

ECHO!

: 한국어 L1 화자를 위한 영어 발음 음소 단위 피드백 서비스

A6 O(1) 설계 발표

2026-04-15

김형준, 김지용, 백지영, 이찬영, 신영환, 최현지

목차

CONTENTS

00	배경	06	결과물 명세
01	성능 지표 이해	07	상세 설계
02	실험 설정	08	업무 분담과 일정
03	음소 인식기 성능	09	달성 목표와 테스트
04	피드백 생성	10	프로젝트 의의
05	문제와 접근		

00

SECTION

배경

00

발음 피드백 서비스의 표준 구현

SECTION

배경

학습자가 "think"를 발음했을 때

1

학습자 발화

s I ŋ k



2

정답

θ I ŋ k

발화

s I ŋ k

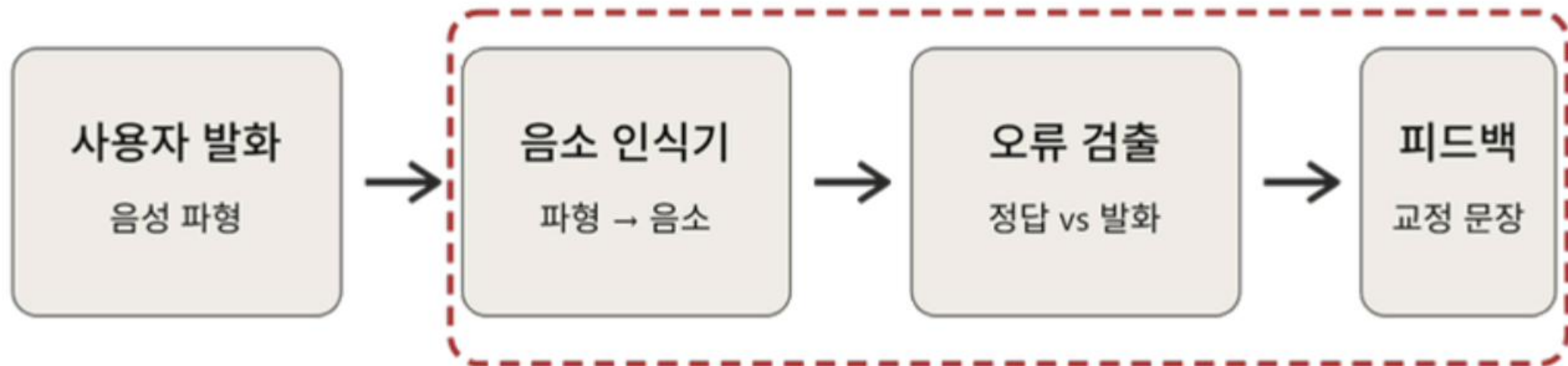


3

혀끝을 윗니와 아랫니 사이에 살짝 내밀고
바람을 불어 보세요

CAPT 시스템의 일반적 구성

Computer-Assisted Pronunciation Training



우리가 개선할 수 있는 영역

인식 성능 · 검출 정확도 · 피드백 품질

00

발음 피드백 서비스의 각 모듈별 기능

SECTION

배경

음소 인식기의 입력과 출력

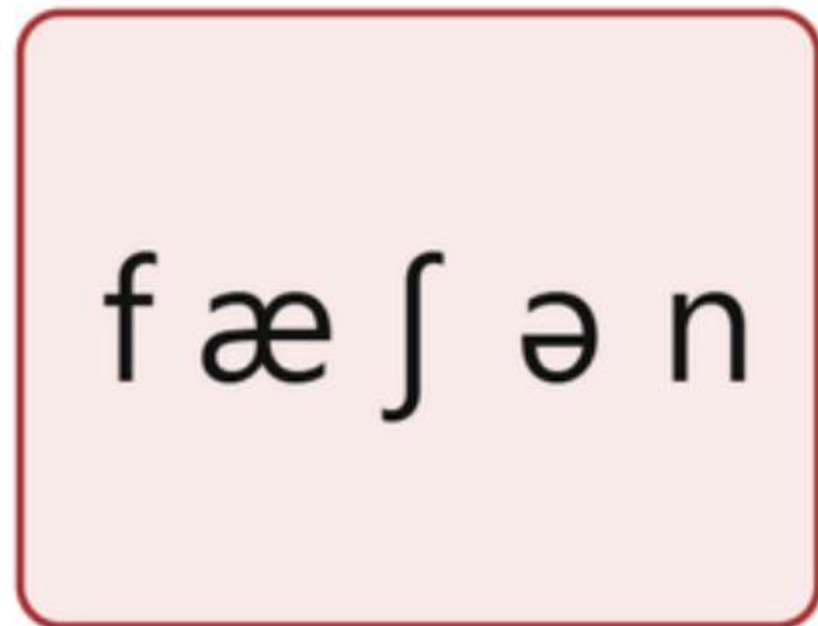
입력



음소 인식기



출력



파형에서 단어가 아닌 음소 단위로 변환된다
단어 단위로는 어느 음소가 틀렸는지 짚을 수 없기 때문

오류는 어떻게 검출되는가

정답 음소열과 인식된 음소열을 나란히 정렬해 다른 위치를 찾는다

정답

(canonical)

f

æ

ɪ

ə

n

⋮

⋮

⋮

⋮

⋮

발화

(perceived)

p

æ

ɪ

ə

n

첫 자리에서 /f/ 가 /p/ 로 대체됨

= 치환 오류 (substitution)

이외에도 음소가 빠지면 삭제(deletion), 끼어들면 삽입(insertion)

오류 유형 ② — 음소가 빠진 경우

예시 단어: "asked"

정답
(canonical)

æ

s

k

t

⋮

⋮

⋮

발화
(perceived)

æ

s

k

—

어말 /t/ 가 탈락

= 삭제 오류 (deletion)

한국어 화자에게 흔한 어말 자음 탈락 패턴

오류 유형 ③ — 없던 음소가 끼어든 경우

예시 단어: "spring"

정답

(canonical)

s	—	p	r	ɪ	ŋ
⋮	⋮	⋮	⋮	⋮	⋮
s	ʍ	p	r	ɪ	ŋ

발화

(perceived)

자음군 사이에 /ʍ/ 가 끼어듦

= 삽입 오류 (insertion)

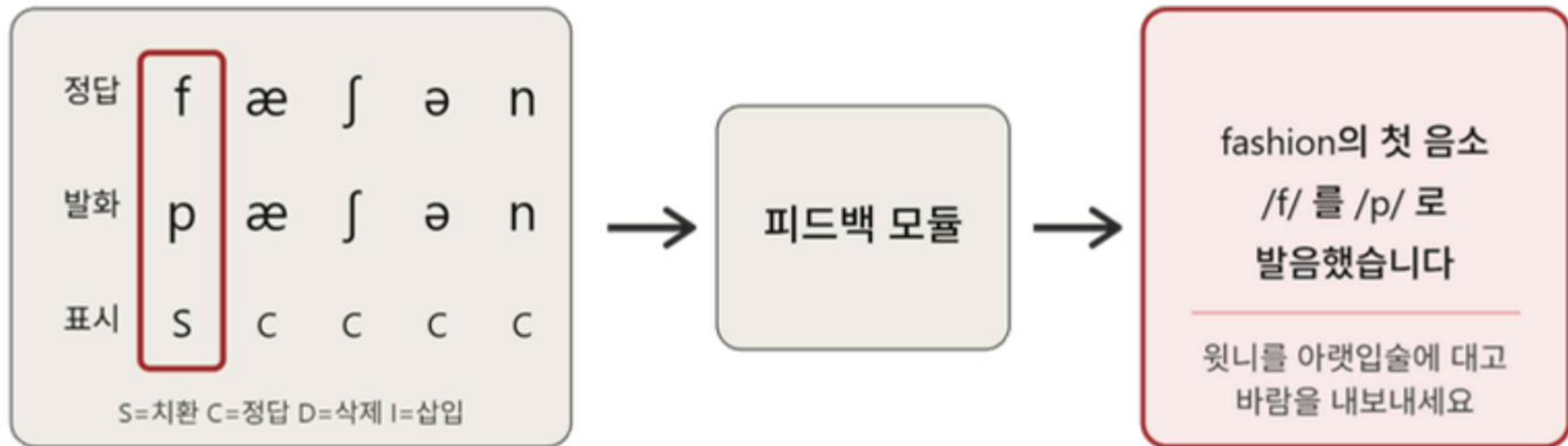
spring → 스프링. 한국어의 모음 삽입 습관에서 비롯

정렬 알고리즘: Levenshtein distance · 표준 방식

피드백 모듈의 입력과 출력

입력: 오류 정렬 결과 전체

출력: 한국어 피드백



전체 정렬 결과를 한꺼번에 받아 문맥을 반영한 피드백을 생성한다

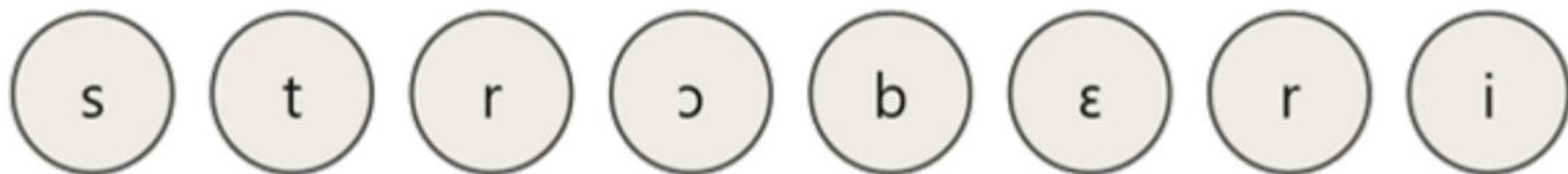
01

SECTION

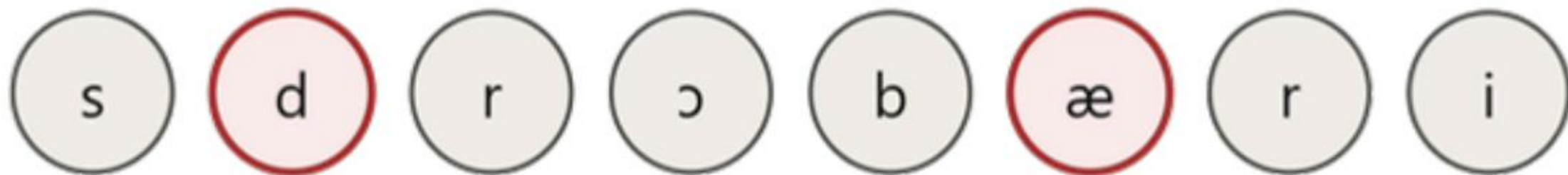
성능 지표 이해

학습자가 "strawberry"를 발음했다

정답



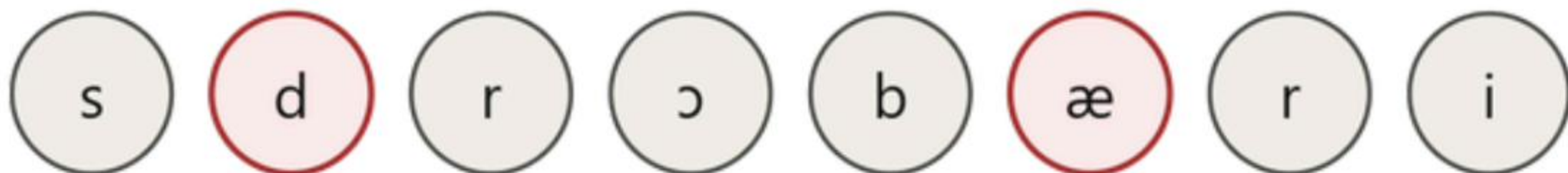
실제 발화



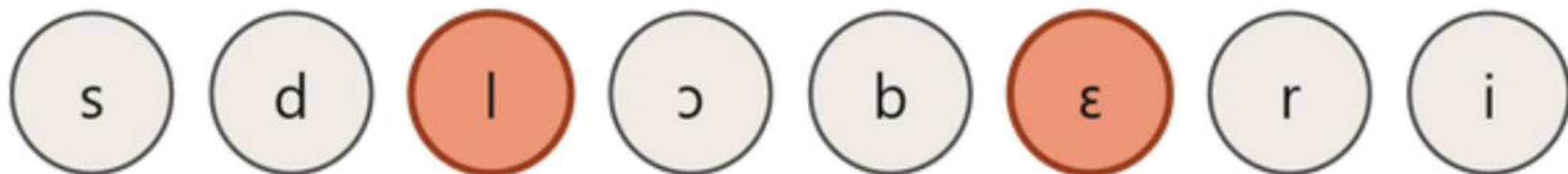
8개 음소 중 2개가 오발음

PER — 인식기의 듣기 정확도

실제 발화



인식기 출력



8개 중 2개 잘못 들음

PER = 25%

MDD F1 — 오류 탐지 종합 점수

정답	s	t	r	ɔ	b	ε	r	i
실제 발화	s	d	r	ɔ	b	æ	r	i
인식기 출력	s	d	l	ɔ	b	ε	r	i
판정	○	○	×	○	○	○	○	○

