Optics and Lasers in Engineering 141 (2021) 106562

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

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

Remote eavesdropping at 200 meters distance based on laser feedback interferometry with single-photon sensitivity

Zhong Xu [a,b,#], Jiyang Li [a,#], Shulian Zhang [a], Yidong Tan [a,*], Xiliang Zhang [b], Xuling Lin [c], Xinjun Wan [d], Songlin Zhuang [d]

[a] The State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instrument, Tsinghua University, Beijing 100084, China
[b] School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China
[c] Beijing Institute of Space Mechanics and Electricity, Beijing 100094, China
[d] University of Shanghai for Science and Technology, Shanghai 200093, China

ARTICLE INFO

Keywords:
Single-photon sensitivity
Remote Eavesdropping
Laser feedback interferometry

ABSTRACT

A remote eavesdropping system based on the laser feedback interferometry is proposed. The diffuse objects neighboring the speaker can work as inconspicuous sensors. The modulated incident beam is scattered back to the laser cavity by the objects, forming the self-mixing interference. The huge amplification caused by the self-mixing interference provides the system with ultra-high sensitivity, even the feedback light whose intensity is -104dB lower than the laser output can still be sensed. The comprehensible voice reconstruction of the object at the distance of 200 m can be realized, although only 0.91 photon in each interference modulation cycle is scattered back into the laser cavity. Additionally, the range of the incident angle of the beam can be up to ±45 deg, and the optimal one is ±30 deg. This system can meet various requirements of acoustic signals detection and be potentially applied in fields as disaster relief and remote surveillance.

1. Introduction

The acquisition and reconstruction of remote voice is of great significance for sound field visualization, audio surveillance and intrusion detection [1-2]. The traditional methods of the sound recovery usually employ conventional speech sensors, such as condenser microphones. Lee S et al. present an ultrathin, conformable, and vibration-responsive electronic skin to sense voices quantitatively by detecting the acceleration of the skin [3]. Such sensors generally need to be placed close to the objects under surveillance, which are easy to be discovered and shielded, also inapplicable in long-distance detection. To enhance the working distance, parabolic microphones and millimeter microwave radar sensor are proposed, which can capture voice in the designated direction [4-5]. However, all the sound over the entire working distance will be captured. The undesirable noise will bring extra difficulty in signal processing. Davis A et al. have reconstructed the voice using only high-speed video of the object [6]. This passive measurement method is limited by the ambient brightness and the short working distance. Because of the complexity in the processing of the video, the real-time eavesdropping cannot be performed.

Due to the limitation of the above-mentioned methods, the usage of laser in voice detection has attracted many research interests. Compared with traditional voice acquisition technology, because of the directionality of the laser beam, the working distance can be greatly improved. Besides, the wavelength of the laser can be selected in the infrared range, which ensures the concealment of detection. Among the methods of the laser remote eavesdropping, laser Doppler vibrometer (LDV) is well-developed with high performance [7-11]. A narrow-linewidth laser is employed to ensure the long working distance, but its susceptibility to ambient temperature restricts the validity of the practical application [9]. Although the working distance of the sound acquisition system based on LDV is reported to be up to 300m, retroreflective tapes [10] or high-power laser (2W) [11] is utilized. Therefore, it is particularly critical to develop the eavesdropping scheme that only uses a relatively low-power laser to achieve long-range detection without target mirrors. Laser feedback interferometry (LFI), also known as self-mixing interferometry, is a demonstration of interferometry technology with reflected or scattered beams from external objects into the laser. Otsuka observed and systematically researched on the dynamic self-mixing modulation in class-B lasers, which lead to versatile self-mixing metrology for different physical effects and applications [12-13]. According to the laser feedback interference theory, the laser intensity, the polarization state and the phase position of a laser could be modified by introducing coherent optical feedback from an external surface. Weak changes of the returning light can be amplified by stimulated radiation automati-

* Corresponding author.
E-mail address: tanyd@tsinghua.edu.cn (Y. Tan).
# Zhong Xu and Jiyang Li contributed equally.

https://doi.org/10.1016/j.optlaseng.2021.106562
Received 26 September 2020; Received in revised form 27 December 2020; Accepted 25 January 2021
Available online 8 February 2021
0143-8166/

cally [13-14]. Ultra-high sensitivity makes LFI widely applied in fields as displacement measurement [15-17], velocity measurement [18-19], microdevice geometric structure measurement [20] and high-precision thermal expansion coefficient (TEC) determination of materials [21]. In addition, there has been a lot of researches on sound reproduction [22-24]. Otsuka et al. used a self-aligned optical feedback vibrometry technique with the self-mixing modulation effect in a laser-diode-pumped microchip solid-state laser to reproduce the nano-vibration measurement and highly sensitive sound of audio frequency speakers as early as 2003 [22]. Then, the vibration of objects placed 2.5 km away from the laser had been achieved through effective long-haul self-mixing interference effect of thin-slice $LiNdP_4O_{12}$(LNP) laser, whose linewidth was narrow and evaluated to be 16 kHz [23].

In this paper, a remote eavesdropping system based on LFI is developed. The laser beam impinges on the object and then is diffused back to laser cavity. The weak diffuse signal is amplified by stimulated radiation in the microchip laser cavity, with the ultrahigh gain factor up to 10^6. Benefiting from this amplification, 0.91 photon sensitivity per feedback cycle is achieved. In this scheme, only heterodyne demodulation method is utilized without any photon-correlation techniques which are widely used in LDV systems. For diffuse objects, clear speech within 100 m and recognizable one within 200 m are successfully obtained and eavesdropped. In addition, this eavesdropping system can support a range of incidence angle in $\pm$ 60° at the distance of 100 m, and the optimal performance of the system can be ensured in the range of $\pm$30° at the distance of 200 m. Based on the different qualities of the reconstructed speech utilizing various objects in the experiments, some suggestions on the selection of appropriate eavesdropping objects are also provided.

2. Principle for Audio Capture

When the voice stimulates an object, weak vibration in scales of micrometers or nanometers will occur. The sound is reconstructed by detecting this vibration. Fig. 1 shows the schematic diagram of the remote eavesdropping system based on LFI.

