# 청각 장애인을 위한 수어 영상-자연어 번역 서비스 및 모바일 어플리케이션 구현

조수민*, 조성연**, 신소연***, 이지항*,+
*상명대학교 지능데이터융합학부 휴먼지능정보공학전공
**서울여자대학교 정보보호학과
***서울여자대학교 디지털미디어학과
jsm3626@naver.com , thdus060978@naver.com, tjdyyysy@gmail.com, jeehang@smu.ac.kr

# Developing a mobile application serving sign-language to text translation for the deaf

Su-Min Cho *, Seong-Yeon Cho**, So-Yeon Shin ***, Jee Hang Lee*,+
*Dept. of Human-Centered Artificial Intelligence, Sangmyung University, Seoul, South Korea
**Dept. of Information Security, Seoul Women's University, Seoul, South Korea
***Dept. of Digital Media, Seoul Women's University, Seoul, South Korea
+Corresponding author: Jee Hang Lee

## 요 약

Covid-19 로 인한 마스크 착용이 청각장애인들의 소통을 더 어렵게 하는 바, 제 3 자의 도움 없이 쌍방향 소통을 가능하게 하는 서비스의 필요성이 커지고 있다. 이에 본 논문은 소통의 어려움을 겪는 청각장애인과 비청각장애인을 위한 쌍방향 소통 서비스에 대한 연구와 개발 과정, 기대 효과를 담는다. 서비스는 GRU-CNN 하이브리드 아키텍처를 사용하여 데이터셋을 영상 공간 정보와 시간 정보를 포함한 프레임으로 분할하는 영상 분류 기법과 같은 딥 러닝 알고리즘을 통해 수어 영상을 분류한다. 해당 연구는 "눈속말" 모바일 어플리케이션으로 제작 중이며 음성을 인식하여 수어영상과 텍스트로 번역결과를 제공하는 청각장애인 버전과 카메라를 통해 들어온 수어 영상을 텍스트로 변환하여 음성과 함께 제공하는 비청각장애인 버전 두 가지로 나누어 구현한다. 청각장애인과 비장애인의 쌍방향 소통을 위한 서비스는 청각장애인이 사회로 나아가기 위한 가장 기본적인 관문으로서의 역할을 할 것이며 사회 참여를 돕고 소통이라는 장벽을 넘어서는 발돋움이 될 것이라 예측된다.

## 1. 서론

Covid-19 로 인해 마스크 착용이 필수화 되면서, 표정과 입 모양이 의사 소통에 많은 부분을 차지하는 청각장애인들에게 불편함이 가중되고 있다[1]. 이에 따라, 청각 장애인이 직접적인 수어와 소리를 통해 눈으로 말을 볼 수 있도록 표현해 주는 서비스에 대한 요구가 증가하는 상황이다.

본 수어-자연어 번역 서비스는 접근성이 높은 모바일 애플리케이션으로 제작하여, 청각장애인과 비장애인이 타인의 도움 없이 소통 가능한 서비스 제공을 목표로 한다. 이를 수행하기 위해 딥러닝 알고리즘과 정보처리 기술을 접목해 시스템을 구현하고 애플리케이션을 통해 서비스를 제공하고자 한다.

## 2. 사용 기술

### 2.1 형태소 분석

자연어를 수어로 번역하기 위해 텍스트의 성분을 분석하여 수어 번역에 필요한 형태소만 남기는 과정이 필요하다. 이때 번역되는 수어는 자연 수어로 한다[2]. 자연 수어는 청각장애인 사이에서 자연 발생한 것으로, 시제와 조사, 문장 종결 법이 없어 국어 문법과 일치하지는 않는 독특한 특성을 지닌다. 하지만 한 문장에 사용되는 수어 단어 수가 적어 직관적이고 시각적으로 이해하기 쉽다.

이러한 요구사항을 충족하기 위해, 분석된 형태소에서 체언(N**), 용언 중 동사(VV), 형용사(VA), 긍정/부정 지정사(VC*), 관형사(MD*), 부사(MA*)를 제외한 나머지 품사를 제거한다. 반환된 용언(V**)의 형태소

에는 ‘다’를 붙여 용언의 기본형으로 가공한다.

이 연구에서 형태소는 서울대학교 IDS (Intelligent Data Systems) 연구실에서 진행한 꼬꼬마(kkma) 프로젝트의 형태소 분석기 라이브러리를 사용하여 분석한다[3]. ‘꼬꼬마 한글 형태소 분석기’는 자바 라이브러리로 배포되어 자바 언어로 작성된 안드로이드 프로젝트에 추가하여 사용하기에 용이하다.

