УДК 681.5.015: 004.056.57

Маракова И.И., Комендантов И.И.
Marakova I.I., Komendantov I.I.

ИССЛЕДОВАНИЕ СИСТЕМ С ЦИФРОВЫМИ ВОДЯНЫМИ ЗНАКАМИ ПРИ ОСНОВНОМ ПОКРЫВАЮЩЕМ СООБЩЕНИИ В ВИДЕ АУДИО-СИГНАЛА

WATERMARKING SYSTEM WITH AUDIO COVER MESSAGES ANALYSIS

Аннотация. Рассмотрены системы с цифровыми водяными знаками при основном покрывающем сообщении в виде аудио-сигнала при атаке аддитивным шумом. Аналитически полученные результаты были проверены с помощью имитационного моделирования. Проведена оптимизация систем с цифровыми водяными знаками в условиях атаки временной десинхронизации.

Summary. We consider private watermarking system at use as the cover message of audio signals with additive noise attack Analytical research results are checked by imitation modeling. The watermarking system optimization in the condition of time resynchronization attack are made.

В настоящее время интенсивное развитие информационных технологий стимулирует поиск и разработку новых подходов обеспечения информационной безопасности. Возникает целый ряд ситуаций в компьютеризированных системах хранения, обработки, приёма, передачи информации, когда применение криптографических методов не решает возникающих проблем. Например, цифровая подпись может быть без труда удалена из электронного документа; шифрование документов во многих странах запрещено на законодательном уровне; к процедурам идентификации нередко предъявляется требование скрытности и т.д.

Одним из выходов в подобных ситуациях является использование методов сокрытия информации, в частности, цифровой стеганографии, которая изучает методы и средства сокрытия факта присутствия секретного сообщения в некотором другом общедоступном основном покрывающем сообщении (ОПС), т.е. формирования стегасообщения. Предполагается, что легальные пользователи системы обмена обладают некоторыми секретными данными (ключами), которые позволяют обнаруживать и читать эти секретные сообщения. Задача *атакующего* – нелегального пользователя – состоит в обнаружении факта присутствия секретного сообщения (пассивная атака) и/или изменения стегасообщения (активная атака). Относящиеся к цифровой стеганографии системы с цифровыми водяными знаками (ЦВЗ) предназначены не столько для сокрытия секретной информации, сколько для передачи с ОПС некоторой дополнительной, возможно и не секретной информации, которую не возможно удалить, не ухудшив значительно надёжность восприятия сообщения. Другими словами, предназначением систем с ЦВЗ относительно решающей аналогичные задачи технологии цифровой подписи является то, что ЦВЗ не увеличивают объём ОПС, не отделимы от ОПС и подвергаются тем же преобразованиям, что и ОПС [1].

Основные отличия систем с ЦВЗ от классических систем связи заключается в следующем: помеха в системе связи, как правило, аддитивна, а преобразования стегасообщения (атакующий канал) описываются значительно более сложно, чем аддитивная помеха; критерием оптимальности дискретных систем связи является минимизация вероятностей ошибок (максимизация отношения сигнал/шум для аналоговых систем), тогда как в системах с ЦВЗ критерием является та же минимизация вероятностей ошибок детектирования и декодирования ЦВЗ, но при расплывчатом критерии сохранения надёжности восприятия стегасигнала. Надёжность восприятия, как правило, оцениваемое экспертом сходство стегасообщения и ОПС, может характеризоваться некоторым косвенным параметром, например, отношением сигнал/шум, причём при использовании в качестве ЦВЗ широкополосных псевдослучайных последовательностей сигналом является ОПС, а шумом – ЦВЗ.

Системы с ЦВЗ, предполагающие передачу одного бита информации (на выходе детектора 0 при отсутствии ЦВЗ либо 1 при наличие ЦВЗ) называются *системами с нулевым битом*. На практике большой интерес представляют системы с ВЗ, модулированные информацией. При теоретическом анализе рассмотрение системы с нулевым битом не умаляет общности исследований.

Относительный рейтинг требований к системам с ВЗ определяется спецификой их применения. Например, при слежении за несанкционированным копированием, если стегасообщение передаётся по аналоговому каналу, то ВЗ должны быть невосприимчивыми к искажениям, обусловленным данным каналом. Однако, если есть возможность достоверно исключить модификацию ВЗ между погружением и детектированием, то данное требование излишне.

