基于分数基音延迟动态搜索的语音隐写算法

田 晖[1,2,3],严 艳[1,2,3],汤莉莉[2,3,4],吴俊彦[1,2,3],王慧东[1,2,3],全韩彧[1,2,3]
(1. 华侨大学计算机科学与技术学院,福建厦门 361021;2. 厦门市数据安全与区块链技术重点实验室,福建厦门 361021;
3. 福建省大数据智能与安全重点实验室,福建厦门 361021;4. 华侨大学机电及自动化学院,福建厦门 361021)

摘 要: 论文提出了一种基于分数基音延迟动态搜索的语音隐写算法. 该算法可根据隐藏容量(x比特/子帧)的需要将分数基音延迟候选值集合划分为2^x个子集,每个子集代表不同的x比特信息. 在闭环基音搜索过程中,可为每个子帧选择既能表示待嵌入隐秘信息且内插后的归一化相关系数最大的分数基音延迟候选值,从而有效降低隐写操作对于原始载体的影响. 以目前IP语音系统中广泛使用的自适应多速率语音编码为例,对该算法从隐藏容量、不可感知性及抗检测性三方面进行了性能评估并与相关工作进行了对比分析. 实验结果表明,本文提出的隐写算法较之现有基于基音延迟的隐写算法可在确保较高隐写容量的同时达到更好隐写安全性(即更好抗检测能力和不可感知性).

关键词: 语音隐写;动态搜索;分数基音延迟;自适应多速率语音编码;隐写安全性
基金项目: 国家自然科学基金(No.61972168)
中图分类号: TP309 **文献标识码:** A **文章编号:** 0372-2112(2023)01-0067-09
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20211473

Speech Steganography Based on Dynamic Search of Fractional Pitch Delay

TIAN Hui[1,2,3], YAN Yan[1,2,3], TANG Li-li[2,3,4], WU Jun-yan[1,2,3], WANG Hui-dong[1,2,3], QUAN Han-yu[1,2,3]
(1. *College of Computer Science and Technology*, *Huaqiao University*, *Xiamen*, *Fujian* 361021, *China*;
2. *Xiamen Key Laboratory of Data Security and Blockchain Technology*, *Xiamen*, *Fujian* 361021, *China*;
3. *Fujian Key Laboratory of Big Data Intelligence and Security*, *Xiamen*, *Fujian* 361021, *China*;
4. *College of Mechatronics and Automation*, *Huaqiao University*, *Xiamen*, *Fujian* 361021, *China*)

Abstract: In this paper, we present a speech steganography algorithm based on dynamic search of fractional pitch delay. The algorithm can divide the candidate value set of fractional pitch delay into 2^x subsets according to the needs of the covert capacity (x bits/subframe), where each subset represents different x bits of information. In the closed-loop pitch search process, the algorithm can select for each subframe the best candidate value of pitch delay that can not only denote the secret information but also make the interpolated normalized correlation coefficient largest. In this way, the impact of steganographic operations on the original carriers can be effectively reduced. Taking adaptive multi-rate speech codec widely used in the current Voice-over-IP systems as an example, the performance of presented algorithm has been evaluated from the aspects of covert capacity, imperceptibility and anti-detection, and compared with related works. Experimental results show that the proposed steganographic algorithm can achieve better steganography security (better resistance to detection and imperceptibility) than the existing steganographic methods based on pitch delay, while maintaining relatively high steganographic capacity.

Key words: speech steganography; dynamic search; fractional pitch delay; adaptive multi-rate speech codec; steganographic security
Foundation Item(s): National Natural Science Foundation of China (No.61972168)

1 引言

隐写术是利用人类的视听等器官的不敏感性,在数字媒介的冗余信息部分嵌入隐秘信息以实现信息隐藏的一种安全技术[1~3]. 相较于传统加密技术,隐写术不仅保护了隐秘消息的内容,而且掩盖了其存在性,从而大大提高了信息传输和存储的安全性[3~5]. 近年来,

收稿日期:2021-11-01;修回日期:2022-09-02;责任编辑:李勇锋

隐写术已从最初以图像为载体发展至几乎所有的多媒体[6~9]. 随着IP语音(Voice over Internet Protocol, VoIP)技术的快速发展和日益普及,以VoIP为载体的隐写成为了信息隐藏领域的一个新兴分支[2,3,6,10].

为了实现语音信号的快速传输,VoIP通常采用语音压缩编码对语音信号进行压缩. 代数码本激励线性预测(Algebraic Code-Excited Linear Prediction, ACELP)是目前大多数低速率语音压缩编码标准(如G.723.1[11],G.729[12]和Adaptive Multi-Rate[13]等)所采用的典型编码算法. 它主要通过线性预测分析、自适应码本搜索和固定码本搜索三部分提取模型参数[13],其中,线性预测分析的主要功能是获得线谱对频率的索引;自适应码本搜索通过基音分析获得基音延迟索引和基音增益索引;固定码本搜索则是为了获得固定码本索引和固定码本增益.

相应地,以ACELP编码算法为基础,现有的语音隐写算法主要围绕三类典型的特征参数进行展开,即线性预测系数(Linear Prediction Coefficient, LPC)[10,14,15]、固定码本(Fixed CodeBook, FCB)[16~18]和基音延迟(Pitch Delay, PD)[19~24]. 在ACELP编码中,准确可靠地预测并提取基音周期对语音的分析合成至关重要,直接影响到合成语音是否能够真实再现原始语音信号. 然而,由于基音周期的变化幅度较大,并且受到说话人的声带特性和背景噪声等因素的影响,因而基音周期预测(也称为基音检测)很难取得非常精准的结果[20,22,23]. 从这个意义上来说,基于自适应码本搜索得到基音延迟参数其实是对于基音周期的近似预测,其误差难以避免. 正是基音周期的这种"测不准"的特性为信息隐藏提供了可能性——对基音延迟进行适当的改动不会对语音质量构成较大影响. 因此,基音延迟被视作语音帧上的主要隐写域之一,并涌现了许多有效的隐写算法.

