Journal of The Korea Society of Computer and Information
Vol. 24 No. 2, pp. 9-16, February 2019
www.ksci.re.kr
https://doi.org/10.9708/jksci.2019.24.02.009

A Computer-Assisted Pronunciation Training System for Correcting Pronunciation of Adjacent Phonemes

Jaesung Lee*

Abstract

Computer-Assisted Pronunciation Training system is considered to be a useful tool for pronunciation learning for students who received elementary level English pronunciation education, especially for students who have difficulty in correcting their pronunciation in front of others or who are not able to receive face-to-face training. The conventional Computer-Assisted Pronunciation Training system shows the word to the user, the user pronounces the word, and then the system provides phoneme or audio feedback according to the pronunciation of the user. In this paper, we propose a Computer-Assisted Pronunciation Training system that can practice on the varying pronunciation according to positions of adjacent phonemes. To achieve this, the proposed system is implemented by recommending a series of words by focusing on adjacent phonemes for simplicity and clarity. Experimental results showed that word recommendation considering adjacent phonemes leads to improvement of pronunciation accuracy.

▸Keyword: Educational Data Mining, Computer-Assisted Pronunciation Training, Automatic Feedback, Word Recommendation, Adjacent Phonemes

I. Introduction

컴퓨터 보조 발음 교육(Computer-Aided Pronunciation Training, CAPT) 시스템이 일대다 교육 환경의 단점을 크게 줄일 수 있기 때문에 CAPT 시스템은 교육계에서 많은 관심을 받고 있다 [1]. 예를 들어, 아시아에서 영어 발음 교육을 받는 학생은 다른 학생 앞에서 발음을 수정하는 것을 두려워하는 경향이 있는데 [2,3], 이러한 상황에서 CAPT 시스템은 학생들에게 스트레스 없는 환경을 제공하는 데 도움이 되기 때문에 좋은 대안이 될 수 있다 [4]. 또한 CAPT 시스템을 통해 학생들은 개인 교사와의 직접 대면 훈련 시간을 찾는 데 어려움을 또한 피할 수 있다 [5]. 한편, 컴퓨터 보조 발음 교육은 인공지능의 한 세부 분야인 전문가 시스템과 밀접한 연관관계가 있으며, 음성 인식, 음성 합성, 음운론 등의 연구를 수반하므로 인공지능 인문학의 발전에도 크게 기여할 수 있다 [6].

일반적인 CAPT 시스템을 통한 발음 연습은 CAPT 시스템이 사용자들이 발음할 테스트 단어 나 문장을 표시하고 테스트 단어의 발음에 필요한 음소 및 음성과 같은 피드백을 제공하여 발음 기술을 향상시킨다. 이 때, 피드백은 CAPT 시스템이 학생의 수용할 수없는 발음을 교정할 수 있는 유일한 기회이므로 효과적인 CAPT 시스템은 피드백 생성 시 수용할 수 없거나 잘못된 발음의 원인을 정확하게 지적할 수 있어야 한다 [7,8]. 또한, 교정을 원하는 잘못된 발음 습관 등과 같은 사용자의 요구사항에 관계없는 무차별적인 피드백은 학습 의욕과 교육 자체의 효율성을 현저하게 떨어뜨릴 수 있기 때문에 특정 유형의 오류에 초점을 맞추는 것이 효과적인 발음 교육을 위한 좋은 전략이 될 수 있다 [4,9].

기존의 CAPT 시스템에서는 피드백 생성이 실제 영어 발음 학습의 전략을 모방하여 구현된다. 이 때, 발음 학습의 방식은 대략 두 가지로 나눌 수 있는데, 하나는 음소 기반 학습(Phonics training)이고 둘째는 전체 단어 기반 학습(Whole

• First Author: Jaesung Lee, Corresponding Author: Jaesung Lee
*Jaesung Lee (curseor@cau.ac.kr), School of Computer Science and Engineering, Chung-Ang University
• Received: 2018. 11. 14, Revised: 2019. 01. 18, Accepted: 2019. 01. 19.
• This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A6A3A01078538).

word training)이다 [10]. 음소 기반 학습은 음소 단위의 오류를 교정하는 것을 주요 목표로 하여 음소 강조 접근법으로 알려져 있다 [11]. 한편, 전체 단어 훈련은 학생들의 암기 효과를 장려하는 의미 강조 접근법으로, 변칙적인 발음 규칙들을 사용자 스스로 학습할 수 있게 하는 장점이 있다 [12,13].

효과적인 피드백 생성을 구현하기 위해 음소 기반 학습 및 전체 단어 기반 학습 전략 모두를 사용할 수도 있지만, 연음 현상의 경우는 음소로 표현되지도 않고 음성으로도 전달되기 어렵기 때문에 사용자 스스로 부적절한 발음을 수정하기 불충분할 수 있다 [11]. 예를 들어, “stop” /s,t,ɑ,p/과 “top” /t,ɑ,p/ 두 단어의 발음에 포함된 음소 /t/는 동일한 음소로 표시되지만, 실제로는 각각 [s,t,ɑ,p]와 [th,ɑ,p]으로 발음된다. 그러나 이러한 발음상의 차이를 음소를 활용해 모두 표현하게 될 경우 피드백이 매우 복잡해지며, 컴퓨터 보조 발음 교육의 특성상 자가 학습의 효율성이 떨어지기 때문에 연음 현상을 학습하기 위한 효과적이고 간단한 피드백의 생성은 어려운 일이다.

