4.3 СВИ для аудиосигналов

(А) Обзор (давно известных) специализированных СВИ для аудио

НЗБ-метод и СВИ (Cvejic)

- НЗБ для аудио: 2-3 бита – нормально
- СВИ (Cvejic): до 4-5 бит на отсчёт
- 3-шаговый подход:
  - 1) Стандартное встраивание в 4 НЗБ непосредственной заменой бит
  - 2) Замена старшего бита с минимизацией ошибки
  - 3) Диффузия ошибки встраивания ЦВЗ

СВИ (Cvejic). Принцип замены старшего бита

- Пусть мы встроили k бит ($k < 16$), тогда $\varepsilon_{max} \leq 2^k - 1$
- $C(n)$ – исходный отсчёт сигнала
- Два варианта $C^W(n)$:
  - $C_{LSB}(n)$ – результат прямого НЗБ-встраивания
  - $C'_{LSB}(n)$ – НЗБ + замена $(k+1)$-го бита
- Расчёт ошибки и выбор варианта
  - $e(n) = |C(n) - C_{LSB}|, e'(n) = |C(n) - C'_{LSB}(n)|$
  - $e(n) \leq e'(n) \rightarrow C^W(n) = C_{LSB}(n)$
  - $e(n) > e'(n) \rightarrow C^W(n) = C'_{LSB}(n)$
- Итог: было $\varepsilon_{max} = 2^k - 1$, стало $\varepsilon'_{max} = 2^{k-1}$

СВИ (Cvejic). Процедура диффузии ошибки

- $C(n+1) \coloneqq C(n+1) + \frac{1}{2}e(n)$
- $C(n+2) \coloneqq C(n+2) + \frac{1}{4}e(n)$
- $C(n+3) \coloneqq C(n+3) + \frac{1}{8}e(n)$
- $C(n+4) \coloneqq C(n+4) + \frac{1}{8}e(n)$

СВИ (Bender-1, метод фазового кодирования)

- Слуховая система человека более чувствительна к изменению амплитуды сигнала, нежели фазы
- Модификация фазы тем более незаметна, если сохраняется характер её изменения во времени
- Основная идея метода фазового кодирования состоит в замене фазы исходного сегмента на опорную фазу, характер изменения которой отражает собой данные, которые необходимо скрыть

СВИ (Bender-1). Встраивание информации

- $C(n), 0 \leq n \leq N-1$
- Сигнал разбивается на K коротких сегментов $\mathrm{c}_k(n),\ k = 0..K-1, n = 0..N/K-1$
- К каждому сегменту применяется ДПФ (N/K-точечные). Далее выделяется фаза - $\varphi_k(m)$ и амплитуда $A_k(m), m = 0..N/K-1, k = 0..K-1$
- Запоминается разность фаз между соседними сегментами:

$$\Delta\varphi_k(m) = \varphi_{k+1}(m) - \varphi_k(m), k = 0..K-2$$

СВИ (Bender-1). Встраивание информации

- Бинарный вектор, подлежащий встраиванию, представляется в виде ступенчатой функции $W(m)$ в частотной области со ступеньками в $\frac{\pi}{2}$ или $-\frac{\pi}{2}$.

$$W(m) = (-1)^{b_m}\pi/2, \qquad \forall m = 0..N/K$$

- Заменяем начальную фазу

$$\varphi_0^W(m) = W(m) \quad \forall m = 0..N/K$$

- Сбор фазы сигнала по φ_0^W и $\Delta\varphi_k(m)$:

$$\varphi_1^W(m) = \varphi_0^W(m) + \Delta\varphi_0(m)$$

...

$$\varphi_{K-1}^W(m) = \varphi_{K-2}^W(m) + \Delta\varphi_{K-1}(m)$$

- Собираем исходную амплитуду и новую фазу и делаем обратное ДПФ

СВИ (Bender-1). Итого

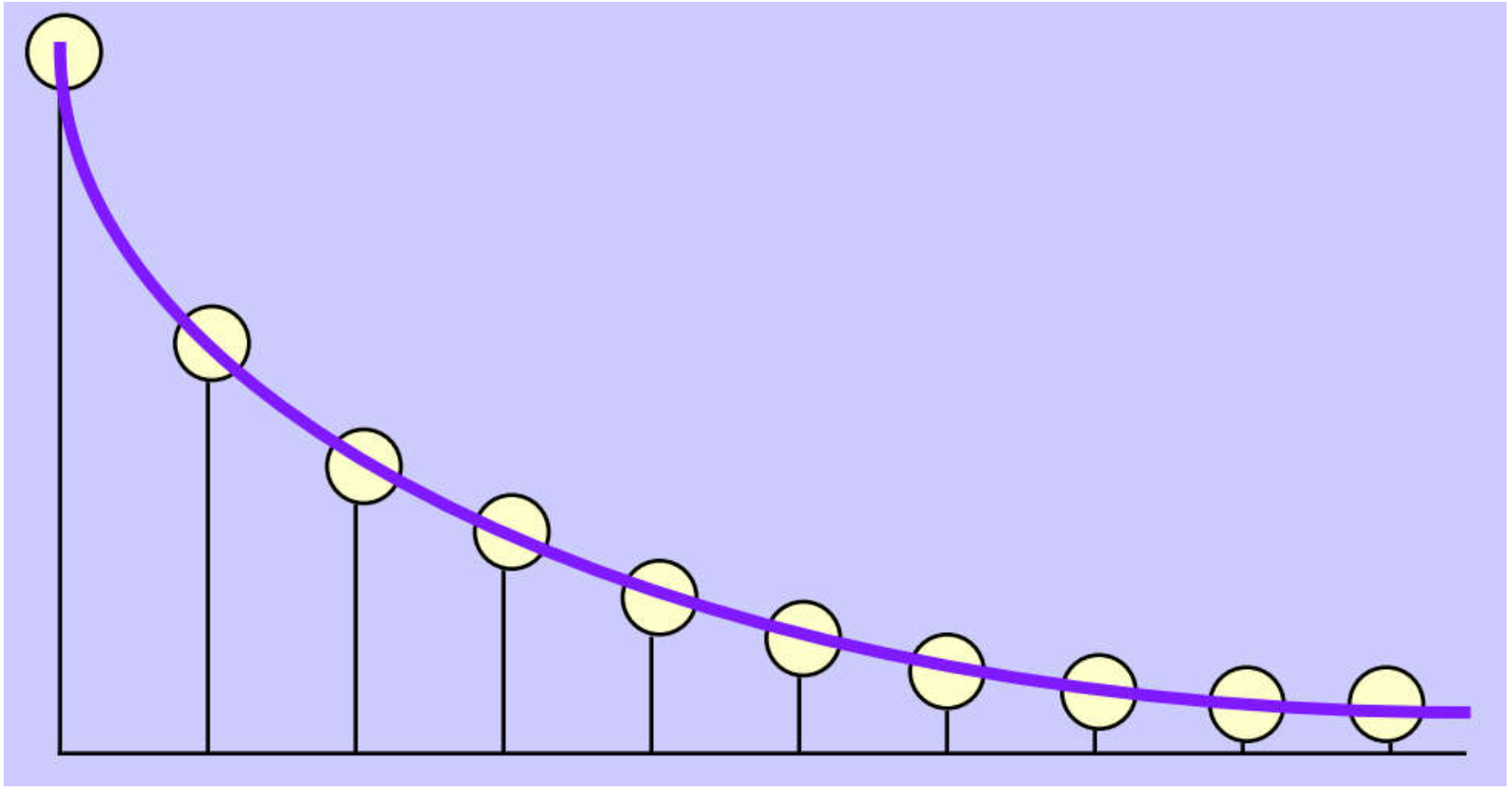
- Для извлечения нужны:
  - Синхронизация сигнала
  - Длина сегмента N/K
- Пропускная способность метода – 8-32 бит/с – сравнительно небольшая
- Принципиально иной подход, не применяющийся в СВИ для изображений

