

改进谱减法结合神经网络的语音增强研究*

姚远 王秋菊 周伟 鲍程毅 彭磊

(华中师范大学物理科学与技术学院 武汉 430079)

摘要: 背景噪声是通信系统噪声干扰的来源之一,语音增强可以降低乃至消除噪声干扰,进而提高语音的可懂度。为了减小复杂噪声环境下谱减法引发的音乐噪声,采取正交的多窗谱估计对语音功率谱平滑处理,有效的减小了信息丢失和估计波动。利用自适应谱减系数调整谱增益和谱下限来控制残留噪声,利用优化的 IMCRA 算法对噪声及时更新来判决语音段和静音段,同时借助特性良好的 BP 神经网络方法进行训练,语音和噪声谱通过谱减后,波形重构获取增强的语音信号。仿真结果表明语音降噪效果好、可懂度高。

关键词: 语音增强;谱减法;噪声谱估计;多窗平滑;深度神经网络

中图分类号: TN911.23 **文献标识码:** A **国家标准学科分类代码:** 510.40

Research on speech enhancement
based on spectral subtraction and neural network

Yao Yuan Wang Qiuju Zhou Wei Bao Chengyi Peng Lei

(College of Physical Science and Technology, Central China Normal University, Wuhan 430079, China)

Abstract: Background noise is one of the noise sources of communication system, speech enhancement can reduce or eliminate noise interference, and improve speech intelligibility. Speech power spectrum is adopted orthogonal multi window spectrum estimation for smoothing processing, to reduce music noise from spectrum subtraction under the complicated noise environment. The loss of information and estimation have effectively reduced. Using adaptive spectral subtraction coefficient to adjust spectrum gain and floor to control the residual noise, using the optimized IMCRA algorithm for noise update to speech and silence, the enhanced speech signal has acquired by spectral subtraction and waveform reconstruction after the deep neural network trained data. The simulation results show that the noise reduction effect of speech is good, speech intelligibility is well.

Keywords: speech enhancement; spectral subtraction; noise spectrum estimation; multi window smooth; deep neural network

1 引言

车载设备、工程设备运行时会受到来自周围环境、电磁干扰、设备内部噪声乃至其他讲话者的干扰,例如安装在汽车、飞机或舰船上的通话系统受到的强背景噪声的干扰,降低了语音的感知程度,影响了通话质量,语音增强显得尤为重要^[1-3]。Boll^[4]提出从带噪语音频谱中减去估计的噪声谱得到增强后的语音功率谱的谱减法(spectral subtraction, SS)实现抑制噪声,该方法简单易行降噪效果好但引来音乐噪声。Berouti 等人^[5]引入过减因子有效平衡了语音失真和噪声残留。Thomson^[6]提出对每帧信号进

行正交窗功率谱估计的多窗谱估计法,比传统的周期图法只用一个数据窗估计带噪语音功率谱误差小。语音端点检测(voice activity detector, VAD)用来分析语音信号的起止点,利用前导无声段的噪声谱作为估计的噪声谱,对语音段进行特征提取得到有用信息^[7]。本文首先借助 Cohen^[8]提出的改进最小值控制递归平均算法来估计噪声谱,对带噪语音帧一阶平滑去除噪声,其次估计不同频带内语音存在概率,以带噪语音局部能量值和该帧窗口最小值的比值来判断该帧为语音是有声段还是无声段,如果当前帧为语音有声段,则该频段内的噪声谱和上一帧的噪声谱相等,否则实时搜索跟踪该频带带噪语音谱,求得最小值再乘偏置补

收稿日期:2016-11

* 基金项目:华中师范大学基本科研业务费专项资金(CCNU16GF0001)资助项目

偿因子得到估计的噪声功率谱值,最后结合深度神经网络进行数据训练,波形重构获取增强的语音信号。

2 改进谱减法结合深度神经网络的语音增强

传统谱减法通常用静音帧的统计平均值代替当前帧的噪声谱估计,会因该窗内噪声分量的不确定性残留部分噪声,频谱显示则为一系列随机的谱峰,在信号变化剧烈时产生“音乐噪声”^[9],此时噪声衰减或消失,但降低了语音可懂度。