<표 1> kkma 한글 형태소 품사 태그 표

| 대분류 | 세종 품사 태그 | | 심광섭 품사 태그 | | KKMA 단일 태그 V 1.0 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 태그 | 설명 | Class | 설명 | 묶음 1 | 묶음 2 | 태그 | 설명 | 확률태그 | 저장사전 |
| 체언 | NNG | 일반 명사 | NN | 명사 | N | NN | NNG | 보통 명사 | NNA | noun.dic |
| | NNP | 고유 명사 | | | | | NNP | 고유 명사 | | |
| | NNB | 의존 명사 | NX | 의존 명사 | | | NNB | 일반 의존 명사 | NNB | simple.dic |
| | | | UM | 단위 명사 | | | NNM | 단위 의존 명사 | | |
| | NR | 수사 | NU | 수사 | | | NR | 수사 | NR | |
| | NP | 대명사 | NP | 대명사 | | | NP | 대명사 | NP | |
| 용언 | VV | 동사 | VV | 동사 | V | VV | VV | 동사 | VV | verb.dic |
| | VA | 형용사 | AJ | 형용사 | | VA | VA | 형용사 | VA | |
| | VX | 보조 용언 | VX | 보조 동사 | | VX | VXV | 보조 동사 | VX | |
| | | | AX | 보조 형용사 | | | VXA | 보조 형용사 | | |
| | VCP | 긍정 지정사 | CP | 서술격 조사 '이다' | | VC | VCP | 긍정 지정사, 서술격 조사 '이다' | VCP | raw.dic |
| | VCN | 부정 지정사 | | | | | VCN | 부정 지정사, 형용사 '아니다' | VCN | |
| 관형사 | MM | 관형사 | DT | 일반 관형사 | M | MD | MDT | 일반 관형사 | MD | simple.dic |
| | | | DN | 수 관형사 | | | MDN | 수 관형사 | | |
| 부사 | MAG | 일반 부사 | AD | 부사 | | MA | MAG | 일반 부사 | MAG | |
| | MAJ | 접속 부사 | | | | | MAC | 접속 부사 | MAC | |
| 감탄사 | IC | 감탄사 | EX | 감탄사 | I | IC | IC | 감탄사 | IC | |
| 조사 | JKS | 주격 조사 | JO | 조사 | J | JK | JKS | 주격 조사 | JKS | |
| | JKC | 보격 조사 | | | | | JKC | 보격 조사 | JKC | |
| | JKG | 관형격 조사 | | | | | JKG | 관형격 조사 | JKG | |
| | JKO | 목적격 조사 | | | | | JKO | 목적격 조사 | JKO | |
| | JKB | 부사격 조사 | | | | | JKM | 부사격 조사 | JKM | |
| | JKV | 호격 조사 | | | | | JKI | 호격 조사 | JKI | |

```
I/zygote: Do full code cache collection, code=245KB, data=151KB
I/zygote: After code cache collection, code=219KB, data=137KB
I/System.out: ================================================  왼쪽은 위험해요
I/System.out: 왼쪽은              => [0/왼쪽/NNG+2/은/JX]
I/System.out: 위험해요        => [4/위험/NNG+6/하/XSV+7/어요/EFN]
I/System.out: ================================================  오른쪽으로 운전 천천히 하세요
I/System.out: 오른쪽으로      => [9/오른쪽/NNG+12/으로/JKM]
I/System.out: 운전            => [15/운전/NNG]
I/System.out: 천천히              => [18/천천히/MAG]
I/System.out: 하세요              => [22/하/XSV+23/세요/EFN]
I/System.out: ======================== resultArr
I/System.out: 0왼쪽 1위험 2오른쪽 3운전 4천천히
I/System.out: resultArray size: 5
D/OpenGLRenderer: HWUI GL Pipeline
```

(그림 1) kkma 를 이용한 형태소 분석 결과

### 2.2 Convolutional Neural Network

CNN(Convolutional Neural Networks)은 대표적 딥러닝 구조로, 특히 영상 및 고차원 데이터 처리에 장점을 보인다. 특징을 추출하는 Convolution layer 와 추출된 특징을 샘플링하는 Pooling layer 가 필요에 따라 fully connected layer 와 연결되어 강력하고 효율적인 분류 성능을 보인다. Convolution layer 에서는 이미지에 필터링 기법을 적용하고 Pooling layer 에서는 이미지를 하나의 대표적인 스칼라값으로 변환하여 이미지의 크기를 줄이는 역할을 수행한다. 이는 시각 데이터를 인식하는데 가장 널리 사용되는 모델이다.

