Course Code 005636 (Fall 2017)

Multimedia

Basics of Digital Audio

Prof. S. M. Riazul Islam, Dept. of Computer Engineering, Sejong University, Korea

E-mail: riaz@sejong.ac.kr

Outline

- Digitization of Sound
  - Sampling and Quantization
  - Nyquist Theorem
  - Signal-to-Noise Ratio (SNR)
  - dB scale and sound measurement
- Overview of MIDI
- Transmission of Audio

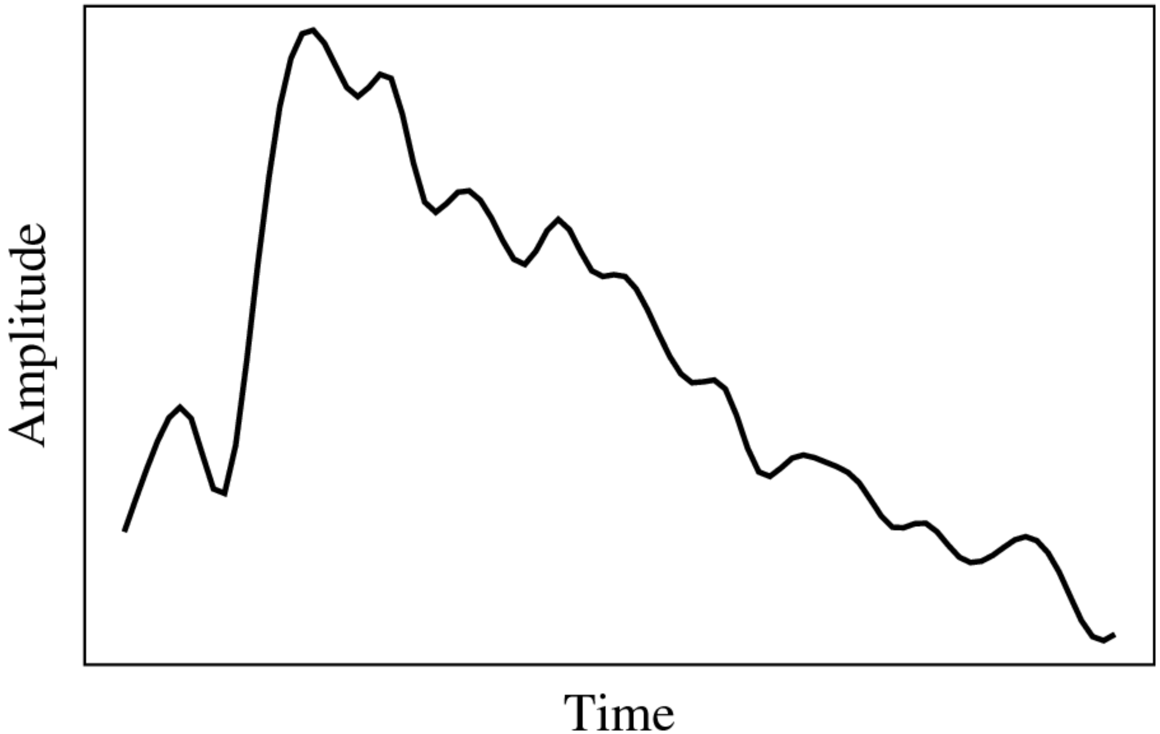
Digitization of Sound

- 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.
  - e.g., a speaker in an audio system vibrates back and forth and produces a *longitudinal* pressure wave that we perceive as sound.
  - Since sound is a pressure wave, it takes on continuous values, as opposed to digitized ones.
  - They 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).
  - To use a digital version of sound waves, we must form digitized their representations.

Digitization

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

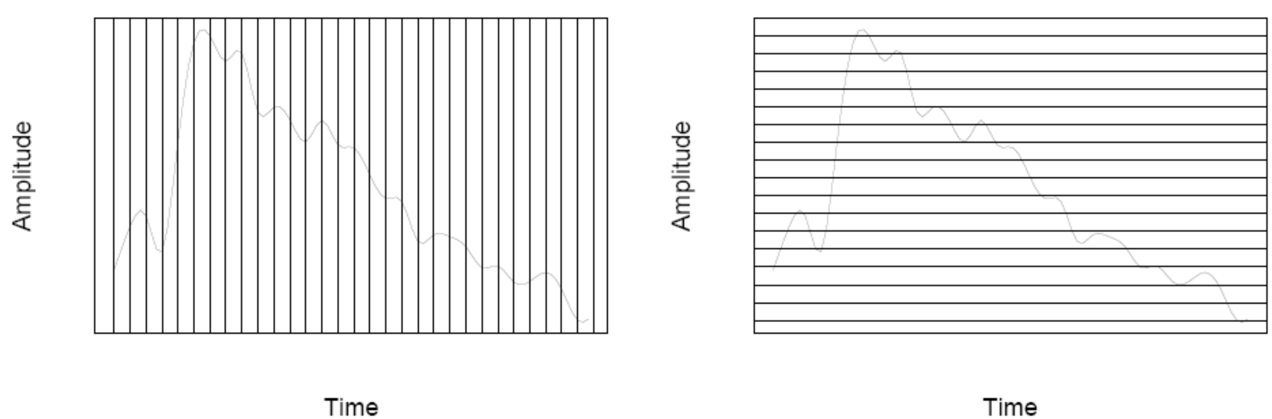
- The 1-dimensional nature of sound: **amplitude** values depend on a 1D variable, time.



