BreathListener: Fine-grained Breathing Monitoring in Driving Environments Utilizing Acoustic Signals

Xiangyu Xu
Shanghai Jiao Tong University
chillex@sjtu.edu.cn

Jiadi Yu*
Shanghai Jiao Tong University
jiadiyu@sjtu.edu.cn

Yingying chen
Rutgers University
yingche@scarletmail.rutgers.edu

Yanmin Zhu
Shanghai Jiao Tong University
yzhu@cs.sjtu.edu.cn

Linghe Kong
Shanghai Jiao Tong University
linghe.kong@sjtu.edu.cn

Minglu Li
Shanghai Jiao Tong University
mlli@sjtu.edu.cn

ABSTRACT

Given the increasing amount of time people spent on driving, the physical and mental health of drivers is essential to road safety. Breathing patterns are critical indicators of the wellbeing of drivers on the road. Existing studies on breathing monitoring require active user participation of wearing special sensors or relatively quiet environments during sleep, which are hardly applicable to noisy driving environments. In this work, we propose a fine-grained breathing monitoring system, *BreathListener*, which leverages audio devices on smartphones to estimate the fine-grained breathing waveform in driving environments. By investigating the data collected from real driving environments, we find that Energy Spectrum Density (ESD) of acoustic signals can be utilized to capture breathing procedures in driving environments. To extract breathing pattern in ESD signals, *BreathListener* eliminates interference from driving environments in ESD signals utilizing background subtraction and Ensemble Empirical Mode Decomposition (EEMD). After that, the extracted breathing pattern is transformed into Hilbert spectrum, and we further design a deep learning architecture based on Generative Adversarial Network (GAN) to generate fine-grained breathing waveform from the Hilbert spectrum of extracted breathing patterns in ESD signals. Experiments with 10 drivers in real driving environments show that *BreathListener* can accurately capture breathing patterns of drivers in driving environments.

CCS CONCEPTS

• **Networks → Cyber-physical networks**.

KEYWORDS

Breathing Monitoring, Driving Safety, Acoustic Sensing

ACM Reference Format:
Xiangyu Xu, Jiadi Yu, Yingying chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. BreathListener: Fine-grained Breathing Monitoring in Driving Environments Utilizing Acoustic Signals. In *The 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '19), June 17–21, 2019, Seoul, Republic of Korea.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3307334.3326074

*Jiadi Yu is the corresponding author

MobiSys '19, June 17–21, 2019, Seoul, Republic of Korea

ACM ISBN 978-1-4503-6661-8/19/06…$15.00
https://doi.org/10.1145/3307334.3326074

1 INTRODUCTION

People spent more and more time traveling on roads with car trips. According to statistics, 76% of people in American spend an average of more than 25 minutes every day driving for commuting[24]. The increasing time spent in driving brings more concern on drivers' physical and mental health, which is essential for building a safe driving environment. Vital signs such as breathing is a good indicator of drivers' health status[5]. And passive breathing monitoring is desirable because the monitoring system should facilitate safe driving instead of a distraction to the driver inside of the vehicle. Once the breathing patterns are obtained, potential health issues of drivers, including drowsy, fatigue, etc., can be detected and alarmed in advance. Therefore, a fully automatic and passive monitoring system to capture drivers' breathing patterns, which should not need drivers' active involvement, is highly desirable.

There have been some existing studies on breathing monitoring leveraging RF-signals[30][32], WiFi signals[1][19][2], acoustic signals[26][27] and sensors on smartphones[15][3]. However, most of these approaches can only provide with coarse-grained breathing patterns such as breathing rate. Moreover, these approaches either require the subject to be relatively stationary(e.g., sleeping)[26][19], or need a relatively quiet environment[32][27] (e.g., home or office). In driving environments, to accurately monitor drivers' body status(i.e, providing clinic-level breathing analysis), we need to estimate fine-grained breathing waveforms from dynamic drivers in noisy vehicular environments. There also have been works monitoring vital signs in driving environments, including breathing [11], EEG[29], and cardiac[23]. These studies are based on pre-deployed infrastructures, which are intrusive to drivers.

Towards this end, a fine-grained breathing monitoring system in driving environments, which does not depend on any pre-deployed infrastructure and additional hardware, is essential to provide ubiquitous healthcare services for drivers. Since acoustic signals have advantages of easily accessible and slow propagating velocity, they have been utilized in recent researches to track the fine-grained movements of objects[21][28][39]. Considering that breathing brings regular movements of chest and abdomen, we study whether it is feasible to utilize acoustic signals with drivers' smartphones for fine-grained breathing monitoring in driving environments. To

realize the acoustic-based breathing monitoring, we face several challenges in practice. First, the approach needs to eliminate the interferences of acoustic signals' multi-path propagation in driving environments. Second, the approach should be robust to various driving environments, including different road types, traffic conditions and smartphone placements of drivers. Third, the approach oughts to generate the fine-grained breathing waveform from the limited acoustic devices on smartphones.

In this paper, we first analyze the feasibility of utilizing acoustic signals to capture breathing patterns during driving, and find that Energy Spectrum Density (ESD) of acoustic signals can be used to sense breathing procedures in driving environments. Based on the observation, we propose a fine-grained breathing monitoring system, *BreathListener*, which leverages acoustic devices on smartphones to estimate the fine-grained breathing waveform in driving environments. *BreathListener* first eliminates the non-movement interferences in ESD signals utilizing background subtraction. Then, *BreathListener* separates breathing from other movements in driving environments utilizing Ensemble Empirical Mode Decomposition (EEMD), and extracts breathing patterns in ESD signals. Since Hilbert spectrum can show the instant frequency of the waveforms, we further transform the extracted breathing pattern is into Hilbert spectrum. Finally, we design a deep learning architecture based on Generative Adversarial Network (GAN), which can generate fine-grained breathing waveform from the Hilbert spectrum of extracted breathing patterns in ESD signals. Our experiments validate the accuracy and robustness of *BreathListener* in various real driving environments.

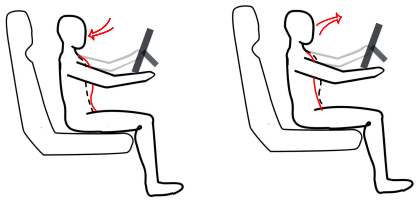
We highlight our main contributions as follows:

- We design a breathing monitoring system, *BreathListener*, which leverages acoustic devices on smartphones to obtain the fine-grained breathing waveform for drivers in real driving environments.
- We find that ESD of acoustic signals can be utilized to capture breathing waveform in driving environments, and further eliminate the interferences of environment noises and extract the breathing pattern in ESD signals leveraging background subtraction and EEMD method.
- We propose a deep learning architecture for *BreathListener* based on GAN, which can be utilized to generate fine-grained breathing waveform from the Hilbert spectrum of extracted breathing patterns in ESD signals.
- We conduct experiments in real driving environments and the results show that *BreathListener* achieves an average error of $0.11bpm$ for breathing rate estimation, and an average correlation coefficient of 0.95 compared to ground truth for breathing waveform estimation.

The rest of the paper is organized as follows. Section 2 introduces the background and preliminaries. Section 3 presents the system design of *BreathListener*. The evaluation results is shown in Section 4. Section 5 gives the discussion. Related works are reviewed in Section 6. Section 7 concludes the paper.

2 BACKGROUND AND PRELIMINARIES

As one of the most significant vital signs of life, breathing patterns play a fundamental role in reflecting the status of the human body.



(a) Inhale (b) Exhale

Figure 1: Illustration of breathing procedure in driving environments.

Therefore, it is of great significance to monitor the breathing of drivers for providing health services in driving environments.

2.1 Breathing Mechanism

Breathing is the process of human body to facilitate gas exchange with the environment[9]. Typically, the breathing procedure consists of repetitive cycles of inhalation and exhalation. During inhalation, the diaphragm contracts and intercostal muscles pull the ribs upwards, which causes the chest to expand. Meanwhile, with the air breathed in, the abdominal muscles stretch, which leads the abdomen to bulge, as shown in Fig.1(a). As for exhalation, all the muscles of inhalation relax, and the chest and abdomen return to resting position by elasticity, as shown in Fig.1(b). Breathing procedure is automatically controlled by the human body, and thus is highly related to activity levels[5].

In driving environments, the body status of drivers can be indicated by their breathing patterns, and the most widely used index is breathing rate. However, breathing rate is a rather coarse-grained indicator. For example, when the breathing rate of a driver becomes a bit slower than normal, it is hard to tell whether the driver is in a relax mode or becomes tired and sleepy. To accurately indicate the body status of drivers, fine-grained breathing information is required. In this case, when a driver becomes tired or sleepy, usually the tidal volume (the volume of air displaced between an inhalation and exhalation) becomes smaller than normal[25], while the tidal volume remains normal when the driver is in a relax mode. Since breathing waveform contains the information of flow rate, tidal volume and airway pressure for each inspiration and expiration. It can be taken as an accurate indicator for a number of diseases(including airway obstruction, congestive heart failure, atelectasis, etc.) and body status(including drowsy, drunk, etc.)[6]. Therefore, it is necessary to monitor the fine-grained breathing waveform of drivers in driving environments for preventing potential health problems and traffic accidents.

