

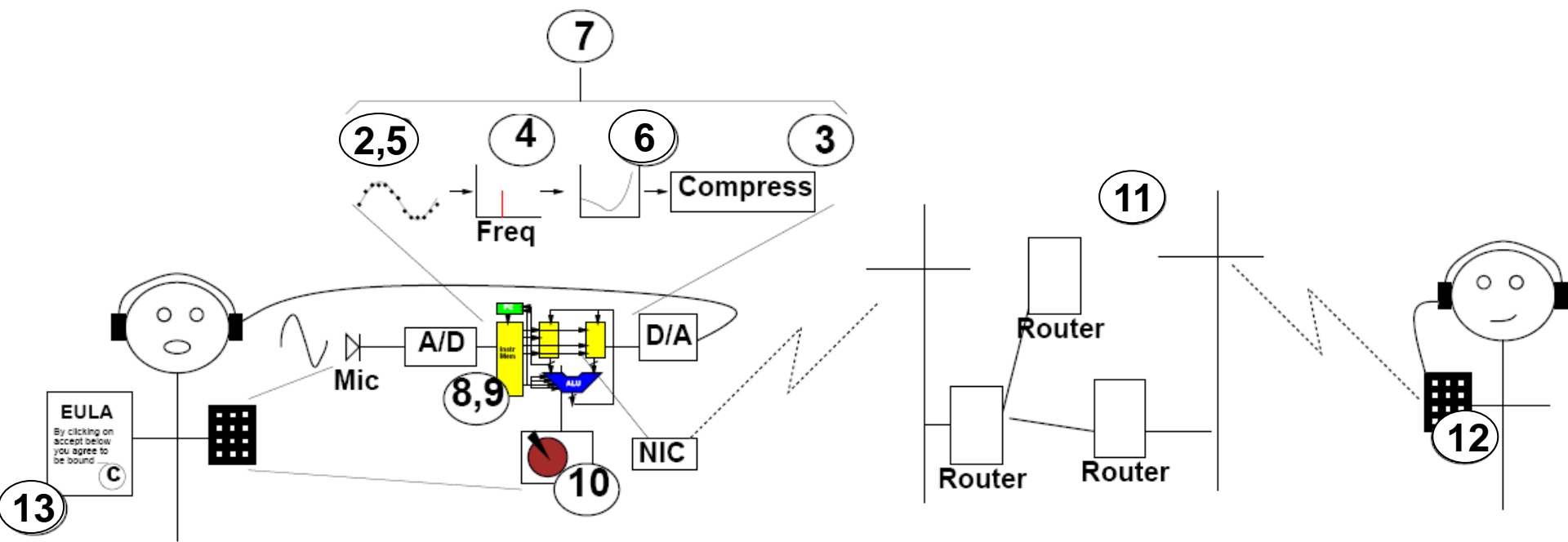
ESE250: Digital Audio Basics

Week 2

January 19, 2012

Sampling

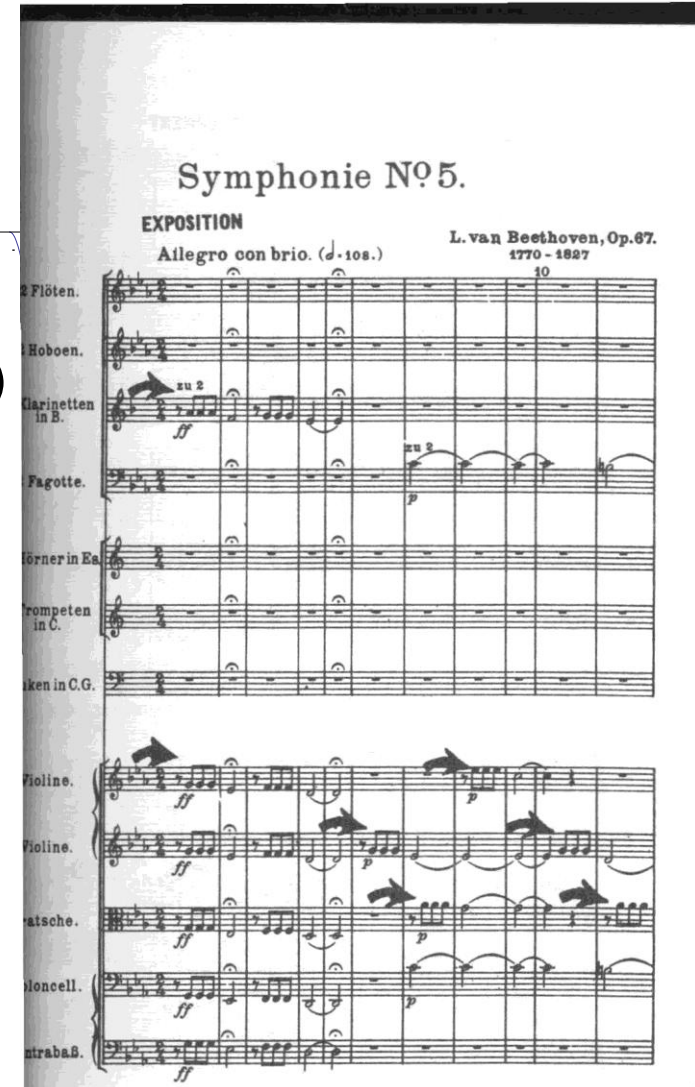
Course Map



Numbers correspond to course weeks

Sound: Physics → Psychoacoustics

- Physical Sound
 - a field of acoustic (local atmospheric pressure)
 - waveforms (structured variations in pressure and time)
- Receivers of interest
 - Microphone
 - Piezo-electric materials
 - Electronic circuits
 - Ear (Psycho-physics)
 - Subdiscipline within psychology
 - Mapping physical stimulus onto measurable human perceptual response
 - Brain (cognitive science/AI)



[Harcourt, Brace & Company, NY, 1935]

Recording: From Copying to Sampling

- Analog Recording: copying

- from tinfoil
- to vinyl
- to metal filings
- ... to history ...



[source: [wikipedia](#)]



[source: [wikipedia](#)]



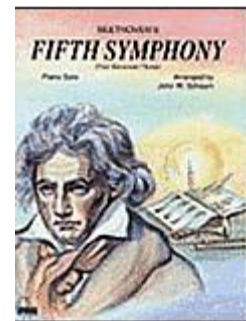
[source: [wikipedia](#)]

- Digital Recording: sampling

- In time (abscissa)
- In value (ordinate)

- Noise: errors due to signal processing

- errors are inevitable either way
- we'll focus on digital recording noise in this class



[source: [sheetmusicplus](#)]

Digital Audio Agenda

