

Article

A Speech Enhancement Algorithm for Speech Reconstruction Based on Laser Speckle Images

Xueying Hao ¹, Dali Zhu ^{2,3}, Xianlan Wang ¹, Long Yang ^{2,3} and Hualin Zeng ^{2,3,*}¹ Graduate Department, Wuhan Research Institute of Posts and Telecommunications, Wuhan 430074, China² Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China³ School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: zenghualin@iie.ac.cn

Abstract: In the optical system for reconstructing speech signals based on laser speckle images, the resonance between the sound source and nearby objects leads to frequency response problem, which seriously affects the accuracy of reconstructed speech. In this paper, we propose a speech enhancement algorithm to reduce the frequency response. The results show that after using the speech enhancement algorithm, the frequency spectrum correlation coefficient between the reconstructed sinusoidal signal and the original sinusoidal signal is improved by up to 82.45%, and the real speech signal is improved by up to 56.40%. This proves that the speech enhancement algorithm is a valuable tool for solving the frequency response problem and improving the accuracy of reconstructed speech.

Keywords: laser speckle image; speech enhancement; frequency spectrum correlation coefficient; frequency response

1. Introduction

In the field of information security and the prevention of social crimes, covertly acquiring clear remote speech signals has become an important research topic [1,2]. For years, researchers have proposed a variety of methods for measuring vibration signals [3–11], many of which are used in speech signal reconstruction and are based on laser. There are two main methods based on remote vibration signals detection: one is based on laser Doppler vibrometer (LDV) [3–8] and the other is optical sensing and image processing technologies [9–11]. LDV has the advantages of non-contact, high spatial and temporal resolution, real-time processing. But LDV cannot fulfill the full field detection. The optical sensing and image processing technology is mainly divided into two types: natural light image processing [12] and coherent light image processing [13–16]. The former has simple system structure and easy to implement. However, it requires high computational cost, and its processing speed is slow. In the 1980s, researchers of [13,14] proposed the method to extract subtle motion from speckle images. Additionally, in recent years, researchers have also reconstructed speech based on speckle images [15,16]. The principle of this method is to irradiate the rough surface object near the sound source with laser, and the reflected light interferes to form secondary speckle [17]. Then the sound makes the surrounding objects vibrate slightly, which makes the speckle pattern shift slightly. The remote speech signal is reconstructed by extracting the small movement between speckle images. The system of remote speech reconstruction based on laser speckle pattern is mainly divided into two parts: the first is the construction of an optical system, which is used to collect continuous speckle image sequences; the second is the design of a reconstruction algorithm, which is used to detect movement from speckle images and reconstruct remote speech signal.

As regards optical system construction, ref. [18] proposed a set of simplified optical equipment, it reduces the cost and realizes full-field non-contact detection. This optical device makes the technology more suitable for remote monitoring. The current laser technology and imaging technology can collect high-quality speckle images at a low price.



Citation: Hao, X.; Zhu, D.; Wang, X.; Yang, L.; Zeng, H. A Speech Enhancement Algorithm for Speech Reconstruction Based on Laser Speckle Images. *Sensors* **2023**, *23*, 330. <https://doi.org/10.3390/s23010330>

Academic Editor: Jinjun Xia

Received: 16 August 2022

Revised: 2 September 2022

Accepted: 3 September 2022

Published: 28 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Regarding speech reconstruction algorithm, researchers have proposed a variety of methods, including digital image correlation (DIC) [19–24], optical flow method [16,25,26], and intensity method [27,28]. Ref. [29] proposed a geometric method to explain the motion of speckle. Ncorr [20] is a digital correlation algorithm specially designed by researchers for two-dimensional images. In recent years, with the rapid development of machine learning and artificial intelligence technology, many neural-network-based methods have been proposed, such as convolutional neural network (CNN) [30–32] and convolutional long short-term memory (LSTM) [33].

At present, the detection method based on laser speckle images is the most suitable for this task, but the problem of frequency response has been found in the actual experiments and applications. We found that the speech signal reconstructed by this optical device [18] has inhomogeneous enhancement or attenuation of speech components at different frequencies, and the frequency response of different vibration objects is different. Any object has a natural frequency, when the vibration sound wave of this natural frequency is transmitted to the object, the vibration amplitude of the object will have the maximum growth. Different objects have different natural frequencies, so the performance of frequency response is also different. This phenomenon greatly affects the accuracy of the reconstructed sound signal.

At present, many reconstruction algorithms ignore the frequency response, fewer algorithm is proposed to solve this problem. In this paper, a speech enhancement algorithm for different vibration objects is proposed to weaken the frequency response and improve the accuracy of speech reconstruction.

This paper has two main contributions.

- The frequency response of long-distance speech reconstruction based on laser speckle image is proposed.
- Our algorithm is a speech enhancement algorithm designed to reduce the influence of frequency response, which greatly improves the accuracy of reconstructed speech signals.

The rest of this paper is organized as follows: Section 2 introduces the methodology of this paper, including DIC method and speech enhancement algorithm. Section 3 present the experimental setup. Section 4 introduces the experimental datasets and evaluation metrics. Section 5 is the result. Section 6 is the conclusion.

2. Methodology

2.1. Digital Image Correlation Method

There is room for improvement in many speech reconstruction algorithms. For example, the intensity method runs fast but has low accuracy. The performance of the six algorithms is compared in [34], and the results show that the cross-correlation method is one of the best choices. In order to accurately detect the frequency response in this optical device, this paper uses the DIC method as the benchmark.

Figure 1 shows the conversion between object vibration and speckle pattern displacement. In this figure, the transversal plane is regarded as the xoy plane, and the axial axis is the z -axis. The motion of this object can be divided into three directions: transverse, axial, and tilt. Ref. [18] proved that when strongly defocus the speckle image captured by the camera, only the tilt motion has the noticeable impact on the displacement of the speckle image, the influence of the other two motions on the shape and displacement of the speckle image can be ignored.

Refs. [35,36] deduced the speckle formation theory in detail, and ref [18] mentioned that making the imaging plane move from Z_1 to Z_2 . $A(x_o, y_o)$ is the amplitude distribution of speckle at Z_2 .

$$A(x_o, y_o) = \left| \iint \exp[i\phi(x, y)] \exp[i(\beta_x x + \beta_y y)] \exp\left[\frac{-2\pi i}{\lambda Z_2}(xx_o + yy_o)\right] dx dy \right| \quad (1)$$

$$\beta_x = \frac{4\pi \tan \alpha_x}{\lambda} \quad (2)$$

$$\beta_y = \frac{4\pi \tan \alpha_y}{\lambda} \tag{3}$$

where λ is the optical wavelength and ϕ is the random phase generated by the rough surface. α_x, α_y are tilting angle in the x -axis and y -axis respectively.

The inverse of the magnification of the imaging system is expressed in M .

$$M = \frac{Z_3 - F}{F} \approx \frac{Z_3}{F} \tag{4}$$

where F is the focal length of the imaging lens.

The speckles pattern displacement d is related to the tilting angle α of the object.

$$d = \frac{Z_2 \alpha}{M} \tag{5}$$

According to the previous description, we know that under appropriate conditions, there is only speckle movement between two consecutive speckle images, but no shape change. And the displacement is in the xoy plane. The relationship between two consecutive speckle images I_t, I_{t+1} is:

$$I_{t+1} = I_t(x + \Delta x, y + \Delta y) \tag{6}$$

where $\Delta x, \Delta y$ are the relative displacements in the x -axis and the y -axis direction.

The calculation steps of DIC method are shown in Figure 2. Take two consecutive speckle images I_t and I_{t+1} . Calculate the correlation coefficient between two images. At the maximum correlation, the corresponding displacement is the relative displacement between two images.

