色噪声下基于白化频谱重排鲁棒主成分分析的语音增强算法

罗勇江* 杨腾飞 赵 冬

(西安电子科技大学电子工程学院 西安 710071)

摘 要：基于鲁棒主成分分析(RPCA)的单通道语音增强算法是高斯白噪声环境下语音增强的一种重要处理手段，但其对低秩语音分量处理效果欠佳且无法较好地抑制色噪声。针对此问题，该文提出一种基于白化频谱重排RPCA的改进语音增强算法(WSRRPCA)，通过优化噪声白化模型，将色噪声语音增强转换成白噪声语音信号处理，利用频谱重排改进RPCA语音增强处理算法，从而获得色噪声环境下语音信号处理性能的整体提升。仿真实验表明，该算法能够较好地实现色噪声环境下的语音增强，且相对于其他算法具有更佳的噪声抑制和语音质量提升能力。

关键词：语音增强；鲁棒主成分分析；色噪声；噪声白化；频谱重排

中图分类号：TN912.35 **文献标识码**：A **文章编号**：1009-5896(2021)12-3671-09

DOI: 10.11999/JEIT200594

Speech Enhancement Algorithm Based on Robust Principal Component Analysis with Whitened Spectrogram Rearrangement in Colored Noise

LUO Yongjiang YANG Tengfei ZHAO Dong

(*School of Electronic Engineering*, *Xidian University*, *Xi'an* 710071, *China*)

Abstract: The Robust Principal Component Analysis (RPCA) based speech enhancement algorithm plays an important role for single channel speech processing in white Gaussian noise environment, but it has a poor processing effect on low-rank speech components and can not well suppress color noise. In view of this problem, an improved speech algorithm based on Whitening Spectrum Rearrangement RPCA (WSRRPCA) is proposed in this paper, which by optimizing the noise whitening model, color noise speech enhancement is converted into white noise speech signal processing, and spectrum rearrangement is used to improve RPCA speech enhancement processing algorithm to obtain an overall improvement in speech signal processing performance in a colored noise environment. Simulation experiments show that this algorithm can better achieve speech enhancement in a colored noise environment, and has better noise suppression and speech quality improvement capabilities than other algorithms.

Key words: Speech enhancement; Robust Principal Component Analysis (RPCA); Colored noise; Noise whitening; Spectrogram rearrangement

1 引言

语音增强技术是语音信号处理的重要手段，用于抑制含噪语音中的背景噪声，单通道语音增强则是最具挑战性的应用需求[1]之一。于是语音增强技术开始被许多学者所研究，并提出了多种传统的单通道语音增强算法，例如频谱减法[2]、最小均方误差估计算法[3]和维纳滤波算法[4]等，这些算法在白噪声下的性能较好，但在色噪声环境中性能下降。

基于子空间的语音增强算法主要以矩阵分析为基础，并在时域中实现语音增强。Ephraim等人[5]提出了针对含噪语音信号协方差矩阵使用特征值分解的语音增强方法，但该方法的前提是假设噪声为白噪声，对于其他噪声并不能取得较好的效果。Yi等人[6]对该方法进行了改进，提出了广义子空间方法，该方法采用非酉变换将含噪信号分别投影到信号加噪声子空间和噪声子空间，通过对噪声子空间中的信号分量进行置零，并保留信号子空间中的分量来估计纯净语音信号，最终实现对语音信号的增强。该方法中的变换具有内置的预白化功能，可以处理含色噪声语音信号，但该方法需要纯净语音信号样本，计算复杂度较大，不易实现。

相对于传统的语音增强算法，鲁棒主成分分析(Robust Principal Component Analysis, RPCA)作为一种实现矩阵低秩稀疏分解的模型，已用于单通道语音增强中，且具有较好的语音增强性能。纯净

收稿日期：2020-07-20；改回日期：2021-03-25；网络出版：2021-06-03
*通信作者： 罗勇江 yjluo@mail.xidian.edu.cn

语音的能量主要集中在共振峰处，在时频域表现出稀疏的特征，背景噪声的频谱总是呈现重复的模式，具有低秩特性。在基于RPCA的单通道语音增强算法[7]中，含噪语音的时频幅度矩阵可以被RPCA分解为代表噪声分量的低秩矩阵和代表语音分量的稀疏矩阵，然后对稀疏矩阵进行时域重建，实现语音和噪声分离。文献[8]对原始的RPCA进行扩展，将含噪语音谱分解为噪声结构矩阵、纯净语音矩阵和剩余噪声矩阵3个子矩阵，直接将语音信号从噪声中分离。文献[9]将RPCA与语音与噪声词典的稀疏表示相结合，使用一种新的交替优化算法和基于互相关的最小角度及相关标准的语音与噪声词典的稀疏编码来对含噪语言进行分解。相对传统的方法，基于RPCA的语音增强算法对于白噪声环境下的普通语音信号增强具有较好的性能，但存在低秩语音分量丢失和色噪声的抑制能力较差的问题，主要表现为：(1)该算法对于低秩语音分量信号处理效果较差，因为语音中的浊音在时域表现出明显的周期性，在时频域中则为低秩性，因此这部分语音会被RPCA错误分解到低秩矩阵中作为噪声[10,11]；(2)该算法在色噪声环境下的语音增强性能较弱，因为白噪声具有较为平坦的时频谱特性且其时频谱的秩近似于1，从而表现为良好的低秩性，但色噪声的时频谱特性并不均匀且其时频谱不具有低秩特性。

为了解决上述问题，本文利用噪声白化和频谱重排改进基于RPCA的单通道语音增强算法，提出一种基于白化频谱重排鲁棒主成分分析(Whitened Spectrogram Rearrangement Robust Principal Component Analysis, WSRRPCA)的单通道语音增强算法，提升色噪声环境下的语音增强效果。

2 基于RPCA的单通道语音增强系统

基于RPCA的单通道语音增强系统由分析-去噪-综合(Analysis-Modification-Synthesis, AMS)框架实现，包括3个阶段：(1)分析阶段，通过短时傅里叶变换(Short Time Fourier Transform, STFT)将时域的含噪语音转换为时频幅度矩阵；(2)去噪阶段，对含噪语音的时频幅度矩阵进行某种修改；(3)合成阶段，对修改后的时频幅度矩阵每一列进行离散傅里叶逆变换(Inverse Discrete Fourier Transform, IDFT)，然后利用叠接相加法(Over-Lap-Add, OLA)合成时域的增强语音系统框图如图1所示。

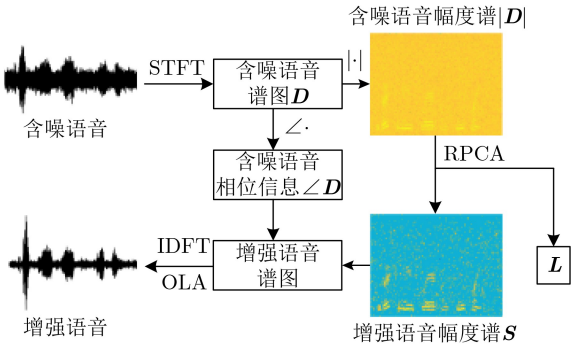

图 1 基于RPCA的单通道语音增强算法的系统框图

RPCA模型的输入数据为矩阵$|\boldsymbol{D}| \in \boldsymbol{R}^{m\times n}$，它是含噪语音的时频幅度矩阵，具有$n$个短时帧，每一帧有$m$个频谱元素。利用RPCA可以将$|\boldsymbol{D}|$分解为两个输出部分：对应于噪声的低秩矩阵$\boldsymbol{L} \in \boldsymbol{R}^{m\times n}$和对应于语音的稀疏矩阵$\boldsymbol{S} \in \boldsymbol{R}^{m\times n}$，3个矩阵的数学关系表示为

$$|\boldsymbol{D}| = \boldsymbol{L} + \boldsymbol{S} \tag{1}$$

未知的矩阵$\boldsymbol{L}$和$\boldsymbol{S}$可以通过解决如下的凸优化问题获得[12]

$$\left.\begin{aligned} \min \quad & \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{S}\|_1 \\ \text{s.t} \quad & |\boldsymbol{D}| = \boldsymbol{L} + \boldsymbol{S} \end{aligned}\right\} \tag{2}$$