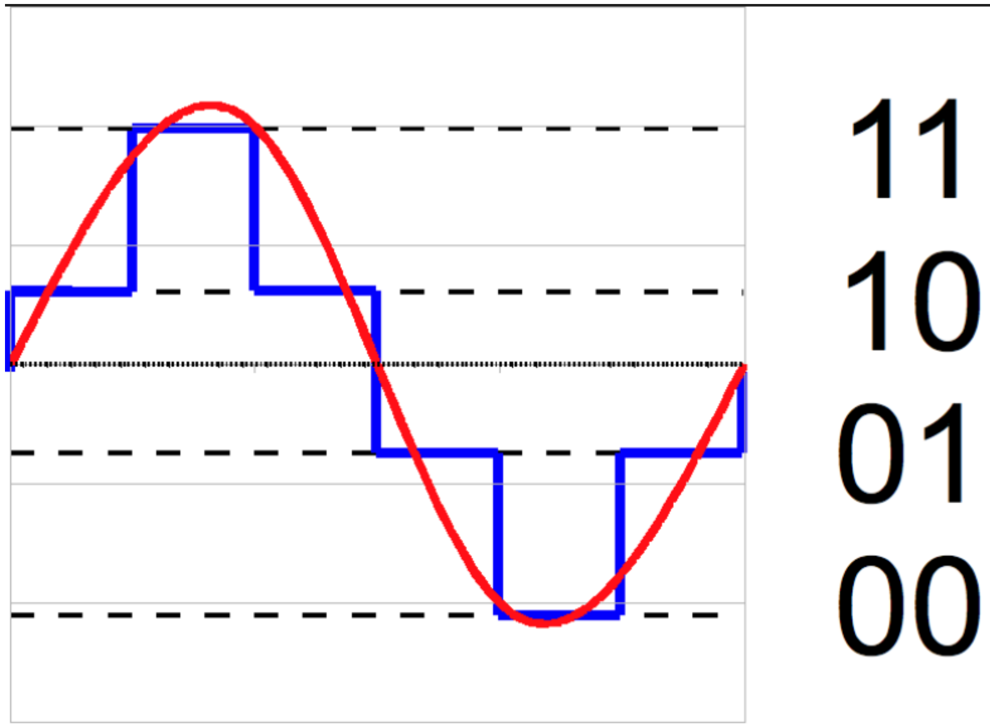
- To digitize, the signal must be **sampled** in each dimension: in time, and in amplitude.
  - Sampling means measuring the quantity (usually) at evenly-spaced intervals.

  - The sampling, using measurements only 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 8 ~ 48 kHz (48,000 samples per second)(determined by the Nyquist theorem).

