

세계한국어한마당 학술대회

[국제한국어교육학회 · 이중언어학회]

한국어 발음 교육의 미래를 말하다

- AI, 음성 인식 기술에 기반한 한국어 발음 교육의 현재와 미래

컴퓨터 기반 한국어 발음 훈련

Computer Assisted Pronunciation Training for Korean

2022년 10월 6일

정민화 교수

서울대학교 언어학과

mchung@snu.ac.kr

목차

AI, 음성인식 기술에 기반한 한국어 발음 교육의 현재와 미래

- 컴퓨터 기반 한국어 발음 훈련 소개
 - MDD (Mispronunciation Detection and Diagnosis) vs. Assessment
 - 모국어 특성에 따른 외국인의 한국어 발음 특징 예
- 컴퓨터 기반 발음 훈련 연구동향 소개
- 컴퓨터 기반 한국어 발음훈련을 위한 연구자원 소개
 - AI Hub Data
 - Montreal Forced Aligner

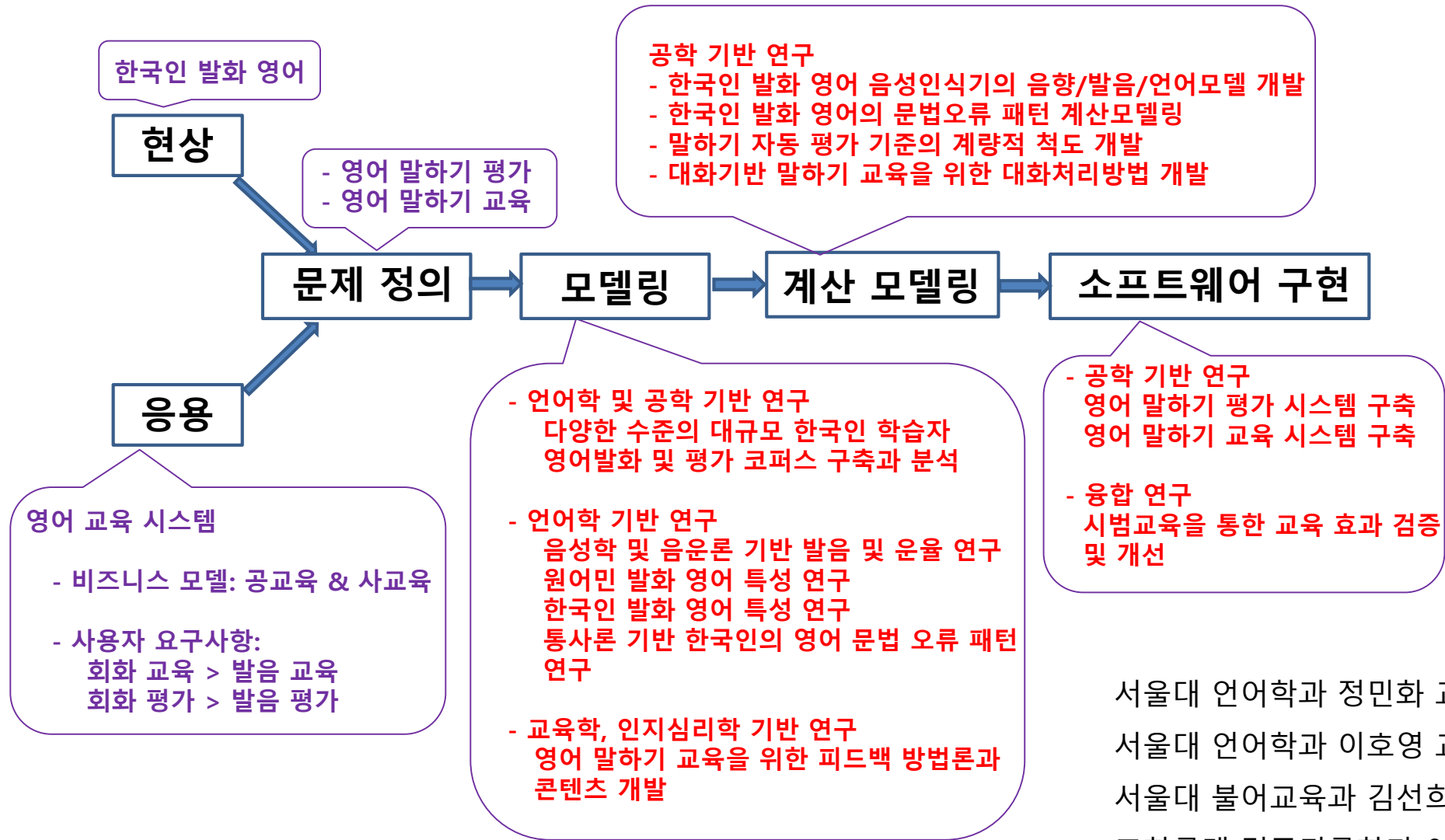
컴퓨터 기반 언어 학습 (CALL), 컴퓨터 기반 발음 훈련 (CAPT)

- CALL: Computer Assisted Language Learning
 - 한국인을 위한 외국어(영어/중국어/일본어/스페인어 등) 학습과 평가
 - 외국인(영어/중국어/일본어/스페인어 등의 모국어 사용자)을 위한 한국어 학습과 평가
 - 다양한 L1, L2 조합
 - 말하기, 듣기, 읽기, 쓰기의 모든 영역
 - **말하기 영역: 발음/말하기/회화 학습 및 평가**
 - 이론 언어학 (음성학, 음운론, 형태론, 통사론, 의미론) + 응용 언어학 (언어습득, 언어교육, 심리언어학)
+ 공학 (음성인식, 음성합성, 자연어처리, 인공지능)
- CAPT: Computer Assisted Pronunciation Training (Computer Aided Pronunciation Training)
 - 컴퓨터 기반 한국어 발음 훈련

컴퓨터 기반 한국어 발음 훈련

- 훈련과 평가는 다른 문제이며 다른 접근 방법을 취하고 있음
- 훈련: MDD (Mispronunciation Detection and Diagnosis) Problem
 - Error Detection and Corrective Feedback Generation Problem
 - 외국인의 한국어 발음 오류 검출
 - ➔ 외국인의 모국어 특성을 반영한 비원어민 발화 한국어 모델 개발 필요
 - 외국인 학습자가 이해하기 쉽고 따라하기 쉬운 오류 설명 및 교정 피드백 생성
 - 학습/훈련을 위해 학습자와 상호작용이 필요함 ➔ 비원어민 발화 한국어 음성인식기 개발 필요
- 평가: Assessment Problem
 - 인간 평가자의 점수와 자동평가 소프트웨어의 점수의 상관관계 제고가 목표
 - 자동평가 소프트웨어의 기계학습 알고리즘이 채택하는 평가요소는 인간 전문가의 평가요소와 크게 다를 수 있음
 - 원어민 발화 한국어 모델을 기준으로 발음 평가