其中，$\|\cdot\|_*$表示矩阵的核-范数，$\|\cdot\|_1$表示矩阵的1-范数，λ是一个正的加权参数。目前已开发出许多算法来解决式(2)中的凸优化问题，如加速近端梯度法[13]、增广拉格朗日乘子(Augmented Lagrange Multiplier, ALM)法[14]和奇异值阈值法[15]。其中，ALM算法既实用又高效，可以通过如下方式解决式(2)中的凸优化问题

$$\begin{aligned} \min \quad & \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{S}\|_1 + \langle \boldsymbol{Y}, |\boldsymbol{D}| - \boldsymbol{L} - \boldsymbol{S}\rangle \\ & + \frac{\mu}{2}\||\boldsymbol{D}| - \boldsymbol{L} - \boldsymbol{S}\|_{\mathrm{F}}^2 \end{aligned} \tag{3}$$

其中，$\|\cdot\|_{\mathrm{F}}$表示矩阵的F-范数，$\langle\cdot,\cdot\rangle$表示矩阵内积，$\boldsymbol{Y}$是拉格朗日乘数，$\mu$是正标量，$\lambda$是加权参数。由于$\boldsymbol{Y}$和$\mu$具有动态调整的能力，因此$\lambda$是唯一需要确定的参数，参数的选择方法将在后续讨论。

对于高斯白噪声环境下的语音增强，基于RPCA的语音增强算法的核心思想是通过低秩稀疏分解，将含噪语音的时频谱分解为代表噪声分量的低秩矩阵和代表语音分量的稀疏矩阵，然后仅对稀疏矩阵进行时域重建，实现语音和噪声分离。因此对于一些具有弱稀疏特性的语音分量将被分解到低秩矩阵中，从而造成分量的丢失。而对于色噪声条件下的语音增强，由于色噪声的时频特性并不具有低秩特性，常规的基于RPCA算法较难抑制色噪声。

3 基于WSRRPCA的单通道语音增强算法

本文提出的基于WSRRPCA的单通道语音增强算法从将色噪声处理转化为白噪声和破坏低秩语音分量信号的角度出发，构建合理的白化色化模型和频谱重排方法，提高色噪声环境中语音增强性能。

3.1 噪声白化

白噪声频谱是近乎平坦的，其时频幅度矩阵的秩可以被认为等于1，具有出色的低秩特性。色噪声时频图中的能量分布不均匀，难以保证较好的低秩性。因此基于RPCA的单通道语音增强算法在色噪声中的性能不如在白噪声中的性能。为获得较好的色噪声环境下的语音增强性能，本文使用噪声白化将色噪声转变为白噪声。色噪声的白化可以通过Cholesky分解或基于线性预测的有限长单位冲激响应(Finite Impulse Response, FIR)滤波器来实现[16,17]。Cholesky分解方法需要将信号修改为Hankel矩阵模型，这会使信号结构改变，并且后续处理步骤无法直接应用。相反，基于线性预测的FIR滤波器仅改变信号的幅度和相位且能保留信号模型，它可以使用过去样本的线性加权组合来预测噪声信号$v(n)$在n时刻的值，预测值$\hat{v}(n)$可以表示为[18]

$$\hat{v}(n)=\sum_{k=1}^{P}a_k v(n-k) \tag{4}$$

其中，a_k和P分别为线性预测器的系数和阶数。

预测均方误差为

$$E(n)=\sum_{n=0}^{N-1}e^2(n)=\sum_{n=0}^{N-1}\left[v(n)-\sum_{k=1}^{P}a_k v(n-k)\right]^2 \tag{5}$$

$v(n)$的自相关为

$$r_{vv}(m)=\sum_{n=0}^{N-1}v(n)\,v(n-m) \tag{6}$$

根据最小均方误差准则，$E(n)$最小时导数为0，即

$$\begin{aligned}\frac{\partial E(n)}{\partial a_j}=&-2\sum_{n=0}^{N-1}v(n)\,v(n-j)\\&+2\sum_{k=1}^{P}a_k\sum_{n=0}^{N-1}v(n-k)\,v(n-j)=0\end{aligned} \tag{7}$$

因此根据式(6)和式(7)可得

$$r_{vv}(j)=\sum_{k=1}^{P}a_k r_{vv}(j-k) \tag{8}$$

利用Levinson-Durbin算法求解上式可以得到用于构建白化滤波器的系数a_k，$e(n)$是色噪声经过白化得到的结果。

利用白化滤波器对含噪语音处理时，不仅要形成对色噪声分量的白化，也要尽可能使得对语音分量的影响最小。语音信号的一个重要特点是具有可预测性，因为它的相邻取样之间存在一定的相关性，可以通过线性预测模型(Linear Prediction Coding, LPC)描述语音信号的特征和产生机理。在LPC中，语音信号可以通过激励一个时变的全极点滤波器产生，语音信号的特性由全极点滤波器获得，激励信号$\mu(n)$通常采用无语音分量的高斯白噪声[19]。此时，语音信号$s(n)$可以表示为

$$s(n)=\sum_{p=1}^{Q}a_p s(n-p)+\mu(n) \tag{9}$$

其中，Q是滤波器阶数，a_p是滤波器的系数。

考虑语音分量信号$s(n)$经过上述白化滤波器时，其输出信号$g(n)$可以表示为

$$\begin{aligned}g(n)=&s(n)-\tilde{s}(n)\\=&\sum_{p=1}^{Q}a_p s(n-p)+\mu(n)-\sum_{k=1}^{P}a_k s(n-k)\\=&\sum_{j=1}^{M}a_j s(n-j)+\mu(n)\end{aligned} \tag{10}$$

其中，$\tilde{s}(n)$为语音信号$s(n)$通过白化滤波器的输出信号，$M=\max(Q,P)$。从式(10)可以看到语音分量经过白化滤波器后得到的输出信号为高斯白噪声与语音信号过去值的线性组合，说明语音信号经过白化滤波器后，仍然符合LPC模型，具有语音信号的稀疏特性。

3.2 频谱重排

在利用RPCA对含噪语音的时频幅度矩阵进行低秩稀疏分解时，纯净语音并非是完全稀疏的。因为含噪语音的频谱中某些语音成分可用有限数量的频谱基来描述，这意味着它们具有低秩特征[10]。因此，可以认为语音信号由稀疏分量和低秩分量组成

$$\boldsymbol{S}=\boldsymbol{S}_{\mathrm{s}}+\boldsymbol{L}_{\mathrm{s}} \tag{11}$$

