

基于自适应冗余字典的语音信号稀疏表示算法

王天荆^{*①②} 郑宝玉^① 杨震^①

^①(南京邮电大学宽带无线通信与传感网技术教育部重点实验室 南京 210003)

^②(南京工业大学理学院 南京 210009)

摘要: 基于冗余字典的信号稀疏表示是一种新的信号表示理论,当前的理论研究主要集中在字典构造算法和稀疏分解算法两方面。该文提出一种新的基于自适应冗余字典的语音信号稀疏表示算法,该算法针对自相关函数为指数衰减的平稳信号,从K-L展开出发,建立了匹配信号结构的冗余字典,进而提出一种高效的基于非线性逼近的信号稀疏表示算法。实验结果表明冗余字典中原子的自适应性和代数结构使短时平稳语音信号稀疏表示具有较高的稀疏度和较好的重构精度,并使稀疏表示算法较好地应用于语音压缩感知理论。

关键词: 语音信号处理; 压缩感知; 稀疏表示; K-L展开; 冗余字典

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2011)10-2372-06

DOI: 10.3724/SP.J.1146.2011.00125

A Speech Signal Sparse Representation Algorithm Based on Adaptive Overcomplete Dictionary

Wang Tian-jing^{①②} Zheng Bao-yu^① Yang Zhen^①

^①(Key Lab of "Broadband Wireless Communication and Sensor Network Technology", Ministry of Education, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)

^②(School of Science, Nanjing University of Technology, Nanjing 210009, China)

Abstract: The sparse representation based on overcomplete dictionary is a new signal representation theory. Recent activities in this field concentrate mainly on the study of dictionary design algorithm and sparse decomposition algorithm. In this paper, a novel speech signal sparse representation algorithm is proposed based on adaptive overcomplete dictionary. Considering stationary signal with autocorrelation function of exponential decay, an adaptive overcomplete dictionary is constructed in terms of the Karhunen-Loève (K-L) expansion. Furthermore, an effective algorithm based on the nonlinear approximation is proposed to obtain sparse decomposition of signal with the adaptive dictionary. The experimental results indicate that short-term stationary speech signal sparse representation based on the adaptability and algebraic structure of atom in the overcomplete dictionary has higher sparsity and better reconstructive precision. The sparse representation algorithm can preferably be used in compressed sensing.

Key words: Speech signal processing; Compressed Sensing (CS); Sparse representation; Karhunen-Loève (K-L) expansion; Overcomplete dictionary

1 引言

在信号处理中,用空间变换有效地表达信号,可提高压缩效率。传统的信号表示是基于“正交基”的展开,如DCT,小波变换等,但这种分解不总能够达到较好的效果,尤其是对时频变化范围很广的信号效果更差。一种更好的信号表示方式应是根据

信号的特点,自适应地选择基函数。文献[1,2]提出信号在过完备原子库(或称为冗余字典)上进行稀疏分解的思想,该原子库比通常的完备正交基大得多,在信号所表示的空间中足够密。因此,信号稀疏表示的结果就是信号线性展开中大部分基函数的系数为零,只有少数基函数具有较大的非零系数。这里称基函数为原子,其集合为字典。信号稀疏表示具有由可揭示信号主要特征与内在结构的少量原子来表示信号的优良特性,因此在信号去噪^[3]、特征识别^[4]、弱信号提取^[5]和数据压缩^[6]等领域得到了广泛应用。针对稀疏分解,已有MP(Matching Pursuit)^[1],

2011-02-21收到, 2011-07-22改回

国家自然科学基金(60971129), 国家973计划项目(2011CB302903), 江苏省自然科学基金(BK2011238), 中国博士后科学基金(20100481167)和江苏省博士后科学基金(1101022B)资助课题

*通信作者: 王天荆 tianjingwang@njupt.edu.cn

BOB(Basis Orthogonal Best)^[7], BP (Basis Pursuit)^[8]和MOF (Method of Frames)^[9]等多种算法, 其中MP算法最常用。

2004年Donoho^[10]建立了一种“边感知边压缩”的新兴采样技术——压缩感知(Compressed Sensing, CS), 该理论将信号稀疏表示提升到了一个新的高度。CS表明: 只要未知信号在某一个空间具有稀疏性, 就能以低于奈奎斯特采样定理要求的频率进行随机采样, 并通过求解优化问题(从少量采样信号)称为观测)中以高概率重构出原信号。因此, 信号稀疏表示有极其重要的理论意义和广泛的应用价值。

