

CMPT 889: Lecture 3
Fundamentals of Digital Audio, Discrete-Time Signals

Tamara Smyth, tamaras@cs.sfu.ca
School of Computing Science,
Simon Fraser University

October 6, 2005

Sound

- Sound waves are longitudinal waves.
- A disturbance of a source (such as vibrating objects) creates a region of compression that propagates through a medium before reaching our ears.
- More generally, the source produces alternating regions of low and high pressure, called *rarefactions* and *compressions* respectively, which propagate through a medium away from the source from some location.

Sound Power and Intensity

- The two physical quantities for a sound wave described so far are the frequency and amplitude of pressure variations.
- Related to the sound pressure are
 1. the sound **power** emitted by the source
 2. the sound **intensity** measured some distance from the source.

- Intensity is given by

$$I = p^2 / (\rho c)$$

where ρ is the density of air and c is the speed of sound.

- The intensity is the power per unit area carried by the wave, measured in watts per square meter (W/m^2)
- Intensity therefore, is a measure of the power in a sound that actually contacts an area, such as the eardrum.
- The range of human hearing is $I_0 = 10^{-12} \text{ W}/\text{m}^2$ (threshold of audibility) and $1 \text{ W}/\text{m}^2$ (threshold of feeling).

Linear vs logarithmic scales.

- Human hearing is better measured on a logarithmic scale than a linear scale.
- On a linear scale, a change between two values is perceived on the basis of the **difference** between the values. Thus, for example, a change from 1 to 2 would be perceived as the same amount of increase as from 4 to 5.
- On a logarithmic scale, a change between two values is perceived on the basis of the **ratio** of the two values. That is, a change from 1 to 2 (ratio of 1:2) would be perceived as the same amount of increase as a change from 4 to 8 (also a ratio of 1:2).

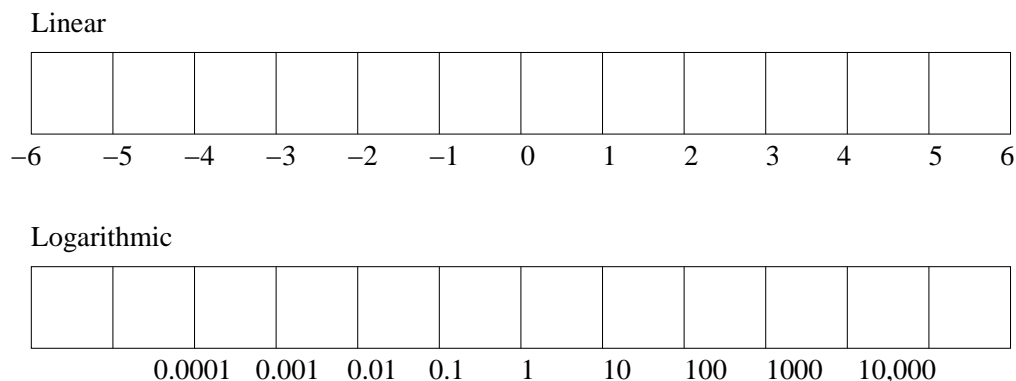


Figure 1: Moving one unit to the right increment by 1 on the linear scale and multiplies by a factor of 10 on the logarithmic scale.

Decibels

- The decibel is defined as one tenth of a bel which is named after Alexander Graham Bell, the inventor of the telephone.
- The decibel is a logarithmic scale, used to compare two quantities such as the power gain of an amplifier or the relative power of two sound sources.
- The decibel difference between two power levels ΔL for example, is defined in terms of their power ratio W_2/W_1 and is given in decibels by:

$$\Delta L = 10 \log W_2/W_1 \quad \text{dB.}$$

- Since power is proportional to intensity, the ratio of two signals with intensities I_1 and I_2 is similarly given in decibels by

$$\Delta L = 10 \log I_2/I_1 \quad \text{dB.}$$

Sound Power and Intensity Level

- Decibels are sometimes used as absolute measurements.
- Though this may seem contradictory (since the decibel is always used to compare two quantities), one of the quantities is really a fixed reference.
- The *sound power level* of a source, for instance, is expressed using the threshold of audibility $W_0 = 10^{-12}$ as a reference and is given in decibels by

$$L_W = 10 \log W/W_0.$$

- Similarly, the *sound intensity level* at some distance from the source can therefore be expressed in decibels by comparing it to a reference—usually $I_0 = 10^{-12} \text{ W/m}^2$:

