

文章编号:1007-5321(2017)06-0057-08

DOI:10.13190/j.jbupt.2017-021

# 基于移动通信电路域的语音信息隐藏算法

薛一鸣<sup>1</sup>, 何宁宁<sup>1</sup>, 陈 鹄<sup>1</sup>, 胡彩娥<sup>2</sup>, 牛少彰<sup>3</sup>

(1. 中国农业大学 信息与电气工程学院, 北京 100083; 2. 国网北京市电力公司, 北京 100031;  
3. 北京邮电大学 计算机学院, 北京 100876)

**摘要:** 通过分析移动通信电路域语音信道的结构特点,提出了一种适用于移动通信电路域信道的抗语音声码器断续传输(DTX)处理和有损压缩的语音信息隐藏算法. 选择经语音端点检测判决为浊音帧的有效语音信号进行3级小波分解,选择第3级小波细节系数做快速傅里叶变换(FFT)后,在表征频带0.5~1 kHz的FFT系数上根据听觉掩蔽理论对秘密信息进行量化抖动调制嵌入,实现秘密信息在语音中的隐藏. 实验结果表明,该算法模型在10 bit/s上限容量时,携密语音质量评价分数高于3.48,经过全球移动通信系统增强型全速率声码器的DTX技术处理、有损压缩以及信道中存在的其他干扰后,误码率保持在0.06左右.

**关键词:** 语音信息隐藏; 断续传输; 语音端点检测; 有损压缩

中图分类号: TN91

文献标志码: A

## Speech Hiding Algorithm Based on Mobile Circuit-Switched Domain

XUE Yi-ming<sup>1</sup>, HE Ning-ning<sup>1</sup>, CHEN Yao<sup>1</sup>, HU Cai-e<sup>2</sup>, NIU Shao-zhang<sup>3</sup>

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China;  
2. State Grid Beijing Electric Power Company, Beijing 100031, China;  
3. School of Computer Science, Beijing University of Posts and Telecommunication, Beijing 100876, China)

**Abstract:** Based on analysis of circuit-switched mobile channels, an algorithm for speech information hiding was proposed, which is robust for the speech vocoder discontinuous transmission (DTX) mechanism and the lossy compression. The algorithm performs 3-level discrete wavelet transform decomposition on the voiced speech frames decided by voice activity detector, and selects the 3rd level detail coefficients to implement fast Fourier transform(FFT). By means of the auditory masking theory, the rational dither modulation is used to embed the secret data on FFT coefficients within the frequency band of 0.5 ~ 1 kHz. The experimental results show that the Perceptual evaluation of speech quality score of stego speech is higher than 3.48, and the bit error rate is about 0.06 under the interference of DTX of the global system for mobile communication enhanced full rate vocoder, the lossy compression and some other attacks in the channels.

**Key words:** speech information hiding; discontinuous transmission; voice activity detector; lossy compression

基于语音的信息隐藏一直是信息安全领域的研究热点<sup>[1-2]</sup>,按照语音通信方式的不同可分为基于互

收稿日期: 2017-03-08

基金项目: 国家自然科学基金项目(U1536121)

作者简介: 薛一鸣(1968—),男,副教授,硕士生导师, E-mail: xueym@cau.edu.cn.

联网协议电话的隐藏和基于移动通信语音电路域交换业务的隐藏. 基于互联网协议电话的隐藏研究多围绕声码器展开, 这些方法通过对声码器的结构或编码码流进行修改能取得较大的嵌入容量<sup>[3]</sup>. 移动通信语音电路域交换信道因其稳定的通信质量和较低的通信延迟成为语音通信中实时信息隐藏的天然信道. 在基于移动通信语音电路域的隐藏研究中, Ahani 等<sup>[4]</sup>采用离散小波变换(DWT, discrete wavelet transform)结合稀疏分解实现信息在语义层次上的隐藏, 有效提高了携密语音的不可感知性; Kazemi 等<sup>[2]</sup>研究抗自适应多速率 12.2 kbit/s 编码速率的改进型扩频隐藏方法, 取得了较好的抗有损压缩性能, 但容量较小; Rekik 等<sup>[5]</sup>通过替换 DWT 和快速傅里叶变换(FFT, fast Fourier transform)后的高频系数实现语音参数的隐藏. 这些传统的基于对语音载体在时域或变换域做修改实现隐藏的方法并没有考虑实际语音通信信道的特性, 除了改进型扩频法以外大都不能抵抗声码器中断续传输(DTX, discontinuous transmission)技术对提取端误码率的影响, 对声码器的有损压缩鲁棒性也表现不佳.

