컴퓨터교육학회 논문지 2025년 제28권 제12호
https://doi.org/10.32431/kace.2025.28.12.009

청각장애인의 발음 교정 지원을 위한 웹 플랫폼 개발*

Development of a Web Platform for Supporting Pronunciation Correction for the Hearing Impaired

이기정† · 윤희욱†† · 오승민†† · 조민서†† · 유수진†††
Keejeong Lee† · Heewook Yoon†† · Seungmin Oh†† · Minseo Cho†† · Sujin Yoo†††

요약

청각은 인간의 언어적 상호작용과 감정 교류를 통해 사회적·정서적 발달에 중요한 역할을 한다. 그러나 청각장애인의 경우 청각 기반 언어 학습 경로에 제약이 존재한다. 본 연구에서는 KoSpeech, Librosa, 거대 언어 모델을 활용하여 청각장애인을 위한 다중 피드백 기반 발음 교정 프로그램을 개발하였다. 제안된 시스템은 발음 정확도, 음조, 발화 속도 및 리듬을 종합적으로 분석하고, 각 항목에 대해 피드백을 제공한다. 초기 테스트에서 처리 지연과 청각장애인 음성 데이터 부족으로 한계가 드러났다. 또한 단일 참여자 대상 정성 평가에 국한되어 있다. 향후에는 처리 성능을 개선하고, 시스템을 고도화하여 정량적, 정성적으로 효과성을 검증할 계획이다.

주제어 청각장애, 발음 교정, 발음 개선, 다중 피드백, 맞춤형 음성 표현 학습

ABSTRACT

Hearing plays an essential role in human social and emotional development by enabling linguistic interaction and emotional exchange. However, individuals with hearing impairment face limitations in auditory-based language learning pathways. This study developed a multi-feedback pronunciation correction program for the hearing impaired using KoSpeech, Librosa, and large language models. The proposed system comprehensively analyzes pronunciation accuracy, pitch, speech rate, and rhythm, and provides feedback for each feature. In the initial test, limitations were observed due to processing delays and insufficient training data for hearing-impaired speech. Furthermore, the evaluation was restricted to a qualitative assessment with a single participant. Future work will focus on improving processing performance, enhancing the system, and validating its effectiveness through both quantitative and qualitative evaluations.

Keywords Hearing Impairment, Pronunciation Correction, Pronunciation Improvement, Multi-modal Feedback, Personalized Speech Expression Learning

†정회원 한성대학교 IT융합공학부
††정회원 한성대학교 컴퓨터공학부
†††종신회원 한성대학교 컴퓨터공학부 조교수 (교신저자)

논문투고 2025년 07월 24일
심사완료 2025년 09월 26일
게재확정 2025년 10월 05일
발행일자 2025년 12월 31일

* 본 연구는 2025년 한성대학교 학술연구비 지원과제임

1. 서론

청각은 인간에게 있어 매우 중요한 감각으로, 언어적 상호작용과 감정 교류의 핵심적인 수단 역할을 해왔다. 우리는 언어를 통해 타인과 소통하고, 주변의 다양한 소리를 들으며 감정과 생각을 공유함으로써 삶을 더욱 의미 있게 영위한다. 이처럼 소리를 인식하는 행위는 단순한 감각 처리 차원을 넘어, 인간의 사회적·정서적 삶에 깊이 관련되어 있다[1].

특히 인간의 언어 습득은 청각 자극을 통한 모방과 반복 학습을 기반[2,3]으로 하며, 듣기 능력은 말하기 능력 향상과 발음 개선에 필수적인 요소로 작용한다[4]. 일반적으로 비장애인은 타인의 발화를 모방하면서 자연스럽게 음운 규칙과 화용적 요소를 체득하지만, 청각장애인은 이러한 경로의 언어 학습이 제한된다[5,6,7]. 청각장애인은 청각 피드백의 결핍으로 인해 발화 조절 능력이 제한되며, 이는 음성의 주파수, 강도, 진폭 변동률 등에서 비장애인과 비교하여 뚜렷한 차이를 보인다[8]. 특히, 선천성 청각장애 아동은 제한된 청각 경험과 우연 학습의 부재로 인해 언어 발달이 지연될 수 있으며[9], 후천적 청력 손실이 발생한 성인은 음정, 억양, 음량 등의 음성 매개변수에서 왜곡이 발생할 수 있다[10].

청각장애인은 주로 수화나 문자 정보, 입 모양과 같은 시각적 단서를 통해 언어를 학습하지만, 이러한 방식만으로는 음정, 운율 등의 음성적 특성을 스스로 인식하거나 교정하기 어렵다[11,12,13]. 그러나 청각장애인은 발화 시 리듬의 부재, 부적절한 억양과 같은 운율 특징을 가지고 있다.

뿐만 아니라 Bennett[14], Fletcher 외[15], Fletcher 외[16], 신은영 외[17], Park과 Seol[18]의 연구에 따르면 청각장애인들은 전통적으로 발음 교정에 개별 자음과 모음(분절음)을 통한 정확한 조음에 초점을 맞춘 학습을 받아왔다. 그러나 많은 청각장애 학습자들은 일정 수준의 명료도를 확보한 후에 '명료도 고원(Intelligibility Plateau)'현상에 부딪힌다. 명료도가 더 이상 유의하게 상승하지 않고[19, 20], 단어의 발음은 비교적 명확한 편이나, 전체적인 말의 흐름이 부자연스럽게 들리는 문제이다[21]. 이러한 부자연스러움은 억양, 강세, 리듬, 말의 속도와 같은 초분절(prosody)적 요소들의 통제 실패에서 기인한다[22]. 예를 들어 "밥 먹었어"라는 동일한 문장도 끝을 올려 발음하면 질문이 되고, 끝을 내리면 평서문이 된다. 이처럼 초분절적 요소는 문장의 의미를 명확하게 하고, 화자의 감정과 의도를 전달하는 핵심 요소이기 때문[23]이다. 효과적인 의사소통은 의미 전달을 넘어서 편안한 상호작용을 통해 이루어진다는 점에서 초분절적 요소의 개선은 발음 교정의 부차적 목표가 아닌 핵심 과제라 할 수 있으며, 초분절 훈련이 명료도 고원을 뛰어넘도록 지원하는 역할이 가능하다[24]. 하지만 청각장애 화자는 발화 연습 과정에서 긴장·부담이 높고 반복 지속이 쉽지 않기 때문에, 학습 맥락 자체를 더 편안하고 자발적 연습이 가능하도록 설계할 필요가 있다. 이때 놀이·게임의 도입은 환경을 부드럽게 하고 참여를 끌어올리는 실천적 해법이 될 수 있다. 실제로 학습에 놀이나 게임을 적용하면 편안하고 자유로운 분위기에서 학습이 제공되어 긍정적인 영향을 미칠 수 있으며[25-27], 이는 유아나 청소년기의 학습자 뿐 아니라 성인 학습자에게도 유효하다[28-31]. Tye-Murray 외[32]는 청각장애 아동 99명을 대상으로 16시간동안 게이미피케이션된 훈련(청각, 시청각) 후 효과를 입증하였고, Spehar 외[33]는 동일 코호트 후속 연구로써 16시간 게이미피케이션 훈련을 하고 난 뒤 4-6주 후에도 효과가 유효하게 지속되고 있음을 보였다. 이외에도 Pimperton 외[34], Parmar 외[35] 역시 청각장애인을 대상으로 한 게이미피케이션의 효과를 보고했으며, Saeedi 외[36]는 아동의 언어 치료용 디지털 게임이 다양한 임상에서 학습동기와 치료 지속 의지를 긍정적으로 증진한다고 보고하였다.

한편, 일반 언어 학습 플랫폼을 이용하는 것도 청각장애인이 언어를 학습하는 대안이 될 수 있지만, 이들 은 주로 오디오 중심이어서, 자신의 음성을 직접 청취하기 어려운 청각장애인의 요구를 충분히 반영하지 못하는 한계[37]가 있다. 이를 보완하기 위해서는 청각이라는 손상된 피드백 회로를 기능적으로 재구성하는데 필요한 핵심 데이터로 시각 피드백을 제공할 필요가 있다[38,39]. 학습자는 자신의 발음에 대한 시각적 출력값과 목표하는 시각적 패턴(예: 목표 억양 곡선)을 비교하여 시각적 오류를 인지하고 이를 바탕으로 수정할 수 있으며[40,41], 이러한 시각 기반 훈련이 축적되면 초분절적 요소의 생성이 점차 자동화되어 학습자는 더 유창하고 여유롭게 의사소통에 집중[42]할 수 있다. 즉, 시각적 단서와 청취 기반 언어 학습의 약점을 상호 보완하는 청각장애인 맞춤형 언어 학습 프로그램의 개발이 요구된다.

이러한 배경 속에서 청각장애인의 언어 학습과 발음 교정을 지원하기 위한 기술적 시도도 다수 이루어졌다. 청각장애인을 위한 발음 교정 모바일 애플리케이션이 개발되었고[43], 립 리딩(Lip-reading) 알고리즘을 적용한 스마트 미러 시스템이 제안되었으며[44], 3D 얼굴 모델과 음성 인식 엔진을 활용한 발음 학습 시스템이 설계되었다[45]. 그러나, 기존 시스템들은 입 모양 중심의 피드백에 의존하거나, 비장애인의 발화를 기반으로 구축된 상용 음성 인식 엔진을 이용하는 경향이 있다. 이러한 구조는 학습자의 발음 오류를 파악하기 어렵게 만들고, 비표준적인 발화에 대한 인식률을 낮추어 청각장애인에게 실질적인 도움을 제공한다고 보기 어렵다.

