Contactless Infant Monitoring using White Noise

Anran Wang, Jacob E. Sunshine, Shyamnath Gollakota
{anranw,jesun,gshyam}@uw.edu
University of Washington

ABSTRACT

White noise machines are among the most popular devices to facilitate infant sleep. We introduce the first contactless system that uses white noise to achieve motion and respiratory monitoring in infants. Our system is designed for smart speakers that can monitor an infant's sleep using white noise. The key enabler underlying our system is a set of novel algorithms that can extract the minute infant breathing motion as well as position information from white noise which is random in both the time and frequency domain. We describe the design and implementation of our system, and present experiments with a life-like infant simulator as well as a clinical study at the neonatal intensive care unit with five new-born infants. Our study demonstrates that the respiratory rate computed by our system is highly correlated with the ground truth with a correlation coefficient of 0.938.

CCS CONCEPTS

• **Applied computing** → **Life and medical sciences**; • **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*;

KEYWORDS

Computational Health; Smart Speakers; White Noise; Physiological Sensing; Active Sonar; Health and Wellness

ACM Reference Format:

Anran Wang, Jacob E. Sunshine, Shyamnath Gollakota. 2019. Contactless Infant Monitoring using White Noise. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom'19), October 21–25, 2019, Los Cabos, Mexico.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3300061.3345453

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MobiCom '19, October 21–25, 2019, Los Cabos, Mexico
© 2019 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.
ACM ISBN 978-1-4503-6169-9/19/10…$15.00
https://doi.org/10.1145/3300061.3345453

Figure 1: Infant monitoring at the Neonatal Intensive Care Unit (NICU) using smart speakers.

1 INTRODUCTION

Sleep plays an integral role in human health and is vitally important for neurological development in children, particularly infants [23, 34, 51]. Consumer sleep products that monitor the vital signs of infants are increasingly popular with parents [7, 9]. Infant monitors that track vital signs such as respiratory rates are frequently being used to monitor children less than one year of age because of their susceptibility to rare and devastating sleep anomalies [26]. The most frightening of these anomalies is Sudden Infant Death Syndrome (SIDS). SIDS is defined as the sudden death of a child less than one year of age usually occurring during sleep; it is the leading cause of death among children between one month to one year old in developed countries [28] and respiratory failure is believed to be a part of the final common pathway [32, 49].

A key problem with infant vital sign monitors, however, is their level of invasiveness. Indeed, these devices use specifically designed sensors and wires that require contact with the infant [5, 9, 12] or their sleep surface [3, 8]. A critical drawback of these contact-based systems is that they have lead to severe complications including rashes, burns and, in rare cases, death from strangulation [4]. Thus, a contactless means of monitoring breathing [35, 61] holds appeal as a safer and less invasive alternative.

In this paper, we ask the following question: can we enable *contactless* motion and respiratory monitoring in infants using white noise? White noise is commonly used for sleeping

infants in order to increase their stimulus threshold, thus allowing for more uninterrupted sleep [17]. Prior studies have shown that moderate amounts of white noise can be beneficial to sleep [39, 50] and have no long-term ill effects on health or hearing [53]. As a result, white noise machines are among the most popular devices to facilitate infant sleep.

A white noise machine that achieves contactless respiratory monitoring could improve sleep quality as well as potentially identify important anomalies in infant breathing. In addition to being contactless in nature, a single device such as a smart speaker (e.g., Amazon Echo) that can integrate both these functions would help reduce the number of monitoring devices as well as the associated cost, while improving sleep quality and potentially reducing the risk of sleep anomalies.

We present *BreathJunior*, the first contactless system that uses white noise to achieve motion and respiratory monitoring. We design algorithms built for smart speakers (e.g., Amazon Echo) that can monitor an infant's sleep using white noise. At a high level, the smart speaker emits white noise which gets reflected off the infant's body; these reflections arrive at the microphone array of the device which are then processed to extract the infant's body and minute chest motion. While prior active sonar systems use custom-designed signals (e.g., 18–20 kHz FMCW [41, 42]) to track breathing *in adults*, these frequencies are audible to infants [37, 52], making them inappropriate for infant sleep monitoring.[1] In contrast to white noise, long-term exposure to sound at these high frequencies in infants may also cause headache, nausea and temporary hearing loss [27, 48]. Thus, using white noise is an appealing approach for infant monitoring.

Achieving contactless infant monitoring using white noise is challenging for multiple reasons.

- White noise is by definition random in both the time and frequency domain. As a result, it is challenging to embed or extract useful information from random white noise signals.
- The signal strength of the reflections corresponding to breathing motion is proportional to the surface area of the chest. Infants not only have a much smaller torso but their chest displacement due to breathing is also much smaller compared to adults. Further, infants require a higher sampling rate as they breathe at a much higher rate (20-60 breaths per minute) compared to adults (12-20 breaths per minute).
- Finally, the white noise signal intensity should be low to minimize the risk of exceeding safe noise levels, yet at the same time the echoes still need to be detected reliably. In particular, while prior work transmits FMCW signals at 90 dB(A) [42], research has shown noise exposure level exceeding 75 dB(A) can cause sleep disturbance in infants [22, 45].

[1]The following link plays this audio for better illustration of the problem: https://youtu.be/tGHU515dFeE.

BreathJunior addresses the above challenges by making two key technical contributions. First, we design an acoustic receive beamforming algorithm that amplifies the minute reflections from the infant's chest by computing its direction with respect to the microphone array at the smart speaker. Our algorithm efficiently computes the infant's direction amongst N different angles using only $O(\log N)$ iterations (see §2.2.4).

We then localize the infant and track their breathing motion from the white noise reflections. To do this, we introduce a novel technique that transforms white noise into multiple FMCW signals at the receiver. Specifically, we prove that we can transform the received white noise reflections into N concurrent FMCW chirps, which are orthogonal in the frequency domain, while preserving the multi-path reflection information with negligible SNR loss (see §2.2.2). We demodulate these orthogonal FMCW chirps and decode the minute respiration motion and compute their distance from the smart speaker, by combining the phase information across the N FMCW chirps. We show that this method of combining phase across these N orthogonal FMCW chirps further increases the signal strength of the minute reflections from the infant.

We prototype our system using an off-the-shelf seven-microphone array which has an identical microphone layout and sensitivity to Amazon Echo Dot[2], but can output raw recorded signals. We first use SimNewB infant simulator [10] to systematically evaluate BreathJunior in various scenarios. SimNewB, co-created by the American Academy of Pediatrics, mimics the physiology of newborn infants, retails for around $25,000 and allows us to set the breathing rate as well as move various parts of the body. Our results show the following:

- Using 59 dB(A) white noise, our system estimates the breathing frequency within 95% and 90% of the baseline at distances of 0.5 m and 0.7 m respectively from the infant. These accuracies remain unaffected by clothing and for different orientations of the smart speaker.
- We can detect apnea (cessation of breathing for more than 15 seconds) with high accuracy. We can also detect body motion including arm or leg movements with a sensitivity and specificity of 95% and 100%.

Finally, we conduct a clinical study at a Neonatal Intensive Care Unit (NICU). We choose this environment because the infants are all connected to wired, hospital-grade respiratory monitors providing ground truth while they sleep. We recruited five infants, with consent from their parents, over the course of a month; recruitment is slow and difficult, given the state of the infants who are admitted to the NICU. We performed a total of seven sessions over a total duration of 280 minutes. Our study shows the following:

• The infants have the breathing rate between 35-65 breaths per minute (BPM) and in some rare instance as high as 70 BPM. The respiratory rate computed by BreathJunior is highly correlated with the baseline system, with an interclass correlation (ICC) of 0.938.

• Using the thresholds from the neonatal simulator experiments, we can identify the arm and leg movements as well as crying accurately with infants in the NICU.

Contributions. To summarize, the goal of our work is to provide a safe and accessible way to monitor infant respiration at home using commodity smartspeaker hardware. To this end, this paper makes four key contributions: (1) We introduce the first contactless system that uses white noise to achieve motion and respiratory monitoring. Using this we designed the first active sonar system that can track breathing in infants, (2) we design algorithms to extract the motion information as well as track the infant distance from random white noise signals. We also present algorithms that use the microphone array to beamform in the direction of the infant to extract the weak breathing signals, that are otherwise not detectable, (3) we evaluate our design using a hardware prototype and systematically evaluate it with the SimNewB infant simulator to understand the various tradeoffs, and (4) we perform a clinical study at the neonatal intensive care unit of a large medical center to demonstrate the feasibility of using our system to accurately track breathing and other movements using white noise in new-born infants.

2 BREATHJUNIOR

Fig. 2 shows the architecture of our system. The speaker transmits pseudo-randomly generated Gaussian white noise that gets reflected off the infant body and received by the circular microphone array. Our algorithms process the signals from all the seven microphones to increase the signal strength of the minute reflections from the infant's chest using receive beamforming algorithms. We then transform the received pseudo-random white noise reflections into five concurrent FMCW chirps, at the receiver, while preserving the multi-path reflection information. We then demodulate these chirps and decode the minute respiration motion by combining the information across the five chirps. To support beamforming and respiration detection, our algorithms also localize the position of the infant. Finally, using the received signals, our algorithms can also monitor body motion as well as detect audible baby sounds like crying using interference cancellation techniques.

In the rest of this section, we first describe white noise generation at the speaker and then explain different components in our receiver algorithm.