정답과 인식기 출력을
자리별로 비교
같으면 ○ 다르면 ×

오발음 2개 중 1개 탐지
정답 6개 중 1개 오판
 $F1 = 0.5$

MDD F1 — 오류 탐지 종합 점수

정답	s	t	r	ɔ	b	ε	r	i
실제 발화	s	d	r	ɔ	b	æ	r	i
인식기 출력	s	d	l	ɔ	b	ε	r	i
판정	○	○	×	○	○	○	○	○

Precision

오류라 판정한 것 중 진짜 오류 비율

$$1 / (1 + 0) \\ = 1.00$$

Recall

실제 오류 중 잡아낸 비율

$$1 / (1 + 1) \\ = 0.50$$

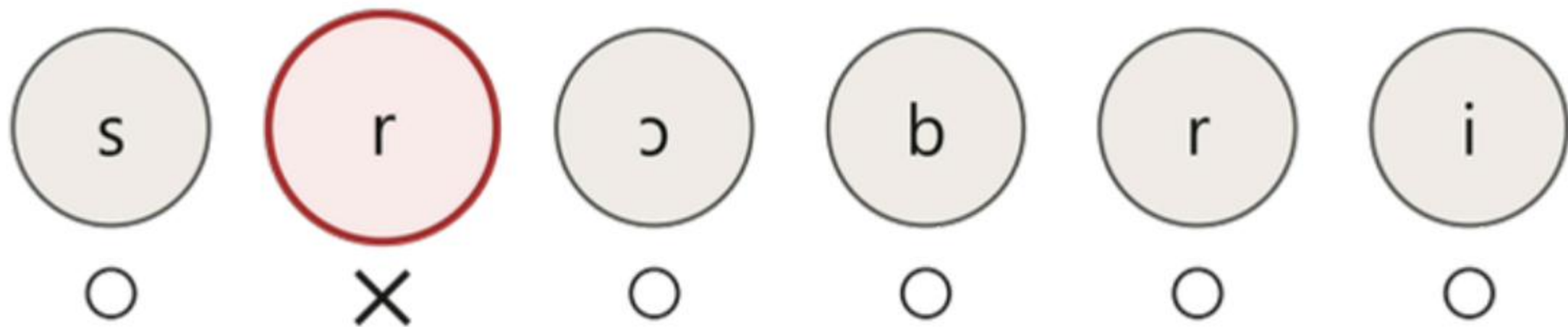
F1

P 와 R 을 합친 종합 점수

$$2 \cdot P \cdot R / (P + R) \\ = 0.50$$

FRR — 억울한 오판 비율

정답이었던 6개 음소만 따로 보면



6개 중 1개가 억울한 오판

$\text{FRR} \approx 16.7\%$

02

SECTION

실험 설정

학습 데이터

L2-ARCTIC

3,496

발화

6개 L1 비원어민
영어 발화 데이터

AI-Hub

61,990

발화

한국인 학습자의
영어 발화 데이터

+

총 65,486 발화 · 화자 906명

베이스라인 모델

Wav2Vec2-base

Facebook AI · 자기지도 학습 음성 표현

94.4M

파라미터

42

음소 분류

CTC

loss function

파형을 입력받아 매 프레임마다 어떤 음소인지 예측하도록 학습

03

SECTION

음소 인식기 성능

베이스라인 결과

테스트 9,835 발화 · 419,743 음소

PER

15.93%

음소 6개 중 1개를
잘못 인식

MDD F1

0.765

오류 탐지
종합 점수

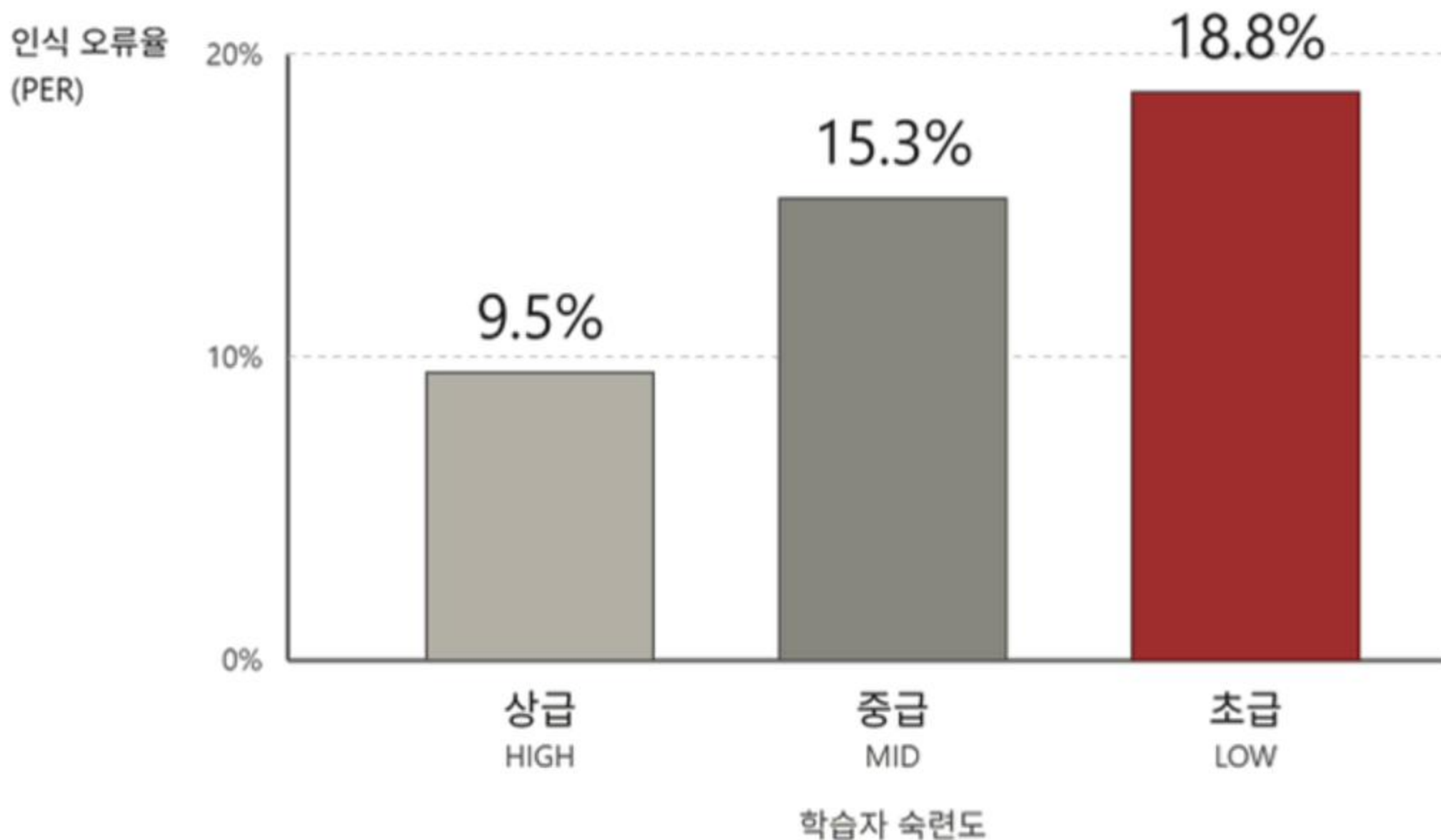
FRR

5.83%

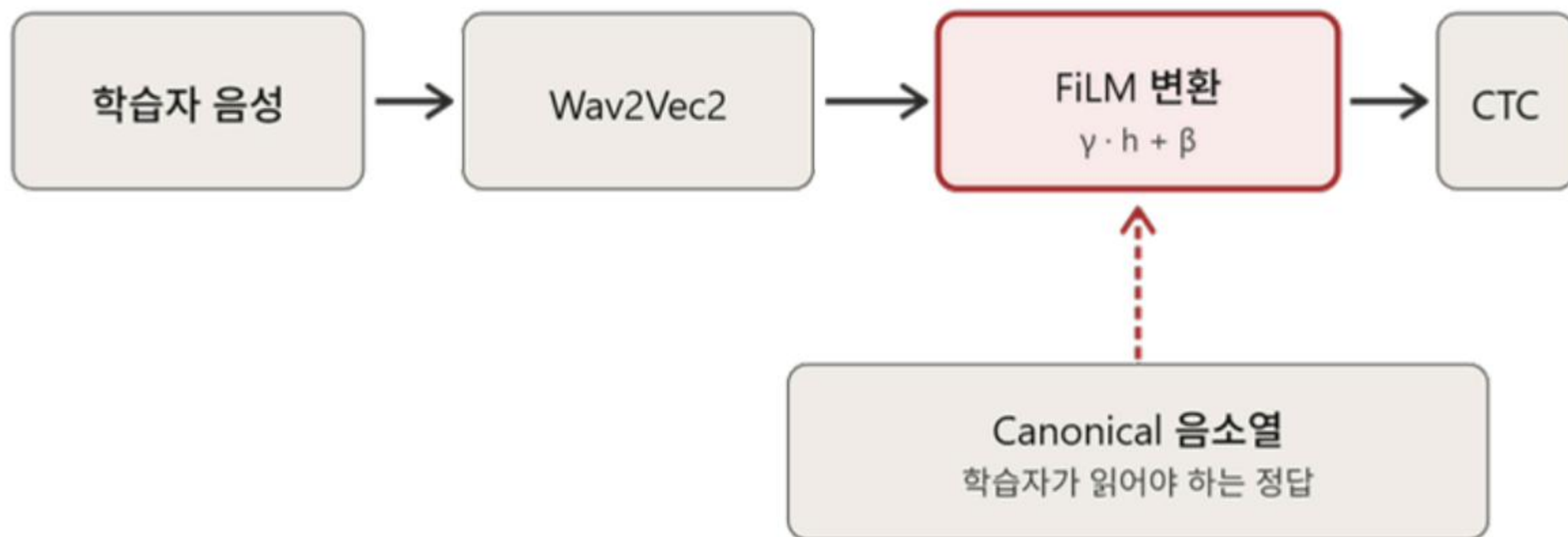
정답 100개 중 약 6개가
억울하게 오류로 판정

우리가 집중해서 줄이려는 지표는 FRR

한국어 억양이 강할수록 인식기가 헛갈린다



접근 — 정답 음소를 인식기에 주입



정답 정보로 표현을 보정해 FRR 저감

04

SECTION

피드백 생성

한국인 학습자 오발음의 전형 패턴

