BuMA: Non-Intrusive Breathing Detection using Microphone Array

Kaiyuan Hou
Columbia University
New York, NY, USA
Kaiyuan.Hou@columbia.edu

Stephen Xia
Columbia University
New York, NY, USA
stephen.xia@columbia.edu

Xiaofan Jiang
Columbia University
New York, NY, USA
jiang@ee.columbia.edu

ABSTRACT

Breath monitoring is important for monitoring illnesses, such as sleep apnea, for people of all ages. One cause of concern for parents is sudden infant death syndrome (SIDS), where an infant suddenly passes away during sleep, usually due to complications in breathing. There are a variety of works and products on the market for monitoring breathing, especially for children and infants. Many of these are wearables that require you to attach an accessory onto the child or person, which can be uncomfortable. Other solutions utilize a camera, which can be privacy-intrusive and function poorly during the night, when lighting is poor. In this work, we introduce BuMA, an audio-based, non-intrusive, and contactless, breathing monitoring system. BuMA utilizes a microphone array, beamforming, and audio filtering to enhance the sounds of breathing by filtering out several common noises in or near home environments, such as construction, speech, and music, that could make detection difficult. We show that BuMA improves breathing detection accuracy by up to 12%, within 30cm from a person, over existing audio filtering algorithms or platforms that do not leverage filtering.

CCS CONCEPTS

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

KEYWORDS

beamforming, real-time systems, embedded systems

ACM Reference Format:
Kaiyuan Hou, Stephen Xia, and Xiaofan Jiang. 2022. BuMA: Non-Intrusive Breathing Detection using Microphone Array. In *1st ACM International Workshop on Intelligent Acoustic Systems and Applications (IASA'22), July 1, 2022, Portland, OR, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3539490.3539598

1 INTRODUCTION

Millions of people suffer from sleep apnea, where the person stops breathing for long periods of time during sleep. This is a serious disorder that has detrimental effects on a person's overall quality of life. Numerous studies link severe sleep apnea to an increased risk of death. Over the course of five years, researchers [1] discovered that people with sleep apnea had a considerably elevated risk of sudden cardiac death. The danger was most significant for those aged 60 and older with moderate to severe sleep apnea (20 episodes an hour). When their oxygen saturation levels fell below 78%, the risk rose by an additional 80%. Additionally, those with severe sleep apnea have a two to fourfold increased risk of irregular cardiac rhythms compared to individuals without sleep apnea. Other researchers[2] found that some people with sleep apnea are more than 2.5 times more likely to pass away due to a heart attack between 12 a.m. and 6 a.m. than those without sleep apnea.

Sleep apnea also commonly occurs in infants. Small preterm infants are most likely to have infant sleep apnea. It sometimes occurs in larger preterm or full-term infants. During the first month after birth, sleep apnea occurs in 25% of infants who weigh less than 5.5 pounds. The risk increases to 84% for infants who weigh less than 2.2 pounds. One particular cause of concern for parents is sudden infant death syndrome (SIDS), where an infant passes away during sleep due to defects in the brain that controls breathing and waking up from sleep [3].

A system that can detect long episodes where no breathing occurs could potentially help. For example, such a system could alert parents of children who stop breathing while sleeping for long periods of time. There are several different technologies that have been developed for monitoring breathing and sleeping. The first technology is touch sensing, which detects body movement through vibration sensors and is used by many existing products to detect long periods where no breathing occurs. To improve detection accuracy, many of these products require the user to directly attach them to the body, which can cause uncomfort and affect sleep quality. Camera-based methods are another set of technologies, which analyzes video streams to infer movement, respiration rate, and other vital signs. Camera-based methods are generally very accurate, but are not as privacy-sensitive as a lower fidelity sensing modality like vibration. Additionally, camera-based methods generally require good lighting, which is usually not satisfied at night when most people are sleeping.

