

文章编号:1004-9037(2014)02-0259-06

稀疏卷积非负矩阵分解的语音增强算法

张立伟<sup>1</sup> 贾 冲<sup>1</sup> 张雄伟<sup>1</sup> 闵 刚<sup>1,2</sup> 曾 理<sup>1</sup>

(1. 解放军理工大学指挥信息系统学院,南京,210016;2. 西安通信学院基础部,西安,710106)

**摘要:**鉴于卷积非负矩阵分解在语音增强算法中的成功应用,进一步考虑语音信号在时频域的稀疏性,提出了稀疏卷积非负矩阵分解(Sparse convolutive nonnegative matrix factorization, SCNMF)的语音增强算法。该算法包括训练和增强两个阶段。训练阶段通过 SCNMF 算法分别对纯净语音和噪声的频谱进行训练,得到纯净语音和噪声字典,并将其作为增强阶段的先验信息。增强阶段首先通过 SCNMF 算法对带噪语音的频谱进行分解,然后利用纯净语音和噪声联合字典以及相应的迭代公式对语音编码矩阵进行估计,重构增强语音。通过实验仿真分析了稀疏因子对增强语音质量的影响。实验结果表明,在非平稳噪声和低信噪比条件下,本文算法增强效果均优于多带谱减、非负矩阵分解和卷积非负矩阵分解等传统的算法。

**关键词:**语音增强;稀疏卷积;非负矩阵;字典训练;稀疏因子

**中图分类号:**TN912.3      **文献标志码:**A

Speech Enhancement Based on Convolutive Nonnegative Matrix Factorization with Sparseness Constraints

Zhang Liwei<sup>1</sup>, Jia Chong<sup>1</sup>, Zhang Xiongwei<sup>1</sup>, Min Gang<sup>1,2</sup>, Zeng Li<sup>1</sup>

(1. Institute of Command Information System, PLA University of Science & Technology, Nanjing, 210007, China;  
2. Department of Basic, Xi'an Communications Institute, Xi'an, 710106, China)

**Abstract:** In recent years, sparse convolutive nonnegative matrix factorization(SCNMF) algorithm has been well used for speech enhancement. Considering the sparsity of speech signals in the frequency domain, a speech enhancement approach based on SCNMF is proposed. The approach consists of a training stage and a denoising stage. During the training stage, the prior information about the spectrum of speech and noise is modeled by SCNMF algorithm and the dictionaries of speech and noise are constructed. In the denoising stage, the spectrum of noisy speech is analyzed by SCNMF algorithm, and the dictionaries of speech and noise and the iterative formulation are combined to evaluate the coding matrix of speech, then the enhanced speech is reconstructed. The impact of sparse factor on enhanced speech quality is analyzed through simulation experiments. Experimental results show that the proposed method outperforms traditional speech enhancement algorithms, such as multi-band spectral subtraction (MSS), nonnegative matrix factorization(NMF), convolutive nonnegative matrix factorization (CNMF), in non-stationary noise and low SNR conditions.

**Key words:** speech enhancement; sparse convolutive; nonnegative matrix; dictionary training; sparse factor

引 言

语音增强是语音信号处理中的重要领域,其主

要目的是从带噪语音信号中尽可能提取纯净的原始语音或原始语音参数,在语音编码、语音识别和听觉辅助等领域具有重要的应用价值。20 世纪 70 年代以后,随着数字信号处理理论的发展成熟,人

