音频数据加密的二值音频可听密码方案

张选平[1]，王旭[2]，邵利平[3]，赵仲孟[1]
(1. 西安交通大学电子与信息工程学院，710049，西安；2. 中航工业洛阳电光设备研究所，471000，河南洛阳；
3. 陕西师范大学计算机学院，710062，西安)

摘要：数字音频文件数据量大、相关性强、冗余度高，传统加密算法难以满足其实时安全性要求. 针对此问题，根据人耳听觉特性提出了一种基于二值音频的可听密码方案. 该方案借鉴图像半色调技术对秘密音频数据进行二值化处理，以降低音频信息的数据量；结合可视密码技术的(k,n)门限方案的基本阵对二值数字音频进行分存和加密，从而破坏了原始音频数据之间的相关性. 在解密时只需同步播放任意k份秘密音频就可直接通过人耳听觉系统解密，少于k份的任意秘密音频同步播放不会暴露原始音频的任何信息. 实验结果表明，加密后的音频具有良好的随机性和安全性. 与原始秘密音频相比，恢复后的秘密语音虽有一些失真，但其语音内容易于辨识、可懂.

关键词：数字音频；人耳听觉系统；(k,n)门限；可听密码
中图分类号：TP309.7 **文献标志码**：A **文章编号**：0253-987X(2012)08-0027-06

An Audio Cryptography Scheme for Binary Audio

ZHANG Xuanping[1], WANG Xu[2], SHAO Liping[3], ZHAO Zhongmeng[1]
(1. School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. Luoyang Institute of Electro-Optical Equipment of AVIC, Luoyang, Henan 471000, China;
3. School of Computer Science, Shaanxi Normal University, Xi'an 710062, China)

Abstract: The characteristics of digital audio such as a large quantity of data, strong correlation and high redundancy, make it difficult to meet real-time security to encrypt digital audio by using conventional encryption algorithms. An audio cryptography scheme based on binary audio is proposed by using the auditory properties of human being. In order to reduce the amount of audio data, the secret audio is transformed into binary audio by halftone as in the image process. The basic array of (k,n) threshold in visual cryptography scheme is used to share the binary audio so that correlations in audio data is removed. The secret audio information can be directly decrypted through human auditory system by synchronously playing any k shares of the binary audio, but synchronously playing any $m(m<k)$ shares does not disclose any information. Experiments show that the encrypted audio by the proposed scheme has high randomicity and security. Comparisons with original secret audio show that although the decrypted secret voice has some distortion, its content is still natural and comprehensible.

Keywords: digital audio; human auditory system; (k,n) threshold; audio cryptography

数字音频加密是保护音频信息安全的重要手段，然而音频文件数据量大、相关性强、冗余度高，传统加密算法很难满足其实时性的要求[1]. 针对音频信息加密问题，研究者提出了包括选择性加密[2]、语音信息隐藏[3-4]、秘密共享[5-8]等多种语音信息保护方法. 这些方案都不同程度地利用了人耳听觉系统

收稿日期：2012-01-07. 作者简介：张选平(1963－)，男，副教授. 基金项目：国家自然科学基金青年科学基金资助项目(61100239)；教育部高等学校博士学科点专项科研基金资助项目(61100239).
网络出版时间：2012-05-15 网络出版地址：http://www.cnki.net/kcms/detail/61.1069.T.20120515.1824.002.html

的某些感知特性[9]，如人耳对音频相位的不敏感性和听觉掩蔽效应等.

可听密码是一种不需要计算机参与，仅依靠人耳听觉系统就能解密语音消息的加密方法. 它首先将原始音频 S 分成 n 份不同的秘密音频，当同时播放其中若干份秘密音频时就可通过人耳直接恢复出原始秘密音频 S.

结合半色调技术中的 Floyd 误差扩散算法[10]，本文给出了一种用于数字音频保护的可听密码方案，首先将高采样数字音频转换为二值半色调音频，然后利用声波的叠加特性和人耳听觉的平均效应，结合可视密码技术的(k,n)门限方案[11]的基本阵对二值音频进行加密. 当同步播放加密后的任意 k 份分存音频，即可直接通过人耳听觉系统解密，少于 k 份的任意秘密音频同步播放不会暴露原始音频的任何信息. 实验结果表明，加密后的音频具有良好的随机性和安全性. 与原始音频相比，恢复后的秘密语音虽有一些失真，但其语音易于辨识、可懂.

