

基于混沌序列的压缩感知语音增强算法

隋 昊,周 萍,沈 昊,杨 青

(桂林电子科技大学 电子工程与自动化学院,广西 桂林 541004)

摘 要: 利用语音在离散余弦变换域下的稀疏性,提出一种将混沌序列和符号函数相结合的观测矩阵,该观测矩阵具有良好的伪随机性,实现了确定性和随机性的统一. 首先将带噪的语音信号在 DCT 域进行稀疏,然后利用提出的观测矩阵进行观测,以保留大部分语音特性,最后在原有重构算法基础上改进重构算法,以加快重构速度. 此算法可以使重构语音信号的可懂度和清晰度得到大幅提升,实现了语音增强.

关键词: 压缩感知;稀疏表示;混沌序列;信号重构

中图分类号: TN912.34

文献标识码: A

文章编号: 1000-7180(2018)01-0096-04

Compressed Sensing Speech Enhancement Algorithm Based on Chaotic Sequences

SUI Hao,ZHOU Ping,SHEN Hao,YANG Qing

(Department of Electric Engineering and Automation,Guilin University of Electronic Technology,Guilin 541004,China)

Abstract: In this paper, present a kind of observation matrix which combines the chaotic sequence with the sign function, which has good pseudo randomness. Firstly, the noisy speech signal is sparse in the DCT domain, then the observation matrix is used to preserve most of the speech features, finally based on the original reconstruction algorithm, this paper improves the reconstruction algorithm to speed up the reconstruction. The algorithm proposed in this paper makes the intelligibility and clarity of reconstructed speech signal be greatly improved, and the speech enhancement is realized.

Key words: compressed sensing; sparse representation; chaotic sequence; signal reconstruction

1 引言

在实际生活中,无处不在的噪声影响着人们接收到的语音的质量,语音增强便是把背景噪声从噪声语音记录中去掉,让带噪语音的质量得以改善,可懂度以及清晰度都能得到提高,这不仅对人接收语音有帮助,而且也促进了机器识别人类语言进程的发展.

近年来兴起的由 Candes^[1]、Donoho^[2]等人提出的以信号稀疏性作为前提的信号压缩感知(CS)采样理论,突破 Nyquist 采样定理的两倍的条件限制,针对可稀疏表示的信号,只用远小于传统采样方法

数据量的测量值也能精确重构出不丢失主要数据的信号,达到语音增强的目的. 这样不仅加快了数据处理的速度,也优化了资源配置. 目前压缩感知主要应用于信息压缩、图像处理、无线传感网络等领域^[3],在语音信号的应用上还不是很广泛. 本文根据压缩感知的主要理论,将 CS 应用于语音信号的增强,采用改进的混沌序列作为观测序列,并且进一步优化重构恢复算法.

2 压缩感知理论

CS 主要由三部分^[4]组成:

(1)信号的稀疏表示;

收稿日期: 2017-03-11; **修回日期:** 2017-04-19

基金项目: 国家自然科学基金资助项目(61462017);广西省自然科学基金资助项目(2014GXNSFAA118353);广西自动检测技术与仪器重点实验室基金项目(YQ15110)

(2)观测矩阵的设计;

(3)信号重构.

2.1 稀疏表示

信号具有或经过变换具有稀疏性是使用压缩感知使用的前提条件,信号本身或者经过稀疏变换后,重构信号所用的观测数目就会变少,重构的效率也会变高.

设 $N \times 1$ 维的离散信号为 $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$, $N \times N$ 维的变换矩阵 $\Psi = [\psi_1, \psi_2, \dots, \psi_N]$, 其中 $\psi_i (i = 1, 2, \dots, N)$ 为 $N \times 1$ 的向量, 信号 $\mathbf{x}$ 可以由以下公式表示:

$$\mathbf{x} = \Psi \mathbf{a} = \sum_{i=1}^N \alpha_i \psi_i \quad (1)$$

式中, $\mathbf{a} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$ 为信号经过变换后的系数, $\mathbf{a}$ 中非零个数为 $\|\mathbf{a}\|_0 = K$, 如果 $K \ll N$, 则信号 $\mathbf{x}$ 为 K -稀疏, 这个过程就是对信号的稀疏表示.

