

离散余弦小波包变换及语音信号压缩感知

张长青, 陈砚圃

(西安通信学院, 陕西西安 710106)

摘要: 针对语音信号压缩感知问题, 在研究语音离散余弦变换(Discrete Cosine Transform, DCT)系数和小波包变换(Wavelet Packet Transform, WPT)特性的基础上构造了离散余弦小波包变换(Discrete Cosine Wavelet Packet Transform, DCWPT)。DCWPT 首先获取语音信号的 DCT 域系数, 结合语音频谱特性选取部分 DCT 系数进行 WPT 变换, 从而得到比 DCT 系数更加稀疏的 DCWPT 系数。为将此变换直接用于压缩感知, 构造了 DCWPT 的正交稀疏分解矩阵并分析了其稀疏表示性能。结合稀疏表示基优化了正交匹配追踪重构算法, 提出了基于 DCWPT 的语音信号压缩感知框架。通过压缩重构对照实验, 采用主客观评价指标, 得出该方法优于传统基于 DCT 的语音压缩感知方法的结论。

关键词: 压缩感知; 语音信号; 离散余弦变换; 离散余弦小波包变换; 稀疏表示

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-3630(2014)-01-0035-06

DOI 编码: 10.3969/j.issn1000-3630.2014.01.008

Discrete cosine wavelet packet transform and compressed sensing for speech signal

ZHANG Chang-qing, CHEN Yan-pu

(Xi'an Communications Institute, Xi'an 710106, Shaanxi, China)

Abstract: Concerning the compressed sensing of speech signal, the discrete cosine wavelet packet transform (DCWPT) for speech signal is proposed on basis of the properties of discrete cosine transform and wavelet packet transform. The coefficients of DCWPT can be obtained by wavelet packet transform (DWT) from the coefficients of discrete cosine transform (DCT), and the coefficients are sparser in DCWPT domain than in DCT domain. In order to apply this new efficient transform to the compressed sensing of speech signal successfully, the sparse decomposition matrix of DCWPT is constructed and its performance analyzed. Also the orthogonal matching pursuit reconstruction algorithm is optimized according to the sparse decomposition matrix, and a new framework of the compressed sensing of speech signal based on DCWPT is put forward. It is concluded by subjective and objective indicators from the experiment that the new method is better than the traditional DCT method.

Key words: compressed sensing; speech signal; discrete cosine transform; discrete cosine wavelet packet transform; sparse representation

0 引言

语音压缩^[1]随着信号处理技术的发展, 经历了从时域到变换域再到模型参数的压缩过程, 高质量、高效率的语音压缩方案一直是人们追求的目标。然而, 以前的语音压缩技术都是基于香农采样定理实现的, 这一体系导致的结果就是先采集大量的信号数据, 变换得到稀疏系数后仅保留少量有用的值, 从而造成很大的资源浪费。

压缩感知^[2](Compressed Sensing, CS)理论的出现提供了一种全新的“边采样边压缩”信号获取和处理方式。其采样方法是通过一组特定波形去感知信号, 即将信号投影到给定波形上(衡量与给定波形的相关度), 感知到一组压缩数据; 最后利用最优化的方法实现对压缩数据解压, 估计出原始信号^[3]。可以看出, CS 的提出为信号“节约型”采样和压缩方法带来了新的曙光。CS 理论自提出以来, 在高维图像采集、雷达成像及无线传感器网络等领域显示出极大的生命力, 然而有关语音信号压缩感知的研究相对较少。Christensen^[4]指出 CS 可作为稀疏分解的预处理方法应用于语音和音频信号处理中, 并强调其性能与信号的稳定性和稀疏特性有直接关系。Sreenivas^[5]将 CS 用到了依赖于未知线性变换的稀疏激励语音信号中, 利用匹配追踪(Matching

收稿日期: 2012-10-15; 修回日期: 2012-12-29

基金项目: 国家自然科学基金资助项目(61072125)

