University of Pennsylvania
Department of Electrical and System Engineering
Digital Audio Basics

ESE150, Spring 2019 — Midterm — Wednesday, March 13

- Exam ends at 5:50PM; begin as instructed (target 4:35PM)
- Do not open exam until instructed to begin exam.
- Problems weighted as shown.
- Calculators allowed.
- Closed book = No text or notes allowed.
- Provided reference materials on next to last page.
- Show work for partial credit consideration.
- Unless otherwise noted, answers to two significant figures are sufficient.
- Sign Code of Academic Integrity statement (see last page for code).

____

I certify that I have complied with the University of Pennsylvania's Code of Academic Integrity in completing this exam.

Name: Solution

1					2	3			4			
a	b	c	d	e		a	b	c	a	b	c	d
3	3	3	3	3	15	5	3	2	2.5	2.5	2.5	2.5

5				6	7	8				**Total**
a	b	c	d			a	b	c		
2	3	3	2	15	10	3	3	9		100

Average: 81

1. Cat hearing ranges from 45Hz to 64KHz. Consider performing sound capture for cats.

   (a) What is the minimum sample rate you should use to capture the full range of frequencies the cat can hear?

   128KHz (Nyquist)

   (b) What is the minimum period between samples at this rate?

   $\frac{1}{128{,}000\,\mathrm{Hz}} = 7.8\mu\mathrm{s}$

   (c) How many Bytes will it take to encode a 3-minute, single-channel (mono) "cat" song with 16b PCM samples?

   $128,000 \times \frac{16}{8} \times 3 \times 60 = 46\mathrm{MB}$

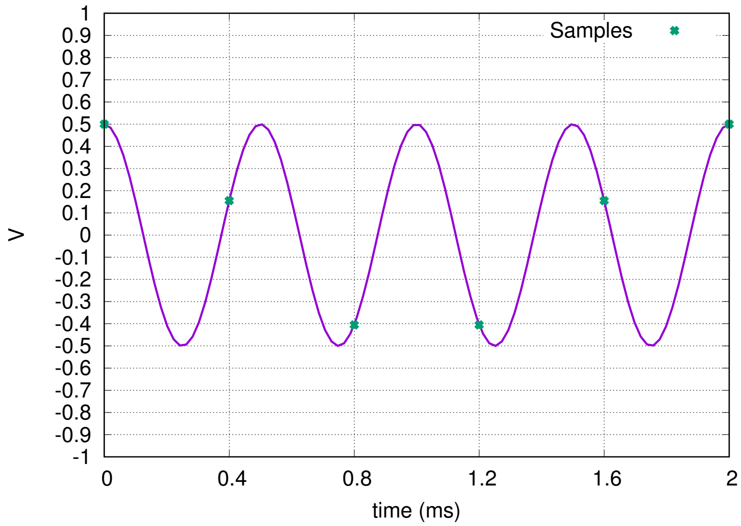
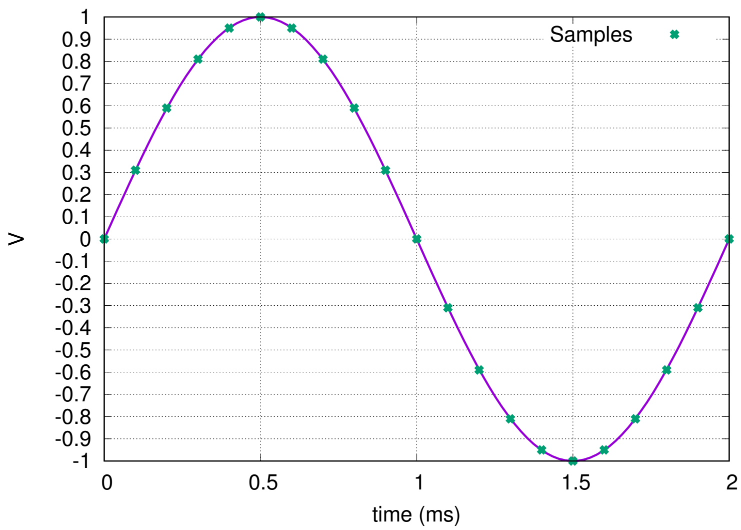
   (d) Given that human hearing goes to 22KHz, how much more PCM data is required for a "cat" song than a human song? (State as a ratio.)

