

# 量化压缩感知在语音压缩编码中的应用

朱俊华,谷 鹏,潘海琦,丁 飞

(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,江苏 南京 210003)

**摘 要:**利用语音信号在离散余弦变换(DCT)域的近似稀疏性和量化压缩感知理论,文中提出一种基于量化压缩感知的语音压缩编码方案。编码端利用压缩感知技术,将语音信号投影成数据量大大减少的观测序列,然后对观测序列采用Lloyd-Max量化得到量化后的观测样值;解码端直接利用量化后的观测样值,结合重构算法重构出原始语音信号的DCT系数,经过DCT反变换得到重构后的语音信号,并采用后置低通滤波器改善重构语音的听觉效果。该编码方案解码端不需要进行反量化,而是直接利用量化后的观测样值进行重构,有效降低了解码端的运算量及复杂度。仿真结果表明:采用量化迭代硬阈值(QIHT)算法重构效果优于迭代硬阈值算法(IHT),重构语音的信噪比能达到20 dB以上,MOS分达到3.26。

**关键词:**离散余弦变换;量化压缩感知;Lloyd-Max量化;量化迭代硬阈值算法

**中图分类号:**TP31

**文献标识码:**A

**文章编号:**1673-629X(2014)11-0155-04

**doi:**10.3969/j.issn.1673-629X.2014.11.039

## Application of Quantized Compressed Sensing in Speech Compression Encoding

ZHU Jun-hua, GU Peng, PAN Hai-qi, DING Fei

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

**Abstract:** Utilizing the sparsity of Discrete Cosine Transform (DCT) coefficients of speech signal and the theory of quantized compressed sensing, a novel speech coding scheme based on Quantized Compressed Sensing (QCS) is proposed in this paper. Based on CS theory, the speech signal is transformed into measurement sequence at encoder side, by which the size of the data set is significantly reduced. After quantizing the measurement sequence by Lloyd-Max scheme, the sample value of measurement is gained. At decoder side, together with reconstruction algorithm, the DCT coefficients can be reconstructed by the measurements. The speech signal can be reconstructed after DCT inverse transform. The quality of reconstructed speech signal can be improved by post low-pass filter. The speech signal can be directly reconstructed by the quantized measurements without inverse quantization, leading to the reduction of computation and complexity in decoder site. The simulation results show that the performance of Quantized Iterative Hard Thresholding (QIHT) algorithm is superior to that of Iterative Hard Thresholding (IHT) algorithm. The Signal-to-Noise Ratio (SNR) of the reconstructed speech signal is about 20 dB while mean opinion score (MOS) is up to 3.26.

**Key words:** discrete cosine transform; quantized compressed sensing; Lloyd-Max quantization; quantized iterative hard thresholding

## 0 引言

传统的语音信号采集系统包含两部分,首先以奈奎斯特采样速率采集语音信号,然后对采样得到的信号进行压缩,以便于数据的传输、存储、计算等。考虑到语音信号的稀疏性,总是以大于等于两倍信号带宽的采样率对语音信号进行采样造成了系统资源的浪费。近年来针对稀疏信号(或可压缩信号),Donoho

等人提出了一种新的采样理论—压缩感知(Compressed Sensing, CS)理论<sup>[1-3]</sup>。该理论指出可以将采样和压缩同时进行,从而有效避免采样的浪费。CS理论能够利用远低于奈奎斯特采样率的采样速率采集信号,采样得到的观测序列包含了原始信号的全部信息,可以通过观测序列高概率精确地重构出原始信号<sup>[4-7]</sup>。目前,国内外的学者对于语音信号的压缩感

收稿日期:2013-12-13

修回日期:2014-03-18

网络出版时间:2014-07-28

基金项目:国家自然科学基金资助项目(60971129,61271335);江苏省普通高校研究生科研创新计划(CXZZ13\_0488)

作者简介:朱俊华(1988-),男,硕士,研究方向为基于压缩感知的语音信号处理技术。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140728.1230.056.html>