1 数字音频的二值化处理

在秘密语音通信环境下，通信双方关注的是通信内容，一般并不在意语音的音质，因此可在保证语音信息可理解的前提下，对音频信号进行二值半色调处理. 一方面可极大地减少音频信息的数据量，另一方面可借鉴二值图像的可视密码技术来方便地实现可听密码方案. 本文将 Floyd 等[10]提出的误差扩散算法应用于数字音频的二值化处理. 首先将连续的 WAV 音频中第 1 个数据与阈值比较，得到一个二值输出，然后将输入数据与输出数据之间的差值按照一定的映射方式扩散到后续未处理的数据序列，依次处理后续数据. 误差扩散算法原理如图 1 所示.

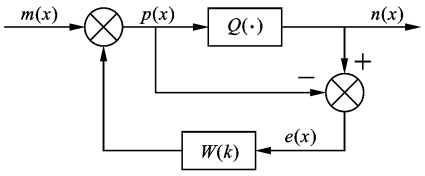


图 1 误差扩散算法原理图

图 1 中，$m(x)$、$p(x)$、$n(x)$和 $e(x)$分别表示 x 处的采样数据、上一步误差扩散后的输入数据、二值化的输出数据以及量化误差；$Q(\cdot)$是阈值量化器；$W(k)$是误差扩散滤波器. 计算过程如下

$$p(x)=m(x)+\sum_{k=1}^{n}w(k)\cdot e(x-k) \tag{1}$$

$$n(x)=Q(p(x)) \tag{2}$$

$$e(x)=n(x)-p(x) \tag{3}$$

这里 $Q(\cdot)$的阈值分别取音频正半轴和负半轴振幅的平均值，$W(k)$要求满足$\sum_{k=1}^{n}w(k)=1$. 在音频序列误差扩散中，经反复实验设定的传递参数为 $w(1)=7/16$，$w(2)=5/16$，$w(3)=3/16$，$w(4)=1/16$，其误差扩散过程如图 2 所示. 其中，序号 0 处为当前处理单元，通过阈值滤波器，当前波形数据的量化误差按上述参数权重分配依次传递给未处理的后续音频采样数据.

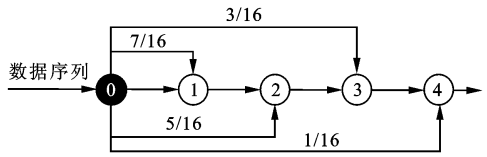


图 2 音频数据量化误差传递过程

经上述二值化过程处理的音频，其连续时域波形被转换成只有二值方波，尽管音频波形发生了变化且引入了噪声，但仍能通过人耳辨识出原始音频.

2 (k,n)门限可听密码方案

人耳听觉系统具有低通滤波特性，人们所听到的声音是一系列音频的局部平均效果. 另外，声波具有叠加特性，当多个数字音频同步播放时，人们听到的是这些音频叠加后的效果，具体表现为多段音频在同一时刻的采样数值之和. 因此，在对音频数据二值化的基础上，可借鉴可视密码技术的(k,n)门限方案，将二值秘密音频分存为 n 份子秘密音频，使得在同步播放其中任意 $k(k\leqslant n)$份子秘密音频时，通过声波相互叠加，可直接被人耳辨识和解密.