한국인을 위한 영어 말하기 평가 및 교육 사례: 서울대 & 포항공대



컴퓨터 기반 한국어 발음 훈련 솔루션 개발을 위한 언어학 연구

- 모국어 특성에 따른 외국인의 한국어 발음 특징 예
 - 영어권 화자는 영어와 한국어 간 다른 음소 체계로 파열음의 상관속 구분, 다양한 환경에서의 올바른 유음([l],[r]) 사용, 모음 /ɪ/, /—/ 발음 등에 어려움을 겪을 수 있음. 해당 요소들은 뜻을 구분하여 표현하는 것으로 이어지기 때문에 주요 평가 요소로 간주되어야 함. 더불어, 영어는 강세언어로서 한국어와 다른 리듬 및 억양 패턴을 가지고 있기 때문에 언어 종속적 요소로 정의하여 평가할 필요 있음.
 - 스페인어를 모국어로 하는 화자는 스페인어와 한국어 간 다른 음소 체계로 파열음의 상관속 구분, 자음 /ㅎ/, /ㅈ/, /ㅉ/, 모음 /ɪ/, /—/ 발음 등에 어려움을 겪을 수 있음. 스페인어는 상대적으로 한국어와 비슷한 리듬 체계를 가지기 때문에, 타언어권 학습자에 비해 리듬 관련 어려움을 겪을 가능성은 적음.
 - 프랑스어를 모국어로 하는 화자는 프랑스어와 한국어 간 다른 음소 체계로 파열음의 상관속 구분, 마찰음의 상관속 구분, 모음 /ɪ/와 /ɒ/의 구분, 자음 /ㅎ/ 및 종성 /ㅇ/의 발음 등에 어려움을 겪을 수 있음. 프랑스어 억양의 음역은 한국어보다 상대적으로 넓고 역동적이라는 특징을 가짐. 이와 같은 모국어 억양 간섭이 유창성에 영향을 미칠 수 있음.
 - 독일어를 모국어로 하는 화자는 독일어와 한국어 간 다른 음소 체계로 파열음 및 마찰음의 상관속 구분, 구개음 /ㅈ/와 /ㅊ/의 발음 등에 어려움을 겪을 수 있음.

컴퓨터 기반 한국어 발음 훈련 솔루션 개발을 위한 언어학 연구

■ 모국어 특성에 따른 외국인의 한국어 발음 특징 예

- 러시아어를 모국어로 하는 화자는 러시아어와 한국어 간 다른 음소 체계로 파열음 및 마찰음의 상관속 구분, /ㅎ/의 탈락, 파찰음 /ㅈ/와 /ㅊ/의 발음, 종성 /ㅇ/의 발음, 모음 /ㅣ/와 /ㅡ/ 및 반모음 /w/의 발음 등에 어려움을 겪을 수 있음.
- 러시아어에는 한국어의 비음화, 경음화, 유음화 등 음소 단위 대치에 해당하는 음운 규칙이 존재하지 않으므로, 이들 음운 규칙을 적용하는 데 어려움을 겪을 수 있음.
- 중국어를 모어로 하는 화자는 중국어와 한국어 간 다른 음소 체계로 파열음의 상관속 구분, 초성 /ㄹ/의 탄설음, 받침 /ㄹ/의 설측음, 모음 /ㅡ/, /ㅣ/ 발음에 어려움을 겪을 수 있음.
- 중국어는 음절 박자 언어(syllable-timed language)로 한국어와 유사한 리듬 패턴 및 억양 패턴을 가지고 있음. 따라서 중국어 화자는 타언어권 학습자보다 한국어 리듬 패턴을 용이하게 학습할 가능성이 높음.
- 그러나 중국어는 한국어와 다르게 성조 언어이기 때문에, 한국어 발음에 간섭할 수 있음. 특히, 중국어의 평서문은 하강의 경계 성조를 가지므로 의미에 따라 다양하게 분화되는 한국어의 문말 억양을 어려워하거나 과적용할 수 있음. 또한, 중국어의 성조 조합이 한국어 한자어 음높이에 영향을 미치기도 함. 부자연스러운 억양 사용은 유창성(fluency)와 이해가능도(comprehensibility)에 영향을 미칠 수 있기 때문에, 발음 평가 요소로 간주되어야 함.

컴퓨터 기반 한국어 발음 훈련 솔루션 개발을 위한 언어학 연구

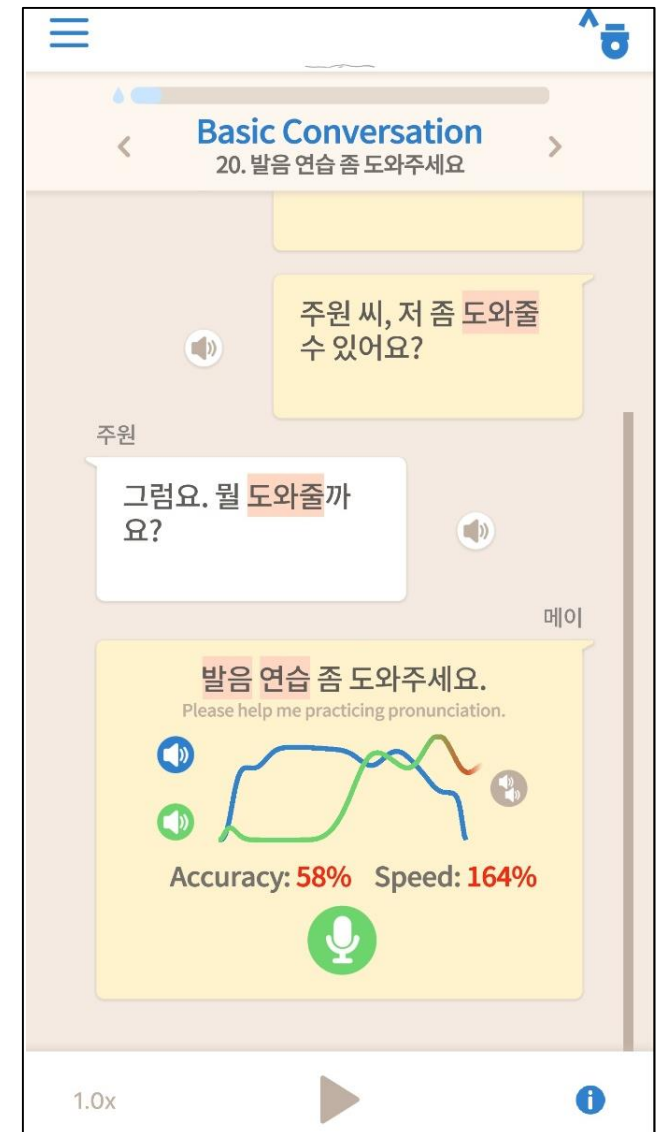
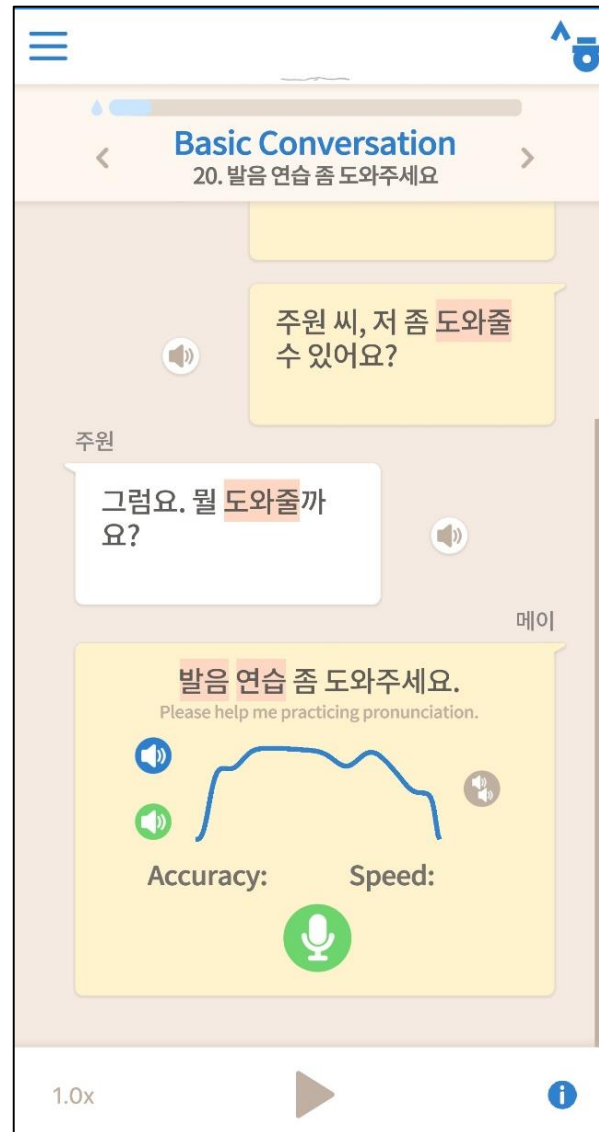
■ 모국어 특성에 따른 외국인의 한국어 발음 특징 예

