10 Speech and Audio Signals

10.1 Introduction

Speech and audio signals are normally converted into PCM, which can be stored or transmitted as a PCM code, or compressed to reduce the number of bits used to code the samples. Speech generally has a much smaller bandwidth than audio.

10.2 PCM parameters

Digital systems tend to be less affected by noise than analogue. The main source of noise is quantization noise, which is caused by the finite number of quantization levels converting to a digital code.

The main parameters in determining the quality of a PCM system are the dynamic range (DR) and the signal-to-noise ratio (SNR).

10.2.1 Quantization error

The maximum error between the original level and the quantized level occurs when the original level falls exactly halfway between two quantized levels. This error will be half the smallest increment or

$$\text{Max error} = \pm\frac{1}{2}\frac{\text{Full scale}}{2^N}$$

10.2.2 Dynamic range (DR)

The dynamic range is the ratio of the largest possible signal magnitude to the smallest possible signal magnitude. If the input signal uses the full range of the ADC then the maximum signal will be the full-scale voltage. The smallest signal amplitude is one which toggles between one quantization level and the level above, or below. This signal amplitude, for an *n*-bit ADC, is the full-scale voltage divided by the number of quantization levels (that is, 2^n). Thus, for a linearly quantized signal:

$$\text{Dynamic range} = \frac{V_{\max}}{V_{\min}}$$

$$\text{Number of levels} = 2^n - 1$$

$$\text{Dynamic range} = 20\log\frac{V_{\max}}{V_{\max}/2^n - 1} = 20\log(2^n - 1)\ \text{dB}$$

if 2^n is much greater that 1, then

$$\text{Dynamic range} \approx 20\log 2^n = 20n\log 2 \approx 6.02n \text{ dB}$$

Table 10.1 outlines the DR for a given number of bits. Normally the maximum number of bits is less than 20. The voltage ratio of a given number of bits is also given in square brackets [*ratio*]. For example an 8-bit system has a DR of 48.18 dB and the largest voltage amplitude is 256 times the smallest voltage amplitude. A 16-bit system has a DR of 96.33 dB and the largest voltage amplitude is 65 536 times the smallest voltage amplitude.

Table 10.1 Dynamic range of a digital system

Number of bits	*DR (dB)* [*ratio*]	*Number of bits*	*DR (dB)* [*ratio*]
1	6.02 [2]	11	66.23 [2 048]
2	12.04 [4]	12	72.25 [4 096]
3	18.06 [8]	13	78.27 [8 192]
4	24.08 [16]	14	84.29 [16 384]
5	30.10 [32]	15	90.31 [32 768]
6	36.12 [64]	16	96.33 [65 536]
7	42.14 [128]	17	102.35 [131 072]
8	48.16 [256]	18	108.37 [262 144]
9	54.19 [512]	19	114.39 [524 288]
10	60.21 [1 024]	20	120.41 [1 048 576]

10.2.3 Signal-to-noise ratio (SNR)

It can be shown that the SNR for a linearly quantized digital system is (See Appendix 7):

$$\text{SNR} = 1.76 + 6.02n \text{ dB}$$

Table 10.2 outlines the SNR for a given number of bits. Normally the maximum number of bits is less than 20. The voltage ratio of a given number of bits is also given in square brackets [*ratio*]. For example, an 8-bit system has an SNR of 49.92 dB and the largest rms voltage is 313.33 times the smallest rms voltage. A 16-bit system has an SNR of 96.33 dB and the largest rms voltage is 80 167.81 times the smallest rms voltage.

Table 10.2 Signal-to-noise ratio of a digital system

Number of bits	*SNR (dB)* [*ratio*]	*Number of bits*	*SNR (dB)* [*ratio*]
7	43.90 [156.68]	14	86.04 [20 044.72]
8	49.92 [313.33]	15	92.06 [40 086.67]
9	55.94 [626.61]	16	98.08 [80 167.81]
10	61.96 [1 253.14]	17	104.10 [160 324.5]
11	67.98 [2 506.11]	18	110.12 [320 626.9]
12	74.00 [5 011.87]	19	116.14 [641 209.6]
13	80.02 [10 023.05]	20	122.16 [1 282 331]

10.3 Differential encoding

Differential coding is a source-coding method which is used when there is a limited change from one value to the next. It is well suited to video and audio signals, especially audio, where the sampled values can only change within a given range. It is typically used in PCM (pulse code modulation) schemes to encode audio and video signals.

