深圳大学 SHENZHEN UNIVERSITY

深圳大学物联网研究中心 IOT Research Center of Shenzhen University

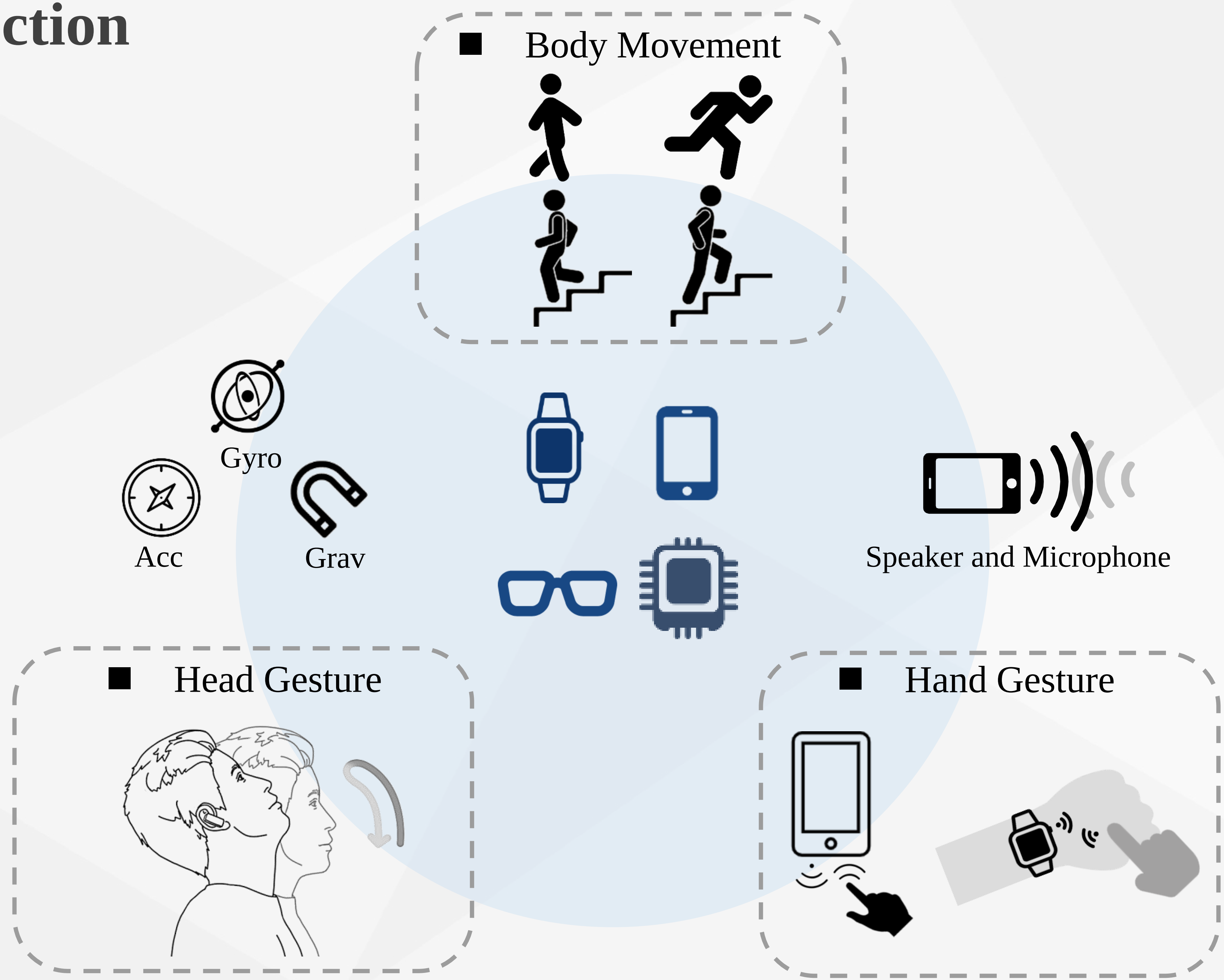
广东省无线大数据与未来网络工程技术研究中心 Wireless Future DataNet Research Centre

CHAR: Composite Head-body Activities Recognition with A Single Earable Device

Peizhao Zhu, Yongpan Zou, Wenyuan Li, Kaishun Wu

College of Computer Science and Software engineering
Shenzhen University

Introduction



Introduction

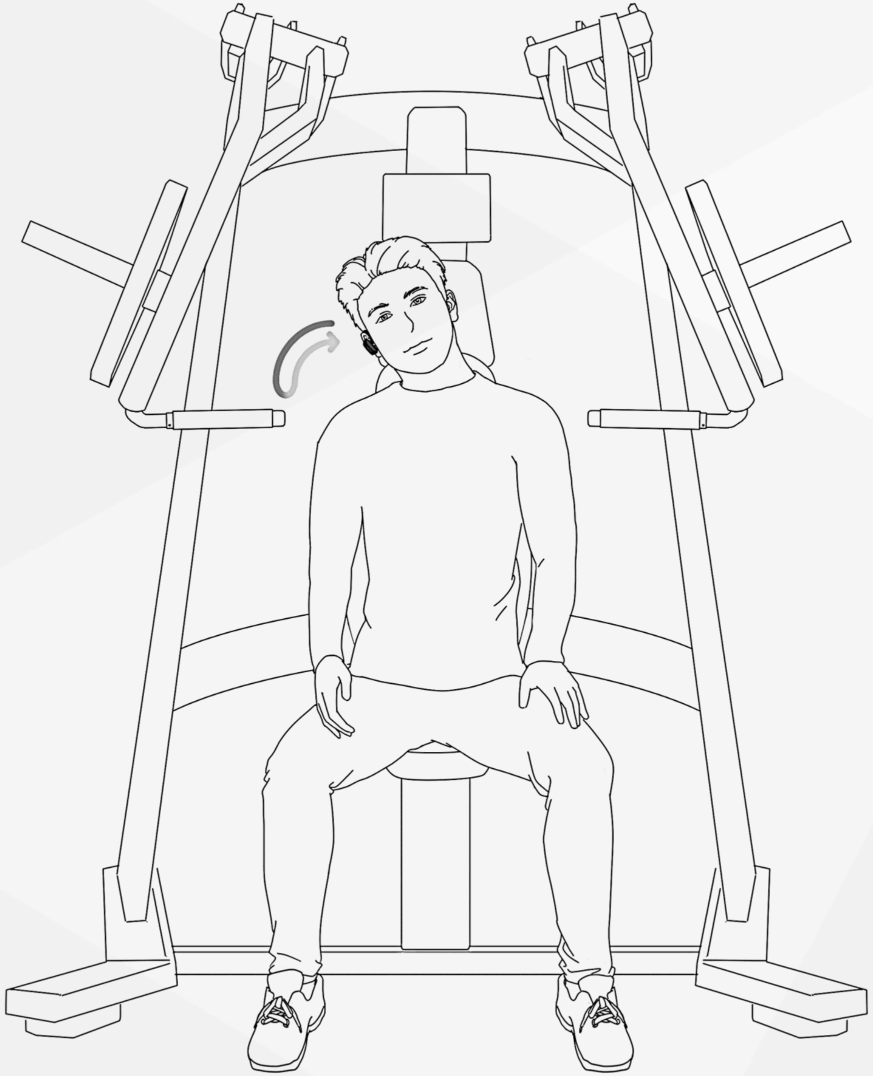


Fig 1. Doing exercises while staying still with hands occupied.

Fig 2. Going upstairs while carrying goods.

Introduction

Motivation

- Composite head-body activities have significant and practical value, since they have significant semantic information for human-computer interaction in real world.
- The commonalities and differences between different tasks are beneficial for boosting the recognition and generalization performance.

Contribution

- We consider a novel HAR problem in which composite head-body activities are recognized.
- We propose an adaptive segmentation method which dynamically adjusts according to different scenarios.
- We design a multi-task learning network to recognize head gestures and body movements simultaneously.
- Extensive experiments have shown that CHAR can recognize 60 composite activities with high accuracy.

