Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

Short: A data-driven respirator fit test model via human speech signal

Jinmiao Chen [a], Zhaohe (John) Zhang [a], Shangqing Zhao [a,*], Song Fang [a], Thomas M. Peters [b], Evan L. Floyd [c], Changjie Cai [c,**]

[a] *University of Oklahoma (OU), 660 Parrington Oval, Norman, OK, 73019, USA*
[b] *University of Iowa, Iowa City, IA 52242, USA*
[c] *University of Oklahoma (OU) Health Sciences Center, 1100 N. Lindsay, Oklahoma City, OK 73104, USA*

ARTICLE INFO

MSC:
41A05
41A10

Keywords:
Respirator fit test
Speech signal
Semi-supervised learning

ABSTRACT

The need for personal protective equipment, such as respirators, has been emphasized by pandemics as they provide protection against infectious diseases. Adequate protection is only possible when respirators fit properly and are worn correctly. Therefore, it is especially critical to closely monitor and ensure respirator fit, particularly during a pandemic. To ensure proper fit and continuous monitoring, we propose a new noninvasive method that uses speech signals to measure the attenuation of sound caused by the respirator. This method provides a quantitative measure of respirator Fit Factor (FF, the ratio of the concentration of a substance in ambient air to its concentration inside the respirator). This method is also cost-effective and easy to implement. By collecting limited labeled and unlabeled speech data, augmenting labeled data, extracting time and frequency domain features, we achieved up to 86.24% accuracy in respirator fit detection using semi-supervised learning model.

1. Introduction

Close monitoring and ensuring the respirator fit become particularly important. For instance, the emergence and rapid spreading of the global Coronavirus Disease 2019 (COVID-19) pandemic has resulted in a significant demand for respirators, which provide an effective solution for protecting people from infectious diseases, particularly for healthcare personnel, who are at great risk of being exposed to the virus (Bartoszko, Malik Farooqi, Alhazzani, & Loeb, 2020). Respirators provide adequate protection only if they fit and are properly worn by the wearer (Ozog et al., 2020). However, due to the global shortage of respirators, the U.S. Centers for Disease Control and Prevention (CDC) suggests the potential reuse of disposable respirators to conserve available supplies (Implementing, 0000). Previous studies found that sterilizing masks might decrease the filtration efficiency of different masks (e.g., N95, KN95, and surgical masks) (Cai & Floyd, 2020), which negatively impacts the fit factor. In addition, respirators are critical in protecting workers from airborne hazards in many occupations, such as miners, construction workers, and doctors.

One issue with the current real-time fit assessment is its lack of availability in the healthcare facilities where it is most needed. The current real-time fit assessment requires a portable particle counter (e.g. TSI PortaCount Portacount, 0000), which is expensive and not wearable. Costs and size associated with the TSI PortaCount severely limit the number of devices available for real-time base risk mitigation decisions. The Condensation Nuclei Counter has also been widely used for fit test (Coffey et al., 2002), but its size and price limit the potential for the personal real-time assessment as well.

[*] Correspondence to: OU-Tulsa, 4502 E 41st St, Tulsa, OK 74135, USA.
[**] Correspondence to: OU-HSC, 801 N.E. 13th Street, Oklahoma City, OK 73104, USA.
E-mail addresses: shangqing@ou.edu (S. Zhao), changjie-cai@ouhsc.edu (C. Cai).

https://doi.org/10.1016/j.smhl.2023.100400
Received 23 March 2023; Accepted 24 March 2023
Available online 31 March 2023

Rather than leveraging the healthcare facilities, recently, extensive studies have been carried out to demonstrate the feasibility of data-driven methods on respirator detection, which feed machine learning classifiers with image or video signals to decide the presence of the respirator (Chandrika, 0000; Militante & Dionisio, 2020). However, these methods can induce privacy concerns since people may not want to be recorded on cameras. Although these image- or video-based classification models cannot be applied to the respirator fit assessment since they fail to characterize the material and penetrability of various respirators, which are significant factors to affect the fit factor, they highlight the need to raise awareness and develop learning-based solutions in the real-time respirator fit assessment.