ML (Microchip Laser) is a solid microchip laser whose microchip material is Nd: YVO_4 (fluorescent lifetime, 9×10^{-5}s; photon lifetime, 3.63×10^{-11}s). The microchip is pumped by a semiconductor laser with a wavelength of 808nm to generate a 1064nm laser, and it works in a single-longitudinal mode. The fluorescence-to-photon lifetime is calculated to be 2.48×10^6 and the relative pumping level is 1.9. The output light of the laser is divided into two beams by a beam splitter (BS). The reflected light is detected by the photodetector (PD) and transferred to the electrical signals. The transmitted light is modulated by the two acousto-optic modulators (AOMs). Since the working frequency of the acousto-optic modulator is usually high, two AOMs are employed, and the frequency difference between them is used as the feedback frequency shift. Suppose the working frequency of the AOM_1 and AOM_2 is $\Omega_1=2\pi f_1$ and $\Omega_2=2\pi f_2$. If the frequency of the initial light emitted is υ,

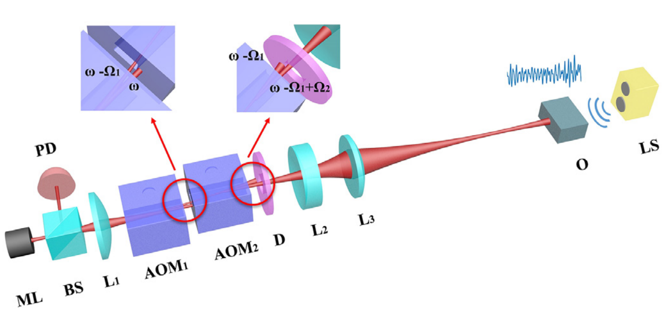

Fig. 1. Schematic diagram of the experimental setup. ML, microchip laser; BS, beam splitter; PD, photodetector; AOM_1 and AOM_2, acousto-optic modulators; D, diaphragm; L_1, L_2 and L_3, optical lens; O, object; LS, loudspeaker.

the transmitted light passes through AOM_1 and is diffracted at -1 order, which means the frequency is ω-Ω_1; then it passes through AOM_2, and is diffracted at +1 order such that the frequency now is ω-Ω_1+Ω_2=ω+Ω,. The measurement beam is reflected by the object and is diffracted again at +1 order in AOM_2 and at -1 order in AOM_1. As a result, the feedback measurement beam has the frequency of ω-Ω_1+Ω_2+Ω_2-Ω_1=ω+2Ω. The diaphragm(D) is used to retain the beam with the frequency shifting of $\Omega=2\pi f_0$. Due to the 10 mrad divergence of the solid-state laser, the effect of divergence is significant when the measurement distance reaches 200 m. Thus, a collimation system consisted of L_1, L_2 and L_3 is designed to improve the divergence angle of the laser beam, ensuring sufficient signal-to-noise ratio (SNR) for the laser feedback system in the case of long-distance measurement. L_1 not only collimates the laser beam, but also condenses the backscattered light from the object into the laser to increase the feedback light intensity, improving the performance of the laser feedback system. The weak light scattered by the object back to the cavity interferes with the light field in the cavity. The feedback light-induced laser intensity modulation ΔI can be expressed as [25]:

$$\frac{\Delta I(2\Omega)}{I} = \sqrt{\kappa}G(2\Omega)\cos(2\Omega t - \varphi + \phi), \tag{1}$$

where ΔI denotes the intensity modulation of the measurement light. I is the laser output power in steady state without feedback, κ is the effective feedback level,$\phi$is a fixed phase shift. The external cavity length can be regarded as the distance between microchip laser and object. The motion of the object is equivalent to the perturbations of the external cavity length. $\varphi$is the phase related to the external cavity length.$G$is the amplification coefficient, which can be expressed as:

$$G(2\Omega) = 2\gamma_c \frac{\left[\eta^2\gamma^2 + (2\Omega)^2\right]^{1/2}}{\left[\eta^2\gamma^2(2\Omega)^2 + \left(4\pi^2 f_r{}^2 - (2\Omega)^2\right)^2\right]^{1/2}}, \tag{2}$$

where γ_c is the decay rate of photons in laser cavity, $\gamma$is the decay rate of population inversion, $\eta$is the relative pumping level of the laser, and $f_r$is the relaxation oscillation frequency of the laser. For the solid-state lasers, the amplification coefficient G can reach to 10^6 [25]; thus, the laser frequency-shifted optical feedback system has ultra-high sensitivity to the feedback signals.

The laser output power modulated by the feedback light is detected by the PD and converted into electrical signals. The reference signal, generated by a mixer and a low-pass filter, is characterized with the frequency difference between two electrical driving signals for AOM1 and AOM2. As a result, the phase signal φ can be obtained. By collecting and demodulating the phase change $\Delta\varphi$, the corresponding change of the external cavity length ΔL can be obtained as:

$$\Delta L = \frac{\lambda \cdot \Delta\varphi}{4\pi}, \tag{3}$$

where λ is the wavelength of the laser. Therefore, by recording and analyzing the external cavity length change with the time, the displacement of the object caused by the sound pressure can be obtained and then be used to reconstruct the remote voice signals.

3. Single Photon Sensitivity

To estimate the feedback level κ, calculations are carried out using the principle of laser propagation. The schematic diagram is shown in Fig. 2.

The power of the laser hitting the object Pican be expressed as:

$$P_i = \frac{\mathrm{P}_o \cdot T_\alpha \cdot \mathrm{K} \cdot \mathrm{S}_o}{S_l}, \tag{4}$$

where $P_o$is the output power of the laser, $K$is the optical transmission efficiency determined by the transmittance of the optical lenses and the diffraction efficiency of the two AOMs, $T_\alpha$is the attenuation coefficient of laser atmospheric transmission,$S_o$is the area of object,$S_l$is the area

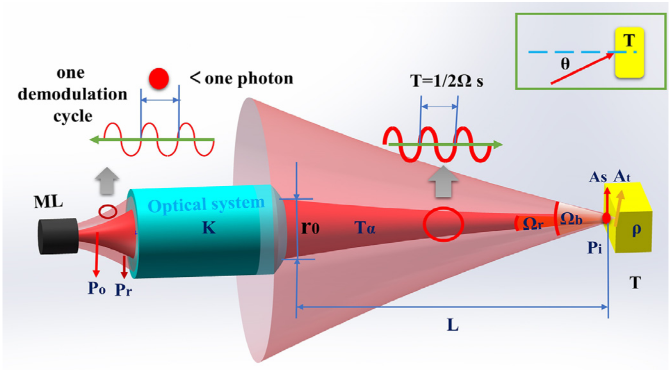

Fig. 2. Optical schematic of the eavesdropping system.ML, microchip laser; O, object.

