# Alarm Sound Classification System in Smartphones for the Deaf and Hard-of-Hearing Using Deep Neural Networks

Yuhki Shirarishi[1], Takuma Takeda[1], Akihisa Shitara[2]

[1]Tsukuba University of Technology, Japan, [2]University of Tsukuba, Japan

E-mail: yuhkis@a.tsukuba-tech.ac.jp

# A Short Resume of the Presenter

Yuhki Shiraishi, Ph.D.

2018.4 – present:
Associate Professor
Faculty of Industrial Technology
Tsukuba University of Technology, Japan

I teach for students who are deaf or hard of hearing (DHH).

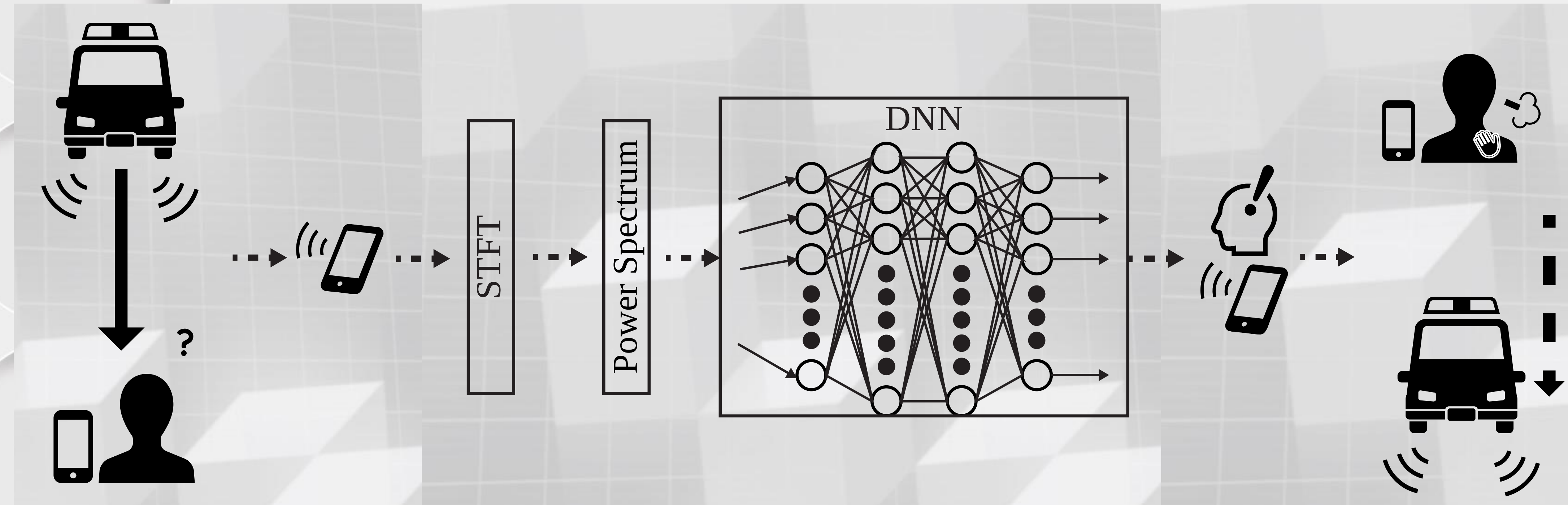
Research Interests:
Intelligent Signal Processing, Machine Learning, Information Support System, etc.

