조음장애인용 Al스피커를 위한 머신러닝 기반 언어교정기 개발

이동헌⁰, 문미경^{*} ⁰동서대학교 컴퓨터공학부,

*동서대학교 컴퓨터공학부

e-mail: zone7907@gmail.com^o, mkmoon@dongseo.ac.kr*

Development of a Machine Learning-based Language Corrector for Al Speakers of Patients with Articulation Disorders

DongHeon Lee^O, Mikyeong Moon*

^oDivision of Computer Engineering, Dongseo University, Korea,

*Division of Computer Engineering, Dongseo University, Korea

• 요 약 •

최근 인공지능의 발달로 인해 AI스피커에 대한 연구가 활발히 이루어지고 있다. 조음장에는 구강 안에서 말소리를 제대로 만들지 못해서 제대로 된 언어를 구시하지 못하는 장애를 말한다. 조음장애인들이 AI스피커를 사용하면 발음을 제대로 인식하지 못하기 때문에 사용의 어려움이 있다. 본 논문에서는 경증 조음장애인들이 AI스피커를 이용할 수 있도록 머신러닝 기반 언어교정기의 개발내용에 관하여 기술한다. 이는 언어로 명령 줄 수 있는 여러 시스템에 활용될 수 있을 것으로 기대한다.

키워드: 머신러닝(Machine Learning), 개인화(Personalization)

I Introduction

최근 4차 산업혁명이 진행되면서 인공자능에 대한 관심이 증가하였다. 그중 가장 대중화된 AI스피커에 대한 연구가 활발히 진행되고 있다. Statista에서는 글로벌 음성인식 시장의 경우 2015년 5억9990만 달러를 기록하고 해마다 점점 성장하고 있으며, 2024년에는 7000억 달러 규모로 성장할 것이라고 추정하고 있다. AI스피커가 제공하는 서비스로는 음악재생, 날씨, 시간, 알람, 타이머 등의 기능이 있다. 그중 가장 대중적으로 사용되는 가능은 음악재생이며 음악을 추천해주는 기능도 존재한다.

조음장에는 발성에 관여하는 근육의 조절능력 장애로 인하여 나타 나는 운동성 언어장애로 언어치료를 받는 환자 중 가장 많은 수를 차지한다. AI스피커는 자연어처리를 통해 명령어를 해석하는데 조음 장애인들은 발음을 정확히 구사하지 못해 원하는 명령을 내리는데 어려움이 있다.

본 논문에서는 경증 조음장애인용 AI스피커 사용을 위한 머신러닝 기반 언어교정기의 개발 내용에 관해 기술한다. 이 AI스피커를 통해서 경증 조음장애인들은 AI스피커를 자기 자신의 발음을 인식시켜 명령할 수 있으며, 이를 통해 삶의 질을 개선시킬 수 있음을 기대한다.

II. Preliminaries

연귀[1]에서는 신경학적 손상에 의한 언어장애인 음성 인식률 개선에 관한 연구가 진행되었다. 해당 연구에서는 언어 장애인 구강 특성에 따른 어려운 발성 부분을 인위적으로 보정해주는 MEMS Microphone 장치를 개선하는 연구를 진행하였다. H/W와 S/W 특성을 감안하여 유사 사전을 만들어 원활한 의사소통을 위한 말 명료도 향상에 기여하였다. 본 논문에서는 머신러닝을 통해 언어 인식률을 향상하는 방법에 대해 연구한다.

III. Machine Learning-based Language Corrector

1. 머신러닝을 위한 데이터 셋 생성 방법

본 논문에서는 음성인식을 KT의 GIGA 지니 음성인식 API를 사용하였다. 그리고 데이터베이스는 MySQL을 사용하였다. 음성인식을 위해 경증 조음장애인의 음성을 Pyaudio를 통해서 음성을 입력받는다. 그리고 저장된 녹음 파일을 KT GIGA 지니 음성인식 API를 통해 음성을 텍스트로 변환한다. 변환된 텍스트를 데이터베이스에 해당 명령어 텍스트에 저장을 하여 데이터 셋을 생성한다. 예를 들어

저장하라는 데이터가 '오케이 구글'이면 사용자가 직접 음성을 입력하여 데이터를 삽입한다. 발음이 부정확하여 오케이 구글이 아닌 '오에이 구그', '오헤히 후을', '오헤이 구그', '호케히 후글', '호헤이 후홀', '오헤기 구흐' 등 다양한 데이터가 쌓이게 된다. 학습을 위해 최소 20개 이상의 데이터가 쌓여야 데이터 모델 생성이 가능하다.

ACKNOWLEDGEMENT

본 연구는 과학기술정통신부 및 정보 통신기획평가원의 SW중심 대학지원사업의 연구결과로 수행되었음(2019-0-01817)

2. 머신러닝를 이용하여 데이터 모델 생성

머신러닝은 Machine Learning for kids를 사용하여 데이터 모델을 생성한다. Machine Learning for kids에서 텍스트를 이용하여 머신러 닝을 진행하는데 기계학습 모델은 Watson Conversation의 API를 사용하여 학습한다. 학습하려는 클래스를 생성하여 해당 클래스가 어떤 데이터가 들어가야 하는지 구분한다. 그리고 앞서 생성한 데이터 셋을 클래스마다 삽입해줌으로써 학습을 통해 데이터 모델을 생성한다.

예를 들어 명령을 내리기 위해 '오케이 구글' 이라는 단어를 학습하기 위해 google 이라는 클래스를 생성하여 클래스에 데이터를 넣어학습을 시켜서 해당하는 데이터 모델을 생성한다. Table 1은 AL스파커에서 가장 많이 사용하는 명령어를 데이터 모델로 생성한 결과를 보여준다. 해당 명령어에 대해서는 부정확한 발음으로 명령을 내려도 인식이 가능하도록 하였다.

Table 1 클래스 명 및 명령어 종류

구분	클래스 명	명령어
명령	google	오케이 구글
날씨	weather	오늘 날씨 어때?
	rain	내일 비 와?
	dust	미세먼지 어때?
날짜	date	오늘 며칠이야?
	time	지금 몇 시야?
음악	music	노래 틀어줘
	next	다음 곡 재생
	pre	이전 곡 재생
	stop	노래 꺼
볼륨	up	볼륨 높여줘
	down	볼륨 줄여줘

IV. Conclusions

본 논문에서는 경증 조음장애인용 AI스피커 사용을 위한 머신라닝 기반 언어교정기에 대해 기술하였다. 본 언어교정기를 통해 경증 조음장애인도 AI스피커에 명령을 전달하여 서비스를 받을 수 있다. 향후 계획으로는 더 많은 데이터를 학습하여 인식률을 높이는 것과 단어 학습을 통해서도 문장 인식이 되도록 하는 연구가 진행될 것이다.

REFERENCES

[1] Hyung-keun Lee, Soonhub Kim, KI-Woong Yang, "A Mobile Emotion Recognition App with Deep learning Model Compression for the Blind and Visually Impaired" The Korea Institute of Information and Communication Engineering, 2019.11, 1397-1406(10 pages)