Applied Acoustics 126 (2017) 136–142

Contents lists available at ScienceDirect

Applied Acoustics

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

The novel role of arctangent phase algorithm and voice enhancement techniques in laser hearing

He-yong Zhang [a,*], Tao Lv [a,b], Chunhui Yan [a,b]

[a] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, State Key Laboratory of Laser Interaction with Matter, Changchun 130033, China
[b] University of the Chinese Academy of Sciences, Beijing 10039, China

ARTICLE INFO

Article history:
Received 9 February 2017
Received in revised form 16 May 2017
Accepted 18 May 2017
Available online 30 May 2017

Keywords:
LDV
Laser hearing
Arctangent phase algorithm
Voice enhancement

ABSTRACT

At present, laser hearing has played more and more important role in the field of anti-terrorism and security defense all around the world. In order to acquire remote voice, a Laser Doppler Vibrometer (LDV) is established, the voice signal demodulation method is based on the arctangent phase algorithm. On the basis of the system, a kind of speech enhancement technology is used to improve the intelligibility of the noisy voice signals detected by the LDV system. First, based on the heterodyne detection theory, the detection principle and method which acquire voice by detecting throat vibration is introduced. Then a kind of speech enhancement technology is used to improve the intelligibility of the noisy voice signals detected by the LDV system. Finally, to validate this system and speech enhancement technology, some experiments are performed and the results indicated that the comprehensible speech signals within the range of 75 m can be obtained by self-made LDV. On the other hand, the speech enhancement technology can improve the intelligibility of the noisy voice signals detected by the LDV system effectively.

1. Introduction

Multimodal/multi-sensor surveillance systems are widely deployed today for security purpose [1]. Although a lot of progress has been made, particularly with the rapid improvements of color and infrared (IR) cameras and the corresponding algorithms for monitoring subjects at a large distance, audio information, as an important data source, has not yet been fully explored. A few systems have been reported to integrate visual and acoustic sensors. But in these systems, the acoustic sensors need to be close to the subjects in monitoring. Parabolic microphones could be used for remote hearing and surveillance, which can capture voice at a fairly large distance in the direction pointed by the microphone. But it is very sensitive to the noise caused by wind or sensor motion, and all the signals on the way get captured. Recently, all kinds of LDV have been widely used in industry inspection [2–7]. The prpducts such as those manufactured by Polytec and B&K Ometron can effectively detect vibration within two hundred meters with sensitivity on the order of 1 μm/s. For example, they have been used to measure the vibrations of civil structures like high-rise buildings, bridges, towers etc. at the distance up to 200 m. However, literature on remote voice detection using LDV is rare. Therefore, the study of the novel application of an LDV for remote voice detection will be the main focus of this paper.

LDV has the characteristics of long distance, non-contact and high sensitivity; it has been widely used in industry and military field [8,9]. Because LDV can detect and measure extremely tiny vibration of a target at a long distance, this motivates us to detect the tiny vibration of a target (vibration caused by the voice energy) at a large distance to acquire remote speech signal by LDV [10–14]. Most of the research results show that the commercial LDV based on 632 nm is applied for voice detection, this is not fit for realistic application due to its short detection range, also not convenient for the integration of demodulation and the voice enhancement algorithm. Therefore, this paper will focus on the arctangent phase algorithm and voice enhancement techniques in laser hearing. In fact, the realistic application system will be based on the near IR and eye-safe laser output, but this laser system has some troubles in target alignment in the wide field. So we will adopt the visible laser at 532 nm for the experiment to verify the demodulation and voice enhancement algorithm, and the all-fiber near IR LDV system is developing now in my LAB.

In this paper, a LDV system based on laser homodyne structure, arctangent phase demodulation algorithm and audio signal enhancement algorithms is developed, and it is used to detect long range audio signal. A few experiments are implemented to test its performance in the end.

* Corresponding author.
E-mail address: zhanghy@ciomp.ac.cn (H.-y. Zhang).

http://dx.doi.org/10.1016/j.apacoust.2017.05.024

2. Experimental setup

Fig. 1 shows the schematic diagram of the LDV we developed. The LDV is composed of transceiver and signal processing units. The main parameters of the system are given in Table 1. A 50-mW single mode CW laser with the line-width less than 1 MHz at wavelength of 532 nm used as transmitted laser source. The output laser is divided into two beams by a beam splitter. One beam is modulated by Acousto-Optical Modulator (AOM) with 80 MHz frequency shift, this beam is taken as the local-oscillator (LO) beam. The other beam is transmitted to vibrating target perpendicularly through an optical circulator and telescope. Due to the vibration of the target (vibration caused by the voice energy), the reflected laser beam carries Doppler frequency shift, and it can be received by the same telescope, this beam is taken as the signal beam.

