基于压缩感知观测序列倒谱距离的语音端点检测算法

叶　蕾[1]　孙林慧[1]　杨　震[2]

（1. 南京邮电大学 通信与信息工程学院，南京，210003；

2. 南京邮电大学 信号处理与传输研究院，南京，210003）

摘　要：本文基于语音信号在离散余弦基上的近似稀疏性，采用稀疏随机观测矩阵和线性规划重构算法对语音信号进行压缩感知与重构。研究了语音信号的压缩感知观测序列特性，根据语音帧和非语音帧压缩感知观测序列频谱幅度分布分散且差异较大的特性，提出基于压缩感知观测序列倒谱距离的语音端点检测算法，并对4dB-20dB下的带噪语音进行端点检测仿真实验。仿真结果显示，基于压缩感知观测序列倒谱距离的语音端点检测算法与奈奎斯特采样下语音的倒谱距离端点检测算法一样具有良好的抗噪性能，但由于采用压缩采样，减少了端点检测算法的运算数据量。

关键词：端点检测；压缩感知；倒谱距离

中图分类号：TN912.3　　**文献标识码**：A　　**文章编号**：1003-0530(2011)01-0067-06

Endpoint Detection Algorithm Based on Cepstral Distance of Compressed Sensing Measurements of Speech Signal

YE Lei[1]　SUN Lin-hui[1]　YANG Zhen[2]

(1. College of Telecommunication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003; 2. Institute of Signal Processing and Transmission, Nanjing University of Posts and Telecommunications, Nanjing 210003)

Abstract: Based on the approximate sparsity of speech signal in discrete cosine basis, Compressed Sensing theory is applied to compress and decompress speech signal in this paper, that is, speech signal is projected to a sparse random measurement matrix and reconstructed by Linear Program. Features of compressed sensing measurements of speech signal is researched in this paper. According to the decentralization characteristic of amplitude spectrum distribution of compressed sensing measurements of speech signal and difference of amplitude spectrum of compressed sensing measurements of voice and non-voice speech, endpoint detection algorithm based on cepstral distance of compressed sensing measurements of speech signal is proposed. Simulation with noisy speech signals ranging from 4dB to 20 dB is made. The simulation results show that endpoint detection algorithm based on cepstral distance of compressed sensing measurements can achieve the same performance of cepstral distance of speech signal sampled by Nyquist Theory in noisy circumstance. The amount of calculation of endpoint detection of speech is reduced by the compressed sensing technology.

Key words: endpoint detection; compressed sensing; cepstral distance

---

收稿日期：2010年7月9日；修回日期：2010年10月26日

基金项目：本文得到国家自然科学基金项目（编号60971129）；南京邮电大学校科研基金青蓝计划项目（编号NY210031、NY208038）及“宽带无线通信与传感网技术”教育部重点实验室资助；江苏省普通高校研究生科研创新计划（CX10B_189Z，CX10B_191Z）

