基于自适应多尺度压缩感知的语音压缩与重构

孙林慧,杨 震,叶 蕾
(南京邮电大学通信与信息工程学院,江苏南京 210003)

摘 要: 本文针对语音信号的压缩感知问题,在系数总长度不超过原信号长度的前提下,推导了 Sym 小波分解合成的矩阵形式,提出了语音信号多尺度压缩感知(MCS)框架.进一步分析语音信号在小波基下不同级的稀疏性,提出了自适应多尺度压缩感知(AMCS)方法,把该方法运用到语音压缩与重构中,对重构语音进行了主客观评价,并进行了说话人识别验证,得出结论:基于 AMCS 比三层 MCS 重构语音的性能好.

关键词: Sym 小波; 多尺度压缩感知; 自适应多尺度压缩感知; 语音压缩与重构; 基追踪
中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112 (2011) 01-0040-06

Speech Compression and Reconstruction Based on Adaptive Multiscale Compressed Sensing Theory

SUN Lin-hui, YANG Zhen, YE Lei
(*College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu* 210003, *China*)

Abstract: In this paper, a matrix form of Sym wavelet decomposition and synthesis is deduced, keeping the length of the coefficient no more than the length of original speech signals, and then we propose a framework of speech Multiscale Compressed Sensing (MCS) and an Adaptive Multiscale Compressed Sensing (AMCS) method by analyzing sparsity of different wavelet levels of speech signals. We compare AMCS with MCS by applying both methods to speech compression and reconstruction, and the reconstructed speech signal evaluated by the objective and subjective evaluation is applied to speaker recognition. The experimental results show that the reconstruction performance of speech signal based on AMCS is superior to MCS.

Key words: Sym wavelet; multiscale compressed sensing; adaptive multiscale compressed sensing; speech compression and reconstruction; basis pursuit

1 引言

Nyquist 采样定理是大多数信号采样所遵循的规律,是信号精确重构的充分条件,但不一定是必要条件.2004 年由 Donoho 等人提出的压缩感知[1~3](Compressed Sensing,CS)是基于信号在某个域的稀疏性建立的信号采样新理论,表明具有稀疏性的压缩感知能获得更好的压缩性能,信号的稀疏性或可压缩性是实现压缩重构的必要条件之一.边压缩边采样使得 CS 理论具有巨大的吸引力和应用前景,其应用研究已经涉及到众多领域,如:CS 雷达、无线传感网络、图像采集设备的开发、医学图像处理、Analog-to-Information 等[4].

随着移动通讯技术的迅速发展,各种掌上设备手机、商务通、PAD、传感节点等得到广泛应用.但由于其便携式的特点,导致存储能力、计算能力有限,完全靠自身进行大数据量的计算、存储是不现实的,因而就必须基于服务器/客户端模型.采用基于压缩感知的语音压缩和重构算法,有以下两方面的优点:客户端压缩方法简单,只要乘以随机观测矩阵即可,无论语音是浊音还是清音,是英文还是中文,压缩方案相同;针对客户端相同的压缩方案,服务器端可以采用不同的改进的重构方案提高信号重构性能.

目前,Gemmeke 等利用 CS 原理对噪声环境下的语音进行识别,语音识别系统的抗噪性能大大提高[5].Giacobello 等把 CS 的理论框架应用到语音编码[6].Xu 等把 CS 理论引入信息隐藏[7].练秋生等把解析轮廓波变换运用到图像稀疏表示和 CS 中[8].他们把 CS 直接应用到某个领域,但是没有详细分析语音信号在多级小波基下的稀疏性,也没有考虑语音信号的多尺度压缩对压缩重构性能的影响.

收稿日期:2010-01-13;修回日期:2010-08-28
基金项目:国家 863 高技术研究发展计划(No.2006AA010102);国家自然科学基金(No.60971129;No.60902065);南京邮电大学青蓝计划(No.NY208038);江苏省普通高校研究生科研创新计划(No.CX09B_148Z,No.CX10B_191Z,No.CX10B_189Z)

由于小波变换多级分解合成过程,系数的总长度总在变化,不方便于实用.本文首先在小波分解系数总长度基本等于原信号长度的前提下,推导了 Sym 小波分解合成的矩阵表示形式,然后对语音信号进行联合采样即把 CS 理论应用到高频系数,而低频系数采用传统的 Nyquist 采样方法.最后基于多尺度压缩感知(MCS,Multiscale Compressed Sensing)分析语音信号的压缩与重构性能.由于不同语音信号在 Sym 正交小波基下的稀疏性不同,提出了自适应多尺度压缩感知(Adapted MCS,AMCS)方法,自适应感知每级高频系数的稀疏度从而得到观测矩阵 $\boldsymbol{\Phi}$,并且自适应决定分解的层数.把 MCS 和 AMCS 方法分别运用到语音压缩与重构中,对利用基追踪重构的语音进行了主客观评价,并进行了说话人识别验证.

