

Uncovering between-categorical details in L2 pronunciation errors using Wav2Vec2.0 code vectors*

Eunsoo Hong¹ · Sunhee Kim² · Minhwa Chung^{1,**}

¹Department of Linguistics, Seoul National University, Seoul, Korea

²Department of French Language Education, Seoul National University, Seoul, Korea

Abstract

L2 pronunciation errors, influenced by L1 phonetic transfer, often span the boundaries of two or more canonical phonemes. This phenomenon necessitates moving beyond the categorical judgments by analyzing pronunciation using units more granular than phonemes. While previous studies proposed phonemic posterior-grams (PPG) in this context, these features remain tied to a categorical framework when derived through phonemic association. This study employs Wav2Vec2.0 code vectors as an alternative description unit, granted that the self-supervised learning (SSL) representation is free from external phonemic stipulation. We first evaluate whether code vectors encode L1 and L2 speech differently and use them to identify patterns among categorically identical error samples. To this end, the used inventory in single-speaker data of L1 (CMU ARCTIC) and L2 (L2 ARCTIC) is first compared. Then, substitution errors are selected from our L2 analysis NIA037 data using the model finetuned with TIMIT. For each error type, sub-categorical inspection concerns deriving dominant patterns from corresponding featural sequences and interpreting them using raw code vector clustering results and phoneme-code vector co-occurrence probabilities, both constructed from the L1 reference data TIMIT. The findings revealed that, beyond discerning L2 speech, code vectors can effectively quantify the between-categorical positions of error patterns in the pronunciation spectrum. Specifically, Euclidean distances between pattern vectors reflected phonetically grounded variation phenomena documented in existing L2 literature.

Keywords: Wav2Vec2.0 code vector, self-supervised learning representation, between-categorical, error pattern

1. 서론

제2 언어 발음은 모국어의 영향으로 목표 언어의 표준 발음과 차이가 나며, 이는 단순한 표준 음소의 교체로 설명하기 어려운 경우가 많다. 모국어와 목표 언어의 음성 체계

가 근본적으로 달라 발생하는 현상이기 때문이다. 예를 들어, 광동어 모어 화자가 발음하는 영어 /n/ 음소는 종종 /n/과 /l/의 중간의 성격을 지닌다(Li et al., 2020; Mao et al., 2018). 두 음소를 음절 초성에서 이음으로 여기는 모국어 체계의 변화에 의한 현상이다. 이러한 발음 변이는 음소 수

* This research was supported by the Linguistics 「SNU 10-10 Initiative」 and is based in part on the first author's 2024 master's thesis.

** mchung@snu.ac.kr, Corresponding author

Received 16 August 2024; Revised 28 October 2024; Accepted 29 October 2024

© Copyright 2024 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.

준의 분석만으로 설명되기 어렵기에, 음소 이하의 세분화된 단위의 도입이 제기된다. 아울러 두 (음소)범주 사이의 오류를 체계적으로 분석하기 위해서는, 반복되는 발음 변이의 규칙성(패턴)을 찾아내야 한다. 그러나 세분화된 단위를 직접 정의하고 이를 바탕으로 발음을 표현하며, 오류의 규칙성을 도출하는 과정은 많은 시간과 전문인력을 요구한다. 이 같은 배경 하, 발음 오류의 세분화된 패턴을 발견하기 위해 비지도 학습 기법을 활용한 연구들이 시도되어왔다(Li et al., 2018, 2020; Mao et al., 2018; Wang & Lee, 2013, 2015).

기존 연구들은 공통적으로 MFCC(mel-frequency cepstral coefficients) 음향 특징과 음소 라벨로 훈련된 모델의 음소 사후 확률(phonemic posterior-gram, PPG)을 세분화된 음성 표현 자질로 삼았다. 또한 음성 신호에서 반복되는 특징적인 패턴을 찾는 음향 패턴탐색 기법(acoustic pattern discovery)을 통해 발음 오류와 관련된 패턴을 추출하였다. 이때 분포된 신호의 유사한 특징을 묶는 군집화(clustering)가 활용된다. 음향 패턴탐색 연구의 지평을 연 Park & Glass(2007)에서는 DTW(dynamic time warping)를 이용해 두 음성 시퀀스 사이에서 유사도가 높은 정렬 경로(alignment path)를 선별하고 인접 그래프(adjacency graph)를 만들어 군집화한 단위를 어휘적 의미를 갖춘 패턴으로 간주했다. 어휘 단위(lexical unit)가 발화에서 반복적으로 사용되듯, L2 발음 오류 또한 음성 체계 간 상호작용에 의해 언어학적 유의성을 가지며 반복된다. 이에 착안하여 Li et al.(2018), Wang & Lee(2013, 2015)에서는 동일한 음소 범주에 속하는 발음 표본들의 PPG를 군집화하여 음소 세부 유형을 분류하고자 했다. 한편 Mao et al.(2018)은 전체 발화의 PPG를 군집화해 /n_l/처럼 기존의 음소 범주에 속하지 않는 새로운 운영 단위를 발견하고자 했다.

그러나 PPG는 지도학습 자질이자 범주적 사고를 탈피하지 못한다는 한계를 지닌다. 모델이 학습한 음소 집합에 대한 각 입력 음성프레임의 소속 확률을 나타내므로, 음소 라벨을 동반한 지도학습 과정이 필수적이다. 이때 충분한 양의 데이터 라벨이 확보되지 않으면, 자질의 음성 표현력이 저하된다. 다시 말해 PPG를 분석자질로 사용하는 것은 비지도 기법으로 노동 비용을 절감하고자 하는 연구 흐름에 반한다. 또한 사전 정의된 음소 범주에 대한 분류 결과이므로, 범주적 사고로부터 자유로울 수 없다. 이는 음소 이상의 특징을 파악하기 위해 다시 음소적 틀로 회귀하는 모순점을 드러낸다. 위 두 한계점들을 개선하고자 본 연구에서는 Wav2Vec2.0(Baevski et al., 2020) 코드백터를 대체 분석자질로 탐색해보고자 한다.

Wav2Vec2.0은 자기지도학습(self-supervised learning, SSL) 모델로, 입력된 레이블 없이 데이터의 자기 상관성을 활용하여 의미있는 표현을 학습한다. SSL 모델의 표현 학습(representation learning)은 문자를 보지 않고 소리만 듣고 단어의 의미를 배우는 신생아의 언어학습 기전을 모방하여

발전했다(Liu et al., 2022). 이 과정은 데이터 내에서 의미 있는 표현 단위를 자동으로 찾아내는 음향 패턴탐색의 지향점과 맞닿아 있다. 따라서 학습된 표현은 이미 패턴탐색의 속성을 반영한다.

음향 패턴 분석이 음성 언어의 음소나 형태소와 같은 기본 단위를 찾아내듯, 자기지도학습(SSL)을 통해 얻은 표현들은 음운론적인 패턴, 어휘적인 의미, 그리고 통사론적인 구조를 반영하여 언어의 다양한 측면을 포착한다. Martin et al.(2023)은 HuBERT(Hsu et al., 2021) 모델이 입력 음성을 음소(phoneme)와 음성(phone) 단위에서 각기 다른 표현으로 효과적으로 인코딩할 수 있음을 입증했다. 모델은 음운 중성화되는 서로 다른 음성을 동일한 음소로 인식하고, 같은 표현으로 인코딩했지만, 같은 음소의 변이음은 다른 표현으로 구분하였다. Pasad et al.(2023)은 HuBERT(Hsu et al., 2021), Wav2Vec2.0(Baevski et al., 2020), WavLM(Chen et al., 2022), FaST-VGS+(Peng & Harwath, 2022) 등 다양한 자기지도 학습 모델의 표현들이 소리 정보뿐만 아니라 단어 수준의 속성도 내포하고 있음을 표준 상관 분석, 음향 단어 판별, 단어 분할의 방법들을 통해 입증했다. 이때 사전 학습 목표에 따라 어휘 정보를 가장 풍부하게 담고 있는 층위가 달랐다. 소리의 국소적 특징을 재구축하는 것을 목표로 하는 모델은 중간 층위에서, 중간 층위의 이산적 단위를 재구축하는 것을 목표로 하는 모델은 더 상위 층위에서 가장 많은 어휘 지식을 보유했다. 나아가, 단어 판별 정보는 각 음소 분절의 중간 지점 프레임에서 가장 명확하게 나타났다. Shen et al.(2023)은 통사론적 정보가 모델의 학습 목표를 최적화하는 데 유용하다면, 학습된 모델이 통사론적 정보를 표현할 수 있다는 가정을 검증했다. Pasad et al.(2023)에서 사용된 것과 동일한 모델들의 통사론적 구조 인코딩을 TreeDepth Probe와 TreeKernel Probe라는 두 가지 통사론적 탐색 기법을 활용하여 분석했다. TreeDepth Probe 분석 결과, 구성소 나무의 최대 깊이는 중간 층위부터 깊은 층위에서 주로 나타났다. 또한, TreeKernel Probe를 활용한 유사도 분석을 통해 모델이 학습한 통사론적 정보는 주로 어휘 정보와 긴밀하게 연결되어 있음을 확인했다. SSL을 통해 얻은 표현들이 음운론적, 어휘적, 통사론적 특징을 반영한다는 사실은 본 연구에서 이를 발음 오류 분석에 활용하는 근거가 된다.

한편, 이 같은 언어학적 유의성에 착안하여 자기지도 표현 학습(self supervised representation learning)을 발음 평가에 활용한 연구들이 존재한다. Kim et al.(2022)은 Wav2Vec2.0(Baevski et al., 2020)의 각 층에서 추출한 표현들을 양방향 LSTM을 이용하여 심층 분석하고, 이를 종합해 음성 맥락 벡터(audio context vector)를 만들었다. 같은 방법으로 텍스트 임베딩을 처리하여 언어학적 맥락 벡터(linguistic context vector)를 생성하고, 이를 음성 맥락 벡터와 결합한 후, 인간 평가자의 점수를 예측하도록 선형 회귀 모델을 학습시켰다. Anand et al.(2023)은 별도의 선형회귀 모델 설계 없이

Wav2Vec2.0(Baevski et al., 2020) 표현 특징을 활용하여 음성 시퀀스를 벡터로 변환한 후, 이를 정답 표현 벡터와의 거리로 비교해 가독성(intelligibility)을 평가했다. 발음열의 시간적 변형을 고려하여 정확한 거리를 측정하기 위해 DTW 알고리즘을 사용하였으며, 이는 Park & Glass(2007)의 연구에서 제시된 패턴탐색 기법과 유사한 접근 방식이다. 한편, Bannò et al.(2023)은 자기지도학습 모델이 다양한 음성 언어 처리 문제에 적용될 수 있다는 점에 착안하여, SSL 모델 위에 별도의 평가 모듈(grader)을 추가해 L2 말하기 능력 점수를 산출했다. 사전학습된 모델이 음성인식, 화자인식, 키워드 스팟팅 등 다양한 음성 관련 하위 작업에 적용될 수 있다는 범용성을 확장한 예시다. 위 범용성을 발음 오류 분석 과제에 적용하고자 한 연구들 또한 존재하지만, 자기지도 학습 기법의 효력을 직접 활용하기보다는, 지도 학습 기반의 파인튜닝에 의존하는 경우가 많았다(Peng et al., 2021; Wu et al., 2021; Xu et al., 2021). 모델이 사전에 정해진 음소 라벨 집합을 기반으로 음성을 분류하도록 학습(파인튜닝)되었기 때문에, 평가 시스템 역시 인식된 음소 시퀀스를 정확한 표준 음소 시퀀스와 비교하는 범주형 평가 방식을 벗어나기 어렵다.

본 연구는 기존 연구들과 달리 자기지도 표현 학습의 효력을 직접 활용하여 비범주적 발음 오류 분석을 수행한다. 자기지도학습을 통해 습득된 표현 자질은 음소 라벨 없이 데이터 속성으로부터 유추된 운영 단위이므로 PPG와 달리 범주적 제약을 받지 않는다. 코드벡터는 Wav2Vec2.0모델의 두 가지 표현 학습 자질 중 하나이다. 이를 자기지도학습 자질의 대표적인 예로 채택한 이유는 회수 가능성, 음성학적 유의성, 그리고 음소보다 더 세분화된 단위라는 점 때문이다.

Wav2Vec2.0은 다층 합성곱(CNN) 특징 인코더와 트랜스포머(transformer) 컨텍스트 인코더로 구성된다. 원시 음성 신호가 CNN 특징 인코더를 통과하면 첫 번째로 음성의 본질적인 특징을 표현한 잠재 자질(latent representation)이 학습된다. 표현된 특징에는 음높이, 음색, 스펙트럼 등의 데이터 내재적 정보가 포함된다. 후속되는 트랜스포머 인코더는 잠재 자질을 입력받아 음성이 실현되는 문맥 정보를 학습한 컨텍스트 벡터(context vector)를 산출한다. 문맥 정보의 학습은 마스킹 된 구역에 대응되는 잠재 자질을 올바르게 예측하는 자기지도학습 목표로 설명된다. 이때 잠재 자질을 감독 신호로 활용하고자 양자화한 결과가 코드벡터(discrete latent speech representation). 따라서 Wav2Vec2.0에는 양자화되지 않은 잠재 자질, 코드벡터, 컨텍스트 벡터가 존재한다. 이 중 컨텍스트 벡터는 약한 감독(weak supervision) 아래 생성되므로, 온전히 비지도적인 잠재 자질보다 발음의 원론적인 특성과 덜 밀접하다. 잠재 자질 중에서도 코드벡터는 양음성 샘플을 구분하는 자기지도학습 목표와 직결되므로, 모델이 정의한 발화 구성의 가장 기본적인 단위로 간주될 수 있다. 또한 Wav2Vec2.0 코드벡터는 타 자기지도

학습 모델의 감독신호 자질들과 달리 모델 내부에서 학습되어 회수 가능한 장점이 있다. 대조적으로 비슷한 전방 업무(upstream task)를 지닌 HuBERT(Hsu et al., 2021)는 오프라인에서 진행된 군집화로 감독 신호가 구성되기에 모델로부터 양자화 벡터를 직접 회수할 수 없다.

아울러 Wav2Vec2.0저서 Baevski et al.(2020)에서 진행된 코드벡터의 음성학 유의 검증 실험(phonetic probing)은 코드벡터가 음소와 연관되어있으며 동시에 음소보다 세분화된 단위임을 보여준다. 코드벡터 종류별 음소 분포의 조건부 확률을 들여다보면, 하나의 코드벡터는 하나의 음소를 대표하지만 하나의 음소를 표현하는 데는 여러 개의 코드벡터가 사용되는 비대칭적인 분포를 보인다. 이것은 자질이 음소와 연관되어 있으면서도, 하나의 음소 내에서 나타나는 발음 변이를 세밀하게 구분할 수 있는 더 작은 단위임을 의미한다. 음성학적 유의성은 다국어 사전학습 모델 XLSR-53저서 Conneau et al.(2020)에서 진행된 빈도 분포 조사에서 다시 한번 검증된다. 모델의 사전학습에 활용된 언어 데이터별 코드벡터 빈도 분포를 조사했을 때, 유사한 언어 쌍일수록 빈도 분포가 비슷하게 나타났다. 이는 코드벡터가 각 언어의 공통된 음성 특징을 잘 반영하고 있음을 의미한다.