СВИ (Bender-2, метод на основе эхо-сигнала)

- Метод заключатся во встраивании секретной информации в аудио контейнер путем добавления эхо-сигнала
- Идея метода состоит в том, что при малом сдвиге одного сигнала относительно другого, человек воспринимает два сигнала как один, а эхо воспринимается как дополнительный резонанс
- При сдвиге T=0,001 секунды два сигнала сливаются в один

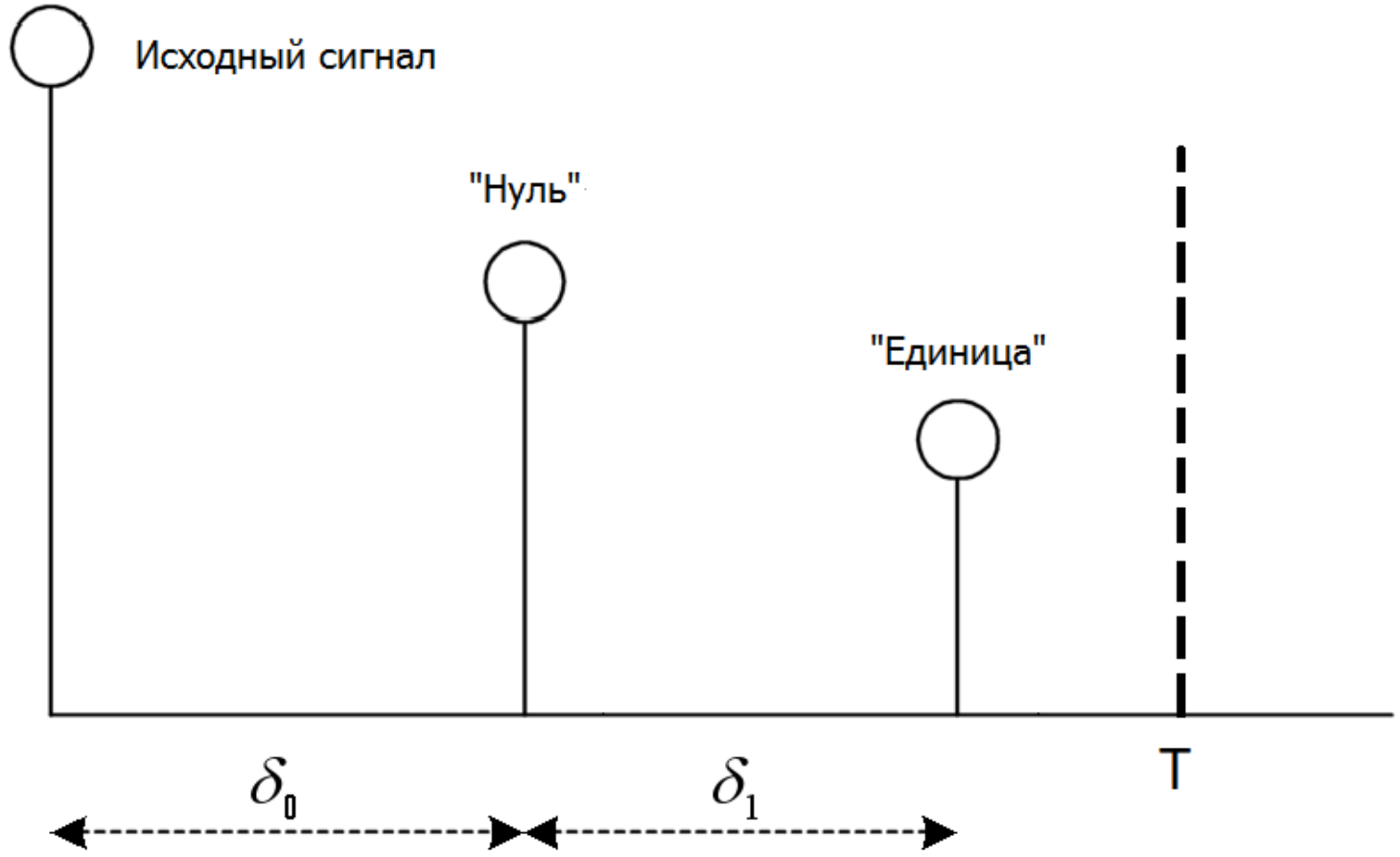
СВИ (Bender-2). Основы

- Натуральный сигнал затухает по экспоненциальному закону
- Значит искусственные эхо-сигналы должны соответствовать этой модели