Системы с ЦВЗ с аудио или речевыми ОПС основаны на несовершенствах человеческой слуховой системы, т.е. использовании перцепционной модели (ПМ) [2, 3]. Однако в данных работах

не проводилось исследование зависимости численной оценки эффективности от длины ЦВЗ при фиксированных параметрах системы, а именно, параметров надежности восприятия в виде отношения сигнал/шум при формировании стегасообщения и преобразований канала атакующего, интенсивности ЦВЗ, пропускной способности. Представляется интересным проверка результатов аналитических исследований при ОПС в виде аудио сигналов, а также оптимизация алгоритмов погружения и детектирования ЦВЗ относительно атак временной десинхронизацией.

Целью данной работы является решение указанных выше задач.

При имитационном моделировании можно непосредственно оценивать надежность восприятия аудио и речевых сообщений, а не косвенно, т.е. на основе какого-либо параметра (среднеквадратического отклонения, коэффициентов корреляции ОПС и стегаобраза и т.д.), не учитывающего ПМ. Рассматриваемые секретные системы с ЦВЗ используют в качестве секретного ключа (СК) собственно ЦВЗ, являющееся широкополосным сигналом (ШПС).

Количественной мерой эффективности являются вероятности ошибок детектирования ЦВЗ (вероятность пропуска ЦВЗ P_m, и вероятность ложного обнаружения ЦВЗ P_{fa}), аналитическая оценка которых как функции основных параметров системы с ЦВЗ может сравниваться с результатами моделирования методом Монте Карло [4]. Расчет вероятностей ошибок системы с ЦВЗ в условиях воздействия аддитивного шума может рассматриваться в качестве нижней оценки эффективности. При использовании в качестве ЦВЗ широкополосных сигналов необходимо обеспечение синхронизации передающей и приемной частей. Использование многократного или избыточного погружения ЦВЗ в какой-то степени решает проблему, но при этом увеличивается вероятность атак с оцениванием ЦВЗ и коллизионных атак.

При имитационном моделировании исследовались следующие структуры секретной системы с ЦВЗ:

– неинформированное устройство погружения (УП), информированный детектор;
– неинформированное УП, неинформированный детектор;
– информированное УП, неинформированный детектор.

Информированный детектор использует при формировании решающего функционала ОПС, для неинформированного детектора ОПС является помехой. Информированное УП не столько использует знание ОПС для применения положений ПМ и обеспечения надежного восприятия стегаобраза, сколько для оптимизации процедуры детектирования.

При формировании стегосигнала аддитивным способом

$$s(n) = c(n) + w(n), \qquad (1)$$

где $c(n)$, $n \in A_N = 1,\ldots,N$ – ОПС с дискретизацией во времени,

$w(n)$, $n \in A_N = 1,\ldots,N$ – ЦВЗ, дискретная во времени последовательность $\{\pm\alpha\}$, надежность восприятия контролируется выбором параметра искажения $\eta_w = \sigma_c^2 \,/\sigma_w^2$,

где σ_c^2 – дисперсия сигнала; σ_w^2 – дисперсия ЦВЗ. Стегасигнал (1) после воздействия аддитивного шума

$$s'(n) = c(n) + w(n) + \varepsilon(n),$$

где $\varepsilon(n)$ – шум с гауссовым распределением, нулевым средним и дисперсией σ_ε^2, значение которой также ограничивается с точки зрения надежности восприятия стегосигнала величиной отношения сигнал/помеха после атаки шумом $\eta_a = \sigma_c^2 \,/(\sigma_\varepsilon^2 + \sigma_w^2)$, причём $\eta = \eta_w / \eta_\alpha = (\sigma_\varepsilon^2 + \sigma_w^2)/\sigma_w^2$.

