

基于 ISD 优化 SSA-VMD 算法的 自适应英文语音增强

高楠¹, 毛露露^{2†}, 梁利利¹

(1. 咸阳师范学院, 陕西 咸阳 712000; 2. 成都纺织高等专科学校, 四川 成都 611731)

摘 要:针对传统语音增强方法存在滤除有效信号且增强性能欠佳的问题,提出了一种基于 ISD 优化 SSA-VMD 算法的自适应英文语音信号增强方法。首先利用 SSA-VMD 算法自适应获取输入语音信号 VMD 关键参数的最优组合,再对其进行分解得到各模态分量,基于 ISD 原理对各模态分量与输入信号的相似度进行评估,进而筛选出噪声模态分量,通过小波阈值法对其进行降噪,最后重构信号得到增强后的语音。实验结果表明,利用所提方法处理英文语音信号,信噪比更高,均方误差更小,语音增强效果也更明显,具有较好的适用性。

关键词:SSA-VMD 算法; ISD 原理; 自适应; 语音增强

中图分类号:TP301.6

文献标识码:A

Adaptive English Speech Enhancement Based on ISD Optimized SSA-VMD Algorithm

GAO Nan¹, MAO Lulu^{2†}, LIANG Lili¹

(1. Xianyang Normal University, Xianyang, Shaanxi 712000, China;

2. Chengdu Textile College, Chengdu, Sichuan 611731, China)

Abstract:Due to traditional speech enhancement methods have the problems of filtering out effective signals and poor enhancement performance, an adaptive English speech signal enhancement method based on ISD optimized SSA-VMD parameters is proposed. Firstly, the SSA-VMD algorithm is used to adaptively obtain the optimal combination of VMD key parameters for the input speech signal, and then it is decomposed to obtain each modal component. The similarity between each modal component and the input signal is evaluated based on the ISD principle, and then the noise modal components are selected, and the noise is reduced by wavelet threshold method. Finally, the signal is reconstructed to obtain enhanced speech. The experimental results show that using the proposed method to process English speech signal, the SNR is higher, the MSE is smaller, and the speech enhancement effect is more obvious, which has good applicability.

Key words:SSA-VMD algorithm; principle of ISD; adaptive; speech enhancement

语音信号高度不稳定,特别是在恶劣噪声环境下,对语音的质量和可懂度会产生严重影响。语音增强在一定程度上能够有效减弱或抑制这些环境噪声,改善语音质量^[1]。

目前主流的语音增强方法有谱减法、统计模型方法、子空间法、神经网络方法等^[2-4]。早期, Boll^[5]提出谱减法,随后经典的维纳滤波法、最小均方差估计法也相继被提出^[6],这几种方法在语音

增强方面确实都具有一定的效果,但存在会产生语音失真和音乐噪声的缺点。近些年,也出现了将循环神经网络、卷积神经网络、生成对抗网络等深度学习用于语音信号增强^[7-9],这类方法对非平稳噪声下的语音增强效果更好,但计算复杂,对硬件的内存资源要求较高。为了适应非线性非平稳信号处理的需求,Poovarasan 等^[10]采用经验模态分解对语音信号进行增强,但该方法存在模态混叠问题,容易产生语音失真,为了解决该问题,由 Dragomiretskiy 等^[11]提出了变分模态分解(VMD)方法,该方法可以有效避免由于端点检测而带来的模态混叠问题。Ram 等^[12]随后基于 VMD 算法进行语音增强,也得到了较好的增强效果,但由于 VMD 的模态数目和惩罚因子选择难度较大,容易产生欠分解或过分解的问题。李宏等^[13]采用灰狼优化(GWO)算法优化 VMD 算法的参数组合,该方法能够有效去除语音中的噪声,但存在可能陷入局部最优解的缺点。另外,现有的语音增强研究多为泛化的语音增强方法,但不同语种的语法、机理、产生方式等导致发音存在很大差异,出现了语音增强效果差,自适应力不足的情况。

针对上述问题,提出了一种适用于英文语言特性的单一语音增强方法,该方法基于麻雀搜索算法(SSA)实现了 SSA-VMD 算法的参数自适应寻优,再通过板仓-斋藤距离(ISD)筛选出噪声模态分量,并利用小波阈值法对其进行去噪。从实验结果可以看出,该方法能够自适应地对输入语音信号进行增强处理,噪声抑制效果明显,信噪比也得到了提升。