본 논문에서는 연음 현상으로 인해 허용되지 않는 발음에 대한 교정을 목표로 하는 새로운 CAPT 시스템을 제안한다. 피드백의 단순성을 유지하기 위해 제안하는 CAPT 시스템은 복잡한 음소를 표시하는 대신 단어를 추천하여 피드백을 생성하도록 고안한다. 연음으로 인한 오류를 고려하기 위해, 제안하는 시스템은 발음에 사용된 인접 음소 집합의 빈도수와 부적절한 발음에 대한 기여도를 고려한다. 연음 현상으로 인한 부적절한 발음 정정의 효용성을 검증하기 위해 제안된 시스템에서 고려한 음소를 추적하여 개별 사용자의 피드백에 대한 추가 분석을 수행한다. 실험 결과를 통해 제안하는 시스템이 사용자의 잘못된 발음을 수정하는 데 효과적인 것을 확인할 수 있었다.

II. Related Works

CAPT 시스템을 통한 일반적인 교육 시나리오는 (1) 시스템이 단어를 표시하고 (2) 사용자가 시스템이 표시한 단어를 발음한 뒤, (3) 시스템이 학생에게 피드백을 제공함으로써 사용자의 잘못된 발음 또는 적어도 수용하기 어려운 발음을 교정한다 [14,15]. 이러한 잘못된 발음의 원인은 참여한 사용자의 모국어에 따라 다를 수 있으므로 과거 영어 발음 교육에서 습득한 사전 지식을 활용하여 분석을 향상시킬 수도 있다 [4].

발음 교육의 전략은 크게 지각적 훈련과 생산적 훈련 두 가지로 나눌 수 있다 [10]. 지각적 교육은 단어의 올바른 발음을 사용자에게 들려줌으로써 발음 정확도를 향상시키고 이를 반복함으로써 모국어와의 혼동으로 인한 오류를 줄여나간다. 반면에 생산적 훈련은 강세, 억양, 어구 및 부정확한 혀의 움직임과 같은 오류를 정정하는 것을 목표로 합니다. 전통적으로 오디오 자료만으로 진행되는 전통적인 생산적 훈련은 학습자를 위한 교정 정보가 부족하다는 비판이 있었기 때문에 연구자들은 CAPT 시스템에서 기존의 생산적 훈련 전략의 약점을 보완하기 위한 다양한 노력을 해왔다.

최근에는 모국어에 따른 잘못된 발음 습관을 고려하여 교정 피드백을 산출하는 방법을 조사하기 위한 다양한 연구가 제안되었다 [4]. 초기의 연구들은 발음 테스트에서 얻은 점수 등의 요약된 피드백을 제공했으나 [14,15,16], 특정 오류가 왜 또는 왜 발생했는지에 대한 정보를 제공하지 않아 발음 훈련에 한계가 있었다 [17]. 학습자는 종종 학습자의 모국어(L1 언어)에서 활용되는 음소와 비슷한 학습할 언어 (L2 언어)의 음소를 발음하는 경향이 있는데 [4,18,19], 이를 활용하면 CAPT 시스템을 통해 주어진 단어를 어떻게 실제로 발음했는지를 지적함으로써 학습의 효과를 증대시킬 수 있다 [20]. 이러한 접근법은 발음 오류 패턴 탐색 기술로 알려져 있는데, 오류 패턴 탐색은 주로 음성 인식 기술을 수반한다 [5,7,15,21,22]. 사용자의 오류 패턴을 바탕으로 로지스틱 회귀 분석 [23] 또는 선형 회귀 분석 [24] 등의 기계학습 알고리즘을 적용하여 피드백을 생성하기도 하였다. 학습 사례에 비해 너무 많은 개수의 오류 패턴으로 인한 기계학습 알고리즘의 오류를 방지하기 위해 문자-음소 변환을 위한 음소 시퀀스 모형을 만들기도 하였다 [9,25]. 발음 훈련의 효과를 높이기 위해 발음의 오디오 정보와 조음 기관의 시각적 정보가 제안 된 멀티 모달 예제를 제공하는 CAPT 시스템 또한 연구되었다 [10,26]. 오디오와 시각적 정보를 동기화하기 위해, 조음 기관의 모양을 생성하는 애니메이션을 생성함으로써 피드백을 생성하였다 [27,28].

기존 연구에서는 사용자의 발음을 음소 단위로 분석하여 각 개별 음소를 기준으로 부적절한 발음을 판별한 뒤, 찾아낸 음소를 바탕으로 피드백이 이루어졌다 [14,18]. 또한, 사용자의 발음을 분석하고 피드백을 생성하는 과정에서 기계학습과 같이 계산량이 많거나 과도하게 복잡한 알고리즘을 사용함으로써 적용 범위에 제한이 있다 [9,15,23]. 한편, 음소를 기준으로 만들어진 피드백은 영어 발음 초보자들이 정확히 이해하기 어려운 면이 있으며, 인접 음소에 의해 달라지는 발음을 음소만을 활용하여 표현할 경우 피드백이 한층 더 어려워지는 단점이 있다 [25,26]. 이를 극복하기 위해 제안하는 시스템은 인접음소 기준으로 부적절한 발음을 분석하는 데에만 집중함으로써 계산량을 활용하여 획기적으로 감소시키되 피드백 단계에서는 단어 추천을 통해 발음 훈련을 수행하도록 한다.

III. Proposed Methods

CAPT 시스템이 사용자의 부정확한 발음의 원인을 확인하고 피드백을 제공하기 위해서 사용자 발음을 테스트하기 위한 일련의 단어 집합이 필요하다 [29]. 제안하는 시스템에서는 영어 발음 훈련 전문가를 통해 선별된 초등학교 수준에서 자주 사용되는 700 단어를 선택함으로써 시스템이 제시한 단어를 사용자가 알지 못해 발생하는 오류를 피할 수 있도록 하였다 [30]. 기존 CAPT 시스템의

단점인 발음 인식 알고리즘의 높은 계산 복잡도 문제를 피하기 위해 제안 된 시스템은 Android와 같이 널리 사용되는 모바일 장치 운영체제에서 번들 응용 프로그램으로 포함되는 Google Voice Search (GVS)를 사용하도록 고안되었다 [31]. 이 때, 사용자가 제시된 단어를 발음하면 GVS는 해당 음성 신호를 구글 서버로 전송한 다음 최대 10개 미만의 인식된 단어 집합을 반환한다. 반환된 단어 목록에 테스트 된 단어가 없는 경우 제안하는 시스템은 해당 단어에 대한 사용자의 발음이 적절하지 않다고 판단한다. 예를 들어, 테스트에 사용된 단어가 100개이고, 이에 대해 사용자가 발음을 하였을 때, 40개의 단어가 GVS에 의해 회신된 리스트에 포함되어 있지 않다면 해당 사용자의 발음 정확도는 60%로 간주한다.