2.1 多窗谱估计的谱减法

基于多窗谱估计(multiple-window-spectrum, MWS)算法的谱减法对同一数据序列加多个正交的数据窗分别求频谱值,然后加权平均得到比周期图法更准确的功率谱估计,有效减少音乐噪声,该方法步骤主要如下^[10]:1)对语音信号采用高通滤波预加重,提升高频幅值;2)根据语音信号短时平稳特性对带噪语音信号 $y(n)$ 分帧加汉明窗得到 n 个帧段,帧间重叠 m ,第 i 帧为 $y_i(n)$;3)对 $y_i(n)$ 作离散短时傅里叶变换得到第 i 帧语音信号频谱值 $Y_i(k)$, k 取值 $0 \sim N$, N 为帧长,分别求得其幅值 $|Y_i(k)|$ 和相角 $\theta(Y_i(k))$,得到第 i 帧信号左右两帧幅度谱均值 $|\bar{Y}_i(k)| = \frac{1}{2M+1} \sum_{j=-M}^{j=M} |Y_{i+j}(k)|$,以 i 帧为中心进行 $2M+1$ 帧平滑,令 $M=1$;4)根据 MATLAB 多窗谱估算函数 $pmtm$ 求

得分帧加窗后的平滑功率谱 $P_y(k, i) = \frac{1}{2M+1} \sum_{j=-M}^{j=M} P_y(k, i+j)$;5)语音开始前有一段静音段时间 IS ,对应帧长 NIS ,则噪声功率谱均值 $DIS(k) = \frac{1}{NIS} \sum_{i=1}^{NIS} P_y(k, i)$;6)由谱减公式计算增益函数 $G(k, i)$,过减因子 α 主要影响语音谱的失真程度,增益补偿因子 β 用来控制残留噪声以及音乐噪声,公式如下:

$$G(k, i) = \begin{cases} \beta P_y(k, i) - \alpha DIS(k) / P_y(k, i), \\ \beta P_y(k, i) - \alpha DIS(k) \geq 0; \\ \beta DIS(k) / P_y(k, i), \\ \beta P_y(k, i) - \alpha DIS(k) < 0. \end{cases} \quad (1)$$

7)谱减后的幅度谱 $|\hat{S}_i(k)| = G(k, i) |\bar{Y}_i(k)|$;8)最后 $\hat{S}_i(k)$ 和相角 $\theta(Y_i(k))$ 进行离散傅里叶逆变换(IDFT),恢复增强后的语音信号:

$$\hat{s}(n) = IDFT[|\hat{S}_i(k)| \exp[j\theta(Y_i(k))]] \quad (2)$$

2.2 自适应谱减系数

当带噪语音的幅度谱(功率谱)与估计出来的噪声谱相减出现负值时,对带噪语音谱引入过减因子 α 和增益补偿因子 β 弥补噪声过估计问题引起的音乐噪声^[11]。表 1 给出 α 和 β 取值对信噪比和 PESQ 值的影响。

表 1 不同谱减系数对信噪比和可懂度的影响

α, β	$\alpha=4, \beta=0.001$		$\alpha=4, \beta=0.098$		$\alpha=2.8, \beta=0.098$	
	信噪比	语音可懂度	信噪比	语音可懂度	信噪比	语音可懂度
White	9.670 3	1.680 7	9.803 6	1.803 6	9.973 1	1.230 9
Babble	7.421 9	1.421 9	7.495 6	1.495 6	8.396 5	1.092 4
Volvo	9.680 8	1.680 8	9.826 8	1.826 8	10.840 2	1.237 6
Machinegun	1.421 2	1.578 8	0.947 9	0.947 9	-0.818 8	0.987 9
F16	8.168 4	1.168 4	8.241 6	1.241 6	9.381 5	1.031 2

2.3 改进的最小值控制递归平均噪声功率谱估计

针对上述传统谱减法的残留噪声问题,本文结合改进的最小值控制递归平均(improved minima controlled recursive averaging, IMCRA)方法多次谱平滑处理,得到语音存在概率和语音起止点判决门限估计噪声谱^[12],对噪声谱更新后,当前第 i 帧噪声谱估计公式为:

$$\bar{P}_d(k, i) = \alpha_{isc} \bar{P}_d(k, i-1) + [1 - \alpha_{isc}] P_y(k, i-1) \quad (3)$$

