Optics Communications 440 (2019) 117–125

Contents lists available at ScienceDirect

Optics Communications

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

The effect of speckles noise on the Laser Doppler Vibrometry for remote speech detection

Tao Lv [a,b,*], Xiyu Han [a,b], Shisong Wu [a,b], Yuanyang Li [a]

[a] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Keywords:
Laser Doppler Vibrometry
Speckles noise
Speech detection

ABSTRACT

Speckles noise generated by the relative in-plane motion between the Laser Doppler Vibrometry (LDV) and the target damages the quality of the LDV-captured speech severely. As such, it is necessary to research on the effect of speckles noise on the LDV for remote voice detection and remove it. In this paper, we comprehensively discuss the statistical properties of dynamic speckles with the combination of theories and experiments, and the results show that the fluctuation of speckles intensity depends strongly on the relative velocity for a given value of the detection distance. Then, we investigate the effects of the speckles noise on the LDV for remote audio detection, and the results indicate that speckles noise can cause spikes and bursts appearing in the LDV-captured voice signals, which degrades the quality of the speech. To address this issue, an algorithm of the speckle noise reduction in the LDV-measured audio signals is proposed and the experimental results demonstrate that the proposed algorithm performs well.

1. Introduction

Remote acoustic signals detection is receiving a growing interest by the scientific community. It can be used for surveillance, such as intrusion detection, abnormal situations detection in public places and underwater acoustic environment monitoring [1–4]. Whereas, because of the quick drop of the speech pressure as it propagates through the air, the working distances of the traditional acoustic sensor is short. The microphone array can be used for long-range voice detection, but it is sensitive to the environment noise, and the detection distance is still very limited [5,6]. Furthermore, these types of sensors need to be fixed at pre-determined locations, if the targets move out of the detection range, they will not obtain any signals.

With high-precision, non-contact, high spatial resolution and wide dynamic range, Laser Doppler Vibrometry (LDV) can measure the displacement and velocity of remote solid objects precisely [7–11]. Because of the advantages mentioned above, LDV offers an effective method alternative to traditional sensing methods such as accelerometers, and it has a wide range of application in commercial, industrial and biomedicine fields. Furthermore, the voice signal can cause vibration of any surrounding objects, although these vibrations are usually at a micrometer or nanometer scale. So, similar to the microphone to obtain the sound signals by detecting the microphone membrane vibration, the LDV can be used to detect the objects vibration caused by the speech of the person close to the target to acquire voice signals. Some studies using LDV for remote speech detection have been published. For instance, Li, Wang [12–17] and et al. have presented their results in detecting and processing voice signals of people from large distances using an LDV, Peng [18] and Yekutiel [19] used the LDV to enhance the quality of speech, and the LDV is used by Deng [20] to identify long range speaker. In previous paper, we also developed a system which consists of an LDV and a PTZ camera to detect human activities [21].

Despite the LDV has ability to detect long range speech, the voice signal quality acquired by the LDV is mainly determined by the surface properties of the selected object nearby the target. So, Rui Li [22] has investigated the effect of the vibration characteristics of targets on LDV remote voice acquisition, and Qu [16] proposed a target surface selection and automatic laser focusing method to improve the LDV's performance for the long-range detection. However, if there is relative motion between the selected target and LDV due to human activity or environmental factors, the LDV-measurement audio may be distorted by dynamic speckles noise. Many studies have attempted to account for the effect of speckles on the LDV system, such as the LDV uncertainty limit due to speckle noise is studied in [23] and the prediction of the vibration amplitude noise floor due to speckle at low frequencies is given by Rice [24,25], Hill [26], Jiang [27] and Dräbenstedt [28]. However, their studies focused on the effect of LDV vibration measurement, and the effect of dynamic speckles noise on the LDV for remote speech detection has not been studied.

In this paper, we present several exact formulations based on Gaussian noise and Speckles theory, and use those to evaluate the effect of

* Corresponding author.
E-mail address: 18767120269@163.com (T. Lv).

https://doi.org/10.1016/j.optcom.2019.02.014
Received 3 January 2019; Received in revised form 24 January 2019; Accepted 4 February 2019
Available online 15 February 2019

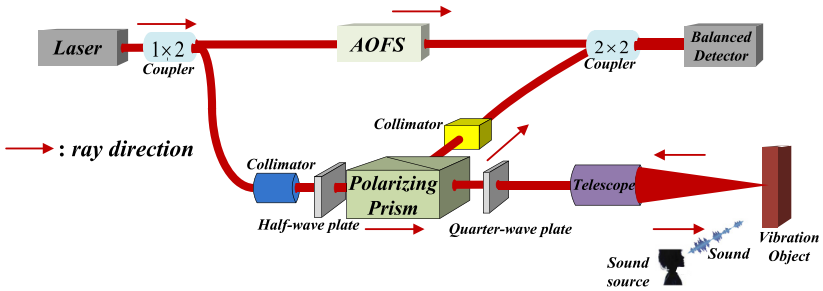

Fig. 1. The diagram of the transceiver unit. The AOFS is a Gooch&Housego T-M040-0.5C8J-3-F2P. The collimator is a Thorlabs PAF-X-2-C. The balanced detector is a Thorlabs PDB570C.

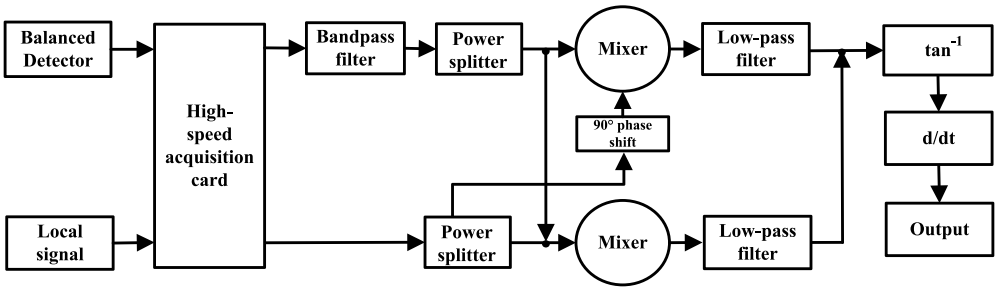

Fig. 2. The diagram of the signal processing unit.

speckles noise on the LDV system for remote speech detection. Besides, a speckles noise removal algorithm is proposed to remove the speckles noise from LDV-captured speech, and the performance superiority of the proposed approach is demonstrated.

The rest of the paper is organized as follows: Section 2 briefly introduces the fundamentals of the LDV and its use for voice detection. In Section 3, the properties of dynamic speckles are discussed theoretically and experimentally. The effect of speckles on the LDV system for remote speech collection, from theoretical analyzed and numerical simulated is presented in Section 4. The speckle noise removal algorithm is presented in Section 5. Concluding remarks are given in Section 6.

2. Long-range speech measurement with LDV

Because the LDV is a relatively new modality for the speech collection, we give a short introduction of LDV system in this section.

