

Обработка и интерпретация речевого сигнала. Введение

П. А. Холявин

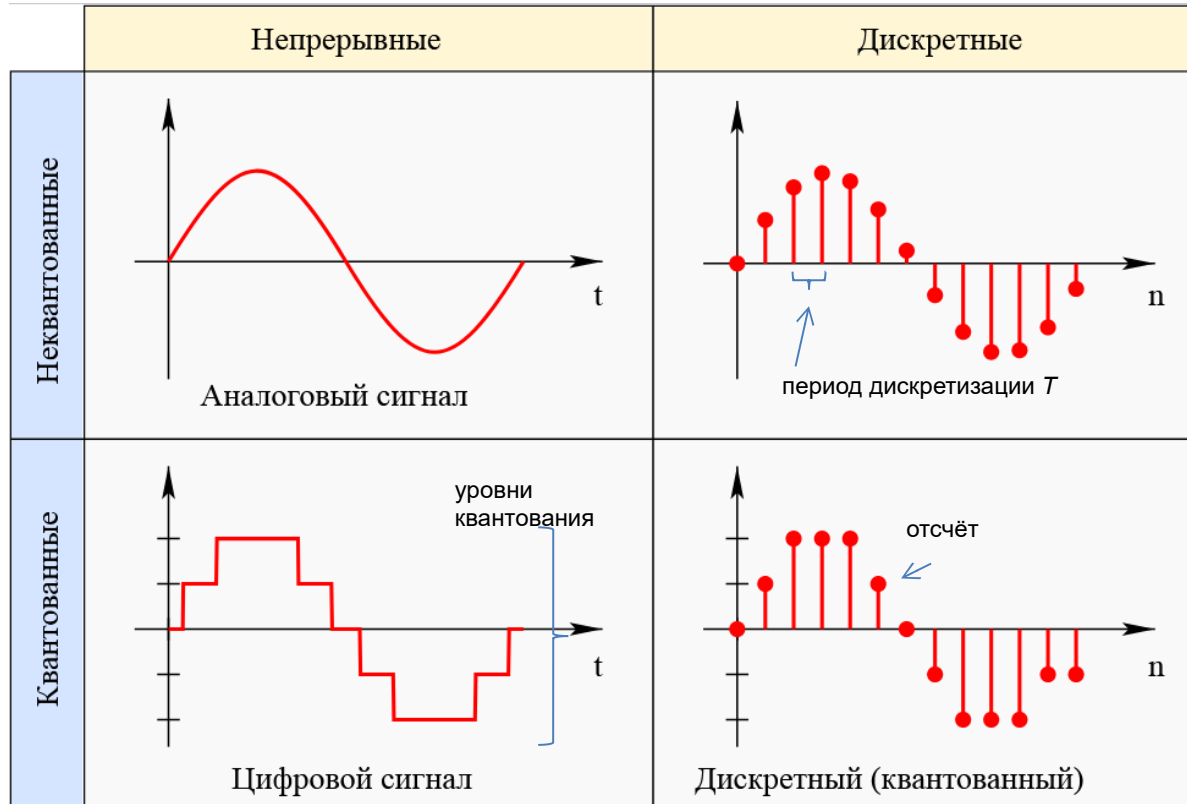
p.kholyavin@spbu.ru

05.09.2024





Аналого-цифровое преобразование

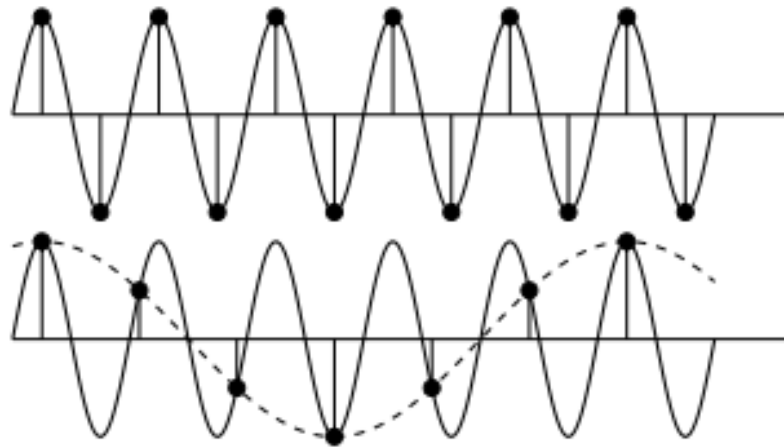


$$F_{\text{дискр}} = 1 / T$$



Теорема Котельникова

Любой сигнал $s(t)$, спектр которого не содержит составляющих с частотами выше некоторого значения f , может быть без потерь представлен в виде дискретного сигнала с частотой дискретизации $F \geq 2f$ (частота Найквиста).