信号稀疏表示的一个关键问题就是如何设计有效的冗余字典。当前已有小波包、小波和正弦函数的级联、局部余弦等多种冗余字典, 但它们不总能保证信号的稀疏性^[11]。文献[12]选取可分离Gabor函数作为语音原子库, 但离散Gabor函数中多个时频参数所得的原子数量巨大, 增加了复杂度。近两年, 通过学习、训练来获得冗余字典的方法也得到了发展, 如文献[13]将聚类法用于图像字典学习; Aharon等人^[14]将K均值聚类法推广为K-SVD, 自适应更新字典。但这些方法需通过训练大量样本来更新字典, 计算量和存储空间巨大。文献[15]提出了基于模板匹配近似KLT域上的语音信号压缩感知重构算法, 算法为了避免对信号求取KLT分解矩阵, 采用训练语音自相关矩阵的非齐次线性组合来近似语音的自相关矩阵, 增大了表示难度; 同时, 清音帧信号在近似KLT域上不具有稀疏性, 不利于实际应用。

本文研究基于K-L展开的自适应冗余字典构造算法及平稳信号稀疏分解算法, 由此获得了短时平稳语音信号的稀疏表示, 并将之应用于语音压缩感知。首先, 针对自相关函数为指数衰减型的连续平稳信号, 提出了自适应正交K-L基的构建算法, 通过离散正交基获得自适应冗余字典; 然后, 验证了语音信号在此字典上的稀疏性及重构性能; 最后, 实现了语音的CS压缩采样和重构。实验结果表明自适应冗余字典的代数结构不仅保证了语音信号的稀疏性, 而且大幅度缩减了稀疏分解时所需的原子个数, 稀疏表示算法提高了稀疏分解效率, 在语音压缩感知中显示了优良的性能。

2 基于K-L展开的自适应冗余字典构造算法

2.1 基于K-L展开的自适应正交K-L基的构建

K-L 展开提供了从独立随机变量集合的角度研究随机过程的可能性^[16], 其思想在于把随机过程展开为一列正交函数的线性组合形式。不失一般性, 假设连续实随机过程 $\{x(t), t \in [0,1]\}$ 是二阶矩过程,

则其 K-L 展开式为

$$x(t) = \sum_{n=1}^{\infty} c_n \varphi_n(t) \quad (1)$$

其中系数 $c_n = \int_0^1 x(t) \varphi_n(t) dt$; 正交K-L基 $\{\varphi_n(t)\}$ 是自相关函数 $R_x(t, u)$ 的特征函数组, $\varphi_n(t)$ 和与之对应的特征值 $\lambda \in R$ 满足Fredholm积分方程:

$$\lambda \varphi_n(t) = \int_0^1 R_x(t, u) \varphi_n(u) du \quad (2)$$

通常, 取式(1)中有限的 N 项作为 $x(t)$ 线性逼近, 即

$$\tilde{x}(t) = \sum_{n=1}^N c_n \varphi_n(t) \quad (3)$$

K-L 展开的关键是获得 $R_x(t, u)$ 的特征值和特征函数。针对平稳随机过程的自相关函数在时延较小时较快衰减的特性, 本文考虑 $R_x(t, u) = r_x(0)e^{-\mu|t-u|}$, 其中参数 μ 反映了衰减速度, 并由式(2)求解如下方程:

$$\begin{aligned} \lambda \varphi(t) &= \int_0^1 R_x(t, u) \varphi(u) du \\ &= r_x(0) \left(e^{-\mu t} \int_0^t e^{\mu u} \varphi(u) du + e^{\mu t} \int_t^1 e^{-\mu u} \varphi(u) du \right) \end{aligned} \quad (4)$$

当 $t = 0$ 和 $t = 1$ 时, 由式(4)分别有

$$\left. \begin{aligned} \lambda \varphi(0) &= r_x(0) \int_0^1 e^{-\mu u} \varphi(u) du \\ \lambda \varphi(1) &= r_x(0) \int_0^1 e^{-\mu(1-u)} \varphi(u) du \end{aligned} \right\} \quad (5)$$

对式(4)两边求导可得

$$\begin{aligned} \lambda \varphi'(t) &= r_x(0) \left(-\mu e^{-\mu t} \int_0^t e^{\mu u} \varphi(u) du + \mu e^{\mu t} \int_t^1 e^{-\mu u} \varphi(u) du \right) \end{aligned} \quad (6)$$

当 $t = 0$ 和 $t = 1$ 时, 由式(6)分别有

$$\left. \begin{aligned} \lambda \varphi'(0) &= r_x(0) \mu \int_0^1 e^{-\mu u} \varphi(u) du \\ \lambda \varphi'(1) &= -r_x(0) \mu \int_0^1 e^{-\mu(1-u)} \varphi(u) du \end{aligned} \right\} \quad (7)$$

由式(5), 式(7)可知

$$\varphi'(0) = \mu \varphi(0), \quad \varphi'(1) = -\mu \varphi(1) \quad (8)$$

对式(6)两边求导可得

$$\begin{aligned} \lambda \varphi''(t) &= r_x(0) \left(\mu^2 \int_0^t e^{-\mu(t-u)} \varphi(u) du \right. \\ &\quad \left. + \mu^2 \int_t^1 e^{-\mu(u-t)} \varphi(u) du - 2\mu \varphi(t) \right) \\ &= (\mu^2 \lambda - 2r_x(0)\mu) \varphi(t) \end{aligned} \quad (9)$$