2.2 Capturing Breathing Waveform with Acoustic Signals

For ubiquitous breathing monitoring, RF and Wi-Fi sensing are developed quickly during these years, but they either need additional hardware implementations or not able to obtain the fine-grained breathing waveform. Since acoustic signals can be easily accessed without special hardware, and have relative slow transmission speed, it has been utilized to track the fine-grained movements

of objects[21][26][39][44]. These approaches are either frequency-related or phase-related, which are designed for quantitatively tracking the movements of objects, so they can accurately track the most notable movement (usually the largest/fastest movement or movements of nearest object) that cause a dominate frequency or phase change in acoustic signals. For those less notable movements (usually the smaller/slower movement comparing to other movements), these approaches take them as background noises. However, since drivers constantly perform notable body movements during driving, such as steering, braking, etc. Comparing to these movements, the breathing of drivers are less notable in frequency or phase changes. Therefore, these approaches based on phase or frequency could be unsuitable to sense body movements brought by breathing in driving environments.

Energy Spectrum Density (ESD) of acoustic signals describes how the energy of acoustic signals is distributed with frequency in space, and thus is able to sense the movements of all objects around as changes in energy distribution. Therefore, to monitor the breathing waveform in noisy driving environments, we utilize ESD of acoustic signals obtained by smartphones.

In the following, we first show that ESD signals can capture chest and abdominal movements brought by breathing of drivers, and further show that the breathing pattern will not be ignored in ESD signals even with the body movements of drivers during driving.

Specifically, an acoustic tone $T(t) = sin(2\pi ft + \phi)$, with the frequency f and the initial phase ϕ, is transmitted continuously through the speaker of a smartphone and then propagates with various paths, and finally gets received by the microphone of the same phone. The received signal at time t can be represented as:

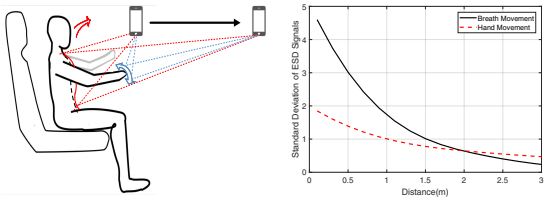
$$R(t) = \sum_{i \in \Omega} A_i sin(2\pi ft + \phi_i), \tag{1}$$

where Ω denotes the set of all paths of acoustic signals, A_i is the coefficient representing the amplitude reduction of the acoustic signal of path i, which approximately satisfies $A_i \propto 1/d_i$, with d_i been the corresponding propagation distance, ϕ_i is corresponding to the initial phase ϕ and the phase change during the propagation in path i. Then, ESD of $R(t)$ can be represented as:

$$\begin{aligned} ESD_{R(t)} &= (\sum_{i \in \Omega} A_i)^2 \\ &= (\sum_{j \in \alpha} A_j)^2 + (\sum_{k \in \Omega \setminus \alpha} A_k)^2 + 2 \sum_{j \in \alpha} A_j \sum_{k \in \Omega \setminus \alpha} A_k, \end{aligned} \tag{2}$$

where α is the set of paths related to the chest and abdomen of a driver, while $\Omega \setminus \alpha$ is the set of other possible paths.

It can be seen from Eq.2, when a driver breathes, the movements of the chest and abdomen cause the changes of the term $(\sum_{j \in \alpha} A_j)^2$, and then reflects on ESD of received signal $ESD_{R(t)}$. Concretely, with a smartphone in the vehicle, when the driver inhales, the distance of each corresponding path $d_j (j \in \alpha)$ becomes smaller, then the coefficient of acoustic signals that propagate in path $A_j (j \in \alpha)$ become larger ($A_i \propto 1/d_i$), resulting in the increase of energy spectrum $ESD_{R(t)}$. On the contrary, $ESD_{R(t)}$ decreases when the driver exhales. Therefore, ESD signals can be used to capture movements of chest and abdominal brought by breathing of drivers.



(a) Acoustic sensing for both breathing and hand movements (b) Standard deviation of ESD signals for both movements

Figure 2: Illustration of hand movement and breathing movement effects on ESD signals.

When driver performs some body movements during driving, such as steering, braking, etc., Eq.2 can be rewritten as

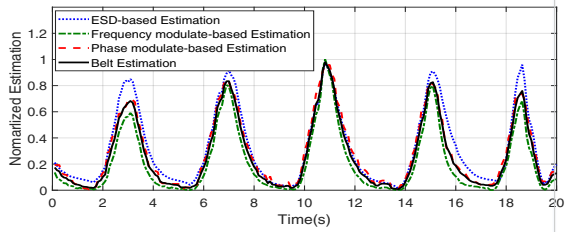
$$ESD_{R(t)} = (\sum_{j \in \alpha} A_j + \sum_{k \in \beta} A_k + \sum_{l \in \Omega \setminus (\alpha \cup \beta)} A_l)^2, \tag{3}$$

where α is the set of paths related to breathing of drivers, β is the set of paths related to body movement like steering, note that α and β may have some overlap paths, and $\Omega \setminus (\alpha \cup \beta)$ is the set of other possible paths for acoustic signals.

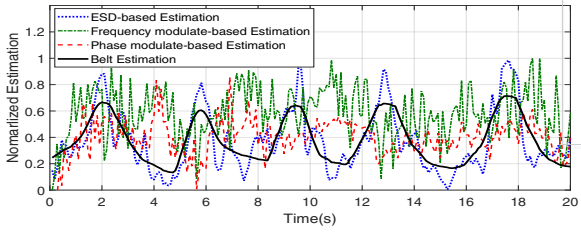
According to Eq.3, it is clear that both breathing and other body movement of a driver can be reflected in ESD signals as changes in term $\sum_{j \in \alpha} A_j$ and $\sum_{k \in \beta} A_k$, respectively. However, it is hard to determine form Eq.3 which term brings the dominate changes in ESD signals. Fig.2(a) shows a case when a driver is breathing while performing a hand movement at the same time. Although the hand movement usually has larger speed (i.e., larger change in individual A_i) comparing to the breathing, the corresponding paths of hand movements may be smaller(i.e., smaller coverage in β compared with α), as shown in Fig.2(a). To exploit the influence of these two movements in ESD signals, we conduct a simple experiment in the lab. During the experiment, the subject sits in a chair, while the smartphone that generates acoustic signals moves away from the subject step by step ($0.2m$ per step). For each step, the smartphone computes the standard deviation of ESD signals for breathing and hand movement(hold the breath when performing) of the subject, respectively. The result is shown in Fig.2(b). It can be seen from Fig.2(b) that breathing brings larger changes in ESD signal when the distance is within $2m$, and becomes the dominate change($2X$ larger than hand movements) when the distance is smaller than $1m$. Since the driving environment is a rather small space ($< 1m$), the changes brought by breathing of a driver is always non-ignorable in ESD signals.

To obtain ESD signals, we use a smartphone to generate continues pilot tones from the speaker and collect acoustic signals from the microphone. The frequency of the pilot is set as $20kHz$ to avoid been interfered by acoustic noises in driving environments (e.g., music, talking, etc.) and heard by people. The sampling rate of the microphone is set at $48kHz$, which is the highest sampling rate for off-the-shelf smartphones. After that, we transform the collected acoustic signal into frequency domain using 2048-point Fast Fourier Transformation (FFT), which corresponds to the time resolution of $40ms$. Then, ESD of received signals can be calculated by:

$$ESD_{R(t)} = \frac{FFT_N(R(t))^2}{N}, \tag{4}$$



(a) Case 1: a driver sits in vehicle without movements



(b) Case 2: a driver in the driving state

Figure 3: Illustration of different methods for estimating breathing in driving environments.

where $N = 2048$. In practice, we only care about ESD signals around $20kHz$ in frequency domain, then Eq.4 turns to:

$$ESD_{R(t)} = \frac{1}{N} \sum_{f=20kHz-\Delta f}^{20kHz+\Delta f} (R(f))^2, \tag{5}$$

where $R(f) = FFT_N(R(t))$ and Δf is set to be $20Hz$ considering the influence of Doppler effect to the pilot.

We further conduct an experiment to capture the breathing pattern of a volunteer utilizing frequency modulate-based approach, phase modulate-based approach and ESD signals, respectively. These three approaches are tested in two cases: case 1 is the scenario that the volunteer sits in a static vehicle without any movements, so that the only movement of the body is caused by breathing; case 2 is the scenario that the volunteer is driving, so the movements of the volunteer contains breathing and other body movements. During the experiment, the smartphone is fixed on the instrument panel of the vehicle, and the subject wears a NEULOG respiration monitor logger sensor[36] to provide the baseline of breathing waveform.

Fig.3 illustrates the results of different acoustic approaches in our experiment, where we randomly sample a 20s slice from each case. It can be seen from Fig.3(a) that all three methods can capture the breathing pattern when the volunteer and vehicle are static, and comparing to baseline, frequency modulate-based and phase modulate-based method performs better than ESD signals because they can quantitatively tracking the movements of breathing. While as shown in Fig.3(b), when the volunteer is driving, the estimating results of frequency modulate-based and phase modulate-based method become much worse because there are multiple body movements of the volunteer. In contrast, although is relatively noisy compared to Fig.3(a), ESD signals can still capture the breathing of the volunteer driver as periodic up-and-down trends. This result shows the potential of utilizing ESD signals for capturing the breathing waveform in driving environments.

