

2025-2 AI 프로젝트Ⅳ HCI 시스템 설계

# 위험소리 감지 알림 UX 서비스 기획 (최종)

인공지능학과 22619027 지정원

# Contents

00 선행 연구와의 차별성

01 피드백 반영

02 데이터셋

03 전처리 및 실험 결과 - 10주차

04 전처리 및 실험 결과 - 14주차

05 최종 성능 비교

06 웹 서비스 구현

# \* 문제 인식

디지털타임스



사회 | 일반

## [SNS, 그후] 손에 스마트폰, 귀엔 이어폰 낀 채 `휘청휘청`... 거리의 `시한폭탄`

박상길 기자 [구독 +](#) 입력 2024-03-06 14:49 수정 2024-03-06 18:59

≡ **경기일보** 사회 | 사회일반

### ‘들고 싶은 것만 들어요’... 교통사고 부른 노이즈캔슬링

슬인 2024-09-30 09:00



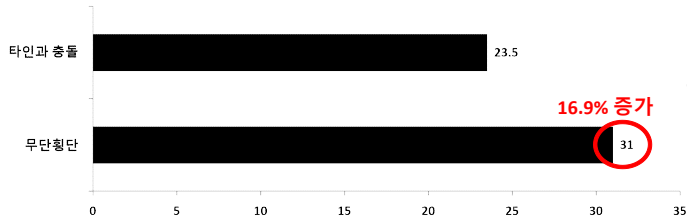
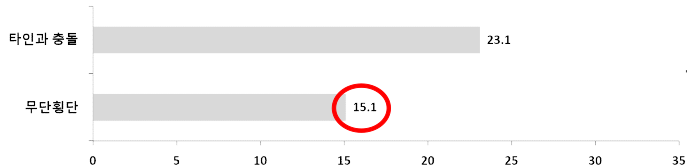
한준호 기자 [hjh1212@kyeongs.com](mailto:hjh1212@kyeongs.com)  
기자메이저 >

무선 이어폰 착용 사고 위험 ↑ ... 도내 보행 '교통사고' 주의보  
최근 3년간 연평균 8천600건... 시민들 '안전의식' 수반 중요

실제 도로교통공단의 교통사고분석시스템에 따르면 2020년부터 2022년까지 발생한 보행자 교통사고는 10만9877건으로 전체 교통사고의 18%를 차지했으며 이로 인한 사망자는 3044명에 달했다. 특히 고령 보행자가 차지하는 비율은 매년 증가했는데 고령 보행사망자의 전체 보행사망자 대비 비율은 2020년부터 2022년까지 2.3%p(포인트) 늘어났다.

이 같은 사망 사고의 가장 큰 원인은 '노이즈 캔슬링' 기능을 활성화한 무선 이어폰을 사용하기 때문이라는 지적이다. 도로교통공단에 따르면 노이즈 캔슬링 기능이 활성화한 이어폰을 끼고 다니는 개인형 이동장치 사고의 경우 2020년 897건에서 2022년 2386건으로 2.6배 급증했다.

## \* 문제 인식



출처 : 삼성교통안전문화연구소

# 00. 선행 연구와의 차별성

김지혁, 박상근. (2024-12-18). 소리 분류 모델을 이용한 골목길에서의 차량-보행자 충돌 위험 방지 시스템. 한국정보과학회 학술발표논문집, 전남.

- 선행 연구: 소리 분류 모델을 이용한 골목길에서의 차량-보행자 충돌 위험 방지 시스템 (2024.12)

## 3.3 충돌 위험 방지 앱 개발



그림 1. 앱 초기 화면(좌), 차량 인식 화면(우)

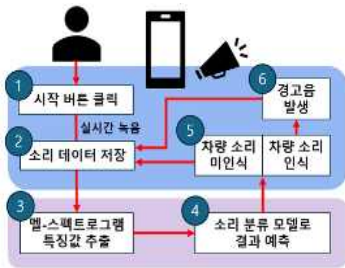


그림 2. 앱 동작 구조

표 1. LSTM, CNN 기반의 소리 분류 모델 성능 비교

	LSTM	CNN
Accuracy	93.3%	96.2%
Precision	93.5%	96.4%
Recall	93.3	96.2%
F1-score	93.3	96.2%

# 00. 선행 연구와의 차별성

조유겸, 황태욱, 배무성, 임태훈, 김정창. (2025-06-18). 청각장애인을 위한 위험 소리 감지 및 알림 시스템. 한국통신학회 학술대회논문집, 제주.

