

음성인식 피드백 프로그램을 활용한 영어 학습 프로그램 개발 연구*

이경량
(세종대학교)

Lee, Kyoung Rang. 2018. Developing English learning program using speech recognition-based feedback. *Korean Journal of English Language and Linguistics* 18-4, 486-508. This study was planned to explore whether a computer program using speech recognition-based feedback (ET-Bot) would improve Korean middle school English learners' speaking. The results of 46 middle school English learners were analyzed in terms of the accuracy and the time of pronouncing words, sentences, and a paragraph which had been carefully chosen, based on the previous studies and existing pronunciation diagnosis tools. Immediate feedback provided by the speech recognition tool right after participants read aloud seemed to motivate them to correct themselves. Even though there were not big differences in terms of gender as well as English performance scores, possible gender differences should be further explored in terms of their attitudes and affective factors to pronouncing English words and sentences. Also, the free speaking task that ET-Bot provided in addition to reading aloud tasks showed its possible usage to diagnose the participants' speaking proficiency. Using a speech recognition-based feedback program, this study hopes to provide a firm basis for future studies about helping English learners correct their own mistakes with the immediate feedback, helping teachers diagnose their students' speaking, and suggesting a new way to research on speaking by helping researchers transcribe easily. Detailed results and discussions are provided.

Keywords: speech recognition, immediate feedback, English learning program, reading aloud, gender differences, ET-Bot

1. 서론

효과적으로 영어를 학습할 수 있도록 테크놀로지를 활용한 역사는 상당히 긴 편이며, 이를 알아보기 위하여 다양한 분석이 시도되었다. 예를 들면, 1912년부터 2012년까지 100년 동안 출간된 논문(*English Journal*)에서 영어 학습을 위해 새로운 테크놀로지를 도입한 연구에 대한 논문을 모두 훑어본 결과, 오디오 미디어를 활용한 경우 117건, 컴퓨터 미디어를 활용한 경우 198건, 활동사진 미디어를 이용한 경우 311건을 포함하여 총 787편으로 놀랍게도 매우 오래전부터 많은 교사와 연구자가 그 당시 새로운 기술을 활용하고 있었다

* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A5A2A01027326)

(McCorkle and Palmeri 2016). 특히 컴퓨터를 활용한 경우는 1980년대 초반부터였고, 컴퓨터 활용 수업에 대한 전문 저널들이 상당수인 것을 감안하면 컴퓨터는 짧은 기간에 급진적으로 도입이 되고 있다. 교사들은 컴퓨터를 학생 중심 수업을 유도하기에 매우 효과적인 도구라고 생각했으며, 이를 도입한 수업의 대부분이 쓰기 등 산출 수업(151건, 76%)이었다. 컴퓨터와 컴퓨터를 활용한 프로그램들은 다수의 학생들이 쓰기 활동을 하는데 매우 효과적이라서(Ulusoy 2006), 영어가 모국어인 학생들뿐만 아니라 영어 학습자들도 쓰기 활동을 하기 위해 컴퓨터를 활용하는 경우가 대부분이었다.

그렇다면 영어 학습을 위해 컴퓨터를 활용하는 이유는 뭘까? 영어와 같은 외국어를 학습할 때 학습자들은 특히 걱정(foreign language anxiety)과 같은 정의적 요인에 의해 영향을 받으며, 그 중에서도 말하기 과업을 해야 할 때 학습자들의 걱정이 극대화된다고 한다(Young 1999). 걱정을 많이 하게 되면 정상적인 수행을 하기 어려워 학습자들의 걱정을 덜어줄 수 있는 환경을 만들어주는 것이 중요하다는 정의적 필터 가설(Krashen 1981)에서도 볼 수 있듯이, 학습자들이 외국어를 구사해야 할 때 걱정을 너무 많이 하게 되면 외국어 학습 과정이나 평가 결과에 부정적인 영향을 미친다는 연구 결과도 많이 찾아볼 수 있다(Hewitt and Stephenson 2011, MacIntyre and Gardner 1991). 그러나 적당한 정도의 걱정은 오히려 학생들에게 동기 부여를 하는 등 긍정적인 영향을 미친다는 연구 결과 또한 있으며(Ehrman and Oxford 1995), 이를 바탕으로 컴퓨터를 이용하여 원어민과 직접 대면하여 말할 때보다 적당한 정도의 걱정을 할 수 있는 환경(low-anxiety environment)을 마련할 수 있어 학생들의 말하기 실력을 향상시킬 수 있을 것(Beauvois 1998, Kern 1995, Wu and Marek 2009)으로 예상된다.

컴퓨터를 활용한 대화 환경이 영어 말하기 연습에 효과적이라고 해도 오디오나 비디오 채팅 프로그램 수준을 넘어서 요즘 상용화되고 있는 가상현실(virtual reality) 프로그램을 활용할 경우 오히려 또 다른 수준의 걱정, 다시 말해 정서적인 거부감(resistance)을 느끼게 된다고도 한다(Childs 2010). 연구자 또한 수업 시간에 학생들에게 *Second Life*와 같은 가상현실 프로그램을 활용하는 과제를 부여했을 때 학생들이 기괴하다며 거부감을 느끼는 것을 보았다. 기술적으로는 이미 가상현실 속의 원어민과 대화가 가능한 수준이지만 학습자들은 아직은 더 편안한 프로그램을 활용하여 영어 말하기 연습을 원하는 것을 보았을 때, 한국의 영어 학습자들의 영어 말하기 연습에 최적화된 컴퓨터 활용 방법을 찾을 필요가 있다고 사료되었다.

약 10년간의 컴퓨터를 활용한 외국어 교육 관련 연구를 메타 분석한 또 다른 연구 결과에 의하면 영어 말하기 연습을 할 때 소리내어읽기(reading aloud)와 같은 단순한 과제가 가장 효과적이었다고 한다(Lin 2014). 본 연구는 이에 기초하여 주어진 단어 및 문장을 소리내어 읽을 수 있는 프로그램을 개발하기로 하였다. 최근 무료로 개방된 음성인식 기술을 도입하여 컴퓨터 프로그램을 가상현실보다는 익숙하고 좀 더 직관적으로 대화가 가능한 대상으로 여길 수 있는 환경을 조성하여 한국의 학습자들은 어떤 소리를 발음하기 어려워하는

지 어떤 단어나 문장을 발화하기 어려워하는지 진단하고 연습할 수 있는 기회를 제공하는 것이 필요하다고 판단되었다. 특히 음성인식 기술을 도입하는 이유는 학습자가 소리내어 읽었을 때 즉각적으로 피드백을 줄 수 있기 때문인데, 학습자들에게 제공하는 피드백은 즉각적으로 제공할 때 매우 효과적이라고 하며(Zaborik 1987) 학습자들 또한 피드백을 받는 것이 학습에 도움이 된다고 생각한다(Saito 1994).

소리내어읽기나 자유주제 대화 및 발표 등 외국어 말하기 관련 과제에 대한 연구를 할 때 연구자들에게 가장 큰 부담은 전사(transcribing)라서, 학습자들에게 한 단어씩 제공하여 읽도록 해서 진단하고자 하는 소리(자음 또는 모음, 대부분은 자음)에 집중할 수 있도록 하는 진단도구들이 대부분이다(Eisenberg and Hitchcock 2010). 한 단어씩 제공하는 표준화된 진단도구들은 학습자들이 각 자음이나 모음을 발음할 수 있는지를 시간상 매우 효율적으로 진단할 수 있도록 도와주고 어른들과의 발음 비교도 쉽게 할 수 있게 한다고 한다(Bernhardt and Holdgrafer 2001). 그러나 이런 진단도구를 제대로 활용하기 위해서는 목표로 하는 소리뿐만이 아니라 그 소리를 포함한 단어 전체를 전사하는 것이 필요하다고 하는데(Klein 1984), 올바른 발음뿐만 아니라 잘못된 발음까지도 모두 듣고 전사해야 하는 과정이 연구자들에게 부담이었던 것을 감안하면, 이러한 전사 과정을 포함한 음성인식 프로그램을 활용하여 연구자나 교사의 부담을 덜어주고 학습자들에게 즉각적인 피드백을 제공할 수 있어 매우 효과적일 것으로 기대된다.

따라서 본 연구는 영어 원어민과 대화할 기회가 상대적으로 많지 않은 우리나라 중학교 영어 학습자들을 대상으로 음성인식 피드백 프로그램을 활용하여 학습자들의 영어 말하기 연습을 돕기 위한 프로그램을 개발하여 그 효과를 알아보기 위하여 계획되었다.