作者简介: 张长青(1986—), 男, 陕西眉县人, 硕士, 研究方向为智能信息处理。

通讯作者: 张长青, E-mail: zhangcq1108@163.com

Pursuit, MP)算法实现了信号恢复。Giacobello^[6]将 CS 与稀疏线性预测相结合用于语音编码,得到了较高感知质量的编码语音。初步研究结果表明基于 CS 的语音处理框架是可行的,其关键在语音稀疏表示方法上。DCT 变换以其简单、高效的优点成为了语音稀疏化的首选工具,相关的研究可参考文献[7,8]。Kekre^[9]采用离散余弦小波变换提取了图像的特征矢量,且得到了比 DCT 变换更好的效果;苟平^[10]将 DCT 变换与小波变换用于语音编码得到了较低比特率的编码效果,但如何将离散余弦变换和离散小波变换(Discrete Wavelet Transform, DWT)有机结合构造出新的稀疏变换方法来提高语音稀疏表示效果,以便更好地用于语音 CS 处理仍需研究。

在前人工作的基础上本文分析了语音信号 DCT 变换和 DWT 变换特性,结合 CS 理论构造了适合语音 CS 的 DCWPT 变换矩阵,获得了更稀疏的语音表示形式;在 CS 重构算法中结合 DCWPT 系数特点优化了正交匹配追踪(Orthogonal Matching Pursuit, OMP)算法。实验结果表明基于 DCWPT 的重构语音质量要高于基于 DCT 的重构语音质量。

1 压缩感知基本框架

CS 指出,只要信号是稀疏的或在某个已知的变换域上是稀疏的,就可以通过满足受限等距映射特性(Restricted Isometry Property, RIP)和不相关特性^[11]的降维观测矩阵实现对信号的感知采样。重构时通过最优化重构算法就可以从低维的采样信号中恢复出高维的原始信号。下面从 CS 的三个关键环节介绍其基本框架。

1.1 稀疏变换

采用 CS 方法对某一信号处理的前提条件是:信号必须有已知的稀疏表示形式。已知某一维离散信号为 $\mathbf{x}=[x(1), x(2), \dots, x(N)]^T \in \mathbf{R}^{N \times 1}$, 正交稀疏变换矩阵可表示为 $\Psi \in \mathbf{R}^{N \times N}$, 则 $\boldsymbol{\theta}=\Psi \mathbf{x}=[\theta(1), \theta(2), \dots, \theta(N)]^T \in \mathbf{R}^{N \times 1}$ 为信号 \mathbf{x} 在变换基 Ψ 下的系数。如果有 $0 < p < 2$ 且 $R > 0$, $\boldsymbol{\theta}$ 满足 $\|\boldsymbol{\theta}\|_p \leq R$, 那么 $\boldsymbol{\theta}$ 就可以认为是稀疏的^[2]。为便于描述和对比,本文采用 Gini 系数^[12]方法对信号表示形式的稀疏性进行度量, Gini 系数介于 0 和 1 之间,其值越大说明越稀疏。

1.2 信号压缩

在 CS 中信号是通过一个从 N 维到 M 维($M \ll N$)的降维观测矩阵 Φ 实现压缩的。压缩过程如式(1)所示:

$$\mathbf{a}=\Phi \mathbf{x}=\Phi \Psi^{-1} \boldsymbol{\theta}=\mathbf{A}^{\text{CS}} \boldsymbol{\theta} \in \mathbf{R}^{M \times 1} \quad (1)$$

