CONTACTLESS FATIGUE DETECTION USING COMMODITY SMART SPEAKER

Zhi Wang, Siheng Li, Fusang Zhang, Beihong Jin

State Key Laboratory of Computer Sciences, Institute of Software, Chinese Academy of Sciences, China
University of Chinese Academy of Sciences, China

ABSTRACT

Currently, fatigue has become one of the biggest health concerns. Existing fatigue detection methods either require dedicated intrusive devices (e.g., EEG, ECG) that suffer high costs or leverage video camera analyses that are light sensitive. In this paper, we present a contactless and low-cost fatigue detection method and develop the corresponding system, recognizing users fatigue state by the acoustic signal from a commodity smart speaker. In particular, leveraging the acoustic signal transmitted by the speaker, we first extract the users respiration and heartbeat waveform from the received reflected signal. Then we employ key features extracted from both respiration and heartbeat data to achieve effective fatigue detection. Fatigue and alertness data from ten participants are used to conduct extensive experiments. The experimental results show that our system can achieve the average detection accuracy of 91%, which is comparable to ECG or visual based fatigue detection systems.

Index Terms— Acoustic sensing, Fatigue detection, Contactless sensing, Vital sign monitoring

1. INTRODUCTION

Fatigue is a feeling of constant tiredness, which afflicts modern people and even affects human health. The complaint of fatigue is high in general population in a range of 18.3−27% [1]. The higher prevalence of fatigue has been reported in many scenarios that may further induce health and safety problems. For example, fatigue impacts on drivers in terms of their reaction time, awareness of hazards around them and their attentions. Drowsy drivers are three times more likely to be involved in a car crash, especially when they are awake over 20 hours [2]. According to statistics, about 20−30% of road accidents and 5−15% of all fatal road accidents involve in driver's fatigue [3]. Other studies [4] [5] [6] show that doctors and nurses with accumulation of fatigue during successive work shifts have a rise in accident rates and errors. Also, fatigue is a sign of some underlying illnesses, such as thyroid disorders, heart diseases or diabetes. Therefore, daily fatigue monitoring is of great significance for human health and work safety.

Existing fatigue detection methods either have to rely on biological signals from specialized electrocardiogram/ electroencephalogram (ECG/EEG) measurements [7][8] or facial vision analyses from video cameras [9][10]. Among them, the EEG based methods are considered to be the most reliable, as they directly measure neurophysiological activities in the human brain [11]. While the achieved accuracy is high, the equipment required is usually expensive and users are required to wear sensors operated by professionals. On the other hand, the camera-based methods capture the eye blink rate by the facial vision analyses to indicate the potential fatigue. However, these methods [9][10] not only bring in severe privacy concerns but also cannot be applied at night. Can we present a cheap contactless method by exploring commodity devices to monitor fatigue? We happen to find smart speakers (such as Amazon Echo and Google Home), which are popular in home environments, have shown the capability of performing many contactless sesnsing tasks [12][13].

In this paper, we infer human fatigue with an off-the-shelf smart speaker. We let the speaker transmit an acoustic Frequency-Modulated Continuous-Wave (FMCW) signal. The signal hits the target nearby, reflected by the chest of the target and then received by the microphone. We carefully analyze the received reflected signal and extract the respiration and heartbeat information of the target. Then we select a key subset from multiple respiration and heartbeat features and employ the appropriate machine-learning classifier to identify the fatigue state. The challenges come from two aspects. Firstly, the reflected signal contains the motions caused by both human breaths and heartbeats, however, the motion induced by heartbeats is much smaller (only about 0.1-0.5mm) than that induced by respiration (about 1-5mm). Therefore, the first challenge is how to distinguish between heartbeat and respiration motion and quantify them. Secondly, human fatigue is undoubtedly reflected in vital signs (e.g., respiration and heartbeat), however, nobody knows what key features in vital signs are. We propose a new fatigue detection method, addressing the above challenges. Moreover, we build the corresponding system [14]. We conduct extensive experiments on different persons. Experimental results show our system achieves the average accuracy of above 91%. Even for the newcomers whose data never occur in the training set of our classifier, our system can still detect their fatigue states with 85% accuracy.

