenizer)를 이용하여 분리된다. 이후 각 토큰을 하나씩 마스크로 대체한 마스킹 문장들을 생성하며, 이러한 문장들을 KoBERT에 순차적으로 입력하여 마스킹된 위치에



Fig. 3. Misrecognized Word Detection Process

대해 KoBERT가 예측하는 Top-k 후보 토큰 리스트를 확인한다. 각 위치의 원래 토큰이 해당 Top-k 후보에 포함되지 않는 경우, 해당 단어는 문맥상 어색하다고 판단하며, 오인식 단어로 탐지된다. 탐지된 단어는 특수 토큰으로 저장되고, 이 마스킹 정보는 후속 단계인 발음 유사도 기반 후보 생성 및 문맥 기반 재선정 과정에 활용된다. 결과적으로 이 방식은 단순한 사전 기반 오타 탐지와 달리, 전체 문장의 의미적 흐름과 통계적 언어 지식을 종합적으로 고려하기 때문에 오인식과 누락을 줄이고, 더 정밀한 오류 탐지를 수행할 수 있다.

### 3. Similarity-Based Word Selection

입력된 텍스트로부터 오인식된 단어가 탐지되고 오인식 단어에 대한 교정 후보 단어 리스트가 생성되면, 그 다음으로 후보 단어 중 가장 적절한 정답 단어를 선택하는 과정을 거친다. 이 과정은 주로 발음 유사도에 기반하여 수행되며, 오인식의 원인 대부분이 발음상의 유사성에 기인한다는 점에서 매우 중요한 역할을 한다.

오인식된 단어와 각 교정 후보 간의 유사도를 계산하기 위해 LD(Levenshtein Distance) 알고리즘을 활용한다. LD 알고리즘은 두 문자열 간의 삽입, 삭제, 치환 연산 최소 횟수를 기반으로 유사도를 정량화하는 방식으로, 비교적 계산이 간단하고 일반적인 문자열 유사도 측정에서 널리 사용된다. 문자열 유사도를 측정하기 위한 자로윙클러(Jaro-Winkler), G2P(Grapheme to phoneme) 등의 기법들이 존재하지만 한국어의 초/중/종성의 단위에 대한 처리를 반영하기 어렵거나 연산 비용이 크기 때문에 실시간 처리속도에 대응하기 위한 경량화가 어렵다는 문제가 존재한다[19-20]. 따라서 본연구에서는 자모기반 LD를 발음 유사도 검사기법으로 채택한다[21].

기존 LD 알고리즘 역시 문자 단위의 물리적 거리만을 계산하며, 실제 발음상의 유사성이나 언어학적 유사성을 반영하지 못한다는 한계가 존재한다. 예를 들어 “리포트”와 “니포트”는 자소 하나 차이(ㄹ, ㄴ)에 불과하지만, 음성 인식 시스템에서는 자주 혼동되는 대표적인 발음 유사 사례이다. 반면, LD 알고리즘은 이러한 미세한 발음 차이보다는 전반적인 철자 구조의 차이를 기준으로 유사도를 판단하기 때문에, 한국어의 음운론적 특성을 고려한 보정이 어렵다.

		ㄹ	ㄴ	ㄷ	ㄱ	ㅇ	ㅡ	ㅣ	ㅁ
	0	1	2	3	4	5	6	7	8
ㄴ	1	1	2	3	4	5	6	7	8
ㄷ	2	2	1	2	3	4	5	6	7
ㄱ	3	3	2	1	2	3	4	5	6
ㅇ	4	4	3	2	1	2	3	4	5
ㅡ	5	5	4	3	2	1	2	3	4
ㅣ	6	6	5	4	3	2	1	2	3
ㅁ	7	7	6	5	4	3	2	1	2
ㅂ	8	8	7	6	5	4	3	2	1
ㅅ	9	9	8	7	6	5	4	3	2

Fig. 4. Jamo-Based Levenshtein Distance Calculation for “리포트” and “니포트”

