文章编号: 1002-1175(2009)04-0474-09

一种基于感知特性的鲁棒性语音认证算法*

古 今 郭 立† 郑东飞

(中国科学技术大学电子科学与技术系, 合肥 230027)

(2009 年 1 月 9 日收稿; 2009 年 3 月 18 日收修改稿)

Gu J, Guo L, Zheng DF. A robust speech authentication algorithm based on perceptual characteristics[J]. Journal of the Graduate School of the Chinese Academy of Sciences, 2009, 26(4): 474~482.

摘 要 提出了一种基于感知域的鲁棒性语音认证算法,将语音的感知特性与签名算法相结合,在满足内容认证和身份认证的同时,能够有效地抵抗通信噪声微扰.算法基于语音的掩蔽效应和非线性效应等人耳感知特性,着重去除其时频域掩蔽阈值下的冗余信息,进行非线性滤波后提取感知参数,并运用改进的 Rainbow 算法对这些语音参数进行签名.实验证明,该算法的唯一性和针对通信噪声的鲁棒性都很好,兼有 Rainbow 签名的安全性保证,可以满足语音通信中的鲁棒性认证要求.

关键词 感知域,鲁棒性,语音认证

中图分类号 TP309.7

1 引言

在目前开放式的通信环境中,篡改窃听传输中数据非常容易,信息来源也较难认证.日渐增长的多媒体数据传输的安全认证需求引起了广泛关注.但迄今为止,大部分认证研究,如签名和水印等,都集中在数字图像方面,对音频认证的相关研究较少.鉴于目前语音通信的迅猛发展,对语音数据的认证研究十分必要.

数字签名和数字水印是目前多媒体认证的 2 种主要形式.由于(半)脆弱水印的认证技术需要修改传输数据本身,在某些情况下会造成不希望的改动[1].所以本文选择研究基于数字签名的方案.语音认证需要解决两方面问题:内容认证和身份认证.数字签名中验证内容完整性的主流算法,如单向哈希函数 MD5 和美国政府的安全哈希算法 SHA,将所有的数据当作二进制比特流计算散列值.但这种方式并不适用于语音信息[2].首先,对于庞大的语音数据,hash 函数计算量过大,效率较低.其次,原始数据任何微小的改变都会产生截然不同的散列值.而语音信息中少量比特的变动往往并不会影响整体听觉,不应该导致认证的失败.因此需要对传统的认证方式进行改进,提取感知特性,削弱人耳难以感知的信息对认证结果的影响,兼顾鲁棒性和篡改敏感性.另外,一般用于身份认证的公钥签名算法,如 RSA 和 ECC,计算量很大,不适用于资源受限的通信终端.而新兴的多变量公钥算法中运算速度快,占用资源较少.因此本文采用多变量公钥算法作为语音数据的签名方案.根据语音应用进行修改后,将其整合到本文的感知认证过程中.

基于以上分析,本文将语音的感知特性与认证算法相结合,提出了一种算法,提取语音数据中声学

*国家自然科学基金项目(60772032)资助

† 通讯联系人, E-mail: lguo@ustc.edu.cn

感知上最重要的信息并对其进行认证. 这将使得经过信道微扰后感知上保持相近的语音可以得到相同的特征摘要, 从而可以通过认证, 而篡改后的语音则得到不同的摘要, 不能通过认证. 目前已有不少感知摘要算法的研究. Haitsma 提出一种基于频带之间能量差阈值的摘要提取算法[3]; Mihcak 提出从时频特征中随机抽取统计特征的算法[4]; Burges 用 MCLT 计算对数谱, 然后用 PCA 对谱参数进行分析[5]; Sukittanon 和 Atlas 运用频域特征调制和小波变换进行特征提取[6]; Foote 和 Logan 则运用 Mel 谱参数作为特征集[7] 等. 但所有这些研究都是利用统计特性区分不同音频, 主要用于音乐数据库的搜索或音频信号分类. 因此只要求与其他音频的统计特性有一定区分度即可, 允许合法操作后的摘要结果出现一定范围内的误差. 而本文的感知摘要针对签名设计, 摘要的一点区别都会导致认证失败, 因此对特征提取的摘要算法在特定干扰下的鲁棒性要求更高.

整个认证算法的设计流程为: 首先根据语音的时频域掩蔽效应, 去除掩蔽阈值之下的冗余信息; 然后根据耳蜗的非线性滤波效应, 采用 MFCC 算法提取语音的 Mel 参数并进行多维矢量量化; 最后改进多变量公钥体系中的 Rainbow 算法, 并用其对提取量化后的参数进行签名.

