Vibration Characteristics of Various Surfaces Using an LDV for Long-Range Voice Acquisition

Rui Li, Tao Wang, *Student Member, IEEE*, Zhigang Zhu, *Senior Member, IEEE*, and Wen Xiao

***Abstract*—Laser Doppler vibrometers (LDVs) as non-contact vibration sensors have an ability of remote voice acquisition. With the assistance of a visual sensor (camera), various targets in the environment, where an audio event takes place, can be selected as reflecting surfaces for collecting acoustic signals by an LDV. The performance of the LDV greatly depends on the vibration characteristics of the selected targets (surfaces) in the scene, on which a laser beam strikes and from which it returns. In this paper, the relations of a target's material and structural properties with its vibration characteristics are studied, and a vibration amplitude model is established. Then the vibration characteristics of several typical surfaces with different materials and structures are explored through both simulations and real-sensor experiments. Based on their responses to the frequencies in the range of human voice, the targets are classified into three categories by the number of fluctuations (zero, one, or two) in their vibration returns in the range of speech. Some short- and long-range experimental results are presented for the speech acquisition from surfaces of these three categories, and their feasibilities in speech acquisition are also evaluated.**

***Index Terms*—Laser Doppler vibrometer (LDV), long-range voice acquisition, multimodal sensors, vibration characteristics.**

I. Introduction

MULTIMODAL sensory systems are attracting more attention for large area surveillance, perimeter protection, search and rescue, and structural health monitoring. In the last few years, a few systems [1], [2] have been reported to integrate visual and acoustic sensors. This multimodal capability would greatly increase the acquired information through both watching and listening of the targets that are presenting in a scene. While visual and thermal sensors have long-range capabilities, in these systems, the acoustic sensors have to be close to the subjects in monitoring. For long-range applications, remote voice acquisition (hearing) is one of the key problems to be solved. A commercial Laser Doppler vibrometer (LDV), for example the one by Polytec, has been used to capture audio signals in long distances, for multimodal long range surveillance [3]. The voice signals could be acquired by capturing the vibrations of target surfaces that are caused by the speech of a person next to the targets.

In the previous work, for both indoor and outdoor experiments, the audio signals have been obtained from both short distances and long distances [3]. It has been shown that the performance of the LDV acoustic detection strongly depends on both the reflectance and the vibration properties of the targets surfaces that are selected for the laser beam to be directed to, as well as the distance between the targets and the sensor. In order to improve the detection performance, different aspects of the problems have been studied and corresponding solution have been proposed. Li [4] proposed a speech processing technique, which is a combination of several typical signal processing techniques (bandpass filtering, Wiener filtering, and volume adaptation), to enhance the audio signals. Although this technique can be used to improve the intelligibility of many of the obtained signals from different targets surfaces, it does not study the causes of the problems that affect the quality of the acquired signal: the reflectance and the vibration properties of the targets surfaces. Hence, in some cases, particularly for long-distance voice detection, this signal processing approach does not work very well because of the poor reflection of the LDV laser beam due to the surface characterization and the poor vibration of the targets due to their mass. In another work, Qu [5] proposed a target surface selection and automatic laser focusing method to improve the LDV's performance for the long-range detection. The authors designed a vision-aided system to the LDV, consisting of a pan–tilt–zoom camera and a mirror on a pan–tilt platform. The reflection properties of surfaces in the camera's field of view are compared by analyzing their reflected laser intensities. The target surface with the highest reflected intensity is chosen and the laser beam is directed and focused on to it. The high reflected intensity leads to a high signal-to-noise ratio (SNR) for the acquired signal, and thus, the improved performance. However, this target surface selection method only considers the surface reflectance, which is not the only factor to determine the LDV's performance. The vibration characteristics of the targets are another key impact factor for LDV voice acquisition, which have not been studied. As one important aspect of the study of scene phenomenology for long-range voice

Manuscript received July 28, 2010; accepted November 10, 2010. Date of publication November 18, 2010; date of current version April 20, 2011. The first author is supported by a State Scholarship Fund from the China Scholarship Council of the Ministry of Education of P. R. China. This work is also partially supported by a PSC-CUNY Award, National Collegiate Inventors and Innovators Alliance (NCIIA) under an Advanced E-Team grant, and by Air Force Office of Scientific Research (AFOSR) under Award #FA9550-08-1-0199. The work was done while R. Li was visiting the Department of Computer Science, CUNY City College, New York, NY. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Okyay Kaynak.

R. Li is with the School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, and the First Research Institute of Ministry of Public Security of P.R.C., Beijing 100048, China (e-mail: lirui5434@gmail.com).

T. Wang and Z. Zhu are with the Department of Computer Science, CUNY City College, New York, NY 10031 USA, and the Department of Compute Science, CUNY Graduate Center, New York, NY 10016 USA (e-mail: twang@cs.ccny.cuny.edu; zhu@cs.ccny.cuny.edu).

W. Xiao is with the School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China (e-mail: xiaow@buaa.edu.cn).

Digital Object Identifier 10.1109/JSEN.2010.2093125

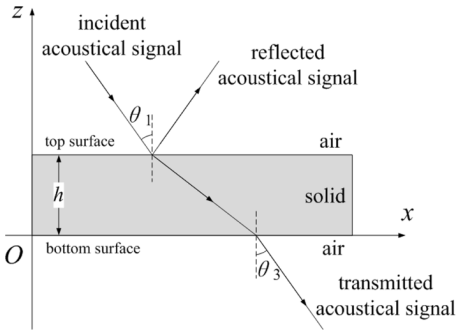


Fig. 1. Interaction between acoustical signal and plate.

acquisition system, the focus of this paper is on the analysis of the vibration characteristics of different targets and their influences on LDV remote voice acquisition, which could give some valuable suggestions on performance improvement through the selection of targets with good vibration characteristics.