1 语音信号增强原理

1.1 VMD 原理

VMD 是一种自适应、非递归的信号处理方法,可以将输入语音分解为多个本征模态函数(IMF)^[14]。设定 $s(t)$ 为输入的语音信号,可通过 VMD 将其自适应分解为 K 个分量 $v_k(t)$, 定义为:

$$v_k(t) = A_k(t) \cos [\varphi_k(t)] \quad (1)$$

式中, $A_k(t)$ 是幅值, $\cos [\varphi_k(t)]$ 是相位函数, t 是时间变量, $k = 1, 2, \dots, K$ 。对各模态进行 Hilbert 变换,可得到对应的单侧谱,经指数因子修正,再进行信号解调得到各模态带宽,则构造出约束变分数学模型为:

$$\begin{cases} \min_{\{v_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * v_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s. t. } \sum_k v_k(t) = s(t) \end{cases} \quad (2)$$

式中, $v_k = \{v_1, v_2, \dots, v_K\}$, 为所有 IMF 分量集合; $\omega_k = \{\omega_1, \omega_2, \dots, \omega_K\}$, 为中心频率集合; ∂_t 是函数对 t 的偏导数; $\delta(t)$ 表示 Dirac 分布; $*$ 表示卷积运算。为了对式(2)模型进行求解,引入二次惩罚函数项与 Lagrange 乘子:

$$\begin{aligned} L(\{v_k\}, \{\omega_k\}, \lambda) = & \\ & \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * v_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \\ & \left\| s(t) - \sum_k v_k(t) \right\|_2^2 + \langle \lambda(t), s(t) - \sum_k v_k(t) \rangle \end{aligned} \quad (3)$$

式中, λ 为 Lagrange 乘子, α 为二次惩罚因子。

进一步,利用交替方向乘子法对 $s(t)$ 进行分解,对 v_k^{n+1} 、 ω_k^{n+1} 和 λ^{n+1} 进行交替更新并进行双重提升,在 $\omega \geq 0$ 时,可得:

$$\hat{v}_k^{n+1}(\omega) = \frac{\hat{s}(\omega) - \sum_{i=1, i \neq k}^K \hat{v}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (4)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{v}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{v}_k(\omega)|^2 d\omega} \quad (5)$$

$$\hat{\lambda}^{n+1} = \hat{\lambda}^n + \tau [\hat{s}(\omega) - \sum_k \hat{v}_k^{n+1}(\omega)] \quad (6)$$

式中, τ 表示噪声容限。在预设阈值判别条件 $\epsilon > 0$ 时,终止条件如下:

$$\frac{\sum_k \|\hat{v}_k^{n+1} - \hat{v}_k^n\|_2^2}{\|\hat{v}_k^n\|_2^2} < \epsilon \quad (7)$$

式中, ϵ 表示收敛阈值,若满足式(7),则迭代结束,否则返回继续迭代寻优。

1.2 SSA-VMD 算法的参数寻优

在 VMD 算法中,分解的模态数 K 和惩罚因子 α 相互制约,共同影响着 VMD 的分解效果。因此,根据输入信号的不同,得到最优参数组合 $[K, \alpha]$ 是实现 VMD 自适应分解的关键。SSA 是 Xue 等^[15]提出的一种群智能优化算法,本文将 SSA 用于语音信号 VMD 的关键参数寻优。

SSA-VMD 算法参数寻优的计算步骤如下:

Step1:初始化麻雀种群,并定义种群数、迭代最大限制次数、寻优边界范围、VMD 参数组合

$[K, \alpha]$ 的优化范围等相关参数。

Step2: 计算初始适应度函数值并进行排序, 确定当前最好与最差个体。

Step3: 对发现者的位置进行更新, 更新描述如下:

$$X_{i,d}^{m+1} = \begin{cases} X_{i,d}^m \cdot \exp\left(-\frac{i}{\beta \cdot \text{iter}_{\max}}\right), & \text{if } R_2 < P \\ X_{i,d}^m + Q \cdot M, & \text{if } R_2 \geq P \end{cases} \quad (8)$$