При ОПС в виде аудио сигнала частотный диапазон составляет 10 … 22000 Гц. Стандарты сжатия речевой (RPLE-LPC, LD-CELP, VCELP, CELP и др.) и аудио (MPEG-1, 2, 3) информации основаны на минимизации перцепционной энтропии [5]. При проведении имитационного моделирования и ОПС в виде аудио-сигнала был реализован алгоритм, в котором перед погружением ЦВЗ выполняется переход в частотную область на основе дискретного Фурье преобразования (ДФП) или дискретного косинусного преобразования (ДКП) (рис. 1). При разделении ОПС на окна (фреймы) длительностью 2048 отсчетов ДФП выполняется для каждого фрейма, т.е.

$$DFT(c(n)) = c(k) = \sum_{n=0}^{N-1} c(n) \exp(-2\pi nkj / N),$$

где $k = 1, \ldots, K$, $K = 2048$, $n = 1,\ldots, N$, $N = N_{\text{фр}} = 2048$.

Перед погружением ЦВЗ в виде ШПС, распределенного по всему фрейму, определялась дисперсия данной части ОПС σ_{Fc}^2 и, если $\sigma_{Fc}^2 \le 0{,}1\sigma_c^2$, фрейм не использовался для формирования

стегосигнала, как не пригодный с точки зрения обеспечения надежности восприятия. Однако для аудио сигналов наличие таких «тихих» фреймов не превышало 1…5%.

