基于 DCT 变换的端到端语音加密算法

陈瑶瑶 郝建华 张子博
(装备学院 北京 101416)

摘 要：为了实现语音的端到端安全加密通信，提出了一种新的可通过 RPE-LTP 压缩编码的语音加解密算法。该算法基于 DCT 变换域，在发送端对语音信号先后进行频域置乱和时域置乱，接收端进行解置乱和重建信号。运算过程限制在实数域，不会涉及相位的估计问题，避免了语音数据的损失。仿真结果表明，加密算法实现了加密的语音信号非噪声化，保持了语音信号的特性，满足通过声码器进行编码的信号输入要求。进一步降低了加密后的语音可懂度，提高了解密后的语音质量和信号重建率。

关键词：规则脉冲激励长时预测；端到端；语音加密；频域置乱；时域置乱
中图分类号：TN918 **文献标识码**：A **国家标准学科分类代码**：510.4040

End-to-end speech encryption algorithm based on DCT transform

Chen Yaoyao Hao Jianhua Zhang Zibo
(Equipment Academy, Beijing 101416, China)

Abstract: For the realization of the end to end secure encrypted communication of voice, this paper proposes a new voice encryption algorithm, which can penetrate RPE-LTP compression encoder. In the sender, the algorithm scrambles with the voice signal in frequency and time domain based on DCT transform domain . In the receiving end, it decrypts and reconstructs the signal. Opreation process is limited in real number domain and does not involve the phase estimation problem, it avoids the loss of voice data. The results show that the encryption algorithm can realize the non-noise signal of the encrypted voice, maintains the characteristics of the speech signal, meets the requirements of input signal to penetrate RPE-LTP compression encoder. Reduce the encrypted speech intelligibility further, improve the quality of the decrypted voice.

Keywords: RPE-LTP; end-to-end; voice encryption ; frequency domain scramble; time domain scramble

1 引 言

在我国，手机已经被广泛使用，而随之而来的窃听事件却从未停止。一些涉及公务、商务活动甚至国家秘密的通信内容一直受到窃听的威胁。美国于 2007 年已经开始实施代号为“棱镜”的绝密电子监听计划，监听内容和覆盖面之广令世界哗然。2014 年 12 月 4 日，有消息再次披露，至少从 2010 年开始，美国启动了代号为“极光黄金”秘密监视计划。通过监视全球手机运营商，发现手机网络漏洞，并秘密植入新的安全漏洞，来窃取手机通话、电子邮件及短信等信息。通信安全面临严峻挑战。

GSM 移动通信系统作为目前应用效果最好、应用范围最广的语音承载方式，它的系统安全机制一直存在隐患[1]。它提供一种仅在无线链路上进行加密的空中接口加密[2]，信息经过基站后进行的是无线透明传输，无法实现端到端的安全通信。

国外已经有基于 GSM 通信系统的能够提供端到端加密[3]通信的产品问世[4]，不过大多是利用数据通道实现，会产生系统延时性和网络互用性等问题，而语音通道在这方面具有优势。

GSM 系统采用的特定的编码器对语音信号进行编解码。语音经过普通加密算法之后成为白噪声，无法透过编码器，从而无法还原语音信号。本文基于 GSM 语音通道，设计了一种抗 RPE-LTP 压缩编码的语音加密算法，加密后的语音信号仍然具有类语音的特性，可以顺利通过声码器进行有效地编解码；同时，时域置乱和基于离散余弦变换(discrete cosine transform，DCT)域的频域置乱提高了加密强度，加密语音可懂度进一步降低，解密后的语音质量得到提高。

收稿日期：2014-12

2　GSM 语音通信特点

2.1　GSM 语音通信过程

在 GSM 通信系统中,以电路域为主要承载网络的移动话音业务的工作流程如图 1 所示。通信过程采用二次编解码工作模式,历经 4 个声码器,进行了四次编解码转换。其中,话音编码主要由规则脉冲激励长时预测[5](regular pulse excitation-long term prediction, RPE-LTP)编码组成。