It is well known that wearing a respirator could block the transmission of speech signals, thereby inevitably altering some acoustic properties (e.g., frequency and amplitude). Recent works (Cohn, Pycha, & Zellou, 2021; Magee et al., 2020; Saunders, Jackson, & Visram, 2021) have shown the variance of acoustic effects of different types of respirators on speech signals. Then, a common question can be raised: can we use the speech signal as the alternative media of the image inputs to a machine learning model for the fit assessment? In fact, recent studies have verified the presence of relationships between the distortion of the speech signal with types the respirators (Wittum, Feth, & Hoglund, 2013). For example, surgical masks and N95 respirators can attenuate higher-frequency sounds between 2000 to 7000 Hz by 3 dB to 12 dB (Goldin, Weinstein, & Shiman, 2020). Ryan et al. further demonstrated that such acoustic attenuation can affect the speech frequency above 1000 Hz and exhibits substantial variation between mask types (Corey, Jones, & Singer, 2020).

This paper focuses on developing a method that indirectly determines the fit factor (the ratio of the particle concentration in ambient air to its concentration inside the respirator) of respirators through easily accessible human speech signals. Specifically, we first collect our own dataset by inviting the human participants to our lab for the fit test and speech recording. Following that, we process the raw speech signal to extract time domain and frequency domain features for the model training. We quantify the fit factor into four levels (1–10, 10–100, 100–500, >500) to represent different protection degrees. We use these four levels as data label in our analysis. Finally, we feed learning models with the speech signal features after processing raw data and choose the model with the best accuracy for the fit factor prediction. Our well-trained model can be easily deployed into portable devices such that people can obtain the fit factor in real-time. Unlike traditional methods, our proposed technique is non-invasive (i.e., no physical body contact is required). Our main contributions can be summarized as follows.

- We are the first to investigate the relationship between the attenuation of acoustic signals and the respirator fit factor, and develop a data-driven method to learn the fit factor, instead of directly measuring it.
- Our method, which leverages easily accessible human speech signals to reverse engineer the respirator fit factor, is a non-invasive and lightweight system. It can be easily deployed into most portable devices with limited computing resources for the fit test. Compared with the traditional image-based learning model, our system gets rid of privacy concerns from traditional camera recording.
- To train semi-supervised learning models for predicting the fit factor label, we collect both labeled and unlabeled data. Based on experimental data, it is clear that our proposed method can reach an accuracy of up to 86.24% in a variety of settings.

To our best knowledge, there is no prior work focusing on developing the data-driven non-invasive architecture for real-time monitoring of individual respirator fit using smart low-cost sensors. The proposed framework can benefit all industry sectors where masking is essential (e.g., Public Safety, Healthcare and Social Assistance, Mining, Manufacturing, Construction, Agriculture, Forestry, and Fishing) by reducing occupational respiratory disease, and promoting safe and healthy work design and well-being.

2. Background and related work

2.1. Respirator fit test

The quantitative fit test aims to provide a numerical fit factor (FF) for evaluating how well the respirator can block the aerosol penetration (Persing, Sietsema, Farmer, & Peters, 2021). The fit test can be performed in a laboratory environment using a benchtop instrument (e.g., TSI PortaCount). During the test, the subject will wear the assigned respirator and perform the following eight test exercises to simulate workplace activities including normal breathing, deep breathing, turning the head side-to-side, grimacing, moving the head up and down, talking, and bending over at the waist (Standard, 0000; Zhuang et al., 2008). For each test exercise, it shall be performed for one minute, and a FF is calculated. Then the overall FF can be derived by taking the harmonic mean

$$\text{Overall Fit Factor} = \frac{\text{Number of exercises}}{1/ff_1 + 1/ff_2 + \cdots + 1/ff_8}, \tag{1}$$

where $ff_i \in [1, +\infty]$ is the FF for each exercise, which is the ratio of ambient particles to particles in respirators.

Existing studies have contributed significantly to the design of fit tests in order to advance their performance (Bergman et al., 2014; Lin & Chen, 2017; Zhuang, Bradtmiller, & Shaffer, 2007; Zhuang, Coffey, & Ann, 2005). For example, the authors in Bergman et al. (2014) developed an advanced respirator fit-test headform. The benchtop healthcare instruments confine the fit test into the laboratory environment. Recent research has contributed efforts to designing low-cost wearable devices which can measure FF when the respirator is in use (Persing et al., 2021). But these methods still cannot remove the requirement of measuring FF through particular devices. Our research focuses on investigating the relationship between the FF and the distortion of the speech signals, which can be easily fetched through our daily used devices such as the cellphone.

