



Fostering learner autonomy in English pronunciation: A convergent, interdisciplinary phonetics approach*

Minjung Son**

Department of Applied English Content, Hannam University, Daejeon, Korea

Abstract

This study adopts a technology-supported, phonetics-based instructional approach to enhance learners' English pronunciation. The course was designed as an integrated curriculum that combines phonetics, phonology, intonation (ToBI), pronunciation training, and pronunciation assessment within a unified framework. Learners engaged in systematic pronunciation practice, conducted acoustic analyses using Praat, recorded speech samples, and participated in two rounds of multiple-forced-choice pronunciation assessment experiments. In addition, they implemented AI-based automatic speech recognition (ASR) using Python code and evaluated their pronunciation through the Word Error Rate (WER). This course was developed from a convergent, integrative, and experimental perspective and intended to promote learner autonomy. Survey responses indicated that approximately 70% of learners perceived overall improvement in their pronunciation skills. A mixed-effects model analysis revealed that short-term exposure to automated ASR-WER feedback was insufficient to systematically alter participants' rating behaviors and had only limited influence on their sensitivity to specific evaluation dimensions.

Keywords: phonetics, phonology, ToBI, multiple forced choice, rating English pronunciation, ASR, WER, Python

1. 서론

두 차례에 걸친 세계 대전을 거치며 영어는 국제적인 환경에서 의사소통에 있어 다양한 분야에서 공용어가 되었다. 견고하고 높은 위상을 가진 영어는 교육학적 측면에서 수많은 영어학습자와 교수자들의 주목을 받아왔다. 특히 1970-1980년대에 외국어 교육 분야에 주류가 되었던 의사소통적 접근법(commun-icative language teaching, CLT)은 학습자 중심 교수법으로, 실생활에서 의

사소통을 위한 목표 언어 사용을 중요시하였다(Candlin & Mercer, 2007; Nunan, 1991). 따라서 제2외국어 습득에서 영어 어휘 및 영문법 습득, 그리고 영어학습 자료 개발 및 학습 현장 활용 분야는 주목받고 활발히 연구되었으나 영어 발음은 부차적인 문제로 간주되며 크게 주목받지 못하였다. 제2외국어로서 영어 사용자들은 그 정도의 차이가 있을 뿐 모국어의 악센트가 영어 발화에 남아 있기 마련이고, 의사소통을 위해서는 어휘, 영문법, 영어학습 자료의 효과적 활용이 더 중요했기 때문이다. 따라

*This research was supported by the Center for Teaching&Learning&Distance Education at Hannam University.

**minjungson@hnu.kr, Corresponding author

Received 14 November 2025; Revised 10 December 2025; Accepted 10 December 2025

© Copyright 2025 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서 영어 발음 연구는 화자-청자 사이의 이해 가능성(intelligibility), 기능적 전달 가능성(functional communicability), 그리고 학습자 본인의 발음 점검(self-monitoring)을 통한 자가 발음교정 등에 집중되었고(Brown, 1991; Morley, 1991), 의사소통을 저해하지 않는 수준의 명료한 발음이면 충분하다고 간주되었다.

그럼에도 여전히 유창한 영어 발음은 의사소통을 더욱 원활히 하고 화자의 발화 자신감을 높일 수 있는 수단이라는 견해가 존재한다. 분절음 발화 정확도와 제2외국어 학습자의 모국어 악센트 사이의 인지적 관련성에 관하여 꾸준히 연구되었고(Levi et al., 2007; Riney et al., 2000; Tsukada et al., 2005), 목표 언어와 모국어의 음소 체계의 차이로 인한 제2외국어 음소 종류에 따른 습득의 난이도 차이도 활발히 연구되었다(Best, 1990; Best et al., 2001). 음성학·음운론 기반의 접근을 통해 분절음 지식뿐 아니라 강세, 리듬, 억양과 같은 초분절음 지식의 습득을 발음 교육에 통합한 연구들 또한 꾸준히 보고되고 있다(Celce-Murcia et al., 2010; Kim, 2018; Kim, 2020, 2024; Kim & Chung, 2023; Sung, 2010). Kim(2018)은 제2외국어로서의 영어학습자와 영어원어민 화자의 발화 자료를 비교하며 분절음과 초분절음에 관한 발음 숙련도를 평가하였다. 영어원어민을 포함한 전문 평가자 4인이 상위 학습자로 분류한 8명에 한정하여 후속으로 음성 분석[문장 기본주파수(f0), 강세/비강세 모음에 나타난 기본주파수(f0), 모음 길이(length) 및 모음 강도(intensity)]를 한 결과 상위 학습자들과 영어원어민 간에 문장 기본주파수는 차이는 없었으나 모음 길이 차이로 인한 강세/비강세 모음의 차이는 관찰되었다. Kim(2020)은 언어학적 요인을 명시적으로 설명하며 영어 발음 교육을 하고 발음평가를 했다. 구체적으로 Dale & Poms(2005)의 문장을 활용하여 녹음한 상급 학습자 음성 자료는 3가지의 언어학적[음성적 요인(모음: normalized perceived vocalic index, nPVI), 음운론적 요인(억양 고저 음절: prominent tonic syllable), 어휘적 요인(기능어 축약: reduction in function words)] 차원에서 정밀하게 분석되었다. 그 결과, 학습자의 발화 자연성은 어휘·음향·음성적 지표들이 상호작용함으로써 형성되는 것으로 나타났다. 따라서 이러한 언어학적 요인들은 학습자의 발음 능력을 체계적으로 향상시키기 위한 교육적 지표로 활용될 가능성을 보여준다.

명시적 영어 발음 교육의 긍정적 측면은 교육과정을 섬세하게 개선하는 것보다 무관하지 않다. Kim(2024)은 음성학 및 음운론에 기초하여 한 학기 동안 영어 발음법을 교육한 후 설문지 조사를 시행한 결과 학습자의 발음은 개선되었고, 자신감이 높아졌으며, 발음 지도법에 대한 전문성 확보와 같은 긍정적 효과를 발견하였다. Derwing & Munro(2005)에서도 유사한 결과가 관찰되었는데, 명시적 발음 교육은 의사소통 능력과 자신감에 긍정적 효과를 가져왔다. 이들 연구는 명시적 발음 지도법을 사범대학 영어과 교육과정에 필수적으로 편성할 필요성을 제기하며, 그 교육적 효과와 예비 교사 양성의 질적 향상을 뒷받침하는 근거를 제시한다.

한편, 영어 발음 교육의 효과성을 논의할 때는 발음 자체의 특성뿐 아니라 이를 어떻게 평가하는가 또한 중요한 고려 요소가 된다. 전통적으로 제2외국어로서의 영어학습자 발음평가는 전문

평가자 측정에 의존해 왔는데(Kim & Chung, 2023; Sung, 2010), 평가 기준 특정 및 신뢰성, 전문 평가자 간의 평가 변이성 등과 같은 문제점이 공존해 왔다(Kim, 2023). 최근에는 인공지능 기술을 기반으로 한 발음 평가 도구의 실질적 활용 방안 등에 관한 연구가 적극적으로 진행되고 있고, 이들 도구가 학습 현장에 도입되어 사용되기 시작했다(Ahn & Nam, 2024; Hong & Nam, 2021; Kim, 2023; Liakin et al., 2015).

