Digital Signal Sampling and Quantization

Prof. Dr. Renato Pajarola
Visualization and MultiMedia Lab
Department of Informatics
University of Zürich

VISUALIZATIONAND
MULTIMEDIALAB

Copyrights

- Most figures of these slides are copyright protected and come from the book ***Fundamentals of Multimedia*** or other indicated sources
- You understand that the slides contain copyright protected material and therefore the following conditions of use apply:
  - The slides may be used for personal teaching purposes only
  - Publishing the slides to any public web site is not allowed
  - Sharing the slides with other persons or institutions is prohibited

Overview

1. Sampling theorem
2. Signal to noise ratio
3. Linear quantization
4. Non-linear quantization

Digitization

- **Digitization** means conversion to a stream of numbers, and preferably these numbers should be integers for efficiency.
- Fig. 6.1 shows the 1-dimensional nature of sound: **amplitude** values depend on a 1D variable, time. (And note that images depend instead on a 2D set of variables, x and y).

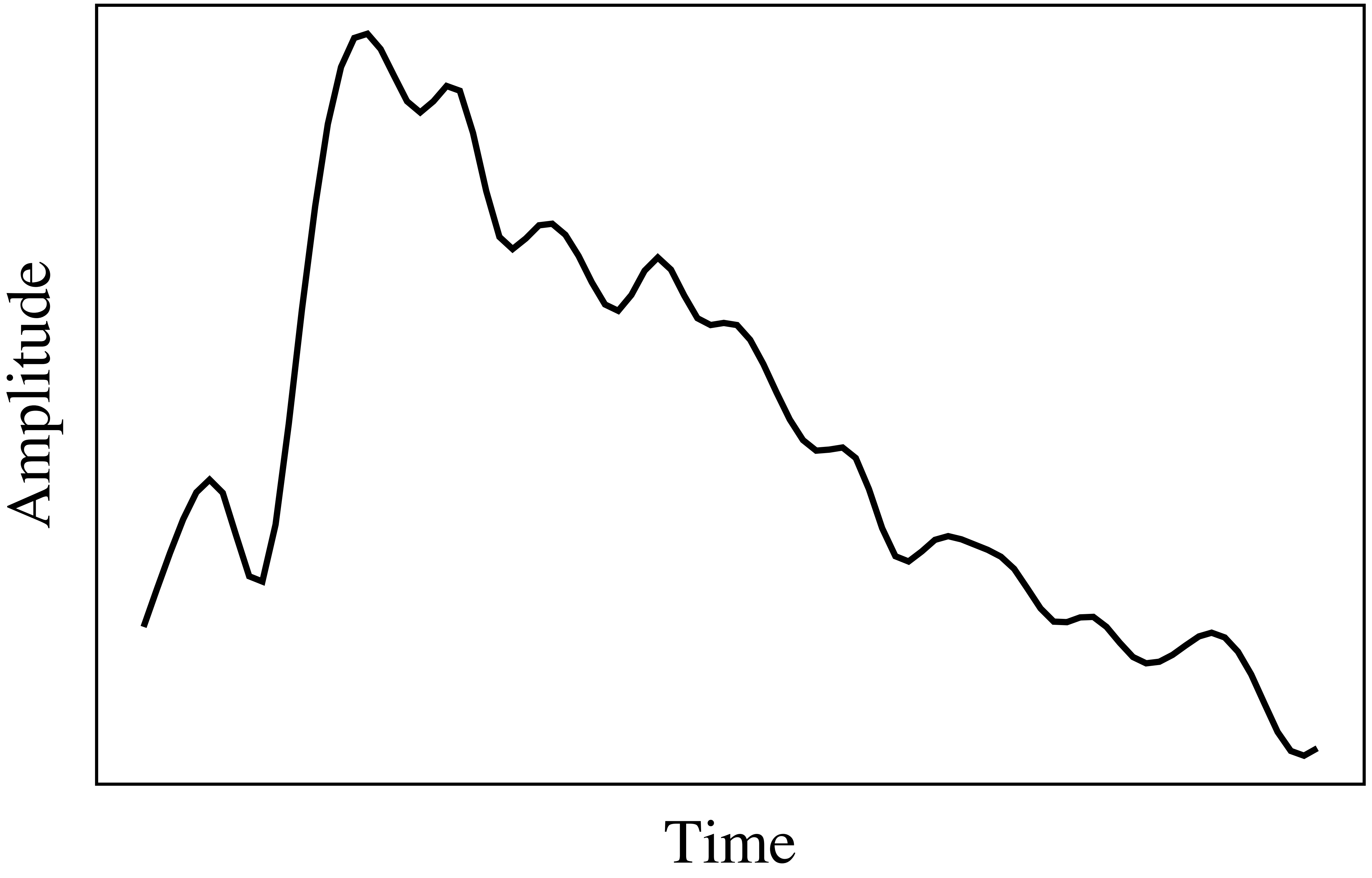


Fig. 6.1: An analog signal: continuous measurement of pressure wave.

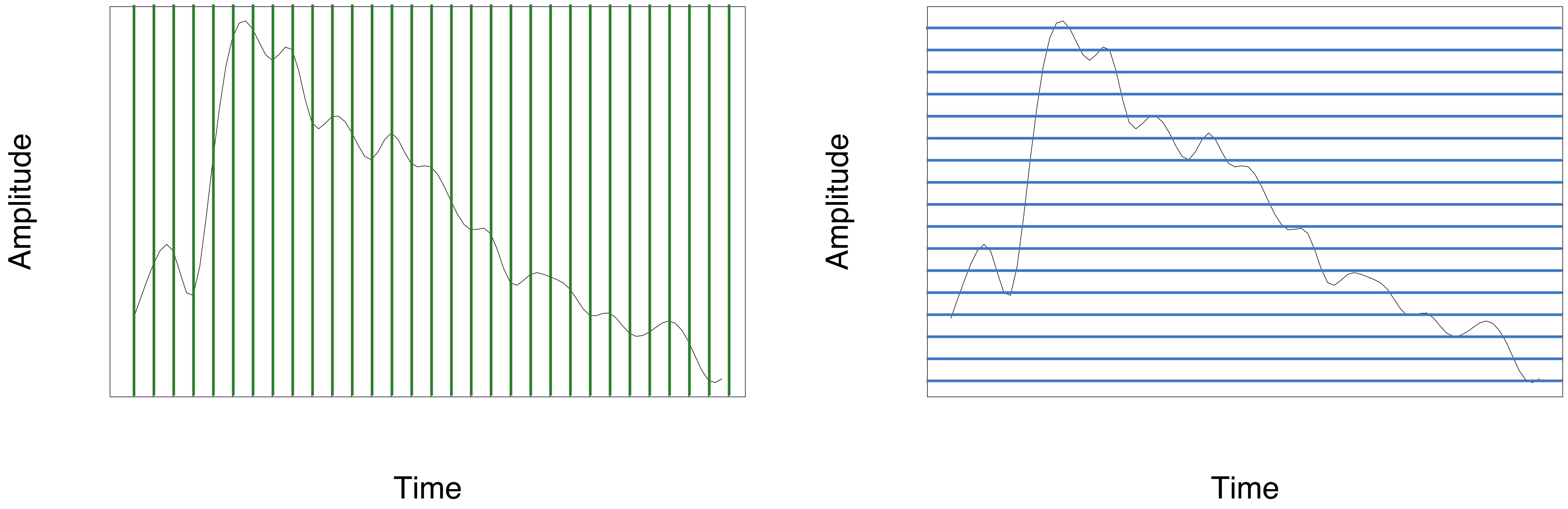
- The graph in Fig. 6.1 has to be made digital in both time and amplitude. To digitize, the signal must be **sampled** in each dimension: in time, and in amplitude.

  (a) Sampling means measuring the quantity we are interested in, usually at evenly-spaced intervals.

  (b) The first kind of sampling, using measurements only at evenly spaced time intervals, is simply called, *sampling*. The rate at which it is performed is called the *sampling frequency* (see Fig. 6.2(a)).