System Design

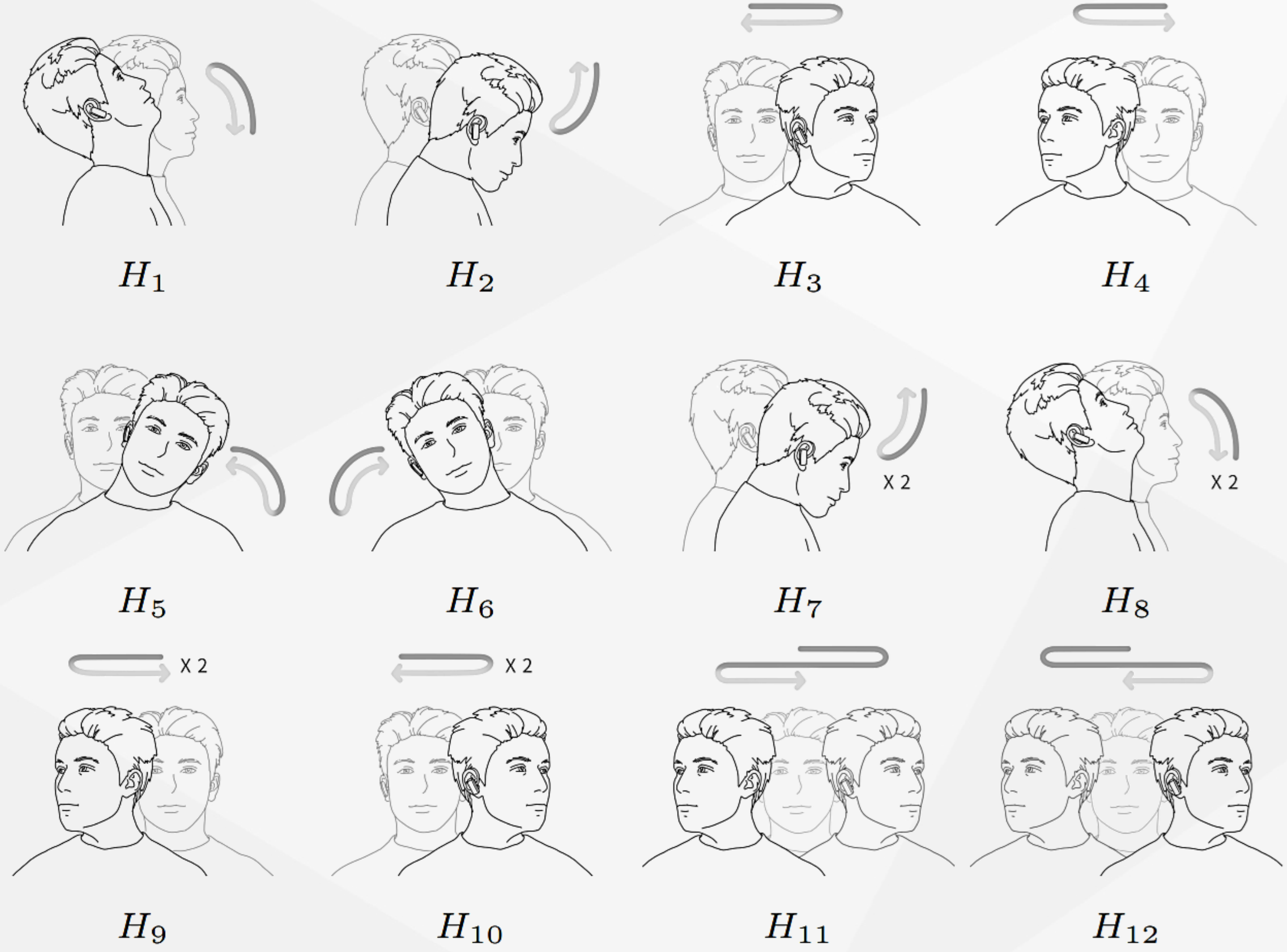


Fig 3. The design of head gestures.

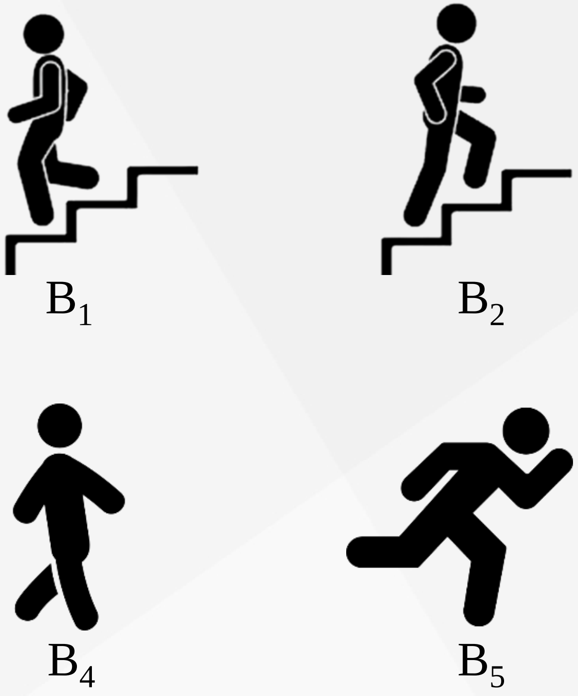


Fig 4. The design of body movements:

- Going downstairs
- Going upstairs
- Staying still
- Life-walking
- Jogging

System Design

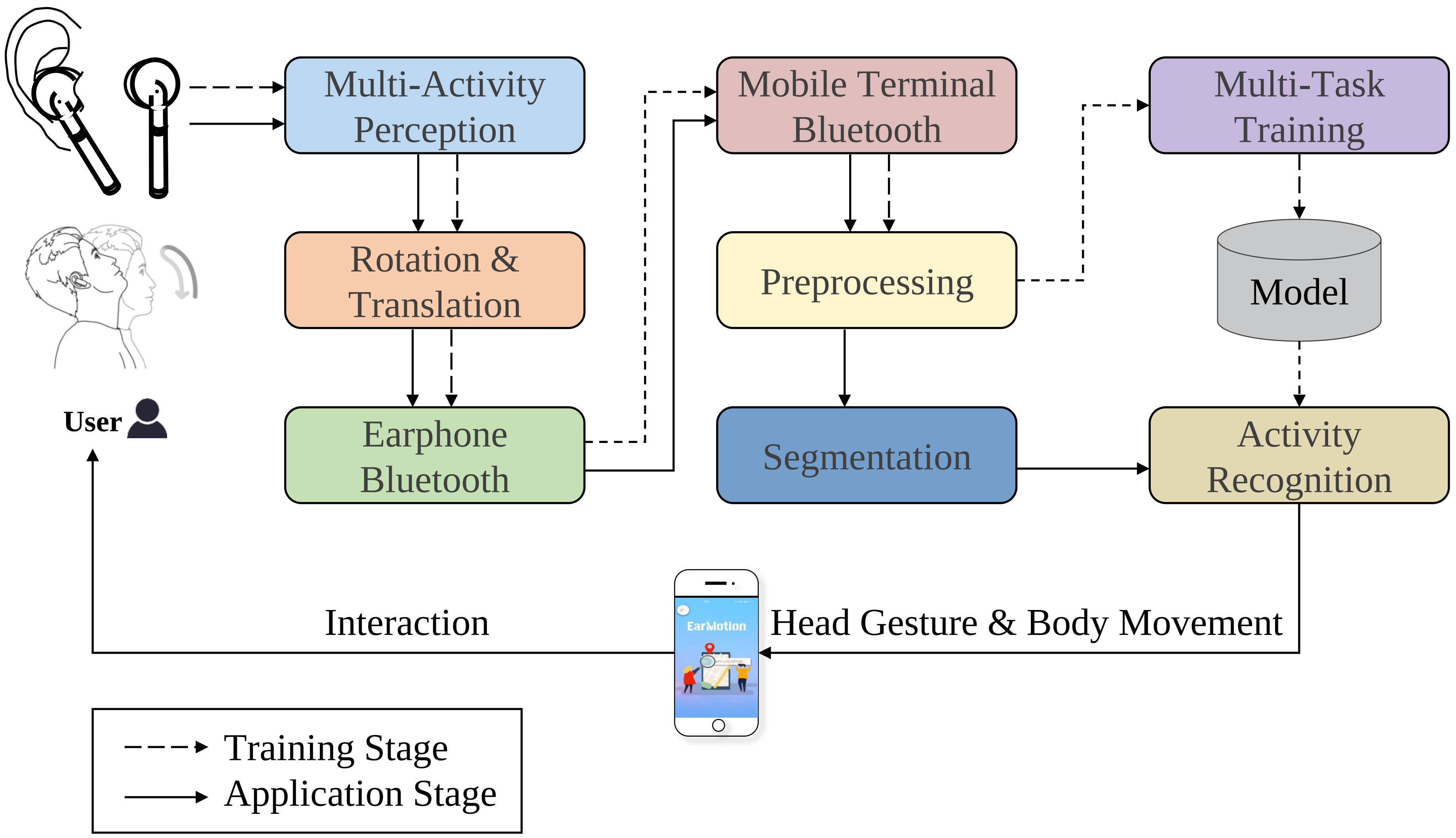


Fig 5. The system overview of CHAR.