当 $\mu^2 \lambda - 2r_x(0)\mu < 0$ 时, 有 $\varphi''(t) + \omega^2 \varphi(t) = 0$, 其通解为 $\varphi(t) = c_1 \cos(\omega t) + c_2 \sin(\omega t)$ ($c_1, c_2 \in R$), 其中 $\omega = \sqrt{[2r_x(0)\mu - \mu^2 \lambda] / \lambda}$ 。联合式(8), 方程特解为 $\varphi_n(t) = (n\pi/\mu) \cos(n\pi t) + \sin(n\pi t)$ ($n \in Z$)。因 $\varphi_0(t)$

= 0 不能作为特征函数, 则正交特征函数组为

$$\varphi_n(t) = \frac{n\pi}{\mu} \cos(n\pi t) + \sin(n\pi t), \quad n \in Z - \{0\} \quad (10)$$

加入 $\varphi_0(t) = 1$, 构建完备的自适应正交 K-L 基函数组 $E_1 = \{\varphi_0(t)\} \cup \{\varphi_n(t), n \in Z - \{0\}\}$, 其中

$$\langle \varphi_n(t), \varphi_m(t) \rangle = \int_{-\infty}^{+\infty} \varphi_n(t) \varphi_m(t) dt = 0, \\ n \neq m, \forall n, m \in Z$$

2.2 基于K-L展开的自适应冗余字典构造算法

针对数字信号处理, 需要离散正交K-L基。首先, 确定式(10)中参数 μ 。信号 $\mathbf{x} \in R^M$ 在时延 τ 的自相关函数无偏估计为 $R_x(\tau) = 1/(M - \tau) \cdot \sum_{i=1}^{M-\tau} x(i)x(i + \tau)$, 通过求解

$$\min_{\mu} G(\mu) = \|R_x - \hat{R}_x\|_2^2 = \sum_{\tau} (R_x(\tau) - \hat{R}_x(\tau))^2 \quad (11)$$

来获得最优解 μ^* , 其中 $\hat{R}_x(\tau) = r_x(0)e^{-\mu|\tau|}$ 。然后, 对 $\varphi_n(t)$ 在 $0 \leq t \leq 1$ 进行均匀采样获得原子 $\mathbf{e}_n = [e_n(1), \dots, e_n(i), \dots, e_n(M)]^T$, 其中 $e_n(i) = n\pi/\mu \cdot \cos\left(\frac{n\pi(i-1)}{M-1}\right) + \sin\left(\frac{n\pi(i-1)}{M-1}\right)$ ($i = 1, \dots, M$), 再联合 $\mathbf{e}_0 = [1, \dots, 1]^T$, 就可构建自适应冗余字典 $E_2 = \{\mathbf{e}_0\} \cup \{\mathbf{e}_n, n \in Z - \{0\}\}$ 。如果遍历 n 所有取值, 则会产生非常庞大的字典。依据原子的几何结构特性, 易见 $\mathbf{e}_{-n} = -\mathbf{e}_n$, 则可将 E_2 缩减为 $E_3 = \{\mathbf{e}_0\} \cup \{\mathbf{e}_n, n \in Z^+\}$, 大幅度减少了原子个数, 以提高稀疏分解效率。

3 基于自适应冗余字典的语音信号稀疏表示

3.1 信号稀疏表示理论

信号稀疏表示源于非线性逼近理论^[17], 即给定一个字典 $D = \{\mathbf{e}_n, n = 1, \dots, N\}$, 其原子可张成整个 Hilbert 空间 $H = R^M$ ($M \leq N$)。在 D 中自适应选取 K 个原子对 $\mathbf{x} \in R^M$ 作非线性逼近

$$\tilde{\mathbf{x}} = F_K \mathbf{x} = \sum_{n \in I_K} \alpha(n) \mathbf{e}_n = \sum_{n \in I_K} \frac{\langle \mathbf{x}, \mathbf{e}_n \rangle}{\langle \mathbf{e}_n, \mathbf{e}_n \rangle} \mathbf{e}_n \quad (12)$$

其中 $\text{Card}(I_K) = K$, $\boldsymbol{\alpha} = [\alpha(1), \dots, \alpha(N)]^T$ 为分解系数向量。逼近绝对误差为 $\varepsilon_K = \|\mathbf{x} - F_K \mathbf{x}\|_2$ 。当 D 是正交基时, 保留 $|\alpha(n)|$ 最大的 K 个基原子, 可得 \mathbf{x} 的最佳 K 项逼近; 当 D 是冗余字典时, 式(12)有多种解答, 信号稀疏表示就是选出分解系数最为稀疏的一个, 这是一个 NP 问题。如果要在冗余字典中获得较优的非线性逼近结果, 可考虑相干系数 $\beta =$