- Problem Statement
 - Given a limited set of resources
 - memory
 - computational power (instruction set, clock speed)
 - And a performance specification
 - What sort of errors
 - Are how damaging
 - Devise an audio signal recording and reconstruction architecture
 - that maximizes performance
 - while remaining true to the resource constraints
- Agenda
 - Weeks 2 – 7: explore the implications of finite (countable) memory
 - Weeks 8 - 11: exploit the capabilities of computational engine

Technology Story: from CD to MP3

- CD (late 20th Century)
 - ~ 600 MB capacity
 - ~ 1 hour of music
 - “Transparent” sound quality
 - Indistinguishable from best analog recording
 - To “almost all” humans

- iPod shuffle (early 21st Century)
 - 2000 MB capacity
 - ~ 50 hours of music
 - Advertised: 500 songs
 - Conversion factor: ~ 6 min/song
 - “Transparent” sound quality



[source: [wikipedia](#)]



[source: [apple](#)]

MP3 performance advantage (song-hr/MB):

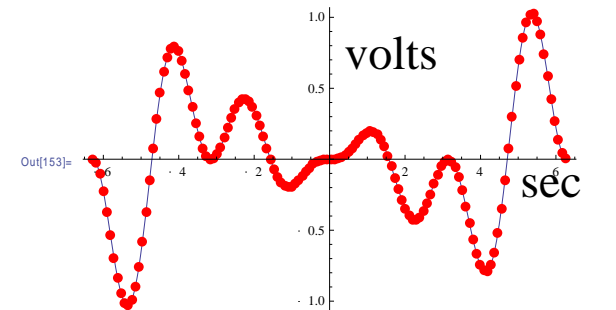
MP3: 50 song-hr/ 2000 MB = 1/40 song-hr/MB
 CD: 1 song-hr/ 600 MB = 1/600 song-hr/MB

MP3 song-hr/MB = **15** × CD song-hr/MB

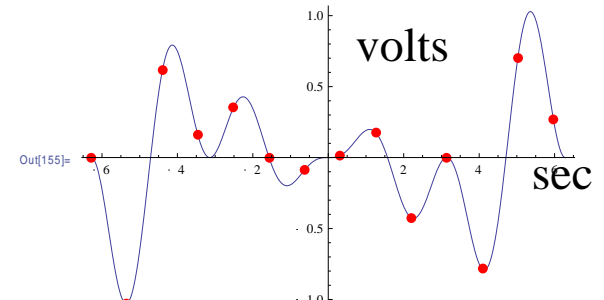
What Changed?