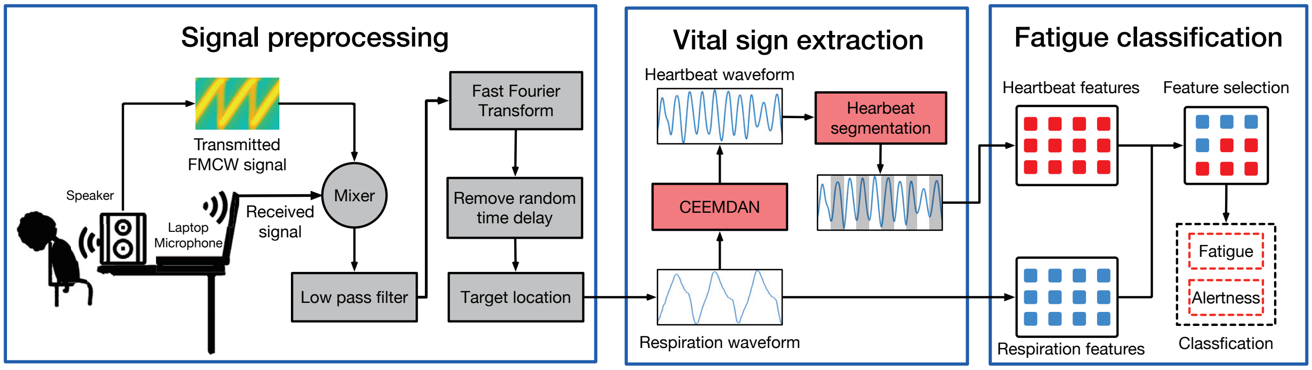


Fig. 1. Overview of fatigue detection system.

2. OUR SYSTEM

Our system can recognize human fatigue states by acoustic signals. As shown in Figure 1, the system mainly consists of three modules.

Signal preprocessing: From the mixed signal (i.e., the received reflected signal multiplied by the transmitted signal), we calculate the distance between target and transceiver device.

Vital sign extraction: From the phase change of the mixed signal at the corresponding distance, we obtain respiration data and then extract the tiny heartbeat waveform. Based on the heartbeat form, we further segment the heartbeat signal and estimate interbeat intervals (IBIs).

Fatigue classification: We select key features from both respiration and heartbeat data, and also choose the most appropriate classifier to recognize users fatigue state.

2.1. Signal Preprocessing

In our system, the speaker and the microphone are required to be placed in the same location, hereafter they are collectively called "devices". The transmitted signal $x_{tx}(t)$ is a FMCW signal whose frequency linearly increases over time and can be expressed as:

$$x_{tx}(t) = Acos(\phi(t)) = Acos(2\pi(f_0 t + \frac{kt^2}{2})) \quad (1)$$

where $k = \frac{B}{T}$ is the slope of frequency change, B is the bandwidth, T is the sweep time, f_0 is the starting frequency, A is the amplitude of transmitted signal.

Let R be the distance between the target and the device, and c be the signal propagation speed in the air, the signal arrives at the target and gets reflected to the receiver after time $\tau = \frac{2R}{c}$. Therefore, the received signal $x_{rx}(t)$ can be treated as the transmitted signal with time delay τ, which is represented as:

$$x_{rx}(t) = Dcos(2\pi(f_0(t-\tau) + \frac{k(t-\tau)^2}{2})) \quad (2)$$

where D is the amplitude of received reflected signal.

