인공지능 데이터 구축·활용 가이드라인
- 데이터 11 고객 응대 데이터 -

인공지능 데이터 구축	사업 총괄	i-Screamedu
	데이터 설계	i-Screamedu
	원천데이터 수집 및 정제	i-Screamedu
	데이터 가공	i-Screamedu edutech KOREA
	데이터 검수	namu i-Screamedu edutech KOREA
	크라우드 소싱	i-Screamedu edutech KOREA
	저작도구 개발	TmaxSoft
	AI모델 개발	TmaxSoft
	응용 서비스 개발	TmaxSoft
가이드라인 작성	아이스크림에듀	김 종 범
가이드라인 버전	버전 1.3.0 작성일 2021.01.20	

목 차

1. 데이터 명세 정보

1.1 데이터 정보 요약

데이터 이름	고객응대 데이터	
활용 분야	• 연구분야: 음성인식, 음성언어처리, 자연어처리, 한국어 음성언어연구, 신호처리 등 • 산업분야: 음성챗봇, 비접촉 음성 AI 키오스크, 비접촉 음성 AI 사이니지 등	
데이터 요약	다양한 도메인에서 주문, 예약, 환불, 정보조회 등의 음성인식으로 서비스에 활용될 수 있는 음성인식 학습용 3,000 시간의 남녀 1:1 비율의 고객응대 음성 데이터셋	
데이터 출처	가상 시나리오에 기반한 크라우드소싱 녹취 데이터	
데이터 이력	배포버전	2.0.0
	개정이력	1.0.0: 신규 1.1.0: 전문가 의견서 반영 2.0.0: TTA 검토 의견 반영 (양식 목차 준수 포함)
	작성자/ 배포자	작성자 : 김 종 범 배포자 : 아이스크림에듀

1.2 데이터 포맷

1.2.1 데이터셋 계층 구조

- 계층1: 데이터셋 – 고객응대 (KrespSpeech)
- 계층2: 카테고리 – 카테고리(domain)을 나타내는 알파벳과 숫자 2자리로 구성된 카테고리 번호
- 계층3: 서브카테고리 – 서브카테고리를 나타내는 영어 알파벳(J)과 숫자 2자리로 구성된 서브카테고리 번호
- 계층4: 원천음원 일련번호: 일련번호(Serial number)를 나타내는 알파벳과 숫자 8자리로 구성된 원천음성파일 일련번호 - S00000001
- 계층5: 음원 파일, 텍스트 파일, 메타데이터 파일
  - 음원 파일 – 발화 단위 RAW 데이터 (WAV)
  - 텍스트 파일 – 발화 단위 TEXT 데이터 (TXT)
  - 메타데이터 파일 : 화자 정보, 원천 음원을 가늠할 수 있는 카테고리 및 인입 유형, 해당 원천으로부터 추출된 발화 단위의 음원, 텍스트 파일 경로 등을 포함하는 메타데이터 (JSON)

1.2.2. 데이터셋 계층 구조 예시

계층	1단계	2단계	3단계	4단계	5단계
	KrespSpeech	D50	J01	S00000017	0001.wav
					0001.txt
					0002.wav
					0002.txt
					0003.wav
					0003.txt
					…
					S00000017.json
		D50	J03	S00000173	0001.wav
					0001.txt
					…
					S00000173.json

1.3 어노테이션 포맷 - 메타데이터

1.3.1. 메타데이터 포맷

ID	항목		타입
	키 명	키 설명	
	dataSet	데이터셋	Dict
1	version	데이터셋 버전	String
3	date	녹취된 날짜	String
4	typeInfo	음원 데이터 상세 정보	Dict
4-1	category	음원 카테고리 정보	String
4-2	subcategory	음원 서브카테고리	String
4-3	place	음원 녹취 장소 : 녹취 장소 불명확시엔, null로 설정	String
4-4	speakers	화자 목록	List
4-4-1	id	화자 아이디	String
4-4-2	type	화자 유형: 고객, 상담원	String
4-4-3	age	나이대: 20대, 30대, 50(추정), null(알수없음) 등	String
4-4-4	gender	화자 성별: 남, 여	String
4-4-5	residence	거주지역: 서울, 대전, 부산, 광주, null(알수없음) 등	String
4-5	inputType	입력형식: 방송, 유선, 모바일, 인터넷 등	String
5	dialogs	전사 데이터 목록: 묵음 기준으로 나누어진 발화 단위로 생성	List
5-1	speaker	화자 아이디: speakers에 등록된 id	String
5-2	audioPath	발화 단위 RAW 데이터 경로	String
5-3	textPath	발화 단위 TEXT 데이터 경로	String

1.3.2. 샘플

메타데이터 파일 (KrespSpeech/D50/J01/S000017/S000017.json)

```
{
        "dataSet": {
                "version": "1.0",
                "date": "2020",
                "typeInfo": {
                        "category": "구매",
                        "subcategory": "카페",
                        "place": "cafe",
                        "speakers": [
                                {
                                        "id": "1",
                                        "type": "고객",
                                        "age": "20대",
                                        "gender": "여",
                                        "residence": null
                                }

                        ],
                        "inputType": "mobile"
                },
                "dialogs": [
                        {
                                "speaker": "1",
                                "audioPath": "KresSpeech/D50/J01/S00000017/0001.wav",
                                "textPath": "KresSpeech/D50/G01/S00000017/0001.txt",
                        },
                        {
                                "speaker": "2",
                                "audioPath": "KresSpeech/D50/J01/S00000017/0002.wav",
                                "textPath": "KresSpeech/D50/J01/S00000017/0002.txt",
                        },
                        이하생략 ...
                ]
         }
}
```

전사파일 (KrespSpeech/D50/J01/S00000017/0001.txt)

무엇을 주문하시겠습니까?

전사파일 (KrespSpeech/D50/J01/S00000017/0001.txt)

아이스아메리카노요

1.4 데이터 구성

	폴더 경로				파일
명칭	데이터셋	도메인	업종	순번	데이터셋 파일
예시	KrespSpeech	D50	J01	S00000001	S00000001.json 0001.wav 0001.txt 0002.wav 0002.txt ...
설명	R(esponse): 고객응대	D50: 구매 D51: 예약 D52: 생활	J01: 카페 J02: 식당 J03: 의류 J04: 소매 J05: 숙박 J06: 학원 J07: 독서실 J08: 미용실 J09: 여행 J10: 민원 J11: 세탁소 J12: 옷수선 J13: 여가오락 J14: 위치정보	일련번호	

1.5 데이터 통계

1.5.1 데이터 구축 규모

데이터의 종류	수집시간	제공방식
구매 도메인	1,000시간	wav 음원파일 txt 전사파일 json 메타파일
예약 도메인	1,000시간	
생활 도메인	1000시간	
총 구축량	**3,000시간**	

1.5.2 데이터 분포

□ 도메인

구매형	예약형	생활형
1000시간	1000시간	1000시간

□ 성별

남성	여성
1000명 이상	1000명 이상

발화시간 기준의 남녀 비율 1:1 (5% 오차 범위 이내로 데이터셋 구축)

1.6 원시데이터 특성

1.6.1 대상분류

분류	설명
시뮬레이션	• 실 녹취된 음원 데이터에 대한 법률 검토 결과, 개인의 바이오 데이터를 외부에 공개하는 것은 법률적인 이슈 발생으로 판단. • 가상 시나리오 스크립트로 크라우드 소싱 녹취 작업

1.6.2 제약조건

분류	설명
제약있음	• 충분한 스크립트 수 확보 • 하나의 스크립트 당 녹취 가능 인원 수 제한 • 한명 당 최대 녹취 가능 시간 제한 (5시간)

1.6.3 속성

파일 형식	코덱	샘플레이트	비트	채널
WAV	PCM	16KHz	16비트	모노

1.7 기타정보

1.7.1 포괄성

도메인		
구매 도메인	예약 도메인	생활 도메인

1.7.2 독립성

분류	설명
해당사항없음	• 가상 시나리오 스크립트로 녹취 진행 하므로, 민감정보와 법적 문제에서 특별한 이슈가 발생하지 않음.

