Sound Intro

Tamara Berg
Advanced Multimedia

Reminder

HW1 due Thurs, Feb 24.

deadline extended to Sunday, Feb 27.

Questions?

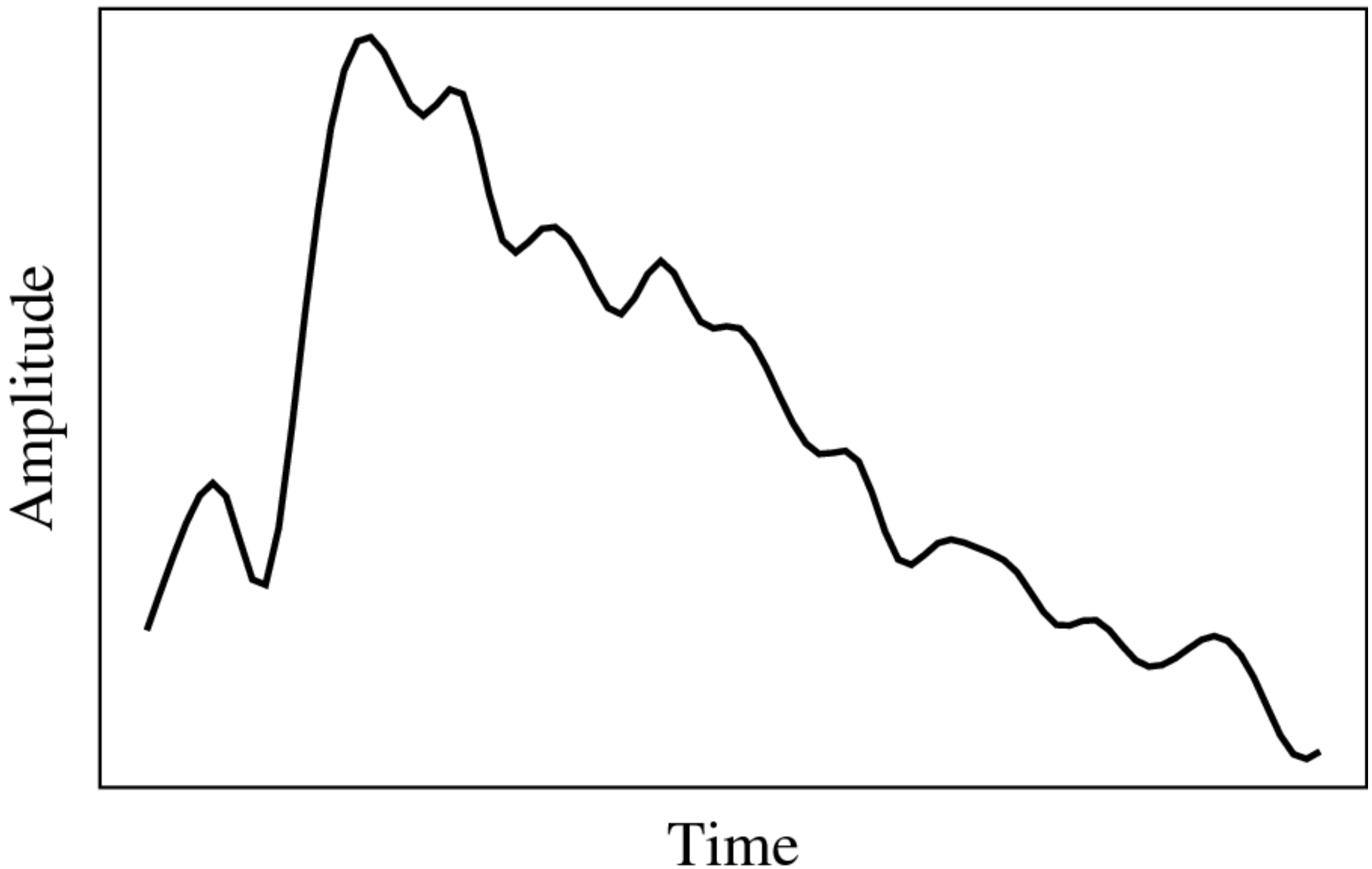
What is sound?

- 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.

  (a) For example, a speaker in an audio system vibrates back and forth and produces a *longitudinal* pressure wave that we perceive as sound.

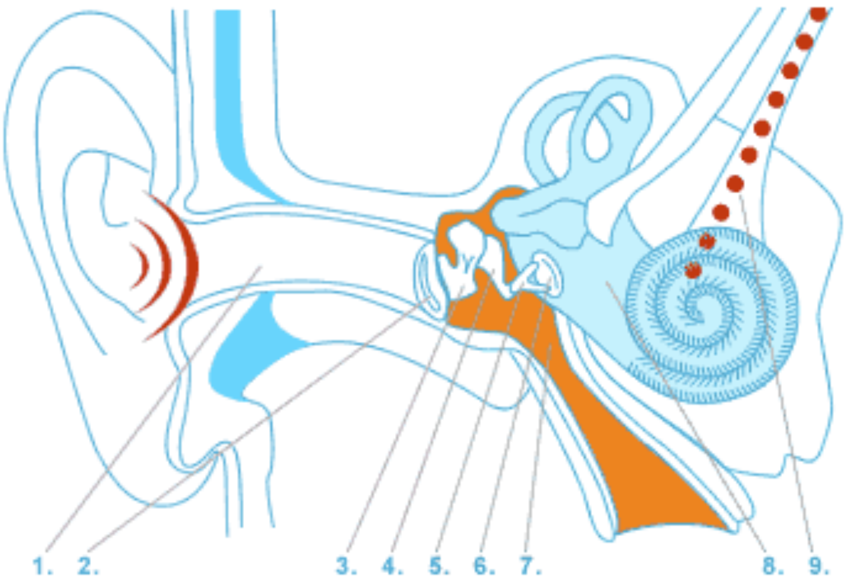
  (b) Since sound is a pressure wave, it takes on continuous values, as opposed to digitized ones.

Sound Wave



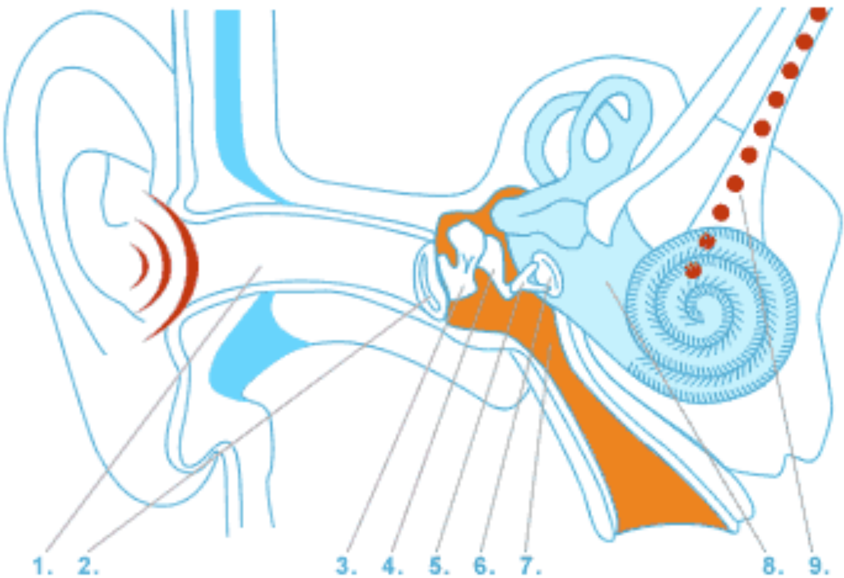
How does the ear work?

As the sound waves enter the ear, the ear canal increases the loudness of those pitches that make it easier to understand speech and protects the eardrum - a flexible, circular membrane which vibrates when touched by sound waves.



1. Ear canal
2. Eardrum
3-5. Ossicles
6. Oval window
7. Canal leading to the nose
8. Cochlea
9. Auditory nerve

How does the ear work?