In this work, we present BuMA, an audio-based, non-contact, and real-time breathing monitoring system using microphone array, built on top of the Raspberry Pi ecosystem. We do not claim nor show that our system is an accurate system for detecting or curing sleep apnea; rather, BuMA is a system for detecting periods where breathing is not occurring. BuMA can be placed in a variety of different places to monitor breathing while sleeping, including on a nearby table, clipped onto a baby crib, or hanging off of a dream catcher. We take an audio-based approach because audio does not require good lighting, is more privacy-sensitive, and does

IASA'22, July 1, 2022, Portland, OR, USA

ACM ISBN 978-1-4503-9403-1/22/07…$15.00
https://doi.org/10.1145/3539490.3539598

not require a device to be in contact with the user. However, normal breathing sounds tend to be soft and low-energy, and there can be many noises in the home or nearby environment that could overpower that of breathing, such as passing vehicles, loud construction sounds, or speech.

BuMA is an improved version of AvA, an audio filtering platform to both enhance breathing sounds while filtering out noises from the environment that may overpower breathing sounds. AvA is an audio filtering platform that allows developers to supply models and detectors of specific sounds they want to enhance (e.g., breathing) or filter out. However, there could be a wide range of different noises in the environment that could overpower breathing sounds; AvA requires users to manually specify different sounds to filter out. To address this challenge BuMA incorporates a multi-class noise detector to dynamically select different noises currently present in the environment for AvA to filter out. We evaluate BuMA in a case study and show that BuMA outperforms other state-of-art filtering schemes for improving breathing detection by up to 12%.

We make the following contributions:

- We introduce BuMA, an audio-based system that detects and monitors breathing. BuMA consists of a edge processing unit that utilizes a six element microphone array to better detect breathing by enhancing breathing sounds and filtering out common noises in or near the home environment, including construction, speech, and music. BuMA accomplishes this in real-time by leveraging both spatial filtering (e.g., beamforming) and data-driven filtering.
- We create a cloud-based system for dynamically selecting noises in the environment to filter out. This system utilizes a multi-class noise detector to periodically detect the composition of overpowering noises in the environment and leverages specific sound models of the detected noises to filter them out.
- We conduct a study and find that BuMA can detect breathing with up to 12% more accuracy than using other methods of audio filtering, or using no filtering at all.

2 RELATED WORKS

2.1 Sleep and Breathing Monitor Platforms

There are a variety of breathing monitoring products and applications commercially available, especially for monitoring infants. Sense-U Baby Breathing Monitor [4] is wearable system that tracks a baby's breathing, movement, temperature, rollover and sleeping position and alerts parents when there are any alarming signs (e.g., the baby stops breathing for long periods of time). This system is contact-based and needs to be clipped onto the diaper to perform its full range of functions.

A second class of commercial products are monitoring systems that use cameras. Miku is one such product that uses a camera to monitor breathing and baby movements [5]. Though this device does not require contact, it costs hundreds of dollars and is more privacy intrusive because of its use of video. Another popular product is Cloud Baby Monitor, which is a mobile application that allows both parents and children to see each other simultaneously through a mobile device [6]. This product also utilizes audio to allow parents to speak with their children, but does not perform audio processing to measure or monitor the child. Vision-based methods have two major weaknesses. First, it is difficult to accurately monitor the state of the child at night when the room is dark. Second, these devices are privacy intrusive because there is a camera continuously watching the child while the parent is away.

Breathing monitoring is also a major research topic of interest. Researchers have explored breathing detection and monitoring using a wide range of sensing modalities. Methods that utilize vibration sensors [7, 8] require the sensor to be installed into the mattress or on the ground where the person sleeps. Though not directly in contact with the person, the vibration sensors are still in very close proximity to the person and cannot easily be moved to another room, space, or bed. There are also a variety of vision-based methods, including video [9], RFID [10], and radar [11]. As mentioned previously, vision-based methods provide rich information about the environment and are more privacy intrusive. In this work, we take an audio-based approach because it allows us to develop a remote sensing solution, while also preserving more privacy.

There are sound-based platforms for sensing breathing [12–14], but many of them require very close proximity (e.g. almost touching the mouth) or need the sensor to be attached to the body. Microphones, although more privacy-sensitive than cameras, can still observe non-breathing sounds occurring in other parts of the home or room. Unlike previous works, we take an audio filtering approach to filter out other potentially privacy-sensitive sounds in the environment, while retaining breathing sounds.