知已有了一定的研究。丹麦奥尔堡大学的 Christensen M. G. 等研究分析了压缩感知技术应用于语音信号处理领域时需考虑及解决的问题,针对使用冗余字典的实时语音信号,利用压缩感知作为预处理器将语音信号进行稀疏分解,有效降低了运算复杂度<sup>[8]</sup>。Griffin A. 等将压缩感知应用于正弦模拟音频信号,分析了 CS 应用于单声道及多声道音频编码的可能性<sup>[9]</sup>。Abrol V. 等则着重比较了语音信号处理中不同稀疏基及测量矩阵的性能,详细分析了性能边界、压缩比、重构误差等<sup>[10]</sup>。上海大学的万旺根课题组针对实际的自然音频,提出了基于 CS 技术的理论架构,分析了不同重构算法的性能,研究结果表明,OMP 算法能够取得最佳的重构性能,且计算速度较快<sup>[11]</sup>。作者所在的南京邮电大学课题组针对语音信号压缩感知,对观测序列进行了研究,分析了不同投影矩阵对重构语音性能的影响<sup>[12]</sup>。

然而,这些工作主要讨论了基于 CS 理论的理想化语音处理系统架构。对于实际的语音处理系统而言,需考虑观测序列量化对系统性能的影响。文中提出一种基于量化压缩感知的语音编码方案,分析比较了 IHT 及 QIHT 算法对于 1~4 bit 量化压缩感知语音处理系统的影响。研究结果表明:采用 QIHT 重构算法,重构得到的语音信号信噪比达到 20 dB 以上,MOS 分达到 3.26。

## 1 压缩感知理论框架

压缩感知理论由 Donoho 等人在 2006 年提出<sup>[1]</sup>,它是一种新的边采样边压缩的理论框架,主要涉及三个核心问题:

(1) 寻找某个正交变换基或者紧框架  $\Psi$ ,使得信号  $\mathbf{x} \in R^N$  在  $\Psi$  下是稀疏的。这一步保证了  $\mathbf{x}$  满足压缩感知理论的应用前提,为接下来的步骤提供理论依据;

(2) 用一个与变换基  $\Psi$  不相关的  $M \times N$  维的观测矩阵  $\Phi$  与  $\mathbf{x}$  相乘,得到  $\mathbf{x}$  在低维空间的表示  $\mathbf{y}$ ,称  $\mathbf{y}$  为观测向量;

(3) 在解码端通过求一个  $l_0$ -范数最优化问题精确或近似重构出原信号  $\mathbf{x}$ 。

假设原始信号  $\mathbf{x} \in R^N$  在正交基  $\Psi = \{\psi_i \mid \psi_i \in R^N, i = 1, 2, \dots, n\}$  上是  $K$  稀疏的,即

$$\mathbf{x} = \Psi^{-1} \boldsymbol{\theta} \quad (1)$$

式中,稀疏系数向量  $\boldsymbol{\theta} = \Psi \mathbf{x} = (\theta_1, \theta_2, \dots, \theta_n)^T$  只有  $K$  个非零系数,即  $\|\boldsymbol{\theta}\|_0 = K$ 。

CS 理论通过利用观测矩阵将原始信号  $\mathbf{x}$  从高维空间投影到低维空间上,得到观测向量  $\mathbf{y} = (y_1, y_2, \dots, y_m)^T$ ,即

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi^{-1} \boldsymbol{\theta} = \Xi \boldsymbol{\theta} \quad (2)$$

式中,  $\Phi$  为  $M \times N$  ( $M < N$ ) 维的观测矩阵,  $\Phi$  与  $\Psi$  互不连贯; $\Xi$  称为 CS 矩阵。如果  $\mathbf{x}$  本身是稀疏的,  $\Psi$  取单位矩阵。

观测矩阵  $\Phi$  需要满足有限等距特性 (Restricted Isometry Property, RIP)<sup>[11]</sup>,即对于任意长度为  $N$  的  $K$ -稀疏信号  $\mathbf{x}$ , 存在常数  $\delta > 0$ ,使得观测矩阵  $\Phi$  满足

$$1 - \delta \leq \frac{\|\Phi \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq 1 + \delta \quad (3)$$