따라서 이러한 문제점을 보완하기 위해, 입력 텍스트를 자모(초성, 중성, 종성) 단위로 분리한 후, 각 자모 간의 LD 값을 계산하는 자모기반 편집 거리(Jamo-level LD) 방식을 적용한다. 자모 단위 분리는 한국어 문자의 발음 구조를 더욱 정밀하게 반영할 수 있으며, 자음 교체, 모음 유사성, 종성 누락 현상 등 실제 음성 오인식에서 빈번히 나타나는 오류 패턴을 민감하게 반영할 수 있다. 이를 통해, 자소 분리 시 한국어의 음절 구조인 초성-중성-종성을 고려하여, 철자뿐만 아니라 발음상의 유사성과 구조적 유사성까지 함께 평가할 수 있는 점에서 기존 LD 알고리즘보다 더욱 효과적인 유사도 측정 수단이 된다. Fig. 4와 같이 “니포트”와 “리포트”의 LD를 계산할 경우, 자모기반 LD 값은 편집 거리가 1로 계산되며, 이는 일반적인 문자 LD 값보다 훨씬 더 발음 유사성을 잘 반영한다.

최종적으로, 모든 후보 단어에 대해 자모기반 LD를 계산하고, 편집 거리가 가장 작은 후보를 최종 교정 단어로 선택하게 된다. 최근에는 음성인식 결과의 신뢰도를 활용하여, 오류 가능성이 높은 단어를 선택적으로 교정하는 방법이 제안되었다[15]. 선행 연구에서 제안된 기법은 ASR 신뢰도 정보를 기반으로 LLM을 활용한 교정 과정을 수행함으로써, 불필요한 수정을 최소화하면서 교정 효율이 향상된 것으로 보고되었다. 본 연구의 자모기반 후보 단어 선택 과정 또한 이러한 선택적 교정 개념을 반영하여, 발음 유사도와 문맥 적합도를 동시에 고려함으로써 안정적인 교정 결과를 도출하도록 설계되었다. 최종 교정이 끝난 문장은 번역, 자막, 명령처리 등 다양한 분야에서 응용할 수 있다.

Table 1. Model Definition for WER Comparison

Model	Configuration
Baseline	Whisper(Large-V3)
Model 1[18]	Whisper + BERT
Model 2 (proposed)	Whisper + MMSE + Pre-emphasis + BERT + LD

## IV. Experiments

본 장에서는 제안된 음성인식 후처리 시스템의 성능을 정량적으로 검증하기 위해, 음성인식 결과 텍스트 내 오인식 단어를 정확하게 탐지할 수 있는지, 탐지된 단어에 대해 올바른 후보를 선택하여 교정할 수 있는지, 그리고 전체 문장 단위에서 최종 결과가 기존 STT 결과보다 낮은 WER를 달성할 수 있는지 평가한다. 이를 위해 인식 단어 오류율(Word Error Rate, WER) 오인식 단어 탐지 정확도, 단어 교정 성공률을 각각 측정한다.

### 1. Word Error Rate (WER) Evaluation

제안한 음성인식 후처리 시스템의 실제 성능 향상을 정량 평가하기 위해, 세 가지 모델을 비교하였다. 비교모델은 Whisper (Large-v3) 단독 모델, Whisper와 BERT 기반 교정모델[18] 그리고 제안모델로 구성되어 있으며 표 1에 각 모델에 대한 정의를 요약하였다.

평가는 동일한 테스트 세트에 대해 WER과 처리시간을 표준 지표로 성능을 산출하고, 제안 시스템 적용 전, 후 결과를 비교 분석하였다. 이를 위해 약 7~9개의 어절로 구성된 2~3초 분량의 한국어 문장 1,000건으로 구성된 평가용 대본을 준비하였다. 각 문장은 모두 문법적으로 올바른 문장이며 해당 대본을 음성인식 모델에 음성을 저자들이 직접 음성으로 녹음하여 각 모델의 입력값으로 사용하였다. 여기서 오인식 단어 교정을 위한 MLM(Masked Language Modeling)의 Top-k는 30으로 설정하였다.

