

●○ 상황별 음성 과제

# 고객 응대 데이터



●○ 개요: 고객응대 데이터란?

- 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 하고 있음.
- 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 비접촉 고객 서비스를 위한 음성인식 모델 및 대화 시나리오를 구성할 수 있는 기본 대화 시나리오를 제공함.
- 본 고객응대 데이터의 사례에 대해서는 아래의 내용을 참고할 수 있다.

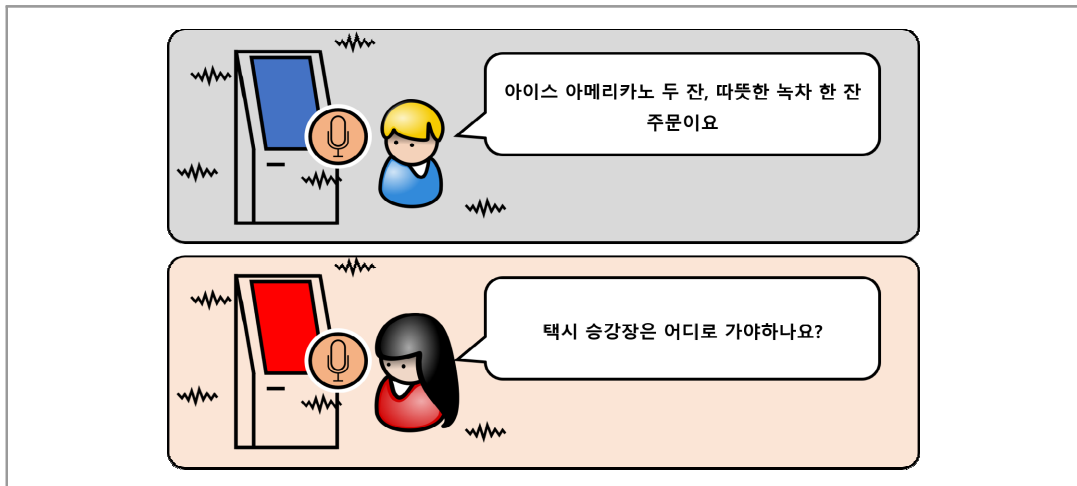


그림 | AI 키오스크 음성인식 서비스 예시

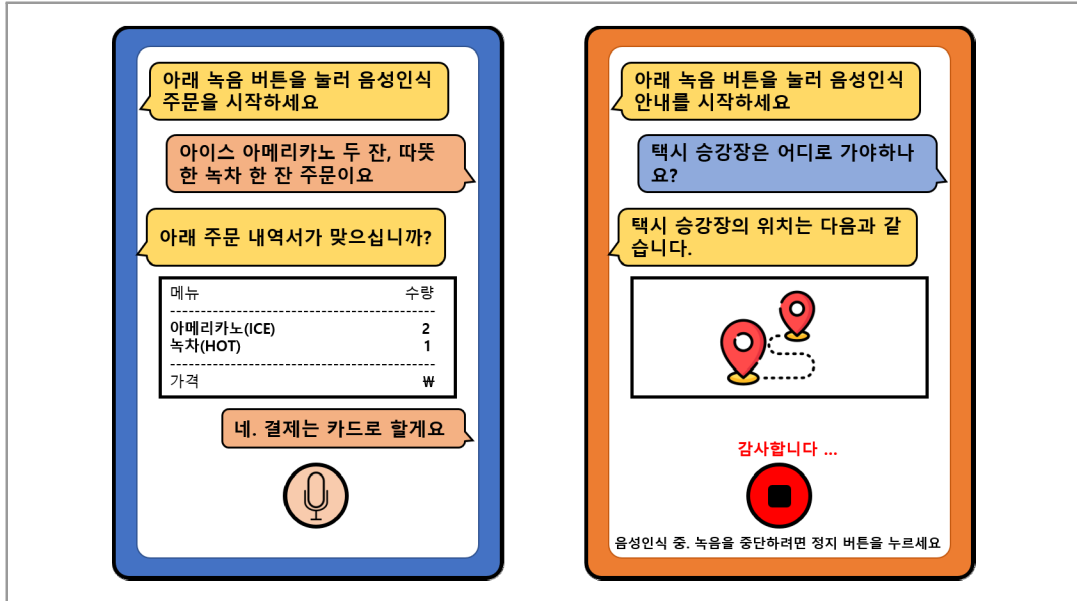


그림 | AI 키오스크 음성인식 서비스 화면 예시

## ●○ 데이터셋의 구성

- 본 데이터셋은 3개 도메인을 기준으로 3,000시간 으로 구성되어 있음.
- 문장별로 전사된 시간 기준으로 데이터량 산정.
- 문장 앞뒤의 0.5초 이내의 묵음 시간이 포함된 Speech Signal로 데이터 구축량 산정.
- Speech Signal 만으로 3000 시간 이상의 데이터 구축.

