

양방향 지화 번역 장치 설계 및 구현

*김주완, 박정주, 정지훈, 김대진

전남대학교 전자컴퓨터공학부

e-mail : *kaka2654@hanmail.net, parkjeongju140@gmail.com,*

wlgns2234@naver.com, djinkim@jnu.ac.kr

Design and Implementation of Bidirectional Finger Language Translation Device

*Ju Wan Kim, Jeong Ju Park, Ji Hun Jeong, Dae Jin Kim

School of Electronic and Computer Engineering

Chonnam National University

Abstract

Sign language interpretation services through an interpreter are limited, so a sign language translator using artificial intelligence can help a lot in communication with the hearing impaired.

In this paper, a portable finger sign language translation device that can be connected to a mobile phone is designed and implemented.

The translation device translates finger motions and outputs them as voices, and recognizes the voice of the other party and outputs them as finger animations.

I. 서론

2020년 기준 청각장애인은 395,789명으로 해마다 꾸준히 증가하고 있다[1]. 청각장애인은 제 1 언어로 수화를 사용하지만, 소수의 일반인만 수화를 할 수 있으며 수화 통역사의 인원이 부족한 현실이다[2]. 통신 기술의 발달과 인공지능이 발전하면서 청각장애인의 의사소통 개선을

위해 수화 번역 장치[3], 스마트 장갑[4] 등 관련 연구가 활발히 진행되고 있다. 하지만 아직까지 완벽하게 인공지능 기반의 한국 수어를 구현하기엔 어려움이 있다. 본 논문에서는 스마트폰과 JETSON NANO 마이크로프로세서 보드를 이용하여 청각장애인이 청인과 대화 시 발생할 수 있는 의사소통 문제를 해결할 수 있는 보조기기로서 휴대용 양방향 지화 번역 장치를 설계하고 구현하였다.

II. 지화 번역 시스템 설계 및 구현

2.1 양방향 지화 번역 시스템

본 연구에서는 사용자인 농인이 왼쪽 손목에 착용한 카메라를 통해 오른손의 지화 동작 손가락 랜드마크 데이터를 추출하고 이를 기반으로 자·모음 모션을 판단하여 단어와 문장 기반의 음성으로 출력한다. 또한 상대방인 청인으로부터 들어오는 음성 데이터를 입력받고, 이를 기반으로 사용자인 농인이 핸드폰을 통해 애니메이션 캐릭터가 지화 동작을 수행하는 모습을 볼 수 있도록 한다. 양방향 지화 번역 시스템은 그림 1과 같이 동작 및 음성 인식 장치, 지화 생성기,

관리 서버로 구성되어 있다.

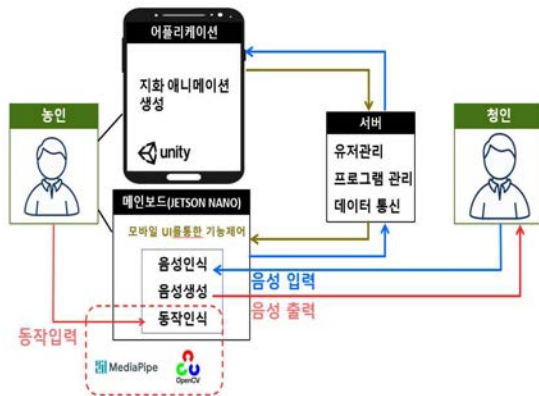


그림 1. 양방향 지화 번역기 시스템 구성도

2.2 동작 및 음성 인식 알고리즘

MediaPipe는 구글에서 제공하는 기계 학습 프레임워크로, 빠른 데이터 값을 추출할 수 있기 때문에 지화 동작 데이터값을 추출하는데 사용한다. 그림 2와 같이 지화를 하는 농아인의 손 좌표값 21 개의 랜드마크(landmark) 데이터를 이용한다. 이때 랜드마크 데이터는 $[x, y, z]$ 의 3 차원 좌표값을 반환한다.