$\sup_{i \neq j} \left| \frac{\langle \mathbf{e}_i, \mathbf{e}_j \rangle}{\|\mathbf{e}_i\| \|\mathbf{e}_j\|} \right|$ 小于某个常数的非相干字典, 其原子间的线性相关性较弱。因 E_3 中原子是离散正交 K-L 基的矢量, 原子间的 β 较小, 即 E_3 为非相干字典, 可利用最佳 K 项逼近获得近似信号 $\tilde{\mathbf{x}}$ 。

3.2 基于自适应冗余字典的语音信号稀疏表示

语音信号具有短时平稳性, 其自相关函数有衰减特性, 所以可利用上述字典来研究它的稀疏表示。我们对采样率为 16 kHz, 长为 320 样点的清音帧、浊音帧和过渡帧信号(图1(a), 1(d), 1(g))进行分解, 发现清音帧信号在 DCT 域上不具有稀疏性(图1(c)), 浊音帧和过渡帧信号的稀疏性也不理想(图1(f), 1(i)), 而它们在自适应冗余字典上都有显著的近似稀疏性(图1(b), 1(e), 1(h))。

由原子的代数式可见: 当 n 增大时, \mathbf{e}_n 中各元素值也增大, 从而迫使语音信号在分解时将能量主要集中在 n 较小的原子上, 即使得分解系数具有近似稀疏性。由图1(b), 1(e), 1(h) 易见, 保留绝对值最大的 K 个 $\alpha(n)$ ($n \in I_K$), 并令 $\alpha(n) = 0$ ($n \notin I_K$), 可得到稀疏分解向量 $\boldsymbol{\alpha}$, 因此无需用优化算法就可进行稀疏表示, 大大降低了计算复杂度, 便于将稀疏表示应用于 CS。

3.3 CS 中基于 MP 的信号重构算法

CS 认为若信号 \mathbf{x} 在变换矩阵 $\boldsymbol{\Psi}$ 上具有稀疏性, 则可用 $I = cK$ (通常 $c = 3 \sim 4$) 个观测 $\mathbf{y} = [y(1), \dots, y(I)]^T$ 以高概率重构出原信号, 即求解如下 l_0 优化问题

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0, \quad \text{s.t. } \mathbf{y} = \boldsymbol{\Phi} \mathbf{x} = \boldsymbol{\Phi} \boldsymbol{\Psi} \boldsymbol{\alpha} = \boldsymbol{\Xi} \boldsymbol{\alpha} \quad (13)$$

其中 $\boldsymbol{\Phi} = [\varphi_1, \dots, \varphi_M]$ ($\varphi_i \in R^I, i = 1, \dots, M$) 服从正态随机分布, 且与 $\boldsymbol{\Psi}$ 不相关; $\|\boldsymbol{\alpha}\|_0 = K$ 。基于自适应冗余字典的讨论, 不妨设 $N = M$, 则 $\boldsymbol{\Psi} \in R^{M \times M}$ 。

因 l_0 问题是 NP 难题, Donoho 等提出将之转化为 l_1 问题(称为 BP 算法), 通过线性规划求解 $\boldsymbol{\alpha}$, 并重构信号 $\tilde{\mathbf{x}} = \boldsymbol{\Psi} \boldsymbol{\alpha}$ 。但 BP 算法计算量巨大, 若采用 MP 算法可降低计算复杂度。本文基于 MP 的信号重构算法如下:

首先求解 \mathbf{y} 在冗余字典 $\boldsymbol{\Xi} = \{\mathbf{g}_\gamma, \gamma = 1, \dots, M\}$ 上的稀疏分解, 即从 $\boldsymbol{\Xi}$ 中选出与 \mathbf{y} 最为匹配的最佳原子 \mathbf{g}_{γ_1} , 满足 $\left| \frac{\langle \mathbf{y}, \mathbf{g}_{\gamma_1} \rangle}{\langle \mathbf{g}_{\gamma_1}, \mathbf{g}_{\gamma_1} \rangle} \right| = \sup_{\gamma} \left| \frac{\langle \mathbf{y}, \mathbf{g}_{\gamma} \rangle}{\langle \mathbf{g}_{\gamma}, \mathbf{g}_{\gamma} \rangle} \right|$, 于是 \mathbf{y} 分

解为在 \mathbf{g}_{γ_1} 上的分量和残余两部分 $\mathbf{y} = \frac{\langle \mathbf{y}, \mathbf{g}_{\gamma_1} \rangle}{\langle \mathbf{g}_{\gamma_1}, \mathbf{g}_{\gamma_1} \rangle} \cdot \mathbf{g}_{\gamma_1} + R^1 \mathbf{y}$ 。对残余可不断进行上面的分解过程, 即