2　算法模型

传统的精确认证算法受传输环境影响很大, 在有噪声存在的信道中性能急剧恶化, 不适用于语音信息的认证. 而人类听觉系统的识别和认证却有着很强的鲁棒性. 因此, 利用人类听觉系统的感知特性, 去除冗余, 从语音中提取重要的感知参数, 可以尽可能减轻非关键信息的变化对认证结果的影响, 增强认证鲁棒性. 算法模型如图 1 所示.

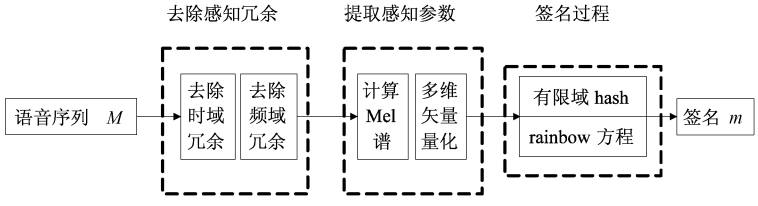

图 1　基于感知域的认证算法流程图

从图 1 的流程可以看到, 本文基于听觉掩蔽效应来去除感知冗余, 基于非线性滤波器效应提取感知参数, 然后再用改进后的 Rainbow 算法对感知参数进行签名认证. 下面分别进行分析.

2.1　掩蔽效应

心理声学的一个重要现象是听觉掩蔽效应, 分为时域掩蔽和频域掩蔽, 如图 2 所示. 时域掩蔽是指当 2 个音信号在时间上相邻时, 强音会对弱音有一定的掩蔽. 时域掩蔽分为前掩蔽和后掩蔽, 其中后掩蔽的时间可长达 200ms, 远大于前掩蔽的时间. 而当掩蔽信号和被掩蔽信号同时发生且处在相近频率上时, 则属于频域掩蔽[8]. 掩蔽效应, 特别是频域掩蔽效应, 一般被用于感知语音编码. 本文则将其应用到语音认证领域, 通过计算时频域掩蔽阈值, 去除或削弱阈值以下的语音信号的影响, 达到选择性认证的效果.

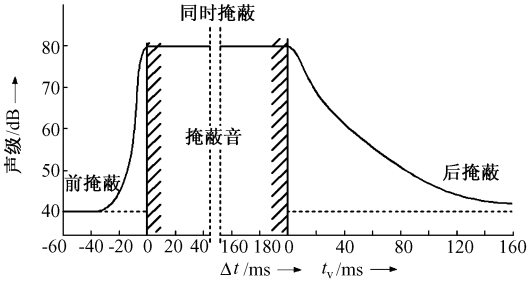

图 2　掩蔽效应

2.2　非线性滤波器效应

在人的听觉系统中, 耳蜗的作用相当于一个滤波器组. 由于声音的不同频率成分沿着基底膜的分布是对数型的, 因此耳蜗的滤波作用相当于在一个非线性频率尺度上进行, 使得人耳对低频信号比对高频信号更敏感.

为此, 人们用 Mel 频率刻度对实际频率轴进行弯折来模拟人耳所听到的声音高低与声音频率之间的非线性关系. Mel 标度与频率的关系可近似表示为[9]: $\mathrm{mel}(f)=1127\times \lg(1+f/700)$.

在 MFCC 特征提取算法中，输入语音帧首先通过傅里叶变换得到频谱，然后在频域上应用一组 Mel 频率上均匀分布的滤波器来获得类似人耳听觉特性的非线性频谱分辨率. MFCC 是特征提取中最重要和最成功的特征，在语音识别中应用广泛. 本文将其与矢量量化相结合，提取鲁棒性感知参数进行后续签名.

2.3 Rainbow 签名算法

传统公钥算法大多建立在大数分解或者离散对数的基础上，不适用于语音通信. 以最广泛使用的 RSA 为例. 首先，运算量较大. 随着目前破解能力的不断提高，RSA 等算法的参数也随之不断增大，对运算资源要求日渐增高. 其次，RSA 并未被证明是一个 NPC 问题. Shor 于 1994 年提出的关于大数因子分解的量子多项式算法表明，一旦量子计算机成为现实，目前应用广泛的 RSA 便不再安全[10].

多变量公钥算法体系(multivariate public key cryptosystems，MPKC)是传统公钥算法的重要替代算法之一. 由于在有限域上解一组多变量多项式方程是一个 NPC 问题，即使将来实用性的量子计算机问世也无法破解，因此安全性较高. 另外 MPKC 运算速度快，耗费的资源也少，尤其适用于经典公钥算法(RSA，ECC 等)所不适合的场合，如智能卡、无线传感器网络以及移动通信终端等[11]. 对于语音通信而言，通话的实时性使得算法的速度尤为重要，同时手机等通信终端上的运算资源也较为有限. 因此本文采用 MPKC 中的 Rainbow 算法对语音信息进行签名认证.

