

一种基于广义奇异值分解的语音增强算法

曹梅双¹, 曾庆宁¹, 陈芙蓉²

(1 桂林电子科技大学 通信与信息工程学院, 广西 桂林 541004;

2 桂林电子科技大学 计算机与控制学院, 广西 桂林 541004)

摘 要: 广义奇异值分解的单通道语音增强算法在加性噪声为白噪声的情况下, 效果比较理想. 加性噪声为有色噪声的情况下, 通常用一种基于熵奇异值分解(QSVD)的方法来处理. 对 QSVD 算法进行衍生, 首先提出了一种基于广义奇异值分解的子空间语音增强算法(GSVD). 为了处理低信噪比时残留的音乐噪声, 结合人耳的听觉掩蔽效应, 进一步提出了一种基于感官抑制的 GSVD(PCGSVD). 试验结果显示, PCGSVD 算法能够明显地提高语音质量、可懂度和识别率, 特别是在加性噪声是有色噪声的情况下实验结果明显优于其他的语音增强算法.

关键词: 听觉掩蔽门限; 广义奇异值分解; 信号子空间; 语音增强

中图分类号: TN4

文献标识码: A

文章编号: 1000-7180(2010)03-0083-04

A Perceptually Constrained GSVD-Based Approach for Speech Enhancement

CAO Mei-shuang¹, ZENG Qing-ning¹, CHENG Fu-rong²

(1 Department of Telecommunications and Information Engineering, Guilin University of Electronic Technology, Guilin 541004, China;

2 College of Computer and Control, Cuilin University of Electronic Technology, Guilin 541004, China)

Abstract: Generalized singular value decomposition is very useful when the additive noise is white. But the residual musical noise is still perceivable under lower signal-to-noise conditions. Therefore, a perceptually constrained GSVD, which is referred to PCGSVD algorithm, is further proposed to incorporate the masking properties of human auditory system to make sure the residual noise to be under the AMTs.

Key words: auditory masking thresholds; generalized singular value decomposition; signal subspace; speech enhancement

1 引言

近年来, 基于奇异值分解的子空间算法在信号处理上的应用越来越广泛^[1]. 通过对带噪语音信号样点构建的 Hank 矩阵进行奇异值分解(SVD), 能够将输入信号的矢量空间分解成为两个正交的子空间. 假定纯净语音信号只出现在信号子空间中, 而噪声则分布于信号子空间和噪声子空间, 我们可以去除噪声子空间, 通过信号子空间来重建纯净语音信号. 这种方法在加性噪声是白噪声的情况下非常有用, 但当噪声是非白噪声时, 常用的方法就是先将噪

声白化, 然后再反向处理. 为了避免这个过程, 应用 QSVD 算法可以将 SVD 和白化处理统一在一个语音增强框架内^[2]. 这种算法在处理有色噪声的时候效果非常理想. 文中对 QSVD 进行扩展, 提出了一种基于 GSVD 的语音增强算法. 在文献[3]中定义的步骤下, 可以精确地确定每一帧带噪语音信号中语音信号的维数. 实际上, 基于 KLT 算法的语音增强算法也应用了类似的概念来处理有色噪声^[4].

尽管 GSVD 算法优于以前的 SVD 算法, 在低信噪比的情况下, 增强得到的信号依然会有一些残留音乐噪声. 这也是文中提出结合人耳听觉掩蔽门

限和 GSVD 到同一框架下的原因^[5],称之为 PCGSVD. 由于 PCGSVD 操作于广义奇异值域,而传统的听觉掩蔽门限 (AMTs) 定义于频域,用现有方法来实现广义奇异域到频域之间的转换,得出了 PCGSVD 算法的闭合解的形式. 试验结果显示(客观与客观评价),无论加性噪声是平稳噪声还是非平稳噪声,PCGSVD 算法都优于以前提出的 SVD 算法,有效地减少了音乐噪声,提高了语音的质量和可懂度,特别是有色噪声的时候效果更佳.

2 基于 GSVD 语音增强算法

假设 y_i 是输入语音信号 y_I 的第 i 个样点,它等于 d_i 和 n_i 之和,它们分别代表纯净语音信号 d_I 和噪声信号 n_I 的第 i 个抽样点.

$$y_i = d_i + n_i \quad (1)$$

目的就是要从 y_i 中估计出 d_i . 图 1 给出了基于 GSVD 语音增强算法的框架.

