

Optimization of Ultrasonic Respiratory Signals based on Supervised Learning

Jiawen Chen¹, Ziyue Dang², Lingkun Li³, Fan Dang¹

¹ Global Innovation Exchange, Tsinghua University, Beijing, China

² Department of Computer Science and Technology, Tsinghua University, Beijing, China

³ School of Software, Beijing Jiaotong University, Beijing, China

cjw22@mails.tsinghua.edu.cn, dzy19@mails.tsinghua.edu.cn, lkli@bjtu.edu.cn, dangfan@tsinghua.edu.cn

Abstract—There are various methods to monitor human respiration. Traditional methods of monitoring the human respiratory process often rely on complex medical equipment, which makes it difficult for users to operate. Nowadays, more and more researchers are focusing on smartphone-based systems that use mobile phones to transmit ultrasound to the chest and abdomen of the human body and use the unique reverse echo of ultrasound to collect respiratory signals. However, this method is easily disturbed by the environment, clothing, equipment, and other factors. Thus, the accuracy is unsatisfactory. This paper presents a method to optimize the respiratory signals collected by ultrasound. This method is based on supervised learning. Piezoelectric sensors and mobile phones are used to monitor human respiratory signals. A Long-Short Term Memory (LSTM) is established to learn the expression from ultrasonic signals to piezoelectric signals to improve the accuracy of signal acquisition. The results show that the model has good performance in both the time and frequency domains, achieving less than 0.05 mean absolute error (MAE) and 0.8779 intersections over union (IoU). The model can be used to optimize the ultrasound respiratory signals.

Index Terms—ultrasonic perception; piezoelectric sensing; supervised learning; signal optimization

I. INTRODUCTION

With the improvement of people's life quality and the development of wearable devices, the measurement and monitoring of individual vital signs have become a topic of great interest. Respiration is an essential indicator among them. As one of the most important vital signs in humans, respiratory monitoring is essential for disease detection and prevention. Respiratory monitoring can detect many chronic diseases, including asthma and chronic obstructive pulmonary disease [1]. Monitoring respiration can also provide insight into a user's sleep and emotional state. The elderly commonly experience abnormal respiratory events, such as obstructive or central sleep apnea [2]. These breathing disorders can reduce sleep quality and even become life-threatening.

Traditional breath monitoring requires users to wear cumbersome and costly equipment on their bodies.

In recent years, some researchers have proposed using acoustic data to monitor human respiration; the system can detect the sounds of cough, sneezing, and sniffing. However, this method is easily disturbed by the surrounding environment, which limits the scale of current acoustic-data-based respiration monitor systems [3], [4].

Other researchers proposed using mobile phones to transmit ultrasonic waves to monitor humans' respiratory signals. This method's advantage is that it is simple to generate transmitted signals, and individuals can carry them with them at all times, making it suitable for long-term monitoring in daily life [1], [2], [5]. In addition, the ultrasonic wave is a kind of mechanical wave. Therefore, users do not need to worry about long-term electromagnetic radiation monitoring.

Inhale and exhale are the two steps of the breathing cycle. Air enters the lungs through the mouth or nose during inhalation, causing the lungs to expand. This will cause the expansion of the chest wall. Simultaneously, the bottom diaphragm and the upper abdomen will contract. Inhalation is the opposite of exhalation. Air will exit the human body via the mouth or nose. As the air leaves the human body, the chest wall will contract, and the diaphragm will relax, resulting in a reduction in the volume of the chest and abdominal cavities [6]. Respiratory chest movement alters the distance between the speaker and the microphone, thereby altering the phase of the reflected signal. After the reflected signal has passed through the high pass filter and been demodulated, it is possible to obtain human respiratory motion.

However, the ultrasonic reflection is susceptible to significant errors, and three factors primarily cause the analysis noise:

- Due to the non-stationary and non-linear nature of the ultrasonic signal, its echo signal contains a great deal of helpful information and numerous abrupt components.
- Ultrasonic waves are reflected on the surface of human clothing during the sampling process. Various textile materials will scatter ultrasonic waves to varying degrees in this process, resulting in microstructure noise.
- The mobile phone will emit dispersive sounds similar to white noise.

In conclusion, it is necessary to consider additional processing of the signal sampled by the system to increase its reliability. Existing de-noising techniques include, among others, wavelet threshold de-noising, empirical mode decomposition (EMD) threshold de-noising, joint wavelet, and EMD threshold de-noising. However, these methods often result in signal distortion during reconstruction. It is more difficult to determine the threshold without prior knowledge of respiratory