### 2.3 Gated Recurrent Unit

GRU 는 RNN(Recurrent Neural Network)의 프레임워크의 일종으로 장기 의존성의 문제를 해소하고 LSTM 보다 간단한 구조로 이루어져 있어 학습할 가중치가 적어 학습속도가 빠르다는 이점이 있다.

## 3. 제안 시스템 구현

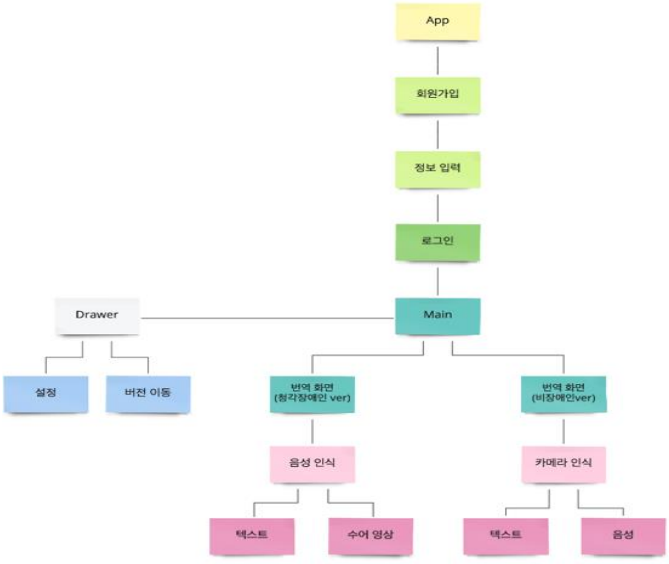
### 3.1 어플리케이션 설계

이 서비스는 하나의 애플리케이션을 청각장애인과 비장애인이 동시에 사용할 수 있다. 모드에 따라 청각장애인용 및 비장애인용 페이지를 다르게 제공한다.

청각 장애인에게 기본으로 설정되는 페이지를 ‘청각장애인 버전’, 비장애인에게 기본 설정되는 페이지를 ‘비장애인 버전’이라고 칭하겠다. 청각장애인 버전은 음성을 수어로 변환하고, 비장애인 버전은 수어를 음성으로 변환하는 기능을 포함한다. 애플리케이션을 처음 실행하고 회원가입을 할 때 장애 여부를 한 번 입력하면 다음 실행할 때도 자동으로 해당 버전으로 이동하도록 하였다. 만약 다른 버전을 사용하기 원하는 경우, 메뉴를 통해 접근이 가능하다. 어플리케이션 ‘눈속말’의 전체 구조는 그림 2 에서 확인 가능하다.

### 3.2 데이터셋 수집

수어 분류 모델을 학습시키기 위해 AiHub 에서 제공하는 수어 영상 데이터를 활용한다. 총 5,000 개의 단어, 3,000 개의 문장, 800 개의 지문자, 200 개의 지숫자 데이터가 있다. 데이터는 MP4 파일의 수어 동영상과 영상의 30fps 분할 이미지에 대한 키포인트를 가공한 JSON 파일, 형태소/비수지 요소를 가공한 JSON 파일로 구성된다. 자세, 왼쪽 손, 오른쪽 손, 표정까지 총 135 개의 좌표를 포즈 인식에 이용할 수 있다. 해당 형태소 구간의 시작 시각과 종료 시각을 나타내는 값을 기반으로 영상을 가공하면 더 매끄럽게 영상을 재생할 수 있다.




(그림 2) 애플리케이션 “눈속말” 구조도

### 3.3 청각장애인 버전 설계

청각장애인 버전은 인식한 음성을 수어 영상으로 변환하여 텍스트와 함께 제공하는 버전이다. 먼저

Google 에서 제공하는 Speech-to-text (STT) API 를 사용하여 음성을 인식하여 텍스트로 변환한다. 변환된 텍스트를 형태소 분석하여 단어로 추출한다. 각 단어에 해당하는 수어 영상을 데이터 셋에서 찾아 순서대로 재생한다. 이때 데이터 셋에 존재하지 않는 고유 명사는 지문자를 이용하여 나타낸다. 정확한 의사 전달을 위해 텍스트도 함께 제공한다.