2.2 观测矩阵

为了能够实现 $M \times N$ 维的观测矩阵 Φ 和变换矩阵 Ψ 是互不相关的矩阵, 并且为了保证从观测值重构出的信号足够精确, 观测矩阵必须要符合 RIP 准则(有限等距性). RIP 准则能够保证经过降维观测之后两个不同的 K -稀疏信号不会映射到同一集合中, 使得原信号空间与稀疏空间之间是一一对应的. 且 $K < M \ll N$, 对信号线性变换得到 M 维观测向量:

$$\mathbf{y} = \Phi \mathbf{x} \quad (2)$$

将式(1)代入式(2)得出:

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{a} = \mathbf{A}_s \mathbf{a} \quad (3)$$

式中, $\mathbf{y}$ 是 $M \times 1$ 维的向量 $\mathbf{y} = [y_1, y_2, \dots, y_M]^T$, $\mathbf{A}_s$ 为传感矩阵大小为 $M \times N$.

2.3 信号重构

当观测矩阵 Φ 符合 RIP 准则时, 信号重构的问题实质就变成了通过 l_0 范数求解的最优化问题:

$$\min_{\mathbf{a}} \|\mathbf{a}\|_{l_0} \quad \text{s. t.} \quad \mathbf{y} = \mathbf{A}_s \mathbf{a} \quad (4)$$

但是由于求解这个 l_0 范数是个 NP 问题, 因此一般都转化为 l_1 范数来解决.

3 基于混沌序列的自适应 CS 语音增强

理论指出语音信号本身并不是稀疏的, 但当对其进行离散余弦变换后就变为近似稀疏, 要应用 CS 技术, 首先要将信号进行稀疏变换. 本文主要采用最简单的离散余弦变换对语音信号进行稀疏, 然后使

用混沌序列和符号函数进行观测, 再对传统 OMP (正交匹配跟踪) 算法^[5]改进, 最后进行信号重构得到信噪比较高的语音, 实现了语音增强.

3.1 变换矩阵的选用

对语音信号, 目前用的比较多的变换矩阵就是正交矩阵如 DCT 基、小波基等, DCT 压缩性能较好, 而且 DCT 基属于固定正交基, 具有输入无关性, 所以普遍适合各种含噪语音^[6]. 一般来讲语音在 DCT 域下稀疏性良好, 而噪声在 DCT 域没有稀疏性. 研究表明, 纯净的语音信号在离散余弦变换基下具有近似稀疏性, 而 white 噪声和 pink 噪声都不是稀疏的.

3.2 Logistic 混沌序列作为观测矩阵

本文提出使用具有较好的伪随机性的 Logistic 混沌序列和符号函数优化矩阵构造测量矩阵.

单 Logistic 混沌映射的表达式为

$$x_{n+1} = x_n \times \mu \times (1 - x_n) \quad (5)$$

式中, $\mu \in [0, 4]$, $x \in (0, 1)$, μ 被称为混沌 (Logistic) 参数, 研究表明, 当 $x \in (0, 1)$ 时, Logistic 映射处于混沌状态, 即在 Logistic 映射作用下产生的序列既不是周期的也不是收敛的, 除此之外当 x 的取值不在上述范围时, 生成的序列收敛于某一个特定的值. 当 $3.569\ 945\ 6 < \mu \leq 4$ 的时候, 特别是 μ 无限靠近 4 时, 经过上述表达式迭代得到的值是符合伪随机状态分布的.

