한국어 발음 훈련을 위한 음성 인식 웹 사이트 - 바름

민정혜*, 강교진O, 김인기*
*인하공업전문대학 컴퓨터정보과,
O인하공업전문대학 컴퓨터정보과
e-mail: jhmin@inhatc.ac.kr*, ksmug1998@naver.comO, dlsrl5515@gmail.com*

Speech Recognition Website for Korean Pronunciation Training - Baleum

Junghye Min*, Gyo Jin KangO, In Gi Kim*
*Dept. of Computer Science, Inha Technical Collage,
ODept. of Computer Science, Inha Technical Collage

● 요 약 ●

본 논문에서는 외국인과 발음에 어려움을 겪고 있는 한국인들을 대상으로 음성 녹음을 진행하여 점수를 반환받는 웹 사이트를 소개한다. 이 웹 사이트의 목적은 사용자들의 발음 향상을 돕는 것이다. 음성 인식 API와 발음 평가 API를 사용하여 사용자의 발음을 정확하게 평가하고 피드백을 제공함으로써, 외국어 학습자와 발음에 어려움을 겪는 한국인들이 보다 원활하게 의사소통할 수 있도록 돕는다. 향후 연구로는 이 시스템의 사용자들에게 학습 성취에 대한 동기 부여를 하는 기능을 추가해 학습 효과를 높이도록 개선할 것이다.

키워드: 발음 평가(pronunciation evaluation), REST API

I. Introduction

오늘날 세계화된 시대에 다양한 언어와 문화 간의 효과적인 의사소통은 점점 중요해지고 있다. 새로운 언어를 배우는 데 있어서 발음 습득은 의도한 메시지를 전달하는데 중요한 역할을 하는 것으로, 특히 외국인들이 한국어를 배우거나 발음에 어려움을 겪는 한국인들에게 중요하다.

이러한 문제를 해결하기 위해, 본 논문에서는 사용자의 발음을 녹음하고 점수를 반환하는 웹 기반 플랫폼을 소개한다. 이 웹 애플리케이션은 외국어 학습자뿐만 아니라 발음에 어려움을 겪는 한국인들의 발음 개선과 의사소통 능력을 향상시키는 것을 목표로 한다. 최신 인공지능을 이용한 음성 인식 API와 발음 평가 API를 사용하여 사용자의 발음을 정확하게 평가하고 피드백을 제공함으로써, 더욱 원활한 의사소통을 도모할 수 있다. 향후에는 이 시스템의 정확도와 효과성을 높여 교육 및 의사소통 도구로 활용할 수 있도록 개선할 것이다.

II. Preliminaries

1. Related works

1.1 Background

현재 한국의 문화 콘텐츠는 전 세계적으로 뻗어 나가고 있다. Fig.1은 한국 방송 프로그램의 수출이 점증하고 있다는 것을 보여주며 이 상승세는 시간과 비례하여 더욱 높아질 것이다.

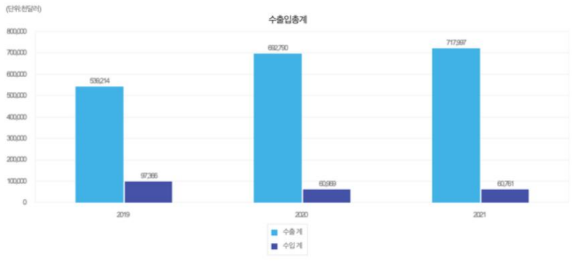

Fig. 1. 2019~2021년 방송 프로그램 수출입 현황

뿐만 아니라 외국인의 한국 유입량도 이전보다 많아지고 있다. 한국을 방문하는 외국인들의 목적에는 여행, 사업, 학업 등 다양하다.

Fig.2에 따르면 2004년을 기점으로 오늘날에 이르기까지 많은 외국인 유학생들이 학업을 위해 한국을 방문하는 것을 확인할 수 있다.

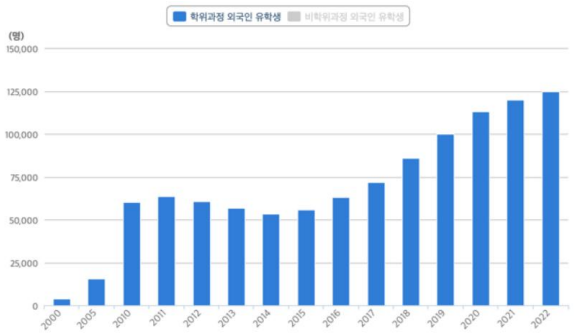

Fig. 2. 연도별 국내 외국인 유학생 현황(1999-2022)

위 자료를 토대로 한국을 바라보는 외국의 시각을 인지하고 그에 따른 한국어에 대한 관심도 향상을 유추해 볼 수 있다.

III. The Proposed Scheme

1. Materials & Method

1.1 Develop Environment

프로젝트 개발 환경은 다음과 같다.