式中: $\bar{P}_d(k, i)$ 表示第 i 个窗口的频谱值为 k 的噪声谱估计, $P_y(k, i-1)$ 表示前一帧的带噪语音平滑功率谱,根据文献^[7, 11]设置理想平滑系数 $\alpha_{isc} = 0.95$, $\bar{P}_d(k, i)$ 除以一个偏差补偿因子 $\beta_m (0 < \beta_m < 1)$,得到补偿后的噪声功率谱估

计为 $\hat{P}_d(k, i) = \bar{P}_d(k, i) / \beta_m$,与传统的 VAD 方法更新噪声相比,IMCRA 算法对噪声谱的突变有更好的适应性。

2.4 基于深度神经网络的语音增强

神经网络^[13](deep neural network, DNN)是一门新兴的模拟人脑结构机制的技术,具有高自适应、鲁棒性、记忆特性,利用 DNN 的声学模型提取语音和噪声的参数进行递归建模和 DNN 训练,通过循环的正向传播与误差反向调节的学习过程,神经元间的权值得到不断的修正^[14],最后进行波形重构,结合之前的 MWS 算法恢复重构后的增强语音信号,消除由谱减法产生的音乐噪声,深度神经网络结合多窗谱估计谱减法(DNN-MWS)算法的具体步骤如图 1 所示。

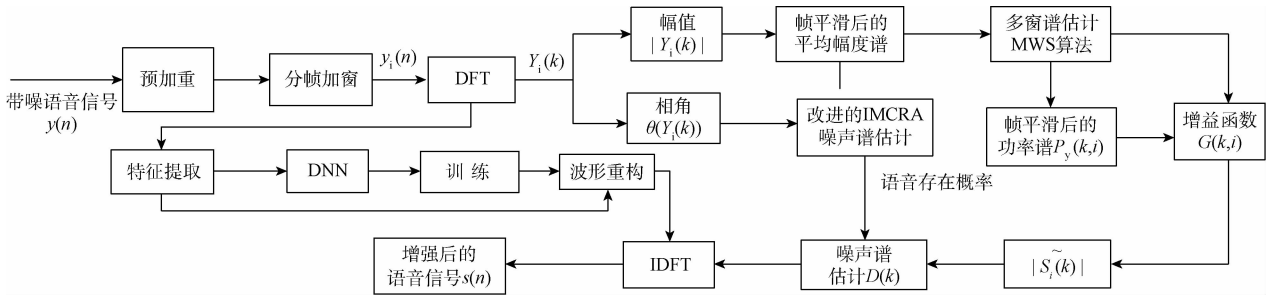


图 1 DNNk-MWS 算法的流程

2.5 仿真分析

实验音频为一段来自语音库的男声纯净语音信号 bluesky1.wav, 读作“蓝天, 白云, 碧绿的大海。”长度 3 s, 比特率 128 kbps, 频率 8 kHz; 分别叠加噪声库 NoiseX92 的白噪声(White)、说话噪声(Babble)、汽车噪声(Volvo)、机器滚轮声(Machinegun)、机舱噪声(F16); 然后用哈明窗对合成的带噪语音分帧加窗处理, 帧长 256 ms, 帧间叠接 128, 分别以 -10、-5、0、5、10 dB 的信噪比进行实验^[4]。以叠加 5 dB 的 babble 噪声为例, 语音时频域仿真图如图 2 所示。

图 2(a) 为采用语音端点检测方法的谱减法(VAD-SS), 增强后语音有失真, 背景噪声较少但伴随大量音乐噪声; 图 2(b) 为采用 IMCRA 算法的多窗谱估计法(IMCRA-MWS), $\alpha=4.0, \beta=0.098$, 增强后语音无明显失真, 噪声完全消除, 有部分音乐噪声; 图 2(c) 为 DNN-MWS 算法, 增强后语音失真小, 残留少量噪声和音乐噪声。

3 种算法的语音增强效果在不同信噪比的噪声强度下作用不同: 5 dB 时, VAD-SS 算法谱减后语音有一些截断效应, 伴随音乐噪声。IMCRA-MWS 算法对相位做出补偿, 信噪比得到提升, 音乐噪声减弱, 语音信号的能量谱衰减较少。DNN-MWS 算法结合递归神经网络模型进行训练, 语音增强效果最好, 有少量能量衰减, 但声音平滑可懂度较好。10 dB 高信噪比时, 噪声谱对语音的影响较小, 出现频谱过减现象, 自适应谱减系数对功率谱产生一定影响, DNN-MWS 算法作用不明显, 并附有时延。

