A First Step Towards On-Device Monitoring of Body Sounds in the Wild

Shyam A. Tailor
University of Oxford
& University of Cambridge

Jagmohan Chauhan
University of Cambridge

Cecilia Mascolo
University of Cambridge

ABSTRACT

Body sounds provide rich information about the state of the human body and can be useful in many medical applications. Auscultation, the practice of listening to body sounds, has been used for centuries in respiratory and cardiac medicine to diagnose or track disease progression. To date, however, its use has been confined to clinical and highly controlled settings. Our work addresses this limitation: we devise a chest-mounted wearable for continuous monitoring of body sounds, that leverages data processing algorithms that run *on-device*. We concentrate on the detection of heart sounds to perform heart rate monitoring. To improve robustness to ambient noise and motion artefacts, our device uses an algorithm that explicitly segments the collected audio into the phases of the cardiac cycle. Our study with 9 users demonstrates that it is possible to obtain heart rate estimates that are competitive with commercial devices, with low enough power consumption for continuous use.

CCS CONCEPTS

• **Computer systems organization** → **Sensor networks**; *Embedded hardware.*

KEYWORDS

Wearables, Audio, Mobile Health

1 INTRODUCTION

Body sounds are a rich source of information about the health of an individual, and can be used for disease diagnosis or tracking. *Auscultation*, the practice of listening to body sounds, has been used for centuries [20] to diagnose cardiac, respiratory and digestive conditions. However, most existing research on auscultation is limited to user studies in a controlled and non-continuous settings which limit the ability of the technique to detect diseases more widely and promptly. Continuous acoustic-based health monitoring has the potential to lead to affordable and scalable monitoring

solutions, as microphones can be manufactured inexpensively and to fit wearable form factors.

Recent works [12, 16, 23] have explored the use of body sounds for medical diagnostics in non-clinical settings. Complementing these works, we make a first step in constructing a *continuous* sensing wearable platform that can monitor body sounds accurately in realistic scenarios.

There are substantial challenges that must be considered to exploit audio collected from the body with a wearable device. The device must be sensitive to the appropriate frequencies, and be designed so it be placed near the source of the sound. Additionally, the device must cope with high ambient noise levels and artefacts introduced by user activities, while running the inference on-device to *maintain user privacy* and *reduce power consumption*. This work focuses on heart rate monitoring as a motivating application. However, the overarching goal of our research is to enable further research into continuously monitoring different vital signs and diseases through the use of wearable devices using an on-device analysis framework. This work makes the following contributions:

(1) We construct a novel device for collecting non-speech body sounds, which can be worn under the user's clothes and which tolerates high noise conditions.
(2) The collection of a new dataset, with 9 users, which assesses the impact of real world conditions on the device, including ambient noise levels and user activities.
(3) We propose an algorithm for accurately calculating heart rate and heart rate variability (HRV) which can run continuously on-device for 2 days.

2 RELATED WORK

Applications of Non-Speech Body Sounds: Several existing works [12, 13, 16, 22, 23] have explored monitoring body sounds using wearables. Larson et al. [12] detect coughs from audio captured by a mobile phone; ApneaApp [13] is a smartphone app which can detect sleep apnea issues in a contactless manner using respiratory sounds. Yatani et al. [22] constructed a device to record sounds from the throat to perform human activity recognition. These works highlight how body sounds can be used to monitor health and well-being. Our work differs in two ways. Firstly, we focus on a continuous heart rate monitoring application rather than activity recognition or monitoring discrete respiration events; this is challenging especially in realistic settings. Secondly, we focus on user privacy by performing all the processing locally on-device.

Heart Rate Monitoring: Different types of sensors such as photoplethysmography (PPG) [15], inertial [8], cameras [10], and wireless [6] have been studied for heart rate monitoring. Hernandez et al. [8] showed that inertial measurement unit (IMU) sensors can measure heart rate accurately from the wrists on some idle

activities. Adib et al. [6] proposed doing heart rate estimation using WiFi signals. Monitoring heart rate can be highly inaccurate with IMU and PPG sensors as they are highly sensitive to human motion, while both camera-based and wireless-based methods are impractical for continuous monitoring. In contrast, acoustic sensors are expected to be less impacted by motion than an IMU or PPG sensors but instead are sensitive to ambient noise. We show that carefully designed hardware and algorithms can yield accurate results while being resilient to ambient noise and motion.