2.2 Audio Filtering

In many applications, other overpowering noises in the environment may greatly impact the detection of the sounds of interest. For example, sounds of nearby construction, passing vehicles, and music can easily be louder than the soft breathing sounds we are interested in observing. Additionally, microphones can record other non-breathing sounds in the environment that could be personal or private, such as speech and conversations. To help improve the detection of breathing and reduce the presence of other non-breathing and potentially privacy-sensitive sounds, we take an audio filtering approach. There are several works that have taken advantage of audio filtering to improve the detection of target sounds in noisy environments [15], such as improving the detection of vehicles for urban safety applications [16] and reducing the detection of privacy sensitive signals, like speech [17].

There are two broad categories of audio filtering works: spatial filtering and content-based filtering. Spatial filtering techniques incorporate different observations in space by installing microphones in various locations. This category includes beamforming [18–20], blind source separation (BSS) [21–23], and adaptive filtering techniques [24–26]. These methods do not take into account the environment's content or sound types, and often require prior knowledge of source locations. If the location of sources are not available in advance, systems that leverage spatial filtering will need to incorporate sound source localization algorithms to detect and localize sounds of interest. On the other hand, content-based filtering often requires only one microphone. These methods, such as deep neural networks (DNN) [27–29], use trained models of specific sounds to filter them out. Because they are trained on specific sounds, it is

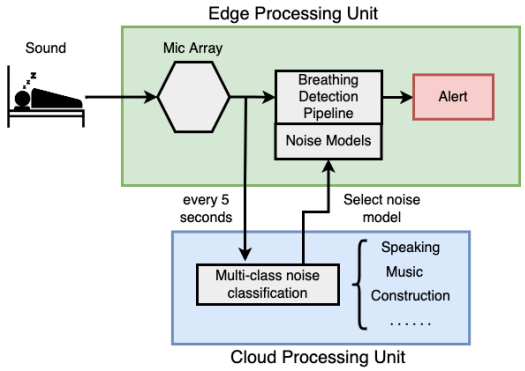


Figure 1: BuMA's system architecture.

difficult to train a single low-power model that can account for all of the possible noises that could exist in our application. To leverage the strengths of both classes of work, while minimizing their weaknesses, we leverage both strategies of filtering.

3 SYSTEM DESIGN

In this section, we introduce the system design of BuMA. Figure 1 shows the system architecture of BuMA, which is made up of two major components: an embedded edge and a cloud component.

First, we introduce the embedded processing unit, which consists of an array of microphones attached to a processing unit that users can place near the sleeping person or clip onto the crib of a baby to monitor breathing. Because sounds of normal breathing are generally soft and low in energy, there are a wide range of other sounds, such as music, speech, or construction sounds, that could overpower the sound of breathing, making detection much more difficult. As such, to perform robust breathing detection on this embedded edge processing unit, we integrate AvA [30], a filtering platform that combines both spatial filtering with content-driven filtering. AvA allows applications and users to select specific sounds in the environment to enhance and noises to filter out using sound models that developers may have created for their own applications.

However, there could be a wide range of noises in the environment, and it may not be enough to filter out one type of sound in the environment. For instance, at one point there could be construction noises in the background, and in the next instance, the construction may stop, but someone could be having a very personal conversation nearby. AvA does not inherently have a mechanism to change which type of sound to filter. To address this challenge, we incorporate a cloud processing component that periodically detects the presence of different types of noises in the environment. Then, the cloud processing component downloads and installs different models onto the edge processing unit to filter out present noises in the environment.

3.1 Edge Processing

Figure 2 shows the full breathing detection pipeline. First, we sample a window of audio from the microphone array. To perform robust breathing detection, we aim to filter out specific noises from the environment that could overpower the sounds of breathing, while also enhancing breathing sounds. To accomplish this, we incorporate the AvA platform [30]. AvA jointly performs spatial filtering and data-driven filtering by combining traditional adaptive beamforming to enhance sounds coming from specific directions, while utilizing data-driven sound detectors and models to enhance specific sounds and filter out specific noises from the environment. AvA accomplishes this utilizing an algorithm called content-informed beamforming (CIBF).