Типы квантования

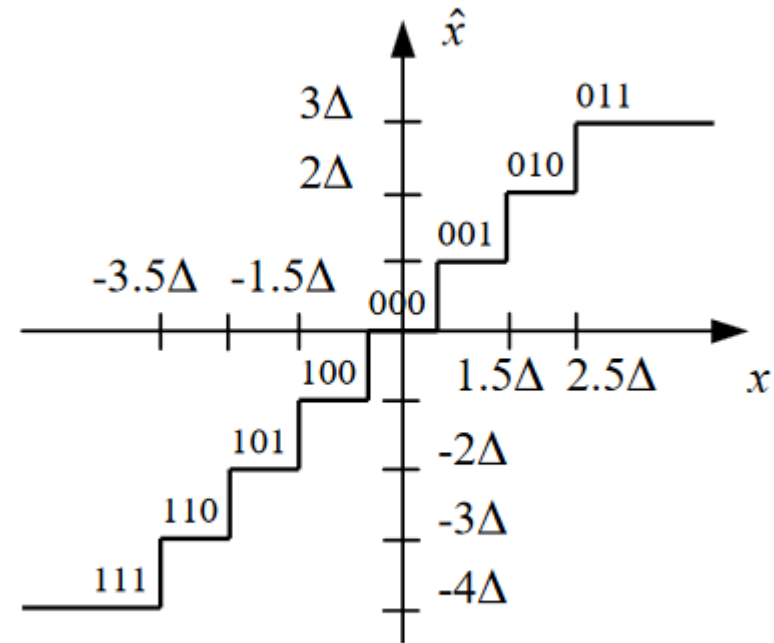
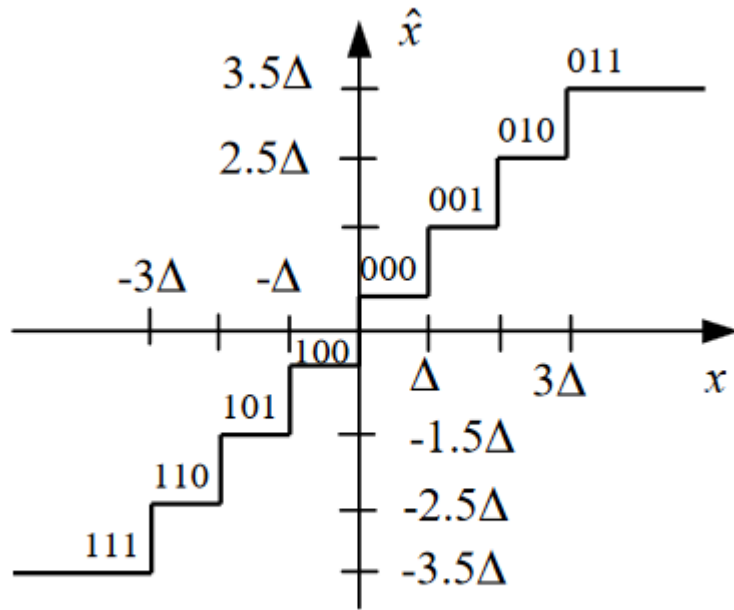


Figure 7.1 Three-bit uniform quantization characteristics: (a) mid-riser, (b) mid-tread.