代表性的工作如,余迟等[19]提出了一种基于Adaptive Multi-Rate Wideband(AMR-WB)语音编码的整数基音延迟隐写算法,它预先筛选出与前后相邻子帧不同的整数基音延迟对应的子帧,然后在筛选出的子帧中采用模除隐藏法选择能够表示秘密信息的候选整数基音延迟构成搜索集合,并从中搜索最大的候选整数基音延迟来实现隐写;Huang等[20]提出了一种基于G.723.1语音编码的整数基音延迟参数隐写算法,该算法首先根据奇偶性划分整数基音延迟搜索范围,并通过在闭环基音搜索过程中根据隐秘信息的奇偶性从对应集合中搜索最优整数基音延迟来实现隐写;严书凡等[21]提出了一种基于G.723.1语音编码的整数基音延迟双层隐写算法,该算法先采用和Huang等人类似的算法在第二和第四子帧的整数基音延迟实现第一层隐写,并进一步利用搜索集合内整数基音延迟取值的任意性限制第四子帧的基音搜索范围实现第二层隐写,有效降低了隐写失真,从而取得了比Huang等人算法更好的不可感知性. 刘程浩等[22]提出了一种基于G.729a语音编码的基音延迟隐写算法,该算法在整数基音延迟上采用和Huang等类似的嵌入算法,而在分数基音延迟上将搜索范围划分为零和非零的搜索区间以分别代表隐秘信息中的"0"和"1",并通过在对应区间搜索最优分数基音延迟来实现隐写. 吴志军等[23]提出了一种基于G.723.1语音编码的整数基音延迟参数隐写算法,该算法利用整数基音延迟的奇偶性表示隐秘信息,并结合矩阵编码在第一、第三和第四子帧的整数基音延迟参数上嵌入隐秘信息,可实现两比特隐秘信息嵌入而仅改动最多一个参数.

作为对抗技术,基于基音延迟的语音隐写分析(隐写检测)近年来也受到了广泛关注. 如Ren等人[25]提出了基于整数基音延迟二阶差分特征的隐写分析算法;Tian等[26]结合整数基音延迟的奇偶特性和二阶差分特性设计了基于混合统计特征的隐写分析算法;Liu等[27]分析了隐写前后整数基音延迟的奇偶性变化,提出了基于奇偶贝叶斯概率的隐写分析算法等. 这些隐写分析算法均能有效检测现有的基于整数基音延迟的隐写算法. 因此,如何提高基于基音延迟隐写算法的抗检测性能是一个亟待解决的关键问题.

事实上,基音延迟不仅包括整数基音延迟还包括分数基音延迟,且分数基音延迟在整个基音延迟中所占的比例要小得多. 例如,某个基音延迟值为$94\frac{1}{6}$,其中整数基音延迟是94,分数基音延迟是$\frac{1}{6}$,而后者在整个基音延迟中所占比例不到0.18%,因此,对分数基音延迟的适当修改不会给基音周期预测造成显著的影响. 鉴于此,Liu等[24]提出了一种基于分数基音延迟的隐写算法,它通过分数基音延迟最低有效位的替换来实现信息隐藏. 同时该算法还引入了部分相似度来衡量隐秘信息和分数基音延迟的相似性,只在相似度达到特定阈值的分数基音延迟上嵌入信息以减少替换操作带来的失真. 此外,考虑到编码过程中分数基音延迟的改变会引起少部分整数基音延迟的变化,该算法采用了整数基音延迟覆盖策略,通过对语音进行两次编码操作,第一次编码用于保存整数基音延迟,第二次编码实现分数基音延迟的替换隐写,并且用第一次编码保存的整数基音延迟覆盖第二次编码得到的整数基音延迟,保证了隐写前后整数基音延迟相同,从而能够有效对抗现有隐写分析算法[25~27]. 然而,该算法仍然存在如下不足:(1)该算法本质上是较为简单直接的最低有效位替换算法,将不可避免地造成语音质量的失真;

(2)该算法需要额外的同步标志位用于指示是否隐藏信息,且需要安全的信道来传递这些标志信息;(3)该算法忽视了有些基音延迟不存在分数部分的事实(例如,在AMR12.2 kb/s编码速率模式下,当第一或者第三子帧的基音延迟范围为[95,143]时,其基音延迟只有整数部分),在嵌入信息后会使得这些基音延迟出现"异常"(即出现了不该有的分数部分),从而降低了算法的隐写安全性. 有鉴于此,本文提出了一种新的基于分数基音延迟动态搜索的隐写算法. 该算法可根据隐藏容量的需要(比如x比特/子帧)将分数基音延迟候选值集合划分为2^x个子集,每个子集代表不同的x比特信息;随后在闭环基音搜索过程中,为每个子帧选择既能表示待嵌入隐秘信息且内插后的归一化相关系数最大的分数基音延迟候选值作为搜索结果. 换言之,该算法可在保证隐秘信息有效嵌入的前提下使得合成语音的加权均方误差达到最小,从而有效地降低了隐写过程对语音质量的影响. 相较于Liu等的算法[24],该算法无需额外的同步信道,且不会出现"异常"的基音延迟,从而具有更好的安全性. 我们以目前VoIP系统中广泛使用的AMR语音编码为例,对该算法从隐藏容量、不可感知性及抗检测性三方面进行了性能评估并与相关工作进行了对比分析. 实验结果表明,本文提出的隐写算法不仅能够有效地对抗现有隐写分析算法,且较之现有基于基音延迟的隐写算法在确保较高隐写容量的同时可达到更好的不可感知性(感知透明性).

2　基于分数基音延迟动态搜索的隐写

为了尽可能减少隐写操作对载体语音质量的影响,本文提出了一种基于分数基音延迟动态搜索的隐写算法. 基于该算法的隐蔽通信过程可描述为:通信双方根据共享的密钥将分数基音延迟候选值集合划分成若干个子集,每个子集表示不同的隐秘信息;发送方在语音编码的自适应基音搜索过程中,通过选择能够表示隐秘信息的最优分数基音延迟候选值作为搜索结果来实现嵌入隐秘信息,并且将压缩编码后的数据通过VoIP信道发送给接收方;接收方对收到的VoIP数据包进行解析并从语音流中提取出基音延迟参数,并根据事先确定的分数基音延迟候选值集合划分结果从中提取出隐秘信息.