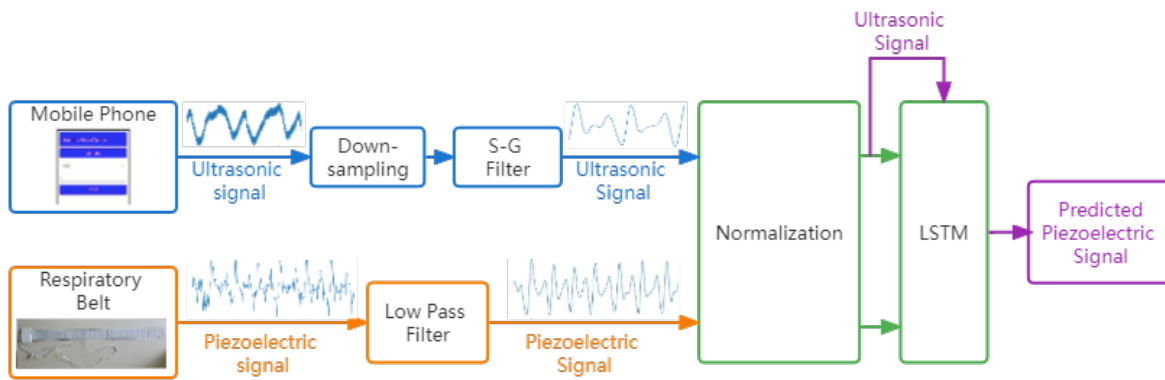


Fig. 1. The system design

signals, so existing noise reduction methods cannot be used.

This paper proposes a method for optimizing the input signal based on supervised learning. Concurrently, the piezoelectric sensor and the ultrasonic method of the mobile phone are used to collect the respiratory signals of users in the same state. It is believed that the data collected by piezoelectric sensors is more precise than by ultrasonic transducers. Then, the supervised learning method is used to input the respiratory signal recorded by ultrasound and output the respiratory signal recorded by the piezoelectric sensor to establish a mapping to learn how to convert the imprecise ultrasonic signal into a more accurate piezoelectric signal. Then, use the piezoelectric signal as the input of our system to refine the input signal.

Our significant contributions can be summarized as follows:

- We use the built-in speakers of off-the-shelf mobile phones to emit stereo ultrasound signals to monitor human respiratory signals. Compared with traditional methods, this method is portable and easy to operate.
- We propose an optimization method of ultrasonic respiratory signal acquisition based on supervised learning. We use the Long Short-Term Memory (LSTM) network to transform ultrasonic signals into more accurate piezoelectric signals. The output piezoelectric signal is used as the input of other systems to improve the system's accuracy.
- Using several indicators, including MAE and Person correlation, evaluate the system in the time and frequency domains. In particular, we define the IoU index in the time domain. The model achieves less than 0.05 MAE, 0.8779 IoU, 0.76 Pearson correlation, and the two signals have a good similarity in the frequency domain. The evaluation result shows the effectiveness of this system.

The rest of this paper is organized as follows. Section II introduces the related work. Section III presents the system design and the working principles, followed by the implementation in Section IV. The evaluation of our system is shown in Section V. Section VI concludes the paper.

II. RELATED WORK

A. Ultrasonic Sensing

In recent years, ultrasonic sensing has been a trendy research topic. Its main principles are ranging and positioning. The ultrasonic sensor system first obtains the object's motion trajectory, then detects the object's motion through the classifier or modeling. Specific application scenarios include speech recognition [7], gesture recognition [8]–[10] and respiratory monitoring [1], [2], [5], [11]–[15]. Specialized ultrasonic sensors were used in past studies to monitor respiration signal [5], [11]. However, since the built-in speakers and microphones in mobile phones can also transmit and capture ultrasound signals caused by motion, more ultrasonic sensing applications are implemented on smartphones to improve accessibility. Active sensing, *i.e.*, emitting and receiving ultrasound signals with the same mobile phone, is commonly used in such cases. Current works on active ultrasonic sensing mainly focus on detecting particular events from respiratory signals [2], [12], [14], or the estimation of certain health metrics [1], [13], [15]. Nevertheless, issues such as human clothing and dispersive sounds emitted by the phone that might cause interference with the desired signal are seldom addressed, and there is still space for further optimization.

B. Piezoelectric Sensing

Piezoelectric sensors can measure physical signals such as displacement, acceleration, velocity, and pressure. Due to its comprehensive frequency response and ideal dynamic characteristics, a piezoelectric sensor is commonly used in dynamic measurement tasks. The principle of piezoelectric sensing is that when the human body contacts the sensor, the sensor converts the small changes in the human body's gravity into a charge signal. Piezoelectric sensors can be made from both organic and inorganic materials. However, compared with inorganic piezoelectric materials such as aluminum nitride (AlN), organic materials such as polyvinylidene fluoride (PVDF) are less expensive and more environmentally friendly [16]. However, as the output impedance of

the organic material is high and the output signal is weak, piezoelectric sensors made from organic material need to go through the charge amplification circuit to output an electrical signal [17]. Among the existing methods of respiratory signal collection, piezoelectric sensing is the most accurate and sensitive method. Furthermore, piezoelectric sensors can be easily manufactured into a flexible thin film, so they can be installed on beds conveniently without interfering with the users' sleep. For these reasons, piezoelectric sensing is widely used to monitor respiration and heartbeat for healthcare and clinical purposes [18]–[20]. Compared with previous methods, we innovatively combine ultrasonic and piezoelectric sensing, taking into account both advantages.