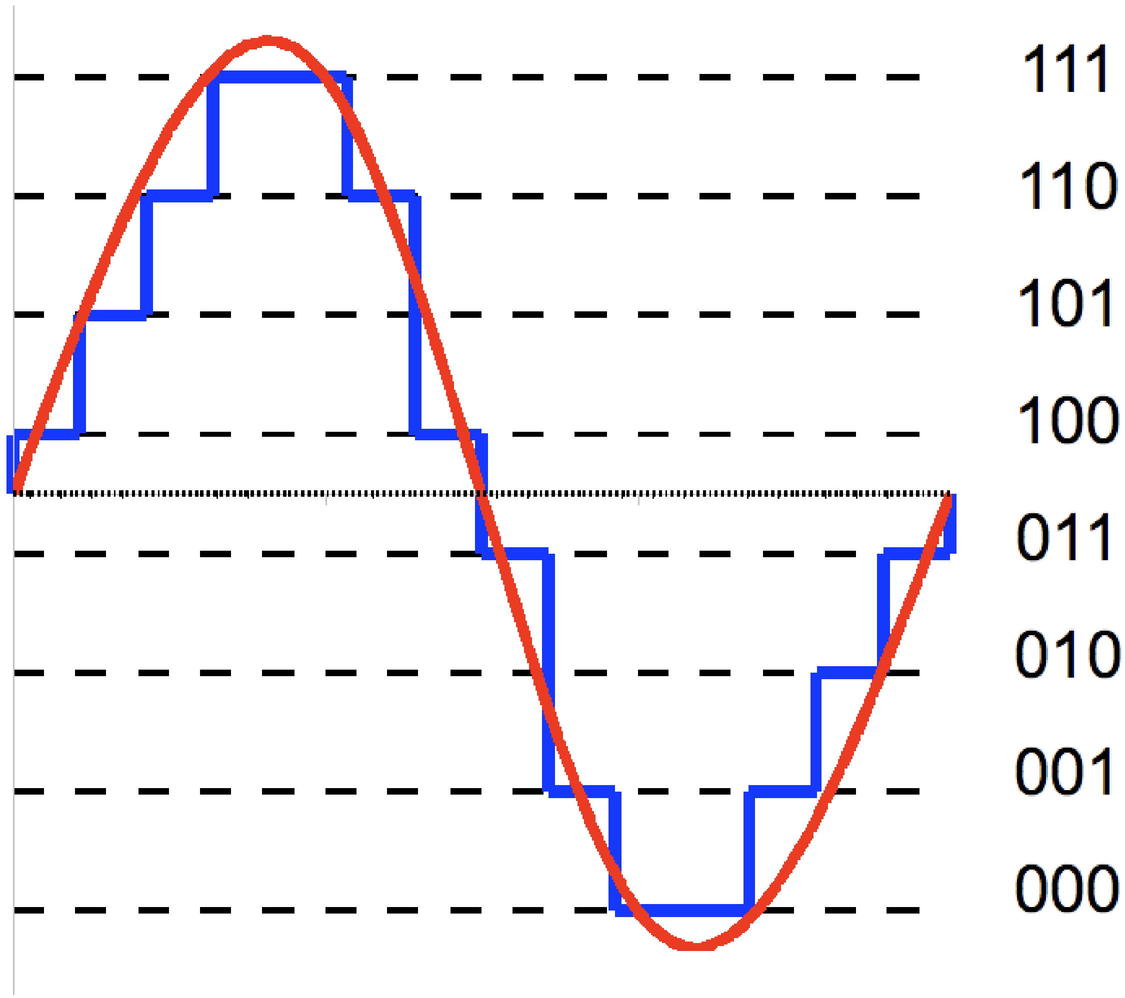
  - Sampling in the amplitude or voltage dimension is called **quantization**.



Quantization in a 2-bit ADC



Quantization in a 3-bit ADC



Step Size and Quantization Error

Suppose, N is the number of bits used for quantization in an ADC

$$\text{Step Size} = \Delta V = \frac{\text{Full Scale Analog Input}}{\text{Number of Steps}} = \frac{V_{max} - V_{min}}{2^N - 1}$$

$$\text{Quantization Error} = E_q = \frac{\text{Full Scale Analog Input}}{2 * \text{Number of Quantization Level}} = \frac{V_{max} - V_{min}}{2^{N+1}}$$

Step Size and Quantization Error

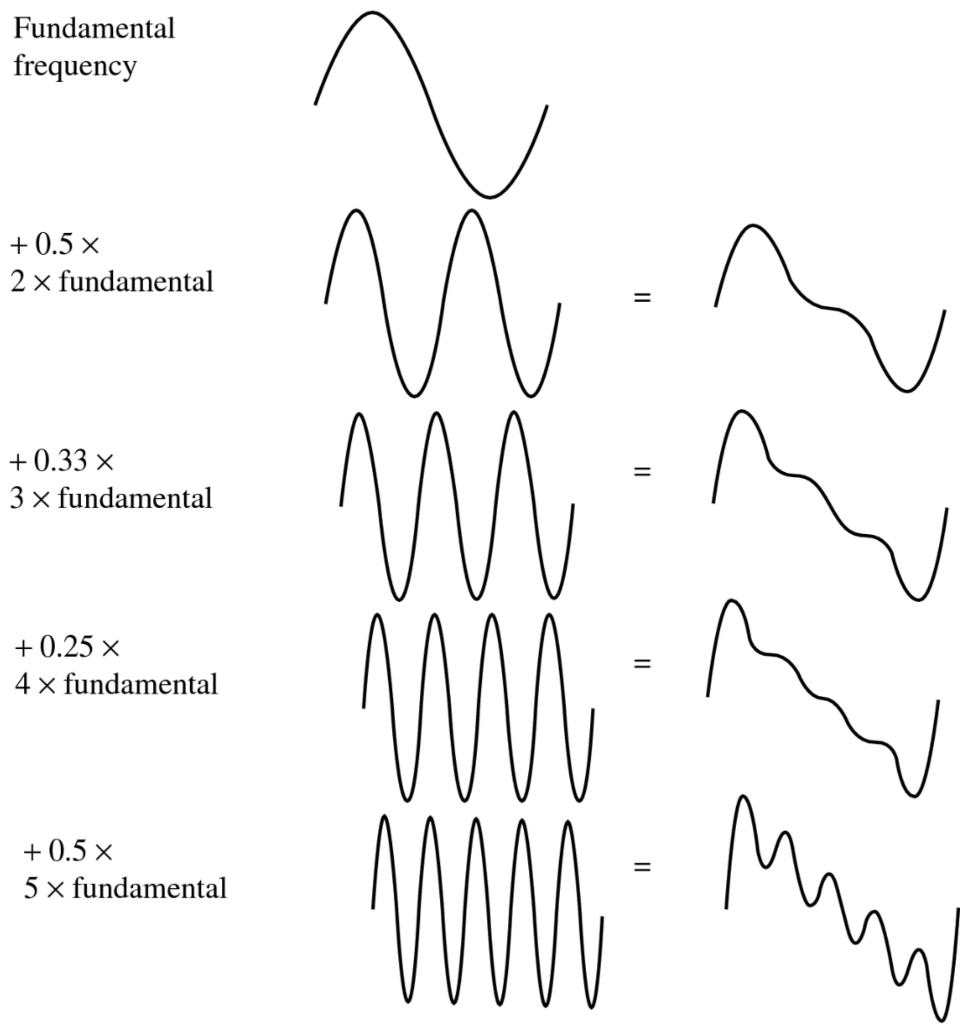
- Let's consider an analog signal which is uniformly distributed between -1 and +1 volt. We want to quantize the signal with a 3-bit ADC.
  - What is the step size?
  - What is the maximum quantitation error during sampling?