이에 본 논문은 청각장애인을 위한 맞춤형 언어 학습 플랫폼 담설(談說)을 제안한다. 플랫폼은 웹 브라우저 기반으로 설계되어 단어와 문장, 한국어 문법 및 상황별 학습 단위로 세분화된 언어 학습 콘텐츠와 구화 연습을 위한 게임형 콘텐츠를 제공한다. 담설은 자체 학습한 자동 음성 인식(Automatic Speech Recognition, ASR) 모델을 통해 문맥 보정 없이 음향 데이터를 텍스트로 변환할 수 있으며, 라이브러리 Librosa[46]를 이용해 발화 속도, 음정, 운율 등 다양한 음성 특징을 추출한다. 이후, 대규모 언어 모델(Large Language Model, LLM)을 활용한 다차원 발음 평가 및 개

선 방향과 음성 성분에 대한 시각적 피드백을 제시한다.

2. 관련 연구

정보 기술의 발달과 함께 언어장애인을 위한 발음 교정 프로그램에 대한 사회적 요구와 기술적 개발이 지속적으로 증가해왔다. 기존 연구들을 바탕으로, 본 연구에서는 청각장애인을 대상으로 한 발음 교정 시스템을 피드백 방식에 따라 두 가지 유형으로 구분하였다. 첫째는 사용자의 입 모양을 분석하여 시각적 피드백을 제공하는 입 모양 기반 피드백 시스템이고, 둘째는 상용 음성 인식 엔진을 활용하여 사용자의 음성을 분석하고 발음 정확도를 평가하는 방식이다.

2.1 입 모양 피드백 중심의 발음 교정 프로그램

입 모양 피드백 중심의 발음 교정 프로그램은 사용자의 입술 움직임을 기반으로 시각적 피드백을 제공하는 구조로 설계된다 [43,44,47]. Zhu[48]에 따르면 조음 음성학 연구에서 입술 모양과 혀의 위치는 언어학자들의 관심분야이며, 3D 입술 형태만으로도 70% 이상 정확도로 모음을 식별할 수 있는 것으로 나타났다. Sato와 Bao[49]는 대화를 통한 의사소통 능력 습득을 위해서는 정확한 발성을 위한 립 리딩(lip-reading)과 모방 연습이 필수[49]라고 언급하였다. Stavness 외[50]에 의하면 입 모양에 따라 발음이 가능한 소리의 종류에 제약 혹은 가능성을 주는 요소에 대한 객관적 탐구 결과를 보고했다. 즉, 입모양은 청각 정보에 의존하지 않고도 발음을 교정할 수 있다는 점에서 청각장애인을 위한 보조 수단으로 적합하며, 기존 연구들은 사용자는 화면이나 실시간 영상을 통해 자신의 입술 움직임을 관찰하고 이를 정적인 이미지나 영상, 또는 단순한 모션 비교를 통해 비교함으로써 사용자 스스로 자신의 발음 상태를 판단하도록 유도하는 방식을 주로 사용한다.

정하윤 외[44]은 스마트 미러 기반의 발음 교정 시스템을 개발하여 립 리딩 알고리즘과 음성 인식 기술을 결합한 실시간 입 모양 비교 기능을 구현하였다. 이미애[47]는 청각장애인의 구화 교육을 위한 립 리딩 알고리즘을 개발하여 광류(Optical Flow)와 조음 발성 특성을 결합한 단음절 분류 시스템을 제안하였다. 이영주 외[43]는 모바일 기기에서 동작하는 애플리케이션을 통해 사용자의 입 모양을 영상으로 인식하고 발음 학습 진도 및 정확도를 피드백하는 기능을 구현함으로써 이동성과 접근성을 고려한 시각 기반 발음 학습 시스템의 가능성을 제시하였다.

기존의 입 모양 피드백 중심 발음 교정 프로그램은 입술의 움직임에 의존하는 구조로 인해 실제 발화 과정에서 나타나는 미세한 차이 또는 음성의 질적 특징을 반영하기 어렵다는 한계를 지닌다. 또한, 유사한 입 모양을 가진 자음 간의 구분이나 성대 및 혀와 같은 비가시적 발음 기관의 움직임에 기반한 음성적 특성은 시각적 피드백만으로 인지하기 어렵다. 이에 따라 사용자는 자신의 발음 오류를 정확히 파악하거나 자율적으로 수정하기 어려우며, 이는 학습자의 발음 교정에 구조적인 제약으로 작용할 수 있다.

담설은 기존 입 모양 피드백 중심 발음 교정 시스템의 한계를 보완하기 위해 LLM을 기반으로 한 다중 피드백 구조로 설계되었다. 시스템은 오디오 신호 처리 라이브러리 Librosa를 활용하여 발화 속도, 피치, 리듬 등의 음향 정보를 정량화하고, 이를 LLM 기반 발음, 음성 성분 평가 모듈과 연계하여 오류 유형, 개선 방안 등을 평가한다. 더불어 반복 학습을 통해 누적된 발화 데이터를 기반으로 사용자의 발음 특성을 분석하고, 개인 맞춤형 피드백을 제공함으로써 자기 주도적 학습이 가능한 발음 교정 환경을 구현하였다.

2.2 상용 음성 인식 엔진 기반 발음 교정 프로그램

상용 음성 인식 엔진 기반 발음 교정 프로그램은 Google Speech-to-Text(STT)[51], Microsoft Azure Speech Service[52], OpenAI Whisper[53] 등과 같은 상용 STT 기술을 활용한다. 이러한 시스템은 사용자의 음성을 텍스트 형태로 변환하고, 이를 정답 텍스트와 비교하여 발음의 정확성을 평가하는 방식으로 동작한다.

강채린 외[54]는 언어 발달 지연 아동을 위한 웹 기반 발음 교정 프로그램 '시나브로(SINABULO)'를 개발하여 Google Speech-to-Text API를 이용해 사용자의 발음 정보에 대한 텍스트 및 시각 자료를 제공하는 기능을 구현하였다. Moxon[55]은 Microsoft Azure Speech Services의 음성 인식(STT), 음성 합성(TTS) 및 자동 발음 평가 기능을 통합한 웹 기반 플랫폼 'ALL-Talk'을 개발하여 발음의 음절 및 음소 단위의 정확도, 유창성, 완성도 등의 평가 결과를 시각적으로 제공하는 자율 발음 훈련 환경을 구현하였다.

상용 음성 인식 엔진은 청각 비장애인의 정규 발화를 기반으로 구축되어, 음소 누락, 억양 불균형 등과 같은 청각장애인의 비표준 발화에 대해서는 낮은 인식 정확도를 보인다 [56]. 이에 따라 기존의 상용 음성 인식 엔진 기반 발음 교정 프로그램은 청각장애인의 실제 발화 내용을 다르게 해석하거나 정확도 평가에 왜곡된 결과를 초래할 수 있어 발음 오류 분석의 신뢰성이 낮고 사용자의 발음 개선 효과를 저해할 수 있다.

본 논문에서는 기존의 상용 음성 인식 기반 발음 교정 프로그램의 한계를 극복하기 위해 한국어 특화 오픈소스 ASR 프레임워크인 KoSpeech[57] 기반 모델을 개발하였다. KoSpeech는 오픈소스 한국어 ASR 프레임워크로, 별도의 언어 모델 결합 없이 작동하며 문맥 기반 보정 없이 발화 내용을 텍스트로 추출할 수 있다. Transformer[58], LAS(Listen Attend and Spell)[59], Deep Speech 2[60] 등 다양한 모델 구조를 지원하며, 한국어 음성 인식 환경에 적합한 학습 파이프라인을 제공한다. 모델은 Deep Speech 2 기반 모델을 선택하고 AI-Hub의 한국어 음성 데이터셋을 활용하여 학습이 수행[61]되었다. 약 1,000시간 분량의 한국어 음성 데이터를 10 에포크(epoch) 학습을 진행하였으며, 학습 완료 후 측정한 WER(Word Error Rate)는 0.27 수준

으로 나타났다. 자체 실험 결과, 조용한 환경과 고품질 마이크를 사용한 조건에서는 비교적 안정적으로 음성을 인식하는 성능을 보였다.

KoSpeech 프레임워크와 학습된 모델은 Flask 서버 내 통합되었으며, 실제 서비스 환경에서는 Python[62]의 subprocess 모듈을 통해 ASR 추론(inference)을 수행한다. CPU 환경에서는 1개 문장 처리에 약 5~10초의 시간이 소요되며, 실시간 처리에는 다소 제약이 있으나 학습자 피드백을 위한 비동기 처리에는 충분한 응답 성능을 제공한다. 또한 비표준적인 발음에 대하여 높은 인식 정확도를 보였다.

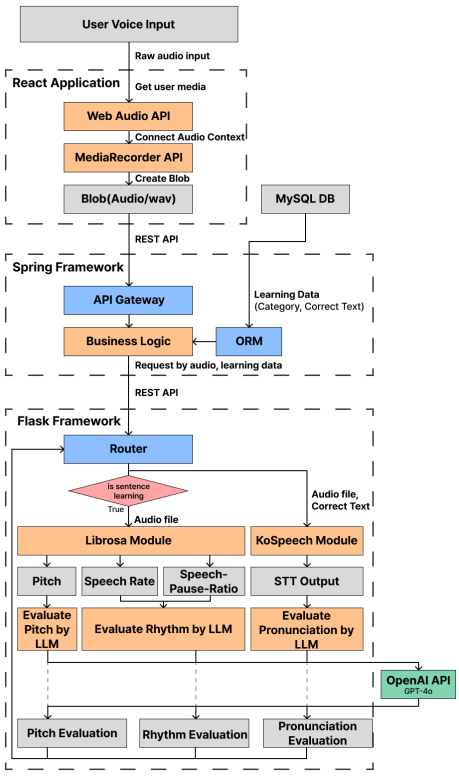

Figure 1. System Structure

3. 제안 시스템

3.1 시스템 구성

담설의 전체 시스템 구조는 Fig. 1과 같다. 사용자는 React 기반 프론트엔드 애플리케이션을 통해 서비스를 이용할 수 있으며, 이 계층에서 인터페이스 구성 및 사용자 입력 처리가 이루어진다. 입력된 요청은 Spring Boot[63] 기반의 백엔드 서버로 전달되며, 서버는 MySQL 데이터베이스를 이용해 데이터 흐름을 제어하고 단어 및 문장, 학습 정보 등과 같은 자료의 관리를 수행한다. 음성 평가 기능은 Flask 프레임워크[64]를 기반으로 별도의 AI 모듈에서 수행된다. 해당 모듈은 음성 인식 모델인 KoSpeech[57], 오디오 분석 라이브러리 Librosa, 자연어 기반 피드백을 위한 OpenAI GPT-4o[65] 모델을 핵심 기술로 활용하며, Flask 서버는 구성 요소 간의 연계 처리를 담당한다. KoSpeech 모델은 AI-Hub의 한국어 음성 데이터셋을 활용하여 사전 학습되었다. 담설의 개발 환경은 Table 1과 같다.

