ARTICLE   OPEN

Contactless cardiac arrest detection using smart devices

Justin Chan [1], Thomas Rea[2,3], Shyamnath Gollakota [1] and Jacob E. Sunshine [4]

Out-of-hospital cardiac arrest is a leading cause of death worldwide. Rapid diagnosis and initiation of cardiopulmonary resuscitation (CPR) is the cornerstone of therapy for victims of cardiac arrest. Yet a significant fraction of cardiac arrest victims have no chance of survival because they experience an unwitnessed event, often in the privacy of their own homes. An under-appreciated diagnostic element of cardiac arrest is the presence of agonal breathing, an audible biomarker and brainstem reflex that arises in the setting of severe hypoxia. Here, we demonstrate that a support vector machine (SVM) can classify agonal breathing instances in real-time within a bedroom environment. Using real-world labeled 9-1-1 audio of cardiac arrests, we train the SVM to accurately classify agonal breathing instances. We obtain an area under the curve (AUC) of 0.9993 ± 0.0003 and an operating point with an overall sensitivity and specificity of 97.24% (95% CI: 96.86–97.61%) and 99.51% (95% CI: 99.35–99.67%). We achieve a false positive rate between 0 and 0.14% over 82 h (117,985 audio segments) of polysomnographic sleep lab data that includes snoring, hypopnea, central, and obstructive sleep apnea events. We also evaluate our classifier in home sleep environments: the false positive rate was 0–0.22% over 164 h (236,666 audio segments) of sleep data collected across 35 different bedroom environments. We prototype our proof-of-concept contactless system using commodity smart devices (Amazon Echo and Apple iPhone) and demonstrate its effectiveness in identifying cardiac arrest-associated agonal breathing instances played over the air.

INTRODUCTION

Out-of-hospital cardiac arrest (OHCA) is a leading cause of death worldwide[1] and in North America accounts for nearly 300,000 deaths annually.[2] A relatively under-appreciated diagnostic element of cardiac arrest is the presence of a distinctive type of disordered breathing: agonal breathing.[3,4] Agonal breathing, which arises from a brainstem reflex in the setting of severe hypoxia,[5,6] appears to be evident in approximately half of cardiac arrest cases reported to 9-1-1. Agonal breathing indicates a relatively short duration from arrest and has been associated with higher survival rates.[7–9] Sometimes reported as "gasping" breaths, agonal respirations may hold potential as an audible diagnostic biomarker, particularly in unwitnessed cardiac arrests that occur in a private residence, the location of 2/3 of all OHCAs.[10,11]

The widespread adoption of smartphones and smart speakers (projected to be in 75% of US households by 2020[12]) presents a unique opportunity to identify this audible biomarker and connect unwitnessed cardiac arrest victims to Emergency Medical Services (EMS) or others who can administer cardiopulmonary resuscitation (CPR). In this study, we hypothesized that existing commodity devices (e.g., smartphones and smart speakers) could be used to accurately identify OHCA-associated agonal breathing instances in a domestic setting. As initial proof-of-concept, we focus on a relatively controlled environment, the bedroom, which is the location of the majority of OHCA events that occur within a private residence.[11,13] A key challenge to algorithm development for this purpose is accessing real-world data on agonal breathing; agonal breathing events are relatively uncommon, lack gold-standard measurements and cannot be reproduced in a lab because of their emergent nature. To overcome this challenge, we leverage a

unique data source, 9-1-1 audio of confirmed cardiac arrest cases, which can include agonal breathing instances captured during the call. As our negative dataset, we use ambient household noise and audio from polysomnographic sleep studies, which include data that share similar audio characteristics to agonal breathing such as snoring and obstructive apnea events. Using real-world audio of agonal instances from OHCA events, we evaluate (1) whether a support vector machine (SVM) can be trained to detect OHCA-associated agonal breathing instances in a bedroom and sleep setting and (2) whether the SVM can be deployed and accurately classify agonal breathing audio in real-time using existing commodity smartphones and smart speakers.

RESULTS

Concept

Our agonal breathing detection pipeline (Fig. 1a) captures audio samples from a smart speaker and smartphone and outputs the probability of agonal breathing in real-time on each 2.5 s audio segment. We use Google's VGGish model[14] as a feature extractor to transform the raw audio waveforms into embeddings which are passed into an SVM. Each segment is transformed from the time-domain into a log-mel spectrogram,[15] and is further compressed into a feature embedding using principal component analysis. These embeddings are then passed into an SVM with a radial basis function kernel that can distinguish between agonal breathing instances (positive data) and non-agonal instances (negative data). An agonal-breathing frequency filter is then applied to the classifier's probability outputs to reduce the false positive rate of the overall system. For comparison, Fig. 1b, c shows example