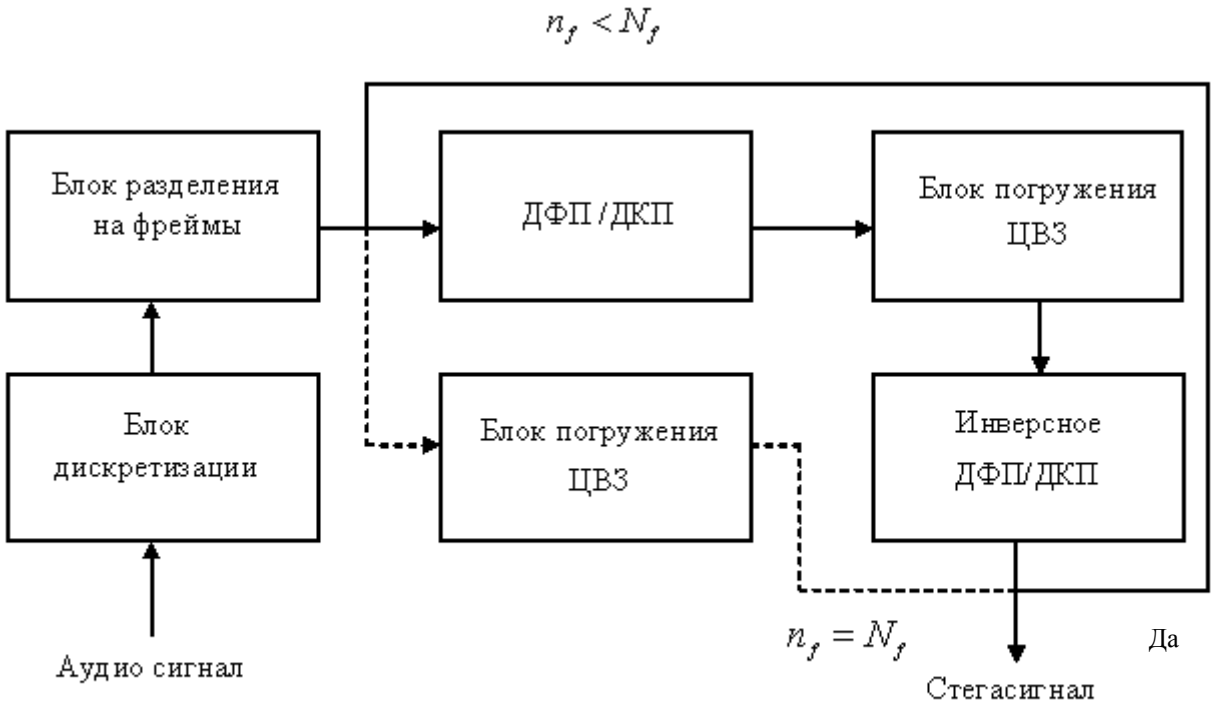

Рисунок 1 – Структурная схема алгоритма погружения ЦВЗ

При переходе в частотную область при помощи ДФП длина фрейма N_f = 2048 (0,05 с), а при ДКП, N_f = 1024 (0,023 с). Погружение ЦВЗ осуществляется в область 100 … 500 Гц для обеспечения устойчивости к преобразованиям сжатия. Амплитуда ЦВЗ выбиралась из условия $\eta_w = \sigma_c^2 \ / \sigma_w^2 = \{100...2500\}$, т.е. $\alpha = \{0{,}1\sigma_c ... 0{,}02\sigma_c\}$, при $\sigma_w = (1{,}1...2)\sigma_\varepsilon$. Если длина ЦВЗ оказывалась больше, чем количество отсчётов фрейма, то длина фрейма увеличивалась. Для сравнения с результатами аналитического моделирования при имитационном моделировании рассматривалась библиотека различных ОПС с последующим усреднением полученных оценок. Полученное расхождение не более 20% требуемой длины ЦВЗ для обеспечения $P_m = P_{fa} = 10^{-3}$ при прочих равных параметрах и в условиях воздействия аддитивного шума продемонстрировало удовлетворительное совпадение [4]. Причем, для структуры с неинформированным детектором при погружении ЦВЗ в несколько фреймов и адаптации интенсивности ЦВЗ к текущей дисперсии ОПС наблюдается некоторое улучшение эффективности системы (рис. 2). Шум атаки является гауссовым некоррелированным процессом с нулевым средним и дисперсией $\varepsilon = N(0, \sigma_\varepsilon^2)$. Детектирование осуществлялось по правилу

$$\begin{aligned} &\text{есть ЦВЗ, если } \Lambda \ge \lambda, \\ &\text{нет ЦВЗ, если } -\lambda < \Lambda < \lambda, \\ &\text{есть ЦВЗ, если } \Lambda \le -\lambda, \end{aligned}$$

где Λ – функционал линейного корреляционного детектора (ЛКД);

λ – порог ЛКД, причем

$$\lambda = \sum_{n=1}^{N_f} s(n) \cdot w(n) = \sum_{n=1}^{N_f - K} c(n) \cdot w(n) + \sum_{n=1}^{K} c(n) \cdot w(n) + \sum_{n=1}^{N_f} w(n) \cdot w(n), \qquad (2)$$

где K – число дискретов ОПС, которые дают нулевую корреляцию с ЦВЗ.