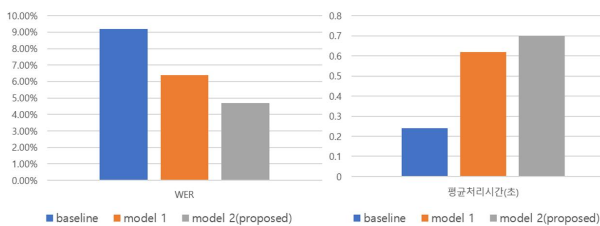


Fig. 5. Comparison of WER and Processing Time by Method

Fig. 5는 baseline 시스템과 제안 기법 간의 WER과 음성 입력부터 디스플레이에 출력되는 문장당 평균 처리시간을 시각적으로 비교한 결과를 나타낸다. 실험 결과, Whisper 단독 사용 시 평균 WER은 9.2%로 측정되었으며, 이는 발음, 억양, 환경 잡음 등 다양한 요인으로 인해 일부 단어가 오인식되었음을 의미한다. 또한 KoBERT의 MLM 교정기능만을 활용한 모델은 6.2%의 WER로 측정되어 MLM 교정이 개선효과가 있음을 보였다. 제안모델의 경우에는 WER이 4.7%로 감소하였으며, 이는 평가 모델 중 가장 우수한 성능이며 전체 오류율의 약 48.9% 감소에 해당하는 성능을 개선하였다.

제안모델이 오인식한 문장 47건 중 32건(68%)은 모음 인식에 의한 오류이며 대부분 “ㅐ”와 “ㅑ”간의 오인식에 집중되어 있다. 자음오인식에 의한 오류 15건(32%)은 “ㄹ”과 “ㄴ”에 집중되어 있는 현상이 발견되었다. 또한 이들 47건의 오인식문장 중 26건(55%)은 오인식단어 탐지에 실패했고 나머지 21건(45%)은 오인식단어를 탐지했으나 교정에 실패한 경우이다.

한편, 처리 속도 측면에서 제안 기법은 문장당 평균 약 0.70초의 처리 시간이 소요되어, 기존 baseline 대비 약 3배 수준(baseline 0.24초)의 연산 지연이 발생하였다. 그러나 해당 처리 속도는 실시간 음성인식 기반 시스템이 요구하는 응답 시간 범위 내에 충분히 수용 가능한 수준으로, 실제 서비스 적용에 있어 실효성을 저해하지 않는 것으로 판단된다.

이러한 결과는 제안된 시스템이 Whisper 모델의 음성인식 결과를 후처리함으로써, 단어 수준의 오류를 효과적으로 탐지 및 교정하고, 최종 문장 수준에서도 인식 품질을 유의미하게 향상할 수 있음을 의미한다. 특히, 단일 단어 오인식이 문장 전체의 WER에 크게 영향을 미치는 짧은 문장 조건에서는 정교한 오류 탐지 및 교정 알고리즘이 WER 개선에 매우 효과적으로 작용함을 확인할 수 있었다.

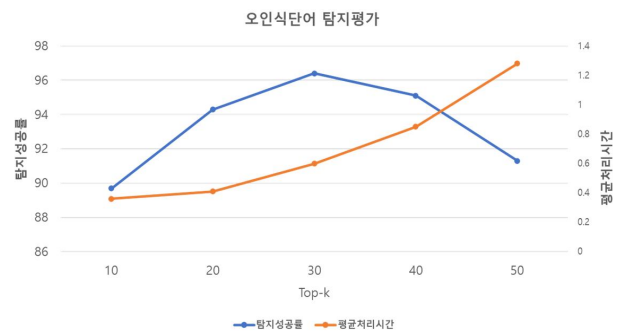


Fig. 6. Performance Evaluation of Misrecognized Word Detection

## 2. Top-k Misrecognition Detection

실험을 위해 1,000개의 테스트 문장을 생성하였으며, 각 문장에는 의도적으로 발음 유사성에 기반한 오인식 단어가 1개씩 포함되어 있다. KoBERT의 MLM에 대한 Top-k의 범위를 10에서 50으로 설정하여 실험을 진행하였다. Fig. 6은 제안된 오인식 단어 탐지 알고리즘의 탐지 성능을 정량적으로 보여준다.