The LO and signal beam are mixed by a beam combiner then detected by a photoelectric detector. The detector output signal $\boldsymbol{u}$ is given as

$$u = \alpha A_o A_s \cos[\omega_{AO} t + \varphi(t) + \varphi_1 - \varphi_2] \quad (1)$$

where A_o and A_s are the amplitude of LO and signal beam, ω_{AO} is the frequency shift caused by AOM, φ_1 and $\varphi_2$are random phase, α is the photoelectric conversion efficiency, $\varphi(t)$ is the Doppler shift, which can be expressed as

$$\varphi(t) = \frac{4\pi S(t)}{\lambda} \quad (2)$$

where λ is the wavelength of laser, $S(t)$ is the vibration displacement. We can't get the object vibration characteristic directly because the detector output signal is optical beat signal. At this moment, we need quadrature demodulation circuit and arctangent phase algorithm to demodulate the beat signal to acquire the object vibration information. Fig. 2 and Fig. 3 show the demodulation and arctangent phase algorithm block diagram.

The detector output signal is halved by power divider in the demodulate process. The first part turns into u_I which mixes with the driving signal of AOM and passed through a low-pass filter. The second part turns into u_Q which has a 90°phase shift and mixes with the driving signal of AOM and passed through a low-pass filter. We use an acquisition card with 0.1 MHz sampling rate to sample u_I and u_Q. We can use arctangent phase algorithm to get the Doppler shift $\varphi(t)$ based on u_I and u_Q, then the vibration displacement can be recovered through $\varphi(t)$, also the speech signal will be output from the processing unit after some digital filtering. The baseband signal u_I and u_Q and the Doppler shift $\varphi(t)$ during in the process above are given as:

$$\begin{aligned} u_I &= \xi \cos[\varphi(t) + \varphi_1 - \varphi_2] \\ u_Q &= \xi \sin[\varphi(t) + \varphi_1 - \varphi_2] \\ \varphi(t) &= \arctan(u_Q/u_I) + m\pi + \Delta\varphi \end{aligned} \quad (3)$$

where ξ is the amplitude coefficient of the both signals, and the ambiguity of the arctangent function can be removed by a phase

Table 1
Main Parameters of the LDV System.

Parameters	Value
Wavelength	532 nm
Power	50 mW
Line-width	less than 1 MHz
Intermediate frequency (IF)	80 MHz
Detector	1 GHz bandwidth
Sampling rate	0.1 MHz
Effective range	50 m

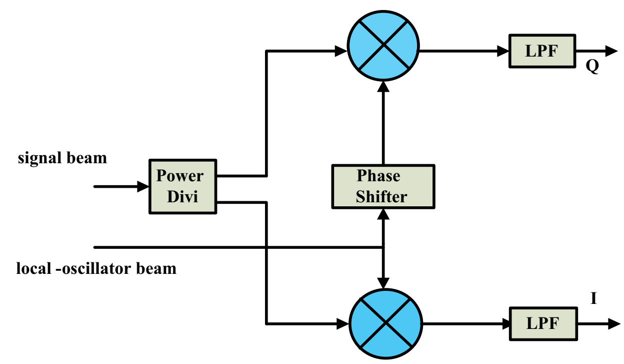

Fig. 2. The block diagram of demodulation.

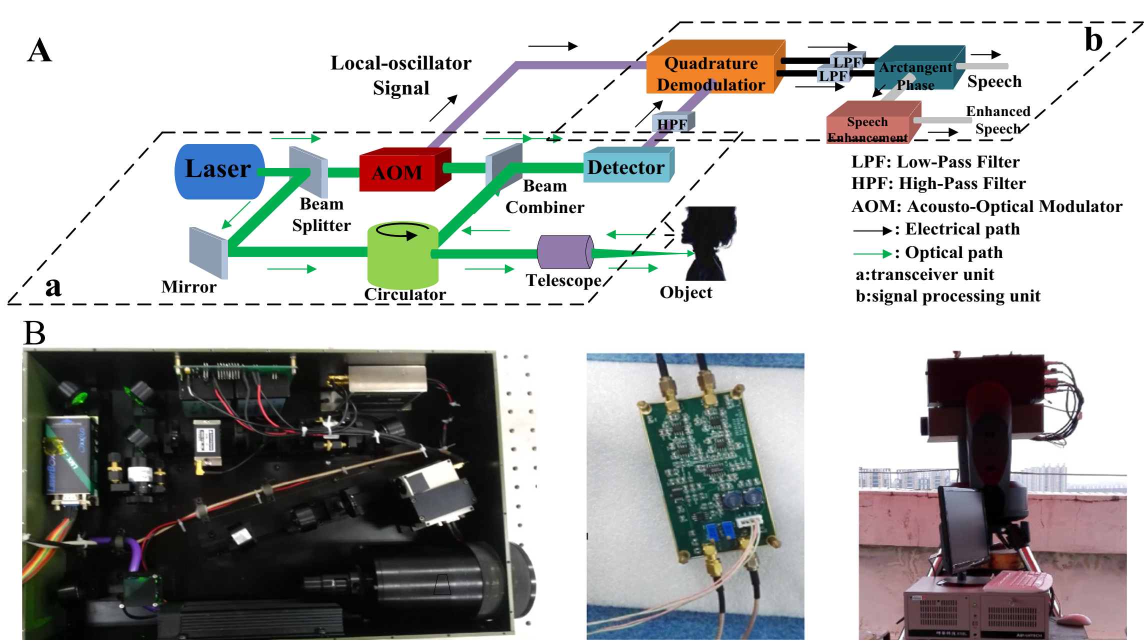

Fig. 1. (A) Schematic diagram of the LDV and (B) prototype of the LDV system.