of the object illuminated by the laser. The laser power received by the system $P_r$can be expressed as:

$$P_r = \frac{P_i \cdot \rho \cdot T_\alpha}{\Omega_b} \cdot \Omega_r \cdot K, \tag{5}$$

where$\rho$is the object reflectance.$\Omega_b$is backscattering angle. The measured object is supposed to be a lambert reflector, besides, the laser optical axis coincides with the normal of the object, making $\Omega_b=\pi$.$\Omega_r=A_r/R^2$is the angle of the pupil, and R is the working of the system. $A_r=\pi \cdot {r_o}^2$is the area of the receiving aperture, and $r_o$represents the emitted light spot radius. Besides, the area of the object is larger than the spot area, making $S_o=S_l$.The effective feedback level κcan be expressed as:

$$\kappa = \frac{{T_\alpha}^2 \cdot K^2 \cdot \rho \cdot {r_0}^2}{R^2}. \tag{6}$$

where T=$1/2f_0$ is feedback optical modulation period, f_0=1 MHz is frequency-shifted frequency. E_1=1.867×10^{-19} J is single photon energy at a wavelength of 1064 nm. At the maximum working distance e.g., 200 m, the effective feedback level is as low as -104dB, equivalent to 10^{-11}. The effective scattering light intensity return to the laser cavity is evaluated to be less than one photon (0.91 photon) per feedback modulation period of $1/2f_0$ μs. The detailed parameters of calculations about single photon sensitivity are provided in Table 1. When the laser beam illuminates the object at an oblique angle, n_p will be smaller. Such low energy is usually undetectable, but in this scheme, due to the ultrahigh sensitivity of LFI, enough signal intensity can still be achieved even if there is only one photon in the modulation period because of the amplification effect. According to Eq. (1), benefiting from the gain of G, the effective feedback level κ can be as low as 10^{-12} [26]. In other words, much higher SNR can be obtained compared with traditional interferometer when detecting low signal in single photon level. In addition, it can be seen from Eq. (6) that a large receiving aperture can theoretically improve the distance of detection. The receiving aperture of the system $2r_o$ is 12.4 mm, which is much smaller than the aperture of 100 mm in common LDV system. Therefore, at the same effective feedback level, a longer working distance can be obtained by increasing the size of the receiving aperture.

Table 1
Details parameters of Calculations about Single Photon Sensitivity

Parameter	Value
output power of the laser P_o	10 mW
transmittance of the optical lens $K_{L1}(K_{L2}, K_{L3})$	99.5%
transmittance of the BS K_{BS}	80%
diffraction efficiency of the AOM K_{AOM}	60%
optical transmission efficiency $K = {K_{L1}}^3 \times K_{BS} \times {K_{AOM}}^2$	0.2837
attenuation coefficient of laser atmospheric transmission T_α	0.9361
backscattering solid angle Ω_b	π
angle of the pupil Ω_r	3.019×10^{-9}
emitted light spot radius r_0	6.2 mm
working distance of the system R	200 m
object (paper box) reflectance ρ	0.5
per feedback modulation period of $T=1/2f_0$	0.5 μs
Single photon energy at a wavelength of 1064nm E_1	1.867×10^{-19} J

4. Experiment and Result

4.1. Sound Recovery for Single Frequency Signal

Voice usually consists of multiple frequency components. The quality of recovered voice depends on the reconstructed performance for each single frequency signal. Based on that, the evaluation system is set up, shown in Fig. 3 (a). To make the test close to real situation, a paper box, works as the surveillance object, which is also the effective diffuse surface of external cavity. Behind that, the loudspeaker under eavesdropping is set 0.5 m away, which plays the sound to motivate the paper box. Note that the loudspeaker and the object are separated into two platforms to avoid direct contact, and the motion of the object is only caused by the sound pressure. Part (1) is the LFI eavesdropping system, the focal lengths of the lenses L1, L2 and L3 in the system are 100mm, -25mm and 300mm respectively. In addition, the working frequency of the AOM1 and AOM2 is 70MHz and 71MHz, respectively, 1MHz is the difference of AOMs' frequency shift. Due to the frequency shift characteristics of the acousto-optic modulators, the laser reflected by the object can obtain a frequency shift of 2MHz in the measurement signal. The tiny responding motion of the paper box induces the change of measurement signal phase φ, which can be detected by the eavesdropping system. The distances of the experiments are 100 m, 150 m and 200 m, respectively. Due to the length limitation of the corridor, the reflector is used to provide a longer measurement distance.

The short-term Fourier transform (STFT) is a function of time and frequency, and it reflects the time-varying characteristics of the frequency spectrum of the speech signal. The height of each bar (color-enhanced) indicates the amplitude of the frequency within that band. The information of the reconstructed sound can be gained intuitively with this kind of visual. Fig .4(a) illustrates the spectrogram obtained by the STFT of the test sound, which consists of a series of tones. Each tone lasts for 0.5 s, ranging from 200 Hz to 3000 Hz with the interval of 100 Hz. The recovered spectrograms with the paper box at the distance of 100 m, 150 m and 200 m are expressed as Fig. 4(b-d). The frequency of tones can be distinguished and coincide with the test sound in spite of some unavoidable noises. In addition, it can be clearly seen from Fig. 4. that the energies of some frequency components are relatively low, which is caused by the frequency response characteristics of the object. This phenomenon will be discussed in detail in Section 5.2.

The frequency error is defined as the difference between the frequencies of recovered signals and the original one at the single frequency.

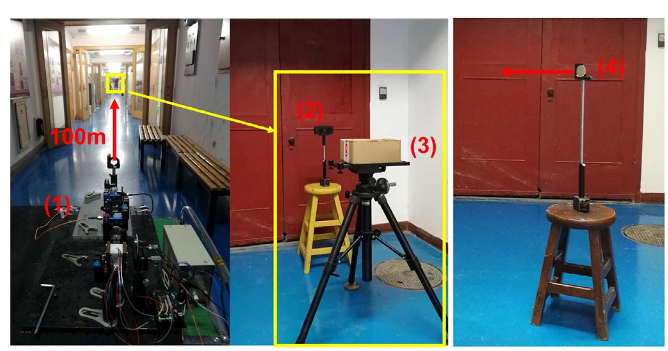

Fig. 3. Experimental setup of the LFI eavesdropping system (1). Sound from an audio source, such as a loudspeaker (2) stimulates an object (3) paper box, and the system (1) recover sound from the object. The reflector (4) deflects the light to the corridor on the other side.

