

文章编号:1004-9037(2014)02-0286-07

稀疏低秩模型下的单通道自学习语音增强算法

李轶南 贾 冲 杨吉斌 吴海佳 张立伟

(解放军理工大学指挥信息系统学院,南京,210007)

摘要:针对现有基于字典学习的增强算法依赖先验信息的问题,基于矩阵的稀疏低秩分解提出一种无监督的单通道语音增强算法。该算法首先通过稀疏低秩分解将带噪语音的幅度谱分解为低秩、稀疏和噪声3部分,然后通过对低秩部分进行自学习构建出噪声字典,最后利用所得噪声字典和乘性迭代准则于低秩和稀疏部分中分离出纯净语音。相较于其他基于字典学习的语音增强算法,本文所提算法无需语音或噪声的先验信息,因而更加方便和实用。实验结果显示,本文算法能够在保留语音谐波结构的同时有效抑制噪声,增强效果明显优于鲁棒主成分分析和多带谱减法。

关键词:语音增强;无监督学习;字典学习;稀疏低秩分解

中图分类号:TN912.3

文献标识码:A

Self-Learning Approach for Monaural Speech Enhancement Based on Sparse and Low-Rank Matrix Decomposition

Li Yinan, Jia Chong, Yang Jibin, Wu Haijia, Zhang Liwei

(College of Command Information System, PLA University of Science and Technology, Nanjing, 210007, China)

Abstract: To resolve the prior dependency of existing enhancement algorithms based on dictionary learning, an unsupervised self-learning approach for speech enhancement in one channel record is presented. Firstly, the algorithm decomposes the magnitude spectrum of noisy speech efficiently into low-rank part, sparse part and noise part. Then, the dictionary of noise is acquired by learning the low-rank part. Finally, the clean speech is separated by using the acquired noise dictionary and multiplicative update rules. As the approach is unsupervised, it is more convenient and practice than other enhancement methods based on dictionary learning. The experiment results show that the approach proposed outperforms other enhancement methods like robust principal component analysis and multiband spectra subtraction in terms of harmonic structure maintaining and noise suppression.

Key words: speech enhancement; unsupervised learning; dictionary learning; sparse and low-rank decompositions

引 言

语音信号在实际的应用中不可避免地会受到来自周围环境噪声的污染,导致语音质量和可懂度的下降。学者们一直致力于研究增强算法以期实现从被噪声污染的语音信号中尽可能准确地恢复出原始语音信号。自20世纪70年代以来,出现了诸如谱减法、卡尔曼滤波法、信号子空间法等诸多

经典算法^[1-2],相应的改进算法也层出不穷,然而这些算法在去除实际环境中的噪声,特别是非平稳噪声以及类语音噪声时,其效果往往难以令人满意。

近年来,字典学习算法不断涌现^[3-4],基于字典学习的语音增强算法成为学者们研究的热点,新的算法不断被提出,这些新算法的出现为解决传统增强算法难以解决的棘手问题带来了新的曙光。

文献[5]对语音和噪声分别进行字典学习,得到二者的非负联合字典,通过将带噪语音在联合字

典上进行投影,分离出纯净语音。文献[6]发展了联合字典的思想,将卷积模型引入字典学习的过程中,使字典中的原子能够更好地反映出语音的时频域结构特征。尽管上述增强算法能够获得较好的增强效果,但是这种基于全监督字典学习的增强算法需要大量语音和噪声的先验知识,使得这些方法难以推广到实际应用之中。

学者们提出了很多新的更加实用的增强算法。文献[7]基于非负稀疏编码(Non-negative sparse coding, NNSC)提出一种半监督增强算法,只需要预先训练出噪声字典,就能够实现对于带噪语音中特定噪声的去除。文献[8]使用 K-SVD 算法预先训练得到语音字典,利用语音激活检测(Voice activity detection, VAD)在无语音期间获取噪声字典,实现了纯净语音信号的提取。这些方法虽然降低了对于先验知识的需求,但是依然无法实现无监督条件下的语音增强,使得上述算法仍然具有其自身的局限性。

