Queen Mary University of London

Imperial College London

Sense Me without Knowing *Me*!

Mohammad Malekzadeh (PhD Student in CS)

A Joint Work with:

Richard G. Clegg, Andrea Cavallaro, and Hamed Haddadi.

Context

1973

- Location (~50m)
- Microphone

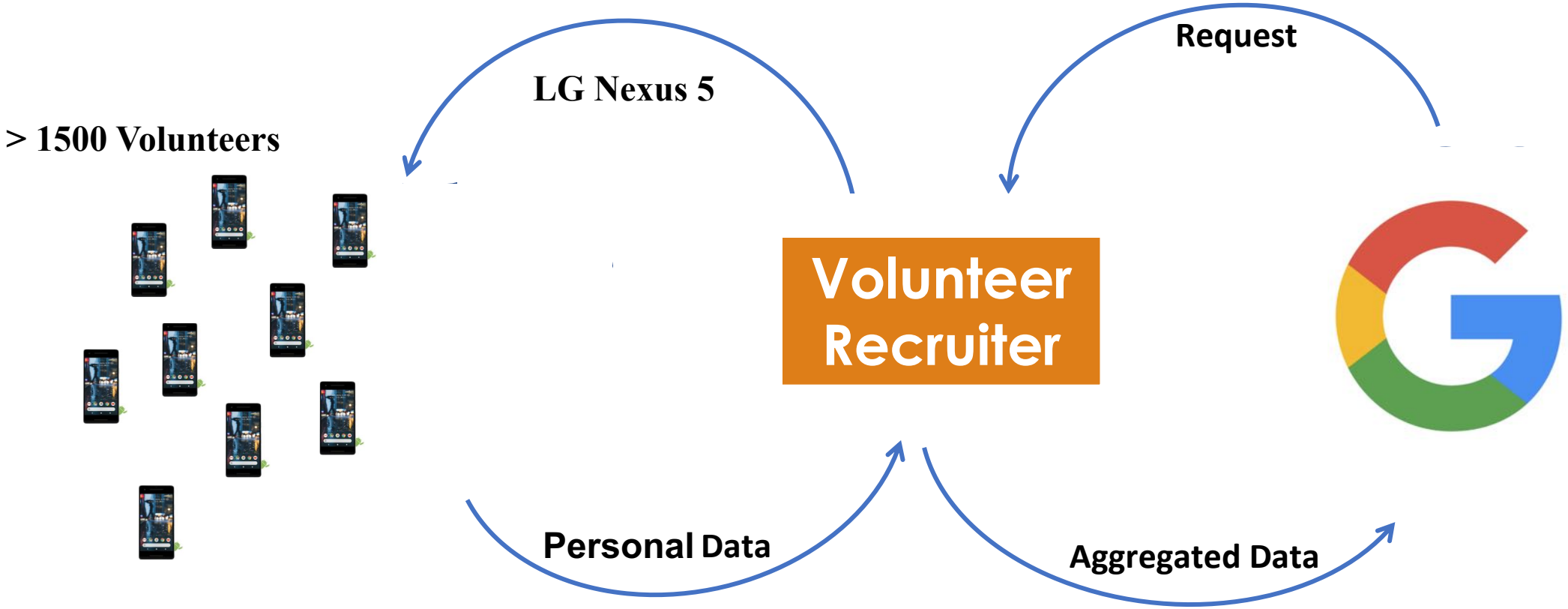
2018

- Location (~3m)
- Microphone
- **Gyroscope**
- **Accelerometer**
- Barometer
- Magnetometer
- Thermometer
- Proximity
- Ambient Light
- Humidity

Smart devices measure more and more data every generation.