This paper is organized as follows. Section II describes the vibration characteristics of target surfaces through modeling and simulations. In Section III, the frequency responses of different targets are analyzed, and the real speech signals acquired from them using an LDV are presented, by adopting several sets of experiments. Finally, we conclude our work and discuss some future research directions in Section IV.

II. Modeling Vibration of Target Surfaces

In the previous work on the acoustic sensing by detecting the target vibration caused by the acoustical signals, most of them focus on signal processing and system design issues. The inherent characteristic of target surface vibration as acquired by a vibration sensor has not been well studied. This will be the focus of this paper. Hence, in this section, a model of the target surface vibrations caused by acoustical signals is established by using the physical principle of the structural acoustic insulation. Then a theoretical analysis for understanding the affect of the vibration for the LDV voice acquisition is presented.

A. Target Surface Vibration Model

For the remote voice acquisition by an LDV, vibration measurements are made at the point where the laser beam strikes the target under the vibration caused by a voice source. Usually, the stricken targets have the structure of *plates*. Such targets include walls, doors, metal boxes, traffic signs, building pillars, containers, and so on [3]. The interaction between acoustical signals and a plate is illustrated in Fig. 1. The structure of the plate is assumed to be a single plate, which is a layer of elastic solid medium bounded by two layers of acoustic flat fluid media (such as air) [6].

The acoustic signal, considered to be a sinusoidal harmonic pressure load, impinges to the top surface of the plate. The incident acoustic pressure potential on and above the top surface can be represented as [6]

$$\varphi_{\text{in}} = Ae^{jk_0[x_1 \sin\theta_1 - (z_1 - h)\cos\theta_1]} \tag{1}$$

where A is the amplitude of the acoustic pressure potential, k_0 is the wave number (which is $2\pi f/c_0$), f is the frequency of the acoustical signal, c_0 is the acoustic velocity in air, x_1 and z_1 are the coordinates of a point in the air layer which is on or above the top surface of the plate, h is the thickness of the plate, and θ_1 is the incident angle of the acoustical signal. The acoustic signal on the top surface of the plate is divided into two parts: one is reflected, and the other one is transmitted through the bottom surface. For acquiring voice signals using an LDV off the top surface of a target, we are interested in understanding the vibration properties of the top surface.

If the reflection rate and transmission rate are denoted as V and D, then the acoustic pressure potential on or above the top surface, which is composed by two parts (incident acoustic pressure potential and reflected acoustic pressure potential), can be expressed as

$$\Phi = \varphi_{\text{in}} + \varphi_{\text{re}} = Ae^{jk_0[x\sin\theta_1 - (z-h)\cos\theta_1]} + AVe^{jk_0[x\sin\theta_1 + (z-h)\cos\theta_1]} \tag{2}$$

where φ_{re} is the reflected acoustic pressure potential. The transmitted acoustic pressure potential on and below the bottom surface is modeled as

$$\varphi_{\text{tran}} = ADe^{jk_0[x_2\sin\theta_3 - z_2\cos\theta_3]} \tag{3}$$

where θ_3 is the emergence angle of the transmitted acoustical signal on the bottom of the plate, x_2 and z_2 are the coordinates of a point in the air layer which is on or below the bottom surface of the plate.

The relationship between the acoustic pressure potential Φ and the vibration amplitude u along the z axis can be calculated by [6]

$$u = -\frac{1}{j\omega}\frac{d\Phi}{dz} \tag{4}$$

where $\omega = 2\pi f$ is angular frequency. Hence, the vibration amplitude of the top surface (where $z_1 = h$) can be derived as

$$u_{\text{top}} = -\frac{A\cos\theta_1}{c_0}(1 - V) = -\frac{p\cos\theta_1}{2\pi f\rho_0 c_0}(1 - V) \tag{5}$$

where p is the sound pressure of the acoustical signal, and ρ_0 is the density of the air. For acquiring voice signals using an LDV from the surface of a target, we are more interested in obtaining the vibration magnitude of the top surface of the target. However, it is very hard to calculate the reflection rate V due to its complicated expression as shown in [6]. Fortunately, a lot of research has been performed [7]–[10] in the so-called *sound transmission loss* (STL) of a single plate, which is expressed as

$$R = -20\log_{10} D. \tag{6}$$

Meanwhile, based on the law of energy conservation, we have

$$V^2 + D^2 = 1. \tag{7}$$

Hence, we can make use of (6) and (7) to simplify the problem by having the key parameter V going through some transformations. Obviously, R is the key parameter for this transformation.

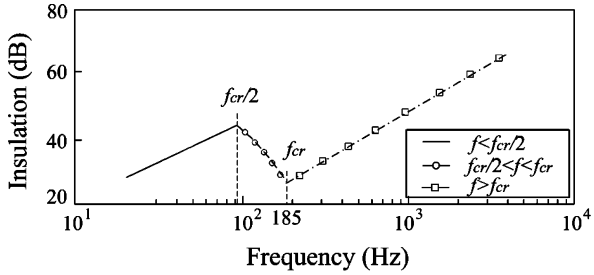


Fig. 2. STL curve of 10-cm-thick concrete.