베이스라인 실험에서 직접 관찰된 패턴 · 치환 오류 59.6%

자음 대체

fashion → pashion

치간음 대체

think → sink

모음 삽입

spring → swpring

어말 자음 탈락

asked → as—

기존 피드백은 모국어와 연결되지 않는다

스픽 등 영어 학습 앱의 피드백 방식

기존 설명

puk → buk

성대를 진동시키며
입술을 다물었다
떼어 보세요

영어 음운 기준

학습자의 반응

한국어 ㅂ 으로는
둘 다 똑같이
[북] 으로 들리는데

뭐가 다른가

모국어 직관과 연결 안됨

학습자의 모국어를 기준으로 삼는 피드백이 필요

구한말에는 영어 b 를 위한 한글이 있었다

1908년 출간된 영어 교재 '아학편' 은
두 가지 ㅂ 소리를 다른 글자로 구분해 적었다

ㅂ

한국어의 일반적 ㅂ

영어 /p/ 에 가까움

성대 진동 없음

ㅃ

시옷계 합용병서

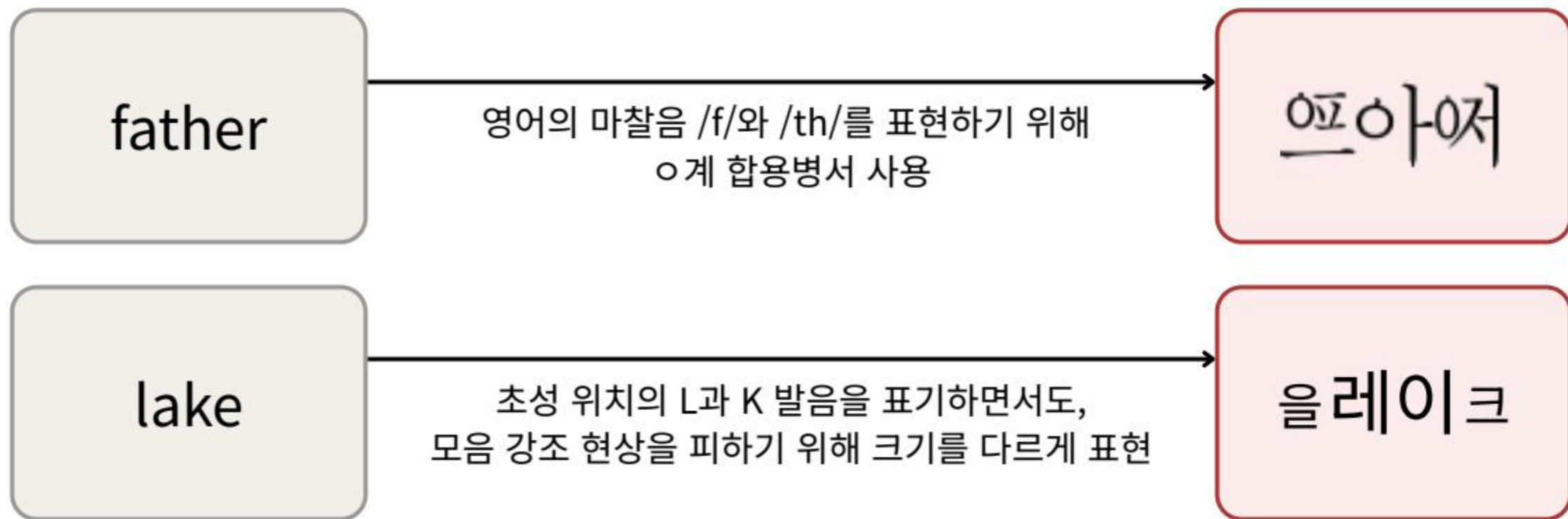
영어 /b/ 에 대응

성대 진동 있음

한국어 ㅂ 만으로는 영어 b 를 표현할 수 없다

100년 전 교사들이 이미 알고 있던 사실

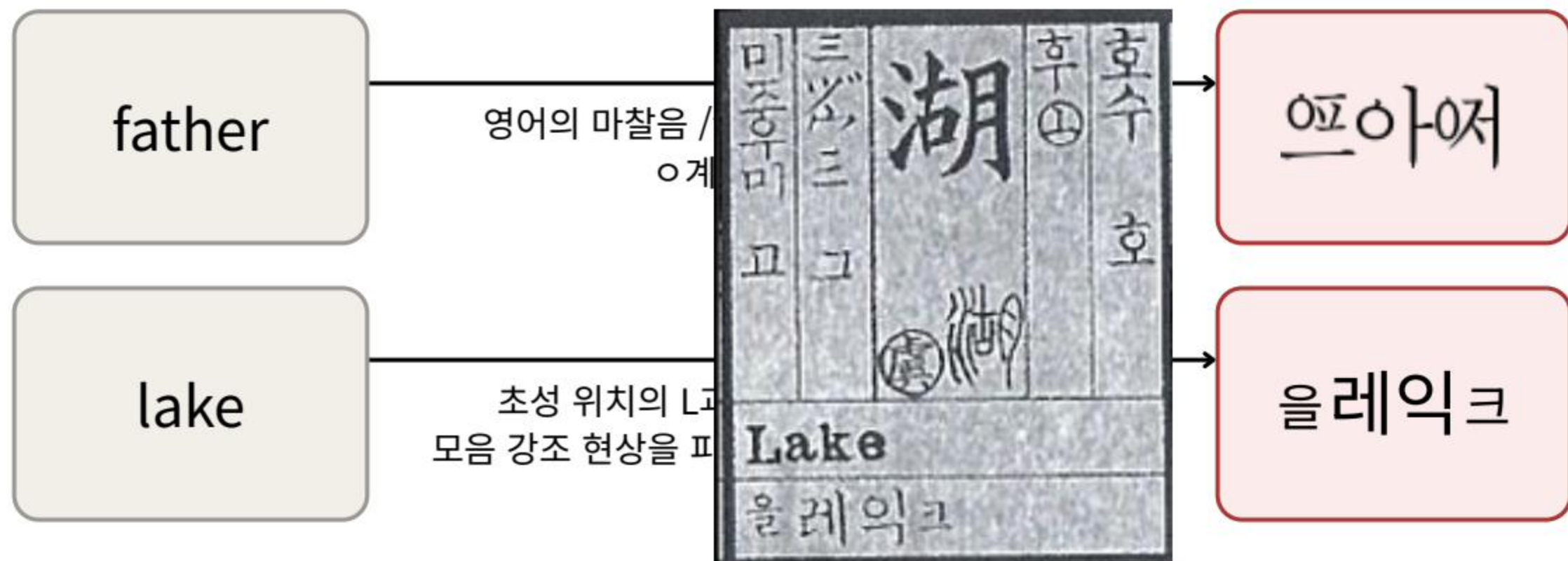
한국인의 영어 오발음은 한국어 내에서 같은 음소로 분류되는
다양한 음성 간 차이를 인지하지 못하는 것이 원인



→ 지석영·전용규의 <아학편>에서는 영어의 음소를 여러 가지 한글 병음으로 세분화하여 표현함.

사용자에게 자칫 비슷하게 느껴질 수 있는 **음성들 간 차이를 인식시키고,**
이에 따라 달라지는 **조음 위치를 인식시키기** 위해 활용

한국인의 영어 오발음은 한국어 내에서 같은 음소로 분류되는
다양한 음성 간 차이를 인지하지 못하는 것이 원인



→ 지석영·전용규의 <아학편>에서는 영어의 음소를 여러 가지 한글 병음으로 세분화하여 표현함.

사용자에게 자칫 비슷하게 느껴질 수 있는 **음성들 간 차이를 인식시키고,**
이에 따라 달라지는 **조음 위치를 인식시키기** 위해 활용

이 원리를 그대로 피드백에 적용한다

상황: 학습자가 book 의 b 를 한국어 ㅂ 으로 발음

학습자 발음

book [ㅂ옥]

한국어 ㅂ 은 영어 p 에 가까움
b 의 성대 진동이 빠짐

교정 지시

book [ㅅ옥]
= [으ㅂ옥]