  (c) For audio, typical sampling rates are from 8 kHz (8,000 samples per second) to 48 kHz. This range is determined by Nyquist theorem discussed later.

  (d) Sampling in the amplitude or voltage dimension is called **quantization**. Fig. 6.2(b) shows this kind of sampling.



(a) (b)

Fig. 6.2: Sampling and Quantization. (a): Sampling the analog signal in the time dimension. (b): Quantization is sampling the analog signal in the amplitude dimension.

- Thus to decide how to digitize audio data we need to answer the following questions:

  1. What is the sampling rate?
  2. How finely is the data to be quantized, and is quantization uniform?
  3. How is audio data formatted? (file format)

Nyquist Theorem

- Signals can be decomposed into a sum of sinusoids. Fig. 6.3 shows how weighted sinusoids can build up quite a complex signal.

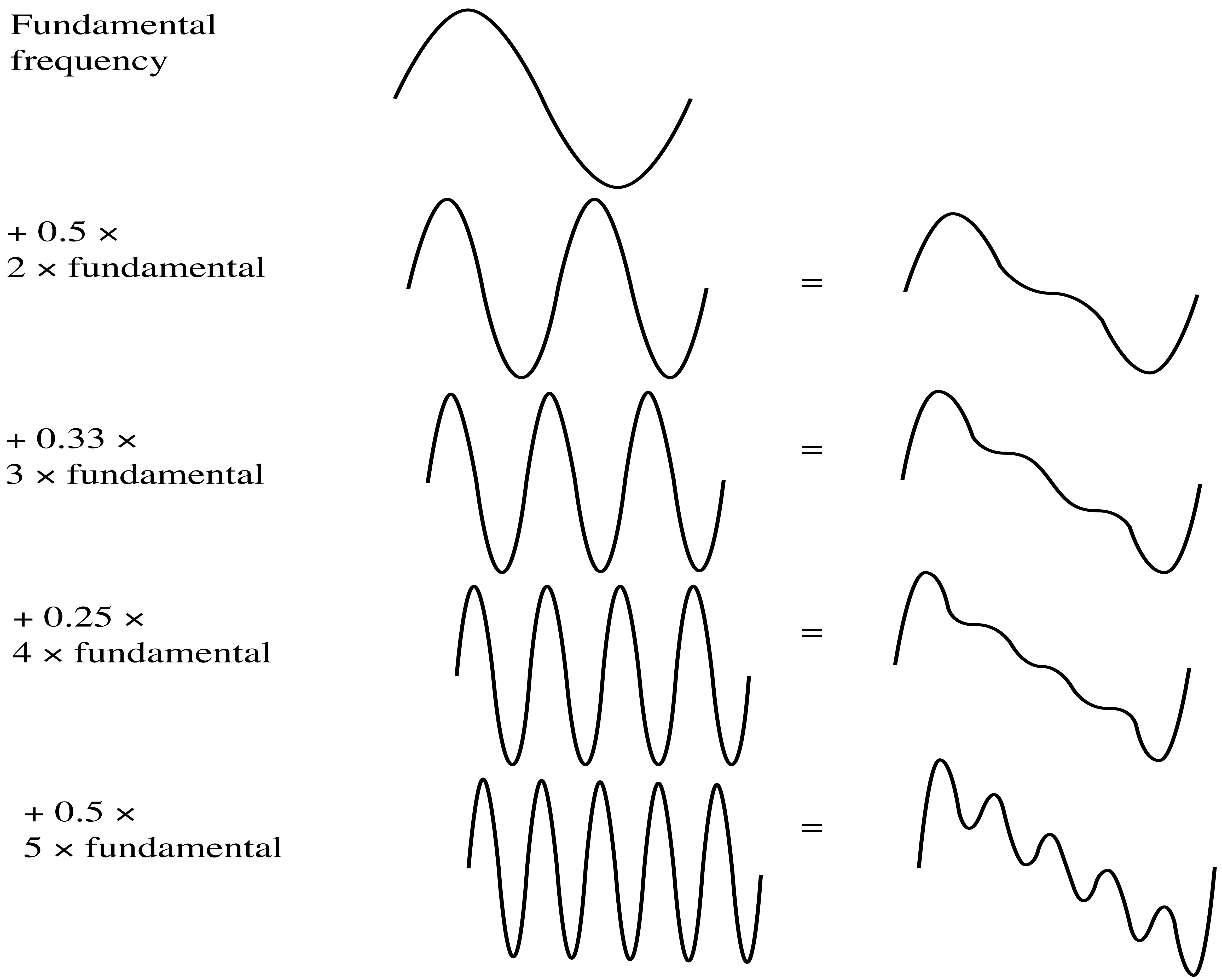


Fig. 6.3: Building up a complex signal by superposing sinusoids

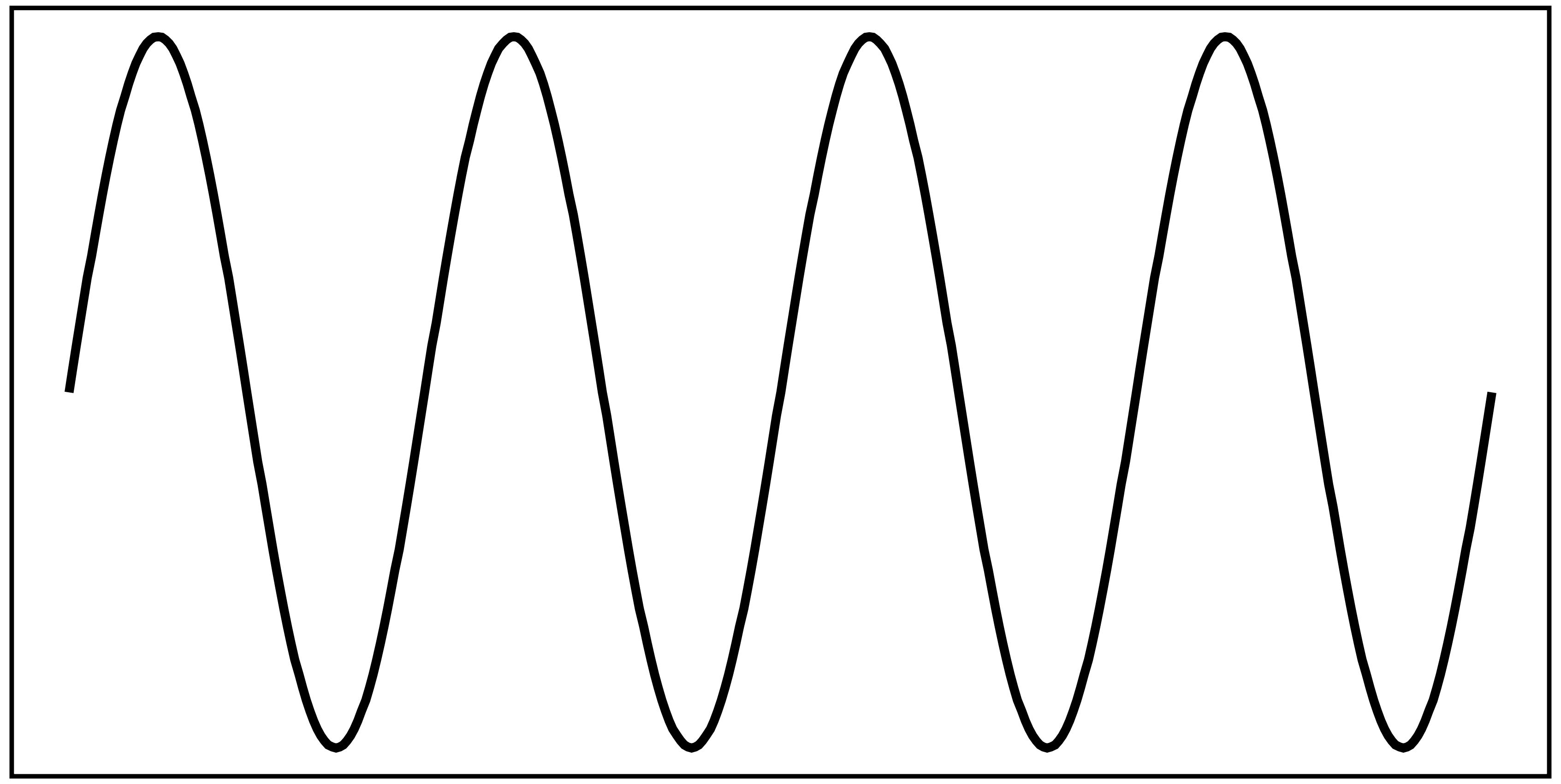
- The Nyquist theorem states how frequently we must sample in time to be able to recover the original sound.