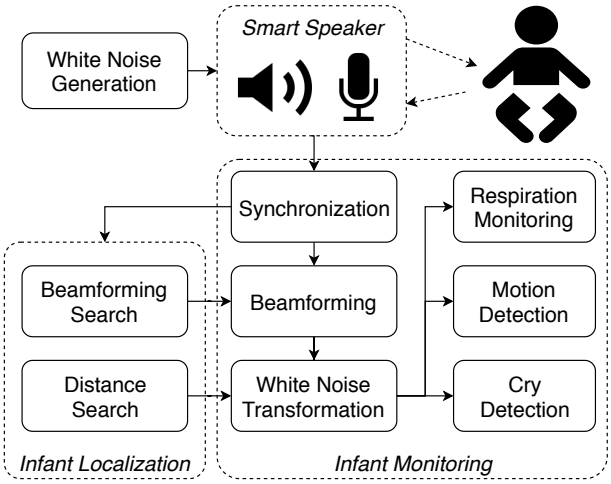


Figure 2: Different components in BreathJunior.

2.1 White Noise Generation at Speaker

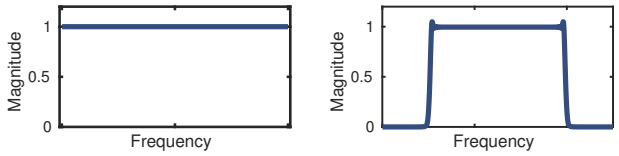
At the transmitter, we generate deterministic white noise using pseudo-random sequences with a known seed, such that it has a flat frequency response. To do this, we encode an impulse signal by shifting the phases of each of its frequency components by a random real sequence uniformly distributed in $[0, 2\pi]$.

The generated signal follows Gaussian white noise for two reasons. First, an impulse signal is flat in the frequency domain, and randomly changing the phase does not affect this. Second, the pseudo-random phase, denoted by ϕ_f, is independent and uniformly distributed in $[0, 2\pi]$. From the central limit theorem, suppose our sampling rate is r, and each time-domain sample, $\frac{1}{\sqrt{r/2}} \sum_{f=1}^{r/2} \exp(-j(2\pi f t + \phi_f))$, follows a normal distribution with a zero mean and constant variance when r is large enough, making it white noise.

In practice, we generate the signal as a stream of *blocks*, each of which has a constant duration. A long duration ensures that we can increase the SNR of the received signal using correlation but would limit the ability to monitor high breathing rates. We use a duration of $T = 0.2s$ and a sampling rate of 48000Hz; so, our frequency range is $1Hz$ to $f_{max} = 24000Hz$. We use a *Mersenne Twister* pseudo-random generator [38] to generate $f_{max}T$ different phase offsets for each block, and perform IDFT to convert it back into the time-domain, which is then played through the speaker:

$$S(t) = \sum_{f=1}^{f_{max}T} e^{-j(2\pi f \frac{t}{T} + \phi_f)} \quad (1)$$

where ϕ_f is the pseudo-randomly generated phase. We note that the same phase is added to the i and $f_{max}T - i$ frequencies so the IDFT results in a real signal.



(a) White noise **(b) FMCW signal**

Figure 3: The similarity of the frequency domain between white noise and FMCW signals.

2.2 Decoding Breathing at Microphone Array

2.2.1 Block-level Synchronization. The first step is to estimate the beginning of each transmitted white noise block as received by the microphone array. To do this, we re-generate the transmitted block using the same seed, at the receiver side. We then perform cross-correlation between the received signal using the center microphone in the array and the re-generated transmitted block. We then identify the peak in the cross-correlation result which corresponds to the direct path from the speaker to the microphone. We use the location of this peak as the start of the first block in the received signal. We need to synchronize once at the beginning as the speaker and all microphones share the same sampling clock.

Note that, we cannot extract respiration from cross correlation, because the sub-millimeter chest motion is much smaller than the granularity of a sample. Instead, we transform the pseudo-random white noise into FMCW signals at the receiver so that we can decode and extract the fine-grained multipath profile using FFT efficiently.

2.2.2 Transforming White Noise into Multi-FMCW. We describe how to transform the received white noise signal into a FMCW chirp. We then explain how to extract breathing motion from the FMCW chirp. Finally, we explain how to improve the SNR by transforming white noise into multiple concurrent FMCW chirps.

Transforming white noise to FMCW chirp. A key step in our receiver algorithm is that we can remove the randomness of the white noise by transforming it into FMCW chirps that can be efficiently decoded to track tiny motions, without losing information about the reflections. In other words, although the speaker transmitted white noise and the reflections from the infant motion correspond to white noise, we can transform the received signal to look like FMCW chirps played through the speaker and reflected off the infants body, rather than white noise.

Our intuition is that in the frequency-domain a FMCW chirp is approximately flat, as shown in Fig. 3. Further, within the FMCW frequency range, the transmitted white noise is also flat. Hence, we can in principle transform white noise in the desired frequency range into an FMCW chirp by shifting the phase of each frequency component of the received signal.

Specifically, consider we want to transform the received white noise block of duration T, within frequencies between f_0 to $f_0 + F$, into an FMCW chirp. We first generate a FMCW chirp template of that duration, $fmcw(t) = exp(-j2\pi(f_0 t + \frac{F}{2T}t^2))$. We then perform a DFT on this time window to get,

$$FMCW(f) = C\alpha_f e^{-j\psi_f} \tag{2}$$

where C is a constant and $\alpha_f \approx 1$. This gives us the phases ψ_f of each of its frequency components within $[f_0T, (f_0 + F)T]$. Since we also know the exact phases ϕ_f we used in the transmitted white noise block in Eq. 1, we can correct the phase of each frequency in the received white noise signal by $\phi_f - \psi_f$, within $[f_0T, (f_0 + F)T]$ to transform it into an FMCW chirp.

We mathematically show that this transform preserves the multi-path reflection information. In particular, in the presence of multiple paths, the received signal within the frequency range $[f_0T, (f_0 + F)T]$ can be written as, $w(t) = \sum_{p\in paths} A_p \sum_{f=f_0T}^{(f_0+F)T} e^{-j(2\pi f \frac{t-t_p}{T} + \phi_f)}$, where A_p and t_p are the attenuation factor and time-of-arrival of path p. Performing a DFT on $w(t)$ gives us, $W(f) = \sum_{p\in paths} A_p e^{-2\pi \frac{t_p}{T} f + \phi_f} = A'_f e^{-j\Phi_f}$. Our proposed phase transformation changes the phase of each frequency as follows, $\hat{\Phi}_f = \Phi_f - \phi_f + \psi_f$. We prove that this converts the white noise into a FMCW chirp without losing multipath information as follows:

$$\begin{aligned}
\hat{w}(t) &= \sum_{f=f_0T}^{(f_0+F)T} \sum_{p\in paths} A_p e^{-j(2\pi f \frac{t-t_p}{T} + \phi_f)} e^{-j(-\phi_f + \psi_f)} \\
&= \sum_{f=f_0T}^{(f_0+F)T} \sum_{p\in paths} A_p e^{-j(2\pi f \frac{t-t_p}{T} + \psi_f)} \\
&= \sum_{p\in paths} A_p \sum_{f=f_0T}^{(f_0+F)T} e^{-j(2\pi f \frac{t-t_p}{T} + \psi_f)} \\
&\approx \frac{1}{C} \sum_{p\in paths} A_p fmcw(t - t_p)
\end{aligned}$$

The final approximation is because $\alpha_f \approx 1$ in Eq. 2. Hence, the multipath reflections from the environment and the infant body in the received white noise signal are preserved after transformed into FMCW chirps. Note that this approximation introduces an SNR loss of around 0.05dB and a constant phase bias that does not affect the monitoring result.

Extracting breathing signal from FMCW. After the signal is transformed to a FMCW chirp, we can perform traditional FMCW demodulation to extract the breathing signal. To do this, we first multiply the received FMCW chirp

by a downchirp signal,

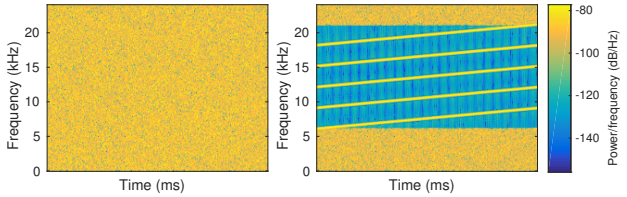
$$e^{-j2\pi(-f_0 t-\frac{F}{2T}t^2)} \sum_{p\in paths} e^{-j2\pi(f_0(t-t_p)+\frac{F}{2T}(t-t_p)^2)}$$
$$= \sum_{p\in paths} e^{-j2\pi(\frac{F}{T}t_p t+f_0 t_p-\frac{F}{2T}t_p^2)} \quad (3)$$

Next we perform an FFT on this signal, where each frequency bin corresponds to reflections at different distances. While this can be used to separate reflections from other environmental sources from that of the infant, it cannot be used to extract the minute breathing motion which has a resolution of a few millimeters (this is because of the resolution we get from Eq. 3 is limited by the bandwidth). However, the phase of each frequency component of the demodulated signal is also a function of distance. Specifically, from Eq. 3, the phase of the FFT bin corresponding to the time-of-arrival t_p is $f_0 t_p - \frac{F}{2T}t_p^2$. In other words, a tiny 1mm displacement will result in a significant 0.185 radian phase difference when $f_0 = 10000$Hz. Hence, we can track tiny motion even if it is much less than the theoretical FMCW resolution limit which is proportional to $\frac{c}{2B}$.