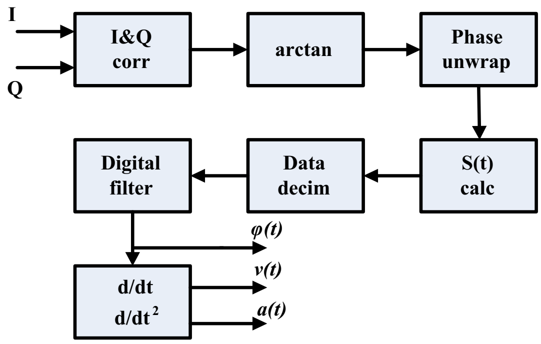

Fig. 3. The block diagram of arctangent phase algorithm.

unwrapping algorithm. The demodulated voice signal will not be fit for hearing due to the large noise in the speech signal acquired directly by LDV. So we use some speech enhancement techniques to improve the intelligibility of the noisy voice signals detected by the LDV system.

3. Demodulation algorithm simulation

In order to check the validity of the arctangent phase algorithm and voice enhancement methods, some simulation analysis and experiment test were carried out in the following sections. The effects of the arctangent phase algorithm can be showed in the demodulation simulation, and the voice enhancement methods will be appeared during the course of realistic voice acquisition.

The demodulation simulation was performed on Matlab software with the conditions of f_{AO} = 80 MHz, λ = 532 nm, $_{\Delta}\varphi = 0$; αA_o-A_s = 1 according to Eqs. (1-3). Figs. 4–7 are the simulation results with single frequency, double frequency and the voice vibration signal based on the arctangent phase algorithm.

The vibration signal's expressions in Figs. 4 and 5 are $S(t) = 10 * \cos(2\pi \times 500t)$, $S(t) = 10 * \cos(2\pi \times 500t) + 10 * \cos(2\pi \times 1000t)$ respectively. The recovered signal showed that the vibration amplitude and frequency are the same with the original signal, because there is no noise added in the simulation. The two figures above demonstrated the demodulation methods based on arctangent phase algorithm has well effects. In fact, the real voice in the world was composed with different kinds of vibrations, so the signal is also more complex than single or double frequency vibrations. In order to validate the algorithm above, the real voice were applied in the numerical simulations, we have tried two kind of human voice "ni hao" in Chinese and "Hello" in English in the simulations, the results can be shown in Fig. 6,7 as follows:

The real voice signal which modulated on the laser beam can be demodulated by the laser coherent detection through LDV and arctangent phase algorithm effectively. The time and frequency domain characters of the recovered signal look the same as original signal. We can confirm that the laser coherent detection with arctangent phase algorithm has the ability to recover complex voice vibration signals.

4. Experimental results

The feasibility of detecting the long range single tone signals with LDV is demonstrated as Fig. 8.

In the experiments, the voice box is regarded as the target, and it was driven by sinusoidal signal generated by signal generator to produce single tone. We can control the single tone frequency by adjusting the frequency of the signal generator. The LDV transmitted the laser beam perpendicularly to the voice box surface (penetrate through glass) in order to obtain the optimum reflected signal with maximum carrier information. In order to verify the capability of speech recovering, we set the signal's frequency from 300 Hz to 3 kHz which is the frequency range of human voice with consistent sound intensity, and the effective detection distance is 75 m.

The experimental results are shown in Fig. 9, we can obviously find the recovered signal waveform and frequency are similar with the single tone from the generator. We compared the data measured by our LDV with the theoretical data, the result showed the measuring error of frequency within 0.081%. The measured data are shown in table 2. In addition, there are some undesirable noise caused by environment and LDV system, we can use band-pass filters to attenuate the noise outside of the signal bandwidth. The experimental results show the LDV system has ability to detect long range single tone signals.

Although frequency range of human voice is about 300 Hz to 3 kHz, the frequency response range of the LDV is much wider than that. Even if we have used the band-pass filters, we still get signals that are subject to large, slowly varying components corresponding to the slow but significant background vibrations near to the targets. The magnitudes of the meaningful acoustic signals are relatively small, adding on top of the low frequency vibration signals. In addition, the inherent "speckle pattern" problem on a normal "rough" surface and the occlusion of the LDV laser beam by passing-by objects both introduce noise spreading across the voice frequency range. This creates undesirably loud noise when we directly listen to the acoustic signal. Therefore, we apply an audio signals enhancement technique to deal with these problems.