自文献[9]提出鲁棒主成分分析(Robust principal component analysis, RPCA)以来,语音的稀疏低秩建模逐渐成为研究热点^[10-11]。受 RPCA 思想的启发,本文将稀疏低秩模型引入到基于字典学习的语音增强中去,提出一种自学习语音增强算法。该算法根据噪声是否易于进行低秩建模将噪声区分为结构化噪声和非结构化噪声两类。首先通过 Go Decomposition(GoDec)算法^[12]将带噪语音幅度谱分解为稀疏、低秩和噪声 3 部分,通过舍弃噪声部分去除非结构化噪声;然后,通过字典学习算法对低秩部分进行自学习,得到结构化噪声的字典;最后,使用所得噪声字典和迭代公式,分离出纯净语音的幅度谱。实验结果表明,本文算法能够在保留语音固有谐波特性的同时有效移除噪声,增强效果显著优于诸如 RPCA 和多带谱减法等无监督增强算法。

1 语音的稀疏低秩模型

主成分分析(Principal component analysis, PCA)是最常用的降维方法,该方法能够较好地挖掘出高维数据样本中潜在的低维特征。然而,当样本数据中存在少数数值偏差较大的样本时,其分解所得的结果往往会严重偏离实际。为此,Candès 等人利用凸优化相关理论提出了一种新的解决方法——鲁棒主成分分析(Robust principal component analysis, RPCA)。RPCA 能够将一个被任

意大小的稀疏噪声所污染的数据矩阵分解为一个稀疏矩阵和一个低秩矩阵之和,从而克服了传统 PCA 算法对于稀疏的高强度噪声敏感的不足。

1.1 鲁棒主成分分析

假设被噪声污染的幅度谱为 \mathbf{Y} ,且 \mathbf{Y} 中的噪声具有潜在的低秩结构,由于语音信号具有时频域上的稀疏性,那么就能够将 \mathbf{Y} 分解为低秩部分 \mathbf{L} 和稀疏部分 \mathbf{S} 之和,如式(1)所示

$$\mathbf{Y} = \mathbf{L} + \mathbf{S} \quad (1)$$

式(1)可以使用下面的优化问题来进一步描述

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \text{rank}(\mathbf{L}) + \lambda \|\mathbf{S}\|_0 \\ \text{s. t. } \mathbf{Y} = \mathbf{L} + \mathbf{S} \end{aligned} \quad (2)$$

式中: λ 是用以权衡矩阵 \mathbf{L} 的秩和矩阵 \mathbf{S} 的 l_0 范数(即矩阵中非零元素的总个数)的常系数,通常取 $\lambda = \frac{1}{\max(n_1, n_2)}$, n_1 和 n_2 为矩阵 \mathbf{Y} 两个维度上的维数; $\|\cdot\|_0$ 表示矩阵中非零元素的个数。由于 l_0 范数非凸,此问题不易求解,是非确定性多项式难题(Non-deterministic polynomial hard, NP-hard)。RPCA 使用矩阵的核范数(矩阵奇异值绝对值之和)和矩阵的 l_1 范数(非零元素的绝对值之和)分别代替矩阵 \mathbf{L} 的秩和矩阵 \mathbf{S} 的 l_0 范数,将式(2)松弛为形如式(3)的凸优化问题。

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{s. t. } \mathbf{Y} = \mathbf{L} + \mathbf{S} \end{aligned} \quad (3)$$

式中: $\|\cdot\|_*$ 和 $\|\cdot\|_1$ 分别表示核范数和 l_1 范数,其中核范数可以认为是奇异值的 l_1 范数。

对于式(3)可以使用增广拉格朗日乘子法(Augmented lagrange multiplier, ALM)^[13]来方便地进行求取。

1.2 稀疏低秩模型在语音分离中的应用

文献[10]首次将 RPCA 算法应用到语音分离中,实现了歌唱语音和音乐伴奏的无监督分离。文献[10]指出,音乐伴奏往往具有明显的重复结构,所对应的幅度谱具备明显的低秩特征;与此同时,歌唱语音具有显著的谐波结构,这使得歌唱语音在整个歌曲的幅度谱上表现地十分稀疏。依据上述事实,文献通过 RPCA 算法分解歌曲的幅度谱,所得的低秩部分代表了音乐伴奏,而稀疏部分则对应于歌唱语音,据此实现了二者的分离。

然而,实际环境中的背景噪声较上面所述场景更为复杂,既有类似于音乐伴奏的低秩噪声,又有诸如高斯噪声一类的高维噪声,因此,直接应用 RPCA 进行增强往往效果并不理想,后面的实验将

会证明这一点。进一步的研究显示,将带噪语音幅度谱分解为低秩、稀疏和噪声 3 部分之和的分解模型更加合理。在此分解模型下,低秩部分代表了噪声中结构性强的部分;语音信号由于其固有的短时平稳性,将会被分解到稀疏部分中去;噪声部分则代表了密集的非结构化噪声,是噪声中结构特征不明显的部分。本文算法依据上述分解模型实现了无监督条件下的单通道语音增强。