Top-k 값이 증가할수록 KoBERT의 후보군 내 정답 단어 포함 가능성이 높아지며, 오인식 단어 탐지율도 함께 상승한다. 특히, Top-30 설정에서 최고 탐지성공률인 96.4%를 기록하였으며, Top-20, Top-40 역시 95% 내외의 높은 정확도를 유지하였다. 그러나 Top-50에서는 후보군 과잉으로 인해 의미적으로 부적절한 단어가 다수 포함되면서 오교정 비율이 증가하고, 전체 교정 성공률이 하락하는 현상이 발생하였다.

결론적으로 Top-k 증가에 따른 커버리지(coverage) 이득과 노이즈 유입의 트레이드오프(Trade-off)가 k=20에서 k=30 부근에 형성되며, 이는 KoBERT MLM의 확률분포 꼬리부에서 저확률, 저품질 후보가 급격히 많아지는 특성과 OOV(Out of Vocabulary)에서의 후보 미포함 문제가 맞물린 결과로 해석된다.

처리 속도 측면에서는 k의 증가에 따라 후보 탐색 및 유사도 계산 연산량이 선형적으로 증가하였다. 결론적으로 본 실험에서는 탐지성공률(96.4%)와 처리시간(0.51초)의 균형 관점에서 Top-30은 최고의 교정 성능을 제공하며, 처리시간 증가를 수용할 수 있는 고정밀 응용에 적합하다. 반면, Top-20은 탐지성공률(94.3%)과 처리 시간(0.31초) 모두 우수한 수준을 유지하여 실시간 교정 시스템에서 가장 합리적인 설정값인 것으로 판단된다.

오인식 단어를 정성적으로 분석한 결과, 세 가지 주요 원인으로 분류할 수 있었다. 첫 번째는 오인식 단어가 문맥상 자연스럽게 수용되는 경우이다. 이는 “고양에 돌아왔습니다”가 “고양에 돌아왔습니다”로 인식된 경우와 같이 오인식된 단어가 실제 단어이며 문맥상 이상이 없다고 판단되는 사례이다. 이 경우 KoBERT는 통계적으로 가능한 문맥으로 인식하여 정상적으로 탐지가 되지 않은 것으로 판단된다.

두 번째는 정상적인 단어임에도 불구하고 KoBERT의 Top-k 후보 단어 리스트에 정답 단어가 포함되지 않아 탐지를 실패한 경우이다. 이는 대부분 외래어이거나 특수 도메인 어휘가 포함된 경우였다. KoBERT는 뉴스 및 위키 기반 말뭉치로 학습되어 있어 도메인 특화 단어에 대해 상대적으로 낮은 예측 확률을 보이는 한계가 있으며, 이로

인해 해당 단어가 Top-k 리스트에 포함되지 않고 탐지에서 제외되었다.

세 번째는 신조어 및 모델 사전학습 범위를 초과하는 단어가 포함된 경우이다. 예를 들어 “온보당”, “챗지피티”, “멀티모달” 등 최근 사용 빈도가 급증한 단어들이 포함되었을 때 KoBERT가 이를 제대로 처리하지 못하는 경우이다. 이러한 단어들은 모델의 사전학습 범위 밖에 있어 의미 파악이나 후보 단어 예측이 제대로 이루어지지 않아 정상적인 단어로 인식되지 않는다.

## 3. Word Correction Success Rate

앞선 실험을 통해 오인식 단어가 성공적으로 탐지된 경우를 대상으로 Top-k가 30일 때, 해당 단어에 대한 자동 교정이 실제로 올바르게 수행되었는지를 평가하였다. 평가 절차는 다음과 같다. 먼저 오인식 단어로 판단된 영역에 대해 시스템이 제안한 교정 단어로 교체하고, 이를 해당 문장의 정답 문장과 비교하였다. 비교 결과가 일치하는 경우를 교정 성공으로 간주하고, 그렇지 않은 경우는 교정 실패로 분류하였다.