其中，$\boldsymbol{S}_{\mathrm{s}}\in R^{m\times n}$表示稀疏语音矩阵，$\boldsymbol{L}_{\mathrm{s}}\in R^{m\times n}$表示低秩语音矩阵。设$r_{\boldsymbol{L}_{\mathrm{s}}}\ll\min(m,n)$为矩阵$\boldsymbol{L}_{\mathrm{s}}$的秩，矩阵$\boldsymbol{L}_{\mathrm{s}}$中包含的频谱基为$\mathbf{ls}_1,\mathbf{ls}_2,\cdots,\mathbf{ls}_{r_{\boldsymbol{L}_{\mathrm{s}}}}\in R^{m\times 1}$，则$\boldsymbol{L}_{\mathrm{s}}$的每一列可以由$\mathbf{ls}_1,\mathbf{ls}_2,\cdots,\mathbf{ls}_{r_{\boldsymbol{L}_{\mathrm{s}}}}$的线性组合表示，即

$$\boldsymbol{L}_{\mathrm{s}}=\left[\sum_{i=1}^{r_{\boldsymbol{L}_{\mathrm{s}}}}\mathbf{ls}_i\alpha_{i,1}\ \sum_{i=1}^{r_{\boldsymbol{L}_{\mathrm{s}}}}\mathbf{ls}_i\alpha_{i,2}\cdots\sum_{i=1}^{r_{\boldsymbol{L}_{\mathrm{s}}}}\mathbf{ls}_i\alpha_{i,n}\right] \tag{12}$$

其中，$\alpha_{i,j}$是第j列中第i个频谱基的加权参数。

由于RPCA模型的固有特性，低秩语音分量将被当作噪声，即通过RPCA得到的低秩矩阵$\boldsymbol{L}$是噪声矩阵$\boldsymbol{L}_{\mathrm{n}}\in\mathrm{R}^{m\times n}$与低秩语音矩阵$\boldsymbol{L}_{\mathrm{s}}$的叠加

$$\boldsymbol{L}=\boldsymbol{L}_{\mathrm{n}}+\boldsymbol{L}_{\mathrm{s}} \tag{13}$$

因此，式(1)可以写为更精确的形式

$$|\boldsymbol{D}| = \boldsymbol{L}_{\mathrm{s}} + \boldsymbol{L}_{\mathrm{n}} + \boldsymbol{S}_{\mathrm{s}} \tag{14}$$

若将$\boldsymbol{L}_{\mathrm{s}}$和$\boldsymbol{S}_{\mathrm{s}}$一起分解到主要包含语音成分的稀疏矩阵$\boldsymbol{S}$中则可以在增强语音中保留较多的语音成分，从而提高算法性能。增加矩阵$\boldsymbol{L}_{\mathrm{s}}$的秩是获得良好分解结果的一种可行方案。因此，本文采用了一种称为频谱重排的技术来增加矩阵$\boldsymbol{L}_{\mathrm{s}}$的秩，频谱重排技术需要对含噪语音的时频幅度矩阵$|\boldsymbol{D}|$的每一列的谱元素进行重新排列，且各列之间的排列规则需要保证互不相同。

为了保证每一列的重排规则不同，且具有可控性和可复现的特点，本文采用哈希函数来为矩阵$|\boldsymbol{D}|$的每一列生成相应的重排规则。哈希函数定义为

$$f(k) = (a \cdot k + b)_m + 1 \tag{15}$$

其中，$a \in [2, m)$，$b \in [0, m)$均为控制参数，且a需要和m互质[20]，$(\cdot)_m$表示对m取模。哈希函数式(15)实现从k到$f(k)$的映射，并且映射关系随着a和b的变化而改变。以一个列向量$\boldsymbol{x} = [x_1\ x_2\ \cdots\ x_m]^{\mathrm{T}} \in R^{m\times 1}$为例说明元素重排的实现过程。将列向量$\boldsymbol{x}$中的各个元素的位置记为序列$(1, 2, \cdots, m)$，则此序列可以通过式(15)映射为新的序列$(f(1), f(2), \cdots, f(m))$，且根据数论中完全剩余系的理论，组成新序列的元素是完备的，即新序列中各个元素无重复、无丢失。然后利用序列$(1, 2, \cdots, m)$和$(f(1), f(2), \cdots, f(m))$生成一个第$k$行第$f(k)$列的元素为1而其余元素为0的重排矩阵$\boldsymbol{T} \in R^{m\times m}$。那么，列向量$\boldsymbol{x}$可以由重排矩阵$\boldsymbol{T}$转化为其重排向量$\boldsymbol{x}_{\mathrm{r}}$

$$\boldsymbol{x}_{\mathrm{r}} = T \cdot \boldsymbol{x} \tag{16}$$

其中，$\boldsymbol{x}_{\mathrm{r}} = [x_{f(1)}\ x_{f(2)}\ \cdots\ x_{f(m)}]^{\mathrm{T}}$。例如，取$m = 5$，$a = 2$，$b = 0$，序列$(1, 2, 3, 4, 5)$被哈希函数$f(k) = (2k + 0)_5 + 1$映射为序列$(3, 5, 2, 4, 1)$，假设重排矩阵$\boldsymbol{T}$为

$$\boldsymbol{T} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{17}$$

因此向量$\boldsymbol{x} = [x_1\ x_2\ x_3\ x_4\ x_5]^{\mathrm{T}}$重排后的结果为$\boldsymbol{x}_r = [x_3\ x_5\ x_2\ x_4\ x_1]^{\mathrm{T}}$。

由于频谱重排需要对矩阵$|\boldsymbol{D}|$的每一列采用不同的重排规则，因此每一列对应的重排矩阵必须互不相同，故需要产生n个重排矩阵$\boldsymbol{T}_j (j = 1, 2, \cdots, n)$，这一需求可以通过改变哈希函数的控制参数$a$和$b$来实现。那么，对矩阵$|\boldsymbol{D}|$的频谱重排可以表示为

$$|\boldsymbol{D}|_{\mathrm{r}}(:, j) = \boldsymbol{T}_j \cdot |\boldsymbol{D}|(:, j), \quad j = 1, 2, \cdots, n \tag{18}$$

其中，$|\boldsymbol{D}|_{\mathrm{r}}$表示$|\boldsymbol{D}|$被重排后的矩阵，$\boldsymbol{T}_j$为第$j$列对应的重排矩阵。与式(14)类似，$|\boldsymbol{D}|_{\mathrm{r}}$也可以表示为3个矩阵的总和

$$|\boldsymbol{D}|_{\mathrm{r}} = \boldsymbol{H}_{\mathrm{l}} + \boldsymbol{V} + \boldsymbol{H}_{\mathrm{s}} \tag{19}$$

其中，$\boldsymbol{H}_{\mathrm{l}}$, $\boldsymbol{V}$和$\boldsymbol{H}_{\mathrm{s}} \in R^{m\times n}$分别是$\boldsymbol{L}_{\mathrm{s}}$, $\boldsymbol{L}_{\mathrm{n}}$和$\boldsymbol{S}_{\mathrm{s}}$经过频谱重排后得到的矩阵。