The LDV used to detect remote audio is composed of transceiver unit and signal processing unit. The diagram of the transceiver unit is shown in Fig. 1. A 20-mW single-frequency (1550 nm) fiber laser with single longitudinal mode and narrow linewidth (less than 10 kHz) is used as the transmitter. An emergent beam from the transmitter is divided into two beams by a 1 × 2 single-mode fiber coupler. One acted as the local oscillator (LO) beam, and the other acted as the transmitted beam. In order to discriminate the direction of target vibration, an acousto-optic frequency shifter (AOFS) is applied to shift the frequency of LO beam 40 MHz offset. The transmitted beam is focused on the target after passing through a half wave plate, a polarization prism, a quarter-wave prism and a zooming telescope. For ultra-long standoff range applications, the long-range focal telescope can be used. Due to the vibration of the target caused by the speech pressure, the backscattered beam carries a Doppler frequency shift $\Phi_d(t)$, which is proportional to the vibration displacement x of the target according to $\Phi_d(t) = 2ks = (4\pi/\lambda)s$. The backscattered beam is received by the same telescope and mixed with LO beam to feed into the balanced detector. Noting that the polarizing prism is used to isolate transmitted beam and backscattered beam.

The output of the balanced photodetector I is given by the time average of the square of the total light amplitude resulting from the combination of the LO and backscattered beams. In order to obtain the acoustic signal, a signal processing unit (as shown in Fig. 2) is needed. The signal processing unit contains a dual-channel high-speed acquisition card (16 bits and 200 M/s) and an in-phase quadrature (IQ) demodulator [29]. First the detector output signal I and local oscillator are sampled by the data acquisition card. Because the center frequency of the balanced photodetector I is 40 MHz, the sampling rate of 200 M/s can ensure that the digital signal completely retains the information in the balanced photodetector I after sampling. Then, the Doppler frequency shift $\Phi_d(t)$ can be acquired by the IQ demodulator. In the end, according to the Doppler frequency shift $\Phi_d(t)$, the speech (vibration displacement x) can be acquired.

3. The properties of dynamic speckles caused by relative in-plane motion

In order to analyze the impact of dynamic speckles noise on the LDV for remote speech detection, it is necessary to consider the properties of dynamic speckles. Because the relative motion between the selected target and LDV which may be caused by human activity or environment factors, is mostly in-plane motion or tilt rather than rotation, we only consider the effect of the translational speckles noise and the boiling speckles noise can be neglected. The propagation process of spot from a diffuse target to the receiving system is illustrated in Fig. 3. We assume that the target has an in-plane velocity $V_\perp$, the distance between the diffuse target and the LDV is R, and the Gaussian beam waist ω_0 is on the target (the laser is focused on the target surface). It is worth noting that if the transceiver system as shown in Fig. 1 is a perfect optical system, the relation between Gaussian beam waist radius ω_0 and the focusing distance R is linear.

In reality, the detection distance satisfies the far field condition, so the detected speckle complex amplitude $E(X, t)$ can be written as:

$$E(X,t) = \frac{1}{i\lambda R}\exp(ikR)\exp\left(i\frac{k}{2R}X^2\right)\int_{-\infty}^{+\infty} E_0(\zeta) \times \exp\left(i\phi\left(\zeta - V_\perp t\right)\right) e\left(-ik\frac{\zeta X}{R}\right) d\zeta \quad (1)$$

where the wave vector is denoted by $k = 2\pi/\lambda$, $\Phi(\zeta)$ is the incident light phase at the detector, and the amplitude of the incident light $E_0(\zeta)$ is

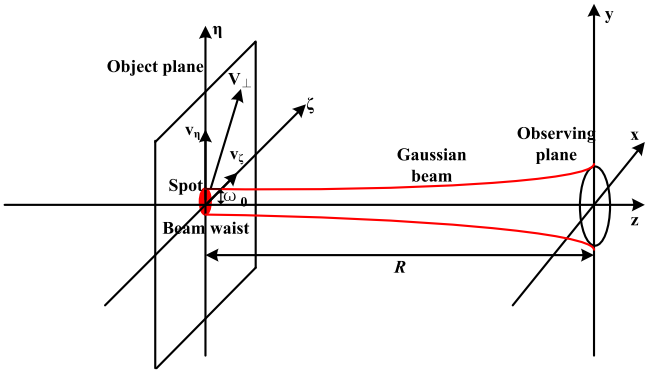

Fig. 3. Transmission model of speckle field.

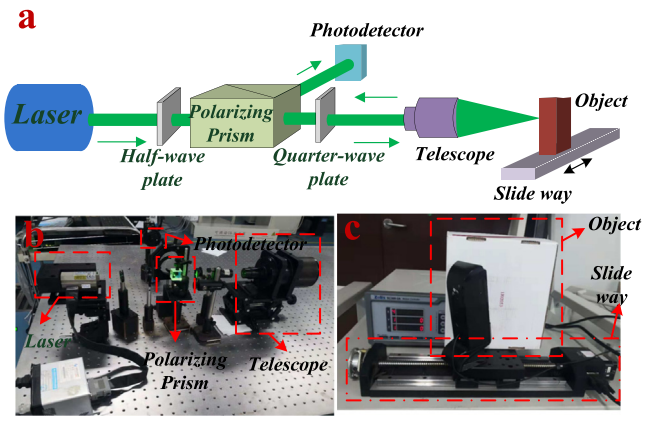

Fig. 4. (a) Experimental setup for detecting speckle intensity. (b) Prototype of the experimental system. (c) Target.

expressed by

$$E_0(\zeta) = \exp\left(-\frac{\zeta^2}{{\omega_0}^2}\right)\exp\left(-i\varphi_{add}\right) \tag{2}$$

where φ_{add} is an additional phase, which can be ignored.

Let the fluctuation of the speckle intensity from the mean value be defined as $\Delta(IX,t) = I(X,t) - \langle I(X,t)\rangle$, where $\langle\ldots\rangle$ stands for an ensemble average. Since the complex amplitude $E(X,t)$ is assumed to obey circular Gaussian statistics with zero mean under the central limiting theorem, the correlation function of the intensity fluctuation $\Gamma_{\Delta I}(X,t_1,t_2)$ is expressed by using the second-order correlation with respect to the complex amplitude and is given by

$$\begin{aligned}\Gamma_{\Delta I}(X,t_1,t_2) &= \left\langle \Delta I(X,t_1)\,\Delta I(X,t_2)\right\rangle = \left|\left\langle E(X,t_1)\,E^*(X,t_2)\right\rangle\right|^2 \\ &= \left|\Gamma_E(X,t_1,t_2)\right|^2\end{aligned} \tag{3}$$

By using (1) and (3), the time correlation function of the complex amplitude $E(X, t)$ becomes

$$\begin{aligned}\Gamma_E(X,t_1,t_2) &= \frac{-1}{\lambda^2 R_1 R_2}\exp\left(ik\left(R_1 - R_2\right)\right) \\ &\times \exp\left(\frac{iX^2}{2}\left(\frac{1}{R_1}-\frac{1}{R_2}\right)\right)\iint_{\infty} E_0(\zeta_1,z_1)\,E^*{}_0(\zeta_2,z_2) \\ &\times\left\langle\exp\left\{i\left[\phi(\zeta_1 - V_\perp t_1) - \phi(\zeta_2 - V_\perp t_2)\right]\right\}\right\rangle \\ &\times \exp\left(-ikX\left(\frac{\zeta_1}{R_1}-\frac{\zeta_2}{R_2}\right)\right)d\zeta_1 d\zeta_2\end{aligned} \tag{4}$$