- 선행 연구: 청각장애인을 위한 위험 소리 감지 및 알림 시스템 (2025.06)

2025년도 한국통신학회 하계종합학술발표회

청각장애인을 위한 위험 소리 감지 및 알림 시스템

조유겸, 황태욱, 배무성, 임태훈, 김정창

국립한국해양대학교

bamppy12@g.kmou.ac.kr, xodnr1130@g.kmou.ac.kr, baemuseong@g.kmou.ac.kr,  
lth3097@g.kmou.ac.kr, jehkim@kmou.ac.kr

Hazardous Sound Detection and Alert System for Hearing Impaired People

Yu-gyeom Cho, Tae-uk Hwang, Museong Bae, Taehun Lim, Jeongchang Kim

Korea Maritime and Ocean University (KMOU)

요약

본 논문에서는 청각장애인이 위험 상황 소리를 감지하기 어려운 문제를 해결하기 위해 스마트폰의 내장 마이크로부터 소리를 수집하여 CNN 기반 모델로 실시간 분석 후, 스마트폰에서 진동/식각 알림으로 사용자에게 위험 상황을 즉각적으로 전달하는 모바일 애플리케이션을 개발하였다.

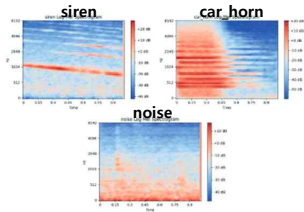


그림 1. 각 상황별 Log-Mel Spectrogram

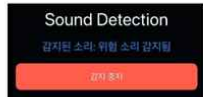
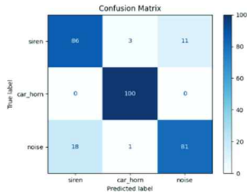


그림 2. 위험 소리 감지 시 애플리케이션 실행 화면.




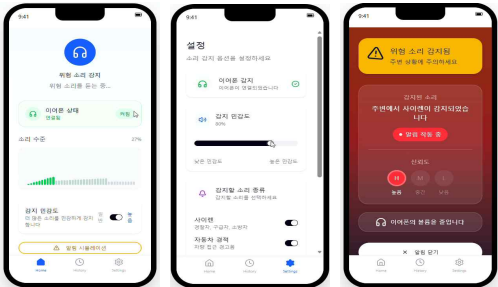
# 00. 선행 연구와의 차별성

## - 선행 연구와의 차이점

	선행 연구	진행 연구
① 학습 데이터	원본 위험음	원본 위험음 + <b>노이즈 합성</b> -> 실제 외부 환경 묘사 (도로 등..)

	선행 연구	진행 연구
② 경고	위험음 감지하면 이어폰에서 단순 “뽵” 소리	<ul style="list-style-type: none"><li>어떤 위험음인지 “종류”까지 구분</li><li>탐지 민감도 조절</li></ul>

# 00. 선행 연구와의 차별성

	선행 연구	진행 연구
③ 디자인	 <p>개발자 중심 설계 디자인 (DCD)</p>	 <p>사용자 중심 설계 디자인 (UCD)</p>



# 01. 피드백 반영

위험음 임계값 정해 **"감지 민감도"** 설정, 낮음/일반/높음으로 분류

ex)

감지 민감도 **낮음**일 때 위험음 탐지율(신뢰도) **40%** -> 경고음 **안올림**

= 큰 소리만 탐지 = 불필요한 경고 x



감지 민감도 **높음**일 때 위험음 탐지율(신뢰도) **40%** -> 경고음 **올림**

= 멀리서 나는 작은 소리까지 탐지

실내와 실외 구분이 어려울 것 같음

처리할 부분이 많아 보여서 힘들어 보임



	실외(도로) 환경	실내(도서관/카페) 환경
민감도 조절	 안전 중요 (작은 위험음까지 탐지)	 잘못 경고 불편

# 01. 피드백 반영

노이즈를 넣었을 때 정확도가 계속 떨어지는데, 이 성능을 높일 수 있는 방안을 고려하지 않은 것 같다.