  (a) Fig. 6.4(a) shows a single sinusoid: it is a single, pure, frequency (only electronic instruments can create such sounds).

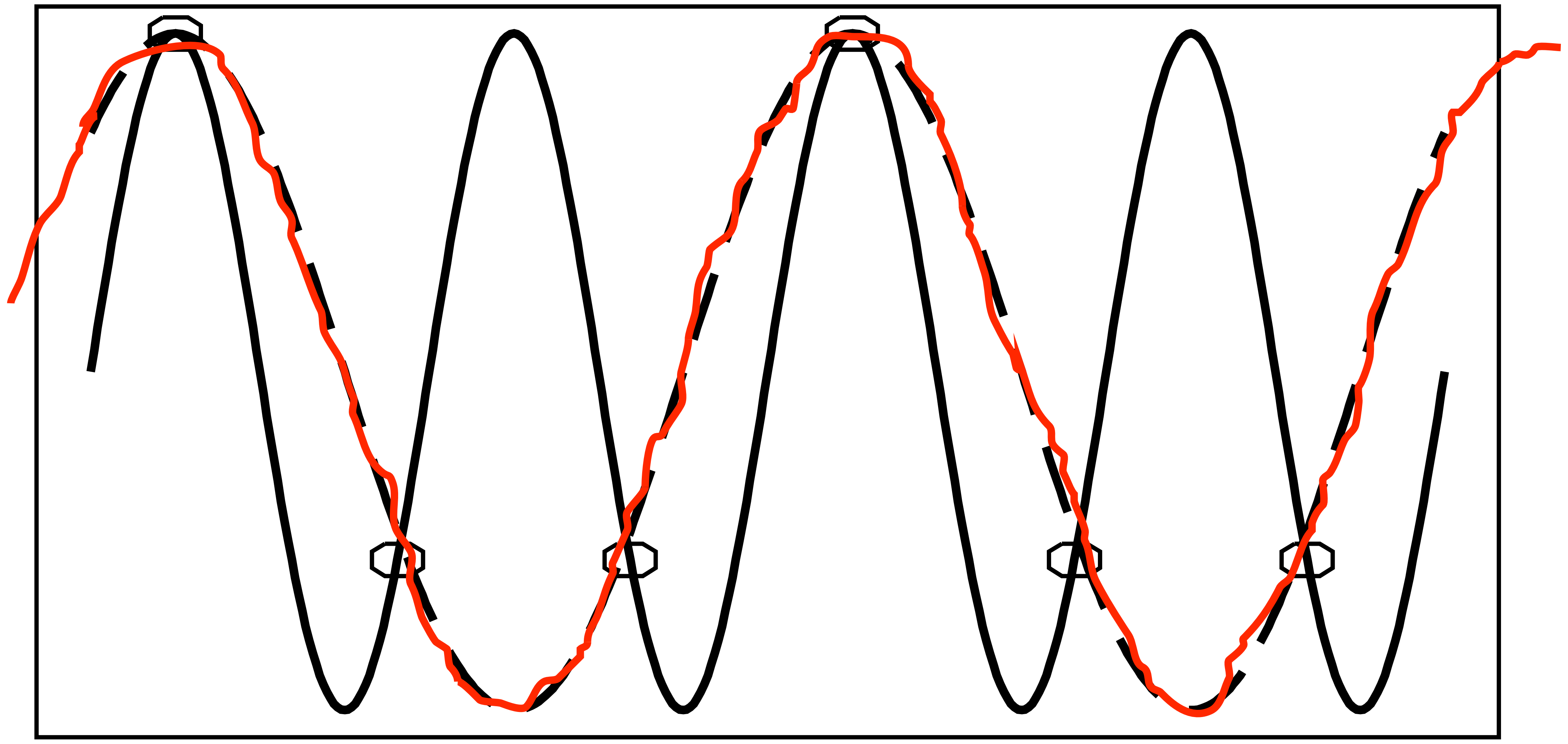
  (b) If sampling rate just equals the actual frequency, Fig. 6.4(b) shows that a false signal is detected: it is simply a constant, with zero frequency.

  (c) Now if sample at 1.5 times the actual frequency, Fig. 6.4(c) shows that we obtain an incorrect (**alias**) frequency that is lower than the correct one — it is half the correct one (the wavelength, from peak to peak, is double that of the actual signal).

  (d) Thus 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**.



(a)



(c)

Fig. 6.4: Aliasing. (a): A single frequency. (b): Sampling at exactly the frequency produces a constant. (c): Sampling at 1.5 times per cycle produces an *alias* perceived frequency.

Nyquist–Shannon Sampling Theorem

- For correct sampling of a signal with maximal frequency $f_{\max}$, we must use a sampling rate $f_{\text{sampling}} \geq 2f_{\max}$ equal to at least twice the maximum frequency content in the signal
  - f_{sampling} Nyquist rate
  - $f_{\max}$: Bandwidth

- **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 $2(f_2 - f_1)$.

- **Nyquist frequency**: half of the Nyquist rate.
  - Since it would be impossible to recover frequencies higher than Nyquist frequency in any event, most systems have an **antialiasing filter** that restricts the frequency content in the input to the sampler to a range at or below Nyquist frequency.

- The relationship among the Sampling Frequency, True Frequency, and the Alias Frequency is as follows:

$$f_{alias} = f_{sampling} - f_{true}, \quad \text{for} \quad f_{true} < f_{sampling} < 2 \times f_{true} \qquad (6.1)$$

- In general, the apparent frequency of a sinusoid is the lowest frequency of a sinusoid that has exactly the same samples as the input sinusoid. Fig. 6.5 shows the relationship of the apparent frequency to the input frequency.