2　基于小波变换的多尺度压缩感知理论

2.1　CS 基本理论

信号 $\boldsymbol{x}\in\boldsymbol{R}^N$ 是 $N*1$ 维列向量,可以用正交基矩阵 $\boldsymbol{\Psi}\in\boldsymbol{R}^{N\times N}$来表示[2]:

$$\boldsymbol{x}=\boldsymbol{\Psi\alpha} \tag{1}$$

其中,投影系数矢量 $\boldsymbol{\alpha}=\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{x}$.显然,$\boldsymbol{x}$ 和 $\boldsymbol{\alpha}$ 是同一信号的等价表示,$\boldsymbol{x}$ 是信号的时域表示,$\boldsymbol{\alpha}$ 是信号的 $\boldsymbol{\Psi}$ 域表示.当信号 $\boldsymbol{x}$ 在正交基 $\boldsymbol{\Psi}$ 上仅有 $K\ll N$ 个非零系数(或远大于零的系数)时,称 $\boldsymbol{\Psi}$ 为信号 $\boldsymbol{x}$ 的稀疏基,信号 $\boldsymbol{x}$ 在 $\boldsymbol{\Psi}$ 上是 K-稀疏的,此时式(1)就是 $\boldsymbol{x}$ 的稀疏表示.

如果 $\boldsymbol{x}$ 是 K-稀疏的,根据压缩感知理论[2]可知,我们可以用一个与正交基 $\boldsymbol{\Psi}$ 不相关的观测矩阵 $\boldsymbol{\Phi}\in\boldsymbol{R}^{M\times N}(M\ll N)$对信号 $\boldsymbol{x}$ 进行线性变换,得到观测向量 $\boldsymbol{y}\in\boldsymbol{R}^M$.

$$\boldsymbol{y}=\boldsymbol{\Phi x}=\boldsymbol{\Phi\Psi\alpha}=\boldsymbol{\Xi\alpha} \tag{2}$$

其中 $\boldsymbol{\Xi}=\boldsymbol{\Phi\Psi}$.

由于观测 $\boldsymbol{y}$ 的维数 M 远远小于信号 $\boldsymbol{x}$ 的维数 N,从而达到压缩的目的.解式(2)的逆问题是一个病态问题,所以无法直接从观测 $\boldsymbol{y}$ 中解出信号 $\boldsymbol{x}$,然而当式(2)中 $\boldsymbol{\alpha}$ 是 K-稀疏的,且 $K<M\ll N$ 时,利用信号稀疏分解理论中已有的稀疏分解算法,可以通过求解式(2)的逆问题得到稀疏系数 $\boldsymbol{\alpha}$,再代入式(1)进一步得到信号 $\boldsymbol{x}$[1~3].重构的最直接方法是通过 l_0 范数下求解式(3)的最优化问题:

$$\min\|\boldsymbol{\alpha}\|_0 \quad \text{s.t.} \quad \boldsymbol{\Xi\alpha}=\boldsymbol{y} \tag{3}$$

从而得到稀疏系数 $\boldsymbol{\alpha}$ 的估计.常用的求解方法有基追踪(BP:Basis Pursuit)、匹配追踪法(MP:Matching Pursuit)和正交匹配追踪法(OMP:Orthogonal Matching Pursuit)等.

2.2　基于小波变换的 MCS 理论

参考文献[3]提出基于 Harr 小波对信号进行联合采样:直接对信号 $\boldsymbol{x}$ 的低频系数 $\boldsymbol{\beta}_{j0}$进行传统的 Nyquist 采样,对高频系数 $\boldsymbol{\alpha}_j$ 进行 CS 采样:

$$\boldsymbol{y}_j=\boldsymbol{\Phi}\,\boldsymbol{x}_j=\boldsymbol{\Phi\Psi}_j\boldsymbol{\alpha}_j=\boldsymbol{\Xi}\,\boldsymbol{\alpha}_j \tag{4}$$

其中 $\boldsymbol{x}_j$ 表示高频系数$\boldsymbol{\alpha}_j$ 对应的信号.这样,如果得到低频系数 $\boldsymbol{\beta}_{j0}$和观测 $\boldsymbol{y}_j$,通过式(3)重构 $\boldsymbol{\alpha}_j$ 后,进一步重构原信号.

由于 Harr 小波 $\psi(x)$已知,式(4)中的 $\boldsymbol{\Psi}_j$ 容易生成,那么对于 $\psi(x)$没有明确的表达形式的其他正交小波,例如 Sym 正交小波,怎样才能得到 $\boldsymbol{\Psi}_j$ 呢? 在下节中我们推导出一种解决方案.