此处要求观测矩阵与稀疏变换矩阵不相干(当观测矩阵与稀疏变换矩阵的相干度为 1 时, CS 即退化为常规采样方式),目前多采用随机矩阵作为 CS 的观测矩阵。

1.3 重构算法

重构就是要从 \mathbf{a} 中恢复出信号 \mathbf{x} 。实际上这是一个未知数个数大于方程个数的病态方程求解问题,庆幸的是 $\boldsymbol{\theta}$ 可以作为由 \mathbf{a} 到 \mathbf{x} 的“桥梁”从而使问题变得可解。若 $\boldsymbol{\theta}$ 是 K -稀疏的,通常有 $N > M = cK \lg(N/K) > K$ (c 为常数,一般取 4 左右),那么就可以很容易地由 M 个方程解出 K 个未知数。问题是, K 个大系数在 $\boldsymbol{\theta}$ 中的位置是自由的,因此该问题的解决就归结到了寻求最优解。该过程可描述为式(2):

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_0 \quad \text{s.t.} \quad \mathbf{a}=\Phi \mathbf{x}=\mathbf{A}^{\text{CS}} \boldsymbol{\theta} \quad (2)$$

然而,这是一个 NP-hard 问题,用 L^1 范数优化目标代替 L^0 范数可以将其转化为线性规划问题,如式(3)所示:

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \quad \mathbf{a}=\Phi \mathbf{x}=\mathbf{A}^{\text{CS}} \boldsymbol{\theta} \quad (3)$$

求解此类问题的代表算法为 OMP 算法。

2 离散余弦小波包域的语音压缩感知

卡洛变换(Karhunen Loeve Transform, KLT)是均方意义下的最优变换,不过 KLT 的变换性能具有数据依赖性,实现比较困难。DCT 作为最接近 KLT 的正交变化具有和离散傅里叶变换(Discrete Fourier Transform, DFT)一样的频率特性和快速算法,因此通常选择 DCT 作为语音信号的首选稀疏变换方法。语音信号的 DCT 系数能较好地体现出信号能量在频域的集中分布特性,它是近似稀疏的。要采用 CS 方法提高语音信号的压缩性能还需寻找更加稀疏的语音表示方式。文献[10]指出,只要能较好地逼近语音信号在频域的能量包络就可以获得比较理想的重构语音质量。本节的主要任务就是在语音 DCT 变换系数的基础上寻找对其再压缩的方法。

2.1 离散余弦小波包变换

小波变换(Wavelet Transform, WT)是一种多尺度变换,在任一尺度上信号都可以被分解为描述信号变化趋势的“近似”和描述信号细微变化的“细节”信息,利用 WT 可以实现对信号不同频段的独立处理, WPT 可得到比 WT 更精细的信号处理能力。结合 DCT 系数分布较集中的特点,我们采用 WPT 对 DCT 系数进行再压缩以获得比 DCT 系数更稀疏的表示形式,称这一复合稀疏变换为 DCWPT。

基于 DCWPT 的语音 CS 压缩框架如图 1 所示，虚线框表示的即为 CS 观测，为了便于理解和描述，把对 x 的观测转化为对稀疏系数 θ 的观测，通过 Gini 系数方法可以计算出 θ 中有效大系数的个数 K ，其值可用于确定观测矩阵的维数。

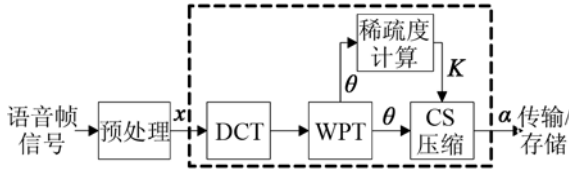


图 1 基于 DCWPT 的语音 CS 压缩框架

Fig.1 The compression structure of CS for speech signal based on DCWPT

具体压缩过程描述如下：

(1) 语音帧预处理：该部分包括去直流和带通滤波两个步骤。实验发现在语音重构时即使去掉 150Hz 以下的频率分量，人耳也能自动“填补”上相关信息而不会使主观听觉感受很差，因此滤波通带选为 150~3400Hz。