An Example

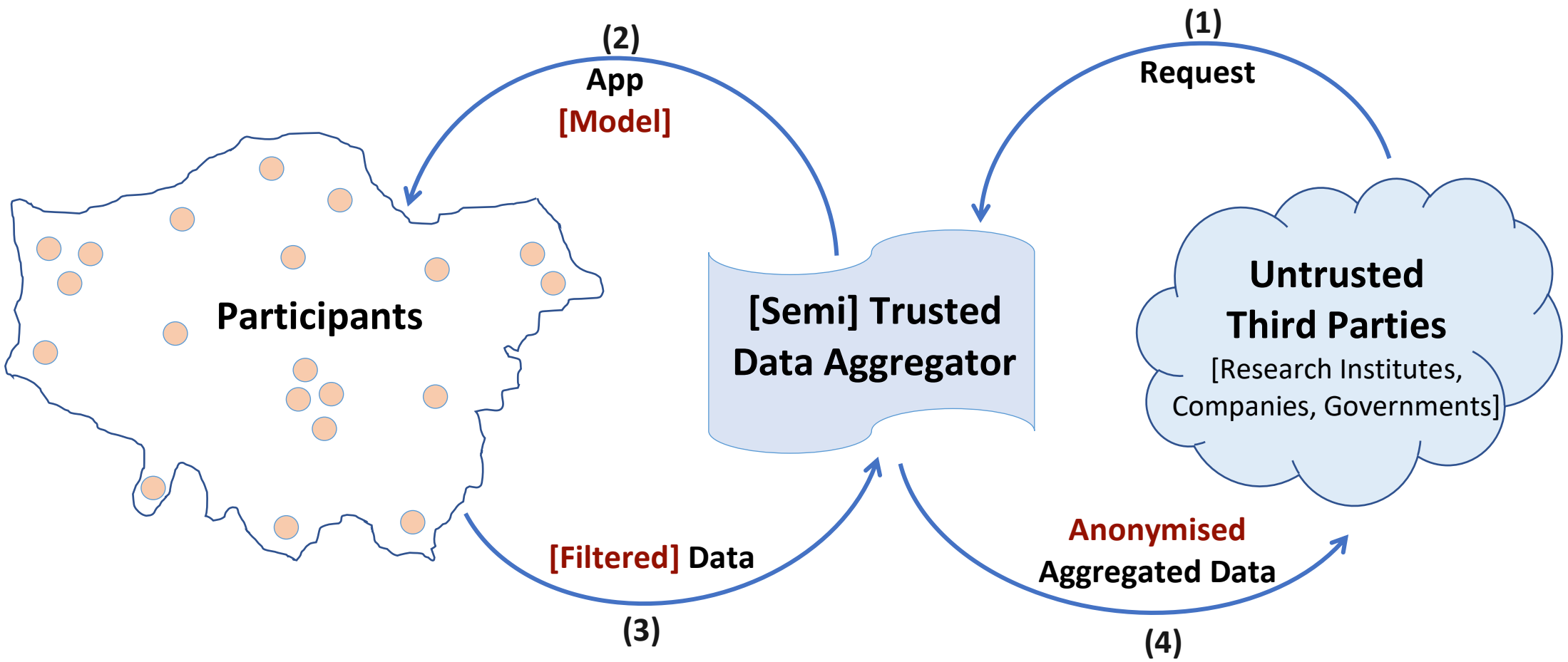
Google ATAP project Abacus



- **Goal:** using biometric patterns, like motion, instead of password

Ref: DOI 10.1109/ACCESS.2016.2557846

Context



Privacy-Preserving Sensing

MotionSense Dataset

- ❑ **Same Activity Set : 6 ADL activities**
- ❑ **Same Place**
- ❑ **Same Phone in the Front Pocket**
- ❑ **Accelerometer and Gyroscope**