$R^k \mathbf{y} = \frac{\langle R^k \mathbf{y}, \mathbf{g}_{\gamma_k} \rangle}{\langle \mathbf{g}_{\gamma_k}, \mathbf{g}_{\gamma_k} \rangle} \mathbf{g}_{\gamma_k} + R^{k+1} \mathbf{y}$, 其中 \mathbf{g}_{γ_k} 满足

$\left| \frac{\langle R^k \mathbf{y}, \mathbf{g}_{\gamma_k} \rangle}{\langle \mathbf{g}_{\gamma_k}, \mathbf{g}_{\gamma_k} \rangle} \right| = \sup_{\gamma} \left| \frac{\langle R^k \mathbf{y}, \mathbf{g}_{\gamma} \rangle}{\langle \mathbf{g}_{\gamma}, \mathbf{g}_{\gamma} \rangle} \right|$ 。经过 K 步迭代后, 有

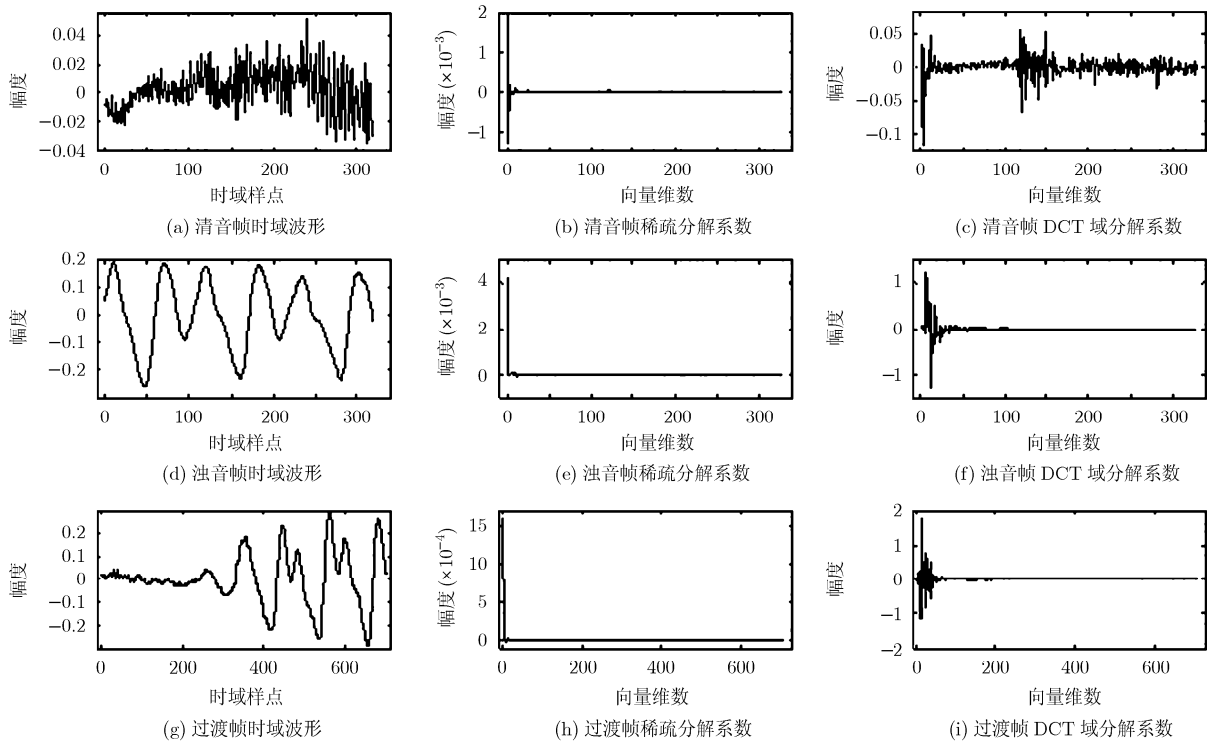


图1 清音帧、浊音帧和过渡帧信号的时域波形及分解系数图

$$\mathbf{y} = \sum_{k=1}^K \frac{\langle R^k \mathbf{y}, \mathbf{g}_{\gamma_k} \rangle}{\langle \mathbf{g}_{\gamma_k}, \mathbf{g}_{\gamma_k} \rangle} \mathbf{g}_{\gamma_k} + R^{K+1} \mathbf{y} \quad (14)$$

由于 $\|R^k \mathbf{y}\|$ 的衰减特性, 用少数原子可表示 \mathbf{y} 的主要成分, 即 $\mathbf{y} \approx \sum_{k=1}^K \frac{\langle R^k \mathbf{y}, \mathbf{g}_{\gamma_k} \rangle}{\langle \mathbf{g}_{\gamma_k}, \mathbf{g}_{\gamma_k} \rangle} \mathbf{g}_{\gamma_k}$, 并可得稀疏系数向量 $\boldsymbol{\alpha}$, 重构信号为 $\tilde{\mathbf{x}} = \boldsymbol{\Psi} \boldsymbol{\alpha}$ 。

