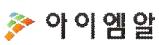
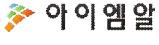
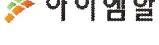
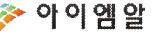
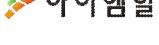
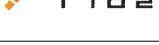
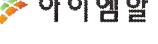


인공지능 데이터 구축·활용 가이드라인

- 데이터 94 위급상황 음성/음향 데이터 -

인공지능 데이터 구축	사업 총괄	 아이엠알
	데이터 설계	 아이엠알
	원천데이터 수집 및 정제	 아이엠알
	데이터 가공	 아이엠알  mforus  TmaxAI
	데이터 검수	 아이엠알  mforus  vtw <small>브이티디플유 Value through Technology & Wisdom</small>
	크라우드 소싱	 아이엠알
	저작도구 개발	 아이엠알
	AI모델 개발	 아이엠알  TmaxAI
	응용 서비스 개발	 아이엠알
가이드라인 작성	(주)아이엠알	조성훈 차장
가이드라인 버전	버전 2.5.0 작성일 2021.02.27	

목 차

1. 데이터 명세 정보	2
1.1 데이터 정보 요약	2
1.2 데이터 포맷	3
1.3 어노테이션 포맷	4
1.4 데이터 구성	8
1.5 데이터 통계	9
1.6 원시데이터 특성	11
1.7 기타정보	17
2. 데이터 구축 가이드	20
2.1 데이터 구축 개요	20
2.2 문제정의	22
2.3 획득·정제	23
2.4 어노테이션/라벨링	42
2.5 검수	52
2.6 활용	56

붙임. 어노테이션 및 라벨링 매뉴얼

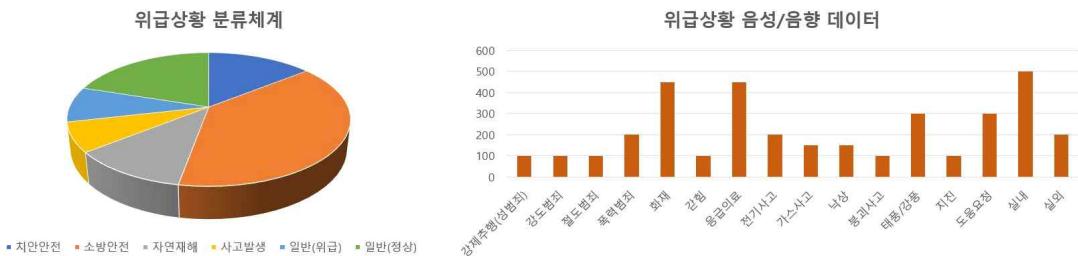
1. 데이터 명세 정보

1.1 데이터 정보 요약

- 본 문서는 위급상황 음성/음향 데이터 구축을 위한 절차와 가공 지침을 제시한다. 위급상황 데이터의 수집단계부터 데이터 정제, 데이터 가공, 데이터 검수 및 절차에 대한 설명을 제공한다. 음성/음향 데이터 및 저작도구의 효율적인 활용을 위해 원천데이터 및 가공데이터의 형태와 구축 절차(1장 및 2장)를 따라야 하며 구축된 데이터의 어노테이션 방법과 검수 기준 및 방법을 제시하여 고품질의 AI학습 데이터를 확보할 수 있도록 하였다.

데이터 이름	위급상황 음성/음향 데이터
활용 분야	<ul style="list-style-type: none"> • 위급상황 음성/음향 데이터를 활용한 응용 서비스 – 이상탐지 솔루션 • 유관기관(경찰청, 소방청, 복지기관, 의료기관 등)과 연계한 취약계층 돌봄 지원 • 치안, 사고, 질병 등 위급상황에서 음성/음향 데이터 분석 기반 상황 판단 기술 활용하여 빠른 상황 판단을 통한 사회 안전 서비스 지원
데이터 요약	<ul style="list-style-type: none"> • 위급 상황 음성/음향 데이터 <ul style="list-style-type: none"> - 성별 : 남성, 여성 - 연령대 : 유아, 청소년, 노인, 기타 - 상황 : 치안안전, 소방안전, 자연재해, 사고활동, 일반(위급), 일반(정상) • 위급 상황 AI 학습용 데이터셋 <ul style="list-style-type: none"> - 16개 클래스(중분류) 1,000건 이상의 상황·3,500시간 - 음성/음향 단일 데이터셋, 음성/음향 복합 데이터셋
데이터 출처	<ul style="list-style-type: none"> • 공개 데이터 <ul style="list-style-type: none"> - 공개 오디오 구축사이트 - DCASE, AudioSet, UrbanSound, FreeSound 등 • 크라우드 소싱 <ul style="list-style-type: none"> - 크라우드 워커 : 직접 녹음(가정), 수집 데이터(실내, 실외), 상황 연출(녹음실) - 크라우드 소싱 데이터 수집 사이트 : http://www.d2ai.co.kr <p>※ 저작권/개인정보 이슈 해결된 데이터 수집 : 녹음 동의서 활용</p>

- 음향 센서, 스마트 폰을 통해 수집된 위급상황 음성/음향 원천 데이터
- 음성, 음향, 위급상황 유형, 화자 분류(json)가 포함된 오디오 파일(.wav 등) 형태로서 CD 품질(16bit, Mono, 44.1kHz)의 위급상황 음성/음향 AI 학습 데이터셋
- 위급상황 음성/음향 원천 데이터 기반으로 구축된 위급상황 음성/음향 AI 학습 데이터셋과 AI 모델을 활용하여 위급상황 음성/음향 데이터셋의 유효성을 검증하고 위급상황 인식률 향상
- 고품질 위급상황 음성/음향 데이터셋을 활용한 응용 서비스 창출하고, 경찰청, 소방청, 복지기관, 의료 기관 등 유관기관과 연계하여 음성/음향 데이터 분석 기반 사회 안전 서비스 지원



1.2 데이터 포맷

○ 데이터 기본 포맷

- 위급상황 음성/음향 데이터는 16bit 44.1kHz의 CD음질인 WAV 파일로서 기본 포맷은 각 WAV + JSON 파일 쌍으로 구성되어 있음

○ 데이터 예시

예시	데이터 항목	JSON형식
	위급상황 오디오 음향 데이터 음성 데이터	<pre>{ "info": { "description": "위급상황 음향/음성 데이터셋", "url": "https://www.aihub.or.kr", "version": "1.0", "year": "2020", "contributor": "NIA2020", "date_created": "2020/12/01" }, "licenses": [{ "url": "https://www.aihub.or.kr", "name": "CC O" }], "audio": [{ "fileName": "1-234-41", "fileSize": "15.2", "duration": 6.45, "fileFormat": "wav", "sampleRate": "44.1KHz", "bitRate": "16", "recordingType": "Mono", "acquisitionType": "제작", "acquisitionDevice": "삼성갤럭시노트9", "acquisitionPlace": "실내" }], "annotations": [{ "audioType": "음성", "audio_id": "1.23.154", "area": { "start": 3.33, "end": 4.39 }, "categorys": ["category_01": "소방안전", "category_02": "화재", "category_03": "화재상황"], "description": "도움요청", "gender": "남자", "generation": "유아", "residence": "서울", "soundQuality": "정상", "dialect": "전라도" }, { "audioType": "음향", "audio_id": "1.23.155", "area": { "start": 5.23, "end": 7.39 }, "categorys": ["category_01": "소방안전", "category_02": "화재", "category_03": "화재상황"], "description": "제지하는소리", "soundQuality": "정상" }] }</pre>
	위급상황 제계 위급상황 분류 상황 음향 분류 화자 음성 분류	
	음성 데이터 화자 유형 정보 화자 원경 정보 화자 설명 정보	
	음향 데이터 음향 출처 정보 음향 원경정보 음향 설명 정보	

<위급상황 음성/음향 데이터 예시>

- 본 데이터셋은 음성과 음향으로 구성된 위급상황 오디오를 수집단계에서 설정된 분류 체계를 기반으로 위급상황 인식을 하기 위하여 음성 데이터의 경우 화자의 성별, 지역, 연령 등의 정보로 구성되어 있고, 음향의 경우 출처 및 환경, 설명 등의 정보로 구성되어 있다.

1.3 어노테이션 포맷

1.3.1 라벨링 정보

구분	라벨링	설명
음향/ 음성	영역	<ul style="list-style-type: none"> - 음향/음성 데이터 Cropping (상황에 오디오 시작시간 및 끝시간을 표시) - 위급상황 추정을 위한

1.3.2 라벨링 핵심 정보

구분	한글명	속성	내용
음향	영역	사운드 영역(area)	시작시간(start), 끝시간(end)
음성	영역	사운드 영역(area)	시작시간(start), 끝시간(end)
		발화자 속성	성별(gender), 연령(generation), 지역(residence) 등

- 위급상황을 이상 유형을 파악하기 위해서는 음향 및 음성의 이상 형태를 파악해야 한다. 녹음된 사운드에서 상황 영역인 오디오의 시작 시간 및 끝 시간을 표기함
- 음향 및 음성을 기반한 상황 검출을 위하여 상황정보와 음성의 경우 발화자의 성별, 연령, 지역 등의 정보를 입력함

1.3.2.1 사운드 영역

- 사운드 영역은 위급 또는 일반 상황 객체 전체가 사운드영역 담기면서 각각의 시작 및 끝선들은 오디오 상황과 맞닿아 있어야 함
- 사운드 영역의 가로 길이는 위급상황 또는 일반상황의 시작과 끝을 나타내는 오디오 파형이어야 함



1.3.2.2 데이터 제공항목 및 항목별 속성정보

구분	NO	제공항목			속성정보		예시
		항목명	필수 선택	설명	허용 데이터	타입	
licenses	1	라이센스 명칭	필수	AI 학습용 데이터를 활용하기 위한 규정		String	CC 0
	2	라이센스 주소	필수	해당 라이센스를 검색, 다운로드할 수 있는 사이트 주소		String	https://www.aihub.or.kr
info	1	정보 설명	필수	AI 학습용 위급상황 음성/음향 데이터 세트 관련 설명 내용		String	위급상황 음향/음성 데이터셋
	2	정보 주소	필수	AI 학습용 위급상황 음성/음향 데이터 세트를 다운로드할 수 있는 주소		String	https://www.aihub.or.kr
	3	정보 버전	필수	AI 학습용 위급상황 음성/음향 데이터 세트 버전		Number	1.0
	4	정보 제작년도	필수	AI 학습용 위급상황 음성/음향 데이터 세트 제작년도		Number	2020
	5	정보 제작자	필수	AI 학습용 위급상황 음성/음향 데이터 세트를 제작한 기관 정보		String	NIA2020
	6	정보 생성일자	필수	AI 학습용 위급상황 음성/음향 데이터를 생성한 일자		String	2020-12-24
audio	1	오디오 명칭	필수	원천데이터 오디오 명칭	16종분류_오디오번호	String	4.폭력범죄_5 44193.wav
	2	오디오 크기	필수	원천데이터 오디오 크기		Number	740924
	3	오디오 길이	필수	원천데이터 오디오 길이		Number	8.4
	4	오디오 종류	필수	원천데이터 오디오 종류 voice = 음성, sound = 음향	voice,sound	String	voice
	5	오디오 포맷	필수	원천데이터 오디오 포맷	wav	String	wav
	6	오디오 샘플레이트	필수	원천데이터 오디오 샘플레이트	44.1Khz	String	44.1Khz
	7	오디오 비트레이트	필수	원천데이터 오디오 비트레이트	8 이상	Number	16
	8	오디오 녹음방식	필수	원천데이터 오디오 녹음방식	Mono,Stereo	String	Mono
	9	오디오 취득방식	필수	원천데이터 오디오 취득방식		String	제작
	10	오디오 취득장치	필수	원천데이터 오디오 취득장치		String	갤럭시탭S6
	11	오디오 취득장소	필수	원천데이터 오디오 취득장소	실내,실외	String	실내

구분	NO	제공항목			속성정보		예시
		항목명	필수 선택	설명	허용 데이터	타입	
annotations	1	어노테이션 식별자	필수	원천데이터에 라벨링, 카테고리, 음성정보를 담고있는 어노테이션 식별자	16종분류_오디오번호_label	String	4.폭력범죄_5_44193_label.wav
	2	영역	필수	어노테이션 영역		Object	
	2-1	시작	필수	어노테이션 시작값		Number	1.92
	2-2	종료	필수	어노테이션 종료값		Number	4.32
	3	카테고리	필수	어노테이션의 카테고리		Object	
	3-1	대분류	필수	카테고리 대분류	치안안전,소방안전, 사고발생,자연재해, 일반(위급),일반(정상)	String	사고발생
	3-2	중분류	필수	카테고리 중분류	강제추행(성범죄), 강도범죄,절도범죄, 폭력범죄,간힘, 전기사고,가스사고, 화재,응급의료, 태풍/강풍,지진,낙상, 붕괴사고,도움요청, 실내,실외	String	화재
	3-3	소분류	필수	카테고리 소분류		String	화재를 발견한 상황
	4	설명	필수	어노테이션 영역에 대한 설명 음성일 경우 문장이, 음향일 경우 상황설명이 들어감		String	불이야
	5	오디오 종류	필수	어노테이션 영역의 오디오 종류	voice,sound	String	voice
	6	성별	선택	어노테이션 음성 주인의 성별 오디오가 음성일때만 존재함	male,female	String	male
	7	연령	선택	어노테이션 음성 주인의 연령대 오디오가 음성일때만 존재함	adult,child, teenager, senior,else	String	adult
	8	지역	선택	어노테이션 음성 주인의 거주지역 오디오가 음성일때만 존재함		String	서울
	9	역양	선택	어노테이션 영역에 대한 음질 오디오가 음성일때만 존재함	standard,gangwon,chungcheong,kyongsang,jeolla,jeju,else	String	standard
	10	음질	필수	어노테이션 영역에 대한 음질	정상, 노이즈	String	정상
	11	거리	필수	오디오가 녹음된 거리	근접, 중간, 원거리	String	중간

1.3.2.3 어노테이션 데이터 포맷

<위급상황 음향/음성 필수 어노테이션 데이터 포맷>

No	항목		필수 여부	길이	타입	비고
	한글명	영문명				
1	정보	info			Object	
1-1	정보 설명	info.description	Y	100	String	
1-2	정보 주소	info.url	Y	100	String	
1-3	정보 버전	info.version	Y	100	Number	
1-4	정보 제작년도	info.year	Y	100	Number	
1-5	정보 제작자	info.contributor	Y	100	String	
1-6	정보 생성일자	date_created	Y	100	String	
2	라이선스	license			Object	
2-1	라이선스 주소	license.url	Y	100	String	
2-2	라이선스 명칭	license.name	Y	100	String	저작권
3	오디오	audio			Object	
3-1	오디오 명칭	audio.fileName	Y	100	String	
3-2	오디오 크기	audio.fileSize	Y	100	Number	
3-3	오디오 길이	audio.duration	Y	10	Number	
3-4	오디오 종류	audio.audioType	Y	100	String	
3-5	오디오 포맷	audio.fileFormat	Y	100	String	
3-6	오디오 샘플레이트	audio.sampleRate	Y	100	String	
3-7	오디오 비트레이트	audio.bitRate	Y	100	Number	
3-8	오디오 녹음방식	audio.recordingType	Y	100	String	
3-9	오디오 취득방식	audio.acquisitionType	Y	100	String	
3-10	오디오 취득장치	audio.acquisitionDevice	Y	100	String	
3-11	오디오 취득장소	audio.acquisitionPlace	Y	100	String	
4	어노테이션	annotations			List	
4-1	어노테이션 식별자	annotations.audio.id	Y	100	String	
4-2	영역	annotations.area	Y	10	Object	
4-3	시작	annotations.area.start	Y	10	Number	
4-4	종료	annotations.area.end	Y	10	Number	
4-5	카테고리	annotations.categories	Y	10	Object	
4-6	대분류	annotations.category_01	Y	100	String	6분류
4-7	중분류	annotations.category.category_02	Y	100	String	16종
4-8	소분류	annotations.category.category_03	Y	100	String	상황
4-9	설명	annotations.note	Y	100	String	음성/ 음향
4-10	오디오 유형	annotations.audioType	Y	100	String	음향/ 음성
4-11	성별	annotations.gender	N	100	String	음성
4-12	연령	annotations.generation	N	100	String	음성
4-13	지역	annotations.residence	N	100	String	음성
4-14	억양	annotations.dialect	N	100	String	음성/ 음향
4-15	음질	annotations.soundQuality	Y	100	String	음성/ 음향
4-16	거리	annotations.soundDistance	Y	100	String	음성/음 향

1.4 데이터 구성

1.4.1 원천 및 학습 데이터 구성

구분	구 성		데이터 구분	비고
음향/음성	오디오 클립	*.wav file	원천 데이터	수집 파일명에 어노테이션 정보 입력
	라벨링	*.json file	학습 데이터	
	오디오 클립	*.wav file	학습 데이터	

항목	내용	
파일명	3.절도범죄_606014.wav	
플레이 시간	20.9	
라이센스	CC 0(저작권없음)	
대분류	치안안전	
중분류	절도범죄	
소분류(상황)	치안안전_도둑이야_비명소리	
내용	도둑이 침입한 상황	
어노테이션 시작 시간	5	
어노테이션 종료 시간	15.9	
유형	음성	음향
성별	여자	-
연령	성인	-
지역	서울	-
억양	표준	-

구성예	정보	<pre>{ "info": { "description": "위급상황 음향/음성 데이터셋", "url": "https://www.aihub.co.kr", "version": 1.0, "year": 2028, "contributor": "NIA2020", "date_created": "2020-12-30" }, "license": { "url": "https://www.aihub.co.kr", "name": "CC 0" } }</pre>	어노테이션	<pre>{ "annotations": [{ "audio_id": "1-3-8_371403_1", "area": { "start": 1.92, "end": 2.86 }, "categories": ["category_01": "치안안전", "category_02": "절도범죄", "category_03": "도둑이 침입한 상황"], "note": "도둑이 침입한 상황", "audioType": "voice", "gender": "male", "generation": "adult", "residence": "서울", "dialect": "standard", "soundQuality": "정상" }, { "audio_id": "1-3-8_371403_2", "area": { "start": 5.59, "end": 6.52 }, "categories": ["category_01": "치안안전", "category_02": "절도범죄", "category_03": "도둑이 침입한 상황"], "note": "도둑이 침입한 상황", "audioType": "voice", "gender": "male", "generation": "adult", "residence": "서울", "dialect": "standard", "soundQuality": "정상" }] }</pre>
	라이센스	<pre>"license": { "url": "https://www.aihub.co.kr", "name": "CC 0" }</pre>		
	오디오	<pre>"audio": { "fileName": "1-3-8_371403.wav", "fileSize": 772676, "duration": 8.76, "audioType": "voice", "fileFormat": "wav", "sampleRate": "44.1kHz", "bitRate": 16, "recodingType": "Mono", "acquisitionType": "제작", "acquisitionDevice": "삼성 갤럭시탭S6", "acquisitionPlace": "실내" },</pre>		

※ Free 음원 사이트 등에서 상황에 맞는 음원을 다운로드했을 경우 라이센스 및 저작 관련 라이센스 표기 필요

1.5 데이터 통계

1.5.1 데이터 구축 규모

유형		구축량	비고
원천 데이터	음향 및 음성	3,500 시간	음향 및 음성 3,500시간
학습용 데이터	상황 객체	50,000 개	어노테이션 되는 객체 수

○ 데이터 분류에 따른 구축 규모

- 위급상황 데이터 수집 시 위급상황을 인지할 수 있는 각각의 음성과 음향이 복합적으로 구성된 16종의 분류체계를 구성하고 그에 따른 시나리오를 구성하여 음성/음향 데이터를 균일하게 구축, 위급상황(일반)에 자유발화 포함

위급상황 분야	분류체계	구축규모 (시간)	비율(%)
치안안전	강제추행	100	2.9
	강도범죄	100	2.9
	절도범죄	100	2.9
	폭력범죄	200	5.7
소방안전	갇힘	100	2.9
	전기사고	200	5.7
	가스사고	150	4.3
	화재	450	12.9
	응급의료	450	12.9
자연재해	태풍/강풍	300	8.6
	지진	100	2.9
사고발생	낙상	150	4.3
	붕괴사고	100	2.9
일반(위급)	도움요청	300	8.6
일반(정상)	실내	500	14.3
	실외	200	5.7
합계		3,500	100%

1.5.2 데이터 분포

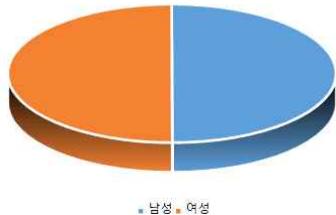
- 위급상황 음성음향 데이터 구축 시간 : 3,500

- 데이터 수집 참여자 성별 및 연령 비율

성별 인원(명)	
남성	여성
250	250

연령대 인원(명)			
유아	청소년	노인	성인
30	100	100	270

성별 인원(명)



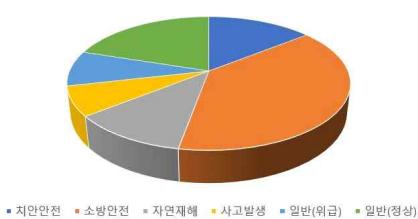
연령대별 인원(명)



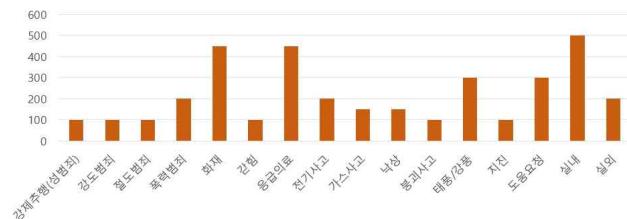
- 6분류 16종 위급상황 음성음향 데이터 시간 : 3,500

분류	치안안전				소방안전				자연재해		사고활동		일반(위급)	일반(정상)		
	강제 추행	강도 범죄	절도 범죄	폭력 범죄	갇힘	전기 사고	가스 사고	화재	응급 의료	태풍/ 강풍	지진	낙상	붕괴 사고			
음성/ 음향	100	100	100	200	100	200	150	450	450	300	100	150	100	300	500	200

위급상황 분류체계



위급상황 음성/음향 데이터



1.5.3 기타 활용 통계

- 해당 사항 없음 -

1.6 원시데이터 특성

1.6.1 위급상황 원시 데이터 특징

- 위급상황 원시데이터는 사운드 AI 기술에 사용될 범용적 청각 인공지능 데이터로 공공데이터 활용, 직접녹음, 크라우드 소싱, 인터넷 공개데이터 등 수집할 수 있으며 기본적으로 오디오 데이터는 그 용량이 고화질 비디오만큼 아니지만, 꽤 큰 편이다. 일반적으로 오디오는 꽤 높은 샘플링주파수를 요구하는데 여기서 샘플링주파수란 소리의 아날로그 정보를 디지털로 바꿀 때 1초를 몇 번 정보를 저장할지를 지칭함
- 환경음의 경우 다양한 주파수 대역으로 정보들이 퍼져 있어서 음성인식보다 훨씬 높은 샘플링주파수가 필요한데, 그 이유는 샘플링주파수가 높을수록 분석 가능한 주파수대역대가 비례해서 넓어지기 때문임
- 또한 녹음 기기별 특성이 매우 아주 다르다는 점도 유의해야 한다. 특성은 주파수 특성이나 감도 등도 있고, 최근에는 음성인식을 위해 간단한 소음 제거 전 처리가 들어가 있는 경우가 있음
- 스마트 폰을 사용하는 경우에도 사용하는 앱마다 기본 세팅이 다를 수 있을 뿐만 아니라 음원 퀄리티에 연관된 수많은 정보가 없는 경우가 많거나 실제와 다르게 메타데이터에 적혀 있는 경우도 많아서 딥러닝의 약점 중 하나인 오버피팅 문제를 언제나 염두하고 시스템을 구성함
- 이렇게 다양한 특성들을 고려해야 하는 이유는 분석하고자 하는 소리가 마치 사람이 소리를 자연스럽게 듣고 이해하듯 환경과 상관없이 어디서든 다양한 소리를 전부 이해할 수 있는 범용적인 청각 인공지능기술을 목표로 만들고 있음

1.6.2 제약조건

- 소방청(119) 및 경찰청(112) 콜센터의 위급상황 녹취록의 경우 개인정보 보호법에 의하여 비공개 데이터로 관리됨이 원칙
- 소방청 및 경찰청의 데이터를 확보하기 위해서는 사전 MOU 등을 체결하고 내부 데이터 공개 위원회를 소집하여 개방 여부 결정이 최소 3개월 이상이 소요되어 본 과제내에서 추진하는 것은 물리적으로 불가하여 사전 MOU체결을 통해 이미 과제를 수행한 기관이 소방데이터를 확보하고 있으나, 이 경우에도 데이터의 외부유출이 엄격하게 규제되고 있는 상황을 고려하여, 기관간의 NDA 체결을 전제로 기관 방문 및 청취를 허가받아 대역을 통한 유사데이터 생성으로 본 사안을 해결하는 것이 최선임
- 음성데이터 중 전문 성우로 부터 취득한 데이터는 일반인의 음성발화와 비교해 볼 때 드라마틱한 발성으로 인해 현실감이 결여되는 경우가 있어 음성 녹화 시 고려 필요
- 음성데이터 중 유아, 아동 및 노년층 연령대의 음성발화는 발성 연기력이 부족하여 높음 품질의 음성데이터를 수집하는데 한계가 있음

※ 소방청 및 경찰청 위급상황 음성 녹취록 공개 요청

구분	아이엠알 컨소시엄 소방청(119) 및 경찰청(112) 데이터 수집을 위한 활동 내역
요청 기관	- 서울 마포경찰서, 서울특별시 경찰청, 광주광역시 경찰청, 소방본청, 116종합상황방재센터, 경기도 북부 소방본부, 광주광역시 소방본부 등
요청 내용	- 위급상황 음향/음성 AI 데이터 구축사업 경찰청/소방당국 협조 요청 1안) 위급상황 음성/음향 데이터의 원천데이터 요청 2안) 위급상황 음성 데이터의 원천 녹취록[문서] 데이터 요청 3안) 소방청 위급상황 음성 데이터 참조를 위한 소방청과 시나리오 협업
결론	- 112, 119콜센터 녹취록 열람 및 반출은 비공개 데이터로 민간기업 공개 불가 - 공공기관 요청시 사전 심의 위원회 소집하여 진행

※ 소방청 및 경찰청 위급상황 음성 녹취록 공개 법률 검토

경찰청 112신고내역 정보공개 관련 법률(시행 2020. 1. 1.)