# Outline

1. Background/Objective
2. Related Work
3. Development System
4. Calcification Algorithm
5. Experiments
6. Discussion
7. Conclusion
8. Acknowledgment

# Background/Objective

- Over 5% of the world's population (466 million people) has disabling hearing loss (40dB/30dB)[1]




- Develop an alarm sound classification system in smartphone using deep neural network (DNN)
- As a result, DHH can go out safely using this system

1) https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

# Related Work

## ➢Ontenna

- an interface focusing on vibration
- let the user know sound by vibration in real-time
- no sound recognition system

https://ontenna.jp

## ➢Google Live Transcribe

- mainly for voice recognition
- recognize environmental sounds
- number of supported sound is limited

## ➢SeeSound

- send to the user via vibration and pop-up notifications
- works only in the home

https://www.see-sound.com

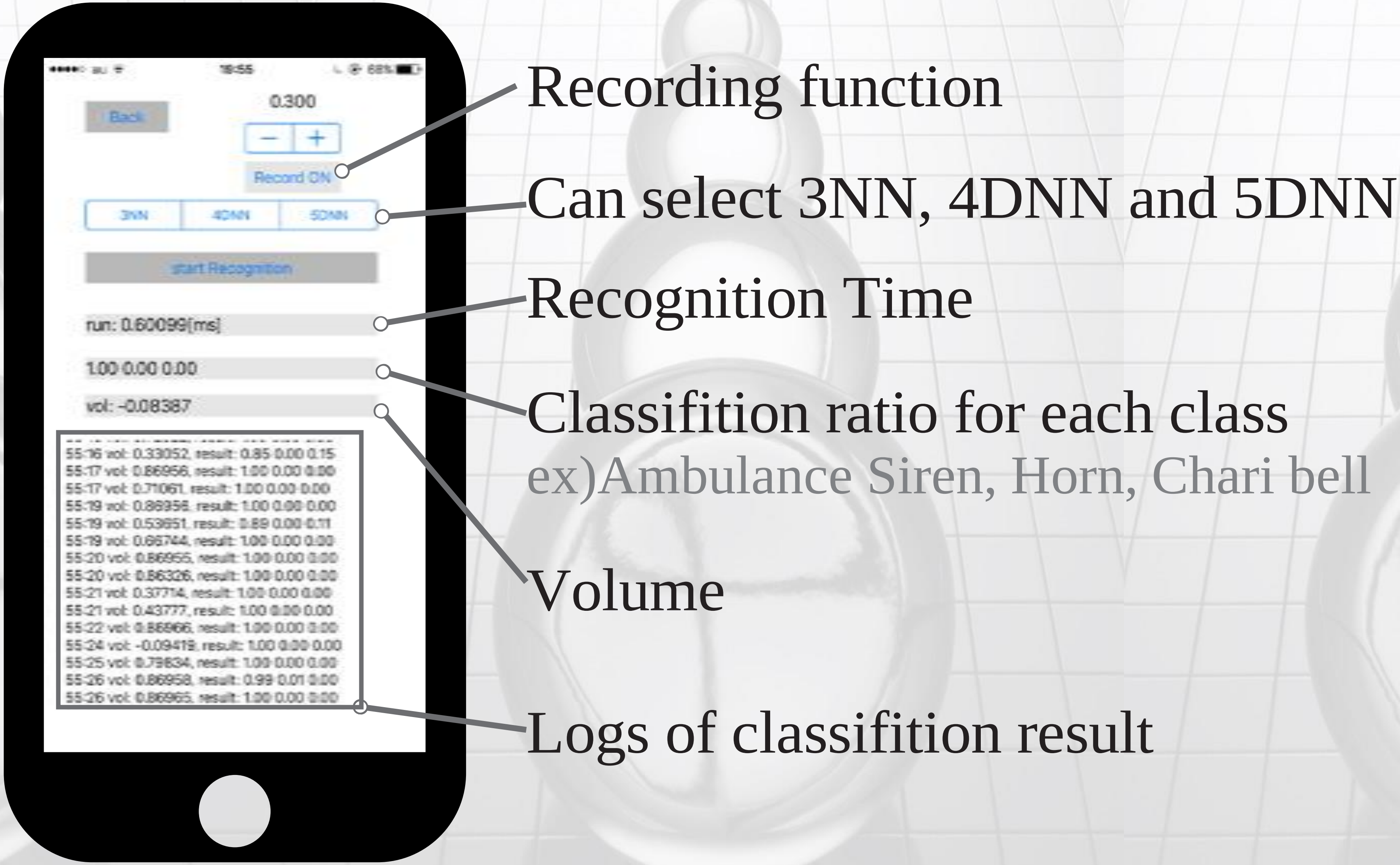
# Development System

- ➢The classification and transmission application run on a smartphone without connecting the internet.