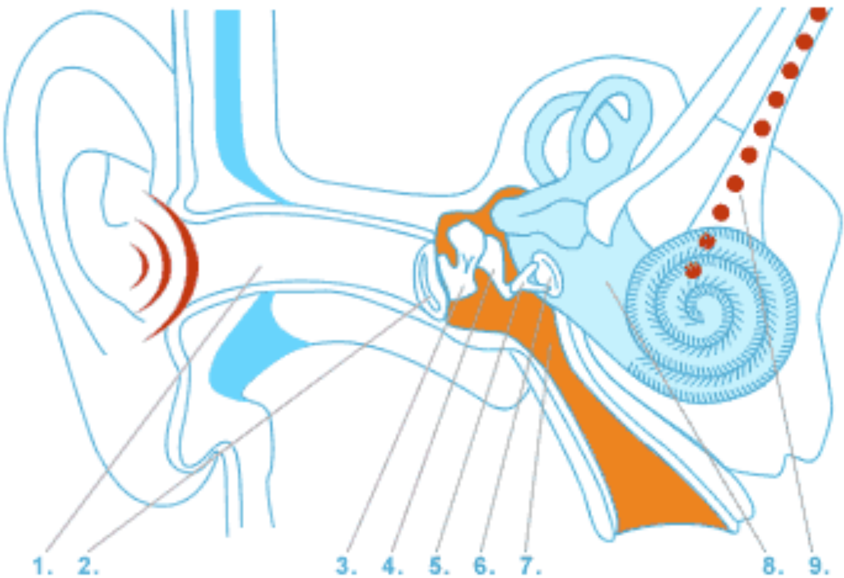


1. Ear canal
2. Eardrum
3-5. Ossicles
6. Oval window
7. Canal leading to the nose
8. Cochlea
9. Auditory nerve

As the sound waves enter the ear, the ear canal increases the loudness of those pitches that make it easier to understand speech and protects the eardrum - a flexible, circular membrane which vibrates when touched by sound waves.

The sound vibrations continue their journey into the middle ear, which contains three tiny bones called the ossicles, which are also known as the hammer, anvil and stirrup. These bones form the bridge from the eardrum into the inner ear. They increase and amplify the sound vibrations even more, before safely transmitting them on to the inner ear via the oval window.

How does the ear work?



1. Ear canal
2. Eardrum
3-5. Ossicles
6. Oval window
7. Canal leading to the nose
8. Cochlea
9. Auditory nerve

As the sound waves enter the ear, the ear canal increases the loudness of those pitches that make it easier to understand speech and protects the eardrum - a flexible, circular membrane which vibrates when touched by sound waves.

The sound vibrations continue their journey into the middle ear, which contains three tiny bones called the ossicles, which are also known as the hammer, anvil and stirrup. These bones form the bridge from the eardrum into the inner ear. They increase and amplify the sound vibrations even more, before safely transmitting them on to the inner ear via the oval window.

The Inner Ear (cochlea), houses a system of tubes filled with a watery fluid. As the sound waves pass through the oval window the fluid begins to move, setting tiny hair cells in motion. In turn, these hairs transform the vibrations into electrical impulses that travel along the auditory nerve to the brain itself.

link

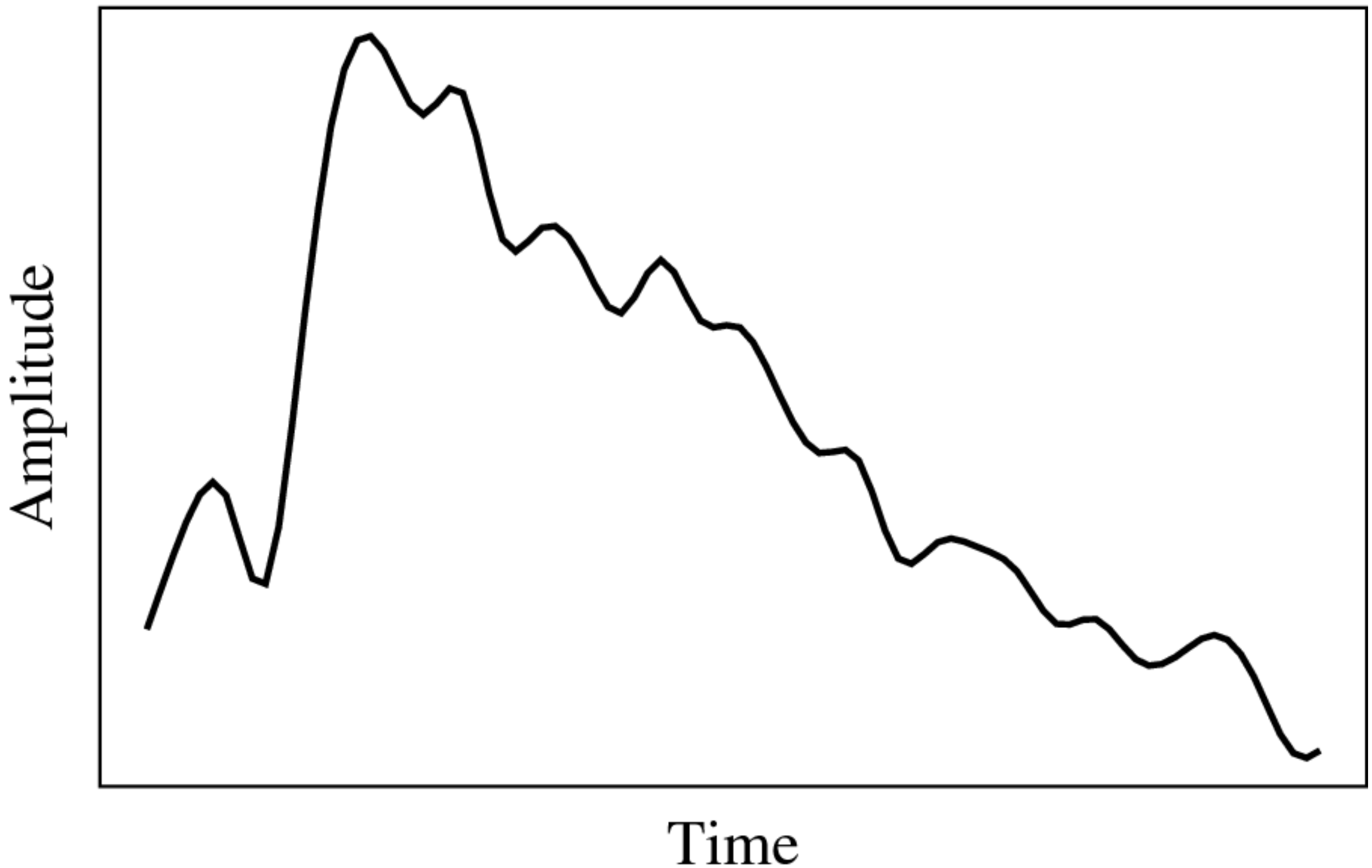
(c) Even though such pressure waves are longitudinal, they still have ordinary wave properties and behaviors, such as reflection (bouncing), refraction (change of angle when entering a medium with a different density) and diffraction (bending around an obstacle).

(d) If we wish to use a digital version of sound waves we must form digitized representations of audio information.

→ Link to physical description of sound waves.

Digitization

- **Digitization** means conversion to a stream of numbers, and preferably these numbers should be integers for efficiency.



An analog signal: continuous measurement of pressure wave.

• Sound is 1-dimensional (**amplitude** values depend on a 1D variable, time) as opposed to images (which are how many dimensions)?

- 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*.

  (c) 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, discussed later.

  (d) Sound is a continuous signal (measurement of pressure). Sampling in the amplitude or voltage dimension is called **quantization**. We quantize so that we can represent the signal as a discrete set of values.