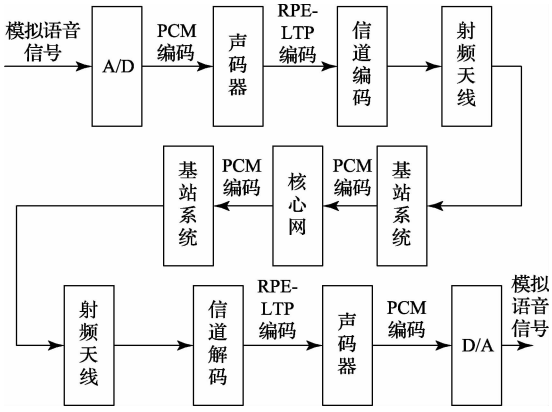

图 1　GSM 语音通信过程

2.2　RPE-LTP 语音编码特点

RPE-LTP 是一种混合参数压缩编码算法,它利用脉冲源和滤波器来代替发声器官的主要特征参数,通过提取输入语音的声音模型参数进行压缩,将这些特征参数进行数字编码,形成数字信号加以传送。在接收端通过这些模型参数恢复原有的语音[6]。这种有损的压缩编码是根据语音的特征提取参数进行传输的,一旦信号不具备声音特征,也将无法进行正确编码或者数据大量丢失,接收端无法正确恢复。所以,端到端加密生成的加密语音信号必须满足类语音的特性,过程如图 2 所示。

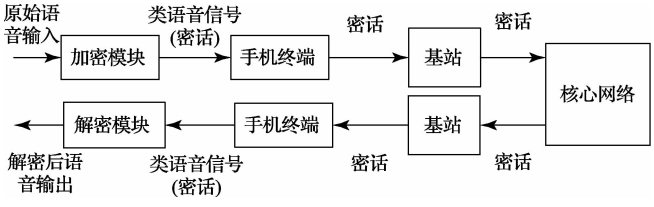

图 2　端到端加密流程

3　语音加密算法设计

3.1　算法介绍

如图 3 所示,端到端的语音加密算法思想主要集中在类语音的调制[7]生成上,这种算法运算量大,容易产生较大延时,不够实用。而对语音信号进行快速傅里叶变换之后进行频域置乱,得到的置乱后序列必然是一簇乱序的复数。而对复数序列进行逆傅里叶变换,使得得到的时域信号为实序列形式是比较困难的[8]。DCT 变换具有信号谱分量丰富、能量集中,且不需要对语音相位进行估算的优点,能在较低的运算复杂度下取得较好的语音增强效果,恰好避免了复数运算。所以,可以采用离散余弦变换替代快速傅里叶变换,结合信号的置乱方法设计了该算法。

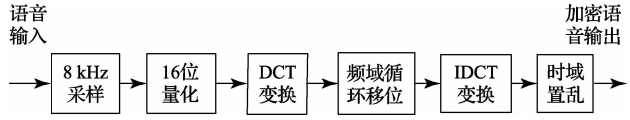

图 3　加密算法流程

3.2　算法基本原理

端到端的加密模块是实现模拟语音信号经过加密处理再到模拟类语音信号的一个转换过程。语音信号时域形式如图 4 所示,代表的语音含义是“尤其是战后的世界秩序的构想上分歧过大互不相让矛盾”,共 4s 的信息长度。

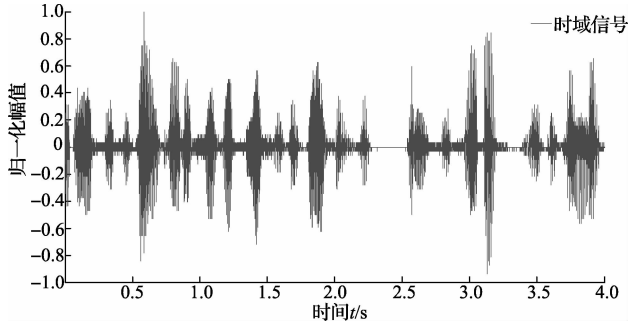

图 4　语音信号时域

实验以帧为单位进行加密操作,一帧长度的选择,也将会影响加密效果和传输延时。本实验以 32 ms 语音信号为一帧单元,将模拟信号进行 8 kHz 采样,16 位量化,得到一帧 256 个采样点,选择 16 帧为一组信号进行仿真实验。