결국 Wav2Vec2.0 코드벡터 활용 근간에는 범주적 사고를 탈피해 음소 이하의 특성을 살펴볼 수 있는 세분화된 접근성과 비지도 패턴탐색에 적합한 속성을 이미 갖추었다는 점이 존재한다. 이에 착안하여, 자질을 활용한 발음 분석은 선행 연구의 오류 패턴탐색 기법들과 모델의 음성학 유의성 검증실험들을 결합해 진행된다. 먼저 분석에 앞서 Conneau et al.(2020)에서 언어별 변별력을 확인하기 위해 쓴 빈도 조사를 같은 언어 내에서도 L1과 L2 음성을 구별할 수 있는지 확인하기 위해 차용했다. L2 발음 변이를 기술하는 데 쓰일 만큼 자질이 소리 변이의 미묘한 차이를 포착할 수 있을지 사전 검증하기 위한 과정이다. 두 번째로 비범주적 특성 탐구 중에서도 본 연구는 같은 음소적 정의 내 다양성을 반환하고자 하는 Li et al.(2018), Wang & Lee(2013, 2015)와 흐름을 같이 한다. 따라서 해당 연구들과 같이 강제 정렬을 통해 나뉜 표준 음소 구간들을 따로 분석하며, 동일 음소 구간별 음성 프레임열 집합에 Li et al.(2018)의 열 분석론을 적용할 예정이다. 다만 모든 구간의 표본들을 분석한 뒤 접근법과 달리, 범주적 오류와 결부된 구간들로 분석 대상을 한정한다. 본 연구는 범주적 탐지에서 간과되는 오류 세부 양상을 훑아보는 것이 목표이기 때문이다. 따라서 코드벡터를 추출할 사전학습 모델을 파인튜닝 해 분석 표본을 이를 교체오류들을 사전 선별했다. 그 후 강제 정렬 정보를 참조해 표준음소 - 인식결과가 불일치하는 음성 구간들을 기록한 뒤, 상응되는 코드벡터열들을 추출했다. 열분석을 적용해 우세한 패턴들이 기록되면 해석을 위해 파인튜닝에 활용한 L1 데이터의 코드벡터 정보를 참조한다. Baevski et al.(2020)의 자질 종류별 함께 나타나는 음소 분포를 그린 확률 그래프를 재구축해, 음성학적 속성

을 유추하는 데 사용하였으며, 데이터에 현존하는 모든 자질들을 추출한 뒤 군집화해 패턴 간 관계성을 파악했다. 군집화는 앞서 언급되었듯이, 유사성 기반으로 패턴을 찾고자 기존 연구들에서 널리 활용된 방법이다(Li et al., 2018; Mao et al., 2018; Wang & Lee, 2013, 2015).

궁극적으로 L2 발음 오류의 단일 음소 범주 이상의 특성을 비지도 방식으로 찾아낼 수 있는지 실험하고자 한다. 자질이 원어인 음성과 다른 L2 음성 체계의 독자성을 인지하며, 오류의 범주 사이 특징을 환원할 수 있을지 조사할 것이다. 본 연구는 분석 대상이 될 범주적 오류를 교체 오류로 한정하였는데, 삽입 혹은 탈락의 경우 강제 정렬 대응 음성 구간이 불명확하기 때문이다. 또한 교체오류는 관련 범주가 명시되었기에 비범주성을 가시화하기 용이하다. 따라서 교체 범주 사이 오류 패턴 분포를 그리며, 그 양상이 음성학적 근거를 수반하는지 점검해 볼 것이다.

2. 방법론

2.1. 코드벡터 빈도조사

잠재 자질의 양자화는 고차원 벡터를 여러 개의 하위 벡터로 나누고 각 하위 벡터를 별도의 코드북으로 매핑하는 PQ(product quantization) 기법을 따른다. 각 분할분의 양자화를 담당하는 다수(G개)의 코드북이 동시에 학습되며, 하나의 코드북은 소리를 대표할 V개의 코드 단어(code word)을 내포한다. CNN 인코더를 통과한 음성 프레임은 G개의 하위 벡터로 분할되어 코드북 별로 하나의 단어를 선택한다. 그 후, G개의 단어를 연쇄(concatenate)하여 음성 프레임을 대표할 표현이 최종적으로 완성된다. 따라서 코드 벡터의 빈도는 1) 개별 코드 단어 수준과 2) 연쇄된 단어 쌍 수준, 이 두 가지 차원에서 살펴볼 수 있다.

2.1.1. 코드 단어 빈도 조사

양자화에 활용되는 코드 단어 수는 $V \times G$ 개로 한정되어있으며, 각 단어 종류는 1부터 V까지의 인덱스로 반환될 수 있다. 따라서 개별 코드 단어 빈도는 각 코드북에서 채택된 인덱스를 기록한 뒤, 한 코드북 내 V개 단어의 발생 기록을 G번 합산해 도출할 수 있다. 일명 각 차수가 g번째 코드북의 v번째 단어 빈도를 나타내는 $V \times G$ 차원의 벡터로 분포를 그려 비교해볼 수 있다. 이는 Conneau et al.(2020)에서 코드 단어 분포의 언어 간 변별력을 확인하기 위해 그림 1과 같이 시도된 바 있다.

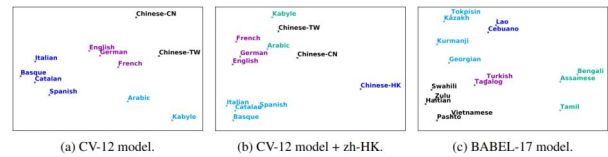


그림 1. Conneau et al.(2020)의 코드 단어 빈도 벡터
Figure 1. Code word frequency vectors in Conneau et al. (2020)

위 예시에서 사전학습 된 다중화자 언어 데이터별로 빈도를 조사했다면, 본 연구에서는 같은 발화목록을 사용하는 L1과 L2 단일화자 데이터에서의 빈도를 조사한다. 개별화자의 빈도 분포가 원어민성을 근거로 나누는지 확인하는 것이 목표다. 이에 조사할 음성파일을 강제 정렬하여 음소 단위로 나누고, 각 음소 구간에서 가장 많이 나타나는 코드벡터 인덱스를 기록하였다. 기록은 인덱스-발생 빈도 쌍의 사전 형태로 json 파일에 저장되었다. 코드북 별로 구축된 빈도 파일은 $V \times 1$ 차원의 확률 벡터로 재구성된다. 이때 사용되지 않은 인덱스는 0의 값이 할당됐다. 그림 2와 3은 빈도 기록 과정과 결과 예시다. 그림 2에서 /s/를 대표할 코드북 1의 22번째 단어와 코드북 2의 234번째 단어가 채택되면 그림3 파일들의 22번째와 234번째 값이 증가할 것이다.

2.1.2. 코드 단어 쌍 빈도 조사

발생 가능한 개별 코드 단어 종류와 달리 발생 가능한 단어 조합의 수는 데이터별로 상이하며, 이론적으로 V^G 종류가 구현될 수 있다. 개별 단어 빈도 조사와 마찬가지로 벡터 구축을 통해 화자 데이터별 자질 이용분포를 비교할 예정이다. L1과 L2 데이터를 아우르는 공통된 공간에서 분포를 비교하고자, 사용된 전체 조합 수를 기준으로 발생 빈도를 측정했다. 분석한 L1 및 L2 발화를 통틀어 5,712쌍의 조합이 관찰되기 때문에 5,712차원의 벡터에 확률분포가 그려졌다.

마찬가지로 강제 정렬된 음성 구간을 대표하는 인덱스를 추출했다. 단, 개별 인덱스 기록이 아닌 쌍으로 회수되므로 그림 4와 같은 기록 형태를 보였다. 그림 4는 그림 2와 같은 음소 /s/에 해당하는 음성 구역의 코드벡터 분포다. 그림 2에서 코드북 1의 22번째 단어와 코드북 2의 234번째 단어가 각 코드북을 대표할 단어로 따로 기록되었다면, 그림 4에서는 두 단어가 [22, 234]라는 짝을 이루며 가장 높은 빈도를 기록한다. 그림 5는 단어 쌍 기록을 저장해 구축된 그림 3과 동일 화자의 빈도 파일이다.

표 1. 분석을 위한 교체 오류 선정

Table 1. Selection of substitution error for the analysis

| 정답 발음 | 인식 결과 | 빈도 | 언어학적 유의성 |
|-------|-------|--------|----------------------------|
| Z | S | 11,204 | 모국어 유성 마찰음 부재 |
| T | *** | 7,668 | - |
| L | R | 5,549 | 모국어 단일 유음 대응 음소 |
| AE | EH | 6,966 | 모국어 모음체계 긴장-이완 대립 부재 |
| *** | IH | 4,791 | 음절 구조 차이 |
| N | *** | 3,386 | - |

*** 대응되는 음소가 부재함.

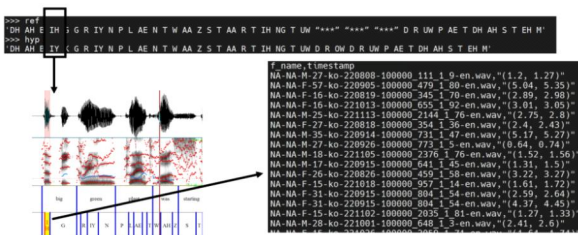


그림 6. IH에서 IY 교체오류의 시간 기록
Figure 6. Recording of time stamps in IH to IY

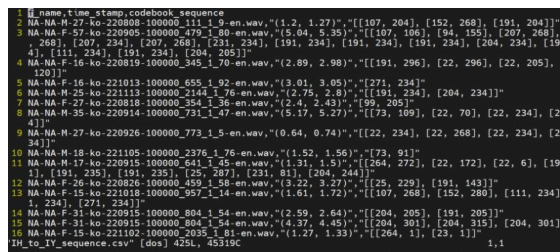


그림 7. IH에서 IY 교체오류 표본 코드벡터열
Figure 7. Code vector sequences of IH to IY substitution

이후, 추출된 코드벡터열 들로부터 우세한 패턴을 도출해 오류의 음소 이하 유형 분화를 시도했다. Li et al.(2018)에서 프레임별 PPG 군집 ID 열로부터 패턴을 탐지한 방법이 사용됐다. 해당 열분석법은 현 연구 과제와 지정된 구역을 표현하는 프레임별 인덱스들로 유형을 도출한다는 목표를 공유한다. Li et al.(2018)은 데이터 전체 음성 프레임들의 PPG를 군집화한 후 부여된 인덱스(cluster ID) 열을 그림 8과 같은 정제 과정을 통해 패턴화했다. 그 과정은 세 단계로 나뉜다. 먼저 1) 최소 빈도를 지닌 인덱스를 제거하고 인근의 동일 인덱스들을 하나로 통합한다. 이렇듯 한 번 정제된 표현 열들로부터 2) 우세한 유형들이 1차적으로 선출된다. 마지막으로 선출된 유형 내에서 3) 하위 부분 열(subsequence)들은 상위 열로 통합된다. 위 단계들을 거쳐 패턴 유형을 도출하는 방법을 본 연구에서는 그림 7과 같은 표본들에 적용했다. 그 결과, 분석한 교체오류 모두 3가지 대표 인덱스로 요약될 수 있었다.

한편 L2 데이터 내부 통계만으로 도출된 패턴의 표현은 자의적이다. 따라서 패턴들의 해석 가능성을 높이기 위해 L1 외부 참조자료를 별도로 구축했다. 참조자료는 L1에서 패턴 인덱스가 1) 어떤 벡터값을 표현하며 2) 확률적으로 어떤 음소들과 대응되는지 알려준다. 이를 구축하기 위해 파인튜닝에 사용된 동일 데이터로부터 현존하는 1) 모든 코드벡터를 수적으로 기록하였으며, 코드벡터 별로 2) 함께 발생하는 음소들의 확률분포를 구했다. 동일 데이터 사용은 범주적 평가와 비범주적 평가 사이의 일관성을 유지해 준다.



그림 8. Li et al.(2018)의 cluster ID 열 정제법
Figure 8. Sequence Filtering and Representation in Li et al. (2018)

또한 코드벡터 참조자료는 단어 수준과 단어 쌍 수준 두 가지 차원에서 구축된다. 단어 단위의 참조자료를 구축하는 이유는 L1에서 사용되지 않는 조합의 속성을 확인해야 하기 때문이다. 선례 없는 조합은 실험 2.1에서 L1과 L2 데이터 사이 빈도 분포 차이를 통해 확인된 L2 특징이다.

그 결과 그림 9와 같은 총 3종류의 벡터 공간이 완성된다. 단어 쌍 벡터는 추가로 39개의 집단으로 군집화된다. 이는 후술할 사용된 음소 집합의 크기이며, 패턴 벡터가 분화되는지 시험할 최소기준이다. 군집화는 수적인 기록이 패턴 간 관계성을 파악하는데 궁극적으로 동원된다는 점과 관련된다.

