Mobile Health

Lecture 4
Photoplethysmography (PPG) and Mobile Health (Part 2)

Cecilia Mascolo
PPG and Sleep

• Heart and respiration are indicative of different sleep stages.

• These, as we have seen, can be characterized with PPG.

• Also audio and movement differ in various phases of sleep (see upcoming audio lectures).
PSG: Sleep Monitoring
Gold Standard

- Polysomnography (PSG) is a multi-sensor approach
  - electroencephalography (EEG),
  - electromyography (EMG)
  - electrooculography (EOG)
- Together facilitate the measurement of brain activity, alongside both muscle and eye movement.
- Measurements of respiratory and cardiac activity are also often included.
Alternatives...

• Basic measures of sleep eg duration, awake episodes, sleeping pattern, bedtime routine, perception after sleep.
• These could be asked through a questionnaire/sleep diary.
• Or
• Smartphones basic features:
  • Has the phone been used (when).
  • Accelerometer of phone placed on mattress can measure movements.
  • Microphone (for sleep apnea).
• Under mattress (acceleration/pressure).
• Contactless Radio (see upcoming lecture).
Sleep Epochs

• There are up to six sleep stages:
  • awake;
  • rapid eye movement sleep (REM);
  • non-rapid eye movement (Non-REM);
  • sleep stage NREM 1 (N1);
  • sleep stage NREM 2 (N2);
  • sleep stage NREM 3 (N3).

• Each night has 4-6 cycles (90mins)
Sleep Apnea

- Sleep Apnea occurs when throat muscles relax and block air flow to the lungs.

- $\tau$ is a measure of IBI (inter beat interval)

---

PPG!

• Wearables with accelerometers and PPG can measure motion as well as Heart Rate and PRW (pulse rate variability).

• PPG-based wearables identify wake and sleep with a performance similar to, or better than, research-grade actigraphy (accelerometer based) devices.

• Sleep stages monitoring is still accurate enough...
Sleep Tracking with PPG
## PPG on wearables vs PSG

<table>
<thead>
<tr>
<th>Wearable</th>
<th>n</th>
<th>PSG</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TST</td>
<td>Light sleep</td>
</tr>
<tr>
<td>Mi Band2</td>
<td>55</td>
<td>r = 0.367**</td>
<td>r = 0.032</td>
</tr>
<tr>
<td>Gear Fit2</td>
<td>54</td>
<td>r = 0.307*</td>
<td>ICC = 0.024</td>
</tr>
<tr>
<td>Fitbit Alta HR</td>
<td>61</td>
<td>r = 0.466***</td>
<td>r = 0.179</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICC = 0.205</td>
<td>ICC = -0.19</td>
</tr>
</tbody>
</table>

Light sleep: sum of N1 and N2 sleep
Deep sleep: N3 sleep
TST: Total sleep time
REM: Rapid eye movement
WASO: Wake time after sleep onset
SE: Sleep efficiency
r: Spearman’s rho
ICC: Intraclass correlation coefficient
* p < 0.05, ** p ≤ 0.01, *** p ≤ 0.001

The deep sleep duration, REM sleep duration, and WASO measured by the Fitbit Alta HR significantly correlated with PSG results, although Fitbit Alta HR underestimated the duration of light sleep when compared to PSG (253 min vs 287 min).

Kim, K., Park, DY., Song, Y.J. et al. Consumer-grade sleep trackers are still not up to par compared to polysomnography. Sleep Breath 26, 1573–1582 (2022).
How do we calculate Sleep Stages from PPG?
An example

• Sleep epochs of 30s.
• Motion features
  • Activity count over epoch (e.g. integrated area under the accelerometer signal).
  • Accelerometer magnitude.
  • Time since last significant movement.
  • Time till next significant movement.

Examples of patterns vs sleep stages

• An activity count feature which includes the magnitude and duration of movement during the 30 s epoch is easily interpreted as being correlated with wake.
• Periods of near-constant heart rate and low movement are associated with deep sleep.
• Periods with a high degree of short-term heart rate variability (e.g. as seen in the LF and HF spectral features) and relatively little movement are associated with REM.
Heart Rate Variability Features

- Inter Beat Interval to calculate Heart Rate Variability:
  - High Frequency (eg through DFT) 0.15–0.4 Hz
  - Low Frequency 0.04–0.15 Hz
  - VLF power (0.015–0.04 Hz)
  - RMSSD: Root mean square of successive differences of IBI
  - pNN50: proportion of successive IBIs that differ more than 50ms over total IBIs
  - Delta IBIs
  - Mean heart rate
  - 90th percentile heart rate
  - 10th percentile heart rate
Breathing Features (see previous lecture)

1s breathing sample: take the frequency spectrum (and limit the power of frequency to plausible breathing frequencies).