1.7.3 유의사항

특정 업종에서 활용하기 위해서는 적용 업종의 제품명 같은 고유명사에 대한 추가 데이터 구축이 이루어져야 하며, 다음의 사항에 유의해서 추가 데이터셋을 구축해야 한다.

- 상품명은 문장으로 구성한다.
- 한 문장에 대한 음성 녹취는 최대 50명을 초과 하지 않는다.
- 한명이 녹취할 수 있는 최대 시간은 5시간으로 제한한다.
- 데이터셋으로 구성 시에 본 과제에서 제공하는 전사 규칙을 준수해서 전사 문장을 구성한다.

2. 데이터 구축 가이드

2.1 데이터 구축 개요

2.1.1 고객응대 데이터셋 구축 목적

본 과제의 데이터 구축 목적은 다양한 매장과 공간의 키오스크, 사이니지 등 기존의 터치 UI로 제공되는 기기들을 음성언어로 주문, 검색, 조작할 수 있는 기술/서비스 개발에 활용할 수 있는 음성 데이터셋 구축을 목적으로 한다. 한국인의 음성을 문자로 바꾸어 주고, 문맥을 이해하는 한국어 음성언어처리 기술 개발을 위한 AI 학습용 한국어 음성 DB 구축을 목표로 한다.

고객응대 데이터셋의 구축량은 다음과 같다.

구축 시간	발화자 수	남녀 비율	도메인 수
3,000시간 이상	2000명 이상	1:1	3개

발화 시간 기준의 남녀 비율 1:1 (5% 오차 범위 이내로 데이터셋 구축)

2.1.2 데이터 구축 절차

AI 키오스크 및 음성챗봇 등의 고객응대 서비스를 위한 학습용 데이터는 데이터 수집, 데이터 정제, 데이터 가공, 데이터셋 검수의 4단계를 거쳐 구축된다.

구축 단계	세부 절차	설명
수집	수집 도메인 선정	구축하려고 하는 서비스 업종 등의 적용 분야를 선정함.
	서비스 시나리오 작성	제공하려는 서비스 시나리오를 수립하여 작성함.
	시나리오별 스크립트 작성	하나의 시나리오에서 다양한 경우의 문장들을 작성하고 상품명 같은 엔티티를 다양하게 적용하여 작성함.
	크라우드소싱을 이용한 녹취	실 서비스 적용을 고려하여 현장 소음이 포함되도록 하며, 동일 문장의 녹취 인원이 최대 50명이 넘지 않도록 함.
정제	원천 데이터 검수	스크립트와 다른 녹취, 과도한 소음이 포함, 녹취자의 목소리 너무 낮은 음원 등을 제거하는 과정을 통해 데이터를 정제
가공	가공 인력 교육	인공지능 학습용 데이터 가공에 필요한 작업 교육과 훈련 수행

	데이터 가공	음원의 녹취와 스크립트 상의 불일치 부분에 대해 스크립트 수정과 음원을 문장 단위로 분리 저장.
데이터 검수	음성 모델 학습	구축된 데이터셋의 음원으로 음성모델 생성
	전수 검수	생성된 음성 모델을 통해 자동 전사된 텍스트와 스크립트와 텍스트 비교

2.2 문제정의

- ❍ 수년 전부터, 터치 UI 기반의 키오스크 기기가 카페, 식당, 극장, 쇼핑몰 등 거의 모든 분야에서 사용되어져 왔음.
- ❍ 하지만, 최근의 코로나19의 팬데믹 상황에서는 터치 기반의 기기는 코로나 바이러스의 전염의 가능성으로 더욱 더 비접속식의 서비스 기기의 필요성이 대두되고 있음.
- ❍ 본 과제의 목적은 이러한 시대적인 상황에서 스타트업, 기존의 중소기업 등 큰 비용 투자를 통해 음성인식 모델 데이터 구축이 어려운 업체들에게 다양한 도메인을 위한 음성인식 학습용 데이터셋을 제공하여 빠르게 응용 서비스를 구현할 수 있도록 함.
- ❍ 많은 업체들이 다양한 분야에서 AI 키오스크, AI 사이니지, 음성챗봇을 위한 음성인식 모델을 생성할 수 있도록 다양한 도메인을 위한 데이터셋을 구축한다.

도메인 종류	설명
구매형 도메인	음료/식사/제품구매 등의 메뉴 혹은 제품을 주문/구매하고 결제를 할 수 있는 구매 서비스 도메인 (카페, 식당, 의류, 소매)
예약형 도메인	호텔, 모텔, 캠핑장, 여행상품 등의 예약하거나 취소할 수 있는 예약 서비스 도메인 (숙박, 학원, 독서실, 미용실, 여행)
생활형 도메인	공공 민원이나 교통, 생활 관련한 문의, 조회할 수 있는 생활 서비스 도메인 (민원, 세탁소, 옷수선, 여가오락, 위치정보)