首先对第一帧信号 S_1 进行 DCT 变换,得到 DS_1:

$$DS_1 = \mathrm{DCT}(S_1) \tag{1}$$

将新产生的 DCT 域频谱序列 DS_1 进行圆周移位,而圆周移位的位数则由密码组决定。该密码组是一串整数长序列,不同的秘钥值将决定长序列的不同起始位置,同时决定了每一帧数据的不同循环移位数。这里 DS_1 所对应的循环移位数设为 C_1,循环移位后得到的频域序列为 EDS_1;然后将加密频域序列 EDS_1 进行 DCT 逆变换,得到加密时域序列 ES_1:

$$ES_1 = \mathrm{IDCT}(EDS_1) \tag{2}$$

这就完成了第一帧的加密操作,加密前后频域效果如图 5 所示。图中也显示出 DCT 变换的能量集中特性。

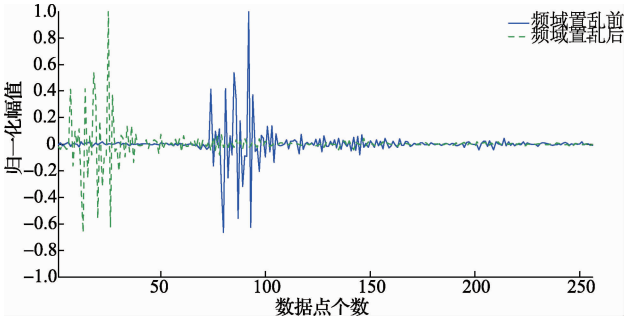

图 5　一帧长度的语音频域置乱前后比较

紧接着的语音帧序列按照此步骤进行加密，完成一组语音信号的保密操作。

将频域置乱后的时域信号 ES_1 进行时域置乱。时域置乱是将 16 帧语音信号，以帧为单位进行时序的调整，另一密码组控制着时域的置乱次序。置乱过程如图 6 所示。

