

선박무선통신 음성인식 서비스 개발 및 활용

† 김광일 · 유상록*

† 제주대학교 해양과학대학, *㈜미래해양정보기술 기업부설연구소 연구소장

Development and Utilization of Speech Recognition Service for Ship Radio Communication

† Kwang-Il Kim · Sang-Lok Yoo*

† Professor, College of Ocean Science, Jeju National University, Jeju 63243, Korea
*Director of Research Institute, Future Ocean Information Technology, Jeju 63243, Korea

요 약 : 선박무선통신장비는 선박이 항해하는데 필요한 안전정보, 선박교통 모니터링 및 관제, 입·출항 정보를 교환하기 위한 필수 장비이므로 선박항해사는 무선통신 내용을 항상 주의 깊게 청취해야 함. 본 연구에서는 선박의 실제 음성 교신데이터 500시간 데이터를 수집 및 학습하고, Wav2Vec 및 Whisper 모델을 활용하여 한글 및 영어(해사영어) 음성인식 모델을 개발하고 실용화를 수행하였다. 음성인식 모델의 성능은 CER(Character Error Rate) 기준 94.5%로 향후 선박 운항 관련 다양한 분야에 적용이 가능할 것으로 사료된다.

핵심용어 : 선박무선통신, 음성인식, Wav2Vec, Whisper, 해사영어

1. 서 론

선박무선통신장비는 선박이 항해하는데 필요한 안전정보, 선박교통 모니터링 및 관제, 입·출항 정보를 교환하기 위한 필수 장비이므로 선박항해사는 무선통신 내용을 항상 주의 깊게 청취해야 한다[1]. 최근 빅데이터, 인공지능 기술 발달로 해양분야에서도 e-Navigation, 자율운항선 등 신기술 도입을 적극 추진중임. 해상통신데이터는 선박 현재상태를 파악할 수 있는 객관적인 데이터로서 이를 문자화(전산화)하여 차세대 시스템과 연계 필요한 실정이다. 또한 향후 자율운항선박 기술이 실용화되면, 자율운항선과 주변선박간 항해의도 파악을 위해 선박무선통신 음성인식 기술이 필수적으로 요구되고 있다.

선박무선통신 교신 내용은 전문분야 대화로서 특수용어, 해사영어(Maritime English), 고유명사(선명 및 지명) 등이 포함되어 있고, 교신 규칙도 포함되어 있어 기존 음성인식 모델로 인식이 어렵다.

이에 본 연구에서는 선박 및 육상 해안무선국에서 송수신되는 선박무선통신 교신데이터를 수집, 가공 및 전처리하여 end-to-end 음성인식 딥러닝 모델을 활용하여 선박무선통신 전용 음성인식 모델을 개발하고자 한다.

2. 선박무선통신 음성인식 모델 개발

본 연구에서 음성인식 모델 개발은 ① 데이터 수집 및 전처리 ② 음향모델(Acoustic Model) ③ 언어모델(Language Model) ④ 후처리 변환 과정으로 수행한다.



그림 1. 선박무선통신 음성인식 모델 개발 과정
Fig. 1. Process of Speech Recognition Model for Ship Radio Communication

2.1 데이터 수집 및 전처리

학습에 사용된 선박 무선통신 음성데이터는 제주대학교 및 실습선에서 약 3년치 데이터를 확보하여, 전문경력 라벨러를 통해 약 200시간 학습을 수행하였다. 또한 음성이 장비마다 8

