Exercise 1: Wearable Sensors

The goal of this task is to extract as much information as possible from a relatively simple data trace acquired by wearable sensors. This involves the consideration of environmental information and creative approaches to combine them with the measured data.

You are given data acquired by wearable sensors worn by a male subject who undertook a day-trip to the popular tourist destination Jungfraujoch in Switzerland. The trip took place sometime between 01.12.2020 00:01 and 23.12.2020 23:59. Data was acquired by 4 wearable devices placed in the following locations:

1. Right above the ankle under a sock
2. On the wrist under a sweatband (in wrist watch position)
3. On the chest attached by medical tape
4. On the forehead under a headband (center between eyebrows and hairline)

The used devices are all identical and continuously measured air pressure at a constant frequency[1] from the time they were turned on until they were turned off at the end of the study. For Task 3 you will also be provided with data acquired by an accelerometer on the board that was worn on the chest.

[1] The four devices have independent real-time clocks (crystal: ECS-.327-7-16-C-TR) with a tolerance of ±20ppm. The sampling rate of the pressure sensor is derived from the real-time clock and therefore subject to the same tolerance which may lead to a drift of up to 144ms between any two devices for every hour of runtime. Additional tolerances due to temperature and aging of the component may increase this effect even more. You may ignore this, no compensation is needed to solve this exercise

1 Data

Note: All provided data is property of SIPLAB. You are not allowed to share it beyond the scope of this lecture. This data is not open-source.

1.1 Data for Task 1 and 2

Download the data here: https://polybox.ethz.ch/index.php/s/YuGYBwI1PBlHObP.
The file contains a pickled Python dictionary containing four numpy arrays (for each sensor) with the data trace of more than 400000 samples each. The measurements are integers representing 0.01 Pa (e.g. 7654321 represents 76543.21 Pa of air pressure). You may load and use the data with the code provided below. Please note that you are expected to use Python 3 in this exercise. You may download a Jupyter Notebook with all snippets here.

```
import numpy as np

# Load data
# .item() is needed as np.load() returns a structured array that needs
# to be converted back to a dict. File should be in same location as code.
data = np.load("ex1_data.dict.npy", allow_pickle=True).item()

print(data.keys())
print("Length of chest data: {}".format(len(data["chest"])))
print("12387th sample of wrist: {}".format(data["wrist"][12387]))
```

1.2 Data for Task 3 (additional Data)

Additional data for Task 3: https://polybox.ethz.ch/index.php/s/wBhGH5VxtqG4ccK
The file contains a pickled Python dictionary containing three numpy arrays (for each axis of the accelerometer) with the data trace of more than 4 Mio. samples each. The measurements are signed integers. The accelerometer has a full range of $\pm 2g$ and a resolution of 20 bit. This means a value of 2^{18} corresponds to $1g$. You may load and use the data as follows:

```
import numpy as np

# Load data
# .item() is needed as np.load() returns a structured array that needs
# to be converted back to a dict. File should be in same location as code.
imu_data = np.load("ex1_data_task3.dict.npy", allow_pickle=True).item()

print(imu_data.keys())
print("Length of trace of X-axis data: {}".format(len(imu_data["X"])))
print("12387th sample of Y-axis: {}".format(imu_data["Y"][12387]))
```

2 Tasks

The tasks for this exercise are split into parts. After the deadline of each part, we will publish relevant results to continue work. Thanks to this, you will be able to solve tasks that rely on earlier results even if you were not able to compute them.
Depending on your approach, you may find it easier to not solve all subtasks in order. Sometimes, an iterative approach may help improve results.

2.1 Part 1 (Deadline: 22.03.2021, 13:00)

1. What percentage of the total recorded time did the subject spend at Jungfraujoch (time between arrival and departure)? You may use the length of the shortest data trace as base value for 100%. (max. 1pt, Full pts tol ±0.2%, 10% Pts tol ±3%)

2. As it is not physically possible to turn on or off all devices at the same time, there is a misalignment between the data traces. Shorten the traces to match the device that was turned on last so that samples of all devices are synchronous by array index. Do this as precisely as possible. How many samples did you remove from the beginning each trace?[2] (max. 2pt, Full pts tol ±8 samples, 10% Pts tol ±50 samples)
The different times of turning off should not affect your work and may be ignored

3. What was the sampling rate of the sensors? (max. 2pt, Full pts tol $\pm 0.6 Hz$, 10% Pts tol $\pm 4.5 Hz$)

Update 22.03.2021
The deadline for Part 1 has passed. We recommend using the following results as a basis for the next two parts of the exercise to avoid error propagation:
Samples removed from each trace to synchronize:
chest: 0
wrist: 56
ankle: 174
head: 119
Sampling frequency of the sensors: 13.7 Hz
(each sensor independently samples at this frequency)

[2] In case you run into issues due to clock drift between devices, try to optimize synchronization close to the center of the data traces.

