

ECHO!

: English Correction & Hearing Optimization

종합설계 프로젝트 제안서

2026. 03. 17.

A6 O(1)팀

김형준, 김지용, 백지영, 이찬영, 신영환, 최현지

Contents

> ☰ 1. 서론

> ☰ 2. 본론

> ☰ 3. 결론

> ☰ 4. 부록

1.1 제안 배경

1.1.1 현황 분석

1.1.2 타사 서비스 분석

1.1.3 음성 인식 정확도 측면

1.1.4 한국어 L1 화자 교정 피드백 제공 한계

1.2 프로젝트 목표

1.3 서비스 개요

왜 발음 피드백 서비스가 필요한가?

1.1.1 제안 배경 : 현황 분석

2025년 상반기 주요 기업 722건 채용 공고 중 535건, 토익스피킹 성적 반영 국내 대표 교육 및 평가 전문 그룹 YBM(회장 민선식) 산하 한국TOEIC위원회가 조사 결과 2025년 상반기(1~7월) **국내 주요 기업의 채용 공고** 722건 중 535건이 전형 과정에서 **토익스피킹(TOEIC Speaking) 성적을 반영**했다고 14일 밝혔다.

한국TOEIC위원회, 스피킹 역량 위한 '토익스피킹' 활용 지속 확대 (세계일보, 2025-11-14)

TOEIC® Speaking Test 시험구성

(11문항 / 약 20분)

구분	문제 유형	문항수	답변 준비 시간	답변시간	평가기준
Questions 1-2	Read a text aloud - 문장 읽기	2	각 45초	각 45초	· 발음 · 억양과 강세



Accuracy

응시자의 답변을 이해하는 데 영향을 미치는 어휘, 문법, 발음, 유창성 등 평가

1.1.1 제안 배경 : 현황 분석

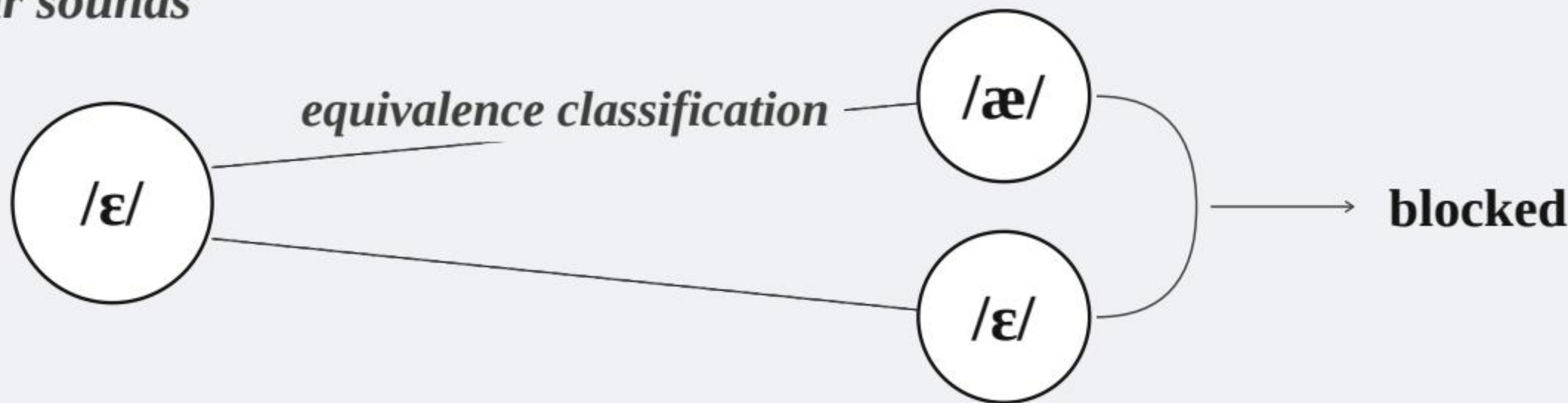
Given the methodological issues in the studies reporting positive results, and the mixed results in the remaining studies, **the overall effect of shadowing on segmental control can be considered inconclusive.**

Whitworth, B., & Rose, H. (2025). A systematic review of research on the use of shadowing for second language pronunciation teaching. Research Synthesis in Applied Linguistics, 1(2), 239–269. <https://doi.org/10.1080/29984475.2025.2546827>

L1 (Korean)

L2 (English)

(a) Similar sounds



1.1.1 제안 배경 : 현황 분석

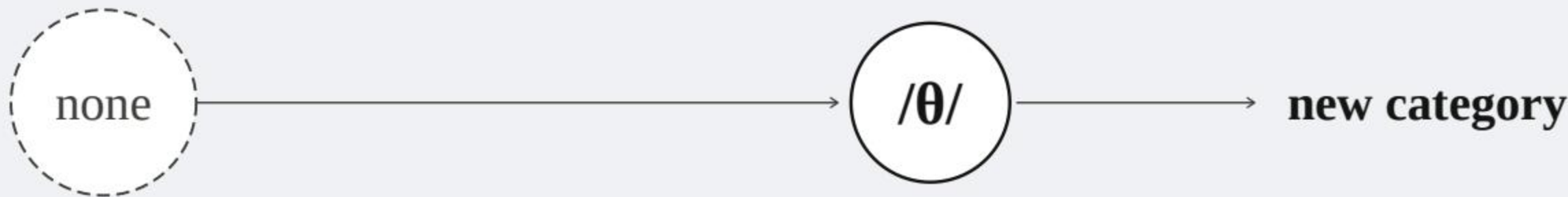
Given the methodological issues in the studies reporting positive results, and the mixed results in the remaining studies, **the overall effect of shadowing on segmental control can be considered inconclusive.**

Whitworth, B., & Rose, H. (2025). A systematic review of research on the use of shadowing for second language pronunciation teaching. *Research Synthesis in Applied Linguistics*, 1(2), 239–269. <https://doi.org/10.1080/29984475.2025.2546827>

L1 (Korean)

L2 (English)