假设发送方经过加密处理后的隐秘信息为$M=\{m_i=0\text{ or }1|i=1,2,\cdots,L_M\}$,其中$m_i$指第$i$比特隐秘信息;$L_M$为隐秘信息的长度;用于隐藏信息的语音子帧集合为$F=\{f_j|j=1,2,\cdots,L_F\}$,其中$f_j$指第$j$子帧的语音;$L_F$指载体的子帧数,语音编码后产生的分数基音延迟参数集合为$D=\{d_j|j=1,2,\cdots,L_F\}$,$d_j$指第$j$子帧的分数基音延迟,$d_j\in U$,$U$为分数基音延迟候选值集合;整数基音延迟参数集合为$H=\{h_j|j=1,2,\cdots,L_F\}$,$h_j$指第$j$子帧的整数基音延迟,$h_j\in T$,$T$为整数基音延迟候选值集合. 在第一或第三子帧中,部分范围内的基音延迟只有整数部分(分数基音延迟恒为0),将该范围内的整数基音延迟集合记为T'. 在进行隐蔽通信之前,通信双方首先协商一组密钥K,并根据实际隐藏容量的要求(记为x比特/子帧),在密钥K的控制下将分数基音延迟候选值集合U平均划分为2^x个互不相交的子集(若U_a和U_b是U的任意两个不同子集($a,b\in\{0,1,\cdots,2^x-1\}$),则$U_a\cap U_b=\varnothing$,且$U_0\cup U_1\cup\cdots\cup U_{2^x-1}=U$),并使得每个子集表示一个唯一的$x$比特二进制数. 不难看出,上述划分方式使得每个子帧(分数基音延迟)可以隐藏x比特隐秘信息,且x的取值范围为$\{1,2,\cdots,X\}$,其中x的最大值X由下式决定,

$$X=\lfloor\log_2 n\rfloor \tag{1}$$

式中,n指分数基音延迟候选值的数量,$\lfloor e\rfloor$表示对e向下取整. 例如,当期望隐藏容量为1比特/子帧(x=1)时,集合U将被平均划分成U_0和U_1两个子集,它们可分别表示隐秘信息"0"和"1";当x=2时,集合U将被平均划分成U_0,U_1,U_2和U_3四个子集,它们可分别表示隐秘信息"00","01","10"和"11". 对于给定的隐藏容量x比特/子帧,隐秘信息M可分组表示为$M=\{\mathfrak{M}_1,\mathfrak{M}_2,\cdots,\mathfrak{M}_r\}$,其中$\mathfrak{M}_k=\{m_{(k-1)x+1},m_{(k-1)x+2},\cdots,m_{kx}\}$,$k=\{1,2,\cdots,r\}$,$r=\lceil L_M/x\rceil$(为便于描述,此处假设$L_M$能够被$x$整除).

2.1　嵌入算法

发送方逐子帧地嵌入隐秘信息,其算法流程包括如下步骤:

Step1:初始化子帧序号$j=1$,隐秘信息分组序号k=1.

Step2:判断子帧f_j能否用于信息隐藏. 首先对于f_j搜索其整数基音延迟h_j. 若f_j是某个语音帧中的第一或第三子帧(即$j \bmod 4=1$ or 3,mod表示求余运算),且其整数基音延迟$h_j\in T'$,则f_j不能用于信息隐藏,即$d_j=0$,执行Step6;否则,f_j能够用于嵌入隐秘信息,执行Step3.

Step3:根据待嵌入的隐秘信息分组$\mathfrak{M}_k=\{m_{(k-1)x+1},m_{(k-1)x+2},\cdots,m_{kx}\}$,确定能够表示隐秘信息分组的分数基音延迟候选值子集Us,其中s指分数基音延迟候选值子集的下标,由式(2)计算得到. 例如,当x=1时,若$\mathfrak{M}_k=\{0\}$,则$s=0$,说明子集U_0中的分数基音延迟候选值能够表示隐秘信息分组$\mathfrak{M}_k$;若$\mathfrak{M}_k=\{1\}$,则$s=1$. 当$x=2$时,若$\mathfrak{M}_k=\{0,0\}$,则$s=0$;若$\mathfrak{M}_k=\{0,1\}$,则$s=1$;若$\mathfrak{M}_k=\{1,0\}$,则$s=2$;若$\mathfrak{M}_k=\{1,1\}$,则$s=3$.

$$s=\sum_{i=1}^{x}m_{(k-1)x+i}\cdot 2^{x-i} \tag{2}$$

Step4:计算分数基音延迟候选值集合U中每个候选值内插后的归一化相关性系数,将所有候选值对应

的系数构成的集合记为$\mathcal{R}$,并定义一个临时集合变量$\mathcal{R}$,令$\mathcal{R}=\varnothing$.

Step5:从$\mathcal{R}=\mathcal{R}-\mathcal{R}$中选择值最大的元素(即内插后的归一化相关性系数最大),记为c_y,并确定c_y对应的分数基音延迟候选值u_y($y\in\{1,2,\cdots,n\}$). 若$u_y\in Us$,则将u_y作为搜索结果,即$d_j=u_y$,$k=k+1$,执行Step6;若$u_y\notin Us$,则$\mathcal{R}=\{c_y\}$,循环执行Step5.

Step6:若$k>r$,则嵌入完毕;否则,继续向下一子帧嵌入隐秘信息,即$j=j+1$,执行Step2.

2.2 提取算法

相应地,接收方根据提前和发送方协商的密钥K和隐藏容量的要求(x比特/子帧)确定分数基音延迟候选值集合U被划分后的2^x个互不相交的子集$\{U_0,U_1,\cdots,U_{2^x-1}\}$,进而从VoIP数据流中逐子帧提取隐秘信息,其算法流程包括如下步骤:

Step1:初始化子帧序号$j=1$,隐秘信息分组序号$k=1$,隐秘信息$M=\varnothing$.

Step2:判断子帧f_j是否被嵌入隐秘信息. 首先从f_j提取其整数基音延迟h_j. 若f_j是某个语音帧中的第一或第三子帧(即$j \bmod 4=1$ or 3),且其整数基音延迟$h_j\in T$,则f_j不包含隐秘信息,继续执行Step6;否则,f_j包含隐秘信息,继续执行Step3.

Step3:提取子帧f_j的分数基音延迟参数d_j,并确定d_j所属子集U_s,其中$U_s\in\{U_0,U_1,\cdots,U_{2^x-1}\}$.