A recording of heart sounds is known as a phonocardiogram (PCG), and techniques for automated analysis have been investigated. Applications include heart rate estimation or identifying heart abnormalities [14, 17, 18, 21]. However, these works are not analysed in the context of a limited computational or energy budget. These techniques require significant computational resources, which may be achieved using offloading from the wearable. However, there are privacy concerns and issues with power consumption. Our work explicitly addresses these issues: we aim to perform continuous, on-device, heart rate monitoring in realistic settings.

3 WEARABLE DEVICE DESIGN

Hardware: We now provide an overview of the hardware.

Microphone Selection: We used an off-the-shelf contact microphone [1] that uses a piezoelectric transducer, rather than electret or condenser microphones, as they offer an excellent low-frequency response, making them sensitive to the frequencies of interest. These microphones pick up vibrations from the skin directly—necessitating the use of an elastic strap to hold the microphone against the skin—but making them less susceptible to ambient noise.

Microcontroller: A Teensy 3.2 development board was used to control the wearable as it has low power consumption and has been successfully used for other audio datalogging projects [3]. Although we used the Teensy to collect the data, we do not foresee it being used in a commercial product due to reasons specified in section 5.

Printed Circuit Board Implementation: The PCB design had dimensions of 31x37mm. Most board area is devoted to audio, but some ancillary functionality was included such as a microSD card slot and an IMU sensor to collect inertial data. The PCB implemented buffering and amplification circuits to amplify the signal and improve the impedance match between the microphone and the Teensy's ADC. The design used a low-power operational amplifier designed for wireless microphones. An enclosure was 3D printed and mounted onto an elastic strap to be worn around the torso.

Placement: Four possible placements, often evaluated in clinical settings, were considered: (1) bottom of sternum, (2) on the back behind the heart, (3) bottom of ribcage (offset towards heart) and (4) at the top of the chest (offset towards heart). The last placement was chosen to align with the bronchus, which was hypothesised to allow better transmission of respiratory sounds. Placement 1 was unstable due to the surrounding curvature on the body, which has a wide inter-person variance. Placement 2 is also problematic: shoulder blade movement displaces the device from the skin. Placement 3 and 4 were evaluated by one of the authors. The top of the chest was selected after assessing the audio recorded while performing 7 different activities, including sitting quietly, deeply breathing, drinking, coughing, sniffing, throat clearing, and talking.

4 HEART RATE ESTIMATION ALGORITHM

Our proposed algorithm uses the sounds present in the cardiac cycle. We first discuss the cardiac cycle and the associated sounds, before proceeding to describe heart rate and HRV estimation algorithms.

Cardiac Cycle: The heart consists of 4 chambers: 2 atria, and 2 ventricles [24]. Between the chambers, there are one-way valves which force the blood into the next chamber upon a contraction; when these valves close, the heart sounds are produced. There are two sounds in healthy hearts: S_1, referred to as "lub", and S_2 ("dub"), which is usually quieter. This process is shown in Figure 1; for more detail, we refer the reader to Gersh [24].
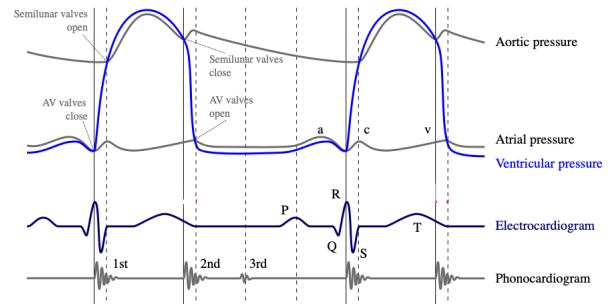

Figure 1: Heart sounds in cardiac cycle [5]

Heart Rate Estimation Using Segmentation: Springer et al. [19] proposed a state-of-the-art algorithm for PCG segmentation, which used a hidden *semi*-markov model (HSMM) to model transitions between 4 states: S_1, systole, S_2 and diastole [19]. Using an HSMM allows for the state residence time to be modelled, as the transition probabilities are dependent on the time that has elapsed since entry to the current state. The emission probabilities for each state is modelled using logistic regression trained from envelope features extracted from the PCG. However, the results on our collected data were extremely poor: envelope-based features are disrupted by transient changes in signal amplitude. Another issue is that these features are computationally expensive. Hence, we propose several modifications to the original algorithm.