4 实验结果及分析

4.1 语音信号的重构

实验环境为安静环境, 对象为两位男性和两位女性。为了说明稀疏分解的质量, 定义稀疏度为 $D = K/M$ 。针对图1, 利用基于自适应冗余字典的稀疏表示算法(Signal Sparse Representation algorithm based on Adaptive Overcomplete Dictionary,

SSRAOD)分别得到3个信号的 $K = [MD]$ ($[\cdot]$ 为取整函数)个系数, 然后进行信号重构。清音帧信号的稀疏性弱于浊音帧和过渡帧信号, 所以需要较多的原子来重构信号。当 $D = 40\%$ 时, 可重构出清音帧信号的绝大部分细节特征(图2(a)), 逼近绝对误差 $\varepsilon_K = 0.0049$ 。浊音帧信号在 $D = 15\%$ 时, $\varepsilon_K = 0.0114$ (图2(b)), 这说明仅用48个原子就可很好地重构出原信号, 达到了稀疏分解的目的。过渡帧信号同时具有清音和浊音信号的特征, 当 $D = 15\%$ 时, 也可精确重构出原信号, $\varepsilon_K = 0.0105$ (图2(c))。图2说明SSRAOD具有较高的稀疏度和较好的重构精度, 且有较高的计算效率, 如表1。当 D 相同时, SSRAOD所花费的计算时间远远少于MP和BP算法。当问题规模较大时, BP算法处理时延较长, 且易于陷入局部最优解, 产生较大的重构误差。

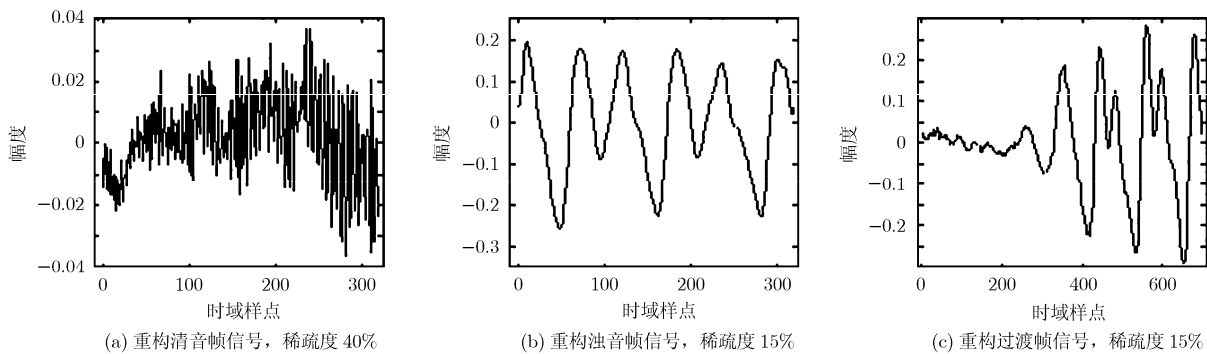


图2 清音帧、浊音帧和过渡帧信号的信号重构波形图

表1 清音帧、浊音帧和过渡帧信号稀疏分解时间和逼近绝对误差比较

信号	SSRAOD		MP		BP	
	$t(s)$	ε_K	$t(s)$	ε_K	$t(s)$	ε_K
清音帧 ($D = 40\%$)	0.0099	0.0049	1.9644	0.0050	12.1704	0.0170
浊音帧 ($D = 15\%$)	0.0092	0.0114	0.3880	0.0095	10.2422	200.11
过渡帧 ($D = 15\%$)	0.0181	0.0105	3.6702	0.0104	11.4159	0.4687

下面实验针对测试语音分别为女声和男声的“大规模集成电路”，采用矩形窗分帧，帧间无重叠，帧长为320样点/帧，定义重构语音信号平均分段信噪比SegSNR (dB):

$$\text{SegSNR} = \frac{1}{N_{\text{frame}}} \sum_{i=1}^{N_{\text{frame}}} 10 \times \lg \left(\frac{\mathbf{x}_i^T \mathbf{x}_i}{(\mathbf{x}_i - \tilde{\mathbf{x}}_i)^T (\mathbf{x}_i - \tilde{\mathbf{x}}_i)} \right) \quad (15)$$

其中 N_{frame} 为信号的总帧数。将重构信号按信噪比不大于 0 dB, 0~10 dB, 10~20 dB, 20~50 dB, 大于 50 dB 分别划分为重构质量很差、差、中等、好、很好。表 2 给出了不同稀疏度对应的信号重构质量, 当 D 越高时, SegSNR 越高。为了重构出信号时域波形的细节, 女声和男声分别由表 2 选择 D 为 14% 和 12%, 如图 3 和图 4。通常女性基频明显高于男性, 所以女声比男声需要更多原子进行信号重构, 以减少重构误差。