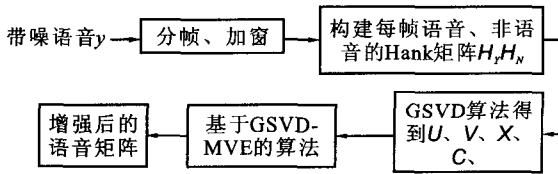


图1 基于 GSVD 语音增强算法框架

2.1 分帧、加窗

将输入语音信号 y_I 分为长度为 M 、重叠的语音帧,而后的增强算法都是针对每一帧进行.

2.2 构建每一帧数据的 Hank 矩阵

为了将子空间概念用于语音增强,必须构建两个 $L \times K$ 维的 Hank 矩阵,带噪语音信号及估计噪声信号的 Hank 矩阵如下:

$$H_Y = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_L & y_{L+1} & y_{L+3} & \cdots & y_{L+K-1} \end{bmatrix}$$

$$H_N^A = \begin{bmatrix} n_1 & n_2 & n_3 & \cdots & n_K \\ n_2 & n_3 & n_4 & \cdots & n_{K+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ n_L & n_{L+1} & n_{L+3} & \cdots & n_{L+K-1} \end{bmatrix}$$

其中, H_Y 为带噪语音信号矩阵, H_N^A 为最新缓冲得到的非语音帧, $M = L + K - 1$, 且 K 远小于 L . H_Y 可以表示为两个 Hank 矩阵 H_D 和 H_N 之和, $H_Y = H_D + H_N$. 在无噪声的情况下, H_N 为 $L \times K$ 维的 0 矩阵,我们必须选择 K ,使得 $\text{rank}(H_Y) < K$. H_D 和 H_N 都是未知的,但是 H_N 可以通过 $H_N^A = \partial H_N$ 估计

出来,其中 ∂ 可以是常量也可以是个变量. 减小 ∂ 的值能够减小估计语音的畸变,但是也能增加残留噪声,所以必须选一个折衷值,使得试验性能最佳. 文中取 $\partial = 1.1$.

通过下面给出的 GSVD 算法,可以得到 H_Y 和 H_N ,也就是说,可以运用子空间概念和 GSVD 算法于亏秩最小二乘问题中,从观察带噪语音信号中得到估计的纯净语音.

2.3 GSVD 算法

GSVD 算法在解决几种受限最小二乘问题上非常有用. 通过 GSVD,可以将 H_Y 和 H_N 分别转换为约束条件的对角矩阵 C 和 S ,如下:

$$U^T H_Y X = C = \text{diag}(c_1, \cdots, c_K), 1 \geq c_K \geq \cdots \geq c_1 \geq 0 \quad (2)$$

$$V^T H_N^A X = S = \text{diag}(s_1, \cdots, s_K), 1 \geq s_1 \geq \cdots \geq s_K \geq 0 \quad (3)$$

满足 $C^T C + S^T S = I_K$ 或 $c_i^2 + s_i^2 = 1, 1 \leq i \leq K$. 其中 $X \in R^{K \times K}$ 为非奇异矩阵, U 和 V 都为 $L \times K$ 的实矩阵,它们都是列正交向量. 由约束条件可以得到 GSVD 算法闭合的数学解析式^[6],能够精确地估计信号子空间和噪声子空间的维数. $c_i/s_i, i = 1 \sim K, i = 1 \sim K$,及 X 的列向量分别为 H_Y 和 H_N 的广义特征值和广义特征向量. 传统的 SVD 算法可以认为 H_N^A 为单位矩阵时的特例.

2.4 基于 GSVD-MVE 的子空间算法

矩阵 C 的对角元素 c_i 可以分为主要集和次要集. 主要集 ($c_i < s_i, i = 1, 2, \cdots, m, m \leq K$),对应的是 H_Y 和 H_N 中信号子空间部分,而次要集 ($c_i > s_i, i = m, \cdots, K, m+1 \leq i \leq K$),对应的是噪声子空间. m 和 $K - m$ 分别为信号子空间和噪声子空间的维数,其中 m 是一个与 c_i/s_i 大小有关的变量. m 与每一帧的瞬时信噪比有关,我们可以据此来灵活地确定每一帧 Hank 形式矩阵中信号子空间及噪声子空间的维数. H_Y 和 H_N 中信号子空间可以通过 X^{-1} 的前 m 个行向量来构建,而剩下的 $K - m$ 行则对应的是噪声子空间. 通过最小方差估计算法(得到最低的残留噪声水平),可以找到一个最佳的转换矩阵 $P \in R^{K \times K}$,使得 $H_Y P$ 和 H_D 的 Frobenius 距离最小,从而估计出 H_D .