$$L_I = 10 \log I/I_0.$$

Sound pressure Level

- Recall that intensity is proportional to sound pressure amplitude squared

$$I = p^2 / (\rho c)$$

- Though ρ and c are dependent on temperature, their product is often approximated by 400 to simplify calculation.
- The *sound pressure level* L_p (SPL) is equivalent to sound intensity level and is expressed in dB by:

$$\begin{aligned} L_p &= 10 \log I / I_0 \\ &= 10 \log p^2 / (\rho c I_0) \\ &= 10 \log p^2 / (4 \times 10^{-10}) \\ &= 10 \log (p / (2 \times 10^{-5}))^2 \\ &= 20 \log p / (2 \times 10^{-5}) \\ &= 20 \log p / p_0. \end{aligned}$$

where $p_0 = 2 \times 10^{-5}$ is the threshold of hearing for amplitude of pressure variations.

Continuous vs. Discrete signals

- A **signal**, of which a sinusoid is only one example, is a set, or sequence of numbers.
- A **continuous-time** signal is an infinite and uncountable set of numbers, as are the possible values each number can have. That is, between a start and end time, there are infinite possible values for time t and instantaneous amplitude, $x(t)$.
- When continuous signals are brought into a computer, they must be digitized or *discretized* (i.e., made *discrete*).
- In a **discrete-time signal**, the number of elements in the set, as well as the possible values of each element, is finite, countable, and can be represented with computer bits, and stored on a digital storage medium.

Analog to Digital Conversion

- A “real-world signal is captured using a microphone which has a diaphragm that is pushed back and forth according to the compression and rarefaction of the sounding pressure waveform.
- The microphone transforms this displacement into a time-varying voltage—an analog electrical signal.
- The process by which an analog signal is digitized is called *analog-to-digital* or “a-to-d” conversion and is done using a piece of hardware called an **analog-to-digital converter (ADC)**.
- In order to properly represent the electrical signal within the computer, the ADC must accomplish two tasks:
 1. Digitize the time variable, t , a process called **sampling**
 2. Digitize the instantaneous amplitude of the pressure variable, $x(t)$, a process called **quantization**

Sampling

- Sampling is the process of taking a sample value, individual values of a sequence, of the continuous waveform at regularly spaced time intervals.

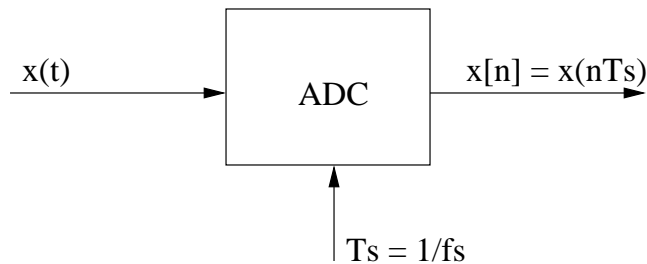


Figure 2: The ideal analog-to-digital converter.

- The time interval (in seconds) between samples is called the sampling period, T_s , and is inversely related to the *sampling rate*, f_s . That is,

$$T_s = 1/f_s \text{ seconds.}$$

- Common sampling rates:
 - Professional studio technology: $f_s = 48 \text{ kHz}$
 - Compact disk (CD) technology: $f_s = 44.1 \text{ kHz}$
 - Broadcasting applications: $f_s = 32 \text{ kHz}$

Sampled Sinusoids

- Sampling corresponds to transforming the continuous time variable t into a set of discrete times that are integer multiples of the sampling period T_s . That is, sampling involves the substitution

$$t \longrightarrow nT_s,$$

where n is an integer corresponding to the index in the sequence.