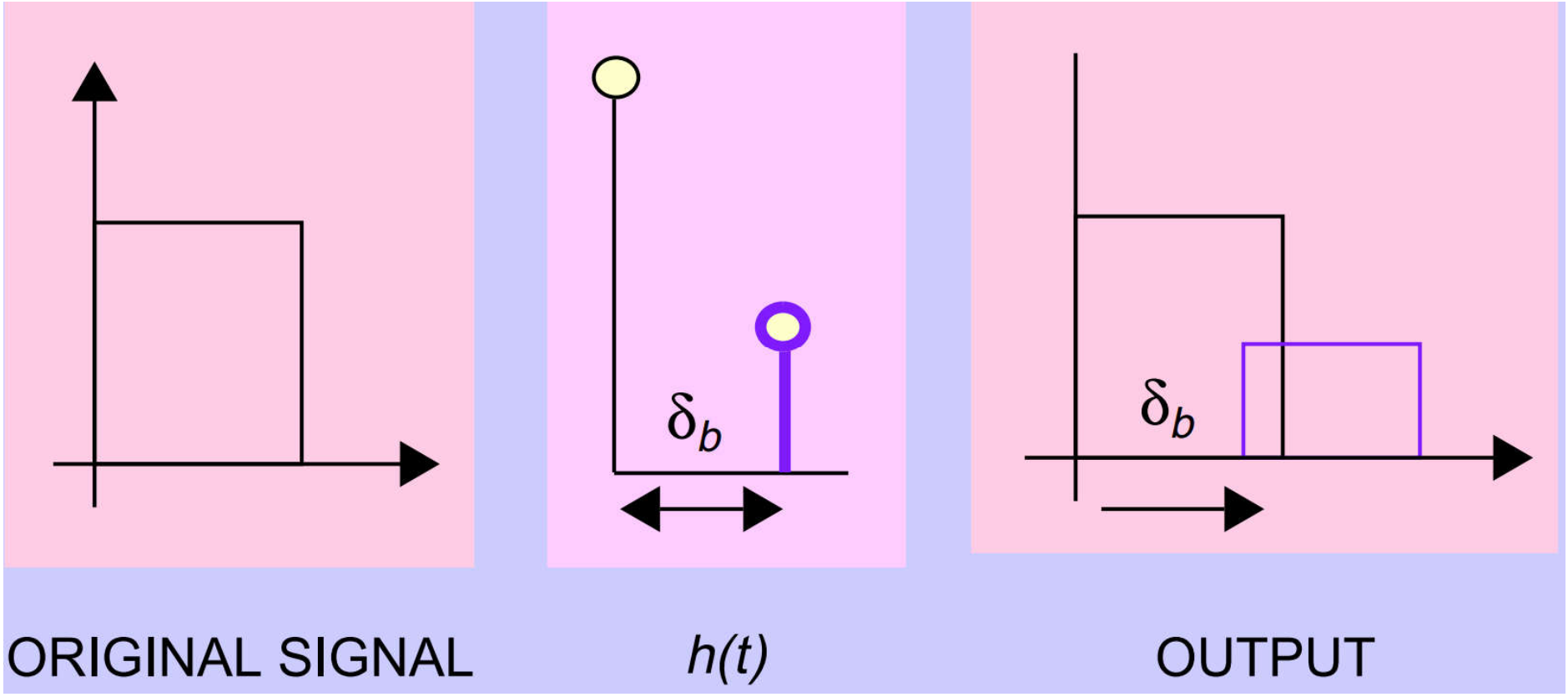


СВИ (Bender-2). Основы

- Принцип кодирования встраиваемой скрытой информации: два разных эхо-сигнала, отличающихся сдвигом δ_b, $b = \{0; 1\}$.