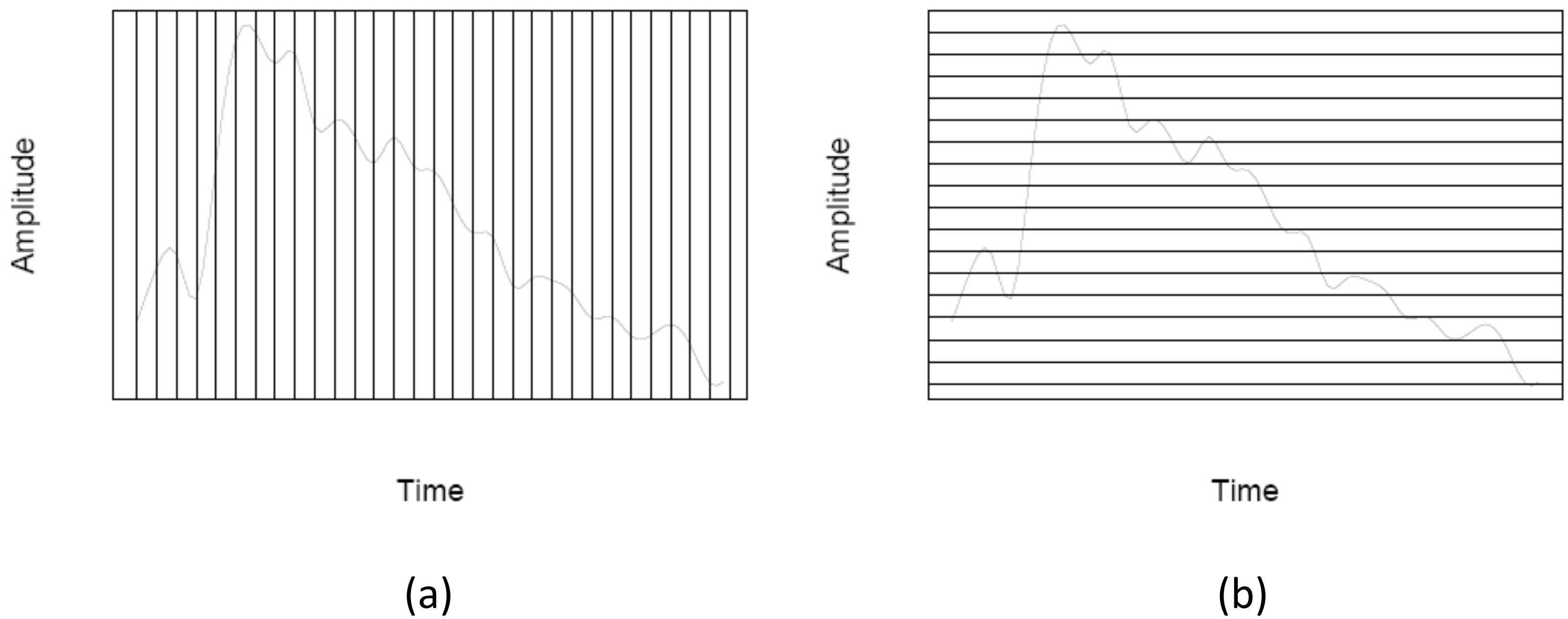
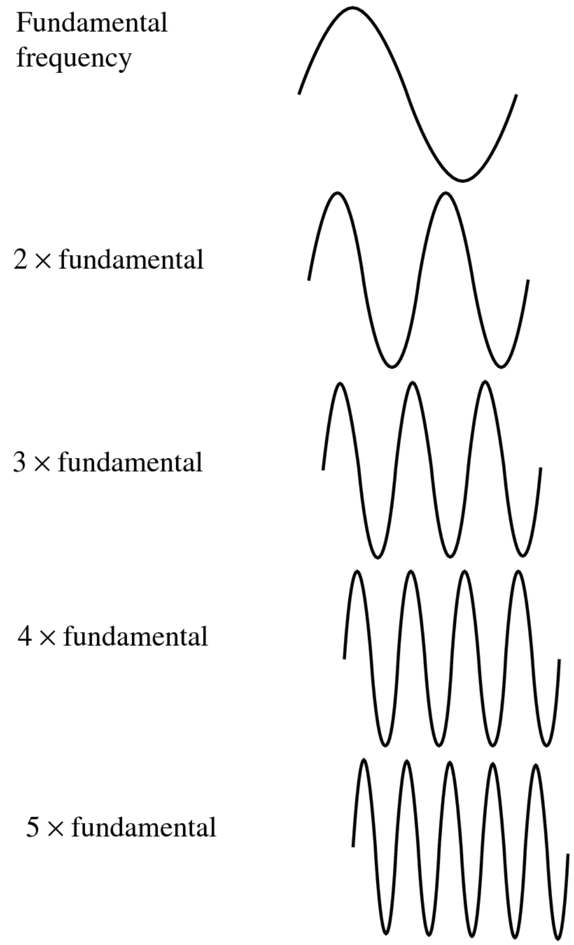
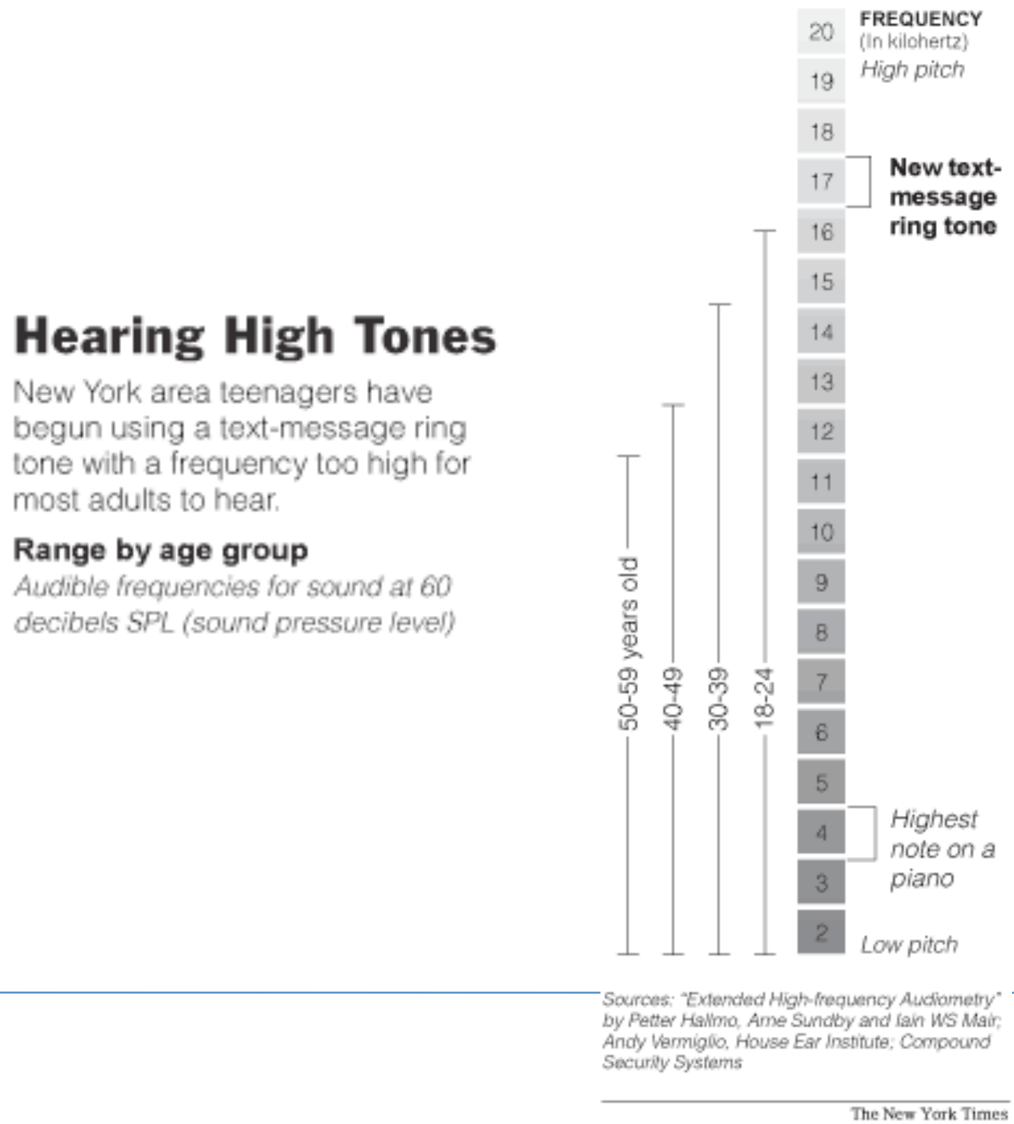


Fig. 6.2: Sampling and Quantization. (a): Sampling the analog signal in the time dimension. (b): Quantization is sampling the analog signal in the amplitude dimension.