式中,  $\delta$  为有限等距常数 (Restricted Isometry Constant, RIC)。

在接收端,希望通过观测向量  $\mathbf{y}$  重构出原始信号  $\mathbf{x}$ ,由于系数向量  $\boldsymbol{\theta}$  是  $K$  稀疏的,可以通过求解最小  $l_0$ -范数问题:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \operatorname{argmin} \|\boldsymbol{\theta}\|_0 \\ \text{s. t. } \Xi \boldsymbol{\theta} &= \Phi \Psi^{-1} \boldsymbol{\theta} = \mathbf{y} \end{aligned} \quad (4)$$

式中,  $\|\cdot\|_0$  表示  $l_0$  范数,表示向量  $\boldsymbol{\theta}$  中非零元素的个数。求解公式 (4) 是一个 NP-hard 问题,通常要转化为  $l_1$  范数问题来求解。

## 2 量化压缩感知及 QIHT 算法

与传统的压缩感知不同,文中提出的量化压缩感知把观测值进行了量化。观测值  $\mathbf{y}$  由下式得到

$$\mathbf{y} = Q_b [\Phi \mathbf{x}] \quad (5)$$

式中,  $\Phi \in R^{M \times N}$  是观测矩阵;  $Q_b$  是量化算子,即对每一个观测值用  $b$  比特来量化。如果不作进一步的编码,那么观测向量  $\mathbf{y}$  一共需要  $B = b \times M$  比特。

对于每一种  $\mathbf{z} = \Phi \mathbf{x} \in R^M$  的分布,假定量化算子  $Q_b$  是最优的。在实际中,考虑随机高斯观测矩阵  $\Phi \sim N^{M \times N}(0, 1)$ ,有  $z_i \sim N(0, \|\mathbf{x}\|^2 = 1)$ ,调整量化算子  $Q_b$ ,得到使量化失真最小的  $b$  比特最优高斯量化器,即 Lloyd-Max 量化器<sup>[13]</sup>。该量化器通过一组阈值  $\{\tau_i \in R : 1 \leq i \leq 2^b + 1\} (-\tau_1 = \tau_{2^b+1} = +\infty)$  定义了  $2^b$  个量化区间  $R_i = [\tau_i, \tau_{i+1})$ 。量化区间内的某一值  $\lambda$  通过该量化器输出相应的量化值  $\{q_i \in R_i : 1 \leq i \leq 2^b\}$ ,即

$$Q_b[\lambda] = q_k \Leftrightarrow \lambda \in R_k \quad (6)$$

其中,  $2\tau = q_{i-1} + q_i, q_i = E[g_x | g_x \in R_i], g_x \sim N(0, 1)$ 。

对于任意  $N$  维矢量  $\mathbf{x} \in R^N$ ,由  $\mathbf{y} = \Phi \mathbf{x}$  得到观测向量  $\mathbf{y} \in R^M$ ,其中  $\Phi$  是  $M \times N$  维的观测矩阵。迭代硬阈值算法<sup>[14]</sup> (Iterative Hard Thresholding, IHT) 就是求解满足约束条件  $\|\mathbf{x}\|_0 \leq K$  的  $\min 0.5 \|\mathbf{y} - \Phi \mathbf{x}\|_2^2$  优化问题。即

$$\min_{\mathbf{x} \in R^N} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 \quad (7)$$

s. t.  $\|\mathbf{x}\|_0 \leq K$

IHT 算法十分简洁,采用如下迭代公式

$$x^{(n+1)} = H_K [x^{(n)} + \mu \Phi^* (y - \Phi x^{(n)})] \tag{8}$$

式中,  $x^{(0)} = 0$ ;  $\mu > 0$  且是常数;  $H_K(a)$  是一个非线性算子, 它将矢量  $a$  中幅值最大的前  $K$  个元素外的所有元素置零。

