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

ESE150, Spring 2018 — Midterm — Wednesday, February 28

- Exam ends at 5:50PM; begin as instructed (target 4:35PM)
- 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	5	6	7	8	9	10	Total
10	10	10	10	10	10	10	10	10	10	100

Average: 79, Std. Dev.: 18

1. Time-Domain data samples:

   (a) How many bits are required to encode a single (mono) track of 4 minutes of 44KHz sample audio with 16b samples? [4 points]

   $4 \times 60 \times 44,000 \times 16 = 170,000,000 = 1.7 \times 10^8$

   (b) If we reduce the sample rate to 32KHz, how much will we also need to reduce the per sample quantization to halve the bits required for encoding? [3 points]

   $\frac{32,000}{44,000} \times \alpha = \frac{1}{2}$

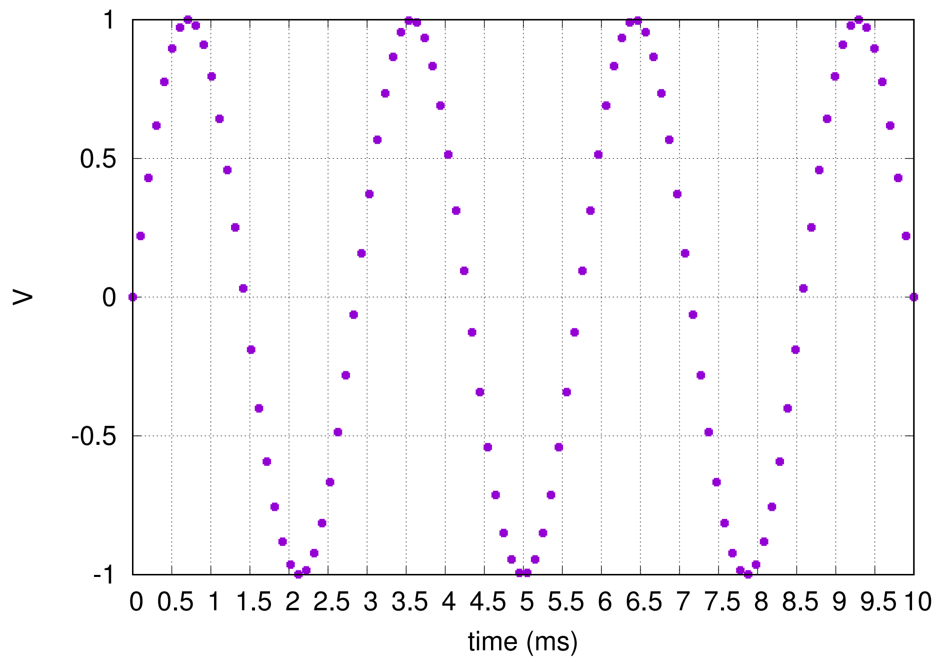
   $\alpha = 0.69$, or quantization must be reduced to $\alpha \times 16 = 11$ bits

   (c) What kind of compression is this and why? [3 points]

   Lossy. We are discarding information in the quantization (losing the distinction among 2^5 different values and losing the ability to accurately capture frequencies between 16KHz and 22KHz.

   -1 if not identify what information is lost

2. For the following samples of a sine wave:



(a) What is the frequency of the sine wave? [5 points]
350Hz
-1 for approximating frequency
-1 for misinterpretting and getting wrong magnitude (e.g. ms vs. s)

(b) What is the sample rate? [5 points]
10KHz (from 10 samples per ms)
-1 for error in last step of calculation
-1 for misinterpretting and getting wrong magnitude (e.g. ms vs. s)

3. Sample period and frequencies:
   The time for analogRead on the arduino is 125μs.

   - Based on this, what is the upper bound for the achievable sample rate on the Arduino? [3 points]
     $\frac{1}{0.000125}$=8KHz

   - What is the upper bound on the highest frequency the Ardunio sampling can accurately capture? [4 points]
     4KHz – Nyquist frequency = half the sample rate

   - Assuming no external filtering, what happens to a 6000 Hz tone? [3 points]
     It becomes aliased to a lower frequency. Specifically, it will show up as a 2KHz tone.
     -1 if get aliasing but not identify where/how the aliased signal ends up.