[1]Paul G. Allen School of Computer Science and Engineering, University of Washington, Washington, WA, USA; [2]Division of General Internal Medicine, University of Washington, Washington, WA, USA; [3]Medic One, Emergency Medical Services, King County, Seattle, WA, USA and [4]Department of Anesthesiology & Pain Medicine, University of Washington, Washington, WA, USA
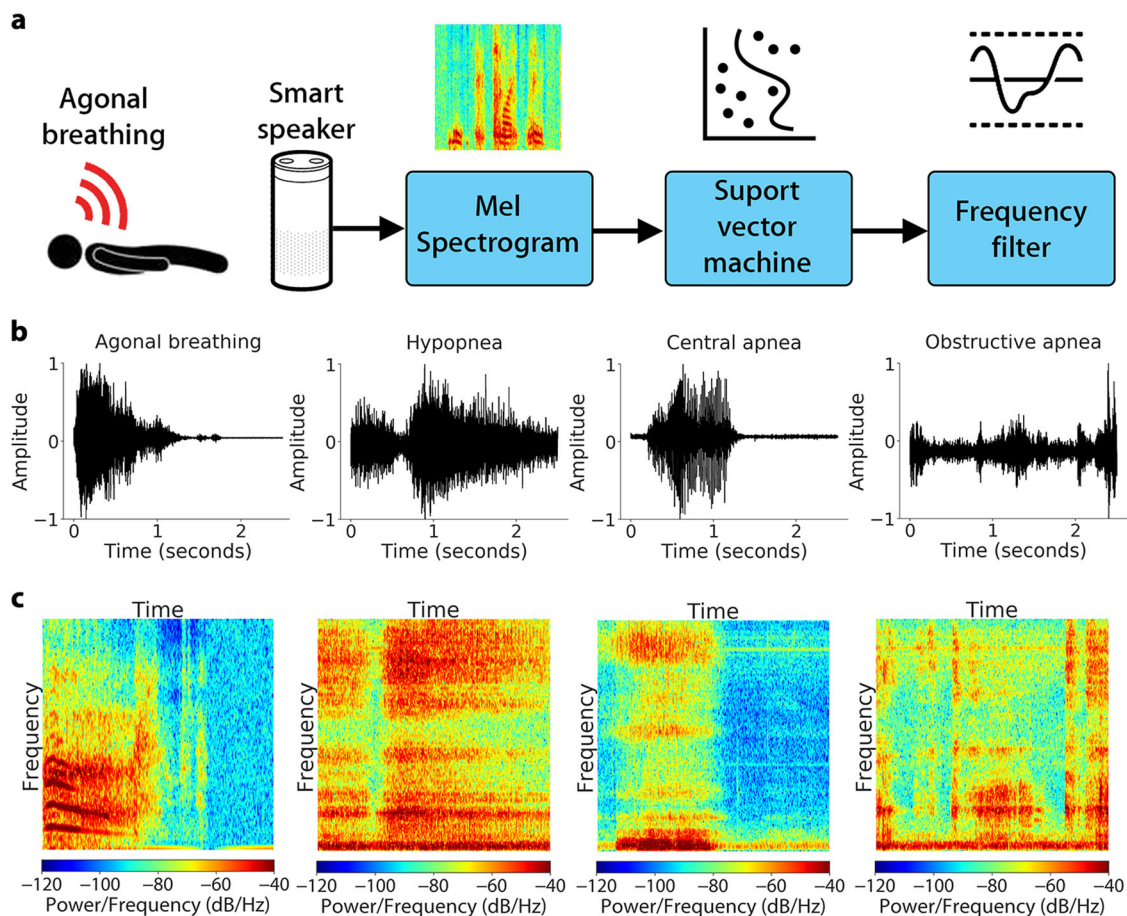Correspondence: Shyamnath Gollakota (gshyam@uw.edu) or Jacob E. Sunshine (jesun@uw.edu)

Fig. 1 Using a smart speaker to detect agonal breathing. **a** Agonal breathing detection pipeline. **b** Audio waveform and **c** spectrogram of agonal breathing, hypopnea, central apnea, and obstructive apnea

audio waveforms and spectrograms for agonal breathing as well as snoring and apnea events, which can sound similar to but are physiologically different from agonal breathing.

Datasets

Our agonal breathing recordings are sourced from 9-1-1 emergency calls from 2009 to 2017, provided by Public Health-Seattle & King County, Division of Emergency Medical Services. The dataset included 162 calls (19 h) that had clear recordings of agonal breathing (Fig. 2a). For each occurrence, we extract 2.5 s worth of audio from the start of each agonal breath. We extracted a total of 236 clips of agonal breathing instances. Given the relatively small size of our agonal breathing dataset, we augment the number of agonal breathing instances with label preserving transformations, a common technique applied to sparse datasets.[16,17] We augment the data by playing the recordings over the air over distances of 1, 3, and 6 m, in the presence of interference from indoor and outdoor sounds with different volumes and when a noise cancellation filter is applied. The recordings were captured on different devices, namely an Amazon Alexa, an iPhone 5s and a Samsung Galaxy S4 to get 7316 positive samples.

Our negative dataset consists of 83 h of audio data captured during polysomnographic sleep studies, across 12 different patients. These audio streams include instances of hypopnea, central apnea, obstructive apnea, snoring, and breathing (Supplementary Table 1). The negative dataset also includes interfering sounds that might be present in a bedroom while a person is asleep, specifically a podcast, sleep soundscape and white noise.

To train our model, we use 1 h of audio data from the sleep study in addition to other interfering sounds. These audio signals were played over the air at different distances and recorded on different devices to get 7305 samples. The remaining 82 h of sleep data (117,985 audio segments) is then used for validating the performance of our model.

Classifier performance

We applied k-fold ($k = 10$) cross-validation and obtained an area under the curve (AUC) of 0.9993 ± 0.0003 (Fig. 2b). We obtain an operating point with an overall sensitivity and specificity of 97.24% (95% CI: 96.86–97.61%) and 99.51% (95% CI: 99.35–99.67%), respectively (Fig. 2c). We ran k-fold ($k = 10$) cross-validation using other machine learning classifiers including k-nearest neighbors, logistic regression and random forests. These classifiers achieved an AUC that was >0.98 but slightly lower than the AUC of the trained SVM (Fig. 2b). Our detection algorithm can run in real-time on a smartphone natively and can classify each 2.5 s audio segment within 21 ms. With the smart speaker, the algorithm can run within 58 ms. We visualized the audio embeddings of our dataset by using t-SNE[18] to project the features into the 2-D space (Fig. 2d). The two clusters represent the abstract features of agonal breathing instances and audio from the polysomnographic studies.