아학편의 ㅅ 처럼
"으" 를 앞에 붙여 성대를 먼저 진동

100년 전 ㅅ 을 현대식 발음 팁으로 옮긴다

[으ㅂ옥] 이라는 표기는 ㅅ 의 성대 진동 원리를 누구나 따라할 수 있게 만든 것

05

SECTION

문제와 접근

정리

문제

접근

① 음소 인식기

FRR 5.83%
정답 발화 오판



FiLM Canonical Injection
정답 정보 주입

② 피드백 생성

피드백 이해 불가
영어 기준 설명



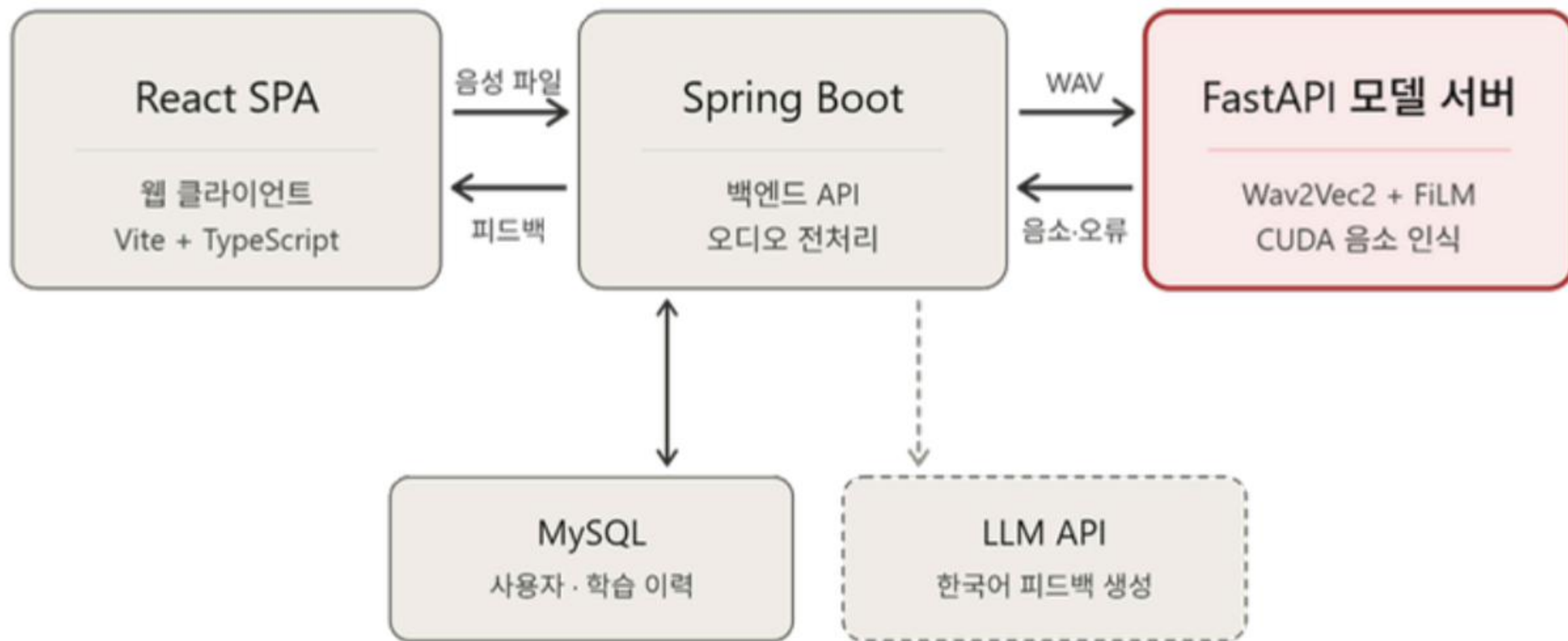
아학편 L1 피드백
한국어 음운 기준

06

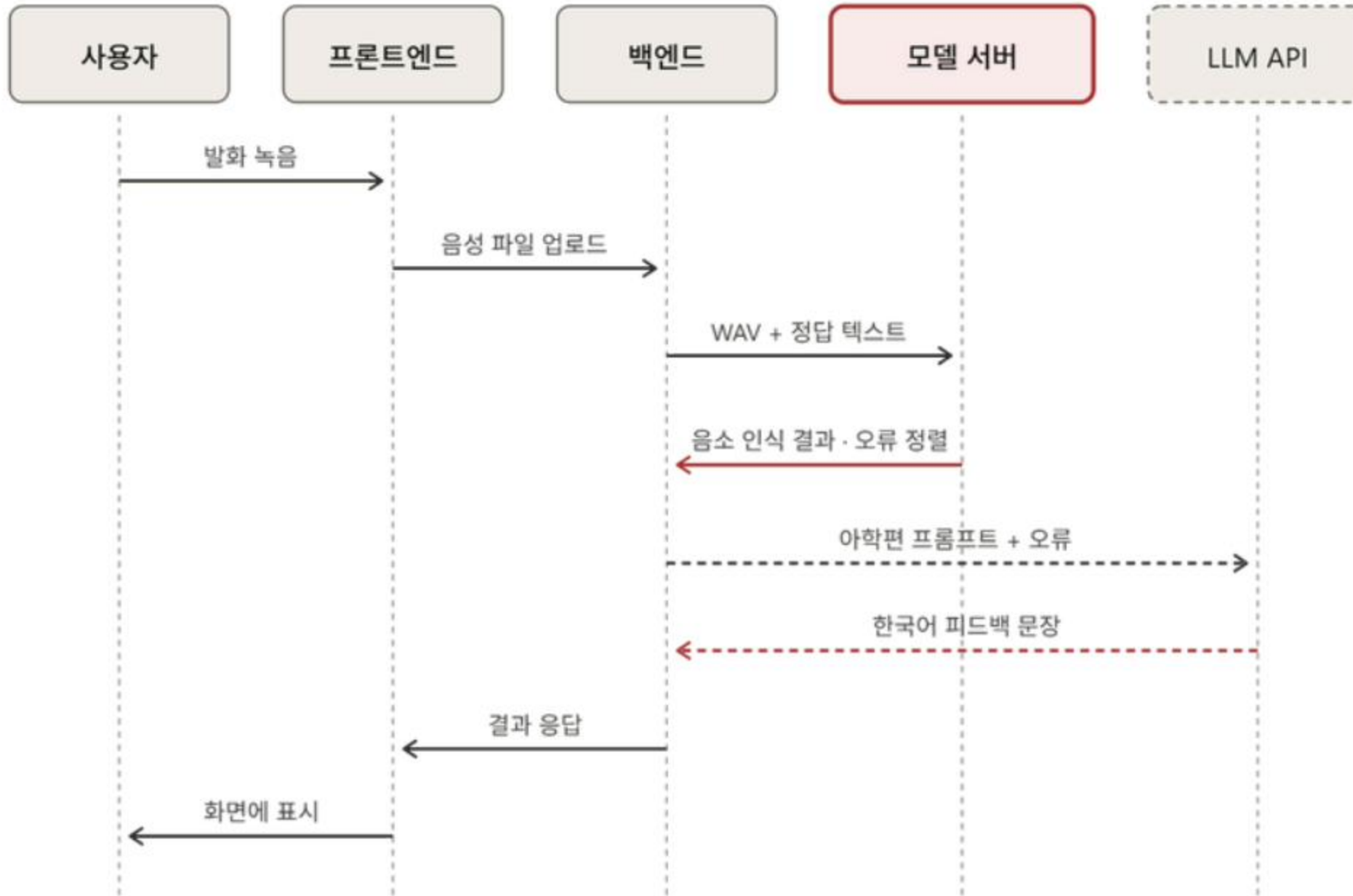
SECTION

결과물 명세

ECHO 시스템 구성도



발음 평가 파이프라인



실선 = 내부 호출 · 점선 = 외부 API

기술 스택

프론트엔드

React 18

TypeScript · Vite

Tailwind CSS

Radix UI · MUI

Recharts

Motion

React Router

SPA · 웹 클라이언트

백엔드

Spring Boot 4

Java 25 · Gradle

Spring Data JPA

Hibernate

Spring Security