C. LSTM

As an improved recurrent neural network, long short-term memory network (LSTM) is widely used in machine learning fields. It can not only solve the problem that RNN cannot deal with long-distance dependence but also solve common problems such as gradient explosion or gradient disappearance in neural networks, which is very effective in processing sequence data. The structure of the LSTM network is shown in Fig. 2. Unlike previous RNN models, LSTM is composed of recurrently connected memory blocks, each containing an input gate, an output gate [21], and a forget gate [22].

The forget gate determines what information to discard from the cell state. It inputs the output h_{t-1} of the previous state and the input information of the current state X_t into a Sigmoid function to generate a value between 0 and 1, which is multiplied by the cell state to determine how much information to discard (keep). 0 indicates *completely discard*, and one indicates *completely retain*. The forget gate, after connecting h_{t-1} and X_t , is multiplied by a weight W_f and biased by b_f , which is the parameter that the network needs to learn. If the size of the hidden state (size of a hidden layer of neurons) is h_{size} , then the size of W_f is $h_{size} * h_{size}$. The value of h_{size} is manually set.

The input gate determines what new information to store in the cell state. It inputs h_{t-1} from the previous state and X_t from the current state into a Sigmoid function, producing a value between 0 and 1 i_t to determine how much new information we need to keep. At the same time, a tanh layer obtains a *candidate new information* C_t to be added to the cell state from the output h_{t-1} of the previous state and the input X_t of the current state. Multiply the value i_t with the *candidate new information* C_t to get the update we want to add to the cell state. The input gate (a Sigmoid function layer) and the tanh layer, both neural network layers, learn their parameters as before the forget gate.

The output gate determines what information to output from the cell state. As before, a Sigmoid function will first produce a number between 0 and 1 o_t to determine how much information in the cell state we need to output. The cell state information is first *activated* (nonlinear transformation) through a tanh layer when multiplied by o_t . So the output of

this LSTM block h_t is got. The output gate also has its weight parameters to be learned.

These gates can control the flow of information that enters or exits the memory block, and the memory block can remember values over arbitrary time intervals. This structural design allows LSTM to overcome the vanishing gradient problem in previous RNN models. Having the capability to maintain temporal memory, LSTM is an advanced method to deal with temporal sequences. Relevant applications include time-series prediction, natural language processing, and image captioning [23]. Furthermore, LSTM has been proven to be helpful in the fields of sound recognition and signal processing, as shown in studies by Laffitte *et al.* [24], Lyu *et al.* [25], and Wang *et al.* [26].

III. METHODOLOGY

A. Overview

The system design is shown in Fig. 1. We simultaneously obtain ultrasonic and piezoelectric respiratory signals using a mobile phone and a respiratory belt. Following downsampling, the Savitsky-Golay filter, and normalization, the ultrasonic signal is fed into the LSTM network. After low-pass filtering and normalization, the processed piezoelectric signal serves as the target output for supervised learning. Once the LSTM is trained, and a model with good performance is obtained, the network can convert a standardized ultrasonic input signal into an output piezoelectric signal.

B. Ultrasonic Data Acquisition

Typically, the built-in speakers of most commercially available smartphones can produce sounds up to 22 kHz. [12], [14]. Therefore, the ultrasonic signal emitted by the speaker can be expressed as:

$$s(t) = \cos(2\pi f_1 t) + \cos(2\pi f_2 t), \quad (1)$$

where $f_1 = 18,000$ and $f_2 = 22,000$.

After the microphone records the reflected signal $m(t)$, the respiratory sampler first uses a high-pass filter to eliminate components below 16 kHz. The reflected signal $m(t)$ can be considered as the product of the ultrasonic signal $s(t)$, and the original respiratory signal $x(t)$ [9], [15]. Therefore,

$$m(t) = x(t) \cdot s(t). \quad (2)$$

The original respiratory signal $x(t)$ is demodulated by multiplying $m(t)$ by $s(t)$, and the result is passed through a low-pass filter with a low cutoff frequency, such as 200Hz. According to Equation (1) and (2), we have:

$$\begin{aligned} m(t)s(t) &= x(t)s^2(t) = x(t)[\cos(2\pi f_1 t) + \cos(2\pi f_2 t)]^2 \\ &= x(t)[\cos^2(2\pi f_1 t) + 2\cos(2\pi f_1 t)\cos(2\pi f_2 t) \\ &\quad + \cos^2(2\pi f_2 t)] \\ &= x(t)\left\{\frac{1}{2}[1 + \cos(2\pi 2f_1 t)] + \cos(2\pi(f_1 + f_2)t) \right. \\ &\quad \left. + \cos(2\pi(f_2 - f_1)t) + \frac{1}{2}[1 + \cos(2\pi 2f_2 t)]\right\}. \end{aligned} \quad (3)$$