(b) New sound

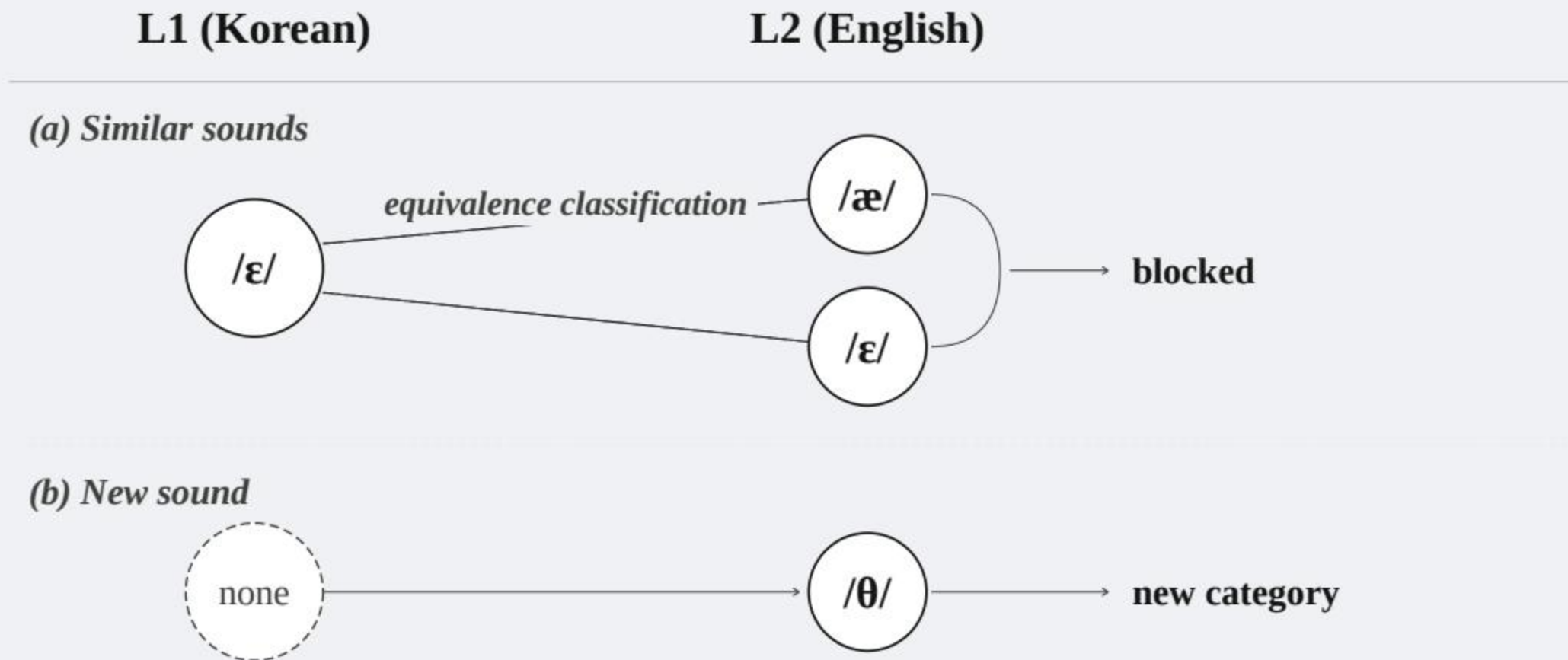


Flege (1995), Speech Learning Model — Hypothesis 5

1.1.1 제안 배경 : 현황 분석

“쉐도잉이 발음 개선에 효과가 있다고 딱 잘라 말할 수 없다”

Whitworth, B., & Rose, H. (2025). A systematic review of research on the use of shadowing for second language pronunciation teaching. *Research Synthesis in Applied Linguistics*, 1(2), 239–269. <https://doi.org/10.1080/29984475.2025.2546827>



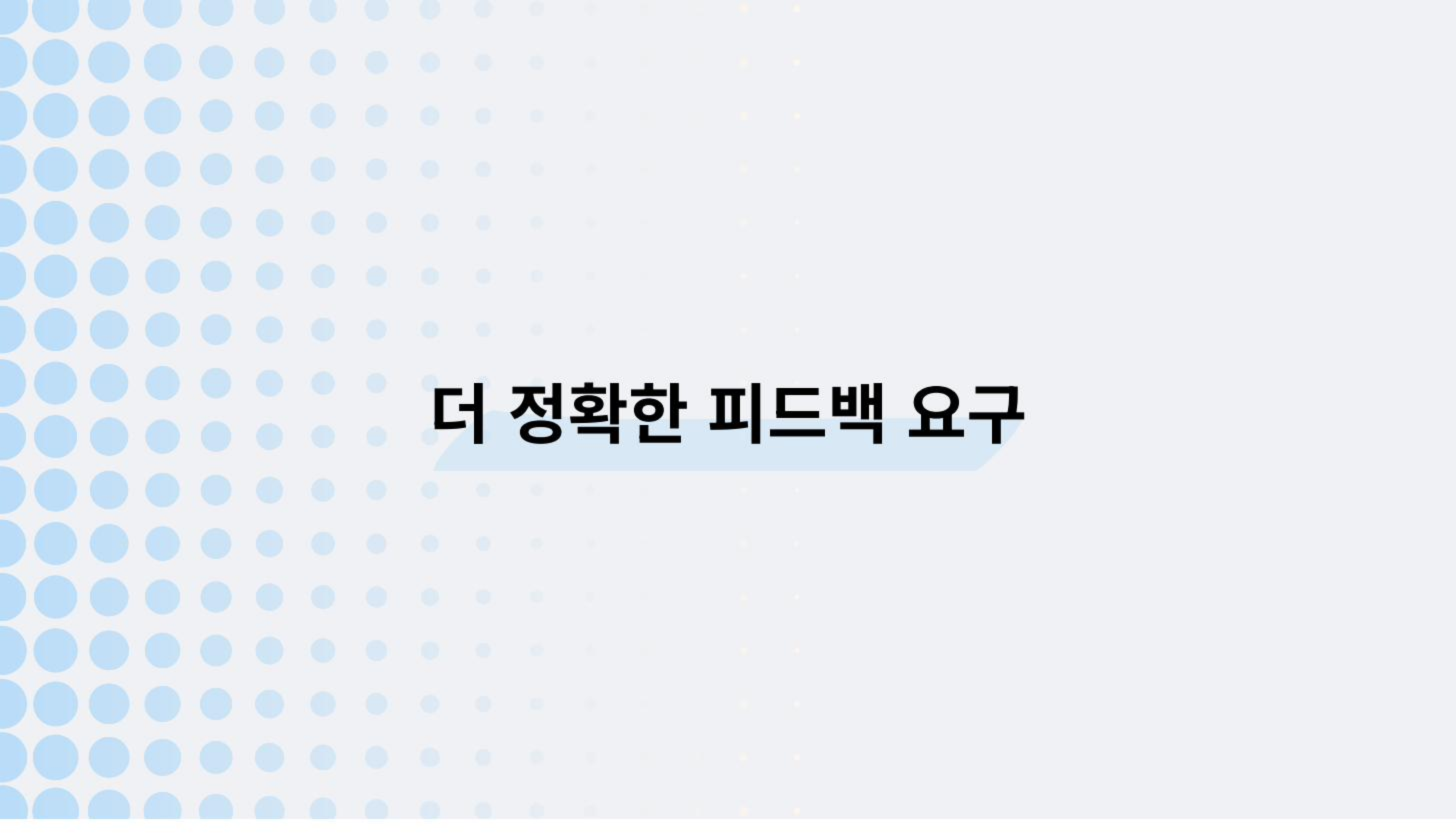
Flege (1995), Speech Learning Model — Hypothesis 5

1.1.1 제안 배경 : 현황 분석



청년층을 중심으로
어학 자격증의 필요성이 대두되지만,
정확한 **피드백 없는 학습**만으로는
발음 개선에 한계가 있다.

타사 서비스 분석



더 정확한 피드백 요구

1.1.2 제안 배경 : 타사 서비스 분석



듀*링고 App store 리뷰

< 불친절한 학습어플
왜 그런지에 대한 설명이 너무 없고
그냥 계속 틀리다 보면
패턴이 생기는 걸 보고
눈치껏 정답 맞추게 됨
틀린 것에 대한 해석이나 설명이 필요함
기본적인 설명이 너무 없음 >



스* App store 리뷰

< 음성인식 최악...
발음이 이상하면 그것에 대해
피드백이 있어야 하는데
그것도 아니고... >

더 자세한 피드백 서비스에 대한 사용자 요구

1.1.2 제안 배경 : 타사 서비스 분석

발음 피드백 항목	말해보카	듀오링고
정확도 점수 제공	O - 레벨 측정	X- 맞음/틀림만
오류 위치 표시	O - 잘못된 부분 알림	△ - 단어 단위 pass/fail만
모범 발음 대조	X	X
교정 가이드	X	X
강세/억양 피드백	X	X

1.1.2 제안 배경 : 타사 서비스 분석

틀린 발음 교정하기 AI 발음 코치



스* 공식 사이트

스픽은 학습자의 발음을 음소 단위로 분석합니다.
학습자의 발음과 모범 발음을 실시간으로 대조 분석해
학습자가 어떻게 발음했는지, 어떻게 발음하면 더 자
연스러울지 상세하게 안내하고 교정합니다.