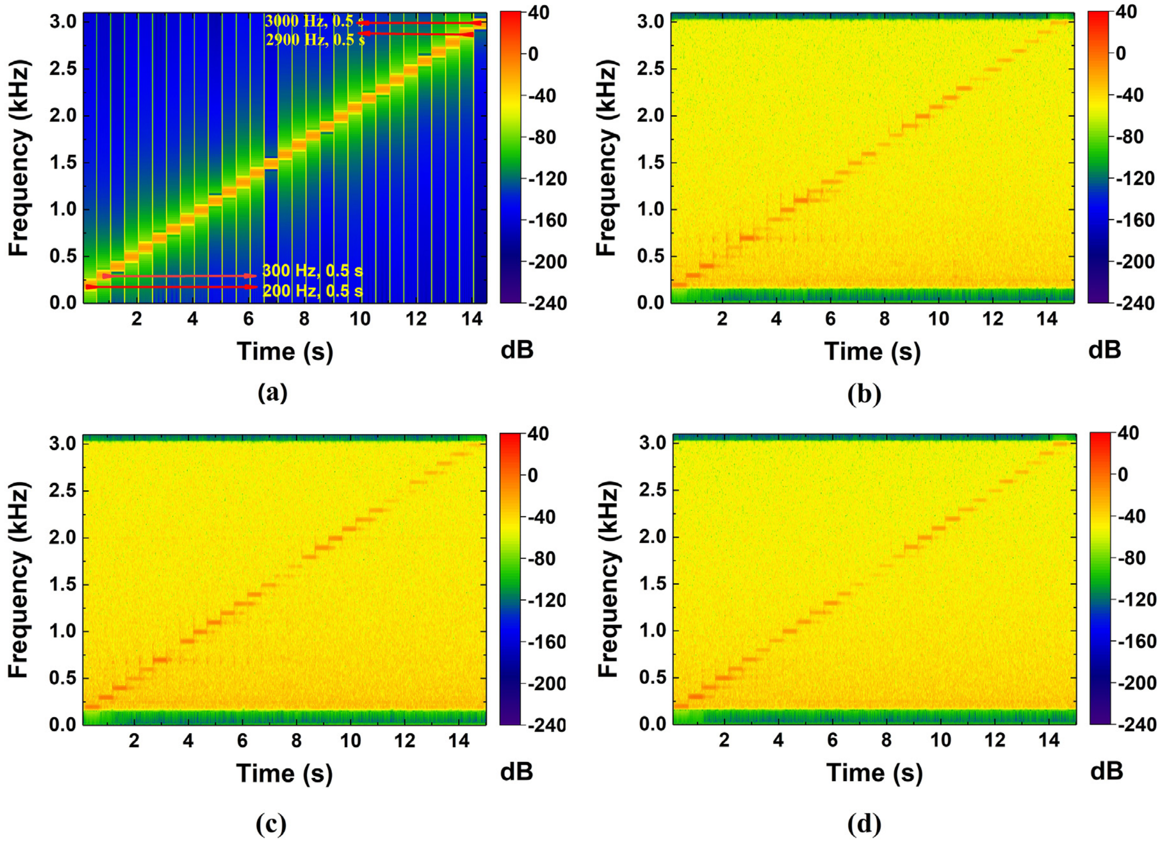

Fig. 4. The spectrograms of the test sound recovered in the different distances. (a) test sound spectrogram; (b) recovered spectrogram at 100 m (c) 150 m (d) 200 m (see **Visualization 1**).

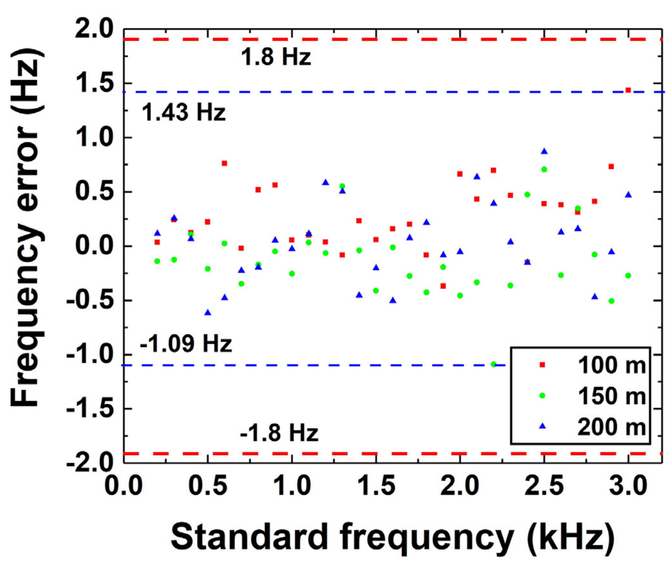

Fig. 5. Frequency error at the distance of 100 m, 150 m, 200 m.

The frequency errors of the signals recovered from different objects are shown in Fig. 5, and all of the frequency measuring errors are within ±1.43 Hz, and the maximum of relative frequency error $|\Delta f/f|$ is less than 0.13 %. Note that the frequency resolution of the human ear is $|\Delta f| \approx 1.8$Hz in the frequency range of 100-500 Hz. While in the frequency range of 500-15000 Hz, the relative frequency resolution is almost constant, which is $|\Delta f/f| \approx 3.5$ %. Therefore, the frequency accuracy of remote eavesdropping satisfies the measurement requirements, which ensures the faithfulness of the recovered sound. The results show that the remote eavesdropping system can accurately acquire single-tone signals at the distance of 200 m.

4.2. Voice Reconstruction

This set of experiments focus on recovering human speech. The loudspeaker plays the sentence “I look at the man and the women angrily", captured from a sample of New Concept English. The waveform and spectrogram of the voice mentioned above is depicted in Fig. 6 (a). The paper box is placed 100 m, 150 m, and 200 m respectively, working as the object under detection. The corresponding motion stimulated by the speaker can be obtained. A bandpass filter is applied to suppress the noise induced by surroundings. The passband expands from 200 Hz to 3000 Hz, which covers the frequency of the human voice. As can be seen from the results shown in Fig. 6 (b-d), with the increase of working distance, it is more difficult for the laser beam to focus on the object, so the returned signal weakens and background noise also increases. The same conclusion can be drawn from the corresponding audio in the appendix (Visualization 2), and the voice recovered directly at the distance of 200 m is still comprehensible. In addition, when the measurement distances are 100 m, 150 m, and 200 m, the number of photons returned to the cavity in a modulation period is 3.64, 1.62, and 0.91, respectively. Owing to the feedback amplification effect, the comprehensible voice can still be detected at the distance of 200 m.