- Number of levels = 8 (which would map to [-1 -.75 -.5 -25 0 .25 .5 .75]).
- Step Size is 0.25 volt.
- The maximum quantization error will be 0.125.

- Thus to decide how to digitize audio data we need to answer:
  - What is the sampling rate?
  - How finely is the data to be quantized, and is quantization uniform?
  - How is audio data formatted? (file format)

Nyquist Theorem

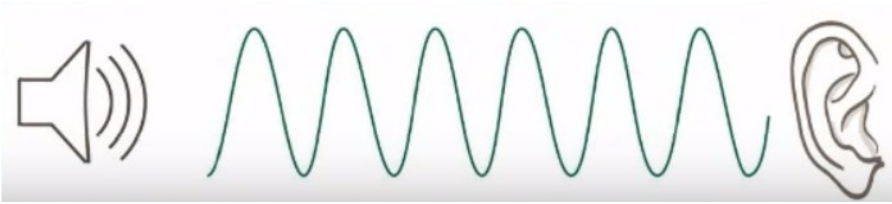
- Signals can be decomposed into a sum of sinusoids.
  - e.g., how weighted sinusoids can build up quite a complex signal.



Nyquist Theorem

- Whereas **frequency** is an absolute measure, **pitch** is generally relative — a perceptual subjective quality of sound.
  - Pitch and frequency are linked by setting the note A above middle C to exactly 440 Hz.
  - 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.
  - **Harmonics**: any series of musical tones whose frequencies are integral multiples of the frequency of a fundamental tone.
  - If we allow non-integer multiples of the base frequency, we allow non-'A' notes and have a more complex resulting sound.