Table 1. Development Environment

Component	Technology Stack	Version
Frontend	React.js	19.0.0
Backend	Spring Boot	3.4.2
Database	MySQL	8.4.4
AI Server	Flask	3.1.0
ASR Engine	KoSpeech	1.3
LLM Module	OpenAI GPT-4o	gpt-4o-2024-11-20
Dataset	AI-Hub Korean Speech Data	1.0

Table 2. Extracted Voice Features for Evaluation

Feature Type	Attributes	Description
Syllables Per Second	duration_sec	Total duration of the utterance (in seconds)
	syllable_count	Number of syllables in utterance
	speech_rate	Syllables spoken per second
	rate_label	Categorized speech rate: Fast / Normal / Slow
Pitch	time	Time stamp of each pitch measurement (in seconds)
	pitch	Fundamental frequency (Hz) at the given time
	voiced	Whether the segment is voice(1) or unvoiced(0)
Speech-Pause-Ratio	time	Time stamp of each measurement (in seconds)
	speech_pause_ratio	Ratio of speech duration to puase duration (0.0-1.0) A lower value indicates more frequent pauses; a higher value indicates fluent and uninterrupted speech.

3.2 단일 음성 평가

시스템의 음성 평가는 기본적으로 '단어 학습' 및 '문장·문법 학습'의 두 가지 로직으로 구성된다. 두 학습 유형 모두 사용자의 발음 정확도를 평가하기 위해 KoSpeech 기반 ASR 모델과 LLM을 함께 활용한다.

단어 학습에서는 사용자가 정답 텍스트를 보고 해당 단어를 발음하면 단어 텍스트와 사용자의 음성 데이터가 백엔드 서버를 거쳐 Flask 기반 AI 서버로 전달된다. Flask 서버에서는 KoSpeech 모델을 통해 입력된 음성을 텍스트로 변환하며, LLM은 정답 텍스트, 사용자의 발화 텍스트, 틀린 음소, 평가 프롬프트를 기반으로 정확도를 평가하고 피드백을

생성한다. 해당 피드백은 발음의 주요 오류 지적, 발음 교정 조언을 포함한다.

문장·문법 학습에서는 위와 동일한 발음 정확도 평가 외에도 발화의 억양과 운율 정보를 분석하여 보다 정밀한 피드백을 제공한다. 이를 위해 Librosa 라이브러리를 활용하여 음성 파일에서 초당 음절 수(Syllables Per Second), 음조(Pitch), 발화 중단 비율(Speech Pause Ratio: 시간 구간별 음성이 차지하는 비율)을 추출하며, 추출된 데이터의 형태는 Table 2에 제시되어 있다.

LLM은 Table 2의 음성 성분과 평가용 프롬프트를 함께 입력받아 사용자의 발화가 자연스러운 억양과 속도로 이루어졌는지 평가한다. LLM 모듈은 각 음성 성분에 특화된 프롬프트를 적용하여 독립적으로 구성되며, 음성 성분별 특성을 정밀하게 분석할 수 있도록 설계되었다. 각 음성 성분을 평가하는 LLM은 일관성을 확보하기 위해 temperature값을 0.3의 낮은 값으로 설정하였다. 0.1과 같이 낮은 값은 매우 일반적이고 보편적인 평가가 나타나고 0.5 이상에서는 같은 파일이더라도 평가가 상이하게 나왔지만, 0.3으로 설정했을 때는 같은 음성 파일로 반복 요청하였을 때에도 적절한 학습 조언을 생성하면서 평가의 일관성을 유지함을 확인하였다. 학습자 피드백 수준에서는 충분히 납득 가능한 수준으로 확인하였다. 이와 같은 성분별 프롬프트 최적화 방식은 LLM의 평가 정확도를 높이고, 사용자에게 구체적인 피드백을 제공하는 데 효과적이다. Fig. 2는 위 일련의 과정을 시각적으로 나타낸 것이다.

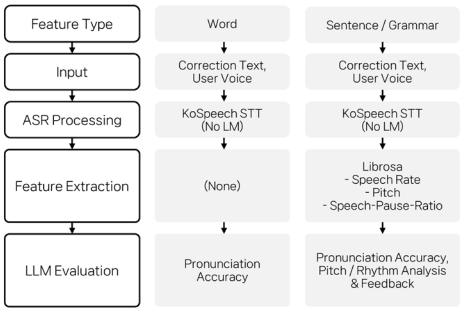

Figure 2. Evaluation Process

3.2.1 음조 (Pitch)

음조는 사용자 발화의 음조가 자연스러운지 평가하기 위한 척도로, 시간에 따른 기본 주파수(Fundamental Frequency)의 변화 양상을 분석한다. 분석에 활용되는 데이터는 일정 간격으로 분할된 시간 배열(time), 각 시점에 대응되는 주파수 값(pitch), 그리고 해당 시점의 음성의 유무 여부를 나타내는 플래그(voiced)로 구성된다. 데이터의 예시는 Table 3과 같다.

Table 3. Example of Extracted Pitch Features from User Utterance

Time (sec)	Pitch (Hz)	Voiced
0.10	127	True
0.15	130	True
0.20	NaN	False

3.2.2 초당 음절 수 (Syllables Per Second)

사용자의 발화 속도를 평가하기 위한 정량적 지표로, 전체 발화 시간 대비 발화된 음절 수를 바탕으로 계산된다. 계산에 사용되는 속성은 녹음된 총 시간(duration_sec), 음절 개수(syllabled_count)와 이를 기반으로 산출된 초당 발화 음절 수(speech_rate)이다. 산출된 speech_rate는 일반적인 한국어 화자의 평균 발화 속도(초당 약 3~6음절[66,67])와 비교하여 느림, 적절, 빠름 중 1개의 평가 라벨(rate_label)로 분류된다. 데이터의 예시는 Table 4와 같다.

Table 4. Example of Extracted Syllables Per Second from User Utterance

Duration (sec)	Syllable Count	Speech Rate (syll/sec)	Rate Label
3.32	13	3.91	Normal
2.10	14	6.67	Fast
4.70	11	2.34	Slow

3.2.3 발화 중단 비율 (Speech-Pause-Ratio)

발화 중단 비율이란 일정 시간 간격 내에서 사용자의 실제 발화가 얼마나 연속적으로 이루어졌는지를 0.0에서 1.0 사이의 값으로 정량화한 지표이다. 이는 초당 음절 수와 함께 발화 속도 평가에 활용되며, 본 연구에서는 이를 위해 직접 정의한 계산식을 사용하였다. 값이 1.0에 가까울수록 해당 구간에서 발화가 끊김이 없이 지속되었음을 의미하고, 0.0에 가까울수록 침묵 구간이 많았음을 나타낸다. 이 값은 다음과 같은 계산식을 통해 도출되며, 적용 예시를 Table 5에 제시하였다.

$$SPR_t = \frac{\sum_{i=1}^{N_t} (\min(e_i, t+\Delta t) - \max(s_i, t))}{\Delta t}$$

t : 현재 분석 구간의 시작 시점
Δt : 분석 프레임의 길이
s_i : 음성 감지 구간의 시작 시점
e_i : 음성 감지 구간의 끝 시점
N_t : 현재 분석 구간 $[t, t+\Delta t]$과 겹치는 음성 구간의 개수

Table 5. Example of Extracted Speech-Pause-Ratio from User Utterance

Time (sec)	Speech-Pause-Ratio
0.00	0.71
0.05	0.93
0.10	0.45

3.3 학습 종합 평가

학습자가 담설에 접속하면 단어 학습, 문장 학습, 문법 연습을 수행할 수 있다. 각 학습 단위는 여러 가지의 발화 과제로 구성된다. 예를 들어, '문장 학습' 카테고리에 포함된 '병원에서 대화' 학습 기능은 무작위로 선정된 세 가지 문장에 대한 발화 과제를 제공한다. 학습자 발화 시, 음성을 녹음하고 해당 음성은 담설의 모델에 의해 평가된다. 평가 결과는 배열 형태로 저장되며 단어 학습의 경우 발음 피드백 필드를, 문장 학습의 경우 발음, 음정, 리듬 피드백 필드를 포함한다.

3.3.1 발음 분석 알고리즘

발음 분석은 Kospeech 모듈로 인식된 발화 텍스트를 정답 텍스트와 비교하는 과정으로 시작한다. LLM에 평가를 요청을 전달할 때 정답 텍스트와 인식 결과를 유저 컨텍스트에 포함한다. 프롬프트는 두 텍스트를 발음 규칙에 맞게 변환하여 표기 차이로 인한 오판을 방지하고, 발음이 사실상 동일한 음소(예: ㅐ/ㅔ)는 정답으로 간주하도록 지시하는 내용이 포함된다. 정확도 점수(0~100점), 오류 유형, 개선 포인트를 출력하도록 구성한다. 이 프롬프트를 통해 음성 인식 결과가 표기-발음 차이를 반영하지 못하더라도 언어 모델 단계에서 표기를 발음으로 변환하는 과정을 거쳐 오판 가능성을 최소화하고 실질적인 개선 방법을 제시하도록 한다.