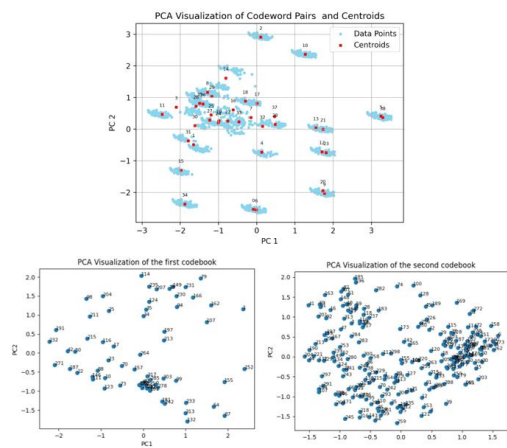
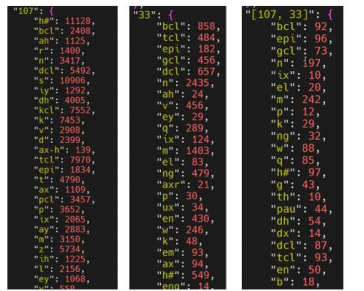


그림 9. L1 참조 자료의 코드 단어 및 단어 쌍 벡터 공간들
Figure 9. Raw code vector spaces in L1 reference material

음소 분포 조건부 확률 또한 개별 단어와 함께 발생하는 음소들의 기록과 연쇄된 형태에서 동시 기록되는 음소 분포 자료가 따로 구축되었다. 사용목록 빈도 조사와 유사하게, 인덱스별로 같이 기록되는 음소들의 빈도를 그림 10과

같이 json 파일에 저장하였다. 기록된 원시 빈도는 최종적으로 확률분포로 정규화된다. 연쇄된 형태의 음소 분포 조사는 Baevski et al.(2020)의 유의성 탐색 $P(\text{phoneme} | q)$ 그래프의 횡단면을 재구축한 자료다. 그 관계성은 그림 11에서 확인할 수 있다. 두 탐색에 사용된 코퍼스 또한 같다. 단, Baevski et al.(2020)에서 라벨 된 기호를 39개로 축약해 확률분포를 통합했다면, 본 연구에서는 전체 61개의 음소 라벨 기호들을 모두 동원하였다. 이는 음소 이하 특성 파악을 위해 더 세밀한 분석을 가능하게 하기 위한 설계이다.



codebook1 codebook2 codeword pair

그림 10. 코드벡터 별 공동 발생 음소 빈도
Figure 10. Frequency of co-occurring phonemes per code vector

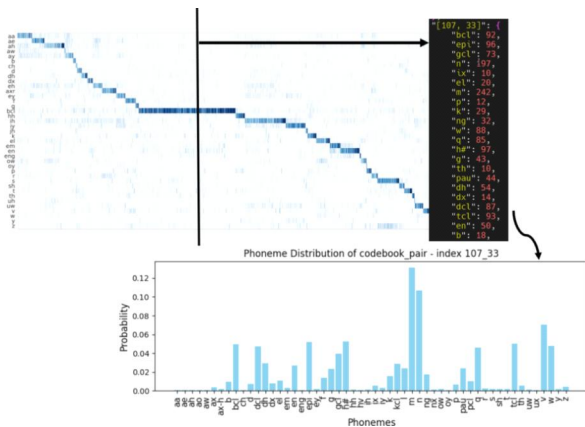


그림 11. 음소분포 참조자료와 Baevski et al.(2020)의 유의성 탐색
Figure 11. Relationship with $P(\text{phoneme} | q)$ plotting in Baevski et al. (2020)

3. 실험

3.1. 데이터

빈도 조사와 오류 분석을 위한 데이터 선정에는 다른 기준이 적용된다. 우선 빈도 조사는 음소 구간 당 대표 인덱스 기록으로 수행되기 때문에, L1과 L2 비교 데이터들이 같은 음소 열을 다뤄야 했다. 따라서 내용적 통제가 이루어진 낭독체 발화여야 하는 것이 첫 번째 선정 기준이었다. 또한 그림 1과 같은 다중화자 데이터 간 거시적 비교가 아닌, 같은 언어 내 특정 화자 집단에서 관찰되는 특성을 조사하고

자 했다. 따라서 화자 단위로 충분한 음성을 확보할 필요가 있었다. 이에 1,132개 문장을 지닌 동일 발화목록 ARCTIC prompt를 낭독하는 단일화자 데이터 CMU ARCTIC(Kominek & Black, 2004)과 L2 ARCTIC(Zhao et al., 2018)이 채택되었다. CMU ARCTIC의 경우 스코틀랜드, 인도 화자를 제외한 북미 원어민 화자의 발화만 사용됐다. 본 연구는 미국영어 표준 발음을 기준으로 측정된 변이양상을 분석하기 때문이다. 사용된 L2 ARCTIC 버전 5.0에는 힌디어, 베트남어, 한국어, 만다린어, 아랍어, 스페인어 모국어 화자들의 영어 발화 데이터가 포함된다. 각 L1 언어마다 4명의 모국어 화자가 낭독한 음성 데이터가 존재했다. 아울러 MFA(Montreal forced aligner; McAuliffe et al., 2017)를 이용해 생성된 Praat (Boersma, 2001) TextGrid 포맷의 강제 정렬 라벨링이 내포돼 있었다. 이는 빈도 조사에서 대표 인덱스를 기록하기 위한 기준으로 사용된다. 사전 생성된 강제 정렬 라벨이 부재한 CMU ARCTIC에서는, L2 ARCTIC에서 사용된 것과 동일한 방법으로 강제 정렬 작업을 별도로 수행했다. 이때 활용된 발음 사전과 음향 모델은 L2 오류 분석 데이터를 강제 정렬할 때도 적용된다.

한편 오류 분석을 통해 보편적인 결과를 도출하기 위해서는 특정 화자의 음성만 포함한 데이터보다 다양한 사람들의 음성을 포괄하는 데이터가 유용했다. 이런 측면에서 NIA037 교육용 한국인 영어 음성 데이터는 다양한 배경과 영어 숙련도를 가진 한국인 학습자들의 발음 데이터를 포함하여, L2 음성 특징을 분석하는 데 적합했다. 따라서 훈련 데이터 중 문장 단위로 발화된 부분을 L2 분석 대상으로 선정하였다. 문단 단위 발화는 MFA 시스템에 활용되기 위해 각 문장에 일일이 라벨을 붙이는 수작업이 필요했으므로, 자동화된 분석에는 적합하지 않아 제외되었다. 아울러 신뢰성 있는 참조자료 구축을 위해서 음소별 시간 정보가 정확하게 표기된 L1 데이터가 필요했다. 이에 다양한 화자의 음성을 포함하고 있으며 음소 단위까지 수동으로 주석된 TIMIT(Garofolo et al., 1993)이 선정되었다. TIMIT은 북미 8개 지역 방언의 630명 화자 발화 데이터를 포함하고 있다. 표 2는 실험 별 사용된 데이터들을 요약한다.

표 2. 실험 별 사용된 L1 및 L2 데이터
Table 2. The used L1 and L2 data per experiment

| 실험 | 2.1 빈도조사 | 2.2 ~ 3 오류 분석 |
|----|--|---|
| L1 | CMU ARCTIC (5명 북미 화자: bd1, slt, jmk, rms, clb) | TIMIT (과인튜닝: train split, 참조자료 구축: train & test split) |
| L2 | L2 ARCTIC v5.0 (6개 모국어 배경을 지닌 24명 화자) | NIA037 교육용 한국인의 영어 음성 데이터 (낭독체 발화의 train split 중 문장 단위 발화) |

3.2. 실험 환경

모든 실험은 fairseq 프레임워크(Facebook AI Research,

2019)를 기반으로 진행되었으며, LibriVox (LV-60k) 데이터에 사전학습된 LARGE 모델을 사용했다. 이는 Baevski et al.(2020)에서 음소적 유의성 검증이 진행된 사전학습모델이기도 하다. 범주적 오류탐지를 위해 TIMIT 학습 데이터를 활용하여 다음과 같은 파라미터 설정으로 파인튜닝을 진행했다: CTC criterion, 40000 max update, 3e-4 learning rate, update frequency=4, adam optimizer(betas: 0.9-0.98, eps:1e-08), tri stage learning rate scheduler(ratio 0.1,0.4,0.5), mask probability 0.65, mask channel probability 0.5, mask channel length 64. 파인튜닝 라벨 생성을 위해 먼저 Lee et al.(2018)의 g2p 모델을 이용하여 26개 알파벳을 문장 강세 기호를 포함한 39개의 ARPAbet 기호로 변환했다. 이 중 문장 강세 정보를 제외한 아래 음소 집합이 최종 활용된다: AA, AE, AH, AO, AW, AY, B, CH, D, DH, EH, ER, EY, F, G, HH, IH, IY, JH, K, L, M, N, NG, OW, OY, P, R, S, SH, T, TH, UH, UW, V, W, Y, Z, ZH. 강세 정렬 시 파인튜닝 라벨과 동일한 음소 집합을 사용하고자 MFA의 English(US) ARPA v3.0.0 발음 사전과 음향 모델을 음성 데이터에 적용한 뒤, 강세를 제거했다. 마지막으로 코드벡터 군집화와 시각화 모두 faiss라이브러리(Facebook Research, 2024)를 활용했다.

3.3. 평가 매트릭

DBI(Davies-Bouldin index)는 군집 간 분리도와 군집 내 응집도의 상대적 크기를 비교하여 군집화 성능을 평가하는 지표다. 본 평가 매트릭은 각 군집의 중심 간 거리와 군집 내 데이터의 평균 거리의 비율로 계산되며, 값이 작을수록 군집화 결과가 우수함을 의미한다. 식 (1)에서 N은 군집의 수, c는 군집 중심(centroid), S는 각 군집 내 평균 산포도를 나타낸다.

$$DBI=1/N \times \sum_{i=1}^N \max_{i \neq j} (S_i + S_j / d(c_i, c_j)) \quad (1)$$

4. 실험 결과

4.1. L1 L2 화자별 코드벡터 이용 현황

L1-L2 음성별로 다른 코드벡터 활용 양상을 통해 자질이 L2 음성 체계를 독자적으로 인코딩하는지 확인할 수 있다. 활용되는 코드 단어(쌍) 종류의 빈도 조사는 빈도 벡터의 군집화를 통한 거시적 검토와 개별화자의 사용목록을 상호 비교하는 미시적 검토로 구성된다. 먼저 그림 12와 13은 군집화 결과의 시각화 자료다. 화자별 빈도 분포 데이터가 산점도 상에 표시되었으며, 각 데이터점 옆에는 해당 화자의 이름이 병기되어 있다. 또한, 각 화자가 속한 군집에 따라 데이터점의 색상을 달리하여 시각적으로 구분하였다.

두 그림의 군집화 결과 모두 원어민과 비원어민 발화를 성공적으로 구분하고 있다. L2 ARCTIC 화자들의 분포 점들은 갈색 군집에, CMU ARCTIC 화자들의 분포 점들은 하

늘색 군집에 속하며, 두 화자 그룹은 서로 다른 군집 중심(centroid)에 연계되어 있다. 이때 군집 간 분리도는 단어 차원에서보다 단어 쌍 차원에서 더 높다. 실제 DBI 점수로 산포도 비율을 살펴보면, 그림 12의 단어 빈도 벡터는 1.038을, 그림 13의 단어 쌍 빈도 벡터는 0.822의 점수를 기록했다. 다시 말해, 개별 단어의 사용보다는 단어 조합의 형태에서 L2 학습자의 특징이 분명하게 드러난다. 따라서 단어 쌍 차원에서 화자별 사용 패턴을 심층적으로 분석해볼 필요가 있다.

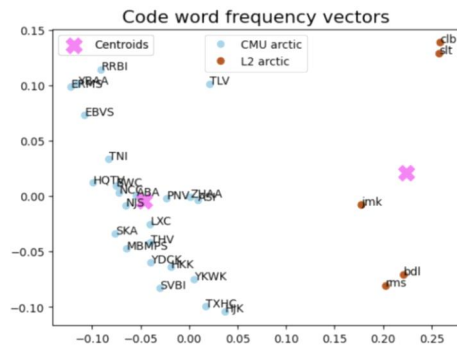


그림 12. 코드 단어 빈도 벡터의 시각화와 군집화
Figure 12. Visualization of code word frequency vectors clustering

단어 빈도 기반의 분류 역시 원어민과 비원어민 간의 차이를 어느 정도 반영하지만, 그림 1에서의 분류만큼 명확한 결과를 보여주는 못한다. 같은 언어 내 발음 변이양상에 따른 코드 단어 인코딩 차이가, 사용되는 음소 집합 자체가 다른 언어들 사이 인코딩 차이에 비해 미세하기 때문에 해석할 수 있다.

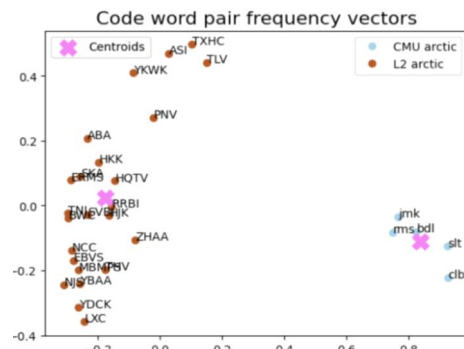
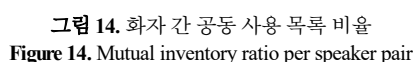


그림 13. 코드 단어 쌍 빈도 벡터의 시각화와 군집화
Figure 13. Visualization of code word pair frequency vectors clustering

화자별 사용 패턴 분석을 위해, 각 화자가 사용한 단어 쌍의 공유 정도를 비율로 계산하여 그림 14의 열지도(heat map)에 나타냈다. 열지도에서 각 셀의 색깔은 두 화자 간의 코드벡터 사용 패턴 유사도를 나타낸다. 이때 비율은 가로 축의 기준 화자가 사용한 모든 단어 쌍 개수를 분모로, 세

두 번째로 주목할 점은 두 집단 소속 화자들 사이 교차점 중에서도 오른편 CMU ARCTIC 화자들을 기준으로 계산된 공유 비율이 하단의 L2 ARCTIC 화자들을 기준으로 계산된 공유 비율보다 낮다는 것이다. 다시 말해, 같은 비교 대상들 사이 계산된 비율이더라도, L2 화자의 사용목록 수가 분모로 올 때 비해 L1 화자의 사용목록 수가 분모로 올 때 더 작은 값이 계산된다. 따라서 L1 화자들의 사용목록 크기가 L2 화자들의 사용목록 크기보다 전반적으로 클 것임이 예상된다.



L2 ARTIC
Low-level speakers

| | BWC | LXC | NCC | TXHC | ABA | SKA | BAA | ZHAA |
|-------------|-----|------|-----|------|------|------|------|------|
| # of tokens | 973 | 1074 | 981 | 1248 | 1165 | 1080 | 1123 | 1403 |

L2 ARTIC
Intermediate-level speakers

| | ASI | RRBI | SVBI | TNI | HJK | HKK | YDCK | YKWK |
|-------------|------|------|------|------|------|------|------|------|
| # of tokens | 1164 | 893 | 1443 | 1180 | 1389 | 1301 | 1491 | 1291 |

L2 ARTIC
High-level speakers

| | EBVS | ERMS | MBMPs | NJS | HQTV | PNV | THV | TLV |
|-------------|------|------|-------|------|------|------|------|------|
| # of tokens | 964 | 1125 | 1085 | 1387 | 1103 | 1163 | 1449 | 1276 |

L1 CMU ARTIC

| | jmk | bdI | rms | clb | slt |
|-------------|------|------|------|------|------|
| # of tokens | 1516 | 1911 | 1633 | 2637 | 3133 |

그림 15. 화자별 코드 벡터 사용목록 크기
Figure 15. The number of code vector inventory per speaker

조사 결과, 5명 CMU ARTIC 화자들의 사용목록이 24명 L2 ARCTIC 화자들의 목록보다 더 큰 점이 확인되었다. 나아가 언어 수준이 높은 그룹일수록 더 다양한 종류의 조합을 사용했다. 화자 RRBI를 제외하면, 파란색 하위 그룹보다는 초록색 중위 그룹이, 초록색 중위 그룹보다는 보라색 상위그룹이 더 많은 수를 기록한다. 다시 말해, 언어 수준에 비례해 증감하는 활용 목록의 다양성을 엿볼 수 있다. 해당 경향성은 원어민 화자들을 실력이 가장 높은 기준점으로 둘 때, 더욱 일관된 근거를 확보한다. 90점 미만의 낮은 발화 수준부터 L1 화자 수준의 표준 발음에 이르기까지, 숙련도가 높아질수록 활용되는 코드백터 목록의 크기가 증가하기 때문이다. 위 관계성을 통해 자질이 발음 변이의 정도를 인지하고 있다고 추론할 수 있다. 코드백터는 사전학습 데이터인 L1을 기준으로 습득된 음성 단위이므로, 표준 발음으로부터 편차가 심한 낮은 수준의 발화일수록, 학습된 단위들을 온전히 구현하기 어려울 것이다. 반면 학습 수준이 높아질수록 L1 음성 체계를 더욱 정확하게 구현할 수 있게 되어, 다양한 코드 단어 조합을 활용하여 더욱 풍부한 표현이 가능해진다. 결국, 자질은 비원어민 발화를 원어민 발화로부터 구분하는 기능뿐만 아니라, 학습 수준에 따른 발음 편차를 점진적으로 식별해낼 수 있는 잠재력을 보유하고 있다. 따라서 오류의 범주적 정의 이하 다양성을 탐색하기 위한 적합한 수단이 된다.

| Speaker | L1 | Gender | TOEFL iBT |
|---------|----------|--------|-----------|
| HKK | Korean | M | 114 |
| YDCK | Korean | F | 110 |
| BWC | Mandarin | M | 80 |
| LXC | Mandarin | F | 86 |
| YBAA | Arabic | M | 100 |
| SKA | Arabic | F | 79 |
| EBVS | Spanish | M | 70 |
| NJS | Spanish | F | 110 |
| RRBI | Hindi | M | 91 |
| TNI | Hindi | F | 99 |

그림 16. Zhao et al.(2018)의 L2 ARTIC 화자 정보
Figure 16. Demographic information of L2 ARTIC in Zhao et al. (2018)

4.2. 범주적 교체 오류 선정

앞서 L2 발음 오류의 범주 이하 세부성을 조사하기 위해, 언어학적 근거를 갖춘 교체 오류 목록을 선정해야 함을 기술한 바 있다. 표 3은 파인튜닝 된 모델과 현존하는 L2 문헌들을 활용해 최종 선정된 오류들이다. 교체에 참여한 음소 쌍들은 마찰음, 유음, 모음의 범주를 아우르며 여섯 가지 경로로 하위 분류된다.