- Hardware advantage
 - Better hardware storage density? (no – why not?)
 - Hint: how is an MP3 player different from a thumb drive?
- Question(s) for this semester:
 - How does computation yield a net storage advantage?
 - What other advantages does computation confer ?
- Question for this week's lecture:
 - (Baseline of technology story)
 - How is sound represented and stored on a CD?
 - Overview Answer:
 - CD recorders sample waveform in time every $1/44000$ sec
 - CD recorders sample waveform in value at 65536 distinct levels
- Follow-on Question: can we do better?

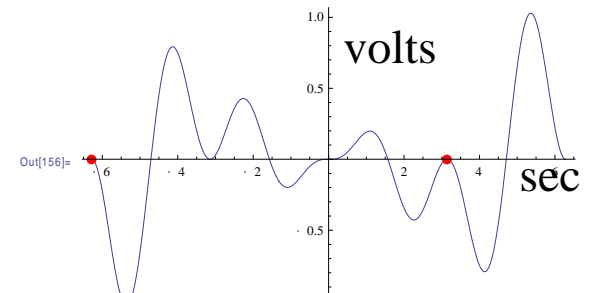
Quantizing in Time



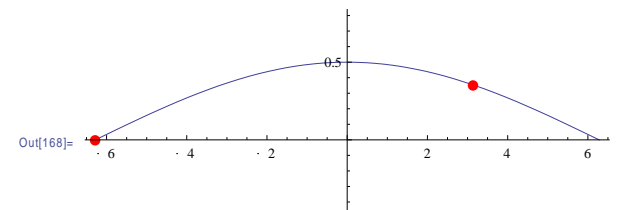
Sampling interval: $T_s = 0.03 \pi \text{ sec}$



Sampling interval: $T_s = 0.3 \pi \text{ sec}$



Sampling interval: $T_s = 3 \pi \text{ sec}$



Sampling interval: $T_s = 3 \pi \text{ sec}$

- Intuition

we ought to get a pretty good impression of a waveform's sound by dense sampling in time

- (some) Questions

- What do we mean by “impression”?
- What do we mean by “dense”?
- Doesn't the answer depend upon the particular waveform?
- What else does it depend upon?

- (some) Answers: week 6

Quantizing in Value

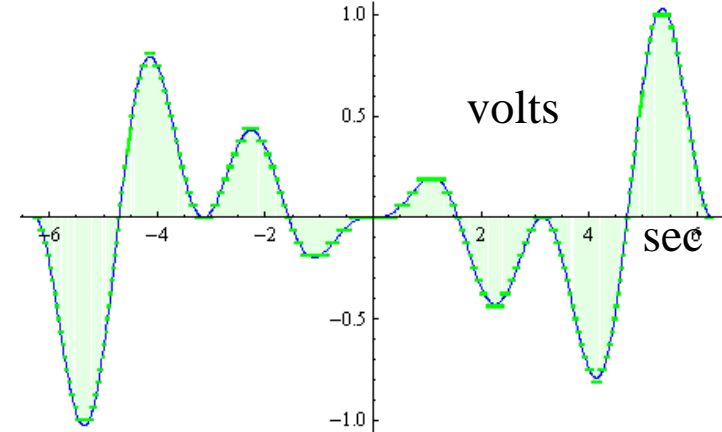
- Intuition

we ought to get a pretty good impression of a waveform's sound by dense sampling of recorded voltage

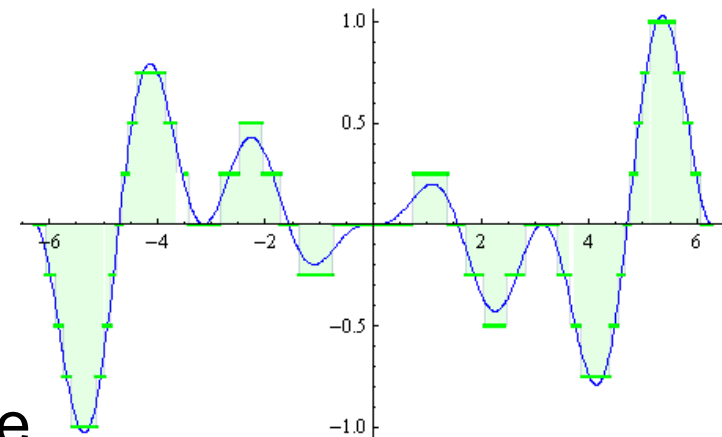
- (some) Questions

- What do we mean by “impression”?
- What do we mean by “dense”?
- Doesn't the answer depend upon the particular waveform?
- What else does it depend upon?