Step4:根据s进一步确定隐秘信息分组$\mathfrak{M}_k=\{m_{(k-1)x+1},m_{(k-1)x+2},\cdots,m_{kx}\}$,其中$m_{(k-1)x+i}$($i=1,2,\cdots,x$)可由下式计算,

$$m_{(k-1)x+i}=\left\lfloor\frac{s}{2^{x-i}}\right\rfloor \bmod 2 \tag{3}$$

例如:当$x=1$时,若$d_j\in U_0$,则$\mathfrak{M}_k=\{0\}$;若$d_j\in U_1$,则$\mathfrak{M}_k=\{1\}$. 当$x=2$时,若$d_j\in U_0$,则$\mathfrak{M}_k=\{0,0\}$;若$d_j\in U_1$,则$\mathfrak{M}_k=\{0,1\}$;若$d_j\in U_2$,则$\mathfrak{M}_k=\{1,0\}$;若$d_j\in U_3$,则$\mathfrak{M}_k=\{1,1\}$.

Step5:计算$M=M+\mathfrak{M}_k$,$k=k+1$,执行Step6.

Step6:若$j=L_F$,则提取完毕,返回隐秘信息M;否则,继续从下一子帧提取隐秘信息,即$j=j+1$,执行Step2.

3 性能评估

为了评估本文隐写算法的性能,采用相关研究工作中普遍使用的语音样本库[28~30]进行实验,该样本库共有2 800条长度为10 s、8 kHz采样、16 bit量化的PCM编码语音样本,其中包含汉语男声、汉语女声、英语男声以及英语女声各700条. 不失一般性,我们以VoIP系统中广泛使用的AMR12.2 kb/s编码速率模式为例对算法进行实验,但从原理上来说本文算法亦可广泛适用于VoIP系统中其他语音编码器. 在AMR12.2 kb/s编码速率模式下,分数基音延迟共有7个候选值,可以划分为2^1($x=1$)个或2^2($x=2$)个互不相交的子集,从而本文算法对应存在4比特/帧和8比特/帧的两种隐藏容量,分别记为隐写模式$S_0(1)$和隐写模式$S_0(2)$. 为便于比较本文算法和已有的相关工作,我们定义了九种隐写模式,如表1所示. 为便于比较和描述,各模式均采取最大容量嵌入方式(即嵌入率为100%).

表1 隐写模式的定义

隐写模式	隐写算法	参数设置
$S_0(1)$	本文算法	$x=1$
$S_0(2)$	本文算法	$x=2$
S_1	Huang等的算法[20]	—
S_2	严书凡等的算法[21]	—
S_3	吴志军等的算法[23]	$k=2$, $N=3$
$S_4(3,4)$	Liu等的算法[24]	$\eta_1=3$, $\eta_2=4$
$S_4(2,4)$	Liu等的算法[24]	$\eta_1=2$, $\eta_2=4$
$S_4(2,3)$	Liu等的算法[24]	$\eta_1=2$, $\eta_2=3$
$S_4(0,0)$	Liu等的算法[24]	$\eta_1=0$, $\eta_2=0$

注:吴志军等的算法[23]中,k和N分别为语音帧中待嵌入的隐秘信息长度和载体长度;Liu等的算法[24]中,η_1和η_2分别为相似度匹配的最小和最大阈值.

以下将从隐藏容量、不可感知性和抗检测性三个方面与相关工作进行实验对比来分析所提出隐写算法的性能.

3.1 隐藏容量

隐藏容量是评价隐写算法性能的一个重要指标,本文采用每个语音帧能够嵌入的比特数(即比特/帧)来衡量隐藏容量. 表2给出了本文的两种隐写模式和Huang等[20]、严书凡等[21]、吴志军等[23]以及Liu等[24]的隐写模式的理论最大隐藏容量和实测平均隐藏容量.

从表2中我们可以得出如下结论:

(1)S_1、S_2和S_3模式的理论最大隐藏容量和实测平均隐藏容量是一致的,其原因是Huang等的算法[20]、严

表2 不同隐写模式的隐藏容量对比

隐写模式	理论最大隐藏容量/(比特/帧)	实测平均隐藏容量/(比特/帧)
$S_0(1)$	4	3.713
$S_0(2)$	8	7.426
S_1	4	4
S_2	3	3
S_3	2	2
$S_4(3,4)$	1.5	1.507
$S_4(2,4)$	3	3.020
$S_4(2,3)$	4	4.014
$S_4(0,0)$	8	8

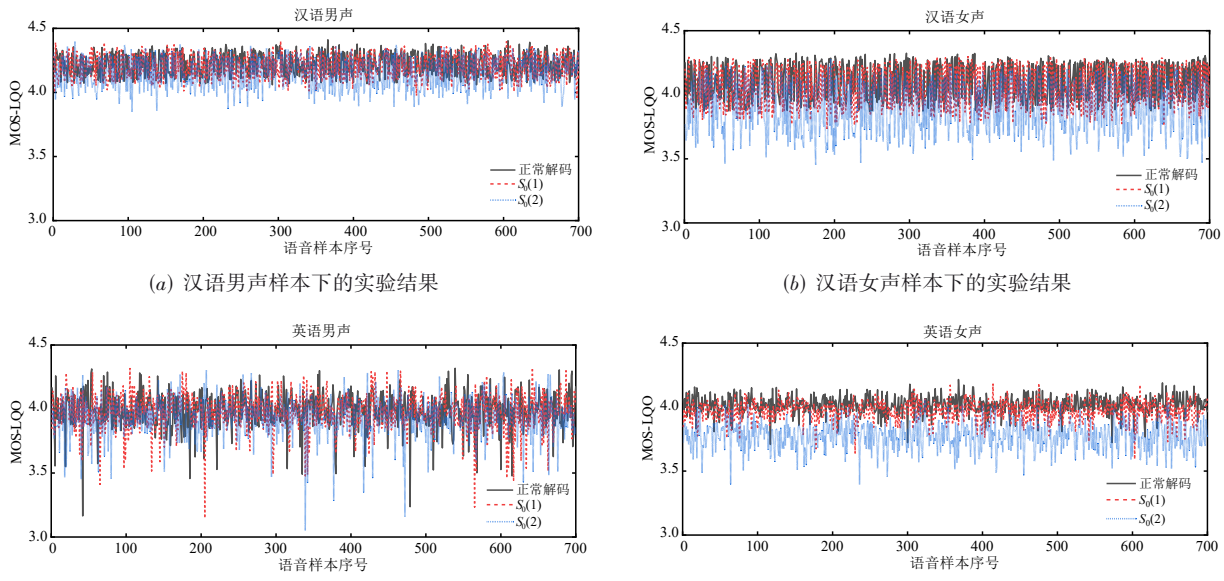
书凡等的算法[21]和吴志军等的算法[23]的嵌入位置相对固定,不会随载体特性发生变化.

