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

ESE150, Spring 2020 — Midterm — Wednesday, March 4

- 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.

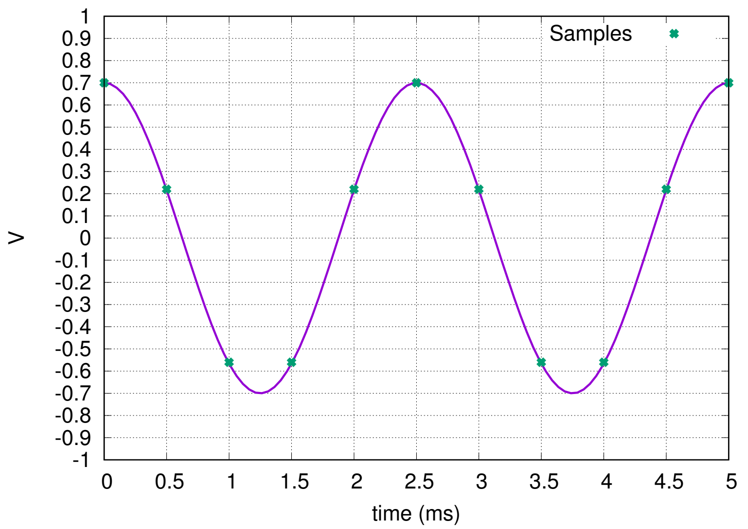
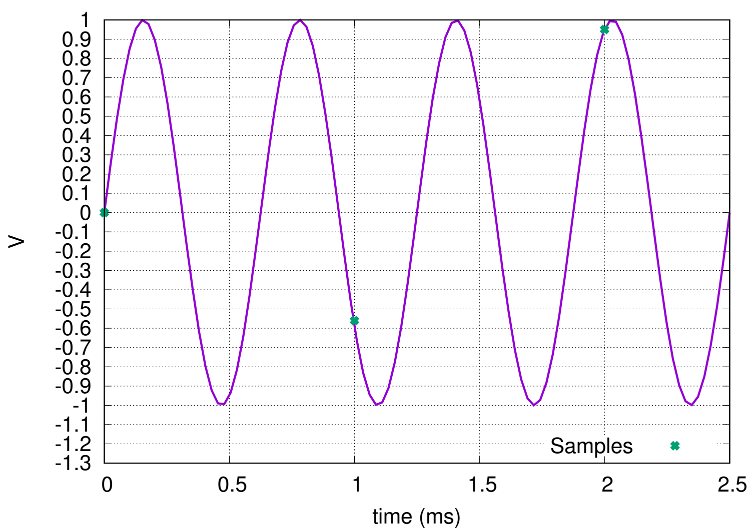
Name: Solution

1	2			3					4	5			
	a	b	c	a	b	c	d	e		a	b	c.i	c.ii
10	4	3	3	2	2	2	2	2	10	2	3	2	3

6	7				8					Total
	a	b	c	d	a	b	c	d		
10	5	5	5	5	6	4	3	7		100

Average: 74

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

Waveform and Sample	Waveform Frequency (Hz)	Sample Frequency (Hz)	Inferred Frequency (Hz) from Samples	Properly Sampled?
	400 Hz	2000 Hz	400 Hz	Y
	1600 Hz	1000 Hz	400 Hz	N

2. Telephone digital voice uses PCM encoding with 8KHz sample rate and 8b amplitude quantization.

   (a) What is the maximum frequency this sample rate can accurately capture?

   4K Hz – Nyquist frequency is one half the sampling rate.

   (b) What is the sample period (length of time between samples)?

   $\frac{1}{8000\text{Hz}} = 0.125\,\text{ms}$

   -1 using 1/4000 instead of 1/8000

   -1 misplace decimal, but otherwise correct

   (c) How many bits does this scheme require to record a 15 second voice-mail message?

   $8\text{K} \times 8\text{b} \times 15 = 960{,}000$

   -2 using 4K instead of 8K

3. Consider the following quote from Lt. Cmd. Vindman:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T	h	i	s		i	s		A	m	e	r	i	c	a	.	