2 单通道自学习语音增强算法

增强算法对于传统的语音加性噪声模型进行了进一步的细化和完善:假设 $y(t)$ 为带噪语音信号, $s(t)$ 为纯净语音信号, $n(t)$ 为噪声信号。根据 $n(t)$ 是否易于进行低秩建模将噪声进一步细化区

分为结构化噪声和非结构化噪声两部分

$$y(t) = s(t) + n(t) = s(t) + n_s(t) + n_u(t) \quad (4)$$

式中: $n_s(t)$ 是噪声中结构性强的部分,即结构化噪声部分; $n_u(t)$ 则噪声中结构性不明显的部分,即非结构化噪声部分。

本文增强算法如图 1 所示。首先对于带噪语音进行短时傅里叶变换(Short time Fourier transformation, STFT),求取带噪语音的幅度谱,然后使用矩阵的稀疏低秩分解将带噪语音的幅度谱分解为低秩、稀疏和噪声 3 部分,通过舍弃噪声部分移除非结构化噪声 $n_u(t)$,通过对于低秩部分进行学习得到结构化噪声部分 $n_s(t)$ 的字典 D_n ;最后利用所得的噪声字典和相应的乘性迭代公式分离出纯净语音信号。

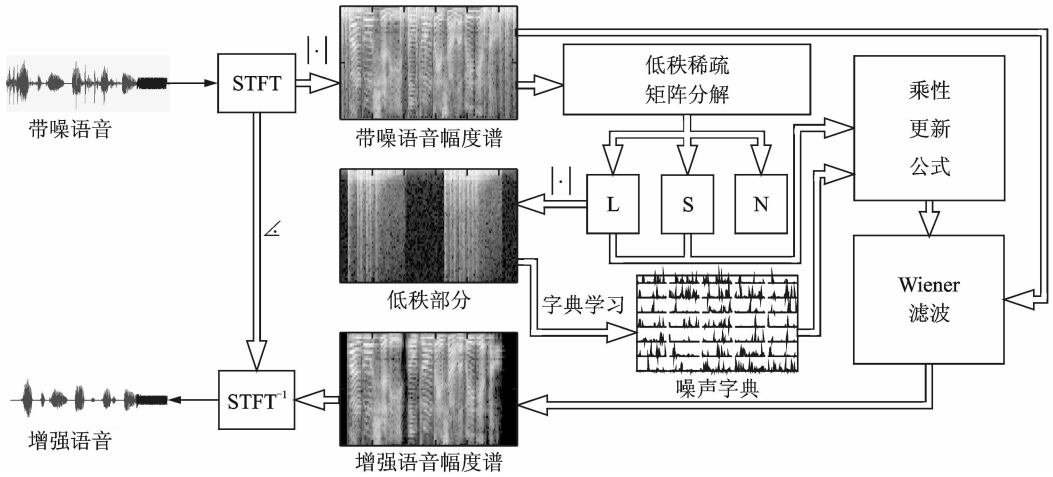


图 1 增强算法流程图

Fig. 1 Pipeline of the proposed algorithm

2.1 语音幅度谱表示

首先对带噪语音信号 $y(t)$ 进行分帧加窗,然后计算每一帧的快速傅里叶变换(Fast Fourier transformation, FFT)得到 $y(t)$ 的短时傅里叶变换(Short time Fourier transformation, STFT)。

$$\dot{Y} = \text{FFT}[h(0) \cdots h(W-1)] \cdot$$

$$\begin{bmatrix} y(t_1) & y(t_2) & \cdots & y(t_N) \\ \vdots & \vdots & \vdots & \vdots \\ y(t_1 + W - 1) & y(t_2 + W - 1) & \cdots & y(t_N + W - 1) \end{bmatrix} \quad (5)$$

式中: W 为所取语音帧帧长和 FFT 的长度, $h(n)$ ($n=0, \dots, W-1$) 为归一化语音窗, $R = t_i - t_{i-1}$ 为相邻语音帧之间交叠的样点个数,取 $R=L/2$ 。若带噪语音信号的帧数为 N ,则经 STFT 变换后所得矩阵的大小为 $W \times N$ 。通过对 \dot{Y} 求取绝对