MySQL · H2

LLM API client

REST API · 오디오 전처리

모델 서버

FastAPI

Python · Uvicorn

PyTorch 2.7

CUDA 12.8

Transformers

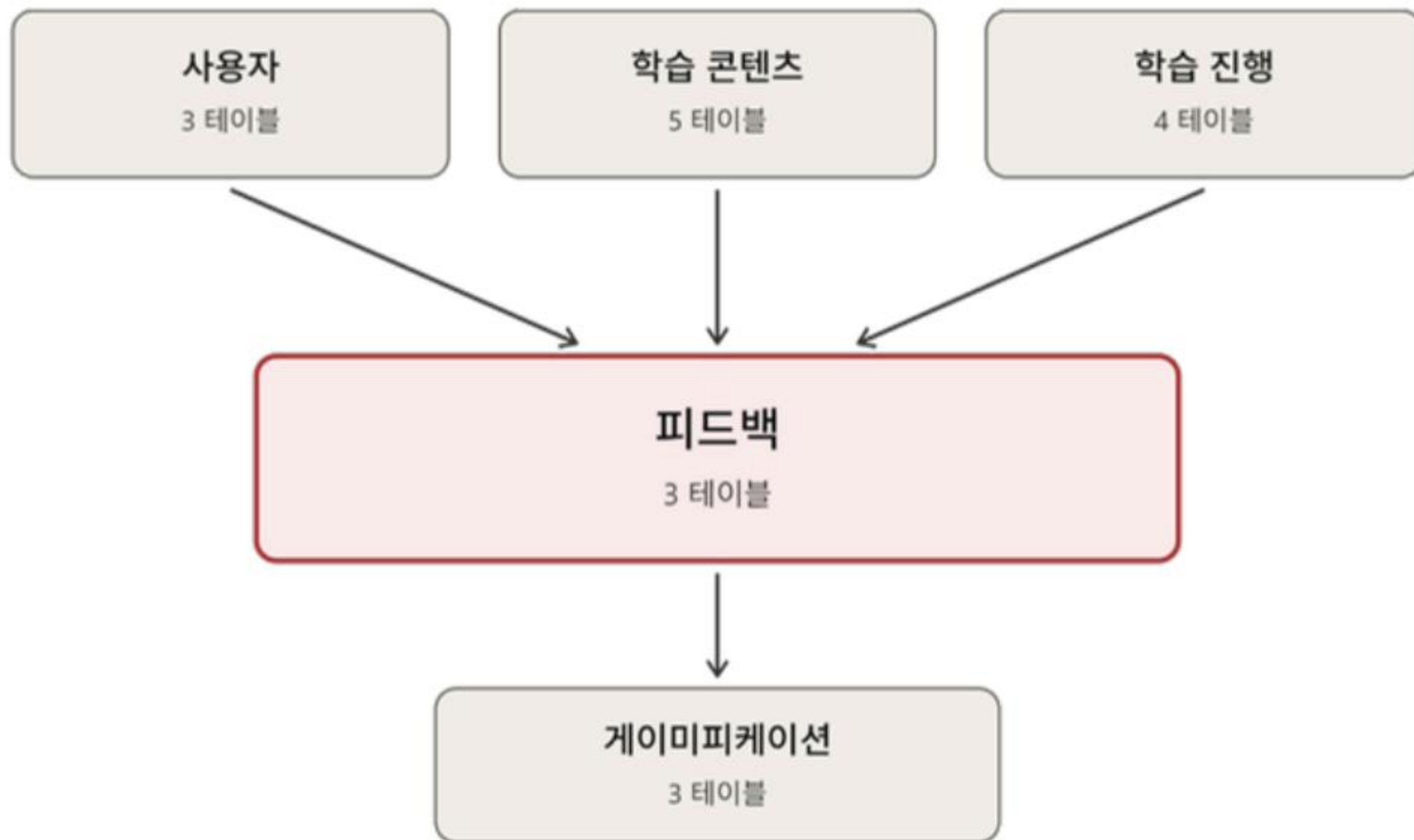
torchaudio

NumPy

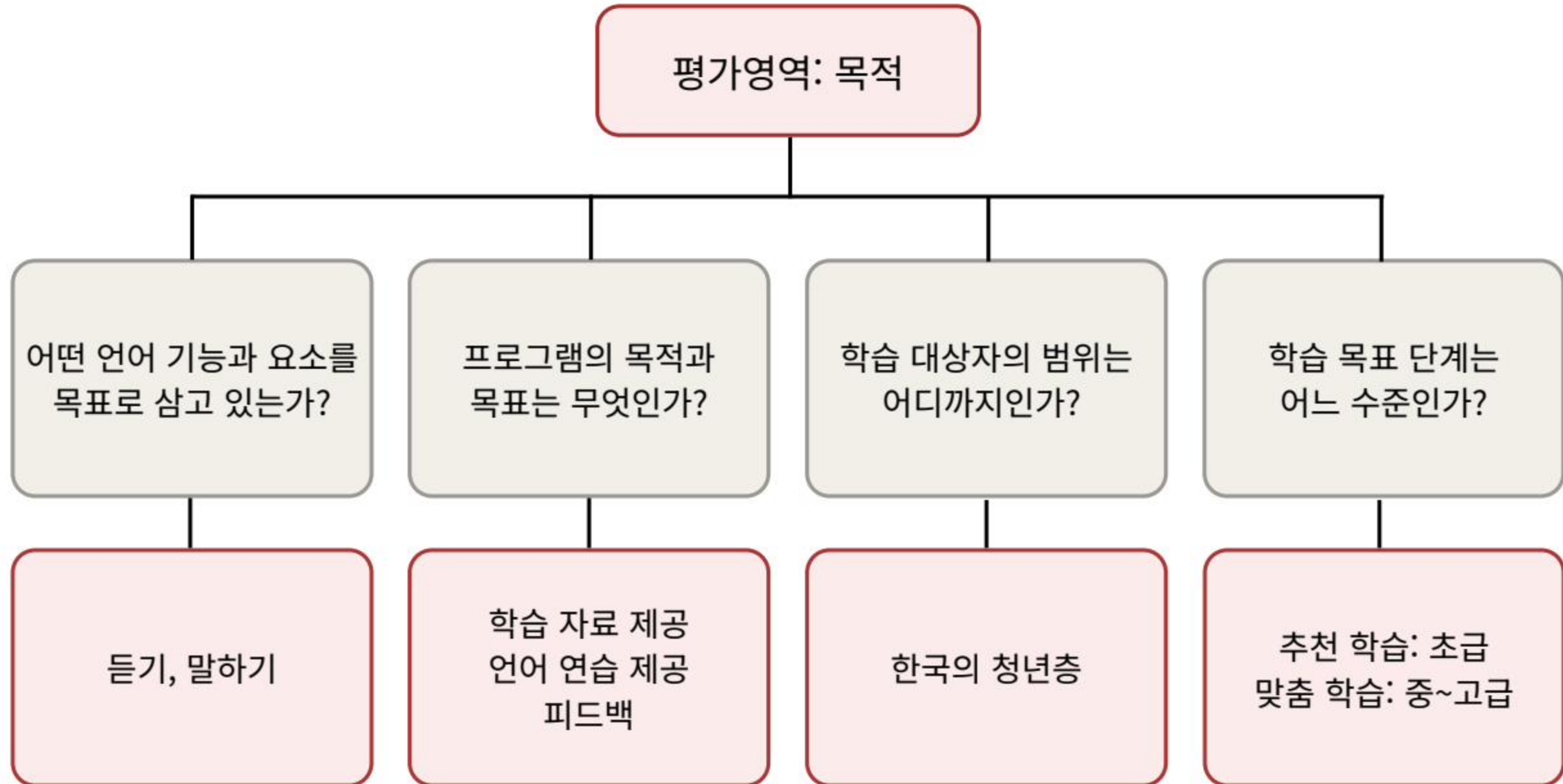
음소 인식 추론 서버

데이터베이스 구조

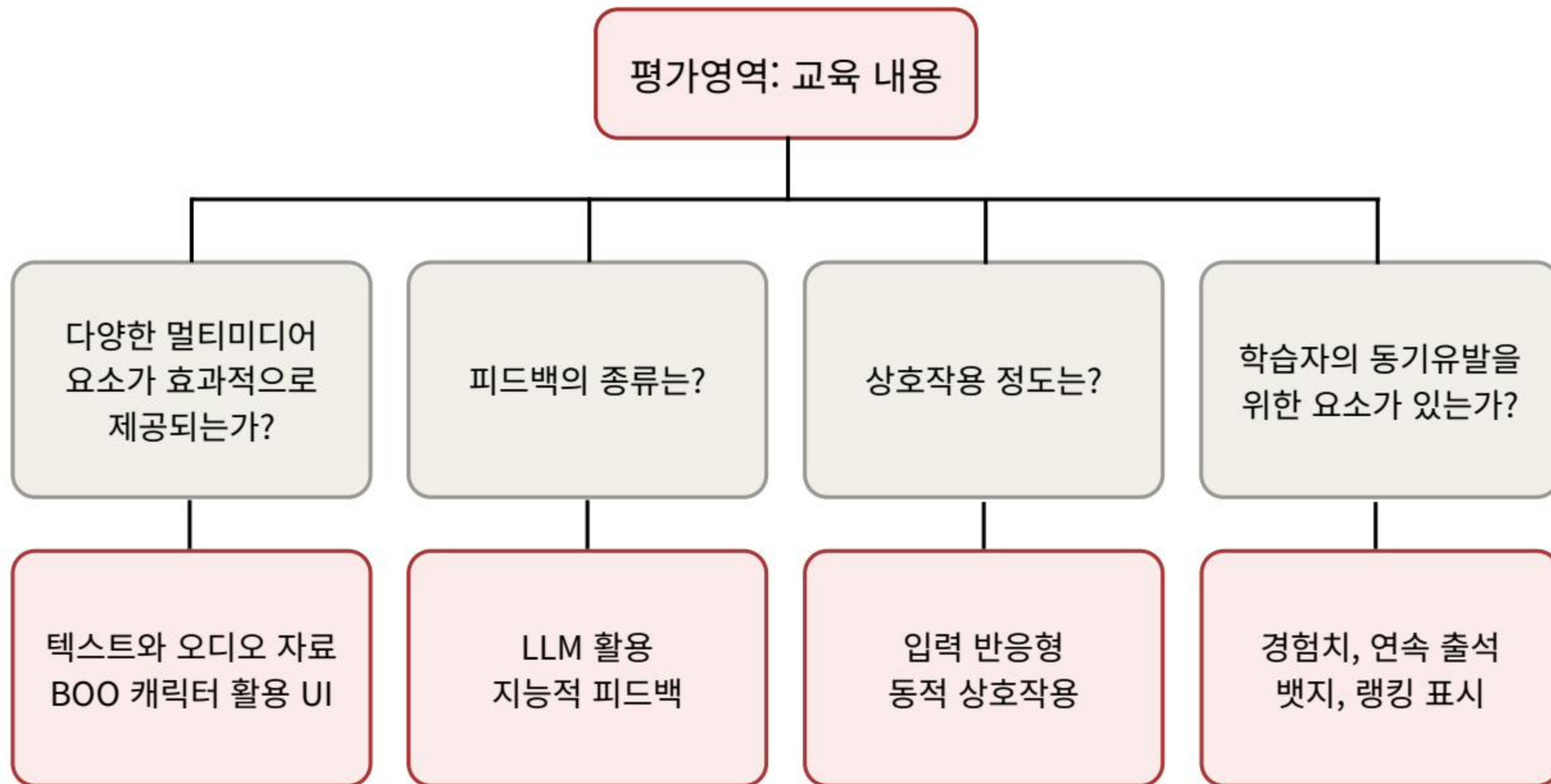
ERD · 18 테이블 5 영역



ECHO 기능 구성의 기준



ECHO 기능 구성의 기준



UI 1: 로그인/회원가입

ECHO!

아이디

비밀번호

로그인

또는



회원가입

© O(1)

회원가입

ECHO!에 오신 것을 환영합니다

아이디

비밀번호

비밀번호 확인

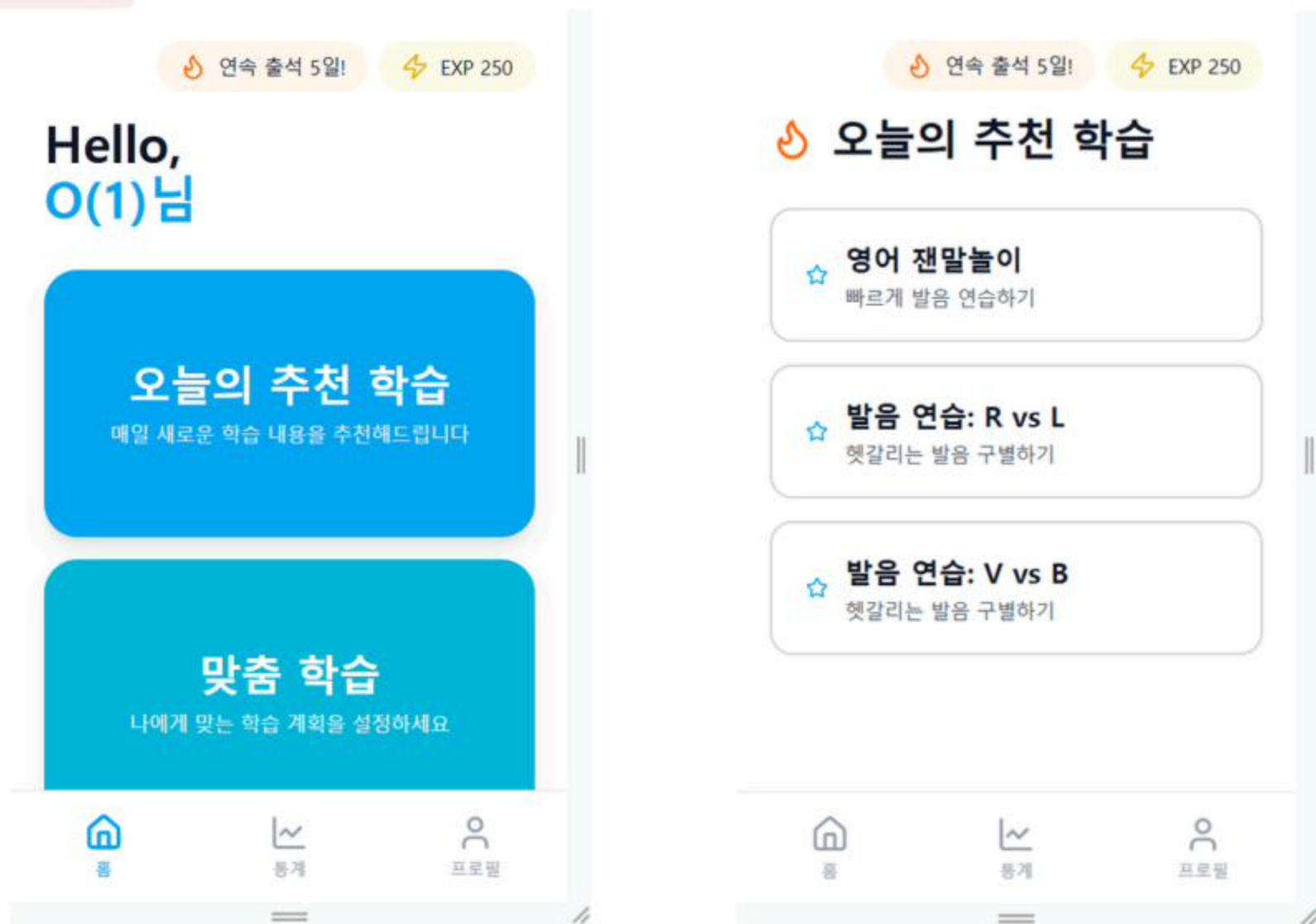
닉네임

이메일

☐ 서비스 이용약관에 동의합니다

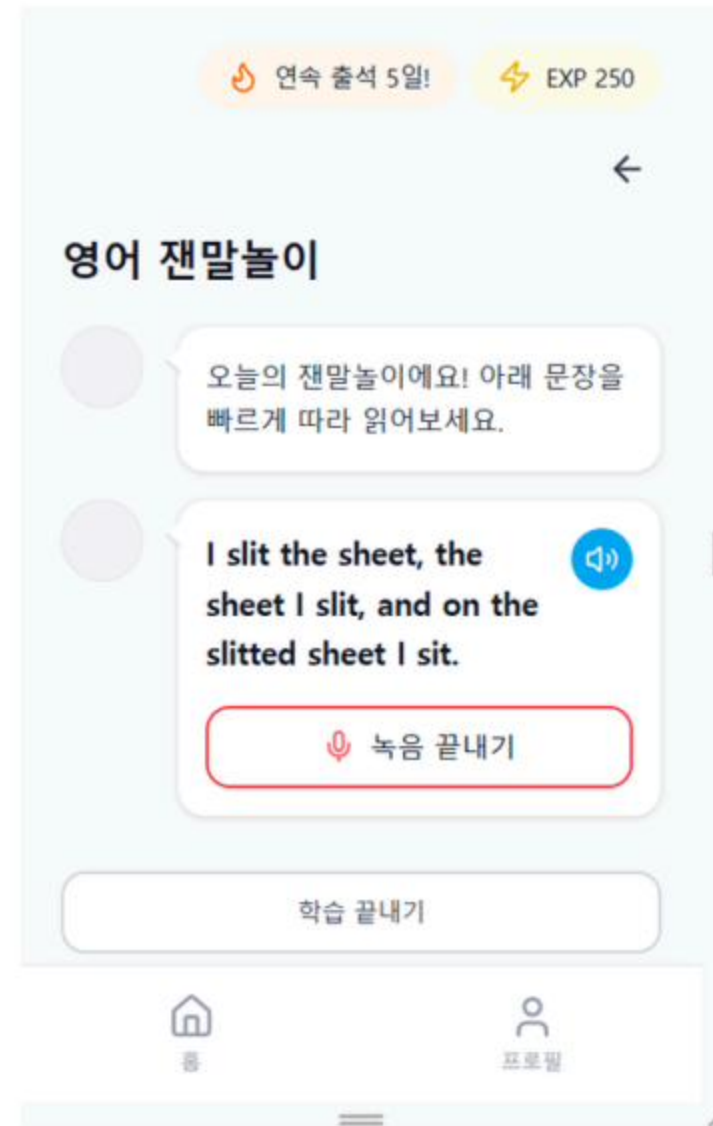
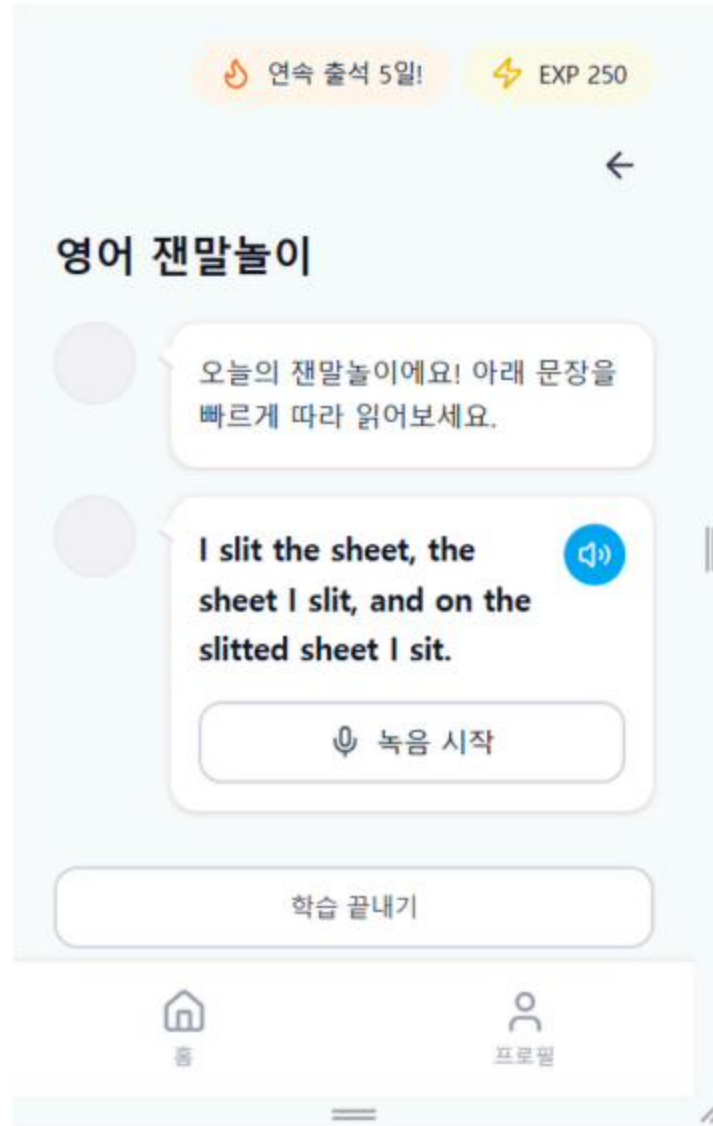
소셜 로그인 구현 예정