Adaptations for Continuous Monitoring: We used log-magnitudes of Short-Time Fourier Transform (STFT) coefficients as features, as different activities introduce power into different parts of the spectrum. Therefore a classifier can learn to ignore known noise profiles. We used audio sampling rate of 500Hz to enable frequencies below 200Hz to be distinguished. The STFT was performed on 16 samples (Hann windowing, hop length of 5 samples) yielding 100 features per second. A plot of the features against time is reported in Figure 2. When the user is stationary, consistent spikes in all features can be observed, corresponding to heart sounds.

We also use Random Forests (RF) instead of logistic regression (LR) as a RF can learn non-linear decision boundaries, while admitting efficient inference on a microcontroller. A RF (10 trees, maximum depth 8) was trained to predict the presence of *either type* of heart sound. This optimisation was made as the two heart sounds are difficult to distinguish. By tying the emission probabilities, the inference of the most likely path through the states relies on duration of the states observed. This is relatively easy as the
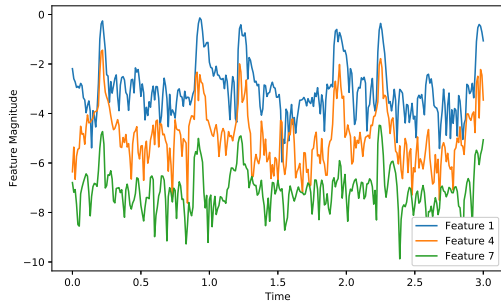
Figure 2: Plot of spectral features used for segmentation against time; in this scenario the user is still. Higher-indexed features correspond to higher frequencies.
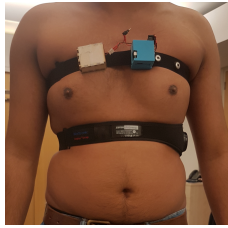

Figure 3: Devices: Ours (top), ground truth (bottom).

S_1-S_2 and S_2-S_1 durations are notably different: 0.128 ± 0.062 and 0.356 ± 0.121 seconds respectively in our dataset.

Estimating Heart Rate and Variability: The segmentation is post-processed by Algorithm 1, which describes steps to obtain the heart rate and HRV. The intervals between heartbeats are found from the labels; intervals are the differences between the *start* of two S_1 states. Kalman filtering was used to estimate heart rate from the intervals. For accurate HRV estimation outliers were rejected using a standard technique [9].

Algorithm 1 HR and HRV estimation.

```
function EstimateHeartRate(labels)
    deltas = FindTimesBetweenBeats(labels)
    bpm = 60 / (deltas / labelSampleRate)
    bpm = KalmanFilter(bpm)
    return bpm
function EstimateVariability(labels)
    deltas = FindTimesBetweenBeats(labels)
    retainedDeltas = List()
    for i in 1...Len(deltas) - 1 do
        window = deltas[Max(i - 4, 0): i]
        m = Mean(window)
        if m × 0.7 ≤ deltas[i] ≤ m × 1.3 then
            retainedDeltas.Append(deltas[i]) // Reject if outlier
    return StandardDeviation(retainedDeltas)
```

---

5 RESULTS

User Study: Nine participants (3 female, 6 male, all healthy) performed the data collection after we obtained ethics approval. Participants wore the device (Figure 3) and perform several activities under 3 noise regimes: silence, music placed 1 meter from the participant (average loudness of 63dB) and background noise from a

