

実生活中的な継続的な生体センシングの難しさ

Difficulties in continuous physiological sensing in-the-wild

荒川豊*¹ 中村 優吾*¹ 松田 裕貴*²
Yutaka Arakawa Yugo Nakamura Yuki Matsuda

*¹九州大学 Kyushu University *²奈良先端科学技術大学院大学
Nara Institute of Science and Technology

To realize feasible services based on physiological sensing, we need to assume that commercially available sensors will be used on a daily basis. Collecting physiological data from ordinary people over a long period of time is quite different from performing highly accurate physiological sensing using dedicated sensors in a controlled environment in a laboratory. In this paper, through two investigations, we share the problems of physiological sensing in real life and clarify the points that should be considered to achieve stable sensing. The first investigation was conducted by three researchers who are accustomed to using wearable devices, and found that even with careful recording, the collected values are different by the devices and applications. The second survey was conducted on 60 general public. Since 15% of them forgot to wear the wearable during the night, it indicates the difficulty of asking people who are not in the habit to sleep with the wearable on. In addition, the number of heart rate data recorded per day differed greatly from person to person, which may have an effect on data analysis.

1. はじめに

2021年8月の発表では、Apple Watchのユーザ数が1億人を突破するとともに、100ドル未満のスマートウォッチの販売は前年比547%も増加している[Counterpoint Research 21]など、ウェアラブルデバイスの利用者は急速に増加している。こうしたウェアラブルデバイスには、加速度センサや光学式脈波センサ等のセンサが搭載されており、種々の生体データを手軽に収集可能になっている。

そして、このようなデバイスから取得した生体データに機械学習を組み合わせ、身体状態や内面状態を計測、あるいは推定する研究が盛んに行われており、今や、加速度センサを用いた歩数計、加速度センサと脈拍センサを用いた睡眠計などどのような機種でも標準的に搭載される機能になりつつある。我々も、メンタルヘルス領域の研究において、産業保健師が利用する質問票と生体データを同時収集し、それらの関係性を見出す研究を4年前から実施[Tani 20]しており、延べ200名以上の一般社員に小型活動量計を配布し、データ収集を行ってきた。そして、抑うつ傾向と睡眠との関係性[Fukuda 20, Fukuda 21]やワーク・エンゲイジメントと脈拍の関係性[Harashima 21]、業務内容とメンタルヘルスの関係性[Nishimura 21]について分析を進めてきた。これらの研究では、Fitbit Charge 4を統一的に利用してきたが、実験室実験ではなく一般社員に長期間装着してもらい実生活中での計測であることから、種々の問題が生じていると考えられる。

例えば、装着し忘れや充電し忘れによるデータ欠損、ずれによる外れ値(異常値)などが挙げられる。また、今後、他のウェアラブルデバイスなどを活用することを想定すると、センサの個体差やメーカーによるアルゴリズムの差も生じてくる。本稿では、このような実生活中での市販民生デバイスを用いた継続的な生体センシングについて、実際のデータをもとに議論を進め、どのような対策が可能かについて検討する。

以降では、2つの調査(A, B)に分けて説明する。調査A
連絡先: 荒川豊, 九州大学大学院システム情報科学研究院情報
知能工学部門, 福岡市西区元岡 744, 092-802-3792,
arakawa@ait.kyushu-u.ac.jp

は、ウェアラブルデバイスを使い慣れた3名の研究者が行ったもので、装着忘れや装着ミスはないものとし、手動で記録した睡眠の真値をもとに、デバイスやアプリ、装着者による違いについて分析する。一方、調査Bは、一般人60名を対象したもので、装着忘れや装着ミスによる取得できたデータの違いについて分析する。

2. 調査A: デバイスやアプリケーションによる生体センシングの品質差

生体センシングのために使用するウェアラブルデバイス、データ処理用アプリケーション、あるいはそれらの組み合わせによって得られるデータに差がある場合、その後の分析・モデル構築に影響を与える可能性がある。本章では、「睡眠ログ」データを対象として、次の2つの観点から分析を行う:異なるデバイスによって得られたデータに差があるかどうか(分析1)、同一デバイスで測定したデータを異なるアプリケーションで処理した場合に差が現れるかどうか(分析2)。