Thus, if we knew the round-trip distance d between the infant and the microphone array (we will discuss this distance estimation in §2.2.4), the FFT bin corresponding to this distance is $f_{resp} = \frac{d*F}{c*T}$, where c is the speed of sound. We extract the breathing signal by tracking the phase φ_i in this frequency bin for each ith demodulated chirp. Note that this phase sequence is confined to $[-\pi, \pi]$, causing sharp transitions from π to $-\pi$ or vice versa. To address this, we can simply compensate for the 2π phase shift by adding or subtracting a 2π when there is a more than π change between adjacent phase measurements.

Improving SNR with multi-FMCW chirps. One approach is to transform white noise into a single large FMCW chirp that spans the whole frequency range of the white noise transmission. A large band FMCW chirp has better spatial resolution because of more fine-grained frequency bins after demodulation and DFT. However, even when using the whole 24kHz band, the resolution is limited to 1.4cm, which is still much larger than the movement of the chest of an infant. On the other side, each FFT bin has less information and thus less SNR, making extracting respiration events difficult.

Instead of transforming the whole band into a single FMCW chirp, we split the band between 6 kHz to 21 kHz into five sub-bands, which are then transformed into five concurrent FMCW chirps independently. Chirp i has a starting frequency $f_0 = 3000 + 3000i$ Hz and bandwidth $F = 3000$ Hz. We get rid of those below 6kHz because of environmental noise, and those above 21 kHz because of low sensitivity. The spectrogram before and after transformation is shown in Fig. 4.



(a) Before transformation **(b) After transformation**

Figure 4: Transforming white noise to multi-FMCW chirps at the receiver.

By doing this, we trade-off resolution for SNR because each transformed chirp has less bandwidth. However, the same frequency bin of each of the five demodulated chirps corresponds to a same time-of-arrival (see Equation 3). Hence, we can fuse the five phases of each FFT bin from each demodulated chirp to improve SNR.

Recall from Eq. 3 that the phase of a FFT bin corresponding to the same time-of-arrival is linear to the beginning frequency of the FMCW chirp. Hence, we average the φ across the five FMCW chirps as

$$\varphi^{(avg)} = \frac{\sum_{i=1}^{5} \varphi^{(i)}/(3000+3000i)}{\sum_{i=1}^{5} 1/(3000+3000i)} \quad (4)$$

where $\varphi^{(i)}$ is the phase at the frequency bin corresponding to the respiration signal, f_{resp}, of the ith demodulated chirp. We use this phase value to extract the minute breathing motion with sufficient SNR.

2.2.3 Respiration, Motion and Cry Monitoring. From this phase data, we can extract minute breathing as well as coarse infant motion information.

- *Respiration rate monitoring.* We apply a finite impulse response (FIR) filter onto the phase sequence with a pass-band of $[0.4Hz, 1.1Hz]$. This corresponds to the normal range of an infant's respiration rate. We count the number of zero-crosses for the filtered signal and divide it by two to compute breathing rate.

- *Apnea detection.* To detect apnea which is a prolonged pause (more than 15 seconds) of the respiration, we first record the average amplitude A of the filtered phase signal during the initial one-minute localization duration. When a duration of 15 seconds has an average amplitude less than βA where β is a constant, we classify it as an apnea event. We empirically choose β using the infant simulator.

- *Motion detection.* The signal change due to movements of legs and hands is much larger than the movement from respiration. Fig. 5 shows the phase changes in the presence of body motion. The plot shows that because reflections from coarse body motion have more energy, we see a large variance in the phase information. Thus, if the total variance

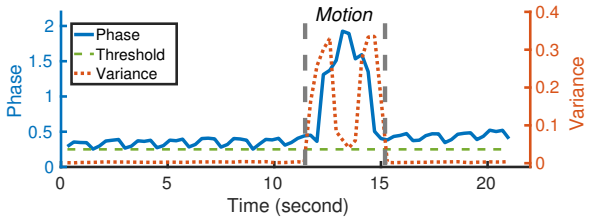


Figure 5: Phase changes resulting from body movements is significantly higher than from breathing.

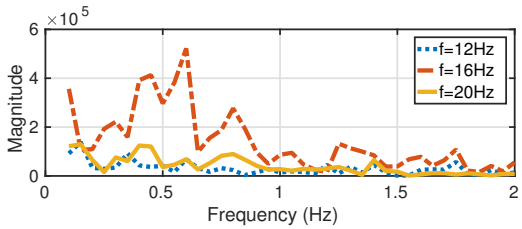


Figure 6: The FFT of H_f over different FFT bins, f.

within the last N phases exceeds a threshold, we classify it as body motion. We empirically choose the threshold using the simulator. Note that because the positions of legs and hands are not far from the chest, their movement leads to interference to the respiration signals. As a result, the system does not monitor respiration during motion periods.

- *Crying and sound detection.* Ideally, we would like to detect crying and other sounds from the infant in the presence of the white noise. We note that infant crying sounds are typically loud in comparison to the white noise signal generated by our smart speaker. We can further improve sound detection by calculating the difference between two adjacent chirps across time. Any sound from the infant will superimpose onto the white noise. The transformation procedure, while transforming white noise into chirps, will transform crying and other sounds into noise signals. Hence, two adjacent chirps will be different, especially at low frequencies. We calculate the L2 norm of the difference between two adjacent transformed chirps, $p(S_{i-1}, S_i) = ||S_{i-1} - S_i||_2^2$. If the value exceeds a threshold, and it occurs frequently within a short time period, the system would classify it as infant sounds. Note that most of the sound from other people in the environment are reduced in amplitude due to the beamforming process described next.

2.2.4 Infant Localization and Beamforming. The above discussion assumes that we know the distance of the infant relative to the smart speaker and hence know the frequency bin, f_{resp}, corresponding to the breathing motion. In this section, we first describe how to localize the infants and identify their distance from the smartspeaker. We then explain how we perform receive beamforming on the microphone array to increase the SNR of the breathing reflections.

Initial Distance computation. After computing an FFT on the FMCW chirps, we find the most likely FFT bin that corresponds to respiratory motion. To this end, we store the complex value of each FFT bin, f of the demodulated chirp, H_i. For each FFT bin f, we perform another FFT over the complex values across all the chirps within the first minute of tracking. We then calculate the SNR_{resp} for each bin f, defined as the energy within $[0.4Hz, 1Hz]$ (corresponding to breathing rates of 20-60 breaths/min) divided by the energy above 1 Hz. This SNR is a good indicator of the quality of the respiration signal in FFT bin f. We select the lowest frequency bin (*i.e.*, nearest to the microphone array), that has a peak SNR comparable with its neighboring frequency bins. We denote this frequency bin as f_{resp}. The round-trip distance between the infant and the smart speaker can then be estimated as $\frac{f_{resp}Tc}{F}$ from Eq. 3. For example, Fig. 6 shows the FFT of H_f on different bins, f, in one of our experiments. We see that FFT bin $f = 16Hz$ has the peak SNR with more energy within $[0.4Hz, 1Hz]$. This bin corresponds to the distance from the infant. Note that at this stage, we could not yet use the accurate phase-based algorithm from §2.2.3, as it assumes that the distance to the infant is known. Further, we have not yet performed beamforming to increase the SNR of the infant reflections.

Receive Beamforming algorithm. Now that we have an initial estimate of the distance, we design a receiver-side beamforming algorithm to suppress other static reflections and increase the SNR of the weak reflections from the infant. At a high level, the signals captured by the seven microphones on the array are added together using the appropriate delays. Suppose we know the angle α of the infant relative to the smart speaker, the delays Δ_i could be calculated based on the angle, α, as, $\Delta_i = ||P_i - P_0||sin(\alpha)$, where P_i is the location of the ith microphone. We can then calculate the beamforming signal $R(t) = \sum_{i=1}^{7} R_i(t - \Delta_i)$, where $R_i(t)$ is the sample at time t received on microphone i.

So the key question is: *how do we find the angle of the infant with respect to the microphone array?* A naïve solution is to exhaustively search over all the possible angles to find the best angle that maximizes the signal strength of the respiratory signal. This however is computationally expensive. Instead, we utilize the wide-band nature of white noise, and design a multi-step beamforming method based on a ternary-search algorithm that progressively reduces both the search range as well as beam width to compute the infant's direction.

We leverage the following property of acoustic beam widths: a signal transmitted from a microphone array at a frequency f has a beam width proportional to $sin^{-1}\frac{C}{f}$, where C is a constant [24]. Said differently, at the higher acoustic frequencies,

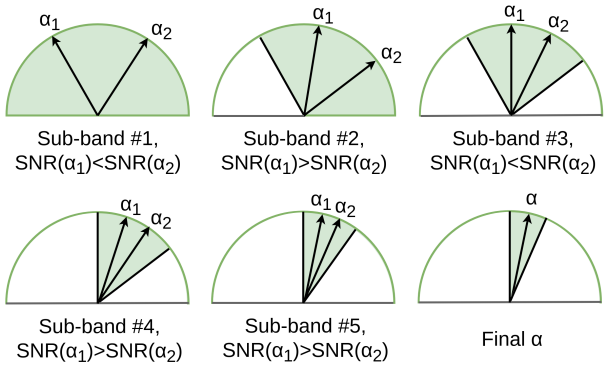


Figure 7: The progressive ternary search algorithm for beamforming search. Green area is the search scope.

a narrower beam width is achieved while beamforming. As a result, we can design a divide and conquer algorithm that starts at the lower frequencies, eliminates directions for the infant and use the higher frequencies to increase the beam resolution and narrow in on the direction of the infant.