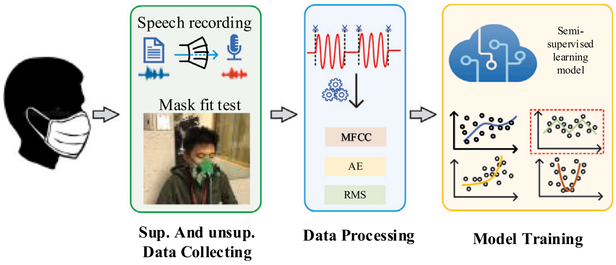
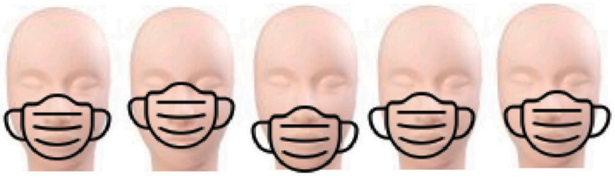


Fig. 1. System architecture of learning the FF.



(a) Perfect (b) Higher (c) Lower (d) Left (e) Right

Fig. 2. Different styles of wearing a respirator.

2.2. Existing data-driven methods

With plenty of data on acoustic voice measures in recording of people producing vocal tasks with and without wearing a mask, the distortion effect of the mask on the voice signal can be explored. Also, different types of masks (such as surgical masks, N95 and KN95 respirators, and cloth masks) may have different high-frequency effects, altering the directivity of speech (Corey et al., 2020). There are tons of research efforts studying the impact of face masks on speech perception (e.g., Corey et al. (2020), Nguyen et al. (2021) and Rahne, Fröhlich, Plontke, and Wagner (2021)), since the speech signal with frequencies above 1 kHz often contains crucial information for comprehension.

However, existing research focuses on the overall impact of masks on voice signals such as speech quality degradation, while our proposed research focuses on impacts of time domain and frequency domain. Features of a signal in the time domain and the frequency domain may account for the vast majority of information content in a speech signal.

3. Speech signal based respirator fit test model

In this section, we present our proposed data-driven model to reverse engineer the FF via speech signals. Our system is divided into three phases: labeled and unlabeled data collecting, data processing, and model training, as shown in Fig. 1. In the first step, we invite human participants to our lab to collect the dataset used for training models. Then we process the raw speech signal to extract features for the model training. Finally, we feed learning models with features, and choose the best model for real-world use.

3.1. Data collection

To build the connection between the FF and the recorded speech signal while wearing the respirator, we conduct a human study to collect our own datasets. We first prepare several types of respirator, then for each respirator, we invite human participants to wear it for the fit test and store the FF, serving as the ground truth label of the learning model. In addition to the normal procedure, we place an extra microphone one meter away in front of the participant to record the speech signal, and the participant is required to read a pre-selected sentence from the Harvard Sentences (Fang et al., 2018; Rothauser, 1969; Yang et al., 2022).

In order to balance the distribution of the collected FF, for each respirator, we require each participant to wear it in five styles (as shown in Fig. 2): (a) wearing it perfectly, (b) wearing it a little higher such that the bottom of the mouse cannot be covered perfectly, (c) wearing it a little lower such that the nose cannot be covered perfectly, (d) wearing it a little left-handed such that the right-hand side of the mouse and nose cannot be covered perfectly, and (e) wearing it a little right-handed such that the left-hand side of the mouse and nose cannot be covered perfectly. As we have a limited number of participants and FF measurement is cumbersome, in addition to the labeled data, we also conduct the data augmentation by collecting a large amount of unlabeled data, that only records the speech signal when wearing different respirators without the ground truth FF label. Denoted by $C_{labeled}$ and $C_{unlabeled}$ the sets of the labeled speech signal and unlabeled speech signal respectively. Our study involved human participants for the respirator fit test and voice recording. The full protocol has been reviewed and meets the criteria for exemption from our Institutional Review Board (IRB) review. Our IRB has determined that the study involves minimal risk for human participants (i.e., the risk is no more than the one that they face during their daily lives). We follow the approved protocol to inform them of the full study procedure and protect their identities without publishing any personally identifiable information.