데이터의 종류	수집시간	제공방식
구매 도메인	1,000시간	JSON 포맷 파일
예약 도메인	1,000시간	JSON 포맷 파일
생활 도메인	1000시간	JSON 포맷 파일
총 구축량	3,000시간	

- 고객응대 데이터셋의 구축량은 다음과 같다.

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

- 데이터셋의 음원 파일은 압축 변환되지 않은 아래의 원본 파일로 구축

파일 형식	코덱	샘플레이트	채널
WAV	PCM	16K	Mono

## ●○ 데이터셋의 설계 기준과 분포

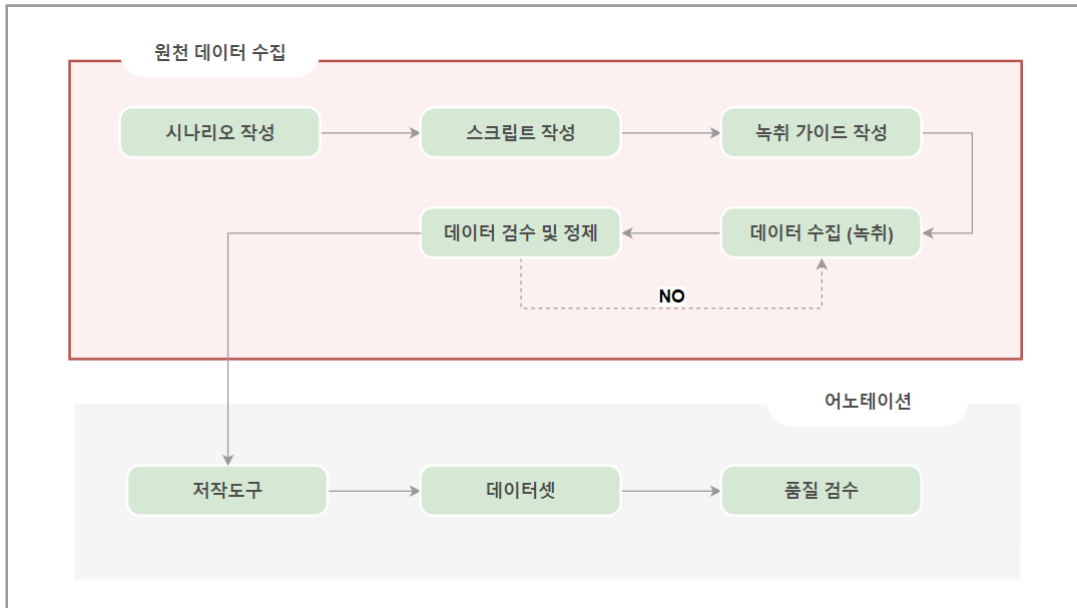
- 데이터셋을 설계할 때 가장 중요하게 고려한 부분은 편향성 방지 및 획득대상의 사실성이다. 고객응대가 예상되는 다양한 도메인 및 다양한 상황을 기반으로 데이터가 분포 되도록 설계하여 응용서비스 단계의 확장성을 강화하고자 했다.

도메인 종류	도메인 구축시간	업종별 데이터량 (시간)
구매 도메인	1000 시간	- 카페: 100 시간 - 식당: 300 시간 - 의류: 300 시간 - 소매: 300 시간
예약 도메인	1000 시간	- 숙박: 370 시간 - 학원: 280 시간 - 독서실: 40 시간 - 미용실: 150 시간 - 여행: 160 시간
생활 도메인	1000 시간	- 민원: 160 시간 - 세탁소: 50 시간 - 옷수선: 40 시간 - 여가오락: 280 시간 - 위치정보: 470 시간

- 고객응대 데이터의 수집은 실제 상황에서 데이터를 수집시 공개에 필요한 법적권리를 획득하는데 현실적인 한계가 있어 가상시나리오를 기반으로 크라우드 워커를 사용해 데이터를 수집하였음.
- 가상 시나리오의 녹취는 스크립트 구성의 다양성 확보가 중요하다. 따라서 녹취 스크립트의 구성은 단순한 상품 목록을 늘리는 구성이 아닌 챗봇 구성 데이터와 콜센터 고객응대 데이터를 바탕으로 다양한 문장을 구성하는 방식으로 스크립트 구축되며, 원천 스크립트 데이터 구성은 대분류(도메인), 중분류(업종), 소분류(문장 카테고리), 발화문장으로 분류하여 작성한다.