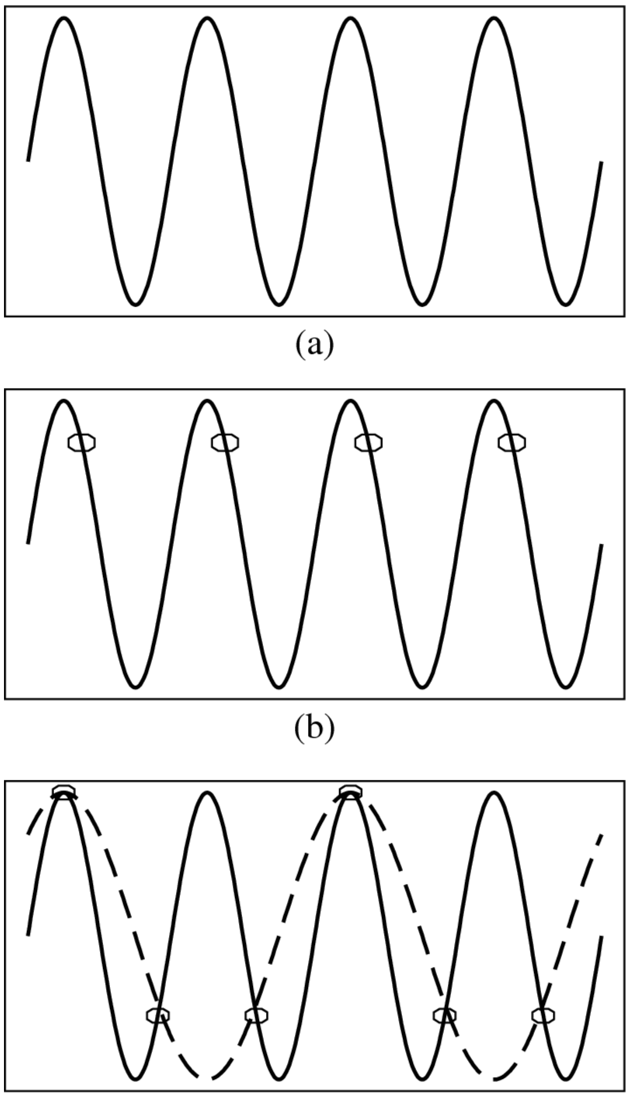
Sound frequency and Pitch



Sound Pitch

Nyquist Theorem

- For correct sampling we must use a sampling rate equal to at least twice the maximum frequency content in the signal -- the **Nyquist rate**.
  - A single sinusoid: a single, pure, frequency.
  - If sampling rate just equals the actual frequency, a false constant signal is detected.
  - If sample at 1.5 times the actual frequency, we obtain an incorrect (**alias**) frequency lower than the correct one — half the correct one (the wavelength, from peak to peak, is double that of the actual signal).



(a) (b) (c)

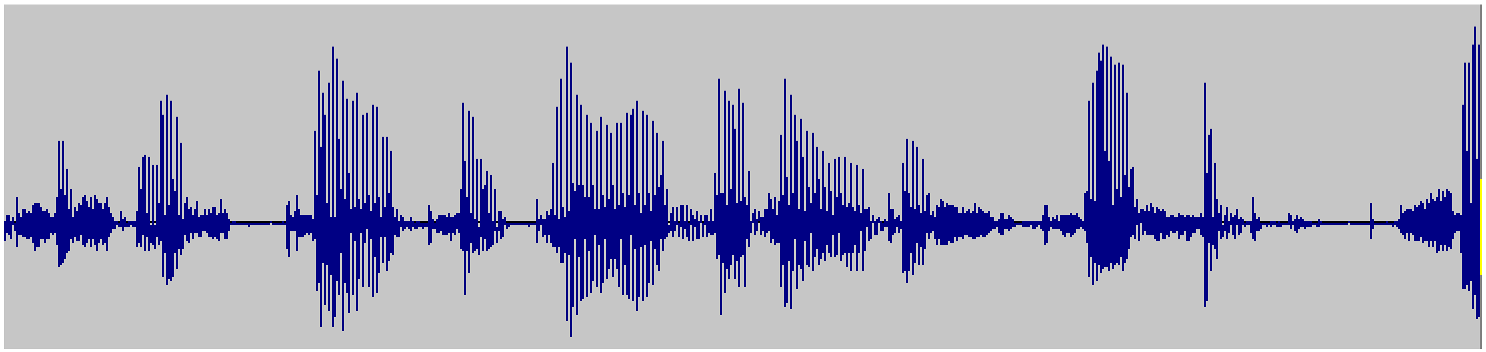
Nyquist Theorem

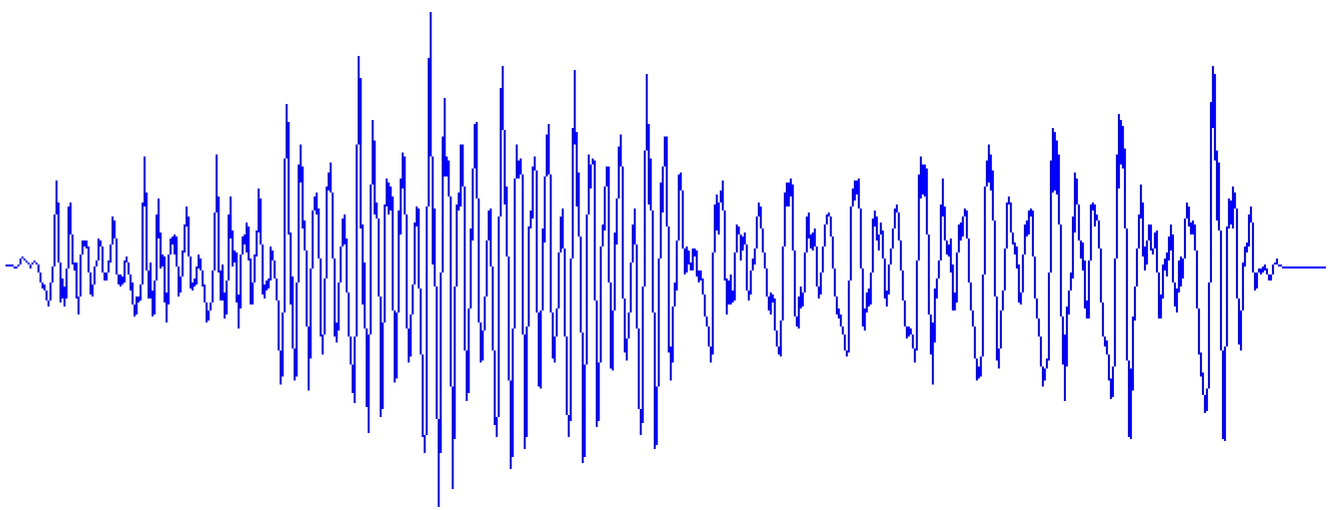
- **Nyquist Theorem**: If a signal is **band-limited**, i.e., there is a lower limit f_1 and an upper limit f_2 of frequency components in the signal, then the sampling rate should be at least $2(f_2 - f_1)$.

