COMP3516: Data Analytics for IoT

Lecture 7: Mobile Sensing

Chenshu Wu

Department of Computer Science

2025 Spring

香 港 大 學
THE UNIVERSITY OF HONG KONG

HKU AIoT LAB
香港大學人工智能物联网實驗室

Contents

- Mobile Sensors and IMUs
- Sensor Data Processing
- Human Activity Recognition
- Mobile Sensing Applications

Mobile sensing vs. Wireless sensing

- Wireless sensing: Hardware installed somewhere in the space
  - Freedom!
- Mobile sensing: Users need to carry a (mobile) device
  - Goes anywhere!

- A key paradigm shift:
  - Device-based vs. device-free
  - Or Active vs. passive
  - Or Contact-based vs non-contact/contactless

Smartphone & Sensors

- Tell a feature on your phone and identify the enabling sensor
  - E.g., Face ID: RGB camera + infrared camera

- *What are the most critical ones (in your opinions)?*

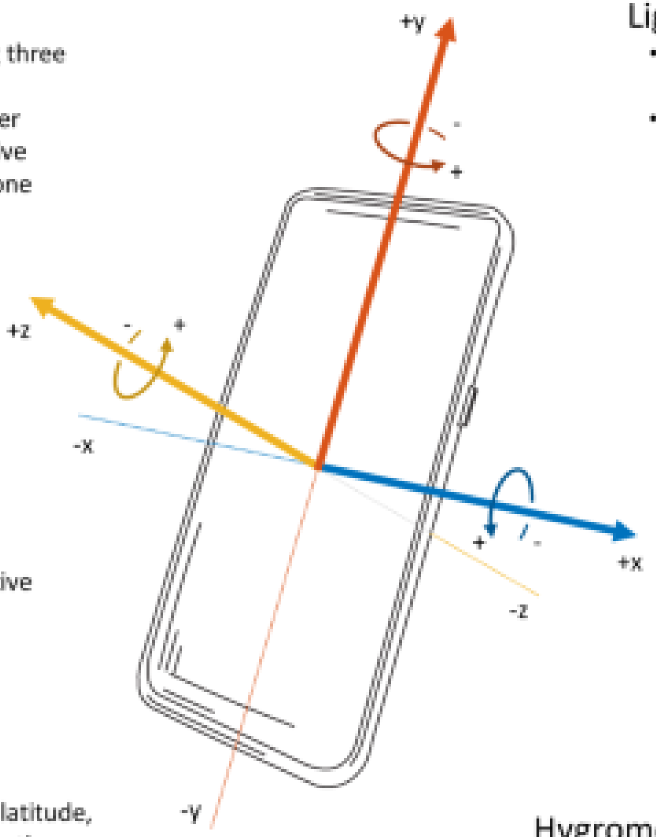
Smartphone sensors

Accelerometer
- Measures rate of change of velocity along three orthogonal axes of smartphone
- Output: gravitational units (g) or meters per seconds squared (m/s^2); positive or negative depending on the orientation of smartphone

Gyroscope
- Measures angular velocity around three orthogonal axes of smartphone
- Output: radians per second (rad/s); positive or negative depending on the direction of rotation

Magnetometer
- Measures strength of Earth's magnetic field relative to three orthogonal axes of smartphone
- Output: microtesla (µT); positive or negative depending on the orientation of smartphone

GPS
- Measures geolocation of smartphone as latitude, longitude, and altitude coordinates on Earth
- Output: decimal degrees (°)



Light sensor
- Measures ambient light level (illuminance) in front of the sensor (screen)
- Output: lux (lx)

Proximity sensor
- Measures distance between the sensor (screen) and the closest visible object
- Output: centimeters (cm)

Barometer
- Measures atmospheric pressure
- Output: hectopascal (hPa) or millibar (mbar)

Thermometer
- Measures ambient air temperature
- Output: Celsius (C)

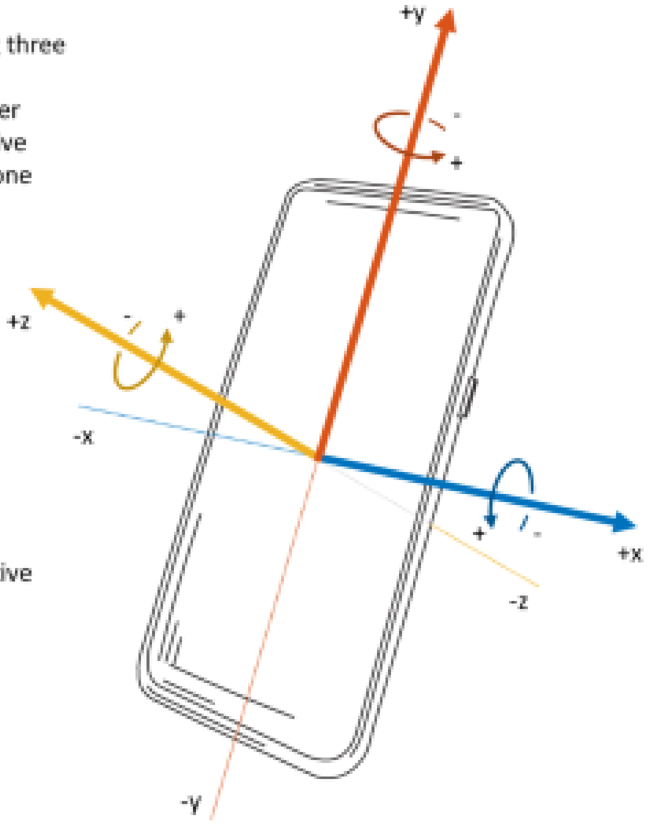
Hygrometer
- Measures ambient relative humidity
- Output: percent (%)

Smartphone sensors

Accelerometer
- Measures rate of change of velocity along three orthogonal axes of smartphone
- Output: gravitational units (g) or meters per seconds squared (m/s^2); positive or negative depending on the orientation of smartphone

Gyroscope
- Measures angular velocity around three orthogonal axes of smartphone
- Output: radians per second (rad/s); positive or negative depending on the direction of rotation

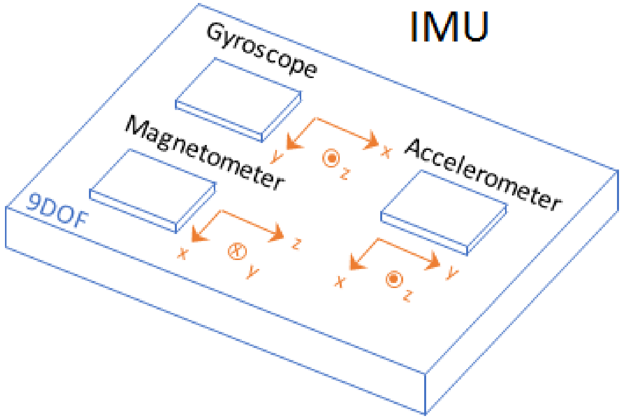
Magnetometer
- Measures strength of Earth's magnetic field relative to three orthogonal axes of smartphone
- Output: microtesla (μT); positive or negative depending on the orientation of smartphone



Barometer
- Measures atmospheric pressure
- Output: hectopascal (hPa) or millibar (mbar)

Inertial Measurement Units (IMUs)

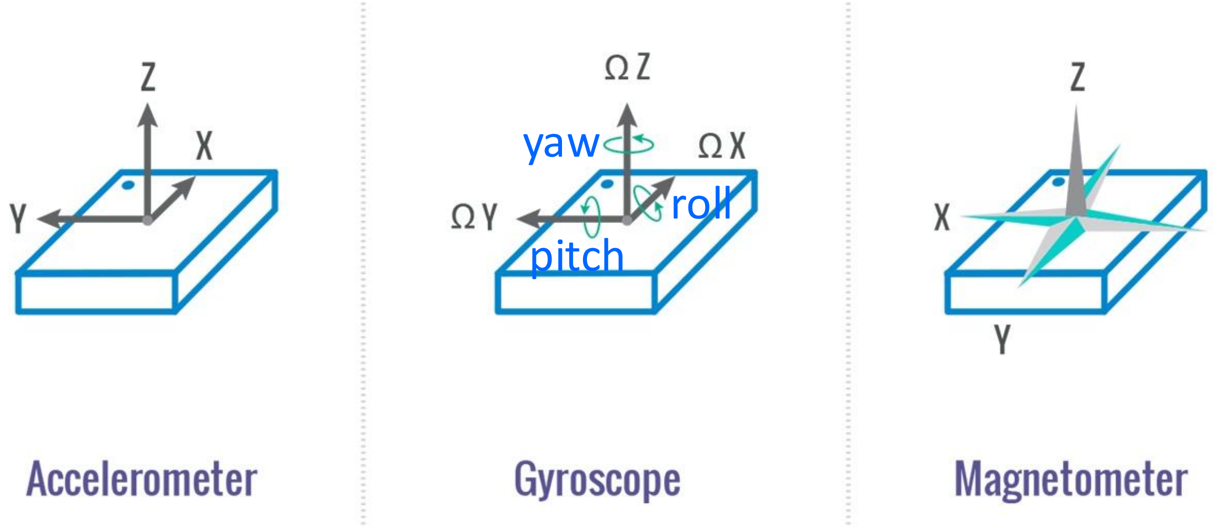
- Accelerometer
  - Measures linear acceleration of movement along three axes in m/s^2
- Gyroscope
  - Measures angular velocity of rotation along three axes in degrees/s
- Magnetometer
  - Measures magnetic field strength in µT (micro-Teslas), indicating orientation



IMUs

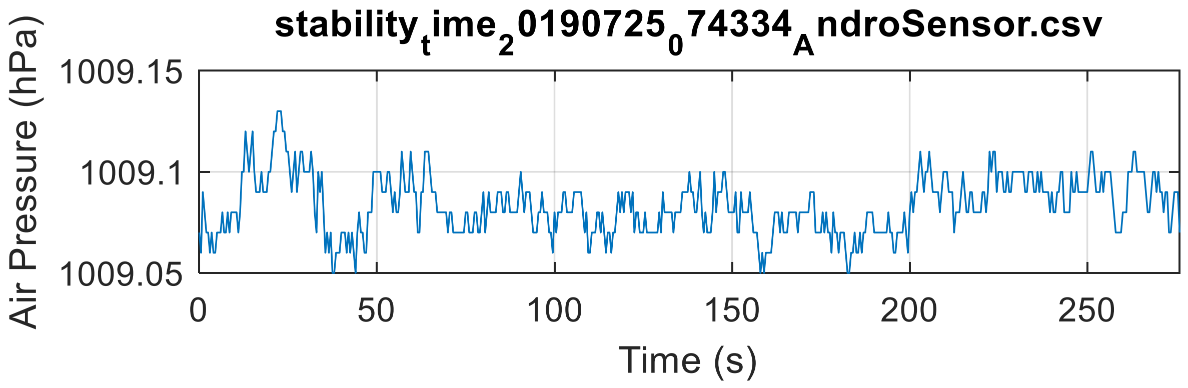
- 6 Degree of Freedom (DoF)
  - Three of linear motion (acceleration): x, y, z
  - Three of rotational motion (angular velocity): roll, pitch, yaw

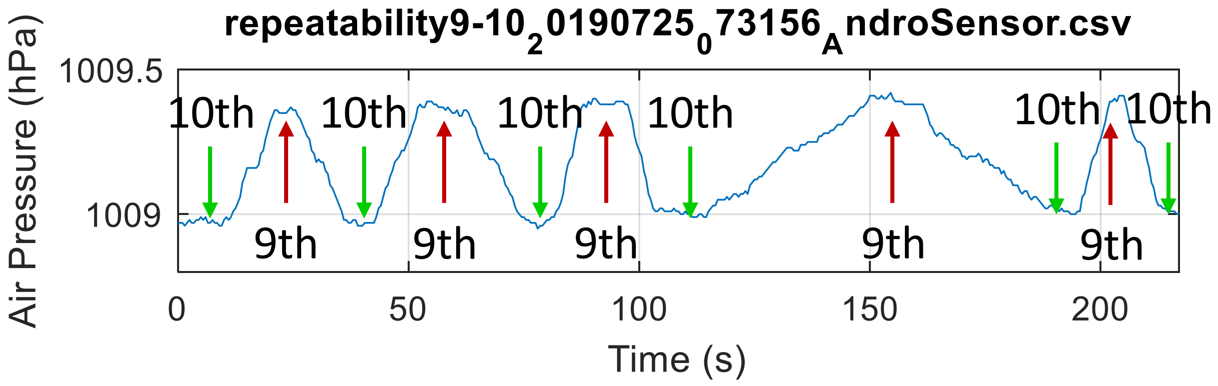
- 9 DoF
  - Fusion of A/G/M
  - Each of 3 axes



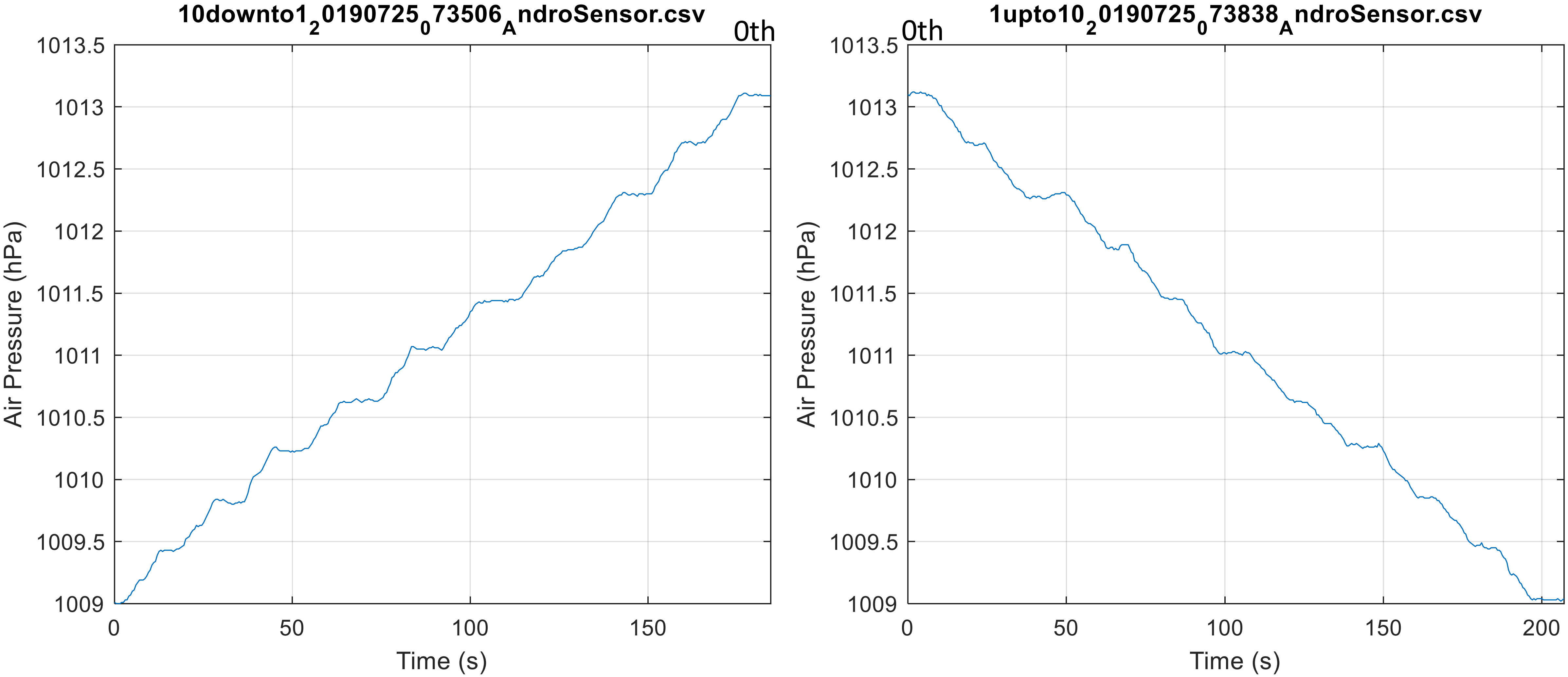
Smartphone Barometer

- Measure atmospheric pressure
  - Can be used to calculate altitude
  - Augment GPS's vertical location
- May not be adequate for absolute altitude estimation, yet sufficient for height change detection
  - Measure pressure difference to infer elevation change

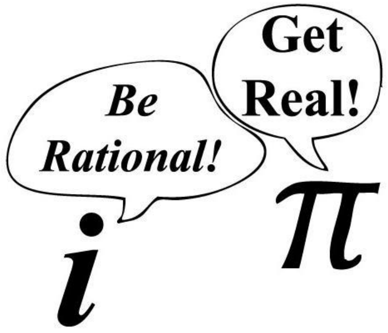




Barometer for floor transition



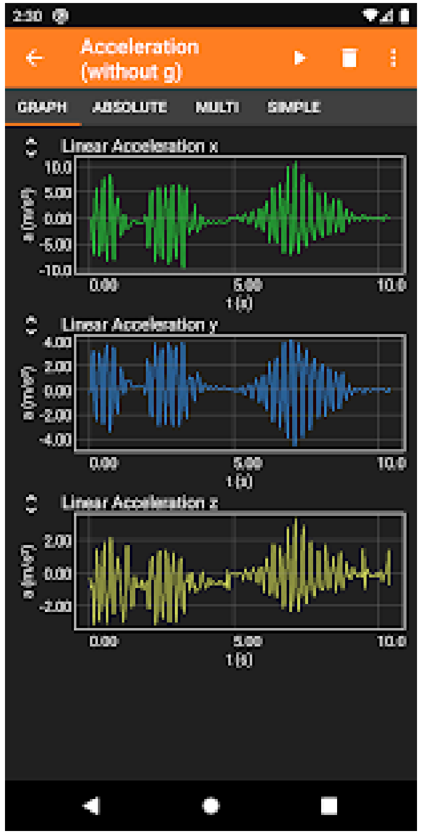
Get Real and Have Fun!