대분류	중분류	소분류	문장
구매도메인	식당	배달문의	지금 배달되나요?
		메뉴주문	세트메뉴 말고 단품으로 추천해 주세요.
예약도메인	호텔	조식문의	아침 식사가 포함되나요?
		할인문의	평일 할인되는 가격으로 묵을 수 있나요?

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



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

검수 항목	설명
카테고리 분류 확인	<ul style="list-style-type: none"> <li>• 대분류의 도메인 분류가 각 업종 그룹별로 맞게 정리가 되었는지 확인한다.</li> <li>• 중분류의 세부 업종의 분류로 맞게 되었는지 확인한다.</li> <li>• 소분류가 너무 세부적으로 나누어지지 않았는지 확인하여 지정된 카테고리 목록 안에서 설정되도록 조정한다.</li> </ul>
문장 확인	<ul style="list-style-type: none"> <li>• 작성된 문장이 고객의 대화인지 확인한다.</li> <li>• 작성된 문장이 구어체인지 확인하고, 문어체이면 구어체로 변환한다.</li> <li>• 작성된 문장에 개인정보식별 데이터가 포함되어 있으면 수정한다.                         <ul style="list-style-type: none"> <li>- 이름: 대표적인 가명으로 변환</li> <li>- 전화번호: 불가능한 조합의 번호 (010-0123-2344)</li> <li>- 주소: 동까지만 작성</li> </ul> </li> </ul>

- 가상 시나리오 기반의 녹취에서 가장 중요한 요소는 고객 입장에서 녹취를 수행하는 클라우드 워커가 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 수집에서는 아래와 같은 내용을 주로 고려하였다.

활용성	설명	비고
서비스와 활용성	<ul style="list-style-type: none"> <li>• 시나리오 작성 시, 다양한 서비스 분야에 적용될 수 있도록 2개 이상의 도메인</li> <li>• 실제 데이터셋 활용을 고려하여 추가 데이터 등록을 최소화 하도록 슬롯 엔티티의 아이템 다양화</li> </ul>	
소음	<ul style="list-style-type: none"> <li>• 실제 활용 환경을 고려한 환경 소음 포함</li> </ul>	
데이터의 균형	<ul style="list-style-type: none"> <li>• 녹취되는 성별의 균형</li> <li>• 한 시나리오에 대한 중복 녹취 50명 이하</li> </ul>	

## ●○ 데이터 구조

데이터셋에 따른 항목과 해당 값은 테이블과 같다.

No	항목		타입
	한글명	영문명	
	데이터셋	dataSet	
1	데이터셋 버전	version	String
2	녹취된 음원의 URL	mediaUrl	String
3	녹취된 날짜	date	String
4	음원 데이터 상세 정보	typeInfo	
4-1	음원 카테고리 정보 : 강의, 회의, 고객응대 등	category	String
4-2	음원 서브카테고리	subcategory	String
4-3	음원 녹취 장소	place	String
4-4	화자 목록	speakers	List
4-3-1	화자 유형 : 강사, 상담사, 고객, 기타	type	String
4-3-2	인입 유형 : 유선, 모바일, 인터넷 등	gender	String
4-5	화자 성별 : 남성, 여성	inputType	String
5	전사 데이터 목록 : 화자가 변경될 때마다 생성	dialogs	List
5-1	화자 아이디 : speakers 에 등록된 순번	speaker	String
5-2	전사된 텍스트	text	String
5-3	전사된 텍스트의 음원 재생 시작 위치	startTime	String
5-4	전사된 텍스트의 음원 재생 끝 위치	endTime	String
5-5	전사된 텍스트 문장과 관련된 태그 리스트	tags	String

## ●○ 데이터 예시

```
{
  "dataSet":
  {
    "version": "1.0",
    "mediaUrl": "http://.../23skskjdsfsks.wav",
    "date": "2020/05/20",
    "typeInfo":
    {
      "category": "conference"
      "speakers":
      [
        {
          "type": "bot", "gender": "female"
        },
        {
          "type": "customer", "gender": "female"
        }
      ],
      "inputType": "mobile"
    }
    "dialogues":
    [
      {
        "speaker": 0,
        "text": "무엇을 주문하시겠습니까?",
        "startTime": "10.212",
        "endTime": "12.432",
        "tags": [ "Question", "Bot" ]
      },
      {
        "speaker": 1,
        "text": "아이스아메리카노요",
        "startTime": "12.569",
        "endTime": "13.698",
        "tags": [ "Answer", "Customer" ]
      }
    ]
  }
}
```

## ●○ 데이터 구축 과정

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.
	클라우드소싱을 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