Following the above intuitions, we go through the five multi-FMCW chirps from 6kHz to 21kHz ordered from low to high frequencies. For each FMCW chirp, we maintain an angle scope, $[\gamma_l, \gamma_r]$, which we initialize to $[-\pi/2, \pi/2]$ for the first FMCW chirp. For the ith FMCW chirp ($i = 1 \cdots 5$), we sequentially set the beamforming angle α to two values of $\alpha_1 = (2\gamma_l + \gamma_r)/3$ and $\alpha_2 = (\gamma_l + 2\gamma_r)/3$. At each of these two beamforming angles, we use the method in §2.2.2 to transform beamformed white noise into the demodulated FMCW signal. We then estimate the distance of the infant using the algorithm in §2.2.4 and calculate the SNR of the respiratory signal, as defined earlier, for the two angles, $SNR(\alpha_1)$ and $SNR(\alpha_2)$. If $SNR(\alpha_1) < SNR(\alpha_2)$, we narrow down the search scope to $[\alpha_1, \gamma_r]$; otherwise, we narrow down the angle search scope to $[\gamma_l, \alpha_2]$. We then move to the higher frequency FMCW chirp and do the above processing again, until we reach the highest FMCW chirp, where we finalize α to be the middle of the search scope. The five steps of the above algorithm are illustrated in Fig. 7. This adaptive beamforming method drastically increases the SNR and the operational range by up to 2x (see §3).

Computational complexity. In comparison to an exhaustive search over N angles, the above ternary-search algorithm reduces the complexity to $O(\log N)$. Further, this beamforming angle search and distance estimation is only done once at the beginning of the tracking process to compute the distance and angle of the infant with respect to the device. We use this distance and angle for the duration of infant monitoring. If we lose the breathing or motion signal for more than 30 seconds, we re-initiate the search process to find the new distance/angle of the infant. If neither the breathing nor the motion signal is found after the search, we can raise an alarm to the caregiver.

2.2.5 Addressing Practical Issues. Finally, we describe in detail the practical issues we addressed in our system.

- **Combating inter-block interference**. One problem when we generate white noise in blocks is the interference between adjacent blocks. Specifically, the latter parts of the echoes of the previous block can be superimposed over the beginning of the current block. Because each block is encoded using different random seeds, these inter-block interference signals are transformed into noise and can reduce sensitivity. To address this issue, for each block, we introduce a *guard interval* at the beginning of each block consisting of a *cyclic prefix*. This is similar to the cyclic prefix used in OFDM transmissions. Specifically, for each white noise block, we insert a guard interval at the beginning of each block, consisting of the last g samples of that block. g is picked to be larger than the maximum possible propagation duration. In our test, a duration of 0.1s is found to be sufficient. To maximize randomness, the duration is also randomly selected between 0.1s and 0.15s, known to both the transmitter and receiver.
- **Adaptive sub-band weighting**. Empirically, the frequency response across a large band can change because of the propagation properties, hardware imperfections and environmental noise. Specifically, lower frequencies attenuate slower than higher frequencies [33]. Further, our microphone array has a $5-10dB$ dip around $12kHz$. Finally, the environmental noise is larger at lower frequencies. To account for these effects, we assign different weights to different frequencies. Specifically, we use the SNR_{resp} of each sub-band, described in §2.2.4, as the weights to each of the five chirps in our multi-FMCW signal. Now, instead of giving equal weights to each of the five FMCW chirps, we modify Eq. 4 to compute a weighted average, $\varphi_f^{(fused)} = \frac{\sum_{i=1}^{5} w_i \varphi_f^{(i)}/(3000+3000i)}{\sum_{i-1}^{5} w_i/(3000+3000i)}$, where w_i is the respiratory signal SNR for the i^{th} FMCW chirp.
- **Adaptive speaker volume adjustment.** A problem with existing white noise machines is that the volume of their speaker cannot be adjusted with different distances to the infant. A fixed volume is challenging because the sound pressure could be either too high if the infant is close to the speaker or too low to be effective at larger distances. To address this, we adjust the speaker volume to be dependent on the distance from the smart speaker and the infant. Specifically, we use the distance estimate in §2.2.4 to adjust the white noise volume. To do this, we empirically found that the attenuation was 5.5dB when the distance from the infant doubles. A user can set a preferred at-ear volume (*e.g.*, 56dB). During monitoring, BreathJunior adaptively re-adjusts the

volume using the estimated distance and the corresponding attenuation values.

3 EVALUATION

We implement BreathJunior using a smart speaker prototype, built with a MiniDSP UMA-8-SP USB microphone array [13], which is equipped with 7 Knowles SPH1668LM4H microphones. They are of identical layout as well as sensitivity as an Amazon Echo Dot [2]. We connect it to an external speaker (PUI AS07104PO-R), and 3D-printed a plastic case that holds the microphone array and speaker together. The microphone array is connected to a Surface Pro laptop. We play dynamically generated pseudo-random white noise and record the 7-channel recordings, using XT-Audio library [14]. We capture the acoustic signals at a sampling rate of 48kHz and 24 bits per sample.

Next, we evaluate the effectiveness and accuracy of BreathJunior. We first conduct extensive experiments with a tetherless newborn simulator. The simulator, designed to train physicians on neonatal resuscitation, mimics the physiology of newborn infants. We systematically evaluate the effect of different parameters, including recording position, orientation and distances, at-ear sound pressure level, interference from other people, respiration strength and rate. We then recruit five infants at a Neonatal Intensive Care Unit (NICU) and conduct a clinical study to verify the validity of our system on monitoring respiration, motion and crying.

3.1 Neonatal simulator experiments

Because of the experimental difficulty and potential ethical problem of placing a wired ground truth monitor on a healthy sleeping infant, we first use an infant simulator (SimNewB®, Laerdal, Stavanger, Norway [10]), co-created by the American Academy of Pediatrics, that mimics the physiology of newborn infants. SimNewB is a tetherless newborn simulator designed to help train physicians on neonatal resuscitation and is focused on the physiological response in the first 10 minutes of life. It comes with an anatomically realistic airway and supports various breathing features including bilateral and unilateral chest rise and fall, normal and abnormal breath sounds, spontaneous breathing, anterior lung sounds, unilateral breath sounds and oxygen saturation. These life-like simulator mannequins, which retail >$25,000, are used to train medical personnel on identifying vital sign abnormalities in infants, including respiratory anomalies. SimNewB is operated and controlled by SimPad PLUS, which is a wireless tablet. We are able to control various parameters of the simulator including a) respiration rate and intensity; b) limb motion; and c) sound generation. We use this to evaluate different aspects of BreathJunior's performance.

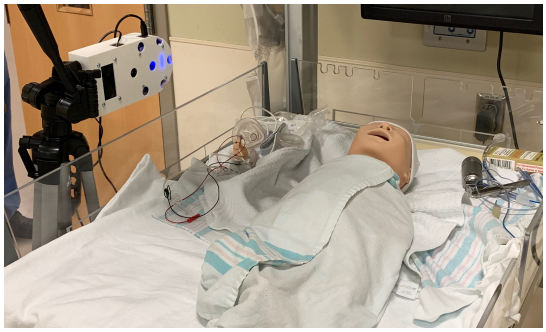


Figure 8: Setup with a neonatal simulator.

Specifically, we perform experiments in the simulator lab at the University of Washington medical school where we put the infant simulator in a 26 inch x 32 inch bassonette by one of the walls shown in Fig. 8. We put the smart speaker prototype on a stand that can adjust the orientation, and put the stand on a table which can adjust its position around the crib. We set its height to 10 cm above the simulator so that the rails of the bassonette will not obstruct the path between the prototype and the simulator.

3.1.1 Effect of distance, orientation and position. We evaluate the effect of the smart speaker distance, orientation and position on the breathing rate accuracy.

Effect of the smart speaker position. We first measure the effect of the smart speaker position with respect to the infant on breathing rate accuracy. To do this, we place the smart speaker hardware in four different positions around the bassonette: left, right, front and rear. This effectively evaluates the effect of placing the smart speaker at different sides of a crib. We place the smart speaker at different distances from the chest of the infant, from 30 cm to 60 cm. At each of the distances, we set the infant simulator to breathe at a breathing rate of 40 breaths per minute, which is right in the middle of the expected breathing rate for infants. As the default, we set the sound pressure to be 56 dB at the infant's ear. The smart speaker transmits the white noise signal and we record the acoustic signals for one minute, which we then use to compute the breathing rate. We repeat this experiment ten times.

Fig. 9 plots the results of these experiments. The plots show the following key trends: First, the average computed respiratory rate across the distances up to 60 cm is around 40 breaths per minute, which is the configured breathing rate of the infant simulator (shown by the dotted line). Second, the position of the smart speaker does not significantly affect the breathing error rate. The only exception is when the smart speaker is placed at the rear, where we have slightly higher variance in the measured breathing rate. This is because there is more obstruction from the abdomen and legs. Finally, as expected, the variance in the measured breathing rate

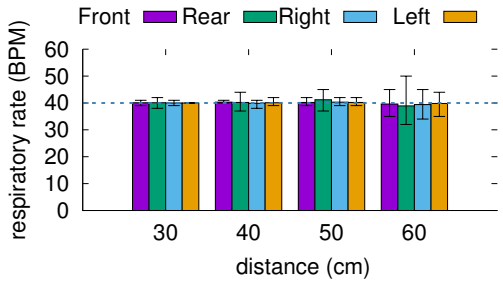


Figure 9: Respiration rate accuracy with different placement locations of the microphone array at 56 dB(A). Error bars represent the min-max interval.