本方案中采用 Shamir 可视密码(k,n)门限方案[11]，使用 2 个 $n\times m$ 基本矩阵 $\boldsymbol{B}_0$ 和 $\boldsymbol{B}_1$ 对二值音频的每一采样值进行加密. 基本矩阵 $\boldsymbol{B}_0$ 和 $\boldsymbol{B}_1$ 的构造算法采用 Droste 基本阵构造算法[12]. 构造的基本阵 $\boldsymbol{B}_0$ 和 $\boldsymbol{B}_1$ 满足 3 个基本条件：①$\boldsymbol{B}_1$ 阵中任意 k 行进行或运算后，海明权重至少为 d；②$\boldsymbol{B}_0$ 阵中任意 k 行进行或运算后，海明权重最多为 $d-\alpha m$，其中 m 为扩展度，α 为对比度，$\alpha\geqslant 1/m$，d 为权重，$d\in\{1,\cdots,m\}$；③$\boldsymbol{B}_0$ 和 $\boldsymbol{B}_1$ 中任意 $q(q<k)$行在进行或运算的叠加过程中，0 和 1 出现次数均相同. 以下是构造的$(3,5)$门限基本阵 $\boldsymbol{B}_0$ 和 $\boldsymbol{B}_1$

$$\boldsymbol{B}_0=\begin{pmatrix}0&1&1&1&1&0&0&0\\0&1&1&1&0&1&0&0\\0&1&1&0&1&1&0&0\\0&1&0&1&1&1&0&0\\0&0&1&1&1&1&0&0\end{pmatrix};$$

$$\boldsymbol{B}_1=\begin{pmatrix}1&0&0&0&0&1&1&1\\0&1&0&0&0&1&1&1\\0&0&1&0&0&1&1&1\\0&0&0&1&0&1&1&1\\0&0&0&0&1&1&1&1\end{pmatrix}\qquad(4)$$

为提高安全性，在对每个采样值进行加密前，需对基本阵分别进行行列同步随机置乱，使得加密后的数据中 0 和 1 能够随机排列. 加密过程如下.

步骤 1　用 Droste 基本阵构造算法构建 $\boldsymbol{B}_0$ 和 $\boldsymbol{B}_1$.

步骤 2　将输入的连续多值波形数字音频转化成二值音频.

步骤 3　对二值数字音频序列中的每个值 s_i 进行如下操作：①若 $s_i=0$，则对 $\boldsymbol{B}_0$ 阵分别进行行列同步置乱，形成 n 个分存值；②若 $s_i=1$，则对 $\boldsymbol{B}_1$ 阵分别进行行列同步置乱，形成 n 个分存值.

步骤 4　输出 n 份分存加密数字音频，结束.

一个长度为 l 的明文序列经过加密后变成 n 个长度为 $m\times l$ 的密文序列. 下面以 Windows XP 操作系统的启动音频 start. wav 中 74 000～75 000 这 1 000 个采样点来说明加密过程，如图 3 所示. 首先，对其进行二值化，生成只包含 0 和 1 的二值序列，二值化后的数字音频波形如图 4 所示.

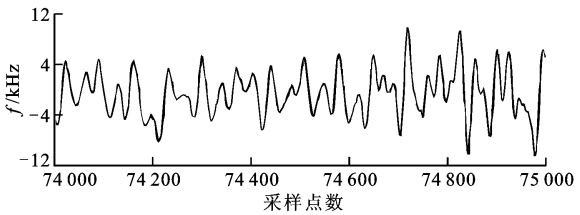


图 3　原始音频

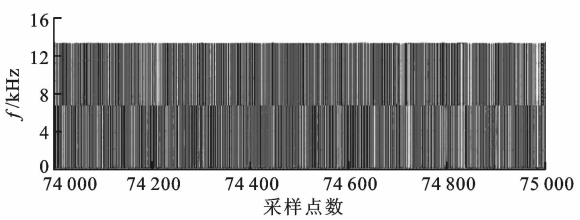
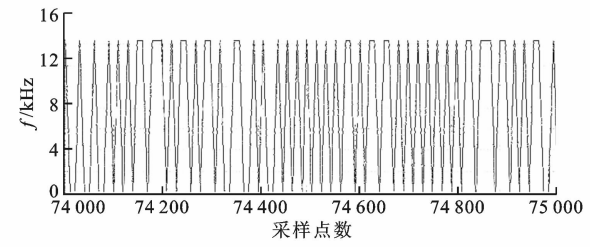
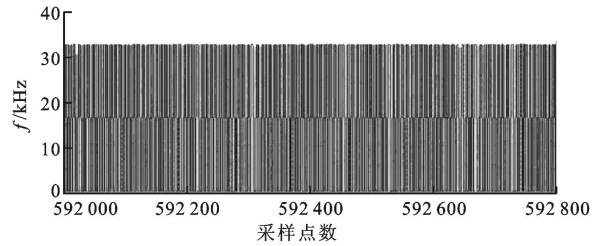


