

Литература

1. Кузнецов Е.П., Витязев В.В. Цифровая обработка сигналов в задачах эхо-компенсации. – "Цифровая обработка сигналов", 2006, № 3, с. 8-19.
2. Ректорис К. Вариационные методы в математической физике и технике. М., Мир, 1985. 590 с.

*Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры для инновационной России» на 2009-2013 гг, гос контракт № 14 740 11 0390*

*Статья поступила 09 12 2011*

**Д.т.н., проф. Е.Г. Жиляков, П.Г. Лихолоб (НИУ «БелГУ»)**

**E.G. Zhilyakov, P.G. Likholob**

**СПОСОБ СКРЫТОГО ИНДЕКСИРОВАНИЯ РЕЧЕВЫХ  
СООБЩЕНИЙ В МУЛЬТИМЕДИЙНЫХ БАЗАХ ДАННЫХ**

**A WAY TO HIDE INDEXING VOICE MESSAGES  
MULTIMEDIA DATABASES**

*В статье предложена модель создания метки индекса элемента мультимедийной базы данных. Описан способ, внедряющий метку цифрового водяного знака в речевые данные.*

*In this paper we propose a model to label the element index multimedynoy database. Describes a method that implements label digital watermark in the speech data. The method meets the proposed criteria characterizing features of the introduction of CEH in speech data.*

*Ключевые слова: стеганография, цифровой водяной знак, цифровая метка, субполосная матрица, собственный вектор, часть энергии, внедрение, устная речь, речевые данные.*

*Key words: steganography, digital watermarking, digital label, subpolosnaya matrix, the eigenvector of the energy, introduction, speaking, speech data.*

Одной из распространенных форм обмена информацией, представляющих собой данные и знания, является устная речь. Для хранения и накопления информации в виде мультимедийных баз, аудио-сигналы устной речи преобразовывают в кодовые комбина-

ции (речевые данные) и индексируют. Задачу индексирования можно решить при помощи методов стеганографии. Методы цифровой стеганографии позволяют внедрить метку без увеличения объема хранимых кодовых комбинаций и заметной потери их субъективного качества звучания.

Под меткой здесь будем понимать неделимый элемент ( $w_i$ ) счетного множества, являющийся частью некоторой малообъемной информации, предназначенной для индексирования речевых данных в мультимедийных базах данных. Объединение меток образует идентификатор для процесса накопления и поиска информации, который в стеганографии принято называть – цифровым водяным знаком (ЦВЗ).

Для внедрения меток в аудио-сигналы разработано множество методик [1-3]. Существующие методики адаптированы под внедрение меток в аудио-сигналы, являющиеся музыкальными произведениями. Отличительной особенностью музыкальных произведений (рис. 1а) от речевых данных (рис. 1б) является наличие большого количества частотных составляющих. Количество частотных составляющих ограничивает пространство вариантов внедрения, и определяет пороги внедрения.