式中, $X_{i,d}^{m+1}$ 表示麻雀位置; m 是目前的迭代次数; β 是随机数, $\beta \in (0, 1]$; $\text{iter}_{\max}$ 代表最大迭代次数; R_2 为预警值, $R_2 \in [0, 1]$; P 为安全值, $P \in [0.5, 1]$; Q 代表服从正态分布的随机数; M 为一行多维矩阵, 其元素均是 1。

Step4: 对加入者位置进行更新, 数学模型表示为:

$$X_{i,d}^{m+1} = \begin{cases} Q \cdot \exp\left(\frac{X_W^m - X_{i,d}^m}{i^2}\right), & \text{if } i > n/2 \\ X_E^{m+1} + |X_{i,d}^m - X_E^{m+1}| \cdot A^+ \cdot M, & \text{otherwise} \end{cases} \quad (9)$$

式中, X_E 、 X_W 分别为当前全局最好和最差位置; A 为一行多维矩阵, 且元素随机为 1 或 -1, $A^+ = A^T (AA^T)^{-1}$ 。

Step5: 对于意识到危险的麻雀位置, 其更新表示如下:

$$X_{i,d}^{m+1} = \begin{cases} X_B^m + \sigma \cdot |X_{i,d}^m - X_B^m|, & \text{if } f_i > f_B \\ X_{i,d}^m + H \cdot \left(\frac{|X_{i,d}^m - X_W^m|}{(f_i - f_W) + \mu}\right), & \text{if } f_i = f_B \end{cases} \quad (10)$$

式中, X_B 为当前全局最好位置; σ 代表步长参数; H 是随机数, $H \in [-1, 1]$; f_i 表示麻雀适应度值; f_B 、 f_W 分别为当前全局最好和最差适应度值; μ 取值为极小的常数。

Step6: 更新适应度值, 并选出目前最优麻雀位置。

Step7: 进行迭代运算并得到最优值, 并判断当前结果是否优于上一次迭代, 如果是, 对其进行更新。否则返回 Step2 并继续执行后续步骤, 当满足停止条件时, 输出 VMD 最优的参数组合 $[K, \alpha]$ 。

在对 VMD 进行参数寻优时, 考虑到 SSA 算法适应度函数构造的关键点, 本文选取最小包络嫡为 SSA 算法的适应度函数值。语音信号处理的 SSA-VMD 算法参数寻优流程如图 1 所示。

1.3 基于 ISD 的信号重构原理

输入语音信号 $s(t)$ 经 SSA-VMD 算法得到 K 个 IMF 分量, 计算其概率密度函数, 再利用 ISD 评

估 $s(t)$ 与各 IMF 分量的概率密度函数间的相似度, 据此筛选出有效分量, 进而完成语音信号重构^[16]。

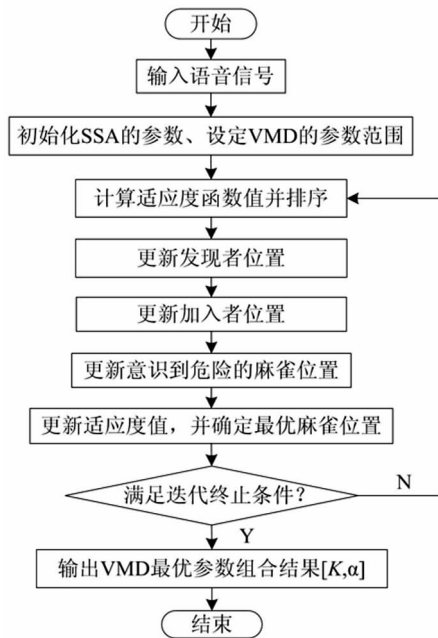


图 1 SSA-VMD 算法参数寻优流程图

ISD 的计算表达式定义如下:

$$\text{ISD}(p, q) = \sum_{i=1}^n \frac{p(i)}{q(i)} - \log \frac{p(i)}{q(i)} - 1 \quad (11)$$

