第 42 卷第 2 期　　声　学　学　报　　Vol. 42, No. 2
2017 年 3 月　　ACTA　ACUSTICA　　Mar., 2017

用于无监督语音降噪的听觉感知鲁棒主成分分析法*

闵　刚[1,2]　邹　霞[1]　韩　伟[1]　张雄伟[1†]　谭　薇[2]

(1　解放军理工大学指挥信息系统学院　南京　210007)

(2　西安通信学院　西安　710106)

2015 年 11 月 26 日收到

2016 年 6 月 16 日定稿

摘要　针对现有稀疏低秩分解语音降噪方法对人耳听觉感知特性应用不充分、语音失真易被感知的问题，提出了一种用于语音降噪的听觉感知鲁棒主成分分析法。由于耳蜗基底膜对于频率感知具有非线性特性，该方法采用耳蜗谱图作为语噪分离的基础。此外，选用符合人耳听觉感知特性的板仓－斋田距离度量作为优化目标函数，在稀疏低秩建模过程中引入非负约束以使分解分量更符合实际物理含义，并在交替方向乘子法框架下推导了具有闭合解形式的迭代优化算法。文中方法在语音降噪时是完全无监督的，无需预先训练语音或噪声模型。多种类型噪声和不同信噪比条件下的仿真实验验证了该方法的有效性，噪声抑制效果较目前同类算法更为显著，且降噪后语音的可懂度和总体质量有所提高、至少相当。

PACS 数：43.60, 43.72

Unsupervised speech denoising via perceptually motivated robust principal component analysis

MIN Gang[1,2]　ZOU Xia[1]　HAN Wei[1]　ZHANG Xiongwei[1]　TAN Wei[2]

(1　*College of Command Information Systems, PLA University of Science and Technology*　Nanjing　210007)

(2　*Xi'an Communications Institute*　Xi'an　710106)

Received Nov. 26, 2015

Revised Jun. 16, 2016

Abstract　To overcome the shortcomings in the existing sparse and low-rank speech denoising method that the auditory perceptual properties are not fully exploited and the speech distortion is easily perceived, a perceptually motivated robust principal component analysis (ISNRPCA) method is presented. To reflect the nonlinear property for frequency perception of the basilar membrane, cochleagram is utilized as inputs of ISNRPCA. ISNRPCA uses the perceptually meaningful Itakura-Saito measure as its optimization objective function. Moreover, nonnegative constraints are also imposed to regularize the decomposed terms with respect to their physical meaning. We propose an alternating direction method of multipliers (ADMM) to solve the optimization problem of ISNRPCA. ISNRPCA is totally unsupervised, neither the speech nor the noise model needs to be trained beforehand. Experimental results under various noise types and different SNRs demonstrate that ISNRPCA shows promising results for speech denoising. Compared to the state of the art baselines, this method achieves better performance on noise suppression and demonstrates at least comparable intelligibility and overall speech quality.

引言

语音降噪旨在抑制嘈杂环境中的干扰噪声，获得尽可能纯净的语音信号，在语音编码、识别、说话人识别等领域有着广阔的应用前景[1−2]，现实环境中广泛存在的背景噪声严重恶化语音编码、说话人识别等系统的性能，造成语音质量和可懂度的降低[3−4]。语

*　国家自然科学基金项目 (61471394, 61402519) 和江苏省自然科学基金项目 (BK20140071, BK20140074) 资助

†　通讯作者：张雄伟，xwzhang9898@163.com

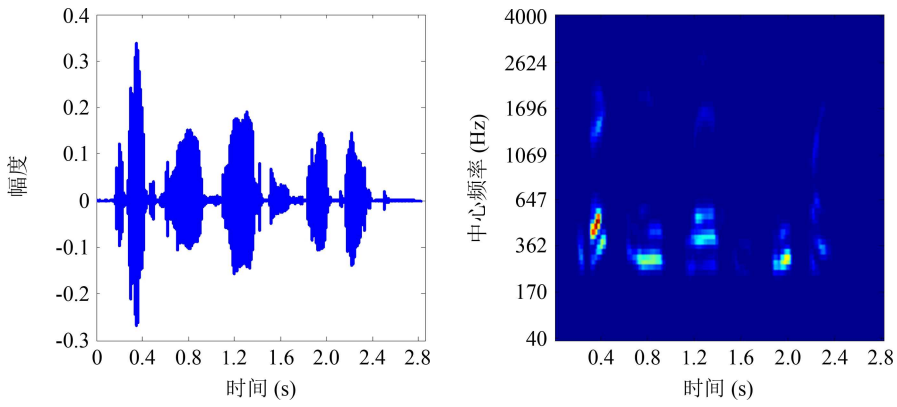
音降噪是语音信号处理领域的一个重要分支，经过多年研究并相继提出了谱减法、维纳滤波法、语音短时谱估计法等诸多成功的语音降噪模型和方法 [5−6]。然而，如何应对现实环境中复杂多变的干扰噪声，特别是在低信噪比、非平稳噪声环境下进行语音降噪仍然是富有挑战性的难题，当只有一个传声器采集到的声音信息可以应用时，该问题将变得更加困难。

传统多带谱减、维纳滤波等语音降噪方法需要对噪声进行可靠地估计，如果噪声估计不准确则语音降噪性能会严重恶化。近年来，随着硬件平台计算能力的不断提升以及机器学习方法在信号处理领域的广泛应用，盲源分离框架下的监督和无监督语噪分离新方法不断涌现，并在语音降噪方面展现出很大的潜力。这些新方法主要包括非负矩阵分解法 (Nonnegative Matrix Factorization, NMF)[7−10]、鲁棒主成分分析法 (Robust Principal Component Analysis, RPCA) 等 [11−12]。其中，RPCA 法既不依赖于噪声估计和跟踪，也无需预先训练语音或噪声模型，其完全无监督的特点对于实用化语音降噪非常具有吸引力，因而受到广泛关注。在语音谱是稀疏而噪声谱是低秩的假设条件下，语音信号和干扰噪声通过短时傅里叶变换 (Short-time Fourier Transform, STFT) 在语谱图上展现出良好的可分离特性 [13−15]，通过 RPCA, Godec 等稀疏低秩分解方法即可以实现语音和噪声的相对分离。

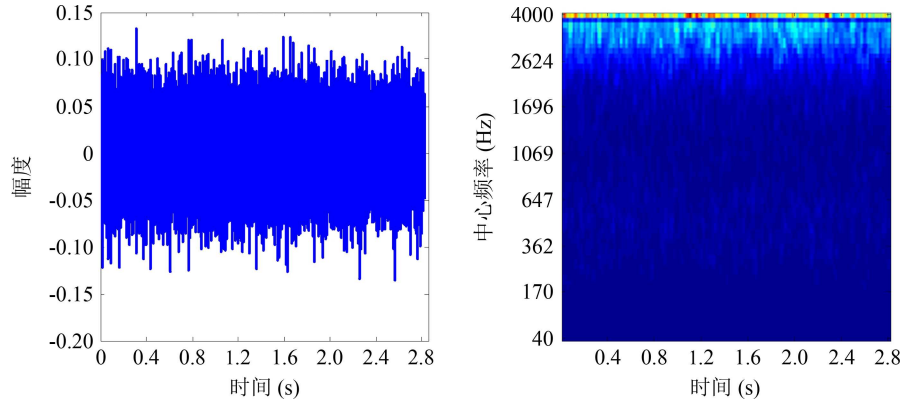
然而，现有稀疏低秩分解语音降噪方法还没有充分考虑人耳的听觉感知特性，其模型也有待进一步完善，主要问题是：一是现有方法通常选用 F 范数距离度量作为优化目标函数，但是该度量倾向于强调大系数的贡献，致使低能量区域的语音失真极易滋生；其次，现有方法由于没添加相关约束，无法保证其稀疏低秩分解结果是非负的，这与语音或噪声时频谱的物理含义不一致。针对上述问题，本文以能够反映人耳主观听觉感受的板仓－斋田 (Itakura-Saito, IS) 距离度量作为优化目标函数。IS 距离度量允许能量较大的语音频带容忍较大的失真，而能量较小的频带容忍较小的失真，这与人耳的听觉掩蔽特性是一致的。此外，在稀疏低秩建模过程中引入非负约束可使分解分量更符合实际物理含义。最后，通过估计理想比值掩蔽模 (Ideal Ratio Mask, IRM) 有效抑制干扰噪声并改善降噪语音质量。