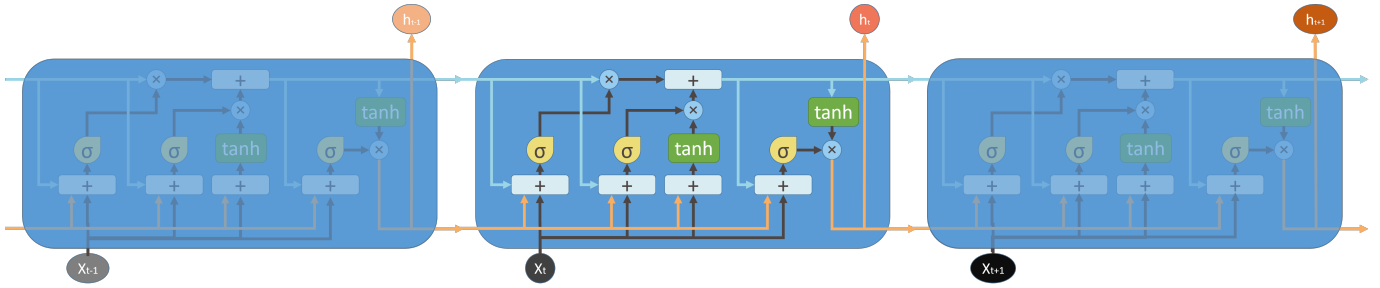


Fig. 2. LSTM network structure

After passing through the low-pass filter with the cut-off frequency of 200Hz, the components $\cos(2\pi \cdot 2f_1t)$, $\cos(2\pi \cdot 2f_2t)$, $\cos(2\pi(f_1 + f_2)t)$, and $\cos(2\pi(f_1 - f_2)t)$ are eliminated. Therefore, we have

$$m(t)s(t) \Rightarrow x(t) \left(\frac{1}{2} + \frac{1}{2} \right) = x(t). \quad (4)$$

The extracted $x(t)$ is used as the respiratory signal for subsequent steps.

C. Piezoelectric Sensor Acquisition

A piezoelectric sensor is a sensor of the piezoelectric effect produced by some dielectrics under force. The so-called piezoelectric effect refers to the phenomenon that some dielectrics will generate charges on their surfaces due to the polarization of internal controls when they are deformed (including bending and stretching deformation) by external forces in a specific direction. Piezoelectric materials can be divided into single piezoelectric crystals, piezoelectric polycrystalline, and organic piezoelectric materials. The most widely used piezoelectric sensors are all kinds of piezoelectric ceramics, including piezoelectric polycrystals and quartz crystals in piezoelectric monocrystals. Other piezoelectric single crystals, suitable for high-temperature radiation gallants, include lithium niobate and bismuth germanate.

The accuracy of the pressure sensor made of a semiconductor core is easily affected by temperature, so the temperature range of the pressure sensor should be considered. Static accuracy refers to the accuracy achieved at a specific temperature. It can be divided into four grades: 0.01% – 0.1% Full Span(FS) is super precision, 0.1–1% FS is precision, 1–2% FS is ordinary precision, and 2–10% FS is low precision.

D. LSTM Network

The LSTM network was built according to the structure shown in Fig. 2, and the dropout probability was set to be 0.3. The network's input is the ultrasonic signal after low-pass filtering and downsampling, and the dimension of each sample is (480, 1). The network's output is the filtered piezoelectric signal, and the dimension of each sample is (250, 1).

The system flow chart is shown in Fig. 1.



Fig. 3. The experimental scenario

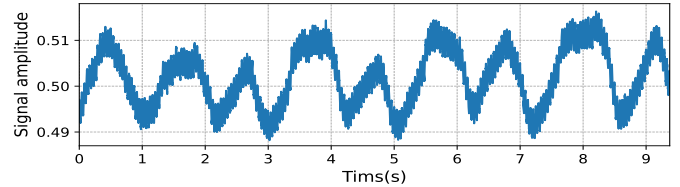


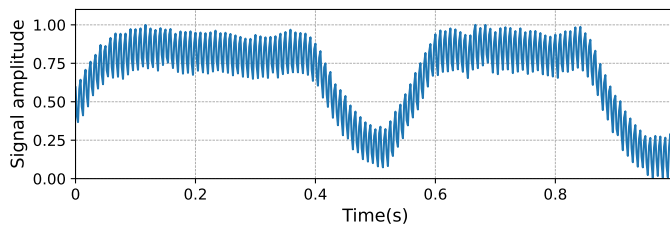
Fig. 4. Ultrasonic respiratory signal collected by mobile phone

IV. IMPLEMENTATION

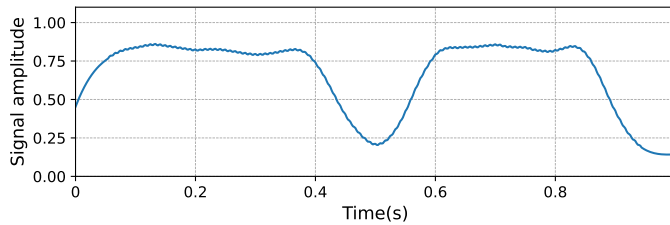
The experimental scenario is depicted in Fig. 3. The study subject is lying flat on the bed with the breathing belt in its exact position directly below the chest and the mobile phone speaker placed directly above the body, perpendicular to the chest. The participant maintains a steady respiration rate while the two devices collect data simultaneously.