矩阵$\boldsymbol{H}_{\mathrm{l}}$通过重新排列$\boldsymbol{L}_{\mathrm{s}}$得到，可以表示为

$$\boldsymbol{H}_{\mathrm{l}} = \left[\sum_{i=1}^{r_{\boldsymbol{L}_{\mathrm{s}}}} \boldsymbol{T}_1 \mathbf{ls}_i \alpha_{i,1}\ \sum_{i=1}^{r_{\boldsymbol{L}_{\mathrm{s}}}} \boldsymbol{T}_2 \mathbf{ls}_i \alpha_{i,2} \cdots \sum_{i=1}^{r_{\boldsymbol{L}_{\mathrm{s}}}} \boldsymbol{T}_n \mathbf{ls}_i \alpha_{i,n}\right] \tag{20}$$

显然，矩阵$\boldsymbol{H}_{\mathrm{l}}$的第$j$列是一组新的频谱基向量$\boldsymbol{T}_j\mathbf{ls}_1, \boldsymbol{T}_j\boldsymbol{ls}_2, \cdots, \boldsymbol{T}_j\mathbf{ls}_{r_{\boldsymbol{L}_{\mathrm{s}}}}$的线性组合。比较两组新的频谱基$\boldsymbol{T}_1\mathbf{ls}_1, \boldsymbol{T}_1\mathbf{ls}_2, \cdots, \boldsymbol{T}_1\mathbf{ls}_{r_{\boldsymbol{L}_{\mathrm{s}}}}$和$\boldsymbol{T}_2\mathbf{ls}_1, \boldsymbol{T}_2\mathbf{ls}_2, \cdots, \boldsymbol{T}_2\mathbf{ls}_{r_{\boldsymbol{L}_{\mathrm{s}}}}$，由于$\boldsymbol{T}_1$和$\boldsymbol{T}_2$不相同，因此$\boldsymbol{T}_1\mathbf{ls}_1, \boldsymbol{T}_1\mathbf{ls}_2, \cdots, \boldsymbol{T}_1\mathbf{ls}_{r_{\boldsymbol{L}_{\mathrm{s}}}}$和$\boldsymbol{T}_2\mathbf{ls}_1, \boldsymbol{T}_2\mathbf{ls}_2, \cdots, \boldsymbol{T}_2\mathbf{ls}_{r_{\boldsymbol{L}_{\mathrm{s}}}}$这两组频谱基不相关。所以，$\boldsymbol{T}_1\mathbf{ls}_1, \boldsymbol{T}_1\mathbf{ls}_2, \cdots, \boldsymbol{T}_1\mathbf{ls}_{r_{\boldsymbol{L}_{\mathrm{s}}}}$与$\boldsymbol{T}_2\mathbf{ls}_1, \boldsymbol{T}_2\mathbf{ls}_2, \cdots, \boldsymbol{T}_2\mathbf{ls}_{r_{\boldsymbol{L}_{\mathrm{s}}}}$是线性独立的，即$\boldsymbol{H}_{\mathrm{l}}$的第1列向量和第2列向量是线性独立的。类似地，$\boldsymbol{H}_{\mathrm{l}}$的所有列向量几乎都是线性无关的，因此$\boldsymbol{H}_{\mathrm{l}}$近似于满秩。

具有低秩特性的语音矩阵$\boldsymbol{L}_{\mathrm{s}}$可以通过频谱重排转换为接近满秩的矩阵$\boldsymbol{H}_{\mathrm{l}}$，但对于低秩噪声矩阵$\boldsymbol{L}_{\mathrm{n}}$则有不同的结果。$\boldsymbol{L}_{\mathrm{n}}$可以用列向量表示为

$$\boldsymbol{L}_{\mathrm{n}} = [\boldsymbol{l}_1\ \boldsymbol{l}_2\ \cdots\ \boldsymbol{l}_n] \tag{21}$$

其中，$\boldsymbol{l}_j$是$\boldsymbol{L}_{\mathrm{n}}$的第$j$个列向量。由于色噪声已经过白化处理且白化噪声的频谱是平坦的，因此$\boldsymbol{l}_j = [l_{1,j}\ l_{2,j}\ \cdots\ l_{m,j}]^{\mathrm{T}}$中的元素具有如式(22)的关系

$$l_{1,j} \approx l_{2,j} \approx \cdots \approx l_{m,j} \tag{22}$$

使用$\boldsymbol{T}_j\ (j = 1, 2, \cdots, n)$对$\boldsymbol{L}_{\mathrm{n}}$进行频谱重排得矩阵$\boldsymbol{V}$

$$\boldsymbol{V} = [\boldsymbol{T}_1 \cdot \boldsymbol{l}_1\ \ \boldsymbol{T}_2 \cdot \boldsymbol{l}_2\ \ \cdots\ \ \boldsymbol{T}_n \cdot \boldsymbol{l}_n] \tag{23}$$

$\boldsymbol{V}$的第j列向量为

$$\begin{aligned} \boldsymbol{T}_j \cdot \boldsymbol{l}_j &= \boldsymbol{T}_j \cdot [l_{1,j}\ l_{2,j} \cdots l_{m,j}]^{\mathrm{T}} \\ &= \left[l_{f_j(1),j}\ l_{f_j(2),j} \cdots l_{f_j(m),j}\right]^{\mathrm{T}} \end{aligned} \tag{24}$$

其中，f_j是$\boldsymbol{T}_j$对应的哈希函数。由于频谱重排只改变元素的位置而不改变其大小，所以有$l_{f_j(1),j} \approx l_{f_j(2),j} \approx \cdots \approx l_{f_j(m),j} \approx l_{1,j} \approx l_{2,j} \approx \cdots \approx l_{m,j}$，则$\boldsymbol{T}_j \cdot \boldsymbol{l}_j \approx \boldsymbol{l}_j$。这意味着$\boldsymbol{L}_{\mathrm{n}}$经过频谱重排，其列向量几乎不会发生变化，因此有

$$\boldsymbol{V} \approx \boldsymbol{L}_{\mathrm{n}} \tag{25}$$

显然，矩阵$\boldsymbol{V}$仍然是低秩的。

对于稀疏且高秩的语音矩阵$\boldsymbol{S}_{\mathrm{s}}$，其重排后的矩阵$\boldsymbol{H}_{\mathrm{s}}$仍然保持稀疏和高秩的特征。

综上所述，频谱重排可以提高低秩语音$\boldsymbol{L}_{\mathrm{s}}$的

秩，而又不改变稀疏语音矩阵$\boldsymbol{S}_{\mathrm{s}}$的稀疏和高秩特征以及噪声矩阵$\boldsymbol{L}_{\mathrm{n}}$的秩。

3.3 WSRRPCA算法的结构

WSRRPCA算法的结构是在AMS框架的基础上实现的。其算法结构图如图2所示。

在分析阶段，首先使用含噪语音开始时的非语音段数据来构建白化滤波器，通过该滤波器将输入的含噪语音$\boldsymbol{d}(n)$白化为$\boldsymbol{d}_{\mathrm{w}}(n)$。然后，使用STFT将$\boldsymbol{d}_{\mathrm{w}}(n)$转换为其时频图$\boldsymbol{D}_{\mathrm{w}}$，并得到时频幅度谱$|\boldsymbol{D}_{\mathrm{w}}|$和相位谱$\angle\boldsymbol{D}_{\mathrm{w}}$，其中$|\boldsymbol{D}_{\mathrm{w}}|$用于去噪阶段，$\angle\boldsymbol{D}_{\mathrm{w}}$将直接用作合成阶段中增强语音的相位信息。