17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
H	e	r	e	,		r	i	g	h	t		m	a	t	t	e	r	s	.

This has 37 symbols from a set of 16 unique symbols.

symbol	A	H	T	.	,		a	c	e	g	h	i	m	r	s	t	**sum**
count	1	1	1	2	1	5	2	1	4	1	2	4	2	4	3	3	37

(a) How many bits to encode this using a uniform encoding where each symbol is encoded using the same number of bits?

$37 \times \log_2(16) = 148$

(b) Is this likely to be compressible with a variable-length symbol encoding? Why? Illustrate with specific symbols.

Yes. Some symbols (like space and 'e') occur much more frequently than others (like ',' and 'c'). We can assign short encodings to the common cases and longer encoding to the uncommon ones.

(c) What is the Shannon Entropy lower bound for for this quote?

$$\text{Lower Bound} = -\sum_i \log_2(P(c[i])) \qquad (1)$$

Hint: there are only 5 different counts, so 5 different P(c[i]) values to calculate.

No one got full credit here. Problem should have been clearer that c[] is the 37 character input array. So this is formulated to add the lower bound encoding for each character.

$$-\log_2\left(P(c[0])\right)-\log_2\left(P(c[1])\right)-\log_2\left(P(c[2])\right)-\log_2\left(P(c[3])\right)-\log_2\left(P(c[4])\right)\ldots$$

$$=-\log_2\left(P('T')\right)-\log_2\left(P('h')\right)-\log_2\left(P('i')\right)-\log_2\left(P('s')\right)-\log_2\left(P('\ \ ')\right)\ldots$$

$$=-\log_2\left(\frac{1}{37}\right)-\log_2\left(\frac{2}{37}\right)-\log_2\left(\frac{4}{37}\right)-\log_2\left(\frac{3}{37}\right)-\log_2\left(\frac{5}{37}\right)\ldots$$

$$=-(5)\log_2\left(\frac{5}{37}\right)-(3\times 4)\log_2\left(\frac{4}{37}\right)-(2\times 3)\log_2\left(\frac{3}{37}\right)-(4\times 2)\log_2\left(\frac{2}{37}\right)-(6\times 1)\log_2\left(\frac{1}{37}\right)$$

$$=139.6\approx 140$$

(d) Consider the following set of variable-length binary encodings. Assign each symbol to an encoding to minimize the encoded length.

encode	000	0010	0011	0100	01010	01011	0110	0111
symbol	spc	.	a	h	A	H	m	t
encode	10000	10001	1001	101	11000	11001	1101	111
symbol	T	,	i	e	c	g	s	r

There are several equivalent assignments.

Space, being the most frequent, must get one of the 3b encodings. Two of {e,i,r}, which each have 4 occurrences, should get the other two. There are seven, 5b encodings; these go to the one occurrence symbols, of which there are also seven. This leaves the 4b encodings for the remaining symbols which have occurrences of 2–4.

(e) For the above assignment, how many bits are required to encode the quote?

$(5+4+4)\times 3+(2+2+2+2+3+3+4)\times 4+(1+1+1+1+1+1+1)\times 5=141$

4. While watching a movie, which sounds are likely to annoy you (interfere with your preception and enjoyment...) during each of two scenes: (a) piano recital (25–4200 Hz) at 40–80dB, (b) 100dB explosion (broad spectrum 20–8,000 Hz) ?

   Classify each as: (U)nnoticeable, (H)ighly annoying, (L)ightly annoying

	piano	explosion
70dB, 45KHz squeak	U	U
60dB, 2-3 KHz child cry	H	U
20dB, 100-300 Hz whisper	L	U
40dB, 5Hz mechanical vibration	U	U
50dB, 11KHz electronic whine	H	H

   - 45KHz is above the highest band for human hearing
   - The child crying is in sensitive auditory bands and likely to be louder than the piano notes in those bands at some points during the piano recital.
   - The explosion is broad spectrum with louder components than the child in the bands in which the cry occurs. As such, it likely masks the child cry.
   - The whisper is low volume, but might be audible during periods when no notes occur in the identified speech bands.
   - The whisper is masked by the loud, broad-spectrum explosion.
   - 5Hz is below the lowest band for human hearing.
   - The 11KHz noise is within the human auditory band, and above the bands for either the piano or explosion, making it audible in both cases.