† 교신저자 : 종신회원, kki@jejunu.ac.kr
* 정회원, sanglokyoo@gmail.com

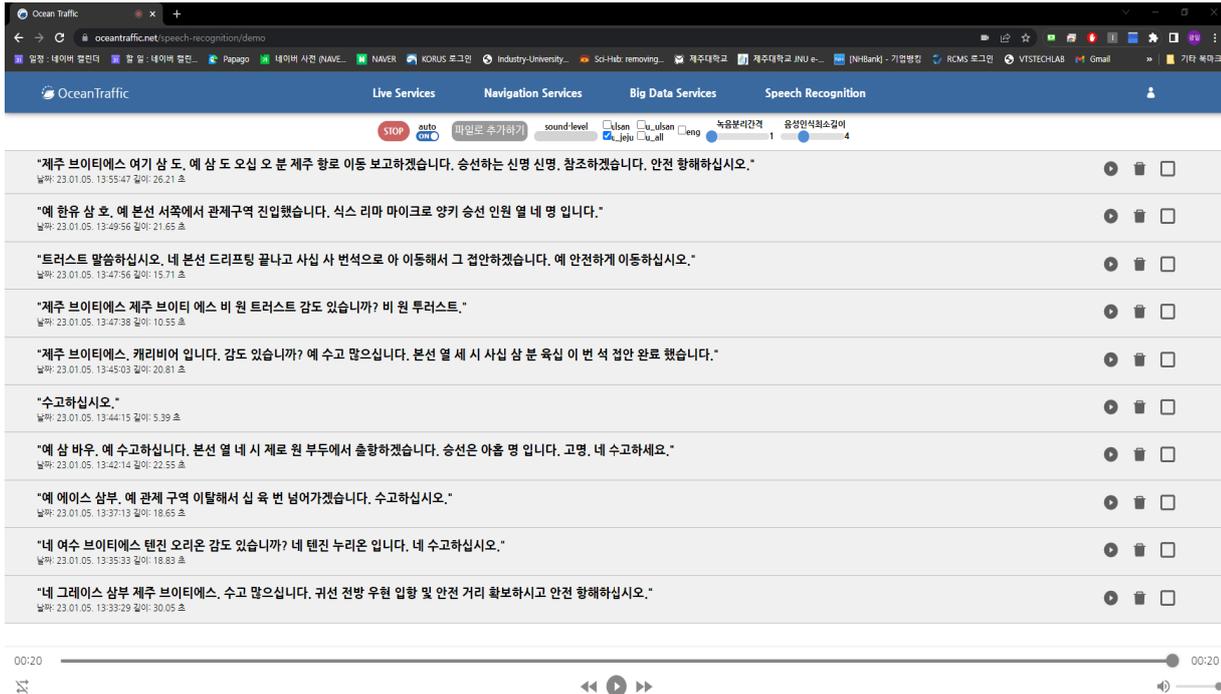


그림 2. 웹 기반 선박무선통신 음성인식 서비스(<https://oceanttraffic.net>)
 Fig. 2. Web based Speech Recognition Service for Ship Radio Communication

khz 또는 16 khz 로 샘플레이트가 차이가 있어서 8 khz 대역으로 통일하였다.

2.2 음향모델 및 언어모델

음향모델은 음성을 만들어내는 음소(phone)와 해당 발음 사이의 관계를 나타내며, 전처리되어 사용되는 데이터와 해당 음성의 대응을 통계적인 정보로 나타낸다. 본 연구에서는 페이스북에서 개발하여 사전 학습된 Wav2Vec 2.0 모델[2]을 사용한다. 이 모델은 적은양의 학습 데이터로도 음성인식모델 개발 장점이 있다.

언어모델은 음성인식에서 발화자의 음성에서 특징 추출하여 음절을 인식하고 글자, 단어 간 확률에 따른 텍스트를 생성을 한다. 언어 모델은 한국어 BERT(Bidirectional Encoder Representations from Transformers) 모델을 적용하여 선박 무선통신 용어에 적합한 텍스트를 생성한다.

2.3 후처리 변환

언어모델의 결과 중 전문용어는 인식 과정에서 오류를 발생하기 쉬우므로 성능 고도화를 위해 선박 분야에서 사용되는 전문용어(해사영어, 지명, 선박이름, 부두번호 등) 표현의 특성을 반영한 N-gram 기반 후처리 모델을 개발하여 특정 단어 추론 향상하였다.

3. 모델 성능 및 활용

개발한 모델에 대한 성능을 검증하기 위해 추가로 수집한 30시간의 음성데이터를 라벨링하여 CER(Character Error Rate) 기준으로 모델의 성능을 평가하였다. 모델 사용 결과 한국어 교신 내용에 대한 인식율은 94.5%로 실무에서 활용하기 충분한 성능을 얻었다. 하지만 해사영어 교신에 대한 음성인식은 정확도가 낮아 향후 영어교신에 대한 추가적인 학습데이터를 수집하여 라벨링을 수행하겠다.

본 연구에서 개발한 한국어 음성인식 기술은 웹(oceanttraffic.net) 기반으로 선박 관련 종사자들에게 서비스 활용할 계획이다.

참 고 문 헌

- [1] 김진태, & 정훈. (2017). 음성인식 기술의 동향과 해군 정보통신분야 적용방안: 음성-문자 변환 (Speech To Text) 중심으로. 국방과 기술, (456), 120-127.
- [2] Baevski, Alexei, et al. "wav2vec 2.0: A framework for self-supervised learning of speech representations." Advances in neural information processing systems 33 (2020): 12449-12460.
- [3] Li, F., Jin, Y., Liu, W., Rawat, B. P. S., Cai, P., & Yu, H. (2019). Fine-tuning bidirectional encoder representations from transformers (BERT) - based models on large-scale electronic health record notes: an empirical study. JMIR medical informatics, 7(3), e14830.