Journal of KIIT. Vol. 17, No. 10, pp. 121-127, Oct. 31, 2019. pISSN 1598-8619, eISSN 2093-7571 121
http://dx.doi.org/10.14801/jkiit.2019.17.10.121

중국어 성조학습을 위한 음성시각화 어플리케이션 개발

이선희*, 김건우**, 강준영***

Development of Speech Visualization Application for Chinese Tone Learning

Sun-Hee Lee*, Kun-Woo Kim**, and Jun-Young Kang***

이 논문 또는 저서는 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2018S1A5A8029882)

요 약

자기주도학습이 가능한 성조 연습 프로그램을 개발하려면 다음의 두 가지가 고려돼야 한다. 첫째는 언제 어디서든 민첩한 방식으로 피드백을 주는 것이고, 둘째는, 1차적으로 얻은 피드백을 타인과 공유하여 2차 피드백을 얻고 배움을 완성하는 것이다. 본 연구는 시각적으로 즉시 피드백을 받을 수 있는 어플리케이션(TONE VIEWER)를 개발하였다. 톤뷰어는 YIN알고리즘을 채택하여 모바일의 전산처리 속도로 인해 원어민음성 재생 시 운율 곡선에 변화가 생기는 문제점을 개선하였고, GMED를 활용하여 원어민과 학습자의 성조 곡선의 유사도를 도출한다. 톤뷰어는 원어민 음성 업로드, 학습자 발음 녹음, 두 음성의 음높이 유사도 측정 등의 기본 기능을 구현하였다.

Abstract

This study took into account the following functions when developing the application: first, it should provide feedback information to the user at anytime; second, the application should allow the user to share his/her information with other users so that the user can receive more interactive feedbacks from teachers or peers. We explain the different development stages of our tone learning application named “ToneViewer” and about the YIN algorithm, which was adopted to detect the pitch of both the Chinese native speakers and the non-native users. The application used GMED to compare the matrix grids of the pitch data between the native speaker and the non-native users. The application is equipped with basic functions such as uploading native speakers' voices, recording learners' pronunciation, and measuring the similarity between the native and non-native voices.

Keywords
Chinese tone learning, speech visualization application, YIN algorithm, pitch anaysis

* 사이버한국외국어대학교 중국어학부 교수 (교신저자)
- ORCID: https://orcid.org/0000-0002-4229-9739
** 스마트잭 대표
- ORCID: https://orcid.org/0000-0001-9249-0227
*** 서울시립대 컴퓨터과학과
- ORCID: https://orcid.org/0000-0001-5510-8121

· Received: Aug. 27, 2019, Revised: Oct. 17, 2019, Accepted: Oct. 20, 2019:
· Corresponding Author: Sun-Hee Lee
Imunro 107, Cyber Hankuk University of Foreign Studies, Dept. of Chinese Dongdaemun-gu, Seoul, Republic of Korea
Tel.: +82-02-2173-3479, Email: lishanxi@cufs.ac.kr

Ⅰ. 서 론

중국어를 이해할 때 성조가 차지하는 비중이 46%에 달하지만 배우는 과정에서는 쉽게 배울 수 있는 언어요소는 아니다[1]. 실제로 선행연구에 따르면 학습자들은 성조 학습의 어려움을 토로하는 경우가 많다. 외국인들의 경우 중국어 성조를 잘못 사용하게 될 가능성이 높다. 이 경우 전체적인 맥락에서 의미 변화가 생기므로 중국인이 듣기에는 당연히 부자연스럽고 어색하게 들릴 수밖에 없다. 그렇다면 더욱 효율적으로 학습자가 자신의 음높이를 파악하고 교정해 나가도록 하는 것은 더 나은 커뮤니케이션을 실현하는 데 도움이 될 것이다. 사람의 지능은 한 가지로 이루어져 있지 않고, 여러 가지가 종합적으로 연결되어 있으며[2], 인간의 인지 과정은 시각, 청각 등 다양한 감각을 동시에 활용하여 진행된다[3]. 따라서 본 연구는 말소리는 귀로 듣고 연습한다는 기존의 교육방식에서 벗어나 '성조 및 운율 발음 학습과 교정에 도움이 되는 시각화된 그래프를 제공'하는 어플리케이션을 개발해 보겠다.