2.2 Part 2 (Deadline: 29.03.2021, 13:00)

This is published for your convenience, if you work on this before the deadline of Part 1, you may need to recompute your results with exact solutions to Part 1 which are provided after its deadline.

1. On what date did the subject conduct the trip to Jungfraujoch? (max. 2pt)

2. Which of the following towns was the starting location of the trip (start of the data trace)?(max. 1pt)

   - Kleine Scheidegg (46.58519, 7.96124)
   - Wengen
   - Grindelwald
   - Lauterbrunnen
   - Stechelberg
   - Interlaken

   All locations listed without coordinates are cities within Switzerland.

3. For which of the following locations is it possible and reasonable that the subject passed through while the sensors were active? (max. 3pt)
   Note: A short reasoning is required for each answer, random answers will be awarded 0 points.

   - Basel
   - Bern
   - Grindelwald
   - Interlaken
   - Kleine Scheidegg (46.58519, 7.96124)
   - Lauterbrunnen
   - Luzern
   - Wengen
   - Zürich

   All locations listed without coordinates are cities within Switzerland.

Update 29.03.2021
The deadline for Part 2 has passed. You may use the following information when solving the last part of the exercise:
Route: The subject started his journey at Grindelwald Terminal, taking the cablecar to Eigergletscher and from there the train to Jungfraujoch. The return journey was identical back to Grindelwald Terminal, from where the subject took a car and drove the fastest route to Zurich (over Brünigpass).
The journey did take place on 16.12.2020.

If you have not done so yet, it may help you to compute the elevation profile of the journey from the air pressure:
This can be done using a version of the barometric formula e.g. the pressure altitude. As we find out that the trip took place on 16.12.2020, we can find the sea-level-equivalent pressure of 1019.4Pa from the weather sation on Jungfraujoch. The pressure can therefore be calculated the following way:

$$elev. = 44307.69 * \left(\frac{p}{1019.4}\right)^{0.190284}$$

or in Python (you may need to adjust the scaling of p depending on pre-processing):

```python
elevation = dict()
for key in data:
    elevation[key] = [44307.69*(1-(p/10194000)**(0.190284)) for p in data[key]]
```

There are also formulas that take into account air temperature but for the tasks at hand, the above is sufficient.

2.3 Part 3 (Deadline: 12.04.2021, 13:00)

This is published for your convenience, if you work on this before the deadline of Part 2, you may need to recompute your results with exact solutions to Part 2 which are provided after its deadline. For this task, we additionally provide you with data from an accelerometer which recorded data on the device that the subject wore on his chest (refer to Section 1.2). You may use all data to solve this task.

1. What is the maximum measured vertical speed of the elevator leading up to the viewing platform of sphinx observatory on Jungfraujoch? Compute the value both for the ride up and down! (max. 2pt)
   For this task, the maximum speed is defined as the maximum average vertical velocity over a period of 3 seconds.

2. What is the sampling rate of the Accelerometer? (max. 1pt)

3. Analyze the walking activity of the subject.
   (a) What percentage of the time did the subject spend walking?[3] You may use the length of the accelerometer data trace as base value for 100%. (max. 3pt)

   (b) How many steps did the subject take throughout the day? (max. 3pt)

   Please assume that the subject was not walking while the first 335000 and the last 100000 samples of the accelerometer data was recorded. The subject taking the device on and off might distort your results otherwise.

[3] We define the state of "walking" as follows: The subject is walking at any time t at which he is involved in a stepping movement within the interval of [t-0.4s, t+0.4s]. This means, the subject is considered to stop walking 0.4s after he last placed his foot on the floor. However, you will be able to solve this exercise without using this exact definition. We only put it here to give you an idea of the expected resolution of your solution.

3 Submission and Grading

You must submit your results by the given deadlines on Moodle. Both team members must enter the solutions seperately. Teammembers may post different solutions if in disagreement. For each answer, there is a textarea in which you have to briefly explain your reasoning of how you computed your solution. Additionally, you need to submit the code (e.g. Jupyter Notebook) needed to arrive at your solutions. Solutions should be printed with some sort of label in the code. Assumptions and manual observations must be labelled as such with an in-line comment.
The submission deadline is a hard deadline, we cannot accept late submissions (this includes technical problems not caused by ETH). We suggest entering partial solutions early into the moodle quiz and latter updating of your answers.

3.1 Grading

In this exercise, only your answers to the actual questions will be graded. No points are awarded for wrong solutions but reasonable explanations. No points are awarded for correct solutions without coherent reasoning. The grading of the answer includes a tolerance where applicable and reasonable. Partial points may be awarded for answers outside of the tolerance window for some questions.
For all exercise work, points will be awarded throughout the semester and result in one grade which contributes 50% towards the total course grade.