The Wiener filtering method is adopted during in the signal enhancement, which is one of the most effective audio signals enhancing techniques that can reduce the noise with frequency

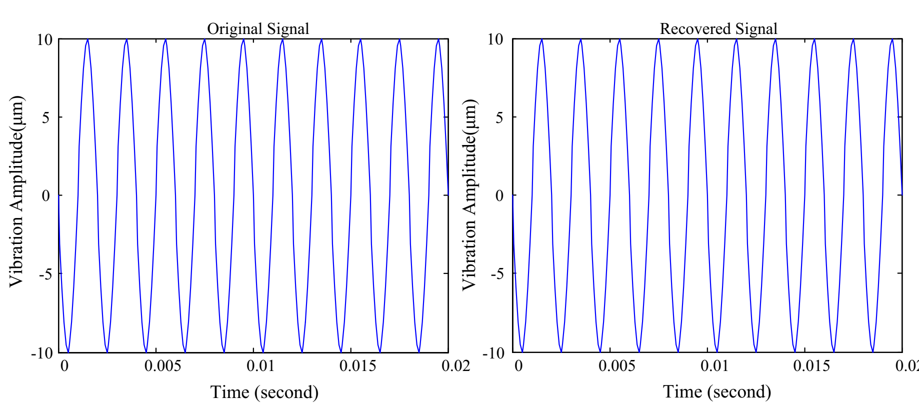

Fig. 4. Single frequency vibration and it's recovered signal.

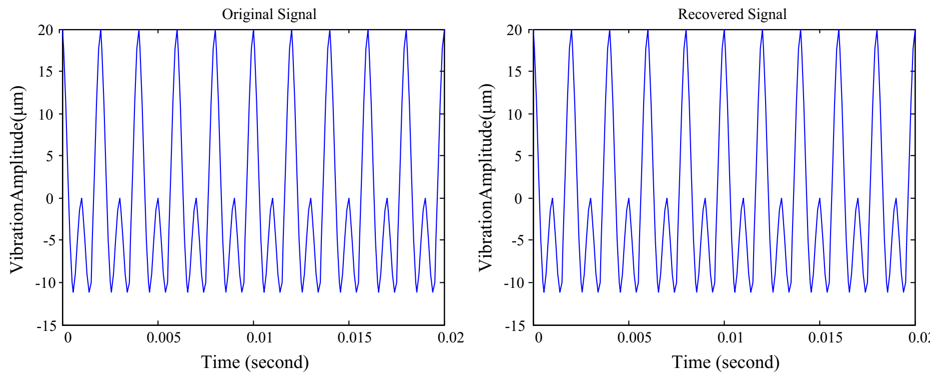

Fig. 5. Double frequency vibration and it's recovered signal.

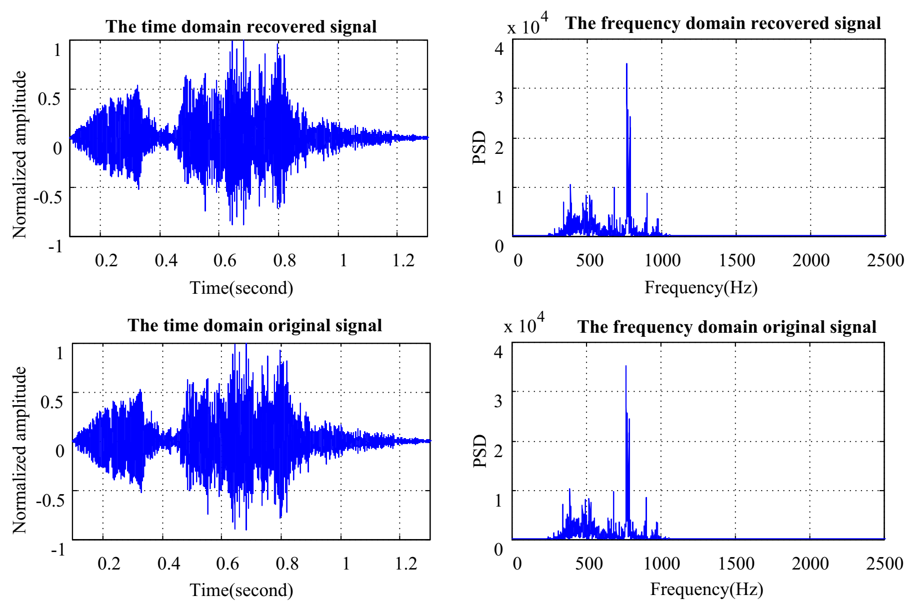

Fig. 6. The time and frequency domain analysis for the recovered and original signal"ni hao".