1　语音和噪声的耳蜗谱图表示差异

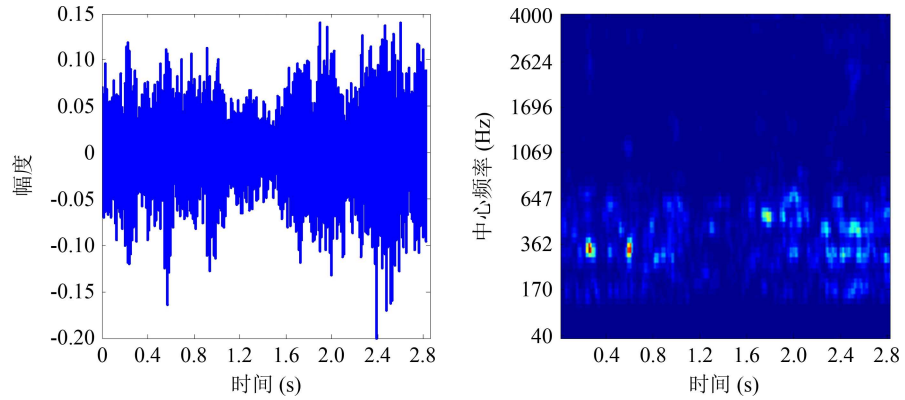
为了模拟耳蜗基底膜对于频率感知的非线性特征，可采用一组伽玛通滤波器对时域信号进行滤波并计算各个通道的短时能量，从而得到具有非均匀时频刻度的能量分布图，即耳蜗谱图 [2]。耳蜗谱图中处于人耳敏感低频区的时频单元相对具有更高的分辨率，这与人耳的听觉感知特性相吻合。在文献 16 中，研究者使用非负矩阵分解通过乘性迭代在耳蜗谱图上将语音从音乐信号中成功分离，并证实了相对于具有均匀时频刻度的语谱图，语音和器乐在耳蜗谱图上具有更强的可分离性。本文将研究应用更广泛的语噪分离问题，以耳蜗谱图作为分离基础，应用语音谱的稀疏特性和噪声谱的低秩特性通过提出的板仓－斋田非负鲁棒主成分分析 (Itakura-Saito Nonnegative RPCA, ISNRPCA) 模型进行无监督语音降噪。相对于器乐信号，噪声信号类型更多样，部分类型噪声时频谱的低秩特性并不十分显著，因此需要引入额外的误差项来补偿稀疏低秩建模的不足。如图 1 所示，语音信号和高斯白噪声、babble 噪声等典型背景噪声在耳蜗谱图上差异显著。在耳蜗谱图上，语音信号的能量集中在少数的时频单元，表现出显著的稀疏特性；然而噪声信号在耳蜗谱图上则呈现出一些相似的频谱结构和模式，用少量的基向量就可有效表示，因此耳蜗谱图上的噪声信号处在



(a)　语音信号



(b) 白噪声



(c) Babble 噪声

图 1 语音信号和典型噪声信号的时域波形和耳蜗谱图示意图 (左侧子图是时域波形，右侧子图是耳蜗谱图)

一个低秩的子空间中。尽管一些相对复杂多变的噪声信号并不遵循低秩模型假设，但可近似地将其建模成低秩分量和稠密误差分量之和。因此，应用稀疏低秩分解方法有望在耳蜗谱图上将语音信号和干扰噪声有效分离，从而达到语音降噪的目的。

2 板仓–斋田非负鲁棒主成分分析

2.1 问题描述

考虑纯净语音信号 $s(t)$ 受到加性背景噪声信号 $n(t)$ 的干扰，则含噪语音信号 $y(t)$ 为：

$$y(t) = s(t) + n(t). \tag{1}$$

通过时频变换，可得到 $y(t)$ 在语谱图或耳蜗谱图上的表示 $\boldsymbol{Y} \in \boldsymbol{R}^{m\times n}$。通常假设语音信号和噪声信号在统计意义下互不相关，此时 $\boldsymbol{Y}$ 可表示成 3 个分量之和：语音分量 $\boldsymbol{S}$、噪声分量 $\boldsymbol{L}$ 以及稠密误差分量 $\boldsymbol{E}$[14−15]，

$$\boldsymbol{Y} = \boldsymbol{S} + \boldsymbol{L} + \boldsymbol{E}. \tag{2}$$

为了能够从 $\boldsymbol{Y}$ 中分离出感兴趣的 $\boldsymbol{S}$ 和 $\boldsymbol{L}$，在语音谱是稀疏而噪声谱是低秩的假设条件下通常选择 F 范数距离作为优化目标函数，可以得到：

$$\begin{aligned} &\arg\min_{\boldsymbol{L},\boldsymbol{S}} \ \|\boldsymbol{Y}-\boldsymbol{S}-\boldsymbol{L}\|_F^2, \\ &\text{subject to} \ \ \operatorname{rank}(\boldsymbol{L}) \leqslant r_L, \ \operatorname{card}(\boldsymbol{S}) \leqslant c_s, \end{aligned} \tag{3}$$

其中，$\|\cdot\|_F$ 表示矩阵的 F 范数，r_L 表示 $\boldsymbol{L}$ 的秩，c_s 表示 $\boldsymbol{S}$ 的势即支撑集中分量的个数。然而，主观听觉实验表明 IS 距离度量与人耳的听觉感受更为一致，因此常被用作语音失真的度量准则。IS 距离度量定义为：

$$\mathrm{IS}(y,x) = \frac{y}{x} - \log\frac{y}{x} - 1, \tag{4}$$

当变量是矩阵时，IS 距离度量为：

$$\mathrm{IS}\left(\boldsymbol{Y},\boldsymbol{X}\right) = \sum_{j=1}^{n}\sum_{i=1}^{m}\mathrm{IS}\left(y_{ij},x_{ij}\right). \tag{5}$$

本文提出的板仓–斋田非负鲁棒主成分分析法，旨在通过优化迭代使得含噪语音谱 $\boldsymbol{Y}$ 和语音分量 $\boldsymbol{S}$ 与噪声分量 $\boldsymbol{L}$ 之和的 IS 距离达到最小。与传统 RPCA 法不同，这里 $\boldsymbol{Y}$, $\boldsymbol{L}$ 和 $\boldsymbol{S}$ 均须添加非负约束以保证和语谱图或耳蜗谱图的物理意义相吻合。因此，ISNRPCA 模型可描述为：

$$\begin{aligned} &\arg\min_{\boldsymbol{L},\boldsymbol{S}} \ \mathrm{IS}(\boldsymbol{Y},\boldsymbol{L}+\boldsymbol{S}) + \lambda\|\boldsymbol{S}\|_1 + \beta\|\boldsymbol{L}\|_*, \\ &\text{subject to} \ \ \boldsymbol{L} \geqslant 0, \ \boldsymbol{S} \geqslant 0, \end{aligned} \tag{6}$$