3 基于感知域的认证算法

3.1 去除感知冗余信息

3.1.1 利用时域掩蔽去除冗余

考虑时域掩蔽时，本文选择掩蔽效应更显著的后掩蔽. 根据掩蔽阈值随时间推移呈指数下降的特性，采用 Jesteadt[12] 的模型进行模拟.

$$FM = 0.7(2.3 - \log_{10}\Delta t)(L - 20), \tag{1}$$

其中，FM 是求得的掩蔽阈值，单位为 dB. Δt 为掩蔽音与被掩蔽音之间的时间差，单位为 ms. L 为掩蔽音的能量，单位为 dB. 将语音分为 20～32ms 范围内固定长度的帧，再将每一帧分为 4 个子帧. 分别对每个子帧计算掩蔽阈值 FM_j，计算 L_j 时包括了前一帧和当前帧前面 $j-1$ 个子帧的所有能量. 再选取 4 个子帧掩蔽阈值的最大值作为整帧的掩蔽阈值. 若某帧的能量值小于该帧的掩蔽阈值，则将该帧的信号置零. 因为被掩蔽的部分多为能量较小的音帧，最易被噪声影响，造成认证失败，因此这一步骤可以去除噪声对不易察觉的声音的影响，提高认证鲁棒性. 原始语音波形与处理后的语音波形比较如图 3 所示. 其中横轴表示时域上的采样点数，纵轴表示时域信号的归一化幅值.

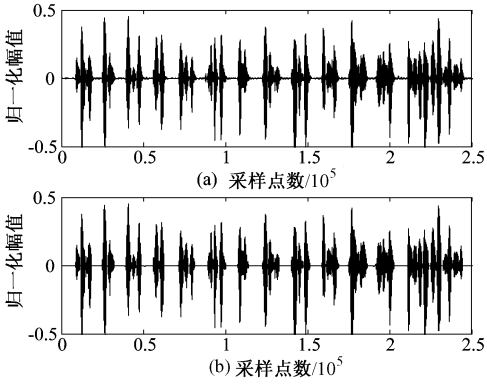

图 3 时域冗余处理前后的波形对比图

从图 3 可以看出，处理前后的音频波形几乎完全一致，说明基于后掩蔽的冗余去除对声音的感知效果完全没有影响. 此外，这一处理使得能量较高的音帧后的静音部分更为平坦，可以有效提高鲁棒性.

3.1.2 利用频域掩蔽去除冗余

图 4 表示一帧数据的频域掩蔽阈值和频域冗余. 其中图 4(a)中的曲线和星号线分别表示该帧数据的掩蔽阈值和功率谱密度. 与之相对，图 4(b)中的曲线和星号点分别表示语音的频谱幅值和冗余表示，纵轴为归一化的频谱幅

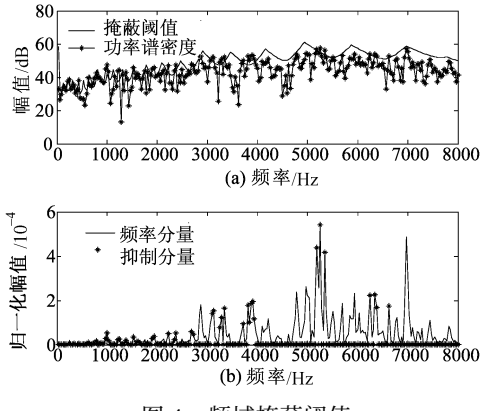

图 4 频域掩蔽阈值

度, 为零的星号点表示属于冗余应该被抑制的部分, 其余星号点表示可感知应该被增强的部分. 由此可见, 一帧频域数据中, 不可感知的冗余占了相当部分, 对其进行抑制从而提高鲁棒性有很大的空间.

根据以上分析, 对冗余信息进行消减. 根据生理声学模型 Model 1 Layer 1 中定义的掩蔽模型, 计算一帧数据的全局掩蔽阈值. 按照下式处理对应频率点的值:

$$F(k) = F(k) \times P(k) / (T(k) - c), \tag{2}$$

其中, $F(k)$为频域上的信号值, 为一帧时域数据经过 fft 变换后, 再经过 hamming 窗以消除边际效应的结果. $P(k)$为功率谱密度, $T(k)$为掩蔽阈值. c 为常数, 根据实验选定. 这种处理压缩了人耳无法感知的频域信息, 增强了听阈之上的信息, 进一步在频域上消减了冗余度. 按照(2)式的方法处理而并不直接将掩蔽阈值以下的信息清零的原因是, 如受到干扰前后的信号在某个频率点分别处于掩蔽阈值的两端, 完全清除会使原本微小的差别大幅度扩张, 反而不利于鲁棒性.