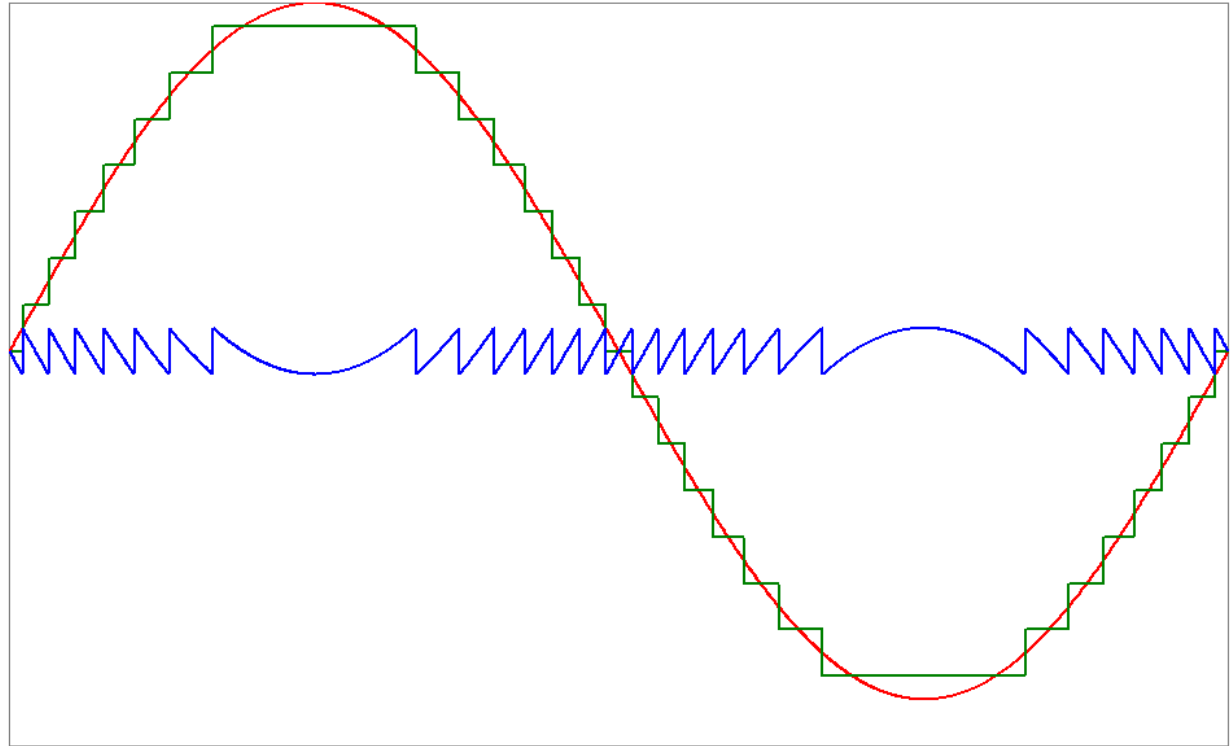
С ненулевой степенью

С нулевой степенью



Шум квантования

$$e[n] = \hat{x}[n] - x[n]$$





Кодирование речевого сигнала

1. Ограничение в частотной области:

Телефонная речь – 300-3400 Гц, ЧД = 8 кГц

Wideband speech – 50-7000 Гц, ЧД = 16 КГц

HF-аудио – ЧД = 44.1 кГц



Цифровые аудиоформаты

1. Без сжатия (WAV, AIFF, RAW)
2. Сжатие без потерь (FLAC, M4A, WMA Lossless)
3. Сжатие с потерями (MP3, AAC, WMA)



Формат .WAV

The Canonical WAVE file format

endian	File offset (bytes)	field name	Field Size (bytes)	
big	0	ChunkID	4	The "RIFF" chunk descriptor
little	4	ChunkSize	4	
big	8	Format	4	
big	12	Subchunk1 ID	4	The "fmt" sub-chunk describes the format of the sound information in the data sub-chunk
little	16	Subchunk1 Size	4	
little	20	AudioFormat	2	
little	22	NumChannels	2	
little	24	SampleRate	4	
little	28	ByteRate	4	
little	32	BlockAlign	2	
little	34	BitsPerSample	2	
big	36	Subchunk2 ID	4	
little	40	Subchunk2 Size	4	
little	44	data		The "data" sub-chunk Indicates the size of the sound information and contains the raw sound data



Формат .WAV

Типичные *ЧД* WAV:

11 025, 22 050, 44 100
8 000, 16 000, 32 000

Типичные значения *количества бит на отсчёт* WAV: 8, 16, 24, 48



Linear Pulse Code Modulation (PCM)

1. Входной сигнал ограничен
2. Шаг квантизации постоянен (равномерное квантование)

B – количество бит, тогда $N = 2^B$ – количество уровней квантования
 Δ – шаг квантования

$$\hat{x}[n] = c[n]\Delta$$

$$\hat{x}[n] = \text{sign}(c[n])\frac{\Delta}{2} + c[n]\Delta$$



Linear Pulse Code Modulation (PCM)

3. Шум квантования:

а) является белым

б) не коррелирует с сигналом

в) равномерно распределён в интервале $(-\Delta / 2, \Delta / 2)$

$$\sigma_e^2 = \frac{\Delta^2}{12} = \frac{X_{\max}^2}{3 \times 2^{2B}}$$

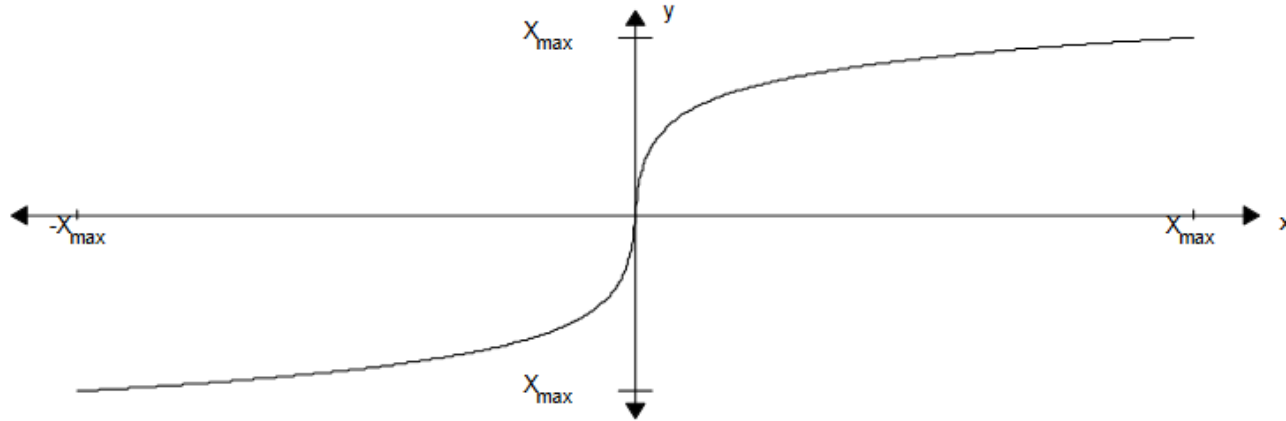
$$SNR(dB) = 10 \log_{10} \left(\frac{\sigma_x^2}{\sigma_e^2} \right) = (20 \log_{10} 2) B + 10 \log_{10} 3 - 20 \log_{10} \left(\frac{X_{\max}}{\sigma_x} \right)$$



μ -law и A-law PCM

$$y[n] = X_{\max} \frac{\log \left[1 + \mu \frac{|x[n]|}{X_{\max}} \right]}{\log[1 + \mu]} \text{sign}\{x[n]\}$$

$$y[n] = X_{\max} \frac{1 + \log \left[\frac{A|x[n]|}{X_{\max}} \right]}{1 + \log A} \text{sign}\{x[n]\}$$



$$\mu = 255$$

$$A = 87.56$$



Адаптивный РСМ

Шаг квантования пропорционален стандартному отклонению сигнала:

$$\Delta[n] = \Delta_0 \sigma[n]$$

Оно меняется медленно, поэтому шаг не нужно передавать с каждым отсчётом.

