

언어 학습 음원 분석 방법 및 언어 학습 음원을 재생하는 전자 디바이스 연구

송규빈*, 오정현**, 황채원**, 유동완***

*강남대학교 컴퓨터공학과

**가톨릭대학교 컴퓨터정보공학부

***성공회대학교 글로벌 IT 학과

E-mail : thd0427__@naver.com

LANGUAGE LEARNING SOURCE ANALYSIS METHOD AND ELECTRONIC DEVICE FOR PLAYING LANGUAGE LEARNING SOURCE RESEARCH

Gyu-Bin Song*, Jeong-Hyeon Oh**, Chae-won Hwang**, ***Dong-Wan Yu

*Dept. of Computer Science, GangNam University

**Dept. of Computer Science Information Engineering, Catholic University of Korea

***Dept. of Global IT, SungKongHoe University

요 약

언어 학습 음원 분석 방법 및 언어 학습 음원을 재생하는 전자 디바이스 연구로, 음원을 문장 단위로 분할하여 스크립트화하는 것을 주요 목표로 한다. 분석과정은 크게 세단계로 나눌 수 있다. 무음 구간 분석, 음원 분할 및 STT 구간, 스크립트 재구성이다. 이런 분석 과정을 통해 나온 결과물의 정확도는 90%로서 본 연구의 목표를 달성한다.

1. 서론

1) 연구의 필요성

본 연구는 언어 학습에 도움을 주는 음원 분석 방법 및 음원 재생 전자 디바이스에 대한 것으로, 보다 구체적으로 토익 리스닝 파트에 대한 학습을 도와주는 언어 학습 음원 분석 방법 및 언어 학습 음원을 재생하는 전자 디바이스에 관한 것이다.

본 연구에서는 공인 영어 능력 평가 시험 중 하나인 토익(TOEIC)에서 리스닝 파트를 연구 자료로 활용하였다. 토익은 현재 가장 많은 수험생들이 시험 본 영어 능력 평가 시험이다. 듣기와 읽기 능력에 대한 평가로 듣기는 part1에서 part4 까지, 읽기는 part5에서 part7 까지 구성되어 있다. 토익은 취업을 위한 필수 관문으로 여겨지며 매년 많은 학생 및 취업준비생들이 응시하는 시험이다. 이러한 토익 시험의 점수 향상을 위해서 오답 노트 작성과 같은 오류 점검 과정이 필수적으로 요구된다.

한편, 현재의 토익 리스닝 학습 시 틀린 문제나 다시 듣고 싶은 문제를 찾을 때, 음원 상태바를 움직여 수동적으로 원하는 부분을 찾아야 되기 때문에 원하는 부분을 찾지 못하는 문제점이 있다. 또한, 학습의 효율성을 위해 파트별 혹은 문제별 분할된 음원은 추가로 비용을 지불해야 획득할 수 있

기 때문에 사용자에게 부담이 될 수 있다.

이러한 불편함에 대응하는 언어 학습 음원 분석 방법 및 기술에 대한 필요성 검증을 위해 20 대 토익 학습 경험자 100 명을 대상으로 설문 조사를 진행하였다. 설문 참여자 중 ‘토익 학습 시, 어떤 애플리케이션을 사용했는가’ 문항에 대한 응답으로 90%가 ‘기본 미디어 플레이어를 사용했다.’를 선택하였다. 이는 대부분의 응답자들이 별도의 애플리케이션 도움 없이 갖고 있는 음원만 이용해 공부해 왔음을 보여준다. 이 중 ‘리스닝 파트 학습 시, 어떤 기능이 필요하다고 생각했는가’ 문항에 대한 응답으로 60%가 ‘원하는 부분으로 찾아가는 기능’을 선택했고 18%가 ‘속도 조절 기능’, 또 다른 18%는 ‘구간 반복 기능’을 선택하였다. 이는 기본 미디어 플레이어를 사용했던 토익 학습자들에게 학습 중 불가피하게 겪게 되는 불편함을 해소하고 학습의 효율성을 증대하는 다양한 기능에 대한 수요가 있다는 것을 보여준다.

이에, 토익을 준비하는 사용자들이 토익 리스닝 공부를 보다 쉽고 편리하게 할 수 있도록 돕는 학습 어플리케이션 및 이를 실행하는 전자 디바이스의 개발이 필요하다. 또한, 파트별 또는 문제별 분할 음원 재생을 제공할 수 있는 기술이 필요하다.

2) 연구의 해결과제

본 연구가 해결하고자 하는 과제는 사용자가 원하는 문제를 빠르게 다시 들을 수 있고 웨도잉(따라 말하기)을 더 편하게 하는 학습 분석 엔진과 이를 적용한 어플리케이션 및 디바이스를 제공하는 것이다. 또한, 본 발명이 해결하고자 하는 과제는 기존의 초 단위 음원 분석이 아닌 문장 단위 재생을 제공하는 것이다. 또한, 본 연구가 해결하고자 하는 과제는 추가 비용없이 파트별 또는 문제별 음원 재생 기능을 제공하는 것이다.

2. 본론

1) 분석과정

본 논문의 서론에서 설명된 서비스에 대한 분석엔진을 구현한 방법과 그 과정의 내용은 다음과 같다. 분석 엔진은 음원을 문장단위로 분할하여 스크립트화 하는 것을 주요 목표로 한다. 분석 과정은 크게 세단계로 나눌 수 있다. 무음 구간 분석(과정 1), 음원 분할 및 STT(Speech To Text)¹ 구간(과정 2), 스크립트 재구성(과정 3) 이다. 해당 과정은 Linux(Ubuntu 18.04)에서 구현되었다. 무음 구간 분석은 ffmpeg 와 sox 를 사용하였다. 또한 STT 는 Mozilla 재단의 오픈소스 DeepSpeech[1]를 활용하였다. 분석에 활용된 음원은 해커스에서 무료로 배포하고 있는 토익 LC 음원을 사용하였으며, 정답셋은 본 연구에서 정의한 토익 스크립트 구조로 재구성하여 만들었다.