2.3 획득정제

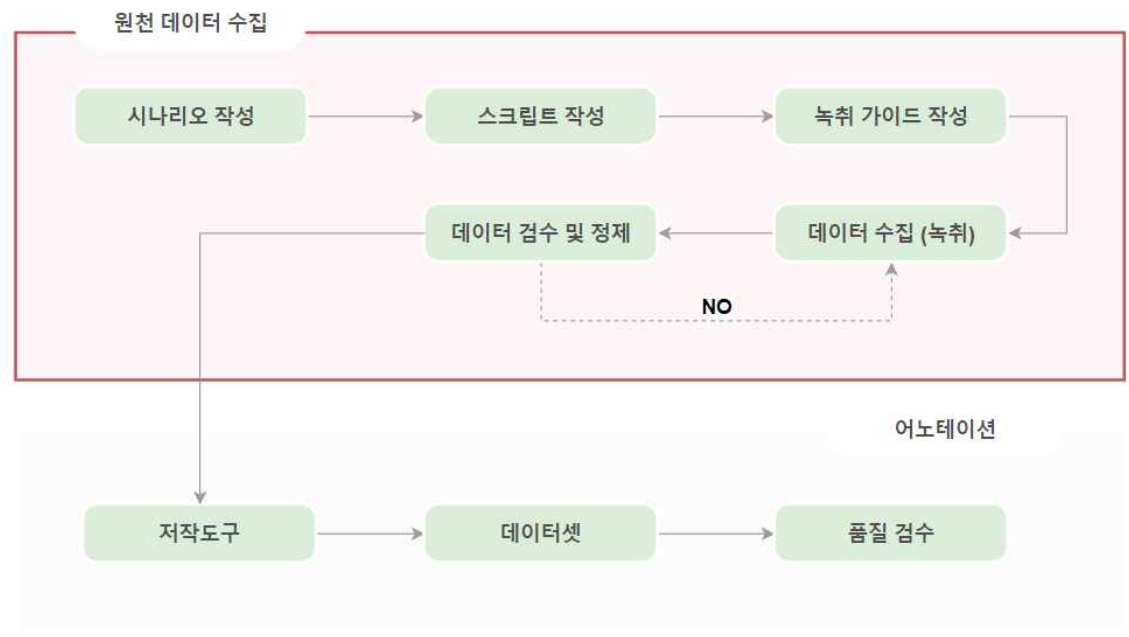
2.3.1 데이터 구축 규모

원천데이터 종류	수집 채널	수집 시간
구매 도메인	크라우드 소싱 녹취 (가상시나리오 기반)	1000시간
예약 도메인	크라우드 소싱 녹취 (가상시나리오 기반)	1000시간
생활 도메인	크라우드 소싱 녹취 (가상시나리오 기반)	1000시간
총 구축량		3,000시간

- ❍ 문장별로 전사된 시간 기준으로 데이터량 산성
- ❍ 문장 앞뒤의 0.5초 이내의 묵음 시간이 포함된 Speech Signal로 데이터 구축량 산정
- ❍ Speech Signal 만으로 3000 시간 이상의 데이터 구축

2.3.2 원천데이터 수집

원천데이터 수집 절차는 데이터 시나리오 및 스크립트 작성, 녹취 가이드 작성, 크라우드소싱을 통한 데이터 녹취, 데이터 정제 단계로 이루어진다.



□ 시나리오 수집 및 작성

고객응대 시나리오는 키오스크 혹은 챗봇, 콜센터로 들어오는 고객 응대 데이터를 수집하여 대분류(예: 음식점), 중분류(예: 배달음식점), 소분류(예:배달문의), 고객문장(예: 배달비도 따로 내야 하나요?)으로 분류하여 작성한다.

대분류	중분류	소분류	문장
음식점	중식	배달문의	지금 배달되나요?
학원	어학원	교육비문의	한 달 수강료는 얼마인가요?

□ 시나리오 검수

다양한 소스를 바탕으로 작성된 시나리오에 대해 검수 인력을 통해 검수 작업을 진행한다.

검수 항목	설명
카테고리 분류 확인	• 대분류의 도메인 분류가 각 업종 그룹별로 맞게 정리가 되었는지 확인한다. • 중분류의 세부 업종의 분류로 맞게 되었는지 확인한다. • 소분류가 너무 세부적으로 나누어지지 않았는지 확인하여 지정된 카테고리 목록 안에서 설정되도록 조정한다.
문장 확인	• 작성된 문장이 고객의 대화인지 확인한다. • 작성된 문장이 구어체인지 확인하고, 문어체이면 구어체로 변환한다. • 작성된 문장에 개인정보식별 데이터가 포함되어 있으면 수정한다. - 이름: 대표적인 가명으로 변환 - 전화번호: 불가능한 조합의 번호 (010-0123-2344) - 주소: 동까지만 작성

□ 자연스러운 녹취 및 검수

가상 시나리오 기반의 녹취에서 가장 중요한 요소는 고객 입장에서 녹취를 수행하는 크라우드워커가 자연스러운 녹취를 진행해야 하며 검수 과정에서도 부자연스러운 녹취 음원을 재 녹취될 수 있도록 한다. 녹취자나 검수자가 어떤 녹취가 정상적인 음원이고 혹은 잘못된 음원인지를 교육될 수 있도록 녹취가 잘된 음원과 잘못된 음원들의 샘플들을 제작하여 작업 전 확인할 수 있도록 한다.

□ 데이터 수집 고려사항

활용성	설명	비고
서비스와 활용성	• 시나리오 작성 시, 다양한 서비스 분야에 적용될 수 있도록 2개 이상의 도메인 • 실제 데이터셋 활용을 고려하여 추가 데이터 등록을 최소화 하도록 슬롯 엔티티의 아이템 다양화	

소음	• 실제 활용 환경을 고려한 환경 소음 포함	
데이터의 균형	• 녹취되는 성별의 균형 • 녹취인 지역의 다양성 • 녹취인 나이의 다양성 • 한 시나리오에 대한 중복 녹취 50명 이하	

2.3.3 원천데이터 정제

인공지능 학습 데이터의 품질을 높이기 위한 절차로써, 녹취 작업자가 아닌 전문 품질 검수 담당자가 검수를 진행한다. 스크립트와 발화 내용의 일치 여부, 소음 여부, 음원 파일의 적합성 등 해당 기준에 부합하지 않으면 다시 녹취 작업을 시행한다.

□ 원천데이터 정제 기준

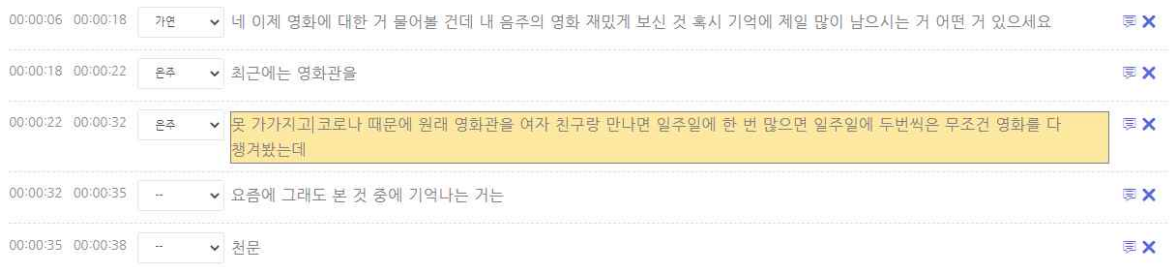
원천데이터 정제 기준	설명	비고
스크립트 정합성 여부	• 스크립트와 발화 내용의 정합성 여부	
소음 여부	• 현장 소음 포함 여부 • 소음의 강도 • 음성의 명확성	
음원 파일의 적합성	• 요구되는 음원 파일의 요구에 부합 여부 ▪ WAV, PCM(코덱), 16KHz, Mono	

2.4 어노테이션/라벨링

2.4.1 어노테이션/라벨링 절차

2.4.1.1 초기 전사 작업

- 초기 전사 작업은 STT(Speech-to-Text)결과를 바탕으로 진행합니다. 오디오와 변환 결과를 비교하여 잘못 분석된 문장이나 어절을 수정합니다.
- STT 변환 중 배경 노이즈가 단어로 인식하는 경우가 있습니다. 해당 문장을 삭제합니다.
- 해당 문장을 말하는 사람을 선택합니다.