4.2 基于自适应冗余字典的语音压缩感知

基于图3和图4考虑基于自适应冗余字典的语音压缩感知。图5通过 $I = 4K = 4[MD]$ 个随机观测, 根据本文基于MP的重构算法, 获得女声和男声的

CS重构信号, 其时域波形图与原信号基本一致, SegSNR分别为22.8444 dB和36.3450 dB。比较图3、图4和图5, 易见: 当 D 相同时, 语音信号的CS重构性能略差于信号稀疏分解的重构性能, 这是因为它既包含了稀疏分解的逼近误差又包含了 \mathbf{y} 基于MP的重构误差, 若适当增加观测个数, 就可提高SegSNR。

由图5可知, 信号稀疏表示是CS的前提, 特别是观测个数 I 与分解原子数 K 有密切关系。图6(a)着重研究 D 与SegSNR的关系, 测试语音为由上述4位说话人分别随机选取10句短语, 共40句短语组成的一长段语音。当 K 固定时, $I = 4K$ 的SegSNR明显高于 $I = 3.5K$ 和 $I = 3K$, 所以增加 I 的确可以提高重构精度。在CS中, 通常用压缩率 $R = I/M$ 作为数据压缩性能优劣的标准, 易知 $R = cK/M = cDM/M = cD$ 。图6(b)反映了 R 影响着语音的CS重构性能, 当 R 越大, SegSNR越高, 重构效果越好。另一方面, R 相同时, $c = 3$ 的SegSNR高于 $c = 3.5$ 和 $c = 4$ 。这是因为 c 越小, D 越高, 重构时所用的原子越多, 重构误差就越小, 此与表2结论一致。

表2 女声和男声信号重构SegSNR(dB)

女声	$D(\%)$	6	7	9	11	12	13	14	15
	SegSNR	19.831	21.346	24.587	27.028	28.458	29.501	30.402	31.649
男声	$D(\%)$	3	4	5	7	9	11	12	13
	SegSNR	19.593	23.829	28.494	32.916	36.942	39.734	41.419	42.441

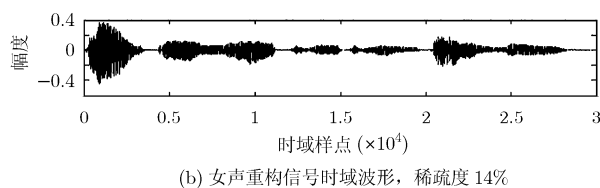
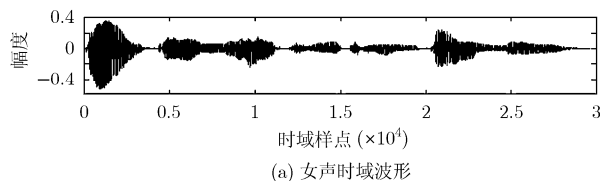


图3 女声语音信号及重构信号时域波形图

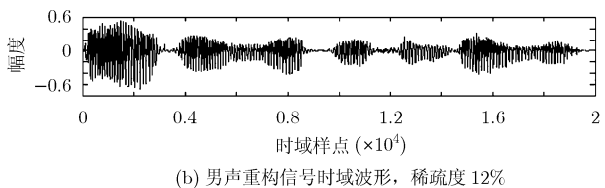
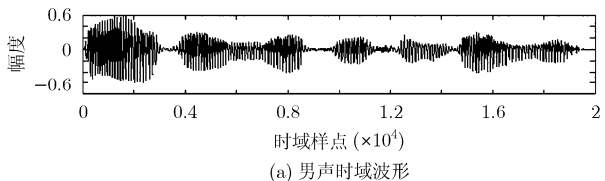