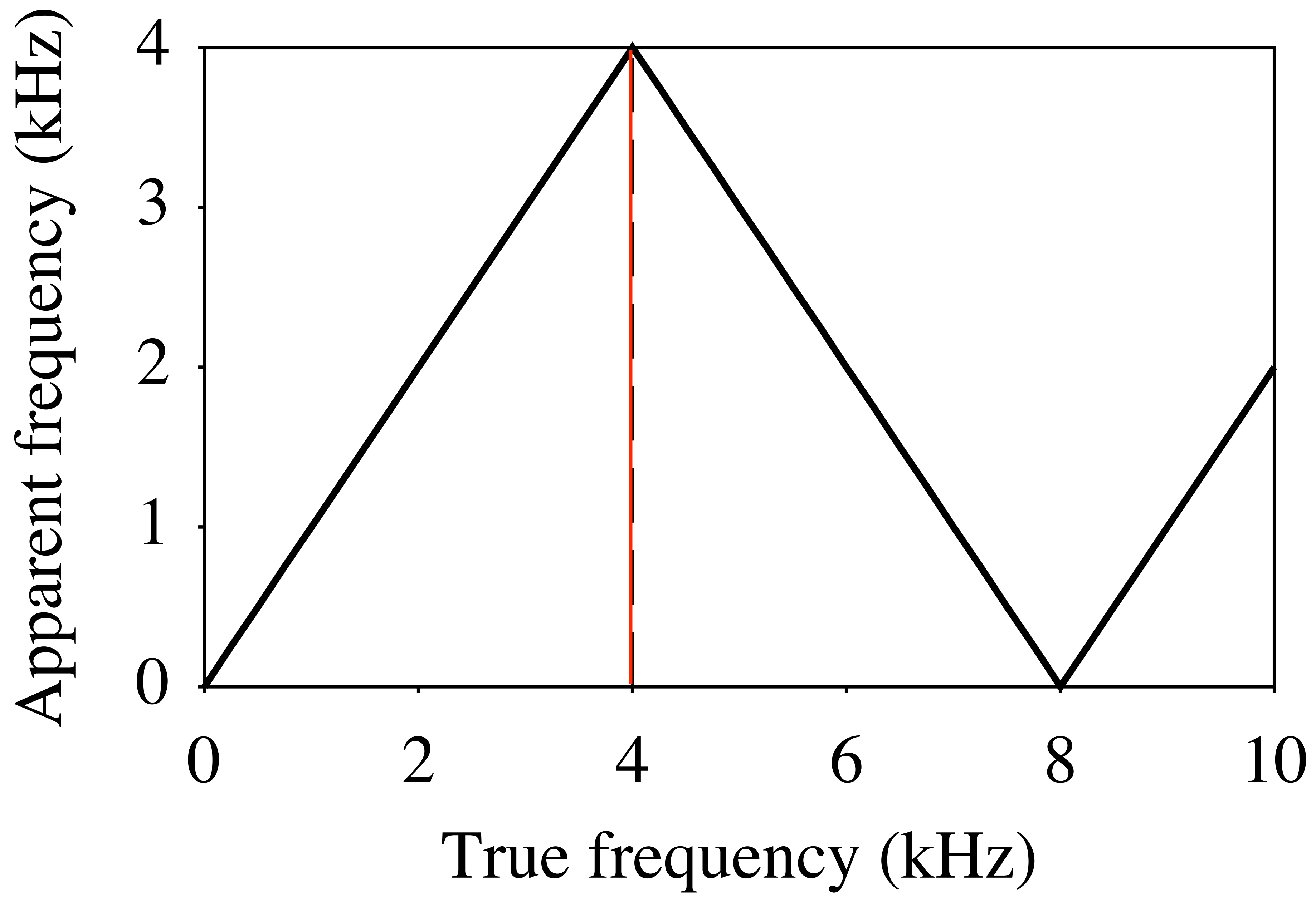
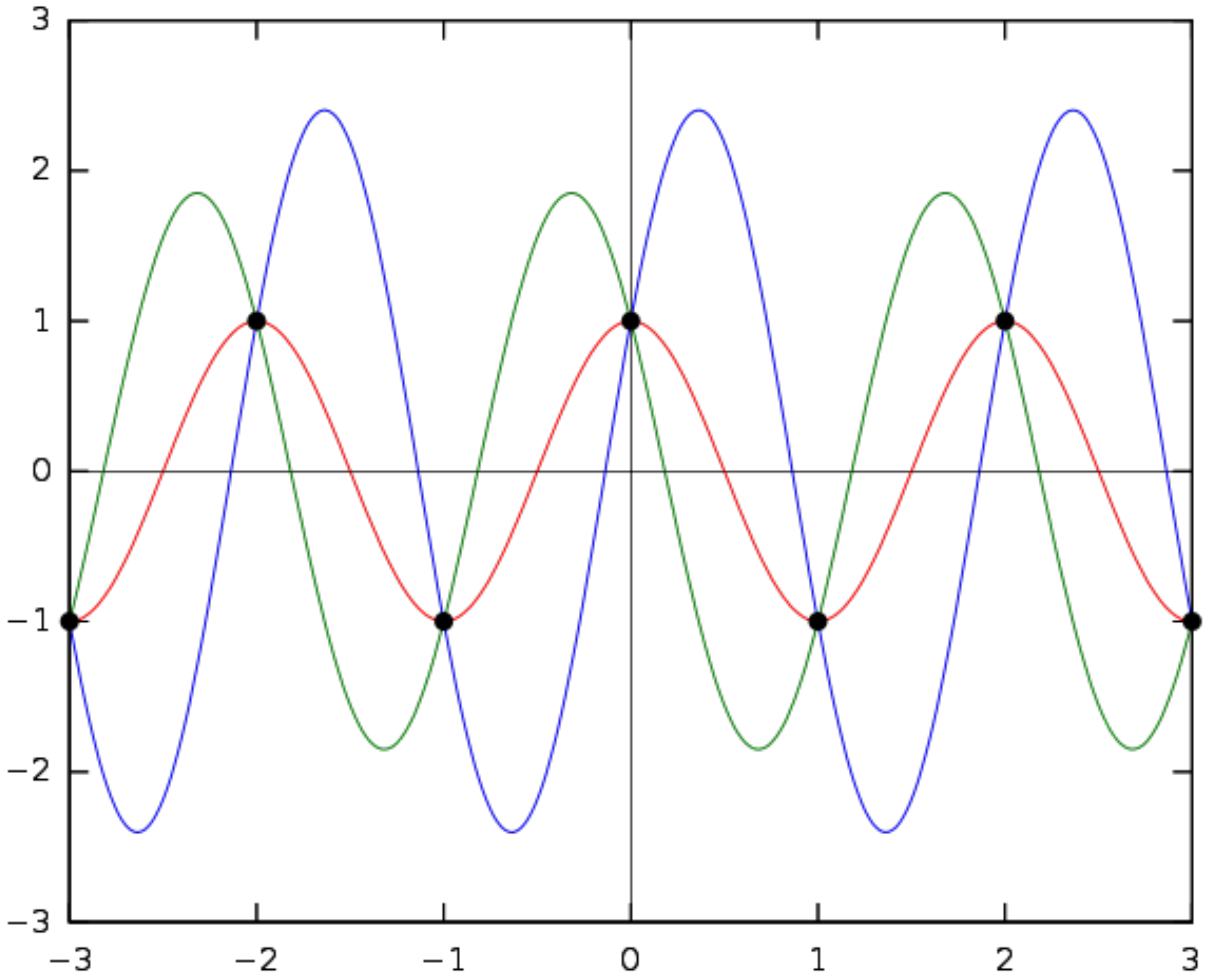
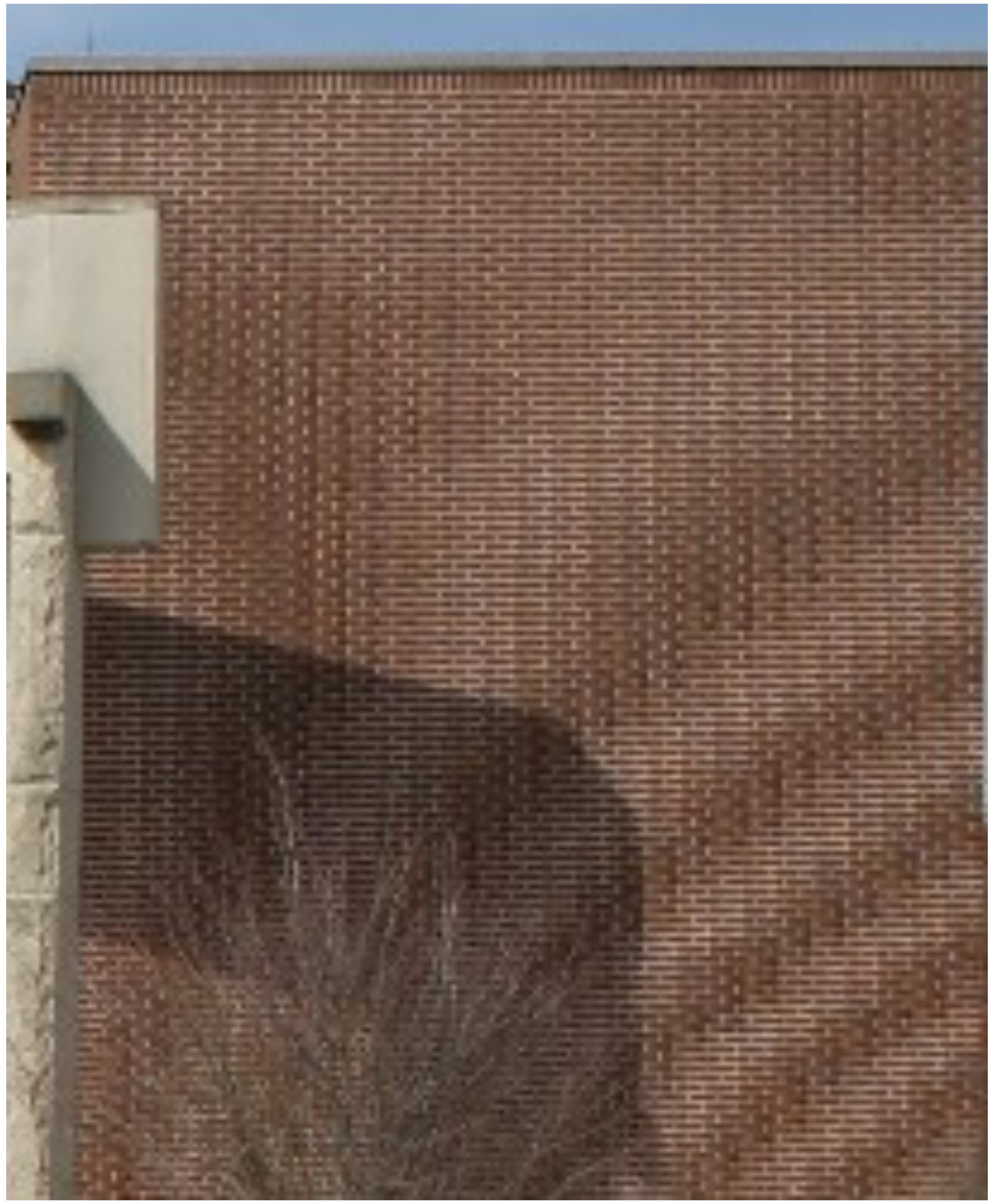
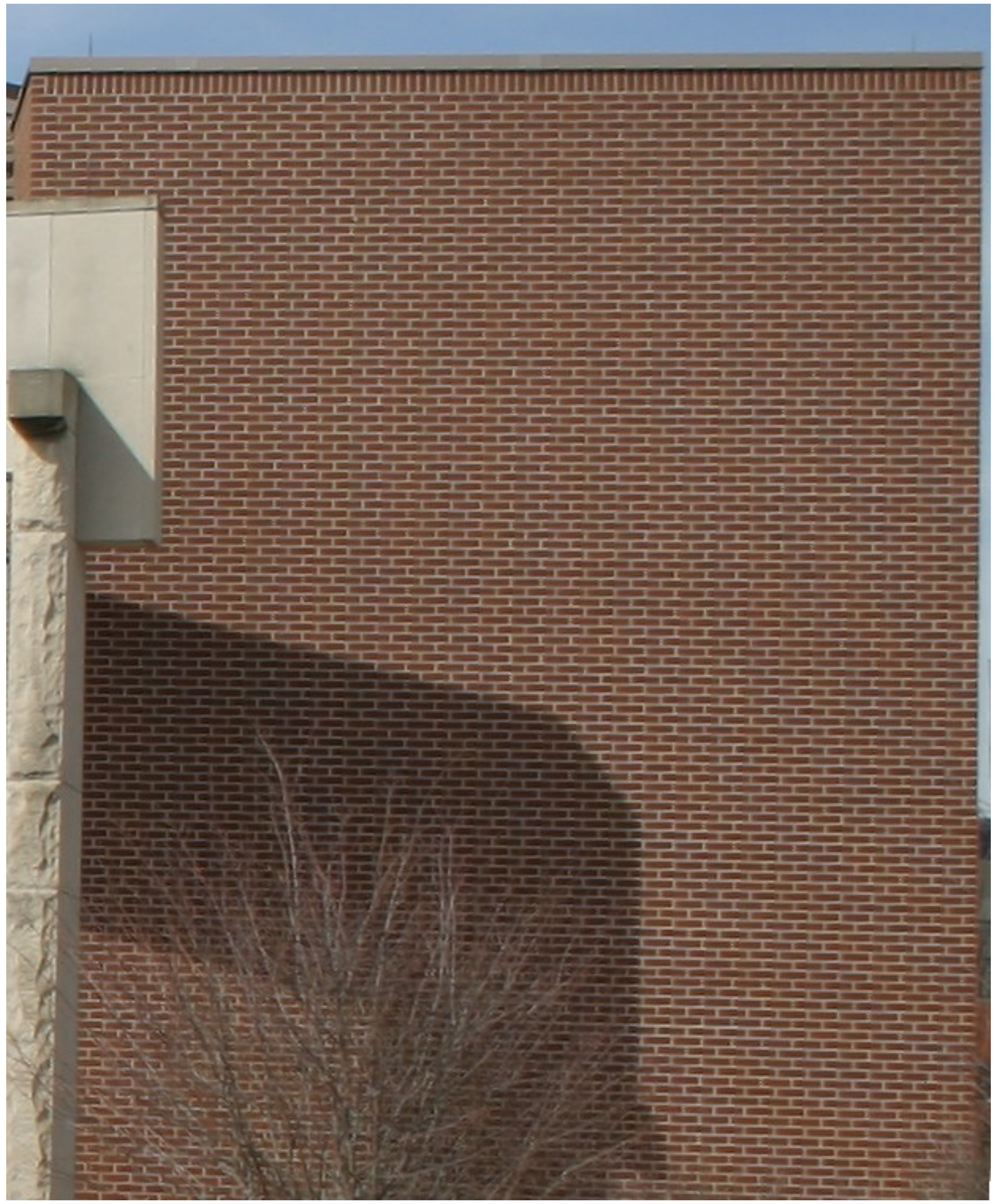


