

文章编号 1004-924X(2017)03-0569-07

## 远距离语音的激光相干获取及增强

吕 韬<sup>1,2\*</sup>, 张合勇<sup>1</sup>, 郭 劲<sup>1</sup>, 晏春回<sup>1,2</sup>

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;  
2. 中国科学院大学, 北京 100049)

**摘要:**基于激光相干探测理论建立了全光纤激光相干测振系统,并结合维纳语音增强技术实现了远距离、高质量语音信号的获取。首先,介绍了该系统的检测原理及光纤器件特性,搭建了全光纤激光相干测振实验系统检测发音者咽喉振动,从而实现远距离语音信号的获取。由于直接获取的语音信号受到严重的噪声干扰,采用基于最小控制递归平均算法估计噪声的维纳滤波来抑制噪声,提高语音质量。实验结果显示:激光相干测振系统可有效获取 70 m 内的语音信号(检测喉咙振动还原语音);语音增强技术可有效抑制噪声,将信噪比提高 5 dB,并增强语音的清晰度。由此表明,该系统具备获取远距离、高质量语音信号的能力。

**关键词:**激光相干测振;语音获取;语音增强;维纳滤波

中图分类号:TN249;TN912.35 文献标识码:A doi:10.3788/OPE.20172503.0569

## Acquisition and enhancement of remote voice based on laser coherent method

LÜ Tao<sup>1,2\*</sup>, ZHANG He-yong<sup>1</sup>, GUO Jin<sup>1</sup>, YAN Chun-hui<sup>1,2</sup>

(1. *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;*  
2. *University of Chinese Academy of Sciences, Beijing 100049, China)*  
\* *Corresponding author, E-mail:18767120269@163.com*

**Abstract:** An all-fiber Laser Doppler Vibrometer (LDV) based on laser coherent method was investigated to acquire remote and high quality speech signals, combining with the Wiener speech enhancement technology. First, the heterodyne detection theory and the characteristic of fiber devices were introduced and the all-fiber LDV was established experimentally to detect the throat vibration, thus the remote voice was acquired. Whereas the direct acquisition of speech signals was prone to be disturbed by serious noise, the Wiener Filtering based on Minima Controlled Recursive Averaging (MCRA) was used to suppress the noise and improve the voice quality. The results indicate that the comprehensible speech signals within the range of 70 m can be obtained by the LDV. The speech enhancement technology can suppress the noise effectively and the Signal to Noise Ratio (SNR) was increased by 5 dB, thus the speech intelligibility was improved. This system has ability to acquire remote voice with high quality.

**Key words:** laser Doppler vibrometer; voice acquisition; speech enhancement; Wiener filtering

收稿日期:2016-08-25;修订日期:2016-11-10.

基金项目:国家自然科学基金资助项目(No. 61205143)

## 1 引言

激光相干检测技术源于 20 世纪 80 年代, 具有高灵敏性、抗干扰性强、非接触式及远距离等优点而在精密测量与工业生产中获得了广泛的应用<sup>[1-2]</sup>。激光相干检测系统能在非接触情况下检测较远距离目标的微小振动, 因此可通过检测由声压引起的物体表面振动或声带振动来获取语音信号。Weihong Li 等人采用德国 Polytec 公司生产的激光多普勒测速仪 (Laser Doppler Vibrometer, LDV) 检测声源附近易振物体表面的振动以获取语音信号, 并结合监视相机搭建了多模监视系统<sup>[3-6]</sup>, 但该系统所用的激光波段为 633 nm, 属于可见光波段且体积重量较大 (体积为 120 mm×80 mm×345 mm, 质量为 3.4 kg)。因此, 该系统获取语音信号的隐蔽性和便携性较差。Yekutieli Avargel 等人通过商业 LDV 检测发言人咽喉振动以获取语音信号<sup>[7]</sup>, 但所用 LDV 的探测距离较近且光源处于可见光波段, 无法满足实际应用需求。尚建华<sup>[8-9]</sup>等人通过自研 LDV 系统获取语音信号, 在 20 m 的距离范围内开展了单音获取实验, 但尚未开展较远作用距离和复杂语音信号获取方面的实验研究。