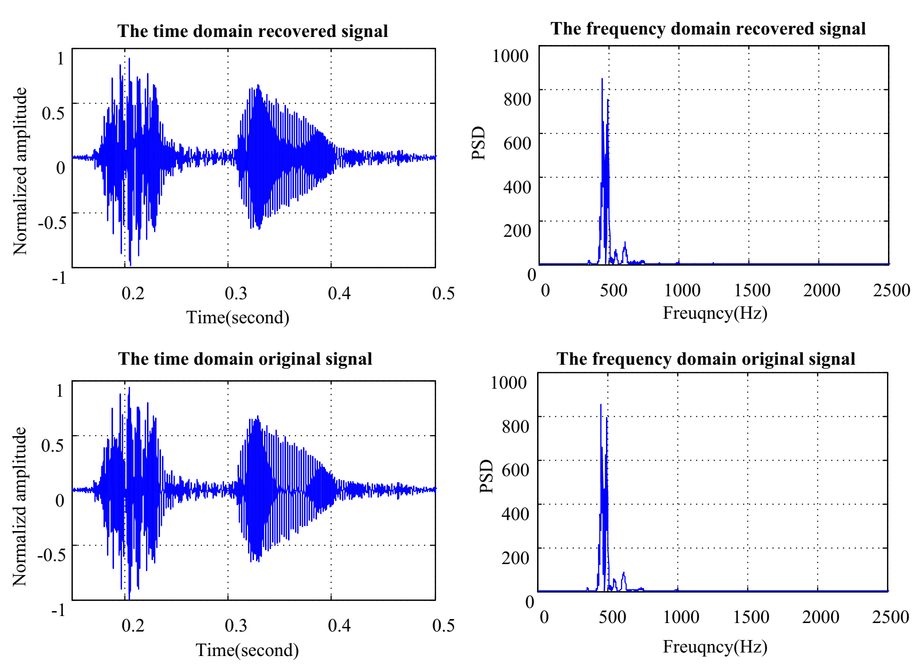

Fig. 7. The time and frequency domain analysis for the recovered and original signal"Hello".

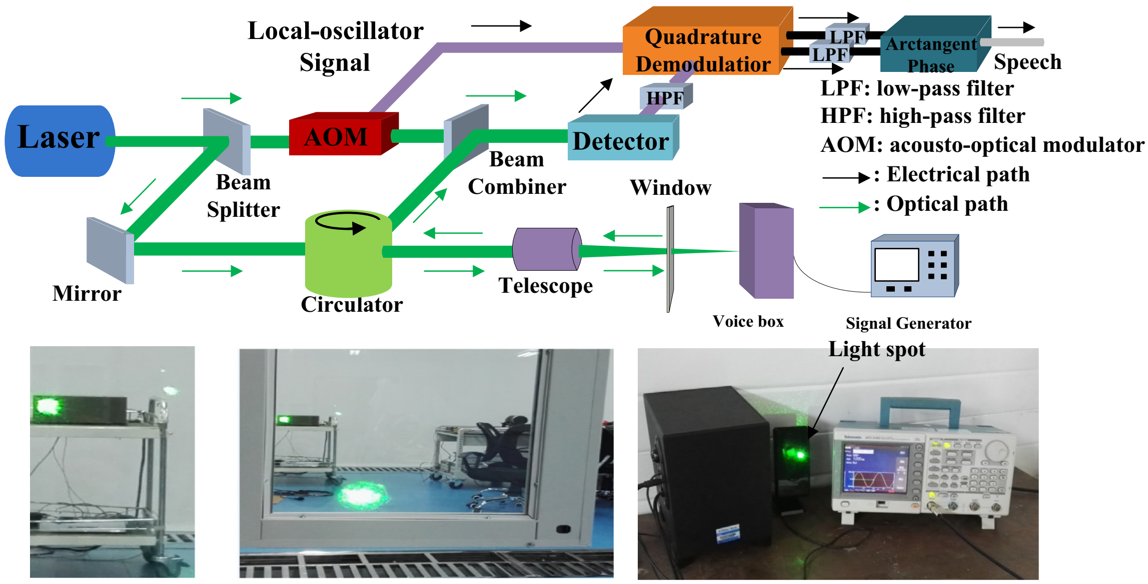

Fig. 8. The experiment setup for single tone signal detecting with LDV.