이렇게 발음해요

laa · tay

이렇게 말하신 것 같아요

laa · dai

dai 대신 **tay**(으)로 발음해 보세요.

설명 더보기 ▾

1.1.2 제안 배경 : 타사 서비스 분석

Speak식 피드백 예시: "pa" 대신 "fa"로 발음해 보세요

학습자: "copy" /kapi/ 목표: "coffee" /kafi/

Korean

ㅂ

bilabial

ㅃ

bilabial asp

/f/ 없음 — ㅍ으로 대체

English

/p/

bilabial stop

/f/

labiodental fric

조음 위치 + 방법이 다름

한국어에 /f/가 없어 ㅍ(/p/)로 대체 → "coffee"가 "copy"로

Flege (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), Speech perception and linguistic experience

더 정확한 인식 성능 요구

1.1.2 제안 배경 : 타사 서비스 분석



스* App store 리뷰

Amazon을 Alphaz로 인식하는
< *** 암만 제대로 말해도 >
인식이 이상하게 되서는
내가 틀렸다고 함 ㅋㅋ
*짜증난다 진짜



말*보카 App store 리뷰

serving the problem
< 이라는 말이 어땠다고 >
자꾸 solving을
serving이라고 인지하고
진짜 이거 말고도
미쳐버릴 것 같은
오류 너무 많음

더 좋은 음성인식 성능에 대한 사용자 요구

1.1.3 제안 배경 : 음성 인식 정확도 측면

낮은 음성 인식 성능



L2 Arctic 기반 실험

낮은 음소인식 성능은 일반적으로
MDD 성능에 악영향을 줄 수 있음

MDD = Mispronunciation
Detection & Diagnosis

가장 많이 쓰는 모델들의
음소인식 성능과 MDD 성능 체크
Wav2Vec2 (base)
Wav2Vec2 (Large)
HuBERT (Large)

1.1.3 제안 배경 : 음성 인식 정확도 측면

L2-ARCTIC (Zhao et al., 2018)

24 non-native speakers

6 L1s: Hindi, Korean, Mandarin,
Spanish, Arabic, Vietnamese
~1,100 utts / speaker (read speech)

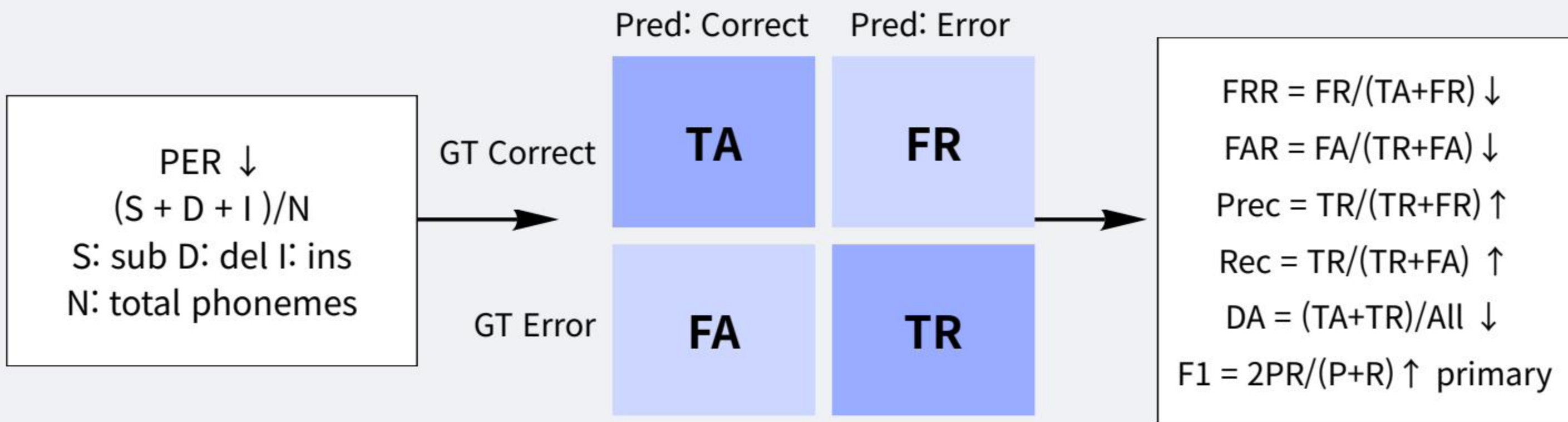
Used subset

6 speakers $\times$ 150 utts
29,087 phonemes
expert phonetic annotation

Per-phoneme annotation

canonical (정답 음소) + perceived (실제 발화 음소)
e.g. "think": canonical /θ/ → perceived /s/ (substitution)

1.1.3 제안 배경 : 음성 인식 정확도 측면



TA 정상 발음 → 정상으로 판정

FR 정상 발음 → 오류로 오판

FA 오류 발음 → 정상으로 놓침

TR 오류 발음 → 오류로 탐지

can: canonical phoneme (정답) / per: perceived phoneme (실제 발화) / pred: model prediction

1.1.3 제안 배경 : 음성 인식 정확도 측면

Position-conditional analysis

PER head output per position

GT correct

GT error

PER_C ↓

정발음 위치 음소 인식 오류율

PER_E ↑

오발음 위치 음소 인식 오류율

$ECR = PER_E / PER_C$ ↑

높을수록 MDD 분별력 ↑

1.1.3 제안 배경 : 음성 인식 정확도 측면

L2-ARCTIC 데이터셋에 대한 세 가지 사전학습 모델로 베이스라인 실험을 수행
(6명 화자, 화자당 150 발화, 총 29,087 음소)

Model	PER	MDD F1	Precision	Recall	FRR	FAR	DA
W2V2-B	14.99%	0.6018	0.5738	0.6329	7.86%	36.71%	88.01%
W2V2-L	14.58%	0.5974	0.5873	0.6082	7.14%	39.18%	88.27%
HuBERT-L	13.66%	0.6079	0.6241	0.5929	5.98%	40.71%	89.04%

1.1.3 제안 배경 : 음성 인식 정확도 측면

PER과 MDD F1 사이의 괴리

position conditional 지표 분석

Model	PER_C	PER_E	ECR	CA_correct	CA_error
W2V2-B	7.34%	57.25%	7.83	92.66%	60.57%
W2V2-L	6.62%	56.27%	8.53	93.38%	59.21%
HuBERT-L	5.60%	58.33%	10.48	94.40%	57.08%

1.1.4 한국어 L1 화자 교정 피드백 제공 한계

한국어 화자(YKWK)를 포함한 Wav2Vec2-Base 모델의 화자별 성능

Speaker (L1)	PER	MDD F1	FRR	FAR	DA
TLV (Vietnamese)	17.19%	0.8003	9.11%	14.69%	89.51%
NJS (Spanish)	13.81%	0.5677	6.96%	42.28%	88.36%
ZHAA (Arabic)	12.80%	0.5258	6.66%	44.22%	89.37%
YKWK (Korean)	14.47%	0.5205	7.22%	46.93%	87.83%
TXHC (Mandarin)	16.06%	0.5185	7.98%	47.10%	86.73%
TNI (Hindi)	15.76%	0.4575	9.37%	48.40%	86.24%

1.1.4 한국어 L1 화자 교정 피드백 제공 한계

한국어 화자(YKWK)에 대한 세 가지 사전학습 모델 비교

Model	YKWK MDD F1	YKWK FAR	YKWK FRR
W2V2-B	0.5205	46.93%	7.22%
W2V2-L	0.5021	50.56%	6.76%
HuBERT-L	0.4998	52.97%	5.86%

그래서 $O(1)$ 팀은 제안합니다...