(2) 稀疏变换：经预处理的语音帧首先在 DCT 变换基下获取 DCT 系数，截取系数中 150~3400Hz 的频率段再做 WPT 压缩。考虑到单频信号 DCT 系数正负交替出现的特点以及小波所对应的分解和重构正交滤波器的最小相位移，经过实验对比选择 db2 小波；综合考虑小波包分解精度和运算的复杂性，分解层数确定为 2 层。在对四个小波包系数压缩处理时区别对待两个低频系数和两个高频系数，采用不同的阈值对其压缩以得到最终的稀疏表示形式。

(3) CS 压缩：此处压缩是通过对小波包系数的观测实现的。为了使观测更加有效，在这里引入了自适应稀疏度计算准则，方法如式(4)所示，通过计算得到稀疏度 K 后就可以用 $M \approx 3K$ 个观测值实现对该帧语音的压缩。

$$K = N(1 - gini_{\theta}) \quad (4)$$

式中 N 为系数长度， $gini_{\theta}$ 为稀疏表示 θ 的 Gini 系数。

为适应 CS 处理方式，在此介绍一下稀疏变换基 Ψ 的构造。DCT 同 DFT 一样有快速算法，其变换基可表示为 Ψ_D ，这里主要说明 WPT 的矩阵构造方法。db2 小波分解矩阵可表示为 Ψ_w ，其中 c_0 、 c_1 、 c_2 、 c_3 为小波滤波器系数。经 Ψ_w 分解后，“近似”系数都在奇数行上而“细节”系数都在偶数行上，再经过“排序”矩阵 S 后就可以将所有“近似”按顺序放在变换系数的前半部分而将所有“细节”按顺序放在变换系数的后半部分，至此就实现了一次小波分解。对得到的“近似”和“细节”分别用

上述步骤再分解就可以实现 2 层 WPT 了。 Ψ_w 和 S 的构造形式如下：

$$\Psi_w = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ c_3 & -c_2 & c_1 & -c_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_0 & c_1 & c_2 & c_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_3 & -c_2 & c_1 & -c_0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & c_0 & c_1 & c_2 & c_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_3 & -c_2 & c_1 & -c_0 \\ c_2 & c_3 & 0 & 0 & 0 & 0 & 0 & 0 & c_0 & c_1 \\ c_1 & -c_0 & 0 & 0 & 0 & 0 & 0 & 0 & c_3 & -c_2 \end{bmatrix}_{N \times N}$$

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{N \times N}$$

综上，DCWPT 变换基可以描述为式(5)：

$$\Psi = \begin{bmatrix} (S\Psi_w)_{N \times N} & 0 \\ 0 & (S\Psi_w)_{N \times N} \end{bmatrix} (\Psi_D)_{2N \times 2N} \quad (5)$$

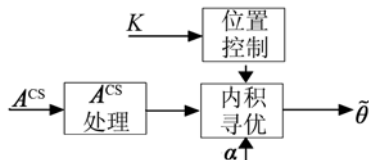
从 Ψ 的构造可以看出 $\Psi^{-1} = \Psi^T$ ，这一特性有助于提高由 $\tilde{\theta}$ 到 \tilde{x} 的重构效率，可以说构造的 DCWPT 与 DFT 一样具有快速算法。

2.2 重构算法的优化

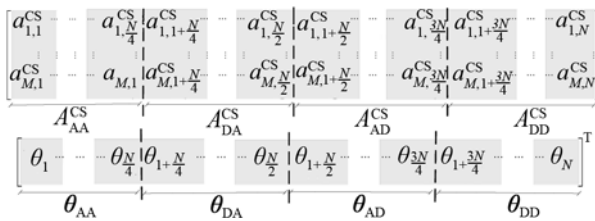
重构是压缩的一个逆过程，其目标是使用较少的观测值和相对小的算法开销获取符合人耳听觉感受需要的重构语音。下面结合图 2 说明基于 DCWPT 的 OMP 重构算法的具体优化改进方案。