- Recall that a sinusoid is a function of time having the form

$$x(t) = A \sin(\omega t + \phi).$$

- In discretizing this equation therefore, we obtain

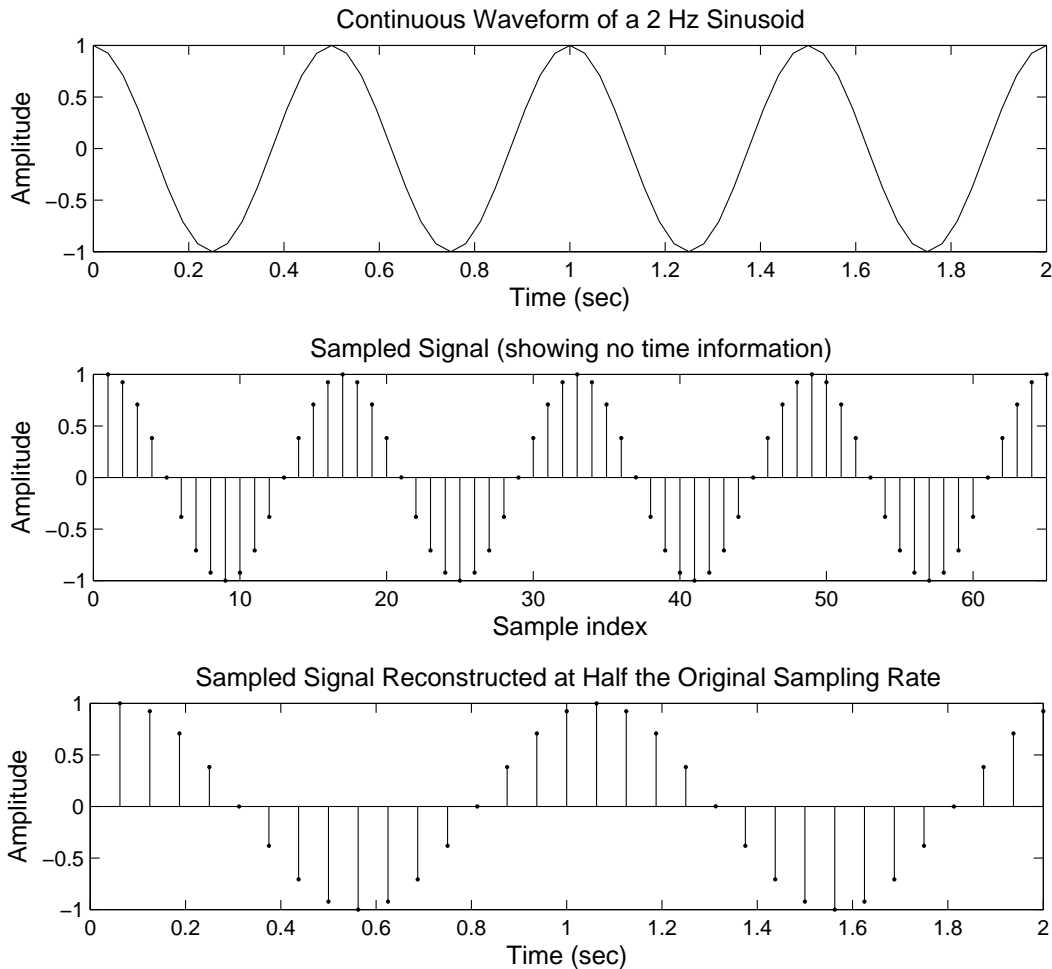
$$x(nT_s) = A \sin(\omega nT_s + \phi),$$

which is a sequence of numbers that may be indexed by the integer n .

- Note: $x(nT_s)$ is often shortened to $x(n)$ (and will likely be from now on), though in some literature you'll see square brackets $x[n]$ to differentiate from the continuous time signal.

Sampling and Reconstruction

- Once $x(t)$ is sampled to produce $x(n)$, the time scale information is lost and $x(n)$ may represent a number of possible waveforms.



- If a 2 Hz sinusoid is reconstructed at half the sampling rate at which it was sampled, it will have half the frequency but twice as long.

Sampling and Reconstruction

- If the sampled sequence is reconstructed using the **same sampling rate** with which it was digitized, the frequency and duration of the sinusoid will be preserved.
- If reconstruction is done using a **different sampling rate**, the time interval between samples will change, as will the time required to complete one cycle of the waveform. This has the effect of not only changing its frequency, but also changing its duration.

Nyquist Sampling Theorem

- What are the implications of sampling?
 - Is a sampled sequence only an approximation of the original?
 - Is it possible to *perfectly* reconstruct a sampled signal?
 - Will anything less than an infinite sampling rate introduce error?
 - How frequently must we sample in order to “faithfully” reproduce an analog waveform?

The **Nyquist Sampling Theorem** states that:

a bandlimited continuous-time signal can be sampled and perfectly reconstructed from its samples if the waveform is sampled over twice as fast as it's highest frequency component.

Nyquist Sampling Theorem

- In order for a bandlimited signal (one with a frequency spectrum that lies between 0 and f_{\max}) to be reconstructed fully, it must be sampled at a rate of $f_s > 2f_{\max}$, called the Nyquist frequency.
- Half the sampling rate, i.e. the highest frequency component which can be accurately represented, is referred to as the Nyquist limit.
- No information is lost if a signal is sampled above the Nyquist frequency, and no additional information is gained by sampling faster than this rate.
- Is compact disk quality audio, with a sampling rate of 44,100 Hz, then sufficient for our needs?