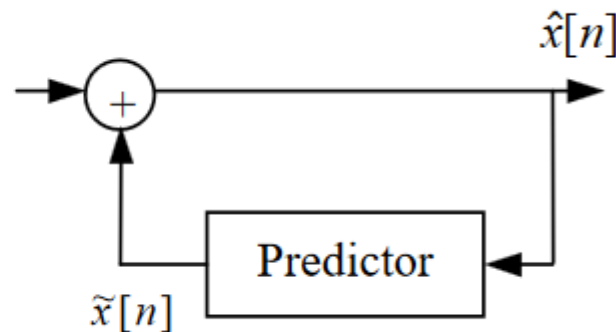
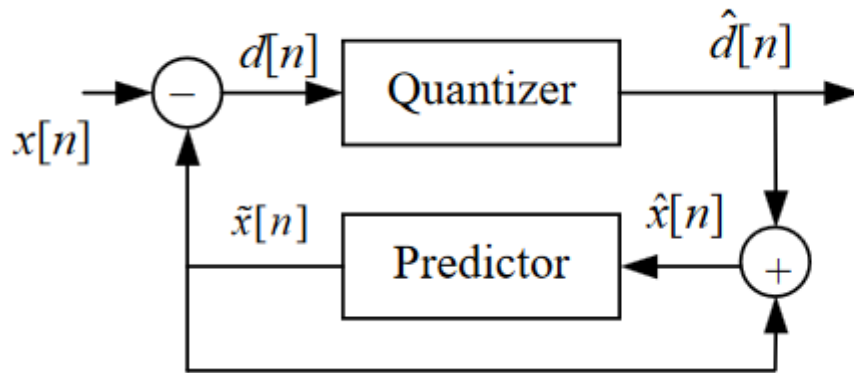


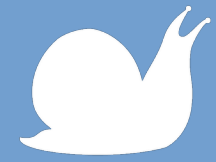
Дифференциальная квантизация (DPCM)

Квантуемая величина – разница между отсчётом и его предсказанным значением:

$$d[n] = x[n] - \tilde{x}[n]$$

$$\hat{d}[n] = Q\{d[n]\} = d[n] + e[n]$$

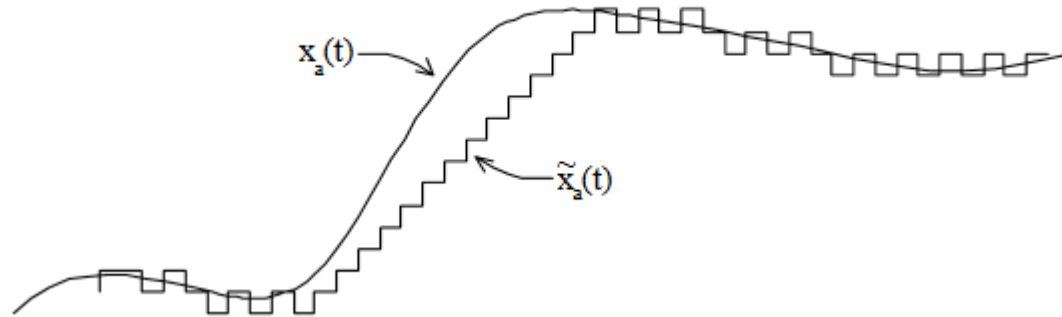




Дельта-модуляция

$$\tilde{x}[n] = x[n-1]$$

$$d[n] = \begin{cases} \Delta & x[n] > x[n-1] \\ -\Delta & x[n] \leq x[n-1] \end{cases}$$





Адаптивная дельта-модуляция

$$\Delta[n] = \begin{cases} \alpha\Delta[n-1] + k_1 & \text{if } e[n], e[n-1] \text{ and } e[n-2] \text{ have same sign} \\ \alpha\Delta[n-1] + k_2 & \text{otherwise} \end{cases}$$

$$0 < k_2 \ll k_1$$

$$0 < \alpha < 1$$



Стандарты кодирования

Standard	Bit Rate (kbits/sec)	MOS	Algorithm	Sampling Rate (kHz)
Stereo CD Audio	1411	5.0	16-bit linear PCM	44.1
WAV, AIFF, SND	Variable	-	16/8-bit linear PCM	8, 11.025, 16, 22.05, 44.1, 48
G.711	64	4.3	μ -law/A-law PCM	8
G.727	40, 32, 24, 16	4.2 (32k)	ADPCM	8
G.722	64, 56, 48		Subband ADPCM	16

Спасибо за внимание!