❖ **24 Different Subjects :**

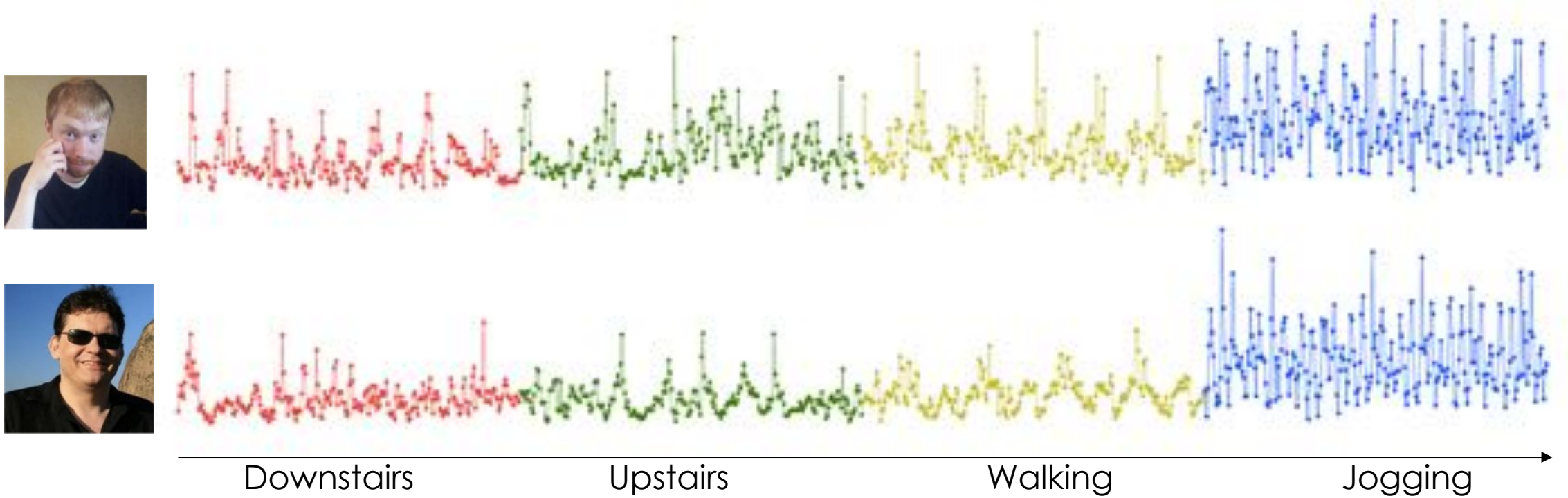
- **Gender:** 14 male - 10 female
- **Age:** [18 – 40] years old
- **Weight:** [45 ,105] kg
- **Height:** [160 , 195] cm

DATABOX 01000100 01100001 01110100 01100001 01100010 01111000

SensingKit

Queen Mary University of London

MotionSense Dataset



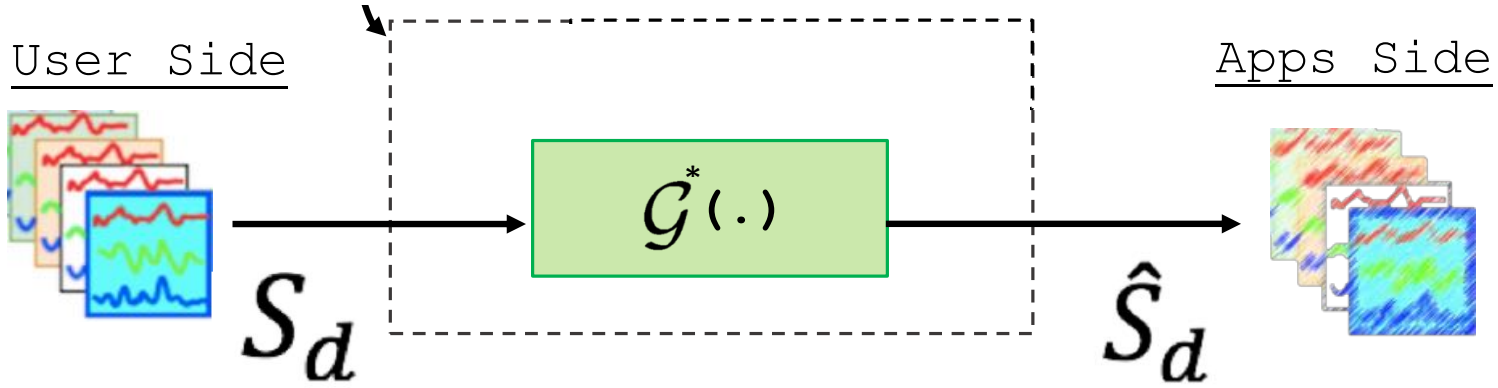
Accelerometer$_{(magnitude)}$ Data

Some Results

- 1-D Accelerometer$_{(magnitude)}$: (50Hz)
- Time-Window 5 second
- Deep Convolutional Network