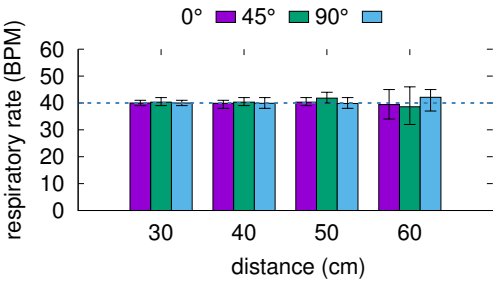


Figure 10: Respiration rate accuracy with different angles of the microphone array at 56 dB(A).

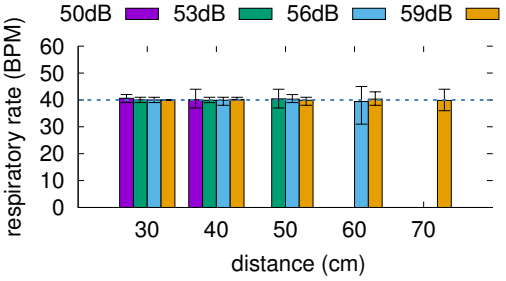


Figure 11: Accuracy of computing respiratory rate with different at-ear sound pressures.

increases with distance. Specifically, the mean absolute error is around 3 breaths per minute when the smart speaker is at a distance of 60 cm, compared to 0.4 breaths per minute at a distance of 40 cm. This is because the reflections from the infant's breathing motion attenuate with distance.

Effect of smart speaker orientation. Next, we run experiments with three different smart speaker orientations. This allows us to evaluate the effectiveness of beamforming as a function of the smart speaker angle. We set the breathing rate of the simulator to 40 BPM and vary the distance of the smart speaker from the infant's chest. We also set the at-ear sound pressure to 56 dB. Fig. 10 shows the detected respiration rates using the three orientations as a function of distance, where 0° is when the microphone array faces the simulator and 90° is when the microphone array faces the ceiling. The plots show that there is no significant difference in the respiratory rate variance across the three orientations. This is because the microphone array is designed to be omni-directional to detect sound across all angles.

3.1.2 Effect of volume, respiration rate & intensity. Next, we evaluate the effect of sound volume, respiration rate and intensity on breathing rate accuracy.

Effect of sound volume. The higher the sound volume from the smart speaker, the better the reflections from the infant breathing motion. However, our target is to keep the white noise volume to be under 60 dB at-ear to be conservatively safe. Here, we evaluate the effect of different at-ear white noise volumes. Specifically, we change the white-noise volume to be between 50-59 dB(A). As before we change the distance between the smart speaker and the infant simulator between 30-70 cm and measure the breathing rate using the white noise reflections at each of these volume levels. The smart speaker is placed at the left and 0° with respect to the infant. As before, we repeat the experiment ten times to compute the mean and variance in the estimated breathing rate while the simulator is set to a breathing rate of 40 breaths per minute.

Fig. 11 shows the results for these experiments. The plots show that when the at-ear sound volume is around 56 dB(A), we achieve low variance in the breathing rate up to distances of 50 cm. When we increase the white noise volume at the infant by 3 dB to 59 dB(A), the breathing rate can be estimated with low variance from a distance of up to 70 cm. This is expected since the reflections from the breathing motion are stronger when the white noise volume is higher.

Effect of respiration rate and intensity. Next, we evaluate the accuracy of the system with varying respiration rates as well as the intensity of each breath. For a typical infant less than one year old, the respiration rate is less than 60 breaths per minute. So, we evaluate the accuracy by varying the breathing rate of the infant simulator between 20-60 breaths per minute. To verify the robustness, we also change the intensity of each breath on the simulator to two different settings: normal and weak. The weak intensity is triggered by a simulated respiratory distress syndrome (RDS), an ailment that can be experienced by infants and particularly those born prematurely. We set the distance of the infant simulator from the smart speaker to 40 cm and the speaker is placed at the left and at 0°.

Fig. 12 shows the results of these experiments with the smart speaker-computed breathing rate as a function of the simulator breathing setting. We also note the results for the two intensity settings. The plots show that we see higher

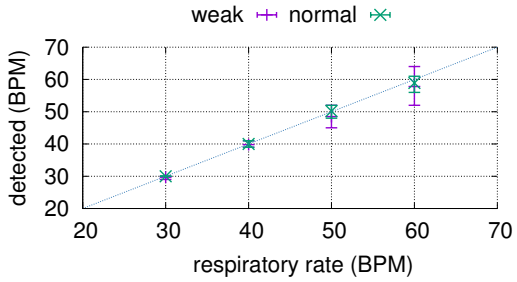


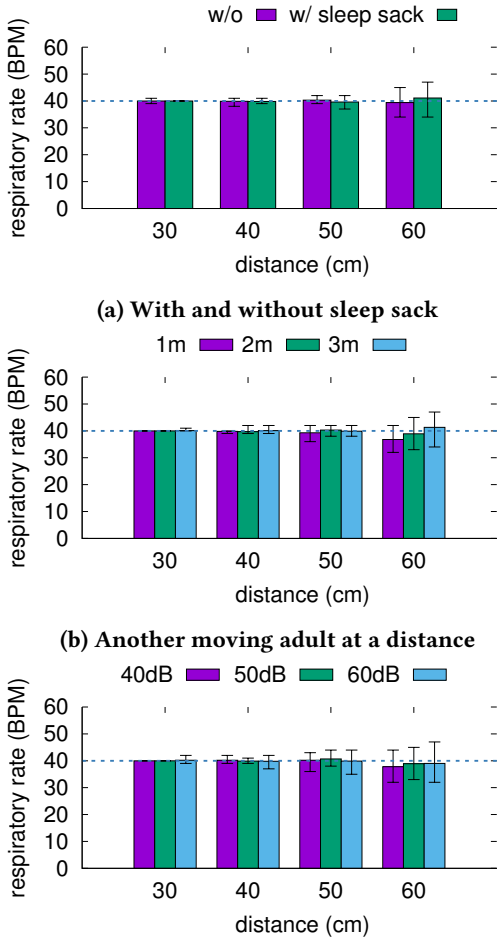
Figure 12: Accuracy *w.r.t.* breathing intensity.

variance in the computed breathing rate as we increase the breathing rate. This is because, as the breathing rate increases, we see more changes within the received signal, which requires higher sampling rates to get the same error resolution. In our implementation, we set the block of each white noise signal to 0.2 s. Thus, as the breathing rate increases, we see less blocks per each breath, which effectively reduces the number of samples per breath, which in turn introduces more errors. As expected, we also see more variance in weak breath situations associated with respiratory distress syndrome. This is because lower intensity results in smaller phase change, resulting in a lower SNR.

3.1.3 Effect of clothes and interference. Finally, we evaluate the effect of blankets and other interfering motion and environmental noise in the environment.

Effect of clothes. We use a typical infant one-piece sleep sack made of cotton which is provided with the simulator to help trainees learn the correct method for putting on this garment that helps swaddle the baby. We repeat the experiments with and without the sleep sack. We run experiments by placing the smart speaker to the left of the infant simulator and at an angle of 0^o, while setting the simulator to breathe at a rate of 40 breaths per minute. We change the distance between the simulator and the smart speaker and compute the breathing rate. Fig. 13a shows the respiratory rate as a function of distance. The plots show that the presence of sleep sack does not significantly affect the breathing rate accuracy. We further evaluate BreathJunior with human infants who are swaddled in blankets in §3.2 and show that it can track their breathing motion.

Effect of interference. The above experiments are all done when an adult is sitting about three meters away from the crib. To further assess if the interference from other people would affect the accuracy, we additionally did the same experiments with an adult sitting at consecutively closer distances. As shown in Fig. 13b, we cannot see much difference except when the distance between the adult and the smart speaker is 1 meter, while the distance between the simulator



(a) With and without sleep sack

(b) Another moving adult at a distance

(c) Ambient interfering sound

Figure 13: Effect of clothes, interference and ambient sound with white noise at 56 dB(A).

and the smart speaker is 60 cm, since the small distance difference leads to spectrum leakage in the FFT of the FMCW demodulation. However, BreathJunior could still extract a breathing rate at this distance.

Effect of ambient noise. We evaluate the effect of ambient noise by playing a clip of pop music using a smartphone placed two meters away from the crib. We set the volume so that the measured sound pressure at the crib is around 40 dB(A), 50 dB(A) and 60 dB(A) respectively. We then turn on the smart speaker playing white noise at 56 dB(A) and report the respiration rate accuracy in Fig. 13c. We see no obvious effect for the ambient noise between 40 and 60 dB(A). This is because frequencies below 6 kHz are filtered out during our white noise transformation algorithm. Further, white noise can be thought of as wide-band spread spectrum which can be resilient to structured acoustic signals like music.

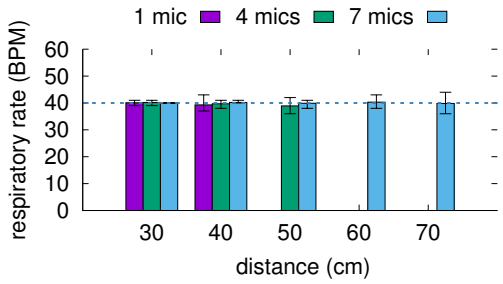


Figure 14: Respiration accuracy *w.r.t.* beamforming.