System Design

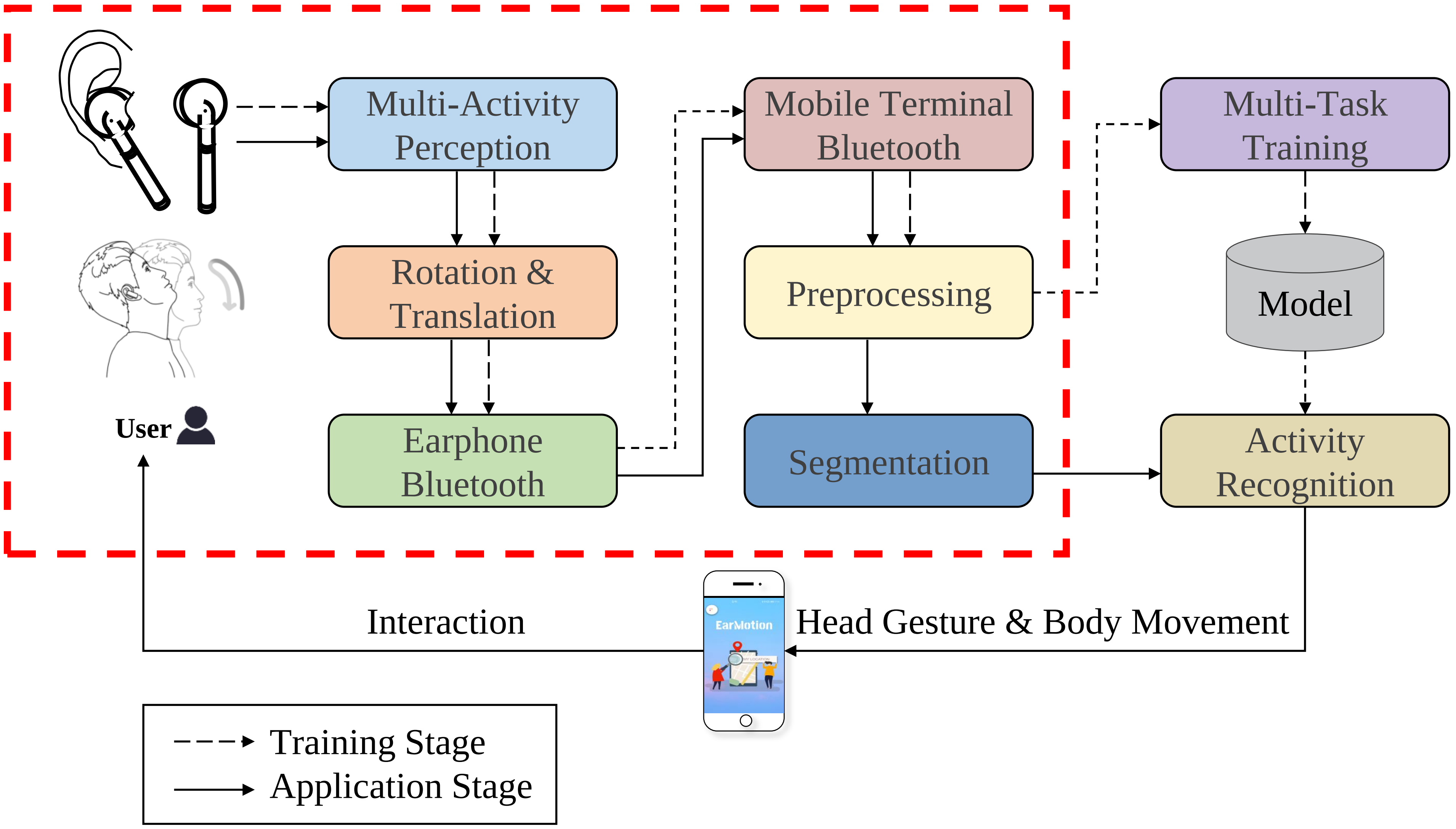


Fig 5. The system overview of CHAR.

System Design

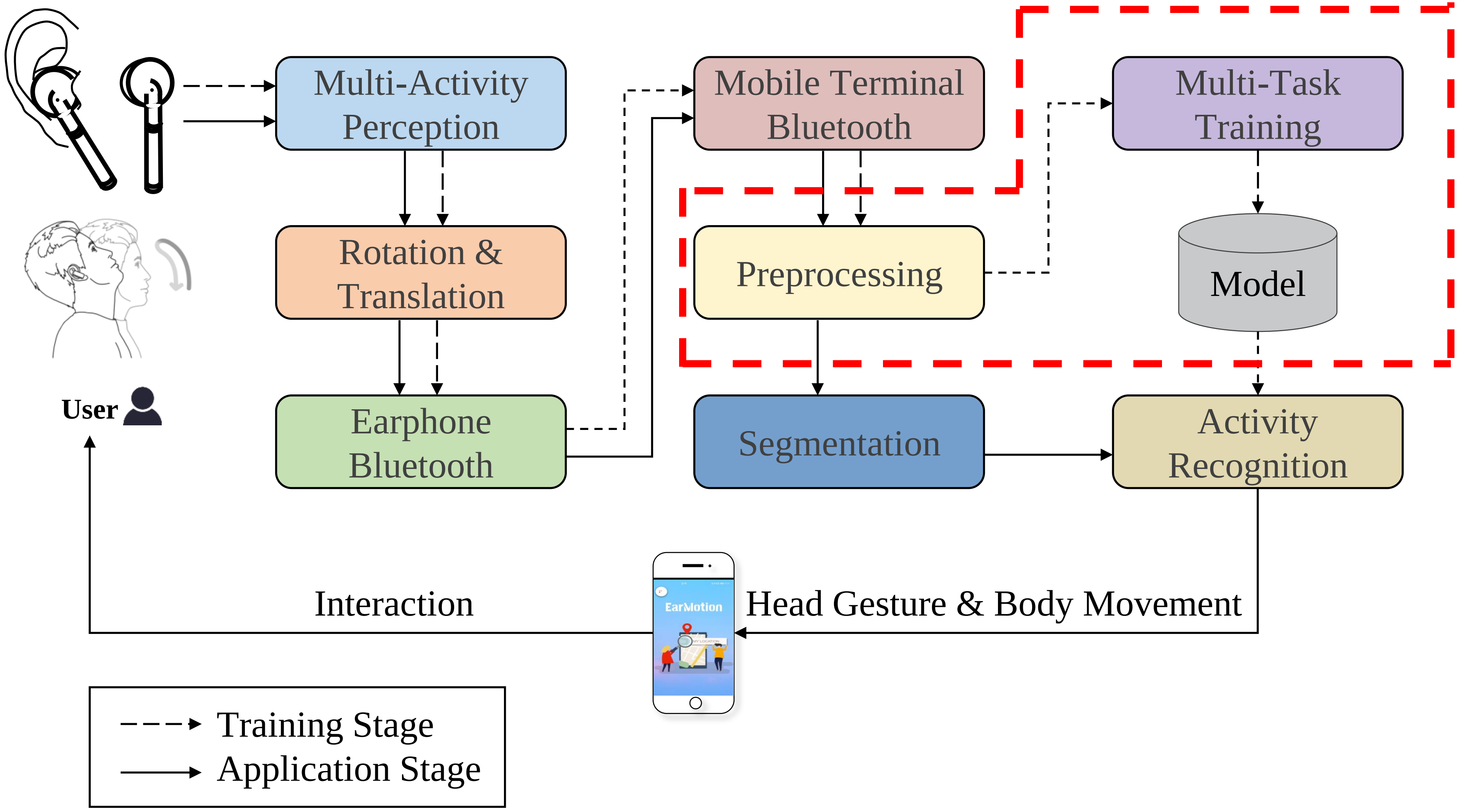


Fig 5. The system overview of CHAR.