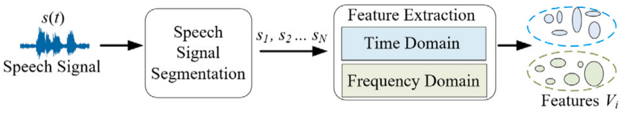


Fig. 3. The flowchart of the speech signal processing.

3.2. Data processing

3.2.1. FF map to label

The FF serves as the label of the speech signals. Our data-driven method is designed to be lightweight for the fit for industry applications. We quantize the FF into four levels, including level 1 (fail), level 2 (weak protective), level 3 (protective), and level 4 (strong protective).

$$l = \begin{cases} \text{Fail}, & \text{if Overall FF} \leq 10 \\ \text{weak}, & \text{if } 10 < \text{Overall FF} \leq 100 \\ \text{protective}, & \text{if } 100 < \text{Overall FF} \leq 500 \\ \text{strong}, & \text{if Overall FF} > 500, \end{cases} \tag{2}$$

where 10,100,500 are three widely-used thresholds used for distinguishing whether the respirator can provide sufficient protection. For example, a FF of 100 is typically used in the medical area of passing the fit test. The FF of 10, 100 and 500 equates to preventing 90%, 99% and 99.8% of particles from entering the mask, respectively.

3.2.2. Speech signal processing

As shown in Fig. 3, the speech signal process consists of two steps: speech signal segmentation and feature extraction, the first step divides the raw speech signal into units, and the second step extracts features from each speech unit.

Speech Signal Segmentation: We first split a raw digital speech signal $s(t)$ into short speech pieces s_i for $i \in \{0, 1, 2, \dots, N\}$, where N is the total number of short speech signal for each recorded speech signal. As the fit factor has little change during a short time, such splitting could make sure stable data distribution.

Feature Extraction: According to existing studies (Duan et al., 2022) in audio engineering, widely-used audio features can be classified into two categories: frequency domain features (i.e., MFCC) and time domain features (i.e., Amplitude Envelop). Our model makes an effort to acquire a feature set V_i for each piece that incorporates features from the frequency domain as well as features from the time domain.

Our model attempts to obtain a feature set V_i for each piece that includes features from both the frequency and time domains.

(i) Frequency Domain. Recent studies have shown that the transmission attenuation of the speech signal is very sensitive with respect to the frequency, for example, N95 respirators can attenuate higher-frequency sounds between 2000 to 7000 Hz by 3 dB to 12 dB (Goldin et al., 2020). Therefore, we can extract the Mel-frequency cepstral coefficients (MFCCs) v_{mf}, which have been widely used to capture the feature of speech (Davis & Mermelstein, 1980; Zhang, Yang, & Fang, 2021). Typically, the MFCC features can be obtained by the following steps: windowing, discrete Fourier transformer (DFT), Mel filter bank, log and Inverse DFT (Muaidi, Al-Ahmad, Khdoor, Alqrainy, & Alkoffash, 2014). The size of the MFCC feature is determined by the filter banks. Higher MFCC means more filter banks, resulting in more feature columns. However, we want to keep the feature dimension at a reasonable number, so we only send the most significant features to the classifiers. We set the number of MFCC to 20 because this number is the default in the HTK MFCC toolkit (Wojcicki, 0000).

(ii) Time Domain. A time-domain speech signal shows the amplitude at each sampling point of speech signal. This can be used to quantify the loudness or energy attenuation when the speech signal penetrates respirators. In the time domain, we extract amplitude envelope (AE) v_{ae} (McFee et al., 2015) and root-mean-square (RMS) energy v_{rms} (Rabiner & Schafer, 2007) features. The AE refers to fluctuations of a sound's amplitude over time, which are primarily dominated by a speech signal's energy, thus allowing us to evaluate the energy fluctuation when piercing a respirator. However, AE is more sensitive to outliers, therefore we also introduce the RMS energy features, which is the averaged energy over a period of time. It has been shown that RMS energy is more resilient when the surrounding environment is noisy (Rabiner & Schafer, 2007). Fig. 4 shows an example of the AE and RMS energy of a speech signal.