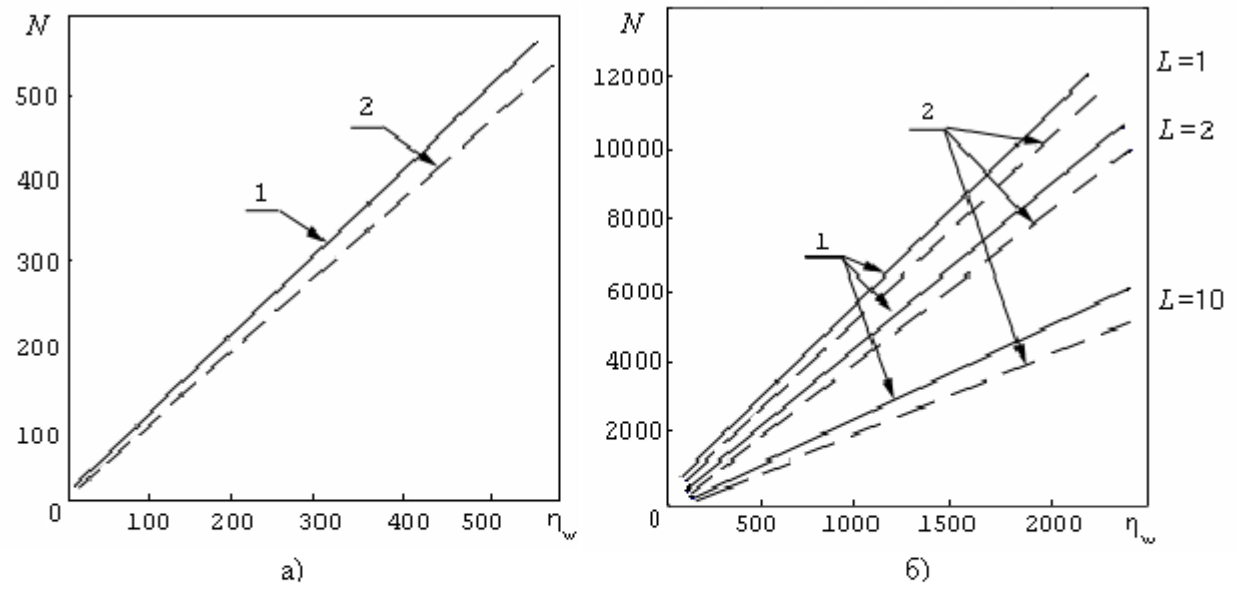

Рисунок 2 — Зависимость числа бит ЦВЗ N от величины η_W, неинформированное УП, информированный детектор (а), неинформированное УП, неинформированный детектор (б, L=1), информированное УП, неиформированный детектор (б, L=2,L=10), где L – объём кодовой книги, $P_m = P_{fa} = 10^{-3}$, η= 1.5 (1), η= 2 (2).

Для информированного ЛКД сравниваемый с порогом (2) функционал определится как

$$\Lambda = \frac{1}{N}\sum_{n=1}^{N}(s'(n) - c(n)) \cdot w(n) = \frac{1}{N}\sum_{n=1}^{N}(w(n) + \varepsilon(n)) \cdot w(n), \tag{3}$$

для неинформированного

$$\Lambda = \frac{1}{N}\sum_{n=1}^{N}s'(n) \cdot w(n) = \frac{1}{N}\sum_{n=1}^{N}(c(n) + w(n) + \varepsilon(n)) \cdot w(n). \tag{4}$$

Однако при погружении ЦВЗ в частотный диапазон 100…500 Гц и соответствующих значениях N использовались дискреты ОПС $c(n)$, которые при перцепционном кодировании, сжатии могут быть утрачены. Реализация алгоритма погружения с учетом комплексной ПМ, позволяет повысить надежность восприятия при фиксированном уровне эффективности и обеспечить устойчивость к преобразованиям канала обработки сигнала и атаки оцениванием, но весьма громоздка с вычислительной точки зрения и для атакующего более эффективными являются преобразования временной десинхронизации.

Известно, что ритмическая характеристика аудио и период голосового тона речи являются такими характеристиками, нарушение которых приводит к существенной потери надежности восприятия. Безусловно, в отдельных фреймах могут быть искажения, однако их доля, как правило, не превышает 5…10%. Используем ритмическую характеристику аудио-сигналов для организации алгоритма погружения ЦВЗ, а именно погружения в такие дискреты ОПС, удаление или изменение которых приводит к потере надёжности восприятия ниже заданного уровня. Получение ритмической характеристики или ее вероятностных мер и являются основой организации временной синхронизации УП и детектора ЦВЗ.

Если положить, что в пределах 12-13 секунд ритмическая структура аудио является неизменной, то при 2048 дискретах в одном фрейме представляется возможным оценивание периода ритма на основе анализа 250 фреймов и получения единого параметра для синхронизированного погружения и детектирования ЦВЗ (рис. 3). Таким образом, наблюдая $c(n')$,

$n' = 1,\ 2,\ ...,\ 2048F,\ n' \in A_N$, где $F = 250$ – число фреймов; $\tau_r(i) = \frac{2048 f}{R}\Delta t \pm \Delta\tau_r$ – длительность ритмов, где $f = 1,... ,F$ – число фреймов с идентичной ритмической структурой; Δt = 0,025 10^{-3}с – интервал дискретизации; $r = 1,... R$ – число ритмов в F фреймах, $i = 1,\ 2,... I,\ I = RF$; $\Delta\tau_r$ – случайное отклонение. Оценке подлежит среднее и дисперсия вероятности $P(m_\tau, \sigma_\tau^2 / \tau_1,\ \tau_2,... \tau_I)$

$$P(C/\tau) = \prod_{i \in A_I} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(c(n) - c(n - i\frac{\tau_i}{\Delta t}))^2}{2\sigma^2} \right)^2. \quad (5)$$