After the reflected signal is received, we perform mixing operation on the transmitted signal and the received signal to obtain mixed signal. The mixed signal is composed of a high-frequency component and a low-frequency component, where the low-frequency component indicates the frequency difference between $x_{tx}(t)$ and $x_{rx}(t)$. After the mixed signal is passed through a low-pass filter to remove the high-frequency component, the left only contains the low-frequency component, which can be represented as:

$$x_m(t) = \frac{AD}{2} cos\Big(2\pi(f_0\tau - \frac{k(\tau^2 - 2t\tau)}{2})\Big) \quad (3)$$

Since the frequency of mixed signal is caused by the delay τ, we can obtain τ by performing Fast Fourier Transform (FFT) on the mixed signal. However, in the real system, the audio signal is firstly put into a buffer before it is sent out, so the system have a random delay and the frequency of mixed signal is determined by the signal propagation time in the air and the system random delay. We calculate the system random delay through few received signals, and eliminate its affect by a corresponding delay compensation on the transmitted signal.

After eliminating the system random delay, we get the frequency of mixed signal $f_b = k\tau$, so the distance R between the target and the device can be calculated as $R = \frac{c \cdot f_b}{2k}$.

2.2. Vital Sign Extraction

After getting the relationship between distance R and frequency of mixed signal f_b, we can obtain the distance between target and device. Then we separate target's vital sign information (amplitude and phase) at a distance of R from the mixed signal. However, due to limitation of signal bandwidth, the amplitudes of respiration and heartbeat waveforms are not greater enough to change the frequency of the mixed signal. Therefore, we have to obtain the respiration waveform by the phase change of mixed signal.

Let Δd be the subtle movement distance of human chest, the mixed signal can also be represented as:

$$x_m(t) = \frac{AD}{2} cos(2\pi f_b t + \frac{4\pi f_0 \Delta d}{c}) \quad (4)$$

where the term $\frac{4\pi f_0 \Delta d}{c}$ is the phase change of mixed signal. Suppose that $\Delta d = 5mm$, $f_0 = 16kHz$, $c = 343m/s$, the corresponding phase change calculated by $\frac{4\pi f_0 \Delta d}{c}$ is $167.4°$. That is, chest displacement of $1 - 5mm$ will cause the signal phase change of $33.5°$ to $167.4°$. Thus, by extracting phase change of mixed signal, we can obtain respiration waveform. Next, we employ a simple peak detection algorithm to extract the beat-to-beat intervals of respiration waveform.

Further, in order to more accurately and comprehensively recognize users state, we detect the heartbeats and obtain heartbeat information.

Since the chest displacement caused by heartbeats is very small, it is submerged in the respiration waveform, what is worse, the respiration waveform also contains other tiny movements of the body. Therefore, we apply Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose the superimposed signal into the superposition of multiple intrinsic mode function (IMF) including respiration, heartbeat, tiny body movements, and other noise components.

Let $x(t)$ be the phase change we get from the mixed signal, $\widetilde{d_i}$ be i-th mode separated by CEEMDAN, $n_i(t)$ be the white noise with zero mean and unit variance, $E_k(\cdot)$ be the operator which produces the k-th mode obtained by EMD, β_i be signal-to-noise ratio coefficient. Thus, in the light of CEEMDAN, the phase change of heartbeat signal can be obtained as follows:
Step 1: For every $i = 1, ..., I$, decompose each $x_i(t) = x(t) + \beta_0 n_i(t)$ by EMD until obtaining its first mode. The first mode of CEEMDAN can be represented as $\widetilde{d_1} = \frac{1}{I}\sum_{i=1}^{I} E_1(x_i(t))$.
Step 2: At the first stage ($k = 1$), calculate the first residue: $r_1 = x(t) - \widetilde{d_1}$.
Step 3: Get $(k + 1)$-th mode by recurrence formula $\widetilde{d_{k+1}} = \frac{1}{I}\sum_{i=1}^{I} E_1(r_k + \beta_k E_k(n_i(t)))$ and $r_{k+1} = r_k - \widetilde{d_{k+1}}$, until the residue cannot be further decomposed by EMD or the max number of IMF is reached.
Step 4: Considering that the frequency of respiration waveform is [0.1-0.5Hz] and the frequency of heartbeat waveform is [0.8- 2Hz], we can obtain heartbeat waveform by differentiating frequency ranges.