- Have some fun playing with the real sensors using phyphox
  - Available on iOS/Android

phyphox
physical phone experiments

https://phyphox.org/download/



Step Count

- Smartphones and wearables already use IMUs to offer activity indicators

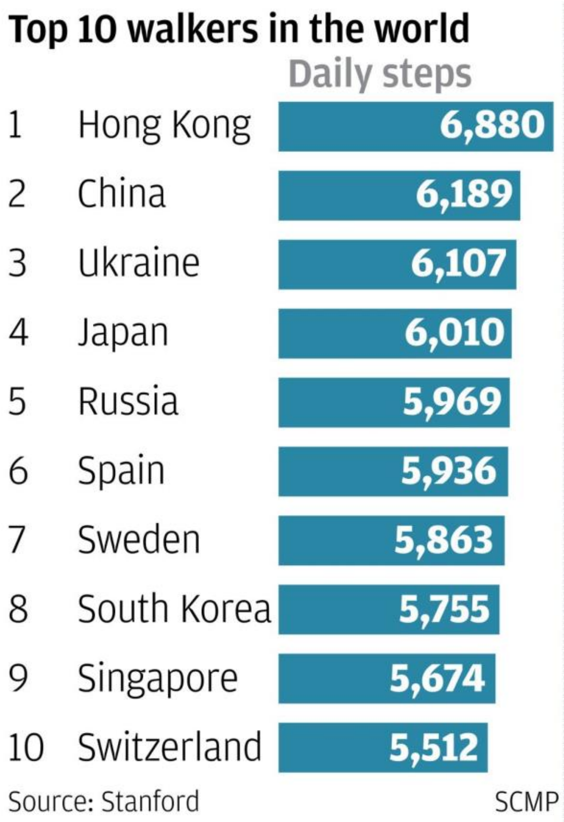
- However, automatic and more refined activity recognition needs more advanced techniques

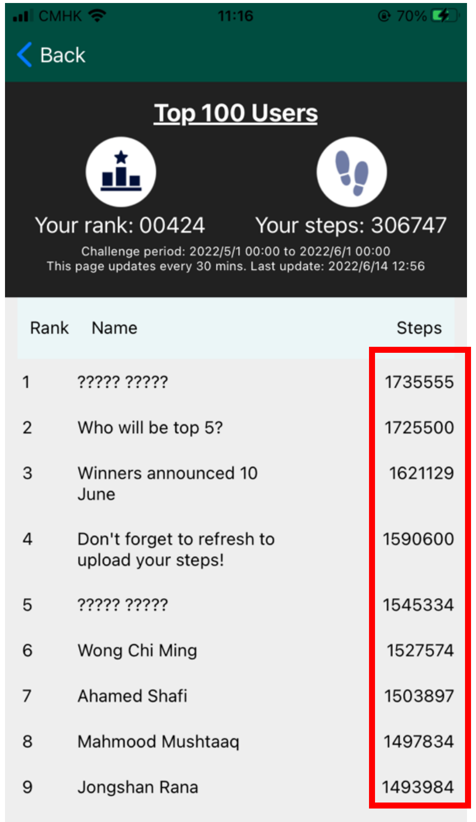
- *What's your daily average?*

Fun Facts

- Hongkongers walk an average of **<u>6,880</u>** steps a day, making them the most active in the world, a study has found.

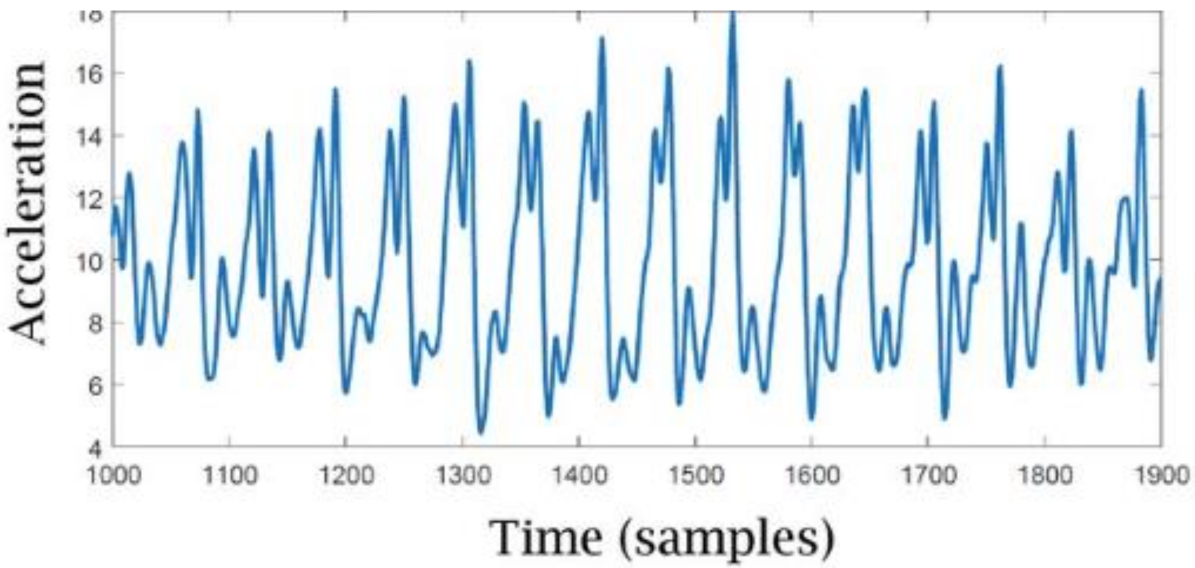
- A Stanford University survey in 2007 ranked Hong Kong first out of 46 countries and regions with 6,880 steps a day, compared with a world average of 4,961.

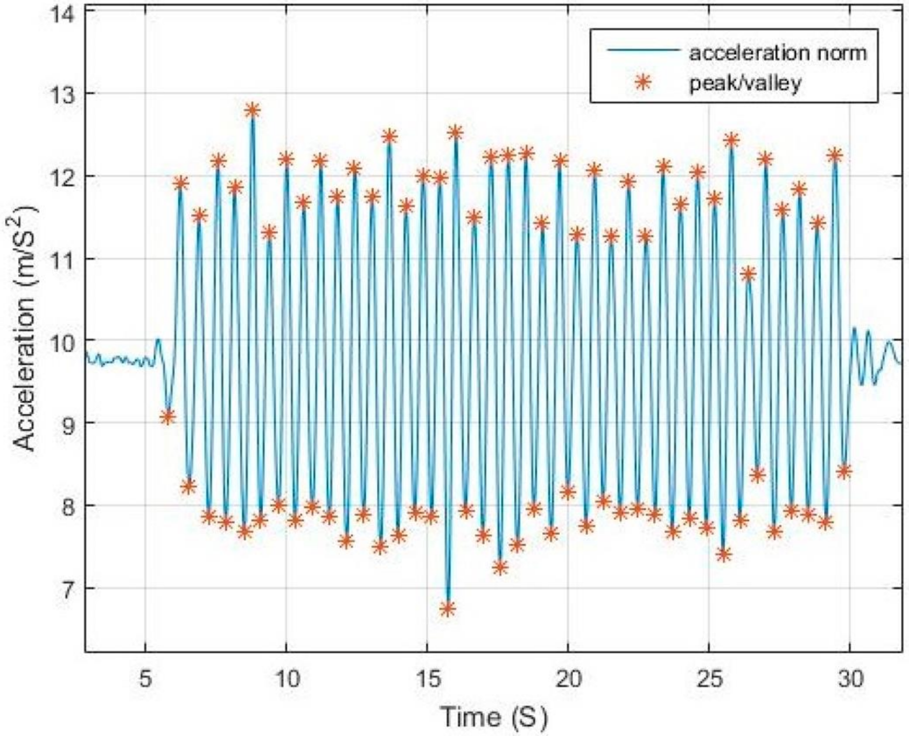




Step Counting

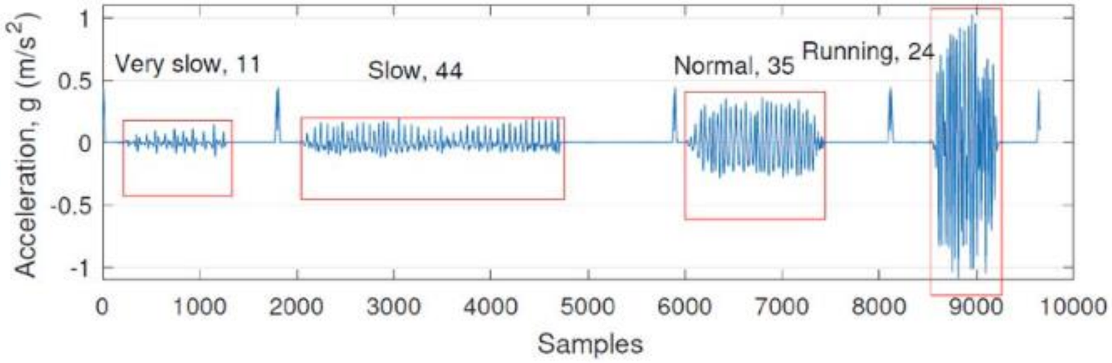
- Walking involves periodic motion sensed by IMU (mainly accelerometer)
  - Walking causes up/down bounce
  - Result in sinusoidal motion in IMU



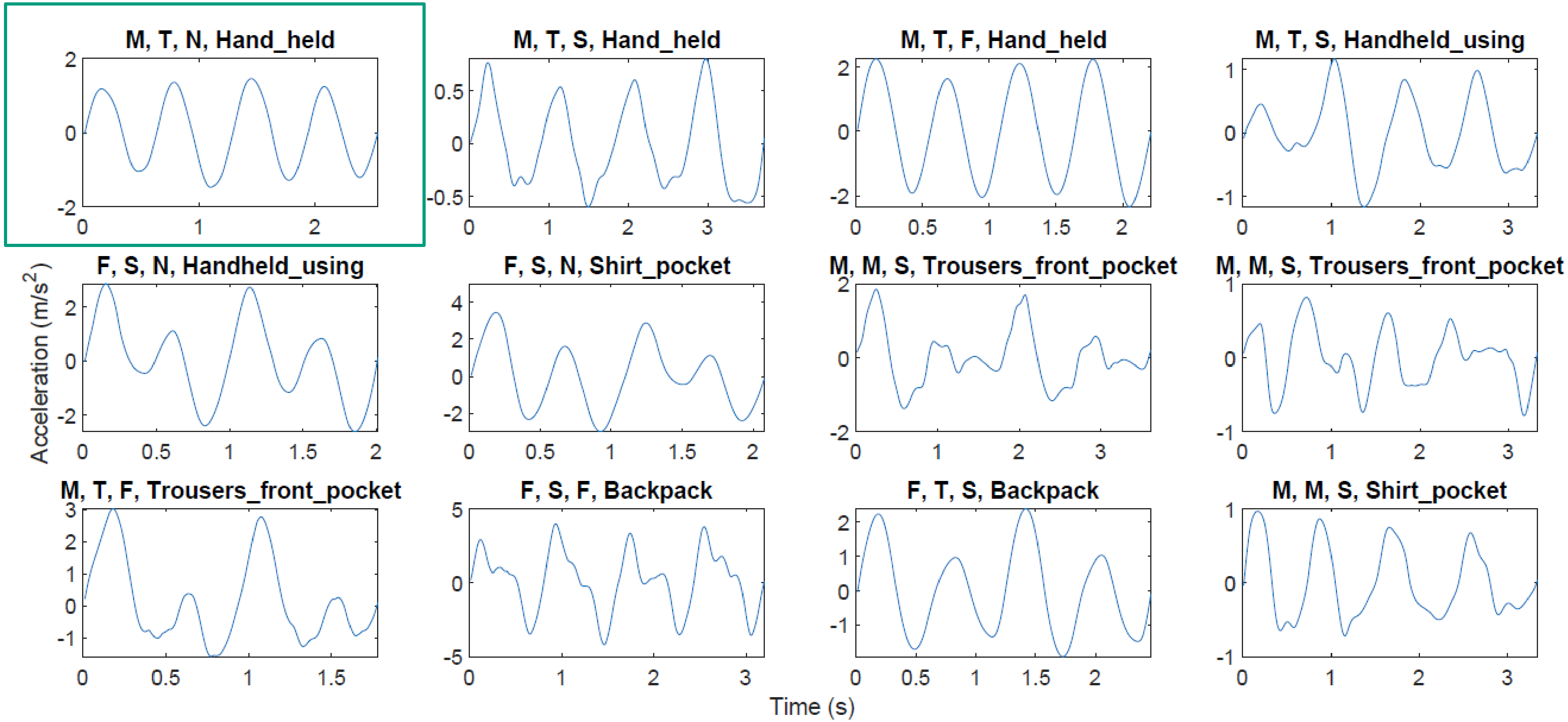


Step Counting

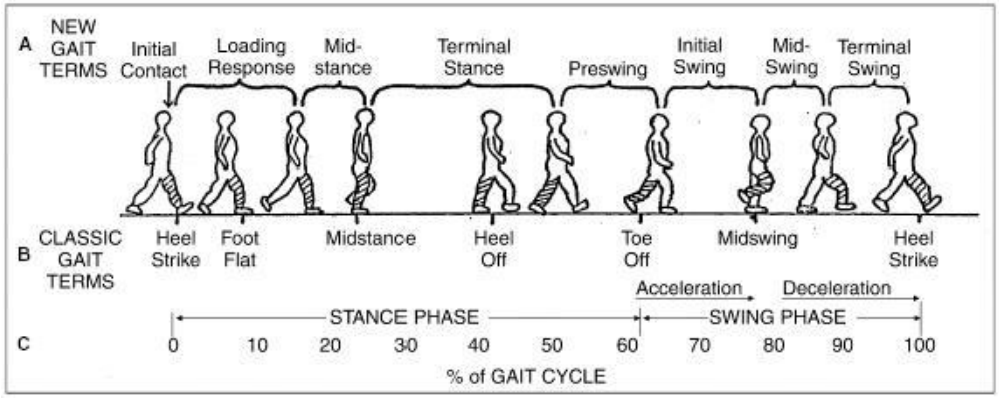
- Methods
  - Zero-crossing, Peak detection, ACF, FFT, …
- What are the problems?
  - IMU senses the sum of walking signal + all other interference
    - Unwanted, unintended motions
  - Phone positions and orientations
  - Diverse user behaviors
  - Different walking speeds