In Eq. (4), $\langle exp(i[\Phi(\zeta_1 - V_\perp t_1) - \Phi(\zeta_2 - V_\perp t_2)])\rangle$ can be simplified into $\Delta S\ \delta[\zeta_1 - \zeta_2 - V_\perp(t_1 - t_2)]$ [17]. Where ΔS is the correlation area of $\Phi(\zeta)$, and in far-field situations, $1/R_1 - 1/R_2$, $k(R_1 - R_2)$ and $1/(\lambda^2 R_1 R_2)$ are similar to constants. In practice, the receiving system is located in the center, so X = 0, and Eq. (4) can simplify to

$$\begin{aligned}\Gamma_E(0,t_1,t_2) &= \Delta S\iint_{\infty} E_0\left(\zeta + \frac{1}{2}V_\perp(t_1 - t_2), z\right) \\ &\times E^*{}_0\left(\zeta - \frac{1}{2}V_\perp(t_1 - t_2), z\right)d\zeta\end{aligned} \tag{5}$$

By using the (3) and (5), the normalized correlation function of the intensity fluctuation $\gamma_{\Delta I}$ (in the far-field) takes the form [30,31]

$$\gamma_{\Delta I}(\tau) = \frac{\Gamma_{\Delta I}(\tau)}{\Gamma_{\Delta I}(0)} = \exp\left(-\frac{\tau^2}{\tau_{c\perp}^2}\right) \tag{6}$$

where $\tau_{c\perp}$ represents the correlation lengths caused by the relative motion, which is equivalent to the time taken for the speckle pattern to advance by one speckle length

$$\tau_{c\perp} = \frac{\omega_0}{|V_\perp|} \tag{7}$$

The formula 6 and 7 are confirmed by an experiment, and the schematic is shown in Fig. 4. In order to simulate relative motion, a beam of laser is focused on the moving object (moving perpendicular to the optical axis) with the reflected light detected by a photodetector. And the normalized correlation function of the intensity fluctuation can be obtained by autocorrelated the output of the photodetector, which has removed the mean.

The results are shown in Fig. 5. We can see that the correlation length $\tau_{c\perp}$ decreases with the increasing velocity $V_\perp$, and the measured normalized correlation function of the intensity fluctuation curves can be fitted by Eq. (6).

4. The effect of dynamic speckles on the LDV system for remote speech detection

4.1. The spectral power density of speckles noise

To consider the effect of the speckles on the LDV system, it is necessary to analyze the output of the instrument photodetector.

If there are M speckles incident on the detector, the total scattered electromagnetic field falling on the detector can be represented as

$$E_s = e^{i2\pi(ft+\phi_d(t))}\sum_{j=1}^{M}\alpha_j e^{i\varphi_j} \tag{8}$$

where f is the optical frequency and f_d is the doppler frequency shift caused by vibration of the target. Assumed that the doppler shift is modulated by a pure tone, the $\Phi_d(t)$ can be written as

$$\phi_d(t) = \frac{4\pi s}{\lambda}\cos\left(2\pi f_s t\right) \tag{9}$$

where s is the amplitude of target vibration and f_s is the frequency of the target vibration.

The individual scattering amplitudes α and phases φ are statistically independent random variables. According to [24], the Eq. (9) can be represented by the product

$$E_s = \varepsilon e^{i[2\pi(ft+\phi_d(t))+\theta]} \tag{10}$$

where ε and θ are the amplitude and phase of a complex Gaussian process.

In heterodyne detection, the scattered electromagnetic field and the local oscillator field are combined at the photodetector, and the photodetector output that removes the dc term can be expressed as follows

$$I = \beta\varepsilon e^{i[2\pi(f_{AOFS}t+\phi_d(t))+\theta]} \tag{11}$$

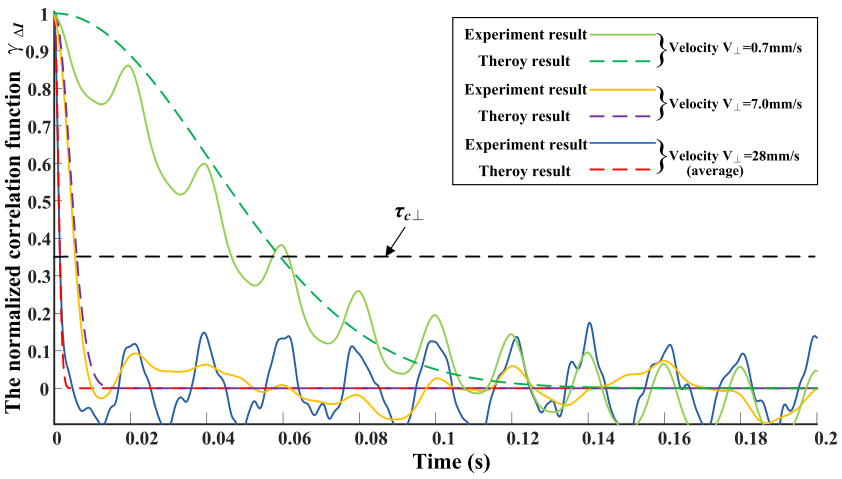

Fig. 5. The normalized correlation function of the intensity fluctuation $\gamma_{\Delta I}$ under different velocity.

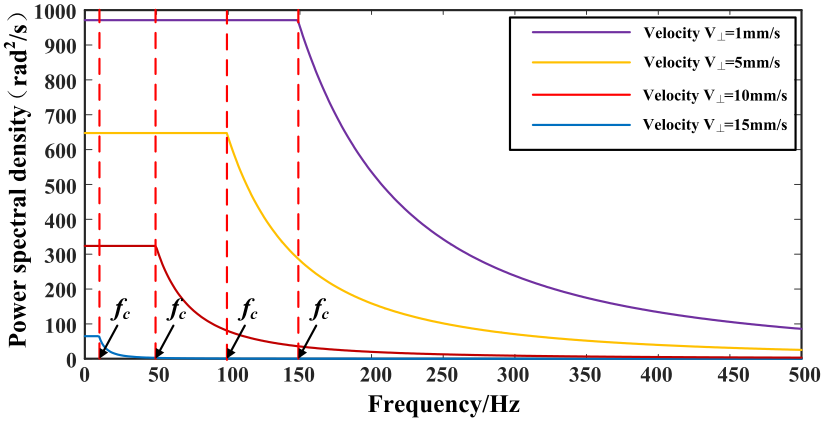

Fig. 6. Speckles noise spectral power density. The parameter used to generate these curves is ω0=0.05 mm.

where f_{AOFS} is the frequency difference between the scattered field and the local oscillator field, which caused by the AOFS.

The output of the I&Q demodulator can be expressed as

$$O = \dot{\phi}_d(t) + \dot{\theta} \tag{12}$$

where $\dot{\theta}$ is the phase derivative of the speckle noise. Thus, the speckles noise is converted from a multiplicative noise at the input to an additive noise at the output of the IQ demodulator.