- 일본어를 모어로 하는 화자는 일본어와 한국어 간 다른 음소 체계로 파열음, 파찰음의 상관속 구분, 비음 /ɔ/, 모음 /ㅡ/, /ㅣ/, 이중모음 발음에 어려움을 겪을 수 있음.
- 일본어는 mora-timed language로 한국어와 상이한 리듬 패턴을 가지고 있음. 이는 두 언어 간 다른 음절 구조로 이어져(모라 vs 음절) 한국어의 음절 발음을 어려워할 수 있음. 예를 들어, 일본어에서는 빈도가 낮은 CVC 음절을 하나의 음절이 아닌 두 개의 음절로 잘못 발음하는 사례가 빈번하게 보고됨.
- 또한 일본어는 한국어와 달리 고저 악센트(pitch accent)를 가진 언어로, 일본인 학습자는 한국어의 억양 대신 모어인 일본어의 악센트를 잘못 사용할 수 있음.
- 베트남어, 몽골어, 미얀마어 등 중국 또는 일본 이외 아시아어는 한국어가 따르는 평음-격음-경음의 삼지적 상관속과 다른 파열음 및 파찰음의 체계를 가지고 있음. 이와 같은 음소 체계 차이로 인해 파열음 및 파찰음을 구분하여 발음하는 것을 어려워할 수 있음.
- 또한 이들 언어에는 대응되는 음소가 잘 존재하지 않는 한국어의 모음 /ㅡ/, ㅣ/ 발음 등에 어려움을 겪을 수 있음.
- 특히 베트남어, 태국어, 미얀마어 등 성조가 존재하는 언어를 모국어로 사용하는 화자의 경우 모국어의 성조 체계의 간섭으로 인해 한국어의 억양 패턴을 학습하고 말하는 데 어려움이 있을 수 있음.

컴퓨터 기반 한국어 발음훈련 사례: 세종학당 한국어 회화학습 초급 앱



발음 연습



컴퓨터 기반 한국어 발음훈련 사례: 세종학당 한국어 회화학습 초급 앱

Pronunciation of "도와줄 수 있어요"

The pronunciation of "도와줄 수 있어요" is [도와줄쑤임써요]. "수" of "-(으)ㄴ 수 있어요" is pronounced as [쑤].

Pronunciation of "발음 연습"

The pronunciation of "발음 연습" is [바름년습]. If there is a final consonant "ㄱ, ㄴ, or ㅇ" in a previous word, and if the following word begins with "이, 야, 여, 요, or 유," the [ㄴ] is inserted in between to pronounce.

Got it!

Sure. What can I help you with?

발음

[바름] pronunciation

This is the word you are learning. See the pronunciation, meaning, and graphic aid.

발음

Practice pronouncing the word by listening, recording, playback, and comparing.

Accuracy: Speed:

연습

[연습] practice

Accuracy: Speed:

연습

[연습] practice

Accuracy: 62% Speed: 79%

컴퓨터 기반 말하기 자동 평가 사례: ETS's SpeechRater

■ ETS SpeechRater

- SpeechRater Version 1 (2006)
 - TOEFP Practice Online 자동 채점
- SpeechRater Version 5 (2019)
 - TOEFL iBT 채점 투입 시작
 - Combination of human and machine scores

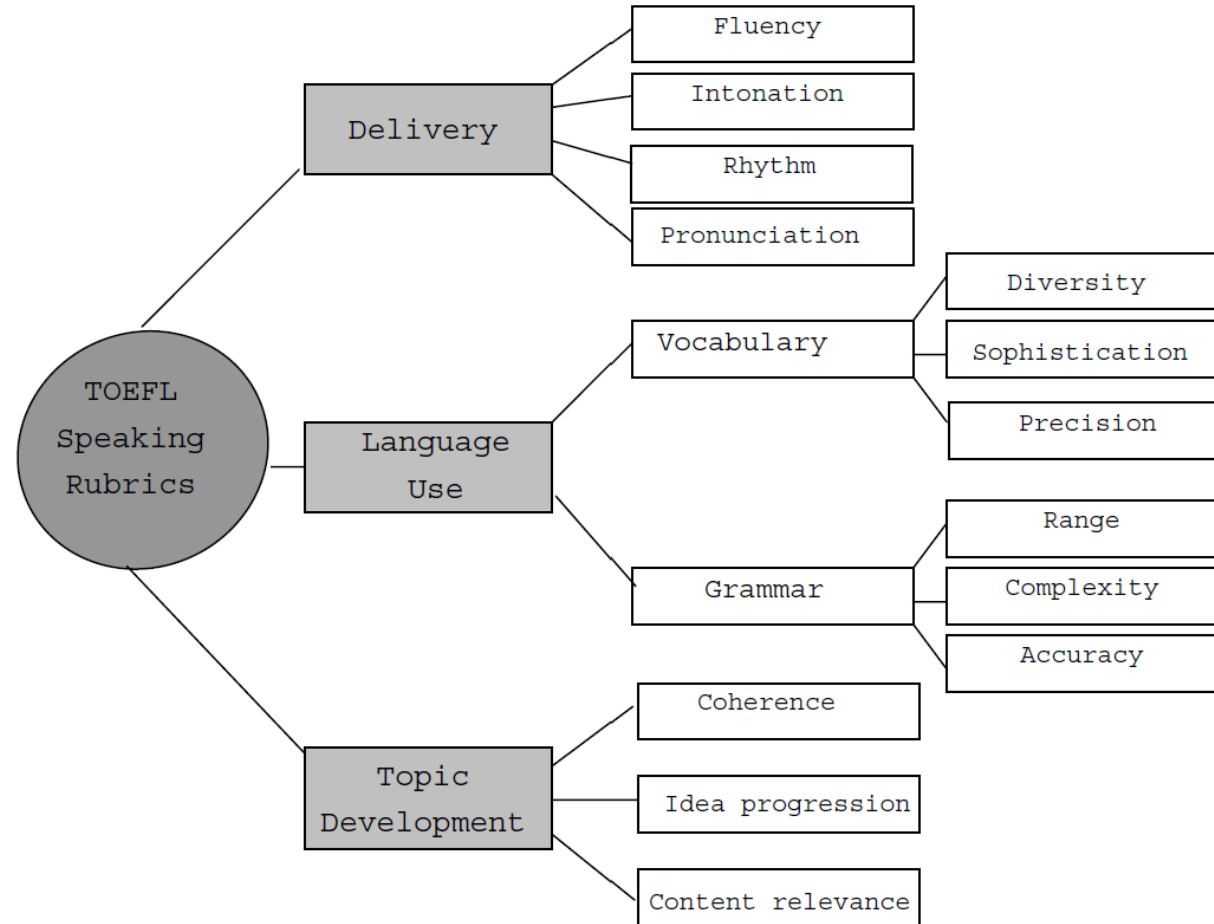


Figure 4. The construct of speech for the TOEFL Internet-based test represented by the scoring rubric.