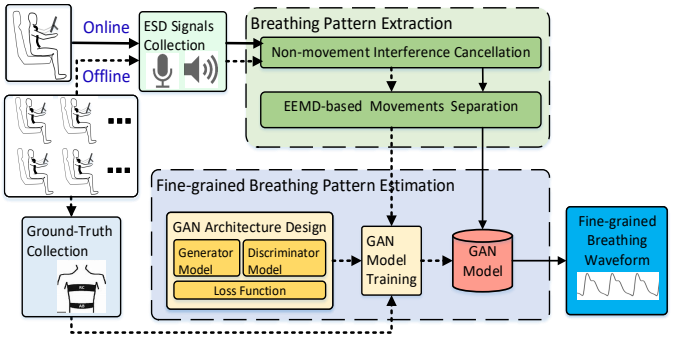


Figure 4: System architecture of *BreathListener*.

However, challenges emerge when applying ESD of acoustic signals for capturing the breathing waveform of drivers. First, acoustic signals transmitted in driving environments are easily interfered by multi-path transmissions, especially the line-of-sight(LOS) signals. Second, breathing patterns need to be extracted from ESD signals with the existence of other body movements and vehicle dynamics. Third, the detailed breathing waveform of drivers need to be constructed from the extracted breathing pattern in ESD signals.

3 SYSTEM DESIGN

We propose a breathing monitoring system, *BreathListener*, which estimates the fine-grained breathing waveform in driving environments leveraging acoustic devices on smartphones.

3.1 System Architecture

Fig.4 shows the architecture of *BreathListener*. The whole system can be divided into offline part and online part.

Offline Procedure. In the offline part, *BreathListener* first collects the acoustic signals of drivers during driving and computes ESD of acoustic signals. After that, in *Breathing Pattern Extraction*, *BreathListener* eliminates the interferences from driving environments and extract the breathing pattern in ESD signals . Particularly, *BreathListener* first reduces the interference of multi-path transmission in ESD signals via *Non-movement Interference Cancellation*. Then, *BreathListener* separates breathing from other movements and dynamics from driving environments in ESD signals leveraging Ensemble Empirical Mode Decomposition (EEMD), and obtain coarse-grained breathing patterns in ESD signals.

Afterwards, in *Fine-grained Breathing Pattern Estimation*, *BreathListener* designs a deep learning architecture based on Generative Adversarial Networks (GAN) for estimating the fine-grained breathing waveforms. In particular, *BreathListener* trains a GAN model utilizing the coarse-grained breathing patterns from *Breathing Pattern Extraction* and the corresponding ground-truth collected from real driving environments.

Online Procedure. In the online part, *BreathListener* obtains the coarse-grained breathing pattern of a test driver through *ESD Signals Collection* and *Breathing Pattern Extraction*, same as described in offline part. Then, in *Fine-grained Breathing Pattern Estimation*, *BreathListener* utilizes the GAN model trained by offline part. The model takes as input the coarse-grained breathing patterns from *Breathing Pattern Extraction*, and generates fine-grained breathing waveforms for the test driver.

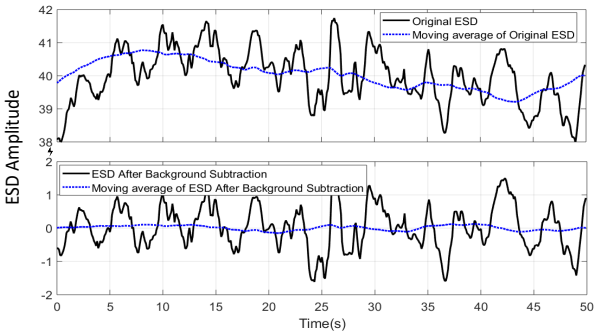


Figure 5: Illustration of non-movement interference cancellation for ESD signals.

3.2 Breathing Pattern Extraction

According to Section 2.2, although ESD of acoustic signals can sense the periodic up-and-down trends of chest and belly brought by breathing procedures of the driver, it is still challenging to extract the breathing pattern from ESD signals due to the interferences in driving environments. Concretely, there are two types of interferences: One is the non-movement interference, which consists of the line-of-sight(LOS) signals that transmitted directly from speakers to microphones on smartphones, and also the multi-path interferences of static objects in driving environments. The other type of interference is the multi-path interferences of movements in driving environments, including movements of drivers and dynamics of vehicles. To extract the breathing pattern from ESD signals, *BreathListener* need to address these two types of interferences.

3.2.1 Non-movement Interference Cancellation. In order to extract the breathing pattern in ESD signals, *BreathListener* first needs to cancel the non-movement interference in ESD signals, which contains line-of-sight(LOS) interference and multi-path interference of static background objects in driving environments. To solve the problem, we utilize the idea of background subtraction[31] to build a background model. Once the smartphone is fixed at a certain position in the vehicle, the background model automatically measures the interferences of LOS signal and static objects in ESD of acoustic signals. Specifically, given k successive ESD measurements denoted as $\{ESD_1, \ldots, ESD_k\}$, the background model is initialized as:

$$B_0 = \frac{1}{k}\sum_{i=1}^{k} ESD_i. \tag{6}$$

Since the position of the smartphone and the corresponding driving environments could change during driving (e.g., open the window, adjust the seats, etc.). To adaptive to the changes, the background model is continuously updated by:

$$B_n = (1-\alpha)B_{n-1} + \alpha ESD_n, \tag{7}$$

where B_n is the background model for ESD signal ESD_n, $n \in \{1, 2, \ldots, k\}$, and $\alpha \in [0, 1]$ is the updating rate. It is clear that the updating rate α is crucial to an accurate measurement of the background model B_n, and *BreathListener* calculates α as:

$$\alpha = \frac{||ESD_n - ESD_{n-1}||}{Max(ESD_n, ESD_{n-1})}. \tag{8}$$

The intuition of Eq.8 is that the combination of LOS signals and static objects reflections are orders of magnitude larger than other reflections. As a result, when ESD signals changes greatly, the background model B_n should adapt to the change as soon as possible for an accurate measurement of non-movement interferences.

Fig.5 shows the original ESD signal and the signal after non-movement interference cancellation. It can be seen from Fig.5 that after the cancellation, the ESD signal is regulated to nearly zero-mean, which means most non-movement interferences are canceled. Moreover, the moving averages of the two signals using a $10s$ window are also given in Fig.5. It can be seen that the moving average of the signal after non-movement interference cancellation is much more flat comparing to the original ESD signals, showing that the long-term change trend in ESD signals, which could be brought by the changes of environments, is also filtered out.

3.2.2 EEMD-based Movements Separating. Besides non-movement interferences, ESD of acoustic signals is also influenced by reflections of moving objects in driving environments. Therefore, after eliminating non-movement interferences, *BreathListener* further separates breathing from other movements in ESD signals.

In driving environments, there are three common categories of movements as follows:

- Vibrations of the vehicle, which are introduced by the movements of the vehicle when driving, usually have a relatively high frequency ($0.5Hz - 10Hz$).
- Movements of the driver, which refer to movements of operating steering wheel, shifting gears, etc., have relatively low frequency ($0.1Hz - 2Hz$).
- Breathing of the driver, which has a certain frequency from 10 to $35bpm$ ($0.16Hz - 0.6Hz$)[25].

Since the frequency ranges of breathing have multiple overlaps with other two categories of movements, we can not separate breathing effectively by applying band-pass filters of its frequency range($0.16Hz - 0.6Hz$). Moreover, since breathing and other body movements can happen at the same distance from the smartphone, it is also hard to separate breathing by distance-based approaches[22].

To separate breathing from other movements under frequency and distance overlaps, we need a filter bank that is adaptive to ESD signals(i.e., the filter bank can figure out the frequency range of each filter automatically according to ESD signals), and the frequency ranges of the filters in the bank are overlapped. Empirical Mode Decomposition (EMD)[13] behaves as such a self-adapting dyadic filter bank that can satisfy the requirement, if the signal consists of white noise which has scales populated uniformly through the whole time-scale or time-frequency space[7][41]. However, uniform distributed white noise are often not available in real-world signals. Since Ensemble Empirical Mode Decomposition(EEMD) method[42] can add white noise to the signal before performing EMD. We use EEMD to make the noises of ESD signals close to uniform distributed white noise. In EEMD method, the original signal is first added with the white noise of standard derivation e, and then decomposed by EMD. The noise-adding and decomposing procedure repeat n times, and the n decompositions are averaged as the final result. It should be notice that there are two key parameters(i.e., e and n) that need to be set properly when applying EEMD in *BreathListener*. For achieving a promising decomposition performance, n need to be sufficiently large (usually $n > 100$), and e should be moderate to the standard derivation

Algorithm 1: EEMD algorithm in *BreathListener*

Require:
ESD signals of a time period $ESD(t)$;
Standard derivation of add white noise in EEMD e
Parameter of repeating time in EEMD n;
Ensure:
Breathing pattern in ESD signals C_x;
1: Initialize $e = (ESD_{max} - ESD_{min})/4$, $n = 1000$;
2: Run EEMD with ESD, e and n;
3: Compute FFT for each component $C_i, i = 1, 2, 3, \ldots$ and find the component $C_x (x \in \{1, 2, 3, \ldots\})$ with frequency ranges within $0.16Hz - 0.6Hz$;
4: Compute the standard derivation of C_x, $Std(C_x)$;
5: **while** $||e - 0.55 \times std(C_x)|| > 0.45 \times Std(C_x)$ **do**
6: $e = std(C_x)$;
7: Repeat step 2 to step 4;
8: **end while**
9: Output C_x as extracted breathing pattern

of the desired signal($0.1 - 1$ times of the standard derivation of the desired signal is considered as proper according to empirical studies[42]). In the case of *BreathListener*, the desired signal is the breathing-related ESD signal. However, we can not directly get the standard derivation the signal. So we design an iterative algorithm to set the proper parameters of EEMD in *BreathListener* for extracting the breathing pattern, described as **Algorithm 1**. As described in line 1 and line 2 of **Algorithm 1**, *BreathListener* first conducts decomposition on ESD signals using EEMD method with $n = 1000$ and $e = (ESD_{max} - ESD_{min})/4$. Fig.6 demonstrates the decomposition result during a 40s driving. In Fig.6, the ESD signal is decomposed to four main components, i.e., C_1, C_2, C_3 and C_4. And then *BreathListener* applies FFT to each component to find the component whose frequency is mainly within breathing range($0.16Hz - 0.6Hz$), i.e., C_3 in Fig.6. After that, as show in line 3 to line 8 of **Algorithm 1**, *BreathListener* computes the standard deviation of C_3 as $std(C_3)$ and compares it to e. If $std(C_3)$ is in the proper range(i.e., $||e - 0.55 \times std(C_x)|| > 0.45 \times Std(C_x)$), then *BreathListener* considers the parameters of EEMD are proper and outputs C_3 as the extracted breathing pattern. If not, *BreathListener* sets $e = std(C_3)$ and re-conduct EEMD for ESD signals.

