基于线性预测分析的语音信号压缩感知*

王红柱，陈砚圃，高 悦，王 浩
（西安通信学院，西安 710106）

摘 要：根据语音信号的特点，提出了一种基于线性预测分析的合成矩阵作为语音信号的稀疏变换域，并验证了语音信号在该域上的稀疏特性。由语音信号和随机高斯矩阵构造相应的观测，采用正交匹配追踪算法重构原始语音信号。实验表明，语音信号在新的变换域上的重构性能要优于 DCT 域，且具有较高的分段信噪比和平均意见得分。

关键词：压缩感知；语音信号；稀疏性；线性预测分析；正交匹配追踪
中图分类号：TN912.3 **文献标志码**：A **文章编号**：1001-3695(2012)04-1335-03
doi:10.3969/j.issn.1001-3695.2012.04.037

Compressed sensing for speech signal based on linear prediction analysis

WANG Hong-zhu, CHEN Yan-pu, GAO Yue, WANG Hao
(*Xi'an Communications Institute*, *Xi'an* 710106, *China*)

Abstract: This paper presented a new speech signal sparse domain—synthesis matrix based on linear prediction technology based on the features of speech signal, and verified the sparsity of speech signal in the new sparse domain. By speech signal and the Gaussian random matrix, used OMP to reconstruct the original speech signal. Experimental results demonstrate that the performance of the speech recovered using compressed sensing for speech signal based on linear prediction analysis is better and reconstructed signal has good segment signal to noise ratio and mean opinion score, compared with DCT domain.

Key words: compressed sensing (CS); speech signal; sparsity; linear prediction analysis; orthogonal matching pursuit (OMP)

近年来，由 Donoho[1] 以及 Candes 等人[2] 提出了一种新颖的信息获取理论，即压缩感知(CS)理论。CS 理论与传统奈奎斯特采样定理不同，它是一种边采样边压缩理论，将传统信号压缩中的采样和压缩两个过程合二为一，在某种意义上突破了奈奎斯特采样定理的限制，是对传统信号采样方法的一次新的革命。该理论指出：只要信号是可压缩的或在某个变换域是稀疏的，那么就可以以远低于奈奎斯特采样速率的方式对其进行采样，通过优化逼近的方法近似无损地恢复出原始信号[3]。压缩感知理论一经提出，就在信息论与编码、信号恢复、有损压缩、光学和雷达成像、无线通信等领域受到高度关注，迅速成为信号处理研究的前沿和热点，并被美国科技评论评为 2007 年度十大科技进展之一。

目前基于 CS 理论的语音信号编码研究还不多，尚属于起步阶段。Sreenivas 等人[4] 基于激励声道模型研究了语音信号的稀疏特性，将 CS 理论成功应用于稀疏激励信号的处理中。郭海燕等人[5] 在 KLT 域验证了语音信号的稀疏特性，提出了基于近似 KLT 域的语音信号压缩感知。孙林慧等人[6] 在 DCT 域分析了语音信号的稀疏特性，给出了基于 CS 理论的分布式语音压缩重构的框架。他们从不同角度研究了语音信号的稀疏特性，为语音信号压缩感知的研究提供了全新的思路和方法。

基于压缩感知理论，针对语音信号的特点，本文提出了一种新的语音信号的稀疏变换矩阵，即基于线性预测分析的合成矩阵。

1 压缩感知基本原理

信号本身或在某个变换域上具有稀疏性，是压缩感知理论应用的基础和前提，只有确保信号的稀疏度，才能保证信号的恢复精度。根据调和分析理论，长度为 N 的一维离散时间信号 $x=[x_1,x_2,\cdots,x_N]$ 可表示为一组标准正交基的线性组合：

$$x=\Psi r=\sum_{i=1}^{K} r_i \varphi_i \tag{1}$$

其中：$\Psi=[\varphi_1,\varphi_2,\cdots,\varphi_N]$，$\varphi_i$ 为 $N\times 1$ 的列向量，向量 r 是信号 x 的加权系数序列。若向量 r 中非零元素的个数为 K，即 $\|r\|_0=K$，则称 r 为 x 的 K 稀疏表示[7]。