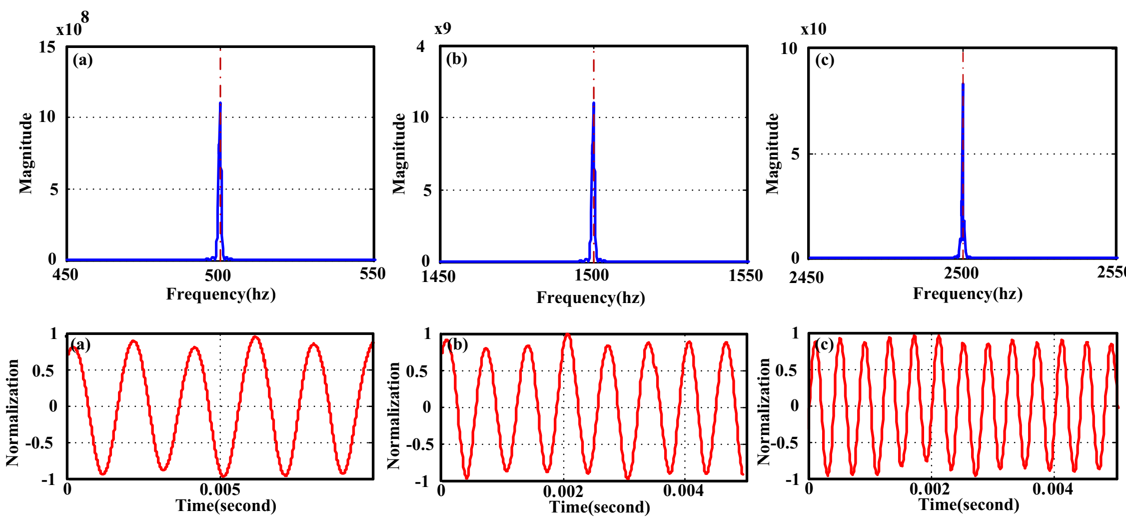

Fig. 9. The sequence diagram and spectrogram of recovered single tone range from 300 to 3000 Hz. (a) the spectrogram and oscillogram of 500 Hz single tone, (b) the spectrogram and oscillogram of 1500 Hz single tone, (c) the spectrogram and oscillogram of 2500 Hz single tone.

inside of speech bandwidth, and increase signal to noise ratio effectively.

The audio signal can be described as the summation of a clean speech signal and additive noise, which is illustrated as

$$y(t) = s(t) + n(t) \tag{4}$$

Assumption $n(t)$ is additive noise signal, and the clean signal $s(t)$ is independent of the noise $n(t)$, so their Fourier transform can be described as

$$Y(k,j) = S(k,j) + N(k,j) \tag{5}$$

where $Y(k,j)$, $S(k,j)$, $N(k,j)$ respectively represent the corresponding Fourier Transformation of time-domain signals in the k frame and the j spectral component.

The Wiener filter frequency response is

$$H(k,j) = \frac{P_s(k,j)}{P_s(k,j) + P_n(k,j)} \tag{6}$$

where $P_s(k,j)$ is the clean speech signal spectrum power and $P_s(k,j) = E[|S(k,j)|^2]$, $P_n(k,j)$ is the noise signal spectrum power and $P_n(k,j) = E[|N(k,j)|^2]$. In Wiener filtering, the output signal $S(k,j)$ can be illustrated as

$$\hat{S}(k,j) = Y(k,j) \times H(k,j) \tag{7}$$

In addition, we use the Minima Controlled Recursive Averaging (MCRA) to estimate noise signal spectrum power.

The feasibility of detecting remote voice with LDV and the speech enhancement technology are demonstrated with the setup shown in Fig. 10. A laser is divided into two beams, one beam is modulated by acousto-optical modulator (AOM) which is taken as the local-oscillator beam (LO), its frequency shifted 80 MHz. The other beam is transmitted to throat (penetrate through glass) perpendicularly through an optical circulator and a telescope, and it can be received by the same telescope, this beam is taken as the signal beam. The LO beam and signal beam are mixed with a com-

Table 2
LDV measuring data and theoretical data.

	Frequency (Hz)	Frequency Measuring Uncertainty		Frequency (Hz)	Frequency Measuring Uncertainty
LDV	300.13	0.0433%	LDV	1501.21	0.081%
Theoretical data,	300.00		Theoretical data,	1500.00	
LDV	500.28	0.056%	LDV	2001.28	0.064%
Theoretical data,	500.00		Theoretical data,	2000.00	
LDV	999.97	0.003%	LDV	2500.20	0.008%
Theoretical data,	1000.00		Theoretical data,	2500.00	