图 3 柱状图表示 3 种算法 MOS 得分情况, 采用 F16 作为背景噪声采集了 6 位试听者对增强后语音信号的主观感受, 实验表明, DNN-MWS 算法的语音增强效果最理想, IMCRA-MWS 次之。

在使用谱减法降噪之前需要进行噪声估计, VAD-SS 方法计算的噪声段平均能量, IMCRA-MWS 方法用的是前导无话段估计噪声功率谱。DNN-MWS 方法用神经网络对语音信号加噪训练提取谱系数, 结合 IMCRA 算法对噪声谱实时更新, 有效地抑制背景噪声和音乐噪声, 图 3 示意, 图 4(a)、(b) 分别为叠加 -5~15 dB 的 White 和 F16 背景噪声信噪比比较。

表 2 3 种算法的输出信噪比比较

背景噪声	信噪比	3 种算法输出信噪比		
		VAD-SS	IMCRA-MWS	DNN-MWS
White	-5	1.807 2	4.795 8	5.947 7
	0	3.001 1	5.342 7	8.457 3
	5	7.204 5	9.973 1	10.850 7
	10	11.968 4	12.691 1	14.122 8
	15	16.003 3	17.401 5	18.801 6
Babble	-5	-0.228 3	0.136 6	0.480 8
	0	2.345 4	3.456 6	4.675 7
	5	7.039 8	8.396 5	8.304 8
	10	13.281 3	12.528 5	12.867 3
	15	15.863 3	17.169 3	17.328 1
Volvo	-5	2.881 7	4.378 6	7.813 3
	0	3.578 6	5.678 5	9.689 6
	5	9.619 2	10.840 2	16.710 3
	10	15.411 7	13.187 8	21.353 7
	15	17.267 7	17.706	25.018 3
Machinegun	-5	0.663	-0.573 8	0.099 3
	0	2.123 4	-0.234 4	2.654 3
	5	7.054 5	-0.818 8	3.468 6
	10	12.191 4	3.801 9	8.161 4
	15	15.942 7	9.594 6	13.52 7
F16	-5	0.983 4	1.496 5	2.901 3
	0	2.890 7	6.564 3	8.078 6
	5	8.300 4	9.381 5	9.872 7
	10	10.660 6	11.864 9	12.755 2
	15	13.395 8	13.910 3	14.050 4

3 结论

针对低信噪比下语音降噪和可懂度的一致性问题, 本文对 VAD-SS、IMCRA-MWS、以及 DNN-MWS 算法进行比较, 从信噪比、PESQ 和 MOS 得分来测评语音增强效果, 采用多窗谱自适应谱减法结合递归神经网络的语音增强算法的降噪效果明显。实验表明, 高信噪比时, 几种算