A simple measure of the performance of the demodulator can be obtained through evaluation of the power spectrum of its output, and the power spectrum can be evaluated from the autocorrelation function. The correlation function can be written as

$$\langle O(t)\, O(t+\tau)\rangle = \left\langle \dot{\theta}(t)\, \dot{\theta}(t+\tau)\right\rangle + \frac{8\pi^2 s^2 f_s^2}{\lambda^2}\cos\left(2\pi f_s \tau\right) \tag{13}$$

where the first term on the right side of the Eq. (13) is the auto-correlation function of the speckle noise, and the second term is the autocorrelation function of the Doppler frequency shift.

The speckle is a Gaussian random process, and the phase θ of a complex Gaussian process with real and imaginary components X_{sp}, Y_{sp} is defined as the arctangent of the ratio Y_{sp}/X_{sp}. The correlation function

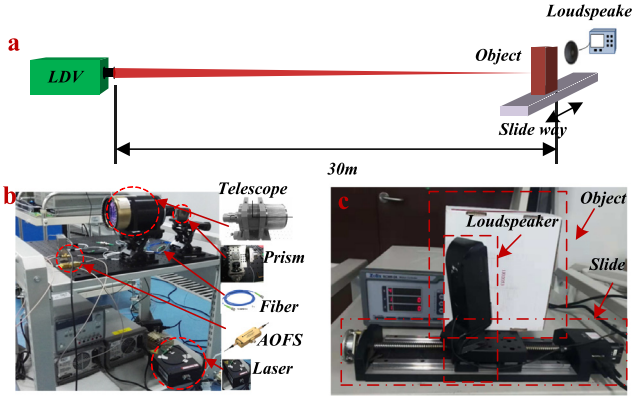

Fig. 7. (a) Experimental setup for verifying the effect of speckles noise. (b) Prototype of the LDV system. (c) Target.

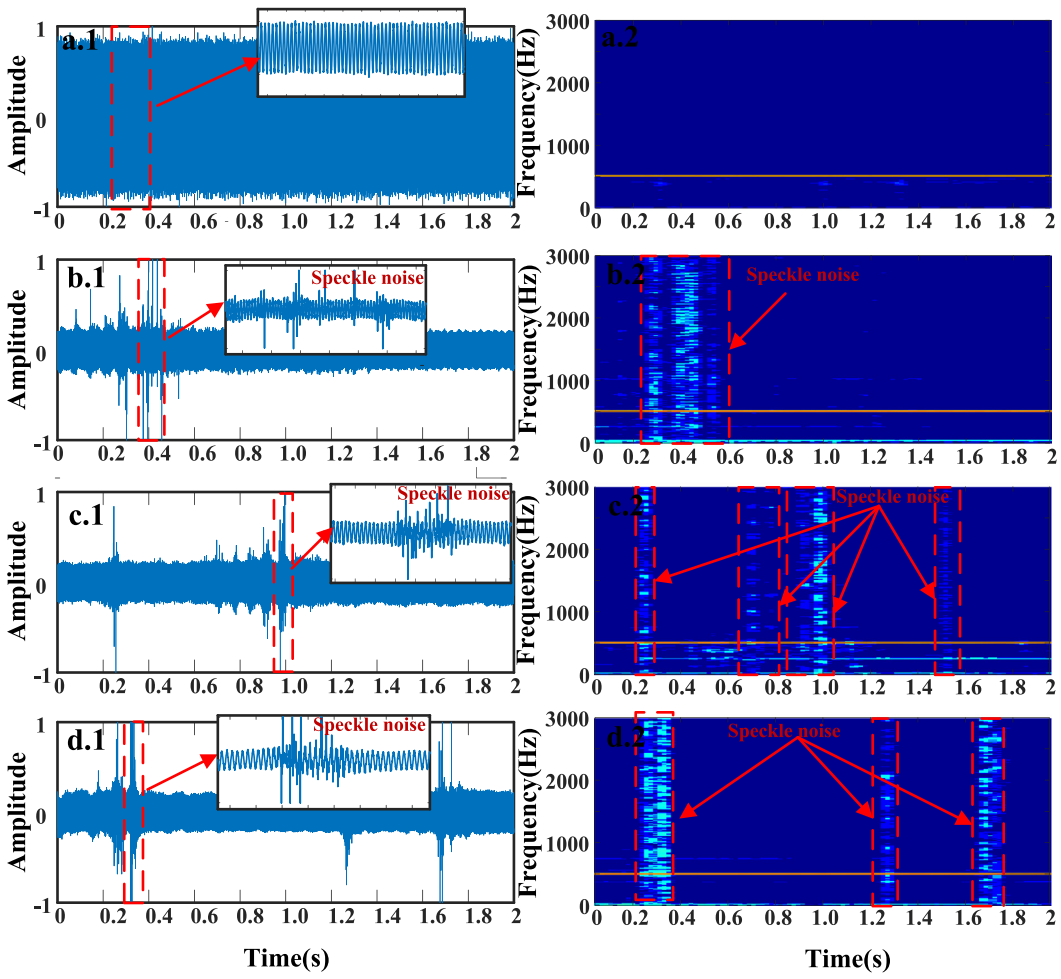

Fig. 8. The Speech waveforms and spectrograms of the LDV-captured single tone with different relative velocity. (a) velocity is zero. (b) Relative velocity is 2.33 mm/s. (c) Relative velocity is 4.66 mm/s. (d) Relative velocity is 7.00 mm/s.

of its derivative may thus be written [24,25]:

$$\begin{aligned}\left\langle \dot{\theta}(0)\,\dot{\theta}(\tau)\right\rangle &= \left\langle \left(\frac{X_{sp}(0)\,\dot{Y}_{sp}(0)-Y_{sp}(0)\,\dot{X}_{sp}(0)}{X_{sp}^2(0)+Y_{sp}^2(0)}\right)\right.\\ &\quad\left.\times\left(\frac{X_{sp}(\tau)\,\dot{Y}_{sp}(\tau)-Y_{sp}(\tau)\,\dot{X}_{sp}(\tau)}{X_{sp}^2(\tau)+Y_{sp}^2(\tau)}\right)\right\rangle\\ &= \left[\frac{g\ddot{g}+h\ddot{h}-\dot{g}^2-\dot{h}^2}{2q}+\frac{\left(\dot{g}h-\dot{h}g\right)^2}{q^2}\right]\ln(1-q)\\ &\quad+\frac{b^2+2b\left(\dot{g}h-\dot{h}g\right)+\left(\dot{g}h-\dot{h}g\right)^2/q}{(1-q)}\end{aligned}\tag{14}$$

where

$$\begin{aligned}b &= \left\langle X_{sp}(0)\,\dot{Y}_{sp}(0)\right\rangle = \left\langle X_{sp}(\tau)\,\dot{Y}_{sp}(\tau)\right\rangle\\ &= -\left\langle \dot{X}_{sp}(0)\,Y_{sp}(0)\right\rangle = -\left\langle \dot{X}_{sp}(\tau)\,Y_{sp}(\tau)\right\rangle\\ g &= \left\langle X_{sp}(0)\,X_{sp}(\tau)\right\rangle = \left\langle Y_{sp}(0)\,Y_{sp}(\tau)\right\rangle\\ h &= \left\langle X_{sp}(0)\,Y_{sp}(\tau)\right\rangle = -\left\langle X_{sp}(\tau)\,Y_{sp}(0)\right\rangle\\ q &= \left(g^2+h^2\right)/\left\langle X_{sp}^2\right\rangle\end{aligned}$$