## ●○ 검수와 품질 확보

- 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 정확도를 위주로 검수를 진행하며, 교차 검증을 통해 데이터의 오류를 최소화함
- 3차 검수자는 대화 주제 및 저작권확인, 목표데이터의 도메인별 수집량 달성 여부를 위주로 검수를 진행하며, 데이터의 다양성을 보장할 수 있도록 함
- 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 SI 모델 학습을 진행하여 데이터의 적합성과 유효성을 최종 판별함
- 어노테이션 결과 피드백 기준

대항목	피드백 목표	비고
음성품질	<ul style="list-style-type: none"> <li>• 음성 파일의 샘플레이트, 채널, 인코딩의 적절성 여부</li> <li>• 음성의 명료함 정도 (SNR)</li> </ul>	전사자와 검수자 주관에 맡기며, 과반수 동의 시 적절함으로 판정
어노테이션 정확도	<ul style="list-style-type: none"> <li>• 음성에 대응하는 텍스트 어노테이션 정확성 여부</li> <li>• 음성에 대응하는 발화자 정보의 정확성 여부</li> <li>• 음성과 텍스트의 싱크 정확성 여부</li> </ul>	상기와 동일
대화 주제 및 저작권확인	<ul style="list-style-type: none"> <li>• 대화 주제의 편향성 여부</li> <li>• 저작권 및 개인정보 침해 여부</li> </ul>	상기와 동일
목표데이터 수집량 달성	<ul style="list-style-type: none"> <li>• 데이터 도메인, 화자 등 수집량의 적절성 여부</li> </ul>	상기와 동일

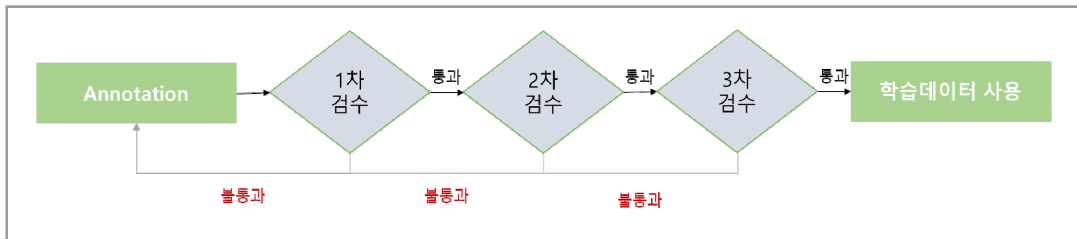


그림 | 어노테이션 결과 검증 절차

- 어노테이션 데이터 항목별 세부 판정 기준

대항목	데이터 항목	어노테이션	판정 기준
음성품질	음성 파일 형식	• N/A	• 전사도구 내 음성 파일 체크를 통한 자동 검수
	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
	싱크	• 음성과 텍스트의 시작과 끝	• 3인 이상의 검수자 판단하에 과반수 통과



대항목	데이터 항목	어노테이션	판정 기준
대화 주제 및 저작권 확인	대화주제	• 강의 전체 도메인에 대한 23종 태깅	• 3인 이상의 검수자 판단하에 과반수 통과
	저작권 및 개인정보	• N/A	• 3인 이상의 검수자에 의한 저작권 및 개인정보 침해여부 판단
목표데이터 수집량 달성	도메인	• 강의 별 음성 길이	• 음성 발화자 및 싱크 정보를 바탕으로 알고리즘을 통한 자동 계산
	화자	• 화자 별 음성 길이	• 음성 발화자 및 싱크 정보를 바탕으로 알고리즘을 통한 자동 계산

승인 이력

번호	상태	내용	날짜
1	승인거절	10:37 ~ 10:48 영어로 -> (0)/(영)으로 11:10 ~ 11:16 마이너스에 (2) 분의 -> 마이너스 (A 분의)/(에이 분의) 15:27 ~ 15:29 최고 왕의 -> 최고 차왕의 18:44 ~ 18:54 숫자 자들의 -> 숫자 들의 19:28 ~ 19:42 세 계급만 접근 -> 격분 33:40 ~ 33:48 (25)/(이 십 오) -> (25)/(이십 오) 옆에 4도 똑같이 34:05 ~ 34:11 위에 것과 똑같이 바꿔주세요 34:11 ~ 34:20 위에 것 똑같이 35:02 ~ 35:08 위에 것 똑같이 36:08 ~ 36:11 37:26 ~ 37:36 (66)/(육십육), (61)/(육 십 일) -> (66)/(육십 육), (61)/(육십 일) 37:39 ~ 37:50 40:28 ~ 40:34 42:15 ~ 42:28	2020.10.29 17:10:04

그림 | 검수자 피드백 예시

## ●○ 데이터 구축 담당자

수행기관(주관) : 티맥스소프트

(전화 : 031-8018-9325, 이메일 : younseong\_kim2@tmax.co.kr)