기존에 개발된 프로그램 중 중국어 성조를 교육의 입장에서 시각화해서 볼 수 있는 프로그램은 일본 성계대학(成蹊大學의) 발음 연습 프로그램, 러시아인의 영어 억양피드백 프로그램 인톤트레이너(Intontrainer), NHK의 중국어 발음교정 프로그램 레벨업 차이니즈(Level up Chinese) 등이 있다. 그러나 이 세 가지 프로그램은 모두 다음의 문제점을 가지고 있어 유비쿼터스 시대에 자기주도적으로 성조연습을 하는 교구재 역할을 하기에는 적합하지 않다. 첫째, 프로그램 자체에 설정되어있는 음원만 연습할 수 있어 자기주도 학습이 불가하다. 둘째, PC기반의 프로그램으로 휴대에 어려움이 있다. 따라서 본고는 모바일에서 구동이 되어 언제 어디서나 사용할 수 있고, 학습자 스스로 연습하고 싶은 음원을 올려 자기주도적으로 학습할 수 있는 어플리케이션을 개발해보겠다.

본 논문의 구성을 다음과 같다. 2장은 기존에 개발된 음성시각화 프로그램 중 학습자 중심의 프로그램을 간략히 소개한다. 3장에서는 본 논문에서 개발하는 TONE VIEWER 이전 프로토타입의 문제점과 이 문제점 개선을 위해 채택한 알고리즘과 유사도 측정 방식에 대하여 논하고, TONE VIEWER 어플리케이션의 사용 방법에 대하여 기술한다. 마지막으로 4장에서는 결론에 대해 기술하고 본 연구가 갖는 한계점 및 향후 연구에 대하여 논한다.

Ⅱ. 성조 학습 관련 음성시각화 프로그램

2.1 성계대학(成蹊大學의) '遊'프로그램

일본 성계대학의 발음 연습 프로그램은 법학부에 홈페이지에 게시된 '遊'이다. 이 프로그램은 홈페이지 자체에서 구동이 되며 그림 1과 같이 디자인되어 있다.

학교 홈페이지에 미리 탑재된 중국어 단어들의 운율곡선을 볼 수 있고, 컴퓨터에 마이크를 연결해 자신의 발음을 녹음한 후 재생하면 원어민의 곡선과 자신의 발음 곡선을 대조해 볼 수 있다. 비록 제공되는 음원만 가지고 연습해야 하지만, 일본의 중국어 교육계에서 인정하는 교육어휘를 약 3천 개 정도 수록하고 있어 어휘는 다양한 연습이 가능하다. 그러나 문장의 경우 제공되지 않으며, 2009년에 개발이 된 이후 업데이트되지 않아 현재 제대로 구동되지 않는다는 문제점이 있다.