图 4　二值化后的音频

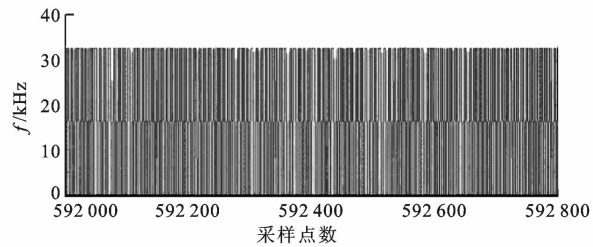
二值化后的音频只包含 0 和 1 两种音频信号，使用 $\boldsymbol{B}_0$ 和 $\boldsymbol{B}_1$ 加密二值音频序列. 以加密前 74 000～74 100的 100 个采样点为例(如图 5a 所示)，加密后实际对应为 592 000～592 800 的 800 个点，选取加密后的 5 个分存音频中的任意 3 个分存音频，如图 5b～5d 所示，对这 3 个音频同步 8 倍速播放，即可通过人耳直接解密，获取秘密语音信息.



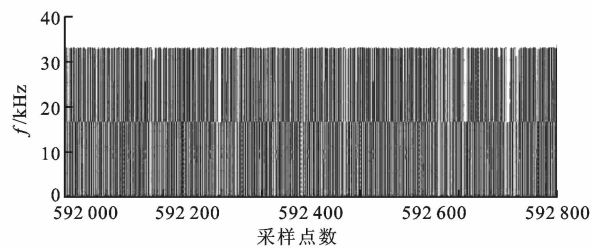
(a)2 音频中 100 个点加密前波形



(b)加密分存音频 1 的波形



(c)加密分存音频 2 的波形



(d)加密分存音频 3 的波形

图 5　(3,5)门限可听密码方案加密前后音频波形图

3　实验结果与分析

在仿真实验中，分类选取 12 个采样率为 44 100 Hz、采样位数为 16 的 WAV 音频文件，其中包含了男声和女声语音各 1 段、5 类典型音乐信号(流行音

乐、蓝调音乐、乡村音乐、民族音乐和古典音乐)各2段.使用本文所提方案对这12段音频文件采用(3,5)门限进行加密和解密,并对结果进行听觉评测和安全性分析.

3.1 听觉效果评测

使用主观评价方法MOS(Mean Opinion Score)[13]和主观评价等级SG(Subjective Grade)[14]对本文方案进行评测.MOS和SG评价等级见表1.参与测试的共20人,男性女性各占一半,其中一半具有声乐专业背景,在隔音环境中对这12段音频的二值化音频、加密音频和解密音频打分并求出平均分,实验结果如表2所示.

表1 MOS和SG评价等级

分值	MOS听觉感受	SG听觉感知质量
5.0	感觉不出语言有任何失真,清澈透明、自然.	不可感知
4.0	语音存在可感知的失真,背景噪声对语音有轻微影响,但语音仍清晰、自然、可懂.	可听见,但不烦人
3.0	明显感觉语音失真,背景噪声偶尔干扰个别字音,清晰度和辨别力下降,但仍自然、可懂.	轻微烦人
2.0	语音失真严重,感觉粗糙,背景噪声令人讨厌,个别字听不清而且有必要重复.	烦人
1.0	语音失真到了不可接受的程度,仅听得出个别字和词,甚至什么都听不懂.	非常烦人

实验结果表明,纯人类语音二值化后,因语言间断性使得噪声比较明显,个别字存在失真但语音仍清晰、自然、可懂;古典音乐具有丰富的细节,二值化后加入了噪声,使得细节流失,但主旋律依然清晰自然;蓝调、乡村和民族音乐二值化后,噪声明显,但不烦人,语音依然清晰、自然、可懂.加密后的所有音频都听不到任何语音信息,噪声极为烦人.加密后的所有分存音频单独8倍速播放时,都听不到任何语音信息,噪声极为烦人.解密后的音频具有一定的失真,噪音轻微烦人,清晰度和辨别力下降,但仍清晰、自然、可懂.