根据 CS 理论，信号的压缩是通过观测矩阵 Φ（$M\times N$ 的矩阵，$M<N$）投影得到的。

$$y=\Phi x=\Phi\Psi r \tag{2}$$

只要观测矩阵 Φ 和变换基 Ψ 极不相似，就可以以高概率从 $M(M\geqslant K\times\log(N/K))$ 个测量值中准确地恢复 K 个系数[8]。研究表明，随机高斯矩阵与大多数正交矩阵不相关[9]。

CS 的重构实际上就是求解欠定方程组 $y=\Phi\Psi r$。在信号是稀疏或者可压缩的前提下，求解该欠定方程问题可转换为最小 0-范数问题：

$$\hat{r}=\arg\min_{r\in R^N}\|r\|_0 \quad \text{s.t. } y=\Phi\Psi r \tag{3}$$

收稿日期：2011-07-28；**修回日期**：2011-09-15 **基金项目**：国家自然科学基金资助项目(61072125)

作者简介：王红柱(1986-)，男，安徽灵璧人，硕士研究生，主要研究方向为智能信息处理(whzhu1986@163.com)；陈砚圃(1963-)，男，河北元氏人，副主任，教授，博士，主要研究方向为信号与信息处理；高悦(1973-)，女，陕西西安人，讲师，博士，主要研究方向为信号与信息处理；王浩(1987-)，男，陕西清涧人，硕士研究生，主要研究方向为智能信息处理.

对于 0-范数的优化问题,实际上是 NP 难问题[8],就是在多项式时间内难以求解,甚至无法验证解的可靠性。Donoho 的研究表明,当$\boldsymbol{\Theta}$ 满足一定条件时,L_1 优化算法与 L_0 优化算法同解[9],所以通常转换为 L_1 优化问题来求解。

$$\hat{r} = \arg\min_{r\in R^N} \| r \|_1 \quad \text{s.t. } y = \boldsymbol{\Phi\Psi} r \tag{4}$$

通过 Lagrange 乘法,凸优化等式(4)可以表示成

$$\hat{r} = \arg\min_{r\in R^N} \| r \|_1 + \gamma \| y - \boldsymbol{\Phi\Psi} r \|_2^2 \tag{5}$$

式(5)可以看做是式(3)的凸化,常用的求解算法有基追踪(basis pursuit, BP)[9]、匹配追踪(match pursuit, MP)[10]和正交匹配追踪(OMP)算法[11]等。

2 语音信号压缩感知

2.1 语音信号的稀疏表示

线性预测(linear prediction,LP)分析是现代语音信号处理中最核心的技术之一,在语音编码、合成、识别等语音处理领域得到了广泛应用。在 AR(autogressive)模型中,由于语音样点之间存在相关性,所以一个语音信号的抽样值 $x(n)$ 可以用过去的 p 个取样值的线性组合来逼近:

$$x(n) = \sum_{i=1}^{p} a_i x(n-i) + e(n) \tag{6}$$

其中:a_i 是预测系数,$e(n)$是自回归滤波器的激励。通过最小均方误差(least mean square,LMS)准则,使实际语音样值与线性预测样值最小,可以确定唯一的一组线性预测系数$\{a_i\}$。由式(6)可得到 LP 分析的残差 $e(n)$为

$$e(n) = x(n) - \sum_{i=1}^{p} a_i x(n-i) \tag{7}$$

将残差写成向量的形式为:$\boldsymbol{e} = \boldsymbol{T}\boldsymbol{x}$。其中:

$$\boldsymbol{e}_{N\times 1} = \begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{bmatrix}, \quad \boldsymbol{x}_{N\times 1} = \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(N) \end{bmatrix}$$

$$\boldsymbol{T}_{N\times N} = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ -a_1 & 1 & 0 & \ddots & & & \vdots \\ -a_2 & -a_1 & 1 & 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ -a_p & \cdots & -a_2 & -a_1 & 1 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & -a_p & \cdots & -a_2 & -a_1 & 1 \end{pmatrix}$$