먼저 마찰음은 오류 빈도가 가장 높은 음소 종류였으며, 선정된 네 가지 오류 Z to S, DH to D, V to B, F to P는 Hong et al.(2014)의 음소인식 기반 코퍼스 분석에서도 한국어 학습자의 자음 발음 변이 현상 중 가장 두드러지게 나타난다고 보고되었다. 영어의 마찰음 /s/, /z/, /ʃ/, /ʒ/, /t/, /v/, /θ/, /ð/, /h/에 비해 한국어의 마찰음 /s/, /s=/, /h/는 종류가 제한적이다. 이러한 음소대립의 차이로 인해 L1-L2 음소 사이 음향적 거리가 멀어 오류가 발생하기 쉽다. 특히 유성 마찰음은 한국어 자음체계에 부재한 유성과 무성의 대립을 포함하고 있어 조음적 어려움을 야기하며(Smith & Swan, 2001), 이로부터 두 가지 교체 경로가 예상된다. 첫 번째로 유성 마찰음 Z가 조음 위치와 방법은 유지하되, 유성성이 결여된 S로 실현된 경우가 있다. 이는 한국어 대응 음소인 치경 마찰음이 모두 무성음(/s/, /s=/)인 점과 관련된다. 두 번째로 조음 위치와 유성성은 유지한 채, 조음 방법만 마찰음에서 파열음으로 바뀌어 DH, V, F가 각각 D, B, P로 실현된 경우가 있다. 한국어 파열음은 마찰음보다 영어 대응 음소와의 음향적 거리가 적은 편이다. 따라서 음향적으로 더 가까운 조음 방법으로 교체가 일어난 예시들이며, 이는 Schmidt(1996)의 한국어 학습자를 대상으로 한 인지 실험에서 DH, V, F를 각각 /d/, /t/, /f/으로 식별했다는 결과와도 상응된다. 따라서 본 예시들은, 모국어 음운 체계의 영향으로 인해 학습자들이 익숙한 조음 방식으로 외국어 소리를 변형시키는 현상을 보여준다.

표 3. 선정된 교체 오류 목록

Table 3. Inventory of selected substitution errors

| 음소 종류 | 교체 경로 | 오류 |
|-------|-----------------------|--|
| 마찰음 | 유성음>무성음 | Z to S |
| | 연속성>불연속성 (마찰음>파열음) | DH to D V to B F to P |
| 유음 | 설측성 +/- | L to R R to L |
| 모음 | 긴장성 +/- | IH to IY IY to IH AE to EH EH to AE |
| | 중모음>저모음 | AH to AA |
| | 이중모음>단모음 | EY to EH OW to AO |

유음 역시 L1과 L2 음성 체계 차이로 인한 발음 변이가 관찰된다. 단일 음소 /l/ 이 분포 환경에 따라 [ɾ]과 [l] 변이

음으로 실현되는 한국어와 달리, 영어의 설측음 /l/과 권설음 /ɫ/은 독립적인 음소들이다(Kim & Rhee, 2019). 따라서 /l/과 /ɫ/사이의 교체는 일관되지 않은 대응 관계에 근원을 둔다. 아울러, 같은 r음일지라도 영어의 권설음 [ɹ]은 탄설음 [t]과 다른, 한국인 화자에게 생소한 조음 방법을 활용한다. 반면, /l/은 두 음성 체계에 모두 현존하는 무표적 음이다. 이와 관련해 Koo(2012)에서는 한국인 학습자들이 /l/의 발음을 /ɫ/보다 쉽게 익힐 수 있다고 분석한 바 있다. 한편 Kim & Rhee(2019)는 학습 수준이 높을수록 유음 발음 정확도가 향상되는 경향을 보인다는 연구 결과를 통해, 유음 발음이 전체 발음 능력을 대변할 평가 지표로 활용될 수 있음을 시사했다. 이러한 언어 습득론적 유의성은 유음 교체오류 표본들이 범주 이하 변이 분석에 적합한 대상임을 알려준다.

선정된 모음 교체오류는 세 가지 이유로 발생한다. 첫 번째는 모국어 모음체계에 없는 긴장과 이완의 대립을 혼동하는 경우다. IH[i]에서 IY [i]로의 변화(혹은 그 반대)와 AE[æ]에서 EH[ɛ]으로의 변화(혹은 그 반대)가 첫 번째 교체 경로의 예시들이다. 양 교체 방향은 긴장성 유무라는 공통된 요인 아래에 발생하기 때문에 짝지어 연구되어왔다(Koo & Oh, 2001; Yang, 2010 등). 그중에서도 Yang(2010)은 국어에 없는 이완 /i/와 긴장 /æ/가 /i/와 /ɛ/보다 더 큰 조음적 어려움을 동반할 것임을 언급했다. 이는 Tsukada et al.(2005)의 지각 실험에서 /i/와 /æ/가 /i/와 /ɛ/로 혼동됐지만, /i/와 /ɛ/는 정확히 식별된 점과도 관련된다. /æ/와 /ɛ/ 사이의 비대칭적인 관계는 Koo & Oh(2001)에서도 확인된다. 연구에 의하면, 한국인의 영어 /æ/ 발음은 일반적인 /æ/의 저모음 특성과 달리 /ɛ/에 가까운 중모음으로 발음되는 경향을 보였다. 두 번째 교체 유형은 후설 모음의 높이에서 일어난다. Tsukada et al.(2005)는 /i/-/i/, /ɛ/-/æ/ 외에도 한국인 학습자들이 지각하기 어려워 혼동하는 영어 모음 쌍으로 /a/-/ʌ/를 제시했다. 이는 Koo & Oh(2001)에서 두 모음/a/와 /ʌ/가 유사한 후설 중모음 구역에서 발음된다고 관찰된 점과 일맥상통한다. 따라서 두 모음 사이 교체 양상 내에서도 본래 중모음의 음향 값을 갖는 /ʌ/로 조음이 편향될 것임이 예상된다. AH/ʌ/에서 AA/a/로의 교체는 타 음소인식 기반 발음 오류 분석 Hong et al.(2014)에서도 가장 현저한 모음 변이 현상 중 하나로 기록되었다. 마지막 모음 교체 유형은 이중모음 EY/eɪ/와 OW/oʊ/가 대응 단모음 EH/ɛ/와 AO/ə/로 축소되어 발음되는 경우다. 이중모음은 하나의 음절 안에 두 개의 모음 소리가 연이어 발음되면서, 조음 기관의 위치가 이동하는 특징을 가진다. 초급 학습자일수록 이중모음 발음 시 혀의 이동 거리가 길어지고, 각 구성 요소를 분절적으로 발음하는 경향이 있다(Oh, 2013; Park, 2001). 이때 길게 끊어 발음되는 두 음을 시간 제약 내에 실현하기 어려울 시, 인접한 단모음으로 대체될 가능성이 있다. Cho & Jeong(2013)의 연구는 이러한 현상을 /oʊ/가 /ə/로 발음되는 경우를 통해 확인하였다. 같은 맥락에서, Choi & Oh(2001)은 한국인 학습자들이 /eɪ/를 /ɛ/와 유사한 /ɛ/로 발음하는 경향이 있다

는 점을 밝혀냈다.

4.3. 범주 하위 오류 패턴 분석 결과

범주 이하 오류 패턴 분석을 통해 다음 세 가지 목표를 달성하고자 한다. 첫째, 오류의 범주를 넘어서는 특성이 두 교체 범주 사이에서 어떻게 구현되는지 확인할 것이다. 둘째, 이를 통해 코드벡터가 비범주성을 어떻게 인지하는지 알아보고자 한다. 셋째, 이러한 구현 양상이 기존에 연구된 L2 발음 변이 현상과 어떤 관련이 있는지 비교 분석할 예정이다. 패턴 분석은 내적 우세 유형을 도출하고, 외적 참조자료를 활용하여 유형 간 관계성과 음성학적 속성을 유추하는 두 단계로 구성된다. 그림 17은 이러한 분석 과정을 시각적으로 보여준다.

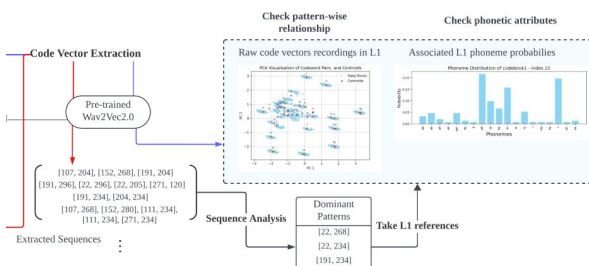


그림 17. 분절 이하 패턴 분석 개요
Figure 17. Sub-segmental pattern analysis schema

일례로 Z to S 오류 표본들은 열분석을 통해 [166, 82], [18, 51], [230, 51]의 세 가지 주요 유형으로 구분된다. [166, 82]와 [230, 51]은 충분한 거리로 분리되는 21번째와 28번째 군집에 속하며, [18, 51]은 L1 참조 데이터에서 관찰되지 않는 조합이다. 패턴 인덱스 간 거리는 그림 18의 군집화 시각 자료로 확인할 수 있다. 나아가 인덱스당 함께 발생하는 음소 확률분포를 통해, 패턴들의 음성학적 속성을 유추했다. [166, 82]와 [230, 51]은 S로의 교체 양상을 반영하며 모두 L1에서 ‘s’와 가장 높은 연계성을 지니지만, 두 번째로 높은 확률에서 차이를 보인다. 그림 19의 분석 결과, [166, 82] 유형은 ‘z’와, [230, 51] 유형은 ‘sh’와 각각 두 번째로 높은 상관관계를 보였다. [18, 51] 조합은 L1 데이터 내 정확한 매칭 항목이 없었기 때문에, 개별 코드북의 음소 확률분포를 분석하는 간접적인 방법을 통해 속성을 추정했다. 그 결과 [166, 82]와 [230, 51]의 속성을 부분적으로 공유하는 중도 유형임을 알 수 있었다. 코드북1의 18번째 인덱스에서는 ‘sh’음소가 높게 나타나 [230, 51]과 유사했지만, 코드북2의 51번째 인덱스는 오히려 [166, 82]와 유사한 음소 분포를 보였다.

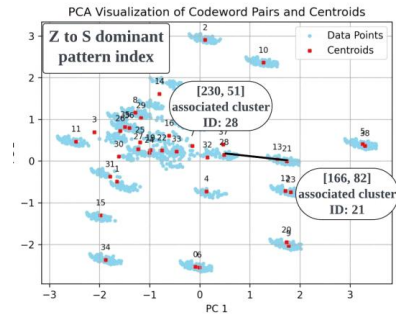
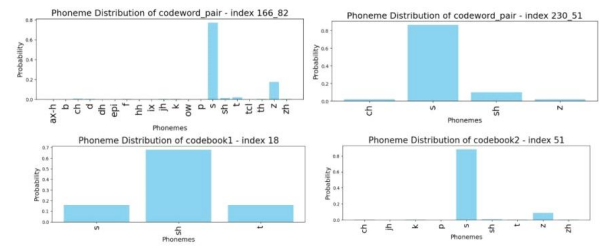


그림 18. 패턴 간 거리 확인 예시
Figure 18. Exemplary probing of pattern-wise distance



의미한다. 이는 한국어 마찰음 목록이 무성음들로만 구성되어 학습자들이 유성 마찰음 발음에 어려움을 겪는다는 4.2의 내용과 일관된다. 이처럼 1) 교체된 특성을 따라 형성된 연속체, 2) 비범주적인 중도 유형, 3) 모국어 음성 체계와 연관된 패턴 분포의 비대칭성은 후술할 세 종류의 교체 경로에서도 공통으로 관찰되는 특징들이다.