本文以输出波长为 1 550 nm、谱线宽度小于 10 kHz、输出功率为 20 mW 的单纵模连续激光为基础, 搭建了用于获取远距离语音信号的全光纤激光相干检测系统, 并结合基于最小控制递归平均算法 (Minima Controlled Recursive Averaging, MCRA) 估计噪声的维纳滤波语音增强技术来提高获取语音信号的质量。实验结果表明: 该激光相干测振系统可有效获取 70 m 内的语音信号 (喉咙振动还原语音); 并且运用语音增强技术可有效抑制噪声, 明显增强语音的清晰度。

## 2 语音获取原理

声带是产生语音的激励源, 因此通过自研系统探测声带振动的位移便可探测语音信号。激光相干测振系统结构如图 1 所示, 激光器输出的激光光束经光纤耦合器分为两束: 一束光经声光调制器 (Acousto-Optic Modulator, AOM) 调制发生频率  $\omega_{AO}$ , 作为本振光  $u_o$ ; 另一路光经待测振动目标 (咽喉) 多普勒调制, 产生多普勒频移  $\varphi(t) = 4\pi S(t)/\lambda$  (其中  $\lambda$  为激光波长,  $S(t)$  为振动位移), 作为信号光  $u_s$ 。将这两路光通过光纤耦合器进行混频, 产生一个频率为  $\omega_{AO} + \varphi(t)$  的拍频信号。该拍频信号通过光电平衡探测器后转换为电压信号, 并通过带通滤波器输出中频信号  $u_{IF}$ <sup>[9]</sup>:

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

其中:  $A_o, A_s$  分别为本振光与信号光的幅度,  $\alpha$  为光电转换效率。

由于光电探测器只输出拍频信号, 因此具体的振动信息需要通过正交解调得到正交基带信号  $u_I$  与  $u_Q$ , 如图 2 所示。其中:

$$\begin{cases} u_I = \xi \cos[\varphi(t) + \varphi_1 - \varphi_2] \\ u_Q = \xi \sin[\varphi(t) + \varphi_1 - \varphi_2] \end{cases} \quad (2)$$

通过反正切鉴相算法<sup>[10]</sup>, 如图 3 所示, 可以获得振动目标的多普勒频移  $\varphi(t)$ :

$$\varphi(t) = \arctan\left(\frac{u_Q}{u_I}\right) + m\pi + \Delta\varphi. \quad (3)$$

由于直接使用反正切鉴相算法进行解调会引起相位跃变, 所以这里使用相位解缠算法来修正相位跃变<sup>[11]</sup>。在获得振动目标引起的相位变化  $\varphi(t)$  后即可获得咽喉振动位移  $S(t)$ , 从而获取语音信号。

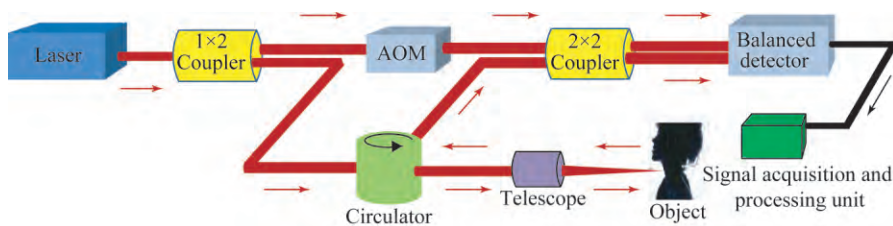


图 1 获取语音的激光相干探测系统原理图

Fig. 1 Schematic diagram of laser coherent detection system for human speech acquisition