To perform the spatial filtering component of CIBF, which beamforms to a specific direction, AvA also incorporates a localization module that localizes significant sources in the environment and utilizes these directions to enhance sounds arriving from a specific direction. To perform the data-driven filtering component of CIBF, developers can supply their own models and detectors of sounds to enhance or noises to filter out.

After filtering out noises and enhancing breathing sounds, we need to detect if breathing sounds are present. We directly utilize the breathing detector we supply to AvA to perform breathing enhancement and noise filtering. To perform breathing detection, we extracted mel-frequency cepstral coefficients (MFCC), a common feature used in many audio applications. These features are then input into a Support Vector Classifier (SVC) with RBF kernel. To train this model, we extract 96 10-second clips of breathing sounds from the Google Audioset dataset [31], and 92 10-second random clips of non-breathing sounds. We use the same parameters and model type (SVC) to train sound detectors for each of three noises (construction, speech, and music).

We build our edge processing unit on a Raspberry Pi 3B+ and use the ReSpeaker Circular Array as our microphone array [32]. This microphone array is a 6 element circular array. All components of our breathing detection pipeline are run on this platform.

One of the shortcomings of AvA is that developers need to select the specific noise to filter out by specifying a model of a sound to detect, but there is no automatic method to switch between different noise models. There could be a wide range of different types of sounds in the environment that could overpower breathing sounds. To account for this challenge, we incorporate a mechanism for dynamically swapping noise models, described in the next section.

3.2 Cloud Processing

Ideally, BuMA should use sound models and detectors of specific noises currently present in the environment to filter them out using AvA. Because loud noises currently present in the environment could change over time, there needs to be a mechanism to select which noises to filter out. To accomplish this, we use a multi-class sound detector to detect what noises are present in the environment at regular intervals. Whenever a new noise is detected in the environment, then the sound model of the newly detected noise is used to filter it out.

Our multi-class sound detector is a gradient boosting classifier that classifies input sounds into 3 classes of potentially different noises that could be commonly found in or near a home environment: speech, music and construction. We extract mel-frequency cepstral coefficients (MFCC) to use as input features. We train our multi-class sound detector to dynamically swap out noise models using audio clips extracted from Google Audioset [31]. For each

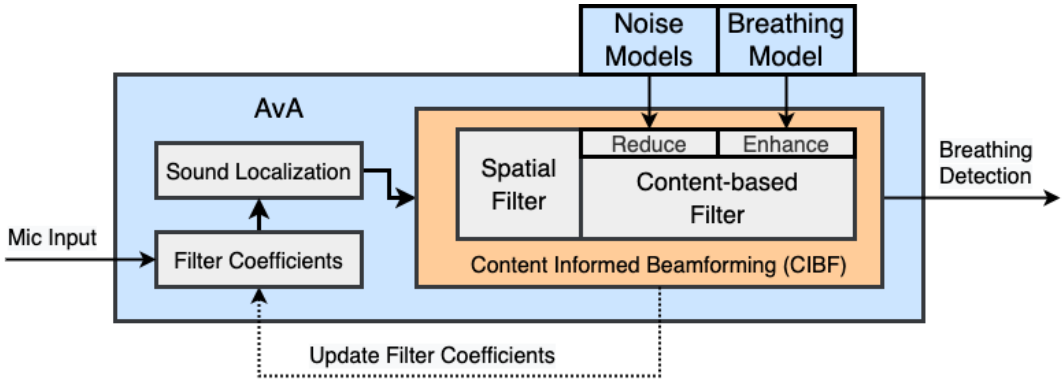


Figure 2: BuMA's breathing detection pipeline.