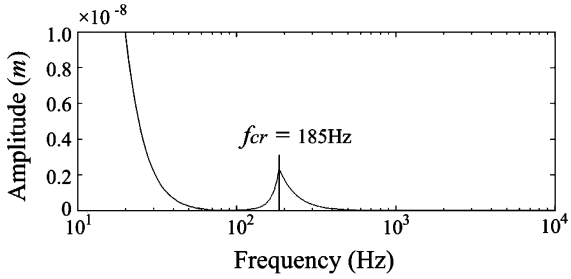


Fig. 3. Vibration amplitude of top surface of a plate.

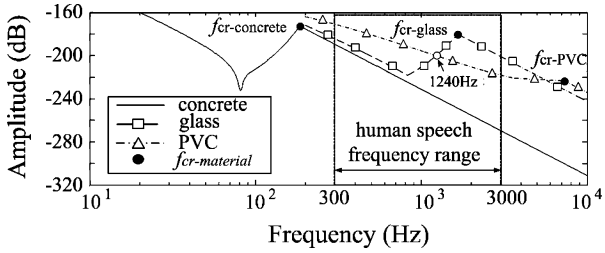


Fig. 4. Vibration amplitudes on the top layers of concrete, glass and PVC.

B. Acoustic Insulation Model

In this part, a Sewell–Sharp–Cremer (SSC) model, which was used by Callister [9] to evaluate the sound transmission loss of a single-layer panel, was adopted into our model to calculate the sound transmission loss R. In the SSC model, the whole frequency range is divided into three parts: $f < f_{\rm cr}/2$, $f \leqslant f_{\rm cr}/2 \leqslant f_{\rm cr}$, and $f > f_{\rm cr}$, where $f_{\rm cr}$ is the *coincidence frequency* [7] and is expressed as

$$f_{\rm cr} = \frac{c_0^2}{2\pi}\sqrt{\frac{12\mu(1-\nu^2)}{Eh^3}} \tag{8}$$

where $\mu = \rho h$ is the mass per unit area of the plate, ρ is the density of mass, ν is the Poisson ratio of the panel, E is Young's modulus, and h is the plate thickness. We will discuss each of the three parts in the following paragraph.

1) $f < f_{\rm cr}/2$: In this part, the sound transmission loss of the plate can be calculated by the Law of Theoretic Mass [7], as

$$R = -10\log_{10}\left\{\frac{\ln(k\sqrt{S}) + 0.16 - U(\Lambda) + \frac{1}{4\pi S k_1^2}}{[(\frac{\mu\pi f}{\rho_0 c_0})(1-\frac{f^2}{f_{\rm cr}^2})]^2}\right\} \tag{9}$$

where $k_1 = 2\pi f/\beta$ is the acoustic wave number in mass, β is the speed of the transverse wave, S is the area of the plate, Λ is the ratio of length and width of the plate (L/W), and $U(\Lambda)$ is a shape factor correction for non-square plates. A useful empirical expression for $U(\Lambda)$ adapted from Sewell [9] is

$$U(\Lambda) = -0.0000311\Lambda^5 + 0.000941\Lambda^4 - 0.0107\Lambda^3 + 0.0526\Lambda^2 - 0.0407\Lambda - 0.00534. \tag{10}$$

2) $f > f_{\rm cr}$: In this part, the sound transmission loss of the plate can be calculated by the following model [9]:

$$R = 20\log_{10}\left(\frac{\mu\pi f}{\rho_0 c_0}\right) + 10\log_{10}\left(\frac{2\eta f}{f_{\rm cr}}\right) - 5 \tag{11}$$

where η is the damping loss factor of the panel.

3) $f \leqslant f_{\rm cr}/2 \leqslant f_{\rm cr}$: In this part, a linear interpolation scheme is used between the mass law STL shown by (9) at one-half of the coincidence frequency and the STL shown by (11) at the coincidence frequency [9].

Fig. 2 shows an STL curve of a 10-cm-thick concrete wall using the above model. Combing the STL estimations within different frequency ranges, the vibration amplitude of the plate's top surface can be calculated. The simulation result for the concrete wall is shown in Fig. 3. In the whole frequency range, the vibration amplitude decreases very quickly with the increase of the frequency. Around the coincidence frequency $f_{\rm cr}$, however, there is a obvious peak in the amplitude with the maximum obtained at the coincidence frequency.

TABLE I
CHARACTERISTIC PARAMETERS OF MATERIALS

Materials / Parameters	Concrete	Glass	PVC
E (Gpa)	29	72	3.7
ν	0.15	0.22	0.4
ρ (×10^{3}kg/m^3)	2.5	2.5	1.44
η	0.006	0.004	0.04
β (m/s)	2250	3436	1200
h (cm)	10	0.7	0.5
S (m×m)	5×3	0.4×0.4	2×1.5

C. Simulation Analysis of Different Materials

As reported in [7] and [8], the coincidence frequency and sound transmission loss will change with different plate materials. Fig. 4 shows the vibration amplitudes on the top surfaces of three plates of different materials: concrete, glass and PVC. The characteristic parameters of these three materials are cited from [7], [8], and [10] and are listed in Table I.