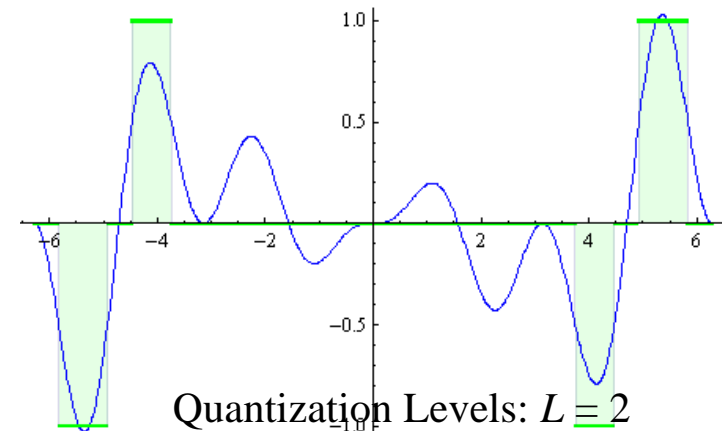
- (some) Answers: now + week 5



Quantization Levels: $L = 2^5$



Quantization Levels: $L = 2^3$



Quantization Levels: $L = 2$

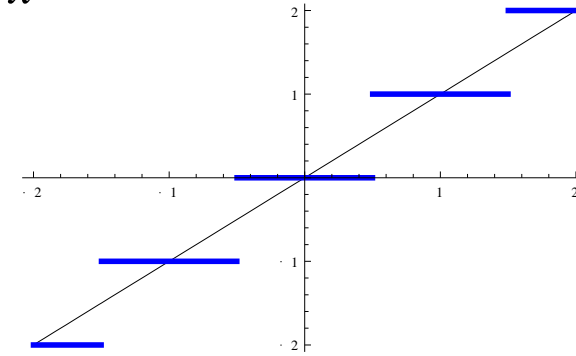
Signal Processing: Questions

- Information Theory
 - Teaches us how to ask certain kinds of questions, e.g.
 - (Q1) Did our signal processing lose us any information?
 - (Q2) If so how much?
 - And sometimes gives us answers to some of them
 - At the “cost” of requiring a more formal view
- We can ask again, a bit more formally:
 - given a signal processing algorithm, $s_f(t) = f[s(t)]$,
 - (Q1) Is f invertible?
 - (i.e., does there exist an algorithm, f^{-1} , such that $s = f^{-1}(p)$)?
 - (Q2) If not, then how big is the noise, $n_f(s) = s - s_f$?

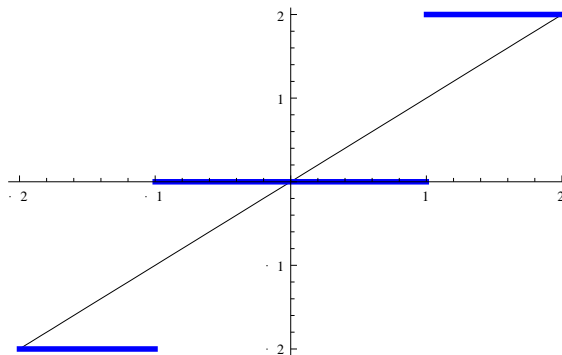
Quantization

- Rounding up
 - $z = \text{Round}(x)$
 - takes the “closest” integer, z , to the real number, x
- Class Exercise: plot the function Round
- A Uniform Quantizer is a function
 - $\text{Quantize}_L(x) = \text{Round}(L \cdot x) / L$
 - From the set of real numbers
 - To the set of integer multiples of $1/L$.
- L the “quantization level”
- Class Exercise: plot the function Quantize_L for $L \in \{1/2, 2, 4\}$

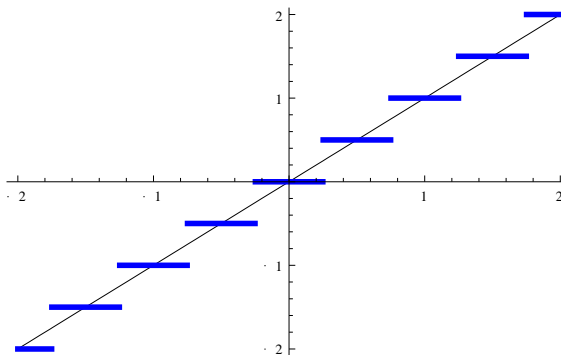
Round(x)