   $\frac{128{,}\mathrm{KHz}}{44{,}\mathrm{KHz}} = \frac{64}{22} = 2.9\times$

   (e) Staying with PCM audio, how could you compress the cat song to guarantee to achieve the same data rate (bits/second) as human PCM audio? Is this lossy or lossless compression?

   Quantize the amplitude to $\frac{16}{\left(\frac{64}{22}\right)} = 5.5$ bits per sample. This is lossless, since we are discarding information.

2. Consider the following sampling cases, complete the table entries.

<table>
<tr><th>Waveform and Sample</th><th>Waveform Frequency (Hz)</th><th>Sample Frequency (Hz)</th><th>Inferred Frequency (Hz) from Samples</th><th>Properly Sampled?</th></tr>
<tr><td>
</td><td>2000</td><td>2500</td><td>500</td><td>no</td></tr>
<tr><td>
</td><td>500</td><td>10,000</td><td>500</td><td>yes</td></tr>
</table>

3. While the cat can hear up to 64KHz, it likely has similar critical band limitations to humans. Consider the hypothetical band structure shown on the facing page, and make the simplifying assumption that we only need to represent the strongest 4 frequencies in each band over a 25 ms time window to 4 Hz resolution. Assume 16b amplitude quantization for each frequency. Let's determine efficient frequency-band representations for discernable cat audio.

   (a) Exploiting this structure, what do you store for each 25 ms window and how many bits does this require?

   Store 4 (frequency, amplitude) pairs in each band, for a total of $4 \times 30 = 120$ frequencies.

   Since we only want 4 Hz resolution, we need $\log_2\left(\frac{64{,}000}{4}\right) =$ 14b to represent each frequency.

   Total: $120 \times (14 + 16) = 3600$

   The above is all we expected for a midterm solution. In practice, since we know we are always saving 4 frequencies per band, we can code the frequencies in each band separately. This can use fewer bits. For example, the first band covers 100-45=55 Hz, so we only need to distinguish $\frac{55}{4} = 14$ different frequencies, which we could do with 4b rather than 16b. The second band only needs to distinguish 25 frequencies, so we can use 5b. So, we actually get something more like:

$$F_{bits} = 4\sum_{1}^{30} \log_2\left(\frac{band_{top} - band_{bottom}}{4}\right) \quad (1)$$

   Then total bits is $120 \times 16 + F_{bits}$. However, since we are already storing 16b for amplitude and that does not change, over half of our encoding bits (1920) calculated above are still needed. This optimization, then, will, at most mean a factor of 2 savings over the simpler case calculated assuming 14b to represent all frequencies.

   (b) How much compression does this achieve over the PCM samples for the same 25 ms window? (State as a ratio.)

   $\frac{128000 \times 16 \times 0.025}{3600} = 14$

   (c) What kind of compression is this (lossless or lossy) and why?

   Lossy – we are discarding potentially masked frequencies. We will recover a waveform that may sound to the cat like the original, but is not identical to the original.

Hypothetical cat auditory critical bands:

Band Number	Low	High
1	45	100
2	100	200
3	200	300
4	300	400
5	400	500
6	500	600
7	600	800
8	800	1200
9	1200	1500
10	1500	2000
11	2000	2500
12	2500	3000
13	3000	4000
14	4000	5000
15	5000	6000
16	7000	8500
17	8500	10000
18	10000	12000
19	12000	15000
20	15000	18000
21	18000	22000
22	22000	25000
23	25000	30000
24	30000	35000
25	35000	42000
26	42000	46000
27	46000	50000
28	50000	56000
29	56000	60000
30	60000	64000

While the cat auditory range to 64,000 Hz is real. This auditory band structure is a synthetic construct generated just for this problem and likely does not represent reality.

