

**ФГКОУ «МОСКОВСКОЕ СУВОРОВСКОЕ ВОЕННОЕ УЧИЛИЩЕ»
МИНИСТЕРСТВА ОБОРОНЫ РОССИЙСКОЙ ФЕДЕРАЦИИ**

Представление текстовой, графической и звуковой информации

**Разработала преподаватель
информатики и ИКТ
Барина Лидия Анатольевна**



Москва 2014 год

I. Представление текстовой информации



Для обработки текстовой информации компьютер использует двоичную систему счисления, состоящую из двух цифр: 0 и 1.

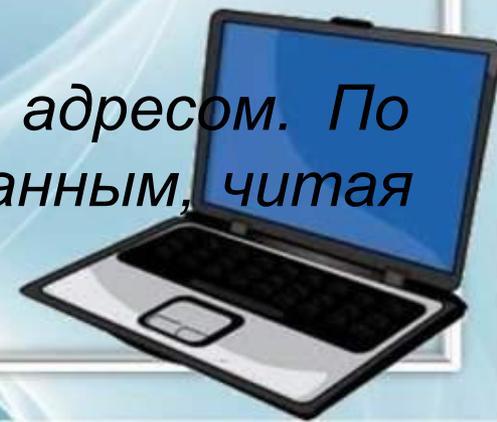
В компьютере используется *байтовый принцип организации памяти*: каждая клеточка – бит памяти.



Бит – это наименьшая единица измерения количества информации, принимающая значение 1 или 0.

Восемь подряд расположенных битов образуют байт.

Порядковый номер байта является его адресом. По этим адресам процессор обращается к данным, читая и записывая их в память.

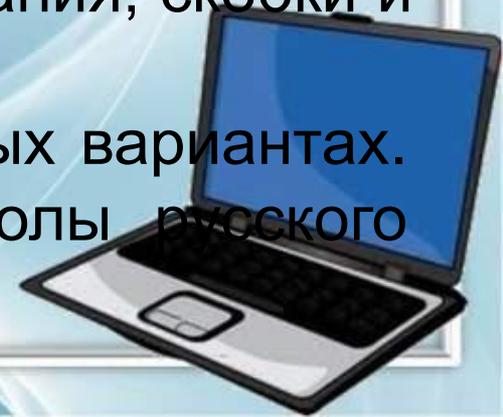


В популярных системах кодировки (ASCII и др.) каждый символ заменяется на 8-разрядное целое положительное двоичное число, т.е. *1 символ компьютерного алфавита «весит» 8 битов.*

Согласно формуле информатики $2^i=N$ можно закодировать: $2^8 = 256$ символов.

Присвоение символу конкретного кода определено таблицей кодировки ASCII. Таблица кодов ASCII делится на две части. Международным стандартом является лишь первая половина таблицы, т.е. символы с номерами от 0 (00000000), до 127 (01111111). Сюда входят буква латинского алфавита, цифры, знаки препинания, скобки и некоторые другие символы.

Остальные 128 кодов используются в разных вариантах. В русских кодировках размещаются символы русского алфавита.



Первая половина таблицы кодов 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
	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



Вторая половина таблицы кодов ASCII

символ	10-Б код	2-Б код									
Ъ	128	10000000	Щ	160	10100000	А	192	11000000	а	224	11100000
Г	129	10000001	Щ̄	161	10100001	Б	193	11000001	б	225	11100001
,	130	10000010	щ	162	10100010	В	194	11000010	в	226	11100010
г	131	10000011	Ј	163	10100011	Г	195	11000011	г	227	11100011
..	132	10000100	о	164	10100100	Д	196	11000100	д	228	11100100
...	133	10000101	Г̄	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
Ѡ	142	10001110	®	174	10101110	О	206	11001110	о	238	11101110
ѡ	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	І	178	10110010	Т	210	11010010	т	242	11110010
“	147	10010011	і	179	10110011	У	211	11010011	у	243	11110011
”	148	10010100	ı	180	10110100	Ф	212	11010100	ф	244	11110100
•	149	10010101	µ	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



В настоящее время получил широкое распространение новый международный стандарт Unicode, который отводит на каждый символ 2 байта. С его помощью можно закодировать ($2^{16} = 65536$) различных символов.

**Пример
представления
текста в памяти
компьютера:**

Слова	Память
<u>file</u>	01100110 01101001 01101100 01100101
<u>disk</u>	01100100 01101001 01110011 01101011



Кодирование текстовой информации

Кодирование – преобразование входной информации в машинную форму (в двоичный код).

Декодирование – преобразование двоичного кода в форму, понятную человеку.