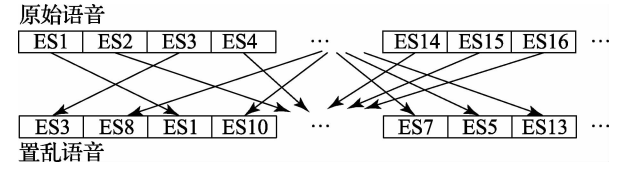

图 6　时域置乱过程

4　仿真结果分析

4.1　仿真结果

实验取图 5 中前 512 ms 语音长度，进行频域置乱，并对两种方法所产生的效果进行了类比。频域置乱前后的语音时域形式如图 7 和 8 所示。

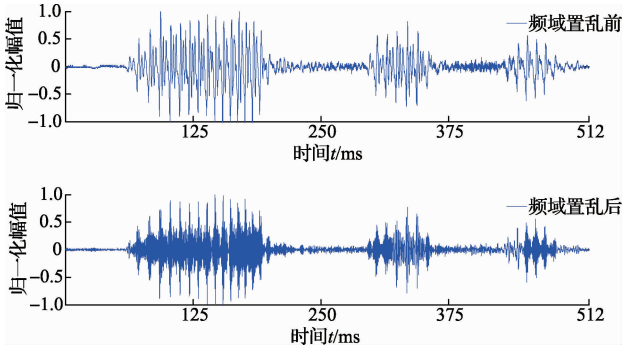

图 7　基于 DCT 时域置乱前后的时域比较

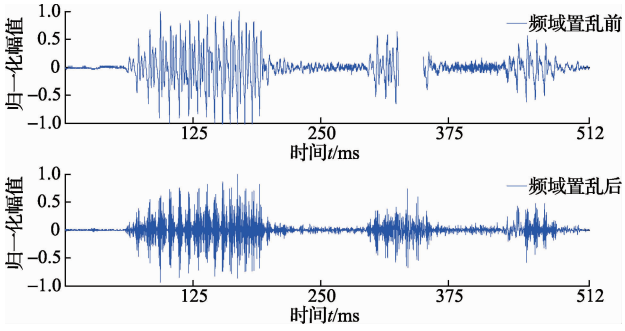

图 8　基于 FFT 时域置乱前后的时域比较

两种方法频域置乱后的语音信号仍旧保持着类语音信号特征。紧接着将频域加密后的语音信号进行时域的置乱，得到时域置乱前后的语音时域对比如图 9 和 10 所示。

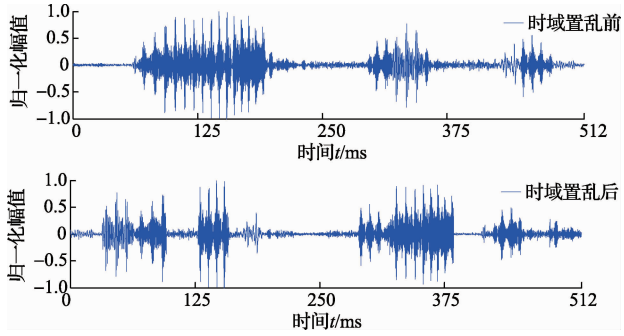

图 9　基于 DCT 时域置乱前后的时域比较

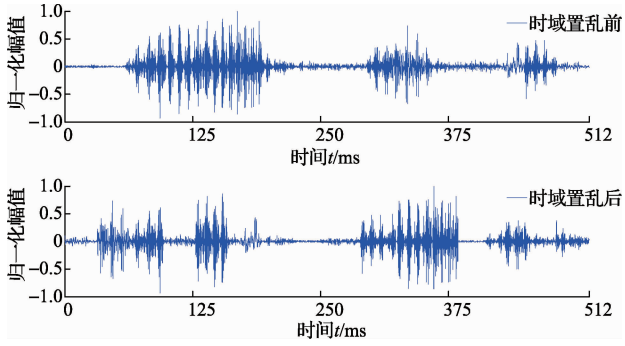

图 10　基于 FFT 时域置乱前后的时域比较

由图中可以明显看到，加密后的信号仍然具有类语音的特征，并非完全随机无规律的伪噪声状态。而且，经过频域和时域的双重加密，语音的时域波形和原始语音有了比较大的区别。

4.2　语音质量评价

4.2.1　主观评价

对加密后的语音信号进行语音可懂度评估，这是一种主观评价方法[9]。评价指标包括频域置乱后的语音可懂度和时域置乱后的语音可懂度。实验请了 24 位评听人员（12 位男士，12 位女士）。在未听过原始语音的情况下，听完频域置乱后的语音后，将听懂的字词写下，与原始字词进行比较，计算可懂度。同样，对于时域置乱后的语音按相同的过程进行评测。测得结果与原始语音进行比较，含义可懂度结果如表 1 所示。

表 1　语音可懂度的比较

不同方法 效果比较	原始语音	频域置乱 后语音	时域置乱 后语音
基于 FFT 方法	100%	20.8%	0
基于 DCT 方法	100%	7.8%	0

评测中，对于频域置乱后的语音，一些实验者能够听对

部分字的音,但是无法理解含义。字音识别正确率结果如表 2 所示。

表 2　字音辨识正确率的比较

不同方法效果比较	原始语音	频域置乱后语音	时域置乱后语音
基于 FFT 方法	100%	83.3%	12.5%
基于 DCT 方法	100%	58.3%	0

4.2.2　客观评价

对加密语音进行时域失真度的客观评价[10]。评价原理是计算一段时间内的噪声与语音的平均功率之比。基于 FFT 的频域置乱算法,会涉及复数域的运算,对于实语音信号来讲,复数部分的数据需要丢掉,这就会导致数据的损失,这在时域中会体现出来。在未加入任何噪声的情况下,两种方法的时域失真情况比较如图 11～13 和表 3 所示。