Tomcat 서버를 기반으로 MVC 디자인 패턴을 이용한다. Fig. 3에서 확인할 수 있듯이 Java Spring boot를 이용하여 Model, Controller를 구현하며 React.js를 이용하여 View를 구현한다. 즉, Java Spring boot로 구동하는 Back end 서버와 node 서버가 분리된 서버로 존재한다.

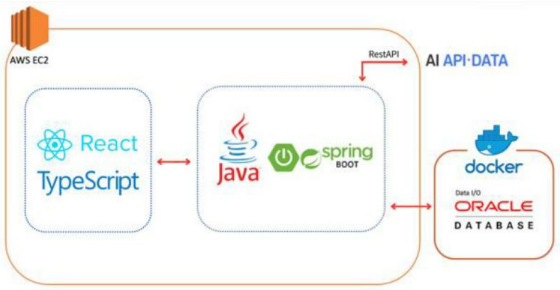

Fig. 3. 시스템 아키텍처

ETRI(한국전자통신연구원)의 음성 인식 및 발음평가 API를 사용하며 Oracle Database를 Docker를 이용하여 관리한다.

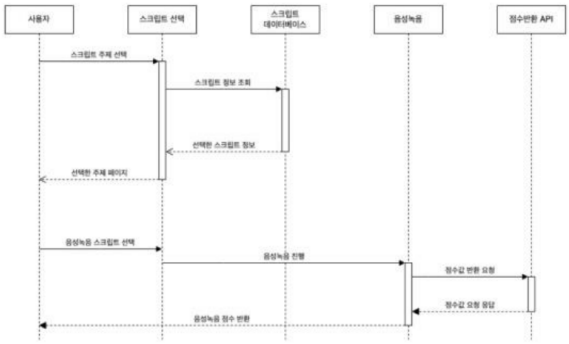

Fig. 4. 순차 다이어그램(Sequence Diagram)

Fig. 4는 시스템을 사용하는 순서를 기준으로 작성한 다이어그램이다. 먼저 사용자는 주제에 따라 항목화된 스크립트(발음할 텍스트)를 선택한다. 선택한 스크립트는 데이터 베이스에서 조회되고 조회가 완료되면 선택한 스크립트에 해당하는 녹음 페이지를 사용자에게 반환한다.

사용자는 녹음 인터페이스를 통해 녹음을 진행한다. 이후 녹음이 완료된 파일을 REST API 통신 방식으로 전송하여 점수를 반환받아 사용자에게 제공한다.

1.2 Pronunciation API

본 연구에서 사용하는 API는 한국전자통신연구원(ETRI)에서 제공하는 발음 평가 API이다.

Fig. 5는 인공지능 기술을 체험할 수 있는 공공 인공지능 오픈 API/DATA 서비스 포털 사이트이다.

Fig. 5. ETRI AI API/DATA

발음 평가 API는 최신 인공지능 기술에 기반하여 한국인을 비롯한 비원어민 영어 발성 및 외국인의 한국어 음성에 대해 높은 정확도의 음성인식 엔진과 높은 수준의 분석 알고리즘에 기반하여 문장별 발음 수준을 측정하여 제공한다. 본 API는 HTTP 기반의 REST API 인터페이스로 JSON 포맷 기반의 입출력을 지원하며 ETRI에서 제공하는 API Key 인증을 통해 사용할 수 있는 Open API이다.

Fig. 6. 다양한 AI기술이 접목된 API

본 연구에 사용된 발음 평가 API 이외에도 ETRI에서는 Fig. 6와 같이 AI 기술이 접목된 다양한 API를 제공한다.

1.3 Details

본 논문의 서비스 파일 처리 부분 및 REST API 통신은 다음 Fig. 7과 같이 이뤄진다.

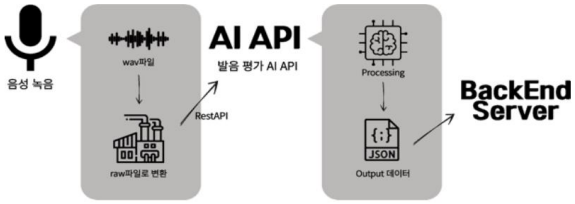

Fig. 7. 세부 처리 로직

녹음 완료된 사용자의 음성 파일은 REST API 통신을 위해 사전에 규정된 파일 포맷(.raw)으로 변경하는 과정을 거쳐야 한다. 녹음 인터페이스에서 wav 포맷의 음성 파일을 서버로 전송하고 서버는 이를 raw 포맷으로 Converting한다.

요구 포맷으로 전처리를 완료한 후 ETRI Open API 서버에 요청을 진행한다. 요청이 성공적으로 처리되면 높은 정확도의 음성인식 엔진과 분석 알고리즘을 탑재한 AI 모델을 이용하여 파일 분석을 통해 녹음된 음성의 발음 평가 점수를 JSON 형식으로 서버에 반환한다.