기술 기반 애플리케이션의 가장 큰 장점은 시간과 장소의 제약 없이 전문 평가자의 도움 없이도 발음평가가 가능하다는 점에 있다. 특히 자동 평가 도구를 발음 훈련에 활용하면 학습자 개인의 습득 속도에 맞춘 학습이 가능해져 자기주도 학습을 촉진하며, 반복 연습과 즉각적인 피드백을 통해 발음 숙련도를 효율적으로 높일 수 있다는 장점이 있다(Kim, 2023). 이러한 즉시적 피드백은 학습자의 발음 향상 과정에도 긍정적으로 기여하는 것으로 보고되었다(Liakin et al., 2015). 그러나 인공지능 기반 자동 음성인식 도구가 보편적으로 활용되기 위해서는 기계 평가 결과가 전문 평가자에 근접할 만큼의 신뢰성을 확보하는 것이 필수적이다. Ahn & Nam(2024)은 다중모드대형언어모델(multimodal large language models)을 활용하여 한국인의 영어 발화 데이터를 분석하고, 전문 평가자와 기계 평가 간의 피어슨 상관계수(Pearson correlation coefficient)를 산출하여 자동 발음평가기의 신뢰도를 정량적으로 검증하였다. 그 결과, 운율 유창성의 경우 전문 평가자와 높은 상관관계를 보였으나, 조음 정확성에서는 상대적으로 상관관계가 약하게 나타났으며, 이는 두 평가 주체 간 평가 기준의 차이에 기인한 것으로 해석하였다.

Kim(2023)은 자동 발음평가기의 활용 가능성을 보다 실용적 관점에서 평가하였다. 학기 초 학습자 발화는 유창도와 이해도에 대해, 학기 말 학습자 발화는 명료도와 정확도에 대해 전문 평가자 4인이 평가한 결과, 네 평가 영역 간에는 완전한 일치가 나타나지 않았다. 그와 동시에 자동음성인식기 Whisper를 사용해 동일 음원에 대한 단어오류율(word error rate, WER)을 산출한 결과, 단어오류율은 전문 평가자의 네 평가 차원과 모두 부적 상관을 보였다. Whisper 기반 평가와 전문가 평가가 모든 차원에서 일치하지는 않았으나, 단어오류율은 학습자의 자기평가 및 자기주도 학습을 지원하는 지표로 활용할 수 있는 잠재성 있기에, 자동 발음평가기가 전문 평가자의 역할을 대체할 도구가 될 수 있다고 하였다. Hong & Nam(2021)의 연구 또한 SpeechPro와 전문 평가자의 비교를 통해 유사한 결과를 보고하며 자동 평가 도구의 교육적 활용 가능성을 시사하였다.

본 연구에서는 학습자는 교실 내 학습에서 영어 발음과 관련된 분절음과 초분절음에 대한 기초이론을 습득하고, 음향 분석을 시행하며, 발음 연습을 한다. 영어 발음평가를 위해서 다중강제 선택(multiple forced choice, MFC) 발음평가 실험, 파이썬(Python) 코드로 인공지능 기반 자동음성인식(automatic speech recognition, ASR)과 단어오류율을 결합하여 사용한다. 일련의 교육과정을 거쳐 학습자가 얻은 교육 효과를 설문조사와 발음평가 실험을 통해 다음 두 가지 연구 질문에 대하여 체계적으로 조사해 본다.

첫째, 영어 발음학습 효능감을 제고하기 위하여 고안한 본 수업 설계가 학습자에게 어떻게 인식되었는지 알아본다.

둘째, 학습자 영어 발음평가에 있어서 AI 기반 도구를 활용한 평가 경험이 어떤 교육적 가능성을 함의하는지 탐색한다.

2. 연구 방법

2.1. 수업 구성

본 연구는 대전광역시 소재 한 대학교에서 2024학년도 2 학기에 개설한 음성학 수업의 성과에 기반하였다. 음성 분석, 발화 녹음, MFC 발음평가 실험, ASR·WER 발음평가 활동, 설문 조사 참여 등 모든 수업 관련 활동은 강의 학점으로 인정(course credit)되어 성적에 반영되었다. 수강 인원은 총 10명(여성 8명과 남성 2명)이었다. 학습자들은 20대 초반 학부생들로서 기본적으로 영어 실력 향상에 관심이 있고, 이전 3학기 동안 다양한 실용 영어 분야 및 자연어처리를 위한 파이썬 언어를 학습하였다. 해당 수업은 인공지능 기반 자동음성인식 애플리케이션이 시중에 다양하게 출시된 점에 착안하여 학습자들이 최첨단 인공지능 기술과 그 기술을 뒷받침하고 있는 음성학 관련 기초이론을 이해하도록 도왔고, 실제 영어 발음 훈련 또한 가능하도록 실용적 관점에서 접근하였으며, 이론-실습-실험 순서로 설계하여 학습 난이도를 조절하였다.

총 15주 수업에서, 학기 초·중반까지 학습 활동은 다양한 이론 습득, 음성 분석, 녹음 순서로 진행하였다. 구체적으로 조음 음성학, 음향음성학, 조음음운론 이론에 기초를 두고, 음성 분석 프로그램인 프랏(Praat 6.4.21 win-intel64) 소프트웨어로 실습하였다(Boersma & Weenink, 2024). 학습자들의 영어 발음을 녹음한 자료는 다중강제선택 발음평가 실험과 ASR·WER 발음평가 활동에서 공통으로 사용되었다. 학기 중반 이후부터는 인공지능 신경망으로서의 인공지능 작동 원리, 디지털 신호처리, 음성으로 입력하여 텍스트로 출력(speech-to-text, STT)하고 텍스트로 입력하여 음성으로 출력(text-to-speech, TTS)하는 과정을 수업에서 다루었다. 파이썬 코드로 인공지능 기반 발음 평가 도구(ASR¹과 WER 측정 조합)를 직접 실행할 때는 학습자 본인의 영어 발음에 한정하여 평가하였다. 구체적으로 학습자가 자기 주도적 학습법을 사용하여 인공지능 기반 영어 발음 평가 도구를 활용해 본인의 발화 음원을 입력 데이터로 사용하고, 정량적 피드백을 얻으며 영어 발음을 향상하도록 하였다.

2.1.1. 주차 수업 계획

학습자들은 총 15주 수업에서 처음 3주간은 영어 분절음의 조음적 특성에 대한 설명을 들었고, 조음 협력 측면에서 음운변

화를 적용한 발화 연습을 하였다. 4-5주차에는 프랏 실습과 음성 분석 역량을 키웠다. 구체적으로 영어 원어민 화자 발음과 영어를 제2외국어로 습득한 학습자 본인의 발음 차이를 프랏을 사용하여 화자 그룹 간 음운적 특징에 대하여 음성 분석법을 배웠다. 이어서 6주차에는 영어 ToBI 억양이론과 프랏 텍스트그리드에 음조와 경계 지수를 라벨링 하는 방법을 배웠다(Silverman et al., 1992). 이어서 MIT OpenCourseWare에 연습 자료로 제공된, ToBI 라벨이 주석으로 첨가된 원어민 화자의 발화 음원을 청취한 후 이를 따라 발화하는 연습을 수행하였다(Veillex et al., 2006). 7주차에는 학습자들이 개별로 “Comma got a cure”² 구문에 나온 21개의 문장을 방송시설이 된 스튜디오에서 녹음하였다. 이때 녹음은 교수자가 하였다. 8주차에는 각 학습자가 본인 발화를 녹음한 음원을 수업 시간 및 방과 후 과제 시간을 사용하여 문장 단위로 분절(segmentation)하여 .wav 파일로 저장하였고, 텍스트그리드 층위를 동반 생성하여 문장 주석을 달았다.