3.2　感知参数提取和量化

对上述处理后的频谱计算 Mel 倒谱参数 MFCC. 将频谱能量乘以一组 24 个三角带通滤波器, 求得每一个滤波器输出的对数能量. 然后再对 24 个对数能量 E_k 进行离散余弦变换, 得到 L 阶 MFCC 参数, 在此 L 取 12. 一般 MFCC 还需计算 L 个一阶差分参数, 甚至二阶差分参数. 但针对本文的认证应用, 12 阶基本参数已经可以满足要求.

实验表明, 对于 Gauss 噪声等正常扰动, 相对帧的每一个 MFCC 参数都会有微小的变化. 而对于不同帧或者剪切、篡改等非法操作, MFCC 参数变动较大. 为了区分正常扰动和恶意篡改, 本文采用矢量量化的方式, 将 MFCC 参数视为 12 维的数据, 进行量化分类. 矢量量化选择自组织特征映射神经网络(SOFM)的方法. 在该方法中, 输出层所有神经元通过相互竞争和自适应学习, 形成空间上的有序结构, 从而实现输入矢量到输出矢量空间的特征映射[13]. 用通话开始后的第一个语音片断训练得到码本, 量化得到每一帧的码字.

为了比较矢量量化与线性量化的效果, 我们模拟原始语音与加噪语音提取参数的结果, 构造与原始数据有较小误差的数据, 然后分别用不同量化方式、相同的量化阶数进行量化, 将加噪参数量化结果与原始参数量化结果相比较, 计算量化结果不同的比例. 若比例较小, 说明该量化方法鲁棒性强, 可以容忍微小的扰动.

如图 5 所示, 横轴为数据对象的误差系数, 从 0.002 到 0.3, 图中用自然对数形式表示. 纵轴为差异比例. 从上到下 3 条曲线分别代表 3 种量化方式: 第 1 条是完全线性量化, 将所有参数无区别地量化为 256 阶; 第 2 条是改进的线性量化, 将 12 个参数组根据各自值域, 分别量化, 考虑到不同参数的取值范围的差异, 因此效果优于前者; 第 3 条是矢量量化, 将 MFCC 参数视作 12 维的矢量进行 256 阶量化, 效果在 3 种方式中最优. 且当数据误差控制在原始数据的 0.01 以下时, 矢量量化的差异比例为 0, 即量化结果相同, 这是鲁棒性认证的前提.

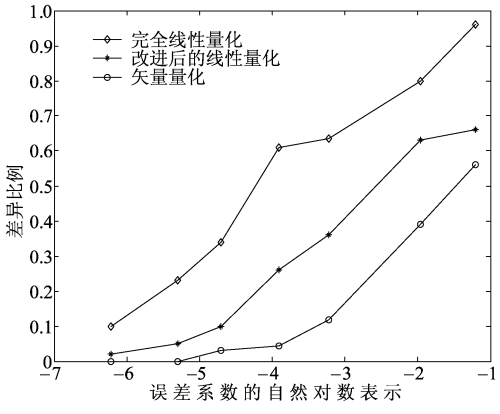

图 5　3 种量化的鲁棒性比较图

3.3　认证签名

认证签名采用多变量公钥体系的 Rainbow 签名算法. 多变量公钥是一个庞大的算法家族, 拥有众多的分支. 特点是在保证较高安全性的前提下, 普遍速度较快. 曾有算法被 NESSIE(new European schemes for signatures, integrity, and encryption)确立为欧洲数字签名标准, 推荐用于智能卡等低端设备. Rainbow 是其中一种建立在单域上的数字签名算法, 有着较好的安全性和效率[14].

多变量公钥加解密的一般形式如下:

$$F(x'_1, \cdots, x'_n) = \boldsymbol{L}_1 \circ \varphi \circ F \circ \varphi^{-1} \circ \boldsymbol{L}_2(x'_1, \cdots, x'_n) = (f_1, \cdots, f_n), \tag{3}$$

$$F^{-1}(y'_1, \cdots, y'_n) = \boldsymbol{L}_2^{-1} \circ \varphi \circ F^{-1} \circ \varphi^{-1} \circ \boldsymbol{L}_1^{-1}(y'_1, \cdots, y'_n), \tag{4}$$

其中, F 为域 L 上的映射, 随具体算法的不同而不同. 域 L 为域 K 的 n 次扩张, K 是一个特征为 2, 基数为 q 的有限域. $\boldsymbol{L}_1$, $\boldsymbol{L}_2$ 为可逆矩阵. φ 为从 L 到 K^n 的线性同态映射. 在 Rainbow 算法中, (3)式中的 F 形式为: $F = (F_1, \cdots, F_{u-1})$, $F_l = \sum_{i \in O_l, j \in S_l} \alpha_{ij} x_i y_j + \sum_{i, j \in S_l} \beta_{ij} x_i x_j + \sum_{i \in S_{l+1}} \gamma_i x_i + \eta (l \in \{0, \cdots, u-1\})$[14], 其中, $\{x_i \mid i \in O_l\}$ 为油变量, $\{x_j \mid j \in S_l\}$ 为醋变量, 一共 μ 层的彩虹式结构.