3.3. Model training

Combining features from both the time and frequency domains, we have $V_i = \{v_{mf}, v_{ae}, v_{rms}\}$. Our dataset contains both labeled and unlabeled data, therefore, we build a semi-supervised learning model based on unsupervised augmentation data (UDA) model (Xie, Dai, Hovy, Luong, & Le, 2020). Fig. 5 shows the overall architecture of the training model, which consists of two domains, i.e., supervised domain and unsupervised domain, designated for processing the labeled and unlabeled data respectively.

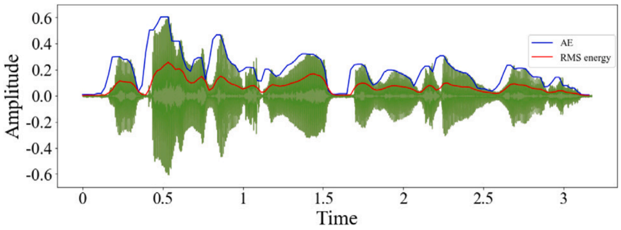


Fig. 4. Example of Root-mean-square energy.

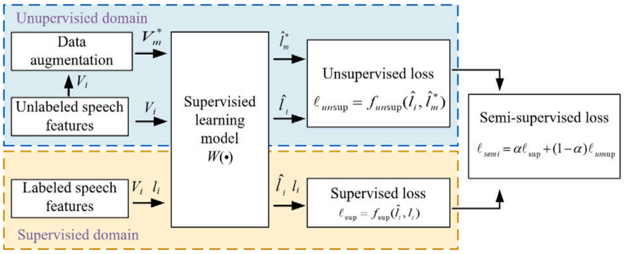

Unupervisied domain
Data augmentation
V_m^*
V_i
Unlabeled speech features
V_i
Supervisied learning model $W(\bullet)$
$\hat{l}_m^*$
$\hat{l}_i$
Unsupervised loss
$\ell_{unsup} = f_{unsup}(\hat{l}_i, \hat{l}_m^*)$
Semi-supervised loss
$\ell_{semi} = \alpha \ell_{sup} + (1-\alpha)\ell_{unsup}$
Labeled speech features
V_i l_i
$\hat{l}_i$ l_i
Supervised loss
$\ell_{sup} = f_{sup}(\hat{l}_i, l_i)$
Supervisied domain

Fig. 5. Architecture of the semi-supervised training model.

$$\ell_{sup} = f_{sup}(l_i, \hat{l}_i), \tag{3}$$

where $\hat{l}_i = W(V_i)$, and $f(\cdot)$ represents a loss function (e.g. cross entropy). $\mathcal{W}$ is a set of supervised learning models.

In the unsupervised domain, we first feed the unlabeled data V_i, obtained from the unlabeled dataset $C_{unlabeled}$, to the data augmentation modular. The unsupervised data will be augmented by adding negligible noise, generating V_m^*, for $m \in 1, 2, \ldots, \beta N$, where β is the augmentation ratio. Then, both V_i and V_m^* will be labeled according the supervised learning model W, and obtain the loss function as

$$\ell_{unsup} = f_{unsup}(\hat{l}_i, \hat{l}_m^*), \tag{4}$$

where $\hat{l}_m^* = W(V_m^*)$. Finally, the full loss function can be written as $\ell_{semi} = \alpha\ell_{sup} + (1-\alpha)\ell_{unsup}$, where α is a scalar to balance the significance of both domains. The objective function of the model training is aiming to find the best learning model $\bar{W}$ and $\bar{\alpha}$ such that minimizes the loss function

$$\{\bar{W}, \bar{\alpha}\} = \underset{W,\,\alpha}{\arg\min}\ \ell_{semi}. \tag{5}$$

4. Experiments and results

In this section, we first describe the experimental settings, and then demonstrate the performance of our proposed method.

4.1. Experimental setup

Our recorded speech dataset contains 177 labeled raw audio samples with ground truth fit factors and 200 unlabeled audio samples from 6 different people where three of them are male and three of them are female with ages spanning from 22 to 35. Our fit test and speech recording are conducted under the conditions of wearing four types of respirators including series 1512 (no valve, Moldex, USA), 2300 (with valve, Moldex, USA), SH3500 (no valve, Uniair, China) and 8822 (no valve, 3M, USA), which are all CDC NIOSH approved N95 respirators. We select 4 sentences from Harvard Sentences for speech signal sampling.