Frequency of sound waves.

Hearing by Age Group



Article

[Mosquito Ringtones]

- Whereas **frequency** is an absolute measure, **pitch** is generally relative — a perceptual subjective quality of sound.

  (a) Pitch and frequency are linked by setting the note A above middle C to exactly 440 Hz.

- Whereas **frequency** is an absolute measure, **pitch** is generally relative — a perceptual subjective quality of sound.

  (a) Pitch and frequency are linked by setting the note A above middle C to exactly 440 Hz.

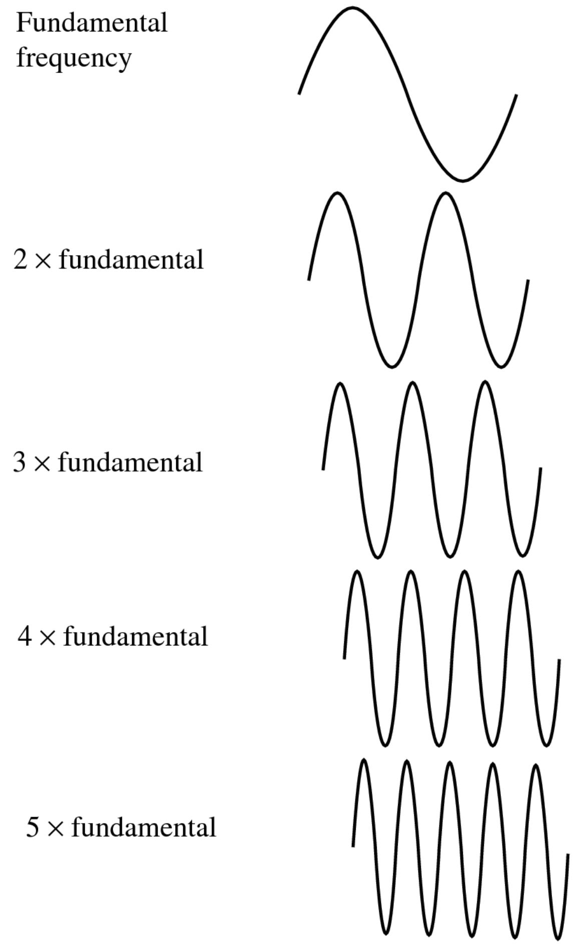
  (b) An **octave** above that note takes us to another A note. An octave corresponds to *doubling the frequency*. Thus with the middle “A” on a piano (“A4” or “A440”) set to 440 Hz, the next “A” up is at 880 Hz, or one octave above.

- Whereas **frequency** is an absolute measure, **pitch** is generally relative — a perceptual subjective quality of sound.

  (a) Pitch and frequency are linked by setting the note A above middle C to exactly 440 Hz.

  (b) An **octave** above that note takes us to another A note. An octave corresponds to *doubling the frequency*. Thus with the middle “A” on a piano (“A4” or “A440”) set to 440 Hz, the next “A” up is at 880 Hz, or one octave above.

  (c) **Harmonics**: any series of musical tones whose frequencies are integral multiples of the frequency of a fundamental tone.



Frequency of sound waves.

- Whereas **frequency** is an absolute measure, **pitch** is generally relative — a perceptual subjective quality of sound.

  (a) Pitch and frequency are linked by setting the note A above middle C to exactly 440 Hz.

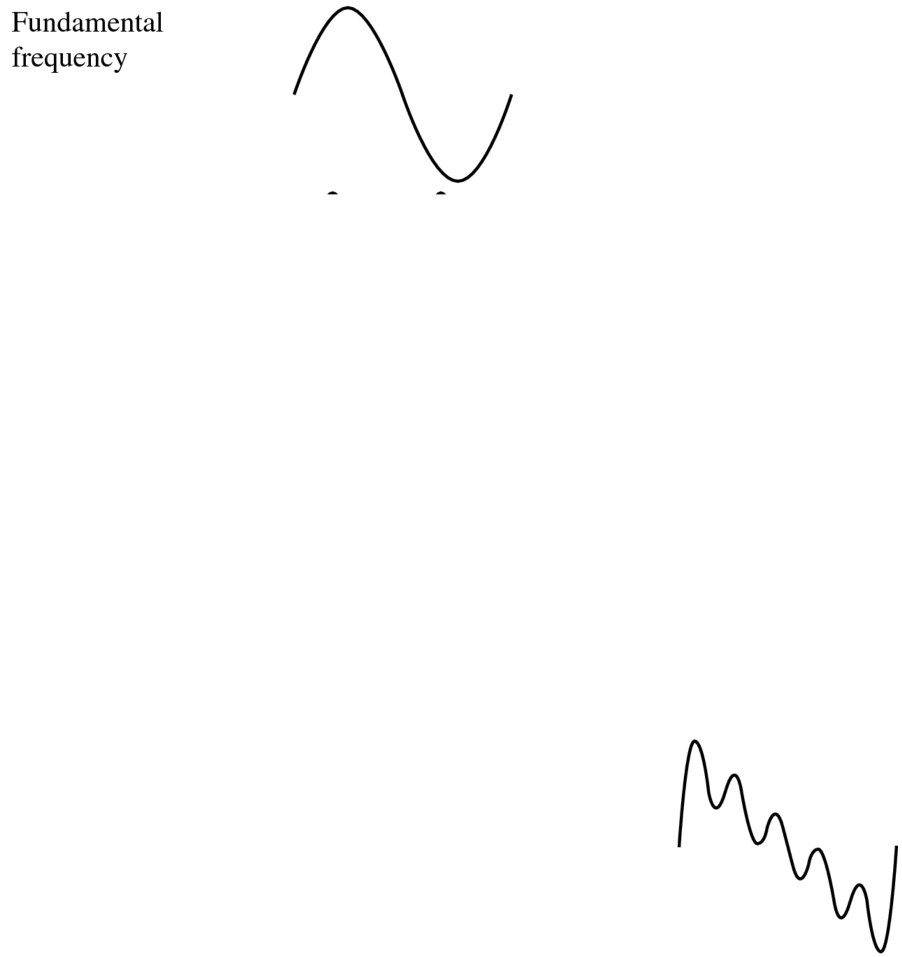
  (b) An **octave** above that note takes us to another A note. An octave corresponds to *doubling the frequency*. Thus with the middle "A" on a piano ("A4" or "A440") set to 440 Hz, the next "A" up is at 880 Hz, or one octave above.

  (c) **Harmonics**: any series of musical tones whose frequencies are integral multiples of the frequency of a fundamental tone.

  (d) If we allow non-integer multiples of the base frequency, we produce a more complex resulting sound.