많은 연구에서 사용자의 발음 오류는 각 음소의 단위를 기준으로 분석을 수행하여 감지되었다. 그러나 단일 음소를 기준으로 분석을 수행할 경우 동일한 음소라도 어떤 음소와 인접해 있느냐에 따라 다르게 발음 될 수 있다 [10]. 예를 들어, "little"이라는 단어의 발음은 /l,ɪ,t,l/로 표현 될 수 있는데, 음소 /l/이 어떤 음운이 선행(/t,l/)하고 후행(/l,ɪ/)했는지에 따라 다르게 발음된다. 그러나 이러한 차이를 음소를 이용해 모두 표현하는 것은 과도하게 복잡한 피드백을 생성하는 결과를 낳으므로 음소 기반 학습의 효과를 현저하기 저하시키는 원인이 될 수 있다. 이를 극복하기 위해 오디오 예제가 사용자에게 제공될 수도 있겠지만, 교사가 없는 상황에서 발음의 차이를 스스로 구별해내는 것을 초급 사용자에게 기대하기는 어렵다.

연음 현상에 의한 발음 오류를 찾고 이에 적합한 피드백을 사용자에게 제안하기 위해 제안하는 시스템은 아래 3가지 기준으로 피드백을 생성하도록 고안되었다.

● 조건1: 연음은 2개 이상의 음소의 발음에서 발생하므로 시스템은 2개 이상의 음소로 구성된 인접 음소 집합을 고려할 수 있어야 한다.

● 조건2: 발음 연습을 무한히 할 수 없으므로 빈도수가 높은 인접 음소를 고려하는 것이 발음 정확도 향상에 많은 기여를 할 것이다. 한편, 빈도수가 낮은 인접 음소는 실생활에서 쓰일 가능성이 낮다. 따라서 자주 사용되지 않는 인접 음소에 대한 발음 연습은 중요도가 낮으므로 출현 빈도가 높은 인접 음소를 우선 고려해야 한다.

● 조건 3: CAPT 시스템의 목표는 사용자의 발음 기술을 교정하기 위한 것이므로 사용자가 부정확하게 발음한 단어에서만 주로 나타나는 인접 음소 집합을 추출해야 한다.

Table 1. Phonemes considered by our system

Index of Phonemes									
1	/ʧ/	2	/ʤ/	3	/aʊ/	4	/aɪ/	5	/eɪ/
6	/oʊ/	7	/ɔɪ/	8	/ə/	9	/ŋ/	10	/Θ/
11	/ð/	12	/ɔ/	13	/ɜ/	14	/æ/	15	/ɪ/
16	/ʌ/	17	/ʊ/	18	/ʃ/	19	/ʒ/	20	/a/
21	/b/	22	/d/	23	/e/	24	/f/	25	/g/
26	/h/	27	/i/	28	/j/	29	/k/	30	/l/
31	/m/	32	/n/	33	/p/	34	/r/	35	/s/
36	/t/	37	/u/	38	/v/	39	/w/	40	/z/

본 연구에서는 위의 조건을 바탕으로 다음과 같이 문제를 정의한다. 먼저, 테스트 단어 집합 X를 활용하여 발음 테스트가 종료되면 $|X|$개의 단어에 대한 발음 테스트 결과 $Y=\{y_1,...,y_{|X|}\}$를 얻을 수 있으며, $y_i \in \{0,1\}$는 i번째 단어 x_i에 대해 부정확한 것으로 판단된 테스트 결과를 1로, 정확하다고 판단된 결과를 0으로 표시한다. 이를 이용하면, 단어 $x \in X$를 발음하기 위해 필요한 음소의 개수가 n개라고 하였을 때, 인접 음소를 고려하여 식(1)과 같이 표현할 수 있다.

$$x_i = \left\{(x_i^{(1)},x_i^{(2)}),...,(x_i^{(n-2)},x_i^{(n-1)}),(x_i^{(n-1)},x_i^{(n)})\right\} \quad (1)$$

이 때, $x_i^{(j)}$는 단어 x_i를 발음하기 위해 필요한 음소 중 j번째 음소를 나타내며, 본 연구에서는 Table 1에서 나열한 총 40개의 음소 집합 중 하나가 될 수 있다.

U를 테스트 단어 집합 X에서 추출할 수 있는 모든 인접 음소들의 집합이라고 하였을 때, 하나의 인접 음소 집합 $u \in U$를 포함하는 단어 집합 $P_u \subseteq X$가 있다고 하자. 그러면 이 단어 집합 P_u는 식(2)와 같이 정의할 수 있다.

$$P_u = \{x_i | u \in x_i\} \quad (2)$$

식(2)를 활용하여 제안하는 시스템에서 단어 추천을 통한 피드백 생성 중 고려할 인접 음소의 등장 빈도수는 식(3)과 같이 계산할 수 있다.

$$f(u) = \frac{|P_u|}{|X|} \quad (3)$$

테스트 단어 집합 X가 고정되어 있을 경우 인접 음소 집합의 빈도수 역시 고정되지만, 실제로는 각 사용자별로 테스트 결과 Y가 바뀌고, 정상 발음으로 판정된 단어에 포함된 인접 음소 집합의 경우 단어 추천에서 제외되므로, 각 사용자별로 고려되는 음소 목록이 달라질 수 있다. 이 때, 사용자가 정확하게 발음한 단어 집합을 식(4)와 같이 정의할 수 있다.