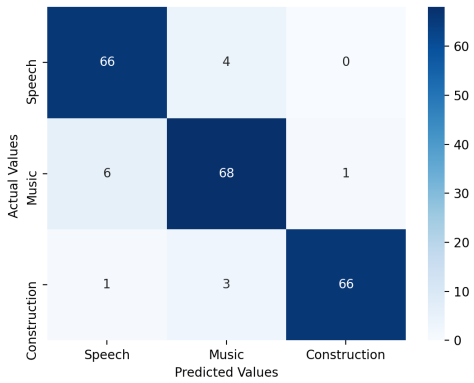


Figure 3: Confusion Matrix of multi-class sound detector

noise/sound class, we use 600 10-second clips extracted from Audioset, which totals 5 hours of audio. The confusion matrix of our multi-class sound detector on the test set is shown in Fig. 3. The total accuracy of our multi-class sound detector is 93%.

We realize that there could be many other noises in the environment, beyond the classes mentioned in this work, and that this multi-class sound detector can be substituted by any multi-class sound detection method, such as [33], to dynamically determine what noises are present in the environment.

Because this multi-class detector is more complex and requires more computation than models and detectors of specific sounds, we move this detector to the cloud, rather than running it at the edge, as shown in the bottom half of Figure 2. We transmit one window of audio every 5 seconds from the edge to the cloud to analyze for the presence of noises. If a new sound is detected, the cloud processing unit alerts the edge processing unit to switch noise models.

We only transmit audio from the edge to the cloud periodically to reduce the amount of data being transmitted. Because much of this audio could be very privacy sensitive, by reducing the amount of data we transmit, we improve the overall privacy of the system by keeping most of the data and computation at the edge.

4 EVALUATION

To evaluate BuMA, we recruited 3 volunteers and compared BuMA against several filtering schemes. Figure 4 shows our set up for collecting samples of breathing from our volunteers. Each volunteer laid down on a mattress in the supine position. We collected breathing samples from each volunteer when BuMA was at the side of the mattress and when BuMA was hanging above the mattress, level with the head of the volunteer. Having BuMA hang above the mattress simulates the case where we may want to embed our platform onto a dream catcher, which are commonly hung above baby cribs for children to play with. We also vary the distance BuMA is away from the volunteer's head from 5 cm to 30 cm. At each position, we played a random clip from Google Audioset [31], that is one of the 3 classes of sounds (construction, music, speech) our multi-class sound classifier detects to dynamically select noise models. At regular intervals of 10 seconds (each clip we played from Audioset is 10 seconds), we randomly play another clip from one of the other classes of noises. We play these clips 150 cm away and at a random angle θ from the person's head, as shown in Figure 4. We varied θ from -45 to 45 degrees.

4.1 Breathing Detection

Figure 5 shows the accuracy of breathing detection versus the distance of the microphone array from the person's head when the microphone array is above the mattress and to the side of the mattress. We plot the accuracy of BuMA compared to existing filtering algorithms, including LCMV beamforming [19] and the AvA filtering platform. We list the average signal-to-noise ratio (SNR) at each configuration. The noise model we select for AvA is the model of the class of the noise that is initially playing. For example, if a clip of construction sound is first playing, then the sound model we use to perform noise filtering is the model of construction noises. Though not explicitly shown, we tune the detectors such that the false positive rate is the same across all comparison methods, meaning all differences in accuracy between the methods is due to differences in the true positive detection rate.

We see that BuMA has the highest accuracy when the microphone array is within 20 cm when the array is at the side of the mattress, and when the microphone array is within 30 cm when the