由 LP 分析可知,浊音的残差信号 $\boldsymbol{e}$ 是以基音周期为周期重复的单位脉冲序列,即残差信号仅在基音周期的整数倍位置上有较大的幅值。这说明矩阵 $\boldsymbol{T}$ 去除了语音信号 $\boldsymbol{x}$ 相邻样点之间的相关性,因此将矩阵 $\boldsymbol{T}$ 称之为分析矩阵。那么语音信号可以近似表示为

$$x = \boldsymbol{T}^{-1}\boldsymbol{e} = \boldsymbol{H}\boldsymbol{e} \tag{8}$$

其中:$\boldsymbol{H}$ 为分析矩阵 $\boldsymbol{T}$ 的逆矩阵,将残差信号 $\boldsymbol{e}$ 映射到语音信号域实现语音信号的重构,因此将 $\boldsymbol{H}$ 矩阵称为合成矩阵。

本文采用合成矩阵 $\boldsymbol{H}$ 作为语音信号的变换矩阵,发现浊音信号具有较好的稀疏性,而清音信号则不具有稀疏性。这是由于清音帧的残差信号是一个高斯白噪声序列,如图 1 所示。图 1 中的每帧语音为 25 ms,预测阶数为 10,采样率为 8 kHz。由于语音的大部分能量都集中在浊音部分,而清音能量通常较小,故清音帧的非稀疏性对语音信号在矩阵 $\boldsymbol{H}$ 上的 CS 性能影响不大。

2.2 语音信号的压缩与重构

Candes 证明了只要信号在某一个正交空间具有稀疏性,就能以较低的频率($M << N$)采样信号,同时以较高的概率近似无损地重构该信号[12]。通过分析可知语音信号的稀疏变换矩阵 $\boldsymbol{H}$ 并非正交矩阵,但是随机高斯矩阵作为观测矩阵与大多数矩阵不相关[9],而且残差信号 $e(n)$是语音信号 $x(n)$在 $\boldsymbol{H}$ 域上的唯一等价表示。这说明可以将 $\boldsymbol{H}$ 矩阵应用于语音信号的压缩感知中。

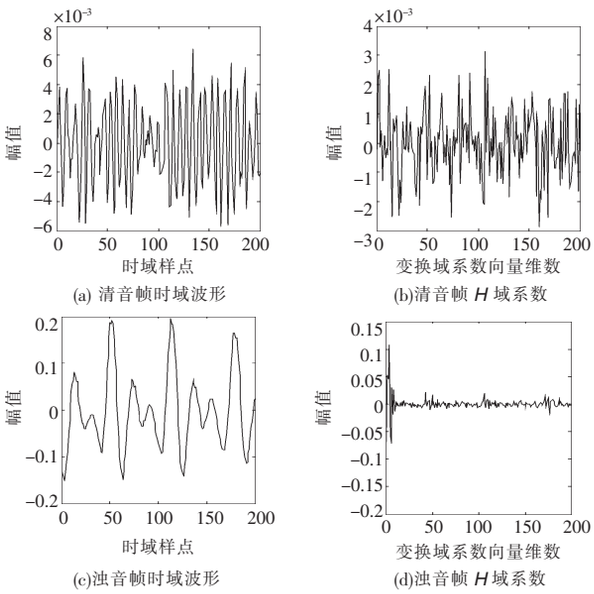

图 1 语音信号的时域波形及在 $\boldsymbol{H}$ 域上的系数

基于 LP 分析的语音信号压缩感知模型如图 2 所示。针对经汉明窗和矩形窗分帧的语音信号,分别利用 LP 分析技术和 CS 技术,提取线性预测系数向量 $\boldsymbol{a}$ 和观测向量 $\boldsymbol{y}$。语音信号的重构以观测向量 $\boldsymbol{y}$ 和线性预测向量 $\boldsymbol{a}$ 为已知条件,利用解优化的方法实现语音在 $\boldsymbol{H}$ 域上的重构。

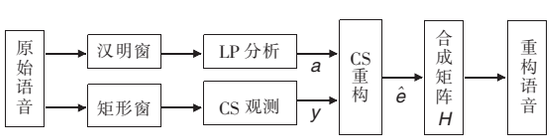

图 2 算法原理

语音信号的压缩分为两个阶段:a)通过观测矩阵$\boldsymbol{\Phi}$ 的 M 个行向量$\{\varphi_i\}$对语音信号进行投影获取观测向量 $\boldsymbol{y}$;b)通过线性预测分析获取每帧语音信号的 p 个线性预测系数,用于构造合成矩阵 $\boldsymbol{H}$。