- ➢The basic flow of the proposed system
  1. Collect environmental sounds with a smartphone
  2. Notify smartphone when an alarm sound is identified

- ➢Deep Learning (DL) is used for classification
  - Keras was used for implementing DL
- ➢The system works in iPhone now

# Development System




Snapshot of develop smartphone application

# Calcification Algorithm

- The alarm classifying flow
  1. Continuous collection of environmental sounds
  2. If volume data exceeding the threshold is detected, record audio data for a certain period.
  3. Specify the alarm class (horns, bicycle bells, ambulance sirens, etc.) of the recorded audio data.

- Because the nature of the alarm sound tends to be monotonous, we apply the short-time Fourier transform (STFT)

$$STFT(t,\omega) = \int_{-\infty}^{\infty} x(\tau)h(\tau - t)e^{-j\omega\tau}d\tau,$$

# Calcification Algorithm

- Operation of the classification application
  1. Use a smartphone microphone and collect sound every 1024 [frames] using 32-bit single-precision floating-point numbers (-1.0 to 1.0).
  2. When the absolute value of the buffered single precision floating-point buffer exceeds the threshold value (0.3), identification processing starts.
  3. Multiply the buffer by $2^{31}$ and change the buffer range to a 32-bit integer type, then execute STFT.
  4. Input of logarithmic power spectrum to DNN.
  5. Display the classification result on the screen.

# Calcification Algorithm

- In a real environment, the target sound would continue to sound.
- Therefore, the final classification result is determined by the following algorithm (called integrated judgment process)

1. Evaluate sounds continuous (1 to 10 times).
2. If there is more than one classification result from a specific sound other than noise,
   A) Calculate the sum of outputs.
   B) The largest of the noise exclusions is used as the final classification result.
3. If all classification results are noise,
   A) Regard the final classification result as noise.

# Experiments

## A. Basic performance of the classification system

- Targets: 5 classes
  - horn, bicycle bell, ambulance siren, fire alarm, noises (footsteps, car driving, voices, door opening/closing, hitting desks, and rubbing plastic bags)
- Evaluation: 5-fold CV
  - 25,000 pieces of training and evaluation data (5,000 pieces × 5 classes) with a maximum of 1,000 epochs (input layer: 513, hidden layer: 128, output layer: 5).

| *Number of layers* | *Classification ratio* |
|---|---|
| 3 | 0.9845 |
| 4 | 0.9867 |
| 5 | 0.9924 |

# Experiments

## B. Performance in a noisy environment

- Targets: 5 classes
  - horn, bicycle bell, ambulance siren, fire alarm, noises
- Evaluation:
  - Data: noisy environment of 50.5 to 100.3 [dB].
  - Method: simple judgement

| | *TP* | *FP* | *FN* | *TN* | *Prec.* | *Recall* | *F-value* | *Max vol[dB]* |
|---|---|---|---|---|---|---|---|---|
| Horn | 545 | 0 | 87 | 2232 | 1.00 | 0.86 | 0.92 | 98.1 |
| Bicycle bell | 502 | 0 | 113 | 2249 | 1.00 | 0.81 | 0.98 | 127.7 |
| Ambulance | 572 | 1 | 56 | 2336 | 0.99 | 0.91 | 0.95 | 90.0 |
| Fire alarm | 631 | 1 | 57 | 2176 | 0.99 | 0.91 | 0.95 | 93.2 |
| Noise | 298 | 262 | 2 | 2563 | 0.53 | 0.99 | 0.69 | 100.3 |