To improve the voice quality, the original signal has been processed in the following. Firstly, enhancement algorithms based on spectral subtraction are used to suppress background noise [27]. Then, the noise-reduction is further improved by using audio processing software

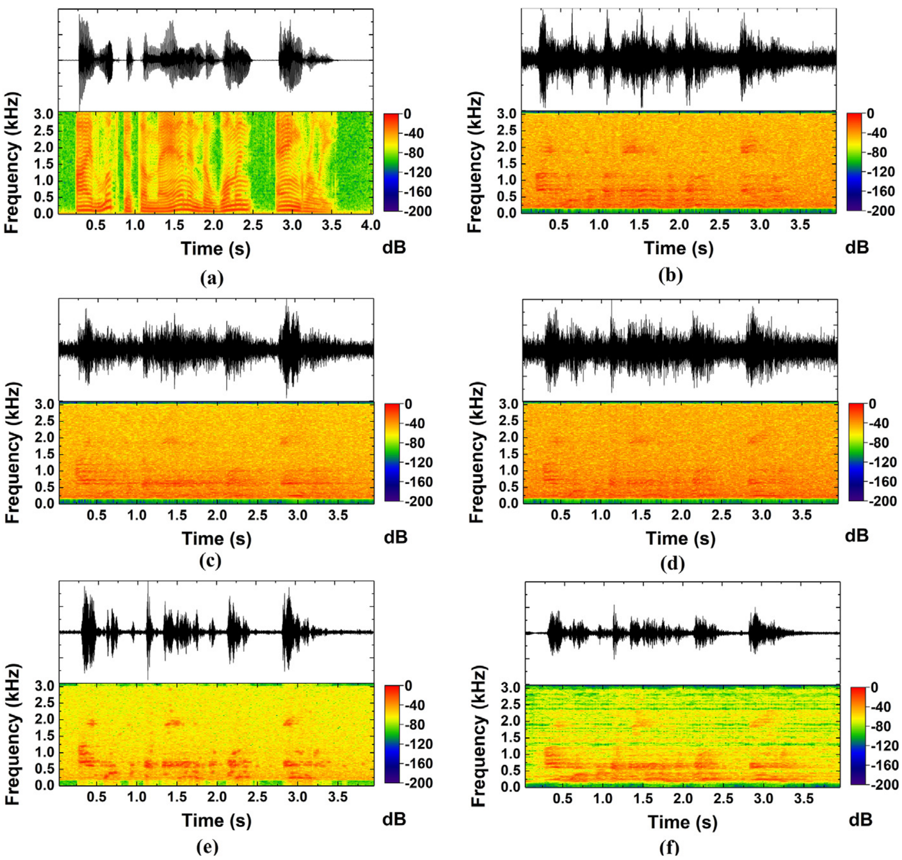

Fig. 6. Waveforms (upper) and speech spectrograms (lower) of the original voice and eavesdropping results. (a) the original voice; (b) the voice recovered directly at the distance of 100 m; (c) 150 m; (d) 200 m. (e) the voice after noise-reduced process with using spectral subtraction about (d); (f) the voice after noise-reduced process with using audio software about (d) (see **Visualization 2**). (All signals correspond to the voice of "I look at the men and women angrily".)

Table 2
The score standard of MOS

MOS score (W_i)	Quality Level	Description
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

(Adobe Audition from Adobe Inc). By Fig. 6(e) and (f), it can be seen that the background noise is significantly reduced by using spectral subtraction. However, owing to the random noise, it is possible that the value of the actual noise spectral line is greater than the value of the average noise spectral fitted by the previous blank sound segment in a certain period of time. The noise cannot be completely eliminated after subtraction, and the peak value of the noise is retained and forms a "music noise". Although with the help of audio software to reduce the noise, some noises still exist, as shown in Fig. 6(f), but with less distortion. In addition, ten volunteers are invited to assess the quality of recovered signal subjectively with the Mean Opinion Score (*MOS*) evaluation criterion, which is a common test in voice evaluation, with a five-point quality scale ranging between one and five [28-29], and this evaluation standard is shown in the Table 2. The final *MOS* score is calculated by:

$$MOS_{LQS} = \frac{1}{N}\sum_{i=1}^{p} W_i N_i, \tag{7}$$

where N is the total number of votes, N_i is the number of a particular score, W_i is the score for each vote, i is the score for each level, and p is the total score level which is 5. This assessment method can directly reflect the subjective feelings of the listeners. The *MOS* scores of the six voices mentioned above are 5, 3, 3, 2.5, 3.5 and 4, and consistent with the analysis above.

5. Discussion on the experimental system

5.1. Angles of Incident Beam

During the eavesdropping process, due to the influences of the surrounding environment such as the obstacles, the remote voice cannot be guaranteed to be acquired at a normal incidence angle each time. As the system can only react to the photons returned to the laser cavity in the direction of the incident light, when the incident angle is too large, the intensity of the signal is weakened, so it is necessary to study the reconstructed characteristics in different incident angles.

To study this issue, the test sound (mentioned in section 4.1) is used to stimulate the paper box at different incident angles, and the results are shown in Fig. 7. Almost all the errors of the measuring frequency are within ±1.8Hz, which indicates that the remote voice detection system is suitable at flexible incident angles from 0 deg to 45 deg at the distance of 200 m. When the working distance is 100 m, the incident angle can reach 60 deg.