在 CS 中 OMP 算法主要是用观测向量余量(第一次为 α)与 CS 算子 A^{CS} 的所有列向量求内积，通过寻找内积最大值对应的列向量位置确定重构 $\tilde{\theta}$ 时大系数的位置(如果某次重构时内积最大对应的是 A^{CS} 中的第 i 列，那么此次重构的即为 $\tilde{\theta}(i)$ 的值)，利用最小二乘法计算该系数的值。从式(1)和式(5)可以看出 DCWPT 系数是由 4 段小波系数构成的，从图 2(b)可以清楚地看到这一点。作为 DCT 系数的 WPT 分解， θ 的 4 段系数对重建 DCT 系数质量的贡献是不尽相同的，因此我们将 θ 中的 K 个大系数“名额”按照 4:2:3:1 的比例分配给 θ_{AA} 、 θ_{DA} 、

θ_{AD} 和 θ_{DD} 。这一过程是通过观测余量与 A^{CS} 列向量求内积时最大值对应的列向量落在 A^{CS} 中对应的四个区域的比例来控制的, 对应于图 2(a)的位置控制模块。此外, 考虑到人耳特性, 我们对 150~3400 Hz 以外的 DCT 系数不予重构, 相对应的就是将 A^{CS} 四个区域两端位置(图中阴影区域)的若干列向量置零, 其操作对应于图 2(a)中的 A^{CS} 处理模块。



(a) 基于 DCWPT 的 OMP 算法优化框图



(b) A^{CS} 和 θ 结构图

图 2 基于 DCWPT 的 OMP 算法优化
Fig.2 Optimization of OMP based on DCWPT

通过上述算法优化过程可以看出: (1) 在内积寻优过程中引入位置控制信息, 将有限的观测值都用于重构对后续恢复语音质量有较大贡献的 K 个稀疏系数上, 这样可以提高观测数据的使用效率; (2) 引入 A^{CS} 处理模块, 仅对 A^{CS} 有效区域的列向量进行内积运算, 缩小了寻优范围, 提高了算法效率。

3 实验仿真

为说明语音信号在 DCWPT 域的稀疏表示性能, 对一帧浊音信号分别采用 DCT 和 DCWPT 两种变换进行稀疏表示及信号重构对照实验加以验证, 实验具体过程描述如下:

(1) 取一帧浊音信号求出其 DCT 系数。

(2) 对 DCT 系数用式(4)计算其大系数个数 K , 并通过阈值法保留 K 个 DCT 大系数作为阈值 DCT 系数, 再通过 IDCT 变换重构原信号并计算信噪比和压缩比。

(3) 截取阈值 DCT 系数在 150~3400Hz 频段内的系数进行 2 层 db2 小波包分解, 对分解得到的四段系数分别计算其阈值并采用硬阈值法压缩以获取 DCWPT 系数。重构时先由小波包系数恢复 DCT 系数, 再对 150~3400 Hz 以外的频率段补零, 最后通过 IDCT 变换得到重构信号, 并计算信噪比和压缩比。

压缩比 R 的定义如式(6)所示:

$$R_{\theta} = \frac{\|\mathbf{x}\|_0}{\|\theta\|_0} \quad (6)$$

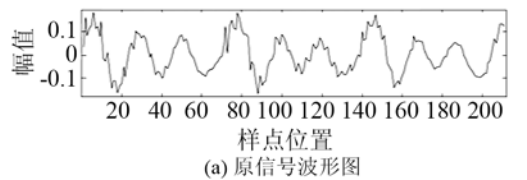
式中 θ 为信号 \mathbf{x} 的稀疏表示形式, $\|\mathbf{x}\|_0$ 表示 \mathbf{x} 中非零元素的个数。实验结果如表 1 和图 3 所示。