Rainbow 的签名认证流程如图 6. 先用 SHA-1 将语音序列 m 映射到定长的位串 $v \| w$, 再用(4)式对其进行签名, 解方程组得到签名 X 随语音进行发送. 接收方收到语音和签名后, 对语音如前依次进行时频域感知处理、参数提取量化、摘要计算, 得 y'. 再将接收到的签名 X 代入(3)式进行计算, 计算结果 Y 与 y' 相比较, 如果相同, 认证成功. 否则认证失败. 本文中 Rainbow 多项式的参数选择为: 层数为 5, 每层的变量数分别为 11、22、27、32、43. 不可约多项式为 $x^6 + x^3 + x^2 + 1$. 该参数的 Rainbow 算法拥有大于 2^{80} 的计算安全性[14].

在传统的 Rainbow 流程中, 需要 3 次 SHA-1, 略显繁琐, 而且在语音终端上硬件实现时需要专门的 SHA-1 运算单元. 为了更好地与整个感知域认证算法以及应用环境相适应, 本文对 Rainbow 算法进行修改, 提出一种基于有限域的方式(galois field, GF-hash)对语音特征提取摘要. 具体流程如图 7 所示.

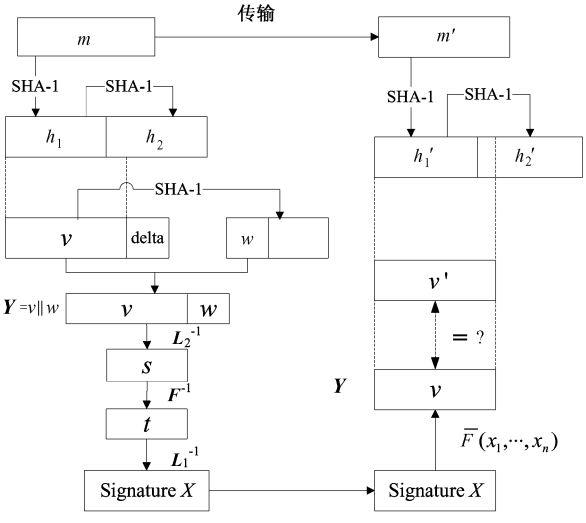

图 6 使用 **SHA-1** 的 **Rainbow** 流程图

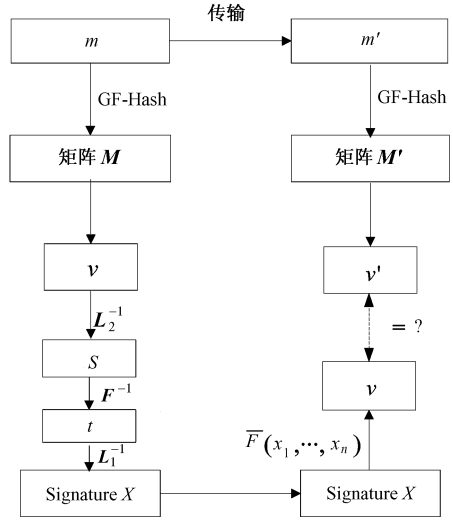

图 7 使用 **GF-Hash** 的 **Rainbow** 流程图

(1)将量化后的码字序列写成连续个 $m \times m$ 矩阵形式. 考虑后续 Rainbow 签名的长度, 将矩阵定为 4×4 大小. 若结果矩阵小于 3 个, 则直接进行签名. 若不足整数个矩阵, 则用开始的码字顺序重复填充.

(2)将矩阵依次计算合并. 如对于矩阵 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$, $\boldsymbol{Y} = f(\cdots f(f(\boldsymbol{A}_1, \boldsymbol{A}_2)\boldsymbol{A}_3), \cdots, \boldsymbol{A}_n)\boldsymbol{A}_1)\boldsymbol{A}_2)$. 其中, 对于每次矩阵操作 $\boldsymbol{Y}_i = f(\boldsymbol{Y}_{i-1}, \boldsymbol{A}_i)$, 有 $c_{ji} = \prod_{k=0}^{m} a_{ik} b_{kj}$. 上式中的乘法运算定义为 GF($2^8$)上的乘法.