2. 이론적 배경 및 선행 연구

2.1 영어 학습을 위한 테크놀로지 활용

구성주의적 학습 이론(constructivist learning theory)에 의하면 언어 학습은 긍정적인 학습 경험의 과정이고 학습자는 관찰, 연습, 탐구, 성찰 등을 통해 객관적인 의미를 구성한다고 했는데, 이는 컴퓨터를 활용하여 학습 환경을 조성했을 때 더 긍정적인 결과를 가져올 수 있다고 한다(Xu 2018). 특히 상호작용 가설(interaction hypothesis)에 따르면 언어 학습이 효과적으로 이루어지기 위해서는 상호작용이 필수적이며 학습자들이 이미 배운 것을 산출해내는 것보다 사회적 환경에서 의미협상(negotiation of meaning)을 하는 상호작용 과정 자체가 언어 학습에 더 중요한 역할을 하기 때문에 상호작용이 활발히 이루어질 수 있는 이해가능한 입력(comprehensible input)을 제공하는 것이 중요하다고 한다(Krashen 1981). 이 때 컴퓨터와 같은 테크놀로지를 활용했을 때, 교실뿐만이 아니라 교실 밖에서도 학습이

지속적으로 일어나야 한다는 모바일 학습이론(theory of mobile learning) (Sharples, Arnedillo-Sanchez, Milrad and Vavoula 2009)에도 부합하며, 전통적인 영어 교육환경에서 보다 더 유의미한 상호작용을 유발하였고 특히 교사의 역할에도 긍정적인 변화를 초래하였다. 이런 변화를 정리하면 다음과 같다(Xu 2018). 전통적인 영어 교육환경에서는 교과서가 교재의 핵심이었다면 테크놀로지를 활용한 교실에서는 교과서 외에도 실시간 블로그나 포드캐스트(podcast)와 같은 교재를 활용하여 교과서 이상의 내용을 다룰 수 있고, 따라서 전통적인 환경에서는 학습 내용이 순차적으로 한 가지 주제에 집중했다면 새로운 환경에서는 동시 다발적으로 모든 자료를 활용하고 나눌 수 있다. 전통적 환경에서 교사는 교실 내 유일한 전문가로서 학생에게 미리 결정된 내용을 가르치고 학생들의 성과에 대해 강조를 했다면, 컴퓨터를 활용하는 환경에서는 그 특성상 교사는 학생들이 목표를 달성하는데 필요한 과정 등을 알아낼 수 있는 도움을 주는 가이드로서의 역할을 하며 학습 성과 자체보다는 학생들의 문제해결 과정이나 방법에 대한 혁신적 시도 능력 등에 대해 강조를 한다.

영어 학습과 교수에 컴퓨터와 같은 테크놀로지가 활용되기 시작한 것은 30여년에 불과하지만 그럼에도 불구하고 영어 교실에서 컴퓨터의 역할은 기술의 발전과 더불어 큰 변화를 거듭하였다. 특히 컴퓨터와 같은 테크놀로지의 발달은 영어교육 현장에서의 학습 자료와 방법에 직접적 영향을 주므로 주목해서 살펴볼 필요가 있다. 컴퓨터를 특히 영어와 같은 언어 학습에 사용하는 방법을 컴퓨터 기반 언어 학습(computer-assisted language learning; CALL)이라고 한다. 1980년대 초반의 문서처리(word processing)나 빈칸 채우기 문제 등의 단순한 기능을 위해 컴퓨터가 사용되었던 반면, 인터넷이나 상호작용이 가능한 웹 기능(web 2.0)과 스마트폰 등 모바일 기술의 발달 등으로 컴퓨터 활용 범위가 극적으로 확장되고 있다(Dudeny and Hockly 2012).

이런 기술의 발달과 컴퓨터의 활용 정도를 비교하기 위해 CALL을 세 가지 단계로 나누었는데, 첫 번째는 행동주의적 CALL(behavioristic CALL), 두 번째는 의사소통중심 CALL(communicative CALL), 세 번째는 통합적 CALL(integrative CALL or integrated CALL)이라고 한다(Dudeny and Hockly 2012). 행동주의적 CALL 시기에는 컴퓨터를 문서처리 기능이나 단순한 게임을 활용하여 제한된 내용을 반복하여 연습할 수 있는 도구로 활용했다면, 의사소통중심 CALL 시기에는 컴퓨터가 제한된 범위지만 학습자에게 피드백을 제시하거나 언어 산출을 도와 개별 활동을 촉진하는 튜터의 기능을 하였다. 통합적 CALL에서는 전 세계적으로 상호작용이 가능한 인터넷과 소셜미디어 발달 등의 발전에 힘입어 앞의 두 시기보다 컴퓨터가 학습자와 훨씬 더 많은 상호작용을 할 수 있게 되었다.

CALL 연구자들은 원어민과 대면하지 않아도 비슷한 상황을 만들 수 있는 사회적 환경을 조성하여 외국어 학습에 필요한 여러 가지 프로그램들을 개발하여 이를 특히 컴퓨터 매개 의사소통(computer mediated communication; CMC)라고 하였는데(Peterson 2009), 온라인 채팅프로그램, 이메일 교환, 게시판 등 대부분 쓰기 활동을 통해 학습자들의 영어 의사소통 능력을 향상시키고자 하였다. 다시 말하면, 원어민과의 대면 의사소통이 어려운 환경

에서 CMC 프로그램을 활용하는 것이 학습자들의 걱정을 낮추고 말하기 능력을 향상시킬 수 있는 가장 적절한 대안이라고 볼 수 있겠다(Lin 2014). 특히 요즘에는 모바일 기술의 발전으로 학습자들이 실제적이고 구체적 내용으로 즉각적인 피드백을 받으며 학습할 수 있는 환경이 조성되어서(Zimmerman and Howard 2013), 학습자들은 개별화된 학습(customized student learning)을 할 수 있고 언제 어디서나 학습에 참여할 수 있다는 장점이 있다.

2.2 영어 말하기 분야에서의 컴퓨터 활용

CALL에 대한 관심이 높아지면서 영어를 가르칠 때 컴퓨터를 활용하면 그렇지 않을 때보다 효과적이라는 연구 결과들이 많이 보고되고 있다(Beauvois 1992, Chappell 2001, Loewen and Erlam 2006, Sauro 2009, Smith 2003, 2004, Sullivan and Pratt 1996). 특히 CMC 프로그램이 원어민과의 대면 의사소통에 대한 적절한 대안이 될 수 있으므로(Lin 2014), 어떤 방식의 CMC 프로그램이 더 효과적일지 CMC 프로그램을 통해 어떤 과제를 제시하는 것이 좋을지 등에 대해 살펴보는 것도 필요하겠다. Lin(2014)은 2000년부터 2012년 동안 출간된 CALL 관련 논문들을 모두 찾아 비교한 후 25편의 논문을 메타분석한 결과, CMC를 활용할 때 원어민과 대면 대화할 때보다 학습자들의 외국어 말하기 능력이 더 효과적으로 향상되었으며(moderate and positive effect), 이 때 자연스러운 자료(naturalistic data)보다는 반응을 유도하는 자료(elicited data)가 더 긍정적인 결과를 초래하였으며, 구체적으로는 주어진 단락을 소리내어읽기와 같은 단순한 과제를 수행할 때 더 효과적이었다는 것을 알게 되었다. 또한 놀랍게도 CMC는 정확성(accuracy)이나 유창성(fluidity)에는 오히려 부정적 영향을 미친다는 것도 알게 되었다. 이러한 메타 분석 결과는 학습자들의 영어 말하기 능력을 향상시키기 위해 컴퓨터를 활용하고자 하는 교사나 연구자들에게 과제나 절차 등에 대한 변화가 있어야 한다는 점을 시사한다.

또한 원어민과의 대면 대화보다 컴퓨터를 활용한 CMC가 학습자들의 긴장과 걱정 정도를 낮추고 적극적인 참여를 유도할 수 있어 더 적합한 대화 환경을 제공한다고 한다. 학습자들의 긴장이나 걱정과 같은 정의적 요소는 나이 변인으로 인한 차이가 있을 것으로 예상되지만, 1990년부터 2012년까지의 CMC 연구를 메타 분석한 결과 모두 대학생(79%)과 고등학생(21%)을 대상으로 하였으며(Ziegler 2016) 영어 학습을 본격적으로 하게 되는 중학생을 대상으로 한 연구는 찾아보기 어려웠다. 또한 말하기(21%)보다는 쓰기(79%)를 대상으로 한 연구가 많았으며(Ziegler 2016), 측정 대상은 유창성, 정확성, 어휘사용정도, 말하는 속도, 문장구조의 복잡성 등이 주를 이루었다(Huang and Hung 2010, Volle 2005).

