

ПРЕДСТАВЛЕНИЕ В КОМПЬЮТЕРЕ НЕЧИСЛОВОЙ ИНФОРМАЦИИ

19.12.2010

ДВОИЧНОЕ КОДИРОВАНИЕ СИМВОЛЬНОЙ ИНФОРМАЦИИ





При двоичном кодировании
текстовой информации
каждому символу ставится в
соответствие своя
уникальная
последовательность из
восьми нулей и единиц, свой
уникальный код
от 00000000 до 11111111
(десятичный код от 0 до 255)

Присвоение символу конкретного двоичного кода - это вопрос соглашения, которое фиксируется в кодовой таблице. Первые 33 кода (с 0 до 32) соответствуют не символам, а операциям (перевод строки, ввод пробела и т.д.). Коды 33 до 127 являются интернациональными и соответствуют символам латинского алфавита, цифрам, знакам арифметических операций и знакам препинания.



Коды с 128 по 255 являются национальными, т.е. в национальных кодировках одному и тому же коду соответствуют различные символы. К сожалению, в настоящее время существует 5 различных кодовых таблиц для русских букв, поэтому тексты созданные в одной кодировке, не будут правильно отображаться в другой.





Хронологически одним из первых стандартов кодирования русских букв на компьютерах был код КОИ - 8 («Код обмена информационный - 8 битный»). Эта кодировка применяется в компьютерах с операционной системой UNIX.



Наиболее распространенная кодировка - это стандартная кириллическая кодировка **Microsoft Windows**, обозначаемая сокращением **CP1251** («CP» означает «Code Page»). Все Windows - приложения, работающие с русским языком, поддерживают эту кодировку.



Для работы в среде операционной системы **MS-DOS** используется «альтернативная» кодировка, в терминологии фирмы Microsoft - кодировка **CP 866**.

Фирма Apple разработала для компьютеров **Macintosh** свою собственную кодировку русских букв (**Mac**)





Международная организация по стандартизации (International Standards Organization, ISO) утвердила в качестве стандарта для русского языка еще одну кодировку под названием **ISO 8859 - 5**.

Стандарты кодировок:

1. КОИ-8 - UNIX
2. CP1251 («CP» означает «Code Page»)
- Microsoft Windows
3. CP 866 - MS-DOS
4. Mac - Macintosh
5. ISO 8859 - 5

ТАБЛИЦА КОДИРОВКИ СИМВОЛОВ

| Двоичный код | Десятичный код | КОИ8 | CP1251 | CP866 | Mac | ISO |
|--------------|----------------|---|--------|---------------|---------------|-----|
| 0000 0000 | 0 | | | | | |
| | | | | | | |
| 0000 1000 | 8 | Удаление последнего символа (клавиша Backspace) | | | | |
| | | | | | | |
| 0000 1101 | 13 | Перевод строки (клавиша Enter) | | | | |
| | | | | | | |
| 0010 0000 | 32 | Пробел | | | | |
| 0010 0001 | 33 | ! | | | | |
| | | | | | | |
| 0101 1010 | 90 | Z | | | | |
| | | | | | | |
| 0111 1111 | 127 | | | | | |
| | 128 | - | Ъ | А | А | К |
| | | | | | | |
| 1100 0010 | 194 | Б | В | - | - | Т |
| | | | | | | |
| 1100 1100 | 204 | Л | М | : | : | Ь |
| | | | | | | |
| 1101 1101 | 221 | Щ | Э | - | Ё | Н |
| | | | | | | |
| 1111 1111 | 225 | Ь | я | Нераз. пробел | Нераз. пробел | п |



В последнее время появился новый международный стандарт **Unicode**, который отводит на каждый символ не один байт, а два, и поэтому с его помощью можно закодировать не 256 символов, $2^{16}=65\ 536$ различных символов. Эту кодировку поддерживает платформа Microsoft Windows&Office97.

ТАБЛИЦА КОДИРОВКИ

Таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера (коды), называется таблицей кодировки.