調査にあたって3名の被験者に2台の異なるデバイスを装着してもらい、2週間に渡ってデータ収集を行った。被験者の属性・データ収集期間を表1、各自が使用したデバイス・アプリを表2に示す。これらのセンサデータに加えて、各被験者は日々の就寝・起床時間(以降、ラベル)を記録している。

表1: 調査Aにおける被験者属性とデータ収集期間

被験者ID	年齢	性別	データ収集期間
P1	44	男性	2022年1月20日~2月3日
P2	29	男性	2022年1月20日~2月3日
P3	29	男性	2022年2月3日~2月17日

2.1 データ収集結果の概要

データ収集の結果を図1に示す。横軸は時間(2週間の各日)、縦軸にデータソースを示す。“Hand-label”は被験者による就寝起床時刻ラベル、“Fitbit”はFitbitから得られたデー

表 2: 使用したデバイスとアプリ

カテゴリ	名称	P1	P2	P3
デバイス	Fitbit ^{*a}			
	... Charge 5	✓	✓	-
	... Charge 4	-		✓
	AppleWatch ^{*b}			
	... Series 7	✓	-	✓
	... Series 6	-	✓	-
アプリ	iPhone ^{*c}			
	... 13 Pro	✓	-	✓
	... 11 Pro	-	✓	-
	Fitbit App ^{*a}	✓	✓	✓
	Apple Health ^{*d*h}			
	... AppleWatch	✓	✓	✓
... iPhone	✓	✓	✓	
Somnus ^{*e*h}	✓	-	✓	
SleepWatch ^{*f*h}	✓	-	✓	
AutoSleep ^{*g*h}	-	✓	-	

^{*a} <https://www.fitbit.com/> ^{*b} <https://www.apple.com/watch/>
^{*c} <https://www.apple.com/iphone/> ^{*d} <https://www.apple.com/ios/health/>
^{*e} <https://somnus.jp/> ^{*f} <https://www.sleepwatchapp.com/>
^{*g} <https://autosleepapp.tantsissa.com/>
^{*h} データは Apple Health アプリを経由してデータエクスポートした。

た、それ以外(“AH...”)は各アプリから AppleHealth に提供・蓄積されたデータをそれぞれ示している。特筆すべき箇所には、緑色の実線円(データ欠損)、赤色の一点鎖線四角形(睡眠判定の誤差大)、橙色の破線四角形(睡眠の誤検出)でマーキングした。

データ欠損および誤検出が生じているサンプルを除外した上で、入眠・起床の推定誤差を算出した結果を表 3 に示す。P1 については、睡眠サイクルが安定しているため、デバイス・アプリ間での誤差には違いが見られなかった。しかしながら P2・P3 に関しては、睡眠時間や睡眠タイミングが日々異なる生活であるため、デバイス・アプリ間の間に大きな差が見られた。

2.2 分析 1: 異なるデバイスによる差

本調査では、デバイスの型番に違いはあるものの、大きく分けて 3 種類のデバイス (Fitbit・AppleWatch・iPhone) を用いたデータ収集を行った。ここでは、それらのデバイスを比較しつつ傾向を整理する。

まず Fitbit は他と比較して広い時間範囲を睡眠と判定する傾向があるが、おおよそラベルに近い判定を下しているように見られる。しかしながら、短い睡眠において睡眠判定に失敗し、データを取り逃しているケースが見られる(緑色の円)。一方で、日中にごく短い期間を睡眠と誤判定しているケースも見られる(橙色の破線四角形)。このことから、ある程度十分な睡眠時間が確保される状況において Fitbit の強みが発揮されると整理できる。

次に AppleWatch は、短い睡眠も比較的正確に捉えているものの、起床時刻を実際よりも早く判定する傾向にある(特に P2 においてこの傾向が見られる)。これは、睡眠判定アルゴリズムの中に加速度などが含まれており、起床前の睡眠が浅くなるタイミングでの体動を「起床」と誤判定している可能性がある。このことから、AppleWatch は欠損なく睡眠イベントを

