Music 270a: Fundamentals of Digital Audio and Discrete-Time Signals

Tamara Smyth, trsmyth@ucsd.edu
Department of Music,
University of California, San Diego

September 30, 2019

Continuous vs. Discrete signals

- A **signal**, of which a sinusoid is only one example, is a sequence of numbers.
- A **continuous-time** signal is an *infinite* and *uncountable* set of numbers, as are the possible values each number can have.
  - 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 (continuous-time) electrical signal.
- The process by which an analog signal is digitized is called *analog-to-digital* or “a-to-d” conversion
  - conversion is done using 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. **sampling**: digitize the time variable t;
  2. **quantization**: digitize the instantaneous amplitude of the pressure variable $x(t)$.

Sampling

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

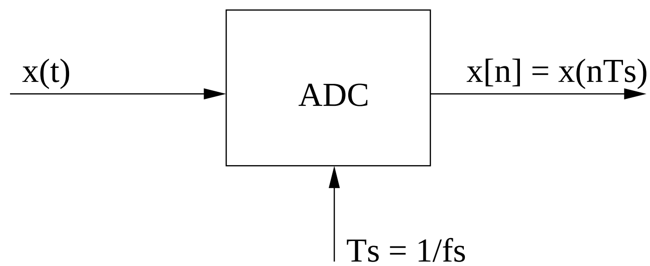


Figure 1: The ideal analog-to-digital converter.

- 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 technolgy: $f_s = 48$ kHz
  - Compact disk (CD) technology: $f_s = 44.1$ kHz
  - Broadcasting applications: $f_s = 32$ 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

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

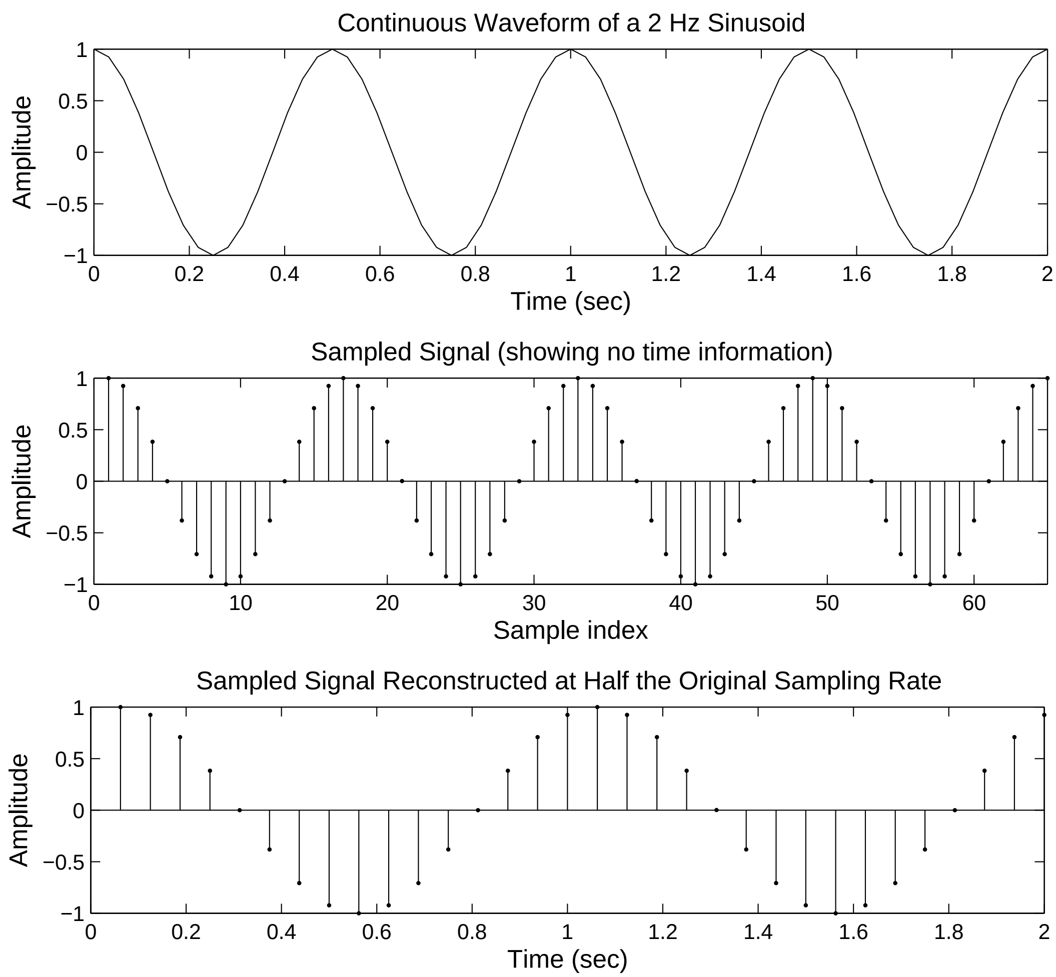
  which is a sequence of numbers that may be indexed by the integer n.