$$P^* = \underset{\substack{P \in R^{K \times K} \\ \text{rank}(H_D) = m}}{\text{argmin}} \| H_Y P - H_D \|_F^2 \quad (4)$$

式(4)的闭合解析式为

$$H_D^A = H_Y P^* = U C' X^{-1} = \sum_{i=1}^m c'_i u_i \tilde{x}_i \quad (5)$$

式中, $\mathbf{U}$ 和 $\mathbf{X}$ 与式(2)和式(3)中的相同, u_i 及 $\tilde{x}_i$ 分别为 $\mathbf{U}$ 和 $\mathbf{X}^{-1}$ 中的第 i 个行向量. 从式(2)、式(3)、式(6)可以看出, $\mathbf{H}_Y$ 、 $\mathbf{H}_N$ 、 $\mathbf{H}_D^\Delta$ 都可以转换到由 $\mathbf{X}^{-1}$ 的行向量张成的空间中. 矩阵 $\mathbf{C}'$ 的对角元素如下:

$$c'_i = \begin{cases} c_i(1 - s_i^2/c_i^2), & 1 \leq i \leq m \\ 0, & \text{else} \end{cases} \quad (6)$$

式中 c_i 和 s_i 与式(2)相同.

估计得出的矩阵 $\mathbf{H}_D^\Delta$ 可能不是 Hank 矩阵形式, 可以简单地将对角线上的元素求平均即可.

2.5 得到估计纯净语音帧

由于 Hank 矩阵的特殊形式, 可以简单地取其第一列和最后一行的元素, 重排得到估计语音帧.

3 基于感官抑制的广义奇异值分解的语音增强算法

尽管广义奇异值分解的语音增强算法性能要优于得到广泛应用的增强算法, 但是在低信噪比的条件下, 还是可以察觉到残留的音乐噪声. 为此, 文中将人耳的听觉掩蔽效应结合 GSVD 算法, 提出了一种基于感官抑制的广义奇异值分解算法, 即 PCGSVD.

3.1 估计人耳的听觉掩蔽门限值

在日常生活中, 经常可以发现当两个声音同时发生时, 一个声音会受另一个声音的影响而减弱的现象, 这种现象叫做听觉掩蔽效应. 被掩蔽音刚能听到的掩蔽音的强度叫做听觉掩蔽门限.

以下步骤为听觉掩蔽门限的计算:

(1) 计算每一 bark 带的能量 $B(i)$

$$B(i) = \sum_{w=w_{bl}}^{w=w_{bh}} X_p(w, i) \quad (7)$$

式中, $X_p(w, i)$ 为第 i 帧语音信号的功率谱, w_{bl} 和 w_{bh} 分别对应的临界带的上下限.

(2) 计算扩展临界带宽功率谱

$$C_j = \sum_{j=1}^{i_{\max}} B(j) * \text{SF}(i, j), j = 1, 2, \dots, i_{\max} \quad (8)$$

其中

$$\text{SF}(i, j) = 15.81 + 7.5 * (i - j + 0.474) - 17.5 * \sqrt{1 + (i - j + 0.474)^2}$$

(3) 由于当纯音和噪音同时存在的时候, 纯音各个部分的频率成分不仅受到邻近频率纯音的掩蔽影响, 也受到噪音的影响. 所以要对上面得到的掩蔽值做进一步修正.

$$T'_i = 10 \lg((C_i) - (O_i/10)) \quad (9)$$

式中, $O_i = \partial(14.5 + i) + 5.5(1 - \partial)$, $\partial = \min(\text{SFM}_{db}/\text{SFM}_{db \max}, 1)$, $\text{SFM}_{db \max} = -60 \text{db}$, $\text{SFM} = \mu_g/\mu_\partial$, μ_g, μ_∂ 分别为每个临界带频谱密度的几何和算术平均值.

(4) 门限值必须要高于人耳的绝对听感, 否则就没有意义, 所以最后要与听感比较.