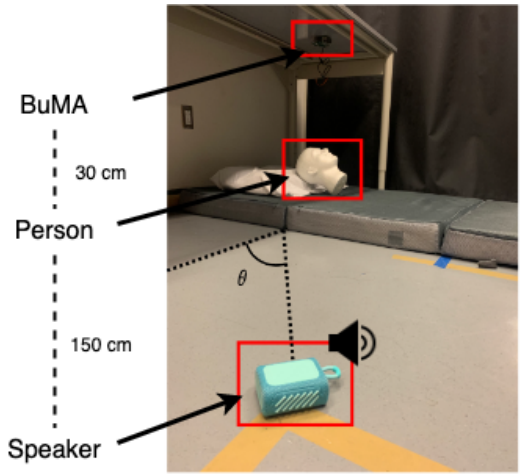
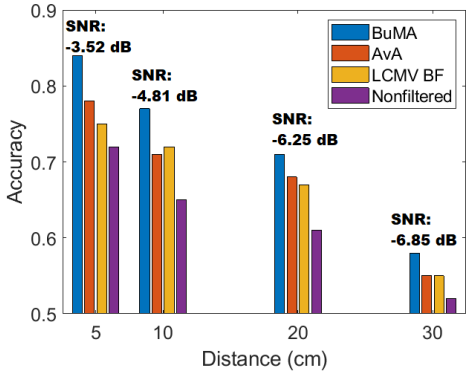
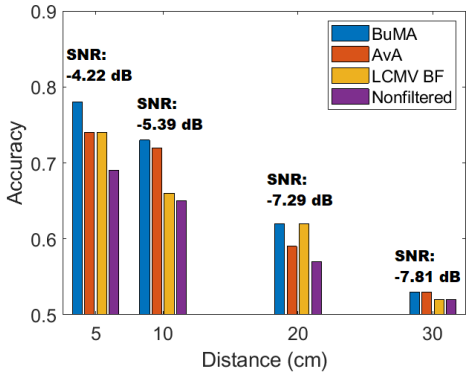


Figure 4: Data collection set up. We placed BuMA either to the side or above the mattress, next to where a person would sleep. Then, we played an interfering noise (construction, speech, or music) from a speaker. In this figure, the distance between speaker and the person's head is 150 cm, and BuMA is 30 cm away right above the person lying down on the mattress in the supine position.



(a) Accuracy vs. distance when BuMA is above the mattress.



(b) Accuracy vs. distance when BuMA is to the side of the mattress.

Figure 5: Accuracy vs. distance BuMA is away from the head of the person. The average SNR is listed above each distance.

array is above the mattress. Past these points, the SNR is too low for our localization and filtering algorithms to significantly detect and enhance breathing. For example, when BuMA is placed to the side of the mattress at 30 cm, filtering does not significantly improve the accuracy of the system. The SNR when the microphone array is above the mattress is higher because the mouth and nose are directly pointing at the array, so the microphones are in the direct path of breathing. BuMA outperforms just normally applying AvA because AvA does not have a mechanism to dynamically switch noise models on the fly. BuMA outperforms the rest of the filtering algorithms because it incorporates both spatial and data-driven filtering, while other methods utilize one or the other.

When BuMA is 10 cm away and above the mattress, BuMA achieves a 73% accuracy, while not applying any filtering results in a 66%, which is a 12% improvement.

4.2 Latency

BuMA samples 500ms windows of audio with 250ms overlap from its microphone array that is attached to the edge processing unit. The full breathing detection pipeline runs on the edge processing unit in 21 ms, which is much less than 250ms.

At the cloud, we receive one window of audio every 5 seconds to determine if new noises are in the environment and that the edge processing unit needs to switch noise models. The cloud processing unit runs its multi-class noise detector to make this determination in an average of 9.44 ms and notifies the edge processing unit. The latency between when the edge processing unit begins to transmit audio windows to the cloud and when the edge processing unit receives a response from the cloud unit (including processing the multi-class noise detector) is on average 18.35ms, which is much lower than 250ms, the amount of time between processing successive windows.

5 DISCUSSION AND FUTURE WORK

Improving breathing detection: Although we improved the detection of breathing by integrating audio filtering and dynamically selecting noises to filter out, we still could not achieve robust detection accuracy (e.g. > 90%). AvA, the filtering platform we incorporated, is a general platform for filtering out and enhancing a wide range of different sounds. We plan to explore more specific methods that take into consideration the salient features of breathing and other common home noises.

Miniaturizing edge processing unit: In this work, we implemented BuMA based on a Raspberry Pi 3b+, which has high relatively high computation capabilities, but relatively high power consumption. In the future, we plan to miniaturize our platform and implement BuMA on a lower power and battery-powered embedded platform or application-specific integrated circuit (ASIC).