The results of the voice signal recovered at multi-angles on the paper box are shown in Fig. 8. The noise of reconstructed signals increases gradually with the increase of the incidence angle. It can be seen from the recovered voice spectrogram that the bright-line, representing the voice harmonic, becomes blurred at the incident angle of 45 deg. However, the audio is still distinguishable. The quality of the reconstructed audio is better when the angle of incidence is less than 30 deg. Furthermore, the voices are also recovered with the detected objects being polyfoam, sponge and milk powder bag at different incident angles, and the voice signal obtained is similar to that of the paper box. Therefore, the optimal range of the incidence angle is within ±30 deg. The sounds

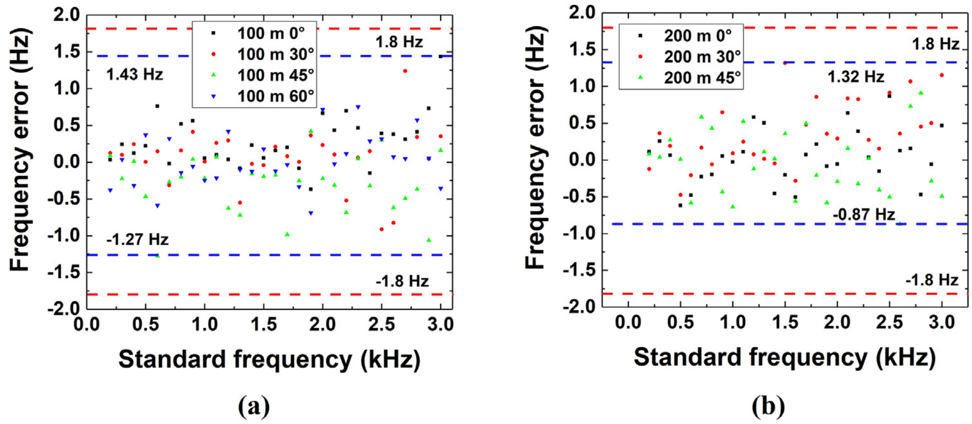

Fig. 7. The frequency error at different detection distance. (a) 100 m; (b) 200 m.

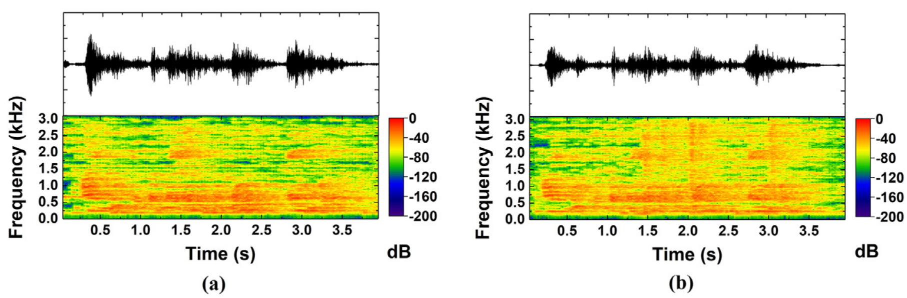

Fig. 8. Waveforms (upper) and speech spectrograms (lower) of the voice recovered from the paper box at the distance of 200 m at different angles. (a) recovered at 30°; (b) recovered at 45° (see **Visualization 3**). (All signals correspond to the voice of “I look at the men and women angrily”.)

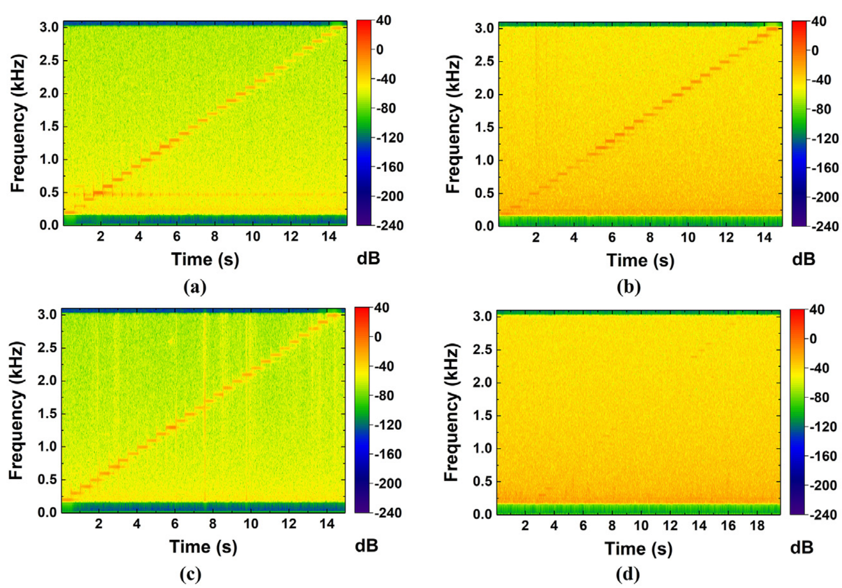

Fig. 9. Sound recovered spectrogram with different objects. (a) polyfoam; (b) sponge; (c) milk powder bag; and (d) door (see **Visualization 4**).

reconstructed from the paper box at different angles at the distances of 100 m are also provided in the appendix (Visualization 3).

5.2. Sound Recovery from Different Objects

To verify the validity of the interception system for different objects, the paper box is replaced by polyfoam, sponge, milk powder bag and wooden door respectively, maintaining the working distance of 200 m. The corresponding spectrograms are listed as Fig. 9(a-d).

By comparison, it can be seen in Fig. 9(d) that the sound can be hardly recovered from the door because the heavy weight restricts the amplitude of the motion. Other objects have good performances in low frequency range. For high frequency recovery, paper box and polyfoam exhibits obvious energy loss. High frequency components of the original sound stimulate the object to produce faster vibrations. It is easier for lighter objects to provide a more definite response, i.e. larger vibration amplitude.

Additionally, voice recovery with different objects is carried out to evaluate their performance in remote voice detection. The results are shown in Fig. 10(a-c), and based on that the properties for each object can be summarized. The polyfoam has a desirable response in the low-mid frequency part. Compared with recovery from the paper box, it contains more details for human hearing and understanding. Though the sponge has a good response throughout the speech frequency bandwidth, the reconstructed speech is partly buried in the noise, since the porous structure induced energy absorption. The milk powder bag has