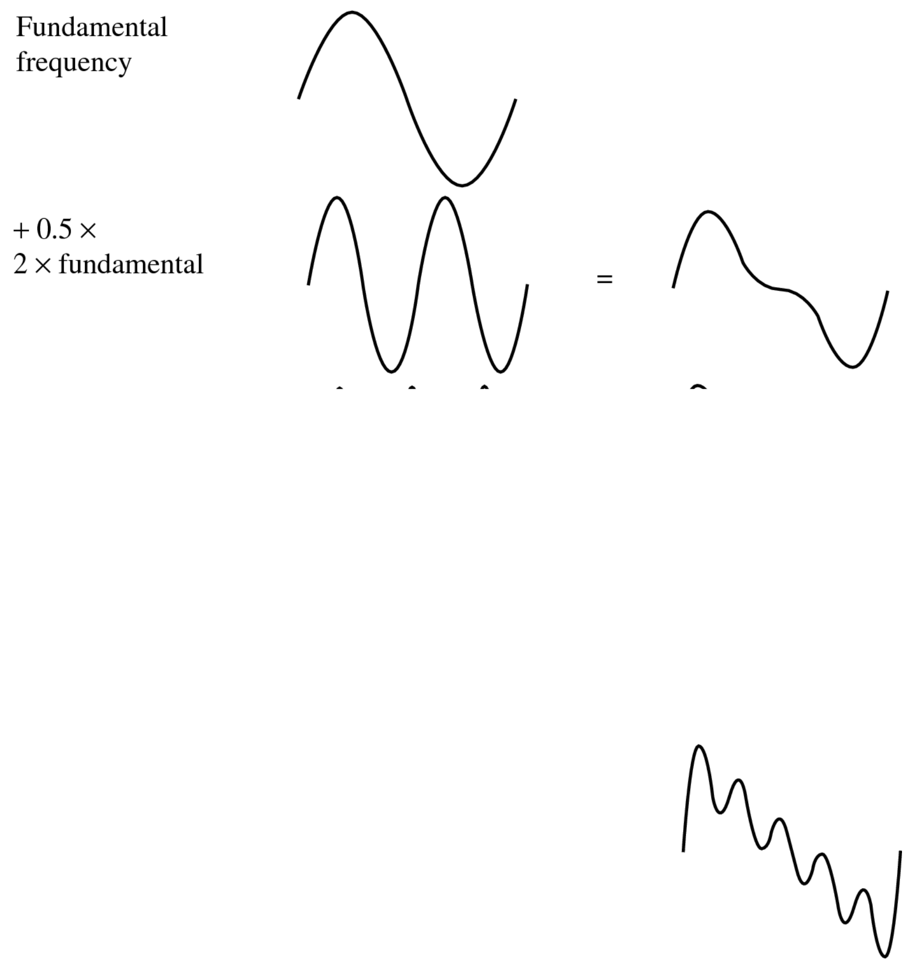
Signals can be decomposed into a weighted sum of sinusoids:

Building up a complex signal by superposing sinusoids



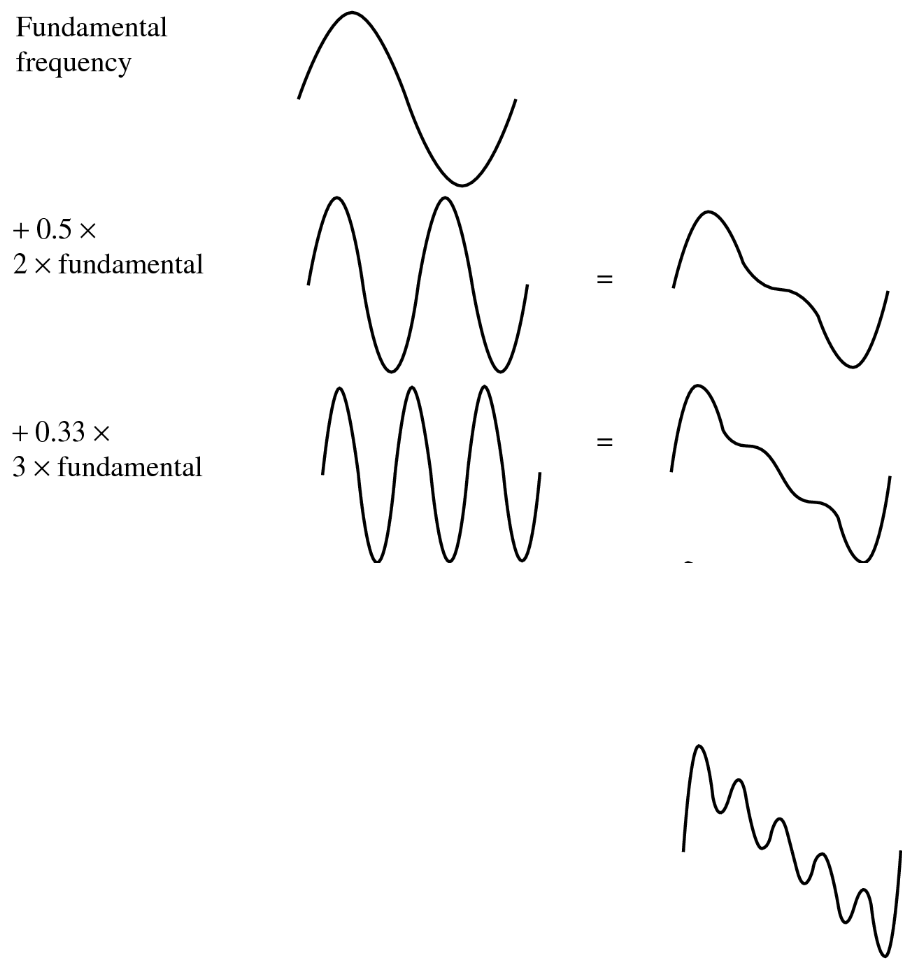
Signals can be decomposed into a weighted sum of sinusoids:

Building up a complex signal by superposing sinusoids



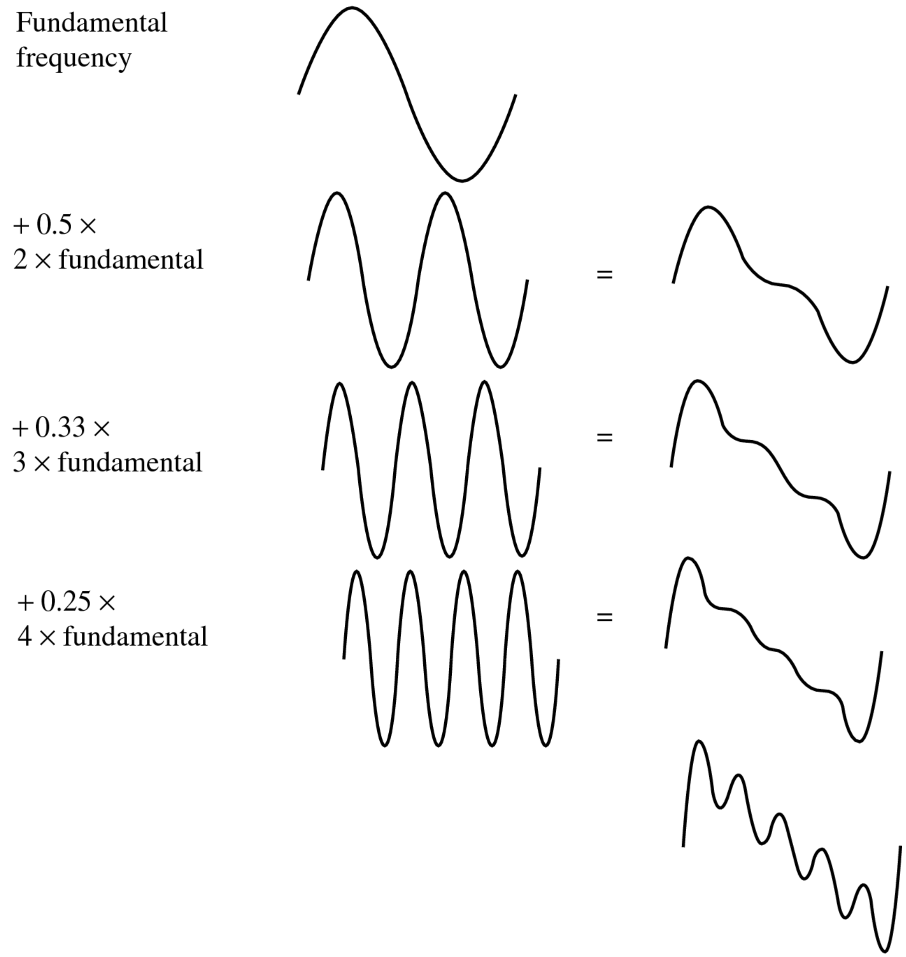
Signals can be decomposed into a weighted sum of sinusoids:

Building up a complex signal by superposing sinusoids



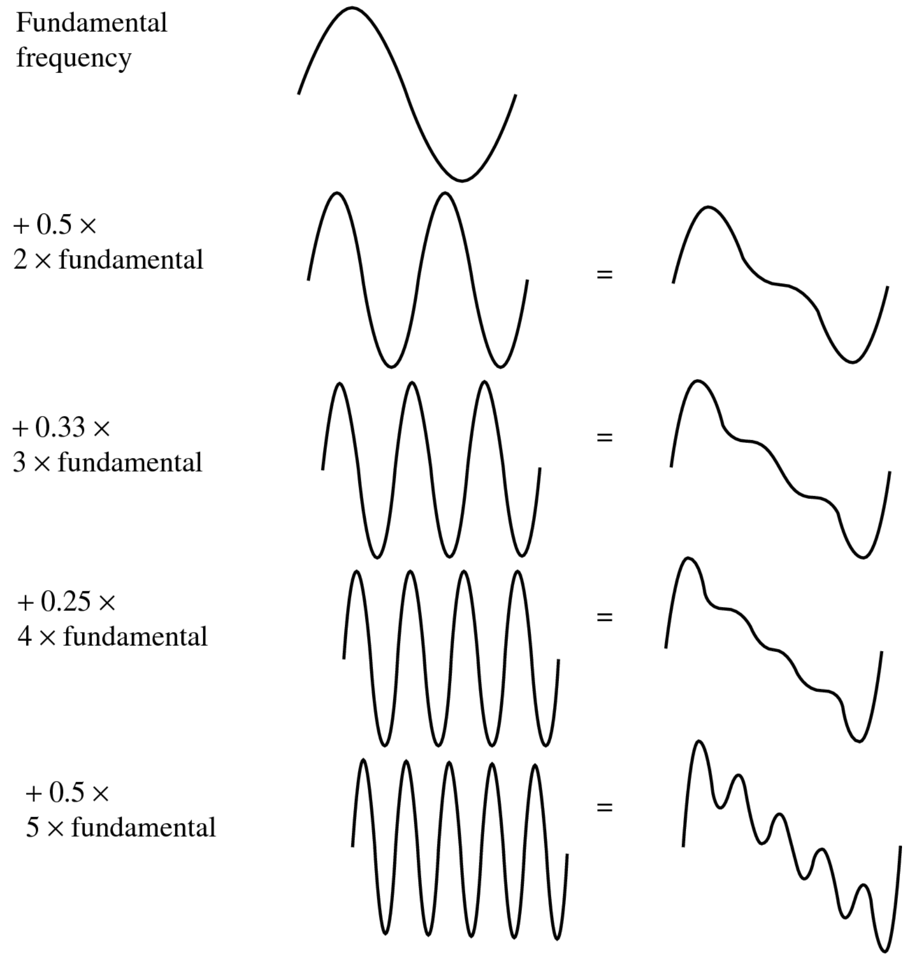
Signals can be decomposed into a weighted sum of sinusoids:

Building up a complex signal by superposing sinusoids



Signals can be decomposed into a weighted sum of sinusoids:

Building up a complex signal by superposing sinusoids



Signals can be decomposed into a weighted sum of sinusoids:

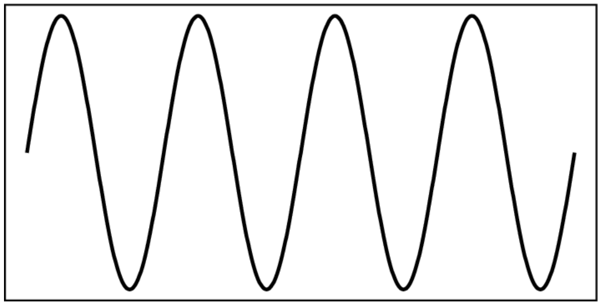
Building up a complex signal by superposing sinusoids

- 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?

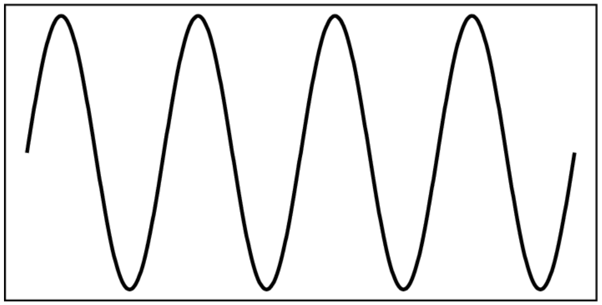
3. How is audio data formatted? (file format)



(a)

Fig. 6.4: Aliasing.

(a): A single frequency.


(a)

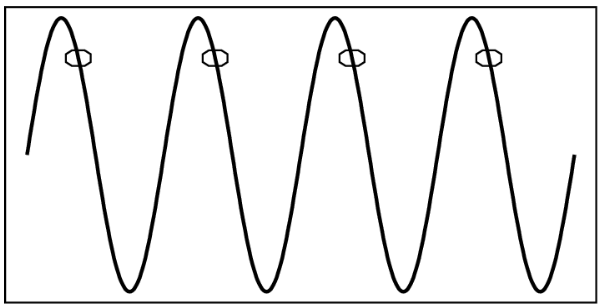
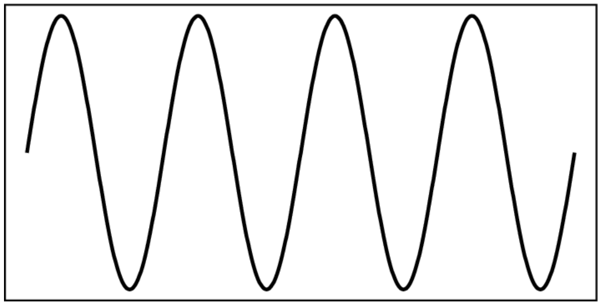
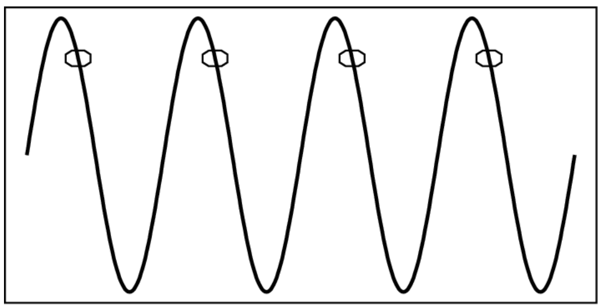

(b)

Fig. 6.4: Aliasing.

(a): A single frequency.



(a)



(b)

Fig. 6.4: Aliasing.

(a): A single frequency.

(b): Sampling at exactly the frequency produces a constant.

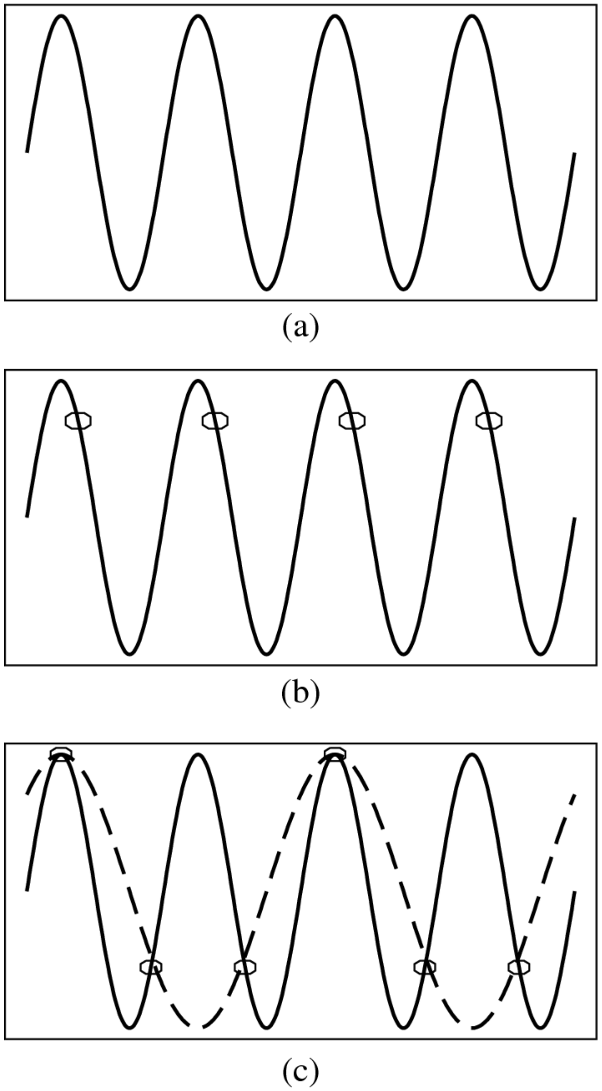


Fig. 6.4: Aliasing.