“

구체적인 조음 위치 정보를 한국어로 제공하는 서비스,



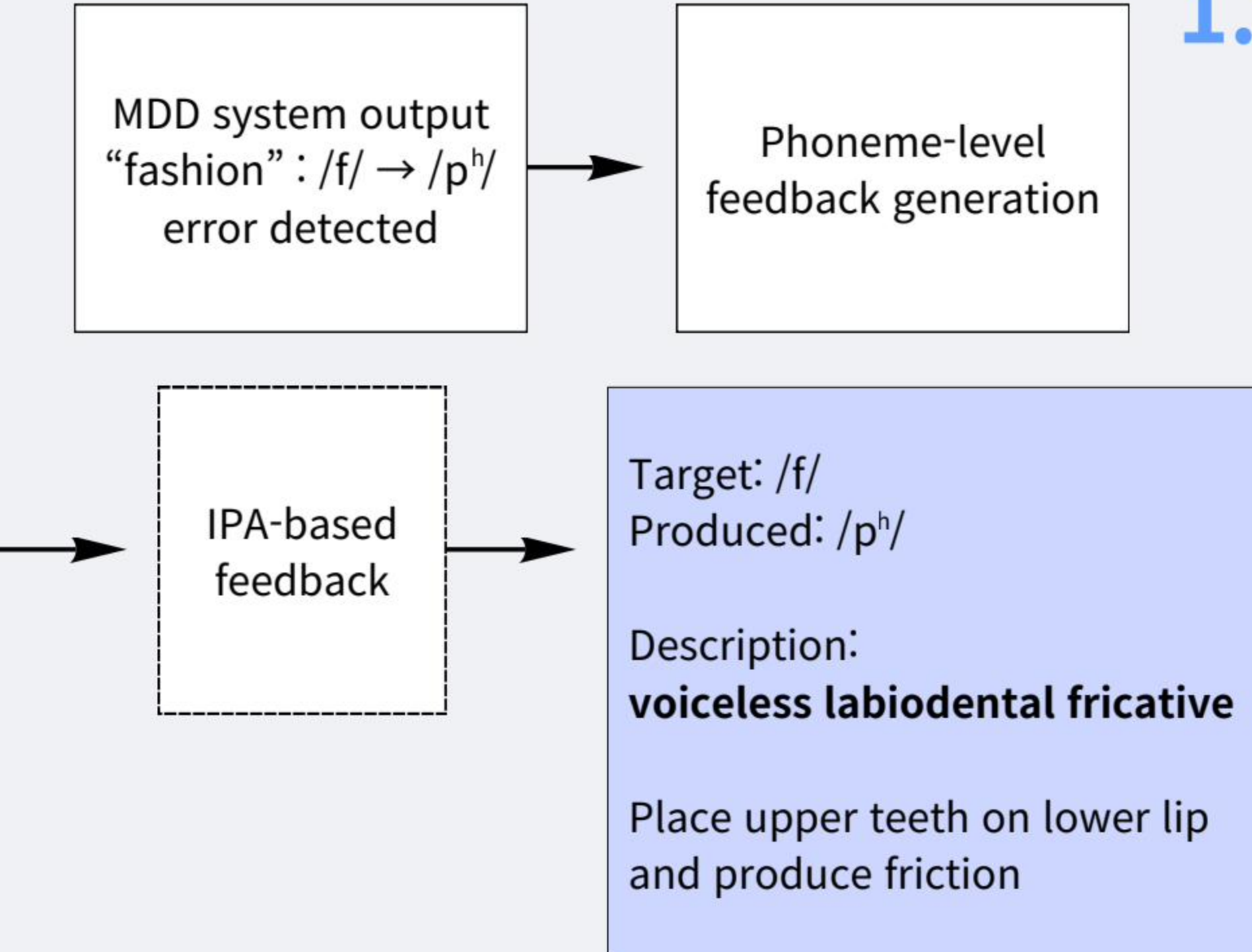
*가상 로고 시안

”

우리가 해결할 문제

**한국인에게 특화된,
더 자세한,
발음 피드백**

1.2 ECHO의 목표



Vegetable /'vɛdʒ.təbl/  Say: "VEJ-tuh-bl".	Rice /'raɪs/  Distinguish from /l/ by pulling the tongue tip back, far from your teeth.
Learn /'lɜ:n/  Pronounced like 'lurn', though spelled with 'ea'.	Book /'bʊk/  The vowel sound is /ʊ/, like in 'look'.

→ **Requires phonetics** knowledge to interpret

1.2 ECHO의 목표

자료 출처 KBS 천상의 컬렉션

숙리나 쇼	ツクリナ ショ	蔬	수	나물 소	이네과 소	イネク ソ	禾	허	벼 화
	魚						歌		
Vegetable					Rice				
배지타불					으라이쓰				
마나부 각구	マナブ カク	學	샷	배울 학	아무 헨	カム ベン	編	썬	책 편
	覺						先		
Learn					Book				
을러언					썬크				

MDD system output
“fashion” : /f/ → /p^h/
error detected

Phoneme-level
feedback generation

L1-grounded
feedback
(ours)

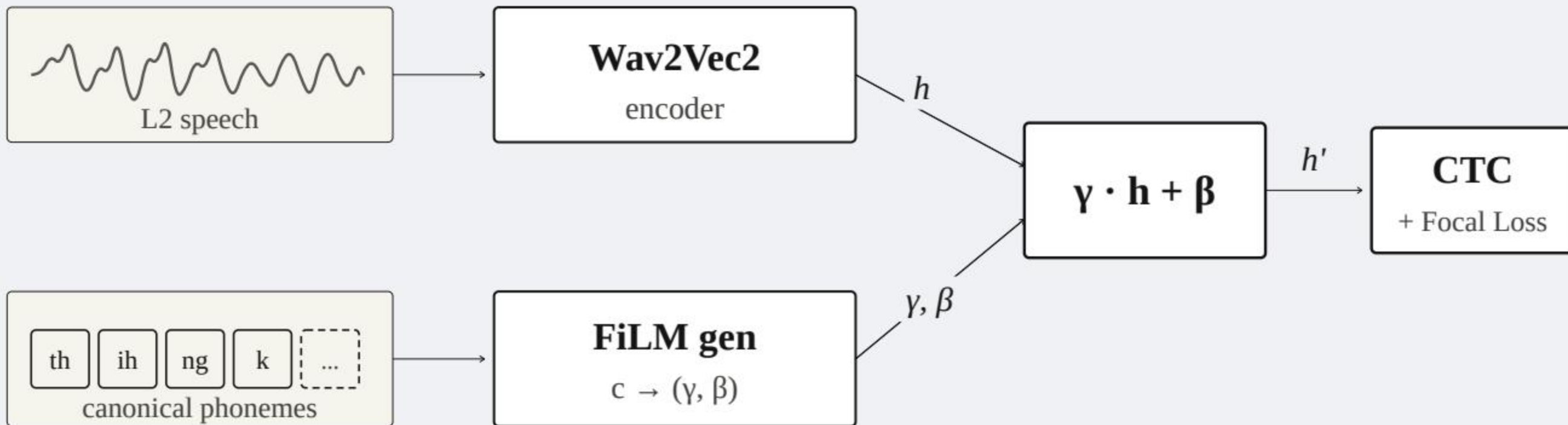
Target: /f/
Produced: /p^h/ (π)

Instruction:
π처럼 입술을 다물지 말고
윗니를 아랫입술에 대고
바람을 내보내세요.