112 신고센터 운영 및 신고처리 규칙 제 4장 13조

① 112신고는 그 긴급성과 출동필요성에 따라 다음 각호와 같이 구분하며, 경찰청장은 각 유형의 세부 기준을 따로 정할 수 있다.

1. code 1 신고 : 다음 각목의 사유로 인해 최우선 출동이 필요한 경우

- 가. 범죄로부터 인명·신체·재산 보호
- 나. 심각한 공공의 위험 제거 및 방지
- 다. 신속한 범인검거

2. code 2 신고 : 경찰 출동요소에 의한 현장조치 필요성은 있으나 제1호의 code1 신고에 속하지 않는 경우

3. code 3 신고 : 경찰 출동요소에 의한 현장조치 필요성이 없는 경우

② 112요원은 신고내용을 합리적으로 판단하여 신고유형을 구분하여 업무처리를 하여야 한다.

112 신고센터 운영 및 신고처리 규칙 제 4장 26조

2항 112 신고 접수 및 무선지령내용 녹음자료는 24시간 녹음하고 3개월간 보존

3항 기타 문서 및 일지는 「공공기관의 기록물 관리에 관한 법률」에서 정하는 바에 따라 보존

- 소방당국 119신고내역 정보공개 관련 법률(시행 2020. 5. 25.)

119 종합상황실 표준 운영규정 제 5장 기록유지 및 장비관리 등 제 25조(자료제공 및 공개)

1. 자료의 제공, 정보의 공개 거부에 대한 사항은 공공기관의 정보공개에 관한 법률(이하 "정보공개법"이라 한다.)에 따라 절차를 준수하여 제공하여야 한다.
2. 119신고접수 등 재난 상황 기록은 소방업무, 소송, 수사, 감사, 국회에서의 요구 외 목적으로 사용하거나 외부에 유출하여서는 아니된다.
4. 119신고통화내역 정보공개 요구 시 녹취록으로 작성하여 제공하는 것을 원칙으로 하되, 단 다른 법률에 특별한 규정이 있거나 통화내역 녹취파일 외에 당해 목적을 달성할 방법이 없다고 판단될 경우에는 목소리를 식별할 수 없도록 변조하여 정보공개법에 따라 절차를 준수하여 제공할 수 있다.

1.6.3 속성

구분	데이터 내용
데이터 품질	- 음성, 음향, 위급상황 유형, 화자 분류(파일명에 분류 가능)가 포함된 오디오 파일(.wav 등) 형태로서 16bit, Mono, 44.1kHz (CD품질) 확보
데이터 크기	- 16클래스(중분류), 1,000건 이상의 상황, 유형(복합 음성/음향)별 3,500시간 이상의 데이터 구축 - 음장의 분포를 기반으로 다양한 위치, 높이, 지향성/비지향성 등을 고려하여 녹음 센서에서 수집된 음성/음향 데이터를 수집

○ 원시데이터 획득에 필요한 녹음 장비

	샘플링	44.1khz
	해상도	16bit
	오디오 포맷	PCM WAV

- 추후 반영사항

- 수집된 음성, 음향 데이터에 대한 주파수 변환을 시행하고, 이를 주파수 정보에 대한 평균, 표준편차, 분산을 계산하여 95% 신뢰도 이내로 들어오는 데이터만을 대상으로 데이터 수집 범위로 설정함
- 오디오 음원에 대한 주파수 변환을 통해 특이한 주파수 대역을 분리하여 분석하고, 이 대역에 대한 별도 청취를 통하여 구분되는 상황과 그렇지 않은 상황으로 분류하여 구분되지 않는 상황데이터에 대한 배제 처리함

○ 위급상황 시 기술적 요소

- 위급 상황에서의 음성/음향 인식 위한 구체적인 기술적 요소를 제시함

구분	대상	기술적 사항
음성	디바이스에서 음성의 크기 (마이크 도달 기준)	70 dB SPL
	키워드의 길이	5음절 이상, 600 ms 이상 (권고: 5음절 이상, 1초)
	주변 소음 대비 음성 SNR	12 dB
	음성 인식 방법	화자 독립
	특정 음원	음성긴급전화 모사 음성/음향, 위급상황 음성 연출 소리
음향	디바이스에서 음향의 크기 (마이크 도달 기준)	70 dB SPL
	음향의 길이	2000 ms 이상 (권고: 2초)
	주변 소음 대비 음성 SNR	12 dB
	음향 인식 방법	음향 독립
	특정 음원	음성긴급전화 모사 음성/음향, 위급상황 음향 연출 소리

- 디바이스에서 음성 및 음향의 크기 (마이크 도달 기준) : 70dB SPL

- 디바이스에서 응답해야 하는 음성의 크기는 일상생활 속에서 발생하는 소음에 대하여 규정한 생활소음 규제를 참고한 환경적 관점 등을 반영함

※ dB SPL(sound pressure level, 음압) : 음압(소리의 매질인 공기의 압력의 실효출력)의 단위는 파스칼(Pa)이지만, $2 \times 10 - 5$ Pa를 기준값(0dB SPL)으로 함

생활소음 규제

생활소음이란 일상 생활 중에 자주 접하게 되는 주변의 소음이라고 생각할 수 있다.
『소음진동·관리법』에 따라 생활소음 규제대상은 확성기에 의한 소음, 공장 및 사업장의 작업소음, 기타 심야의 계속적 또는 반복적 소음으로 구성된다.

대상지역	소음원	시간대별		
		야침, 저녁 (05:00~07:00, 18:00~22:00)	주간 (07:00~18:00)	야간 (22:00~05:00)
가. 주거지역, 녹지지역, 관리지역 中 취락지구·주거개발전용지구 및 관광휴양개발전용지구, 자연환경 보전지역, 그 밖의 지역에 있는 학교·종합병원·공공도서관	확성기	옥외설치 온내에서 옥외로 소음이 나오는 경우	60이하 50이하	65이하 55이하
		공장	50이하	55이하
	사업장	동일 건물	45이하	40이하
		기타	50이하	55이하
		공사장	60이하	65이하
나. 그 밖의 지역	확성기	옥외설치 온내에서 옥외로 소음이 나오는 경우	65이하 60이하	70이하 60이하
		공장	60이하	65이하
	사업장	동일 건물	50이하	55이하
		기타	60이하	65이하
		공사장	65이하	70이하

- 음성 키워드의 길이 : 5음절 이상, 600ms이상 (권고 : 5음절 이상, 1초)
 - "TTA 스마트 폰 음성인식 긴급전화 서비스 해설서"을 참조하여 최소한으로 5음절 이상의 600ms이상의 음성키워드 기술규격으로 도출함
- 음향의 길이 : 2000ms 이상 (권고 : 2초 이상)
- 주변 소음 대비 음성 SNR : 12dB
 - 신호대잡음비(SNR)는 신호음과 잡음 사이의 비율관계를 의미하며, dB로 표시한다. 본 표준에서는 음성과 잡음 사이의 비율이라 볼 수 있으며, "TTA 스마트폰 음성인식 긴급전화 서비스 해설서" 의견을 바탕으로 기술규격은 12dB로 협의됨
 - 예를 들어, 음성이 70dB라면 잡음이 58dB 이상이 됐을 경우에 신호대잡음비가 12dB이하가 되기 때문에 단말에서 음성 및 음향 인식이 안 될 가능성성이 있음

주변 소음 대비 음성 SNR 계산근거

$$\text{신호음(음성)dB} - \text{잡음dB} = \text{신호대잡음비dB}$$

$$70\text{dB} - 58\text{dB} = 12\text{dB}$$

- 음성 인식 방법 : 화자 독립
 - 재난·사고 상황에서는 최대한 짧은 시간 동안 빠른 구조 지원이 요구되어지고 범죄 상황처럼 신고자의 음성만을 인식하는 기술이 상황적으로 필요하지 않기 때문에 스마트 폰 단말의 음성 인식 방법으로 '화자 독립'을 채택하였다.
- 특정음원 : 예) '음성인식긴급전화'
 - 앞서 용어정의 4.1에서 정의하였듯이 특정음원은 음성 인식 긴급전화 서비스를 통하여 연결된 전화라는 음성 멘트로서 긴급구조기관 및 경찰청 응답원이 처음 전화를 수신하는 시점에서 재생되는 음원
 - 특정음원의 음성멘트는 국내 3대 이동통신사의 의견 수렴결과, 음원을 전송하여 응답원이 통화 수신하는 시점에서 재생하는 것이 기술적으로 구현 가능하며 전송 멘트에 대한 규정은 없음

○ 음향 및 음성 주파수 범위

- 주파수 음역

주파수 범위	설명	주파수 범위	설명
16Hz 이하	초저주파수	16 ~ 20,000Hz	인간의 가청 음역
200 ~ 600Hz	저주파수	125 ~ 8,000Hz	일상생활 음역
600 ~ 1,500Hz	중간주파수	1,000 ~ 5,000Hz	감도 좋은 음역
1,500 ~ 20,000Hz	고주파수	250 ~ 4,000Hz	회화 음역
20,000Hz 이상	초음파	500 ~ 2,000Hz	중요 회화 음역

- 환경음의 소리크기 수준

데시벨	환경음 설명	환경음 소리 크기 수준	
-10dB ↑	청각 과민자만 듣는 소리	청각과민 ↑	
-5dB ↑	청각 과민자만 듣는 소리		
0dB	일반인이 들을 수 있는 최저 소리		
10dB	나뭇잎 바스락 소리		
20dB			
30dB	속삭이는 소리		
40dB	가정의 평균 생활 소음		
50dB	교실, 사무실, 냉장고 소리		
60dB	일상대화, 에어컨 소리		
70dB	전화벨 소리		
80dB	지하철, 대도시의 거리, 헤어드라이기 소리	듣기 고통스러운 수준 ↓	
90dB	기차, 공장, 전기톱 소리, 오토바이, 잔디 깍는 소리		
100dB			
110dB	자동차 경적, 공사장 소리		
120dB	락 악단, 비행기 이륙 소리		
130dB	제트기 엔진, 폭죽 소리, 총 소리		
140dB			
150dB ↓		고막파열 수준 ↓	

- Mel-Spectrogram 변환 이미지로 변환하여 MFCC Feature값 추출한 후 Feature 값에 따른 상황 분류
 - 수집된 음성 음향 원천 데이터(.wav)를 Mel-Spectrogram 변환 이미지로 변환 후 MFCC Feature 값 추출한 후 Feature 값에 따른 상황 분류 진행함

Mel spectrogram 이미지 변환	Labeling	MFCC 이미지 추출 feature 값
주파수	Mel spectrogram 이미지	직접 청취

- Feature값 추출 결과 분류체계와 1:1 맵핑되는 Feature값 찾기 어려울 뿐 아니라, 같은 상황으로 분류되는 음원파일임에도 불구하고 서로 다른 Feature값 나타나 정량적 수치에 따른 상황 분류보다는 직접 청취에 따른 상황 분류가 더 적합할 것으로 보여 크라우드 워커 레이블러를 통한 위급 상황 분류를 진행함
- 여러 분류에 해당하는 음원이 혼재된 환경에서 수집하였을 경우 주요 음원과 서브 음원 분류에 따른 음원 정확도 분석을 진행, 시간, 주파수, 스펙트럼 음압레벨 및 실제 음원 청취에 따라 분류함

○ 위급상황 상황 소방청 데이터 수집계획(안)

① 추진목표

- 기 수행과제 원천데이터를 통한 위급상황 실제 데이터 청취/녹취록 작성
- 원천 데이터 기반 유사 시나리오 작성/크라우드 소싱을 통한 다양성 확보

② 추진사항

- 협력 : 세종대학교 AI센터
- 내용 : 원천 데이터 청취 및 녹취록 작성
- 추진근거
 - 경찰청/소방당국의 정보공개 불가로 인한 유사 수행과제 Data 활용
 - 연구명 : 음성/음향분석 기반 상황판단 솔루션 기술개발
 - 주관 : 한국과학기술연구원
 - 참여연구기관 : 세종대학교, 한양대학교 등

③ 유사 데이터 구축

- 유사 시나리오/데이터 제작
 - 다양한 유사 시나리오를 통한 폭넓은 가용성(Availability) 확보¹⁾
 - 다양성(Variety) 확보를 위한 크라우드 소싱 활용

1.7 기타정보

1.7.1 포괄성

○ 치안안전, 소방안전, 자연재해, 사고발생, 일반(위급), 일반(정상) 5가지 대분류로 분류함

○ 주요 위급상황

- 강제추행, 강도범죄, 절도범죄, 폭력범죄, 갇힘, 전기사고, 가스사고, 화재, 응급의료, 태풍/강풍, 지진, 도로교통사고, 낙상, 봉괴사고, 폭발사고, 도움요청 등 행정안전부 재난안전 및 위급상황에 대한 통계가 높음

구분	데이터 내용
포괄성	<ul style="list-style-type: none"> - 16종 1,000건 이상의 상황에 대하여 항목, 대상(음향/음성), 형식(단독 데이터, 조합 데이터, 소음 데이터), 지역(억양, 실내, 실외), 사람(남성, 여성, 노인, 성인, 아동, 기타), 공간(안방, 거실, 주방, 화장실, 사무공간) 등에 대한 내용으로 위급상황별 시나리오를 구성하여 음향/음성 데이터셋 구축

1.7.2 독립성

- 데이터 획득은 온라인 음원 프리 라이센스 데이터, 크라우드 워커, 직접 녹음을 통하여 수집되었으며 개인정보 및 민감정보는 포함되어 있음
- 본 데이터 수집 과제를 통하여 수집한 음향/음성은 라이센스 프리 및 개인정보활용 등의 정식 계약 및 공개 협약을 통해 수집된 음향 및 음성이므로 원천 공개 가능함

- 저작권 침해를 방지하기 위하여 수집된 데이터는 AI Hub에 공개된다는 내용을 수집 대상 양식장에 고지하고 동의하에 데이터를 수집 및 공개하였음
- 본 데이터 획득은 본 과제의 수행기관이 법·제도, 규정, 지침, 시행규칙 등을 참고하여 도메인 전문가 와의 협업을 통해 정의를 하고 직접 수집한 데이터이므로 독립성이 있음을 고지함

○ 데이터 내용

구분	데이터 내용
독립성	<ul style="list-style-type: none"> - 통신비밀보호법으로 인해 녹취기기 등 설치를 위해서 녹음 대상자의 허가가 필요하므로 이를 위한 녹음 허가 동의서 작성 및 요청 - 특정인 음성, 성명, 전화번호, 주소 등의 개인정보는 음향/음성 데이터셋 구축에 있어서 배제하는 것을 원칙으로 수행 - 녹음동의서에 대한 충분한 설명 및 서명 날인 확보를 통하여 개인정보에 대한 기본적인 이슈를 해결 - 저작권자의 허가가 필요하므로 데이터의 사용 및 공개 가능 여부에 확인을 선행하여 침해 이슈 해결 - 저작권자의 권리를 침해할 수 있는 데이터는 사전에 배제하는 것을 원칙으로 수행 - 출처를 명시하여 사용이 가능한 제한적인 저작권은 필수사항을 이행하여 기본적인 이슈를 해결

○ 이슈 해결 방안

구분	이슈 해결 방안
개인 정보	<ul style="list-style-type: none"> - 통신비밀보호법으로 인해 녹취기기 등 설치를 위해서 녹음 대상자의 허가가 필요하므로 이를 위한 녹음 허가 동의서 작성 및 요청 - 특정인 음성, 성명, 전화번호, 주소 등의 개인정보는 음향/음성 데이터셋 구축에 있어서 배제하는 것을 원칙으로 수행 - 녹음동의서에 대한 충분한 설명 및 서명 날인 확보를 통하여 개인정보에 대한 기본적인 이슈를 해결
저작권	<ul style="list-style-type: none"> - 저작권자의 허가가 필요하므로 데이터의 사용 및 공개 가능 여부에 확인을 선행하여 침해 이슈 해결 - 저작권자의 권리를 침해할 수 있는 데이터는 사전에 배제하는 것을 원칙으로 수행 - 출처를 명시하여 사용이 가능한 제한적인 저작권은 필수사항을 이행하여 기본적인 이슈를 해결
유해성	<ul style="list-style-type: none"> - 유해성을 확인 및 검증하여 사회적 민감 정보에 대한 이슈 해결 - 성차별, 정치적 성향, 종교 등 사회적 민감 정보를 포함하는 경우 배제하는 것을 원칙으로 수행

1.7.3 유의사항

- 본 데이터를 사용함에 있어서, 본 데이터는 공개데이터, 크라우드소싱, 직접녹음 등을 기반으로 수집 및 라벨링이 되고, 데이터 검수 기관에서 데이터 검수를 통과한 데이터를 공개를 하였지만, 특정 데이터는 예러를 포함할 수 있음을 유의하며 데이터를 사용하기를 권장함

- 위급상황(치안안전, 소방안전 등)을 라벨링 하는 데이터는 본 지원과제에서 제작한 데이터가 세계 최초이기에, 특정 세계 표준을 따르지 않고, 본 지원과제의 가이드라인에 맞춰서 제작되었음을 유의하며 사용하기를 권장함

1.7.4 관련 연구

- 음성 및 음향 데이터를 기계 학습 모델에 적용하여 인식률을 높이기 위한 다양한 믹싱 및 증강기술을 이용함

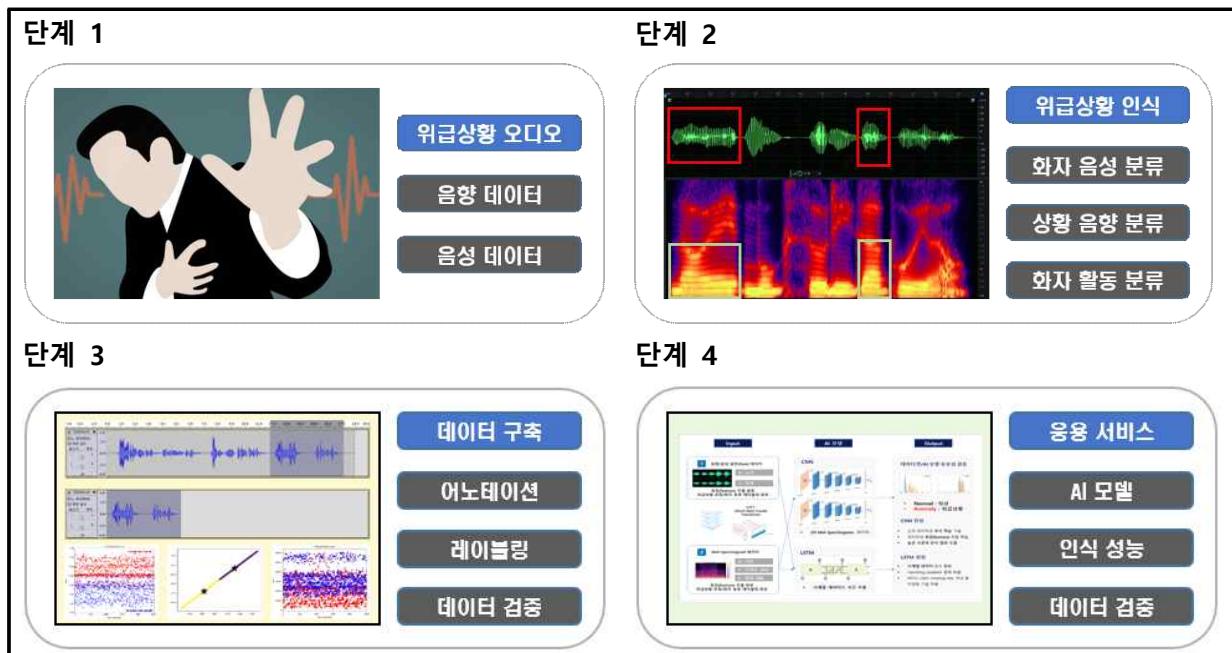
순번	분류	제목	저자	출처/학회	내용
1	계재 논문	Learning from between-class examples for deep sound recognition	Yuji Tokozume 외 2명	ICLR2018	<p>Deep sound의 인식을 위해 서로 다른 class에 속한 두 가지 소리를 랜덤하게 혼합(mixing)하여 Between-class 소리를 생성한 후 Deep learning을 적용한 학습 모델에 적용 및 훈련하여 mixing 비율을 출력. 결과적으로 다양한 소리 인식 성능이 개선됨</p>
2	발표 논문	Shuffling and mixing data augmentation for environmental sound classification	Tadanobu Inoue 외 7명	Detection and classification of Acoustic Scenes and Event 2019	<p>스마트 스피커를 이용해 독거노인등의 비정상적 활동 등을 알릴 수 있는 기능을 위해 기계 학습을 이용해 데이터셋의 두가지 기준 소리를 섞어서 믹싱한 데이터를 CNN기법의 기계 학습 모델에 적용시켜 기본 84.5%에서 F1 score 89.95%로 향상시킨 데이터 증강기술을 개발함</p>
3	발표 논문	SpeechMix – Augmenting deep sound recognition using hidden space interpolations	Amit Jindal 외 5명	Interspeech 2020	<p>자동 음성인식을 위한 데이터 증강기법으로서 Neural 모델의 hidden layer로부터 2~3개의 음성 샘플을 조합하여 새로운 증강 음성 샘플을 생성, 훈련기반의 interpolation을 적용하지 않은 방법을 제시</p>

2. 데이터 구축 가이드

2.1 데이터 구축 개요

- 독거노인, 중증 장애인 등 사회 내 안전 취약계층을 포함한 안전 범죄로부터 위험에 노출될 수 있는 모든 국민들의 위급 상황 발생 시 안전 확보 기술 개발을 위해 위급상황에 따른 음성 및 음향 데이터 확보가 필요하며 이 가이드라인의 목적은 지능형 위급상황 인식의 학습 개발에 활용하기 위한 데이터 셋을 구축하고 공개하여 이를 바탕으로한 선순환 생태계를 조성하며, 이를 적용하기 위한 가이드라인을 제공하기 위함
- 가이드라인의 적용은 음향 디바이스로부터 입력되는 음향 및 음성에서 이상상황을 탐지하는 지능형 위급상황 인식 AI의 학습 개발에 활용하기 위한 데이터셋을 구축하고 시범 모델을 개발하여 데이터 셋을 검증하는 것이 범위이고, 최종적으로 인공지능 학습을 위한 최적의 위급상황 음향/음성 데이터셋 구축 및 공개를 목표로 함
- 활용범위로 범죄신고 및 긴급사고, 화재·구조·구급 등과 같은 영역에서 음향/음성 분석기반 상황 판단 기술이 활용될 경우 긴급 상황에서 빠른 상황판단을 통한 인명구조 및 안전 사회 유지에 효율적임
- 인공지능 학습용 데이터셋을 구축하고 검증, 그리고 이를 공개하고 활용하는 방법에 대한 지침을 서술함
- 인공지능 학습용 데이터셋을 구축에 관한 기술과 구축한 데이터셋에 대한 데이터 수집, 정제, 가공, 검사, 활용에 대한 방안을 제공함