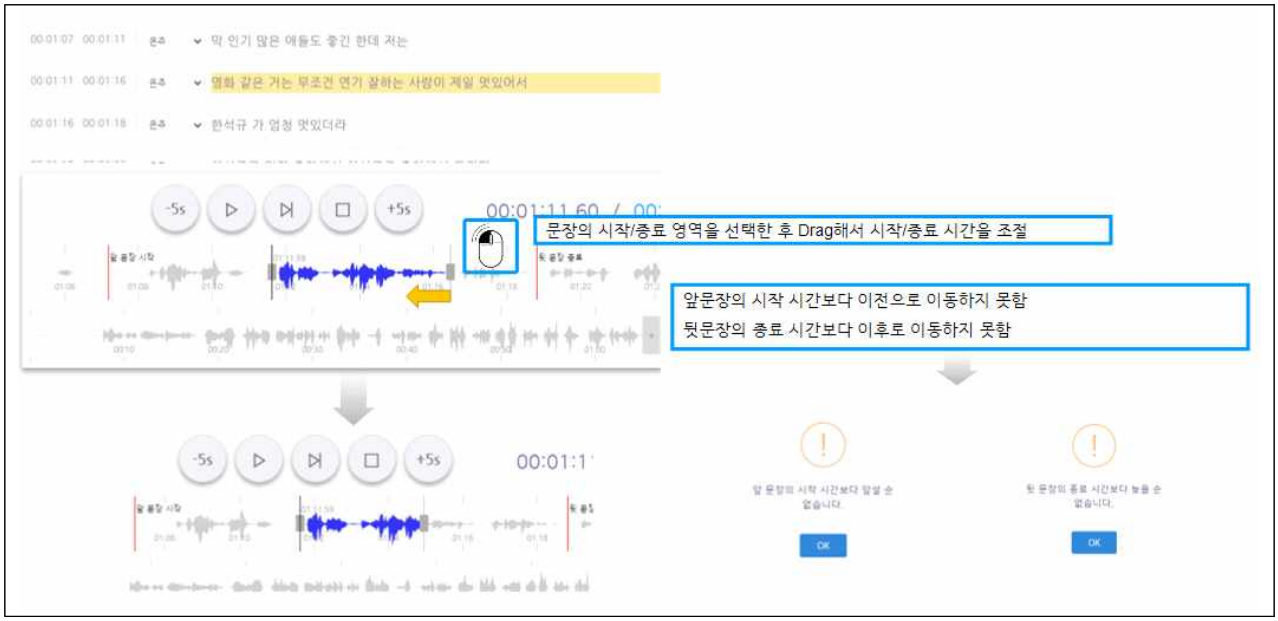


2.4.1.2 잘못된 어절 수정

2.4.1.3 과게 잘려진 문장의 병합

2.4.1.4 구분된 음성 구간의 시작점과 끝 지점 보정

- 음성인식 과정에서 잡음이나 묵음 구간이 포함되었을 수 있기 때문에 이를 보정합니다.


문장의 시작/종료 영역을 선택한 후 Drag해서 시작/종료 시간을 조절
앞문장의 시작 시간보다 이전으로 이동하지 못함
뒷문장의 종료 시간보다 이후로 이동하지 못함
OK
OK

2.4.2 어노테이션/라벨링 기준

1.1. 개요

1.1.1. 표준발성에서 벗어나거나 같은 전사에 대하여 두 가지 이상 발음이 가능한 경우 발음전사와 철자전사를 병행하며, 이 경우 (철자전사)/(발음전사)로 표기한다 (이 문서에서 향후 이를 '이중전사'라 칭한다). 예) (컴퓨터)/(컴퓨타)

1.1.2. 발음전사: 발성된 내용을 소리 값에 최대한 가깝게 표기한다. 이는 음성인식의 음향 모델링을 주된 목적으로 한다.

1.1.3. 철자전사: 표준어법에 맞게 표기한다. 이는 음성인식의 언어모델링 등을 주된 목적으로 한다.

1.1.4. 숫자, 외래어, 기호, 도량형 및 온도 단위는 발음 전사를 수행하되, 별도의 목록표를 생성하여 발음 전사별로 해당되는 표준 표기를 명시한다 (1.3, 1.7, 1.8절 참조).

1.1.5. 이중전사를 할 때, 이중전사의 범위를 표시하기 위해 괄호('(', ')')를 사용한다.

1.1.6. 이중전사, 잡음, 중복 발성 등을 나타내기 위한 특수 기호(meta symbol, 예: '/', '(', ')', '*', '+')는 원래의 목적으로만 표기되어야 한다. 특수기호가 실제 발성된 경우에는 발성된 형태를 반영하여 발음전사 한다. 분수 표기도 풀어서 표기한다.

예)

- 1/3 -> 삼 분의 일
- 슬래시, 작대기, slash
- 별표, star sign, asterisk
- 덧셈기호, 더하기, plus

1.1.7. 전사 과정에서 삽입되는 모든 기호('(), /'등)는 아스키코드만 사용하도록 한다.

1.1.8. 이중전사 할 때 '/'와 앞 뒤 괄호 사이에는 space를 두지 않는다.

1.1.9. 단일 발화 문장은 최소 2개 이상의 어절 혹은 5글자 이상으로 이루어져야 한다.

1.1.10. 단일 발화의 음성 길이는 최대 20초를 넘지 않도록 한다.

1.1.11. 발화 간 묵음(노이즈 포함)구간이 0.5초 이상 있을 경우 별개의 발화로 취급하며, 발화별 음성 싱크를 조절 후 전사한다. (이 기준으로 발화가 잘렸을 경우 상기된 최소 발화 기준(1.1.9)을 만족하지 않아도 된다.)

1.2. 잡음

1.2.0. 잡음은 대표 발화자의 음성을 제외한 전사된 다른 모든 소리로 정의한다.

1.2.1. 단어의 앞과 뒤에 거의 붙어 발생된 잡음은 단어와 분리하여 표기한다.

1.2.2. 잡음이 있는 상황에서 사람에게서 발생하는 잡음은 명확히 구분될 정도로 큰 것만 표기해도 좋다.

1.2.3. 다음에 정의된 잡음 이름 뒤에 '/'를 붙여 표기한다.

- b : 숨소리
- l : 웃음 소리(laugh)

- o : 다른 사람의 말소리가 포함된 경우 문장의 맨 앞에 표기

- n : 주변의 잡음 (상기에 명시된 요소 이외의 잡음)

1.2.4 문장 전체에 배경음악이 있는 경우 문장 맨 앞에 n/을 표기한 후 전사한다.

1.3. 숫자 표현

1.3.1. 기본적으로 숫자는 모두 숫자 기호가 아닌 문자로 표현하며, 필요한 경우 별도의 목록표를 작성한다.

- 숫자는 발음한 형태를 반영하여 문자로 표현하며, 한국어 및 영어에 대해 동일하게 적용한다.

- 한국어의 경우 십진 단위로 띄어 쓴다. 숫자를 하나씩 발음한 경우에도 띄어 쓴다.

- 단위를 나타내는 '년', '월', '일', '시', '분' 등은 숫자와 띄어 쓴다.

- 경계 내부에 설령 간투어 또는 잡음 등이 포함되어 있다고 하더라도 포함된 간투어를 포함한 상태로 경계를 표시해 준다.

- 기본적으로 발음전사로 수행하나 표준발음과 발성이 다른 경우 이중전사한다.