→ Interpretable without
specialized knowledge

**더 정확한,
올바른 발음을 더 잘 인식하는,
더 좋은 인식 정확도**

1.2 ECHO의 목표



$$h' = \gamma \cdot h + \beta \rightarrow \text{PER_C} \downarrow \rightarrow \text{FRR} \downarrow$$

canonical 시퀀스로 γ, β 생성 $\rightarrow$ encoder feature를 affine 변환

$\rightarrow$ 정발음 위치에서 canonical 인식 확률 향상

1.2 ECHO의 목표

AI-Hub
한국인 영어 발화
데이터셋을 추가 학습
데이터 활용

CPU 환경에서도
수용 가능한 응답 속도를
확보하기 위한
모델 경량화

Contents

> ☰ 1. 서론

> ☰ 2. 본론

> ☰ 3. 결론

> ☰ 4. 부록

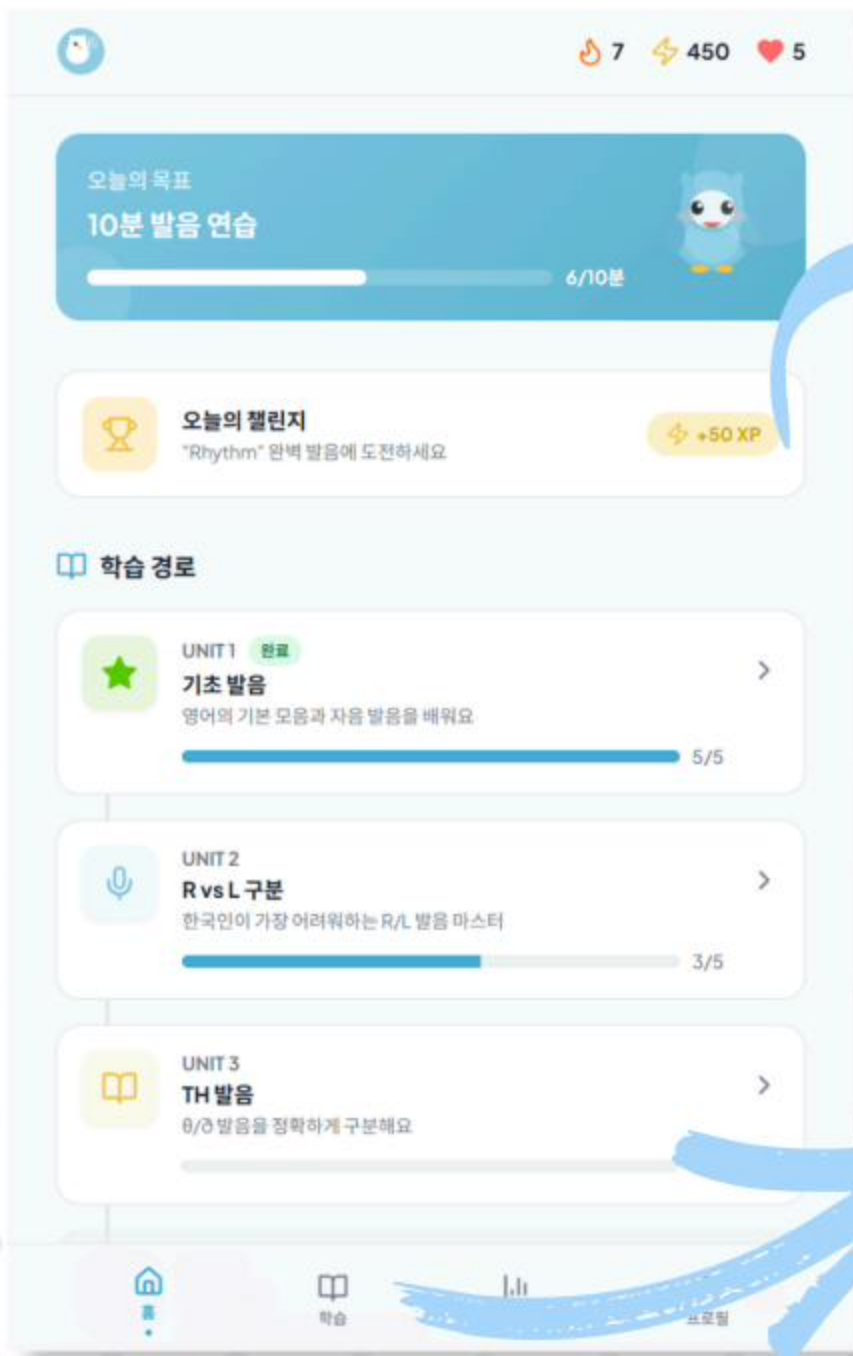
2.1 결과물 명세

2.2 기타 구현

2.3 사용 사례

예상 UI와 관련 기능은...