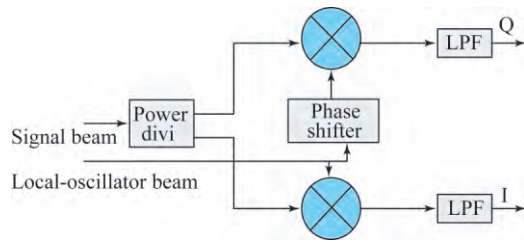


图 2 正交解调框图

Fig. 2 Block diagram of demodulation

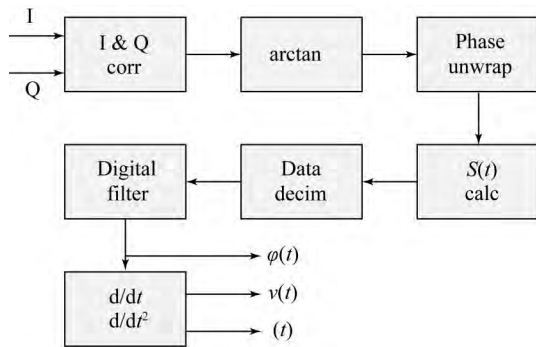


图 3 反正切鉴相框图

Fig. 3 Block diagram of arctangent phase algorithm

### 3 语音增强技术

由于激光束照射在粗糙物体表面会不可避免地存在光束散斑效应,背景噪声以及后续信号处理时引入的暗电流噪声等会严重干扰获取的语音信号。语音增强技术是减小噪声干扰的重要手段,该技术可提高获取语音信号的质量。短时谱估计语音增强算法又称为短时傅里叶分析-修正合成(Analysis-Modification-Synthesis, AMS)增强算法,这类算法不需要从带噪语音信号中生成模型参数,限制较少、适应性较好,所以得到了广泛应用。其代表性算法有谱减法、维纳滤波、谱幅度最小均方误差估计法等。本文选用维纳滤波来进行信号增强。

#### 3.1 维纳滤波

维纳滤波器是统计意义上的最优滤波器<sup>[12,15]</sup>。假设带噪语音信号为  $y(t) = s(t) + n(t)$ , 其中  $s(t)$  为纯净语音信号,  $n(t)$  为噪声, 两者相互独立。由于语音信号是一个非稳态的时变信号, 因此需对带噪语音进行分帧处理, 然后再进行短时傅里叶变换(Short-Time Fourier Transform, STFT)。其输出为:

$$Y_i(k) = S_i(k) + N_i(k), \quad (4)$$

其中:  $Y_i(k)$ ,  $S_i(k)$ ,  $N_i(k)$  分别表示带噪语音、纯净语音以及噪声第  $i$  帧第  $k$  个频谱分量。

维纳滤波方法就是设计一个数字滤波器  $H_i(k)$ , 当输入为  $Y_i(k)$  时, 滤波器的输出为  $S_i(k)$  的估算值, 即有:

$$\hat{S}_i(k) = H_i(k)Y_i(k), \quad (5)$$

根据正交性原理可知维纳滤波  $H_i(k)$  为:

$$H_i(k) = 1 - \frac{\lambda_n(i, k)}{E[|Y_i(k)|^2]}, \quad (6)$$

其中  $\lambda_n(i, k)$  表示第  $i$  帧的噪声方差。

#### 3.2 噪声估计算法

由于用于语音增强的维纳滤波器是建立在噪声估计基础上, 因此, 噪声估计是语音增强技术的关键环节。简单的噪声估计可通过带噪语音的前几帧纯噪声段估计得到, 但此类方法只适用于平稳噪声。基于语音激活检测(Voice Activity Detection, VAD)的噪声估计算法可在无语音帧的条件下更新噪声, 但是在低信噪比或者非平稳噪声的环境下, 这种噪声估计的效果不理想<sup>[13]</sup>。基于最小统计的最小控制递归平均算法(MCRA)<sup>[13]</sup>则可以动态更新噪声功率谱, 适合非平稳噪声环境。实际环境中系统多受非平稳噪声的干扰, 因此, 本文采用 MCRA 算法估计噪声。

##### 3.2.1 最小控制递归平均算法

在 MCRA 算法中, 噪声方差  $\lambda_n(i, k)$  的估计值可表示为:

$$\hat{\lambda}_n(i+1, k) = \hat{\lambda}_n(i, k)P'(i, k) + [\partial_n \hat{\lambda}_n(i, k) + (1 - \partial_n)|Y_i(k)|^2](1 - P'(i, k)), \quad (7)$$