The mobile phone speaker is positioned perpendicular to the participant's chest. The left channel transmits 18 kHz high-frequency signals, while the right channel transmits 22 kHz signals. The distance between the mobile phone speaker and the human body is about 5–10 cm. After the chest wall reflects the signal, the mobile phone's microphone receives the signal. The respiratory signal of the human body can be obtained after demodulation. The respiration sampler for this experiment is a Samsung Galaxy Z flip 5G smartphone running Android 12 at a sampling rate of 48 kHz. Ten volunteers' respiratory signals are collected. Each sample is collected for five minutes. The sampled signals are shown in Fig. 4.

Fig. 5(a) shows that the signal has a carrier frequency. A smoothing workflow is required to filter out this carrier and make the transition along the curve more gradual. We



(a)



(b)

Fig. 5. (a) Original ultrasonic signal (b) Filtered ultrasonic signal

use a Savitsky-Golay filter to process the ultrasonic signal. The filtering results are depicted in Fig. 5(b), and the curve becomes smooth.

We use a sleep monitoring tape to collect piezoelectric signals. Its piezoelectric thin-film sensor is sensitive to weak dynamic signals, including respiratory and heart rate signals. The sleep monitoring belt is laid flat on the bed when in use. When a person is lying on a bed, the piezoelectric film sensor can detect the heartbeat and breathing fluctuations. The respiratory and heart rates can be calculated and obtained. In this experiment, we only require the respiration-related piezoelectric signal.

Due to the hypersensitivity of the piezoelectric film, it is impossible to accurately determine whether or not a person is on or off the bed. A minor disturbance at the bedside could trigger a false alarm. This system's breathing belt has a shielding function, providing excellent environmental anti-interference. The membrane pressure sensor will only output a pressure signal when the human body is supine, reducing the likelihood of error.

The thin-film pressure sensor and piezoelectric thin-film sensor are integrated into a 1.5 mm thick cloth belt with a Universal Asynchronous Receiver/Transmitter (UART) signal acquisition circuit board, as shown in Fig. 6. The Microcontroller Unit (MCU) of the circuit board contains an algorithm that can directly output heart rate, respiratory rate, and other information.

The input voltage of the breathing belt is 5V (DC). We acquire data using the UART interface with a baud rate of 115200. First, the original 57-byte signal data packet is returned. The first 5 bytes are the American Standard Code for Information Interchange (ASCII) characters of `Odata`, followed by 50 bytes containing 25 groups of data, with a 2-byte, signed hexadecimal integer in each group (-32768-

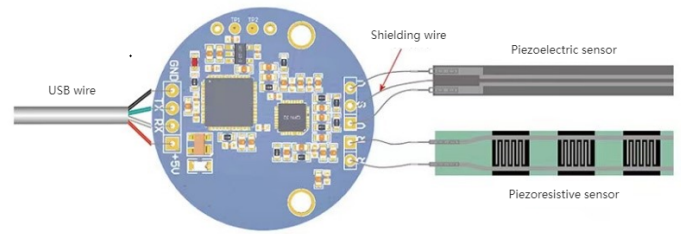


Fig. 6. UART signal acquisition circuit board

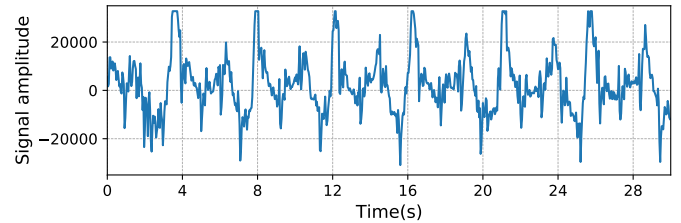


Fig. 7. the waveform of respiratory signal

32767). Representing the piezoelectric signal, the lower bit is in front, and the higher bit is in the back. The rate of sampling is 25 Hz. The final two bytes are also a signed hexadecimal integer, but the effective range is 0–4096, and the sampling rate is 1 Hz. They represent the piezoresistive signal.

A 9-byte result packet is then returned, with the first 5 bytes containing the ASCII characters for `Bdata`. A 1-byte unsigned integer serial number ranging from 0 to 59 is followed. The next 1-byte status value is an unsigned integer, with 0 indicating that the user is in bed, 1 indicating that the user has left the bed, 2 indicating that the user's body is moving, 3 indicating weak respiration, 4 indicating that a heavy object is on the bed (not a person), and 5 indicating that the user is snoring. The eighth 1-byte value is an unsigned integer representing the heart rate, followed by a 1-byte respiration rate. When respiration is weak, the output respiration rate is 0. After the user lies in bed for 25 seconds, the monitoring tape activates and returns the heart and respiratory rates via serial communication.

The data returned by the breathing belt is retrieved using a serial port debugging tool. Every second, 66 bytes of data will be returned. As depicted in Fig. 7, the returned data is processed to obtain the respiratory signal waveform. The ordinate represents the signal's amplitude, while the abscissa represents time. The piezoelectric respiratory signal is passed through a low-pass filter with a cut-off frequency of 1 Hz, and the waveform is shown in Fig. 8.