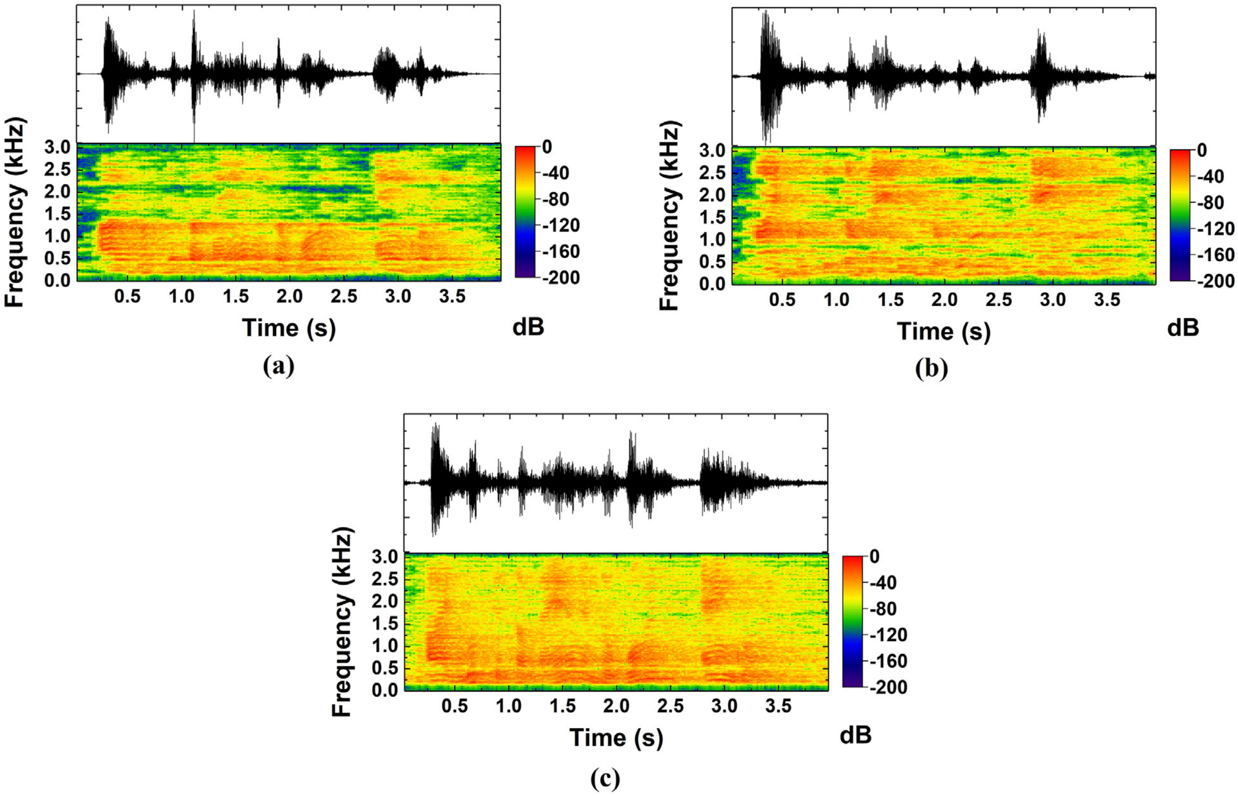

Fig. 10. Waveforms (upper) and speech spectrograms (lower); The noise-reduced voice recovered from (a) the polyfoam, (b) the sponge, (c) the milk powder bag (see **Visualization 5**). (All signals correspond to the voice of "I look at the men and women angrily".)

the best performance in the whole frequency range. For thin-film objects similar to the milk powder bag, the reconstructed signal maintains both strong background and necessary details, which make it easier to be understood. These conclusions could provide guidance for eavesdropping. The reconstructed voice from the objects like polyfoam and milk powder bag will give better performances. Especially for the milk powder bag, it has a better response in whole frequency bandwidth, which means that the voice recovered from thin-film objects will have higher intelligibility. For remote interception, the milk powder bag is more proper. However, on account of the highest intensity voice band ranging from 300 to 500Hz, so the objects that respond to low-mid frequencies still meet the requirements of monitoring.

5.3. Influence of the Atmospheric environment

When the working distance is far enough, such as 200 meters, we shall take the environmental disturbance into consideration. According to the Edlen formula, the air refractive index should not value as 1 considering the changes of environmental temperature, pressure and humidity during the optical path. The laser wavelength drift also affects the vibration accuracy of the voice source. Therefore, the measurement uncertainty is calculated as:

$$u(\Delta L) = \sqrt{\left[\frac{\partial \Delta L}{\partial \Delta \varphi} \cdot u(\Delta\varphi)\right]^2 + \left[\frac{\partial \Delta L}{\partial \lambda} \cdot u(\lambda)\right]^2 + \left[\frac{\partial \Delta L}{\partial n} \cdot u(n)\right]^2}, \quad (8)$$

where n denotes the air refractive index.

The measured wavelength uncertain of the used laser is small as to 10^{-12}, while the air refractive index uncertainty is 3×10^{-8} in the normal laboratory environment. When the eavesdropping environment is not quiet enough, to be specific, in the noisy industry or market, the system will have difficulty in reconstructing the target voice. On the other hand, this also proves the high sensitivity of the proposed instrument.

Actually, our system proves to be effective in a normal environment and the retrieval process has employed some ways such as applying a bandpass filter to decrease the background environmental influence. However, environmental disturbance always exists when using an eavesdropping sensor. There are two feasible solutions. One is to use the laser beam which has only undergone single frequency shift as a reference signal shown in formula (9), and select some objects which are not so easily disturbed by vibration around the object to measure the phase fluctuation of the reference signal as the interference of atmosphere for compensation. The compensation process is detailed in formula (10). For the measured and reference signals are quasi-common-path, the air refractive index and external cavity length are basically the same, thus the environmental influence will be greatly decreased after the compensation. The modified sound vibration $\Delta L$is given by the formula (11).

$$\frac{\Delta I}{I} \propto G(2\Omega)\cos(2\Omega t - \varphi_m + \phi_0) + G(\Omega)\cos(\Omega t - \varphi_r + \phi_0), \quad (9)$$

where φ_m is the measured phase and φ_r is the reference phase.

$$\Delta\phi_m - \Delta\phi_r = \frac{2n\omega}{c}\Delta L + \left(\frac{2\omega L_1}{c}\Delta n_1 - \frac{2\omega L_2}{c}\Delta n_2\right) + \frac{2n(L_1 - L_2)}{c}\Delta\omega, \quad (10)$$

where ω denotes the laser angular frequency.

$$\Delta L = \frac{c}{2n\omega}(\Delta\phi_m - \Delta\phi_r) = \frac{\lambda}{4\pi}(\Delta\phi_m - \Delta\phi_r), \quad (11)$$

The second solution is to use two semiconductor lasers to pump the same microchip to obtain two parallel beams of measurement light. Same as the above principle, one of the measurement signals is used as compensation to improve the sound recovery quality.