5. Assume for simplicity in this problem that each band has 30 discernable frequencies. We allocate 180b to encode each band. Assume 16b quantization is essentially perfect (zero error). The error for a frequency is a value between 0 and 1 equal to

$$Error(f) = \frac{|\text{ActualFrequency} - \text{EncodedFrequency}|}{2^{16}} \quad (2)$$

Encoding to k-bit quantization means an encoded value of e will be interpreted as an EncodedFrequency of $e \times 2^{16-k}$.

(a) If we assign equal quantization to every frequency in the band, how many bits can we use to represent the amplitude of each frequency?

$$\frac{180}{30} = 6$$

(b) Ignoring masking, what is the maximum possible total error across the entire band? (sum up the errors across all frequencies in the band.)

Maximum difference is $2^{16-6}/2 = 2^9$; divide by two because can round up or down.

$30 \times \frac{2^9}{2^{16}} = 0.23$

(c) Assuming anything with amplitude 20% below the maximum frequency in the band is masked and contributes no subjective error. Assume frequencies within the 20% bound contribute a subjective error equal to the full Error(f) stated above. Assume for simplicity omitted frequencies do not cost encoding bits.

i. What is the maximum number of loud (amplitude within 20% of maximum amplitude frequency in the band) frequencies that a band can have and achieve a zero subjective error encoding?

$\frac{180}{16} = 11$

ii. What is the maximum possible subjective error when there are 15 loud frequencies (amplitude within 20% of maximum amplitude frequency in the band) and you assign equal quantization to these loud frequencies.

$\frac{180}{15} = 12\text{b each}$

$30 \times \frac{2^{16-12-1}}{2^{16}} = 0.0037$

6. Given: $f(t) = 0.3\cos(2\pi \cdot 400t) + \sin(2\pi \cdot 600t) + 0.2\sin(2\pi \cdot 700t)$
give the second, tenth, and twenty-third time-sample values of $f(t)$ for a 4KHz sample rate.

sample	2	10	23
time (t)	0.5 ms	2.5 ms	5.75 ms
value $f(t)$	1.2	0.10	0.25
$\cos(2\pi \cdot 400t)$	0.31	1.0	-0.31
$\sin(2\pi \cdot 600t)$	0.95	0.0	0.31
$\sin(2\pi \cdot 700t)$	0.81	-1.0	0.16

You only need to complete the **value** row. Other rows are likely useful to assembling your solution and showing your work for partial credit.

reasonable to treat first sample as 0, so times of 0.25 ms, 2.25 ms, 5.5 ms

-2 if one value wrong

-3 if looks like mistakenly computed sin/cos on degrees for values that are actually in radians.

7. Early telephones used DTMF (dual-tone, multiple frequency)-signalling to send signals including phone number digits. They used pairs of frequencies (the dual tone) to represent each symbol. The table below shows how pairs formed from 2 sets of 4 frequencies were used to represent 16 symbols:

	1209 Hz	**1336 Hz**	**1477 Hz**	**1633 Hz**
697 Hz	**1**	**2**	**3**	**A**
770 Hz	**4**	**5**	**6**	**B**
852 Hz	**7**	**8**	**9**	**C**
941 Hz	*	**0**	**#**	**D**

Assume a tone must be present for at least 100 ms to be registered as a symbol and an absence of dual tone must occur for at least 20 ms to denote the separation between one symbol and the next.

(a) How would DTMF symbol detection and extraction work for a modern time-sampled digital system listening to an analog line from an analog phone producing DTMF signals?

- Sample in time.
- Use Fourier Transform to convert to frequencies.
- (use 10 ms windows; so guaranteed to get an empty window between DTMF symbols.)
- Look for DTMF signal pairs in the frequency conversion.