10.3.1 Delta modulation PCM

PCM coverts analogue samples into a digital code. Delta PCM uses a single-bit code to represent an analogue signal. With delta modulation a '1' is transmitted (or stored) if the analogue input is higher than the previous sample or a '0' if it is lower. It must obviously work at a higher rate than the Nyquist frequency, but because it uses only 1 bit, it normally uses a lower output bit rate. Figure 10.1 shows a delta modulation transmitter.

Initially the counter is set to zero. A sample is taken and if it is greater than the analogue value on the DAC output, the counter is incremented by 1, or it is decremented. This continues at a time interval given by the clock. Each time the present sample is greater than the previous sample, a '1' is transmitted; otherwise a '0' is transmitted. Figure 10.2 shows an example signal. The sampling frequency is chosen so that the tracking DAC can follow the input signal. This results in a higher sampling frequency, but because it only transmits one bit at a time, the output bit rate is normally reduced. Figure 10.3 shows that the receiver is almost identical to the transmitter except that it has no comparators.

Two problems with delta modulation are:

- **Slope overload**. This occurs when the signal changes too fast for the modulator to keep up; see Figure 10.4. It is possible to overcome this problem by increasing the clock frequency or increasing the step size.
- **Granular noise**. This occurs when the signal changes slowly in amplitude, as illustrated in Figure 10.5. The reconstructed signal contains a noise which is not present at the input. Granular noise is equivalent to quantization noise in a PCM system. It can be reduced by decreasing the step size, though there is a compromise between smaller step size and slope overload.

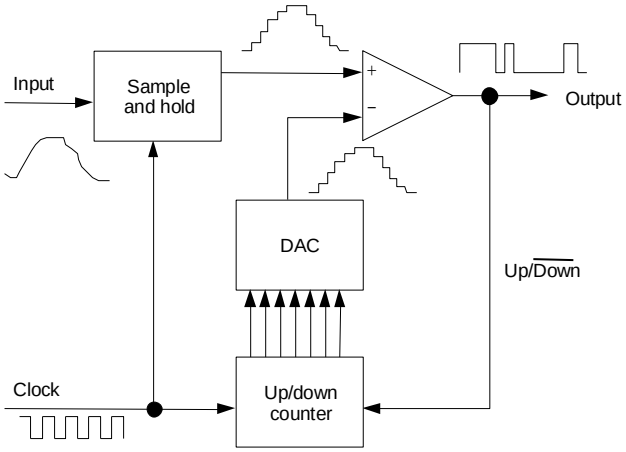
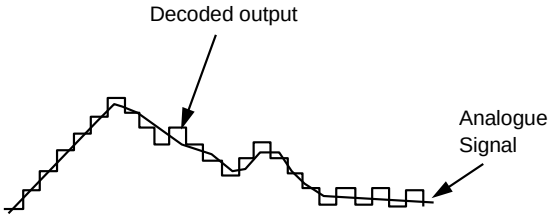


Figure 10.1 Delta modulation



Code: 11111110001000110000010101

Figure 10.2 Delta modulator signal

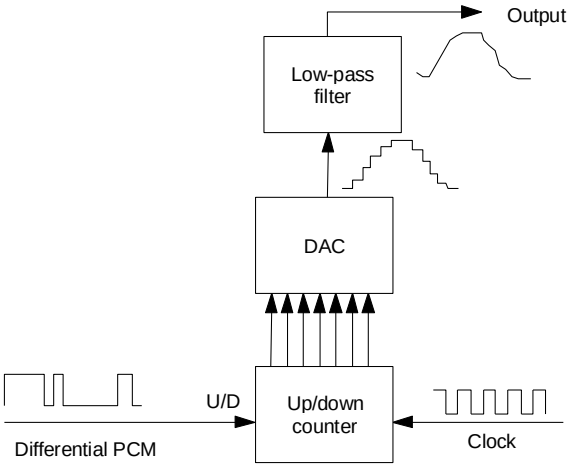
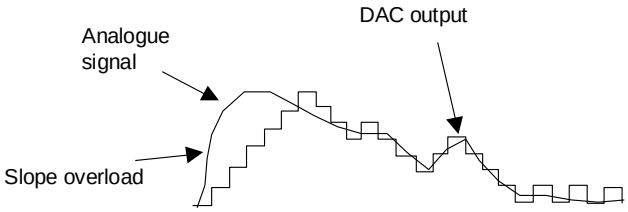