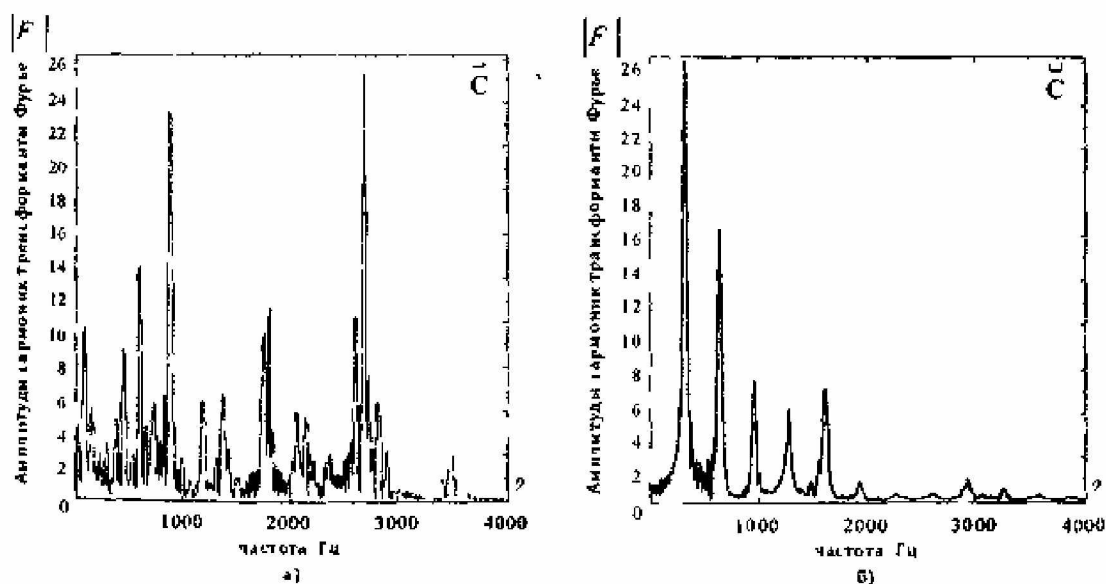


Рис 1

Огибающая амплитудного спектра окна аудио-сигнала (а), соответствующего фрагменту концерта Антонио Вивальди, огибающая амплитудного спектра окна аудио-сигнала (б), соответствующего звуку «о» в слове «компьютерные»

Для сравнения на рис. 1а представлен фрагмент музыкального произведения оркестрового сопровождения, соответствующего фрагменту концерта Антонио Вивальди, и представлен отрезок данных речевого аудио-сигнала, соответствующий звуку «о». Оба аудио-сигнала имеют частоту дискретизации 8000 Гц и уровень квантования 16 бит.

Анализ рис. 1б показывает, что присутствие в аудио-сигнале речевых данных меньшего количества частотных компонент, ограничивает выбор частоты внедрения ( $f_0$ ). Важно отметить, что в аудио-сигнале речевых данных, в отличие от музыкального произведения, энергия большинства звуков русской речи сосредоточена в узком частотном интервале. Для речевого аудио-сигнала можно выделить малое количество частотных компонент, несущих подавляющую долю энергии, что существенно ограничивает поиск позиции внедрения, обеспечивающей высокую скрытность. Также в большинстве способов внедрения меток ЦВЗ [1-3] частотную ось разбивают на равные интервалы, что не позволяет более гибко использовать способ внедрения. Эти особенности существенно ограничивают пространство вариантов внедрения, что, в свою очередь, приводит к уменьшению скрытности внедряемых меток. Сосредоточение энергии в узком частотном интервале также приводит к тому, что необходимо выбирать порог внедрения каждому окну для обеспечения минимальных искажений.

Исходя из вышесказанного, можно предложить ряд критериев, выполнение которых позволит повысить гибкость и скрытность алгоритмов внедрения меток в речевые данные: возможность осуществить внедрение в узкополосный частотный интервал произвольной ширины; возможность дискретно выбирать центральную частоту внедрения (выполняя уход от привязки к равномерному разделению на частотные интервалы); автоматически выбрать порог внедрения, учитывая особенности частотных компонент; обеспечить минимальное воздействие на частотные компоненты, находящиеся вне частотного интервала, в котором осуществлялось внедрение. На основе сформулированных выше критериев предлагается модель внедрения метки в аудио-сигнал речевых данных на основе частотных представлений (рис. 2).

В данной модели метка  $w$ , ЦВЗ кодируется автоматически выбираемым и регулируемым перепадом между частями энергии

$\pm \Delta P$ , сосредоточенными в соседних частотных интервалах ( $D_l$  и  $D_r$ ) окна аудио-сигнала. Метка  $w_i$  кодируется в знаке фронта между частями энергии.

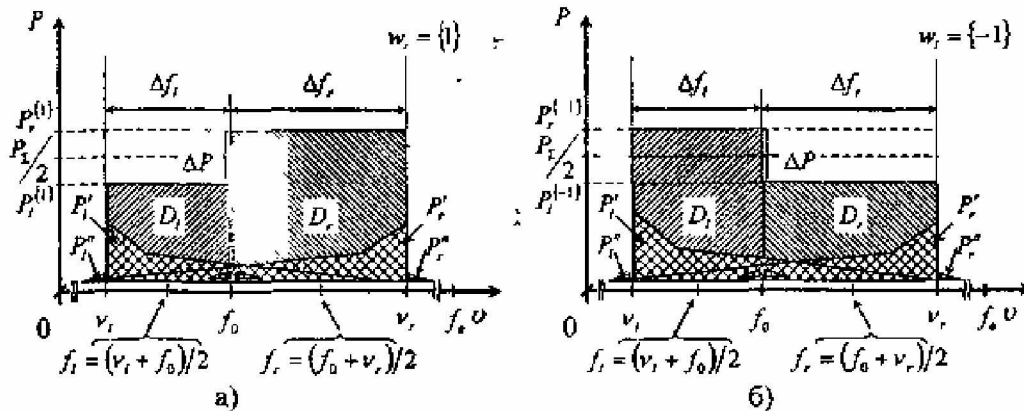


Рис 2

Модель внедрения одного бита ЦВЗ

Для разработки метода внедрения был использован математический аппарат [4], позволяющий выделить часть энергии ( $P$ ) в заданном частотном интервале:

$$P = X_n^T \mathbf{A} X_n, \quad (1)$$

где:  $X_n$  – вектор значения амплитуд аудио-сигнала,

$T$  – знак транспонирования,

матрица  $\mathbf{A} = \{a_{i,k}\}$  с элементами вида (2):

$$a_{i,k} = (\sin(\nu_r(i-k)) - \sin(\nu_l(i-k))) / (\pi(i-k)), \quad i, k = 1, \dots, N \quad (2)$$

Также важно отметить свойство оптимальной фильтрации в заданном частотном интервале, которым обладает матрица  $\mathbf{A} = \{a_{i,k}\}$ :

$$Y_n = (\mathbf{I} - \mathbf{A}) X_n, \quad (3)$$

где:  $\mathbf{I}$  – единичная матрица, имеющая размерность  $N \times N$ ;

$Y_n$  – вектор-результат оптимальной фильтрации.

Определение закодированной метки  $w_i$  ЦВЗ в стего-файл-контейнере осуществляется путём оценки разницы в частях энергии соседних частотных интервалов ( $D_l$  и  $D_r$ ), находящихся слева и справа относительно центральной частоты внедрения ( $f_0$ ):

$$w_i = \begin{cases} 1, & P_l > P_r \\ -1, & P_l < P_r \end{cases} \quad (4)$$

Внедрение предлагается реализовывать путём замещения существующей разницы частей энергии в соседних частотных интервалах на разницу, формируемую частями энергии собственных векторов ( $Q_l$  и  $Q_r$ ) субполосных матриц  $A_l$  и  $A_r$ , рассчитанных для правого ( $D_r$ ) и левого частотных интервалов ( $D_l$ ). Субполосные матрицы ( $A_l$  и  $A_r$ ) являются симметричными и неотрицательно определенными. Поэтому они обладает полной системой ортонормальных собственных векторов ( $Q_k$ ), соответствующих неотрицательным собственным числам ( $\lambda_k$ ) и удовлетворяющих соотношениям [4]:

$$A = \sum_{k=1}^N \lambda_k Q_k Q_k^T. \quad (5)$$

Для замещения целесообразно использовать собственные вектора ( $Q_k$ ) субполосных матриц ( $A$ ), соответствующие собственным числам ( $\lambda_k=0.9-1$ ), т.к. часть энергии этих векторов максимально сосредоточена в заданных частотных интервалах. Стоит отметить, что доля концентрации энергии в частотном интервале пропорциональна значению собственного числа, а просачивание части энергии ( $P''_r$  и  $P''_l$ ) пропорционально  $(1-\lambda_k)$  [4].

Внедрение осуществляют модификацией отрезка аудиосигнала, являющегося окном файла-контейнера ( $C_n, n=1, \dots, N$ ):

$$S_n = (I - (A_l + A_r)) C_n + K_l \cdot Q_l + K_r \cdot Q_r, \quad (6)$$

где:  $K_r, K_l$  – нормирующие коэффициенты собственных векторов  $Q_l$  и  $Q_r$ .

$S_n$  – вектор аудио- сигнала содержащего метку (окно стего- файла-контейнера).

Важным моментом является вычисление нормирующих коэффициентов ( $K_r, K_l$ ). Данные коэффициенты: учитывают часть энергии, содержащейся в частотных интервалах до внедрения ( $P_\Sigma$ ); учитывают и минимизируют влияние друг на друга частей энергии, оставшейся в замещаемых частотных интервалах после оптимальной фильтрации  $P'_r$  и  $P'_l$ ; определяют значение разницы в частях энергии  $\Delta P$  (для создания перепада, положительного или отрицательного фронта, между частями энергии).

Для обеспечения скрытности внедрения метод предполагает сохранение части энергии, содержащейся до внедрения:

$$P_\Sigma = C_n^T \cdot (A_l + A_r) \cdot C_n, \quad (7)$$

где  $P_{\Sigma}$  – часть энергии, которая содержалась в частотных интервалах до внедрения.