1 引言

语音信号端点检测又称语音活动检测(Voice Activity Detection:VAD),是语音信号预处理阶段的关键技术,其准确性在某种程度上直接影响整个语音处理(如语音识别、语音编码、语音增强)系统的性能。准确的语音信号端点检测能提高识别系统的精度、语音编码效率及语音增强系统的信噪比。但目前各种端点检测算法(短时能量、过零率、自相关参数、时频参数、频带方差、谱熵、倒谱、高阶统计量、分形维、独立分量分析、高阶累量)[1]都是针对奈奎斯特采样的语音而言的,数据量和运算量大,并且多数算法对于噪声敏感,这对多数情况下语音信号实时分析处理的需求是不利的。压缩感知(Compressive Sensing,CS)[2-8]技术是近几年出现的一种新兴的采样技术,该理论指出,只要信号是可压缩的或在某个变换域上是稀疏的,那么就可以用一个与变换基不相关的观测矩阵将高维信号投影到一个低维空间上,然后通过求解一个优化问题就可以从这些少量的投影中以高概率重构原信号。具体而言,当原信号在某个变换域上具有稀疏性时,可以对原信号进行某种变换得到相应的观测序列,此时观测序列的样值个数将远远小于用奈奎斯特采样频率采样得到的信号样值个数,并且,通过这些观测序列,可以精确地重构原信号[2-8]。自然界的多数信号都具有稀疏性,语音信号也不例外,我们能够构建语音信号的稀疏基或近似稀疏基,得到语音信号的稀疏表示,从而将压缩感知原理用于语音信号。基于压缩感知的语音处理系统,突破了传统的奈奎斯特采样的限制,利用样值数大大减少的观测序列来对语音信号进行分析和处理,从而减少了语音信号处理的运算量和复杂度,因此将语音信号处理与压缩感知技术结合具有重要的理论意义和应用价值。目前国内外将 CS 用于语音信号处理领域的研究尚处于起步阶段。Daniele Giacobello 等将 CS 用于稀疏线性预测,得到稀疏的激励参数来进行有效编码[9],J. F. Gemmeke, B. Cranen 利用 CS 原理对噪声环境下的语音进行识别[10],Tao Xu,Wenwu Wang 利用 CS 实现盲源分离[11],显示了压缩感知原理与语音信号处理结合的巨大应用前景。为了区别出语音信号的有效段与无声段,我们将压缩感知原理用于语音信号处理时,需要通过压缩感知观测序列对语音信号进行端点检测。本文研究了语音压缩感知观测序列的特征,并提出对压缩感知观测距离采用倒谱距离算法进行语音信号的端点检测,并将该算法的效果与传统的奈奎斯特采样下的倒谱距离算法的效果进行比较。实验结果显示本文算法端点检测的正确率与奈奎斯特采样下语音的倒谱距离端点检测算法相当,且大大减少了运算数据量。

2 压缩感知基本原理

由于语音信号在离散余弦 DCT(Discrete Cosine Transform)域上是近似稀疏的,根据压缩感知基本原理,如果将 k-稀疏语音信号 $x_0 \in R^n$ 随机投影到与基不相关的观测矩阵 Φ 上,可以产生 m 个观测值。通常观测向量的维数要远远小于原始信号的维数($m<n$),可将观测向量 y 看作原始信号抽样和压缩后的结果。解压缩时,根据观测向量,应用数学优化算法重构原始信号[2-8]。

重构采用 BP(Basis Pursuit)算法[12],通过求解 l_1 最优化问题得到 DCT 系数 θ:

$$\min_{\theta} \| \theta \|_1 \quad \text{subject to } y=\Phi \quad \theta=\Xi\theta \tag{1}$$

其中 $\theta = \quad ^T x_0 = (\theta_1, \theta_2, \ldots, \theta_n)^T$ 为 DCT 系数向量,正交基 $= \{ \quad _i \mid \quad _i \in R^n, i=1,2,\ldots,n \}$ 为 DCT 基,$\Xi = \Phi \quad$ 称为 CS 矩阵。

由于 θ 为 n 维无约束变量,为了保证线性规划标准形式里变量的非负性,可令 $\theta = u - \nu$,其中 u、ν 均为 n 维非负约束变量,即 $u \geq 0, \nu \geq 0$。由此,将(1)问题转化为如下的线性规划问题[13]:

$$\min_{x} c^T x \quad \text{subject to} \quad Ax=b, \quad x \geq 0 \tag{2}$$

其中 $C=(1,\cdots,1)^T, A=(\Xi,-\Xi), b=y, x=\begin{matrix} u \\ \nu \end{matrix}, \theta = u-\nu$。

针对等式线性规划问题(2),可以通过内点法 IP(Interior Point Algorithm)或单纯形(simplex Algorithm)算法求解出最优解 x^*[14,15],并进一步得出 θ^* 或原始信号 x_0 的重构信号 x_0^*。