考虑到公式(6)的标量量化模型, 对 IHT 算法中更新余量的求解进行优化, 得到量化迭代硬阈值算法<sup>[12]</sup> (Quantized Iterative Hard Thresholding, QIHT)。QIHT 的迭代公式如下

$$x^{(n+1)} = H_K [x^{(n)} + \mu \Phi^* (y - Q_b(\Phi x^{(n)}))] \tag{9}$$

式中,  $x^{(0)} = 0$ ;  $\mu > 0$  且是常数。

3 基于量化压缩感知的语音压缩编码方案及其性能分析

文中基于语音信号在 DCT 域的稀疏性, 将语音信号进行压缩感知得到观测样值, 然后对观测样值进行量化, 并利用量化后的观测样值和压缩感知重构算法来重构原始语音信号。相较于文献[9]中的编解码方案而言(即对观测序列采用码激励线性预测技术进行编解码, 根据解码后的观测序列和压缩感知重构算法进行重构), 文中提出的方案, 对观测序列进行量化后, 不需要反量化, 而是直接由量化后的观测样值结合压缩感知重构算法重构出原始的语音信号。

图 1 所示为编解码系统框图, 原始语音信号经压缩采样后的观测值通过 Lloyd-Max 量化器, 利用 IHT 或 QIHT 算法对量化后的观测值进行重构。此外, 为优化重构语音的听觉效果, 对重构的语音信号进行低通滤波处理。

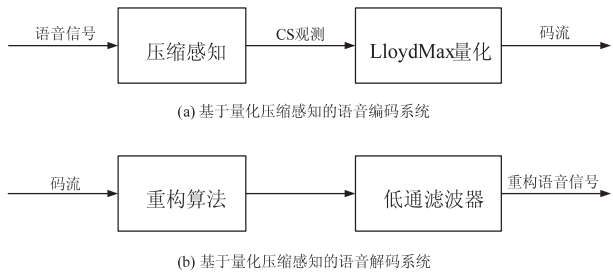


图 1 基于量化压缩感知的语音压缩编解码系统

实验中的数据库采用中国科学院自动化所录制的语音库, 取 50 段不同说话人的语音, 共 1 000 个语句, 以下给出的是平均结果。原始语音的采样率为 8 kHz。解码后的重构语音质量用分段信噪比和 MOS 分来度量, 文中 MOS 分均采用 P. 862 标准算出。

图 2 为 IHT 重构算法对 1~4 bit 量化压缩感知系统性能的影响。

由图可知, 当观测点数增加时, 重构语音信号的信噪比随之增加; 随着量化阶数的增加, 系统性能明显提升。当观测点数为 640 时, 采用 1 bit 量化, 重构语音