Figure 10.3 Delta modulator receiver



PCM 11111110001000110000010101

Figure 10.4 Slope overload

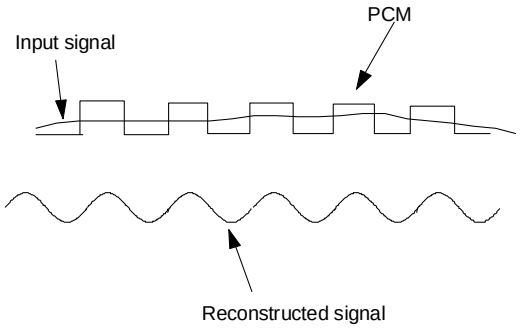


Figure 10.5 Granular noise

10.3.2 Adaptive delta modulation PCM

Unfortunately, delta modulation cannot react to very rapidly changing signals and will thus take a relatively long time to catch them up (known as slope overload). It also suffers when the signal does not change much as this ends up in a square wave signal (known as granular noise). One method of reducing granular noise and slope overload is to use adaptive delta PCM. With this method the step size is varied by the slope of the input signal. The larger the slope, the larger the step size; see Figure 10.6. The algorithms usually depend on the system and the characteristics of the signal. A typical algorithm is to start with a small step and increase it by a multiple until the required level is reached. The number of slopes will depend on the number of coded bits, such as 4 step sizes for 2 bits, 8 for 3 bits, and so on.

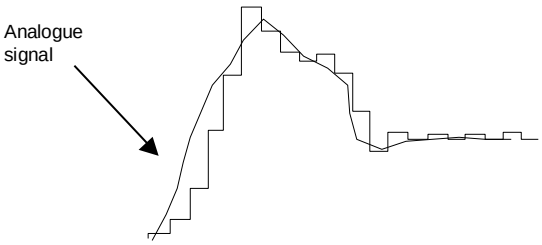


Figure 10.6 Variation of step size

10.3.3 Differential PCM (DPCM)

Speech signals tend not to change much between two samples. Thus similar codes are sent, which leads to a degree of redundancy. For example, in a certain sample it is likely the signal will only change within a range of voltages, as illustrated in Figure 10.7.

DPCM reduces the redundancy by transmitting the difference in the amplitude of two consecutive samples. Since the range of sample differences is typically less than the range of individual samples, fewer bits are required for DPCM than for conventional PCM.

Figure 10.8 shows a simplified transmitter and receiver. The input signal is filtered to half the sampling rate. This filter signal is then compared with the previous DPCM signal. The difference between them is then coded with the ADC.