• HF power (0.15–0.4 Hz)
• LF power (0.04–0.15 Hz)
• VLF power (0.015–0.04 Hz)
Results

<table>
<thead>
<tr>
<th>Actual (%) stage</th>
<th>Wake (15.7%)</th>
<th>REM</th>
<th>Light (50.8%)</th>
<th>Deep (14.4%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wake</td>
<td>6116</td>
<td>640</td>
<td>2035</td>
<td>33</td>
</tr>
<tr>
<td>REM (19.1%)</td>
<td>424</td>
<td>7653</td>
<td>2566</td>
<td>47</td>
</tr>
<tr>
<td>Light (50.8%)</td>
<td>1995</td>
<td>3598</td>
<td>19681</td>
<td>3179</td>
</tr>
<tr>
<td>Deep (14.4%)</td>
<td>145</td>
<td>314</td>
<td>2583</td>
<td>5056</td>
</tr>
</tbody>
</table>

[Graph showing the relationship between true duration and estimated duration for different stages.]
## Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall sensitivity (in detecting sleep)</td>
<td>94.6</td>
</tr>
<tr>
<td>Specificity (in detecting wake)</td>
<td>69.3</td>
</tr>
<tr>
<td>PSG-tracker agreement (light sleep)</td>
<td>69.2</td>
</tr>
<tr>
<td>PSG-tracker agreement (deep sleep)</td>
<td>62.4</td>
</tr>
<tr>
<td>PSG-tracker agreement (REM sleep)</td>
<td>71.6</td>
</tr>
</tbody>
</table>

Cohen’s Kappa is used to rate agreement between PSG and tracker.

$$\kappa = \frac{\%\text{Observed Agreements} - \%\text{Agreements by chance}}{1 - \%\text{Agreements by chance}}.$$
New wearables and sleep...

- Oura Ring
- Oura underestimated TST and overestimated WASO.
- Oura significantly underestimated REM sleep and light sleep (stage N1+N2), and overestimated time spent in deep sleep (stage N3)
Deep Learning approaches over PPG Sleep

- Works that use **transfer learning**: model trained on a large database of heart rate variability (HRV) measures and then fine-tuned to a smaller database of pulse rate variability (PRV) measures derived from the IBIs detected on the PPG.
- ECG can be used to calculate HRV and HRV can be correlated with sleep stages (LSTM models seem good)
- There is a lot of ECG data: train on that!
- Then transfer to lower data regime.
Deep learning approaches over PPG Sleep

Cohen’s kappa: how much is the agreement of each method with PSG?
PPG. When does it not work: Motion...
Bland-Altman Plots

Polar 7 strap as ground truth
Heart Rate Measures

PPG on Earables. When does it not work: Micro motion...

(a) Shake.  
(b) Brow Raiser.  
(c) Lip Puller.  
(d) Mouth Stretch.

Red line: PPG in stationary case. Gray line: (same user) in various motion sessions for that movement.

PPG. When it does not work: Skin colour

• Paper [1] discusses:

  • That Paper [2] found no significant difference in accuracy across skin tones but did find differences by devices in response to changes in activity.
  • Previously reported studies [3] finding wearables using green light technology had larger errors rates in tracking heart rate and energy expenditure for individuals with darker skin tones especially if exercising.
  • Racial biases and limitations of Fitzpatrick Skin Type Scale: originally used for propensity skin to burn :)
  • Too few people with the darkest skin tones were included \((n = 9\) in FST Type 6) in paper [2].

References on the next slide.

References from previous page


Skin colour affects oxygen-sensor accuracy

COVID-19 broadened the use of pulse oximeters for rapid blood-oxygen readings, but it also highlighted the fact that skin pigmentation alters measurements. Two groups of researchers analyse this issue, and its effects on people with dark skin.

Matthew D. Keller
& Brandon Harrison-Smith
Pulse-oximetry errors affect patient outcomes

Since Sjöding and colleagues’ report, several large retrospective studies have confirmed that darker-skinned people (those self-identifying as Black, Asian, Hispanic or a combination of these) are more likely than white people to experience occult hypoxaemia\textsuperscript{1,2}. In one study of people with COVID-19, had equivalent arterial blood-gas values\textsuperscript{3}. A more comprehensive analysis showed that, even when baseline health conditions are taken into account, people with occult hypoxaemia are prone to organ dysfunction and in-hospital mortality, and that Black people in this group have the worst organ dysfunction\textsuperscript{4}.

Although clinical reports of skin-colour bias in pulse oximetry were not widespread until the COVID-19 pandemic, evidence for this issue has been accumulating for decades\textsuperscript{5,6}. A comparison reported in February found that pulse-oximeter readings from nine devices were consistently less accurate for darker-skinned people than for lighter-skinned people\textsuperscript{7}. But the study also found that testing healthy individuals under carefully controlled laboratory conditions resulted in fewer cases of occult hypoxaemia than are measured in hospitals. In fact, none of the 491 people who were tested by the authors had readings consistent with occult hypoxaemia, whereas Sjöding and colleagues tallied 187 cases out of 3,527 measurements from a...
**Figure 1** | **Pulse-oximetry accuracy varies with skin tone.**

**a,** Devices known as pulse oximeters estimate the oxygen concentration in a person’s blood by shining red and infrared light through their fingertip. Oxygenated haemoglobin absorbs infrared light more efficiently than it does red light, whereas the opposite is true for deoxygenated haemoglobin. **b,** These signals are affected by melanin, which is distributed through the skin in structures, known as melanosomes, that are produced by cells called melanocytes. Melanosomes in dark skin are both larger and more numerous than are those in light skin. Long-standing oximetry theory does not fully account for the way in which photons are scattered by the biomolecular content and structure of the tissue, and thus imprecisely corrects for the effect of pigmentation. Calibration studies compound this problem, because they typically oversample light-skinned people. This has led to overestimation of the oxygen concentration in some Black individuals’ blood, and therefore to missed diagnoses of dangerously low oxygen levels.
Questions