表 1 稀疏表示方法性能指标对比
Table 1 Comparison of sparse representations

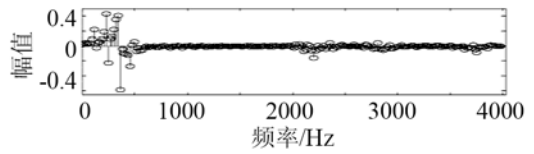
信号形式	gini 系数	重构信噪比/dB	压缩比
原时域信号	0.36	inf	1
DCT 系数	0.75	inf	1
阈值 DCT 系数	0.87	17.5	3.7
DCWPT 系数	0.91	14.8	7.2

说明: (1) 信噪比为“inf”代表没有重构误差;

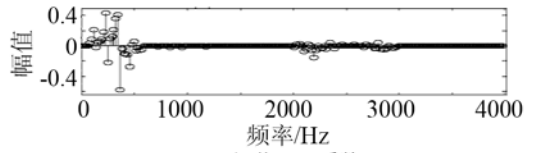
(2) 表中压缩比仅为原时域信号中非零元素个数与稀疏系数中非零元素个数的比值, 并未考虑 CS 观测与重构。



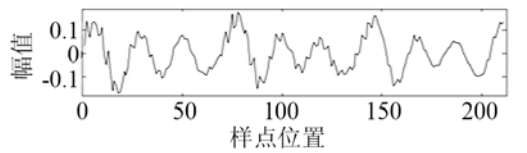
(a) 原信号波形图



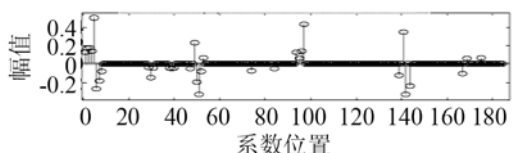
(b) 原信号 DCT 系数



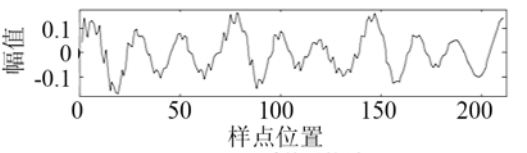
(c) 阈值 DCT 系数



(d) 阈值 DCT 系数重构波形



(e) DCWPT 系数



(f) DCWPT 系数重构波形

图 3 浊音帧信号在 DCT 和 DCWPT 变换下的压缩重构图

Fig.3 Sparse representations of a frame of voiced speech under DCT and DCWPT

从表 1 可以看出 DCT 系数是比较稀疏的, 而 DCWPT 稀疏表示方法要比 DCT 和阈值 DCT 更加

有效。从压缩重构效果图看，DCWPT 系数可以准确地恢复出语音的轮廓和绝大部分细节信息，从而能得到比较满意的重构语音。经实验分析，主要是因为 DCWPT 系数较好地提取了 DCT 系数的频谱变化趋势信息和少量细节信息。需要说明的是这里没有采用 CS 框架压缩重构，若将 DCWPT 用于语音 CS，还需经过随机观测和优化重构，这将使得压缩比和重构信号质量均略有下降。

为了验证 DCWPT 变换基在语音信号 CS 中的实际应用效果，本文选取中国科学院自动化所语音库中的 10 句男声和 10 句女声作为测试语音，采用 DCWPT 作为稀疏表示基，采用降维($M \approx 3K$)随机矩阵作为观测矩阵，采用基于 DCWPT 的优化 OMP 算法作为重构算法构建语音 CS 框架。利用该框架对测试语音做压缩处理，在重构端选用 P.862 标准的主观平均意见分(Mean Opinion Score, MOS)对重构语音做主观评价，为确保评分结果的可比较性采用 GL Voice Quality Testing(VQT)打分软件获取 MOS 分，测试时对每一个测试组(一句语音为一个测试组，共 20 组)分别采用上述方法进行压缩和重构，通过保留稀疏表示系数中非零元素个数的多少来控制压缩比，在不同的压缩比下求取重构语音的 MOS 分。每组实验重复 3 次，最终的 MOS 分取其平均值。此外，还采用平均帧重构信噪比(用 AFSNR 表示)说明重构语音性能。

$$AFSNR = \frac{1}{N_f} \sum_{n=1}^{N_f} 10 \lg \left(\frac{\|x_n\|_2^2}{\|x_n - \tilde{x}_n\|_2^2} \right) \quad (7)$$