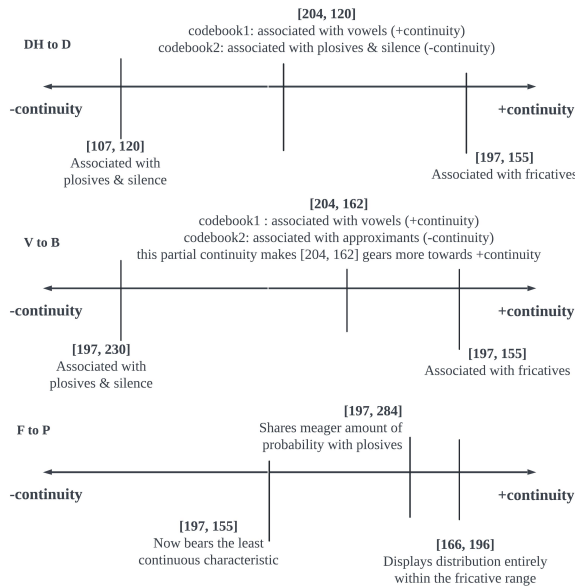


그림 21. DH to D, V to B, F to P 교체오류 패턴 분포
Figure 21. Discovered patterns in DH to D, V to B, F to P substitution

마찰음에서 파열음으로 조음 방법이 변하는 두 번째 교체 경로는 연속성(continuity)의 결여로 요약될 수 있다. 이와 관련된 세 오류 DH to D, V to B, F to P 모두 1) 파열음과 무음 특징이 나타나는 불연속적인 패턴에서 마찰음과 모음 특징이 나타나는 연속적인 패턴으로 이어지는 연속체를 형성한다. 2) 이때 각 연속체 내 중도 유형들은 코드북1에서는 연속적인 특성을, 코드북 2에서는 불연속적인 특성을 나타내며 양면적인 성격을 띤다. 3) 또한 DH to D와 V to B 교체오류의 패턴 분산도가 F to P 교체오류의 패턴 분산도보다 컸다. 이 같은 비대칭성은 무성 마찰음에 비해 큰 유성 마찰음의 조음적 어려움을 반영한다. 목표하는 표준 발음이 발음하기 어려울수록, 의도된 특성과 상이한 값이 실현될 확률이 높기 때문이다. 따라서 DH와 V의 오류 패턴은 F의 오류 패턴보다 표준 발음에서 더 멀리 떨어진 영역에 분산되어 분포하는 경향을 보이게 된다. 그림 21은 무성 마찰음 F의 발음 오류가 의도된 연속성을 가장 가깝게 구현했음을 보여준다. 유성 마찰 연속체 내 가장 연속적 성격의 패턴 [197, 155]이 F to P 연속체에서는 가장 비연속적 성격의 패턴을 이룬다. 또한 F to P 연속체 내 가장 연속적인 패턴 [166, 196]은 대부분의 분포를 같은 무성 마찰음인 ‘s’와 공유하고 있다(그림 22). /s/는 모국어 마찰 음소 목록 /s/ /ʃ/와 가장 유사한 대응 음소기에, 오류 패턴 분산도의

차이가 학습자의 L1 음성 체계와 관련된다는 주장을 뒷받침한다. 다시 말해, 패턴의 분포는 대응하는 데 생기는 어려움을 반영한다. V와 DH보다 F는 /s/ /ʃ/와 음향적 거리가 적기 때문에, 연속성을 향한 응집력이 더 강하다. 이러한 분산도 차이는 그림 23의 벡터 공간상 유클리디안 거리와도 일관된다. 패턴을 대표하는 코드벡터들의 소속 군집들 사이 거리를 비교해 보면, DH to D와 V to B에 비해 F to P가 더 작다.

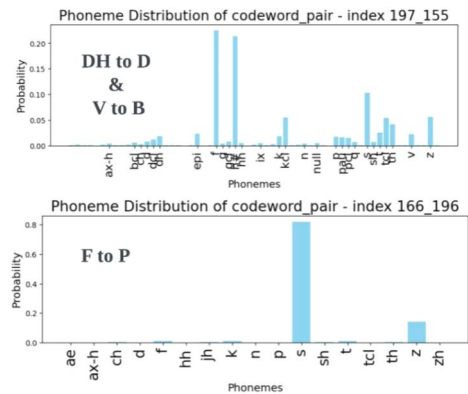


그림 22. 마찰음에서 파열음 교체 연속체 내 연속적 패턴들
Figure 22. [+continuity] end in the fricative to plosive error continuum

그림 22, 24, 25는 그림 21에 언급된 불연속적, 연속적, 중도 유형 패턴들의 음성학적 속성을 보여준다. 연속성 척도는 각 오류 연속체 내 상대적인 크기로 측정되므로, 서로 다른 연속체 간의 절대적인 수치를 동일선상에서 평가하기 힘들 수 있다. 예를 들어 그림 25를 보면, V to B의 중도 유형 패턴 [204, 162]는 코드북 2에서 접근음과 연계된 상대적 불연속성을 구현한다. 그러나 이는 DH to D 중도 유형 패턴 [204, 120]을 구성하는 코드북2의 파열음 목록 연계보다는 연속적일 수 있다.

[204, 162]가 [204, 120]에 비해 교체 연속체 내에서 연속적인 패턴과 더 가까운 점은 그림 23의 유클리디안 거리 간 상대적 비교로도 확인할 수 있다. V to B에서 [204, 162]는 [197, 230]에 비해 [197, 155]와 가깝다. 반면, DH to D의 [204, 120]은 다른 두 패턴 벡터 [107, 120], [197, 155]와 유사한 거리를 유지한다.

한편 상대적 특성 비교를 통해 얻은 공통적인 발견은 중도 유형의 상극이 클수록 L1 발생 빈도가 낮아진다는 것이다. 상반된 특성의 조합이 곧 비범주성을 의미할 때, 비범주적인 L2 발음 오류의 특성이 L1에 나타나지 않기 때문이라 해석될 수 있다. 예를 들어, 두 코드북에서 연속성 값의 차이가 가장 큰 [204, 120]은 TIMIT 데이터 전체에서 ‘n’ 음소를 표현하는 데 단 한 번만 사용되었다. 이는 비강에서 연속적인 공명을 일으키고 구강에서는 불연속적인 폐쇄를 동반하는 /n/ 음소의 복합적인 조음 기제와도 일관된다.

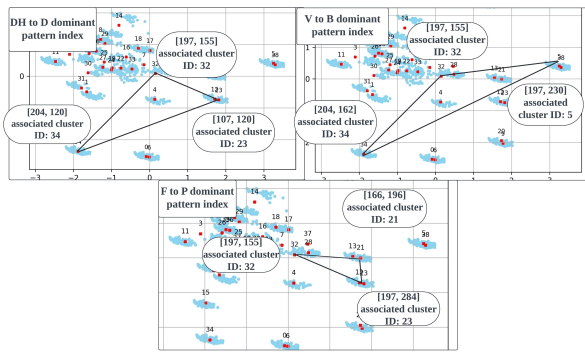


그림 23. DH to D, V to B, F to P 교체오류 패턴들 사이 유클리디안 거리
Figure 23. Euclidean distance among discovered patterns in DH to D, V to B, and F to P substitution

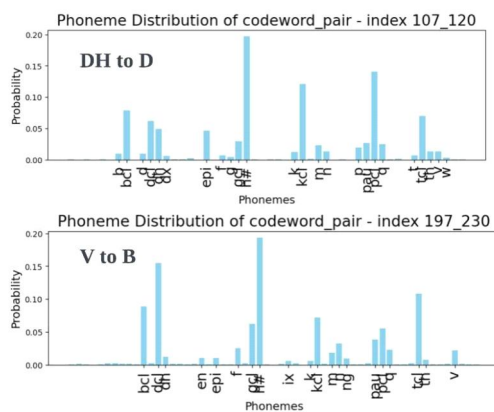


그림 24. 마찰음에서 파열음 교체 연속체 내 불연속적 패턴들
Figure 24. [-continuity] end in the fricative to plosive error continuum

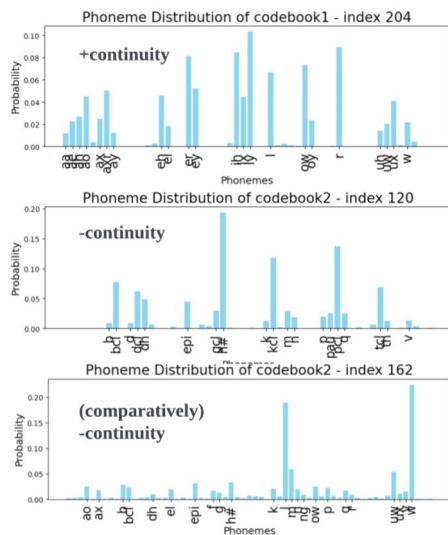


그림 25. 마찰음에서 파열음 교체 연속체 내 중도 유형 패턴들
Figure 25. Intermediaries of the fricative to plosive error continuum

유음 교체는 공기 이동 통로의 변화로 설명된다. 설측음 /l/이 혀끝을 치경에 댄 채, 양측 통로로 공기를 통과시키는

반면, 권설음 /ɹ/은 혀뿌리를 말아 올리고 혀끝을 내려 중앙 부로 공기를 통과시킨다. 이러한 조음 위치 차이는 공기 흐름의 저항을 감소시켜 /l/에서 더 강한 공명을 유발하고, 결과적으로 /l/이 모음과 유사한 넓은 저주파 성분을 가지게 한다. 따라서 패턴을 대표하는 코드벡터가 모음과 높은 확률을 공유하면, 설측성이 결여되어 권설음의 특성에 더 가까워진 것으로 해석될 수 있다.

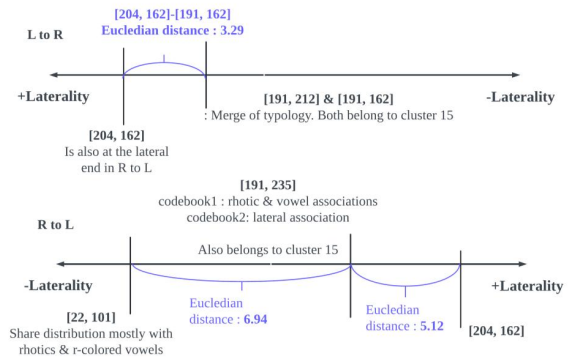


그림 26. L to R, R to L 교체오류 패턴 분포
Figure 26. Discovered patterns in L to R, R to L substitution

1) 그림 26는 설측음과 연계된 패턴에서 시작하여 권설음, 모음의 특징을 더 많이 포함하는 패턴으로 이어지는 오류 연속체들을 보여준다. 양방향의 교체 모두 동일한 L1-L2 음성 체계들의 상호작용 하에 일어나는 만큼, 패턴 분포가 상당수 겹친다. [204, 162]는 두 연속체에서 모두 가장 설측성의 성격을 지닌 패턴을 대표하고 있다. 또한 L to R의 중도 유형 패턴 벡터 [191, 212], [191, 162]와 R to L의 중도 유형 패턴 벡터 [191, 235] 모두 15번 군집에 속하므로, 유사한 특징을 가질 것으로 예상된다. L to R에서는 두 패턴 벡터가 동일한 군집에 할당되어 유형 합류 현상을 보인다. 그림 27과 28은 그림 26에 기술된 연속체 양 끝 패턴들의 음성학적 속성을 보여준다.

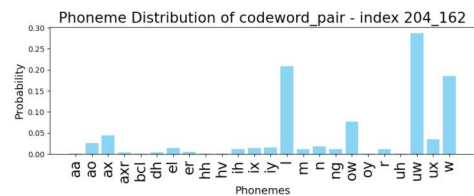


그림 27. 유음 교체 연속체 내 설측성 패턴
Figure 27. [+laterality] end in the liquid error continuum

2) 유음 교체의 중도 유형 또한 비범주적인 속성을 보이며, L1 데이터에서 희박하다. 그림 29의 분석 결과, 코드북1의 191번 인덱스는 권설음과 r-화된 모음과 관련이 깊은 것으로 나타나지만, 코드북2의 162번과 235번 인덱스는 설측

음과 높은 상관관계를 보인다. 하단 코드북2 벡터 공간을 보면, 162, 212, 235번 단어들은 유클리디안 거리가 가까워서 밀집해 있다. 이를 통해 세 코드 단어 모두 설측성과 관련된 음향적 특성을 공유하고 있음을 추측할 수 있다. 궁극적으로 코드북1의 권설음 특징과 코드북2의 설측음 특징을 동시에 갖는 중도 유형 조합들은 L1 TIMIT에서 매우 드물게 나타난다. [191, 162]는 ‘l’을 표현하는데 두 번, ‘uw’를 표현하는데 한 번 등장하며, [191, 235]는 세 모음 ‘ih’, ‘uh’, ‘uw’을 표현하는데 각각 한 번씩 등장한다.

3) 아울러 L to R과 R to L의 중도 유형 모두 설측성 끝으로 치우친 분포를 보인다. 분포의 비대칭성은 그림 26에서 파란색 글씨와 기호로 표기된 벡터 간 유클리디안 거리 계산을 통해 확인 가능하다. 해당 계산치는 그림30의 벡터 공간 위 매핑을 통해 시각화된다. 패턴들의 소속 군집 위치를 기점으로 세로축을 따라 형성된 위계는 그림 26의 설측성 연속체와 일치한다. 교체 오류 내에서도 발음 변이가 설측음 쪽으로 더 기울어진 경향을 보이는 것은, 모국어에 없는 권설음을 발음하는 것이 /l/을 발음하기보다 더 어렵기 때문이라 해석할 수 있다. 이는 4.2절에서 Koo(2012)가 제시한 연구 결과와도 일치한다.

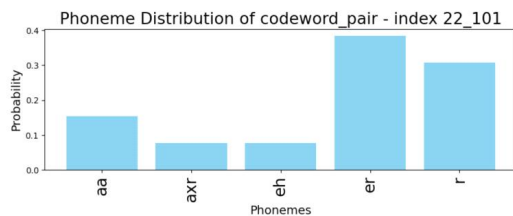


그림 28. 유음 교체 연속체 내 비설측성 패턴
Figure 28. [-laterality] end in the liquid error continuum

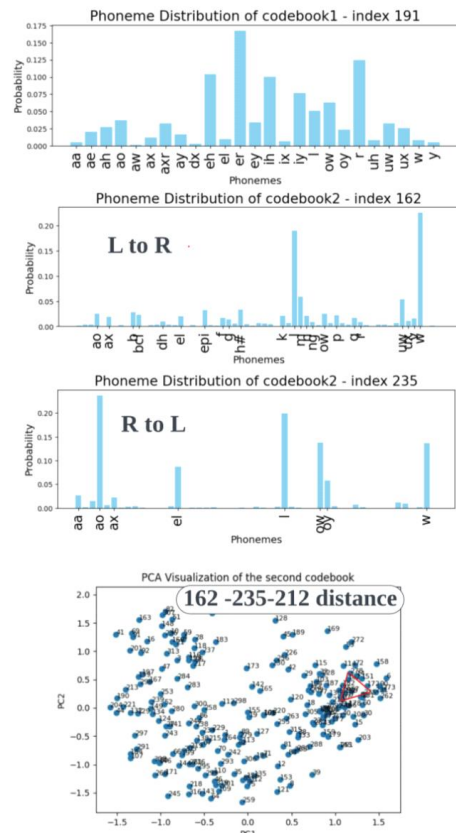


그림 29. 유음 교체 연속체 내 중도 유형 패턴들
Figure 29. Intermediaries of the liquid error continuum

긴장-이완 모음 교체와 이중모음 축소 오류는 패턴 표본이 부족한 특수성으로 인해 5장에서 논의될 예정이다.

세 번째로, 1) AH to AA 오류 패턴들 역시 교체된 속성인 모음 높낮이를 따라 형성된 연속체를 이룬다. 그림 31을 보면, 패턴 벡터들이 AH와 같은 높이인 AO에 비해 상대적으로 얼마나 더 AA와 공동 발생할 확률이 높은지에 따라 연속체 내 위상이 결정된다. 이때 2) [264, 6]과 [264, 131]은 같은 군집에 속하므로 하나의 유형으로 통합되며, 중도 유형인 만큼 코드북1과 2에서 AH 대비 AA와 연계되는 상대적 비율이 다르다. 그림 32에서 확인할 수 있듯이, 코드북 1의 264번째 단어는 AO 비율이 더 높지만, 코드북 2의 131번째 단어는 AA 비율이 더 높다. 범주 사이 위치 내에서도 3)중도 유형 [264, 6], [264, 131]은 그림 33의 높은 패턴 [232, 131]과 인접해 있으며, 따라서 패턴들의 분포는 높은 값 쪽으로 치우쳐 있다.