不可感知性、不可检测性、鲁棒性和嵌入容量是语音信息隐藏算法常用的4个评价指标. 不可感知性作为基本指标表示携密语音与原始载体语音相比没有引入人耳听觉上的失真. 不可检测性表征算法抗隐写分析的强度, 目前主流的隐写分析方法有倒谱系数、支持向量机以及动态贝叶斯网络等<sup>[6-7]</sup>, 基本原理是通过设计各种分类器来实现对原始语音和携密语音的区分. 鲁棒性表示算法抵抗各种常见数字信号处理技术攻击的健壮性. 容量则用于衡量载体语音所能隐藏秘密信息的多少, 常用每秒传送位数(bit/s)表示.

在基于移动通信语音电路域信道的信息隐藏研究中, 信道环境的高复杂性使得降低秘密信息的提取误码率成为研究的重点和难点, 因此鲁棒性(这里表现为抵抗信道中各种干扰的性能)是除了不可感知性之外衡量该应用场景下算法性能的重要指标. 限于篇幅, 将围绕不可感知性和鲁棒性的提高展开讨论, 分析移动通信语音电路域交换信道对语音的传输处理机制, 充分考虑抗增益干扰、加性高斯白噪声干扰、全球移动通信系统(GSM, global system for mobile communication)增强型全速率(EFR, enhanced full rate)声码器 DTX 技术和有损压缩等鲁

棒性要求, 在此基础上提出基于语音端点检测-VAD-DWT-FFT(VAD-DWT-FFT, voice activity detector-DWT-FFT)的新型信息隐藏与提取模型, 并根据人耳听觉掩蔽理论和秘密信息对语音 0.5 ~ 1 kHz 频带的能量进行量化抖动调制(RDM, rational dither modulation). 实验结果表明, 所提出的隐藏模型在嵌入容量上限为 10 bit/s 时, 携密语音感知质量评价(PESQ, perceptual evaluation of speech quality)分数达到 3.48 以上, 同时保持误码率在 0.06 左右.

## 1 信道特性分析及算法介绍

### 1.1 电路域信道特点与 DTX 技术

如图1所示, 移动通信语音电路域交换信道中, 语音从传输到接收的完整过程可简化为手机A—基站A—移动交换中心A—移动交换中心B—基站B—手机B这一框架.

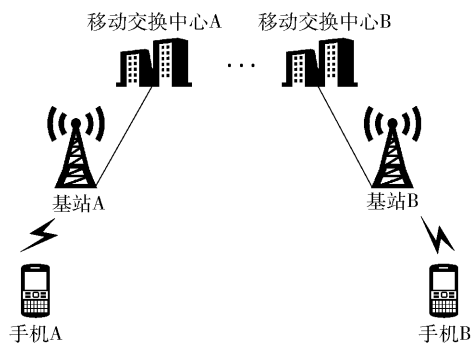


图1 移动通信语音电路域传输框架

进一步分析得到图2所示的移动通信语音电路域传输信道结构, 其中手机包括模块1 模数/数模(AD/DA, analog digital/digital analog)转换器和模块2 语音转码器及检测器, 模块1 实现模拟语音和脉冲编码调制(PCM, pulse code modulation)格式数字语音的相互转换, 而模块2 则实现 PCM 格式的数字语音和声码器压缩编码格式码流的相互转换. 对于发送端, 手机将声码器压缩编码后的码流经天线射频发送给基站系统, 在接收端则通过天线接收基站系统传输的待声码器解压缩的码流. 值得注意的是, 模块2 中采用了 DTX 技术, 对经 VAD 判决后标识为有效语音帧的 PCM 格式数字语音做声码器压缩并正常传输, 而对噪声帧和静音帧仅简单提取能表征能量和谱结构的参数进行 DTX.