9-10주차에는 음성신호 처리 및 인공지능경망으로서의 인공지능 구동 원리, 인공지능 스피커에서 소리를 텍스트로 변환시키는 것과 텍스트를 소리로 변환시키는 방법에 대하여 배웠다. 11주차에는 인공지능 기반 발음평가 도구를 사용하기 위하여, 학습자들은 해당 작업 수행하며 클라우드 기반 주피터 노트북 코랩(Colab)에서 파이썬 코드 실습을 하였고, 자력으로 인공지능 발음평가 도구(자동음성인식과 연계된 단어오류율 측정 파이썬 코드 실행)를 사용할 만큼 익숙해졌다.

이후 3개의 영어 발음평가 실험은 12-14주차에 걸쳐 순차적으로 실시되었다. 먼저 1차 다중강제선택 발음평가 실험을 12주차에 하였다. 다음으로 13주차에는 인공지능 기반 발음평가 도구를 사용하여 단어오류율을 측정하였는데, 이때 7주차에 녹음된 학습자 본인 발화(녹음 파일) 21개 문장을 입력 음원으로 사용하였다. 구체적으로 SpeechRecognition 라이브러리를 설치하여 입력된 음성을 텍스트로 변환하였고, 그 결과를 단어오류율 측정을 위해 가설 문장(hypothesis text)으로 사용하면서 학습자 본인 발화 평가에 활용하였다. 마지막으로 14주차에는 2차 다중강제선택 발음평가 실험을 하였다.

15주차에는 모든 학습자는 온라인 구글폼(Google Forms)을 사용하여 이론 시험을 쳤다. 이어서 온라인 구글폼 설문조사에 참여하였는데, 구체적으로 인공지능 기반 영어 발음평가 도구 사용을 포함하여, 조음음성학, 음향음성학, 조음음운론, ToBI 억양이론, 프랏을 활용한 음성의 물리적 측정법 습득 등이 본인의 영어 발음학습에 어떤 영향을 미쳤는지, 어떤 방법이 학습자 영어 발음 향상에 도움을 주었는지 등을 응답하며 학기를 마무리하였다. 교육은 강의 계획서에 따라 충실히 진행되었다.

1 구글사에서 제공하는 다중모드대형언어모델(Google MLLMs)은 하루에 일정 정도 무료로 음성 인식 서비스를 제공한다. 구체적으로 구글 코랩에서 파이썬 코드를 사용하여 SpeechRecognition 라이브러리를 설치한다. 여기에 속한 recognize_google(audio) 메서드는 인공지능과 클라우드에 기반한 Google Speech Recognition API(application programming interface)를 사용한다.

2 <https://research.ncl.ac.uk/necte2/documents/comma.pdf>

2.2. 발음평가 자료 수집

2.2.1. 발화 녹음

총 15주차 수업 중 7주차에 학습자 10명이 방음 스튜디오 녹음 세션에 참여하였다. 모든 화자는 녹음에 앞서 “comma got a cure” 단락이 인쇄된 자료를 받고, 대기실에서 10분간 모르는 단어를 찾아보고 미리 읽어보며 준비하는 시간을 가졌다. 녹음 세션에서 각 화자는 편안하게 의자에 똑바로 앉은 상태에서 움직임을 최소화하며 컴퓨터 모니터에 제시된 문장을 하나씩 보통 속도로 읽었다. 교수자는 2017년에 출시된 맥북에어(MacBook Air)를 사용하여 컴퓨터 모니터에 읽기 자료를 제시하였고 프랏(Praat 6.1.16)의 녹음 기능을 사용하였다(Boersma & Weenink, 2020). 맥북에어에 설치된 내장 마이크를 사용하여 샘플링 주파수는 44,100 Hz, 스테레오 채널로 학생 영어 발화를 녹음하였고, 화자는 제시된 단락의 21개 문장을 보통 속도로 차례대로 읽으며 1회차와 2회차 녹음을 연이어서 하였다. 피실험자가 문장을 읽고 녹음하는 동안 실험자의 개입은 없었다.

2.2.2. 다중강제선택 영어 발음평가 실험

12주차와 14주차에 프랏이 제공하는 다중강제선택 스크립트를 변형하여 영어 발음평가 때 사용하였다. 발음평가 실험의 음원은 각 화자의 2회차 녹음 자료에서 두 개 문장을 선택하여, 1차 다중강제선택 발음평가 실험에서는 “Finally, she administered ether.” 문장을, 2차 실험에서는 “Her efforts were not futile.” 문장을 대상으로 1단계(초급)-2-3-4(중급)-5-6-7단계(고급) 범위에서 평정하였다. 비전문 평가자인 학생들은 개인의 직관과 역량에 의존하여 자신의 영어 발음뿐 아니라 타인의 영어 발음도 함께 평가하였다. 그림 1은 다중강제선택 발음평가 실험에서 사용된 화자1의 발음 파형과 스펙트로그램 예시이다.

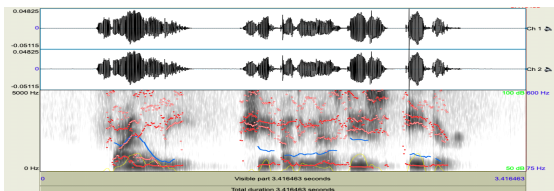


그림 1. 자극 예시(화자1) “Finally she administered ether.”
Figure 1. Example of sound stimuli from Subj1 “Finally, she administered ether.”

다중강제선택 발음평가 실험은 한 학기 동안 강의실로 사용

한 컴퓨터 실습실에서 단체로 실시하였다. Kim(2023)이 사용한 4개의 평가항목을 동일하게 적용하며 1차·2차 다중강제선택 발음평가 실험을 4개 세션[이해도(Comprehensibility)-유창도(Fluency)-정확도(Accuracy)-명료도(Intelligibility)]에 걸쳐 진행하였다.3 실험에 참여한 학습자는 각자 탁상용 컴퓨터에 Intel64 윈도우용 Praat 6.4.21를 직접 설치하였고(Boersma & Weenink, 2024), 유선 이어폰 착용 후 다중강제선택 발음평가 실험에 참여하였다. 각 세션 첫 화면에서 평가 영역을 한글 병기로 정확히 안내하였고, 해당 화면은 컴퓨터 키보드 ‘Enter’를 치면 실험 페이지로 바로 이동하였다. 피실험자는 먼저 상단에 평가 영역 탭을 마우스로 클릭하여 발음 등급 탭을 활성화한 후 평정하였다. 학습자는 세션 종료 후 발음평가 실험 결과를 CSV(comma separated value)인 .Table 포맷 파일로 컴퓨터에 순차적으로 저장하였다. 발음평가 실험을 종료한 후에는 전자우편에 모든 결과 파일을 첨부하여 교수자에게 송신하였다.