由上述 Logistic 混沌序列产生的混沌伪随机序列 $\{x_n\}$ 经过符号函数变换为 $\{a_n\}$, 将得到的序列取 N 长形成 $M \times N$ 维测量矩阵 Φ . 用此方法构造出的测量矩阵比随机矩阵更稳定, 如本文方法生成的混沌序列(图 1)和混度序列中 -1 和 1 的个数统计(图 2), 以生成长度为 1 000 的混沌序列为例可以看出, 本文方法构造出的矩阵 1 和 -1 的个数比趋近于为 1:1, 其他常见的高斯随机矩阵没有这个特点.

3.3 改进语音重构算法

当前流行的重构算法大多采用的是匹配追踪类的算法, 具有代表性的是 MP (Matching Pursuit, 匹配追踪法)、在 MP 的基础上对原子正交化处理而改进的 OMP^[7] (Orthogonal Matching Pursuit, 正交匹配追踪法)、在 OMP 基础上改进的 ROMP^[8] (Regularized Orthogonal Matching Pursuit, 正则化正交匹配追踪法). 但这些算法都要求稀疏度是已知的, 对于稀疏度无法确定的信号就不再适用, 本文提出的算法借鉴了文献[9]提出的稀疏度自适应匹配追踪 (Sparsity Adaptive Matching Pursuit, SAMP)

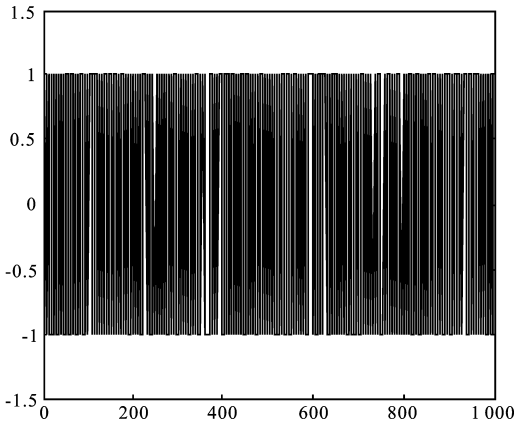


图1 生成的混沌矩阵

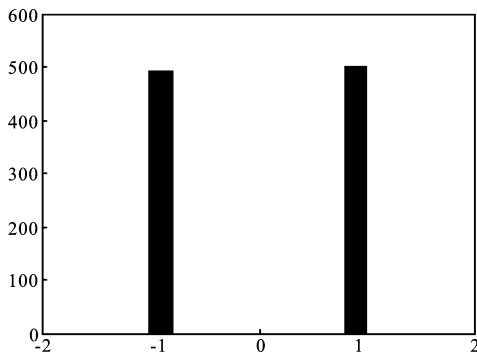


图2 混沌序列中1和-1个数统计图

算法,该算法不但不需要事先知道稀疏度,并且对传统 SAMP 的固定步长进行改进,设置可变步长.这种可变的步长,可以实现一步一步对信号稀疏度进行估计,又引入了 ROMP 的正则化过程,对原子进行二次筛选.由于 ROMP 是在 OMP 基础上进行改进的,比 OMP 重构花费时间更短,效果更好.

ROMP 算法主要通过实际观察值与估计值之间的余量 r 与传感矩阵 $\mathbf{A}_s$ 中各列内积,然后求模值得到相关系数 u :

$$u = \{u_j | u_j = |\langle r, \varphi_j \rangle|, j = 1, 2, \dots, N\} \quad (6)$$

并筛选出 K 个最大相关系数所对应的 $\mathbf{A}_s$ 的列序号构成集合进行正则化二次筛选.通过迭代计算后得到信号重建的支撑集,再采用最小二乘法对信号进行逼近和余量的更新:

$$\hat{x} = \operatorname{argmin} \|\mathbf{Y} - \Phi_{\Lambda} x\|_2 \quad (7)$$

$$r_{\text{new}} = \mathbf{Y} - \Phi_{\Lambda} \hat{x} \quad (8)$$

结合 ROMP 和 SAMP 的优势,并对其不足进行改进,提出的改进算法具体的操作步骤如下所示.

