라즈베리파이를 활용한 개인 맞춤형 음성인식 엔진조사

장서연O, 이강희*
O숭실대학교 글로벌미디어학부,
*숭실대학교 글로벌미디어학부
e-mail: syhoneyjam@naver.comO, kanghee.lee@ssu.ac.kr*

A Survey on Personalized Voice Recognition Engine Using Raspberry Pi

Seo-Yeon JangO, Kang-Hee Lee*
OGlobal School of Media, Soongsil University,
*Global School of Media, Soongsil University

● 요 약 ●

라즈베리파이는 교육용 프로젝트의 일환으로 개발된 임베디드 보드로서 모듈 확장성이 용이하여 사용자가 원하는대로 용도를 변경하거나 기능을 확장할 수 있다. 현재 한국 소프트웨어 교육은 프로그래밍 언어와 이론 위주의 경향이 짙은데, 이는 초기 학생들의 프로그래밍에 대한 흥미를 저하시키는 데 가장 큰 걸림돌이 되고 있다. 본 논문에서는 라즈베리파이를 활용하여 사물인터넷의 대표적 입력수단으로 활용되고 있는 음성인식을 구현하는 방법에 대해 논한다. 또한 향후에는 TPU(Tensor Processing Unit)의 전용 소프트웨어를 사용하여 대용량의 실시간 음성인식을 GPU를 이용하여 구현함으로써 실제 물리적인 장치들을 프로그래밍으로 제어함으로써 소프트웨어의 현실 통제 가능성을 직접 체험하여 음성인식뿐만 아니라 동작 원리 및 기저 기술들에 대한 관심을 불러일으키는 하나의 좋은 교육 방법이 될 수 있다.

키워드: 라즈베리파이(Raspberry Pi), 음성인식(Voice Recognition), 맞춤형(Customized)

I. 서론

인공지능과 빅데이터 등의 기술을 기반으로, 인간의 생활패턴을 디지털화하여 정보를 활용하고 공유하는 것뿐만 아니라 주어진 정보를 가공하여 새로운 형태의 데이터를 창조해내는 시대의 흐름에 있다. 이러한 기술들은 산업 전체에 새로운 비즈니스 모델을 제시하며, 자체의 산업적 중요성뿐 아니라 전체 산업과 사회, 국가시스템을 움직이는 '회로도'이자 '알고리즘' 역할을 한다는 점에서 중요성이 갈수록 커질 전망이다[1]. 그에 따라 소프트웨어 교육 문화 확산과 활용 능력 교육은 미래 국가경쟁력을 결정지을 주요 요소이기에, 전 세계 주요국의 중요한 교육 정책으로 자리 잡고 있다. 하지만 소프트웨어적 한계인 '다소 현실로 와닿지 못하다' 라는 측면과 한국의 경우, 프로그래밍 언어에 초점을 맞춘 교육으로 인해 여전히 접근이 어려운 분야로 인식되는 것이 실정이다. 이를 극복하기 위해서는 모니터 속에서 그치는 교육이 보다는 실제 현실의 문제를 해결할 수 있는 다양한 측면의 접근방법과 응용 능력을 키우는 교육을 지향해야한다[2]. 본 논문에서는 효과적인 소프트웨어 교육 도구로 라즈베리파이를 제안하며, 현재 사물 인터넷의 입력수단으로 많이 활용되고 있는 음성인식을 활용한 인공지능 스피커를 모듈 확장성이 용이한 라즈베리파이와 무료 오픈 소스이며 가장 기본이 되는 음성인식 엔진을 활용하여 자신만의 인공지능 스피커를 제작할 수 있는 방법을 논의할 것이다. 또한 소프트웨어의 현실 통제 가능성을 직접 체험하여 음성인식 그 자체뿐만 아니라 동작 원리 및 기저 기술들에 대한 관심을 더욱 강화하기 위한 효과적인 방법을 모색해보고자 한다.

II. 라즈베리파이의 음성인식엔진

1. Pocketsphinx

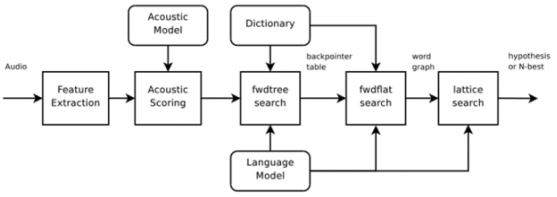

Fig. 1. Pocketsphinx 다이어그램

CMU(Carnegie Mellon University)에서 개발한 오픈 소스 음성인식 소프트웨어인 스핑크스의 경량 버전이다. 음향 모델과 언어 모델 및 단어 사전 또는 문법 파일을 이용하여 처리하기에 모델과 파일들이 필요하다. 음향 모델은 직접 녹음된 음성 파일을 훈련하여 빌드하거나 기존의 빌드된 모델 파일들을 활용한다. 다음 사이트에서 다운로드할 수 있다.

https://sourceforge.net/projects/cmusphinx/files/Acoustic%20and%20Language%20Models/

적극적으로 개발 중이며 GMM(Gaussian Mixture Model) 계산을 위한 고정 포인트 산술 및 효율적인 알고리즘과 같은 기능을 통합한다. 모든 처리는 라즈베리파이에서 이루어지므로 느린 편이나, 오프라인 사용이 가능하며 실시간 음성인식을 지원한다.