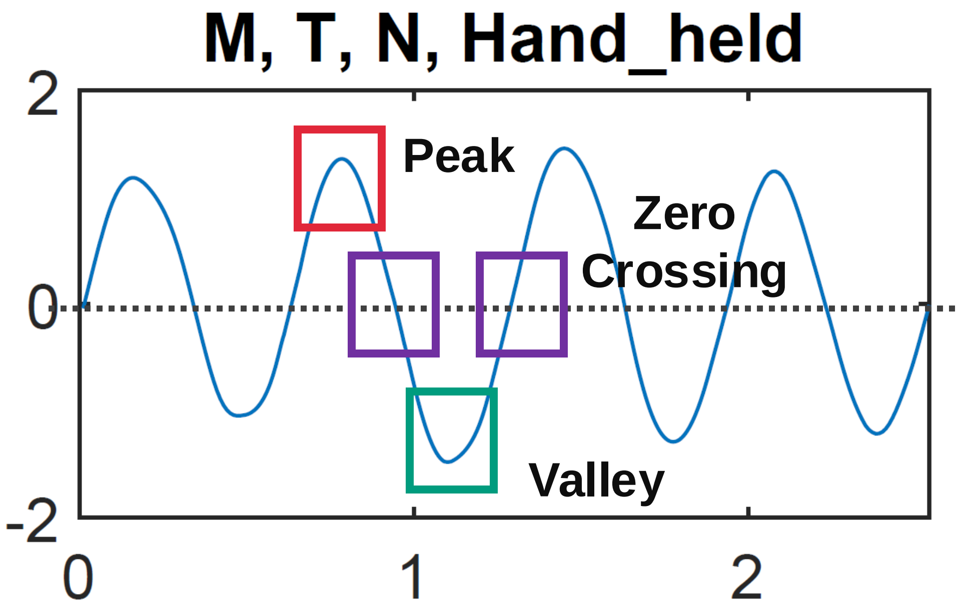
Step Counting

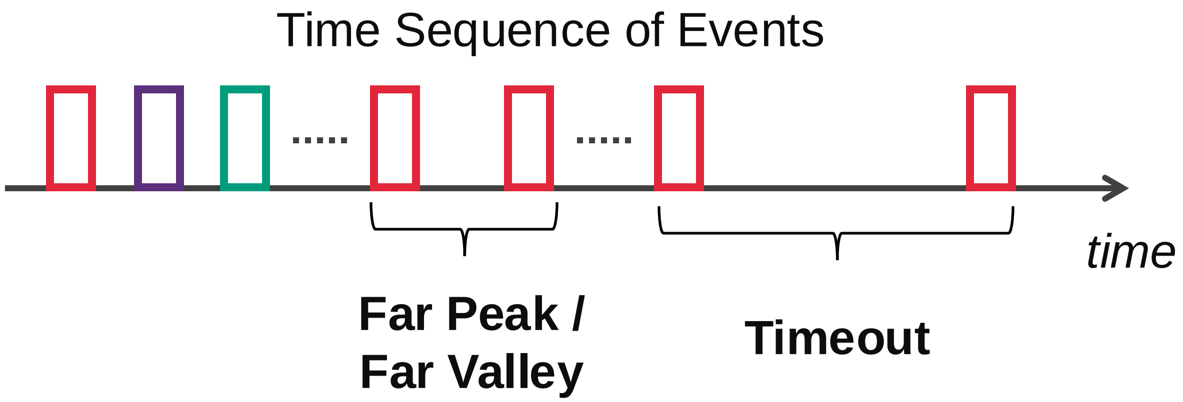


A FSM Approach

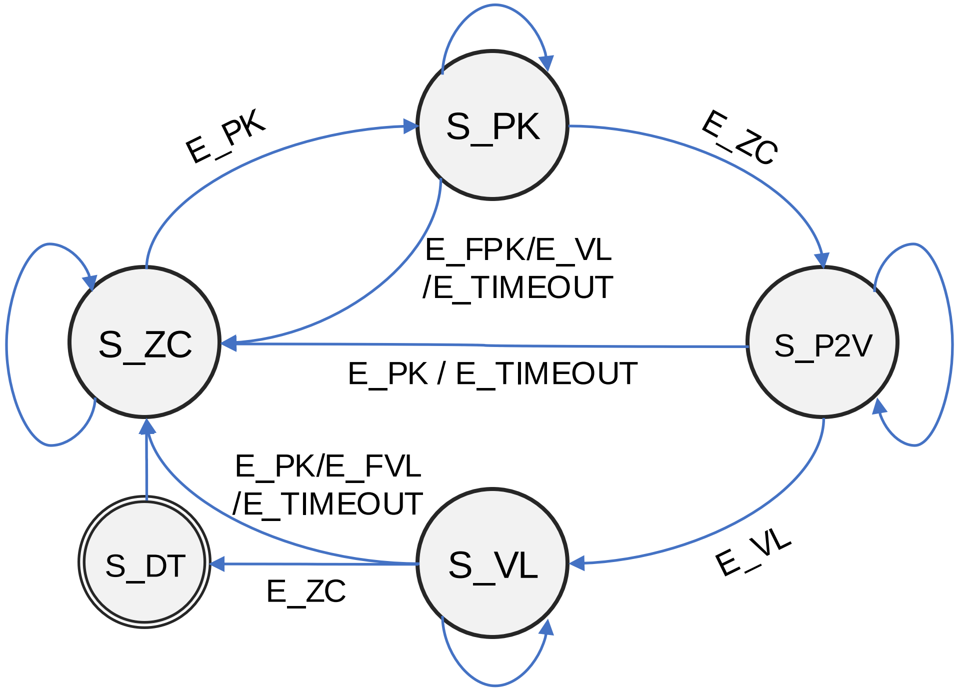


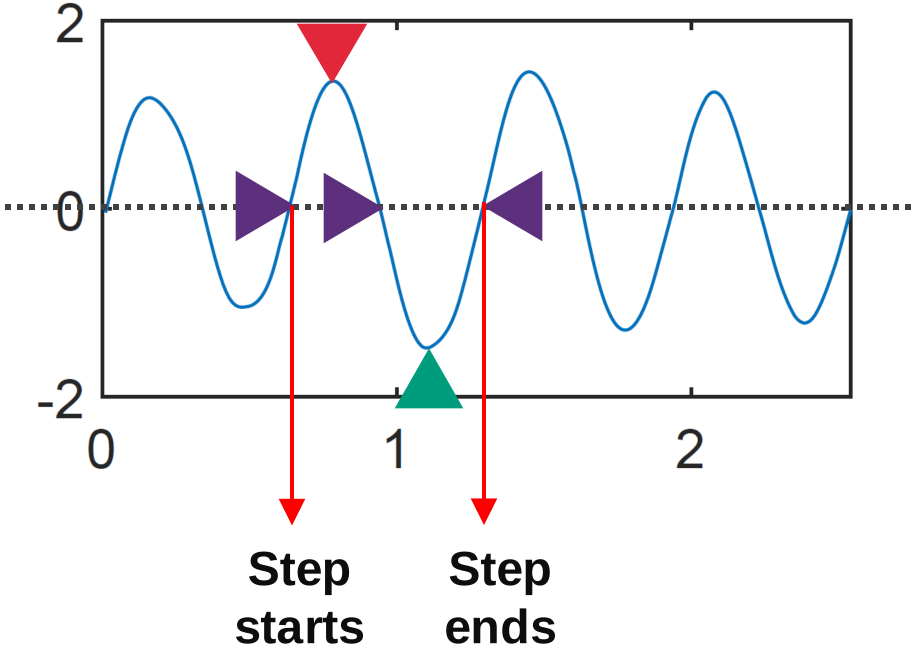
- Using a Finite State Machine (FSM) to characterize the acceleration transitions during a step cycle
- Transform raw accelerations into atomic events

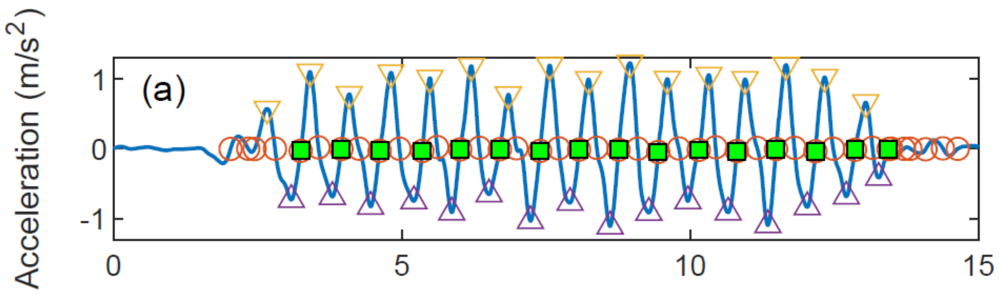




FSM Design





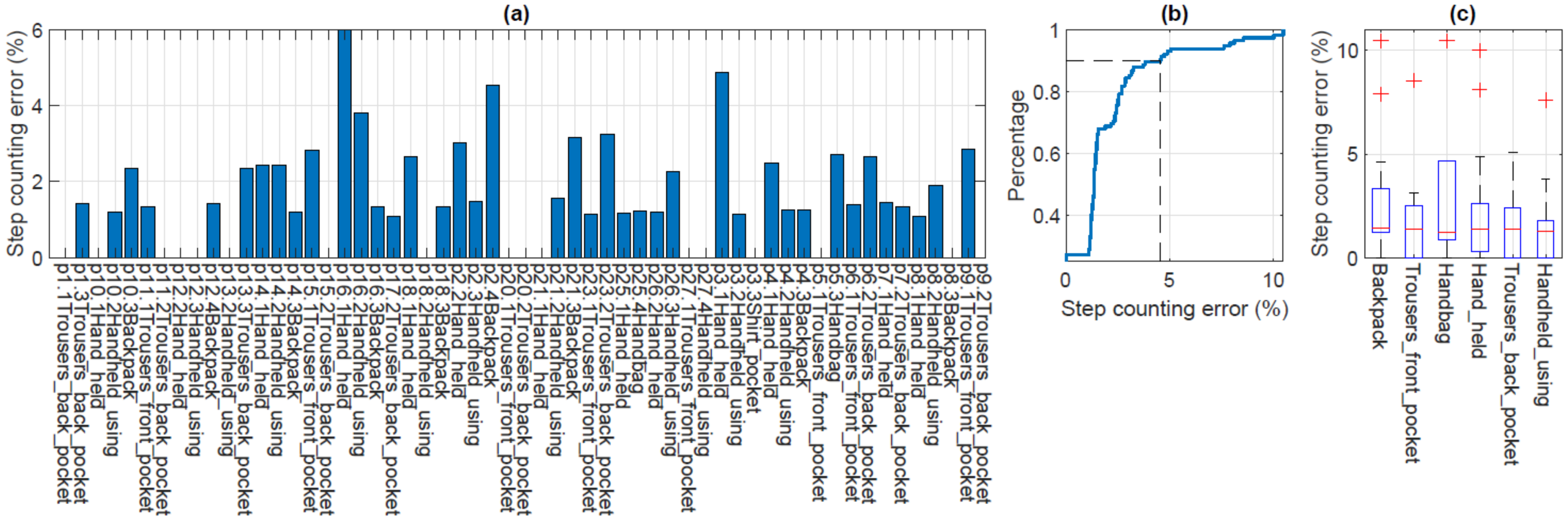


STATES:
S_ZC: Zero Crossing
S_PK: Peak State
S_P2V: ZC from Peak to Valley
S_VL: Valley State
S_DT: Step Detected

EVENTS:
E_PK: Peak detected
E_VL: Valley detected
E_ZC: Zero crossing
E_FPK/E_FVL: Far peak/valley
E_TIMEOUT: State timeout

Step Count Accuracy

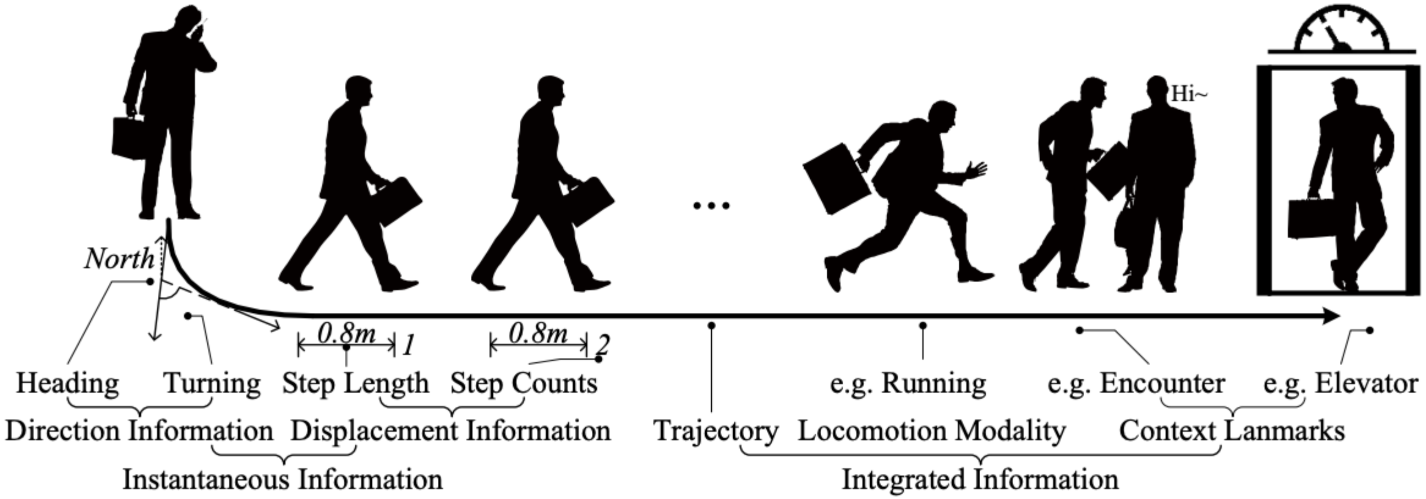
- Public dataset
  - The step counting errors for most of the traces are within 3%, while the 90%tile error is less than 5%, outperforming all the 9 different approaches evaluated on the same dataset [1].



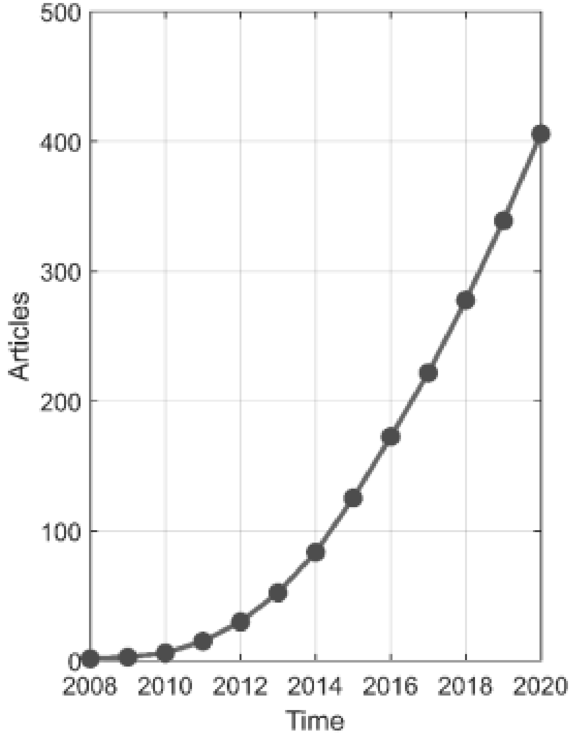
[1] A. Brajdic and R. Harle. Walk detection and step counting on unconstrained smartphones. In Proceedings of ACM UbiComp. ACM, 2013.

