ISSN 1000-0054 清华大学学报(自然科学版)2011 年 第 51 卷 第 9 期 16/33
CN 11-2223/N J Tsinghua Univ (Sci & Tech), 2011, Vol. 51, No. 9 1234-1238

基于压缩感知过程的语音增强

周小星[1], 王安娜[1], 孙红英[2], 杨鸿武[2]

(1. 东北大学 信息科学与工程学院, 沈阳 110819; 2. 西北师范大学 物理与电子工程学院, 兰州 730070)

摘 要: 压缩感知(compressive sensing, CS)是一种基于信号稀疏性的采样方法,可以有效提取信号中所包含的信息。该文提出了一种基于 CS 过程的语音增强新算法。算法利用语音在离散余弦变换(discrete cosine transform, DCT)域下的稀疏性,采用 Hadamard 矩阵对带噪语音进行压缩测量,通过改进的正交匹配跟踪(orthogonal matching pursuit, OMP)算法恢复语音信号,实现语音增强。与经典谱减法和子空间算法进行实验对比分析,结果表明:该算法在降噪性能上优于经典谱减法和子空间算法。

关键词: 语音增强; 压缩感知; 离散余弦变换; Hadamard 矩阵; 正交匹配跟踪

中图分类号: TN 912.35 **文献标志码**: A

文章编号: 1000-0054(2011)09-1234-05

Speech enhancement based on compressive sensing

ZHOU Xiaoxing[1], WANG Anna[1], SUN Hongying[2], YANG Hongwu[2]

(1. College of Information Science and Engineering, Northeastern University, Shenyang 110819, China;
2. College of Physics and Electronic Engineering, Northwest Norm University, Lanzhou 730070, China)

Abstract: Compressive sensing (CS) which is a sampling method based on the signal sparseness, can efficiently extract information contained in a signal. This paper presents a speech enhancement algorithm based on CS using a Hadamard matrix to compress the noisy speech using the sparseness of the speech in a discrete cosine transform (DCT). The speech signal is then restored using the modified orthogonal matching pursuit (OMP) by setting the iterative threshold to realize speech enhancement. Tests show that the algorithm outperforms both the spectral subtraction speech enhancement method and the subspace speech enhancement method for noise removal.

Key words: speech enhancement; compressive sensing; discrete cosine transformation; Hadamard matrix; orthogonal matching pursuit

语音是非平稳、时变的信号。通过语音传递信息是人类最重要、最常用的信息交换形式之一。通常,研究者们是在语音信号相对纯净的条件下,对信号进行各种处理。但现实生活中的语音不可避免地要受到周围环境噪声的影响。这些噪声的存在会严重影响语音信号的质量与可懂度。在这种实际需要的推动下,早在 60 年代,语音增强这个课题作为语音信号处理的一个重要分支就已经引起了人们的注意;70 年代曾形成了一个研究高潮,并取得一些基础性成果。目前,语音增强方法[1]主要有谱减法、Weiner 滤波、Kalman 滤波,以及相继发展起来的子空间增强、小波变换和这些增强方法的各种改进算法。

近年来兴起了一种充分利用信号稀疏性的全新信号采集、编解码理论——压缩感知(compressive sensing, CS)[2]。有别于 Nyquist 采样定理,在 CS 理论中,采样率只依赖于信号的稀疏性和等距约束性(restricted isometry property, RIP)[3],不再受信号带宽的限制。CS 只用较少的压缩测量值表示信号而不丢失恢复信号所需要的数据,在大大降低信号采样率的同时,实现信号的准确或近似重建。这样不仅可以更高效地处理数据,还可以节省存储成本。现阶段,研究者们已经将 CS 广泛应用于信息压缩编码、信号恢复、数字图像处理及遥感图像处理等[4]诸多领域,但在语音处理方面的应用还非常少。本文分析了 CS 的基本原理,将 CS 应用于语音增强处理,采用 Hadamard 矩阵对带噪语音进行压缩测量,通过改进的正交匹配跟踪(orthogonal matching pursuit, OMP)算法[5]恢复语音信号,实现语音增强。与经典谱减法和子空间算法进行实验对比分析,结果表明:该算法在降噪性能上优于经典谱减法和子空间