们开始采用信号处理的方法研究和实现语音增强问题。最具代表性的算法是 Boll 提出的谱减法 (Spectral subtraction, SS)<sup>[1]</sup>, 其思想简单直观、易于实现, 但该方法假设在有声段噪声是固定的, 且依赖话音激活检测技术 (Voice activity detection, VAD), 在非平稳噪声环境和低信噪比条件下, 效果并不理想。多带谱减法 (Multi-band spectral subtraction, MSS) 是将噪声频谱划分为互不重叠的频带, 分别在不同频带采用不同的谱减参数进行增强处理<sup>[2]</sup>。该算法效果要明显好于传统谱减法, 但仍需要 VAD 过程。基于非负矩阵分解 (Non-negative matrix factorization, NMF)<sup>[3]</sup> 的语音增强算法是通过训练构造语音和噪声的字典并将其组合成一个联合字典, 利用非负矩阵分解更新带噪声语音在联合字典下的投影系数, 实现语音和噪声的分离。然而, 非负矩阵分解算法只是在频域进行字典学习, 学习得到的字典中的原子通常只是一维信号, 缺乏有效表征原始信号的结构特征, 也没有考虑帧与帧之间的相关性。针对上述不足, Smaragdis 提出一种新的矩阵分解算法——卷积非负矩阵分解 (Convolutional nonnegative matrix factorization, CNMF)<sup>[4]</sup>。利用 CNMF 算法分解语音信号, 得到的时频字典能够更好地保留语音信号中的个人特征信息及帧间相关性。孙健等率先将该算法运用到语音转换中, 提出一种基于卷积非负矩阵分解的语音转换方法<sup>[5]</sup>, 获得了较好的转换效果。随后, 黄建军等人将该算法运用到语音增强中, 提出了一种时频学习单通道语音增强算法<sup>[6]</sup>。虽然该算法取得了较好的增强效果, 但是对语音信号中的一些时频特征描述仍不够完善。本文在卷积非负矩阵分解的基础上, 考虑到语音信号在时频域的稀疏性, 提出了基于稀疏卷积非负矩阵分解 (Sparse convolutional nonnegative matrix factorization, SCNMF) 的语音增强算法。

## 1 稀疏卷积非负矩阵分解

### 1.1 卷积非负矩阵分解

卷积非负矩阵分解是非负矩阵分解的二维扩展, 可以用如下数学模型来表示

$$\mathbf{V} \approx \sum_{t=0}^{T-1} \mathbf{W}(t) \overset{t \rightarrow}{\mathbf{H}} \quad (1)$$

式中:  $\mathbf{V} \in \mathbf{R}^{\geq 0, L \times N}$  为待分解的矩阵;  $\mathbf{W}(t) \in \mathbf{R}^{\geq 0, L \times R}$  和  $\mathbf{H} \in \mathbf{R}^{\geq 0, R \times N}$  分别为时频原子和相应的时变增益系数;  $(\cdot) \overset{t \rightarrow}{\phantom{}}$  运算符表示向右移动矩阵  $i$  列并将最左边的  $i$  列补零。CNMF 的数学模型可以理解为由

非负矩阵集  $\mathbf{W}$  和非负矩阵  $\mathbf{H}$  的卷积来表示待分解非负矩阵  $\mathbf{V}$ , 故称为卷积非负矩阵分解。CNMF 的目标就是寻找一系列的非负矩阵  $\mathbf{W}(t)$  和  $\mathbf{H}$ , 使其卷积的结果尽可能地逼近矩阵  $\mathbf{V}$ 。使用 KL 散度作为目标函数即

$$D(\mathbf{V} | \hat{\mathbf{V}}) = \sum_{ik} \left( V_{ik} \log \left( \frac{V_{ik}}{\hat{V}_{ik}} \right) + V_{ik} - \hat{V}_{ik} \right) \quad (2)$$

式中:  $\hat{\mathbf{V}}$  为  $\mathbf{V}$  的估计,  $\hat{V}_{ik} = \left( \sum_{t=0}^{T-1} \mathbf{W}(t) \overset{t \rightarrow}{\mathbf{H}} \right)_{ik}$ 。

KL 散度是在泊松噪声假设条件下求解非负矩阵  $\mathbf{W}(t)$  和  $\mathbf{H}$  的最大对数似然解<sup>[3]</sup>, 用于描述  $\hat{\mathbf{V}}$  与  $\mathbf{V}$  的逼近程度。通过对目标函数的优化可以得到乘性迭代规则<sup>[4]</sup>

$$\mathbf{W}(t) = \mathbf{W}(t) \otimes \frac{\mathbf{V} \cdot \overset{t \rightarrow}{\mathbf{H}}^T}{\mathbf{E} \cdot \overset{t \rightarrow}{\mathbf{H}}^T} \quad (3)$$

$$\mathbf{H} = \mathbf{H} \otimes \frac{\mathbf{W}(t)^T \cdot \left[ \overset{t \rightarrow}{\mathbf{V}} \right]}{\mathbf{W}(t)^T \cdot \mathbf{E}} \quad (4)$$