System Design

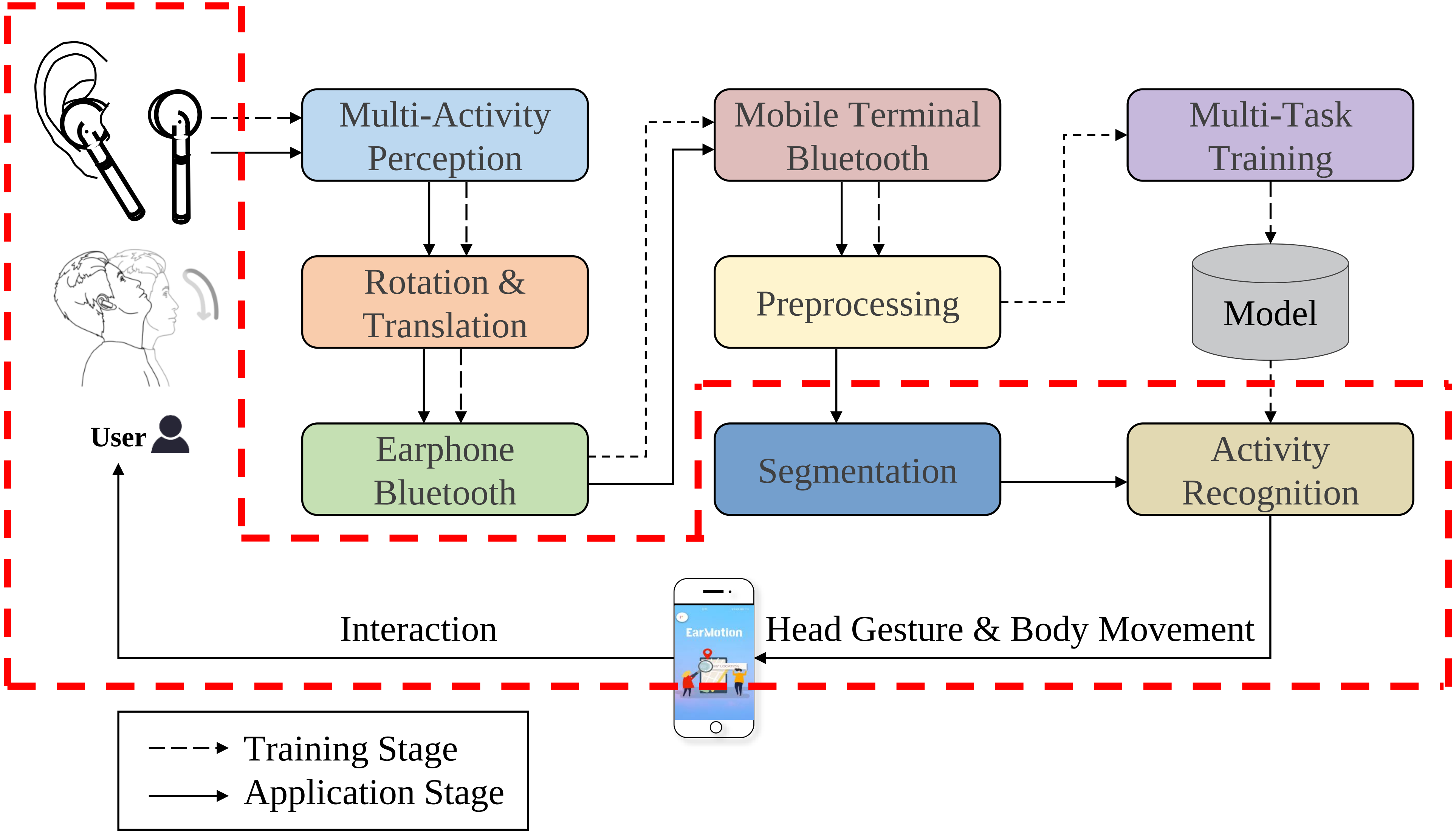
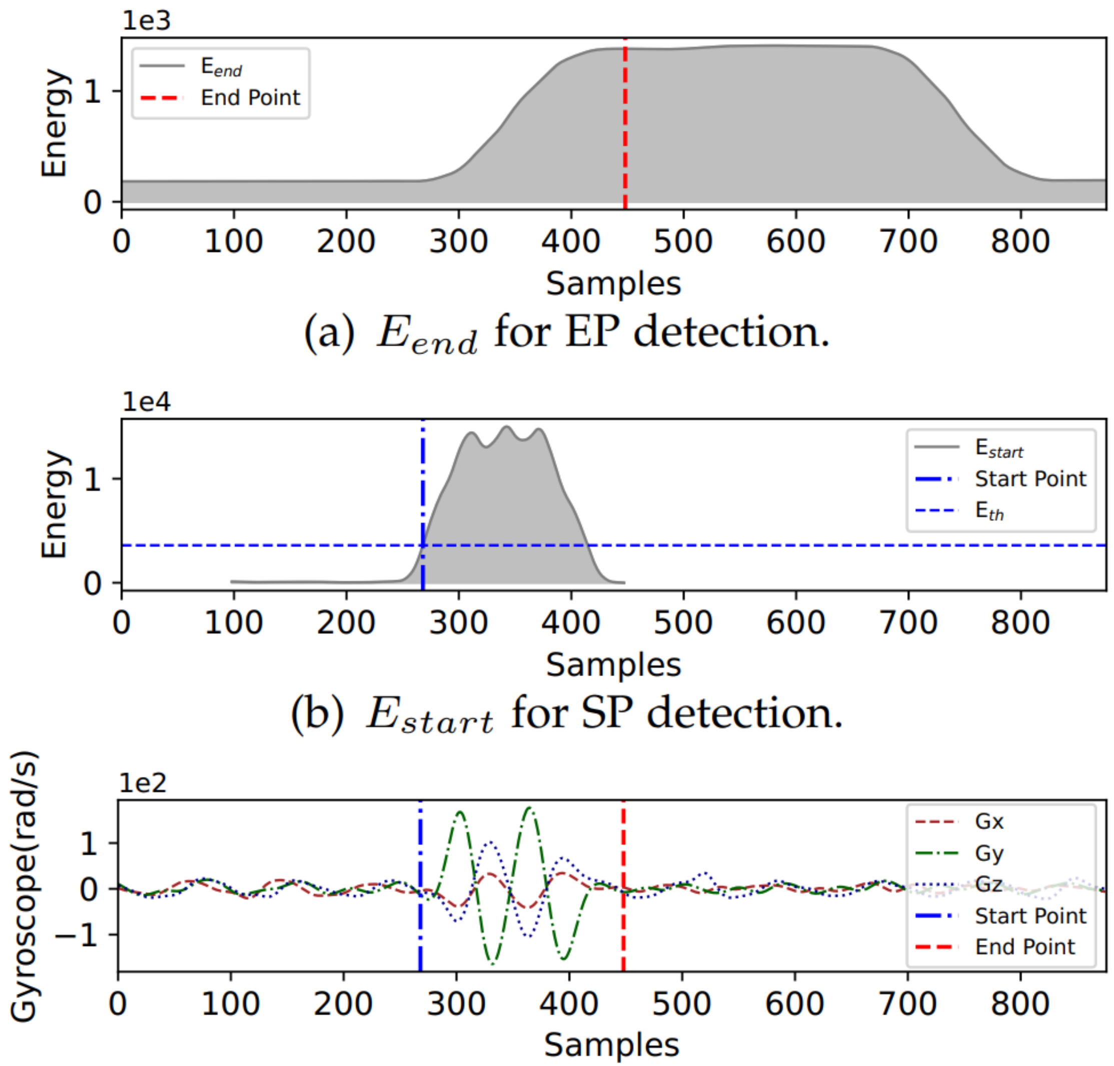


Fig 5. The system overview of CHAR.

System Design



(a) E_{end} for EP detection.

(b) E_{start} for SP detection.

(c) Origin signal and segmentation result.

Fig 6. Example of activity segmentation.

Algorithm 1: Activity segmentation algorithm

Input: 3-axis signal $Gyro = \{G_x, G_y, G_z\}$; Window length for EP and SP detection L_{end}, L_{start}; Framing stride S; Threshold for EP detection θ; Scaling coefficient for SP detection k

Output: SP; EP

W_{signal} = Frame($Gyro$, ($L_{end} + L_{start}$), S);
W_{end} = $W_{signal}(L_{start}$:, :, :);
W_{cont} = W_{signal}(: L_{start}, :, :);
for *i = 1 to FrameNum(W_{end})* **do**
    for *j = 1 to ChannelNum(W_{end})* **do**
        $E_{ch}(i, j)$ = Energy($W_{end}(i, j, :)$)
    end
    $ch(i)$ = FindMaxIndex($E_{ch}(i, :)$);
    $E_{end}(i)$ = Energy($W_{end}(i, ch(i), :)$);
    E_d = Diff(E_{end});
    EP = FindFirstIndexLess($E_d(i)$, θ);
    if *EP is not None* **then**
        E_{cont} = Energy($W_{cont}(i, ch(i), :)$);
        W_{start} = Frame($W_{end}(i, ch(i), :)$, L_{start}, S);
        for *j = 1 to FrameNum(W_{start})* **do**
            $E_{start}(j)$ = Energy($W_{start}(j, :)$);
        end
        SP = FindFirstIndexGreater(E_{start}, $k \times E_{cont}$);
    end
end
return SP, EP