To evaluate false positive rate, we run our classifier trained over the full audio stream collected in the sleep lab (Fig. 2e). The sleep audio used to train each model was excluded from evaluation. By relying only on the classifier's probability outputs,
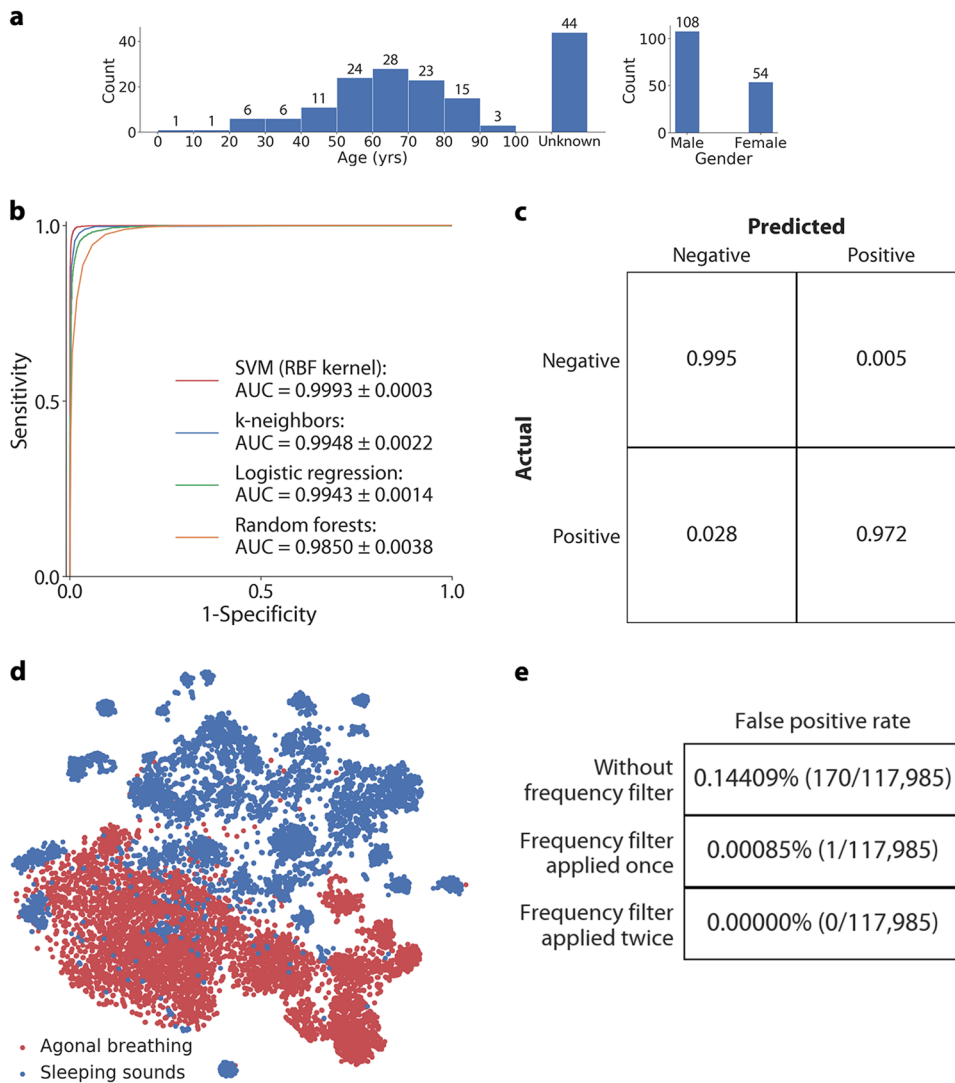
Fig. 2 Performance of agonal breathing classifier. **a** Demographic summary of subjects with agonal breathing during 9-1-1 calls showing distribution of age and gender. **b** ROC curve for our support vector machine classifier, cross-validated on sounds collected from a sleep study, and domestic interfering sounds. **c** Confusion matrix of agonal breathing and sleeping/domestic interfering sounds indicating the operating point on the ROC curve. **d** t-SNE algorithm is applied to visualize the audio embeddings in 2-D. The point clouds show clusters representing the abstract features learned to represent both agonal breathing and negative sound instances. **e** The false positive rate when running the classifier across an 82-h stream of sleep data without and with the frequency filter. By applying a frequency filter to check if the rate of positive predictions matches the rate of agonal breathing, we can decrease the false positive rate

we obtain a false positive rate of 0.14409%, which corresponds to 170 of 117,985 audio segments. To reduce false positives, the classifier's predictions are passed through a frequency filter that checks if the rate of positive predictions is within the typical frequency at which agonal breathing occurs (i.e., within a range of 3–6 agonal breaths per minute[19,20]). This filter reduced the false positive rate to 0.00085%, when it considers two agonal breaths within a duration of 10–20 s. When it considers a third agonal breath within a subsequent period of 10–20 s, the false positive rate reduces to 0%. In our proposed use case a static smart speaker or smartphone would be able to operate on the entire duration of agonal breathing which has been estimated to last for ~4 min,[19] in the early phase of cardiac arrest. Because it is from a sleep lab, we note that the audio stream used in this analysis is captured from a relatively quiet sleep environment, without loud interfering noises.