○ 데이터 구축 단계별 가이드



○ 데이터 구축 프로세스 가이드



○ 데이터 구축 상세 프로세스 가이드

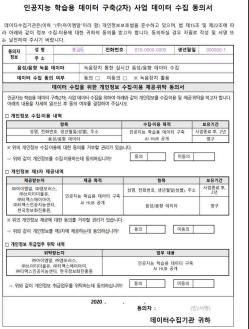
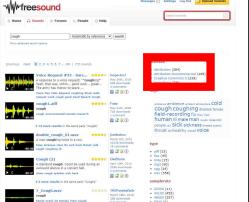
단계	세부 절차	설명
1. 데이터 수집	1-1. 원천 데이터 선정	원천 데이터를 수집하는 대상 설정 (저작권 확인, 위급상황 여부 검토)
	1-2. 원천 데이터 수집	음성음향 포맷 원천 데이터를 수집 (위급상황 기반 6분류 및 16종)
2. 데이터 정제	2-1. 수집 데이터 폐기	수집 데이터 음질/잡음 기준 사용/폐기 결정 (2인 1조 작업, 필요 시 인원 추가)
	2-2. 수집 데이터 정제	데이터 구축 목적에 부합하도록 수정 (위급상황 기반 분류 16종)
3. 데이터 가공	3-1. 정제 데이터 어노테이션	정제 데이터 기반 특징(음성음향) 추출 (wav 파일 및 스펙트로그램)
	3-2. 정제 데이터 라벨링	특징(음성음향) 데이터 기반 메타데이터 생성 (AI 학습용 데이터셋 후보군 구축)
4. 데이터 검증	4-1. 검수기관 및 전문기관 검증	컨소시엄 내 검수기관 및 외부 전문기관 검증 (검증 지침 기준 구축 데이터 평가 및 검증)
	4-2. 검증 결과 보고 및 반영 조치	검증 결과 기반 구축 데이터 반영 및 개선 (반영 및 개선 데이터 재평가 및 재검증)

2.2 문제정의

2.2.1 임무 정의

- 공공안전을 위해 위급상황시, 발생되는 소리의 크기와 특정 음향/음성데이터를 입력받아, 소리크기와 특정 음향/음성데이터를 기준으로 위급상황인지 여부를 비교 분석한 후, 위급상황에 따른 상황판단 정확도를 높이는 것임
- 범죄 및 긴급사고, 화재·구조·구급 등과 같은 영역에서 음향/음성 분석기반 상황판단 기술이 활용될 경우 긴급 상황에서 빠른 상황판단을 통한 인명구조 및 안전 사회 유지에 효율성을 높일 수 있음
- CCTV 또는 IoT 디바이스, 전화로부터 입력되는 오디오에서 이상상황을 탐지하는 지능형 위급상황 음향/음성 인식 AI의 학습 개발에 활용하기 위한 데이터셋을 구축하고 시범 모델을 개발하여 데이터셋을 검증하는 것이 범위이고, 최종적으로 인공지능 학습을 위한 최적의 위급상황 음향/음성 데이터셋 구축 및 공개를 목표로 함

2.2.2 데이터 구축 유의사항

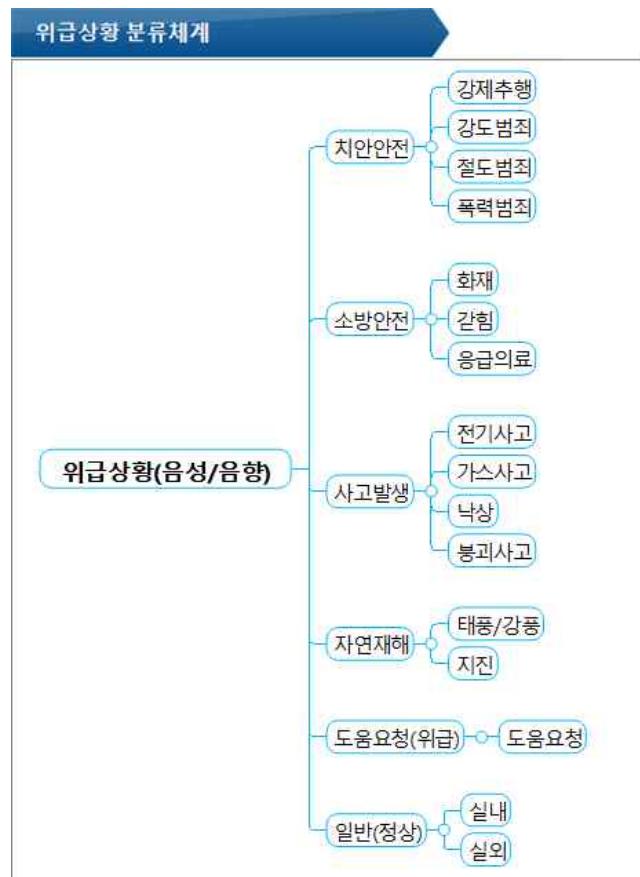
구분	이슈 해결 방안	비고
개인 정보	<ul style="list-style-type: none"> - 통신비밀보호법으로 인해 녹취기기 등 설치를 위해서 녹음 대상자의 허가가 필요하므로 이를 위한 녹음 허가 동의서 작성 및 요청 - 특정인 음성, 성명, 전화번호, 주소 등의 개인정보는 음향/음성 데이터셋 구축에 있어서 배제하는 것을 원칙으로 수행 - 녹음동의서에 대한 충분한 설명 및 서명 날인 확보를 통하여 개인정보에 대한 기본적인 이슈를 해결 	
저작권	<ul style="list-style-type: none"> - 저작권자의 허가가 필요하므로 데이터의 사용 및 공개 가능 여부에 확인을 선행하여 침해 이슈 해결 - 저작권자의 권리를 침해할 수 있는 데이터는 사전에 배제하는 것을 원칙으로 수행 - 출처를 명시하여 사용이 가능한 제한적인 저작권은 필수사항을 이행하여 기본적인 이슈를 해결 	
유해성	<ul style="list-style-type: none"> - 유해성을 확인 및 검증하여 사회적 민감 정보에 대한 이슈 해결 - 성차별, 정치적 성향, 종교 등 사회적 민감 정보를 포함하는 경우 배제하는 것을 원칙으로 수행 	

2.3 획득·정제

2.3.1 원천데이터 선정

2.3.1.1 위급상황 기준 선정

- 위급상황 검출을 위한 음성 및 음향 데이터셋 구축을 위해 먼저 위급상황에 대해 정의하는 것이 필요함
- 위급상황을 다루는 법규 및 법률과 정부부처 시행령, 여러 가지 논문등을 분석하여 정의함
- 『119구조·구급에 관한 법률』을 비롯하여『경찰법』,『소방기본법』,『재난 및 안전관리 기본법』,『소방 기본법』에 따라 화재, 재난·재해 및 테러, 범죄 등의 긴급한 상황을 위급한 상황으로 정의함
- 위급상황의 객관적 및 실제적인 정의 및 분류위해, 행정안전부, 보건복지부, 국토교통부, 환경부, 보건복지부, 해양수산부 등 많은 정부 부처와 관계있으나 그 중 경찰청과 소방청을 외청으로 두고 있는 행정안전부의 위급상황 분류체계 참고함
- 행정안전부 생활안전지도 분류체계 및 통계정보, 국민재난안전 분류체계 정보 반영하고 각종 논문, 행정안전부 생활안전지도, 국민재난안전 포털, 소방청 및 경찰청 통계 데이터 등 기준으로 16종의 위급상황 선정함



2.3.1.2 위급상황 기준

대분류	중분류	상황	설명	비고
치안안전	강제추행	강제추행, 성폭력 등	강제추행 및 성폭력 등에 따른 위급상황 음향 및 음성	
	강도범죄	신체 등을 위협하는 상황	강도범죄에 의해 발생하는 위급상황 음향 및 음성	
	절도범죄	도둑 등	도둑, 절도에 의해 발생하는 위급상황 음향 및 음성	
	폭력범죄	신체, 기물 등으로 타격하는 상황욕설 등 싸우는 상황	싸움, 폭력 등에 의해 발생하는 위급상황 음향 및 음성	
소방안전	화재	화재 등	화재 등에 따른 위급상황 음향 및 음성	
	갇힘	승강기, 실내 감금 등	승강기 갇힘 등에 의한 위급상황 음향 및 음성	
	응급의료	부상, 지병 등에 의한 응급의료 상황	앰뷸런스 호출 등 응급 의료 구급요청 신고 음향 및 음성	
사고발생	전기사고	감전, 누전 및 합선 등	감전사고 등에 발생하는 위급상황 음향 및 음성	
	가스사고	가스누수, 가스에 의한 등 폭발하는 상황 등	가스 누출 등에 발생하는 위급상황 음향 및 음성	
	낙상	넘어짐 상황	낙상사고에 따른 위급상황 음향 및 음성	
	붕괴사고	산사태 및 건물 붕괴 상황	산사태, 건물 붕괴에 의한 위급상황 음향 및 음성	
자연재해	태풍/강풍	태풍 및 강풍상황	태풍에 따른 위급상황 음향 및 음성	
	지진	지진상황	지진 발생에 의한 위급상황 음향 및 음성	
일반 (위급)	도움요청	도움요청 상황(음성)	소방서(119) 등에 도움요청 신고 음향 및 음성	
일반 (정상)	실내	일반적인 실내 상황	위급상황과 비교 대조하여 위급상황 인식 향상 위한 음향 및 음성	
	실외	일반적인 실외 상황	위급상황과 비교 대조하여 위급상황 인식 향상 위한 음향 및 음성	

○ 원천 데이터 획득 출처

- 위급상황 음성/음향 데이터셋 구축을 위해 다음과 같이 인터넷 공개 사이트를 이용한 데이터, 다양한 연령별, 지역별 음성을 수집하기 위한 크라우드 소싱 및 성우 연기자 및 자체 녹음을 통해 원천 데이터를 획득함
- 하기 방안 외 필요시 공공기관(소방재청 119 데이터 등)에 협조를 하여 법률 검토 후 데이터 취득 가능함

획득 방안	출처	데이터 보유량	라벨링 유무	비고
공개 데이터	DCASE	• 성능이 검증된 음향 데이터 파일 및 평가 데이터 제공	유	인터넷 공개 사이트 활용
	AudioSet	• 2,084,320개의 라벨링된 사운드 파일	유	
	UrbanSound	• 1,302개의 라벨링된 사운드 파일	유	
	FSD	• 600개 이상의 사운드 클래스, 297,144개의 사운드 파일	유	
	Freesound	• 400,000개 정도의 라벨링된 사운드 파일, 약 120,000개의 저작권 이슈 해결이 가능한 공개 데이터 제공	유	
	etc.	• 추가적인 사이트 확보 진행 중	유	
크라우드 소싱	불특정 다수	• 참여 작업자 수에 따라 차이가 있음 • 음성의 경우, 작업 가이드에 대표적인 예시를 제시하고, 상황에 따른 각 화자별 자유 발화를 포함하여 녹음	유	
자체 구축 (녹음)	자체 녹음팀	• 설치 대수 및 녹음 시간에 따라 차이가 있음 • 음성의 경우, 작업 가이드에 대표적인 예시를 제시하고, 상황에 따른 각 화자별 자유 발화를 포함하여 녹음	유	

○ 원천 데이터 획득 환경 구축

획득방안	획득 환경 구축 내용	비고
공개 데이터	<ul style="list-style-type: none"> • 해외 사이트(DCASE, AudioSet, Urban Sound, FSD, Freesound)는 주로 음향 데이터 확보가 목적으로, 다운로드 후에 16종 분류체계와 목적에 부합하도록 음향/음성 분석을 통하여 유효성이 확보된 음향/음성데이터를 선정 • 온라인 저작도구를 통한 음성/음향 데이터 등록(http://www.d2ai.co.kr) • 데이터 선정기준 <ul style="list-style-type: none"> - 저작권 불분명한 데이터는 제외 - 음원데이터 샘플링을 통해 수초 이하 및 초저음 음향 데이터 제외 - 잡음이 많고 각 뒤섞인 음향 데이터 제외 	공개 데이터, 크라우드 소싱, 녹음 데이터는 상호 보완을 통하여 위급상황 데이터셋 확보.
크라우드 소싱	<ul style="list-style-type: none"> • 크라우드 소싱 도메인을 구축하여 불특정 다수로부터 음성/음향 데이터를 획득 • 녹음파일 형식 기준 (WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) • 녹음파일 장비: 개인용 휴대폰(기본녹음 앱 이용) 	

	<ul style="list-style-type: none"> 데이터 수집은 시나리오의 상황에 적합한 장소에서 하도록 공지 온라인 저작도구를 통한 음성/음향 데이터 등록(http://www.d2ai.co.kr) 통신비밀보호법을 위한 녹음 허가 동의서 작성 Reward 시스템을 통해 크라우드 워커에게 보상지급 	
자체 구축 (녹음)	<ul style="list-style-type: none"> 녹음 환경 구축 <ul style="list-style-type: none"> - 녹음파일 형식 기준(WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) - 녹음장비: 태블릿 이용, 녹음앱은 기본 앱외에 다른 앱도 사용 가능하나 형식에 벗어나지 않도록 함 - 녹음 목표시간 및 수집량: 일평균 최소 40분 이상 및 150건 이상 - 설정된 시나리오를 활용하여 데이터 수집 - 수집 데이터 다양성을 위해 녹음기와 참여자의 거리 별 녹음, 녹음기 설치 장소(실내 및 실외)별 녹음 시행 - 3자에 대한 녹음 시 녹음 허가 동의서 작성 온라인 저작도구를 통한 음성/음향 데이터 등록(http://www.d2ai.co.kr) 데이터 선정 기준 <ul style="list-style-type: none"> - 녹음팀 사내 전문 검수팀에 의해 등록된 음성/음향 데이터에 대한 전수검사를 통해 데이터 최종 선정 	

2.3.2 획득·정제 절차

2.3.2.1 데이터 획득

○ 데이터 획득 수행조직

- 데이터는 성능이 상이한 스마트 폰 또는 태블릿을 통해 취득한 위급상황 음향/음성 데이터를 상황 정보, 환경정보와 결합하는 것으로 정의함
- 이 중 원천데이터의 획득 및 정제에 관한 책임과 역할은 아래 참조
 - 원천 데이터 수집 : (주)아이엠알
 - 원천 데이터 정제 : (주)아이엠알 / (주)엠포러스

2.3.2.2 데이터 획득 분류 및 녹음 시나리오

대분류	중분류	상황	시나리오 예	비고
치안안전	강제추행	강제추행, 성폭력 등	<ul style="list-style-type: none"> • 음성 <ul style="list-style-type: none"> - 누, 누구세요? 누구야, 당신! 뭐 하는 거야? 저리 가! 당신 미쳤어? 어딜 만져? 만지지 마! 가만 안 들 거야. 경찰 부를 거야. - 살려주세요. 제발, 가게 해 주세요! - 왜 이래요? 이러지 마요. 보내주세요. 대체 나한테 왜 이래요? 하지 마요. 싫어! 저리 가! 하지 마! - 도와주세요! 거기 누구 없어요? 저 좀 살려주세요 • 음향 <ul style="list-style-type: none"> - 뒷걸음질할 때 신발 끄는 소리, 상대방 밀치며 팔 휘젓는 소리, 두려울 때 내는 신음, 입 막았을 때 손바닥 사이로 흘러나오는 비명, 밀쳤을 때 넘어지는 소리, 저항할 때 몸부림치는 소리 	
	강도범죄	신체 등을 위협하는 상황	<ul style="list-style-type: none"> • 음성 <ul style="list-style-type: none"> - 이, 이러지 마세요! 그거 치워요. 위험해요. - 나 죽일 거예요? 그럼 당신 살인자 되는 거예요. - 잘못했어요. 그러니까 제발 사, 살려 주세요. - 엄마(아빠)! 무서워요. 	

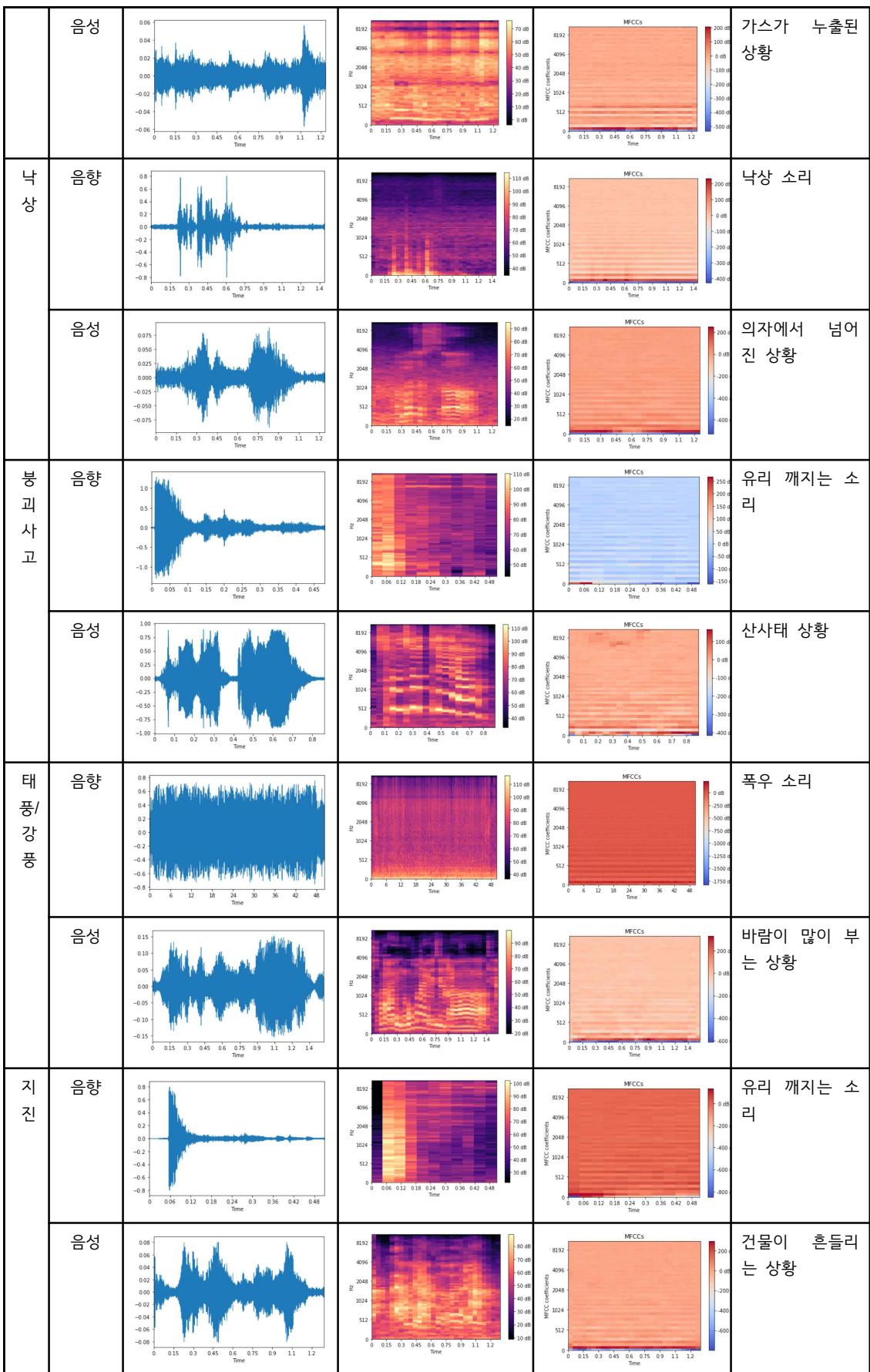
			<ul style="list-style-type: none"> • 음향 <ul style="list-style-type: none"> - 맞는 소리, 비명 소리, 물건이 깨지는 소리 등 - 아이들 울음소리, 빌 때 손바닥 비비는 소리, 목 졸릴 때 버둥거리는 소리 	
	절도범죄	도둑 등	<ul style="list-style-type: none"> • 음성 <ul style="list-style-type: none"> - 집에 도둑 들었어요, 도둑 들었나봐요, 도둑 좀 확인해 주세요, 집이 다 털렸어요, 도어락 뜯겨있어요, 도둑이야, 내 통장 사라졌어 등 - 거기 누구야? 누군데 남의 집에... - 뭐야? 도둑이야? 무서워. 어떡해! - 무슨 소리야? 저거 문 따는 소리 아냐? - 가만, 쉿! 안에서 이상한 소리가 나. 도둑 듣 것 같아. • 음향 <ul style="list-style-type: none"> - 물건 뒤지는 소리, 잠금장치 절단하는 소리 등 	
	폭력범죄	신체, 기물 등으로 타격하는 상황 욕설 등 싸우는 상황	<ul style="list-style-type: none"> • 음성 <ul style="list-style-type: none"> - 때리지 마세요.., 잘못했어요, (울먹이는 소리) 저한테.., 왜... 이러세요, 아, 아파요. 때리지 마요, 저 잘못한 거 없어요. 제발 때리지 마세요.. • 음향 <ul style="list-style-type: none"> - 구타 소리, 소리, 폭력에 따른 신음소리 등 	
	화재	화재 등	<ul style="list-style-type: none"> • 음성 <ul style="list-style-type: none"> - 불났어! 불, 방안에 연기가 가득해요, 연기를 너무 많이 마셔서, 산불이 크게 났심더, , 불이 났어요, 여기 불 났어요 불, 콜록거리고 다들 울어요, 앞이 아무것도 안 보여요, 지금 연기가 엄청 까맣고.. 등 - 부, 불이야! - 어떤해? 불났어! 물! 물! 빨리 물 가져와. - 119 전화해. 소방차 불러. - 연기가 너무 심해. 숨 막혀. 숨을 못 쉬겠어. - 어디로 나가지? 연기 때문에 아무것도 안 보여. 무서워. - 사방이 불이야. 뜨거워. 못 나가겠어. • 음향 <ul style="list-style-type: none"> - 불 탄는 소리, 화재경보 소리 등 - 연기에 콜록거리는 소리 	
소방안전	감침	승강기, 실내 감금 등	<ul style="list-style-type: none"> • 음성 <ul style="list-style-type: none"> - 안에 갇힌 것 같아요, 아까 소리 들리냐고 물었더니, 엘리베이터 안에 사람이 갇히셔서, 갇힌 지 좀 오래된 것 같아요, 아무리 두드려도 사람이 없어요, 제발 문 좀 열어주세요, 나는 문 못 열겠어, 문이 안 열려요 등 - 엘리베이터가 멈췄어요. 안 움직여요. - 엘리베이터에 갇혔어요. - 문이 잠겼어요. 안 열려요. 못 열겠어요. - 안에 갇혔어요. 밖에 아무도 없는 것 같아요. - 저 좀 꺼내주세요. • 음향 <ul style="list-style-type: none"> - 거칠게 문 두드리는 소리 등 	
	응급의료	부상, 지병 등에 의한 응급의료 상황	<ul style="list-style-type: none"> • 음성 <ul style="list-style-type: none"> - 죽을 것 같아요, 나 어떡해, 아뇨.. 그냥 쓰러져 있어요 - 음식물이 목에 걸렸어요. - 숨을 못 쉬겠어요. - 갑자기 쓰러졌어요. - 숨을 안 쉬어요. • 음향 <ul style="list-style-type: none"> - 구토하는 소리, 기침하는 소리, 신음 소리, 호흡곤란 소리 등 	
	전기사고	감전, 누전 및 합선	<ul style="list-style-type: none"> • 음성 	