3　语音信号的多尺度压缩感知

3.1　基于 MCS 的语音信号的压缩与重构框架

基于 MCS 理论建立的语音信号压缩重构框图如图 1 所示,图 1 中画出了基于 MCS 理论的三级小波压缩重构过程.首先在客户端对语音信号分帧,然后进行多级小波分解,每级高频系数采用 CS 理论得到对应的观测序列,所有观测序列和低频系数(联合采样)一起编码传输,最后在服务器端多级重构高频系数,联合低频系数一起合成语音,接下来可以对重构语音进行说话人识别等应用.

基于 MCS 理论分析语音信号处理主要涉及以下三方面的内容:

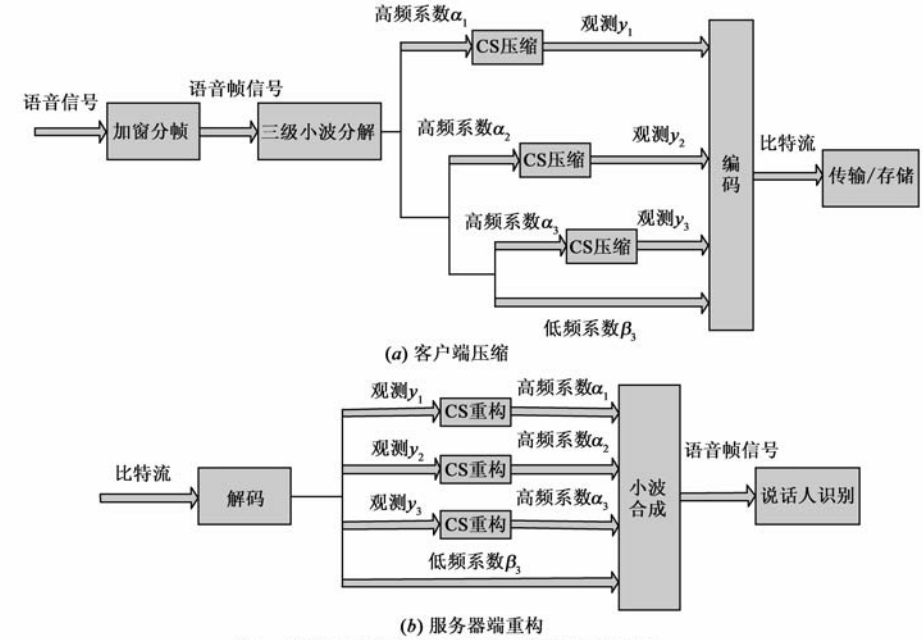

图1　基于MCS的语音信号的压缩与重构框图

(1)客户端对于语音信号的每帧数据 $\boldsymbol{x}\in\boldsymbol{R}^N$,采用多级小波分解,那么分几级可以基本重构,重构时每级的小波基如何构造,这是本文重点分析的部分.

(2)客户端如何设计一个平稳的、与变换基 $\boldsymbol{\Psi}$ 不相关的 $M*N$ 维的观测矩阵 $\boldsymbol{\Phi}$,保证稀疏向量从 N 维降到 M 维时重要信息不遭破坏,即信号低速采样问题.本文选择高斯随机矩阵作为观测矩阵.

(3)服务器端采用哪种快速重构算法可以从线性观测中恢复原信号,即语音信号重构问题.本文采用 BP 算法.

3.2 基于 Sym 正交小波变换的语音信号的 MCS

信号的稀疏表示已经得到广泛研究[9],而 CS 理论的前提条件是信号具有稀疏性或可压缩性.语音信号不满足严格稀疏的要求,但是语音信号具有可压缩性,那么合理地选择稀疏基 $\boldsymbol{\Psi}$,使得信号的稀疏系数个数尽可能少,不仅有利于提高采集信号的速度,而且有利于减少存储、传输信号所占用的资源.常用的稀疏基有:正(余)弦基、小波基、chirplet 基以及 curvelet 基等.对于能量有限信号空间,小波基是一类具有时间局部性或频谱局部性的正交基[10].由 Haar 在 1909 年给出的 Haar 小波基,是 dbL(L 表示阶数)中最简单的一种,它的 $\psi(x)$已知,但是它的光滑性很差,限制了它的适用性.而 SymL 是对 dbL 的改进,不仅具有正交性而且具有紧支撑性[11],在实际应用中大多情况采用,有更好的适用性,但是它的 $\psi(x)$没有明确的表达形式,怎样才能得到 $\boldsymbol{\Psi}_j$ 呢? 接下来我们推导出一种解决方案.