值就可以得到带噪语音的幅度谱

$$Y = |\dot{Y}|, \quad Y \in R^{W \times N} \quad (6)$$

记录下 \dot{Y} 中相应的相位信息 $\angle \dot{Y}$,以便在语音重构时对语音波形进行恢复。

2.2 语音幅度谱的稀疏低秩分解

GoDec 算法是在 RPCA 的基础上优化了矩阵的稀疏低秩分解模型而得到的一种矩阵分解算法。使用该算法可以将带噪语音的幅度谱分解为 3 个矩阵之和,即

$$Y = L + S + N, \quad \text{rank}(L) \leq r, \quad \text{card}(S) \leq k \quad (7)$$

式中: $\text{rank}(L)$ 表示的是矩阵 L 的秩, $\text{card}(S)$ 表示的是矩阵 S 的势,即矩阵中非零元素的数目。

为了求解式(7),将其转化为两个最优化的子问题,其本质是在残差最小化的条件下,对低秩和

稀疏矩阵分别进行估计,如式(8)

$$\begin{cases} \mathbf{L}_t = \arg \min_{\text{rank}(\mathbf{L}) \leq r} \|\mathbf{Y} - \mathbf{L} - \mathbf{S}_{t-1}\|_F^2 \\ \mathbf{S}_t = \arg \min_{\text{card}(\mathbf{S}) \leq k} \|\mathbf{Y} - \mathbf{L}_t - \mathbf{S}\|_F^2 \end{cases} \quad (8)$$

在求解式(8)中两个子问题的迭代过程中计算开销最大的运算为 SVD 运算,GoDec 算法采用双边随机投影(Bilateral random projection, BRP)来代替 SVD 运算,大大提高了运算效率,显著降低了算法的计算复杂度。

使用 GoDec 算法将所得的幅度谱矩阵 \mathbf{Y} 分解为低秩 \mathbf{L} ,稀疏 \mathbf{S} 和噪声 \mathbf{N} 三个部分。由于非结构化噪声部分很难用稀疏部分 \mathbf{S} 或是低秩部分 \mathbf{L} 来表示,此类噪声通常被 GoDec 算法分解到噪声部分即矩阵 \mathbf{N} 中,直接去除此部分即可移除非结构化噪声。

根据上述分析,直接将带噪语音的幅度谱分解为低秩、稀疏和噪声 3 部分似乎就能实现对于语音信号的增强。然而,实际的实验结果显示直接将矩阵进行稀疏低秩分解难以实现对非平稳噪声环境下 3 部分的完美分离,这主要是因为单纯的稀疏低秩分解所得的低秩部分更关注语音信号在时频域上的重复性,而并不侧重于研究这些重复信号所具有的具体特征。为了克服这个缺点,本文算法引入字典学习算法来学习这些不断重复的噪声信号所具有的独特特征,从而进一步提升语噪分离的性能。

2.3 稀疏低秩模型下的自学习语音增强

结构化噪声部分通常具有比语音信号更加明显的重复和冗余结构。通过选取合适的秩,能够在矩阵分解所得的低秩部分,即在矩阵 \mathbf{L} 中获取到此类噪声的结构信息,通过对低秩部分进行字典学习,就能够获取相应的噪声字典,从而实现自学习。

相关研究表明,由局部到整体的累加方式更符合人类感官从局部到整体的认知过程且人耳对声音相位不敏感,这里首先对 \mathbf{L} 求模值,然后通过非负矩阵分解(Non-negative matrix factorization, NMF)^[14]的方式获得结构化噪声的归一化非负字典 \mathbf{D}_n 。设结构化噪声字典中原子的个数为 num ,则有

$$|\mathbf{L}| \approx \mathbf{D}_n \mathbf{C}'_n \quad (9)$$

式中: $|\mathbf{L}| \in \mathbf{R}^{W \times N}$ 为稀疏低秩分解所得的低秩矩阵的模值, $\mathbf{D}_n \in \mathbf{R}^{W \times num}$ 和 $\mathbf{C}'_n \in \mathbf{R}^{num \times N}$ 分别为结构化噪声的归一化非负字典和对应的增益系数矩

阵。

假设移除非结构化噪声以后的语音幅度谱为 $\bar{\mathbf{Y}}$,纯净语音的字典为 \mathbf{D}_s ,对应的增益系数矩阵为 \mathbf{C}_s ,结构化噪声字典为 \mathbf{D}_n ,相应的增益系数矩阵为 \mathbf{C}_n 。语音增强问题就转化为

$$\bar{\mathbf{Y}} = \mathbf{L} + \mathbf{S} = \mathbf{D}\mathbf{C} = [\mathbf{D}_s \quad \mathbf{D}_n] \begin{bmatrix} \mathbf{C}_s \\ \mathbf{C}_n \end{bmatrix} \quad (10)$$