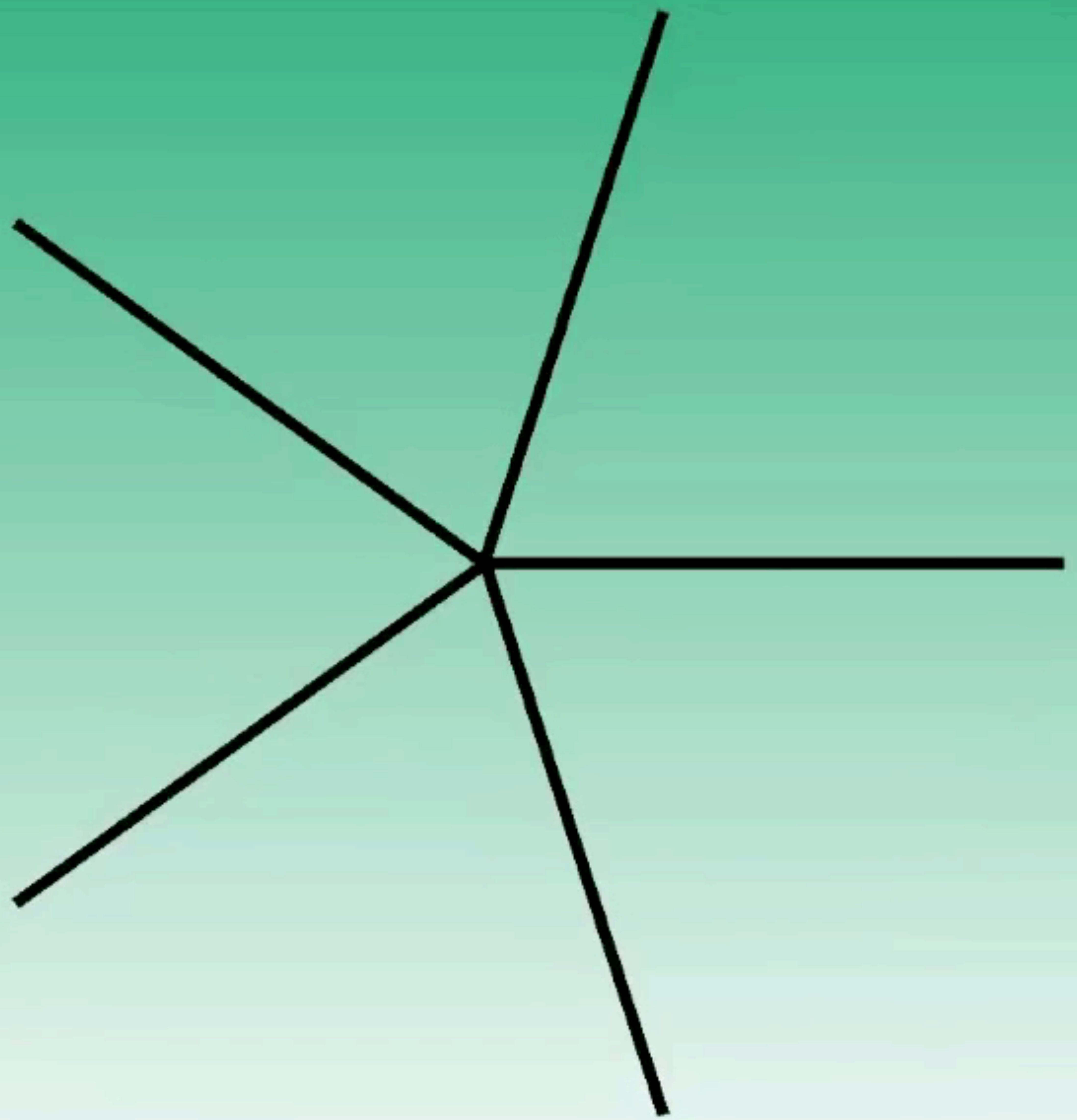
Fig. 6.5: Folding of sinusoid frequency which is sampled at 8,000 Hz. The folding frequency, shown dashed, is 4,000 Hz.