从 SymL 小波分解合成过程推导出小波基矩阵.采用 SymL 时相应的低频系数和高频系数都是有限长度的,为简单起见,设低频系数与高频系数都是实数.可以通过系数之间的关系来构造基矩阵.对有限长度的输入数据序列:$x_n;n=1,2,\cdots,N$,令 $\beta_{0,n}=x_n$,SymL 小波变换高频系数 $\alpha_{j,k}$、低频系数 $\beta_{j,k}$之间存在关系:

$$\beta_{j,k}=\sum_{n=1}^{N}l_n\beta_{j-1,2k+n}=\sum_{n=1}^{N}\beta_{j-1,n}l_{n-2k}=\beta_{j-1}*l(2k) \tag{5}$$

$$\alpha_{j,k}=\sum_{n=1}^{N}h_n\beta_{j-1,2k+n}=\sum_{n=1}^{N}\beta_{j-1,n}h_{n-2k}=\beta_{j-1}*h(2k) \tag{6}$$

其中 $j=1,2,\cdots,J,k=1,\cdots,N/2^j$.其中 J 为实际中要求分解的步数,最多不超过 $\log_2 N$.这里,$\{l(n)\}$与$\{h(n)\}$分别为分解低通与高通滤波器系数.由式(5)、(6)可见,每级一维 DWT 与一维线性卷积计算很相似.所不同的是:在 DWT 中,输出数据下标增加 1 时,权系数在输入数据的对应点下标增加 2,这称为"间隔取样".注意到系数和输入序列都是有限长度的序列,上述和实际上都只有有限项.若完全按照上述公式计算,在经过 J 步分解后,所得到的 $J+1$ 个序列 $\boldsymbol{\alpha}_j,j=1,\cdots,J$ 和 $\boldsymbol{\beta}_J$ 的非零项的个数之和一般要大于 N,超过了输入序列的长度.如图 2 所示.一帧语音数据 512 点,进行 5 级分解,1~5 级的高频系数个数分别为 263、139、77、46、30,第 5 级低频系数的个数为 30,系数总个数为 585 > 512.在数据压缩等应用中,希望总的长度基本不增加,这样可以提高压缩比、减少存储量并减少实现的难度.

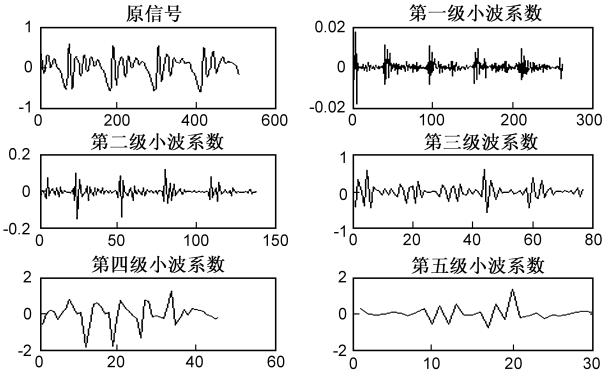

图2 一段语音信号五级小波分解

式(5)、(6)中输出序列长度等于输入序列长度加上滤波器系统响应序列长度减 1,而一般情况系统响应序列长度大于 1,所以输出长度大于输入长度,使得五级分解后总系数大于原信号长度.于是把输入序列和系统响应序列进行周期延拓构建周期序列,进行周期卷积,最后取主值序列作为输出序列,即进行循环卷积,那么输出和输入序列长度则相等.分析第 j 步分解,式(5)、(6)我们采用类似循环卷积的方法近似.当就 $j-1$ 步低频系数长度为 8,滤波器的长度为 4 时,令第 j 级高频系数 $\alpha_j(n)=\{\alpha_{j,1},\alpha_{j,2},\cdots,\alpha_{j,8}\}$,分解高通滤波器系数 $h(n)=\{h_1,h_2,h_3,h_4,0,0,0,0\}$,第 $j-1$ 级低频系数 $\beta_{j-1}(n)=\{\beta_{j-1,1},\beta_{j-1,2},\cdots,\beta_{j-1,8}\}$,则 $\alpha_j(n)\approx h(n)\otimes\beta_{j-1}(n)$,其中$\otimes$表示 N 点的循环卷积,得到结果是 N 点的,然后抽取偶数位置的值将得到第 j 级的高频系数.矩阵表示形式:

$$\boldsymbol{\alpha}_j=\boldsymbol{H}_d\boldsymbol{\beta}_{j-1} \tag{7}$$

其中 $\boldsymbol{H}_d$ 为分解矩阵,同理可得 $\boldsymbol{\beta}_j=\boldsymbol{L}_d\boldsymbol{\beta}_{j-1}$,$\boldsymbol{L}_d$ 具有与 $\boldsymbol{H}_d$ 类似的形式.我们重点需要构造逆变换矩阵 $\boldsymbol{L}_r$ 和 $\boldsymbol{H}_r$.小波逆变换为:

$$\beta_{j-1,n}=\sum_{k\in Z}\beta_{j,k}l_{n-2k}+\sum_{k\in Z}\alpha_{j,k}h_{n-2k}\quad j=J,\cdots,1 \tag{8}$$