UI 2: 메인 화면 (오늘의 추천 학습 / 맞춤 학습)



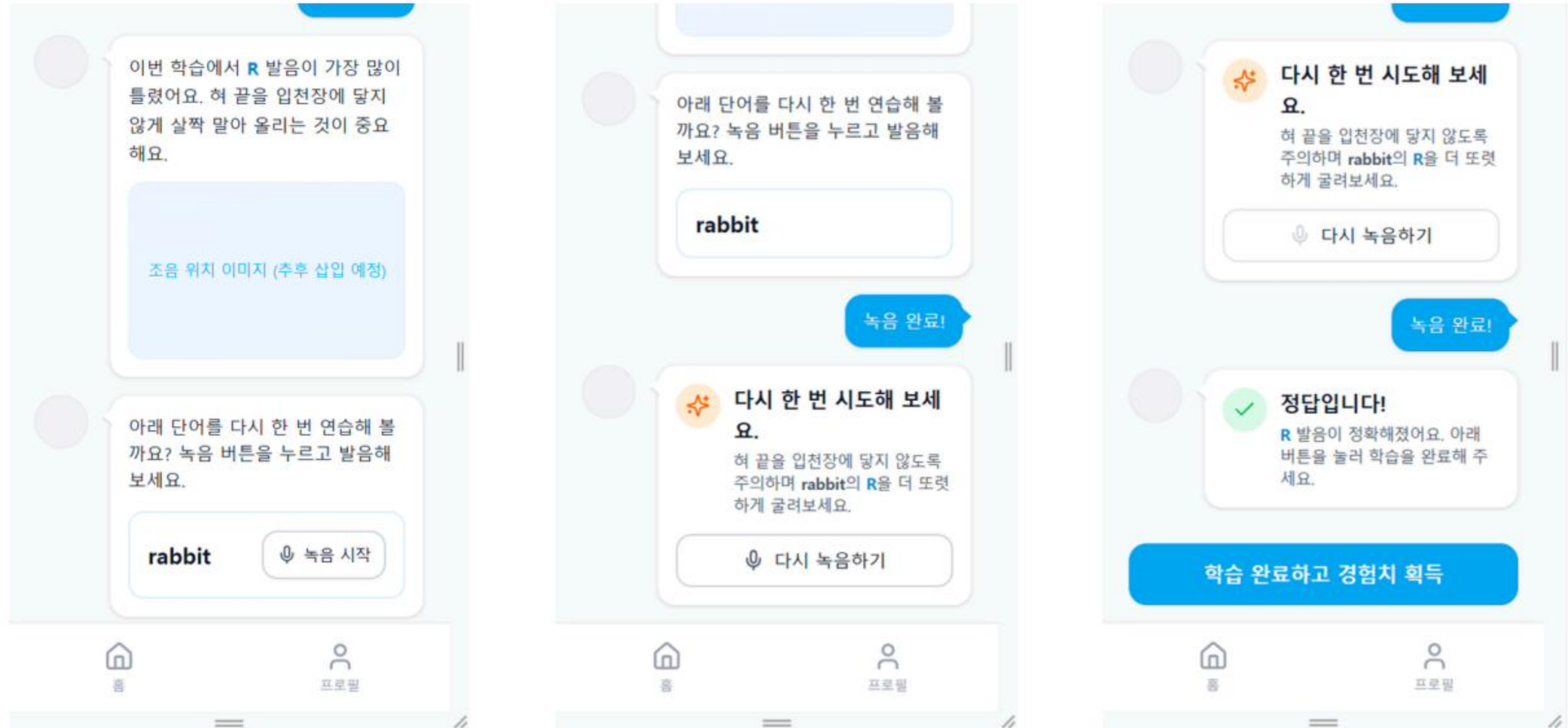
오늘의 추천 학습과 사용자 맞춤 학습
추천 학습 (초급) : 사용자 공통 유닛 / 맞춤 학습 (중~고급) : 사용자 개별 입력 유닛

UI 3: 오늘의 추천 학습 (학습 채팅창)



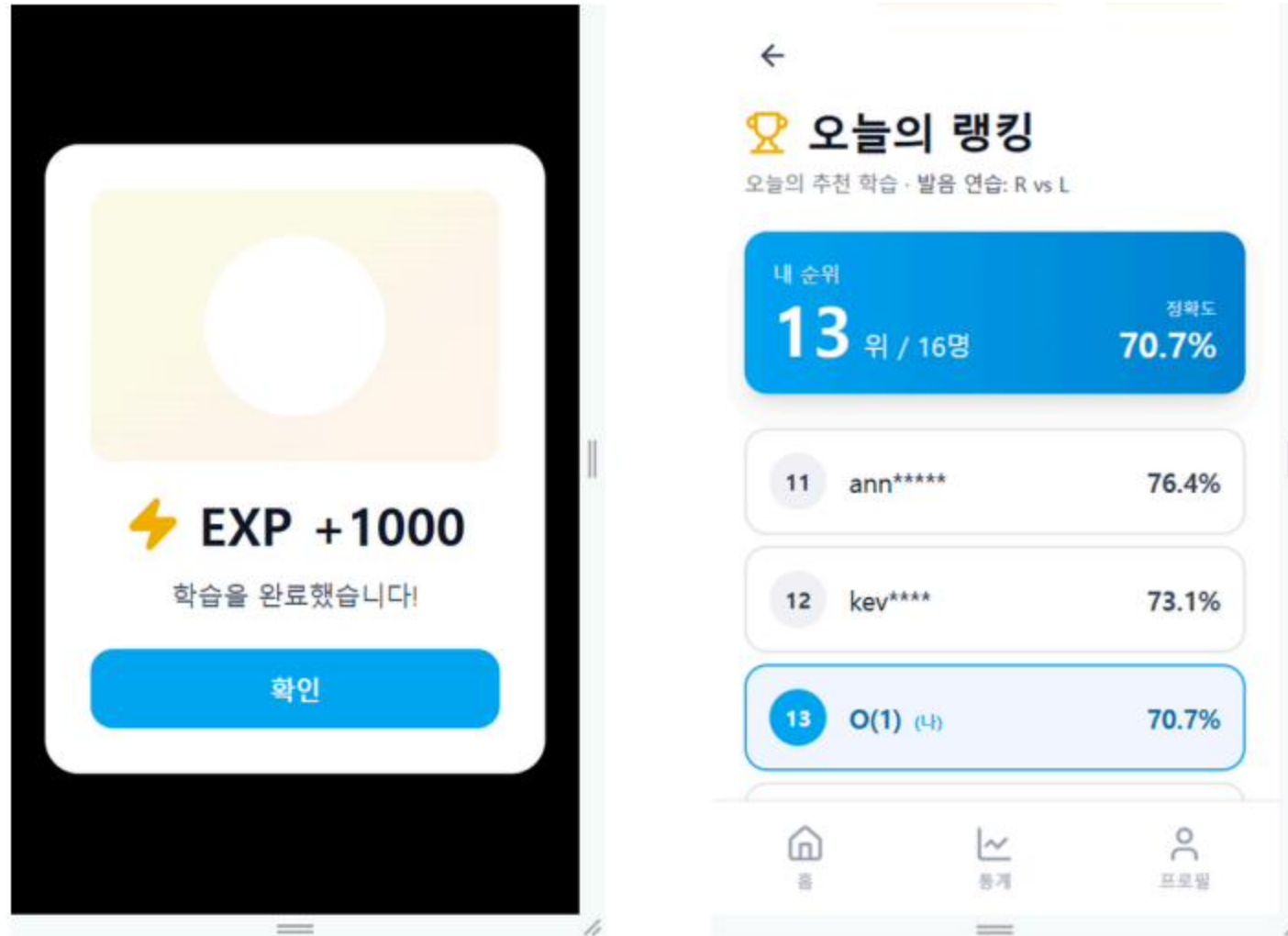
녹음 시작 버튼을 눌러 음성 녹음 시작, 끝내기 버튼으로 제출

UI 4: 피드백 확인



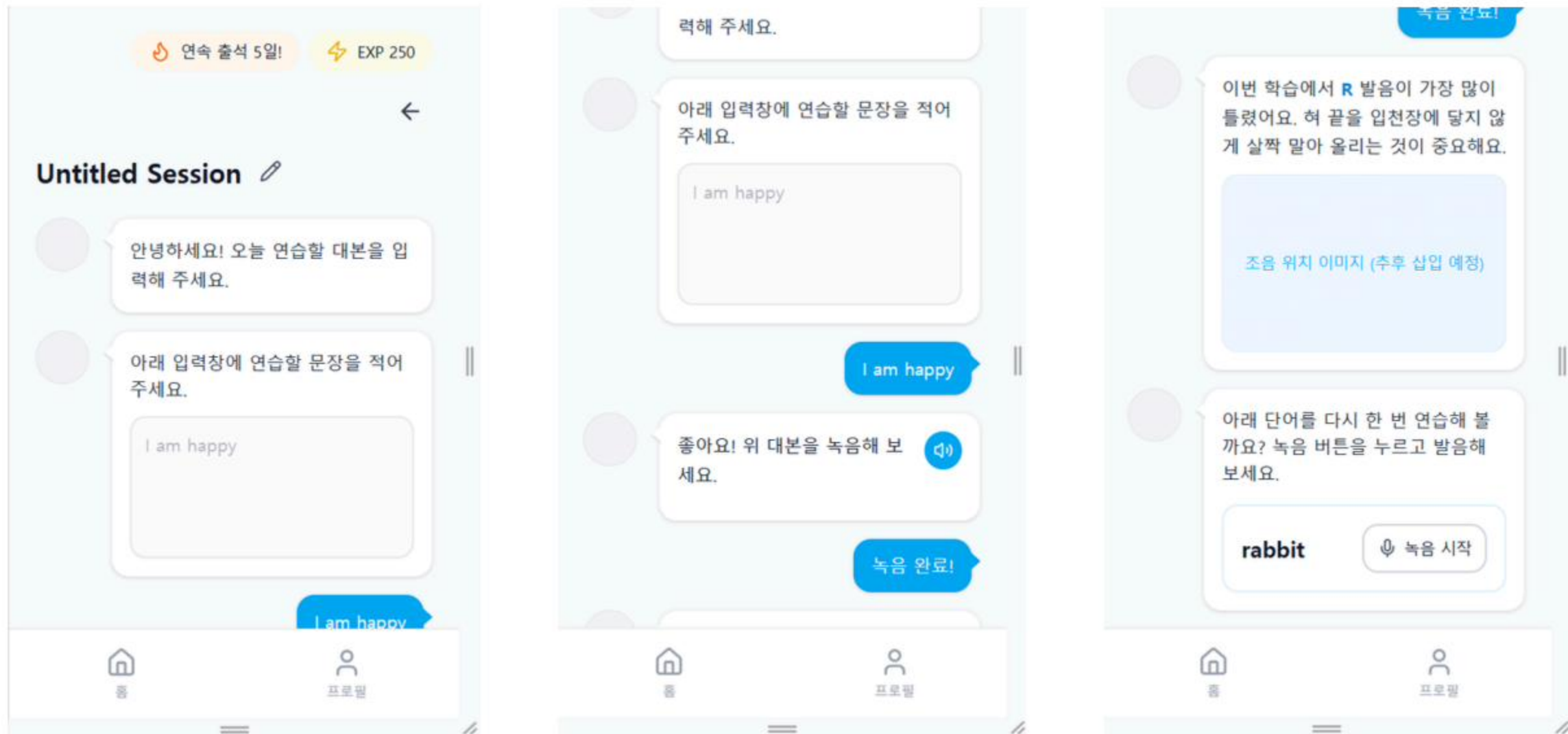
사용자 음성을 분석해 생성된 피드백을 읽고 정확하게 발음할 때까지 재녹음