예)

- (5대)/(오 대) 그룹이 모여, 자동차 (5대)/(다섯 대)를

- (24시간)/(이십 사 시간), (24시간)/(스물 네 시간)

- (867-860-2437)/(팔 육 칠 팔 육 공 에 이 사 삼 칠)

- (14시)/(십 사 시), (14시)/(열 네 시)부터

- (1999년)/(천 구백 구십 구 년)에, (1999년)/(일천 구백 구십 구 년)에

- (스물)/(스무) 시간이요 <- 발성이 표준 발음과 다르므로 이중전사 함.

1.3.2. 숫자만으로 이루어진 기념일 등 특정 의미가 있는 단어들을 목록을 별도 작성한다. 이때, 아라비아 숫자에 붙는 단위, 조사나 어미는 붙인다.

예) 팔 일 오 8.15 사 일 구 4.19 오 칠 오 공 부대 5750부대

1.3.3. 숫자와 접미사의 표기는 표준어법을 따른다.

1.3.4. 서수의 경우 문자만을 표기한다. (첫번째, 두번째 등)

1.4. 간투어 표현

1.4.1. 발성자가 다음 발성을 준비하기 위해서 소요되는 시간을 벌기 위해서 발성하는 것으 로 의미 없는 것을 말한다. 간투어 뒤에 '/'를 붙여 표기한다.

예) 아/, 그/, 어/, 그/, 아/, 음/, 저/, 저기/, 에/, 으/, 응/

1.4.2. 감탄사는 별도로 취급하지 않고 간투어(1.4.1)와 동일하게 처리한다.

1.5. 외국어/외래어/약자

1.5.1. 일반적으로 외국어 문자로 표기하는 경우, 통상의 발음대로 읽은 경우는 통상의 표기를 따른다. 예) KBS, MBC, AT&T, ETRI, OPEC, FIFA 등

1.5.2. 우리말로 표기하여 자연스러운 것은 한글로 표기한다. 애매한 경우도 한글로 표기한다. 예) 뉴욕, 시카고, 파티, 버스, 핸드폰, 모바일, 인터넷, 호텔 등등

1.5.3. 통상적인 발음으로 읽는 외국어/외래어/약자들에 대한 목록표를 별도 작성한다.

1.5.4. 외국어 문자로(1.5.1) 표기하는 경우 발음전사 한다.

예) (KBS)/(케이비에스), (ATM)/(에이티엠), (ETRI)/(에트리), (FIFA)/(피파)

1.6. 문장 부호

1.6.1. 문맥적인 의미를 파악하여 표기하며, 한 문장이 끝나면 반드시 문장부호(마침표, 물 음표, 느낌표)를 표기하며, 발화자의 말 멈춤이나 문맥상 쉼표의 존재가 파악이 가능할 경우) 쉼표 ',’는 허용을 한다 (맞춤법 상 생략이 가능한 경우 생략한다.).

1.6.2. 상기에 명시된 것 이외의 문장 부호는 사용하지 않는다.

1.6.3. 인용문의 경우도 따옴표를 사용하지 않고 그대로 전사한다.

1.6.4. ‘시’ 등의 문장 부호가 사용되지 않는 어구의 전사는 원문의 표기에 따라 문장 부호를 사용하지 않는다.

1.7. 도량형 및 온도, 단위

1.7.1. 온도 등의 단위는 발음을 반영하여 한글/영어 문자로 적어준다.

1.7.2. “Degree Celsius”와 같이 띄어 쓰는 것이 분명한 경우를 제외하고는 붙여 쓴다. 예) kilometer (O), kilo-meter (X), kilo meter (X)

1.7.3. 모든 도량형은 목록표를 별도 작성한다. 이 때, 목록표에는 유로, 프랑 등 키보드에 없는 기호도 포함한다.

예) 밀리미터 mm 미리미터 mm 미리 mm 밀리메터 mm 킬로그램 kg

1.7.4. 숫자, 기호, 영문표기에 대하여 (영문표기)/(실제발음)으로 반드시 이중전사한다. 한국어는 (철자표기)/(실제발음)로 표기한다. 이 경우는 한국어는 전사문 작성자가 귀로 들었을 때 발음 자체는 명확히 들리는 경우이며, 발음 자체가 불명확한 경우는 1.9절을 참고한다.

예)

(UNESCO)/(유네코) : '유네스코'를 '유네코'로 잘못 발성한 경우

(UNESCO)/(유 엔 이 에쓰 씨 오) : ‘유네스코’를 ‘유 엔 이 에쓰 씨 오’로 알파벳으로 읽은 경우 (한국어)

(UNESCO)/(유네스코) // ‘유네스코’라고 통상의 방식대로 발성한 경우

(다섯)/(다싯) 대 // '다섯 대'를 '다싯 대'로 잘못 발성한 경우

1.7.5. 한글의 자음 및 모음의 경우 (철자표기)/(실제 발음)을 따른다. 예) (ㄱ)/(기역)

1.8. 띄어쓰기

1.8.1. 띄어쓰기는 표준어법에 맞추어 하되 표준어법으로 명확히 결정할 수 없는 경우에는 띄운다.

1.8.2 화자가 의도적으로 띄어서 발음하는 경우 띄어서 표기하나 이중전사한다.

예) (폭 탄 세 일)/(폭탄세일)

1.9. 알아듣기 힘든 발음

1.9.1. 화자가 발음한 내용을 잘 알아 듣기 힘들 때 어절의 뒷부분에 '*'를 붙여 이중전사한다. 즉 전후 문맥을 보고는 알 수 있으나 한 단어만을 놓고 볼 때 발음을 잘못하여 분명히 알 수 없을 때 붙여준다. 명확히 발성된 경우는 '*'를 붙이지 않는다.

예)

나는(이렇게)/(이럴끄*) 그것을 해결하였다. (청취시 '이럴끄'와 비슷하게는 들리지만, 분명히 알 수 없을 때)

나는 (이렇게)/(이럴끄) 그것을 해결하였다. (청취시 '이럴끄'가 분명히 들리는 경우)

1.9.2. 방언에 해당하는 발성은 다음과 같이 이중 전사를 한다. 예) (장의사)/(장으사), (학교)/(핵교)

1.9.3. 문맥을 고려해봐도 전혀 알아들을 수 없는 어절은 'u/' 으로 표기한다.

1.9.4. 발성과 동시에 발생하는 잡음은 어절 끝에 '*'를 붙여 표기한다. 예) 기차 타는 곳이* 어디입니까? // '곳이' 가 발성될 때 외부잡음이 크게 섞임

1.9.5.1. 반복 발성이나 잘못된 발성은 반드시 표기 한다. 이때 불필요하게 중복 또는 잘못 발성된 부분은 뒤에 '+'를 붙인다. 예) 아침에 학교+ 학교에 갔다.
I don't have sta+ stati+ statistical knowledge.

1.9.5.2. 통상적인 표현으로 굳어진 반복 발성(형용사의 반복 표현 등)은 일반적으로 표기한다.

예) 너무너무, 많이많이

1.9.6. 반복 발성의 발음이 불분명할 때 * 와 + 를 병기한다. 예: "학교*+ 학교에 갔다."

1.9.7. 대화체 문장은 문장 자체가 이상하더라도 그대로 전사한다.

1.10 발화자의 표기

1.10.1. 각 발화 문장의 발화자는 모두 표기한다.