固定从低秩部分中学习得到的结构化噪声字典 \mathbf{D}_n ,并使用下面的基于梯度下降的乘性更新公式^[15]更新剩下的 3 个矩阵

$$\mathbf{C}_s \leftarrow \mathbf{C}_s \otimes \frac{\mathbf{D}_s^T \bar{\mathbf{Y}}}{(\mathbf{D}_s^T \mathbf{D})\mathbf{C} + \lambda_s} \quad (11)$$

$$\mathbf{C}_n \leftarrow \mathbf{C}_n \otimes \frac{\mathbf{D}_n^T \bar{\mathbf{Y}}}{(\mathbf{D}_n^T \mathbf{D})\mathbf{C} + \lambda_n} \quad (12)$$

$$\mathbf{D}_s \leftarrow \mathbf{D}_s \otimes \frac{\bar{\mathbf{Y}}\mathbf{C}_s^T + \mathbf{D}_s \otimes (\mathbf{1}(\mathbf{D}_s \otimes (\mathbf{D}(\mathbf{C}\mathbf{C}^T))))}{\mathbf{D}(\mathbf{C}\mathbf{C}^T) + \mathbf{D}_s \otimes (\mathbf{1}(\mathbf{D}_s \otimes (\bar{\mathbf{Y}}\mathbf{C}_s^T)))} \quad (13)$$

式中: $\mathbf{1}$ 为全 1 矩阵, λ_s 和 λ_n 是控制增益矩阵稀疏度的常数,符号 \otimes 表示矩阵或向量中对应元素的乘。

2.4 增强语音波形重构

经过式(11-13)的若干次的迭代后,就能够分离出纯净语音和对应噪声。本文采用 Wiener 滤波法进行后处理,来进一步提升算法的增强效果,使得增强后的语音听起来更自然。

Wiener 滤波可以认为是对于语音时频域上的最小均方误差估计,其频域表达式为

$$\hat{\mathbf{Y}} = (\angle \hat{\mathbf{Y}}) \cdot \frac{E[|\mathbf{Y}_s|^2]}{E[|\mathbf{Y}_s|^2 + |\mathbf{Y}_n|^2]} \bar{\mathbf{Y}} \quad (14)$$

式中: $\mathbf{Y}_s = \mathbf{D}_s \mathbf{C}_s$, $\mathbf{Y}_n = \mathbf{D}_n \mathbf{C}_n$, $\angle \hat{\mathbf{Y}}$ 为 STFT 时记录下的带噪语音相位信息。

将估计出的语音频谱 $\hat{\mathbf{Y}}$ 进行逆 STFT 就可以重构出增强后的语音波形。

3 增强算法性能评估

实验选用的纯净语音来自 TIMIT 标准语音库中的男女语音片段各 5 句,噪声则来自 Noisex-92 标准噪声库^[16],将二者下采样到 8 kHz,信噪比分别选取 -5, 0, 5 和 10 dB 对算法分别进行评估。在稀疏低秩分解中,设定低秩部分的秩为 1,来获取重复性明显的噪声部分,同时避免语音信号过多地泄露到低秩部分中。纯净语音和噪声字典的原子个数均设定为 40。在字典更新的过程中,乘性

迭代公式进行不超过 200 次的迭代。

测试实验着重选取了 Pink, F16, Machinegun 和 Babble 4 种具有代表性的噪声进行测试。其中 Pink 是自然界中最常见的噪声,其频率分量功率主要分布在中低频段,并没有明显的时频结构,代表了非结构化噪声;F16 为美军双座 F16 战斗机巡航过程中座舱内的噪声,能量集中在在 0~700 Hz 和 2 750 Hz 频带附近且呈现出均匀的变化特性;Machinegun 为机枪扫射时所发出的噪声,能够代表瞬时噪声;Babble 为有容纳大约 100 个人个同时在讲话的人的餐厅中的背景噪声,能量主要集中在低频段,是类语音噪声的代表)

采用 BSS-EVAL 评价体系^[17]、短时客观可懂度测量(Short-Time objective intelligibility measure, STOI)^[18]和语音质量客观评估方法(Perceptual evaluation of speech quality, PESQ)^[19]对语音的性能进行评估。其中,BSS-EVAL 是目前公认的性能比较好的盲源分离算法评估体系,该评估体系通过计算信源引入噪声比(Signal artifacts ratio, SAR)、信干比(Signal to interference ratio, SIR)、信源失真比(Signal to distortion ratio, SDR),从不同方面反映了增强算法的效果;STOI 是一种比较新的机器驱动的可懂度客观评估方法,其计算值与人对于语音的实际可懂度高度相关;PESQ 是用来评价语音的主观试听效果的客观计算方法,能够很好地近似平均意见得分(Mean opinion score, MOS)。