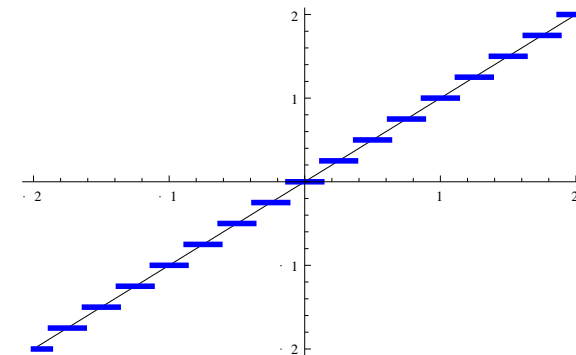
Quantize_{1/2}(x)



Quantize₂(x)



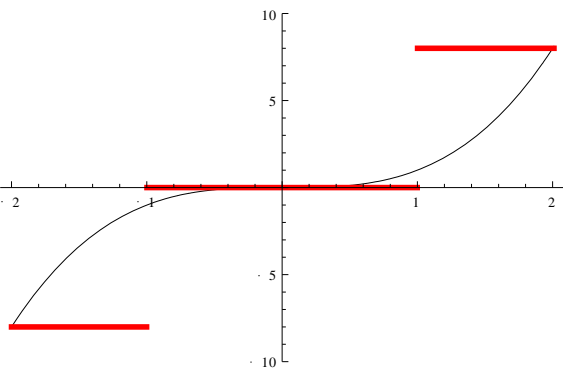
Quantize₄(x)



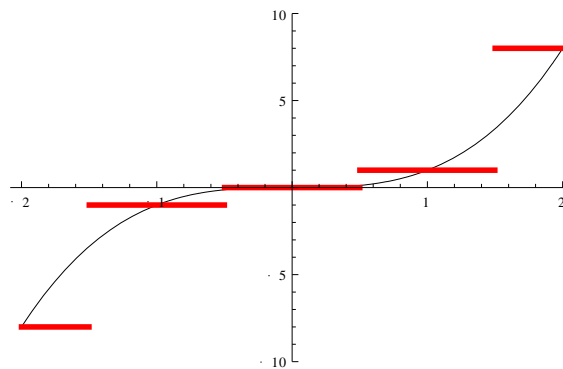
Sampling

- Quantization in time: a “sampler” is an input quantizer
 - $s_T(t) = \text{Sample}_T[s(t)] = s[\text{Quantize}_{1/T}(t)] = s[\text{Round}(t/T) \cdot T]$
 - T the “sampling interval”; $1/T$ the “sampling frequency”
 - Sample_T is a function from the set of
 - real-valued signals-varying-in-realtime
 - to the set of real-valued signals jumping at discreet, uniform time intervals
- Class Exercise: for $s(t) = t^3$ plot $\text{Sample}_T[s(t)]$, $T \in \{1/2, 1, 2\}$

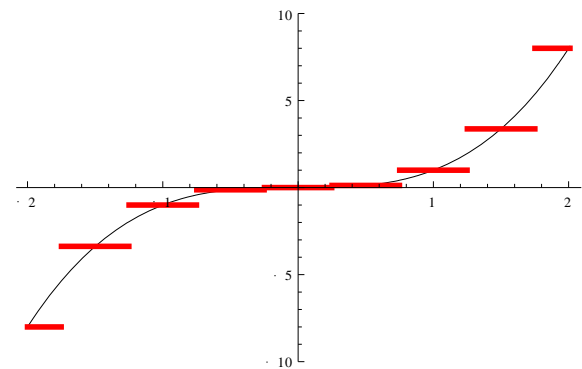
$\text{Sample}_2(t^3)$



$\text{Sample}_1(t^3)$



$\text{Sample}_{1/2}(t^3)$



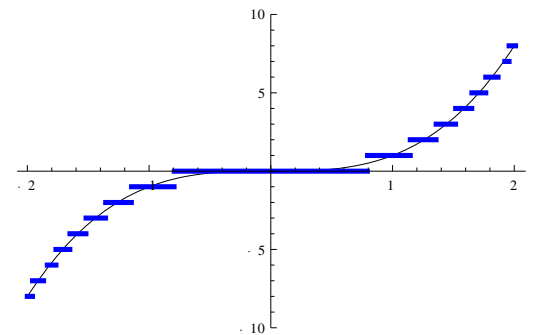
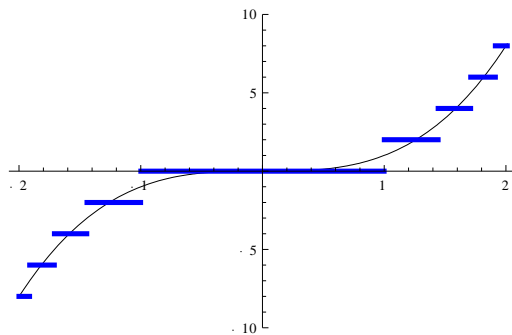
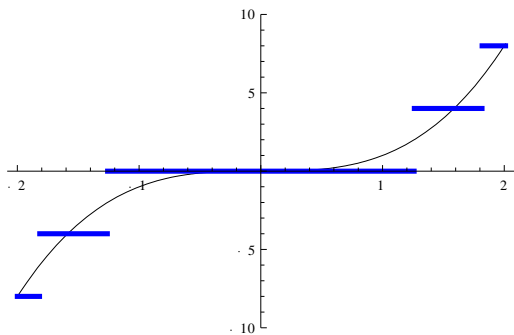
Uniform Coding