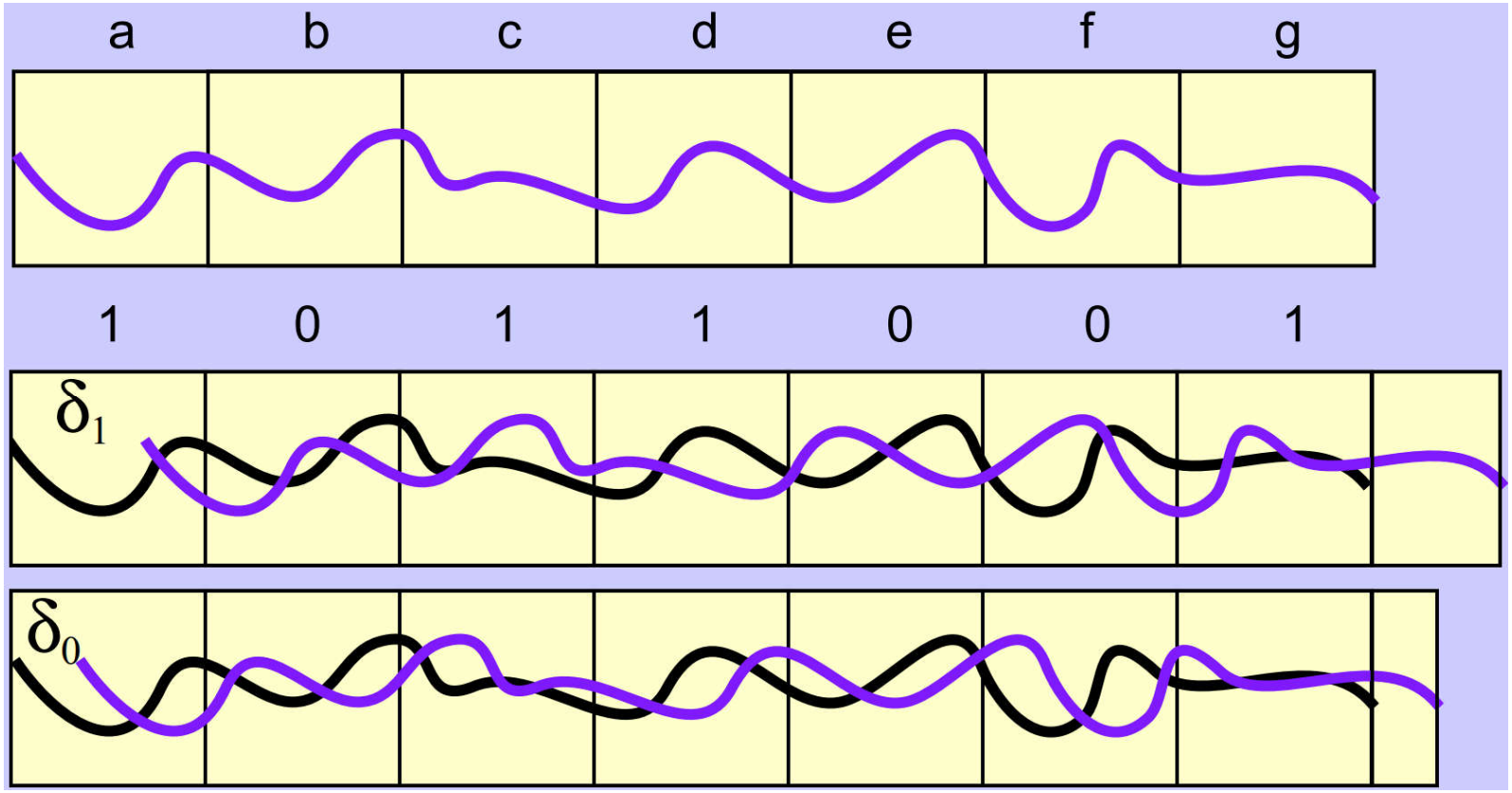


СВИ (Bender-2). Формирование эхо-сигнала



СВИ (Bender-2). Формирование эхо-сигнала

- Сигнал разделяется на фрагменты, в которые встраивается один бит:
- $C(n)$
- $C_1(n)$
- $C_0(n)$



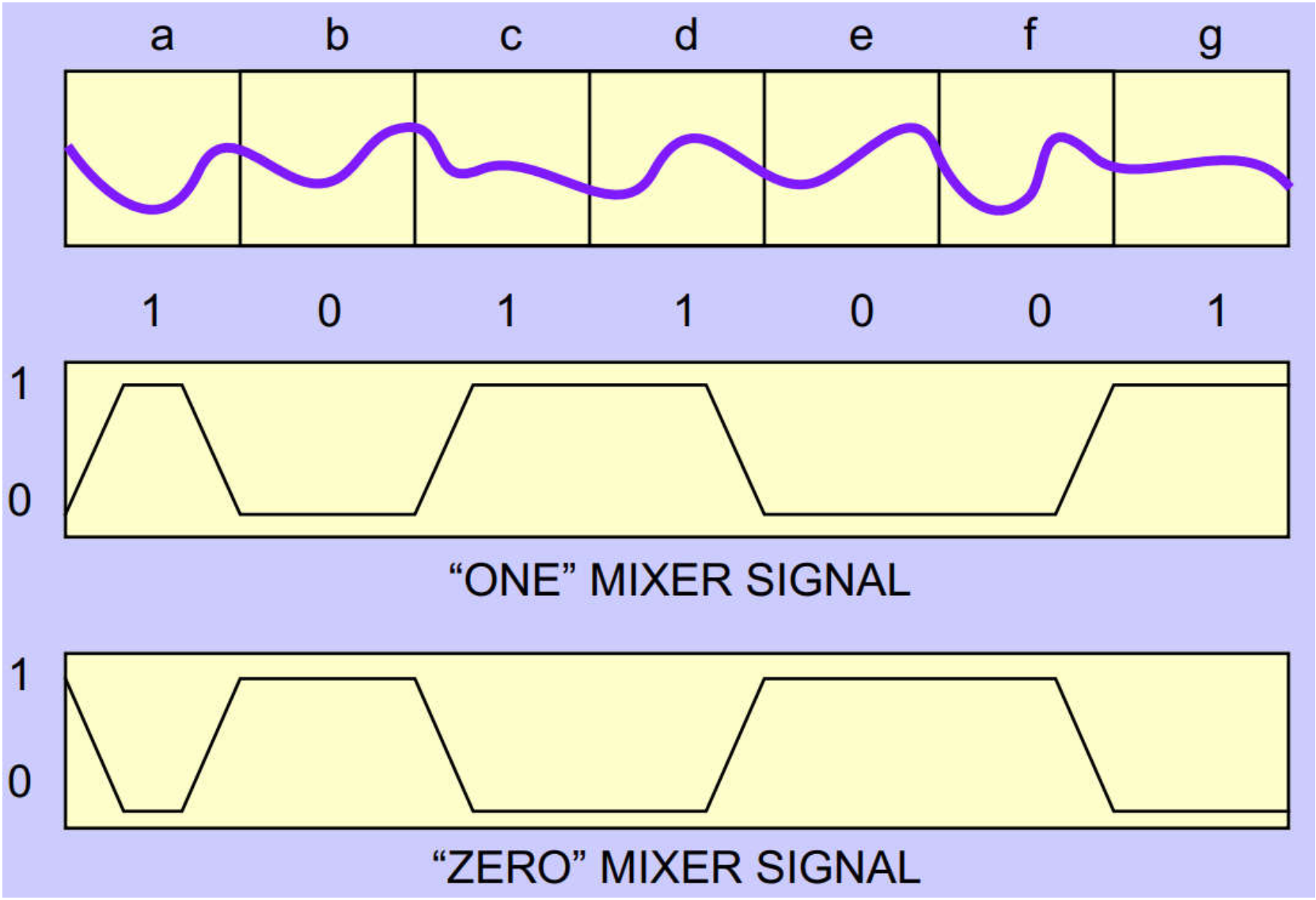
СВИ (Bender-2). Встраивание

- Итоговое встраивание

$$C^W(n) = C(n) + \alpha_0 \cdot zero(n) \cdot C_0(n) + \alpha_1 \cdot one(n) \cdot C_1(n),$$