- **Nyquist frequency**: half of the Nyquist rate.
  - Since it would be impossible to recover frequencies > 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 <= Nyquist frequency.

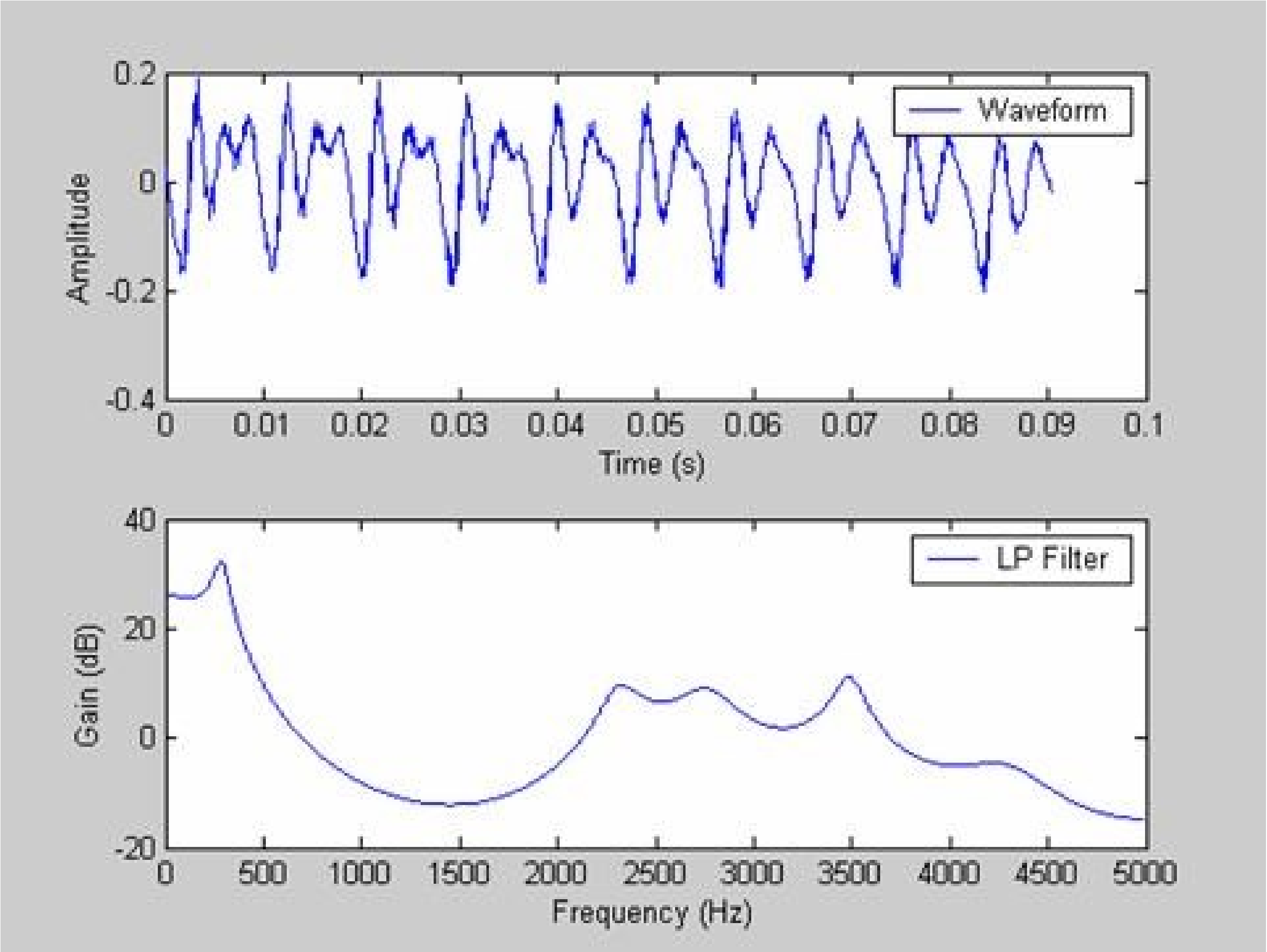
http://www.dsptutor.freeuk.com/aliasing/AliasingDemo.html