최근 기술의 발달로 학생들의 학습을 향상시키기 위해 학생 개개인에 맞추어(individualized), 적응형(adaptive) 학습 도구를 개발하고 있으며, 특히 어휘 학습 분야에서는 웹사이트 등을 활용한 연구들이 많이 수행되었다(Abraham 2008, Li 2010). 우리나라나

일본처럼 영어 원어민과의 접촉 빈도가 상대적으로 낮은 환경에서 컴퓨터나 모바일 기술을 활용한 학습 도구를 개발하여 긍정적인 효과를 보고하는 경우가 많다(Lee and Choi 2010, Lee and Kwon 2014).

교실에서 컴퓨터나 모바일 기술을 활용하는 경우는 학습자들에게 피드백을 즉각적으로 제공할 수 있기 때문이다. 특정 행동의 결과에 대해 제공하는 모든 정보인 피드백은 즉각적으로 제공했을 때 효과적이라고 하는데, 학습을 하는 동안 본인의 행동이 효과적인지 바로 알 수 있게 되어 불필요한 행동을 교정하는 등 학습스타일을 바꾸어 궁극적으로는 학습 성취도가 매우 높아진다고 하였으며(Zaborik 1987), 학습 행동에 문제가 없을 경우에도 즉각적인 피드백은 현 학습스타일을 지속적으로 활용하여 미래에도 긍정적인 효과를 도출할 수 있다고 한다. *Grammarly*와 같은 잘 알려진 소프트웨어를 활용한 CALL 방식의 피드백을 도입했을 때 듣기나 쓰기 영역에서도 효과가 있었지만, 영어 학습자들의 문법 능력이 크게 향상되어 문법 영역에 가장 효과가 있었다고 한다(Mehrgan 2010, Nutta 2001). 그러나 영어 학습자의 말하기 영역에 음성인식 피드백을 활용한 경우는 드물었다.

기술의 발달과 관련해서 중국에서는 언어학습은 아니어도 문자와 음성 변환 기술, 음성인식, 필체인식, 자연어 처리 등과 같이 인간과 컴퓨터가 좀 더 직관적인 상호작용을 할 수 있는 기술을 연구하고 있다. 미국 MIT에서는 사용자가 특별한 지식이 없어도 언어나 시각을 활용하여 자연스러운 상호작용을 할 수 있는 기술을 개발하고 있고, 구글이 음성인식 툴을 무료로 개방하고 있어 국내에서도 음성인식을 통해 직관적으로 기계와 대화할 수 있는 기술을 활발히 개발하고 있다. 따라서 음성인식과 같은 직관적 상호작용을 가능하게 하는 기술의 첨단적인 발전이 영어교육 분야에도 적극적으로 시도되어야 할 필요가 있겠다.

다시 말해, 컴퓨터를 활용한 영어 교육 분야에서 상대적으로 배제되어왔던 중학생들을 대상으로, 최근 무료로 개방된 음성인식 기술 등을 활용하여 긴장이나 걱정 정도는 낮추면서 직관적으로 원어민과 대면 대화를 하는 듯한 환경을 마련하고, CMC 분야에서 가장 효과적이었던 소리내어읽기 과제를 통해 학습자들의 단어 및 문장 발음을 진단 및 연습할 수 있는 프로그램을 개발하여 그 효과를 알아볼 필요가 있겠다. 이를 위해 본 연구는 다음과 같은 연구 질문을 설정하였다.

- 1) 우리나라 중학교 영어 학습자들은 음성인식 피드백 프로그램을 이용할 때 단어 및 문장 등의 소리내어읽기 과제를 어떻게 수행하는가?
- 2) 우리나라 중학교 영어 학습자들은 음성인식 피드백 프로그램을 이용할 때 단어 및 문장 등의 소리내어읽기 과제 수행 정도에 성적이나 성별 등의 차이가 있는가?
- 3) 우리나라 중학교 영어 학습자들은 음성인식 피드백 프로그램을 이용할 때 그림 묘사 자유 발화 과제를 어떻게 수행하는가?

3. 연구 방법

3.1 연구 참여자

서울에 위치한 중학교 3학년 총 46명이 본 연구에 참여하였다. 해당 학교의 영어 교사는 학습자들의 영어 발음 향상을 위한 프로그램 개발에 관심이 있어 참여하였으며, 해당 교사가 가르치는 학생들을 대상으로 음성인식 프로그램을 활용하여 단어, 문장, 문단을 소리내어 읽을 수 있도록 지도하였다. 총 2개 반 58명이 참여하였으나, 프로그램을 휴대폰에 설치하여 시행할 수 없었던 학생들을 제외한(학생의 휴대폰 버전이 너무 낮아 프로그램이 시행되지 않았고, 다른 학생들은 모두 휴대폰으로 음성인식을 시행하였으므로 환경이 달랐던 학생들의 결과는 포함하지 못하였다) 46명의 결과를 분석 대상으로 하였다.

본 연구에 참여하기 직전 학습자들이 1학기 중간고사를 본 결과가 있어 동의하에 성적 비교를 한 후, 상, 중, 하 세 그룹으로 나누어 비교하였으며 각 수준 및 성별에 따른 참여자의 수는 표 1과 같다.

표 1. 참여자 현황

	합	상	중	하
남성	22	10	4	8
여성	24	11	10	3
합	46	21	14	11

3.2 연구 도구

음성인식 프로그램(ET-Bot 음성버전, 그림 1)을 이용하여 소리내어읽기 과제를 수행하도록 하기 위해 여러 가지 소리를 포함(그림 2)하고 있는 문단(Weinberger 2015)을 골랐다. 이 문단은 미국의 표준 영어(General American English)를 기본으로 전 세계 영어 화자들의 말하기 억센트를 비교할 수 있도록 말하기 코퍼스를 구축하기 위해 개발되었다. 모든 참여자가 동일한 문단을 읽은 것을 발음기호로 표기하여 나이, 성별, 국적, 모국어, 참여 당시 나이, 영어 학습 시작 나이, 영어 학습 방법(학교 또는 일상생활), 영어권 국가 거주 기간 등에 따라 어떤 말하기 억센트를 보이는지 동일한 문단으로 비교할 수 있다. 그림 2에서 볼 수 있듯이 이 사이트에서 제시하는 문단은 대부분의 말하기 진단도구들이 목표로 하는 소리(Eisenberg and Hitchcock 2010)를 포함하고 있으므로, 본 연구에서도 동일한 문단을 활용하였다(그림 3).

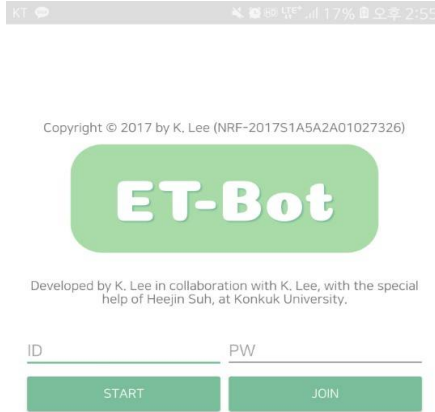


그림 1. ET-Bot(음성인식 피드백 프로그램) 휴대폰 실행 화면

single consonants		vowels	clusters	
initial	final		initial	final
k (3)	z (5)	i (12)	pl (2)	sk (1)
t (3)	l (4)	a (4)	st (4)	ŋz (2)
ð (6)	ŋ (1)	ε (4)	bj (2)	ks (1)
θ (3)	θ (1)	æ (10)	fj (3)	nz (2)
w (5)	m (1)	ɪ (11)	sp (1)	bz (1)
s (2)	ʃ (5)	ʌ (2)	sn (3)	nd (3)
f (3)	v (3)	e (10)	sl (1)	dz (1)
tʃ (1)	f (1)	u (5)	bl (1)	gz (1)
n (1)	k (4)	oo (3)	sm (1)	
b (3)	b (1)	aɪ (1)	sk (1)	
l (1)	d (2)	eɪ (5)	θj (1)	
ʃ (2)	g (2)	ɔ (3)	tj (1)	
d (1)	n (4)	ɔɪ (1)		
ʃ (1)	p (1)			
g (1)	t (2)			
m (2)				
h (4)				

그림 2. 소리내어읽기 과제에 포함된 소리 (괄호 안 숫자는 빈도. <http://accent.gmu.edu>에서 인용)

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

그림 3. 소리내어읽기 과제 단락(<http://accent.gmu.edu>에서 추출)

본 연구 참여자들은 음성인식 프로그램에서 제시하는 위의 문단을 소리내어읽기 전에, 학습자는 음성인식 프로그램에 익숙해지고 음성인식 프로그램은 각 학습자들의 발음을 제대로

인식하는지 훈련할 수 있는 기회를 제공받았다. 두 명의 영어교육 전공자가 기존 발음 진단 도구(e.g., *Goldman Fristoe Test of Articulation*) 등을 검토하고, 많이 쓰이는 발음 진단 도구 11개를 비교 분석한 결과(Eisenberg and Hitchcock 2010)에 기초하여 18개의 단어와 10개의 문장을 제시한 후, 목표 문단을 문장 단위로 끊어 읽을 수 있도록 하였다. 파일럿 테스트에서 문단 전체를 제시하고 소리내어읽기를 시도하였는데, 무선인터넷을 사용하는 프로그램의 특성상 긴 지문을 읽는 동안에 인터넷이 끊겨 음성인식이 제대로 되지 않는 경향이 있었다. 따라서 프로그램 운영의 효율성을 위해 문장 단위로 끊어 읽도록 제시하였다. 마지막으로 그림(그림 4)을 제시한 후 자유 발화를 하도록 하여 본 연구에 쓰인 음성인식 프로그램의 확장 가능성을 타진해보기로 하였다.