式中： N_f 表示总帧数； x_n 和 \tilde{x}_n 分别表示第 n 帧语音和重构语音，在压缩比一定时 AFSNR 越大说明重构语音越好。

为了便于说明基于 DCWPT 的语音 CS 框架性能，实验用基于 DCT 的语音 CS 框架作为对照。图 4 为其中一段女声在不同方法下的重构语谱图和分段压缩比，表 2 为不同压缩比时的重构语音 MOS 得分，图 5 为不同压缩比下的平均帧重构信噪比。

从重构语谱图来看，基于 DCT 的语音 CS 能重构出能量较大的语音频率成份，从而保证了重构语音质量。在 DCT 压缩系数的基础上获取 DCWPT 压缩系数，并采用基于 DCWPT 的 OMP 优化重构算法重构语音，从语谱图可以看出其与阈值 DCT 重构质量接近，在保证一定重构语音质量的前提下，经测试，当 DCT 方法压缩比为 2.7:1 时 MOS 得分为 3.41，同等条件下 DCWPT 方法的 MOS 得分为 3.13 而压缩比高达 4.4:1。从分段压缩比可以看出 DCWPT 框架的压缩性能要明显优于 DCT 框

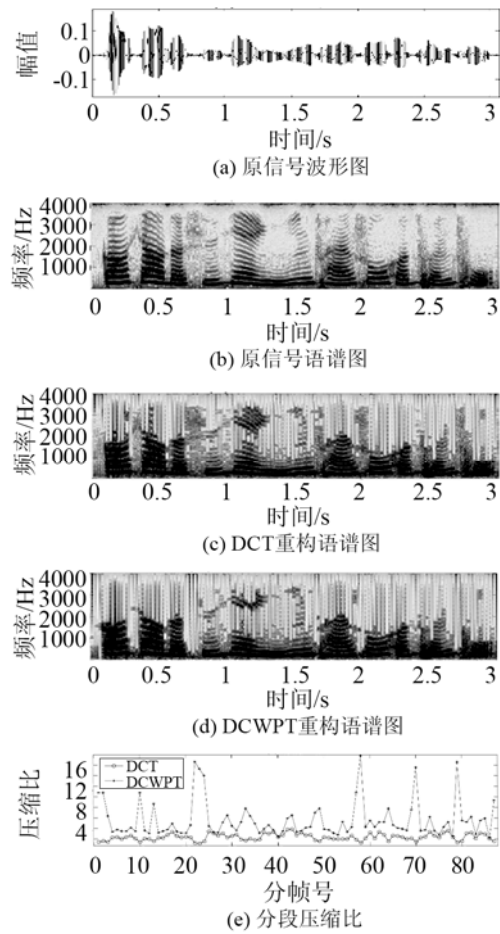


图 4 基于 DCT 和 DCWPT 的语音 CS 性能对比图
Fig.4 Performances of CS for speech based on DCT and DCWPT

表 2 不同压缩比时重构 MOS 分
Table 2 MOS scores for different compression ratios

方法	MOS 分						
	压缩比=1:1	2:1	3:1	4:1	5:1	6:1	7:1
DCT	4.05	3.61	3.34	3.04	2.69	2.23	1.82
DCWPT	4.14	3.75	3.54	3.28	2.87	2.48	2.16

