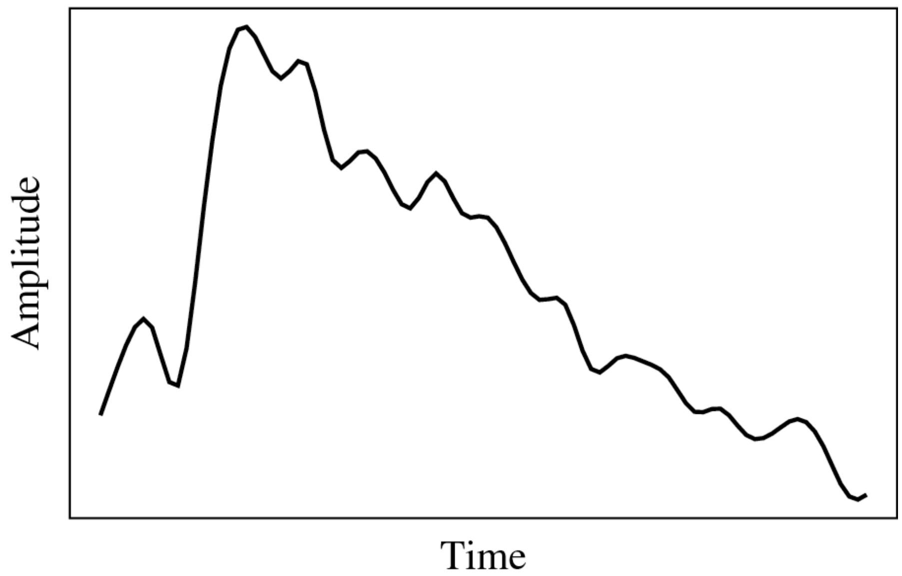
What is Sound?

- Sound is a wave phenomenon like light, but is macroscopic and involves molecules of air being compressed and expanded under the action of some physical device
  - Since sound is a pressure wave, it takes on continuous values, as opposed to digitized ones



Digitization

- **Digitization** means conversion to a stream of numbers, and preferably these numbers should be integers for efficiency
  - Sampling means measuring the quantity we are interested in, usually at evenly-spaced intervals
  - Measurements at evenly spaced time intervals is called sampling. The rate at which it is performed is called the *sampling frequency*. For audio, typical sampling rates are from 8 kHz (8,000 samples per second) to 48 kHz. This range is determined by the Nyquist theorem
  - Sampling in the amplitude or voltage dimension is called **quantization**

Quality	Sample Rate (Khz)	Bits per Sample	Mono / Stereo	Data Rate (uncompressed) (kB/sec)	Frequency Band (KHz)
Telephone	8	8	Mono	8	0.200-3.4
AM Radio	11.025	8	Mono	11.0	0.1-5.5
FM Radio	22.05	16	Stereo	88.2	0.02-11
CD	44.1	16	Stereo	176.4	0.005-20
DAT	48	16	Stereo	192.0	0.005-20
DVD Audio	192 (max)	24(max)	6 channels	1,200 (max)	0-96 (max)

Nyquist theorem

- The Nyquist theorem states how frequently we must sample in time to be able to recover the original sound. For correct sampling we must use a sampling rate equal to at least twice the maximum frequency content in the signal. This rate is called the **Nyquist rate.**
- **Nyquist Theorem**: If a signal is **band-limited**, i.e., there is a lower limit f_1 and an upper limit f_2 of frequency components in the signal, then the sampling rate should be at least $2(f_2 - f_1)$

Signal to Noise Ratio (SNR)

- The ratio of the power of the correct signal and the noise is called the *signal to noise ratio* (**SNR**)
  - a measure of the quality of the signal.