3.2 估计出 AMTs, 然后将其转换到奇异值域

由于上述 AMTs 是在频域内得出的, 然而基于 GSVD 的算法操作与广义奇异值域, 所以必须将 AMTs 转换到广义奇异值域. 因为纯净语音信号的功率谱并不知道, 可以由式(6)、式(7)得出估计语音的自相关矩阵:

$$\mathbf{R}_{dd}^\Delta \cong (1/M - 1)(\mathbf{H}_D^\Delta)^T \mathbf{H}_D^\Delta = (1/M - K - 1)\mathbf{X}^{-T}(\mathbf{C}')^2\mathbf{X}^{-1} \quad (10)$$

式中, M 是 $\mathbf{H}_Y$ 的行维, 而 K 是列维. 从 $\mathbf{R}_{dd}^\Delta$ 可以估计出语音信号的功率谱:

$$\hat{\Gamma} \cong (1/M - K - 1) * \sum_i^m [c_i(1 - s_i^2/c_i^2)]^2 w_i^T \quad (11)$$

式中, c_i/s_i 与式(2)、式(3)中相同, m 为信号子空间的维数, $w_i \in R^{1 \times J}$ 为 $\mathbf{X}^{-1}$ 的第 i 行的 J 点 DFT. 根据估计功率谱 $\hat{\Gamma}$, 可以由上一小节给出的步骤推导出频率域的 AMTs $\theta \in R^{J \times 1}$, 进一步可以推导出奇异值域的 AMTs $\theta \in R^{J \times 1}$:

$$\gamma = \lceil (1/J) \mathbf{G} \theta \rceil \quad (12)$$

式中, $\mathbf{G} \in R^{K \times J}$ 为一个变换矩阵, 它的行向量为矩阵 $\mathbf{X}$ 的 J 点 DFT.

3.3 基于 PCGSVD 子空间算法及其闭合解的戒.

PCGSVD 算法和 GSVD 算法除了步骤 4 不同, 其余的基本相同, 所以在这里只需要比较这一点就可以了. 选取的矩阵 $\mathbf{P}$, 不仅要使 $\mathbf{H}_p \mathbf{P}$ 和 $\mathbf{H}_D$ 之间的 Frobenius 距离最小, 还要使估计还有使估计噪声前 m 个归一化能量小于人耳的听觉掩蔽门限值.

$$\mathbf{P}^* = \underset{\substack{\mathbf{P} \in R^{K \times K} \\ \text{rank}(\mathbf{H}_D) = m}}{\text{argmin}} \|\mathbf{H}_Y \mathbf{P} - \mathbf{H}_D\|_F^2 \quad (13)$$

满足:

$$\left\{ \begin{array}{l} \|v_i^T \mathbf{H}_N^\Delta \mathbf{p}\|^2 / \|\tilde{x}_i\|^2 \leq \gamma_i, 1 \leq i \leq m \\ \|v_i^T \mathbf{H}_N^\Delta \mathbf{p}\|^2 / \|\tilde{x}_i\|^2 = 0, M+1 \leq i \leq K \end{array} \right\},$$

式中, $v_i \in R^{L \times 1}$ 和 $\tilde{x}_i \in R^{1 \times K}$ 分别为 $\mathbf{V}$ 的第 i 个列向量及 $\mathbf{X}^{-1}$ 的第 i 个行向量.

由式(12)、式(13)可以得出 PCGSVD 的闭合解:

$$H_D^A = H_Y P^* = UC'X^{-1} \quad (14)$$

矩阵 C' 的对角元素：

$$c'_i = \begin{cases} \min[c_i(1 - s_i^2/c_i^2), c_i * \sqrt{\gamma/s_i}], & 1 \leq i \leq m \\ 0, & m+1 \leq i \leq K \end{cases} \quad (15)$$

注意到式(15)与式(6)唯一的不同在于,式(15)结合了听觉掩蔽门限效应,即残留噪声要低于人耳的听觉掩蔽门限^[7].

4 试验结果分析

采用的纯净语音为抽样频率为 8kHz,人为加入 0db 的高斯白噪声, M 取值 512,即一帧语音长度,设定 $J = M, \vartheta = 1.1, L = 473, K = 40$. 试验结果如图 2、图 3 所示.