A.과정 1. 무음 구간 분석

먼저 음원에서 특정 소리의 크기와 구간을 지정하여 무음 구간을 분석하고자 하였다. 예를 들어, -80dB 이하의 소리가 0.3 초 이상 지속될 경우, 해당 구간을 무음 구간이라고 정의하였다. 본 연구에서는 소리의 크기가 분석 결과에 큰 영향을 끼치는 요인으로 예상하였다. 이에 따라 다양한 설정값으로 시도하였으며, 이후 결과에서 정리하도록 한다.

B.과정 2. 음원 분할 및 STT

과정 1에서 분석된 무음 구간을 통해 과정 2에서는 실제로 음원을 분할하고 STT 하여 스크립트 결과를 얻었다. 해당 과정은 멀티프로세스 방식으로 구현되었다. 각 프로세스는 CPU bound 작업을 수행하며, 기존 단일 프로세스로 구현할 때보다 훨씬 좋은 성능을 보였다. 토익 음원 내 a, b, c, d 와 같은 선택지는 1 초 이하의 재생 시간을 가지므로 STT 결과가 제대로 출력되지 않아, 음원 양 끝에 1 초 정도의 패딩구간을 붙여서 사용하였다.

C.과정 3. 스크립트 재구성

해당 과정에서는 과정 2 의 STT 결과를 본 연구에서 정의한 토익 스크립트 구조로 매핑하는 작업을 하였다.

<예시 1. 매핑 전후 비교>

매핑 전	매핑 후
part one	PART 1
number twenty two	Number 22
a	(A)

해당 과정은 분기문으로 구현되었다.

2) 결과

분석엔진 결과의 정확도에 가장 주요한 영향을 끼친 요인으로 과정 1 의 소리의 크기를 예상하였다. 해당 요인에 대한 설정값과 정확도를 정리하면 다음과 같다.

<표 1. 소리의 크기에 따른 정확도 비교>

소리의 크기	정확도
-80dB	94.56%
-70dB	81.98%
-90dB	94.56%

위의 정확도는 저자가 정의한 토익 스크립트 구조로 정답셋을 만들고, 분석엔진의 결과와 단순비교를 통해 계산하였다.

표 1 에 나타나는 결과로 보아 특정 이상의 임계값을 넘으면 정확도는 더 이상 변하지 않는다.

-70dB 에서는 문장이 제대로 분할되지 않아 정확도가 떨어지는 문제가 발생했다.

현재 STT 기술은 다양한 언어(미국식 영어, 영국식 영어, 호주식 영어 등등)이 섞여 있을 경우 결과가 완벽하지 않고, 사람 이름이나 회사 이름 등의 고유명사를 정확하게 분석할 수 없으므로 100 에 가까운 정확도를 내는데 한계가 있다.

본 연구에서 사용한 음원에서의 최적 크기는 -80dB 이지만 해당 값은 음원 품질에 따라 달라질 수 있다. 값은 음원의 최고/최저 소리 크기에 따라 조절될 수 있다.

¹ STT(Speech To Text) : 음성을 텍스트로 변환하는 기술.

이외에도 결과의 성능에 영향을 끼친 주요한 요인은 과정 2의 구현방식이다. 단일 프로세스 방식과 멀티 프로세스 방식으로 구현했을 때의 차이를 다음 표에서 보여준다.

참고문헌

[1] <https://github.com/mozilla/DeepSpeech>

<표 2. 구현 방식에 따른 성능 비교>

구현방식	수행 시간
단일 프로세스	41 분
멀티 프로세스	5 분

해당 비교에 사용된 음원은 45 분의 완전한 토익 LC 음원이다.

멀티 프로세스로 구현한 엔진의 수행시간은 서버 환경에 따라 충분히 달라질 수 있다. 그러나 표 2 결과에 미루어 보아 단일 프로세스보다 멀티 프로세스로 구현하였을 때에 확연한 성능의 차이가 있다.

3. 결론

본 논문에서는 언어 학습, 구체적으로는 토익 리스닝 파트에 대한 학습을 도와주는 방법을 파악하였다. 리스닝 공부에는 반복 듣기와 따라 말하기가 중요하다. 일반 음원 플레이어에서는 듣고 싶은 부분을 음원 상태바를 움직여 수동으로 찾아야 했기에 원하는 부분을 한 번에 찾기 힘들다. 그렇기 때문에 듣고 싶은 부분을 한 번에 찾아서 들을 수 있는 방법에 관해 연구한 것이다.

연구 결과는 무음 구간 분석과 STT 기술을 활용하여 분석엔진을 통한 애플리케이션을 제안한 것이다. 자세한 결과는 다음과 같다.

무음 구간을 정의하여 음원을 분할하고 STT 하역 스크립트 결과를 얻었다. 멀티 프로세스 방식으로 구현하였다. 각 프로세스는 CPU bound 작업을 수행하였다. 정의한 무음 구간을 기준으로 음원을 한 문장 구간 단위로 분할하고 STT 요청을 보낸다. 그 후 스크립트를 추출하여 애플리케이션에 보여준다.

지금으로서는 토익 리스닝에 중점을 두었지만 차후 다른 언어시험에 대해 데이터를 쌓고 연구를 진행한다면 다양한 언어 듣기시험에 대한 더 좋은 결과를 얻을 수 있다.