Performance on real-world sleep data
In order to evaluate our classifier outside of the sleep lab, we measure the false positive rate of our classifier on real-world recordings of sleep sounds that occur within the home (snoring, breathing, movement in bed). We recruited 35 individuals to record themselves while sleeping using their smartphone for a total duration of 167 h (Fig. 3a). The recordings were manually checked to ensure the audio corresponded to sleep sounds. We then retrained our classifier with an additional 5 min of data from each subject, with a comparable operating point with a sensitivity and specificity of 97.17% (95% CI: 96.79–97.55%) and 99.38% (95% CI: 99.20–99.56%), respectively. The false positive rate of the classifier without a frequency filter is 0.21761%, corresponding to 515 of the 236,666 audio segments (164 h) used as test data. After applying the frequency filter, the false positive rate reaches 0.00127% when considering two agonal breaths within a duration
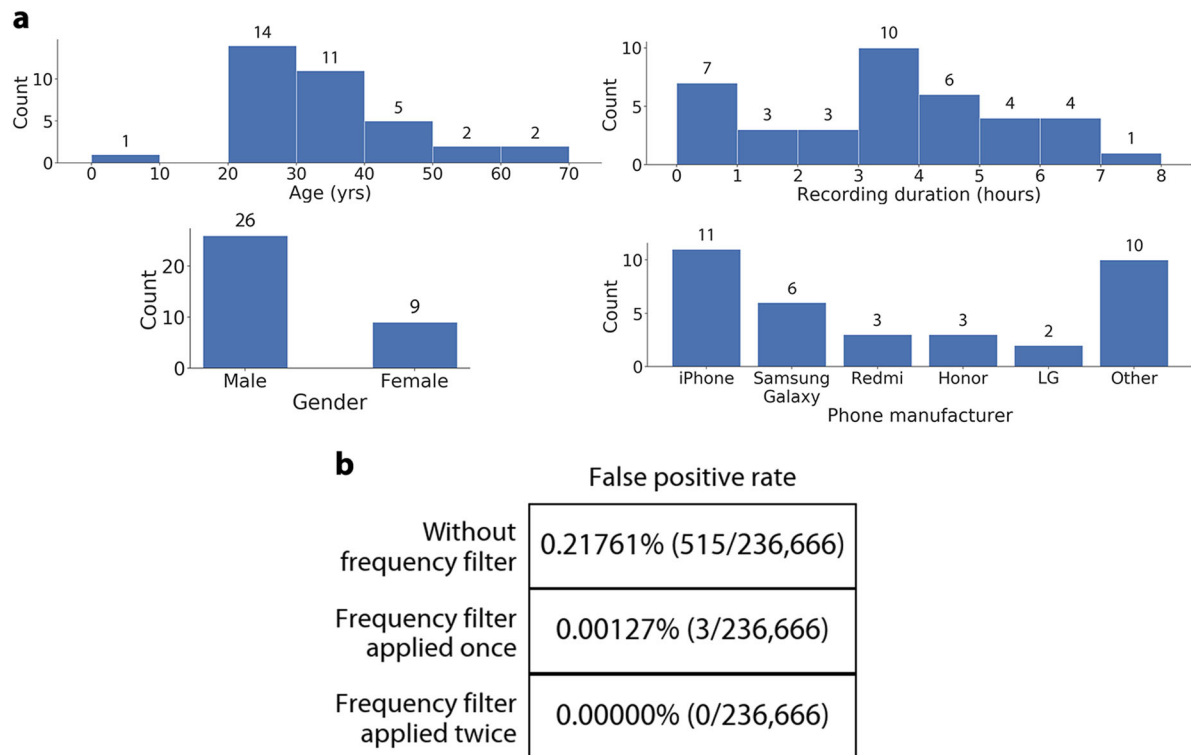
Fig. 3 Performance on real-world sleep data. **a** Demographic summary of subjects showing distribution of age, gender, audio recording duration, and smartphone manufacturer. One participant submitted three unique recordings. **b** The false positive rate when running the classifier trained with the real-world sleep data

of 10–20 s, and 0% after considering a third agonal breath within a subsequent period of 10–20 s (Fig. 3b).

Benchmark performance

Finally, we benchmark the performance of our classifier. For these experiments we played the audio clips of agonal breathing over the air from an external speaker and captured the audio on an Amazon Echo and Apple iPhone 5s. In Fig. 4, we show the detection accuracy of our classifier in a domestic setting on a smart speaker and smartphone. We evaluate detection accuracy using the $k = 10$ validation folds in our dataset such that no audio file in the validation set appears in any of the different recording conditions in the training set. Figure 4a shows the detection accuracy of our classifier in ambient conditions at distances of 1, 3, and 6 m on the Echo and iPhone 5s. Both devices achieve >96.63% mean accuracy at distances up to 3 m. We also evaluated the effect of placing the smartphone in a pocket, with the subject supine on the ground and the speaker next to the head, and obtain a mean detection accuracy of 93.22% ± 4.92%. Figure 4b shows our system performance, using the same experimental setup, but in the presence of indoor interfering sounds (cat, dog, air conditioner) and outdoor interfering sounds (traffic, construction and human speech). Across all interfering sound classes the smart speaker and smartphone achieve a mean detection accuracy of 96.23%. Finally, we evaluate how a smartphone or smart speaker can use acoustic interference cancellation to reduce the interfering effects of its own audio transmissions and improve detection accuracy of agonal breathing (Fig. 4c, d). We set the smartphone to play sounds one might play to fall asleep including a podcast, sleep soundscape (i.e., river current) and white noise. We play them at a soft (45 dBA) and loud (67 dBA) volume. Simultaneously, we play the agonal breathing audio clips. Without any audio cancellation, the detection accuracy is consistently poor,

with an average accuracy of 22.46 and 4.76% across distances and sounds for soft and loud interfering volumes. When the audio cancellation algorithm is applied, our detection accuracy achieves an average accuracy of 98.62 and 98.57% across distances and sounds for soft and loud interfering volumes, respectively.