将本文所提出的基于稀疏低秩模型的自学习语音增强算法与 RPCA 算法和多带谱减法(Multi-band spectral subtraction, MBSS)^[20]两种无监督算法进行比较来对本文算法进行评估。其中,RP-CA 算法是一种基于稀疏低秩分解的无监督分离方法,在分离歌唱和伴奏时,体现出很好的性能;MBSS 是一种性能比较好的增强算法,在多种噪声环境下均能获得较好的增强效果。

图 2 给出了 4 种噪声环境下 BSS-EVAL 评价体系中 3 个关键指标的平均测量值。在这 3 个指标中,SAR 对于分离过程中引入噪声的大小进行评估,SIR 是表征增强算法对噪声抑制程度的一个值,SDR 则反映了分离算法的总体性能。3 个指标越高的算法,相应的性能也就越好。

由图 2 中可以看出,本文算法的 SAR 测度平均比 RPCA 方法高出 3.3 dB,比 MBSS 高出

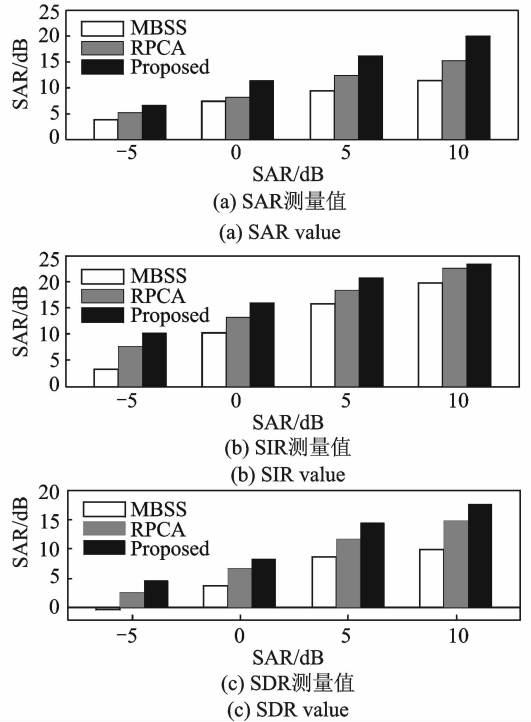


图 2 BSS-EVAL 评价体系平均测量值
Fig. 2 Performance of MBSS, RPCA and the proposed algorithm in terms of BSS-EVAL metrics

5.5 dB。在 SIR 测度方面,本文算法平均比 RPCA 高出 2.3 dB,比 MBSS 高出 5.8 dB。SDR 测度显示,本文算法比 RPCA 高出 2.1 dB,比 MBSS 高出 5.3 dB。这些指标说明相较于其他两种算法,本文算法引入更少的噪声,具有更好的噪声抑制能力,其综合性能指标也更加优越。

图 3 列出了 3 种增强方法的 STOI 测量值。可以看出,本文算法具有比其他两种方法更高的测

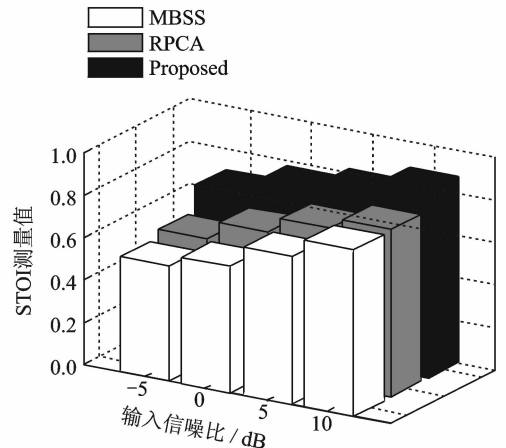


图 3 STOI 平均测量值
Fig. 3 Average short-time objective intelligibility measures of different algorithms