Для разных типов ЭВМ используются различные кодировки. С распространением IBM PC международным стандартом стала таблица кодировки **ASCII** (American Standard Code for Information Interchange) - Американский стандартный код для информационного обмена.

| символ | 10- Б код | 2-Б код | символ | 10- Б код | 2-Б код | символ | 10-Б код | 2-Б код | символ | 10-Б код | 2-Б код |
|--------|-----------------|----------|--------|-----------------|----------|--------|-------------|----------|--------|-------------|----------|
| | 32 | 00100000 | 8 | 56 | 00111000 | P | 80 | 01010000 | h | 104 | 01101000 |
| ! | 33 | 00100001 | 9 | 57 | 00111001 | Q | 81 | 01010001 | i | 105 | 01101001 |
| " | 34 | 00100010 | : | 58 | 00111010 | R | 82 | 01010010 | j | 106 | 01101010 |
| # | 35 | 00100011 | ; | 59 | 00111011 | S | 83 | 01010011 | k | 107 | 01101011 |
| \$ | 36 | 00100100 | < | 60 | 00111100 | T | 84 | 01010100 | l | 108 | 01101100 |
| % | 37 | 00100101 | = | 61 | 00111101 | U | 85 | 01010101 | m | 109 | 01101101 |
| & | 38 | 00100110 | > | 62 | 00111110 | V | 86 | 01010110 | n | 110 | 01101110 |
| ' | 39 | 00100111 | ? | 63 | 00111111 | W | 87 | 01010111 | o | 111 | 01101111 |
| (| 40 | 00101000 | @ | 64 | 01000000 | X | 88 | 01011000 | p | 112 | 01110000 |
|) | 41 | 00101001 | A | 65 | 01000001 | Y | 89 | 01011001 | q | 113 | 01110001 |
| * | 42 | 00101010 | B | 66 | 01000010 | Z | 90 | 01011010 | r | 114 | 01110010 |
| + | 43 | 00101011 | C | 67 | 01000011 | [| 91 | 01011011 | s | 115 | 01110011 |
| , | 44 | 00101100 | D | 68 | 01000100 | \ | 92 | 01011100 | t | 116 | 01110100 |
| - | 45 | 00101101 | E | 69 | 01000101 |] | 93 | 01011101 | u | 117 | 01110101 |
| . | 46 | 00101110 | F | 70 | 01000110 | ^ | 94 | 01011110 | v | 118 | 01110110 |
| / | 47 | 00101111 | G | 71 | 01000111 | _ | 95 | 01011111 | w | 119 | 01110111 |
| 0 | 48 | 00110000 | H | 72 | 01001000 | ` | 96 | 01100000 | x | 120 | 01111000 |
| 1 | 49 | 00110001 | I | 73 | 01001001 | a | 97 | 01100001 | y | 121 | 01111001 |
| 2 | 50 | 00110010 | J | 74 | 01001010 | b | 98 | 01100010 | z | 122 | 01111010 |
| 3 | 51 | 00110011 | K | 75 | 01001011 | c | 99 | 01100011 | { | 123 | 01111011 |
| 4 | 52 | 00110100 | L | 76 | 01001100 | d | 100 | 01100100 | | 124 | 01111100 |
| 5 | 53 | 00110101 | M | 77 | 01001101 | e | 101 | 01100101 | } | 125 | 01111101 |
| 6 | 54 | 00110110 | N | 78 | 01001110 | f | 102 | 01100110 | ~ | 126 | 01111110 |
| 7 | 55 | 00110111 | O | 79 | 01001111 | g | 103 | 01100111 | □ | 127 | 01111111 |