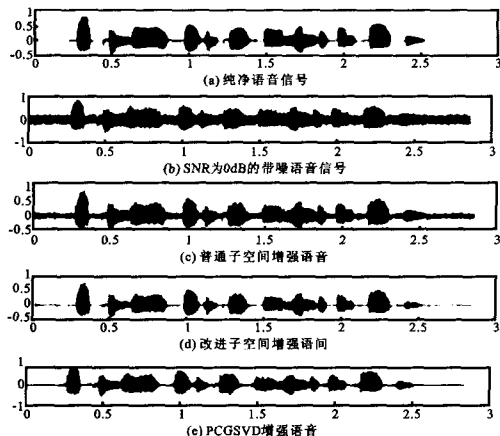


图 2 一段语音(s)对应的归一化幅度

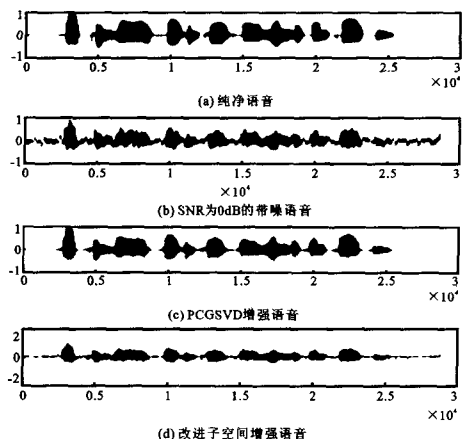


图 3 一段语音(s)对应的语谱图

采用的纯净语音为 8kHz 采样,人为加入工厂噪声录音(有色噪声), M 取为 512,即一帧长度,设定 $J = M, O = 1.1, L = 473, K = 40$. 试验结果见图 4、图 5.

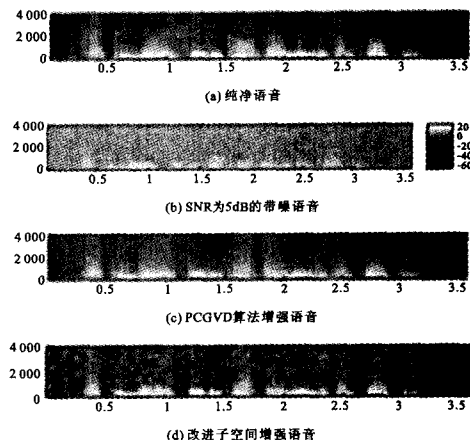


图 4 一段语音(s)对应的归一化幅度

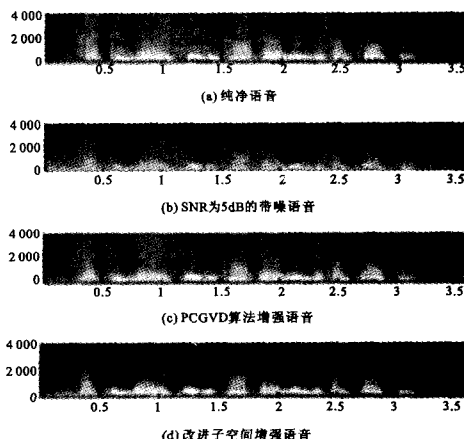


图 5 一段语音(s)对应的语谱图

5 结束语

在加入白噪声时,从语谱图中可以看到,语音的高频分量有一定程度的衰减.这是因为类噪声语音信号无法掩蔽宽频白噪声,所以这部分频率成分被当成噪声减掉.在加入有色噪声时,因为噪声可以被掩蔽效应所掩蔽,增强语音要明显好于其他子空间算法.也说明现实中的窄带及低通噪声可以有效被滤除,体现了算法的实用性.而且主观的听觉测试显示,PCGSVD 算法较之以前的子空间算法得到的增强语音的残留音乐噪声现象要好很多,有效地提高了带噪声语音的质量和可懂度.文中算法可以广泛应用于自动语音识别、人与人及人机通信系统中去.

参考文献:

[1] 胡玉平. 基于神经网络的奇异值分解域公开水印算法[J]. 微电子学与计算机, 2008, 25(2): 1-4.