For the experiment, ten participants' data were collected. The volunteers, aged between 20 and 50, were five men and five women in good health. During the experiment, they were asked to lie flat on the bed at rest and breathe evenly. We had all the volunteers wear either a shirt or a T-shirt during the experiment.

In the same state, each data consists of an ultrasonic signal

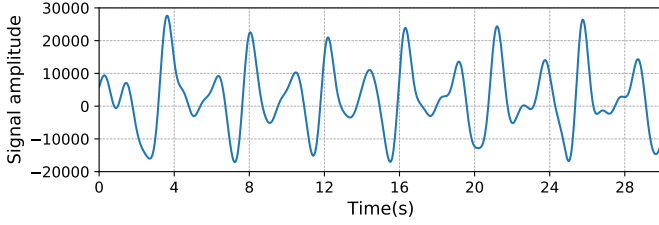


Fig. 8. the waveform of filtered piezoelectric respiratory signal

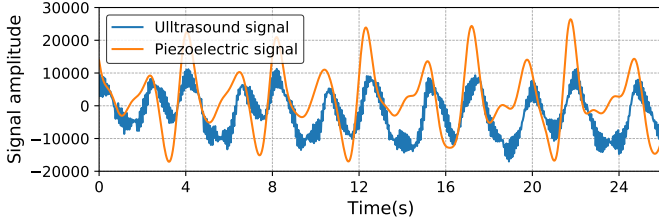


Fig. 9. The waveform of the piezoelectric and ultrasonic respiratory signal

and a piezoelectric signal with a duration of one minute. Within 10 seconds, they were aligned according to time and randomly cut into fragments. A total of 2,000 data groups were collected. The clipped ultrasonic signal segment serves as the input to the network, while the piezoelectric signal segment serves as the desired output. To train and evaluate the network, we randomly selected 1,600 groups of data as the training set and the remaining 400 groups as the test set.

As shown in Fig. 9, the piezoelectric and ultrasonic signals are collected simultaneously in the same state, and it could be roughly observed that the data are correlated.

V. EVALUATION

This section will evaluate the model's performance in both the time and frequency domains.

A. Time Domain Analysis

We use the mean absolute error (MAE), the intersection over union (IoU), and the Pearson correlation coefficient in the time domain to evaluate the model.

1) *Mean Absolute Error (MAE)*: The mean absolute error is calculated as

$$\text{MAE} = \frac{\sum_{i=1}^N |\hat{y}_i - y_i|}{N}, \quad (5)$$

where N is number of test cases, \hat{y}_i is the output given by the model, and y_i is the correct output. It shows the average value of the absolute error between the predicted value and the observed value. In general, the lower the MAE, the better.

We perform 1,000 iterations of testing on the LSTM model. As shown in Fig. 10, the model achieves less than 0.05 MAE.

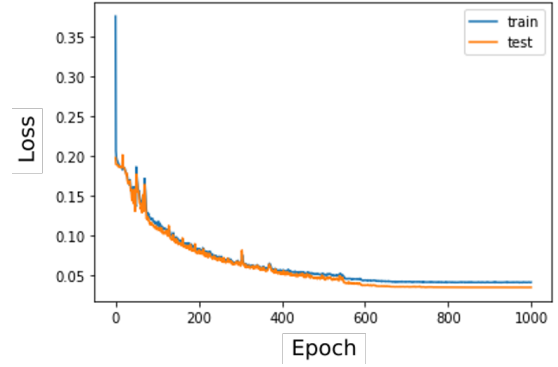


Fig. 10. The loss curve

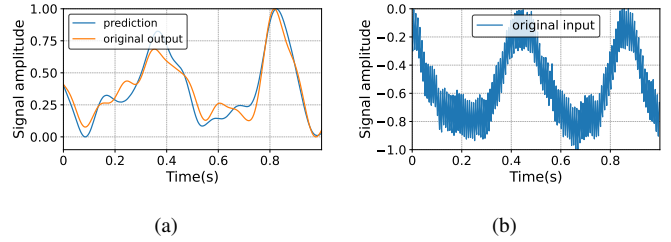


Fig. 11. (a) The model output results and the corresponding piezoelectric signal results (b) Ultrasonic input data

2) *Intersection over Union (IoU)*: To further verify the model's performance, we take a group of data with a length of 10 seconds as the model's input, which does not belong to the training set or the test set. The ultrasonic input data is shown in Fig. 11(b). The model output results and the corresponding piezoelectric signal results are shown in Fig. 11(a).

We use the intersection over union (IoU) in the time domain to measure the similarity between the predicted signal and the ground truth. IoU is a standard for measuring the accuracy of detecting corresponding objects in a specific data set.