其中:  $\hat{\lambda}_n(i, k)$  是第  $i$  帧噪声方差的估计值,  $\partial_n$  ( $0 < \partial_n < 1$ ) 是平滑参数,  $P'(i, k)$  为条件语音存在概率。

由式(6)可知, 条件语音存在概率的估计是估计噪声的重要一环, 它是每个频带中含噪语音功率谱和局部最小值的比值。为了得到局部最小值, 需要对含噪语音的功率谱进行频域平滑。其中, 频域平滑使用一阶递归平均, 即:

$$S(i, k) = \partial_s S(i-1, k) + (1 - \partial_s)S_i(i, k). \quad (8)$$

式中:  $S_i(i, k) = \sum_{l=-\omega}^{\omega} b(l)|Y(i, k-l)|^2$ , 是功率谱的频域平滑;  $b(l)$  是一个归一化的汉明窗,  $\partial_s$  是平滑参数。在一个长为  $L$  帧的窗内比较含噪语

音平滑谱和局部最小值,以获取局部最小值  $S_{\min}(i,k)$ 。

含噪语音平滑谱与局部最小值的比率定义为  $S_r(i,k) = S(i,k)/S_{\min}(i,k)$ 。

通过求解一个贝叶斯最小成本函数可以得到语音的判决式:

$$I(i,k) = \begin{cases} 1, S_r(i,k) \geq \delta \\ 0, \text{else} \end{cases}, \quad (9)$$

其中:  $I(i,k) = 1$ , 代表语音存在;  $I(i,k) = 0$ , 代表语音不存在。使用该判决和一个平滑参数  $\partial_p$  就可以得到对语音存在概率的估计:

$$\hat{P}'(i,k) = \partial_p \hat{P}'(i-1,k) + (1-\partial_p) I(i,k). \quad (10)$$

在得到语音存在概率后可进行噪声方差更新。

### 3.3 算法仿真

用 MATLAB 对基于 MCRA 算法估计噪声

的维纳滤波进行仿真。仿真中纯净语音选自实验录音“你好”。由于在实际系统中,散粒噪声是噪声干扰的主要成分,因此在纯净语音中加入散粒噪声生成带噪语音。这里使用波形图、语谱图以及信噪比(Signal to Noise Ratio, SNR)作为评判语音增强效果的标准,即:

$$SNR = 10 \log_{10} \frac{\sum_{t=0}^{T-1} s^2(t)}{\sum_{t=0}^{T-1} n^2(t)}. \quad (11)$$

图 4(a)为输入信噪比为 0 dB 的带噪语音的语谱图与波形图,图 4(b)为经维纳滤波增强后语音的语谱图与波形图,图 4(c)为纯净语音的频谱图与波形图。表 1 为加入不同 dB 噪声的维纳滤波输出语音的信噪比。