其中，$\|\cdot\|_1$ 和 $\|\cdot\|_*$ 分别表示矩阵的 ℓ_1 范数与核范数；λ 和 β 为正则化参数，用于均衡 3 个分量贡献的大小。

从式 (4) 可以看出，IS 距离度量是尺度不变的，这使得该距离度量下大系数和小系数的贡献同等重要，从而与人耳听觉系统相吻合。因此，ISNRPCA 可被认为是一种受听觉感知启发的鲁棒主成分分析法。下节我们将描述如何通过交替方向乘子法 (Alternating Direction Method of Multipliers, ADMM) 求解式 (6) 中的 ISNRPCA 优化问题。

2.2 ISNRPCA 的优化求解

引入辅助变量 $\boldsymbol{X}$, $\boldsymbol{S}_+$ 和 $\boldsymbol{L}_+$, 式 (6) 可改写为：

$$\begin{aligned}&\arg\min_{\boldsymbol{L},\boldsymbol{S},\boldsymbol{L}_+,\boldsymbol{S}_+,\boldsymbol{X}} \ \mathrm{IS}(\boldsymbol{Y},\boldsymbol{X})+\lambda\|\boldsymbol{S}_+\|_1+\beta\|\boldsymbol{L}_+\|_*,\\ &\text{subject to} \quad \boldsymbol{X}=\boldsymbol{L}+\boldsymbol{S},\\ &\qquad\qquad \boldsymbol{S}_+=\boldsymbol{S},\ \boldsymbol{L}_+=\boldsymbol{L},\\ &\qquad\qquad \boldsymbol{L}_+\geqslant 0,\ \boldsymbol{S}_+\geqslant 0.\end{aligned} \tag{7}$$

式 (7) 的增广拉格朗日函数可写成：

$$\begin{aligned}&L_\rho(\boldsymbol{X},\boldsymbol{L},\boldsymbol{S},\boldsymbol{L}_+,\boldsymbol{S}_+,\boldsymbol{\Omega}_X,\boldsymbol{\Omega}_S,\boldsymbol{\Omega}_L)=\mathrm{IS}(\boldsymbol{Y},\boldsymbol{X})+\\ &\frac{\rho}{2}\|\boldsymbol{X}-\boldsymbol{L}-\boldsymbol{S}+\boldsymbol{\Omega}_X\|_F^2+\frac{\rho}{2}\|\boldsymbol{S}-\boldsymbol{S}_++\boldsymbol{\Omega}_S\|_F^2+\\ &\frac{\rho}{2}\|\boldsymbol{L}-\boldsymbol{L}_++\boldsymbol{\Omega}_L\|_F^2+\lambda\|\boldsymbol{S}_+\|_1+\beta\|\boldsymbol{L}_+\|_*,\end{aligned} \tag{8}$$

其中，$\boldsymbol{\Omega}_X$, $\boldsymbol{\Omega}_S$ 和 $\boldsymbol{\Omega}_L$ 分别表示 $\boldsymbol{X}$, $\boldsymbol{L}$ 和 $\boldsymbol{S}$ 的归一化对偶变量，ρ 是归一化参数用于控制算法的收敛速率。由于式 (8) 中的目标函数是可分离的且每个子问题都是相对易求解的优化问题，故可采用 ADMM 算法对其求解 [17]。

在 ADMM 框架下具体求解式 (8) 时，所有待求解变量均通过求解对应的子问题依次交替优化。通过对两个主变量 $\boldsymbol{S}$ 和 $\boldsymbol{L}$, 3 个对偶变量 $\boldsymbol{\Omega}_X$, $\boldsymbol{\Omega}_S$ 和 $\boldsymbol{\Omega}_L$ 以及 3 个辅助变量 $\boldsymbol{X}$, $\boldsymbol{S}_+$ 和 $\boldsymbol{L}_+$ 分别求偏导，即可按照梯度下降法进行迭代优化。以主变量 $\boldsymbol{S}$ 和 $\boldsymbol{L}$ 为例，

$$\begin{aligned}\boldsymbol{S}=\arg\min_{\boldsymbol{S}}\ &\frac{\rho}{2}\|\boldsymbol{X}-\boldsymbol{L}-\boldsymbol{S}+\boldsymbol{\Omega}_X\|_F^2+\\ &\frac{\rho}{2}\|\boldsymbol{S}-\boldsymbol{S}_++\boldsymbol{\Omega}_S\|_F^2,\end{aligned} \tag{9}$$

$$\begin{aligned}\boldsymbol{L}=\arg\min_{\boldsymbol{L}}\ &\frac{\rho}{2}\|\boldsymbol{X}-\boldsymbol{L}-\boldsymbol{S}+\boldsymbol{\Omega}_X\|_F^2+\\ &\frac{\rho}{2}\|\boldsymbol{L}-\boldsymbol{L}_++\boldsymbol{\Omega}_L\|_F^2.\end{aligned} \tag{10}$$

对 $\boldsymbol{S}$ 和 $\boldsymbol{L}$ 分别求偏导，可得：

$$\boldsymbol{S}=(\boldsymbol{X}-\boldsymbol{L}+\boldsymbol{\Omega}_X-\boldsymbol{\Omega}_S+\boldsymbol{S}_+)/2, \tag{11}$$

$$\boldsymbol{L}=(\boldsymbol{X}-\boldsymbol{S}+\boldsymbol{\Omega}_X-\boldsymbol{\Omega}_L+\boldsymbol{L}_+)/2. \tag{12}$$

用于更新辅助变量 $\boldsymbol{X}$ 的子问题为：

$$\boldsymbol{X}=\arg\min_{\boldsymbol{X}\geqslant 0}\ \mathrm{IS}(\boldsymbol{Y},\boldsymbol{X})+\frac{\rho}{2}\|\boldsymbol{X}-\boldsymbol{L}-\boldsymbol{S}+\boldsymbol{\Omega}_X\|_F^2. \tag{13}$$

文献 18 中的定理 2 对类似式 (13) 的子问题已经求解，应用该结果并作相应改造即可得到 $\boldsymbol{X}$ 的更新过程为：

$$\begin{aligned}&\boldsymbol{A}=\boldsymbol{\Omega}_X-\boldsymbol{L}-\boldsymbol{S},\ B_{ij}=\frac{1}{3\rho}-\frac{A_{ij}^2}{9},\\ &C_{ij}=-\frac{A_{ij}^3}{27}+\frac{A_{ij}}{6\rho}+\frac{Y_{ij}}{2\rho},\ D_{ij}=B_{ij}^3+C_{ij}^2,\\ &V_{ij}=\begin{cases}\left(C_{ij}+\sqrt{D_{ij}}\right)^{1/3}+\left(C_{ij}-\sqrt{D_{ij}}\right)^{1/3}, & D_{ij}\geqslant 0,\\ 2\sqrt{-B_{ij}}\cos\left(\dfrac{1}{3}\cos^{-1}\left(\dfrac{C_{ij}}{\sqrt{-B_{ij}^3}}\right)\right), & D_{ij}<0,\end{cases}\\ &X_{ij}=V_{ij}-\frac{A_{ij}}{3}.\end{aligned} \tag{14}$$

因此，在 ADMM 框架下求解 ISNRPCA 问题时可总结如算法 1 所示。其中，$S_\lambda(\cdot)$ 表示软门限算子，定义为：

$$S_\lambda(x)=\begin{cases}x-\lambda, & x\geqslant\lambda,\\ 0, & -\lambda<x<\lambda,\\ x+\lambda, & x\leqslant-\lambda,\end{cases} \tag{15}$$