3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3







Source video: ***https://www.youtube.com/watch?v=SFblNinFsxk***

Audio Filtering

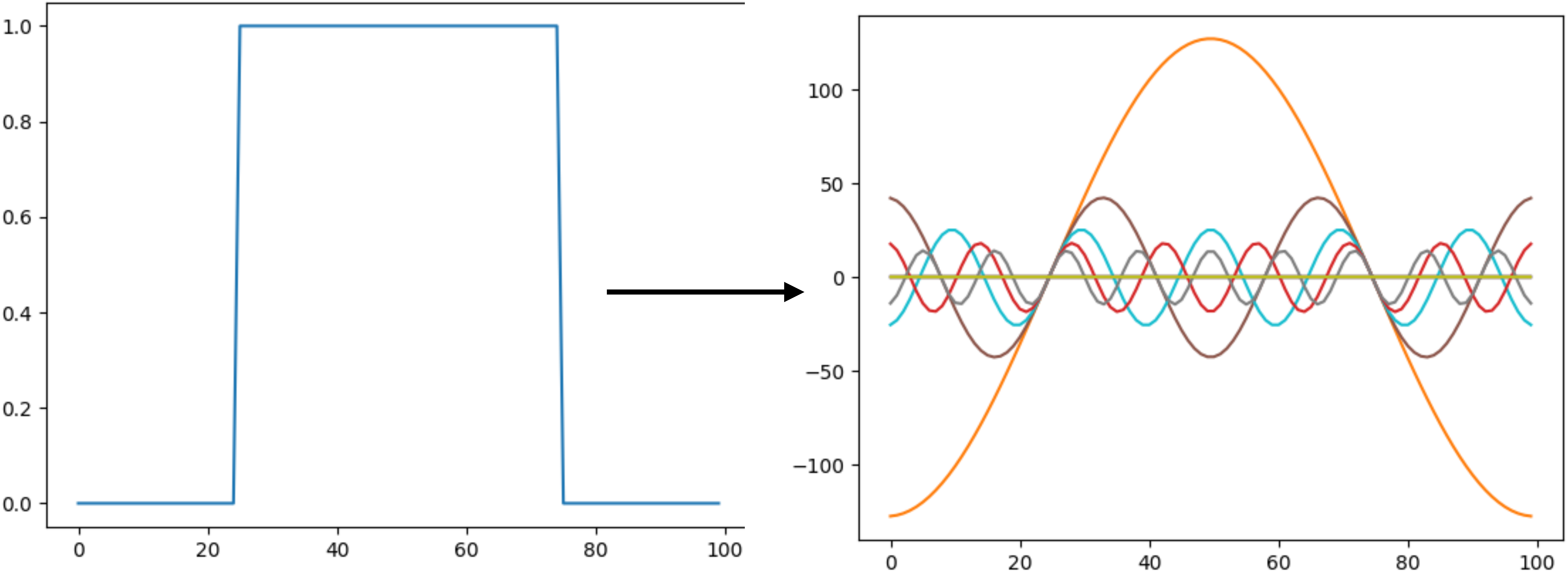
- Prior to sampling and AD conversion, the audio signal is also usually *filtered* to remove unwanted frequencies. The frequencies kept depend on the application:

  (a) For speech, typically from 50Hz to 10kHz is retained, and other frequencies are blocked by the use of a **band-pass filter** that screens out lower and higher frequencies.

  (b) An audio music signal will typically contain from about 20Hz up to 20kHz.

  (c) At the DA converter end, high frequencies may reappear in the output — because of sampling and then quantization, smooth input signal is replaced by a series of step functions containing all possible frequencies.

  (d) So at the decoder side, a **lowpass** filter is used after the DA circuit.


1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
100
50
0
−50
−100
0
20
40
60
80
100

Recap

- **Sampling rate:** how often a given signal is sampled
- **Bandwidth:** maximal frequency present in a given signal
- **Nyquist rate:** how often a given signal should be sampled to avoid aliasing
- **Nyquist frequency:** half of the Nyquist rate
- **Aliasing:** when two signals are perceived the same
- **Folding frequency:** half of the sampling rate
- **Apparent frequency:** frequency at which a signal will be perceived
- **Alias frequency:** apparent frequency for aliased frequencies (i.e. those above the Nyquist frequency)
- **Antialiasing filter:** removes a signal's frequencies that go above the Nyquist frequency

- Text book [7] *Fundamentals of Multimedia* Chapter(s): 6.1

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.