- The SNR is usually measured in decibels (**dB**), where 1 dB is a tenth of a **bel**. The SNR value, in units of dB, is defined in terms of base-10 logarithms of squared voltages, as follows:

$$SNR = 10\log_{10}\frac{V_{signal}^2}{V_{noise}^2} = 20\log_{10}\frac{V_{signal}}{V_{noise}}$$

Common sounds

Threshold of hearing	0
Rustle of leaves	10
Very quiet room	20
Average room	40
Conversation	60
Busy street	70
Loud radio	80
Train through station	90
Riveter	100
Threshold of discomfort	120
Threshold of pain	140
Damage to ear drum	160

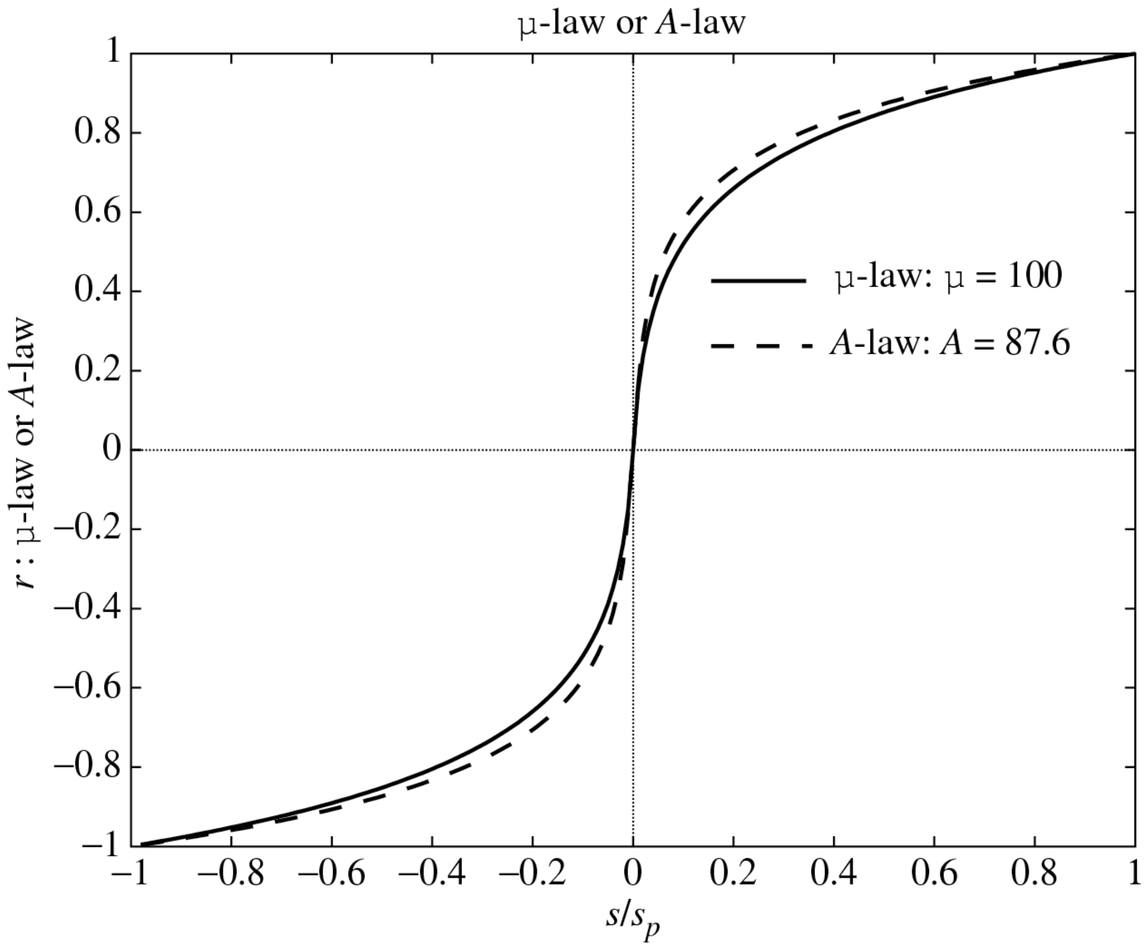
Signal to Quantization Noise Ratio (SQNR)