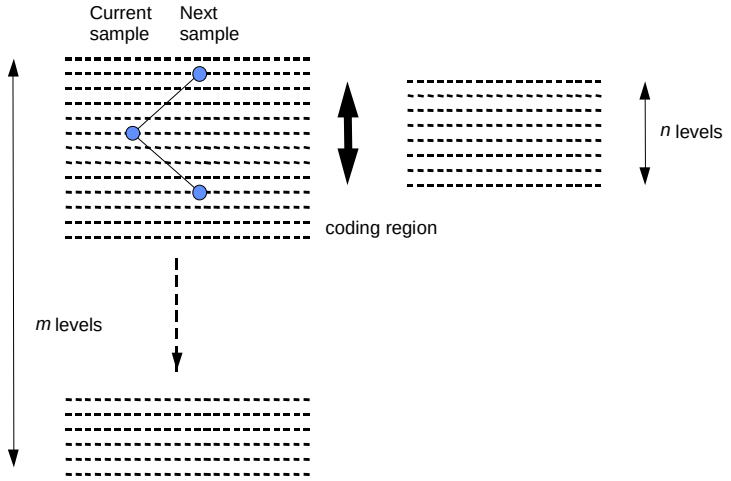


Figure 10.7 Normal and differential quantization

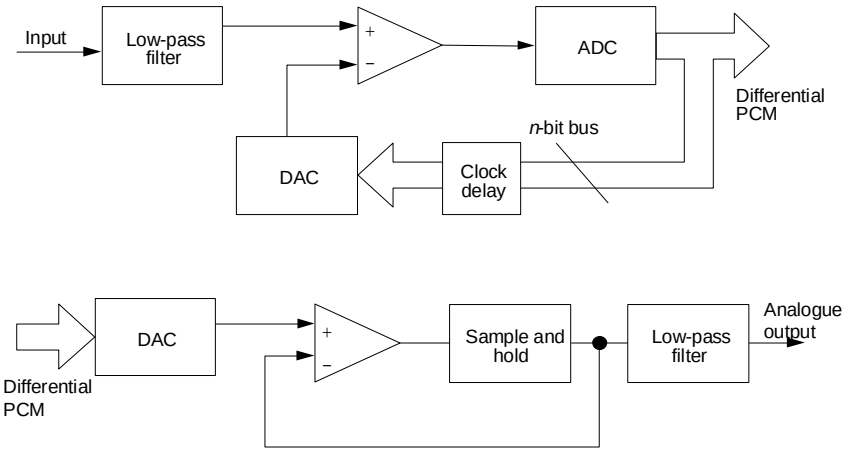


Figure 10.8 DPCM transmitter/receiver

10.3.4 Adaptive differential PCM (ADPCM)

ADPCM allows speech to be transmitted at 32 kbps with little noticeable loss of quality. As with differential PCM the quantizer operates on the difference between the current and previous samples. The adaptive quantizer uses a uniform quantization step *M*, but when the signal moves towards the limits of the quantization range, the step size *M* is increased. If it is around the center of the ranges, the step size is decreased. Within any other regions the step size hardly changes. Figure 10.9 illustrates this operation with a signal quantized to 16 levels. This results in 4-bit code.

The change of the quantization step is done by multiplying the quantization level, M, by a number slightly greater, or less, than 1 depending on the previously quantized level.

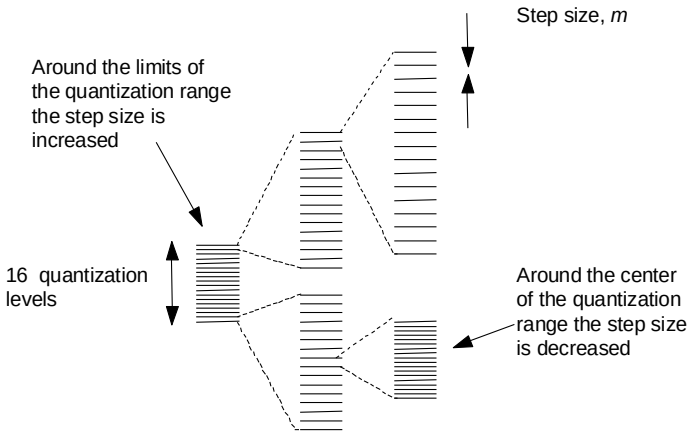


Figure 10.9 ADPCM quantization

10.4 Speech compression

Subjective and system tests have found that 12-bit coding is required to code speech signals, which gives 4096 quantization levels. If linear quantization is applied then the quantization step is the same for quiet levels as for loud levels. Any quantization noise in the signal will be more noticeable at quiet levels than at loud levels. When the signal is loud, the signal itself swamps the quantization noise, as illustrated in Figure 10.10. Thus, an improved coding mechanism is to use small quantization steps at low input levels and a higher one at high levels. This is achieved using non-linear compression.

The two most popular types of compression are A-Law (in European systems) and μ-Law (in the USA). These laws are similar and compress the 12-bit quantized speech code into an 8-bit compressed code. An example compression curve is shown in Figure 10.11. As an approximation, the two laws are split into 16 line segments. Starting from the origin and moving outwards, left and right, each segment has half the slope of the previous.

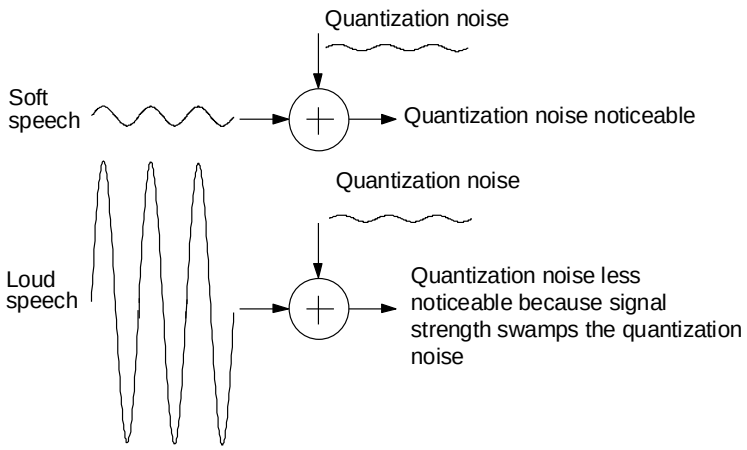


Figure 10.10 Quantization noise is more noticeable with low signal levels

Using an 8-bit compressed code at a sample rate of 8000 samples per second gives a bit rate of 64 kbps. ISDN uses this bit rate to transmit digitized speech. Figure 10.12 shows a basic transmission system.