4. Consider the following set of sounds that occur **simultaneously**. Identify which sounds a human is likely to hear and explain why or why not. (read through all 4 sounds before starting your answers.)

   (a) Soft (25dB) cat purrs at 150 Hz

   (likely) heard. While this is soft, it is in the human auditory range and it is in a different band from the loud noise, so won't be masked.

   (b) Loud (80dB) mechanical squeak at 30 KHz

   Not heard. This is above the human auditory range.

   (c) Moderate (60dB) child yelling at 2200 Hz

   Heard. This is in the human auditory range and is the loudest sound in its critical band.

   (d) Soft (30dB) cat meow at 2100 Hz

   Not Heard. This is in the same critical band as the child yelling, but 30dB (1000×) lower amplitude. It will be masked by the child yelling.

5. Consider the following Claude Shannon quote:

```
Thus we may have knowledge of the past but cannot control it;
we may control the future but have no knowledge of it.
```

Including spaces and punctuation, the quote has 116 total symbols (and 26 unique symbols).

(a) If we give every symbol the same number of bits, how many bits do we need to encode each symbol?

$\lceil \log_2(26) \rceil = 5$

(b) For a variable length symbol encoding intended to minimize the encoded length, which symbol should get the fewest bits? (and why?)

The space, because it is the most frequently ocurring symbol in the quote.

(c) For a variable length symbol encodingintended to minimize the encoded length, which symbol could get the most bits? (and why?)

The captial T (or any of the characters that occurs only once: period, semi, p) can get the longest encoding because it (they) are the least frequently occuring (the most uncommon) symbols.

(d) Is variable length symbol encoding a lossy or lossless compression scheme? (why?)

Loessless. We can perfectly reconstruct the original message.

6. Given a signal with 3 frequencies, 250 Hz (amplitude 1/2), 1400 Hz (amplitude 1/4), and 8000 Hz (amplitude 1), that can be represented as sine waves starting at $t = 0$, what is the amplitude of the temporal waveform at each of the following time points?

Time	**Amplitude**
0.02 ms	0.90
1.00 ms	0.65
2.37 ms	-0.30

$0.5\times\sin(2\pi \times 250 \times t)+0.25\times\sin(2\pi \times 1400 \times t)+\sin(2\pi \times 8000 \times t)$

7. A curious engineer with several cats would like to know what her cats hear (are they reacting to sounds she cannot hear? do they talk to each other above the range of her hearing?) Making the same assumptions about cat hearing from Problems 1 and 3, describe how she could build a cat-sound translator that converts the full range of cat-audible sounds into human-audible sounds.

   Sample sound at 128KHz (as per Problem 1).

   Fourier Transform into frequency domain.

   Scale the frequency domain results into the human hearing range. Roughly divide each frequency spike by 3.

   It would also be fine to leave sound up to 22 KHz alone, and simply scale down the frequencies between 22 and 64 KHz down into the human audio range and add them in along with the original data from 0 to 22 KHz.

   Inverse Fourier Transform to reproduce an audio waveform for human consumption.

8. Audio forensics (think Sherlock Holmes, CSI) may not find conventional MP3 encodings as useful as standard Compact-Disc (44KHz sample rate) PCM encodings or even analog recordings. [Assume the analog recording is a perfect capture of the original continuous sound waveform at the point of the recording microphone.]

   (a) Of the three, which would capture a high pitched sound (e.g. 30KHz)? For each, why or why not?

      i. MP3 encoded
      No. This is above human hearing and would be filtered out before sampling when constructing an MP3.

      ii. PCM encoded with 44KHz samples, anti-alias input filtering
      No. This is above the Nyquist frequency of 22 KHz for 44 KHz sampling. Proper anti-aliasing would filter out all frequencies above 22 KHz before sampling.

      iii. analog recording
      Yes. We are assuming the analog recording is perfect and would properly capture these high frequencies.