量值。这说明,本文算法的可懂度要高于其他两种算法,并且随着信噪比的下降,本文算法的可懂度下降相对缓慢,这说明本文算法在低信噪比下仍具有较好的可懂度。

表 1 列举了 3 种增强方法在 4 种不同噪声环境和输入信噪比下的 PESQ 测量值,这些测量值进一步验证本文算法的性能。

表 1 不同算法和噪声下的 PESQ 值

Table 1 PESQ scores of different algorithms and noises

噪声	SNR /dB	增强算法		
		MBSS	RPCA	Proposed
Pink	-5	1.98	2.11	2.32
	0	2.28	2.37	2.61
	5	2.44	2.51	2.79
	10	2.63	2.69	2.82
F16	-5	2.04	2.24	2.46
	0	2.27	2.48	2.58
	5	2.36	2.62	2.76
	10	2.61	2.79	2.90
Machinegun	-5	2.34	2.44	2.70
	0	2.59	2.62	2.81
	5	2.72	2.84	2.92
	10	2.86	2.91	3.08
Babble	-5	1.63	2.04	2.38
	0	1.82	2.27	2.31
	5	2.23	2.36	2.46
	10	2.61	2.68	2.77

然而,无论是 BSS-EVAL 评价体系、STOI 度还是 PESQ 值,都只能从宏观上反映算法的性能,为了更好地观察出增强语音信号的细节特征,本文给出了 3 种增强算法对于输入信噪比为 5 dB,被 F16 飞机噪声所污染的纯净语音进行增强前后的语谱图,如图 4 所示。

可见,RPCA 比 MBSS 更好地去除了噪声,然而也带来了更大的语音损伤。在 2 750 Hz 的频带附近,使用 RPCA 进行增强的结果依然存在少量噪声残留,导致了类似音乐噪声的试听感受。本文算法在继承了 RPCA 优点的同时,很好地克服了 RPCA 的缺点,能够在有效去除噪声的前提下,较好地保持语音信号的固有谐波特性,增强效果明显优于前两种算法。

4 结束语

本文基于矩阵的稀疏低秩分解提出一种单通道自学习语音增强算法,实现了对于语音信号的无监督增强,该算法能够在保留语音信号固有谐波特

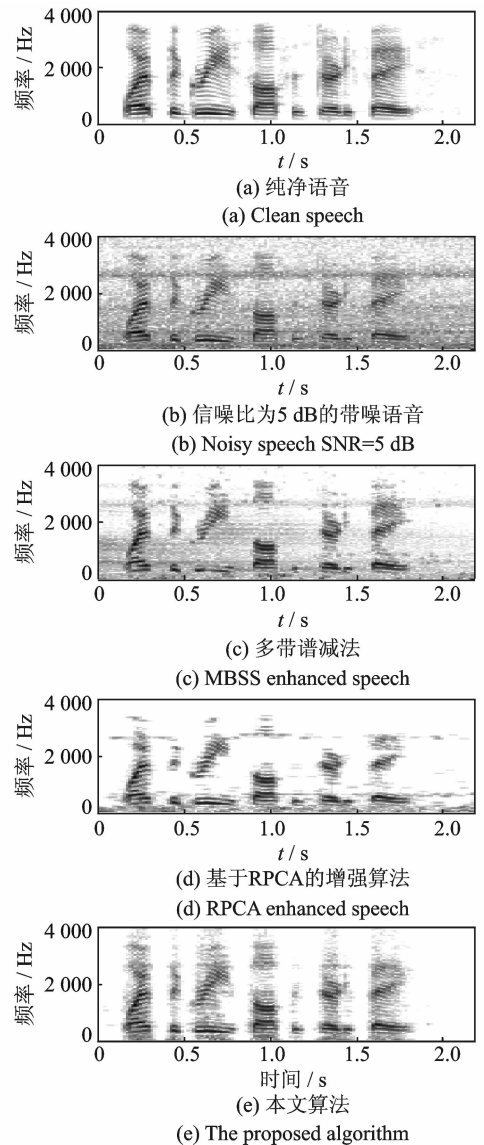


图 4 语音语谱图

Fig. 4 Spectrograms

性的同时很好地移除噪声。在低信噪比条件下,本文算法的增强结果依然能够保持较高的可懂度,其增强性能显著优于鲁棒主成分分析方法和多带谱减法。

参考文献:

- [1] Mohammadiha N, Smaragdis P, Leijon A. Supervised and unsupervised speech enhancement using nonnegative matrix factorization[J]. IEEE Transactions on Audio, Speech and Language Processing, 2013, 21(10): 2140-2151.
- [2] 张丽艳, 殷福亮. 一种改进的奇异值分解语音增强方法[J]. 电子与信息学报, 2008, 30(2): 357-361. Zhang Liyan, Yin Fuliang. An improved speech enhancement method based on SVD[J]. Journal of E-