Here, we define the IoU of the two signals $f(t)$ and $g(t)$ as:

$$\text{IoU} = \frac{\int_0^T \min(f(t), g(t)) dt}{\int_0^T \max(f(t), g(t)) dt} \quad (6)$$

That is, subtracting the intersection of two signals from the union of two signals. The definition of the union and intersection of signals is shown in Fig. 12. The larger the IoU, the better. When the IoU equals 1, the two signals are the same.

The IoU of the two signals in Fig. 11(a) is calculated to be 0.8779, indicating that the signal we predicted is highly similar to the ground truth.

3) *Pearson correlation coefficient*: The Pearson correlation coefficient measures the degree of correlation between two variables, and its value is between -1 and 1. When the Pearson coefficient equals 1, it means a complete positive correlation, 0 means no correlation, and -1 means an absolute negative correlation.

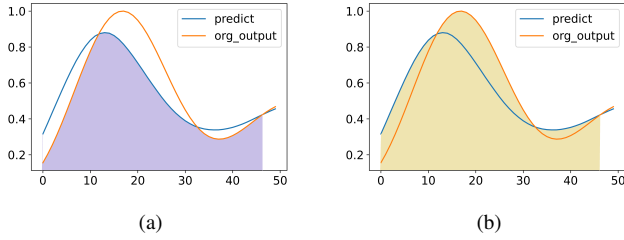


Fig. 12. (a) Intersection of two signals (b) Union of two signals

The Pearson correlation coefficient between two variables is defined as the quotient of covariance and standard deviation between two variables:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}. \quad (7)$$

The above formula represents the overall correlation coefficient, and ρ is often used as the representative symbol. The Pearson correlation coefficient can be obtained by estimating the covariance and standard deviation of the sample, which is commonly expressed as γ :

$$\gamma = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (8)$$

The greater the absolute value of the Pearson correlation coefficient, the greater the degree of correlation. The Pearson correlation coefficient of the two curves in 11(a) is 0.76. It shows that the predicted value correlates well with the ground truth.

B. Frequency Domain Analysis

We perform a fast Fourier transform (FFT) on the two signals in Fig. 11(a) to obtain the frequency domain diagram of the two signals, as shown in Fig. 13. After filtering the DC component, the first harmonic of the two signals is observed to be 0.4 Hz. The second harmonic frequency is 0.6 Hz. Therefore, the two signals have a high frequency domain similarity.

VI. CONCLUSION

Numerous systems have begun attempting to use mobile phones to emit ultrasound to monitor vital human health metrics due to their portability and usability. However, mobile phone ultrasonic waves are easily perturbed, resulting in significant errors. This paper uses piezoelectric sensors and supervised learning methods to propose a method for optimizing ultrasonic signals for monitoring human respiratory signals using ultrasonic waves emitted by mobile phones. The trained model performs admirably. This technique can increase the reliability of ultrasonic signals to improve the performance of other systems.

ACKNOWLEDGMENT

This work is supported in part by the Talent Fund of Beijing Jiaotong University (No. 2022XKRC013).

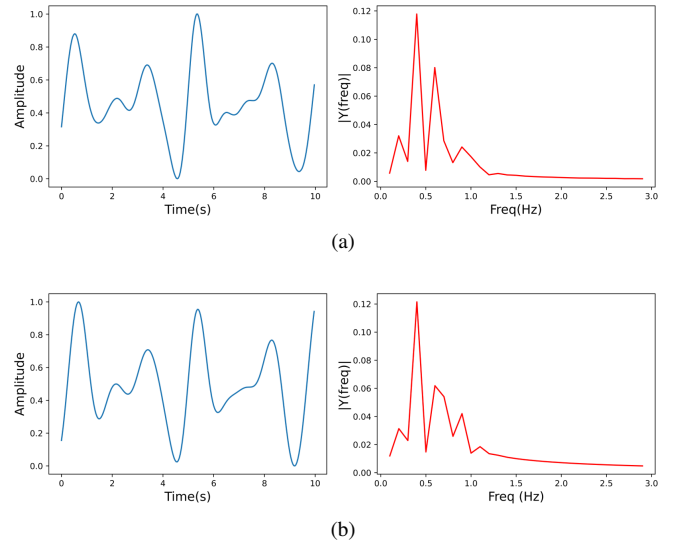


Fig. 13. (a) FFT results of piezoelectric signals (b) FFT result of model output signal