Corr2 is a function of finding the correlation coefficient between two matrices. The correlation coefficient r can be expressed as:

$$r(A, B) = \text{corr2}(A, B) = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right) \left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}} \tag{7}$$

where A and B are two-dimensional matrices of the same size, \bar{A} is the mean of the matrix A , and \bar{B} is the mean of the matrix B .

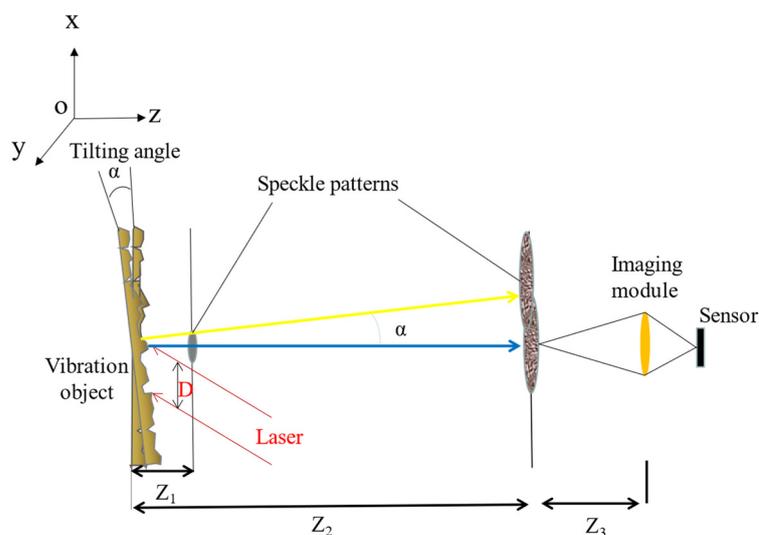


Figure 1. Theoretical description of the system. The conversion between object vibration and speckle pattern displacement.

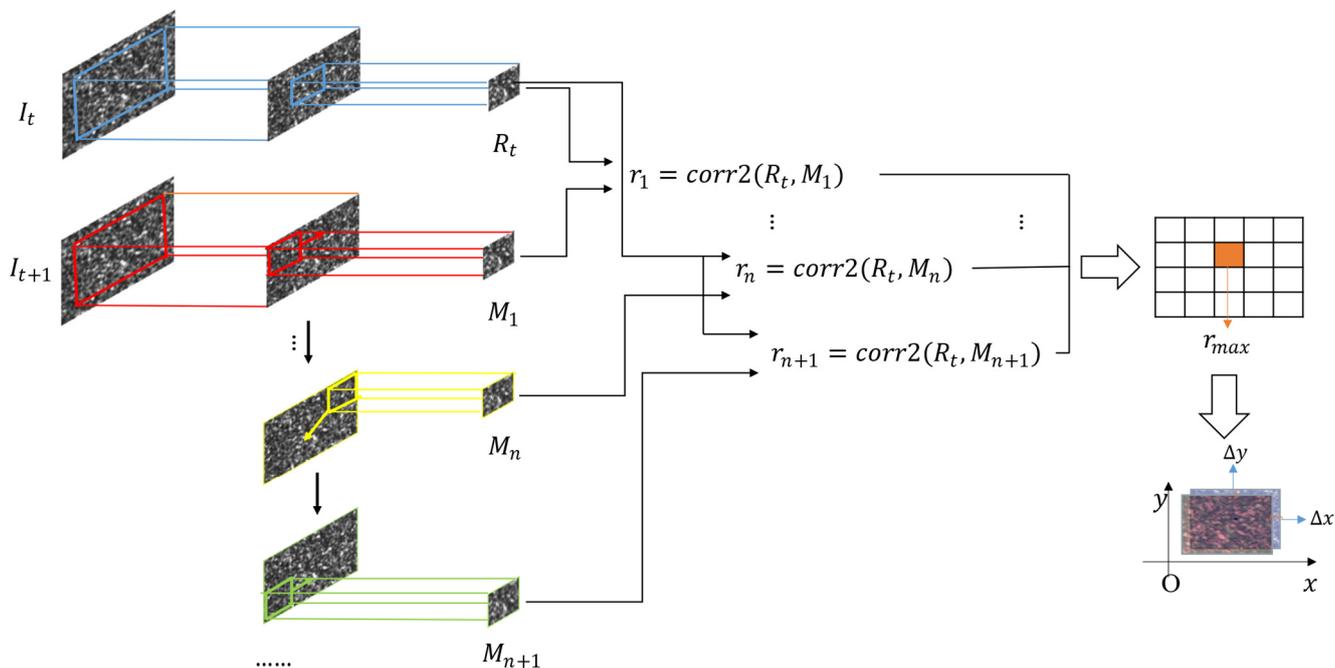


Figure 2. Digital image correlation method.

2.2. Speech Enhancement Algorithm

We use a sinusoidal signal which frequency varies from 80 to 1600 Hz and amplitude remains constant as the audio at the sound source (Figure 3 shows the waveform and frequency spectrum of the sinusoidal signal), and collect the corresponding speckle image dataset. Then use the algorithm in Section 2.1 to reconstruct the sinusoidal signal when the vibration object is a carton (Figure 4 shows the waveform and frequency spectrum of the reconstructed signal). By comparing Figure 3 with Figure 4, we found that the amplitude of the reconstructed signal is different at different frequencies, which seriously affects the accuracy of the reconstructed speech signal. Figure 5 is the reconstructed signal when the vibration object is a paper cup. By comparing Figure 4 with Figure 5, we found that the frequency responses of two different vibration objects are also different.

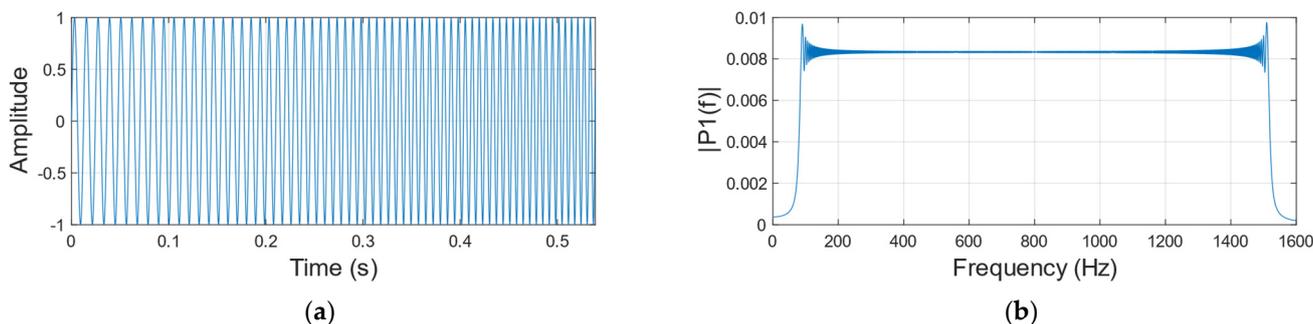


Figure 3. Original sinusoidal signal. (a) Time domain diagram. (b) Frequency domain diagram.

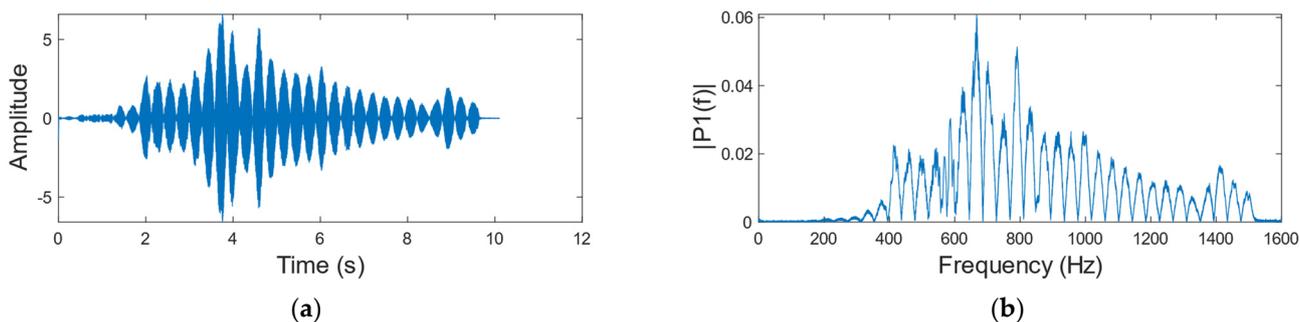


Figure 4. Reconstructed sinusoidal signal when a carton is the vibration object. (a) Time domain diagram. (b) Frequency domain diagram.