실험 결과, 전체 964건의 오인식 단어가 포함된 문장 중 총 932건(약 96.7%)의 성공적인 교정이 수행되었음을 보여준다. 이는 오인식 탐지 후 자모기반의 LD 알고리즘이 높은 정확도로 적절한 단어를 추천하고 반영했음을 의미하며, 시스템이 실제 서비스 환경에서 오인식 개선을 위한 후처리 단계로써 효과적으로 기능할 수 있음을 입증하는 결과이다.

반면, 32건(약 3.3%)의 경우에는 교정 단어가 정답과 일치하지 않아 실패로 기록되었다. 이들 대부분은 정답 단어 자체가 후보 리스트에 포함되지 않은 경우이다. 이는 음성 발화가 특정 고유명사, 신조어, 또는 도메인 특화 용어일 때 주로 발생하였으며, 발음 기반 후보 생성기에서 수용 가능한 범위를 확장하거나 사용자 사전을 도입함으로써 개선이 가능하다.

## V. Conclusion

본 연구에서는 음성인식 모델인 Whisper와 같은 고성능 STT 시스템의 한국어 음성인식 품질을 향상시키기 위해, MMSE 기반 로그 스펙트럼 전처리와 KoBERT 기반 문맥 분석과 자모 단위 편집 거리 후처리를 결합한 통합 개선 기법을 제안하였다. 제안 기법은 입력 음성신호의 주파수 영역 노이즈 제거를 통해 SNR을 개선하고, 음성인식

결과 텍스트에서 문맥상 부적절하거나 발음 유사성에 기인한 오인식 단어를 탐지·교정함으로써 최종 출력 텍스트의 정확도를 높였다.

실험 결과, Whisper 모델의 단독 사용 대비 제안 기법은 WER을 9.2%에서 4.7%로 약 48.9% 감소시켰으며, Top-30 후보 기반 설정에서 96% 이상의 오인식 탐지, 성공률을 달성하였다. 이는 제안 기법이 단어 수준의 정밀한 오류 음운 유사도를 동시에 고려한 품질 개선을 보였다.

그럼에도, 본 연구는 고유명사·신조어·도메인 특화의 어휘에 대한 한계를 여전히 지니고 있으며, 학습 데이터에 포함되지 않은 발화에 대한 탐지 및 교정 성능이 저하될 수 있다. 향후 연구에서는 사용자 사전 기반의 단어 보정, 도메인 적응형 언어모델, 발음 사전 확장을 도입하여 이러한 한계를 극복할 수 있다. 또한, 다국어 환경으로의 확장, 기계 번역, 자막 생성 등 다양한 NLP 파이프라인과 연계로 통해 제안 기법의 범용성을 검증할 수 있다.

결론적으로, 본 연구는 한국어 기반의 STT 환경에서 전처리와 후처리의 결합을 통한 실질적 인식 품질 개선 가능성을 입증하였고, 이를 통해 음성 기반 번역·질의응답 등 고품질 언어 입력을 요구하는 응용 시스템의 성능 향상에 기여할 수 있음을 확인하였다.

## ACKNOWLEDGEMENT

This work was supported by the Digital Innovation Hub project supervised by the Daegu Digital Innovation Promotion Agency(DIP) grant funded by the Korea government(MSIT and Daegu Metropolitan City) in 2025(No.225C000366, 2025년 R&BD협업 프로젝트(상용화 지원)).