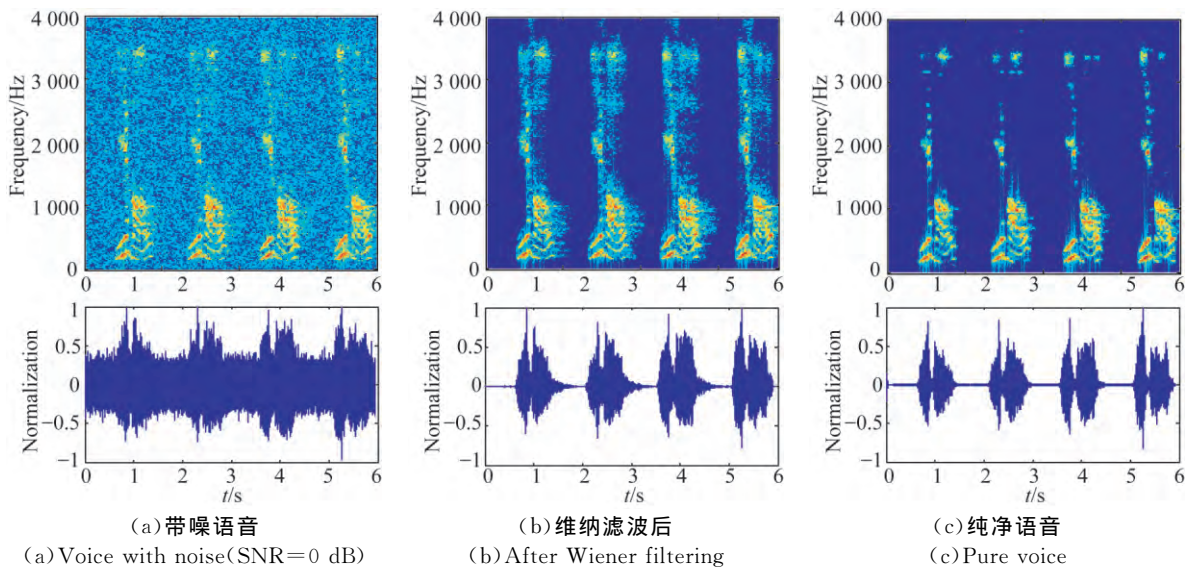


图 4 仿真得到的语音信号的波形图与语谱图

Fig. 4 Spectra and sequence diagrams of speech signal in simulation

表 1 原始语音和维纳滤波后的信噪比

Tab.1 SNRs of original and filtered voices (dB)

输入信噪比	0	3.000	6.000	9.000	12.000
维纳滤波	8.649	10.257	12.212	13.577	14.316
信噪比增加量	8.649	7.257	6.212	4.577	2.316

比较维纳滤波增强后(如图 4(b))与纯净语音(如图 4(c))的时域波形图与谱域语谱图可发现,两者十分接近,虽然由于噪声估计不足使得维纳滤波时谱减过程中产生了音乐噪声,从而导致

滤波后语谱图中非语音部分的颜色比纯净语音深,但是维纳滤波可有效抑制背景噪声,明显提高语音质量。由表 1 可知,维纳滤波可有效提高语音信号的信噪比,并且在信噪比相对较低的情况下对语音信号信噪比的提升效果更显著。

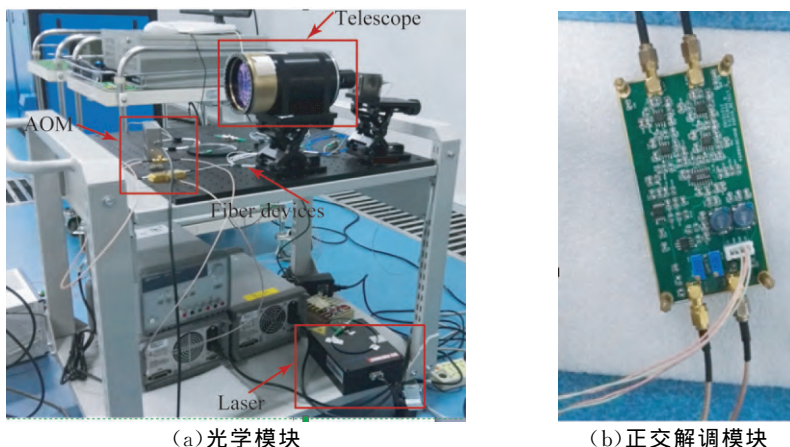
## 4 实验及结果分析

基于上述理论,以输出波长为 1 550 nm 的纵模连续光纤激光器,80 MHz 的声光调制器为



基础,搭建了激光相干测振系统,如图 5 所示,用于获取语音信号;并通过基于 MCRA 估计噪声的

维纳滤波对自研系统直接获取的语音信号进行语音增强。



(a)光学模块 (a)Optical module (b)正交解调模块 (b)Quadrature demodulation module

图 5 激光相干测振系统实物图

Fig. 5 Photo of proposed Laser Doppler Vibrometer(LDV)