3.3.2 음성 성분 분석

음조(Pitch)는 Librosa의 pyin 알고리즘을 이용해 기초주파수를 산출한다. 탐색 범위는 fmin=80Hz, fmax=400Hz로 설정하여 성인 화자의 일반적인 음역대를 포함하였다. 분석은 0.03초 단위의 프레임 간격으로 수행하여 각 프레임에 기록된 Pitch 값을 출력한다.

LLM 기반 평가 단계에서는 문장 유형별 피치 패턴 규칙 준수(평서문: 변화 적음, 의문문: 말미 상승, 감탄: 큰 변화), 일관성과 문맥 적절성 반영되었는지 평가하도록 구성하였다. 유저 컨텍스트로는 이전 단계에서 추출한 pitch 요약치와 정답 텍스트가 전달되며 결과로Pitch 평가(나쁨|보통|좋음), 점수(n/5점), 평가한 이유를 출력한다.

발화 속도(Speech Rate)는 Kospeech 인식 결과에서 산출된 음절 개수를 총 발화 시간으로 나누어 계산하였다. 발음한 총 음절 수가 20개이고 발화 시간이 5초인 경우 20 (Syllables) / 5 (Seconds) = 4 Syllables Per Seconds로 계산한다.

발화 중단 비율(Speech-Pause-Ratio)은 librosa.effects.split 함수를 이용하여 무성 구간을 검출한다. 임계값은 top_db=20으로 설정하였으며, 전체 음성 길이 대비 실제 발화가 이루어진 구간의 비율을 계산하여 산출한다.

발화 속도와 발화 중단 비율은 LLM에서 두 요소를 동시에 평가한다. 이전 단계에서 제공된 수치(Speech Rate, Speech-Pause-Ratio)에 근거해 논리적인 판정, 정상 범위(4-6음절/초), 문장 중간에 2초 이상의 정적은 긴 정적으로 판단, 녹음 구간의 처음과 마지막의 정적은 평가 제외(사용자 조작 시간 반영)가 프롬프트에 포함된다. 결과로는 리듬 점수(나쁨|보통|좋음), 점수(n/5점), 평가 이유를 출력한다.

3.3.3 종합 평가

위 단계는 하나의 학습을 이루는 1개 문제 단위로 각각 이루어지고, 문제를 모두 학습한 뒤에는 각 문제별로 이루어진 평가를 배열로 나열한 것을 유저 컨텍스트로 전달하여 최종적인 평가를 생성한다. 이 때 프롬프트에는 각 문제별로 이루어진 모든 평가가 유저 컨텍스트에 포함되어 들어가며, 시스템 지시로는 사용자 발음 평가 내용을 기반으로 특정 자모나 발음 규칙 등에서 나타나는 공통적인 오류 분석, 리듬/피치에서 나타나는 공통적 문제점 분석, 학습자가 수행해야 할 학습 조언 중점 제시가 포함된다. 이 과정의 수행 결과로는 각 파트 별 학습 조언이 1-2줄 단위로 출력되며, 그 형태는 { 발음: 'ㄱ'발음을 'ㅂ'과 혼동하고 있어요. 'ㄱ'과 'ㅂ'의 발음 원리에 대해 학습해보세요.\n리듬: 발음 속도가 전반적으로 느리고, 중간에 발화가 끊기는 시간이 길어요. 발음 속도를 조금 더 빠르게 연습해보세요.\n음정: 전반적으로 음정이 안정적이지만 문장이 끝날 때 점점 음정이 내려가는 경향이 있어요. 문장 끝까지 음정을 안정적으로 유지하는 연습이 필요해요. }와 같다. Fig. 3은 이 절차를 플로차트로 나타낸 것이다.

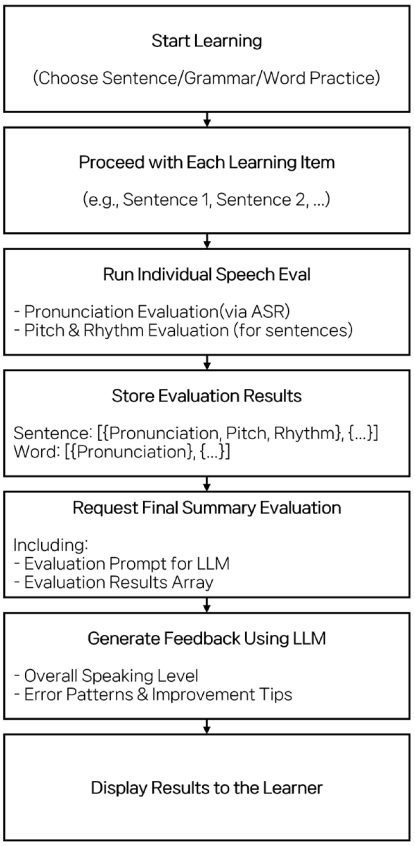

Figure 3. Workflow of Learning Evaluation and LLM-Based Summary Feedback

4. 시스템 기능

본 연구에서 제안하는 담설은 단어 및 문장 학습 모듈, 게이미피케이션 모듈, 학습 결과 리포트의 세 가지 구성 요소로 이루어진다. 각 모듈은 음성 입력을 시작으로, AI 분석과 피드백 제공, 그리고 자가 교정으로 이어지는 순환적 학습 구조를 통해 청각장애 학습자의 발화 명료도 및 반복 학습 효과를 높인다. Fig. 4는 담설 접속 후 로그인을 하게 되면 학습자가 보게 되는 첫 화면이다.

Figure 4. Initial Screen

4.1 단어 및 문장 학습 모듈

'단어 학습' 모듈은 음운, 음운 변동, 사잇소리 현상 카테고리 아래 13가지의 서브카테고리를 기반으로 구성되며, 발음 교육의 기초를 담당한다. 먼저, 화면의 단어를 사용자가 발화하면 음성 데이터는 텍스트 형태로 변환된다. 이후, 이를 해당 단어의 실제 발음과 비교하여 사용자의 발음 정확도를 산출한다. 인식된 텍스트, 정확도, 오류 음소는 시각적으로 강조되며, 시스템은 이를 기반으로 추천 학습 자·모음을 함께 제시하여 학습자의 취약 지점을 보완할 수 있도록 한다. Fig 5는 담설의 단어 학습 화면을 보여준다.

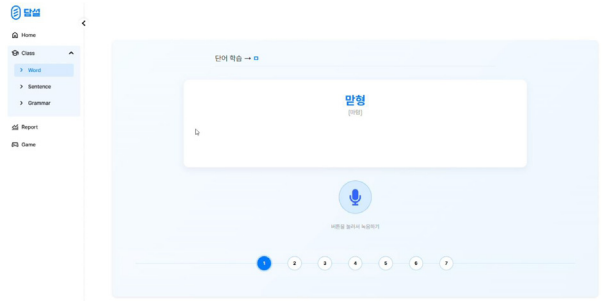

Figure 5. Word Training Screen

'문장 학습' 모듈은 특별한 상황, 비즈니스, 문법 연습을 주제로 한 19가지 카테고리로 구성되며, 학습자는 상황에 맞는 표현을 선택하여 문장 학습을 수행할 수 있다. 마찬가지로, 발화된 문장은 음성 인식 엔진을 통해 텍스트로 변환된 후 Librosa 및 GPT 기반 발음 평가 모듈에 전달되어 문장 단위 발화에 대한 피치 변화 및 일관성, 발음 정확도, 리듬, 발화 속도 등의 음성 특징 요소들이 종합적으로 평가된다. Fig 6은 문장 학습 수행 직후 보게 되는 피드백 화면이다.

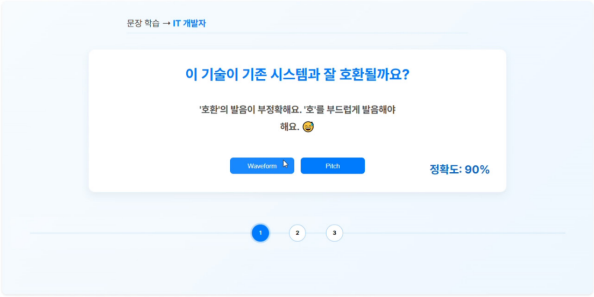

Figure 6. Feedback Screen during sentence training

Fig. 7은 문장 학습 피드백 화면에서 각각 Waveform, Pitch 버튼을 누르면 보이는 화면이다.

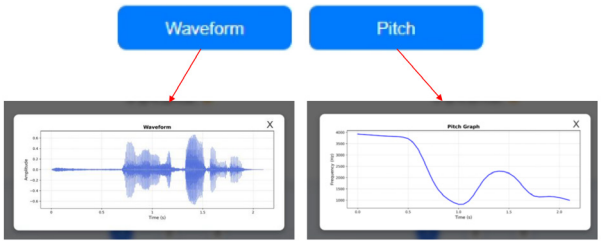

Figure 7. Waveform, Pitch Screen

Waveform(파형)은 학습자에게 소리의 세기와 길이를 보여준다. 소리가 클수록 파형의 높이가 커지고, 작을수록 낮아진다. 이를 통해 학습자 본인이 발화를 약하게 했는지 강하게 했는지 파악할 수 있다. Pitch(피치)는 소리의 높낮이와 억양을 보여준다. 목소리가 올라갈 때(의문문)는 곡선이 위로, 내려갈 때(평서문)는 곡선이 아래로 움직인다. 학습에 주어진 문장이 의문문이었는지 평서문이었는지에 따라 Pitch 그래프를 대조해보면 자연스러운 발화에 도움을 받을 수 있다.

Figure 8. Visualization of sentence-level pronunciation evaluation results including accuracy, pitch, and rhythm

평가 결과는 Fig. 8과 같이 정확도, 억양(Pitch), 리듬을 포함한 시각적 피드백 형태로 제공되며, 학습자는 이를 기반으로 자신의 발화 습관을 분석 및 개선할 수 있다.