2.1 결과물 명세



추천 학습 기능

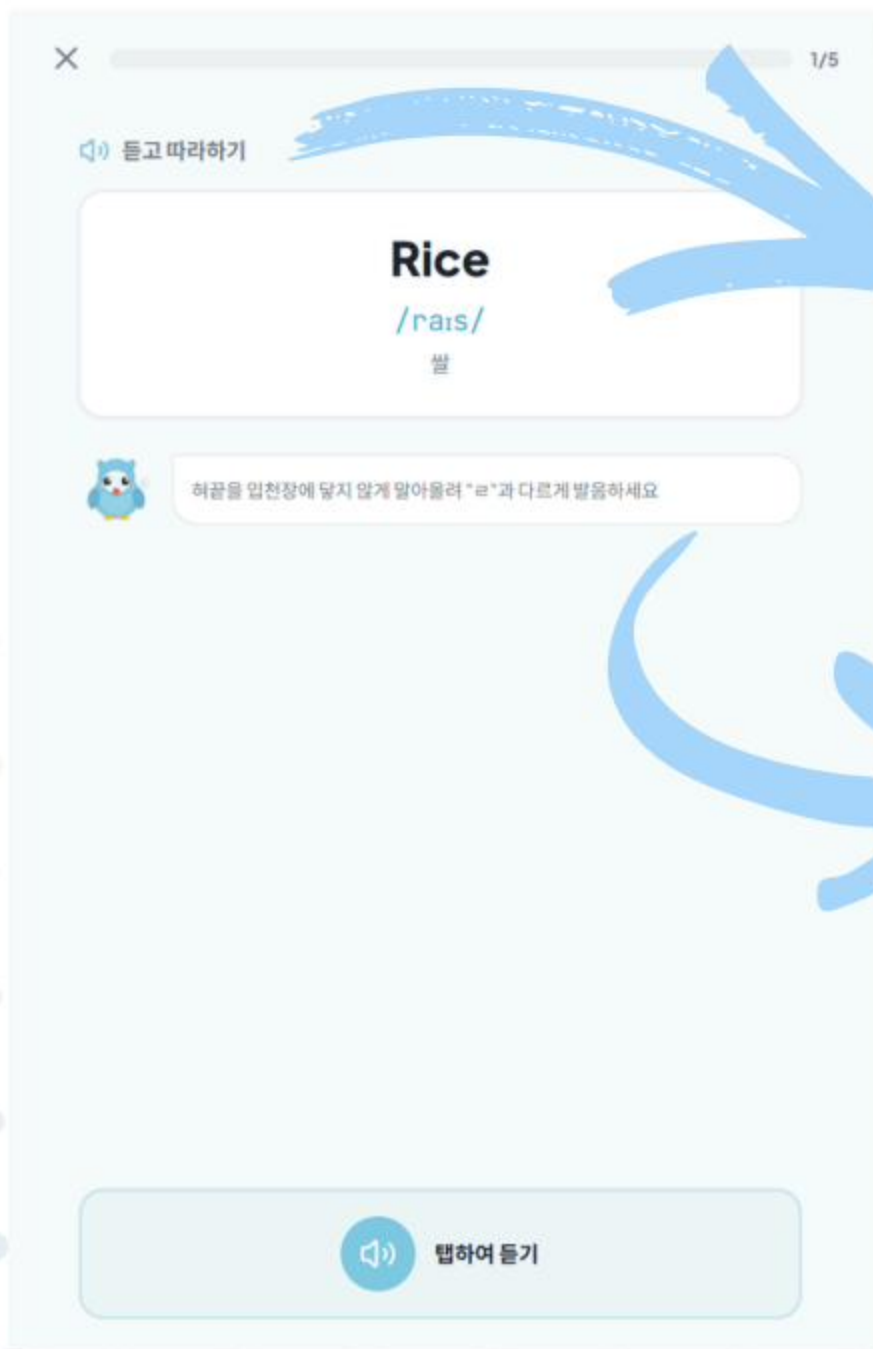


사용자 지정 연습 스크립트

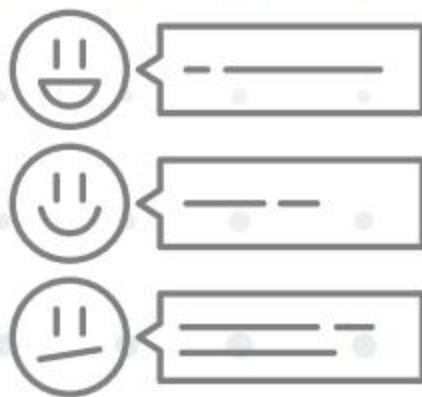


음성 녹음 및 전송

2.1 결과물 명세



표준 발음 제공

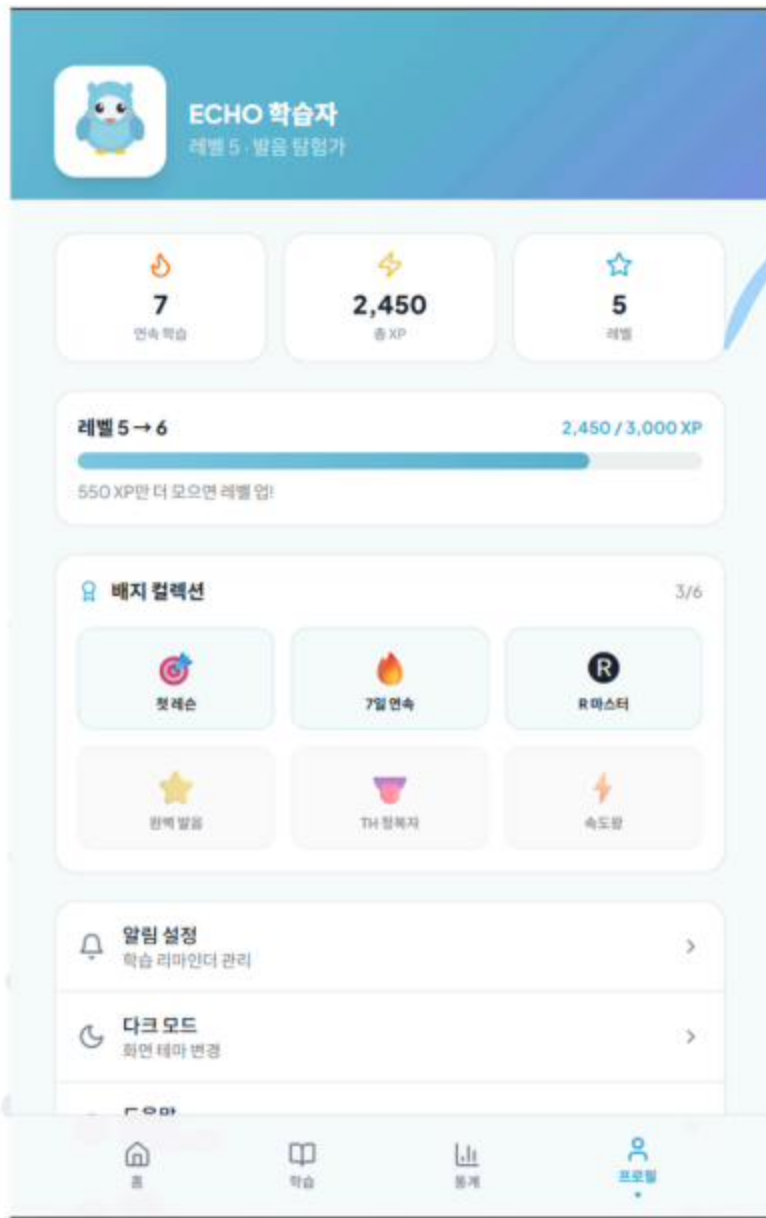


LLM 기반 피드백 시스템



음성 분석 및 오발음 검출

2.1 결과물 명세



실사용자 확보와 유지를 위한
게이미피케이션 및
환급 이벤트 활용 마케팅 전략

2.2 기타 구현

배지 컬렉션

3/6



첫 레슨



7일 연속



R 마스터



완벽 발음



TH 정복자



속도왕

‘속도’와 ‘보상’에 집중하여 학습자 흥미 유지 강조

- 사용자 리텐션을 위한 속도 챌린지
- 손실 회피 심리를 이용한 출석 환급 이벤트
- 환급을 통해 비용부담 해소



학교 마스코트 부의 컬러와 캐릭터성을 활용한 UI 디자인

- 하늘색 기반 컬러 팔레트
- 한국외대 공식 캐릭터 BOO를 활용한 캐릭터 제작

USE CASE는...

2.3 사용 사례 1: 기본 발음 교정 세션

OPIc IM1에서 막힌 막학기 취준생 A의 사용 시나리오



현황

- 토익 850점, 문법 독해 탄탄
- OPIc을 세 번 응시했지만 IM1에서 IM2로 오르지 않는다.
→ 스크립트도 외웠고 문법 오류도 거의 없는데
왜 점수가 안 오르는지 이유를 모른다.

2.3 사용 사례 1: 기본 발음 교정 세션

OPIc IM1에서 막힌 막학기 취준생 A의 사용 시나리오



현황

- 토익 850점, 문법 독해 탄탄
- OPIc을 세 번 응시했지만 IM1에서 IM2로 오르지 않는다.
→ 스크립트도 외웠고 문법 오류도 거의 없는데
왜 점수가 안 오르는지 이유를 모른다.