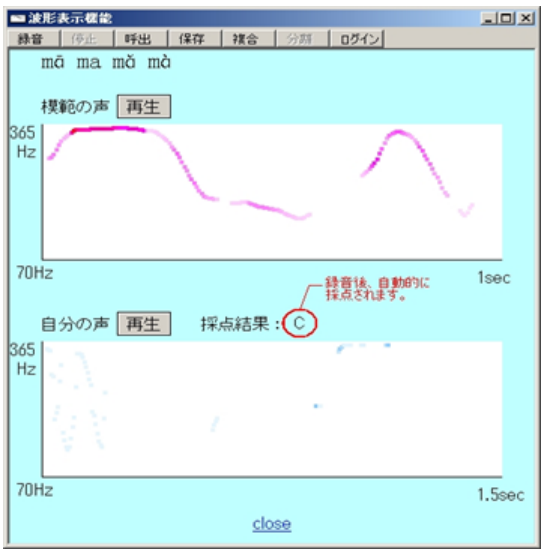


그림 1. 성계대학(成蹊大學의) 발음 연습 프로그램
Fig. 1. YOU program

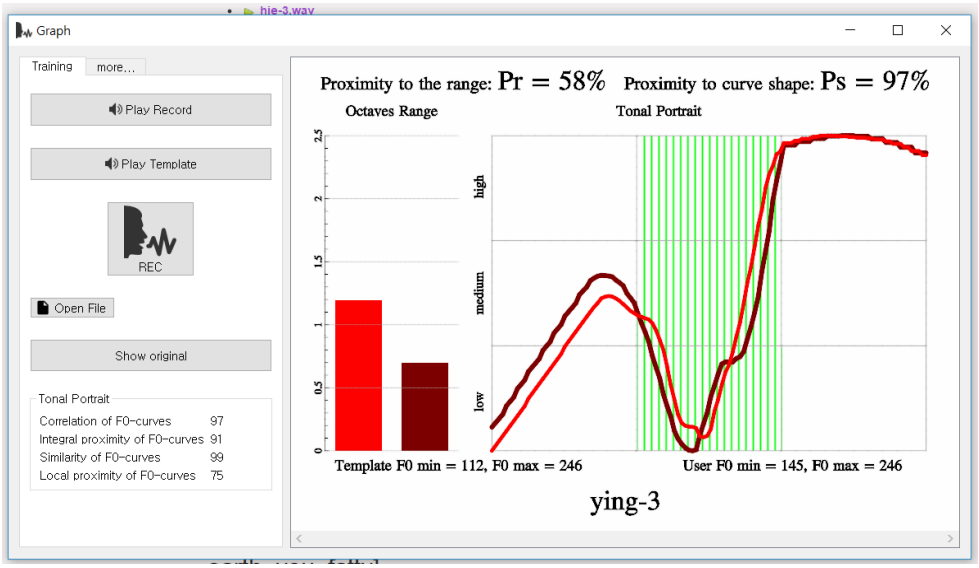


그림 2. 인톤트레이너
Fig. 2. Intontrainer

2.2 인톤트레이너

인톤트레이너(그림 2)는 음성분석을 전공한 러시아 학자 Boris Lobanov가 기안한 억양 모니터링 프로그램이다. 이 프로그램은 그림 2와 같이 실행된다. 밝은 빨간색이 학습자의 발음으로 원어민의 발음과 유사도를 측정해서 수치로 제시해 준다. 현재 프로토타입 개발 상태로 더 개발이 진행되지 않는 점이 아쉽다. 또한, 프로그램 자체에 제공되는 30개 정도의 표현 이외에는 더 연습할 수 없는 한계가 있다.

2.3 NHK 레벨업 차이니즈

NHK의 레벨업 차이니즈의 경우 앞에서 소개한 두 프로그램보다는 플랫폼 디자인은 훨씬 사용자 친화적이다. NHK의 레벨업 차이니즈는 학습자가 연습하고자 하는 발음의 운율 곡선이 그림 3과 같이 제시된다. 제시되고 있는 발음이 어떤 문장인지 텍스트를 통해 간체자 표현과 한어 병음이 모두 제시되므로 학습자는 자신이 연습하는 내용이 무엇인지 더욱 정확히 알 수 있다. 그리고 그 텍스트에 해당하는 발음의 성조 곡선이 원어민 것은 보라색, 학습자 것은 노란색으로 제시되어 연습한 이후 시각적인 모니터링이 가능하다.

그림 3. NHK Level-Up Chinese 실행 화면
Fig. 3. NHK level-up Chinese

그러나 이 프로그램 역시 NHK에서 방송하고 있는 내용만을 연습할 수 있도록 제공하는 한계를 지닌다. 즉, 학습자는 NHK에서 제공하는 음성과 텍스트만을 보고 연습할 수 있어 학습 내용 선택의 폭이 좁다. 이 밖에 세 개의 프로그램 모두 PC 기반의 프로그램으로 장소에 구애받지 않고 연습하기는 어렵다.