To benchmark the classifier's performance against negative audio sounds, we played a stream of negative sounds over the air: snoring, a podcast, a sleep soundscape and white noise, and recorded them on a smart speaker and smartphone. We repeat the benchmark experiments above and record these sounds at different distances and in the presence of indoor and outdoor interfering sounds (Fig. 5). The smart speaker and smartphone achieve a mean detection accuracy of 99.57% at a distance of 3 m; a 100% accuracy corresponds to the classifier correctly identifying that the sounds are from the negative dataset. Across all interfering sounds, the mean detection accuracy was 99.29%.

DISCUSSION
Out-of-hospital cardiac arrest is a widespread public health concern. Early CPR is a core treatment, underscoring the vital importance of timely detection, followed by initiation of a series of time-dependent coordinated actions which comprise the chain of survival.[21] Hundreds of thousands of people worldwide die annually from unwitnessed cardiac arrest, without any chance of survival because they are unable to activate this chain of survival and receive timely resuscitation. Non-contact, passive detection of agonal breathing represents a novel way to identify a portion of previously unreachable victims of cardiac arrest, particularly those who experience such events in a private residence. As the US population ages and more people become at risk for OHCA, leveraging commodity smart hardware for monitoring of these emergent conditions could have public health benefits. Other domains where an efficient agonal breathing classifier could have
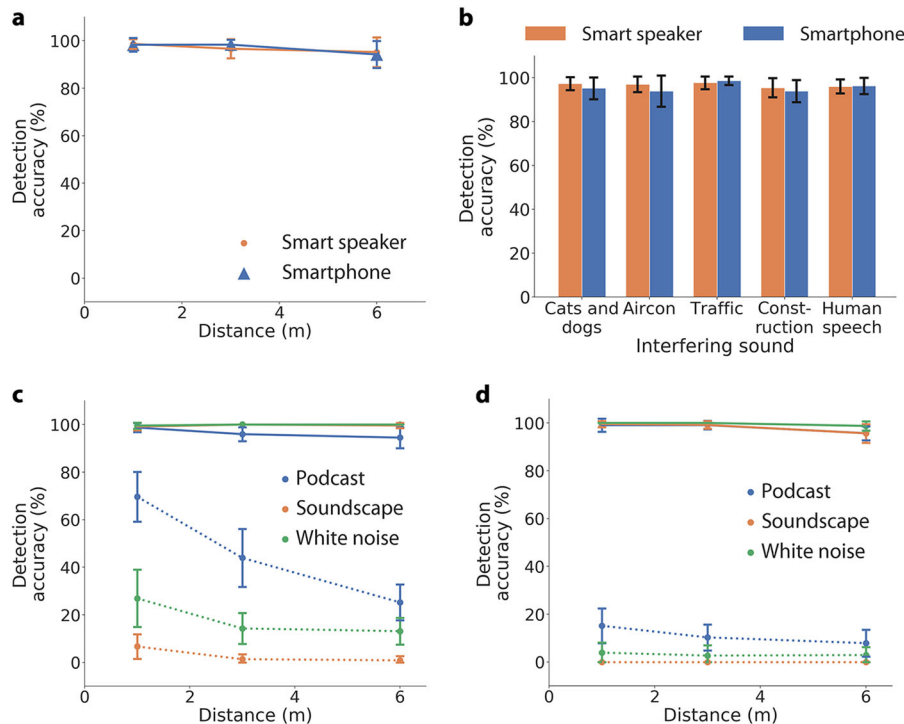
Fig. 4 Benchmark testing of agonal breathing sounds across different scenarios. Mean detection accuracy of smart speaker and smartphone (**a**) across distance (**b**) in the presence of other interfering indoor and outdoor sounds. **c, d** With acoustic interference cancellation a smartphone or smart speaker can reduce the effects of its own audio transmissions and become more sensitive at detecting agonal breathing signals. The left and right subplots show the detection accuracy when interfering sounds are played at soft (45 dBA) and loud (67 dBA) volumes, respectively. Solid and dashed lines indicate detection accuracy with and without interference cancellation, respectively. Error bars indicate the standard deviation accuracy across validation folds
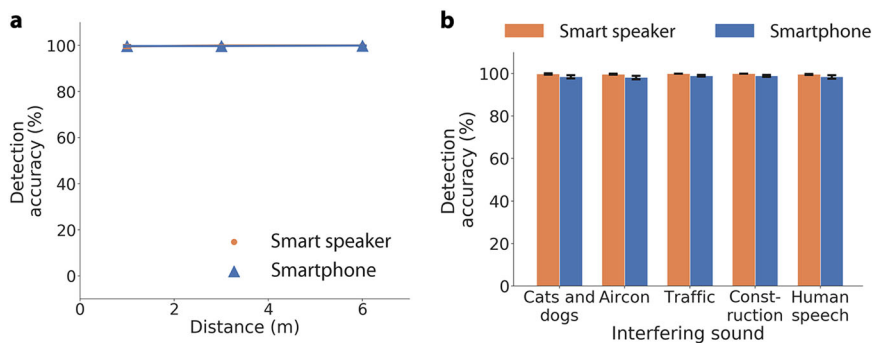

Fig. 5 Benchmark testing of negative sounds across different scenarios. Mean detection accuracy of smart speaker and smartphone (**a**) across distance (**b**) in the presence of other interfering indoor and outdoor sounds. Error bars indicate the standard deviation accuracy across validation folds

utility include unmonitored health facilities (hospital wards and elder care environments[22]), EMS dispatch,[23] and when people have greater than average risk, such as people at risk for opioid overdose-induced cardiac arrest.[24]

An immediate concern of a passive agonal breathing detector is privacy. For this use case, intentional activation of the device (i.e., "Hey Alexa" or "Hey Siri") immediately prior to classification is not feasible because diagnosis involves an unconscious individual in an emergent situation. To address privacy concerns, we envision our system to run locally on the smart devices and not store any data.