To verify the algorithm, we compare the extracted breathing pattern C_3 to actual breathing waveform collected by specialized sensors and show the result in Fig.7. It can be seen that the extracted breathing pattern roughly captures the breathing procedure of the driver in driving environments.

3.3 Fine-grained Breathing Waveform Estimation

In driving environments, *BreathListener* aims to obtained drivers' fine-grained breathing waveforms, which can accurately indicate the body status of drivers.

Although *BreathListener* can extract the breathing pattern of a driver after eliminating the interferences of driving environments in ESD signals, it can be seen from Fig.7 that the fine-grained characters of actual breathing waveform, such as the amplitude and

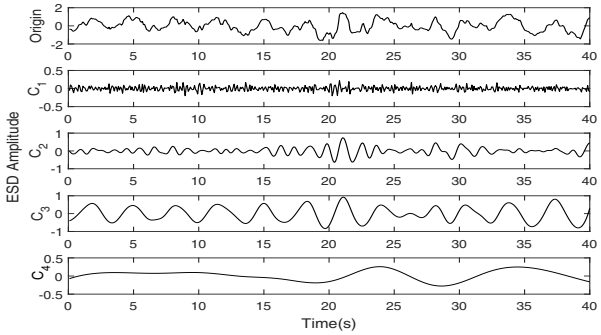


Figure 6: Illustration of EEMD decomposition for ESD signals with $n = 1000$ and $e = 0.8$.

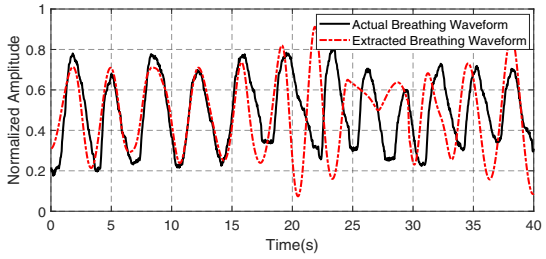


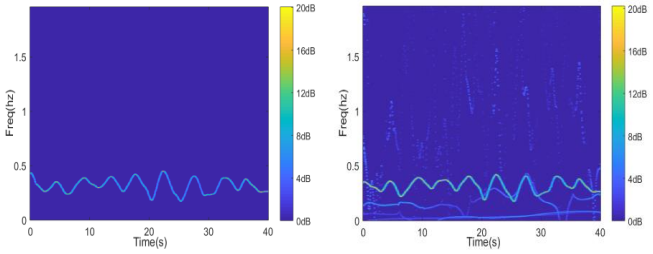
Figure 7: Breathing waveform extracted from ESD signals & actual breathing waveform during a 40s driving.

interval of breathing, are not accurately captured in the extracted breathing waveform.

Since Hilbert spectrum[12] describes the instant frequency of the waveforms, we transform the extracted breathing waveform and actual breathing waveform to Hilbert spectrum to exploit the relationship between the two waveforms, as shown in Fig.8. It can be seen that the Hilbert spectrum of extracted breathing waveform (shown in Fig.8(a)) only contains the main frequency component comparing to the Hilbert spectrum of actual breathing waveform (shown in Fig.8(b)), and thus it is coarse-grained. In order to obtain the fine-grained waveform, we need to transform the coarse-grained Hilbert spectrum(Fig.8(a)) to fine-grained Hilbert spectrum(Fig.8(b)). This problem is similar as the image super-resolution problem to obtain high-resolution images from low-resolution images. And the image super-resolution has nice solutions by Generative Adversarial Network[8](GAN)-based deep learning methods[35][17][40]. Therefore, we propose a GAN-based deep learning architecture to obtain the fine-grained breathing waveform.

3.3.1 GAN Background. Generative Adversarial Network (GAN) is a generative model that aims to generate data that follows a target distribution (e.g., images) from the input distribution(e.g., noises). GAN consists of two key components: a generator model G that learns to generate data $G(z)$ that approximately satisfies a target distribution P_Y, from the input z sampled from some distributions P_Z; and a discriminator model D that learns to distinguish between the generated data $G(z)$ and the data y sampled from the actual target data distribution P_Y. In other words, when training GAN, D and G play the two-player minimax game[8] as:

$$\min_G \max_D \mathbb{E}_{y \sim P_Y} \log D(y) + \mathbb{E}_{z \sim P_Z} \log(1 - D(G(z))), \quad (9)$$



(a) Hilbert Spectrum of Extracted Breathing Waveform (b) Hilbert Spectrum of Actual Breathing Waveform

Figure 8: Hilbert spectrum of the extracted breathing waveform and actual breathing waveform in Fig.7.

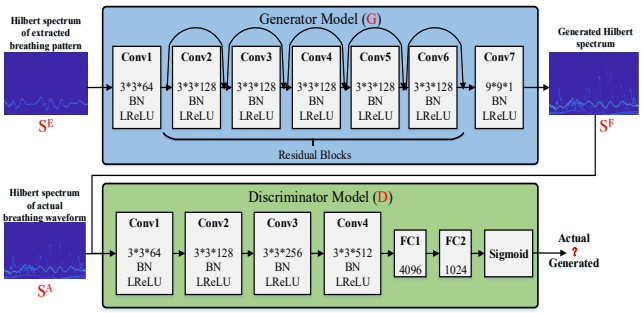


Figure 9: GAN architecture of *BreathListener*.

where $D(x)$ represents the probability that x comes from the actual target data distribution P_Y.

3.3.2 Building a GAN Architecture for BreathListener. In *BreathListener*, we construct a GAN model to obtain fine-grained breathing waveform. Fig.9 shows the architecture of GAN model in *BreathListener*. It can be seen from Fig.9 that, given the Hilbert spectrum of extracted breathing pattern S^E as input, the generator model G aims to output the Hilbert spectrum of fine-grained breathing waveform S^F. Then S^F is fed to the discriminator model D, together with the Hilbert spectrum of corresponding actual breathing waveform S^A (S^E, S^F and S^A can be denoted as real-valued matrices of size $W \times H$). The discriminator model D aims to discriminate whether the spectrum comes from the real dataset or is generated by G. With the GAN architecture, the generator model G aims to fool the discriminator model D, while the discriminator model D aims not to be fooled by the generator model G. By training G and D iteratively, the generator model G is able to generate Hilbert spectrum of breathing waveform S^F, which can be close to actual breathing waveform S^A.

Generator Model Architecture: the generator model G of *BreathListener* is basically a Convolutional Neural Network (CNN)[16], as illustrated in Fig.9. As we can observe from the figure, the generator model is designed with 7 convolution blocks for capturing the fine-grained breathing patterns in Hilbert spectrum. The first block is to unfold the input to features, and the last block is to produce the output. Between them are 5 identical residual blocks[10], which are designed to learn the deep representation of features in Hilbert spectrum. Within each residual block, we implement a convolutional layers with 3×3 convolution kernels and 128 feature channels followed by a Batch-Normalization (BN)[14] layer, and then a Leaky RuLU (LReLU)[20] layer as the activation function. With this complex CNN architecture, the generator model G is capable to learn the sophisticated mapping from the extracted coarse-grained breathing pattern to fine-grained breathing waveform in Hilbert spectrum.

Discriminator Model Architecture: Similar to the generator model G, the discriminator model D is also a CNN as shown in Fig.9. Since the task of D is relatively light, we design D with only 4 convolution blocks, each with a 3×3 convolution kernels followed by a BN layer and an LReLU layer. The feature channels of these blocks increase by a factor of 2 from 64 to 512, which is inspired by the famous VGG network[38]. Then the 512 channels are sent through two Fully-Connected (FC)[16] layers to merge the features, and finally a sigmoid layer to output the probability that current input Hilbert spectrum comes from the dataset of actual breathing pattern.