信号的信噪比可达 6.68 dB; 而采用 4 bit 量化, 重构语音信号的 SNR 则明显提升, 为 19.37。

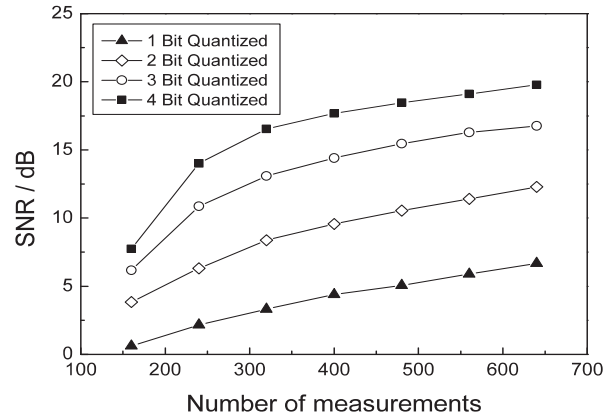


图 2 采用 IHT 算法平均重构信噪比

图 3 为 QIHT 重构算法对 1~4 bit 量化压缩感知系统性能的影响。

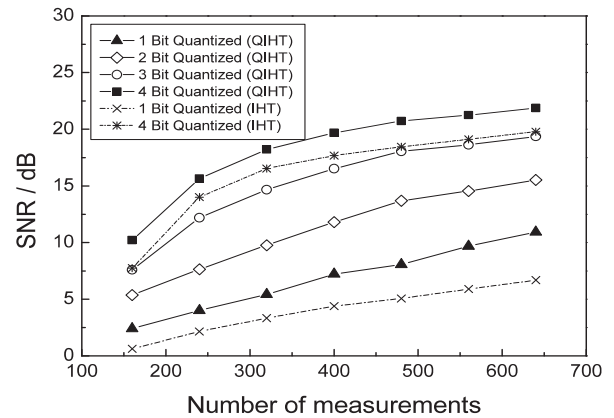


图 3 采用 QIHT 算法平均重构信噪比

由图可知, 当观测点数增加时, 重构语音信号的信噪比随之增加; 另一方面, 随着量化阶数的增加, 系统性能明显提升。当观测点数为 640 时, 采用 1 bit 量化, 重构语音信号的信噪比可达 10.92 dB; 而采用 4 bit 量化, 重构语音信号的 SNR 则明显提升, 为 21.87 dB。从图中也可以看出, QIHT 算法性能要优于 IHT 算法。

文中采用 ITU P. 862 标准的 MOS 分计算法来分析语音信号的重构效果。表 1 为在观测点数分别为 160, 320, 480 及 640 时, 对观测序列进行 1~4 bit 量化所得到的重构语音信号的质量 MOS 分。

表 1 基于 QIHT 算法的量化压缩感知语音系统 MOS 分性能

观测点数	1 bit 量化	2 bit 量化	3 bit 量化	4 bit 量化
160	1.186	1.413	1.736	1.987
320	1.380	1.992	2.357	2.854
480	1.972	2.290	2.637	3.044
640	2.138	2.613	2.978	3.263

由表中数据可以看出, 随着观测点数的增加, 语音信号的重构效果不断提升。在观测点数相同时, 重构

语音信号的 MOS 分性能随着量化阶数的增加而提高。当观测点数为 640 时,对观测序列进行 4 bit 量化,重构语音信号的 MOS 分可达 3.263。

## 4 结束语

考虑到实际语音处理系统的量化问题,文中提出一种基于量化压缩感知理论的语音信号处理系统模型,分析了不同量化阶数及重构算法对该系统性能的影响。通过数值模拟验证了基于量化压缩感知语音信号处理系统的可行性。仿真结果表明,原始语音信号经过该系统后能够被准确地重构。

## 参考文献:

- [1] Donoho D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289–1306.
- [2] Baraniuk R G. Compressive sensing[J]. IEEE Signal Processing Magazine, 2007, 24(4): 118–121.
- [3] Candès E J. Compressive sampling[C]//Proceedings of the international congress on mathematicians. Madrid: [s. n.], 2006: 1433–1452.
- [4] 焦李成, 杨淑媛, 刘芳, 等. 压缩感知回顾与展望[J]. 电子学报, 2011, 39(7): 1651–1662.
- [5] 王伟刚, 胡海峰. 基于压缩感知的协作频谱检测[J]. 计算机技术与发展, 2012, 22(12): 241–244.
- [6] 张爱华, 薄禄裕, 盛飞, 等. 基于小波变换的压缩感知在图像加密中的应用[J]. 计算机技术与发展, 2011, 21(12): 145–147.