ТАБЛИЦА
РАСШИРЕННО
ГО КОДА ASCII

| символ | 10-й код | 2-й код | символ | 10-й код | 2-й код | символ | 10-й код | 2-й код | символ | 10-й код | 2-й код |
|--------|----------|----------|--------|----------|----------|--------|----------|----------|--------|----------|----------|
| Б | 128 | 10000000 | | 160 | 10100000 | А | 192 | 11000000 | а | 224 | 11100000 |
| Г | 129 | 10000001 | Ў | 161 | 10100001 | Б | 193 | 11000001 | б | 225 | 11100001 |
| , | 130 | 10000010 | ў | 162 | 10100010 | В | 194 | 11000010 | в | 226 | 11100010 |
| ı | 131 | 10000011 | Ј | 163 | 10100011 | Г | 195 | 11000011 | г | 227 | 11100011 |
| „ | 132 | 10000100 | о | 164 | 10100100 | Д | 196 | 11000100 | д | 228 | 11100100 |
| … | 133 | 10000101 | Ѓ | 165 | 10100101 | Е | 197 | 11000101 | е | 229 | 11100101 |
| † | 134 | 10000110 | ı | 166 | 10100110 | Ж | 198 | 11000110 | ж | 230 | 11100110 |
| ‡ | 135 | 10000111 | § | 167 | 10100111 | З | 199 | 11000111 | з | 231 | 11100111 |
| € | 136 | 10001000 | Є | 168 | 10101000 | И | 200 | 11001000 | и | 232 | 11101000 |
| ‰ | 137 | 10001001 | © | 169 | 10101001 | Й | 201 | 11001001 | й | 233 | 11101001 |
| Љ | 138 | 10001010 | € | 170 | 10101010 | К | 202 | 11001010 | к | 234 | 11101010 |
| ‹ | 139 | 10001011 | « | 171 | 10101011 | Л | 203 | 11001011 | л | 235 | 11101011 |
| Њ | 140 | 10001100 | ¬ | 172 | 10101100 | М | 204 | 11001100 | м | 236 | 11101100 |
| Ќ | 141 | 10001101 | - | 173 | 10101101 | Н | 205 | 11001101 | н | 237 | 11101101 |
| Ћ | 142 | 10001110 | ® | 174 | 10101110 | О | 206 | 11001110 | о | 238 | 11101110 |
| Ќ | 143 | 10001111 | İ | 175 | 10101111 | П | 207 | 11001111 | п | 239 | 11101111 |
| ђ | 144 | 10010000 | ° | 176 | 10110000 | Р | 208 | 11010000 | р | 240 | 11110000 |
| ‘ | 145 | 10010001 | ± | 177 | 10110001 | С | 209 | 11010001 | с | 241 | 11110001 |
| ’ | 146 | 10010010 | ı | 178 | 10110010 | Т | 210 | 11010010 | т | 242 | 11110010 |
| “ | 147 | 10010011 | ı | 179 | 10110011 | У | 211 | 11010011 | у | 243 | 11110011 |
| ” | 148 | 10010100 | г | 180 | 10110100 | Ф | 212 | 11010100 | ф | 244 | 11110100 |
| • | 149 | 10010101 | и | 181 | 10110101 | Х | 213 | 11010101 | х | 245 | 11110101 |
| – | 150 | 10010110 | ¶ | 182 | 10110110 | Ц | 214 | 11010110 | ц | 246 | 11110110 |
| — | 151 | 10010111 | · | 183 | 10110111 | Ч | 215 | 11010111 | ч | 247 | 11110111 |
| □ | 152 | 10011000 | ë | 184 | 10111000 | Ш | 216 | 11011000 | ш | 248 | 11111000 |
| ™ | 153 | 10011001 | № | 185 | 10111001 | Щ | 217 | 11011001 | щ | 249 | 11111001 |
| љ | 154 | 10011010 | € | 186 | 10111010 | Ъ | 218 | 11011010 | ъ | 250 | 11111010 |
| › | 155 | 10011011 | » | 187 | 10111011 | Ы | 219 | 11011011 | ы | 251 | 11111011 |
| њ | 156 | 10011100 | ј | 188 | 10111100 | Ь | 220 | 11011100 | ь | 252 | 11111100 |
| ќ | 157 | 10011101 | š | 189 | 10111101 | Э | 221 | 11011101 | э | 253 | 11111101 |
| ћ | 158 | 10011110 | s | 190 | 10111110 | Ю | 222 | 11011110 | ю | 254 | 11111110 |
| џ | 159 | 10011111 | ï | 191 | 10111111 | Я | 223 | 11011111 | я | 255 | 11111111 |

ОБРАТИТЕ ВНИМАНИЕ!