III. ToneViewer 개발

3.1 개발 알고리즘

본 연구에서 개발하는 어플리케이션은 성조를 눈으로 볼 수 있다는 의미로 'TONE VIEWER'라고 명

명하였다. 연구 초기에 개발한 프로토타입의 경우 학습자가 발음 연습을 하는 공간이 반드시 조용한 공간은 아니라는 점을 간과하여 소음이 없는 방에서 연습하는 경우 운율 그래프가 안정적으로 나오지만, 주변 소음이 있는 경우에는 녹음 후 플레이했을 때 적절해 보이지 않는 그래프가 제시되기도 하였다.

또, 최초 업로딩한 음성을 시각적으로 보도록 실행하면 처음에는 그림 4의 좌측과 같이 운율곡선이 정상적으로 제시되나 두 번 혹은 세 번 연거푸 같은 실행을 하도록 하면, 3회 이후에는 그림 4의 우측과 같이 적절해 보이지 않는 운율곡선이 제시되기도 하였다[4].

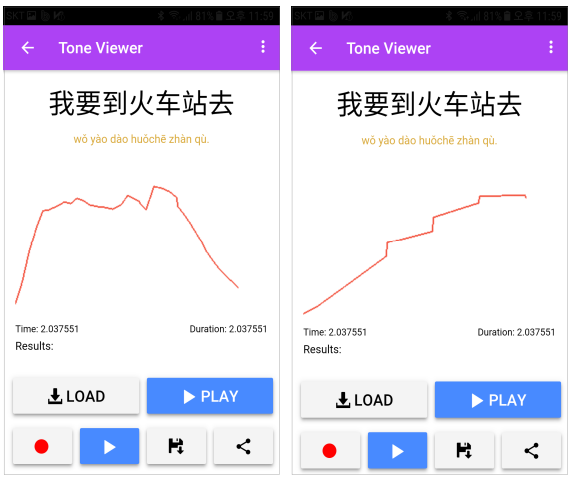


그림 4. 프로토타입 실행 화면
Fig. 4. ToneViewer Prototype

3.2 유사도 측정

우리는 현재 개발에 사용한 알고리즘이 너무 많은 양의 주파수를 동시에 계산해 내려고 하여 전산속도가 나오지 않는 상황을 개선하기 위하여 다른 접근법을 모색하였다. 즉, 성조 학습을 위한 어플리케이션 개발을 위해 인간의 음성을 단선율 멜로디라고 고려하고, 이 단선율 멜로디의 기본주파수를 수집하는 MIR(Music Information Retrieval) 방식[5]을 어플리케이션에 도입해 보았다. 음악을 검색하는 방식 중 기존 연구에 따르면 오차율이 2% 이하로 낮은 알고리즘으로 YIN 알고리즘[6]이 있다.

피치(Pitch)는 음의 주파수가 높고 낮음을 뜻하는 것으로 사람의 청각 인지에 따른 용어이다. 성조는 인지적으로는 피치이지만 물리학적으로는 결국 음의 기본주파수(F0)로 표현이 될 수 있다. 따라서 발화 시간 순서에 따른 F0 값을 얻어내면, 발화자의 음높이 곡선을 그릴 수 있다. 우리는 Alain이 제시한 YIN알고리즘[7]에 따라 먼저 AMDF(Average Magnitude Difference Function)방식으로 식 (1)과 같이 목소리를 단선율로 가정해 피치탐색을 수행하였다. 이 경우 후행 처리가 없어 피치 확보의 오류율이 3배 이상 개선된다.