Loss Function Design: Since GAN model of *BreathListener* is a supervised learning model, the target of training is to minimize the loss function, which represents the error that *BreathListener* want to reduce after training. In *BreathListener*, we formulate the loss function as:

$$l = l_c + l_r + l_{ad}. \tag{10}$$

l_c is the content loss, representing the error between the generate Hilbert spectrum S^F and the Hilbert spectrum of actual breathing pattern S^A, and is defined as

$$l_c = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} (S^A_{x,y} - G(S^E)_{x,y})^2, \tag{11}$$

where $G(S^E)$ is same as S^F. It can be seen from Eq.11 that we calculate the differences element-wisely between S^A and S^F to build the the content loss, which can help to capture the details of actual breathing waveform in Hilbert spectrum. l_r is the reconstruction loss, representing the error between the reconstructed breathing waveform W^F from S^F and the actual breathing waveform W^A, and is denoted as:

$$l_r = \frac{1}{L} \sum_{i=1}^{L} (W^A_i - H^{-1}(G(S^E))_i)^2, \tag{12}$$

where L is the length of W^A, and $H^{-1}(\cdot)$ represents the reverse Hilbert transformation, which transforms the Hilbert spectrum to breathing waveform. l_{ad} is the adversarial loss for training GAN model, which can be denoted as[8]:

$$l_{ad} = \log[1 - D(G(S^E))], \tag{13}$$

where $D(G(S^E))$ represents the probability that the generated Hilbert spectrum $G(S^E)$ is recognized as the Hilbert spectrum of actual breathing pattern S^A by the discriminator model D.

By design loss function, *BreathListener* can train the GAN model to generate Hilbert spectrum of fine-grained breathing waveform.

3.3.3 Training GAN Model for BreathListener. We collect data from real driving environments to train the GAN model of *BreathListener*. Specifically, we develop an Android-based program to generate and collect acoustic signals. The program is designed as an Android App and installed on 5 different types of smartphones, which are Samsung Galaxy S6, Samsung Galaxy S7, Google Pixel, HTC U

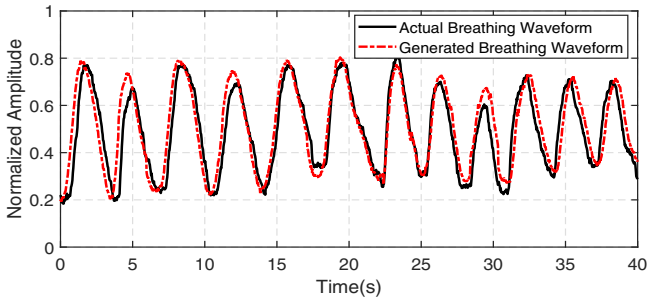


Figure 10: Breathing waveform generated by GAN and actual breathing waveform collected by specialized sensors during a 40*s* driving.

Ultra, and Huawei Mate8, respectively. 20 volunteer drivers (11 males, 9 female, ages $19-55$) are recruited to the data collection, with an extra NEULOG Respiration Monitor Logger Sensor worn by each driver to collect the ground-truth. The smartphones are placed in different places in vehicles, including the instrument panel, cab door, cupholder, etc. The data collection spans 5 days and we collect 100 hours data for training.

After data collection, we separate the acoustic signal into slices with a sliding window of 30*s* length and 1*s* sliding step, which produces nearly 350000 slices. Then *BreathListener* calculates the ESD of the slices and further extract the breathing pattern in ESD signals utilizing methods in Sec.3.2. After that, the extracted breathing patterns are transformed to Hilbert spectrum as training dataset X. By combing samples in X and corresponding ground truth as pairs of training samples, the parameters of generator model G and discriminator model D are updated iteratively until the cost function Eq.10 converges. The training progress is implemented and conducted on Google cloud server.

3.3.4 Generating fine-grained breathing waveforms. After training, *BreathListener* utilizes the GAN model to generate Hilbert spectrum that contains fine-grained breathing characters based on the Hilbert spectrum of coarse-grained breathing patterns. Then *BreathListener* obtains the fine-grained breathing waveform by transforming the Hilbert spectrum to the time domain. Fig.10 shows the breathing waveform generated by *BreathListener* compared to the corresponding actual breathing waveform during a 40*s* driving. It can be observed that the breathing waveform generated by GAN is very close to the actual breathing waveform, even in the detailed trends during each inhale and exhale.

4 EVALUATION

In this section, we evaluate the performance of *BreathListener* in real driving environments.

4.1 Setup

We recruit 10 volunteer drivers (6 males, 4 females, ages $24-47$, different from the volunteer drivers in Sec.3.3.3.) to conduct experiments for evaluating the performance of *BreathListener*, which is implemented as an Android APP that can estimate the breathing waveform in driving environments. The APP is installed in 10 smartphones, which are of the same 5 types as described in Sec.3.3.3, 2 for each type. During the experiments, drivers drive with their own vehicles, and the smartphones implemented with *BreathListener* APP are placed in different places as the wish of drivers, including the instrument panel, cab door, cupholder, etc., and drivers are also free to play music, open the window, etc., during driving. Meanwhile, the ground truth for the breathing pattern of drivers is monitored by the NEULOG respiration monitor logger sensor. We conduct our experiments from 9th April to 22nd April 2018, during which all the daily driving, such as commuting, shopping, etc., is recorded.

For the purpose of evaluation, we also conduct breathing rates estimation based on *BreathListener* system. Specifically, once *BreathListener* obtains the breathing waveform, we compute the breathing rate based on peak identification algorithm[19]. And the ground-truth of breathing rate is achieved by the NEULOG sensor.

4.2 Metrics

To evaluate the performance of *BreathListener*, we define metrics as follows:

- **Rate Estimation Error**: The error of the estimated breathing rate R^E from actual breathing rate R^A, which is defined as the absolute difference between R^E and R^A, i.e., $|R^E - R^A|$.
- **Waveform Estimation Error**: The error between estimated breathing waveform W^E and actual breathing waveform W^A, which is defined as the same form of reconstruction loss in Eq.12, where $H^{-1}(G(S^E))_i$ in Eq.12 is replaced by W_i^E here. The range is $[0, 1]$.
- **Correlation Coefficient**: The shape similarity between estimated breathing waveform W^E and actual breathing waveform W^A, which is defined as $\frac{Cov(W^E, W^A)}{\sigma(W^E)\sigma(W^A)}$, where $Cov(\cdot)$ and $\sigma(\cdot)$ computes the covariance and standard deviation, respectively. The range is $[-1, 1]$.

The claim of being able to track the breathing pattern would make more sense if results show that breathing patterns are recognized at transitions as well i.e., from normal to fatigue or slow to fast. As of now, results are not sufficient for the claim. It is more of tracking a type of periodic waveform. It is necessary to see how the proposed signal decomposition and identification algorithms work during transitions.

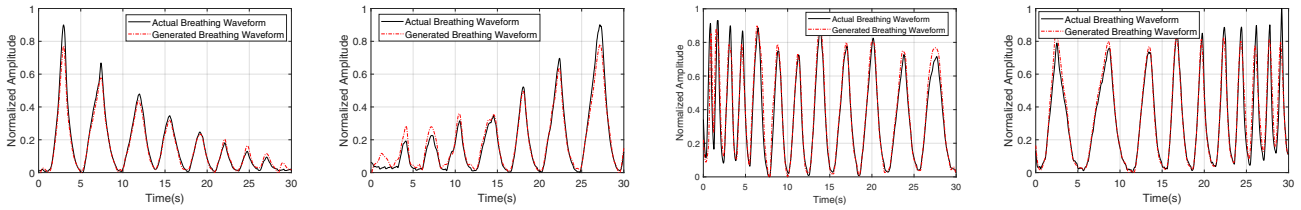
4.3 Micro Benchmark

We provide several micro benchmarks of *BreathListener*, including the demonstration of breathing waveform estimation and running time of each component in *BreathListener*.

4.3.1 Breathing transitions evaluation. Since *BreathListener* aims at monitoring the breathing of drivers, we care about the performance for both normal and abnormal breathing. The demonstration of normal breathing is shown in Fig.10, and we then focus on abnormal breathing. We ask two drivers to intentionally change their breathing pattern in four cases during driving, including breathing from deep to shallow, from shallow to deep, from fast to slow and

Table 1: Running time of *BreathListener*

Time cost	ESD	Intf. cancel	EEMD	GAN	Total
Samsung Galaxy S6	0.19s	0.08s	1.81s	0.93s	3.01s
Samsung Galaxy S7	0.05s	<0.01s	0.67s	0.26s	0.98s



(a) From deep to shallow (b) From shallow to deep (c) From fast to slow (d) From slow to fast

Figure 11: Breathing waveform estimation of *BreathListener* in abnormal transitions.

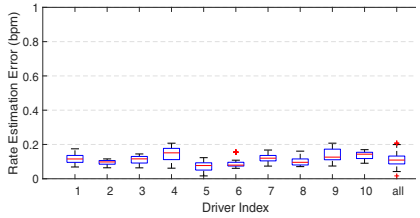


Figure 12: Box plot of breathing rate estimation error for drivers.

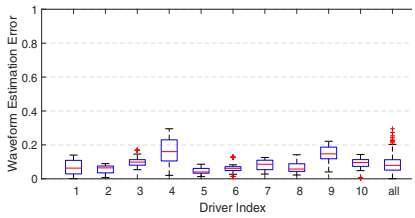


Figure 13: Box plot of breathing waveform estimation error for drivers.

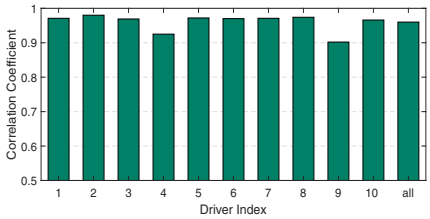


Figure 14: Correlation coefficient of breathing waveform for drivers.

from slow to fast. The result is shown in Fig.11. It can be observed that the estimated breathing waveform of *BreathListener* is close to the actual breathing waveform for all these abnormal breathing transitions. So *BreathListener* is able to obtain fine-grained breathing waveform in abnormal breathing transitions. In addition, it also can be seen from Fig.11 the amplitude of estimated breathing waveform could deviate a bit from the actual breathing waveform when the breathing is abnormal. The reason is that the training data of GAN model in *BreathListener* consists much less information for abnormal breathing than normal breathing. The performance of *BreathListener* can be further improved with more abnormal breathing data for training.