表2 音频二值化与加解密后的MOS和SG测试结果

WAV文件	未加密二值音频		加密二值音频		解密二值音频	
	MOS	SG	MOS	SG	MOS	SG
男声	3.2	3.9	≪1	≪1	2.9	3.0
女声	3.4	3.8	≪1	≪1	3.1	3.1
流行1	4.5	4.6	≪1	≪1	2.9	2.8
流行2	4.6	4.6	≪1	≪1	3.0	3.1
蓝调1	4.2	3.9	≪1	≪1	2.9	3.1
蓝调2	4.2	4.0	≪1	≪1	3.0	3.1
乡村1	4.5	4.2	≪1	≪1	3.0	3.1
乡村2	4.4	4.0	≪1	≪1	3.2	3.1
民族1	4.2	3.9	≪1	≪1	3.3	3.0
民族2	4.1	3.8	≪1	≪1	3.2	3.0
古典1	3.9	4.0	≪1	≪1	2.9	3.0
古典2	3.7	4.1	≪1	≪1	2.8	2.9

3.2 相关性分析

音频文件相邻数据具有很高的相关性,为防止统计攻击,加密算法必须破坏相邻数据的相关性.对上述12段音频进行相关性测试,在每段音频中随机选取2 000对相邻音频数据(x_i,x_{i+1}),分别计算两者的相关系数,实验结果如表3所示.

表3 原始音频数据和加密音频数据的相邻数据相关系数

WAV文件	相邻数据相关系数	
	原始数据	加密数据
男声	0.971 8	0.011
女声	0.992 6	0.023
流行1	0.986 1	0.026
流行2	0.984 4	0.022
蓝调1	0.992 4	0.013
蓝调2	0.991 9	0.033
乡村1	0.992 7	0.022
乡村2	0.993 5	0.420
民族1	0.989 7	0.033
民族2	0.990 3	0.066
古典1	0.992 8	0.037
古典2	0.992 6	0.032

由表 3 可以看出，加密前原始音频文件相邻数据的相关系数都在 0.9 以上，而加密后的相关系数基本都在 0.05 以下，因此加密过程破坏了原始音频数据的相关性，具有较高的安全性.

3.3　性能比较分析

目前可听密码方案大多数只实现了$(2,n)$或(k,k)方案，而(k,n)方案并不多见. 这里将本文方案与文献[7]中的(k,n)方案从安全性、解密的复杂性和恢复的音质质量 3 方面进行对比分析，结果如表 4 所示.

表 4　安全性、复杂性和解密质量对比

方案	安全性	解密复杂性	解密质量
本文	高	简单	较差
文献[7]	一般	较复杂	较高

从安全性方面分析，文献[7]方案在分存原始音频时需公开所使用的矩阵，同时各个分存信息之间存在相关性，而本文方案的每个分存信息是从基本阵中随机选取，并且基本阵不需公开，因此有较高的安全性.

从解密复杂性方面分析，本文方案只需同步播放任意 k 个子音频就可由人耳直接解密，而在文献[7]方案中，需根据参与解密的子音频为每一个音频计算一个放大因子，按放大因子放大子音频的音量，通过同步播放而解密，因此解密过程稍显复杂.

从解密的音频质量分析，由于本文方案需要对原始音频进行二值化处理，因此使得原始音频发生了一定的变化，恢复后的音频质量稍差，而文献[7]方案中直接对原始音频进行分存，所以恢复后的音频与原始音频基本接近，音质较好.