收稿日期: 2011-07-15

基金项目: 国家自然科学基金资助项目(60875015)

作者简介: 周小星(1986 -),男(汉),江西,硕士研究生。

通信作者: 王安娜,教授,E-mail: wanganna@mail.neu.edu.cn

算法。

1 压缩感知理论

CS 理论主要包括信号的稀疏表示、测量矩阵和重建算法 3 部分[6]：

1) 信号的稀疏表示。

信号具有稀疏性是应用 CS 理论的前提。稀疏性是指信号自身或者经过变换后，仅含有少数非零值，绝大部分值均等于零，但这种严格的稀疏性要求在多数情况下很难满足。因此，只要信号是近似稀疏的，绝大部分值接近零，具有可压缩性，则 CS 同样适用。设离散时间信号 $\boldsymbol{x}=[x_1,x_2,\cdots,x_N]^{\mathrm{T}}$，$\boldsymbol{\Psi}=[\boldsymbol{\psi}_1,\boldsymbol{\psi}_2,\cdots,\boldsymbol{\psi}_N]$ 为 $N\times N$ 维正交变换矩阵，$\boldsymbol{\psi}_i=[\psi_i(1),\psi_i(2),\cdots,\psi_i(N)]^{\mathrm{T}}$，且 $\langle\boldsymbol{\psi}_i,\boldsymbol{\psi}_j\rangle=0$，$i\neq j$。$\boldsymbol{x}$ 满足：

$$\boldsymbol{x}=\boldsymbol{\Psi}\boldsymbol{y}=\sum_{i=1}^{N}y_i\boldsymbol{\psi}_i. \tag{1}$$

其中 $\boldsymbol{y}=[y_1,y_2,\cdots,y_M]^{\mathrm{T}}$，$y_i=\langle\boldsymbol{x},\boldsymbol{\psi}_i\rangle$。$\boldsymbol{y}$ 中非零元素个数称为 $\boldsymbol{y}$ 的 $\boldsymbol{l}_0$ 范数，记做 $\|\boldsymbol{y}\|_0$。如果 $\|y\|_0=K$ 且 $K\ll N$，则称信号 $\boldsymbol{x}$ 为 K 稀疏的。

2) 测量矩阵。

测量矩阵用来对高维信号进行低维投影来获取采样信号，是 CS 理论的实现方式。合适的测量矩阵应该保证低维采样得到 M 个观测值，完整包含或者尽可能多地包含原信号信息。为了确保信号的线性投影能够保持信号的原始结构，CS 中的测量矩阵必须满足 RIP，或者与变换域矩阵具有不相关性。压缩测量过程的数学描述如下：当 $\boldsymbol{x}$ 为 K 稀疏信号时，可以利用 $M\times N$ 维测量矩阵 $\boldsymbol{\Phi}$，将 N 维数据压缩成 M 维测量信号 $\boldsymbol{s}=[s_1,s_2,\cdots,s_M]^{\mathrm{T}}$：

$$\boldsymbol{s}=\boldsymbol{\Phi}\boldsymbol{x}. \tag{2}$$

其中 $M\geqslant CK\log_2(N/K)$，C 为常数。

3) 重建算法。

重建算法是 CS 理论的核心。目的是利用尽可能少的 M 个观测值快速稳定、高概率地恢复出长度为 N 的原始信号。即根据式(2)，求解线性方程组得到 $\boldsymbol{x}$。将式(1)代入式(2)得到：

$$\boldsymbol{s}=\boldsymbol{\Phi}\boldsymbol{x}=\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{y}=\tilde{\boldsymbol{\Phi}}\boldsymbol{y}. \tag{3}$$

其中 $\tilde{\boldsymbol{\Phi}}=[\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_N]$，称为传感矩阵或恢复矩阵，维数为 $M\times N$。