컴퓨터 기반 말하기 자동 평가 사례: ETS's SpeechRater

■ Speech Proficiency 자동 평가에 사용된 특징

Table 2: *Speaking Proficiency Features Extracted by SpeechRater*

Category	Sub-category	# of Features	Example Features
Prosody	Fluency	24	This category includes features based on the number of words per second, number of words per chunk, number of silences, average duration of silences, frequency of long pauses (≥ 0.5 sec.), number of filled pauses (<i>uh</i> and <i>um</i>). See [14] for detailed descriptions of these features.
	Intonation & Stress	11	This category includes basic descriptive statistics (mean, minimum, maximum, range, standard deviation) for the pitch and power measurements for the utterance.
	Rhythm	26	This category includes features based on the distribution of prosodic events (prominences and boundary tones) in an utterance as detected by a statistical classifier (overall percentages of prosodic events, mean distance between events, mean deviation of distance between events) [14] as well as features based on the distribution of vowel, consonant, and syllable durations (overall percentages, standard deviation, and Pairwise Variability Index) [15].
Pronunciation	Likelihood-based	8	This category includes features based on the acoustic model likelihood scores generated during forced alignment with a native speaker acoustic model [16].
	Confidence-based	2	This category includes two features based on the ASR confidence score: the average word-level confidence score and the time-weighted average word-level confidence score [17].
	Duration	1	This category includes a feature that measures the average difference between the vowel durations in the utterance and vowel-specific means based on a corpus of native speech [16].
Grammar	Location of Disfluencies	6	This category includes features based on the frequency of between-clause silences and edit disfluencies compared to within-clause silences and edit disfluencies [18],[19].
Audio Quality	—	2	This category includes two scores based on MFCC features that assess the probability that the audio file has audio quality problems or does not contain speech input [20].

컴퓨터 기반 말하기 자동 평가 사례: ETS's SpeechRater

- 피드백 생성에 사용된 특징

Feature name	Construct area	Description
Speaking Rate	Delivery-Fluency	Words per second
Sustained Speech	Delivery-Fluency	Number of words without disfluencies
Pause Frequency	Delivery-Fluency	Pauses per word
Repetitions	Delivery-Fluency	Number of repetitions
Vowels	Delivery-Pronunciation	Vowel sounds compared to a native speaker model
Rhythm	Delivery-Pronunciation	Stressed syllables
Vocabulary depth	Language Use-Vocabulary	Use of infrequent words

Research Trends on CAPT

- **Pronunciation Scoring** and **Mispronunciation Detection and Diagnosis (MDD)** is an indispensable component of the CAPT system, as it provides an instant feedback to the users.
- Traditional methods for automatic pronunciation assessment
 - Based on Automatic Speech Recognition (ASR)
 - Features extracted from the Hidden Markov Models (HMMs) of ASR system
 - HMM likelihood, posterior probability, pronunciation duration features
 - a variation of the posterior probability, or Goodness of Pronunciation (GOP)
 - GOP optimized based on Deep Neural Networks (DNNs)
 - Accuracy of the assessment depends on these manually-engineered features.

Previous Studies on CAPT

■ **GOP (Goodness of Pronunciation)**

- GOP is the most famous method among traditional scoring methods
- $P(p|o^p)$: posterior probability of phoneme p given pronunciation o
- $NF(p)$: the number of pronunciation frames of phoneme p
- Phonemes of an utterance are forced-aligned using Kaldi ASR system

$$GOP(p) = \frac{|\log(P(p|o^p))|}{NF(p)} = \frac{|\log(\frac{P(o^p|p)P(p)}{\sum_{q \in Q} P(o^q|q)P(q)})|}{NF(p)}$$

Previous Studies on CAPT

■ **GOP (Goodness of Pronunciation)**

- Steps of GOP:
 - processes forced alignment b/t canonical phone sequence (p) and audio signal (o)
 - computes the likelihood of $P(p|o^p)$ given the alignment information
 - classifies the phoneme as a mispronunciation if the likelihood does not exceed a pre-defined threshold
 - GOP of a phoneme is normalized by duration (frames)
- GOP-like phoneme posteriorgrams are mainly based on DNN-HMM
- Disadvantages
 - No thorough mispronunciation diagnosis
 - Manual and discretized steps

Previous Studies on CAPT

■ Extended Recognition Network

- utilizes phonological rules to derive mispronunciation patterns → formulate ERN
- drawbacks
 - no guarantee whether **all mispronunciation possibilities** from all language learners are covered
 - overly bushy recognition network → lower performance on AM models

■ Solutions

- All possible alternative phones are considered (= all possible mispronunciations are covered)
- State-level Acoustic Model (S-AM)
- Acoustic-Phonemic Model (APM)
- Acoustic-Graphemic Model (AGM)
- Acoustic-Phonemic-Graphemic Model (APGM)