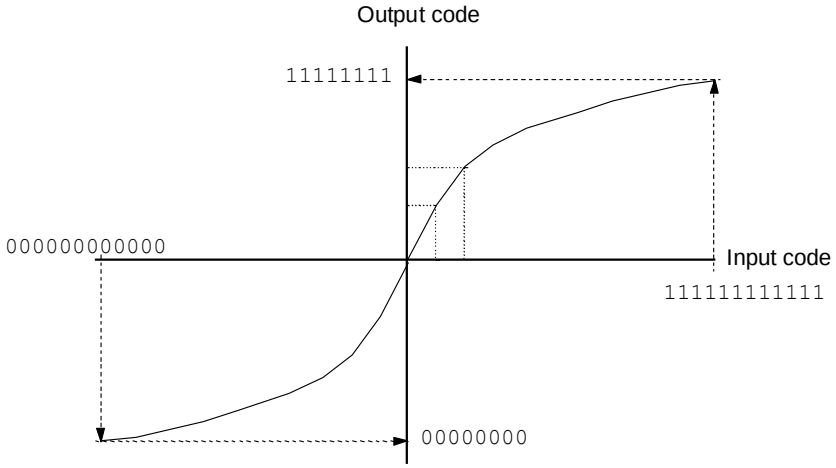


Figure 10.11 12-bit to 8-bit non-linear compression

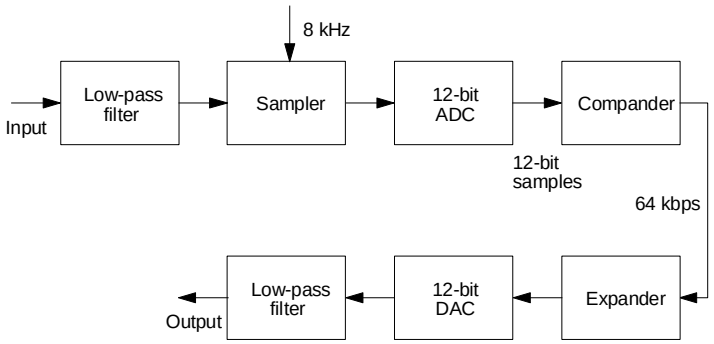


Figure 10.12 Typical PCM speech system

10.5 A-Law and μ-Law companding

The companding and expansion encoding is normally implemented using either μ-Law or A-Law. A-Law is used in Europe and in many other countries, whereas μ-Law is used in North America and Japan. Both were defined by the CCITT in the G.711 recommendation and both use non-uniform quantization step sizes which increase logarithmically with signal level. μ-Law uses the compression characteristic of:

$$y = \frac{\log(1+\mu x)}{\log(1+x)} \quad \text{for } x \geq 0$$

where y is the output magnitude

x is the input magnitude
μ is a positive factor which is chosen for the required compression characteristics

Figure 10.13 shows an example of μ-Law using $\mu = 1$, $\mu = 50$ and $\mu = 255$. Using $\mu = 0$ gives uniform conversion (linear quantization). Normally speech systems use $\mu = 255$ as this characteristic is well matched to human hearing. An 8-bit implementation can achieve a small SNR a and dynamic range equivalent to that of a 12-bit uniform system.