式中:  $\otimes$  为矩阵元素相乘运算符;  $\mathbf{E}$  为元素都为 1 的矩阵。

### 1.2 稀疏卷积非负矩阵分解

由于语音信号在时频域存在稀疏性, 因此采用稀疏非负矩阵分解算法能够更好地反映语音的这种特征, 且该算法收敛速度快, 存储空间小<sup>[7]</sup>。Paul 等在 CNMF 算法中加入稀疏性约束, 即对编码矩阵  $\mathbf{H}$  进行稀疏性约束<sup>[8]</sup>, 得到的时频字典  $\mathbf{W}(t)$  能够更为详细地包含语音的特征信息。通过控制  $\mathbf{H}$  的稀疏度可以使  $\mathbf{W}(t)$  成为过完备时频字典。对  $\mathbf{H}$  进行稀疏性约束, 式(2)可以改写为

$$D(\mathbf{V} | \hat{\mathbf{V}}) = \sum_{ik} \left( V_{ik} \log \left( \frac{V_{ik}}{\hat{V}_{ik}} \right) + V_{ik} - \hat{V}_{ik} \right) + \lambda \mathbf{H} \quad (5)$$

式中  $\lambda$  为稀疏因子, 用于控制编码矩阵的稀疏性和重构时语音的失真度。

增加稀疏性约束后带来一个新问题: 增加的稀疏约束项  $\lambda \mathbf{H}$  是单调递增函数, 当  $\mathbf{W}(t)$  增大,  $\mathbf{H}$  减小时,  $(\mathbf{W}(t) \rightarrow \alpha \mathbf{W}(t), \mathbf{H}(t) \rightarrow (1/\alpha) \mathbf{H}(t))$ , 其中  $\alpha \geq 1$ , 目标函数值将会减小。在这种情况下, 等式右边的第一项并不会受到影响, 而第二项则会减小, 必然导致  $\mathbf{W}(t)$  的元素值趋于无穷大,  $\mathbf{H}$  的元素值趋于零。因此, 为了控制  $\mathbf{W}(t)$  和  $\mathbf{H}$  中元素值大小, 需要引入另一个约束条件, 即令  $\|\mathbf{W}(t)\|_2$  实现归一化。

与 CNMF 算法的乘性迭代规则相似, 利用梯

度下降法对目标函数(5)优化可以得到如下的乘性迭代规则<sup>[7]</sup>

$$\mathbf{H} = \mathbf{H} \otimes \frac{\mathbf{W}(t)^T \cdot \left[ \frac{\mathbf{V}}{\hat{\mathbf{V}}} \right]}{\mathbf{W}(t)^T \cdot \mathbf{E} + \lambda \cdot \mathbf{E}} \quad (6)$$

$\mathbf{W}(t) =$

$$\mathbf{W}(t) \otimes \frac{\mathbf{V} \hat{\mathbf{H}}^T + \mathbf{W}(t) \otimes (\mathbf{E}(\mathbf{W}(t) \otimes (\hat{\mathbf{V}} \hat{\mathbf{H}}^T)))}{\hat{\mathbf{V}} \hat{\mathbf{H}}^T + \mathbf{W}(t) \otimes (\mathbf{E}(\mathbf{W}(t) \otimes (\mathbf{V} \hat{\mathbf{H}}^T)))} \quad (7)$$

SCNMF 算法中对  $\mathbf{W}(t)$  和  $\mathbf{H}$  的求解采用迭代求解方式,具体过程如下:

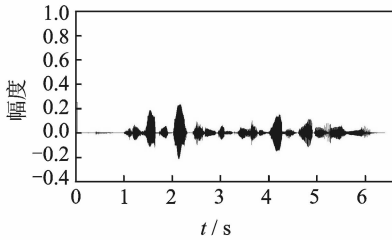
- (1) 开始;
- (2) 初始化  $\mathbf{W}(t)$  和  $\mathbf{H}$ ;
- (3) 重复;
- (4)  $k \leftarrow k+1$ ;
- (5) 更新  $\mathbf{W}$ :  $\mathbf{W}(t) =$

$$\mathbf{W}(t) \otimes \frac{\mathbf{V} \hat{\mathbf{H}}^T + \mathbf{W}(t) \otimes (\mathbf{E}(\mathbf{W}(t) \otimes (\hat{\mathbf{V}} \hat{\mathbf{H}}^T)))}{\hat{\mathbf{V}} \hat{\mathbf{H}}^T + \mathbf{W}(t) \otimes (\mathbf{E}(\mathbf{W}(t) \otimes (\mathbf{V} \hat{\mathbf{H}}^T)))}$$

$$(6) \text{ 更新 } \mathbf{H}: \mathbf{H} = \mathbf{H} \otimes \frac{\mathbf{W}(t)^T \cdot \left[ \frac{\mathbf{V}}{\hat{\mathbf{V}}} \right]}{\mathbf{W}(t)^T \cdot \mathbf{E} + \lambda \cdot \mathbf{E}}$$

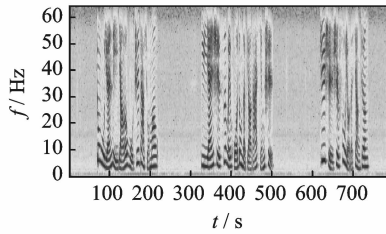
(7) 直到  $\| \mathbf{V}_{train} - \mathbf{WH} \|_F < \epsilon$ ,  $\epsilon =$

$$\sqrt{\sum (v_{ij} - \hat{v}_{ij})^2} \text{ 或 } k \geq I_{\max};$$



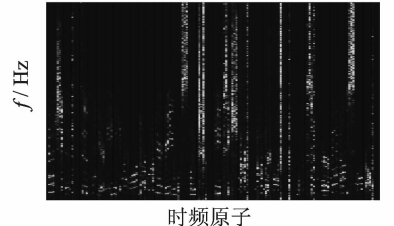
(a) 纯净语音

(a) Clean speech



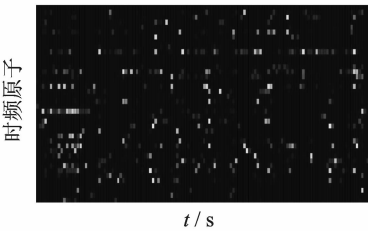
(b) 纯净语音时频谱

(b) Spectrograms of clean speech



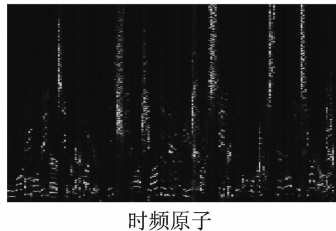
(c) SCNMF分解得到的 $\mathbf{W}$

(c) Matrix  $\mathbf{W}$  factorized using SCNMF algorithm



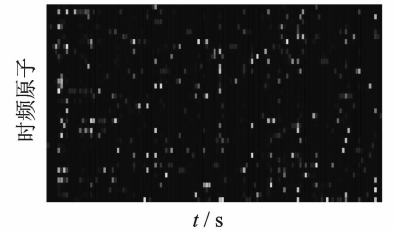
(d) SCNMF分解得到的 $\mathbf{H}$

(d) Matrix  $\mathbf{H}$  factorized using SCNMF algorithm



(e) CNMF分解得到的 $\mathbf{W}$

(e) Matrix  $\mathbf{W}$  factorized using CNMF algorithm



(f) CNMF分解得到的 $\mathbf{H}$

(f) Matrix  $\mathbf{H}$  factorized using CNMF algorithm

图 1 “SA1”纯净语音及其时频谱以及分别用 CNMF 和 SCNMF 算法分解得到的时频字典和时变增益

Fig. 1 Waveforms and spectrograms of speech “SA1”, and basis matrices and coefficient matrices factorized from “SA1” using CNMF and SCNMF algorithms respectively