In the special case in which the spectrum of O is symmetric, the Eq. (14) can be simplified to

$$\left\langle \dot{\theta}(0)\,\dot{\theta}(\tau)\right\rangle = \frac{1}{2}\left(\frac{\rho\ddot{\rho}-\dot{\rho}^2}{\rho^2}\right)\ln\left(1-\rho^2\right)\tag{15}$$

where, ρ is the correlation function of the speckle field. So, the ρ can be expressed by Eq. (6).

$$\rho = \exp\left(-\frac{\tau^2}{\tau_{c\perp}^2}\right)\tag{16}$$

According to Eq. (15), the speckle noise spectral power density S(f) in rad^2/s can be expressed by

$$S(f) = \int_0^\infty\left(\frac{\rho\ddot{\rho}-\dot{\rho}^2}{\rho^2}\right)\ln\left(1-\rho^2\right)e^{i\omega t}dt = 2\int_0^\infty\frac{\dot{\rho}^2}{1-\rho^2}e^{i\omega t}dt\tag{17}$$

After the speckle noise floor which is generated by the transverse motion between the target and LDV measured carefully, Dräbenstedt [28] found that below the cutoff frequency $f_c = 1/(2\tau_{c\perp})$, the speckle noise spectral density is constant. Whereas, beyond that cutoff frequency the spectral noise density falls off with *1/f*. So, the speckle noise spectral power density S(f) can be rewritten as

$$S(f) = \begin{cases} 2\displaystyle\int_0^\infty\frac{\dot{\rho}^2}{1-\rho^2}dt = \frac{3.27}{\tau_{c\perp}} = 3.27\frac{|V_\perp|}{\omega_0} & f \le f_c\\ \dfrac{2f_c{}^2}{f^2}\displaystyle\int_0^\infty\frac{\dot{\rho}^2}{1-\rho^2}dt = \frac{f_c{}^2}{f^2}\frac{3.27}{\tau_{c\perp}} = 3.27\frac{f_c{}^2}{f^2}\frac{|V_\perp|}{\omega_0} & f > f_c\end{cases}\tag{18}$$

The formula 18 is consistent with Rice [24,25], Hill [26], Jiang [27] (with a factor difference 2) and Dräbenstedt [28] (with a factor difference 4) at high CNR.

The functional form of the speckles noise spectral power density with different velocities (for a given Gaussian beam waist) is shown in Fig. 6, the speckle noise spectral power density increases with the increase of

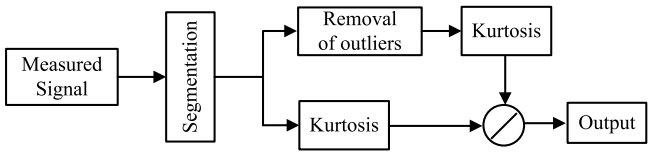

Fig. 9. Block diagram of the algorithm for speckle noise location.

the velocity of the relative motion. The knee of the speckles noise floor occurs at $f_c = 1/(2\tau_{c\perp})$, and the knee frequency increases as the velocity increases.

4.2. The effects of speckle noise

As shown in Fig. 6, The speckles noise is mainly located in the low frequencies region. Whereas, the frequency range of human voice is about 300 Hz to 3 kHz, the speckles noise may be introduced spreading across the voice frequency range. As shown in Eq. (12), due to the speckles noise, the phase of the Doppler signal changes discontinuously from $-\pi$ to π, which will cause spikes and bursts appearing in the output of demodulator signals. The spikes and bursts appear as sharp and harsh squeaks and pops in the voice signal, which have a great influence on the quality of the speech.

This phenomenon can be confirmed by a simple experiment (as shown in Fig. 7). This experiment is carried in a lab that has an unobstructed line-of-sight of 30 m. We placed the LDV at one end and the test target at the other. Because we can control the loudspeaker to generate any sound wave that we want, the loudspeaker is selected as a sound source, and the carton is selected as measurement target. To simulate the process of the relative motion, the loudspeaker and the carton are placed on the 1 DOF motorized rail. In this experiment, the sound source is the single tone with a frequency of 500 Hz.

As shown in Fig. 8.a, the LDV-captured voice is consistent with the sound source. Whereas, from Fig. 8.b to Fig. 8.d, we can find that when the target begins to move, the LDV-captured voice will begin to appear bursts, spikes and distortions, which creates sharp and harsh squeaks and pops when we listen to the LDV-captured voice (can be seen in from Fig. 8.b.2 to Fig. 8.d.2). Besides, it is worth noting that the relative velocity seems to promote the speckles noise. Hence, it is necessary to make sure the detection target without relative motion (in-plane) with the LDV. Whereas, in practice, it is difficult to find this type of target. Therefore, an algorithm for attenuating this noise is presented in Section 5

5. Speckles noise removal

Many speckles noise removal algorithms have been proposed, but they have been mainly used for vibration measurement [32,33], for instance, J. Vass [32] use a statistical indicator kurtosis ratio to select a portion which is undistorted by speckles noise. However, if this algorithm is directly applied to LDV-captured voice, it will cause some segments of voice lost. In this paper, we proposed a speckles noise removal algorithm to attenuate the speckles noise.

5.1. Speckles noise detection

Since the distribution of speckle noise appears to be random (see Fig. 8.b to Fig. 8.d), the speckle noise location must be confirmed before we remove it.

Motivated by the impulsive nature of speckle noise, a scalar indicator segment kurtosis ratio is used by us to determine the position of the speckle noise. This method was put forward by J. Vass [32] in 2008, and it is extended to speech signals. The block diagram of the speckle noise detection algorithm is depicted as Fig. 9.

First, the raw (measured) signal x is segmented by using a rectangular window.

Next, the lth segment trimmed signal x_t^l is achieved by thresholding the lth segment raw signal x^l

$$x_t^l = \left(thr_{low} < x^l[n] < thr_{high}\right) \tag{19}$$

where thr_{low} and thr_{high} are the lower and the upper threshold respectively. In other words, the trimmed signal x_t^l only consists of samples $x^l[n]$ satisfying formula 19.

Besides, the kurtosis of x_t^l and x^l are acquired by Eq. (20).

$$\beta_2(x) = \frac{\mu_4}{\left(\mu_2\right)^2} = \frac{E\left\{\left(x-\mu_x\right)^4\right\}}{\sigma_x^4} \tag{20}$$

where μ_x and σ_x are the mean value and the standard deviation of x respectively, and $\mathrm{E}\{\ldots\}$ stands for an ensemble average.