그림 4. 자유 발화 과제

참여자들은 음성인식 프로그램에서 제시하는 각 단어 및 문장들을 읽은 후 즉각적인 음성 인식 결과를 볼 수 있어, 제대로 인식이 되는 발음을 했을 경우(그림 5) 다음으로 넘어가고, 제대로 인식이 되지 않아 목표와 다른 발음의 단어나 문장이 제시되었을 경우(그림 6) 다시 소리내어읽기를 시도할 수 있는 기회를 제공받았다. 따라서 발화 시간도 분석의 대상에 포함하였다.

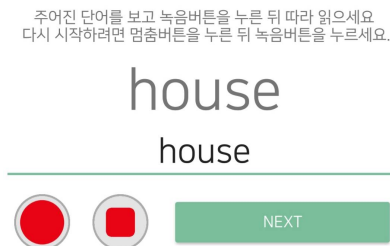


그림 5. 참여자가 발화한 발음이 맞았을 때 음성인식 화면

주어진 단어를 보고 녹음버튼을 누른 뒤 따라 읽으세요.
다시 시작하려면 멈춤버튼을 누른 뒤 녹음버튼을 누르세요.

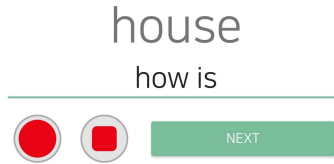


그림 6. 참여자가 발화한 발음이 틀렸을 때 음성인식 화면

3.3 자료 수집 절차

본 연구 참여자들에게 ET-Bot 앱을 활용하여 영어 단어 및 문장을 소리내어 읽을 수 있는지에 대해 설명을 하고 참여할 수 있도록 하였다. 영어 교사가 학습자들이 동시에 참여할 수 있도록 컴퓨터 실습실 이용 등 환경을 제공하였으나, 컴퓨터 실습실에서 동시에 녹음을 진행하기 어려운 상황이었어서 학생들의 휴대폰에 해당 프로그램을 다운받아 교사의 지시 하에 개별 녹음을 진행하였다.

참여자들은 주어진 단어와 문장을 보고 소리내어 읽은 후 즉각적으로 인식되어 제시되는 단어와 문장이 맞았을 경우(그림 5), 다음으로 넘어가기 버튼을 눌러 다음 과제를 하였다. 또는 그림 6과 같이 제대로 발음하지 못해 다른 표현으로 제시되었을 때 녹음 버튼을 다시 눌러 새로 시도를 할 수 있었다. 여러 번 발음을 해도 제대로 인식되지 못했을 경우 다음 과제로 넘어갈 수도 있으며, 이 때 시간과 함께 마지막 발음으로 인식된 표현이 저장되었다.

연습 단어와 문장이 끝나면 목표 문단이 나와 같은 방식으로 음성인식을 하였으며, 맨 마지막에는 그림(그림 4)이 제시되고 한 문장으로 그림에 대한 묘사를 하는 과제를 제시하여, 음성인식 프로그램으로 자유로운 대화에 대한 전사도 가능할지에 대한 가능성을 타진해보았다.

3.4 자료 분석 절차

ET-Bot은 학습자들의 발화 시간과 발음한 내용을 저장하도록 설계되어서 학습자들의 참여 후에 전사 과정이 필요 없다는 장점이 있었다. 발화 시간과 발음한 내용을 참여자별로 엑셀로 추출한 후 SPSS 18.0을 활용하여 비교하였다($p < .10$).

참여자마다 저장된 내용의 발음이 맞았는지 틀렸는지에 대한 분석은 학원에서 영어를 가르친 경력이 약 2년인 영어교직전공자와 본 연구 참여자들의 영어교사가 각자 하였으며, 비교한 결과 두 명의 결과는 전혀 다르지 않았다.

각 단어는 일치하면 1, 위의 그림 6에서처럼 문제는 'house'이었으나 저장된 산출물이 'how is'로 일치하지 않으면 0으로 코딩하였다. 문장이나 문단에서는 구성 단어들을 제시된 것과 산출된 것을 각각 비교하여 맞는 단어 수로 코딩하였다. 예를 들면 'He caught the big fish.'이라는 문장이 제시되었을 때 'he caught the big fish'라고 저장되어 있으면 다섯 개의 단어이므로 5, 'he caught the bick pish'라고 저장되어 있으면 3(he caught the)으로 코딩하였다. 이 때 음성인식 프로그램에서는 대소문자나 따옴표, 쉼표 등의 구두점은 인식되지 않으므로 이에 대한 구분 없이 채점하였다.

또한 참여자들의 영어 성적을 비교하기 위해 본 연구를 시작하기 전 시행한 1학기 중간고사 점수를 기준으로 활용하였다. 100점 만점에 평균이 74점이었는데 표준편차가 꽤 큰 편(24점, 최대값 100점, 최소값 8점)이어서 학습자들의 성적에 따라 비교하기로 하였으며, 학습자들의 성적을 모두 살펴보았는데 평균을 기준으로 상하그룹을 나눌 수 있는 구간 점수를 받은 학생들이 아무도 없어(중간고사에서 86-75, 60-46 구간의 점수를 받은 학생은 아무도 없었다), 상(100-87), 중(74-61), 하(45-0) 그룹으로 나누어 비교하였다(표 1).

이렇게 코딩된 숫자와 자동으로 저장된 발화 시간을 성별에 따라 비교할 때는 독립표본 t-검정을, 성적에 따라 비교할 때는 일원분산분석(ANOVA)을 이용하였다. 자유 표현 과제 등은 해당 산출물을 참여자에 따라 비교하였다.

4. 결과 및 논의

4.1 중학생 참여자들의 음성인식 프로그램 활용 결과

우리나라 중학생들은 즉각적인 피드백을 제공하는 음성인식 프로그램을 활용하여 소리내어 읽기 과제를 어떻게 수행하는지 알아보기 위하여 음성인식 프로그램 ET-Bot을 개발하여 알아보았다. ET-Bot은 기존의 발음 진단도구들과는 달리 음성인식 프로그램을 이용하여 즉각적인 인식 결과를 알아볼 수 있으므로, 학생에게도 교사에게도 매우 편리한 프로그램이라는 것을 알 수 있었다.

각 단어에 대한 발음을 ET-Bot이 제대로 인식하여 목표 단어와 같은 단어로 제시했을 경우 1, 다른 단어로 제시했을 때는 0으로 코딩한 결과 각 단어별 평균은 표 2와 같다. 최소값과 최대값을 함께 비교할 경우 흥미로운 결과를 알 수 있었다. 본 연구 참여학생들은 'fish, monkey, spider' 세 단어를 발음하는데 전혀 문제가 없었고(최저점이 1점으로 모두 맞게 발음함) 'duck, go, quack'은 상대적으로 잘 발음하지 못하는 경향이 있었으나, 전체적으로 주어진 단어를 발음하는데 큰 문제는 없어보였다($M=16.07$). 발음이 쉽다고 알려진 입술소리(bilabial m)를 포함한 'monkey'는 쉬울 것으로 기대했고, 이중모음으로 인해 'quack'은 발음이 어려울 것으로 기대했는데, 기대와 일치하는 경향을 보였다. 그러나 'fish'

와 같이 우리말에는 없는 발음(초성 f, 종성 sh)을 포함한 단어를 모든 학생들이 어려움 없이 발음했다는 것과, 상대적으로 어려울 것이라고 예상했던 모음 사이의 d(flap)를 포함한 'spider'는 잘 발음하고, 상대적으로 쉬울 것이라 예상했던 'duck'과 'go'는 오히려 잘 발음하지 못했다는 것이 흥미로웠다. 문맥이 없이 발음을 한 경우라 향후 비슷한 소리가 문장 내에서 어떻게 발음되는지 살펴보아야 할 것이다.