- где
  - $one(n) = 1 - zero(n)$,
  - α_0, α_1 устанавливаются совместно на основе значений сдвигов δ_0, δ_1 и параметров затухания

СВИ (Bender-2). Встраивание



СВИ (Bender-2). Извлечение

- Кепстр сигнала:

$$K = \mathcal{F}^{-1}\left(ln\left(\mathcal{F}\left(\widetilde{C^{W}}\right)\right)^{2}\right)$$

- Расчёт автокорреляционной функции кепстра
- Это нужно для выравнивания спектра (он быстро затухает). Поэтому АКФ кепстра предпочтительнее, чем АКФ самого сигнала.
- Поиск двух локальных всплесков АКФ – отыскание δ_0, δ_1
- На каждом фрагменте корреляция K с двумя сдвинутыми копиями на δ_0, δ_1. Выбор значения бита по максимуму
- Пропускная способность метода – около 16 бит/с при сохранении высокого качества

(B) Сравнительное экспериментальное исследование применительно к MP3

Экспериментальное исследование. Постановка задачи. Исследуемые методы.

Цель – сравнить различные по принципу работы СВИ для аудио по пропускной способности, стойкости к MP3-сжатию, качеству результирующего аудио.

Метод НЗБ:

- использует один отсчет для передачи одного бита встроенной информации;
- параметр: номер затрагиваемой битовой плоскости.

Метод "Patchwork":

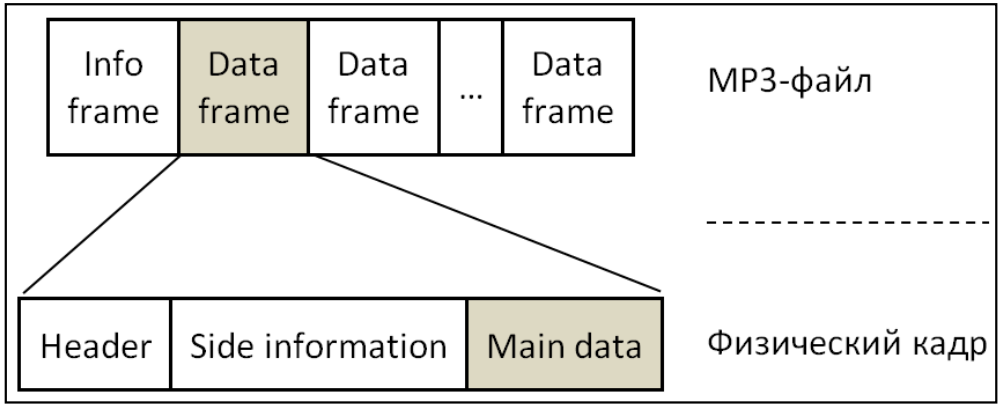
- использует множество отсчетов для передачи одного бита информации;
- параметры:
  - L – количество отсчетов, передающих один бит встроенной информации;
  - k – коэффициент, определяющий долю отсчетов, подлежащих изменению;
  - α – коэффициент, определяющий величину изменения отсчетов.

Формато-зависимый метод:

- использует области файла, которые не несут аудиоинформации.

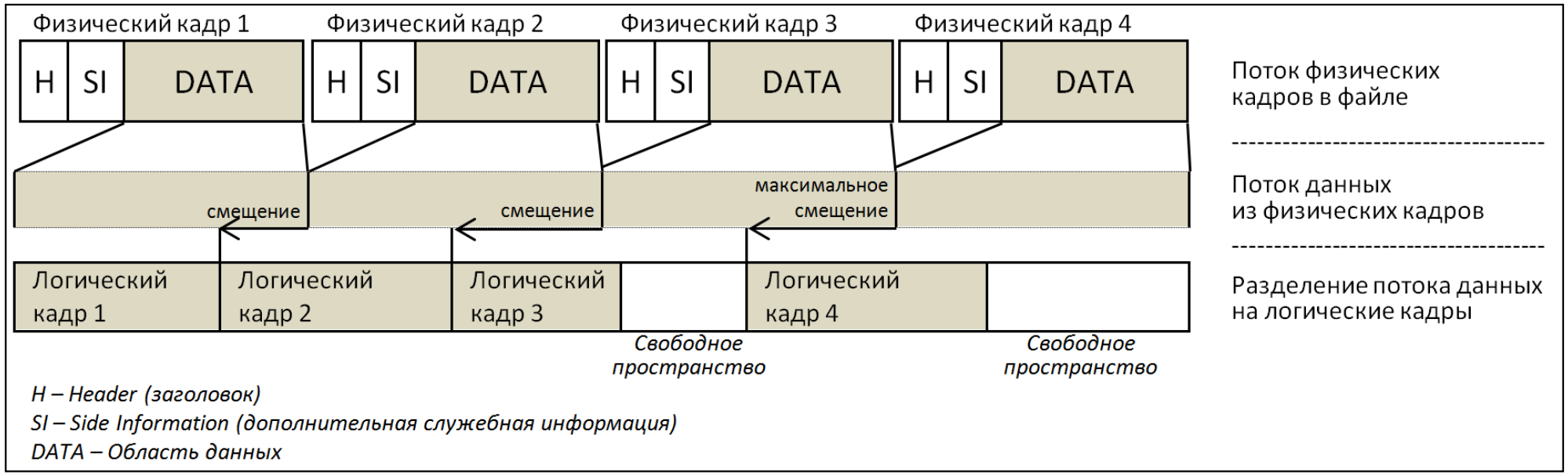
Структура MP3-файла как основа формато-зависимого метода

- Главные этапы алгоритма преобразования: переход в спектральную область; фильтрация спектральных компонент; переквантование; масштабирование; кодирование Хаффмана.
- Длина логического кадра – 576 отсчётов – фиксирована
- Длина кода Хаффмана после квантования переменная. Этот код пишется в поле Main data.
- К каждому кадру прилагается служебная информация для декодирования: Header (4 байта) и Side Information (17/32 байта).
- Длина физического кадра в режиме CBR фиксирована: 626 байт, включая Header и Side Information.