Aliasing

- To ensure that all frequencies entering into a digital system abide by the Nyquist Theorem, a low-pass filter is used to remove (or attenuate) frequencies above the Nyquist limit.



Figure 3: Low-pass filters in a digital audio system ensure that signals are bandlimited.

- Though low-pass filters are in place to prevent frequencies higher than half the sampling rate from being seen by the ADC, it is possible when processing a digital signal to create a signal containing these components.
- What happens to the frequency components that exceed the Nyquist limit?

Aliasing cont.

- If a signal is undersampled, it will be interpreted differently than what was intended. It will be interpreted as its **alias**.

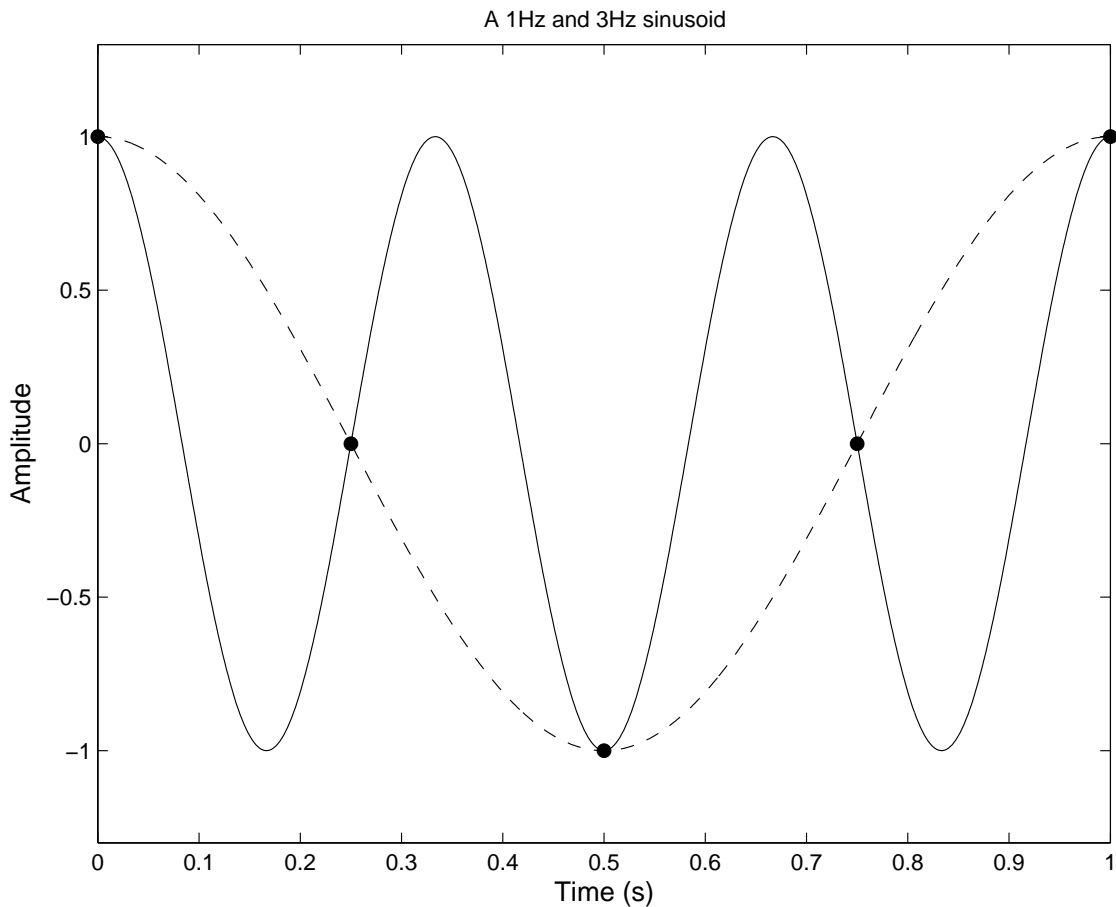


Figure 4: Undersampling a 3 Hz sinusoid causes its frequency to be interpreted as 1 Hz.

What is the Alias?