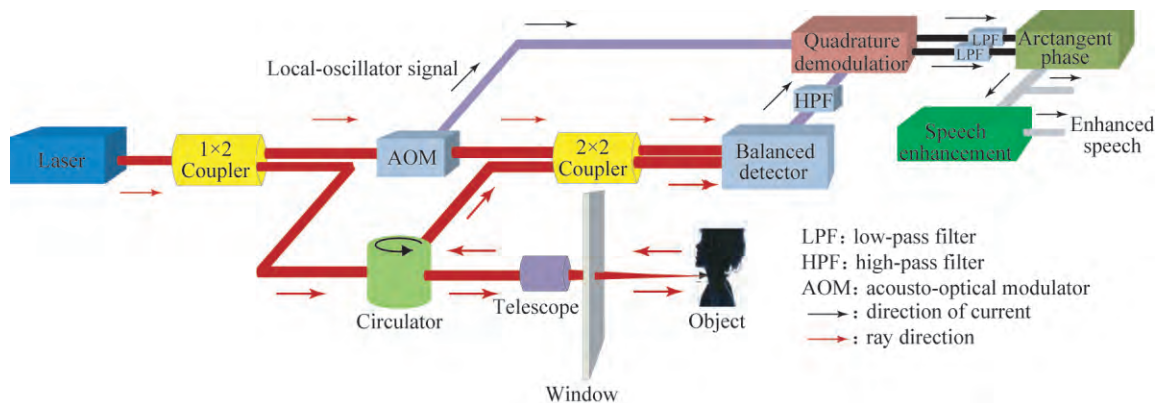


图 6 激光相干测振实验系统示意图

Fig. 6 Schematic diagram of Laser Doppler Vibrometer(LDV) experiment setup

如图 6 所示,在室内环境下,发射激光光束通过保偏光纤耦合器分为两束:一束经声光调制器(AOM)调制发生 80 MHz 频移,作为本振光;另一束穿透玻璃聚焦在 70 m 处的振动目标(咽喉)上,其表面漫反射的激光回波受到振动目标(咽喉)的多普勒调制,产生多普勒频移,作为信号光。信号光被望远系统接收后与本振光通过光纤耦合器进行相干叠加,产生拍频信号。该拍频信号通过平衡光电探测器转换为电压信号。平衡探测器输出的拍频信号经自研的正交解调电路处理,输出两路基带信号,利用采集卡(采样率为 0.05 MHz)同时采集两路基带信号,使用反正切鉴相解调算法将这两路基带信号还原为目标振动

信号,从而获取语音信号。随后,再通过基于 MCRA 估计噪声的维纳滤波对自研系统直接获取的语音信号进行增强。本实验所使用的激光功率为 20 mW,线宽小于 10 kHz。望远系统用来聚焦出射光束与接收反射光,保偏光纤环行器用来实现激光发射与接收信号的高效隔离。反正切鉴相解调算法与基于 MCRA 估计噪声的维纳滤波均在 MATLAB 环境下运行,运行参数为  $\delta_n = 0.95, \delta_s = 0.8, \delta_p = 0.2, \delta = 5$ 。