Структура MP3-файла как основа формато-зависимого метода

- Чтобы не возникало в каждом кадре неиспользуемое пространство, разделяют физические/логические кадры и применяют сдвиги
- Текущий физ. кадр содержит Header и SI по текущему логическому кадру, но данные его могут уже начаться в Main data предыдущего физического кадра.
- Смещение записано в SI, но это максимум 511 байт. Если данные предыдущего кадра закончились раньше, до всё-таки возникает незаполненное пространство.



Формато-зависимый метод

Порядок функционирования:

- Расчёт всех смещений логических кадров
- Восстановление последовательности отрезков кадров, не содержащих данные
- Последовательное заполнение этих отрезков встраиваемой информацией.

Свойства

- – Полноценной стеганографической защиты нет
- – При анализе файла отсутствие пустых областей будет очевидным
- + Эту проблему можно решить неполным заполнением контейнера: встраивание можно производить только в начало свободных областей
- – Однако это не проблема для умного злоумышленника
- – При перекодировании вся встроенная информация будет потеряна
- + Полное отсутствие искажений
- + Гарантия извлечения из MP3
- + Относительно большой объём встраивания и скрытость от посторонних глаз

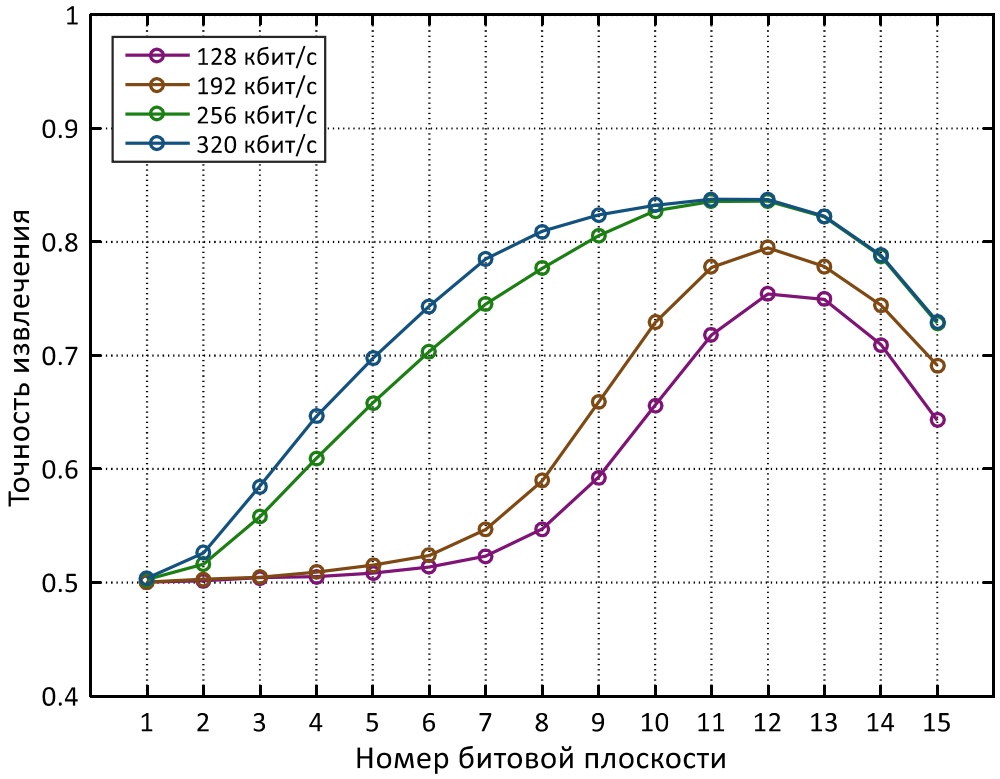
Критерии для сравнения методов

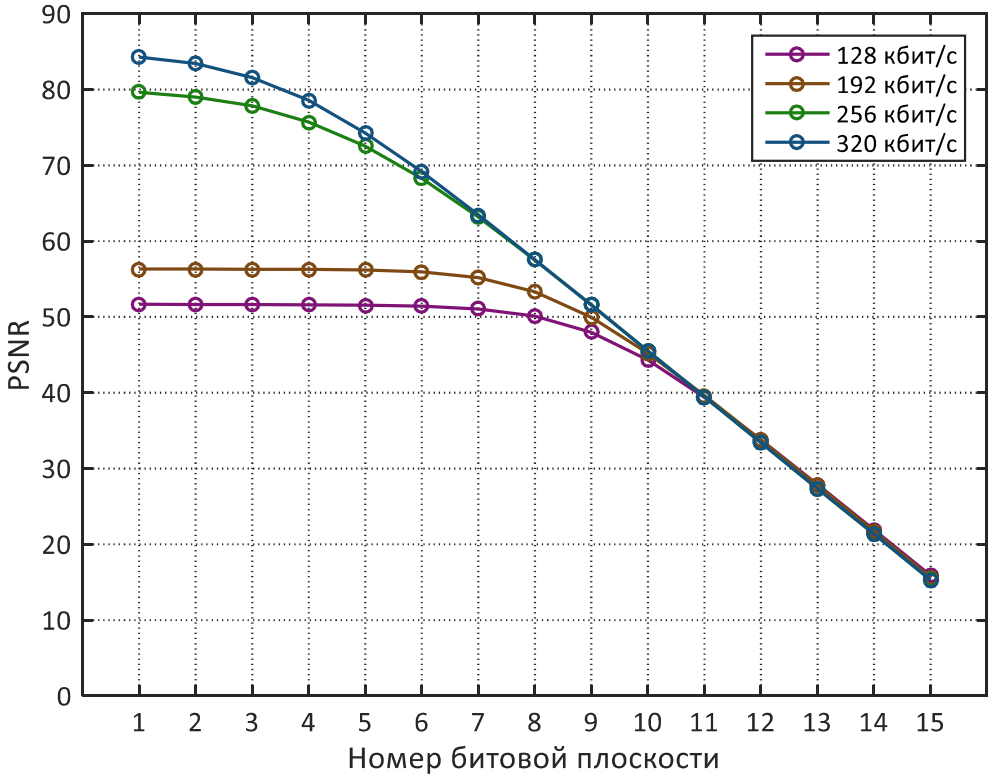
- Объем встраивания:
  - скорость передачи встроенной информации (VD, Volume of Data);
  - доля контейнера, занимаемая встроенными битами;
- точность извлечения;
- качество аудио после встраивания:
  - PSNR, пиковое соотношение сигнал-шум;
  - MSE-HAS, частотно-взвешенный среднеквадратичный показатель.