Beyond Step Count

- Human activities recognition (HAR)
- Inertial tracking (Dead-reckoning)



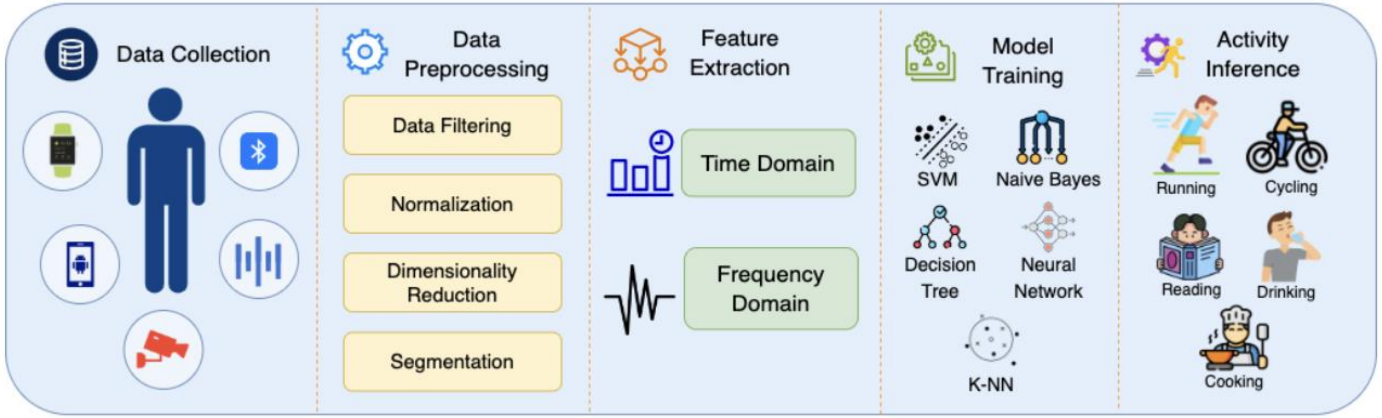
- *A general pipeline of processing?*



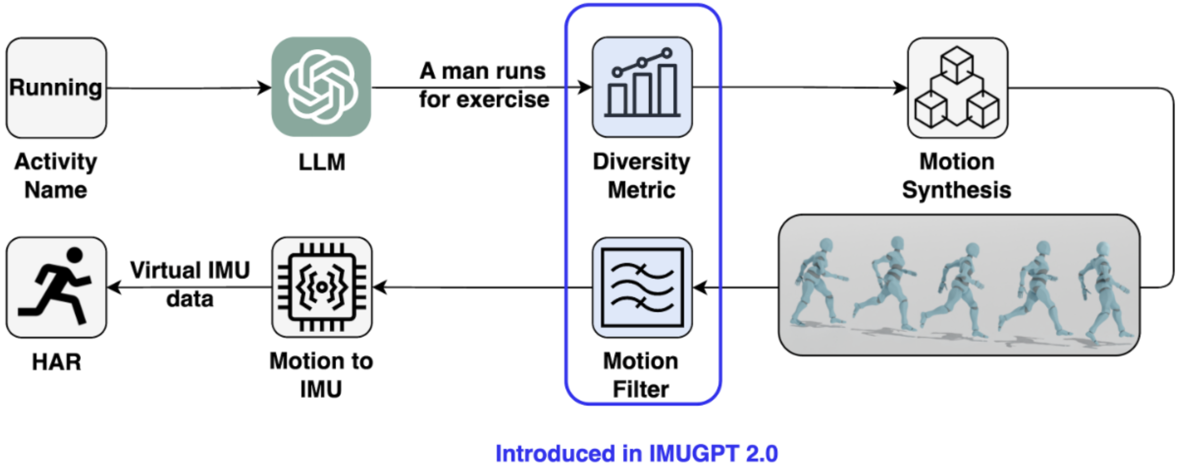
Cumulative number of peer-reviewed articles on human activity recognition (HAR) using smartphones published between January 2008 and December 2020, based on a search of PubMed, Scopus, and Web of Science databases

Human Activity Recognition Framework

Before LLM



After LLM



Sampling

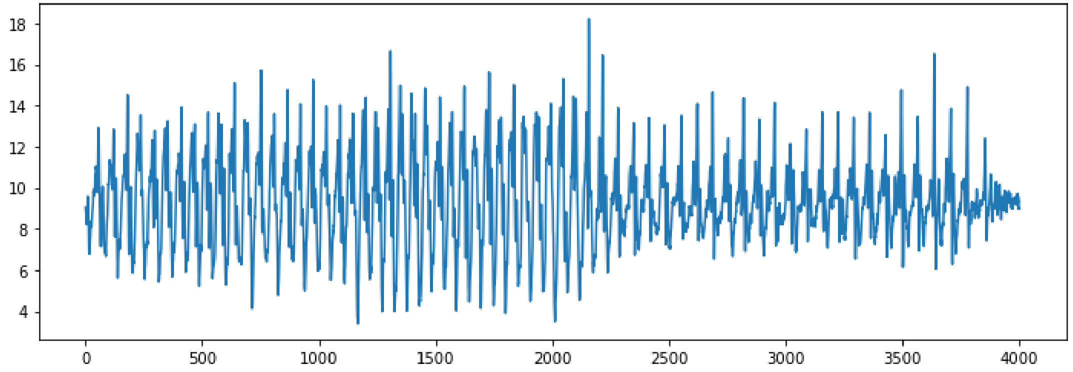
- Generally between 20 to 30 Hz
  - Can be 100 Hz or even higher, depending on the devices and OS

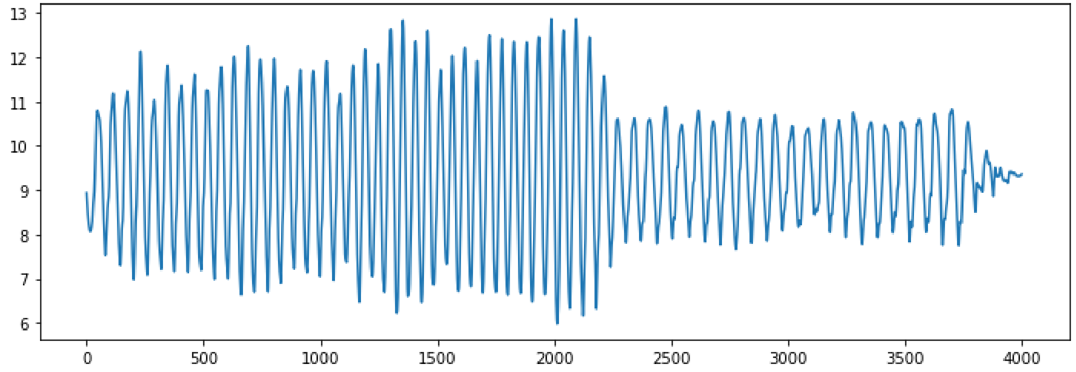
- Sampling rate affects responsiveness
  - (Android API specifications)
    - `SENSOR_DELAY_FASTEST`: best effort
    - `SENSOR_DELAY_GAME`: 20 ms (50 Hz)
    - `SENSOR_DELAY_NORMAL`: 200 ms (10 Hz)
    - `SENSOR_DELAY_UI`: 60 ms (~17 Hz)

- Sampling rate affects battery life
  - A lower sampling rate is preferred, especially for long-term, background data collection

Filtering

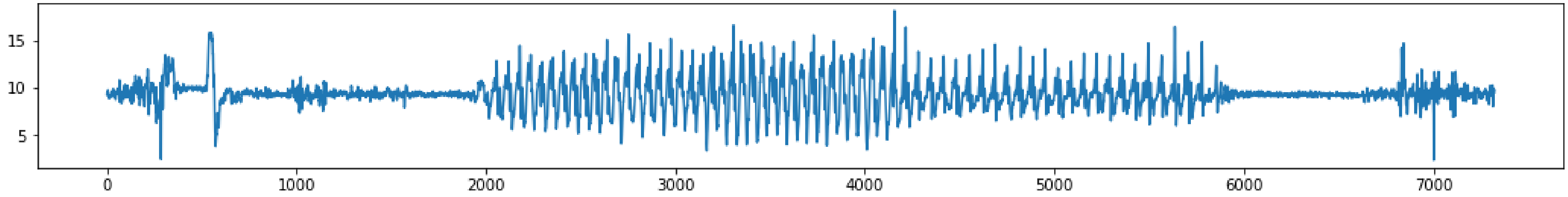
- Remove certain (unwanted, known) frequencies
  - “known”: Some prior knowledge of unwanted frequencies to design the filter
- Example
  - Low-pass filter: pass low frequencies and attenuate high frequencies
  - Band-pass filter: pass only frequencies in a frequency band





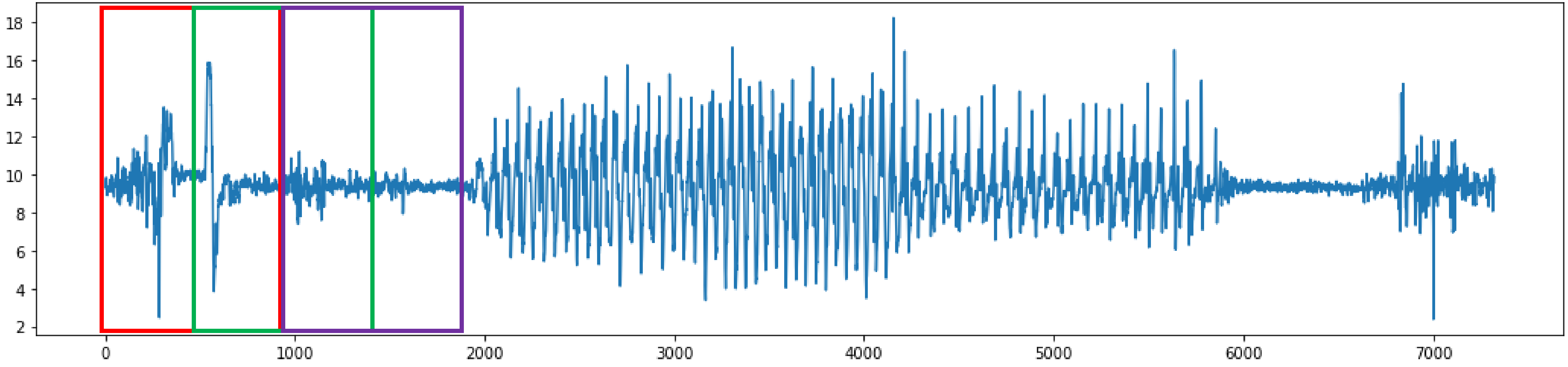
Windowing

- Pick a segment of the sensory time series for processing
  - How to localize some temporal patterns of interest?
  - Where to start/stop and what they are?
- Apply a window for Data Segmentation
  - Break down streaming data into segments for processing
  - Sliding window approach
    - Window length, overlap, window label…
    - Window-based preprocessing, global preprocessing



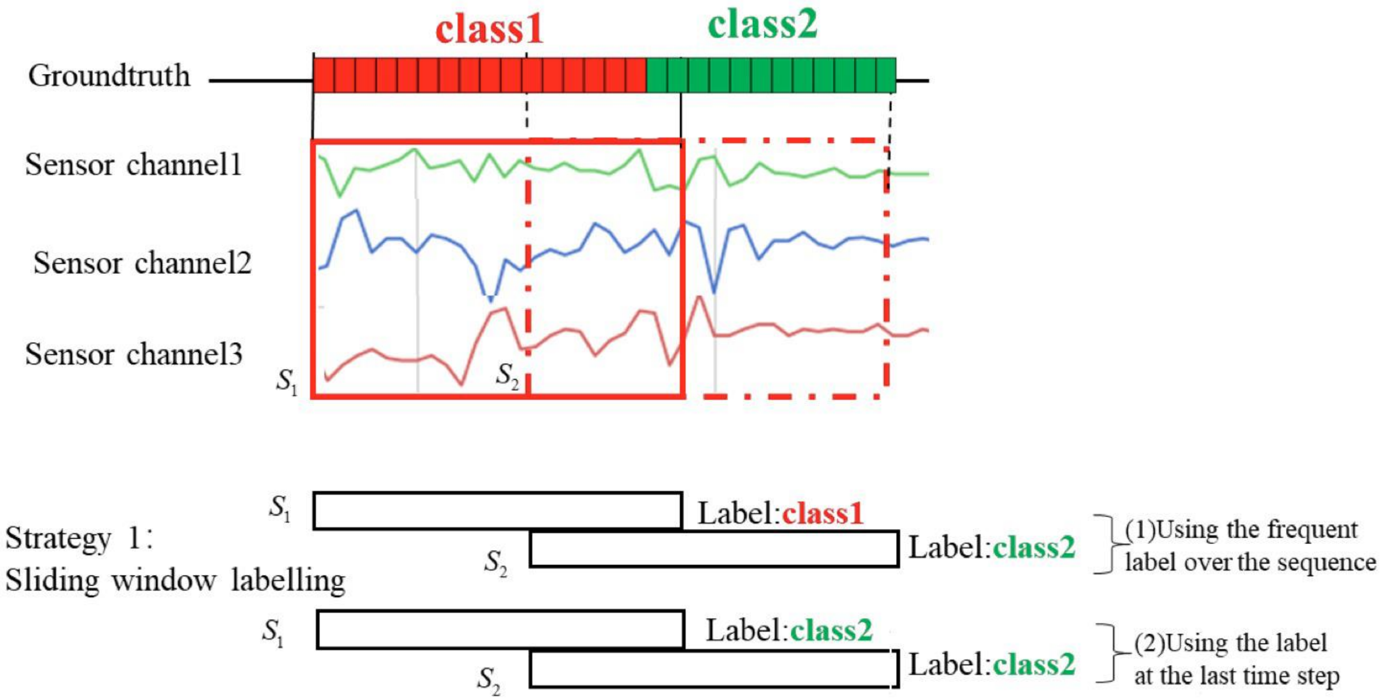
Windowing

- Choose a fix window size & an overlap (sliding step), e.g., 50%
- Preprocessing
  - Offline: global preprocessing
  - Online: window-based (no future data; past data discarded)