   (b) Of the three, which might be able to recover a whispered conversation happening further from the microphone than a loud conversation? For each, why or why not?

      i. MP3 encoded
      Unlikely. The whispered conversation likely occurs in the same critical bands as the loud conversation and will be mostly masked by the loud conversation.

      ii. PCM encoded with 44KHz samples, anti-alias input filtering
      Yes. We can convert to the frequency domain and identify the loud conversation by its high amplitude frequency components. We could then subtract out those frequencies and amplify the remaining frequencies to bring out the whispered conversation.

      iii. analog recording
      Yes, we can use a similar strategy as above for the PCM samples.

   (c) Assuming you are willing to store information at a variable data rate (unlike MP3s), want to preserve any potential signal up to 120 KHz, but also want to minimize bits used (equivalently, maximize the time period you can store on 1TB of storage), describe a sampling, transformation, compression, and storage strategy.

   Anti-alias filter at 120 KHz. Sample at 240 KHz.

   Perform Fourier Transform of time windows to convert into frequency domain for each time window.

Only keep and store (frequency, amplitude) pairs for non-zero amplitude frequencies that exist in each time window. Since only a few frequencies will typically be present (particularly during times when nothing is happening), we get our compression from only storing the present frequencies. However, when there is a large amount of sound occurring, we store more frequencies and hence more data so all the sounds are available.
Run a lossless compression (like Huffman) over the (frequency, amplitude) pairs that you need to store. During periods of low activity, we will likely only be recording the same, repetative frequencies. These are likely to be compressible since they are common.

Human auditory critical bands:

Band Number	Low	High
1	20	100
2	100	200
3	200	300
4	300	400
5	400	510
6	510	630
7	630	720
8	720	920
9	920	1080
10	1080	1370
11	1270	1480
12	1480	1720
13	1720	2000
14	2000	2320
15	2320	2700
16	2700	3150
17	3150	3700
18	3700	4400
19	4400	5300
20	5300	6400
21	6400	7700
22	7700	9500
23	9500	12000
24	12000	15500

Code of Academic Integrity

Since the University is an academic community, its fundamental purpose is the pursuit of knowledge. Essential to the success of this educational mission is a commitment to the principles of academic integrity. Every member of the University community is responsible for upholding the highest standards of honesty at all times. Students, as members of the community, are also responsible for adhering to the principles and spirit of the following Code of Academic Integrity.*

Academic Dishonesty Definitions

Activities that have the effect or intention of interfering with education, pursuit of knowledge, or fair evaluation of a students performance are prohibited. Examples of such activities include but are not limited to the following definitions:

A. Cheating Using or attempting to use unauthorized assistance, material, or study aids in examinations or other academic work or preventing, or attempting to prevent, another from using authorized assistance, material, or study aids. Example: using a cheat sheet in a quiz or exam, altering a graded exam and resubmitting it for a better grade, etc.

B. Plagiarism Using the ideas, data, or language of another without specific or proper acknowledgment. Example: copying another persons paper, article, or computer work and submitting it for an assignment, cloning someone elses ideas without attribution, failing to use quotation marks where appropriate, etc.

C. Fabrication Submitting contrived or altered information in any academic exercise. Example: making up data for an experiment, fudging data, citing nonexistent articles, contriving sources, etc.

D. Multiple Submissions Multiple submissions: submitting, without prior permission, any work submitted to fulfill another academic requirement.

E. Misrepresentation of academic records Misrepresentation of academic records: misrepresenting or tampering with or attempting to tamper with any portion of a students transcripts or academic record, either before or after coming to the University of Pennsylvania. Example: forging a change of grade slip, tampering with computer records, falsifying academic information on ones resume, etc.

F. Facilitating Academic Dishonesty Knowingly helping or attempting to help another violate any provision of the Code. Example: working together on a take-home exam, etc.

G. Unfair Advantage Attempting to gain unauthorized advantage over fellow students in an academic exercise. Example: gaining or providing unauthorized access to examination materials, obstructing or interfering with another students efforts in an academic exercise, lying about a need for an extension for an exam or paper, continuing to write even when time is up during an exam, destroying or keeping library materials for ones own use., etc.

* If a student is unsure whether his action(s) constitute a violation of the Code of Academic Integrity, then it is that students responsibility to consult with the instructor to clarify any ambiguities.