	Classification Accuracy
activity	~ 98%
gender	~ 96%
Identity	~ 89%

Sensor Data Anonymisation

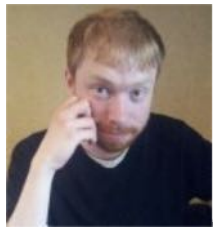


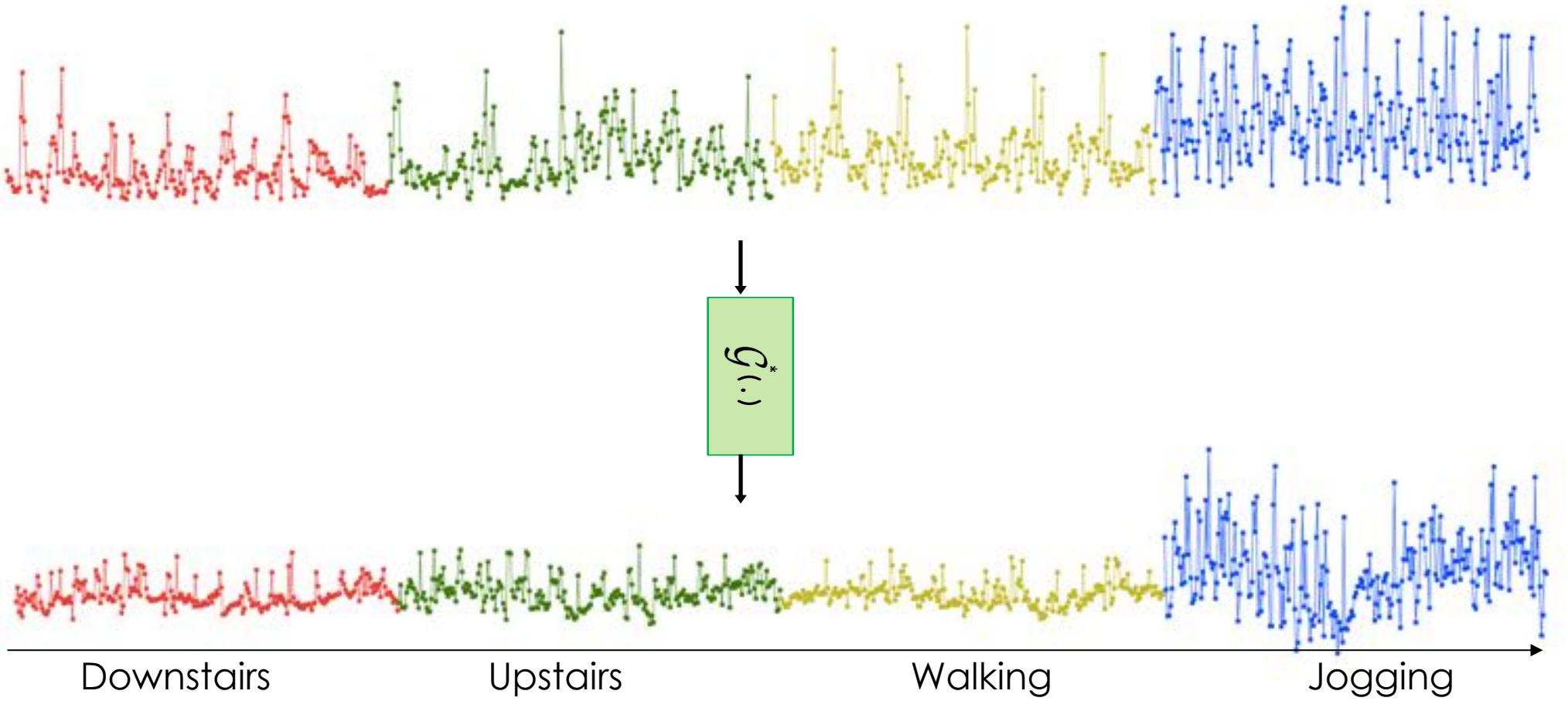
$$\mathcal{G}^*(.) = \underset{\mathcal{G}(.)\in\mathcal{F}}{\operatorname{argmin}} \left(p\left(I_{\mathbf{s}}\left(\hat{S}_d\right)\right) - p\left(I_{\mathbf{n}}\left(\hat{S}_d\right)\right) \right)$$