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

Sampling and Reconstruction

- Once $x(t)$ is sampled to produce $x(n)$ (a finite set of numbers), the time scale information is lost and $x(n)$ may represent a number of possible waveforms.
- To preserve the **frequency** and **duration** of the sinusoid, the sampled sequence must be reconstructed using the *same sampling rate* with which it was digitized.
- If reconstruction is done using a **different sampling rate**,
  - the time interval between samples will change (changing duration);
  - the time required to complete one cycle of the waveform will change (changing frequency);

Sampling and Reconstruction



- If a 2 Hz sinusoid is reconstructed at half the sampling rate at which is was sampled, it will have a frequency of 1 Hz, and will be twice as long.

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 $f_s > 2f_{\text{max}}$, 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 2: 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.**

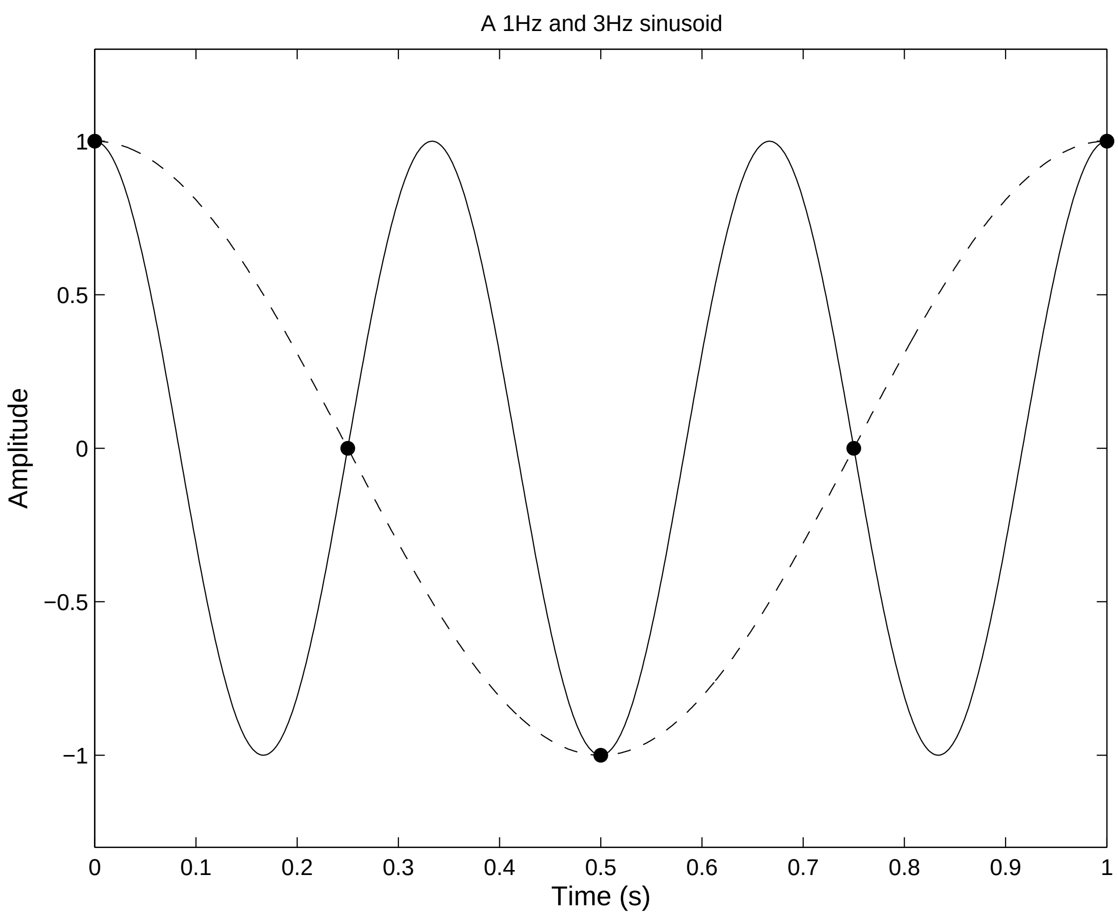


Figure 3: Undersampling a 3 Hz sinusoid causes it's frequency to be interpreted as 1 Hz.

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

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

- 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

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

What is an Alias? cont.

- Infinite discrete sinusoids will give the same sequence with respect to the sampling frequency.