- If voltages are actually in 0 to 1 but we have only 8 bits in which to store values, then effectively we force all continuous values of voltage into only 256 different values. This introduces a roundoff error. It is not really "noise". Nevertheless it is called **quantization noise** (or quantization error)
- **Linear and Non-linear Quantization**
  - **Linear format**: samples are typically stored as uniformly quantized values
  - **Non-uniform quantization**: set up more finely-spaced levels where humans hear with the most acuity
    - Weber's Law stated formally says that equally perceived differences have values proportional to absolute levels:
      - ΔResponse $\propto$ ΔStimulus/Stimulus

Nonlinear quantization

- Nonlinear quantization works by first transforming an analog signal from the raw s space into the theoretical r space, and then uniformly quantizing the resulting values. Such a law for audio is called μ-**law** encoding, (or **u-law**). A very similar rule, called A-**law**, is used in telephony in Europe

- The μ-law in audio is used to develop a nonuniform quantization rule for sound: uniform quantization of r gives finer resolution in s at the quiet end



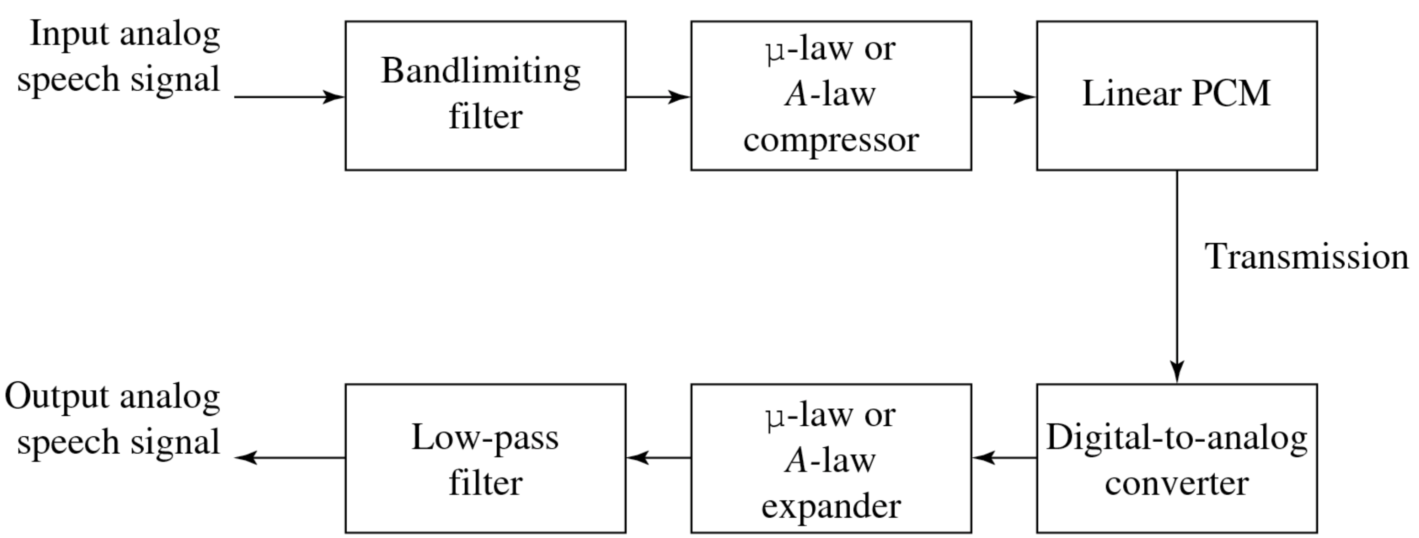
Synthetic sounds

- Frequency modulation (with a magnitude envelope)
- Wav table: the actual digital samples of sounds from real instruments are stored. Since wave tables are stored in memory on the sound card, they can be manipulated by software so that sounds can be combined, edited, and enhanced

- MIDI is a scripting language — it codes “events” that stand for the production of sounds. E.g., a MIDI event might include values for the pitch of a single note, its duration, and its volume.