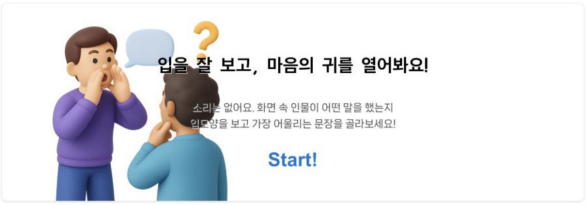

Figure 9. Lip-Sync Gamified Home

4.2 게이미피케이션 모듈

청각장애인 사용자의 구화(口話) 말읽기 능력을 강화하기 위해, 담설은 립싱크(lip-sync) 기반 게이미피케이션 모듈을 포함하였다. Fig 9는 담설의 게이미피케이션 도입 화면을 보여준다. 저작권에 문제가 없는 영상들에 FaceFusion[68]을 적용하여 게이미피케이션에서 제공하고자 하는 문장과 영상 속 인물의 입 모양을 맞추어 제공하였다. Fig. 10과 같이, 학습자는 음성이 제거된 영상 속 인물의 입 모양을 관찰한 후, 제시된 보기 중 가장 적절한 문장을 선택한다. 이는 학습자의 시각적 단서에 기반한 발화 추론 능력을 증진하며, 실제 발화 상황에 대한 인지력을 높인다. Fig. 11과 같이 게임 종료 후에는 정답률, 오답 통계가 제공되며, 학습자는 자신의 인지 방식과 판단 근거를 되돌아볼 수 있다.

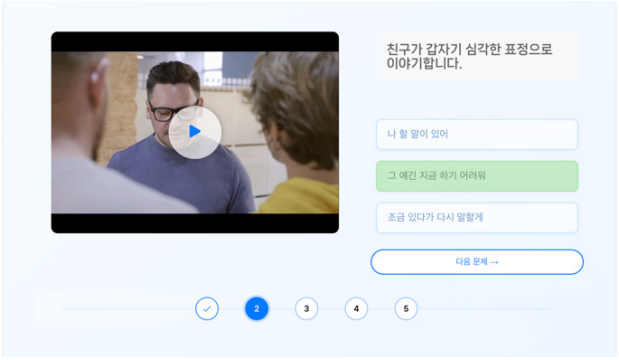

Figure 10. Interface of the lip-sync based visual speech recognition game

4.3 학습 결과 리포트

담설은 학습자의 누적 학습 결과를 기반으로 '학습 결과 리포트' 페이지를 제공한다. 시스템은 발음 정확도, 피치, 리듬에 대한 평가 항목을 수치화하고, 이를 주 단위로 누적하여 궤적 그래프 형태의 피드백으로 제공한다. 사용자의 발화 내에서 반복적으로 발생하는 오류 음소는 추천 학습 항목으로 제시된다. 이를 바탕으로 학습자는 자신의 발화 패턴을 체계적으로 이해하고, 자기 주도적 학습 계획을 수립할 수 있으며, 지속적인 발음 교정을 도모할 수 있다.

Figure 11. Game Results Summary

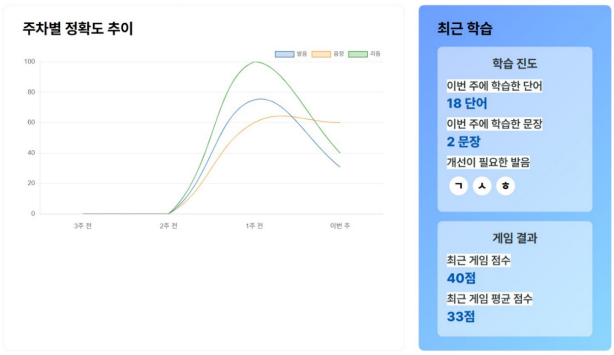

Figure 12. Week-by-Week Pronunciation Accuracy Change

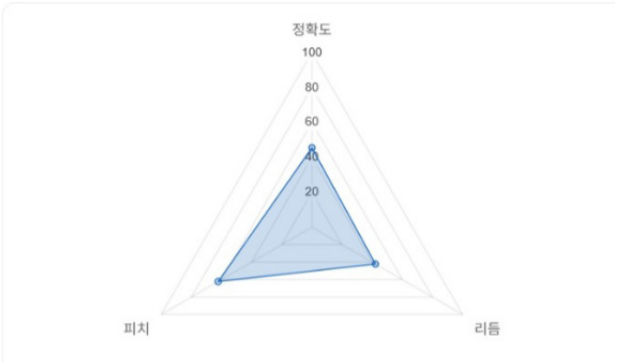

Figure 13. Pronunciation Accuracy, Pitch, Rhythm Triangle

Fig. 12, 13, 14는 담설을 통해 학습을 완료하면, 사용자가 보게 되는 리포트 화면이다. 상단 이미지는 이전에 학습한 기록이 있다면 주차별로 학습자의 발음 정확도 추이를 그래프로 정리해서 보여주며, 최근에 어느 부분까지 학습했는지 '학습 진도'로 정리해서 보여준다. 하단 이미지는 학습자의 발음 정확도, 피치, 리듬이 어느 정도인지 시각화해서 각각 보여준 것이다.

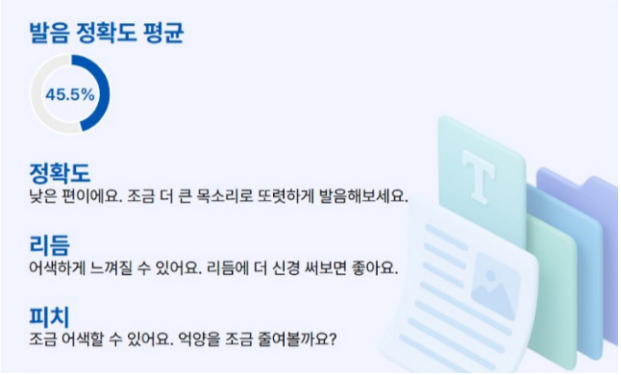

Figure 14. Overall Training Evaluation

5. 결과

본 논문에서 제안한 웹 기반 플랫폼 담설(談說)의 사용성 평가를 위해 중증(2급) 청각장애인 1명(이하 참여자)과 테스트를 수행하고 이후에 인터뷰를 수행하였다. 연구진 소속 학교의 장애학생지원센터의 도움을 받아 연결되었으며, 테스트 이후 참여자에게 소정의 사례를 하였다.

참여자는 선천적인 농인으로 수어를 유창하게 사용할 수 있는 30대 여성이었다. 비장애인과의 소통을 위해서 STT 기능이 내장된 스마트폰 어플리케이션을 사용하며, 담설의 사용성 테스트를 수행할 때에도 해당 어플리케이션의 도움을 받아 연구진과 소통하였다. 테스트는 튜토리얼 안내부터 인터뷰 종료까지 총 70여 분 가량 소요되었다.

테스트 순서는 다음과 같다. 담설에서 제공하는 기능을 정리한 튜토리얼의 인쇄본을 참여자에게 제공하여 시스템의 전체 기능이 무엇인지 익히도록 하고 각 기능에 대해 궁금한 점을 질문하는 시간을 가졌다. 이후 '단어 학습', '문장 학습', '고요 속의 외침(게이미피케이션)'을 수행하고, 리포트를 확인하는 순서로 진행되었다.

단어 학습 내에는 13가지의 카테고리, 문장 학습 내에는 19가지 카테고리가 각각 포함되어 있으며 한 개의 카테고리당 7개의 문제로 구성되어 있어 시간적 제약으로 인해 단어 학습 내의 두 개의 카테고리(모음, 음운의 변동), 문장 학습(특별한 상황-식당에서 주문)을 수행하였다. 단어 학습의 경우 총 14문제를 풀었으며, 인식이 잘 되지 않았던 1문제를 제외한 13문제의 평균 점수는 100점 만점에 32.3점이었다.

문장 학습의 경우 단어 학습에 비해 다소 난이도가 있는 관계로 2문제를 풀고 참여자의 요청에 의해 테스트를 중단하였다.

게이미피케이션의 경우, 기본적으로 주어진 5문제 중 3문제만 풀고 참여자의 요청에 의해 테스트를 중단하였다.

Table 6. Interview Question List

#	Category	Question
1	Training Process	Did you find the process of measuring your pronunciation score difficult when you first used the system?
2		As you practiced pronunciation at the word and sentence level, what was the most engaging part, and what was the hardest part?
3		How convenient was the voice recording and playback? What, if anything, was inconvenient?
4	Perceived Training Effectiveness	While using the system, did your pronunciation get better? What words or sentences made the improvement most noticeable?
5		How accurate is the system's pronunciation score as a measure of your pronunciation proficiency?
6		To what extent did the pre- vs. post-training score change boost your motivation? Please describe your feelings.
7	Gamification	Did you enjoy the activity where you watched an AI video and guessed the sentence from the lip movements?
8		How much did this gamified approach improve your training effectiveness or motivation?
9		Were there any inconveniences or areas you wanted to see improved while solving the tasks?
10	Usability & Pain Points	What was the single most helpful feature for your learning—visualization, pronunciation scoring, results summary, or something else? Why?
11		Were there features you considered redundant or not helpful? Please explain.
12		How likely are you to recommend this system to others? What's the main reason?
13		What would you like to see added or improved in future versions of the system?

테스트가 끝나고 난 뒤, Table 6와 같은 질문에 따라 인터뷰를 진행하였다.

참여자는 담설의 전반적인 기능을 사용해보고 해당 부분에 대해 다음과 같은 긍정적인 피드백을 주었다.

"단어만 보면 발음을 어떻게 해야할지 좀 막막했는데, 발음기호가 같이 나와서 따라할 수 있어서 좋았습니다."

"청각장애인들은 'ㅅ'발음을 어려워합니다. 실제로 'ㅅ'과 'ㅎ'의 구분이 어려워요. 학습 문제에서 'ㅅ'발음을 연습할 수 있어 좋은 것 같습니다."

"발음 점수를 보니, 새로고침해서 다시 연습해보고 싶은 생각이 듭니다. 아마 복습하다보면 더 나아지지 않을까 생각합니다."

"내 발음이 몇 점이라고 나타나는 부분이 있어서, 정확도가 객관화되어 알아볼 수 있어서 좋았습니다."