(a): A single frequency.

(b): Sampling at exactly the frequency produces a constant.

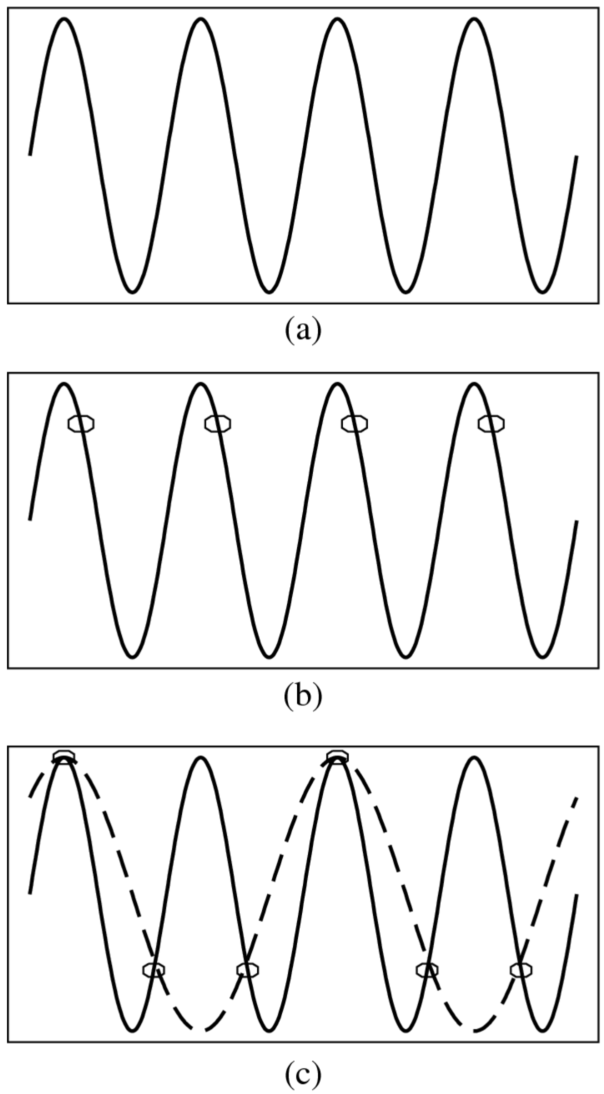


Fig. 6.4: Aliasing.

(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.

Aliasing

The relationship among the Sampling Frequency,

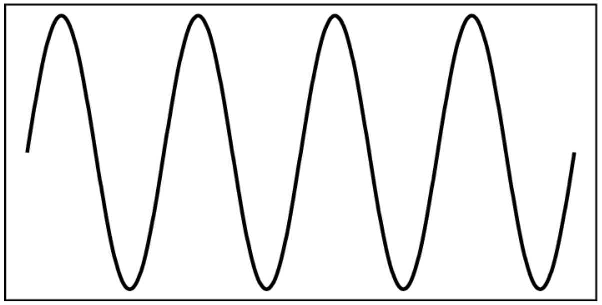
True Frequency, and the Alias Frequency is as

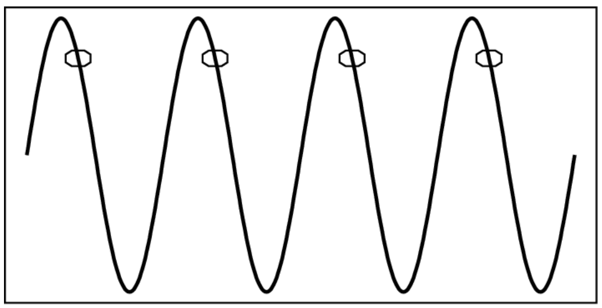
follows:

$$f_{alias} = f_{sampling} - f_{true}, \text{ for } f_{true} < f_{sampling} < 2 \times f_{true}$$

If true freq is 5.5 kHz and sampling freq is 8 kHz.

Then what is the alias freq?


(a)


(b)

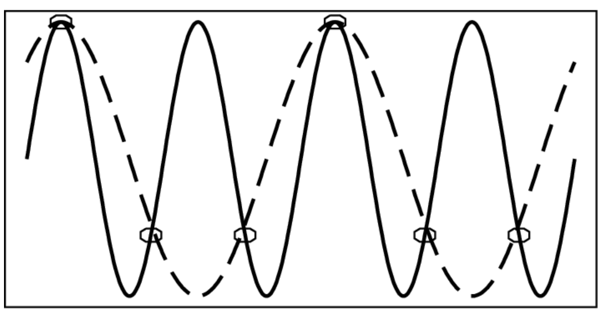
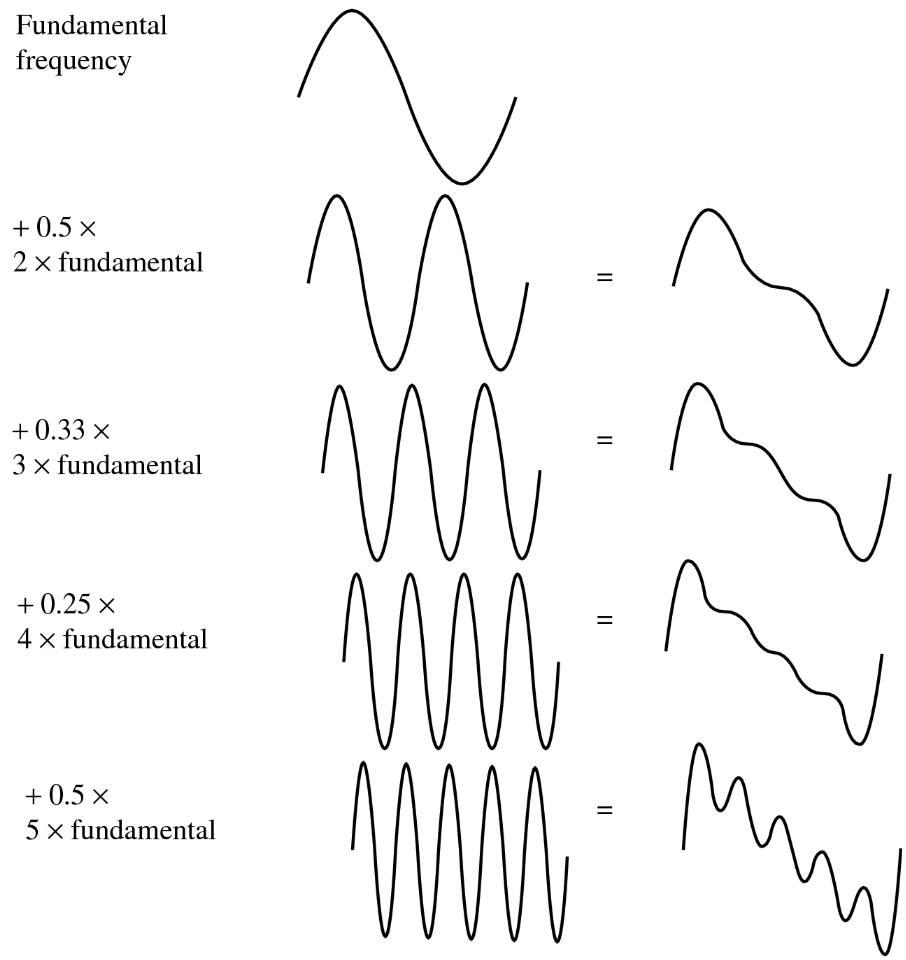

(c)

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.