3 语音压缩感知观测序列特征分析

本文研究的原始语音采样率为 16kHz,采用 20ms 分帧处理,即每帧 320 个原始样点。稀疏基采用离散

余弦基,观测矩阵采用文献[16]中的稀疏随机矩阵以满足保证信号完全重构的有限等距性质(Restricted Isometry Property,RIP)。该随机矩阵的每列元素中只有两个为“1”(位置随机),其它均为“0”,有利用运算量的降低。每帧压缩感知观测序列的样值数分别取 40(压缩比为 1:8)和 80(压缩比为 1:4)进行实验。由于 CS 观测序列的特征与原始语音序列大有不同,目前各种传统的基于原始语音奈氏采样的端点检测算法不一定适合观测序列,所以本文首先研究 CS 观测序列的特征,然后根据其特征确定利用 CS 观测序列进行语音端点检测的算法。由于大部分端点检测算法都是基于时域波形或频域能量的,我们对 30000 帧语音压缩感知序列波形及幅度谱特性进行分析,得到如下结论:

1)无论是非语音帧(如图 1)还是语音帧(图 2、图 3),CS 观测序列都不再具有周期性,而是显示出较强的随机特性(图中压缩比为 1:4,原语音信噪比为 20dB),这使传统端点检测中的某些时域算法(如双门限、自相关、分形维等)难以对语音段和非语音段的 CS 观测序列进行有效区别。

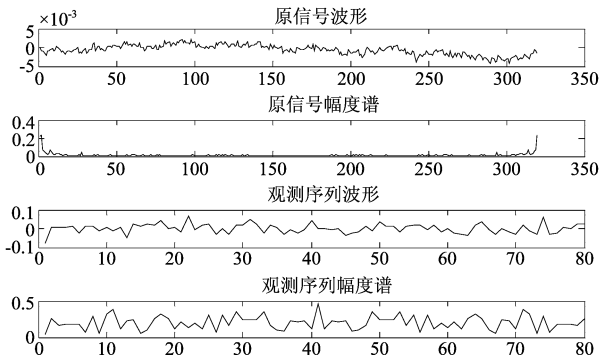


图 1　非语音帧 CS 前后序列波形及幅度谱对比

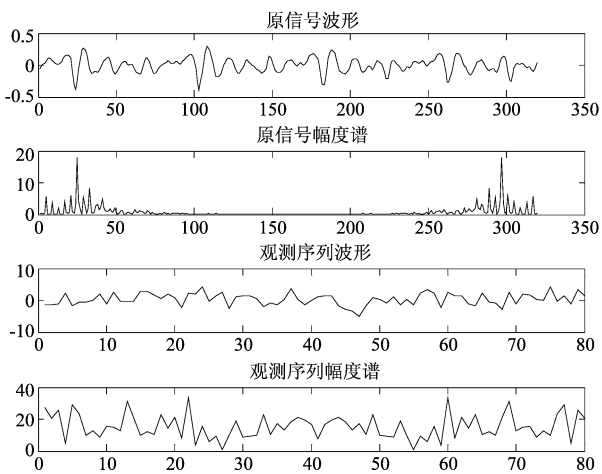


图 2　语音帧 CS 前后序列波形及幅度谱对比(a)

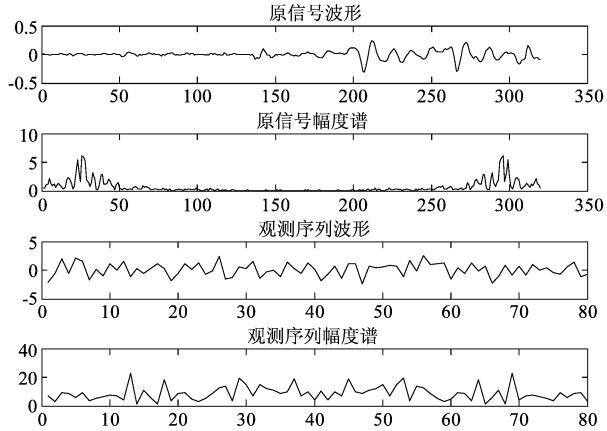


图 3　语音帧 CS 前后序列波形及幅度谱对比(b)

2)非语音帧和语音帧的频谱能量分布都比较分散,不再集中于某个子带,使传统端点检测算法中某些基于频谱的算法(如频带方差、谱熵等)在用 CS 观测序列进行语音端点检测时失效。