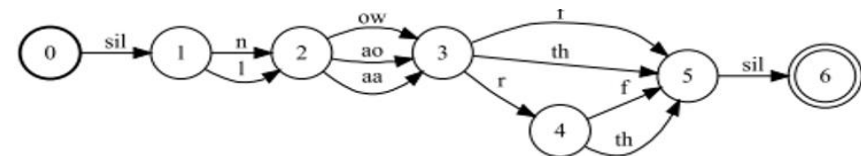


Figure 3: *Extended recognition network of "north"*

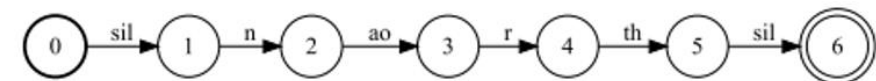
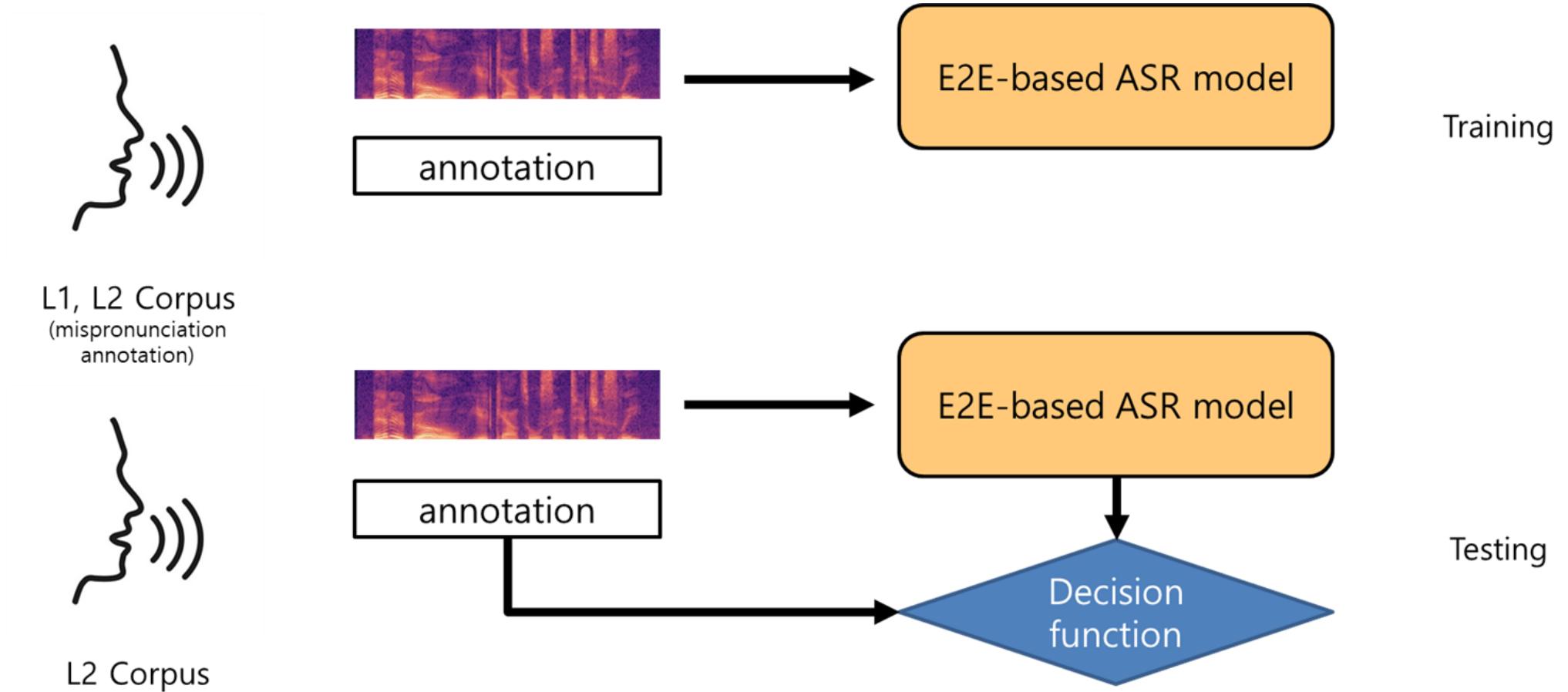


Figure 2: *Standard recognition network of "north"*

Recent Studies on CAPT

- E2E-based CAPT System



Recent Studies on MDD for CAPT

Title	Conferences	Model	Features	Dataset & Performance
CNN-RNN-CTC based E2E MDD	ICASSP 2019	CNN-RNN-CTC	Spectrogram	CU-CHLOE - PER 12.07% - F1 74.62%
SED-MDD: Towards sentence dependent E2E MDD	ICASSP 2020	CNN-RNN + Attention-GRU s equential labeling	Character embeddings Mel-spectrogram	L2-ARCTIC - PER 13.65%
E2E MDD for L2 English Speech Leveraging Novel Anti-Phone Modeling	Interspeech 2020	Hybrid CTC-Attention	Mel-filter-bank	L2-ARCTIC - F1 56.02%
Transformer based E2E MDD	Interspeech 2021	wav2vec 2.0	Raw wave input	CU-CHLOE - PER 5.97% - F1 80.98%
Exploring non-autoregressive E2E neural modeling for English MDD	ICASSP 2022	ConformerEncoder-CTC- TransformerDecoder	Mel-filter-bank	L2-ARCTIC - PER 22.6% - F1 67.19%
Approach to MDD with acoustic, phonetic and linguistic (APL) embeddings	ICASSP 2022	ASR-based pretraining + MD D APL fine-tuning CNN-RNN-Attention Encoder -Decoder	Mel-filter-bank Phonetic embeddings	L2-ARCTIC - PER 16.96% - F1 53.62%
Phoneme MD by jointly learning to align	ICASSP 2022	wav2vec 2.0 acoustic encoder acoustic-phonemic attention + CNN Multitask Learning (MTL)	Raw wave input Canonical phonemes	L2-ARCTIC - F1 63.04%

Recent Studies on CAPT

- Dataset
 - L2-ARCTIC is the most used (TIMIT dataset is employed for L1 fine-tuning) L2 dataset
 - For Chinese-native's L2 English utterances, CU-CHLOE is chosen
- Acoustic Features
 - Spectrogram-based, FBANK-based acoustic features are given as input
 - In case of using wav2vec 2.0 as AM, raw waveform is used as input
- Model Evaluation Metrics
 - Phone Error Rate
 - F1 score (with Precision and Recall score provided as well)

Recent Studies on CAPT

■ Models

- Development of neural network led to the era of end-to-end (E2E) ASR model for MDD tasks.
- Recent studies utilize ASR paradigms such as
 - Connectionist Temporal Classification (CTC) alignment (*Leung et al, 2019; CNN-RNN-CTC BASED END-TO-END MISPRONUNCIATION DETECTION AND DIAGNOSIS*)
 - Attention-based (ATT) method (*Lin et al, 2022; PHONEME MISPRONUNCIATION DETECTION BY JOINTLY LEARNING TO ALIGN*)
 - A hybrid CTC-ATT method (*Zhang et al, 2020; End-to-End Automatic Pronunciation Error Detection Based on Improved Hybrid CTC/Attention Architecture*)
 - Pretrained models are used with fine-tuning to bring better performance (*Baevski et al, 2020; wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations*)

Recent Studies on CAPT

- The Disadvantages of E2E Models
 - As the task relies on ASR performance, it follows the disadvantages of E2E models.
 - When training with custom ASR models, mispronunciation labels should be included in the dictionary beforehand.
 - Slow inference speed due to the autoregressive manner of deep neural models.
 - Although named 'end-to-end', to get high-level results, manually engineered features are sometimes still needed.
 - Lack of task related data: a serious lack of annotated L2 data for deep learning training.

Recent Studies on CAPT: Other Issues

- The impact of forced alignment in L2 speech recognition
 - GOP score (the most common pronunciation assessment algorithm) is computed using phoneme level posterior probabilities estimated using GMM-HMM or DNN-HMM-based acoustic model.
 - While computing the GOP, the target phoneme boundaries are located using a forced-alignment algorithm. Other (classification-based) pronunciation assessment systems similarly rely on forced alignment.
 - For both systems, forced-alignment errors can have downstream consequences.
 - Susceptibility to error particularly increases when used on atypical speech due to acoustic mismatch between the utterance and the acoustic model.

Recent Studies on CAPT: Other Issues

- ASR vs. MDD
 - MDD needs to catch all the speech variation as is, whereas ASR needs to map all the speech variation to the canonical phoneme (the goal discordance between MDD and ASR).
 - Wav2vec2-base-960h produces **higher precision but lower recall** than wav2vec2-base in *Yang et al, 2022* → more tolerant judge by rejecting less L2 pronunciations; makes sense given the goal of ASR
 - The same result appears in *Ye et al, 2022*, with larger model having higher precision but lower recall → with a deeper model trained on a larger dataset, the model is noise-tolerant and speaker-normalized **but may lose some useful information for MDD**.
 - **Over-robustness** of the ASR pretrained features may not be desirable for MDD tasks!
 - How we can overcome such tendency is another key issue.