청각장애인을 위한 위험  
소리 감지 및 알림 시스템

[선행연구 위험음 원본]

	precision	recall	f1-score
siren	0.82	0.86	0.84
car_horn	0.96	1.00	0.98
accuracy			<b>0.89</b>

[진행연구 위험음 원본]

	precision	recall	f1-score
siren	0.99	1.00	1.00
car_horn	1.00	0.96	0.98
accuracy			<b>0.99</b>

[위험음 원본 + 도로 소음]

	precision	recall	f1-score
siren	0.93	0.92	0.92
car_horn	0.90	0.90	0.90
accuracy			<b>0.92</b>

위험음 + 도로 소음 / 위험음 + 사람 말소리 / 위험음 + 공공데이터 사람 말소리 / 위험음 + 음악  
(0.92) (0.81) (0.77) (0.67)



- clear sound가 아니기 때문에 성능 떨어질 수밖에 없음.
- 실제 사용 환경 흉내 필요

# 01. 피드백 반영

현재는 대상이 시민들이지만, 만약 사회적 약자들과 같이 특  
| 특수한 사람들이라면 어떤식으로 진행할 예정인가요?



청각장애인을 위한 위험 소리 감지  
및 알림 시스템

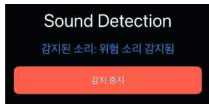


그림 2. 위험 소리 감지 시 애플리케이션 실행 화면.

중 siren 또는 car\_horn 클래스의 예측 확률이 0.6 이상일 경우 해당 소리를 위험으로 간주하며, noise 클래스가 감지된 경우에는 위험 소리로 판단하지 않는다. 또한, 동일한 위험 소리가 3 회 예측 주기 동안 연속으로 감지되었을 때, 푸시 알림이 발생하도록 하여 중복 알림을 방지한다.

실험 결과에서 도로소음의 성능이 제일 좋게 나왔다면 ( 나머지 나머지 노이즈들은 잘 구분하지 못한다는 결과 아닌가요?



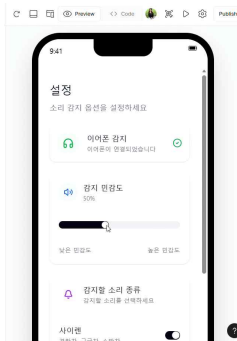
- 14주차 실험 : noise 전처리 (X)
- 이번 실험 : **time window 증가**  
**Mel-Spectrogram Smoothing**

=> 합성된 소음 유지 + 특징만 안정화  
(denoising X)

# 01. 피드백 반영



성능 개선 후 앱 개발까지 완료되면 더욱 좋을 것 같습니다  
앱도 만들면 좋을 것 같다

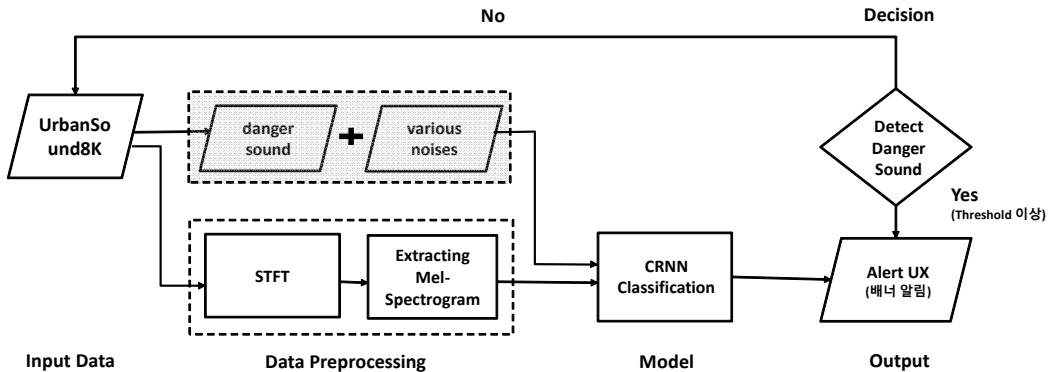


[Figma 초안 디자인]