### 3.4 비장애인 버전 설계

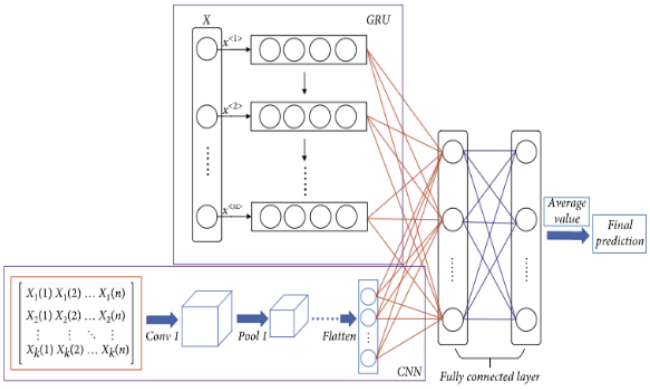
비장애인 버전은 카메라를 통해 들어온 수어 영상을 텍스트로 변환하여 음성과 함께 제공하는 버전이다. 먼저 수어 데이터 셋을 이용해 수어를 분류해 주는 CRNN(CNN-GRU) 딥러닝 모델을 학습시킨다.

그 후 Android Studio 에서 제공하는 카메라 기능을 사용하여 수어 장면을 취득하여 수어 분류 학습이 완료된 CRNN 모델에 넣어 실시간으로 수어를 분류하는 작업을 진행한다.

분류된 단어는 조사와 접사 등이 생략된 원형의 상태이다. 단어의 순서와 의미를 파악해 품사를 파악하고, 해당 품사에 맞는 조사는 SKT Brain 에서 제공하는 KoBERT 프로그램을 이용하여 튜닝한 후, 앱 화면에 출력한다. Google 에서 제공하는 Text-to-Speech(TTS) API 를 이용하여 음성 서비스도 함께 제공한다.

### 3.4 CRNN 모델

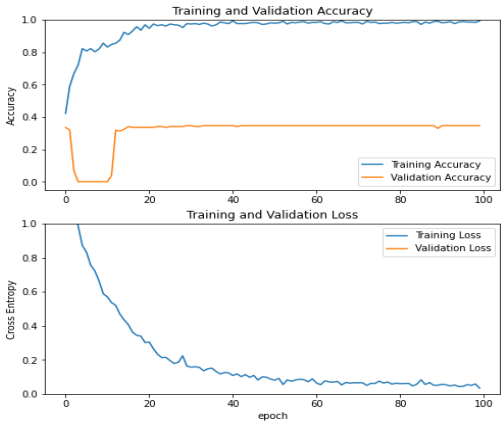
획득한 수어 영상은 공간 정보와 시간 정보를 포함한 프레임으로 나누어져야 한다. 두 가지 측면을 모두 모델링 하기 위해 고차원 데이터를 처리하는 데에 이상적인 CNN 모듈과 시간 시퀀스 데이터를 잘 처리하는 GRU(Gated Recurrent Unit) 모델의 장점을 결합한 GRU-CNN 하이브리드 구조를 사용한다.



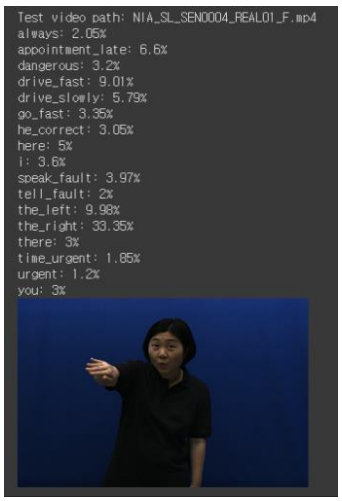

(그림 3) CRNN (GRU-CNN) hybrid neural networks

영상을 OpenCV 의 Video Capture()를 이용해 최대 프레임 수에 도달할 때까지 프레임을 추출하고, 영상의 길이가 최대 프레임 수보다 적은 경우 비디오를 0 으로 채워준다. 이를 통해 영상을 같은 수의 프레임 이미지로 만들어준 뒤 CRNN 모델에 넣어 학습시킨다. CRNN 모델은 KERAS 에서 제공하는 "*Video Classification with a CNN-RNN Architecture*"의 github 공개 코드를 이용해 구현하였다.

그림 4 는 CRNN 모델의 학습 곡선 및 학습 정확도를 보여준다. Epoch 는 100, BATCH_SIZE 는 64 로 설정하고, 수어 데이터셋을 모델에 입력으로 하여 학습하였다. 그림 5 는 학습이 완료된 모델의 test 결과이다. 해당 영상은 the_right 라는 수어 예시인데, test 결과에서 33.35%로 가장 높은 우도를 확보하고 the right 로 추론한 것을 확인할 수 있다.




