Big Idea (Week 2):
Signals, Symbols, and Infinities
Sampling, Quantization, and Noise

A “sound” is a field of acoustic (local atmospheric pressure) waveforms (structured variations in space and time). A computer is a symbol processor with a finite memory and limited computational power (respecting both its model of arithmetic and the speed with which it performs that arithmetic). Digital audio engineering starts with the problem of acquiring and storing in the computer’s memory a symbolic representation of those continuous waveforms amenable to processing (minimally: internal copying and reconstruction via external speakers) using its model of arithmetic.

The nearly universal solution to this first problem found in contemporary digital audio technology is recourse to “pulse code modulation” (PCM), the process of converting acoustic sound pressure waves into analog voltage signals using analog electro-mechanical components, and then sampling those voltage signals in time and value using analog-to-digital electronics. The question now arises as to what level of fidelity to the original sound can be expected from this electromechanically transformed and double-digitally (both in time and value) sampled signal. That is, how much noise does a particular sampling rate or quantization interval introduces compared to the original signal? More formally, we introduce the notion of a quantization function,

$$\text{Quantize}_L(x) = \text{Round}(L \cdot x)/L,$$

where Round is the function that rounds real numbers to the nearest integer.

Using this formalism, we can define the PCM function as a composition of quantization steps as follows. A sampler quantizes time to convert a continuous signal $s(t)$ to one varying discretely between constant intervals of duration $T$, $s_T(t)$ defined as

$$s_T(t) = \text{Sample}_T[s(t)] = s[\text{Quantize}_{1/T}(t)] = s[T \cdot \text{Round}(t/T)].$$

The PCM function follows the time sampled signal with a quantization in value to some level $L$, yielding $s_{(L,T)}(t)$, defined as

$$s_{(L,T)}(t) = \text{PCM}_{(L,T)}[s(t)] = \text{Quantize}_L \left( \text{Sample}_T[s(t)] \right) = \text{Quantize}_L \left( s[\text{Quantize}_{1/T}(t)] \right).$$

The noise introduced in sampling and quantization can then be formally defined as:

$$n(t) = s(t) - \text{PCM}_{(L,T)}[s(t)].$$

Since there are an uncountably infinite number of $s(t)$’s, but only a countable number of $s_{d}(t)$’s, this double digital sampling process always discards information, introducing non-zero noise. However, if we have a model for the structure that exists within $s(t)$, or, at least, the structure that we care about in $s(t)$, and this structure is appropriately restricted, we can represent $s(t)$ perfectly—or, at least, so that the noise,
$n(t)$, is irrelevant to our application. For sound, we will be exploiting a model of the human auditory system that allows us to differentiate the noise characteristics that humans can and cannot hear.

An empirical answer to this question can be found in the last decade’s digital audio technology (e.g., compact discs and digital audio tapes). Only very, very rare human beings can discern any difference (perceive any noise) between the sound recorded and played back on the best analog audio tape player and the same sound (i.e., the transformed voltage signal) sampled in time every $1/44 \cdot 10^{-3} = 2.27 \cdot 10^{-5}$ seconds and sampled in value with $2^{16} = 65536$ different voltage levels at each time step.