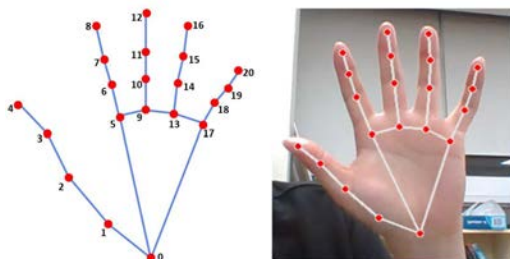


그림 2. Mediapipe 손모양 랜드마크

k-NN은 데이터가 속한 라벨이 이미 주어졌 있는 상황에서 이루어지는 인공지능의 지도 학습 방법으로, 예측하고자 하는 대상과 가장 가까운 k개의 대상을 찾는 기계학습 방법이다. 입력으로는 벡터와 레이블(label)을 사용한다. 본 연구에서는 Mediapipe를 통해 추출된 21개의 랜드마크 값을 통해 각 랜드마크 사이의 벡터 값 20 개를 추출하였다. 추출된 20 개의 벡터 값을 이용해 손가락 사이의 각도 데이터를 추출하였다. 각도 데이터는 손가락의 관절이 꺾이는 부분 15 개와, 엄지와 각각 나머지 4 개 손가락(검지, 중지, 약지, 새끼)과의 각도 5 개를 추출하였고, 주

가적으로 각도만으로 구별하기 힘든 지화 동작의 방향벡터 값을 추가하여 총 21 개의 입력 벡터 값을 학습에 사용하였다. 입력은 0~180 도의 값들을 0~1 사이로 정규화하였고, 출력 레이블 값은 지화 동작 데이터 31 개를 대상으로 0~1로 정규화 후 분류 학습을 진행하였다. 31 개의 지화 동작에 대하여 각각 400 개씩의 데이터를 제작하였다. 총 12,600 개의 데이터 중 10,000 개의 데이터를 학습하고 2,600 개의 데이터로 k의 값에 따른 정확도를 테스트 해본 결과 k값에 상관없이 모두 98% 이상의 정확도를 보여주었다. 이 값은 손모양 랜드마크 데이터가 추출된 상태라고 가정하고 k-NN 모델 정확도를 측정한 것이다. 분리된 자·모음은 Jamo 라이브러리를 이용해 단어로 변환하고, Google Speech API를 통해 음성으로 출력한다. 마찬가지로 마이크를 통해 들어온 음성데이터를 Google Speech API가 텍스트 문장으로 변환한다. 변환된 텍스트 형식의 문자를 Jamo 라이브러리를 통해 자·모음으로 분절하고 분절된 자·모음에 맞게 사용자가 지화 동작을 확인할 수 있도록 모바일 어플리케이션으로 전송한다.

2.3 지화 애니메이션 생성기

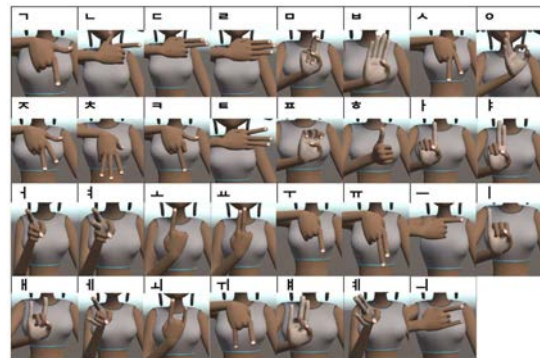


그림 3. 31개 지화 애니메이션 동작

유니티(Unity) 3D를 이용하여 지화를 애니메이션으로 시각화하여 제공하고, 번역기를 제어하는 UI를 제작하는 데에 이용한다. 그림 3)은 애니메이션 모델에 지화 모션을 적용하여 자·모음을 표현한 모습이다.