4. Categorize the following as lossy or lossless:

   (a) storing (frequency, amplitude, phase) triples for the non-zero frequency elements [2 points]

   lossless – can reconstruct by infering the non-stored cases are zero

   (b) starting with 16b time-sampled data, and converting to recording of (time,new amplitude) when changes occur [2 points]

   lossless – can reconsruct the waveform holding the value constant between changes

   (c) starting with 16b samples, add adjacent sample pairs and storing a single 17b value for each original pair of 16b samples [2 points]

   lossy – not enough information to uniquely restore the two original pairs

   (d) starting with 16b time-sampled data and converting to store, for each sample, the difference relative to the value of the previous time sample. [2 points]

   lossless – can recover the original by integrated the values (summing together the sequene of changes)

   (e) reporting all answers to 2 significant figures [2 points]

   lossy – cannot recover lower order figures

5. Which encoding uses the fewest bits to encode this quote and why?

t	h	e		b	e	s	t		i	s		y	e	t		t	o		c	o	m	e

[Correct choice – 5 points; Mention common case short encdoing – 5 points]

C – make common case inexpensive; C gives the commonly occurring symbols short encodings, while allowing less common symbols to take have longer encodings. (B also makes more frequent cases shorter, but doesn't optimally assign lengths based on frequency; it gives some symbols too short an encoding, forcing too many symbols to be longer than necessary. Technically, C uses encodings close to the Shannon optimal length of $-\log(p)$.)

A: $4 \times 23 = 92$

B: $1 \times 5 + 10 \times 1 + 9 \times 1 + 3 \times 4 + 7 \times 1 + 6 \times 1 + 10 \times 1 + 5 \times 2 + 4 \times 2 + 2 \times 4 + 10 \times 1 = 95$

C: $2 \times 5 + 5 \times 1 + 5 \times 1 + 3 \times 4 + 5 \times 1 + 5 \times 1 + 5 \times 1 + 3 \times 2 + 4 \times 2 + 2 \times 4 + 5 \times 1 = 74$

D: $5 \times 5 + 3 \times 1 + 3 \times 1 + 5 \times 4 + 3 \times 1 + 3 \times 1 + 3 \times 1 + 5 \times 2 + 5 \times 2 + 5 \times 4 + 3 \times 1 = 103$

symbol	A	B	C	D
(space)	0000	0	01	11111
b	0001	00000000001	10100	000
c	0010	000000001	10101	001
e	0011	001	111	11101
h	0100	00000001	11010	010
i	0101	0000001	11011	011
m	0110	0000000001	11000	100
o	0111	00001	100	1100
s	1000	0001	1011	1101
t	1001	01	00	11100
y	1011	0000000000	11001	101

As given had two identical encodings. One should have had another 0 as shown. Did not effect choice of correct answer. You could either ignore the fact two were same and just trust lengths, or you could make one of those longer.

6. Given: $f(t) = 0.5\cos(2\pi \cdot 800t) + \sin(2\pi \cdot 1000t)$ give the first 5 time-sample values of $f(t)$ for a 4KHz sample rate. [per sample 2 points]

sample	value
0	$0.5 \cdot \cos(2\pi \cdot 800 \cdot 0.00025 \cdot 0) + \sin(2\pi \cdot 1000 \cdot 0.00025 \cdot 0)$ =0.5+0=0.5
1	$0.5 \cdot \cos(2\pi \cdot 800 \cdot 0.00025 \cdot 1) + \sin(2\pi \cdot 1000 \cdot 0.00025 \cdot 1)$ =0.15+1=1.15
2	$0.5 \cdot \cos(2\pi \cdot 800 \cdot 0.00025 \cdot 2) + \sin(2\pi \cdot 1000 \cdot 0.00025 \cdot 2)$ =-0.40+0=-0.40
3	$0.5 \cdot \cos(2\pi \cdot 800 \cdot 0.00025 \cdot 3) + \sin(2\pi \cdot 1000 \cdot 0.00025 \cdot 3)$ =-0.4-1=-1.4
4	$0.5 \cdot \cos(2\pi \cdot 800 \cdot 0.00025 \cdot 4) + \sin(2\pi \cdot 1000 \cdot 0.00025 \cdot 4)$ = 0.15+0=0.15

partial credit if show calculation and one of two components is correct.