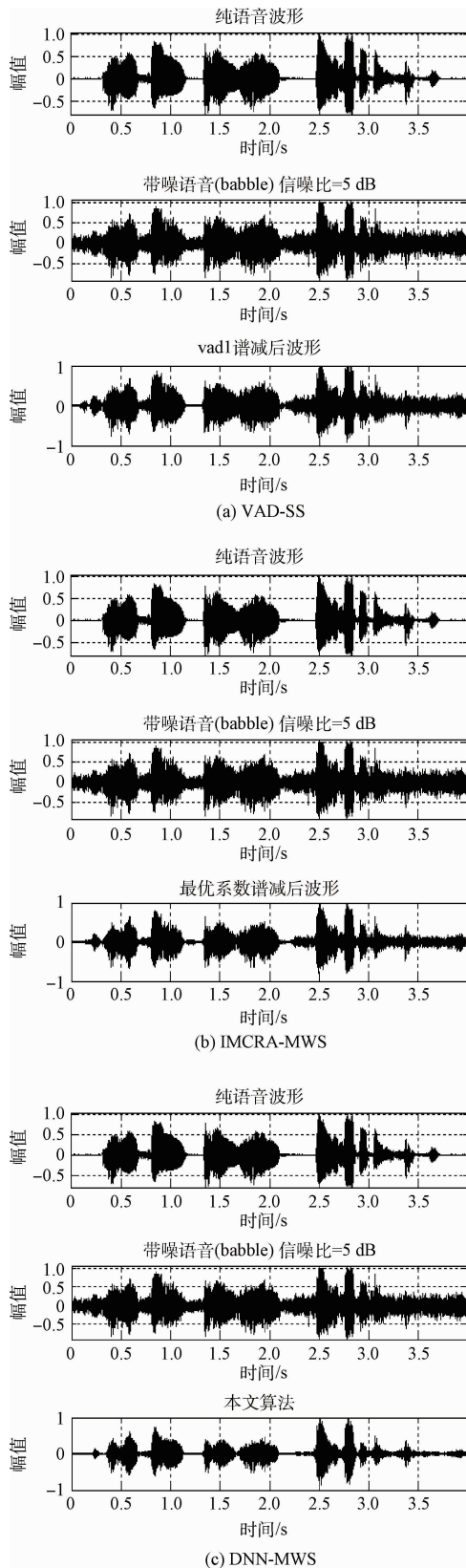


图 2 叠加 5 dB 噪声时,3 种算法的语音时频域仿真

法均有效去除了背景噪声,而低信噪比时,由于估计的噪

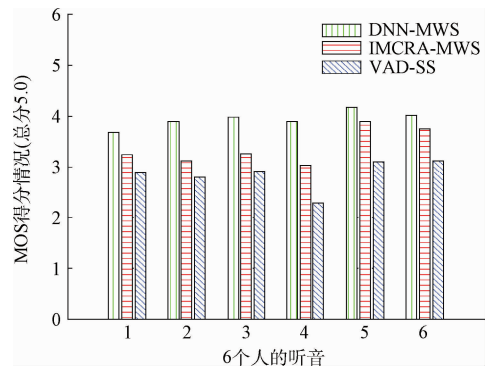


图 3 F16 机舱噪声下 MOS 分

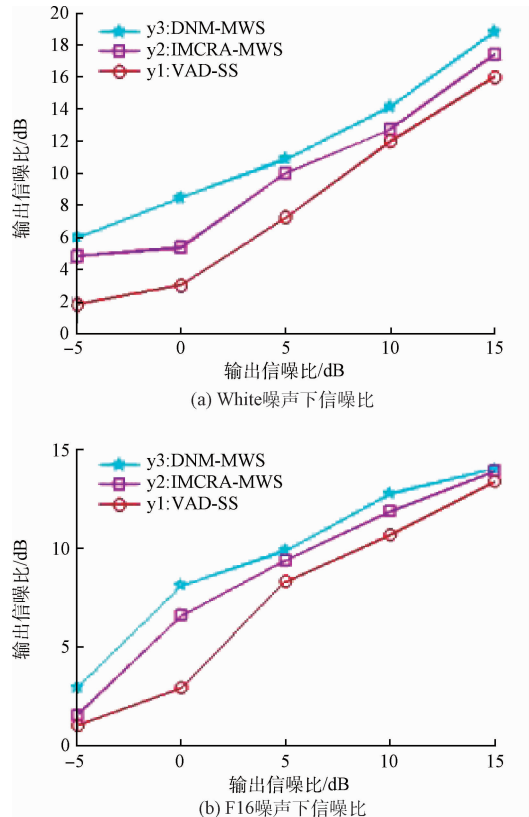


图 4 叠加 -5~15 dB 的 2 种噪声下的信噪比

声谱有一定的时延,采用 DNN-MWS 算法语音保真度更好,可懂度更高。

参考文献

- [1] 赵力. 语音信号处理 [M]. 北京: 机械工业出版社, 2003.
- [2] YAO Y, YI B SH, YAO Y Q. Speech enhancement under aviation noise[C]. International Conference on Wireless Communications, Networking, and Mobile Computing, 2006: 1-4244-0517-3.
- [3] 姚勇强, 易本顺, 姚远. 航空噪声背景下的语音端点