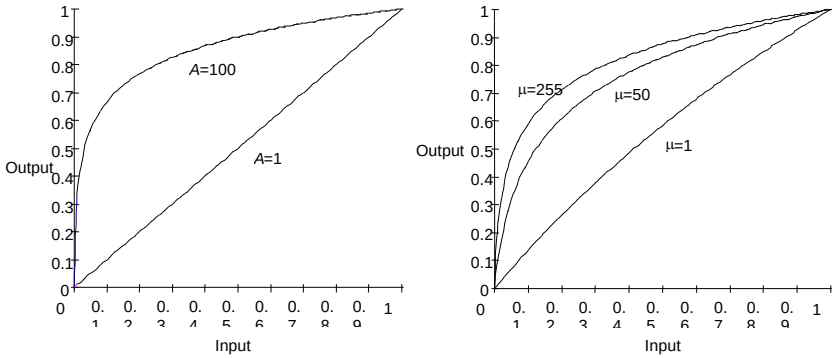


Figure 10.13 A-Law and μ-Law characteristics

The A-law also uses quantization characteristics that vary logarithmically. Figure 10.13 shows an example of A-Law using $A = 1$ and $A = 100$. Most A-Law speech systems use $A = 87.56$. The compression characteristic is:

$$y = \begin{cases} \dfrac{Ax}{1+\log A} & \text{for } 0 \le |x| \le \dfrac{1}{A} \\ \dfrac{1+\log(Ax)}{1+\log A} & \text{for } \dfrac{1}{A} \le |x| \le 1 \end{cases}$$

where A is a positive integer.

Figure 10.14 shows two input waveforms, 1 V peak to peak and 0.1 V peak to peak. It can be seen that the companding processes amplifies the lower amplitudes more than the large amplitudes. This causes low-amplitude speech signals to be boosted compared with loud speech. Also notice that the waveform has been distorted because the low amplitudes are amplified more than the large amplitudes.

10.5.1 Digitally linearizable log-companding

The mathematical formulas for A-Law and μ-Law are normally approximated to a series of linear segments. This permits more precise control of the quantization characteristics. The chosen approximation used is to make the step sizes in consecutive segments change by a factor of 2. Figure 10.15 shows the characteristic of the piecewise linear conversion. It can be seen that the slope of each segment is twice the slope of the previous segment (although in A-Law 98.56, segment 0 and segment 1 have the same slope). Each segment has 16 quanti-

zation levels and there are 16 segments (8 for positive inputs and 8 for negative inputs). Thus, 1 bit identifies the sign bit, 3 bits identify the segment (in the positive or negative part) and 4 bits identifies the quantization level. The 8-bit companded values thus take the form:

```
SLLLQQQQ
```

where `S` is the sign bit, `LLL` is the segment number and `QQQQ` is the quantization level within the segment.

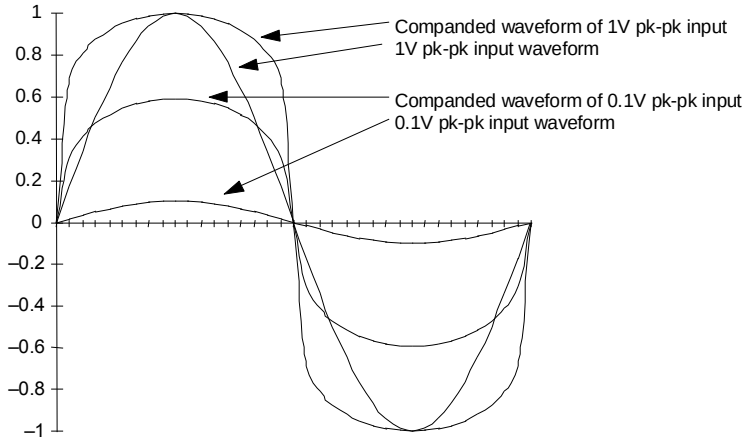


Figure 10.14 Effects of waveforms with μ-255 encoding.

Table 10.3 shows the conversion for A-Law 87.56. For example, if the input value is between 16 and 17, the companded value will be 001 0000. If this value is positive then the most significant bit will be a 1, thus the companded value will be 1001 0000.

Table 10.3 shows that the step sizes for the first two segments are the same (unity step size). Table 10.4 shows the μ-Law encoding table.

Consider A-Law with the input range between +5 V and –5 V. An input voltage of +1 V will correspond to the input level of:

$$\text{Input} = \frac{1}{5} \times 2048 = 409.6$$

Referring to Table 10.3, this is within the segment from 256 to 512. The code will thus be S101XXXX. The level within the segment will be:

$$\text{Level} = \frac{409.6 - 256}{16} = 9.6$$

which corresponds to quantization level 9. Thus, the companded value is:

```
01011001
```