1.10.2. 발화자의 성별 및 각 성별의 등장 순서에 따라 남성의 경우 'M + 등장순서', 여성의 경우 'F + 등장순서'로 표시한다.

예) 해당 음성의 첫번째로 등장한 남자: 'M1', 해당 음성의 첫번째로 등장한 여성: 'F1', 해당 음성의 3번째로 등장한 남자: 'M3'

1.10.3. 해당 발화자 판별이 어려울 시 'etc'로 표현한다.

1.10.4. 하나의 대화 안에서 같은 사람일 경우 화자 정보를 동일하게 유지한다.

예)

- 30분 길이 강의: 처음 선생님의 화자 정보를 F1으로 정한 경우, 강의 끝까지 해당 선생님에 대해 F1 유지

- 고객 응대/상담에서 대화: 고객 화자 정보를 M2으로 정한 경우, 대화 끝까지 동일 고객에 대해서 M2 유지

1.10.5. 발화자 지역, 나이대, 성별, 녹음방식, 잡음환경, 수집디바이스에 대한 정보가 있을 경우 발화자 정보에 추가로 표시한다.

1.10.6. 동일한 목소리라도 음성 내 역할이 다를 경우 다른 발화자로 처리한다.

1.10.7. 두 명 이상의 발화자가 동시에 말을 할 경우 주 발화자(먼저 말을 한 대화의 주체)가 대표 발화자가 되며 이를 기준으로 잡음 처리 규칙을 따른다.

1.10.8. 복수의 대화에서 같은 발화자라는 것을 알 수 있는 경우, 별도의 메타데이터를 이용하여 해당 정보를 제공한다.

예)

- 서로 다른 강의임에도 발화자(강의 선생님)가 동일한 경우

- 서로 다른 고객 상담 대화임에도 발화자(고객 또는 상담사) 동일한 경우

1.10.9. 화자가 사람이 아닌 사물 등의 목소리를 낼 경우 원 화자로 표기한다. (F1이 컴퓨터의 목소리를 낼 경우 F1으로 표기한다.)

2.4.3 어노테이션/라벨링 도구

ㅇ **프로젝트 관리**

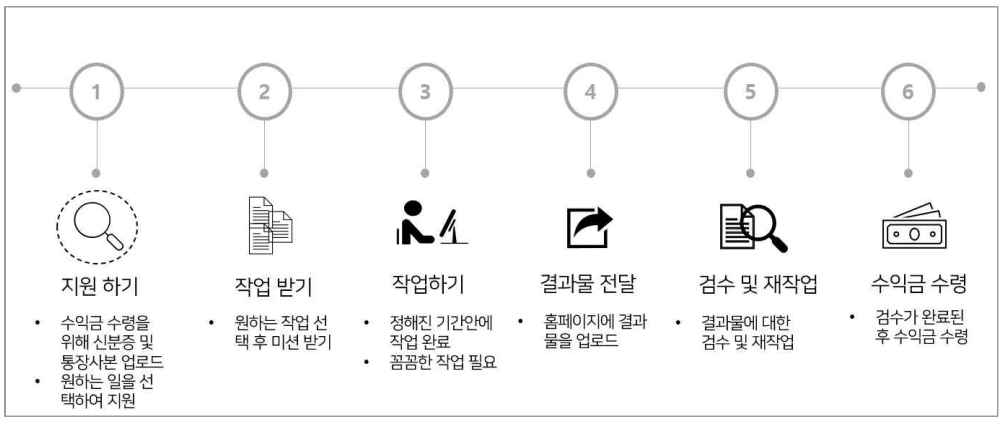
- 음성데이터를 여러 기능이나 목적 등으로 세분화 하여 프로젝트로 구성하여 개별 전사 프로젝트로 진행 하도록 지원

<프로젝트 생성 화면>

ㅇ **작업자 관리**

- 작업자를 선정하고 작업자 별 진행상황이나 진척도 혹은 관련 내용 일체를 확인하고 관리하는 기능 제공

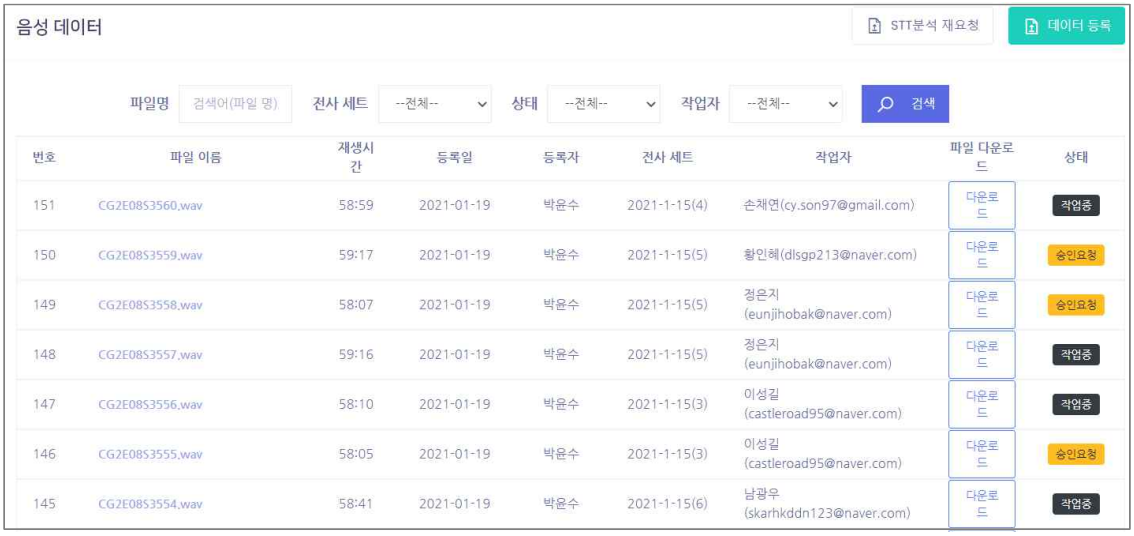
<작업자 프로세스>



○ **전사 작업 지원**

- 프로젝트 별 작업자들의 전사 작업 진행도를 파악하는 등의 기능을 지원하는 관리자용 기능 지원
- 프로젝트 별 작업자들이 직접 자신들의 작업 진행도 및 참여 실적, 성공 실적 등을 확인하며 작업자 지원 기능 지원