2.2.3. 음성의 텍스트화와 단어오류율 측정

학습자들은 1·2차 다중강제선택 실험 사이 기간인 13주차에 기술 기반 영어 발음평가 피드백에 단기간 노출되었다. ASR·WER 발음평가 때 사용한 파이썬 코드는 두 부분으로 나누어졌다. 먼저 자동음성인식 도구 부분은 .wav 포맷으로 저장된 학생 발화를 입력 데이터로 사용하여 텍스트로 출력할 때 사용되었다. 다음으로 단어오류율 도구 부분은 전사된 텍스트를 가설 문장(hypothesis text)으로 사용하여 참조 문장(reference text)과 비교할 때 사용되었다. 아래 수식 (1)은 단어오류율을 측정하는 수식이다. 구체적으로, 삽입된 오류 단어 수(I), 탈락된 오류 단어 수(D), 그리고 대체된 오류 단어 수(S)의 합을 분자로 하고 총 단어 수(N)를 분모로 하여 나눈 값이 단어오류율이다.

$$WER = \frac{I + D + S}{N} \quad (1)$$

2.2.4. 통계 분석

통계 분석은 R(4.5.2)을 사용하여 수행하였다(R Core Team, 2025). ASR·WER 기반 영어 발음평가 이전에 실시한 다중강제선택 발음평가는 사전 실험으로, 이후에 실시한 평가는 사후 실험으로 분류하였다. 총 9명의 피실험자는 각 회기에서 7점 척도의 평정(goodness rating)과 반응시간(reaction time, RT)을 보고하였으며, 본 연구에서는 평정 자료만을 분석 대상으로 하였다. 입력 데이터는 총 2,160개 토큰으로 구성되었다[2(실험순

3 다중강제선택 발음평가 실험에 앞서 각 평가항목에 대하여 수업 시간에 구체적으로 설명하는 시간을 가졌다. 이와 함께 다중강제선택 발음평가 실험 첫 안내 페이지에 한글 설명을 병기하였다. “Comprehensibility”는 “외국인 엑센트가 있다 하더라도 청자가 화자의 말을 잘 이해하여 알아들을 수 있음”으로, “Fluency”는 “화자 발음이 영어원어민과 비슷하여 자연스럽게 유창함”으로, “Accuracy”는 “화자의 자모음 발음, 강세, 억양의 정확함”으로, “Intelligibility”는 “화자 발음이 명료하고 적당한 속도로 문장은 발음 없이 발화함”으로 안내문을 병기하였다(cf., Munro & Derwing(2015)은 모국어 발화 평가에서 고려되는 지각 기반 평가 요소들을 표로 정리하며 “Accentedness(How different is the speech from a local variety?); Comprehensibility(How easy is it understand the speech?); Fluency(To what degree is the speech free of pauses, repetitions, hesitations, false starts, etc.); Intelligibility(How much does the listener actually understand?)”라고 제안하였다(p. 5).].

서)×10(자극)×4(평가항목)×3(반복)×9(피실험자)]⁴

선형 혼합효과 모형(linear mixed-effects model)은 평정(goodness rating)을 종속변수로 설정하였고, 실험순서(experiment order), 평가항목(response category), 양자의 상호작용(experiment order × response category), 그리고 반복(repetition)을 고정효과(fixed effects)로 지정하였다. 무선효과(random effects)는 피실험자에 무선 절편(random intercept)과 실험순서에 대한 무선 기울기(random slope)를, 자극에 무선 절편을 포함하는 구조로 설정하였다. 우도비 검정(likelihood ratio test)을 통해 무선 기울기를 포함한 모형이 더 적합한 것으로 나타나, 해당 모형을 최종 분석에 사용하였다. 분석에는 tidyverse, lme4, lmerTest, performance, emmeans 등의 라이브러리를 활용하였다.

3. 결과

설문조사에서 음성학 수업 수강생 10명(여성 8명과 남성 2명)이 응답하였다. 설문은 총 14개의 문항으로 구성하였으나 아래에서는 본 연구와 직접적 관련이 있는 설문 문항 및 결과만 보고한다.

3.1. 학습 분야 효과 조사

본 연구에서 영어 발음 향상을 위해 기초로 삼은 총 7개의 학습 분야는 i) 조음음성학, ii) 음향음성학, iii) 조음음운론, iv) 음조와 경계 지수(tone and break indices, ToBI) 억양이론, v) 프랏 활용 음성 분석, vi) 다중강제선택 발음평가 실험, vii) 자동 음성 인식과 단어오류율(ASR · WER)이다. (i)에서 (v) 단계까지 학습자들은 발음 연습을 동반하였다. 표 1에 1번-2번은 기초 질문이고, 표 2 3번-5번은 심층 질문이다. 각 표에는 설문 문항과 함께 학습자 응답률(%)이 표시되어 있다.

표 1의 설문지 기초 조사 결과는 다음과 같다. 학습자들은 음성학 수업 수강을 통해서 프랏을 사용하여 음성을 분석해 본 경험이 가장 도움이 되었다고 응답하였다(10명 중 6명). 조음음운론, 음향음성학, ASR · WER 발음평가가 도움이 되었다고 응답한 학습자는 각각 1명씩이었다. 반면 조음음성학이나 음조와 경계 지수 억양이론 분야는 영어 발음 향상을 위한 음성학 수업에서 학습자들의 선택을 받지 못하였다. 모든 학습 분야가 도움이 되었다고 응답한 1명의 학생을 제외하고는 이 두 분야가 음성학 교육에서 학습자들이 학습 효능감을 크게 느끼는 분야는 아니었다. 다음으로 2번 문항에서는 기술 기반 시대를 살아가는 학습자들이 과학적 · 논리적 사고력 향상에 도움받은 부분을 조사했을 때, 학습자들은 프랏을 사용하여 음성을 분석해 본 경험을 가장 높이 평가했고(10명 중 5명), 조음음성학과 음향음성학은 20%의 응답률을 기록하며 다음으로 꼽혔다. 이 두 개의 설문 문항 결과를 종합하면, 음조와 경계 지수 억양이론 및 관련 발음

연습은 본 수업에서 다른 학습 요소에 비해 상대적으로 낮은 학습 효능감을 제공한 것으로 나타났다.

표 1. 설문지 기초 조사 문항
Table 1. Exploratory survey items

1. 이번 학기 수업 내용 중 가장 도움이 된 부분은 무엇입니까?	
①	조음음성학 (0%)
②	조음음운론 (10%)
③	음향음성학 (10%)
④	ToBI 인토네이션 (0%)
⑤	ASR · WER 발음평가 (10%)
⑥	Praat 분석 사용법 (60%)
⑦	모든 분야 (10%)
2. 과학적이고 논리적 사고에 도움을 가장 많이 받은 분야는 무엇입니까?	
①	조음음성학 (20%)
②	조음음운론 (0%)
③	음향음성학 (20%)
④	ToBI 인토네이션 (0%)
⑤	ASR · WER 발음평가 (0%)
⑥	Praat 분석 사용법 (50%)
⑦	모든 분야 (10%)

표 2의 설문지 심층 조사 결과는 다음과 같다. 문항 3번 질문인 조음음성학, 조음음운론, 음조와 경계 지수 억양이론 및 실습과 관련하여 각 분야에서 학습자들이 가장 교육적 효과를 경험한 분야를 좀 더 자세히 알아보았다.