当 $\|\boldsymbol{y}\|_0=K$，$K<M$，且 $\tilde{\boldsymbol{\Phi}}$ 满足 RIP 条件，重建问题转化为求解最优 $\boldsymbol{l}_0$ 范数问题：

$$\hat{\boldsymbol{y}}=\arg\min\|\boldsymbol{y}\|_0,\quad \text{s.t.}\quad \boldsymbol{s}=\tilde{\boldsymbol{\Phi}}\boldsymbol{y}. \tag{4}$$

该问题是一个 NP 难问题。有 3 类求解方法：第 1 类为贪婪算法，主要有匹配跟踪及其改进算法；第 2 类为凸优化算法，主要有基追踪算法等；第 3 类为其他算法，包括最小全变分方法、迭代阈值法以及各种改进算法。

2 基于压缩感知的语音增强

本文利用语音和噪声在离散余弦变换(discrete cosine transform，DCT)域下稀疏性的不同，通过改进 OMP 重建算法，设置相似度阈值实现语音增强。

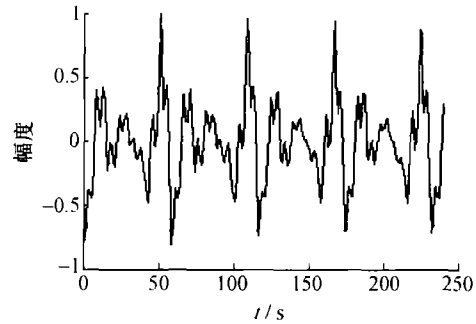
语音信号在 DCT 域具有稀疏性，而噪声则不具有。图 1 中，图 1a 和 1b 分别为语音和 Gauss 白噪声，图 1c 和 1d 分别为图 1a 和 1b 经过 DCT 后的取值分布直方图。对比图 1c 和 1d 可知：在图 1c 中，绝大部分取值接近零，舍弃那些接近零值的数据，逆 DCT 后的语音失真比较小，即语音在 DCT 域具有稀疏性；在图 1d 中，所有值分布都比较均匀，舍弃部分值恢复的信号失真比较大，因此 Gauss 白噪声在 DCT 域不具有稀疏性。根据 CS 理论，对带噪语音进行低维投影，当观测维数足够包含语音信息时，由于噪声不具有稀疏性，投影后将丢弃部分噪声信息，这部分噪声在信号重建时无法恢复。因此利用语音和噪声在 DCT 域稀疏性的不同可以实现部分的噪声去除。

进行低维投影的测量矩阵分为随机测量矩阵和确定性测量矩阵。目前，普遍采用随机测量矩阵。对于不同的测量矩阵，如果恢复相同质量的信号所需数据维数越低，则称此测量矩阵性能比其他矩阵的更优。根据上述分析，在保证包含足够语音信息的前提下，当观测数据维数越低，丢弃噪声信息越多，语音增强效果就越好。因此，进行语音增强时需要选用性能最优的测量矩阵。常用的 4 种随机测量矩阵的性能由高到低次序为：Hadamard 矩阵、Toeplitz 矩阵、Bernoulli 分布随机矩阵、Gauss 分布随机矩阵(其中后两者性能相同)[7]。因此在进行语音增强时，理论上采用 Hadamard 矩阵去噪效果要比其他随机矩阵的好。

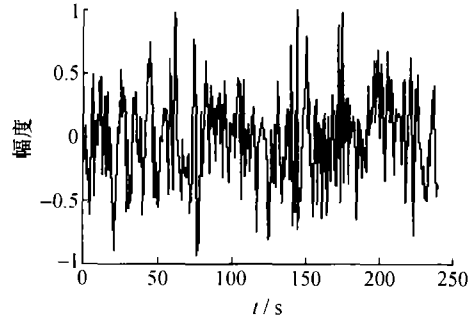
本文采用改进的 OMP 算法进行语音信号重建。基本的 OMP 算法原理[8]如下：