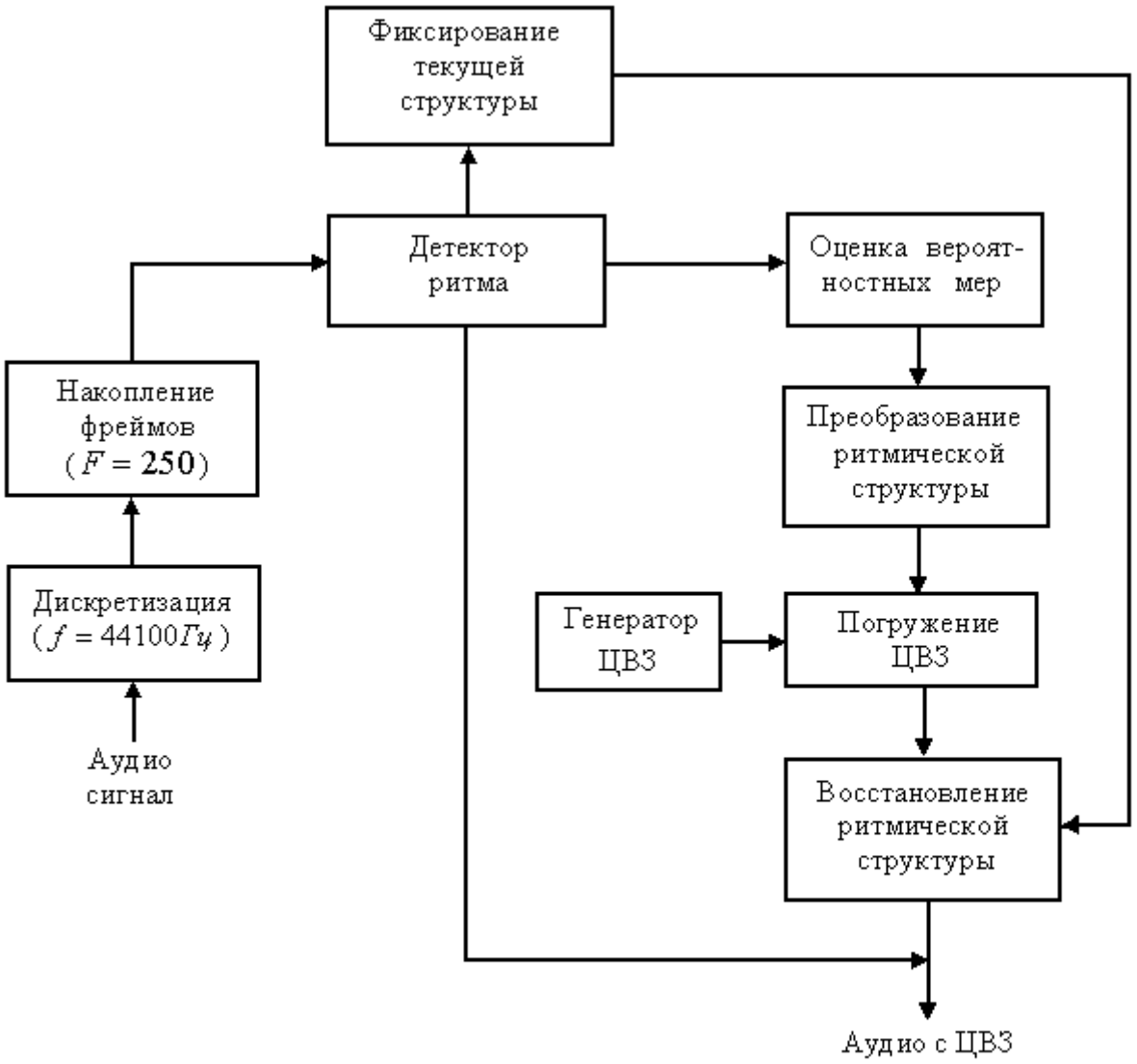

Рисунок 3 — Структурная схема алгоритма оценки ритмической структуры ОПС (вероятностных мер ритма ОПС)

Для оценки средней длительности ритма по некоторой выборке значений используется алгоритм оценки и оптимизации (ОО)[6]. Реализация оценки включает два шага:

- моделирование на заданном объёме статистики;
- оценка параметров при помощи метода наибольшего правдоподобия.