Finally, the lth segment kurtosis ratio KR^l is obtained by dividing the kurtosis of the x^l by the kurtosis of the x_t^l (Eq. (21)). If the KR^l exceeds the threshold thr_K, the x^l has speckle noise.

$$KR^l = \frac{\beta_2\left(x^l\right)}{\beta_2\left(x_t^l\right)} \tag{21}$$

The experimental results are shown in Fig. 10, we can obviously find that the location of speckle noise can be confirmed accurately by using the segment kurtosis ratio algorithm.

5.2. Speckles noise removal

Due to the correlation between the adjacent sampling points of voice signals, a method based on linear prediction is developed to remove the speckle noise. The linear prediction is first proposed by Wiener, and it is widely used in speech coding, speech synthesis and speech recognition [34]. The basic idea of linear prediction is to predict the values of the current or future sampling points by using the values of several adjacent past sampling points. The linear prediction can be expressed by

$$x(n) = \sum_{i=1}^{p} a_i x(n-i) + e(n) = \hat{x}(n) + e(n) \tag{22}$$

where, $\hat{x}(n)$ is the estimate of the $x(n)$, $e(n)$ is the linear prediction error, $\{a_i\}$ is the linear prediction coefficients, and p is the number of coefficients $\{a_i\}$. According to Eq. (22), if the coefficients $\{a_i\}$ are obtained, the speech can be estimated by using several adjacent past sampling point values. The coefficients $\{a_i\}$ can be obtained by the minimum mean square error criterion. The mean square error can be written as

$$E_x = \frac{1}{n}\sum_n e^2(n) = \frac{1}{n}\sum_n [x(n) - \hat{x}(n)]^2 = \frac{1}{n}\sum_n \left[x(n) - \sum_{i=1}^{p} a_i x(n-i)\right]^2 \tag{23}$$

In order to obtain the minimum mean square error, the partial derivate of the mean square error E_x with respect to the coefficients $\{a_i\}$ is zero.

$$\frac{\partial E_x}{\partial a_j} = \sum_n x(n)\,x(n-j) - 2\sum_{i=1}^{p} a_i \sum_n x(n-i)\,x(n-j) = 0 \tag{24}$$

where, $j = 1, 2 \ldots p$ and $i = 1, 2 \ldots p$. Let

$$\phi(i,j) = \sum_n x(n-i)\,x(n-j) \tag{25}$$

The Eq. (24) can be simplified to

$$\sum_{i=1}^{p} a_i \phi(i,j) = \phi(0,j) \tag{26}$$

The autocorrelation function of $x(n)$ can be expressed as

$$r(j) = \sum_{n=0}^{N-1} x(n)\,x(n-j) \tag{27}$$

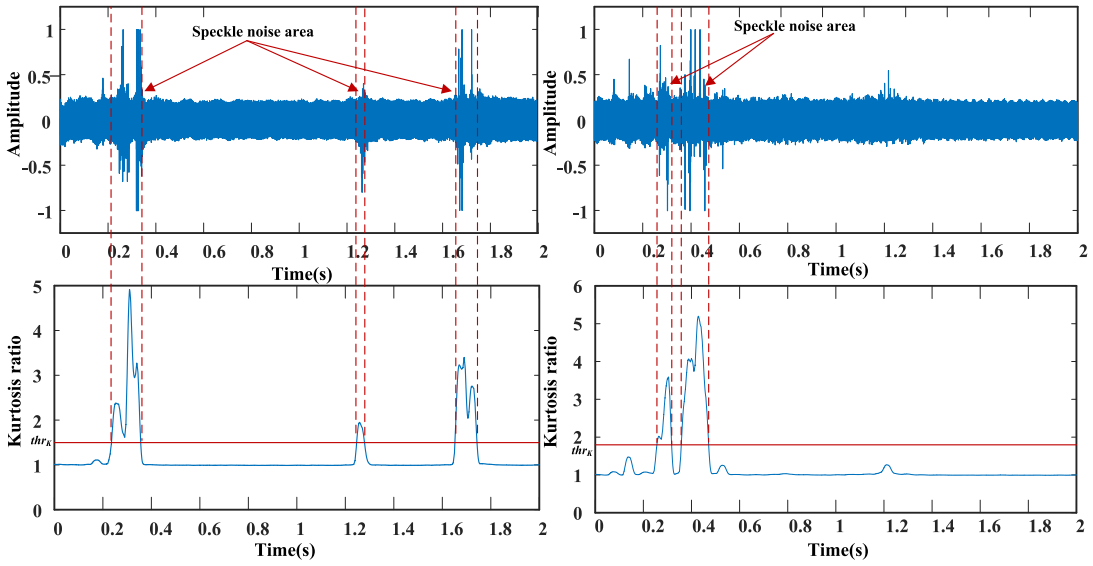

Fig. 10. Confirmation of the speckle noise area.

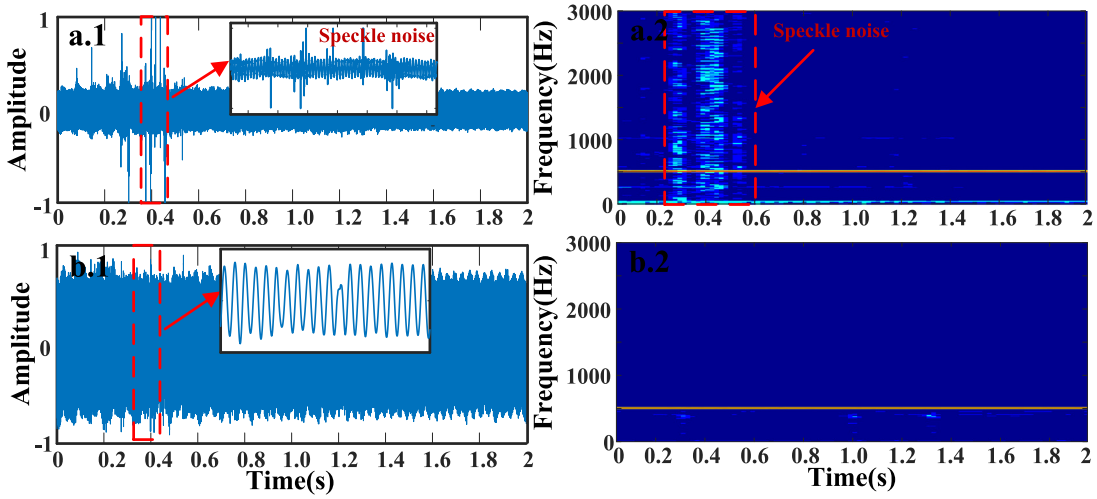

Fig. 11. Spectrograms and waveforms. (a) the LDV-captured tone signal with speckles noise. (b) the enhanced signal.

where, N is the length of the signal. According to Eq. (27), the Eq. (25) can be rewritten as

$$\phi(i,j) = r(i-j) \tag{28}$$

Since $r(j)$ is an even function, the Eq. (28) can be rewritten as

$$\phi(i,j) = r(|i-j|) \tag{29}$$

According to Eq. (29), the Eq. (25) can be expressed as

$$\sum_{i=1}^{p} a_i r(|i-j|) = r(j) \Rightarrow$$
$$\begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(p-1) \\ r(1) & r(0) & r(1) & \cdots & r(p-2) \\ r(2) & r(1) & r(0) & \cdots & r(p-3) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ r(p-1) & r(p-2) & r(p-3) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ r(3) \\ \vdots \\ r(p) \end{bmatrix} \tag{30}$$

So, we can acquire the coefficients $\{a_i\}$ by solving the Eq. (30).