표 2. 각 단어 발음의 정확도와 발화 시간

목표 단어	N	발음의 정확도				발화 시간(초)			
		최소값	최대값	M	SD	최소값	최대값	M	SD
apple	46	.00	1.00	.96	.21	3.67	119.76	11.84	20.68
boy	46	.00	1.00	.96	.21	3.44	2250.80	57.25	330.73
cup	46	.00	1.00	.87	.34	3.38	138.75	22.66	31.12
door	46	.00	1.00	.83	.38	3.59	128.16	22.33	23.01
drum	46	.00	1.00	.85	.36	3.15	38.21	13.02	11.27
duck	46	.00	1.00	.70	.47	3.39	265.99	36.01	53.09
fish	46	1.00	1.00	1.00	.00	3.18	13.17	5.43	2.23
go	46	.00	1.00	.80	.40	3.35	116.35	21.02	26.70
hammer	46	.00	1.00	.93	.25	2.89	149.13	10.07	21.50
house	46	.00	1.00	.98	.15	4.07	176.53	25.64	33.99
monkey	46	1.00	1.00	1.00	.00	3.25	32.40	6.22	4.95
pig	46	.00	1.00	.93	.25	3.24	146.49	19.15	32.47
plate	46	.00	1.00	.89	.31	.80	47.74	10.62	11.71
quack	46	.00	1.00	.67	.47	4.61	213.61	43.46	39.00
spider	46	1.00	1.00	1.00	.00	2.83	50.25	8.02	8.92
table	46	.00	1.00	.91	.28	2.94	124.68	12.40	18.67
watch	46	.00	1.00	.93	.25	3.16	73.10	9.38	13.87
web	46	.00	1.00	.85	.36	2.94	342.80	28.60	52.69
all words	46	9.00	18.00	16.07	2.59	4.52	143.83	20.17	21.31

ET-Bot의 운영 방식에 대해 위에 언급한 것처럼 발음을 제대로 인식하지 못했을 경우(그림 6), 음성인식되어 제시된 단어가 목표 단어와 다른 것을 즉각적으로 확인하여 다시 시도해 볼 수 있는 기회를 주기 때문에 단어와 정확도 이외에도 단어의 발화 시간을 확인해 봐야 할 필요가 있었다. 표 2에 제시된 시간을 함께 살펴보면, 흥미로운 점은 학생들이 발음하기 쉬웠던 세 단어는 예상대로 발음 평균 시간도 짧았고(fish 5.43, monkey 6.22, spider 8.02), 발음하기 어려워했던 'duck(36.01)'과 'quack(43.46)'은 발음 평균 시간도 긴

편이었다. 그러나 'go'는 전체 단어 발음 평균(20.17)에 비추어 발음 평균 시간이 그리 길지는 않았고(21.02), 오히려 'boy(57.25), web(28.60)'과 같이 모음이나 유성자음으로 끝나는 경우가 여러 번 시도한 결과 발음 시간이 길었다. 특히 'web'은 우리나라 학생들이 발음할 경우 주로 'wep'으로 인식이 되어 다시 시도하는 경우가 많았으나, 'boy'의 경우 쉽게 접하는 단어이고 무성자음과 혼동되지 않는 상황임에도 놀랍게도 매우 긴 시간(최대값 2250.80)을 옹게 발음하려고 한 경우가 있었다. 왜 이런 현상이 있었는지에 대한 질적인 후속 연구가 더 필요할 것으로 사료된다.

정리하면, 학습자들은 모든 단어를 생각보다 긴 시간 동안 발화하여 대부분 단어를 옹게 발음하였는데, 이런 결과에 비추어보아 음성인식 프로그램에서 제공하는 즉각적인 피드백이 본인의 발음을 교정하는데 매우 효과적이었다고 판단된다.

문장의 발음을 알아보기 위하여 총 10개의 문장을 발음하도록 하였는데 그 결과는 표 3과 같다. 표 3에서 볼 수 있듯이 각 문장을 제대로 발음했을 경우, ET-Bot이 목표 문장과 같은 문장으로 인식하여 제시하였으며 각 문장의 최대값은 각 문장의 모든 단어를 모두 잘 발음한 학생들이 있다는 것을 알려준다. 또한 틀린 단어는 0으로 처리하여 문장마다의 합계 점수로 계산하였으며 최소값이 0인 경우는 한 단어도 제대로 발음하지 못했다고 볼 수 있다. 예를 들면, 첫 번째 문장인 'Don't sleep on the deck.'는 5개의 단어로 이루어져 있으므로 'dont sleep on the deck'처럼 음성인식할 수 없는 대소문자나 구두점을 제외하고 목표 문장대로 인식되어 제시되면 5점으로, 'dont slip on the deck'처럼 한 단어(slip)가 틀리게 인식되어 제시되면 4점으로 코딩한 것을 분석한 결과이다.

표 3. 문장 발음의 정확도와 발화 시간

목표 문장	N	발음의 정확도				발화 시간(초)			
		최소값	최대값	M	SD	최소값	최대값	M	SD
S1(5)	46	2.00	5.00	3.96	.94	5.83	259.90	37.54	43.23
S2(6)	46	.00	6.00	5.26	1.12	1.88	67.80	17.83	15.02
S3(5)	46	.00	5.00	3.47	1.24	4.58	684.32	54.87	107.92
S4(5)	46	.00	5.00	4.33	1.02	.94	124.90	22.59	21.76
S5(4)	46	.00	4.00	3.07	1.48	4.01	112.72	21.86	20.26
S6(4)	46	3.00	4.00	3.56	.50	4.09	125.90	18.24	22.01
S7(8)	46	1.00	8.00	6.76	1.48	5.77	117.97	24.33	22.68
S8(4)	46	2.00	4.00	3.60	.65	3.97	260.08	15.08	38.06
S9(4)	46	.00	4.00	3.76	.74	4.27	61.36	16.96	14.80
S10(5)	46	.00	5.00	3.96	1.40	4.23	260.34	24.48	39.65
All(50)	46	22.00	50.00	41.69	6.98	8.41	207.64	33.97	33.72

Note. S1: Don't sleep on the deck. S2: The man will be here soon. S3: He sat on his cot. S4: He caught the big fish. S5: The sign says 'Pool'. S6: Look at the sheep. S7: He was hurt when he hit the rack. S8: I need a cup. S9: He said he would. S10: Use that for the vase.

평균 점수가 상대적으로 낮고 발화 시간이 상대적으로 긴 문장(10 문장을 모두 말한 평균 발화 시간보다 한 문장의 발화 시간이 더 긴 경우)은 표 3에서 볼 수 있듯이 S1(Don't sleep on the deck.), S3(He sat on his cot.), S8(I need a cup.), S10(Use that for the vase.)이었다. 첫 번째 문장에서는 동사의 장모음을 단모음으로 발음하거나(slip), 'deck'을 'back, dack, bed, desk' 등으로 다르게 발음한 경우가 많았다. 여덟 번째 문장은 'cup' 대신 'cop, call' 등으로 잘못 발음하였고, 특히 세 번째 문장은 앞의 두 문장에서처럼 모음 구분을 명확히 하지 못하는 우리말 특성으로 인해 'cot'를 'cut, coat, cup, cock, court, quote, hot' 등으로 발음하였다. 마지막 문장은 다른 문장들과는 달리 우리에게 익숙하지 않은 자음(v)을 발음하는 걸 어려워하였다('vase' 대신 'face, base, best' 등으로 잘못 발음). 다시 말해 'f'와 조음 방법이 같은 'v' 소리는 상대적으로 중학생들에게 많이 쓰이지 않고 강조되지 않는 경향이 있어, 쉽게 발음하였던 'fish'와는 달리 'vase'를 어려워한 점이 흥미로웠다.

정리하면, 참여 학생들은 우리말 발음에서 명확히 구분하지 않는 모음이 끝소리 무성 자음과 함께 있는 경우(deck, cot, caught, hurt, rack, cup 등)에 발음을 어려워하는 경향이 있었다. 이에 대해서는 후속 연구를 통해서 해당 모음과 무성 자음 및 유성 자음 끝소리를 함께 조합하여 발음을 비교해 볼 필요가 있겠다.