手机对语音的处理过程要求信息隐藏提取算法模型能有效抵抗 AD/DA 转换、采样、量化、增益改

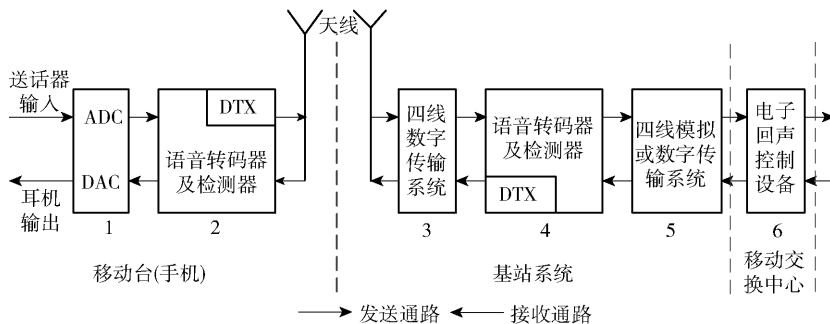


图2 移动通信语音电路域传输信道结构

变以及 GSM EFR 声码器 DTX 技术和有损压缩的干扰。语音声码器中的 DTX 技术决定了噪声帧和静音帧中的语音无法作为隐藏载体。这意味着隐藏端需要有效语音的判决机制来选择能兼容声码器 DTX 技术的有效语音帧位置进行秘密信息隐藏。基站系统由模块 3 四线数字传输系统、模块 4 语音转码器及检测器和模块 5 四线模拟或数字传输系统组成,移动交换中心则主要包括模块 6 电子回声控制设备。模块 3、模块 5 和模块 6 可认为是语音传输的主要路由,会带来通信中语音的音量改变和回声叠加问题;模块 4 同模块 2 原理一致,两者配合实现语音信号的声码器编解码。基站系统和移动交换中心对通信语音的处理要求信息隐藏提取算法模型除了能抵抗手机所带来的干扰类型外,还需对回声具有一定的健壮性。

## 1.2 DWT

DWT 是分析语音这类非平稳信号的有效技术。DWT 具有多分辨率的特性,每一级的 DWT 都能将对应频带的信号分解为低频和低频 2 个子带信号,并分别称为近似系数和细节系数:

$$C^K(i) = \sum_{j=0}^{J_\omega-1} h(j-2i)C^{K-1}(j) \quad (1)$$

$$D^K(i) = \sum_{j=0}^{J_\omega-1} g(j-2i)D^{K-1}(j) \quad (2)$$

其中:  $C^K(i)$  和  $D^K(i)$  分别为第  $K$  级 DWT 后得到的近似系数和细节系数,  $h(j)$  和  $g(j)$  分别为正交小波的低通滤波器系数和高通滤波器系数,  $J_\omega$  为小波滤波器的长度。进行了三级 DWT, 并选择频率覆盖范围为  $[0.5f_s/2^3, 0.5f_s/2^2]$  的第 3 级的细节系数  $D^3(i)$  作为秘密信息的嵌入频带, 其中  $f_s$  为语音信号的 A/D 采样率, 这里  $f_s = 8 \text{ kHz}$ 。

## 1.3 听觉掩蔽

根据听觉掩蔽理论<sup>[8]</sup>, 只要秘密信息嵌入后对

原始载体语音所造成的能量失真小于掩蔽阈值, 则携密语音不会带来人耳听觉感知的异常。选择频带的中心频率值  $f_{\text{rep}}$  计算 Bark 谱如下:

$$z_{\text{rep}} = 13 \arctan(0.00076f_{\text{rep}}) + 3.5 \arctan((f_{\text{rep}}/7500)^2) \quad (3)$$

由此可得到听觉掩蔽阈值的估计值为

$$a(z_{\text{rep}}) = \lambda a_{\text{tinn}}(z_{\text{rep}}) + (1 - \lambda) a_{\text{nnn}}(z_{\text{rep}}) \quad (4)$$

其中:  $\lambda$  为介于 0 ~ 1 之间的音调参数;  $a_{\text{tinn}}$  为音调掩蔽索引,  $a_{\text{tinn}}(z) = -0.275z - 15.025$ ;  $a_{\text{nnn}}$  为噪声掩蔽索引, 通常取固定值 -9。将上述参数代入式(4)求得听觉掩蔽阈值  $a(z_{\text{rep}}) \geq a_{\text{tinn}}(z_{\text{rep}})$ , 即在人耳听觉系统的异常感知范围内, 由中心频率  $f_{\text{rep}}$  所确定的频带的能量改变最大允许等级为  $a_{\text{tinn}}(z_{\text{rep}})$  dB。用  $E_{\text{ori}}$  表示频带的原始能量, 则最大能量失真  $E_\Delta$  表示为<sup>[9]</sup>

$$E_\Delta = 10^{\frac{a_{\text{tinn}}(z_{\text{rep}})}{10}} E_{\text{ori}} \quad (5)$$

## 2 VAD-DWT-FFT 信息隐藏算法

根据移动通信系统中声码器对输入语音的处理机理, 提出适用于移动通信语音电路域信道的 VAD-DWT-FFT 隐藏和提取算法模型, 如图 3 所示。

### 2.1 VAD 判决算法

采用 VAD 判决算法的目的是从语音流中挑选出可用于信息隐藏的有效语音帧, 实现秘密信息对声码器 DTX 技术良好的通透性。VAD 判决算法的设计可分为 2 种思路, 表示为式(6)和式(7)。

$$\{x_{\text{VAD}}\} \subset \{x_{\text{DTX}}\} \quad (6)$$

式(6)表示 VAD 判决算法所挑选的有效语音帧集合是声码器 DTX 技术判决语音帧的真子集, 此方法可有多种实现方式, 如简单逆滤波跟踪法和平均幅度差函数法等, 设计灵活, 复杂度也较低, 对有效语音中的浊音具有很高的判决准确度。