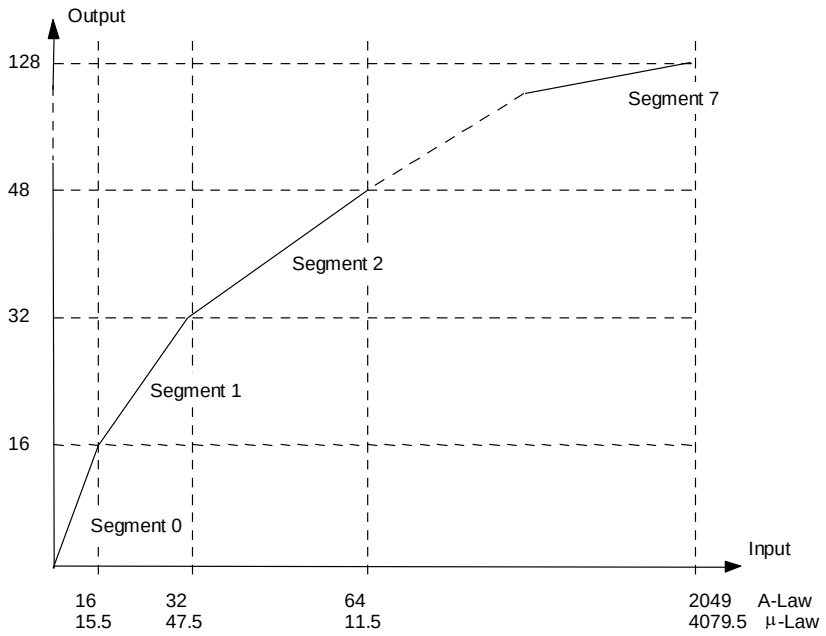


Figure 10.15 Piecewise linear compression for A-Law and μ-Law

Table 10.3 A-Law 87.56 encoding/decoding

Input	*Companded*	*Decoder level*	*Decoded level number*	*Step size*
0–1	000 0000	0	0.5	1
…	…	…	…	
15–16	000 1111	15	15.5	
16–17	001 0000	16	16.5	1
…	…	…	…	
31–32	001 1111	31	31.5	
32–34	010 0000	32	33	2
…	…	…	…	
62–64	010 1111	47	63	
64–68	011 0000	48	66	4
…	…	…	…	
124–128	011 1111	63	126	
128–136	100 0000	64	132	8
…	…	…	…	
248–256	100 1111	79	252	
256–272	101 0000	80	264	16
…	…	…	…	
496–512	101 1111	95	504	
512–544	110 0000	96	528	32
…	…	…	…	
992–1024	110 1111	111	1008	
1024–1088	111 0000	112	1056	64
…	…	…	…	
1984–2048	111 1111	127	2016	

Table 10.4 μ-255 encoding/decoding

Input	*Companded*	*Decoder level*	*Decoded level number*	*Step size*
0–0.5	000 0000	0	0	1
…	…	…	…	
14.5–15.5	000 1111	15	15	
15.5–17.5	001 0000	16	16.5	2
…	…	…	…	
45.5–47.5	001 1111	31	46.5	
47.5–51.5	010 0000	32	49.5	4
…	…	…	…	
107.5–111.5	010 1111	47	109.5	
111.5–119.5	011 0000	48	115.5	8
…	…	…	…	
231.5–239.5	011 1111	63	235.5	
239.5–255.5	100 0000	64	247.5	16
…	…	…	…	
479.5–495.5	100 1111	79	487.5	
497.5–527.5	101 0000	80	511.5	32
…	…	…	…	
975.5–1007.5	101 1111	95	991.5	
1007.5–1071.5	110 0000	96	1039.5	64
…	…	…	…	
1967.5–2031.5	110 1111	111	1999.5	
2031.5–2159.5	111 0000	112	2095.5	128
…	…	…	…	
3951.5–4079.5	111 1111	127	4015.5	

10.6 Speech sampling

With telephone-quality speech the signal bandwidth is normally limited to 4 kHz, thus it is sampled at 8 kHz. If each sample is coded with 8 bits then the basic bit rate will be:

Digitized speech signal rate = 8×8 kbps = 64 kbps

Table 10.5 outlines the main compression techniques for speech. The G.722 standard allows the best-quality signal. The maximum speech frequency is 7 kHz rather than 4 kHz in normal coding systems; this is equivalent of 14 coding bits. The G.728 allows extremely low bit rates (16 kbps).

10.7 PCM-TDM systems

Multiple channels of speech can be sent over a single line using time division multiplexing (TDM). In the UK a 30-channel PCM system is used, whereas the USA uses 24.