(8) 结束: 最终得到  $\mathbf{W}(t)$  和  $\mathbf{H}$ 。

### 1.3 CNMF 和 SCNMF 处理语音信号的结果对比

取 TIMIT 标准语音库中“SA1”纯净语音, 通过短时傅里叶变换 (Short-time Fourier transform, STFT) 得到幅度谱  $\mathbf{V}$ 。当语音字典原子数目  $R=40$ , 时间跨度  $T=8$ , 稀疏因子  $\lambda=0.05$  时, 通过 CNMF 和 SCNMF 算法分别对  $\mathbf{V}$  进行分解, 结果如图 1 所示。

通过图 1, 可以看出 SCNMF 算法相对于 CNMF 算法, 更多地保留了信号的特征信息, 特别是中高频段的信息, 并将这些信息反映在时频字典上, 如图 1(c, e) 所示。正是由于采用 SCNMF 算法对语音信号分解时生成了过完备时频字典, 使得其时变增益相较于 CNMF 算法而言更加稀疏, 如图 1(d, f) 所示。根据信号稀疏表示原理, 信号表示得越稀疏, 分离去噪效果越好, 同时也会造成更多的语音失真<sup>[9]</sup>。尽管 CNMF 算法分离的时变增益也具有稀疏性, 但是相对而言, SCNMF 算法中加入了稀疏因子  $\lambda$ , 通过调整  $\lambda$  的取值, 能够在去除残余噪声和减小语音失真方面得到折中, 从而能得到更加准确的增强语音。

2 基于稀疏卷积非负矩阵分解的语音增强

与传统的基于字典学习的语音增强算法相似, 本文提出的基于稀疏卷积非负矩阵分解的语音增强算法分为字典学习和增强两部分。假设语音信

号和噪声信号均为加性信号, 则带噪语音信号 $y(t)$  可以表示为

$$y(t) = s(t) + n(t) \tag{8}$$

式中: $s(t)$ 为纯净语音信号; $n(t)$ 为噪声信号, 二者相互独立。算法可划分为学习和增强两个阶段, 总体框架如图 2 所示。

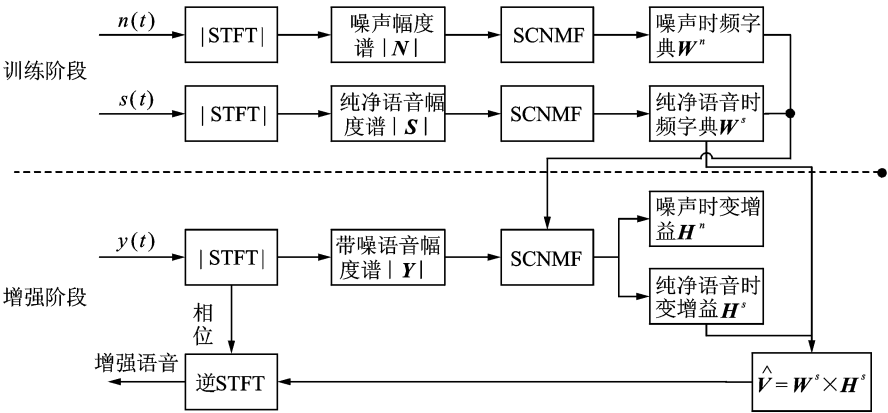


图 2 SCNMF 语音增强算法总体框架

Fig. 2 Framework of speech enhancement based on SCNMF algorithm

字典学习阶段:首先通过 STFT 得到纯净语音和噪声的幅度谱,然后利用稀疏卷积非负矩阵分解算法分别将纯净语音和噪声幅度谱分解为过完备时频字典 $W^s, W^n$ 和时变增益 $H^s, H^n$ ,将过完备时频字典 $W^s$ 和 $W^n$ 保存下来作为增强阶段的先验信息。

增强阶段:通过 STFT 得到带噪语音幅度谱,将幅度谱作为稀疏卷积非负矩阵分解的输入,用训练阶段保存的纯净语音和噪声过完备时频字典 $W^s$ 和 $W^n$ 作为先验信息,通过分解得到语音的时变增益 $H^s$ 和噪声时变增益 $H^n$ 。在重构出语音和噪声的幅度谱之后,运用逆 STFT 分析得到增强后的语音信号。