An advantage of a contactless detection mechanism is that it does not require a victim to be wearing a device while asleep in the bedroom, which can be inconvenient or uncomfortable. Such a solution can be implemented on existing wired smart speakers,

and as a result would not face power constraints and could scale efficiently. Potential downsides include that, to date, agonal breathing has been identified in ~50% of cardiac arrest victims, so people experiencing an unwitnessed cardiac arrest without agonal breathing would go undetected by our system. With that said, it is worth noting that prior incidence estimates of cardiac arrest-associated agonal breathing events have been based on 9-1-1 calls, which likely biases estimates and underestimates the true incidence of agonal breathing during cardiac arrest.[25]

Our proof-of-concept study has the following limitations. The number of agonal instances in this study was from one geographic community over an 8-year period and is relatively sparse containing 10 min of clearly captured agonal breathing sounds. Additional agonal events are needed to ensure our model

generalizes to variations of agonal breathing. Moreover, additional audio of agonal instances, which likely reside in 9-1-1 databases around the world, would also contribute to a more accurate detection system. Evaluation on an alternative positive dataset, such as recordings from patients in hospice care or inpatient end-of-life settings, could help conclusively validate the real-world performance of the classifier. In addition, evaluation over a longer duration is needed to gauge the real-world clinical value of the classifier. Another classification consideration is whether conditions such as seizure, hypoglycemia, severe stroke or drug overdoses with disordered breathing (but not agonal breathing) are distinct or similar to agonals. Further work is needed in this area, yet it is worth noting that all of these instances represent acute conditions requiring prompt medical intervention. In addition, our current system is focused on detection within the controlled environment of the bedroom. Building and evaluating a general detection system that works reliably in different environments is an area of future work. Finally, this proof-of-concept study did not involve EMS activation. A real-world implementation would sound an alarm and require a user-interface that provides a cancellation opportunity before the emergency medical response system was activated, so as to further minimize false positives.

Technology is rapidly evolving and in turn providing opportunities to improve human health.[26,27] The increasing adoption of commodity smart speakers in private residences[12] and hospital environments[28] may provide a wide-reaching means to realize the potential of a contactless cardiac arrest detection system.

METHODS
This study was approved by the University of Washington Institutional Review Board. The methods were performed in accordance with University of Washington's ethical, professional and legal standards. The 9-1-1 dataset was provided by Public Health-Seattle & King County, Division of Emergency Medical Services. For the sleep apnea dataset, human participants from the polysomnography studies provided written informed consent.

Datasets
The data represents a subset of 9-1-1 calls which (a) contained a known cardiac arrest and (b) had been identified to contain cardiac arrest-associated agonal breathing instances. The negative data consist of recordings of 12 patients sleeping in a sleep lab recorded on a Samsung Galaxy S4.

Our agonal breathing recordings are sourced from 9-1-1 emergency calls from 2009 to 2017 provided by Public Health-Seattle & King County, Division of Emergency Medical Services. There are 729 calls totaling to 82 h (Fig. 2a). The provided recordings include only calls involving cardiac arrest and specifically those determined to contain occurrences of agonal breathing, either by audible identification of agonal breathing or by description of the breathing from the caller. Each call is further rated by the 9-1-1 operator and an EMS quality assurance reviewer with a confidence score indicating the presence of audible agonal instances. We train our classifier on audio from calls that are rated with high confidence by both the operator and reviewer to contain audible agonal instances. These instances predominantly occur when the 9-1-1 operator asked the caller to place the phone next to the victim's mouth for the purposes of breathing identification. A clinician who has experience identifying agonal breathing listened to a subset of recordings with the researcher and pointed out instances of agonal breathing. The researcher then identified all instances of agonal breathing that did not co-occur with other interfering sounds such as human speech. The trained researcher did this by listening to the 162 calls (19 h) and manually recorded timestamps where agonal breathing was heard during the call. For every timestamp annotation, we extract 2.5 s worth of audio from the start of each agonal breath. We extracted a total of 236 clips of agonal breathing instances. The female to male ratio was 0.5 and the median age was 62 (IQR: 21).

Two independent researchers confirmed the presence of agonal breathing sounds. They were first trained with examples of agonal breathing sounds. They then listened to the 236 clips and were instructed to mark clips that did not contain agonal breathing. The first researcher

marked 1 of the 236 clips as not agonal breathing (classifying it as a cough sound), but marked all remaining 235 clips as containing agonal breathing. The second researcher marked all 236 clips as containing agonal breathing.

Our negative dataset consists of 83 h of audio from polysomnographic studies across 12 different patients (Supplementary Table 1). The female to male ratio was 1 and the median age was 57.5 (IQR: 10.25). The mean number of hypopnea, central apnea, and obstructive apnea events across patients was 41, 24, and 26, respectively. The mean apnea–hypopnea index (AHI) was 13, where a value of 0–5 is considered as 'no apnea', 5–15 is considered as 'mild apnea', 15–30 is considered as 'moderate apnea', and values > 30 are considered as 'severe apnea'.[29] AHI annotations were identified and calculated by trained sleep technicians. The negative dataset also includes interfering sounds that might be played while a person is asleep: podcast, sleep soundscape, and white noise.