컴퓨터 기반 한국어 발음훈련을 위한 연구자원 소개

- 데이터

- AI Hub 데이터: 인공지능 학습을 위한 외국인 한국어 발화 음성
- AI Hub 데이터: 언어교육용 서양어, 아시아어 사용자의 한국어 음성 데이터

- 소프트웨어 툴

- Montreal Forced Aligner



AI Hub 데이터 (한국지능정보사회진흥원)

<https://aihub.or.kr>

AI 허브는 AI 기술 및 제품·서비스 개발에 필요한 AI 인프라(AI 데이터, AI SW API, 컴퓨팅 자원)를 지원함으로써 누구나 활용하고 참여하는 AI 통합 플랫폼입니다.

AI 허브의 사용자를 위해 개발 및 활용을 위한 인프라 서비스와 AI 활성화를 위한 서비스를 제공하고 있습니다.



AI Hub

AI 데이터찾기

AI 개발지원

참여하기

정보공유

고객지원

AI 허브소개

로그인 회원가입



한국어
데이터 93종

영상이미지
데이터 78종

헬스케어
데이터 67종

재난안전환경
데이터 59종

농축수산
데이터 41종

교통물류
데이터 46종

AI Hub 데이터: 인공지능 학습을 위한 외국인 한국어 발화 음성

[AI 데이터찾기](#)[AI 개발자원](#)[참여하기](#)[정보공유](#)[고객지원](#)[AI 허브소개](#)[마이페이지](#)[로그아웃](#)

데이터 분야

[AI 데이터찾기](#) > 데이터 분야

#한국어 음성 인식

#외국인

#이주민

#한국어

#한국어 학습

#음성 전사

인공지능 학습을 위한 외국인 한국어 발화 음성

분야 한국어

유형 오디오

갱신년월 : 2022-07

구축년도 : 2021

조회수 : 496

다운로드 : 22

용량 : 379.34 GB



관심데이터 등록

2

수행기관(주관) : (주) 씨에스리

책임자명	전화번호	대표이메일	담당업무
윤희우	070-4756-4580	hwyoona@cslee.co.kr	· 사업관리 및 전체 공정 관리

수행기관(참여)

기관명	담당업무
세종대학교	· 데이터 설계 및 수집 관리
이화여자대학교	· 데이터 설계 및 수집 관리
드림비트	· 데이터 정제 및 가공
디그랩	· 데이터 가공
액션파워	· 학습모델 구현

<https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=505>

AI Hub 데이터: 인공지능 학습을 위한 외국인 한국어 발화 음성

2021년도 구축, 2022.07 공개

데이터 종류	데이터 형태	원천데이터 규모			메타 데이터 규모	
		사이즈(GB)	시간(h)	건수	사이즈(MB)	건수
베트남어	녹음음성 및 전사 결과	82.07	764.93	194,082	343.25	388,164
영어		13.30	123.90	31,554	55.51	63,108
일본어		81.55	760.08	216,902	379.69	433,804
중국어		114.23	1,064.66	292,593	516.53	585,186
태국어		57.70	537.82	139,692	244.89	279,384
기타		112.79	1,051.00	284,842	501.98	569,684
총계		461.64	4,302.39	1,159,665	2,041.85	2,319,330

AI Hub 데이터: 인공지능 학습을 위한 외국인 한국어 발화 음성

언어분류별 분포

언어분류	녹음 시간	비율(%)
베트남어	764.93	17.78
영어	123.90	2.88
일본어	760.08	17.67
중국어	1,064.66	24.75
태국어	537.82	12.50
기타	1,051.00	24.43
총계	4,302.39	100

세트별 분포

세트	녹음 시간	비율(%)
한국일반	947.73	22.03
한국생활I	836.76	19.45
한국생활II	895.77	20.82
한국문화I	827.47	19.23
한국문화II	794.67	18.47
총계	4,302.39	100

유형별 분포

세트	녹음 시간	비율(%)
대본 읽기	3,101.09	72.08
질문에 답변하기	1,201.30	27.92
총계	4,302.39	100

AI Hub DB: 언어교육용 서양어, 아시아어 사용자의 한국어 음성 데이터

- 2022년 현재 구축중



미국, 영국, 호주, 캐나다 등 영어권 외국인 발화 음성 1,000시간 이상

1

영어 모국어 사용자 한국어 음성 데이터 구축



스페인어, 프랑스어, 독일어, 러시아어 모국어 사용자 발화 음성 1,000시간 이상

2

유럽어 모국어 사용자 한국어 음성 데이터 구축



중국어, 일본어 모국어 사용자 발화 음성 1,000시간 이상

3

중·일어 모국어 사용자 한국어 음성 데이터 구축



베트남어, 인도네시아어, 태국어, 인도어 모국어 사용자 발화 음성 1,000시간 이상

4

아시아어 모국어 사용자 한국어 음성 데이터 구축

AI Hub DB: 언어교육용 서양어, 아시아어 사용자의 한국어 음성 데이터

- 2022년 현재 구축중



Data			평가 루브릭 5단계(1~5)		
번호	과업	개수	발음정확성 Accentedness	유창성 Fluency	이해가능도 Comprehensibility
1	단어 읽기	180	O	-	O
2	문장 읽기	88	O	O	O
3	문단 읽기	7	O	O	O
4	이야기 읽기	2	O	O	O



Data			평가 루브릭 6단계(0~5)			
번호	과업	개수	내용 Content	언어사용의 정확성과 적합 성	전달력 Delivery	척도 Scale
5	문장 듣고 따라하기	20	O(0,1,3,5)	O	O	O(1~5)
6	질문에 답하기	20	O	O	O	O(1~5)
7	개인적인 주제에 관해 말하기(단답형 10, 열린 질문 10)	20	O	O	O	O(1~5)
8	언어기능에 맞게 말하기(거절하기 3, 요청하기 3, 조언하기 3)	9	O	O	O	O(1~5)
9	그림과 그래프 설명하기	6	O	O	O	O(1~5)
10	의견 말하기	3	O	O	O	O(1~5)

AI Hub DB: 언어교육용 서양어, 아시아어 사용자의 한국어 음성 데이터

■ 발음평가: 단어 읽기

- 학습자의 모어/제1언어와 수준을 고려하여 단어 선정

no.	언어권	숙달도	범주	세부	단어	오류양상	비고
1	영어권	초급	음운현상	비음화	작년에	<u>작.년에</u>	미적용
2	영어권	초급	음운현상	비음화	박물관	<u>박.물관</u>	미적용
3	영어권	중급	음운현상	비음화	기억나는	<u>기억.나는</u>	미적용
4	영어권	중급	음운현상	비음화	옛날	<u>옴날</u>	미적용
5	영어권	중급	음운현상	비음화	빛난다	<u>빈.난다</u>	미적용
6	영어권	중급	음운현상	비음화	집집마다	<u>지프마다</u>	미적용
7	영어권	중급	음운현상	비음화	즐겁네요	<u>즐겁.네요</u>	미적용
8	영어권	초급	초성	파열음	바다	<u>파다</u>	ㅂ-ㅍ
9	영어권	초급	초성	파찰음	주문해요	<u>추문해요</u>	ㅈ-ㅊ
10	영어권	고급	음운현상	연음	법원	<u>버프원</u>	연음
11	영어권	중급	초성	마찰음	설명	<u>썰명</u>	ㅅ-ㅆ
12	영어권	초급	음운현상	탈락	걸려요	<u>거려요</u>	ㄹ탈락
13	영어권	중급	음운현상	격음화	만형	<u>만.형</u>	미적용
14	영어권	고급	중성	미파	충격	<u>충겨크</u>	파열
15	영어권	고급	음운현상	비음화	맥락	<u>맥.라크</u>	미적용, 파열 2개
16	영어권	고급	중성	단모음	거론	<u>고론</u>	ㄱ-ㄴ
17	영어권	고급	음운현상	경음화	폭발	<u>포크발, 포크팔</u>	미적용, ㅂ-ㅍ