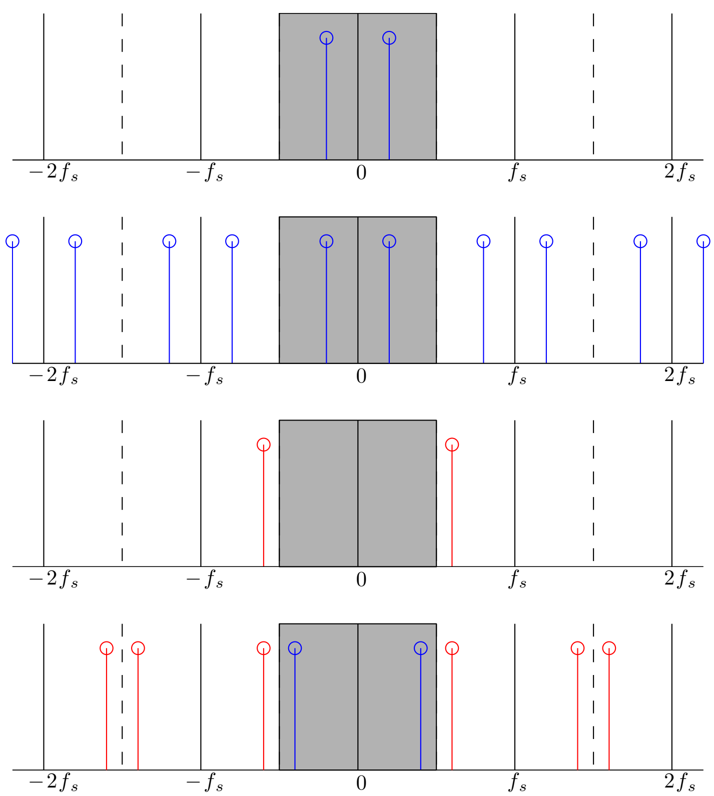


Figure 4: The shaded area represents the sounding bandwidth.

- A signal exceeding $f_s/2$ will have a negative frequency component with an alias falling within the sounding bandwidth (the shaded area).

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,
$$f_{\text{out}} = f_{\text{in}}.$$
  - If f_{in} is greater than Nyquist but less than the sampling rate,
$$f_{\text{out}} = f_s - f_{\text{in}}.$$

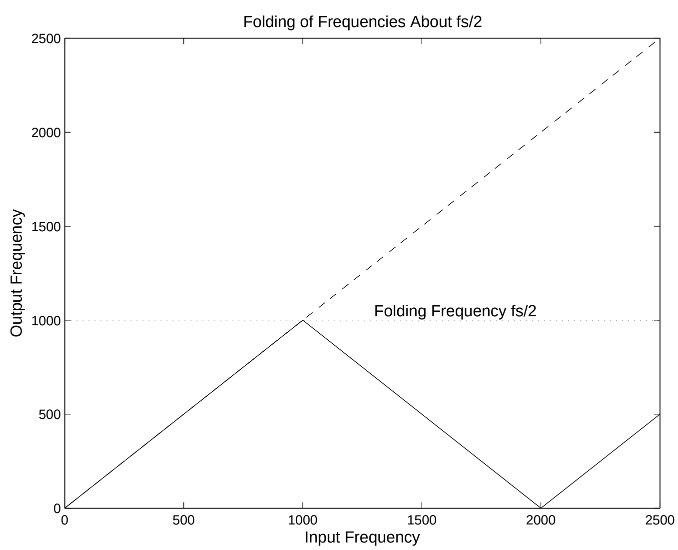


Figure 5: Folding of a sinusoid sampled at $f_s = 2000$ samples per second.

- 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 a finite number of possible values to the amplitude of the signal at that sample.
- If amplitude values are represented using n bits, there will be 2^n possible values that can be represented.
- If $n = 16$ (CD quality),
  - each sample can have $2^{16} = 65,536$ possible values;
  - the highest possible amplitude is $2^{15} = 32,768$, since *audio signals are both positive and negative*.
- Since the original signal is continuous and can have infinite possible values, **quantization error** will be introduced in the approximation.
- There are two related characteristics of a sound system that will be effected by how accurately we represent a sample value:

Dynamic Range and SNR

- Two sound system characteristics effected by how accuratley 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)**, the ratio 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.
- If a system has a dynamic range of 80 dB, a signal of 30 dB below maximum would yield a SNR of 50 dB.
- The dynamic range predicts the maximum SNR under ideal conditions.

Quantization Error (Linear Converter)

- When noise is a result of quantization error, audibility is determined using the **signal-to-quantization-noise-ratio (SQNR)**.
- If amplitude values are quantized by rounding to the nearest integer (*the quantizing level*) using a *linear converter*, the error will be uniformly distributed between 0 and 0.5.
- The SQNR of a linear converter is typically determined by the ratio of
  1. the maximum amplitude (2^{n-1}) to
  2. the maximum quantization noise (0.5 or 2^{-1})

  and is expressed in decibels (dB) as

$$20\log_{10}\left(\frac{2^{n-1}}{2^{-1}}\right) = 20\log_{10}(2^n) \text{ dB} \quad (1)$$

$$= 96 \text{ dB, if } n = 16. \quad (2)$$

- A sound with an amplitude 40dB below maximum would have a SQNR of only 56 dB.
- Matlab uses 64 bits (double-precision floating point)—convert to 16-bit when using `audiowrite`.