마지막으로, 표준화된 발음 비교 문단(Weinberger 2015)을 어떻게 발음하는지 알아보았다. 위에서 언급한대로 이 문단은 여러 가지의 소리를 포함하고 있어 다양한 참여자에 따라 액센트를 비교해 볼 수 있도록 만든 것이며, 따라서 앞의 단어나 문장들보다 학습자들이 어려워하는 발음이 상당 수 있었다. 표 4에서 볼 수 있듯이 여러 번 시도를 하여 발화 시간이 늘어났으며, 그럼에도 불구하고 잘못 인식된 단어들도 많이 있었다. 문단에서는 오히려 'thick, train'처럼 첫소리 발음을 잘 못하거나, 'slab, Bob, frog, bag'처럼 유성 자음이 끝소리로 왔을 때 무성 자음으로 발음하는 등 자음의 발음에 어려움이 많았다. 단어, 문장, 문단의 발화 시간과 정확도를 함께 보았을 때, 참여 학생들은 즉각적으로 제공된 피드백을 보고 스스로 옳게 발음하기 위해 여러 번 시도하여 연습을 한 것을 알 수 있었다. 이는 본 프로그램이 스스로 영어를 읽어보고 연습해볼 수 있는 동기를 부여하는 환경을 제공하기에 적절하다는 가능성을 보여준 것이라고 볼 수 있겠다.

표 4. 단락 발음의 정확도와 발화 시간

목표 문장	N	발음의 정확도				발화 시간(초)			
		최소값	최대값	M	SD	최소값	최대값	M	SD
S11(3)	46	.00	3.00	2.93	.45	3.89	82.14	11.41	12.54
S12(11)	46	5.00	11.00	10.40	1.44	7.09	120.31	21.02	19.17
S13(20)	46	.00	20.00	13.25	6.20	10.56	717.75	92.73	143.99
S14(15)	46	.00	15.00	12.86	3.39	7.72	111.63	29.72	23.80
S15(20)	46	1.00	20.00	15.61	5.03	11.01	1714.02	96.99	260.84
All(69)	46	11.00	69.00	55.05	13.70	46.26	2068.05	251.87	357.58

Note. S11: Please call Stella. S12: Ask her to bring these things with her from the store. S13: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. S14: We also need a small plastic snake and a big toy frog for the kids. S15: She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

4.2 성적과 성별에 따른 음성인식 프로그램 활용 결과

본 연구에 참여한 우리나라 중학교 영어 학습자들의 소리내어읽기 과제 수행 정도가 성적이나 성별 등에 따라 차이가 있었는지 알아보았다. 본 연구에서는 성적이나 성별이 다른 중학생 집단이 음성인식 프로그램을 이용하여 소리내어 읽었을 때 영어 단어 및 문장의 발음과 발화 시간에 차이가 있었는지도 알아보았다.

우선 참여자들의 성별에 따른 발음 정도를 비교하기 전에 참여자들의 중간고사 영어 성적을 성별에 따라 비교한 결과, 남녀의 성적 차이는 유의미하지 않았다(남: 22명, $M=68.32$, $SD=29.25$ / 여: 24명, $M=79.21$, $SD=19.39$, $t=1.50$, $df=44$, $p=.141$). 통계적으로 유의미한 차이는 없었으나 표 1에서 살펴본 것처럼 성적에 따른 남녀의 분포가 매우 다른 것 주의 깊게 살펴보아야 할 것이다. 다시 말해, 영어를 잘 하는 학생들의 경우, 남녀의 수가 거의 비슷한데(상: 남 10명, 여 11명), 중간 성적의 학생들(남 4명, 여 10명)과 영어를 잘 못하는 학생들(하: 남 8명, 여 3명)의 분포에서 남자와 여자가 매우 다른 양상을 보이는 것을 알 수 있다. 물론 참여자들의 수가 너무 적어서 통계적으로 유의미한 결과가 나오지 않았을 수 있으나 여학생들의 경우 중간이나 잘하는 학생들이 대부분이며, 남학생들의 경우 표준편차가 큰 것으로도 또한 알 수 있듯이 중간 성적의 남학생들은 거의 없었다.

이런 분포의 차이가 발음을 하는데 어떤 차이를 만드는지 여부를 확인하기 위해, 통계적으로 남녀의 성적 차이가 없었음에도 불구하고 각 단어나 문장, 목표 문단을 발음하는데 남녀 차이가 있는지 알아보았다. 표 5에서 볼 수 있듯이 총 18개의 단어에 대한 발음은 영어 성적이 낮은 하 그룹($t=1.94$, $df=8.84$, $p=.084$)을 제외하고는 모두 성별에 따른 차이는 없었다. 다시 말해, 여학생보다 남학생의 수가 더 많았던 하 그룹의 경우 남학생들은 여학생들보다 평균 3개 정도의 단어를 발음하지 못했다.

표 5. 성별에 따른 단어 발음 차이

	성별(N)	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p.</i>
모든 학생	남(22)	15.45	2.84	1.56	44	.126
	여(24)	16.63	2.24			
상	남(10)	15.70	2.21	1.20	19	.245
	여(11)	16.91	2.39			
중	남(4)	17.25	1.50	.79	12	.444
	여(10)	16.20	2.44			
하	남(8)	14.25	3.65	1.94	8.84	.084
	여(3)	17.00	1.00			

참여자 수가 많지는 않지만 실제로 각 단어를 비교해봤을 때 상과 중 그룹에서는 성별 차이가 없었다. 하 그룹에서 여학생들이 남학생들보다 더 잘 발음한 것은, 다시 말해 모든 하 그룹의 여학생들은 발음했을 때 인식이 잘 되었으나, 남학생들은 반 정도만 제대로 발음할 수 있었던 것은 'cup(남 .63, 여 1.00, $t=2.05$, $df=7.00$, $p=.80$), duck(남 .50, 여 1.00, $t=2.65$, $df=7.00$, $p=.033$), quack(남 .38, 여 1.00, $t=3.42$, $df=7.00$, $p=.011$), web(남 .63, 여 1.00, $t=2.05$, $df=7.00$, $p=.084$)'이었다.

각 단어의 발음 시간을 성별에 따른 비교를 해 보았을 때, 'plate'만 여학생들이 남학생들보다 긴 시간 동안 연습하여 발음하였다(남 7.06, 여 13.87, $t=2.11$, $df=27.36$, $p=.044$). 성적별로 비교해보았을 때는 상과 하 그룹에서는 성별 차이가 없었으며, 중 그룹에서 'hammer(남 4.03, 여 8.37, $t=3.30$, $df=9.85$, $p=.008$), spider(남 4.95, 여 9.23, $t=1.84$, $df=11.70$, $p=.091$), table(남 4.51, 여 15.28, $t=3.01$, $df=9.06$, $p=.015$)'을 여학생들이 더 많이 연습하여 발음한 것을 알 수 있었다. 그러나 전체적으로 단어를 발음하는 정확도와 발화 시간에서 성적이나 성별에 따른 특별한 차이가 있지는 않았다.

단어의 발음과 발화 시간에서 남녀 차이가 없었던 상 그룹이 문장의 발음에서는 S10에서만 차이를 보였다. 다시 말해, 'vase'가 포함되어 있어 전체적으로 발음하기 어려워졌던 마지막 문장(남 3.80, 여 4.73, $t=1.90$, $df=19$, $p=.073$)은 여학생들이 해당 문장을 옹계 발음하기 위해 더 많은 시간을 들였다는 것을 알 수 있었다. 문장 발화 시간은 상과 중 그룹은 별 차이가 없었고, 하 그룹에서만 'cot'가 포함된 세 번째 문장(남 16.86, 여 33.74, $t=1.85$, $df=9$, $p=.097$)을 잘 발음하기 위해 여학생이 더 많은 시간을 들였다. 문장 발음과 발화 시간에서는 여학생들이 더 많이 노력했다는 모습을 보였다.

또한 문단에 대해서도 성별이나 성적에 따른 차이가 거의 없었다. 네 번째 문장(We also need a small plastic snake and a big toy frog for the kids.)만 역시 남자보다 여자가 더 긴 시간을 들여 연습을 한 것만 차이가 있었다(남 23.58, 여 35.59, $t=1.71$, $df=31.54$, $p=.098$).

다시 말해, 단어, 문장, 문단의 발음과 발화 시간에 대해 성적을 기준으로 성별에 따른 차

이가 있는지 알아보았을 때, 여학생들이 남학생들보다 많은 시간을 들여 옳게 발음이 되어 목표 단어나 문장이 인식되도록 하는데 노력을 한 경우들을 볼 수 있었다. 이로 미루어보아 여학생들이 영어를 더 잘한다는 속설이 실제 성적 차이로 반영이 된다고보다는 더 잘하기 위해 노력하는 모습을 여학생들에게서 더 자주 볼 수 있었던 것에 기인하는 것으로 생각할 수 있겠다. 성별에 따라 영어를 대하는 태도나 정의적 측면에 어떤 차이가 있는지에 대한 후속 연구가 필요하다고 생각된다.