$S_{+\lambda}(\cdot)$ 表示经过 $S_\lambda(\cdot)$ 运算后再取其正的部分。

算法 1: ISNRPCA 问题的 ADMM 优化算法

(1) 输入：$\boldsymbol{Y}$
(2) 输出：$(\boldsymbol{S},\boldsymbol{L})$ 的估计
(3) 初始化：$k=0$, $M=300$, $\rho=1$, random($\boldsymbol{L}^{(0)}$), random($\boldsymbol{S}^{(0)}$), $\boldsymbol{L}_+^{(0)}=\boldsymbol{S}_+^{(0)}=\boldsymbol{0}$, $\boldsymbol{\Omega}_X^{(0)}=\boldsymbol{\Omega}_S^{(0)}=\boldsymbol{\Omega}_L^{(0)}=\boldsymbol{0}$
(4) 当 $k\leqslant M$ 时进行循环
(5) // 更新辅助变量，$\boldsymbol{X}$:
(6) $\boldsymbol{X}^{(k+1)}=\arg\min_{\boldsymbol{X}\geqslant 0}\ \mathrm{IS}(\boldsymbol{Y},\boldsymbol{X})+\frac{\rho}{2}\left\|\boldsymbol{X}-\boldsymbol{L}^{(k)}-\boldsymbol{S}^{(k)}+\boldsymbol{\Omega}_X^{(k)}\right\|_F^2$
(7) // 更新主变量，$\boldsymbol{S}$ 和 $\boldsymbol{L}$:
(8) $\boldsymbol{S}^{(k+1)}=(\boldsymbol{X}^{(k+1)}-\boldsymbol{L}^{(k)}+\boldsymbol{\Omega}_X^{(k)}-\boldsymbol{\Omega}_S^{(k)}+\boldsymbol{S}_+^{(k)})/2$
(9) $\boldsymbol{L}^{(k+1)}=(\boldsymbol{X}^{(k+1)}-\boldsymbol{S}^{(k+1)}+\boldsymbol{\Omega}_X^{(k)}-\boldsymbol{\Omega}_L^{(k)}+\boldsymbol{L}_+^{(k)})/2$
(10) // 更新辅助变量，$\boldsymbol{S}_+$ 和 $\boldsymbol{L}_+$:
(11) $\boldsymbol{S}_+^{(k+1)}=\boldsymbol{S}_{+\lambda/\rho}(\boldsymbol{S}^{(k+1)}+\boldsymbol{\Omega}_S^{(k)})$
(12) $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}=\mathrm{svd}(\boldsymbol{L}^{(k+1)}+\boldsymbol{\Omega}_L^{(k)})$; $\boldsymbol{L}_+^{(k+1)}=\boldsymbol{U}\boldsymbol{S}_{+\beta/\rho}(\boldsymbol{\Sigma})\boldsymbol{V}$
(13) // 更新对偶变量，$\boldsymbol{\Omega}_X$, $\boldsymbol{\Omega}_S$ 和 $\boldsymbol{\Omega}_L$
(14) $\boldsymbol{\Omega}_X^{(k+1)}=\boldsymbol{\Omega}_X^{(k)}+(\boldsymbol{X}^{(k+1)}-\boldsymbol{L}^{(k+1)}-\boldsymbol{S}^{(k+1)})$
(15) $\boldsymbol{\Omega}_L^{(k+1)}=\boldsymbol{\Omega}_L^{(k)}+(\boldsymbol{L}^{(k+1)}-\boldsymbol{L}_+^{(k+1)})$
(16) $\boldsymbol{\Omega}_S^{(k+1)}=\boldsymbol{\Omega}_S^{(k)}+(\boldsymbol{S}^{(k+1)}-\boldsymbol{S}_+^{(k+1)})$
(17) $k=k+1$
(18) 循环结束
(19) $\widehat{\boldsymbol{S}}=\boldsymbol{S}_+^{(k)}$, $\widehat{\boldsymbol{L}}=\boldsymbol{L}_+^{(k)}$
(20) 输出 $(\widehat{\boldsymbol{S}},\widehat{\boldsymbol{L}})$

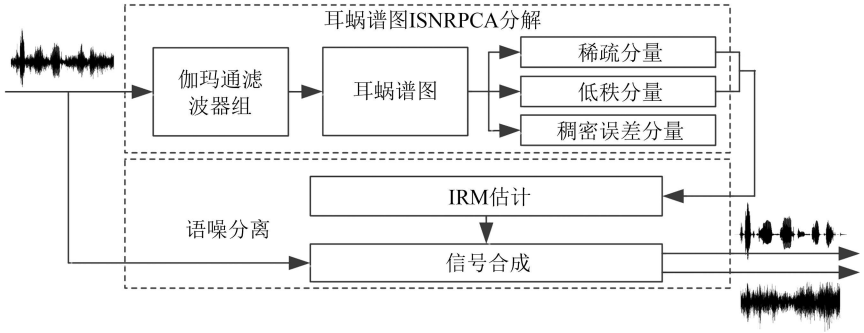


图 2 基于 ISNRPCA 的语音降噪过程示意图

3 基于 ISNRPCA 的无监督语音降噪

应用 ISNRPCA 算法进行语音降噪的主要过程如图 2 所示，可分为耳蜗谱图上进行 ISNRPCA 分解和语噪分离两个阶段。首先，采用伽玛通滤波器组对输入的含噪语音进行耳蜗分析，并计算含噪语音的耳蜗谱图表示 $\boldsymbol{Y}$；其次，采用 ISNRPCA 算法对 $\boldsymbol{Y}$ 进行稀疏低秩分解，得到稀疏语音分量的估计 $\widehat{\boldsymbol{S}}$，低秩噪声分量的估计 $\widehat{\boldsymbol{L}}$ 以及稠密误差分量的估计 $\widehat{\boldsymbol{E}}$。在听觉场景分析中为了提高分离语音的总体质量，对理想二值掩蔽模 (Ideal Binary Mask, IBM) 的估计可推广至对 IRM 的估计 [19−20]，

$$\boldsymbol{R}=[R_{ij}]=\left[\frac{\widehat{S}_{ij}}{\widehat{S}_{ij}+\widehat{L}_{ij}}\right]. \tag{16}$$

然而，IRM 不可避免存在估计误差，这会造成频谱的不连续并产生音乐噪声问题。因此，为了提升降噪语音的主观听觉感受，可以对估计的 IRM 进行一阶平滑，

$$R_{ij}=\alpha R_{ij}+(1-\alpha)\,R_{ij-1}. \tag{17}$$

通过大量仿真实验，α 可取经验值 0.5。在得到 IRM 的估计 $\boldsymbol{R}$ 后，采用比值掩蔽的方法对含噪语音的耳蜗谱图进行加权并得到降噪语音的耳蜗谱图表示 $\widetilde{\boldsymbol{S}}$，如式 (18) 所示。其中，$\odot$ 表示点乘即矩阵对应元素相乘。最后，结合伽玛通滤波器组的相位时延分别合成出语音和噪声信号 [2]。

$$\widetilde{\boldsymbol{S}}=\boldsymbol{Y}\odot\boldsymbol{R}. \tag{18}$$

4 仿真实验及结果分析

本节将在 6 种混合信噪比 (Signal to Noise Ratios, SNRs)、5 种不同类型噪声条件下对 ISNRPCA 算法的语音降噪性能进行评估。所有仿真实验均在硬件配置为 Intel 双核 2.93 GHz CPU、3 G 内存的个人计算机上运行 matlab 2013b 软件完成。