图4 男声语音信号及重构信号时域波形图

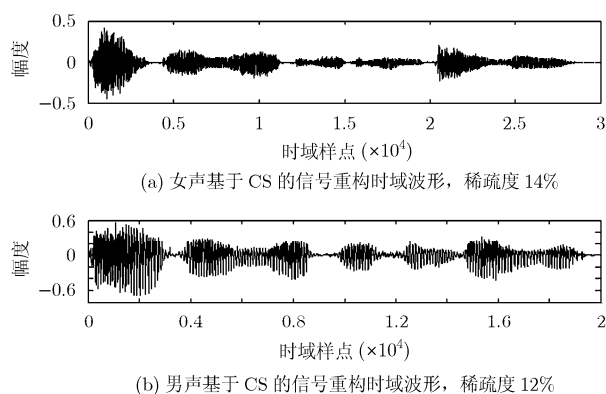


图5 女声和男声基于压缩感知的信号重构时域波形图

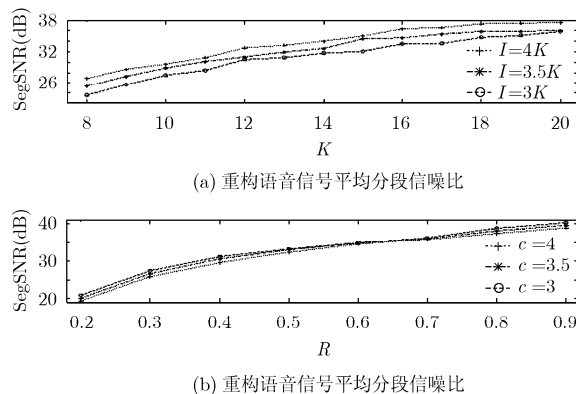


图6 语音压缩感知的信号重构性能图

5 结束语

本文依据K-L展开, 通过求解指数衰减型自相关函数的Fredholm积分方程, 构建了自适应冗余字典, 并由字典的代数结构设计了基于非线性逼近的信号稀疏表示算法, 降低了计算复杂度。实验验证了语音信号基于自适应冗余字典的稀疏性, 测试了基于自适应冗余字典的语音压缩感知的性能, 从而说明了本文算法的有效性、可行性和应用性。

参考文献

- [1] Mallat S and Zhang Z. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 1993, 41(12): 3397-3415.
- [2] Bergeau F and Mallat S. Matching pursuit of images. In: *Proceedings of IEEE-SP, USA*, 1994: 330-333.
- [3] 成萍, 赵家群, 司锡才, 等. 基于稀疏表示的被动毫米波L-R成像算法. *电子与信息学报*, 2010, 32(7): 1706-1711. Cheng P, Zhao J Q, Si X C, et al. L-R imaging algorithm for passive millimeter wave based on sparse representation. *Journal of Electronics & Information Technology*, 2010, 32(7): 1706-1711.
- [4] Wright J and Yang A Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227.
- [5] 傅霆, 尧德中. 稀疏分解的加权迭代方法及其初步应用. *电子学报*, 2004, 32(4): 567-570. Fu T and Yao D Z. Iterative weighted method of sparse decomposition and preliminary application. *Acta Electronica Sinica*, 2004, 32(4): 567-570.
- [6] Mairal J and Bach F. Non-local sparse models for image restoration. *IEEE International Conference of Computer Vision*. Kyoto, Japan, 2009: 2272-2279.
- [7] Coifman R and Wickerhauser M. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 1992, 38(2): 1713-1716.
- [8] Chen S and Donoho D L. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 1999, 20(1): 33-61.
- [9] Daubechies I. Time-frequency localization operators: a geometric phase space approach. *IEEE Transactions on*

Information Theory, 1988, 34(4): 605-612.

- [10] Donoho D L. Compressed sensing. *IEEE Transactions on Information Theory*, 2006, 52(4): 1289-1306.
- [11] 孙玉宝, 肖亮. 基于Gabor感知多成份字典的图像稀疏表示算法研究. *自动化学报*, 2008, 34(11): 1379-1386. Sun Y B and Xiao L. Sparse representations of images by a multi-component Gabor perception dictionary. *Acta Automatica Sinica*, 2008, 34(11): 1379-1386.
- [12] 井爱雯, 刘云. 基于MP算法的语音信号稀疏分解. *计算机工程与应用*, 2009, 45(5): 144-146. Jing A W and Liu Y. Speech signal sparse decomposition based on matching pursuit algorithm. *Computer Engineering and Applications*, 2009, 45(5): 144-146.
- [13] Rubinstein R and Elad M. Double sparsity: learning sparse dictionaries for sparse signal approximation. *IEEE Transactions on Signal Processing*, 2010, 58(3): 1553-1564.
- [14] Aharon M, Elad M, and Bruckstein A M. The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 2006, 54(11): 4311-4322.
- [15] 郭海燕, 杨震. 基于近似KLT域的语音信号压缩感知. *电子与信息学报*, 2009, 31(12): 2948-2952. Guo H Y and Yang Z. Compressed speech signal sensing based on approximate KLT. *Journal of Electronics & Information Technology*, 2009, 31(12): 2948-2952.
- [16] 李杰, 刘章军. 基于标准正交基的随机过程展开方法. *同济大学学报*, 2006, 34(10): 1279-1283. Li J and Liu Z J. Expansion method of stochastic processes based on normalized orthogonal bases. *Journal of Tongji University (Natural Science)*, 2006, 34(10): 1279-1283.
- [17] Cohen A and Dales J. Nonlinear approximation of random functions. *SIAM Journal on Applied Mathematics*, 1997, 57(2): 518-540.

王天荆: 女, 1977年生, 讲师, 博士后, 研究方向为信号处理和无线传感器网络。

郑宝玉: 男, 1945年生, 博士生导师, 主要研究方向为现代通信中的智能信号与信息信号处理、量子信号与信息处理技术。

杨震: 男, 1961年生, 教授, 博士, 博士生导师, 研究方向为无线通信与网络信号处理、语音处理与现代语音通信技术、信息安全技术。