# Experiments

## B. Performance in a noisy environment

- Targets: 5 classes
  - horn, bicycle bell, ambulance siren, fire alarm, noises
- Evaluation:
  - Data: noisy environment of 50.5 to 100.3 [dB].
  - Method: integrated judgment process

| | *TP* | *FP* | *FN* | *TN* | *Prec.* | *Recall* | *F-value* | *Max vol[dB]* |
|---|---|---|---|---|---|---|---|---|
| Horn | 100 | 0 | 0 | 400 | 1.00 | 1.00 | 1.00 | 98.1 |
| Bicycle bell | 100 | 0 | 0 | 400 | 1.00 | 1.00 | 1.00 | 127.7 |
| Ambulance | 100 | 0 | 1 | 400 | 1.00 | 0.99 | 0.99 | 90.0 |
| Fire alarm | 100 | 0 | 1 | 400 | 1.00 | 0.99 | 0.99 | 93.2 |
| Noise | 100 | 2 | 0 | 398 | 0.99 | 1.00 | 0.99 | 100.3 |

# Experiments

## C. Performance for unlearned horn sounds

- Targets: 5 classes
  - horn, bicycle bell, ambulance siren, fire alarm, noises
- Evaluation:
  - Data: new type of horn sound (20 times * 7 types) different from the learning data in a noisy environment.
  - Method: simple judgement

| | *TP* | *FN* | *Classification rate* |
|---|---|---|---|
| Horn 1 | 62 | 16 | 0.79 |
| Horn 2 | 43 | 11 | 0.80 |
| Horn 3 | 42 | 8 | 0.84 |
| Horn 4 | 55 | 23 | 0.71 |
| Horn 5 | 67 | 8 | 0.89 |
| Horn 6 | 36 | 29 | 0.55 |
| Horn 7 | 50 | 33 | 0.60 |

# Experiments

## C. Performance for unlearned horn sounds

- Targets: 5 classes
  - horn, bicycle bell, ambulance siren, fire alarm, noises
- Evaluation:
  - Data: new type of horn sound (20 times * 7 types) different from the learning data in a noisy environment.
  - Method: integrated judgment process

| | *TP* | *FN* | *Classification rate* |
|---|---|---|---|
| Horn 1 | 20 | 0 | 1.00 |
| Horn 2 | 20 | 0 | 1.00 |
| Horn 3 | 20 | 0 | 1.00 |
| Horn 4 | 20 | 0 | 1.00 |
| Horn 5 | 20 | 0 | 1.00 |
| Horn 6 | 20 | 1 | 0.95 |
| Horn 7 | 20 | 1 | 0.95 |

# Experiments

## D. Adding new type of data from the web

- Targets: 5 classes
  - horn, bicycle bell, ambulance siren, fire alarm, noises
- Evaluation:
  - Data: new 428 car horn sounds, 169 bicycle bell sounds, and 929 ambulance siren sounds
  - Method: 5-fold CV (simple judgement)

| *Number of layers* | *classification Ratio* |
|---|---|
| 3 | 0.9367 |
| 4 | 0.9498 |
| 5 | 0.9714 |
| 6 | 0.9710 |

# Discussion

- Data collection
  - lacks of labeled plenty clean data
  - the crowdsourcing is one example of a solution
- Recognition start timing
  - the fast response time is vital because of the dangerous situation
- Direction of the sound source
  - DHH peoples are generally hard to notice the direction
  - This problem would be resolved by using a microphone array and direction estimate algorithms

# Conclusion

- We have proposed and developed an alarm sound classification system using DNN by smartphones.
- Evaluation experiments were performed to verify the effectiveness of the system.
- We also discuss the limitation of the developed system and the expectation of the improved system by overcoming these limitations.

# Acknowledgement


- The authors would like to thank Nobuyoshi Hata and Kazuki Yano, who have partially worked on the project.
- This work was partially supported by JSPS KAKENHI Grant Numbers #16K16460, #19K11411, and Promotional Projects for Advanced Education and Research in NTUT.
- One of the authors, Takuma Takeda, is now working at NEC Fielding, Ltd., Japan.