(3)将最后得到的 2 个结果矩阵共 32 个元素视为输入变量, 进行 Rainbow 签名.

相比原先的 SHA-1, 这样的摘要构造更适合于整个算法环境: 第一, 前后一致性好. 计算摘要前的输入为范围从 0 到 255 的码字, 而后续签名的对象是 GF(2^8)的元素. 摘要的计算仍然采取有限域 GF(2^8)的形式有助于运算的一致性. 而且, 涉及的运算皆为有限域上的运算, 与后续的 Rainbow 签名相同. 因此若在应用平台上实现时无需引入其他指令和硬件模块, 复用率高. 第二, 在保证安全的情况下速度较快, 适用于语音终端资源有限的环境. 所有的运算全为连续乘法, 而有限域上的元素可看成是生成元 α 的幂次: $K = \{1, \alpha, \cdots, \alpha^{2^m - 2}\}$, 相乘运算相当于其对应幂次相加之和模 $(2^m - 1)$ 的余数, 即: $\alpha^x \times \alpha^y = \alpha^{(x+y) \bmod 127}$[15]. 所以连续乘法可化成映射后的连续加法, 运算速率非常快. 以加法的速度得到乘法的效

果, 性能时间比较高.

对该摘要算法的扩散性进行分析. 首先考虑算法结构, 以单次乘法而论, 有限域上乘法的混淆扩散效应大于普通乘法. 并且两轮这样的矩阵运算之后, 单个元素的差异就会扩散到所有元素, 扩散性和安全性较好. 最后一轮重复对 $\boldsymbol{A}_1$ 进行处理, 保证了即使差异出现在 $\boldsymbol{A}_n$, 也能得到足够的扩散.

其次进行统计分析. 理想的摘要值散布结果是初值的微小扰动将导致摘要值每个 bit 都以 50%的概率发生变化. 本文对 100 个 15s 的语音片断进行测试, 统计如下指标:

平均变化 bit 数 $B=\frac{1}{n}\sum_{n=1}^{N}B_n$, 平均变化率 $P=(B/256)\times 100\%$, 变化 bit 数 B 的均方差 $\Delta B=\sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(B_n-B)^2}$, P 的均方差 $\Delta P=\sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(B_n/256-P)^2}$. 其中, n 为统计次数, B_n 为第 n 次明文改变 1bit 后摘要结果的变化比特数. 将每一bit 进行变化, $N=3944$. 所得结果见表 1. 从表中数据可以看出, 基于有限域的摘要值的平均变化比特数和每比特的平均变化概率都非常接近理想状态下的 128bit 和 50%, 均匀利用了明文空间, 从统计效果上保证了攻击者无法在已知明文密文对的情况下伪造其他明文密文. 同时 ΔB 和 ΔP 都很小, 表明算法对明文的混乱与散布能力强而稳定. 因为该摘要算法仅限于对语音提取的特征数据的应用, 提取语音特征的过程加大了篡改的难度, 再加上语音传输的实时性限制和 Rainbow 算法的安全性保证, 可以认为该算法能够基本满足本文应用环境中防碰撞的要求.

表 1　摘要值变化的统计参数列表

n	B/bit	ΔB/bit	P/(%)	ΔP/(%)
3944	128.26	0.501	11.59	0.045

4　实验结果

本文认证算法的安全性主要由 Rainbow 签名算法保证[14]. 特别的, 对于感知认证, 有 2 个重要性能指标: 鲁棒性和唯一性[16]. 鲁棒性要求对于感知上相同的语音, 认证算法应该得到同一个序列值. 也就是说对于一些正常的处理和扰动, 如重采样、微小噪声、压缩等, 结果应该保持一致. 唯一性则要求对于两个语意和感知上不相同的语音, 结果应该不同, 并且具有很好的扩散性. 唯一性保证了语意上的篡改会导致认证失败. 语音通信过程中, 主要的正常扰动表现为通信过程中的噪声, 这是本文的主要研究目标.

4.1　鲁棒性分析

选取一系列实验语音片断进行实验. 语音数据为 16bit, 采样速率为 16K, 分帧后每帧长度大约 32ms, 基本满足语音短时平稳的条件. 模拟传输信道, 对语音片断掺杂噪声, 进行基于感知域的特征提取.

信道中常见的随机噪声主要是脉冲干扰和起伏噪声. 表 2 表示对语音片断加上随机的脉冲噪声的情况. 设定每次噪音冲击改变一个音频数据的数值, 按照不同的改变位分别统计通过率. 每次实验次数 $n=1024$, i 表示数据被改变位的序号. 表 3 表示对语音片断加上不同功率的高斯白噪声后的认证结果. 其中, 改变率定义为认证失败时摘要结果被改变的比例. 表 4 则比较多种方法对加上高斯白噪声后的语音认证时通过率的不同, 分别使用一般的 MFCC 提取参数, 本文算法 RSAA (robust speech authentication algorithm), 以及 RSAA 中去掉感知冗余消除步骤后的算法 MFCC *.