4.3.2 Running time evaluation of each components. For evaluating the real-time performance of *BreathListener*, we test the running time for each component in *BreathListener* on two types of smartphone of different computing power, i.e., Samsung Galaxy S6 with Mali-T760 CPU and Samsung Galaxy S7 with MSM8996 CPU. MSM8996 is much more powerful than Mali-T760. Table.1 shows the running time results. It can be seen from Table.1 that ESD signal computation and non-movement interference cancellation together can be finished within $300ms$ for these two types of smartphones. While the running time of EEMD-based movements separation and GAN model is relatively longer. In total, for Samsung Galaxy S7, the running time of *BreathListener* is less than $1s$, which is appropriate for a real-time breathing monitoring system. For Samsung Galaxy S6, the running time of *BreathListener* is about $3s$, which shows that less-powerful smartphones could influent the real-time performance of *BreathListener*.

4.4 Overall Performance

We evaluate the overall performance of *BreathListener* for different drivers and different breathing rate ranges of drivers. Both normal and abnormal breathing of drivers are included in the following evaluation.

4.4.1 Evaluation for Different Drivers. Since breathing rate directly shows a coarse-grained result of breathing estimation, we compute the breathing rate of drivers base on the breathing waveform obtained by *BreathListener*. Fig.12 shows the breathing rate estimation error of each driver. It can be seen from Fig.12 that the breathing rate estimation errors for each of the 10 drivers is no larger than $0.3bpm$ (breath per minute), and the average estimation error for all drivers is as low as $0.11bpm$, showing that *BreathListener* can achieve accurate breathing rate estimation.

We show the breathing waveform estimation errors for the 10 drivers in Fig.13. As we can observe, the waveform estimation error for each driver is always lower than 0.28, and the average estimation error for breathing waveform is as low as 0.08, showing that the breathing waveform estimation of *BreathListener* is accurate and robust to various drivers. We also show the correlation coefficient of breathing waveform estimation in Fig.14, with all correlation coefficient larger than 0.9 and the average correlation coefficient larger than 0.95. Correlation coefficient larger than 0.8 indicates a strong positive relationship. So the breathing waveform estimation of *BreathListener* is highly related to actual breathing waveform.

It also can be seen from Fig.12, Fig.13 and Fig.14 that the estimation results for driver 4 and driver 9 is not as good as other drivers. The reason is that driver 4 prefers to carry the smartphone in the pocket of trousers during driving, and driver 9 likes wearing over-sized coats during driving. We will show the impact of smartphone placements and clothing of drivers to *BreathListener* in the following evaluations.

4.4.2 Evaluation for Different Breathing Rages. Since breathing rate of people ranges from 10 to $35bpm$[25], including slow breathing ($10-16bpm$), normal breathing ($16-25bpm$) and fast breathing ($25-35bpm$), we further evaluate the performance of *BreathListener* under different breathing rate range. Fig.15 shows breathing rate estimation error under different breathing rate range. It can be seen that although the variance of rate estimation error increases

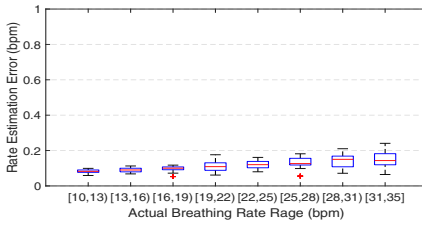


Figure 15: Box plot of breathing rate estimation error under different breathing rate ranges.

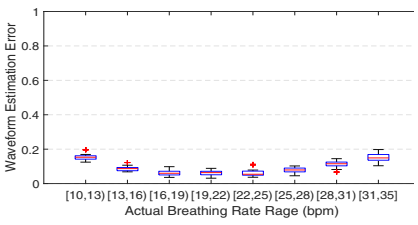


Figure 16: Box plot of breathing waveform estimation error under different breathing rate ranges.

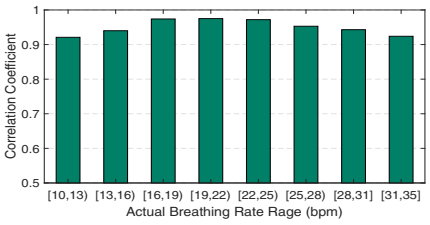
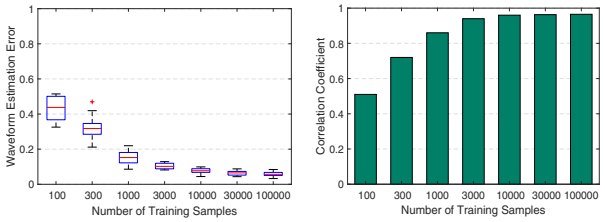
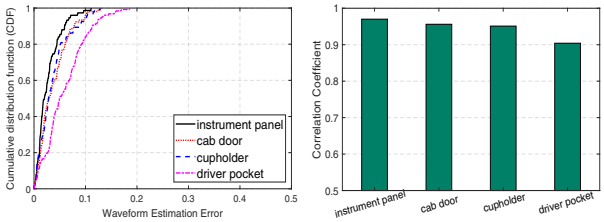


Figure 17: Correlation coefficient of breathing waveform under different breathing rate ranges.



(a) Waveform estimation error. (b) Correlation coefficient.

Figure 18: Impact of training set size.



(a) CDF of Waveform estimation error. (b) Correlation coefficient.

Figure 19: Impact of smartphone placements.

a bit as the breathing rate increases, the rate estimation error is always lower than 0.22*bpm*, showing *BreathListener* can estimate the breathing rate accurately no matter the breathing rate of a driver is normal, fast or slow.

Fig.16 shows the waveform estimation error of *BreathListener* for different breathing rate ranges. We can observe from Fig.16 the waveform estimation error becomes slightly larger than normal cases when the actual breathing is slow (lower than 16*bpm*) or fast (higher than 25*bpm*). This is because the training data of normal breathing are much more than slow or fast breathing for training the GAN model. However, even in the worst case, the waveform estimation error is no larger than 0.2. We also show the correlation coefficient for different breathing rate ranges in Fig.17. It shows similar results as in Fig.16, with the correlation coefficient always larger than 0.92.

Since *BreathListener* can always achieve accurate breathing rate estimation (lower than 0.3*bpm*) in the above evaluation, and breathing waveform is a more fine-grained result comparing to breathing rate, we thus focus on the performance of breathing waveform estimation(i.e., the waveform estimation error and correlation coefficient) in the following evaluation.

4.5 Impact of Training Set Size

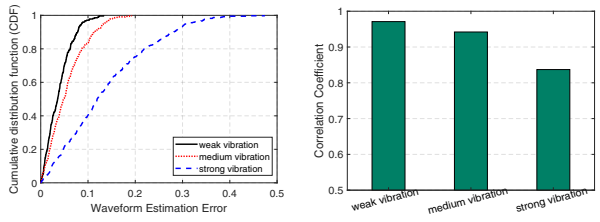
According to Section 3.3.3, we collect about 100 hours driving data (about 350000 samples) for training GAN model of *BreathListener*. Fig.18 shows the impact of training set size on the performance of breathing waveform estimation. It can be seen that as training set size increases, the waveform estimation error first decreases and then goes stable (as shown in Fig.18(a)), and the correlation coefficient (as shown in Fig.18(b)) first increases and then goes stable. Specifically, *BreathListener* obtain the waveform error less than 0.1 and correlation coefficient larger than 0.95 with 10000 training samples, which corresponds to only 3 hours driving data. More data improves little to the performance.

4.6 Impact of Smartphone Placements

In our experiments, we study the impact of smartphone placements for *BreathListener* by placing smartphones on four different places during driving, i.e., instrument panel, panel near cab door, cup-holder and pocket of drivers. Fig.19(a) shows the cumulative distribution function (CDF) of waveform estimation error. It can be seen that *BreathListener* achieves less than 0.14 error for more than 90% cases in all smartphone placements. The error of placing smartphones in drivers' pockets is slightly larger than other cases, because acoustic could be blocked and interfered by the driver's body. In addition, Fig.19(b) shows the correlation coefficient in different smartphone placements. As we can see, the correlation coefficient for instrument panel, panel near cab door and cup-holder are about 0.97, 0.955 and 0.95, respectively. For driver's pocket, the correlation coefficient is still larger than 0.9.