2. Implementation results

Fig. 8은 사용자가 스크립트 목록 페이지 url로 접근한 화면이다. 왼쪽 사이드 바에서 원하는 항목을 선택할 수 있고 해당 항목에 matching 되는 스크립트들을 query를 이용하여 추출한 후 화면에 노출한다.

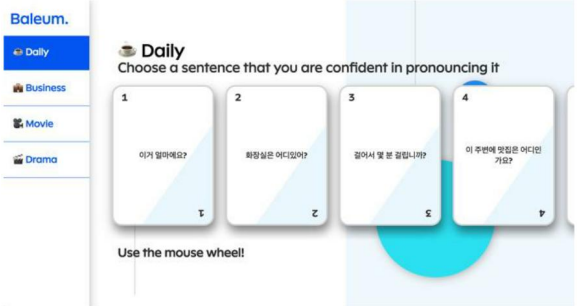

Fig. 8. 항목화된 스크립트 선택 화면

스크립트를 선택하면 Fig. 9 modal이 등장한다. 이는 사용자가 스크립트를 직접 읽고 발음을 녹음하는 기능을 수행한다. 서비스의 주요 타깃이 외국인이라는 점을 고려하여 TTS(Text To Speech) 기능을 추가하였다. “Listen to the sentence” 버튼을 클릭하면 스크립트 텍스트를 TTS 변환하여 음성으로 제공한다.

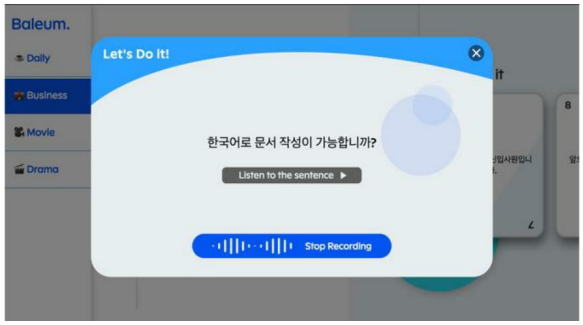

Fig. 9. 음성 녹음 인터페이스 화면

하단의 “Recording” 버튼을 이용하여 녹음을 진행한다. 버튼을 클릭하면 음파 UI가 움직이며 녹음 진행 상태 정보를 제공하며 만약 버튼을 클릭한 뒤 3초 안에 ETRI Open API 서버로부터 데이터를 응답받지 못할 경우 0점을 반환한다. 이를 통해 서버의 응답이 장시간 지연되거나 네트워크에 문제가 발생했을 경우에 사용자가 비정상적인 서비스를 제공받지 않도록 예외 처리를 수행한다.

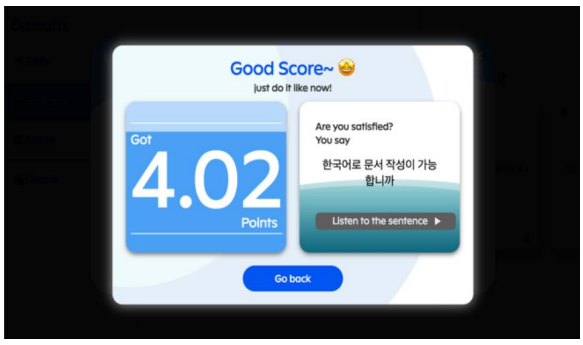

Fig. 10. 응답 점수 제공 화면

성공적으로 녹음을 마치면 보이는 Fig. 10과 같이ETRI Open API 서버로부터 응답받은 데이터를 화면에 노출한다. 왼쪽 섹션은 점수를 시각화하고 오른쪽 섹션은 사용자의 음성 파일을 STT(Speech To Text) 변환을 통해 텍스트로 제공한다. 이는 한글을 일정 수준 이상 학습하지 못한 사용자들에게 자신이 실제로 어떻게 발음했는지 알려주는 지표로 사용된다.

IV. Conclusions

본 논문에서는 세계적으로 높아져가는 한국의 문화 콘텐츠 인식을 파악하여 외국어 학습자와 발음에 어려움을 겪는 한국인을 위한 웹 기반 한국어 발음 평가의 장을 제공하였다. 해당 서비스는 현재 한국어라는 하나의 언어에 국한되어 있지만 지속적인 개발을 통해

한국어 이외의 다양한 언어들을 발음하고 학습할 수 있는 세계 언어 학습 플랫폼으로 확장될 수 있다고 기대된다.

REFERENCES

[1] E-나라지표, “Current Status of Export and Import of Broadcasting Programs” https://www.index.go.kr/unity/potal/eNara/main/EnaraMain.do?cdNo=000

[2] KESS(교육통계서비스) , “International students in Korea by year (1999-2022)” https://kess.kedi.re.kr/index

[3] ETRI(한국전자통신연구원) AI API.DATA, “발음 평가 API, 음성 인식 API” https://www.etri.re.kr/kor/main/main.etri, https://aiopen.etri.re.kr/demo/Pronunciation, https://aiopen.etri.re.kr/guide/Recognition