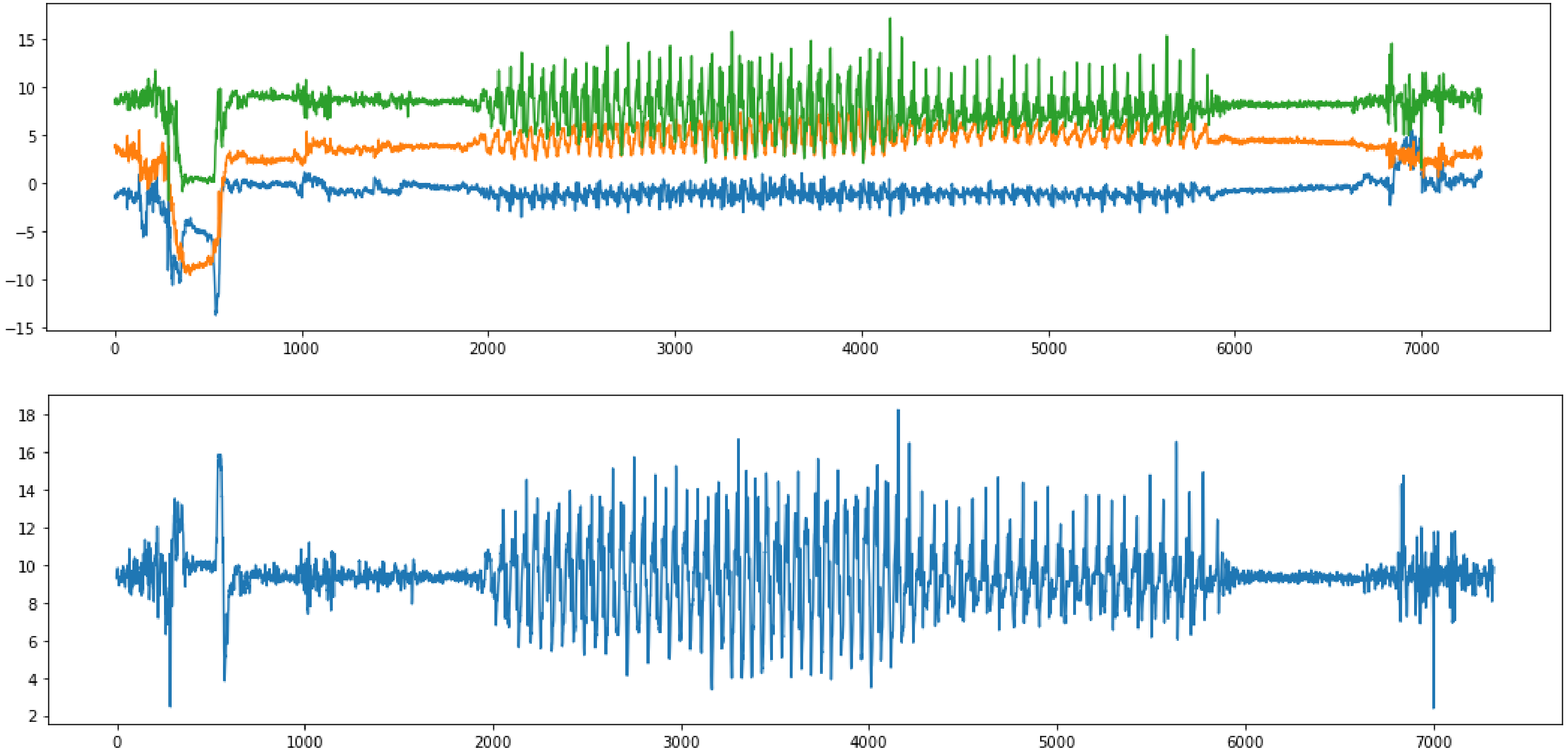
Labeling

- Mapping segments with classes



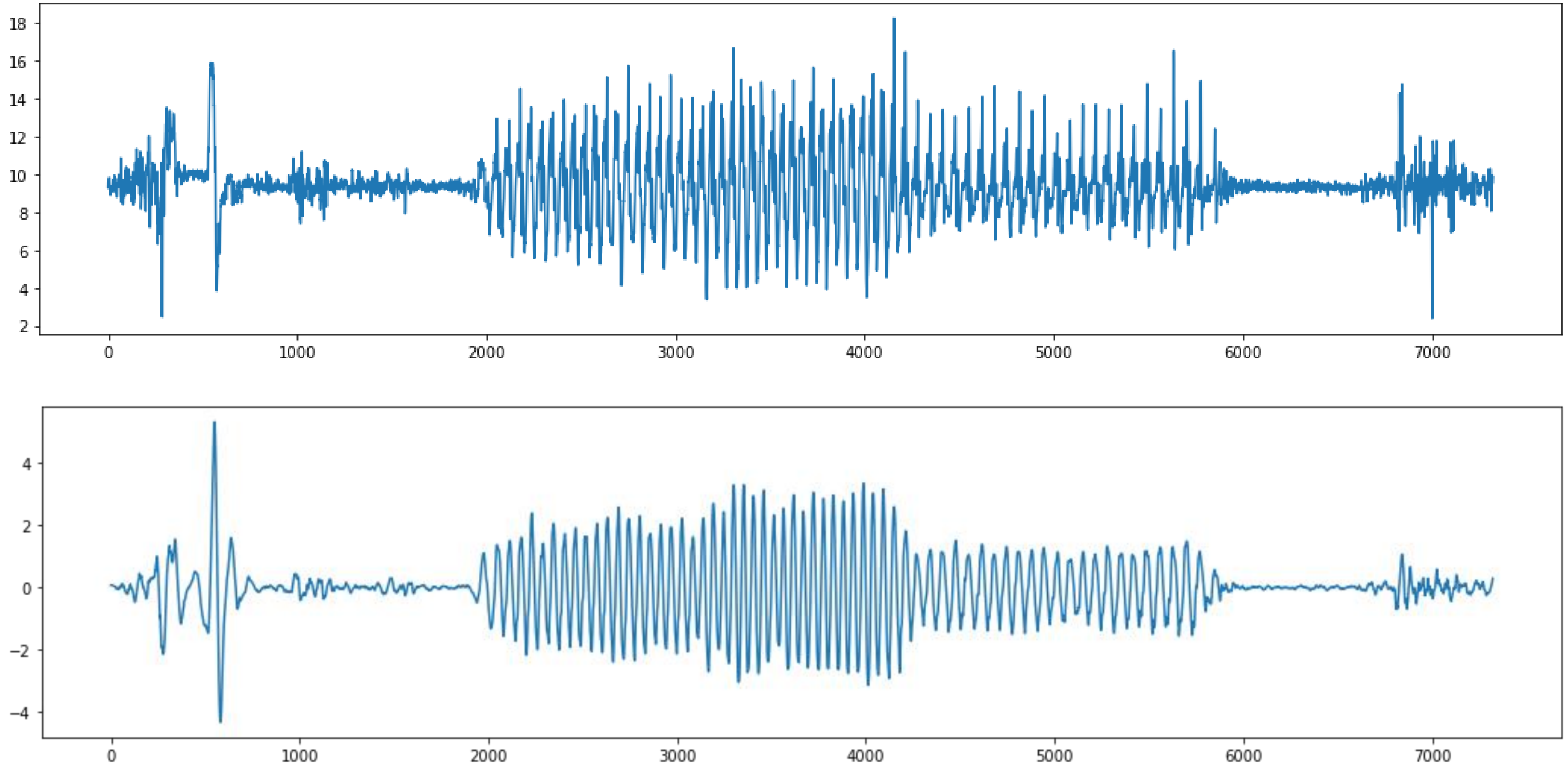
Coordinate Transformation

- (Where are) How to handle the 3-axis data?



$$\forall i : m_i = \sqrt{x^2 + y^2 + z^2}$$

Gravitational and Body Force Separation



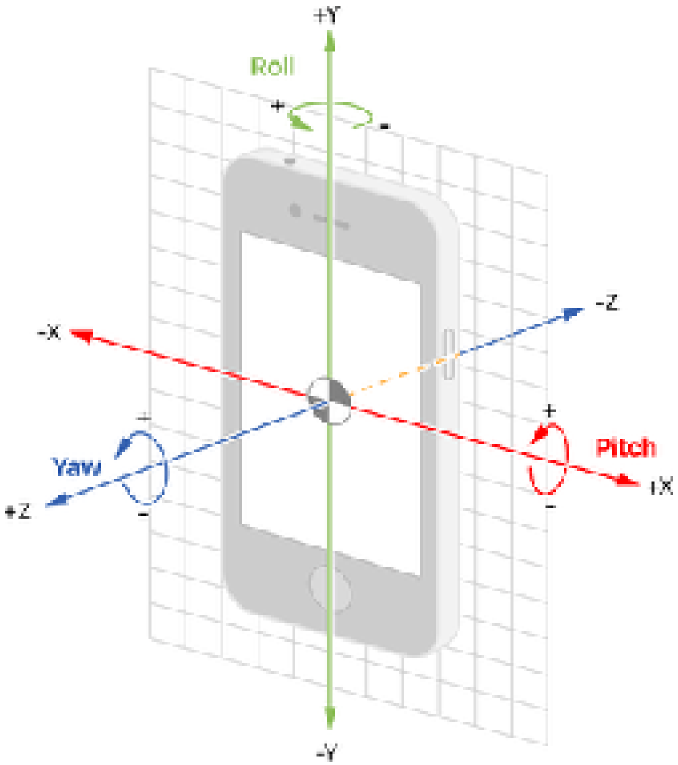
Gravity removal

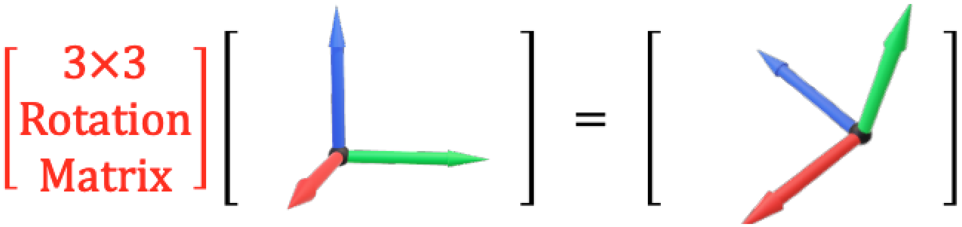
Coordinate Transformation

- Cartesian frame of reference (x,y,z)
- Rotations represented by Euler angles (yaw, pitch roll)

- Frames of reference
  - Local coordinate frame
  - Global coordinate frame

- Coordinate transformation
  - Rotation Matrix
  - (Euler angles)
  - (Quaternions)
  - Direct API available on mobile OS



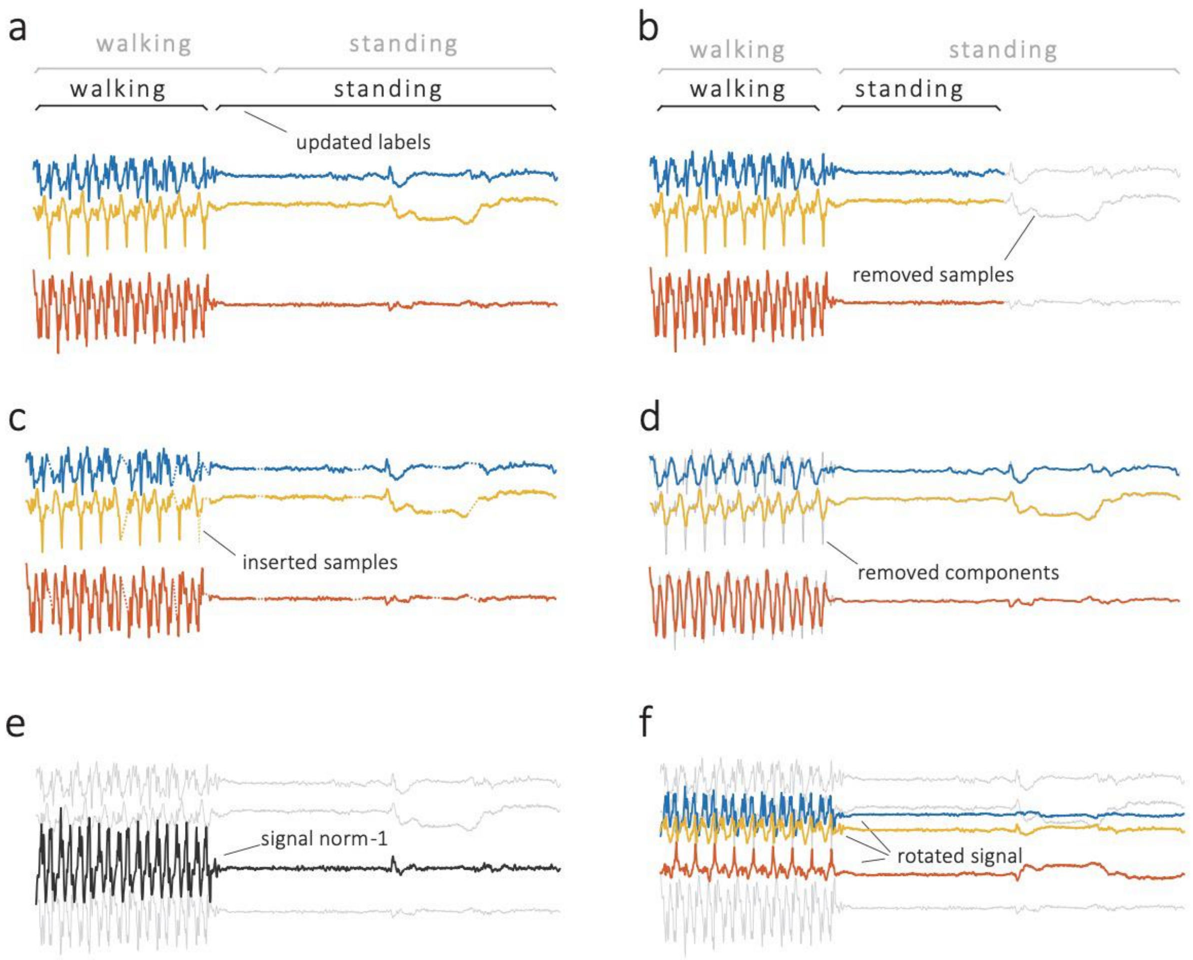


Preprocessing

A: labels are realigned (by eye)

C: missing data is filled with adjacent data

E: aggregation (magnitude)