Метод наибольшего правдоподобия состоит в том, что в качестве оценки параметра из выборки $(x_1,\ x_2,...\ x_n)$ берётся такое x, при котором оцениваемая функция L достигает своего максимума, т.е. $\frac{\partial L}{\partial x} = 0$ или в более удобной с вычислительной точки зрения форме

$$\frac{1}{L}\frac{\partial L}{\partial x} = \frac{\partial \ln L}{\partial x} = 0\ . \tag{6}$$

Если плотность распределения оцениваемой функции зависит от нескольких параметров, то наиболее правдоподобная оценка системы параметров является решением системы уравнений

$$\frac{\partial \ln L}{\partial x^i} = 0,\ \ x^i \in A_x,\ \ i = 1,...\ I.$$

Предполагая, что случайные значения длительности ритма $\tau \in A_\tau$ характеризуются гауссовой плотностью распределения со средним m_τ и дисперсией σ_τ^2

$$L(\tau_1,\ \tau_2,...,\ \tau_{RF};\ m_\tau,\ \sigma_\tau^2) = \left(\frac{1}{\sqrt{2\pi\sigma_\tau^2}}\right)^n \exp\left[-\frac{1}{2\sigma_\tau^2}\sum_{r=1}^{RF}(\tau_r - m_\tau)^2\right]$$

получим

$$\ln L = -\frac{RF}{2}\ln 2\pi - \frac{RF}{2}\ln\sigma_\tau^2 - \frac{1}{2\sigma_\tau^2}\sum_{r=1}^{RF}(\tau_r - m_\tau)^2 \tag{7}$$

и согласно (6) уравнения для определение вероятностных мер ритмической структуры определятся как

$$\frac{\partial \ln L}{\partial m_\tau} = \frac{1}{2\sigma_\tau^2}\sum_{r=1}^{RF}(\tau_r - m_\tau)^2 = 0,$$

$$\frac{\partial \ln L}{\partial(\sigma_\tau^2)} = -\frac{RF}{2\sigma_\tau^2} + \frac{1}{2\sigma_\tau^2}\sum_{r=1}^{RF}(\tau_r - m_\tau)^2 = 0,$$

откуда следует

$$\begin{aligned} \overline{m}_\tau &= \frac{1}{RF}\sum_{r=1}^{RF}\tau_r, \\ \overline{\sigma}^2 &= \frac{1}{RF}\sum_{r=1}^{RF}(\tau_r - \overline{m}_\tau)^2. \end{aligned} \tag{8}$$

При оценке на основе 200 фреймов реального аудио сигнала получено $\overline{m}_\tau = 428$, $\sigma_\tau^2 = 185$ и среднеквадратическая ошибка (СКО) оценки составляет 3,5% (табл. 1). При рассмотрении 10000 фреймов СКО уменьшается до 2% и оценка вероятностных мер ритмической структуры исследуемого аудио ОПС становится более точной. Оптимизация алгоритмов погружения к атакам временной десинхронизации на основе использования ритмической структуры конкретного ОПС при погружении ЦВЗ позволяет обеспечивает требуемый уровень устойчивости.

Таблица 1 – Определение вероятностных мер ритмической структуры аудио-сигнала

r	1	2	3	4	5	6	7	8	9	10
τ, мс	414	428	428	441	427	456	414	414	428	442
$\Delta\tau$	14	0	0	-13	1	-28	14	14	0	-14
r	11	12	13	14	15	16	17	18	19	20
τ, мс	414	428	414	456	442	428	427	414	414	442
$\Delta\tau$	14	0	14	-28	-14	0	1	14	14	-14

На основании полученных оценок вероятностных мер ритмической структуры ОПС преобразуется к форме с детерминированным ритмом (рис. 4). После этого осуществляется погружение ЦВЗ и при необходимости восстановление случайной ритмической структуры исходного файла (рис. 3).