The basic steps in the speckle noise removal algorithm are: First, the signal segments where the speckle noise occurs are ascertained by segment kurtosis ratio. Secondly, the noisy signal segments are regarded as the data loss. Thirdly, based on the several sampling point values adjacent to the lost data $x(n-p)\ldots x(n-1)$, using the linear prediction to estimate the lost data $x(n)$. In this paper, the number of coefficients $p = 256$.

$$x(n) = \sum_{i=1}^{256} a_i x(n-i) \tag{31}$$

The spectrograms and waveforms of LDV-captured tone signals with speckles noise and their enhanced signal are depicted in Fig. 11. Comparing Fig. 11.a with Fig. 11.b, we can obviously find that the speckle noise can be eliminated by this method effectively.

Besides, to prove that this algorithm is also suitable for speech signals, an experiment is set up. The experimental set up is similar to Fig. 7. The major difference is that the excitation source is no longer a single tone signal but a voice.

The spectrograms and waveforms of LDV speech signals with speckles noise, their enhanced signal, and the corresponding original clean speech are depicted in Fig. 12. It can be seen from the spectrograms and waveforms that the LDV speech signals (Fig. 12a) are disturbed by the speckles noise seriously with the speckles noise concentration distribution in the relatively low frequency range (under 1600 Hz). Fortunately, we find that the algorithm proposed by us can eliminate noise effectively.

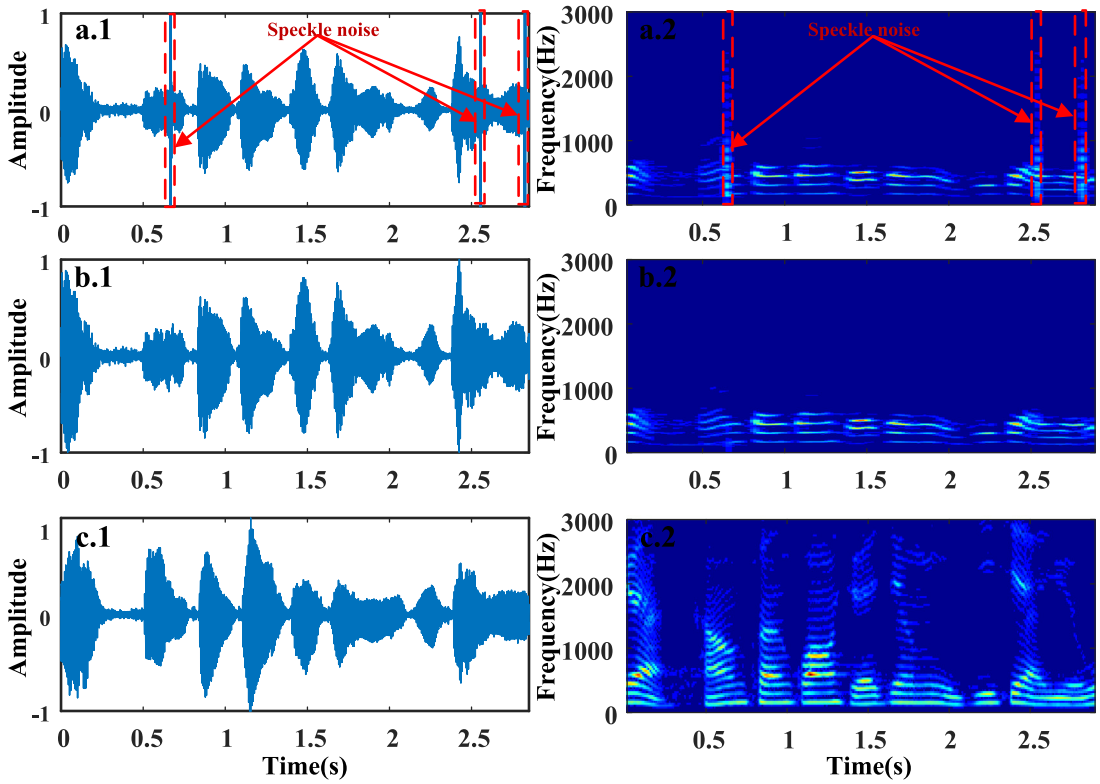

Fig. 12. Spectrograms and waveforms. (a) Speech signals measured by the LDV. (b) Speech enhanced using speckles removal algorithm. (c) Clean speech signals measured by the cell phone.

Meanwhile, the performance of speckles noise removal algorithm proposed by us is evaluated using the Perceptual evaluation of speech quality (PESQ). PESQ is a method for evaluating voice quality first proposed by ITU-T (International Telecommunication Union Standardization Organization) in 2001 [35]. Unlike other objective evaluation methods, the PSEQ algorithm, which based on auditory and cognitive models, simulates subjective evaluation by combining physical characteristics such as speech frequency and loudness with human perception characteristics. Therefore, this method is currently recognized as the most relevant to subjective evaluation. The PSEQ is given as follows:

$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind} \tag{32}$$

where, D_{ind} is the average disturbance value, A_{ind} is the asymmetric disturbance value and the specific steps for solving the two variables are mentioned in [35,36]. Another Coefficients are set as follows: $a_0 = 4.5$, $a_1 = -0.1$, $a_2 = -0.0309$ according to [36]. The PESQ score range is from 0.5 to 4.5, and the measured speech signals with high quality observe a high PESQ score.

The PESQ score of the LDV speech signals using the above formula and experimental data is 3.1933, while the PESQ score of the enhanced signals is 3.4021. The PESQ score increased by 6.54 percent compared to the former, which demonstrates that the quality of the speech signal is significantly improved. Experiment results indicated that the used speckles noise removal algorithm can eliminate speckles noise effectively.

6. Conclusions

In this paper, we have done a comprehensive study on the effect of speckles noise on the LDV for remote speech detection. The contribution of this paper can be concluded into three aspects. First, we investigate the statistical properties of dynamic speckle due to relative in-plane motion theoretically and experimentally, and the results show that the speckles intensity fluctuation increases with the increase of the relative velocity for a given value of the detection distance. Second, both experimental and theoretical analyses are given to explore the effect of the speckles intensity fluctuation on the LDV system. The results indicate that the spectral power density of the speckles noise increases with the increase of the relative velocity, and the speckles noise can cause spikes and bursts appearing in the LDV-captured voice signals. Due to the spikes and bursts, sounds like sharp and harsh squeaks, which have a great influence on the quality of the speech. At last, a speckles removal algorithm based on kurtosis ratio and linear prediction is introduced to suppress the speckles noise and improve the quality of the LDV-captured speech. The experiment results demonstrate the feasibility and effectiveness of the speckles removal algorithm.

This work is supported by the National Natural Science Foundation of China under Grant No. 61205143.