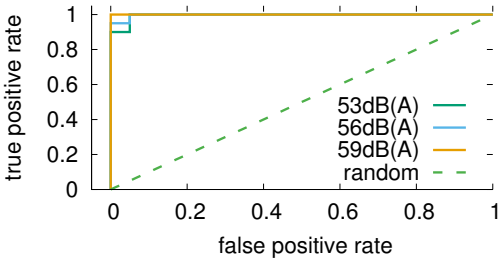


Figure 15: Apnea event detection ROC curves.

3.1.4 Effect of receive beamforming. Here, we quantitatively evaluate the benefits of using receive beamforming. As before, we run experiments by placing the smart speaker to the left of the infant simulator and at an angle of 0^o, while setting the simulator to breathe at a rate of 40 breaths per minute. We keep at-ear sound pressure at 59dB and change the distance of the smart speaker and the infant simulator and collect the data on the smart speaker. We then extract the breathing signals using a) only a single center microphone on the smart speaker without using our receive beamforming algorithm; b) four microphones on the top and bottom of the smart speaker; and c) all seven microphones to decode the signal. We plot the three results in Fig. 14. The plot shows that receive beamforming improves the range by approximately 1.75x — without beamforming, BreathJunior with a single microphone can support up to 40 cm range, whereas receiver beamforming with seven microphones improves the range to 70 cm. Moreover, while using four microphones reduces the variance in the estimated respiratory rate it does not significantly increase the distance compared to using only a single microphone without beamforming.

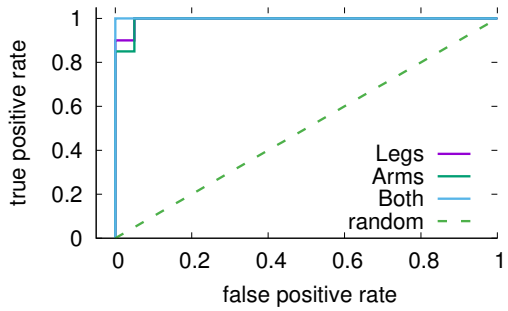
3.1.5 Apnea, motion and sound detection. Here we evaluate BreathJunior's ability to identify apnea events, body motion as well as audible sound.

Apnea detection. An apnea event is defined as a 15-second respiratory pause [18]. While it is difficult to run experiments with human infants that also have apnea events, we can simulate them on our infant simulator. Specifically, we simualte a 15 second central apnea event by remotely pausing the respiration of the infant simulator and resuming it after 15 seconds. We use the thresholding method in §2.2.3 to detect the presence of an apnea event during the 15 second. We use the 15-second duration before the apnea event where the infant simulator breathes normally to evaluate the false positive rate (FP). We place the smart speaker 50 cm left of the simulator at an angle of zero degree. The simulator is set to breathe at a rate of 40 breaths per minute. We repeat this experiment 20 times to generate the receiver operating characteristic (ROC) curve by different values of the threshold by computing the sensitivity and specificity of the algorithm in identifying apnea events. Fig. 15 shows the ROC curves when we vary the volume of white noise between 50-59 dB(A). As expected, the accuracy improves at higher volume.

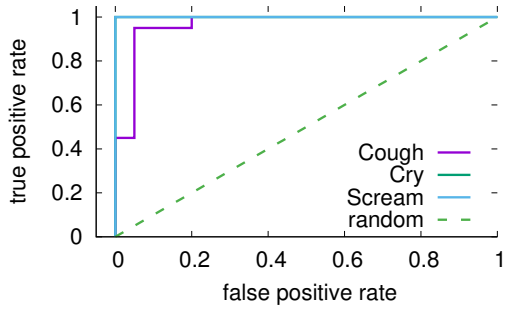
Motion detection. Next, we evaluate BreathJunior's ability to detection body movements such as hand and leg motion. We can remotely control the infant simulator to move its arms and legs. Specifically, for each movement, the arm or leg rotates around the shoulder joint away from the body for an angle of approximately 30°, than rotates back to its original position. Each movement takes approximately two seconds. We perform each of these movements 20 times and record the true positive events. Like before, we also use 20 2-second clips of normal breathing motion under the same condition. We set the distance between the infant simulator and the smart speaker to 50 cm and set the simulator to breath at 40 breaths per minute.

Fig. 16a shows the ROC curves for each of the three movements: arm motion, leg motion and arm+leg motion. The AUC for the three movements was 0.9925, 0.995 and 1 respectively. The plots show that BreathJunior's accuracy for motion detection is high. For instance, the operating point for arm motion had an overall sensitivity and specificity of 95% (95% CI: 75.13% to 99.87%) and 100% (95% CI: 83.16% to 100.00%), respectively. This is expected because these movements reflect more power than the minute breathing motion and hence can be readily identified.

Sound detection. Finally, we evaluate BreathJunior's ability to detect infant audible sounds. The infant simulator has an internal speaker that plays realistic recorded sounds of infant crying, coughing and screaming, which are frequent sounds from infants. The volume is to set to be similar to an infant sound. As before, we record 20 2-second clops of each sound type and use 20 2-second clips where the simulator was breathing but was silent. The infant simulator was set to breathe at 40 BPM and the distance from the smart speaker was 60 cm. Fig. 16b shows the ROC curves for each of the three infant sounds. The area under the curve (AUC) for detecting the three sounds was 1, 0.965, 1 respectively.



(a) Motion detection



(b) Sound detection

Figure 16: Motion/sound detection ROC curves.

3.2 Clinical Study with Infants

The American Academy of Pediatrics strongly recommends against any wired systems in an infant's sleep environment, making ground truth collection of respiratory signals on *healthy infants* at home unsafe and potentially ethically challenging [1]. To overcome this challenge, we conduct clinical studies at the Neonatal Intensive Care Unit (NICU) of a major medical center. The vast majority of infants in this NICU are born prematurely (i.e., before 38 weeks gestation). We choose this environment because the infants are all connected to wired, hospital-grade respiratory monitors providing respiratory rates while they sleep in their bassinets. Each infant is treated in individual bassinets in a separate room, where their parents and nurses are also sitting around 1.5 meters away from the bassinet, most of the time. We recruited five infants, with consent from their parents, over the course of a few months. This study was approved by University of Washington's Institutional Review Board and followed all the prescribed criteria.

Clinical study setup. Since infants at this age sleep intermittently between feedings, our recording sessions ranged from 20 minutes to 50 minutes. All infants, because they were in the NICU, were connected to hospital grade respiratory monitoring equipment (Phillips LTD). Fig. 1 shows the setup with our study smart speaker. The smart speaker prototype is placed outside the crib to ensure safety, and the distance between the prototype and the monitored infant is kept between 40-50cm. We ensure that the at-ear sound pressure is 59dB(A). We performed a total of 7 sessions over a total duration of 280 minutes. Of these, the nurses or parents were interacting or feeding the infant for 62 minutes. We perform our algorithms over the remaining 218 minutes.

Infant	Total session	Total duration	Effective duration	Sleep duration
1	1	40min	33min	20min
2	3	125min	90min	63min
3	1	40min	35min	9min
4	1	30min	20min	11min
5	1	45min	40min	33min

Table 1: Statistics across the recruited infants.

Respiratory rate accuracy. We could access respiratory rate measurements from the Phillips hospital system with minute-to-minute granularity. We synchronize the clocks between the logging computer in the hospital and our laptop to align the start of each minute. Note that the precision of the respiratory rate from the Phillips system is 1 BPM, and we use it as ground truth and compare the error of our system with it. Our breathing rate experiments had infants with a minimum weight of 3.5 kg and a maximum weight of 4.5 kg. This is within the weight range for our target application population of normal infants above the age of 1 month. Note that BreathJunior only monitors breathing when the infant is not moving. We note that while infants can move their limbs to varying degrees in the post-natal period, they are generally unable to roll over (back-to-front) until approximately 6 months of age [6]. Further, when the infant is moving or crying the ground truth breathing rate signal is also affected. So we focus on the time duration when the infant is not moving or crying but is either stationary or sleeping.

Fig. 17 shows the respiratory rates detected by our system compared to that reported by the groundtruth. The plot shows multiple key trends.

- Unlike adults, the respiratory rate for infants is significantly higher. In the NICU, the population is typically premature babies, many of whom have respiratory problems, often with breathing rates above 35 BPM and in some instances as high as 70 BPM.
- At a breathing rate above 65 BPM, we see larger errors. This is expected because the various parameters in our system are designed for a maximum breathing rate of 60 BPM and for non-NICU infants. This limitation can be further addressed, and the algorithm improved, by using a combination of shorter block duration and a band-pass filter that adaptively adjusts its pass band to the frequency range of the

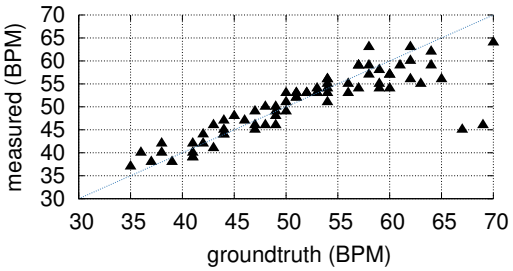
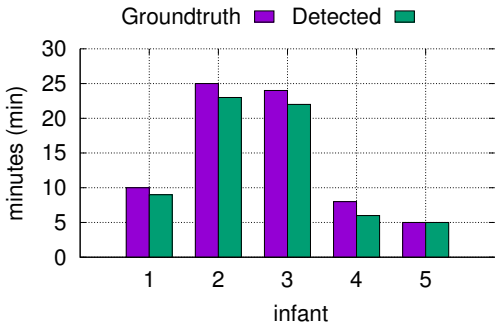


Figure 17: Comparison between respiratory rate from BreathJunior and groundtruth with infants at NICU.