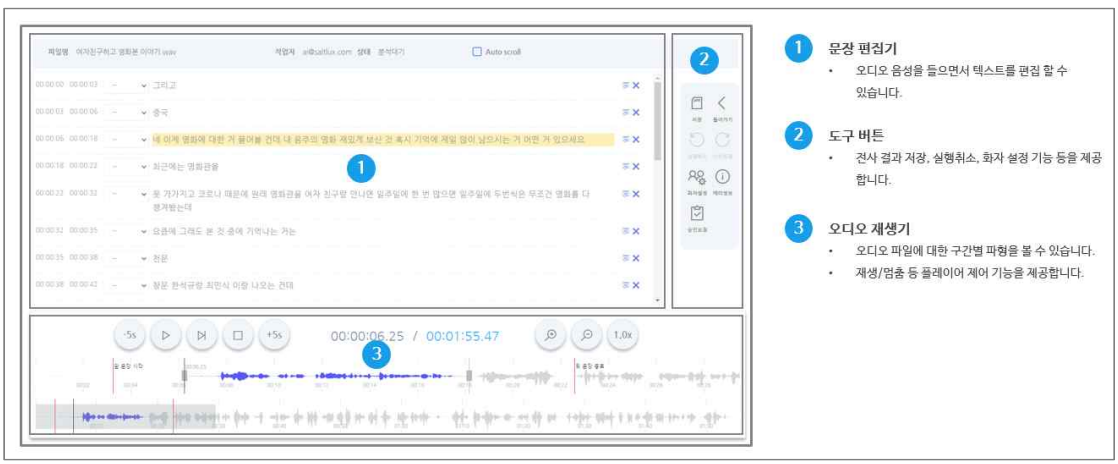
<전사관리화면>



○ **전사작업 지원 플랫폼**

- 음성 자료 전사 시 작업 효율을 높일 수 있는 전사 프로그램을 활용하여 전사 지침를 준수하며 학습데이터를 구축

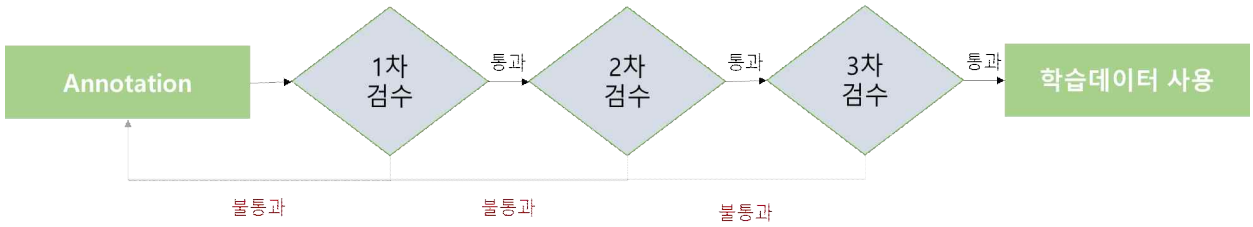
<음성데이터 전사도구>



- 모든 기능을 단축키로 조작할 수 있도록 하여 전사 작업 시간을 최소화
- 세그먼트 합치기/나누기 기능으로 문장 발화의 정확한 시간 정보 저장 가능
- 재생 속도 조절 가능으로 화자가 많거나 말이 빠른 경우 전사 작업에 용이
- 매뉴얼 기능으로 반복되는 태깅 작업의 오류를 최소화
- 자동 UTF-8 저장 방식으로 파일 포맷 오류 가능성 제거

2.5 검수

2.5.1 검수 절차



<어노테이션 결과 검증 절차>

- 데이터 검수는 양질의 데이터를 얻는 데 필요한 작업으로 정성적/정량적 평가를 통해 데이터의 유효함을 판별함
- 웹 저작도구를 사용하여 전사한 어노테이션 결과에 대해 1차, 2차, 3차 검수자를 걸쳐 음성품질 / 어노테이션 정확도 / 대화 주제 및 저작권확인 / 목표데이터 수집량 달성 여부를 확인함
- 1, 2차 검수자는 음성품질 및 어노테이션 정확도, 대화 주제 및 저작권 확인 위주로 검수를 진행하며, 교차 검증을 통해 데이터의 오류를 최소화함
- 1, 2차 검수는 전수검사로 모든 데이터에 대하여 진행함
- 3차 검수자는 목표데이터의 음성품질 및 도메인별 수집량 달성 여부 위주로 검수를 진행하며, 데이터의 다양성을 보장할 수 있도록 함
- 3차 검수는 검수도구를 통해 자동검수를 진행하며, 통과한 데이터에 대해서도 샘플링 검사를 진행하여 데이터 완결성을 보장함
- 각 차수의 검수자는 피드백을 통해 데이터가 한쪽에 치우치지 않도록 하며, 전사자가 숙달할 수 있도록 지원함
- 3차 검수 완료 이후 TTA를 통한 데이터 검증 절차를 걸치며, 동시에 AI 모델 학습을 진행하여 데이터의 적합성과 유효성을 최종 판별함

2.5.2 검수 기준

대항목	피드백 목표	비고
음성품질	• 음성 파일의 샘플레이트, 채널, 인코딩의 적절성 여부 • 음성의 명료함 정도 (SNR)	전사자와 검수자 주관에 맡기며, 과반수 동의 시 적절함으로 판정
어노테이션 정확도	• 음성에 대응하는 텍스트 어노테이션 정확성 여부 • 음성에 대응하는 발화자 정보의 정확성 여부 • 음성과 텍스트의 싱크 정확성 여부	상기와 동일
대화 주제 및 저작권확인	• 대화 주제의 편향성 여부 • 저작권 및 개인정보 침해 여부	상기와 동일
목표데이터 수집량 달성	• 데이터 도메인, 화자 등 수집량의 적절성 여부	검수도구를 통한 자동 검수 및 샘플링 검사

대항목	데이터 항목	어노테이션	판정 기준
음성품질	음성 파일 형식	• N/A	• 검수도구를 통한 자동 검수 및 샘플링 검사
	명료함	• N/A	• 전사자와 검수자에 의한 음성 명료함 판단
어노테이션 정확도	텍스트	• 전사규칙에 따라 전사된 발화 내용	• 3인 이상의 검수자 판단하에 과반수 통과
	발화자	• 화자 고유번호, 화자 성별, 나이	• 3인 이상의 검수자 판단하에 과반수 통과, 검수도구를 통한 자동 검수, 샘플링 검사
	싱크	• 음성과 텍스트의 시작과 끝	• 검수도구를 통한 자동 검수 및 샘플링 검사
대화 주제 및 저작권 확인	대화주제	• 고객 응대 도메인에 대한 23종 태깅	• 3인 이상의 검수자 판단하에 과반수 통과
	저작권 및 개인정보	• N/A	• 3인 이상의 검수자에 의한 저작권 및 개인정보 침해 여부 판단
목표데이터 수집량 달성	도메인	• 고객 응대 도메인 별 음성 길이	• 검수도구를 통한 자동 검수 및 샘플링 검사
	화자	• 화자 별 음성 길이	• 검수도구를 통한 자동 검수 및 샘플링 검사

2.5.3 검수 도구

- 도메인 수집량 검수 도구: 전사파일에 기초하여 음원을 발화 단위로 분리하고, 파일명 파싱을 통해 도메인/세부도메인 정보 추출 후 수집량을 측정하여 검수함
- 발화자 성별 검수 도구: 전사파일에 기초하여 발화자의 성별이 제대로 작성되어 있는지 검수함
- 어노테이션 정확도 검수 도구: 전사파일에 기초하여 unknown으로 표시된 발화가 전체 발화의 5%를 넘지 않도록 검수함

