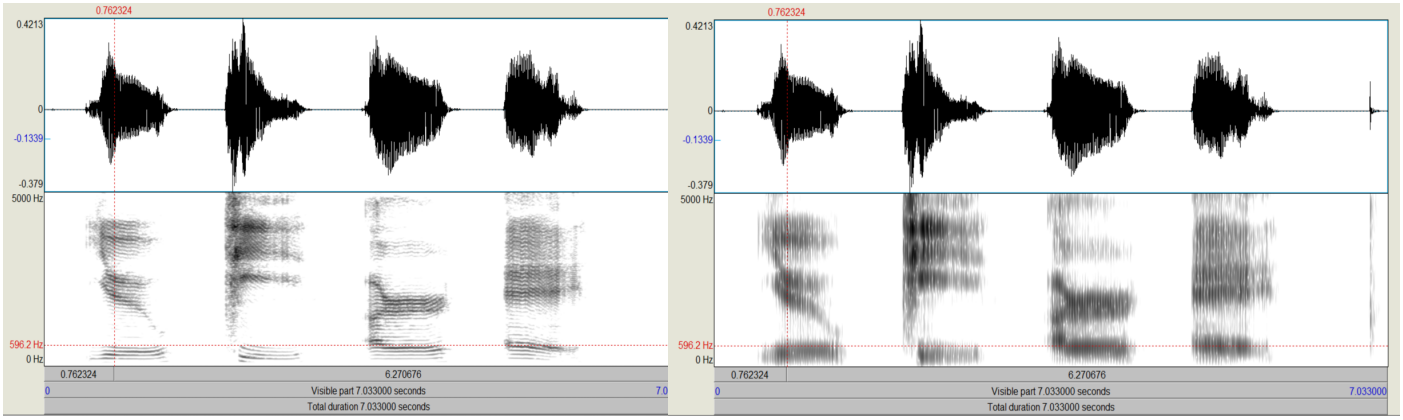
UCLA — Electrical and Computer Engineering Dept.
ECE114: Speech and Image Processing — Take-home Speech Exam
Due November 6, 2020 at 4pm (pacific standard time)

This exam has 4 questions, for a total of 100 points.

Open book. Calculators allowed, but you must show all work. When asked to explain your reasoning, please give a TYPED explanation. Full credit will not be given without proper justification where asked.
Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page.

Question	Points	Score
1	30	
2	24	
3	30	
4	16	
Total:	100	

Please go to next page...



(a) One spectrogram. **(b)** Another spectrogram.

Figure 1: Two spectrograms

1. Fig. 1 shows two spectrograms taken using different window lengths of a person saying "UCLA." However, the person misspoke, and two of the phonemes are incorrect.
   (a) (4 points) Transcribe the phrase "UCLA" in the Arpabet.
   (b) (6 points) Mark the time regions that correspond to each phoneme in the spectrogram. You can do this by drawing vertical lines on the spectrogram that segment it into regions containing one phoneme each.
   (c) (6 points) Identify each region you drew as voiced or unvoiced.
   (d) (4 points) Which phonemes were said incorrectly? How do you know?
   (e) (4 points) Which spectrogram is a wideband and which is a narrowband spectrogram? How do you know?
   (f) (6 points) Identify the pitch of the speaker? Which spectrogram did you use, and how did you identify the pitch?

2. Construct a filter to approximate the vocal tract transfer function of an /i/ sound.

   (a) (2 points) List the first three formants of an /i/ sound.

   (b) (6 points) Assuming that we sample at 44 100 Hz, write a stable transfer function that has poles at the locations of these formants. You may select any reasonable magnitude for the poles.

   (c) (6 points) Draw a zero pole diagram for the system.

   (d) (10 points) What is the smallest filter order possible such that if this vocal tract impulse response, $v(n)$, were convolved with a glottal pulse signal, $x(n)$, from someone with a pitch of 150 Hz, the pitch harmonics would be resolved. Assume that the filter order is the size of the DFT of $x(n)$.