System Design

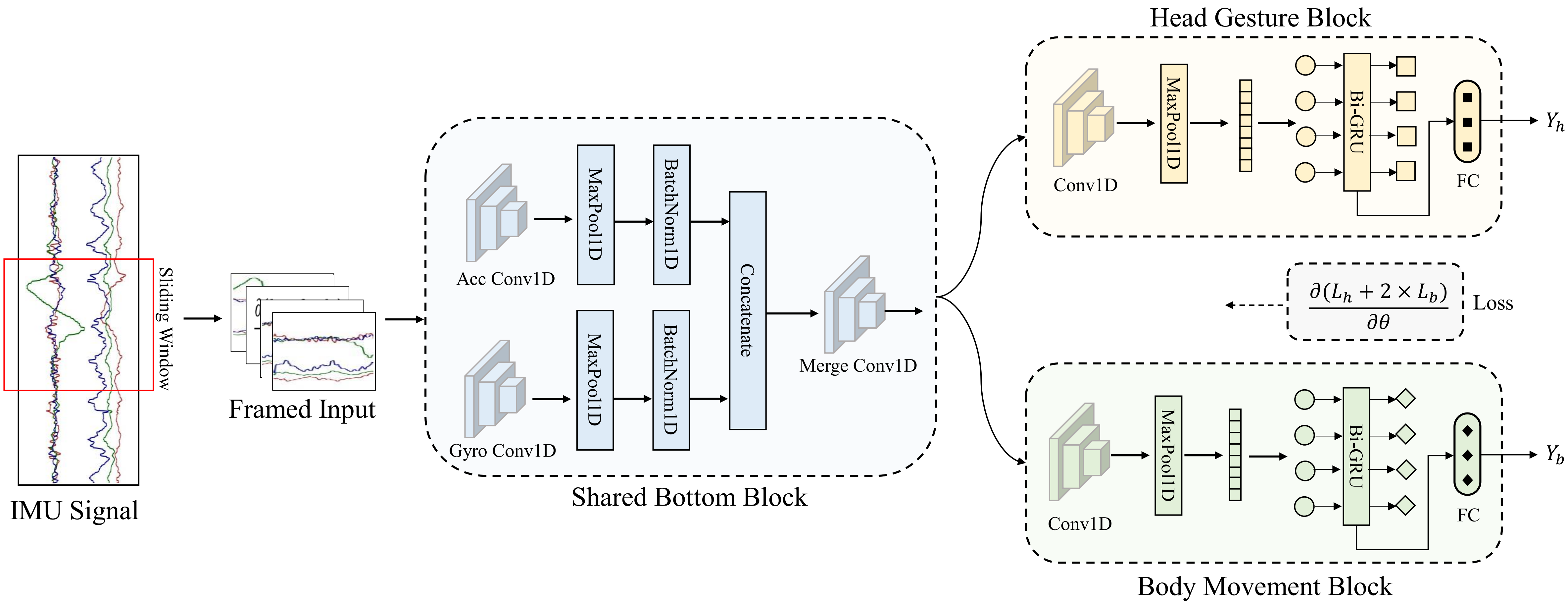


Fig 7. The architecture of our composite-activity recognition network (CARN).

System Design

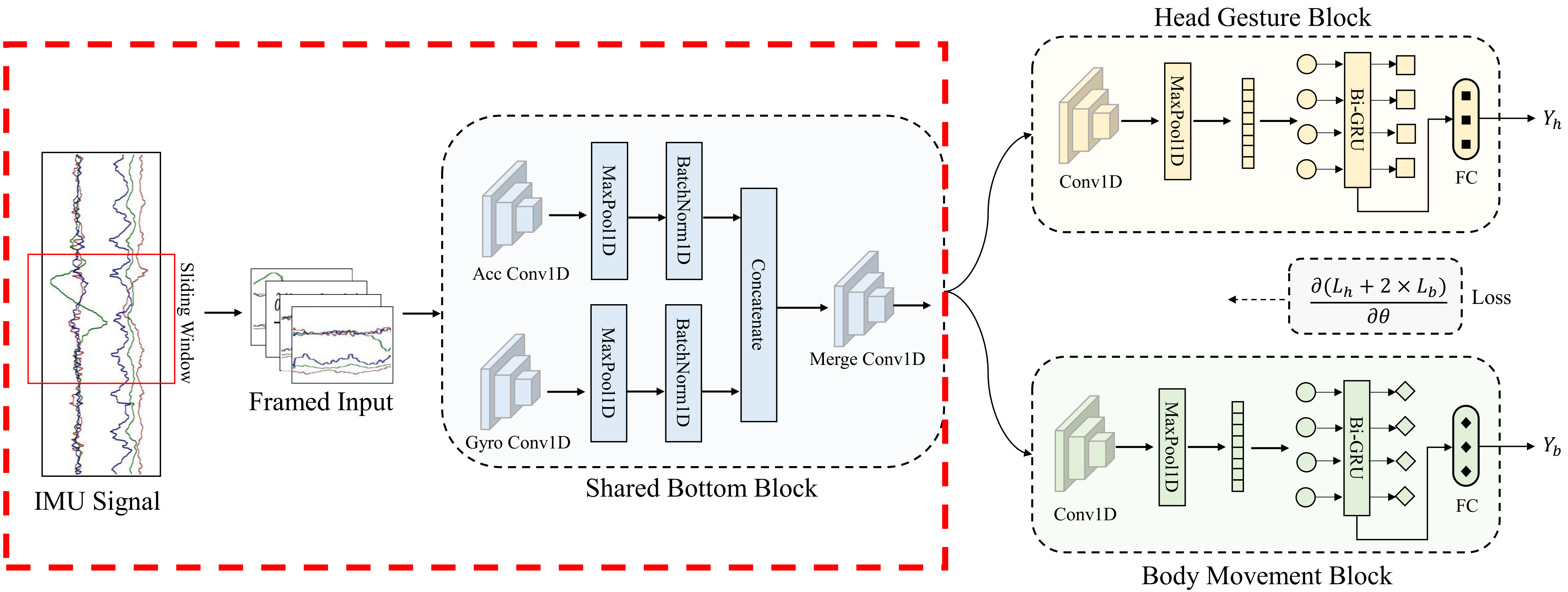


Fig 7. The architecture of our composite-activity recognition network (CARN).

System Design

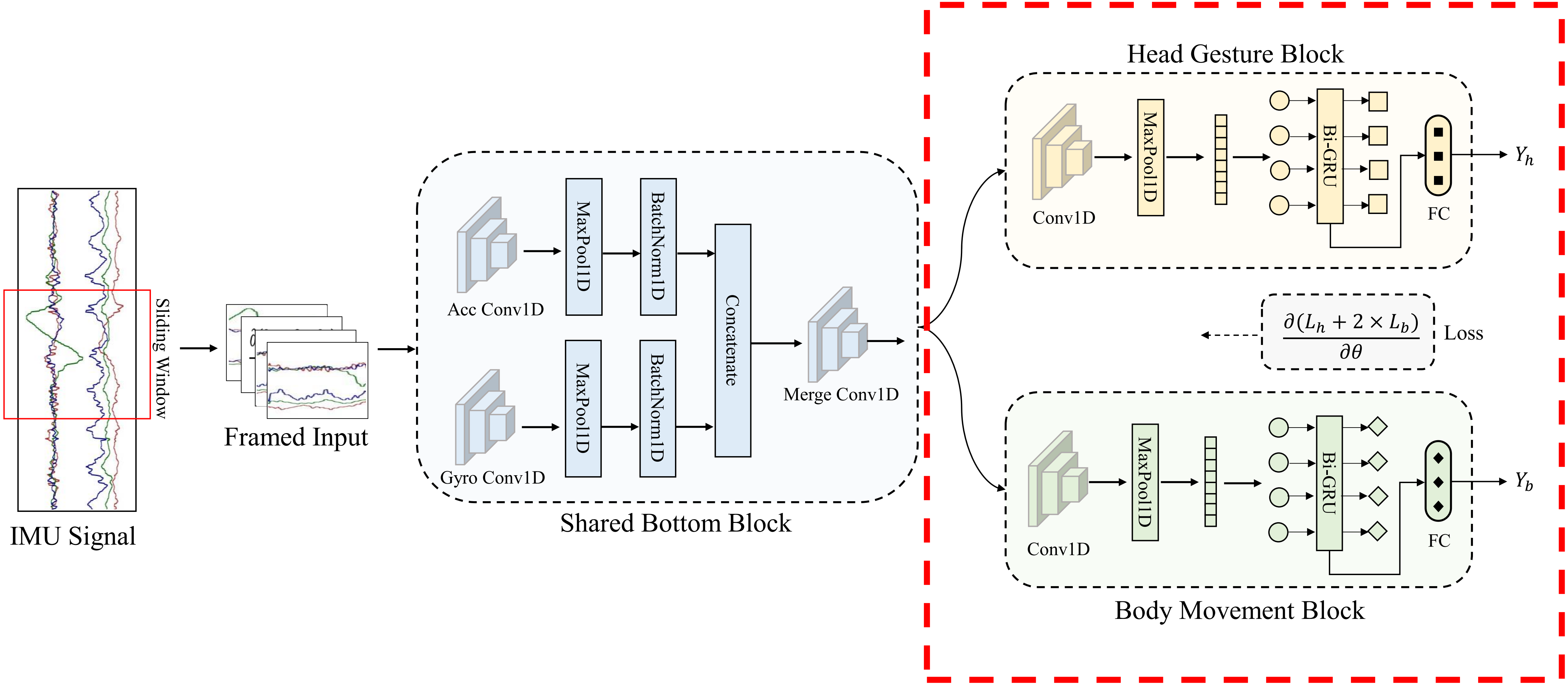


Fig 7. The architecture of our composite-activity recognition network (CARN).