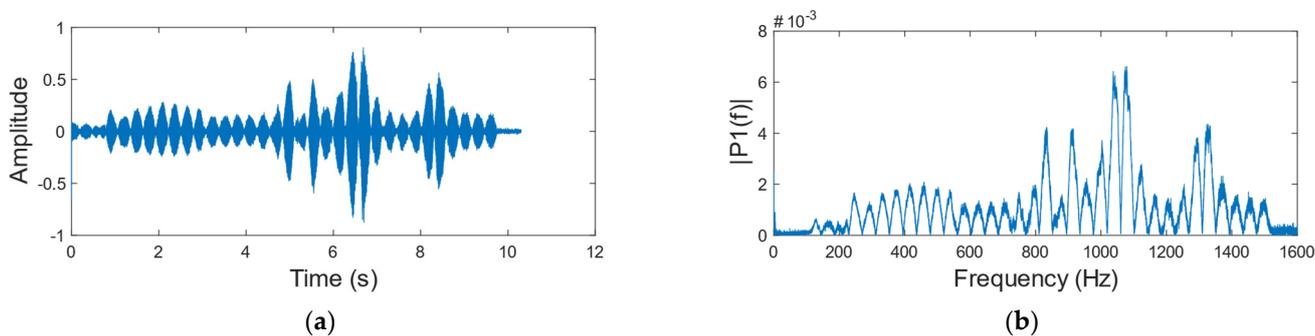


Figure 5. Reconstructed sinusoidal signal when a paper cup is the vibration object. (a) Time domain diagram. (b) Frequency domain diagram.

According to this phenomenon, we calculated the amplitude changes of several common vibration objects at different frequencies. According to the degree of change, the reconstructed speech signals of different frequencies are enhanced to different degrees. The above method can reduce the influence of the frequency response. The design flow of the algorithm is shown in Figure 6. To more clearly explain the steps involved in this speech enhancement algorithm, Figure 7 shows the results of each step of the speech enhancement algorithm when the vibration object is a carton.

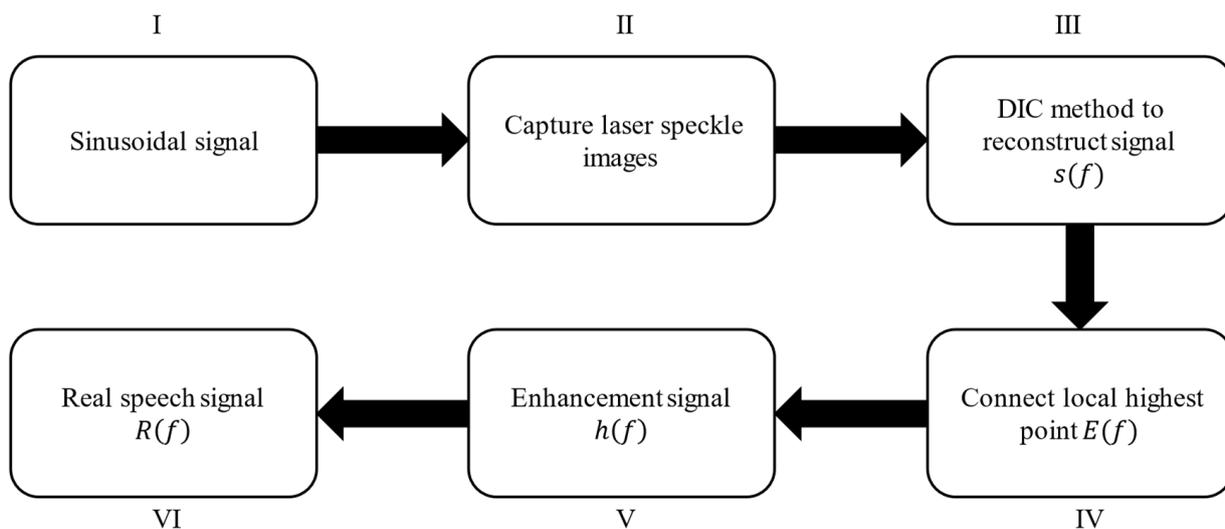


Figure 6. Speech enhancement algorithm.

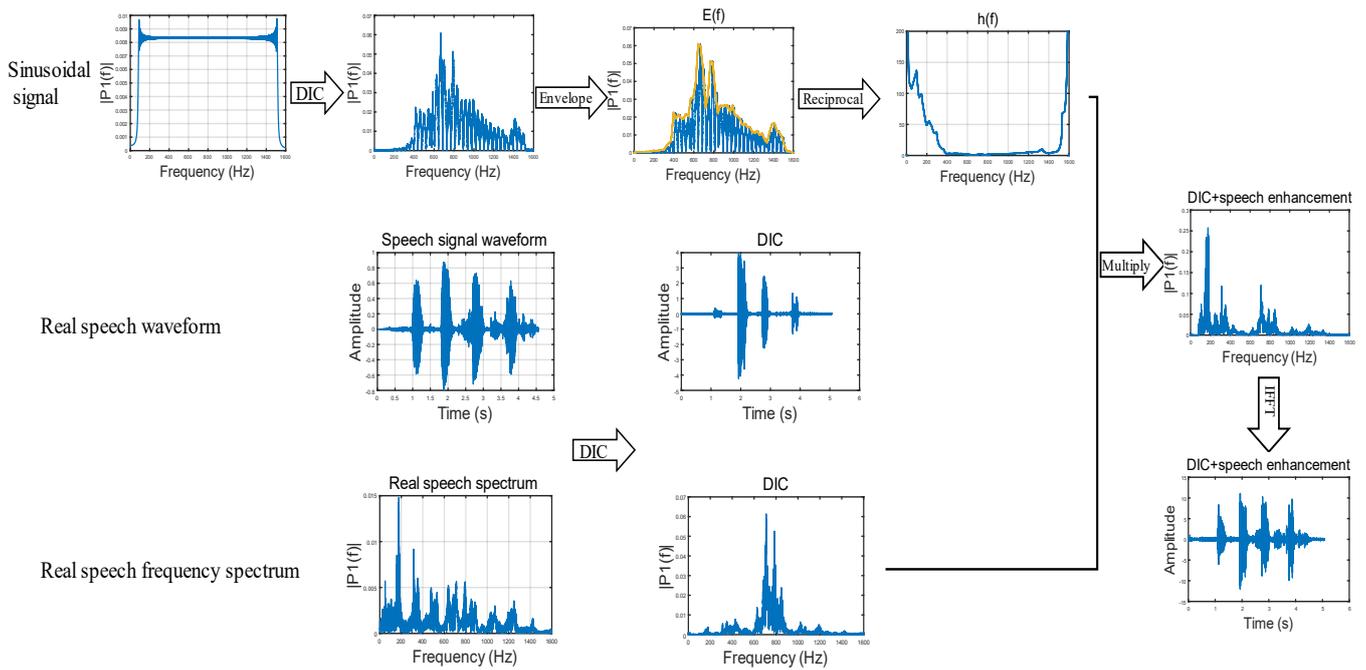


Figure 7. The process of speech enhancement when the vibration object is a carton.