- Quantization in value: a “uniform coder” is an output quantizer
 - $s_L(t) = \text{Quantize}_L[s(t)] = \text{Round}(L \cdot s(t)) / L$
 - This defines a function From the set of
 - real-valued signals-varying-in-realtime
 - to the set of integer-valued signals varying-in-realtime
- Class Exercise:** for $s(t) = t^3$ plot $\text{Quantize}_L[s(t)]$, $L \in \{1/4, 1/2, 1\}$

$\text{Quantize}_{1/4}(t^3)$

$\text{Quantize}_{1/2}(t^3)$

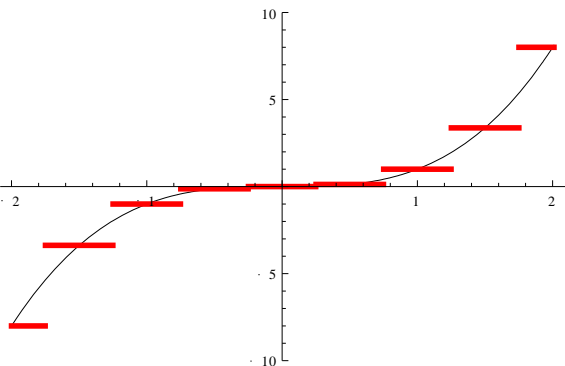
$\text{Quantize}_1(t^3)$



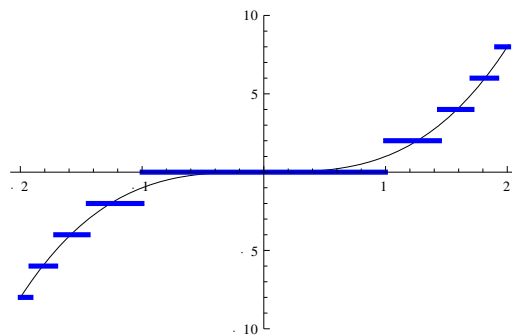
Pulse Code Modulation

- PCM: quantization in time and value (input & output)
 - $s_{(L,T)}(t) = \text{PCM}_{(L,T)} [s(t)] = \text{Quantize}_L(s[\text{Quantize}_{1/T}(t)])$
 $= \text{Quantize}_L(\text{Sample}_{1/T}[s(t)])$
 - From the set of
 - real-valued signals-varying-in-realtime
 - to the set of integer-valued functions-varying-in-integers
- Class Exercise: for $s(t) = t^3$ plot $\text{PCM}_{(1/2,1/2)} [s(t)]$

Sample_{1/2}(t³)

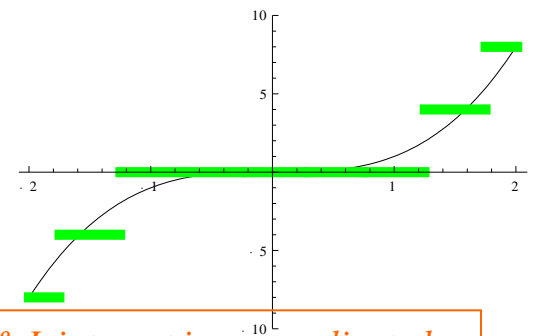


Quantize_{1/2}(t³)