2.6 활용

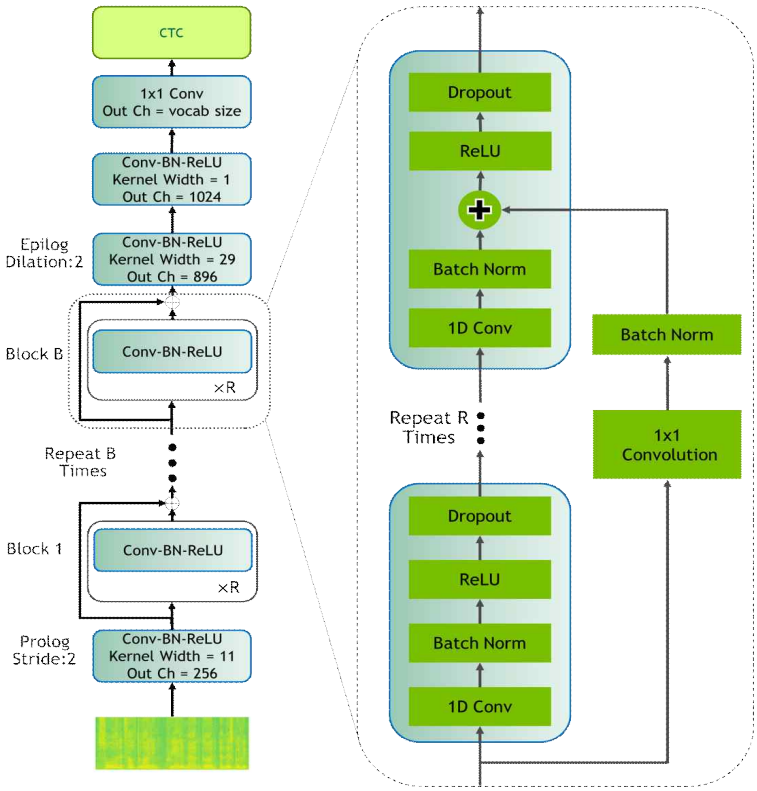
활용 방안	설명
데이터 공개	AIhub를 통한 구축 데이터셋 공개
데이터 활용 서비스	AI 음성인식 모델 기반 회의 대화록 자동 생성 서비스
모델 프로토 타입	Jasper 모델 기반 오프라인 음성인식 모델 개발

2.6.1 음성인식 모델 및 파이프라인

2.6.1.1 음성인식 모델

- Jasper (Just Another Speech Recognizer)은 2019년 nvidia에서 발표한 음성인식 모델로 LAS와 parameter 수를 비교 했을 때 크게 차이 나지 않지만, nvidia TensorRT를 활용한 모델 최적화를 고려하여 설계되었음
- 순차적인 연산을 필요로 하는 autoregreesive 방식의 LAS 및 RNN-Transducer 모델과 달리 Jasper는 non-autoregreesive 방식으로 parallel하게 동작하여 빠른 추론이 가능함
- 일반적으로 non-autoregreesive 방식은 이전 state의 context를 이용하지 않기에 autoregreesive 방식에 비해 성능이 떨어지나, Jasper는 state-of-the-art에 근접한 성능을 보임
- 따라서 Jasper 모델 최적화를 진행할 경우 추론 속도와 성능을 모두 만족시킬 수 있으며 온라인/오프라인 음성인식에도 구분 없이 적용 가능하다는 특성을 가짐

<Jasper 모델 아키텍처>



2.6.1.2 음성인식 파이프라인

- 전처리단계에서는 모델 학습을 위한 데이터의 음성 특징 벡터를 추출
- 음향모델학습 단계에서는 추출된 데이터를 이용하여 음향 모델을 학습
- 언어모델학습 단계는 음향 모델의 성능을 향상시키기 위해 언어 모델을 추가적으로 사용하는데, 텍스트(스크립트)를 이용하여 학습
- 최종적으로 추론과정 을 통해 학습된 모델의 유효성을 검증

- 모델 학습과 검증, 평가에 사용하는 데이터는 서로 중복되지 않으며, 통상적으로 학습:검증:평가 (8:1:1)의 비율로 나눠 제공하나, 비율로 나눌 경우 데이터량이 지나치게 많음
- 본 과제에서는 3000시간 데이터 중 2950시간을 학습(train) 데이터, 30시간 분량의 데이터를 검증(Valid) 데이터로 20시간 분량의 데이터를 평가(Test) 데이터로 사용하며, 평가 데이터는 학습/검증 데이터셋과 발화자가 겹치지 않는 데이터로 구성됨

2.6.2 데이터 활용

❍ 키오스크는 최근에 모든 분야에게 가장 유용하게 사용되고 기기로 사용되고 있다. 카페, 식당, 극장, 호텔 등 거의 모든 분야에서 터치 기반의 키오스크를 사용하여 주문, 예약, 발급, 정보 조회 등의 유용한 서비스를 제공받고 있다.

❍ 하지만 지금은 코로나 19로 인해 터치 기반의 기기(엘리베이터, 키오스크) 들은 항상 코로나 바이러스의 전염 위험을 가지고 있어서 비접촉 고객 서비스의 필요성이 대두되었음.

❍ 기존의 터치 기반의 키오스크 업체들은 본 과제로 구축된 데이터셋으로 다양한 도메인 서비스 환경의 비접속 고객 서비스를 위한 음성인식 모델 및 대화 시나리오를 구성할 수 있는 기본 대화 시나리오를 제공함.

❍ 본 구축 가이드를 통해 서비스를 개발하고자 하는 업체의 특화된 시나리오의 발화 음성들을 통합하여 해당 업체에 특화된 음성인식 학습용 데이터셋을 구축할 수 있음.

2.6.3 데이터 제공

- ai-hub 홈페이지(https://www.aihub.or.kr)를 통해 데이터를 제공하고자 함
- 데이터는 기본적으로 공개 데이터세트로서 인공지능 산업, 기술개발을 하고자하는 경우 신청을 통해 데이터 사용 권한을 획득할 수 있다.
- 데이터의 사용을 위한 준수사항을 서약 받은 후 제공을 하며, 준수사항은 다음과 같음

ㅇ 1. AI데이터 등은 권리자로 데이터 제공 처를 명시할 경우 자유로운 이용 및 변경이 가능하며 2차적 저작물에도 동일하게 출처를 명시하여야 한다. ㅇ 2. 제공받은 AI데이터 등에 대하여 승인을 얻은 연구자가 아닌 제 3자에게 열람하게 하거나 제공, 양도, 대여, 판매하지 아니한다. ㅇ 3. AI데이터 등에 기반을 둔 제품·서비스 개발 또는 기술연구에 활용할 때에는 논문 등 결과물에 데이터의 출처를 반드시 명시 하여야 한다. ㅇ 4. AI데이터 등의 이용 및 그에 따른 연구로 인하여 발생하는 모든 책임은 AI데이터를 이용한 개인 및 해당 기관의 기관장에게 있다. ㅇ 5. 향후 데이터의 제공 처에서 데이터의 활용사례 등 실태조사를 수행할 경우 성실하게 임하여야 한다.

2.6.4 데이터 유지보수

- 유지 보수는 크게 데이터 유지 보수와 서비스 유지 보수로 나누어 제공함
- 데이터 유지보수는 데이터를 제공하는 aihub 홈페이지를 통해 사용자로부터의 이슈를 취합함
- 데이터 유지보수는 오탈자, 메타데이터 불일치 등과 같은 데이터의 오류사항에 대한 수정 작업을 의미하며, 분기/반기 단위로 취합된 이슈사항에 대응함
- 서비스의 유지보수는 깃허브를 통하여 이슈를 정리하여 분기/반기 단위로 수정 사항을 일괄 적용하여 지속적으로 보강함