3 仿真结果

本节将通过实验仿真,首先讨论稀疏因子对 SCNMF 算法的影响,从中找出稀疏因子 $\lambda$ 的最优值。然后,再将本文提出的语音增强算法与多带谱减法、基于 NMF 和 CNMF 的语音增强算法进行比较,评估本文算法的性能。

3.1 稀疏因子对算法的影响分析

从 1.1 节中可知,影响 CNMF 算法增强效果

主要有 2 个参数,即语音字典原子数目 $R$ 和时间跨度 $T$ 。在实验中发现,40 个时频原子就足够描述语音的声学特征,更多的原子数目会增强算法计算量,性能提升却很有限,因此语音字典原子数目 $R$ 设置为 40 个。与文献[6]中的参数设置一致,在这里时间跨度 $T=8$ 。而对于 SCNMF 算法,影响其增强效果主要有 3 个参数,即语音字典原子数目 $R$ 、时间跨度 $T$ 和稀疏因子 $\lambda$ 。 $R$ 和 $T$ 的选取,根据 CNMF 算法中的取值不变,本节主要讨论稀疏因子 $\lambda$ 的选取。根据文献[7]可知,稀疏因子 $\lambda$ 为大于零的实数,它主要用于调节算法总体误差和稀疏度的关系。其取值越大,时变增益矩阵 $H$ 越稀疏,反之亦然。由 1.3 节可知,时变增益矩阵 $H$ 的稀疏度与增强后得到的语音质量具有紧密联系。 $\lambda$ 值越大, $H$ 越稀疏,则增强后的语音将包含更少的残余噪声。但是, $H$ 越稀疏在增强过程中将会有更多的语音成分丢失,从而造成更大的语音失真。为了实现对语音失真和残留噪声之间的权衡,得到更好的增强效果,本节设计了以下实验,在总体性能最优的条件下对稀疏因子 $\lambda$ 进行优化选择。

实验中,将纯净语音中分别加入 F16, M109, Babble 三种典型噪声,分别得到信噪比为 $-5, 0, 5, 10$  dB 的带噪语音。获取带噪语音的时频谱时

采用 Hamming 窗函数,窗长 512,帧移 256。在 3.2 节的算法性能比较时采用同样的参数设置。稀疏因子  $\lambda$  的取值范围设定为  $[0, 1]$ 。在实验中发现,算法性能在  $\lambda$  较小时比较敏感,因此  $\lambda$  步长采用不等间隔,在  $[0, 0.1]$  时采取较小步长(0.01),在  $[0.1, 1]$  时采取较大步长(0.1)。对数频谱距离(Log-spectral distance, LSD)反映了语音频谱失真的度量, LSD 值越小,说明算法的增强效果越好。因此本文采用 LSD 测度作为参数优化过程中的评价指标。

图 3-5 曲线为所有实验后得到的 LSD 值。显然,对于不同信噪比,在  $[0, 0.05]$  区间, LSD 值先上升而后降至最低点,在  $[0.05, 1]$  区间, LSD 值上

升至一定程度逐渐趋于平稳(当然存在个别奇异点,这里不再考虑)。这一结果表明在不同信噪比下  $\lambda=0.05$  是一个较理想的最优值。

3.2 与其他算法的性能比较

为了验证本文算法的有效性,将其与多带谱减法、基于 NMF 和 CNMF 的语音增强算法进行比较。实验中,取纯净语音为 8 kHz 采样、16 b 量化的标准女声信号。噪声取自 Noisex-92 噪声库中的 F16 飞机噪声、M109 坦克噪声和 Babble 噪声。通过 MATLAB 将纯净语音信号与噪声信号进行混合,分别生成信噪比为  $-5, 0, 5, 10$  dB 的带噪语音。

本文采用 3.1 节所述的对数频谱距离 LSD 和客观质量评估方法(Perceptual evaluation of speech quality, PESQ)对语音增强算法的性能进行比较。PESQ 取自 2001 年国际电信联盟(ITU-T)推出的 P. 862 标准,用来评价语音的主观试听效果,能够很好地反映语音信号的感知质量, PESQ 得分越高表示增强效果越好。