Implementation

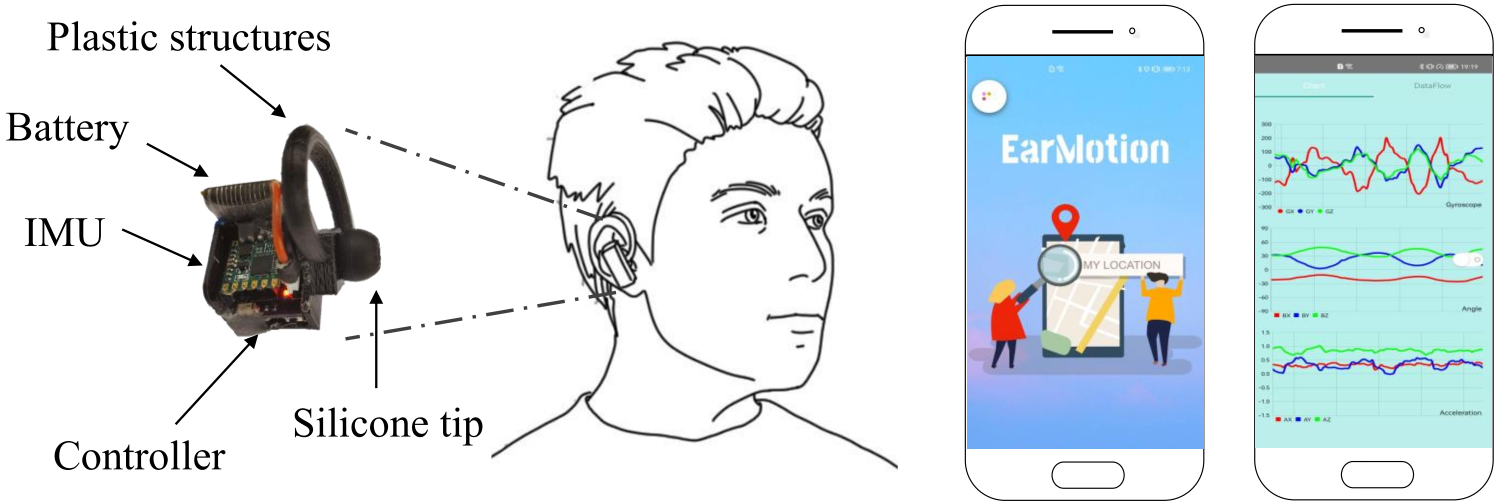


Fig 8. The hardware and mobile application of CHAR.

■ Hardware
- JY901 IMU
- ESP32 as the microcontroller
- 3.7 V lithium battery
- 3D printed plastic enclosure
- HUAWEI Mate40 Pro

■ Dataset
- 15 participants
- 12 head gestures
- 5 body movements
- 5 repetitions
- 15×12 ×5 ×5 = 4500 instances

■ Server
- Intel(R) Xeon(R) Platinum 8260 CPU
- NVIDIA GeForce RTX 2080Ti GPU

Evaluation

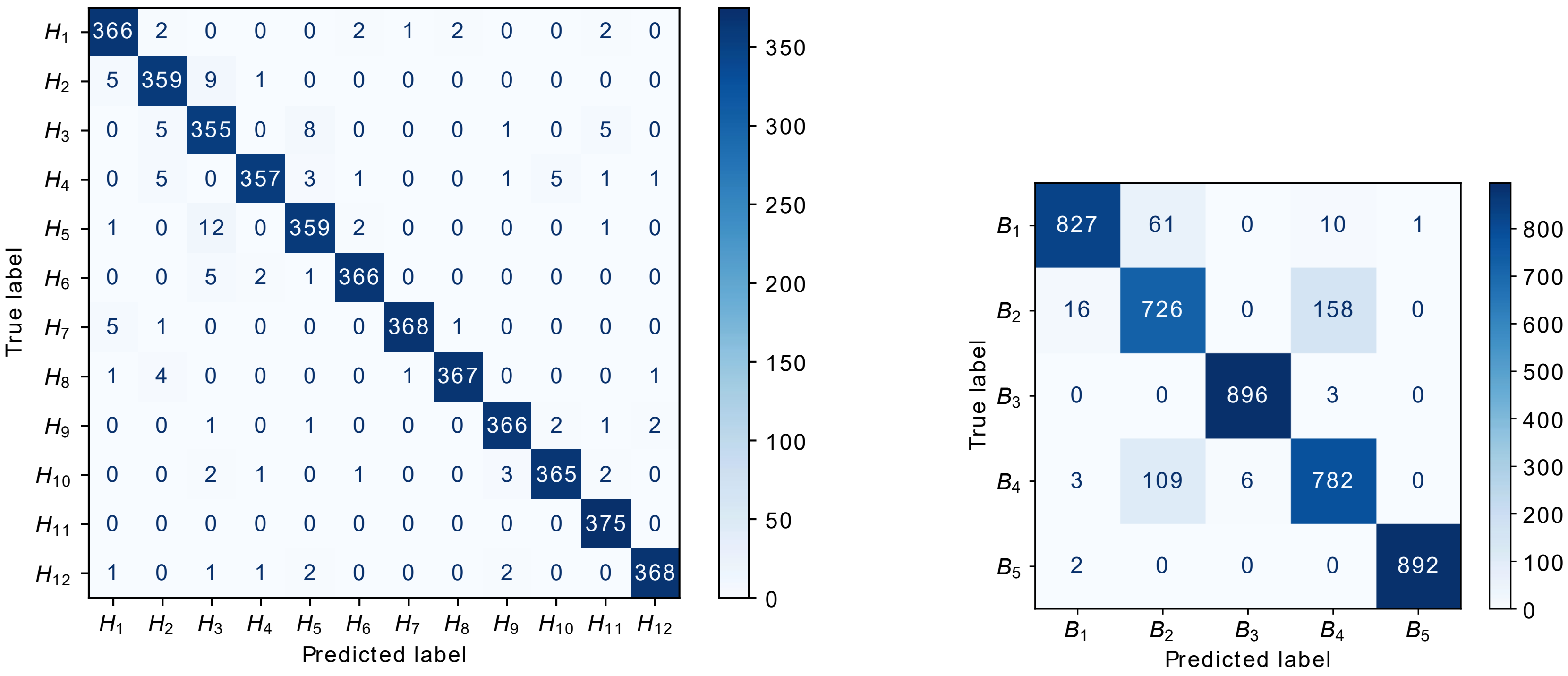


Fig 9. Confusion matrices of head gestures and body movements recognition in user-independent cases.

- Activity Segmentation: MDR = 2.8%, FDR = 1.2%
- Activity Recognition: 97.7% for head gestures, 92.0% for body movements

Evaluation

Tab 1. Accuracy using different kinds of data segment.

Training / Testing	Head gesture	Body movement
D_h / D_h	97.7%	92.0%
D_h / D_a	93.7%	91.9%
$\boldsymbol{D_a}$ / $\boldsymbol{D_a}$	**94.9%**	**89.8%**

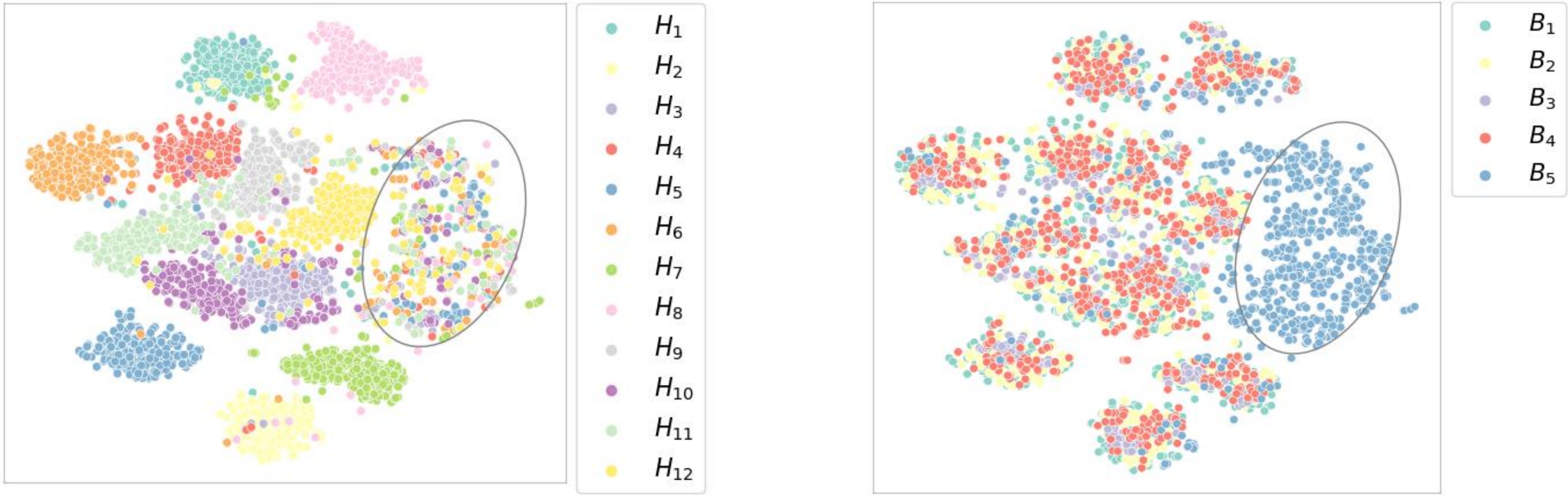
- Cascade Performance
  - 94.6% for head gestures
  - 89.8% for body movements

Tab 2. Accuracy of benchmarks and our network.