首先对 $\beta_{j,k}$、$\alpha_{j,k}$插值补零,然后利用一维循环卷积方式构造合成矩阵.令第 j 级高频系数 $\alpha_j(n)=\{0,\alpha_{j,2},0,\alpha_{j,4},0,\alpha_{j,6},0,\alpha_{j,8}\}$,合成高通滤波器系数 $h(n)=\{0,0,0,0,h_4,h_3,h_2,h_1,\}$,第 j 级低频系数 $\beta_j(n)=\{0,\beta_{j,2},0,\beta_{j,4},0,\beta_{j,6},0,\beta_{j,8}\}$,合成低通滤波器系数 $l(n)=\{0,0,0,0,l_4,l_3,l_2,l_1\}$,第 $j-1$ 级低频系数 $\beta_{j-1}(n)=$

$\{\beta_{j-1,1},\beta_{j-1,2},\cdots,\beta_{j-1,8}\}$,则式(8)可表示为

$$\beta_{j-1}(n)=l(n)*\beta_j(n)+h(n)*\alpha_j(n)\approx l'(n)\otimes\beta_j(n)+h'(n)\otimes\alpha_j(n) \quad (9)$$

其中 $l'(n)$、$h'(n)$是 $l(n)$、$h(n)$循环右移一位的结果(实验结果显示,这样重构误差小).合成过程的矩阵形式为:

$$\boldsymbol{\beta}_{j-1}=\boldsymbol{L}_r\boldsymbol{\beta}_j+\boldsymbol{H}_r\boldsymbol{\alpha}_j \quad (10)$$

其中 $\boldsymbol{\beta}_j$,$\boldsymbol{\alpha}_j$ 是分别是第 j 级分解的低频系数矢量和高频系数矢量,矩阵 $\boldsymbol{L}_r$,$\boldsymbol{H}_r$ 是相应的合成矩阵.$\boldsymbol{L}_r$ 具有类似形式.如果采用 CS 对语音信号压缩重构时,令 $\boldsymbol{\Psi}=\boldsymbol{H}_r$ 可以重构上一级低频系数.客户端对所有高频系数 $\boldsymbol{\alpha}_j$ 分别通过式(4)进行 CS 采样得到一系列 n_j 个观测 $\boldsymbol{y}_j$.服务器端解决式(3)得到 $\boldsymbol{\alpha}_j$,结合 $\boldsymbol{\beta}_j$ 采用式(10)一起重构 $\boldsymbol{\beta}_{j-1}$,重复此过程,直至重构出 $\boldsymbol{\beta}_0$,即信号.进行 j_0 级分解,所有的系数采样后得到总的样值个数为:

$$n_{ms}=N/2^{j0}+n_{j0}+n_{j0-1}+\cdots+n_1 \quad (11)$$

而原来低频系数和高频系数总个数为:

$$m=N/2^{j0}+N/2^{j0}+N/2^{j0-1}+\cdots+N/2^1=N \quad (12)$$

压缩比为 n_{ms}/m,体现了采用 MCS 压缩与传统的 Nyquist 采样相比压缩的程度.

基于 MCS 的语音信号的压缩与重构编程思想.

客户端:(1)语音信号加窗分帧,我们前期研究中已经得出结论,帧长取 512 点性能较好,本文取帧长 512 点;(2)对帧信号多级小波分解;(3)对每级系数对应的信号 $\boldsymbol{\beta}_\alpha=\boldsymbol{H}_r\boldsymbol{\alpha}$ 采用 CS 压缩采样得到观测 $\boldsymbol{y}_\alpha=\boldsymbol{\Phi}\boldsymbol{\beta}_\alpha=\boldsymbol{\Phi}\boldsymbol{H}_r\boldsymbol{\alpha}$;(4)最后一级的高频系数矢量 $\boldsymbol{\beta}$ 联合所有的 $\boldsymbol{y}_\alpha$ 矢量,采用传统的 LBG 矢量量化(Vector Quantization,VQ)方法编码,取 1bit/系数.观测矩阵 $\boldsymbol{\Phi}$ 的随机种子、帧长、每级压缩比在数据头编码.

服务器端:解码后重构,采用文献[12]中的方法由观测重构每级高频系数,我们通过 Matlab 优化工具箱中内点法"linprog"求解得到高频系数的最优解,结合当前级的低频系数采用式(10)一起重构上一级的低频系数,直到最后得到语音帧信号.