语音信号的重构就是解优化的过程,即通过求解 L_1 优化问题获得语音信号在稀疏域上的稀疏系数,利用合成矩阵将稀疏系数投影到语音信号域,从而实现信号的重构,如式(9)。

$$\hat{e} = \arg\min_{r\in R^N} \| e \|_1 + \gamma \| y - \boldsymbol{\Phi H} e \|_2^2$$
$$\hat{x} = \boldsymbol{H}\hat{e} \tag{9}$$

本文将使用 OMP 算法来重构语音信号 $\boldsymbol{x}$ 在合成矩阵 $\boldsymbol{H}$ 上的稀疏残差信号$\hat{e}$。该算法根据观测向值 y_i 与恢复矩阵 $\boldsymbol{\Theta}$($\boldsymbol{\Theta} = \boldsymbol{\Phi H}$)列向量的相似程度来确定向量 $\boldsymbol{e}$ 中非零元素的位置,再通过解最小二乘问题得到向量中的非零值。

3 实验结果分析

本文定义信号的压缩比 d、分段信噪比 SNRseg 作为衡量

重构语音质量的客观评价方法:

$$d = M/N \tag{10}$$

$$\mathrm{SNRseg} = \frac{1}{M}\sum_{i=1}^{M}\lg\left(\|x_i\|_2^2/\|x_i-\hat{x}_i\|_2^2\right) \tag{11}$$

其中:M 为原始语音信号的总帧数。实验采用 ITU P.862 标准对重构语音信号的质量作出主观评价。

实验对象为不同说话人的语音信息,采样率为 8 kHz,取线性预测阶数 $p=10$,$N=200$,$M=4$ K,观测矩阵Φ 为 $M\times N$ 的随机高斯矩阵。实验主要研究语音信号在 H 上的 CS 性能以及不同变换矩阵上重构语音信号每帧的信噪比和分段信噪比,并给出重构语音的主观评价。

实验 1 分析语音信号在 H 域上的 CS 性能。在压缩比为 0.5 条件下,实验对不同说话人的语音进行了分析,在此仅选取具有代表性的一帧浊音信号。

如图 3 所示,除两端重构语音与原始语音有一定误差外,基本上可以实现语音信号的精确重构,实验算出其重构信噪比为 23.15 dB。这是由于采用自相关法获取线性预测系数时引入的误差,为了尽可能减小两端的误差,在线性预测阶段使用两端具有平滑过渡特性的汉明窗,在 CS 观测阶段使用能够保证语音信号时域特性的矩形窗。

图 4 给出了语音信号的部分帧在不同稀疏变换域上重构的信噪比。可以看出,H 矩阵上语音信号的重构信噪比要高于 DCT 域,并且实验算出在 H 矩阵和 DCT 域上重构的分段信噪比分别为 15.72 dB 和 8.95 dB。

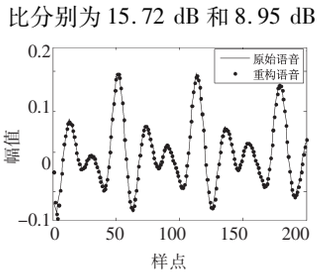

图 3　浊音信号的重构

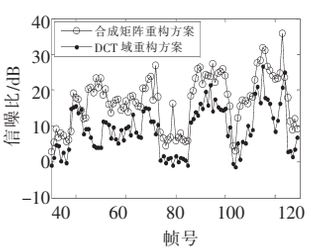

图 4　各方案下的各帧重构信噪比

实验 2 分析合成矩阵 H 作为变换矩阵在不同压缩下重构语音信号的分段信噪比。DCT 域上的语音压缩方案采用矩形窗连续分帧。

图 5 描述了不同压缩比下 H 域重构语音信号的分段信噪比,并将其与 DCT 域的重构性能相比较。在不同压缩比下,本方案下重构信号的分段信噪比明显高于 DCT 域。因而在通信双方具有相同观测矩阵的情况下,发送方仅需多发 10 个线性预测系数,其获得的重构语音的分段信噪比要高于 DCT 域重构的分段信噪比。