(2)Liu等的算法[24]根据嵌入参数的不同,可以提供多种嵌入容量,但由于它采用的是载体自适应隐写方式,因而,除最大容量模式$S_4(0,0)$外,其他模式下的实测隐藏容量与理论隐藏容量略有差别(在理论隐藏容量附近波动);此外,Liu等的算法在所有的分数基音延迟中都无差别地嵌入隐秘消息,忽略了有些基音延迟不存在分数部分的事实,因而存在着较大的安全风险.

(3)本文算法可以提供两种不同隐藏容量的隐写模式$S_0(1)$和$S_0(2)$,其中$S_0(2)$能达到与$S_4(0,0)$一致的最大容量,而$S_0(1)$的隐藏容量与S_1和$S_4(2,3)$两种模式相当,但大于S_2、S_3、$S_4(3,4)$和$S_4(2,4)$四种模式;另外,本文算法的实测平均隐藏容量均略小于其理论最大隐藏容量,其原因是本文算法避免了在不存在分数部分的基音延迟中隐藏信息. 具体来说,在AMR 12.2 kb/s编码速率模式下,当第一或者第三子帧的基音延迟范围为[95,143]时,该基音延迟只有整数部分,因而不适合进行信息嵌入. 经实验统计,2 800条10 s语音样本中出现此类情况的比例约为7.17%,因而,$S_0(1)$和$S_0(2)$两种模式的实测平均隐藏容量较之理论最大隐藏容量会略小约7.17%,但有效避免了Liu等的算法存在的安全漏洞.

3.2　不可感知性

为了评估隐写算法的不可感知性(感知透明性),本文采用MOS-LQO(Mean Opinion Score-Listening Quality Objective)值来评估隐写操作对于语音质量的影响. MOS-LQO值是依据ITU-T P.862标准,即客观语音质量评估(Perceptual Evaluation of Speech Quality,PESQ)[31]标准给出的一种语音质量客观评价分数,其取值范围为[1.017,4.549],MOS-LQO值越大表明语音质量越好. 我们给出了各类语音样本在正常解码及分别采用$S_0(1)$和$S_0(2)$模式隐写后的MOS-LQO值对比(如图1所示),统计了2 800个样本在正常解码和以不同隐写模式隐写后的MOS-LQO均值(如图2所示),以及不同模式隐写后MOS-LQO的平均下降幅度(如表3所示). 从中我们可以得出如下结论.



(a) 汉语男声样本下的实验结果　(b) 汉语女声样本下的实验结果

(c) 英语男声样本下的实验结果　(d) 英语女声样本下的实验结果

图1　正常解码及分别采用$S_0(1)$和$S_0(2)$模式隐写后语音样本的MOS-LQO值对比

(1)由于女性语音的基音频率范围比男性语音的更广[20],且基音频率的变化速度也更快,因而,在相同隐写模式下,女性样本的语音质量受隐写操作的影响更大.

(2)对于本文隐写算法的两种隐写模式而言,隐写容量越大,MOS-LQO变化越大,即$S_0(2)$的MOS-LQO曲线较之$S_0(1)$的曲线偏离正常解码曲线的范围更大. 但总体而言,语音质量影响不大,$S_0(1)$模式隐写后的各类样本MOS-LQO均值在3.97至4.20之间(平均下降幅度在0.012至0.050之间),$S_0(2)$模式隐写后的各类样本MOS-LQO均值在3.78至4.15之间(平均下降幅度在0.060至0.241之间),说明本文隐写算法能够实现较好的不可感知性.

(3)对于理论隐写容量均为8比特/帧的隐写模式$S_0(2)$和$S_4(0,0)$来说,前者的语音质量明显高于后者,在各类语音样本下前者MOS-LQO的下降幅度仅为后者的30.77%至57.38%;对于理论隐写容量均为4比特/帧的隐写模式$S_0(1)$、S_1和$S_4(2,3)$,$S_0(1)$的语音质量明

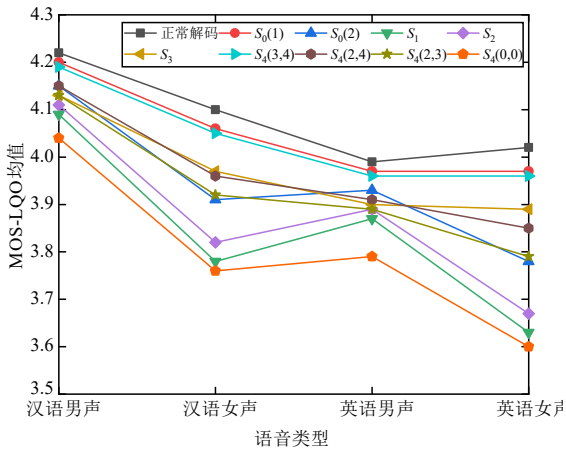


图2 不同隐写模式下所有语音样本MOS-LQO均值的对比

显高于另两种模式，在各类语音样本下$S_0(1)$模式MOS-LQO的下降幅度仅为S_1模式($S_4(2,3)$模式)的10.43%至13.75%(12.50%至23.53%)；不仅如此，较之理论隐写容量均为3比特/帧的隐写模式S_2和$S_4(2,4)$、理论隐写容量为2比特/帧的隐写模式S_3及理论隐写容量为1.5比特/帧的隐写模式$S_4(3,4)$，$S_0(1)$模式隐写后的语音质量依然更好，在各类语音样本下$S_0(1)$模式MOS-LQO的下降幅度仅为$S_4(3,4)$模式的38.71%至84.62%. 由此看出，较之已有隐写算法，本文隐写算法能够在保证相同、甚至更高的隐写容量前提下获得更好的语音质量，即实现更好的不可感知性.

表3 不同隐写模式下所有语音样本MOS-LQO的平均下降幅度

隐写模式	MOS-LQO的平均下降幅度			
	汉语男声	汉语女声	英语男声	英语女声
$S_0(1)$	0.014	0.044	0.012	0.050
$S_0(2)$	0.065	0.193	0.060	0.241
S_1	0.124	0.320	0.115	0.391
S_2	0.108	0.287	0.097	0.345
S_3	0.090	0.136	0.088	0.120
$S_4(3,4)$	0.025	0.052	0.031	0.061
$S_4(2,4)$	0.068	0.143	0.074	0.172
$S_4(2,3)$	0.090	0.187	0.096	0.224
$S_4(0,0)$	0.181	0.346	0.195	0.420

3.3 抗检测性

为进一步评估各算法的抗检测性能，本文分别采用Ren等[25]、Tian等[26]和Liu等[27]提出的基于基音延迟参数的隐写分析算法对表1中各隐写模式进行检测实验. 实验中，对于每种隐写模式，构造由2 800个原始语音样本和对应隐写样本组成的数据集，并按4:1的比例将其划分为训练集和测试集. 表4~6分别给出了Ren等[25]、Tian等[26]和Liu等[27]的隐写分析方法对于各种隐写模式检测正确率的统计结果.