- I. Sinusoidal signal: Table 1 shows the common frequency range of speech. The sinusoidal signal which frequency varies from 80 to 1600 Hz and amplitude remains constant at 1.
- II. Capture laser speckle images: Use the high-speed camera to collect the corresponding laser speckle images.
- III. DIC method: Speech signals are reconstructed from speckle images using the DIC method described in Section 2.1. The discrete frequency domain sequence of the sinusoidal signal reconstructed is $s(f)$.
- IV. Connect local highest point: Connect the local highest points of the amplitude signal in the reconstructed speech spectrum to obtain its envelope signal. The envelope signal is named $E(f)$.
- V. Enhancement signal: Equation (8) shows the calculation formula of the signal $h(f)$. Note that $E(f)$ at some frequencies is very small, which will make $h(f)$ very large. When $h(f)$ is greater than 1000, set it to 1.

$$h(f) = \frac{E(f)_{max}}{E(f)} \quad (8)$$

- VI. Real speech domain: Multiply the discrete frequency domain sequence of the reconstructed real speech $r_1(f)$ with the signal $h(f)$. The enhanced speech discrete frequency domain sequence is $R(f)$. According to the frequency range of speech, we process the speech with the frequency of 80–1400Hz.

$$R(f) = r_1(f) \times h(f) \quad (9)$$

After the following steps, the discrete frequency domain sequence $R(f)$ of the real speech signal will be obtained, and the speech waveform after speech enhancement can be obtained by using the inverse discrete Fourier transform (Equation (10)):

$$r_n = \frac{1}{N} \sum_{k=1}^N R_k e^{2\pi i(n-1)(k-1)/N} \quad (10)$$

where N is the total number of samples of the R sequence, R is the speech discrete frequency domain sequence after speech enhancement, and r is the speech discrete time domain sequences after speech enhancement.

Table 1. Frequency range of common speech (Hz).

Sound Area	Male	Female
Bass	82–392	82–392
Midrange	123–493	123–493
Treble	164–698	220–1100
Fundamental frequency range	64–523	160–1200

3. Experimental Setup

According to previous research results, the formation and acquisition process of laser speckle image can be set (as Figure 8). Speckle is a three-dimensional ellipsoidal shape with its long axes facing the light propagation direction. Figure 9 shows the experimental platform. Figures 10 and 11 are the equipment simulation diagrams of two kinds of lasers, and Figure 12 is the equipment physical map. The equipment used mainly includes:

1. The high-speed camera MVCAM AI-030U815M, with a maximum frame rate of 3200 frames per second (fps);
2. He-Ne laser (detailed parameters are shown in Table 2);
3. Fiber laser (detailed parameters are shown in Table 3);
4. Machine vision experiment frame, with fine-tuning camera clip and universal clip;
5. One personal computer (PC) with universal serial bus 3.0 (USB3.0) interface.

In this experiment, the high-speed camera is used to collect speckle images. The frame rate of the camera is closely related to the exposure interval. The process of speckle image acquisition by high-speed charge coupled device (CCD) camera can be regarded as uniform sampling of continuous speckle video. The sampling frequency f_s of the high-speed camera and the highest frequency of speech f_m need to satisfy the Nyquist theorem (Equation (11)).

$$f_s \geq 2f_m \quad (11)$$

Equation (11) shows that the frame rate of the high-speed camera should be greater than or equal to two times of the highest frequency of speech. The data in Table 1 show the speech frequency range of male and female, and $f_m = 1200$ Hz. This paper considers that all the frequency ranges in the speech can be restored to meet the actual use requirements, so the frame rate of the high-speed camera is greater than or equal to 2400 fps. The frame rate of the high-speed camera used in this experiment is 3200 fps, which meets the basic requirements of speech reconstruction.

In conjunction with Figure 1, we describe various physical parameters of speckle images. The first is the resolution of the speckle patterns in the Z_2 plane:

$$\delta_x = \frac{\lambda Z_2}{D} \cdot \frac{1}{M} = \frac{\lambda F}{D} \cdot \frac{Z_2}{Z_3} \quad (12)$$

where D is the diameter of the laser beam, Z_2 , Z_3 are different distances in Figure 1. F is the focal length of the lens. λ is the optical wavelength.

The optical system has requirements for the focal length F . The size of the pixel in the detector is Δs . It is assumed that every speckle in this plane will be observed at least by K pixels. The requirements for F are as follows:

$$F = \frac{K\Delta s Z_3 D}{Z_2 \lambda} \quad (13)$$

The distance Z_2 needs to be satisfied:

$$Z_2 > \frac{D^2}{4\lambda} \tag{14}$$

Finally, the number of speckles in every dimension of the spot is N :

$$N = \frac{\phi}{M\delta_x} = \frac{\phi D}{\lambda Z_2} = \frac{F \cdot D}{F_{\#} \lambda Z_2} \tag{15}$$

where ϕ is the diameter of the aperture of the lens, $F_{\#}$ is the F number of the lens, $M\delta_x$ represents the speckle size obtained on the Z_2 -plane.

Table 2. Basic parameters of the He-Ne laser.

Parameter	Quantity	Unit
Wavelength	632.8	nm
Operating current	4~6	mA
Rated voltage	220 ± 22	V
Rated frequency	50	Hz
Rated input power	<20	W

Table 3. Basic parameters of the fiber laser.

Parameter	Quantity	Unit
Center wavelength	635	nm
Continuous output power	60	mW
Operating voltage	2.55	V
Threshold current	70	mA
Operating current	190	mA

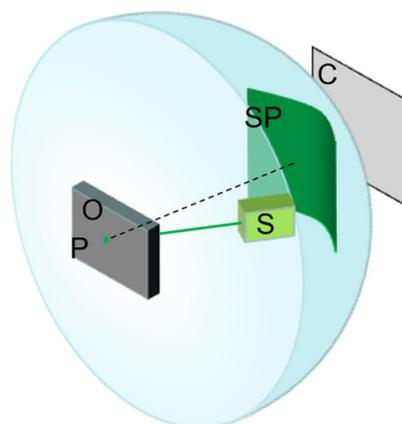


Figure 8. Schematic diagram of speckle image formation and collection. The laser beam generated by laser S irradiates on object O with a rough surface near the sound source (P is a point on object O surface), and then the reflected light interferes to form speckle. C is the camera sensor (the size is exaggerated).

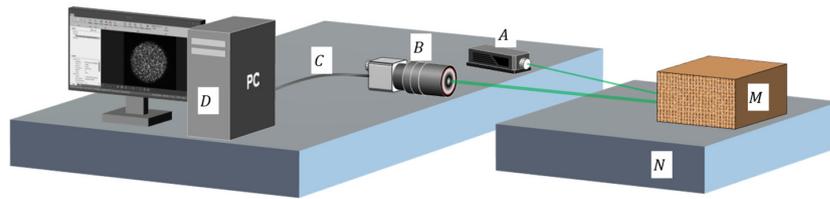


Figure 9. Experimental setup diagram. The laser beam generated by laser A is projected onto object M on the N platform. The change of speckle is collected by B camera, and these images are transmitted to the D computer through the C interface to extract the movement between speckle images.

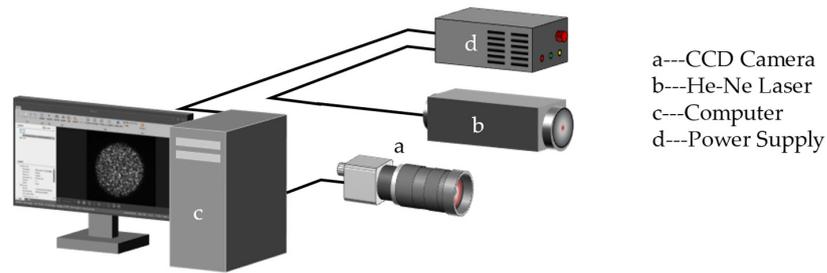


Figure 10. Simulation diagram of the He-Ne laser system.