- The relationship between the signal frequency f_0 and the sampling rate f_s can be seen by first looking at the continuous time sinusoid

$$x(t) = A \cos(2\pi f_0 t + \phi).$$

- Sampling $x(t)$ yields

$$x(n) = x(nT_s) = A \cos(2\pi f_0 nT_s + \phi).$$

- A second sinusoid with the same amplitude and phase but with frequency $f_0 + lf_s$, where l is an integer, is given by

$$y(t) = A \cos(2\pi(f_0 + lf_s)t + \phi).$$

- Sampling this waveform yields

$$\begin{aligned} y(n) &= A \cos(2\pi(f_0 + lf_s)nT_s + \phi) \\ &= A \cos(2\pi f_0 nT_s + 2\pi lf_s nT_s + \phi) \\ &= A \cos(2\pi f_0 nT_s + 2\pi ln + \phi) \\ &= A \cos(2\pi f_0 nT_s + \phi) \\ &= x(n). \end{aligned}$$

What is an Alias? cont.

- There are an infinite number of sinusoids that will give the same sequence with respect to the sampling frequency (as seen in the previous example, since l is an integer (either positive or negative)).
- If we take another sinusoid $w(n)$ where the frequency is $-f_0 + lf_s$ (coming from the negative component of the cosine wave) we will obtain a similar result: it too is indistinguishable from $x(n)$.

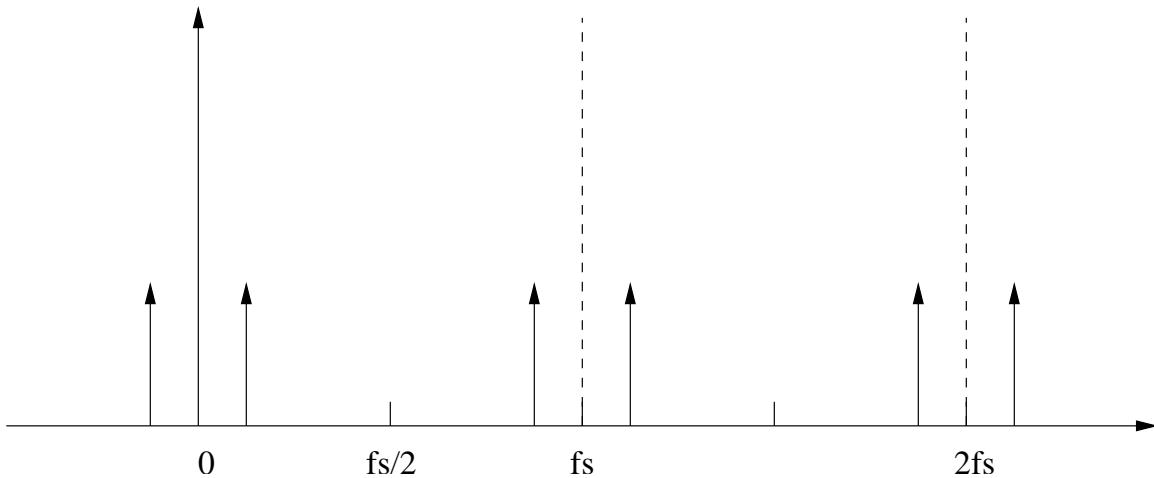


Figure 5: A sinusoid and its aliases.

- Any signal above the Nyquist limit will be interpreted as its alias lying within the permissible frequency range.

Folding Frequency

- Let f_{in} be the input signal and f_{out} be the signal at the output (after the lowpass filter). If f_{in} is less than the Nyquist limit, $f_{out} = f_{in}$. Otherwise, they are related by