PCM_(1/2,1/2)[s(t)] =

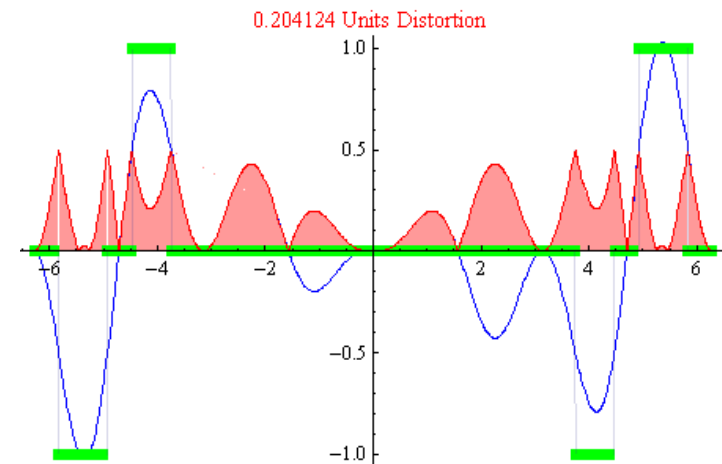
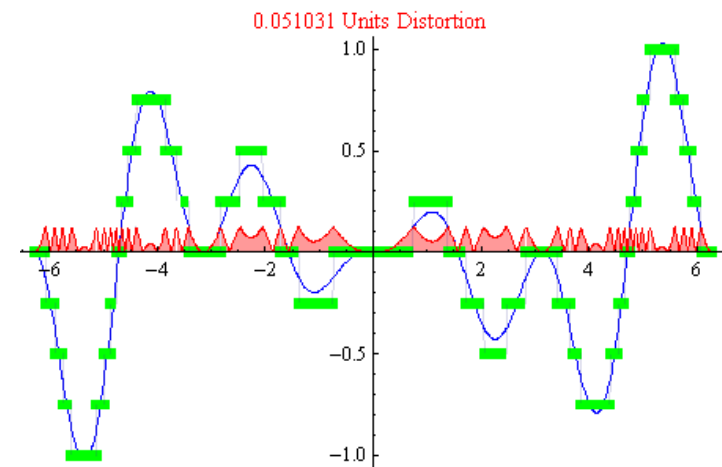
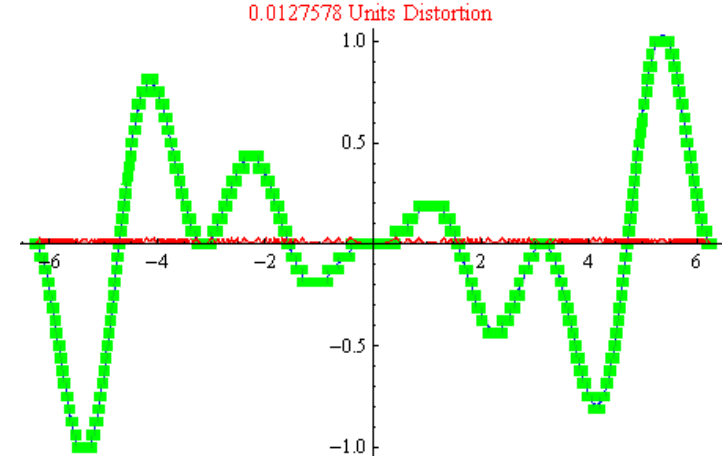
Quantize_{1/2}[Sample_{1/2}(t³)]



Moral: T & L interact in a complicated manner – we need a theory!!

Quantization Noise

- Traditionally quantization noise is called “distortion”
 - $s_L(t) = \text{Quantize}_L[s(t)]$
 $= \text{Round}[L \cdot s(t)] / L$
 - $n_q(t) = s(t) - s_L(t)$
- Distortion varies with the number of quantization levels
 - typically increases
 - as number of levels decreases



Interlude: Quantum Audio?

- The Universe is “really” digital: use Quantum Physics!
 - Planck [1900]: $E = \hbar \nu$
 - the energy, E , of any mass
 - is an **integer multiple**
 - of its atoms’ (properly scaled via \hbar) oscillating frequency, ν
 - Millikan [1923]: $\hbar \approx 6.26 \cdot 10^{-24}$ Js
- [[Makowski](#):] Fundamental quanta in MKS units
 - 10^{-35} m; 10^{-8} kg; 10^{-44} s;
 - Volt: $V = J/C$
 - $\approx J \cdot 10^{19} / e = \text{kg} \cdot \text{m}^2 \cdot \text{s}^{-2} \cdot 10^{19} / e$
 - $= 10^{-8} \cdot 10^{-70} \cdot 10^{88} \cdot 10^{19} / e$
 - $= 10^{29} / e$
- Conclusion: **Quantum Audio Engineering** requires sampling at
 - Sampling interval: $T_s = 10^{-44}$ sec
 - Quantization Levels: $L = 10^{29} \approx 2^{100}$

Naïve Information Theory

- Intuition:
 - clearly, s_L is lossy for finite L (class: show why)
 - but “dense enough” quantization eventually becomes essentially lossless
- Naïve Claim
 - Informally
 - distortion should vanish
 - as the number of levels gets very large
 - Formally
 - for any signal, $s(t)$,
 - $\lim_{L \rightarrow \infty} n_q(t) = s(t) - s_L(t) = 0$
- Naïve Implication:
 - can we, eventually, think of Quantize_L as an invertible function
 - and play back our recording of $s_L(t) = \text{Quantize}_L[s(t)]$ via $\text{Quantize}_L^{-1} [s_q(t)]$?
- Actual Situation: the essential signal/symbol divide
 - $L \in \mathbb{N}$ is “only” a counting number
 - there are far too many real signals ($s : \mathbb{R} \rightarrow \mathbb{R}$)
 - to be captured by the far more limited number of quantized signals, $s_L(t)$
- It turns out that there are even far more real numbers than integers!