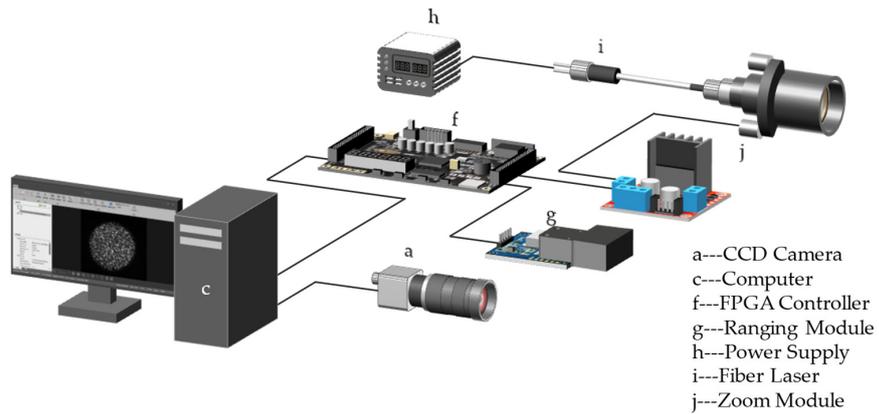


Figure 11. Simulation diagram of the fiber laser system.

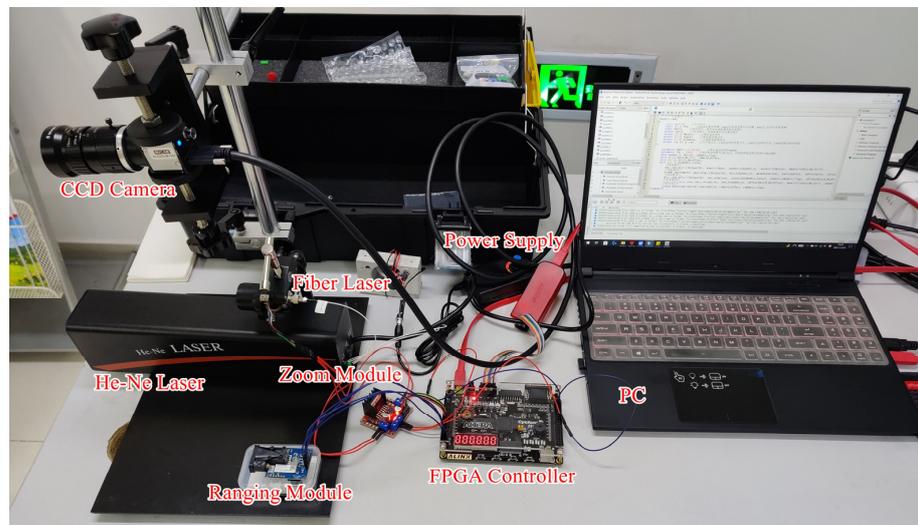


Figure 12. Equipment physical map.

4. Experiment Datasets and Evaluation Metrics

4.1. Data Collection

4.1.1. Sinusoidal Datasets

According to the speech frequency range in Table 1, we set the frequency of the sinusoidal signal to gradually change from 80 to 1600 Hz, with an amplitude of 1. Irradiate the laser generated by the He-Ne laser on five vibration objects near the sound source: carton, A4 paper, plastic cup, paper cup, and leaf. Use the CCD camera to collect the speckle image sequence generated. The number of speckle images is shown in Table 4.

Table 4. The number of speckle images acquisitions for sinusoidal datasets.

Vibration Object	Number of Speckle Images	Acquisition Time (s)
Carton	32,344	10.1
A4 paper	35,952	11.2
Plastic cup	31,941	10.0
Paper cup	32,970	10.3
Leaf	33,427	10.4

4.1.2. Two Laser Datasets

In order to verify the versatility of the algorithm under different laser systems, we use the He-Ne laser and the fiber laser to irradiate the carton near the sound source, and the number of speckle images collected is shown in Table 5.

Table 5. The number of speckle images acquisitions for two laser datasets.

Laser	Number of Speckle Images	Acquisition Time (s)
He-Ne laser	16,213	5.1
Fiber laser	18,975	5.9

4.1.3. Five Vibration Object Datasets

To verify the performance of the algorithm under different vibration objects, we use real speech as the sound source signal, project the laser onto different vibration objects, and collect the corresponding laser speckle image sequences respectively. The number of speckle images collected by five vibration objects is shown in Table 6.

Table 6. The number of speckle images acquisitions for five vibration object datasets.

Vibration Object	Number of Speckle Images	Acquisition Time (s)
Carton	18,140	5.7
A4 paper	16,785	5.2
Plastic cup	19,443	6.1
Paper cup	19,247	6.0
Leaf	18,647	5.8

4.2. Evaluation Metrics

This paper needs to calculate the correlation between the original audio signal and the speech signal reconstructed from speckle images. Since the start time of the two signals cannot be accurately aligned, the waveform correlation coefficient in the time domain is meaningless. To evaluate the accuracy of a reconstructed speech, we use the frequency spectrum correlation coefficient as the evaluation metrics. The frequency spectrum correlation coefficient here refers to the amplitude correlation coefficient of the frequency spectrum in detail.

As we know, scaling speech amplitude will affect the loudness of the speech, not the timbre. Therefore, using the frequency spectrum correlation coefficient can quantify the accuracy of speech reconstruction.

The value of the correlation coefficient is between -1 and 1. When it is 1, the two speeches are equal or proportional. When it is -1, one speech is equal to or proportional linearly to another negative speech.

$$\rho(x, y) = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{x_i - \mu_x}{\sigma_x} \right) \left(\frac{y_i - \mu_y}{\sigma_y} \right) \quad (16)$$

where x and y represent the amplitude signal of discrete frequency spectrum signals with the same sampling points. N is the total number of samples. μ_x, μ_y is the mean value of x and y accordingly. σ_x, σ_y is the standard deviation of x and y accordingly. ρ is the correlation coefficient.

5. Results

5.1. Performance on Sinusoidal Datasets

We recorded the changes of the frequency spectrum correlation coefficient between the reconstructed speech and the original speech after using the speech enhancement algorithm. We used the speech enhancement algorithm in this paper to process the sinusoidal datasets to verify whether the algorithm can increase the frequency spectrum correlation coefficient between the reconstructed sinusoidal signal and the original sinusoidal audio. Table 7 shows the sinusoidal signal reconstruction accuracy in detail. Figure 13 shows the line graph of the frequency spectrum correlation coefficient of reconstructed sinusoidal signal on different vibration objects. The existence of frequency response has a great impact on the reconstruction of sinusoidal signal, which will make the spectral correlation coefficient low. The sinusoidal signal dataset only verifies whether our proposed speech enhancement algorithm can reduce the frequency response; it cannot be verified whether the algorithm can also play a positive role in real speech. Therefore, the following experimental results are used to verify the enhancement effect of this algorithm on real speech.

Table 7. Comparison of sinusoidal signal reconstruction accuracy.

Vibration Object	DIC	DIC + Speech Enhancement	Improve
Carton	0.2968	0.4548	53.23%
A4 paper	0.3348	0.4561	36.23%
Plastic cup	0.2593	0.4731	82.45%
Paper cup	0.2844	0.4409	55.03%
Leaf	0.2460	0.3963	61.10%

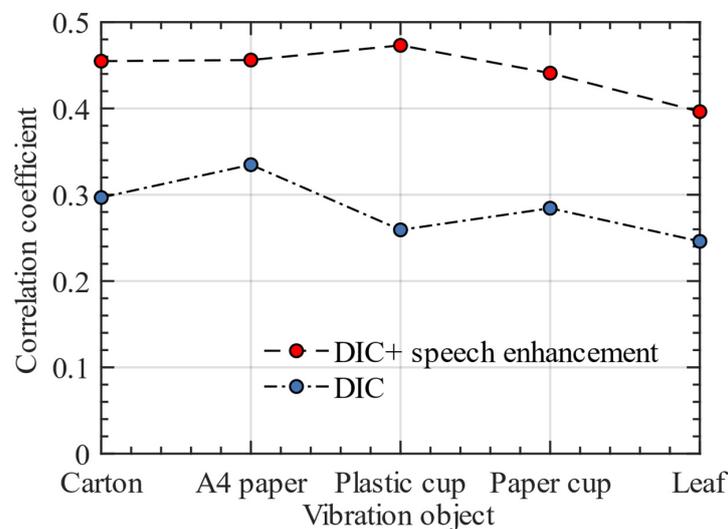


Figure 13. Accuracy comparison of sinusoidal signal reconstruction with and without enhancement.