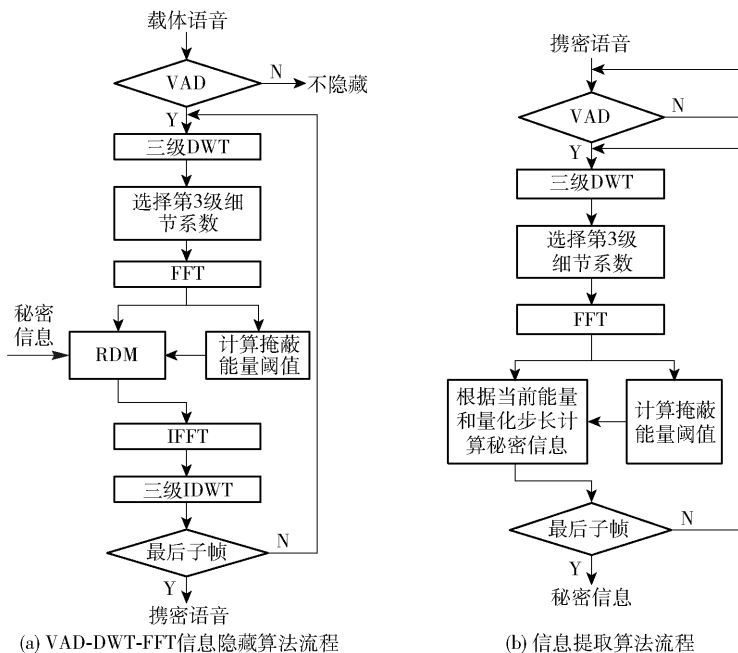


图3 VAD-DWT-FFT 信息隐藏和提取模型

$$\{x_{\text{VAD}}\} = \{x_{\text{DTX}}\} \quad (7)$$

式(7)表示 VAD 判决算法所挑选的有效语音帧集合与声码器 DTX 技术判决的语音帧集合相等. 显然, 这要求 VAD 判决算法与 DTX 技术所采用的算法完全一致. 相比第 1 种设计思路, DTX 技术中的判决算法往往是多种语音检测算法的组合, 能实现基音检测、音调检测和复杂信号检测等, 对语音信号和噪声信号区分效果较好.

由于有效语音帧中清音的频谱结构过于简单, 在清音段做隐藏会对其频谱造成明显改变, 不仅语音质量下降明显, 而且容易引起第三方窃密者的怀疑. 所以笔者采用第 1 种设计思路, 选择有效语音帧中的浊音帧作为实际的载体语音, 能提高 VAD-DWT-FFT 隐藏提取模型的不可感知性能.

## 2.2 信息隐藏与提取算法

在采用 VAD 判决算法对载体语音进行预处理, 获得对抗信道中声码器 DTX 技术的鲁棒性后, 信息隐藏提取模型的设计关键在于解决声码器混合编码方式所带来的有损压缩问题. 声码器混合编码的本质是提取语音的特征参数(基频、共振峰等)进行传输, 在经过完整的声码器编解码恢复原始语音后, 语音的基频能量及其谐波结构保持较好. 经过大量的样本实验发现, 上述理论分析表现为对于 0.1 ~ 2 kHz 频段内的信号能量和频谱结构, 在声码器处理前后差异不大.

根据这一原理及实验结果, 将待传输的秘密信息嵌入载体语音的 0.5 ~ 1 kHz 频带能量上(见图 3), 对 VAD 判决为有效语音的帧执行三级 DWT, 选择第 3 级的细节系数  $D^3(i)$  所表征的信号作为实际载体. 值得注意的是, 由于小波滤波器组的特性并不理想, 经过 DWT 处理后分解产生的细节成分和近似成分频谱都会伴生类似于谐波的虚假谱峰. 虚假谱峰的位置与语音信号以及所用的小波簇类型相关, 并且常常出现在 2 kHz 甚至 3 kHz 以上频带内. 若直接对上述的细节系数  $D^3(i)$  做能量调制, 进行信息隐藏, 将等价于对处于 0.5 ~ 1 kHz 频带的信号以及虚假谱峰做调制, 而 2 kHz 以上频带的信号能量在经过声码器处理后损失严重, 这直接导致对细节系数做能量调制的方法无法抵御声码器有损压缩攻击.

针对这一问题, 对细节系数  $D^3(i)$  做了进一步的 FFT 处理, 选择处于 0.5 ~ 1 kHz 内的频点计算频带能量, 并做 RDM, 能有效避免对虚假谱峰进行操作, 保证了对声码器有损压缩的鲁棒性.