表 1-2 分别给出了 4 种算法在不同信噪比条件下增强语音的 LSD 值和 PESQ 值的比较情况,由表可以看出,相对于多带谱减法、基于 NMF 和 CNMF 的语音增强算法,本文提出的基于 SCNMF 的语音增强算法具有更好的增强效果。

同时可以看出,当在低信噪比情况下,本文提出的基于 SCNMF 的语音增强算法较其他 3 种算法优势更加明显,而在高信噪比情况下,本文的算法提升性能效果不明显。这是由于在低信噪比情况下,相对于其他 3 种算法,基于 SCNMF 的语音增强算法分解得到的过完备时频字典能够在充分去除残余噪声的条件下,更多地保留语音成分。

最后,对本文所提出算法的复杂度进行简单分析。本文算法的计算量包括 STFT、语音过完备时频字典与时变增益更新和逆短时傅里叶变换(ISTFT)。由 SCNMF 算法的迭代流程可知,整个算法的计算量主要集中在过完备时频字典和时变增益的迭代求解,即式(6-7)的矩阵运算上。设带噪语音幅度谱频率点数为  $L$  和整个语音帧数目为  $N$ ,为矩阵分解的秩为  $R$ ,时频原子大小为  $T$ ,根据文献[4]分析, CNMF 的时间复杂度为  $O(TLNR)$ ,而本文算法的时间复杂度与 CNMF 算法处于同一级别,因此其时间复杂度也是  $O(TLNR)$ 。

图 3 混入 F16 噪声时稀疏因子  $\lambda$  对算法性能影响  
Fig. 3 Performance of the algorithm affected by sparse factor  $\lambda$  mixed with F16 noise

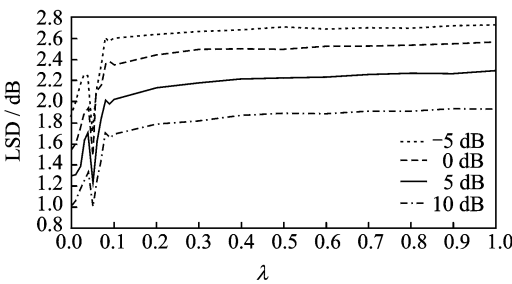


图 4 混入 M109 噪声时稀疏因子  $\lambda$  对算法性能影响  
Fig. 4 Performance of the algorithm affected by sparse factor  $\lambda$  mixed with M109 noise

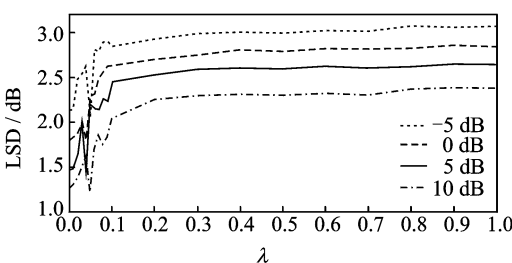


图 5 混入 Babble 噪声时稀疏因子  $\lambda$  对算法性能影响  
Fig. 5 Performance of the algorithm affected by sparse factor  $\lambda$  mixed with Babble noise

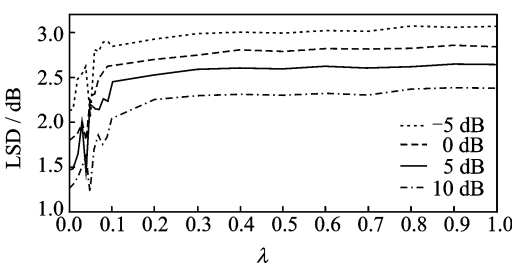


表 1 增强语音的对数谱失真  
Table 1 LSD of enhanced speech

SNR/dB	F16 噪声				M109 噪声				Babble 噪声			
	MSS	NMF	CNMF	SCNMF	MSS	NMF	CNMF	SCNMF	MSS	NMF	CNMF	SCNMF
-5	2.483	2.415	2.395	2.136	2.532	2.524	2.492	1.888	2.652	2.604	2.557	2.412
0	2.241	2.158	2.027	1.801	2.306	2.261	2.225	1.546	2.375	2.395	2.213	2.043
5	1.803	1.812	1.686	1.522	1.754	1.895	1.888	1.249	1.968	1.917	1.851	1.683
10	1.420	1.467	1.378	1.279	1.356	1.558	1.535	1.015	1.677	1.549	1.504	1.354