Table 10.5 Speech compression standards

ITU standard	*Technology*	*Bit rate*	*Description*
G.711	PCM	64 kbps	Standard PCM
G.721	ADPCM	32 kbps	Adaptive delta PCM where each value is coded with 4 bits
G.722	SB-ADPCM	48, 56 and 64 kbps	Subband ADPCM allows for higher-quality audio signals with a sampling rate of 16 kHz
G.728	LD-CELP	16 kbps	Low-delay code excited linear prediction for low bit rates

With a PCM-TDM system, several voice band channels are sampled, converted to PCM codes, these are then time division multiplexed onto a single transmission media.

Each sampled channel is given a time slot and all the time slots are built up into a frame. The complete frame usually has extra data added to it such as synchronization data, and so on. Speech channels have a maximum frequency content of 4 kHz and are sampled at 8 kHz. This gives a sample time of 125 μs. In the UK, a frame is built up with 32 time slots from TS0 to TS31. TS0 and TS16 provide extra frame and synchronization data. Each of the time slots has 8 bits, therefore the overall bit rate is:

Bits per time slot = 8
Number of time slots = 32
Time for frame = 125 μs

$$\text{Bit rate} = \frac{\text{No of bits}}{\text{Time}} = \frac{32 \times 8}{125 \times 10^{-6}} = 2048 \text{ kbps}$$

In the USA and Japan this bit rate is 1.544 Mbps. These bit rates are known as the primary rate multipliers. Further interleaving of several primary rate multipliers increases the rate to 6.312, 44.736 and 139.264 Mbps (for the USA) and 8.448, 34.368 and 139.264 Mbps (for the UK).

The UK multiframe format is given in Figure 10.16. In the UK format the multiframe has 16 frames. Each frame time slot 0 is used for synchronization and time slot 16 is used for signaling information. This information is sub-multiplexed over the 16 frames. During frame 0 a multiframe-alignment signal is transmitted in TS16 to identify the start of the multiframe structure. In the following frames, the eight binary digits available are shared by channels 1–15 and 16–30 for signaling purposes. TS16 is used as follows:

Frame 0 `0000XXXX`
Frames 1–15 `1234 5678`

where `1234` are the four signaling bits for channels 1, 2, 3, …, 15 in consecutive frames, and `5678` are the four signaling bits for channels 16, 17, 18, … 31 in consecutive frames.

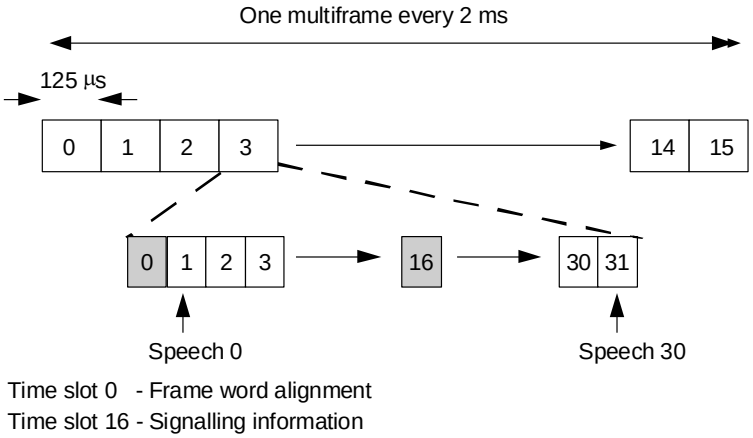


Figure 10.16 PCM-TDM multiframe format with 30 speech channels

Thus in the first frame the `0000XXXX` code word is sent, in the next frame the first channel and the 16th channel appear in TS16, the next will contain the second and the 17th, and so on. Typical 4-bit signal information is:

`1111` – circuit idle/busy
`1101` – disconnection

TS0 contains a frame-alignment signal which enables the receiver to synchronize with the transmitter. The frame-alignment signal (`X0011011`) is transmitted in alternative frames. In the intermediate frames a signal known as a not-word is transmitted (`X10XXXXX`). The second binary digit is the complement of the corresponding binary digit in the frame-alignment signal. This reduces the possibility of demultiplexed misalignment to imitative frame-alignment signals.

Alternative frames:

TS0: `X0011011`
TS0: `X10XXXXX`

where `X` stands for don't care conditions.