検出することには長けているものの、睡眠時間帯の推定には向かないと整理できる。

最後に iPhone では、スマートフォン搭載センサによって得られたデータに基づくものであるため、睡眠判定をしておらず、動きが止まったタイミング(すなわち、InBed)のみを取得しているものと考えられる。そのため、睡眠開始の判定が困難である事がわかる(赤色の一点鎖線四角形)。また、日中スマートフォンを机に放置しているタイミングなども睡眠と誤判定しているケースが見られる(橙色の破線四角形)。このことから、スマートフォンを睡眠判定に用いることは適切でないと整理できる。

2.3 分析 2: 異なるアプリ処理による差

本稿では、AppleWatch から得られるデータに基づいて複数のアプリケーションでの分析データも収集している。そこで、異なるアプリ処理に基づく差について傾向を整理する。

データの粒度について AppleHealth からエクスポートされたデータ群(AH...)を見ると、Apple から提供されている睡眠カテゴリ(Awake・Asleep・InBed)が、アプリ毎に異なる割合となっていることがわかる。Somnus や SleepWatch などの各アプリでは、より詳細な睡眠ステージに関する分析が示されているが、AppleHealth と連携する際に情報を抽象化する処理を施しており、そのアルゴリズムの差からこのような違いが生まれているものと考えられる。このことから、AppleHealth といった共通基盤を経由することで本来得られているはずのデータ粒度が抽象化されてしまう課題が示唆されている。

データの欠損について 各被験者はいずれの日においても、AppleWatch および Fitbit を装着し、iPhone を携帯していたにもかかわらず、データの欠損が見られる(緑色の円)。P1 については Somnus アプリ (Day3~6)、P2 については AutoSleep アプリ (Day2)、P3 については Somnus アプリ (Day14) にて欠損が見られる。これは、アプリによって同期方法に差があり、ユーザの操作タイミングに同期の成否が依存している可能性がある。

睡眠判定について AppleHealth (AppleWatch) に比べて、サードパーティアプリ (Somnus, SleepWatch, AutoSleep) が、ごく短い時間幅でも睡眠判定を下す傾向があることがわかる。特に P3 では顕著で、デスクワークや動画視聴といった体動が減るタイミングを睡眠と判定している。サードパーティアプリではより細やかな睡眠状況が捉えられることを独自性としていることも考えられ、睡眠判定アルゴリズムの感度が高く設定されている可能性が示唆された。

2.4 分析 3: デバイスおよび装着した腕の左右による違い

図 2 は、被験者 P1 について、心拍データと歩数データを示したものである。P1 は右利きで、左手に Apple Watch を装着し、右手に Fitbit を装着して生活している。当然、それぞれのアプリ上において装着している腕については正しく設定している。

まず、心拍については、Apple Watch と Fitbit で下限値は概ね似ているように思われるが、上への振れ幅は Fitbit の方が大きい印象である。これは、機種によっては、ハードウェアあるいはアプリ内部で平滑化処理が働いている可能性が示唆される。

次に、歩数については、Apple Watch の方が多くカウントすることが多い。Day1, Day6, Day13 は、昼食のため外出し 1 キロ程度歩いている。Day1 と Day6 は同じエリアである。また、Day2, Day9 は土曜日で、毎週車で大型ショッピングモー