6.3 Quantization and Transmission of Audio

- producing quantized sampled output for audio is called **PCM** (Pulse Code Modulation). The differences version is called **DPCM** (and a crude but efficient variant is called **DM**). The adaptive version is called **ADPCM**



Differential coding

- If a time-dependent signal has some consistency over time ("temporal redundancy"), the difference signal, subtracting the current sample from the previous one, will have a more peaked histogram, with a maximum around zero

ADPCM

- **ADPCM** (Adaptive DPCM) takes the idea of adapting the coder to suit the input much farther. The two pieces that make up a DPCM coder: the quantizer and the predictor.
  - In Adaptive DM, adapt the quantizer step size to suit the input. In DPCM, we can change the step size as well as decision boundaries, using a non-uniform quantizer.
  - We can carry this out in two ways:

    (a) **Forward adaptive quantization**: use the properties of the input signal.

    (b) **Backward adaptive quantization**: use the properties of the quantized output. If quantized errors become too large, we should change the non-uniform quantizer.

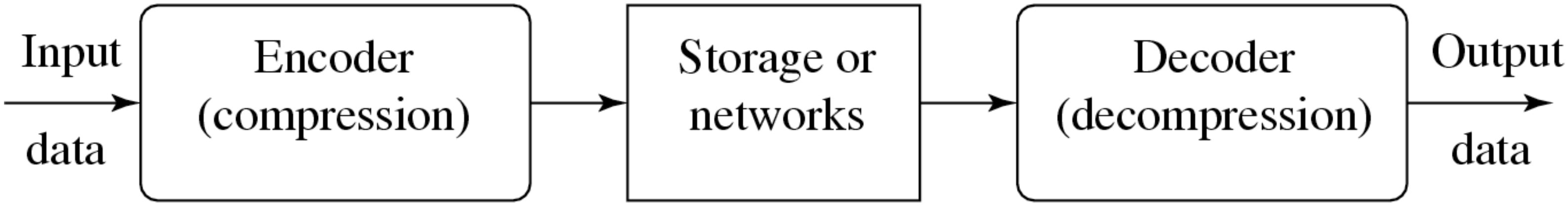
    We can also **adapt the predictor**, again using forward or backward adaptation. Making the predictor coefficients adaptive is called *Adaptive Predictive Coding* (APC)

Chapter 7: Lossless compression

- **Compression**: the process of coding that will effectively reduce the total number of bits needed to represent certain information.

If the compression and decompression processes induce no information loss, then the compression scheme is **lossless**; otherwise, it is **lossy**.

Compression ratio:



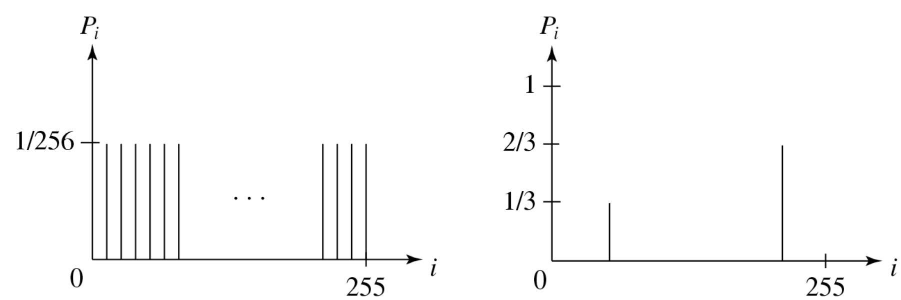
Shannon’s theory

- The *entropy* η of an information *source* with alphabet $S = \{s_1, s_2, \ldots, s_n\}$ is:

$$= -\sum_{i=1}^{n} p_i \log_2 p_i \qquad (7.3)$$

p_i – probability that symbol s_i will occur in S.

Compression is not possible for a) because entropy is 8 (need 8 bits per value)



(a) (b)

Run length coding

- • **Memoryless Source**: an information source that is independently distributed. Namely, the value of the current symbol does not depend on the values of the previously appeared symbols.
- • Instead of assuming memoryless source, *Run-Length Coding* (*RLC*) exploits memory present in the information source.
- • **Rationale for RLC**: if the information source has the property that symbols tend to form continuous groups, then such symbol and the length of the group can be coded.