"단어 학습에는 확실히 효과가 있다고 생각합니다."

“요즘 젊은 청각장애인들은 우리(20대 후반~30대 초반)보다 훨씬 발음이 좋아요. 수어도 안배우고 지내는 친구도 많습니다. 사회생활을 하다가 소통을 하기 위해 뒤늦게 수어를 배우는 경우도 있고요. 그래서 저희같은 20대 후반~30대 초반 혹은 그보다 나이가 많은 연령대의 사람들에게 담설을 추천하고 싶습니다.”

참여자 피드백에 따르면, 발음 기호 제시는 단어 발음을 직관적으로 학습할 수 있도록 도와 학습 효과를 높였으며, 특히 청각장애인이 구분하기 어려운 자음(예: ‘ㅅ’, ‘ㅎ’)의 훈련 가능성을 긍정적으로 인식하였다. 또한 발음 점수와 같은 객관적 지표 제공은 자기 학습의 동기를 강화하고 반복 학습을 유도하는 촉진 요인으로 작용하였다. 더 나아가, 청각장애인의 발음 능력에 세대별 차이가 있음을 지적하며, 본 시스템이 20대 후반 이상 연령층에 특히 유용하게 기여할 수 있음을 시사하였다.

하지만 개선이 필요한 부분, 아쉬운 점에 대해서는 다음과 같은 피드백을 주었다.

“학습으로 넘어갔을 때, 아무런 지시문 없이 단어나 문장만 나와 있어서 뭘 하면 되는지 잘 모르겠다는 생각이 들었습니다. 화면에 ‘소리 내어 발음해보세요’ 문구가 추가되면 조금 더 직관적일 것 같습니다.”

“음성 녹음을 하기 위해서 마이크 버튼을 클릭하고, 녹음한 후 다시 정지 버튼을 클릭해서 녹음을 멈추는 과정이 조금 번거로웠습니다. 녹음을 멈춰야 하는 줄 학습자 입장에서 안내 말풍선이나 문구가 없어서 말하고 계속 기다리게 되는 것 같습니다.”

“AI 영상의 입모양을 보고 문장이 무엇인지 맞추는 문제(게이미피케이션)는 발음 교정과 어떤 관련이 있는지 잘 모르겠습니다. 시각장애인 입장에서 사실 와닿지 않고, 영상의 등장인물이 외국인인데다가 입모양이 작아서 무슨 발음을 하는지 알기 어려웠습니다. 차라리 타자연습할 때 단어가 화면에 무작위로 나오는 것을 따라치듯, 여기에서는 단어 누르고 말하고 점수를 획득하게 구성하는게 어떨까요.”

“문장은 솔직히 너무 길고 말하기 꺼려졌습니다. 단어를 더 많이 해보고 싶었습니다. 그리고 가능하다면 가, 나, 다…와 같은 낱 글자를 정확하게 말하는 방법을 기능으로 추가해주면 좋을 것 같습니다. 예를 들어 ‘나’ 할 때 이빨에 혀를 대고 소리낸다 와 같은 안내가 같이 나오도록 학습을 구성하는거죠.”

“문장을 연습한다면 내 발음이 ‘정확히 어디’가 틀렸는지 보여주면 좋겠습니다. 단어보다 긴데 틀린 곳이 어딘지 나오지 않아서 어떻게 고쳐서 다시 시도할지 알기 어렵습니다.”

“청각장애인들은 비장애인들과 달리, 한 글자 한 글자 또박또박 말하려고 노력합니다. 하지만 비장애인들은 빠르게 말하는 과정에서 저희만큼 발음을 또박또박 말하려 하지 않는 것 같습니다. 정확도를 측정하는 부분에서 청각장애인들의 이러한 특성이 잘 반영되지 않은 것 같습니다.”

“음성 녹음을 할 때, 청각장애인들은 자기가 크게 말하는지 작게 말하는지 감이 없습니다. 목소리 크기에 따라 화면에 별도로 시각화되어 나오는 무언가가 있으면 도움이 될 것 같아요.”

참여자는 학습 화면에서 구체적인 지시문 부재로 인해 학습 흐름이 직관적으로 이해되지 않는 점을 지적하였으며, 발화 유도 문구 제시와 같은 명확한 안내가 필요함을 제안하였다. 또한 음성 녹음 과정에서 버튼 조작이 번거롭고, 녹음을 종료해야 한다는 점에 대한 안내 부족으로 학습자가 혼란을 겪을 수 있음을 언급하였다. 이와 더불어, 녹음 후 분석 결과가 나타나기까지 단어는 8~20초, 문장은 20~38초가 소요되어 결과 대기 시간이 과도하게 길었으며, 연구진이 로그 메시지를 통해 내부 분석이 진행됨을 확인할 수 있었음에도 참여자는 아무런 안내 없이 기다려야 했다는 점이 주요 개선 과제로 확인되었다.

발화 인식률과 평가 정확도 측면에서도 한계가 드러났다. KoSpeech로 학습된 모델은 비장애인의 경우 인식률이 100%에 근접하였으나, 청각장애인 음성이 포함되지 않아 참여자의 발화 인식률과 평가 점수가 낮게 측정되었다. 이는 본 시스템이 청각장애인의 발음을 비장애인과 유사한 수준으로 지원한다는 초기 목표와 달리 실제 사용자에게는 불리하게 작용했음을 보여준다. 따라서 향후 청각장애인 음성을 포함한 추가 학습을 통해 발음 인식과 평가 체계를 개선할 필요가 있다.또한 참여자는 긴 문장보다는 단어 중심 학습을 선호하며, 문장 발화에서는 부담감을 호소하였다. 이는 선행연구에서 보고된 바와 같이, 단어 발음 정확도가 높더라도 연속적인 발화나 문장 단위 정확도와의 상관성이 낮다는 점과 일치한다. 실제로 문장 발화 결과에서 피치는 양호하게 측정되었으나 리듬과 속도는 부정적으로 평가되었다. 따라서 문장 발음을 학습할 때는 학습자가 현재 발화 중인 문장의 특정 부분과 권장 속도 대비 실제 발화 속도를 시각적으로 제공하고, 평가 결과에서도 문장의 어느 부분에서 개선이 필요한지를 구체적으로 제시하는 방식이 보다 효과적인 학습을 지원할 수 있을 것이다.

마지막으로, 참여자는 청각장애인이 한 글자씩 또박또박 발화하는 특성을 강조하며, 현 시스템의 정확도 측정 방식이 이를 충분히 반영하지 못하고 있다고 지적하였다. 또한 음량에 대한 자기 인지가 어려운 특성을 고려할 때, 발화 크기를 시각적으로 피드백하는 기능이 제공된다면 학습 효용성이 더욱 향상될 것으로 기대된다.

6. 논의 및 결론

본 논문에서는 청각장애인의 발음 교정을 위한 웹 기반 플랫폼 담설(談說)을 개발하여 기존 입 모양 피드백 중심 프로그램과 상용 STT 엔진 기반 발음 교정 시스템의 한계를 보완하고자 하였다. 기존 접근은 주로 입술 움직임에만 의존하거나 비장애인 발화를 기반으로 학습된 음성 인식 엔진

을 활용함으로써, 청각장애인의 발화 특성을 충분히 반영하지 못하고 정확한 발음 오류 분석에도 제약이 있었다. 이에 본 연구에서는 자체 학습한 한국어 ASR 모델과 대규모 언어모델(LLM)을 연계하여 비표준 발화에 대한 인식 성능을 확보하고, 음성 성분 분석을 통한 다차원적 피드백을 제공하는 구조를 구현하였다. 이를 통해 청각장애인의 실제 발음 개선을 유도할 수 있는 학습 환경을 마련한 점에서 학문적·실천적 의의를 지닌다.

담설의 음성 인식 및 피드백 모듈은 KoSpeech 기반 음성 인식 모델, Librosa 라이브러리, GPT-4o 모델을 연계하여 구성되었으며, 이를 통해 사용자의 발음을 정확도, 피치, 발화 속도, 리듬 등 다양한 차원에서 분석하였다. 또한 음성 성분별 특화된 평가 프롬프트를 적용하여 학습자가 자신의 발화 습관을 파악하고 반복 학습을 통해 교정할 수 있도록 지원하였다. 실제 단어·문장 학습 모듈의 적용 결과, 조용한 환경에서 고품질 마이크를 사용하는 조건에서는 높은 인식 정확도와 자연스러운 피드백 제공이 가능함을 확인하였다.

그러나 본 연구에는 몇 가지 한계가 존재한다. 첫째, 실험 참여자의 경험에 따르면 단어는 8~20초, 문장은 20~38초에 달하는 지연이 발생하여 실시간 학습 몰입에 제약이 있었다. 둘째, KoSpeech 모델이 초기에는 비장애인 음성만을 기반으로 학습되어 청각장애인 발화의 인식률과 평가 점수가 낮게 측정되는 문제가 확인되었다. 이는 향후 청각장애인 음성 데이터를 포함한 추가 학습을 통해 개선될 필요가 있다. 셋째, 참여자는 긴 문장 발화에 부담을 느끼고 정확도 측정에서도 리듬과 속도가 낮게 평가되었는데, 이는 단어 정확도와 문장 정확도의 낮은 상관성을 재확인한 결과였다. 따라서 문장 발화 학습 시 발화 위치, 속도, 강세를 시각적으로 제공하고 오류 구간을 구체적으로 제시하는 보완이 요구된다. 마지막으로, 청각장애인의 특성인 한 글자씩 또박또박 발화하는 습관이나 자기 음량 인지의 어려움이 충분히 반영되지 않았다는 점에서, 향후 발화 크기 시각화 등 장애 특성을 고려한 맞춤형 피드백 기능이 추가적으로 필요하다.

또한 본 연구는 정성적 평가 위주로 진행되었고, 청각장애인 한 명의 의견을 일반화하기 어렵다는 점에서 한계가 있다. 따라서 다양한 환경과 대상자를 포함한 추가 데이터 수집과 정량적 효과 검증이 후속 연구로 수행되어야 할 것이다.