We augment the data by playing the recordings over the air at distances of 1, 3, and 6 m, in the presence of interference from indoor and outdoor sounds with different volumes and when a noise cancellation filter is applied. The recordings were captured on different devices, namely an Amazon Alexa, an iPhone 5s and a Samsung Galaxy S4. Similarly, for the negative dataset, portions of the sleep data from all patients were played over the air and recorded on different devices as well as over a phone connection. We play a 5 min portion of audio data from each patient over the air at different distances and record the data on an Amazon Alexa, iPhone 5s and over a phone connection. The entire dataset for cross-validation consists of 14,621 data points with 7316 agonal breathing instances and 7305 instances of negative data.

The classifier's false positive rate is evaluated on a set of real-world sleep sounds that occur in bedroom settings. We recruited 35 subjects to record themselves sleeping in their own bedrooms with a smartphone. Subjects were recruited from the Amazon Mechanical Turk platform. Subjects were asked to record themselves sleeping with their smartphone. All recordings submitted by subjects were manually reviewed to assure the presence of sleep sounds. The female to male ratio was 0.35, the median age was 33.00 (IQR: 13.00), the median recorded time was 4.48 (IQR: 3.12) h, and 28 unique smartphones were used across all subjects (Fig. 3a).

Data preparation
We note that the audio clips were sampled at a frequency of 8 kHz which is standard for audio received over a telephone. All audio clips are normalized between a range of −1 and 1. The audio clips are then passed through Google's VGGish[14] model for extracting feature embeddings from an audio waveform. The VGGish model transforms the waveforms into a compact embedding. The model resamples all audio waveforms at 16 kHz then computes a spectrogram using the Short-Time Fourier Transform. A log-mel spectrogram is generated and PCA is applied on the spectrogram to produce a 256-dimensional embedding.

Training algorithm
We performed k-fold validation ($k = 10$). For any given fold, none of the breathing instances in the validation set occurred in the training set. We evaluate detection accuracy such that no audio file in the validation set appears in any of the different recording conditions in the training set (e.g., if a file played at 6 m is present in the validation set, the same file played at 1 m is not present in the training set). We use a support vector machine with a radial basis function kernel and a regularization (C parameter) of 10. To reduce bias in our classifier we partitioned the data such that recordings from the same call did not straddle the training and validation set split. During cross-validation there was never an instance where a subject in the training set occurred in the validation set or vice versa.

Benchmark experiments
To record audio indefinitely on the Echo we used Echo's Drop In feature which streams audio to another smartphone. That smartphone was plugged into a laptop which recorded audio data that was received on the smartphone's audio interface. Audio from the Echo is streamed at 16 kHz and recorded at 44.1 kHz. The iPhone recorded data at 44.1 kHz. Each of the 236 audio clips is prepended with a frequency modulated continuous wave (FMCW) chirp. An FMCW chirp has good auto-correlation properties, as a result we can cross-correlate the recordings from the Echo and iPhone with the chirp to determine the exact timestamp of each audio clip. Each audio clip can then be extracted and transformed into an input for the classifier.

In our benchmark scenarios we evaluate the detection accuracy of our classifier across different distances on a second generation Amazon Echo and an iPhone 5s. We played the 236 audio clips of agonal breathing from a AmazonBasics Wireless Bluetooth speaker and recorded the audio on the Echo and iPhone. The sound intensity of the recordings were ~70 dBA at a distance of 1 m. We fixed the location of the Echo and iPhone and placed the speaker at different distances.

To evaluate the audio interference cancellation algorithm we set the iPhone 5s to play music at two different volumes (45 and 67 dBA), while simultaneously recording audio. We then ran an acoustic interference cancellation algorithm that allowed the smartphone to locally reduce the interference of its own audio transmissions. We used an adaptive least mean squares filter to reduce the dissimilarity between the device's transmission and the received audio recording. Our filter uses the Sign-Data LMS algorithm with 100 weights and a step size of 0.05.

When evaluating system performance in the presence of interfering sounds we use two external speakers, one which plays the agonal breathing recordings and another that plays the interfering noise. The interfering noise is played with a sound intensity of ~55 dBA at a distance of 1 m. The interfering sounds are played outside the room containing the agonal breathing speaker and the recording device to simulate sounds that would be heard from outside a bedroom.

Run-time analysis
The most time consuming operations within the detection pipeline are the fast Fourier transforms (FFTs) required to generate the spectrogram and running inferences on the audio embeddings. Our iPhone 7 implementation of the detection algorithm used the Accelerate frameworks to perform the FFTs and Monte Carlo sampling to approximate the radial basis function kernel. On an iPhone 7 performing the FFTs to generate a single log-mel spectrogram takes 16 ms and running inferences on the support vector machine takes 5 ms. While the classifier can in principle run locally on the Echo device, Amazon currently does not allow third party programs to locally analyze data. Thus, to estimate the performance of our system when run natively on an Amazon Echo, we ran our pipeline on an iPhone 4, which shares the same Cortex-A8 processor as the Echo. On an iPhone 4, computing the spectrogram takes 40 ms and making predictions takes 18 ms.

Reporting summary
Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

DATA AVAILABILITY
All data necessary for interpreting the manuscript have been included. The datasets used in the current study are not publicly available due to the sensitivity of 9-1-1 calls but are available from the corresponding authors on reasonable request and with the permission of Public Health-Seattle & King County, Division of Emergency Medical Services.

CODE AVAILABILITY
The code used in this study is available at https://cardiacalert.cs.washington.edu/.