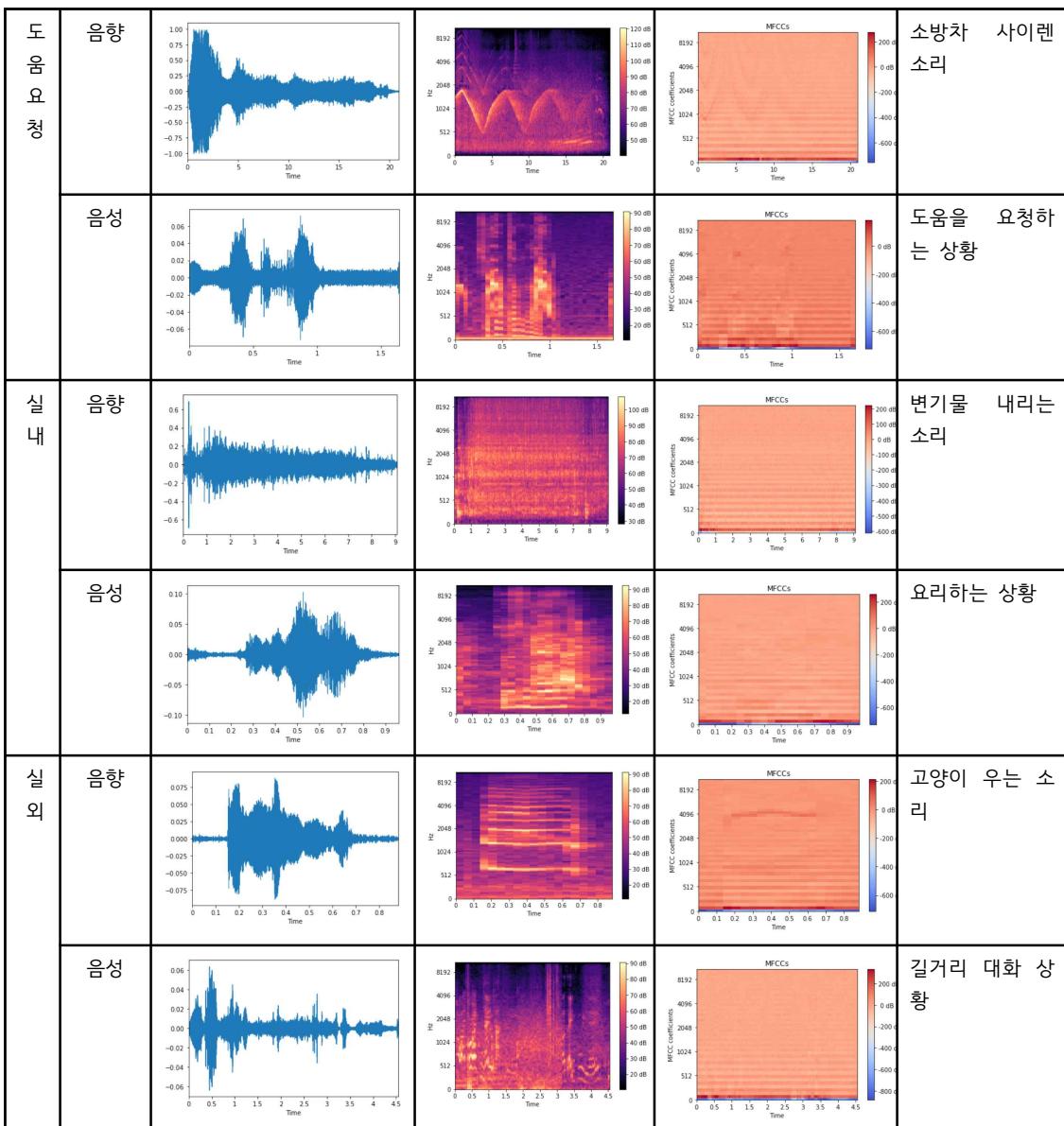
		등	<ul style="list-style-type: none"> - 누전됐나 봐요. 불이 완전히 다 나갔어요. - 두꺼비집 만질 줄 몰라요. - 아무것도 안 보여요. 뭔가 타는 냄새가 나요. • 음향 - 스파크 소리, 전기 터지는 소리, 기계작동이 갑자기 멈추는 소리 등 	
	가스사고	가스누수, 가스에 의한 등 폭발하는 상황 등	<ul style="list-style-type: none"> • 음성 - 방금 가스 폭발같아, 가스 새는거 같아요 - 이상한 냄새가 나요. 가스 냄새 같아요. 속이 메슥거려요. • 음향 - 가스 새는 소리, 폭발음 소리, 호흡 곤란 소리 등 	
사고발생	낙상	넘어짐 상황	<ul style="list-style-type: none"> • 음성 - 넘어지면서 발목을 뻣 것 같아요, 다리가 부러진것 같아요 - 미끄러졌어요. 다리(어깨, 팔)가 아파요. 부었어요. - 욕실에서 넘어졌어요. 머리에서 피가 나요. - 계단 내려오다 헛발 디뎠어요. . • 음향 - 바닥에 넘어지는 소리, 굴러 떨어지는 소리 등 	
	붕괴사고	산사태 및 건물 붕괴 상황	<ul style="list-style-type: none"> • 음성 - 산이 무너져 내렸어요. 흙더미에 갇혔어요. - 산사태가 났어요. 흙이 집을 덮쳤어요. • 음향 - 돌이나 흙이 쏟아지는 소리, 철근이 휘는 소리, 돌이나 흙이 바스라지는 소리 등 	
자연재해	태풍/강풍	태풍 및 강풍상황	<ul style="list-style-type: none"> • 음성 - 바람이 너무 세요. 가만히 서 있을 수가 없어요. 금방이라도 날아갈 것 같아요. - 바람에 다 날아가요. 다 잠겨버렸어요. • 음향 - 강한 바람이 부는 소리, 폭우가 쏟아지는 소리, 천둥이 치는 소리, 창문 깨지는 소리, 창문 흔들리는 소리 등 	
	지진	지진상황	<ul style="list-style-type: none"> • 음성 - 어떡해 바닥이 흔들려요, 방금 지진인거 같아요, 지진! 지진이야, 지진이야, 땅이 흔들려요, • 음향 - 물건이 흔들리는 소리, 물건이 떨어지는 소리, 땅이 울리는 소리 등 	
일반 (위급)	도움요청	도움요청 상황(음성)	<ul style="list-style-type: none"> • 음성 - 도와주세요, 살려주세요, 신고해 주세요, 너무 불안해요 등 	
일반 (정상)	실내	일반적인 실내 상황	<ul style="list-style-type: none"> • 음성 - 일상 대화 등 • 음향 - 샤워 소리, 물끓는 소리, 청소기 소리, 세탁기소리, 변기 물 내리는 소리, 설거지 소리, 키보드 소리, 도마 소리 등 	
	실외	일반적인 실외 상황	<ul style="list-style-type: none"> • 음성 - 버스, 기차, 버스터미널 안내방송 상황 등 • 음향 - 오토바이소리, 개 짖는 소리, 아이들이 뛰어노는 소리, 자동차가 지나가는 소리, 기차가 지나가는 소리 등 	

2.3.2.3 데이터 획득 분류 스펙트럼 예시

분류	음성/음향	주파수	Mel-spectrogram	MFCC	설명
강제 추행	음향				호흡 곤란 소리
	음성				성적 추행을 당한 상황
강도 범죄	음향				창문 두드리는 소리
	음성				강도에게 흥기로 협박(위협) 당하는 상황
절도 범죄	음향				도둑이 침입한 상황
	음성				도둑이 침입한 상황
폭력 범죄	음향				매 휘두르는 소리
	음성				부부 싸움하는 상황

화재	음향				기침 소리 (성인 남자)
	음성				대피하는 상황
간 험	음향				문 두드리는 소리
	음성				집안에 간한 상황
위급의료	음향				호흡 곤란
	음성				비명 소리(유아 여성)
전기사고	음향				스파크 소리
	음성				누전된 상황
가스사고	음향				폭발음 소리





2.3.2.4 데이터 획득 방법

○ 데이터 획득 시 기술적 공통상황

1) 위급상황에서의 음성 및 음향 인식 위한 구체적인 기술적 요소

구분	대상	기술적 사항
음성	디바이스에서 음성의 크기 (마이크 도달 기준)	70 dB SPL
	키워드의 길이	4음절 이상, 600 ms 이상 (권고: 5음절 이상, 1초)
	주변 소음 대비 음성 SNR	12 dB
	음성 인식 방법	화자 독립
	특정 음원	음성긴급전화 모사 음성/음향, 위급상황 음성 상황 연출 소리
음향	디바이스에서 음향의 크기 (마이크 도달 기준)	70 dB SPL
	음향의 길이	2000 ms 이상 (권고: 2초)
	주변 소음 대비 음성 SNR	12 dB
	음향 인식 방법	음향 독립
	특정 음원	음성긴급전화 모사 음향, 위급상황 음향 연출 소리

- 디바이스에서 음성 및 음향의 크기 (마이크 도달 기준) : 70dB SPL
 - 디바이스에서 응답해야하는 음성/음향의 크기는 일상 생활 속에서 발생하는 소음에 대하여 규정한 생활소음규제를 참고한 환경적 관점 등을 반영함

※ dB SPL(sound pressure level, 음압) : 음압(소리의 매질인 공기의 압력의 실효출력)의 단위는 파스칼(Pa)이지만, 2×10^{-5} Pa를 기준값(0dB SPL)으로 함

생활소음규제

생활소음이란 일상 생활 중에 자주 접하게 되는 주변의 소음이라고 생각할 수 있다.

『소음진동·관리법』에 따라 생활소음 규제대상은 학성기에 의한 소음, 공장 및 사업장의 작업소음, 기타 심야의 계속적 또는 반복적 소음으로 구성된다.

대상지역	소음원	시간대별		
		야침, 저녁 (05:00~07:00, 18:00~22:00)	주간 (07:00~18:00)	야간 (22:00~05:00)
가. 주거지역, 농지지역, 관리지역 中 취락지구·주거개발진흥지구 및 관광휴양개발진흥지구, 자연환경 보전지역, 그 밖의 지역에 있는 학교·종합병원·공공도서관	학성기	옥외설치	60이하	65이하
		옥내에서 옥외로 소음이 나오는 경우	50이하	55이하
	사업장	공장	50이하	55이하
		동일 건물	45이하	50이하
		기타	50이하	55이하
	공사장	60이하	65이하	50이하
나. 그 밖의 지역	학성기	옥외설치	65이하	70이하
		옥내에서 옥외로 소음이 나오는 경우	60이하	65이하
	사업장	공장	60이하	65이하
		동일 건물	50이하	55이하
		기타	60이하	65이하
	공사장	65이하	70이하	55이하

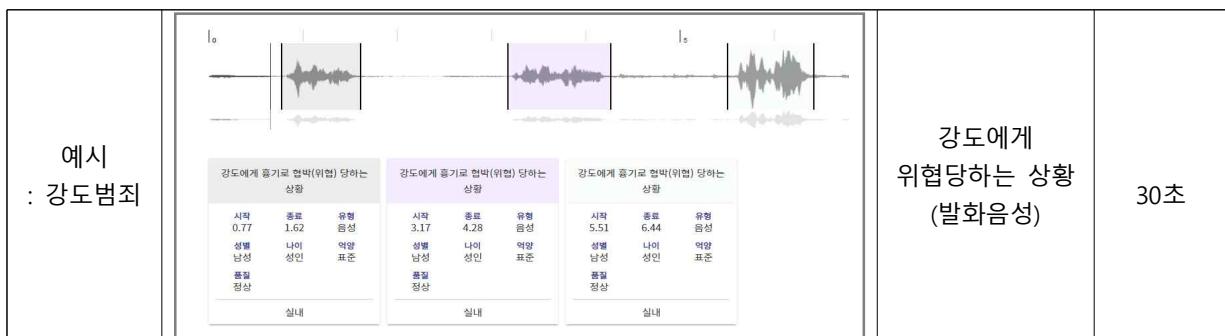
2) 수집 데이터 분류

구분	획득 방법
공개 데이터	<ul style="list-style-type: none"> 해외 사이트(DCASE, AudioSet, Urban Sound, FSD, Freesound) 검증을 통해 다량의 데이터를 보유하고 있으면서 데이터 저작권 사용이 가능한 사이트 선정(예: FreeSound) 저작권 분석과 청각에 의한 음향 샘플링을 통해 음향 데이터 선별 및 수집 수집 시 분류별 카테고리 딕렉토리를 구축하여 해당 딕렉토리에 구축 <ul style="list-style-type: none"> - 예: 대분류→중분류→음성/음향/분류중→ 소분류→저작권→출처 사이트
자체 녹음	<ul style="list-style-type: none"> 녹음 파일 형식 준수(WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) 녹음 목표시간 및 수집량: 일평균 최소 40분 이상 및 150건 이상 설정된 시나리오를 활용하여 성우 및 녹음팀 자체에 의한 데이터 수집 <ul style="list-style-type: none"> - 수집 데이터 다양성을 위해 녹음기와 참여자의 거리 별 녹음, 녹음기 설치 장소(실내 및 실외)별 녹음 시행 - 2인1조 또는 3인 1조에 의한 아이디어 공유 및 롤플레잉 녹음 - 현장 및 음향실에서 시나리오 기반 상황 재연 및 음성/음향 녹음 - 음성의 다양성을 위해 음향실 스튜디오 내에서 총 46명의 성우를 활용하여 하루 약 최대 2시간 이내로 연기하여 1,400시간 녹음 음향실 스튜디오 환경 <ul style="list-style-type: none"> - 음향실 1번: 2.1m x 3.0m, 녹음을 위한 4개의 태블릿 PC 이용 - 음향실 2번: 2.0m x 3.0m, 녹음을 위한 5개의 태블릿 PC 이용

- 녹음업체를 섭외하여 시나리오를 기반으로 실제 녹음 내용을 참조하여 위급상황을 재현하여 직접 제작
- 최대한 현실과 근접한 음성 및 음향 제작을 위해 실제 위급상황 녹취록 또는 위급상황 공개된 영상을 참고하여 시나리오를 제작하여 촬영

○ 녹음환경 예시

구분	획득 환경	수집데이터	속성
위급상황 음향/음성 녹음환경		위급상황 음향/음성	연령별 발화자



○ 녹음 획득 도구

구분	획득도구	수집데이터	수량
고정형 위급상황 음향/음성 수집장치	  <p>Samsung Galaxy Tab 6 Lite - 샘플링 : 44.1khz - 해상도 : 16bit / Mono - 감도 : -45db - 최대 입력음압 : 140dB SPL - 포맷 : PCM WAV 포맷</p>	위급상황 음향/음성	60식 (1셋트 구성 : 마이크 + 고정틀, 60식 동일 구성)

3) 크라우드 소싱

구 분	획득 방법
크라우드 소싱	<ul style="list-style-type: none"> 녹음파일 형식 준수(WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) 성별 및 다양한 연령별로 원천데이터 내 빈도 균등하게 배분하여 모집 온라인 저작도구를 통해 데이터 수집 및 녹음방법, 어노테이션/라벨링 가능할 수 있게 16종 분류체계와 목적에 맞게 음성/음향 등록 (http://www.d2ai.co.kr)   <ul style="list-style-type: none"> 개인정보보호를 위한 녹음 허가 동의서 작성 Reward 시스템을 통해 크라우드 워커에게 등록 데이터에 대한 보상지급 정제 시 사내 전문 검수 팀에 의해 등록된 음성/음향 데이터에 대한 전수검사를 통해 데이터 최종 선정

- 음성 데이터 수집 크라우드 워커는 웹사이트를 통해 다양한 지역 및 연령대를 고려하여 모집
- 음향의 경우 위급상황을 재현하여 녹화 가능한 크라우드 워커 모집
- 음향 등의 배역을 맡아 이상행동을 재현
- 현재 데이터 분류를 토대로 정의한 16가지의 위급상황의 종류 (폭행, 싸움, 절도, 기물파손, 실신, 배회, 침입, 투기, 데이트폭행 및 추행, 주취행동, 납치, 강도)로 총 1000여개의 시나리오로 녹화 종류 업데이트 중이므로 추후에 이상행동이 추가되거나 제외될 수 있음)

- 녹음 대상자에게 반드시 개인정보 활용 동의서 작성 등의 법적 절차를 거쳐 녹음 및 데이터를 공개 하며, 법적 의무사항을 준수할 수 있도록 절차와 동의서 등에 대한 법률 자문을 구함

○ 녹음환경 예시

구분	획득 환경	수집데이터	속성																																																																																																												
위급상황 음향/음성 녹음환경			위급상황 음향/음성 연령별 발화자																																																																																																												
예시 : 폭력범죄	<table border="1"> <thead> <tr> <th colspan="4">때리지마</th> </tr> <tr> <th>시작</th> <th>종료</th> <th>유형</th> <th>음성</th> </tr> </thead> <tbody> <tr> <td>0.59</td> <td>1.39</td> <td>유형</td> <td>음성</td> </tr> <tr> <td>성별</td> <td>나이</td> <td>역량</td> <td>표준</td> </tr> <tr> <td>남성</td> <td>성인</td> <td></td> <td></td> </tr> <tr> <td>환경</td> <td>장소</td> <td></td> <td></td> </tr> <tr> <td>침실</td> <td>침실</td> <td></td> <td></td> </tr> <tr> <td>정상</td> <td>정상</td> <td></td> <td></td> </tr> <tr> <td colspan="4">실내</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="4">때리지마</th> </tr> <tr> <th>시작</th> <th>종료</th> <th>유형</th> <th>음성</th> </tr> </thead> <tbody> <tr> <td>2.74</td> <td>3.4</td> <td>유형</td> <td>음성</td> </tr> <tr> <td>성별</td> <td>나이</td> <td>역량</td> <td>표준</td> </tr> <tr> <td>남성</td> <td>성인</td> <td></td> <td></td> </tr> <tr> <td>환경</td> <td>장소</td> <td></td> <td></td> </tr> <tr> <td>침실</td> <td>침실</td> <td></td> <td></td> </tr> <tr> <td>정상</td> <td>정상</td> <td></td> <td></td> </tr> <tr> <td colspan="4">실내</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="4">때리지마</th> </tr> <tr> <th>시작</th> <th>종료</th> <th>유형</th> <th>음성</th> </tr> </thead> <tbody> <tr> <td>5.15</td> <td>5.9</td> <td>유형</td> <td>음성</td> </tr> <tr> <td>성별</td> <td>나이</td> <td>역량</td> <td>표준</td> </tr> <tr> <td>남성</td> <td>성인</td> <td></td> <td></td> </tr> <tr> <td>환경</td> <td>장소</td> <td></td> <td></td> </tr> <tr> <td>침실</td> <td>침실</td> <td></td> <td></td> </tr> <tr> <td>정상</td> <td>정상</td> <td></td> <td></td> </tr> <tr> <td colspan="4">실내</td> </tr> </tbody> </table>	때리지마				시작	종료	유형	음성	0.59	1.39	유형	음성	성별	나이	역량	표준	남성	성인			환경	장소			침실	침실			정상	정상			실내				때리지마				시작	종료	유형	음성	2.74	3.4	유형	음성	성별	나이	역량	표준	남성	성인			환경	장소			침실	침실			정상	정상			실내				때리지마				시작	종료	유형	음성	5.15	5.9	유형	음성	성별	나이	역량	표준	남성	성인			환경	장소			침실	침실			정상	정상			실내				폭력을 당하는 상황 (발화음성 또는 음향) 35초	
때리지마																																																																																																															
시작	종료	유형	음성																																																																																																												
0.59	1.39	유형	음성																																																																																																												
성별	나이	역량	표준																																																																																																												
남성	성인																																																																																																														
환경	장소																																																																																																														
침실	침실																																																																																																														
정상	정상																																																																																																														
실내																																																																																																															
때리지마																																																																																																															
시작	종료	유형	음성																																																																																																												
2.74	3.4	유형	음성																																																																																																												
성별	나이	역량	표준																																																																																																												
남성	성인																																																																																																														
환경	장소																																																																																																														
침실	침실																																																																																																														
정상	정상																																																																																																														
실내																																																																																																															
때리지마																																																																																																															
시작	종료	유형	음성																																																																																																												
5.15	5.9	유형	음성																																																																																																												
성별	나이	역량	표준																																																																																																												
남성	성인																																																																																																														
환경	장소																																																																																																														
침실	침실																																																																																																														
정상	정상																																																																																																														
실내																																																																																																															

○ 녹음 획득 도구

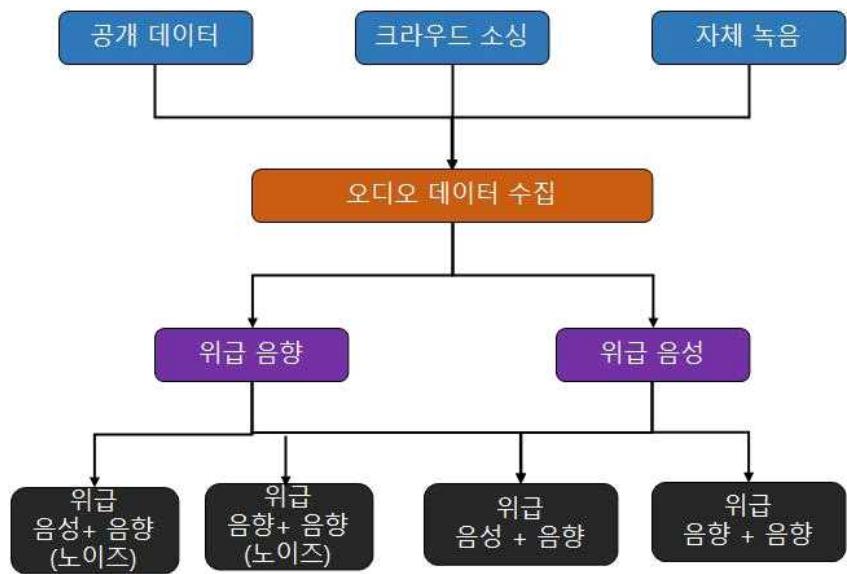
구분	획득도구	수집데이터	녹음환경구성
고정형 위급상황 음향/음성 수집장치		<ul style="list-style-type: none"> - 오디오 녹음 설정 기준을 준수하는 다양한 스마트폰 [오디오 녹음 설정 기준] - 샘플링 : 44.1khz - 해상도 : 16bit / Mono - 감도 : -45db - 최대 입력음압 : 140dB SPL - 포맷 : PCM WAV 포맷 	위급상황 음향/음성 샘플링, 해상도, 포맷 등 동일 환경 구성

4) 오디오 믹싱

○ 획득 데이터 믹싱

- 수집된 음성 및 음향 데이터를 이용한 오디오 믹싱을 이용하여 위급상황 시나리오에 맞는 적정한 음성/음향 데이터 구축
- Deep sound의 인식을 위해 서로 다른 class에 속한 두 가지 소리를 랜덤하게 혼합(mixing)하여

Between-class 소리를 생성한 후 Deep learning을 적용한 학습 모델에 적용 및 훈련하여 mixing 비율을 출력. 결과적으로 다양한 소리 인식 성능이 개선될수 있도록 믹싱 데이터 수집



○ 데이터 획득 시 직접 취득 불가 시 녹음 방안

구분	분류	상황	녹음 방안
	직접 녹음	일반(위급), 일반(정상)	<ul style="list-style-type: none"> ▶ 일반(위급) <ul style="list-style-type: none"> - 도움 요청 <ul style="list-style-type: none"> • 크라우드 소싱 활용하여 성우 등 크라우드 워커 고용 • 크라우드 워커가 가정 내에서 구조 요청하는 상황을 규정하여 직접 녹음 가능
음성	상황 연출 또는 유사데이터 생성	치안안전, 소방안전, 사고발생, 자연재해	<ul style="list-style-type: none"> ▶ 치안안전 <ul style="list-style-type: none"> - 강제추행, 강도범죄, 절도범죄, 폭력범죄 <ul style="list-style-type: none"> • 현재 경찰청 콜센터의 위급상황 녹취록이 비공개되어 있어 직접 녹음이 불가능 • 상황별 시나리오 마련하고 해당 상황을 연출하여 데이터 수집 불가피 ▶ 소방안전 <ul style="list-style-type: none"> - 화재, 전기사고, 가스사고 <ul style="list-style-type: none"> • 자주 발생하는 상황이 아닐뿐더러 사고발생 시 직접 녹음하기엔 위험 <ul style="list-style-type: none"> • 현재 소방청과의 협조를 통해 데이터를 확보 노력 중에 있으나 개인정보보호법 등 규제로 인해 직접 녹음 어려움 • 크라우드 소싱 활용하여 성우 등 크라우드 워커 고용 • 상황별 시나리오 마련하고 해당 상황을 연출하여 데이터 수집 불가피 - 응급의료 <ul style="list-style-type: none"> • 크라우드 워커가 직접 의료기관에 방문하거나 가족, 친구, 친지 등을 통해 직접 녹음하는 것도 가능하지만, 직접 녹음하기에는 상황 발생 빈도가 낮아 상황을 포착한 직접 녹음이 어려움 <ul style="list-style-type: none"> • 상황별 시나리오 마련하고 상황을 연출하여 데이터 수집 불가피 ▶ 사고발생