Цифры кодируются по стандарту ASCII в двух случаях - при вводе-выводе и когда они встречаются в тексте. Если цифры участвуют в вычислениях, то осуществляется их преобразование в другой двоичный код.

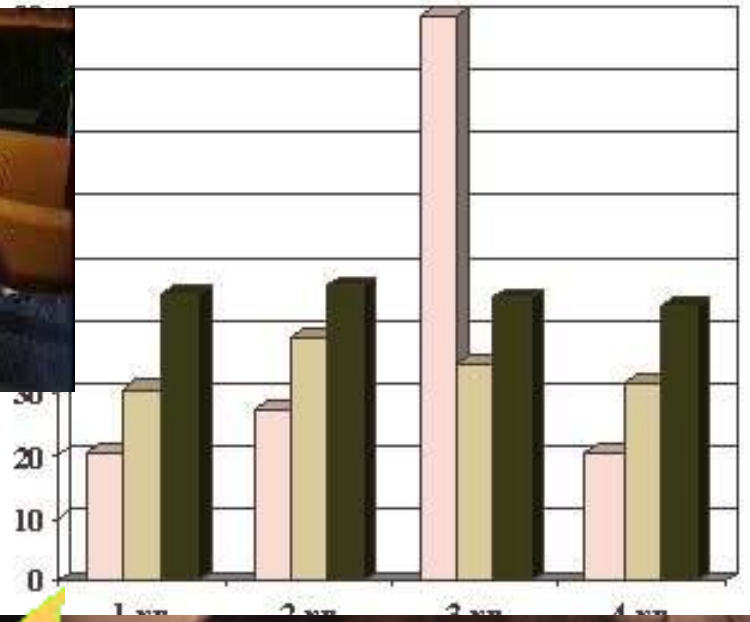
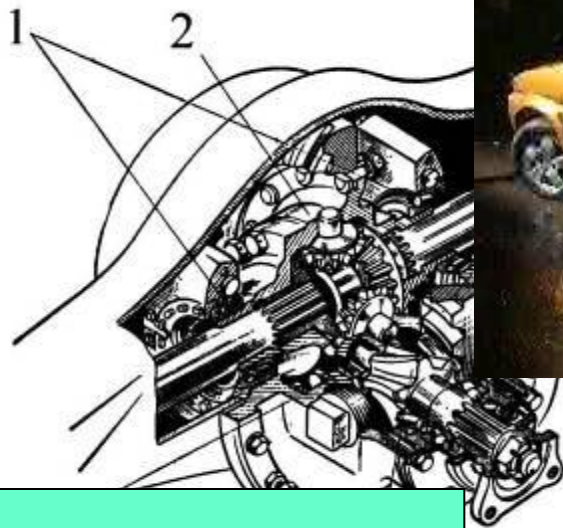
Возьмем число **57**.

При использовании в тексте каждая цифра будет представлена своим кодом в соответствии с таблицей ASCII. В двоичной системе это - **00110101 00110111**.

При использовании в вычислениях код этого числа будет получен по правилам перевода в двоичную систему и получим - **00111001**.

ДВОИЧНОЕ КОДИРОВАНИЕ ГРАФИЧЕСКОЙ И ЗВУКОВОЙ ИНФОРМАЦИИ

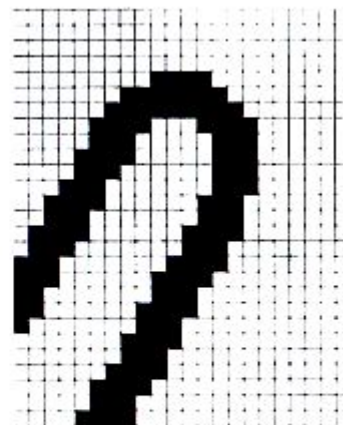




а) кодирование черно-белого изображения:

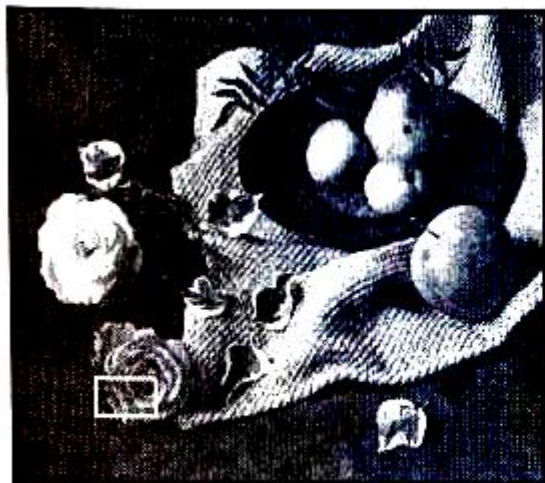


фрагмент для кодирования
подсказка: рассмотрим фрагмент



ТСЯ

б) кодирование цветного изображения или изображения с полутонами:





Фрагмент





Увеличенный фрагмент



ТЬ

Монохромное изображение (черно-белый монитор)

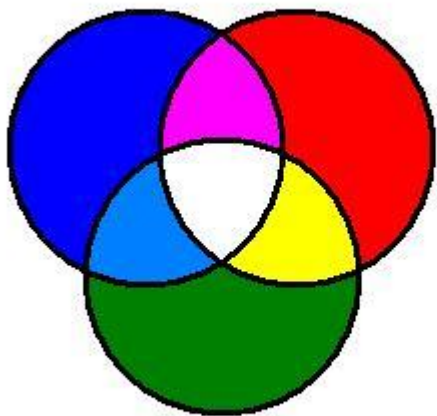
| | | |
|---|---|----------------------|
|  | 0 | 1 бит видеопамяти |
|  | 1 | |

| | | |
|---|----|-----------------------|
|  | 00 | 2 бита видеопамяти |
|  | 01 | |
|  | 10 | |
|  | 11 | |

Цветные изображения могут иметь различную глубину цвета (бит на точку 4, 8, 16, 24). Каждый цвет можно рассматривать как возможные состояния точки, и тогда по формуле $N=2^I$ может быть вычислено количество цветов отображаемых на экране монитора.

| Глубина цвета I | Количество отображаемых цветов N |
|------------------------|---|
| 4 | $2^4=16$ |
| 8 | $2^8=256$ |
| 16 (High Color) | $2^{16}=65\ 536$ |
| 24 (True Color) | $2^{24}=16\ 777\ 216$ |

Аддитивная модель
RGB (сложение цветов)



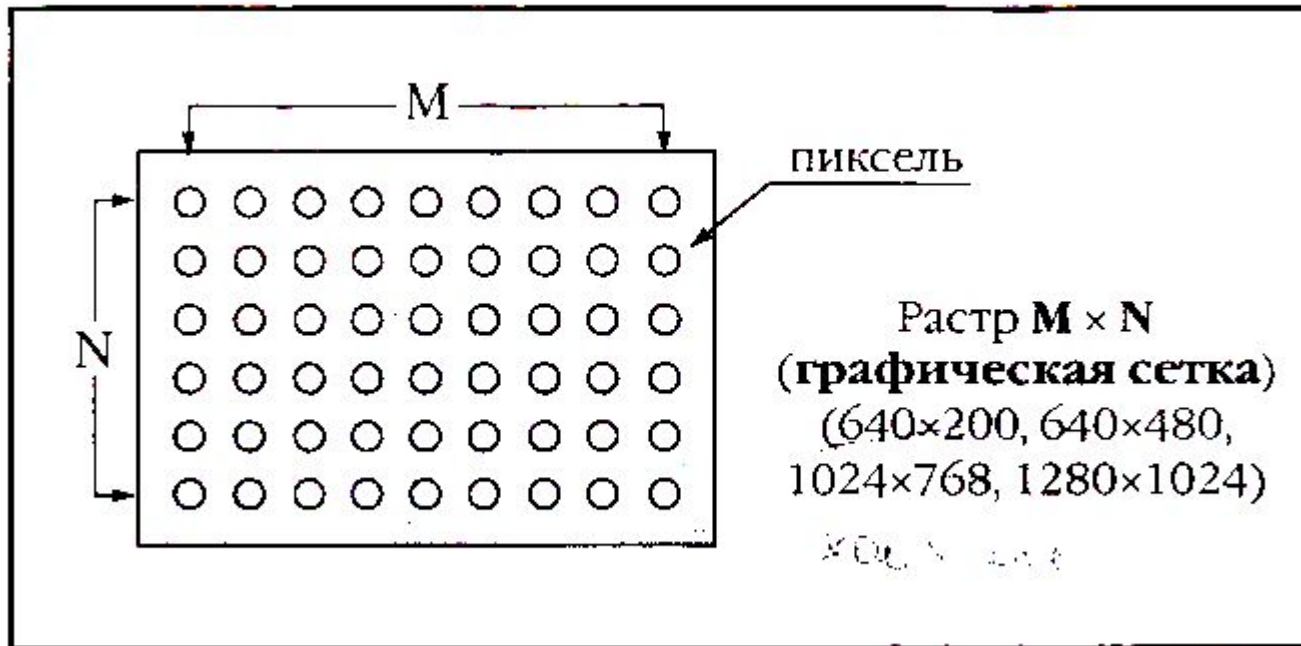
Восьмицветная палитра (на основе базовых цветов)