这里使用语谱图、波形图以及信噪比作为自研系统获取语音信号质量的判据,同时还作为基于 MCRA 估计噪声的维纳滤波的语音信号增强效果的评判依据。

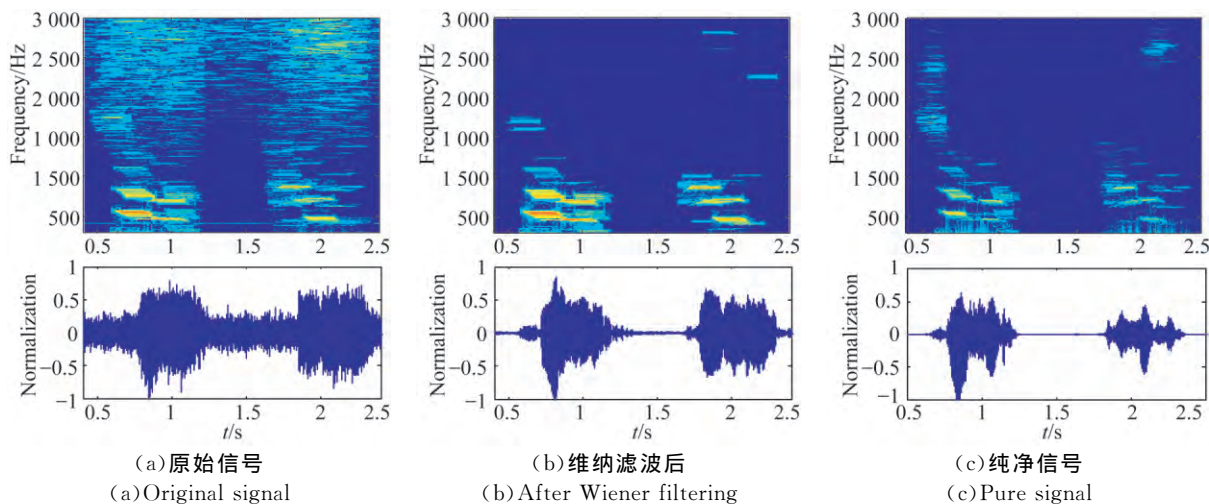


图 7 实验得到的语音信号语谱及波形图

Fig. 7 Spectra and sequence diagrams of speech signal in experiment

如图 7(a) 为自研系统获取语音信号“树木, 丛林”的语谱图及波形图, 图 7(b) 为经维纳滤波增强后语音信号“树木, 丛林”的语谱图与波形图, 图 7(c) 为纯净语音“树木, 丛林”(通过录音设备所得)的语谱图与波形图。比较图 7(a) 与图 7(c) 的语谱图可以发现, 自研系统获取的语音信号在整个频域范围内存在较大的背景噪声, 这是因为激光束照射在粗糙物体表面时存在散斑效应, 并且后续信号处理时还会引入暗电流噪声。但是图 7(a) 包含的语音频率成分与纯净信号(图 7(c)) 相同, 并且从波形图也可以发现, 波形轮廓大致相同, 说明自研激光相干测振系统具备获取 70 m 处语音信号的能力。比较图 7(b) 与图 7(c) 可以发现, 虽然经过维纳滤波增强后的语谱图的非语音部分颜色比纯净语音的语谱图深一些(这是由于噪声估计不足, 导致在谱减过程中产生了音乐噪声), 但是仍跟纯净信号的语谱图十分接近, 并且波形也与纯净信号十分相似。与直接获取语音信号的语谱图(图 7(a)) 相比, 经过维纳滤波后的整个频域干净了许多, 这说明维纳滤波可以有效滤除噪声, 提高语音质量。通过信噪比来评估自研系统获取语音信号的能力以及维纳滤波的表现, 结果如表 2 所示。系统直接获取的语音信号的信噪比为 6.079 dB, 经维纳滤波增强后语音信

号的信噪比为 11.140 dB。实验结果表明, 自研系统可有效获取 70 m 内的语音信号, 基于 MCRA 估计噪声的维纳滤波可有效抑制噪声, 将信噪比提高 5 dB, 语音清晰度明显增强。

表 2 距离 70 m 的语音信号信噪比

Tab. 2 SNR of speech signal at distance of 70 m (dB)

参数名称	原始信号分段信噪比	维纳滤波后分段信噪比
取值	6.079	11.140

## 5 结 论

本文基于激光相干探测理论搭建了激光相干测振系统, 用于获取远距离语音信号, 并通过语音增强技术来提高语音信号质量。实验结果表明, 该系统可有效获取 70 m 内的语音信号, 语音增强技术可有效抑制噪声, 并将信噪比提高 5 dB, 语音清晰度得到明显增强。不过, 通过探测声带振动来获取语音信号的方式存在一定局限性, 在实际应用中会受到诸如发言人身体移动等因素的限制, 因此下一步工作将通过比较分析各种物体在声场环境下的振动特性, 选择合适的易振物体作为探测目标以获取更高质量的语音信号。