$$Q = \{x_i | y_i = 0\} \quad (4)$$

조건 3에 의해 사용자가 부정확하게 발음한 단어에서만 주로 나타나는 인접 음소 집합을 고려해야 하므로 아래의 식(5)를 만족하는 인접 음소 집합 v는 제외해야 한다.

$$V = \{v \in U | P_v \cap Q \neq \varnothing\} \quad (5)$$

따라서 제안하는 시스템에 의해 고려될 인접 음소 집합 R은 식(6)과 같이 정의될 수 있다.

$$R = \{U - V\} \quad (6)$$

이제 R에 포함된 각각의 인접 음소 $r_1,...,r_{|R|}$과 빈도수 $f(r_1),...,f(r_{|R|})$를 활용하여 i번째 단어 x_i의 중요도를 식(7)을 이용하여 계산할 수 있다.

$$g(x_i) = \sum_{u \in x_i} f(u) \quad (7)$$

식(7)을 이용하여 모든 단어에 대한 추천 중요도 $g(\cdot)$을

구할 수 있다. 제안하는 시스템은 추천 중요도 점수가 높은 단어를 위주로 총 70개의 단어를 사용자에게 추천한다.

IV. Example of Word Recommendation

제안하는 시스템의 추천 단어 선택 과정을 보여주기 위해 단어 5개로 이루어진 데이터에 대한 예제를 준비하였다. 본 예제에서 등장하는 단어는 “boat", "flow", "goat", "nose", 그리고 "throat"로 이 단어들의 발음은 각각 /b,oʊ,t/, /f,l,oʊ/, /g,oʊ,t/, /n,oʊ,z/, 그리고 /Θ,r,oʊ,t/로 표기할 수 있다. 각각의 인접 음소들을 별도로 떼어내어 표기함으로써 어떤 인접 음소가 각 단어의 발음에 필요한지 알 수 있는데, 이를 Table 2를 통해 표시하였다. 예를 들어, "boat"는 인접 음소 집합 {(b,oʊ), (oʊ,t)}로 표현할 수 있고, "throat"의 경우 인접 음소 집합 {(Θ,r), (r,oʊ), (oʊ,t)}로 표현할 수 있다. 이를 바탕으로 인접 음소를 포함하는 단어 집합을 추출할 수 있는데, 이를 위해 Table 2를 살펴보면 5개의 단어를 표현하기 위해 필요한 인접 음소 집합 U의 크기는 9인 것을 알 수 있으며, 각각의 인접 음소가 포함되어 있을 경우 1로, 없을 경우 0으로 표시한 것을 확인할 수 있다. 예를 들어 인접 음소 (oʊ,t)를 포함하는 단어 집합 P(oʊ,t)={"boat", "goat", "throat"}가 되며, 따라서 인접 음소 (oʊ,t)의 빈도수 f(oʊ,t) $=3/5=0.6$임을 알 수 있다.

다음으로 영어 발음 교육을 위한 추천에서 정확하게 발음된 단어들은 제외해야 한다. Table 2를 살펴보면 Test Result가 0인 단어들이 있는데, 이를 고려하면 Q는 {“flow", "nose"}임을 알 수 있다. 따라서 정확하게 발음한 단어에 포함된 인접 음소들은 (f,l), (l,oʊ), (n,oʊ), (oʊ,z)이므로 V={(f,l), (l,oʊ), (n,oʊ), (oʊ,z)}이며, 결과적으로 제안하는 시스템에 의해 고려될 인접 음소 집합 R={(b,oʊ), (oʊ,t), (g,oʊ), (Θ,r), (r,oʊ)}로 총 5개가 되는 것을 알 수 있다.

마지막으로 집합 R에 포함된 인접 음소의 빈도수에 따라 최종적으로 각 단어의 점수를 측정한다. 예를 들어, “throat”의 경우 {(Θ,r), (r,oʊ), (oʊ,t)}로 표현되는데, 각 인접 음소의 빈도수는 0.2, 0.2, 0.6이므로 “throat"의 중요도 점수는 1.0으로 평가된다. 이와 마찬가지로 "boat", "goat" 역시 중요도 점수가 0.8로 동일하게 평가되는 것을 확인할 수 있는데, 이에 따라 제안하는 시스템은 "throat", "boat", "goat" 단어들을 연습 과정에서 추천하되, ”flow"와 "nose"는 추천하지 않게 된다.

V. Experimental Results

본 연구에서는 한국어를 모국어로 하는 5명의 사용자를 대상으로 실험을 실시하였다. 모든 사용자는 25.2±4.2세의 남성이며, 한국에서 통상적으로 요구되는 고등학교 수준의 영어 교육을 마쳤다. 제안하는 시스템에 의한 영어 발음 교육의 효과성을 검증하기 위해 각 사용자에게 제안된 시스템에서 추천하는 단어를 연습 할 기회를 제공한 후 각 사용자의 발음 정확도 향상을 측정하였다. 이 때, 실험 1일차에는 각 사용자가 제안하는 시스템이 추천한 각 단어를 발음하고, 실험 2일차에는 각 사용자가 시스템에서 추천하는 단어를 발음하여 연습을 수행하였다. 마지막으로 실험 3일차에는 각 사용자가 실험 1일차에 사용한 단어를 다시 발음하여 발음 정확도를 계산하였다.