## REFERENCES

- [1] X. L. Dong, S. Moon, Y. E. Xu, K. Malik, and Z. Yu, "Towards next-generation intelligent assistants leveraging llm techniques," in Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 5792-5793, 2023. DOI: 10.1145/3580305.3599572
- [2] Y. Guan, D. Wang, Z. Chu, S. Wang, F. Ni, R. Song, L. Li, J. Gu, and C. Zhuang, "Intelligent virtual assistants with llm-based process automation," arXiv preprint arXiv:2312.06677, 2023. DOI: 10.48550/arXiv.2312.06677
- [3] R. Sarikaya, "The technology behind personal digital assistants: An overview of the system architecture and key components," IEEE Signal Processing Magazine, vol. 34, no. 1, pp. 67-81, 2017. DOI: 10.1109/MSP.2016.2617341
- [4] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp. 4580-4584, 2015. DOI: 10.1109/ICASSP.2015.7178838
- [5] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in 2013 IEEE international conference on acoustics, speech and signal processing. IEEE, pp. 6645-6649, 2013. DOI: 10.1109/ICASSP.2013.6638947
- [6] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in International conference on machine learning. PMLR, pp. 28492-28518, 2023.
- [7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, vol. 1, pp. 4171-4186, 2019. DOI: 10.18653/v1/N19-1423
- [8] X. Chen, K. Hui, B. He, X. Han, L. Sun, and Z. Ye, "Incorporating ranking context for end-to-end bert re-ranking," in European Conference on Information Retrieval. Springer, pp. 111-127, 2022.
- [9] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal processing magazine, vol. 29, no. 6, pp. 82-97, 2012. DOI: 10.1109/MSP.2012.2205597
- [10] A. Das, J. Li, R. Zhao, and Y. Gong, "Advancing connectionist temporal classification with attention modeling," in 2018 IEEE International conference on acoustics, speech and signal processing (ICASSP). IEEE, pp. 4769-4773, 2018. DOI: 10.1109/ICASSP.2018.8461558
- [11] J. Li, R. Zhao, H. Hu, and Y. Gong, "Improving rnn transducer modeling for end-to-end speech recognition," in 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE, pp. 114-121, 2019. DOI: 10.1109/ASRU46091.2019.9003906
- [12] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach, "An overview of noise robust automatic speech recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 4, pp. 745-777, 2014. DOI: 10.1109/TASLP.2014.2304637
- [13] H. Seong, N.-J. Kim, H. G. Ryu, and H.-J. Lee, "Extending whisper for korean-english code-switching speech recognition," in 2025 IEEE International Conference on Consumer Electronics

- (ICCE). IEEE, pp. 1-4, 2025. DOI: 10.1109/ICCE63647.2025.10929894
- [14] Y. Fang, B. Cheng, J. Peng, X. Li, Y. Xi, C. Zhang, and G. Zhong, "Fewer hallucinations, more verification: A three-stage llm-based framework for asr error correction," arXiv preprint arXiv:2505.24347, 2025. DOI: 10.48550/arXiv.2505.24347
- [15] A. Hernandez, T. A. Vergara, A. Maier, and P. A. P'erez-Toro, "Confidence-guided error correction for disordered speech recognition," arXiv preprint arXiv:2509.25048, 2025. DOI: 10.48550/arXiv.2509.25048
- [16] H. Yu, Y. Cho, G. Park, and M. Kim, "Krongbert: Enhanced factorization based morphological approach for the korean pretrained language model," *Information Processing & Management*, vol. 62, no. 3, p. 104072, 2025. DOI: 10.1016/j.ipm.2025.104072
- [17] T. Kim, S. Jeong, and Y. Song, "Kogec: Korean grammatical error correction with pre-trained translation models," arXiv preprint arXiv:2506.11432, 2025. DOI: 10.48550/arXiv.2506.11432
- [18] D. Min, S. Nam, and D. Choi, "A Study on Improving the Accuracy of Korean Speech Recognition Texts Using KcBERT," *Journal of KIISE*, vol. 51, no. 12, pp. 1115-1124, 2024. DOI: 10.5626/JOK.2024.51.12.1115
- [19] W. E. Winkler, "Overview of record linkage and current research directions," *Research in Official Statistics*, vol. 1, no. 1, pp. 57-64, 1997.
- [20] H. Y. Kim, J. H. Kim, and J. M. Kim, "Fast Bilingual Grapheme-To-Phoneme Conversion," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Industry Track)*, pp. 289-296, 2022. DOI: 10.18653/v1/2022.naacl-industry.32
- [21] W. I. Cho, S. M. Kim, and N. S. Kim, "Investigating an effective character-level embedding in Korean sentence classification," in *Proceedings of the 33rd Pacific Asia Conference on Language, Information and Computation*, pp. 10-18, 2019. DOI: 10.48550/arXiv.1905.13656
- [22] B. Kim and Y. Kwon, "Searching for effective preprocessing method and cnn based architecture with efficient channel attention on speech emotion recognition," *Scientific Reports*, vol. 15, no. 1, p. 32689, 2025.
- [23] P. Cherukuru and M. B. Mustafa, "Cnn-based noise reduction for multi-channel speech enhancement system with discrete wavelet transform (dwt) preprocessing," *PeerJ Computer Science*, vol. 10, p. e1901, 2024. DOI: 10.7717/peerj-cs.1901
- [24] Y. Iqbal, T. Zhang, T. S. Gunawan, A. Pratondo, X. Zhao, Y. Geng, M. Kartiwi, N. Saleem, and S. Bourious, "A hybrid speech enhancement technique based on discrete wavelet transform and spectral subtraction," *IEEE Access*, 2025. DOI: 10.1109/ACCESS.2025.3546434