AUTHOR CONTRIBUTIONS
S.G., J.E.S., and J.C. designed the experiments; J.C. conducted the experiments; J.C. and S.G. designed the algorithms and models; J.C. conducted the analysis with technical supervision by S.G.; J.C., S.G., and J.E.S. interpreted the results; J.E.S., S.G., and J.C. wrote the manuscript; T.R. provided substantial clinical feedback on the manuscript. J.E.S. conceptualized the study using the 9-1-1 dataset, S.G. conceptualized the sleep dataset analysis and J.C. conceptualized the real-world sleep data study.

REFERENCES
1. Myat, A., Song, K.-J. & Rea, T. Out-of-hospital cardiac arrest: current concepts. Lancet **391**, 970 (2018).
2. McNally, B. et al. Out-of-hospital cardiac arrest surveillance—cardiac arrest registry to enhance survival (cares), United States, October 1, 2005–December 31, 2010. Morb. Mortal. Wkly. Rep.: Surveill. Summ. **60**, 1 (2011).
3. Schultz, A. M., McCoy, M. A. & Graham, R. Strategies to Improve Cardiac Arrest Survival: a time to act. (National Academies Press, Washington, DC, 2015).
4. Rea, T. D. Agonal respirations during cardiac arrest. Curr. Opin. Crit. Care **11**, 188 (2005).
5. Poets, C. F., Meny, R. G., Chobanian, M. R. & Bonofiglo, R. E. Gasping and other cardiorespiratory patterns during sudden infant deaths. Pediatr. Res. **45**, 350 (1999).
6. Lumsden, T. Observations on the respiratory centres in the cat. J. Physiol. **57**, 153 (1923).
7. Hauff, S. R. et al. Factors impeding dispatcher-assisted telephone cardiopulmonary resuscitation. Ann. Emerg. Med. **42**, 731 (2003).
8. Clark, J. J., Larsen, M. P., Culley, L. L., Graves, J. R. & Eisenberg, M. S. Incidence of agonal respirations in sudden cardiac arrest. Ann. Emerg. Med. **21**, 1464 (1992).
9. Bang, A., Herlitz, J. & Martinell, S. Interaction between emergency medical dispatcher and caller in suspected out-of-hospital cardiac arrest calls with focus on agonal breathing. A review of 100 tape recordings of true cardiac arrest cases. Resuscitation **56**, 25 (2003).
10. van Diepen, S. et al. Multistate 5-year initiative to improve care for out-of-hospital cardiac arrest: Primary results from the HeartRescue Project. J. Am. Heart Assoc. **6**, e005716 (2017).
11. Eisenberg, M. S. et al. Identification of cardiac arrest by emergency dispatchers. Am. J. Emerg. Med. **4**, 299 (1986).
12. Gartner says worldwide spending on VPA-enabled wireless speakers will top $2 billion by 2020. https://www.gartner.com/en/newsroom/press-releases/2016-10-03-gartner-says-worldwide-spending-on-vpa-enabled-wireless-speakers-will-top-2-billion-by-2020 (2016).
13. Kiyohara, K. et al. Out-of-hospital cardiac arrest at home in Japan. Am. J. Cardiol. **123**, 1060–1068 (2019).
14. Models for audioset: a large scale dataset of audio events. https://github.com/tensorflow/models/tree/master/research/audioset (2019).
15. Muda, L., Begam, M. & Elamvazuthi, I. Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (DTW) techniques. Preprint at https://arxiv.org/abs/1003.4083 (2010).
16. Cui, X., Goel, V. & Kingsbury, B. Data augmentation for deep neural network acoustic modeling. IEEE/ACM Trans. Audio, Speech Lang. Process. (TASLP) **23**, 1469 (2015).
17. Wong, S. C., Gatt, A., Stamatescu, V. & McDonnell, M. D. Understanding data augmentation for classification: when to warp? Preprint at https://arxiv.org/abs/1609.08764 (2016).
18. Maaten, Lvd & Hinton, G. Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579 (2008).
19. Roppolo, L. P. et al. Dispatcher assessments for agonal breathing improve detection of cardiac arrest. Resuscitation **80**, 769 (2009).
20. Lewis, M., Stubbs, B. A. & Eisenberg, M. S. Dispatcher-assisted CPR: time to identify cardiac arrest and deliver chest compression instructions. Circulation **128**, 1522 (2013).
21. Neumar, R. W. et al. Part 1: executive summary: 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation **132**, S315 (2015).
22. Pape, M. et al. Survival after out-of-hospital cardiac arrest in nursing homes—a nationwide study. Resuscitation **125**, 90 (2018).
23. Blomberg, S. N. et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation **138**, 322–329 (2019).

24. Nandakumar, R., Gollakota, S. & Sunshine, J. E. Opioid overdose detection using smartphones. *Sci. Transl. Med.* **11**, eaau8914 (2019).

25. Bobrow, B. J. et al. Gasping during cardiac arrest in humans is frequent and associated with improved survival. *Circulation* **118**, 2550 (2008).

26. Gambhir, S. S., Ge, T. J., Vermesh, O. & Spitler, R. Toward achieving precision health. *Sci. Transl. Med.* **10**, eaao3612 (2018).

27. Steinhubl, S. R., Muse, E. D. & Topol, E. J. The emerging field of mobile health. *Sci. Transl. Med.* **7**, 283rv3 (2015).

28. Siwicki, B. Special report: AI voice assistants making an impact in healthcare. http://provisionnetworks.net/special-report-ai-voice-assistants-making-an-impact-in-healthcare/ (2018).

29. Nieto, F. J. et al. Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. *Jama* **283**, 1829 (2000).