RDM 作为量化索引调制的改进方法, 具有很好的抵抗固定增益攻击的特性<sup>[10]</sup>. 考虑到语音通话信道中存在如手机接听音量调节、线路增益等诸多类型的增益攻击对携密语音的影响, 采用了 RDM 法对 VAD-DWT-FFT 处理后的频带进行秘密信息嵌入操作. 具体执行 RDM 时根据人耳的听觉掩蔽效



应来自适应调整量化步长.

频带能量可表示为

$$E_k = \sum_{i=0}^{\frac{L}{N}} f_i(d) \quad (8)$$

其中:  $L$  为 VAD 判决时信号的帧长度,  $N$  为做 DWT 的子帧长度,  $f_i(d)$  为第  $i$  子帧 FFT 后第  $d$  个频点的系数.

对能量进行量化后, 最大的能量失真小于等于量化步长的一半, 利用 1.3 节所述掩蔽阈值与能量改变的关系可得量化步长为

$$\Delta_k = 2 \left( \frac{1}{L_{\text{buf}}} \sum_{i=1}^{L_{\text{buf}}} \hat{E}_{k-i} \right) \times 10^{\frac{a(z_{\text{rep}}) - \alpha}{10}} \quad (9)$$

其中:  $\Delta_k$  为当前帧的量化步长,  $L_{\text{buf}}$  为参与量化步长计算的预存帧数量,  $\hat{E}_{k-i}$  为第  $k-i$  帧能量的量化值,  $\alpha$  为表征秘密信息嵌入强度的系数. 随着语音载体的不同, 由掩蔽阈值得到的量化步长也是不同的, 根据人耳听觉掩蔽理论的量化步长设置能够确保在使用不同语音载体时, 携密语音都不会引起人耳听觉感知上的失真.

此时根据秘密信息  $b_k$  和量化步长  $\Delta_k$  对当前帧能量进行 RDM 后的结果为

$$\hat{E}_k = \begin{cases} \text{floor} \left( \frac{E_k}{\Delta_k} + 0.5 \right) \Delta_k, & b_k = 1 \\ \text{floor} \left( \frac{E_k}{\Delta_k} \right) \Delta_k + \frac{\Delta_k}{2}, & b_k = 0 \end{cases} \quad (10)$$

对应的 FFT 系数量化值为

$$\hat{f}_i(d) = \sqrt{\frac{\hat{E}_k}{E_k}} f_i(d), \quad i=0, 1, \dots, \frac{L}{N} \quad (11)$$

考虑到 FFT 系数在数值上具有共轭对称特性, 用  $\text{conj}()$  表示取共轭操作, 则在通过式 (11) 对  $d$  点系数进行量化后还需对  $d$  点的共轭对称点  $N+2-d$  做如下处理:

$$\hat{f}_i(N+2-d) = \text{conj}(\hat{f}_i(d)), \quad i=0, 1, \dots, \frac{L}{N} \quad (12)$$

完成 FFT 系数的量化处理后, 执行 FFT 逆变换 (IFFT, inverse FFT) 和 DWT 逆变换 (IDWT, inverse DWT), 得到携密语音信号. 在通信接收端对携密语音分析提取秘密信息的过程如图 3 所示, 执行与隐藏过程类似的 VAD-DWT-FFT 提取方案, 此时提取的秘密信息可表示为

$$b'_k = \begin{cases} 0, & \left| \frac{E'_k}{\Delta'_k} - \text{floor} \left( \frac{E'_k}{\Delta'_k} \right) - 0.5 \right| \leq 0.25 \\ 1, & \text{其他} \end{cases} \quad (13)$$

其中:  $E'_k$  为提取端当前帧指定频带的能量,  $\Delta'_k$  为提取端计算得到的当前帧的量化步长.

### 2.3 算法分析

提出的 VAD-DWT-FFT 模型重点解决在移动通信的语音信道中应用信息隐藏算法所遇到的高误码率挑战. 采用 VAD 算法选择浊音帧进行隐藏, 通过牺牲一定的嵌入容量来对抗语音信道中声码器的 DTX 技术处理. 在语音的 0.5 ~ 1 kHz 频带隐藏, 以此抵抗声码器有损压缩对秘密信息的干扰. 采用 RDM 作为嵌入方式来获得对信道中固定增益干扰的鲁棒性.