Computational Representation

- Big Question: what is a “number” and how should a computer represent it?
- Numbers coming from the physical world seem to be “analog”
 - 19th Century Mathematicians formalized this idea
 - Finally defining the set of “real numbers,” \mathbb{R}
 - As an axiomatically precise version of “continuous” (geometric) extent
 - Resolving the ancient confusion about “irrational” numbers
 - the “length” of the diagonal of a unit cube
 - cannot be expressed as the ratio of any two counting numbers
- Numbers represented by a computer must be “symbolic”
 - Boolean algebra (generic hardware-level model of computation)
 - takes its universe in the basic symbols
 - $\mathbb{B} = \{\perp, \top\} \approx \{\emptyset, \{\emptyset\}\} = \emptyset \cup \{\emptyset\} \cup \{\emptyset \cup \{\emptyset\}\}$
 - Set theory (and principle of “induction” or “...”)
 - gives the set of “natural” (or “counting” numbers)
 - $\mathbb{N} = \{0, 1, 2, 3, \dots\} \approx \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset, \{\emptyset, \{\emptyset\}\}\}, \dots\}$
 $= \mathbb{B} \cup \{\mathbb{B}\} \cup \{\mathbb{B} \cup \{\mathbb{B}\}\} \cup \dots$
 - More or less straightforward set theory gives the
 - Integers: $\mathbb{Z} \approx \mathbb{N} \cup \neg \mathbb{N}$
 - Rationals: $\mathbb{Q} \approx \mathbb{Z} \times \mathbb{Z} = \{(m,n) \mid m \in \mathbb{N} \ \& \ n \in \mathbb{N}\}$
 - But there is an (historically + conceptually) important “gulf” to cross before \mathbb{R}

Countable and Uncountable Sets

- Imagine a “perfect” sampling function
 - implemented on a “symbolic” computer with “infinite” memory
 - that samples losslessly, e.g., $s_{\infty}(t) = \text{Quantize}_{\infty}[s(t)] = s(t)$
- This implies, among other consequences, that
 - at each time, t ,
 - for each signal, s , there is a unique level, L , such that
 - $s(t) = L$
- Can we “count” the infinity of real numbers?
 - pick a particular time, say $t=0$,
 - and enumerate all the functions $s(t)$
 - by their “order” at time 0, given by the unique $L = s(0)$ for each s .
- Georg Cantor’s demonstration [[1891](#)] that \mathbb{R} is uncountable:
 - after you have ordered your list of all the real numbers, $\{s_1, s_2, s_3, \dots\}$
 - Cantor will come along and find you at least one, say r , still uncounted:

Toward a Model-Driven Theory

- Why did we bother to try to represent *all* signals?
 - No physical source could produce some mathematically conceivable signals
 - No animal organ could transduce some physically plausible signals
 - No human listener could hear some perceptually active sounds
- Let's assume-away the irrelevant
 - The more we constrain the class of signals
 - The more efficiently we will be able to process them
- CD technology approach:
 - Pulse Code Modulation (PCM)
 - Sample in time at 44 kHz or every $T_s = 22.7 \cdot 10^{-6}$ sec
 - Sample in voltage at $L = 2^{16} = 65536$ distinct levels
 - Implicit set of assumptions about audio-relevant signals

Models: “generative” information

- A “model”
 - is a set of (computationally expressed) assumptions
 - about the sender and/or the receiver
 - of any signal to be processed
- By thus delimiting the class of signals we achieve
 - (in theory) the possibility of exact reconstruction
 - (in practice) the ability to predict
 - how quickly the processing noise will diminish
 - as our computational resources are increased
 - for the “cost” of
 - keeping around (and then executing in use) some “side information”
 - (a mathematical/computational representation of the model)
 - that is systematically used in reconstructing the signal from its stored record
- Examples:
 - Audio processing: we will model the receiver (the human auditory system)
 - Speech processing: we add a model of the sender (the properties of language)

ESE250: Digital Audio Basics

End Week 2 Lecture

Sampling