4.1 实验数据及评估方法

在仿真实验中，选择 NOIZEUS 语音库中的纯净语音和 Noisex-92 噪声库中的噪声作为实验数据 [21−22]，所有语音信号和噪声信号均被下降采样率至 8 kHz。具体地，测试纯净语音包含 3 名男性说话人和 3 名女性说话人，共 30 条语句，每条语句时长约 3 s。测试噪声包含平稳噪声和非平稳噪声两大类，其中有：F16, hf-channel, white, babble 和 factory 噪声。将纯净语音与噪声按照不同大小 SNRs 进行混合，具体为 −10 dB, −5 dB, 0 dB, 5 dB, 10 dB 和 15 dB。

为了客观评估 ISNRPCA 等不同算法的语音降噪效果，本文采用不同类型准则评估其性能。第 1 种是使用 BSS EVAL package V3.0 计算信号失真比 (Signal to Distortion Ratio, SDR)[23]，该准则在盲源分离中广泛使用；第 2 种是采用 ITU-T P.862 标准即语音质量感知评估 (Perceptual Evaluation of Speech Quality, PESQ) 和 COMPOSITE(COSI) 模型来评估降噪语音的总体质量 [24−25]；第 3 种是采用短时客观可懂度得分 (Short-time Objective Intelligibility, STOI) 评估降噪语音的可懂度 [26]。SDR 用于衡量语音和噪声的分离性能，表现为对噪声的抑制能力；PESQ, COSI 和 STOI 得分衡量降噪语音的质量。具体应用时，SDR 分值越大表示语音降噪算法对噪声的抑制能力越强，PESQ, COSI 和 STOI 得分越高表示降噪语音质量越好。

4.2 ISNRPCA 算法正则化参数设置

在 RPCA 这类算法中，正则化参数的选取是一个重要环节，可在理论的指导下结合大量实验取适当的经验值。在 RPCA 算法中，Candès 等人证明了在绝大多数应用场合，λ 取典型值 $(\max(m,n))^{-1/2}$ 时即可得到理想的稀疏和低秩分离效果。对于本文的

ISNRPCA 算法，正则化参数 λ 和 β 可在其基础上进行微调，本文设定 λ 的取值为 $ab(\max(m,n))^{-1/2}$。其中，当 SNRs 较低时，$a=1$，否则 $a=0.94$；处理低频带耳蜗谱图时 $b=1.15$，处理高频带耳蜗谱图时 $b=0.95$。由于 ISNRPCA 建模过程与稳健主成分基追踪法 (Stable Principal Component Pursuit, SPCP) 一致 [27]，受 SPCP 问题中正则化参数选择启发，我们设定 $\lambda\beta=c$，其中 c 为经验常数。也就是当稀疏分量权重增加时，低秩分量权重按倒数减小。需要说明的是，语音降噪算法的噪声抑制能力和降噪语音质量之间往往是矛盾的，对噪声的抑制能力越强时也常会对语音带来较大的损伤，低 SNRs 条件下语音降噪和高 SNRs 条件下语音降噪也需要均衡。为了合理确定经验常数 c 的取值，分别计算 5 种类型噪声、6 种 SNRs 条件下降噪语音的平均 PESQ, COSI, SDR 和 STOI 得分，结果如图 3 所示。可以看出，当 $c=0.16$ 时 PESQ 平均得分最高，$c=0.175$ 时 COSI 平均得分最高，SDR 评估和 STOI 评估趋势完全相反。因此，为了在不同指标之间取得均衡，可经验性地设定 $c=0.165$。在后续的性能评估实验中，将会看到本文正则化参数设置对于不同种类噪声和不同高低 SNRs 具有较好的适应性。

4.3 ISNRPCA 算法收敛性

尽作者们所知，当 RPCA 的分解分量多于 3 个时，算法全局收敛性的理论证明目前还未被证实。但在实际应用中，通过 300 次迭代 ISNRPCA 算法足以收敛至一个统计不动点，且与优化变量的初始值选取无关。定义第 k 次迭代的相对误差如式 (19) 所示，该指标常被用作描述算法的收敛特性。此外，也可用 $\boldsymbol{Y}$ 和分解得到的 $\boldsymbol{S}$ 与 $\boldsymbol{L}$ 之和的 IS 距离，也就是 $\mathrm{IS}(\boldsymbol{Y},\boldsymbol{L}+\boldsymbol{S})$ 来描述算法的收敛情况。

$$\mathrm{RelErr}(k)=\frac{\left\|\boldsymbol{S}^{(k)}-\boldsymbol{S}^{(k-1)}\right\|_F^2}{\left\|\boldsymbol{S}^{(k-1)}\right\|_F^2}+\frac{\left\|\boldsymbol{L}^{(k)}-\boldsymbol{L}^{(k-1)}\right\|_F^2}{\left\|\boldsymbol{L}^{(k-1)}\right\|_F^2}. \tag{19}$$

图 4 给出了应用 ISNRPCA 算法对一段典型语音进行降噪时的收敛曲线。可以看出，迭代 300 次后 ISNRPCA 算法已经较好地收敛，能够稳定地估计出语音分量 $\boldsymbol{S}$ 和噪声分量 $\boldsymbol{L}$。

4.4 算法性能分析

本节将进一步采用 SDR, PESQ, COSI 和 STOI 4 种准则详细评估 ISNRPCA 算法、近期报道的同类算法和经典语音降噪算法的性能。参考算法为语谱图上稀疏低秩分解法 (SLTF)[15]，鲁棒主成分分析法 (RPCA)[12]；经典算法为多带谱减法 (MBSS)[5]。具体实现时，RPCA 采用增广拉格朗日乘子算法 (程序网址：http://perception.csl.illinois.edu/matrix-rank/sample_code.html)，SLTF 采用 Semi-soft Godec 算法 (程序网址：http://sites.google.com/site/godecomposition/)。设置正则化参数时，RPCA 算法中的 λ 取典型值 $(\max(m,n))^{-1/2}$；根据文献 15 中作者的建议，