목소리(멜로디)

$$S_m(t) = F_m(t) \times A_m(t) \tag{1}$$

또한 피치 값과 해당 피치가 측정된 시간을 각각 f와 x로 나타낸 뒤, 식 (2)처럼 보간법을 사용하여 빠르게 피치 값을 확보하도록 하였다.

$$f(x) = f(0) + \frac{(f(1) - f(0)) \times (x - x_0)}{(x_1 - x_0)} \tag{2}$$

다음으로 YIM 알고리즘을 통해 검출된 원어민과 학습자의 피치 데이터 간의 유사도를 측정하기 위하여, 우리는 음고를 나타내는 곡률을 매트릭스 위에 표현하고 공간상에서 이를 대조하였다.

즉, 추출한 정보를 Grid-based Matrix상에 표현하고 이를 공간 대조하는 방식을 고려하였고, 이를 위해 진폭과 시간의 왜곡에 견고한 GMED(Grid Matrix Euclidean Distance)와 GMDTW(Grid Matrix Dynamic Time Warping)[8] 방식을 활용했다. 시계열 데이터 사이의 유사도 비교에는 STS3(Set-based similarity)[9] 알고리즘을 사용하였다. STS3는 그리드를 이용해 시계열 데이터를 집합으로 표현하고, 집합으로 표현된 시계열을 기반으로 유사성을 측정하는 알고리즘이다. 그리드가 주어지면 시계열을 자신이 지나는 그리드 셀 인덱스의 집합으로 변환한다.

격자형 매트릭스에 운율 데이터를 제시하고 점유한 공간을 1, 점유하지 않은 공간은 0의 정보를 갖는 행렬로 표시하게 되면, 행렬의 0이 아닌 요소만 비교가 가능해 빠른 전산처리에 용이하다.

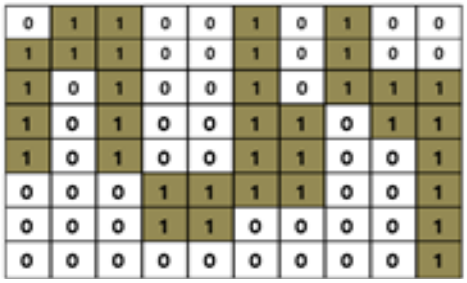


그림 5. Grid-based matrix형태 시계열
Fig. 5. Grid-based matrix time series

주파수로 변환된 음성 데이터의 경우 화자에 따라 주 음역대가 다르다. 화자에 따른 음역대 차이를 고려하여 각 음성 데이터에 대해 집합으로 변환하기 전에 Min-max 정규화(Normalization)를 사전 처리(Pre-processing)한다.

$$X' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{3}$$

마지막으로 시계열 t_1과 t_2의 각 집합표현 $S(t_1)$과 $S(t_2)$ 사이의 자카르드 유사도를 다음과 같이 정의하였다.

$$\begin{aligned} sim(t_1,t_2) &= JS(S(t_1),S(t_2)) \\ &= \left|\frac{(S(t_1) \cap S(t_2)}{S(t_1) \cup S(t_2)}\right| \end{aligned} \tag{4}$$

3.3 ToneViewer 실행 인터페이스 구성

3.2절에서 구상한 알고리즘으로 우리는 안드로이드 어플리케이션 'TONE VIEWER'를 개발하였다. 본 연구가 개발한 어플리케이션은 음성의 피치를 분석하여 유사도를 측정하는 기본 기능을 구현한다.

먼저 어플리케이션 아이콘을 클릭하면 처음 열리는 인터페이스에 중국어의 가장 성조 1성, 2성, 3성, 4성으로 녹음된 'ma' 음성 파일이 탑재되어 있다.

그림 6에서와 같이 기본 음성파일을 제공한 이유는 처음에 음성을 업로드하기 어려운 사람도 본 프로그램이 어떤 형태로 진행을 하는지 알 수 있도록 하기 위해서이다. 편의를 위해 일단 프로그램에서 제공하고 있는 기본 음성을 선택하면 그림 7의 상단과 같은 형태의 원어민의 피치 그래프 곡률이 나타난다.