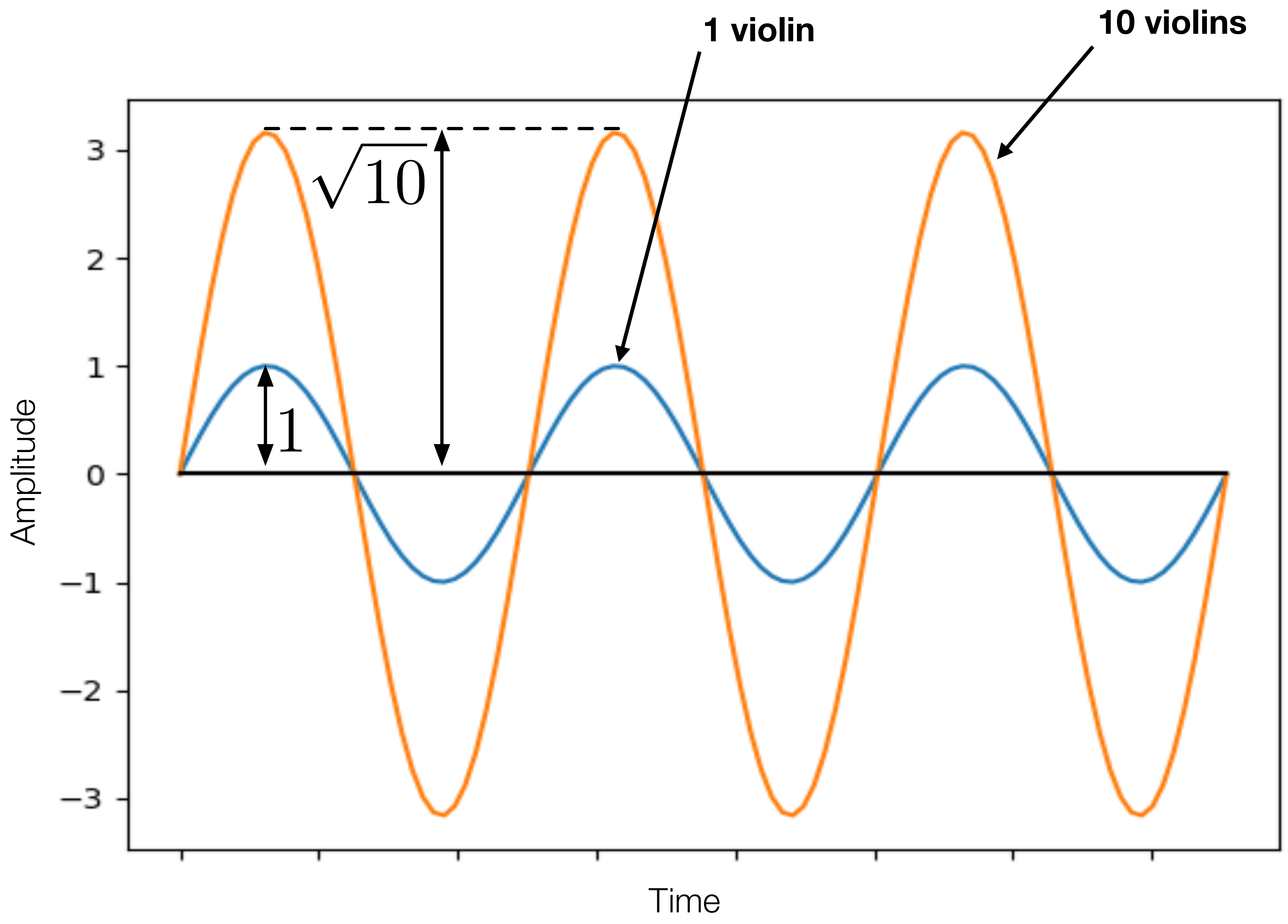
- The SNR is usually measured in decibels (**dB**), where 1 dB is a tenth of a **bel**. The SNR value, in units of dB, is defined in terms of base-10 logarithms of squared voltages, as follows:

$$SNR = 10 \log_{10} \frac{V_{signal}^2}{V_{noise}^2} = 20 \log_{10} \frac{V_{signal}}{V_{noise}} \tag{6.2}$$

a) The power in a signal is proportional to the square of the voltage. For example, if the signal voltage V_{signal} is 10 times the noise, then the SNR is $20 * \log_{10}(10) = 20$dB.

b) In terms of power, if the power from ten violins is ten times that from one violin playing, then the ratio of power is 10dB, or 1B.

c) *To know:* Power — 10; Signal Voltage — 20.



1 violin
10 violins
$\sqrt{10}$
1
3
2
1
0
−1
−2
−3
Amplitude
Time

Threshold of Hearing

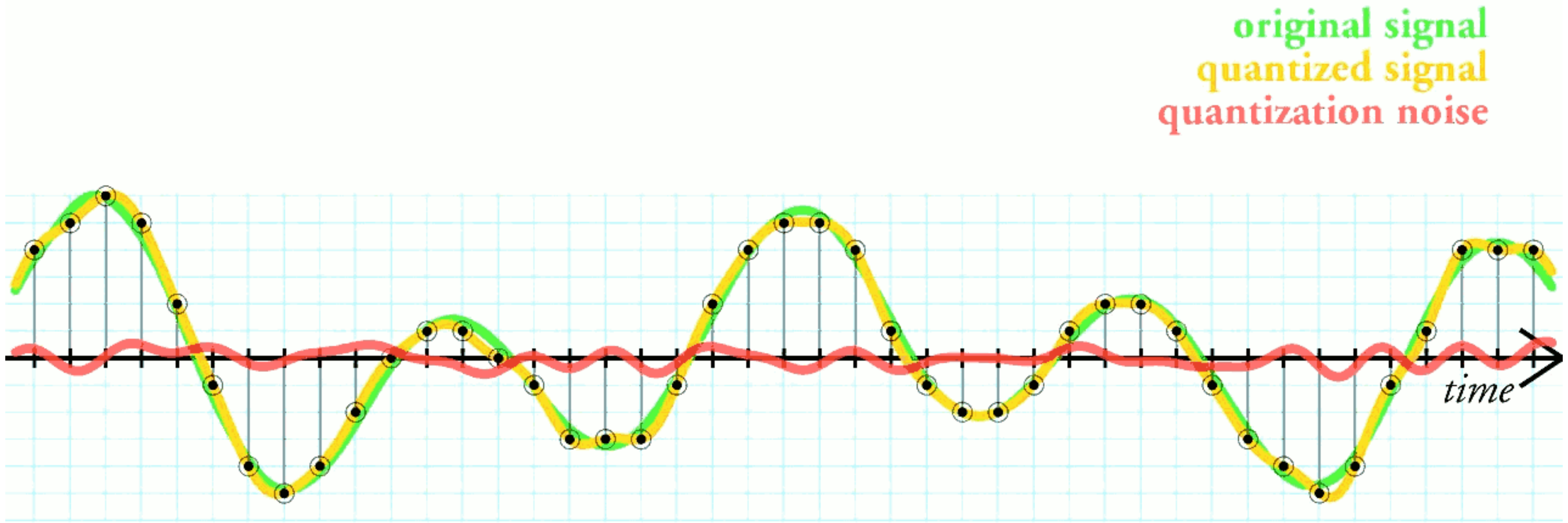
- The **absolute threshold of hearing** (ATH) is the minimum sound level of a pure tone that an average ear with normal hearing can hear with no other sound present. The absolute threshold relates to the sound that can just be heard by the organism.[1][2] The absolute threshold is not a discrete point, and is therefore classed as the point at which a response is elicited a specified percentage of the time.[1] This is also known as the auditory threshold.
- The threshold of hearing is generally reported as the RMS sound pressure of 20 μPa (micropascals) = 2×10^{-5} pascal (Pa). It is approximately the quietest sound a young human with undamaged hearing can detect at 1,000 Hz.[3] The threshold of hearing is frequency dependent and it has been shown that the ear's sensitivity is best at frequencies between 1 kHz and 5 kHz.[3]

- The usual levels of sound we hear around us are described in terms of decibels, as a ratio to the quietest sound we are capable of hearing. Table 6.1 shows approximate levels for these sounds.

Table 6.1: Magnitude levels of common sounds, in decibels

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

Quantization



- Sampled signal values are represented as discretized digital binary numerical values
  - A binary numerical value with limited number of bits has a limited value range it can represent
- Quantized amplitude values inherently introduce an error, also revered to as quantization noise
- Uniform quantization linearly distributes binary numbers over a given value range interval

Signal to Quantization Noise Ratio (SQNR)

- Aside from any noise that may have been present in the original analog signal, there is also an additional error that results from quantization.

(a) 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.

(b) This introduces a roundoff error. It is not really "noise". Nevertheless it is called **quantization noise** (or quantization error).

- The quality of the quantization is characterized by the Signal to Quantization Noise Ratio (**SQNR**).

  (a) **Quantization noise**: the difference between the actual value of the analog signal, for the particular sampling time, and the nearest quantization interval value.

  (b) At most, this error can be as much as half of the interval.

(c) For a quantization accuracy of N bits per sample, the SQNR can be simply expressed:

$$SQNR = 20 \log_{10} \frac{V_{signal}}{V_{quan\_noise}} = 20 \log_{10} \frac{2^{N-1}}{\frac{1}{2}}$$

$$= 20 \times N \times \log 2 = 6.02\, N (\mathrm{dB}) \qquad (6.3)$$

- Notes:

(a) We map the maximum signal to $2^{N-1} - 1$ ($\simeq 2^{N-1}$) and the most negative signal to -2^{N-1} .

(b) Eq. (6.3) is the *Peak* signal-to-noise ratio, PSQNR: peak signal and peak noise.