$$PSNR(u,v) = 10lg\frac{sup_n{}^2\, u(n)}{\varepsilon^2_{\text{кв}}(u,v)},$$ где $sup\, u(n)$ – максимально возможное значение сигнала $u(n)$; $\varepsilon^2_{\text{кв}}(u,v)$ – среднеквадратичная ошибка для сигналов u и v.	$$MSE\_HAS = \frac{1}{N}\sum_{n=0}^{N-1} W(\omega_n) \times \frac{1}{M}\sum_{m=0}^{M-1}\left\lvert E(m, e^{i\omega_n})\right\rvert^2,$$ где N – количество частотных интервалов; $W(\omega_n)$ – неотрицательная весовая функция; M – количество временных интервалов; $E(m, e^{i\omega_n})$ – кратковременный спектр разностного сигнала

Исследование метода НЗБ

Графики зависимости точности извлечения и критерия качества PSNR от номера затрагиваемой битовой плоскости для файлов с разными битрейтами.

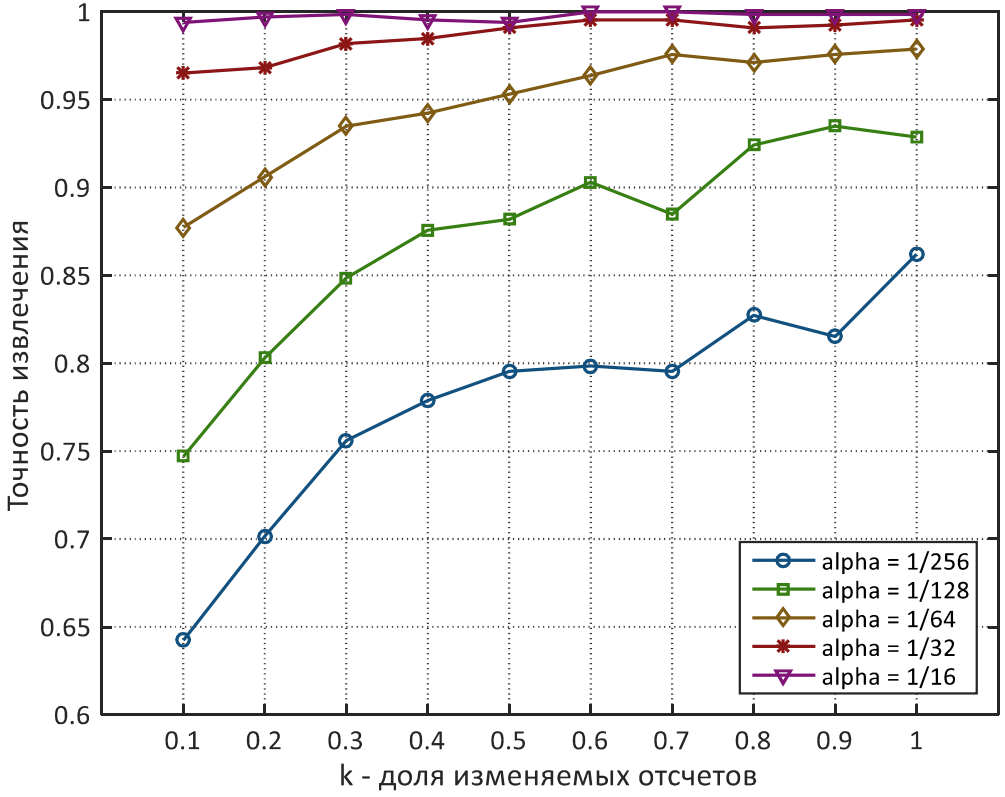


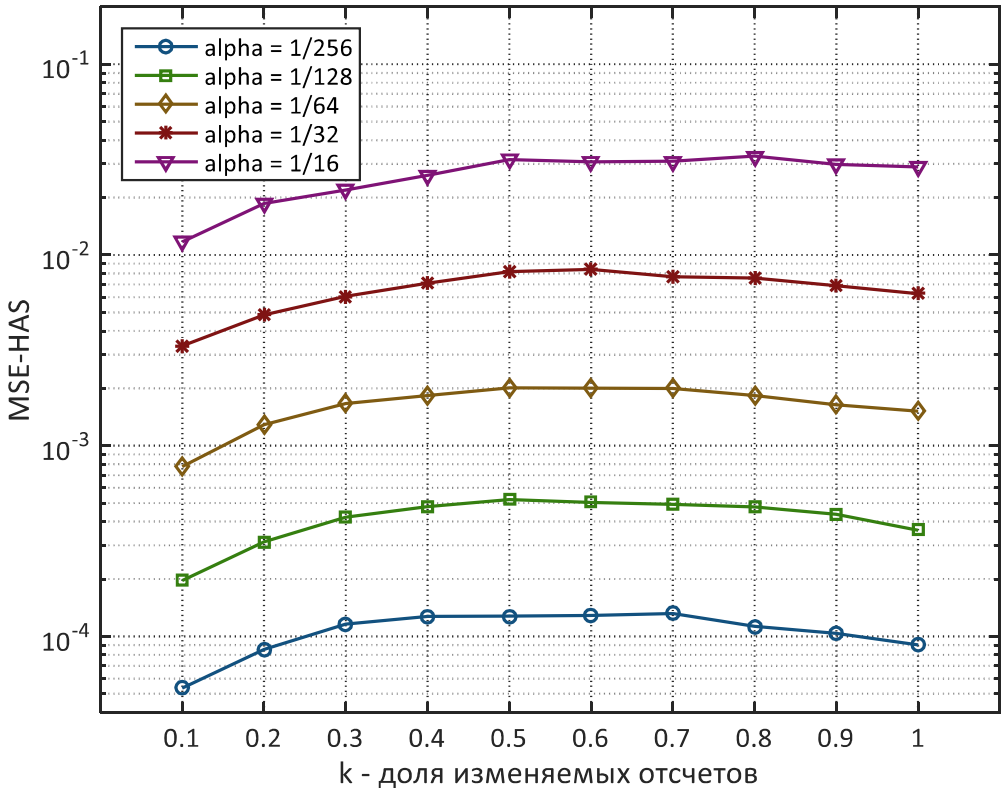


1 - младшая битовая плоскость
15 - старшая битовая плоскость
16 - знак (не используется)

Исследование метода "Patchwork"

Графики зависимости точности извлечения и критерия качества MSE-HAS от доли изменяемых отсчетов при разных значениях коэффициента альфа, определяющего величину изменения отсчетов. Объем встраивания фиксирован и равен 4,41 бит/с.