VAD-DWT-FFT 以一个 VAD 帧作为 1 bit 信息的隐藏单元, 在信道环境恶劣的情况下, 可以通过使用  $Y$  帧语音传递 1 bit 信息来进一步降低误码率. 此时嵌入容量表示为

$$C = \frac{\text{每秒浊音帧数}}{\text{每秒总语音帧数}} \frac{f_s}{L} \frac{1}{Y} \quad (14)$$

## 3 实验结果

为了对所提出的 VAD-DWT-FFT 信息隐藏提取模型做不可感知性、鲁棒性测试. 选择了 DWT-FFT-幅度替换法 (DWT-FFT-MR, DWT-FFT-magnitude replace)<sup>[5]</sup>、DWT-离散余弦变换 (DWT-DCT, DWT-discrete cosine transform)<sup>[11]</sup> 和改进的扩频-M 元双正交广义最大似然法 (ISS-MGML, improved spread spectrum-m-ary bi-orthogonal generalised maximum likelihood)<sup>[2]</sup> 进行对比评估实验. 选择 DWT-FFT-MR 做对比是因为 DWT-FFT-MR 方法和所提方法都是基于 DWT 域的隐藏方法, 区别在于 DWT-FFT-MR 方法对载体语音分帧后做一级 DWT 分解, 并对细节系数进行 FFT 分析, 通过直接替换 FFT 高频系数的幅度值来实现秘密信息的隐藏; 而所提方法则先对语音进行 VAD 判决, 挑选出浊音帧作为真正的载体语音, 对浊音帧做 3 级 DWT, 再根据声码器对语音的处理特性, 选择第 3 级的细节系数做 FFT, 最后根据听觉掩蔽理论将秘密信息通过 RDM 方式嵌入到表征 0.5 ~ 1 kHz 频带的 FFT 系数的能量上. DWT-DCT 是基于 DWT 域的经典音频水印算法.

ISS-MGML 是一种应用于保密通信的改进型扩频隐藏法,具有对抗自适应多速率声码器 12.2 kbit/s 有损压缩的良好鲁棒性,故也被选为实验的对比对象之一. 实验包括不可感知性分析与评价、抗信道攻击(GSM 声码器的 DTX 技术及有损压缩处理、固定增益攻击、加性高斯白噪声)的误码率性能评价. 实验环境及实验参数设定如下.

硬件环境: Intel(R) Core(TM) i3-4130 CPU@3.4 GHz 4.00 GB RAM, 软件环境: Microsoft Windows 7 旗舰版、Matlab R2010b, 载体语音选取 TIMIT 语料库中采样率为 16 kHz、6 bit 量化的 WAV 格式语音,对语音做降采样处理,使采样率转换为 $f_s = 8\text{ kHz}$ 后,组合成 8 段长度为 20 s 的载体语音,并划分为 4 组,每组 2 段. 秘密信息是二进制伪随机序列,声码器采用 GSM EFR 声码器,语音质量评估模型为 PESQ.

实验参数: VAD-DWT-FFT 模型中各参数选取为  $L = 160, N = 32, L_{\text{buf}} = 16, \alpha = 14$ ,小波簇为“db8”,取  $Y = 5$ ,得到容量  $\leq 10\text{ bit/s}$ ; DWT-FFT-MR 和 DWT-DCT 模型中选择容量  $= 10\text{ bit/s}$ ; ISS-MGML 模型中选择帧容量  $= 4\text{ bit/s}$ (考虑误码率因素,这里选择小于  $10\text{ bit/s}$  的容量,不影响后续的实验数据比较).

3.1 不可感知性实验

ITU-T P. 862.3 提供了 PESQ 语音评价方法进行语音客观质量的测评. 本实验将秘密信息分别经过 DWT-FFT-MR、DWT-DCT、ISS-MGML、VAD-DWT-FFT 4 种方法嵌入 8 段载体语音中生成携密语音. 对携密语音和原始载体语音做 PESQ 得到测试结果如表 1 所示. 表 1 中的 PESQ 测试数据表明,VAD-DWT-FFT 方法的 PESQ 分数平均为 3.487,与 DWT-FFT-MR 和 DWT-DCT 方法生成的携密语音质量相

表 1 PESQ 测试结果

语音编号		DWT-FFT-MR	DWT-DCT	ISS-MGML	VAD-DWT-FFT
1 组	F1. 1	3.462	3.771	3.333	3.452
	F1. 2	3.356	3.407	2.912	3.568
2 组	F2. 1	3.549	3.51	3.210	3.429
	F2. 2	3.485	3.455	3.207	3.511
3 组	F3. 1	3.492	3.548	2.699	3.326
	F3. 2	3.345	3.523	3.207	3.637
4 组	F4. 1	3.262	3.263	3.142	3.560
	F4. 2	3.421	3.659	3.203	3.412
平均		3.422	3.517	3.114	3.487

当,比 ISS-MGML 方法高 0.373,具有良好的不可感知性.