표 2. 설문지 심층 조사 문항
Table 2. Detailed follow-up survey items

3. 조음음성학 중에서 가장 도움이 된 분야는 무엇입니까?	
①	자음 조음위치 및 조음방법 (0%)
②	모음 조음위치 및 조음방법 (30%)
③	자음과 모음 조음위치 및 조음방법 모두 (70%)
4. 영어 자음 및 모음과 관련한 조음음운론 내용 중에서 가장 도움이 된 분야는 무엇입니까?	
①	조음방법과 협동조음(gestural coordination) (30%)
②	음운환경에 따른 음운변화 (40%)
③	조음방법, 협동조음, 분절음 환경에 따른 음운 규칙 모두(30%)
5. ToBI 억양이론을 배우고 실습하는 것을 통해 가장 도움이 된 분야는 무엇입니까?	
①	다양한 조합의 어조와 경계 지수가 조합된 음원을 들어본 경험 (70%)
②	ToBI 주석이 표기된 것을 보며 영어 발화 연습을 한 경험 (20%)
③	무작정 따라 읽은 경험 (10%)

학습자들은 조음음성학 분야에서 자음 발화보다는 모음의 조음위치 및 조음방법을 정확히 배운 것에 가장 만족하였다. 자음과는 달리[예외: 활음(/j/, /w/)과 반전음(/r/)] 모음은 구강 내 조음 협착이 일어나지 않기 때문에 조음(gesture)적 관점에서는 정

4 ASR · WER 발음평가 활동 전 MFC 발음평가 실험은 「사전」으로 처리하였고, ASR · WER 발음평가 활동 후 MFC 발음평가 실험은 「사후」로 처리하였다. ASR · WER 발음평가에 참여하지 않은 1인은 선형 혼합효과 모형을 사용한 통계 분석에서 제외하였다.

확한 조음위치를 직관적으로 습득하는 것이 쉽지 않기 때문에 교육적 효과가 높은 것으로 보인다(Browman & Goldstein, 1986, 1989, 1992, 1995). 문항 4번 질문인 조음음운론과 관련하여, 음절 위치에 따른 조음방법과 협동조음에 대하여 배운 것이 도움이 되었다는 응답률(30%)과 음운환경에 따른 음운변화를 배운 것이 도움이 되었다는 응답률(40%), 그리고 모든 분야에서 도움이 되었다는 응답률(30%) 사이에는 거의 차이가 없었다. 마지막으로 문항 5번 질문인 음조와 경계 지수 억양이론과 관련하여 음조와 경계 지수의 여러 조합이 가능하다는 점과 그를 실제로 적용한 음원을 다양하게 청취해 본 경험이 가장 도움이 되었다는 응답이 가장 많았다(10명 중 7명). 비록 이들 세 분야가 기초 조사 결과상으로 학습자들의 응답이 저조한 항목이었다 하더라도, 심층 조사 결과에서 학습자가 경험한 긍정적 효과를 특정할 수 있었다. 첫째, 조음음성학은 학생들이 자음의 조음위치 및 방법과 모음의 발화 방법 습득 시 도움을 준 것으로 볼 수 있다. 둘째, 음운환경에 의한 음운변화를 조음적 측면에서 배운 것은 유익했으며, 다양한 조합의 음조와 경계 지수 조합의 청취 경험은 다양한 영어 억양 양상 인식에 도움을 준 것으로 볼 수 있다.

3.2. 비전문 평가자로서의 신뢰성

설문조사 6번 항목과 7번 항목에서는 음성학 수강 전·후에 학습자 본인이 느끼는 영어 발음 수준을 직관적으로 평가해 보라고 요청하였다. 아래 그림 2(a)는 음성학 과목 수강 전의 결과로 10명의 학습자 평균은 7단계 중 4.00 단계였고, 그림 2(b)는 영어음성학 과목 수강 후의 결과로 학습자 평균은 3.23 단계였다.

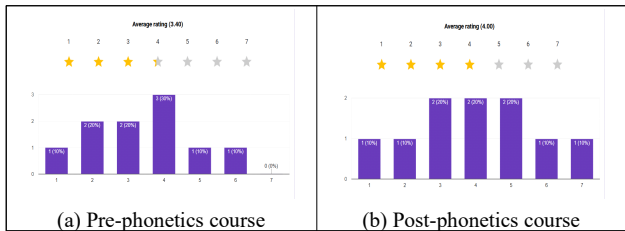


그림 2. 학습자 본인 발음 수준에 대한 직관적 자가 평가

Figure 2. Learners' self-rated perception of their pronunciation proficiency

설문조사 13번 항목에서는 영어음성학 수강 후 학습자 본인 발화에 향상이 있는지 조사하였고, 10명 중 7명이 영어 발음에 향상이 있었다고 응답하였고(70%), 3명은 없었다고 응답하였다(30%).

위 설문 문항 6번, 7번, 13번의 결과를 종합해 보면, 학습자들

은 음성학 수강 이후 전반적인 발음 향상이 있었다고 인식하는 동시에, 발음을 객관적으로 평가하기 위한 다양한 음향적·음성적 지식을 습득함으로써 자신에 대한 평가가 이전보다 더 엄격해졌을 가능성도 있음을 시사한다.

3.3. 비전문 평가자의 발음 평정에 영향을 주는 요인

선형 혼합효과 모형 분석 결과는 다음과 같다.⁵ 실험순서에 대한 무선 기울기를 포함한 모형이 유의하게 더 적합한 것으로 나타났다($\chi^2(2) = 16.33, p < .001$; (표 3)). 고정효과 분석 결과는 표 4에 제시되어 있다. 평정은 실험순서(사전-사후) 간 유의한 차이를 보이지 않았다($\beta = 0.165, SE = 0.457, p > .05$). 반복측정 Cohen의 d (Morris & DeShon, 2002)를 사용하여 실험순서 간 비교에 대한 효과 크기를 산출했을 때, 두 조건 간의 평균 차이는 매우 작았으며(사전-사후 = 0.058), 차이 점수의 표준편차는 비교적 큰 편으로($SD = 1.00$), 반복측정 효과 크기는 매우 작은 수준이었다($d = 0.06$). 이러한 결과는 전체적으로 피실험자의 평정은 두 실험 세션 간에 실질적으로 변화하지 않았으며, 집단 수준에서 유의미한 평정 수치의 상승도 나타나지 않았음을 시사한다. 정확도를 기준으로 비교하였을 때, 평가항목 중에서 명료도에서 평정만이 유의한 결과를 나타내며, 사전·사후 모두에서 가장 높은 평정을 유지하였다($\beta = 0.449, SE = 0.099, p < .001$). 실험순서와 평가항목의 상호작용은 이해도에서만 유의하였으며, 사후에 비해 사전 평가에서 평정이 더 높은 것으로 나타났다($\beta = 0.346, SE = 0.135, p < .05$).

무선효과 분석 결과는 표 5에 제시되어 있다. 무선효과 분산은 자극($\sigma^2 = 0.951$)과 피실험자($\sigma^2 = 0.680$)에 기인하는 변동이 상당함을 보여주었고, 실험순서의 무선 기울기 분산은 비교적 작았다($\sigma^2 = 0.078$). 마지막으로 표 6에서 확인할 수 있듯이, 고정효과만으로는 설명력이 낮았고(주변 $R^2 = 0.017$), 무선효과가 전체 변동을 설명하는 핵심 요소였다(조건부 $R^2 = 0.628$). 이는 청자의 발음 평정이 실험순서(사전-사후) 간 변화보다 자극 고유의 음성적 특성과 청자 개인의 음향적·음성적 지식에 의해 좌우되었을 수 있다는 점을 시사한다.