从表4~6中可以得出如下结论.

表4 不同隐写模式在Ren等隐写分析方法[25]下的检测正确率

	$S_0(1)$	$S_0(2)$	S_1	S_2	S_3	$S_4(3,4)$	$S_4(2,4)$	$S_4(2,3)$	$S_4(0,0)$
汉语男声	52.50%	48.93%	97.50%	99.64%	68.93%	50.00%	50.00%	50.00%	50.00%
汉语女声	48.93%	51.07%	100.00%	100.00%	72.14%	50.00%	50.00%	50.00%	50.00%
英语男声	51.43%	52.86%	96.79%	98.93%	67.14%	50.00%	50.00%	50.00%	50.00%
英语女声	52.86%	57.50%	100.00%	100.00%	86.07%	50.00%	50.00%	50.00%	50.00%

表5 不同隐写模式在Tian等隐写分析方法[26]下的检测正确率

	$S_0(1)$	$S_0(2)$	S_1	S_2	S_3	$S_4(3,4)$	$S_4(2,4)$	$S_4(2,3)$	$S_4(0,0)$
汉语男声	49.29%	64.64%	98.93%	100.00%	70.00%	50.00%	50.00%	50.00%	50.00%
汉语女声	52.50%	64.64%	99.64%	100.00%	75.71%	50.00%	50.00%	50.00%	50.00%
英语男声	54.64%	60.36%	97.14%	99.64%	70.71%	50.00%	50.00%	50.00%	50.00%
英语女声	55.71%	66.43%	100.00%	100.00%	89.64%	50.00%	50.00%	50.00%	50.00%

表6 不同隐写模式在Liu等隐写分析方法[27]下的检测正确率

	$S_0(1)$	$S_0(2)$	S_1	S_2	S_3	$S_4(3,4)$	$S_4(2,4)$	$S_4(2,3)$	$S_4(0,0)$
汉语男声	52.86%	60.00%	94.64%	94.29%	71.79%	50.00%	50.00%	50.00%	50.00%
汉语女声	55.00%	65.71%	97.86%	96.43%	75.00%	50.00%	50.00%	50.00%	50.00%
英语男声	52.50%	63.21%	96.07%	97.86%	66.43%	50.00%	50.00%	50.00%	50.00%
英语女声	54.29%	68.93%	99.64%	99.64%	82.50%	50.00%	50.00%	50.00%	50.00%

(1)三种隐写分析方法对于S_1和S_2模式的检测正确率均在94%以上，且有多种情况下甚至达到100%，对于S_3模式的检测正确率在66.43%至89.64%之间，说明Huang等[20]和严书凡等[21]的隐写算法已经无法对抗现有的检测方法，吴志军等[23]的隐写算法由于隐藏容量相比于Huang等[20]的算法减少了一半，因而被检测的正确率也相应降低，但仍存在较大的被成功检测的风险.

(2)Liu等的算法由于只在分数基音延迟部分隐

写,并且在嵌入操作后还执行了“整数基音延迟覆盖”操作,使得整数基音延迟隐写前后完全一致,因而,现有三种基于整数基音延迟特性的隐写分析方法完全失效(检测正确率为50%),说明Liu等的算法能够有效对抗现有的隐写分析方法.

(3)在不同语种样本下三种隐写分析方法对于$S_0(1)$模式的检测正确率均在50%左右(最高不超过56%),对于$S_0(2)$模式的检测正确率在48.93%至68.93%之间,说明本文算法也能够有效对抗现有的隐写分析方法.检测正确率波动的原因是修改分数基音延迟参数对整数基音延迟参数产生了影响.检测正确率接近50%说明整数基音延迟隐写前后几乎未改变,而检测正确率达到68%说明对整数基音延迟有一定的影响,整数基音延迟隐写前后发生了部分变化.当然,为了消除这一影响,本文算法亦可融入“整数基音延迟覆盖”操作,并对应存在可保证隐写前后的整数基音延迟参数完全一致的两种隐写模式$S'_0(1)$和$S'_0(2)$,其对抗已有隐写分析方法性能的实验结果如表7所示.不难看出,现有隐写分析方法对于$S'_0(1)$和$S'_0(2)$两种模式的检测正确率均为50%,表明由于隐写前后的整数基音延迟参数完全一致,目前的隐写分析方法[25-27]均无法有效检测$S'_0(1)$和$S'_0(2)$模式.

表7　$S'_0(1)$和$S'_0(2)$隐写模式在现有隐写分析方法下的检测正确率

	Ren等[25]的隐写分析方法		Tian等[26]的隐写分析方法		Liu等[27]的隐写分析方法	
	$S'_0(1)$	$S'_0(2)$	$S'_0(1)$	$S'_0(2)$	$S'_0(1)$	$S'_0(2)$
汉语男声	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
汉语女声	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
英语男声	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
英语女声	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%

进一步,我们随机选择了30个10 s语音样本,对以上所有隐写模式满嵌时的算法时间开销进行了测试,结果如图3所示.从中不难看出:

(1)吴志军等的S_3模式、Liu等的四种模式($S_4(3,4)$、$S_4(2,4)$、$S_4(3,4)$、$S_4(2,4)$)和本文算法融合“整数基音延迟覆盖”操作的$S'_0(1)$、$S'_0(2)$模式的隐写算法时间开销明显比其他模式大了将近一倍.S_3模式的算法时间开销较大的原因是吴志军等人的算法首先在矩阵编码之前需要对语音进行一次正常编码得到整数基音延迟参数,然后在第二次语音编码过程中根据矩阵编码的结果决定是否需要重新进行闭环基音搜索,也就是需进行两次编码操作;Liu等的四种模式和本文算法融合“整数基音延迟覆盖”操作的$S'_0(1)$、$S'_0(2)$模式的算法时间开销较大的原因则出在“整数基音延迟覆盖”操作上,因为它同样需要对语音进行两次编码,即用第一次正常编码保存的整数基音延迟覆盖第二次编码得到的整数基音延迟.