음소, 구강구조 단위별 피드백
(한국어 화자 특유의) 오류 교정 방법 제시

2.3 사용 사례 2: 맞춤형 스크립트 학습

해외 파견 준비 중인 6년차 직장인 B의 사용 시나리오



현황

- 6개월 후 미국 법인으로 파견
- 독해/작문 문제 X
- 실제 발표에서 문제 발생

2.3 사용 사례 2: 맞춤형 스크립트 학습

해외 파견 준비 중인 6년차 직장인 B의 사용 시나리오



현황

- 6개월 후 미국 법인으로 파견
- 독해/작문 문제 X
- 실제 발표에서 문제 발생



원하는 스크립트 전용 발음 교정
맞다/틀리다 수준 이상의 피드백

Contents

> ☰ 1. 서론

> ☰ 2. 본론

> ☰ 3. 결론

> ☰ 4. 부록

3.1 달성 목표

3.2 프로젝트 의의

발음 피드백 모듈 평가

3.1 달성 목표

→ L2-ARCTIC 데이터셋 기준의 성능 평가와 소프트웨어 파이프라인의 정상 동작 여부를 기준으로 성공 판단

 **1 FRR 4%대로 감소,
MDD F1하락 없이 유지**

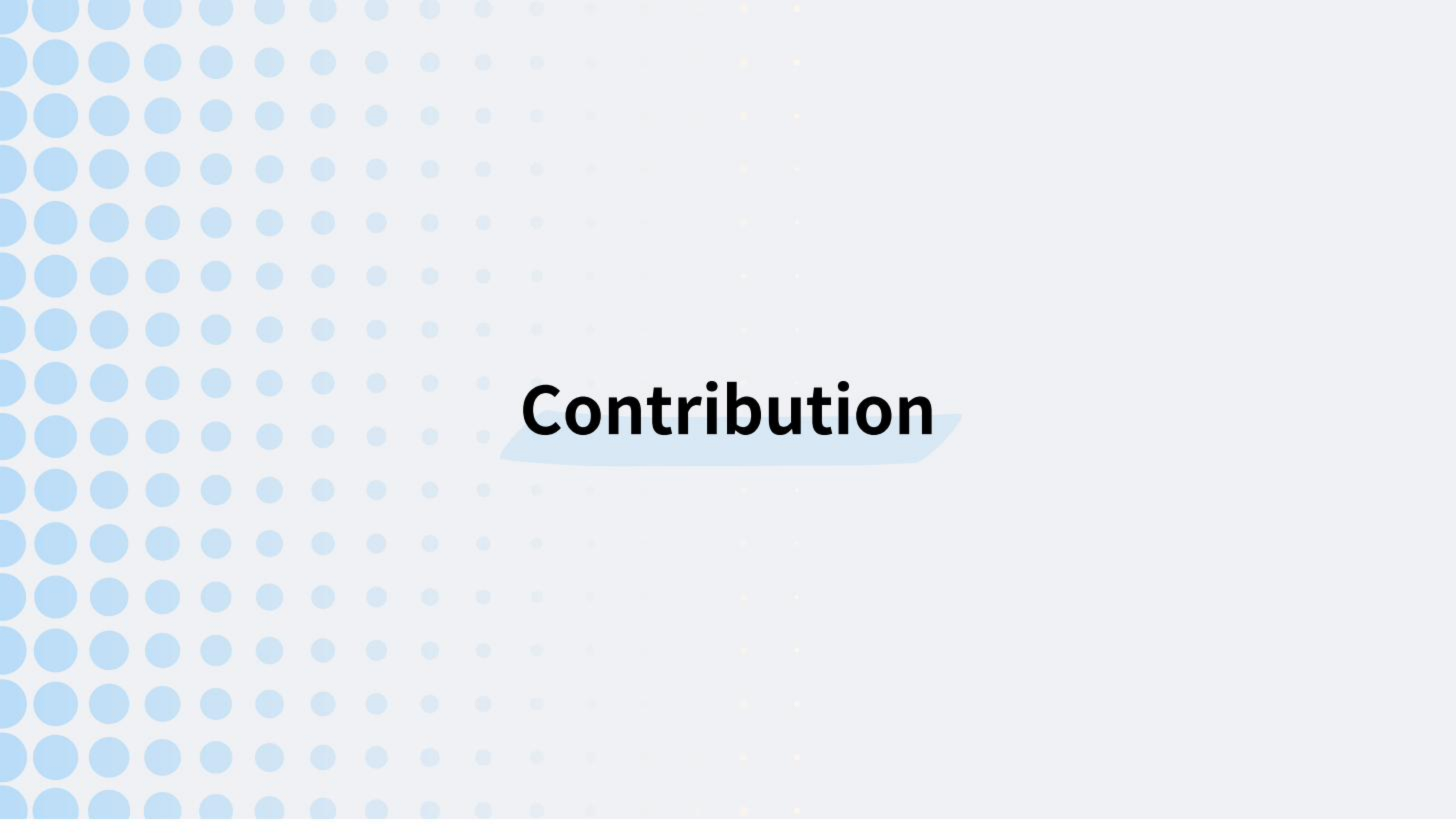
FRR은 학습자의 의욕을 가장
빠르게 꺾는 오류

Canonical Injection이
정발음 위치 인식 오류율을
낮출수 있는가?

 **2 자체 제작 벤치마크로
LLM피드백의 타당성 검증**

벤치마크를 자체 구축하여
피드백 타당성 평가

올바른 피드백을
고를 줄 아는 능력이
LLM에게 있는가?