Signals can be decomposed into a weighted sum of sinusoids:

Building up a complex signal by superposing sinusoids

- **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.

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 amplitudes, 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) For example, if the signal amplitude A_{signal} is 10 times the noise, then the SNR is

a) For example, if the signal amplitude A_{signal} is 10 times the noise, then the SNR is

$20 * \log_{10}(10) = 20\text{dB}$.

b) *dB always defined in terms of a ratio.*

- 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

Merits of dB

* The decibel's logarithmic nature means that a very large range of ratios can be represented by a convenient number. This allows one to clearly visualize huge changes of some quantity.

Merits of dB

* The decibel's logarithmic nature means that a very large range of ratios can be represented by a convenient number. This allows one to clearly visualize huge changes of some quantity.
* The mathematical properties of logarithms mean that the overall decibel gain of a multi-component system (such as consecutive amplifiers) can be calculated simply by summing the decibel gains of the individual components, rather than needing to multiply amplification factors. Essentially this is because log(A × B × C × ...) = log(A) + log(B) + log(C) + ...

Merits of dB

* The decibel's logarithmic nature means that a very large range of ratios can be represented by a convenient number. This allows one to clearly visualize huge changes of some quantity.
* The mathematical properties of logarithms mean that the overall decibel gain of a multi-component system (such as consecutive amplifiers) can be calculated simply by summing the decibel gains of the individual components, rather than needing to multiply amplification factors. Essentially this is because $\log(A \times B \times C \times ...) = \log(A) + \log(B) + \log(C) + ...$
* The human perception of sound is such that a doubling of actual intensity causes perceived intensity to always increase by the same amount, irrespective of the original level. The decibel's logarithmic scale, in which a doubling of power or intensity always causes an increase of approximately 3 dB, corresponds to this perception.

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.

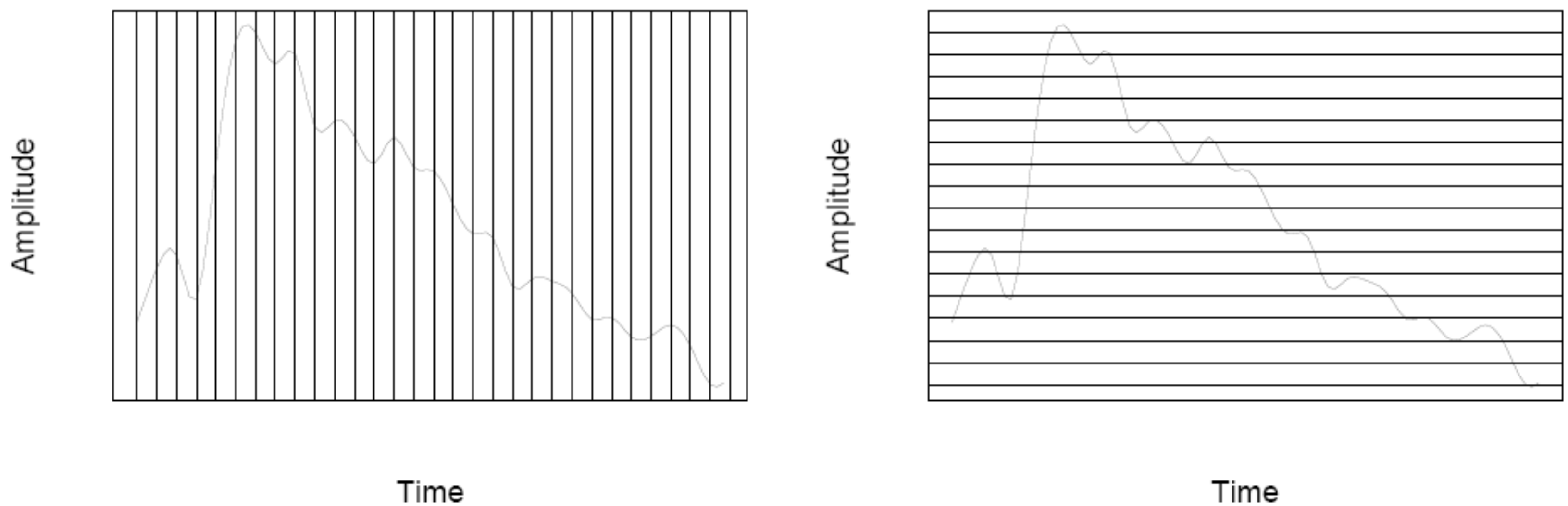
  (a) If voltages are actually in 0 to 1 but we have only 8 bits in which to store values, then effectively we force all continuous values of voltage into only 256 different values.

  (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.



(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.

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

$$SQNR = 20\log_{10}\frac{V_{signal}}{V_{quan\_noise}}$$

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

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

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

In the worst case

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.

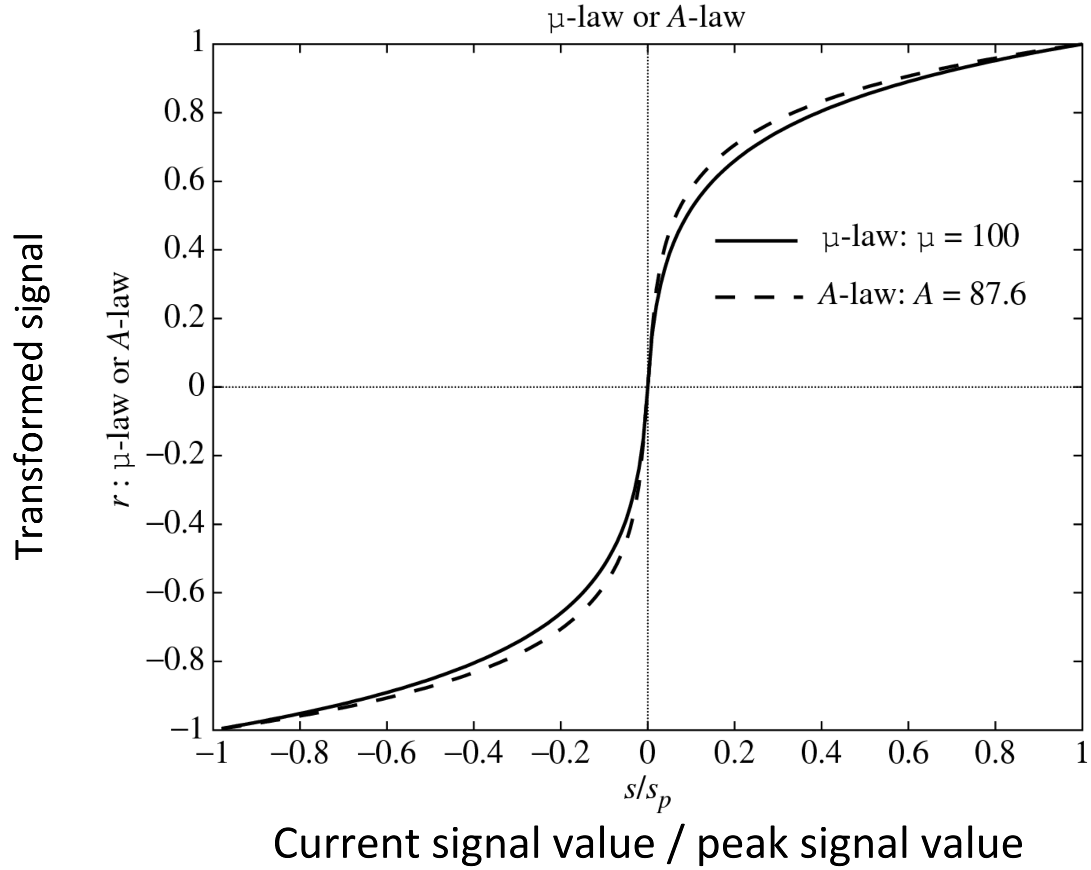
Linear and Non-linear Quantization

- **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.

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. A very similar rule, called A-**law**, is used in telephony in Europe.



Values in s get mapped to values in r non-uniformly. "Perceptual coder" – allocates more bits to intervals for which a small change produces a large change in perception.

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 (s/sp near 0).