(2)就本文算法而言,隐写模式$S_0(1)$和$S_0(2)$的时间开销比$S'_0(1)$和$S'_0(2)$低得多.因此,在实时性要求较高的场景下,推荐使用隐写模式$S_0(1)$和$S_0(2)$,而在实时性要求不高但是对抗检测能力要求更高时,建议选用隐写模式$S'_0(1)$和$S'_0(2)$模式.

此外,如前文所述,Liu等的算法由于忽略了部分基音延迟不存在分数值的事实,在所有的分数基音延迟中都无差别地嵌入隐秘消息,使得隐写操作事实上比较容易被发现.我们统计了本文算法和Liu等人的算法所涉及的各种模式隐写后存在“异常”分数基音延迟值(即第一或者第三子帧的基音延迟在[95,143]时

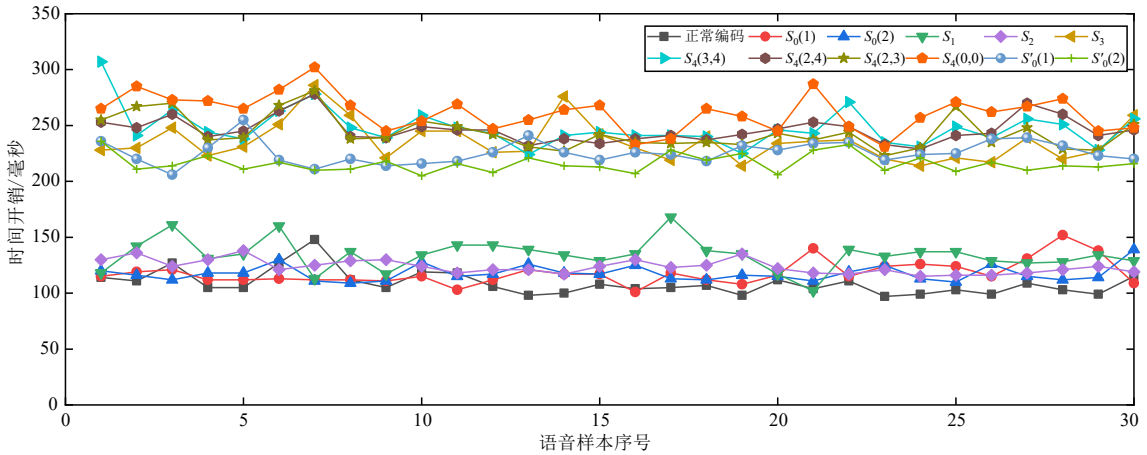


图3　不同隐写模式的算法时间开销对比

存在分数部分)的样本比例,如表8所示.不难看出,$S_4(2,3)$和$S_4(0,0)$模式隐写后的所有样本都包含“异常”分数基音延迟值,能够100%被检测出来;$S_4(3,4)$和$S_4(2,4)$模式隐写后的样本可以被成功检测的概率在94%以上;相较之下本文两种模式隐写后均不存在这类“异常”样本,而无法被检测.

表8 不同隐写模式下存在“异常”分数基音延迟值的样本比例

	$S_0(1)$	$S_0(2)$	$S_4(3,4)$	$S_4(2,4)$	$S_4(2,3)$	$S_4(0,0)$
汉语男声	0	0	94.57%	94.57%	100%	100%
汉语女声	0	0	94.57%	94.57%	100%	100%
英语男声	0	0	94.29%	94.57%	100%	100%
英语女声	0	0	94.43%	94.57%	100%	100%

综上,不管是现有的三种基音延迟隐写分析方法还是基于异常分数基音延迟值的检测方法均无法有效检测本文隐写算法,而其他隐写算法均存在不同程度上被检测的可能.由此可以说明,本文提出的隐写算法较之现有算法具有更强的抗检测性能.

4 总结

基于基音延迟的隐写是语音隐写的重要分支之一.然而,现有算法在隐写透明性和抗检测能力上存在不同程度的问题.鉴于此,本文提出了一种新的基于分数基音延迟动态搜索的隐写算法.该算法可根据隐藏容量的需要将分数基音延迟候选值集合划分为若干个子集,每个子集代表不同的比特信息;在闭环基音搜索过程中,为每个子帧选择既能表示待嵌入隐秘信息且内插后的归一化相关系数最大的分数基音延迟候选值作为搜索结果以减少隐写操作对语音信号的影响.以目前VoIP系统中广泛使用的AMR语音编码为例,对该算法从隐藏容量、不可感知性及抗检测性三方面进行了性能评估并与相关工作进行对比分析.实验结果表明,本文提出的隐写算法不仅能够有效地对抗现有隐写分析算法,且较之现有基于基音延迟的隐写算法在确保较高隐写容量的同时可达到更好的不可感知性.值得指出的是,虽然本文以基音延迟域为对象描述基于动态最优搜索的隐写算法,但从语音编码原理的角度看,该思路还可进一步扩展到线性预测参数域和固定码本参数域,这也是我们未来的重要研究工作之一.同时,由于该方法目前尚无有效检测手段,从避免其“滥用”的角度出发,研究对应的隐写分析方法也是一个值得深入的研究课题.

参考文献

[1] PROVOS N, HONEYMAN P. Hide and seek: An introduction to steganography[J]. IEEE Security&Privacy, 2003, 1(3): 32-44.

[2] 田晖, 郭舒婷, 秦界, 等. 基于可量化性能分级的自适应IP语音隐写方法[J]. 电子学报, 2016, 44(11): 2735-2741.
TIAN H, GUO S T, QIN J, et al. Adaptive voice-over-IP steganography based on quantitative performance ranking[J]. Acta Electronica Sinica, 2016, 44(11): 2735-2741.(in Chinese)

[3] MAZURCZYK W. VoIP steganography and its detection—a survey[J]. ACM Computing Surveys, 2013, 46(2): 1-21.

[4] ZIELIŃSKA E, MAZURCZYK W, SZCZYPIORSKI K. Trends in steganography[J]. Communications of the ACM, 2014, 57(3): 86-95.

[5] XU T T, YANG Z. Simple and effective speech steganography in G.723.1 low-rate codes[C]//Proceedings of the 2009 International Conference on Wireless Communications &Signal Processing. Nanjing: IEEE, 2009: 1-4.

[6] HUANG Y F, YUAN J, CHEN M C, et al. Key distribution over the covert communication based on VoIP[J]. Chinese Journal of Electronics, 2011, 20(2): 357-360.