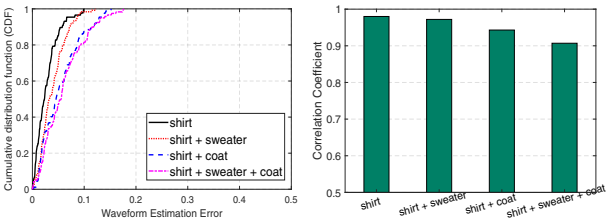
4.7 Impact of Smartphone Vibrations

Vibrations could bring movements to smartphones, and further influent the performance of *BreathListener*. To evaluate the impact, we divide smartphone vibrations for a time period into three levels, i.e., weak vibration, medium vibration and strong vibration. Specifically, we use the standard derivation of accelerometer readings as an indicator to determine the level of smartphone vibrations. During a time period, if the largest standard derivation is smaller than 1 for all three axis, we take the case as weak vibration. If the largest standard derivation is between $1-5$, which could be brought by slight shakes of vehicles during driving, we take the case as medium vibration. And if the largest standard derivation is larger than 5, which could occur when driving on successive uneven roads, we take the case as strong vibration. We plot the CDF of the waveform estimation error and the correlation coefficient of breathing waveform estimation under different levels of smartphone vibration in Fig.20(a) and Fig.20(b), respectively. It can be seen from Fig.20(a) and Fig.20(b) that the waveform estimation error of *BreathListener* is lower than 0.1 for more than 80% cases



(a) CDF of Waveform estimation error. (b) Correlation coefficient.

Figure 20: Impact of smartphone vibrations.



(a) CDF of Waveform estimation error. (b) Correlation coefficient.

Figure 21: Impact of drivers' clothing.

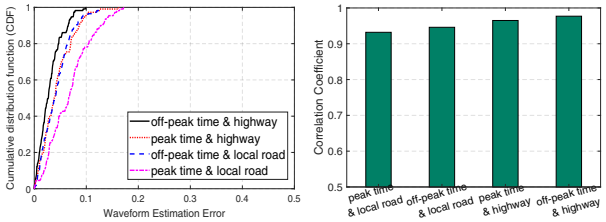
of weak and medium vibration, and the corresponding correlation coefficient is larger than 0.94. The result shows that *BreathListener* can still obtain accurate breathing waveform estimation under weak or medium smartphone vibration. For strong smartphone vibration, the waveform estimation error is larger than 0.15 for 40% cases, and the correlation coefficient is about 0.83. Therefore, strong smartphone vibration(which could occur when driving on successive uneven roads) could influent the accuracy of estimated breathing waveform in *BreathListener*.

4.8 Impact of Clothing

We conduct an extensive experiment to evaluate the performance of *BreathListener* under different clothing of drivers. In particular, we evaluate four different typical clothing: shirt, shirt+sweater, shirt+coat, and shirt+sweater+coat. The result is shown in Fig.21. It can be seen that for all kinds of clothing, *BreathListener* achieves that 80% waveform estimation errors are less than 0.1. Moreover, it can be seen from Fig.21(a) that *BreathListener* performs slightly better when drivers wear less clothes. This is because clothes, especially loose clothes like coats, could block acoustic signals and partly hide the body movement brought by breathing. In addition, it can be seen from Fig.21(b) that for the worst case when the driver wears 'shirt+sweater+coat', the correlation coefficient is still larger than 0.91. Therefore, *BreathListener* can estimate the breathing waveform accurately even when drivers wear a lot.

4.9 Impact of Road Types and Traffic Conditions

Traffic conditions and road types may influent drivers' driving behaviors and vehicle conditions, thus could impact the performance of *BreathListener*. We analyze the collected traces of different traffic conditions (during peak time and off-peak time) and different road types (on local road and highway), respectively. Fig.22(a) shows the CDF of the waveform estimation error under all four combinations of road types and traffic conditions. It can be seen that the



(a) CDF of Waveform estimation error. (b) Correlation coefficient.

Figure 22: Impact of road types and traffic conditions.

waveform estimation error is lower than $0.11bpm$ for more than 80% cases at any combination of road types and traffic conditions. Fig.22(b) shows the correlation coefficient for breathing waveform estimation under different road types and traffic conditions, it can be seen that the correlation coefficient is always larger than 0.93. Moreover, during peak time, the estimation error is slightly larger because drivers may perform more driving operations during heavy traffic, which brings interferences to *BreathListener*. For different road types, the estimation error on local roads is slightly larger because on local roads, vehicles may suffer from poor road conditions, such as bumpy roads, which brings extra interferences.

5 DISCUSSION

We further exploit breathing monitoring problem when there are multiple people in a vehicle. Considering the blockage of seats to acoustic signals, the breathing and movements of rear seats passengers can be ignored. Therefore, we only consider the case when driving with a co-pilot. According to Section.3.2.2, *BreathListener* decomposes ESD signals through EEMD and considers the component mainly within the breathing frequency range ($0.16Hz-0.6Hz$), as the breathing pattern of the driver in ESD signals.

Fig.23 shows the EEMD result when there is a co-pilot during diving, and the smartphone is placed on the operating panel. It can be seen that two candidate components, i.e., C_3 and C_4, fall into the breathing frequency range, which could related to the driver and the co-pilot. Considering the ESD amplitude of C_3 is much larger than C_4, and the smartphone is placed much closer to the driver than the co-pilot, C_3 is much likely related to the driver. However, since breathing movements of the driver and the co-pilot are close in frequency, there exists co-channel interference and adjacent-channel interference[33] in ESD signals. Therefore, when apply EEMD, the result components of the two breathing movements could interfere to each other so that they will looks similar in the trends, as C_3 and C_4 shown in Fig.23.

To extract the accurate breathing pattern for the driver, we need to reduce the interference of the co-pilot. We take advantage of the amplitude differences between the breathing-related EEMD component for driver and that of the co-pilot. According to the algorithm of EEMD, we can reduce the interferences in EEMD result by making the parameter e (standard deviation of white noise) too large for the breathing of the co-pilot in ESD signals, but still appropriate for the driver. Specifically, when *BreathListener* detects two candidate components C_A and C_B in the frequency range ($0.16Hz - 0.6Hz$), it first calculates their standard deviation as Std_{C_A} and Std_{C_B}. Then, *BreathListener* changes the parameter as $e = \sqrt{Std_{C_A} \times Std_{C_B}}$ to perform the EEMD again. Fig.24 shows

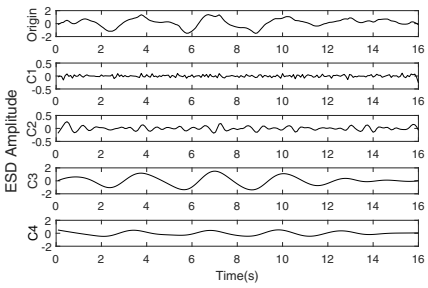


Figure 23: EEMD result of ESD signals when driving with a co-pilot.

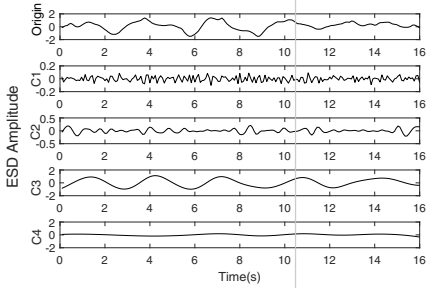


Figure 24: EEMD result after reduction of interference from the co-pilot.

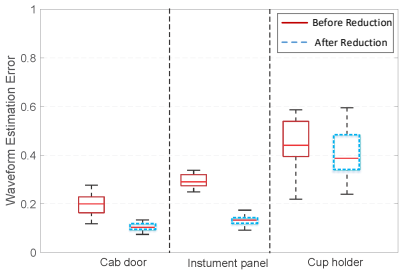


Figure 25: Waveform estimation error when driving with a co-pilot.

the EEMD result after reducing interference, there is only one component, i.e., C_3, related to the breathing.

It is also should be noticed that the interference of the co-pilot grows when the smartphone placed closer to the co-pilot. Fig.25 shows the breathing waveform estimation error for the driver when driving with a co-pilot, with the smartphone in different placements. It can be seen from Fig.25 that when the smartphone moves closer to the co-pilot (from cab door to the cup-holder), the waveform estimation error increases. Specially, when the smartphone is placed on the cup-holder, the error increases to nearly 0.5, showing that *BreathListener* can not distinguish breathing from the driver and the co-pilot when the smartphone is placed right in the middle of the driver and the co-pilot, which is the 'blind spot' of our system. However, in other smartphone placements, such as cab door and instrument panel, the breathing waveform estimation error is lower than 0.18 after interference reduction. The result shows that besides the 'blind spot', *BreathListener* is able to accurately capture the fine-grained breathing waveform of a driver even with a co-pilot in the vehicle.

6 RELATED WORK

In this section, we review the existing works related to *BreathListener*. There have been existing works on monitoring the vital signals of drivers. Among all kinds of vital signals, breathing pattern[11][43], cardiac pattern[23][37], and EEG pattern[29][18] are most commonly monitored and exploited. However, these solutions rely on pre-deployed infrastructure or additional hardware, which are intrusive and also incur a high cost.

To provide ubiquitous breathing estimation without contact, researchers have proposed solutions utilizing RF signals [4] [32] and WiFi signals[1][19][2]. For RF-based approaches, radar modules[32] and UWB radars[4] are leveraged to detect the chest motion for breathing estimation. For WiFi-based approaches, the breathing pattern is traced leveraging the received signal strength (RSS)[1], channel state information (CSI)[19] or phase information[2] of WiFi signals. However, the above approaches require additional RF and WiFi deployment, which is not convenient in driving environments. Although several types of smartphones are embedded with NFC model that support RF reading, currently the RF card on smartphones is not capable to serve as an active RF radar that is powerful enough to track the breathing pattern. Moreover, they also require relatively stationary state and relatively quiet environment.