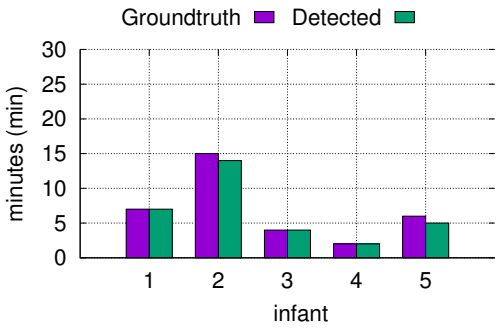
respiration. We also note that these respiration rates were observed in atypical infants (*i.e.*, born prematurely, under weight, or with underlying respiratory problems hence their admission in an NICU).

- The respiratory rate computed by BreathJunior is highly correlated with the baseline — the interclass correlation (ICC) between them was 0.938.

Motion and crying detection accuracy. Finally, we compare BreathJunior's motion and sound detection capabilities with the ground truth. We used the threshold values from the simulator experiments which gave us the best sensitivity and specificity (top-left points of Fig. 16a and Fig. 16b) for this purpose. We manually note the duration, on a minute resolution, when the infant is crying and moving; we use this as the ground truth for these experiments. Figs. 18 show the results for both the ground truth as well as BreathJunior for both body movements (e.g., arms/legs) as well as crying, for each of the five infants. The figures show that there is a good correlation with the ground truth.



(a) Movement



(b) Sound

Figure 18: Accuracy for detecting motion as well as sounds with infants at NICU.

4 RELATED WORK

Physiological monitoring solutions. Wired vital sign monitors are traditionally used for both hospital and home use [44]. By definition, they require physical contact of the sensors with the infant's body or on their sleep surface. These sensors include pulse oximeters [31], and thoracic impedance monitors [19]. A critical drawback of these wired systems is that they may interrupt sleep and can lead to severe complications including death from strangulation [4]. More recently, wireless wearable solutions are being designed to track vital signs. These require wearables in contact with the infant body including smart socks [9], wristbands [7] or other probes [21, 60] which track the heart rate of infants. Sleep surfaces embedded with sensors have also been designed for tracking physiological signals [3, 8]. All these solutions however require contact with the infant body. In contrast, our design is the first contactless solution that uses white noise, which can facilitate sleep, to track breathing and other infant movements.

There has also been a renewed interest in designing contactless solutions that utilize cameras and radar [35, 61]. [40] uses cameras to recognize respiration and heart rate, however cameras are sensitive to light conditions especially during sleep. [61] use ultra-wideband radar to track respiration and heart rate in adult participants. [11, 56] use millimeter wave radar to track heart rate and respiration in infants. Radar solutions however require specialized hardware and ultra wide bandwidth which are not available on existing Wi-Fi radios or smart speakers. [35] uses WiFi signals to track respiration in adult participants. Wi-Fi based breathing monitoring has not yet been demonstrated for infants. Further, Wi-Fi based tracking solutions are prone to interference from other moving objects given their long range and are affected by ambient Wi-Fi data transmissions [29, 46]. We take an alternate approach that operates at short-range using white noise as an active sonar system.

Acoustic sensing. Acoustic signals are widely studied for motion tracking and localization because of their slow propagation speed and ease of use in commodity devices. [43, 55, 57] track finger motion, [15, 25, 58] track gestures using acoustic signals. Acoustic signals are also used to track devices [59] such as smartwatches and smartphones using a

microphone array; [36, 54] tracks smartphones using multiple speakers. Acoustic reflections have also been used for detecting middle ear fluid using smartphones [20].

The closest to our approach is prior work on active sonar that uses 18-20 kHz acoustic transmissions from a phone speaker to track breathing in adult participants for diagnosis of sleep apnea [42] and opioid overdose [41]. [47] uses sound between 17-19 kHz to detect respiration as well as heart rate. While adults generally cannot hear 18–20 kHz acoustic signals, infants have much better sensitivity compared to adults at higher frequencies up to 20 kHz [37, 52], which makes those high frequency sounds potentially audible and thus inappropriate for infant sleep monitoring. Long-term exposure to ultrasound in infants may also cause headache, nausea and temporary hearing loss [27, 48]. Our approach differs in three key ways: 1) we explore the use of microphone arrays on smart-speaker devices such as Amazon echo to achieve contactless respiratory monitoring and 2) we use white noise as a signal source and develop algorithms to extract the breathing motion from reflections of these white noise transmissions and 3) we show for the first time that an active sonar system can be used for tracking the minute breathing motion from infants.

5 CONCLUSION AND DISCUSSION

We present a contactless solution that can monitor infants using white noise. From a clinical utility perspective, there are several potential use cases for a smart speaker-based respiratory monitor. These include respiratory rate monitoring for the purposes of identifying early signs of incipient infection, non-invasive monitoring for respiratory changes of chronic diseases (e.g., asthma, COPD, congestive heart failure), non-invasive monitoring of older kids with epilepsy or recurrent central apneas, and even monitoring for the purposes of wellness. All these are areas that require further inquiry. The use case presented here is compelling for children and parents because it provides two functionalities: white noise to facilitate sleep and respiratory monitoring. And the system can do these tasks at low cost, using a commodity smart speaker. While the use of consumer infant vital sign monitoring devices is a source of debate [16], these systems remain a fixture among many parents who make a conscious choice to monitor their children while they sleep.

There are a few studies about the effects of noise on infants as well as adults. Although 50 dB(A) is recommended for a hospital nursery, there is significant related work that notes that there is no known negative consequences of white noise exposure as long as the sound pressure is less than 75 dB(A) [22, 45]. For adults, the WHO recommends a noise limit of 85dB (A) on an average of 8 working hours. White noise machines currently on the market have an average noise level of 63.3 dB (A) at a distance of 2 m [30]. As a result, 59 dB(A) is considered safe and within normal limits for a clinical as well as home environment.

While we focus on white noise, using other noise types including pink noise, brown noise and natural sounds (*e.g.*, raindrops, fan noise) is worth exploring as well. Further, we may use shorter block duration to support higher respiratory rates greater than 65 BPM and combine them with adaptive filters that dynamically infer the range of respiratory rates. Finally, BreathJunior achieves an operational range of 0.7 m using white noise with 59 dB(A) at-ear sound pressure. However, longer ranges can be achieved using microphones with higher sampling rate and bit resolution. Further, our breathing experiments were limited to a minimum infant weight of 3.5 kg. Evaluating the system with infants with lower weight is a worthwhile research direction.

6 ACKNOWLEDGMENTS

We thank Vikram Iyer, Mehrdad Hessar, Ali Najafi, Justin Chan, Rajalakshmi Nandakumar, Nick Mark and our shepherd Krishna Chintalapudi for feedback on the manuscript. The authors are funded in part by NSF award CNS 1812559. The authors also hold equity in Sound Life Sciences Inc.

REFERENCES

[1] Sids and other sleep-related infant deaths: Updated 2016 recommendations for a safe infant sleeping environment. Pediatrics 138, 5 (2016).
[2] Amazon Echo Dot 2nd Generation. https://www.amazon.com/All-New-Amazon-Echo-Dot-Add-Alexa-To-Any-Room/dp/B01DFKC2SO, 2019.
[3] Angelcare Movement Sound Monitor. https://www.amazon.com/Angelcare-Movement-Sound-Monitor-White/dp/B00GU07FLQ, 2019.
[4] Angelcare recalls baby monitors after 2 deaths. https://www.cnn.com/2013/11/22/health/baby-monitor-recall/index.html, 2019.
[5] Baby Vida Oxygen Monitor. https://www.amazon.com/Baby-Vida-Oxygen-Monitor-White/dp/B00VBI42HM, 2019.
[6] CDC's Developmental Milestones. https://www.cdc.gov/ncbddd/actearly/milestones/index.html, 2019.
[7] Fitbit Official Site for Activity Trackers. https://www.fitbit.com/home, 2019.
[8] New Babysense 7. https://www.amazon.com/New-Babysense-Under-Mattress-Non-Contact/dp/B075XQHMVT, 2019.
[9] OwletCare - Baby Monitor. https://owletcare.com/, 2019.
[10] SimNewB. https://www.laerdal.com/us/doc/88/SimNewB, 2019.
[11] Single chip radar sensors with sub-mm resolution. https://www.xethru.com/, 2019.
[12] The Best Movement Monitor. https://www.babygearlab.com/topics/health-safety/best-movement-monitor, 2019.
[13] UMA-8-SP USB Microphone Array. https://www.minidsp.com/products/usb-audio-interface/uma-8-sp-detail, 2019.
[14] XT-Audio. https://sjoerdvankreel.github.io/xt-audio/, 2019.
[15] Aumi, M. T. I., Gupta, S., Goel, M., Larson, E., and Patel, S. Doplink: Using the doppler effect for multi-device interaction. In Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (2013), UbiComp '13.