-1 point for not mention window size or use too large of a window size.

(b) If we just wanted a system to decode DTMF signals:

i. What is the minimum sample rate required for our time-sampled digital system?

$2 \times 1633 \approx 3300$ Hz

Nyquist Sampling Rate—twice the maximum frequency component trying to capture.

ii. What analog filtering would be needed to support operation at this sample rate?

Low pass filter to remove frequencies above 1633 Hz to prevent aliasing.

-1 point for not being specific about filtering frequency.

(c) If you had an MP3 of a phone conversation where someone inadvertantly pushed a button on the phone that generated one or more DTMF signals, how could you clean it up? (produce a better MP3 that removed the audible DTMF tones)

- Look for DTMF frequency pairs in audio frames.
- If pair is present in a frame, remove (zero out) DTMF frequencies.
- Save out modified MP3.

The MP3 already represents things in frequency. -1 point for stating need to convert into (and out of) frequency.

(d) Assume you can place a digital processor between the microphone and analog output line on a phone (and similarly between the analog input line and the speaker). For legacy signalling over an analog line, how could you insert DTMF tones on the originating end and remove them at the destination so that the humans never heard them?

- Insert
  - Capture with A2D
  - Sample in time
  - Perform inverse Fourier on the DTMF tone pair to create time samples for DTMF tones
  - Add DTMF tone samples into time samples
  - Render time samples back to analog through D2A
- Remove
  - Capture with A2D
  - Sample in time
  - Perform Fourier Transform on time samples windows
  - Look for DTMF signal in each window
  - If present, note the symbol and remove the frequency pair from the frequency represenation
  - Perform inverse Fourier Transform on each window to convert back to time samples
  - Render time samples back to analog through D2A

8. Corporate sonic branding creates short audio snippets intended to identify the specific brand. For simplicity let's model the space of potential sounds as composed of four 100 ms chords, where each chord is composed of up to 3 frequencies.

   - Assume at most 30 distinguishable frequencies per critical band.
   - Restrict to the 20 bands from 2–21 to assure most people can hear the frequencies.

   (a) Assuming we want to avoid masking in the design of the brand snippet, what restrictions should we place on the frequencies composing each chord (composition of 3 frequencies composed in each of the four 100 ms segments)?

   Use a single frequency per critical band.

   (b) Given this restriction, and assuming you don't want to differentiate brand snippets only by the amplitude of frequencies, how many potential 400 ms audio snippets are there?

   For each 100 ms chord, we can choose 3 bands, hence the $\binom{20}{3}$. Within each of the 3 bands, there are 30 frequencies to choose from. Each 100 ms chord is independent, so we have the same range of choices for each of the four.

   $$\left(\left(\binom{20}{3}\right)30^3\right)^4 \approx 9.0 \times 10^{29}$$

   Technically, we can consider chords of 2, 1, or 0 notes, but that doesn't change the magnitude significantly.

   $$\left(\left(\binom{20}{3}\right)30^3 + \left(\binom{20}{2}\right)30^2 + \left(\binom{20}{1}\right)30 + 1\right)^4 \approx 9.2 \times 10^{29}$$

   (c) How many bits of information does this represent?

   $$\log_2\left(9.0 \times 10^{29}\right) \approx 100$$

   (d) Assuming 25 ms sample windows, 4b amplitude quantization, and that you can model the MP3 as encoding the present frequencies and their associated amplitudes in each 25 ms window, how large (in bits) will the MP3s be to respresent these audio snippets?

   Each window has 3 frequencies.
   Each frequency might need 15b (could be 11b, if think about 1152 sample windows).
   Each frequency pair takes about 15b+4b=19b.
   16, 25 ms windows.
   $19 \times 3 \times 16 = 912$b

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 student's 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 person's paper, article, or computer work and submitting it for an assignment, cloning someone else's 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 student's 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 one's 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 student's 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 one's 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 student's responsibility to consult with the instructor to clarify any ambiguities.