Исследование формато-зависимого метода

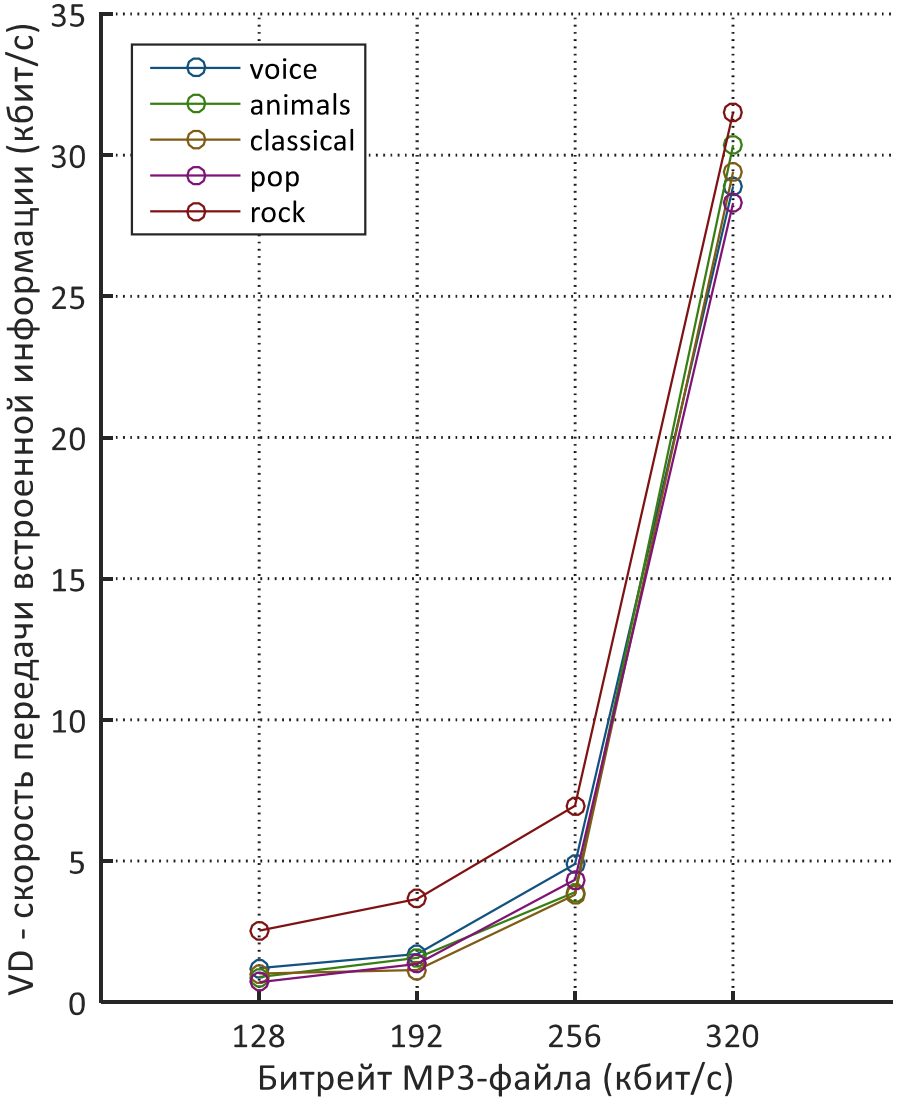


График зависимости скорости передачи встроенной информации от битрейта MP3-файла для разных категорий данных.

Скорость передачи встроенной информации для метода НЗБ: 44,1 кбит/с, для метода "Patchwork" меньше 12 бит/с.

Экспериментальное исследование. Выводы (применительно к защите MP3)

Сравнительная таблица для исследованных методов:

Методы	Преимущества	Недостатки	Применение
НЗБ	максимальный объем данных	низкая точность извлечения	стеганографический канал связи с применением помехоустойчивого кодирования
"Patchwork"	настраивается на выигрыш в точности и/или качестве аудио по сравнению с НЗБ	небольшой объем данных	защита авторских прав, защита от несанкционированного распространения
Формато-зависимый	неизменное качество аудио и абсолютная точность извлечения	легко разрушить и обнаружить встроенную информацию	защита от модификаций контейнера, стеганографический канал связи