3. Consider the three-tube model of Fig. 2 (the asterisk indicates the position of the vocal cords). Assume that the length of the vocal tract is $l = l_1 + l_2 + l_3 = 17.5\,\text{cm}$ and the speed of sound is $c = 360\,\text{m/s}$. The cross-sectional areas are assumed to be $A_1 = A_3 = \pi d_1^2/4$ and $A_2 = \pi d_2^2/4$.

Figure 2: Three-tube model for vowels.

(a) (10 points) Compute the nomogram for $0 < l_1 < 12.5\,\text{cm}$, $l_2 = 5\,\text{cm}$, with the choice of $d_1 = 3.4\,\text{cm}$ and $d_2 = 1\,\text{cm}$. For this problem, compute the resonance frequencies F_1, F_2, and F_3 of the entire system for each value of l_1 from 0 to 12.5cm in increments of 0.5 cm. Make a table of the resonances from each decoupled tube. Clearly label the columns and rows of the table and submit it. Then plot the values in the table corresponding to each of the first three formants to form a nomogram as shown in class. You may find the min() function helpful for this. Label which curve corresponds to which tube and to which formant frequency in the plot. You may label it how you wish as long as it is clear. You may do this problem by hand or in a graphing program like MATLAB or Excel. You need not include any code, but you must explain your calculations and write out all equations used. We assume no acoustic coupling in this part of the problem.

(b) (5 points) On your plot, revise the nomogram taking into account the phenomenon of acoustic coupling. Draw (by hand is OK) the lines corresponding to where you expect the three formants F_1, F_2, and F_3 to be for all values of l_1. You do not need to be precise.

(c) (10 points) As you did in part a, plot another nomogram for $0 < l_1 < 12.5\,\text{cm}$, $l_2 = 5\,\text{cm}$, with the choice of $d_1 = 3.4\,\text{cm}$ and $d_2 = 2.4\,\text{cm}$. How is this nomogram different from what you drew before?

(d) (5 points) The formants of the vowels /a/, /o/, /u/, and /i/ can be estimated using this three-tube model with $d_2 = 1\,\text{cm}$ for values of $l_1 = 0.5\,\text{cm}, 2.5\,\text{cm}, 7.5\,\text{cm}, 9.1\,\text{cm}$, respectively. The vowel /e/ is obtained with the same parameters as the vowel /i/, but with a wider constriction, i.e., $d_2 = 2.4\,\text{cm}$. Based on these considerations, list the formants F_1 and F_2 for the vowels /a/, /o/, /u/, /i/, and /e/.

4. Consider the speech segment
$s(n) = [2, -1, \underline{1}, 3, -1, -2, -1, 1, 3, -1, -2, -1, 1, 3, -1, -2, -1, 2, 1, 1, 3, -2, 1, 2, -3, -1, 1]$. The correlation function is given by

$$R(i) = \sum_{n=i}^{N_{\mathrm{w}}-1} s^{\mathrm{w}}(n) s^{\mathrm{w}}(n-i), \quad 0 \le i \le N_{\mathrm{w}} - 1,$$

where $s^{\mathrm{w}}(n) = s(n)w(n)$, $w(n)$ is a rectangular window of length $N_{\mathrm{w}} = 15$. Let the model order be $p = 3$.

(a) (4 points) Compute $R(i)$, $i = 0, \ldots, p$

(b) (4 points) Find the 3nd-order prediction coefficients, a_1, a_2, and a_3, using the correlation method of linear prediction analysis.

(c) (4 points) Find the corresponding error, $E_{\min}$. Write the expression for the vocal tract's transfer function, $V(z)$.

(d) (4 points) Compute the expression for the poles of the corresponding vocal tract model. Are they real or complex conjugate? How many formants are there? Explain. Let the sampling frequency be $F_{\mathrm{s}} = 8000\,\mathrm{Hz}$.

Please go to next page...