1) 初始化：$\boldsymbol{s}$ 为观测信号，K 为信号稀疏程度。设置残差 $\boldsymbol{r}(0)=\boldsymbol{s}$，索引集合 $\boldsymbol{\Lambda}(0)=\varnothing$，原子集合矩阵 $\boldsymbol{\Theta}=[]$为空矩阵，迭代序号 $i=0$。

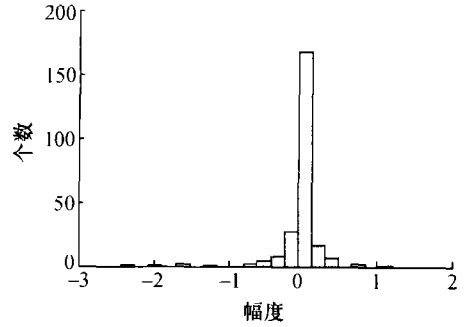
2) 循环开始：计算残差 $\boldsymbol{r}(i)$与传感矩阵 $\tilde{\boldsymbol{\Phi}}$ 的列向量 $\boldsymbol{\varphi}_j$ 内积，也即相似度 $m_j=\langle\boldsymbol{r}(i),\varphi_j\rangle$，$1\leqslant j\leqslant N$。寻找对应相似度最大列的索引 λ，$\lambda=$



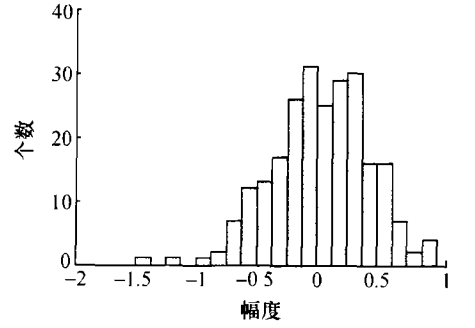
(a) 语音



(b) 高斯白噪声



(c) 语音 DCT 变换后取值分布直方图



(d) 白噪声 DCT 变换后取值分布直方图

图 1 语音和噪声在 DCT 域的分布对比

$\arg\max_{j=1,\cdots,N}|m_j|$，相应最大相似度值为 m_λ。迭代序号 $i=i+1$。

3) 更新索引集合 $\boldsymbol{\Lambda}_i=\boldsymbol{\Lambda}_{i-1}\cup\{\lambda\}$，扩充原子集合矩阵 $\boldsymbol{\Theta}=[\boldsymbol{\Theta},\boldsymbol{\varphi}_\lambda]$。

4) 计算信号估计值 $\hat{\boldsymbol{y}}(i)=(\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{s}$，并更新残差 $\boldsymbol{r}(i)=\boldsymbol{s}-\boldsymbol{\Theta}\hat{\boldsymbol{y}}(i)$。判断 $i\leqslant K$，若满足则转至步骤 2 继续迭代，否则迭代终止。

5) 结束：求解 $\hat{\boldsymbol{x}}=\boldsymbol{\Psi}\hat{\boldsymbol{y}}(K)$，算法结束。

利用基本的 OMP 算法进行语音信号重建的过程中，主要依据残余信号与传感矩阵列向量之间的相似度。在迭代初始阶段，语音分量占残余信号的主要部分，相似度比较大，恢复信号成分是语音；随着迭代的继续，语音分量不断被提取，相似度不断减小。当噪声分量和残余语音分量相当时，恢复成分开始包含部分噪声。如果一直迭代直到结束，显然恢复信号会包含大量噪声，影响增强效果。因此需要对基本的 OMP 算法进行一些改进，以实现语音增强处理。可以通过设置相似度阈值来控制迭代次数，阈值大小可以通过实验获取。在重建恢复迭代过程中，当计算得到的相似度低于此阈值时，此帧语音处理结束，进行下一帧处理。

3 实验结果与对比分析

为了验证本文利用 CS 理论进行语音增强的效果，对不同噪声类型和不同信噪比情况下的带噪语音分别采用本文算法、经典谱减法和子空间算法进行去噪处理，对比分析实验结果。