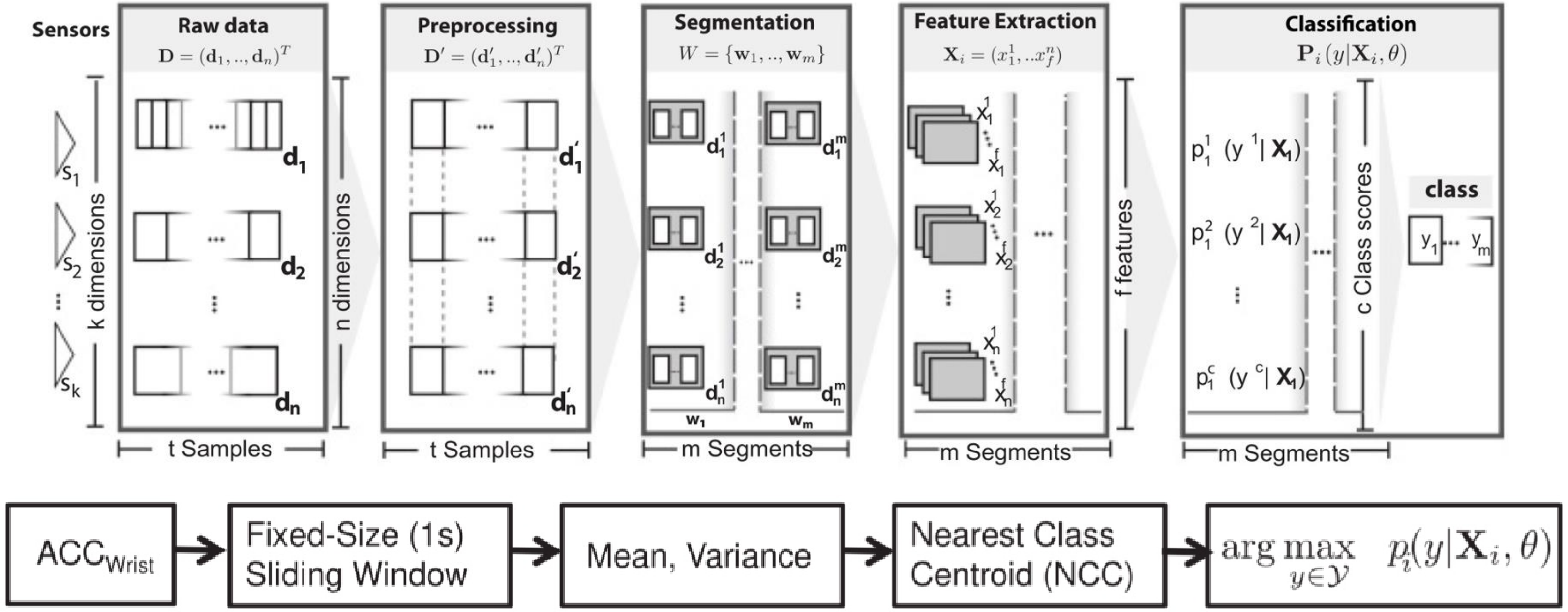
B: samples are removed to balance data

D: removing components, denoising: high frequency noise cancellation

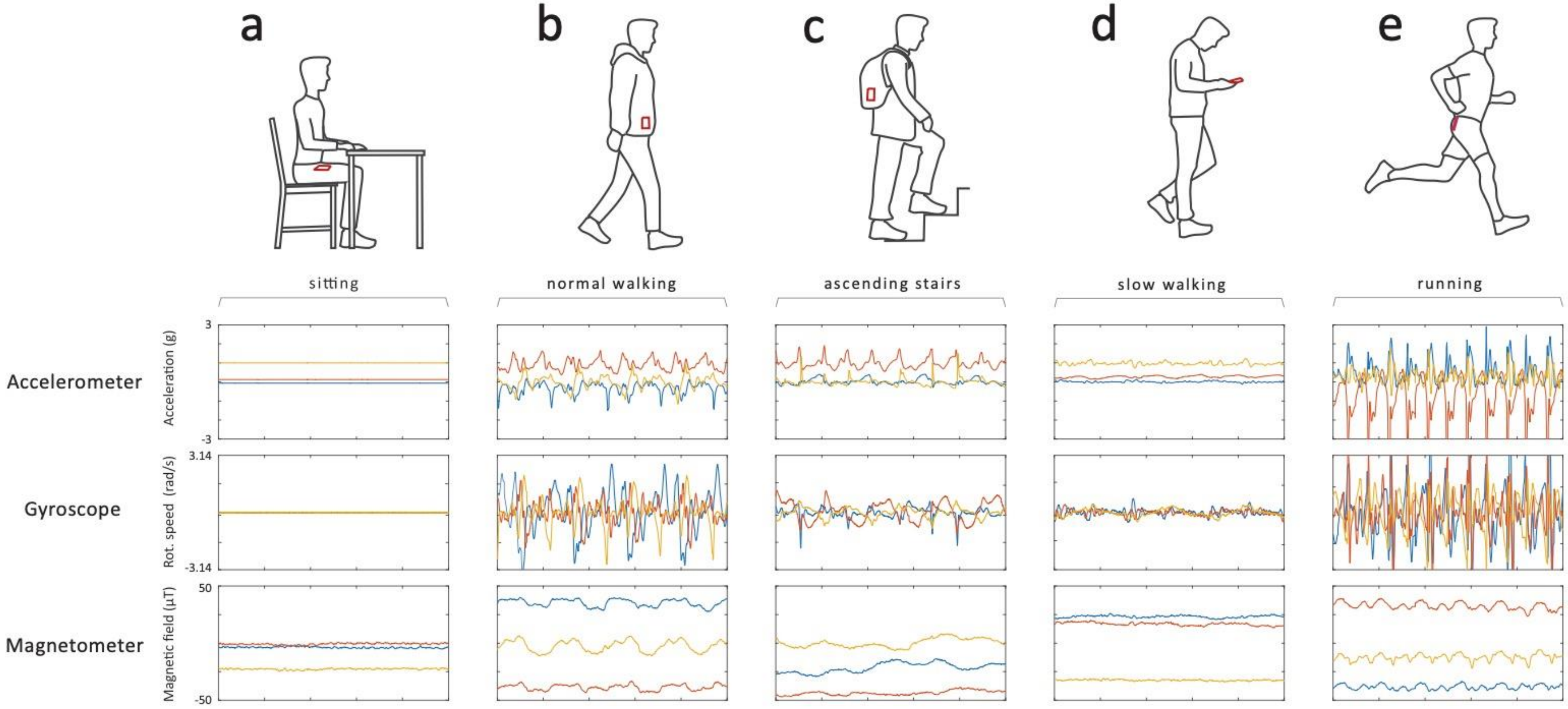
F: rotate to different coordinate system

Segmented Samples to Prediction

- For each sample further analysis is applied to reach a prediction, for example:
  - Extract features on a sample and build a classifier to use them to decide on the class label for a sample



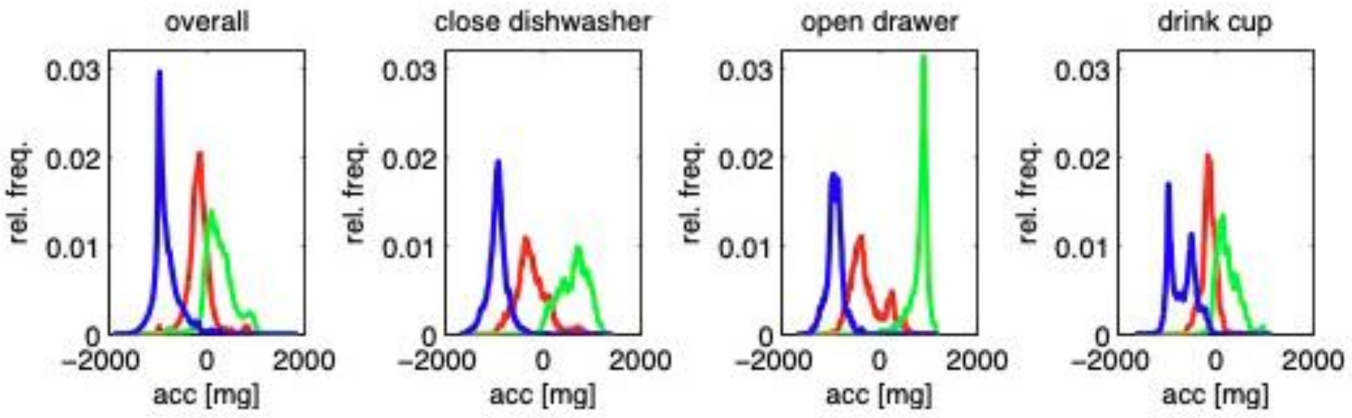
Human Activities Recognition



From A systematic review of smartphone-based human activity recognition methods for health research.
Marcin Straczkiewicz, Peter Jamess, Jukka- Pekka Onnela. Nature NPJ Digital Medicine.

Classification

- Feature extraction produces a feature vector
  - Feature engineering: Manual (and more explainable) features
  - Deep learning: Automatic (yet less/not explainable) features
- Classification: Match a feature vector to a pre-defined set of classes



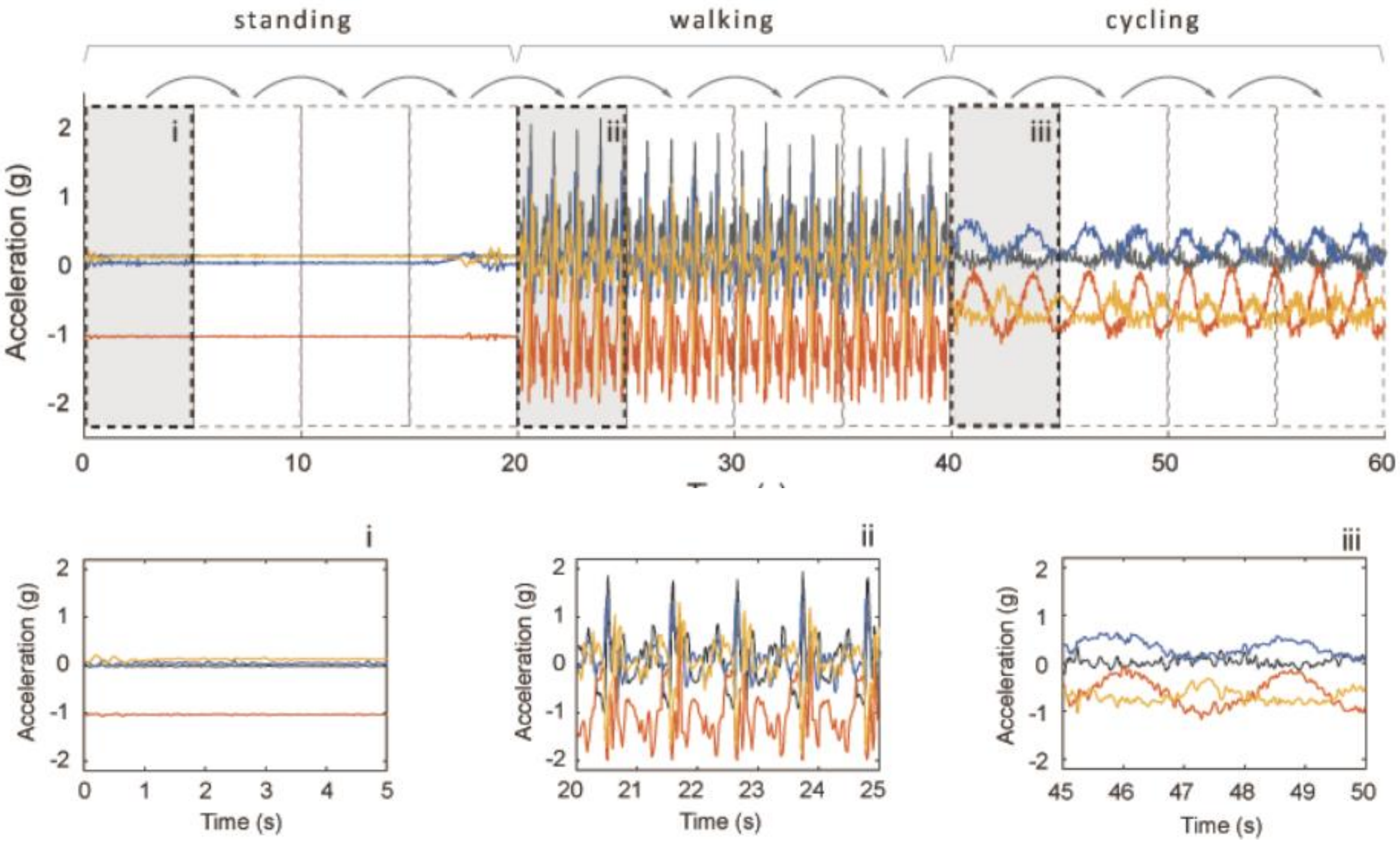
Distribution of x,y, z axis acceleration per window for various activities

Class Prediction Problem

- Predict activity given a window of movement data.
- Predict activity given multiple windows of movement data.
- Predict the activity sequence given multiple windows of movement data.
- Predict activity given a sequence of movement data for a presegmented activity.
- Predict activity cessation or transition given a window of movement data.
- Predict a stationary or non-stationary activity given a window of movement data

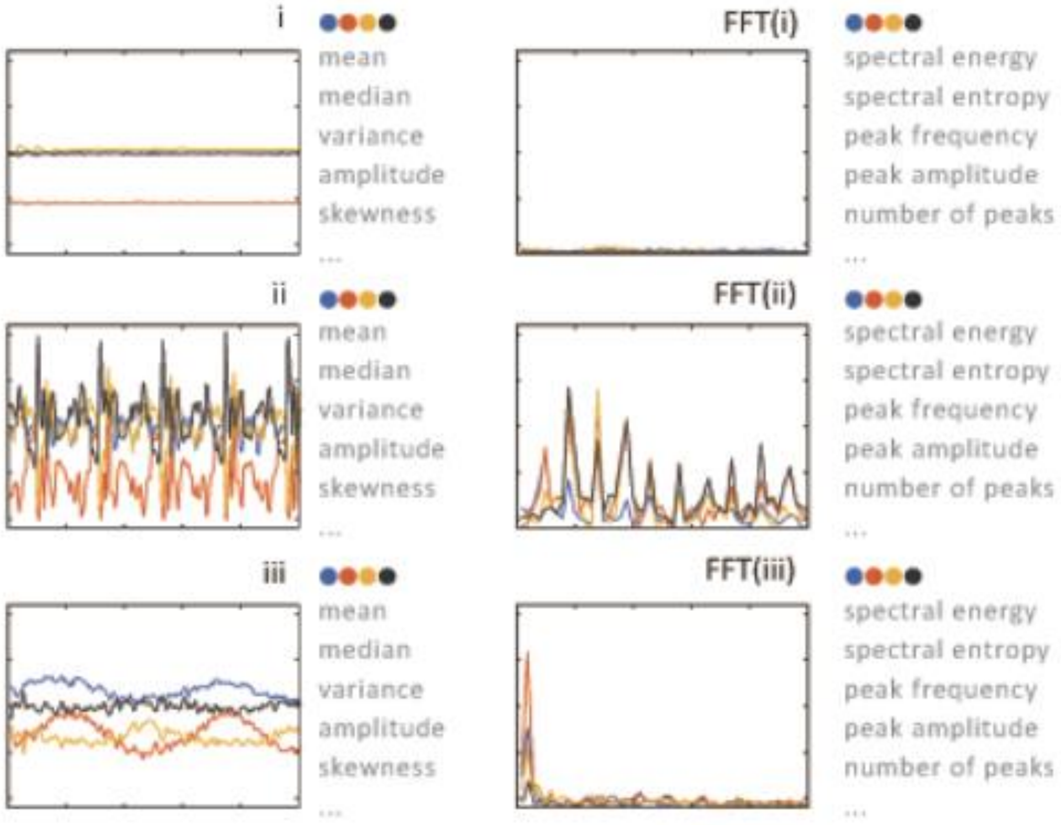
Step 1: Data Preparation

- Activities: sitting, standing, waking, cycling
- Preprocessing:
  - Filtered
  - (Coordinates transformed)
  - Gravity removed
  - Segmented



Step 2: Feature Extraction

- Feature Engineering
  - Mean (can help distinguish between standing and sitting)
  - Standard deviation
  - Number of peaks (can help distinguish between waking and running)
- More possible features
  - median, skewness, kurtosis, max, min…
  - Spectral energy/entropy, peak amplitude, #peaks, harmonic ratio…