2. Google Speech API

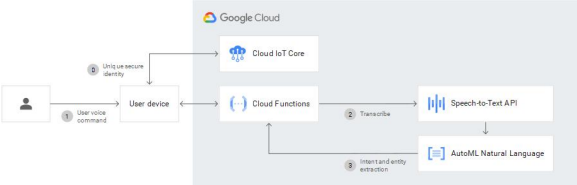

Fig. 2. Google Speech API 다이어그램

음성 정보를 네트워크를 통해 구글 서버로 전송하여 인식 결과를 전달받는 API이다. 구글의 머신러닝 기술을 이용하여 오디오를 텍스트로 변경하는 것뿐만 아니라 노이즈 켄슬링과 스트리밍, 멀티채널 인식, 그리고 전 세계 120가지 이상의 언어와 방언을 인식한다[3].

3. Julius

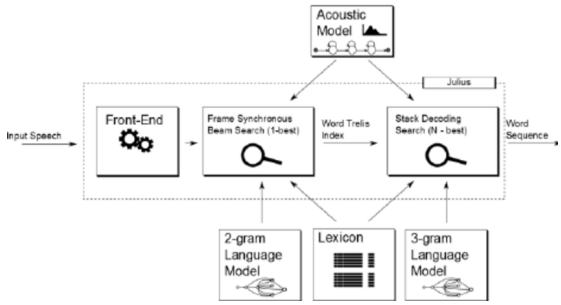

Fig. 3. Julius 다이어그램

일본 내 대부분의 음성인식, 음성인터페이스 연구기관에서 이용되고 있는 오픈 소스 소프트웨어이다. 2패스 탐색에 기초하는 인식 알고리즘을 이용하고 있으며, 수만 개의 어휘 인식을 실시간으로 실행할 수 있다. 프로젝트 초중기에는 CMU Sphinx과 유사하게 음성인식을 위한 일반적인 기술과 전통적인 기계학습 방식으로 개발이 되었으나, 후반기에 DNN(Deep Neural Network)을 통한 딥러닝을 접목하며 인식률이 상승했다. 현재 지원되는 언어는 C/C++이며, 언어모델로는 일본어와 영어를 지원한다[4].

다음 사이트에서 다운로드가능하다.

http://julius.osdn.jp/en_index.php

https://github.com/julius-speech/julius

4. 자신만의 라이브러리 운영하기

자체 음성인식 라이브러리를 구축하는 방법이다. PortAudio 를 통해 ALSA 를 사용하여 USB 마이크에서 오디오를 녹음하고 libsndfile을 통해 FLAC 파일에 저장한 후 Google로 전송하는 처리 과정을 거친 뒤 JSON 파일을 받아 가공한다. 굉장한 시간과 노력이 드는 작업이나, 전적으로 개발자가 모든 것을 통제할 수 있다.

III. 결론

본 논문에서는 라즈베리파이와 음성인식 엔진을 활용하여 인공지능 스피커의 기반을 다지는 방법을 다루었다. 실제 물리적인 현실을 설계하고 제어하며 소프트웨어와 하드웨어 사이의 데이터 처리 과정을 고려하며 프로그래밍하는 교육은 현재의 모니터에만 머무는 교육의 훌륭한 보완재이다. 기존 인공지능 스피커와 달리 확장성이 뛰어난 라즈베리파이를 이용하여 자신의 프로젝트 목표에 부합하는 다양한 엔진들을 사용함으로써 본인이 원하는 기능만을 가진 하나의 최적화된 인공지능 스피커를 만들 수 있는 기반을 갖출 수 있다는 점에서 의의가 있다.

향후에는 GPU와 구글의 오픈소스인 TPU(Tensor Processing Unit)[5]의 전용 소프트웨어를 사용하여 대용량의 실시간 음성인식을 구현하고자 한다.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음(2018-0-00209)

REFERENCES

[1] http://www.dt.co.kr/contents.html?article_no=2019112902101631650002

[2] YoungSun Park, SangJin An, YoungJun Lee. (2014). A Study on the Programming Education Using Raspberry Pi. Proceedings of the Korean Society of Computer Information Conference , 22(1), 451-453.

[3] https://cloud.google.com/speech-to-text?hl=ko

[4] http://julius.osdn.jp/en_index.php?q=index-en.html#documentation

[5] Sanghong Kim, Bowon Lee. (2018). Korean Speech Recognition System Based on Neural Network Training Using Raspberry pi. The Institute of Electronics and Information Engineers, (), 349-352.