| R | G | B | Цвет |
|---|---|---|------------|
| 0 | 0 | 0 | черный |
| 0 | 0 | 1 | синий |
| 0 | 1 | 0 | зеленый |
| 0 | 1 | 1 | голубой |
| 1 | 0 | 0 | красный |
| 1 | 0 | 1 | розовый |
| 1 | 1 | 0 | коричневый |
| 1 | 1 | 1 | белый |

Шестнадцатичетная палитра (И - бит интенсивности)

| И | R | G | B | Цвет |
|---|---|---|---|--------------|
| 0 | 0 | 0 | 0 | черный |
| 0 | 0 | 0 | 1 | синий |
| 0 | 0 | 1 | 0 | зеленый |
| 0 | 0 | 1 | 1 | голубой |
| 0 | 1 | 0 | 0 | красный |
| 0 | 1 | 0 | 1 | розовый |
| 0 | 1 | 1 | 0 | коричневый |
| 0 | 1 | 1 | 1 | серый |
| 1 | 0 | 0 | 0 | темно-серый |
| 1 | 0 | 0 | 1 | ярко-синий |
| 1 | 0 | 1 | 0 | ярко-зеленый |
| 1 | 0 | 1 | 1 | ярко-голубой |
| 1 | 1 | 0 | 0 | ярко-красный |
| 1 | 1 | 0 | 1 | ярко-розовый |
| 1 | 1 | 1 | 0 | ярко-желтый |
| 1 | 1 | 1 | 1 | белый |

Изображение может иметь различный размер, которое определяется количеством точек по горизонтали и вертикали.



В современных ПК обычно используются 4 основных размера изображения или разрешающих способностей экрана: 640x480, 800x600, 1024x768, 1280x1024 пикселя.



Графический режим вывода изображения на экран определяется разрешающей способностью экрана и глубиной (интенсивностью) цвета.

Полная информация о всех точках изображения, хранящаяся в видеопамяти, называется битовой картой изображения.

Для того чтобы на экране монитора формировалось изображение, информация о каждой его точке храниться в видео памяти ПК.

Рассчитаем объем видеопамяти для наиболее распространенного в настоящее время графического режима (800x600 точек, 16 бит на точку):

Всего точек на экране **800x600 = 480 000**
точек

480 000x16 бит = 7 680 000 бит = 960 000
байт = 937,5 Кбайт \approx **938 Кбайт**