Table 2. Pronunciation accuracy of each user at each testing round

Test round	Users					Avg.
	S1	S2	S3	S4	S5	
1st round	70%	61%	58%	78%	56%	65%
2nd round	76%	65%	69%	85%	60%	71%
Difference	6%	4%	11%	7%	4%	6%

5.1. Analysis on correction process

Table 2는 각 테스트에 따라 사용자의 발음 정확도 변화를 보여주고 있다. Table 2의 두 번째 행에는 1차 테스트에서 모든 테스트 단어 대비 각 사용자가 올바르게 발음한 단어의 백분율 값(정확도)을 표시하고 있다. 예를 들어, 사용자 5 (S5)는 첫 번째 테스트에서 56%의 단어에 대해 정확한 발음을 하여 다른 사용자들과 비교하였을 때 가장 낮은 발음 정확도를 달성한 것을 알 수 있다. 제안한 시스템에 의해 추천된 단어를 활용하여 연습한 후, 2차 테스트에서의 사용자별 변화된 발음 정확도는 세 번째 행에 나타내었으며, 1차 테스트와 2 차 테스트의 발음 정확도 차이를 Table 2의 마지막 행에 나타내었다. 전체적인 실험 결과를 보았을 때, 발음 정확도가 평균 6 % 증가하였으므로, 제안하는 시스템이 사용자의 발음 정확도 향상에 기여한 것을 알 수 있다.

제안하는 시스템의 우수성을 검증하기 위해 추가적으로 비교 실험을 수행하였다. 이 실험에서는 한국어를 모국어로 하는 남자 4명, 여자 1명으로 구성된 5명의 사용자가 참여하였다. 모든 과정은 제안하는 시스템의 검증 조건과 동일하게 진행되도록 하였으나 인접 음소 교정에 의한 효과성을 검증하기 위해 비교 실험에서는 피드백 과정에서 사용자들의 발음 결과와 무관하게 무작위로 단어를 추천하여 발음 연습을 수행하도록 하였다. 이와 같은 실험을 수행한 결과 1차 테스트에서 측정된 발음 정확도와 2차 테스트에서 측정된 발음 정확도에 차이가 거의 없었으며, 평균적으로는 약 1.8%의 발음 정확도 하락이 관측되어 제안하는 시스템에 의한 단어 추천 전략이 무작위 단어 추천에 비해 효과적임을 알 수 있었다.

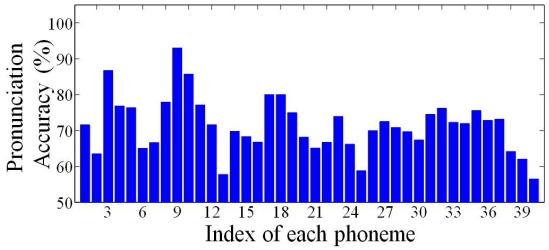

Fig. 1. Pronunciation accuracy of each phoneme averaged over 5 users at 1st round test

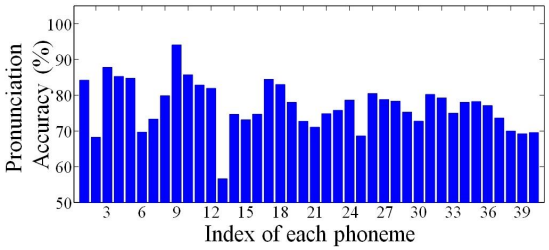

Fig. 2. Pronunciation accuracy of each phoneme averaged over 5 users at 2nd round test

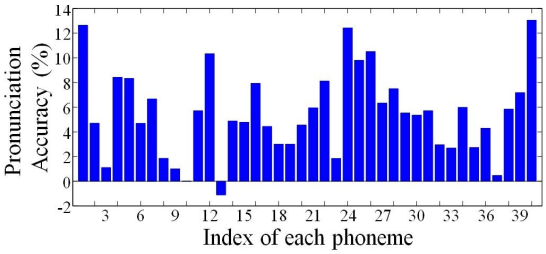

Fig. 3. Improvement of pronunciation accuracy according to each phoneme averaged over 5 users

각 음소를 기준으로 발음 정확도를 분석하기 위해, 단어를 발음하기 위해 필요한 음소를 포함하는 단어별로 발음 정확도를 측정하였다. 단, 인접 음소를 고려함으로써 분석이 복잡해지는 것을 피하기 위하고 가시성을 확보하기 위해 본 분석에서는 단일 음소를 기준으로 분석을 수행하였다. 예를 들어 사용자가 /z/와 같은 음소를 잘못 발음하는 경향이 있는 경우 “zebra", "zoo", "zone"과 같은 단어의 발음 정확도가 낮을 것이다. Fig. 1은 1차 테스트에서 각 음소를 포함하는 단어 집합을 기준으로 평균 발음 정확도를 보여주고 있다. 가로축은 각 단어 집합에 공통적으로 포함되는 각 음소의 인덱스를 나타내고, 세로축은 Table 1에서 제시한 각 음소의 인덱스를 바탕으로 해당 단어 집합에 대한 5명의 사용자의 평균 발음 정확도를 나타낸다. 실험 결과, 각각의 음소에 따라 발음 정확도가 균등하지 않아 음소에 따른 전략적인 교정이 필요하다는 것을 보여준다.

Table 3은 Fig. 1에 표현된 음소에 따른 평균 발음 정확도가 가장 낮은 음소 10를 나타낸 것이다. 기존 연구에서 알려진 바와 같이 한국인이 영어 발음을 할 경우, /z/, /g/, /ʤ/, /b/와 같은 음소를 한국어에서 활용되는 유사한 자음과 혼동하여 잘못 발음하는 경우가 종종 있다 [32, 33, 34]. 더욱이, 한국인들은 /f/와 /v/와 같은 마찰음의 발음을 정확하게 발음하기가 어려워 한다는 점도 잘 알려져 있다 [35]. 또한, 영어에 비해 한국어의 모음 수가 적다는 점에서 한국어 사용자가 모음 발음을 덜 구분하는 것 기존 연구에서 잘 연구된 바가 있다. 예를 들어 한국인 사용자는 음소 /æ/와 /ɜ/ 사이에 거의 차이를 두지 않는다 [36]. 또한 한국인 사용자를 대상으로 한 영어 발음 교육 연구에서 자주 /oʊ/와 /o/를 혼동한다고 보고된 바 있다 [37]. 요약하면 한국인 사용자의 입장에서 원어민과 비교했을 때 영어 모음을 정확하게 발음하기 어렵다는 점이 알려져 있으며, 실험 결과 역시 한국어 사용자의 입장에서 발음하기 어려운 음소들이 주로 포함되어 있음을 나타내고 있다.