Our training model set $\mathcal{W}$ includes deep neural network (DNN) and convolutional neural network (CNN). For the DNN model, we build a four-layer neural network. The first layer is the input layer. The following are the two hidden layers, and the last layer is the output layer with the size of four for “fail”, “weak protective”, “protective”, and “strong protective” as all possible classifications. The CNN is designed as one input layer, three convolutional layers, two maxpooling layers, one fully connected layer, and one output layer. We set convolution kernel size is 3 × 3, pooling kernel size as 2 × 2. The first three layers have 6 feature maps. The next two layers have 20 feature maps. We configure our training process using adam optimizer, and we choose cross-entropy as the loss function for both f_{sup}, f_{unsup} for the classification tasks. Unless otherwise specified, by default, we set $\alpha = 0.4$ and $\beta = 20$.

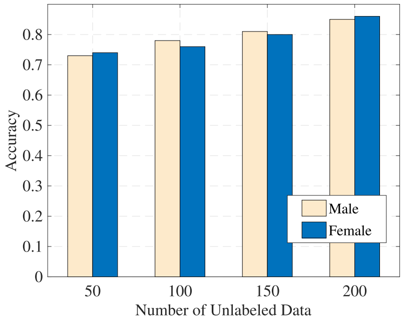


Fig. 6. Accuracy v.s. number of unlabeled data.

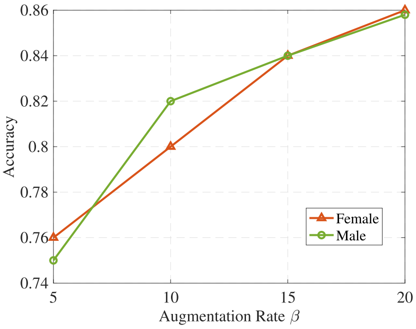


Fig. 7. Accuracy v.s. augmentation rate β.

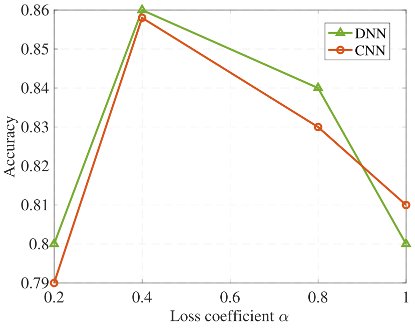


Fig. 8. Accuracy with respect to different α.

4.2. Results and analysis

Our data contain both the labeled and unlabeled data for the model training. We first evaluate impact from different number of unsupervised data on the classification accuracy. Fig. 6 shows the accuracy when we increase the number of unsupervised data from 50 to 200. In this experiment, the supervised learning model W is DNN. We can observe that the accuracy continuously increases as we feed more unsupervised data into our learning model. It indicates that unlabeled data add additional information for classifying model. In addition, we observe that there is no evident difference in the performance of the speech signal between females and males. For example, the accuracy is 86.21% for female's speech and 85.88% for the speech of males when number of unlabeled data is 200.

The unsupervised learning contains a data augmentation modular based on a augmentation rate β. Fig. 7 shows how the augmentation rate affects the final classification accuracy. It can be observed that as we increase the augmentation rate from 5 to 20, the accuracy monotonously increase for both female and male, and can achieve 86.21% for female and 85.88% for male. It indicates that the data augmentation modular is effective for providing more data for the training purpose.

We also evaluate the impact from the loss coefficient α that is used to balance the significance between the supervised and unsupervised data. Fig. 8 shows the relationship between α and the final classification accuracy. When $\alpha = 0.4$, we see the most

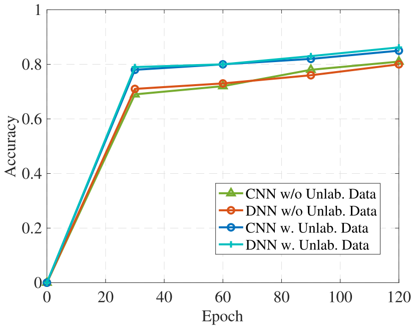


Fig. 9. Accuracy with and without unlabeled data.

improvement in performance. Both labeled and unlabeled data might achieve a balance in this manner, which would result in improved performance.

Fig. 9 compares the results between using and without using the unlabeled data for both DNN and CNN. It clearly shows that the accuracy is better when using the semi-supervised learning. For example, the accuracy with supervised learning achieves 86.24% at the end of the epochs for DNN, which is significantly better than that without using supervised data, i.e., 81.11%.