구분	분류	상황	녹음 방안	
			<ul style="list-style-type: none"> - 낙상, 붕괴사고 <ul style="list-style-type: none"> • 자주 발생하는 상황이 아닐뿐더러 사고발생 시 직접 녹음하기엔 위험 • 현재 소방청과의 협조를 통해 데이터를 확보 노력 중에 있으나 개인정보보호법 등 규제로 인해 직접 녹음 어려 • 상황별 시나리오 마련하고 해당 상황을 연출하여 데이터 수집 불가피 ▶ 자연재해 <ul style="list-style-type: none"> - 태풍/강풍, 지진 <ul style="list-style-type: none"> • 자주 발생하는 상황이 아닐뿐더러 사고발생 시 직접 녹음하기엔 위험 • 현재 소방청과의 협조를 통해 데이터를 확보 노력 중에 있으나 개인정보보호법 등 규제로 인해 직접 녹음 어려 • 상황별 시나리오 마련하고 해당 상황을 연출하여 데이터 수집 불가피 	
	직접 녹음	일반(위급), 일반(정상)		
음향	상황 연출 또는 유사데이터 생성	치안안전, 소방안전, 사고발생, 자연재해	<ul style="list-style-type: none"> ▶ 일반(정상) 중 실내/실외 <ul style="list-style-type: none"> - 누구나 겪는 일상적인 상황의 경우 굳이 상황을 연출하지 않아도 크라우드 워커가 가정이나 실외에서 직접 녹음 가능 ▶ 일반(정상) 중 산업현장 <ul style="list-style-type: none"> - 크라우드 워커가 직접 산업 현장을 방문하여 직접 녹음 가능 	
			<ul style="list-style-type: none"> ▶ 치안안전 <ul style="list-style-type: none"> - 강제추행, 강도범죄, 절도범죄, 폭력범죄 <ul style="list-style-type: none"> • 현재 경찰청 콜센터의 위급상황 녹취록이 비공개되어 있어 직접 녹음이 불가능 • 크라우드 소싱 활용하여 성우 등 고용 • 상황별 시나리오 마련하고 해당 상황을 연출하여 데이터 수집 불가피 ▶ 소방안전 <ul style="list-style-type: none"> - 화재, 전기사고, 가스사고 <ul style="list-style-type: none"> • 자주 발생하는 상황이 아닐뿐더러 사고발생 시 직접 녹음하기엔 위험 • 현재 소방청과의 협조를 통해 데이터를 확보 노력 중에 있으나 개인정보보호법 등 규제로 인해 직접 녹음 어려 • 상황별 시나리오 마련하고 해당 상황을 연출하여 데이터 수집 불가피 ▶ 응급의료 <ul style="list-style-type: none"> • 크라우드 워커가 직접 의료기관에 방문하거나 가족, 친구, 친지 등을 통해 직접 녹음하는 것도 가능하지만, 직접 녹음하기에는 상황 발생 빈도가 낮아 상황을 포착한 직접 녹음이 어려움 • 상황별 시나리오 마련하고 상황을 연출하여 데이터 수집 불가피 ▶ 사고발생 	

구분	분류	상황	녹음 방안
			<ul style="list-style-type: none"> - 낙상, 붕괴사고 <ul style="list-style-type: none"> • 자주 발생하는 상황이 아닐뿐더러 사고발생 시 직접 녹음하기엔 위험 • 현재 소방청과의 협조를 통해 데이터를 확보 노력 중에 있으나 개인정보보호법 등 규제로 인해 직접 녹음 어려 • 상황별 시나리오 마련하고 해당 상황을 연출하여 데이터 수집 불가피 ▶ 자연재해 <ul style="list-style-type: none"> - 태풍/강풍, 지진 <ul style="list-style-type: none"> • 자주 발생하는 상황이 아닐뿐더러 사고발생 시 직접 녹음하기엔 위험 • 현재 소방청과의 협조를 통해 데이터를 확보 노력 중에 있으나 개인정보보호법 등 규제로 인해 직접 녹음 어려 • 상황별 시나리오 마련하고 해당 상황을 연출하여 데이터 수집 불가피
기타	무음, 음성+음향	소방안전	<ul style="list-style-type: none"> ▶ 화재, 전기사고, 가스사고 <ul style="list-style-type: none"> • 음성 음향 결합하여 위험 상황 인식

2.3.2.5 공공기관 데이터 활용 방법

- 소방청 등 위급상황에 해당되는 녹화 음향 및 음성을 참고
- 데이터셋 공개 시 개인정보보호 이슈가 발생할 수 있으므로, 수집된 녹음 데이터는 기관 내부적으로 시나리오 작업 및 검증용 데이터로만 활용
- 개인정보보호법 등 관련 법규를 검토하고 결과 제출

※ 데이터 수집의 경우 실내외 녹음이 정상적인 데이터 구축 과정이며, 소방청 등의 비식별화된 녹음 데이터만 활용함 (녹음 데이터는 활용은 아이엠알과 세종대의 업무 협약으로 이뤄진 사항임)

※ 검증용 데이터를 기반으로 시나리오를 만들었기 때문에 학습용 데이터와 검증용 데이터 사이의 일관성에 문제없음

2.3.3 정제 절차

- 공개 데이터, 크라우드 소싱 및 자체 녹음 및 믹싱 등을 통해서 획득한 데이터는 품질관리를 위해 공통적으로 아래와 같은 절차를 거쳐 정제함
- 불필요하거나 무의미한 데이터 제거하는 과정을 통해 음향 및 음성 데이터 정제함
- 음향 및 음성 데이터가 이상행동 시나리오에 맞게 녹음 되었는지 확인함

○ 음향 및 음성 데이터의 품질 검수 및 정제

정제 절차	정제 방법
1. 오디오 청취	숙련 검수자가 음향 청취
2. 음질/잡음 확인	오디오 음질의 명확성 및 위배되는 잡음의 존재 여부를 확인
3. 획득 규정 준수	오디오가 규정된 획득 조건과 방식의 준수 여부를 확인
4. 획득 품질 확인	수집 오디오의 획득 정보 오류 확인 (수집/녹음, 유의미한 객체 존재 유무 등)
5. 다수결 원칙	2인 1조로 상기 과정을 통하여 검수하고 검수 결과 2명의 의견이 상이할 때는 제 3의 검수자를 투입하여 다수의 의견을 취합하여 오디오 데이터의 사용/폐기를 결정
정제 결과	

- 개인을 특정할 수 있는 정보가 있는 경우 폐기
- 음성을 듣고 레이블러 또는 전문가가 판단할 수 없는 음성은 폐기
- 데이터의 잡음 유무(음성 대비 잡음이 클 경우 제외)
- 적정한 소리 크리(식별이 어려운 작은 소리 제외)
- 발음의 정확성(발음이 너무 부정확한 경우 제외)

2.3.4 획득/정제 기준

○ 획득 기준

항목	내용
공개 데이터	<ul style="list-style-type: none"> • 해외 공개 사이트(DCASE, AudioSet, Urban Sound, FSD, Freesound)를 이용하여 주로 음향 데이터 획득 • 데이터 선정기준 <ul style="list-style-type: none"> - 파일 형식 기준 (WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) - 16종 분류기준 및 설정 시나리오에 적합하지 않은 데이터는 제외

항목	내용																								
	<ul style="list-style-type: none"> - 저작권 불분명한 데이터는 제외 - 음원데이터 샘플링을 통해 수초 이하 및 초저음 음향 데이터 제외 - 잡음이 많고 각 뒤섞인 음향 데이터 제외 <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>설명</th><th>약칭</th><th>저작자 표시</th><th>2차 저작물</th><th>영리 목적</th><th>자유저작물 여부</th></tr> </thead> <tbody> <tr> <td>퍼블릭 도메인 (Creative Commons 0)</td><td>CC 0</td><td>필요 없음</td><td>예</td><td>예</td><td>예</td></tr> <tr> <td>저작자 표시 (Attribution)</td><td>CC BY</td><td>예</td><td>예</td><td>예</td><td>예</td></tr> <tr> <td>저작자 표시 + 비영리 (Attribution Noncommercial)</td><td>CC BY-NC</td><td>예</td><td>예</td><td>아니오</td><td>아니오</td></tr> </tbody> </table> <ul style="list-style-type: none"> • 온라인 저작도구를 통한 음성/음향 데이터 등록(http://www.d2ai.co.kr) • 데이터 등록 시 출처 등의 환경정보 입력 	설명	약칭	저작자 표시	2차 저작물	영리 목적	자유저작물 여부	퍼블릭 도메인 (Creative Commons 0)	CC 0	필요 없음	예	예	예	저작자 표시 (Attribution)	CC BY	예	예	예	예	저작자 표시 + 비영리 (Attribution Noncommercial)	CC BY-NC	예	예	아니오	아니오
설명	약칭	저작자 표시	2차 저작물	영리 목적	자유저작물 여부																				
퍼블릭 도메인 (Creative Commons 0)	CC 0	필요 없음	예	예	예																				
저작자 표시 (Attribution)	CC BY	예	예	예	예																				
저작자 표시 + 비영리 (Attribution Noncommercial)	CC BY-NC	예	예	아니오	아니오																				
크라우드 소싱	<ul style="list-style-type: none"> • 녹음파일 형식 준수 (WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) • 녹음장비: 개인 휴대폰 이용(기본녹음 앱 이용) • 데이터 수집은 시나리오의 상황에 적합한 장소에서 하도록 공지 • 데이터 수집 시 16종 분류기준 및 설정 시나리오에 적합하지 않은 데이터는 제외 • 온라인 저작도구를 통한 음성/음향 데이터 등록(http://www.d2ai.co.kr) • 데이터 등록 시 성별, 나이, 억양 등의 환경정보 입력 • 개인정보보호를 위한 녹음 허가 동의서 작성 																								
자체 구축 (녹음)	<ul style="list-style-type: none"> • 녹음파일 형식 기준(WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) • 녹음장비: 태블릿 PC 이용, 녹음 앱은 기본 앱 외에 다른 앱도 사용 가능하나 WAV 파일 형식에 벗어나지 않도록 함 • 녹음 목표시간 및 수집량: 일평균 최소 40분 이상 및 150건 이상 • 수집 데이터 다양성(노이즈 간섭, 장막 등)을 살리기 위해 녹음기와 참여자의 거리 별 녹음, 녹음기 설치 장소(실내 및 실외)별 녹음 시행 • 3자에 대한 녹음 시 녹음 허가 동의서 작성 • 16종 분류기준 및 설정 시나리오에 적합하지 않은 데이터는 제외 • 온라인 저작도구를 통한 음성/음향 데이터 등록(http://www.d2ai.co.kr) • 데이터 등록 시 성별, 나이, 억양 등의 환경정보 입력 																								

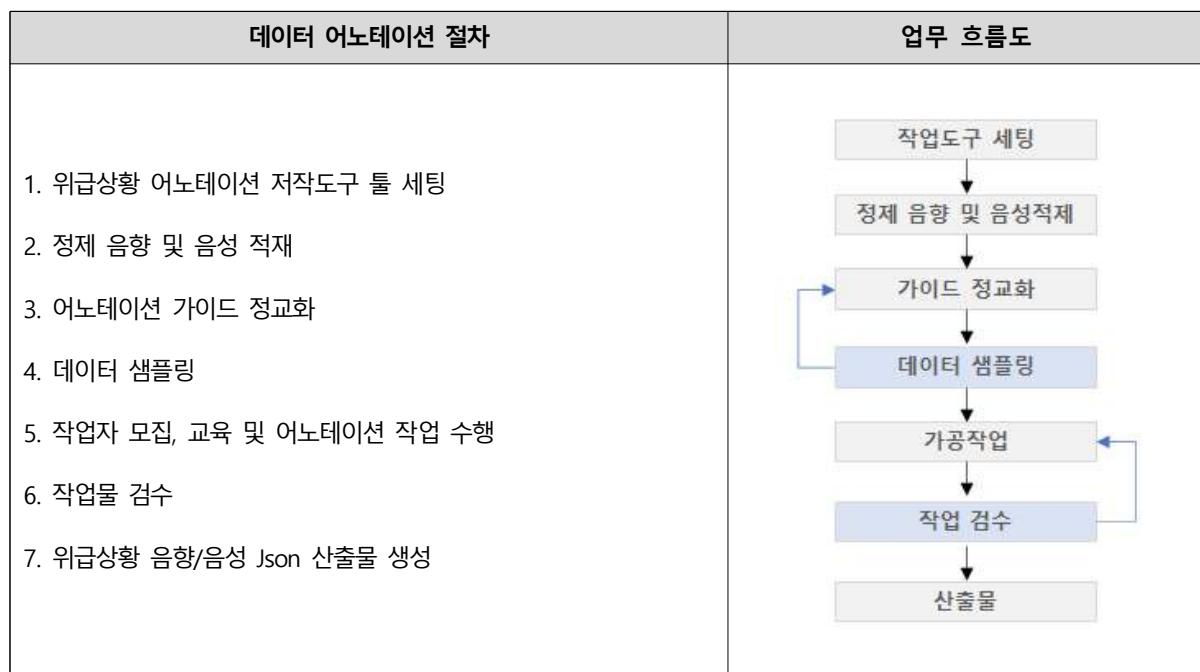
○ 정제 기준

획득 절차	내용
공개 데이터	<ul style="list-style-type: none"> • 파일 형식 준수 여부 (WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) • 저작권 분석과 청각에 의한 음향 샘플링을 통해 음향 데이터 선별 및 폐기 <ul style="list-style-type: none"> - 일반 성인 청각 능력을 기준으로 데이터와 분류체계 합치 여부, 소리 크기 적정성을 이용하여 판별 - 16종 분류체계 및 설정 시나리오에 따른 수집 데이터 수집데이터 적정성 판별 • 2인 1조 및 3인 1조 등 전문 검수팀에 의한 교차 검증을 이용하여 최종 판별 <ul style="list-style-type: none"> - 1인 O, 다른 1인 X 시 : 추가 1인 투입하여 판단 후 사용 및 폐기
크라우드 소싱	<ul style="list-style-type: none"> • 녹음파일 형식 준수 여부 (WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) • 품질기준 여부 <ul style="list-style-type: none"> - 데이터의 잡음 유무(음성 대비 잡음이 클 경우 제외) - 적정한 소리 크기(식별이 어려운 작은 소리는 제외) - 발음의 정확성(발음이 너무 부정확한 경우 제외) - 데이터의 다양성(녹음기와 거리, 다양한 실내 위치 등 고려) • 개인정보보호에 대한 내용 준수 여부

획득 절차	내용
자체 녹음	<ul style="list-style-type: none"> 유해한 데이터 수집 여부 정제 시 3인 이상의 전문 검수 팀에 의해 등록된 음성/음향 데이터에 대한 전수검사 실시 녹음 파일 형식 준수(WAV파일, 16bit, 44.1kHz Mono(CD녹음 음질)) 품질기준 여부 <ul style="list-style-type: none"> - 데이터의 잡음 유무(음성 대비 잡음이 클 경우 제외) - 적정한 소리 크기(식별이 어려운 작은 소리는 제외) - 발음의 정확성(발음이 너무 부정확한 경우 제외) - 데이터의 다양성(녹음기와 거리, 다양한 실내 위치 등 고려) 16종 분류체계 및 설정 시나리오와 수집 음성을 비교하여 적정성 판단 개인정보보호에 관한 내용 준수 여부 유해한 데이터 수집 여부 정제 시 3인 이상의 전문 검수팀에 의해 등록된 음성/음향 데이터에 대한 전수검사 실시

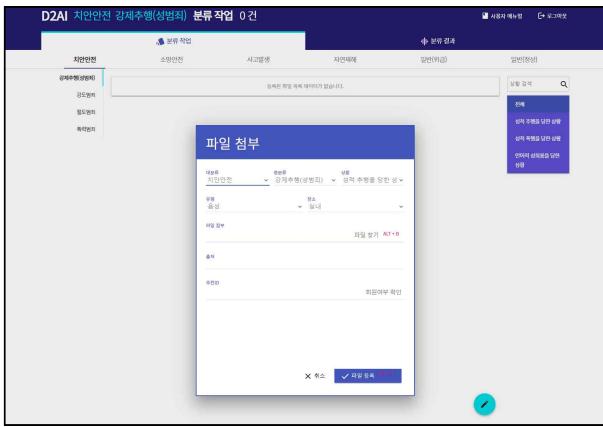
2.4 어노테이션/라벨링

2.4.1 데이터 어노테이션 절차

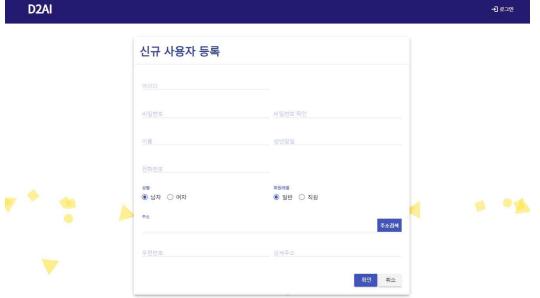
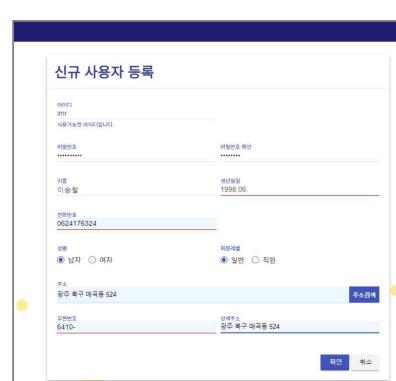


2.4.1.1 데이터 적재

○ 정제 데이터 가공 제작 툴 적재

정제 녹음파일 적재 리스트 예시 화면	데이터 흐름도
	<pre> graph TD A[원천데이터] --> B[정제데이터] B --> C[데이터 선별] C --> D[데이터 샘플링] D --> E[플랫폼 적재] E --> F[작업 검수] F --> G[산출물] B <--> C </pre>

2.4.1.2 크라우드 소싱 및 레이블링 플랫폼

크라우드 소싱 및 레이블링 플랫폼 일반 매뉴얼 > 회원가입	
회원가입 > d2ai.co.kr 사이트 접속	회원가입 > 회원가입 메뉴
	
회원가입 > 정보입력	회원가입 > 가입완료
	

크라우드 소싱 및 레이블링 플랫폼 일반 매뉴얼 > 어노테이션/라벨링 작업

로그인> d2ai.co.kr 사이트 로그인



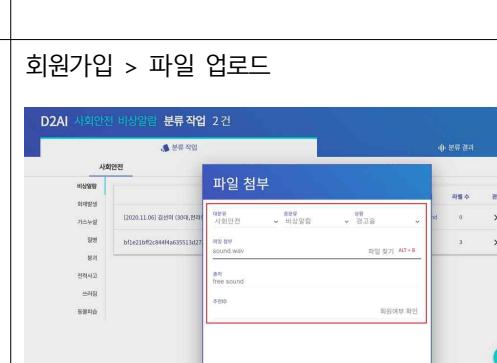
로그인> 로그인 후 사용자 매뉴얼



라벨링 작업> 분류 작업



회원가입 > 파일 업로드



라벨링 작업 > 레이블링 추가

시작인덱스: 0.00
종료인덱스: 0.00
라벨: 경고음
시작: 3.84 종료: 6.99
유형: 경고음
설명: 경고음

라벨링 작업 > 레이블링 유형선택

시작	종료	유형	설명
3.84	6.99	경고음	경고음

라벨링 작업 > 레이블링 정보입력

시작	종료	유형	설명
3.84	6.99	경고음	경고음

라벨링 작업 > 레이블링 완료

시작	종료	유형	설명
3.84	6.99	경고음	경고음

2.4.2 어노테이션/라벨링 기준

2.4.2.1 어노테이션/라벨링 일반 기준

항목	내용
1. 휴면 에러 방지	<ul style="list-style-type: none"> - 데이터 가공 기준은 알고리즘 개발 방향 및 학습방법론에 따라 차이가 있으며, 학습용 데이터는 해당 기준에 부합되는 기준(Labeling 방법, 메타데이터 유형/종류/단계 등...)에 따라 가공이 되도록 진행 - 가공에 대한 기준이 명확히 숙지되지 않으면 기준에 맞지 않는 가공에 따른 휴면 에러가 발생하므로 작업전 기준 숙지 - 작업자의 컨디션, 주변상황 들에 의한 Error의 발생 가능. 다량의 데이터에 대한 Annotation 진행에 따른 Error 발생 가능성 있으므로 일정 시간 업무 후 휴식 시간을 두어 에러 방지 - 다량의 데이터를 가공함에 있어 학습방법론 변경 등에 따른 기준변경 등이 발생 할 수 있으며, 이를 작업에 반영하는 과정에서 Error가 발생할 수 있으므로 학습 방법론 변경 최소화
2. 리스크 대비 방안	<ul style="list-style-type: none"> - 작업자의 숙련도 부족에 따른 Error 발생을 대비하여 작업자에 대한 업무적합도 확인 절차 진행 - 라벨링은 집중력, 판단력, 지속력이 중요한 작업으로서, 적정한 인원 선발을 위해 인성면접 후 Pilot 교육/평가를 통해 학습용 데이터 가공업무에 적합한 인원선발 진행 교육 및 실습 - 작업자에 대한 학습용 데이터 가공 기준 사전 교육을 진행하여 기준에 대한 지식 베이스의 숙련도를 확보 - 교육진행 후 해당 내용에 대한 가공대상 객체별 실습 및 피드백을 통하여 지식 베이스의 숙련도를 실제 작업 간 판단에 대한 숙련도로 확장
3. 정기 평가 실시	<ul style="list-style-type: none"> - 라벨링 작업자의 숙련도 확인 및 숙련도 향상을 확인하기 위한 정기평가 실시 - 단위로 평가하며, 평가 결과에 따라 인원교체 및 업무 수량 조정에 활용 - 학습용 데이터 가공 작업 간 작업자 2인 이상을 1개의 단위(조, 파트 등) 해당 단위 내에서 작업자 간 교차검수를 통하여 검수 진행
4. 기준 변경	<ul style="list-style-type: none"> - 다량의 학습용 데이터의 경우 라벨링 작업도중 기준변경 등이 발생하였을 경우 이미 작업된 학습용 데이터에 대하여 다시 한 번 전수 검수/수정 등을 진행 - 가공 기준을 기준으로 한 검수도구를 활용하여 Error 발생 위치 및 변경된 기준에 따른 수정이 필요한 항목을 파악하고 수정/변경작업 진행 - 속성 값 변경 등의 사항은 도구 내에서 자동변경 될 수 있도록 지원 (스크립트 활용도 가능)

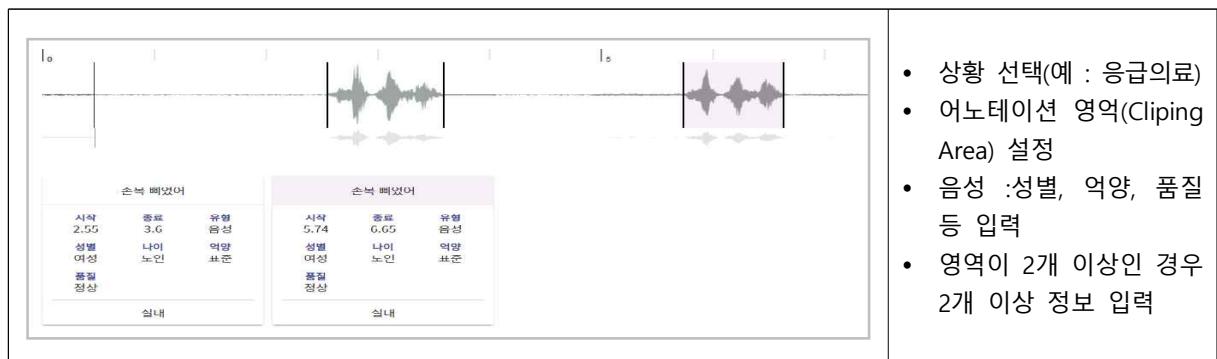
2.4.2.2 어노테이션/라벨링 표준 (예시)