[Figma 시안을 실제 웹 프로토타입으로 구현]

# \* FlowChart



## 02. 데이터셋

[원본 위험음]

classID:  
A numeric identifier of the

0 = air\_conditioner

1 = car\_horn

2 = children\_playing

3 = dog\_bark

4 = drilling

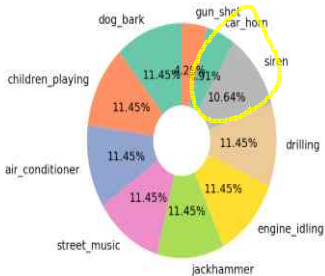
5 = engine\_idling

6 = gun\_shot

7 = jackhammer

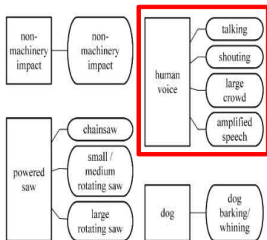
8 = siren

9 = street\_music



UrbanSound8K (car\_horn, siren)

[합성할 noise]



SONYC-UST-V2  
(only human voice)

음성 메모

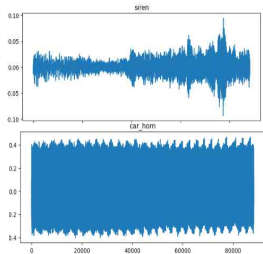
<input type="radio"/> 새로운 녹음 4	오전 12:05	00:05
<input type="radio"/> 새로운 녹음 5	이제	00:07
<input type="radio"/> 새로운 녹음 5	이제	00:07
<input type="radio"/> 새로운 녹음	이제	00:58
<input type="radio"/> 새로운 녹음 3	이제	00:06
<input type="radio"/> 새로운 녹음 2	이제	00:02
최근 삭제된 항목 6개의 파일		

직접 수집한 다양한 noise  
(도로 소음, 음악, 사람 말소리)