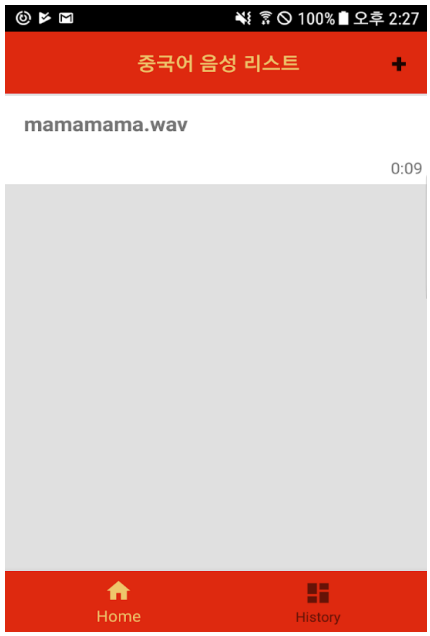


그림 6. 실행 첫 화면
Fig. 6. Starting screen of ToneViewer

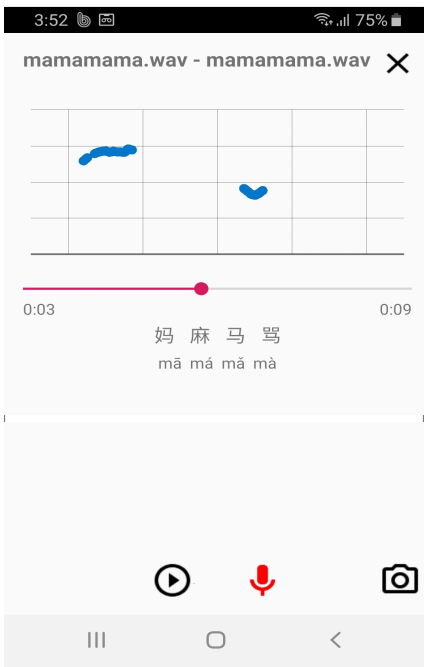


그림 7. 음성파일 실행화면
Fig. 7. Screen after playing wav file

학습자가 업로드하는 파일의 명칭이 화면 위에 그대로 표기되도록 하여 제공되는 음성파일에 자막이 함께 제시되지 않더라도 파일명을 통해 학습자가 자신이 연습하는 발음이 무엇인지 알 수 있도록 했다.

그림 7의 하단에 있는 빨간색 마이크 아이콘을 누르면 그림 8의 왼쪽과 같은 화면으로 전환되며, 다시 하얀 마이크 아이콘을 클릭하면 '1, 2, 3 시작' 이라는 멘트가 떠서 학습자가 녹음 전에 인지적으로 준비할 수 있도록 한다. 녹음을 진행하는 화면은 그림 8의 오른쪽과 같으며, 원어민의 발화는 하늘색 곡선 학습자의 발화는 노란색 곡선으로 표시된다.