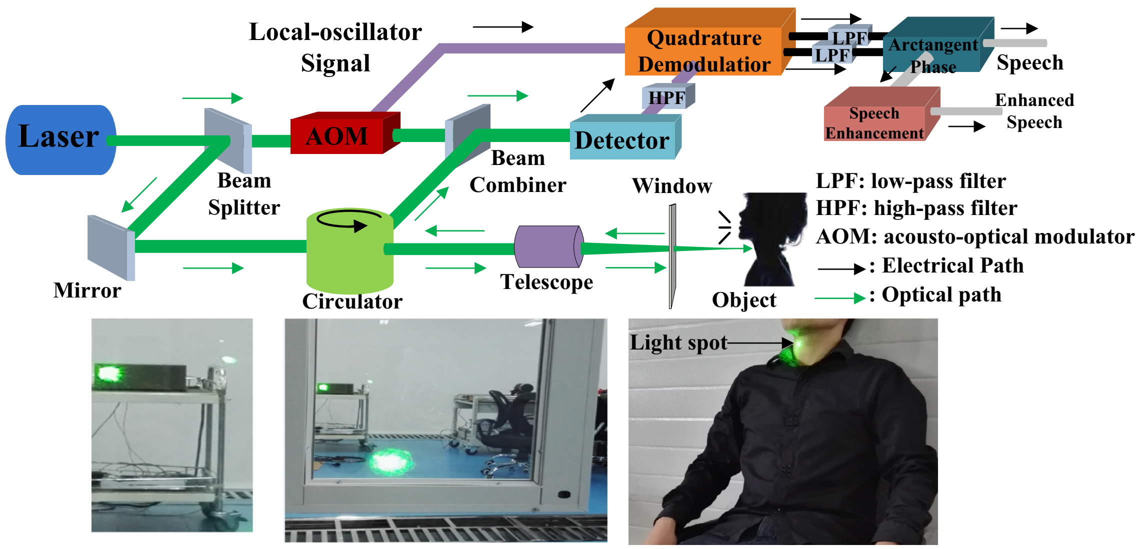

Fig. 10. The experiment setup for human voice acquisition with LDV.

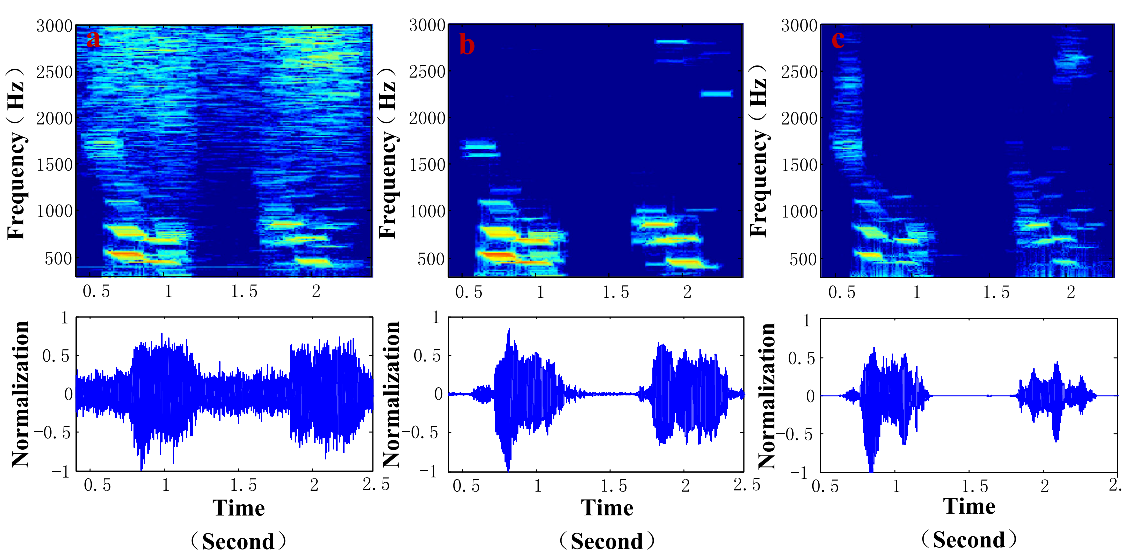

Fig. 11. The spectrogram and oscillogram of original LDV signal (a), the spectrogram and oscillogram of Wiener filtered signal (b), the spectrogram and oscillogram of clean signal (c). All correspond to the speech of “tree, jungle” (in Chinese).

biner and detected by photoelectric detector. Because the detector output signal is beat between the two beams, we demodulate the beat signal to acquire the remote voice by using quadrature demodulation circuit and arctangent phase algorithm. Then the speech enhancement technology can be used to improve the intelligibility of the noisy voice signals from the LDV. In this experiment, the distance between throat and LDV system is about 75 m.

To evaluate the experiment result, we compare the oscillogram and spectrograms of LDV audio, its enhanced speech, and a corresponding clean signal captured at the same time with a recording device (such as cell phone). The distance between throat and recording device is about 5 cm. These figures are shown in Fig. 11.

Compare Fig. 11.a with Fig. 11.c, we can find that most of the high frequency energy in clean signal (Fig. 11.c) disappeared in LDV audio signal (Fig. 11.a). It is because the throat could not

Table 3
The score standard of MOS.