实验语音样本来自 NOIZEUS 带噪语音库[9]，语音采样频率为 8 kHz。该库中包含 30 条句子，分别被 8 种不同类型的噪声以不同的信噪比(SNR)进行干扰。这些噪声包括展览馆场景声、火车场景声、汽车场景声、车站场景声、街道场景声、餐馆场景声、咿呀作声、机场场景声。由于 NOIZEUS 库不包含白噪声，实验中采用 Gauss 白噪声干扰纯净语音，以产生不同 SNR 的噪声语音。

3.1 白噪声干扰

语音信号受 Gauss 白噪声干扰时，分别选取 SNR 为 5 dB 与 −5 dB 的带噪语音，通过 3 种增强算法的增强处理，对比分析增强效果。图 2 中，图 2d、2e、2f 分别为对 SNR = 5 dB 的带噪语音使用经典谱减法、子空间法、本文算法增强的结果；图 2g、2h、2i 分别为对 SNR = −5 dB 的带噪语音使用经典谱减法、子空间法、本文算法增强的结果。

由图 2 可知：当 SNR = 5 dB 时，采用本文算法与经典谱减法和子空间法的增强效果相当，都能很好地去除噪声干扰；但在 SNR = −5 dB 时，信号完全被噪声所淹没，此时本文算法比其他 2 种算法增强效果好。经典谱减法虽然做法简单，实现容易，但其增强效果取决于对噪声功率谱估计的准确程度。一般情况下噪声功率谱不容易估计，在强噪声环境下更困难。因此图 2g 中存在“音乐噪声”，去噪效果