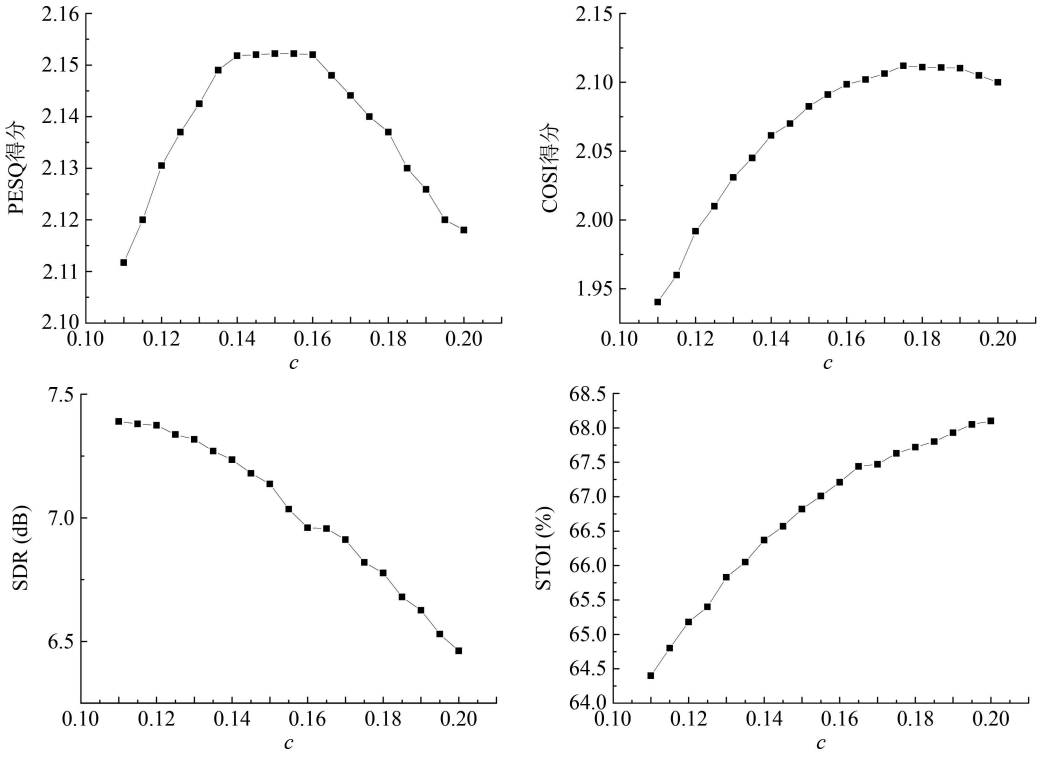


图 3　c 取不同数值条件下降噪语音的 PESQ, COSI, SDR 和 STOI 平均得分

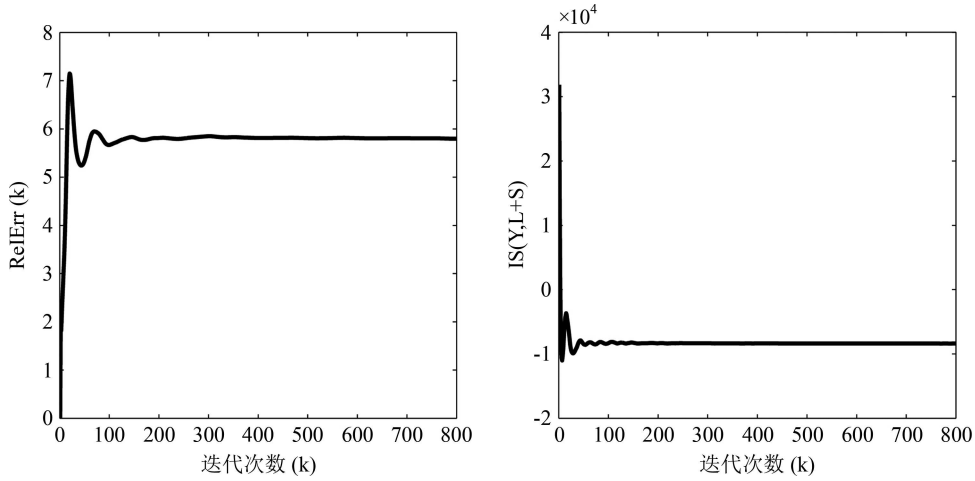


图 4 0 dB babble 噪声条件下，应用 ISNRPCA 算法对语句 “The birch canoe slid on the smooth planks (NOIZEUS 数据库中的 sp01 语句)” 进行降噪时的算法收敛性示意图

SLTF 算法中设定 $r_L=2, \lambda=0.05$。令 L 表示用于语音分帧时的窗长，R 表示帧移，N 表示 DFT 的窗长，单位都为样点数，f 表示频率范围，K 表示伽玛通滤波器的数目。通过时频变换得到语谱图和耳蜗谱图时，上述参数的设置如表 1 所示。为了便于读者从事相关研究，我们提供了 ISNRPCA 算法的具体实现 (程序网址：https://cn.mathworks.com/matlabcentral/fileexchange/50427-isnrpca)。

表 1 语谱图和耳蜗谱图参数设置

类型	L	R	N	K	f (Hz)
语谱图	256	128	256	—	—
耳蜗谱图	256	128	—	64	40～4000

首先分析语谱图和耳蜗谱图上 ISNRPCA 算法的性能差异。图 5 给出了传统语谱图和考虑了人耳听觉感知特性的耳蜗谱图上进行语噪分离的结果，可以看出在测试的所有 SNRs 条件下，耳蜗谱图上进行语噪分离的平均 SDR 得分均优于语谱图上语噪分离的结果，这说明 ISNRPCA 算法在耳蜗谱图上可将语音和噪声更有效分离，从而实现对噪声的有效抑制；在描述降噪语音质量的平均 PESQ 得分方面，除了极低 −10 dB 外，其它 SNRs 条件下耳蜗谱图上分离的得分也较高，这说明耳蜗谱图上降噪语音质量更好。因此，选择耳蜗谱图进行语噪分离不仅符合直观分析，仿真实验结果也表现出优势。

然后评估 IRM 平滑对 ISNRPCA 算法语音降噪的影响。IRM 估计不可避免存在误差从而造成降噪语音频谱的不连续，采用 IRM 平滑的方法能够补偿由此造成的性能损失并抑制音乐噪声。从图 6 可以看出，在所有测试 SNRs 条件下加入 IRM 平滑后能够有效改善降噪语音质量，PESQ 平均得分均得到明显提高；但中高 SNRs 条件下随着 IRM 估计误差的减小，IRM 平滑的负面影响开始显现：噪声抑制性能有所损失，表现在 SDR 平均得分开始下降。由于 IRM 平滑对提高降噪语音质量的作用十分明显，因此在后续的实验中将继续采用该项技术。

进一步对比 ISNPRCA 算法与参考算法的噪声抑制性能，表 2 给出了不同噪声类型条件下各种算

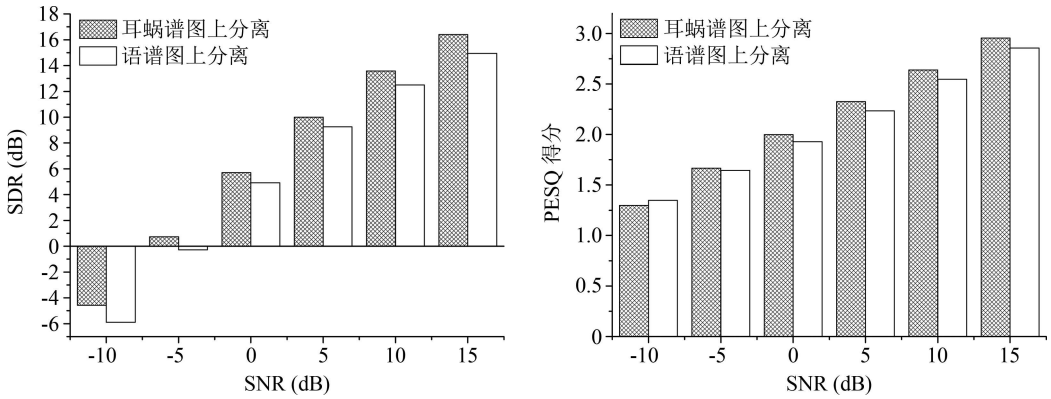


图 5 语谱图和耳蜗谱图上 ISNRPCA 算法的 SDR 和 PESQ 得分对比

法的 SDR 评估结果。可以看出，相对于其它噪声类型， babble 噪声条件下的 SDR 分值相对较低，这说明 ISNRPCA 算法去除 babble 噪声的能力相对较弱。这是由于 babble 噪声是与语音最为接近的人群噪声，结构性分量较少且低秩特性不明显。但是，在几乎所有噪声类型和 SNRs 条件下本文 ISNPRCA 算法均获得了最高的 SDR 分值，优于 RPCA 算法，显著优于 SLTF 算法和 MBSS 算法，且优势在 SNRs 高于 10 dB 的高信噪比条件下更为明显。因此，在所有评估算法中 ISNRPCA 算法的语噪分离能力最强，对噪声的抑制效果最好。

最后对比 ISNPRCA 算法与参考算法的降噪语音质量情况。在反映降噪语音总体质量的 PESQ 和 COSI 得分、可懂度的 STOI 得分方面，从表 3 — 表 5 可以看出 ISNRPCA 算法也更具竞争力，与参考算法相比得分更高或至少相当。在 −10 dB 极低 SNRs 条件下虽然 ISNRPCA 算法的 STOI 得分略低于 SLTF 算法和 RPCA 算法，但其它 SNRs 条件下 INSRPCA 算法的 STOI 得分最高，且在高 SNRs 条件下优势明显。反映语音总体质量的 PESQ 和 COSI 评估结果与可懂度评估结果相似：在设置的所有 SNRs 条件下，ISNRPCA 算法几乎都给出最高的 PESQ 和 COSI 得