In Fig. 4, the horizontal axis is the frequency (in Hertz) and the vertical axis is the amplitude (in decibels) in order to highlight the fluctuation. The curves of vibration amplitudes on the top surfaces of different panels are shown in different styles. And their coincidence frequencies are indicated in dots, and marked as $f_{\rm cr}$-[material], for example $f_{\rm cr-concrete}$. A rectangular box is plotted to indicate the main frequency range of human speech, from 300 to 3000 Hz. From Fig. 4, we have the following observations.

First, only the coincidence frequency of the glass plate is in the range of speech. The coincidence frequency of the concrete

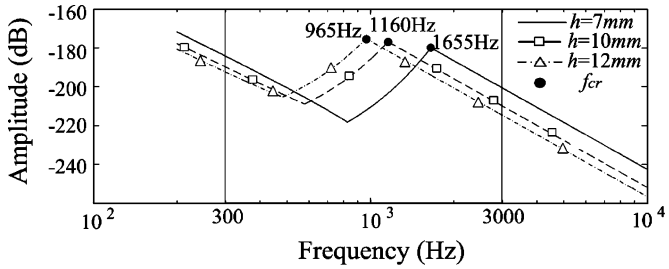


Fig. 5. Vibration amplitudes of glass panels in different thicknesses.

plate is smaller than the lower end of speech frequency (300 Hz); and the coincidence frequency of the PVC plate is larger than the upper end (3000 Hz). Hence, only the glass plate has an amplitude fluctuation in the speech frequency range. The amplitudes of both concrete and PVC plates just decrease monotonically with the increase of frequency in the speech range.

Second, in the frequency range of speech, the amplitudes of the concrete plate are the smallest. It means that the LDV sensor will have the lowest signal magnitudes of speech if a concrete wall is chosen to be the reflecting target. Furthermore, the slope of its amplitudes (the solid line) is the largest, indicating worse sensitivity at higher frequency.

Third, the PVC plate has larger vibration amplitudes than the glass plate below 1250 Hz in the speech frequency range. It means that the acquired signals by the LDV off a PVC surface has stronger signal magnitude levels in the lower frequency components. However, the acquired signals will have low resolution because of the attenuation in the higher frequency components. In the range from 1250 to 3000 Hz, the vibration amplitudes of the glass are larger than the PVCs because of the coincidence effect. This means that the acquired voice signals from the glass plate will have a better performance on the resolution and intelligibility.

D. Simulation Analysis of Different Thicknesses

For the structural property of the target, only the thickness is considered in this paper. Fig. 5 gives the vibration amplitudes on the top surface of the glass plate of different thickness. The coincidence frequencies of the glass plates with different thicknesses decrease with their thicknesses (and therefore mass) as predicted in (8). In the frequency ranges below 300 Hz and above 3000 Hz, the vibration amplitudes curves are parallel to each other. This is because their sound transmission losses decrease 6 dB for each reduction in mass by half [9]. Meanwhile, in the speech range from 300 to 3000 Hz, there is a fluctuation for each curve. And the parts of the fluctuation curves are also approximately parallel. Hence, we can conclude that the amplitude curves of different thicknesses have the same shape, which is determined only by the material of the panels, and the only difference is the overall levels of magnitudes. In this paper, we will use this material invariance as the key property to targets classification.

III. Experimental Results and Targets Classification

In the previous section, the theoretical analysis using both modeling and simulation shows that within the frequency range

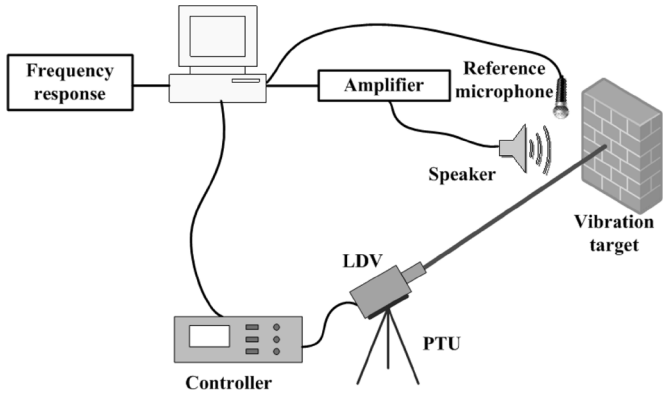


Fig. 6. Experimental setup.

of speech, there are three types of *characteristic curves of vibrations* caused by speech (Fig. 4). They correspond to materials when the coincidence frequencies are below, within and above the frequency range of speech. For the surfaces with the same material, the vibration characteristic curves have the same shape. Guided by these observations, some experiments will be carried out in this section. Based on these experimental results, the surfaces we test in our experiments will be classified. This classification is useful for target selection for speech acquisition.

A. Experimental Setup

The setup of our remote voice acquisition sensor system (together with the system test component) is shown in Fig. 6. The LDV voice acquisition system consists of an LDV sensor, a pan–tilt unit (PTU) and a personal computer (PC). The LDV from Polytec includes a controller OFV-5000 with a digital velocity decode card VD-6 and a sensor head OFV-505. The sensor head of the LDV uses a He–Ne red laser with a wavelength of 632.8 nm and is equipped with a super-long-range lens. It sends the interferometry signals to the controller. The controller box processes signals received from the sensor head of the LDV, and then output signals to computer using S/P-DIF output [3]. Meanwhile, as the test component for the LDV system, a speaker and a reference microphone are used. The speaker controlled by the PC is used to generate acoustic source signals. The driving signals produced by the PC are a series of sinusoidal harmonic signals whose frequencies range from 300 to 3000 Hz, with an interval step of 100 Hz. In order to correct the nonlinear effect of the speaker (whose frequency response is not uniform in the speech range), the reference microphone is used to calibrate the speaker.