The score of MOS	Intelligibility	Distortion level
5	Excellent	Don't feel distortion
4	Good	Just feel distortion
3	Okay	Feel distortion and have a little disgust
2	Difficult	Obviously feel distortion but can bear it
1	Bad	Unacceptable

vibrate in such high frequency. In Fig. 11.a we also see strong noise in the whole frequency domain, this because the inherent "speckle pattern" problem on a normal "rough" surface and circuit noise will create undesirably loud noise. However, the valid frequency involved in the clean signal (Fig. 11c) are same with the LDV audio signal (Fig. 11a), the oscillogram of original LDV signal are similar to the clean signal. Compare Fig. 11b with c and a, when we apply the Wiener filtering, the background noise is decreased significantly, and the spectrogram and oscillogram with Wiener filtered signal are more similar to the clean signal. Experiments indicated that the comprehensible speech signals within the range of 75 m can be obtained by the LDV developed by us, and speech enhancement technology can improve the intelligibility of the noisy voice signals detected by the LDV system effectively. In addition we use SNR values and invite five volunteers to assessment the recovered signal quality with the Mean Opinion Score (MOS) evaluation criterion which evaluation standard is shown in the Table 3. The MOS without Wiener filtering is 3.4, the SNR is 6.079 dB, and the MOS with wiener filtering is 4.0, the SNR is 11.140 dB.

5. Conclusion

In summary, we develop a LDV system which can detect the long range audio, also with the audio signal enhancement techniques to improve audio quality effectively. This system has a 50 mW single mode CW laser with the line-width less than 1 MHz at wavelength of 532 nm. Quadrature demodulation circuit and arctangent phase algorithm are used to demodulate the signal. In addition, the audio signals enhancement techniques are applied in this system to improve the signal quality. Experiments indicated that the comprehensible speech signals within the range of 75 m can be obtained with the LDV, and speech enhancement technology can improve the intelligibility of the noisy voice signals detected by the LDV effectively. In the domestic related research field, the speech detection performance of our LDV system is in the lead. In the future, the fiber laser with better coherence and higher power can be used to improve the detection range, a better experiment results will be acquired at that moment. And this kind of LDV can be used in various applications such as rescue in disaster and remote area surveillance.

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

References

[1] Weihong Li, Ming Liu, Zhigang Zhu, Huang Thomas S. LDV remote voice acquisition and enhancement. The 18th International Conference on Pattern Recognition, 2006.
[2] Rzasa John R, Cho Kyuman, Davis Christopher C. Long-range vibration detection system using heterodyne interferometry. Appl Opt 2015;54(20):6230–6.
[3] Chiu Ming-Hung, Chen Wei-Chou, Tan Chen-Tai. Small displacement measurements based on an angular-deviation amplifier and interferometric phase detection. Appl Opt 2015;54(10):2885–90.
[4] Li Rui, Madampoulos Nicholas, Zhu Zhigang, Xie Liangping. 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(21):5011–8.
[5] Zhang Xin, Diao Weifeng, Liu Yuan, Zhu Xiaopeng, Yang Yan, Liu Jiqiao, et al. Eye-safe single-frequency single-mode polarized all-fiber pulsed laser with peak power of 361 W. Appl Opt 2014;53(11):2465–9.
[6] Weifeng Diao, Xin Zhang, Jiqiao Liu, Xiaopeng Zhu, Yuan Liu, Decang Bi, et al. All fiber pulsed coherent lidar development for wind profiles measurements in boundary layers. Chin Opt Lett 2014;12(7):072801.
[7] Jianhua Shang, Shuguang Zhao, Yan He, Weibiao Chen, Ning Jia. Experimental study on minimum resolvable velocity for heterodyne laser Doppler vibrometry. Chin Opt Lett 2011;9(8):081201.
[8] Jianhua Shang, Yan He, Dan Liu, Huaguo Zang, Weibiao Chen. Laser doppler vibrometer for real-time speech-signal acquirement. Chin. Opt. Lett. 2009;7(8):080732.
[9] Martin Bauer, Frank Ritter, Georg Siegmund. High-precision laser vibrometers based on digital Doppler-signal processing. Proc of SPIE 2002;4827:50–61.
[10] ZHANG Yu-ze, LI Li-yan, TONG You-wan, ZENG Hua-lin, ZHOU Yan. Vibration characteristics of aluminum material and its influences on laser doppler voice acquisition 2013. 8907.89074B-1.
[11] Tao Wang, Zhigang Zhu, Ajay Divakaran. Long range audio and audio-visual event detection using a laser doppler vibrometer. Proc of SPIE 2010. 7704. 77040J-1.
[12] Li Rui, Wang Tao, Zhu Zhigang, Xiao Wen. Vibration characteristics of various surfaces using an LDV for long-range voice acquisition. IEEE Sens J 2011;11(6):1415–22.
[13] Lutzmann P, Göhlera B, van Putten F, Hill CA. Laser vibration sensing: overview and applications. Proc of SPIE 2011. 8186. 818602-1.
[14] Yufu Qu, Wang Tao, Zhu Zhigang. Vision-aided laser doppler vibrometry for remote automatic voice detection. IEEE/ASME Trans Mechatron 2011;16(6):1110–9.