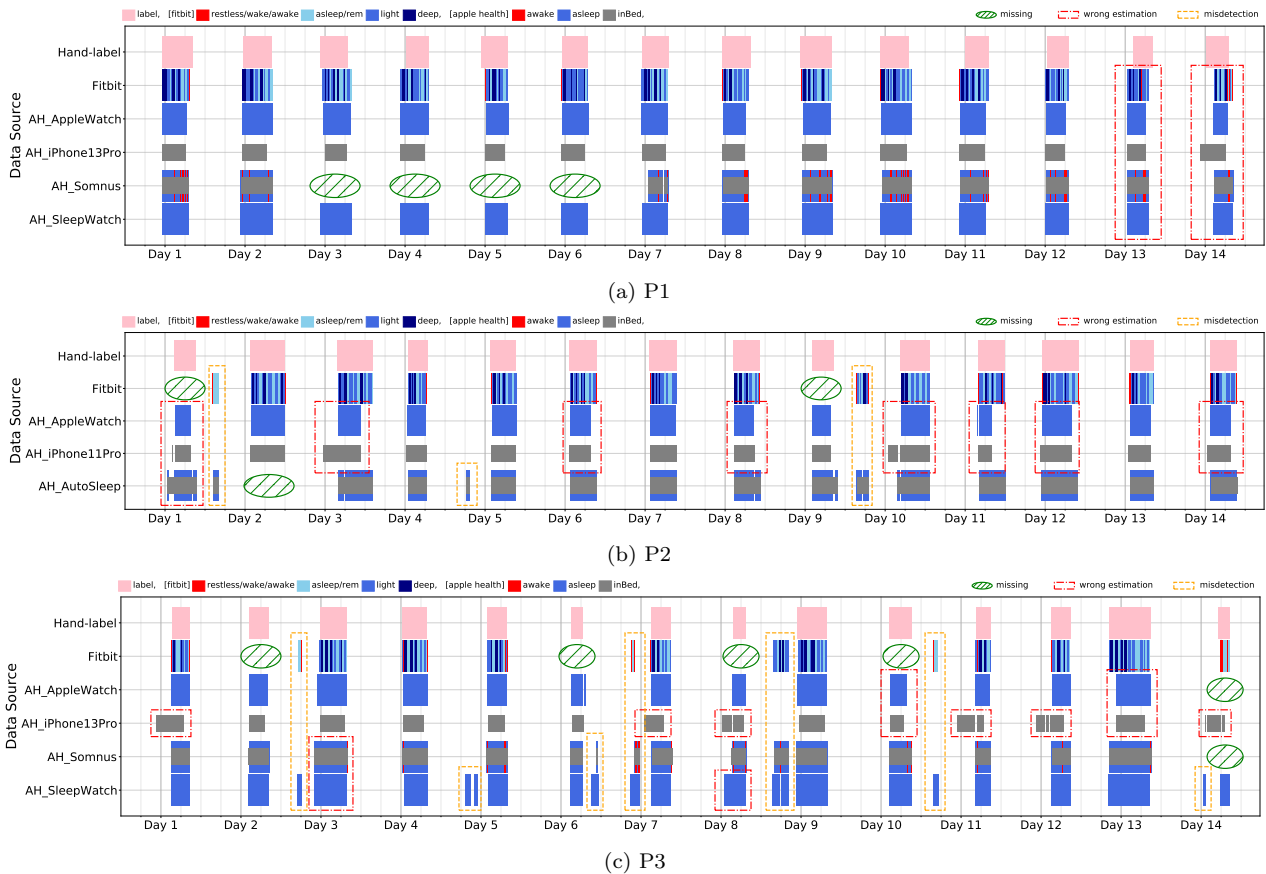


図 1: 睡眠データ

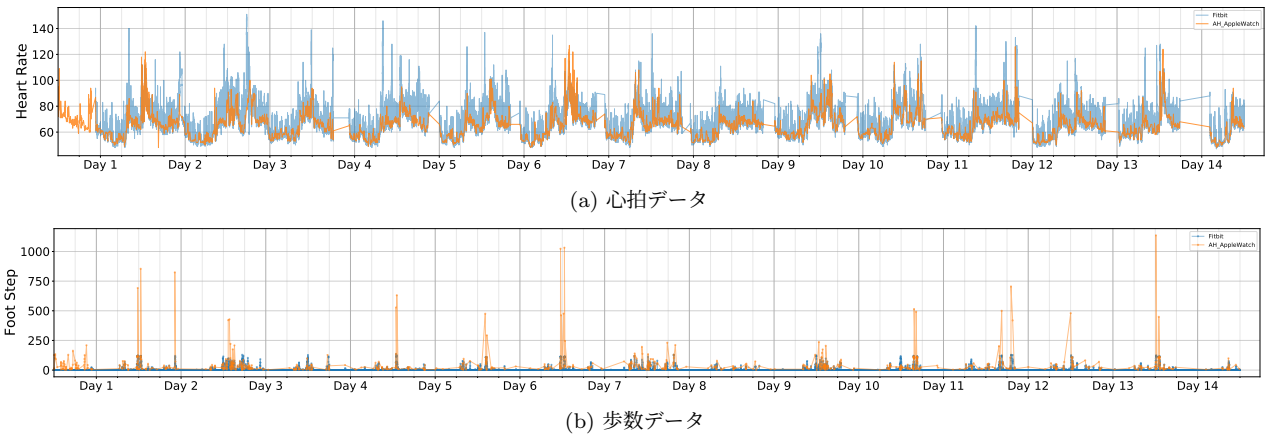


図 2: P1 (Apple Watch: 左手, Fitbit : 右手 (利き手))

ルにしている。これら記録のあるデータをもとに考えると、Apple Watch による歩数カウントの方が真値に近く、利き手に装着した Fitbit の歩数は低めに抑えられていると言える。

左右の違いについては、利き手の方が、料理（特に包丁を使う場合）をする際に歩数とカウントされやすくなる可能性が考えられるが、今回、左右の機器を入れ替えた実験を行っていないため、検証することはできなかった。また、デバイス間の差異についても、正確には腕の同じ場所に装着して、歩数や心拍を計測し比較する必要があるが、今回は、3名の被験者ともに左右の腕にそれぞれ装着したため、厳密な比較にはなっていない。

3. 調査 B：一般会社員 60 名を対象とした生体センシングのデータ品質

調査 A では、真値を正確に記録するためにすべての被験者は研究者であった。研究者たちは、普段からウェアラブルを装着していることから、つけ忘れ、充電し忘れといった問題点は起きなかったが、一般的な被験者を対象とした実験ではそうした問題も生じる。そこで、これまでに集めたデータのうち 2018 年に一般会社員 60 名を対象として行った実験データを振り返ってみた。このときの実験では、企業 5 社の総務・人

表 3: デバイス・アプリごとの入眠・起床時間推定誤差

入眠時間の推定誤差 (分)						
ソース	P1		P2		P3	
	Avg.	Std.	Avg.	Std.	Avg.	Std.
Fitbit	49.5	44.7	8.9	3.7	10.9	7.5
AppleWatch	46.6	41.3	13.7	5.9	23.8	34.3