在去噪阶段，对时频幅度谱$|\boldsymbol{D}_{\mathrm{w}}|$进行频谱重排得到$|\boldsymbol{D}_{\mathrm{w}}|_{\mathrm{r}}$，然后利用RPCA对矩阵$|\boldsymbol{D}_{\mathrm{w}}|_{\mathrm{r}}$进行低秩稀疏分解以获得低秩矩阵$\boldsymbol{L}$和稀疏矩阵$\boldsymbol{S}$。

在合成阶段，首先，将对稀疏矩阵$\boldsymbol{S}$执行逆重排获得$\boldsymbol{S}'$。然后，使用幅度谱$\boldsymbol{S}'$和含噪语音的相位谱图$\angle\boldsymbol{D}_{\mathrm{w}}$合成增强语音的时频图$\boldsymbol{X}_{\mathrm{w}}$

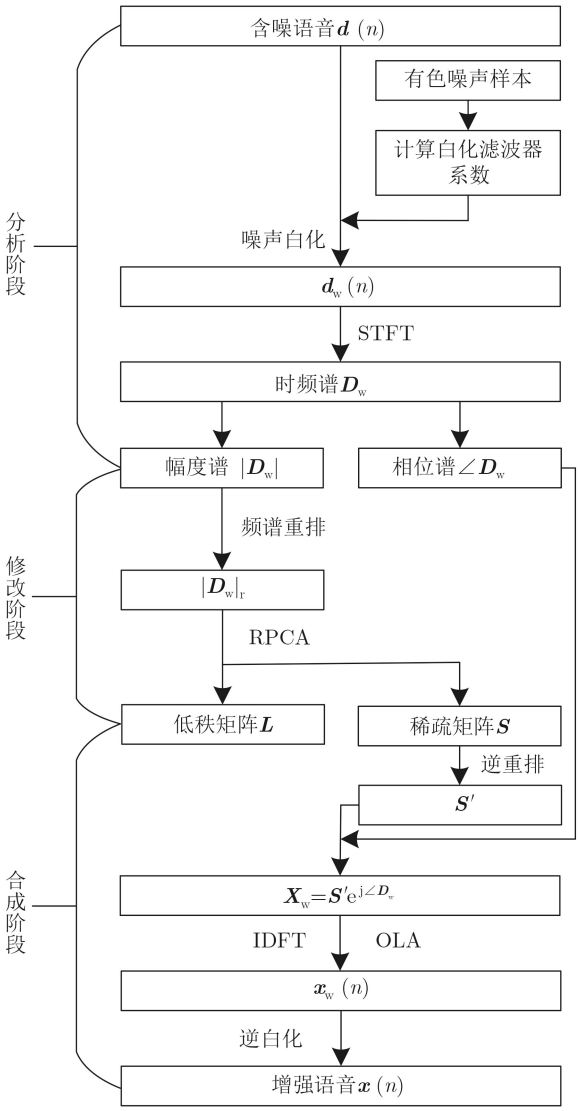

图 2 WSRRPCA算法的结构框图

$$X_{\mathrm{w}}=\boldsymbol{S}'\mathrm{e}^{\mathrm{j}\angle\boldsymbol{D}_{\mathrm{w}}} \tag{26}$$

传统的语音增强算法认为相位差较小，且不易被听觉系统感知到，因此重构信号时可以采用带噪信号的相位[21]。然后，对$\boldsymbol{X}_{\mathrm{w}}$的每一列做IDFT，并使用OLA将多帧语音合成时域信号$\boldsymbol{x}_{\mathrm{w}}(n)$。最后对$\boldsymbol{x}_{\mathrm{w}}(n)$逆白化得到最终的增强语音信号$\boldsymbol{x}(n)$。

3.4 WSRRPCA算法的参数

在WSRRPCA算法中参数λ会影响稀疏矩阵$\boldsymbol{S}$的稀疏度。λ越大，矩阵$\boldsymbol{S}$的稀疏度越大，矩阵$\boldsymbol{S}$中残留的噪声更少，但使用RPCA进行分离后，语音成分会丢失更多，导致更高的语音失真。相反$\boldsymbol{S}$的稀疏度小，则语音信号的失真会更少，但会包含更多的残留噪声[8]。因此，选择合适的λ值可以使语音分量和噪声分量之间获得理想的平衡。

为了确定参数λ的取值，完成了相应的仿真实验。在实验中，选择背景噪声为白噪声，信噪比分别为–5 dB, 0 dB, 5 dB和10 dB的含噪语音。然后将λ以0.001的间隔从0.001增大到0.15，并使用每一个λ的取值作为算法参数，利用本文提出的WSRRPCA算法对含噪语音进行增强，最后对每一次得到的增强语音计算其源失真比(Source-to-Distortion Ratio, SDR)，得到λ与相应的SDR的关系，结果如图3所示。

SDR是衡量增强语音中语音能量和噪声能量比值的指标，表示的是算法的噪声抑制能力，单位是dB[22]。SDR越大表示算法的噪声抑制性能越好。从图3可以看出，当λ分别约等于0.144, 0.114, 0.094和0.076时，4个SDR曲线达到其峰值。在本文中，将使SDR得分位于曲线最高点的λ确定为算法的最优参数，并将其表示为λ_{opt}。显然，在实际应用中搜索λ_{opt}是不合理的。本文将λ_{opt}的估计值$\hat{\lambda}_{\mathrm{opt}}$用作WSRRPCA算法的参数。从图3可以看出，$\lambda_{\mathrm{opt}}$随含噪语音的信噪比变化而变化，因此，可以根据信噪比来设置$\hat{\lambda}_{\mathrm{opt}}$。尽管含噪语音的信噪比在实践中仍然是未知的，但其粗略估计eSNR可以通过式(27)计算

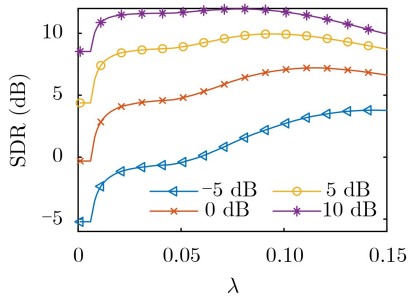

图 3 不同信噪比下SDR与λ的关系曲线

$$\mathrm{eSNR}=10\lg\left(\frac{\sum_{j=1}^{n}\sum_{i=1}^{m}|\boldsymbol{D}|^2(i,j)}{\frac{n}{8}\sum_{j=1}^{8}\sum_{i=1}^{m}|\boldsymbol{D}|^2(i,j)}-1\right) \tag{27}$$

通过$|\boldsymbol{D}|$的前8个非语音帧来估计噪声，然后，使用对应的eSNR和λ_{opt}进行数据拟合，得到$\hat{\lambda}_{\mathrm{opt}}$和eSNR之间的关系

$$\hat{\lambda}_{\mathrm{opt}}=\mathrm{c}\cdot\mathrm{eSNR}+\mathrm{d} \tag{28}$$

其中c和d是常数。本文进行了足够的数值实验，确定了合理的c和d的值，即$\mathrm{c}=-0.004$, $\mathrm{d}=0.1181$。显然，当噪声为色噪声时，使用相同的c和d来估算$\hat{\lambda}_{\mathrm{opt}}$是合理的，因为在估算最优参数$\hat{\lambda}_{\mathrm{opt}}$之前已使用了噪声白化。