Contribution

3.2 프로젝트 의의



개선된 발음 피드백

더 정확한 음소 단위 인식기를
활용하여 더 정확한
오류 유형 탐지 및 피드백 제공

이해하기 쉬운 피드백

한국어 L1 간접 패턴 고려하여
구체적인 교정 방법을 LLM으로 생성

사용자 맞춤 스크립트 피드백

학습 목적에 따라 연습 스크립트
선택 또는 직접 입력 가능

Contents

> ☰ 1. 서론

> ☰ 2. 본론

> ☰ 3. 결론

> ☰ 4. 부록

4.1 한국어와 영어 음운 체계의 근본적인 차이

4.2 역할 분담

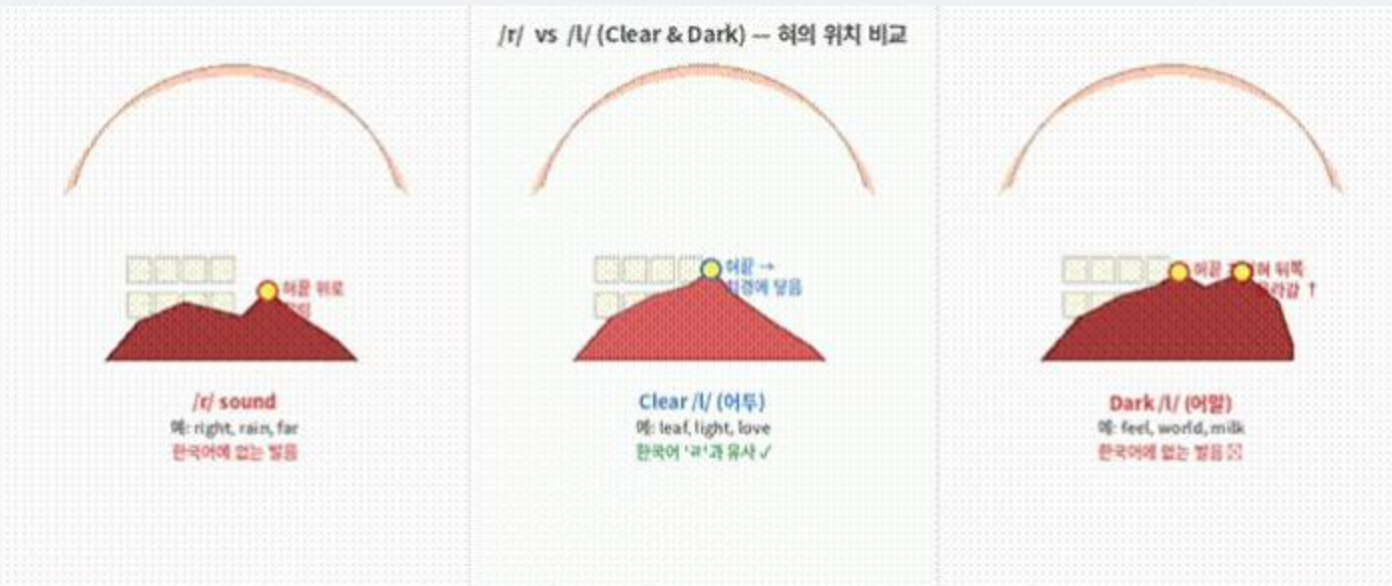
4. 부록

1) 한국어와 영어 음운 체계의 근본적인 차이

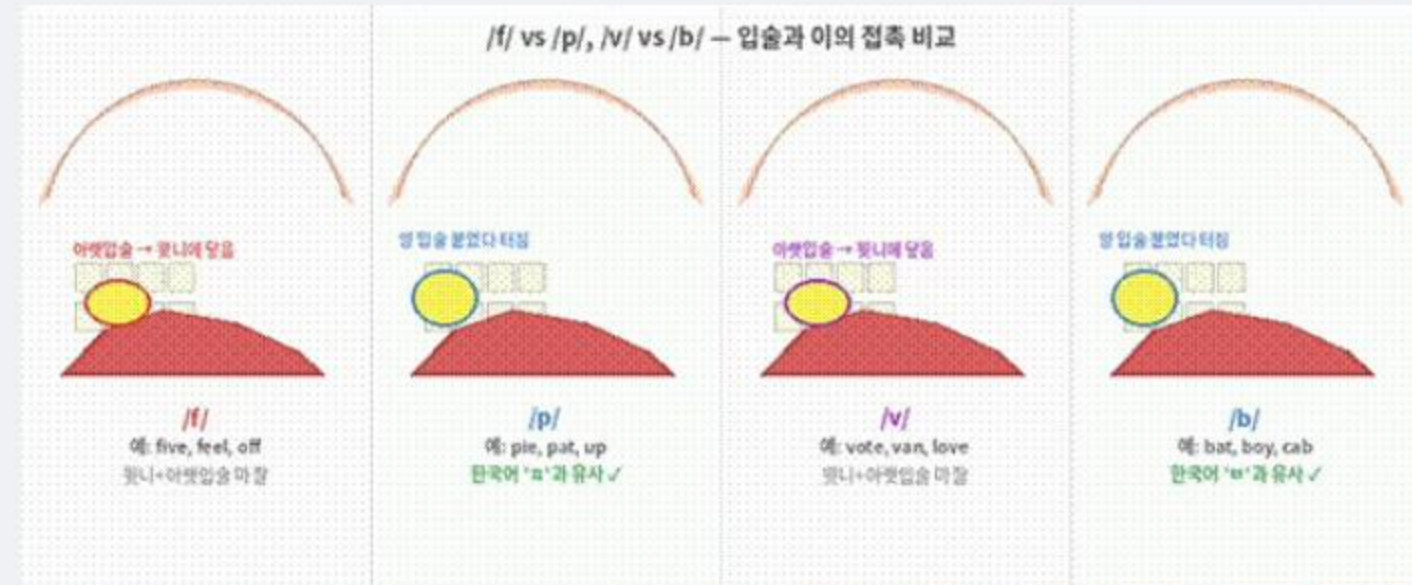
<한국인의 영어 발음 오류 사례>

오류 유형	영어에서 구별	한국인의 발음	실제 예시
/r/ 대 /l/	완전히 다른 두 소리	혼용 (어말 Dark /l/ 불가)	right → light, feel의 어말 /l/ 탈락
/f/ 대 /v/	이와 입술의 마찰음	/ㅍ/, /ㅂ/ 로 대체	five → 파이프, vote → 보트
/θ/, /ð/	혀 끝이 이 사이로 나오는 소리	/ㅅ/, /ㄷ/ 로 대체	think → sink, this → dis

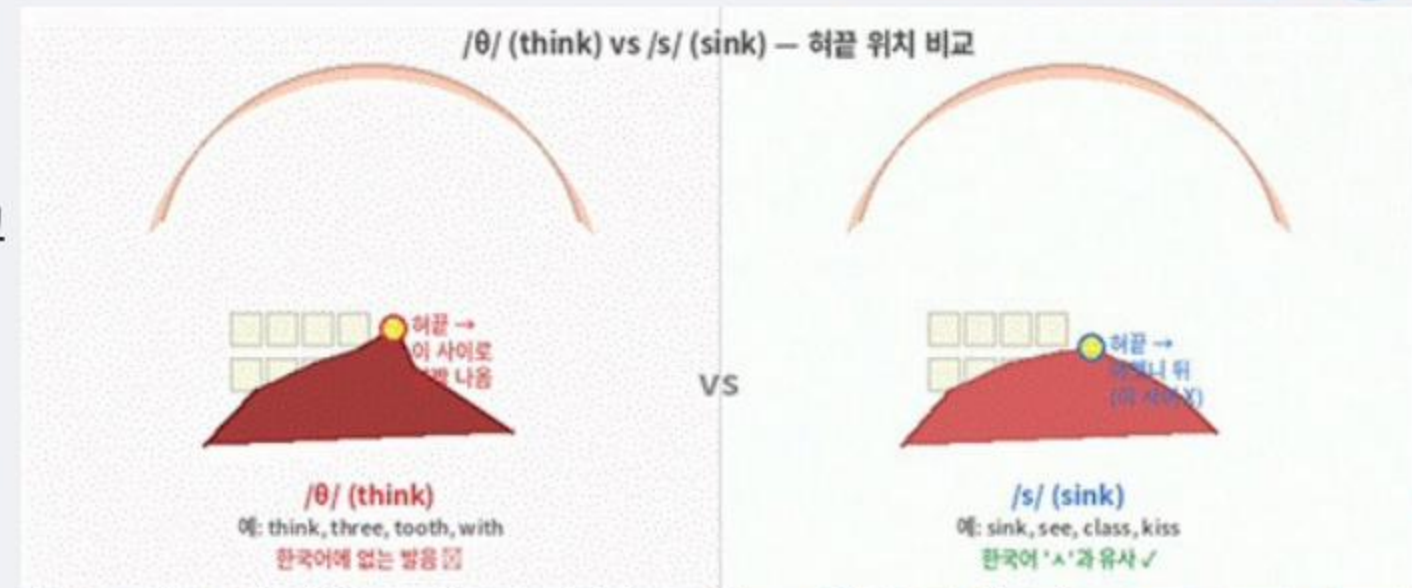
4. 부록



▲ 그림 1. /r/ vs Clear /l/ (어두) vs Dark /l/ (어말) — 혀의 위치 비교



▲ 그림 2. /f/ vs /p/, /v/ vs /b/ — 입술과 이의 접촉 방식 비교



▲ 그림 3. /θ/ (think) vs /s/ (sink) — 혀끝 위치 비교

4. 부록

1) 한국어와 영어 음운 체계의 근본적인 차이

<한국인의 영어 발음 오류 사례>

오류 유형	영어에서 구별	한국인의 발음	실제 예시
모음 삽입	자음이 연속되는 구조 유지	자음 사이에 모음 삽입	spring → 스프링 strike → 스트라이크
강세, 리듬	강세 받는 음절만 두드러지게	모든 음절 비슷하게 발음	비강세 모음을 약하게 처리하지 못함

4. 부록

2) 역할 분담

1. 음소 탐지 모델 학습 및 실험 - 신영환
2. 백엔드 API 서버와 오디오 처리 - 김지용
3. 프론트엔드 인터페이스와 피드백 시각화 - 이찬영, 최현지
4. QA와 데이터 처리 - 김형준, 백지영