Network	Head gesture	Body movement	Composite activity
DCNN	91.2%	81.0%	73.4%
DeepSense	92.4%	76.5%	69.8%
DeepConvLSTM	88.1%	84.9%	74.4%
CARN	**97.7%**	**92.0%**	**89.7%**

- Comparison
  - CARN outperforms existing classical networks.

Evaluation



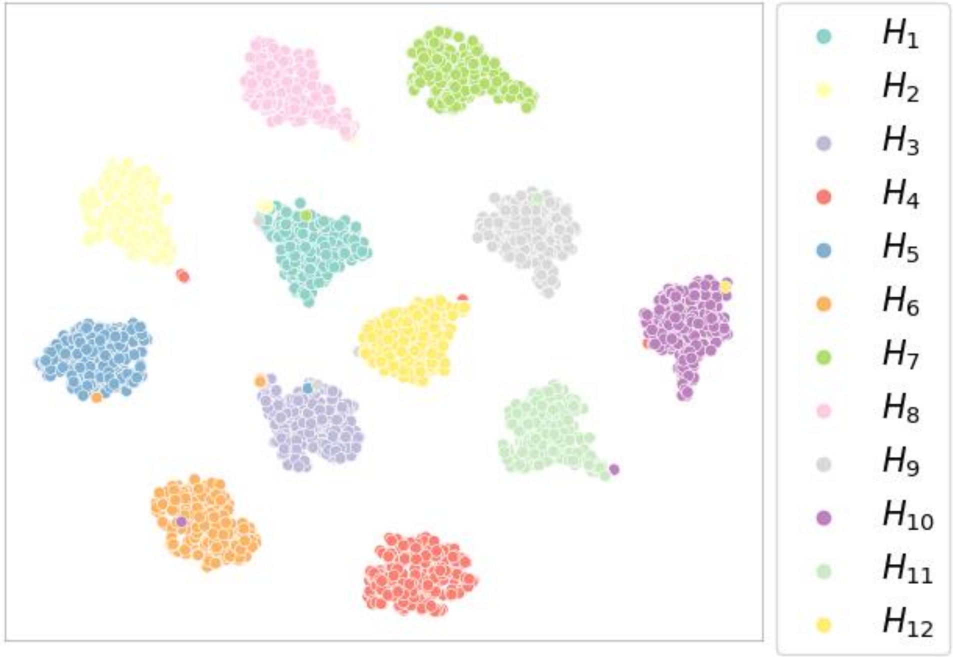
(a) Colored by head gestures.

(b) Colored by body movements.

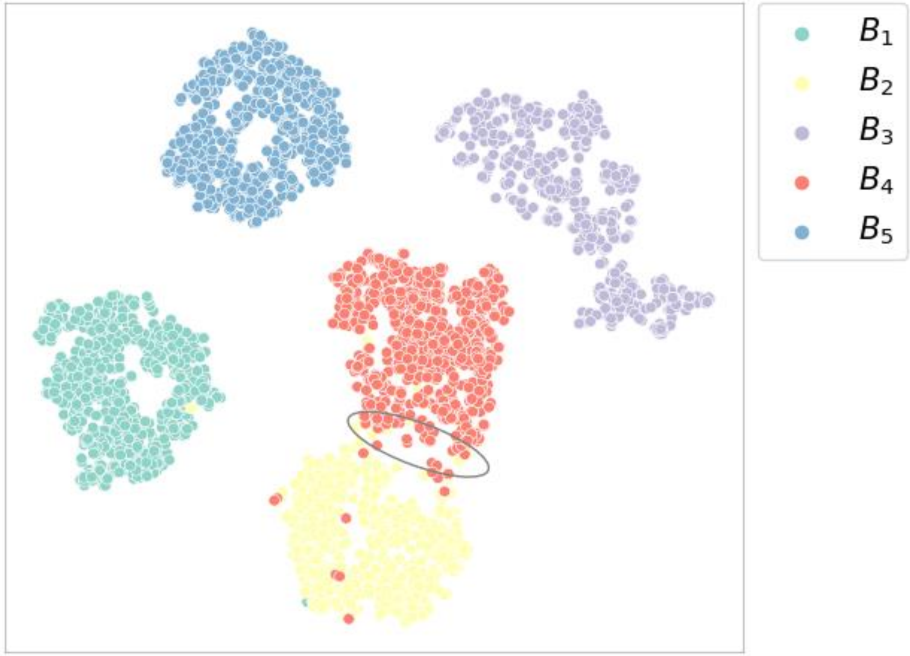
Fig 10. Embedding visualization with t-SNE algorithm for features extracted by Shared Bottom Block of CARN. The figure corresponds to the same features and has been colored in terms of different types of activities.

- The shared bottom block of the network distinguishes activities to a certain extent, but cannot achieve fine-grained composite activity recognition

Evaluation



(a) Head gesture embedding.



(b) Body movement embedding.

Fig 11. Embedding visualization with t-SNE algorithm for features further extracted by Head Gesture Block (a) and Body Movement Block (b) of CARN, respectively.

- The task-specific top blocks of the network have the ability to further extract unique features and achieve better composite activity recognition

Evaluation

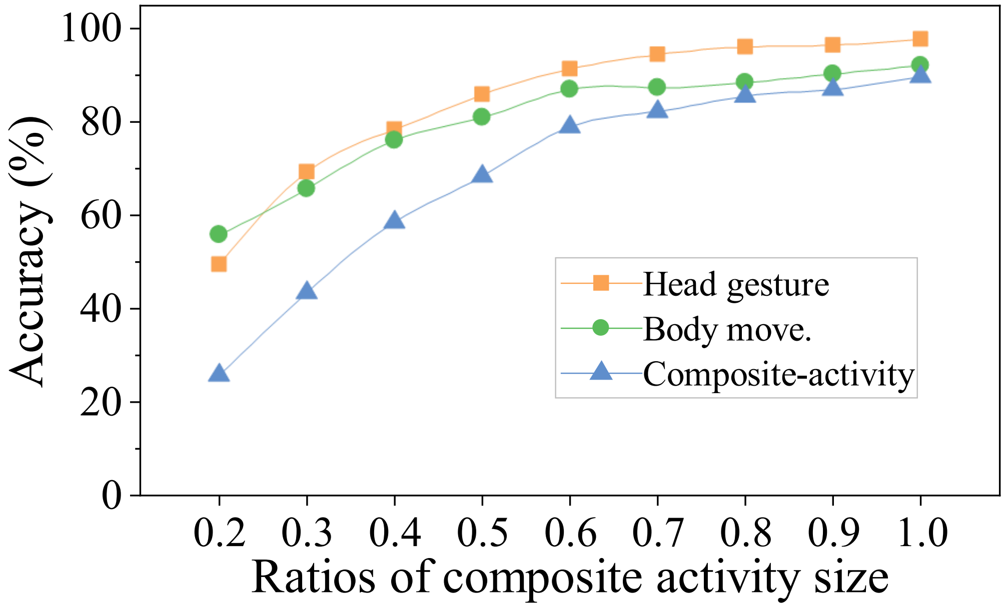


Fig 12. Recognition accuracies with different number of training activity types.

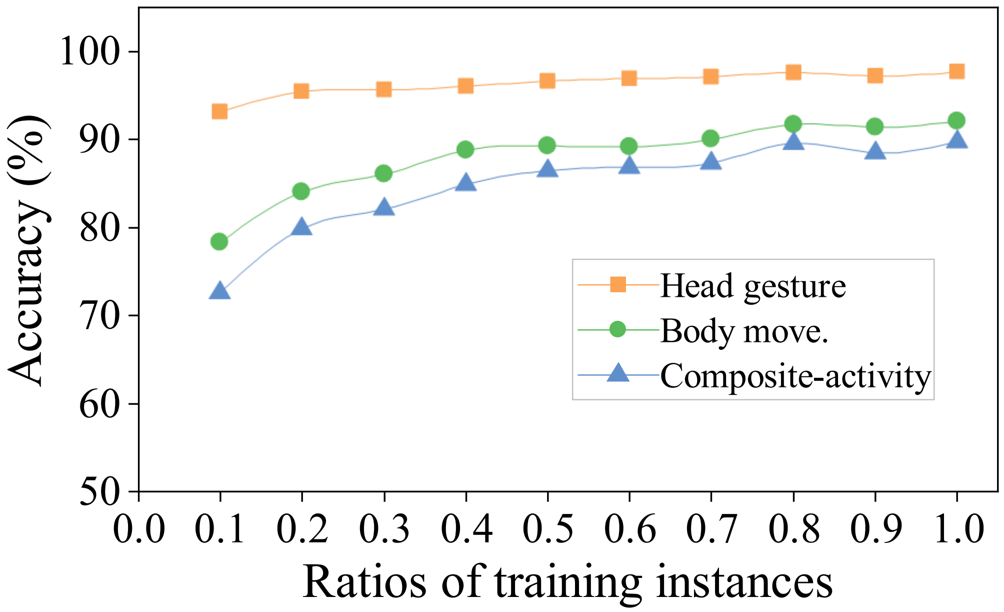


Fig 13. Recognition accuracies with different training dataset sizes.