$$f_{out} = f_s - f_{in}.$$

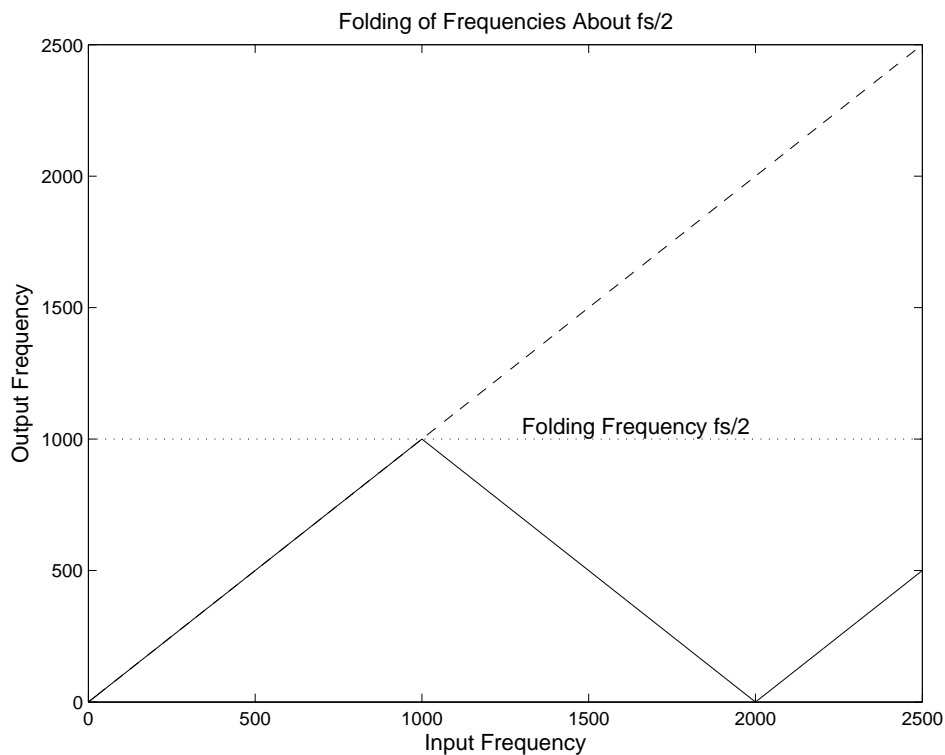


Figure 6: Folding of a sinusoid sampled at $f_s = 2000$ samples per second.

- The *folding* occurs because of the negative frequency components.

Quantization

- Where sampling is the process of taking a sample at regular time intervals, quantization is the process of assigning an amplitude value to that sample.
- Computers use bits to store such data and the higher the number of bits used to represent a value, the more precise the sampled amplitude will be.
- If amplitude values are represented using n bits, there will be 2^n possible values that can be represented.
- For CD quality audio, it is typical to use 16 bits to represent audio sample values. This means there are 65,536 possible values each audio sample can have.
- Quantization involves assigning one of a finite number of possible values (2^n) to the corresponding amplitude of the original signal. Since the original signal is continuous and can have infinite possible values, **quantization error** will be introduced in the approximation.

Quantization Error

- There are two related characteristics of a sound system that will be effected by how accurately we represent a sample value:
 1. **The dynamic range**, the ratio of the strongest to the weakest signal)
 2. **The signal-to-noise ratio (SNR)**, which compares the level of a given signal with the noise in the system.
- The dynamic range is limited
 1. at the lower end by the noise in the system
 2. at the higher end by the level at which the greatest signal can be presented without distortion.
- The SNR
 - equals the dynamic range when a signal of the greatest possible amplitude is present.
 - is smaller than the dynamic range when a softer sound is present.

Quantization Error cont.

- If a system has a dynamic range of 80 dB, the largest possible signal would be 80 dB above the noise level, yielding a SNR of 80dB. If a signal of 30 dB below maximum is present, it would exhibit a SNR of only 50 dB.
- The dynamic range therefore, predicts the maximum SNR possible under ideal conditions.

Quantization Error cont.

- If amplitude values are quantized by rounding to the nearest integer (called *the quantizing level*) using a *linear converter*, the error will be uniformly distributed between 0 and 1/2 (it will never be greater than 1/2).
- When the noise is a result of quantization error, we determine its audibility using the signal-to-quantization-noise-ratio (SQNR).
- The SQNR of a linear converter is typically determined by the ratio of the maximum amplitude (2^{n-1}) to maximum quantization noise (1/2).
- Since the ear responds to signal amplitude on a logarithmic rather than a linear scale, it is more useful to provide the SQNR in decibels (dB) given by

$$20 \log_{10} \left(\frac{2^{n-1}}{1/2} \right) = 20 \log_{10} (2^n).$$

- A 16-bit linear data converter has a dynamic range (and a SQNR with a maximum amplitude signal) of 96dB.
- A sound with an amplitude 40dB below maximum would have a SQNR of only 56 dB.

Quantization Error cont.

- Though 16-bits is usually considered acceptable for representing audio with good SNQR, its when we begin processing the sound that error compounds.
- Each time we perform an arithmetic operation, some rounding error occurs. Though the operation for one error may go unnoticed, the cumulative effect can definitely cause undesirable artifacts in the sound.
- For this reason, software such as Matlab will actually use 32 bits to represent a value (rather than 16).
- We should be aware of this, because when we write to audiofiles we have to remember to convert back to 16 bits.