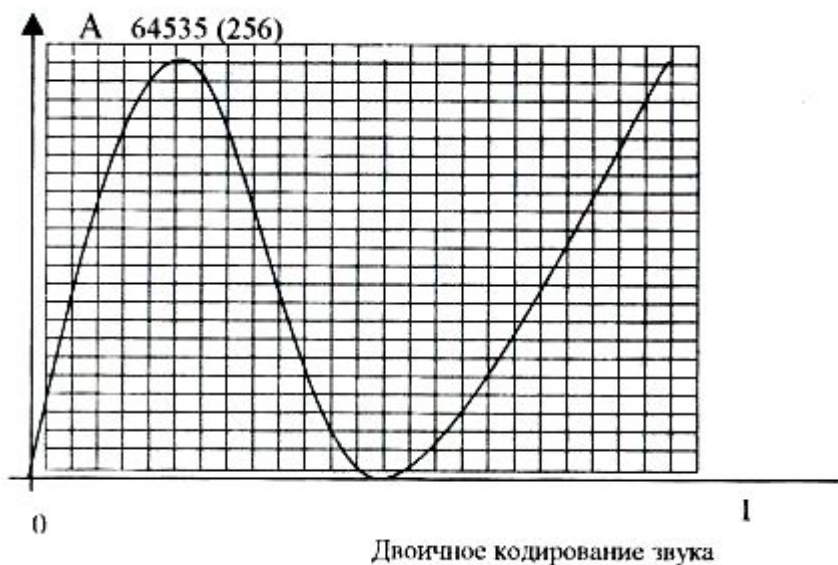
С начала 90-х годов ПК получили возможность работать со звуковой информацией. Каждый ПК, имеющий звуковую плату, микрофон, наушники или колонки, может записывать, сохранять и воспроизводить звуковую информацию.

С графической информацией мы работаем посредством графических редакторов, то со звуковой информацией с помощью редакторов аудиофайлов.



Звуковой сигнал - это непрерывная волна с изменяющейся амплитудой и частотой.

При двоичном кодировании непрерывного звукового сигнала он заменяется серией его отдельных выборок - отсчетов.

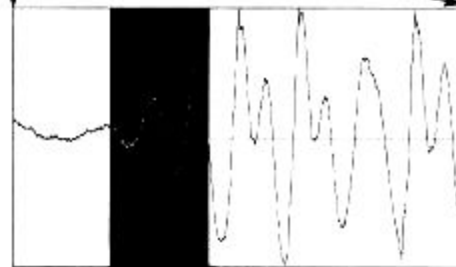


в) звук:

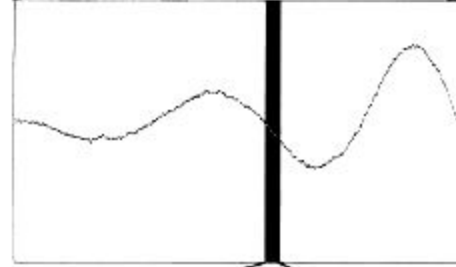
Перед вами фонограмма слова «Мама»:



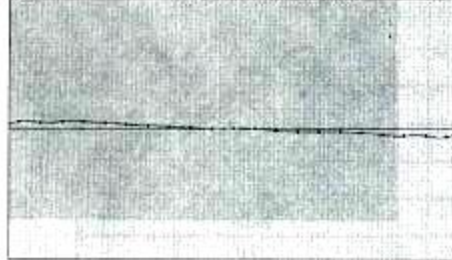
выделим фрагмент
и увеличим масштаб



выделим фрагмент
и увеличим масштаб



выделим фрагмент
и увеличим масштаб



Современные звуковые карты могут обеспечить кодирование 65 536 различных уровней сигнала или состояний. Для определения количества бит, необходимых для кодирования, решим показательное уравнение:

$$65\,536 = 2^l, \text{ то } l = 16 \text{ бит.}$$

Таким образом, современные звуковые карты обеспечивают 16-битное кодирование звука. При каждой выборке значению амплитуды звукового сигнала присваивается 16 битный код.

Количество выборок в секунду может быть в диапазоне от 8 000 до 48 000, т.е. Частота дискретизации аналогового звукового сигнала может принимать значения от 8 до 48 КГц.

При частоте 8 КГц качество дискретизированного звукового сигнала соответствует качеству радиотрансляции, а при частоте 48 КГц - качеству звучания аудио-CD. Следует учитывать, что возможны как моно- так стерео- режимы.

Можно оценить информационный объем моноаудиофайла длительностью звучания 1 секунду при среднем качестве звука (16 бит, 24 КГц). Для этого количество бит на одну выборку необходимо умножить на количество выборок в 1 секунду:

$$16 \text{ бит} * 24\,000 = 384\,000 \text{ бит} = 48\,000 \text{ байт} = 46,875 \text{ Кбайт}$$