Speech Signal





Speech Signal Spectrum



Signal to Noise Ratio (SNR)

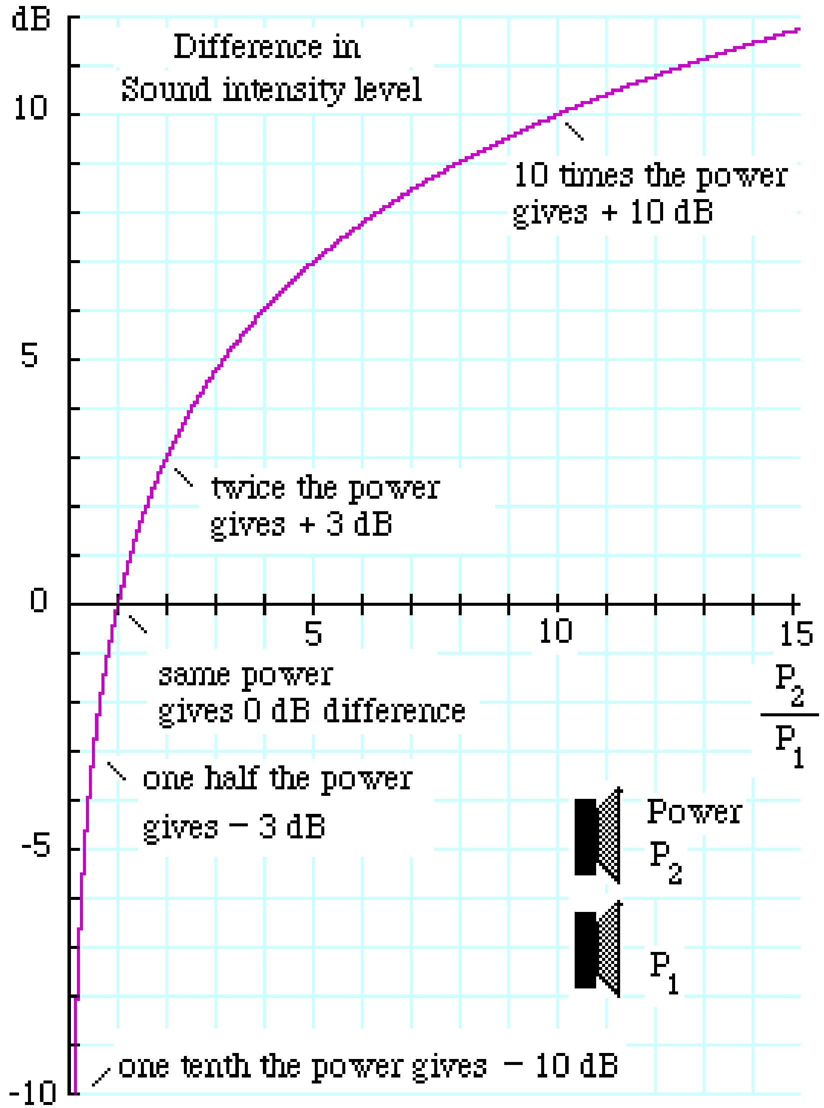
- The ratio of the power of the correct signal and the noise is called the *signal to noise ratio* (**SNR**) — a measure of the quality of the signal.
- The SNR is usually measured in decibels (**dB**), where 1 dB is a tenth of a **bel**. The SNR value, in units of dB, is defined in terms of base-10 logarithms of squared voltages, as follows:

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

- 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.
- 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.

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

Understanding dB Scale



- 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.
  - 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

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.
  - 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.
  - The quality of the quantization is characterized by the Signal to Quantization Noise Ratio (**SQNR**).
  - **Quantization noise**: the difference between the actual value of the analog signal, for the particular sampling time, and the nearest quantization interval value.
  - At most, this error can be as much as half of the interval.

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

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

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

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:
  - 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.
  - An audio music signal will typically contain from about 20Hz up to 20kHz.
  - 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.
  - So at the decoder side, a **lowpass** filter is used after the DA circuit, making use of the same cutoff as at the high-frequency end of the coder's band-pass filter.