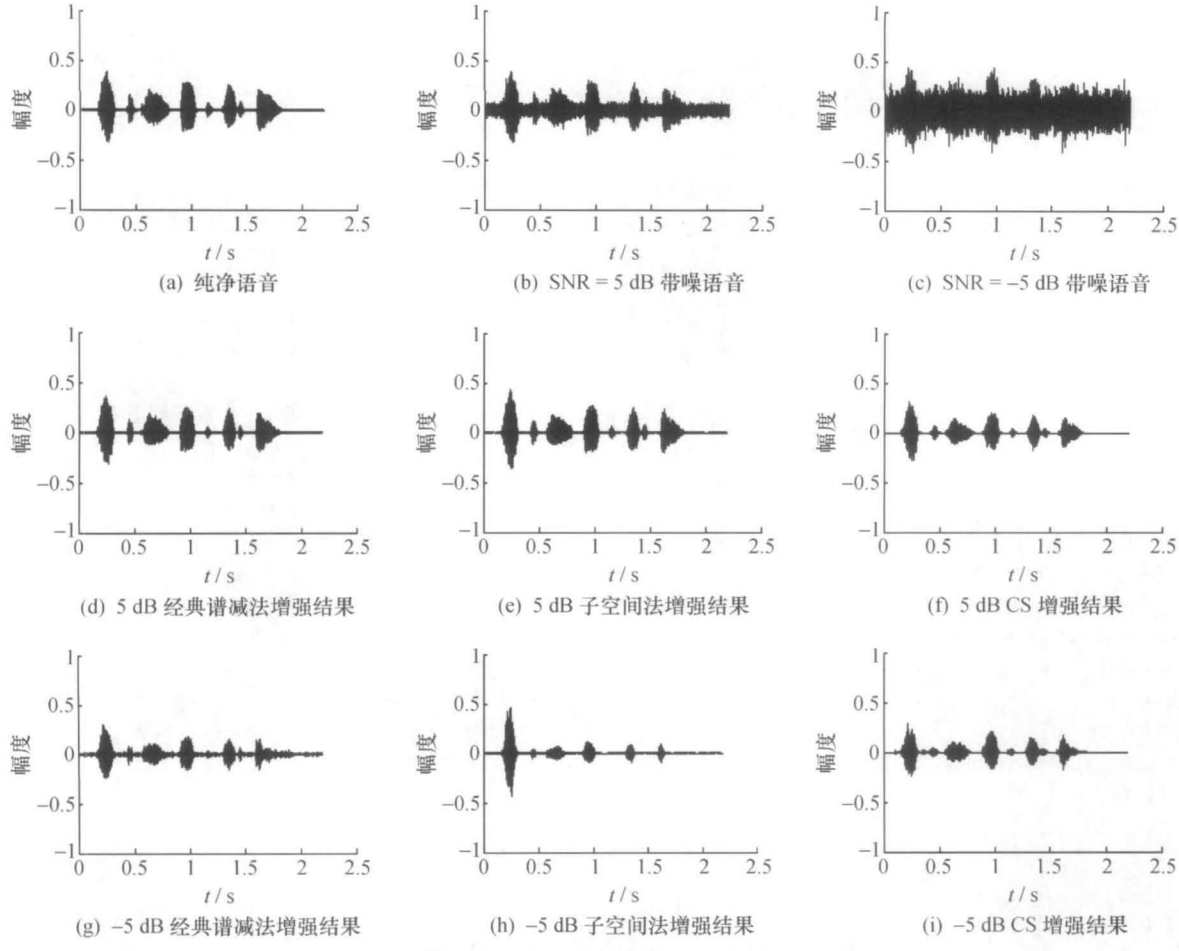


图 2 白噪声干扰下的 3 种算法语音增强效果比较

不太理想。由于子空间算法在将带噪语音分解为信号和噪声 2 个正交子空间时，对噪声和语音有各种假设，而在强噪声干扰下，这些假设很难满足。因此图 2h 显示的子空间算法去噪效果很不理想，有明显的失真。本文算法利用语音与噪声信号在 DCT 域稀疏性的不同以及噪声与语音特性之间的差异实现语音增强，所以效果较好。

3.2 色噪声干扰

当信号受色噪声干扰时，选用 SNR＝0 dB 的不同类型带噪语音进行增强处理，针对每一种噪声类型，求得增强结果的平均 SNR。表 1 为各种不同噪声下采用 3 种增强算法处理后的平均 SNR。

表 1 不同噪声干扰下应用 3 种算法增强后语音的平均 SNR

噪声	平均 SNR/dB		
	经典谱减法	子空间法	本文算法
展览馆场景声	9.80	12.54	10.62
火车场景声	11.00	11.36	11.12
汽车场景声	9.72	10.34	13.68

（续表）

噪声	平均 SNR/dB		
	经典谱减法	子空间法	本文算法
车站场景声	8.28	11.11	11.33
街道场景声	9.38	10.55	12.82
餐馆场景声	9.77	10.65	8.50
咿呀作声	10.60	12.75	8.96
机场场景声	12.25	13.80	10.55

由表 1 中的数据可以看出：当噪声为展览馆和火车场景声的情况下，本文算法效果介于经典谱减法与子空间之间；当噪声为汽车、车站和街道场景声的情况下，本文算法效果要优于经典谱减法与子空间算法；而当噪声为餐馆场景声、咿呀作声和机场场景声时，本文算法增强效果稍差于经典谱减法与子空间法。这种增强差异主要源于本文算法是利用语音与噪声信号在 DCT 域稀疏性的不同以及两者之间特性的差异，实现语音增强；当干扰噪声类似于语音时，语音和噪声之间的差异不明显，恢复语音中将包含大量噪声，去噪效果下降。如何改进这些不足是下一步要开展的工作。

4 结 论

本文分析了CS的基本原理,并将CS应用到语音增强处理中。给出了语音信号在DCT域的近似稀疏表示,采用Hadamard矩阵对带噪语音进行压缩测量,通过改进的OMP算法对语音信号进行恢复,实现语音增强处理。在不同类型以及不同强度的噪声干扰下,将本文算法与经典谱减法和子空间算法进行实验比较。实验结果表明:当干扰为低SNR白噪声或某些色噪声时,本文算法在降噪性能上优于经典谱减法和子空间算法。

参考文献 (References)

[1] Loizou P C. Speech Enhancement: Theory and Practice [M]. USA CRC Press, 2007.