5.2. Performance on Two Laser Datasets

Using the same speech as the as the audio at the sound source, two lasers are used to irradiate the same vibration object (carton), and the CCD camera is used to collect the laser speckle image sequences. Table 8 is the comparison of the frequency spectrum correlation coefficient with and without speech enhancement. Figure 14 shows the reconstructed speech and original speech waveforms using the He-Ne laser. Figure 15 shows the reconstructed speech and original speech waveforms using the fiber laser. In the waveform comparison, we can also clearly find that the waveform of the speech is significantly improved after the speech enhancement.

Table 8. Comparison of the speech reconstruction accuracies of two laser datasets.

Laser	DIC	DIC + Speech Enhancement	Improve
He-Ne laser	0.3624	0.5668	56.40%
Fiber laser	0.4155	0.5849	40.77%

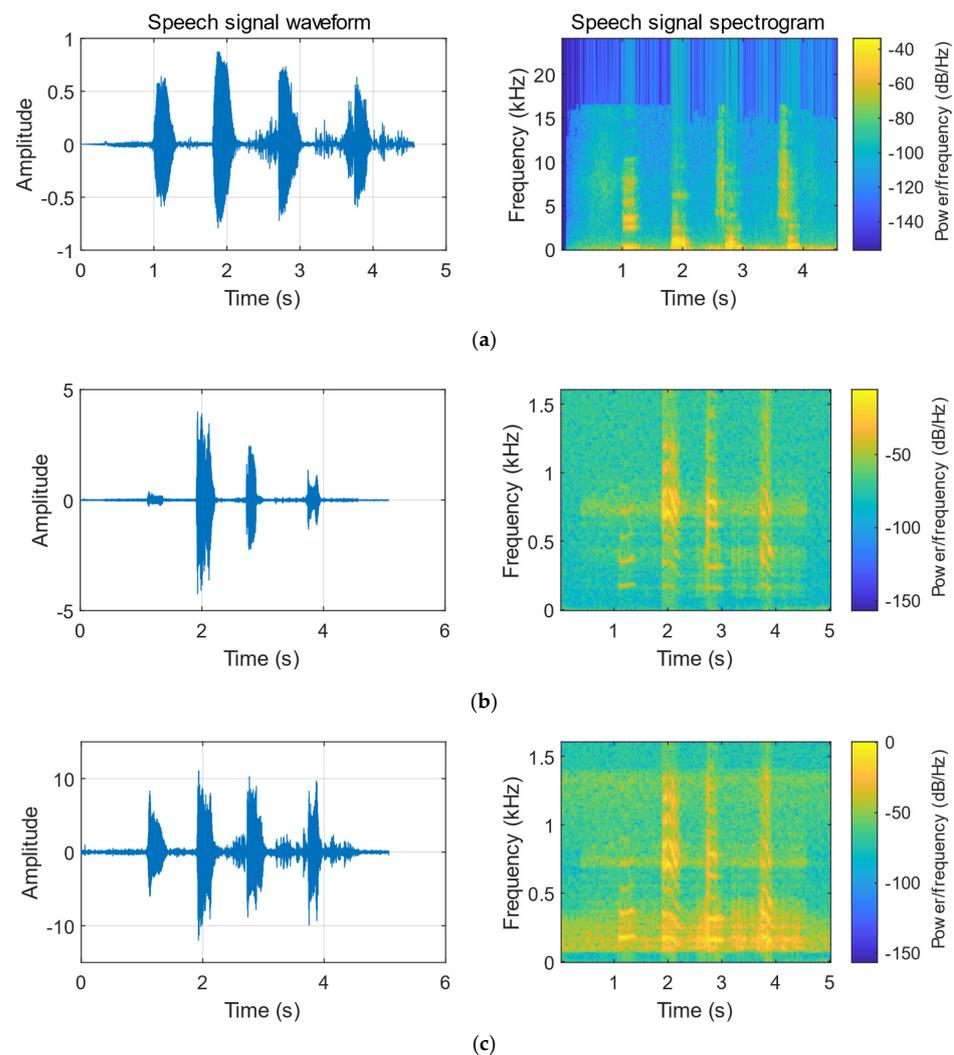


Figure 14. Comparison of speech signals reconstructed using the He-Ne Laser. (a) Original speech signal. (b) Speech signal reconstructed by DIC. (c) Speech signal reconstructed by DIC + speech enhancement.

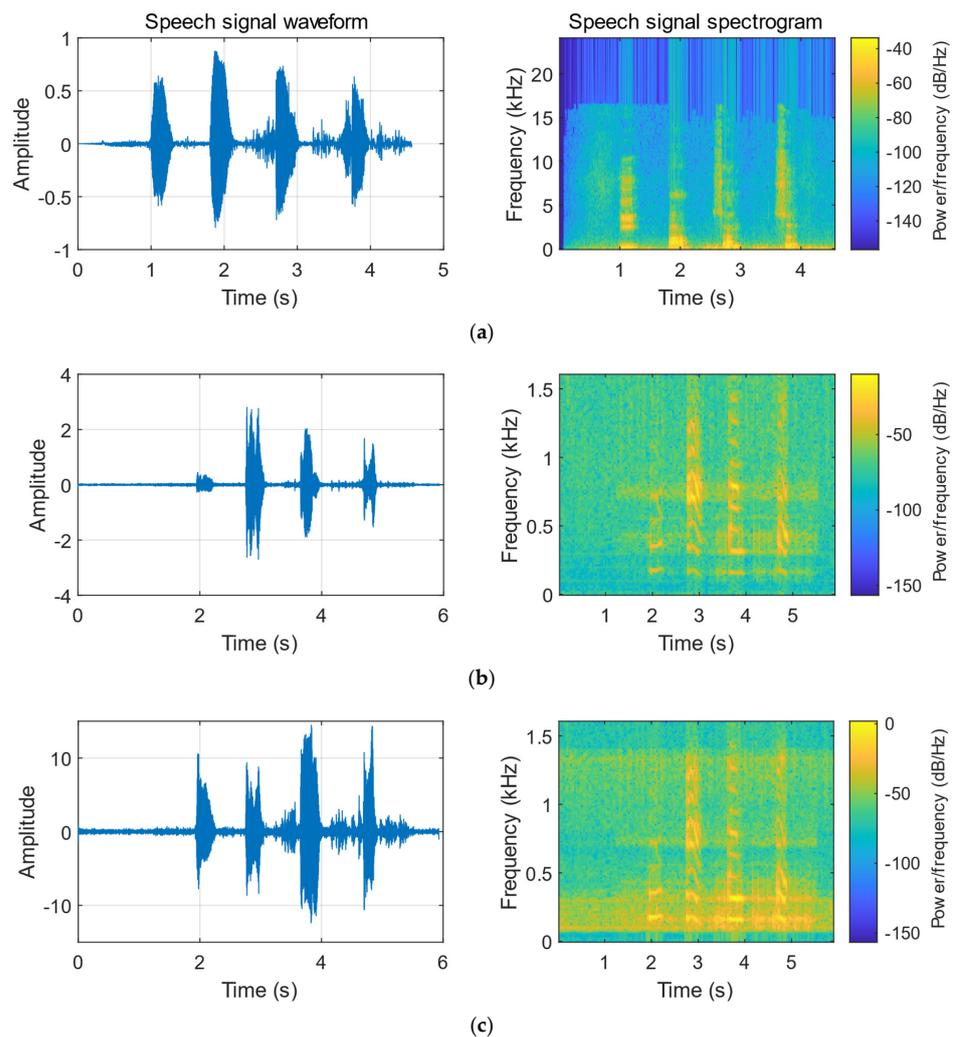


Figure 15. Comparison of speech signals reconstructed using the fiber laser. (a) Original speech signal. (b) Speech signal reconstructed by DIC. (c) Speech signal reconstructed by DIC + speech enhancement.