[16] Bonafide, C. P., Jamison, D. T., and Foglia, E. E. The emerging market of smartphone-integrated infant physiologic monitors. Jama 317, 4 (2017), 353–354.

[17] Borkowski, M. M., Hunter, K. E., and Johnson, C. M. White noise and scheduled bedtime routines to reduce infant and childhood sleep disturbances. The behavior therapist (2001).

[18] Brouillette, R. T., Fernbach, S. K., and Hunt, C. E. Obstructive sleep apnea in infants and children. The Journal of pediatrics 100, 1 (1982), 31–40.

[19] Brouillette, R. T., Morrow, A. S., Weese-Mayer, D. E., and Hunt, C. E. Comparison of respiratory inductive plethysmography and thoracic impedance for apnea monitoring. The Journal of pediatrics 111, 3 (1987), 377–383.

[20] Chan, J., Raju, S., Nandakumar, R., Bly, R., and Gollakota, S. Detecting middle ear fluid using smartphones. Science Translational Medicine 11, 492 (2019).

[21] Chung, H. U., Kim, B. H., Lee, J. Y., Lee, J., Xie, Z., Ibler, E. M., Lee, K., Banks, A., Jeong, J. Y., Kim, J., et al. Binodal, wireless epidermal electronic systems with in-sensor analytics for neonatal intensive care. Science 363, 6430 (2019), eaau0780.

[22] GÄDEKE, R., DÖRING, B., KELLER, F., and VOGEL, A. The noise level in a childrens hospital and the wake-up threshold in infants. Acta Pædiatrica 58, 2 (1969), 164–170.

[23] Gómez, R. L., Bootzin, R. R., and Nadel, L. Naps promote abstraction in language-learning infants. Psychological science 17, 8 (2006), 670–674.

[24] Goodwin, M. M., and Elko, G. W. Constant beamwidth beamforming. In 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing (1993), vol. 1, IEEE, pp. 169–172.

[25] Gupta, S., Morris, D., Patel, S., and Tan, D. Soundwave: using the doppler effect to sense gestures. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2012), ACM, pp. 1911–1914.

[26] Hall, K. L., and Zalman, B. Evaluation and management of apparent life-threatening events in children. American family physician 71, 12 (2005).

[27] Hanson, M. A. Health effects of exposure to ultrasound and infrasound: report of the independent advisory group on non-ionising radiation, 2010.

[28] Heron, M. P. Deaths: leading causes for 2010.

[29] Huang, D., Nandakumar, R., and Gollakota, S. Feasibility and limits of wi-fi imaging. In Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems (New York, NY, USA, 2014), SenSys '14, ACM, pp. 266–279.

[30] Hugh, S. C., Wolter, N. E., Propst, E. J., Gordon, K. A., Cushing, S. L., and Papsin, B. C. Infant sleep machines and hazardous sound pressure levels. Pediatrics 133, 4 (2014), 677–681.

[31] Kamlin, C. O. F., Dawson, J. A., O'donnell, C. P., Morley, C. J., Donath, S. M., Sekhon, J., and Davis, P. G. Accuracy of pulse oximetry measurement of heart rate of newborn infants in the delivery room. The Journal of pediatrics 152, 6 (2008), 756–760.

[32] Kinney, H. C., and Thach, B. T. The sudden infant death syndrome. New England Journal of Medicine 361, 8 (2009), 795–805. PMID: 19692691.

[33] Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. Fundamentals of acoustics. Fundamentals of Acoustics, 4th Edition, by Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, James V. Sanders, pp. 560. ISBN 0-471-84789-5. Wiley-VCH, December 1999. (1999), 560.

[34] Krueger, J. M., Rector, D. M., Roy, S., Van Dongen, H. P., Belenky, G., and Panksepp, J. Sleep as a fundamental property of neuronal assemblies. Nature Reviews Neuroscience 9, 12 (2008), 910.

[35] Liu, X., Cao, J., Tang, S., Wen, J., and Guo, P. Contactless respiration monitoring via off-the-shelf wifi devices. IEEE Transactions on Mobile Computing 15, 10 (2016), 2466–2479.

[36] Mao, W., He, J., and Qiu, L. Cat: high-precision acoustic motion tracking. In Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (2016), ACM, pp. 69–81.

[37] Mari, U., Kaoru, A., and Hironobu, T. How high-frequency do children hear? Inter-noise (2016).

[38] Matsumoto, M., and Nishimura, T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation (TOMACS) 8, 1 (1998), 3–30.

[39] Messineo, L., Taranto-Montemurro, L., Sands, S. A., Oliveira Marques, M. D., Azabarzin, A., and Wellman, D. A. Broadband sound administration improves sleep onset latency in healthy subjects in a model of transient insomnia. Frontiers in neurology 8 (2017), 718.

[40] Nam, Y., Kong, Y., Reyes, B., Reljin, N., and Chon, K. H. Monitoring of heart and breathing rates using dual cameras on a smartphone. PloS one 11, 3 (2016), e0151013.

[41] Nandakumar, R., Gollakota, S., and Sunshine, J. E. Opioid overdose detection using smartphones. Science translational medicine 11, 474 (2019), eaau8914.

[42] Nandakumar, R., Gollakota, S., and Watson, N. Contactless sleep apnea detection on smartphones. In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services (2015), ACM, pp. 45–57.

[43] Nandakumar, R., Iyer, V., Tan, D., and Gollakota, S. Fingerio: Using active sonar for fine-grained finger tracking. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (2016), ACM, pp. 1515–1525.

[44] of Pediatrics, A. A., et al. Apnea, sudden infant death syndrome, and home monitoring. Pediatrics 111 (2003), 914–917.

[45] Philbin, M. K. The influence of auditory experience on the behavior of preterm newborns. Journal of Perinatology 20, S1 (2000), S77.

[46] Pu, Q., Gupta, S., Gollakota, S., and Patel, S. Whole-home gesture recognition using wireless signals. In Proceedings of the 19th Annual International Conference on Mobile Computing & Networking (New York, NY, USA, 2013), MobiCom '13, ACM, pp. 27–38.

[47] Qian, K., Wu, C., Xiao, F., Zheng, Y., Zhang, Y., Yang, Z., and Liu, Y. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. In IEEE INFOCOM 2018-IEEE Conference on Computer Communications (2018), IEEE, pp. 1574–1582.

[48] Repacholi, M. H. Ultrasound: characteristics and biological action, vol. 19244. National Research Council of Canada, NRC Associate Committee on Scientific âĂę, 1981.

[49] Siren, P. M. A., and Siren, M. J. Critical diaphragm failure in sudden infant death syndrome. Upsala journal of medical sciences 116, 2 (2011), 115–123.

[50] Stanchina, M. L., Abu-Hijleh, M., Chaudhry, B. K., Carlisle, C. C., and Millman, R. P. The influence of white noise on sleep in subjects exposed to icu noise. Sleep medicine 6, 5 (2005), 423–428.

[51] Touchette, É., Petit, D., Séguin, J. R., Boivin, M., Tremblay, R. E., and Montplaisir, J. Y. Associations between sleep duration patterns and behavioral/cognitive functioning at school entry. Sleep 30, 9 (2007), 1213–1219.

[52] Trehub, S. E., Schneider, B. A., Morrongiello, B. A., and Thorpe, L. A. Developmental changes in high-frequency sensitivity: Original papers. Audiology 28, 5 (1989), 241–249.

[53] Wachman, E. M., and Lahav, A. The effects of noise on preterm infants in the nicu. Archives of Disease in Childhood-Fetal and Neonatal Edition 96, 4 (2011), F305–F309.

[54] Wang, A., and Gollakota, S. Millisonic: Pushing the limits of acoustic motion tracking. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2019), CHI '19, ACM, pp. 18:1–18:11.

[55] Wang, W., Liu, A. X., and Sun, K. Device-free gesture tracking using acoustic signals. In Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (2016), ACM, pp. 82–94.

[56] Yang, Z., Pathak, P. H., Zeng, Y., Liran, X., and Mohapatra, P. Monitoring vital signs using millimeter wave. In Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing (2016), ACM, pp. 211–220.

[57] Yun, S., Chen, Y.-C., Zheng, H., Qiu, L., and Mao, W. Strata: Fine-grained acoustic-based device-free tracking. In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (2017), ACM, pp. 15–28.

[58] Zhang, C., Waghmare, A., Kundra, P., Pu, Y., Gilliland, S., Ploetz, T., Starner, T. E., Inan, O. T., and Abowd, G. D. Fingersound: Recognizing unistroke thumb gestures using a ring. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 3 (2017), 120.

[59] Zhang, C., Xue, Q., Waghmare, A., Jain, S., Pu, Y., Hersek, S., Lyons, K., Cunefare, K. A., Inan, O. T., and Abowd, G. D. Soundtrak: Continuous 3d tracking of a finger using active acoustics. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 2 (2017), 30.

[60] Zhang, J., Chen, D., Zhao, J., He, M., Wang, Y., and Zhang, Q. Rass: A portable real-time automatic sleep scoring system. In Real-Time Systems Symposium (RTSS), 2012 IEEE 33rd (2012), IEEE, pp. 105–114.

[61] Zito, D., Pepe, D., Mincica, M., Zito, F., Tognetti, A., Lanatà, A., and De Rossi, D. Soc cmos uwb pulse radar sensor for contactless respiratory rate monitoring. IEEE Transactions on Biomedical Circuits and Systems 5, 6 (2011), 503–510.