Expanding library of noises: We integrated a multi-class noise detector to periodically detect the presence of noises in the environment. Currently, we distinguish between three different classes

of noises, and we plan to expand this number. Additionally, we envision a large library of noise sound models that can allow BuMA to filter and account for numerous potential noises that could occur in the environment.

Breathing monitoring for infants in real environments: Sleep apnea commonly occurs in babies, where they stop breathing for periods of time and potentially passing away (SIDS). We plan to explore how BuMA can be used to monitor the breathing of babies in real environments and alert parents if their child has stopped breathing for long periods of time.

6 CONCLUSION

We present BuMA, a non-contact platform for breathing monitoring. Unlike existing products and works for monitoring breathing, which require devices to be mounted on the person sleeping, or onto the mattress, BuMA can be placed at distance to monitor breathing remotely, without the using cameras that are more privacy intrusive and require sufficient lighting to function properly. BuMA utilizes spatial filtering, beamforming, and data-driven models to filter out overpowering noises commonly found in or near home environments that make breathing detection difficult. We incorporate a mechanism for dynamically selecting noise models to filter out specific noises in the environment and intelligently partition computation between the edge and cloud. We demonstrate that BuMA can improve breathing detection accuracy by up to 12% within 30cm. BuMA is a first step towards robust, contactless, and privacy-sensitive breathing monitoring.

ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation under Grant Numbers CNS-1704899, CNS-1815274, CNS-11943396, and CNS-1837022. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Columbia University, NSF, or the U.S. Government or any of its agencies.

REFERENCES

[1] Apoor S Gami, Eric J Olson, Win K Shen, R Scott Wright, Karla V Ballman, Dave O Hodge, Regina M Herges, Daniel E Howard, and Virend K Somers. Obstructive sleep apnea and the risk of sudden cardiac death: a longitudinal study of 10,701 adults. *Journal of the American College of Cardiology*, 62(7):610–616, 2013.

[2] Reena Mehra, Katie L Stone, Paul D Varosy, Andrew R Hoffman, Gregory M Marcus, Terri Blackwell, Osama A Ibrahim, Rawan Salem, and Susan Redline. Nocturnal arrhythmias across a spectrum of obstructive and central sleep-disordered breathing in older men: outcomes of sleep disorders in older men (mros sleep) study. *Archives of internal medicine*, 169(12):1147–1155, 2009.

[3] Mayo Clinic. Sudden infant death syndrome (sids) - symptoms causes, 2020.

[4] Sense-U. The most complete baby monitoring system, 2022.

[5] Miku. Real-time breathing and sleep tracking – mikucare, 2022.

[6] VIGI Limited. Cloud baby monitor, 2022.

[7] Carlo Massaroni, Joshua Di Tocco, Daniela Lo Presti, Umile Giuseppe Longo, Sandra Miccinilli, Silvia Sterzi, Domenico Formica, Domenico Formica, Domenico Formica, Paola Saccomandi, Paola Saccomandi, and Emiliano Schena. Smart textile based on piezoresistive sensing elements for respiratory monitoring. *IEEE Sensors Journal*, 2019.

[8] Ailton Siqueira, Amanda Franco Spirandeli, Raimes Moraes, and Vicente Zarzoso. Respiratory waveform estimation from multiple accelerometers: An optimal sensor number and placement analysis. *IEEE Journal of Biomedical and Health Informatics*, 2019.

[9] Xiaobai Li, Jie Chen, Jie Chen, Jie Chen, Jie Chen, Jie Chen, Guoying Zhao, and Matti Pietikäinen. Remote heart rate measurement from face videos under realistic situations. *null*, 2014.

[10] Yanwen Wang and Yuanqing Zheng. Tagbreathe : Monitor breathing with commodity rfid systems. *IEEE Transactions on Mobile Computing*, 2020.

[11] Nir Regev, Dov Wulich, and Dov Wulich. Multi-modal, remote breathing monitor. *Sensors*, 2020.