항목	내용
파일명	3.절도범죄_606014.wav
플레이 시간	20.9
라이센스	cc 0(저작권없음)
대분류	치안안전
중분류	절도범죄
상황	치안안전_도둑이야_비명소리
내용	도둑이 침입한 상황
플레이 시작 시간	5
플레이 종료 시간	15.9
유형	음성
성별	여자
연령	성인
지역	서울
역량	표준
구성 예시	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <div style="background-color: #0070C0; color: white; padding: 5px; text-align: center;">정보</div> <pre>{ "info": { "description": "위급상황 음향/음성 데이터셋", "url": "https://www.aihub.co.kr", "version": 1.0, "year": 2020, "contributor": "NIA2020", "date_created": "2020-12-30" }, "license": { "url": "https://www.aihub.co.kr", "name": "CC 0" } }</pre> <div style="background-color: #0070C0; color: white; padding: 5px; text-align: center;">라이선스</div> <pre>{ "audio": { "fileName": "1-3-8_371403.wav", "fileSize": 772676, "duration": 8.76, "audioType": "voice", "fileFormat": "wav", "sampleRate": "44.1kHz", "bitRate": 16, "recodingType": "Mono", "acquisitionType": "제작", "acquisitionDevice": "삼성 갤럭시탭S6", "acquisitionPlace": "실내" } }</pre> <div style="background-color: #0070C0; color: white; padding: 5px; text-align: center;">오디오</div> <pre>[{ "audio_id": "1-3-8_371403_1", "area": { "start": 1.92, "end": 2.86 }, "categories": { "category_01": "치안안전", "category_02": "절도범죄", "category_03": "도둑이 침입한 상황" }, "note": "도둑이 침입한 상황", "audioType": "voice", "gender": "male", "generation": "adult", "residence": "서울", "dialect": "standard", "soundQuality": "정상" }, { "audio_id": "1-3-8_371403_2", "area": { "start": 5.59, "end": 6.52 }, "categories": { "category_01": "치안안전", "category_02": "절도범죄", "category_03": "도둑이 침입한 상황" }, "note": "도둑이 침입한 상황", "audioType": "voice", "gender": "male", "generation": "adult", "residence": "서울", "dialect": "standard", "soundQuality": "정상" }] }</pre> </div> </div>

2.4.2.3 상황별 어노테이션/라벨링 기준

위급상황	어노테이션/라벨링 기준
강제추행	강제추행 관련 위급상황 오디오가 들리기 시작하는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
강도범죄	강도관련 특징 오디오가 들리기 시작하는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
절도범죄	도둑 관련 특징 오디오가 들리는 지점을 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
폭력범죄	폭력발생 특징 오디오가 들리기 시작하는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
화재	화재 관련 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
갇힘	갇힘 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
응급의료	응급의료 특징 오디오가 들리기 시작하는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
전기사고	전기사고 특징 오디오가 들리기 시작하는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
가스사고	가스사고 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
낙상	낙상 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
붕괴사고	붕괴사고 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
태풍/강풍	태풍/강풍 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
지진	지진 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
도움요청	도움요청 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
실내	정상 실내 활동 관련 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링
실외	정상 실외 활동 특징 오디오가 들리는 지점 기준으로 파형 내 이벤트 발생 구간을 지정 후 라벨링

[어노테이션 예시]



2.4.2.3.1 상황별 위급상황에 대한 진위 파악을 위한 기준

- 음성/음향 데이터로부터 위급상황을 판단하는 가이드라인 설정함에 있어 영상과 달리 음성신호는 속이는 것이 훨씬 수월하다는 전제를 두어야 함 (예를 들어, 웃으면서 슬퍼하는 음성 연기가 얼마든지 가능하고, 친구와 놀고 있으면서 도서관에 있다고 하는 것도 가능)
- 속이려고 하는 사람의 의도를 파악하는 것은 종합적 판단에 의존할 수밖에 없으며, 이 판단에 고려되어야 할 부분을 아래와 같이 정의하며, 원데이터의 상황과 불일치하는 부데이터가 하나라도 나오면 거짓으로 판정하고 제외

- 1) 원데이터 (음성) : 속이려고 하는 화자의 텍스트
- 2) 부데이터1 (음성) : 주변 사람의 목소리로 화자가 말하는(또는 외치는) 텍스트와의 상관성이 있으면 포함, 없으면 제외
- 3) 부데이터2 (음향) : 주변 환경음(또는 배경음)으로 화자의 텍스트와 상관성이 있으면 포함, 없으면 제외
- 4) 부데이터3 (음향) : 순간적인 음향이 화자의 텍스트와 상관성이 있으면 포함, 없으면 제외

2.4.2.3.2 상황별 위급상황에 대한 진위 파악을 위한 가이드라인 예시

1) 강제추행 : 강제추행 및 성폭력 등에 따른 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 도와주세요 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 강제추행 및 성폭력 상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 강제추행 및 성폭력 상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 강제추행 및 성폭력 상황과 상관성이 있으면 포함, 없으면 제외

2) 강도범죄 : 강도범죄에 의해 발생하는 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 강도야, 살려주세요 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 강도 상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 강도 상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 강도 상황과 상관성이 있으면 포함, 없으면 제외

3) 절도범죄 : 도둑, 절도에 의해 발생하는 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 도둑이야, 훔쳐갔어요 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 절도상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 절도상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 절도상황과 상관성이 있으면 포함, 없으면 제외

4) 폭력범죄 : 싸움, 폭력 등에 의해 발생하는 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 싸움 났나봐, 싸움났다, 욕설 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 폭력상황과 상관성이 있으면 포함, 없으면 제외

- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 폭력상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 폭력상황과 상관성이 있으면 포함, 없으면 제외

5) 화재 : 화재 등에 따라 발생하는 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 불이야 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 화재상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 화재상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 화재상황과 상관성이 있으면 포함, 없으면 제외

6) 간힘 : 승강기 간힘 등에 의해 발생하는 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 갇혔어요, 열어주세요 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 간힘상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 간힘상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 간힘상황과 상관성이 있으면 포함, 없으면 제외

7) 응급의료 : 앰뷸런스 호출 등에 의해 발생하는 응급 의료 구급요청 신고 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 신고해주세요, 거기 119죠? 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 응급의료상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 응급의료상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 응급의료상황과 상관성이 있으면 포함, 없으면 제외

8) 전기사고 : 감전사고 등에 발생하는 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 감전이야, 누전이야 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 전기사고상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 전기사고상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 전기사고상황과 상관성이 있으면 포함, 없으면 제외

9) 가스사고 : 가스누출 등에 따라 발생하는 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 가스 새는 냄새난다, 터진거 아냐? 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 가스사고상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 가스사고상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 가스사고상황과 상관성이 있으면 포함, 없으면 제외

10) 낙상 : 낙상사고 등에 따라 발생하는 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 넘어졌어요 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 낙상상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 낙상상황과 상관성이 있으면 포함, 없으면 제외

- 부데이터3 (음향) : 순간적인 음향이 낙상상황과 상관성이 있으면 포함, 없으면 제외

11) 붕괴사고 : 폭발상황에 따른 위급상황 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 무너진다, 부서진다 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 붕괴상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 붕괴상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 붕괴상황과 상관성이 있으면 포함, 없으면 제외

12) 태풍/강풍 : 태풍/강풍에 의한 위급사항 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 태풍이다, 강풍이야 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 태풍/강풍상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 태풍/강풍상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 태풍/강풍상황과 상관성이 있으면 포함, 없으면 제외

13) 지진 : 지진 발생에 의한 위급사항 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 지진이다, 지진이야 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 지진상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 지진상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 지진상황과 상관성이 있으면 포함, 없으면 제외

14) 도움요청 - 음성 : 위급상황에서 119등에 도움요청에 의해 발생되는 신고 음향 및 음성

- 원데이터(음성) : 속이려고 하는 화자의 텍스트 (예 : 신고해주세요 등 다급한 음성)
- 부데이터1 (음성) : 주변 사람의 목소리로 도움요청상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터2 (음향) : 주변 환경음(또는 배경음)이 도움요청상황과 상관성이 있으면 포함, 없으면 제외
- 부데이터3 (음향) : 순간적인 음향이 도움요청상황과 상관성이 있으면 포함, 없으면 제외

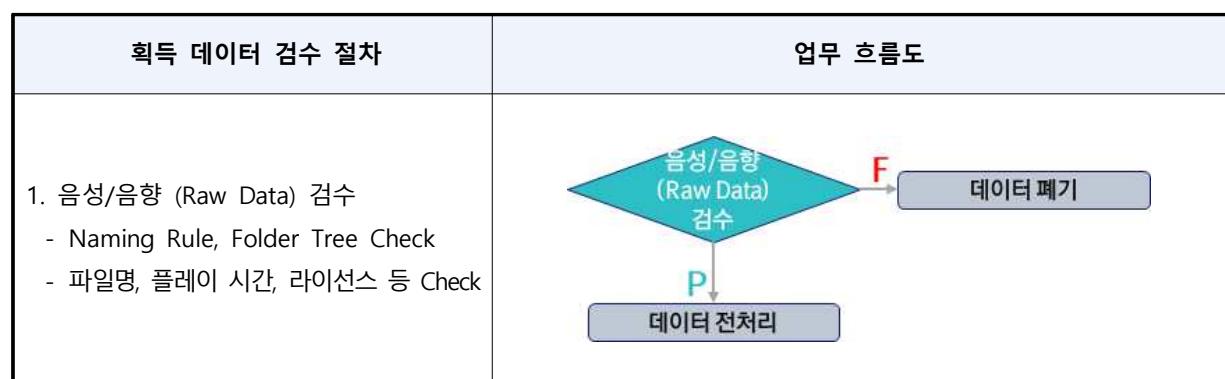
2.5 검수

2.5.1 수행 단계 및 조직



2.5.2 검수 절차

2.5.2.1 획득 데이터 검수 절차



※ 정제작업이 이루어지기 이전단계인 획득 단계의 Raw Data에 대해서는 각 파일의 식별자, 파일명칭, 파일폴더, 생성일자, 라이선스, 분류, 종류, 크기, 시작, 종료 등을 숙련 검수자(전문가)에 의하여 점검

2.5.2.2 정제 데이터 검수 절차

정제 데이터 검수 절차	업무 흐름도
<p>2. 데이터 교차 검수(담당자)</p> <ul style="list-style-type: none"> - 대분류, 중분류, 상황 Check - 시작시간, 종료시간, 유형 등 Check 	<pre> graph TD A[전처리 데이터(작업자)] --> B{데이터 교차 검수(담당자)} B -- F --> C[기준 미달] C --> D[추가 데이터 전처리] D --> A B -- P --> E[데이터 사용] </pre>

※ 가공작업이 수행되기 이전단계인 정제작업 내용에 맞는 품질검증이 이루어지며, 이미지 데이터의 경우에는 식별자, 파일명칭, 파일폴더, 생성일자, 라이선스, 분류, 종류, 크기, 시작, 종료 등의 항목들에 대한 품질 점검을 숙련 검수자(전문가)에 의하여 진행

2.5.2.3 가공 데이터 검수 절차

가공 데이터 검수 절차	업무 흐름도
<p>1. 검수용 알고리즘 전수 검수</p> <ul style="list-style-type: none"> - 구문적 정확성을 Check하고 오류 사항을 Annotation팀에 전달 - 구문적 정합성을 프로그램을 활용하여 체크 <p>2. 2인 교차 수작업 검수</p> <ul style="list-style-type: none"> - 의미적 정확성을 수작업 샘플 검증후 결과 Annotation팀에 전달 - 목표 수준까지 검수자 순환을 통한 중복 검사 실시 	<pre> graph TD A[어노테이션(작업자)] --> B{검수용 알고리즘 전수 검수} B -- F --> C[오류 발견] C --> D[상호 피드백 및 반복 수행] D --> A B -- P --> E[1차 검수 통과] E --> F{2인 교차 수작업 검수} F -- F --> G[오류 발견] G --> H[상호 피드백 및 반복 수행] H --> A F -- P --> I[2차 검수 통과] </pre>

※ Annotation이 완료된 데이터셋과 메타데이터에 대한 품질검증을 숙련 검수자(전문가)에 의하여 진행
 ※ 우선 검증용 알고리즘으로 전수 데이터를 검증한 후, 오류로 판정된 데이터에 대해서는 수작업 검증 팀에 의해 육안 검수를 진행
 ※ 검증결과는 수작업 Annotation팀과 공유하여 상호 feedback을 주고받으면서 목표 품질수준에 도달할 때까지 iterative하게 반복
 ※ 수작업 검증 작업은 2인 이상 번갈아 가며 교차검증을 반복함으로써 품질 확보 달성

2.5.2.4 활용 데이터 검수 절차

활용 데이터 검수 절차	업무 흐름도
<p>1. 유효성 결과 검수</p> <ul style="list-style-type: none"> - 활용시 발생하는 오류 수정 - 오류 원인 분석 후 Annotation팀 전달 - 데이터 품질 확보 후에는 AI 알고리즘에 대한 개선 검토 	<pre> graph TD A[데이터 개발 · 유통] --> B[에러 발생] B --> C[오류대장 취합] C --> D{오류 원인 분석} D -- P --> E[오류 이력관리] D -- F --> G[Annotation 팀 전달] G --> H[오류 해결] H --> A </pre>

- ※ 데이터가 유통된 후 활용되면서 발생하는 에러들은 오류대장으로 취합하여 관리하며, 오류 원인을 분석한 후 Annotation팀에 전달
- ※ 동일한 오류가 반복되지 않도록 오류 이력관리 등의 품질통제관리 진행

2.5.3 검수 기준

1) 검수 기준 개요

- 일관된 고품질의 AI 학습용 위급상황 음성/음향 데이터 세트를 구축하기 위하여 다음과 같은 품질검수 기준을 정의함

① 원천데이터 수집 검수

- 위급상황 음성/음향 데이터 품질기준 충족 시
- 위급상황 음성/음향 데이터 클래스 판단 기준(자체기준) 충족 시
 - * 데이터 가공(Annotation)을 통해 만들어진 Json파일과 Meta data의 각 항목들을 시스템 프로그램을 통해 검수하여 비어있거나 초과되어 있으며, 잘못된 형식의 정보가 들어있지 않을 경우 검수 통과

② 원천데이터 정제(비식별화) 검수

- 위급상황 음성/음향 데이터 내 개인정보, 공간정보, 조사정보에 대한 비식별화(RGB 0.0.255) 100% 진행 여부

③ 원천데이터 가공(라벨링/어노테이션) 검수

- 음성/음향 데이터 당 1개 이상의 라벨링(바운딩 박스) 존재 여부
- 라벨링 작업 결과의 “2.5.2 어노테이션/라벨링 기준” 충족
- “1.2.2 데이터 제공항목 및 항목별 속성정보” 및 “1.3 어노테이션 포맷” 충족

④ 최종검수

- 구축활용 가이드라인 준수여부 최종 검수

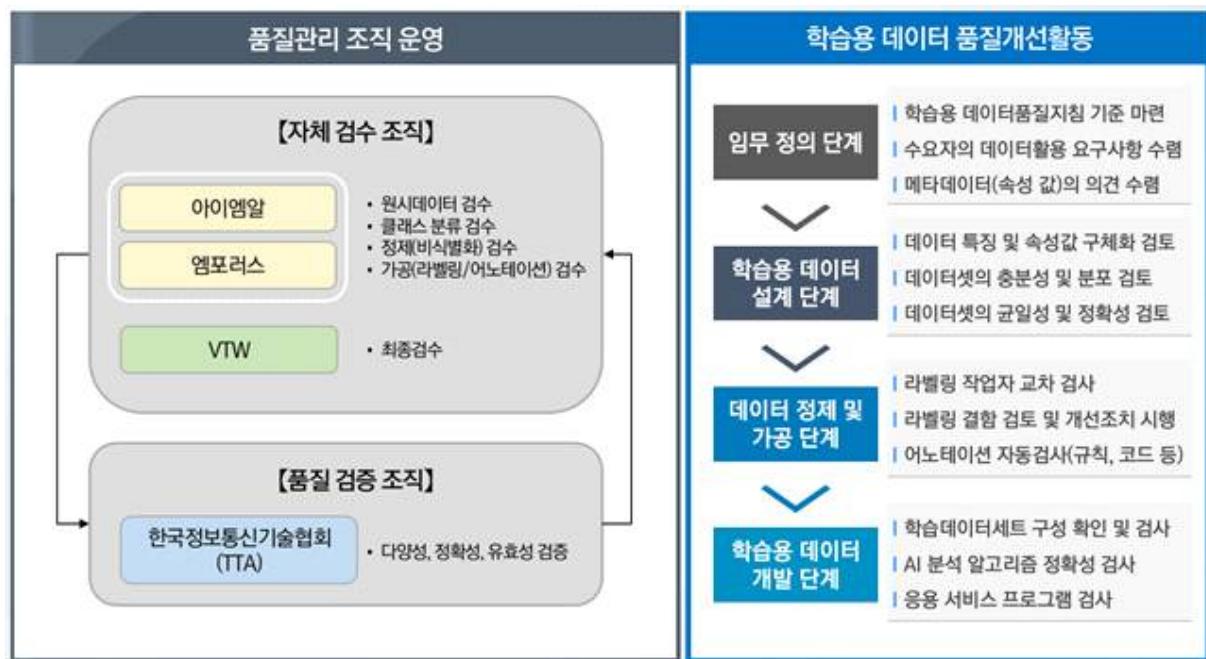
⑤ 데이터 품질검증

- 외부기관 TTA 검증
- 데이터의 다양성과 정확성, 유효성을 검증하기 위해 아래와 같은 검증 항목 및 세부 계획서를 외부검증 기관인 TTA에 제출하여 데이터 품질검증을 진행함

<AI 학습용 데이터 품질검수 항목 및 기준>

품질특성	항목명	측정 지표	정량 목표
다양성	위급 상황 시나리오	분류	16종
정확성	구문적 정확성	정확도	100%
	의미적 정확성	정확도	80%
유효성	위급상황 인식률	F1-score	0.9 이상

2.5.4 검수 조직

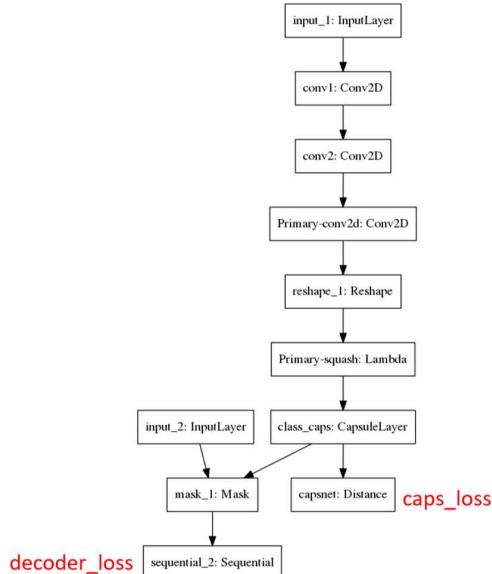


2.6 활용

2.6.1 활용 모델

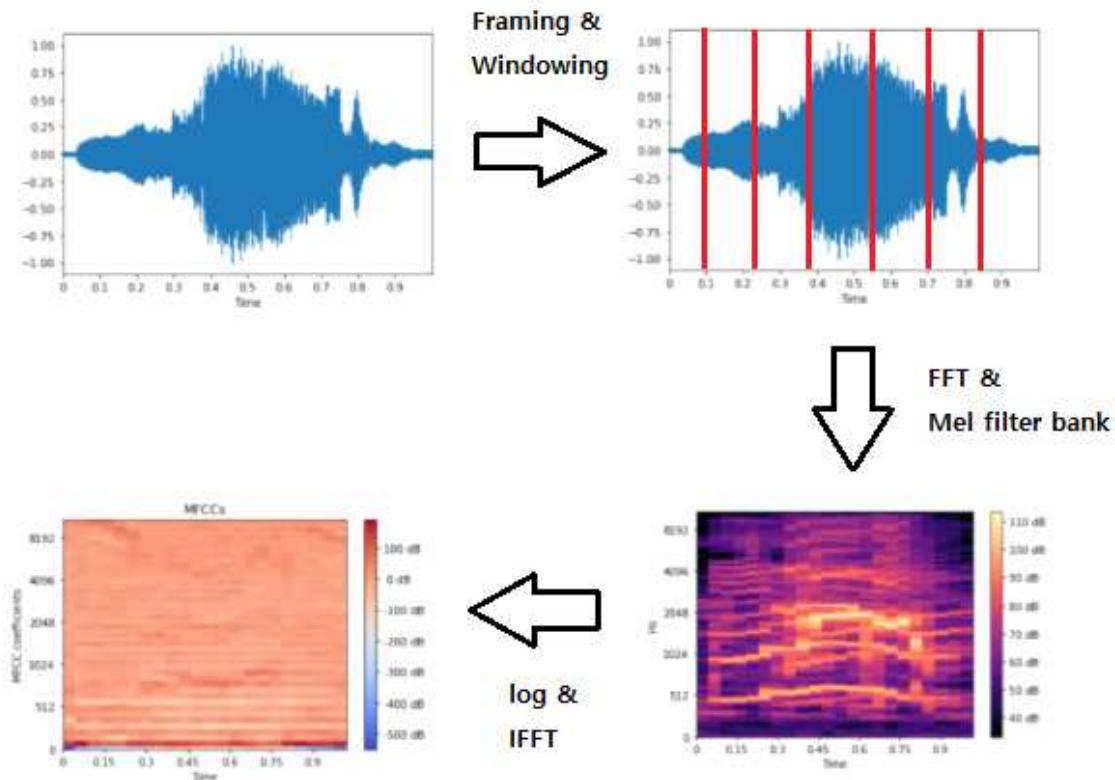
2.6.1.1 모델 학습

CapsNet based structure



- 취득된 음성·음향 오디오 데이터를 Image데이터 MFCC(Mel-Frequency Cepstral Coefficient)로 변환하여 Capsulenet 학습 데이터로 사용
- Capsulenet은 CNN이 위치 정보에 취약하다는 점을 고 차원의 entity를 통해 보완
- Network를 capsulation하여 더 많은 entity property 정보를 획득
- Dynamic routing을 이용하여 resampling하므로써 더 복잡한 entity를 자유롭게 표현 가능

[학습 모델 : Capsulenet(Capsnet)]



[이미지 변환 과정 : MFCC(Mel-Frequency Cepstral Coefficient)]

명칭	내용
Windowing	신호의 구간을 선택
FFT(Fast Fourier Transform)	신호를 주파수 성분으로 변환
Mel filter bank	스펙트럼을 Mel spectrum으로 변환
IFFT(Inverse Fast Fourier Transform)	역 FFT

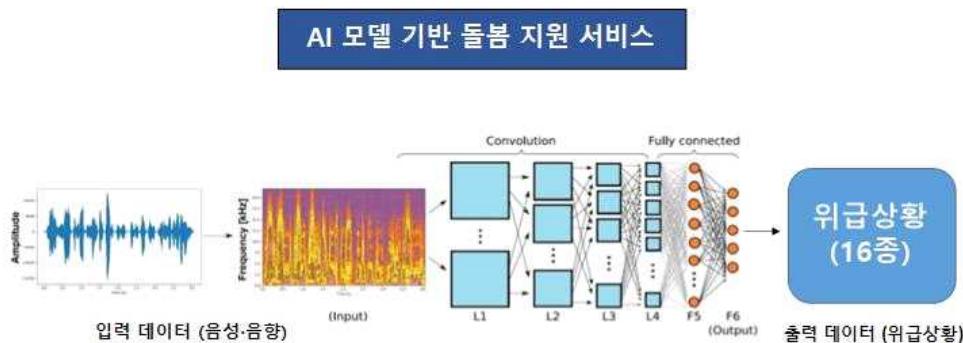
[용어설명 : MFCC(Mel-Frequency Cepstral Coefficient)]