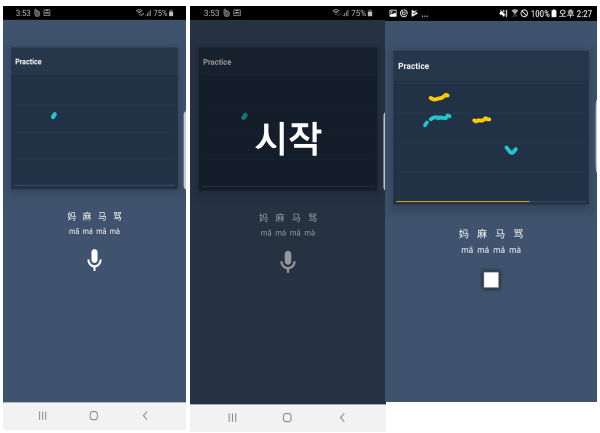


그림 8. 녹음화면
Fig. 8. Screen of recording session

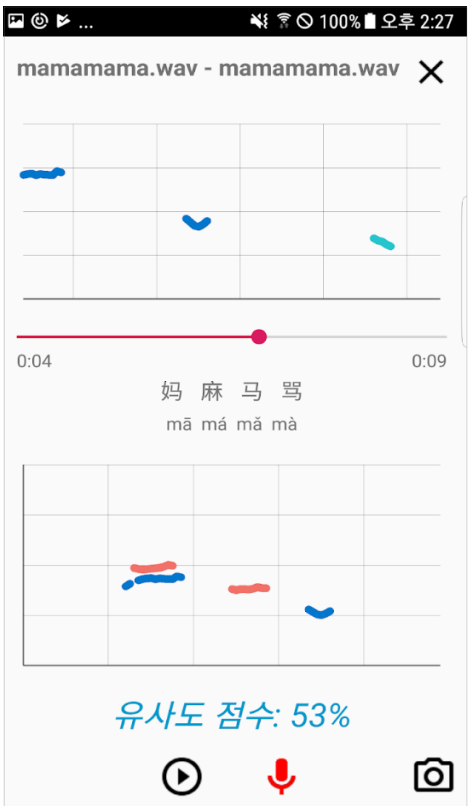


그림 9. 녹음 후 유사도 대조화면
Fig. 9. Screen of similarity comparison result

녹음이 완료된 이후에는 다시 그림 9와 같은 화면으로 전환되며 대조된 유사도 점수가 제시된다. 어플리케이션을 설치하면 자동적으로 [ToneViewer] 라는 폴더가 생성되고, 오른쪽 하단의 카메라 아이콘을 클릭하면 화면이 캡쳐되어 생성된 폴더에 자동저장된다. 파일명칭은 화면을 스크랩한 순간의 일자와 시간으로 지정이 되어, 이후에 언제라도 학습자가 연습한 순서대로 시각데이터를 정렬할 수 있도록 하였다.

Ⅳ. 결론 및 향후 과제

본 연구는 한국인 학습자가 어려워하는 성조학습을 개선하기 위해서 첫째, 언제 어디서든 즉각적이고 민첩한 방식으로 피드백을 주어야 하고, 둘째 1차적으로 얻은 피드백을 타인과 공유하여 2차 피드백을 얻고 배움을 완성하게 해야 한다고 판단했다. 그리고 이 두 가지를 기초적으로라도 실현하기 위하여 휴대가 간편한 모바일에서 구동하는 음성시각화 어플리케이션을 개발했다.

TONE VIEWER 어플리케이션은 원어민 음성 업로드, 학습자 발음 녹음, 두 음성의 음높이 유사도 측정 등의 기본 기능을 구현하고 있다. 본 어플리케이션은 2019년도 중국어문연구회 국제학술대회 참가자들에게 소개되고 현장에서 직접 다운로드 받아 중국어 전문가들이 간단하게 검토를 해 보았다. 당시 학습자가 자신이 학습하고 싶은 음성 파일을 업로드할 수 있다는 점이 가장 좋은 점으로 평가받았다. 그러나 현재 유사도 측정의 정밀도가 높지 않은 점은 단점으로 지적되었다. 특히, 당시 녹음을 하지 않아도 유사도가 50%로 나오기도 하였는데, 이는 학회장이 비교적 소음이 많이 있는 공개된 대형 강의실이어서 주변 소음 역시 데이터로 잡혀 발생한 현상이다. 유사도의 정밀도가 낮으면 당연히 피드백에 대한 신뢰도가 떨어질 수 밖에 없고, 이를 학습에 바로 활용하기는 어렵다.