4 实验结果及分析

4.1 实验设置

用于仿真的30个纯净语音来自NOIZEUS数据库，色噪声则来自NOISEX-92数据库，分别设置这些纯净语音在–5 dB, 0 dB, 5 dB和10 dB这4种信噪比下被色噪声污染，从而得到含噪语音信号。此外，语音和噪声都以8 kHz进行重新采样。在噪声白化过程中，含噪语音的前1024个点用于构建白化滤波器。在使用STFT时，通过汉明窗将含噪语音分帧，长度为256个样本，移位为128点，然后对每个帧应用256点快速傅里叶变换。ALM算法用于进行矩阵低秩稀疏分解。为了衡量算法的性能，使用SDR以及语音质量的感知评估(Perceptual Evaluation of Speech Quality, PESQ)作为评估指标。PESQ是衡量增强语音质量的指标[23]，分数越高表示语音质量越好。

4.2 WSRRPCA与RPCA算法分解结果对比

图4显示了当语音被f16噪声污染的情况下，RPCA和WSRRPCA算法对含噪语音的直观分解结果对比。在图4中，图4(a)和图4(d)分别为纯净语音和f16噪声的时频图；图4(b)和图4(e)为由RPCA分解含噪语音时频图而得到的语音分量和噪声分量的时频图；图4(c)和图4(f)为由WSRRPCA分解含噪语音时频图而得到的语音分量和噪声分量的时频图。对比图4(d)和图4(e)可知，采用RPCA处理后，许多低秩语音成分被分解到噪声矩阵中。而对比图4(e)和4(f)可以观察到，采用WSRRPCA获得的噪声矩阵中，低秩语音残留较少。同样，对比图4(b)和图4(c)可知，WSRRPCA的处理结果在RPCA的语音分量结果的基础上保留了更多的低秩语音分量，这将使得WSRRPCA能够获得更好的语音质量。

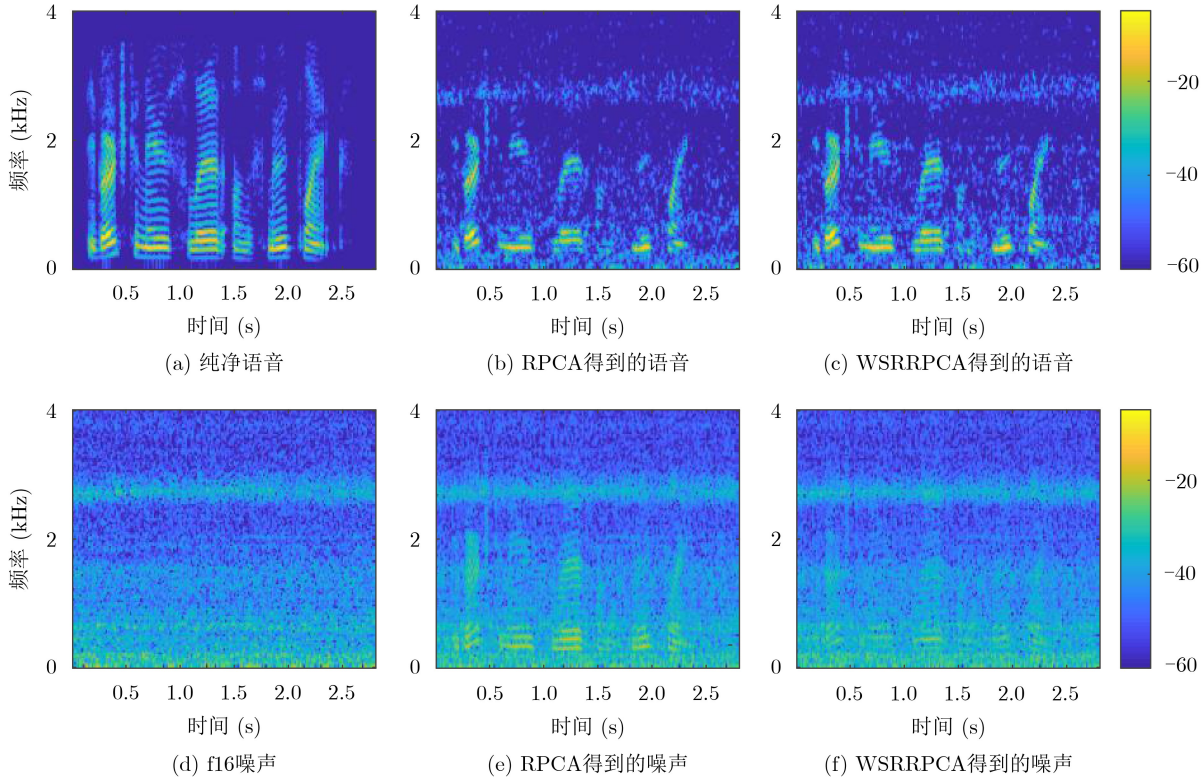

图 4 WSRRPCA与RPCA处理结果对比图

4.3 WSRRPCA与其他语音增强算法的性能对比

为了验证提出的WSRRPCA算法在色噪声环境下具有较好的语音增强效果，本文利用数值仿真实验，将该算法与基于RPCA的单通道语音增强算法[6]、基于约束低秩稀疏矩阵分解(Constrained Low-rank and Sparse Matrix Decomposition, CLSMD)的单通道语音增强算法[24]、几何谱减(Geometric Approach to Spectral Subtraction, GASS)算法[25]、结合信号存在不确定性的对数最小均方误差(Minimum Mean Square Error of the Log-spectra under Signal Presence Uncertainty, LogMMSE-SPU)算法[26]等进行了性能对比分析。

首先，测试语音sp01被buccaneer1，buccaneer2，f16，factory1，hfchannel和pink噪声破坏，其中信噪比为0 dB，数值结果如表1所示。实验结果表明，当语音信号被不同色噪声破坏后，在5种色噪声类型中WSRRPCA获得了最高的SDR分数。就PESQ指标而言，当噪声为buccaneer1，buccaneer2，factory1和hfchannel时，WSRRPCA的得分略低于RPCA的得分。

大多数的语音增强方法在减少背景噪声的同时也会引入语音失真，因此，语音增强的目的是不明显引入信号失真的前提下，对其中的噪声进行有效抑制。综合SDR和PESQ的结果来考虑，在保证增强语音质量不发生明显降低的同时，本文提出的WSRRPCA算法在多种色噪声环境中具有更好的噪声抑制性能。

为了进一步说明所提算法具有更优的性能，本文用了30组语音信号，语音sp01-sp30分别受到6种类型的色噪声的破坏，且信噪比分别为–5 dB, 0 dB, 5 dB和10 dB。对于每种情况，例如sp01被buccaneer1噪声污染，信噪比为–5 dB，进行了100次的蒙特卡罗实验。对于每种色噪声，将30个语音的SDR和PESQ分数的平均值作为最终结果，直观的条形图对比结果如图5和图6所示。

从SDR的结果可以看出，在多种色噪声的低信噪比环境中(–5 dB, 0 dB和5 dB)，本文提出的算法比其他算法具有更高的得分。尤其是当信噪比为–5 dB时，相比于其余算法，提出的算法可以实现优异的噪声抑制效果。而在信噪比为10 dB的高信噪比情况下，RPCA算法、CLSMD算法和本文提出的算法获得了几乎相同的SDR分数。对于PESQ，当信噪比为–5 dB时，所提出算法的得分仅略低于GASS算法和RPCA算法的得分。在其他信噪比时，所提出算法的PESQ得分与GASS和RPCA的结果基本相同，甚至稍高一些，综合评估SDR和PESQ这两种指标，当背景噪声为色噪声时，本文提出的算法在尽可能不降低语音质量的同时，可以有效地抑制背景噪声，比其他4种方法更具优势。