Activity	Noise Regime	Median Absolute Error / BPM	Mean Absolute Error / BPM	Median Percentage Error	Mean Percentage Error
Rest	Silence	0.23 ± 0.01	1.18 ± 0.08	0.33 ± 0.02	1.49 ± 0.09
Rest	Music	0.20 ± 0.01	2.91 ± 0.11	0.26 ± 0.02	3.34 ± 0.13
Rest	Conversation	0.14 ± 0.04	2.54 ± 0.12	0.22 ± 0.07	2.77 ± 0.13
Deep Breathing	Silence	5.18 ± 0.27	7.15 ± 0.11	6.04 ± 0.32	7.80 ± 0.12
Deep Breathing	Music	11.84 ± 0.43	14.20 ± 0.21	13.91 ± 0.56	15.07 ± 0.24
Deep Breathing	Conversation	8.47 ± 0.74	12.29 ± 0.30	9.94 ± 0.82	13.25 ± 0.34
Coughing	Silence	13.75 ± 1.45	13.61 ± 0.61	14.24 ± 1.38	14.33 ± 0.67
Coughing	Music	5.81 ± 0.40	8.59 ± 0.23	6.98 ± 0.49	9.53 ± 0.29
Coughing	Conversation	5.32 ± 0.32	8.73 ± 0.19	6.75 ± 0.57	9.57 ± 0.23
Clearing Throat	Silence	5.40 ± 0.57	7.90 ± 0.35	6.71 ± 0.85	8.96 ± 0.41
Clearing Throat	Music	4.14 ± 0.50	7.27 ± 0.33	5.70 ± 0.67	8.68 ± 0.49
Clearing Throat	Conversation	1.40 ± 0.18	4.78 ± 0.36	1.91 ± 0.19	5.31 ± 0.40
Swallowing	Silence	2.51 ± 0.78	5.65 ± 0.27	2.93 ± 0.90	6.28 ± 0.32
Swallowing	Music	3.24 ± 0.86	5.42 ± 0.31	3.98 ± 0.81	6.19 ± 0.35
Swallowing	Conversation	3.09 ± 0.53	8.11 ± 0.18	4.33 ± 0.76	9.46 ± 0.24
Drinking	Silence	7.07 ± 0.27	7.97 ± 0.30	8.23 ± 0.38	9.18 ± 0.34
Sniffing	Silence	2.89 ± 0.41	4.63 ± 0.29	3.63 ± 0.34	5.73 ± 0.44
Sniffing	Music	1.02 ± 0.22	3.85 ± 0.39	1.45 ± 0.35	4.79 ± 0.55
Sniffing	Conversation	1.18 ± 0.10	3.77 ± 0.22	1.64 ± 0.11	4.31 ± 0.26
Speech	Silence	7.55 ± 0.41	10.81 ± 0.26	11.11 ± 0.56	13.85 ± 0.37
Walking	Silence	6.11 ± 0.10	8.56 ± 0.09	7.35 ± 0.17	9.29 ± 0.11
Running	Silence	37.61 ± 0.40	38.99 ± 0.25	31.62 ± 0.31	30.64 ± 0.20

Table 1: Accuracy of segmentation-based algorithm.

coffee shop placed 1 meter from the participant (average loudness of 42dB). Participants did the following activities: breathing normally and deeply, coughing, clearing throat and sniffing 10 times, swallowing 5 times, drinking water, reading a news article, and walking and jogging for 5 minutes. A Zephyr Bioharness 3 was also worn to provide ground truth ECG data.

Accuracy: The results reported are an average of 10 runs using leave-one-person-out cross validation.

Processing ECG based Ground Truth Data: Raw ECG data was processed using BioSPPY [4] and Neurokit [2] libraries to obtain heart rate estimates from the ground truth data. Using the raw signal derived from the inter-beat-intervals yields noisy results because of HRV. To compensate, we applied exponential smoothing.

Heart Rate Estimation: The heart rate estimates were compared to the ground truth every 2 seconds and are shown in Table 1 The results shows that accuracy when resting is competitive with commercially available chest-mounted heart rate trackers: one study indicated a popular device had a mean percentage error of 0.8% [7]. We obtained a mean percentage error of approximately 3.34% in the worst case, but median percentage error was below 0.33% for each scenario. Different noise regimes have minimal impact, demonstrating our approach's noise robustness.

Reasonable heart rate estimates can also be obtained when doing other activities: while walking, the median percentage error was 7.35%. This result confirms our hypothesis that the classifier can learn to classify heart beats from the spectral coefficients when there is noise. Running, however, introduces sufficient noise for the classifier to degrade substantially. Performance also degrades during speech, for both genders. We believe that this is due to low frequency respiratory noise rather than the vocal frequencies.