마지막으로 학습자들의 성적별 차이가 있는지 알아본 결과, 단어나 문장의 발음 및 발화 시간에서는 별 차이가 없었고, 문단의 발음에서만 상, 중, 하 그룹의 차이가 유의미하게 나타난 것이 있었다. 발음의 정확도에서는 문단의 네 번째 문장(상 13.65 > 중 13.31 > 하 10.91, $F(2, 43)=2.67, p=.081$)과 다섯 번째 문장(상 17.60 > 중 15.00 > 하 12.73, $F(2, 43)=3.95, p=.027$)이 성적별로 유의미한 차이가 있었다. 또한 이제까지 그룹별 차이가 전혀 없었던 문단의 첫 번째 문장인 'Please call Stella.' 발화 시간이 하 그룹이 가장 길었고 상 그룹이 가장 짧았다(상 7.67 < 중 10.49 < 하 19.64, $F(2, 43)=3.76, p=.032$)는 점도 흥미로웠다. 문단의 문장들이 더 길어서 참여 학생들이 발음을 하기 어려워하여 발음의 정확도와 발화 시간에서 기대했던 성적별 차이가 몇 가지 나타날 수 있었던 것 같다. 후속 연구에서 이처럼 긴 문장을 제시하여 소리내어읽어 볼 수 있도록 한 직후, 참여 학생들과의 인터뷰를 진행하여 여러 번 읽어서 연습을 해야 했던 문장은 어떤 요소 때문에 그런지를 알아 볼 수 있으면 더 유의미할 것으로 기대된다.

4.3 자유 발화

본 연구에서는 기존 연구(Lin 2014)에서 효과적이었던 소리내어읽기 과제를 중심으로 진행하였지만 음성인식 프로그램의 확장 가능성을 타진해보기 위해 그림을 묘사하는 자유 발화 과제를 마지막에 추가하였다. 정지(stop) 표지판을 보고 한 문장으로 설명하라고 했는데, 참여 학생들은 생각보다 긴 문장을 발화하지는 않았으나, 학생들의 영어 수준 등을 진단하기에 효과적이라고 생각되었다. 예를 들면 성적이 낮은 학생들은 'stop sign can see', 'stop must', 'red sign see', 'station is stop', 'stop is'처럼 해당 학생에게 어떤 수업이 필요할 지에 대한 정보를 주는 발화를 하였다. 이 학생들은 우리말의 발화 순서대로 영어 단어를 나열하고 있어 교사들이 이를 보고 해당 학생들의 수업을 계획하는데 도움이 될 수 있겠다.

그리고 생각보다 많은 학생들이 'stop'이라고만 발화하여 그림을 묘사하는 문장 발화에 어려움을 겪고 있다는 것을 알 수 있었다. 교사들은 수업 중에 문장 형태로 발화하거나 더 많은 양을 발화하도록 이끌어낼 수 있는 과제를 제공하고, 본 연구의 ET-Bot을 통해서도 그림 묘사 과제를 추가하여 학생들이 보다 편하게 문장 형태로 발화할 수 있는 연습 기회를 제공하는 것이 바람직할 것으로 사료된다. 영어를 잘하는 학생들은 'the sign says stop',

'you must stop here', 'it means stop', 'I can see the stop sign' 등으로 짧지만 자연스러운 발화를 하여 이 학생들을 위한 적절한 추가 과제를 제공하여 실력 향상을 도모해야 할 것이다. 그리고 흥미롭게도 한 학생만 'there is a stop sign and sky is blue'하고 정지 표지판과 함께 배경인 하늘도 묘사하였는데, 다른 학생들은 배경 정보에 대해 어떤 생각이 들었는지, 왜 그 학생만 배경 정보까지 포함했는지 등에 대해 알아보는 것도 필요하겠다. 또한 학생은 정지 표지판에 대한 언급 없이 'blue sky'라고만 하였는데 왜 배경 정보만 묘사했는지에 대한 인터뷰도 필요할 것으로 보인다.

본 연구는 소리내어읽기 과제를 위주로 참여 중학생들이 음성인식 프로그램을 활용하여 단어, 문장, 문단을 어떻게 발음하는지, 남녀별, 성적별 차이는 있는지를 알아보기 위해 계획되었지만, 발화 직후에 발음에 대한 즉각적인 피드백을 줄 수 있는 음성인식 프로그램의 확장 가능성을 알아보기 위해 추가한 그림 묘사 자유 발화 과제가 참여 학생들의 영어 수준 진단 및 수업 설계에 도움을 줄 수 있다는 가능성을 볼 수 있었다는 것에도 의의가 있다.

5. 결론

본 연구는 우리나라 중학교 영어학습자를 대상으로 컴퓨터를 활용한 영어 말하기 연습 프로그램을 개발하기 위해 계획되었다. 최근 급속한 기술의 발달로 음성인식 툴이 무료로 공개되어 있어 이를 활용하여 학습자들이 개별적이고 즉각적인 피드백을 받을 수 있는 프로그램의 개발이 필요하다고 사료되었다. 개별적이고 즉각적인 피드백이 학습자에게 긍정적인 학습 효과를 유발하기 때문에(Zaborik 1987) 이를 효과적으로 제공하기 위해서는 컴퓨터를 활용하는 것이 필요하였다. 또한 컴퓨터를 활용한 영어 학습 프로그램에 관한 연구에서는 컴퓨터를 활용해야 한다는 요건 때문에 주로 대학생이나 고등학생을 대상으로 하고 있어, 더 어린 연령대의 중학생을 대상으로 그 효과성을 타진해보는 것이 필요하였다. 컴퓨터를 이용하여 원어민과 쉽게 대면할 수 없는 교실에서도 원어민과 의사소통하는 듯한 환경을 만들어주기 위한 프로그램(CMC 등)들은 영어 학습자들의 산출 능력 향상에 효과적이라고 하였는데(Peterson 2009), 대부분은 전사 과정이 필요 없는 산출 능력인 쓰기에 대한 연구였으며(Ulsoy 2006), 컴퓨터를 활용한 피드백은 학습자의 문법 능력에 긍정적인 영향을 미친다고도 하였다(Mehrgan 2010, Nutta 2001). 그러나 중학생을 대상으로 음성인식 프로그램을 활용하여 즉각적인 피드백을 제공함으로써 영어 학습자들의 말하기 관련 능력이 향상되는지에 대한 연구는 상대적으로 드물었다.

따라서 본 연구를 통해 음성인식 피드백 프로그램을 활용하여 학습자들에게 가장 효과적이었던 컴퓨터 과제인 소리내어읽기(Lin 2014)를 제시하여 우리나라 중학교 영어 학습자들이 어떻게 수행하는지, 성별이나 성적에 따라서 차이가 있는지를 알아보았다. 결과적으로 본 연구에서 개발한 음성인식 피드백 프로그램인 ET-Bot을 이용함으로써, 연구자나 교사

입장에서는 말하기 연구에서 가장 소모적인 과정인 전사 과정을 음성인식이 대신할 수 있어 말하기 연구에 대한 새로운 방법을 제시하였다 볼 수 있으며, 학습자들의 입장에서는 단어 및 문장에 대한 발화 피드백을 즉각적으로 제공받아 옳은 발음으로 인식되고자 충분히 노력하도록 하는 동기 부여 효과가 있다고 판단되었다. 다시 말해, 대부분의 학습자들이 맞는 발음으로 단어나 문장들을 발화하였는데, 발화된 시간을 보면, 해당 단어나 문장을 읽는 일반적인 발화 시간보다 긴 것을 알 수 있었는데, 이를 통해 본인이 발화한 것이 잘못 인식되어 화면에 제시되면 다음으로 즉각 넘어가기 보다는 여러 번 다시 시도하여 맞는 발화로 인식되어 제시될 수 있도록 노력하였다는 것을 알 수 있었다. 이런 점 때문에 성별이나 성적에 따라 발음의 정확성 등에 큰 차이는 없었으며, 이것이 개별화된 학습을 촉진하는데 필요한 요소라고 기대된다.

성별에 따른 결과 차이가 크지는 않았지만, 상대적으로 길고 어려운 발음을 포함한 문단 읽기 과제에서는 여학생이 더 오랜 시간 정확한 발화를 하기 위해 노력했던 것으로 보아, 영어를 대하는 태도나 정의적 측면에 대하여 성별에 따른 질적 연구 등이 필요할 것으로 보인다.