7. Sound Perception

   (a) Assuming the following frequency components exist simultaneously, which has the least effect on perceived sound quality and why? [5 points]

      i. amplitude 1, frequency 1500
      ii. amplitude 0.3, frequency 1400
      iii. amplitude 0.3, frequency 1600

   1600 Hz tone. This is in the same critical band as the dominant, 1500 Hz so likely to be masked. The 1400 Hz tone is in a different critical band, so will not be masked.

   (b) Assuming the following tones all occur at 40dB, which will sound the loudest? [5 points]

      i. 100 Hz
      ii. 1,000 Hz
      iii. 10,000 Hz

   1,000Hz – human hearing is most sensitive here. The lower and higher frequencies will be perceived as less loud.

8. In music video games (e.g., RockBand, Karaoke, or Guitar Hero), a singer earns points by matching the tune for a lyric track.

   (a) Using what you know from this course, how can the game process recorded sound input to identify how well the singer is performing? (quality of singing = ability to sing the right notes at the right time) [6 points]

   Perform DFT on each time window to identify the frequencies at each point in time. Score based on match of frequency at time.

   -2 for trying to do this in time domain

   -1 for only trying to match to critical band

   (b) The singer will typically be singing along with background instrument tracks played by the game. The sound from these tracks will also be picked up by the microphone into which the singer sings. How can the game cope with the composite sound that includes both the background instruments and the singers input? [4 points]

   Subtract out the expected frequencies and amplitudes for the instruments. Ideally, design the accompanyment so the instrument frequencies do not overlap with the intended vocal frequencies, so subtracting them out will not effect the singer frequencies.

   -1 if simply assume voice is louder than instruments

9. Compare an SMS text message to cell phone audio.

   - Assume a single SMS text message is 160 characters, where each character is 8b ASCII.
   - Assume the 160 character message is equivalent to 2 seconds of spoken sound.
   - Telephone quality audio is 8b samples at 8KHz

   (a) How much more compact is the SMS text message (ratio of bits required)? [6 points]

   $\frac{2\times 8000\times 8}{8\times 160} = 100$; The SMS text message uses only 1% of the bits required by the telephone quality audio sample.

   (b) What information is lost when you substitute the SMS text message for the cell phone audio? [4 points]

   Timing, speed, pauses; voice (speaker recognition); emotional intent (happy, sad, angry, humorous, ...)

10. Two of your friend both recorded a live historic speech (e.g., Jason Kelce in front of the Art Museum earlier this month?). During a key point in the speech, the person next to them yells loudly (40dB above the speech) around 1500 Hz.

- One friend is recording raw, PCM samples at CD-quality (16b, 44KHz)
- The other is recording directly to an MP3

(a) For the CD-quality recording [5 points]

i. Can you repair it? (remove the loud noise so listeners can hear the entire speech, including the key point when the person is yelling) yes

ii. If not, why not? If you can, outline how? Take DFT; identify frequency components for the yell and subtract them out. Leave the rest of the frequencies, particularly the speech, alone. Take inverse-DFT to convert back to PCM samples if appropriate (or store as DFT samples).
-2 if not say how subtract out; -1 if try do in analog level rather than frequency domain

(b) For the MP3-encoded recording [5 points]

i. Can you repair it? (see above) no (or maybe, but not as well)

ii. If not, why not? If you can, outline how? No: During MP3 encoding, the yell will mask softer sound in the band. The MP3 encoder will remove these other frequencies during encoding since they cannot be heard. The information is lost.
Maybe: You could do the same thing as for the CD-quality encoding – remove the yell frequencies. You would only lose the frequencies in the same critical band as the yell. So, it should result in a *better* recording, but may lose sounds than the CD-quality recording does not lose.
-2 if not identify masking

This page intentionally left mostly blank for pagination.
Feel free to use for work space.

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.