(1) 输入:初始余量 r_0 等于语音信号经过 DCT 转换后投影到混沌序列观测矩阵上得到的 M 维观

测向量 $\mathbf{y}$,步长 $\text{size} = 1$,阶段 $\text{stage} = 1$,迭代次数 $n = 1$,索引值集合 $\Lambda = \varphi$;

(2) 设定阈值 ϵ ,当 $\|r\|_2 \leq \epsilon$ 时,就停止迭代,若不符合条件继续步骤(3);

(3) 运用前面的式(6) 计算相关系数 u ,并在 u 中找出 size 个最大值,将其所对应的索引值存入 J 中;

(4) 正则化:对 J 中每个索引值所对应原子的相关系数进行正则化处理,结果中的原子相关系数必须满足式 $|g_i| \leq 2|g_j|, (i, j \in J)$,并将正则化得到的结果并入集合 Λ ,更新支撑集 Φ_{Λ} ;

(5) 应用 OLS 进行信号逼近并更新余量,利用式(7) 得到信号的估计值 $\hat{x}$,将 $\hat{x}$ 中前 size 个最大值所对应的索引值并入集合 Λ ,更新支撑集 Φ_{Λ} ,利用式(8) 更新余量;

(6) 当 $\|r_{\text{new}}\|_2 \geq \|r_{n-1}\|_2$ 时, $\text{stage} = \text{stage} + 1$, $\text{size} = \text{size} \cdot \text{stage}$,转到步骤(3) 执行,否则 $r = r_{\text{new}}, n = n + 1$,转到步骤(2) 执行.

(7) 最后将输出的信号估计 $\hat{x}$ 后面的元素补零得到稀疏系数向量,再通过 DCT 逆变换求出原始信号.

由上面的步骤很容易看出,实验过程中步长是自适应变化的,根据判断条件来决定迭代和阶段的跳转,因此该算法在不需要已知稀疏度 K 的前提下,就可以进行信号重建.因为步长可变,所以也不会因为固定步长的过小导致迭代多次依旧无法达到迭代条件或固定步长过大出现的过估计现象.正则化过程在保证重构信号的质量的同时又大大缩短了重构语音所需的时间.

3.4 低通滤波器

人说话所发出的声音能量主要集中在低频段,在高频部分的分量很少,而噪声在高频部分依旧大量存在,就导致了高频部分重构出的基本都是噪声信号,所以需要运用一个低通滤波器对重构后的语音进行平滑滤波,这样可以有效去除高频部分的噪声分量,改善了输出语音的可懂度和信噪比以及清晰度.传递函数离散形式表示为

$$H(z) = \frac{1-G}{1-Gz^{-1}} \quad (9)$$

式中,参数 G 取 0.9.

4 实验仿真与分析

仿真实验在 matlab 上实现,实验用的语音样本选自 CASIA 汉语语音库,采样率 16 kHz,在测试库

中选取 30 个不同说话人的语音,每人 10 句,共 300 句,每句时长大约 1~5 s 不等. 语音帧长 256,帧移 128,压缩比 0.5. 利用 MOS 值来衡量系统语音质量,让 30 位不同听众接受听觉感受的测试,听众各自根据在 matlab 上用本文算法处理过的语音的清晰度、可懂度以及噪声情况综合给出平均主观意见分,对比传统压缩感知方法的 DCT 变换下稀疏表示,采用随机的 Gauss 观测矩阵,传统 OMP 算法在不同信噪比下不同环境噪声打出的 MOS 记录下来如表 1 所示. 表 2 所示为输入信号在背景噪声分别为 White、Babble、Factory 下,SNR 为 0 dB 时,经过两种算法增强处理后达到的平均 SNR. 将一段在雨中录的带噪语音“桂林电子科技大学”,采用对比谱减法和本文方法进行语音增强,效果如图 3 所示.