- $6.02N$ **is the worst case**. If the input signal is sinusoidal, the quantization error is statistically independent, and its magnitude is uniformly distributed between 0 and half of the interval, then it can be shown that the expression for the SQNR becomes:

$$SQNR = 6.02N + 1.76(dB) \tag{6.4}$$

(c) The *dynamic range* is the ratio of maximum to minimum absolute values of the signal: V_{max}/V_{min}. The max abs. value V_{max} gets mapped to $2^{N-1} - 1$; the min abs. value V_{min} gets mapped to 1. V_{min} is the smallest positive voltage that is not masked by noise. The most negative signal, $-V_{max}$, is mapped to -2^{N-1}.

(d) The quantization interval is $\Delta V = (2V_{max})/2^N$, since there are 2^N intervals. The whole range V_{max} down to $(V_{max} - \Delta V/2)$ is mapped to $2^{N-1} - 1$.

(e) The maximum noise, in terms of actual voltages, is half the quantization interval: $\Delta V/2 = V_{max}/2^N$.

Example

What is the dynamic range of the human ear?

- Minimum magnitude: 0dB (absolute threshold of hearing)
- Maximum magnitude: 160dB (hearing loss)
- Maximum power / minimum power = 10^{16}
- Dynamic range = maximum voltage / minimum voltage = 10^8 (hundred millions)
- We need at least $\log_2 10^8 + 1 \approx 28$ bits to quantize this range (using linear quantization)

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:

$$\Delta\text{Response} \propto \Delta\text{Stimulus}/\text{Stimulus} \tag{6.5}$$

  - Inserting a constant of proportionality k, we have a differential equation that states:

$$dr = k\,(1/s)\,ds \tag{6.6}$$

with response r and stimulus s.

Review Weber's Law

- Difference in observed response proportional to relative change in signal values
  - Change of observed response Δr
  - Change in signal values Δs, signal value (before change) s
  - Relative change in values $\dfrac{\Delta s}{s}$
- Observed perceptual (e.g. visual) effect $\Delta r = k \cdot \dfrac{\Delta s}{s}$
  - Proportionality konstant k

– Integrating, we arrive at a solution

$$r = k \ln s + C \tag{6.7}$$

with constant of integration C.
Stated differently, the solution is

$$r = k \ln(s/s_0) \tag{6.8}$$

s_0 = the lowest level of stimulus that causes a response ($r = 0$ when $s = s_0$).

- 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 equations for these very similar encodings are as follows:

μ-law:

$$r = \frac{\text{sgn}(s)}{\ln(1+\mu)} \ln\left\{1 + \mu\left|\frac{s}{s_p}\right|\right\}, \qquad \left|\frac{s}{s_p}\right| \le 1 \tag{6.9}$$

A-law:

$$r = \begin{cases} \frac{A}{1+\ln A}\left(\frac{s}{s_p}\right), & \left|\frac{s}{s_p}\right| \le \frac{1}{A} \\ \frac{\text{sgn}(s)}{1+\ln A}\left[1 + \ln A\left|\frac{s}{s_p}\right|\right], & \frac{1}{A} \le \left|\frac{s}{s_p}\right| \le 1 \end{cases} \tag{6.10}$$

$$\text{where} \quad \text{sgn}(s) = \begin{cases} 1 & \text{if} \quad s > 0, \\ -1 & \text{otherwise} \end{cases}$$

- Fig. 6.6 shows these curves. The parameter μ is set to $\mu = 100$ or $\mu = 255$; the parameter A for the A-law encoder is usually set to $A = 87.6$.

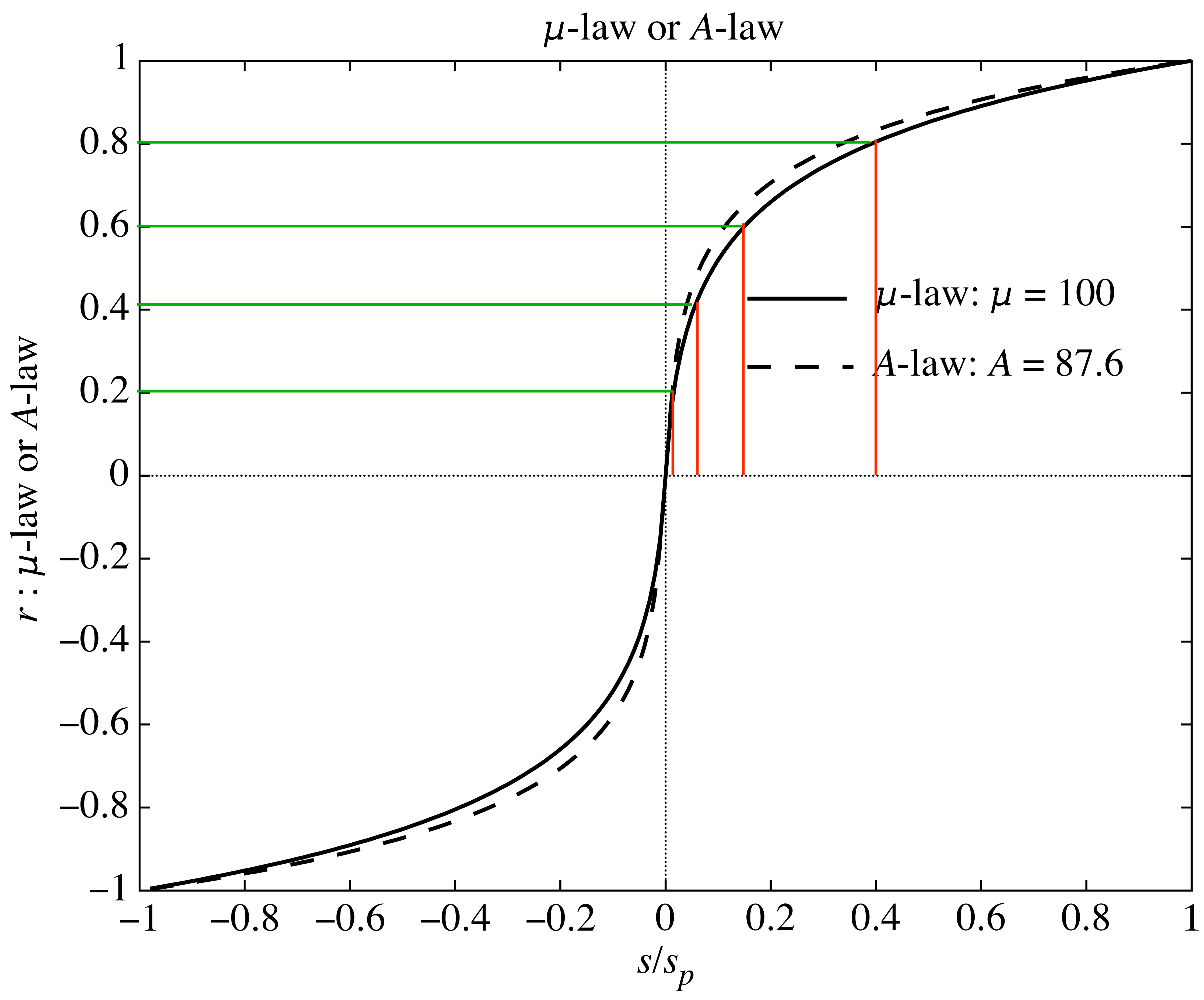


Fig. 6.6: Nonlinear transform for audio signals

- 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.

Recap

- **Amplitude:** maximal signal value (height of the wave)
- **Voltage:** same as amplitude, in the electric world
- **Power:** (proportional to the) squared voltage
- **Quantization noise:** error introduced in a signal when quantizing its voltage
- **Dynamic range:** Maximal voltage divided by minimal (closest to 0, but not 0) voltage
- **Quantization interval:** Maximal voltage * 2 / 2^N
- **SQNR:** SNR formula for the quantisation noise
- **Absolute threshold of hearing:** minimal audible signal (set at 0dB)
- **Weber's Law:** perceived magnitude scales *logarithmically* with respect to input stimulus
- **μ-law and A-law:** non-linear quantization schemes (motivated by Weber's law)

- Text book [7] *Fundamentals of Multimedia* Chapter(s): 6.1