4　结　论

本文给出了一种基于人耳听觉特性的(k,n)门限可听密码安全方案. 结合图像半色调技术中的 Floyd 误差扩散算法，完成了数字音频的二值半色调处理工作. 在二值化数字音频的基础上，利用声波的叠加特性和人耳听觉的低通滤波效应，结合可视密码技术的(k,n)基本阵，构造了可听密码(k,n)门限方案. 对该方案进行了仿真实验和听觉主观评测，实验结果表明，经可听密码(k,n)门限方案加密后的数字音频，在低于门限情况下独立或同时加速播放时，听不到任何有意义的语音信息，噪声极为烦人，达到了加密效果. 任意大于或等于门限值的多个分存音频同时加速播放时，可恢复出秘密语音信息，虽然背景噪声轻微烦人，但仍自然、可懂. 本文方案可通过人类听觉系统直接对同步加速播放的多个加密数字音频进行解密，易于实现，具有高安全性、低代价和简单易用等特点.

参考文献：

[1] 张选平，王旭，邵利平. 基于可逆元胞自动机的二值音频加密算法[J]. 高技术通讯，2011，21(2)：117-123.
ZHANG Xuanping，WANG Xu，SHAO Liping. Binary audio encryption based on reversible cellular automata [J]. High Technology Letters，2011，21(2)：117-123.

[2] WANG Honggang，HEMPEL M，PENG Dongming，et al. Index-based selective audio encryption for wireless multimedia sensor networks [J]. IEEE Transactions on Multimedia，2010，12(3)：215-220.

[3] 谭良，吴波，刘震，等. 一种基于混沌和小波变换的大容量音频信息隐藏算法[J]. 电子学报，2010，38(8)：1812-1818.
TAN Liang，WU Bo，LIU Zhen，et al. An audio information hiding algorithm with high-capacity which based on chaotic and wavelet transform [J]. Acta Electronica Sinica，2010，38(8)：1812-1818.

[4] KEKRE H B，ATHAWALE A，RAO S，et al. Information hiding in audio signals [J]. International Journal of Computer Applications，2010，8(9)：14-19.

[5] NASKAR P K，KHAN H N，ROY U，et al. Shared cryptography with embedded session key for secret audio [J]. International Journal of Computer Applications，2011，26(8)：1-9.

[6] HASSAN N F，HASSAN R F，ALI A E. Encrypting audio data hiding by visual secret sharing [J]. Journal of Engineering and Technology，2009，27(13)：2352-2364.

[7] EHDAIE M，EGHLIDOS T，M R AREF. A novel secret sharing scheme from audio perspective[C]// IEEE International Symposium on Telecommunications. Piscataway，NJ，USA：IEEE，2008：13-18.

[8] SOCEK D，MAGLIVERAS S S. General access structures in audio cryptography [C]// Proceedings of the IEEE International Conference on Electro Information Technology. Piscataway，NJ，USA：IEEE，2005：1-6.

[9] ZAMANI M，AHMAD R B，MANAF A B A，et al. An approach to improve the robustness of substitution techniques of audio steganography [C]// Proceedings

of the IEEE International Conference on Computer Science and Information Technology. Piscataway, NJ, USA: IEEE, 2009: 5-9.

[10] FLOYD R W, STEINBERG L. An adaptive algorithm for spatial greyscale [J]. Journal of the Society for Information Display, 1976, 17(2):75-77.

[11] NAOR M, SHAMIR A. Visual cryptography[C]// Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques. Berlin, Germany: Springer, 1995:1-12.

[12] DROSTE S. New results on visual cryptography [C] // Proceedings of the 16th Annual International Cryptology Conference on Advances in Cryptology. Berlin, Germany: Springer-Verlag, 1996: 401-415.

[13] 温泉，王树勋，年桂君. DCT 域音频水印：水印算法和不可感知性测度[J]. 电子学报，2007，35(9)：1702-1705.
WEN Quan, WANG Shuxun, NIAN Guijun. Audio watermarking in DCT domain: algorithm and measurement of imperceptibility [J]. Acta Electronica Sinica, 2007, 35(9):1702-1705.

[14] 全笑梅，张鸿宾. 基于小波包域听觉感知分析的统计音频水印算法[J]. 电子学报，2007，35(4)：673-678.
QUAN Xiaomei, ZHANG Hongbin. Statistical audio watermarking algorithm based on auditory analysis in wavelet packet domain [J]. Acta Electronica Sinica, 2007, 35(4):673-678.

（编辑 武红江）