(下转第 92 页)

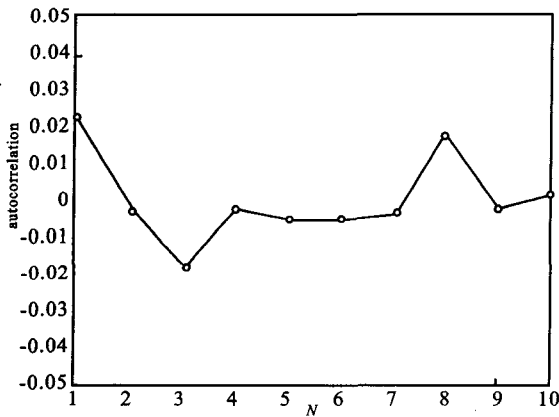


图 11 随机序列的自相关性分析

表 1 NIST SP800-22 测试标准测试结果

SP800-22 Test	Length 20000	Num 10	Length 4000	Num 50
	通过率		通过率	
Frequency	1.000		0.960	
Block Frequency	1.000		1.000	
Runs	0.900		1.000	
Spectral DFT	1.000		1.000	
Non-overlapping Templates	1.000		1.000	
Serial	1.000		0.960	
Approximate Entropy	1.000		0.980	
Cumulative Sums(1, 2)	0.900		0.970	

说明了其产生随机数的原理,并采用 HJTC 0.18 μ m CMOS 混合信号工艺设计实现,系统具有很好的随机性和很高输出比特率.对该 TRNG 的整体仿真和验证表明,在未加任何后处理电路修正的情况下生成的随机数通过了标准随机性测试,输出比特率可达 200Mb/s,非常适合用于集成电路 CMOS 主流工艺设计环境下的高速加密嵌入式系统中.

参考文献:

- [1] 辛茜,曾晓洋,张国权,等.基于电阻热噪声的真随机数发生器设计[J].微电子学与计算机,2004,21(7):143-146.
- [2] Bucci M, Germani L, Luzzi R, et al. A high-speed oscillator-based truly random number source for cryptographic applications on a smart card IC[J]. IEEE Trans. on Computers, 2003, 52(4):403-409.
- [3] Stojanovski T, Kocarev L. Chaos-based random number generators - Part I: analysis[J]. IEEE Trans. on Circuits and System I, 2001, 48(3):281-288.
- [4] 王欣,周童,王永生,等.一种基于混沌原理的真随机数发生器[J].微电子学与计算机,2009,26(2):135-139.
- [5] Callegari S, Rovatti R, Setti G. Embeddable ADC-based true random number generator for cryptographic applications exploiting nonlinear signal processing and chaos[J]. IEEE Trans. on Signal Processing, 2005, 53(2):793-805.
- [6] Poli S, Callegari S, Rovatti R, et al. Post-processing of data generated by a chaotic pipelined ADC for the robust generation of perfectly random bitstreams[C]//IEEE International Symp. on Circuits and Systems. Italy: Bologna, 2004:585-588.
- [7] 刘秉正,彭建华.非线性动力学[M].北京:高等教育出版社,2004.
- [8] Setti G, Mazzini G, Rovatti R, et al. Statistical modeling of discrete-time chaotic processes - basic finite-dimensional tools and applications[J]. the IEEE, 2002, 90(5):662-690.

作者简介:

张勇男,(1984-),硕士.研究方向为数模混合集成电路设计.

王永生男,(1974-),博士,副教授.研究方向为数模混合集成电路设计.

(上接第 86 页)

- [2] Jenson S H, Hanson P C, Hanson S D. Reduction of broad-band noise in speech by truncate QSVD[J]. IEEE Trans. Speech audio process, 1995,3(6):439-448.
- [3] Ju G H, Lee L S. Speech enhancement based on generalized singular value decomposition approach[C]//Pro. IC-SLP. Taiwan: National Taiwan University, 2002:1801-1804.
- [4] Mittal U, Phamdo N. Signal/noise KLT based approach for enhancing speech degraded by colored noise[J]. IEEE Trans. Speech audio process, 2000,8(2):159-167.
- [5] Ju G H, Lee L S. Perceptually constrained generalized singular value decomposition-based approach for enhancing

speech corrupted by colored noise[C]//Pro. ISCA 2003. Geneva, Switzerland, 2003:533-536.

- [6] Golub G H, Van C F. Loan matrix computation[M]. 2nd ed. Baltimore, MD: Jonns Hopkins, univ. press, 1996.
- [7] Jablou F, Champagne B. A perceptual signal subspace approach for speech enhancement in colored noise[C]//Acoustics, Speech, and Signal Processing, 2002 IEEE International Conference. USA: Orlando, 2002.

作者简介:

曹梅双男,(1981-),硕士研究生.研究方向为语音增强、说话人识别.