For the resultant heartbeat waveform, we adopt EM algorithm to segment it. We deem each heartbeat is continuous in time domain and has the same morphology i.e., template. Thereby, the segmentation of the heartbeat waveform can be converted into an optimization problem for each heartbeat. Through iterative optimization of the heartbeat template and the heartbeat segmentations, we can obtain the optimal heartbeat segmentations. On the basis of heartbeat segmentations, we can easily obtain IBIs of heartbeats.

2.3. Fatigue Detection

We use the heartbeat IBI sequences along with the respiration interval sequences to recognize the user's fatigue state. The fatigue detection module includes feature extraction, feature selection and classification.
Feature Extraction: We extract 42 features in total from both heartbeat IBI sequences and respiration interval sequences, as listed in Table 1. The detailed descriptions of these features can be found from medical information and references [15] [16] [17].
Feature Selection: In the context of limited training data, using all 42 features is likely to lead to over-fitting, ultimately resulting in poor classification performance. Therefore, we use l_1-SVM [18] to select the most relevant features of user's fatigue state. l_1-SVM is a feature selection approach which selects the features best contributing to the classifier's performance for training a SVM model.
Classification: We divide user's fatigue state into two categories: alertness and fatigue. We train SVM to predict the user's state, and compare its performance with other eight different machine learning algorithms, e.g., KNN, Linear Discriminant Analysis (LDA). The performance of classification is evaluated in Section 3.3. As a result, we adopt SVM as the classifier.

We have implemented our system [14] using an off-the-shelf smart speaker (JBL Jembe, 6 Watt, 80 dB) which is connected to a laptop (MacBook Pro 2.6GHz with an Intel Core i7, 16 GB RAM) via the 3.5mm Audio Interface (AUX). The smart speaker is employed to transmit acoustic signals and the microphone built in the laptop is used to receive the signals. The transceivers are placed 60cm away from the user. We adopt $f_0 = 16kHz, B = 5kHz, T = 0.02s$ to generate acoustic FMCW signals. The laptop employs a 48kHz sampling rate and processes the signals in real time to monitor the target vital signs and determine his/her fatigue state.

Table 1. Features

Domain	Features
Time	Mean, Median, SDNN, pNN50, RMSSD, SDNNi, MeanRate, SdRate, HRVTi, SD1, SD2, SD2/SD1
Frequency	Welch PSD: LF/HF, peakLF, peakHF Burg PSD: LF/HF, peakLF, peakHF Lomb-Scargle PSD: LF/HF, peakLF, peakHF

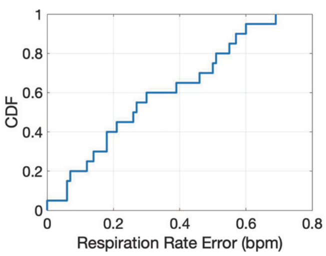
3. EVALUATION

3.1. Experiment Setup

Participants: We recruited 10 participants, aging from 21 to 33 years old. Each participant is monitored for 50 minutes, including 25-minute alertness data and 25-minute fatigue data. Thus, we have totally collected 500-minute signal data.

We adopt the 1-minute sliding window with a step of 30 seconds to separate these data into 900 samples.

Ground Truth: We employ a 3-lead ECG monitor, i.e., Heal Force PC-80B to get heart rates (HRs) and inter-beat intervals (IBIs), and employ a respiration monitor belt logger sensor NUL-236 [19] to get the respiration data. For fatigue detection, we collect alertness data in the morning and fatigue data after the lunch when people tend to feel tired. Each participant reports his/her fatigue state before and after data collection. These labels are used for classification.