In our experiments, the vibration targets selected include the following objects that could be found in our laboratory (and any typical indoor environments). They are concrete wall, door, glass plate, white board, metal plate, metal box, and paper box. Similar kinds of materials can also be easily found in an outdoor environment. Of course, the quality of the return signals from a surface of each object depends on both the reflectance and the vibration of the surface. Since this paper mainly study the vibration properties of the surfaces, we put a small piece of retro-reflective tape on the surface of each target, on which the laser

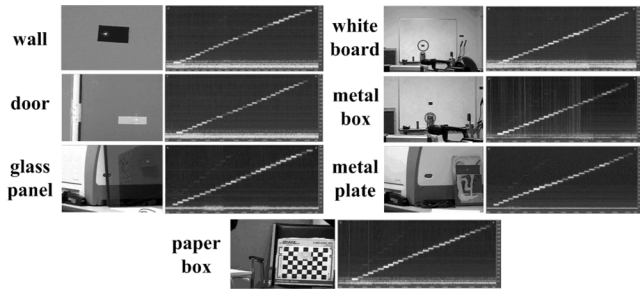


Fig. 7. Vibration targets and acquired signals.

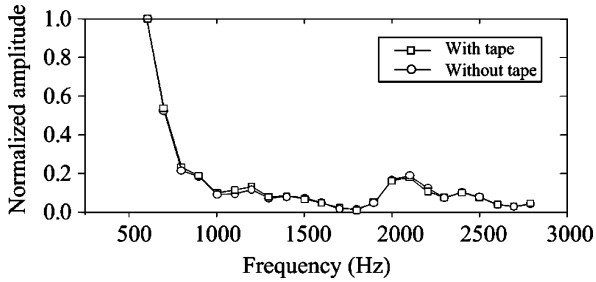


Fig. 8. Experimental results of white board with and without retro-reflective tape.

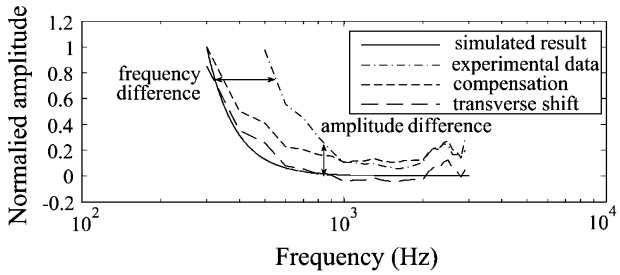


Fig. 9. Simulated and experimental results of the wall.

beam will be projected. (The affect of this retro-reflective tape on the vibration of the target will be discussed below.) Fig. 7 shows the images of these objects and the laser spots, the spectrograms of the received signals by the LDV. The ladder-like slashes indicates the responses of the LDV to voice signals at different frequency levels of 100-Hz intervals. The brightness values of the spectrograms represent signal levels, i.e., the vibration amplitudes of the target surfaces caused by the source signals.

B. Analysis of Experimental Results

In our experimental setup, a retro-reflective tape is attached on each target surface to increase the signal levels. In this part, the affect of the retro-reflective tape on the vibration of a target is estimated at first. Then, the vibration characteristics of different targets are analyzed. The experimental results are compared with the simulated results generated from the theoretical model to validate the characteristic curves of vibration amplitudes (captured by the LDV) with various targets materials.

1) The Affect of the Retro-Reflective Tape on the Target Vibration: For wall acoustic insulation measurement, the sensors are usually attached on a wall surface. As long as the weight of the sensor is much smaller than the weight of the wall, its affect on the acoustic insulation measurements can be neglected [6]. For our experiment, the self weight of the retro-reflective tape is light. Furthermore, the projected laser point on the target surface has been focused. Its diameter is so small that we can design the size of the retro-reflective tape to make its weight much smaller than the weight of the chosen target. Hence, its affect on the vibration characteristic of the target also can be neglected. In order to experimentally validate this analysis, we take the white board as the target. The laser directs to the same place on the surface, with and without retro-reflective tape, respectively. The vibrations are measured by the LDV and compensated by the reference microphone (the compensation method will be introduced below). The results are shown in Fig. 8. We have found that the two vibration curves almost identical, which indicates that the attached retro-reflective tape on the white board surface has little affect on its vibration characteristic.

In the above experiment, when the target surface was not treated with the retro-reflective tape, we observed that the acquired signals with a decreased reflected light intensity and an increased noise level compared to the results obtained with the retro-reflective tape. However, for our analysis, we are only concerned with the vibration responses of the target to different frequencies of our test audio source. Hence, even the noise level increases, as long as the amplitudes of the vibration responses are higher than the amplitudes of the noise, we will obtain the correct results. And almost the same vibration characteristic curves are obtained, as validated in Fig. 8.

As described and analyzed above, since the retro-reflective tape does not affect the vibration characteristics of the target surface and helps to improve the effective responses, which makes our analysis correct, in our following experiments, the retro-reflective tape is adopted to simplify the analysis.