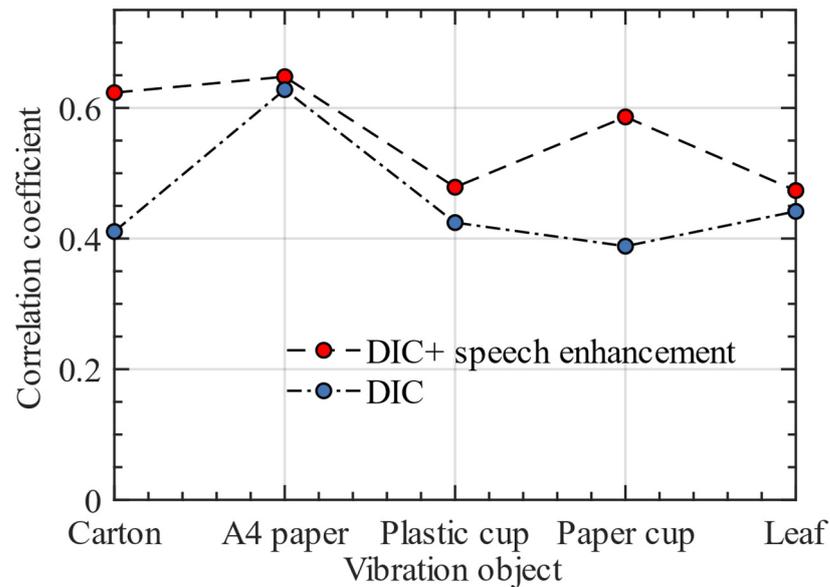
5.3. Performance on Five Vibration Object Datasets

Table 9 shows the difference in frequency spectrum correlation coefficient between the reconstructed speech signal and the original speech using DIC and DIC + speech enhancement. The results show that after using the speech enhancement algorithm, the accuracy of speech signals reconstructed from different vibration objects is improved. This further proves that our algorithm has strong adaptability to different environments and has the practical application value.

The results in Table 9 show that the accuracy of speech reconstruction for some datasets is improved slightly after using speech enhancement algorithm. We analyzed the reasons for this phenomenon as follows: when collecting these datasets, the surrounding environment has little interference, which ultimately makes the accuracy of speech reconstruction high. This can also be proved by the fact that the frequency spectrum correlation coefficient before speech enhancement is higher than that of other datasets. Figure 16 shows the line graph of the frequency spectrum correlation coefficient of reconstructed speech signals on different vibration objects.

Table 9. Comparison of the speech reconstruction accuracies of five vibration object datasets.

Vibration Object	DIC	DIC + Speech Enhancement	Improve
Carton	0.4107	0.6234	51.79%
A4 paper	0.6280	0.6478	3.15%
Plastic cup	0.4244	0.4786	12.77%
Paper cup	0.3883	0.5862	50.97%
Leaf	0.4416	0.4735	7.22%

**Figure 16.** Accuracy comparison of speech signal reconstruction with and without enhancement.

5.4. Discussions

Our speech enhancement algorithm is the speech process after the speech reconstruction algorithm, so whether the algorithm can be used in real time depends on the speed of the speech reconstruction algorithm. In the following content, we discuss some key factors that affect the performance of speech enhancement algorithm:

- (1) Loudspeaker performance: In order to provide a benchmark for frequency response, we use a sinusoidal signal which frequency varies from 80 to 1600 Hz and amplitude remains constant at 1 as the audio at the sound source. However, there will be errors in the sinusoidal signal played by the loudspeaker. These errors will be fewer if the loudspeaker performance is better.
- (2) Environmental noise and platform vibration: In the experiment, we find that the environmental noise and the vibration of the experimental platform seriously interfere with the sinusoidal signal reconstruction, and then affect the frequency spectrum correlation coefficient between the original sinusoidal signal and the reconstructed signal. The reconstructed sinusoidal signal contains environmental noise and vibration of the experimental platform, but the original signal is a clean and impurity-free speech, and the correlation coefficient between the two is low. Therefore, a quiet environment and a stable experimental platform are conducive to the verification of experimental results.

6. Conclusions

The main innovation of this paper is to propose a speech enhancement algorithm to solve the frequency response problem of speech signal reconstruction from laser speckle images. The experimental results in this paper show that our speech enhancement algorithm can effectively improve the accuracy of reconstructed speech, and the comparison of

the reconstruction accuracy of different vibration object datasets verifies the practicability of the algorithm. In the experimental results, we found that when the ambient noise is small and the vibration interference of the experimental platform is small, the reconstructed speech accuracy is high. Our algorithm achieves a major advance of reconstructing speech from laser speckle images.

Although our speech enhancement algorithm has an obvious effect on reducing the frequency response problem, it also introduces some noise signals while increasing the accuracy of the reconstructed speech. In order to obtain higher-definition speech signals, we also need to make further improvements to the speech enhancement algorithm. In the future, we will focus on reconstructing speech with a small frequency response and high definition.