3)语音帧与非语音帧的 CS 观测序列幅度谱有很大区别,非语音帧相对语音帧来说,其 CS 观测序列幅度谱很小,而倒谱距离可以反映出两者谱的区别,因此我们提出基于 CS 观测序列倒谱距离的语音信号端点检测算法。

4　基于语音压缩感知观测序列倒谱距离的端点检测算法

对于两个不同的语音 CS 观测序列 $y_0(n)$ 和 $y_1(n)$,倒谱距离表示其倒谱差异的均方值:

$$d_{cep}^2=\frac{1}{2}\int_{-}\left|\lg Y_1(\omega)-\lg Y_0(\omega)\right|d\omega$$

$$=\sum_{n=-}\left[c_1(n)-c_0(n)\right]^2 \qquad (3)$$

3)式中 d_{cep} 为倒谱距离,$c_0(n)$ 和 $c_1(n)$ 分别是 $y_0(n)$ 和 $y_1(n)$ 的倒谱系数,$Y_0(\omega)$ 和 $Y_1(\omega)$ 分别为两个序列的谱密度函数。用 P 阶倒谱系数近似无限阶倒谱系数,式(3) 可以近似为

$$d_{cep}=4.34\sqrt{\left[c_1(0)-c_0(0)\right]^2+2\sum_{n=1}^{p}\left[c_1(n)-c_0(n)\right]^2} \qquad (4)$$

本文基于 CS 观察需要的倒谱进行语音端点检测的具体算法步骤如下:

1)假设 CS 观测的前 5 帧序列对应的原信号是背景噪声,求出这些帧 CS 观测倒谱系数的平均值作为背景噪声 CS 观测倒谱系数的估计值。

2)计算 CS 观测的倒谱系数,利用(4)式计算每帧 CS 观

测倒谱系数与噪声 CS 观测倒谱系数估计值的倒谱距离,并将得到的数据平滑以避免数据突变导致的误判,得到倒谱距离轨迹。

3)确定倒谱距离门限,将每帧 CS 观测倒谱距离与该门限比较,大于门限值的帧判为语音帧,否则判为非语音帧。

步骤 3)中倒谱距离门限根据大量实验来确定。实验结果表明起点处的门限较高(5.0),终点处的门限较低(3.3)。图 4 给出某字的 CS 观测序列倒谱距离轨迹,得到该字的起点和终点分别为第 13 帧和第 39 帧,图 5 给出相应的端点检测结果(信噪比 20dB)。