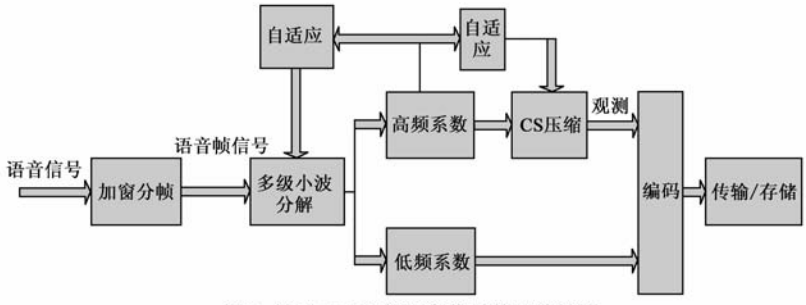

图3　基于AMCS的语音信号的压缩框图

3.3　基于 Sym 正交小波变换的语音信号自适应 MCS

我们采用 SymL 正交小波基,语音信号是近似稀疏的,那么到底分解多少级,重构性能好呢? 每帧语音信号的每级高频系数的稀疏程度都不同,如果采用同样的压缩比不能反映每级高频系数的稀疏程度.由 CS 理论知道,当信号的稀疏性为 $K(K\ll N)$时,采用 CS 压缩得到观测为 $M=4K$ 时可以重构.也就是说随机观测矩阵 $\boldsymbol{\Phi}$ 的选择与每级高频系数的稀疏性 K 有关.那么,每级高频系数 K 具体取多大呢? 不同级 K 不同,所以我们必须首先定义每级高频系数的 K,然后根据具体级高频系数的稀疏性自适应选择分解的级数、自适应选择每级的观测数,以便在信号稀疏的前提下保证好的重构性能.我们提出 AMCS,在客户端进行自适应多级分解联合采样(如图 3),服务器端与 MCS 完全相同.

在客户端自适应多级分解编程思想:(1)语音信号加窗分帧,帧长取 512 点;(2)对帧信号多级小波分解;(3)对高频系数多级分解,每级高频系数绝对值大于阈值门限的个数定义为 K,得到每级的稀疏性 K,观测数大小 $M=4K$,当 M 大于该级高频系数长度时终止分解,信号最多可分解到上一级,对每级系数对应的信号 $\boldsymbol{\beta}_\alpha=\boldsymbol{H}_r\boldsymbol{\alpha}$ 采用 CS 压缩采样得到观测 $\boldsymbol{y}_\alpha=\boldsymbol{\Phi}\boldsymbol{\beta}_\alpha$;(4)最后一级的高频系数矢量 $\boldsymbol{\beta}$ 与所有的 $\boldsymbol{y}_\alpha$ 矢量,采用传统的 LBG 矢量量化方法编码,取 1bit/系数,然后传输.观测矩阵 $\boldsymbol{\Phi}$ 的随机种子、帧长在数据头编码,每级高频系数的稀疏度 K 在帧头编码.

4　实验仿真

接下来我们对本文提出的 MCS 和 AMCS 两种压缩重构方法进行性能仿真.

4.1　数据库描述和评价语音质量的主客观方法

在本文的实验中采用的是中国科学研究院自动化研究所的 CASIA 汉语语音库.在语音库中选 100 人(50 个男性 50 个女性),每人 60 句,句长 1s－5s 不等,采样频率 16kHz,进行实验仿真.本文采用平均帧重构信噪比 AFSNR(式(13))和 PESQ MOS 分来评价重构语音的质量,AFSNR 越大重构性能越好,PESQ MOS 取值范围为 0(最差)~4.5(最好).

$$\mathrm{AFSNR}=\frac{1}{P}\sum_{i=1}^{P}10\log10\left(\frac{\|x_i\|_2^2}{\|x_i-\hat{x}_i\|_2^2}\right) \quad (13)$$

其中 P 是总帧数,x_i 是第 i 帧的数据序列矢量.

4.2　性能仿真

(1)实验 1　通过分析不同的 Sym 下的 AFSNR 和 MOS 分,考察基于 MCS 理论

对语音信号压缩重构的性能.帧长取 512 点,即 32ms,每级的压缩比为 0.1.从我们的实验中发现,基于 MCS 进行多级分解,分解级数超过三级重构语音的质量太差,因此我们着重研究基于 MCS 进行三级分解,对三级高频系数进行压缩采样,对第三级低频系数进行线性采样,总样值为 $n_{ms}=2^6+12+25+51=152$,而直接传输所有系数 $m=512$.压缩比为 152/512,近似为 0.3,也就是利用联合采样样值数只是 Nyquist 采样样值数的 0.3 倍,接下来采用传统的 LBG 矢量量化方法编码,取 1bit/系数的话,数码率为 $152\times1/32=4.75$kbit/s.图 4、5 是实验 1、2 数据的柱状图,实验数据为 100 人实验数据的平均值,不进行 VQ 和进行 VQ 分别标记为 MCS 和 MCS + VQ,可以看出:当压缩比为 0.3 时,基于 MCS 采用 Sym8 重构性能最好,不进行 VQ 和进行 VQ,重构语音的 AFSNR 分别为 21.26dB 和 20.07dB,MOS 分别为 3.55 和 3.48,远优于采用 DCT 基对语音信号进行压缩重构的性能(压缩比为 0.3 时,在我们的实验条件下 AFSNR 和 MOS 分别为 14.347dB 和 2.35),且达到 4.8kb/s 的 CELP 语音编码质量[13].