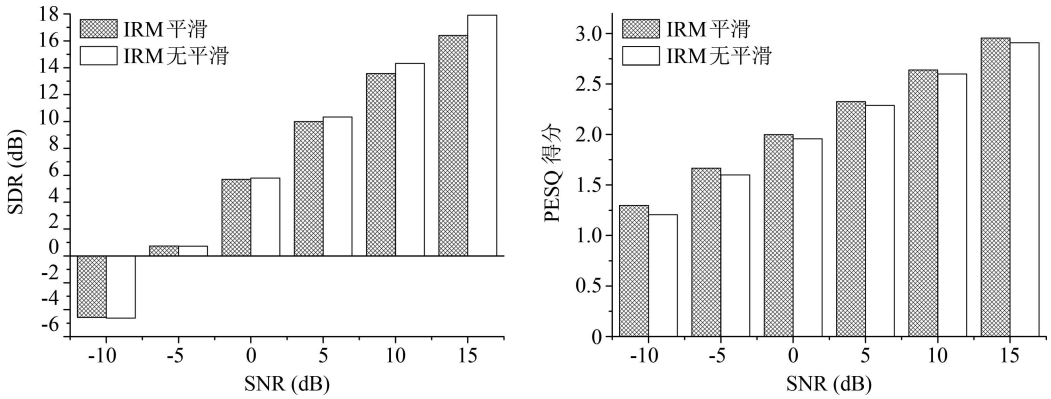


图 6　是否采用 IRM 平滑条件下 ISNRPCA 算法的 SDR 和 PESQ 得分对比

表 2　不同算法 SDR 分值对比

噪声类型	算法	SNRs (dB)					
		−10	−5	0	5	10	15
Babble	MBSS	−8.86	−2.71	2.79	7.27	11.5	15.2
	SLTF	−6.61	−1.59	3.11	7.12	9.52	10.7
	RPCA	−7.46	−1.84	3.57	8.24	11.6	13.5
	ISNRPCA	**−6.54**	**−1.08**	**4.28**	**9.11**	**13.0**	**16.2**
Factory	MBSS	−7.50	−2.26	2.96	7.00	10.8	14.5
	SLTF	−5.28	−0.23	4.70	8.56	10.1	11.0
	RPCA	−5.69	−0.04	5.14	9.43	12.3	13.9
	ISNRPCA	**−4.50**	**1.02**	**6.12**	**10.3**	**14.0**	**16.7**
F16	MBSS	−7.12	−1.56	3.51	7.55	10.8	14.9
	SLTF	−4.87	0.48	4.79	8.64	10.3	11.1
	RPCA	−5.38	0.20	5.35	9.56	12.4	13.9
	ISNRPCA	**−4.59**	**1.11**	**6.16**	**10.4**	**13.9**	**16.6**
Hf-channel	MBSS	−6.14	−1.39	2.90	6.45	10.1	13.9
	SLTF	−5.40	0.18	4.88	8.97	10.5	11.2
	RPCA	−4.68	1.07	6.13	**10.1**	12.6	14.0
	ISNRPCA	**−3.54**	**1.48**	**6.15**	**10.1**	**13.5**	**16.2**
White	MBSS	−6.20	−1.34	2.37	6.04	8.92	13.0
	SLTF	−5.18	0.31	5.51	9.81	11.3	11.4
	RPCA	−4.16	**1.59**	**6.58**	**10.4**	12.8	14.1
	ISNRPCA	**−3.64**	1.15	5.76	9.88	**13.2**	**16.0**

表 3 不同算法 PESQ 得分对比

噪声类型	算法	SNRs (dB)					
		−10	−5	0	5	10	15
Babble	MBSS	0.85	1.44	1.90	2.22	2.54	2.81
	SLTF	**1.09**	**1.62**	1.92	2.22	2.40	2.55
	RPCA	1.05	1.61	1.99	2.31	2.58	2.81
	ISNRPCA	1.02	**1.62**	**2.00**	**2.35**	**2.66**	**2.96**
Factory	MBSS	1.09	1.40	1.79	2.15	2.48	2.78
	SLTF	1.32	1.65	2.00	2.31	2.50	2.67
	RPCA	1.31	1.68	2.02	2.34	2.61	2.83
	ISNRPCA	**1.33**	**1.69**	**2.04**	**2.37**	**2.69**	**3.01**
F16	MBSS	1.21	1.52	1.86	2.22	2.70	2.86
	SLTF	1.36	1.73	2.05	2.34	2.26	2.69
	RPCA	1.37	1.73	2.07	2.38	2.66	2.86
	ISNRPCA	**1.49**	**1.79**	**2.10**	**2.40**	**2.79**	**3.01**
Hf-channel	MBSS	1.17	1.32	1.59	1.98	2.37	2.71
	SLTF	1.32	1.54	1.79	2.10	2.39	2.64
	RPCA	**1.38**	1.56	1.83	2.13	2.43	2.69
	ISNRPCA	1.32	**1.60**	**1.88**	**2.20**	**2.53**	**2.87**
White	MBSS	1.14	1.31	1.58	1.95	2.30	2.66
	SLTF	1.29	1.53	1.81	2.16	2.46	2.71
	RPCA	1.31	1.55	1.84	2.15	2.44	2.70
	ISNRPCA	**1.32**	**1.62**	**1.92**	**2.28**	**2.60**	**2.91**

表 4 不同算法 COSI 得分对比

噪声类型	算法	SNRs (dB)					
		−10	−5	0	5	10	15
Babble	MBSS	**1.03**	**1.39**	1.92	**2.40**	2.85	3.23
	SLTF	1.00	1.25	1.61	1.96	2.22	2.47
	RPCA	1.00	1.32	1.82	2.29	2.69	3.02
	ISNRPCA	**1.03**	**1.39**	**1.93**	**2.40**	**2.86**	**3.28**
Factory	MBSS	1.04	1.30	1.75	2.25	2.70	3.11
	SLTF	**1.07**	1.35	1.77	2.10	2.26	2.50
	RPCA	1.04	1.34	1.82	2.27	2.66	2.98
	ISNRPCA	1.04	**1.36**	**1.87**	**2.35**	**2.79**	**3.20**
F16	MBSS	**1.10**	1.35	1.82	2.32	2.77	3.13
	SLTF	**1.10**	**1.42**	1.83	2.17	2.38	2.52
	RPCA	1.06	1.38	1.87	2.32	2.70	3.03
	ISNRPCA	**1.10**	**1.42**	**1.88**	**2.35**	**2.79**	**3.19**
Hf-channel	MBSS	**1.18**	1.39	1.79	2.30	2.79	3.21
	SLTF	1.13	1.39	1.73	2.13	2.45	2.70
	RPCA	1.14	1.40	1.83	2.27	2.68	3.03
	ISNRPCA	1.14	**1.49**	**1.95**	**2.41**	**2.88**	**3.30**
White	MBSS	**1.02**	1.15	1.47	1.99	2.47	2.96
	SLTF	**1.02**	1.16	1.53	2.01	2.39	2.65
	RPCA	1.01	1.17	1.55	2.01	2.44	2.81
	ISNRPCA	**1.02**	**1.20**	**1.64**	**2.13**	**2.58**	**2.98**

表 5　不同算法 STOI 得分对比 (%)