향후 연구에서는 스트리밍 기반 추론과 데이터 증강을 통해 지연을 줄이고 다양한 소음·장치 환경에서의 인식 안정성을 확보하며, 피치·발화 속도·리듬 등 핵심 초분절 지표를 통합한 피드백을 제공하도록 알고리즘을 개선한다. 또한 테스트 수행 시 수집한 사용자 의견을 반영해 UI/UX를 재설계하고, 평가 화면에서 음절·단어·문장 단위로 어느 부분이 틀렸고 어느 부분이 정확했는지를 색상 하이라이트·타임라인 마커·예시 발화 비교로 정밀 피드백을 제공할 예정이다. 더불어 실시간 음성 시각화(음량 레벨·피치 곡선·템포/리듬·강세 표시)를 도입하여 학습자가 발화 순간의 상태를 직관적으로 확인하고 즉시 수정할 수 있도록 할 것이다.

효과 검증은 사전등록된 설계에 따라 파일럿 후 본시험(무작위 비교)을 진행한다. 정량 분석의 신뢰구간을 확보하기 위해 단일군 기준 최소 25명을 모집하고, 표본은 연령, 청각장애 유형, 학습 맥락을 포괄하도록 할 예정이다. 또한 데이터 수집·활용 전 과정에서 윤리·보안 절차를 준수할 것이다.

참고문헌

[1] Chung, J., & Kwak, M. (2018). Perspectives in Auditory Rehabilitation. *J Clinical Otolaryngol,* 29(1), 5-10. https://doi.org/10.35420/jcohns.2018.29.1.5

[2] Ghazi-Saidi, L., & Ansaldo, A. I. (2017). Second language word learning through repetition and imitation: Functional networks as a function of learning phase and language distance. F*rontiers in human neuroscience, 11,* 463. https://doi.org/10.3389/fnhum.2017.00463

[3] Lester, N., Moran, S., Küntay, A., Allen, S., Pfeiler, B., & Stoll, S. (2022). Detecting structured repetition in child-surrounding speech: Evidence from maximally diverse languages. *Cognition*, 221, 104986. https://doi.org/10.1016/j.cognition.2021.104986

[4] Ahmadi, S. (2016). The importance of listening comprehension in language learning. *International Journal of Research in English Education*, 1(1), 7-10.

[5] Shojaei, E., Jafari, Z., & Gholami, M. (2016). Effect of early intervention on language development in hearing-impaired children. Iranian journal of otorhinolaryngology, 28(84), 13.

[6] Nagaraja, J., Hamid, B., & Maamor, N. (2024). Phonological Acquisition Process in Hearing-Impaired Children: A Systematic Review. *GEMA Online Journal of Language Studies*, 24(3). https://doi.org/10.17576/gema-2024-2403-08

[7] Bowe, F. (1998). Language development in deaf children. *The Journal of Deaf Studies and Deaf Education*, 3(1), 73-77. https://doi.org/10.1093/oxfordjournals.deafed.a014342

[8] Stadio, A., Sossamon, J., Luca, P., Indovina, I., Motta, G., Ralli, M., Brenner, M., Frohman, E., & Plant, G. (2025). “Do You Hear What I Hear?” Speech and Voice Alterations in Hearing Loss: A Systematic Review. *Journal of Clinical Medicine,* 14(5), 1428. https://doi.org/10.3390/jcm14051428

[9] Yeom, S., & Lee, E. (2020). Research Trends on Pragmatic Language Ability of Individuals with Hearing Impairment. *Journal of Special Education,* 36(2), 81-204. https://doi.org/10.31863/JSE.2020.08.36.2.81

[10] Lee, S., Kim, H., Sim, H., Nam, C., Choi, J., & Park, E. (2010). Auditory-Perceptual Evaluation of the Speech of Adults with Hearing Impairment Based on Suprasegmental Factors, Speech Intelligibility, and Speech Acceptability.

Communication Sciences and Disorders, 15(4), 477-493.
[11] Spaai, G., Derksen, E., Hermes, D., & Kaufholz, P. (1996). Teaching intonation to young deaf children with the intonation meter. *Folia phoniatrica et logopaedica*, 48(1), 22-34. https://doi.org/10.1159/000266379
[12] Deroche, M., Kulkarni, A., Christensen, J., Limb, C., & Chatterjee, M. (2016). Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants. *Frontiers in Neuroscience,* 10, 73. https://doi.org/10.3389/fnins.2016.00073
[13] now, D., & Ertmer, D. (2012). Children's development of intonation during the first year of cochlear implant experience. *Clinical linguistics & phonetics*, 26(1), 51-70. https://doi.org/10.3109/02699206.2011.588371
[14] Bennett, C. (1978). Articulation training of profoundly hearing-impaired children: A distinctive feature approach. *Journal of Communication Disorders*, 11(5), 433-442. https://doi.org/10.1016/0021-9924(78)90036-9
[15] Fletcher, S., Dagenais, P., & Critz-Crosby, P. (1991). Teaching vowels to profoundly hearing-impaired speakers using glossometry. *Journal of Speech, Language, and Hearing Research*, 34(4), 943-956. https://doi.org/10.1044/jshr.3404.943
[16] letcher, S., Dagenais, P., & Critz-Crosby, P. (1991). Teaching consonants to profoundly hearing-impaired speakers using palatometry. Journal of Speech, *Language, and Hearing Research*, 34(4), 929-943. https://doi.org/10.1044/jshr.3404.929
[17] Shin, E., Cho, S., & Lee, H. (2022). A preliminary study on standardization of phoneme perception test for school-aged children : Focused on hearing impaired children. *The Journal of the Acoustical Society of Korea*, 41(1), 99-107. https://doi.org/10.7776/ASK.2022.41.1.099
[18] Park, K., & Seol, H. (2025). A Review on the Evaluation of Speech Acceptability in Individuals with Hearing Impairment. Audiology and Speech Research, 21(1), 1-8. https://doi.org/https://doi.org/10.21848/asr.240172
[19] Tomblin, J., Peng, S., Spencer, L., & Lu, N. (2008). Long-term trajectories of the development of speech sound production in pediatric cochlear implant recipients. J*ournal of Speech, Language, and Hearing Research*, 51(5), 1353-1368. https://doi.org/10.1044/1092-4388(2008/07-0083)
[20] Tobey, E., Geers, A., Sundarrajan, M., & Shin, S. (2011). Factors influencing speech production in elementary and high school-aged cochlear implant users. *Ear and hearing*, 32(1), 27S-38S. https://doi.org/10.1097/AUD.0b013e3181fa41bb
[21] Kang, J., & Yoon, M. (2020). A Comparison of the Speech Production Ability of Children With Cochlear Implants and Children With Normal Hearing. *Journal of Speech-Language & Hearing Disorders*, 29(1), 13-21. https://doi.org/10.15724/jslhd.2020.29.1.013
[22] Karimi-Boroujeni, M., Dajani, H., & Giguère, C. (2023). Perception of prosody in hearing-impaired individuals and users of hearing assistive devices: An overview of recent advances. *Journal of Speech, Language, and Hearing Research*, 66(2), 775-789. https://doi.org/10.1044/2022_JSLHR-22-00125
[23] Larrouy-Maestri, P., Poeppel, D., & Pell, M. (2025). The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspectives on Psychological Science*, 20(4), 623-638. https://doi.org/10.1177/17456916231217722
[24] Sobhy, O., Abdou, R., Ibrahim, S., & Hamouda, N. (2021). Effects of a prosody rehabilitation program on expression of affect in preschool children with hearing impairment: a randomized trial. *The Egyptian Journal of Otolaryngology*, 37(1), 60. https://doi.org/10.1186/s43163-021-00119-4
[25] Gwak, S., Baek, J., & Yoo, S. (2022). Exploring the Application of Playful Learning in SW Liberal Education to Enhance Learning Motivation : Focusing on non-CS student. *Journal of the Korean Association of information Education, 26*(5), 327-340. http://dx.doi.org/10.14352/jkaie.2022.26.5.327
[26] Rojabi, A., Setiawan, S., Munir, A., Purwati, O., Safriyani, R., Hayuningtyas, N., ... & Amumpuni, R. (2022). Kahoot, is it fun or unfun? Gamifying vocabulary learning to boost exam scores, engagement, and motivation. *Frontiers in Education*. 7, 939884. https://doi.org/10.3389/feduc.2022.939884
[27] Darwich, L., DeBay, D., Forbes, L., & Mahfouz, J. (2025). Play, reflect, cultivate social and emotional learning: a pathway to pre-service teacher SEL through playful pedagogies. *Frontiers in Education*. 9, 1478541. https://doi.org/10.3389/feduc.2024.1478541
[28] Alotaibi, M. (2024). Game-based learning in early childhood education: a systematic review and meta-analysis. *Frontiers in psychology*, 15, 1307881. https://doi.org/10.3389/fpsyg.2024.1307881
[29] Oh, S., Won, Y., & Lee, Y. (2025). A Meta-Analysis of the Effects of Game-Based Learning in English Education in South Korea. *Journal of Educational Technology*, 41(2), 577-600. http://dx.doi.org/10.17232/KSET.41.2.577
[30] Gentry, S., Gauthier, A., Ehrstrom, B., Wortley, D., Lilienthal, A., Car, L., ... & Car, J. (2019). Serious gaming and gamification education in health professions: systematic review. *Journal of medical Internet research*, 21(3), e12994. https://doi.org/10.2196/12994
[31] Lee, M., Shin, S., Lee, M., & Hong, E. (2024). Educational outcomes of digital serious games in nursing education: a systematic review and meta-analysis of randomized controlled trials. *BMC Medical Education*, 24(1), 1458. https://doi.org/10.1186/s12909-024-06464-1
[32] Tye-Murray, N., Spehar, B., Sommers, M., Mauzé, E., Barcroft, J., & Grantham, H. (2022). Teaching children with hearing loss to recognize speech: Gains made with computer-based auditory and/or speechreading training. *Ear and hearing*, 43(1), 181-191. https://doi.org/10.1097/AUD.0000000000001091
[33] Spehar, B., Tye-Murray, N., Mauzé, E., Sommers, M., & Barcroft, J. (2024). Speech Perception Training in Children: The Retention of Benefits and Booster