表 2 单个随机脉冲干扰后的认证通过率 (%)

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
通过率	100	100	100	100	100	100	100	100	100	100	100	99	99	96	94	82
改变率	—	—	—	—	—	—	—	—	—	—	—	1	1	1	1	1

表 3 高斯白噪声干扰后的认证通过率 (%)

SNR/dB	80	70	60	50	40	30	25	20
通过率	100	100	100	100	100	93	60	0
改变率	—	—	—	—	—	0.9	5	26

表 4 多种方法认证通过率的比较 (%)

SNR/dB	80	70	60	50	40	30	25	20
MFCC	78	32	0	0	0	0	0	0
MFCC *	100	100	100	98	89	52	0	0
RSAA	100	100	100	100	100	93	60	0

由表 2、表 3 数据可得, 对于绝大部分脉冲干扰, 该鲁棒性认证都可以通过. 而对于高斯白噪声, 只要控制在人耳不易察觉的范围下, 认证通过率也都比较高. 一般地, 当信噪比低于 30dB, 噪声可以被人耳明显感知. 因此, 以上结果可以说明, 对于典型的通信信道噪声, 本文的感知认证算法具有良好的鲁棒性. 从表 4 则可以看出, 普通的 MFCC 提取参数不适用于鲁棒性签名认证, 本文的算法设计, 特别是时频域的冗余消除等感知处理, 对鲁棒性有很大的改进.

4.2 唯一性分析

接着, 对 300 个不同的语音样本进行唯一性测试. 实验证明, 对于不同音频(相同说话人与不同说话人, 相同语句与不同语句), 认证检测率都为 100%(即通过率为 0%), 改变率在 80%以上. 对于原始音频与篡改后的音频(进行了添加、删除、替换等操作, 或者加入明显的噪声), 认证检测率也都为 100%, 改变率在 20%以上. 且根据前文关于有限域 hash 扩散性的分析, 平均变化比特数和每比特的平均变化概率都接近理想状态, 扩散性能很好. 因而充分证明了该感知认证算法的唯一性.

另外对改变率进行分析. 从表 2 和表 3 可得, 不易被人耳感知的干扰(所有的脉冲干扰, 信噪比大于 30dB 的高斯白噪声干扰)发生时, 即使个别情况下认证失败, 摘要改变率也基本低于 5%. 而所有的可感知的篡改(如添加, 删除, 修改语音片断, 以及可被感知的噪声), 改变率都在 20%以上. 因此除去本文中语音签名的应用环境. 本文认证算法的前半部分——计算感知摘要还可以单独提出, 根据改变率来精确区分合法和非法修改, 用于数字水印的嵌入、数据库搜寻等.

4.3 效率分析

下面对本文算法与其他常见算法的效率进行比较. 首先分析公钥签名部分. 表 5 分别在 Windows XP 和 Linux 操作系统下, 使用 VC. net2003 和 gcc 2.96 的编译环境对本文算法和其他几种常见算法进行验证. 表中的数字代表时钟数. 其中, RSA-PSS 是一种基于 RSA 的签名算法, ECDSA-GF 是一种基于椭圆曲线算法 ECC 的签名算法. Rainbow 栏为本文的实验数据, 其余算法数据来自于 NESSIE 评估结果. 从表 5 可以看出, 相比传统的 RSA、ECC, Rainbow 签名算法在签名和验证速度上有了很大的提高.

表 5 不同算法的运算速度 M

算法	Pentium4 2.8G Windows XP		AMD 4000+Linux	
	签名	验证	签名	验证
RSA-PSS	82	1.59	53	1.17
ECDSA-GF(2^{163})	5.9	7.8	4.5	6.1
Rainbow	10.3	0.79	8.2	0.57

其次分析摘要部分. 目前尚无其他用于数字签名的感知摘要算法的报道, 无法直接比较. 至于本文第 1 节中提到的用于数据库搜索的感知算法同样转换到频域进行, 且涉及到后期匹配, 特征抽取通常比

本算法更为复杂. 而同样用于感知认证的数字水印算法较杂, 各种性能相差较大, 不易比较. 仅就同样使用掩蔽效应的水印算法[17] 相比, 本文算法直接从处理过的频域参数中提取特征用于签名, 而不需像水印的计算那样再恢复成时域信号, 运算复杂度有一定降低.

总体而言, 本文感知认证算法的效率无论在感知摘要还是公钥签名上, 都要优于相应的传统算法.