式中, p, q 是信号的概率密度函数。当相邻两个 ISD 的增量最大时, 表示该 IMF 分量之后的相似度急剧降低, 噪声较大, 有效信号较少。因此, 以增量最大处作为转折点来筛选噪声分量。若 ISD 增量最大时处于 IMF_a 与 IMF_{a+1} 间, 则前 a 个 IMF 分量被作为有效模态分量, 其余分量被作为噪声分量进行处理。

对于噪声分量, 利用小波阈值法对其进行去噪, 选用 Sym8 为小波基, 分解层数设定为 3, 根据 Heursure 阈值准则对小波系数进行处理, 最后重构信号, 可表示如下:

$$s'(t) = \sum_{i=1}^a \text{IMF}(t) + \sum_{i=a+1}^K \text{WT}[\text{IMF}(t)] \quad (12)$$

其中, $s'(t)$ 为重构语音; WT 代表小波阈值处理。

2 实验及性能评估

2.1 实验验证

为了对本文所提方法进行验证, 实验选用来自 NOIZEUS 语料库的英文语音数据^[17]。选取该语料库中的一段发音为: “Smoke poured out of every crack” 的英文语音信号, 采样频率为 8 kHz, 对其

叠加随机高斯白噪声,信噪比为 5 dB。纯净语音及含噪语音的波形分别如图 2、图 3 所示。

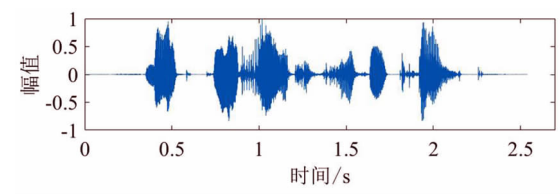


图 2 原始纯净语音信号波形图

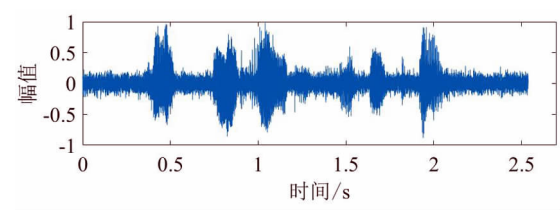


图 3 高斯白噪声干扰的语音信号波形图

利用本文所提方法对含噪语音进行增强处理,需要基于 SSA-VMD 算法得到分解模态数 K 与惩罚因子 α 的最优参数组合,根据 SSA-VMD 算法的参数寻优流程,以最小包络熵作为适应度值得到的适应度函数迭代搜索变化曲线如图 4 所示。从图 4 中可以看出,在第 15 次迭代时最小包络熵的结果为 9.2835,并开始趋于稳定,在迭代寻优到 30 代后,收敛稳定值保持不变,此时,输出的分解模态数 K 与惩罚因子 α 的最佳组合为[10,2301]。

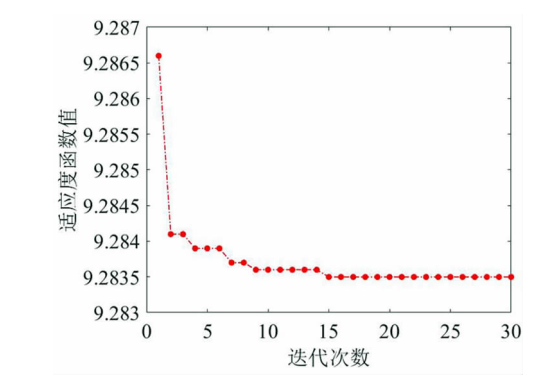


图 4 适应度函数迭代搜索变化曲线

将上述 $[K, \alpha]$ 用于 VMD 分解, $K = 10$ 时,得到的各 IMF 分量如图 5 所示。由图 5 可知,语音信号被分解为 10 个 IMF 分量,且语音信号中的有效成分主要被集中在前 6 个 IMF 分量中,IMF7 至 IMF10 分量以噪声成分为主。因此,有必要对 VMD 分解得到的各 IMF 分量做进一步的筛选及信号降噪处理。