Table 3. Worst 10 phonemes with lowest average pronunciation accuracy at the 1st round test

Index	40	13	25	39	2
Phoneme	/z/	/ɜ/	/g/	/w/	/ʤ/
Accuracy	57%	58%	59%	62%	64%
Index	38	6	21	24	7
Phoneme	/v/	/oʊ/	/b/	/f/	/ɔɪ/
Accuracy	64%	65%	65%	66%	67%

Table 4. Top 10 phonemes in the viewpoint of improved pronunciation accuracy shown in Fig. 3

Index	Phoneme	1st test	2nd test	Increment
40	/z/	57%	70%	13%
1	/ʧ/	72%	84%	13%
24	/f/	66%	79%	12%
26	/h/	70%	81%	11%
12	/ɔ/	72%	82%	10%
25	/g/	59%	69%	10%
4	/aɪ/	77%	85%	8%
5	/eɪ/	76%	85%	8%
22	/d/	67%	75%	8%
16	/ʌ/	69%	75%	8%

Fig. 2는 5명의 사용자에 대한 2차 테스트에서 각 음소를 포함하는 단어의 평균 발음 정확도를 보여주고 있다. Fig. 3은 Fig. 1과 Fig. 2를 바탕으로 각 음소에 따른 발음 정확도 향상을 도표로 보여주고 있는데, 대부분의 음소에 대해 각 음소의 발음 정확도가 향상되었음을 알 수 있다. Table 4는 Fig. 3에 나열된 음소 중 발음 정확도 향상 정도를 기준으로 상위 10개의 음소를 추려낸 것이다. Table 4를 살펴보면, 1차 테스트에서 발음 정확도가 낮았던 /z/, /f/ 및 /g/와 같은 음소의 발음 정확도가 피드백 및 연습 후 약 10 % 향상된 것을 알 수 있다.

5.2. In-depth Analysis on users

Table 5는 제안하는 시스템이 사용자 1의 추천 단어를 추천할 때 고려한 인접 음소들의 목록을 나타내고 있다. 실험 결과에 따르면 사용자 1은 자음 “f”를 포함하는 인접 음소들, 예를 들어 /f,i/, /f,ə/, /f,eɪ/ 등이 포함된 단어인 “breakfast", "feel", "fun" 및 "favorite"들에 대해 잘 발음하지 못한다는 것을 알 수 있다. 발음된 것을 이 모든 단어는 발음 될 때 f에 대한 스트레스를 포함합니다. 반면에 사용자 1은 ”shelf"나 “yourself"와 같이 f가 마지막에 발음되는 단어들은 올바르게 발음했음을 확인하였다. 또한 사용자 1은

/v,eɪ/나 /v,æ/와 같이 발음에서 /v/를 포함하는 단어들을 잘 발음하지 못한 것을 확인할 수 있다. 요약하면, 사용자 1은 1차 테스트에서 잘못 발음한 351 단어 중 85개를 정확히 발음하는데 성공했지만 /f/ 및 /v/의 발음에는 큰 변화가 없었다. 이는 /f/와 /v/가 한국어에서 사용되지 않는 발음이기 때문인 것으로 추정되며, 기존 연구에서는 그 원인으로 /f/와 /v/의 발음을 한국어의 'ㅍ'이나 'ㅂ' 같은 자음의 발음으로 자주 바꾸기 때문이라고 지적한바 있다 [38].

사용자 2의 발음 연습 전 정확도는 61.14%였으나 2차 테스트에서의 발음 정확도는 64.57%로 측정되어 전반적으로 3.43%의 향상도을 기록하였다. Table 6은 제안하는 시스템이 사용자 2를 위한 추천 단어를 생성하기 위해 고려한 인접 음소의 목록을 나열한 것이다. 사용자 2는 "game", "gate", "get"과 같이 /g,eɪ/ 또는 /g,e/과 같은 인접 음소를 발음에서 요구하는 경우 정상적으로 발음하는 것을 확인하였다. 그러나 /æ,g/ 또는 /e,g/ 등의 인접 음소를 발음에서 요구하는 단어들, 예를 들어 "bag", "tag", "beg"의 경우 정확하게 발음하지 못 하였다. 즉, 사용자 2는 음소 /g/가 자음으로 위치하는 경우 이를 올바르게 발음했지만, 다른 모음 뒤에 따라오는 경우에는 정상적으로 발음하지 못한 것을 의미한다.

Table 5. Important Adjacent phonemes for User 1 and a list of words possibly recommended

Adjacent Phonemes	$f(\cdot)$	Possible words to be recommended
/l,d/	0.86	cold, field, fold, hold, old, shoulder, yield, scalder, ...
/d,u/	0.86	due, duke, dune, duration, duty, duplicate, dual, ...
/f,i/	0.71	feed, feel, feet, fever, field, fifteen, fifty, fig, fill, ...
/r,ɪ/	0.57	gorilla, green, hungry, hurry, orange, ostrich, read, ...
/v,e/	0.43	very, vest, vet, verify, vendor, vessel, vesting, ventilate, ...
/f,ə/	0.43	breakfast, elephant, facilitate, facility, facilitator, ...
/v,æ/	0.29	value, van, vanity, vanish, valid, validation, vantage, ...
/v,eɪ/	0.29	vain, vase, vail, vane, vein, vapour, vains, ...
/f,ʌ/	0.14	fudge, fun, fund, fungi, fungus, funnel, funnier, funny, ...
/f,eɪ/	0.14	face, facing, fade, fail, failure, faint, faintest, faith, fake, ...