[2] Baraniuk R G. Compressive sensing [J]. *IEEE Signal Processing Magazine*, 2007, **24**(4): 118-124.

[3] Candès E, Wakin M. An introduction to compressive sampling [J]. *IEEE Signal Processing Magazine*. 2008, **25**(2): 21-30.

[4] 石光明, 刘丹华, 高大化, 等. 压缩感知理论及其研究进展 [J]. 电子学报, 2009, **37**(5): 1070-1081.
SHI Guangming, LIU Danhua, GAO Dahua, et al. Advances in theory and application of compressed sensing [J]. *Chinese journal of electronics*, 2009, **37**(5): 1070-1081. (in Chinese)

[5] Blumensath T, Davies M E. Gradient pursuits [J]. *IEEE Transaction on Signal Processing*, 2008, **56**(6): 2370-2382.

[6] 金坚, 谷源涛, 梅顺良. 压缩采样技术及其应用 [J]. 电子与信息学报, 2010, **32**(2): 470-475.
JIN Jian, GU Yuantao, MEI Shunliang. An introduction to compressive sampling and its applications [J]. *Journal of Electronics & Information Technology*, 2010, **32**(2): 470-475. (in Chinese)

[7] 李小波. 基于压缩感知的测量矩阵研究 [D]. 北京: 北京交通大学, 2010.
LI Xiaobo. Research on Measurement Matrix Based on Compressed Sensing [D]. Beijing: Beijing Jiaotong University, 2010. (in Chinese)

[8] Tropp J, Gilbert A. Signal recovery from random measurements via orthgonal matching pursuit [J]. *Transaction on information theory*, 2007, **53**(12): 4655-4666.

[9] Yi Hu, Loizou P C. Subjective comparison and evaluation of speech enhancement algorithms [J]. *Speech Communication*, 2007, **49**(7-8), 588-601.

(上接第1233页)

参考文献 (References)

[1] Young S. The HTK Book [EB/OL]. [2010-01-01]. http://htk.eng.cam.ac.uk.

[2] Hong Y, Abeer A, Kazemzadeh A, et al. Pronunciation variations of Spanish-accented English spoken by young children [C]// In Proceedings of INTERSPEECH'2005. Lisbon, 2005: 749-752.

[3] YANG Jian, WU Peishan, XU Dan. Mandarin speech recognition for nonnative speakers based on pronunciation dictionary adaption. [C]// In Proceedings of ISCSLP'08. Kunming, 2008: 1-4.

[4] 潘复平, 赵庆卫, 颜永红. 一种用于方言口音语音识别的字典自适应技术 [J], 计算机工程与应用, 2005, 41(23): 5-6.
PAN Fuping, ZHAO Qingwei, YAN Yonghong. Pronunciation dictionary adaptation based accent modeling for large vocabulary continuous speech recognition [J]. *Computer Engineering and Applications*, 2005, **41**(23): 5-6. (in Chinese)

[5] Hoste V, Daelemans W, Gillis S. Using rule induction techniques to model pronunciation variation in Dutch [J]. *Computer Speech and Language*, 2004, **18**(1): 1-23.

[6] Wester M. Pronunciation modeling for ASR knowledge-based and data-derived methods [J]. *Computer Speech and Language*, 2003, **17**(1): 69-85.

[7] 刘林泉, 郑方, 吴文虎. 基于小数据量的方言普通话语音识别声学建模 [J]. 清华大学学报(自然科学版), 2008, 48(4): 604-607.
LIU Linquan, ZHENG Fang, WU Wenhu. Research on a small data set based acoustic modeling for dialectal Chinese speech recognition [J]. *J Tsinghua Univ (Sci and Tech)*, 2008, **48**(4): 604-607. (in Chinese)

[8] 米娜瓦尔. 维吾尔语方言和语言调查 [M]. 北京: 民族出版社, 2004.
Minawar. Uyghur Dialect and Language Investigation [M]. Beijing: The ethnic publishing house, 2004. (in Chinese)