More recently, a number of approaches that leveraging the built-in sensors of off-the-shelf smartphones for breathing estimation have been proposed[15] [3][34][26][27]. Jonathan et al.[15] estimate the breathing pattern by extract the PPG signal from the photos taken by cameras in a smartphone. Zephyr[3] utilizes the accelerometer and gyroscope of the smartphone to estimate the breathing rate. Ren et al.[34] monitor the breathing rate leveraging the breathing sounds when people are sleeping. These approaches can only estimation the breathing rate, but not the fine-grained breathing waveform. Nandakumar et al.[26][27] uses FMCW of acoustic signals to detect several breathing-related diseases. Although these works can track the breathing waveform of subjects, they do not consider separating breathing from other body movements of a subject, which makes them unsuitable in tracking breathing waveform of dynamic drivers in noisy driving environments.

Recent years have witnessed tremendous development for acoustic-based tracking technology. AAmouse[45] realizes a virtual mouse for VR/AR games leveraging acoustic devices of smartphones. CAT[21], LLAP[39], FingerIO[28] and Strata[44] track movements of a finger/hand with acoustic devices on smartphones, which can achieve sub-centimeter error. However, all of these acoustic-based approaches can only capture movements of the object nearest to acoustic devices. Since drivers constantly perform body movements during driving, such as steering, braking, etc., these approaches can not accurately capture the movements brought by breathing.

Unlike all of the existing works, *BreathListener* estimate the fine-grained breathing waveform from dynamic drivers in noisy driving environments leveraging only acoustic devices on smartphones.

7 CONCLUSIONS

In this paper, we address the problem of providing fine-grained breathing monitoring in driving environments. Particularly, we propose a fine-grained breathing monitoring system, *BreathListener*, which utilizes acoustic devices on smartphones to estimate fine-grained breathing waveform in driving environments. *BreathListener* first captures the breathing pattern of drivers leveraging the energy spectrum density (ESD) of acoustic signals, then filters the influence of environmental noises, and further extract the breathing pattern in ESD signals. After that, we propose a deep learning architecture based on Generative Adversarial Network (GAN) to generate fine-grained breathing waveform. Experiments in real driving environments validates the accuracy of *BreathListener* for fine-grained breathing waveform estimation.

ACKNOWLEDGMENT

This research was sponsored by NSFC (No.61772338).

REFERENCES

[1] Heba Abdelnasser, Khaled A Harras, and Moustafa Youssef. 2015. UbiBreathe: A ubiquitous non-invasive WiFi-based breathing estimator. In *Proc. ACM MobiHoc'15, Hangzhou, China.*

[2] Fadel Adib, Hongzi Mao, et al. 2015. Smart homes that monitor breathing and heart rate. In *Proc. ACM CHI'15, Seoul, Korea.*

[3] Heba Aly and Moustafa Youssef. 2016. Zephyr: Ubiquitous accurate multi-sensor fusion-based respiratory rate estimation using smartphones. In *Proc. IEEE INFOCOM'16, San Francisco, CA, USA.*

[4] Yifan Chen and Predrag Rapajic. 2008. Human respiration rate estimation using ultra-wideband distributed cognitive radar system. *International Journal of Automation and Computing* 5, 4 (2008), 325–333.

[5] M Elliott et al. 2012. Critical care: the eight vital signs of patient monitoring. *British Journal of Nursing* 21, 10 (2012), 621–625.

[6] Jin Fei and Ioannis Pavlidis. 2006. Analysis of breathing air flow patterns in thermal imaging. In *International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 946–952.

[7] Patrick Flandrin, Gabriel Rilling, and Paulo Goncalves. 2004. Empirical mode decomposition as a filter bank. *IEEE signal processing letters* 11, 2 (2004), 112–114.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. 2014. Generative adversarial nets. In *NIPS'14, Montreal, Canada.*

[9] John E Hall. 2010. *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proc. IEEE CVPR'16, Las Vegas, USA,.*

[11] Jennifer Healey and Rosalind Picard. 2000. SmartCar: detecting driver stress. In *Proc. IEEE ICPR'00, Barcelona, Spain.*

[12] David Hilbert. 1989. Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. In *Integralgleichungen und Gleichungen mit unendlich vielen Unbekannten*. Springer, 8–171.

[13] Norden E Huang et al. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences* 454, 1971 (1998), 903–995.

[14] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).

[15] E Jonathan and Martin Leahy. 2010. Investigating a smartphone imaging unit for photoplethysmography. *Physiological measurement* 31, 11 (2010), 70–79.

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *NIPS'12, Lake Tahoe, USA.*

[17] Christian Ledig, Lucas Theis, et al. 2016. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802* (2016).

[18] Wei Li, Qi-chang He, Xiu-min Fan, and Zhi-min Fei. 2012. Evaluation of driver fatigue on two channels of EEG data. *Neuroscience letters* 506, 2 (2012), 235–239.

[19] Jian Liu, Yan Wang, et al. 2015. Tracking vital signs during sleep leveraging off-the-shelf wifi. In *Proc. ACM MobiHoc'15, Hangzhou, China.*

[20] Andrew L Maas, Awni Y Hannun, et al. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML'13, Atlanta, USA.*

[21] Wenguang Mao, Jian He, and Lili Qiu. 2016. CAT: high-precision acoustic motion tracking. In *Proc. ACM MOBICOM'16, New York City, USA.*

[22] Wenguang Mao, Mei Wang, and Lili Qiu. 2018. AIM: Acoustic Imaging on a Mobile. In *Proc. ACM Mobisys'18, Munich, Germany.* 468–481.

[23] Toshiyuki Matsuda et al. 2008. ECG monitoring of a car driver using capacitively-coupled electrodes. In *Proc. IEEE EMBC'08, Vancouver, Canada.*

[24] Brian McKenzie and Melanie Rapino. 2011. *Commuting in the united states, 2009.* US Department of Commerce, Economics and Statistics Administration, US Census Bureau Washington, DC.

[25] J. P. Mortola. 2004. Breathing around the clock: an overview of the circadian pattern of respiration. *European Journal of Applied Physiology* 91, 2-3 (2004), 119–129.

[26] Rajalakshmi Nandakumar, Shyamnath Gollakota, et al. 2015. Contactless sleep apnea detection on smartphones. In *Proc. ACM Mobisys'15, Florence, Italy.*

[27] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Jacob E Sunshine. 2019. Opioid overdose detection using smartphones. *Science translational medicine* 11, 474 (2019), eaau8914.

[28] Rajalakshmi Nandakumar, Vikram Iyer, et al. 2016. Fingerio: Using active sonar for fine-grained finger tracking. In *Proc. ACM CHI'16, San Jose, USA.*

[29] G Pastor et al. 2006. Rear-view mirror use, driver alertness and road type: an empirical study using EEG measures. *Transportation research part F: traffic psychology and behaviour* 9, 4 (2006), 286–297.

[30] Neal Patwari, Lara Brewer, et al. 2014. Breathfinding: A wireless network that monitors and locates breathing in a home. *IEEE Journal of Selected Topics in Signal Processing* 8, 1 (2014), 30–42.

[31] Massimo Piccardi. 2004. Background subtraction techniques: a review. In *Proc. IEEE SMC'04, The Hague, Netherlands.*

[32] Tauhidur Rahman, Alexander T Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N Patel, Julie A Kientz, and Tanzeem Choudhury. 2015. Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar. In *Proc ACM Ubicomp'15, Osaka, Japan.*

[33] Theodore S Rappaport et al. 1996. *Wireless communications: principles and practice*. Vol. 2. prentice hall PTR New Jersey.

[34] Yanzhi Ren, Chen Wang, Jie Yang, and Yingying Chen. 2015. Fine-grained sleep monitoring: Hearing your breathing with smartphones. In *Proc. IEEE INFOCOM'15, Kowloon, Hong Kong.*

[35] Christopher G Scully, Jinseok Lee, Joseph Meyer, Alexander M Gorbach, Domhnull Granquist-Fraser, Yitzhak Mendelson, and Ki H Chon. 2012. Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Transactions on Biomedical Engineering* 59, 2 (2012), 303–306.

[36] NeuLog Logger Sensors. [n. d.]. NEULOG Respiration Monitor Logger Sensor. [Online]. Avaliable: http://www.neulog.com/.

[37] Hugo Silva, André Lourenço, and Ana Fred. 2012. In-vehicle driver recognition based on hand ECG signals. In *Proc. IUI'12.* ACM, 25–28.

[38] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[39] Wei Wang, Alex X Liu, et al. 2016. Device-free gesture tracking using acoustic signals. In *Proc. ACM MOBICOM'16, New York City, USA.*

[40] Yifan Wang, Federico Perazzi, Brian McWilliams, Alexander Sorkine-Hornung, Olga Sorkine-Hornung, and Christopher Schroers. 2018. A Fully Progressive Approach to Single-Image Super-Resolution. *arXiv preprint arXiv:1804.02900* (2018).

[41] Zhaohua Wu and Norden E Huang. 2004. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 460, 2046 (2004), 1597–1611.

[42] Zhaohua Wu and Norden E. Huang. 2005. Ensemble Empirical Mode DecompositionïijŽa Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis* 1, 01 (2005), 1–41.

[43] Guosheng Yang, Yingzi Lin, and Prabir Bhattacharya. 2010. A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Information Sciences* 180, 10 (2010), 1942–1954.

[44] Sangki Yun et al. 2017. Strata: Fine-Grained Acoustic-based Device-Free Tracking. In *Proc. ACM Mobisys'17, Niagara Falls, USA.*

[45] Sangki Yun, Yi-Chao Chen, and Lili Qiu. 2015. Turning a mobile device into a mouse in the air. In *Proc. ACM Mobisys'15, Florence, Italy.*