6. Conclusion

Remote eavesdropping system with ultrahigh sensitivity based on LFI has been demonstrated. The effective feedback level is below one photon per feedback cycle. The high-quality voice signals at the distance of 100 m and comprehensible voice signals at the distance of 200 m are obtained successfully. The system does not require any electric photon-correlation techniques since the microchip laser itself acts as a mixer-oscillator to amplify the weak feedback light. In the future, we will focus on the performance enhancement of remote laser voice acquisition system, such as improving the receiving aperture, taking full advantage of effective feedback levels, which can be as low as 10^{-12}, and so on, to meet various requirements of voice detection and toward broad application prospects in engineering.

Author Statement

Yidong Tan put forward the concept of the eavesdropping techniques by laser feedback interferometry. Zhong Xu and Yidong Tan designed the experiments. Jiyang Li helped to construct and modify the prototype of the experiments. Zhong Xu, Jiyang Li, Xiliang Zhang, Shulian Zhang, Yidong Tan, Xuling Lin, Xinjun Wan and Songlin Zhuang performed experiments and analysis. Zhong Xu, and Jiyang Li wrote the paper.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the National Science Fund for Excellent Young Scholars under Grant 51722506, in part by the National Natural Science Foundation of China under Grant 51961130387, in part by the Royal Society-Newton Advanced Fellowship under Grant 191072, in part by the grant (Code No. 2019M660624) from China Postdoctoral Science Foundation, and in part by the CASC youth talent support program(201981).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.optlaseng.2021.106562.

References

[1] Rothberg S J, Allen M S, Castellini P, et al. An international review of laser Doppler vibrometry: Making light work of vibration measurement[J]. Optics and Lasers in Engineering 2017;99:11–22.
[2] Zieger C, Brutti A, Svaizer P. Acoustic based surveillance system for intrusion detection[C]. In: 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE; 2009. p. 314–19.
[3] Lee S, Kim J, Yun I, et al. An ultrathin conformable vibration-responsive electronic skin for quantitative vocal recognition[J]. Nature communications 2019;10(1):2468.
[4] Jiao M, Lu G, Jing X, et al. A novel radar sensor for the non-contact detection of speech signals[J]. Sensors 2010;10(5):4622–33.
[5] Tian Y, Li S, Lv H, et al. Smart radar sensor for speech detection and enhancement[J]. Sensors and Actuators A: Physical 2013;191:99–104.
[6] Davis A, Rubinstein M, Wadhwa N, et al. The visual microphone: passive recovery of sound from video[J]. ACM Trans. Graph 2014;33(4) Article 79.
[7] Vanherzeele J, Vanlanduit S, Guillaume P. Acoustic source identification using a scanning laser Doppler vibrometer[J]. Optics and lasers in engineering 2007;45(6):742–9.
[8] Avargel Y, Cohen I. Speech measurements using a laser Doppler vibrometer sensor: Application to speech enhancement[C]. In: 2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays. IEEE; 2011. p. 109–14.
[9] Lv T, Guo J, Zhang H, et al. Acquirement and enhancement of remote speech signals[J]. Optoelectronics Letters 2017;13(4):275–8.
[10] Zhu Z, Li W, Wolberg G. Integrating LDV Audio and IR Video for Remote Multimodal Surveillance[C]. Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on. IEEE; 2005.
[11] Xiao W. Research on Long-distance Laser Vibration Measurement Technology. University of Electronic Science and Technology of China; 2018. Ph.D.2018.
[12] Otsuka K. Effects of external perturbations on $LiNdP_4O_{12}$ lasers[J]. Quantum Electronics, IEEE Journal of 1979;15(7):655–63.
[13] Otsuka K. Self-Mixing Thin-Slice Solid-State Laser Metrology[J]. Sensors 2011;11(2):2195–245.
[14] Ko JY, Otsuka K, Kubota T. Quantum-noise-induced order in lasers placed in chaotic oscillation by frequency-shifted feedback[J]. Physical review letters 2001;86(18):4025.
[15] Chen J, Zhu H, Xia W, et al. Synthetic-wavelength self-mixing interferometry for displacement measurement[J]. Optics Communications 2016;368:73–80.
[16] Zhu K, Guo B, Lu Y, et al. Single-spot two-dimensional displacement measurement based on self-mixing interferometry[J]. Optica 2017;4(7):729–35.
[17] Tan Y, Zhang S, Zhang S, et al. Response of microchip solid-state laser to external frequency-shifted feedback and its applications[J]. Scientific reports 2013;3:2912.
[18] Otsuka K. Self-mixing thin-slice solid-state laser Doppler velocimetry with much less than one feedback photon per Doppler cycle[J]. Optics letters 2015;40(20):4603–6.
[19] Wu S, Wang D, Xiang R, et al. All-fiber configuration laser self-mixing Doppler velocimeter based on distributed feedback fiber laser[J]. Sensors 2016;16(8):1179.
[20] Xu C, Tan Y, Zhang S, et al. The structure measurement of micro-electro-mechanical system devices by the optical feedback tomography technology[J]. Applied Physics Letters 2013;102(22):221902.
[21] Zheng F, Tan Y, Lin J, et al. Study of non-contact measurement of the thermal expansion coefficients of materials based on laser feedback interferometry[J]. Review of Scientific Instruments 2015;86(4):043109.
[22] Abe K, Otsuka K, Ko JY. Self-mixing laser Doppler vibrometry with high optical sensitivity: application to real-time sound reproduction[J]. New Journal of Physics 2003;5:8.
[23] Otsuka K. Long-haul self-mixing interference and remote sensing of a distant moving target with a thin-slice solid-state laser. Optics letters 2014;39(4):1069–72.
[24] Sudol S, Miyasaka Y, Otsuka K, et al. Quick and easy measurement of particle size of Brownian particles and plankton in water using a self-mixing laser[J]. Optics Express 2006;14(3):1044–54.
[25] Zhang S, Zhang S, Tan Y, et al. Common-path heterodyne self-mixing interferometry with polarization and frequency multiplexing[J]. Optics letters 2016;41(20):4827–30.
[26] Lacot E, Day R, Stoeckel F. Coherent laser detection by frequency-shifted optical feedback[J]. Physical Review A 2001;64(4):043815.
[27] Boll S. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Transactions on acoustics, speech, and signal processing 1979;27(2):113–20.
[28] Salza P L, Foti E, Nebbia L, et al. MOS and pair comparison combined methods for quality evaluation of text-to-speech systems[J]. Acta Acustica united with Acustica 1996;82(4):650–6.
[29] MolinaGarcia Mariano, Alonso JI. Evolution of mean opinion Score due to handover events in IEEE 802. 11 Wlan networks[C]. ieee Radio wireless Symposium IEEE 2009;63(5):207–10.