# 03. 전처리 및 실험 결과 - 10주차

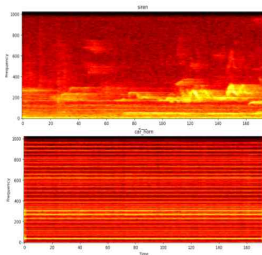
## - 원본 데이터 전처리

### 0. Waveform



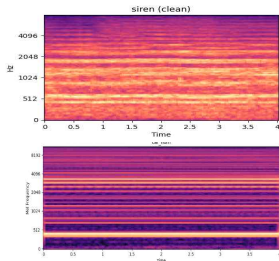
원본 음성 데이터 시각화

### 1. STFT (Short Time Fourier Transform)



서로 다른 음성 분리,  
그 시점마다의 주파수 성분을 파악

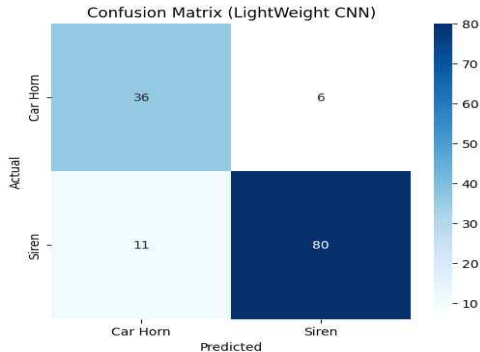
### 2. Mel-Spectrogram



사람이 민감한 주파수의 소리는 확대하고,  
민감하지 않은 주파수의 소리는 축소

### 03. 전처리 및 실험 결과 - 10주차

- 위험음 원본



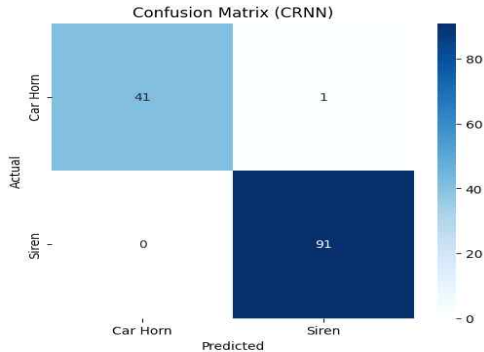
	precision	recall	f1-score
siren	0.92	0.88	0.90
car_horn	0.90	0.85	0.87
accuracy			0.89
macro avg	0.91	0.87	0.88
weighted avg	0.91	0.89	0.89

LightWeight CNN



# 03. 전처리 및 실험 결과 - 10주차

- 위험음 원본



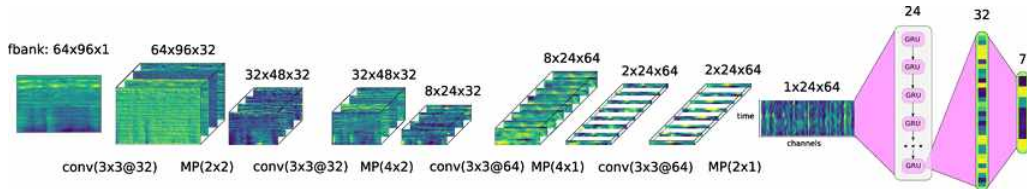
CRNN 모델 성능 더 좋으므로  
LightWeight CNN 사용 x

	precision	recall	f1-score
siren	0.99	1.00	1.00
car_horn	1.00	0.96	0.98
accuracy			0.99
macro avg	1.00	0.98	0.99
weighted avg	0.99	0.99	0.99

**CRNN**  
(Convolutional Recurrent Neural Network)

# \* CRNN

[ **CRNN** (Convolutional Recurrent Neural Network) ]



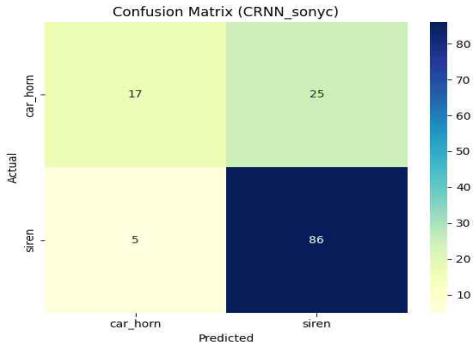
CNN + RNN -> 시각적 및 시간적 특징 동시 학습 신경망 구조

CNN : 시간-주파수 영역에서 중요한 **공간적 특징** 추출

RNN : 시퀀스 데이터의 **시간적 패턴** 학습

### 03. 전처리 및 실험 결과 - 10주차

- 위험음 + SONYC-UST-V2 (human voice)



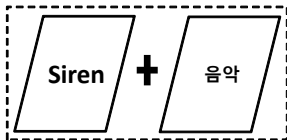
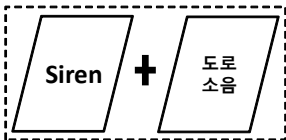
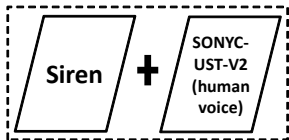
	precision	recall	f1-score
siren	0.77	0.95	0.85
car_horn	0.77	0.40	0.53
accuracy			0.77
macro avg	0.77	0.67	0.69
weighted avg	0.77	0.77	0.75

**CRNN**  
(Convolutional Recurrent Neural Network)

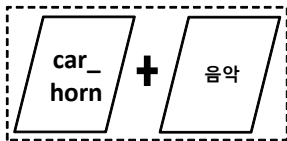
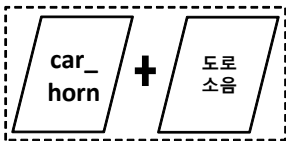
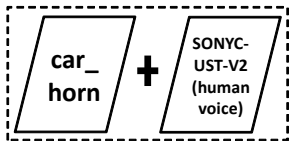
## 04. 전처리 및 실험 결과 - 14주차