- 检测和语音增强[J]. 电声技术, 2006(1):36-39.
- [4] BOLL S. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Transactions on Acoustics Speech & Signal Processing, 1979, 27(3): 113-120.
- [5] BEROUTI M, SCHWARTZ R, MAKHOUL J. Enhancement of speech corrupted by acoustic noise[C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1979: 208-211.
- [6] THOMSON D J. Spectrum estimation and harmonic analysis[C]. Proceedings of the IEEE, 1982, 70(9): 1055-1096.
- [7] 韦晓东, 胡光锐, 任晓林. 应用倒谱特征的带噪语音端点检测方法[J]. 上海交通大学学报, 2000, 34(1): 185-188.
- [8] COHEN I. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging[C]. IEEE Transactions on Speech and Audio Processing, 2003, 11(5): 466-475.
- [9] 崔兆国, 周萍. 基于 TEO 能量谱减法的语音增强技术的研究[J]. 计算机应用与软件, 2014, 31(1): 2-3.
- [10] 韩芳, 靳宗信. 低信噪比下的端点检测算法研究[J]. 西北师范大学学报. 2016, 52(5): 988-1001.
- [11] 齐立萍, 孙昊, 杨鹏, 等. 基于参数自适应的改进谱减法[J]. 科学技术与工程, 2016, 16(3): 2-5.
- [12] 张建伟, 陶亮, 周健, 等. 基于改进谱平滑策略的 IMCRA 算法及其语音增强[J]. 计算机工程与应用, 2017, 53(1): 153-157.
- [13] XU Y, DU J, DAI L R, et al. A regression approach to speech enhancement based on deep neural networks[J]. IEEE Transactions on Audio Speech & Language Processing, 2015, 23(1): 7-19.
- [14] 黎林, 朱军. 基于小波分析与神经网络的语音端点检测研究[J]. 电子测量与仪器学报, 2013, 27(6): 528-534.

作者简介

姚远, 1979 年出生, 博士, 副教授, 研究方向为信号处理与感测技术。

E-mail: yaoyuan@mail. ccnu. edu. cn

王秋菊(通讯作者), 硕士研究生, 研究方向为语音信号处理。

E-mail: wqj20100802060@163. com

(上接第 66 页)

- [3] 张伟昆. 测试性分析与评估体系的研究[J]. 国外电子测量技术, 2015, 34(5): 38-43.
- [4] 张曼宁, 赵恒凯, 郑国莘. 毫米波段折射率结构常数的建模及敏感性分析[J]. 电子测量技术, 2016, 39(12): 150-154.
- [5] 韩先平. 大气折射误差快速修正方法研究[J]. 电子测量技术, 2016, 39(5): 57-60.
- [6] 陈沛, 贺旭东. 基于灵敏度分析的隔振器模型修正方法[J]. 国外电子测量技术, 2016, 35(4): 38-41.
- [7] 宁飞, 贺庚贤, 葛欣宏. 星载设备电源线传导发射的测量不确定度评定方法[J]. 电子测量技术, 2016, 39(1): 80-83.
- [8] 贾云涛, 胡耀元, 张建永, 等. 基于 MATLAB 的改进仪表测量精度的研究[J]. 电子测量技术, 2015, 38(9): 44-46.
- [9] 刘军. 一种多端口矢量网络分析仪误差校准简化方法[J]. 国外电子测量技术, 2016, 35(9): 29-33.
- [10] 史永彬, 于蒙, 李迪. 示波器测量脉冲信号测量结果的不确定度分析与评定[J]. 国外电子测量技术, 2016, 35(3): 50-53.
- [11] 王延年, 牛飞婷, 刘婷. 指纹图像质量评估方法[J]. 国外电子测量技术, 2015, 34(10): 32-36.
- [12] 邓欣, 刘晓东, 李保军, 等. 隔水管振动导致的 ADCP 流速测量误差分析[J]. 国外电子测量技术, 2015, 34(9): 26-28.
- [13] 王瑞宝. 基于蒙特卡洛法的微波功率测量不确定度评定[J]. 国外电子测量技术, 2015, 34(7): 28-31.
- [14] 周锋, 王瑞宝. 一种选取有效测量点简化测量不确定度计算的方法[J]. 国外电子测量技术, 2015, 34(6): 37-39.
- [15] 葛欣宏, 刘遒, 蔚素升, 等. 25 Hz~10 kHz 电源线传导发射的测量不确定度评定方法研究[J]. 国外电子测量技术, 2015, 34(7): 43-46.

作者简介

李凌梅, 1982 年出生, 硕士, 工学硕士, 主要研究方向为测试计量技术及仪器、长度几何量计量与检定。

Email: tjllm333@126. com

戴庆生, 1964 年出生, 副高, 主要从事长度几何量计量与检定。

张欣, 1984 年出生, 工程师, 主要从事长度几何量计量与检定。

杨佳, 1964 年出生, 硕士, 主要从事能源计量与检定。

王宜, 1990 年出生, 本科, 主要从事长度几何量计量与检定。