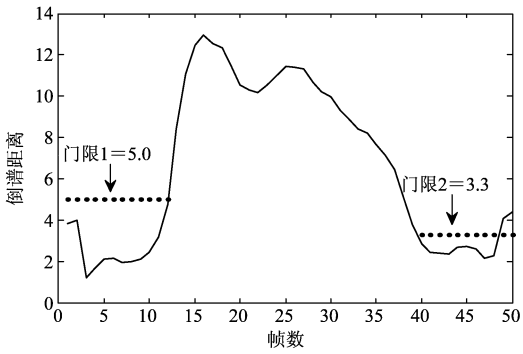


图 4　某字的 CS 观测序列倒谱距离轨迹

本文对不同信噪比下的基于 CS 观测序列倒谱距离的端点检测效果进行了实验,CS 压缩比不同时取得的结果是近似的,实验部分给出压缩比为 1:8 时的结果。

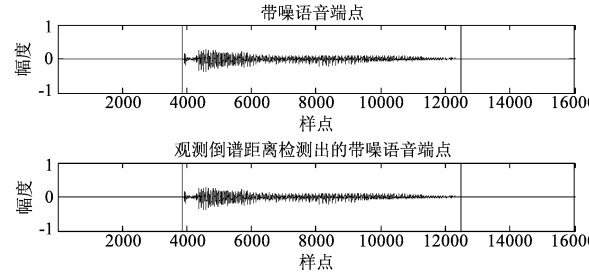


图 5　基于 CS 观测序列倒谱距离的语音端点检测结果

5　仿真实验

实验中的数据库采用中国科学院自动化所录制的语音库,共 600 个语句,原始语音的采样率为 16kHz,20ms 分帧进行压缩感知及端点检测,稀疏基采用离散余弦基,观测矩阵采用高斯随机矩阵,压缩比为 1:8。端点检测的正确率为正确判断语音帧及非语音帧的帧数与总帧数的比值,以手工标注的端点作为判断标准。表 1 至表 3 给出不同信噪比时,高斯白噪声、粉红噪声及汽车噪声(噪声源来自标准噪声库 noisex-92)三种不同噪声下基于 CS 观测序列倒谱距离的端点检测的正确率,并将此结果与奈奎斯特采样下的倒谱距离端点检测算法进行了比较,两者性能相当。

表 1　高斯白噪声下两种端点检测算法正确率

端点检测算法 \ 信噪比	20dB	16dB	12dB	8dB	4dB
基于 CS 观测倒谱距离	0.9693	0.9671	0.9474	0.8750	0.7753
奈奎斯特采样下倒谱距离	0.9737	0.9698	0.9539	0.8553	0.7763

表 2　粉红噪声下两种端点检测算法正确率

端点检测算法 \ 信噪比	20dB	16dB	12dB	8dB	4dB
基于 CS 观测倒谱距离	0.9715	0.9631	0.9524	0.8904	0.7976
奈奎斯特采样下倒谱距离	0.9747	0.9652	0.9684	0.8999	0.8091

表 3　汽车噪声下两种端点检测算法正确率

端点检测算法 \ 信噪比	20dB	16dB	12dB	8dB	4dB
基于 CS 观测倒谱距离	0.9688	0.9609	0.9570	0.8797	0.7925
奈奎斯特采样下倒谱距离	0.9553	0.9532	0.9492	0.8914	0.8003

在现有的语音端点检测算法中,奈奎斯特采样下的倒谱距离检测抗噪性能良好,而本文基于 CS 观测序列倒谱距离的端点检测算法与奈奎斯特采样下语音的倒谱距离端点检测算法性能相当,但由于是对经过 1∶8压缩的 CS 观测序列进行端点检测运算,减少了运算量。表 4 给出两种端点检测算法每帧运算量的对比。基于 CS 观测序列倒谱距离的端点检测算法涉及到求观测序列、求倒谱系数和求倒谱距离的运算,而奈奎斯特采样下语音的倒谱距离端点检测算法涉及到求倒谱系数和求倒谱距离的运算。前者由于采用 320 * 40 的稀疏矩阵做观测矩阵 Φ(即每列只有两个元素为“1”,位置随机,其它元素均为“0”),由(5)式,由原始语音 x_0 求观测序列 y 实际只涉及加法运算,且最大运算量为 2 * 319 即 638 次加法。

$$y=\Phi \quad \theta=\Phi x_0 \tag{5}$$

求倒谱系数公式为

$$c(n)=ifft(\log(|fft(x_0)|)) \tag{6}$$

根据(4)、(5)和(6),将两种算法的每帧的运算量进行对比,如表 4 所示。

表 4　两种端点检测算法运算量对比(每帧)

运算量 \ 算法	基于 CS 观测倒谱距离	奈奎斯特采样下倒谱距离
求观测	最多 638 次加法	无
求倒谱系数	40 点 fft　1 次 求模　40 次 求对数　40 次 40 点 ifft　1 次	320 点 fft　1 次 求模　320 次 求对数　320 次 320 点 ifft　1 次
求倒谱距离	平方　40 次 加法　80 次	平方　320 次 加法　640 次

可见,基于 CS 观测倒谱距离的端点检测运算量大大少于奈奎斯特采样下的倒谱距离端点检测算法。

6　结束语

压缩感知理论给信号采样算法带来一次新的革命,采样数据的大量减少使其应用到许多领域。语音信号具有稀疏性,因此可以将压缩感知技术应用到语音信号处理领域,压缩感知技术与语音信号处理结合,可以使语音编码、语音识别、语音合成、语音增强等领域的技术更加完善。本文研究了基于压缩感知观测序列的语音端点检测算法,为压缩感知与语音信号处理技术结合的研究提供了一定的思路。

参考文献

[1] 马静霞. 带噪语音端点检测方法的研究 [D]. 秦皇岛燕山大学电气工程学院,2007.
MA Jing xia. Research on Voice Activity Detection Method [D]. College of Electronical Engineering of QIN Huangdao Yanshan University,2007.