5. Conclusion

Due to the global COVID-19 pandemic, close monitoring and ensuring the respirator fit become particularly crucial. In this paper, we develop a data-driven real-time respirator fit test system using a non-invasive acoustic sensor, which is commonly available on a mobile phone. We first collect our own labeled data and unlabeled data to extract time domain and frequency domain features. Next, we train supervised and semi-supervised models in a set with these features. Finally, we choose the optimal one with the highest accuracy for the FF label prediction. We also attempt to use other environmental factors such as pressure and motion to infer the FF under the condition that the speech signal is unavailable. Our experimental results show that we can achieve detection accuracy of respirator up to 86.24% under different circumstances. Our work demonstrates the feasibility of real-time respirator fit testing to protect people in various environments.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The work was supported in part by NSF 1948547, 2155181, and CDC 75D30121C10674.

References

Bartoszko, Jessica J., Malik Farooqi, Mohammed Abdul, Alhazzani, Waleed, & Loeb, Mark (2020). Medical masks vs n95 respirators for preventing covid-19 in healthcare workers: A systematic review and meta-analysis of randomized trials. *Influenza Other Respiratory Viruses*, *14*(4), 365–373.

Bergman, Michael S., Zhuang, Ziqing, Hanson, David, Heimbuch, Brian K., McDonald, Michael J., Palmiero, Andrew J., et al. (2014). Development of an advanced respirator fit-test headform. *Journal of Occupational and Environmental Hygiene*, *11*(2), 117–125.

Cai, Changjie, & Floyd, Evan L. (2020). Effects of sterilization with hydrogen peroxide and chlorine dioxide solution on the filtration efficiency of n95, kn95, and surgical face masks. *JAMA Network Open*, *3*(6), Article e2012099.

Chandrika, Deb Face mask detection. https://github.com/chandrikadeb7/Face-Mask-Detection.

Coffey, P. J., Girman, S., Wang, S. M., Hetherington, L., Keegan, D. J., Adamson, P., et al. (2002). Long-term preservation of cortically dependent visual function in rcs rats by transplantation. *Nature Neuroscience*, *5*(1), 53–56.

Cohn, Michelle, Pycha, Anne, & Zellou, Georgia (2021). Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech. *Cognition*, *210*, Article 104570.

Corey, Uriah, Jones, Ryan M., & Singer, Andrew C. (2020). Acoustic effects of medical, cloth, and transparent face masks on speech signals. *The Journal of the Acoustical Society of America*, *148*(4), 2371–2375.

Davis, Steven, & Mermelstein, Paul (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *28*(4), 357–366.

Duan, Rui, Qu, Zhe, Zhao, Shangqing, Ding, Leah, Liu, Yao, & Lu, Zhuo (2022). Perception-aware attack: Creating adversarial music via reverse-engineering human perception. In *Proceedings of the 2022 ACM SIGSAC conference on computer and communications security* (pp. 905–919). New York, NY, USA.

Fang, Song, Markwood, Ian, Liu, Yao, Zhao, Shangqing, Lu, Zhuo, & Zhu, Haojin (2018). No training hurdles: Fast training-agnostic attacks to infer your typing. In *Proceedings of the 2018 ACM SIGSAC conference on computer and communications security* (pp. 1747–1760). New York, NY, USA.

Goldin, A., Weinstein, B. E., & Shiman, N. (2020). How do medical masks degrade speech perception?

Implementing filtering facepiece respirator (ffr) reuse, including reuse after decontamination, when there are known shortages of n95 respirators. https://www.cdc.gov/coronavirus/2019-ncov/hcp/ppe-strategy/decontamination-reuse-respirators.html.

Lin, Yi-Chun, & Chen, Chen-Peng (2017). Characterization of small-to-medium head-and-face dimensions for developing respirator fit test panels and evaluating fit of filtering facepiece respirators with different faceseal design. *PLoS One*, *12*(11), Article e0188638.

Magee, Michelle, Lewis, Courtney, Noffs, Gustavo, Reece, Hannah, Chan, Jess C. S., Zaga, Charissa J., et al. (2020). Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols. *The Journal of the Acoustical Society of America*, *148*(6), 3562–3568.