Training. *Ear and hearing*, 45(1), 164-173. https://doi.org/10.1097/AUD.0000000000001413

[34] Pimperton, H., Kyle, F., Hulme, C., Harris, M., Beedie, I., Ralph-Lewis, A., ... & MacSweeney, M. (2019). Computerized speechreading training for deaf children: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research*, 62(8), 2882-2894. https://doi.org/10.1044/2019_JSLHR-H-19-0073

[35] Parmar, B. J., Salorio-Corbetto, M., Picinali, L., Mahon, M., Nightingale, R., Somerset, S., ... & Vickers, D. (2024). Virtual reality games for spatial hearing training in children and young people with bilateral cochlear implants: the "Both Ears (BEARS)" approach. *Frontiers in Neuroscience*, 18, 1491954. https://doi.org/10.3389/fnins.2024.1491954

[36] Saeedi, S., Bouraghi, H., Seifpanahi, M., & Ghazisaeedi, M. (2022). Application of digital games for speech therapy in children: a systematic review of features and challenges. *Journal of healthcare engineering*, 2022(1), 4814945. https://doi.org/10.1155/2022/4814945

[37] Porcar-Gozalbo, N., López-Zamora, M., Valles-González, B., & Cano-Villagrasa, A. (2024). Impact of hearing loss type on linguistic development in children: A cross-sectional study. *Audiology Research, 14(6)*, 1014-1027. https://doi.org/10.3390/audiolres14060084

[38] Huang, W., Wong, L., & Chen, F. (2022). Just-noticeable differences of fundamental frequency change in mandarin-speaking children with cochlear implants. *Brain Sciences*, 12(4), 443. https://doi.org/10.3390/brainsci12040443

[39] Choi, A., Kim, H., Jo, M., Kim, S., Joung, H., Choi, I., & Lee, K. (2024). The impact of visual information in speech perception for individuals with hearing loss: a mini review. *Frontiers in psychology*, 15, 1399084. https://doi.org/10.3389/fpsyg.2024.1399084

[40] Hitchcock, E., Ochs, L., Swartz, M., Leece, M., Preston, J., & McAllister, T. (2023). Tutorial: Using visual–acoustic biofeedback for speech sound training. *American journal of speech-language pathology*, 32(1), 18-36. https://doi.org/10.1044/2022_AJSLP-22-00142

[41] Hao, S., Wang, Q., Zhang, Y., Miao, Y., & Shan, Y. (2024). The effect of different visual feedback interfaces of music training games on speech rehabilitation in hearing-impaired children: An fNIRS study. *Neuroscience Letters*, 843, 138010. https://doi.org/10.1016/j.neulet.2024.138010

[42] Renckens, M., Raeve, L., Nuyts, E., Mena, M., & Bessemans, A. (2021). Visual prosody supports reading aloud expressively for deaf readers. *Visible Language*, 55(1). https://doi.org/10.34314/vl.v55i1.4603

[43] Lee, Y., Lim, S., Choi, Y., & Moon, B. (2015). A Mobile App(See&Speech) of Correcting Pronunciation for Hearing-Impaired Persons. *The Journal of Korean Association of Computer Education*, 18(4), 11–18. https://doi.org/10.32431/kace.2015.18.4.002

[44] Jeong, H., Jeong, D., Lee, J., & Kim, B. (2017). Development of Smart Mirror System for Hearing Deaf's Pronunciation Training. *Journal of Digital Contents Society*, 18(2), 267–274. https://doi.org/10.9728/dcs.2017.18.2.267

[45] Eo, S., & Kim, Y. (2006). Pronunciation Learning System Development for The Language Impairment. *Proceedings of the Korea Multimedia Society Conference,* Korea, 2006(5), 685-688.

[46] Librosa. (2025, March 11). (Version 0.11.0). [Computer software]. https://zenodo.org/records/15006942

[47] Lee, M. (2018). A Lip-reading Algorithm Using Optical Flow and Properties of Articulatory Phonation. *Journal of Korea Multimedia Society*, 21(7), 745-754. https://doi.org/10.9717/kmms.2018.21.7.745

[48] Zhu, S. (2022). Research on the dynamic viseme of the lip shape based on facial motion capture technology. *Frontiers in Neurorobotics*, 16, 922756. https://doi.org/10.3389/fnbot.2022.922756

[49] Sato, Y., & Bao, Y. (2022). Identification of 3D Lip Shape during Japanese Vowel Pronunciation Using Deep Learning. *Applied Sciences*, 12(9), 4632. https://doi.org/10.3390/app12094632

[50] Stavness, I., Nazari, M., Perrier, P., Demolin, D., & Payan, Y. (2013). A biomechanical modeling study of the effects of the orbicularis oris muscle and jaw posture on lip shape. *Journal of Speech, Language, and Hearing Research*, 56(3), 878-890. https:// doi.org/10.1044/1092- 4388(2012/12-0200)

[51] Google LLC. (n.d.). Speech-to-Text (API) [Computer software]. Google Cloud. https://cloud.google.com/speech-to-text

[52] Microsoft Corporation. (2025, March 10). Azure AI Speech service [Computer software]. Microsoft Azure. https://azure.microsoft.com/en-us/products/ai-services/ai-speech

[53] OpenAI. (2022, September 21). Whisper (Version large-v2) [Computer software]. https://github.com/openai/whisper

[54] Kang, C., Lee, Y., & Chung, M. (2023). SINABULO: pronunciation correction program to improve delayed speech development. *Annual Conference of KIPS, Korea, 30*(2), 757-758. https://doi.org/10.3745/PKIPS.Y2023M11A.757

[55] mon Moxon. (2024). All-Talk: Enhancing EFL Pronunciation With Microsoft Azure Speech Services. *ABAC Journal*, 44(4), 139–161. https://doi.org/10.59865/abacj.2024.58

[56] Glasser, A., Kushalnagar, K., & Kushalnagar, R. (2017). Feasibility of using automatic speech recognition with voices of deaf and hard-of-hearing individuals. *In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 373-374. https://doi.org/10.1145/3132525.3134819

[57] Kim, S. (2020). KoSpeech (Version latest) [Computer software]. https://github.com/sooftware/KoSpeech

[58] ETRI. (2019, May 15). Korean speech data (Version 1.0). [Audio Data]. Daegu: AI-Hub. https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100

&aihubDataSe=realm&dataSetSn=123
[59] Spring Boot. (2025, January 23). (Version 3.4.2) [Computer software]. https://spring.io/projects/spring-boot
[60] Flask. (2024, November 13). (Version 3.1.0) [Computer software]. https://flask.palletsprojects.com/en/stable/
[61] OpenAI. (2024, November 20). GPT-4o [Large Language Model]. https://platform.openai.com/docs/models/gpt-4o
[62] Chun, E., Jeong, Y., Kim, H., Kim, N., Kim, S., & Lee, Y. (2024). Perception of Voice Attractiveness: Effects of Speaking Rate, Gender, and Age. *Communication Sciences and Disorders*, 29(2), 462-472. https://doi.org/10.12963/csd.240029
[63] Won, Y. (2022). A study on the change of prosodic units by speech rate and frequency of turn-taking. *Phonetics and Speech Sciences*, 14(2), 29-38. http://doi.org/10.13064/KSSS.2022.14.2.029
[64] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. https://doi.org/10.48550/arXiv.1706.03762
[65] Chan, W., Jaitly, N., Le, Q., & Vinyals, O. (2016, March). Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. *In 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4960-4964). IEEE. https://doi.org/10.1109/ICASSP.2016.7472621
[66] Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., ... & Zhu, Z. (2016, June). Deep speech 2: End-to-end speech recognition in english and mandarin. *In International conference on machine learning* (pp. 173-182). PMLR.
[67] Python. (2023, June 6). (Version 3.10.12). [Computer software]. https://www.python.org/downloads/release/python-31012/
[68] FaceFusion. (2025, March 28). (Version 3.1.2). [Computer software]. https://docs.facefusion.io/

이기정

· 2019년~현재 한성대학교 IT융합공학부 사이버보안트랙 재학

관심분야 : LLM, LVM, 웹 백엔드 개발
keejungyutiyub@gmail.com

윤희욱

· 2020년~현재 한성대학교 컴퓨터공학부 웹공학트랙 재학

관심분야 : 웹 개발, HCI, 기술 접근성, 인공지능, UI/UX
hw010603@naver.com

오승민

· 2021년~현재 한성대학교 컴퓨터공학부 모바일소프트웨어트랙 재학

관심분야 : 웹 개발, 인공지능, 보조공학, UI/UX
osmksh1004@naver.com

조민서

· 2021년~현재 한성대학교 컴퓨터공학부 모바일소프트웨어트랙 재학

관심분야 : 생성형 AI, 사이버 보안, HCI
minseojo1@naver.com

유수진

· 2010년 고려대학교 컴퓨터교육과(이학사)
· 2012년 고려대학교 일반대학원 컴퓨터교육학과 (이학석사)
· 2020년 고려대학교 일반대학원 컴퓨터학과 (공학박사)
· 2015년 2월~2017년 5월 롯데카드 IT기획팀, IT비즈팀
· 2017년 10월~2018년 12월 Stipop 풀스택 개발자, CTO
· 2021년 3월~2021년 12월 한양대학교 SW융합원 SW교육전담교수
· 2022년 2월~2022년 12월 성균관대학교 소프트웨어융합대학 초빙교수
· 2023년 1월~2024년 8월 고려대학교 데이터과학원 연구교수
· 2024년 9월~현재 한성대학교 컴퓨터공학부 웹공학트랙 조교수

관심분야 : 지식그래프, 링크드데이터, 컴퓨터교육, 인공지능 윤리, 디지털격차
sujin.yoo@hansung.kr