UI 5: 경험치 획득과 랭킹



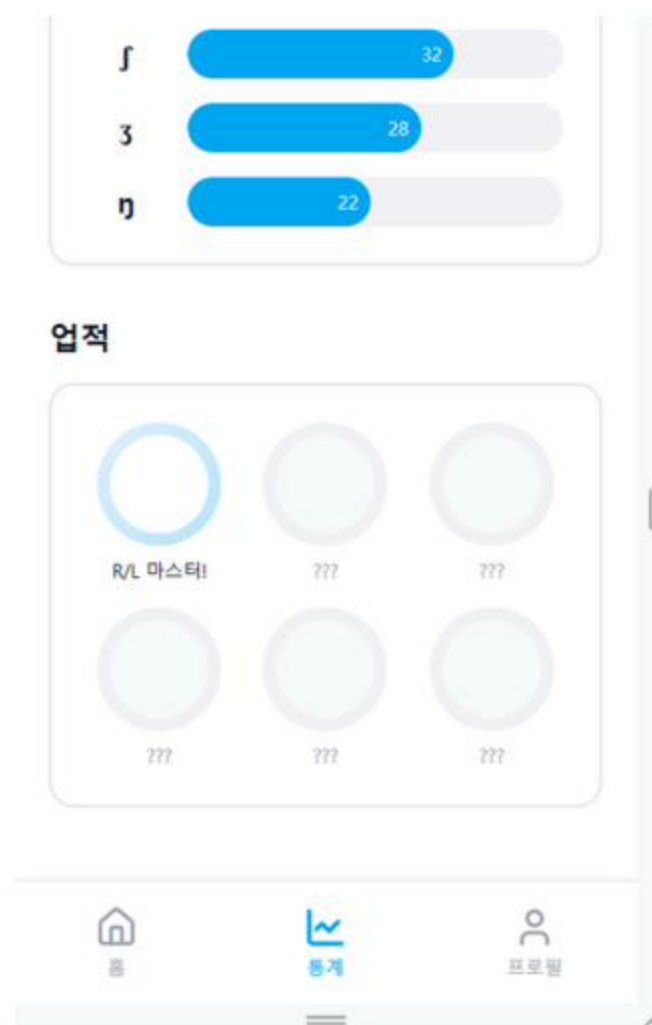
학습을 끝내면 경험치를 획득하고, 종합 지표 기반 사용자 랭킹 표시

UI 6: 맞춤 학습 (사용자 스크립트 입력)



맞춤 학습 탭에서는 사용자 입력 스크립트 기반 발음 피드백 제공

UI 7: 통계 및 프로필



경험치, 연속 출석 체크, 배지 컬렉션, 지난 일주일간 많이 틀린 발음 표시

사용 사례 — 학습자 A · 추천 학습

OPIc IM1에서 막힌 막학기 취준생 · 발음 때문에 점수가 오르지 않음

1

앱 실행

메인 화면에서 오늘의 추천 학습 진입

2

발음 녹음

"I slit the sheet, the sheet I slit..." 따라 읽기

3

한국어 피드백 수신

"R 발음에서 혀끝이 입천장에 닿지 않게 살짝 말아 올리세요"

4

재녹음 후 통과

정답 처리 · 경험치 +50 · 연속 출석 5일 유지

사용 사례 — 학습자 A · 맞춤 학습

저녁에 OPIc 시험에서 외운 답변 스크립트로 직접 연습

1

맞춤 학습 진입

메인 화면에서 맞춤 학습 탭 선택

2

스크립트 직접 입력

"My favorite hobby is reading novels..." 본인 답변 입력

3

반복 오류 패턴 발견

/f/와 /v/ 를 한국어 ㅍ, ㅂ로 발음하는 패턴 누적 탐지

4

통계 확인

프로필 탭에서 일주일 누적 발음 정확도와 약점 음소 확인

07

SECTION

업무 분담과 일정

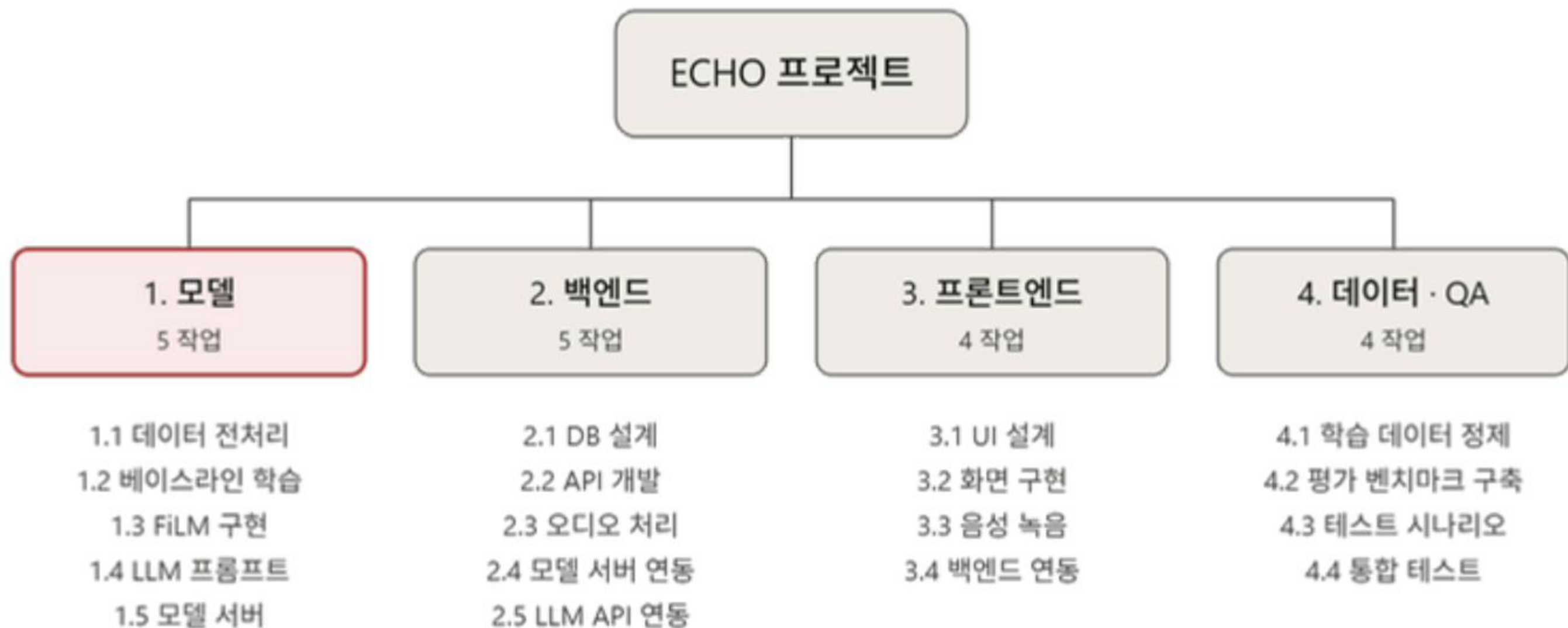
팀 역할 분담

A6 O(1)팀 · 6명 · 4개 파트

모델	백엔드	프론트엔드	데이터 · QA
신영환	김지용	이찬영 최현지	김형준 백지영
FiLM 음소 인식기 학습 아학편 LLM 피드백 생성	API 서버 DB 설계 오디오 처리 모델 서버 연동	UI 구현 학습 화면 피드백 시각화	데이터 전처리 벤치마크 구축 테스트 시나리오

WBS 구조도

4 영역 · 18 작업



간트 차트

종합설계 간트차트

프로젝트 이름	ECHO 프로젝트	팀명	O(1)
프로젝트 관리자	신영환, 김형준, 김지용, 백지영, 이찬영, 최현지	날짜	26년 3월 4일 ~ 26년 6월 10일

[illegible]

종합설계 간트차트

[illegible]

종합설계 간트차트

[illegible]

간트 차트

[illegible]

08

SECTION

달성 목표와 테스트

달성 목표 ① 인식기 FRR 저감

베이스라인

5.83%

현재 측정값

→

목표

4%대

FiLM Canonical Injection

자체 구축 데이터셋에서 평가

L2-ARCTIC + AI-Hub 한국인 영어 발화 통합

화자 단위 분리 · train/test 중복 없음

정답 100개 중 약 4개 미만이 억울하게 오판되는 수준

달성 목표 ② 피드백 평가 벤치마크 구축

발음 오류 케이스

50 개

비교 시스템

4 가지

한국인 평가자

N 명

비교 시스템

IPA 설명

음성학 용어

음소 매핑

스픽식 단순 지시

제로샷 LLM

프롬프트 없이

아학편 기반

한국어 음운 체계

달성 목표 ③ 테스트 시나리오 전부 통과

테스트 케이스

17 개

검증 영역

5 영역

목표 통과율

100 %

검증 영역

발음 인식

정상 · 오발음
묵음

피드백 교정

모음 구분
반복 오류

사용자 적응

추천 · 맞춤
신규 온보딩

엣지 케이스

소음 · 권한
네트워크

학습 데이터

이력 저장
통계 리포트

	항목	입력	기대 출력
TC-01	정상 발음 인식	“apple”을 정확히 발음	정확도 90% 이상, 통과 표시
TC-02	오발음 감지	“apple”을 /æ/ 대신 /a/로 발음	오류 감지 후 교정 피드백 제공
TC-03	묵음 처리	"knight"의 k 발음	묵음 처리 올바른지 확인
TC-04	소음 환경	배경 잡음이 있는 상태로 녹음	분석 성공/실패
TC-05	무음 입력	아무 말 없이 녹음 종료	“다시 시도해보세요” 안내 메시지

	항목	입력	기대 출력
TC-06	모음 구분	/i:/ vs /I/ (sheep vs ship)	i 발음법 안내 및 예시 제공
TC-07	연음 처리	"want to" → "wanna" 허용 여부	구어체 허용 범위 정책대로 처리 발음별 정확도 추이 그래프 표시
TC-08	반복 오류 추적	같은 발음을 n회 연속 틀림	맞춤 LLM 피드백 제공 (조선시대 영어교재 아학편 기반)
TC-09	추천학습 사용자	?지표 OO 이상	정답처리
TC-10	맞춤학습 사용자	?지표 OO 이상	정답처리
TC-11	신규 사용자 온보딩	최초 실행	학습 목록 표시

	항목	입력	기대 출력
TC-12	한국어 억양 간섭	한국어식 영어 발음 입력	한국인 공통 오류 패턴 피드백
TC-13	매우 빠른 발음	정상 속도의 2배로 발음	속도 경고 또는 정상 처리
TC-14	마이크 권한 거부	마이크 접근 차단 상태	권한 요청 안내 화면 표시
TC-15	네트워크 끊김	오프라인 상태에서 분석 요청	오프라인 모드 안내 또는 캐시 처리
TC-16	학습 이력 저장	세션 종료 후 재접속	이전 진도, 오류 이력 유지
TC-17	통계 리포트	7일 사용 후 리포트 열람	출석 달력, 틀린 발음 순위표 제공