Segmentation Accuracy: The segmentation quality is assessed by evaluating the precision and recall of the predicted S_1 states, relative to the ground truth ECG. If the *start* of the S_1 state occurs within 100ms of an R-Peak in the corresponding ECG signal, then the segmentation algorithm is deemed to have correctly predicted the S_1 state. The results are reported in Table 2 and are comparable to Springer et al. [19] (F1 0.956) when users are at rest—despite their data coming from a clinical digital stethoscope. We observe a larger than expected difference between precision and recall for deep breathing: this may be due to a phenomenon where the heart

Activity	Noise Regime	Precision	Recall	F1
Rest	Silence	0.976 ± 0.001	0.960 ± 0.003	0.968 ± 0.002
Rest	Music	0.937 ± 0.004	0.893 ± 0.004	0.915 ± 0.004
Rest	Conversation	0.963 ± 0.004	0.934 ± 0.006	0.948 ± 0.005
Deep Breathing	Silence	0.864 ± 0.004	0.772 ± 0.004	0.815 ± 0.004
Deep Breathing	Music	0.776 ± 0.007	0.637 ± 0.008	0.699 ± 0.007
Deep Breathing	Conversation	0.819 ± 0.004	0.703 ± 0.005	0.757 ± 0.005
Coughing	Silence	0.693 ± 0.011	0.568 ± 0.010	0.624 ± 0.011
Coughing	Music	0.768 ± 0.009	0.682 ± 0.011	0.722 ± 0.010
Coughing	Conversation	0.749 ± 0.011	0.655 ± 0.011	0.699 ± 0.011
Clearing Throat	Silence	0.725 ± 0.012	0.658 ± 0.011	0.690 ± 0.011
Clearing Throat	Music	0.820 ± 0.013	0.765 ± 0.012	0.791 ± 0.013
Clearing Throat	Conversation	0.783 ± 0.011	0.736 ± 0.009	0.759 ± 0.010
Swallowing	Silence	0.902 ± 0.015	0.865 ± 0.021	0.883 ± 0.018
Swallowing	Music	0.866 ± 0.006	0.790 ± 0.010	0.827 ± 0.008
Swallowing	Conversation	0.829 ± 0.013	0.752 ± 0.013	0.789 ± 0.013
Drinking	Silence	0.801 ± 0.007	0.717 ± 0.005	0.756 ± 0.006
Sniffing	Silence	0.832 ± 0.005	0.797 ± 0.006	0.814 ± 0.005
Sniffing	Music	0.822 ± 0.008	0.790 ± 0.008	0.805 ± 0.008
Sniffing	Conversation	0.818 ± 0.008	0.785 ± 0.008	0.801 ± 0.008
Speech	Silence	0.652 ± 0.008	0.599 ± 0.007	0.624 ± 0.008
Walking	Silence	0.548 ± 0.005	0.494 ± 0.005	0.519 ± 0.005
Running	Silence	0.456 ± 0.004	0.309 ± 0.003	0.368 ± 0.003

Table 2: Accuracy of S_1 heart sound localisation.

Activity	Noise Regime	Median Absolute Error / ms	Mean Absolute Error / ms	Median Percentage Error	Mean Percentage Error
Rest	Silence	4.01 ± 0.79	8.28 ± 1.30	6.14 ± 1.19	14.51 ± 2.66
Rest	Music	7.50 ± 6.10	19.18 ± 2.77	14.74 ± 9.37	39.25 ± 10.13
Rest	Conversation	7.78 ± 3.27	9.81 ± 1.43	13.26 ± 4.75	20.38 ± 3.10

Table 3: Accuracy of HRV estimates.

sounds split during deep breathing. Despite the poor results for segmentation during walking, the heart rate estimates for walking are reasonably accurate. This is because of the emission probability estimates for footsteps being high enough that the Viterbi algorithm takes them for a heart sound, if there is a plausible time difference between the footstep and an actual heart sound. This leads to the true S_1 sound being reported as the S_2 sound, and the S_2 sound not being identified, as it has a lower energy than the S_1 sound.

Heart Rate Variability: Table 3 only gives the results for HRV estimation when participants were at rest as HRV estimation requires accurate segmentation. To the best of the authors' knowledge, no empirical survey has assessed HRV accuracy for commercially available heart rate monitors which makes it difficult to compare. Values reported for HRV in healthy adults had an inter-quartile range of approximately 30ms [11]; the accuracy obtained by our approach is therefore likely sufficient for indicative readings.

Power Consumption and Latency: An STMicro Nucleo L496ZG-P board was used to run latency and power consumption experiments as it is a realistic microcontroller that could be used in a commercial product. We used the Teensy to reduce development time, but we do not expect a commercial device to be based upon it.

Latency: It took 3ms to extract 1 second of audio features, approximately 1.2ms to calculate 1 second of emission probabilities and 695ms to run Viterbi algorithm. Therefore, assuming overheads, the segmentation algorithm requires approximately 700ms of computation for 1s of audio.