1) 아바타 출처 : <https://www.mixamo.com>

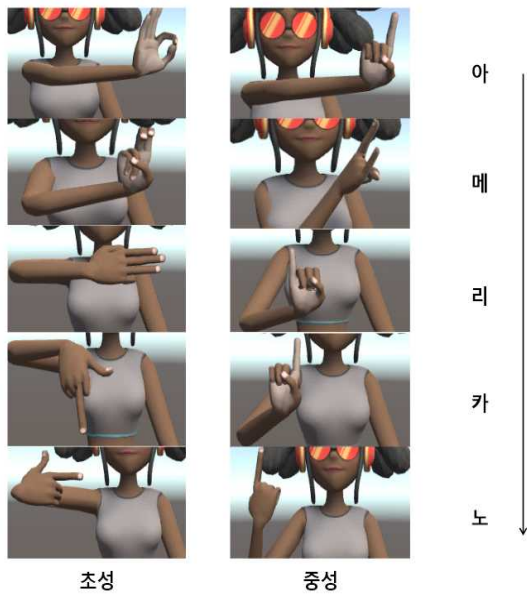


그림 4. 아메리카노 지화 표현

31 개의 지화 애니메이션 모델을 제작하여 서버로부터 받은 자·모음 데이터에 따라 한글 초성, 중성, 종성에 맞게 동작하도록 제작하였다. 본 연구에서는 단어의 길이가 최대 5 개까지만 동작하도록 설정하였다. 하나의 단어가 시작할 때는 왼쪽에서 지화가 시작하고 점차 오른쪽으로 이동하여 마지막 글자는 맨 오른쪽에서 끝나도록 설계하였다. 그림 4는 “아메리카노” 단어의 지화가 왼쪽에서 시작하여 오른쪽으로 끝나는 예시를 보여준다.

2.4 TCP/IP 소켓통신

사용자가 착용하는 메인 MCU 보드와 어플리케이션 간에 소켓 통신 방식을 이용하였다. 서버를 통한 실시간 데이터 통신을 통해 자·모음 데이터를 전송하고 프로그램을 관리할 수 있다. 또한 서버 방식의 관리를 통해 서비스 사용자의 지화 모션데이터, 프로그램 등을 관리한다.

2.5 구현한 사용자 화면

MCU 보드는 NVIDIA사 JETSON NANO 4GB, JETSON NANO 호환 무선랜 카드 Intel 8265AC, JETSON NANO T208 UPS, USB 카메라, USB 마이크, USB 스피커로 구성되어 있다. 카메라는 사용 시 사용자의 손만 인식할 수 있게 왼쪽 손목 위에 부착하였다. 사용자는 손목에 장착된 카메라를 통해 지화 동작을 수행하고

합성된 문장을 스피커를 통해 출력한다. 또한 마이크에서 상대의 음성을 받고 어플리케이션에서 분절된 자·모음에 맞는 지화 애니메이션을 확인할 수 있다. 이 두 가지 동작에 대해 핸드폰 모바일 환경에서 어떤 기능을 사용할 것인지 사용자에게 유니티 3D UI를 제공한다.



그림 5. 모바일 UI 화면

IV. 결론 및 향후 연구 방향

본 논문에서는 최근 연구되고 있는 수화 인공지능 번역과 애니메이션 모델링, 실시간 소켓 통신을 이용하여 휴대용 지화 번역 장치를 구현하였다. 각각 자신들의 언어로 소통을 할 수 있어서 서로의 정체성과 문화를 해치지 않고 의사소통을 할 수 있게 된다. 또한 휴대성을 갖추으로써 야외 활동 시에도 도움을 줄 수 있을 것이다.

개발된 장치는 k-NN 모델뿐만 아니라 CNN 등과 같은 딥러닝 기반의 모델을 이용하여 MCU 보드가 아닌 서버 환경에서 구동시킨다면 좋은 성능을 기대할 수 있을 것이다. 또한 자연어 전처리, 오타 수정 등의 다양한 알고리즘을 적용시켜 활용성을 높일 수 있다. 향후 연구 방향으로 표정과 움직임까지 확장하여 데이터를 추출하고 실시간 번역 서비스를 구축하여 휴대용 수화 번역 장치를 개발할 계획이다.

참고문헌

- [1] 2021 장애통계연보, 한국장애인개발원, 2021.
- [2] 박종엽, “농인의 경험을 통한 수화통역센터

- 역할과 기능에 관한 해석학적 연구”, 한국장애인복지학, 제41권 제40호 pp.221-251, 2018.
- [3] 임현준, “딥러닝 기반의 음성-수화 번역 시스템 설계 및 구현”, 한양대학교 석사학위 논문, 2019.
- [4] 스마트 장갑을 이용한 수화 인식 인터랙션 시스템 연구, 임재철, 김광현, 반영환, 박가영, 정령, 2015.
- [5] 한국판 농인 문화적응 척도 개발 및 타당화, 한국사회복지학, 제66권 제15호 pp.55-73, 음영지, 손진훈, 박지은, 손선주, 임진섭, 2014.