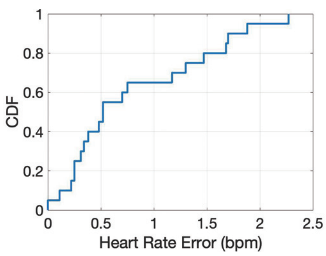


Fig. 2. Respiration Estimation **Fig. 3**. Heartbeat Estimation

3.2. Evaluation of Respiration and Heartbeat Extraction

We first evaluate the accuracy of our system in extracting respiration and heartbeat. Figure 2 shows the cumulative distribution function (CDF) of respiration rate error. The median error is 0.27 bpm and the largest error is 0.69 bpm. Figure 3 shows the cumulative distribution function (CDF) of heart rate error. It can be seen that the median error is 0.52 bpm. These results show that our system can achieve high accuracy in respiration and heartbeat extraction.

Table 2. Fatigue detection with user-dependent classifier

User	Accuracy	Precision	Recall	F1-score
1	96.67%	96.00%	97.78%	96.88%
2	98.89%	97.78%	100.00%	98.88%
3	83.63%	84.01%	83.74%	83.87%
4	82.22%	80.46%	89.06%	84.54%
5	85.74%	88.93%	76.67%	82.35%
6	96.67%	96.36%	98.00%	97.17%
7	91.11%	93.56%	88.83%	91.13%
8	81.11%	79.64%	84.61%	82.05%
9	98.89%	97.78%	100.00%	98.88%
10	**100.00%**	**100.00%**	**100.00%**	**100.00%**

3.3. Evaluation of Fatigue Detection

As mentioned earlier, we use l_1-SVM to select the most relevant features to fatigue detection. Among all the selected features, meanRate (i.e., mean heart/respiration rate measured from heartbeat/respiration interval sequences) has the highest weighted score for both heartbeat and respiration, suggesting it is the most relevant feature. It is reasonable as the heart/respiration rate is apparently slow when people feel fatigued or tend to fall asleep.

To evaluate the accuracy of fatigue detection, we train two types of classifiers, namely user-dependent classifier and user-independent classifier.

User-dependent Classifier: Each user-dependent classifier is trained and tested on the data from the same user. We have trained 10 user-dependent classifiers with 5-fold cross validation.

The performance of all the user-dependent classifiers is listed in Table 2. The mean accuracy for all the classifiers is 91.49% . The highest and lowest accuracy is 100% from user No.10 and 81.11% from user No.8, respectively.

User-independent Classifier: Each user-independent classifier is trained on 8 users and tested on the remaining 2 users. We employ 9 different classifiers, i.e., SVM, KNN, Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), Decision Tree (DT), Gradient Boost Decision Tree (GBDT), Adaboost (Ada) and Linear Discriminant Analysis (LDA), to compare their performance and find the best classifier.

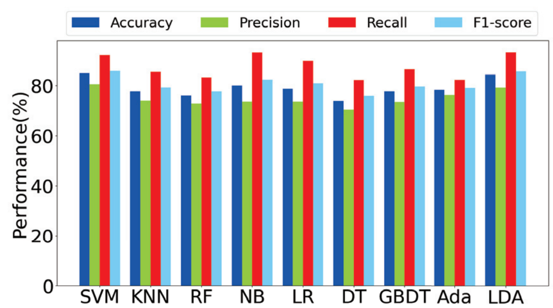


Fig. 4. Performance of user-independent classifiers

The performance of all the user-independent classifiers is listed in Figure 4. SVM has the highest accuracy (85%), the highest precision (80.58%), second highest recall (92.22%), and highest F1-score (86.01%). The results indicate that SVM is the best classifier for our fatigue detection system.

We find that the accuracy of video-based approach [20] and ECG-based approach [21] is 85.47% and 83.9%, respectively, which indicates that our system achieves comparable performance with the fatigue detection systems using ECG sensors and cameras.

4. CONCLUSION

In this work, we propose a contactless and cost effective fatigue detection system using a smart speaker. To the best of our knowledge, we are the first to explore a commodity smart speaker to demonstrate the feasibility of contactless fatigue detection using acoustic signal. As the next step, we will enhance the ability of adapting to different devices (speakers and microphones). Besides, we will provide our system to users with different occupations (e.g., drivers, doctors) and further improve the system in real applications.