表 1　不同噪声下多种算法的性能对比

噪声类型	语音增强算法	SDR (dB)	PESQ
buccaneer1	GASS	1.1148	1.5057
	logMMSE-SPU	–2.9566	0.9222
	RPCA	4.8432	**1.6275**
	CLSMD	5.3583	1.0624
	WSRRPCA	**6.2530**	1.6106
buccaneer2	GASS	–0.4980	1.5690
	logMMSE-SPU	–3.0210	1.1192
	RPCA	3.8481	**1.7261**
	CLSMD	4.6147	0.9079
	WSRRPCA	**4.9989**	1.6944
f16	GASS	1.4805	1.7816
	logMMSE-SPU	–2.3210	1.1926
	RPCA	4.3886	1.8461
	CLSMD	5.4681	1.1948
	WSRRPCA	**6.2030**	**1.8751**
factory1	GASS	0.3133	1.4930
	logMMSE-SPU	–2.7692	1.1512
	RPCA	4.0886	**1.8264**
	CLSMD	4.2691	1.2895
	WSRRPCA	**5.1138**	1.7905
hfchannel	GASS	1.4168	1.3519
	logMMSE-SPU	–3.0150	1.0336
	RPCA	5.1769	1.6378
	CLSMD	**6.7771**	1.1689
	WSRRPCA	6.1418	**1.6441**
pink	GASS	1.0008	1.6570
	logMMSE-SPU	–1.4077	1.2425
	RPCA	4.0835	1.8472
	CLSMD	3.9805	1.4122
	WSRRPCA	**7.0699**	**1.9045**

5　结束语

在基于RPCA的单通道语音增强算法的局限性的驱使下，本文提出一种基于白化频谱重排RPCA的语音增强算法，该算法通过合理优化噪声白化和色化模型改善原有算法只能适用高斯白噪声场合的能力，并利用频谱重排改善信号的时频低秩特性，从而提升噪声分量和语音分量分离性能，最终实现色噪声环境下的语音增强性能的提升。多种色噪声环境中的数值实验结果表明，与多种算法相比，本文提出的算法表现出更优的噪声抑制性能和更好的语音可懂度。

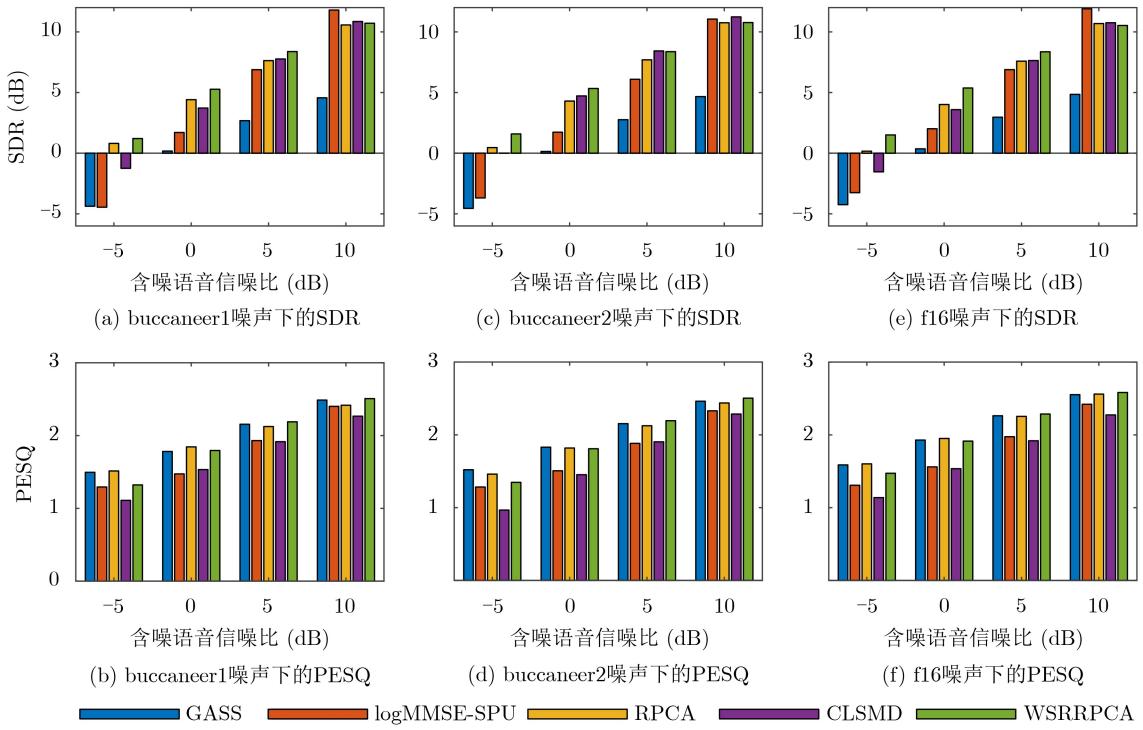

图 5 buccaneer1, buccaneer2和f16噪声环境下不同算法的性能对比图

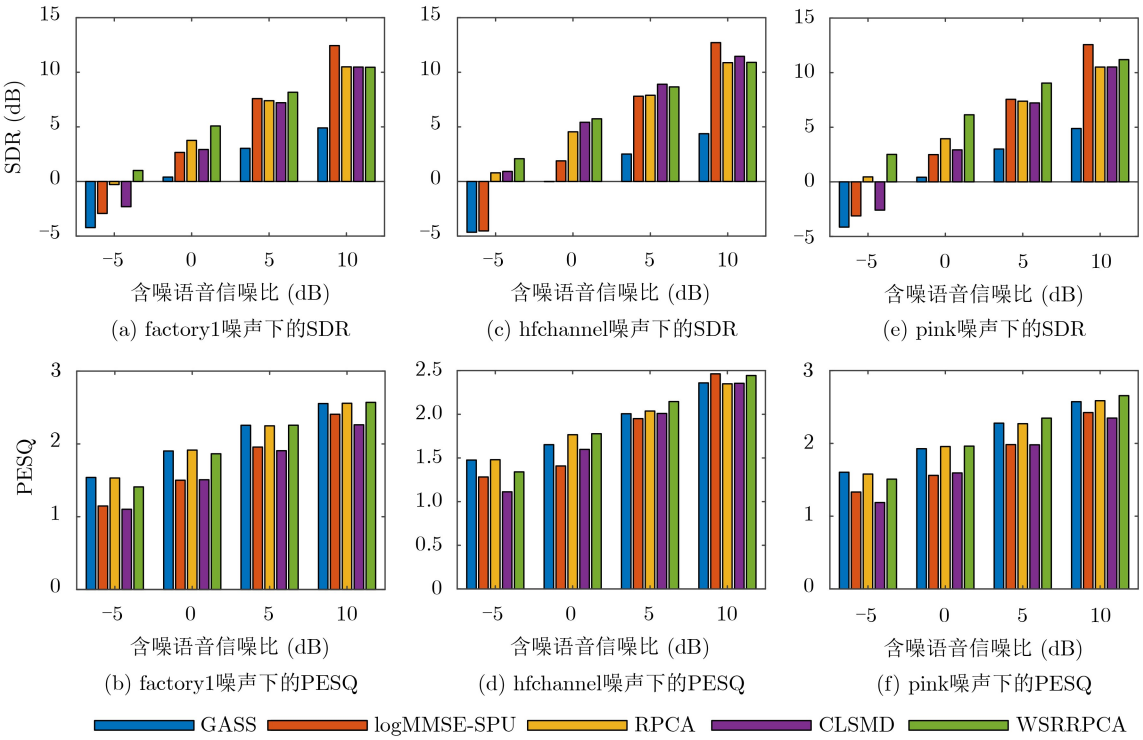

图 6 Factory1, hfchannel和pink噪声环境下不同算法的性能对比图

参 考 文 献

[1] LOIZOU P C. Speech Enhancement: Theory and Practice[M]. 2nd ed. London: CRC Press, 2013: 1–2.