REFERENCES

- [1] L. Ge, J. Zhang, and J. Wei, "Single-frequency Ultrasound-based Respiration Rate Estimation with Smartphones," *Computational and Mathematical Methods in Medicine*, vol. 2018.
- [2] T. Wang, D. Zhang, L. Wang, Y. Zheng, T. Gu, B. Dorizzi, and X. Zhou, "Contactless Respiration Monitoring Using Ultrasound Signal with Off-the-shelf Audio Devices," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2959–2973, 2018.
- [3] Y. Wu, F. Li, Y. Xie, Y. Wang, and Z. Yang, "SymListener: Detecting Respiratory Symptoms via Acoustic Sensing in Driving Environments," *ACM Trans. Sen. Netw.*, feb 2022. [Online]. Available: <https://doi.org/10.1145/3517014>
- [4] X. Chang, C. Peng, G. Xing, T. Hao, and G. Zhou, "ISleep: A Smartphone System for Unobtrusive Sleep Quality Monitoring," *ACM Trans. Sen. Netw.*, vol. 16, no. 3, jul 2020. [Online]. Available: <https://doi.org/10.1145/3392049>
- [5] S. D. Min, D. J. Yoon, S. W. Yoon, Y. H. Yun, and M. Lee, "A Study on a Non-contacting Respiration Signal Monitoring System Using Doppler Ultrasound," *Medical & biological engineering & computing*, vol. 45, no. 11, pp. 1113–1119, 2007.
- [6] H. Kaneko and J. Horie, "Breathing Movements of the Chest and Abdominal Wall in Healthy Subjects," *Respiratory care*, vol. 57, no. 9, pp. 1442–1451, 2012.
- [7] J. Tan, C.-T. Nguyen, and X. Wang, "SilentTalk: Lip Reading through Ultrasonic Sensing on Mobile Phones," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.
- [8] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "Fingerio: Using Active Sonar for Fine-grained Finger Tracking," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 1515–1525.
- [9] W. Wang, A. X. Liu, and K. Sun, "Device-free Gesture Tracking Using Acoustic Signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 82–94.
- [10] Z. Wang, Y. Hou, K. Jiang, W. Dou, C. Zhang, Z. Huang, and Y. Guo, "Hand Gesture Recognition Based on Active Ultrasonic Sensing of Smartphone: a Survey," *IEEE Access*, vol. 7, pp. 111 897–111 922, 2019.
- [11] S. D. Min, J. K. Kim, H. S. Shin, Y. H. Yun, C. K. Lee, and M. Lee, "Noncontact Respiration Rate Measurement System Using an Ultrasonic Proximity Sensor," *IEEE sensors journal*, vol. 10, no. 11, pp. 1732–1739, 2010.
- [12] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless Sleep Apnea Detection on Smartphones," in *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, 2015, pp. 45–57.

- [13] J. Chauhan, Y. Hu, S. Seneviratne, A. Misra, A. Seneviratne, and Y. Lee, "BreathPrint: Breathing Acoustics-based User Authentication," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 278–291.
- [14] X. Wang, R. Huang, and S. Mao, "SonarBeat: Sonar Phase for Breathing Beat Monitoring with Smartphones," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2017, pp. 1–8.
- [15] X. Song, B. Yang, G. Yang, R. Chen, E. Forno, W. Chen, and W. Gao, "SpiroSonic: Monitoring Human Lung Function via Acoustic Sensing on Commodity Smartphones," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.
- [16] M. T. Chorsi, E. J. Curry, H. T. Chorsi, R. Das, J. Baroody, P. K. Purohit, H. Iliès, and T. D. Nguyen, "Piezoelectric Biomaterials for Sensors and Actuators," *Advanced Materials*, vol. 31, no. 1, p. 1802084, 2019.
- [17] J. Hou, Y. Zhang, S. Zhang, X. Geng, Y. Wang, C. Chen, and H. Zhang, "Respiration Signal Extraction from Pulse Wave Collected by PVDF Sensor," *IEEE Access*, vol. 8, pp. 149 878–149 886, 2020.
- [18] N. Bu, N. Ueno, and O. Fukuda, "Monitoring of Respiration and Heartbeat During Sleep Using a Flexible Piezoelectric Film Sensor and Empirical Mode Decomposition," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 1362–1366.
- [19] Y.-Y. Chiu, W.-Y. Lin, H.-Y. Wang, S.-B. Huang, and M.-H. Wu, "Development of a Piezoelectric Polyvinylidene Fluoride (PVDF) Polymer-based Sensor Patch for Simultaneous Heartbeat and Respiration Monitoring," *Sensors and Actuators A: Physical*, vol. 189, pp. 328–334, 2013.
- [20] S. So, D. Jain, and N. Kanayama, "Piezoelectric Sensor-based Continuous Monitoring of Respiratory Rate during Sleep," *Journal of Medical and Biological Engineering*, vol. 41, no. 2, pp. 241–250, 2021.
- [21] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [22] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to Forget: Continual Prediction with LSTM," *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [23] G. Van Houdt, C. Mosquera, and G. Nápoles, "A Review on the Long Short-term Memory Model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, 2020.
- [24] P. Laffitte, Y. Wang, D. Sodoyer, and L. Girin, "Assessing the Performances of Different Neural Network Architectures for the Detection of Screams and Shouts in Public Transportation," *Expert systems with applications*, vol. 117, pp. 29–41, 2019.
- [25] C. Lyu, Z. Liu, and L. Yu, "Block-sparsity Recovery via Recurrent Neural Network," *Signal Processing*, vol. 154, pp. 129–135, 2019.
- [26] Q. Wang, P. Du, J. Yang, G. Wang, J. Lei, and C. Hou, "Transferred Deep Learning Based Waveform Recognition for Cognitive Passive Radar," *Signal processing*, vol. 155, pp. 259–267, 2019.