표 3. 선형 혼합효과모형 분산분석 결과

Table 3. ANOVA results from linear mixed-effects models

	npar	AIC	BIC	logLik	-2*Log(L)	Chisq	Df	Pr(> t)
m_basic	13	5,167.6	5,238.8	-2,570.8	5,141.6			
m_adv	15	5,155.3	5,237.4	-2,562.6	5,125.3	16.332	2	0.0002841***

* $p < .05$, *** $p < .001$

5단계 분석과는 별도로, 정확도·이해도·유창도·명료도의 각 평정 영역에 대해 개별적인 선형 혼합효과 모형 분석을 수행하였다. 네 개의 모든 영역에서 모형 비교 결과 무선 기울기를 포함한 복잡한 모형은 적합도를 유의하게 향상시키지 않았으며, 피실험자와 자극에 대해 무선 절편만을 포함한 기본 무선효과 구조가 데이터에 적합한 것으로 나타났다[정확도, $\chi^2(2) = 2.515$; 이해도, $\chi^2(2) = 3.378$; 유창도, $\chi^2(2) = 1.647$; 명료도, $\chi^2(2) = 2.367$, 모두 $p > .05$]. 그 결과 4개 영역 모두에서 실험순서 사후와 사전에 차이가 보이지 않았다[정확도, $\beta = -0.084, SE = 0.517$; 이해도, $\beta = 0.304, SE = 0.408$; 유창도, $\beta = 0.047, SE = 0.463$; 명료도, $\beta = -0.168, SE = 0.475$, 모두 $p > .05$].

표 4. 선형 혼합효과모형 고정효과 추정치

Table 4. Fixed effects estimates from linear mixed-effects models

	Estimate	Std.Error	df	t-value	Pr(> t)
(Intercept)	3.647e+00	4.206e-01	2.540e+01	8.671	4.59e-09***
experiment_pre	-1.654e-01	4.569e-01	2.284e+01	-0.362	0.7208
response_comp	1.105e-01	9.745e-02	1.732e+03	1.133	0.2572
response_fl	-9.2663e-02	9.636e-02	1.730e+03	-0.962	0.3364
response_in	4.487e-01	9.935e-02	1.731e+03	4.516	6.73e-06***
repetition_rep2	4.723e-04	5.753e-02	1.730e+03	0.008	0.9935
repetition_rep3	-2.294e-02	5.805e-02	1.731e+03	-0.395	0.6928
experiment_pre:response_comp	3.455e-01	1.354e-01	1.732e+03	2.552	0.0108*
experiment_pre:response_fl	1.435e-01	1.357e-01	1.732e_03	1.058	0.2903
experiment_pre:response_in	1.329e+02	1.358e-01	1.733e+03	0.098	0.9220

* $p < .05$, ** $p < .01$, *** $p < .001$

표 5. 선형 혼합효과모형 무선효과 분산 성분

Table 5. Random effects variance components from linear mixed-effects models

Groups	Name	Variance	Std.Dev	Corr
stimuli	(Intercept)	0.95127	0.99753	
subject	(Intercept)	0.67970	0.8244	
	experiment_pre	0.07752	0.2784	-0.21
Residual		0.98677	0.9934	
Number of obs: 1765		groups:	stimuli, 20	subject, 9

표 6. 선형 혼합효과모형 설명력 지표

Table 6. Model fit indices for linear mixed-effects models

Conditional R^2	Marginal R^2
0.628	0.017

4. 결론 및 논의

본 연구는 조음음성학, 음향음성학, 조음음운론, 음조와 경계 지수 억양이론에 기반하여 학습자가 영어 발음을 과학적으로 분석하고 체계적으로 개선할 수 있는 통합적 교육을 경험하는 과정을 조명하였다. 학습자는 원어민 발화와 비원어민 발화 간의 차이를 조음음운론 측면의 이론적 배경에 근거하여 음향학적 관점에서 객관적으로 이해하고, 자신의 발화를 실험적 방법을 통해 점검하며, 정량적 피드백에 기반한 자기 주도적 발음교정 과정을 경험하였다. 이 과정에서 얻은 구체적인 성과를 기술하면 다음과 같다. 첫째, 학습자는 자신의 발화 데이터를 녹음하고 실험 자료로 활용하는 경험을 통하여 학습자 중심의 참여적 수업 모델 구현에 동참하였다. 둘째, 학습자는 프랏 음성 분석 도구를 활용하여 본인의 영어 발화를 음향적·인지적 관점에서

진단하고 스스로 개선 방향을 도출하는 역량을 길렀다. 셋째, 학습자는 음운환경에 의한 음운변화를 조음적 측면에서 배운 것이 유의하다고 생각했다. 넷째, 학습자는 다양한 음조와 경계 지수를 조합한 음원을 청취한 경험이 다양한 영어 억양 양상을 인식하는 데에 도움이 된다고 보았다.

이와 더불어 본 연구는 인공지능 기반 발음 평가 도구의 사용을 통해 융복합적·통합적 학습 경험을 확장하고, 4차 산업혁명 시대에 요구되는 디지털 리터러시와 언어 데이터 해석 능력을 함양을 도모하였다. 구체적인 성과를 정리하면 다음과 같다. 첫째, 시중에 유료로 제공되는 발음평가를 구독하지 않고도, 학습자의 디지털 역량을 활용하여 최신 기술을 자유롭게 사용할 수 있게 하였다. 둘째, 영어 발음교정을 하기 위해 전문가의 코칭을 받지 않고도, 실험음성학과 컴퓨터 언어의 통합적 소양을 동시에 갖추고 독자적으로 발음 연습을 할 때 꾸준한 영어 발음 향상 효과가 있을 것으로 기대한다(Kim, 2023; Liakin et al., 2015). 셋째, 평가 방식 또한 이론 시험, 발화 녹음 과제 수행, 다중강제선택 영어 발음평가 실험 참여, ASR·WER 파이썬 코딩 스크립트 실행 및 결과 제출, 설문 참여 등 다양하게 구성되었기에 학습자의 다면적 성장을 정량 및 정성적으로 평가할 수 있는 근거를 마련하였다. 위와 같은 일련의 학습 경험은 학습자들이 장기적으로 자기주도적 발음 향상을 이어갈 수 있는 학습 생태 환경 조성에 지속적으로 기여할 것이다.

4.1. ASR·WER 이 청자 평정의 차원별 민감도 변화에

끼치는 제한적 역할 및 연구의 한계성

본 연구에서는 ASR·WER 활용 발음평가 경험을 전통적 인지평가를 기반으로 한 다중강제선택 발음평가에 통합시켰을 때 청자의 평정 방향 변화에 대하여 조사해 보았다. 실험순서(사전-사후)에 따른 전체적 평정의 상승 혹은 하락은 확인되지 않았던 점이 시사하는 바는 자동화된 피드백에 단기적 노출만으로 청자의 평정을 체계적으로 변화시키기에는 충분하지 않다는 것이다. 그러나 이해도 평정에서 나타난 실험순서와 평가 항목의 상호작용은 ASR·WER 활용 발음평가 경험이 특정 평가 차원에서 선택적으로 영향을 미칠 수 있다는 것을 확인시켜 주었다. 교육적·평가적 관점에서 볼 때, ASR·WER 발음 평가 활동 경험이 청자의 전체적인 평정 변화를 이끌지는 않더라도, 특정 평가 차원에 대한 민감도를 높이는 데 도움을 줄 가능성은 다소 열려 있는 것으로 해석할 수 있다.