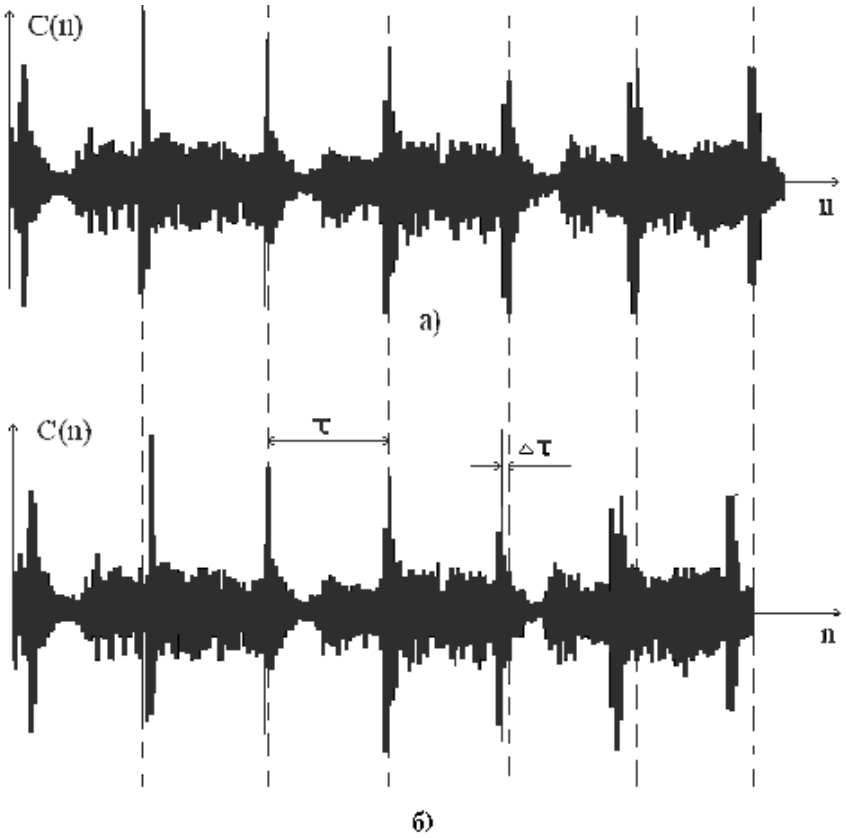

Рисунок 4 – Временные диаграммы исходного аудио (а) и с преобразованной ритмической структурой (б)

Моделирование с несколькими ОПС при погружении с учетом ритмической структуры продемонстрировало невозможность для атакующего без знания СК ухудшить вероятности ошибок более, чем на порядок в рамках заданных ограничений на надежность восприятия стегасообщения.

В заключении можно отметить, что подтверждение результатов аналитических исследований эффективности систем с ЦВЗ в условиях воздействия аддитивного шума имитационным моделированием при использовании в качестве ОПС аудио-сигналов позволяет надеется на ценность теоретических исследований для разработки практических систем. Оптимизация собственно алгоритмов погружения и детектирования ЦВЗ возможна после конкретизации практического применения системы. Не менее важным является исследование и оптимизация системы с ЦВЗ при ОПС в виде аудио сигналов относительно атак оцениванием ЦВЗ, что будет выполнено в последующих работах.

Литература

1. *Маракова И.И., Сафронов А.С.* Проблематика и перспективы методов сокрытия информации// Труды Одесск. нац. политехн. ун.-та. – 2003. – Вып.1. – С. 184–188.
2. *Bassia P., Pitas I.* Robust Audio Watermarking in the Time Domaine – EUSIPCO – Vol. 1. – Rodos, Greece. – 1999. – P.177-183.
3. *Cox I, Kilian J., Leighton F.T., Shamoon T.* Secure Spread Spectrum Watermarking for Images, Audio and Video // IEEE Int. Conference on Image Processing. – Vol. 3. –1996. – P. 243-246.
4. *Маракова И.И., Мараков Д.А.* Оценка эффективности систем со скрытыми цифровыми метками// Труды Одесск. нац. политехн. ун-та. – 2002. – Вып.2. – С. 110-115.
5. *Painter T., Spanias A.* Perceptual Coding of Digital Audio // Proceedings of the IEEE. – Vol.88(4). – 2000. – P. 413-451.
6. *Akaike H.* (1973). Information theory and an extension of the maximum likelihood principle, in B.N. Petrox and F. Caski, Second international symposium on information theory. – Budapest: Akademiai Kiado. – P. 267-681.