2.6.1.2 시험 계획

항목	내용					
	분류 체계	상황 (시나리오)	성별 구분	연령 구분		
1. 다양성	16종	1000개	2개	4개		
	<div style="border: 1px solid black; padding: 5px; text-align: center;"> 2개 상황이 믹싱 된 경우 (16종 100가지상황에서 3가지 상황 활용 예시) </div>					
	화재 상황	음성 A + 음향 B (도움요청 상황 + 불타는 상황)	시간 / 상황			
	응급 신고 상황	음성 C + 음성 D (상태 및 의식을 확인하는 상황 + 119 신고 상황)	1~300초/상황(시나리오)			
	폭력 상황	음향 E + 음성 F (물건이 깨지는 상황 + 신고를 요청하는 상황)	상황 (시나리오)			
- 16종 분류체계 하위에 다양한 경우의 수를 고려하여 인식률 향상을 통한 데이터 구축						
2. 정확성	<ul style="list-style-type: none"> • 구문적 정확성 <ul style="list-style-type: none"> - 어노테이션 정의(rule)에 따라 모든 json, 구문이 적합하게 작성되었는지(구축 데이터에 대해 프로그램 적으로 검증) • 의미적 정확성 <ul style="list-style-type: none"> - 전체 데이터 중 무작위로 1% 이상 샘플링하여 16종 분류에 정확하게 라벨링이 되었는지 확인 					
3. 유효성	<ol style="list-style-type: none"> 1) 종류 및 수량 : Train(8) : Val(1) : Test(1) 2) 개발 언어, 프로그램 : Python3.6.11, Tensorflow1.15.0, Keras2.2.4, CUDA10.2 3) Docker Image 용량 : 7GB 4) 학습네트워크 개념 : CNN, Capsulenet 중 선별적용 5) 학습 조건 : <ul style="list-style-type: none"> - Epoch : 30 - Batch : 100 - Optimizer : ADAM 6) 검증/평가 방법 : <ul style="list-style-type: none"> - Traning, Validation, Test로 데이터를 구별 - 학습에 Training, Validation을 사용하여 검증하고 Test 데이터로 평가 7) 유효성 목표 : F1-score 0.9이상 					

	<ul style="list-style-type: none"> - F1-score는 모델의 성능을 측정할 때 클래스 데이터가 불균형한 경우 유용하게 사용되는 평가방법 - F1-score는 정밀도(Precision)와 재현율(Recall)의 조화평균 - 정밀도(Precision)는 모델이 참으로 분류한 것 중에서 실제 참인 것의 비율 - 재현율(Recall)은 실제 참인 것 중에서 모델이 참이라고 예측한 것의 비율
4. 시험모델	<ul style="list-style-type: none"> - 목적 : 16종의 위급상황 음성/음향 분류를 통한 인식 - 학습 및 검증/평가 데이터 : 음성 및 음향으로 구성된 위급상황 오디오 파일 20% (700 시간)을 8: 1로 구분 - AI 학습 모델 : CapsNet (훈련 데이터로 해당 모델 학습) - 분류 및 인식 : ① 입력 데이터 : 위급상황 음성/음향 데이터 ② 출력 데이터 : 16종의 위급상황 분류 및 인식 결과

2.6.1.3 서비스 활용 시나리오



입력 (오디오)	출력 (위급상황)	연계 서비스	비고
음성/음향 데이터	강제추행	경찰서	
	강도범죄	경찰서	
	절도범죄	경찰서	
	폭력범죄	경찰서	
	화재	소방서, 병원	
	갇힘	소방서, 병원	
	응급의료	병원	
	전기사고	소방서, 병원	
	가스사고	소방서, 병원	
	낙상	소방서, 병원	
	붕괴사고	소방서, 병원	
	태풍/강풍	소방서, 병원	
	지진	소방서, 병원	
	도움요청	소방서, 경찰서	
	실내	복지기관	
	실외	복지기관	

2.6.2 데이터 활용 분야

- 총 3,500시간의 음성/음향 데이터는 아래와 같은 활용 가능

활용분야	세부 활용 방안
공공분야	<ul style="list-style-type: none"> - 정부나 지자체의 사회 안전에 대한 방안을 모색 - 사회복지시설, 치안 유기기관 등에서 빠른 위급상황 파악
연구분야	<ul style="list-style-type: none"> - 위급상황 시 안전에 대해서 연구자료 제공 - 각종 범죄에 대한 분석 및 연구자료 제공 - 보건 복지 문제에 대한 연구 자료 제공 - 사회 문제 연구에 대한 통계 자료 제공
산업분야	<ul style="list-style-type: none"> - 의료/요양 산업 발전을 위한 인공지능 관련 기업들의 활성화를 도모 - 위험상황을 인지하여 보다 안전한 경비시스템 구축산업에 활용 - AI스피커를 활용한 고령화 인구에 대한 돌봄 서비스 제공
교육분야	<ul style="list-style-type: none"> - AI교육 실습 시 샘플 데이터 제공 - AI 모델 기초 분석에 대한 알고리즘 제공 - 음성/음향에 관한 교육 자료로 제공

2.6.3 데이터 활용 정책

- 음성/음향 데이터는 아래와 같은 활용 가능

활용정책	세부 내용
배포방안	<ul style="list-style-type: none"> - 음향 및 음성 데이터 셋은 AI Hub 사이트를 통하여 다운로드 가능하며, 한국지능정보사회진흥원(NIA)의 정책에 따름 - AI Hub 사이트를 통하여 데이터 공유 시스템을 구축하고 연구자 혹은 기업에서 비교적 자유로운 데이터 학습을 가능 - 음성의 경우 개인정보에 민감한 데이터로, 개인정보 보호를 위해 화자 정보는 공개하지 않으며 데이터 취득 시 개인정보활용 동의서를 음성녹음자에게 서명을 받았기에 배포를 위해서는 반드시 적법한 절차를 거쳐야 함
데이터 활용시 필요조건	<ul style="list-style-type: none"> - 개인정보보호를 위해 피 실험자에게 개인정보동의서 상으로 연구 목적 활용에 대한 공유 동의를 받고 있으며, 상업적 목적으로는 데이터베이스를 직접 공유하지 않고, 학습모델만 제공함. - 연구 목적 활용의 경우, 기관 및 활용 목적 등에 대한 검토 후 데이터셋을 제공하도록 하며, 상업적목적으로 활용하는 기업의 경우에는 학습하고자 하는 인공지능 네트워크를 제공 받음.

2.6.4 인공지능 학습용 데이터 공개·개방 방안

- 구축한 위급상황 음향/음성 데이터는 NIA(한국지능정보사회진흥원)의 AI Hub 사이트(<https://www.aihub.or.kr>)를 통해 공개 예정으로, 일반인 누구나 AI Hub 사이트를 통해 다운로드할 수 있음
- 웹 인터페이스를 기반으로 데이터셋 플랫폼을 구현하여 데이터셋 생산자와 사용자 모두 웹 GUI를 통해 데이터셋을 추가하거나, 이용할 수 있음
- 데이터셋 플랫폼에 대한 사용 설명서 및 사용 예 등을 종합적으로 기술하여 배포

2.6.3 유지보수

- 이메일을 통해 사용자들의 문의사항을 확인하고, 자주 묻는 질문을 정리하여 FAQ로 웹사이트를 통해 제공하며 주기적으로 업데이트 진행 예정

불임. 어노테이션 및 라벨링 매뉴얼

오디오 레이블링 페이지 사용 매뉴얼

2021. 01. 13





목 차

1. 오디오 레이블링 페이지 사용 매뉴얼	1
1.1 회원 가입	1
1.2 로그인	4
1.3 분류 작업	5
1.3.1 파일 업로드	5
1.3.2 라벨링 추가	9
1.3.3 라벨링 수정	14
1.3.4 파일 삭제	15
1.4 분류 결과	16
1.5 음향 상황 관리	17
1.5.1 음향 상황 목록 조회	17
1.5.2 음향 상황 추가	18
1.6 로그아웃	19



제목 : 데이터셋 분류 페이지 사용 매뉴얼

버전
1.3.0

페이지
1(21)

1. 오디오 레이블링 페이지 사용 매뉴얼

1.1 회원가입



<로그인 화면>

- 브라우저에서 d2ai.co.kr로 이동합니다.
- 로그인 페이지로 이동합니다.
- 창 중앙 테이블 하단의 '신규 사용자 등록' 버튼을 클릭합니다.



<회원가입 화면>



<우편번호 찾기 창>

- '신규 사용자 등록'창이 뜨면 정보를 입력합니다.
- 아이디를 넣으면 자동으로 아이디 중복확인이 진행됩니다.
- 주소버튼을 누르면 '우편번호 찾기 창'이 뜨면서 주소를 검색하여 입력합니다.
- 나머지 정보를 기입하고 '확인' 버튼을 눌러 등록을 완료 합니다.

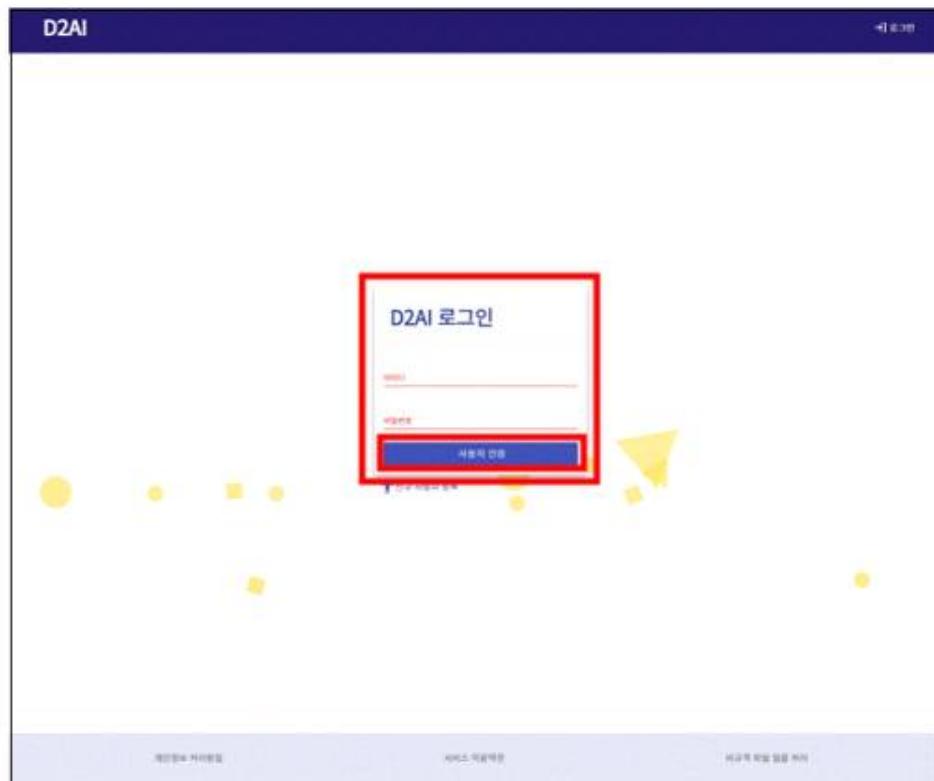


제목 : 데이터셋 분류 페이지 사용 매뉴얼

버전
1.3.0

페이지
4(21)

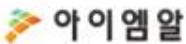
1.2 로그인



The screenshot shows the D2AI login page. At the top, there is a dark blue header bar with the D2AI logo on the left and a search bar on the right. Below the header is a light gray background area containing several yellow circular icons of varying sizes. In the center, there is a white rectangular form with a red border. Inside the form, the text "D2AI 로그인" is displayed above two input fields. Below the input fields is a blue button with the text "사용자 인증". At the bottom of the page, there is a light blue footer bar with three small text links: "개인정보 처리방침", "서비스 이용약관", and "비교적 저렴한 물품 찾기".

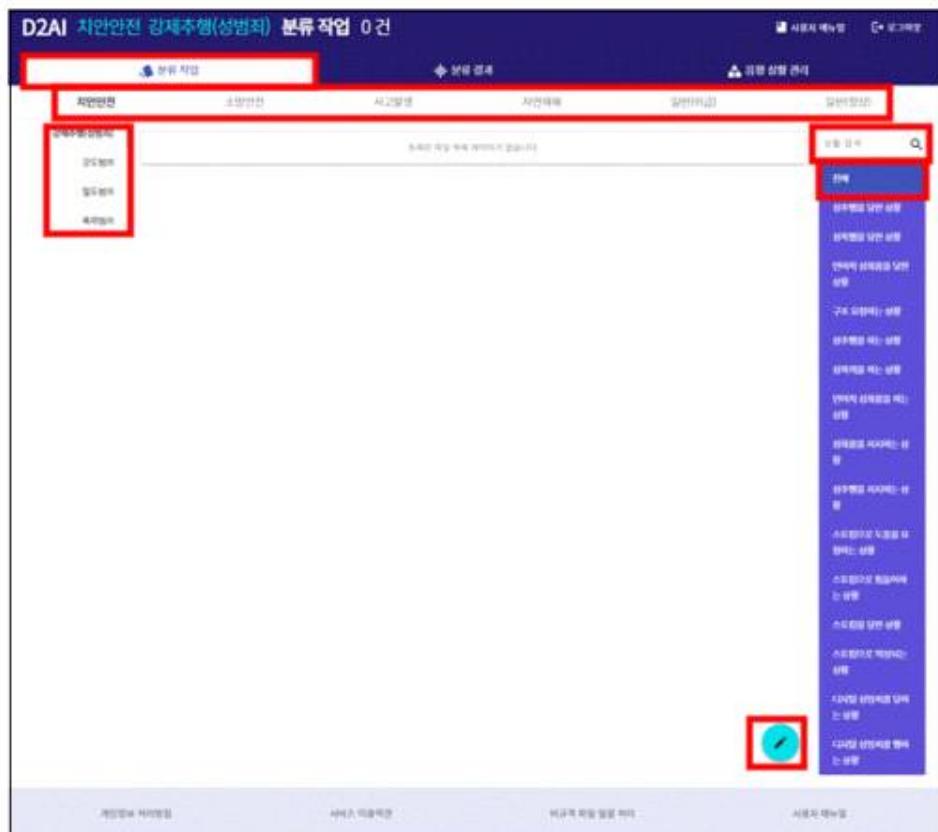
<로그인 화면>

- 로그인 창에 아이디와 비밀 번호를 넣고 '사용자 인증' 버튼을 누릅니다.

 아이엠알	제목 : 데이터셋 분류 페이지 사용 매뉴얼	버전 1.3.0	페이지 5(21)
--	-------------------------	-------------	--------------

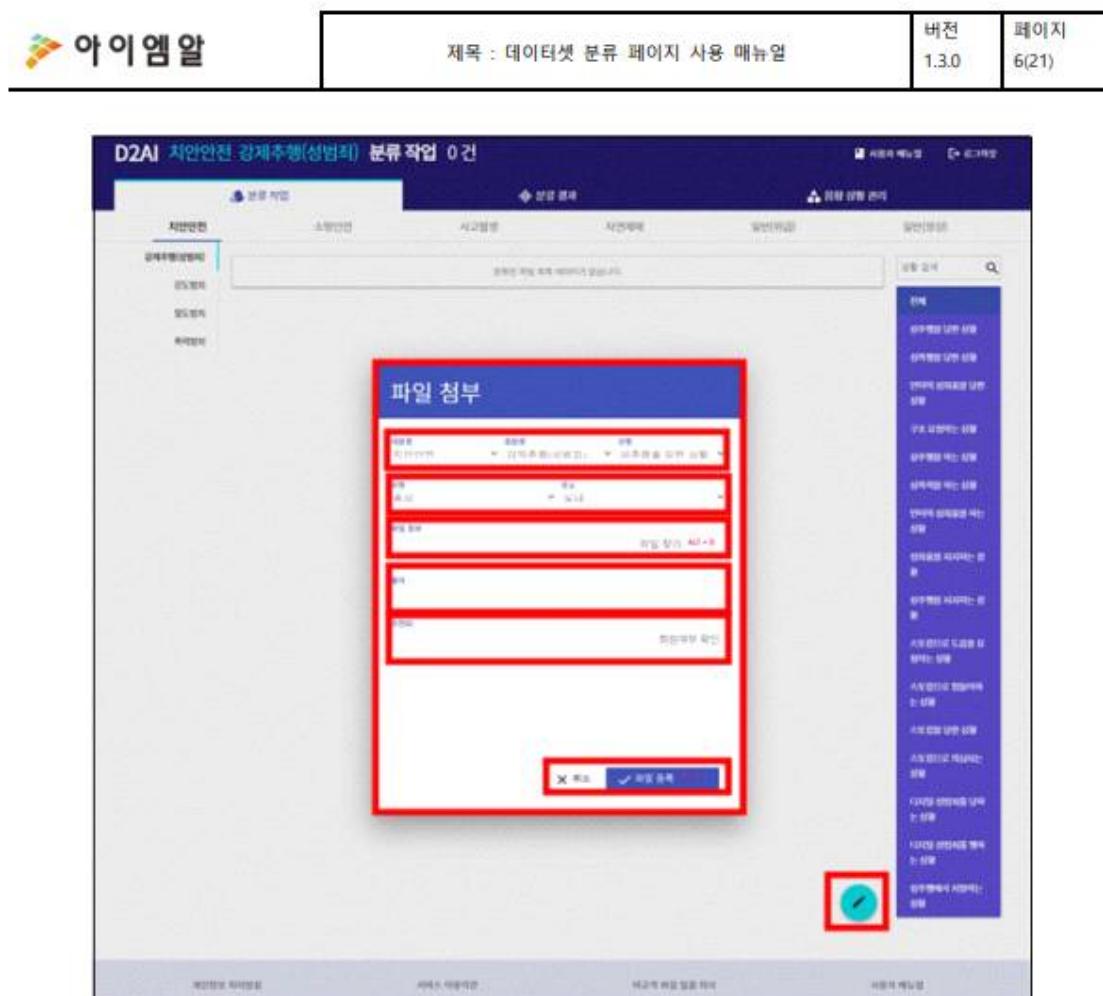
1.3 분류 작업

1.3.1 파일 업로드



<분류 작업 화면>

- 로그인 이후 상단의 ‘분류작업’ 탭을 클릭하여 분류작업 페이지로 들어가고 등록하기 원하는 상단의 탭 (대분류), 좌측의 탭 (중분류), 우측의 탭 (상황)을 클릭한 후 하단의 버튼을 클릭하여 업로드 창을 엽니다.
- 우측의 탭 (상황)은 검색 가능합니다.



<파일 첨부 화면>

- 하단의 버튼을 누르면 파일업로드 창이 뜹니다.
- 파일업로드 창의 상단 셀렉트 박스에서 대분류, 중분류, 상황 설정이 가능합니다. (위에서 언급한 대분류, 중분류, 상황 설정 권장)
- 파일업로드 창의 두 번째 셀렉트 박스에서 유형(음성, 음향)과 장소(실내, 실외)를 선택합니다.
- 세 번째인 ‘파일첨부’ 부분에서 우측 파일 찾기 버튼을 누르고 업로드 할 파일을 선택합니다.
- 하단의 ‘출처’ 부분에 출처관련 정보를 입력합니다.
- 최하단의 추천ID에서 추천하고자 하는 ID를 입력하시고 ‘회원여부 확인’ 버튼을 누르면 입력한 ID가 회원인지 확인이 됩니다. (회원이 아니면 추천 ID가 등록되지 않습니다.)



제목 : 데이터셋 분류 페이지 사용 매뉴얼

버전
1.3.0

페이지
7(21)

- 정보입력이 완료되면 '파일등록' 버튼을 눌러 업로드를 실행합니다.
- 파일은 wav 파일만 업로드 되며 확장자가 wav 파일이고 포맷이 pcm인 경우에 업로드가 가능하며(adpcm 불가), 세부적으로 bit rate는 16bit를 권장합니다.

아이엠알

제목 : 데이터셋 분류 페이지 사용 매뉴얼	버전 1.3.0	페이지 8(21)
-------------------------	-------------	--------------

D2AI 치안안전 강제추행(성범죄) 분류 작업 1 건

분류 작업 분류 관리 사용자 메뉴 등록 관리

지연면전 조정면전 서고방영 자연재해 긴급구조

강제추행(성범죄) 고도경계 청도경찰 출동경비 관리

222537...open-world...woman-to-woman.wmv

화면 검색

화면

신수현을 담은 상황

신수현을 막는 상황

언어적 대화로 막는 상황

기운

구조 요청하는 상황

신수현을 막는 상황

신제작을 막는 상황

언어적 대화로 막는 상황

기운

언제면 차단하는 상황

신수현을 차단하는 상황

스포츠으로 도움을 드리는 상황

스포츠로 활동하여 드는 상황

스포츠로 남한 상황

스포츠로 역사문화 상황

언어를 선택하여 드는 상황

언어를 선택하여 막아 드는 상황

신수현에게서 차단하는 상황

수정

제작 정보 확인 및 수정

서비스 이용약관

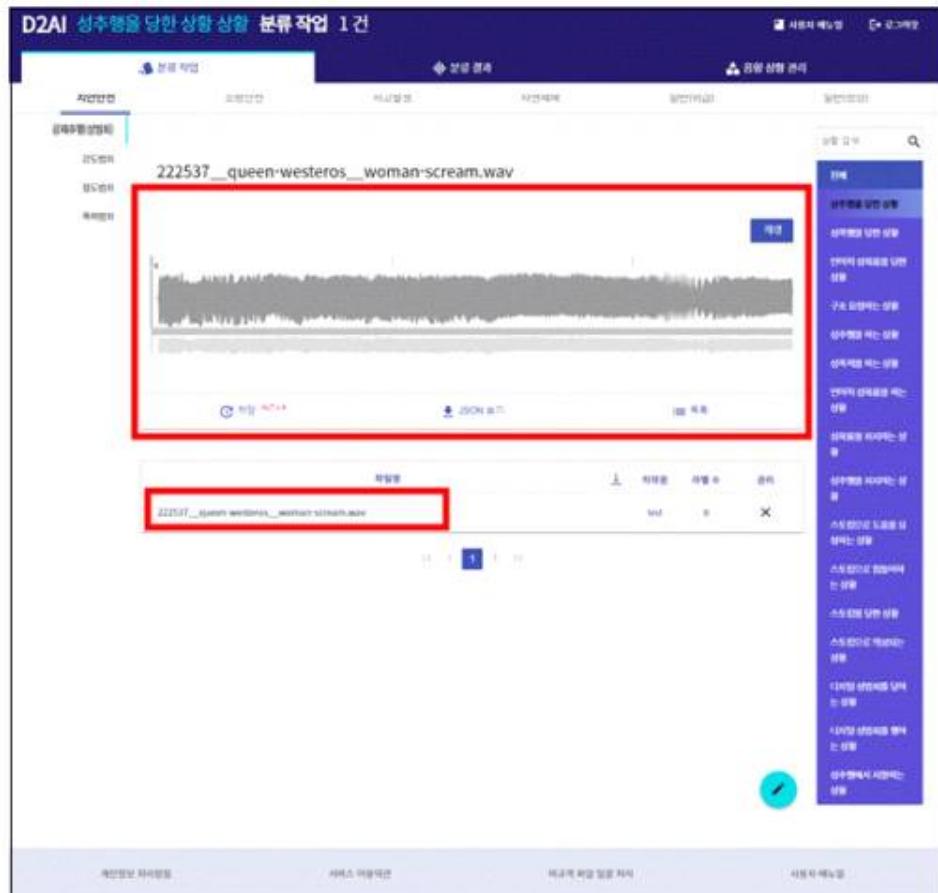
비교적 높은 점수 차이

사용자 메뉴얼

<업로드 이후 화면>



1.3.2 라벨링 추가



<라인>

- 업로드 이후 파일 목록이 뜨고 파일명을 클릭하면 해당 파일을 라벨링 할 수 있습니다.
 - 음성 파일 wave 화면에 드래그를 하면 범위가 지정됩니다.
 - 드래그된 범위를 클릭하면 입력폼이 뜹니다.

The screenshot shows the D2AI platform interface for a classification task. At the top, there's a header with the logo '아이엠알' (Aiml), the title '제목 : 데이터셋 분류 페이지 사용 매뉴얼', version '버전 1.3.0', and page count '페이지 10(21)'. Below the header, the main area has tabs for '분류 작업' (Classification Task), '분류 결과' (Classification Result), and '결과 상황 관리' (Result Status Management). The '분류 작업' tab is active. A file path '222537__queen-westeros__woman-scream.wav' is displayed. On the left, there's a sidebar with categories like '음성' (Voice), '음악' (Music), '인사' (Greeting), etc. The main content area shows a waveform with a red box around a specific segment. Below the waveform is a table with columns '시작 시간' (Start Time), '종료 시간' (End Time), '주제' (Subject), and '설명' (Description). A large red box surrounds this table. To the right of the table are two buttons: '재생' (Play) and '저장' (Save). At the bottom, there are buttons for 'JSON 보기' (View JSON) and '다른 파일' (Other File). The bottom navigation bar includes icons for '파일' (File), '설정' (Setting), '도움말' (Help), and '로그아웃' (Logout).