表 1 平均主观意见分

输入 SNR/dB	传统压缩感知		本文算法	
	white	babble	white	babble
-5	2.7	2.4	3.1	2.9
0	3.6	2.8	3.8	3.2
5	3.5	3.0	3.9	3.7
10	3.9	3.6	4.1	4.0
15	4.0	3.8	4.2	4.1

表 2 语音增强后的平均信噪比

算法	环境噪声		
	White	Babble	Factory
传统算法	8.12	7.22	10.34
本文算法	10.12	9.78	12.36

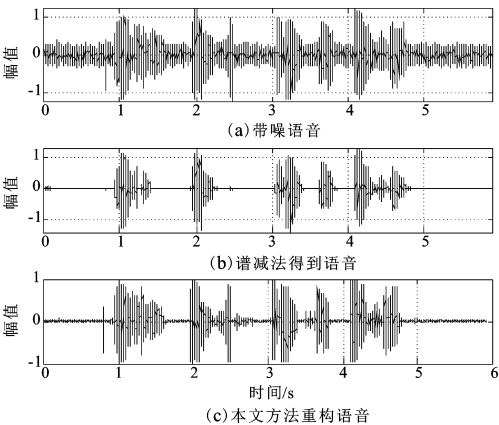


图 3 谱减法与本文方法去噪效果图

由表 1 和表 2 可以看出,本文提出的算法不管在白噪声还是在人谈论的背景噪声下都取得良好效果,相比传统压缩感知的算法,本文算法具有更高的可懂度和清晰度,即便是在低信噪比的条件下本文算法增强效果依旧十分明显. 在 Babble 背景下 MOS 分相对 White 低,这是因为 Babble 是以人说

话的声音作为背景噪声的,内部结构和语音信号基本一样,容易当成语音分量重构出来,而 Factory 是工厂工作的噪声,相对比较平稳,在重构过程中容易被区分出来. 由图 3 可以明显地看出,采用谱减法进行语音去噪增强时会出现音乐噪声和把语音当做噪声减去的现象,而经过本文的方法重构出的语音,不但去掉了噪声,而且不会出现把语音当做背景噪声去掉的现象.

5 结束语

本文通过分析传统压缩感知理论,对理论深入认识并改进,提出了基于符号函数和混沌序列构成的新的具有稳定性的观测矩阵. 将需要增强的语音经过 DCT 基稀疏,结合本文提出的混沌序列结合符号函数构造的观测矩阵进行观测,优化重构算法,区分噪声分量和语音分量来对语音信号进行重构恢复,抑制噪声从而实现语音增强. 实验表明在背景环境噪声不同或同种噪声不同信噪比的情况下,本文算法在语音增强方面都优于传统压缩感知算法以及传统的谱减法.

参考文献:

[1] Candes E J, Tao T. Near-optimal signal recovery from random projections: universal encoding strategies [J]. IEEE Transactions on Information Theory, 2006, 52 (12):5406-5425.

[2] Donoho D L. Compressed sensing [J]. IEEE Transactions on Information Theory, 2006, 52(4):1289-1306.

[3] 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.

[4] 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.

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

[6] 毛静,尹鹏亮,马琦,等. 自适应压缩感知的语音压缩重构算法研究[J]. 电子技术应用, 2015, 41(2):72-74.

[7] Tropp J, Gilbert A. Signal recovery from random measurements via orthogonal matching pursuit [J]. IEEE Transactions on Information Theory , 2008, 53 (12):4655 - 4666.

[8] Needell D, Vershynin R. Greedy signal recovery and un-certainty principles[C]// Proceedings of the Conference on Computational Imaging. San Jose, USA, SPIE, 2008: 1 - 12.

通过表 1 可以看出,将基 2 DIT-FFT 算法并行化映射到 4 个 PE 上时,加速比提升的不是很明显,这是由于 PE 的个数和访问存储、PE 间通信所消耗的时间没有达到一个理想的平衡点. 当映射到 8 个 PE 上时,加速比有了明显的提升,这说明 PE 的个数和访问存储、PE 间通信所消耗的时间基本达到了较理想的平衡点. 但是从 8 个 PE 增加到 16 个 PE 时,执行周期数的减少也开始减缓,因为随着使用 PE 个数的增多,访问存储和 PE 之间通信的时间也在加大,这会影响到最终的执行速度. 因此,就目前实验结果来看,使用 8 个 PE 是较理想的结果,这样可以在加快执行速度的同时尽量降低其他工作所消耗的时间.