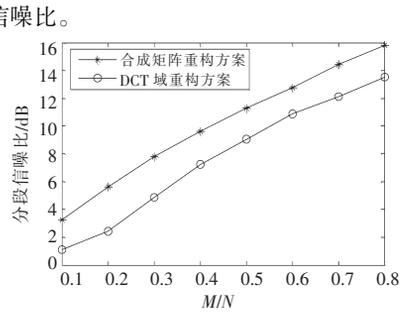

图 5　各方案下的重构信号分段信噪比

实验 3 采用 ITU P.862 标准分别对实验 2 中由 H 域和 DCT 域作为变换域重构出的语音信号作主观评价,得到不同压缩比下的平均意见分(mean opinion score, MOS),如表 1 所示。

表 1　各方案下语音信号重构 MOS 对比

M/N	H 矩阵	DCT 域
0.1	2.56	0.83
0.2	2.81	1.21
0.3	2.98	1.68
0.4	3.15	2.12
0.5	3.21	2.39
0.6	3.33	2.72
0.7	3.43	2.89
0.8	3.61	3.02

从表 1 中可以看出,在不同压缩比下,由合成矩阵 H 作为稀疏变换域重构出的语音信号的 MOS 明显高于 DCT 域。非正式主观听力测试也表明合成矩阵 H 重构的语音在压缩比小于 0.3 时仍有很好的可懂度和自然度,而 DCT 域重构的语音质量较差,能感觉明显的听觉质量下降。

4　结束语

本文提出了一种基于线性预测分析的语音信号压缩感知方法。首先结合线性预测模型给出了基于线性预测分析的合成矩阵作为语音信号的稀疏变换域;然后对语音信号在新的变换域上的稀疏特性进行了实验验证;最后分析了语音信号在该域上的 CS 性能。实验表明,语音信号在合成矩阵上具有近似的稀疏特性和较好的重构性能。实验也证明了只要信号的稀疏矩阵与观测矩阵不相关且可逆,就可以将该矩阵作为变换矩阵应用于信号的压缩感知中。

参考文献:

[1] DONOHO D. Compressed sensing[J]. IEEE Trans on Information Theory, 2006, 52(4): 1289-1306.

[2] CANDES E J, ROMBERG J, TAO T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information[J]. IEEE Trans on Information Theory, 2006, 52(2): 489-509.

[3] CANDES E, TAO T. Near optimal signal recovery from random projections and universal encoding strategies [J]. IEEE Trans on Information Theory, 2006, 52(12): 5406-5425.

[4] SREENIVAS T V, BASTIAAN K W. Compressive sensing for sparsely excited speech signals[C]//Proc of IEEE ICASSP. Washington DC: IEEE Computer Society, 2009: 4125-4128.

[5] 郭海燕,杨震. 基于近似 KLT 域的语音信号压缩感知[J]. 电子与信息学报, 2009, 31(12): 2948-2952.

[6] 孙林慧,杨震. 基于压缩感知的分布式语音压缩与重构[J]. 信号处理, 2010, 26(6): 824-829.

[7] 石光明,刘丹华,高大化,等. 压缩感知理论及研究进展[J]. 电子学报, 2009, 37(5): 1070-1081.

[8] BARANIUK R. A lecture on compressive sensing [J]. IEEE Signal Processing Magazine, 2007, 24(4): 118-121.

[9] CHEN S S, DONOHO D L, SAUNDES M A. Atomic decomposition by basis pursuit [J]. SIAM Review, 2001, 43(1): 129-159.

[10] NEFF R, ZAKHOR A. Very low bit rate video coding based on matching pursuits[J]. IEEE Trans on Circuits and Systems for Video Technology, 1997, 7(1): 158-171.

[11] NEEDELL D, VERSHYNIN R. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit [J]. Found Comput Math, 2009, 9(3): 317-334.

[12] CANDES E. Compressive sampling[C]//Proc of the International Congress of Mathematicians. Madrid, Spain: European Mathematical Society Publishing House, 2006: 1433-1452.