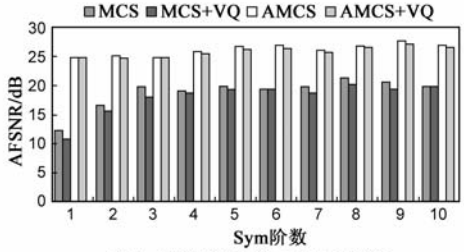

图4 重构语音的AFSNR比较

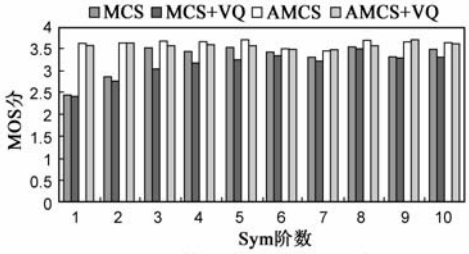

图5 重构语音的MOS分比较

分析采用 Sym8 重构性能优于其他 SymL 的原因主要是由于随着 L 的增大,小波函数和尺度函数的光滑性愈来愈好,但支撑域随之增大,即时域的定域性变差,$1\leqslant L\leqslant10$ 时,对于我们研究的语音信号 Sym8 从光滑性和支撑域来说达到最佳,语音信号在 Sym8 下三级分解,系数最稀疏,因而固定的压缩比时重构语音的质量最好.

分析 VQ 前后实验结果不同的原因.由于 VQ 本身有一定误差,进行 VQ 后重构质量降低一些,但是采用 Sym8 重构语音质量仍可以被接受,未采用 VQ 的实验结果更能说明基于 MCS 对语音信号压缩重构的性能.

(2)实验 2 考察基于 AMCS 理论对语音信号压缩重构的性能.基于 AMCS,首先根据每级高频系数的稀疏度进行自适应压缩采样,然后对最后级的低频系数进行线性采样.根据实验分析阈值取 0.01,实验数据为 100 人实验数据的平均值,不进行 VQ 和进行 VQ 分别标记为 AMCS 和 AMCS + VQ.从图 4、5 可以看出:VQ 前后基于 AMCS,采用 Sym1 ~ 10 重构语音都可以达到 25dB 左右,MOS 分都可以达到 3.5 左右.

分析 VQ 前后实验结果.由于 VQ 本身有一定误差,进行 VQ 后重构质量降低一些,但是无论从 AFSNR 还是 MOS 角度评价重构语音质量,基于 AMCS 比基于 MCS 的固定三级分解重构性能好,主要是由于基于 AMCS 的语音压缩与重构在压缩比、重构语音的性能方面总体达到最优,实现了自适应采样.

当采用 Sym8 时,512 个样点的语音帧采用 AMCS 时其平均采样样点数是 157,压缩比为 157/512,近似为 0.31,采用传统的 LBG 矢量量化方法编码,取 1bit/系数的话,数码率为 $157\times1/32=4.91$kbit/s.与采用 MCS 时码率相近,但是重构语音的 AFSNR 高出 5dB,MOS 分也提高了,算法复杂度仍为 $O(n^3)$,但是系数长度随着级数的增加越来越短,即每级高频系数是上级的长度的一半,算法复杂度可以得到改善.

4.3 对重构语音进行说话人识别测试

为了进一步测试上述基于 CS 理论进行压缩重构的语音质量,特别是语音信号中说话人的声纹特征保真度,实验 3 把实验 1、2 重构的语音用于说话人识别.

首先考虑到实际中模型训练和识别在不同时间且针对不同内容,我们把每个人语音的 1/2 用来训练,另外 1/2 用来识别,在图 6 和表 1 中分别标记为 MCS/2,AMCS/2.预先在服务器端对 100 人的所有原语音的 1/2 进行预处理、特征提取 24 维的 MFCC(帧长为 20ms,帧间重叠 10ms),训练 GMM 模板.为了减小人数增多对识别结果的影响,把 100 人随机分成 10 组,每组 10 人.把采用 Sym8 基于 MCS 和 AMCS 重构的语音(采用 VQ)作为测试语音进行说话人识别,图 6 显示了 10 组的结果,表 1 为平均正确率.从图 6 可以看出,除了个别组,大多数基于 AMCS 重构语音比基于 MCS 三级压缩重构语音说话人识别性能好,主要是由于基于 AMCS 的压缩重构充分考虑了不同语音在小波基下的稀疏程度,根据其在不同级高频系数的稀疏度进行了自适应采样,更好的保留了语音信号的个性特征.个别组例外,分析原因主要是由于 AMCS 采用的阈值是固定的,对个别人来说不一定最佳所造成的.

接下来排除由于训练识别不匹配带来的影响,验证是否能够较高程度保留原说话人的声纹特征.预先在服务器端对 100 人的所有原语音训练模板,把采用 Sym8 基于 MCS 和 AMCS 重构的语音(采用 VQ)作为测

试语音进行说话人识别,从图 6 和表 1 可以看出,识别性能较 MCS/2 和 AMCS/2 有不同程度的提高,从平均值来看基于 AMCS 重构语音比基于 MCS 三级压缩重构语音说话人识别性能好.