- 원본 데이터 + 직접 수집한 노이즈  $\longrightarrow$  다양한 환경 대비

Siren

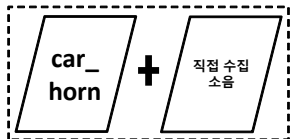
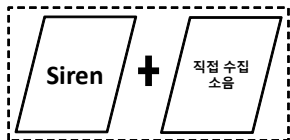


car\_horn



# 04. 전처리 및 실험 결과 - 14주차

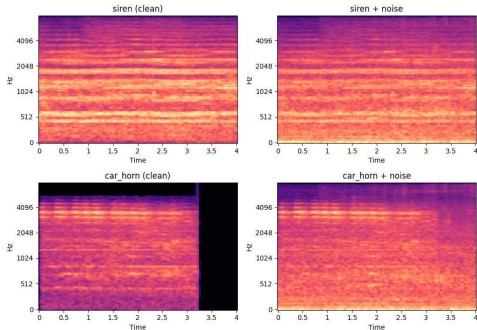
- (원본 데이터 + 노이즈) Mel-Spectrogram 추출



인간의 청각 시스템 특성 모방  
데이터 차원 축소



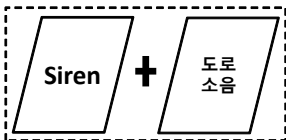
Mel-Spectrogram



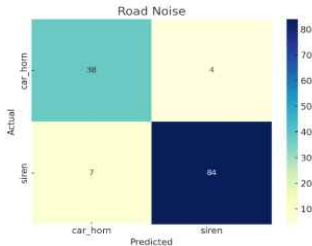
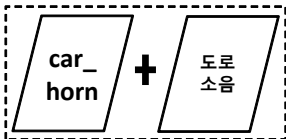
## 04. 전처리 및 실험 결과 - 14주차

- 위험음 + 도로 소음

siren + 도로 소음



car\_horn + 도로 소음



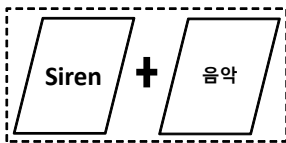
	precision	recall	f1-score
siren	0.93	0.92	0.92
car_horn	0.90	0.90	0.90
accuracy			0.92

Accuracy 0.92

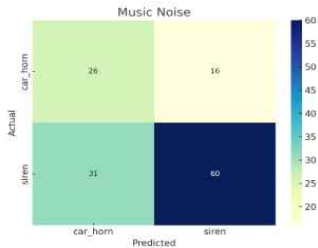
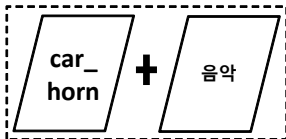
## 04. 전처리 및 실험 결과 - 14주차

- 위험음 + 음악

siren + 음악



car\_horn + 음악



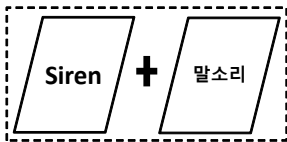
	precision	recall	f1-score
siren	0.72	0.66	0.69
car_horn	0.62	0.62	0.62
accuracy			0.67

Accuracy 0.67

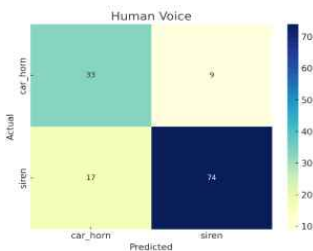
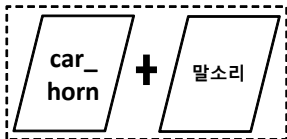
## 04. 전처리 및 실험 결과 - 14주차

- 위험음 + 사람 말소리

siren + 사람 말소리



car\_horn + 사람 말소리



	precision	recall	f1-score
siren	0.84	0.81	0.82
car_horn	0.77	0.79	0.78
accuracy			0.81