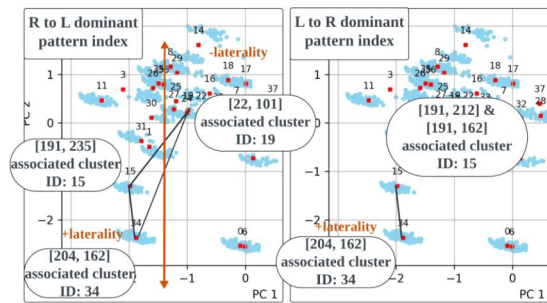


그림 30. L to R, R to L 교체 오류 패턴들 사이 유클리디안 거리
Figure 30. Euclidean distance among discovered patterns in L to R and R to L substitution

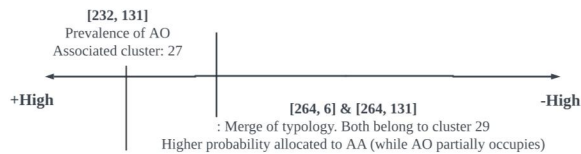


그림 31. AH to AA 교체오류 패턴 분포
Figure 31. Discovered patterns in AH to AA substitution

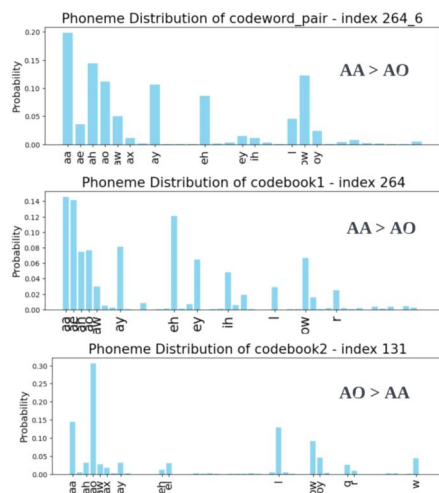


그림 32. 후설 모음 높이 교체 연속체 내 중도 유형 패턴
Figure 32. Intermediary of the vowel height error continuum

해당 비대칭성은 패턴 벡터들 사이 유클리디안 거리와 음성학적 분포 속성이라는 두 가지 차원에서 모두 확인 가능하다. 우선 그림 34를 참조해보면, [264, 6]이 속한 29번째 군집과 [232, 131]이 속한 27번째 군집은 서로 매우 가까운 거리에 위치한다. 나아가 [264, 6]는 AA와의 연계성이 AO보다 높지만 AO와도 확률을 일정량 공유하고 있다. 따라서 [232, 131]과 음성학적 거리가 크지 않을 것이 예상된다. 다시 말해 음성학적 이해로 유추된 패턴 사이 거리와 패턴을 대표하는 코드벡터 사이의 유클리디안 거리가 유사하다. 교

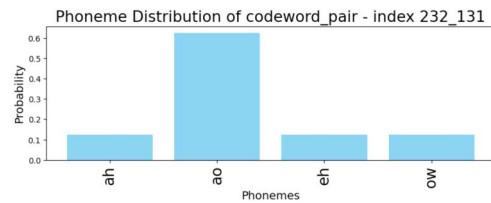


그림 33. 후설 모음 높이 교체 연속체 내 높은 패턴
Figure 33. [+high] end in the vowel height error continuum

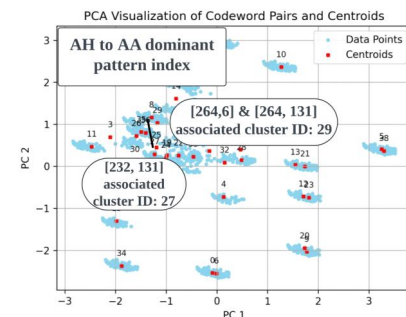


그림 34. AH to AA 교체 오류 패턴들 사이 유클리디안 거리
Figure 34. Euclidean distance among discovered patterns in AH to AA substitution

5. 논의

4.1절의 주요 발견은 L1과 L2 발화에서 코드백터의 활용 방식이 서로 다르다는 것이었다. 화자별 코드백터 사용 목록 분석 결과, L1과 L2 두 화자 집단은 빈도뿐 아니라 자질의 종류에서도 차이를 보였다. 특히 L2 화자의 학습 수준이 높아질수록 L1 화자와 공유하는 코드 단어 조합이 증가하는 경향이 관찰되었다. 따라서 학습 수준이 낮은 L2 화자일수록 L1 화자들이 사용하지 않는 비표준적인 코드 단어 조합을 사용할 가능성이 클 것이 예상된다. 이러한 현상은 4.3절에서 비범주적 중도 유형으로 분석된 바 있다. 두 코드북 단어들에서 상반된 값을 가지는 조합들은 L1 참조 데이터에 거의 나타나지 않았으며, 나타나더라도 매우 낮은 빈도로 관찰되었다.

한편 긴장-이완 교체 오류와 이중모음-단모음 오류는 중도 유형 패턴뿐만 아니라 모든 오류 표본에서 상반된 코드 단어 조합을 보였다. 따라서 L1 데이터 부족으로 인해, 개별 코드 단어 분석을 통해 조합된 패턴의 특징을 추론하였다.

긴장-이완 교체 오류의 패턴 벡터들은 코드북 1에서 긴장성을, 코드북 2에서 이완성을 더 강하게 나타내는 공통적인 특징을 보였다. 그중 EH to AE는 예외적인 활용 양상을 보였으나, 후에 더 자세히 논의될 예정이다. 긴장성과 이완성

값은 표 4의 긴장/이완 모음과 코드 단어의 연관성을 비율로 계산하여 도출되었다. 각 코드 단어의 긴장 모음과의 공동 출현 확률을 이완 모음과의 공동 출현 확률로 나누어 긴장-대-이완 비율을 계산했다. 이 비율은 표 5 상단에 나타나 있다. 긴장-이완 교체 패턴들을 구성하는 코드 단어들은 코드북 1보다 코드북 2에서 높은 긴장-대-이완 비율을 갖는 점을 확인할 수 있다.

코드 단어 쌍의 긴장성 비율을 종합하여 비교하면, 그림 35 상단에 3개의 연속체가 완성된다. 해당 연속체들은 긴장성이라는 음성학적 이해로 계산된 위계를 나타낸다. 그리고 본 위계는 벡터들 사이 유클리디안 거리를 반영한다. 그림 36의 상단 눈금은 세 교체오류를 구성하는 패턴들의 위계를 종합하고 있다. 긴장성 대소 비교를 바탕으로 구축된 거리 들은 파란색 유클리디안 거리와 비례한다. 예를 들어, 코드북 1과 2에서 각각 0.315와 9.744의 비율을 가지는 [22, 234]와 0.511과 9.744의 비율을 가지는 [191, 234] 사이의 거리는 4.715다. 이는 조합된 값들 사이의 간극이 큰 [22, 268] (0.511-9.744) 와 [191, 234](0.315-2.918) 사이의 거리 7.44보다 작다.

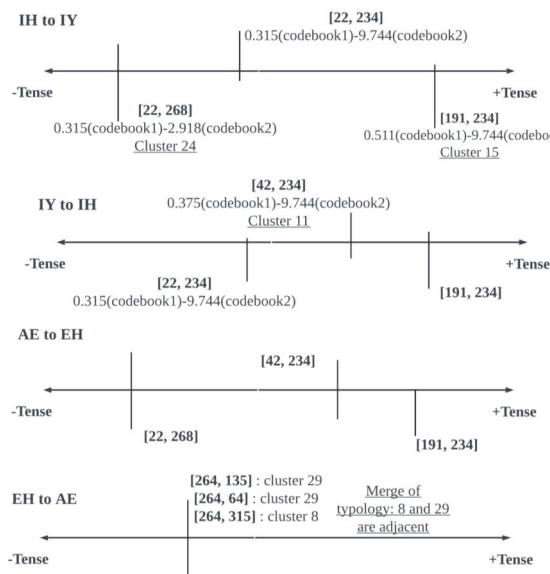


그림 35. 긴장-이완 교체 오류 패턴 분포
Figure 35. Discovered patterns in tenseness substitution

한편, EH to AE는 다른 교체 패턴과 달리, L1 데이터가 풍부한 조합들로 구성된다. 해당 조합들은 매우 인접한 8번째 또는 29번째 군집에 속하며, 하나의 유형으로 통합된다. 이 유형은 긴장성이 강한 15번 군집([191, 234])과는 가장 멀리 떨어져 있고, 이완성이 강한 24번 군집([22, 268])과는 가장 가깝다. 또한, 긴장성 연속체에서 중간 정도의 긴장도를 나타내는 11번 군집과는 중간 거리를 유지한다. 따라서 통합된 패턴은 이완성의 성격이 강하다 볼 수 있으며, 그림 35 하단의 연속체에서 이완성 끝으로 치우쳐 있다. 8번째-

29번째 군집과 11, 15, 24번 군집 간의 유클리디안 거리는 그림 36 하단에 시각적으로 나타나 있다.

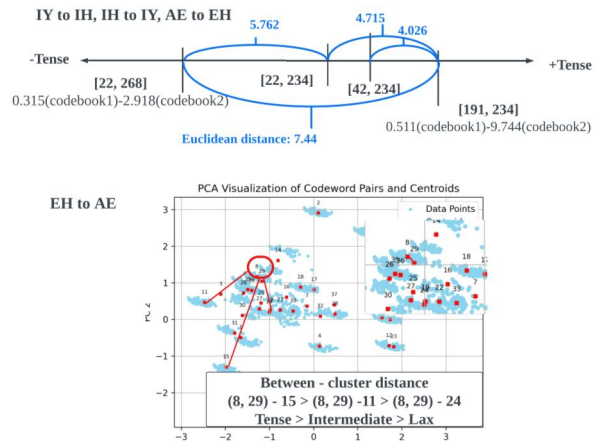


그림 36. 긴장-이완 교체 오류 패턴들 사이 유클리디안 거리
Figure 36. Euclidean distance among discovered patterns in tenseness substitution

그림 35에 제시된 4개의 연속체들을 종합해 볼 때, IH와 IY, EH와 AE 음소 쌍들은 교체 방향 사이 비대칭성을 보인다. 즉, 한 방향의 교체 패턴들이 표준 발음을 향한 더 높은 응집력을 보인다면, 반대 방향의 오류 패턴들에서는 더 분산된 경향이 관찰된다. 이러한 현상은 4.2절에서 언급된 한국인 학습자들이 발음하기 어려워하는 모음과 관련이 있다. IH/i/와 AE/æ/은 IY/i/와 EH/ε/에 비해 학습자들이 어려움을 겪는 발음들이다. 따라서 IH/i/와 AE/æ/가 목표음인 IH/i/ to IY/i/와 AE/æ/ to EH/ε/ 오류에서 패턴들이 더 큰 변동성을 보인다고 해석할 수 있다. 반면 발음하기 상대적으로 쉬운 IY/i/와 EH/ε/가 목표 음인 IY/i/ to IH/i/와 EH/ε/ to AE/æ/의 오류 연속체에서는, 패턴들이 의도된 속성과 비슷한 값을 상대적으로 일관되게 구현한다.

아울러 EH/ε/ to AE/æ/의 특수성은 Yang(2013)의 연구에서 제시된 미국인과 한국인의 /æ/와 /ε/ 발음 차이와 관련이 있다. 두 L1과 L2 화자 그룹 사이 /æ/와 /ε/의 상대적 발음 편차는 상반된 경향성을 보인다. 연구에 의하면, AE/æ/는 미국인의 발음 편차가 한국인보다 크지만, EH/ε/의 발음 편차는 한국인이 더 크다. 따라서 미국인 발음의 LibriVox 데이터로 사전 훈련된 모델은 /æ/와 비슷한 발음 오류 들을 더 넓은 범위에서 /æ/로 인식할 것이다. AE/æ/를 표현하는 코드벡터들의 종류가 다양할 것이므로, L2 오류와 대응되는 표본들이 타 긴장-이완 교체 오류들보다 많았을 것이라 이해된다. 또한 AE/æ/로 인식되는 기준이 유연하므로, 패턴 벡터들 모두 인접한 하나의 유형으로 통합될 수 있었을 것이다. 반대로 모델의 EH/ε/ 인식 범위는 한국인 학습자의 발음 변이 범주보다 상대적으로 협소하다. 미국인의 /ε/ 발음은 더 일관된 경향을 보이기 때문에, 이에 맞추어 학습된

모델은 /e/와 유사한 소리들을 더 좁은 범위에서만 /e/로 인식할 것이다. 따라서 AE/æ/ to EH/ε/ 오류에서는 패턴 벡터의 발생 빈도와 벡터들 사이 분산 측면에서 반대의 경향이 관찰되었을 것이다.

표 4. 긴장성 및 전후설성 계산 기준 음소 목록

Table 4. Associated phonemes for tenseness/backness calculation

| 지표 | 긴장성 |
|-------|---|
| 긴장모음 | ae, aw, ay, ey, iy, ow, oy, uw |
| 이완모음 | aa, ah, ao, ax, axr, eh, er, ih, ix, uh, ux |
| 지표 | 전후설성 |
| 전설 음소 | eh, er, ey, ih, iy, y, l, ix, el, dx, b, bcl, f, p, n, m, v, dcl, dh, pcl, tcl, t, z, nx, em, th, s |
| 후설 음소 | aa, ae, ah, ao, aw, ow, uh, uw, ay, oy, w, ux, r, hh, q, k, hv, ng, g, gcl |

표 5. 코드 단어별 긴장성 및 전후설성 계산 수치

Table 5. Calculated tenseness and backness values of code words

| 지표 | 긴장-대-이완 비율(tense-to-lax ratio) |
|------|--|
| 코드북1 | 191 (0.511)>42 (0.375)>22 (0.315) [-tense] |
| 코드북2 | 234 (9.744)>268 (2.981) [+tense] |
| 지표 | 후설-대-전설 비율(back-to-front ratio) |
| 코드북1 | 191 (1.687)>232 (0.7)>22 (0.559) [-back] |
| 코드북2 | 212 (2.084) [+back] |
| 지표 | 전설-대-후설 비율(front-to-back ratio) |
| 코드북1 | 42 (2.666)>22 (1.788) [-back] |
| 코드북2 | 234 (11.363)>319 (3.606) [+back] |

앞서 이중모음은 두 조음 위치 사이의 이동을 동반하며, 이동 경로를 빠르게 실행하는 어려움이 단모음 분절로 이어짐을 설명했다. Oh(2013)을 비롯한 선행 연구들에서는 이동 거리를 전후설성 지표인 제2 포먼트 값의 변화량으로 측정해 분석했다. F2 변화가 클수록 분절이 예상되기에, 자질이 전설음과 후설음의 특징을 얼마나 양 코드북에서 일관되게 구현하는지(전후설성 결의도)를 분석하여 단모음 교체 가능성을 탐색해볼 수 있다. 이중모음의 이동 출발점인 음절핵의 상대적 위치를 반영해, OW 오류 패턴 벡터들의 후설성으로의 결의 여부와, EY 오류 패턴 벡터들의 전설성으로의 결의 여부를 각각 조사했다. 패턴을 구성하는 두 코드 단어의 값이 조화를 이룰수록, 이중모음이 성공적으로 발음됐을 것이며, 반대로 차이가 클수록 단모음으로 대체될 가능성이 클 것이 예측된다.