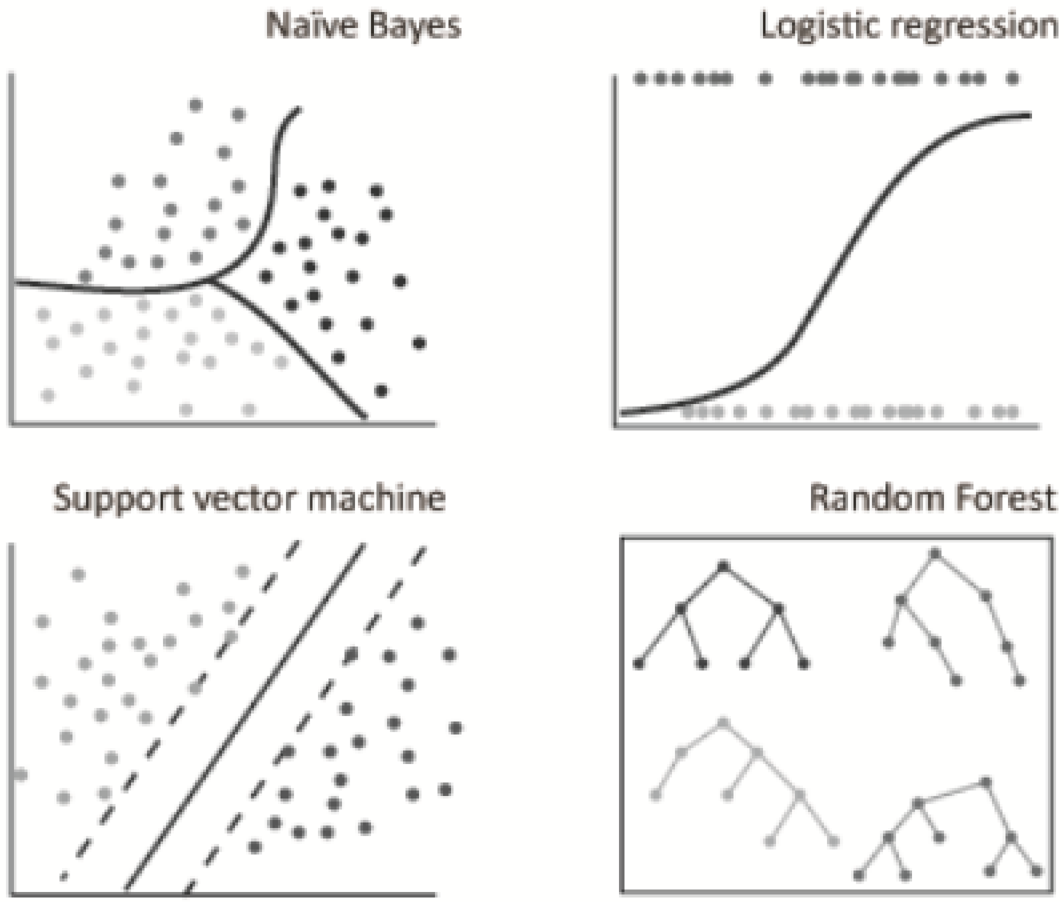


time-domain

frequency-domain

Step 3: Classification Model

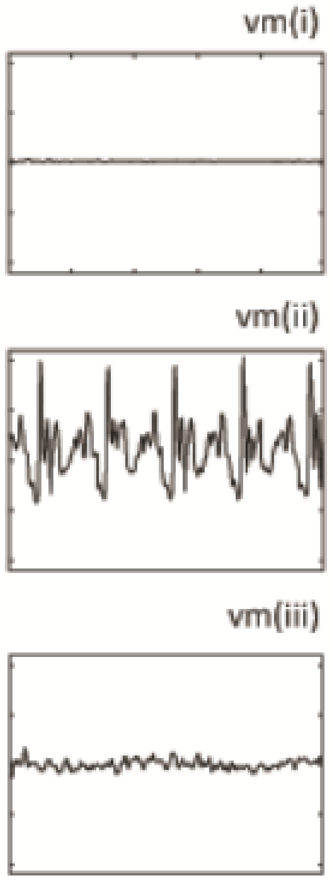
- Classical machine learning
  - K Nearest Neighbour
  - Naïve Bayes classifier
  - Decision Trees
  - Hidden Markov Models
  - Support Vector Machine



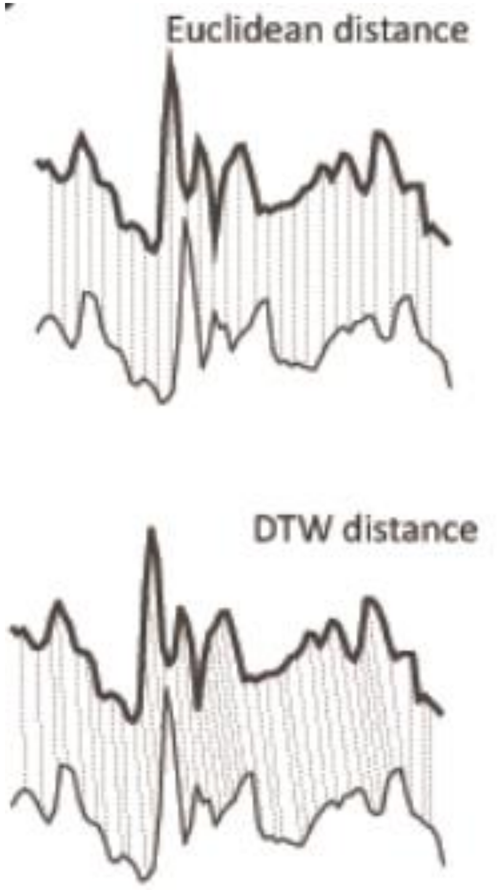
Step 2/3: Alternative (1)

- Segment serving as a template
  - May build from many segments/samples to reduce noise

- Template matching
  - Euclidean distance
  - Dynamic Time Warping (DTW)
  - EMD (Earth Mover's Distance)



Activity template

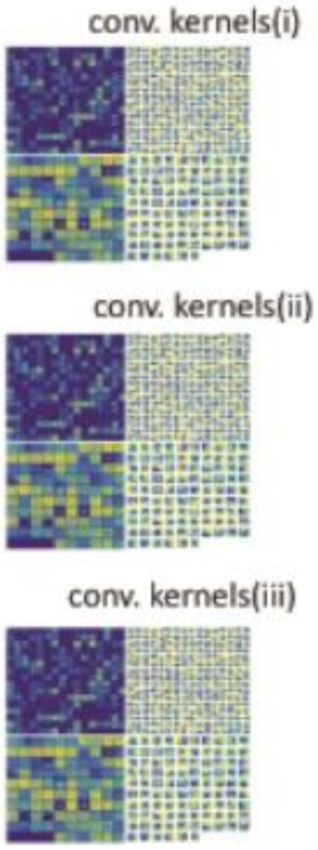


Template matching

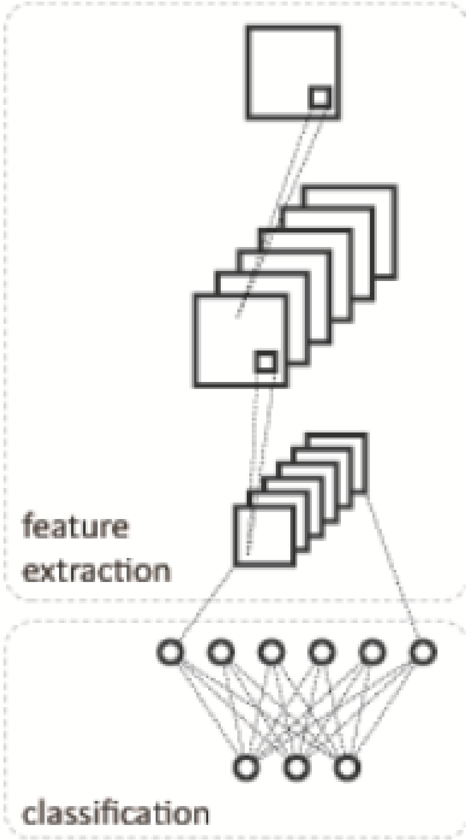
Step 2/3: Alternative (2)

- Deep neural networks for feature extraction
  - Input segments into deep learning networks that compute high-dimensional deep/hidden features

- Deep learning models
  - CNN, RNN, GAN, Transformer…



"Deep" features



Deep learning

An HAR Example: Gait Analysis

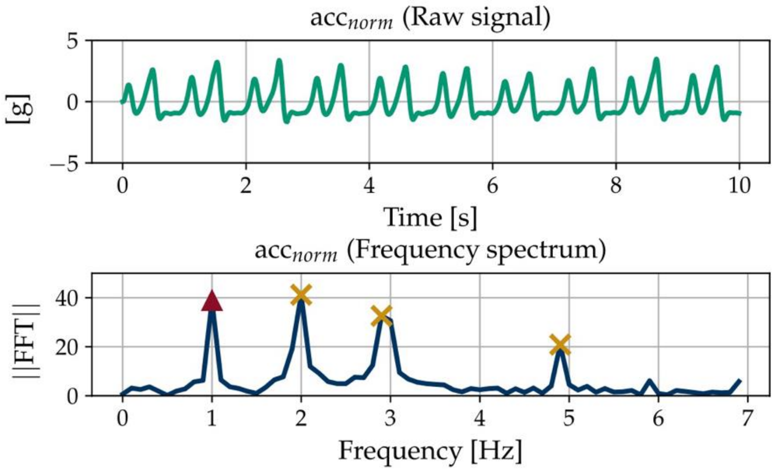
- Training set: 150 gait analysis recordings of 121 patients in hospital
- Validation set: 203 gait recordings from 7 PD patients at their home
- Exercises:
  - 1. 2x10 m walk with a break at the turning point (2x10m) 2)
  - 2. 4x10 m walk without stops at turning points
  - 3. (4x10m) 2-minute walk back and forth along a straight path of 25 m (2min)
  - 4. Tapping on the ground with the heel (heel)
  - 5. Tapping on the ground with heel and toes alternately (heel-toe)
  - 6. Circular movement of the foot (circling)

M. Ullrich, A. Kuderle, J. Hannink, S. Del Din, H. Gaßner, F. Marxreiter, J. Klucken, B. Eskofier,F. Kluge. Detection of Gait From Continuous Inertial Sensor Data Using Harmonic Frequencies. JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS. 2020.

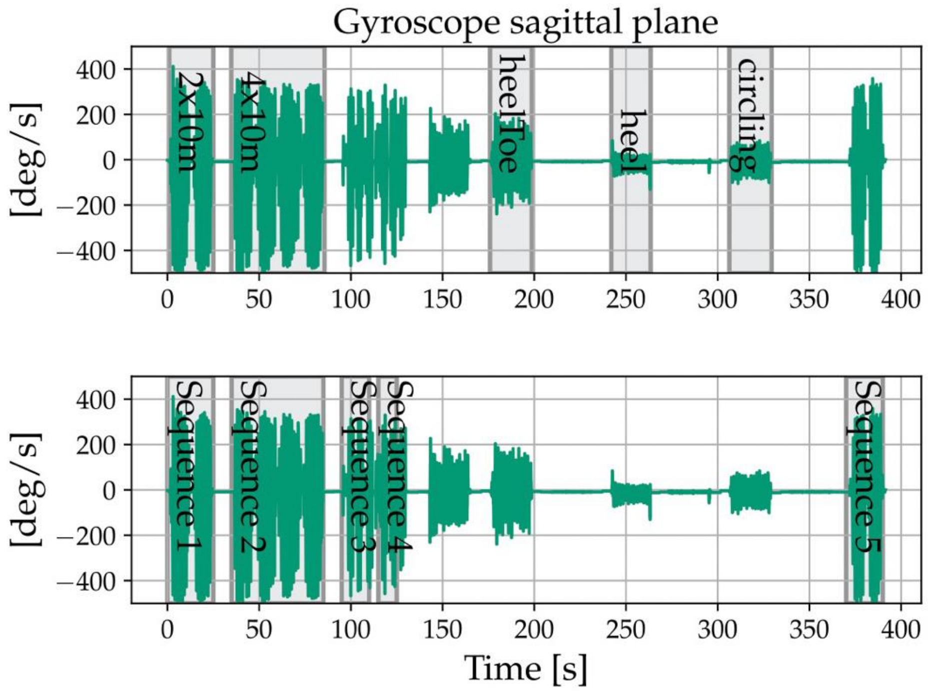
An HAR Example

- Norm of accelerometer and gyroscope for window used to detect movement. If above a threshold accept sequence.
- Low pass filter (cut off 6Hz).
- Use FFT to find important frequencies.
- Uses autocorrelation to measure peaks (and harmonic frequencies)
- Use these to decide if to keep this window.

$$|s^{3d}| = \sqrt{s_x^2 + s_y^2 + s_z^2},$$



Some results

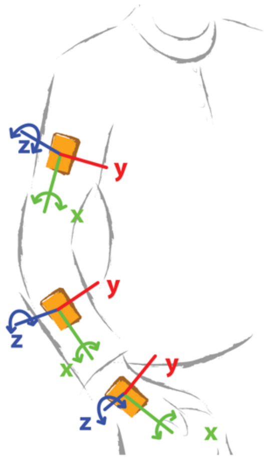


	acc_v	acc_{norm}	gyr_{ml}	gyr_{norm}
Lab Data Set				
Sensitivity	0.97 (0.03)	0.94 (0.04)	0.98 (0.01)	0.89 (0.04)
Specificity	0.95 (0.02)	0.96 (0.01)	0.96 (0.02)	0.81 (0.04)
Youden index	0.92 (0.02)	0.90 (0.04)	0.94 (0.01)	0.70 (0.06)
Opt. Peak Prom.	8	13	17	11
Val. Data Set				
Sensitivity	0.50	0.70	0.97	0.89

Sensitivity: TPR of detecting existing gait
Specificity: TNR regarding the rejection of non-gait movements
acc_v : Acceleration along the vertical axis
gyr_{ml} : Rate of rotation around the medio-lateral axis

Another Example

- Confusion matrix



Activities

opening a window
closing a window
watering a plant
turning book pages

drinking from a bottle
cutting with a knife
chopping with a knife
stirring in a bowl

forehand
backhand
and smash

classification (columns) / groundtruth (rows)

	NULL	Open window	Drink	Water plant	Close window	Cut	Chop	Stir	Book	Forehand	Backhand	Smash	*recall*
NULL	24267	216	444	3228	48	24	60	75	45		3		**85.42**
Open window	3849	1938	453	291	48	12	9		24				**29.26**
Drink	3984	927	3780	321	3	9							**41.89**
Water plant	3984	726	774	3735	21	57	15						**40.11**
Close window	3891	381	1173	945	1533								**19.35**
Cut	2940		264	450		6585	456		3				**61.55**
Chop	2895	168	435	153		909	5742		126				**55.06**
Stir	4947	39	135	42	21	474	561	4392	207				**40.60**
Book	4560	27	144	951		354	1725	60	6687				**46.09**
Forehand	3195	330		144	609	9	66		3	969	6	3	**18.17**
Backhand	3003	207	21		21	3	6	24	33		1302		**28.18**
Smash	1860	57		78	185		42	45		1567	137	230	**5.47**
precision	**38.29**	**38.64**	**49.59**	**36.13**	**61.59**	**78.06**	**66.14**	**95.56**	**93.81**	**38.21**	**89.92**	**98.71**	

LLM for Mobile Sensing

- LLMs Understand Sensor Data!
- LLMs Generate Sensor Data!

IMUGPT 2.0: Language-Based Cross Modality Transfer for Sensor-Based Human Activity Recognition

ZIKANG L
AMITRAJI
HRUDHAI
LIZHE ZH
ELIZABET
HYEOKHY
THOMAS

SensorLLM: Aligning Large Language Models with Motion Sensors for Human Activity Recognition

Zechen Li[1], **Shohreh Deldari**[1],
[1]University of New South Wales,
[2]University of Tokyo

Large Language Models Memorize Sensor Datasets! Implications on Human Activity Recognition Research

Harish Haresamud
Georgia Institute of Tech
Atlanta, USA

Nikhil Murlidhar Shan
Georgia Institute of Tech
Atlanta, USA

LLaSA: Large Multimodal Agent for Human Activity Analysis Through Wearable Sensors

Sheikh Asif Imran, Mohammad Nur Hossain Khan, Subrata Biswas, Bashima Islam
Worcester Polytechnic Institute
Worcester, USA
simran@wpi.edu, mkhan@wpi.edu, sbiswas@wpi.edu, bislam@wpi.edu

Foundation Models

Google Research Google DeepMind *2024-10-03*

Scaling Wearable Foundation Models

Girish Narayanswamy[°,1], Xin Liu[°,†,1], Kumar Ayush[1], Yuzhe Yang[1], Xuhai Xu[1], Shun Liao[1], Jake Garrison[1], Shyam Tailor[1], Jake Sunshine[1], Yun Liu[1], Tim Althoff[1], Shrikanth Narayanan[1], Pushmeet Kohli[2], Jiening Zhan[1], Mark Malhotra[1], Shwetak Patel[1], Samy Abdel-Ghaffar[1] and Daniel McDuff[†,1]
°Co-first, †Corresponding Author, [1]Google Research, [2]Google DeepMind

Wearable sensors have become ubiquitous thanks to a variety of health tracking features. The resulting continuous and longitudinal measurements from everyday life generate large volumes of data; however, making sense of these observations for scientific and actionable insights is non-trivial. Inspired by the empirical success of generative modeling, where large neural networks learn powerful representations from vast amounts of text, image, video, or audio data, we investigate the scaling properties of sensor foundation models across compute, data, and model size. Using a dataset of up to 40 million hours of in-situ heart rate, heart rate variability, electrodermal activity, accelerometer, skin temperature, and altimeter per-minute data from over 165,000 people, we create `LSM`, a multimodal foundation model built on the largest wearable-signals dataset with the most extensive range of sensor modalities to date. Our results establish the scaling laws of `LSM` for tasks such as imputation, interpolation and extrapolation, both across time and sensor modalities. Moreover, we highlight how `LSM` enables sample-efficient downstream learning for tasks like exercise and activity recognition.