[7] 王继军, 李国祥, 夏国恩, 等. 图像插值空间完全可逆可分离密文域信息隐藏算法[J]. 电子学报, 2020, 48(1): 92-100.
WANG J J, LI G X, XIA G E, et al. A separable and reversible data hiding algorithm in encrypted domain based on image interpolation space[J]. Acta Electronica Sinica, 2020, 48(1): 92-100. (in Chinese)

[8] YANG Z L, ZHANG S Y, HU Y T, et al. VAE-Stega: Linguistic steganography based on variational auto-encoder[J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 880-895.

[9] RANA S, KAMRA R, SUR A. Motion vector based video steganography using homogeneous block selection[J]. Multimedia Tools and Applications, 2020, 79(9): 5881-5896.

[10] XIAO B, HUANG Y, et al. An approach to information hiding in low bit-rate speech stream[C]//Proceedings of the IEEE GLOBECOM 2008. New Orleans: IEEE, 2008: 1-5.

[11] ITU-T G.723.1: Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s[S/OL]. [2021-04-20].http://www.rosoo.net/Files/UpFiles/RsProduct/unsorted/20071/2007189302247358.pdf.

[12] ITU-T G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP) [S/OL]. [2021-04-20]. https://www.doc88.com/p-5911700978695.html.

[13] ETSI EN 301 704 V7.2.1: Adaptive multi-rate (AMR) speech transcoding[S/OL]. [2021-04-20].https://www.etsi.org/deliver/etsi_en/301700_301799/301704/07.02.01_60/en_301704v070201p.pdf.

[14] LIU P, LI S, et al. Steganography integrated into linear

predictive coding for low bit-rate speech codec[J]. Multimedia Tools and Applications, 2017, 76(2): 2837-2859.

[15] 孙鑫昊, 王开西. 基于最短欧氏距离替换码元的VoIP隐写算法[J/OL]. 计算机工程与应用, 2021: 1-9. http://kns.cnki.net/kcms/detail/11.2127.TP.20210419.1512.086.html.
SUN X H, WANG K X. The codeword replacement based on shortest euclidean distance for VoIP steganography[J/OL]. Computer Engineering and Applications, 2014: 1-9. http://kns.cnki.net/kcms/detail/11.2127.TP.20210419.1512.086.html. (in Chinese)

[16] GEISER B, VARY P. High rate data hiding in ACELP speech codecs[C]//Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. Las Vegas: IEEE, 2008: 4005-4008.

[17] MIAO H B, HUANG L S, CHEN Z L, et al. A new scheme for covert communication via 3G encoded speech[J]. Computers & Electrical Engineering, 2012, 38(6): 1490-1501.

[18] SU Z P, LI W W, ZHANG G F, et al. A steganographic method based on gain quantization for iLBC speech streams[J]. Multimedia Systems, 2020, 26(2): 223-233.

[19] 余迟, 黄刘生, 等. 一种针对基音周期的3G信息隐藏方法[J]. 小型微型计算机系统, 2012, 33(7): 1445-1449.
YU C, HUANG L S, et al. A 3G speech data hiding method based on pitch period[J]. Journal of Chinese Computer Systems, 2012, 33(7): 1445-1449. (in Chinese)

[20] HUANG Y F, et al. Steganography integration into a low-bit rate speech codec[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(6): 1865-1875.

[21] 严书凡, 等. 基于基音周期预测的低速率语音隐写[J]. 计算机应用研究, 2015, 32(6): 1774-1777.
YAN S F, et al. Steganography for low bit-rate speech based on pitch period prediction[J]. Application Research of Computers, 2015, 32(6): 1774-1777. (in Chinese)

[22] 刘程浩, 柏森, 黄永峰,等. 一种基于基音预测的信息隐藏算法[J]. 计算机工程, 2013, 39(2): 137-140.
LIU C H, BAI S, HUANG Y F, et al. An information hiding algorithm based on pitch prediction[J]. Computer Engineering, 2013, 39(2): 137-140. (in Chinese)

[23] 吴志军, 等. 基于随机位置选择和矩阵编码的语音信息隐藏方法[J]. 电子与信息学报, 2020, 42(2): 355-363.
WU Z J, et al. Speech information hiding method based on random position selection and matrix coding[J]. Journal of Electronics & Information Technology, 2020, 42(2): 355-363. (in Chinese)

[24] LIU X K, TIAN H, et al. A novel steganographic method for algebraic-code-excited-linear-prediction speech streams based on fractional pitch delay search[J]. Multimedia Tools and Applications, 2019, 78(7): 8447-8461.

[25] REN Y Z, YANG J, WANG J W, et al. AMR steganalysis based on second-order difference of pitch delay[J]. IEEE Transactions on Information Forensics and Security, 2016, 12(6): 1345-1357.

[26] TIAN H, et al. Steganalysis of adaptive multi-rate speech using statistical characteristics of pitch delay[J]. Journal of Universal Computer Science, 2019, 25: 1131.

[27] LIU X K, TIAN H, LIU J, et al. Steganalysis of adaptive multiple-rate speech using parity of pitch-delay value[C]// Proceedings of the International Conference on Security and Privacy in New Computing Environments. Tianjin, China: Springer, 2019: 282-297.

[28] TIAN H, WU Y P, et al. Distributed steganalysis of compressed speech[J]. Soft Computing, 2017, 21(3): 795-804.

[29] TIAN H, WU Y P, CHANG C C, et al. Steganalysis of adaptive multi-rate speech using statistical characteristics of pulse pairs[J]. Signal Processing, 2017, 134(C): 9-22.

[30] TIAN H, WU Y P, CHANG C C, et al. Steganalysis of analysis by-synthesis speech exploiting pulse-position distribution characteristics[J]. Security and Communication Networks, 2016, 9(15): 2934-2944.

[31] ITU-T Recommendation P.862. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs[S/OL]. [2021-04-20]. https://www.itu.int/rec/T-REC-P.862.

作者简介

田　晖　男,1982年10月出生,湖北赤壁人.博士,教授,博士生导师.主要研究领域为网络与信息安全、数据安全、人工智能安全、信息隐藏及检测、数字取证等.
E-mail: htian@hqu.edu.cn

严　艳　女,1997年2月出生,江西赣州人.华侨大学计算机科学与技术学院硕士研究生.主要研究方向为信息隐藏及检测、深度学习.