[12] Shih-Hong Li, Bor-Shing Lin, Chen-Han Tsai, Cheng-Ta Yang, and Bor-Shyh Lin. Design of wearable breathing sound monitoring system for real-time wheeze detection. *Sensors*, 17(1), 2017.

[13] Chenyu Huang, Huangxun Chen, Lin Yang, and Qian Zhang. Breathlive: Liveness detection for heart sound authentication with deep breathing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1), mar 2018.

[14] Takahiro Emoto, Udantha R. Abeyratne, Kenichiro Kawano, Takuya Okada, Osamu Jinnouchi, and Ikuji Kawata. Detection of sleep breathing sound based on artificial neural network analysis. *Biomedical Signal Processing and Control*, 41:81–89, 2018.

[15] Stephen Xia and Xiaofan Jiang. Improving acoustic detection and classification in mobile and embedded platforms: Poster abstract. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)*, IPSN '21, page 402–403, New York, NY, USA, 2021. Association for Computing Machinery.

[16] Stephen Xia, Jingping Nie, and Xiaofan Jiang. Csafe: An intelligent audio wearable platform for improving construction worker safety in urban environments. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)*, IPSN '21, page 207–221, New York, NY, USA, 2021. Association for Computing Machinery.

[17] Stephen Xia and Xiaofan Jiang. Pams: Improving privacy in audio-based mobile systems. In *Proceedings of the 2nd International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things*, AIChallengeIoT '20, page 41–47, New York, NY, USA, 2020. Association for Computing Machinery.

[18] Anastasios Alexandridis, Anthony Griffin, and Athanasios Mouchtaris. Capturing and reproducing spatial audio based on a circular microphone array. *JECE*, 2013, January 2013.

[19] O. L. Frost. An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60(8):926–935, 1972.

[20] L. Griffiths and C. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, 30(1):27–34, 1982.

[21] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411 – 430, 2000.

[22] Siow Yong Low, S. Nordholm, and R. Togneri. Convolutive blind signal separation with post-processing. *IEEE Transactions on Speech and Audio Processing*, 12(5):539–548, 2004.

[23] Shoko Araki, Hiroshi Sawada, Ryo Mukai, and Shoji Makino. Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors. *Signal Process.*, 87(8):1833–1847, August 2007.

[24] A.V. Oppenheim, E. Weinstein, K.C. Zangi, M. Feder, and D. Gauger. Single-sensor active noise cancellation. *IEEE Transactions on Speech and Audio Processing*, 2(2):285–290, 1994.

[25] Ying Song, Yu Gong, and S.M. Kuo. A robust hybrid feedback active noise cancellation headset. *IEEE Transactions on Speech and Audio Processing*, 13(4):607–617, 2005.

[26] Jordan Cheer and Stephen Elliott. Multichannel control systems for the attenuation of interior road noise in vehicles. *Mechanical Systems and Signal Processing*, 60-61:753–769, 08 2015.

[27] Emad M. Grais and Hakan Erdogan. Single channel speech music separation using nonnegative matrix factorization and spectral masks. In *2011 17th International Conference on Digital Signal Processing (DSP)*, pages 1–6, 2011.

[28] Yu Ting Yeung, Tan Lee, and Cheung-Chi Leung. Using dynamic conditional random field on single-microphone speech separation. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 146–150, 2013.

[29] Emad M Grais, Gerard Roma, Andrew J. R. Simpson, and Mark Plumbley. Two-stage single-channel audio source separation using deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(9):1469 – 1479, 2017.

[30] Stephen Xia and Xiaofan Jiang. Ava: An adaptive audio filtering architecture for enhancing mobile, embedded, and cyber-physical systems. In *Proceedings of the 21st International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2022) (to appear)*, IPSN '22. IEEE, 2022.

[31] Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. Audio set: An ontology and human-labeled dataset for audio events. In *Proc. IEEE ICASSP 2017*, New Orleans, LA, 2017.

[32] Ltd. Seeed Technology Co. Respeaker 6-mic circular array kit for raspberry pi, 2021.

[33] Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron Weiss, and Kevin Wilson. Cnn architectures for large-scale audio classification. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017.