2) Wall, Door and White Board: A part of the simulation result of the 10-cm wall (Fig. 3) in the speech range (300–3000 Hz) is redrawn with a thick line in Fig. 9 (indicated as "simulated result" in the figure). Meanwhile, the curve of experimental data is shown as a thin line. The dotted–dashed line is the compensated result by the reference microphone (labeled as "compensation"). It is obvious that both frequency and amplitude shifts exist between the compensated result and the simulated result; this fact was also reported in [9]. Based on the analysis in [9] and [10], the differences could be caused by the neglect of the plate boundary in the STL models discussed in Section II. As the compensated result parallel shifted to the dashed line (labeled as "parallel shift"), it can be clearly seen that the shifted compensated result is remarkably close to the simulated result.

With the same approach, the experimental data of the door and the white board are processed and shown in Fig. 10. It is concluded that their vibration characteristic curves have very similar shapes as the wall within the speech range, whose amplitudes decrease monotonically with the increase of frequency in the speech range. Hence, these three kinds of materials (targets) are classified into the same category, and their characteristic curves of frequency responses can be expressed by the model of

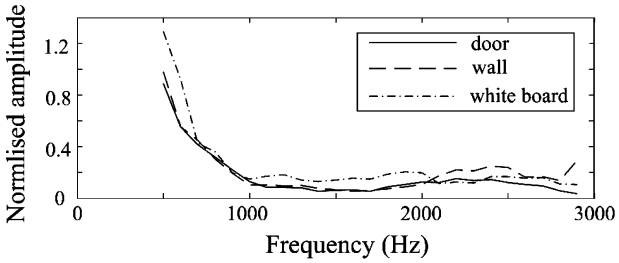


Fig. 10. Experimental results of wall, door and white board.

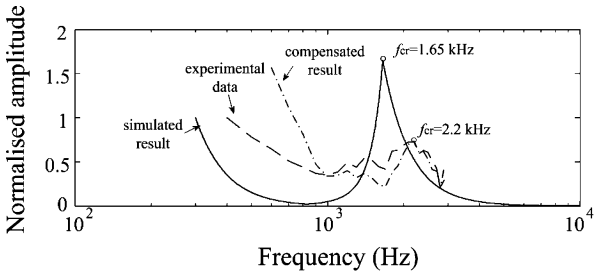


Fig. 11. Simulated, experimental and compensated results of glass panel.

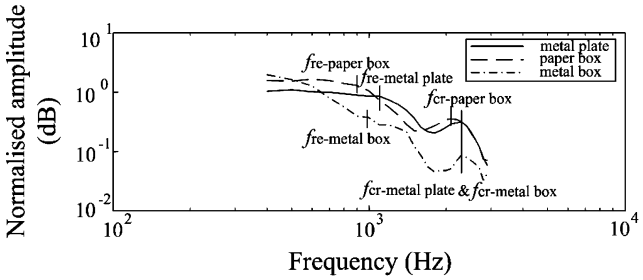


Fig. 12. Experimental results of metal plate, metal box, and paper box.

the wall described in Section II, whose coincidence frequency is smaller than 300 Hz.

3) Glass Plate: Fig. 11 shows the simulated, experimental, and compensated results of the glass plate, which has a thickness of 4 mm. Its theoretical coincidence frequency is 1.65 kHz, which is within the speech range. Hence, there is a fluctuation around coincidence frequency. Using the fluctuation as a signature, the coincidence frequency in the real data captured by the LDV can be identified to be 2.2 kHz. There are also coincidence frequency difference and amplitude difference, as also reported in [8].

4) Metal Box, Metal Plate, and Paper Box: Fig. 12 shows the compensated results of the metal box, the metal plate, and the paper box. In order to show the vibration characteristic curves more clearly, the unit of vertical axis is in decibels. Different from the above results, the three curves have two fluctuations, as marked in the figure (with short vertical lines). As described in [8], they are caused by the structure of these boxes, which is the plate of multi-laminated panels: mass–air–mass. For this kind of structure, the model reported in [8] describes that there are two fluctuations in its insulation curve. One is caused by the mass–air–mass resonance, and the other one is caused by the coincidence effect [8]. Correspondingly, there will be two fluctuations in the vibration characteristic curves of the top layer.

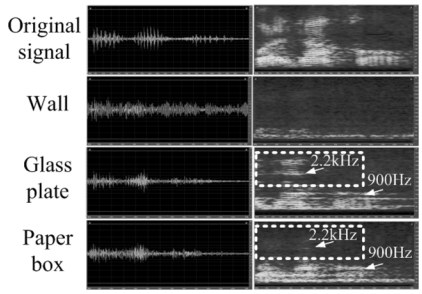


Fig. 13. Acquired speech signals and their spectrograms at a short distance.

This is validated by our experimental results. Since the metal box and the metal plate (the cover of the metal box) have the same material, they share the same coincidence frequency, as described in [8]. Although the structure of the metal plate is different from the metal box and the paper box, they still have very similar frequency responses. It may be caused by the chamber formed by the metal plate and its supporting backboard.

In conclusion, based on the experimental data and analysis, these targets are divided into three categories by the number of the fluctuations in their vibration characteristics curves to frequencies within the speech range. For the targets in the first category (including wall, door, and whiteboard), there is no fluctuation in their response curves. Hence, they have a very strong response to the low-frequency part. Then, the response strengths decrease very quickly with the frequency increases. In this case, if these targets are chosen for the speech acquisition by the LDV, the low-frequency components of the acquired signal will be so strong that the other frequency components (especially the high frequency components) will be submerged, which could make the signals having a strong background but less details for human hearing and understanding. For the targets in the second (glass plate) and the third (metal and paper) categories, there are one fluctuation and two fluctuations in their response curves, respectively. Because of the coincidence effect, their responses to the frequencies around the coincidence frequency are reinforced. Hence, the acquired signals from these targets not only have a strong background, but also have more details, which will improve the signal's intelligibility. They are recommended for LDV listening.