■ 발음평가: 문장 읽기

번호	숙달도	범주	평가 문장	비고
1	초급	연음, 비음화	가족이 몇 명이에요?	
2	초급	연음, ㄴ 첨가	우리 집 옆 큰 시장에 구경을 자주 가요.	아랍어, 일본
3	초급	유기음화	제 동생은 수영을 잘하지만 저는 잘 못해요.	
4	초급	비음화, 경음화	십만 원만 빌려줄 수 있어요?	
5	초급	경음화, 연음, ㅎ 탈락	학교 근처 식당에는 항상 사람이 많아요.	일본

51	중·고급	경음화, 연음, 억양	내가 먹던 라면이 어디로 갔지?	
52	중·고급	유기음화, 연음, 비음화, 억양	며칠 앓더니 살이 쏙 빠졌네요.	중국
53	중·고급	경음화, 유음의 비음화, 연음	요즘 갑자기 기억력이 나빠져서 자꾸 뭘 잊어버려요.	중국
54	중·고급	비음화 경음화, 연음	실컷 먹고 잤더니 얼굴이 퉁퉁 부었어요.	동남아
55	중·고급	ㅎ 탈락, 연음 유음의 비음화, 비음화	좋은 의견이 있으시면 의견란에 꼭 써 주시기 바랍니다.	

AI Hub DB: 언어교육용 서양어, 아시아어 사용자의 한국어 음성 데이터


■ 발음평가: 단락 읽기

수준	지시문
초급	안녕하세요, 여러분. 저는 줄리앙이라고 합니다. 프랑스에서 왔습니다. 한국 문화와 한국 음식 그리고 한국 역사에 관심이 있어서 한국어를 혼자 조금 공부했습니다. 앞으로 1년 동안 한국어도 배우고 한국의 여러 곳에 다녀 보려고 합니다. 아직 한국어를 <u>잘 못하고</u> 한국 친구도 <u>없지만</u> 한국 생활이 재미있을 것 같습니다. 여러분과 같이 공부하게 되어 정말 기쁩니다.
중고급	지난여름 저는 대학 동창들과 함께 부산 여행을 갔습니다. 각각 다른 직장에 다니는 우리는 일 년 전부터 여행 계획을 짜고 <u>준비를</u> 했습니다. 그런데 여행 첫날 태풍이 올라오는 바람에 우리는 숙소 밖으로 한 발자국도 나갈 수가 없었습니다. “우리가 어떻게 준비한 여행인데...” 하늘이 원망스러웠습니다. 다행히도 하루가 지나고 나니 언제 그랬냐는 듯이 하늘은 맑게 개고 바람도 잔잔해졌습니다.

■ 발음평가: 이야기 읽기


수준	지시문
초급	오늘은 한국어 수업이 있어요. 수업이 오전 9시에 시작해요. 집에서 학교까지 30분이 걸려요. 7시에 일어나서 아침을 먹었어요. 8시에 집에서 나왔어요. <u>지하철 역에</u> 도착했어요. 그런데 지갑이 없었어요. 다시 집에 갔어요. 집에도 지갑이 없었어요. 지갑이 가방에 있었어요. 지하철역까지 뛰어서 갔어요. 학교에 8시 55분에 도착했어요. 수업에 안 늦었어요. 그렇지만 정말 힘들었어요.
중고급	나는 아름다운 색을 사랑한다. 예전 우리 유치원 선생님이 주신 색종이 같은 빨간색, 보라색, 자주색, 녹색, 이런 색깔을 나는 좋아한다. 나는 우리나라 가을 하늘을 사랑한다. 나는 오래된 가구의 색을 좋아한다. 늙어가는 학자의 희끗희끗한 머리카락을 좋아한다. 나는 이른 아침의 새 소리를 좋아하며, 봄 시냇물 흐르는 소리를 즐긴다. 갈대에 부는 바람 소리를 좋아하며, 바다의 파도 소리를 들으면 아직도 가슴이 뒹다. 나는 골목을 지나갈 때 발을 멈추고 한참이나 서 있게 하는 피아노 소리를 좋아한다. 나는 젊은 웃음소리를 좋아한다. 다른 사람 없는 방 안에서 내 귀에다 귓속말하는 내 딸 서영이의 말소리를 좋아한다. 나는 비 오는 날 <u>저녁</u> 때 뒷골목 술집에서 나는 불고기 냄새를 좋아한다. 새로운 책 냄새, 커피 끓이는 냄새를 좋아한다. 봄 <u>흙</u> 냄새를 좋아하며 친구와 향기로운 차 마시기를 좋아한다.

소프트웨어 툴: Montreal Forced Aligner

 Montreal Forced Aligner

Getting started User guide API reference Changelog Pretrained MFA models


Search the docs ...



Digital Ocean: Create your world-changing apps on the cloud developers love **Try now with a \$100 Credit**

Ad by EthicalAds · Host these ads


Montreal Forced Aligner documentation



Getting started

Install the Montreal Forced Aligner and get started with examples and tutorials.


Install MFA



User guide

The User Guide gives more details on input formats, available commands, and details on the various workflows available.

User guide




First steps

Have a particular use case for MFA?

Check out the first steps tutorials.

First steps



API reference


The API guide lists all the inner workings of MFA, the modules and classes that you can import and use in your own scripts and projects, along with details about the Kaldi functionality used.

Reference guide

<https://montreal-forced-aligner.readthedocs.io/en/latest/>


소프트웨어 툴: Montreal Forced Aligner

Montreal Forced Aligner Models




Dictionaries
Pronunciation dictionaries for use with MFA

[Browse dictionaries](#)




Acoustic models
Pretrained acoustic models trained on ASR corpora

[Browse acoustic models](#)



Grapheme-to-phoneme models
G2P models can supplement dictionaries with new pronunciations

[Browse G2P models](#)



Language models
Language models alongside the acoustic models

[Browse language models](#)

[Next >](#)
Pronunciation dictionaries

<https://mfa-models.readthedocs.io/en/latest/>

소프트웨어 툴: Montreal Forced Aligner for Korean

■ Dictionaries

Korean

Show 10 ▼ entries Columns Copy Excel PDF Search:

ID ▲	Language ◆	Dialect ◆	Phoneset ◆	License ◆
Korean (Jamo) MFA dictionary v2_0_0	Korean	Jamo	MFA	CC BY 4.0
Korean MFA dictionary v2_0_0	Korean	N/A	MFA	CC BY 4.0
Korean MFA dictionary v2_0_0a	Korean	N/A	MFA	CC BY 4.0

Showing 1 to 3 of 3 entries Previous 1 Next

■ G2P Models

Korean

Show 10 ▼ entries Columns Copy Excel PDF Search:

ID ▲	Language ◆	Dialect ◆	Phoneset ◆	License ◆
Korean (Jamo) MFA G2P model v2_0_0	Korean	Jamo	MFA	CC BY 4.0
Korean (Jamo) MFA G2P model v2_0_0a	Korean	Jamo	MFA	CC BY 4.0
Korean MFA G2P model v2_0_0	Korean	N/A	MFA	CC BY 4.0
Korean MFA G2P model v2_0_0a	Korean	N/A	MFA	CC BY 4.0

Showing 1 to 4 of 4 entries Previous 1 Next

소프트웨어 툴: Montreal Forced Aligner for Korean

■ Acoustic Models

Korean

Show entries Search:

ID	Language	Dialect	Phonset	License
Korean MFA acoustic model v2_0_0	Korean	N/A	MFA	CC BY 4.0
Korean MFA acoustic model v2_0_0a	Korean	N/A	MFA	CC BY 4.0

Showing 1 to 2 of 2 entries Previous Next

■ Language Models

Korean

Show entries Search:

ID	Language	Dialect	License
Korean language model v2_0_0a	Korean	N/A	CC BY 4.0

Showing 1 to 1 of 1 entries Previous Next

소프트웨어 툴: Montreal Forced Aligner for Korean

■ Benchmarks

Korean alignment benchmarks

Dataset

The dataset used for this benchmark is the [Seoul Corpus](#). The Seoul Corpus was modeled off of the [Buckeye Corpus](#) to create a phonetically/phonemically hand-aligned corpus of Seoul Korean. The corpus consists of 40 speakers of Seoul Korean with 20 male speakers and 20 female speakers, along with 10 speakers each in their teens, twenties, thirties and forties. Similar to the Buckeye Corpus, socio-economic class was also not controlled, but the setting of academic sociolinguistic interviews will bias towards middle to upper class.

The corpus was transcribed in Hangul and aligned in HTK, and then corrected by hand. The transcription is more phonemic than the Buckeye Corpus's phone set (though, even the final Buckeye phone set is not as as phonetic as the original TIMIT-based set they used).

The dataset is freely available on [OpenSLR](#). The [reorganization script here](#) is the basis of the testing data, and creates input TextGrids to align and reference textgrids to compare against in [the alignment evaluation script](#), along with the necessary mapping files to the Seoul Corpus phone set from [MFA's phone set](#) and [GlobalPhone's phone set](#).

■ Corpora

Korean

Show 10 entries

Columns

Copy

Excel

PDF

Search:

ID	Language	Dialect	License
Deeply Korean read speech corpus public sample	Korean	N/A	CC BY-NC-ND 4.0
GlobalPhone Korean v3_1	Korean	N/A	ELRA
Pansori TEDxKR	Korean	N/A	CC BY-NC-ND 4.0
Seoul Corpus	Korean	N/A	CC BY-NC 2.0
Zeroth Korean	Korean	N/A	CC BY 4.0

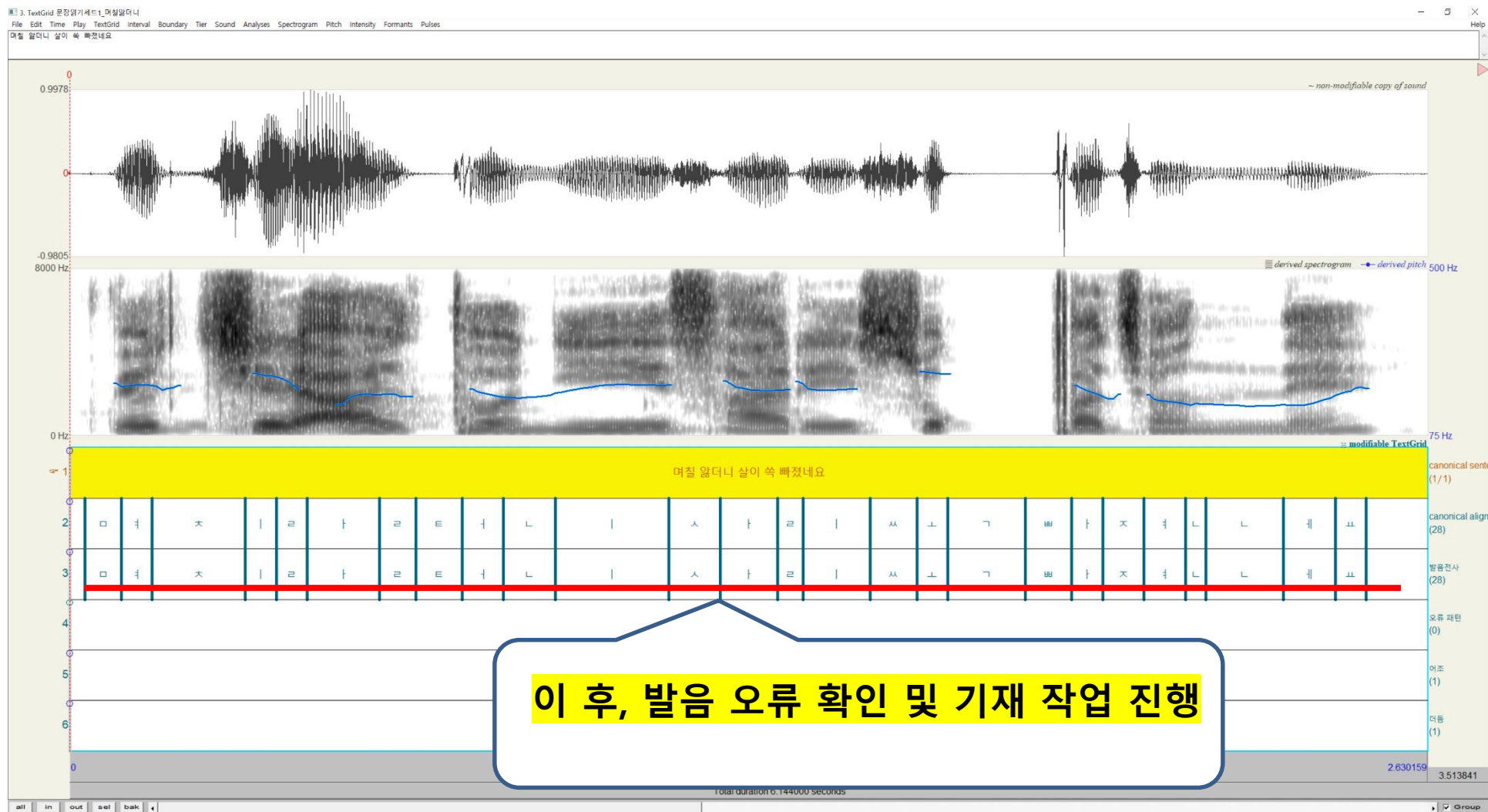
Showing 1 to 5 of 5 entries

Previous

1

Next

소프트웨어 툴: Montreal Forced Aligner for Korean



결론

- 컴퓨터 기반 한국어 발음 훈련 솔루션 개발을 위해서 필요한
 - 기본 개념, 연구동향, 이용 가능한 데이터와 소프트웨어 툴 소개
- CAPT는 언어학과 공학의 학제적 연구 분야
 - 이론 및 응용 언어학이 주 역할을 하며, 기술은 보조 역할을 함
- 모국어 특성에 따른 외국인 한국어 학습자의 한국어 발음 특성 연구와 이를 바탕으로 CAPT에 적용할 수 있는 계산 모델링 연구가 시급함.

감사합니다!

질의 · 응답