반면, 평정은 무선효과가 전체 변동의 핵심적인 설명 요인임을 보여주었는데, 구체적으로는 개별 피실험자의 음향적·음성적 발음평가 기준과 자극에 영향을 받는 것으로 나타났다. 이러한 결과는 ASR·WER 활용 발음평가 활동을 수업에서 실제 적용할 시에는 다음과 같은 점을 고려해야 함을 시사한다. 교육적 맥락에서 발음평가 피드백의 효과를 실질적으로 얻기 위해서는 ASR·WER 활용 발음평가에 단기적 노출보다는 지속적이고 반복적인 교육적 적용이 필요할 것으로 보인다. 하지만, 학습자가 ASR·WER 활용 발음평가 활동에만 국한되어 노출되기보다는 전문가 발음평가에 단기적으로나마 함께 노출될 필

요성도 고려해 봐야 할 것이다. 구체적으로는 전문가 평가를 인공지능 기반 발음평가의 결과에 타당성을 부여하는 하이브리드 결합 방식을 생각해 볼 수 있다. 장기적으로 비전문 평가자의 평가 타당도와 일관성 향상에 있어 효과적 수단이 될 수 있을 것이다.

다만, 수업 설계 단계에서 ASR·WER 활용 발음평가 활동에 노출되지 않은 대조군이 포함되지 않았다는 점은 인공지능 기반 발음평가 경험이 전통적 인지 평가를 기반으로 한 발음평가에 미친 영향을 객관적으로 해석하는 데 걸림돌이 되었고, 연구의 내적 타당성을 약화하였다. Cook et al.(2002)는 단일 집단 사전·사후 설계는 연구 과정에서 불가피하게 개입하는 성숙, 검사 반복, 측정 도구 변화와 같은 내적 타당도 위협 요인들에 취약하므로, 해당 설계만으로는 특정 중재 효과를 명확히 입증하기 어렵다고 하였다. 본 실험에서 피실험자의 수가 경우에 따라 9-10명으로 소수이기에 실험군과 대조군을 나누어 차별된 교육 과정을 실시하지 못한 한계가 존재한다. 후속 연구에서는 대조군을 포함한 실험 설계를 실제 구현함으로써 본 연구에서 부족한 내적 타당성 문제를 보완할 필요가 있다.

이 외에도 설문지의 항목을 좀 더 세밀하게 고안하지 못했던 점과 다른 영어 발음 교육 방법과의 비교가 이루어지지 않은 점은 본 연구의 또 다른 한계점이다. 이와 함께, 심사자가 지적하였듯이, 본 연구에서는 다중강제선택 실험에서 자극으로 사용한 발화에 대하여 정확한 음성학적 분석을 함께 제시하지 않은 점과 영어원어미이나 전문 평가자의 영어 발음평가 등급을 기준으로 제시하지 못한 점은 기준 평정이 처음부터 적절히 제시되지 못한 결함으로 이어진다. 이 부분은 향후 연구에서 필수적으로 보완해야 할 점이다.

여러 가지 미진한 점에도 불구하고 본 연구는 통합적·실천적인 수업 방식을 실천하여 음성학/음운론 이론을 활용한 발음 교육, 디지털 도구(프랏, 파이썬 코드, ASR·WER), 평가 방법론(다중강제선택)을 하나의 교과안에서 포함하는 시도를 하였고, 융합적·통합적 교육의 결과를 보고함으로써, 교육 현장에서 영어 발음 교과 관련 교수자들이 참고할만한 실용적인 수업 모형을 제공한 것에 의의를 두고자 한다.

4.2. 인공지능 기반 발음평가기의 지속적인 성능 향상

또 다른 논의 지점은 인공지능 기반 발음평가기의 신뢰성 문제이다. Kim(2023)의 연구는 Whisper를 활용하여 전문 발음 평가자와 유사한 수준의 평가 결과를 도출하였으며, 이를 근거로 인공지능 기반 평가기가 전문 평가자를 효과적으로 대체할 수 있다고 주장하였다. 본 연구에서도 학습자는 클라우드 기반 자동음성인식 시스템을 이용하여 발화를 텍스트로 변환하고 단어 오류율을 산출함으로써 발음에 대한 즉각적 피드백을 얻을 수 있었다. 그러나 ASR 기술의 최근 발전은 예기치 않은 역설적 문제를 제기한다. 전통적인 GMM-HMM(Gaussian mixture model-hidden markov model) 기반 시스템에서 DNN-HMM(deep neural network-hidden markov model), RNN(LSTM), 그리고 Transformer 기반 엔드투엔드 모델(OpenAI Whisper)로 발전하면

서, 음성인식기는 잡음 환경이나 비표준 발화에 대해 더욱 강한 인식 능력을 제공하게 되었다(Dong et al., 2018; Graves et al., 2013; Hinton et al., 2012; Hochreiter & Schmidhuber, 1997; Rabiner, 2002). 이러한 기술적 고도화는 부정확한 발음조차 목표 단어로 자동으로 보정되어 인식되는 현상을 초래할 수 있고, 결과적으로 실제 발음 오류가 WER에 반영되지 않는 문제가 발생할 수 있다. 그러나, 심사자가 지적하였듯이, 최신 ASR 시스템에서 관찰되는 부정확한 발화의 과대 인식 현상은 모델이 단어 출력을 판단할 때 사용하는 임계값(threshold) 조정을 통해 완화할 수 있기에 이를 학습자에게 적절히 안내하는 것만으로 간단히 해결할 수도 있다. 이와 함께 교수자는 학습자에게 자동음성인식 기술이 전문 발음 평가자를 대신할 수 있을 만큼 발전하여 학습자 본인 발화 평가에서 편리하게 사용할 수 있지만, 인공지능 기반 발음 인식기가 추구하는 기술력의 방향성을 분명히 전달하여 학습자들이 자동음성인식 기술을 발음평가에 사용할 때 좀 더 신중한 견지를 유지하도록 도와야 한다. 전문가 발음평가나 세부 음향학적 지표에 대한 정확한 이해는 발음평가에 객관성과 타당성을 항상 보완하는 방법이기때문에, 앞으로 더더욱 향상될 자동음성인식 기술을 이용하여 편리하게 발음평가를 시행하되, 전통적 방법을 병행하여 활용할 때 발음평가의 객관성과 타당성이 보완될 수도 있을 것이다.

유료 프로그램은 학습자의 비용 부담이나 교육 기관의 비용 정책에 따라 사용 여부가 결정되기 때문에, 필요에 따라 교실 학습 환경에서 자유롭게 사용하는 것에 한계는 있다. 또한 성능이 뛰어나지 않은 공용 컴퓨터에서 사전 학습 데이터 다운로드가 어려울 수도 있다. 이를 극복하기 위해서 간단한 코딩 기초 지식을 갖고 사전 학습 데이터 다운로드는 필수가 아닌 인공지능 기반 자동음성인식 도구 활용이 가능하다면, 교육 현장에서 첨단 도구를 간편하게 사용할 수 있어 유리하다.

비록 인공지능 자동 발음평가 프로그램은 학습자의 디지털 리터러시 능력이 요구되어 기초적인 컴퓨터 언어 사용이 선행되어야 가능하다는 한계점이 존재하지만(Kim, 2023), 교수자는 파이썬 언어와 같은 학습자가 비교적 배우기 쉬운 고급 컴퓨터 언어(high-level computer language)를 사용하여 음성학에 코딩 교육을 결합하는 시도를 해볼 만한 가치가 있다. 일련의 학습 과정을 통하여 학습자는 기술 기반 학습 도구 사용을 적극적으로 할 수 있고, 자기 효능감을 높이는 학습 활동을 통하여 자기 주도적으로 영어 발음을 향상할 수 있을 것이다.