따라서 학습에 필요한 피드백 제공하고, 실제 학습 모니터링 효과를 가지는 교구재로 활용하기 위해서는 향후 매트릭스를 더욱 세분화하여 유사도의 정확성을 확보해야 한다. 매트릭스 해상도를 높였을 때 주변 소음 역시 더 잘 잡게 되는데, 따라서 소음

의 채집을 막을 수 있는 노이즈 필터링이 추가로 진행되어야 한다. 또한 모든 음성파일이 자막을 가질 수 없다는 단점을 극복하기 위하여 발화데이터를 수집해 음성인식을 통한 강제정렬(Forced alignment) 자동 전사 방식을 채택해야 할 것이다.

References

[1] Kwon, Young-sil, "The necessity of HL Tone teaching for effective communication of Mandarin Chinese", Chinese Language Education and Research, Vol. 13, pp. 83-94, Nov. 2011.

[2] Gardner, H. "Frames of mind: The theory of multiple intelligences", New York: Basic Books. pp. 3-13, 1983.

[3] Richard Mayer, "Applying the Science of Learning", American Psychologist, Vol. 63, No. 8, pp. 760-769, Dec. 2008. https://result.uit.no/basiskompetanse/wp-content/uploads/sites/29/2016/07/Mayer.pdf

[4] Sunhee Lee, "Study on the development speech visualization application", Journal of Chinese Humanities, Vol. 71, pp. 419-436, Apr. 2019.

[5] Graham E. Poliner, Daniel P.W. Ellis, Andreas F. Ehmann, Emilia Gomez, Sebastian Streich, and Beesuan Ong, "Melody Transcription From Music Audio: Approaches and Evaluation", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 4, 1247-1256, May 2007.

[6] Sang-un Gum and Ju-han Nam, "How to retrieve Music information with using Melody extracting Algorithm", Magazine of the IEEK, Vol. 43, No. 5, pp. 41-49, May 2016.

[7] Alain de Cheveigné and Hideki Kawahara, "YIN, a fundamental frequency estimator for speech and music", Journal of Acoustic Society of America, Vol. 111, No. 4, pp. 1917-1947, Apr. 2002. http://recherche.ircam.fr/equipes/pcm/cheveign/ps/2002_JASA_YIN_proof.pdf.

[8] Yanqing YeEmail author, Jiang Jiang, Bingfeng Ge, Yajie Dou, and Kewei Yang, "Similarity measures for time series data classification using grid representation and matrix distance", Knowledge and Information Systems, Vol. 60, No. 2, pp. 1105-1134, Aug. 2019.

[9] Jinglin Peng, Hongzhi Wang, Jianzhong Li, and Hong Gao, "Set-based similarity search for time series", In Proceedings of the 2016 International Conference on Management of Data, SIGMOD '16, ACM, San Francisco, California, USA, pp. 2039-2052, Jun. 2016.

저자소개

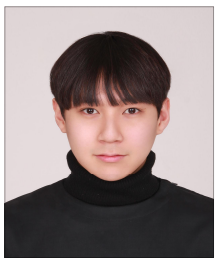
이 선 희 (Sun-Hee Lee)

2001년 2월 : 고려대학교 중어중문학과(문학사)
2005년 2월 : 한국외국어대학교 통번역대학원 한중과 (통번역학석사)
2010년 10월 : 북경어언대학교 인문학원(문학박사)
2019년 10월 현재 : 사이버한국외국어대학교 중국어학부장, 미래교육연구소장
관심분야 : 중국어교육, 음성분석, 음성인식

김 건 우 (Kun-Woo Kim)

2006년 2월 : 고려대학교 화공생명공학과(공학사)
2005년 12월 ~ 2017년 6월 : 삼성전자/디스플레이 상품기획(책임연구원)
2017년 7월 ~ 현재 : (주)스마트잭 대표이사
관심분야 : IT, S/W개발, Material Infomatics

강 준 영 (Jun-Young Kang)



2017년 2월 : 서울시립대학교 컴퓨터과학부(공학사)
2019년 2월 : 서울시립대학교 컴퓨터과학과(공학석사)
2019년 10월 현재 : (주)스마트잭 엔지니어
관심분야 : 데이터마이닝, 시계열 분석, 블록체인