09

SECTION

프로젝트 의의

Contributions

01

AI-Hub 한국인 발화
추가 학습

L2-ARCTIC에 더해
AI-Hub 한국인 발화
6만 발화로 학습

02

FiLM Canonical
Injection

정답 음소 정보를
인식기에 주입해
FRR을 직접 낮춘다

03

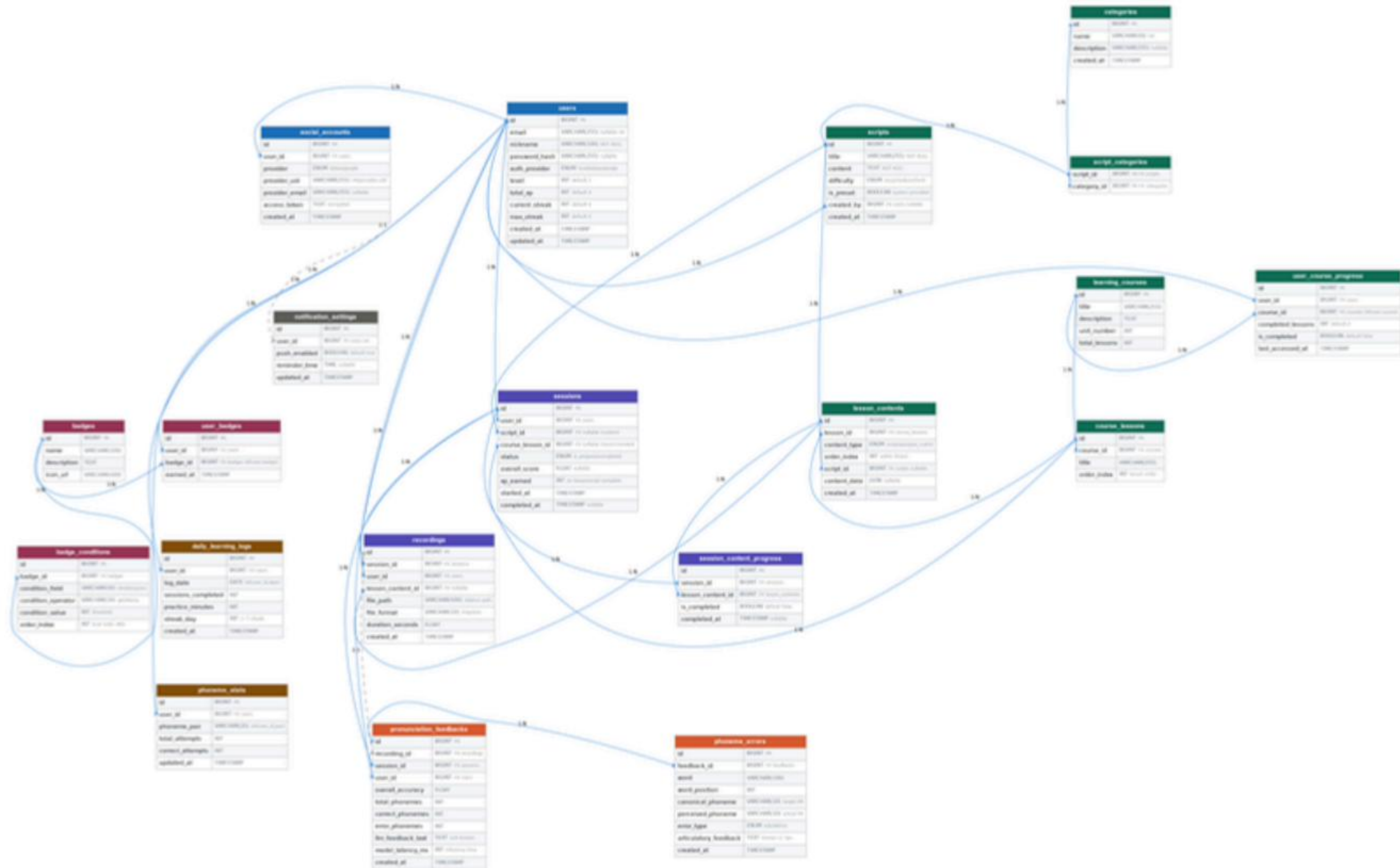
아학편 기반
L1 피드백 생성

한국어 음운 체계로
영어 발음 차이를
설명하는 LLM 피드백

A

APPENDIX

부록



BOO 캐릭터 일러스트 소스



IMG_1600.png



IMG_1601.png



IMG_1602.png



IMG_1603.png



IMG_1604.png



IMG_1605.png



IMG_1606.png



IMG_1607.png



IMG_1608.png



IMG_1609.png



IMG_1610.png



IMG_1611.png



IMG_1612.png



IMG_1613.png



IMG_1614.png



IMG_1615.png



IMG_1616.png



IMG_1617.png



IMG_1618.png

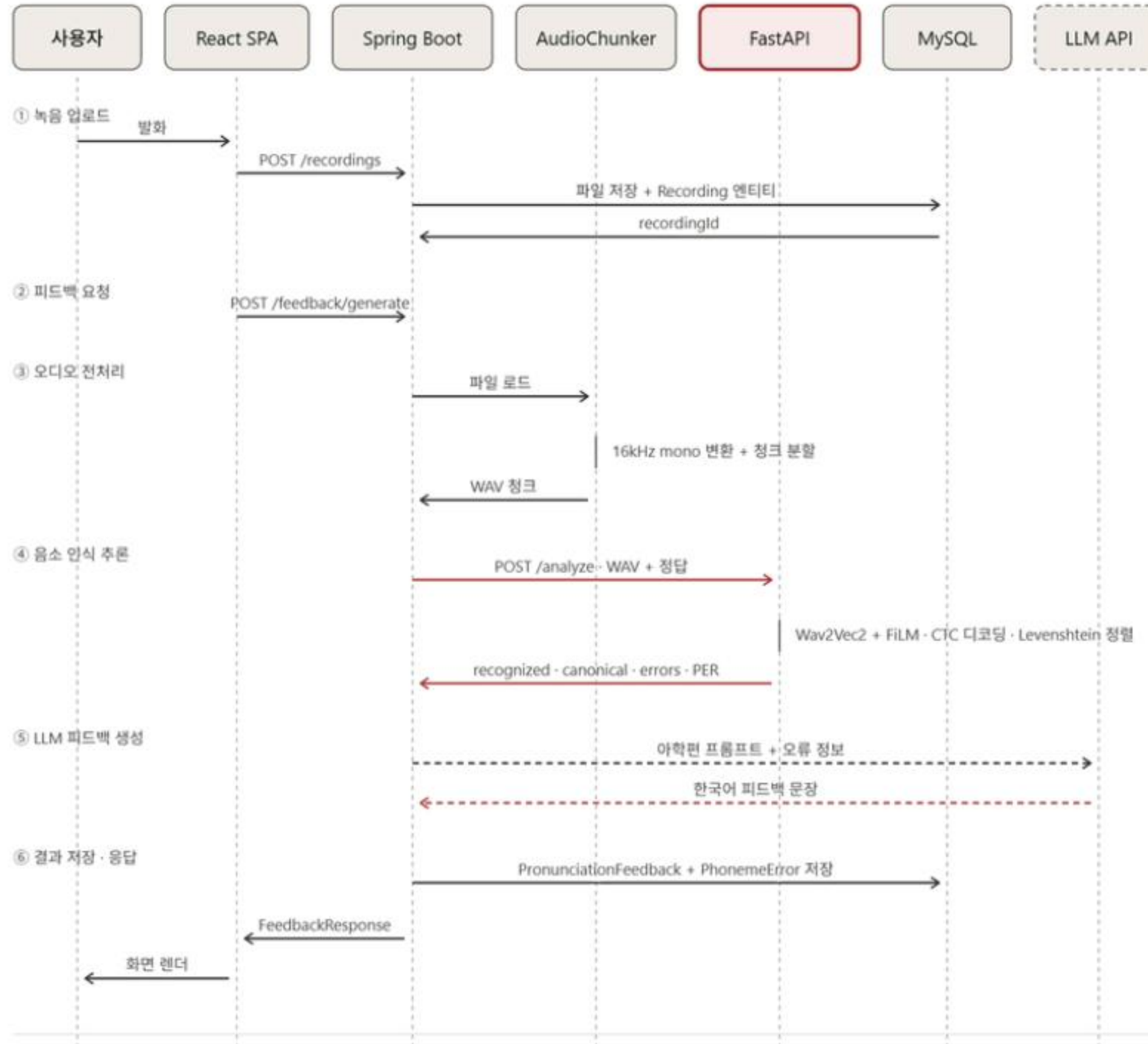


IMG_1619.png



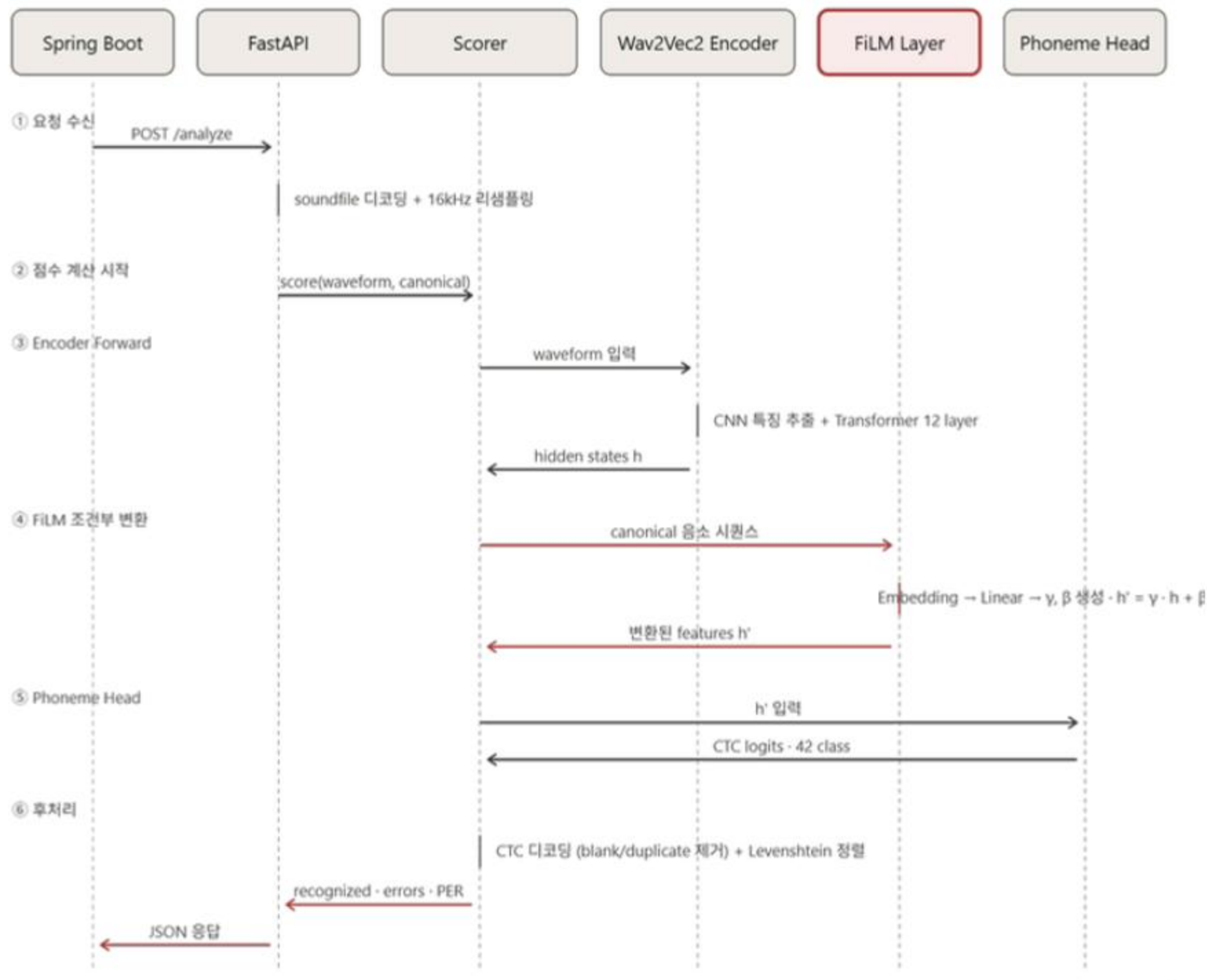
IMG_1622.png

발음 평가 파이프라인 · 상세



실선 = 내부 · 점선 = 외부 API · 빨강 = 모델 서버 · 아학편 관련

모델 서버 내부 추론 · Wav2Vec2 + FiLM



학습 세션 전체 사용자 흐름



용어 정리

분야

CAPT	Computer-Assisted Pronunciation Training · 컴퓨터 보조 발음 학습
MDD	Mispronunciation Detection and Diagnosis · 오발음 검출과 진단
L1 · L2	학습자의 모국어 · 학습 대상 제2언어

평가 지표

PER	Phoneme Error Rate · 음소 인식 오류율
MDD F1	오류 탐지의 정밀도와 재현율을 합친 종합 점수
FRR	False Rejection Rate · 정답 발화를 오류로 잘못 판정한 비율
FAR	False Acceptance Rate · 오발음을 정답으로 놓친 비율

모델

Wav2Vec2	Facebook AI 의 자기지도 학습 기반 음성 표현 모델
CTC	Connectionist Temporal Classification · 정렬 없이 시퀀스를 학습하는 손실 함수
FiLM	Feature-wise Linear Modulation · $h' = \gamma \cdot h + \beta$ 형태의 조건부 변환
canonical · perceived	정답 음소 · 학습자가 실제로 발음한 음소