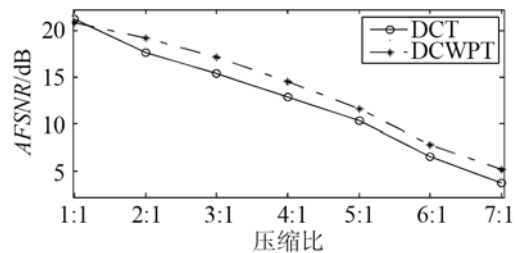


图 5 基于 DCT 和 DCWPT 的平均帧重构信噪比
Fig.5 The AFSNRs based on DCT and DCWPT

架，特别在清音段的压缩比更大，通过两种框架下重构语音主观感受对比发现，清音段的这种大比例压缩对重构语音整体质量的影响并不大。表 2 和图 5 列出了不同压缩比下 DCT 和 DCWPT 两种 CS 框架下重构语音的 MOS 得分和平均帧重构信噪比，可以看出基于 DCWPT 的语音 CS 性能优于基于

DCT 的语音 CS 性能。

4 结 论

本文在 DCT 和 WPT 基础上提出了 DCWPT 稀疏表示方法并构造了快速 DCWPT 的稀疏分解矩阵, 基于 DCWPT 变换提出了语音 CS 新框架并优化了 OMP 重构算法。实验证明, DCWPT 用于语音稀疏表示是有效的, 且在同一压缩比下能获得比 DCT 更高的语音重构质量, 而 DCWPT 方法的计算复杂度与 DCT 方法相当。然而, 在 DCWPT 变换中只是简单地保留了 150~3400 Hz 的频谱, 并没有充分利用心理声学模型展开更精细的研究。相信在稀疏域关键参数的选取问题上引入心理声学模型将更有助于稀疏性能的改善而不使重构语音质量恶化, 因此如何将心理声学模型引入稀疏表示方法中并构造出快速的矩阵算法是一个新的研究方向。

参 考 文 献

- [1] 时磊, 尚秋峰, 李建宇. 语音压缩编码技术研究进展[J]. 声学技术, 2008, 27(4): 553-556.
SHI Lei, SHANG Qiufeng, LI Jianyu. The research progress in speech compression coding technology[J]. Technical Acoustics, 2008, 27(4): 553-556.
- [2] Donoho D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.
- [3] 戴琼海, 付长军, 季向阳. 压缩感知研究[J]. 计算机学报, 2011, 34(3): 425-434.
- DAI Qionghai, FU Changjun, JI Xiangyang. Research on compressed sensing[J]. Chinese Journal of Computers, 2011, 34(3): 425-434.
- [4] Christensen M G, Stergaard J, Jensen S H. On compressed sensing and its application to speech and audio signals[C]// Proceedings of the Forty-Third Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2009: 356-360.
- [5] Sreenivas T V, Kleijn W B. Compressive sensing for sparsely excited speech signals[C]// IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, 2009: 4125-4128.
- [6] Giacobello D, Christensen M G, Murthi M N, 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] Kassim Liban A, Khalifa Othman O, Gunawan T S. Evaluation of sparsifying algorithms for speech signals[C]// International Conference on Computer and Communication Engineering, Kuala Lumpur, 2012: 308-313.
- [8] 郭海燕, 王天荆, 杨震. DCT 域的语音信号自适应压缩感知[J]. 仪器仪表学报, 2010, 31(6): 1262-1268.
Guo Haiyan, Wang Tianjing, Yang Zhen. Adaptive speech compressed sensing in the DCT domain[J]. Chinese Journal of Scientific Instrument, 2010, 31(6): 1262-1268.
- [9] Kekre H B, Sonawane Kavita. Retrieval of Images using DCT and DCT wavelet over image blocks. international journal of advanced computer science and applications, 2011, 2(10): 98-106.
- [10] 苟平. 基于语音信号 DCT 系数的小波变换编码研究[D]. 成都: 四川大学, 2005: 29-34.
GOU Ping. A wavelet transform coding for voice's DCT[D]. Sichuan: Sichuan University, 2005: 29-34.
- [11] Baraniuk R. G. Compressive sensing[J]. IEEE Signal Processing Magazine, 2007, 24(4): 118-121.
- [12] Hurley N, Rickard S. Comparing measures of sparsity[J]. IEEE Transactions on Information Theory, 2009, 55(10): 4723-4741.