5 结束语

本文首先对 FFT 算法的基本原理、基 2 DIT-FFT 算法的运算特点、及其并行点进行了分析和总结;其次提出了一种基于 PAAG 阵列机的基 2DIT-FFT 并行实现算法,最后该算法分别映射到 1 个 PE、4 个 PE、8 个 PE 和 16 个 PE 进行加速实现,并对实验结果进行了分析. 实验结果表明,随着点数的增加会使 PE 与 PE 之间、PE 与共享存储之间的访问也会增加,从而导致加速比下降,所以提高此类算法访存效率将是今后工作的重点.

参考文献:

[1] 丁玉美,高西全. 数字信号处理[M]. 2 版. 西安:西安电子科技大学出版社,2000.

[2] 兰海洋,林晓焕. 快速傅里叶变换的硬件实现[J]. 通信技术,2012, 45(8): 115-117.

[3] Chang Y N, Parhi K K. An efficient pipelined FFT architecture[J]. IEEE Transactions on Circuits & Systems II Analog & Digital Signal Processing, 2003, 50 (6):322-325.

[4] Baas B M. A low-power, high-performance, 1024-Point FFT Processor[J]. IEEE Journal of Solid-State Circuits ,1999, 34(3):380-387.

[5] 方维,孙广中,吴超,等. 一种三维快速傅里叶变换并行算法[J]. 计算机研究与发展, 2011, 48 (3): 440-446.

[6] Shin M, Lee H. A high-speed four-parallel radix-2 4, FFT/IFFT processor for UWB applications [C] // IEEE International Symposium on Circuits and Systems, IEEE, 2008:960-963.

[7] Huang S J, Chen S G. A high-throughput radix-16 FFT processor with parallel and normal input/output ordering for IEEE 802. 15. 3c systems[J]. Circuits & Systems I Regular Papers IEEE Transactions on, 2012, 59(8):1752-1765.

[8] 李涛,杨婷,易学渊,等. 萤火虫 2:一种多态并行机的硬件体系结构[J]. 计算机工程与科学, 2014, 36(2): 191-200.

[9] 李涛,肖灵芝. 面向图形和图像处理的轻核阵列机结构[J]. 西安邮电大学学报, 2012, 17(3):41-47.

[10] Cooley J W, Tukey J W. An algorithm for the machine calculation of complex Fourier series[J]. Mathematics of Computation, 1965, 19(19):297-301.

[11] LI T. PAAG: A polymorphic array architecture for graphics and image processing [C] // PAAP'2012, IEEE Computer Society CPS. IEEE,2012:242-249.

[12] Marowka A, Ramat Gan. Back to thin-core massively parallel processors[J]. IEEE Computer, 2011,44(12): 49-54.

作者简介:

常立博 男,(1985-),博士研究生,助教. 研究方向为计算机视觉、计算机系统结构. E_mail:chang_libo@163. com.

王时雨 男,(1984-),博士研究生. 研究方向为图像处理、计算机系统结构.

(上接第 99 页)

[9] Do T T, Lu Gan, Nguyen N, et al. Sparsity adaptive matching pursuit algorithm for practical compressed sensing[C] // Asilomar Conference on Signals, Systems and Computers. Pacific Grove, California: IEEE Press,2008:581-587.

作者简介:

隋 昊 女,(1992-),硕士研究生. 研究方向为语音识别. E-mail:Emily_Sui@163. com.

周 萍 女,(1961-),硕士,教授. 研究方向为语音识别与智能控制.

沈 昊 男,(1992-),硕士研究生. 研究方向为语音识别.

杨 青 女,(1976-),硕士,副教授. 研究方向为语音识别算法与仿真.