결의도는 표 4에 명시된 전설 또는 후설 음소와 코드 단어가 함께 나타날 확률을 비율로 계산하여 도출된다. 전설성 결의도는 전설 음소와 코드 단어가 함께 나타날 확률의 합을 후설 음소와 코드 단어가 함께 나타날 확률의 합으로 나누는 값이며, 표 5 하단의 전설 대 후설 확률 비로 표현된다. 반대로, 후설성 결의도는 후설 음소와 공동으로 나타날 확률의 합을 전설 음소와 공동으로 나타날 확률의 합으로 나누어 계산되며, 표 5 중단 옆에 후설 대 전설 확률 비로 표현된다. 이처럼 전후설성 결의 척도로 계산된 위계는, 패턴 벡터의 L1 발생 빈도와 긴장성 등급(tenseness ranking)과

유의미한 상관관계를 보였다.

그림 37은 후설 이중모음 OW/ou/가 AO/o/로 교체되는 오류의 패턴 분포를 보여준다. 상단의 오류 연속체는 조합된 코드 단어들의 후설-대-전설 비율을 종합해 후설성 결의 척도를 비교하고 있다. 연속체를 이루는 패턴 중 [191, 212]는 가장 높은 후설-대-전설 비율을 두 코드북에서 일관되게 구현하며 후설성 결의의 끝에 위치한다. 코드북1의 191번째 단어 비율 1.687은 코드북2의 212번째 단어 비율 2.084와 가장 조화를 이룬다. 이는 [232, 212]에서 0.7과 2.084 사이의 간극과, [22, 212]에서 0.599와 2.084 사이의 간극보다 훨씬 적은 수치다. 앞서 언급되었듯이, 두 코드 단어의 상반된 값은 곧 표본이 적은 비범주성을 의미한다. 결과적으로 가장 조화를 이루는 [191, 212]는 다른 두 패턴 벡터들보다 월등히 높은 L1 발생 빈도를 가지게 된다.

발생 빈도의 차이는 코드벡터들 사이 유클리디안 거리와 비례한다. 하단의 벡터 공간 그림을 보면 [232, 212]가 속한 27번째 군집과 [22, 212]가 속한 24번째 군집은 서로 인접해 있다. 그러나 두 군집은 발생 빈도 차이가 큰 [191, 212]가 속한 15번째 군집과는 멀리 떨어져 있다. 결의도는 곧 발생 빈도와 연관되기에, 이 같은 발견은 코드벡터가 패턴의 전후설성 결의도를 수치적으로 인지하고 있음을 시사한다. 궁극적으로 발생 빈도에 따라 벡터 간 거리가 달라진다는 것은, 자질이 음성의 전후설성 결의 여부와 같은 음성학적 이해를 반영하고 있음을 드러낸다.

더불어 가장 이중모음 조음과 속성이 비슷한 [191, 212]은 앞서 긴장성과 결부되었던 군집 15에 속하며, 가장 단모음 조음과 속성이 비슷한 [22, 212]는 이완성과 결부되었던 군집 24에 속한다. 이 같은 긴장성과 이중모음 조음의 관계성은 본 분석법이 음성학적 특징들을 일관되게 반영하고 있음을 보여준다.

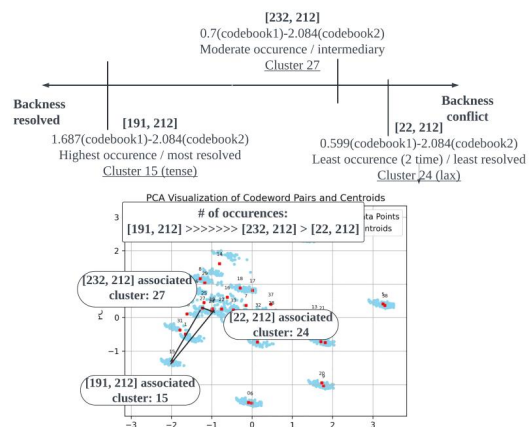


그림 37. OW to AO 교체오류 패턴 분포
Figure 37. Discovered patterns in OW to AO substitution

두 번째로 전설 이중모음 EY/ei/가 EH/ε/로 교체되는 오류의 패턴들은 전설성 결의도를 따라 오류 연속체를 형성

했다. 그림 38의 상단 눈금은 조합된 코드 단어들의 전설-대-후설 비율을 종합해 패턴들의 연속체 내 위치를 표현한다. 후설 이중모음 오류와 유사하게, 패턴들의 위계는 긴장성 등급과 일관되었다.

먼저 가장 높은 전설-대-후설 값을 두 코드북에서 구현하는 패턴 [42, 234]는, IY/i/ to IH/i/ 오류 연속체 속 중도 유형 패턴과 일치한다. 나아가 중간 순위의 결의도를 기록하는 [22, 234]는 IH/i/ to IY/i/에서 [42, 234]에 비해 긴장성의 성격을 덜 띠는 조합이다. 마지막으로 가장 전설성으로 결의되지 못한 [22, 319]는 IH/i/ to IY/i/에서 가장 이완성과 결부된 24번째 군집 소속이다. 이로써 긴장-이완 연속체 속 위계와 EY/ei/ to EH/e/ 오류 패턴의 전설성 결의도 사이에 강한 상관관계가 존재함을 알 수 있다. 다시 말해, 두 코드 단어의 값이 전설성으로 결의된 정도는 긴장성의 성격을 띠는 정도와 비례한다. 그림 38은 두 종류의 연속체들을 동일선상에서 비교해, 그 관계성을 시각화하고 있다.

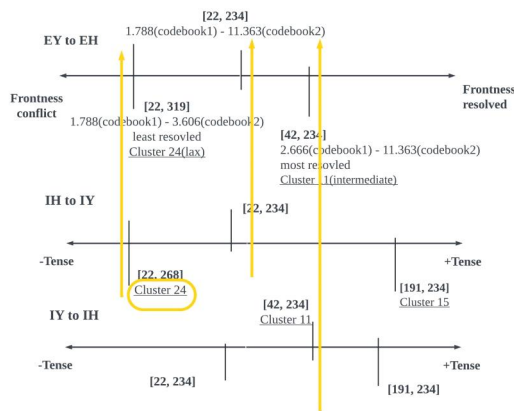


그림 38. EY to EH 교체오류 패턴 분포
Figure 38. Discovered patterns in EY to EH substitution

그러나 전설 이중모음 오류 패턴은 후설 이중모음 오류와 달리 모든 유형에서 L1 발생 빈도가 현저히 낮았다. 이는 Koo & Oh(2001)의 연구에서 분석한 한국인 학습자들의 모음 조음 특성과 관련이 있는 것으로 보인다. 본 연구 결과에 의하면, 한국인의 영어 모음 공간은 미국인에 비해 좁게 형성되어 있으며, 특히 전설 모음 구역에서 그 차이가 두드러진다. 상대적으로 협소한 L2 조음 범위는, L1 발음과의 차이로 인해, 참조 데이터의 대응 코드벡터 표본을 부족하게 만든다.

아울러, 제한적인 조음 공간은 L2 학습 시 새로운 소리를 정확하게 구분하고 생성하는 데 방해가 되어, 소리 간의 대립이 불분명해지는 현상을 초래할 수 있다. 이와 관련해 IH to IY, IY to IH, AE to EH, EY to EH 오류 패턴들은 서로 다른 목표 음소를 지님에도 불구하고 상당수 겹쳤다. 그림 35와 38을 참조해보면, 연속체 위 패턴 벡터들이 반복되는 것을 확인할 수 있다.

결국 L1과 L2 모음 공간의 차이로 인해, 긴장-이완 교체와 이중모음 축소 오류 패턴들은 L2 학습자만의 독특한 코드 단어 조합을 사용하게 된다. 모음은 자음보다 발음 변이에 더 취약하므로, 이처럼 L2 고유의 자질 목록을 사용할 가능성이 크다. 일각에서는 AH to AA 패턴들이 목표 발음과 비슷한 속성을 지니며 L1에 풍부한 코드벡터들로 구성된 점을 근거로 반론을 제기할 수 있다. 그러나 AH는 중모음으로, 분석된 타 모음 오류들에 비해 모음 공간 축소의 영향을 상대적으로 덜 받는다.

본 장에서는 대부분의 패턴 벡터가 비범주적인 특성을 보여 4.3절에서 제시된 단어 쌍 수준의 분석이 어려운 사례들을 다루었다. 논의된 두 교체 경로는 긴장성과 전후설성이라는 음성학적 특징을 기반으로 분석되었으며, 각기 다른 특징을 중심으로 분석된 오류 패턴들의 순위는 긴장성과 이중모음 조음의 상관관계를 반영하였다. 즉, 긴장성이 높은 패턴에서 전후설성이 가장 결의되었으며, 긴장성이 낮아질수록 결의도 또한 낮아졌다. 이는 코드벡터가 다양한 음성학적 특징들을 일관되게 반영하고 있다는 것을 의미한다. 또한 후설 이중모음 오류에서는 발생 빈도와 코드벡터의 위치가 비례했는데, 이는 낮은 발생 빈도의 원인인 비범주성을 자질이 수치적으로 인지하고 있다는 것을 보여준다. 결론적으로, 개별 단어의 음성학적 특징만으로도 오류 패턴을 예측할 수 있었으며, 예측된 패턴의 분포는 단어 쌍 차원의 분석과 유사하게 코드벡터 사이의 거리와 비례했다.

6. 결론

본 연구는 L2 발음 오류의 음소 이하 특징을 살펴보고자 자가지도 학습 표현 자질, Wav2Vec2.0 코드벡터를 활용했다. 코드벡터는 모델 저서의 음성학 유의 검증실험으로 알 수 있듯이, 언어별 음성 체계의 특징을 반영하며 음소와 밀접하게 연관되어 음성학적으로 의미 있는 단위이다. 동시에 외부 지식 없이 모델 자체적으로 학습된 단위이기에, 음소에 의해 제약받지 않는다. 이러한 특성을 응용해, 범주적 정의만으로 설명되기 어려운 발음 오류의 특징을 구체화했다. 그리고 그 과정에서 코드벡터가 표현 가능한 발음 변이의 범위를 탐색했다.

연구 결과 첫째, 코드벡터가 L1과 L2 음성을 다르게 인코딩하여, 원어민과 비원어민 음성을 구분할 수 있음을 확인했다. 이 같은 L2 식별력은 L1과는 달리 L2 음성에서만 발견되는 독특한 자질 목록에 의해 발생했다. 해당 목록들을 구성하는 코드 단어 조합들은 상반된 음성학적 특징을 두 코드북에서 구현했다. 이들은 어떠한 특징의 범주에도 속하지 못하기 때문에 비범주성의 전형이라고 볼 수 있다. 비범주성이 L2 발음 오류의 특징인 만큼, 발음 숙련도가 낮을수록, 비범주적 조합 표본들의 사용이 증가하는 경향을 보였다. 나아가 패턴 분석을 통해, 같은 음소로 잘못 발음하는 오류일지라도, 오류의 정도에 따라 점진적으로 분화될

수 있음을 찾아냈다. 이는 발음 오류가 단순히 하나의 범주로 묶일 수 있는 것이 아니라, 다양한 하위 유형으로 분류될 수 있음을 시사한다. 범주적으로 동일한 오류 표본들은 변화된 음성학적 속성의 구현 척도를 따라 오류 연속체를 형성했다.

분석한 교체오류 예시 중 마찰음, 유음, 중모음 오류들은, 패턴을 대표하는 코드 단어 조합이 L1 참조 데이터에 충분했다. 따라서 단어 쌍 차원에서 분석이 가능했으며, 조합된 단어들의 음소 분포 조건부 확률을 참조해, 패턴들의 음성학적 속성을 유추했다. 그 결과 교체된 속성을 기준으로 오류의 유형이 분화될 수 있음을 확인했다. 다음은 네 가지 교체 경로의 예시들이다. 1) 유성 마찰음에서 무성 마찰음으로 교체가 일어난 Z to S는 유성성의 연속체 내에서, 2) 마찰음에서 파열음으로 교체가 일어난 DH to D, V to B, F to P는 연속성의 연속체 내에서, 3) 유음 교체 L to R 과 R to L은 설측성의 연속체 내에서, 4) 모음 높이가 교체된 AH to AA는 모음 높이의 연속체 내에서 패턴들이 분화됐다. 이때 연속체 속 중도 유형은 변화된 속성(유성성, 연속성, 설측성, 모음 높이)에 대해 두 코드북에서 상반된 값을 구현한다. 이는 앞서 설명된, 두 범주에 걸쳐있는 L2 발음의 특징을 보여준다. 따라서 비범주적 특징을 갖는 중도 유형 패턴 벡터들은, L1 참조자료에서 잘 관찰되지 않았다. 중요한 것은, 코드벡터와 공동 발생하는 음소 확률분포로 유추된 패턴들 사이의 관계가 벡터 공간 위 유클리디안 거리와 일관된다는 점이었다.

한편, 긴장/이완 및 이중모음 오류의 패턴들은 비범주적 목록들로만 정의되었다. 이는 마찰음, 유음, 중모음 오류에서 비범주적 자질의 등장이 중도 유형에 국한되었던 것과 대조적이다. 결과적으로 패턴의 음성학적 속성을 분석하기 위한 단어 쌍 단위의 L1 자료가 부족했다. 따라서 대안으로 두 코드북의 소속 단어들을 따로 분석했다. 코드북 1과 2 단어들의 긴장성, 후설성, 전설성 값을 각각 계산한 후, 이를 종합하여 패턴의 연속체 내 위치를 추정하였다. 이러한 분해 분석의 방법으로 예측한 음성학적 위계 역시 패턴 벡터 사이 유클리디안 거리를 반영했다. 아울러 전후설성 결의도와 긴장성이라는 다른 기준으로 산출된 위계가 음성학적으로 일관된 관계를 보였다. 두 코드북에서 가장 일관된 전설 또는 후설성의 값을 갖는 패턴은 가장 긴장성의 값을 갖는 패턴과 동일 군집 소속이었으며, 반대로 가장 상반된 값을 갖는 패턴은 가장 이완성의 값을 갖는 패턴과 동일 군집 소속이었다. 이는 코드벡터가 다양한 음성학적 특징을 포괄적으로 이해하고 있음을 보여준다.

무엇보다 도출된 패턴들의 분포는 선행 연구의 발음 변이 현상 기술점과 맞닿아 있었다. 분석한 여섯 경로의 교체 예시들 전반적으로, 연속체 내 패턴의 분포는 학습자의 L1 음성 체계와 관련되었다. 다시 말해, 패턴의 분산도 및 목표음을 향한 응집력은 기존 L2 문헌에서 다른 모국어 전이 현상에 의한 조음적 어려움을 반영했다. 이는 오류 연속체

내적 비교와 연속체들 사이 외적 비교라는 두 가지 차원에서 조망해볼 수 있다.