3.2 抗信道干扰实验

为评估所提出方法在信道中所存在的声码器有损压缩、增益攻击和高斯噪声干扰下的整体误码率性能,对 DWT-FFT-MR、DWT-DCT、ISS-MGML 和 VAD-DWT-FFT 方法做 3 类干扰测试. 第 I 类为信道中存在的固定增益和高斯白噪声干扰,设置参数为 3 dB 增益干扰 + 加性高斯白噪声干扰(信噪比为 30 dB). 第 II 类为 GSM EFR 声码器有损压缩 + DTX 技术干扰,目的在于分析 4 种方法对声码器有损压缩和 DTX 技术的鲁棒性. 第 III 类为 GSM EFR 声码器有损压缩 + DTX 技术干扰 + 3 dB 增益干扰 + 加性高斯白噪声干扰(信噪比为 30 dB). 第 III 类干扰包含第 I 类和第 II 类干扰,更接近于真实信道的环境.

从表 2 中可看出,4 种方法对于第 I 类干扰表现均较好;对于第 II 类干扰,VAD-DWT-FFT 方法和 ISS-MGML 方法在 GSM-EFR 声码器干扰下表现良

表 2 不同抗信道干扰类型的误码率测试结果

组	第 I 类				第 II 类				第 III 类			
	DWT-FFT-MR	DWT-DCT	ISS-MGML	VAD-DWT-FFT	DWT-FFT-MR	DWT-DCT	ISS-MGML	VAD-DWT-FFT	DWT-FFT-MR	DWT-DCT	ISS-MGML	VAD-DWT-FFT
1	0	0.022 2	0.045 5	0.014 7	0.207 1	0.459 6	0.108 0	0.051 8	0.289 1	0.458 6	0.108 0	0.070 3
2	0	0.044 9	0	0.013 5	0.225 5	0.485 1	0	0.025 0	0.315 6	0.468 4	0	0.075 0
3	0	0.053 3	0	0	0.170 7	0.427 5	0.038 5	0.051 3	0.307 7	0.417 3	0.038 5	0.035 7
4	0	0.034 1	0	0	0.166 3	0.485 0	0.029 4	0.070 7	0.292 1	0.480 4	0.029 4	0.067 0
平均	0	0.038 6	0	0.006	0.192 4	0.464 3	0.044 0	0.049 7	0.301 1	0.456 2	0.044 0	0.057 0

好,能保持低于 0.05 的误码率,而 DWT-FFT-MR 和 DWT-DCT 方法对声码器干扰的鲁棒性较差;在更复杂的第 III 类信道环境中,DWT-FFT-MR 和 DWT-DCT 方法因过高的误码率已经失去在移动通信语音电路域信道中进行信息隐藏的实用价值,ISS-MGML 方法保持着和第 II 类信道环境中一样的误码率,VAD-DWT-FFT 方法的性能略有下降,但仍维持在 0.06 左右,与 ISS-MGML 仅有 0.013 的差距。

不可感知性实验和抗信道干扰实验结果表明,在 VAD-DWT-FFT 方法的 PESQ 分数与 DWT-FFT-MR 相当,且优于 ISS-GMGL 方法的情况下,VAD-DWT-FFT 方法相比传统的变换域方法(DWT-FFT-MR)在抗声码器干扰(有损压缩及 DTX 技术处理)性能上有明显提升,在复杂信道环境下(第 III 类),误码率性能和 ISS-MGML 方法相当,具有较高实用价值。

为进一步分析 VAD-DWT-FFT 方法中嵌入容量与抗信道干扰能力的关系,选择的信道干扰类型为第 III 类,携密语音 PESQ 分数平均值在 3.487 左右,得到不同嵌入容量条件下误码率性能变化曲线,如图 4 所示。通过图 4 中信息提取误码率与嵌入容量的变化曲线可以看出,在嵌入容量为 10 bit/s 时,误码率在 0.06 左右;随着嵌入容量的减小,误码率也有所下降;在嵌入容量减小至 2 bit/s 时,信息提取的误码为 0,整个隐蔽通信过程能保证信息无失真传输。

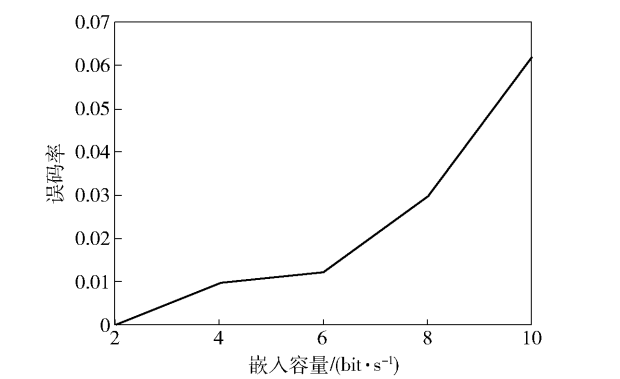


图 4 VAD-DWT-FFT 方法中误码率与嵌入容量的关系

3.3 计算复杂度评价

计算复杂度包括时间复杂度和空间复杂度,在实时的语音信息隐藏应用场景中,时间复杂度是算法实用性评价的重要指标。采用嵌入时间(EMT, embedding time)和提取时间(EXT, extracting time)2