iPhone	50.5	43.1	48.5	76.4	119.0	119.7
Somnus	57.6	44.4	-	-	20.6	27.3
SleepWatch	44.9	43.6	-	-	31.1	44.4
AutoSleep	-	-	24.5	34.8	-	-

起床時間の推定誤差 (分)						
ソース	P1		P2		P3	
	Avg.	Std.	Avg.	Std.	Avg.	Std.
Fitbit	31.2	28.7	3.3	3.1	5.2	4.9
AppleWatch	44.9	42.0	80.7	75.6	15.9	23.3
iPhone	75.2	33.2	80.2	76.1	77.4	35.7
Somnus	36.9	28.2	-	-	8.0	6.9
SleepWatch	33.1	26.8	-	-	4.4	3.8
AutoSleep	-	-	13.6	20.1	-	-

事部門が主導し、各社内でボランティア被験者を募り、Fitbit Charge 3 を貸与して 2~3 週間、データを計測するものであった。金銭的報酬はないが、同時に計測されたさまざまな調査票(ワーク・エンゲイジメントなど産業衛生分野で一般的に利用されているもの)の分析結果を最後に集計して提供することを報酬としていた。

表 4 は、実験で収集したデータの概要を示したものである。母数は、被験者 n の計測日数を D_n とするとして、 $\sum_{n=1}^{60} D_n$ となり、被験者 60 名で延べ 1,240 レコードであった。参加に日数については、平均で 20 日以上参加しているものの、わずか 2 日で離脱した人があることもある。次に心拍データ数についてみると、1 日あたり、1 件/日のレコードしかない人から、34,414 件/日のレコードが記録されている人もいた。個人間で見ても、実験期間中、最大で 5,672 件/日のレコードしか計測されていない人もいれば、最小でも 10,038 件/日のレコードがある人があり、かなりばらつきがあった。