[2] BOLL S. Suppression of acoustic noise in speech using spectral subtraction[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1979, 27(2): 113–120. doi: 10.1109/tassp.1979.1163209.

[3] EPHRAIM Y and MALAH D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1985, 33(2): 443–445. doi: 10.1109/TASSP.1985.1164550.

[4] SCALART P and FILHO J V. Speech enhancement based on a priori signal to noise estimation[C]. International

Conference on Acoustics, Speech, and Signal Processing, Atlanta, 1996: 629–632. doi: 10.1109/ICASSP.1996.543199.

[5] EPHRAIM Y and VAN TREES G L. A signal subspace approach for speech enhancement[J]. *IEEE Transactions on Speech and Audio Processing*, 1995, 3(4): 251–266. doi: 10.1109/89.397090.

[6] YI Hu and LOIZOU P C. A generalized subspace approach for enhancing speech corrupted by colored noise[J]. *IEEE Transactions on Speech and Audio Processing*, 2003, 11(4): 334–341. doi: 10.1109/TSA.2003.814458.

[7] SUN Chengli, ZHANG Qin, WANG Jian, *et al*. Noise reduction based on robust principal component analysis[J]. *Journal of Computational Information Systems*, 2014, 10(10): 4403–4410. doi: 10.12733/jcis10408.

[8] HUANG Jianjun, ZHANG Xiongwei, ZHANG Yafei, *et al*. Speech denoising via low-rank and sparse matrix decomposition[J]. *ETRI Journal*, 2014, 36(1): 167–170. doi: 10.4218/etrij.14.0213.0033.

[9] MAVADDATY S, AHADI S M, and SEYEDIN S. A novel speech enhancement method by learnable sparse and low-rank decomposition and domain adaptation[J]. *Speech Communication*, 2016, 76: 42–60. doi: 10.1016/j.specom.2015.11.003.

[10] SUN Pengfei and QIN Jun. Low-rank and sparsity analysis applied to speech enhancement via online estimated dictionary[J]. *IEEE Signal Processing Letters*, 2016, 23(12): 1862–1866. doi: 10.1109/lsp.2016.2627029.

[11] LUO Yongjiang and MAO Yu. Single-channel speech enhancement based on multi-band spectrogram-rearranged RPCA[J]. *Electronics Letters*, 2019, 55(7): 415–417. doi: 10.1049/el.2018.8131.

[12] CANDÈS E J, LI Xiaodong, MA Yi, *et al*. Robust principal component analysis?[J]. *Journal of the ACM*, 2011, 58(3): 11. doi: 10.1145/1970392.1970395.

[13] NAZIH M, MINAOUI K, and COMON P. Using the proximal gradient and the accelerated proximal gradient as a canonical polyadic tensor decomposition algorithms in difficult situations[J]. *Signal Processing*, 2020, 171: 107472. doi: 10.1016/j.sigpro.2020.107472.

[14] FENG Peihua, LING B W K, LEI Ruisheng, *et al*. Singular spectral analysis-based denoising without computing singular values via augmented Lagrange multiplier algorithm[J]. *IET Signal Processing*, 2019, 13(2): 149–156. doi: 10.1049/iet-spr.2018.5086.

[15] LEI Yunwen and ZHOU Dingxuan. Analysis of singular value thresholding algorithm for matrix completion[J]. *Journal of Fourier Analysis and Applications*, 2019, 25(6): 2957–2972. doi: 10.1007/s00041-019-09688-8.

[16] JARAMILLO A E, NIELSEN J K, CHRISTENSEN M G, *et al*. A study on how pre-whitening influences fundamental frequency estimation[C]. International Conference on Acoustics, Speech and Signal Processing, Brighton, England, 2019: 6495–6499. doi: 10.1109/ICASSP.2019.8683653.

[17] VASEGHI S V. Advanced Digital Signal Processing and Noise Reduction[M]. 4th ed. Hoboken: John Wiley & Sons, 2008: 229–230.

[18] SMITH III J O. Spectral Audio Signal Processing[M]. W3K Publishing, USA, 2011: 298–301.

[19] 张明, 刘祥楼, 姜峥嵘. 基于LPC的语音信号预测仿真分析[J]. 光学仪器, 2015, 37(1): 71–74. doi: 10.3969/j.issn.1005-5630.2015.01.015.

ZHANG Ming, LIU Xianglou, and JIANG Zhengrong. Simulation analysis of speech signal prediction based on LPC[J]. *Optical Instruments*, 2015, 37(1): 71–74. doi: 10.3969/j.issn.1005-5630.2015.01.015.

[20] KAPRALOV M. Sparse Fourier transform in any constant dimension with nearly-optimal sample complexity in sublinear time[C]. The Forty-eighth Annual ACM Symposium on Theory of Computing, Virtual Event, Italy, 2016: 264–277. doi: 10.1145/2897518.2897650.

[21] WANG D L and LIM J S. The unimportance of phase in speech enhancement[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1982, 30(4): 679–681. doi: 10.1109/TASSP.1982.1163920.

[22] VINCENT E, GRIBONVAL R, and FEVOTTE C. Performance measurement in blind audio source separation[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006, 14(4): 1462–1469. doi: 10.1109/TSA.2005.858005.

[23] RAM R and MOHANTY M N. Use of radial basis function network with discrete wavelet transform for speech enhancement[J]. *International Journal of Computational Vision and Robotics*, 2019, 9(2): 207–223. doi: 10.1504/IJCVR.2019.10019996.

[24] SUN Chengli, ZHU Qi, and WAN Minghua. A novel speech enhancement method based on constrained low-rank and sparse matrix decomposition[J]. *Speech Communication*, 2014, 60: 44–55. doi: 10.1016/j.specom.2014.03.002.

[25] LU Yang and LOIZOU P C. A geometric approach to spectral subtraction[J]. *Speech Communication*, 2008, 50(6): 453–466. doi: 10.1016/j.specom.2008.01.003.

[26] COHEN I. Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator[J]. *IEEE Signal Processing Letters*, 2002, 9(4): 113–116. doi: 10.1109/97.1001645.

罗勇江：男，1979年生，副教授，研究方向为低秩稀疏分解、高速信号处理.

杨腾飞：男，1993年生，硕士，研究方向为语音增强、语音信号处理.

赵　冬：男，1996年生，硕士生，研究方向为非高斯信号处理.

责任编辑：马秀强