3.2 Rules

ETH rules on graded work apply. This includes but is not limited to: You are not allowed to share any aspects of your work outside of your group (this includes ideas).

4 Remarks and Scope Limitations

- The data is raw and may include erroneous measurements. The measurements are subject sensor tolerances.
- You do not need additional information about the hardware (e.g. sensor type) to solve the given tasks.
- During his time at Jungfraujoch, the subject spent his time in the publicly accessible complex made up of an underground railway sation, the "Top of Europe" building and the Sphinx observatory. See Section 5.1 for wikipedia links for all these locations.
- The subject only relied on public transportation and/or car rides as means of transportation. All used means of transportation worked as planned (no major traffic jams, train delays, ...).
- The subject did not take voluntary detours or extended breaks on his way to and from Jungfraujoch. This excludes a time period at the beginning after turning the sensors on. During this period, the subject was in the process of mounting the sensors to the specified locations on his body. The subject was stationary during this process.
- The starting location and end location of the trip may not be identical.
- During the stay at Jungfraujoch, the Subject only engaged in activities during which he stood still, walked or sat in chairs/on benches.
- You may assume that all sensors of a device are activated at once (upon turning on) and deactivated at the same time again (upon turning the device off).

5 Tips and Tricks

5.1 Research Resources

To solve tasks of this exercise, you need to acquire additional information about external factors. All required information can be found online on websites with English language support.
Possible Resources are listed below. This list may not cover all needed resources and you may be able to solve the tasks without using these resources:

- General information about Jungfraujoch:
  https://en.wikipedia.org/wiki/Jungfraujoch
- Jungfraujoch railway station:
  https://en.wikipedia.org/wiki/Jungfraujoch_railway_station
- Top of Europe building at Jungfraujoch:
  https://en.wikipedia.org/wiki/Top_of_Europe
- Sphinx Observatory at Jungfraujoch:
  https://en.wikipedia.org/wiki/Sphinx_Observatory
- Topographical Maps of Switzerland:
  https://s.geo.admin.ch/8e79b774a7
- Resource to look up possible train and cable car schedules:
  https://www.sbb.ch/en/home.html
- Resource to plan and calculate car rides:
  http://maps.google.ch/
- Measurements from weather stations in Switzerland:
  https://www.meteoswiss.admin.ch/home/measurement-values.html

5.2 External Code and Algorithms

You may use any python package and algorithm that is publicly available. If you use algorithms that are not completely your own development, you must label your source with a comment in the code with a link to the resource.

5.3 Data Visualization

Data visualization is a very important aspect of data analysis. We recommend you to look at the data very thoroughly before implementing any algorithms. If you are able to solve tasks purely through visual inspection of the signal, that is totally fine (you just have to provide the reasoning of what you observed and what it implies).

Especially with data traces of several hundred thousand samples, interactive plots that allow responsive zooming can be very helpful. We recommend using Matplotlib in Jupyter Notebook or JupyterLab with the interactive features provided through ipympl. Pay attention to installing compatible versions according to the instructions on the linked ipympl page.

When having installed everything correctly, this sample code should get you an interactive plot with two subplots on a shared x-axis. You may use this as a starting point for your data visualizations. You may download a Jupyter Notebook with all snippets here.

```
import matplotlib.pyplot as plt
%matplotlib widget
fig,ax=plt.subplots(nrows=2,ncols=1,figsize=(10, 6), sharex=True)

for key in data:
    ax[0].plot(data[key], label=key)
ax[0].legend(loc='lower right')

# you can plot to the other subplot using this pattern
#ax[1].plot(...)

#programmatic zoom options
#ax[0].set_xlim([123, 543])
#ax[0].set_ylim([9000000, 9500000])

plt.tight_layout()
plt.show()
```

6 Any Questions?

Please ask any questions regarding this work in the public Teams channel of the lecture. Make sure to not give away parts of the solution when asking a question. If you have a question that does include parts of a solution or include ideas on how to solve a task, you may ask through E-Mail. Send your E-Mail to manuel.meier@inf.ethz.ch, the subject must start with *[Ubicomp21 Exercise]*.

7 Changelog

Here we will keep a changelog of this document. Corrections may be released during the exercise. Please check the Teams channel to stay up-to-date.

08.03.2021 original release

18.03.2021 Corrected all the deadlines (it's on monday instead of sunday. All got shifted by 24h). Added the grading tolerances for Part1. Change in definition of maximum Speed in Part 3: It's the "maximum average speed over a period of 3 seconds" and not "at least 3 seconds".

21.03.2021 Corrected error in the previous edit: The deadline for Part three is on Monday, 12th instead of 13th.

22.03.2021 Added partial results for Part 1 to avoid error propagation.

29.03.2021 Added partial results for Part 2 to avoid error propagation. Added a bullet point in chp. 4 that all sensors of a device are turned on and off simultaneously.