사용자 1과 마찬가지로, 실험 결과에 따르면 사용자 2는 발음 연습 후 발음 중 강세가 포함되어 있지 않은 단어에 대한 발음을 교정하였으며, 이는 전체 단어 대비 약 10% 정도에 해당한다. 또한 인접 음소 /ʤ,a/와 /aʊ,a/를 포함하는 단어의 경우, 약 80 %의 발음 오류가 수정된 것을 확인할 수 있었다. 전체적인 관점에서는 발음 연습 전 정상적으로 발음하지 못했던 단어 272개 중 102개의 단어가 수정되어 37.5%의 향상을 보였다.

VI. Conclusions

본 연구에서는 영어 발음 중 연음에 초점을 맞추어 효과적인 CAPT 전략을 제안하였다. 특히, 제안하는 시스템은 음소의 위치에 따른 발음의 변화를 구별하기 위해 복잡한 음소를 표시하는 대신 사용자의 연음에 대한 오류를 고려하여 추천 단어 목록을 생성하고 이를 사용자에게 피드백함으로써 피드백의 단순성을 유지하였다. 실험 결과에 따르면 모든 사용자가 제안하는 시스템을 사용하여 발음 기술을 향상시킬 수 있었다. 또한, 각 사용자에 대한 분석은 연음에 대한 부정확한 발음에 대해 상당한 교정 효과가 있음을 보여주었다.

본 연구에서는 단어 추천을 기반으로 연음에 따른 발음의 변화와 이에 대한 교정을 목표로 하는 효과적인 피드백 전략을 고안 하였지만, 해당 정보는 여전히 추천된 단어 속에 숨겨져 있으므로 사용자 본인은 연음에 의한 발음 차이를 포착하지 못할 수 있는 한계가 있다. 이를 극복하기 위해 조성 기관에 대한 시청각 정보를 제공하면 교육의 효율성을 향상시킬 수 있으리라 기대된다. 한편, 비록 본 연구에서는 연음을 목표로 삼았지만, 발음 오류의 다양한 원인을 목표로 한 새로운 컴퓨터 보조 발음 훈련 시스템의 연구 역시 필요하다. 추후 연구에서는 위와 같은 주제로 연구를 진행하고자 한다.

Table 6. Important Adjacent phonemes for User 2 and a list of words possibly recommended

Adjacent Phonemes	$f(\cdot)$	Possible words to be recommended
/f,ɔ/	0.71	fog, fork, forty, four, fourteen, fall, false, fault, foil, ...
/k,oʊ/	0.71	coach, coal, coast, code, cold, coat, cone, cope, cobra, ...
/æ,g/	0.57	bag, baggage, drag, dragon, flag, nag, tag, wag, ...
/e,g/	0.57	egg, beg, leg, leggings, peg, mega, regular, begging, ...
/eɪ,d/	0.43	afraid, aid, arcade, bathe, blade, maid, raid, paid, ...
/v,e/	0.43	adventure, available, very, vest, vet, conservation, ...
/w,æ/	0.43	wack, wacky, wag, waggle, wagtail, quack, wax, ...
/ʤ,a/	0.29	hedgehog, job, jog, jockey, jogging, jock, ...
/a,ʊə/	0.29	power, shower, flower, tower, empower, bower, ...
/aɪ,v/	0.29	drive, derive, five, arrive, survive, strive, alive, ...

REFERENCES

[1] M. Pennington and J. Richards, "Pronunciation revisited,"

TESOL quarterly, Vol. 20, No. 2, pp. 207-225, 1986
[2] J. Smith and B. Beckmann, "Improving pronunciation through Noticing-Reformulation Tasks, University College London, 2005
[3] S. Shaik, "Computer assisted English pronunciation training to undergraduate students," Journal of English Language and Literature, Vol. 4, No. 2, pp. 117-121, 2015
[4] H. Liao, Y. Guan, J. Tu, and J. Chen, "A prototype of an adaptive Chinese pronunciation training system," System, Vol. 45, No. 1, 2014
[5] R. Hincks, "Speech technologies for pronunciation feedback and evaluation," ReCALL, Vol. 15, No. 1, pp. 3-20, 2003
[6] C.-S. Park, "Understanding Artificial Intelligence Technology for Artificial Intelligence Humanities," Journal of AI Humanities, Vol. 1, No. 1, pp. 173-182, 2018
[7] G. Demenko, A. Wagner, N. Cylwik, "The use of speech technology in foreign language pronunciation training," Archives of Acoustics, Vol. 35, No. 3, pp. 309-329, 2010
[8] G. Kartal, "Working with an imperfect medium: Speech recognition technology in reading practice," Journal of Educational Multimedia and Hypermedia, Vol. 15, No. 3, pp. 303-328, 2006
[9] X. Qian, H. Meng, and F. Soong, "Capturing L2 segmental mispronunciations with joint-sequence models in computer-aided pronunciation training," In proceedings of 7th International Symposium on Chinese Spoken Language Processing, pp. 84-88, Tainan, Taiwan, 2010
[10] K. Wong, W. Leung, W. Lo, and H. Meng, "Development of an articulatory visual-speech synthesizer to support language learning," In proceedings of 7th International Symposium on Chinese Spoken Language Processing, pp. 139-143, Tainan, Taiwan, 2010
[11] K. Wong, W. Lo, and H. Meng, "Allophonic variations in visual speech synthesis for corrective feedback in CAPT," In proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, pp. 5708-5711, Prague, Czech, 2011
[12] K. Goodman, "Reading: A psycholinguistic guessing game," Literacy Research and Instruction, Vol. 6, No. 4, pp. 126-135, 1967
[13] P. Gough, C. Juel, and P. Griffith, "Reading, spelling, and the orthographic cipher," Lawrence Erlbaum Associates, Inc, 1992
[14] H. Franco, H. Bratt, R. Rossier et al., "EduSpeak: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications," Language Testing, Vol. 27, No. 3, pp. 401-418, 2010
[15] S. Witt and S. Young, "Phone-level pronunciation scoring and assessment for interactive language learning," Speech Communication, Vol. 30, No. 2, pp. 95-108, 2000
[16] X. Xi, D. Higgins, K. Zechner, and D. Williamson, "A comparison of two scoring methods for an automated speech scoring system," Language Testing, Vol. 29, No. 1, pp. 371-394, 2012
[17] H. Liao, J. Chen, S. Chang, et al., "Decision tree based tone modeling with corrective feedbacks for automatic Mandarin tone assessment," In proceedings of 11th Annual Conference on the International Speech Communication Association, pp. 602-605, Chiba, Japan, 2010
[18] M. Harrison, W. Lau, H. Meng, and L. Wang, "Improving mispronunciation detection and diagnosis of learners' speech with context-sensitive phonological rules based on language transfer," In proceedings of 9th Annual Conference on International Speech Communication Association, pp. 2787-2790, Brisbane, Australia, 2008
[19] M. Harrison, W. Lo, X. Qian, and H. Meng, "Implementation of an extended recognition network for mispronunciation detection and diagnosis in computer-assisted pronunciation training, In proceedings of ISCA Workshop Speech and Language Technology in Education, pp. 45-48, Warrickshire, UK, 2009
[20] L. Wang, X. Feng, H. Meng, "Mispronunciation detection based on cross-language phonological comparisons," In proceedings of International Conference on Audio, Language and Image Processing, pp. 307-311, Shanghai, China
[21] A. Neri, C. Cucchiarini, H. Strik, and L. Boves, "The pedagogy-technology interface in computer assisted pronunciation training," Computer assisted language learning, Vol. 15, No. 5, pp. 441-467, 2002
[22] F. Zhang F, C. Huang, F. Soong, M. Chu, and R. Wang, "Automatic mispronunciation detection for Mandarin," In proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 5077—5080, Las Vegas, USA, 2008
[23] J. Doremalen, C. Cucchiarini, H. Strik, "Automatic pronunciation error detection in non-native speech: The case of vowel errors in Dutch," The Journal of the Acoustical Society of America, Vol. 134, No. 2, pp. 1336-1347, 2013
[24] H. Strik, K. Truong, F. De Wet, C. Cucchiarini, "Comparing different approaches for automatic pronunciation error detection," Speech communication, Vol. 51, No. 10, pp. 845-852, 2009
[25] L. Wang, X. Feng, H. Meng, "Automatic generation and pruning of phonetic mispronunciations to support