McFee, Brian, Raffel, Colin, Liang, Dawen, Ellis, Daniel P. W., McVicar, Matt, Battenberg, Eric, et al. (2015). librosa: Audio and music signal analysis in python.

Militante, S. V., & Dionisio, N. V. (2020). Deep learning implementation of facemask and physical distancing detection with alarm systems. In *2020 third intl. conference on vocational education and electrical engineering* (pp. 1–5).

Muaidi, Hasan, Al-Ahmad, Ayat, Khdoor, Thaer, Alqrainy, Shihadeh, & Alkoffash, Mahmud (2014). Arabic audio news retrieval system using dependent speaker mode, mel frequency cepstral coefficient and dynamic time warping techniques. *Research Journal of Applied Sciences, Engineering and Technology*, *7*(24), 5082–5097.

Nguyen, Duy Duong, McCabe, Patricia, Thomas, Donna, Purcell, Alison, Doble, Maree, Novakovic, Daniel, et al. (2021). Acoustic voice characteristics with and without wearing a facemask. *Scientific Reports*, *11*(1), 1–11.

Ozog, David M., Sexton, Jonathan Z., Narla, Shanthi, Pretto-Kernahan, Carla D., Mirabelli, Carmen, Lim, Henry W., et al. (2020). The effect of ultraviolet c radiation against different n95 respirators inoculated with sars-cov-2. *International Journal of Infectious Diseases*, *100*, 224–229.

Persing, Allison J., Sietsema, Margaret, Farmer, K. R., & Peters, Thomas M. (2021). Comparing respirator laboratory protection factors measured with novel personal instruments to those from the portacount. *Journal of Occupational and Environmental Hygiene*, *18*(2), 65–71.

Portacount® respirator fit tester 8038. https://tsi.com/products/respirator-fit-testers/portacount%C2%AE-respirator-fit-tester-8038/.

Rabiner, Lawrence R., & Schafer, Ronald W. 2007.

Rahne, Torsten, Fröhlich, Laura, Plontke, Stefan, & Wagner, Luise (2021). Influence of surgical and n95 face masks on speech perception and listening effort in noise. *PLoS One*, *16*(7), Article e0253874.

Rothauser, E. H. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*, 225–246.

Saunders, Gabrielle H., Jackson, Iain R., & Visram, Anisa S. (2021). Impacts of face coverings on communication: an indirect impact of covid-19. *International Journal of Audiology*, *60*(7), 495–506.

Standard interpretations. https://www.osha.gov/laws-regs/standardinterpretations/publicationdate/2004.

Wittum, Lawrence, Feth, Kelsi J., & Hoglund, Evelyn (2013). The effects of surgical masks on speech perception in noise. *Proceedings of Meetings on Acoustics*, *19*(1).

Wojcicki, Kamil Htk mfcc matlab - functions.

Xie, Qizhe, Dai, Zihang, Hovy, Eduard, Luong, Thang, & Le, Quoc V. (2020). Unsupervised data augmentation for consistency training. In *NeuRIPS*.

Yang, Edwin, Fang, Song, Markwood, Ian, Liu, Yao, Zhao, Shangqing, Lu, Zhuo, et al. (2022). Wireless training-free keystroke inference attack and defense. *IEEE/ACM Transactions on Networking*, *30*(4), 1733–1748.

Zhang, Zhaohe, Yang, Edwin, & Fang, Song (2021). Commandergabble: A universal attack against asr systems leveraging fast speech. In *Annual computer security applications conference* (pp. 720–731).

Zhuang, Ziqing, Bradtmiller, Bruce, & Shaffer, Ronald E. (2007). New respirator fit test panels representing the current us civilian work force. *Journal of Occupational and Environmental Hygiene*, *4*(9), 647–659.

Zhuang, Ziqing, Coffey, Christopher C., & Ann, Roland Berry (2005). The effect of subject characteristics and respirator features on respirator fit. *Journal of Occupational and Environmental Hygiene*, *2*(12), 641–649.

Zhuang, Ziqing, Groce, Dennis, Ahlers, Heinz W., Iskander, Wafik, Landsittel, Douglas, Guffey, Steve, et al. (2008). Correlation between respirator fit and respirator fit test panel cells by respirator size. *Journal of Occupational and Environmental Hygiene*, *5*(10), 617–628.