Accuracy 0.81



## 04. 실험 결과 - 14주차

### - 성능 비교

	precision	recall	f1-score
siren	0.93	0.92	0.92
car_horn	0.90	0.90	0.90
accuracy			<b>0.92</b>

	precision	recall	f1-score
siren	0.84	0.81	0.82
car_horn	0.77	0.79	0.78
accuracy			<b>0.81</b>

	precision	recall	f1-score
siren	0.77	0.95	0.85
car_horn	0.77	0.40	0.53
accuracy			<b>0.77</b>

	precision	recall	f1-score
siren	0.72	0.66	0.69
car_horn	0.62	0.62	0.62
accuracy			<b>0.67</b>

위험음 + 도로 소음  
(0.92)

>

위험음 + 사람 말소리  
(0.81)

>

위험음 + 공공데이터  
사람 말소리  
(0.77)

>

위험음 + 음악  
(0.67)

성능 개선 필요

# 05. 최종 성능 비교

- (위험음 + 노이즈) time window ↑ + smoothing

- 위험음 : 고주파 / 도로 소음 : 저주파  
=> 패턴이 겹치지 않아 분류 쉬움
- 도로 소음 변동성 ↓ => 학습 안정성 ↑
- 실제 사용 환경과 가장 근접

• 14주차 실험 : noise 전처리 (X)

• 이번 실험 : time window 증가  
Mel-Spectrogram Smoothing  
(Gaussian Filter)

=> 합성된 소음 유지 + 특징만 안정화  
(denoising X)

위험음 + 도로 소음	0.92	→	0.94
-------------	------	---	------

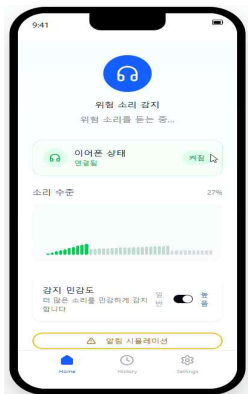
위험음 + 사람 말소리	0.81	→	0.86
--------------	------	---	------

위험음 + 공공데이터 사람 말소리	0.77	→	0.83
-----------------------	------	---	------

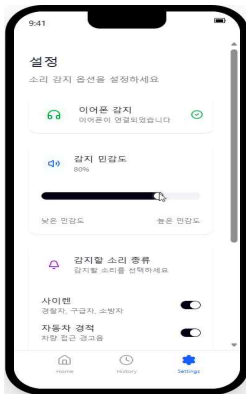
위험음 + 음악	0.67	→	0.78
----------	------	---	------

## 06. 웹 서비스 구현 - Figma 시안

이어폰 착용 유무 / 사운드 감지 민감도 설정 화면



감지 민감도 상세 설정 / 감지 사운드 종류 설정 화면



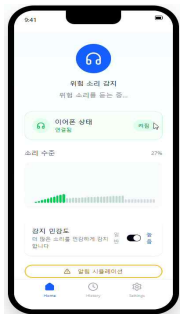
위험 소리 감지 후 경고 화면



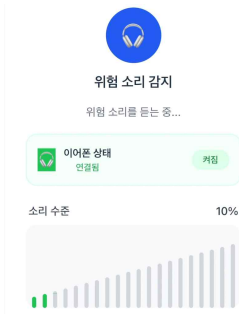
# 06. 웹 서비스 구현 - 실제 웹페이지



이어폰 착용 유무 / 사운드 감지 민감도 설정 화면



이어폰 상태 꺼짐 전환 시 소리 수준 0% 유지



감지 민감도 높음으로 설정 시 위험 경고 화면으로 전환



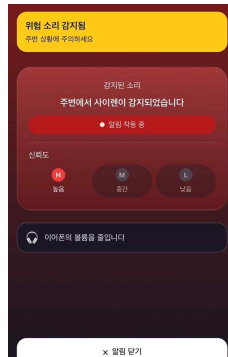
## 06. 웹 서비스 구현 - 실제 웹페이지



위험 소리 감지 후 경고 화면



신뢰도 높음/중간/낮음 화면 전환



**감사합니다.**