Число символов алфавита (мощность алфавита) : $N = 2^i$

где **i** - информационный вес одного символа;

Информационный объем текста: $I = i * K$

i – информационный вес одного символа (*количество бит на кодирование одного символа*);
K - ЧИСЛО СИМВОЛОВ В ТЕКСТЕ.

Объём данных при передаче: $I = V * t$,

где **t** — время передачи данных,

V — скорость передачи данных.

Справочная таблица единиц измерения объема информации:

<i>Название</i>	<i>Соотношение с другими единицами</i>
Бит	Двоичная цифра (0 или 1)
Байт	8 бит
Килобайт	1 Кб = 2^{10} байт = 1024 байт
Мегабайт	1 Мб = 2^{10} Кб = 2^{20} байт
Гигабайт	1 Гб = 2^{10} Мб = 2^{20} Кб = 2^{30} байт
Терабайт	1 Тб = 2^{10} Гб = 2^{20} Мб = 2^{30} Кб = 2^{40} байт



Задача 1.

В одном из вариантов кодировки Unicode каждый символ кодируется двумя байтами. Определите количество символов в сообщении, если информационный объём сообщения в этой кодировке равен 480 бит.

1) 60 2) 40 3) 240 4) 30

Решение.

1). Количество символов: $K = I / i = 480 / 16 = 30$

Ответ: 4) 30

Задача 2.

Информационное сообщение объёмом 5 Кбайт содержит 8192 символа. Сколько символов содержит алфавит, при помощи которого было записано это сообщение?

Решение:

1). $I = i * k$

Объём сообщения: $I = 5 \text{ Кб} = 5120 \text{ байт} = 40960 \text{ бит}$

2). Сообщение содержит 8192 символа, следовательно:

$i = I / k$ $i = 40960 : 8192 = 5 \text{ бит}$ (вес одного символа).

3). $N = 2^5 = 32$

Ответ: 32



II. Представление графической информации



Аналоговый и дискретный способы представления изображений и звука

Графическая и звуковая информация может быть представлена в **аналоговой** или **дискретной** форме.

- При **аналоговом представлении** физическая величина принимает бесконечное множество значений, которые изменяются непрерывно.
- При **дискретном представлении** физическая величина принимает конечное множество значений.

Дискретизация – это преобразование непрерывных изображений и звука в набор дискретных значений в форме двоичных кодов.



Графическая информация

Вся графическая информация в компьютере представлена в виде растровой и векторной графики.

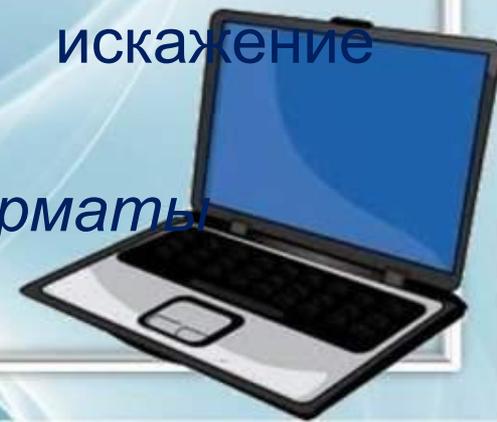
*В **растровой графике** точечный рисунок состоит из пикселей.*

Графические редакторы растрового типа в основном ориентированы не для создания изображений, а на их обработку.

***Достоинство** растровой графики – эффективное представление изображений фотографического качества.*

***Недостаток** – большой объем памяти, искажение изображения при его масштабировании.*

*Растровые графические файлы имеют форматы **JPEG, BMP, TIFF** и др.*



Графическая информация

В **векторной графике** графическая информация – это графические примитивы, составляющие рисунок: *прямые, дуги, прямоугольники и пр.*

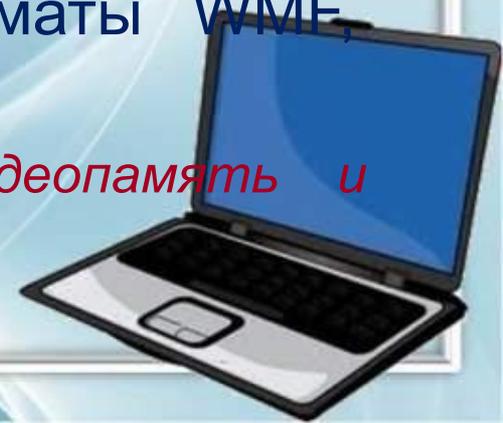
Рисунок представляется в системе экранных координат.

Достоинство векторной графики – небольшой объем памяти и масштабирование изображений без потери качества.