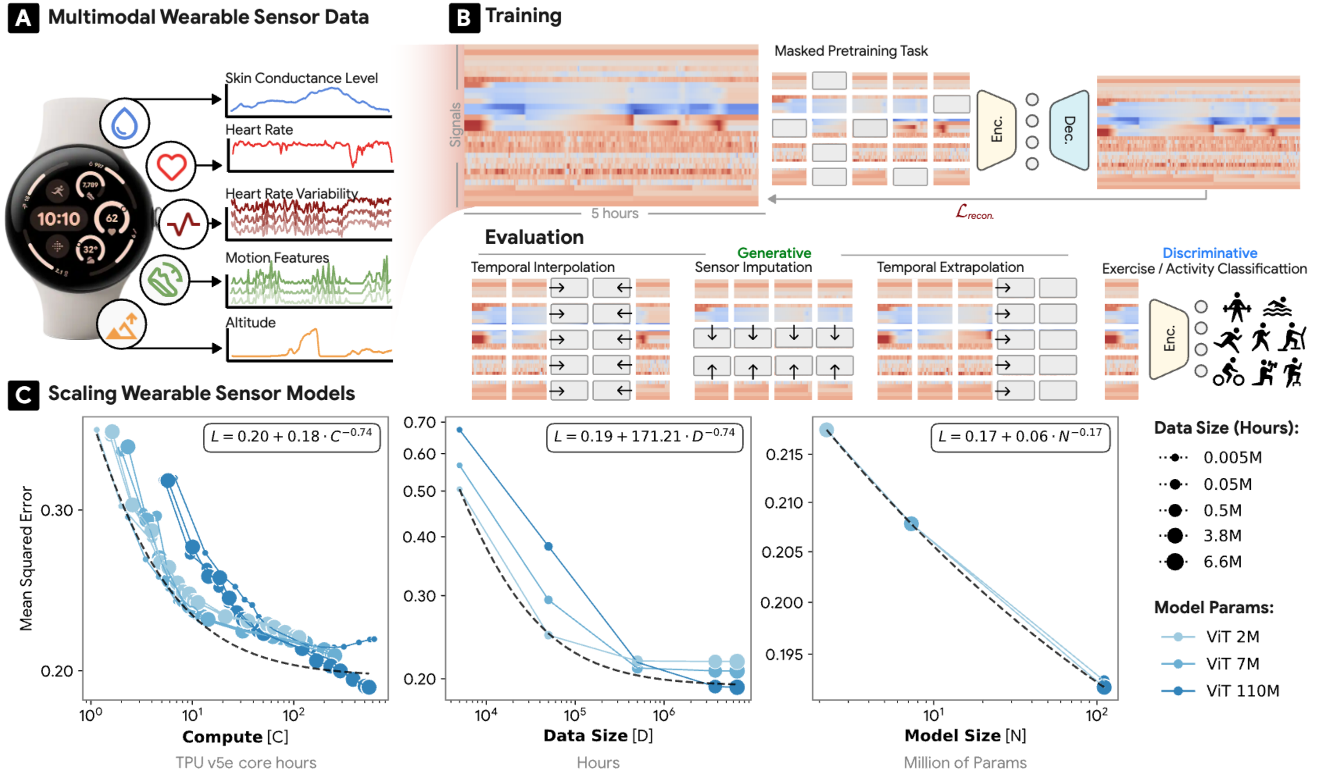


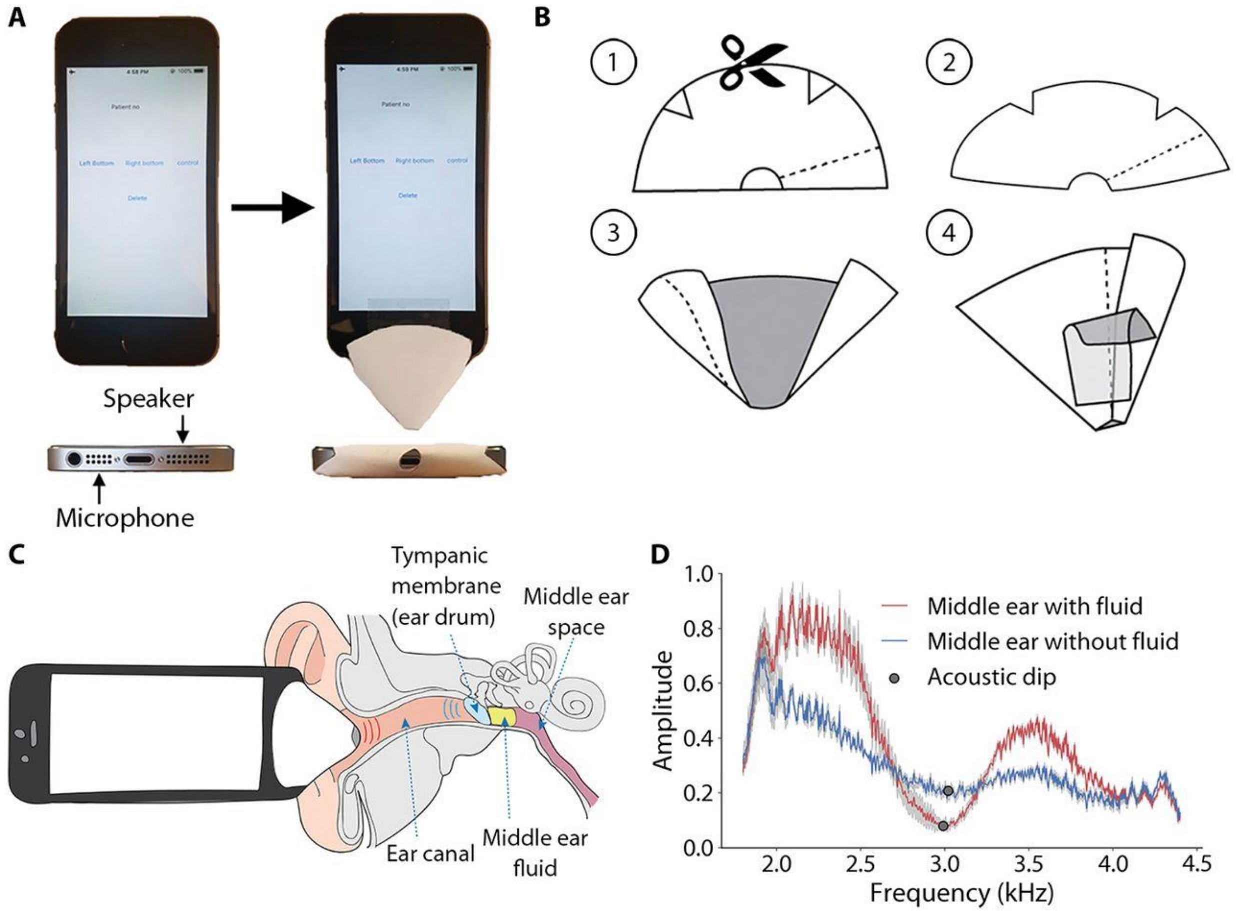
Figure 1 | **Scaling foundation models on wearable data.** Making sense of physiological and behavioral signals derived from wearables is challenging. **(A)** We present a systematic scaling analysis of sensor models using up to 40 million hours of multimodal data from over 165,000 people. **(B)** Using a random masking pretext task, we evaluate on tasks of imputation, forecasting, and downstream classification. **(C)** Experiments show scaling compute, data, and model size are all effective. Scaling is shown on the random imputation task.

More Applications of Mobile Sensing

- So many innovative applications, built upon a combination of various sensors

- Mobile Health
- Interaction
- Transportation

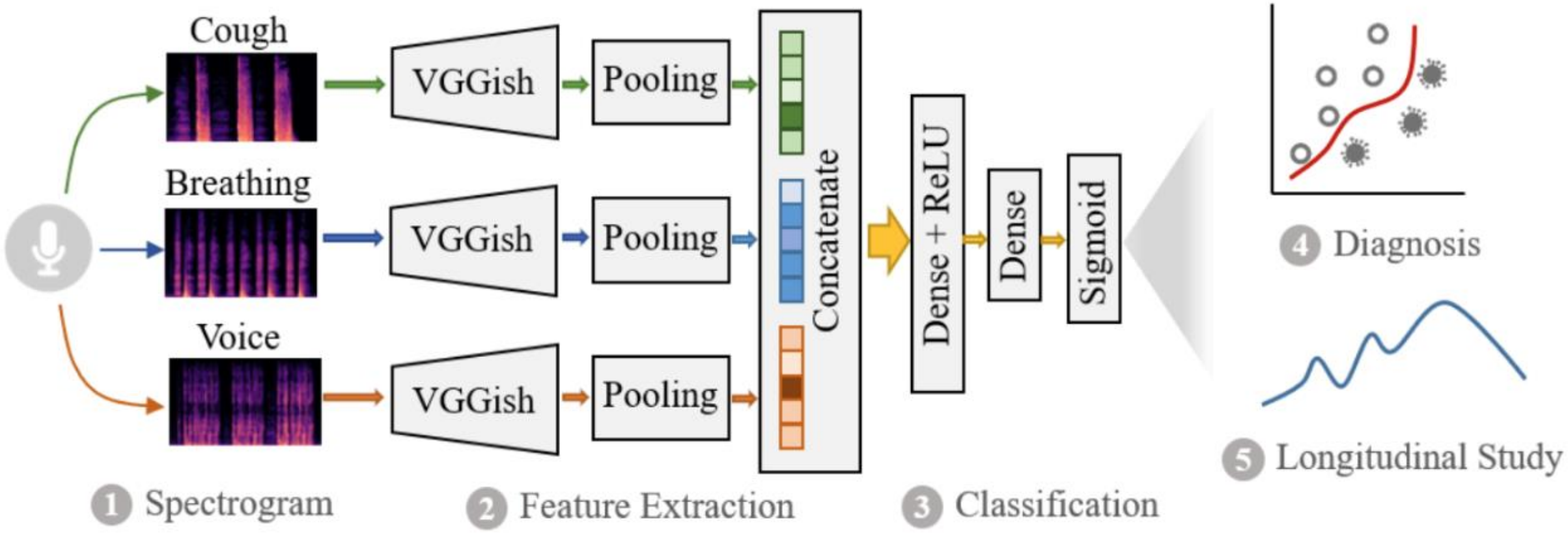
Middle ear fluid detection



Chan, Justin, et al. "Detecting middle ear fluid using smartphones." *Science translational medicine* 11.492 (2019): eaav1102.

COVID-19 Detection

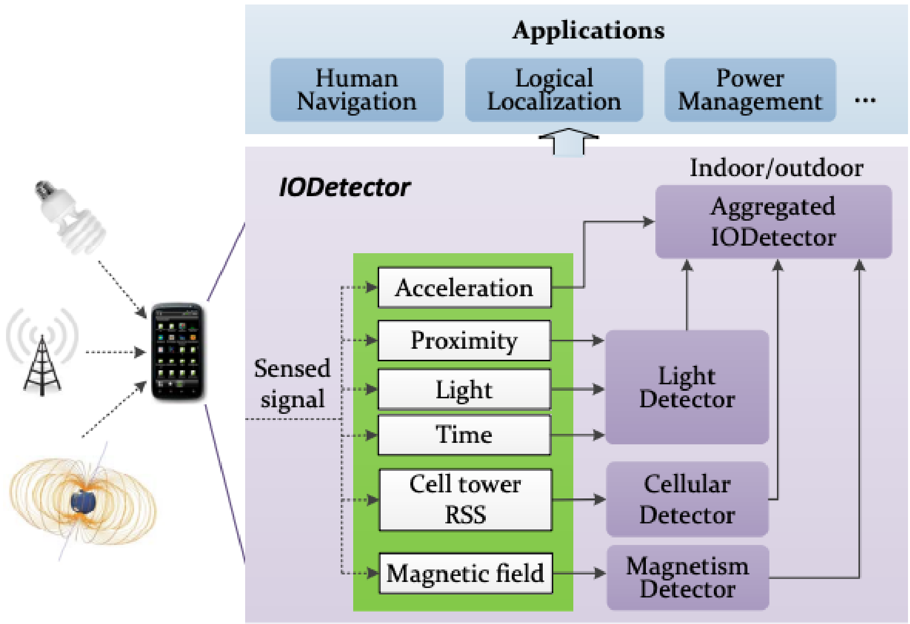
- Data analysis over a large-scale crowdsourced dataset of respiratory sounds collected to aid diagnosis of COVID-19.



Brown, Chloë, et al. "Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data." *arXiv preprint arXiv:2006.05919* (2020).

Indoor/Outdoor Detection

- IODetector: Indoor/Outdoor detection service
  - Adopted to support the real world business of on-demand food delivery.



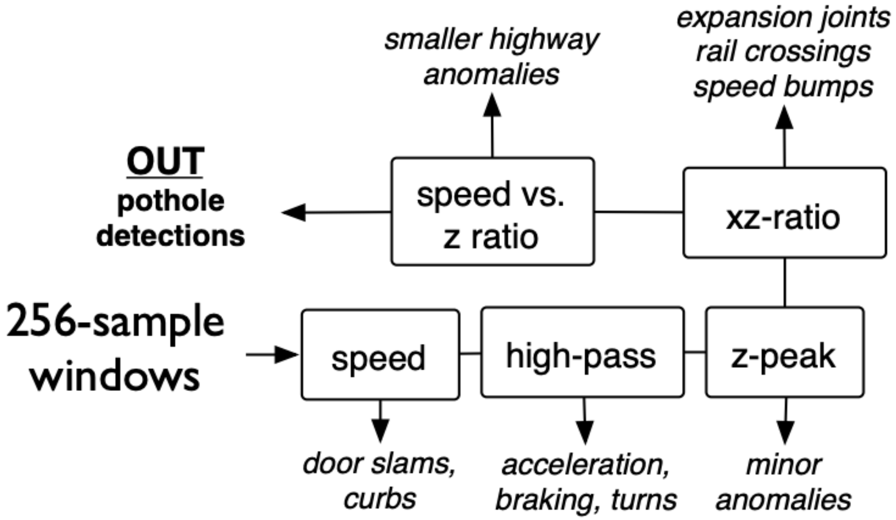
IODetector: A Generic Service for Indoor Outdoor Detection, SenSys'12

Experience: Adopting Indoor Outdoor Detection in On-demand Food Delivery Business, MobiCom'22

The Pothole Patrol

- uses the inherent mobility of the participating vehicles, opportunistically gathering data from vibration and GPS sensors, and processing the data to assess road surface conditions.

pothole v. not pothole



Eriksson, Jakob, et al. "The pothole patrol: using a mobile sensor network for road surface monitoring." *Proceedings of the 6th international conference on Mobile systems, applications, and services*. 2008.

Questions?

- Thank you!