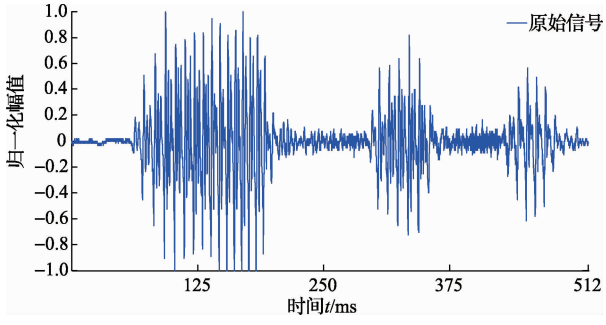

图 11　时域置乱前后的时域比较

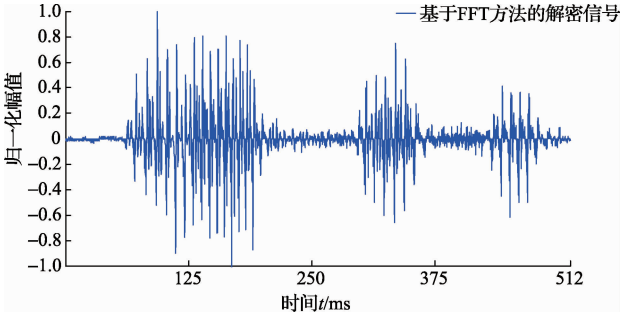

图 12　时域置乱前后的时域比较

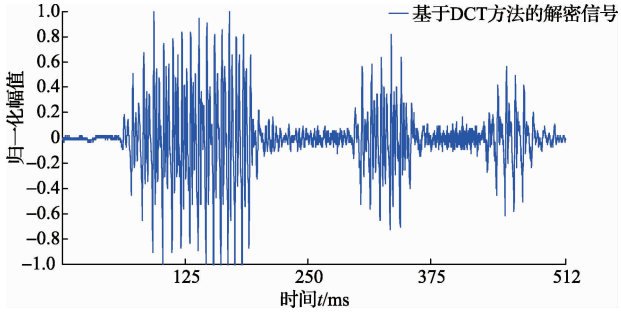

图 13　时域置乱前后的时域比较

表 3　时域失真度的比较

失真度比较	原始语音	基于 FFT 方法	基于 DCT 方法
域失真度	0.000 0	10.019 8	0.875 5

由于基于 DCT 的置乱算法没有理论上的数据丢失,所以可以完全还原原始信号。此外,从最后的解密效果来看,基于 FFT 的置乱方法由于丢失了一部分数据,噪声比较大,听起来极不舒服。

5　结　　论

本文提出了一种基于 DCT 变换的频域和时域的联合置乱方法,加密语音保持有类语音的特性,可以透过压缩编码器后进行通信传输。加密后的语音可懂度几乎为零,安全性得到了提高,解密语音质量较好。该算法计算量相对较小,实际应用中,不会产生较大的时延,实用性比较强。下一步将会在硬件上进行检验,实现端到端通信的整个过程,并应用于手机终端。

参考文献

[1]　王经星,孔维祥,高一,等. 面向移动通信网络的语音安全传输系统[J]. 信息网络安全,2013(2):47-52.

[2]　江斌,陈义和,徐建良. 移动通信系统空中接口安全及防护研究[J]. 微波学报(增刊),2010:735-737.

[3]　杨于村. 基于公众移动通信网的端到端加密语音传输技术研究[D]. 广州:华南理工大学,2009.

[4]　朱卫东,周长林,杨旭,等. 基于 DSP5416 的语音采集与保密通信的实现[J]. 电子测量技术,2012,35(4):37-41.

[5]　江晓文. 基于 DSP 的 RPE-LTP 语音压缩算法的研究与改进[D]. 长沙:中南大学,2012.

[6]　王红军,钟子发,陈润洁. GSM 数字移动通信系统语音信源编解码技术[J]. 电讯技术,2004(1):25-29.

[7]　铁启龙. 基于正交频分复用的 GSM 加密语音通信[J]. 信息通信,2014(3):193-194.

[8]　孙源,罗挺. 基于 FFT 算法的语音加密技术过程实现研究[J]. 中国储运,2011(3):97-98.

[9]　沈晓东. 语音增强技术研究[D]. 北京:清华大学,2011.

[10]　方凡泉,李心广,王桂珍,等. 语音质量客观评价研究与实现[J]. 广州大学学报:自然科学版,2011(1):65-69.

作者简介

陈瑶瑶,在读硕士研究生。主要研究方向为信息与通信系统。

郝建华,副教授。主要研究方向为信息与通信系统。

张子博,在读硕士研究生。主要研究方向为信息与通信系统。