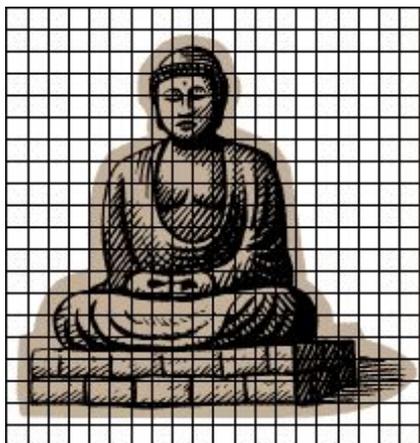
Векторные графические редакторы предназначены для создания оформительских, чертежных и проектно-конструкторских работ.

Векторные графические файлы имеют форматы WMF, CGM.

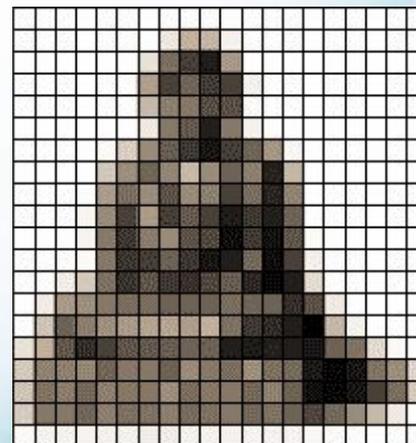
Все графические данные, помещаемые в видеопамять и выводимые на экран, имеют растровый формат.



Растровое кодирование по шагам:



Шаг 1. Дискретизация:
разбивка на *пиксели*.



Шаг 2. Для каждого пикселя
определяется **единый цвет**.

Пиксель – это наименьший элемент рисунка, для которого можно установить цвет.

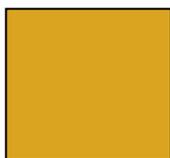


Растровое кодирование (True Color)

Шаг 3. От цвета – к числам: модель RGB

цвет = **R** + **G** + **B**

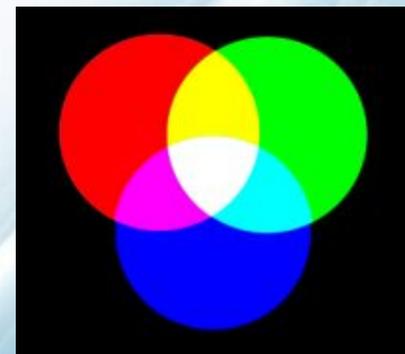
<i>red</i>	<i>green</i>	<i>blue</i>
красный	зеленый	синий
0..255	0..255	0..255



R = 218
G = 164
B = 32



R = 135
G = 206
B = 250



Шаг 4. Числа – в двоичную систему.

Каждый цвет (красный, зеленый и синий) имеет 256 уровней интенсивности. Поэтому можно закодировать:

256 · 256 · 256 = 256³ = 16 777 216 цветов

Количество памяти для хранения цвета 1 (точки) пикселя?

R: 256 = 2⁸ вариантов, нужно 8 бит = 1 байт

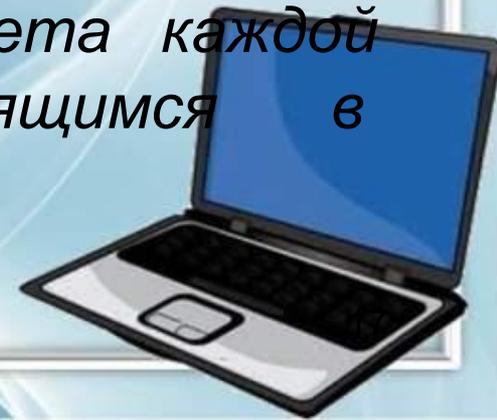
R G B: 256³ = 2²⁴ вариантов, нужно 24 бита = 3 байта



Двоичное кодирование графической информации

В современных компьютерах используются 3 разрешающие способности экрана: **800x600**, **1024x768** и **1280x1024** точки.

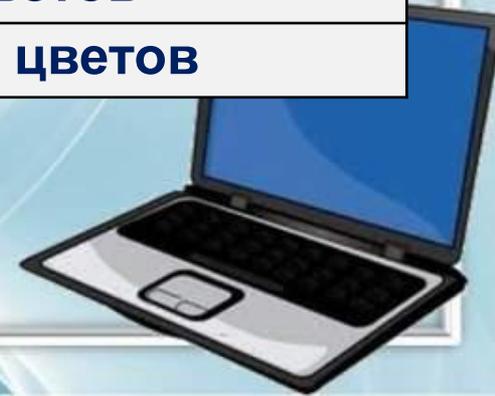
- Для черно-белого изображения, без градаций серого цвета, каждая точка экрана может иметь одно из двух состояний – «**черная**» или «**белая**», то есть для хранения ее состояния необходим **1 бит**.
- Цветные изображения формируются в соответствии с двоичным кодом цвета каждой точки (**глубиной цвета**), хранящимся в видеопамяти.