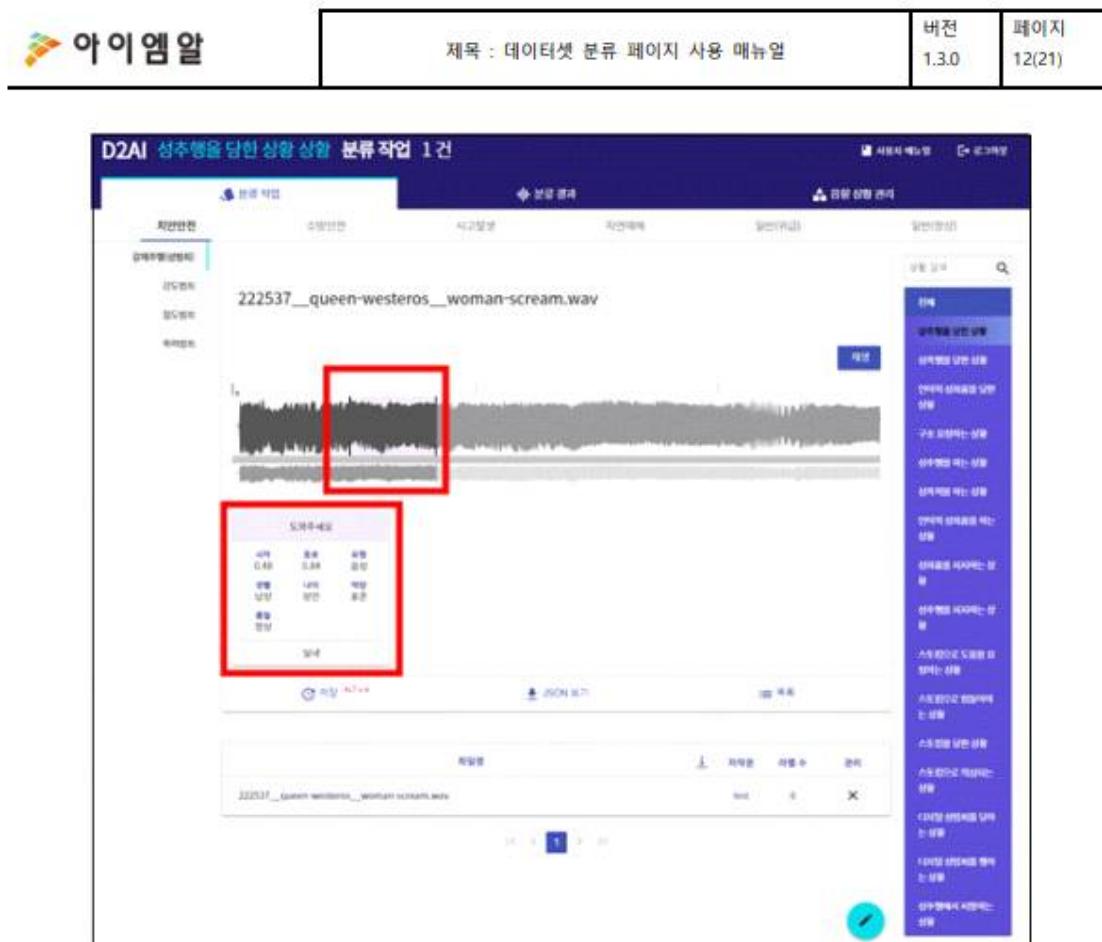
<파일 첨부 시 유형을 '음성'으로 선택한 경우>

- 파일 첨부 시 '유형' 부분을 '음성'으로 선택하면 'note', '성별', '나이', '의양', '품질'의 입력폼이 됩니다. (장소는 파일 첨부 시 선택한 대로 결정)
- '유형'에 맞는 정보를 입력 및 선택하십시오. (* note 값 예시 : 음성일 경우 – 살려주세요 / 음향일 경우 – 개가 짖는 소리)
- 정보입력이 완료되면 하단의 '라벨 저장'버튼을 눌러 저장합니다.
- 상단의 재생버튼을 누르면 재생이 되고 한번 더 누르면 정지가 됩니다. Alt키도 마찬가지 기능을 합니다. '라벨'을 클릭하면 라벨 부분의 소리를 재생할 수 있습니다.

The screenshot shows the D2AI system interface for classifying sounds. At the top, there's a header with the logo '아이엠알' (Aiml), the title '제목 : 데이터셋 분류 페이지 사용 매뉴얼', and version information '버전 1.3.0' and '페이지 11(21)'. Below the header, the main area displays a waveform for the file '222545__queen-westeros__woman-scream.wav'. A red box highlights the waveform itself. Another red box highlights the '재생' (Play) button. A third red box highlights the '라벨' (Label) button, which is highlighted in blue. To the right of the waveform, there's a sidebar with various classification categories listed. At the bottom of the interface, there are buttons for '이전 차례' (Previous) and '다음 차례' (Next), along with other controls like '재생' and '종료'.

<파일 첨부 시 유형을 '음향'으로 선택한 경우>

- 파일 첨부 시 '유형'부분을 '음향'으로 선택하면 'note', '품질'의 입력폼이 뜹니다. (장소는 파일 첨부 시 선택한 대로 결정)
- '유형'에 맞는 정보를 입력하십시오. (* note 값 예시 : 음성일 경우 – 살려주세요 / 음향일 경우 – 개가 짖는 소리)
- 정보입력이 완료되면 하단의 '라벨 저장'버튼을 눌러 저장합니다.
- 상단의 재생버튼을 누르면 재생이 되고 한번 더 누르면 정지가 됩니다. Alt키도 마찬가지 기능을 합니다. '라벨'을 클릭하면 라벨 부분의 소리를 재생할 수 있습니다.



<'라벨저장'버튼을 누른 이후>

- ‘라벨 저장’ 버튼을 누르면 저장한 라벨 목록이 보입니다.
 - ‘라벨 범위’를 클릭한 후 ‘라벨 삭제’버튼을 누르면 ‘라벨 범위’와 해당 라벨 목록이 삭제 됩니다.
 - 추가 수정을 원하면 ‘라벨 범위’를 클릭하여 입력폼이 열리면 수정 후 ‘라벨 저장’ 버튼을 눌러 수정합니다.
 - 추가를 원하면 ‘라벨 범위’를 드래그 하여 클릭 후 내용을 넣고 ‘라벨 저장’버튼을 클릭합니다.

아이엠알

제목 : 데이터셋 분류 페이지 사용 매뉴얼	버전 1.3.0	페이지 13(21)
-------------------------	-------------	---------------

The screenshot shows the AIML labeling interface. At the top, there are three tabs: '결정 작업' (Decision Task), '분류 결과' (Classification Result), and '정답 상황 관리' (Correct Answer Status Management). The '결정 작업' tab is active. On the left, there's a sidebar with categories: '기본인원' (Basic Personnel), '감독방법' (Supervision Method), '감독방식' (Supervision Style), and '작업방식' (Work Method). The main area displays a waveform and some data. Below the waveform is a table with columns '시작' (Start), '중간' (Middle), and '종료' (End) with values 0.40, 0.64, 0.85; and '장별' (Type), '단위' (Unit), '세부' (Detail) with values '내부', '국민', '내부'. A '결과' (Result) section shows '도록주제내' (Topic) with '성우' (Narrator) and '성우명' (Narrator Name). Three buttons at the bottom are highlighted with red boxes: '저장' (Save), 'JSON 열기' (Open JSON), and '목록' (List). To the right, there's a sidebar with various classification labels like '성우에게 보이는 상황' (Situation visible to the narrator), '성우에게 달린 상황' (Situation attached to the narrator), etc.

```

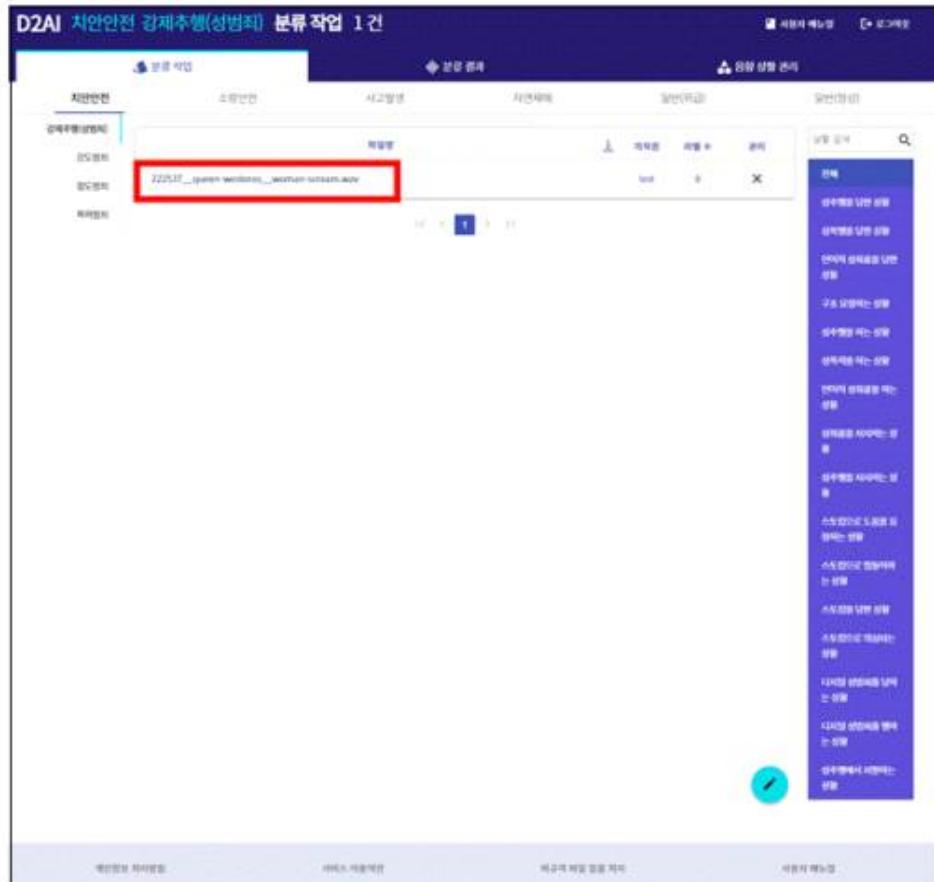
{
  "annotations": [
    {
      "audio": {
        "acquisitionDevice": "설정값을준서열0",
        "acquisitionPlace": "설정",
        "acquisitionType": "설정",
        "audioType": "oice",
        "bitRate": 16,
        "duration": 2.87,
        "fileBase": "seq",
        "filename": "1-1-1_342792",
        "fileSize": 235974,
        "recognitionRate": "99%",
        "sampleRate": "44.1kHz"
      },
      "info": {
        "contributor": "MIA2020",
        "description": "위급상황 음향/음성 데이터",
        "url": "https://www.aiaml.co.kr",
        "version": 1,
        "year": 2020
      },
      "license": {
        "label": "CC 0",
        "url": "https://www.aiaml.co.kr"
      }
    }
  ]
}
  
```

<라벨링 화면>

- 라벨링이 완료되면 하단의 '저장' 버튼을 눌러 라벨링을 등록합니다.
- 중간의 'JSON 열기' 버튼을 누르면 해당 파일의 json 파일을 조회할 수 있습니다. 다시 버튼을 누르면 json 조회 창이 없어집니다.
- 우측의 '목록' 버튼을 누르면 다시 파일 목록으로 돌아갑니다.

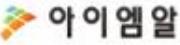
 아이엠알	제목 : 데이터셋 분류 페이지 사용 매뉴얼	버전 1.3.0	페이지 14(21)
--	-------------------------	-------------	---------------

1.3.3 라벨링 수정

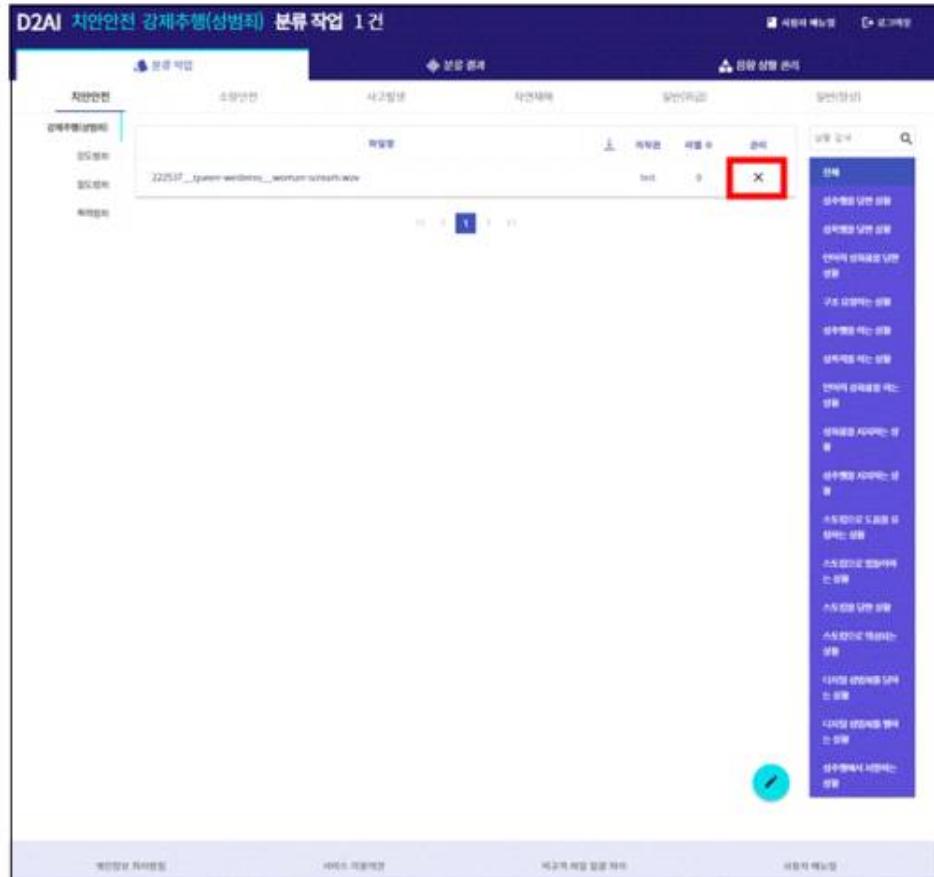


<분류 작업 화면>

- 분류 작업 화면에서 기존에 처리한 파일 목록이 뜹니다.
- 화면에 뜬 파일 목록에서 라벨을 수정할 파일명을 클릭합니다.
- 파일명을 눌러 수정모드로 진입하게 되면 ‘라벨범위’와 라벨 목록이 뜹니다.
- 라벨링 추가와 마찬가지 방식으로 수정이 가능합니다.
- 수정이 완료된 이후 ‘저장’버튼을 눌러 수정을 마칩니다.

 아이엠알	제목 : 데이터셋 분류 페이지 사용 매뉴얼	버전 1.3.0	페이지 15(21)
--	-------------------------	-------------	---------------

1.3.4 파일 삭제

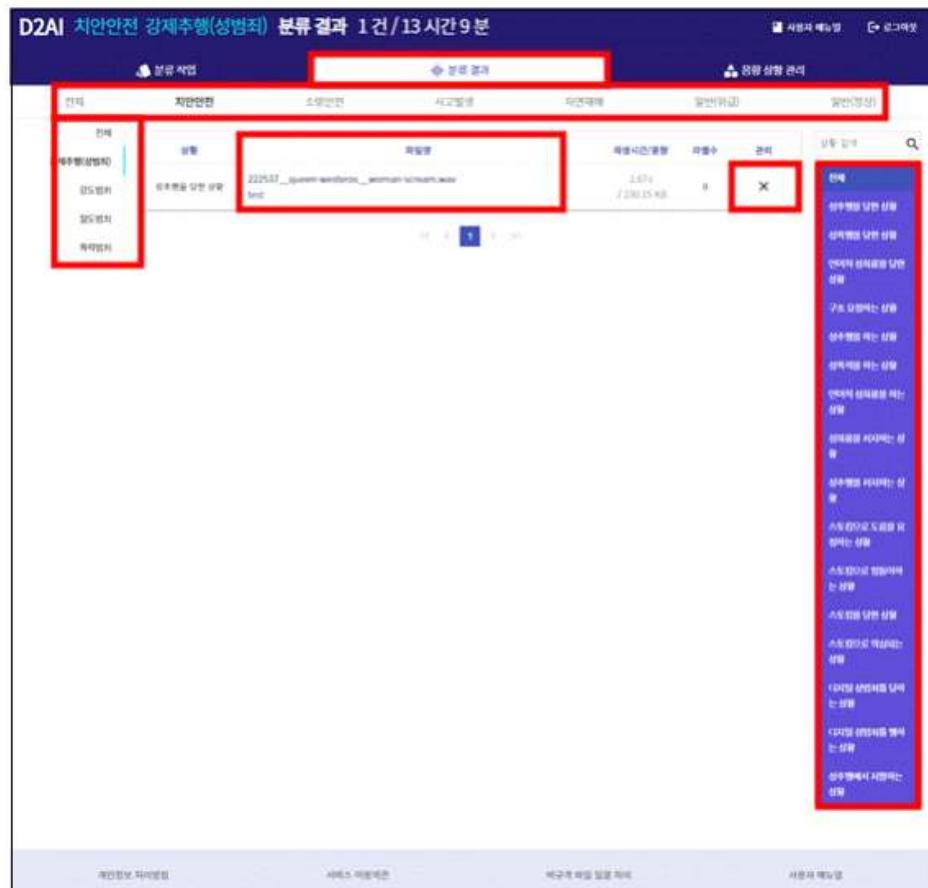


<분류 작업 화면>

- 분류 작업 초기 페이지의 목록에서 삭제하고자 할 파일의 우측 'X'버튼을 클릭하고 파일 삭제를 확인하면 파일 삭제가 완료 됩니다.

	제목 : 데이터셋 분류 페이지 사용 매뉴얼 버전 1.3.0 페이지 16(21)
---	---

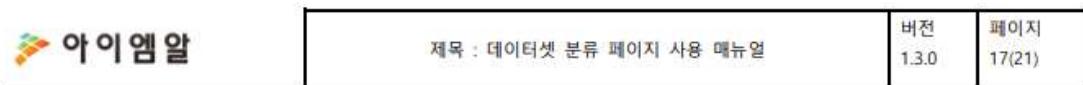
1.4 분류 결과



The screenshot shows the D2AI Classification Result page. At the top, it says 'D2AI 치안안전 강제추행(성범죄) 분류 결과 1건 / 13시간9분'. The main area displays a file named '2229.37_sqwenn_wasteland__woman_scream.wav' with a red box around it. To the left is a sidebar with categories like '전체' (All), '제주행(성범죄)', '감도있어', '정도있어', and '화자있어'. On the right, there's a sidebar with various classification terms such as '성우행을 당한 상황', '성우행을 하는 상황', '성악자를 하는 상황', etc. A red 'X' button is located next to the file name.

<분류 결과 화면>

- 페이지 상단의 '분류 결과' 탭을 클릭하면 회원 자신이 올린 파일 및 라벨링 등의 정보를 확인할 수 있습니다.
- 상단과 좌측의 '분류' 탭을 클릭하면 업로드 시 설정한 '대분류'와 '중분류'에 맞는 정보를 볼 수 있습니다. ('전체'를 누르면 전체 파일 정보가 나옵니다.)
- 우측의 'X'버튼을 클릭하면 파일이 삭제됩니다.
- 파일 명을 클릭하면 해당 파일의 분류작업 페이지로 이동합니다. (수정가능)



1.5 음향 상황 관리

1.5.1 음향 상황 목록 조회

D2AI 치안안전 강제추행(성범죄) 음향 상황 관리 16 건

■ 사용자 메뉴얼 | ☰ 로그아웃

◆ 본문 작업		◆ 보조 경과		◆ 공급 상황 관리	
치안안전	수령안전	사고발생	자연재해	일반위기	일반상황
강제추행(성범죄)					
강도범죄	대분류	중분류	상황		제한여부
절도범죄	치안안전	강제추행(성범죄)	성추행을 당한 상황		제한여부
폭력범죄	치안안전	강제추행(성범죄)	성추행을 당한 상황		제한여부
	치안안전	강제추행(성범죄)	언어적 성희롱을 당한 상황		제한여부
	치안안전	강제추행(성범죄)	구조·도움받는 상황		제한여부
	치안안전	강제추행(성범죄)	성추행을 하는 상황		제한여부
	치안안전	강제추행(성범죄)	성추행을 하는 상황		제한여부
	치안안전	강제추행(성범죄)	언어적 성희롱을 하는 상황		제한여부
	치안안전	강제추행(성범죄)	성희롱을 저지르는 상황		제한여부
	치안안전	강제추행(성범죄)	성추행을 저지르는 상황		제한여부
	치안안전	강제추행(성범죄)	스트레스로 도움을 요청하는 상황		제한여부
	치안안전	강제추행(성범죄)	스트레스로 도움이하는 상황		제한여부
	치안안전	강제추행(성범죄)	스토레일 당한 상황		제한여부
	치안안전	강제추행(성범죄)	스토레일 당하는 상황		제한여부
	치안안전	강제추행(성범죄)	디지털 성범죄를 당하는 상황		제한여부
	치안안전	강제추행(성범죄)	디지털 성범죄를 행하는 상황		제한여부
	치안안전	강제추행(성범죄)	성추행에서 저지르는 상황		제한여부

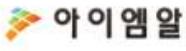
면역학과

- 상단의 ‘음향 상황 관리’탭을 클릭합니다.
 - 상단의 대분류, 좌측의 중분류에서 해당하는 항목을 선택하여 상황을 조회합니다.

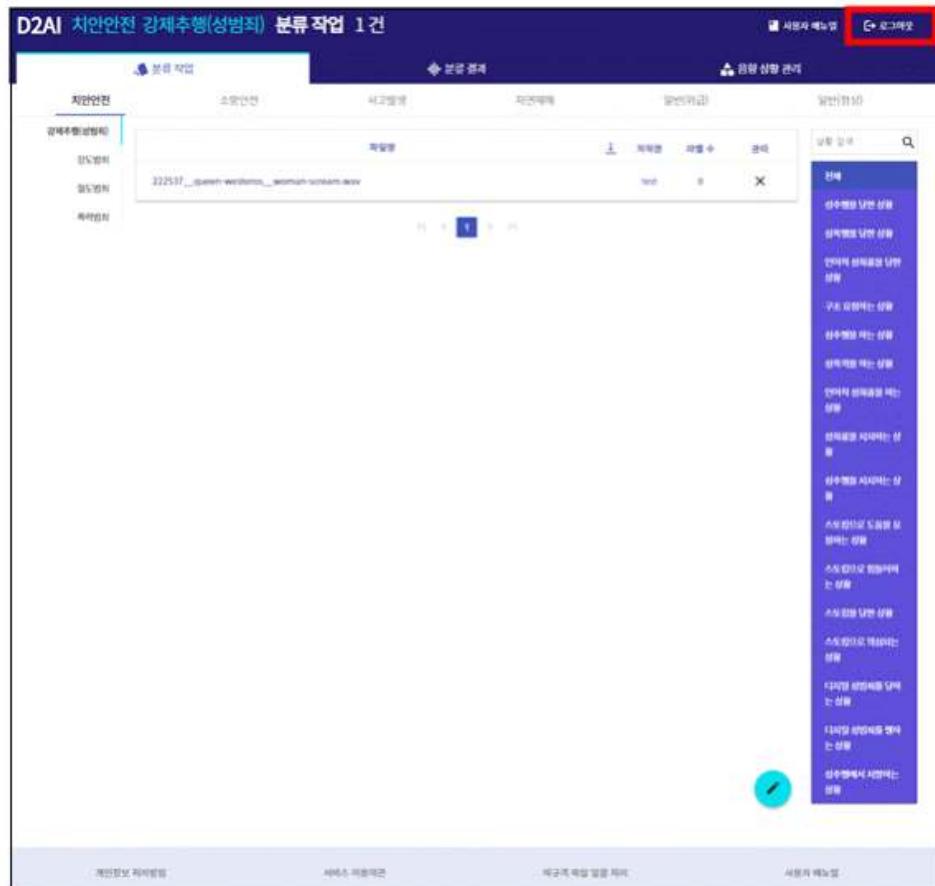


〈음향 상황 등록 화면〉

- 하단의 버튼을 눌러 '상황등록'창을 엽니다.
 - 대분류/중분류/상황을 선택 및 입력하여 하단의 확인 버튼을 누르면 음향 상황이 등록됩니다.
 - 음향 상황의 수정 및 삭제 또는 제한은 관리자에 의해서만 가능합니다.

	제목 : 데이터셋 분류 페이지 사용 메뉴얼 버전 1.3.0 페이지 19(21)
---	---

1.6 로그아웃



<로그아웃 화면>

- 페이지 최상단 우측의 버튼을 누르고 '확인'버튼을 누르면 로그아웃이 완료 됩니다.