表 2 增强语音的 PESQ 得分  
Table 2 PESQ of enhanced speech

SNR/dB	F16 噪声				M109 噪声				Babble 噪声			
	MSS	NMF	CNMF	SCNMF	MSS	NMF	CNMF	SCNMF	MSS	NMF	CNMF	SCNMF
-5	1.800	1.957	2.158	2.360	2.014	2.094	2.151	2.469	1.681	1.750	1.917	2.066
0	2.086	2.305	2.454	2.593	2.358	2.431	2.479	2.692	2.075	2.157	2.241	2.386
5	2.458	2.591	2.692	2.889	2.706	2.716	2.784	2.935	2.446	2.496	2.533	2.667
10	2.795	2.846	2.936	3.117	3.094	2.958	3.074	3.182	2.736	2.781	2.807	2.920

4 结束语

本文提出了一种基于 SCNMF 的语音增强算法,该算法通过训练得到的先验知识对带噪语音进行增强。实验表明,在非平稳噪声和低信噪比条件下,相对于传统的增强算法,在不提升算法复杂度的同时能够更好地抑制背景噪声,取得了良好的增强效果。虽然本文提出的方法较传统的增强算法有一定的改进,但基于 SCNMF 的语音增强方法仍存在一些不足:部分参数值的选取依赖于实验结果或经验值,特别是本文对于稀疏因子的选取,缺乏理论验证;本文方法在字典学习阶段仍然需要语音信息,这在一定程度上制约了算法的适用范围。在今后的工作中,将以上述问题为方向,对语音增强算法进一步进行研究和改进。

参考文献:

[1] 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.

[2] Kamath S, Loizou P. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise[J]. IEEE Trans Acoust, Speech Signal Process, 2002, 8(4):164-168.

[3] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(10): 788-791.

[4] Smaragdis P. Convolutional speech bases and their application to supervised speech separation[J]. IEEE Trans on Audio, Speech and Language Processing, 2007, 15(1):1-12.

[5] 孙健,张雄伟,曹铁勇,等. 基于卷积非负矩阵分解的语音转换方法[J]. 数据采集与处理, 2013, 28(2): 141-148.

Sun Jian, Zhang Xiongwei, Cao Tiejong, et al. Voice conversion based on convolutive nonnegative matrix factorization[J]. Journal of Data Acquisition and Processing, 2013, 28(2):141-148.

[6] 黄建军,张雄伟,张亚非,等. 时频学习单通道语音增强算法[J]. 声学学报,2012,35(5):539-547.

Huang Jianjun, Zhang Xiongwei, Zhang Yafei, et al. Single channel speech enhancement via time-frequency dictionary learning[J]. Acta Acustica, 2012, 35(5):539-547.

[7] Ohoyer P. Non-negative matrix factorization with sparseness constraints[J]. Journal of Machine Learning Research,2004(5):1457-1469.

[8] O'Grady P D, Pearlmutter B A. Convolutional non-negative matrix factorization with sparseness constraint[C] // International Workshop on Machine Learning for Signal Processing. Maynooth, Ireland: IEEE, 2006:427-432.

[9] Zibulevsky M, Pearlmutter B A. Blind source separation by sparse decomposition in a signal dictionary [J]. Neu Comp,2001,13(4):863-882.

作者简介:张立伟(1986-),男,硕士研究生,研究方向:语音增强、信号处理, E-mail: 125289051 @ qq. com; 贾冲(1979-),男,讲师,研究方向:语音信号处理;张雄伟(1965-),男,教授,博士生导师,研究方向:语音增强、信号处理、物联网、军事化信息栅格;闵刚(1983-),男,博士研究生,研究方向:语音与图像处理;曾理(1985-),男,博士研究生,研究方向:语音信号处理、压缩感知理论及应用。