Для создания перепада между частями энергии (фронта)  $P_l$  и  $P_r$ , можно использовать сумму частей энергии ( $\Delta P$ ), вычисляемую по формулам:

$$P'_l = ((I - (A_l + A_r)) \cdot C_n + Q_l)^l \cdot A_r \cdot ((I - (A_l + A_r)) \cdot C_n + Q_l), \quad (8)$$

$$P'_r = ((I - (A_l + A_r)) \cdot C_n + Q_r)^r \cdot A_l \cdot ((I - (A_l + A_r)) \cdot C_n + Q_r), \quad (9)$$

$$\Delta P = P'_l + P'_r. \quad (10)$$

Исходя из вышеприведенных рассуждений, нормирующие коэффициенты имеют вид:

$$K_r = P_{\Sigma} - \xi \cdot \Delta P \cdot w_r, \quad (11)$$

$$K_l = P_{\Sigma} - K_r, \quad (12)$$

где  $\xi$  - численная величина стойкости ЦВЗ (робастность ЦВЗ).

В качестве стега-ключа  $K()$  для данного метода выступает функционал  $K(N, \Delta f_r, \Delta f_l, f_0, \xi)$ , определяемый: длиной окна во временной области ( $N$ ); центральной частотой внедрения меток ( $f_0$ ); шириной частотного интервала ( $\Delta f_r$ ), расположенного справа относительно  $f_0$ ; шириной частотного интервала ( $\Delta f_l$ ), расположенного слева относительно  $f_0$ ; величиной, позволяющей регулировать робастность ЦВЗ ( $\xi$ );

Для визуальной оценки эффективности внедрения на рис. 3 представлены огибающая амплитудного спектра отрезка данных речевого аудио-сигнала, соответствующая звуку «О» и огибающая амплитудного спектра этого же отрезка с внедренной меткой. Для определения границ внедрения на рис. 3 представлены огибающими амплитуд гармоник трансформанты Фурье собственных векторов ( $Q^L$  и  $Q^R$ ) субполосных матриц  $A_l$  и  $A_r$ .

Анализ рис. 3. показывает, что изменения амплитуд гармоник трансформанты Фурье были существенно затронуты в замещаемых частотных интервалах. При этом сохранилась форма спектра. Несущественные изменения в амплитудах гармоник трансформанты Фурье уменьшались при удалении от замещаемых частотных интервалов. Корреляция сигналов во временной области для рассмотренного окна данных речевого аудио-сигнала, соответствующая звуку «О» с меткой и без метки составляет 0.9994.

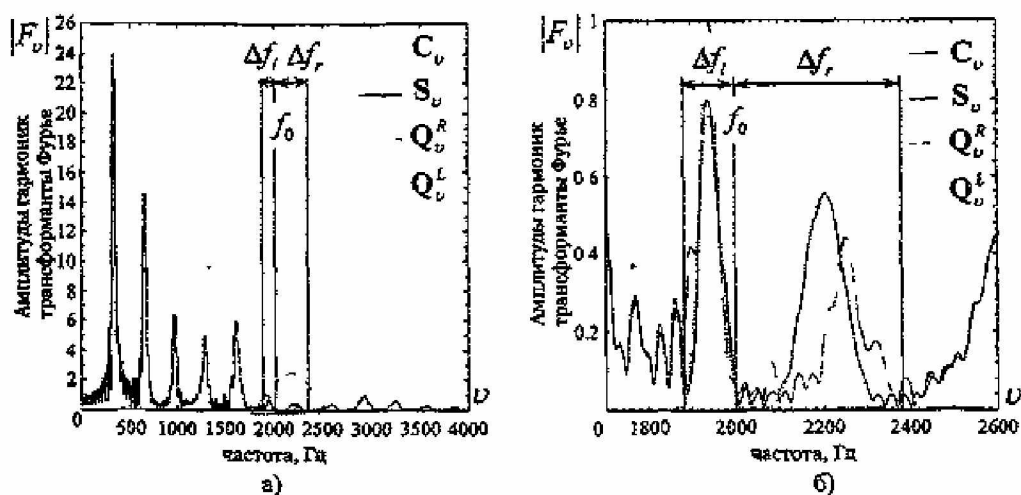


Рис. 3

Огибающие амплитудных спектров (а); огибающая амплитудный спектр, соответствующая частотному интервалу  $\nu \in [1700, 2600]$ , Гц (б)  
 ( — - окно исходного аудио-сигнала, представленное огибающей амплитуд гармоник трансформанты Фурье  $C_\nu$  (файл-контейнер); — — - окно аудио-сигнала с внедренной меткой, представленное огибающей амплитуд гармоник трансформанты Фурье  $S_\nu$  (стего-файл-контейнер); - - - - собственный вектор субполосной матрицы  $A_\nu$ , представленный огибающей амплитуд гармоник трансформанты Фурье  $Q_\nu^R$ , замещающий правый частотный интервал, - - - - собственный вектор субполосной матрицы  $A_\nu$ , представленный огибающей амплитуд гармоник трансформанты Фурье  $Q_\nu^L$ , замещающий левый частотный интервал)