오류 연속체 내적으로는, 학습자의 L1에 존재하는 가장 근접한 소리를 향해 패턴들이 편향되었다. Z/z to S/s에서 S/s를 향한 편향, L/l to R/r 과 R/r to L/l에서 L/l을 향한 편향, AH/ʌ to AA/ɑ에서 AH/ʌ를 향한 편향이 교체 내적 예시들이다. 나아가 연속체들 사이의 패턴 분산도를 비교해볼 때, 더 생소한 조음 특성을 가진 목표 음이 더 큰 분산을 유발했다. F/f to P/p가 DH/ð to D/d와 V/v to B/b 보다 분산이 작은 이유는 F/f가 DH/ð, V/v에 비해 L1 대응 음소와 가깝기 때문이었다. 마찬가지로 IY/i/ to IH/i/와 EH/ɛ/ to AE/æ/가 IH/i/ to IY/i/와 AE/æ/ to EH/ɛ/보다 분산이 작은 점은 IY/i/, EH/ɛ/가 IH/i/, AE/æ/에 비해 L1 대응 음소와 가깝기 때문이다. 또한 EH/ɛ/ to AE/æ/ 오류 패턴들의 특수성은 미국인 L1 화자와 한국인 L2 화자 사이 상반된 AE/æ/와 EH/ɛ/의 상대적 발음 분포 편차로 이해되기도 했다.

L1 표본이 부족해 단어별 분석을 진행했던 오류 예시들은, 특히 표준 발음과 학습자의 L2 발음이 다르다고 알려진 목표 음소들을 다루었다. 대표적으로 긴장-이완 교체와 전설 이중모음 축소 오류의 패턴들은 모순되는 코드 단어 조합들로만 구성되었었다. 이는 해당 오류들의 목표 음소인 영어 전설 모음을 한국인 학습자들이 특히 발음하기 어려워한다는 기존 연구 결과와 부합한다. 아울러 모음은 전반적으로 자음에 비해 발음 변이에 취약하므로, 자음보다는 모음 오류 패턴들에서 본 경향성이 관찰되었다고 해석된다.

결론적으로 본 연구의 발견은 다음의 세 단계로 요약될 수 있다. 1) 음성학적 속성으로 유추된 패턴 분포가 2) 패턴을 대표하는 코드벡터들 사이 유클리디안 거리와 일관된다는 점을 찾았으며, 그 일관성은 3) 기존 L2 문헌들의 발견과도 연관된 언어학적 유의성을 내포했다. 이는 코드벡터 간 수치적 계산을 통해 오류 패턴의 음소 이하 다양성을 정량화할 수 있음을 보여준다. 그리고 본 연구는 그 가능성을 탐색했다는 점에서 의의가 있다.

References

- Anand, N., Sirigiraju, M., & Yarra, C. (2023). Unsupervised speech intelligibility assessment with utterance level alignment distance between teacher and learner Wav2Vec-2.0 representations. Retrieved from <https://arxiv.org/abs/2306.08845>
- Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. ArXiv, abs/2006.11477.
- Bannò, S., Knill, K. M., Matassoni, M., Raina, V., & Gales, M. (2023, August). Assessment of L2 oral proficiency using self-supervised speech representation learning. *Proceedings of 9th Workshop on Speech and Language Technology in Education (SLaTE)* (pp. 126-130). Dublin, Ireland.

- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9), 341-345.
- Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., Li, J., ... Wei, F. (2022). WavLM: Large-scale self-supervised pre-training for full stack speech processing. *IEEE Journal of Selected Topics in Signal Processing*, 16(6), 1505-1518.
- Cho, M., & Jeong, S. (2013). Perception and production of English vowels by Korean learners: A case study. *Studies in Phonetics, Phonology and Morphology*, 19(1), 155-177.
- Choi, H., & Oh, M. (2021). Asymmetrical production of English diphthongs /eɪ/ and /oo/ by Korean learners of English. *Studies in Linguistics*, 58, 19-42.
- Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M. (2020). Unsupervised cross-lingual representation learning for speech recognition. Retrieved from <https://arxiv.org/abs/2006.13979>
- Facebook AI Research. (2019). *Fairseq: Facebook AI Research's sequence-to-sequence toolkit* [Computer software]. GitHub. <https://github.com/facebookresearch/fairseq>
- Facebook Research. (2024). *Faiss: A library for efficient similarity search and clustering of dense vectors* [Computer software]. Retrieved from <https://github.com/facebookresearch/faiss>
- Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., & Pallett, D. S. (1993). DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM. NIST speech disc 1-1.1. NASA STI/Recon technical report n, 93, 27403.
- Hong, H., Kim, S., & Chung, M. (2014). A corpus-based analysis of English segments produced by Korean learners. *Journal of Phonetics*, 46, 52-67.
- Hsu, W. N., Bolte, B., Tsai, Y. H. H., Lakhota, K., Salakhutdinov, R., & Mohamed, A. (2021). HuBERT: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 3451-3460.
- Kim, E., Jeon, J. J., Seo, H., & Kim, H. (2022). Automatic pronunciation assessment using self-supervised speech representation learning. Retrieved from <https://arxiv.org/abs/2204.03863>
- Kim, R. E., & Rhee, S. C. (2019). A study on English liquids in the rated L2 English speech corpus of Korean learners. *Korean Journal of English Language and Linguistics*, 19(1), 53-75.
- Kominek, J., & Black, A.W. (2004, June). The CMU Arctic speech databases. *Proceedings of 5th ISCA Speech Synthesis Workshop* (pp. 223-224). Pittsburgh, PA.
- Koo, H. (2012). A Study of production ability of English liquid /l/ and /r/ by Korean learners of English. *Korean Education Inquiry*, 30(2), 115-126.
- Koo, H. S., & Oh, Y.J. (2001). An analysis of English vowels of Korean learners of English and English native speakers. *Korean Education Inquiry*, 16, 1-12.
- Lee, Y., Shon, S., & Kim, T. (2018). Learning pronunciation from a foreign language in speech synthesis networks. Retrieved from <https://arxiv.org/abs/1811.09364>
- Li, X., Mao, S., Wu, X., Li, K., Liu, X., & Meng, H. (2018, September) Unsupervised discovery of non-native phonetic patterns in L2 English speech for mispronunciation detection and diagnosis. *Proceedings of Interspeech 2018* (pp. 2554-2558). Hyderabad, India.
- Li, X., Wu, X., Liu, X., & Meng, H. M. (2020). Deep segmental phonetic posterior-grams based discovery of non-categories in L2 English speech. Retrieved from <https://arxiv.org/abs/2002.00205>
- Liu, S., Mallol-Ragolta, A., Parada-Cabaleiro, E., Qian, K., Jing, X., Kathan, A., Hu, B., & Schuller, B. W. (2022). Audio self-supervised learning: A survey. *Patterns*, 3(12), 100616-100616.
- Mao, S., Li, X., Li, K., Wu, Z., Liu, X., & Meng, H. (2018, April). Unsupervised discovery of an extended phoneme set in L2 English speech for mispronunciation detection and diagnosis. *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6244-6248). Calgary, AB.
- Martin, K., Gauthier, J., Breiss, C., & Levy, R. (2023). Probing self-supervised speech models for phonetic and phonemic information: A case study in aspiration. Retrieved from <https://arxiv.org/abs/2306.06232>
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017, August). Montreal forced aligner: Trainable text-speech alignment using Kaldi. *Proceedings of the 18th Interspeech 2017* (pp. 498-502). Stockholm, Sweden.
- Oh, E. (2013). Learning dynamic transitions of the American English vowels [eɪ] and [oo]: A comparison between Korean-English bilingual speakers and Korean learners of English. *Linguistic Research*, 29(2), 323-343.
- Park, A. S., & Glass, J. R. (2007). Unsupervised pattern discovery in speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(1), 186-197.
- Park, H. (2001). An experimental study on the lengths of English diphthongs. *Speech Science*, 8(3), 7-14.
- Pasad, A., Chien, C. M., Settle, S., & Livescu, K. (2023). What do self-supervised speech models know about words? Retrieved from <https://arxiv.org/abs/2307.00162>
- Peng, L., Fu, K., Lin, B., Ke, D., & Zhang, J. (2021, August-September). A study on fine-tuning wav2vec2.0 model for the task of mispronunciation detection and diagnosis. *Proceedings of INTERSPEECH 2021* (pp. 4448-4452). Brno, Czechia.

- Peng, P., & Harwath, D. (2022). Self-supervised representation learning for speech using visual grounding and masked language modeling. Retrieved from <https://arxiv.org/abs/2202.03543>
- Schmidt, A. M. (1996). Cross language identification of consonants. Part 1. Korean perception of English. *The Journal of the Acoustical Society of America*, 99(5), 3201-3211.
- Shen, G., Alishahi, A., Bisazza, A., & Chrupala, G. (2023). Wave to syntax: Probing spoken language models for syntax. Retrieved from <https://arxiv.org/abs/2305.18957>
- Smith, B., & Swan, M. (2001). *Learner English: A teacher's guide to interference and other problems* (2nd ed.). Cambridge: Cambridge University Press.
- Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, 33(3), 263-290.
- Wang, Y. B., & Lee, L. S. (2013, May). Toward unsupervised discovery of pronunciation error patterns using universal phoneme posteriorgram for computer-assisted language learning. *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8232-8236). Vancouver, BC.
- Wang, Y. B., & Lee, L. S. (2015). Supervised detection and unsupervised discovery of pronunciation error patterns for computer-assisted language learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(3), 564-579.
- Wu, M., Li, K., Leung, W. K., & Meng, H. (2021, August-September). Transformer based end-to-end mispronunciation detection and diagnosis. *Proceedings of INTERSPEECH 2021* (pp. 3954-3958). Brno, Czechia.
- Xu, X., Kang, Y., Cao, S., Lin, B., & Ma, L. (2021, August-September). Explore wav2vec 2.0 for mispronunciation detection. *Proceedings of INTERSPEECH 2021* (pp. 4428-4432). Brno, Czechia.
- Yang, B. (2010). College students' production and perception of English vowels. *English Language Teaching*, 22(4), 165-184.
- Yang, B. (2013). A comparative study of relative distances among English front vowels produced by Korean and American speakers. *Phonetics and Speech Sciences*, 5(4), 99-107.
- Zhao, G., Sosaat, S., Silpachai, A., Lucic, I., ChukharevHudilainen, E., Levis, J., & Gutierrez-Osuna, R. (2018, September). L2-arctic: A nonnative english speech corpus. *Proceedings of IINTER SPEECH 2018* (pp. 2783-2787). Hyderabad, India.

• **홍은수 (Eunsoo Hong)**

서울대학교 언어학과 석사과정

서울시 관악구 관악로1

Tel: 02-880-6164

Email: stxlla13@snu.ac.kr

관심분야: 음성인식, 음성언어처리, 언어교육

• **김선희 (Sunhee Kim)**

서울대학교 불어교육과 교수

서울시 관악구 관악로1

Tel: 02-880-7693

Email: sunhkim@snu.ac.kr

관심분야: 프랑스어 음성학, 음성언어처리

• **정민화 (Minhwa Chug)** 교신저자

서울대학교 언어학과 교수

서울시 관악구 관악로1

Tel: 02-880-9195

Email: mchung@snu.ac.kr

관심분야: 음성인식, 음성언어처리, 컴퓨터 기반 언어교육

Wav2Vec2.0 코드벡터를 활용한 음소 범주 사이 L2 오류 패턴 탐구*

홍 은 수¹ · 김 선 희² · 정 민 화¹

¹서울대학교 언어학과, ²서울대학교 불어교육과

국문초록

L2 발음 오류는 모국어 전이 현상에 의해 둘 이상의 표준 음소 범주를 아우르는 특성을 띤다. 이 같은 비범주성을 고려한 평가를 위해서는 음소보다 세분화된 발음 분석 단위가 필요하다. 선행 연구들에서는 그 예시로 음소 사후 확률 (phonemic posterior-gram, PPG)을 제안했지만, 사전 정의된 음소 집합과의 연계성으로 소리를 표현하므로 범주적 개입을 온전히 벗어나지 못한다. 이에 본 연구는 자기지도 학습(self-supervised learning, SSL) 모델의 표현 학습이 외부 음소적 규제로부터 자유롭다는 점에 착안해, Wav2Vec2.0 코드벡터를 비범주적 양상 분석의 새로운 단위로 활용한다. 먼저 자질이 L1과 L2 음성을 다르게 인지하는지 확인하기 위해, 내용적으로 통제된 L1 (CMU ARCTIC)과 L2 (L2 ARCTIC) 화자별 데이터에서 사용되는 코드벡터 목록들을 비교했다. 그 후 범주적 정의 내 세부 변이양상들을 조사하고자, 선차적으로 음소 단위 오류탐지를 실시해 분석할 교체오류들을 선별했다. 이때 모델 파인튜닝에 사용될 데이터는 L1 TIMIT이며, 분석 대상은 NIA 037 교육용 한국인의 영어 음성 데이터다. 이후 교체 유형별로 오류 표본들의 음성 프레임에 대응되는 코드벡터열을 추출한 뒤, 우세한 패턴들을 도출했다. 도출된 패턴들은 L1 TIMIT의 코드벡터별 음소 분포 조건부 확률과 현존하는 코드벡터 집합의 군집화 결과를 참조해 최종 해석된다. 분석 결과, 코드벡터는 L2 변별력과 더불어 발음 연속체 속 오류 패턴의 범주 간 위치를 수치화할 능력을 보유하고 있었다. 이른바, 각 오류 패턴을 대표하는 자질들 사이 유클리디안 거리와 패턴 분포는 현존하는 L2 발음 연구에서 기술된 현상들을 반영했다.

핵심어: Wav2Vec2.0 코드벡터, 표현 학습, 비범주성, 오류 패턴

참고문헌

- 구희산. (2012). 영어학습자의 유음 /l/과 /r/의 발성 능력 연구. *한국교육문제연구*, 30(2), 155-126.
- 구희산, 오연진 (2001). 한국인 영어학습자와 영어원어민 영어모음 발음 분석. *한국교육문제연구*, 16, 1-12.
- 김려은, 이석재. (2019). 한국인 영어학습자의 영어 발음 평가 등급에 따른 유음 발화 양상 연구. *영어학*, 19(1), 53-75.
- 양병곤. (2010). 대학생들의 영어모음 발음과 지각. *영어교육연구*, 22(4), 165-184.
- 양병곤. (2013). 한국인과 미국인이 발화한 영어전설모음의 상대적 거리 비교. *말소리와 음성과학*, 5(4), 99-107.
- 오은진 (2013). 미국 영어 모음 [eɪ]와 [oo]의 동적 전이 학습: 한국어-영어 이중언어 구사자와 한국인 영어 학습자 간 비교. *언어연구*, 29(2), 323-343.
- 최혜란, 오미라. (2021). 한국인 영어 학습자의 영어 이중모음 /eɪ/와 /oo/에 대한 비대칭적 산출연구. *언어학 연구*, (58), 19-42.

* 본 연구는 서울대학교 언어학과 SNU 10-10 프로젝트의 지원을 받아 수행되었으며, 제 1저자의 2024년 석사논문의 일부를 바탕으로 작성되었음.