## 参考文献:

- [1] 刘立生,张合勇,王挺峰,等. 激光外差探测对振动目标多普勒频谱成像[J]. 光学精密工程, 2015, 23(6): 1508-1515.  
LIU L SH, ZHANG H Y, WANG T F, *et al.*. Doppler spectrum imaging of vibrating target using laser heterodyne detection [J]. *Opt. Precision Eng.*, 2015, 23(6): 1508-1515. (in Chinese)
- [2] 刘立生,张合勇,郭劲,等. 用激光外差技术高精度测量目标速度[J]. 光学精密工程, 2011, 19(10): 2366-2372.  
LIU L SH, ZHANG H Y, GUO J, *et al.*. High precise measurement of target velocity using laser heterodyne technology [J]. *Opt. Precision Eng.*, 2011, 19(10): 2366-2372.
- [3] QU Y F, WANG T, ZHU Z G. Remote audio/video acquisition for human signature detection [C]. 2012 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2009: 66-71.
- [4] LI W, LIU M, ZHU Z, *et al.*. LDV remote voice acquisition and enhancement [C]. *International Conference on Pattern Recognition*, 2006: 262-265.
- [5] WANG T, ZHU Z G, DIVAKARAN A. Long range audio and audio-visual event detection using a laser doppler vibrometer [J]. *SPIE*, 2010, 7704: 7704J-7704J-6.
- [6] QU Y F, WANG T, ZHU Z G. An active multi-modal sensing platform for remote voice detection [C]. *Proceedings of 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, *IEEE*, 2010: 627-632.
- [7] AVARGEL Y, COHEN I. Speech measurements using a laser Doppler vibrometer sensor: Application to speech enhancement [C]. *Proceedings of 2011 Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, *IEEE*, 2011: 109-114.
- [8] SHANG J H, HE Y, LIU D, *et al.*. Laser Doppler vibrometer for real-time speech-signal acquirement [J]. *Chinese Optics Letters*, 2009, 7(8): 732-733.
- [9] JIAN J H, ZHAO SH G, HE Y, *et al.*. Experimental study on minimum resolvable velocity for heterodyne laser Doppler vibrometry [J]. *Chinese Optics Letters*, 2011, 9(8): 081201.
- [10] BAUER M, RITTER F, SIEGMUND G. High-precision laser vibrometers based on digital Doppler signal processing [J]. *SPIE*, 2002, 4827: 50-61.
- [11] 张东方. 基于短时谱估计的低信噪比语音增强技术研究[D]. 郑州:解放军信息工程大学, 2012.  
ZHANG D F. *Speech Enhancement Technique Research in Low SNR Conditions Based on Short-Time Spectrum Estimation* [D]. Zhengzhou: PLA Information Engineering University, 2012. (in Chinese)
- [12] 袁文浩. 基于噪声估计的语音增强方法研究[D]. 上海:华东理工大学, 2013.  
YUANG W H. *Study on Noise Estimation-based Speech Enhancement Approaches* [D]. Shanghai: East China University of Science and Technology, 2013. (in Chinese)
- [13] 靳晨升. 语音增强算法的研究[D]. 太原:太原理工大学, 2010.  
JIN CH SH. *The Research of Speech Enhancement Algorithm* [D]. Taiyuan: Taiyuan University of Technology, 2010. (in Chinese)
- [14] 宋知用. MATLAB在语音信号分析与合成中的应用[M]. 北京:北京航空航天大学出版社, 2013.  
SONG ZH Y. *Application of MATLAB in Speech Analysis and Synthesis* [M]. Beijing: Beihang University Press, 2013. (in Chinese)

## 作者简介:



吕 韬(1991—),男,浙江绍兴人,博士研究生,2014年于浙江工业大学获得学士学位,主要从事相干微振动测量信号处理技术的研究。E-mail: 18767120269@163.com

## 导师简介:



郭 劲(1964—),男,吉林长春人,研究员,博士生导师,2007年于中科院长春光机所获得博士学位,主要从事激光与物质相互作用技术的研究。E-mail: guojin1964@126.com