Author Contributions: Conceptualization, H.Z.; Data curation, X.H.; Formal analysis, X.H.; Funding acquisition, D.Z. and H.Z.; Investigation, X.H.; Methodology, X.H.; Project administration, H.Z.; Resources, H.Z.; Software, X.H.; Supervision, D.Z., X.W., L.Y. and H.Z.; Validation, L.Y. and H.Z.; Visualization, X.H.; Writing—original draft, X.H.; Writing—review & editing, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Clavel, C.; Ehrette, T.; Richard, G. Events Detection for an Audio-Based Surveillance System. In Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, 6 July 2005; pp. 1306–1309.
2. Zieger, C.; Brutti, A.; Svaizer, P. Acoustic Based Surveillance System for Intrusion Detection. In Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova, Italy, 2–4 September 2009; pp. 314–319.
3. Castellini, P.; Martarelli, M.; Tomasini, E.P. Laser doppler vibrometry: Development of advanced solutions answering to technology's needs. *Mech. Syst. Signal Process.* **2006**, *20*, 1265–1285. [[CrossRef](#)]
4. Li, R.; Wang, T.; Zhu, Z.; Xiao, W. Vibration characteristics of various surfaces using an LDV for long-range voice acquisition. *IEEE Sens. J.* **2010**, *11*, 1415–1422. [[CrossRef](#)]
5. Li, R.; Madampoulos, N.; Zhu, Z.; Xie, L. Performance comparison of an all-fiber-based laser Doppler vibrometer for remote acoustical signal detection using short and long coherence length lasers. *Appl. Opt.* **2012**, *51*, 5011–5018. [[CrossRef](#)] [[PubMed](#)]
6. Rothberg, S.J.; Allen, M.S.; Castellini, P.; Maio, D.D.; Dirckx, J.; Ewins, D.J.; Halkon, B.J.; Muysshondt, P.; Paone, N.; Ryan, T. An international review of laser doppler vibrometry: Making light work of vibration measurement. *Opt. Lasers Eng.* **2017**, *99*, 11–22. [[CrossRef](#)]
7. Wu, S.S.; Lv, T.; Han, X.Y.; Yan, C.H.; Zhang, H.Y. Remote audio signals detection using a partial-fiber laser Doppler vibrometer. *Appl. Acoust.* **2018**, *130*, 216–221. [[CrossRef](#)]
8. Garg, P.; Nasimi, R.; Ozdagli, A.; Zhang, S.; Mascarenas, D.; Taha, M.R.; Moreu, F. Measuring transverse displacements using unmanned aerial systems laser doppler vibrometer (UAS-LDV): Development and field validation. *Sensors* **2020**, *20*, 6051. [[CrossRef](#)]
9. Matoba, O.; Inokuchi, H.; Nitta, K.; Awatsuji, Y. Optical voice recorder by off-axis digital holography. *Opt. Lett.* **2014**, *39*, 6549–6552. [[CrossRef](#)]
10. Ishikawa, K.; Tanigawa, R.; Yatabe, K.; Oikawa, Y.; Onuma, T.; Niwa, H. Simultaneous imaging of flow and sound using high-speed parallel phase-shifting interferometry. *Opt. Lett.* **2018**, *43*, 991–994. [[CrossRef](#)]
11. Bianchi, S. Vibration detection by observation of speckle patterns. *Appl. Opt.* **2014**, *53*, 931–936. [[CrossRef](#)]
12. Davis, A.; Rubinstein, M.; Wadhwa, N.; Mysore, G.J.; Durand, F.; Freeman, W.T. The visual microphone: Passive recovery of sound from video. *ACM Trans. Graph.* **2014**, *33*, 1–10. [[CrossRef](#)]
13. Peters, W.H.; Ranson, W.F. Digital imaging techniques in experiment stress analysis. *Opt. Eng.* **1982**, *21*, 427–431. [[CrossRef](#)]
14. Yamaguchi, I. Advances in the laser speckle strain gauge. *Opt. Eng.* **1988**, *27*, 214–218. [[CrossRef](#)]
15. Zhu, G.; Yao, X.R.; Qiu, P.; Mahmood, W.; Yu, W.K.; Sun, Z.B.; Zhai, G.J.; Zhao, Q. Sound recovery via intensity variations of speckle pattern pixels selected with variance-based method. *Opt. Eng.* **2018**, *57*, 026117. [[CrossRef](#)]
16. Wu, N.; Haruyama, S. The 20k samples-per-second real time detection of acoustic vibration based on displacement estimation of one-dimensional laser speckle images. *Sensors* **2021**, *21*, 2938. [[CrossRef](#)] [[PubMed](#)]

17. Goodman, J.W. *Speckle Phenomena in Optics: Theory and Application*; Roberts and Company: Placerville, CA, USA, 2007.
18. Zalevsky, Z.; Beiderman, Y.; Margalit, I.; Gingold, S.; Teicher, M.; Mico, V.; Garcia, J. Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern. *Opt. Express* **2009**, *17*, 21566–21580. [[CrossRef](#)] [[PubMed](#)]
19. Hua, T.; Xie, H.; Wang, S.; Hu, Z.; Chen, P.; Zhang, Q. Evaluation of the quality of a speckle pattern in the digital image correlation method by mean subset fluctuation. *Opt. Laser Technol.* **2011**, *43*, 9–13. [[CrossRef](#)]
20. Blaber, J.; Adair, B.; Antoniou, A. Ncorr: Open-source 2d digital image correlation matlab software. *Exp. Mech.* **2015**, *55*, 1105–1122.
21. Li, L.; Gubarev, F.A.; Klenovskii, M.S.; Bloskhina, A.I. Vibration measurement by means of digital speckle correlation. In Proceedings of the 2016 International Siberian Conference on Control and Communications (SIBCON), Moscow, Russia, 12–14 May 2016; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2016; pp. 1–5.
22. Hu, W.; Miao, H. Sub-pixel displacement algorithm in temporal sequence digital image correlation based on correlation coefficient weighted fitting. *Opt. Lasers Eng.* **2018**, *110*, 410–414. [[CrossRef](#)]
23. Duadi, D.; Ozana, N.; Shabairou, N.; Wolf, M.; Zalevsky, Z.; Primov-Fever, A. Non-contact optical sensing of vocal fold vibrations by secondary speckle patterns. *Opt. Express* **2020**, *28*, 20040–20050. [[CrossRef](#)]
24. Liushnevskaya, Y.D.; Gubarev, F.A.; Li, L.; Nosarev, A.V.; Gusakova, V.S. Measurement of whole blood coagulation time by laser speckle pattern correlation. *Biomed. Eng.* **2020**, *54*, 262–266. [[CrossRef](#)]
25. Wu, N.; Haruyama, S. Real-time sound detection and regeneration based on optical flow algorithm of laser speckle images. In Proceedings of the 2019 28th Wireless and Optical Communications Conference (WOCC), Beijing, China, 9–10 May 2019; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2019; pp. 1–4.
26. Wu, N.; Haruyama, S. Real-time audio detection and regeneration of moving sound source based on optical flow algorithm of laser speckle images. *Opt. Express* **2020**, *28*, 4475–4488. [[CrossRef](#)] [[PubMed](#)]
27. Chen, Z.; Wang, C.; Huang, C.; Fu, H.; Luo, H.; Wang, H. Audio signal reconstruction based on adaptively selected seed points from laser speckle images. *Opt. Commun.* **2014**, *331*, 6–13. [[CrossRef](#)]
28. Zhu, G.; Yao, X.-R.; Sun, Z.-B.; Qiu, P.; Wang, C.; Zhai, G.-J.; Zhao, Q. A High-Speed Imaging Method Based on Compressive Sensing for Sound Extraction Using a Low-Speed Camera. *Sensors* **2018**, *18*, 1524. [[CrossRef](#)]
29. Heikkinen, J.; Schajer, G.S. A geometric model of surface motion measurement by objective speckle imaging. *Opt. Lasers Eng.* **2020**, *124*, 105850. [[CrossRef](#)]
30. Zhu, D.; Yang, L.; Li, Z.; Zeng, H. Remote speech extraction from speckle image by convolutional neural network. In Proceedings of the 2020 IEEE Symposium on Computers and Communications (ISCC), Rennes, France, 7 July 2020; pp. 1–6.
31. Ma, C.; Ren, Q.; Zhao, J. Optical-numerical method based on a convolutional neural network for full-field subpixel displacement measurements. *Opt. Express* **2021**, *29*, 9137–9156. [[CrossRef](#)] [[PubMed](#)]
32. Liu, C.; Li, L.; Huang, X.; Wang, X.; Wang, C. Audio signal extraction and enhancement based on cnn from laser speckles. *IEEE Photonics J.* **2021**, *14*, 1–5. [[CrossRef](#)]
33. Zhu, D.; Yang, L.; Zeng, H. Remote recovery of sound from speckle pattern video based on convolutional LSTM. In Proceedings of the International Conference on Information and Communications Security (ICICS), Chongqing, China, 29 November 2021; pp. 110–124.
34. Barcellona, C.; Halpaap, D.; Amil, P.; Buscarino, A.; Fortuna, L.; Tiana-Alsina, J.; Masoller, C. Remote recovery of audio signals from videos of optical speckle patterns: A comparative study of signal recovery algorithms. *Opt. Express* **2020**, *28*, 8716–8723. [[CrossRef](#)]
35. Boas, D.A.; Dunn, A.K. Laser speckle contrast imaging in biomedical optics. *J. Biomed. Opt.* **2010**, *15*, 011109. [[CrossRef](#)]
36. Bi, R.; Du, Y.; Singh, G.; Ho, J.H.; Zhang, S.; Attia, A.B.E.; Li, X.; Olivo, M.C. Fast pulsatile blood flow measurement in deep tissue through a multimode detection fiber. *J. Biomed. Opt.* **2020**, *25*, 055003. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.