根据 ISD 原理选择模态分量,计算输入语音

信号 $s(t)$ 与各 IMF 分量的概率密度函数间的 ISD 值,得到 ISD 值的变化趋势如图 6 所示。

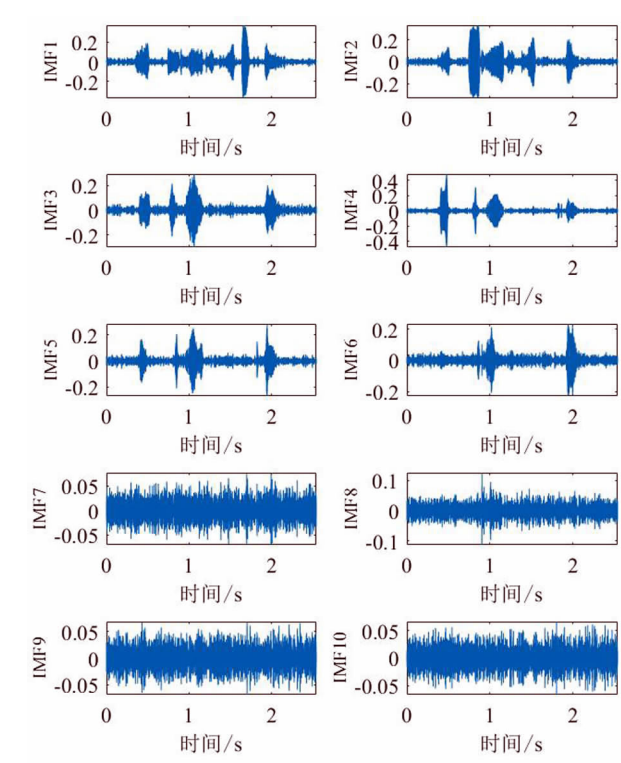


图 5 VMD 分解获取的各 IMF 分量

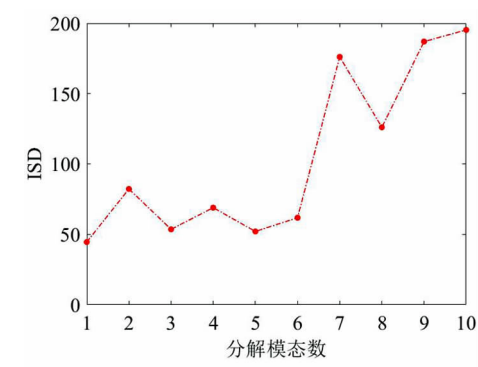


图 6 概率密度函数的 ISD

从图 6 可以看出,语音信号经 VMD 分解后得到低频至高频 IMF 分量的 ISD 值整体呈上升趋势,且 ISD 增量最大值出现在 IMF6 与 IMF7 之间,因此,选取 IMF1 至 IMF6 为有效分量,IMF7 至 IMF10 被作为噪声分量进行小波降噪。

对噪声分量完成降噪后,再将其与有效分量结合实现语音重构,得到的重构结果如图 7 所示。从图 7 可以看出,重构语音信号较好地保留了原始语音的基本特征,且相比于加噪语音信号,高频噪声部分明显被滤除。

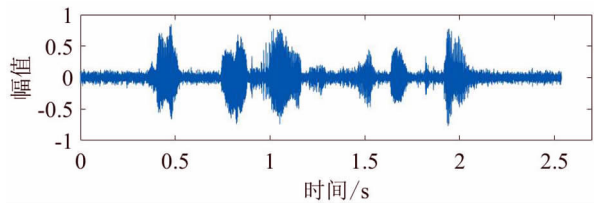


图 7 重构的增强语音信号波形图

2.2 性能评估

为了进一步验证所提方法的语音增强性能,使用信噪比(SNR)和均方误差(MSE)这两种评价指标进行评估,定义如下:

$$SNR = 10\lg \frac{\sum_{i=1}^N \tilde{s}_i^2}{\sum_{i=1}^N [s'_i - \tilde{s}_i]^2} \tag{13}$$

$$MSE = \frac{1}{N} \sum_{i=1}^N [s'_i - \tilde{s}_i]^2 \tag{14}$$

式中, s'_i 为重构信号; $\tilde{s}_i$ 为纯净信号; N 为信号长度; $i = 1, 2, \dots, N$ 。通常, SNR 相对越大, MSE 相对越小,说明语音信号增强的效果越好。

为评估本文方法的性能,进行了语音信号的对比实验测试,分别采用小波阈值法、VMD 相关系数法和本文方法对加噪语音信号进行增强处理,将不同方法实验得到的 SNR 和 MSE 进行对比,结果如表 1 所示。