个指标进行时间复杂度评价<sup>[12]</sup>。由抗信道干扰实验可知,DWT-FFT-MR 方法不适用于笔者所研究的应用场景。因此在表 1 的实验环境下,仅对 VAD-DWT-FFT 和 ISS-MGML 方法的计算复杂度进行测试并比较,得到结果如表 3 所示。

表 3 不同方法下时间复杂度对比

方法	EMT/s	EXT/s	EMT + EXT/s
ISS-MGML	50.574	0.073	50.647
VAD-DWT-FFT	9.845	9.671	19.516

从表 3 中可以发现,对于 8 段 20 s 时长的测试语音,VAD-DWT-FFT 方法在引入 VAD 算法后的 EMT 和 EXT 相近,而 ISS-MGML 方法由于在嵌入时存在长序列的自相关矩阵计算,EMT 远大于 EXT。在总时间上,VAD-DWT-FFT 的 19.516 s 远小于 ISS-MGML 方法的 50.647 s,显然所提出的方法具有比 ISS-MGML 方法更低的时间复杂度。

4 结束语

首先分析了移动通信电路域语音信道的结构特点,并重点研究其对语音的处理特性,为基于该信道进行语音信息隐藏的研究提供了误码率性能优化的理论基础;其次为兼容上述信道中语音声码器的 DTX 技术,并抵抗声码器对语音的有损压缩,提出一种基于 VAD-DWT-FFT 框架的新型信息隐藏和提取模型。实验结果表明,该算法模型在 10 bit/s 上限容量时,携密语音质量 PESQ 分数高于 3.48,经过 GSM EFR 声码器的 DTX 技术、有损压缩以及信道中存在的其他攻击后,能保持误码率在 0.06 左右,算法的时间复杂度优于 ISS-MGML 方法。

下一步的研究工作是改进算法模型,降低误码率,提高嵌入容量,使算法在保持低误码率的同时还能适用于高嵌入容量的场景。

参考文献:

[1] 白剑,曹鹏. 语音信息隐藏容量研究[J]. 北京邮电大学学报, 2016, 39(S1): 76-80.  
Bai Jian, Cao Peng. Capacity estimation research for speech information hiding[J]. Journal of Beijing University of Posts and Telecommunications, 2016, 39(S1): 76-80.  
[2] Kazemi R, Rezaei R, Akhaee M A, et al. Covert communications through mobile voice channels [J]. IET Information Security, 2016, 10(3): 156-164.

- [3] Yan Shu-fan, Tang Guang-ming, Chen Yan-ling. Incorporating data hiding into G. 729 speech codec[J]. *Multimedia Tools & Applications*, 2016, 75(18): 11493-11512.
- [4] Ahani S, Ghaemmaghami S, Z. Jane Wang. A sparse representation-based wavelet domain speech steganography method [J]. *IEEE/ACM Transactions on Audio Speech & Language Processing*, 2015, 23(1): 80-91.
- [5] Rekik S, Guerchi D, Selouani S A, et al. Speech steganography using wavelet and Fourier transforms[J]. *Eurasip Journal on Audio Speech & Music Processing*, 2012(1): 20.
- [6] Paulin C, Selouani S, Hervet É. Audio steganalysis using deep belief networks[J]. *International Journal of Speech Technology*, 2016, 19(3): 1-7.
- [7] Ghasemzadeh H, Khass M T, Arjmandi M K. Audio steganalysis based on reversed psychoacoustic model of human hearing [J]. *Digital Signal Processing*, 2016(51): 133-141.
- [8] He Xing, Scordilis M S. An enhanced psychoacoustic model based on the discrete wavelet packet transform[J]. *Journal of the Franklin Institute*, 2006, 343(7): 738-755.
- [9] Hu Hwai Tsu, Hsu L Y. Robust, transparent and high-capacity audio watermarking in DCT domain[J]. *Signal Processing*, 2015(109): 226-235.
- [10] Hu Hwai Tsu, Hsu L Y. A DWT-based rational dither modulation scheme for effective blind audio watermarking[J]. *Circuits, Systems, and Signal Processing*, 2016, 35(2): 553-572.
- [11] Wang Xiang-yang, Zhao Hong. A novel synchronization invariant audio watermarking scheme based on DWT and DCT [J]. *IEEE Transaction on Signal Processing*, 2006, 54(12): 4835-4840.
- [12] Wang Xinkai, Wang Pengjun, Zhang Peng, et al. A blind audio watermarking algorithm by logarithmic quantization index modulation [J]. *Multimedia Tools and Applications*, 2014, 71(3): 1157-1177.