Identity

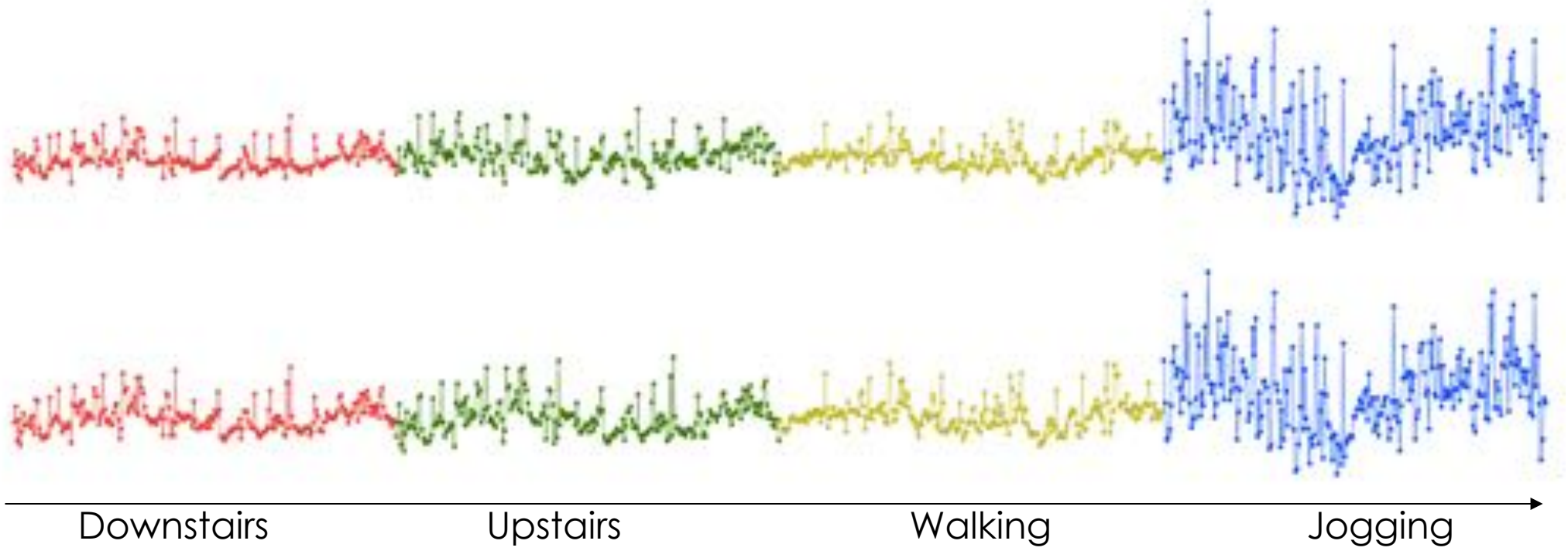
Activity

After Transformation

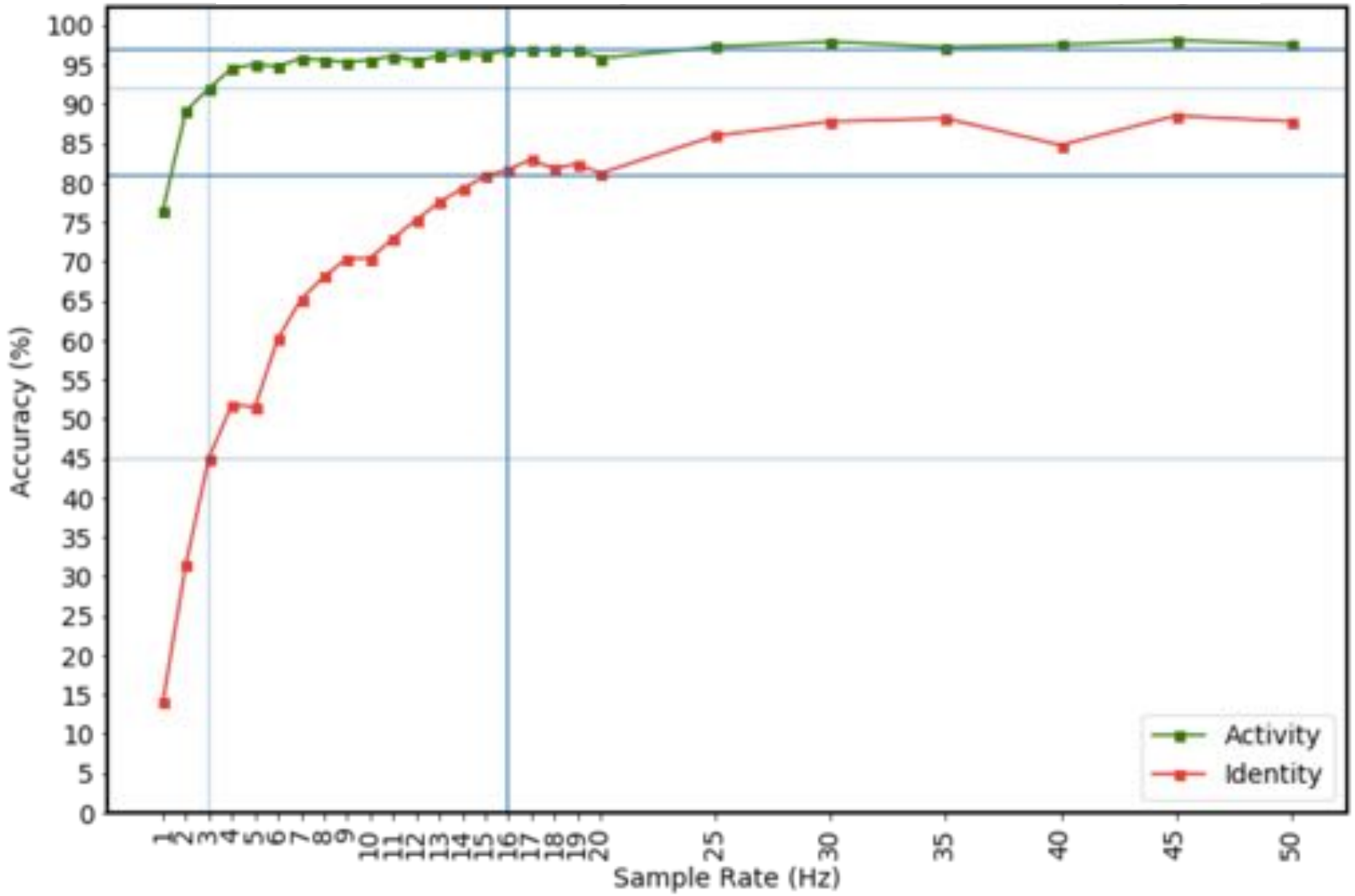




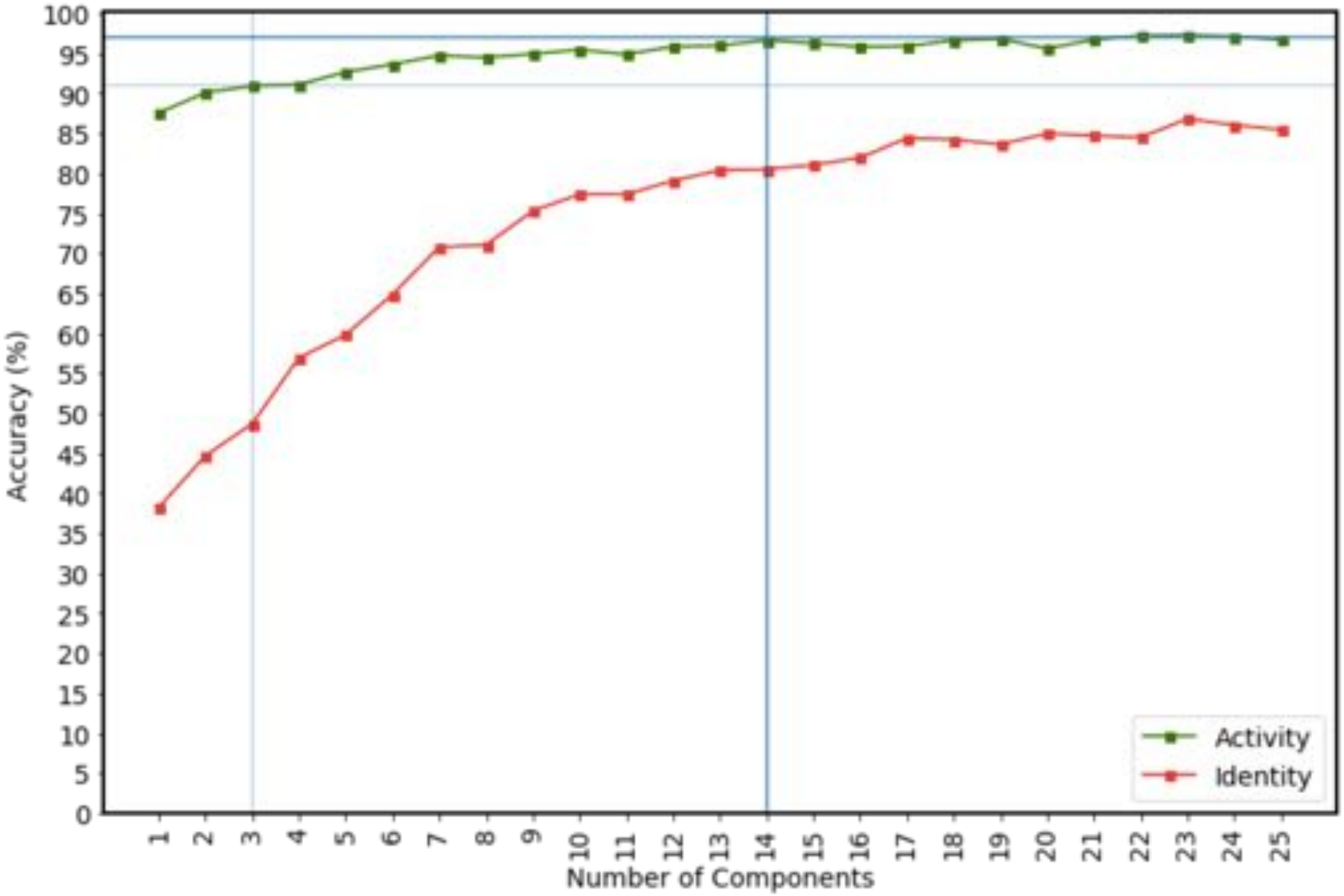
Transformed Data



G0: Downsampling



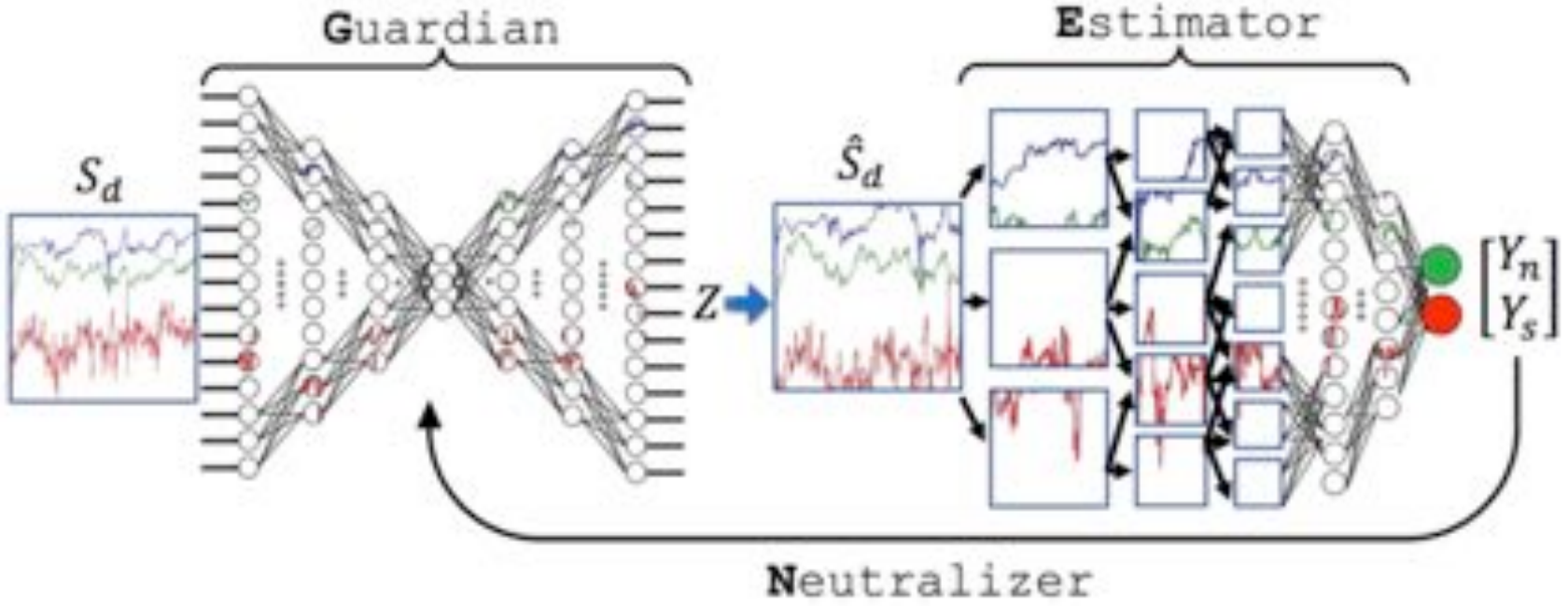
G1: Singular Spectrum Analysis



G*: Feature-Based Reconstruction

I. **Encode** original data into the **feature set**.