## Authors



Yonghun Jang received the B.S. degree in Computer Engineering from Yeungnam University College, Daegu, Republic of Korea, in 2012, and the Ph.D. degree in Computer Engineering from Yeungnam

University, Gyeongsan, Korea, in 2020. He is currently the Director of the Research Institute (Head of the R&D Center) at Nearnetworks, Co., Ltd. His research interests include data analysis, deep learning, and large language models.



Jung Min Lim received B.S. and M.S. degrees in the Department of Computer Engineering from Yeungnam University, Korea, in 2023 and 2025, respectively. He is currently a Ph.D. student in the Department

of Computer Engineering at Yeungnam University. His current research interests include advanced processor architecture, operating systems, and storage systems.



Seong-Guk Nam received the B.S. and M.S. degrees in Computer Engineering from Yeungnam University, Gyeongsan, Korea, in 2021 and 2023, respectively. Since 2023, he has been with the R&D Center at

Nearnetworks, where he is currently a research engineer. He is interested in deep learning and Large Language Model.



Minhyung Ryu received a B.S. degree in the Department of Statistics from Yeungnam University, Gyeongsan, Korea, in 2021 and an M.S. degree in computer engineering from Yeungnam University, in 2023.

She is currently a Researcher in the R&D Center at Nearnetworks, She is interested in healthcare, vision systems, artificial intelligence, and data analysis.



Eunjin Yoo received a B.S. degree in the Department of Statistics from Yeungnam University, Gyeongsan, Korea, in 2021 and an M.S. degree in computer engineering from Yeungnam University, in 2023.

She is currently a Researcher in the R&D Center at Nearnetworks, She is interested in blockchain, recommender system and data analysis.



Myung-Sub Lee received the M.S and Ph.D. degrees in Computer Science and Engineering from Yeungnam University, Gyeongsan, Korea, in 2000 and 2003, respectively. He worked as a professor from 2002 to 2003 at

Hosan University, Gyeongsan. He also worked as a Research Professor at Yeungnam University from 2004 to 2008. He is currently a professor in the Division of Software&Contents at Yeungnam University College. His research interests include machine learning and wireless and sensor networks,.



Jong Wook Kwak received a B.S. degree in Computer Engineering from Kyungpook National University, Daegu, Korea in 1998, a M.S. degree in Computer Engineering from Seoul National University, Seoul, Korea in

2001, and a Ph.D. degree in Electrical Engineering and Computer Science from Seoul National University, Seoul, Korea in 2006. From 2006 to 2007, he worked as a Senior Engineer in the SoC R&D Center, at Samsung Electronics Co., Ltd. During 2011~2012, he was a Guest Researcher at the Research Institute of Advanced Computer Technology, Seoul National University. During 2012~2013, he was a Visiting Scholar at the Georgia Institute of Technology, Atlanta, GA, USA. As a Head Director, he led DREAM Software Human Resource Training Center from 2014~2015. During 2018~2019, he was a Visiting Scholar at Arizona State University, Tempe, AZ, USA. He is currently a professor in the Department of Computer Engineering, Yeungnam University. His research interests include advanced processor architecture, low-power mobile embedded systems, and high performance parallel and distributed computing.