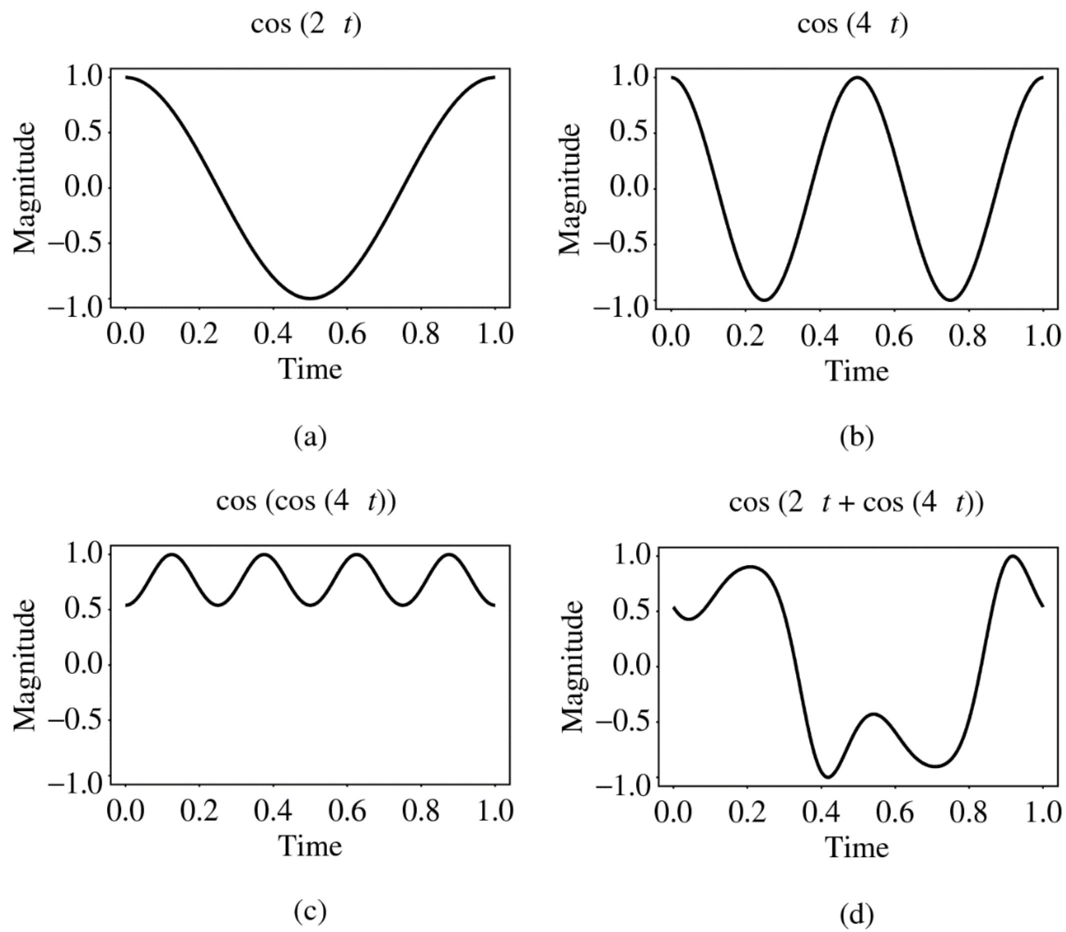
Audio Quality vs. Data Rate

- The uncompressed data rate increases as more bits are used for quantization. Stereo: double the bandwidth. to transmit a digital audio signal.

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

Synthetic Sounds

1.FM (Frequency Modulation): one approach to generating synthetic sound:



- Consider a more complex signal

$$x(t) = A(t)\cos[\omega_c \pi t + I(t)\cos(\omega_m \pi t + \phi_m) + \phi_c]$$

  - it uses a basic carrier frequency ω_c and a modulating frequency ω_m
  - the phase constants ϕ_m and ϕ_c create time-shifts for a more interesting sound
  - *envelope* A(t) specifies overall loudness over time and is used to fade in and fade out the sound (a guitar string has an attach, decay, sustain, release period)
  - I(t) is used to produce a feeling of harmonics by changing the amount of modulation frequency heard

2.Wave Table synthesis: A more accurate way of generating sounds from digital signals. Also known, simply, as **sampling**.

- In this technique, 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: Musical Instrument Digital Interface

- Use the sound card's defaults for sounds: ⇒ use a simple scripting language and hardware setup.

- **MIDI Overview**
  - 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.

  - a standard adopted by the electronic music industry for controlling devices, such as synthesizers and sound cards, that produce music.

  - supported by most synthesizers, so sounds created on one synthesizer can be played and manipulated on another synthesizer and sound reasonably close.

  - Computers must have a special MIDI interface, but this is incorporated into most sound cards. The sound card must also have both D/A and A/D converters.

Quantization and Transmission of Audio

- **Coding of Audio**: Quantization and transformation of data are collectively known as **coding** of the data.
  - The μ-law technique for companding audio signals is usually combined with an algorithm that exploits the temporal redundancy present in audio signals.
  - Differences in signals between the present and a past time can reduce the size of signal values and also concentrate the histogram of pixel values (differences, now) into a much smaller range.
  - The result of reducing the variance of values is that lossless compression methods produce a bitstream with shorter bit lengths for more likely values.
- In general, 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**.

Q&A