- Network can be trained using data from some composite activities and applied to all

- With the percentage increasing, the accuracies initially increase and then remain stable
- It requires more training data to recognize body movements

Evaluation

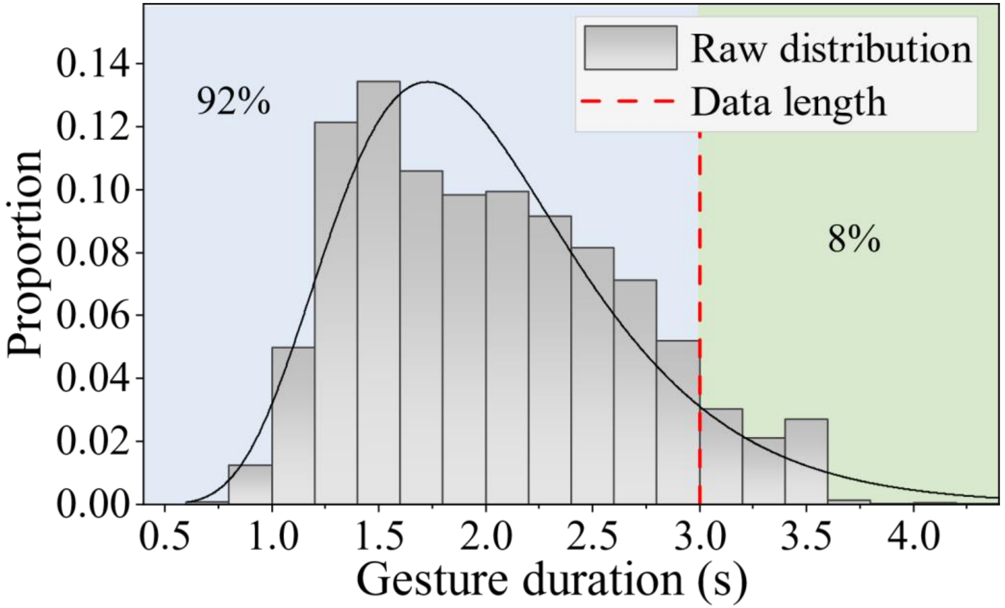


Fig 14. Statistical histogram of head gestures performing time.

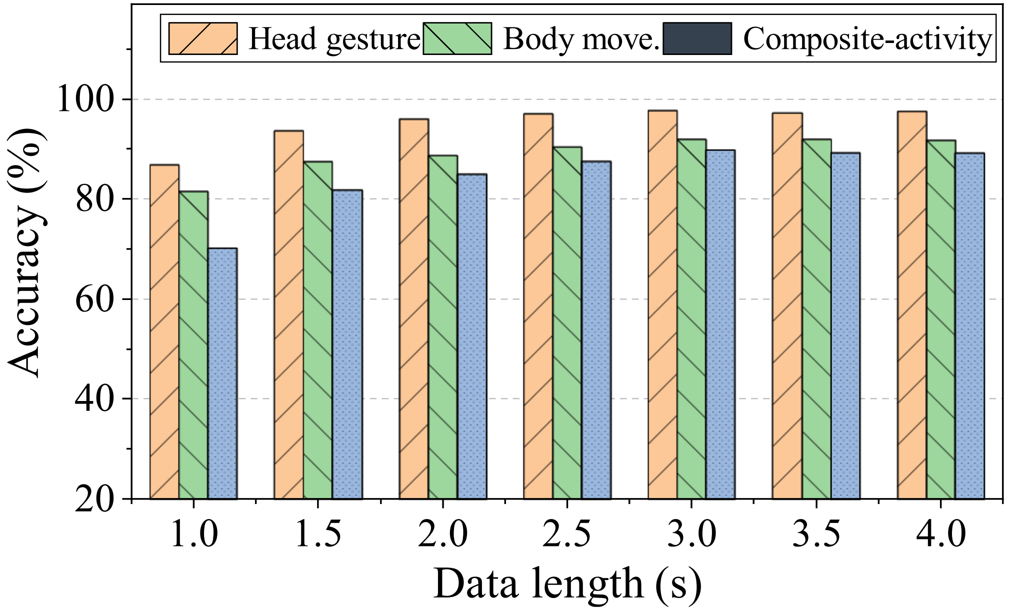


Fig 15. Recognition accuracies with different data lengths.

- More than 90% of the head gestures last less than 3.0 seconds.
- Indicates a proper range of data length.

- Reduce the data length appropriately for mobile device with limited resources.

Evaluation

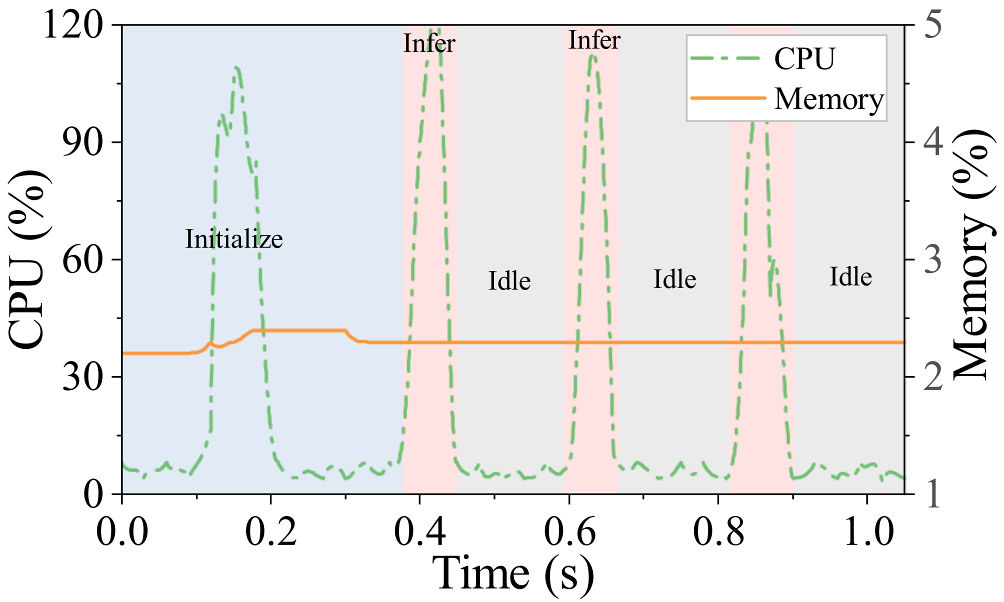


Fig 16. CPU and memory occupation.

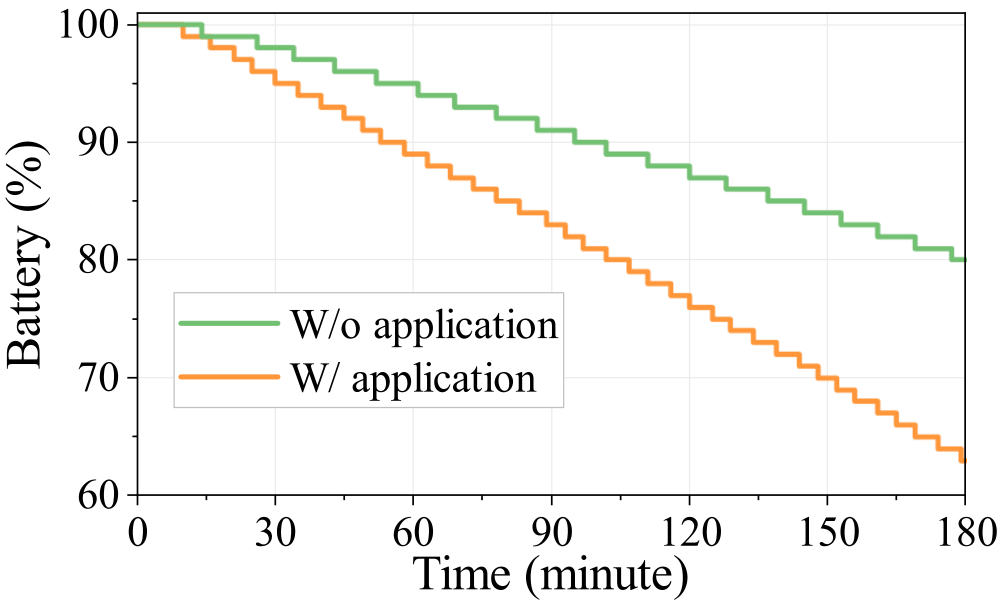


Fig 17. Energy consumption.

- CPU: 10% when idle
- Memory: 2.4%

- Energy: 4.18 mAh/minute
- Response Time: 52 ms

Conclusion

- Consider a novel HAR problem and design a multi-task learning network to recognize head-body activities.

- Implement an earphone-based real-time prototype system with low-cost hardware and a self-developed mobile application.

- Demonstrate CHAR can recognize 60 head-body composite activities with a high accuracy even in the user-independent case.

Thanks

Q&A