C. Experiments on Speech Acquisition

In this section, we use the characteristic curves and the conclusions described above as guidance to choose targets for LDV speech acquisition. Several experiments on speech acquisition by different targets are carried out, both at a short distance (about 5 m) and a long distance (about 140 m). The original test signal, "Good evening" (in Chinese), is captured from a sample of China Central Television (CCTV) News Report. The targets hit by the laser beam are the representations of above three categories, including the wall, the glass plate, and the paper box. Acquired signals at a short distance of about 5 m and their spectrograms are shown in Fig. 13. All the signals are highly intelligible, as shown in both time and frequency domains.

In the original signal, it is obvious that there are plenty of low- and high-frequency components by examining its spectrogram. Hence, the waveform in the time domain is quite sharp. There

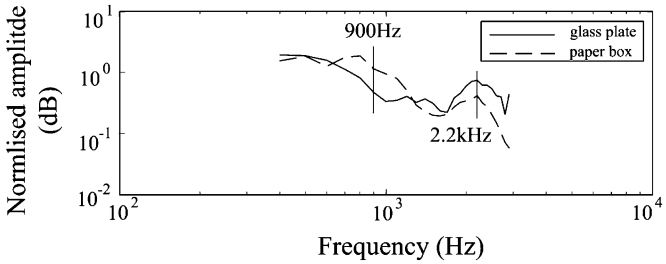


Fig. 14. Frequency responses of glass and paper box.

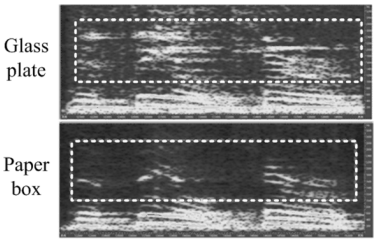


Fig. 15. Acquired speech signals and their spectrograms at a long distance.

are many peaks when the speech happens. In the LDV signal acquired from the concrete wall, the low-frequency components captured are much better than the high-frequency components. This verified that both our theoretical analysis and real-sensor experiments are correct. Correspondingly, the waveform in the time domain is much more flat and has fewer peaks. Note that these peaks in time domain correspond to the high-frequency components in the frequency domain.

In order to compare the signals acquired from the glass plate and the paper box more clearly, their frequency response signals are shown in Fig. 14. We can easily see that the paper box, compared against the glass plate, has larger responses at a frequency around 900 Hz and smaller responses at the frequency around 2.2 kHz. A similar conclusion can also be obtained from Fig. 13. As shown by the high-frequency components highlighted in the dashed rectangular box, the glass plate has much better responses in the high-frequency components. Correspondingly, the waveform of the glass plate has more peaks and is sharper than the waveform of the paper box in the time domain, which means that higher intelligibility and more detailed information can be obtained by the glass plate.

For the long-range voice acquisition, since signals from the wall is obviously worse than the ones from the glass plate and paper box, only the acquired signals from the glass plate and the paper box, which are intelligible, are focused on here. In fact, the vibration characteristic curves can guide us to select the right surfaces for long-range voice detection. (In this experiment, we selected the glass plate or even the paper box instead of the concrete wall.) Their spectrograms are shown in Fig. 15. As shown by the high-frequency components highlighted in the dashed rectangular box, the glass plate also has a much better response in the high-frequency components, which is as same as the results at a short distance. Meanwhile, the noise levels (the bright spots on the background of the spectrogram) in Fig. 15 are higher than the noise levels in the Fig. 13, because of the decreases of the reflected light intensity in the long-range experiment. However, we still obtain the same observations as the one from the short-range experiment. It verifies again that, as long as the intensities of the captured reflected signals by the LDV are strong enough to maintain the noise level lower than the signal level, our conclusions on the vibration characteristics of the targets can still be used as guidance for the targets selection at a large distance.

The analysis of experimental speech acquisition signals at both the short and long distances further verifies the observations we obtained in Sections II and III-B, showing the model of the vibration characteristics of various targets can be applied to real sensor data. Furthermore, for speech acquisition by the LDV, we found that targets like wall are not good choices because of their poor response to the high-frequency components. However, our analysis showed that we can hear voices from a concrete wall if no other choices are available. Comparing the other two kinds of targets, like the paper box and the glass plate, a better signal can be obtained by a glass plate because of its higher responses to high-frequency components. These observations could be good guidance for actual target selection for LDV listening.

IV. Conclusions and Discussions

For LDV voice acquisition, particularly the acoustic events occur from a large distance to the sensor, finding the right vibration surfaces close to the acoustic sources (humans, vehicles, etc.) is very important. In this work, the vibrations of target surfaces caused by speech signal are studied through both theoretical modeling/simulation and real-sensor experimental analysis. Based on the vibration responses of various targets of different materials and structures to different frequencies in the speech range, the targets in our experiments are classified into three categories. The potentials of these three kinds of targets selected for LDV listening are discussed by comparing their characteristic curves. Both short- and long-range LDV voice detection experiments with these three kinds of targets are also carried out. Experimental results verified our conclusion that the acquired signals from the targets in the second and the third categories, like glass plate and paper box, give better performances and are recommended for LDV listening. Especially for glass plate, it has a better response to high-frequency components, which means a higher intelligibility in speech signals.