次に装着し忘れや充電し忘れによるデータの欠損についてであるが、分析の簡略化のため、殆どの被験者が就寝しているであろう深夜 3~4 時と、就業しているであろう 13~14 時のデータの有無をもとに、データ欠損状況を分析した。その結果も、興味深いもので、昼間も夜中も全くつけ忘れることがない人がある反面、夜間の 95% は装着しなかった人もいた。また、昼間装着していない日や、終日装着していない日が 22% に達する被験者たちもいた。延べレコード数 1240 に対して、平均としては、夜のデータが存在しないものが 15%、昼間のデータが存在しないものが 7%、そして終日データが存在しないものが 3% あることがわかった。

これらの結果から、普段ウェアラブルを装着しない人達に対する実験においては、特に夜間での装着をどのように習慣化してもらうかということが課題になるといえる。同時に、計測データ数が人によって大きく異なる点は、装着ズレ等に起因していると考えられ、機器の装着方法についても適切に伝える必要がある可能性が示唆される。

表 4: 60 名の被験者の計測結果

	参加日数	1 日あたりの心拍データ数			装着忘れ率		
		最小	最大	平均	夜間	昼間	終日
最小	2	1	5672	4778.9	0	0	0
最大	27	10038	34414	27335.8	0.94	0.22	0.22
平均	20.65	3352.4	15193.93	10631.39	0.15	0.07	0.03
標準偏差	5.2	2581	8482.9	4084	0.2	0.1	0

4. まとめ

本稿では、実生活中での市販民生デバイスを用いた継続的な生体センシングに関する 2 つの調査を実施し、発生しうる可能性が高い問題を特定するとともに、実環境で安定したセンシングを実現するために考慮すべき点を明らかにした。本稿で紹介した調査 A は、ウェアラブルデバイスを使い慣れた 3 名の研究者を対象とした予備的な実験設定であったため、今後は、より多くの被験者の被験者を募り、利き手や年齢・性別の違いがデバイス間の差異に与える影響を網羅的に調査する予定である。また、本稿では取り扱わなかった、より詳細なデータの品質(睡眠ステージなどの判定品質)についても、長期的な収集データを収集するとともに継続的に調査を進める。

謝辞 本研究の一部は、JSPS 科研費 JP18H03233 の支援を受けて実施したものである。

参考文献

- [Counterpoint Research 21] Counterpoint Research, : Smartwatch Market Grows 27% YoY in Q2 2021; Apple Watch User Base Crosses 100 Million, <https://www.counterpointresearch.com/smartwatch-market-grows-27-yoy-q2-2021-apple-watch-user-base-crosses-100-million/> (2021)
- [Fukuda 20] Fukuda, S., Tani, Y., Matsuda, Y., Arakawa, Y., and Yasumoto, K.: Predicting Depression and Anxiety Mood by Wrist-Worn Sleep Sensor, in *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom) Workshops*, pp. 1-6 (2020)
- [Fukuda 21] Fukuda, S., Matsuda, Y., Arakawa, Y., and Yasumoto, K.: Statistical Analysis between Sleep Status and Occupational Health Indicators for Detecting Depression Signs in Healthy Workers, in *The 13th International Conference on Mobile Computing and Ubiquitous Networking (ICMU '21)*, pp. 1-2 (2021)
- [Harashima 21] Harashima, H., Arakawa, Y., Ishida, S., and Nakamura, Y.: Estimating Work Engagement with Wrist-Worn Heart Rate Sensors, in *The 13th International Conference on Mobile Computing and Ubiquitous Networking (ICMU2021)*, pp. 1-6 (2021)
- [Nishimura 21] Nishimura, Y., Hossain, T., Sano, A., Isomura, S., Arakawa, Y., and Inoue, S.: Toward the Analysis of Office Worker's Mental Indicators Based on Activity Data, in *The 3rd International Conference on Activity and Behavior Computing (ABC2021)*, pp. 1-26 (2021)
- [Tani 20] Tani, Y., Fukuda, S., Matsuda, Y., Inoue, S., and Arakawa, Y.: WorkerSense: Mobile Sensing Platform for Collecting Physiological, Mental, and Environmental State of Office Workers, in *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 1-6 IEEE (2020)