表 1 不同语音增强方法的 SNR 和 MSE 结果对比		
方法	SNR	MSE
小波阈值法	5.4147	0.0715
VMD 相关系数法	7.6470	0.0553
本文方法	8.6087	0.0495

由表 1 可知,使用本文所提方法,有效提高了语音信号的质量,相比于其他两种方法,得到了更大的 SNR 和更小的 MSE,证明了使用基于 SSA-VMD 算法自适应增强英文语音信号的有效性和优越性。

3 结 论

为了提高语音增强性能,可靠地滤除语音信号中的噪声分量,实现自适应的英文语音信号增强,提出了一种基于 ISD 优化 SSA-VMD 算法关键参数的自适应英文语音信号增强方法。该方法可根据输入语音信号,利用 SSA-VMD 算法自适应获取 VMD 关键参数的最优组合 $[K, \alpha]$,完成信号

的 VMD 分解,再根据 ISD 原理对 IMF 分量进行筛选,基于小波阈值法对选出的噪声分量进行降噪,实现语音信号重构。相比于小波阈值法、VMD 相关系数法,本文所提的自适应语音增强方法能够有效地对英文语音信号进行降噪,信噪比和均方误差结果均为最优,语音增强效果也更明显,具有较好的适用性。

参考文献

[1] LEE S, HAN D K, KO H. Single-channel speech enhancement method using reconstructive NMF with spectrotemporal speech presence probabilities[J]. Applied Acoustics, 2016, 117(B): 257–262.

[2] NICOLSON A, PALIWAL K K. Deep learning for minimum mean-square error approaches to speech enhancement[J]. Speech Communication, 2019, 111:44–55.

[3] JADDA A, PRABHA I S. Speech enhancement via adaptive Wiener filtering and optimized deep learning framework[J]. International Journal of Wavelets, Multiresolution and Information Processing, 2023, 21(1):2250032.

[4] YOU C H, MA B. Spectral-domain speech enhancement for speech recognition[J]. Speech Communication, 2017, 94: 30–41.

[5] BOLL S F. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Transactions on Acoustics Speech and Signal Processing, 1979, 27(2):113–120.

[6] 董胡, 刘刚, 马振中. 基于自适应 MMSE-LSA 与 NMF 的语音增强算法[J]. 探测与控制学报, 2021, 43(4): 81–85.

[7] 王志杰, 张学良. 基于双路径循环神经网络的单通道语音增强[J]. 信号处理, 2021, 37(10): 1872–1879.

[8] CHENG J M, LIANG R Y, ZHAO L. DNN-based speech enhancement with self-attention on feature dimension[J]. Multimedia Tools and Applications, 2020, 79: 32449–32470.

[9] 尹文兵, 高戈, 曾邦, 等. 基于时频域生成对抗网络的语音增强算法[J]. 计算机科学, 2022, 49(6): 187–192.

[10] POOVARASAN S, CHANDRA E. Speech enhancement using sliding window empirical mode decomposition and hurst-based technique[J]. Archives of Acoustics, 2019, 44 (3): 429–437.

[11] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531–544.

[12] RAM R, MOHANTY M N. Performance analysis of adaptive variational mode decomposition approach for speech enhancement[J]. International Journal of Speech Technology, 2018, 21(2): 369–381.

[13] 李宏, 李定文, 朱海琦, 等. 一种优化的 VMD 算法及其在语音信号去噪中的应用[J]. 吉林大学学报(理学版), 2021, 59 (5): 1219–1227.

[14] LU J Y, YUE J K, ZHU L J, et al. Variational mode decomposition denoising combined with improved Bhattacharyya distance[J]. Measurement, 2020, 151(6):107283.

[15] XUE J K, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science & Control Engineering, 2020, 8(1): 22–34.

[16] 王冬梅, 肖超利, 路敬祎. 基于 VMD-ISD 的天然气管道泄漏信号去噪研究[J]. 压力容器, 2021, 38(9): 46–54.

[17] HU Y, LOIZOU P C. Subjective evaluation and comparison of speech enhancement algorithms[J]. Speech Communication, 2007, 49(7):588–601.