In addition, the characteristic curves of frequency responses of these targets, can not only be used to make a better selection of appropriate targets for LDV voice detection, but also have the potential to be utilized for both signal enhancement and signal interpretation for the signals captured by the LDV off these targets. The algorithm of signal interpretation and speech enhancement adopting the characteristic curves of targets will be developed in the future. We will also look into the vibration characteristics of the bottom surfaces of plates of various materials for applications like through-wall hearing.

References

[1] D. Zotkin, R. Duraiswami, H. Nanda, and L. Davis, "Multimodal tracking for smart videoconferencing," in *Proc. Int. Conf. Multimedia Expo*, Tokyo, Japan, 2001, pp. 36–39.

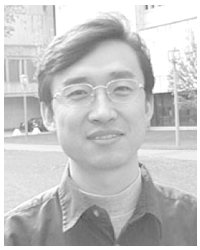
[2] X. Zou and B. Bhanu, "Tracking humans using multimodal fusion," in *Proc. 2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, 2005, pp. 4–4.
[3] Z. Zhu and W. Li, "Integration of laser vibrometer and infrared video for multimedia surveillance display," CS Dept., CUNY Graduate Center, New York, TR-2005006, Jan. 13, 2010 [Online]. Available: http://tr.cs.gc.cuny.edu/tr/files/TR-2005006.pdf
[4] W. Li, M. Liu, and Z. Zhu, "LDV remote voice acquisition and enhancement," in *Proc. 2006 IEEE Pattern Recognition Conf. (ICPR)*, Hong Kong, China, 2006, pp. 262–265.
[5] Y. Qu, T. Wang, and Z. Zhu, "An active multimodal sensing platform for remote voice detection," presented at the IEEE/ASME Int. Conf. Advanced Intelligent Mechatronics (AIM 2010), Montreal, QC, Canada, Jul. 6–9, 2010.
[6] L. M. Brechovskich and O. A. Godin, *Acoustics of Layered Media*. New York: Springer, 1990.
[7] A. Tadeu and J. M. P. Antodnio, "Acoustic insulation of single panel walls provided by analytical expressions versus the mass law," *J. Sound Vibrat.*, vol. 257, no. 1, pp. 457–475, Oct. 2002.
[8] A. Tadeu, J. Antonio, and D. Mateus, "Sound insulation provided by single and double panel walls—A comparison of analytical solutions versus experimental results," *Appl. Acoust.*, vol. 65, no. 1, pp. 15–29, Sep. 2004.
[9] J. R. Callister, A. R. George, and G. E. Freeman, "An empirical scheme to predict the sound transmission loss of single-thickness panels," *J. Sound Vibrat.*, vol. 222, no. 1, pp. 145–151, Apr. 1999.
[10] A. Osipov, P. Meesb, and G. Vermeif, "Low-frequency airborne sound transmission through single partitions in buildings," *Appl. Acoust.*, vol. 52, no. 3, pp. 213–288, Dec. 1997.

Rui Li received the B.S. degree in opt-electronic engineering from Beihang University, Beijing, China, in 2005. He is working towards the Ph.D. degree at Beihang University.

He is currently a visiting research student in the Visual Computing Laboratory, the City College of the City University of New York, supported by State Scholarship Fund from the China Scholarship Council. His research interests include fiber acoustic sensors and laser vibration sensors.

Tao Wang (S'09) received the B.S. degree in computer science from Stony Brook University, New York, in 2002 and the M.Eng. degree in civil engineering from Cornell University, New York, in 2004. He is currently working towards the Ph.D. degree at the Graduate Center of City University of New York.

Since 2006, he has been a Research Assistant in the City College Visual Computing Laboratory, working on multimodal sensor design and integration, and video surveillance.



Zhigang Zhu (S'94–AM'97–M'99–SM'05) received the B.E., M.E., and Ph.D. degrees, all in computer science, from Tsinghua University, Beijing, China, in 1988, 1991, and 1997, respectively.

Previously, he was Associate Professor at Tsinghua University and Senior Research Fellow at the University of Massachusetts, Amherst. He is currently a Full Professor in the Department of Computer Science, the City College and the Graduate Center, the City University of New York. He is Director of the City College Visual Computing Laboratory (CcvcL), and Co-Director of the Center for Perceptual Robotics, Intelligent Sensors and Machines (PRISM) at City College of New York. His research interests include 3-D computer vision, multimodal sensing, virtual/augmented reality, video representation, and various applications in education, environment, robotics, surveillance, and transportation. He has published over 100 technical papers in the related fields.

Dr. Zhu is a Senior Member of the ACM, an Associate Editor of the *Machine Vision Applications Journal*, and a Technical Editor of the IEEE/ASME TRANSACTIONS ON MECHATRONICS.

Wen Xiao received the B.S. degree in laser physics from Northwest University China in 1984 and the Ph.D. degree in optics from the Chinese Academy of Sciences in 1995.

He conducted research on fiber optic sensors as a Postdoctoral Fellow at Zhejiang University and worked in the Chinese Academy of Launch Vehicle (CALV) as a Senior Engineer. He is currently a Professor at Beihang University, Beijing, China, and Vice Dean of the School of Instrument Science and Opto-electronics Engineering at the same university.