II. **Decode** based on features corresponding to **non-sensitive task.**



Comparing the Tradeoffs

	Original	G*: FBR	G1: SSA (14 Components)	G0: DownSampling (16 Hz)
Activity	**~ 98**	~ **92**	~ 92	~ 92
Identity	**~ 89**	**~ 18**	~ 60	~ 55

- The accuracy(%) on validation data during the training process

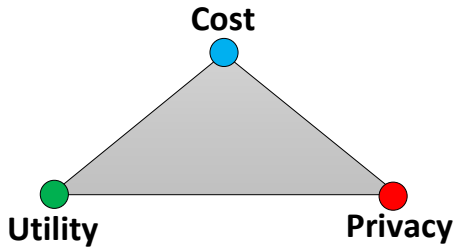
- Related Work:
  - [Xiao et al. 2018] : mutual information based feature selection and sampling rate adjustment

	Original	Best Sub Set of Features	Most Private One
Activity	**~ 95**	**~ 95**	**~ 70**
Identity	**~ 95**	**~ 33**	**~ 16**

DOI 10.1186/s13673-018-0137-6

Next Steps

- **Practical:**
  - The **Cost** of the solution on **Edge** devices?



- **Theoretical:**
  - Provide a **statistical guarantee** (probabilistic bound)
    - Differential Privacy : **Composition Theorem?**
    - Mutual Information : **Joint Distributions?**

Thanks!

Link to the Dataset and Paper's Repository: bit.ly/eli-dw18

Imperial College London

DATABOX
01000100 01100001 01110100 01100001 01100010 01111000

Queen Mary University of London

CIS centre for intelligent sensing