- [7] 刘洋, 季薇, 侯晓赞. 一种改进的基于 OMP 重建的宽带频谱感知算法[J]. 计算机技术与发展, 2013, 23(1): 99–102.
- [8] Christensen M G, Stergaard J, Jensen S H. On compressed sensing and its application to speech and audio signals[C]//Proc of 2009 conference record of the forty-third Asilomar conference on signals, systems and computers. Pacific Grove: IEEE, 2009: 356–360.
- [9] Griffin A, Hirvonen T, Tzagkarakis C, et al. Single-channel and multi-channel sinusoidal audio coding using compressed sensing[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(5): 1382–1395.
- [10] Abrol V, Sharma P, Budhiraja S. Evaluating performance of compressed sensing for speech signals[C]//Proc of 3rd international conference on advance computing conference. [s. l.]: [s. n.], 2013: 1159–1164.
- [11] Yu S, Wang R, Wan W, et al. Compressed sensing in audio signals and it's reconstruction algorithm[C]//Proc of international conference on audio, language and image processing. [s. l.]: [s. n.], 2012: 947–952.
- [12] 叶蕾, 杨震, 郭海燕. 基于小波变换和压缩感知的低速率语音编码方案[J]. 仪器仪表学报, 2010, 31(7): 1569–1575.
- [13] Gray R M, Neuhoff D L. Quantization[J]. IEEE Transactions on Information Theory, 1998, 44(6): 2325–2383.
- [14] Blumensath T, Davies M E. Iterative hard thresholding for compressed sensing[J]. Applied and Computational Harmonic Analysis, 2009, 27(3): 265–274.

(上接第 154 页)

- [3] Nikitin A, Stocks N G, Morse R P. Enhanced information transmission with signal-dependent noise in an array of non-linear elements[J]. Physical Review E, 2007, 75: 021121.
- [4] Gammaitoni L. Stochastic resonance in multi-threshold systems[J]. Phys Lett A, 1995, 208: 315–322.
- [5] Stocks N G. Information transmission in parallel threshold arrays: supra-threshold stochastic resonance[J]. Physical Review E, 2001, 63: 041114.
- [6] Wang Youguo, Wu Lenan. Stochastic resonance and noise-enhanced fisher information[J]. Fluctuation Noise Letters, 2005, 5(3): 435–442.
- [7] Wang Youguo, Wu Lenan. Stochastic resonance in nonlinear signal detection[J]. International Journal of Signal Processing, 2006, 2(3): 108–113.
- [8] 张雷, 宋爱国. 随机共振在信号处理中应用研究的回顾与展望[J]. 电子学报, 2009, 37(4): 811–818.

- [9] 王友国, 吴乐南. 离散时间系统中的噪声辅助信号传输[J]. 电子学报, 2009, 37(10): 2273–2276.
- [10] 王友国, 刘沁雨. 多阈值系统中高斯混合噪声改善信息的传输[J]. 计算机技术与发展, 2011, 21(4): 120–122.
- [11] 王友国, 吴乐南. 极大阈值网络中的噪声提高 Fisher 信息[J]. 南京邮电大学学报: 自然科学版, 2007, 27(2): 68–71.
- [12] 王友国, 吴乐南. 随机谐振在信号接收中的应用研究[J]. 信号处理, 2007, 23(5): 768–770.
- [13] 王友国, 刘洪伟, 罗辑. 基于互信息的多阈值系统中随机谐振现象研究[J]. 计算机技术与发展, 2010, 20(6): 89–92.
- [14] Lewand R. Cryptological mathematics[M]. America: the Mathematical Association of America, 2000.
- [15] 曹雪虹, 张宗橙. 信息论与编码[M]. 第 2 版. 北京: 清华大学出版社, 2009.

# 量化压缩感知在语音压缩编码中的应用

作者：[朱俊华](#)，[谷鹏](#)，[潘海琦](#)，[丁飞](#)，[ZHU Jun-hua](#)，[GU Peng](#)，[PAN Hai-qi](#)，[DING Fei](#)  
作者单位：[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](#)，[江苏 南京](#)，[210003](#)  
刊名：[计算机技术与发展](#)  
英文刊名：[Computer Technology and Development](#)  
年，卷(期)：[2014\(11\)](#)

引用本文格式：[朱俊华](#).[谷鹏](#).[潘海琦](#).[丁飞](#).[ZHU Jun-hua](#).[GU Peng](#).[PAN Hai-qi](#).[DING Fei](#) [量化压缩感知在语音压缩编码中的应用](#)[期刊论文]-[计算机技术与发展](#) 2014(11)