噪声类型	算法	SNRs (dB)					
		−10	−5	0	5	10	15
Babble	MBSS	38.3	48.8	60.0	71.0	80.4	86.7
	SLTF	**40.3**	50.3	60.8	69.7	75.7	79.5
	RPCA	39.8	50.3	61.2	70.8	77.9	82.5
	ISNRPCA	40.0	**51.0**	**62.5**	**72.8**	**81.4**	**87.1**
Factory	MBSS	39.2	47.9	58.0	68.3	77.3	85.3
	SLTF	**42.3**	52.2	62.2	70.7	77.2	81.6
	RPCA	42.0	52.3	62.5	71.2	77.9	82.8
	ISNRPCA	42.2	**52.9**	**63.6**	**73.1**	**81.3**	**87.2**
F16	MBSS	43.1	51.2	60.5	70.1	78.8	85.9
	SLTF	**43.3**	53.3	62.8	71.6	77.5	80.0
	RPCA	42.9	53.0	63.3	72.0	78.6	83.0
	ISNRPCA	43.2	**53.5**	**64.3**	**73.6**	**81.6**	**87.1**
Hf-channel	MBSS	43.0	50.7	59.7	69.3	78.7	86.8
	SLTF	**43.4**	54.3	64.9	74.4	79.4	83.7
	RPCA	43.3	54.5	65.4	74.4	80.9	85.2
	ISNRPCA	42.8	**54.9**	**66.9**	**77.5**	**86.1**	**91.1**
White	MBSS	41.9	48.9	56.6	64.8	73.1	82.3
	SLTF	**44.1**	53.0	62.2	71.0	77.2	81.5
	RPCA	**44.1**	53.4	62.7	70.9	77.3	82.0
	ISNRPCA	43.4	**53.5**	**63.7**	**73.1**	**80.5**	**86.8**

分，低 SNRs 条件下 PESQ 和 COSI 评估优势微弱，但高 SNRs 条件下优势明显。这说明本文提出的 ISNRPCA 模型相对于同类算法中的 RPCA, Godec 模型具有优势，一方面非负约束使得分离结果符合实际物理含义，另一方面优化 IS 距离度量与人耳的听觉感知特性相一致，避免了 RPCA, Godec 模型出现的弱频谱分量区域降噪语音失真过大且被人耳易感知的问题，从而改善降噪语音质量。

5　结论

本文提出了一种在耳蜗谱图上进行语噪分离的听觉感知鲁棒主成分分析法。ISNRPCA 算法以符合人耳主观听觉的板仓－斋田距离度量作为优化目标函数，对分解项施加非负约束并推导了迭代优化算法。该算法能够稳定地估计出耳蜗谱图中的稀疏语音分量和低秩噪声分量，并将两者有效分离。通过多种类型噪声和信噪比条件下的仿真实验，验证了 ISNRPCA 算法用于语音降噪时的良好性能，并进一步证实了语音信号和干扰噪声在耳蜗谱图上的显著差异。ISNRPCA 算法无需预先训练语音或噪声模型，是无监督的降噪算法。下一步工作目标是在低于 −5 dB 的极低 SNRs 条件下，进一步提高降噪语音的可懂度和总体质量。

参　考　文　献

1　Loizou P C. Speech enhancement: theory and practice, Boca Raton, FL: CRC Press, 2007

2　Wang D L, Brown G J. Computational auditory scene analysis: principles, algorithms, and applications. Hoboken, NJ: Wiley/IEEE Press, 2005

3　吴迪，陶智，张晓俊等. 感知听觉场景分析的说话人识别. 声学学报， 2016; **41**(2): 260—272

4　杨琳，张建平，颜永红. 单通道语音增强算法对汉语语音可懂度影响的研究. 声学学报， 2010; **35**(2): 248—253

5　Kamath S, Loizou P. A multi–band spectral subtraction method for enhancing speech corrupted by colored noise. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2002: **IV**-4164

6　吴红卫，俞一彪，吴镇扬. 基于 Laplace–Gauss 模型和简化相位判别的离散余弦变换域语音增强. 声学学报， 2008; **33**(3): 244—251

7　HUANG J J, ZHANG X W, ZHANG Y F *et al.* Single channel speech enhancement via time-frequency dictionary learning. *Chinese Journal of Acoustics*, 2013; **32**(1): 90—102

8　Wang D, Vipperla R, Evans N *et al.* Online non-negative convolutive pattern learning for speech signals. *IEEE Trans. on Audio, Speech, and Language Processing*, 2013; **21**(1): 44—56

9　Mohammadiha N, Smaragdis P, Leijon A. Supervised and

unsupervised speech enhancement using non-negative matrix factorization. *IEEE Trans. on Audio, Speech, and Language Processing*, 2013; **21**(10): 2140—2151

10 Smaragdis P, Févotte C, Mysore G J *et al*. Static and dynamic source separation using nonnegative matrix factorizations: a unified view. *IEEE Signal Processing Magazine*, 2014; **31**(3): 66—75

11 Candès E J, Li X D, Ma Y *et al*. Robust principal component analysis?. *Journal of the Acm*, 2011; **58**(3): 1—37

12 Sun C L, Zhang Q, Wang J *et al*. Noise reduction based on robust principal component analysis. *Journal of Computational Information Systems*, 2014; **10**(10): 4403—4410

13 Chen Z, Ellis D P W. Speech enhancement by sparse, low-rank, and dictionary spectrogram decomposition. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013: 1—4

14 Sun C, Zhu Q, Wan M. A novel speech enhancement method based on constrained low-rank and sparse matrix decomposition. *Speech Communication*, 2014; **60**: 44—55

15 Huang J J, Zhang X W, Zhang Y F *et al*. Speech denoising via low-rank and sparse matrix decomposition. *ETRI Journal*, 2014; **36**(1): 167—170

16 Gao B, Woo W L, Dlay S S. Unsupervised single-channel separation of nonstationary signals using Gammatone filter-bank and Itakura-Saito nonnegative matrix two–dimensional factorizations. *IEEE Trans. on Circuits and System I*, 2013; **60**(3): 662—675

17 Boyd S, Parikh N, Chu E *et al*. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 2011; **3**(1): 1—122

18 Sun D L, Févotte C. Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014: 6201—6205

19 Li Y, Wang D L. On the optimality of ideal binary time-frequency masks. *Speech Communication*, 2009; **51**(3): 230—239

20 Liang S, Liu W J, Jiang W. Integrating binary mask estimation with MRF priors of cochleagram for speech separation. *IEEE Signal Processing Letters*, 2012; **19**(10): 627—630

21 Hu Y, Loizou P C. Subjective evaluation and comparison of speech enhancement algorithms. *Speech Communication*, 2007; **49**(7–8): 588—601

22 Rice University digital signal processing (DSP) group, Noisex92 noise database, 2001. http://spib.rice.edu/spib/select_noise.html

23 Vincent E, Gribonval R, Févotte C. Performance measurement in blind audio source separation. *IEEE Trans. on Audio, Speech, and Language Processing*, 2006; **14**(4): 1462—1469

24 Rix A W, Beerends J G, Hollier M P *et al*. Perceptual evaluation of speech quality (pesq) – a new method for speech quality assessment of telephone networks and codes. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2001: 749—752

25 Hu Y, Loizou P C. Evaluation of objective quality measures for speech enhancement. *IEEE Trans. on Audio, Speech, and Language Processing*, 2008; **16**(1): 229—238

26 Taal C H, Hendriks R C, Heusdens R *et al*. An algorithm for intelligibility prediction of time-frequency weighted noisy speech. *IEEE Trans. on Audio, Speech, and Language Processing*, 2011; **19**(7): 2125—2136

27 Zhou Z, Li X D, Wright J *et al*. Stable principal component pursuit. IEEE International Symposium on Information Theory (ISIT), 2010: 1518—1522