5. REFERENCES

[1] Drew Dawson, Janine Chapman, and Matthew JW Thomas, "Fatigue-proofing: a new approach to reducing fatigue-related risk using the principles of error management," *Sleep medicine reviews*, vol. 16, no. 2, pp. 167–175, 2012.

[2] Ward Vanlaar, Herb Simpson, Dan Mayhew, and Robyn Robertson, "Fatigued and drowsy driving: A survey of attitudes, opinions and behaviors," *Journal of safety research*, vol. 39, no. 3, pp. 303–309, 2008.

[3] Ebrahim Razmpa, Khosro Sadegh Niat, and Babak Saedi, "Urban bus drivers sleep problems and crash accidents," *Indian Journal of Otolaryngology and Head & Neck Surgery*, vol. 63, no. 3, pp. 269–273, 2011.

[4] Alexa K Stuifbergen and Sharon Rogers, "The experience of fatigue and strategies of self-care among persons with multiple sclerosis," *Applied Nursing Research*, vol. 10, no. 1, pp. 2–10, 1997.

[5] Khosro Sadeghniiat Haghighi, Ali Montazeri, Ahmad Khajeh Mehrizi, Omid Aminian, Ania Rahimi Golkhandan, Maryam Saraei, and Mojtaba Sedaghat, "The epworth sleepiness scale: translation and validation study of the iranian version," *Sleep and Breathing*, vol. 17, no. 1, pp. 419–426, 2013.

[6] John A Caldwell, Melissa M Mallis, J Lynn Caldwell, Michel A Paul, James C Miller, and David F Neri, "Fatigue countermeasures in aviation," *Aviation, space, and environmental medicine*, vol. 80, no. 1, pp. 29–59, 2009.

[7] Rongrong Fu and Hong Wang, "Detection of driving fatigue by using noncontact emg and ecg signals measurement system," *International journal of neural systems*, vol. 24, no. 03, pp. 1450006, 2014.

[8] Suganiya Murugan, Jerritta Selvaraj, and Arun Sahayadhas, "Detection and analysis: driver state with electrocardiogram (ecg)," *Physical and Engineering Sciences in Medicine*, pp. 1–13, 2020.

[9] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 545–557, 2017.

[10] W. Sun, X. Zhang, S. Peeta, X. He, and Y. Li, "A real-time fatigue driving recognition method incorporating contextual features and two fusion levels," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 12, pp. 3408–3420, 2017.

[11] Mary E Harrington, "Neurobiological studies of fatigue," *Progress in neurobiology*, vol. 99, no. 2, pp. 93–105, 2012.

[12] Wei Wang, Lei Xie, and Xun Wang, "Tremor detection using smartphone-based acoustic sensing," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, UbiComp '17, pp. 309–312.

[13] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie, "Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, MobiCom '18, pp. 591–605.

[14] "Demo video," `https://youtu.be/QxIzOo_j1zY`.

[15] Jonghwa Kim and Elisabeth André, "Emotion recognition based on physiological changes in music listening," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.

[16] Peter Walter Kamen, Henry Krum, and Andrew Maxwell Tonkin, "Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans," *Clinical science*, vol. 91, no. 2, pp. 201–208, 1996.

[17] U Rajendra Acharya, K Paul Joseph, Natarajan Kannathal, Choo Min Lim, and Jasjit S Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[18] Ji Zhu, Saharon Rosset, Robert Tibshirani, and Trevor J Hastie, "1-norm support vector machines," in *Advances in neural information processing systems*, 2004, pp. 49–56.

[19] "Respiration monitor belt logger sensor nul-236," `https://neulog.com/respiration-monitor-belt/`, 2017.

[20] Guanglong Du, Tao Li, Chunquan Li, Peter X Liu, and Di Li, "Vision-based fatigue driving recognition method integrating heart rate and facial features," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[21] Jianmei Lei, Fangli Liu, Qingwen Han, Yunyang Tang, Lingqiu Zeng, Min Chen, Lei Ye, and Li Jin, "Study on driving fatigue evaluation system based on short time period ecg signal," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2466–2470.