(그림 4) CRNN learning curve

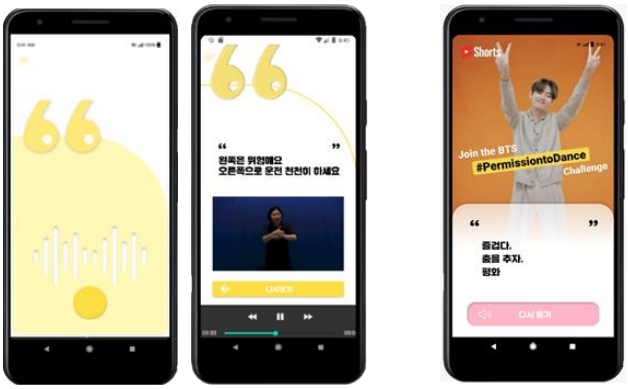



(그림 5) 모델 test 결과

## 4. 결론 및 추가 연구

본 프로젝트는 "눈속말"이라는 앱으로 제작 중이다. 팀원은 3 명의 백엔드 개발자 1 명의 프런트엔드 개발자, 1 명의 디자이너 총 5 명의 학부생으로 구성되어 있다. 현재 텍스트나 음성을 수어로 변경해 주는 비청각장애인 버전은 백엔드 개발과 프런트엔드 개발을 완료하였다. 청각장애인 버전은 CRNN 아키텍처 이용하여 수어 영상 분류 모델의 train 을 마친 상태이고, Validation data 를 통해 학습된 모델의 하이퍼 파라미터를 수정하여 정확도와 성능을 올리는 과정에 있다. 이후 실시간으로 앱을 통해 들어오는 영상을 모델 입력으로 넣고 test 하는 기술과 Google TTS API 를

이용하여 음성 서비스를 지원하는 코드를 추가한 뒤 프런트엔드와 연결할 예정이다.




비장애인버전 청각장애인버전

(그림 6) 비청각장애인 버전 음성 인식 화면 (좌), 인식된 음성의 수어 번역 결과 (중) 및 청각장애인 버전 예시 (우)

청각장애인과 비장애인의 쌍방향 소통을 위한 서비스는 청각장애인이 사회로 나아가기 위한 가장 기본적인 관문의 역할을 할 것이며 사회 참여를 돕고 소통이라는 장벽을 넘어서는 발돋움이 될 것이라 기대된다. 그뿐만 아니라 해당 연구로 진행된 알고리즘은 후에 필담을 할 수없는 상황이거나, 필담이 비효율적이라 생각하는 사람들을 위한 서비스로의 응용이 가능하다.

국적을 불문한 자유로운 소통을 원활하게 하기 위해 외국어 번역 서비스로 확대할 수 있다. 외국인과 소통할 경우 번역할 경우, [3.3 청각장애인 버전]과 [3.4. 비청각장애인 버전]에 번역 API 를 결합해 구현 할 수 있다. 국제 수어로 소통하기 위해선 국제 수어 데이터셋으로 변경하여 [3. 시스템 설계]의 과정을 진행하는 방법이 있다. 추가로 한국 수어 및 국제 수어 데이터를 모아 수어 사전 서비스를 제공하거나, 기억하고 싶은 수어 단어 또는 사전에서 찾은 단어를 저장하여 자신만의 수어 단어장을 만드는 서비스를 제공한다면 수어 교육의 기회가 확대될 것이라 기대된다.

## Acknowledgement

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1G1A1102683). 본 연구는 삼성미래기술육성센터의 지원을 받아 수행하였음 (No. SRFC-TC1603-52). 본 결과물은 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 사회맞춤형 산학협력 선도대학(LINC+) 육성사업의 연구결과임.

## 참고문헌

[1] 오영준, and 장훈. "수화통역시스템 설계 및 구현." 한국정보과학회 학술발표논문집 29.2II (2002): 691-693.

[2] 오영준, and 장훈. "수화통역시스템 설계 및 구현." 한국정보과학회 학술발표논문집 29.2II (2002): 691-693..

[3] 이동주, et al. "꼬꼬마: 관계형 데이터베이스를 활용한 세종 말뭉치 활용 도구." *정보과학회논문지: 컴퓨팅의 실제 및 레터* 16.11 (2010): 1046-1050.

[4] Yang, Xiaodong, Pavlo Molchanov, and Jan Kautz. "Multilayer and multimodal fusion of deep neural networks for video classification." Proceedings of the 24th ACM international conference on Multimedia. 2016.

[5] Kondratyuk, Dan, et al. "Movinets: Mobile video networks for efficient video recognition." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[6] Paul, sayak, editor. Video Classification with a CNN-RNN Architecture. Keras, 28 May 2021, keras.io/examples/vision/video_classification/.

[7] Wu, Lizhen, et al. "A short-term load forecasting method based on GRU-CNN hybrid neural network model."*Mathematical Problems in Engineering* 2020 (2020).