В табл. 1 приведены значения среднеквадратической относительной погрешности (СКО) для рассматриваемого окна данных речевого аудио-сигнала. При этом варьировалась центральная частота внедрения ( $f_0$ ) и ширина интервалов замещения ( $\Delta f_r$  и  $\Delta f_l$ ).

Результаты, приведенные в табл. 1, свидетельствуют о том, что искажения вносимые в результате внедрения метки сильно зависят от распределения энергии по частотным интервалам. Но возможно подобрать параметры, при которых вносимые искажения будут несущественны. Описанный способ осуществляет внедрение в аудио-сигнал без перехода к частотному представлению сигнала метку цифрового водяного знака. Искажения, вносимые ЦВЗ, вызывают минимальные искажения во временном и частотном представлении аудио-сигнала, а также без заметной потери субъективного качества звучания, при правильном подборе параметров внедрения.

Таблица 1

$f_0, \Gamma\text{ч}$	450	600	1000	1500	2000	3000	3600
СКО ( $\Delta f_i = 120 \Gamma\text{ч}, \Delta f_f = 400 \Gamma\text{ч}$ )	0.1028	0.0536	0.0014	0.0041	0.0014	$5.5 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$
СКО ( $\Delta f_i = 120 \Gamma\text{ч}, \Delta f_f = 120 \Gamma\text{ч}$ )	0.1504	0.0556	0.0039	0.0084	$8.9 \cdot 10^{-4}$	$1.2 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$
СКО ( $\Delta f_i = 400 \Gamma\text{ч}, \Delta f_f = 60 \Gamma\text{ч}$ )	0.996	0.697	0.0447	0.0107	0.0086	0.0018	$9.2 \cdot 10^{-4}$



Литература

1. Савченко Н.А. Аналитический обзор алгоритмов квантования как эффективного инструмента для внедрения скрытой информации. – В сб.: Инфоком-4 инфокоммуникационные технологии в науке, производстве и образовании: Тез. докл. IV междунар. НТК, Ставрополь, 2010, т. 2, с. 15-17.
2. Кошкина Н.В. Внедрение ЦВЗ в аудиосигналы на основе пакетной вейвлет-декомпозиции и частотного маскирования. – "Искусственный интеллект", 2010, вып. 4, с. 381-387.
3. Рублёв Д.П., Макаревич О.Б., Федоров В.М. Метод стеганографического встраивания сообщений в аудиоданные на основе вейвлет-преобразования. "Известия ЮФУ. Технические науки", 2010, с. 119-206.
4. Жилияков Е.Г., Белов С.П., Черноморец А.А. Вариационные методы синтеза сигналов на основе частотных представлений. – "Вопросы радиоэлектроники", сер. ЭВТ, 2010, вып.1.

Статья поступила 09.12.2011

Д.т.н., проф. Е.Г. Жилияков, к.т.н. Е.И. Прохоренко,  
А.А. Фирсова, Д.К. Секисов (НИУ «БелГУ»)

E.G. Zhilyakov, E.I. Prokhorenko, A.A. Firsova, D.K. Sekisov

КОМПЬЮТЕРНАЯ СЕГМЕНТАЦИЯ РЕЧЕВЫХ СИГНАЛОВ

COMPUTER SEGMENTATION OF SPEECH SIGNALS

*В статье рассмотрены решающие функции алгоритмов сегментации речевых сигналов, основанные на анализе распределения энергии по частотным интервалам. Проведен анализ возможности использования предложенных решающих функций для осуществления сегментации речевого сигнала.*

*The article deals with the crucial function of the segmentation algorithm of speech signals based on an analysis of the distribution of energy over the frequency intervals. The analysis of the possibility of using the proposed decision functions for the segmentation of the speech signal.*

*Ключевые слова: речевой сигнал, распределение энергии по частотным интервалам, сегментация речевых сигналов, решающие функции.*

*Key words: speech signal energy distribution over the frequency range segmentation of speech signals, the crucial function.*