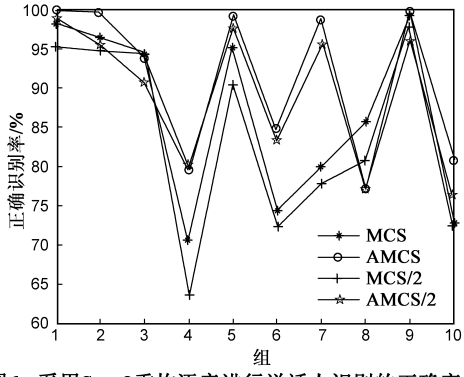

图6　采用Sym8重构语音进行说话人识别的正确率(%)

表 1　说话人识别的平均正确率(%)

类型	MCS/2	AMCS/2	MCS	AMCS
正确率(%)	83.645	89.009	86.622	91.312

5　总结

由于小波变换多级分解合成过程,系数的总长度总在变化,不方便实用,本文首先在系数总长度基本等于原信号长度的前提下,推导了 Sym 小波分解合成的矩阵表示形式,然后对语音信号进行联合采样即把 CS 理论应用到高频系数,而低频系数采用传统的 Nyquist 采样方法.最后基于 MCS 分析语音信号的压缩与重构性能.由于不同语音信号在 Sym 正交小波基下的稀疏性不同,提出了自适应多尺度压缩感知(AMCS)方法,自适应感知每级高频系数的稀疏度得到观测矩阵,自适应决定分解的层数.把 MCS 和 AMCS 方法分别运用到语音压缩与重构中,对利用基追踪重构的语音进行了主客观评价,并进行了说话人识别验证,得出结论:基于 AMCS 的语音压缩重构,实现了自适应采样,且重构语音性能达到最优.后续工作将寻找更适合语音的重构方法,进一步提高重构语音的性能,以便更好的应用于说话人识别等领域.

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

[4] 石光明,等.压缩感知理论及研究进展[J].电子学报,2009,37(5):1070 – 1081.
Shi Guang-ming, et al. Advances in theory and application of compressed sensing[J]. Acta Electronica Sinica, 2009, 37(5): 1070 – 1081. (in Chinese)

[5] Gemmeke J F, hamme H V, et al. Compressive sensing for missing data imputation in noise robust speech recognition[J]. IEEE-Journal of Selected Topics in Signal Processing, Special Session on Compressive Sensing, 2010, 4(2): 272 – 287.

[6] Giacobello D, Christensen M G, et al. Retrieving sparse patterns using a compressed sensing framework: applications to speech coding based on sparse linear prediction[J]. IEEE Signal Processing Letters, 2010, 17(1): 103 – 106.

[7] Xu T T, Yang Z, Shao X. Novel speech secure communication system based on information hiding and compressed sensing [A]. Proceedings of the Fourth International Conference on Systems and Networks Communications[C]. Washington, DC, USA: IEEE Computer Society, 2009. 201 – 206.

[8] 练秋生,等.基于解析轮廓波变换的图像稀疏表示及其在压缩传感中的应用[J].电子学报,2010,38(6):1293 – 1298.
Lian Qiu-sheng, et al. Sparse image representation using the analytic contourlet transform and its application on compressed sensing[J]. Acta Electronica Sinica, 2010, 38(6): 1293 – 1298. (in Chinese)

[9] 李映,等.基于信号稀疏表示的形态成分分析:进展和展望[J].电子学报,2009,37(1):146 – 152.
LI Ying, et al. Advances and perspective on morphological component analysis based on sparse representation[J]. Acta Electronica Sinica, 2009, 37(1): 146 – 152. (in Chinese)

[10] 王大凯,等.小波分析及其在信号处理中的应用[M].北京:电子工业出版社,2006.

[11] S Mallat. A Wavelet Tour of Signal Processing[M]. 2nd ed. New York: Academic Press, 1998.

[12] Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit[J]. SIAM Journal on Scientific Computing, 1998, 20(1): 33 – 61.

[13] 鲍长春.高质量的 4kb/s 散布脉冲 CELP 语音编码算法[J].电子学报,2003,31(2):1 – 5.
BAO Chang-chun. A high quality dispersed-pulse CELP speech coding algorithm at 4kb/s.[J]. Acta Electronica Sinica, 2003, 31(2): 1 – 5. (in Chinese)

作者简介:

孙林慧　女,讲师,1979 年 4 月出生于山西临汾.现为南京邮电大学博士研究生,从事语音信号处理技术方面的研究.
E-mail: sunlh@njupt.edu.cn

杨　震　男,教授、博士生导师,1961 年 11 月出生于江苏苏州.现为南京邮电大学教授,主要研究方向为语音处理与现代语音通信及网络通信技术.