감사의 말

논문의 완성도를 높이는 데 귀중한 시간과 전문적 식견을 바탕으로 비평과 조언을 주신 세 분의 심사위원께 깊이 감사드립니다. 또한 발음평가 인지 실험에 사용한 MFC 스크립트를 지원해 주신 충남대학교 윤수연 교수님께 감사드리며, 통계 분석과 관련하여 부족한 부분을 보완할 수 있도록 조언해 주신 서울대

학교 전종호 교수님께도 감사드립니다. 아울러 혁신 교과 수업에 성실히 임해준 한남대학교 응용영어콘텐츠학과 2023학번 10명의 학생들에게도 감사의 뜻을 전합니다. 본 논문의 모든 오류는 저자의 책임입니다.

References

- Ahn, T., & Nam, H. (2024). Fine-tuning-free English pronunciation assessment system using multimodal large language models. *Studies in Foreign Language Education*, 38(4), 25-38.
- Best, C. T. (1990). Adult perception of nonnative contrasts differing in assimilation to native phonological categories. *The Journal of the Acoustical Society of America*, 88(S1), S177-S178.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, 109(2), 775-794.
- Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (version 6.1.16, MacOS 64-bit). [Computer program]. Retrieved from <https://www.praat.org>
- Boersma, P., & Weenink, D. (2024). Praat: Doing phonetics by computer (version 6.4.21, Windows 64-bit). [Computer program]. Retrieved from <https://www.praat.org>
- Browman, C. P., & Goldstein, L. M. (1989). Articulatory gestures as phonological units. *Phonology*, 6, 201-251.
- Browman, C. P., & Goldstein, L. M. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4), 155-180.
- Browman, C. P., & Goldstein, L. M. (1995). Gestural syllable position effects in American English. In *Producing speech: contemporary issues* (pp. 19-33). American Institute of Physics.
- Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3, 219-252.
- Brown, A. (1991). *Pronunciation models*. Singapore, Singapore: NUS Press.
- Candlin, C. N., & Mercer, N. (2007). *Communicative language teaching: An introduction*. Cambridge, UK: Cambridge University Press.
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation hardback with audio CDs (2)*. Cambridge, UK: Cambridge University Press.
- Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (Vol. 1195). Boston, MA: Houghton Mifflin.
- Dale, P., & Poms, L. (2005). *English pronunciation made simple*. London, UK: Pearson Education.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379-397.
- Dong, L., Xu, S., & Xu, B. (2018, April). Speech-transformer: A no-recurrence sequence-to-sequence model for speech recognition. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5884-5888). Calgary, AB, Canada.
- Graves, A., Mohamed, A., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6645-6649). Vancouver, BC, Canada.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., Senior, A., ... Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), 82-97.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Hong, Y., & Nam, H. (2021). Evaluating score reliability of automatic English pronunciation assessment system for education. *Studies in Foreign Language Education*, 35(1), 91-104.
- Kim, J. E. (2018). An analysis of English pronunciation for high-level proficiency adult learners. *Phonetics and Speech Sciences*, 10(2), 39-44.
- Kim, H., & Chung, H. (2023). Phonetic features influencing native English speakers and Korean EFL teachers' comprehensibility and accentedness judgments. *Studies in Foreign Language Education*, 37(4), 83-107.
- Kim, M. (2020). A study of rhythm improvements and relevant linguistic factors in the pronunciation of English learners. *Foreign Language Education Research*, 34(1), 237-261.
- Kim, M. (2023). Digital enhancement of pronunciation assessment: Automated speech recognition and human raters. *Phonetics and Speech Sciences*, 15(2), 13-20.
- Kim, M. (2024). Efficacy of explicit pronunciation training: Perspectives of pre-service English teachers. *Studies in Foreign Language Education*, 38(4), 1-24.
- Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *The Journal of the Acoustical Society of America*, 121(4), 2327-2338.
- Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French/yl. *CALICO Journal*, 32(1), 1-25.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25(3), 481-520.

- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105-125.
- Munro, M. J., & Derwing, T. M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 375-396). Hoboken, NJ: John Wiley & Sons.
- Nunan, D. (1991). Communicative language teaching: A critical overview. *TESOL Quarterly*, 25(3), 367-375.
- R Core Team. (2025). R: A language and environment for statistical computing (version 4.5.2) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rabiner, L. R. (2002). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286.
- Riney, T. J., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly*, 34(4), 711-737.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992, October). ToBI: A standard for labeling English prosody. *Proceedings of the International Conference on Speech Language Processing* (pp. 867-870). Banff, Canada.
- Sung, E. (2010). Perspectives on an online English pronunciation course and pronunciation error analysis. *Studies in Foreign Language Education*, 24(1), 175-208.
- Tsukada, K., Birdsong, D., Mack, M., Sung, H., Bialystok, E., & Flege, J. (2005). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica*, 61(2-3), 67-83.
- Veilleux, N., Shattuck-Hufnagel, S., & Brugos, A. (2006). Transcribing prosodic structure of spoken utterances with ToBI. MIT OpenCourseWare. Retrieved from <https://ocw.mit.edu/courses/6-911-transcribing-prosodic-structure-of-spoken-utterance-s-with-tobi-january-iap-2006/>

- **손민정 (Son, Minjung)** 교신저자
 한남대학교 응용영어콘텐츠학과 교수
 대전광역시 대덕구 한남로 70 문과대학 218호
 Tel: 042-629-7342
 Email: minjungson@hnu.kr
 관심 분야: 음성학, 음운론, 언어습득, 자연어처리

융복합적 통합 음성학 교육을 통한 학습자 영어 발음학습 자율성 증진

손민정

한남대학교 응용영어콘텐츠학과

국문초록

본 연구는 기술 기반의 음성학적 접근을 통해 학습자의 영어 발음 능력을 향상시키는 것을 목표로 한다. 본 교과는 음성학, 음운론, 억양(ToBI), 발음 훈련, 발음 평가를 통합한 단일 교육 설계로 구성되었다. 학습자들은 체계적인 발음 연습을 수행하고 프랏을 활용한 음성 분석을 하였으며, 발화 샘플을 녹음하고 다중강제선택 영어 발음평가 실험에 참여하였다. 또한 파이썬 코드를 사용하여 인공지능(AI) 기반 음성인식을 구현하고, 단어오류율(WER) 지표를 사용하여 발음을 평가하였다. 본 교과는 융합적·통합적·실험적 관점에서 설계되어 학습자의 자기주도적 학습을 촉진할 수 있도록 고안되었다. 설문조사 결과 약 70%의 학습자가 전반적인 발음 능력의 향상이 있었다고 보고하였다. 그러나 인공지능 기반으로 자동화된 ASR·WER 피드백에 단기적 노출은 피실험자의 평정을 체계적으로 변화시키기에는 충분하지 않았고, 특정 평가 차원에 대한 민감도에 제한적인 영향을 주는 것으로 그쳤다.

핵심어: 음성학, 음운론, 음조와 경계 지수, 다중강제선택, 영어 발음평가, 자동음성인식, 단어오류율, 파이썬

참고문헌

- 김미란. (2020). 영어학습자 발음의 리듬향상과 관련된 언어요인 연구. *외국어 교육 연구*, 34(1), 237-261.
- 김지은. (2018). 발음 숙련도 상위 성인 학습자들의 영어 발음에 대한 분석. *말소리와 음성과학*, 10(2), 39-44.