5　结论

本文针对语音通信中的内容认证和身份认证需求, 提出了一种基于感知域的鲁棒性语音认证算法, 将语音感知特性与签名算法相结合. 首先分析语音的时频域掩蔽效应, 去除不可感知的冗余信息; 然后进行非线性滤波提取 Mel 参数并量化; 最后用嵌入有限域摘要的改进 Rainbow 算法进行签名. 对该认证算法的鲁棒性和唯一性进行实验和分析, 效果良好.

参考文献

[1] Cox P J, Miller M L, Bloom J A. Digital Watermarking[M]. New York: Morgan Kaufmann. 2001. 225-230.

[2] Zhu B B, Swanson M D, Tewfik A H. When seeing isn't believing-multimedia authentication technologies[J]. IEEE signal Processing Magazine, 2004, 21(2): 40-49.

[3] Kaller T, Haitsma J, Oostveen J. Robust audio hashing for content identification[C] //Content Based Multimedia Indexing 2001. Brescia, Italy: IEEE, 2001.

[4] Mihcak M K, Venkatesan R. A perceptual audio hashing algorithm a tool for robust audio identification and information hiding[C] //Information Hiding. Berlin/Heidelberg: Springer Pittsburgh, 2001: 51-65.

[5] Burges C J, Patt J C, Jana S. Distortion discriminant analysis for audio fingerprinting[J]. IEEE Trans. Speech Audio Processing, 2003, 11(3): 165-174.

[6] Sukittanon S, Atlas L E. Modulation frequency feature for audio fingerprinting[C] //International Conference Acoustics, Speech, Signal Processing. Orlando, Fla, USA: IEEE, 2002: 1773-1776.

[7] Foote J T. Content-based retrieval of music and audio[C] //Kuo C C J, *et al*. Multimedia Storage and Archiving Systems Ⅱ, Proc of SPIE, 1997, 3229: 138-147.

[8] 韩纪庆, 冯　淘, 郑贵滨, 等. 音频信息处理技术[M]. 北京: 清华大学出版社, 2007.

[9] 丁爱明. 基于 MFCC 和 GMM 的说话人识别系统研究[D]. 南京: 河海大学, 2006.

[10] Peter W S. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer[J]. SIAM Review, 1999, 41(2): 303-332.

[11] Ding J, Gower J E, Schmidt D S. Multivariate Public Key Cryptosystems[M]. New York: Springer-Verlag, 2006: 2-15.

[12] Jesteadt W, Bacon S P, Lehman J R. Forward masking as a function of frequency, masker lever, and signal delay[J]. Journal of Acoustic Society of America, 1982, 71(4): 950-962.

[13] Kohonen T, Barna G, Chrisley R. Statistical pattern recognition with neural networks: benchmarking studies[C] //Neural Networks. California: IEEE, 1988: 61-68.

[14] Ding J, Schmidt D S. Rainbow, a new multivariate polynomial signature scheme[C] //Applied Cryptography and Network Security. Berlin/Heidelberg: Springer, 2005: 164-175.

[15] Xiang Q, Liu Z. Simplest accomplishment of arithmetic on Galois fields[J]. Journal of University of Electronic Science and Technology of China, 2000, 29(1): 5-8(in Chinese).
向　茜, 刘　钊. 伽华罗域上代数运算的最简实现[J]. 电子科技大学学报, 2000, 29(1): 5-8.

[16] Özer H, Sankur B, Memon N. Perceptual audio hashing functions[C] //EURASIP Journal on Applied Signal Processing. New York: Hindawi, 2005: 1780-1793.

[17] Quan X M, Zhang H B. Statistical audio watermarking algorithm based on perceptual ayalysis[C] //Proceedings of the 5th ACM Workshop on Digital Rights Management. Alexandria, VA, USA: ACM, 2005: 112-118.

A robust speech authentication algorithm based on perceptual characteristics

GU Jin GUO Li ZHENG Dong-Fei
(*Department of Electronics Science and Technology, University of Science and Technology of China, Hefei* 230027, *China*)

Abstract At present, there is a growing need for multimedia information authentication, but studies on speech authentication are rare. Based on perceptual characteristics, a robust speech authentication algorithm is proposed to combine the psychoacoustic properties with signature algorithm. It meets the requirements of entity authentication and content authentication, and resists the channel noise as well. Analyzing the perceptual properties, such as masking effect and non-linear effect, speech redundancy in both temporal field and frequency field is eliminated, and perceptual parameters are extracted. Then an improved Rainbow algorithm is used to sign the extracted data. Experiment results demonstrate good robustness and uniqueness of the algorithm when applied to the robust authentication of audio communication.

Key words perceptual, robustness, speech authentication