Двоичное кодирование графической информации

Глубина цвета (i) – это количество битов, которое используется для кодирования цвета одной точки.

Глубина цвета одной точки (i)	Количество отображаемых цветов (N)
3 бита	$2^3 = 8$ цветов
4 бита	$2^4 = 16$ цветов
8 битов	$2^8 = 256$ цветов
16 битов	$2^{16} = 65536$ цветов
24 бита	$2^{24} = 16777216$ цветов
32 бита	$2^{32} = 4294967296$ цветов



Двоичное кодирование графической информации

Количество цветов на экране:

$$N = 2^i$$

где i - глубина цвета одной точки,

N - количество цветов.

Объем видеопамяти:

$$I = K \cdot i$$

где i глубина цвета точки,
 $K = X \cdot Y$, где X, Y – размеры экрана по горизонтали и по вертикали



Задача 1. Рассчитать необходимый объем видеопамяти для графического режима с разрешением 800x600 точек и глубиной цвета 24 бита на одну точку.

Решение:

1). Всего точек на экране: $K = 800 \cdot 600 = 480000$.

2). Необходимый объем памяти: $I = 24 \text{ бит} \cdot 480000 = 11520000 \text{ бит} = 1440000 \text{ байт} = 1406,25 \text{ Кб} = 1,37 \text{ Мб}$.

Ответ: 1,37 Мб.



Задача 2:

Сколько секунд потребуется для передачи сообщения со скоростью $V=14400$ бит/сек, чтобы передать цветное изображение размером 800×600 пикселей, при условии, что в палитре $N=65536$ цветов?

Решение:

- 1). При $N = 65536$, $i = 2$ байта = 16 битов.
- 2). Общее количество пикселей в изображении: $K = 800 \times 600 = 480000$.
- 3). Вычислим объем памяти для 480000 пикселей:

$$I = K * i = 480\,000 * 2 = 960\,000 \text{ байт} * 8 = 7\,680\,000 \text{ бит.}$$

- 4) Найдем время передачи сообщения: $t = I / V =$
 $7\,680\,000 / 14\,400 \approx 533$ секунд.

Ответ: 533 секунды.



III. Представление звуковой информации



Двоичное кодирование звуковой информации

Звук представляет собой непрерывный сигнал — звуковую волну с меняющейся амплитудой и частотой.

Чем больше амплитуда сигнала, тем он громче для человека.

Чем больше частота сигнала, тем выше тон.

Частота звуковой волны выражается числом колебаний в секунду и измеряется в герцах (Гц).



Двоичное кодирование звуковой информации

Ввод звука в компьютер производится с помощью звукового устройства (микрофон и др.), выход которого подключается к порту звуковой карты.

Звуковая карта производит измерения уровня звукового сигнала (преобразованного в электрические колебания) и результаты записывает в память компьютера в виде последовательности электрических импульсов (двоичных нулей и единиц). Этот процесс называется **оцифровкой звука**.

Промежуток времени между двумя измерениями называется периодом измерений **T** (сек).

Обратная величина **$K=1/T$** (герц) называется частотой дискретизации.



Двоичное кодирование звуковой информации

Оцифровка (перевод в цифровую форму):



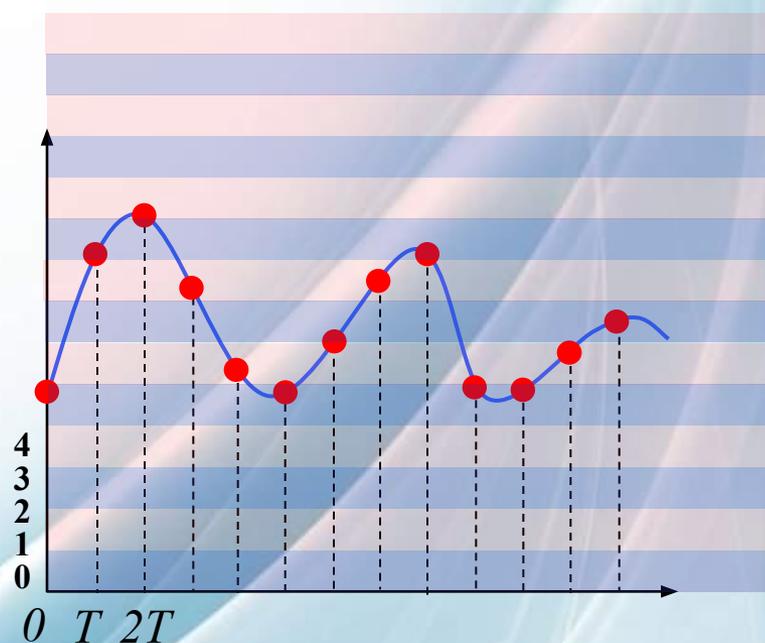
Дискретизация по уровню звука:

У всех точек в одной полосе одинаковый код!

8 бит = $2^8 = 256$ уровней

16 бит = $2^{16} = 65536$ уровней

24 бита = 2^{24} уровней



*«Глубина» кодирования
(разрядность звуковой карты)*



Двоичное кодирование звуковой информации

Частота дискретизации (К) — это количество измерений уровней сигнала за 1 секунду.

Количество бит, отводимое на один звуковой сигнал, называют глубиной кодирования звука (i).

Современные звуковые карты обеспечивают 16-, 32- или 64-битную глубину кодирования звука.

Качество двоичного кодирования звука определяется глубиной кодирования звука (i) и частотой дискретизации (К).

Чем больше частота и глубина дискретизации звука, тем более качественным будет звучание оцифрованного звука. Чем выше качество цифрового звука, тем больше информационный объем звукового файла.

Глубина кодирования звука (i)	Частота дискретизации (К)
16 бит	от 8 до 48 кГц. (от 8000 до 48000 Гц. в сек)
32 бита	
64 бита	

Двоичное кодирование звуковой информации

Количество уровней звука: $N = 2^i$	Информационный объём аудиофайла $I = t \cdot K \cdot i \Rightarrow t = I / (K \cdot i)$
где i - глубина звука (бит),	где t - время звучания (секунд), K - частота дискретизации (Гц), i - глубина кодирования (бит).

Задача 1: Оценить информационный объем стерео-аудиофайла длительностью звучания 1 секунда при высоком качестве звука ($i = 16$ битов, $K = 48 \text{ кГц} = 48000 \text{ Гц}$).

Решение: Для этого количество выборок в 1 секунду умножить на количество битов, приходящихся на одну выборку и умножить на 2 (стерео):

$$I = t \cdot K \cdot i = 48000 \cdot 16 \cdot 2 \text{ бит} = 1536000 \text{ бит} / 8 = 192000 \text{ байт} = 187,5 \text{ Кбайт.}$$

Ответ: 187,5 Кбайт.



Задача 2:

Производится двухканальная (стерео) звукозапись с частотой дискретизации $K=48$ кГц и $i=32$ -битным разрешением, результаты записываются в файл, сжатие данных не используется. Размер файла с записью не может превышать $I=16$ Мбайт.

Какая из приведённых ниже величин наиболее близка к максимально возможной продолжительности записи?

- 1) 17 секунд 2) 44 секунды 3) 65 секунд 4) 177 секунд

Решение:

1). Частота дискретизации $K=48$ кГц = 48 000 значений сигнала за секунду.

2). Глубина кодирования $i=32$ бит = 4 байта.

3) Поскольку запись двухканальная, объём памяти, необходимый для хранения данных одного канала, умножается на 2, поэтому, так как размер файла $I=16$ Мб, один канал занимает $8 \text{ Мб} = 8 \cdot 2^{20}$ байт.

4). Продолжительность записи $t = (8 \cdot 2^{20}) / (48000 \cdot 4) \approx 43,69$ сек.

Ответ: 2) 44 секунды



Задача 3:

Проводилась (моно) звукозапись с частотой дискретизации $K=16$ кГц = 16000 Гц и $i = 32$ бита. В результате был получен файл размером $I = 20$ Мбайт.

Какая из приведенных ниже величин наиболее близка к времени, в течение которого проводилась запись?

- 1) 1 мин 2) 2 мин 3) 5 мин 4) 10 мин

Решение.

- 1) Глубина кодирования: $i = 32$ бита = 4 байта,
- 2) Размер файла: $I = 20$ Мб = $20 \cdot 1024 \cdot 1024 = 20\,971\,520$ байт.
- 3) Время записи: $t = I / (K \cdot i) = 20971520 / (16000 \cdot 4) = 328$ секунд / 60 = 5,5 минут, что близко к 5 минутам.

Ответ: 3) 5 мин