Power Consumption: The biggest consumer of power was the microcontroller: 4.5mA during run-mode and 276 μA during sleep (with ADC enabled), at 3.3V. As a contact microphone is a passive

sensor it does not need external power, and amplification uses negligible power. Assuming a 3.7V 200mAh lithium-polymer battery and a safety margin of 0.7, the battery life is approximately *48 hours*.

6 CONCLUSION

Body sounds are an excellent source of information for understanding the state of the human body and are being increasingly explored in the area of digital health. We designed a wearable device to assess the viability of monitoring heart rate continuously using heart sounds. This work has explicitly considered the difficult challenges associated with our goal: *noisy measurements* and *limited computational and energy budgets*. Our work makes an important step towards enabling heart sounds to be monitored continuously in non-clinical conditions, and we believe that similar approaches can be adopted for other types of body sounds.

ACKNOWLEDGMENTS

REFERENCES

[1] 2019. High Sensitivity Piezo Film Sensor. https://www.te.com/usa-en/product-CAT-PFS0013.html
[2] 2019. NeuroKit. https://github.com/neuropsychology/NeuroKit.py
[3] 2019. Tympan. https://tympan.org/
[4] 2019. Welcome to BioSPPy - BioSPPy 0.6.1 Documentation. https://biosppy.readthedocs.io/en/stable/
[5] DanielChangMD. 2012. A Wiggers Diagram. https://commons.wikimedia.org/wiki/File:Wiggers_Diagram.svg
[6] Adib et al. 2015. Smart homes that monitor breathing and heart rate. In *CHI*. 837–846.
[7] Gillinov et al. 2017. Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. *Medicine and science in sports and exercise* 49, 8 (2017), 1697–1703.
[8] Hernandez et al. 2015. Biowatch: estimation of heart and breathing rates from wrist motions. In *PervasiveHealth*. 169–176.
[9] Kemper et al. 2007. Heart Rate Variability: Impact of Differences in Outlier Identification and Management Strategies on Common Measures in Three Clinical Populations. 62, 3 (2007), 337–342.
[10] Kwon et al. 2012. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In *EMBC*. 2174–2177.
[11] Keet et al. 2013. Short-term heart rate variability in healthy adults. *Anaesthesia* 68, 7 (2013), 775–777.
[12] Larson et al. 2011. Accurate and Privacy Preserving Cough Sensing Using a Low-Cost Microphone. In *UbiComp* (Beijing, China). 375–384.
[13] Nandakumar et al. 2015. Contactless sleep apnea detection on smartphones. In *MobiSys*. 45–57.
[14] Negar et al. 2016. A wearable device for continuous cardiorespiratory System Monitoring. In *LCN Workshops*. 230–235.
[15] Phan et al. 2015. Smartwatch: Performance evaluation for long-term heart rate monitoring. In *ISBB*. 144–147.
[16] Rahman et al. 2014. BodyBeat: A Mobile System for Sensing Non-Speech Body Sounds. In *MobiSys* (Bretton Woods, New Hampshire, USA). 2–13.
[17] Rubin et al. 2017. Recognizing Abnormal Heart Sounds Using Deep Learning. (2017). arXiv:1707.04642 [cs]
[18] Springer et al. 2014. Robust Heart Rate Estimation from Noisy Phonocardiograms. *Computing in Cardiology*, 613–616.
[19] Springer et al. 2016. Logistic Regression-HSMM-Based Heart Sound Segmentation. *IEEE Transactions on Biomed Eng* 63, 4 (2016), 822–832.
[20] Tavel et al. 2006. Cardiac Auscultation: A Glorious Past–and It Does Have a Future! *Circulation* 113, 9 (2006), 1255–1259.
[21] Wang et al. 2007. Phonocardiographic Signal Analysis Method Using a Modified Hidden Markov Model. *Ann Biomed Eng* 35, 3 (2007), 367–374.
[22] Yatani et al. 2012. BodyScope: a wearable acoustic sensor for activity recognition. In *UbiComp*. 341–350.
[23] Zhang et al. 2019. PDVocal: Towards Privacy-preserving Parkinson's Disease Detection using Non-speech Body Sounds. In *MobiCom*.
[24] Bernard J. Gersh. 2000. *Mayo Clinic Heart Book*.