References

[1] C. Zieger, A. Brutti, P. Svaizer, Acoustic based surveillance system for intrusion detection, in: Sixth IEEE International Conference on Advanced Video & Signal Based Surveillance, IEEE Computer Society, 2009.
[2] C. Clavel, T. Ehrette, G. Richard, Events detection for an audio-based surveillance system, in: IEEE International Conference on Multimedia & Expo, IEEE, 2005.
[3] R. Radhakrishnan, A. Divakaran, P. Smaragdis, et al., Audio analysis for surveillance applications, in: IEEE Workshop on Applications of Signal Processing to Audio & Acoustics, IEEE, 2005.
[4] L. Antonelli, F. Blackmon, Experimental demonstration of remote, passive acousto-optic sensing, J. Acoust. Soc. Am. 116 (6) (2004) 3393.
[5] I.A. McCowan, J. Pelecanos, S. Sridharan, Robust speaker recognition using microphone arrays, in: A Speaker Odyssey-the Speaker Recognition Workshop, 2011.
[6] L. Wang, Robust distant speaker recognition based on position-dependent CMN by combining speaker-specific GMM with speaker-adapted HMM, Speech Commun. 49 (6) (2007) 501–513.
[7] JOHN R. RZASA, KYUMAN CHO, CHRISTOPHER C. DAVIS1, Long-range vibration detection system using heterodyne interferometry, Appl. Opt. 54 (20) (2015) 6230–6236.
[8] Ming-Hung Chiu, Wei-Chou Chen, Chen-Tai Tan, Small displacement measurements based on an angular-deviation amplifier and interferometric phase detection, Appl. Opt. 54 (10) (2015) 2885–2890.

[9] Xin Zhang, Weifeng Diao, Yuan Liu, Xiaopeng Zhu, Yan Yang, Jiqiao Liu, Xia Hou, Weibiao Chen, Eye-safe single-frequency single-mode polarized all-fiber pulsed laser with peak power of 361, W. Appl. Opt. 53 (11) (2014) 2465–2469.
[10] Weifeng Diao, Xin Zhang, Jiqiao Liu, Xiaopeng Zhu, Yuan Liu, Decang Bi, Weibiao Chen, All fiber pulsed coherent lidar development for wind profiles measurements in boundary layers, Chin. Opt. Lett. 12 (7) (2014) 072801.
[11] Jianhua Shang, Shuguang Zhao, Yan He, Weibiao Chen, NingJia, Experimental study on minimum resolvable velocity for heterodyne laser Doppler vibrometry, Chin. Opt. Lett. 9 (8) (2011) 081201.
[12] Y. Qu, T. Wang, Z. Zhu, Remote audio/video acquisition for human signature detection, in: IEEE Computer Society Conference on Computer Vision & Pattern Recognition Workshops, IEEE, 2009.
[13] W. Li, M. Liu, Z. Zhu, et al., LDV Remote voice acquisition and enhancement, in: International Conference on Pattern Recognition, IEEE, 2006.
[14] A.T. Wang, Z. Zhu, A. Divakaran, Long range audio and audio-visual event detection using a laser Doppler vibrometer, in: Evolutionary and Bio-Inspired Computation: Theory and Applications IV, Vol. 7704, 2010, 77040J-77040J-6.
[15] Y. Qu, T. Wang, Z. Zhu, Remote audio/video acquisition for human signature detection, in: Cvpr'09 Biometrics, 2009, pp. 66–71.
[16] Y. Qu, T. Wang, S. Member, et al., An active multimodal sensing platform for remote voice detection, in: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, IEEE, 2014.
[17] Rui Li, Nicholas Madampoulos, Zhigang Zhu, Liangping Xie, Performance comparison of an all-fiber-based laser Doppler vibrometer for remote acoustical signal detection using short and long coherence length lasers, Appl. Opt. 51 (21) (2012) 5011–5018.
[18] R. Peng, C. Zheng, X. Li, Bandwidth extension for speech acquired by laser Doppler vibrometer with an auxiliary microphone, in: International Conference on Information, IEEE, 2016.
[19] Y. Avargel, I. Cohen, Speech measurements using a laser Doppler vibrometer sensor: Application to speech enhancement, in: Hands-free Speech Communication & Microphone Arrays, IEEE, 2011.
[20] Y. Deng, Long range standoff speaker identification using laser Doppler vibrometer, in: IEEE International Conference on Biometrics Theory, IEEE, 2016.
[21] TaoLV, ZHANG He-yong, YAN Chun-hui, Double mode surveillance system based on remote audio/video signals acquisition, Appl. Acoust. 129 (2018) 316–321.
[22] R. Li, T. Wang, Z. Zhu, et al., Vibration characteristics of various surfaces using an LDV for long-range voice acquisition, IEEE Sens. J. 11 (6) (2011) 1415–1422.
[23] T. Pfister, A. Fischer, Czarske, Jürgen, Cramér–Rao lower bound of laser Doppler measurements at moving rough surfaces, Meas. Sci. Technol. 22 (5) (2011) 055301.
[24] K.D. Ridley, E. Jakeman, FM Demodulation in the presence of multiplicative and additive noise, Inverse Problems 15 (4) (1999) 989–10021002.
[25] K.D. Ridley, E. Jakeman, Signal-to-noise analysis of FM demodulation in the presence of multiplicative and additive noise, Signal Process. 80 (9) (2000) 1895–1907.
[26] C.A. Hill, M. Harris, K.D. Ridley, et al., Lidar frequency modulation vibrometry in the presence of speckle, Appl. Opt. 42 (6) (2003) 1091-100.
[27] L.A. Jiang, M.A. Albota, R.W. Haupt, et al., Laser vibrometry from a moving ground vehicle, Appl. Opt. 50 (15) (2011) 2263–2273.
[28] A. Dräbenstedt, Quantification of displacement and velocity noise in vibrometer measurements on transversely moving or rotating surfaces, Opt. Meas. Syst. Ind. Insp. V. (2007).
[29] M. Bauer, F. Ritter, G. Siegmund, High-precision laser vibrometers based on digital Doppler signal processing, Proc. Spie 4827 (4827) (2002) 50–61.
[30] T. Iwai, N. Takai, T. Asakura, Dynamic statistical properties of laser speckle produced by a moving diffuse object under illumination of a Gaussian beam, J. Opt. Soc. Amer. 72 (4) (1982) 460–467.
[31] WU. YingLi, Statistical Properties of Laser Speckles and Detection Micro-motion for Arbitrary Targets with Rough Surface, Xidian University, 2013.
[32] J. Vass, R. Šmíd, R.B. Randall, et al., Avoidance of speckle noise in laser vibrometry by the use of kurtosis ratio: Application to mechanical fault diagnostics, Mech. Syst. & Signal Process. 22 (3) (2008) 647–671.
[33] Lal. Amit, Multi-beam laser Doppler vibrometer for landmine detection, Opt. Eng. 45 (10) (2006) 104302.
[34] J.D. Markel, A.H.G. Jr, Linear prediction of speech, Commun. & Cybern. 53 (3) (1976) 266–267.
[35] A.W. Rix, J.G. Beerends, M.P. Hollier, et al., Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs, IEEE Proc. 2 (2001) 749–752.
[36] Wenli Zhang, Algorithm Research and Implementation of PESQ Speech Quality Assessment System, Dalian University of Technology, 2007.