또한 소리내어읽기 과제가 주된 과제였지만 그림 등으로 자유 산출을 독려하는 상황에서 학습자들은 본인의 발화에 대해 음성인식된 피드백을 어떻게 활용할 수 있을지 등을 알아보기 위하여, 본 연구에서 개발된 ET-Bot의 확장 가능성을 알아볼 수 있는 자유 산출 과제를 추가하였다. 같은 그림을 보고 'the sign says stop'과 같이 자연스러운 문장을 구사하거나 'stop sign can see'처럼 우리말 어순으로 단어를 나열하는 등 학습자들의 현재 영어 말하기 능력에 대한 진단을 할 수 있는 가능성이 보였다. 교사들은 자유 산출 과제 결과를 참고하여 학습자들의 현재 상황에 맞는 교수 계획을 세울 수 있고, 학습자들은 본인이 자유롭게 발화한 표현들이 어떻게 인식되어 제시되는지 확인할 수 있어 흥미를 유발하고 궁극적으로는 학습 성취도를 향상시킬 수 있을 것으로 기대된다.

본 연구에서 개발된 ET-Bot의 타당도 검증을 위하여 비슷한 대상의 학생들에게 같은 절차로 수행할 수 있는 후속 연구가 필요하다고 사료된다. 또한 ET-Bot과 비슷한 환경으로 원어민과의 대면 말하기 연습을 함께 실시하여 그 효과성을 비교해보는 것도 필요하겠다. 본 연구에서 문단으로 제시하려고 했던 표준 문단(Weinberger 2015)이 음성인식 되기에 너무 길어 문장으로 끊어 제시한 것이 문단 읽기와 비교하여 어떤 차이를 유발할지에 대한 후속 연구도 필요할 것이다.

참고문헌

- Abraham, L. (2008). Computer-mediated glosses in second language comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*

- 21, 199–226.
- Beauvois, M. H. 1992. Computer-assisted classroom discussion in the foreign language classroom: Conversation in slow motion. *Foreign Language Annals* 25, 455–464.
- Beauvois, M. H. 1998. Conversations in slow motion: Computer-mediated communication in the foreign language classroom. *The Canadian Modern Language Review* 54(2), 198–217.
- Bernhardt, B. H. and G. Holdgrafer. 2001. Beyond the basics I: The need for strategic sampling for in-depth phonological analysis. *Language, Speech, and Hearing Services in Schools* 32, 18–27.
- Chappell, C. 2001. *Computer Application in Second Language Acquisition: Foundations for Teaching, Testing, and Research*. New York: Cambridge University Press.
- Childs, M. 2010. *Learners' Experience of Presence in Virtual Worlds*. Doctoral dissertation, University of Warwick, UK.
- Dudeny, G. and N. Hockly. 2012. ICT in ELT: How did we get here and where are we going? *ELT Journal* 66(4), 533–542.
- Ehrman, M. E. and Oxford, R. 1995. Cognition plus: Correlates of language learning success. *The Modern Language Journal* 79, 67–89.
- Eisenberg, S. L. and E. R. Hitchcock. 2010. Using standardized tests to inventory consonant and vowel production: A comparison of 11 tests of articulation and phonology. *Language, Speech, and Hearing Services in Schools* 41, 488–503.
- Hewitt, E. and J. Stephenson. 2011. Foreign language anxiety and oral exam performance: A replication of Phillip's *MLJ* study. *The Modern Language Journal* 96(2), 170–189.
- Huang, H.-T. D. and S.-T. A. Hung. 2010. Effects of electronic portfolios on EFL oral performance. *Asian EFL Journal* 12(2), 192–212.
- Kern, R. G. 1995. Restructuring classroom interaction with networked computers: Effects on quantity and characteristics of language production. *The Modern Language Journal* 79(4), 457–476.
- Klein, H. B. 1984. Procedures for maximizing phonological information from single-word responses. *Language, Speech, and Hearing Services in Schools* 15, 267–274.
- Krashen, S. D. 1981. *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.
- Lee, J. and S. Choi, 2010. The effects of using multimedia title on preschool children's English listening skills, vocabulary, story recall ability and affective domains.

- Multimedia-Assisted Language Learning* 13(3), 237–252.
- Lee, K. R. and S. Kwon. 2014. Effects of vocabulary memorizing tools on L2 learners' vocabulary size. *The Journal of Asia TEFL* 11(2), 125–148.
- Li, J. 2010. Learning vocabulary via computer-assisted scaffolding for text processing. *Computer Assisted Language Learning* 23, 253–375.
- Lin, H. 2014. Computer-mediated communication (CMC) in L2 oral proficiency development: A meta-analysis. *ReCALL* 27(3), 261–287.
- Loewen, S. and R. Erlam. 2006. Corrective feedback in the cat room: An experimental study. *Computer Assisted Language Learning* 19(1), 1–14.
- MacIntyre, P. D. and R. C. Gardner. 1991. How does anxiety affect second language learning? A reply to Sparks and Ganschow. *The Modern Language Journal* 79(1), 90–99.
- McCorkle, B. and J. Palmeri. 2016. Lessons from history: Teaching with technology in 100 years of *English Journal*. *English Journal* 105(6), 18–24.
- Mehrgan, K. 2010. Computer-assisted language learning: A panacea for grammar development. *Advances in English Linguistics* 1(2), 25–29.
- Nutta, J. 2001. *Is Computer-based Grammar Instruction as Effective as Teacher-directed Grammar Instruction for L2 Students?* University of South Florida: USA.
- Peterson, M. 2009. Learner interaction in synchronous CMC: A sociocultural perspective. *Computer Assisted Language Learning* 22(4), 303–321.
- Saito, H. 1994. Teachers' practices and students' preferences for feedback on second language writing: A case study of adult ESL learners. *TESL Canada Journal* 11(2), 46–70.
- Sauro, S. 2009. Computer-mediated corrective feedback and the development of second language grammar. *Language Learning & Technology* 13, 96–120.
- Sharples, M., I. Arnedillo-Sanchez, M. Milrad and G. Vavoula. 2009. Mobile learning: Small devices, big issues. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder and S. Barnes, eds., *Technology-enhanced Learning: Principles and Products*, 233–249. Dordrecht, Netherlands: Springer.
- Smith, B. 2003. Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal* 87(1), 38–57.
- Smith, B. 2004. Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition* 26(3), 365–398.
- Sullivan, N. and E. Pratt. 1996. A comparative study of two ESL writing environments:

- A computer assisted classroom and a traditional oral classroom. *System* 29, 491–501.
- Ulusoy, M. 2006. The role of computers in writing process. *The Turkish Online Journal of Educational Technology* 5(4), 58–66.
- Volle, L. M. 2005. Analyzing oral skills in voice e-mail and online interviews. *Language Learning & Technology* 9(3), 146–163.
- Weinberger, S. 2015. *Speech Accent Archive*. George Mason University. Retrieved from <http://accent.gmu.edu>
- Wu, W. and M. Marek. 2009. The impact of teleconferencing with native English speakers on English learning by Taiwanese students. *International Journal on E-learning* 8(1), 107–126.
- Xu, Y. 2018. Construction of a multiple English teaching mode based on cloud technology. *iJET* 13(8), 239–253.
- Young, D. J. 1999. *Affect in Foreign Language and Second Language Learning: A Practical Guide to Creating a Low-anxiety Classroom Atmosphere*. Boston: McGraw-Hill.
- Zahorik, J. A. 1987. Reacting. In M. J. Dunkin, ed., *International Encyclopedia of Teaching and Teacher Education*, 416–423. Oxford: Pergamon Press.
- Ziegler, N. 2016. Synchronous computer-mediated communication and interaction. *Studies in Second Language Acquisition* 38, 553–586.
- Zimmerman, S. and B. Howard. 2013. Implementing iPads into K–12 classrooms: A case study. In R. McBride and M. Searson, eds., *Proceedings of Society for Information Technology & Teacher Education International Conference 2013*, 2512–2516. Chesapeake, VA: AACE.

예시언어(Examples in): 영어(English)
 적용가능 언어(Applicable Languages): 영어(English)
 적용가능 수준(Applicable Level): 중등(Secondary)

이경량(Lee, Kyoung Rang)
 교수(Professor), 영어영문학과(Dept. of English language and literature)
 세종대학교(Sejong University)
 서울특별시(Seoul, 05006, Korea)
 광진구 군자동 98(Gunjadong 98, Gwangjingu)
 Tel: 02) 3408–3118
 E-mail: kranglee@sejong.ac.kr

논문 접수(Received): 2018년 10월 30일

논문 수정(Revised): 2018년 12월 12일

게재 확정(Accepted): 2018년 12월 16일