computer-aided pronunciation training, In proceeding of 9th Annual Conference on the International Speech Communication Association, pp. 1729-1732, Brisbane, Australia, 2008

[26] J. Zhao, H. Yuan, W. Leung, H. Meng, J. Liu, S. Xia, "Audiovisual synthesis of exaggerated speech for corrective feedback in computer-assisted pronunciation training," In proceedings of 2013 IEEE International Confernece on Acoustics, Speech and Signal Processing, pp. 8218-8222, Vancouver, Canada, 2013

[27] P. Badin, A. Ben Youssef, G. Bailly, F. Elisei, and T. Hueber, "Visual articulatory feedback for phonetic correction in second language learning. In proceedings of ISCA Workshop Speech and Language Technology in Education, pp. 1-10, Tokyo, Japan, 2010

[28] Y. Iribe, S. Manosavanh, K. Katsurada, R. Hayashi, C. Zhu, T. Nitta, "Generating Animated Pronunciation from Speech Through Articulatory Feature Extraction," In proceedings of 12th Annual Conference on International Speech Communication Association pp. 1617-1620, Florence, Italy, 2011

[29] W. Lo, A. Harrison, and H. Meng, "Statistical phone duration modeling to filter for intact utterances in a computer-assisted pronunciation training system, In proceedings of The 35th IEEE International Conference on Acoustics Speech and Signal Processing, pp. 5238—5241, Dallas, USA, 2010

[30] J. Lee, C.-H Lee, D.-W. Kim, B.-Y. Kang, "Smartphone-assisted pronunciation learning technique for ambient intelligence," IEEE Access, Vol. 5, No. 1, pp. 312-325, 2017

[31] J. Schalkwyk, D. Beeferman, F. Beaufays et al. ""Your word is my command": Google search by voice: a case study," In collection of Advances in speech recognition, pp. 61-90, 2010

[32] H. Koo, "A study of the effects of vowels on the pronunciation of English sibilants," Speech Science, Vol. 15, No. 1, pp. 31-38, 2008

[33] Y. Yun and N. Lee, "Research on the effect of pronunciation training of English unaspirated stops for Koreans," Language and Linguistics, Vol. 57, No. 1, pp. 141-158, 2012

[34] J. Kim, "Korean speakers' pronunciation and pronunciation training of English stops," Phonetics Speech Science, Vol. 2, No. 1, pp. 29-36, 2010

[35] H. Koo, "A study of production difficulties of English bilabial stops and labiodental fricateives by Korean learners of English," Phonetics Speech Science, Vol. 1, No. 1, pp. 11-15, 2009

[36] Y. Yun, "The learning effect of English vowels using the phonological information of Korean vowels," Journal of Modern British American Language Literature, Vol. 30, No. 1, pp. 75-91, 2012

[37] J. Kim and K. Yoon, "The formant frequency difference of English vowels as a function of stress and its application on vowel pronunciation training," Phonetics Speech Science, Vol. 5, No. 1, pp. 53-58, 2013

[38] K.-Y. La, "Improvement methods for teaching primary school English pronunciation in the EFL environment," Studies in English Education, Vol. 6, No. 2, pp. 5-31, 2001

Authors

Jaesung Lee is currently an assistant professor in the School of Computer Science and Engineering, Chung-Ang Univ. in Seoul, Korea. Prior to coming to CAU, he did his postdoc, Ph.D., M.S. and B.S. at Chung-Ang Univ., Korea. His research interest includes advanced machine learning algorithms with innovative applications to music emotion recognition, euducational data mining, affective computing, and robot interaction.