- lectronics & Information Technology, 2008, 30(2): 357-361.
- [3] 曾理, 张雄伟, 陈亮, 等. 基于压缩感知的 K-L 分解语音稀疏表示算法[J]. 数据采集与处理, 2013, 28(3): 267-273.
- Zeng Li, Zhang Xiongwei, Chen Liang, et al. Compressed-sensing-based speech sparse representation with K-L expansion[J]. Journal of Data Acquisition and Processing, 2013, 28(3): 357-361.
- [4] 王天荆, 郑宝玉, 杨震. 基于自适应冗余字典的语音信号稀疏表示算法[J]. 电子与信息学报. 2011, 33(10): 2372-2377.
- Wang Tianjing, Zheng Baoyu, Yang Zhen. A speech signal sparse representation algorithm based on adaptive overcomplete dictionary[J]. Journal of Electronics & Information Technology, 2011, 33(10): 2372-2377.
- [5] Wilson K, Raj B, Smaragdis P, et al. Speech denoising using nonnegative matrix factorization with priors[C]// ICASSP. Las Vegas; IEEE, 2008; 4029-4032.
- [6] Smaragdis P. Convolution speech bases and their application to supervised speech separation[J]. IEEE Transactions on Audio, Speech and Language Processing, 2007, 15(1):1-12.
- [7] Mikkel N S, Jan L, Fu-Tien, et al. Wind noise reduction using non-negative sparse coding[C]// IEEE Workshop on Machine Learning for Signal Processing. Thessaloniki; IEEE, 2007:431-436.
- [8] Christian D S, Tomas D, Joachim M, et al. Speech enhancement using generative dictionary learning [J]. IEEE Transactions on Audio, Speech and Language Processing, 2012, 20(6):1698-1712.
- [9] Emmanuel J C, Xiaodong Li, Yi Ma, et al. Robust principal component analysis [J]. Journal of the ACM, 2011, 58(3):1-37.
- [10] Huang Po-Sen, Chen S D, Smaragdis P, et al. Singing-voice separation from monaural recordings using robust principal component analysis[C]// ICASSP. Kyoto; IEEE, 2012; 57-60.
- [11] Sprechmann P, Bronstein A, Bronstein M, et al. Learnable low rank sparse models for speech denoising [C] // ICASSP. Vancouver, Canada; IEEE, 2013:136-140.
- [12] Zhou Tianyi, Tao Dacheng. GoDec: randomized low-rank & sparse matrix decomposition in noisy case [C] // 28th International Conference on Machine Learning. Bellevue, Washington; Springer Berlin Heidelberg, 2011; 33-40.
- [13] Lin Z, Chen M, Ma Y. The augmented Lagrange multiplier method for exact recovery of a corrupted low-rank matrices[EB/OL]. <http://arxiv.org/abs/1009.5055>, 2010-9-26.
- [14] Lee D D, Seung H S. Learning parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755):788-791.
- [15] Kristian T A. Wind noise reduction in single channel speech signals[D]. Technical University of Demark, 2008; 25-37.
- [16] Rice University Digital Signal (DSP) group. Noisex92 Noise Database[EB/OL]. http://spib.rice.edu/spib/select_noise.html. 1996-8-16.
- [17] Vincent E, Gribonval R, Fevotte C. Performance measurement in blind audio source separation [J]. IEEE Transaction on Audio, Speech, and Language Processing, 2006, 14(4):1462-1469.
- [18] Cees H T, Richard C, Hendriks, et al. An algorithm for intelligibility prediction of time-frequency weighted noisy speech [J]. IEEE Transactions on Audio, Speech and Language Processing, 2011, 19(7):2125-2136.
- [19] International Telecommunication Union. Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, recommendation-862[S]. P. 862. Switzerland Geneva; ITU-T, 2001.
- [20] Philipos C, Loizou. Speech enhancement: Theory and practice[M]. Boca Raton, Florida; Taylor and Francis, 2007; 120-125.

作者简介:李轶南(1988-) 男, 硕士研究生, 研究方向: 语音与图像处理、宽带频谱感知, E-mail: 649447934@qq.com; 贾冲(1979-) 男, 讲师, 研究方向: 语音信号处理; 杨吉斌(1978-) 男, 副教授, 研究方向: 语音增强、语音编码; 吴海佳(1986-) 男, 博士生, 研究方向: 语音与图像处理、云计算、深度学习; 张立伟(1986-) 男, 硕士研究生, 研究方向: 语音增强。