[2] E Candès. Compressive sampling. Proceedings of the International Congress of Mathematicians [C]. Madrid, Spain,2006,3:1433-1452.

[3] E Candès,J Romberg, Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information [J]. IEEE Trans. on Information Theory,2006,52 (2):489-509.

[4] E Candès and J Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions [J]. Foundations of Comput Math,2006,6 (2):227-254.

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

[6] E J Candès, J Romberg. Practical signal recovery from random projections [OL]. http://www. acm. caltech. edu/~ emmanuel/ papers/ PracticalRecovery. pdf.

[7] D L Donoho, Y Tsaig. Extensions of compressed sensing [J]. Signal Processing. 2006,86 (3):533-548.

[8] 石光明,刘丹华,高大化,刘哲,林杰,王良君. 压缩感知理论及其研究进展[J]. 电子学报,2009 5(37):1070-1081.
SHI Guang ming,LIU Dan hua1,GAO Da hua,LIU Zhe, LIN Jie,WANGLiang jun. Advance s in Theory and Application of Compressed Sensing[J]. Chinese Journal of Electronics, 2009, 5(37):1070-1081.

[9] Giacobello, D., Christensen, M. G., Murthi, M. N., Jensen, S. H., Moonen, M. Retrieving Sparse Patterns Using a Compressed Sensing Framework: Applications to Speech Coding Based on Sparse Linear Prediction[J]. Signal Processing Letters, IEEE. 2010,17(1): 103-106.

[10] J. F. Gemmeke and B. Cranen. Using sparse representations for missing data imputation in noise robust speech recognition [C]. European Signal Processing Conf. (EU-

SIPCO), Lausanne, Switzerland, August 2008.

[11] Tao Xu, Wenwu Wang. A compressed sensing approach for underdetermined blind audio source separation with sparse representation[J]. 2009-Sept: 493-496.

[12] S. Chen, D. L. Donoho, M. A. Saunders, "Atomic decomposition by basis pursuit"[J], SIAM J. Sci. Comp, 1999: 20(1)33-61.

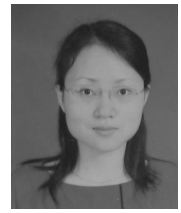
[13] M. Andrecut, R. A. Este, S. A. Kauffman, "Competitive optimization of compressed sensing"[J], Journal of Physics A: Mathematical and Theoretical, 2007(40): 299-305.

[14] 陈宝林. 最优化理论与算法[M]. 第2版,北京:清华大学出版社,2005:26-85.
CHEN Bao ling. Optimal Theories & Algorithms[M]. Edition 2, Beijing: Tsinghua University publishing house, 2005:26-85.

[15] 何坚勇,最优化算法[M]. 第1版,北京:清华大学出版社,2007:1-74.
HE Jian yong. Optimal Algorithms[M]. Edition 1, Beijing: Tsinghua University publishing house, 2007:1-74.

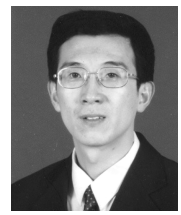
[16] R. Berinde and P. Indyk, Sparse recovery using sparse random matrices. MIT-CSAIL Technical Report, 2008. Available at http://people.csail.mit.edu/indyk/report.pdf.

作者简介



叶　蕾(1978-),女,浙江宁波人。南京邮电大学通信与信息工程学院讲师,信号与信息处理专业博士研究生。目前研究方向为语音处理与现代语音通信。
E-mail:yel@njupt.edu.cn

孙林慧(1979-),女,山西临汾人。南京邮电大学通信与信息工程学院信号与信息处理专业讲师,博士研究生。目前主要研究方向是语音信号处理、信号处理。



杨　震(1961-),男,江苏苏州人。南京邮电大学信号处理与传输研究院教授,博士生导师。目前研究方向为现代网络通信、语音处理与现代语音通信。